A Data-Driven Perspective on ENSO Diversity

Impacts, Definition, and Forecasting

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Jakob Schlör, M. Sc.
aus Würzburg

Tübingen
2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 30.07.2024

Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Dr. Bedartha Goswami
2. Berichterstatter:	Prof. Dr. Philipp Hennig

Disclaimer: This thesis uses Federico Marotta's `kaobook` template based on Ken Arroyo Ohori's doctoral thesis and many stylistic features added by Frank Schneider and Felix Dangel. I would like to thank Jemma Jeffree, Jannik Thuemmel, and Felix Strnad for their help to improve the text.

Acknowledgments

I want to express my deepest gratitude to Bedartha Goswami for his guidance, which allowed me to transition into the field of climate science. Your scientific curiosity and meticulous attention to detail have been inspiring. I am grateful for your advice, support, and the opportunities to pursue my interests, including my visit to Boulder and my participation in various conferences.

Also, I am grateful to Philipp Hennig and Georg Martius for serving on my thesis committee and for providing valuable feedback and support throughout my journey. Thank you and Jakob Macke for being a part of my examination committee.

Great thanks to Sebastian Schwenk, Elena Siziliani, and Shyamala Subramanian (Cluster of Excellence - ML in Sciences), as well as Leila Masri and Sara Sorce (International Max Planck Research School for Intelligent Systems) for their help in administrative matters.

I'm deeply thankful to Antonietta Capotondi and Matt Newman at NOAA for their advice on climate-related topics and for the wonderful experience I had in Boulder. I really enjoyed working with the large team there, with special thanks to Kinya Toride, Rochelle Worsnop, Nicola Maher, and Jemma Joffrey.

This work was made possible by a fantastic team. I am grateful to Felix Strnad, Jannik Thuemmel, and Jan Prosi for our countless discussions in the office, over lunch and coffee, for your patience with me and your friendship. I extend these thanks to everyone on the 4th floor who contributed to a pleasant work environment. I am fortunate to have met all of you.

I enjoyed supervising Jakob Unterholzer, Benedict Roder, Miriam Raquel Rodriguez Carrasco, and Felix Bötte. Thank you for your contributions to this work. I hope you learned as much as I did.

During my time in Tübingen, I was fortunate to be surrounded by wonderful people. I want to express my gratitude to Leonie John, Corinna Schulz, Chris Rhode, Jan Lause, Auguste Schulz, Eric Miele, Florian Ebermeier, Francesco Chini, Katharina Anderer, and everyone else not mentioned here. You all made this an amazing time. Special thanks to Adrian Heilemann, Moritz Frankerl, Christoph Rohrmeier, and Alexander Neef for your support and all the trips and adventures we had both during and before my PhD journey.

I am thankful to my family, particularly my parents, my brother, and my grandparents, for sparking my curiosity in life, placing their trust in me, and consistently providing unconditional love and support throughout my journey.

Finally, I cannot thank Ronja Weissinger enough for always being there for me and being such a wonderful person. Thank you for making me laugh every day, and just everything. I hope for many more joyful moments we can share.

Thank you!

Jakob Schlör
Tübingen, Aug 4, 2024

Abstract

El Niño Southern Oscillation (ENSO) is the dominant mode of interannual variability of the global climate and is characterized by anomalously warm (El Niño) and cold (La Niña) sea surface temperatures (SST) in the tropical Pacific. El Niño and La Niña exhibit a large event-to-event variation in terms of temperature intensity, spatial pattern, and temporal evolution, known as ENSO diversity. ENSO diversity is commonly described by two distinct types — Eastern Pacific (EP) and Central Pacific (CP), based on the location of peak SST anomalies — exhibiting different impacts on weather conditions worldwide, also called teleconnections. While the coupled atmosphere-ocean feedback processes of ENSO are known, the mechanisms contributing to its diversity are not clear.

This thesis introduces data-driven approaches to model various aspects of ENSO diversity, assess ENSOs global impacts, refine its definition, and improve its forecasting accuracy. My contribution is three-fold:

i) I introduce a novel tool to visualize teleconnections of ENSO diversity worldwide, suggesting that EP El Niño events mainly impact surface temperatures in the tropics whereas CP El Niño events exhibit only minor impacts on temperature changes.

ii) Studying the impacts of El Niño events revealed inconsistencies between conventional definitions of ENSO diversity. Consequently, I propose that ENSO diversity should be defined as a continuous phenomenon, rather than the binary separation into CP and EP events. This perspective allows for a more nuanced estimation of onset dynamics and low-frequency changes of ENSO.

iii) I propose a hybrid model for ENSO forecasting, that exhibits skillful forecasts up to 18 months with uncertainty estimates. The combination of linear model and recurrent neural network is data efficient and enables interpretable analysis, highlighting potential mechanisms of ENSO diversity.

With anthropogenic climate change projected to intensify El Niño events, this work contributes to enhancing our understanding and predictive capabilities of ENSO diversity, which is crucial for agriculture, energy production, and disaster mitigation.

Zusammenfassung

El-Niño-Southern-Oscillation (ENSO) ist die größte Klimaveränderung von einem Jahr zum Nächsten. ENSO ist durch ungewöhnlich warme (El Niño) und kalte (La Niña) Meeresoberflächentemperaturen (SST) im tropischen Pazifik gekennzeichnet. Sowohl El Niño als auch La Niña zeigen große Unterschiede zwischen Ereignissen in ihrer Intensität und Temperaturverteilung im Pazifik auf, was als ENSO Diversität bezeichnet wird. Üblicherweise wird zwischen zwei Typen unterschieden: dem Eastern Pacific (EP) und dem Central Pacific (CP) Typ, je nachdem, ob die Temperaturanomalien im Ost- oder Zentralpazifik am größten sind. Die Unterschiede zwischen EP- und CP-Ereignissen betreffen nicht nur den Pazifik, sondern beeinflussen das globale Wettergeschehen. Diese langreichweitigen Auswirkungen auf andere Regionen werden als Telekonnektionen bezeichnet. Trotz umfassender Kenntnisse über die grundlegenden atmosphärischen und ozeanischen Prozesse von ENSO sind die Mechanismen, die zu den Unterschieden zwischen den Ereignissen führen, noch nicht vollständig geklärt. Klimamodelle sind derzeit nicht in der Lage, die ENSO-Diversität präzise abzubilden, was die Genauigkeit von Vorhersagen erheblich beeinträchtigt.

Diese Arbeit untersucht verschiedene Aspekte der ENSO-Diversität mithilfe datengetriebener Modelle:

i) Klimanetzwerke zeigen die Korrelationsstrukturen zwischen Stationen auf der Erde. Da vor allem benachbarte Punkte starke Korrelationen aufweisen, sind die meisten Verbindungen in einem Klimanetzwerk lokal. Wir schlagen eine Metrik vor, die basierend auf der Topologie des Klimanetzwerks Telekonnektionen identifiziert. Damit können wir Unterschiede in globalen Auswirkungen von EP- und CP-El Niño-Ereignissen aufzeigen.

ii) Die binäre Unterteilung von ENSO-Ereignissen in CP- und EP-Typen ist in vielerlei Hinsicht willkürlich, da die Verteilung der El Niño- und La Niña-Ereignisse eher kontinuierlicher Natur ist. Diese Unterteilung wurde eingeführt, um Eigenschaften mehrerer Ereignisse gemeinsam zu untersuchen. Statt einer diskreten Unterteilung in CP und EP, schlagen wir jedoch vor, die Ereignisse mit einer multimodalen Verteilung zu beschreiben, bei der bestimmte Kategorien häufiger auftreten als andere. So können wir El Niño- und La Niña-Jahre mit Wahrscheinlichkeiten für bestimmte Kategorien beschreiben und deren langfristige Veränderung über Jahrzehnte hinweg untersuchen.

iii) Zuletzt schlagen wir einen hybriden Ansatz zur ENSO-Vorhersage vor. Das hybride Modell besteht aus einem linearen Modell und einem rekurrenten neuronalen Netz und sagt nicht nur die Wasseroberflächentemperaturen im Pazifik bis zu 18 Monate im Voraus vorher, sondern ermöglicht es auch, die Rolle der nichtlinearen Dynamik im Pazifik zu untersuchen. Wir zeigen, dass Nichtlinearität vor allem im Westpazifik zu besseren Vorhersagen führt. Zudem können wir mit unserem hybriden Ansatz die Vorhersageunsicherheit quantifizieren und Bedingungen im Pazifik identifizieren, die zu einer höhere Vorhersagbarkeit führen.

Der anthropogene Klimawandel erhöht die Häufigkeit von El-Niño-Ereignissen und damit die Relevanz genauerer Vorhersagen und eines besseren Verständnisses von ENSO. Vor allem für die Planung von Wasserressourcen, Landwirtschaft und Energieerzeugung ist eine frühzeitige Vorhersage von ENSO unabdingbar.

List of publications

J. Schlör, M. Newman, J. Thuemmel, A. Capotondi, and B. Goswami (2024). 'A Hybrid Model for ENSO Dynamics in the Low-Data Regime'. *In Preparation*

R. P. Worsnop, M. Scheuerer, T. M. Hamill, T. Smith, and **J. Schlör** (2024). 'RUFCO: a deep-learning framework to post-process subseasonal precipitation accumulation forecasts'. *In Preparation*

K. Toride, M. Newman, A. Hoell, A. Capotondi, **J. Schlör**, and D. Amaya (2024). 'Using Deep Learning to Identify Initial Error Sensitivity of ENSO Forecasts'. arXiv:2404.15419

J. Schlör, F. Strnad, A. Capotondi, and B. Goswami (2024). 'Contribution of El Niño Southern Oscillation (ENSO) Diversity to Low-Frequency Changes in ENSO Variance'. *Geophysical Research Letters* 51(14), e2024GL10917

F. M. Strnad, **J. Schlör**, R. Geen, N. Boers, and B. Goswami (2023). 'Propagation Pathways of Indo-Pacific Rainfall Extremes Are Modulated by Pacific Sea Surface Temperatures. *Nature Communications* 14(5708), 1–16

F. M. Strnad*, **J. Schlör***, C. Fröhlich, and B. Goswami (2022). 'Teleconnection Patterns of Different El Niño Types Revealed by Climate Network Curvature'. *Geophysical Research Letters* 49(17), e2022GL098571

*: Authors contributed equally.

Contents

Overview 1.

1.1. Introduction

At the time of writing, global sea surface temperature (SST) are the highest on record. This unprecedented rise is largely driven by global warming, caused by the burning of fossil fuels, and has led to land temperatures being 1.5 degrees warmer than in pre-industrial times (Copernicus, 2024). Given that the ocean takes up 90% of heat associated with anthropogenic global warming, the rise in ocean temperatures is expected (Durack et al., 2014). Nevertheless, the jump in global ocean temperatures to two standard deviations higher than the 40-year average came as a surprise in the scientific community (Fig. 1.1).

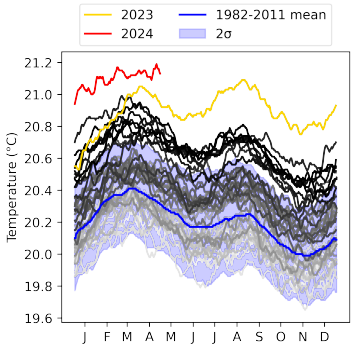


Figure 1.1.: Daily SST averaged globally (60°S-60°N) for the years 1982 to present. The current and last years are highlighted, as well as the mean and 2σ between 1982 and 2011. Data are taken from Reanalyzer (2024).

This unexpected ocean temperature jump highlights the interplay between anthropogenic climate change and natural climate variability. The largest contributor to natural interannual variability is El Niño Southern Oscillation (ENSO), which not only alters ocean temperatures but also the atmospheric conditions worldwide (Taschetto et al., 2020). ENSO is a coupled ocean-atmosphere phenomenon, that is characterized by periods of warm (El Niño) and cold (La Niña) sea surface temperature anomaly (SSTA) in the equatorial Pacific. The recent jump in global SST coincides with a transition to a warm El Niño phase in 2023-2024, after three consecutive years of cold La Niña conditions.

The recent El Niño has amplified the severity of weather events worldwide. For example, Hurricane Otis, the strongest Pacific hurricane that made landfall on records, rendered hundreds of people homeless in Acapulco Mexico (Blake, 2023). At the same time, Australia faced severe droughts and destructive wildfires (Magramo et al., 2024), and marine ecosystems are at risk, with coral bleaching likely to occur on a widespread scale this year (Leslie et al., 2024).

El Niño occurs semi-periodically every 2-8 years, altering SST, sea level pressure (SLP), and atmospheric circulation in the tropical Pacific. ENSO does not only include the warm El Niño phase but also a neutral phase and a cold phase, known as La Niña. Both El Niño and La Niña exhibit variations in intensity, temporal evolution, and spatial distribution of SSTA, resulting in a wide range of climatic impacts. For example, the El Niño events of 1982-83 and 1997-98 recorded exceptionally high SSTA in the eastern equatorial Pacific, termed Eastern Pacific (EP) events, and led to severe precipitation and flooding in Peru. In contrast, the 2002-03 El Niño event displayed moderate SSTA concentrated in the central equatorial Pacific, denoted as Central Pacific (CP), and was associated with significantly less precipitation in South America (McPhaden, 2004).

To minimize the global impacts of ENSO events on society and agriculture, accurate forecasts that account for diversity in intensity and spatial distribution of SSTA are required. To date, most ENSO models are physics-based, solving the differential equations that describe ocean and atmosphere dynamics (Kirtman et al., 2014). These models, however,

are limited in their representation of ocean-atmosphere processes due to finite resolution and approximation errors (e.g. Liu et al., 2022a). For instance, the most advanced physics-based models, such as those within the Coupled Model Intercomparison Project (CMIP), exhibit overly periodic SST variability in the Pacific and exaggerate anomalies in western tropical Pacific - a key region for ENSO dynamics (e.g. Capotondi et al., 2020a; Beverley et al., 2023). Additionally, these coupled ocean-atmosphere models capture only a fraction of the spatial diversity of ENSO events (Karamperidou et al., 2017; Dieppois et al., 2021; Cai et al., 2018).

Over the past decades, the availability of observational data has substantially increased, thanks in large part to real-time satellite observations. This data availability allows the adoption of data-driven models as an alternative to physics-based ENSO models. These models approximate the generative processes of the data, and can thereby capture processes that are unresolved in physics-based models.

This thesis employs data-driven approaches to identify remote impacts of ENSO diversity, define it in terms of latent categories, and improve its forecast accuracy.

1.2. Key Contributions

The key contributions of this thesis are as follows:

i. Impacts: This thesis develops a visual knowledge discovery tool based on the topology of climate networks, that identifies differences in global impacts between El Niño events. We demonstrate that different categories of events exhibit more connections to tropical regions than others.

ii. Definition: We propose an alternative definition of ENSO diversity by approximating the diversity of events with a multi-modal distribution, that allows the assignment of membership probabilities to events instead of grouping them into fixed categories.

iii. Forecasting: This thesis introduces a hybrid neural network-based forecasting model that shows skillful forecasts up to 18 months in advance, provides ensemble-based uncertainty estimates, and asses its predictability. The hybrid approach further allows disentangling linear from nonlinear dynamics in the tropical Pacific.

1.3. Outline

This thesis is structured into four parts: Part I introduces main concepts and sets the stage for the scientific contributions detailed in Part II. Part III discusses the impact of these contributions and outlines potential directions for future research. Supplementary material and additional analyses are contained in Part IV.

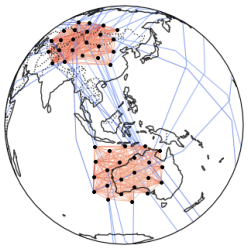
Part I provides the background of this thesis, summarizing related work and motivating the objective of this thesis. Chapter 2 introduces the properties and challenges of climate data analysis, alongside methods for pattern extraction and forecasting. Following this, Chapter 3 introduces

ENSO and its diversity, providing details on its definition, impacts, and forecasting models.

Part II contain the main contributions of this manuscript by addressing challenges in defining ENSO diversity, identifying its global impacts, and producing accurate forecasts.

Chapter 4 introduces a novel network measure designed to identify teleconnections — links that connect remote regions — using climate networks. Unlike previous studies, which focused on individual teleconnections, this approach provides a comprehensive global view of teleconnection patterns. When applied to networks constructed from surface temperature data of EP and CP El Niño events, it is found that EP events exhibit more teleconnections, predominantly confined to the tropics, than CP events.

Chapter 4: Impacts of ENSO diversity



jakob-schloer/netcurvature

Chapter 4 is based on the peer-reviewed journal publication with the following co-author contributions:

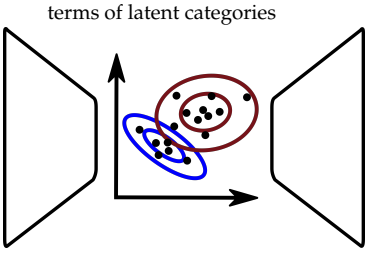
F. M. Strnad, J. Schlör, C. Fröhlich, and B. Goswami (2022). 'Teleconnection Patterns of Different El Niño Types Revealed by Climate Network Curvature'. In: *Geophysical Research Letters* 49.17

	Ideas	**Experiments**	**Analysis**	**Writing**
F. Strnad*	25 %	40 %	45 %	40 %
J. Schlör*	25 %	40 %	45 %	40 %
C. Fröhlich	25 %	20 %	0 %	10 %
B. Goswami	25 %	0 %	10 %	10 %

* Authors contributed equally.

While selecting events for identifying the impacts of EP and CP El Niño events in Chapter 4, I identified inconsistencies in their conventional definitions, which do not account for the continuous nature of these events. Chapter 5 challenges the traditional binary separation by modeling the distribution of ENSO events using a mixture of Gaussians. It shows that ENSO diversity should rather be described by a continuous distribution, with certain states showing higher probability densities. The properties of these states referred to as "fuzzy categories", are analyzed by assigning membership probabilities to each event, revealing shifts in their decadal variability, which are attributed to changes in the likelihood of strong La Niñas and extreme El Niños.

Chapter 5: Defining ENSO diversity in terms of latent categories



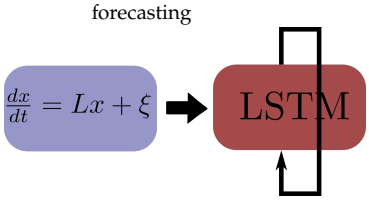
jakob-schloer/LatentGMM

Chapter 5 is based on the peer-reviewed journal publication with the following co-author contributions:

J. Schlör, F. Strnad, A. Capotondi, and B. Goswami (2024). *Contribution of El Niño Southern Oscillation (ENSO) Diversity to Low-Frequency Changes in ENSO Variance*

	Ideas	Experiments	Analysis	Writing
J. Schlör	40 %	80 %	70 %	60 %
F. Strnad	5 %	10 %	10 %	5 %
A. Capotondi	15 %	0 %	10 %	15 %
B. Goswami	30 %	10 %	10 %	20 %

Chapter 6: A hybrid model for ENSO forecasting



jakob-schloer/HybridLIM

Chapter 6 proposes a hybrid model that combines a multivariate linear model with a recurrent neural network to forecast SSTA in the tropical Pacific. This approach allows disentangling linear from nonlinear dynamics within the system, assessing their relative contributions to ENSO dynamics — an open question in the development of physics-based ENSO models. The hybrid model not only offers a degree of interpretability and access to its predictability but also achieves state-of-the-art forecast accuracy. It provides uncertainty estimates by generating a set of ensemble forecasts and maintains high accuracy even in scenarios with limited data.

Chapter 6 is based on the manuscript in preparation with the following co-author contributions:

J. Schlör, M. Newman, J. Thuemmel, A. Capotondi, and B. Goswami (2024). 'A Hybrid Model for ENSO Dynamics in the Low-Data Regime'. In Preparation

	Ideas	Experiments	Analysis	Writing
J. Schlör	50 %	80 %	70 %	60 %
M. Newman	20 %	0 %	10 %	15 %
J. Thuemmel	10 %	20 %	0 %	5 %
A. Capotondi	10 %	0 %	10 %	10 %
B. Goswami	10 %	0 %	10 %	10 %

Part III summarizes the findings of this thesis, discussing their impact and relevance to recent advances in the field. It concludes by proposing directions for future research that build on the insights and methodologies developed throughout this work.

Part I.

Background & Motivation

2. Climate Data Analysis

Thousands of sites worldwide routinely gather operational surface, upper-air, and ocean observations, supplemented by radar, aircraft, and satellite data. Processing the sheer volume of this data is impractical for the human mind; therefore, our objective is to develop data-driven models that effectively summarize and capture the patterns within the data, approximating their underlying *generative process*.

Typically, the *climate system* is modeled by describing the physical processes using differential equations. Despite these models' detailed nature, accurately modeling all influences of this complex, chaotic system across all scales remains infeasible (Lorenz, 1969; Stephenson et al., 2012). Data-driven models, in contrast, extract the relevant dynamics from the data by approximating known and unknown processes without explicitly accounting for them.

Observational data from the climate system are characterized by unique structural properties that have to be considered for their analysis:

i) Generally a climate dataset consists of several variables, N_{var}, with each variable having three spatial dimensions — latitude N_{lat}, longitude N_{lon}, and height N_z — and a temporal dimension N_t. In this thesis, I denote a dataset by $X = \{\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_{N_t})\}$, where $\mathbf{x}(t) \in \mathbb{R}^{N_{\text{var}} \times N_t \times N_{\text{lon}} \times N_{\text{lat}} \times N_z}$. Most surface variables, such as SST, are defined for only one z-dimension.

ii) The data live on the surface of a sphere, with their x-y-coordinates typically represented as angles. Latitude describes the north-south position, ranging from -90 to 90 degrees, with 0 degrees at the Equator, and longitude specifies the east-west position, ranging from -180 to 180 degrees, with 0 degrees defined as passing near the Royal Observatory in Greenwich. Since points on the longitude-latitude grid are not equidistant, alternative *equidistant representations* that reflect the SO(3) symmetry have been proposed.

iii) Climate data exhibit strong local correlation structures, both spatially and temporally. This means that measurements taken close in time and space either influence each other or stem from the same generative process without influencing one another. Such correlations must be accounted for in models to avoid spurious relationships and ensure robust statistical inference.

iv) In the climate system, dynamics occur at all spatial and temporal scales which interact with each other. Averages computed over larger areas can yield different spatial patterns and autocorrelation structures. For example, local weather extremes may be lost in broader regional averages.

v) The dynamics of the climate system are forced by the sun and anthropogenic climate change, leading to periodicity and trends in the data. These forcings can artificially inflate correlations and forecasting accuracy. For example, predicting temperatures for next August based on historical August temperatures typically shows high accuracy because it captures a recurring pattern rather than

Generative process: refers to a model that describes how observations x are generated in terms of underlying hidden structures or typically low dimensional random variables z. Generative models aim to capture the stochastic process that defines the joint probability distribution $p(x, z) = p(x|z)p(z)$ over observed and latent variables, where $p(z)$ is the prior distribution and $p(x|z)$ the likelihood.

Climate system: I refer to the climate system as the complex interaction between the atmosphere, hydrosphere, cryosphere, lithosphere, and biosphere, governed by the dynamic exchange of energy and moisture that determines Earth's climate. Weather and climate correspond to different timescales in these dynamics and are not distinguished here.

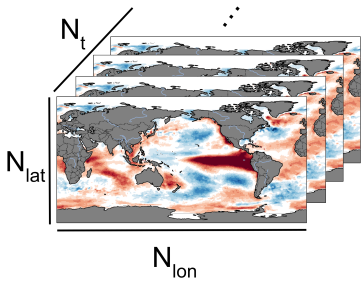


Figure 2.1.: Climate data are high dimensional. One variable (here, sea surface temperature anomalies) has longitude N_{lon}, latitudes N_{lat}, and temporal dimension N_t.

novel information. To further improve predictions, it is necessary to consider the deviation from these recurring patterns and work with *anomalies*.

vi) The climate system's chaotic behavior, exemplified by the butterfly effect Lorenz (1963) and Lorenz (1969), indicates that small changes in one location can grow exponentially, leading to large effects elsewhere. This sensitivity to initial conditions makes forecasting inherently challenging and underscores the need for uncertainty estimates.

Given a set of observations, our objective is to infer a model that approximates the data's generative process. This inference, known as inductive learning or learning from data in machine learning, can be categorized into three types: unsupervised, supervised, and self-supervised learning, where each type targets a distinct aspect of the generative process (Bishop, 2006; Murphy, 2013):

Unsupervised learning aims to approximate the probability distribution of the dataset, $p(X)$ without using labeled outputs. Essentially, it seeks to learn patterns or features from the data itself to extract knowledge about the underlying processes. Unsupervised learning can be used to identify clusters or patterns within the data that might signify different regimes of the climate system.

Supervised learning seeks to approximate a conditional distribution of the data, $p(Y|X)$, where Y serves as external labels. This approach aims to learn a function that, given a new input, can predict the appropriate output. For example, the probability of rain in a specific location using input features such as temperature, humidity, and pressure.

Self-supervised learning focuses on approximating a conditional distribution $p(\hat{X}|X)$, where X and $\hat{X}$ are two different views of the same set. This approach is particularly relevant to sequence prediction tasks, such as forecasting a future state $\hat{\mathbf{x}}$ from the current state $\mathbf{x}$.

In the following, I provide a theoretical outline of *data-driven models* used throughout this thesis. These include unsupervised approaches for identifying patterns in the data (Ch. 4 and Ch. 5) and self-supervised approaches for spatio-temporal sequence forecasting (Ch. 6).

Equidistant projections: There are several methods for generating points that are approximately equidistant on the surface of a sphere:

- **Reduced Gaussian grid** modifies the typically used latitude-longitude grid by reducing the number of points when moving toward the poles (Kurihara, 1965; Hortal et al., 1991).
- **Geodesic grid** is constructed by subdividing the faces of a polyhedron into smaller triangles, which are then projected onto the sphere (Sadourny et al., 1968; Williamson, 1968).
- **HEALPix grid** provides an equal area projection of points on a sphere by subdivisions into rhombic dodecahedrons (Górski et al., 2005).
- **Fibonacci grid:** Points are distributed using the Fibonacci spiral and are projected onto the sphere, resulting in a quasi-uniform distribution (Chukkapalli et al., 1999; Swinbank et al., 2006).
- **Fekete grid** numerically optimizes the distribution of points on a sphere using spherical harmonics (Bendito et al., 2007).

Anomalies: Anomalies are calculated as deviations from the climatology, which is the long-term average of the respective month or season. For instance, the monthly temperature anomaly for this year January is obtained by subtracting the observed temperature from the mean over January temperatures in the last 30 years.

Data-driven model: A model that synthesizes and summarizes information from data to approximate their underlying generative processes. The term "data-driven" is used here in preference to "statistical" or "machine learning" models to emphasize the reliance on data for model parameter inference.

2.1. Pattern Extraction

In this part, we introduce dimensionality reduction methods employed in climate science, as used in Ch. 5, and outline climate network analysis employed in Ch. 4.

2.1.1. Dimensionality Reduction

Dimensionality reduction techniques are designed to transform high-dimensional data, into a lower-dimensional space while preserving their essential structure. The corresponding generative process reads $p(\mathbf{x}) = \int d\mathbf{z} p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$, where $p(\mathbf{z})$ represents the prior distribution in the latent space $\mathbb{R}^M$ with $M << N_{\text{lat}} \times N_{\text{lon}}$ and $p(\mathbf{x}|\mathbf{z})$ is its likelihood. The primary objective is to infer the posterior distribution $p(\mathbf{z}|\mathbf{x})$, ensuring

that the approximated generative process closely matches the true data distribution $p(\mathbf{x})$.

Empirical Orthogonal Function (EOF) Analysis

One of the most widely used dimensionality reduction techniques in climate science is Empirical Orthogonal Function (EOF) analysis, also known as Principal Component Analysis (PCA) (Lorenz, 1956; Wilks, 2019). EOF analysis is a linear dimensionality reduction technique for spatial-temporal data, represented as $X = \{\mathbf{x}(t)\}_{t=1}^{N_t}$, where $x_s(t)$ denotes the value at time $t \in [1, \dots, N_t]$ and grid point $s \in [1, \dots, N_{\text{lat}} \cdot N_{\text{lon}}]$ of the flattened field vector $\mathbf{x}(t) \in \mathbb{R}^{N_{\text{lat}} \cdot N_{\text{lon}}}$. The principal components (PCs) are defined by the latent vector $\mathbf{z}(t) \in \mathbb{R}^M$, where each component $z_r(t), r \in [1, \dots, M]$ is a linear combination of the field values, $z_r(t) = x_s(t) u_{rs}$. Here, $\mathbf{u}_r \in \mathbb{R}^{N_{\text{lat}} \cdot N_{\text{lon}}}$ represents the EOF corresponding to the rth eigenvector of the covariance matrix $\langle XX^T \rangle$, associated with eigenvalue λ_r. Redundancies in the original dataset X allow for significant data variance capture through only the most dominant directions of their joint variations. Consequently, most of the informational content of X can be represented using a significantly reduced number of components, $M \ll N_{\text{lat}} \cdot N_{\text{lon}}$, by focusing on the leading eigenvectors, sorted by their eigenvalues. The original data can be reconstructed using the inverse transformation, $\mathbf{x}(t) = \sum_{r=1}^{M} \lambda_r z_r(t) \mathbf{u}_r$.

EOF analysis is frequently used in climate science, particularly in defining indices such as the Madden-Julian Oscillation and ENSO. While EOFs effectively reduce the dimensionality of complex data sets, it is important to note that they do not necessarily represent dynamically meaningful modes of the system. Instead, EOFs are influenced by boundary conditions and may produce biased patterns, often referred to as "buell patterns" (Wilks, 2019).

While EOF analysis is initially described as a deterministic linear transformation, it can also be conceptualized as a probabilistic method with both the prior and the likelihood assumed to be Gaussian (Bishop, 2006). Moreover, instead of maximizing variance through the eigenvectors of the covariance matrix, EOF analysis can alternatively be formulated to minimize the reconstruction error. This involves calculating the root mean squared error (RMSE) between the original data vector $\mathbf{x}$ and its reconstructed version $\hat{\mathbf{x}}$ (Bishop, 2006). Thereby, EOF analysis is equivalent to a linear autoencoder, as we will see in the next section.

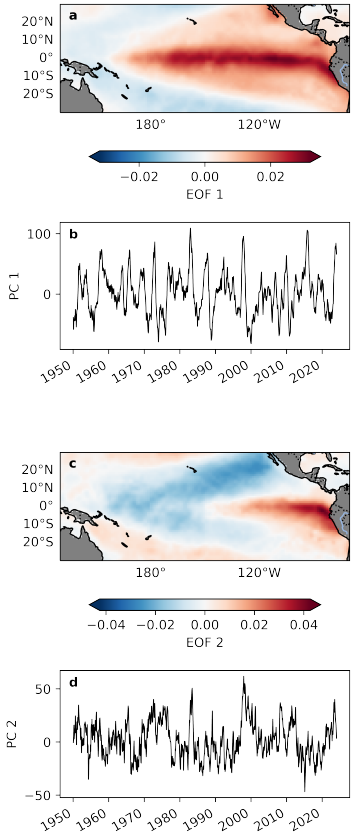


Figure 2.2.: First two leading EOFs (a,c) obtained from tropical Pacific SSTA alongside their corresponding PCs (b,d). The EOF analysis is performed on the HadISST dataset, see Tab. 3.2.

Autoencoder

Autoencoder (AE) allow non-linear dimensionality reduction by using neural networks as encoder and decoder functions, denoted by e_{θ_d} and d_{θ_d} respectively, where θ_e and θ_d represent the network parameters. The encoder takes a datapoint $\mathbf{x}(t) \in \mathbb{R}^{N_{\text{lon}} \times N_{\text{lat}}}$ as input and transforms it into a lower-dimensional latent vector $\mathbf{z}(t) \in \mathbb{R}^M$. Conversely, the decoder creates a reconstruction $\hat{\mathbf{x}}(t) \in \mathbb{R}^{N_{\text{lon}} \times N_{\text{lat}}}$ from the latent vector. The training of these networks focuses on minimizing information loss at the bottleneck (Fig. 2.3), typically by reducing the RMSE between $\mathbf{x}(t)$ and $\hat{\mathbf{x}}(t)$, as

$$argmin_{\theta_e,\theta_d} \frac{1}{N} \sum_{i=1}^{N} (\widehat{\mathbf{x}}(t) - \mathbf{x}(t))^2$$
$$= argmin_{\theta_e,\theta_d} \frac{1}{N} \sum_{i=1}^{N} \left(d_{\theta_d}(e_{\theta_e}(\mathbf{x}(t))) - \mathbf{x}(t)\right)^2,$$

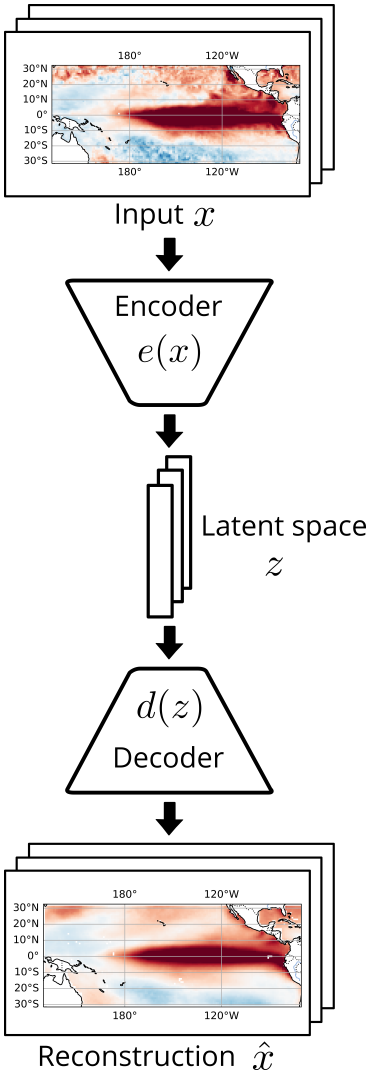


Figure 2.3.: Sketch of an autoencoder network. The input x is first transformed to a lower dimensional latent vector z and then projected back to the input space. The encoder e and decoder d are neural networks that are optimized to minimize the information loss by the dimensionality reduction.

with N is the number of time points used for training.

In Ch. 5, we explore the use of a simple autoencoder for dimensionality reduction. While the autoencoder is a deterministic dimensionality reduction method, it can be seen as the *maximum a posteriori* estimate of a probabilistic autoencoder, specifically the Variational Autoencoder (VAE). The VAE employs both an encoder and a decoder to parameterize the likelihood and the approximate posterior distribution, respectively. Training the VAE minimizes the Kullback-Leibler (KL) divergence in its variational inference approach to quantify the difference between the learned distribution and the prior distribution.

Clustering

Clustering serves as a specialized form of dimensionality reduction in which the one-dimensional latent variable z can only take discrete values $k = 1, \ldots, K$ with prior probability $p(z = k) = \pi_k$. This method utilizes the degree of similarity and differences among individual observations $\mathbf{x}$ to define groups and assign membership. Within a dataset X, clustering analysis categorically assigns data points to clusters $k = 1, \ldots, K$ based on their proximity. For example, daily weather observations were clustered into synoptic types by Kalkstein et al., 1987.

The concept of distance is fundamental to clustering. Ideally, clusters consist of points that are closely spaced relative to the distances between different clusters. There are numerous potential definitions of distance, and the choice significantly influences the outcomes of the analysis. While Euclidean distance is the most intuitive and commonly used metric, it is not the only option, and sometimes may not be the best choice. For climate data for instance the dimensionality of the data might be much larger than the number of datapoints, i.e. $N_t \ll N_{\text{lat}} \cdot N_{\text{lon}}$. In this case, many clustering algorithms struggle due to the curse of dimensionality (Wilks, 2019). As the number of dimensions in a dataset increases, distance measures become increasingly meaningless (Parsons et al., 2004). Additional dimensions spread out the points until, in very high dimensions, they are almost equidistant from each other. To circumvent that issue, either only certain features should be selected or the dimensionality of the features should be reduced first (Ch. 5).

2.1.2. Climate Networks

Another unsupervised approach to uncovering patterns in high-dimensional data is climate networks. Climate networks structure data into a set of nodes and edges which allows for the discovery of key communities and pathways within the system.

Given spatio-temporal data X, where each spatial location i represents a time series (Fig.2.4). A graph G is constructed, comprising a set of nodes $V = \{v_1, \ldots, v_{N_{lat} \cdot N_{lon}}\}$, each corresponding to a spatial location. Nodes are interconnected by a set of edges E, with an edge e_{ij} signifies the interdependence between pairs of time series at nodes v_i and v_j, determined using a statistical similarity measure. The network is represented by a weight matrix $\mathbf{W}$, defined as:

$$\mathbf{W}_{ij} = \begin{cases} w_{ij}, & e_{ij} \in E \\ 0, & \text{otherwise} \end{cases},$$

where the edges e_{ij} are assigned weights w_{ij}. A threshold is often applied to these weights, resulting in an adjacency matrix $\mathbf{A}$:

$$\mathbf{A}_{ij} = \begin{cases} 1, & e_{ij} \in E \\ 0, & \text{otherwise} \end{cases},$$

The selection of input data, such as variables, time resolutions, and the similarity measure, is critical for constructing a meaningful climate network. The chosen similarity measure determines whether the graph's edges are directed ($e_{ij} \neq e_{ji}$) or undirected ($e_{ij} = e_{ji}$). Common similarity measures for continuous variables include the correlation coefficient and mutual information, while for discretely distributed data, event synchronization is often used.

Thorough testing for statistical significance is required when creating unweighted climate networks (Wilks, 2019). This involves two steps:

1. **Null Model Definition:** To ascertain if connections occur by chance, a null model is created. For each location (node), links are randomly rewired. A Gaussian kernel density estimator (KDE) is then used to determine the likelihood of a link existing to the chosen location. An observed link is statistically significant if its spatial likelihood, exceeds the 99.9th percentile in the local null model link distribution.
2. **Multiple Testing Corrections:** Considering the vast number of links in climate networks, it's crucial to correct for multiple testing to avoid false positives by chance. Techniques like the Bonferroni correction (Holm, 1979), which adjusts p-values to account for the number of tests, must be employed.

Climate networks, once constructed, allow us to analyze their topological properties. For instance, investigating network communities can uncover regions with similar climatic behaviors and the network's centrality metrics help to identify the most influential nodes, indicative of regions critical for information flow. Additionally, studying how the network responds to perturbations can be crucial for understanding the system's stability and resilience to climatic anomalies. Furthermore, analyzing the temporal evolution of the network's topology enables us to study how *teleconnections* change over time. In Ch. 4, we employ climate networks to discern the global impacts of El Niño events, highlighting their changes in teleconnection patterns between event types.

Climate networks serve as a visual knowledge-generating tool by condensing both time and variable dimensions of high-dimensional climate

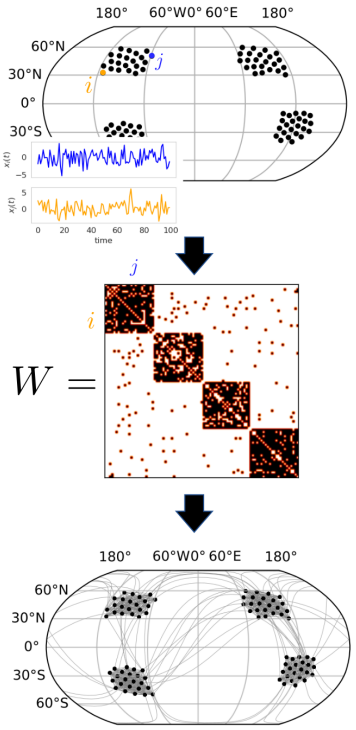


Figure 2.4.: Schematic representation of a climate network construction. The adjacency matrix, W, is obtained from the pair-wise correlation between time series at each location i, j.

Teleconnections: refer to anomalous climate conditions that are related across large geographical distances. Teleconnections are often characterized by atmospheric changes such as temperature, pressure, wind, or rain that occur simultaneously or sequentially across different regions.

data. However, this comes with certain limitations. Most similarity measures used in climate networks capture simultaneous changes, while in reality, physical interactions in the climate system often involve propagation time. Another limitation is that network links represent statistical relationships but not causal interdependencies. For instance, a significant link between two nodes might be induced by dependencies to a third node, with no causal relationships between the first two.

Ideally, climate networks would be causal graphs with lagged dependencies. Probabilistic graphical models, which model conditional probabilities between nodes, have been developed to represent causal structures. Learning these causal relationships from data is an active research field in machine learning (Kaddour et al., 2022; Schölkopf, 2022).

Granger causality: Granger causality serves as a statistical test for determining whether a time series is useful in forecasting another, but it has been argued that regressions reflect "mere" correlations, rather than causation (Thurman et al., 1988; Wilks, 2019).

Runge et al. (Runge et al., 2019a; Runge et al., 2019b) introduced PCMCI for climate data, a method using *Granger causality* to determine links. Although this method facilitates the detection of directed and lagged links, it's limited to a small number of nodes due to the computational complexity of the vast amount of regression tasks. While the network links obtained from PCMCI are referred to as causal, it is debatable if Granger causality reflects true causality. Unfortunately, actual causality, typically derived from interventions and randomized control experiments, is not feasible in the climate system. Within climate models, single interventions can be tested through 'nudging' experiments, but these are limited by computational complexities and aren't widely scalable.

2.2. Data-Driven Forecasting

Forecasting can be either a self-supervised learning task $p(\hat{X}|X)$, when the target outcome is obtained from the same distribution as the input, such as in sequence-to-sequence predictions, or a supervised learning task $p(Y|X)$, as for instance in predicting a specific weather regime that requires externally labeled data y.

Forecasting climate data is essentially a multivariate sequence prediction task. This task requires estimating the future state $\hat{\mathbf{x}}(t+\tau)$ at a given lag time τ, based on initial states $\{\mathbf{x}(t-h), \ldots, \mathbf{x}(t)\}$. The goal is to find a good estimate of the transition probability $p(\hat{\mathbf{x}}(t+\tau)|\mathbf{x}(t-h), \ldots, \mathbf{x}(t))$.

2.2.1. Model Design

Given the structural properties of climate data, outlined at the beginning of the chapter, certain model design choices are beneficial. Here, I outline design choices for data-driven models that are effective for sequence prediction of climate data (Thuemmel et al., 2024). The focus will be primarily on neural network models and include data selection, learning objectives, loss functions, model architecture, and optimization strategies.

i. Data selection: Data selection covers the choice of variables and the selection of spatial and temporal resolutions for both the input and target of the model. Specific considerations for selecting data for ENSO forecasting will be discussed in Sec. 3.1.3.

ii. Learning objective: Learning objectives address the high-level functionality of the model, specifying the forecast mode and addressing how to handle uncertainty. These objectives determine whether the forecast task is classification, regression, or generative modeling. For sequence prediction, two primary approaches can be considered:

- Recurrent (iterative) forecast: This method involves predicting future states iteratively, where each prediction is used recursively as input for subsequent forecasts.
- Lag-time forecast: This approach uses direct forecast models to generate predictions for specific target lag times, without producing any intermediate predictions.

iii. Loss functions Loss functions are used to quantify the fit quality of model predictions to the data and enable gradient-based optimization. The choice of a loss function depends on the learning objective and the nature of the data. I categorize loss functions into deterministic and probabilistic scores:

- Deterministic losses: Common metrics include the L1-norm and L2-norm, corresponding to mean absolute error (MAE) and RMSE, respectively. The selection of a deterministic loss function should align with the assumed probability distribution of the target; for example, RMSE is appropriate for a Gaussian distribution, while L1 is suitable for a Poisson distribution (Murphy, 2013).
- Probabilistic losses: For continuous variables, commonly used losses are the negative log-likelihood and the *Continuous Ranked Probability Score (CRPS)*, the latter of which can be applied in both parametric and non-parametric forms. For discretized variables, such as rainfall measured in bins, Cross-Entropy loss has proven effective.

Numerical physics-based models incorporate uncertainties through ensemble predictions, generated by varying initial conditions and model parameters. These ensemble-based forecasts are beneficial as each member represents a possible forecast trajectory, allowing to capture of rare events and bifurcation. Ensemble-based uncertainty estimates in neural networks are proposed in Sec. 6.

iv. Model architectures Neural network architectures consist of stacked computational primitives, each defined by learnable parameters tailored to specific objectives. These architectures are typically arranged in a sequence of layers, forming the overall network design.

The layers used in this thesis are:

- Fully Connected (Dense) Layer: Each input in a layer connects to all inputs in the subsequent layer. Stacked dense layers with a nonlinear activation function are called Multilayer Perceptron (MLP) (e.g. Rumelhart et al., 1986).
- Convolutional Layer: Applies convolution of learnable parameters and the input. Convolutional layers are particularly effective for processing data with grid-like topologies, such as images (e.g. Lecun et al., 1998). Stacked convolutional layers are called Convolutional Neural Network (CNN).

