Exploring Information Exchange in Climate System and Climate Models.

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 11 Geowissenschaften/Geographie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Praveen Kumar Pothapakula
aus Bapatla, India

Frankfurt 2022
(D30)

vom Fachbereich 11 Geowissenschaften/ Geographie der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:
Prof. Dr. Jürgen Runge

Gutachter:
Prof.Dr. Bodo Ahrens
Prof.Dr. Henning Rust

Datum der Disputation: 03-11-2022

Abstract

The climate system is one of the classical examples of a complex dynamical system consisting of interacting sub-systems through mass, momentum, and energy exchange across various spatial and temporal scales. This thesis aims to detect and quantify sub-component interactions from an information exchange (IE) perspective. For this purpose, IE estimators derived from information theory are explored and applied to the available climate data obtained from observations, reanalysis, global and regional climate models. Specifically, this thesis investigates the usefulness of information theory methods for process-oriented climate model evaluation.

Firstly, methods derived from the concepts of information theory such as transfer entropy and information flow along with their linear and non-linear estimation techniques are initially tested and applied to idealized two-dimensional dynamical systems. The results revealed an expected direction and magnitude of IE providing insights into underlying dynamics. However, as expected the linear estimators are robust for linear systems but fail for non-linear systems. Though the non-linear estimators (kernel and kraskov) showed expected results for all the idealized systems, their free tuning parameters are to be tested for consistent results. Moreover, these methods are sensitive to the available time series length.

A real world example case study involving the dynamics between the Indian and Pacific oceans revealed a physically consistent bi-directional IE. However, unexpected IE was detected in the example of North Atlantic and European air temperatures indicating hidden drivers. Though IE provides insights into system dynamics, the availability of time series length and the system at hand must be carefully taken into account before inferring any possible interpretations of the results.

Quantifying the IE from El-Niño southern oscillation (ENSO) and Indian Ocean Dipole (IOD) to the Indian Summer Monsoon Rainfall (ISMR) with the observational and reanalysis data sets revealed that both ENSO and IOD are synergistic predictors for the inter-annual variability of the ISMR over central India i.e., the monsoon core region. Though the investigated three Global Climate Models (GCM) could not reveal the underlying IE dynamics of ENSO, IOD, and ISMR, a Regional Climate Model (RCM) simulation downscaling one of the GCMs with realistic large scale signals across the lateral boundaries showed good agreement with the observations.

Evaluating a coupled regional climate modeling system driven by two differ-

ent global data sets with IE estimators revealed significant differences between the process chains linking the north-west Mediterranean sea surface temperatures, evaporation, wind speed, and the Vb-cyclone induced precipitation over Danube, Odra, and Elbe catchments in the historical period (1951-2005). Detailed investigation revealed that the north-west Mediterranean Sea in the coupled regional simulation driven by ERA-20C reanalysis corresponded to the Vb-cyclone precipitation over the three catchments while no such correspondence is noted in the EC-EARTH driven simulation. This discrepancy is attributed to the inheritance of the simulation biases from GCM into the RCM. In the future period (1965-2099), no significant changes in the processes are noted from the simulation.

Overall, this thesis used IE estimators in investigating the underlying dynamics of climate system and climate models. The estimators proved useful in providing insights into climate system dynamics assisting in a process based climate model evaluation.

Kurzzusammenfassung

Das Klimasystem ist ein typisches Beispiel für ein komplexes dynamisches System, das aus verschiedenen Subsystemen besteht, die durch Massen-, Impuls- und Energieaustausch auf verschiedenen räumlichen und zeitlichen Skalen miteinander interagieren. Ziel dieser Arbeit ist es, die Wechselwirkungen zwischen den Subsystemen aus der Perspektive des Informationsaustauschs (engl. Information Exchange, IE) zu erkennen und zu quantifizieren. Zu diesem Zweck werden aus der Informationstheorie abgeleitete Maße für den Informationsfluß untersucht und auf die verfügbaren Klimadaten bestehend aus Beobachtungen, Reanalysen, globalen Klimamodellen und regionalen Klimamodellen angewendet. Insbesondere zeigt die vorgelegte Arbeit die Nützlichkeit informationstheoretischer Methoden für die prozessorientierte Klimamodellbewertung.