Continuous Ranked Probability Score (CRPS):
The CRPS is a proper probabilistic score defined as the integral of the squared difference between the cumulative distribution function (CDF), $F(\hat{x})$, of the forecast and the Heaviside step function at the observed value x_o. Mathematically, it can be expressed as:

$$\mathrm{CRPS}(F, x_o) = \int_{-\infty}^{\infty} (F(\hat{x}) - \mathbf{1}(x \geq o))^2 dx. \tag{2.1}$$

For a Gaussian distributed forecast, $\hat{x}\ \mathcal{N}\left(\hat{x}, \mu, \sigma\right)$, the CRPS can be computed analytically, by

$$\mathrm{CRPS}(F, x_o) = \sigma[\frac{x_o - \mu}{\sigma}(2\Phi(\frac{x_o - \mu}{\sigma}) - 1) + 2\phi(\frac{x_o - \mu}{\sigma}) - \frac{1}{\sqrt{\pi}}] \tag{2.2}$$

where Φ and ϕ are the CDF and PDF of the standard normal distribution, respectively. Additionally, the CRPS can be formulated for empirical distributions, as

$$\mathrm{CRPS}(F, x_o) = E|\hat{X} - x_o| - \frac{1}{2}E\left|\hat{X} - \hat{X}'\right|, \tag{2.3}$$

where E is the statistical expectation, and $\hat{X}$ and $\hat{X}'$ are independent realizations from F.

- Recurrent Layer: Used for sequential data, these layers maintain memory states over time. Prominent examples are Long-Short Term Memory network (LSTM) (Hochreiter et al., 1997) units and Gated Recurrent Units (Cho et al., 2014).
- Normalization Layer: Normalize the inputs to stabilize and speed up training. Batch and Layer Normalization for instance are widely used (e.g. Wu et al., 2018).
- Activation Function: Non-linear functions such as Rectified Linear Unit, Sigmoid, and Tanh is applied to the outputs of layers to learn non-linearities in the data (e.g. Bishop, 2006).
- Embedding Layer: Transform categorical variables into high-dimensional vector spaces (Mikolov et al., 2013).
- Residual Connection: Enhances the training of deeper networks by allowing gradients to flow through a shortcut connection, used in architectures like residual networks (He et al., 2015).

Other common layers:

- Pooling Layer: Reduces the spatial dimensions (width, height) of the input for subsequent convolutional layers, with Max Pooling and Average Pooling being the most common forms.
- Dropout Layer: Serves as a regularization technique to prevent overfitting by randomly omitting a subset of features during training (Srivastava et al., 2014).
- Attention Mechanism: Attention enables one to focus selectively on only parts of the input sequence. This mechanism computes the relevance of different parts of the data, facilitating tasks that require contextual understanding, such as natural language processing. Attention is the building block in Transformers (Vaswani et al., 2023).
- Message Passing: Crucial in architectures designed for data represented by graphs. Here, it enables the integration and updating of information between nodes in a graph (Micheli, 2009; Scarselli et al., 2009). Architectures with blocks of message passing are typically called Graph Neural Network (GNN).

For sequence prediction of climate data, networks should have large receptive fields to capture relevant features across various scales effectively. Encoder-decoder architectures (Sec. 2.1), for example, are beneficial as they reduce their effective internal resolution relative to the actual data resolution, enhancing computational efficiency. Additionally, incorporating recurrence and memory capabilities is beneficial for sequence prediction. These enable models to operate over larger temporal steps compared to traditional numerical methods, thus improving computational efficiency.

v. Optimization: Optimizing model parameters defines the learning process, which involves strategies such as mini-batching, learning-rate adaptation, and scheduling of objectives. Neural networks are typically trained using stochastic gradient descent (SGD) (e.g. Robbins et al., 1951), which optimizes the model by iteratively updating parameters to minimize the loss function. This optimization occurs over batches of data — small subsets known as mini-batches — that represent the overall dataset during each iteration. In deep learning libraries, such as pytorch (Paszke et al., 2017), the optimization of model parameters, including the automatic gradient computation, backpropagation, and parameter updating, is computed automatically in an efficient manner.

2.2.2. Inductive Biases

Each design choice from steps i - v introduces structural assumptions about the data and the modeled processes, known as inductive biases (Wolpert et al., 1997; Battaglia et al., 2018). These biases can accelerate model convergence and enhance both computational efficiency and the model's generalization to unseen data. For instance, a convolutional deep learning model introduces the bias that adjacent pixels are likely to contain correlated information, making it particularly effective for processing images compared to a standard MLP, which does not account for spatial proximity. Essentially, every decision in the learning pipeline embeds implicit inductive biases that significantly influence model performance (Thuemmel et al., 2024).

To summarize, the key inductive biases for training neural networks on climate data are:

i) **Large Receptive Field:** Essential for capturing salient features across all scales is to reduce the effective internal resolution of the model relative to that of the data.
ii) **Recurrence and Memory:** Incorporating recurrence and memory within the model is required to forecast on longer temporal scales.
iii) **Ensemble-based uncertainty:** Effectively capturing uncertainties is crucial for forecasting the climate system. Especially, ensemble-based forecasts are beneficial to obtain coherent trajectories and capture extreme events.

El Niño Southern Oscillation 3.

3.1. What is ENSO and Why is it Diverse?

When I discuss El Niño with friends and family, they often recognize it from school or news coverage, usually perceiving it as a kind of storm in South America. This phenomenon, known to Peruvian fishermen since the 19th century for its warm, southward ocean currents and heavy rains around Christmas, was named "El Niño", Spanish for "Christ child" (Carillo, 1892). Our understanding evolved with contributions from Sir Gilbert Walker (Walker, 1925) and Jacob Bjerknes (Bjerknes, 1966), who identified El Niño as part of the larger El Niño Southern Oscillation (ENSO) — a complex ocean-atmosphere phenomenon in the tropical Pacific. The phases of ENSO include El Niño, marked by warmer than average SSTs, La Niña with cooler SSTs, and Neutral for average conditions. These phases occur irregularly every 2 to 8 years, not in a periodic manner as the term “oscillation” might imply (McPhaden et al., 2020). Detailed explanations of the mechanisms driving these phases are presented subsequently.

3.1.1. Mechanisms of ENSO

Matt Newman, my advisor at National Oceanic and Atmospheric Administration (NOAA), once broke down the mechanism of ENSO with a simple analogy: Imagine the Pacific Ocean as a gigantic bathtub that slowly absorbs heat from the warmer atmosphere. Over time, this heat builds up below the surface, potentially over several years. When the ocean gets sufficiently warm, small changes in the atmospheric conditions, such as strong winds, can trigger an abrupt release of heat, resulting in an El Niño event. The amount of heat released to the atmosphere often exceeds what was accumulated, leading to cooler temperatures - a phase we know as La Niña. The ocean then begins to balance these cooler temperatures by gradually taking up heat from the atmosphere once more. While this "bathtub" analogy is a simple picture, it helps in understanding the complex mechanisms that will be detailed next.

Upwelling: Upwelling is a phenomenon where deep, cold, and nutrient-rich water rises to the surface of the ocean. This process typically occurs along coastlines and in the open ocean due to the wind-driven movement of surface waters away from an area.

Thermocline: In oceanography, the thermocline refers to the layer within a body of water, where the temperature gradient is substantially greater than that of the warmer layer above and the colder layer below. In the Pacific Ocean, these vertical temperature jump occurs approximately from 17 to 23°C, centered on the 20°C isotherm, which is typically referred to as the main thermocline.

Neutral Phase

Around the equator, winds blow from the East to the West, driven by temperature gradients between the equator and in the mid-latitudes. In the Pacific, these easterly trade winds generate a surface current that flows from Peru’s coast to the Maritime Continent in the West. As this water travels westward, it warms up under the sun, culminating in the Western Pacific warm pool, which results in the warmest open ocean water temperatures on Earth (Fig. 3.1). Conversely, in the Eastern Pacific, equatorial *upwelling* brings cold water from the ocean’s depths to the surface, forming a "cold tongue" of SST that stretches to the international dateline. The *thermocline* is a region that separates the warm

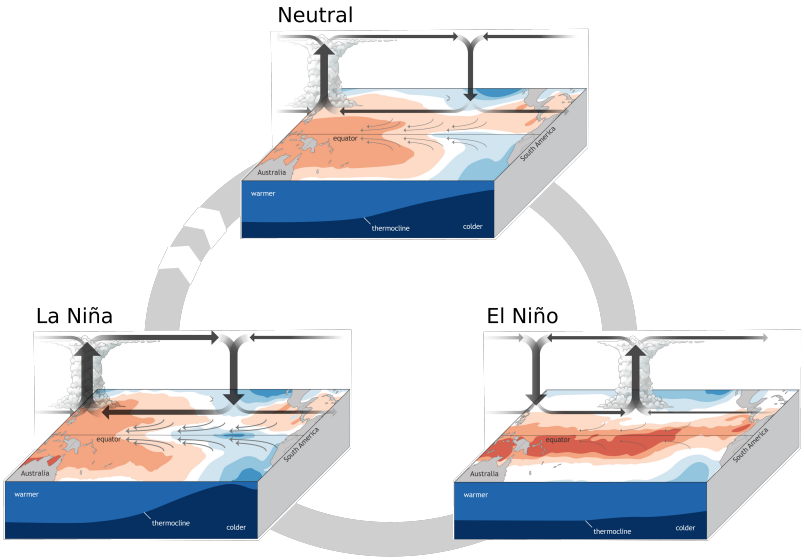


Figure 3.1.: Schematic ocean and atmosphere conditions in the tropical Pacific during normal (**Neutral**), warm (**El Niño**) and cold (**La Niña**) phase of ENSO. The figure is adapted from L'Heureux (2020).

upper ocean from the cold deep waters. Its tilt in the east-west direction is related to the strength of upwelling and the trade winds' intensity. Warm surface waters in the western Pacific, cause air to rise, leading to deep atmospheric convection and heavy rainfall. After condensation in the upper atmosphere, the dry air descends over the cooler waters of the Eastern Pacific. These elements collectively define what we call the Neutral phase of the ENSO.

El Niño Phase

ENSO phase locking: El Niño and La Niña events typically peak in boreal winter, a phenomenon called ENSO phase locking. This seasonal dependency has been predominantly attributed to the seasonal variation of the cold tongue region in the East and the variations of the warm pool extend in the West Pacific.

An El Niño event typically begins with a deeper-than-average zonal thermocline, warmer upper ocean temperatures, and weaker trade winds. While these are necessary but not sufficient onset conditions, the reason for their emergence, and their predictability remain a topic of ongoing research. As trade winds weaken, the thermocline slope levels out, and upwelling in the cold tongue reduces. These changes lead to a decrease in east-to-west water transport, causing the warm pool to expand eastwards. As the central and eastern Pacific warm up, the ascending air masses responsible for deep convection and heavy rainfall in the western Pacific migrate eastwards with the warm water. This shift further weakens the trade winds, which in turn enhances surface warming. The interplay of the ocean and the atmosphere becomes locked in a positive feedback loop, known as the Bjerknes feedback (Bjerknes, 1966), where weakened winds and warming SSTs intensify each other. El Niño typically reaches its peak in the boreal winter months (*ENSO phase locking*) and is characterized by warmer-than-average SSTs, elevated sea levels in the central and eastern Pacific, and anomalously westerly winds (Fig. 3.1).

Sverdrup transport: The surface winds drive the ocean currents. In oceanography, the wind force parallel to the ocean surface is called wind stress. Sverdrup transport describes the meridional water transport, i.e. water transport towards or away from the poles, which results from wind stress on a rotating sphere (Sverdrup, 1947).

These deviations from the normal conditions initiate a counteracting process. The primary balancing force is the *Sverdrup transport*, which redistributes heat from the equator towards the extratropics, drawing cooler water from deeper ocean layers to the surface. This heat discharge at the equator leads to a decay of El Niño back to the Neutral phase but often overshoots into the cold La Niña phase.

La Niña Phase

La Niña conditions are typically initiated by anomalously strong temperature differences between the warmer western and cooler eastern Pacific along with cooler than average upper ocean heat content in the Pacific. This leads to stronger easterly trade winds and enhanced upwelling in the east, which subsequently leads to further cooling in the east and warming in the west. This feedback loop of stronger trade winds, a steeper westward inclination of the thermocline, increased upwelling in the cold tongue region, and the westward shift of the warm pool drives a mature La Niña event (Fig. 3.1). Similar to El Niño, La Niña peaks in boreal winter, and its decay back to Neutral conditions is driven by Sverdrup transport.

3.1.2. The Diversity of ENSO

The previous section outlined the physical feedback processes of ENSO, which can be described by oscillator-like differential equations (see Sec. 3.3.2). While this mechanistic perspective has been substantial for our understanding of ENSO, it is unable to explain why the El Niño in 1997-98 (Fig. 3.2a) was different from the El Niño 2002–2003 (Fig. 3.2b). Already in 1975, Wyrtki highlighted the uniqueness of events by stating that "no two El Niño events are quite alike" (Wyrtki, 1975). The characteristics of event-to-event differences are:

- **Stochastic Occurrence:** Unlike a periodic cycle (Neutral → El Niño → La Niña → Neutral), ENSO phases occur stochastically. This means that neither El Niño nor La Niña follow a predictable, regular pattern but occur in succession or at irregular intervals (Fig. 3.5b). For instance, three La Niña events have occurred successively between 2020 and 2023, termed a triple-dip La Niña.
- **Temporal variation:** The start season of ENSO events and their duration are highly variable. La Niña events, for instance, are more likely to last for multiple years than El Niño events.
- **Intensity variation:** The strength of El Niño and La Niña events differ substantially. Some events may be mild, barely altering global weather patterns, while others can be extremely strong, causing extreme weather events.
- **Spatial variability:** The zonal location and extent of warming (in El Niño) and cooling (in La Niña) in the tropical Pacific are not consistent across events (Fig. 3.2). This spatial variability can lead to different impacts in different regions, such as altered rainfall patterns, droughts, or floods.
- **Asymmetry:** There is a notable asymmetry in the intensity of warm (El Niño) and cold (La Niña) events. Typically, warm events are stronger than cold ones, leading to a positively skewed distribution of SSTA.

The event-to-event variability, known as ENSO diversity (Capotondi et al., 2015) or ENSO complexity (Timmermann et al., 2018), poses significant challenges to the impacts and predictability of events. In the early 2000s, the work of Larkin et al. (2005) drew much attention to the spatial diversity of ENSO as they demonstrated how the zonal location of El

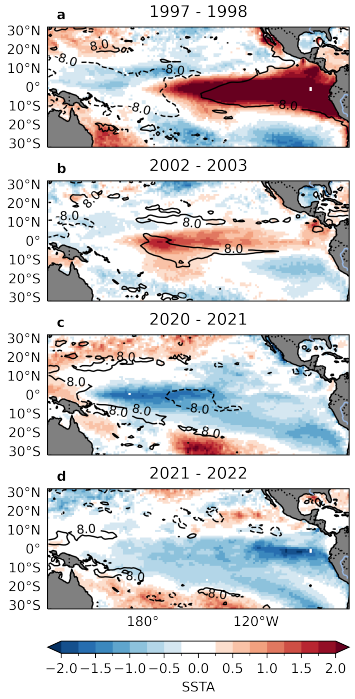


Figure 3.2.: SSTA and sea surface height anomaly (SSHA) during December-January-February (DJF) for the El Niño events of (a) 1997–1998, (b) 2002–2003, and La Niña events of (c) 2020–2021, and (d) 2021–2022. Monthly SST data were obtained from the ORAS5 data set (ORAS5, 2021).

Niño's warming influences temperatures and precipitation in the United States. Their work stimulated intense research activity on ENSO diversity (Kug et al., 2009; Kao et al., 2009; Takahashi et al., 2011). ENSO events have been divided into different categories (Ashok et al., 2007; Kug et al., 2009; Takahashi et al., 2011) to more easily identify differences in the leading dynamical processes, precursors, and impacts (Capotondi et al., 2015; Capotondi et al., 2020b).

Several indices and grouping criteria have been proposed to capture El Niño's diverse characteristics. While these definitions vary, most identify two primary types of El Niño events: *Central Pacific (CP)* and *Eastern Pacific (EP)* El Niño[1]. While EP events exhibit warming in the Eastern and Central Pacific (Fig. 3.3c), CP events typically show warming around the dateline (Fig. 3.4) and tend to have a smaller SSTA amplitude compared to EP events (Fig. 3.3b). EP events are further characterized by a deeper thermocline in the eastern and shallower thermocline in the western equatorial Pacific, contrasted with the weaker positive thermocline depth anomalies of CP events, which extend further west. Additionally, westerly wind anomalies during CP conditions are weaker and located further west compared to those under EP conditions. Although La Niña events also exhibit variations in intensity and spatial pattern (Fig. 3.2c, d), these differences are not as pronounced as those observed in ENSO's warm events (Kug et al., 2011). The common characteristics of EP and CP events are derived from averages over a handful of events (Fig. 3.3a). However, the mechanisms leading to the diversity of events remain unclear.

Despite a large amount of research efforts, to date, the understanding of ENSO diversity is separated into two interpretations. The first, suggests that nonlinear processes play a central role in understanding ENSO and proposes that both types of ENSO are different expressions of a singular phenomenon (Takahashi et al., 2011; Takahashi et al., 2016; Okumura, 2019). The alternative interpretation, supported by Bejarano et al. (2008) and others (Newman et al., 2011b; Newman et al., 2011a; Vimont et al.,

[1] *Central Pacific* El Niños have also been called Dateline El Niños, El Niño Modoki, or warm pool El Niño. Similarly, *Eastern Pacific* El Niños are sometimes termed conventional or cold tongue El Niño.

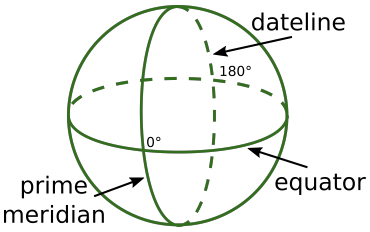


Figure 3.4.: The international dateline generally follows the 180° line of longitude and serves as the boundary where one calendar day ends and the next begins. At the equator, the dateline passes through the central Pacific.

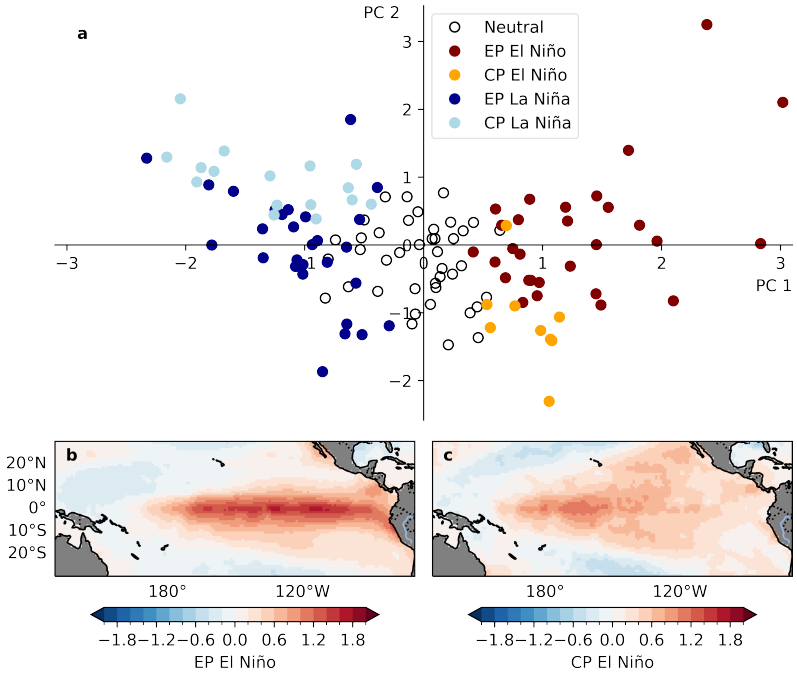


Figure 3.3.: ENSO events are projected on the first two EOFs of tropical Pacific SSTA (a). El Niño, La Niña, and Neutral events are identified by DJF average of Niño3.4 > 0.5, Niño3.4 < -0.5 and -0.5 < Niño3.4 < 0.5, respectively. Using the definition by Kug et al. (2009), El Niño and La Niña events are classified as CP when Niño3 > Niño4 and as EP for Niño4 > Niño3, highlighted by different colors. Averages over events classified as EP (b) and CP (c) El Niño exhibit different locations of peak SSTA.

2014; Vimont et al., 2022; Chen et al., 2015) describes ENSO as a linear system subject to stochastic forcing that exhibit different types through unique patterns of optimal growth. Despite the insights offered by both schools of thought, a comprehensive understanding of ENSO diversity remains elusive, with each interpretation revealing only part of the complex phenomena. Furthermore, the influence of external elements such as the Pacific's background state, interbasin interactions, and the impact of extratropical regions, shape ENSO diversity (Capotondi et al., 2020a).

In representing ENSO events within a lower-dimensional space, particularly through the PC1-PC2 plot in Fig. 3.3a, we observe not distinct regimes but rather a continuous distribution of events. This finding leads to a fundamental question: are EP and CP events distinct entities, or do they represent extremes of a continuous spectrum? This question, raised by Capotondi et al. (2015), will be addressed in Ch. 5.

3.1.3. ENSO in Data

In this section, I summarize how ENSO is characterized in data. Based on the mechanisms of ENSO, I will motivate and outline how ENSO is defined, and which variables and regions are relevant.

Definition

Since meteorological data is inherently high-dimensional, climate scientists reduce the dimensionality by representing complex phenomena like ENSO through a single time series or by grouping time points for creating representative maps. For ENSO, various indices have been established, using different variables and regions. These indices are broadly classified into two types: those describing ENSO's intensity and those characterizing its spatial variability.

Among the various measures for ENSO intensity, the Niño3.4 index is the most popular and widely used time series. The Niño3.4 index is obtained from the average SSTA in the Niño3.4 region (5°N-5°S, 170°W-120°W). While other region-specific indices such as Niño1+2 (0°-10°S, 90°W-80°W), Niño3 (5N-5S, 150W-90W), Niño4 (5N-5S, 160E-150W) are also in use, the Niño3.4 index represents average equatorial SSTs across the Pacific

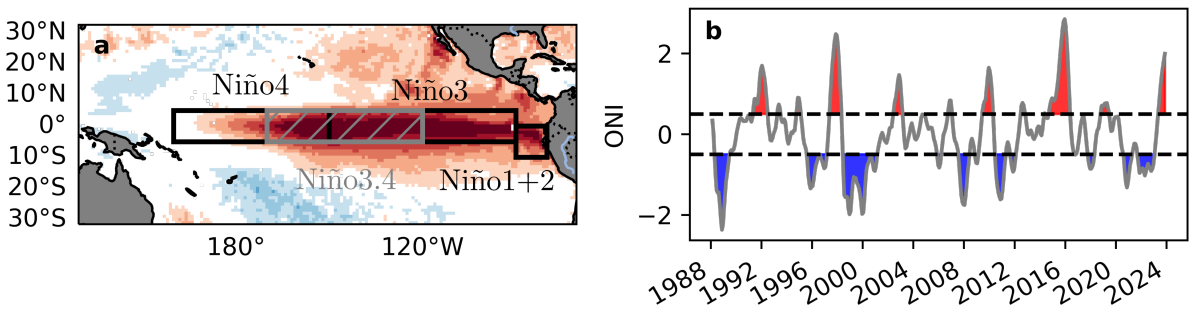


Figure 3.5.: The Niño1+2, Niño3, Niño4, and Niño3.4 index are defined as the average SSTA over the respective boxes (a). El Niño (La Niña) phase is defined as the Oceanic Niño Index (ONI) (b), obtained as the three-month running average of the Niño3.4 index, which is larger than 0.5°C (smaller than -0.5°C). SSTA are used from ORAS5 (ORAS5, 2021).

(Fig. 3.5)a. The Oceanic Niño Index (ONI), introduced by NOAA is based on the three-month running mean of the Niño3.4 index and explicitly defines that the climatology for the SSTA calculation uses a 30-year base period updated every 5 years (Fig. 3.5b). Besides the indices based on SSTA, the Southern Oscillation Index (SOI) provides a measure of ENSO intensity using the atmospheric variable sea level pressure. All these indices are highly correlated with each other and with the leading PCs of SSTA in the tropical Pacific (Fig. 2.2b).

Niño3–Niño4 Approach: Kug et al. (2009) and Yeh et al. (2009) proposed to use the Niño3 and Niño4 indices to differentiate between EP and CP El Niño events. EP events are identified when the boreal winter Niño3 index exceeds 0.5°C and is greater than the Niño4 index. Conversely, CP events are determined when the boreal winter Niño4 index exceeds 0.5°C and surpasses the Niño3 index.

ENSO phases are characterized by grouping time points based on a threshold. The most recognized definition, established by NOAA, identifies an El Niño (or La Niña) event when the ONI index exceeds 0.5 K (or falls below -0.5 K) for a minimum of five consecutive months (Fig.3.5b). Conditions are defined as Neutral when the ONI index lies between -0.5 K and 0.5 K. In numerical climate models, where SSTA variability can be larger, an alternative definition employs the standard deviation of the ONI index as the threshold for categorizing ENSO events.

Similarly, indices that track ENSO's spatial diversity have been established by comparing SSTA between different regions, such as between Niño3 and Niño4 (*Niño3-Niño4*) (Kug et al., 2009), or by considering the first two PCs of the tropical Pacific (*EC*)(Takahashi et al., 2011). For instance, the former identifies an El Niño event as EP when Niño3 is greater than Niño4, and as CP when the opposite is true (Fig. 3.3a). Comprehensive comparisons of different definitions are provided in the works of Yu et al. (2013), Capotondi et al. (2020a), and Karamperidou et al. (2017).

E and C Indices:: Defined by Takahashi et al. (2011), the E and C indices are based on the two leading PCs of SSTAs within the 10°S–10°N tropical Pacific band. The C index is calculated as $C = \frac{(PC1+PC2)}{\sqrt{2}}$, and the E index as $E = \frac{(PC1-PC2)}{\sqrt{2}}$. These indices are designed to be independent and effectively differentiate between moderately warm CP events ($C > E$) and extreme EP events ($E > C$).

Data Design

ENSO is a coupled ocean-atmosphere phenomenon with potential feedbacks globally (Sec. 3.2). To model ENSO one would ideally consider a large range of variables on high spatial and temporal resolution, to capture the global climate dynamics on different scales, and thereby also the dynamics of ENSO. Current physics-based global climate models (Sec. 3.3.2), pursue this comprehensive approach, however, besides their computational limitations, encounter a range of issues and biases. Data-driven models, on the other hand, are limited by the short observational record. These limitations necessitate design choices on the variables, regions, and resolution to use for capturing the complex dynamics of ENSO. While there is no conclusive selection, I will outline and motivate my data design choices for developing data-driven ENSO models, which is summarized in Tab. 3.1.

i) Variables:

SST is central to ENSO modeling due to its dual role in capturing ocean dynamics and integrating information from the fast-varying atmosphere. Since 1980 satellites have enabled almost real-time SST measurements of the whole ocean, while earlier data was derived from point measurements made by buoys and ships.

Probably the second most important predictor for ENSO is the subsurface ocean temperature within the top 300 meters. This parameter is often

Variable	Region	Temporal Resolution	Comments
Sea surface temperature (SST)	Pacific (Indian Ocean and Atlantic)	Monthly	
Sea surface height (SSH)	Pacific	Monthly	SSH is a proxy for thermocline depth but is observed from satellites.
10M zonal-component of wind (u10)	Pacific, especially the western Pacific	Daily	Alternatively, zonal wind stress.
Sea level pressure (SLP)	Pacific	Monthly	Captures the Southern Oscillation

Table 3.1.: Relevant variables, regions, and temporal resolution for data-driven models of ENSO.

described by the thermocline depth or the ocean heat content (OHC), obtained by vertically integrating the subsurface temperatures. The TOGA project (McPhaden et al., 1998) marked the beginning of systematic subsurface temperature measurements using buoys. Alternatively, sea surface height (SSH) exhibits a strong correlation with the thermocline depth (Rebert et al., 1985) and can be measured via satellite. For this reason, I use SSHA as a predictor for ENSO in Ch. 6.

While sea level pressure (SLP), associated with the Southern Oscillation, could be a relevant quantity, its strong correlation with SSTA often limits its additional predictive value. Surface winds, which are related to SLP as they blow from high to low-pressure regions, exhibit significant high-frequency variability. The high-frequency zonal winds in the western Pacific, known as westerly wind burstss (WWBs), is a stochastic forcing for ENSO, that can trigger El Niño and La Niña events through momentum transfer to the ocean surface (e.g. Vecchi et al., 2006; Capotondi et al., 2018). Consequently, the zonal wind component of surface winds, 10m zonal-component of wind (u10), is an important predictor of ENSO. Zonal wind stress, τ_x, which reflects the surface wind's force on the ocean may also be used. However, it is not directly accessible through satellite observations.

The seasonal nature of the ENSO requires focusing on anomalies independent of the variable. To calculate anomalies, we first remove a linear trend from the dataset, establish the climatological mean, and then subtract this mean from the data. In forecasting models, the method of computing climatology requires additional consideration to avoid introducing artificial skills. Specifically, the climatology used should not contain information unavailable at the time of forecast initiation. This issue can be addressed by employing a *fair-sliding* method (Risbey et al., 2021) for anomaly calculation.

Fair-sliding: Fair-sliding is a method to compute anomalies specifically suited for forecasting models. This involves using a 30-year sliding window for the climatology prior to the forecast initialization time. For instance, a forecast made in 2000 would use climatology from the 1969–1999 period, thereby excluding any post-1999 data that would not have been available initially. This approach ensures that our forecasts are based solely on relevant, contemporaneous data, avoiding any hindsight biases.

ii) Spatial Regions and Resolution:
The spatial region of warming and cooling SST spans large parts of the tropical Pacific. For this reason, the resolution of the data fields is less critical, with resolutions up to 2.5° generally sufficient to capture the primary spatial characteristics of ENSO.

Including the tropical Pacific, stretching from South America in the east to the Maritime Continent in the west [120°W-70°E], is crucial for describing ENSO. The impact of off-equatorial SSTA on ENSO, and consequently, the determination of the appropriate latitudinal range for inclusion, remains a subject of active research. Following insights from Richter et al. (2022), incorporating areas like the North Pacific Meridional Mode (NPMM)and the South Pacific Meridional Mode (SPMM), spanning 30°S - 30°N, should enhance model predictability.

North/South Pacific Meridional Mode (NPMM/SPMM:) The NPMM/SPMM is characterized by changes in the north-south SST gradient in the North/South East Pacific Ocean. This mode of variability impacts the atmospheric circulation and the trade winds, and thus affects ENSO.

While the focus has predominantly been on the Pacific, the potential roles of other ocean basins, notably the Indian and Atlantic Oceans, as predictors for ENSO are subjects of active research, hinting at possible future expansions in our understanding of ENSO dynamics (Sprintall et al., 2020).

iii) Temporal Resolution:
Given that ENSO is the predominant mode of interannual variability, substantial shifts in SSTA and SSHA primarily occur on a monthly timescale. This suggests that a monthly temporal resolution is adequate to encapsulate the necessary oceanic variability linked to ENSO dynamics. However, when it comes to atmospheric elements, a monthly resolution may not be sufficiently granular. Specifically, abrupt shifts in westerly trade winds are known to trigger El Niño and La Niña events. To effectively capture these critical changes, I recommend including zonal winds at a daily resolution (Tab. 3.1).

Data Sources

TOGA-project: Tropical Ocean-Global Atmosphere (TOGA) project, a ten-year international study (1985–1994) to monitor the ENSO phenomenon, significantly enhanced the observational infrastructure across the tropical Pacific. Key elements such as the TOGA-TAO moored array, which consists of 70 stationary buoys, provided real-time measurements of wind and ocean thermodynamics.

Observational Data: Since the 1980s, satellites have provided automatic and almost real-time measurements of oceanic surface variables. Prior to this, our insights came from station data gathered by buoys and ships. The *TOGA project* (McPhaden et al., 1998), and later the Argo Project (Feder, 2000), with its array of floating sensors, has further enhanced the information we get from the upper ocean layers. The primary advantage of observational data lies in its minimal biases. However, missing data points, the absence of gridded data for earlier records and subsurface ocean measurements, and a relatively short observational history are a strong limiting factor for applying data-driven models.

Reanalysis Data: Reanalysis data offers a solution to some of the limitations of observational data. They essentially represent interpolated observational data, laid out on a grid and compiled from diverse sources. This interpolation approach, called assimilation, uses a numerical model, which is nudged to the observations by using ensemble Kalman filter techniques. The most notable advantage of reanalysis data is the creation of a continuous, gridded dataset, devoid of missing data points. However, this approach bears its own set of challenges. Notably, the biases of the underlying numerical model, especially in regions where observational data are sparse, can skew results. Furthermore, the short observational record also limits the length of the reanalysis datasets.

Climate Model Data: Data from numerical coupled ocean-atmosphere models provide another data source. These models solve the governing

Reanalysis product	Resolution	Period	Variables
ERSSTv5 (Zhang et al., 2019)	$2° \times 2°$	1854-present	SST
ERA5 (Hersbach et al., 2020)	$0.25° \times 0.25°$	1940-present	SST, τ_x
COBE2 (COBE, 2006)	$1° \times 1°$	1891-present	SST
HadISST (Rayner et al., 2003)	$1° \times 1°$	1870-present	SST
ORAS5 (ORAS5, 2021)	$0.25° \times 0.25°$	1958-present	SST, SSH, τ_x
SODA (Giese et al., 2011)	$0.25° \times 0.25°$	1980-present	SST
GODAS (Behringer et al., 1998)	$0.33° \times 1.0°$	1980-present	SST, SSH, τ_x
CERA-20C (Laloyaux et al., 2018)	$1° \times 1°$	1901-2009	SST, SSH, τ_x

Table 3.2.: Reanalysis products. A non-comprehensive list of reanalysis products of ocean variables. The products contain one or more variables: SST, SSH, and surface wind stress (τ_x). The listed datasets are used in the analysis in Ch. 5.

physical equations on a gridded duplicate of our Earth system. A significant advantage of using climate model data is the sheer volume of data available. These models can run for over a hundred years, generating numerous ensemble members that offer insights into uncertainties and potential future scenarios, including conditions like a 4-degree warmer climate. Especially, the standardization of variables and scenarios for CMIP simulations is a great source of openly available data. However, climate models often exhibit large biases due to their resolution and the absence or approximation of certain processes. For instance, SSTA in the Pacific shows exaggerated variability in most global climate models (e.g. Capotondi et al., 2020a; Beverley et al., 2023), which in turn, is reflected in the data-driven models based on these data.

3.2. Why Do We Care About ENSO?

Changes in heat transfer between the ocean and the atmosphere during the different ENSO phases alter the atmospheric circulation and affect weather patterns far beyond the tropics. Hydrological extremes such as floods and droughts are common with ENSO and are expected to increase with global warming (Trenberth, 2011). For instance, the strong 2015–2016 El Niño event triggered widespread coral bleaching (Hughes et al., 2017) and led to anomalously warm winters in Canada and the northeastern United States (e.g. Taschetto et al., 2020). Similarly, South America often experiences significant rainfall increases during El Niño phases, leading to flooding in countries like Ecuador and Peru. Furthermore, the frequency and intensity of tropical cyclones are substantially enhanced by La Niña (e.g. Taschetto et al., 2020).

These far-reaching effects of ENSO, known as teleconnections, highlight ENSO's essential role in global seasonal weather predictability. For

instance, teleconnections allow estimating likelihoods of increased or decreased precipitation in the Midwest of the US months ahead. Such forecasts are critical for decision-making in sectors like agriculture and water management. Callahan et al. (2023) estimated that in the decade following a potent El Niño event, the global economy could incur costs of up to 84 trillion dollars. Fortunately, the El Niño of 2023-24 manifested with only moderate intensity.

Rossby waves: Rossby waves, also known as planetary waves, occur in rotating fluids due to the conservation of potential vorticity that changes with latitudes. In the atmosphere, they are large-scale meanders in high-altitude winds that transport heat between the equator and the poles.

The primary mechanism driving ENSO's atmospheric teleconnections involves shifts in convective activity, due to changes in SSTAs. These shifts trigger atmospheric waves, called *Rossby waves*, that propagate ENSOs' signal to distant regions. Numerous teleconnections for both El Niño and La Niña have been documented, as outlined in the chapter by Taschetto et al. (2020), which is not reiterated here. Instead, I will present the typical approach used to determine ENSO teleconnections.

Identifying Teleconnections

Identifying teleconnections in climate science is essentially a pattern recognition task, like noticing trends of increased rainfall that coincide with El Niño conditions. The standard methodology in climate science discovering these links involves the following steps:

1. **Pattern Extraction:** Commonly, scientists start by examining EOFs or correlations. For example, to investigate precipitation patterns in California, one would analyze weather station data for high correlations with indices that represent global circulation patterns, such as the ONI index for ENSO.
2. **Causality Assessment:** Beyond identifying correlations, it's crucial to establish causation. This is usually done by grouping time periods according to correlation intensity and examining composite variables like atmospheric pressure. Such analysis can reveal underlying propagation pathways, potentially involving Rossby waves, jet streams, or other climatic factors.
3. **Climate Modeling:** After formulating a possible mechanism, it can be tested in climate models. This often involves "pacemaker experiments", where for instance SSTs in the Pacific are prescribed to suppress typical El Niño conditions, allowing scientists to observe the effect on the proposed teleconnection.

This methodology, while comprehensive and demanding thorough data analysis, statistical validation, and complex climate modeling, is a standard approach for identifying teleconnections in climate science. A crucial initial step in this process is the careful selection of the region or phenomenon of interest. In Ch. 4, climate network analysis is used to identify global teleconnections of EP and CP El Niño events.

3.3. How Well Can We Predict ENSO?

Given ENSO's large impact on weather and climate conditions worldwide, it is an important source of predictability for long-range weather forecasts, specifically on seasonal to annual scales. The effort to improve long-range ENSO forecasting began following the devastating 1982-83 El

Niño event, which triggered severe drought and heatwaves in Australia (Taschetto et al., 2009), highlighting the need for reliable long-term rainfall predictions. Subsequent initiatives, like the TOGA program, have led to the development of models that accurately simulate ENSO dynamics. Currently, the International Research Institute for Climate and Society (IRI) provides a list of operational ENSO forecasting models (IRI, 2024), demonstrating forecast skills for the Niño3.4 index up to 18 months in advance (Zhou et al., 2023). The IRI distinguishes between dynamical and statistical models, which I refer to as physics-based models and data-driven models. While models of the first type describe the physical mechanisms of the system explicitly with differential equations, the latter infers the dynamics from the data.

3.3.1. Predictability

Lorenz's seminal work in 1963 and 1969 on the concept of the 'butterfly effect', has shaped our understanding of the weather as an inherently chaotic system. The sensitivity of the weather system to initial condition perturbations and the interaction between small and large spatial scales limits the predictability of weather and climate (Palmer et al., 2014). For instance, in the mid-latitudes information from the atmospheric initial state decays in less than two weeks. For this reason, weather forecasting is not reliable beyond 14 days. Given these constraints, how is it possible to generate seasonal forecasts for ENSO?

Several factors contribute to the predictability of ENSO. Unlike in the mid-latitudes, the tropics allow for a slower decay of atmospheric initial state information, extending up to about 20 days (Judt, 2020). Moreover, ENSO has an oceanic component and the ocean exhibits a slower rate of variability compared to the atmosphere. The ocean effectively acts as a low-pass filter on the atmospheric dynamics. The most important factor, however, is the large-scale and quasi-periodic characteristics of ENSO which substantially reduces its sensitivity to the chaotic component of the atmosphere, allowing it to be predicted up to two years in advance (e.g. DiNezio et al., 2017).

Despite ENSO's predictability on longer timescales, uncertainties remain that constrain its forecast accuracy, namely:

1. **Initial state uncertainties:** Even minor inaccuracies in determining the current oceanic and atmospheric states can amplify over time, resulting in significant forecast errors.
2. **Model uncertainties:** Models often lack key physical processes or represent them imprecisely, usually due to their occurrence at smaller spatial or temporal scales than what model resolutions can capture.

Besides these general limitations of predictability, ENSO forecasting skill exhibits seasonal fluctuations, present in both physics-based and data-driven models. Forecasts initialized in the boreal spring show significantly reduced skill called the *spring predictability barrier* (Duan et al., 2013). This phenomenon is attributed to the strongly damped SSTA in the equatorial cold tongue during this period (Jin et al., 2007; Stein et al., 2014). Furthermore, ENSO predictability shows decadal variations, a pattern consistent across different models. Hypotheses suggest that

Spring predictability barrier: The term "spring predictability barrier", denoting the reduced forecasting skill of ENSO during the boreal spring, might be somewhat misleading, as predictions initialized in the preceding winter often exhibit higher accuracy. Therefore, the phenomenon should rather be called "spring predictability minimum".

these variations may be due to the nature of ENSO events and associated changes in atmospheric forcing (McPhaden, 2012; Neske et al., 2018) or alterations in the ocean's background state (Capotondi et al., 2017; Lou et al., 2023).

In addition to the decadal variability in forecasting skills, certain ENSO states demonstrate higher predictability than others. Notably, forecasts following strong El Niño events tend to be more skillful, as these conditions often transition into La Niña in the subsequent year, enhancing predictability (Gonzalez et al., 2016; DiNezio et al., 2017; Lou et al., 2023). These periods of enhanced forecast accuracy are often termed "windows of opportunity".

Other sources of ENSO predictability are a current research focus, involving interbasin interactions, notably the Indian Ocean and the Atlantic Ocean, as well as influences from the extratropics (e.g. L'Heureux, 2020).

3.3.2. Physics-Based Models

ENSO forecasting accuracy is hindered not only by uncertain initial conditions but also by the often limited representation of dynamic processes in models. This section provides an overview of physics-based ENSO models and discusses their advances and limitations. Physics-based models have significantly contributed to our understanding of ENSO dynamics over the centuries. They rely on differential equations to simulate physical processes and require numerical methods for solutions. These models are categorized hierarchically based on their complexity: (i) simple oscillators, which capture the cyclic nature and key parameters of ENSO; (ii) intermediate models, focusing on the fluid dynamics and thermodynamics of the equatorial ocean and atmosphere with certain simplifications; and (iii) Global Circulation Model (GCM), offering detailed global climate simulations utilizing advanced computing resources.

Oscillator Models

The simplest ENSO models are harmonic oscillators constructed using ordinary differential equations. These 1D models capture ENSO's oscillatory nature, characterized by periods ranging from 2 to 7 years, and describe the dynamics of the spatial average SSTA and thermocline depth in the tropical Pacific. The delayed oscillator (Suarez et al., 1988), and the *recharge/discharge oscillator* model (Jin, 1997) are notable examples. These models exhibit rapid positive feedback leading to the growth of El Niño and delayed negative feedback, which reverses the oscillation's phase. The positive feedback represents the Bjerknes feedback (Sec. 3.1.1) in both models. In contrast, the delayed oscillator captures the negative feedback through delayed equatorial oceanic wave propagation and reflection. The recharge/discharge model, meanwhile, interprets the negative feedback via the meridional transport of equatorial heat content (Jin et al., 2020). Subsequent advancements in these oscillator models have incorporated nonlinearities (Timmermann et al., 2003; Guckenheimer et al., 2017), integrated atmospheric interactions through stochastic forcing (Jin et al.,

Recharge/discharge oscillator: In its simplest form, the recharge/discharge oscillator reads

$$\frac{dT}{dt} = aT + bH$$
$$\frac{dH}{dt} = cH + dT,$$

where T is a spatial average of SSTA, H is a spatial average of thermocline depth anomalies, and a, b, c, d are fixed parameters.

2007; Levine et al., 2010; Bianucci et al., 2018), and explored shifts in the Walker Circulation (Thual et al., 2023).

These simplistic models have formed much of our mechanistic understanding of ENSO. They offer intuitive illustrations of its fundamental aspects, including causes and evolution, and allow testing of new physical hypotheses. However, their practical application in forecasting is limited. This limitation stems primarily from their low-dimensional structure and lack of interaction with other climate components, like the atmosphere. While important for theoretical understanding, these models do not fully capture the complex behavior of ENSO in observations.

Intermediate Complexity Models

Following simple oscillator models in the hierarchy of ENSO models are the earth model of intermediate complexity (EMIC). EMICs are two-dimensional models, depending on variations in longitude, latitude, and time, while their vertical structure is fixed in both the ocean and atmosphere. These models are typically based on the shallow water equation for the oceanic component which distinguishes a dynamic upper layer from the passive deeper ocean, separated by a stable thermocline. Similarly, atmospheric dynamics in EMICs are also often represented through shallow water equations, considering the atmosphere as a fast-responsive layer compared to the ocean (Gill, 1980). A prominent example of EMICs for ENSO is the Cane-Zebiak model (Cane et al., 1986; Zebiak et al., 1987), along with its evolved variants summarized in the Lamont-Doherty Earth Observatory model (LDEO) (Gao et al., 2020) that are used for operational ENSO forecasting[2].

[2]Operational forecast of the LDEO model: https://iri.columbia.edu/our-expertise/climate/forecasts/enso/archive/200907/models/Lamont_Doherty.html

The advantage of EMICs lies in their computational efficiency and complexity, allowing sensitivity experiments and control over the simulated ENSO dynamics. For instance, conducting a century-long simulation with the Cane-Zebiak model on a single CPU can be completed in less than two hours. This allows for conducting extensive ensemble experiments over long durations, yielding robust statistics of ENSO. Furthermore, EMICs can be initialized with observational data, making operational forecasts possible. Remarkably, their forecasting skill is on par with more complex models, including comprehensive GCMs (Barnston et al., 2012), which are described in the next section.

Nevertheless, the accuracy and efficiency of EMICs heavily rely on the parametrization of unresolved processes. These empirical approximations limit the models' ability to simulate diverse climate scenarios. The LDEO, for example, primarily models the tropical Pacific and is unable to fully represent the spatial diversity of ENSO (e.g. Geng et al., 2022).

Coupled Ocean-Atmosphere Models

Coupled ocean-atmosphere models that solve the *primitive equations* of fluid dynamics and thermodynamics across multiple vertical pressure levels globally are the gold standard for weather and climate prediction. This category includes Global Circulation Model (GCM) and numerical weather prediction (NWP). While GCMs and NWPs are designed for different objectives — weather models focus on initial value problems

and climate models on boundary value problems — they solve the same physical dynamics of the earth system. Both discretize the ocean and atmosphere into boxes and numerically solve the governing differential equations. GCMs used for ENSO modeling need to account for slow-varying factors like the cryosphere, biosphere, ocean currents, and volcanic activity. These models usually run for a hundred to several hundred years and simulate different climate scenarios by changing their boundary conditions like greenhouse gases and solar activity.

Primitive equations: The primitive equations of ocean/atmosphere dynamics are:

i. **Continuity equation:** Representing the conservation of mass.
ii. **Conservation of momentum:** Consisting of the Navier–Stokes equations that describe hydrodynamical flow on the surface of a sphere under the hydrostatic assumption.
iii. **Thermal energy equation:** Relating the overall temperature of the system to heat sources and sinks.
iv. **The Equation of State for Air/Water:** Links pressure, temperature, in the atmosphere or salinity, temperature, and pressure in the ocean to the density of air/seawater.

Due to their finite resolution, GCMs approximate the impact of small-scale processes that are not resolved on the larger grid. This approximation, known as parametrization, often relies on heuristic or empirically derived models (Palmer, 2001). Notable examples include the parametrization of convection and cloud formation, processes that occur on a kilometer-scale but significantly alter the energy balance — the difference between incoming and outgoing radiation — of the larger grid boxes (e.g. Arakawa, 2004). These parameterizations represent the primary sources of error in GCMs (e.g. Brenowitz et al., 2018; Rasp et al., 2018).

Since the early 2000s, GCMs have begun to generate realistic simulations of ENSO (Guilyardi et al., 2020). An important factor in these improvements has been the availability of observational data from TOGA (McPhaden et al., 1998) and the Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016). Today, a wide range of models contribute to ENSO forecasting, with the ensemble average from the North American Multi-Model Ensemble (NMME, Kirtman et al., 2014) representing the state-of-the-art in predictive capability.

The primary advantage of using GCMs for ENSO modeling is their ability to generate realistic dynamics of the tropical Pacific that capture the complex interactions of ENSO with other global processes. These models produce extensive simulations with multiple ensemble members, providing insights into associated uncertainties. Furthermore, GCMs generate all relevant variables, allowing researchers to trace mechanisms that emerge from complex interactions between them. Despite their high computational cost, GCMs enable specialized experiments, such as nudging techniques where experts can investigate the atmospheric responses to specific temperature conditions in the tropical Pacific (Fedorov et al., 2000).

[3]For the Coupled Earth System Model 2 (CESM2), for instance, to run one year of global simulations on a resolution of 100×100 km requires 3456 CPU-hours (Danabasoglu et al., 2020).

However, the GCMs also present large challenges for ENSO modeling. (i) Due to their complexity, GCMs can appear as "black boxes". Users can only analyze model outputs rather than interact directly with the models. (ii) Running GCMs are computationally expensive[3], requiring large supercomputers. This also limits the ability to generate large ensembles for operational forecasts. (iii) GCMs exhibit notable biases, such as the eastward displacement of warm pool anomalies and overestimation of SST anomalies in the tropical Pacific, as seen in the latest CMIP5 and CMIP6 models (Capotondi et al., 2020a; Beverley et al., 2023). These biases are challenging to address due to the intricate ocean-atmosphere interactions and the models' limited resolution. Efforts to refine the physical representation within models sometimes lead to poorer performance, likely due to the interaction of parametrizations and the chosen validation standards.

3.3.3. Data-Driven Models

With the availability of observational data from satellites (Sec. 3.1.3), data-driven models have entered the field of sub-seasonal to seasonal forecasts, and thereby also the task of ENSO prediction. In contrast to physics-based models (Sec. 3.3.2) data-driven models do not rely on solving differential equations describing the known physical processes, but infer its dynamics from data (see Sec. 2.2 for a general introduction into data-driven forecasting models). This section provides an overview of data-driven models for ENSO forecasting which motivates the hybrid model approach in Ch. 6.

The first operational data-driven models for ENSO forecasting were based on multivariate linear statistical techniques, such as canonical-correlation analysis (e.g. Barnett et al., 1988), principal oscillation pattern analysis (e.g. von Storch et al., 1990), and Markov models (e.g. Xue et al., 1994). These models predicted ENSO states with lead times of six to twelve months (Latif et al., 1994). The Linear Inverse Model (LIM) (Penland et al., 1995), derived from the Fokker-Planck equation, is a probabilistic linear model that describes the linear dynamics of the ocean under stochastic atmospheric forcing, approximated as white noise (see Ch. 6 for a detailed description). Subsequent enhancements to these linear models have incorporated seasonality (Shin et al., 2021), state-dependent noise (Martinez-Villalobos et al., 2018), and extended historical data (Chapman et al., 2015). Furthermore, Chen et al. (2015) introduced nonlinear elements by including higher-order polynomials.

Another early data-driven approach involves analog methods, initially proposed by Lorenz (1969) for weather forecasting. These methods identify historical states similar to the current state and use their subsequent evolution to forecast future conditions. Barnston et al. (1989) applied this technique to ENSO forecasting; however, the limited length of the observational record limited the prediction accuracy (Dool, 1989). More recently, the introduction of "model-analogs" — analogs derived from GCM simulation runs — has enabled skillful ENSO forecasts extending beyond 12 months (Ding et al., 2018; Wang et al., 2020; Toride et al., 2024).

While analogs require explicit structuring of the data, neural networks learn the structure in the data implicitly. First applied to ENSO forecasting in the 1990s (Tangang et al., 1997), neural networks initially did not show an improvement in skill upon linear models due to the small network sizes and sparse data availability. However, with observations from satellites and increased computational power, Ham et al. (2019) demonstrated that a CNN could predict the Niño3.4 index up to 17 months in advance, surpassing both data-driven and physics-based models. This seminal work has sparked extensive research on deep learning-based ENSO forecasting. Tab. 3.3 provides a non-comprehensive overview, highlighting differences in input variables, data sources, objectives — fixed lag-time models or recurrent models — and whether they predict a single index or forecast entire fields. The maximum lag time, where the ACC of the Niño3.4 index exceeds 0.5, is a commonly reported metric for these models. However, comparisons should be taken with caution due to differences in the reanalysis products and time periods used for evaluation.

Table 3.3.: Deep learning-based ENSO forecasting models show differences in input variables, data sources, model objectives, architectures, and skills. For the model objective, I differentiate between models that produce the forecast recurrently and models that forecast fixed lag times, as well as, whether the forecast objective is the Niño3.4 time series or the full spatial fields. The most frequently reported measure of skill is the maximum forecast lag-time, where the anomaly correlation coefficient (ACC) of the Niño3.4 index exceeds 0.5. This list of forecasting models is not comprehensive.

Paper	Variables	Sources	Objective	Architecture	Skill [month] (ACC>0.5)
Ham et al. (2019)	SSTA, OHC	CMIP5, SODA, GODAS	Lag-time of Niño3.4	CNN	17
Cachay et al. (2021)	SSTA, OHC	CMIP5, SODA, GODAS	Lag-time of Niño3.4	GNN	15
Zhou et al. (2023)	SSTA, τ_x, τ_y, OHC	CMIP6, SODA, ORAS5	Lag-time of full field	Transformer	18
Wang et al. (2023)	SSTA, OHCA	CMIP5/6, SODA, GODAS	Lag-time of Niño3.4	CNN & Attention	24
Lyu et al. (2023)	SSTA	CMIP6, ERSSTv5	Lag-time of Niño3.4	CNN & Attention	18
Taylor et al. (2022)	SSTA, T2MA	ERA5	Recurrent of full field	U-Net & LSTM	-

Neural networks are nowadays considered state-of-the-art in ENSO forecasting (e.g. Ham et al., 2019; Zhou et al., 2023) because they are capable of learning dynamics that are unresolved in physics-based models. Due to their fast inference, they are also well-suited for operational forecasts, avoiding the slow integration steps required by numerical models. However, to date, most approaches predict only a single index and their forecasts are deterministic (Tab. 3.3). Additionally, the available observational monthly data is too short for training large neural networks (Ch. 6). Long simulations from GCMs suggest that at least 200 to 300 years of data may be necessary to robustly capture changes in ENSO properties (Wittenberg, 2009), necessitating the use of GCM simulations for training neural networks. When trained on these simulations, neural networks inherit the biases of the GCMs. Attempts to bridge the domain gap between simulations and observations, such as fine-tuning models on observational data that have been pre-trained on GCM simulation, have been applied (Ham et al., 2019), but as Zhou et al. (2023) report, these methods do not always lead to improved skill.

Part II.

Research Contributions

Teleconnection Patterns of Different El Niño Types Revealed by Climate Network Curvature

4.

Abstract

The diversity of El Niño events is commonly described by two distinct flavors, the Eastern Pacific (EP) and Central Pacific (CP)-type. While the remote impacts, i.e., teleconnections, of EP and CP events have been studied for different regions individually, a global picture of their structure is still lacking. Here, we use Forman-Ricci curvature applied on climate networks constructed from surface air temperature data to distinguish regional links from teleconnections. Our results confirm that both El Niño types influence the teleconnection patterns, however, with different spatial manifestations. Our analysis suggests that EP El Niños alters the general circulation which changes the teleconnection structure to primarily tropical teleconnections. In contrast, the teleconnection pattern of CP El Niños shows only subtle changes to normal conditions. Moreover, this work identifies the dynamics of the Eastern Pacific as a proxy for the remote impact of both El Niño types.

4.1. Introduction

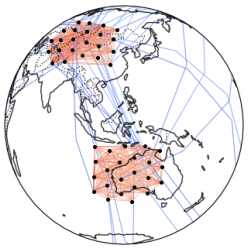
The El Niño-Southern Oscillation (ENSO) is the most dominant interannual variation in the global climate system. It is a dynamical atmospheric and oceanic phenomenon characterized by anomalously warm (El Niño) or cold (La Niña) phases of sea surface temperature (SST) in the equatorial Pacific. Both phases are known to impact Earth's climate significantly on large spatial scales, typically referred to as teleconnections (Trenberth, 1997) and thus have been investigated in many studies over the past two decades (Capotondi et al., 2015; Timmermann et al., 2018; Capotondi et al., 2020a).

Yet, significant differences in the downstream impacts of El Niño events are reported (Shi et al., 2019), depending on the amplitude and spatial position of SST anomalies. Impacts can range from devastating floods and droughts during strong El Niño events to only mild increases in rainfall and temperature changes during moderate to weak El Niños.

These differences can be partly related to the type of El Niño. The diversity of El Niño events is typically characterized by two modes: The "canonical" or Eastern Pacific (EP) El Niño (Rasmusson et al., 1982) with peak SST anomalies in the eastern equatorial Pacific, and the "El Niño Modoki" (Ashok et al., 2007) or Central Pacific (CP) El Niño with peak SST anomalies in the central equatorial Pacific (Kao et al., 2009).

Although the effect of both El Niño types on different locations of the Earth — such as the Indian Ocean (IO) (e.g. Klein et al. (1999)), maritime continent (e.g. Wang et al. (2007)), tropical Atlantic (e.g. Huang (2004)), and Northern America (e.g. Yu et al. (2012)) — has been studied thoroughly (see Okumura (2019) and Taschetto et al. (2020) for an

Code and experiments available at the Github repository:
jakob-schloer/netcurvature



overview), previous work has mainly focused on single teleconnections of the El Niño types.

In comparison, little is known about differences in the spatial extent of global teleconnection patterns between EP and CP events. In this study, we address this issue by introducing a novel machine-learning approach that employs complex climate networks combined with Ricci-curvature, an abstract tool from the theory of complex networks. Climate networks (Dijkstra et al., 2019) have gained increasing interest in the analysis of spatial dependencies of climatic variables through their ability to reduce data to relevant climatic patterns, and therefore, have been widely used in the analysis of ENSO. Tsonis et al. (2008) investigated the topology of El Niño and La Niña networks of surface air temperature. The global impact of El Niño on various geographical zones (Yamasaki et al., 2008), on geographical long-range teleconnections (Donges et al., 2009a; Zhou et al., 2015), and its diversity (Radebach et al., 2013; Kittel et al., 2021) have been analyzed by evolving climate network analyses. Wiedermann et al. (2016) employ them to find a robust way to distinguish between different types of El Niños flavors and Lu et al. (2020) estimate the expected El Niño impacts by using climate networks.

Ricci-curvature of complex networks is a recent approach to visualize the structure of a network intuitively by highlighting whether an edge of the network connects nodes within a community (i.e., a group of densely connected nodes) or bridges two communities (Ollivier, 2009; Sreejith et al., 2016). It has been proven useful, for example in the analysis of financial markets (Sandhu et al., 2016), gene expressions (Sandhu et al., 2015; Pouryahya et al., 2018), brain connectivity (Farooq et al., 2019), urban transportation (Gao et al., 2019), power grids (Jonckheere et al., 2019), and epidemiology (Souza et al., 2021).

We show that our approach is an intuitive yet informative tool to analyze the spatial organization of teleconnections that outlines structural differences between EP and CP El Niño impacts.

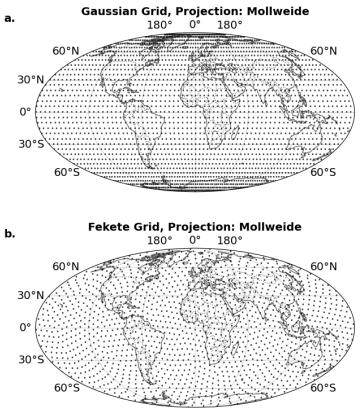


Figure 4.1.: Gaussian Grid and Fekete Grid. In the Gaussian grid (a) points at the poles are much closer to each other than at the equator. To avoid artificially higher correlation values at the poles due to the shorter geographical distances, we interpolate the data on the Fekete grid (b) with approximately uniformly distributed grid points. For visual reasons, we show a grid resolution of 5° instead of the 2.5° resolution used in the analysis of this work.

4.2. Data and Methods

4.2.1. Data

We use daily surface air temperature (SAT, 2-meter air temperature) data for the years 1959–2020 from the ERA5 Global Reanalysis database (Hersbach et al., 2020). We first detrend each time series, then subtract the daily climatology of the whole time period resulting in anomaly time series with respect to the day of the year. We use next-neighbor interpolation to map the data to a grid of spatially approximately uniformly distributed points using the Fekete algorithm (Bendito et al., 2007) to avoid spurious correlation patterns close to the poles (Ebert-Uphoff et al., 2012) (Fig. 4.1). The distance between grid points in the Fekete grid corresponds to the distance between two points at the equator of a 2.5° Gaussian grid, resulting in a total of ≈ 6000 grid points.

4.2.2. Classification of EP and CP El Niño Conditions

We use Dec–Feb daily SAT anomalies and select EP El Niño events, CP El Niño events, and 'Normal' winters based on the average DJF sea surface temperature anomalies in the Niño 3 (N3) and Niño 4 (N4) region (Trenberth et al., 2001). We classify a winter as EP (CP) event if N3 is greater (less) than N4 and N3 (N4) larger than 0.5 (Capotondi et al., 2020a). Winters with N3 and N4 between -0.5 and 0.5 are labeled as 'Normal'. The SST anomalies are calculated using a centered sliding-window 30-year base period, successively updated in 5-year steps. Although the onset and duration of El Niño events show high variability, we restrict our analysis to the Dec–Feb period where El Niño events show in general the highest intensities to avoid seasonality effects in our analysis.

4.2.3. Ricci Curvature of Correlation-Based Climate Networks

We demonstrate the process of computing the curvature of a climate network using a toy dataset. Fig. 4.2 a denotes a set of time series distributed over a sphere. Correlations between all pairs of time series are *predefined* (in contrast to our actual SAT dataset) with *a priori* fixed covariance structure created from a stochastic block model.

Toy dataset: We use a stochastic block model (SBM) to generate a random graph with four equal-sized communities. We set their edge probabilities to

$$P = \begin{bmatrix} 0.7 & 0.02 & 0.02 & 0.03 \\ 0.02 & 0.6 & 0.02 & 0.01 \\ 0.02 & 0.02 & 0.6 & 0.02 \\ 0.03 & 0.01 & 0.02 & 0.7 \end{bmatrix}.$$

The adjacency matrix $A_{\text{toy}} \in \mathcal{R}^N$ of the SBM graph are used to specify the covariance structure of a multivariate Gaussian, $\mathcal{N}\left(x, 0, A_{\text{toy}}\right)$, with N is the number of nodes/data points. We draw $T = 1000$ samples from the Gaussian which forms the N time series of length T of our toy dataset $X_{\text{toy}} \in \mathcal{R}^{NxT}$ with known covariance structure.

In order to calculate the Ricci curvature, we first need to create a climate network, which signifies the most related grid points over the entire globe. The network is determined by considering the 2% strongest ($\rho_{0.98}$) statistically significant correlations ρ_{ij} between pairs of time series $x_i(t), x_j(t)$ which is described by the weighted adjacency matrix $\mathbf{W}_{ij}$ (Fig. 4.2 b):

$$\mathbf{W}_{ij} = \begin{cases} |\rho_{ij}|, & |\rho_{ij}| > \rho_{0.98}\ , \\ 0, & \text{otherwise}\ . \end{cases} \tag{4.1}$$

This is equivalent to a network G whose nodes $V = \{v_i : i = 1, .., N\}$ correspond to spatial locations $i = 1, \ldots, N$ with edge e_{ij} connecting the nodes $(v_i, v_j) \in V$ if $|\rho_{ij}| > \rho_{0.98}$ (Fig. 4.2 c) weighted by ρ_{ij} (for details see Sec. A.1).

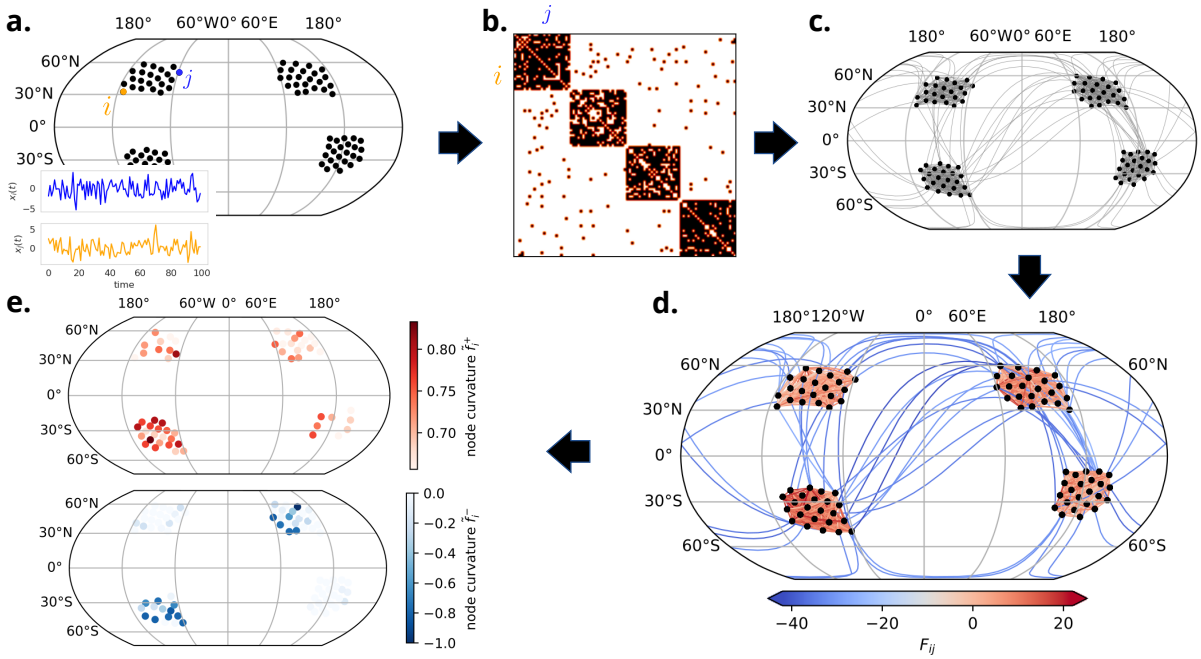


Figure 4.2.: Construction of a climate network and its curvature measure for a toy dataset. By computing correlations of time series between all pairs of locations (a), we obtain the adjacency matrix $\mathbf{W}_{ij}$ (eq. 4.1), orange dots denote points of significant correlations between time series $x_i(t), x_j(t)$ (b), visualized as a network graph (c). The Forman curvature F_{ij} of each edge (d) reveals the property of an edge as within-community (positive curvature) or bridging communities (negative curvature). Positive (negative) node curvature hotspots f_{ij}^+ (f_{ij}^-) (e) are obtained by aggregating the 10% most positively and 10% most negatively curved edges on the nodes.

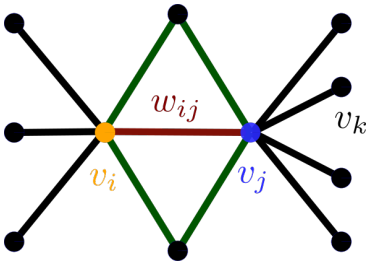


Figure 4.3.: Forman-curvature of an edge e_{ij} with weight w_{ij} connecting nodes v_i and v_j is obtained by Eq. 4.2.

Forman curvature: The Ricci-curvature of a network link describes how the connectivity of its network neighborhood differs from the connectivity of a regular grid. Forman curvature of edge e_{ij} in an undirected network with weight $w_{ij} \in \mathbf{W}$ is estimated as,

$$F_{ij} = w_{ij}\left(|\mathcal{T}_{ij}| \cdot w_{ij} + \frac{2}{w_{ij}} - \sum_{\substack{k=1 \\ w_{ik}w_{kj}=0}}^{N} \sum_{\substack{l\in\{i,j\} \\ w_{lk}>0}} \frac{1}{\sqrt{w_{ij} \cdot w_{lk}}}\right), \tag{4.2}$$

where $\mathcal{T}_{ij} := \{v_k : w_{ik}w_{jk} > 0\}$ denotes the set of nodes in the neighborhood of v_i and v_j which form triangles containing edge e_{ij} (green edges in Fig. 4.3) and $|\cdot|$ denotes set cardinality. The last term in Eq. 4.2 counts the number of edges adjacent to node v_i and v_j which do not form triangles with edge e_{ij} (black edges in Fig. 4.3). Equation 4.2 approximates the "augmented" Forman curvature (cf. Samal et al. (2018), Eq. 9) by considering only triangles, no node weights, and no cycle weights.

Ricci curvature provides a continuous measure over network links (i.e., independent from the spatial grid) using the connectivity between points to describe their relationship in the network. Fig. 4.2 d shows the network of our toy example with their edges e_{ij} colored by their curvature F_{ij}. We see that positive curvature indicates an edge being embedded within a community of nodes. The neighborhood of a positively curved edge is more densely connected than a regular graph which is a network with each node containing the same number of neighboring nodes. Positive

curvature thus indicates a community of nodes with similar dynamics. Negative curvature corresponds to an edge connecting two communities, with its own neighborhood containing fewer connections than a regular graph. It thus can indicate teleconnections as shown in Sec. 4.3.1.

Since multiple edges are difficult to visualize ($\approx 10^6$ links for the SAT network), we define node curvature f_i of the node v_i as the summation of the edge-based measure F_{ij} of all edges e_{ij} connected to v_i. The node curvature is easier to visualize on a map and allows identifying geographical locations connected to strongly negatively or positively curved links.

More precisely, we define the upper 90th (lower 10th) percentile of all edges as F_{ij}^{+} (F_{ij}^{-}) and analogously, the aggregation of the upper 90th (lower 10th) percentile of the curvature values as f_i^{+} (f_i^{-}), which we denote as 'hotspots' (Fig. 4.2 e). As the value ranges differ between networks, we normalize them by using the min-max transformation to $(-1, 1)$, denoted by $\tilde{F}_{ij}$ and analogously, $\tilde{f}_i$.

The intensity of the positive node-curvature hotspots $\tilde{f}_i^{+}$ reveals the communities with the highest edge probabilities in the stochastic block model outlining a community of nodes that behave similarly in time. Negative node-curvature hotspots $\tilde{f}_i^{-}$ coincide with the underlying probabilities used for constructing the dataset indicating locations that are highly teleconnected on the globe.

Of the two numerical approximations of Ricci curvature on networks — Forman-Ricci curvature (Forman, 2003; Sreejith et al., 2016) and Ollivier-Ricci curvature (Ollivier, 2010) — we use Forman-Ricci curvature (henceforth simply Forman curvature) as it is computationally cheaper. Both definitions are highly correlated, barring slight differences in extreme values (Samal et al., 2018).

Node curvature: The Forman-curvature of a network node v_i, is defined as

$$f_i := \frac{\sum_j F_{ij}}{k_i}, \tag{4.3}$$

with F_{ij} is the edge-based measure of all edges connected to that node, normalized by the number of edges k_i.

Positive/Negative curvature: We define the 10% most positive (F_{ij}^{+}) and most negative F_{ij}^{+} link curvature, as

$$F_{ij}^{+} := \{F_{ij} : F_{ij} > Q_F(0.9)\}, \tag{4.4}$$

$$F_{ij}^{-} := \{F_{ij} : F_{ij} < Q_F(0.1)\}. \tag{4.5}$$

4.3. Results

4.3.1. Spatial Organization of Teleconnections Depends on El Niño Type

We compute correlation-based climate networks and their Forman curvature using global SAT data for EP, CP, and Normal conditions. Fig. 4.4 a–c show the distribution of spatial link lengths as the shortest distance between two connected points on the globe, the so-called great-circle lengths. Shown are all links of the network (black markers), the most positively curved F_{ij}^{+} (red) and most negatively curved F_{ij}^{-} (blue) network links for Normal, EP, and CP conditions.

The spatial distribution of F_{ij}^{+} (Fig. 4.4 d–f) shows that positive curvature occurs at only regional scales $\leq 5 \cdot 10^3$ km, resulting in the enforcement of local community structures.

While F_{ij}^{-} links occur at all spatial scales, for lengths $\geq 10^4$ km, the F_{ij}^{-} and 'All' link distributions almost overlaps for CP and Normal conditions.

Long-range teleconnections are here thus modulated predominantly via negatively curved links. As the curvature estimation of climate network

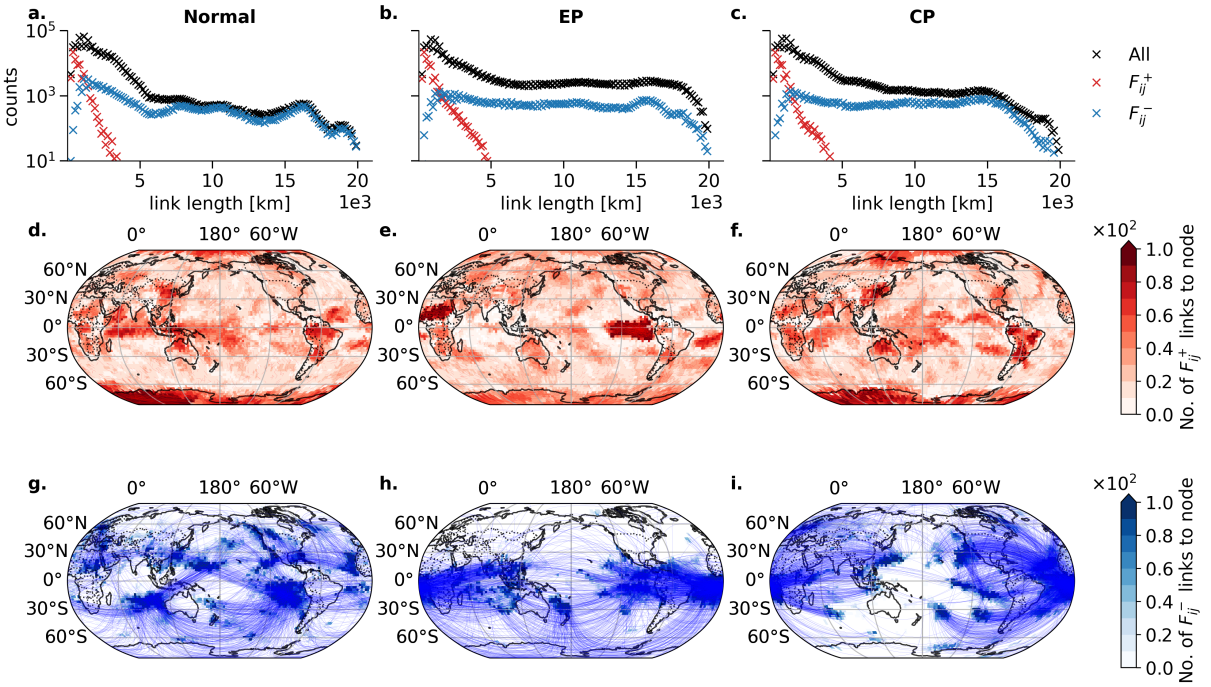


Figure 4.4.: Forman curvature of Normal, EP, and CP El Niño climate network links. The networks are computed from SAT anomalies for Normal (left column), EP (middle column), and CP (right column) conditions. The first row (**a**, **b**, **c**) depicts the spatial link length distribution for all (black), most positively F_{ij}^+ (red) and most negatively curved F_{ij}^- (blue) edges. The second row shows F_{ij}^+ (**d**, **e**, **f**), the third row F_{ij}^- (**g**, **h**, **i**). Colorbars indicate the number of incoming edges to a node. For visual reasons only every 20th edge is plotted in **d**-**i**.

links does not include any information on the spatial length, this suggests that the relationship between negative climate network link curvature and long spatial scales is an intrinsic topological property of the SAT dynamics. Therefore, negative curvature results in link bundles that are related to well-known teleconnection patterns which are discussed in Sec. 4.3.3.

In contrast to CP and Normal conditions, for EP El Niño conditions the distribution of F_{ij}^- does not describe all spatially long-range links (Fig. 4.4 b) and the most negative network links undergo a drastic spatial reorganization (compare Fig. 4.4 h to Fig. 4.4 g, i).

In particular, we observe that the connection between the tropical Pacific and the southern Atlantic is strengthened during EP conditions (Fig. 4.4 h), while during CP El Niños the tropical Pacific shows an enhanced connection to the extratropical Pacific as well as to the mid-latitude North Atlantic region (Fig. 4.4 i, Sec. 4.3.3).

Changes in most positively curved links F_{ij}^+ are more subtle between Normal (Fig. 4.4 d, f), CP (Fig. 4.4 f) and EP conditions (Fig. 4.4 e), such as the weakening of regional correlation structures in the tropical Atlantic (CP conditions) and the strengthening of correlation in the Eastern Pacific and the West African monsoon belt (EP conditions).

Note, we mainly discover teleconnections within the global oceans because correlations are generally higher over oceans than over land due to slower oceanic SAT variability (Lambert et al., 2011).

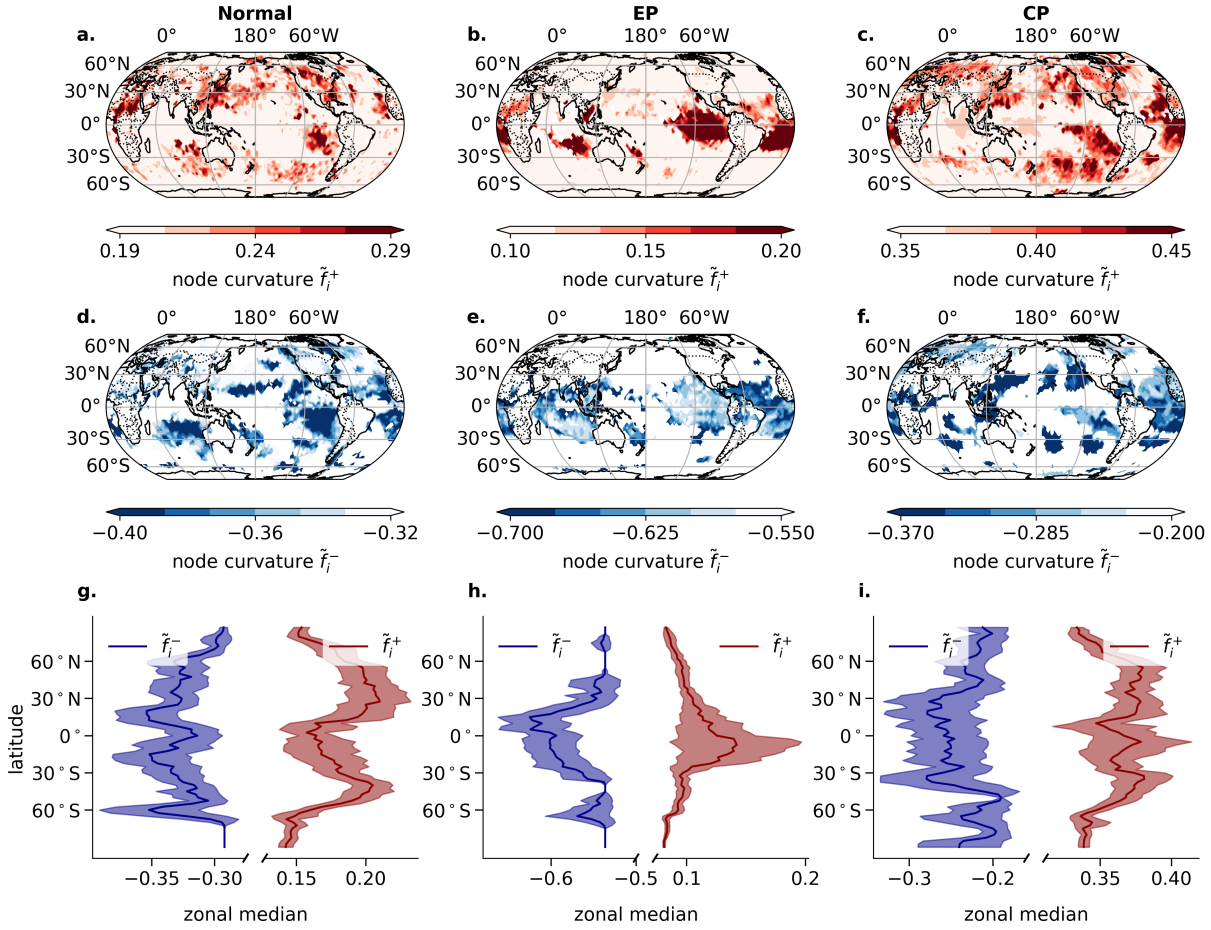


Figure 4.5.: Node curvature of Normal, EP El Niño, and CP El Niño climate networks. Node curvature hotspots for Normal (left column), EP (middle column) CP (right column) hotspots are shown summing the most positively curved $\tilde{f}_i^+$ (**a-c**), most negatively curved $\tilde{f}_i^-$ (**d-f**) links, as well as the aggregated zonal medians of $\tilde{f}_i^+$ (red) and $\tilde{f}_i^-$ (blue) (**g-i**) are shown.

4.3.2. EP El Niño Conditions Lead to Enforcement of Teleconnections to the Tropics

The global teleconnection pattern is under EP conditions strongly alternated compared to Normal conditions (Fig. 4.4, Fig. 4.5). We suggest that the stronger alternation can be attributed to the intensity of the events. EP events tend to have higher temperature anomalies and more stable warming locations in the tropics than CP events leading to a stronger alternation of the global atmospheric circulation which results in the confinement of hotspots around the tropics for both regional links and teleconnections (Fig. A.3a).

In contrast, CP conditions show unstable warming locations (Fig. A.3b) and a weaker amplitude and thus cannot alter the general circulation. Therefore, Normal and CP networks reveal various similarities resulting in positive and negative curvature hotspots over all latitudes (Fig. 4.5). This is further confirmed by the zonal medians (Fig. 4.5 g–i).

We repeat our analysis for strong and moderate El Niño events[4] separately and find that the EP curvature results (Fig. 4.5) are mainly dominated by the strong EP El Niños (not shown). We thus conclude that the differences between EP conditions to CP and Normal conditions are driven by strong

[4]We consider El Niño events to be "moderate" $0.5\,K \leq N3(N4) \leq 1\,K$, and "strong" when N3 (N4) ≤ 1 K.

EP El Niño events. This corroborates previous works by (Wiedermann et al., 2016; Lu et al., 2020) who report a strong localization of climate network links during EP conditions as well.

For EP conditions, positive curvature hotspots $\tilde{f}_i^+$ reveal the well-known ENSO tongue (Fig. 4.5 b) typically observed in empirical orthogonal function analysis of SST data (Johnson, 2013). We also find pronounced regions of $\tilde{f}_i^+$ in the Indian Ocean (IO) and the tropical Atlantic which are known to be affected by strong EP El Niño events (Klein et al., 1999; Wang et al., 2014; Zhang et al., 2015; Rodrigues et al., 2015).

Under CP conditions, $\tilde{f}_i^+$ is spread over all latitudes and over different regions of the globe (Fig. 4.5 c). For instance, we observe a hotspot in the tropical Pacific similar to the El Niño tongue, which is however shifted towards the dateline and also extended southwards.

Negative node-curvature hotspots, $\tilde{f}_i^-$, for EP conditions (Fig. 4.5 e) show enhanced teleconnections in the South China Sea, tropical IO, eastern tropical Pacific, and the tropical Atlantic. This coincides with a decrease in teleconnections in the extratropical Pacific, southern IO, North Atlantic (near Greenland), and the Southern Ocean.

Under CP conditions (Fig. 4.5 f), we find negative node curvature hotspots in the extratropics and mid-latitudes similar to $\tilde{f}_i^+$. While some similarities between CP and Normal conditions are found (Fig. 4.5 d, f), for example at the US-West Coast (Capotondi et al., 2019), in the northern tropical Pacific, and in the tropical Atlantic, the major difference can be found over the European continent, the Northern Atlantic, and the Labrador Sea (see as well Fig. 4.4 g,i).

We repeat our analysis using 1000 years of pre-industrial run of the UKESM1-0-LL model from the Coupled Model Intercomparison Project (CMIP)6 project (Eyring et al., 2016) which shows a comparable variation of SSTA in the tropical Pacific to observational data (Dieppois et al., 2021). We obtain a localization of positive and negative curvature hotspots to the tropics under EP conditions and teleconnections to the mid-latitudes under CP conditions (Fig. A.4), similar to our analysis of observational data (Fig. 4.5). The long time period of the pre-industrial run allows us to study the effect of the small number of EP and CP events in the reanalysis dataset, discussed in Sec. A.4. The curvature analysis obtained from only 7 EP/CP events tends to overestimate the number of negatively curved links. However, qualitatively the spatial patterns of positive and negative curved links are robust in comparison to the 297 EP events obtained in the whole pre-industrial time period (Fig. A.5).

4.3.3. EP and CP El Niño Teleconnection Patterns of Eastern and Central Pacific Ocean, Indian Ocean, and the Labrador Sea

During EP conditions, teleconnections link the N3 region (Fig. 4.6a) to the tropical Atlantic, supporting earlier work showing that strong El Niños can lead to warming in the tropical Atlantic mediated by the tropospheric temperature mechanism (Chang et al., 2006) and the atmospheric bridge via the Pacific North American (PNA) pattern (Alexander et al., 2002; Rodrigues et al., 2011).

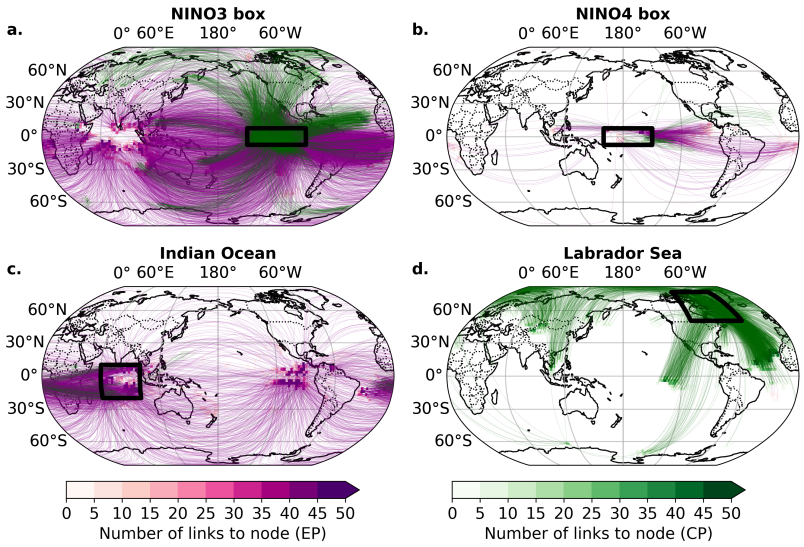


Figure 4.6.: Teleconnections from eastern and Central Pacific Ocean, IO, and the Labrador Sea. $F_{\overline{ij}}$, i.e., teleconnections, for EP (purple) and CP (green) El Niño events connected to the four selected regions (black rectangles): Niño 3 (**a**), Niño 4 (**b**), IO (**c**), and Labrador Sea (**d**). For visual clarity, only a third of all links are shown.

The links connecting N3 to the IO are likely apparent because of the influence of eastern tropical Pacific SSTs on the IO during and after El Niño events attributed to net heat flux anomalies due to changes in the atmospheric circulation of ENSO (Klein et al., 1999).

During CP conditions, we observe links from the N3 box as well, primarily connecting to the extratropical Pacific. These are likely due to the North Pacific Meridional Mode (NPMM) and South Pacific Meridional Mode (SPMM), as atmospheric and oceanic anomalies in the extratropics associated with SPMM and NPMM affect the intensity and flavor of El Niño (You et al., 2018).

Capotondi et al. (2021) showed that the dominant patterns of wind variability in the Northern and Southern Hemispheres, which represent the atmospheric expressions of the NPMM and SPMM, have a stronger relationship with CP events than EP events (Capotondi et al., 2021; Capotondi et al., 2019; Amaya, 2019), consistent with the extratropical links observed during CP events (Fig. 4.6 a).

The teleconnections of the N4 region (Fig. 4.6b) are far fewer compared to the N3 region (Fig. 4.6a). It is rather surprising that the N4 region is not well connected even during CP conditions, although a CP El Niño is primarily characterized by higher SST anomalies in the N4 region (Fig. A.3b).

These findings are in apparent disagreement with Barsugli et al. (2002) and Shi et al. (2019) who showed that the PNA and the northern Pacific are most sensitive to changes of SST in the central tropical Pacific using Global Circulation Model (GCM). GCMs are known to produce SST anomalies too far in the west (Capotondi et al., 2006; Kug et al., 2010; Ham et al., 2012), however, this can only partially explain the missing teleconnections in the Central Pacific. We find that although anomalies in the Central Pacific are stronger during CP conditions, the Eastern Pacific shows significantly higher local correlations (Fig. A.6).

Strong anomalies do not necessarily imply high correlations, therefore, we attribute the weak correlations in the Central Pacific under CP conditions to the higher spatial variability of anomalously warm temperatures

between CP El Niño events than between EP El Niño events (Wang et al., 2013; Chen et al., 2021). During CP conditions, anomalies in the tropical Eastern Pacific are weaker but behave more consistently in time than in the Central Pacific. We thus suggest that the teleconnections from the N3 region are artifacts of the correlation but proxies for the stimulation by the Central Pacific.