Aus den Konzepten der Informationstheorie abgeleitete Methoden wie die Transferentropie, der sogenannte Informationsfluss, sowie deren lineare und nichtlineare Schätzverfahren werden zunächst getestet und auf idealisierte zweidimensionale Systeme angewendet. Die Ergebnisse liefern eine Abschätzung für die Richtung und Größe des Informationsflusses, was Einblicke in die zugrundeliegende Dynamik bietet. Wie erwartet sind die linearen Schätzer zwar robust für lineare Systeme, versagen aber bei nichtlinearen Systemen. Obwohl die nichtlinearen Schätzer (Kernel und Kraskov) robuste Ergebnisse für alle idealisierten Systeme zeigen, müssen ihre freien Tuning-parameter auf konsistente Ergebnisse getestet werden. Außerdem sind diese Methoden empfindlich gegenüber den Eigenschaften der verwendeten Zeitreihen (z.B. der Zeitreihenlänge). Die nichtlinearen Schätzer ergaben einen physikalisch konsistenten bidirektionalen Informationsaustausch zwischen dem Indischen und dem Pazifischen Ozean. Am Beispiel der Lufttemperaturen im Nordatlantik und in Europa wurde jedoch ein unerwarteter Informationsaustausch zwischen beiden Regionen festgestellt, der auf versteckte Einflussfaktoren hinweist. Obwohl die Informationstheorie wichtige Einblicke in die Systemdynamik gewährt, müssen die Länge der verfügbaren Zeitreihen und das jeweilige System sorgfältig berücksichtigt werden, bevor mögliche Interpretationen der Ergebnisse abgeleitet werden können.

Die Quantifizierung des Informationsaustauschs von El-Niño und der Southern Oscillation (kurz ENSO) und dem Dipol des Indischen Ozeans (IOD) auf den Indischen Sommermonsunregen (IMSR) mithilfe von Beobachtungs- und Reanalysedaten ergab, dass sowohl ENSO als auch IOD synergetische

Prädiktoren für die inter-annuelle Variabilität des ISMR über Zentralindien, d.h. der Monsunkernregion, sind. Obwohl die drei untersuchten globalen Klimamodelle (GCM) die zugrundeliegende Dynamik von ENSO, IOD und ISMR nicht aufzeigen konnten, zeigte eine einzelne regionale Klimamodell-Simulation (RCM), bei der eines der GCMs mit realistischen großräumigen Signalen entlang der lateralen Grenzen regionalisiert wurde, eine gute Übereinstimmung mit den Beobachtungen. Dieses Ergebnis unterstreicht die Nützlichkeit der informationstheoretischen Schätzer für die prozessorientierte Evaluierung von Klimamodellen.

Die Bewertung gekoppelter regionaler Klimamodellsysteme mit informationstheoretischen Maßen ergab signifikante Unterschiede zwischen den Prozessketten, die die Meeresoberflächentemperaturen des nordwestlichen Mittelmeers, die Verdunstung, die Windgeschwindigkeit und den von Vb-Zyklonen verursachten Niederschlag über den Einzugsgebieten von Donau, Oder und Elbe im historischen Zeitraum (1951-2005) verbinden. Detaillierte Untersuchungen ergaben, dass das nordwestliche Mittelmeer in der gekoppelten regionalen Klimasimulation auf der Grundlage der ERA-20C-Reanalyse den Vb-Zyklonenniederschlag über den drei Einzugsgebieten beeinflusst, während in der EC-EARTH-Simulation keine derartigen Zusammenhänge festgestellt wurden. Diese Diskrepanz wird auf die Vererbung der Fehlern in der Klimasimulation vom GCM auf das RCM zurückgeführt. In der zukünftigen Periode (1965-2099) werden keine signifikanten Änderungen der Prozesse in der Simulation festgestellt.

Insgesamt hat diese Arbeit gezeigt, dass die Schätzer für den Informationsaustausch zusätzliche Einblicke in die zugrundeliegende Systemdynamik liefern und damit eine prozessbasierte Klimamodellbewertung unterstützen.