This is further supported by the curvature analysis on the UKESM1-0-LL (Sec. A.4). Teleconnections to the northern Pacific and west American coast are found in the CP network while the Central Pacific does not exhibit strong teleconnections. Further analysis using model runs with nudged SSTA could therefore allow disentangling the role of the central and Eastern Pacific on teleconnections to the northern Pacific and North America.

The IO has a large number of most negatively curved links during EP conditions but not for CP (Fig. 4.5 c, d). The EP event teleconnections link the IO to the tropical Atlantic and Pacific basins (Fig. 4.6 c). The links to the tropical Atlantic could be either attributed to indirect links mediated by the impact of the Ninõ 3 region on the tropical Atlantic or might resemble direct impact between the oceans as recently described by Zhang et al. (2021). CP conditions do not result in teleconnections in the IO as CP events are generally weaker (Zhang et al., 2015).

The Labrador Sea (Fig. 4.6d) is another pronounced area of most negative node-curvature in the CP network not present in the EP network (compare Fig. 4.5 c,d. Edges with the most negative curvature adjacent to the Labrador Sea connect to the extra topical Atlantic and, in contrast to Normal year conditions (Fig. S13), to the eastern tropical Pacific. This pattern may be attributed to the North Atlantic Oscillation (NAO), which refers to sea level pressure changes in the Arctic and subtropical Atlantic (Jiménez-Esteve et al., 2018). El Niño is known to cause a negative NAO pattern driven by the PNA pattern where a negative NAO phase is attributable to CP El Niño events via the subtropical bridge (Graf et al., 2012; Domeisen et al., 2019).

Note that we can identify the teleconnection patterns primarily because we have the results from the curvature-based climate network analysis to identify differences in the teleconnection structure and to guide our interpretations. Without the curvature analysis, for instance using classical complex network measures like betweenness centrality (Sec. A.2), it is not trivial to figure out the most important regions for each El Niño flavor.

4.4. Discussion

We presented a new approach to estimate global teleconnection patterns of surface air temperature and used it to investigate the teleconnections of Eastern Pacific and Central Pacific El Niño events. Our approach involves the combination of correlation-based climate networks with a recently established network measure based on Ricci-curvature. In particular, we used Forman-Ricci curvature to distinguish links related to small-scale regional structures (positively curved links) from long-range

teleconnections connecting regions from different parts of the globe (negatively curved links).

We showed that El Niño diversity drastically impacts the spatial organization of teleconnections. We identified teleconnection hotspots for both EP and CP conditions and showed that EP conditions strongly alternate the teleconnection structure of the climate network to be mainly confined to the tropics, whereas CP network teleconnections were found in northern and southern mid-latitudes as well. A comparison of the results from the application of our approach on model output from the UKESM-LL2 Model from the CMIP6 project further corroborates our findings.

We further investigated the impact of ENSO diversity on the teleconnection patterns of four specific regions: the Niño 3 region, the Niño 4 region, the northern IO, and the Labrador Sea. We found that the Niño 3 region in the Eastern Pacific has a large number of teleconnections irrespective of whether we consider EP or CP conditions, whereas the Niño 4 region in the Central Pacific has, in comparison, very few teleconnections under both EP and CP conditions and our analysis suggests the higher spatio-temporal variability of anomalous SSTs in the Central Pacific to be related. We thus conclude that the Eastern Pacific is the primary mediator of El Niño impacts irrespective of the El Niño type, and acknowledge that further work on the role of the Eastern Pacific during CP El Niños is needed.

We found that the northern IO and the Labrador Sea show teleconnections almost only under EP and CP conditions respectively. While the teleconnections of the IO region to the Niño 3 region during EP conditions are well-known, the links between the Labrador Sea to the eastern tropical Pacific and northern tropical Atlantic are not fully understood.

Open Research

Datasets for this research are available from the Copernicus Climate Change Service. The data from 1959 till date was taken from Hersbach et al. (2020). The code for generating and analyzing the networks is made publicly available under (Schlör et al., 2022a). The code for reproducing the analysis of the network curvature described in this paper is publicly available under (Schlör et al., 2022b).

Acknowledgements

The authors would like to thank the reviewers for their helpful comments and suggestions. The authors acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Jakob Schlör and Felix Strnad.

Contributions of El Niño Southern Oscillation Diversity to Decadal Variability

5.

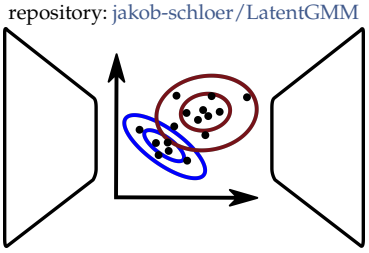
Abstract

El Niño Southern Oscillation (ENSO) diversity is characterized based on the longitudinal location of maximum sea surface temperature anomaly (SSTA) and amplitude in the tropical Pacific, as Central Pacific (CP) events are typically weaker than Eastern Pacific (EP) events. SSTA pattern and intensity undergo low-frequency modulations, affecting ENSO prediction skill and remote impacts. Yet, how different ENSO types contribute to these decadal variations and long-term variance trends remain uncertain. Here, we decompose the low-frequency changes of ENSO variance into contributions from ENSO diversity categories. We propose a fuzzy clustering of monthly SSTA to allow for non-binary event category memberships. Our approach identifies two La Niña and three El Niño categories and shows that the shift of ENSO variance in the mid-1970s is associated with an increasing likelihood of strong La Niña and extreme El Niño events.

5.1. Introduction

The El Niño-Southern Oscillation (ENSO), characterized by anomalous sea surface temperatures (SSTs) in the tropical Pacific, exhibits notable diversity in its amplitude, temporal evolution, and spatial pattern. The El Niño events of 1982-83 and 1997-98, for instance, recorded exceptionally high SST anomaly (SSTA) values in the eastern equatorial Pacific, whereas the El Niño of 2002-03 was less extreme and exhibited the largest anomalies in the central equatorial Pacific (McPhaden, 2004). In order to describe these event-to-event differences, El Niño events have been generally categorized as Eastern Pacific (EP), and Central Pacific (CP) types (Capotondi et al., 2015). EP El Niño events typically have their peak SSTA in the eastern Pacific, may exhibit stronger intensities, and a largely reduced zonal thermocline slope, resulting in the pronounced discharge of warm water from the equatorial thermocline. In contrast, CP events show peak SSTA in the central Pacific and are comparatively weaker with smaller changes in zonal thermocline slope and warm water discharge (Kug et al., 2009; Capotondi, 2013).

Code is available at the Github repository: jakob-schloer/LatentGMM



These different types of ENSO events have substantially different downstream impacts on the global climate (Strnad et al., 2022; Beniche, 2023). For example, extreme drought conditions were recorded in eastern Australia in 2002, while a minor impact on precipitation was detected during the extreme 1997 event (Wang et al., 2007). Weaker and shorter-lived CP events are associated with warm conditions in the equatorial Atlantic during boreal winter, while stronger and persistent EP events lead to cold anomalies in that area, with different impacts on precipitation over northeastern Brazil (Kao et al., 2009; Rodrigues et al., 2011). Thus, a

deeper understanding of ENSO diversity is critical to support predictions of ENSO impacts.

ENSO characteristics, including amplitude and spatial pattern, exhibit decadal variations, which are mediated by changes in the background state of the tropical Pacific (Capotondi et al., 2023). Notable decadal phase transitions were observed in the late 1970s (Miller et al., 1994) and around the year 2000 (McPhaden, 2012). Paleoclimate data also indicate an increase in ENSO amplitude over recent decades (Grothe et al., 2020), consistent with modeling results showing a significant increase in ENSO amplitude after 1960, which was attributed to anthropogenic forcing (Cai et al., 2023). Analysis of a large number of observationally-based datasets revealed that the longitudinal location of the maximum SST anomalies, as well as the intensity of both El Niño and La Nina events, undergo decadal fluctuations (Dieppois et al., 2021), which can, in turn, modulate ENSO predictions (Lou et al., 2023). However, event location and intensity were considered separately by Dieppois et al. (2021), so that the contribution of different ENSO types to these decadal changes and long-term trends in ENSO variance remains unclear. Paleoclimate data (Lawman et al., 2022) indicate that extreme ENSO events may contribute to increases in ENSO variance. Similarly, climate models that capture relevant aspects of ENSO nonlinearities project an increase in ENSO variance that is linked to an increase in the frequency of extreme ENSO events (Cai et al., 2021). However, a more comprehensive understanding of these changes from an ENSO diversity perspective is still missing.

One main reason hindering the estimation of the contribution of different ENSO categories to its decadal modulation is that ENSO classifications often depend on the chosen definitions (Pascolini-Campbell et al., 2015; Yu et al., 2013; Capotondi et al., 2020a; Abdelkader Di Carlo et al., 2023). The disagreement between different ENSO classification methods is likely due to the assumption that ENSO events can be classified into binary types, based on indices capturing the location of the highest SSTA in the Tropical Pacific (like the Niño3 and Niño4 regions), or using Empirical Orthogonal Functions (EOFs) (Ashok et al., 2007; Kug et al., 2009; Kao et al., 2009; Takahashi et al., 2011). However, ENSO events are continuously distributed in the space spanned by the two leading principal components (PCs) (Takahashi et al., 2016; Cai et al., 2018; Capotondi et al., 2020a). Approaches that use a more continuous distribution of SSTAs to identify diversity show that events occur over multiple locations but with enhanced probabilities over the central and eastern tropical Pacific (Dieppois et al., 2021; Shin et al., 2021). In addition, both EP- and CP-type events appear to share common underlying dynamical processes, albeit with varying relevance depending on the longitude (Capotondi, 2013).

A notable example of an El Niño that eludes a binary classification is the 2015/16 events. While its SSTA and several of its impacts were typical of extreme EP events (Santoso et al., 2017), this event was not associated with a significant impact on California precipitation and was not followed by a strong La Niña, as other events of this type (i.e., 1982-83 and 1997-98). Indeed, this event was considered a mixture of EP and CP types (Paek et al., 2017; Capotondi et al., 2020a).

Here, we propose a new approach for characterizing ENSO diversity to better understand its relationship with decadal changes in ENSO variance.

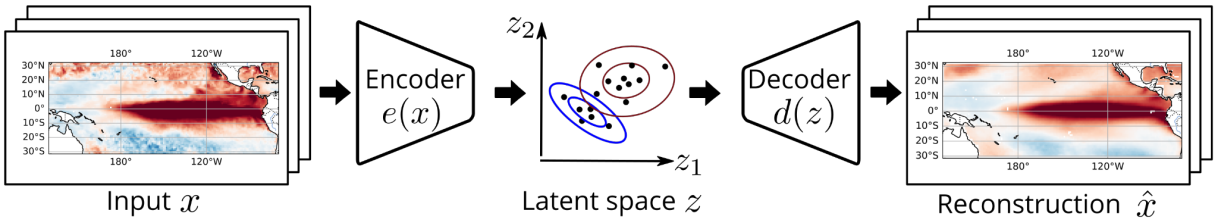


Figure 5.2.: Sketch of the experimental setup. We transform the high-dimensional input fields to a low-dimensional latent space which is denoted by the encoder and decoder function. The distribution of the data in the latent space is then approximated by a GMM.

To that end, we develop a fuzzy clustering of the low-dimensional representation of monthly SSTA in the PC1-PC2 space to achieve a non-binary event categorization (i.e., events belonging to one cluster or not). Instead, in this fuzzy clustering approach, individual events are assigned a probability of belonging to a given cluster. Such membership probabilities are then used to determine their relative contributions to the low-frequency ENSO variance.

5.2. Data and Methods

5.2.1. Data

Our analysis is conducted on monthly SSTA using eight reanalysis datasets (Tab. 3.2). Anomalies are computed by linearly detrending each dataset after 1950 and removing the monthly climatology. We select the Pacific region (130°E-70°W, 30°S-30°N) and interpolate each dataset to a $1° \times 1°$, resolution using the 'nearest-neighbor' technique. In order to ensure the same number of data points per time step, we randomly select four data points without replacement at each month between 1901–2022.

While the EOF analysis is calculated over all months in the merged dataset, we use only El Niño and La Niña winter months for fitting the GMM in the lower dimensional space. El Niño and La Niña boreal winters (December-January-February (DJF)) are selected when the average SSTA in the Niño3.4 region is larger than 0.5 K or smaller than -0.5 K.

If not stated differently, we use ORAS5 for plotting Hovmöller diagrams and composites of ocean variables (SSTA and SSHA). The atmospheric variables used in this study, namely the 6-hourly eastward component of the wind at 10 meters above the surface are taken from ERA5. Their anomalies are computed as described above. We select the zonal wind component over the tropical Pacific (130°E - 70°W, 5°S-5°N) and decompose it into HF (periods longer than 250 days) and LF (periods between 250 - 5 days) using Fourier filtering, as described by Capotondi et al. (2018). For the multivariate GMM analysis, we use SST and SSH which are only available in a subset of the aforementioned datasets, namely, in ORAS5, GODAS, and CERA-20C.

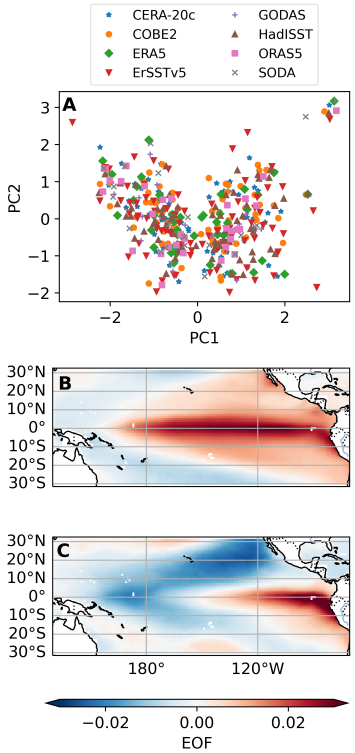


Figure 5.1.: Averages of monthly boreal winter (DJF) SSTAs of eight reanalysis products (A) are projected on EOF1 (B) and EOF2 (C). El Niño (La Niña) boreal winters are selected when the DJF average Niño3.4 is larger than 0.5 K (smaller than -0.5 K).

5.2.2. Fuzzy clustering

Clustering algorithms are typically based on a distance measure between data points which becomes ill-posed in high-dimensional spaces (Parsons et al., 2004). To mitigate this issue, we first reduce the dimensionality of the geospatial fields before applying fuzzy clustering using EOF analysis.

Our input data, $X = \{\mathbf{x}(t)\}_{t=1}^{N_t}$, which consists of tropical Pacific SSTA fields $\mathbf{x}(t) \in \mathbb{R}^{N_{\text{lon}} \cdot N_{\text{lat}}}$, are projected onto the first two EOF patterns to obtain the two PCs $\mathbf{z}(t) \in \mathbb{R}^2$, referred to as a latent vector. Besides EOFs, we also use a nonlinear dimensionality reduction method, specifically a convolutional autoencoder neural network (Fig. 5.2, Sec. B.1.1).

We apply the Gaussian Mixture Model (GMM), a probabilistic unsupervised clustering approach to identify different types of ENSO events. GMMs describe each cluster via a multivariate Gaussian distribution, accommodating overlapping probabilities. Mathematically, GMMs assume that the probability distribution of latent states, $p(\mathbf{z})$, comprises a mixture of Gaussians:

$$p(\mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mathbf{z} \mid \mu_k, \Sigma_k\right), \tag{5.1}$$

with K representing the number of Gaussians, $\mathcal{N}\left(\mathbf{z} \mid \mu_k, \Sigma_k\right)$, each characterized by a mean μ_k and covariance Σ_k. The probability of each Gaussian is denoted as $p\left(c_k\right) = \pi_k$ with $\sum_{k=1}^{K} \pi_k = 1$. The parameters of the Gaussians μ_k, Σ_k are estimated iteratively using the Expectation Maximization (EM) algorithm (Yu et al., 2010). A detailed description of the EM algorithm for estimating GMM parameters can be found for instance in Murphy (2013).

Category membership: The GMM approximation of the low - dimensional distribution enables us to assign conditional probabilities that an event $\mathbf{z}(t)$ belonging to category c_k, using Bayes' theorem:

$$p\left(c_k|\mathbf{z}(t)\right) = \frac{p\left(c_k\right) p\left(\mathbf{z}(t)|c_k\right)}{\sum_l p\left(c_l\right) p\left(\mathbf{z}(t)|p(c_l)\right)} = \frac{\pi_k N\left(\mathbf{z}(t); \mu_k, \Sigma_k\right)}{\sum_l \pi_l N\left(\mathbf{z}(t); \mu_l, \Sigma_l\right)}. \tag{5.2}$$

These category membership probabilities, denoted as $p\left(c_k|\mathbf{z}(t)\right)$, form a time series for each category, indicating the likelihood of occurrence. Notably, the conditional probabilities for a given event $\mathbf{z}(t)$ across the categories sum up to one, i.e. $p\left(c_k|\mathbf{z}(t)\right) = 1$. This indicates that $p\left(c_k|\mathbf{z}(t)\right) = 1$ signifies the event belonging exclusively to category c_k. Conversely, events with probabilities shared among multiple categories cannot be assigned to a single category but exhibit some probability of belonging to two or more categories.

In order to estimate the optimal number of clusters k, we apply the Bayesian Information Criterion (BIC) (Schwarz, 1978). The likelihood of the GMM can generally be increased using more clusters which on the other hand increases the complexity of the model and is therefore prone to overfitting. BIC is a model selection criterion that when minimized balances the model complexity to the increase in the likelihood. BIC

is defined as $BIC(k) = Pln(N) - 2ln\left(p\left(Z|\theta\right)\right)$ with P is the number of parameters of the GMM, $\theta = \left(\pi_k, \mu_k, \Sigma_k\right)$, N is the number of data points and $p\left(X|\theta\right)$ is the likelihood. In our analysis, we fit 100 GMMs with random initialization for each value of k (number of Gaussians) and compute the BIC. The model with the lowest BIC represents the best trade-off between model complexity and fit quality. Additionally, the spread of the BIC values across different initializations serves as a measure of the robustness of the fit, indicating whether the Gaussian parameters consistently converge to the same values regardless of initialization.

5.2.3. Variability estimation

The GMM allows assigning a probability, $p\left(c_k|\mathbf{z}(t)\right)$, to each data point, $\mathbf{z}(t)$, that quantifies its likelihood of belonging to category c_k. These are the category memberships of the data point (Eq. 5.2), which are inherently fuzzy and non-binary (i.e. they are probabilities between 0 and 1), and allow us to model ENSO events in terms of their likelihood of occurrence. As a consequence, we can decompose a variable $\mathbf{y}(t)$, for instance, SSHA, into the contributions of each category, by

$$\mathbf{y}(t) = \sum_k p\left(c_k|\mathbf{z}(t)\right) \cdot \mathbf{y}(t) := \sum_k \mathbf{y}_k(t) \tag{5.4}$$

where $\mathbf{y}_k(t)$ is defined as the contribution of category c_k. Averaging $\mathbf{y}_k(t)$ over time corresponds to a weighted average, with the weights being the categorical memberships (Eq. 5.3). Similarly, we can write the variance $\langle \mathbf{y}^2 \rangle$, as

$$\langle \mathbf{y}^2 \rangle = \langle \left(\sum_k \mathbf{y}_k \right)^2 \rangle = \sum_k \langle \mathbf{y}_k^2 \rangle + \sum_{l \neq m} \langle \mathbf{y}_l, \mathbf{y}_m \rangle, \tag{5.5}$$

where $\langle \mathbf{y}_k^2 \rangle$ is the variance contribution of category c_k and $\langle \mathbf{y}_l, \mathbf{y}_m \rangle$ is the co-variability of categories l and m.

Probabilistic weighted averages: In our approach, we determining composites for each category c_k of a desired variable, $\mathbf{y}$, through their weighted average:

$$\underline{\mathbf{y}}_k = \sum_{t=1}^{N} p\left(c_k|\mathbf{z}(t)\right) \mathbf{y}(t), \tag{5.3}$$

where the weights are the conditional probabilities $p\left(c_k|\mathbf{z}(t)\right)$ given for each event $t = 1, \ldots, N$.

5.3. Results

5.3.1. ENSO Diversity is Well Explained by Five Categories

The two most dominant EOFs of all eight SSTA datasets combined (Tab. 3.2) present the well-known spatial patterns associated with ENSO, i.e. EOF1 depicts the typical ENSO pattern with anomalies in the central-eastern Pacific, while EOF2 exhibits an east-west dipole structure (Fig. 5.1B, C). Projection of the monthly SSTA of all boreal winter (December-January-February) El Niño and La Niña events onto EOF1 and EOF2, produces a distribution in the corresponding PC1–PC2 space that exhibits a wide, boomerang-like shape (Fig. 5.3A, Fig. 5.1A). This nonlinear relationship between PC1 and PC2 has been considered an expression of key ENSO dynamics (Cai et al., 2018; Takahashi et al., 2016; Ham et al., 2012; Karamperidou et al., 2017), and used in the selection of models to consider for examining future projections (Cai et al., 2018;

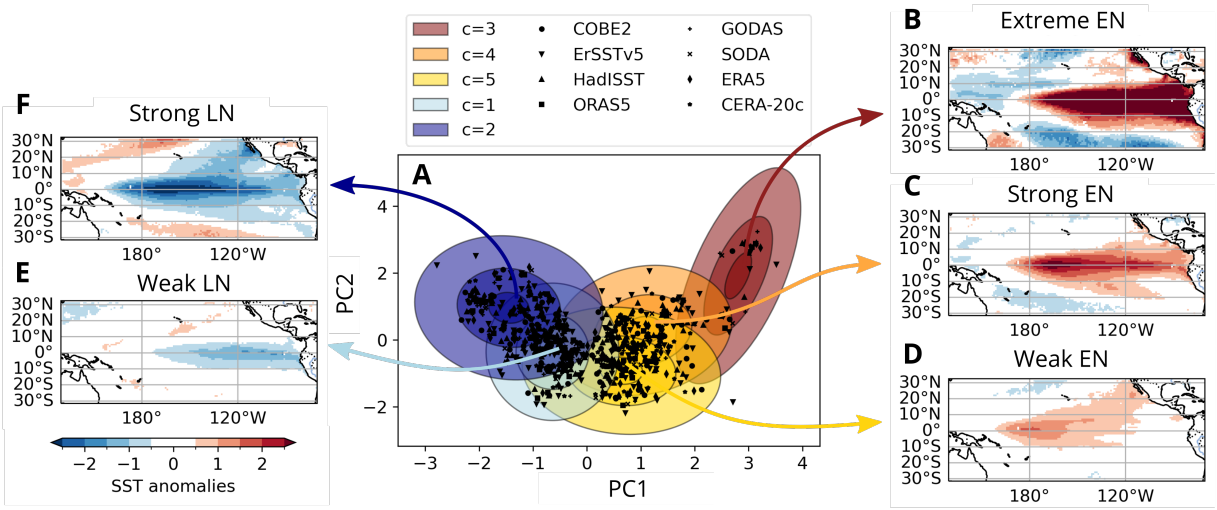


Figure 5.3.: El Niño and La Niña categories in PC1-PC2 space. Monthly boreal winter (DJF) SSTAs of El Niño and La Niña events of all reanalysis products (Tab. 3.2) projected onto the PC1-PC2 space and fitted by a Gaussian Mixture Model (GMM). Each event (DJF averages are shown as black dots in A) has a probability of belonging to each of the five categories (colored Gaussians in A). Pacific SSTA composites for each category are obtained by using the category membership as weights for the averages, depicted in panels (B-F). We obtain three El Niño-like patterns: *Extreme EN* (B), *Strong EN* (C), and *Weak EN* (D), while La Niña events form two categories: *Weak LN* (E) and *Strong LN* (F).

Cai et al., 2021). The nonlinear relationship itself can be accounted for by using a nonlinear transformation in place of the linear EOF-based transform, e.g., an autoencoder, and in that case, the boomerang-like shape is replaced by a simple linear relation between the two latent dimensions (Fig. B.2).

The distribution of boreal winter El Niño and La Niña months in the PC1-PC2 space do not exhibit clear gaps visually, however, their density varies, suggesting a categorical structure (Takahashi et al., 2016). We model this distribution using a GMM with $k = 5$ categories (Eq. 5.1), a number determined using the BIC to ensure parsimony (Fig. B.1A). The five Gaussians are arranged in a row along the boomerang-shaped distribution from the coldest events at the leftmost tip to the warmest events at the rightmost end (Fig. 5.3A).

We use the category memberships, $p\,(c_k|\mathbf{z}(t))$, as weights for averaging the Pacific SSTAs of each category (Eq. 5.3) and find three El Niño-like patterns (Fig. 5.3B-D), and two La Niña-like patterns (Fig. 5.3E-F). Besides the different zonal locations of maximum warming/cooling, their defining factor is the Tropical Pacific SSTA intensity. Hence, we will refer to the three El Niño categories as *Extreme EN* (Fig. 5.3B), *Strong EN* (Fig. 5.3C) and *Weak EN* (Fig. 5.3D), and correspondingly to *Weak LN* (Fig. 5.3E) and *Strong LN* (Fig. 5.3F) for the La Niña categories. We find that the overall clustering of the SSTA patterns into the five categories is robust (SI Sec. B.2)upon changing the number of EOFs (SI Fig. B.1B), incorporating SSHA data along with the SSTA, or varying the latitudinal range of the input. The mean category membership for each boreal winter averaged across the datasets results in five time series of probabilities — one for each category — which reflect the likelihood of their occurrences (Fig. 5.4). We find that the indices are not sensitive to the dataset used for their estimation, as seen in the relatively small spread of the membership probabilities over the datasets at each time point. As with the number of clusters, the category membership is also robust to changes in the number

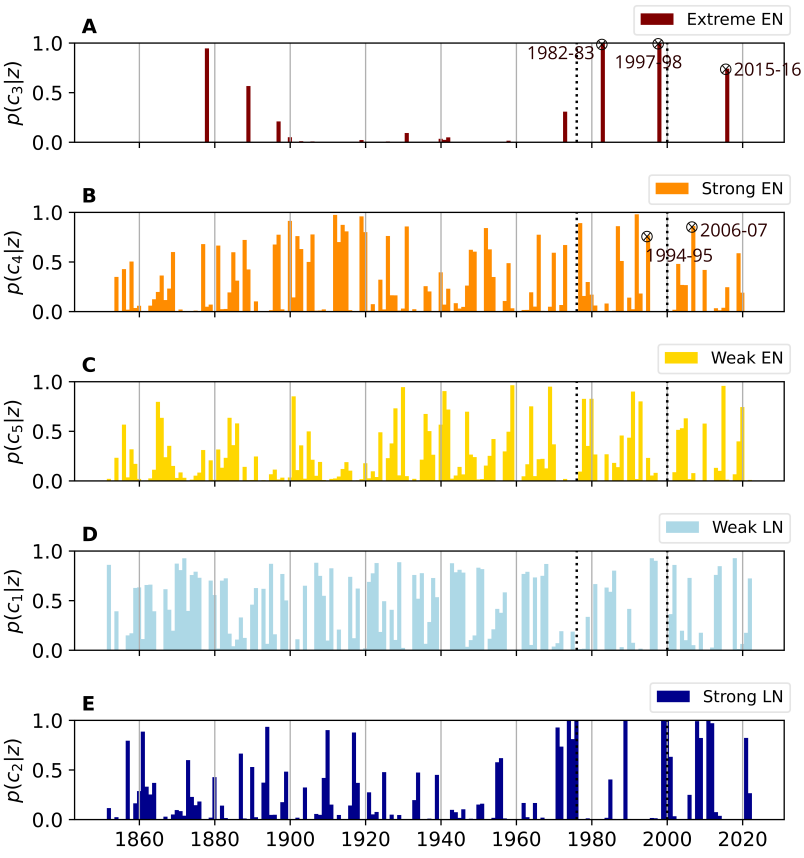


Figure 5.4.: Probabilistic category membership. The GMM in Fig. 5.3 allows us to estimate the likelihood, $p\,(c_k|\mathbf{z}(t))$, of each El Niño and La Niña winter month, $\mathbf{z}(t)$, to belong to each of the categories, c_k (Eq. 5.2). An event belongs only to one category when its probability is 1. However, many events have shared probabilities across several categories. The categories are sorted in the following order (top to bottom): *Extreme EN* (A), *Strong EN* (B), *Weak EN* (C), *Weak LN* (D), and *Strong LN* (E). For visual reasons, we average the monthly probabilities over each winter (DJF) and over reanalysis products. The dashed lines indicate the reported shifts in ENSO variability in 1976-77 and 2000.

of variables, spatial domain, and the use of nonlinear encoding.

Comparing the category membership series for *Strong* and *Weak EN* events (Fig. 5.4B, C) to two conventional classifications – the Niño3-Niño4 classification by Kug et al. (2009) and the EOF-based E/C classification by Takahashi et al. (2011) – we find that our classification mostly agrees with them (Tab. B.1, Sec. B.3). Differences, when they do occur, correspond to differences between the other classifications as well (Pascolini-Campbell et al., 2015; Yu et al., 2013; Capotondi et al., 2015). The two El Niños of 1994-95 and 2006-07, however, are classified by both conventional classifications as CP events, whereas we find them to be *Strong EN* events (Tab. B.1). This is likely because the warm water anomalies span from the central to the eastern Pacific during the duration of these events, leading to possible ambiguities in the event definition. We also note that unlike the 1982-83 and 1997-98 events, which unambiguously belonged to the *Extreme EN* category, the 2015-16 event also has a nonzero membership in the *Strong EN* category, confirming its mixed nature.

5.3.2. Extreme EN Events are Different from Strong EN Events

A key distinction of our ENSO categorization from conventional methods is the identification of the extreme El Niños as a separate class. While our *Weak EN* category corresponds to the conventional CP El Niño type, the conventional EP El Niño type is split into *Extreme EN* and *Strong EN* categories (Sec. B.3, Tab. B.1). SSTA composites of conventional EP El

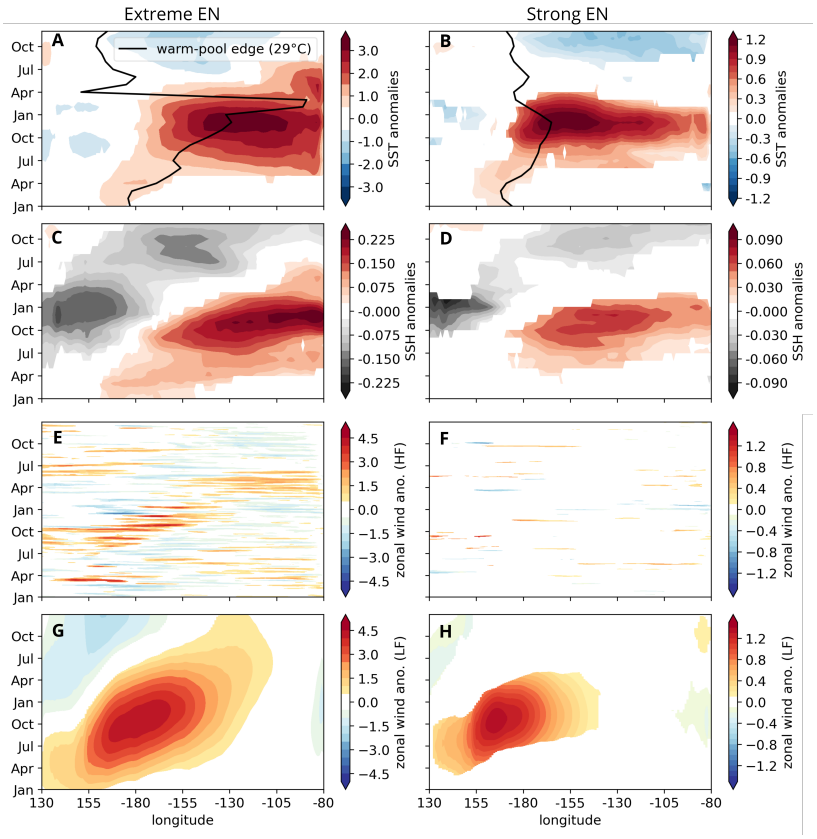


Figure 5.5.: *Extreme EN* and *Strong EN* category. Hovmöller diagrams of SSTA (A, B), SSHA (C, D), high-frequency (HF) zonal wind anomalies (E, F), and low-frequency (LF) zonal wind (G, H) anomalies are obtained by meridional averages (5°S - 5°N) of each month in the year preceding and succeeding El Niño events. Each two-year period is weighted by the corresponding DJF average category membership probability (Fig. 5.4). The black line in (A) and (B) indicates the warm-pool edge, i.e., the 29°C SST isotherm (Sec. B.1.2). Only values that are statistically significant above the 95th percentile are displayed (Sec. B.1.3). SSTA and SSHA are taken from ORAS5 (1958–2022), while 10-meter zonal winds, with their HF- and LF components (Sec. 5.2.1) are computed from ERA5.

Niño events exhibit the maximum warming in the eastern Pacific. The maximum warming is however strongly influenced by the few extreme EN events (1982/83, 1997/98, 2015/16), with an eastward shift of the peak towards the central Pacific when we exclude the extreme events.

The two categories also differ in their evolution, analyzed using Hovmöller diagrams for the year preceding and succeeding an event (Fig. 5.5). *Extreme EN* events show significant warm water volume anomalies, as described by SSHA, around the dateline already in the spring before an event, corresponding with a shift of the warm pool edge near the dateline (black line Fig. 5.5A, C). The *Extreme EN* onset phase also shows strong positive HF and LF zonal wind components to the west of the dateline, during the preceding spring, extending further east as the event develops to its mature phase (Fig. 5.5E, G). *Strong EN* events, on the other hand, do not demonstrate a consistent onset pattern in the preceding spring (Fig. 5.5D, F, H).

These findings corroborate ideas presented in prior research on the impact of the Walker circulation's zonal shift (Thual et al., 2023) and the influence of stochastic high-frequency winds, called Westerly Wind Bursts (WWBs), on ENSO Diversity (Fedorov et al., 2015; Capotondi et al., 2018; Puy et al., 2019).

5.3.3. Interdecadal ENSO Variability is Driven by Strong and Extreme events

The membership probabilities of each category (Fig. 5.4A, B) encode a distinct pattern of decadal-to-multidecadal ENSO variability. In par-

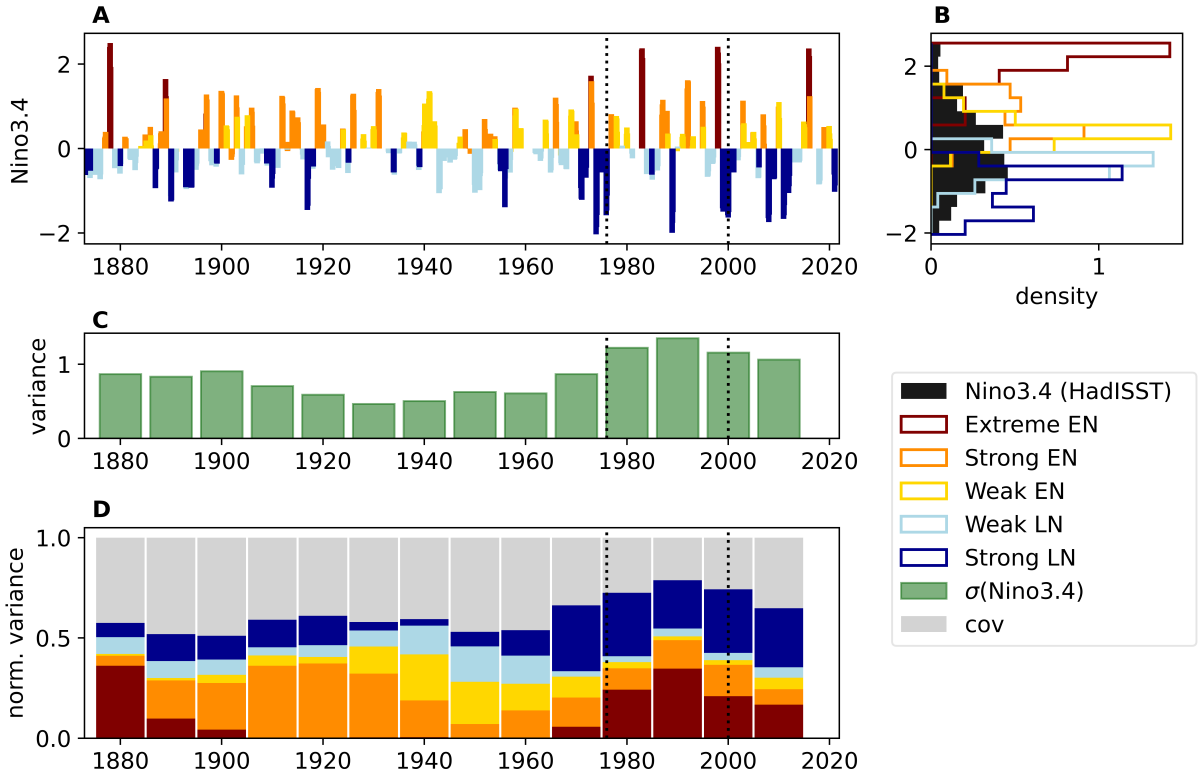


Figure 5.6.: Low-frequency changes of ENSO variance. For each category, the Niño3.4 index is multiplied by their category membership probabilities (A). The histogram of Niño3.4 intensities (B), highlights different SSTA amplitudes between categories. The 20-year running variance of Niño3.4 every 10 years (C), is used to normalize the 20-year running variance of each category (D). *Extreme EN* and *Strong EN* categories dominate the Niño3.4 variance in the early and late 20th century. The variance shift in 1976-77 and 2000 (dashed lines) highlight reported changes in ENSO variability. The Niño3.4 index is taken from HadISST (Rayner et al., 2003).

ticular, the *Strong LN* and the *Extreme EN* categories show a markedly prominent low-frequency variability, which is less evident in other categories. To quantify the low-frequency variation of ENSO, we multiply the Niño3.4 index (from HadISST) by the corresponding membership probabilities for each category (Eq. 5.4, Fig. 5.6A) and calculate their 20-year running variances every 10 years (Eq. 5.5). The variance of each category is normalized by the 20-year running variance of the Niño3.4 index (Fig. 5.6C) to determine their relative contributions.

The total variance of the Niño3.4 index shows a low-frequency modulation, with a minimum around 1920-1960, followed by an increasing trend, which is consistent with the shift in the mean and variance of ENSO after the so-called "1976-77 climate shift" (e.g. Miller et al., 1994), (dashed line in Fig. 5.6C). A slight decrease in variance is noticeable after the year 2000, in line with the other reported climate shift (e.g. McPhaden, 2012). While changes are noticeable in the variance of all categories, only minor contributions to the total Niño3.4 variance are observed for the *Weak LN* and *Weak EN* categories (Fig. 5.6D). The *Strong LN* category exhibits a substantial contribution starting around 1970, and the *Strong EN* category dominates the variance from 1890 to ~1930. Meanwhile, the contribution of the *Extreme EN* category is particularly visible towards the end of the 19th century and after 1980.

Between 20-40% of the decadal variance is due to the covariance between the different ENSO categories (denoted as *cov* in Fig. 5.6D). The covariance

captures mainly the contribution of co-occurrences of different categories to the overall variance in a decade. However, other factors such as events with intermediate values of category memberships and statistical errors in estimating the GMM categories also influence the covariance.

Our statistical approach does not allow us to assess the dynamics underpinning these decadal changes in the contribution of the different ENSO categories. ENSO decadal modulation is often associated with changes in the mean state, but it is not clear whether mean state changes are induced by influences from the extratropical Pacific or other ocean basins, or result as a residual of random variations of ENSO itself (Capotondi et al., 2023). Some studies, however, suggest that some ENSO events could induce decadal phase transitions through either nonlinear dynamical heating (Liu et al., 2022b) or by inducing a discharge of upper ocean heat content in the off-equatorial Western Pacific (Meehl et al., 2021).

Our results provide novel insights into low-frequency changes in ENSO variance, which combines changes in frequency and intensity. We find that although the magnitude of total variance contributed by Eastern Pacific warming events (i.e. *Extreme EN* and *Strong EN events* combined) since the turn of the 20th century is comparable to that seen during the period from the 1880s to 1940s, the variance in the recent decades has been dominated by the *Extreme EN* category, a result consistent with the statistically significant change in ENSO properties in the late 1970s (Capotondi et al., 2017). This shift was associated with a weakening of the easterly trade winds and a zonal reduction of the equatorial thermocline slope, conditions favoring stronger El Niño events. Linked to the increased contributions from *Extreme EN events* is the concurrent increase in the contributions of *Strong LNs* since the 1970s (dark blue in Fig. 5.6D), manifested in very strong, multi-year events starting around 1970 (Fig. 5.6A). These results are consistent with a higher likelihood of stronger CP La Niñas following the heat discharge of strong El Niños, as exemplified by the 1998-99 La Niña event after the extreme 1997-98 El Niño (Cai et al., 2015; Geng et al., 2023). While the increasing contribution of *Extreme EN* and *Strong LN* to ENSO variance starting around 1970 aligns with results from climate model simulations (Cai et al., 2021; Gan et al., 2023), our findings more specifically highlight the key role played by *Strong LN* in the ENSO variance changes after 1970.

We also find that during the 'quiescent' period of ENSO, roughly from the 1930s to the 1960s, most of the ENSO variability originated from Central Pacific warming and Eastern Pacific cooling events, i.e., from the weaker event types, while the influence of central Pacific warming on the ENSO variance in recent decades has been almost negligible. This latter result is in apparent disagreement with the reported increase in intensity and frequency of CP events since 1980 (Lee et al., 2010), and especially after 2000 (dashed line in Fig. 5.6D; (McPhaden et al., 2011)). This discrepancy may be related to the 20-year running variance we used to construct Fig. 5.6D, and our inclusion of the 2015/16 El Nino, an event that was missing in the records used by the earlier studies.

5.4. Discussion

We present a new approach for studying the influence of ENSO diversity on low-frequency changes in ENSO variance. In particular, we use a Gaussian Mixture Model (GMM) within the low-dimensional PC1-PC2 space of monthly SSTA, which enables the assignment of non-binary event category memberships. We identify two La Niña categories (*Weak LN*, *Strong LN*) and three El Niño categories (*Extreme EN, Strong EN, Weak EN*), which combine the two dimensions of ENSO diversity — longitudinal location of maximum SSTA and its intensity. A key contribution of our work involves utilizing the membership probabilities to determine how each of the five categories individually affects the overall decadal variability in the Niño3.4 region. We find that the increasing frequency of both *Extreme EN* and *Strong LN* are the primary drivers of the increased ENSO variance post-1970. While these findings are consistent with previous studies that also detected an increase in extreme ENSO events in the second half of the 20th century (Cai et al., 2018; Cai et al., 2023), our results further highlight the key role played by the increasing frequency of *Strong LN* in these ENSO variance changes.

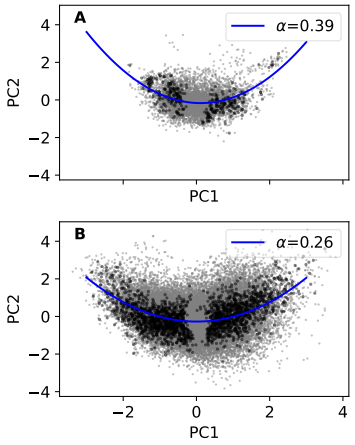


Figure 5.7.: Nonlinearity of ENSO diversity. Monthly tropical SSTA (10°S-10N) exhibits a parabolic shape when projected in PC1-PC2 space. The leading coefficient α of a quadratic fit, $PC2 = \alpha PC1^2 + bPC1 + c$, is a measure of ENSO nonlinearity (Dommenget et al., 2013; Karamperidou et al., 2017; Cai et al., 2018). Grey dots indicate monthly data and black dots show averages over DJF of winters identified as either El Niño or La Niña. The quadratic curve (blue) shows the fit to the monthly data from (**A**) all reanalysis products and (**B**) CESM2.

The proposed fuzzy clustering approach could also be used to quantify how well climate models represent ENSO diversity, akin to Dieppois et al. (2021) and Ayar et al. (2023). A preliminary analysis of the Community Earth System Model version 2 (CESM2) shows that this model only exhibits four ENSO categories instead of five (SI Sec. B.4, Fig. B.3), a discrepancy which cannot be explained by the different number of samples or other technical choices. Specifically, the *Extreme EN* category is missing in the model. While ENSO in CESM2 has an amplitude larger than observed, modeled SSTAs tend to occur preferentially in the central equatorial Pacific, with a more limited range of ENSO spatial patterns (Capotondi et al., 2020a), a behavior that seems to align with the model's inability to simulate extreme events in the eastern Pacific. In addition, the model appears to underestimate ENSO's nonlinearities, as quantified by the quadratic fit coefficient, α, in PC1-PC2 space (Dommenget et al., 2013; Karamperidou et al., 2017; Cai et al., 2018), which is smaller in CESM2 compared to reanalysis (Fig. 5.7). An extensive examination of several climate models using our methodology is beyond the scope of this paper. However, our analysis of CESM2 demonstrates the potential value of our approach for achieving a detailed assessment of ENSO diversity in climate models. Such an investigation, as well as the analysis of possible changes in ENSO types in future climate scenarios, will be considered in subsequent studies.

Open Research

The data on which this article is based are publicly available in Zhang et al. (2019), Hersbach et al. (2020), COBE (2006), Rayner et al. (2003), ORAS5 (2021), Giese et al. (2011), Behringer et al. (1998), and Laloyaux et al. (2018), with their details are listed in Tab. 3.2. Our code is publicly available under Schlör (2023) and `https://github.com/jakob-schloer/LatentGMM.git`.

Acknowledgments

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting J. Schlör and F. Strnad. Furthermore, we express our gratitude to the NOAA Physical Science Laboratory for making their resources available for this study. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. We acknowledge support from the Open Access Publication Fund of the University of Tübingen. A. Capotondi was supported by the NOAA Climate Program Office's Climate Variability and Predictability (CVP) and Modeling, Analysis, Predictions and Projections (MAPP) programs and by DOE Award No. DE-SC0023228.

A Hybrid Model for ENSO Dynamics in the Low-Data Regime

6.

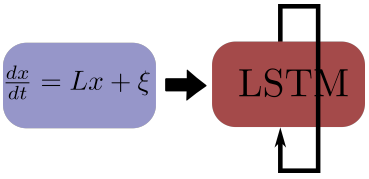
Abstract

While deep-learning models have demonstrated skillful El Niño Southern Oscillation (ENSO) forecasts up to one year in advance, they are predominantly trained on climate model simulations that provide thousands of years of training data at the expense of introducing climate model biases. Simpler Linear Inverse Model (LIM) trained on the much shorter observational record also make skillful ENSO predictions but do not capture predictable nonlinear processes. This motivates a hybrid approach, combining the LIM's modest data needs with a deep-learning non-Markovian correction of the LIM. For O(100 yr) datasets, our resulting Hybrid model is more skillful than the LIM while also exceeding the skill of a full deep-learning model. Additionally, while the most predictable ENSO events are identified in advance by the LIM, they are better predicted by the Hybrid model, especially in the western tropical Pacific for leads beyond about 9 months, by capturing the subsequent asymmetric (warm versus cold phases) evolution of ENSO.

6.1. Introduction

In the past year, deep learning has revolutionized weather forecasting with models such as GraphCast (Lam et al., 2023), Pangu (Bi et al., 2023), and FourCastNet (Pathak et al., 2022) now outperforming traditional state-of-the-art numerical weather prediction models (Ben-Bouallegue et al., 2023). These advancements are largely due to the availability of millions of data points from reanalysis data on hourly resolution, which allows the training of large neural networks with negligible generalization error. While these models demonstrate exceptional medium-range forecasting skills, it remains unclear if these capabilities extend to seasonal or annual predictions. Unlike medium-range forecasting, which is mainly dependent on initial conditions, long-range forecasts of the climate are primarily shaped by boundary forcing.

Code is available at the Github repository: jakob-schloer/HybridLIM



A prominent example of seasonal to annual forecasting is El Niño-Southern Oscillation (ENSO). Characterized by tropical Pacific sea surface temperature anomaly (SSTA), ENSO, significantly influences global climate patterns and is thereby a primary source of seasonal to annual predictability. Its events are characterized by anomalously warm (cold) tropical SSTA, which exhibit a rich diversity in their spatial structure, temporal evolution (Capotondi et al., 2015; Timmermann et al., 2018), and impact on extreme weather conditions worldwide (Taschetto et al., 2020; Strnad et al., 2022). Subsequently, early forecasts of not only the ENSO event likelihood but also its spatial structure are of great value for agriculture and society globally (Callahan et al., 2023).

Deep learning models have demonstrated the capability to produce skillful ENSO forecasts (Ham et al., 2019; Petersik et al., 2020; Cachay

et al., 2021; Zhou et al., 2023). However, the short observational record hampers training these models directly on observational data. To address this, models are trained on simulations from global climate models like models of the CMIP5 and CMIP6 suite. While providing thousands of years of data, these simulations possess inherent biases due to their resolution and the parametrization of unresolved processes, for instance, by overestimating the variability in the Western tropical Pacific (Capotondi, 2013; Chen et al., 2021; Beverley et al., 2023). When trained on simulations the climate model biases transfer to the neural network. Transfer learning, used by Ham et al. (2019), presents a potential workaround for bridging the gap between simulation data and observations. Nevertheless, Zhou et al. (2023) found no significant improvement in model performance when fine-tuning their transformer with reanalysis data, indicating the need for more research in this domain.

The Linear Inverse Model (LIM), first introduced by Penland and Sardeshmukh (Penland et al., 1995), offers an alternative data-driven approach for ENSO forecasting. When fitted on the short observational record alone, LIM demonstrates annual-to-seasonal forecasting skill on par with the North American Multi-Model Ensemble (NMME) (Kirtman et al., 2014) mean forecast, a set of coupled ocean-atmosphere models (Newman et al., 2017). The LIM describes the slower-varying ocean dynamics as a linear system of SSTA and SSHA driven by stochastic atmospheric forcing. The noise forcing is assumed to be white in time but correlated in space, which is modeled by a multivariate Gaussian distribution. Refinements to the LIM, such as the cyclostationary (CS)-LIM to account for the annual cycle (Shin et al., 2021), the inclusion of ocean memory (Chen et al., 2016), and state-dependent noise (Martinez-Villalobos et al., 2018), have improved its predictive skill. However, nonlinear processes in the system are treated as part of LIM's white noise component and are not captured deterministically.

The tropical Pacific SSTA distribution, however, is skewed, with warmer events being more intense and frequent than cooler ones. This asymmetry, not captured in the linear dynamics of the LIM, has been attributed to nonlinear processes in the region (Takahashi et al., 2011; Takahashi et al., 2016; Okumura, 2019; Geng et al., 2022). This raises the question, of whether we can model these nonlinear dynamics by capturing the residual dynamic between the LIMs prediction and observations.

Hybrid models, which combine numerical or empirical models with neural networks, offer a promising approach to capturing unresolved dynamics in the low-data regime (Irrgang et al., 2021). Instead of learning the complex system dynamics of the full system with a neural network, hybrid approaches are more data efficient due to their simpler learning objective. Applications include post-processing of weather forecasts (Gneiting et al., 2005; Bauer et al., 2015) and machine learning parametrization in coupled ocean-atmosphere models (Rasp et al., 2018; Watt-Meyer et al., 2021; Kochkov et al., 2023).

In the context of ENSO forecasting, Goel et al. (Goel et al., 2017) combined an RNN with a vector autoregressive model for time-series forecasting. Similarly, Wang et al. (Wang et al., 2021) and Zhou et al. (Zhou et al., 2022) integrated the principal oscillation pattern, akin to the LIM, with an RNN for forecasting ENSO indices. Expanding beyond ENSO index prediction, Rodrigues et al. (Rodrigues et al., 2021) developed a hybrid

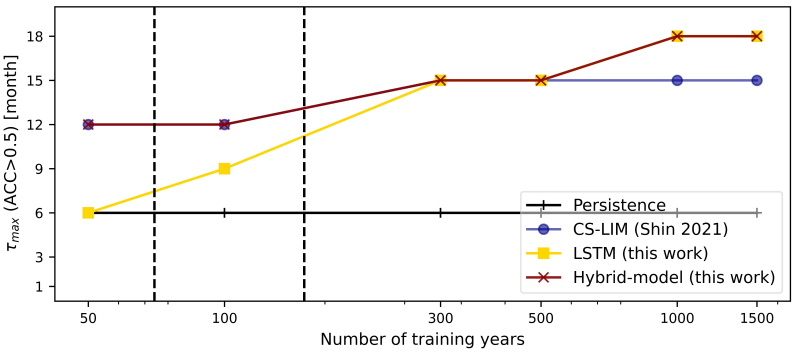


Figure 6.1.: Variations in forecast skill over training data length. The maximum lead time of a forecast to be skillful is defined by its ACC exceeding 0.5. Training on 70 to 150 years of monthly data (equivalent to ORAS5 and CERA-20C, shown with dashed lines), both the CS-LIM and the Hybrid-model outperform the pure LSTM model, which is only marginally better than a persistent forecast. LSTM reaches a 15-month skillful lead time, matching CS-LIM and Hybrid, with 300 years of data. Beyond 500 years of data, both the LSTM and the Hybrid-model surpass the CS-LIM, showing a skillful forecast of up to 18 months.

neural network using the LIM operator within a ResNet-like architecture for global SSTA prediction. A different approach is taken by Wang and Huang (Wang et al., 2024), their fusion model consists of a CNN that is first trained to forecast tropical Pacific SSTA and then perform a weighted average of the CNN with the NMME forecast. Except for the work by Wang and Huang, these Hybrid-models have not yet achieved state-of-the-art forecasting skills, and their potential advantages over fully deep learning models, such as interpretability and data efficiency, remain underexplored.

Here, we propose a Hybrid deep-learning model for ENSO forecasting that combines the LIM with a Long-Short Term Memory (LSTM) network (Hochreiter et al., 1997). This model is designed to capture the residuals between LIM forecasts and target data, thereby improving seasonal forecast accuracy. We diverge from existing methodologies by including seasonality in both the LIM and the LSTM. Furthermore, we adapt both the Hybrid-model and the LSTM to generate probabilistic ensemble forecasts of the full field variables, SSTA and SSHA. This is achieved by employing a set of output layers that generate ensemble members, designed to match the nonparametric distribution of the target data (Lessig et al., 2023).

We conduct a comparative analysis between our Hybrid-model and our fully deep-learning model, LSTM, of ENSO dynamics, focusing on their data requirements. Training on 70 to 150 years of monthly data, comparable in length to reanalysis products like ORAS5 and CERA-20C, both the CS-LIM and the Hybrid-model exhibit a skillful forecast up to 12 months, as indicated by an anomaly correlation coefficient (ACC) exceeding 0.5 (Fig. 6.1). In contrast, the pure LSTM model's forecast skill, marginally better than a persistent forecast, extends only to 6-9 months. The LSTM model requires 300 years of data to reach a forecast skill of 15 months on par with the CS-LIM and Hybrid-model. With the training dataset size exceeding 500 years, both the LSTM and the Hybrid-model surpass the CS-LIM, showing a skillful forecast of up to 18 months.

Our evaluation is performed on 2000 years of pre-industrial control simulations from the Coupled Earth System Model 2 (CESM2) (Danabasoglu et al., 2020), a state-of-the-art coupled global climate model known for its reasonable representation of ENSO diversity (Capotondi et al., 2020a). All metrics are computed for a 200-year test period as detailed in Sec. 6.2.2.

6.2. Data and Methods

6.2.1. Data

Training DL models for ENSO prediction with monthly data is limited by the short observational record. To circumvent this, we use the 2000-year CESM2 pre-industrial control simulation (Danabasoglu et al., 2020), focusing on monthly SST and SSH data in the tropical Pacific region (130°E - 70°W, 31°S - 32°N), which we linearly interpolate to a resolution of 1°x1°. SSH is a proxy for the upper ocean temperatures and the thermocline depth. Despite the lack of external forcing in the control simulation, we observe a trend in SST data. We linearly detrend the data and remove the seasonal cycle by subtracting the monthly climatology which is obtained over the training set (0-1500). As SSTA and SSHA differ in units and scales, we perform a z-score normalization before model training. Both the LIM and Hybrid-model require us to reduce the dimensionality, which is achieved using Empirical Orthogonal Function (EOF) analysis. The dataset is divided into training (75%, year 1-1500), validation (15%, years 1500-1800), and test set (10%, 1800-2000), where the validation set is used for refining the hyperparameters of our models.

6.2.2. Methods

The objective of our study is to accurately predict SSTA and SSHA fields for a specified forecast lag time τ. We define the stacked variable fields at a given time t as $\mathbf{x}(t) = (\mathbf{x}_{\text{SSTA}}, \mathbf{x}_{\text{SSHA}})(t)$, where each field spans the tropical Pacific $\mathbf{x}_{\text{SSTA/SSHA}}(t) \in \mathbb{R}^{N_{lat} \times N_{lon}}$. We estimate a function f that predicts the future state, $\hat{\mathbf{x}}(t+\delta)$ at an incremental time step δ, by

$$\mathbf{x}_i(t+\delta) = f_{m(t)}\left(\mathbf{x}(t), f_{m(t-\delta)}\left(\mathbf{x}(t-\delta), \ldots, f_{m(t-t_{\text{hist}})}\left(\mathbf{x}(t-t_{\text{hist}})\right)\right)\right), \quad (6.1)$$

The autoregressive prediction is based on the previous states, $\mathbf{x}(t-\delta), \ldots, \mathbf{x}(t-t_{\text{hist}})$, referred to as history. All our model forecasts consist of i ensemble members, which allows us to estimate the model uncertainty. The dynamics of the tropical Pacific Ocean show a strong seasonal phase locking. We therefore condition f on the month of the year, $m(t)$. The month conditioning depends on the model architecture and will be discussed for each model separately.

Empirical Orthogonal Function (EOF)

Estimating the linear evolution operator of the LIM requires a matrix inversion of the number of input dimensions. The matrix inversion is intractable for the full spatial fields of SSTA and SSHA. For this reason, each state $\mathbf{x}(t)$ is transformed into a lower-dimensional state $\mathbf{z}(t) = (\mathbf{z}_{\text{SSTA}}, \mathbf{z}_{\text{SSHA}})$. Dimensionality reduction of the SSTA and SSHA fields in the tropical Pacific is achieved through EOF analysis, utilizing the first 20 Principal Components (PCs) for SSTA and the first 10 PCs for SSHA. Including higher-order PCs does not affect our results. Forecasting in the lower dimensional space is then equivalently conducted on these

PCs, as formulated in Eq. 6.1. For analysis and evaluation, we transform our forecast back to grid space. To adequately replicate the high spatial frequencies of the input fields, we add variability by randomly sampling loadings of the higher-order PCs (20-300) (citation) for both SSTA and SSHA at each time step. These random features are then combined with their respective EOFs and added to the forecast fields, ensuring a closer match to the spatial intricacies of the original data.

Linear Inverse Model

The Linear Inverse Model (LIM) describes the dynamic of the tropical Pacific as a multivariate linear system subject to stochastic forcing from the atmosphere. The underlying dynamic of such a system is described by a linear stochastic differential equation,

$$\frac{d\mathbf{z}}{dt} = \mathbf{L}\mathbf{z} + \zeta \tag{6.2}$$

where $\mathbf{L}$ is the linear operator describing the dynamics of $\mathbf{z}$ and $\zeta \sim \mathcal{N}(0, \mathbf{Q})$ is a noise vector that is uncorrelated over time but spatially correlated, as encoded in the covariance matrix $\mathbf{Q}$. Forecasts of $\mathbf{z}$ for a lag time τ are given by the transition probability as,

$$p\left(\mathbf{z}(t+\tau)|\mathbf{z}(t)\right) = \mathcal{N}\left(\mathbf{z}(t+\tau); \mu_\tau(t), \Sigma_\tau\right) \tag{6.3}$$

$$\text{with} \qquad \mu_\tau(t) := e^{\mathbf{L}\tau}\mathbf{z}(t) \qquad \text{and} \qquad \Sigma_\tau := \int_0^\tau e^{\mathbf{L}s}\mathbf{Q}\mathbf{Q}^T e^{\mathbf{L}^T s} ds, \tag{6.4}$$

where $\mu_\tau(t)$ is the infinite ensemble mean forecast and Σ_τ is the forecast covariance matrix.

Penland et al. (1995) show that the linear operator $\mathbf{L}$ and the noise covariance $\mathbf{Q}$ can be estimated from the data under two assumptions. First, the system has to be statistically stationary which allows us to write the Fluctuation-Dissipation relationship as

$$\mathbf{L}\mathbf{C}(0) + \mathbf{C}(0)\mathbf{L}^T + \mathbf{Q} = 0, \tag{6.5}$$

where $\mathbf{C}(0) = \langle \mathbf{z}(t)\mathbf{z}^T(t) \rangle$ is the spatial covariance matrix. Secondly, the autocorrelation of the system decays with lag time τ which can be expressed using the time-lag covariance matrix $\mathbf{C}(\tau) = \langle \mathbf{z}(t+\tau)\mathbf{z}^T(t) \rangle$ as

$$\lim_{\tau \to \infty} \mathbf{C}(\tau)\mathbf{C}(0) = 0 \Rightarrow \mathbf{C}(\tau) = e^{\mathbf{L}\tau}\mathbf{C}(0), \tag{6.6}$$

where $\mathbf{G}(\tau) := \exp(\mathbf{L}\tau)$ is the Greens function that must tend to zero for long lag times. Typically, both assumptions hold for detrended anomaly data of a chaotic system.

Once $\mathbf{L}$ and $\mathbf{Q}$ are estimated from the data, we obtain forecast trajectories from an initial time t to $t + T$, by numerically integrating Eq. 6.2 using the forward Euler-method with incremental update steps δ as

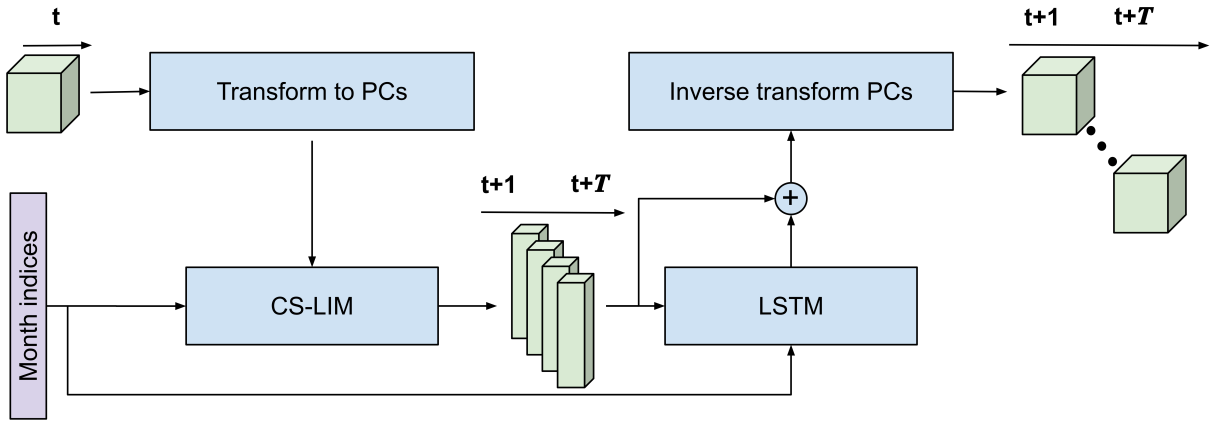


Figure 6.2.: Schematic Representation of the Hybrid-Model. First, the initial state at time t is projected onto the PCs. This is followed by an ensemble forecast using the CS-LIM which is conditioned on the forecast months. Subsequently, the LSTM adjusts each ensemble member of the linear CS-LIM forecast. Finally, the refined forecast is transformed back to grid space by multiplying the PCs with the respective EOF patterns.

$$\mathbf{z}(t+\delta) = \mathbf{z}(t) + \mathbf{L}\mathbf{z}(t)\delta + \zeta\sqrt{\delta}, \tag{6.7}$$

where $\zeta \sim \mathcal{N}(0, \mathbf{Q})$ is a random sample from the noise distribution. We create n ensemble member trajectories from t to $t+T$ when integrating the system n-times. The infinite ensemble member mean is given by $\mu(t)$ in Eq. 6.3.

In the equatorial Pacific, the variance in SSTA shows a distinct annual pattern with low variance during the boreal spring and high variance in the boreal winter. This peak in winter variance aligns with the occurrence of the most intense warm and cold ENSO events, a phenomenon referred to as "ENSO phase locking" (Rasmusson et al., 1982). Shin et al. (2021) showed that including seasonality in the LIM improves its forecast reliability. Their cyclostationary (CS)-LIM involves estimating unique linear operators and noise covariances for each month, indexed using $m(t) = 1, 2, \ldots, 12$. The numerical integration of the stationary (ST)-LIM outlined in Eq. 6.7 changes to

$$\mathbf{z}(t+\delta) = \mathbf{z}(t) + \mathbf{L}^{CS}_{m(t)}\mathbf{z}(t)\delta + \zeta^{CS}_{m(t)}\sqrt{\delta}, \tag{6.8}$$

where $\zeta^{CS}_j \sim \mathcal{N}\left(0, \mathbf{Q}^{CS}_j\right)$. The CS-LIM forms the base model of our Hybrid-model outlined in the following.

Hybrid-model

We introduce a novel Hybrid-model that combines the LIM with an LSTM network. While the LIM captures the predictable linear dynamics, the LSTM learns the residuals between the LIM predictions and the actual data, thus the nonlinear dynamics. Our methodology is schematically detailed in Fig. 6.2.

During inference, we project the initial state of the tropical Pacific, $\mathbf{x}(t)$, onto the leading EOFs and employ the LIM to predict future states over

$\tau = 1, \ldots, T$ timesteps. For each timestep, $t + \tau$, we predict a correction, $\hat{\mathbf{z}}_{\text{res}}(t + \tau)$, to the LIM forecast, $\hat{\mathbf{z}}_{\text{LIM}}(t + \tau)$. The final forecast is thus defined as:

$$\hat{\mathbf{z}}(t + \tau) = \hat{\mathbf{z}}_{\text{LIM}}(t + \tau) + \hat{\mathbf{z}}_{\text{res}}(t + \tau). \tag{6.9}$$

The nonlinear correction is modeled by the LSTM as $\hat{\mathbf{z}}_{\text{res}}(t + \tau) = \text{LSTM}(\hat{\mathbf{z}}_{\text{LIM}}(t + \tau), \mathbf{h}(t))$, where $\mathbf{h}(t)$ is the latent state of the LSTM which aggregates the information of previous states. The LSTM is selected not only for its ability to capture nonlinear relationships inherent in deep neural networks but also for learning non-Markovian dynamics.

To include seasonality within the LSTM, we introduce a learned affine transformation to its latent state (Perez et al., 2017), $\mathbf{h}(t)$, as follows:

$$\mathbf{h}_{m(t)}(t) = (1 + \alpha_{m(t)})\mathbf{h}(t) + \beta_{m(t)}, \tag{6.10}$$

where $\alpha_{m(t)}$ and $\beta_{m(t)}$ represent embeddings for each month, enabling the network to adapt its latent state dynamically to seasonal variations.

The LSTM component is configured to process the forecast from each CS-LIM ensemble member, represented as $[\hat{\mathbf{z}}(t + 1), \ldots, \hat{\mathbf{z}}(t + T)]$. At each time step, the input to the LSTM network is linearly projected into a higher-dimensional latent space where it is processed by two consecutive LSTM layers. By combining the LIM forecast at each timestep with the LSTM's hidden state from the previous time step, the model iteratively accumulates information across the entire forecast sequence. Finally, the predicted latent states are linearly projected back onto the PCs and added to the original LIM prediction. The combined prediction is then transformed to grid space by multiplication with the respective EOFs, as described in Sec. 6.2.2.

Purely deep learning baselines

To verify the utility of our Hybrid-model, we provide a comparison against fully neural network-based approaches. Similar to the structure of our Hybrid-model, we construct an LSTM that operates in the PC space. Additionally, we explore the application of a Convolutional LSTM (ConvLSTM (Shi et al., 2015)) architecture in grid space. We employ a custom variant of ConvLSTM, termed SwinLSTM, that has been specifically tailored to perform forecasts on fields with large-scale spatial structures.

Both the PC-LSTM and the SwinLSTM have a standard Encoder-Decoder structure used for sequence-to-sequence modeling (Sutskever et al., 2014), as depicted in Fig. C.1. Unlike the LIM and Hybrid-model, these models incorporate information from time points preceding the initialization time. The Encoder network is designed to aggregate this historical information into a latent state which initializes the Decoder model. It begins with a downsampling block that transforms the history into a higher-dimensional latent space, which is then processed by two LSTM layers, where the input is added to the hidden state from the previous time step. The hidden state of the second LSTM layer at time t is then passed to the Decoder network. The Decoder mirrors the Encoder with

two LSTM layers of its own, which do not require any additional input other than the hidden state and can thus be rolled out over the full prediction horizon $t + T$, transferring the hidden state autoregressively for each successive prediction time. This is followed by an upsampling block that transforms the hidden state back to the input grid space. To generate 16 ensemble members, we employ a separate upsampling block for each member. Similar to the Hybrid-model, we incorporate seasonal information through a learned affine transformation in both the Encoder and Decoder networks.

PC-LSTM In the PC-LSTM, the downsampling consists of an EOF-truncation of the entire SSTA and SSHA fields onto their respective PCs (Sec. 6.2.2), followed by a learned linear layer that projects the data into the latent space. The latent states are recursively predicted using LSTM layers, in both the encoder and the decoder network. Equivalently to the downsampling, the upsampling block starts with a linear layer to project the latent space back to the PCs, followed by a projection on their respective EOFs to yield a forecast of the variables in grid space.

SwinLSTM We employ a second DL model that does not involve EOF-truncation of data. This model, which we refer to as the shifting window (Swin) LSTM, was inspired by the popular Swin-Transformer architecture (Liu et al., 2021) a state-of-the-art DL architecture for spatial and spatiotemporal data. The model incorporates a ConvLSTM-like Encoder-Decoder architecture with a customized LSTM cell (Fig. C.2). Unlike the PC-LSTM, the input, latent, and hidden states in our model maintain dimensions of channel, height, and width, albeit of varying sizes. For model input, both variables are stacked along the channel dimension. The encoder initially downscales the height and width dimensions to reduce computational costs, while simultaneously expanding the channel dimension via a strided convolutional layer. Following the methodology of Liu et al. (2021) and Liu et al. (2022b) we use equal stride and kernel sizes for this initial encoding, effectively partitioning the input into small patches of equal size (4x4 grid steps in latitude/longitude direction). This method of dimensionality reduction differs from the EOF as it is learned end-to-end with the forecasting model, as well as being based on local patches, rather than global features as encoded in EOFs. Our approach further adapts the standard ConvLSTM layer, in a similar fashion to Liu et al. (2022b), by splitting the convolution into spatial and channel mixing components, interspersed with layer normalizations, as depicted in Fig. C.2. This modification facilitates two major improvements over standard ConvLSTMSs, (i) the spatial mixing component enables a substantially larger receptive field while reducing the number of parameters and (ii) the added normalization stabilizes training and supports conditioning on monthly embeddings, implemented through the affine transformation outlined in Eq. 6.10. Finally, the mirrored Decoder network, which also consists of two SwinLSTM layers and a strided and transposed convolution, transforms the aggregated hidden state back into the full-resolution grid space. An ensemble of 16 separate final projection layers is used to generate a probabilistic prediction and the model is trained using the Ensemble CRPS as detailed below.

Optimization procedure

Each model is designed to produce a probabilistic prediction by generating an ensemble of 16 sequences. We train the models to enhance both the accuracy of individual predictions and the spread among ensemble members. This is achieved by employing the CRPS for optimization. The CRPS is a probabilistic metric that compares the cumulative probability distribution (CDF) of the forecast to the CDF of the target. The target, $\mathbf{x}(t)$, is a single observation, and thus its CDF is a step function. A perfect CRPS score of zero would be a Dirac delta-like predictive probability density function centered at the target, for which the CDF would be the same step-function as is for the target. The CRPS has an analytic expression for parametric distributions, like the Gaussian distribution, but also a statistical form for empirical distributions (Hersbach, 2000; Gneiting et al., 2007).

For our M-ensemble member prediction, we compute the pixel-wise CRPS for empirical distributions as,

$$\mathrm{CRPS}(\hat{X}_s(t), x_s(t)) = E[|\hat{X}_s(t) - x_s(t)|] - \frac{1}{2} \cdot E\left[\left|\hat{X}_s(t) - \hat{X}_s{}'(t)\right|\right], \quad (6.11)$$

where $\hat{X}_s(t)$ and $\hat{X}_s{}'(t)$ are the M-ensemble member prediction and $x_s(t)$ the target value at each location $s = 1, \ldots, 2 \cdot N_{lat} \cdot N_{lon}$. We optimize the parameters of each model by minimizing the averaged CRPS between the observed data, $\mathbf{x}(t+\tau)$, and its ensemble hindcast, $\hat{\mathbf{X}}(t+\tau)$, across all locations and lag $\tau \in [1, T]$. Using Eq. 6.11, we define our tailored loss function as:

$$l\left(\hat{X}(t), \mathbf{x}(t)\right) = \sum_{\tau=1}^{T} \gamma^{\tau} \sum_{s=1}^{2 \cdot N_{lat} \cdot N_{lon}} \mathrm{CRPS}\left(\hat{X}_s(t+\tau), x_s(t+\tau)\right) \quad (6.12)$$

$$(6.13)$$

To address the greater loss values at longer forecast lags, we introduce a power-law decaying weight over lag time, γ^{τ}, where γ is an empirically set hyper-parameter.

In addition, we use the AdamW adaptive gradient algorithm (Loshchilov et al., 2019) in conjunction with cosine annealing (Loshchilov et al., 2017) to dynamically adjust the learning rate during the training phase.

Evaluation metrics

Our analysis of the models, all of which generate ensemble member predictions, is based on probabilistic metrics as well as deterministic metrics of their ensemble mean. We evaluate all models on the test set (200 years) using SSTA and SSHA in the tropical Pacific.

Anomaly correlation coefficient (ACC) The ACC is the temporal correlation between the model forecast time series $\hat{x}_s$ and the target time series x_s at spatial location s. The ACC is defined as,

$$\mathrm{ACC}_s = \frac{\mathrm{cov}(\hat{x}_s(\tau), x_s)}{\hat{\sigma}_{s,\tau}\sigma_s} \tag{6.14}$$

where $\mathrm{cov}(x_s(\tau), x_s) = \langle \hat{x}_s(\tau), x \rangle$ is the covariance between the time series and σ_s their respective variance.

Root mean squared error (RMSE) The RMSESS is a deterministic metric that compares the RMSE of the model to the RMSE of a reference forecast. Throughout this work, we choose the climatology as our reference forecast. The RMSESS can be written as

$$\mathrm{RMSESS}_s(\tau) = 1 - \frac{\mathrm{RMSE}_{\mathrm{model},s}(\tau)}{\mathrm{RMSE}_{\mathrm{ref},s}(\tau)} = 1 - \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{x}_s(t+\tau) - x_s(t+\tau))^2}}{\sigma_s} \tag{6.15}$$

where σ_s is the variance of the data at point s. An RMSESS of 1 is a perfect model forecast and 0 is as good as the climatology forecast.

Continuous Ranked Probability Score (CRPS) Equivalently to the skill score of the RMSE, we define the CRPS skill score using Eq. 6.11 as, $\mathrm{CRPSS}(\hat{X}_s, x_s) = 1 - \mathrm{CRPS}_{\mathrm{model}}(\hat{X}_s, x_s)/\mathrm{CRPS}_{\mathrm{ref}}(\hat{X}_{\mathrm{ref},s}, x_s)$, where the reference forecast $\hat{X}_{\mathrm{ref},s}$ is the climatology. A CRPSS of 1 is a perfect model forecast and 0 is as good as the climatology forecast.

6.3. Results

6.3.1. Improved Skill due to Nonlinearities

The CS-LIM estimated from the PCs of SSTA and SSHA using CESM2 pre-industrial control run in the tropical Pacific, as described in Sec. 6.2.2, captures all predictable linear dynamics of the system. Our Hybrid-model is designed to have the LSTM learn the residuals between the LIM's predictions and the target data. The improvement of the Hybrid-model upon the LIM itself can thus be attributed to the ability of the Hybrid-model to capture the nonlinearities of the tropical Pacific Ocean dynamics.

As an example, we compare a 24-month hindcast of tropical SSTs conducted with the LIM and Hybrid-model (Fig. 6.3) for a chosen El Niño exemplar. The forecasts are initialized in December, 12 months before the peak of the El Niño event (dashed line in Fig. 6.3a). El Niño events are identified when the Niño3.4 index exceeds its standard deviation for at least three consecutive months. The evolution of the average SSTA in the Niño4 region of the target is presented as a black line in Fig. 6.3a. Throughout this work, we analyze the Niño4 instead of the Niño3.4 region because of the west Pacific bias in CESM2. Both the CS-LIM

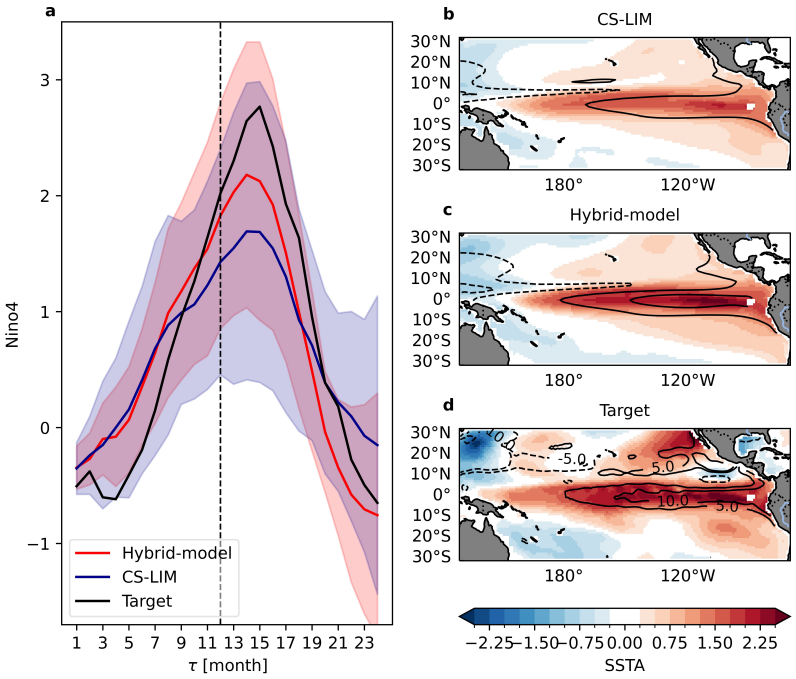


Figure 6.3.: Example El Niño hindcast. Example of a hindcast initialized 12 months prior to an EN event exemplar. The Niño4 mean (solid line) and standard deviation (shading) of the 16 ensemble members of the CS-LIM and Hybrid-model forecast are shown in a. The dashed line in (a) indicates the 12-month lag. The mean SSTA forecast at τ=12 months for the CS-LIM (b) and Hybrid-model (c) show warming patterns similar to the target (d). Mean SSTA are shown as color shadings while mean SSHA are depicted as contour lines.

(red line) and Hybrid-model (blue line) forecasts predict the observed warming, as depicted by the ensemble members' mean, with the shading indicating their respective standard deviations. The SSTA and SSHA field forecasts at τ=12 months lag time (Fig. 6.3c, d), exhibit the El Niño warming in the equatorial Pacific, although they lack some smaller spatial structure evident in the target fields. Notably, the Hybrid-model forecast magnitude is closer to the CESM2 target data than the CS-LIM forecast. The Hybrid-model's uncertainty range, given by the standard deviation between 16 ensemble member forecasts, is also narrower than that of the CS-LIM forecast, yet still encompasses the target data, which might suggest a more accurate ensemble spread.

We evaluate the RMSE and CRPS skill scores of the CS-LIM and Hybrid-model on 200 years of test data to quantify the improvement in skill. Both skill scores are obtained with respect to monthly climatology as a baseline model. Various LIM variants, including different influences on the tropical Pacific dynamics, are constructed to ensure that we use the best linear model.

The initial LIM variant, formulated by Penland et al. (1995), is solely fitted to the first 30 PCs of SSTA in the Pacific and is termed stationary (ST)-LIM (ssta). An advancement to this is the CS-LIM (ssta), introduced by Shin et al. (2021), which includes seasonal variation and substantially surpasses the skill of ST-LIM (ssta), as shown by the RMSE (a) and CRPS (b) skill scores of the Niño4 index in Fig. 6.4. For reference, we present the skill of the persistence forecast which is worse than a climatological forecast (dashed line at zero) after $\tau = 6$ months forecast lag time.

In line with Chen et al. (2016)'s insight on the role of ocean variables in forecasting, a third variant, the CS-LIM (ssta, ssha), incorporating the first 10 PCs of SSHA in the tropical Pacific is employed. We select SSH for its model and observational accessibility and its strong relationship with important ENSO precursors, namely the upper ocean heat content and

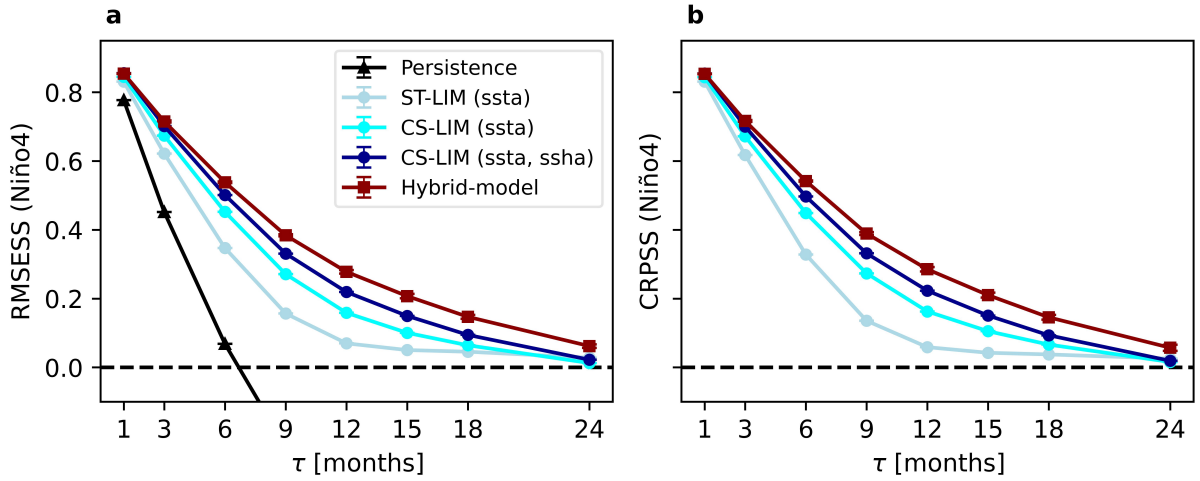


Figure 6.4.: RMSE and CRPS skill scores of the LIM versions and our Hybrid-model. Skill scores for RMSE (a) and CRPS (b) across various LIM versions and the Hybrid-model are evaluated over forecast lag time (τ) using the average SSTA in the Niño4 index region on the test set. The progression in LIM versions from the stationary (ST)-LIM, which uses SSTA data and does not include seasonally-varying operators, to the more advanced cyclostationary (CS)-LIM incorporating seasonality, and then to the CS-LIM (ssta, ssha) that includes both seasonality and SSH factors is depicted. Enhancing the CS-LIM (ssta, ssha), our Hybrid-model utilizes an LSTM to effectively learn and adjust for its residuals.

thermocline depth. This version leads to additional skill enhancement relative to the CS-LIM (ssta). Progressive improvements are evident in each LIM version, with the inclusion of seasonality and ocean variables contributing significantly to enhanced skill(Fig. 6.4). The CS-LIM (ssta, ssha) (from now on only CS-LIM) exhibits the highest skill upon the linear models and thus captures all known predictable linear dynamics.

Our Hybrid-model builds upon the CS-LIM by using an LSTM that learns to correct the error between the LIM forecast and the target data. The LSTM takes 16 ensemble member forecasts of the CS-LIM as input and learns their residuals to the target data by minimizing the CRPS loss function detailed in Eq. 6.13. To ensure the robustness of our findings, we repeat the model training five times with varied weight initialization and data shuffling, whose variability is depicted through error bars in Fig. 6.4. The skill improvement of the Hybrid-model upon the CS-LIM is significant at lag times larger than 6 months. We argue that the skill improvements can be attributed to predictable nonlinearities because all known linear dynamics are captured by the CS-LIM.

We conducted a further examination of the seasonal dependency (Sec. C.2) and spatial distribution of skill for both the LIM and the Hybrid-model. The ensemble mean RMSESS for a 12-month CS-LIM forecast of SSTA exhibits higher skill around the equator compared to the Extratropics (Fig. 6.5a). In contrast, the RMSESS of SSHA demonstrated high skill in both the Eastern and North Western tropical Pacific (Fig. 6.5c), where the largest centers of SSH variability associated with ENSO occur (Capotondi et al., 2020a).

When comparing the Hybrid-model with the CS-LIM (Fig. 6.5b, d), there is a discernible skill improvement around the equator, with the most significant enhancement observed in the western tropical Pacific. This improvement is consistent with the patterns obtained using the CRPSS (not shown).

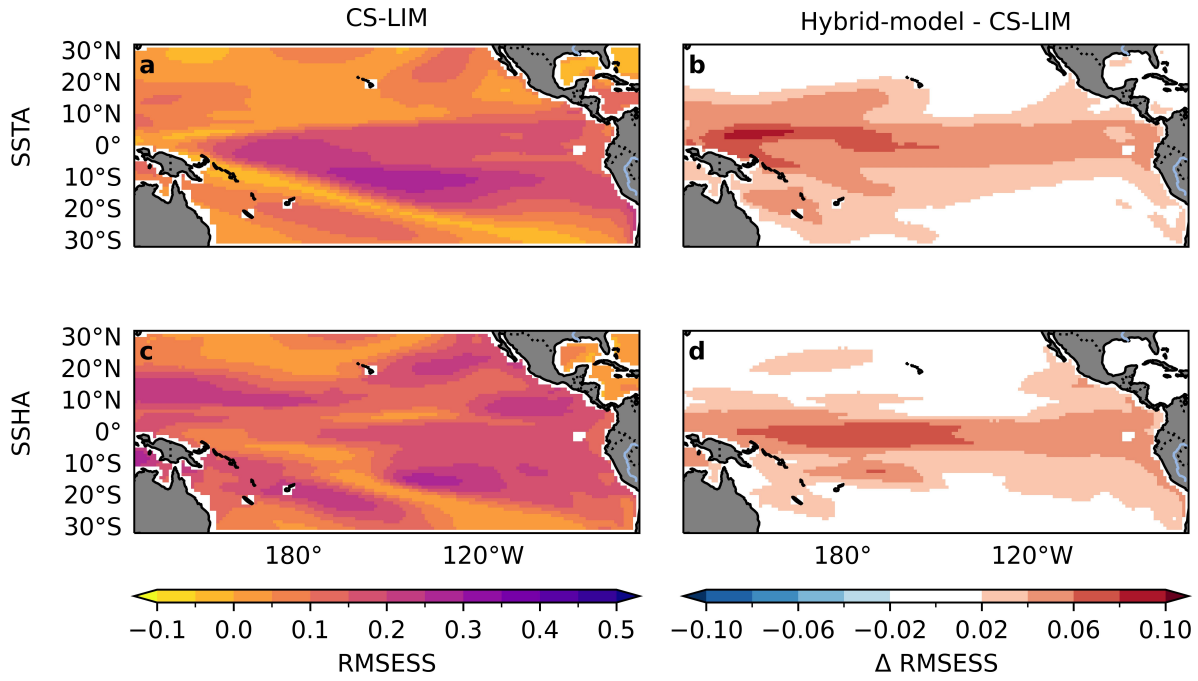


Figure 6.5.: Spatial distribution of skill improvement. RMSE skill score of SSTA and SSHA for the $\tau = 12$ month hindcast of CS-LIM (a, c) and the differences in RMSE skill scores relative to the Hybrid model (b, d). Red colors indicate an improvement in skill in the Hybrid-model, while blue colors indicate a decrease in skill. Using a two-sided t-test, we evaluate the significance of the difference between the 1000 randomly bootstrapped means of CS-LIM and the 95% confidence interval threshold.

6.3.2. Comparison to Purely Deep Learning Baselines

We conducted a comparative analysis between the Hybrid-model and the two fully deep learning models outlined in Sec. 6.2.2, the PC-LSTm, and the SwinLSTM. To ensure the robustness of our findings, each model underwent five separate training sessions with varied weight initialization and data shuffling, the variability of which is depicted through error bars in Fig. 6.6. When trained on 1500 years of data, the deep learning models have similar RMSE and CPRS skill scores to the Hybrid-model (Fig. 6.6a, b), though the SwinLSTM exhibits a slight improvement at the 12-month forecast horizon. This suggests that the Hybrid-model successfully captures most of the predictable dynamics, with the marginal gains of the SwinLSTM likely attributable to the PC truncation. Crucially, the Hybrid-model achieves this level of forecasting skill with significantly fewer parameters compared to the full deep learning models. This aspect is particularly beneficial in scenarios with limited data, as shown in Fig. 6.6c and d. With less than 500 years of monthly data, both the LIM and Hybrid-model exhibit higher 12-month RMSES and CRPS skill scores than the LSTM. This indicates that pure deep learning models trained on the observational record alone (≈70 years) have substantially worse performance than the LIM. It is only with more than 500 years of data that the forecast skill of both the Hybrid-model and the LSTM substantially improve upon the LIM.

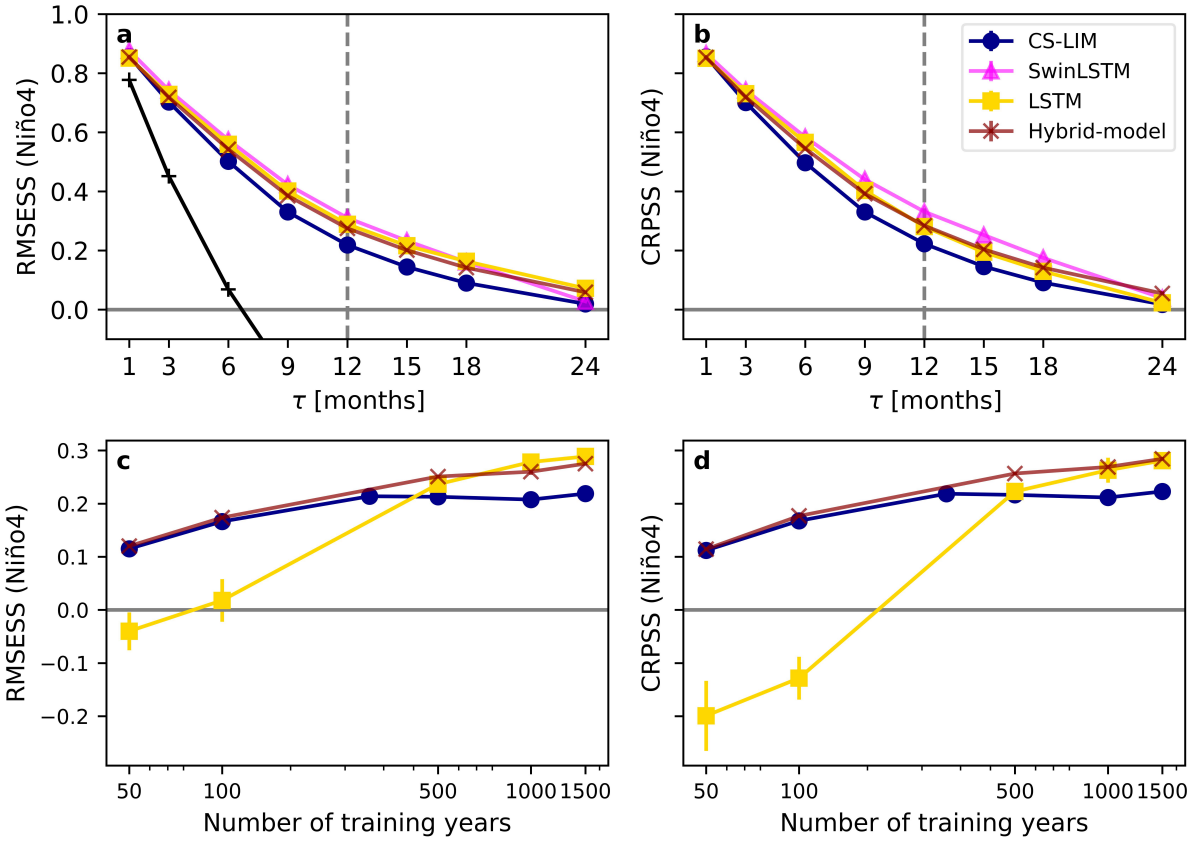


Figure 6.6.: Skill of purely deep learning baselines. When trained on 1500 years of CESM2 data, LSTM and SwinLSTM show deterministic RMSE (a) and probabilistic CRPS (b) scores comparable to the Hybrid model at all lag times τ. However, with limited data (50 or 150 years), the CS-LIM and Hybrid models surpass LSTM forecast skill, shown here for τ = 12 month (c, d). LSTM achieves parity with CS-LIM and Hybrid at 300 years, and exceeds CS-LIM with over 500 years, indicating the extensive data requirements for the pure deep learning models. Skill scores are based on monthly climatology (Sec. 6.2.2) with error bars reflecting model training runs with varied weight initialization and data shuffling.

6.3.3. Predictability Assessment in the Hybrid-model

Within the linear framework of the LIM, we can calculate the optimal initial condition (OIC) for a forecast lag time τ. The OIC, also referred to as optimal precursor, is the singular vector that aligns with the largest singular value of the forecast operator $G(\tau) = \exp(\mathbf{L}\tau)$. This condition is optimal in the sense that of all possible initial conditions of unit amplitude, it evolves into the largest state vector after time τ (Penland et al., 1995). For CS-LIM, we obtain a different OIC for every month and lag time.

The CS-LIM's OIC for a 12-month forecast in April (Fig. 6.7a) exhibits an SST and SSH structure that aligns closely with earlier research (Newman et al., 2011b; Vimont et al., 2014). Key features of the OIC are a band of positive SST anomalies in the northern subtropics, stretching diagonally from approximately 0°, 180° northeastward to around 30°N, 120°W; positive SST anomalies in the southern subtropics, predominantly east of 120°W; enhanced thermocline depth anomalies along the equator; and comparatively weak negative thermocline depth anomalies located at approximately 10°N in the eastern tropical Pacific. We obtain its subsequent evolution after τ = 12 months by applying the LIM operator to the OIC (Fig 6.7b). It is important to note that the magnitude of these

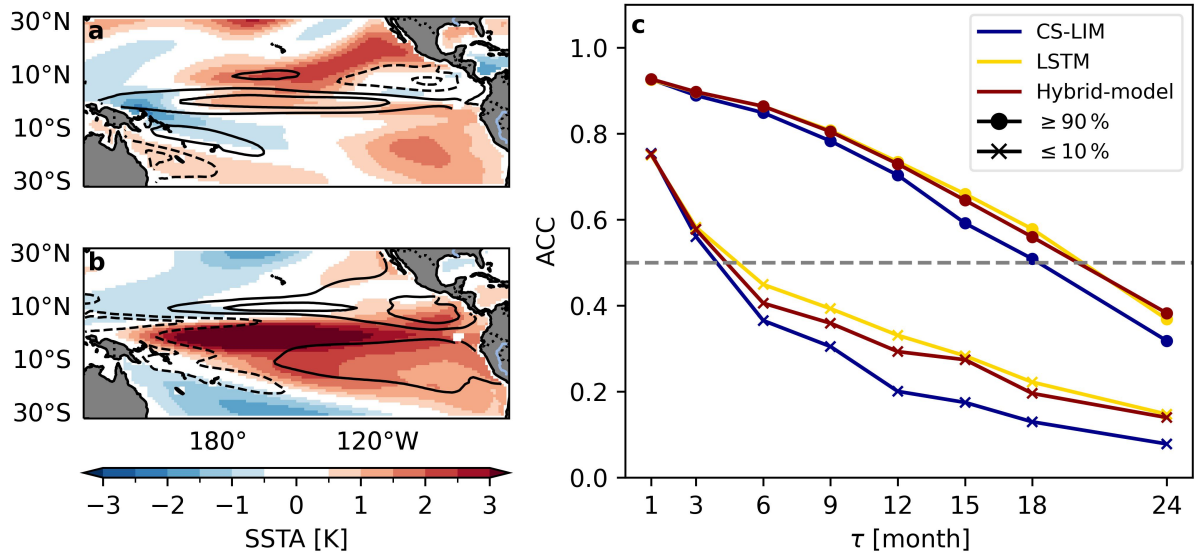


Figure 6.7.: Predictability is determined by linear optimals. The optimal initial condition (OIC) (a) of the CS-LIM for a 12-month lag forecast initialized in April evolves into an El Niño-like pattern after 12 months (b). The projection of the data on the OIC does not only indicate the potential skill of the CS-LIM but also of the Hybrid-model and LSTM (c). We compute the anomaly correlation coefficient over the CS-LIM, Hybrid-model, and LSTM forecasts where the absolute projection of the data on the optimal initial pattern is the lowest (0-10%) and the highest (90-100%).

patterns is arbitrary, a result of the unit norm of the singular vector.

For any given initial forecast state, the CS-LIM's predictability can be estimated by projecting the state onto its OIC. The strength of this projection, or how well the first singular vector aligns with the initial state – termed as optimal initial growth – is a key determinant of the potential forecast skill of the CS-LIM (Sardeshmukh et al., 2000; Newman et al., 2003). To illustrate this point, we select initial states from the test set that have either the lowest (0-10%) or highest (90-100%) projections on the OIC. The average ACC, illustrated in Fig. 6.7c, is substantially larger for hindcasts initialized from states with the highest optimal initial growth than for states with the lowest optimal initial growth.

The ACC of the Hybrid-model forecast initialized from states with the lowest and highest optimal initial growth exhibit both a clear difference in skill and also a substantial skill increase compared to the CS-LIM (Fig. 6.7c). This result implies that optimal initial growth, a property derived from the CS-LIM, influences not only linear predictability but also significantly impacts the predictability of nonlinear dynamics. This hypothesis is further supported by our analysis of the LSTM forecast, where a marked increase in skill is observed for initial states with high optimal initial growth as opposed to those with low initial growth.

6.3.4. ENSO Asymmetry is Nonlinearly Predictable

The initial states projected onto the OIC, shown in Fig. 6.7a, can yield either positive or negative optimal initial growth which evolves into warm and cold patterns, respectively. We examine the average $\tau = 12$ month hindcast of April initial states with the absolute highest optimal initial growth (>90%) for our CS-LIM, Hybrid-model, and LSTM model.

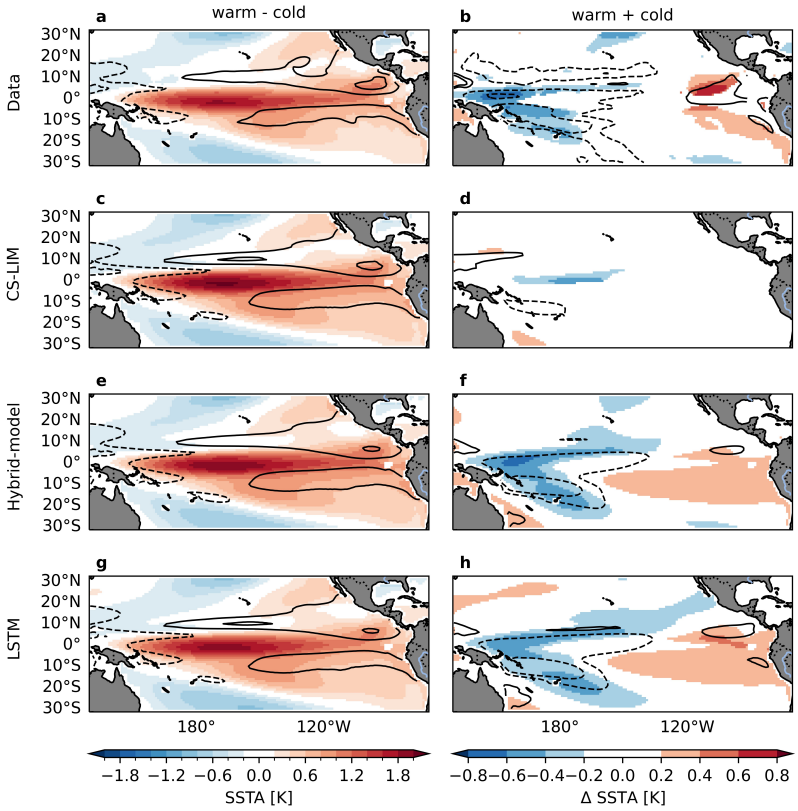


Figure 6.8.: Nonlinear models capture ENSO asymmetrie. The 12-month evolutions of states initialized in April with the absolute highest optimal initial growth (>90%) show warm and cold patterns. The average target state of warm-cold patterns (a) and warm+cold patterns (b) are depicted, as well as their hindcasts of the CS-LIM (c, d), Hybrid-model (e, f) and LSTM (g, h). While the warm-cold pattern delineates the average hindcast pattern's magnitude, the warm+cold pattern depicts their asymmetry. Using a two-sided t-test, we evaluate the significance of the difference between the means of the cold and the warm patterns. Only those values that surpass the 95% confidence interval threshold are shown.

The average hindcast pattern's magnitude and their asymmetry are presented by warm-cold patterns (Fig. 6.8, first column) and warm+cold patterns (Fig. 6.8, second column), respectively. A two-sided t-test is utilized to ascertain if there is a statistically significant difference between the means of the two distributions. We report only those values that surpass the 95% confidence interval threshold. The warm-cold pattern in the target data (Fig. 6.8a) closely resembles the evolved optimal pattern of the CS-LIM (Figure 6.7b) and its average empirical hindcast (Fig. 6.8c).

An evident east-west dipole structure is observed in the asymmetry of warm versus cold events of the target data, see Fig. 6.8b. Cold events exhibit higher SSTA and SSHA magnitudes in the western tropical Pacific, while warm events show greater magnitudes in the Eastern Pacific. This asymmetry is not present in the warm+cold patterns of CS-LIM hindcast (Fig. 6.8d), which is due to its linear and therefore symmetric evolution of cold and warm events. The remaining subtle differences between warm and cold patterns of CS-LIM hindcast likely originate from the asymmetrical distribution of initial conditions.

In contrast, the Hybrid-model and LSTM hindcasts accurately capture both the magnitude and asymmetry of warm and cold events. Both nonlinear model hindcasts exhibit the zonal dipole structure present in the target data, as shown in Fig. 6.8f and h. This finding highlights the ability of nonlinear models to predict the asymmetry between warm and cold patterns. They also underscore the Hybrid-model's potential to disentangle linear and nonlinear predictable dynamics, setting the stage for future systematic analysis of nonlinearities in subsequent work.

6.4. Discussion

In this study, we introduce a Hybrid-model specifically tailored for forecasting SST and SSH in the tropical Pacific, critical factors in seasonal forecasting worldwide. We start from the LIM, an empirical model describing the dynamics of the slower-varying ocean as stochastically forced by the rapidly varying atmosphere with its deterministic dynamics assumed to be linear. However, while the LIM produces ENSO foreScasts comparable to state-of-the-art numerical models, it is unable to capture observed asymmetries of ENSO that may also be important to its predictability.

We combine an LSTM with the LIM to capture predictable nonlinearities and non-Markovianity in the evolution of monthly tropical SSTA. This Hybrid-model is trained and tested on SSTA and SSHA data from the CESM2 pre-industrial control run, where we observe that modeling nonlinearities significantly enhances the forecast accuracy, particularly in the western tropical Pacific within the 9 to 18-month range.

Our Hybrid-model facilitates disentangling linear from nonlinear dynamics. Our findings provide initial evidence that the asymmetry between warm and cold events is a key source of nonlinearity that improves forecasting skills. This first insight lays the groundwork for a more comprehensive follow-up investigation of the potential sources of nonlinearity of ENSO forecasting.

Moreover, we demonstrate that the predictability of the Hybrid-model is strongly related to the theoretical expected skill of the LIM which allows us to reliably assess its predictability. In contrast, neural networks typically struggle to provide accurate predictability assessments on seasonal to annual scales, primarily hindered by their weak spread-to-skill relationship. While our Hybrid-model shows accurate predictability in the tropical Pacific, this potentially offers predictability for other climate oscillation and variables on sub-seasonal to seasonal scales, a promising avenue for future research.

A notable feature of our Hybrid-model is its relatively modest data requirements for training, particularly when compared to more data-intensive deep learning models like the LSTM network. This aspect is crucial given the limited span of available oceanic observational data. For a fair comparison, we utilized data from global circulation models in our training, acknowledging their inherent biases as discussed in our study. The use of domain adaptation methods from deep learning emerges as a promising strategy to close the gap between Global Circulation Model (GCM) data and observational data. However, the field still needs more research to fully understand how neural network models can be adjusted to observational data when pre-trained on simulated data.

Acknowledgements

The authors acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. Furthermore, we express our gratitude to the NOAA Physical Science Laboratory for

making their resources available for this study. The authors thank the International Max Planck Research School for Intelligent Systems for supporting Jakob Schlör.

Data Availability Statement

All data used in this study are publicly available and are referenced in the main text or the supplementary materials. Our code is available at `https://github.com/jakob-schloer/HybridLIM`.

Part III.

Conclusion & Future Directions

Conclusion & Future Directions 7.

ENSO alters weather and climate conditions worldwide and is therefore the largest source of seasonal predictability. ENSO events vary in their intensity, spatial, and temporal characteristics, necessitating accurate representation of this diversity. However, the mechanisms driving ENSO diversity remain unclear, making it challenging for physics-based models to provide reliable forecasts. The main objective of this thesis was to develop models that capture the diversity of ENSO events from data. This work presents new data-driven approaches for defining ENSO diversity, identifying its global impacts, and forecasting it accurately.

In this final chapter, I summarize the contributions of this thesis and lay out possible future directions.

7.1. Summary & Impact

i. Impacts of ENSO Diversity

Besides the scientific curiosity to understand the dynamical properties that yield to the large diversity of ENSO events, it's their impacts on weather conditions worldwide that are important for society. While teleconnections differ between EP and CP El Niño events, their analysis is usually based on a certain region. Chapter 4 introduces a graphical representation to identify teleconnections globally based on climate networks. The analysis shows that teleconnections of EP are stronger and more constrained to the tropics than CP events. Visualizing global patterns of teleconnections is not just important for identifying unknown teleconnections, it can also be used by climate modelers to check where GCMs are not representing teleconnections of ENSO correctly.

ii. Definition of ENSO Diversity

ENSO and its diversity are generally described by single indices. For instance, ENSO intensity is characterized by averages over a certain region in the Pacific, and its spatial diversity is distinctly separated into EP and CP-type events. In Chapter 5, we argue that these indices and thresholds are a too strong simplification, and propose a more nuanced description of the intertwined intensity and spatial distribution of events. The proposed approach approximates the distribution of ENSO events in a lower dimensional space, assigning membership probabilities to events instead of splitting events into binary types. We show how composites and Hovmoeller diagrams, the favorite tools of climate scientists, can still be used with our approach. Providing the tools to work with the probabilistic definition of ENSO diversity, we want scientists to embrace the continuous stochastic nature of ENSO.

iii. Forecasting ENSO Diversity

The hybrid model proposed in Chapter 6, is the first neural network-based model that produces skillful forecasts for up to 18 months with uncertainty estimates based on ensemble members. The combination of

the LIM with the LSTM allows skillful forecasts in the low-data regime, while also enabling the estimation of the nonlinear dynamics. This hybrid approach further allows distinguishing between predictable and non-predictable cases and identifying windows of opportunity. Modification to the neural network baselines in Chapter 6, such as multiple output heads capturing the forecast uncertainty and conditioning the latent space via affine transformations, seem to improve skill across architectures. Our hybrid model is openly available and might become operational in the PSL-NOAA suite.

7.2. Future Work

Working with computer scientists and meteorologists during my PhD, I believe that deep learning approaches in climate science are still in their infancy. To leverage their full potential, we require easier data availability by having well-documented datasets and more efficient access to data. Moreover, a collaborative environment where climate scientists and computer scientists can work together is needed to ensure that relevant questions are addressed with the best suitable techniques. Below, I outline future research directions that I find promising and interesting:

Large-Scale Phenomena Beyond ENSO

ENSO is the most extensively studied large-scale climate phenomenon, however, a large potential exists for analyzing and forecasting other large-scale atmospheric and oceanic oscillations using similar methodologies to the ones developed in this work. Examples of these include:

- The Madden-Julian Oscillation (MJO) (Madden et al., 1971): a recurring band of rainfall that moves along the equator, significantly influencing monsoon patterns.
- The North Atlantic Oscillation (NAO) (e.g. Stephenson et al., 2003): comprises high and low-pressure systems in the North Atlantic that alter the jet stream, affecting weather patterns, such as blocking events, over Europe.
- The Polar Vortex (Baldwin et al., 1999): involves circulating winds around the North Pole that confine cold air over the polar region. The strength and collapse of the Polar Vortex are critical determinants of cold spells over the Northern Mid-latitudes.
- The Atlantic Niño (Zebiak, 1993): often considered the smaller counterpart of the Pacific El Niño, it impacts rainfall in the Sahel and interacts with other ocean basins.

Causal mechanism of ENSO diversity

The data-driven methods proposed in this thesis focus on identifying structure within the data, thereby representing statistical properties. Statistical properties hint at underlying processes, but they do not establish causation. The approach used to identify teleconnections (Ch. 4) relies on instantaneous correlations and does not reflect causal relationships. Similarly, the GMM used to approximate the distribution of ENSO

events (Ch. 5) does not identify physical attractors, and the LSTM model (Ch. 6), with its hundred-thousand parameters, captures nonlinearities and ENSO asymmetry but lacks interpretability.

A logical next step would be to study the physical mechanisms underlying the findings in this thesis. As a starting point, I would suggest improving the parameterization of air-sea interactions in models (e.g. Liu et al., 2022a). This likely enhances the representation of skewness and spatial distribution of SSTA in models. Alternatively, modifying our hybrid approach could allow distinguishing state-dependent noise from the deterministic dynamics that drive ENSO diversity (Geng et al., 2022). This could be achieved by explicitly learning terms in the stochastic differential equation (SDE) that describes the noise-driven ocean dynamics. However, such Neural SDE models can be challenging to train due to the complexity of the integration steps required during training.

Data-Driven Climate Models

An unexplored research question in this thesis is how ENSO will change in the future. Given that data-driven methods are unable to extrapolate to different climate conditions, physics-based models are a natural choice to study climate scenarios.

However, in the last year, deep learning weather prediction (DLWP) models have achieved forecasting skills on par with numerical weather prediction (NWP) models (Ben-Bouallegue et al., 2023). Observations from large language models suggest that these large models can often generalize beyond their training data envelope (e.g. Kaplan et al., 2020). This raises the question: can sufficiently large neural networks learn the underlying physical processes and generalize to warmer climate conditions? Are these models capable of realistically simulating the climate a century ahead, and can we rely on their projections?

One approach to constructing deep-learning climate models is training them on simulations of future climate scenarios derived from existing GCMs. However, this model would inherit the biases of the GCMs used for training. Although training across multiple GCMs could cancel individual GCM biases, this does not address potential biases shared among them (Wittenberg et al., 2006; Capotondi et al., 2020a; Beverley et al., 2023; Lin et al., 2023).

Alternatively, constructing deep-learning climate models based on observational data alone requires training on high temporal resolution datasets, similar to the DLWP models. The hope is that despite the short observational record, the hourly data is sufficiently detailed to enable the model to learn a physical representation of the system, thus allowing generalization to different climate conditions.

I am excited to see how climate and weather modeling will evolve in the future, leveraging the advancements in deep learning.

Part IV.

Appendix

Additional Material for Chapter 4 A.

A.1. Network Construction

Assume complex network graph G, defined by its set of nodes V, connected by its set of edges E, where e_{ij} denotes an edge connecting node v_i to node v_j. This type of graph is described by its adjacency matrix $\mathbf{A}$:

$$\mathbf{A}_{ij} = \begin{cases} 1, & e_{ij} \in E \\ 0, & \text{otherwise} \end{cases},$$

If the edges e_{ij} are weighted by weights w_{ij}, the weighted graph is described by

$$\mathbf{W}_{ij} = \begin{cases} w_{ij}, & e_{ij} \in E \\ 0, & \text{otherwise} \end{cases},$$

We estimate the weighted adjacency matrix $\mathbf{W}$ of the climate network by placing links between pairs of locations which have a correlation value among the 2% strongest absolute correlations. Thus, for the correlation threshold $\rho_{0.98} = Q_{|\rho|}(0.98)$, where $Q_X(\cdot)$ denotes the quantile function for X, we define

$$\mathbf{W}_{ij} = \begin{cases} |\rho_{ij}|, & |\rho_{ij}| > \rho_{0.98} \\ 0, & \text{otherwise} \end{cases}, \tag{A.1}$$

where $i, j \in \{1, \ldots, N\}$ index spatial locations and ρ_{ij} is the Dec–Feb correlation between locations i and j. $\mathbf{W}_{ij}$ thus defines a network G by a set of edges (or links) $e_{ij} \in E$, where e_{ij} connects the nodes $(v_i, v_j) \in V$ with link weight $w_{ij} = |\rho_{ij}|$. We compute instantaneous correlations using Spearman's rank-order correlation to capture nonlinear behavior between all pairs of time series. For a detailed introduction to climate networks, we refer the reader to Dijkstra et al. (2019).

Accounting for autocorrelation in the data, a two-sided test for non-random correlations at a confidence level of 1 % yields a threshold $|\hat{\rho}| = 0.1$. While our threshold is far higher, due to the high number of hypotheses tests (3.6×10^7), we nevertheless expect a non-negligible number of network links to be false positives. We thus additionally use a spatial null model which assumes that correlations caused by physical mechanisms are likely to occur in bundles of links (Boers et al., 2019).

For each spatial location, we randomly rewire its corresponding network links 2000 times and use a Gaussian kernel density estimator (KDE) to get the likelihood of a link to the chosen location. An observed network link to the chosen location is considered statistically significant if the spatial likelihood of the link (also obtained via Gaussian KDE) is above the 99.9-th percentile of the local null model link distribution. For our SAT climate networks, we found that 2 % of all links were identified as spurious links. In Fig. A.1 we visualized an example of the correction

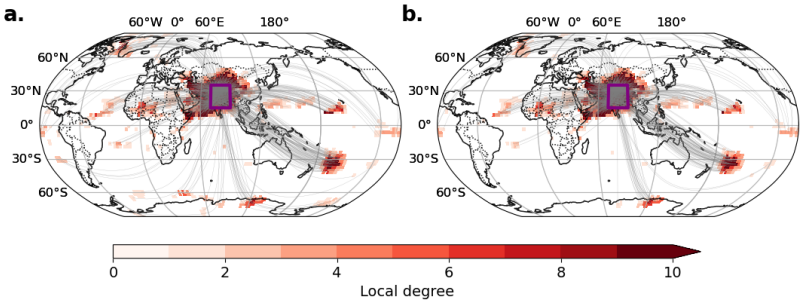


Figure A.1.: Link bundle correction to climate networks. We show a network for CP conditions. The network is further corrected by a Gaussian KDE to avoid spurious links that occur out of random coincidence. We analyze an example for links of the location of the Indian Subcontinent (70°E-90°E, 15°N-35°N, marked by the violet rectangle). **a** shows the uncorrected network, **b** shows the corrected network. For visual reasons, only every 3rd link is plotted.

method by the Gaussian KDE for the location of the Indian subcontinent. In this example, spurious links in the Equatorial and Northern Atlantic Ocean and the Southern Indian Ocean are removed.

A.2. Forman-Ricci Curvature and Betweenness Centrality

Betweenness centrality is a measure of information flow within the network, quantifying the importance of a node based on the number of the shortest paths that pass through it. Betweenness Centrality is defined for nodes v and edges e as,

$$BC_v(v_i) = \sum_{s,t}^{N} \frac{\sigma(v_s, v_t|v_i)}{\sigma(v_s, v_t)} , \tag{A.2}$$

$$BC_e(e_{ij}) = \sum_{s,t}^{N} \frac{\sigma(v_s, v_t|e_{ij})}{\sigma(v_s, v_t)} , \tag{A.3}$$

where $\sigma(v_s, v_t)$ denotes the number of the shortest paths between nodes v_s and v_t and $\sigma(v_s, v_t|v_i) \leq \sigma(v_s, v_t)$ the number of all shortest paths that include node v_i. Similarly, for edge betweenness centrality $\sigma(v_s, v_t|e_{ij}) \leq \sigma(v_s, v_t)$ yields the number of all shortest paths that include edge e_{ij}. BC_n is often referred to as the pathway of a variable through the network and therefore taken as an indicator for the flow of the variable of interest.

Betweenness centrality is typically used to identify edges that connect communities (Freeman, 1977; Donges et al., 2009b; Boers et al., 2013; Ciemer et al., 2018) and is thereby similar to Forman-Ricci curvature. We compare those two measures on a random network with four communities (Fig. A.2) generated using a stochastic block model from NetworkX (Hagberg et al., 2008).

Forman-curvature separates the between-community links from the within-community links. Within-community links are typically part of triangles, indicating local convergence of the shortest paths, i.e. positive curvature. Conversely, links connecting nodes with a high degree that are not part of triangles indicate local divergence of the shortest paths and negative curvature. Forman-curvature provides a continuous measure over network links that indicates if an edge is inside a community or if it straddles two communities (Fig. A.2a). By comparison, the edge betweenness centrality fails to identify many between-community links (Fig. A.2b). This is likely due to the binary notion of the shortest path -

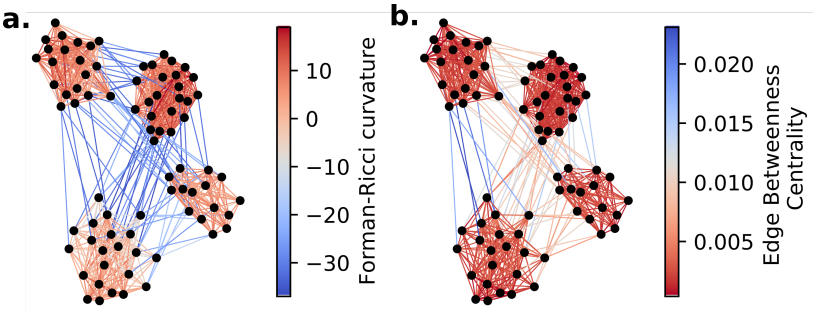


Figure A.2.: Betweenness centrality vs. Forman curvature. In the synthetic network constructed with a Stochastic Block Model, Forman-curvature (**a**) provides values to distinguish within-community links (red) from between-community links (blue) in contrast to betweenness centrality (**b**) with only a few links with high scores. The network with four communities is visualized using the spring layout from NetworkX (Hagberg et al., 2008).

either a path is the shortest or it is not, implying that "almost-shortest" paths are not considered.

A.3. Variability within El Niño Types

We are interested in the center of the anomalies during El Niños. The contours for every year are computed based on the quantiles Q of the seasonal spatial maps for each individual El Niño event. The aggregation of the yearly $Q(0.85)$-DJF contours yields Fig. A.3 suggesting a more extensive spatial diversity of the 8 CP events compared to the 7 EP events.

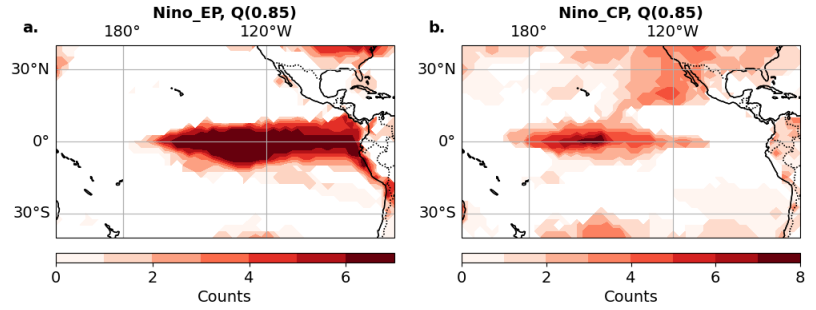


Figure A.3.: Counts of anomaly center during El Niño. We compute composite anomalies for the DJF time series for all years between 1959 and 2020. The year is indicated in which the event mainly occurred. In total 7 events have been identified as EP (a) and 8 events as CP (b) El Niños. The colorbars indicate the number of counts at a spatial location. The contour was chosen according to the $Q(0.9)$ quantile of the spatial map values for each year separately.

A.4. Teleconnections in Climate Models

In order to estimate the error due to the finite amount of EP and CP El Niño events available in the reanalysis data, we repeat our analysis on data from Global Circulation Model (GCM). Model data have the advantage that they can run for a longer period of time and thereby in principle produce an infinite amount of data to compare to. Here, we use the pre-industrial control run of the UKESM1-0-LL model from the CMIP6 which covers 1000 years. We use UKESM1-0-LL because it is known to reproduce similar sea surface temperature variability in the tropical Pacific than found in observations (Dieppois et al., 2021). We again use surface air temperatures on a daily resolution interpolated on an equidistant Fekete grid. Using sea surface temperature data from the same model run, we identify 296 EP El Niño, 168 CP El Niño, and 653 Normal winters in the pre-industrial run.

We compute the correlation-based networks and Forman curvature on the pre-industrial model run. The most positive and most negative

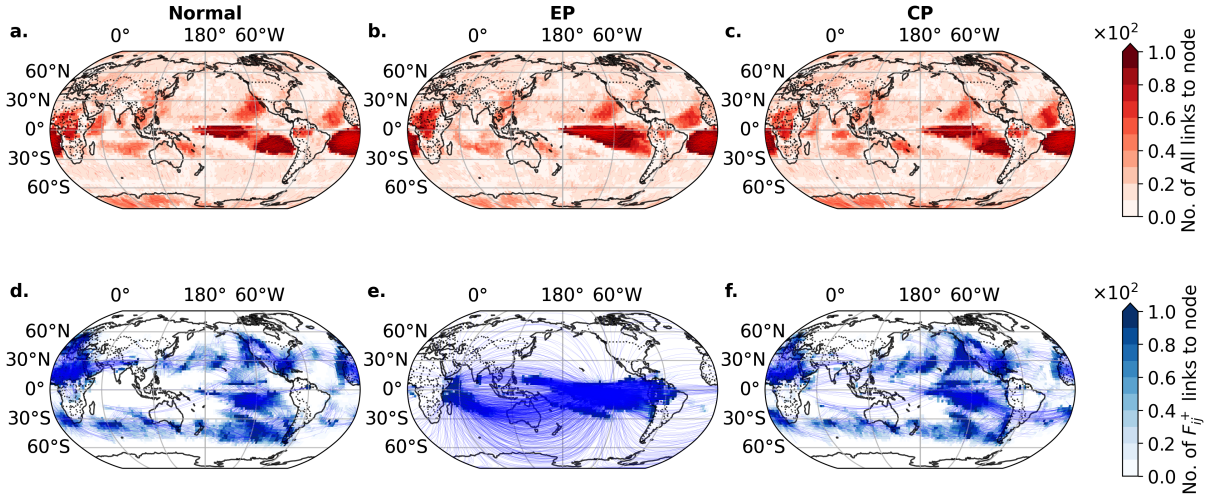


Figure A.4.: Same as Fig. 4.5 but for daily SAT data of the pre-industrial control run of the UKESM1-0-LL model.

curvature hotspots are shown in Fig. A.4. The UKESM1-0-LL model resembles similar teleconnection patterns found in the reanalysis data. Teleconnections are more constrained to the tropical ocean basins under EP conditions while under CP conditions teleconnections are found in the mid-latitudes.

With 296 EP events from the pre-industrial run at hand, we can estimate the error due to the finite amount of events in the observational period. We randomly sample 7 from the 296 available EP events and calculate the curvature. We repeat the sampling procedure 10 times. We compare the obtained distribution of curvature values of the 7 events to the distribution obtained with all 296 EP events (see Fig. A.5 b). The curvature distribution of two samplings are shown in Fig. A.5 c. We find that the distributions cover the same curvature range and show a peak close to zero which is also the case for the curvature distribution obtained from ERA5 (shown in Fig. A.5 a.). This suggests that the UKESM1-0-LL model represents the topology of the SAT reanalysis data well.

However, the curvature distributions obtained from 7 events show higher densities at negative curvatures than the distribution obtained from all EP events. The overestimation of negatively curved links in the network due to the small number of events is also found in the analysis of ERA5 data (see Fig. A.5 a c). We hypothesize that a shorter time period leads to spurious pair-wise correlations which form links over long spatial distances and are thus negatively curved.

This hypothesis is supported by comparing the spatial extent of the curvature obtained by using all 296 EP events from the pre-industrial run (Fig. A.5 d-e) and the average over curvatures obtained from repeated sampling with 7 EP events each (Fig. A.5 f-g). While the most spatial patterns of positively curved edges are similar, we obtain more negatively curved edges to the extra-tropics in the networks constructed from only 7 events. However, pronounced regions of negative curvature are consistent between the analysis using 296 EP events and 7 randomly sampled events. We also find an overestimation of negative curvature using only 7 randomly sampled CP events out of the 168 events in the pre-industrial run (not shown).

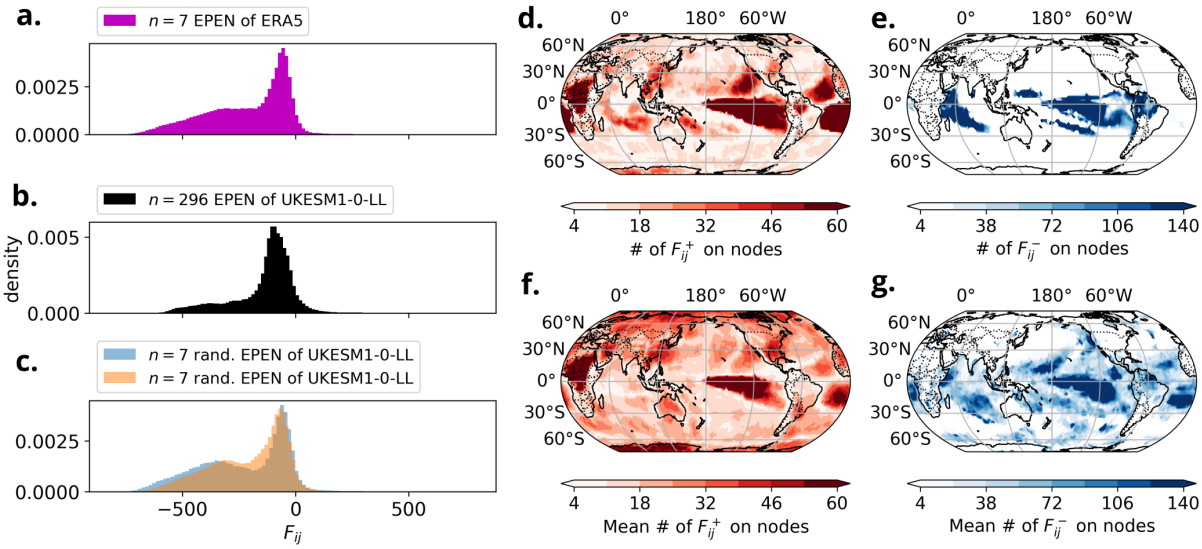


Figure A.5.: Effect of the finite amount of EP El Niño events. The distribution of curvature obtained from all 296 EP El Niño events in the pre-industrial control run of the UKESM1-0-LL model **(b)** is compared to the distributions obtained from 7 randomly sampled EP events from the pre-industrial control run (c) and the distribution obtained from EP events in ERA5 (a). The number of links with positive and negative curvature on each node is shown for the network constructed from 296 EP El Niño events of the pre-industrial control run **(d, e)**. The average number of positively/negatively curved links on each node over the networks constructed from 7 randomly sampled events are displayed in **(f/g)**,

A.5. Local Correlation Analysis

We introduce the average local correlation R_i to better understand the regional dynamics and variability of spatially close points. We define the local correlation as the average correlation of the time series $x_i(t)$ associated with the node v_i to every time series of the node v_b in the set of nodes $\mathcal{B}_i$, where $\mathcal{B}_i$ describes the set of the $B = 200$ spatially closest next-neighbor nodes to v_i:

$$R_i = \frac{1}{B} \sum_{v_b \in \mathcal{B}_i} \rho(x_i(t), x_b(t)) \, . \tag{A.4}$$

Here, ρ describes the Spearman's rank order correlation function between the time series $x_i(t)$ and $x_b(t)$. Note, as we have interpolated the grid to a Fekete Grid, the nodes are approximately uniformly distributed, and therefore no north-south bias due to higher latitudes is included Heitzig et al., 2012. The results are depicted in Fig. A.6.

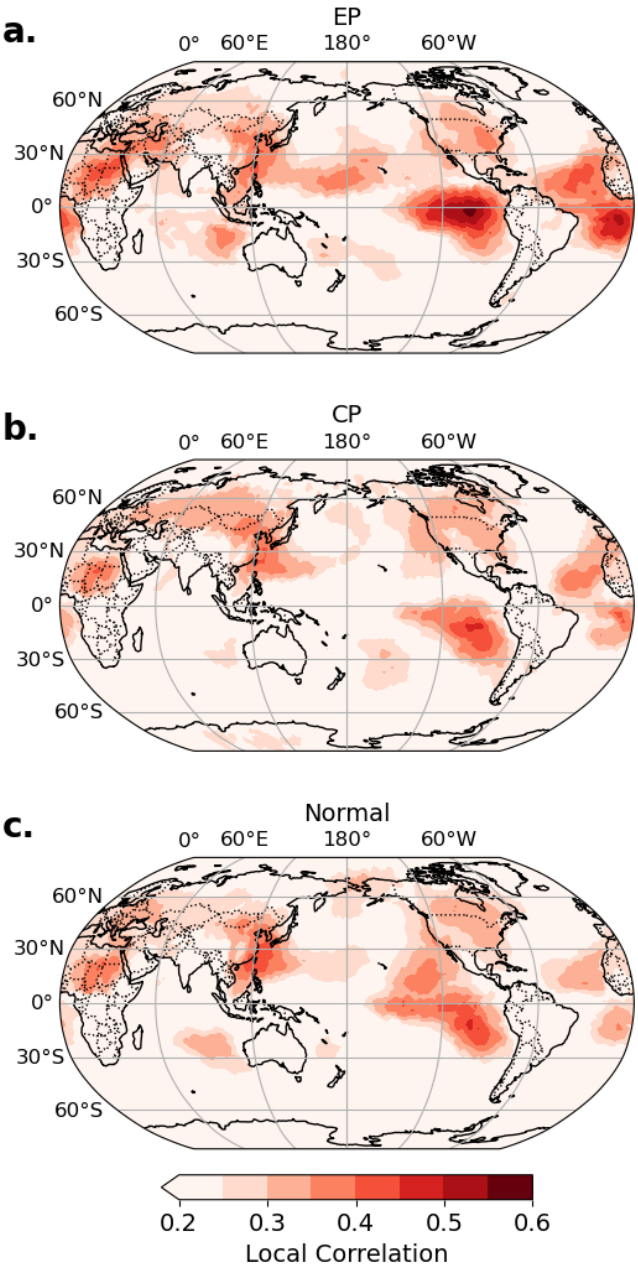


Figure A.6.: Local Correlation Analysis of ENSO conditions. We analyze the local correlation of nodes for the different ENSO conditions. The local correlation (Eq. A.4) takes the $Nb = 200$ next neighbor points and uses Spearman's rank correlation. Displayed are local correlations for (**a**) EP (**b**) CP and (**c**) normal year DJF conditions.

Additional Material for Chapter 5 B.

B.1. Methodological Details

B.1.1. Autoencoder

We denote the data by $X = \{\mathbf{x}(t)\}_{t=1}^{N_t}$, with each time point t is a field $\mathbf{x}(t) \in \mathbb{R}^{N_{\text{lon}} \times N_{\text{lat}}}$. We reduce the dimensionality using a transformation $\mathbf{z}(t) = e\left(\mathbf{x}(t)\right)$ that maps $\mathbb{R}^{N_{\text{lon}} \times N_{\text{lat}}} \rightarrow \mathbb{R}^{M}$ with $M << N_{\text{lon}} \cdot N_{\text{lat}}$. The downscaling function is called an encoder, e. The inverse transformation from latent to data space is described by the function $\hat{\mathbf{x}}(t) = d\left(\mathbf{z}(t)\right)$, called decoder. When reducing the dimensionality, we lose information such that $\hat{\mathbf{x}} \neq \mathbf{x}$, i.e. the entire encoding-decoding pipeline can also be seen as a lossy compression-decompression method.

We use an autoencoder neural network to obtain a non-linear dimensionality reduction (Sec. 2.1.1). Our encoder network consists of four convolutional layers followed by a linear layer. Each convolutional layer halves the longitude and latitude dimension of the input by using a kernel size of 3, a stride of 2, and padding of zero. The output of each convolutional layer is transformed by a ReLU function. We double the number of channels in the first two layers and keep it fixed for the last two layers. After the last convolutional layer, we flatten the output and pass it through a linear layer whose output has the dimensionality M. The decoder is designed to be symmetric to the encoder. A linear layer is followed by four deconvolutional layers with each layer doubling the dimension of x and y. The number of channels is fixed in the first two convolutional layers but halve in the last two layers respectively. This way the output dimensionality matches the dimension of the encoder input. Again we use the ReLU activation function after each layer except the last one. Different variables (SSTA, SSHA) are stacked in the channel dimension.

B.1.2. Warm-Pool Edge

Thual et al. (2023) propose that the diversity in ENSO is linked to the zonal shift of the Walker circulation over the tropical Pacific. They suggest that the edge of the warm-pool serves as an indicator for deep convection shifts, characterizing it as the 29°C isotherm where SSTs saturate over the warm pool. The warm-pool edge is identified as the earliest longitude moving from 80°W (East) to 120°E (West), where the equatorial SST exceeds 29°C for a subsequent 10° westward.

B.1.3. Testing for Statistical Significance of Averages

We utilize a bootstrapping approach to test the hypothesis that the means associated with ENSO categories are statistically distinct from the ENSO

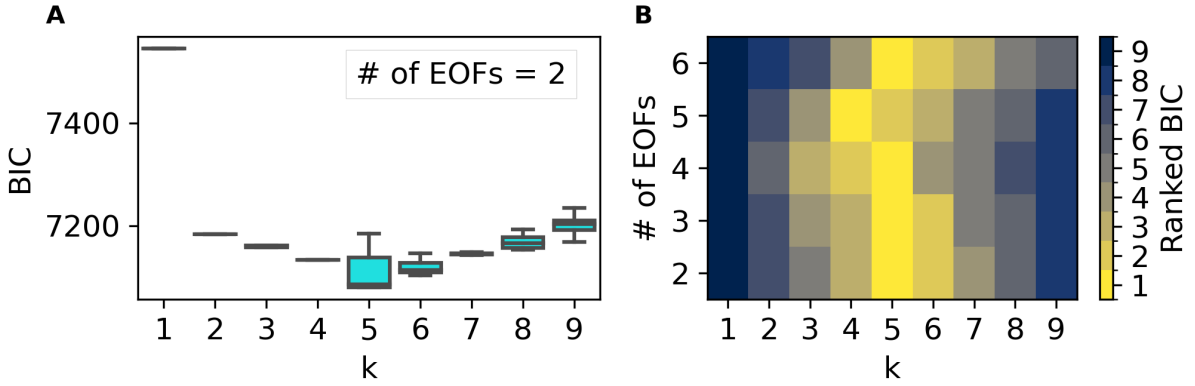


Figure B.1.: Determining the optimal number of clusters. The distribution of BIC is computed for each number of clusters k by fitting 100 GMMs which are randomly initialized. The mean, 10th, and 90th percentile as well as the median are shown as boxplots for the two-dimensional EOF latent representation of SSTA **(A)**. The same analysis is repeated by increasing the number of EOFs **(B)**. The mean BIC over the 100 GMMs for each k is ranked for each number of EOF to show that our results are robust with the number of EOFs used.

Neutral conditions, i.e. the DJF Niño3.4-index is between -0.5 and 0.5. This technique is used for both weighted and unweighted averages.

Statistical significance of the mean $\underline{\mathbf{y}}_k$ over N data points is obtained by generating a null distribution through resampling. We draw N data points with replacement from *Neutral* conditions and calculate the mean of these samples. The process is repeated 1000 times to produce a distribution of the random means. When considering weighted averages, the resampling procedure is adapted by using the shuffled weights as probabilities of the random sampling across the Neutral years. Significance is then quantified by the percentile of the mean against the null distribution which yields the p-values.

Given that our statistical evaluation spans multiple grid points, one has to adjust the p-values to counteract the risk of type-1 errors from multiple comparisons, i.e. false positives arising from repeated testing (Wilks, 2019). We account for multiple comparisons using the test procedure by Holm (1979) combined with the Sidak (1967) correction, ensuring the p-values are corrected appropriately. Throughout this work, statistically significant values are those exceeding the $\alpha = 1 - p \geq 95\%$ confidence threshold.

B.2. Robustness Analysis

To assess the robustness of our approach, we conduct a sensitivity analysis by varying the number of EOFs. We performed the analysis for 2-6 EOFs and visualized the BIC as a function of categories in Fig. B.1B. Since the BIC values exhibit changes with an increasing number of EOFs, we sort the mean BIC values for each number of EOFs and plot their ranks. Our results consistently indicated that the 5-dimensional latent space yielded the lowest BIC values, except in the case of the 4-dimensional latent space, where the minimum BIC was observed at $k = 6$. It is worth noting that the BIC values displayed a wider distribution at $k = 6$ compared to $k = 5$ in the 4-dimensional latent space which explains the deviation (not shown).

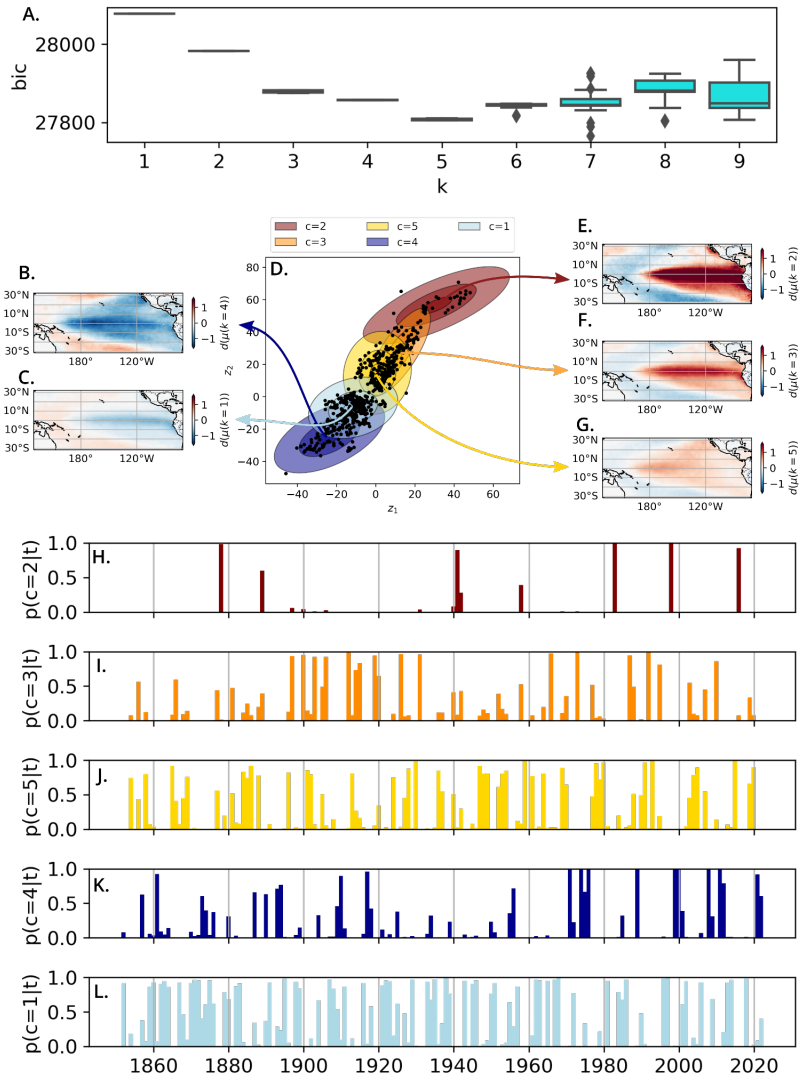


Figure B.2.: GMM applied in latent space of Autoencoder. Boreal winter SSTA of El Niño and La Niña events are projected on the two-dimensional latent space of an autoencoder (black points in D). The nonlinear encoder-decoder neural networks are trained to minimize the reconstruction loss of monthly SSTA across all reanalysis datasets, details are described in Sec. B.1.1. Applying the GMM in this reduced space yields five categories (A), minimizing the BIC. The decoded Gaussian means (B, C, E, F, G) and their category probabilities (H, I, J, K, L) align closely with the results in Fig. 5.3 A and Fig. 5.4.

To examine the potential bias introduced by linear dimensionality reduction in our approach, we adopt an alternative method using a convolutional autoencoder. Unlike the EOF-truncation, this approach allowed for nonlinear dimensionality reduction of the SSTA fields. Here, we employ a symmetric autoencoder as described in Sec. B.1.1. Subsequently, the GMM was fitted based on the low-dimensional representation obtained from the autoencoder. The encodings of the optimal GMM means (Fig. B.2B-G) exhibited similar patterns to those obtained from the EOF analysis. These results imply that a linear decomposition of the SSTA fields suffices to capture their diversity, minimizing concerns regarding bias introduced by the linear dimensionality reduction.

B.3. Comparison to Conventional ENSO Categorization Approaches

In order to compare our category membership with conventional index-based classifications of El Niño events, we present a comparison between the Niño3-Niño4 classification by Kug et al. (2009), the EOF-based E/C

Table B.1.: Classification of El Niño events. Listed are two conventional used indices-based classification approaches of El Niño events as well as our probabilistic clustering. We use E to denote Eastern Pacific, C for Central Pacific type events, W for our *Weak EN*, S for our *Strong EN*, and E* for our *Extreme EN* category. The Niño3-Niño4 approach by Kug et al. (2009) classifies events as E when the DJF average Niño3 > Niño4 and C when Niño4 < Niño3. Takahashi et al. (2011) use the first two PCs of SSTA to create an E- and C-index. When DJF average E-index > C-index events are considered of type E and events are of type C when C-index > E-index. The last column shows the largest category membership and their probability in brackets, $p\,(c|x)$. In cases where the probabilities are shared (i.e. $0.4 < p\,(c|x) < 0.6$), we show the largest two category memberships. The blank entries denote cases in which the respective conditions were not met to classify an event.

	Niño3-Niño4	**E/C**	$p(c_k\|x)$
1951-1952	E	E	E (0.87)
1953-1954	C		C (0.45) / E (0.33)
1957-1958	E	C	C (0.54) / E (0.44)
1958-1959	C		C (0.97)
1963-1964	E	C	C (0.80)
1965-1966	E	C	E (0.81)
1968-1969	C	C	C (0.97)
1969-1970	C		E (0.60) / C (0.37)
1972-1973	E	E	E (0.98)
1976-1977	E		E (0.92)
1977-1978	C		C (0.84)
1979-1980	C		C (0.89)
1982-1983	E	E	**E*(1.00)**
1986-1987	E	E	E (0.88)
1987-1988	C	C	E (0.52) / C (0.47)
1991-1992	E	C	E (0.99)
1994-1995	C	C	E (0.89)
1997-1998	E	E	**E*(1.00)**
2002-2003	C	C	C (0.51) / E (0.49)
2004-2005	C	C	C (0.77)
2006-2007	C	C	E (0.93)
2009-2010	C	C	C (0.58) / E (0.42)
2014-2015	C		C (0.98)
2015-2016	E	C	**E*(0.83)**
2018-2019	C	C	E (0.59) / C (0.40)
2019-2020	C	C	C (0.79)

classification by Takahashi et al. (2011), and our category membership (Tab. B.1).

The events 1983/84, 1997/98, and 2015/16 have a high membership probability in the *Extreme EN* category, while they are classified as EP-type events in the conventional methods. The 1951/52, 1972/73, 1976/77, and 1986/87 events show a high likelihood of being part of the *Strong EN category* which concurs with conventional classifications. We identified the events of 1958/59, 1968/69, 1969/70, 1977/78, 1979/80, 2004/05, 2014/15, and 2019/20 as having a high probability of belonging to the *Weak EN category*, also aligning with conventional classification.

We observe inconsistencies in indices-based classifications for the 1994/95 and 2006/07 events, displaying high membership probability for the Strong EN category, despite being labeled as CP-type by Niño3-Niño4 and E/C definitions. The conventional EP- and CP-type definitions, however, lack consistency as summarized by Pascolini-Campbell et al. (2015), Yu et al. (2010), and Capotondi et al. (2020a).

B.4. ENSO diversity in CESM2

Given the limited number of extreme El Niño events in the observational record, we have extended our analysis to the Coupled Earth System Model 2 (CESM2), chosen because of its large-ensemble simulations and the 2000-year pre-industrial control simulation. Monthly tropical

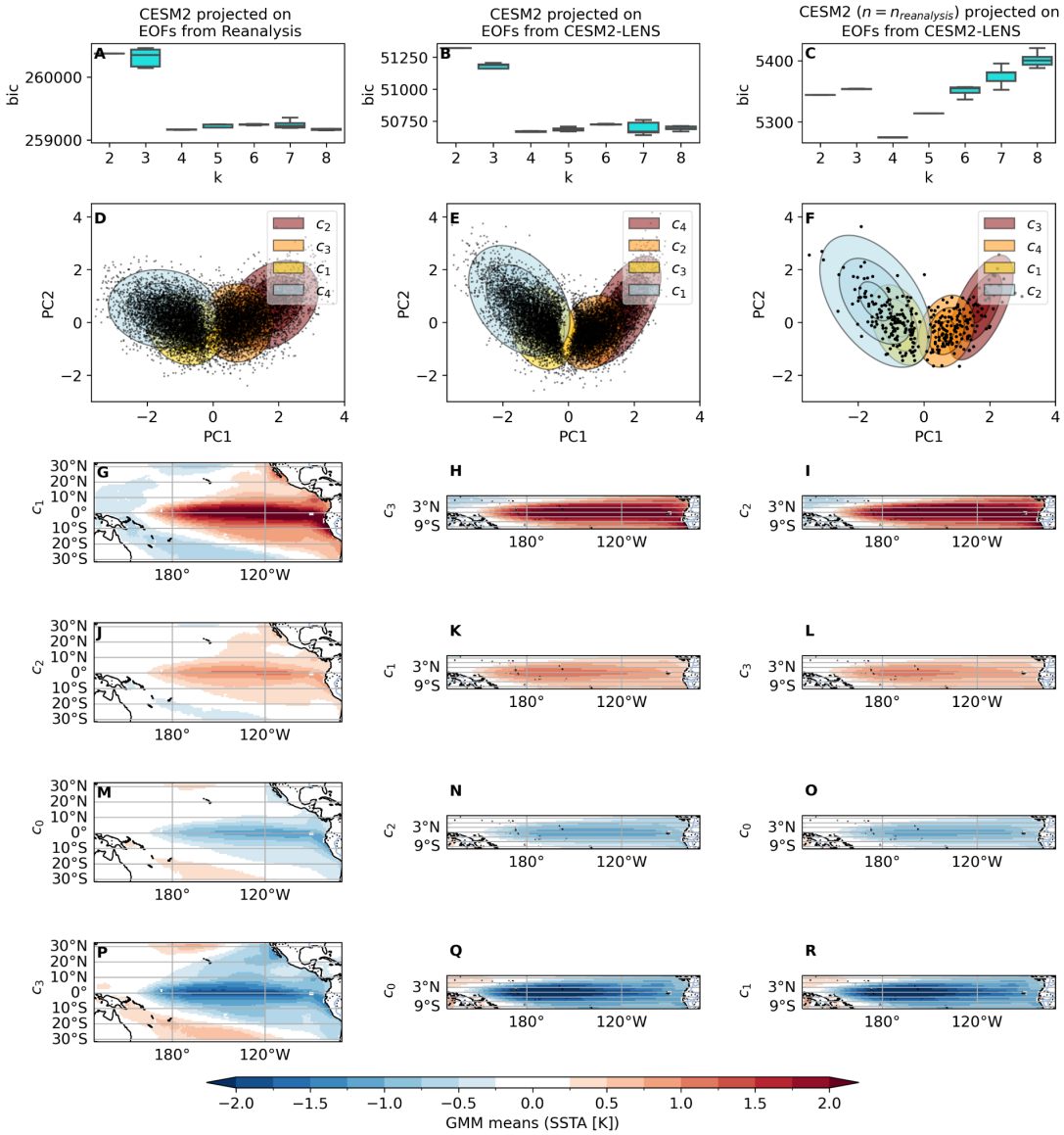


Figure B.3.: GMM analysis on model data from CESM2-LENS historical simulations. We project monthly SSTA during El Niño and La Niña winters (DJF) of all CESM2-LENS historical runs on the first two EOFs from the joined reanalysis dataset (black points in **D**) and from CESM2 historical runs selecting only the equatorial domain (10°S to 10°N) (black points in **E**). For comparison to the shorter observational period, we fit the GMM on a randomly selected subset of EN and LN winters of CESM2-LENS data such that the number of events matches the ones in the joined reanalysis data **(F)**. The optimal number of categories of the GMM is based on the smallest BIC using 100 randomly initialized GMMs for each number of clusters k. The mean, 10th and 90th percentile as well as the median are shown as boxplots (**A, B, C**). In contrast to the five clusters obtained from the reanalysis products, the mean BIC is minimized for k=4 categories using the CESM2-LENS. The decoded means of each Gaussian is projected on EOF1 and EOF2 resemble the typical Strong El Niño (G, H, I), Weak El Niño (J, K, L), Weak La Niña (M, N, O) and Strong La Niñaz (P, Q, R) categories.

Pacific SSTA of CESM2 are selected analogously to the reanalysis data. However, our EOF analysis uses SSTA which is confined within a narrower latitudinal range of 10°S to 10°N, as proposed by (Takahashi et al., 2016). This modification is due to the fact that the second EOF from CESM2 over the full latitudinal range diverges in pattern from EOF2 from reanalysis. Equivalently to the analysis on reanalysis data, we select the boreal winters of El Niño and La Niña, project them onto the first two EOFs which are then utilized for fitting the GMM (Sec. 5.2.1).

We employ the GMM approach on the SSTA from CESM2's historical simulations of boreal winter El Niño and La Niña events, projected

onto the first two EOF modes. By minimizing the BIC, we obtain solely four categories, separating El Niño and La Niña into strong and weak categories, respectively (Fig. B.3). These results are consistent when projecting the CESM2 events on EOF1 and EOF2 from reanalysis or using the 2000-year pre-industrial control simulation of CESM2. Additionally, we repeated the analysis by randomly selecting the same number of events from the CESM2 historical simulations as in the reanalysis dataset, which still resulted in only four clusters (Fig. B.3F). This result suggests that the cluster formation in reanalysis data is not an artifact of the limited number of samples. Instead, we hypothesize that CESM2 may not fully capture the distribution of El Niño and La Niña events as observed in reanalysis data. Moreover, the model's symmetric event distribution and categorization suggest that Extreme El Niños are not represented in CESM2.

The reasons for the model's inability to capture the *Extreme EN* category remain speculative, but it may be related to model biases in upper-ocean processes which impact the evolution and intensity of extreme ENSO events in the model (Wei et al., 2021). Other factors, including influences from the extratropical Pacific (Chiang et al., 2004; Zhang et al., 2014) or from other oceans (Cai et al., 2019), could affect the evolution and distribution of El Niño and La Niña events in CESM2 compared to observational data.

Given that CESM2 fails to simulate extreme El Niño events limits our ability to rigorously test the robustness of the results on the reanalysis data. Despite this limitation, we have thoroughly tested for statistical significance using a bootstrapping method (Sec. B.1.3). Our approach yields robust statistical evidence even with the short observational record.

Additional Material for Chapter 6 C.

C.1. LSTM Architecture

The LSTM and SwinLSTM neural network have the same encoder-decoder structure (Fig. C.1). The encoder network starts with a downsampling operation, which is represented by the EOF-truncation for the PC-LSTM and a strided convolution for the SwinLSTM. After downsampling, the input is linearly projected onto a higher-dimensional latent representation. This is sequentially processed through two LSTM (SwinLSTM) layers, integrating the transformed input with the preceding hidden state. The resulting hidden state at time t is passed to the decoder network. Reflecting the encoder's design, the decoder is composed of two LSTM (SwinLSTM) layers, extending over the future prediction span $t + T$, transferring the hidden state across time without further input. A final upsampling phase, using individual upsampling layers for each m-ensemble, members reverts the hidden state to the original input space. Monthly conditioning is incorporated into both encoder and decoder LSTM layers via affine transformations (Eq. 6.10) of the month embedding.

Inspired by the popular Swin-Transformer architecture (Liu et al., 2021), we modify the classical ConvLSTM cell (Fig. C.2B) to separate spatial and channel mixing into two convolutions. The convolutions are separated by a group normalization layer and a FiLM-layer (Eq. 6.10).

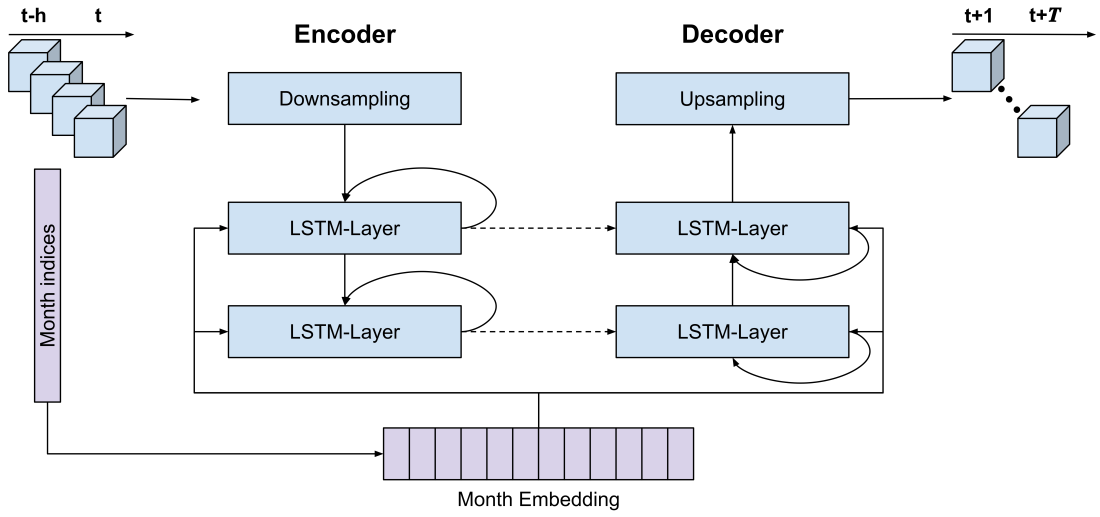


Figure C.1.: Encoder-Decoder architecture of the LSTM and SwinLSTM. Both the LSTM and SwinLSTM neural networks have the same encoder-decoder architecture. The encoder network downsamples the input into a hidden state, which is then sequentially processed through two LSTM-(SwinLSTM)-layers. The hidden state is passed to the decoder network which is rolled out to some future hidden state. Finally, the hidden state is projected back to the input space using an upsampling layer. A affine transformation is used to include monthly information into the LSTM-layers.

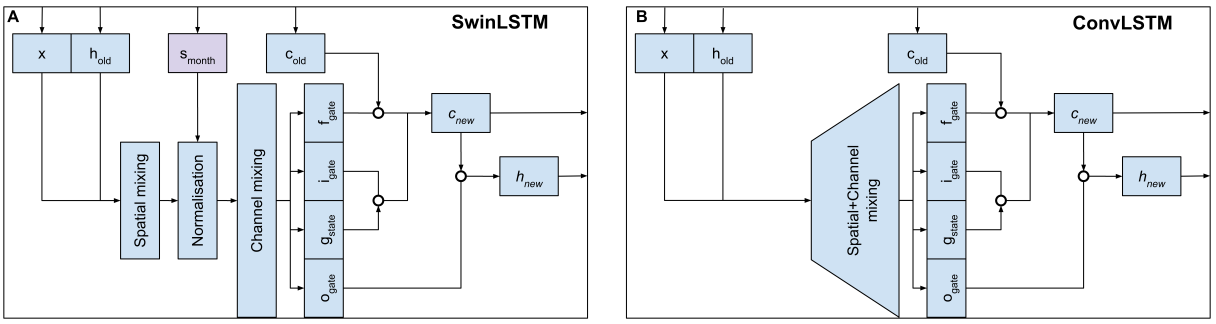


Figure C.2.: Schematic representation of the SwinLSTM and the ConvLSTM cell. Input to our adapted ConvLSTM cell, the SwinLSTM (A), and the ConvLSTM cell by Shi et al. (2015) (B) is the concatenated latent state x and hidden state from the previous time step h_{old}. While in the ConvLSTM cell spatial and channel mixing is performed at once, in the SwinLSTM, we separate spatial and channel mixing into two convolutions, which drastically reduces the number of parameters. We further apply a layer normalization and conditioning on the monthly embedding (Eq. 6.10) in between the spatial and channel mixing. The remainder of the cells are equivalent to the standard LSTM.

C.2. Seasonal Skill Dependency

In addition to assessing the spatial distribution of skill (Sec. 6.3.1), we investigate the seasonal skill variation. The average RMSESS of the Niño4 SSTA from the CS-LIM forecast, evaluated over different lag times and verification months, indicates that late winter and spring months are better predicted than the late summer and fall (Fig. C.3a). These results are consistent with the findings by Shin et al. (2021), and align with the phenomenon known as the spring predictability barrier, characterized by a notable drop in the autocorrelation of the tropical Pacific SSTA in boreal spring.

For the Hybrid-model forecast, we observe the most significant RMSESS improvements upon the CS-LIM at lag times ranging between 9 and 18 months (Fig. C.3B). Specifically, the maximum enhancements are seen at 15 and 18 months during the winter months (December to February), while in spring, the peak skill improvement is at a lag time of 12 months.

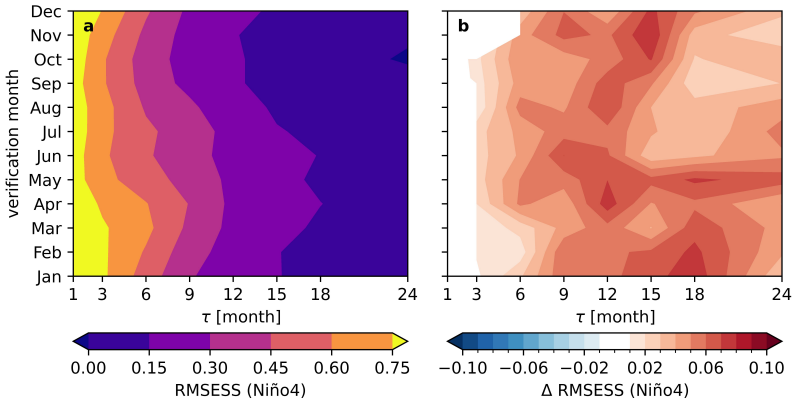


Figure C.3.: Seasonal skill dependency of Niño4. The average RMSE skill score of the CS-LIM forecast for Niño4 SSTA is analyzed over various lag times and verification months (a). The Hybrid-model forecast significantly improves relative to the CS-LIM, particularly for lag times between 9 and 18 months (b). In both panels, statistical significance is determined using a two-sided t-test on 1000 bootstrapped means of CS-LIM and Hybrid-model forecasts. Results displayed exceed the 95% confidence interval.

Bibliography

Here are the references in alphabetic order.

Abdelkader Di Carlo, I. et al. (2023). 'Different Methods in Assessing El Niño Flavors Lead to Opposite Results'. In: *Geophysical Research Letters* 50.15, e2023GL104558. DOI: 10.1029/2023GL104558.

Alexander, M. A. et al. (2002). 'The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air–Sea Interaction over the Global Oceans'. In: *Journal of Climate* 15.16, pages 2205–2231. DOI: 10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

Amaya, D. J. (2019). 'The Pacific Meridional Mode and ENSO: A Review'. In: *Curr Clim Change Rep* 5.4, pages 296–307. DOI: 10.1007/s40641-019-00142-x.

Arakawa, A. (2004). 'The Cumulus Parameterization Problem: Past, Present, and Future'. In: *Journal of Climate* 17.13, pages 2493–2525. DOI: 10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2.

Ashok, K. et al. (2007). 'El Niño Modoki and Its Possible Teleconnection'. In: *Journal of Geophysical Research: Oceans* 112.11, page C11007. DOI: 10.1029/2006JC003798.

Ayar, P. V. et al. (2023). 'A Regime View of ENSO Flavors Through Clustering in CMIP6 Models'. In: *Earth's Future* 11.11, e2022EF003460. DOI: 10.1029/2022EF003460.

Baldwin, M. P. and T. J. Dunkerton (1999). 'Propagation of the Arctic Oscillation from the Stratosphere to the Troposphere'. In: *Journal of Geophysical Research: Atmospheres* 104.D24, pages 30937–30946. DOI: 10.1029/1999JD900445.

Barnett, T. et al. (1988). 'On the Prediction of the El Niño of 1986-1987'. In: *Science* 241.4862, pages 192–196. DOI: 10.1126/science.241.4862.192.

Barnston, A. G. and R. E. Livezey (1989). 'An Operational Multifield Analog / Anti-Analog Prediction System for United States Seasonal Temperatures. Part II: Spring, Summer, Fall, and Intermediate 3-Month Period Experiments'. In: *Journal of Climate* 2.6, pages 513–541. JSTOR: 26194039.

Barnston, A. G. et al. (2012). 'Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing?' In: *Bulletin of the American Meteorological Society* 93.5, pages 631–651. DOI: 10.1175/BAMS-D-11-00111.1.

Barsugli, J. J. and P. D. Sardeshmukh (2002). 'Global Atmospheric Sensitivity to Tropical SST Anomalies throughout the Indo-Pacific Basin'. In: *Journal of Climate* 15.23, pages 3427–3442. DOI: 10.1175/1520-0442(2002)015<3427:GASTTS>2.0.CO;2.

Battaglia, P. W. et al. (2018). *Relational Inductive Biases, Deep Learning, and Graph Networks*. DOI: 10.48550/arXiv.1806.01261. arXiv: 1806.01261 [cs, stat].

Bauer, P., A. Thorpe, and G. Brunet (2015). 'The Quiet Revolution of Numerical Weather Prediction'. In: *Nature* 525.7567, pages 47–55. DOI: 10.1038/nature14956.

Behringer, D. W., M. Ji, and A. Leetmaa (1998). 'An Improved Coupled Model for ENSO Prediction and Implications for Ocean Initialization. Part I: The Ocean Data Assimilation System'. In: *Mon. Weather Rev.* 126.4, pages 1013–1021. DOI: 10.1175/1520-0493(1998)126<1013:AICMFE>2.0.CO;2.

Bejarano, L. and F.-F. Jin (2008). 'Coexistence of Equatorial Coupled Modes of ENSO'. In: *Journal of Climate* 21.12, pages 3051–3067. DOI: 10.1175/2007JCLI1679.1.

Ben-Bouallegue, Z. et al. (2023). *The Rise of Data-Driven Weather Forecasting*. DOI: 10.48550/arXiv.2307.10128. arXiv: 2307.10128 [physics].

Bendito, E. et al. (2007). 'Estimation of Fekete Points'. In: *Journal of Computational Physics* 225.2, pages 2354–2376. DOI: 10.1016/j.jcp.2007.03.017.

Beniche, M. (2023). *A Distinct and Reproducible Teleconnection Pattern over North America during Extreme El Niño Events*. https://www.researchsquare.com. DOI: 10.21203/rs.3.rs-3377657/v1.

Beverley, J. D., M. Newman, and A. Hoell (2023). 'Rapid Development of Systematic ENSO-Related Seasonal Forecast Errors'. In: *Geophysical Research Letters* 50.10, e2022GL102249. DOI: 10.1029/2022GL102249.

Bi, K. et al. (2023). 'Accurate Medium-Range Global Weather Forecasting with 3D Neural Networks'. In: *Nature* 619.7970, pages 533–538. DOI: 10.1038/s41586-023-06185-3.

Bianucci, M. et al. (2018). 'Estimate of the Average Timing for Strong El Niño Events Using the Recharge Oscillator Model with a Multiplicative Perturbation'. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.10, page 103118. DOI: 10.1063/1.5030413.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bjerknes, J. (1966). 'A Possible Response of the Atmospheric Hadley Circulation to Equatorial Anomalies of Ocean Temperature'. In: *Tellus* 18.4, pages 820–829. DOI: 10.1111/j.2153-3490.1966.tb00303.x.

Blake, F. (2023). *Hurricane OTIS*. https://www.nhc.noaa.gov/archive/2023/ep18/ep182023.discus.012.shtml.

Boers, N. et al. (2013). 'Complex Networks Identify Spatial Patterns of Extreme Rainfall Events of the South American Monsoon System'. In: *Geophysical Research Letters* 40.16, pages 4386–4392. DOI: 10.1002/grl.50681.

Boers, N. et al. (2019). 'Complex Networks Reveal Global Pattern of Extreme-Rainfall Teleconnections'. In: *Nature* 566.7744, pages 373–377. DOI: 10.1038/s41586-018-0872-x.

Brenowitz, N. D. and C. S. Bretherton (2018). 'Prognostic Validation of a Neural Network Unified Physics Parameterization'. In: *Geophysical Research Letters* 45.12, pages 6289–6298. DOI: 10.1029/2018GL078510.

Cachay, S. R. et al. (2021). *The World as a Graph: Improving El Ni\~no Forecasts with Graph Neural Networks*. arXiv: 2104.05089 [physics, stat].

Cai, W. et al. (2015). 'Increased Frequency of Extreme La Niña Events under Greenhouse Warming'. In: *Nature Clim Change* 5.2, pages 132–137. DOI: 10.1038/nclimate2492.

Cai, W. et al. (2018). 'Increased Variability of Eastern Pacific El Niño under Greenhouse Warming'. In: *Nature* 564.7735, pages 201–206. DOI: 10.1038/s41586-018-0776-9.

Cai, W. et al. (2019). 'Pantropical Climate Interactions'. In: *Science* 363.6430, eaav4236. DOI: 10.1126/science.aav4236.

Cai, W. et al. (2021). 'Changing El Niño–Southern Oscillation in a Warming Climate'. In: *Nat Rev Earth Environ* 2.9, pages 628–644. DOI: 10.1038/s43017-021-00199-z.

Cai, W. et al. (2023). 'Anthropogenic Impacts on Twentieth-Century ENSO Variability Changes'. In: *Nat Rev Earth Environ* 4.6, pages 407–418. DOI: 10.1038/s43017-023-00427-8.

Callahan, C. and J. S. Mankin (2023). *Persistent Effect of El Niño on Global Economic Growth | Science*. https://www.science.org/doi/10.1126/science.adf2983.

Cane, M. A., S. E. Zebiak, and S. C. Dolan (1986). 'Experimental Forecasts of El Niño'. In: *Nature* 321.6073, pages 827–832. DOI: 10.1038/321827a0.

Capotondi, A. et al. (2020a). 'ENSO and Pacific Decadal Variability in the Community Earth System Model Version 2'. In: *Journal of Advances in Modeling Earth Systems* 12.12, e2019MS002022. DOI: 10.1029/2019MS002022.

Capotondi, A. (2013). 'ENSO Diversity in the NCAR CCSM4 Climate Model: Enso Diversity in the NCAR CCSM4'. In: *J. Geophys. Res. Oceans* 118.10, pages 4755–4770. DOI: 10.1002/jgrc.20335.

Capotondi, A. and L. Ricciardulli (2021). 'The Influence of Pacific Winds on ENSO Diversity'. In: *Sci Rep* 11.1, page 18672. DOI: 10.1038/s41598-021-97963-4.

Capotondi, A. and P. D. Sardeshmukh (2017). 'Is El Niño Really Changing?' In: *Geophysical Research Letters* 44.16, pages 8548–8556. DOI: 10.1002/2017GL074515.

Capotondi, A., P. D. Sardeshmukh, and L. Ricciardulli (2018). 'The Nature of the Stochastic Wind Forcing of ENSO'. In: *J. Climate* 31.19, pages 8081–8099. DOI: 10.1175/JCLI-D-17-0842.1.

Capotondi, A., A. Wittenberg, and S. Masina (2006). 'Spatial and Temporal Structure of Tropical Pacific Interannual Variability in 20th Century Coupled Simulations'. In: *Ocean Modelling*. Oceanic Results from a New Generation of Coupled Climate Models 15.3, pages 274–298. DOI: 10.1016/j.ocemod.2006.02.004.

Capotondi, A. et al. (2015). 'Understanding Enso Diversity'. In: *Bulletin of the American Meteorological Society* 96.6, pages 921–938. DOI: 10.1175/BAMS-D-13-00117.1.

Capotondi, A. et al. (2019). 'Predictability of US West Coast Ocean Temperatures Is Not Solely Due to ENSO'. In: *Sci Rep* 9.1, page 10993. DOI: 10.1038/s41598-019-47400-4.

Capotondi, A. et al. (2020b). 'ENSO Diversity'. In: *El Niño Southern Oscillation in a Changing Climate*. American Geophysical Union (AGU). Chapter 4, pages 65–86. DOI: 10.1002/9781119548164.ch4.

Capotondi, A. et al. (2023). 'Mechanisms of Tropical Pacific Decadal Variability'. In: *Nat Rev Earth Environ* 4.11, pages 754–769. DOI: 10.1038/s43017-023-00486-x.

Carillo, C. N. (1892). 'Desertacion Sobre Las Corrientes y Estudios de La Corriente Peruana de Humboldt'. In: *Boletín de La Sociedad Geográfica de Lima*. Volume t.2 (1892-1893). 11. Lima: Sociedad Geográfica de Lima, pages 72–110.

Chang, P. et al. (2006). 'The Cause of the Fragile Relationship between the Pacific El Niño and the Atlantic Niño'. In: *Nature* 443.7109, pages 324–328. DOI: 10.1038/nature05053.

Chapman, D. et al. (2015). 'A Vector Autoregressive ENSO Prediction Model'. In: *Journal of Climate* 28.21, pages 8511–8520. DOI: 10.1175/JCLI-D-15-0306.1.

Chen, C. et al. (2016). 'Diversity, Nonlinearity, Seasonality, and Memory Effect in ENSO Simulation and Prediction Using Empirical Model Reduction'. In: *Journal of Climate* 29.5, pages 1809–1830. DOI: 10.1175/JCLI-D-15-0372.1.

Chen, D. et al. (2015). 'Strong Influence of Westerly Wind Bursts on El Niño Diversity'. In: *Nature Geosci* 8.5, pages 339–345. DOI: 10.1038/ngeo2399.

Chen, H. et al. (2021). 'Warm and Cold Episodes in Western Pacific Warm Pool and Their Linkage With ENSO Asymmetry and Diversity'. In: *Journal of Geophysical Research: Oceans* 126.12, e2021JC017287. DOI: 10.1029/2021JC017287.

Chiang, J. C. H. and D. J. Vimont (2004). 'Analogous Pacific and Atlantic Meridional Modes of Tropical Atmosphere–Ocean Variability'. In: *Journal of Climate* 17.21, pages 4143–4158. DOI: 10.1175/JCLI4953.1.

Cho, K. et al. (2014). *Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation*. DOI: 10.48550/arXiv.1406.1078. arXiv: 1406.1078 [cs, stat].

Chukkapalli, G., S. R. Karpik, and C. R. Ethier (1999). 'A Scheme for Generating Unstructured Grids on Spheres with Application to Parallel Computation'. In: *Journal of Computational Physics* 149.1, pages 114–127. DOI: 10.1006/jcph.1998.6146.

Ciemer, C. et al. (2018). 'Temporal Evolution of the Spatial Covariability of Rainfall in South America'. In: *Clim Dyn* 51.1, pages 371–382. DOI: 10.1007/s00382-017-3929-x.

COBE (2006). *COBE Sea Surface Temperature: NOAA Physical Sciences Laboratory COBE Sea Surface Temperature*. https://psl.noaa.gov/data/gridded/data.cobe.html.

Copernicus (2024). *Copernicus: 2023 Is the Hottest Year on Record, with Global Temperatures Close to the 1.5°C Limit | Copernicus*. https://climate.copernicus.eu/copernicus-2023-hottest-year-record.

Danabasoglu, G. et al. (2020). 'The Community Earth System Model Version 2 (CESM2)'. In: *Journal of Advances in Modeling Earth Systems* 12.2, e2019MS001916. DOI: 10.1029/2019MS001916.

Dieppois, B. et al. (2021). 'ENSO Diversity Shows Robust Decadal Variations That Must Be Captured for Accurate Future Projections'. In: *Commun Earth Environ* 2.1, page 212. DOI: 10.1038/s43247-021-00285-6.

Dijkstra, H. A. et al. (2019). *Networks in Climate*. Cambridge: Cambridge University Press.

DiNezio, P. N. et al. (2017). 'A 2 Year Forecast for a 60–80% Chance of La Niña in 2017–2018'. In: *Geophysical Research Letters* 44.22, pages 11,624–11,635. DOI: 10.1002/2017GL074904.

Ding, H. et al. (2018). 'Skillful Climate Forecasts of the Tropical Indo-Pacific Ocean Using Model-Analogs'. In: *Journal of Climate* 31.14, pages 5437–5459. DOI: 10.1175/JCLI-D-17-0661.1.

Domeisen, D. I., C. I. Garfinkel, and A. H. Butler (2019). 'The Teleconnection of El Niño Southern Oscillation to the Stratosphere'. In: *Reviews of Geophysics* 57.1, pages 5–47. DOI: 10.1029/2018RG000596.

Dommenget, D., T. Bayr, and C. Frauen (2013). 'Analysis of the Non-Linearity in the Pattern and Time Evolution of El Niño Southern Oscillation'. In: *Clim Dyn* 40.11, pages 2825–2847. DOI: 10.1007/s00382-012-1475-0.

Donges, J. F. et al. (2009a). 'Complex Networks in Climate Dynamics'. In: *Eur. Phys. J. Spec. Top.* 174.1, pages 157–179. DOI: 10.1140/epjst/e2009-01098-2.

– (2009b). 'The Backbone of the Climate Network'. In: *EPL* 87.4, page 48007. DOI: 10.1209/0295-5075/87/48007.

Dool, H. M. van den (1989). 'A New Look at Weather Forecasting through Analogues'. In: *Monthly Weather Review* 117.10, pages 2230–2247. DOI: 10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2.

Duan, W. and C. Wei (2013). 'The 'Spring Predictability Barrier' for ENSO Predictions and Its Possible Mechanism: Results from a Fully Coupled Model'. In: *International Journal of Climatology* 33.5, pages 1280–1292. DOI: 10.1002/joc.3513.

Durack, P. J. et al. (2014). 'Quantifying Underestimates of Long-Term Upper-Ocean Warming'. In: *Nature Clim Change* 4.11, pages 999–1005. DOI: 10.1038/nclimate2389.

Ebert-Uphoff, I. and Y. Deng (2012). 'A New Type of Climate Network Based on Probabilistic Graphical Models: Results of Boreal Winter versus Summer'. In: *Geophysical Research Letters* 39.19. DOI: 10.1029/2012GL053269.
Eyring, V. et al. (2016). 'Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization'. In: *Geoscientific Model Development* 9.5, pages 1937–1958. DOI: 10.5194/gmd-9-1937-2016.
Farooq, H. et al. (2019). 'Network Curvature as a Hallmark of Brain Structural Connectivity'. In: *Nat Commun* 10.1, page 4937. DOI: 10.1038/s41467-019-12915-x.
Feder, T. (2000). 'Argo Begins Systematic Global Probing of the Upper Oceans'. In: *Physics Today* 53.7, pages 50–51. DOI: 10.1063/1.1292477.
Fedorov, A. V. and S. G. Philander (2000). 'Is El Niño Changing?' In: *Science* 288.5473, pages 1997–2002. DOI: 10.1126/science.288.5473.1997.
Fedorov, A. V. et al. (2015). 'The Impact of Westerly Wind Bursts and Ocean Initial State on the Development, and Diversity of El Niño Events'. In: *Clim Dyn* 44.5, pages 1381–1401. DOI: 10.1007/s00382-014-2126-4.
Forman (2003). 'Bochner's Method for Cell Complexes and Combinatorial Ricci Curvature'. In: *Discrete Comput Geom* 29.3, pages 323–374. DOI: 10.1007/s00454-002-0743-x.
Freeman, L. C. (1977). 'A Set of Measures of Centrality Based on Betweenness'. In: *Sociometry* 40.1, pages 35–41. DOI: 10.2307/3033543. JSTOR: 3033543.
Gan, R. et al. (2023). 'Greenhouse Warming and Internal Variability Increase Extreme and Central Pacific El Niño Frequency since 1980'. In: *Nature Communications* 14.1, page 394. DOI: 10.1038/s41467-023-36053-7.
Gao, L. et al. (2019). 'Measuring Road Network Topology Vulnerability by Ricci Curvature'. In: *Physica A: Statistical Mechanics and its Applications* 527, page 121071. DOI: 10.1016/j.physa.2019.121071.
Gao, Y. et al. (2020). 'An Extension of LDEO5 Model for ENSO Ensemble Predictions'. In: *Clim Dyn* 55.11, pages 2979–2991. DOI: 10.1007/s00382-020-05428-7.
Geng, L. and F.-F. Jin (2022). 'ENSO Diversity Simulated in a Revised Cane-Zebiak Model'. In: *Frontiers in Earth Science* 10.
Geng, T. et al. (2023). 'Increased Occurrences of Consecutive La Niña Events under Global Warming'. In: *Nature* 619.7971, pages 774–781. DOI: 10.1038/s41586-023-06236-9.
Giese, B. S. and S. Ray (2011). 'El Niño Variability in Simple Ocean Data Assimilation (SODA), 1871-2008'. In: *Journal of Geophysical Research: Oceans* 116.2, page C02024. DOI: 10.1029/2010JC006695.
Gill, A. E. (1980). 'Some Simple Solutions for Heat-Induced Tropical Circulation'. In: *Quarterly Journal of the Royal Meteorological Society* 106.449, pages 447–462. DOI: 10.1002/qj.49710644905.
Gneiting, T. and A. E. Raftery (2007). 'Strictly Proper Scoring Rules, Prediction, and Estimation'. In: *Journal of the American Statistical Association* 102.477, pages 359–378. DOI: 10.1198/016214506000001437.
Gneiting, T. et al. (2005). 'Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation'. In: *Monthly Weather Review* 133.5, pages 1098–1118. DOI: 10.1175/MWR2904.1.
Goel, H., I. Melnyk, and A. Banerjee (2017). *R2N2: Residual Recurrent Neural Networks for Multivariate Time Series Forecasting*. DOI: 10.48550/arXiv.1709.03159. arXiv: 1709.03159 [cs, stat].
Gonzalez, P. L. M. and L. Goddard (2016). 'Long-Lead ENSO Predictability from CMIP5 Decadal Hindcasts'. In: *Clim Dyn* 46.9, pages 3127–3147. DOI: 10.1007/s00382-015-2757-0.
Górski, K. M. et al. (2005). 'HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere'. In: *ApJ* 622.2, page 759. DOI: 10.1086/427976.
Graf, H.-F. and D. Zanchettin (2012). 'Central Pacific El Niño, the "Subtropical Bridge," and Eurasian Climate'. In: *Journal of Geophysical Research: Atmospheres* 117.D1. DOI: 10.1029/2011JD016493.
Grothe, P. R. et al. (2020). 'Enhanced El Niño–Southern Oscillation Variability in Recent Decades'. In: *Geophysical Research Letters* 47.7, e2019GL083906. DOI: 10.1029/2019GL083906.
Guckenheimer, J. et al. (2017). '(Un)Predictability of Strong El Niño Events'. In: *Dynamics and Statistics of the Climate System* 2.1, dzx004. DOI: 10.1093/climsys/dzx004.
Guilyardi, E. et al. (2020). 'ENSO Modeling'. In: *El Niño Southern Oscillation in a Changing Climate*. American Geophysical Union (AGU). Chapter 9, pages 199–226. DOI: 10.1002/9781119548164.ch9.
Hagberg, A. A., D. A. Schult, and P. J. Swart (2008). 'Exploring Network Structure, Dynamics, and Function Using NetworkX'. In: *Proceedings of the 7th Python in Science Conference*. Edited by G. Varoquaux, T. Vaught, and J. Millman. Pasadena, CA USA, pages 11–15.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo (2019). 'Deep Learning for Multi-Year ENSO Forecasts'. In: *Nature* 573.7775, pages 568–572. DOI: 10.1038/s41586-019-1559-7.
Ham, Y.-G. and J.-S. Kug (2012). 'How Well Do Current Climate Models Simulate Two Types of El Nino?' In: *Clim Dyn* 39.1-2, pages 383–398. DOI: 10.1007/s00382-011-1157-3.
He, K. et al. (2015). *Deep Residual Learning for Image Recognition*. DOI: 10.48550/arXiv.1512.03385. arXiv: 1512.03385 [cs].
Heitzig, J. et al. (2012). 'Node-Weighted Measures for Complex Networks with Spatially Embedded, Sampled, or Differently Sized Nodes'. In: *Eur. Phys. J. B* 85.1, page 38. DOI: 10.1140/epjb/e2011-20678-7.
Hersbach, H. (2000). 'Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems'. In: *Weather and Forecasting* 15.5, pages 559–570. DOI: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.
Hersbach, H. et al. (2020). 'The ERA5 Global Reanalysis'. In: *Q. J. R. Meteorolog. Soc.* 146.730, pages 1999–2049. DOI: 10.1002/qj.3803.
Hochreiter, S. and J. Schmidhuber (1997). 'Long Short-Term Memory'. In: *Neural Computation* 9.8, pages 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
Holm, S. (1979). 'A Simple Sequentially Rejective Multiple Test Procedure'. In: *Scandinavian Journal of Statistics* 6.2, pages 65–70. JSTOR: 4615733.
Hortal, M. and A. J. Simmons (1991). 'Use of Reduced Gaussian Grids in Spectral Models'. In: *Monthly Weather Review* 119.4, pages 1057–1074. DOI: 10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2.
Huang, B. (2004). 'Remotely Forced Variability in the Tropical Atlantic Ocean'. In: *Climate Dynamics* 23.2, pages 133–152. DOI: 10.1007/s00382-004-0443-8.
Hughes, T. P. et al. (2017). 'Coral Reefs in the Anthropocene'. In: *Nature* 546.7656, pages 82–90. DOI: 10.1038/nature22901.
IRI (2024). *IRI – International Research Institute for Climate and Society | April 2024 Quick Look*.
Irrgang, C. et al. (2021). 'Towards Neural Earth System Modelling by Integrating Artificial Intelligence in Earth System Science'. In: *Nat Mach Intell* 3.8, pages 667–674. DOI: 10.1038/s42256-021-00374-3.
Jiménez-Esteve, B. and D. I. V. Domeisen (2018). 'The Tropospheric Pathway of the ENSO–North Atlantic Teleconnection'. In: *Journal of Climate* 31.11, pages 4563–4584. DOI: 10.1175/JCLI-D-17-0716.1.
Jin, F.-F. (1997). 'An Equatorial Ocean Recharge Paradigm for ENSO. Part I: Conceptual Model'. In: *Journal of the Atmospheric Sciences* 54.7, pages 811–829. DOI: 10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2.
Jin, F.-F. et al. (2007). 'Ensemble-Mean Dynamics of the ENSO Recharge Oscillator under State-Dependent Stochastic Forcing'. In: *Geophysical Research Letters* 34.3. DOI: 10.1029/2006GL027372.
Jin, F.-F. et al. (2020). 'Simple ENSO Models'. In: *El Niño Southern Oscillation in a Changing Climate*. American Geophysical Union (AGU). Chapter 6, pages 119–151. DOI: 10.1002/9781119548164.ch6.
Johnson, N. C. (2013). 'How Many Enso Flavors Can We Distinguish?' In: *Journal of Climate* 26.13, pages 4816–4827. DOI: 10.1175/JCLI-D-12-00649.1.
Jonckheere, E. and E. Grippo (2019). 'Ollivier-Ricci Curvature Approach to Cost-Effective Power Grid Congestion Management'. In: *2019 Chinese Control And Decision Conference (CCDC)*. Nanchang, China: IEEE, pages 2118–2123. DOI: 10.1109/CCDC.2019.8832819.
Judt, F. (2020). 'Atmospheric Predictability of the Tropics, Middle Latitudes, and Polar Regions Explored through Global Storm-Resolving Simulations'. In: *Journal of the Atmospheric Sciences* 77.1, pages 257–276. DOI: 10.1175/JAS-D-19-0116.1.
Kaddour, J. et al. (2022). *Causal Machine Learning: A Survey and Open Problems*. DOI: 10.48550/arXiv.2206.15475. arXiv: 2206.15475 [cs, stat].
Kao, H. Y. and J. Y. Yu (2009). 'Contrasting Eastern-Pacific and Central-Pacific Types of ENSO'. In: *Journal of Climate* 22.3, pages 615–632. DOI: 10.1175/2008JCLI2309.1.
Kaplan, J. et al. (2020). *Scaling Laws for Neural Language Models*. DOI: 10.48550/arXiv.2001.08361. arXiv: 2001.08361 [cs, stat].
Karamperidou, C., F.-F. Jin, and J. L. Conroy (2017). 'The Importance of ENSO Nonlinearities in Tropical Pacific Response to External Forcing'. In: *Clim Dyn* 49.7, pages 2695–2704. DOI: 10.1007/s00382-016-3475-y.
Kirtman, B. P. et al. (2014). 'The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction'. In: *Bulletin of the American Meteorological Society* 95.4, pages 585–601. DOI: 10.1175/BAMS-D-12-00050.1.

Kittel, T. et al. (2021). 'Evolving Climate Network Perspectives on Global Surface Air Temperature Effects of ENSO and Strong Volcanic Eruptions'. In: *Eur. Phys. J. Spec. Top.* 230.14, pages 3075–3100. DOI: 10.1140/epjs/s11734-021-00269-9.

Klein, S. A., B. J. Soden, and N.-C. Lau (1999). 'Remote Sea Surface Temperature Variations during ENSO: Evidence for a Tropical Atmospheric Bridge'. In: *Journal of Climate* 12.4, pages 917–932. DOI: 10.1175/1520-0442(1999)012<0917:RSSTVD>2.0.CO;2.

Kochkov, D. et al. (2023). *Neural General Circulation Models*. DOI: 10.48550/arXiv.2311.07222. arXiv: 2311.07222 [physics].

Kug, J.-S. and Y.-G. Ham (2011). 'Are There Two Types of La Nina?' In: *Geophysical Research Letters* 38.16. DOI: 10.1029/2011GL048237.

Kug, J. S., F. F. Jin, and S. I. An (2009). 'Two Types of El Niño Events: Cold Tongue El Niño and Warm Pool El Niño'. In: *Journal of Climate* 22.6, pages 1499–1515. DOI: 10.1175/2008JCLI2624.1.

Kug, J.-S. et al. (2010). 'Warm Pool and Cold Tongue El Niño Events as Simulated by the GFDL 2.1 Coupled GCM'. In: *Journal of Climate* 23.5, pages 1226–1239. DOI: 10.1175/2009JCLI3293.1.

Kurihara, Y. (1965). 'NUMERICAL INTEGRATION OF THE PRIMITIVE EQUATIONS ON A SPHERICAL GRID'. In: *Monthly Weather Review* 93.7, pages 399–415. DOI: 10.1175/1520-0493(1965)093<0399:NIOTPE>2.3.CO;2.

Laloyaux, P. et al. (2018). 'CERA-20C: A Coupled Reanalysis of the Twentieth Century'. In: *J. Adv. Model. Earth Syst.* 10.5, pages 1172–1195. DOI: 10.1029/2018MS001273.

Lam, R. et al. (2023). *GraphCast: Learning Skillful Medium-Range Global Weather Forecasting*. DOI: 10.48550/arXiv.2212.12794. arXiv: 2212.12794 [physics].

Lambert, F. H., M. J. Webb, and M. M. Joshi (2011). 'The Relationship between Land–Ocean Surface Temperature Contrast and Radiative Forcing'. In: *Journal of Climate* 24.13, pages 3239–3256. DOI: 10.1175/2011JCLI3893.1.

Larkin, N. K. and D. E. Harrison (2005). 'Global Seasonal Temperature and Precipitation Anomalies during El Niño Autumn and Winter'. In: *Geophysical Research Letters* 32.16, pages 1–4. DOI: 10.1029/2005GL022860.

Latif, M. et al. (1994). 'A Review of ENSO Prediction Studies'. In: *Climate Dynamics* 9.4, pages 167–179. DOI: 10.1007/BF00208250.

Lawman, A. E. et al. (2022). 'Unraveling Forced Responses of Extreme El Niño Variability over the Holocene'. In: *Science Advances* 8.9, eabm4313. DOI: 10.1126/sciadv.abm4313.

Lecun, Y. et al. (1998). 'Gradient-Based Learning Applied to Document Recognition'. In: *Proceedings of the IEEE* 86.11, pages 2278–2324. DOI: 10.1109/5.726791.

Lee, T. and M. J. McPhaden (2010). 'Increasing Intensity of El Niño in the Central-Equatorial Pacific'. In: *Geophysical Research Letters* 37.14, page L14603. DOI: 10.1029/2010GL044007.

Leslie, J. and K. Rodgers (2024). *NOAA Confirms 4th Global Coral Bleaching Event | National Oceanic and Atmospheric Administration*. https://www.noaa.gov/news-release/noaa-confirms-4th-global-coral-bleaching-event.

Lessig, C. et al. (2023). *AtmoRep: A Stochastic Model of Atmosphere Dynamics Using Large Scale Representation Learning*. arXiv: 2308.13280 [physics].

Levine, A. F. Z. and F.-F. Jin (2010). 'Noise-Induced Instability in the ENSO Recharge Oscillator'. In: *Journal of the Atmospheric Sciences* 67.2, pages 529–542. DOI: 10.1175/2009JAS3213.1.

L'Heureux, M. L. (2020). *The Rise of El Niño and La Niña | NOAA Climate.Gov*. http://www.climate.gov/news-features/blogs/enso/rise-el-ni%C3%B1o-and-la-ni%C3%B1a.

Lin, Y.-S., L.-C. Wang, and J.-L. F. Li (2023). 'Effects of Equatorial Ocean Current Bias on Simulated El Niño Pattern in CMIP6 Models'. In: *Geophysical Research Letters* 50.8, e2023GL102890. DOI: 10.1029/2023GL102890.

Liu, B. et al. (2022a). 'Will Increasing Climate Model Resolution Be Beneficial for ENSO Simulation?' In: *Geophysical Research Letters* 49.11, e2021GL096932. DOI: 10.1029/2021GL096932.

Liu, C. et al. (2022b). 'Equatorial Origin of the Observed Tropical Pacific Quasi-Decadal Variability From ENSO Nonlinearity'. In: *Geophysical Research Letters* 49.10, e2022GL097903. DOI: 10.1029/2022GL097903.

Liu, Z. et al. (2021). *Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows*. DOI: 10.48550/arXiv.2103.14030. arXiv: 2103.14030 [cs].

Lorenz, E. N. (1963). 'Deterministic Nonperiodic Flow'. In: *Journal of the Atmospheric Sciences* 20.2, pages 130–141. DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

– (1969). 'The Predictability of a Flow Which Possesses Many Scales of Motion'. In: 21.3, page 289. DOI: 10.3402/tellusa.v21i3.10086.
Lorenz, E. (1956). *Empirical Orthogonal Functions and Statistical Weather Prediction*. Scientific Report. Massachusetts Institute of Technology, Department of Meteorology.
Loshchilov, I. and F. Hutter (2017). *SGDR: Stochastic Gradient Descent with Warm Restarts*. DOI: 10.48550/arXiv.1608.03983. arXiv: 1608.03983 [cs, math].
– (2019). *Decoupled Weight Decay Regularization*. DOI: 10.48550/arXiv.1711.05101. arXiv: 1711.05101 [cs, math].
Lou, J., M. Newman, and A. Hoell (2023). 'Multi-Decadal Variation of ENSO Forecast Skill since the Late 1800s'. In: *npj Clim Atmos Sci* 6.1, pages 1–14. DOI: 10.1038/s41612-023-00417-z.
Lu, Z. et al. (2020). 'On the Impacts of El Niño Events: A New Monitoring Approach Using Complex Network Analysis'. In: *Geophysical Research Letters* 47.6, e2019GL086533. DOI: 10.1029/2019GL086533.
Lyu, P. et al. (2023). *ResoNet: Robust and Explainable ENSO Forecasts with Hybrid Convolution and Transformer Networks*. DOI: 10.48550/arXiv.2312.10429. arXiv: 2312.10429 [physics].
Madden, R. A. and P. R. Julian (1971). 'Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific'. In: *Journal of the Atmospheric Sciences* 28.5, pages 702–708. DOI: 10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.
Magramo, K. and M. Ogunbayo (2024). *Australian State Orders 30,000 People to Evacuate Due to 'Catastrophic' Fire Risk*. https://www.cnn.com/2024/02/28/australia/australia-victoria-wildfire-intl-hnk/index.html.
Martinez-Villalobos, C. et al. (2018). 'Calculating State-Dependent Noise in a Linear Inverse Model Framework'. In: *Journal of the Atmospheric Sciences* 75.2, pages 479–496. DOI: 10.1175/JAS-D-17-0235.1.
McPhaden, M. J., T. Lee, and D. McClurg (2011). 'El Niño and Its Relationship to Changing Background Conditions in the Tropical Pacific Ocean'. In: *Geophysical Research Letters* 38.15. DOI: 10.1029/2011GL048275.
McPhaden, M. J. (2004). 'Evolution of the 2002/03 El Niño'. In: *Bull. Am. Meteorol. Soc.* 85.5, pages 677–696. DOI: 10.1175/BAMS-85-5-677.
– (2012). 'A 21st Century Shift in the Relationship between ENSO SST and Warm Water Volume Anomalies'. In: *Geophys. Res. Lett.* 39.9. DOI: 10.1029/2012GL051826.
McPhaden, M. J., A. Santoso, and W. Cai (2020). 'Introduction to El Niño Southern Oscillation in a Changing Climate'. In: *El Niño Southern Oscillation in a Changing Climate*. American Geophysical Union (AGU). Chapter 1, pages 1–19. DOI: 10.1002/9781119548164.ch1.
McPhaden, M. J. et al. (1998). 'The Tropical Ocean-Global Atmosphere Observing System: A Decade of Progress'. In: *Journal of Geophysical Research: Oceans* 103.C7, pages 14169–14240. DOI: 10.1029/97JC02906.
Meehl, G. A. et al. (2021). 'The Role of Interannual ENSO Events in Decadal Timescale Transitions of the Interdecadal Pacific Oscillation'. In: *Clim Dyn* 57.7, pages 1933–1951. DOI: 10.1007/s00382-021-05784-y.
Micheli, A. (2009). 'Neural Network for Graphs: A Contextual Constructive Approach'. In: *IEEE Transactions on Neural Networks* 20.3, pages 498–511. DOI: 10.1109/TNN.2008.2010350.
Mikolov, T. et al. (2013). 'Distributed Representations of Words and Phrases and Their Compositionality'. In: *Advances in Neural Information Processing Systems*. Volume 26. Curran Associates, Inc.
Miller, A. J. et al. (1994). 'The 1976-77 Climate Shift of the Pacific Ocean'. In: *Oceanography* 7.1, pages 21–26. JSTOR: 43925525.
Murphy, K. P. (2013). *Machine Learning: A Probabilistic Perspective*. MIT Press.
Neske, S. and S. McGregor (2018). 'Understanding the Warm Water Volume Precursor of ENSO Events and Its Interdecadal Variation'. In: *Geophysical Research Letters* 45.3, pages 1577–1585. DOI: 10.1002/2017GL076439.
Newman, M., M. A. Alexander, and J. D. Scott (2011a). 'An Empirical Model of Tropical Ocean Dynamics'. In: *Clim Dyn* 37.9-10, pages 1823–1841. DOI: 10.1007/s00382-011-1034-0.
Newman, M. and P. D. Sardeshmukh (2017). 'Are We near the Predictability Limit of Tropical Indo-Pacific Sea Surface Temperatures?' In: *Geophysical Research Letters* 44.16, pages 8520–8529. DOI: 10.1002/2017GL074088.
Newman, M., S. I. Shin, and M. A. Alexander (2011b). 'Natural Variation in ENSO Flavors'. In: *Geophysical Research Letters* 38.14, page L14705. DOI: 10.1029/2011GL047658.
Newman, M. et al. (2003). 'A Study of Subseasonal Predictability'. In: *Monthly Weather Review* 131.8, pages 1715–1732. DOI: 10.1175//2558.1.
Okumura, Y. M. (2019). 'ENSO Diversity from an Atmospheric Perspective'. In: *Curr Clim Change Rep* 5.3, pages 245–257. DOI: 10.1007/s40641-019-00138-7.

Ollivier, Y. (2009). 'Ricci Curvature of Markov Chains on Metric Spaces'. In: *Journal of Functional Analysis* 256.3, pages 810–864. DOI: 10.1016/j.jfa.2008.11.001.
– (2010). 'A Survey of Ricci Curvature for Metric Spaces and Markov Chains'. In: *Probabilistic Approach to Geometry*. Volume 57. Mathematical Society of Japan, pages 343–382. DOI: 10.2969/aspm/05710343.
ORAS5 (2021). *ORAS5 Global Ocean Reanalysis Monthly Data from 1958 to Present*. DOI: 10.24381/cds.67e8eeb7.
Paek, H., J.-Y. Yu, and C. Qian (2017). 'Why Were the 2015/2016 and 1997/1998 Extreme El Niños Different?' In: *Geophysical Research Letters* 44.4, pages 1848–1856. DOI: 10.1002/2016GL071515.
Palmer, T. N. (2001). 'A Nonlinear Dynamical Perspective on Model Error: A Proposal for Non-Local Stochastic-Dynamic Parametrization in Weather and Climate Prediction Models'. In: *Quarterly Journal of the Royal Meteorological Society* 127.572, pages 279–304. DOI: 10.1002/qj.49712757202.
Palmer, T. N., A. Döring, and G. Seregin (2014). 'The Real Butterfly Effect'. In: *Nonlinearity* 27.9, R123. DOI: 10.1088/0951-7715/27/9/R123.
Parsons, L., E. Haque, and H. Liu (2004). 'Subspace Clustering for High Dimensional Data: A Review'. In: *SIGKDD Explor. Newsl.* 6.1, pages 90–105. DOI: 10.1145/1007730.1007731.
Pascolini-Campbell, M. et al. (2015). 'Toward a Record of Central Pacific El Niño Events since 1880'. In: *Theoretical and Applied Climatology* 119.1-2, pages 379–389. DOI: 10.1007/s00704-014-1114-2.
Paszke, A. et al. (2017). 'Automatic Differentiation in PyTorch'. In.
Pathak, J. et al. (2022). *FourCastNet: A Global Data-driven High-resolution Weather Model Using Adaptive Fourier Neural Operators*. DOI: 10.48550/arXiv.2202.11214. arXiv: 2202.11214 [physics].
Penland, C. and P. D. Sardeshmukh (1995). 'The Optimal Growth of Tropical Sea Surface Temperature Anomalies'. In: *Journal of Climate* 8.8, pages 1999–2024. DOI: 10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2.
Perez, E. et al. (2017). *FiLM: Visual Reasoning with a General Conditioning Layer*. DOI: 10.48550/arXiv.1709.07871. arXiv: 1709.07871 [cs, stat].
Petersik, P. J. and H. A. Dijkstra (2020). 'Probabilistic Forecasting of El Niño Using Neural Network Models'. In: *Geophys. Res. Lett.* 47.6. DOI: 10.1029/2019GL086423.
Pouryahya, M. et al. (2018). 'Characterizing Cancer Drug Response and Biological Correlates: A Geometric Network Approach'. In: *Sci Rep* 8.1, page 6402. DOI: 10.1038/s41598-018-24679-3.
Puy, M. et al. (2019). 'Influence of Westerly Wind Events Stochasticity on El Niño Amplitude: The Case of 2014 vs. 2015'. In: *Clim Dyn* 52.12, pages 7435–7454. DOI: 10.1007/s00382-017-3938-9.
Radebach, A. et al. (2013). 'Disentangling Different Types of El Ni\~no Episodes by Evolving Climate Network Analysis'. In: *Phys. Rev. E* 88.5, page 052807. DOI: 10.1103/PhysRevE.88.052807.
Rasmusson, E. M. and T. H. Carpenter (1982). 'Variations in Tropical Sea Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation/El Niño'. In: *Monthly Weather Review* 110.5, pages 354–384. DOI: 10.1175/1520-0493(1982)110<0354:VITSST>2.0.CO;2.
Rasp, S. and S. Lerch (2018). 'Neural Networks for Postprocessing Ensemble Weather Forecasts'. In: *Monthly Weather Review* 146.11, pages 3885–3900. DOI: 10.1175/MWR-D-18-0187.1.
Rayner, N. A. et al. (2003). 'Global Analyses of Sea Surface Temperature, Sea Ice, and Night Marine Air Temperature since the Late Nineteenth Century'. In: *J. Geophys. Res. Atmos.* 108.D14, page 4407. DOI: 10.1029/2002JD002670.
Reanalyzer, C. (2024). *Climate Reanalyzer*. https://climatereanalyzer.org/clim/sst_daily/.
Rebert, J. P. et al. (1985). 'Relations between Sea Level, Thermocline Depth, Heat Content, and Dynamic Height in the Tropical Pacific Ocean'. In: *Journal of Geophysical Research: Oceans* 90.C6, pages 11719–11725. DOI: 10.1029/JC090iC06p11719.
Richter, I. et al. (2022). 'Disentangling the North Pacific Meridional Mode from Tropical Pacific Variability'. In: *npj Clim Atmos Sci* 5.1, pages 1–9. DOI: 10.1038/s41612-022-00317-8.
Risbey, J. S. et al. (2021). 'Standard Assessments of Climate Forecast Skill Can Be Misleading'. In: *Nat Commun* 12.1, page 4346. DOI: 10.1038/s41467-021-23771-z.
Robbins, H. and S. Monro (1951). 'A Stochastic Approximation Method'. In: *The Annals of Mathematical Statistics* 22.3, pages 400–407. DOI: 10.1214/aoms/1177729586.
Rodrigues, E. et al. (2021). *Decadal Forecasts with ResDMD: A Residual DMD Neural Network*. arXiv: 2106.11111 [cs].

Rodrigues, R. R., E. J. D. Campos, and R. Haarsma (2015). 'The Impact of ENSO on the South Atlantic Subtropical Dipole Mode'. In: *Journal of Climate* 28.7, pages 2691–2705. DOI: 10.1175/JCLI-D-14-00483.1.

Rodrigues, R. R. et al. (2011). 'The Impacts of Inter–El Niño Variability on the Tropical Atlantic and Northeast Brazil Climate'. In: *Journal of Climate* 24.13, pages 3402–3422. DOI: 10.1175/2011JCLI3983.1.

Rumelhart, D. E., J. L. McClelland, and P. R. Group (1986). *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. The MIT Press.

Runge, J. et al. (2019a). 'Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets'. In: *Science Advances* 5.11, eaau4996. DOI: 10.1126/sciadv.aau4996.

Runge, J. et al. (2019b). 'Inferring Causation from Time Series in Earth System Sciences'. In: *Nat Commun* 10.1, page 2553. DOI: 10.1038/s41467-019-10105-3.

Sadourny, R., A. Arakawa, and Y. Mintz (1968). 'INTEGRATION OF THE NONDIVERGENT BAROTROPIC VORTICITY EQUATION WITH AN ICOSAHEDRAL-HEXAGONAL GRID FOR THE SPHERE'. In: *Monthly Weather Review* 96.6, pages 351–356. DOI: 10.1175/1520-0493(1968)096<0351:IOTNBV>2.0.CO;2.

Samal, A. et al. (2018). 'Comparative Analysis of Two Discretizations of Ricci Curvature for Complex Networks'. In: *Sci Rep* 8.1, page 8650. DOI: 10.1038/s41598-018-27001-3.

Sandhu, R. et al. (2015). 'Graph Curvature for Differentiating Cancer Networks'. In: *Sci Rep* 5.1, page 12323. DOI: 10.1038/srep12323.

Sandhu, R. S., T. T. Georgiou, and A. R. Tannenbaum (2016). 'Ricci Curvature: An Economic Indicator for Market Fragility and Systemic Risk'. In: *Science Advances* 2.5, e1501495. DOI: 10.1126/sciadv.1501495.

Santoso, A., M. J. Mcphaden, and W. Cai (2017). 'The Defining Characteristics of ENSO Extremes and the Strong 2015/2016 El Niño'. In: *Reviews of Geophysics* 55.4, pages 1079–1129. DOI: 10.1002/2017RG000560.

Sardeshmukh, P. D., G. P. Compo, and C. Penland (2000). 'Changes of Probability Associated with El Niño'. In: *J. Climate* 13.24, pages 4268–4286. DOI: 10.1175/1520-0442(2000)013<4268:COPAWE>2.0.CO;2.

Scarselli, F. et al. (2009). 'The Graph Neural Network Model'. In: *IEEE Transactions on Neural Networks* 20.1, pages 61–80. DOI: 10.1109/TNN.2008.2005605.

Schlör, J. (2023). *LatentGMM: An Unsupervised Fuzzy Clustering Approach for ENSO*. Zenodo. DOI: 10.5281/zenodo.10144391.

Schlör, J. and F. M. Strnad (2022a). *Climnet: Creating and Working with Climate Networks*. Zenodo. DOI: 10.5281/zenodo.6325661.

– (2022b). *Netcurvature: Curvature Applied to Climate Networks to Study Teleconnection Patterns*. Zenodo. DOI: 10.5281/zenodo.6325300.

Schlör, J. et al. (2024). 'A Hybrid Model for ENSO Dynamics in the Low-Data Regime'. In Preparation.

Schlör, J. et al. (2024). *Contribution of El Niño Southern Oscillation (ENSO) Diversity to Low-Frequency Changes in ENSO Variance*. DOI: 10.1029/2024GL109179.

Schölkopf, B. (2022). 'Causality for Machine Learning'. In: pages 765–804. DOI: 10.1145/3501714.3501755. arXiv: 1911.10500 [cs, stat].

Schwarz, G. (1978). 'Estimating the Dimension of a Model'. In: *The Annals of Statistics* 6.2, pages 461–464. DOI: 10.1214/aos/1176344136.

Shi, J., A. V. Fedorov, and S. Hu (2019). 'North Pacific Temperature and Precipitation Response to El Niño-like Equatorial Heating: Sensitivity to Forcing Location'. In: *Clim Dyn* 53.5, pages 2731–2741. DOI: 10.1007/s00382-019-04655-x.

Shi, X. et al. (2015). *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*. DOI: 10.48550/arXiv.1506.04214. arXiv: 1506.04214 [cs].

Shin, N.-Y. et al. (2021). 'The Double-Peaked El Niño and Its Physical Processes'. In: *Journal of Climate* 34.4, pages 1291–1303. DOI: 10.1175/JCLI-D-20-0402.1.

Sidak, Z. (1967). 'Rectangular Confidence Regions for the Means of Multivariate Normal Distributions'. In: *Journal of the American Statistical Association* 62.318, pages 626–633. DOI: 10.2307/2283989. JSTOR: 2283989.

Souza, D. B. de et al. (2021). 'Using Discrete Ricci Curvatures to Infer COVID-19 Epidemic Network Fragility and Systemic Risk'. In: *J. Stat. Mech.* 2021.5, page 053501. DOI: 10.1088/1742-5468/abed4e.

Sprintall, J. et al. (2020). 'ENSO Oceanic Teleconnections'. In: *El Niño Southern Oscillation in a Changing Climate*. American Geophysical Union (AGU). Chapter 15, pages 337–359. DOI: 10.1002/9781119548164.ch15.

Sreejith, R. P. et al. (2016). 'Forman Curvature for Complex Networks'. In: *J. Stat. Mech.* 2016.6, page 063206. DOI: 10.1088/1742-5468/2016/06/063206.

Srivastava, N. et al. (2014). 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting'. In: *Journal of Machine Learning Research* 15.56, pages 1929–1958.
Stein, K. et al. (2014). 'ENSO Seasonal Synchronization Theory'. In: *Journal of Climate* 27.14, pages 5285–5310. DOI: 10.1175/JCLI-D-13-00525.1.
Stephenson, D. B. et al. (2003). 'The History of Scientific Research on the North Atlantic Oscillation'. In: *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*. American Geophysical Union (AGU), pages 37–50. DOI: 10.1029/134GM02.
Stephenson, D. B. et al. (2012). 'Statistical Problems in the Probabilistic Prediction of Climate Change'. In: *Environmetrics* 23.5, pages 364–372. DOI: 10.1002/env.2153.
Strnad, F. M. et al. (2022). 'Teleconnection Patterns of Different El Niño Types Revealed by Climate Network Curvature'. In: *Geophysical Research Letters* 49.17, e2022GL098571. DOI: 10.1029/2022GL098571.
Suarez, M. J. and P. S. Schopf (1988). 'A Delayed Action Oscillator for ENSO'. In: *Journal of the Atmospheric Sciences* 45.21, pages 3283–3287. DOI: 10.1175/1520-0469(1988)045<3283:ADAOFE>2.0.CO;2.
Sutskever, I., O. Vinyals, and Q. V. Le (2014). *Sequence to Sequence Learning with Neural Networks*. DOI: 10.48550/arXiv.1409.3215. arXiv: 1409.3215 [cs].
Sverdrup, H. U. (1947). 'Wind-Driven Currents in a Baroclinic Ocean; with Application to the Equatorial Currents of the Eastern Pacific'. In: *Proc Natl Acad Sci U S A* 33.11, pages 318–326.
Swinbank, R. and R. James Purser (2006). 'Fibonacci Grids: A Novel Approach to Global Modelling'. In: *Quarterly Journal of the Royal Meteorological Society* 132.619, pages 1769–1793. DOI: 10.1256/qj.05.227.
Takahashi, K. et al. (2011). 'ENSO Regimes: Reinterpreting the Canonical and Modoki El Niño'. In: *Geophysical Research Letters* 38.10, page L10704. DOI: 10.1029/2011GL047364.
Takahashi, K. and B. Dewitte (2016). 'Strong and Moderate Nonlinear El Niño Regimes'. In: *Climate Dynamics* 46.5-6, pages 1627–1645. DOI: 10.1007/s00382-015-2665-3.
Tangang, F. T., W. W. Hsieh, and B. Tang (1997). 'Forecasting the Equatorial Pacific Sea Surface Temperatures by Neural Network Models'. In: *Climate Dynamics* 13.2, pages 135–147. DOI: 10.1007/s003820050156.
Taschetto, A. S. et al. (2020). 'ENSO Atmospheric Teleconnections'. In: *El Niño Southern Oscillation in a Changing Climate*. American Geophysical Union (AGU). Chapter 14, pages 309–335. DOI: 10.1002/9781119548164.ch14.
Taschetto, A. S. and M. H. England (2009). 'El Niño Modoki Impacts on Australian Rainfall'. In: *J. Clim.* 22.11, pages 3167–3174. DOI: 10.1175/2008JCLI2589.1.
Taylor, J. and M. Feng (2022). 'A Deep Learning Model for Forecasting Global Monthly Mean Sea Surface Temperature Anomalies'. In: *Front. Clim.* 4. DOI: 10.3389/fclim.2022.932932.
Thual, S. and B. Dewitte (2023). 'ENSO Complexity Controlled by Zonal Shifts in the Walker Circulation'. In: *Nat. Geosci.* 16.4, pages 328–332. DOI: 10.1038/s41561-023-01154-x.
Thuemmel, J. et al. (2024). 'Inductive Biases in Deep Learning Models for Weather Prediction'.
Thurman, W. N. and M. E. Fisher (1988). 'Chickens, Eggs, and Causality, or Which Came First?' In: *American Journal of Agricultural Economics* 70.2, pages 237–238. DOI: 10.2307/1242062.
Timmermann, A., F.-F. Jin, and J. Abshagen (2003). 'A Nonlinear Theory for El Niño Bursting'. In: *Journal of the Atmospheric Sciences* 60.1, pages 152–165. DOI: 10.1175/1520-0469(2003)060<0152:ANTFEN>2.0.CO;2.
Timmermann, A. et al. (2018). 'El Niño–Southern Oscillation Complexity'. In: *Nature* 559.7715, pages 535–545. DOI: 10.1038/s41586-018-0252-6.
Toride, K. et al. (2024). *Using Deep Learning to Identify Initial Error Sensitivity of ENSO Forecasts*. DOI: 10.48550/arXiv.2404.15419. arXiv: 2404.15419 [physics].
Trenberth, K. E. (1997). 'The Definition of El Niño'. In: *Bulletin of the American Meteorological Society* 78.12, pages 2771–2778. DOI: 10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.
– (2011). 'Changes in Precipitation with Climate Change'. In: *Climate Research* 47.1-2, pages 123–138. DOI: 10.3354/cr00953.
Trenberth, K. E. and D. P. Stepaniak (2001). 'Indices of El Niño Evolution'. In: *Journal of Climate* 14.8, pages 1697–1701. DOI: 10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2.
Tsonis, A. A. and K. L. Swanson (2008). 'Topology and Predictability of El Ni\~no and La Ni\~na Networks'. In: *Phys. Rev. Lett.* 100.22, page 228502. DOI: 10.1103/PhysRevLett.100.228502.
Vaswani, A. et al. (2023). *Attention Is All You Need*. DOI: 10.48550/arXiv.1706.03762. arXiv: 1706.03762 [cs].

Vecchi, G. A. et al. (2006). 'Weakening of Tropical Pacific Atmospheric Circulation Due to Anthropogenic Forcing'. In: *Nature* 441.7089, pages 73–76. DOI: 10.1038/nature04744.
Vimont, D. J., M. A. Alexander, and M. Newman (2014). 'Optimal Growth of Central and East Pacific ENSO Events'. In: *Geophysical Research Letters* 41.11, pages 4027–4034. DOI: 10.1002/2014GL059997.
Vimont, D. J. et al. (2022). 'The Role of Seasonality and the ENSO Mode in Central and East Pacific ENSO Growth and Evolution'. In: *J. Clim.* 35.11, pages 3195–3209. DOI: 10.1175/JCLI-D-21-0599.1.
von Storch, H. and J. Xu (1990). 'Principal Oscillation Pattern Analysis of the 30- to 60-Day Oscillation in the Tropical Troposphere'. In: *Climate Dynamics* 4.3, pages 175–190. DOI: 10.1007/BF00209520.
Walker, G. T. (1925). 'CORRELATION IN SEASONAL VARIATIONS OF WEATHER—A FURTHER STUDY OF WORLD WEATHER'. In: *Monthly Weather Review* 53.6, pages 252–254. DOI: 10.1175/1520-0493(1925)53<252:CISVOW>2.0.CO;2.
Wang, C. and X. Wang (2013). 'Classifying El Niño Modoki I and II by Different Impacts on Rainfall in Southern China and Typhoon Tracks'. In: *Journal of Climate* 26.4, pages 1322–1338. DOI: 10.1175/JCLI-D-12-00107.1.
Wang, G. and H. H. Hendon (2007). 'Sensitivity of Australian Rainfall to Inter–El Niño Variations'. In: *Journal of Climate* 20.16, pages 4211–4226. DOI: 10.1175/JCLI4228.1.
Wang, S., L. Mu, and D. Liu (2021). 'A Hybrid Approach for El Niño Prediction Based on Empirical Mode Decomposition and Convolutional LSTM Encoder-Decoder'. In: *Computers & Geosciences* 149, page 104695. DOI: 10.1016/j.cageo.2021.104695.
Wang, T. and P. Huang (2024). 'Superiority of a Convolutional Neural Network Model over Dynamical Models in Predicting Central Pacific ENSO'. In: *Adv. Atmos. Sci.* 41.1, pages 141–154. DOI: 10.1007/s00376-023-3001-1.
Wang, X. and C. Wang (2014). 'Different Impacts of Various El Niño Events on the Indian Ocean Dipole'. In: *Clim Dyn* 42.3, pages 991–1005. DOI: 10.1007/s00382-013-1711-2.
Wang, X., J. Slawinska, and D. Giannakis (2020). 'Extended-Range Statistical ENSO Prediction through Operator-Theoretic Techniques for Nonlinear Dynamics'. In: *Sci Rep* 10.1, page 2636. DOI: 10.1038/s41598-020-59128-7.
Wang, Y., Y. Zhang, and G.-G. Wang (2023). 'Forecasting ENSO Using Convolutional LSTM Network with Improved Attention Mechanism and Models Recombined by Genetic Algorithm in CMIP5/6'. In: *Information Sciences* 642, page 119106. DOI: 10.1016/j.ins.2023.119106.
Watt-Meyer, O. et al. (2021). 'Correcting Weather and Climate Models by Machine Learning Nudged Historical Simulations'. In: *Geophysical Research Letters* 48.15, e2021GL092555. DOI: 10.1029/2021GL092555.
Wei, H.-H. et al. (2021). 'Tropical Pacific Air-Sea Interaction Processes and Biases in CESM2 and Their Relation to El Niño Development'. In: *Journal of Geophysical Research: Oceans* 126.6, e2020JC016967. DOI: 10.1029/2020JC016967.
Wiedermann, M. et al. (2016). 'A Climate Network-Based Index to Discriminate Different Types of El Niño and La Niña'. In: *Geophysical Research Letters* 43.13, pages 7176–7185. DOI: 10.1002/2016GL069119.
Wilks, D. S. (2019). *Statistical Methods in the Atmospheric Sciences*. Fourth edition. Amsterdam, Netherlands ; Cambridge, MA: Elsevier.
Williamson, D. L. (1968). 'Integration of the Barotropic Vorticity Equation on a Spherical Geodesic Grid'. In: *Tellus* 20.4, pages 642–653. DOI: 10.1111/j.2153-3490.1968.tb00406.x.
Wittenberg, A. T. (2009). 'Are Historical Records Sufficient to Constrain ENSO Simulations?' In: *Geophysical Research Letters* 36.12, page L12702. DOI: 10.1029/2009GL038710.
Wittenberg, A. T. et al. (2006). 'GFDL's CM2 Global Coupled Climate Models. Part III: Tropical Pacific Climate and ENSO'. In: *Journal of Climate* 19.5, pages 698–722. DOI: 10.1175/JCLI3631.1.
Wolpert, D. and W. Macready (1997). 'No Free Lunch Theorems for Optimization'. In: *IEEE Transactions on Evolutionary Computation* 1.1, pages 67–82. DOI: 10.1109/4235.585893.
Wu, Y. and K. He (2018). *Group Normalization*. DOI: 10.48550/arXiv.1803.08494. arXiv: 1803.08494 [cs].
Wyrtki, K. (1975). 'El Niño—The Dynamic Response of the Equatorial Pacific Oceanto Atmospheric Forcing'. In: *Journal of Physical Oceanography* 5.4, pages 572–584. DOI: 10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2.
Xue, Y. et al. (1994). 'On the Prediction of ENSO: A Study with a Low-Order Markov Model'. In: 46.4, page 512. DOI: 10.3402/tellusa.v46i4.15641.

Yamasaki, K., A. Gozolchiani, and S. Havlin (2008). 'Climate Networks around the Globe Are Significantly Affected by El Ni\~no'. In: *Phys. Rev. Lett.* 100.22, page 228501. DOI: 10.1103/PhysRevLett.100.228501.

Yeh, S. W. et al. (2009). 'El Nino in a Changing Climate'. In: *Nature* 461.7263, pages 511–514. DOI: 10.1038/nature08316.

You, Y. and J. C. Furtado (2018). 'The South Pacific Meridional Mode and Its Role in Tropical Pacific Climate Variability'. In: *Journal of Climate* 31.24, pages 10141–10163. DOI: 10.1175/JCLI-D-17-0860.1.

Yu, G., G. Sapiro, and S. Mallat (2010). 'Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity'. In: *arXiv*. DOI: 10.48550/arXiv.1006.3056.

Yu, J. Y. and S. T. Kim (2013). 'Identifying the Types of Major El Niño Events since 1870'. In: *International Journal of Climatology* 33.8, pages 2105–2112. DOI: 10.1002/joc.3575.

Yu, J.-Y. et al. (2012). 'The Changing Impact of El Niño on US Winter Temperatures'. In: *Geophysical Research Letters* 39.15. DOI: 10.1029/2012GL052483.

Zebiak, S. E. (1993). 'Air–Sea Interaction in the Equatorial Atlantic Region'. In: *Journal of Climate* 6.8, pages 1567–1586. DOI: 10.1175/1520-0442(1993)006<1567:AIITEA>2.0.CO;2.

Zebiak, S. E. and M. A. Cane (1987). 'A Model El Ni&ntilde–Southern Oscillation'. In: *Monthly Weather Review* 115.10, pages 2262–2278. DOI: 10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2.

Zhang, H.-M. et al. (2019). *NOAA Global Surface Temperature Dataset (NOAAGlobalTemp), Version 5.0*. DOI: 10.25921/9qth-2p70.

Zhang, H., A. Clement, and P. D. Nezio (2014). 'The South Pacific Meridional Mode: A Mechanism for ENSO-like Variability'. In: *Journal of Climate* 27.2, pages 769–783. DOI: 10.1175/JCLI-D-13-00082.1.

Zhang, L. and W. Han (2021). 'Indian Ocean Dipole Leads to Atlantic Niño'. In: *Nat Commun* 12.1, page 5952. DOI: 10.1038/s41467-021-26223-w.

Zhang, W. et al. (2015). 'Impact of Different El Niño Types on the El Niño/IOD Relationship'. In: *Geophysical Research Letters* 42.20, pages 8570–8576. DOI: 10.1002/2015GL065703.

Zhou, D. et al. (2015). 'Teleconnection Paths via Climate Network Direct Link Detection'. In: *Phys. Rev. Lett.* 115.26, page 268501. DOI: 10.1103/PhysRevLett.115.268501.

Zhou, L. and R.-H. Zhang (2022). 'A Hybrid Neural Network Model for ENSO Prediction in Combination with Principal Oscillation Pattern Analyses'. In: *Adv. Atmos. Sci.* 39.6, pages 889–902. DOI: 10.1007/s00376-021-1368-4.

– (2023). 'A Self-Attention–Based Neural Network for Three-Dimensional Multivariate Modeling and Its Skillful ENSO Predictions'. In: *Sci. Adv.* 9.10, eadf2827. DOI: 10.1126/sciadv.adf2827.