Acknowledgements

Completing my Ph.D. thesis would not have been possible without the support from my supervisors and colleagues. Firstly I thank my supervisor, Prof. Bodo Ahrens, for introducing me to the information entropy world. Discussions with him provided many new insights and ideas to work upon. His constant support and efforts to fund my Ph.D period are duly acknowledged. Without my co-supervisor Dr. Cristina Primo, I would not have finished my tasks on time. Her help and ideas while developing codes in R were insightful and valuable.

I express my sincere gratitude to my senior colleagues Dr. Shakeel Ashraf, and Dr. Naveed Akthar. Their help during my first days in Frankfurt was helpful. I should specifically mention Dr. Naveed Akthar's help while I struggled to compile and run the coupled regional climate model. I am also thankful to my colleagues Dr. Anika Obermann Hellhund, Stephan Herzog, Dr. Erwan Brisson, Nora Leps, Dr. Amelie Hoff (Krug), Christopher Purr, Danny Risto, Dr. Fanni Kelemen, and Dr. Stamen Dolaptchiev for the fruitful discussions. Thanks to my collaborators Dr. Silje Soerland, Dr. Andreas Prien, and the COMSO-CLM members for their help during publication and conference discussions. Special mention to my dear friends Amelie and Christopher for their help with the German translations. I thank Prof. Joaquim Pinto, KIT for his support and encouragement.

I should also thank our technical staff Timo Keber for assisting me with the model data, data transfer, etc. The technical team from Goethe HLR high-performance computing, Lichtenberg High-Performance Computing center, a big thank you for helping me solve technical problems with climate models. Thanks to Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main for funding my initial phase of research.

I thank my family and friends back in India for their constant support and love. Finally, I thank my wife Anusha for her support, specially during my stressful times with Ph.D work.

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Nomenclature

Acronyms

CMI	Conditional Mutual Information
CORDEX	Coordinated Regional Downscaling Experiments
COSMO-CLM	Consortium for Small-scale Modelling in Climate Mode
ECMWF	European Center for Medium-range weather forecast
ENSO	El-Niño Southern Oscillations
ERA-20C	ECMWF twentieth century reanalysis
GCM	Global Climate Models
GPCC	Global Precipitation Climatology Centre
HadISST	Hadley Centre Sea Ice and Sea Surface Temperature dataset
IE	Information Exchange
IF	Information Flow
IOD	Indian Ocean Dipole
ISMR	Indian Summer Monsoon rainfall
MI	Mutual Information
NAO	North Atlantic Oscillation
NEMO	Nucleus for European Modeling of the Ocean
NWMS	North-Western Mediterranean Sea
PDF	Probability Density Functions
PID	Partial Information Decomposition
RCM	Regional Climate Models
SST	Sea Surface Temperature
TE	Transfer Entropy
TRIP	Total Runoff Integrating Pathways

Deutsche Zusammenfassung

Die Erde ist ein komplexes, dynamisches System, welches aus nichtlinearen Wechselwirkungen zwischen seinen Subsystemen besteht. Diese Subsysteme umfassen die Atmosphäre, Biosphäre, Kryosphäre. Hydrosphäre, Lithosphäre und Anthroposphäre. Sie interagieren auf verschiedenen räumlichen und zeitlichen Skalen durch Austausch von Masse, Impuls und Energie. Um zu verstehen, welche Prozesse das Klima der Erde bestimmen, ist es nötig, diese Wechselwirkungen zu erkennen und zu quantifizieren.

Zusätzlich zu umfassenden Beobachtungsnetzwerken, stellen Wetter- und Klimamodelle ein hilfreiches Werkzeug dar, um das Klima der Erde in der Vergangenheit, der Gegenwart und der Zukunft zu verstehen. Bedingt durch den technischen Fortschritt der letzten Jahrzehnte haben sowohl Klimamodelle als auch Beobachtungssysteme beträchtliche Datenmengen produziert.

Methoden wie Korrelationsanalyse, Regression, empirische orthogonale Funktionen und zeitversetzte Kreuzkorrelation werden häufig genutzt, um Wechselwirkungen zwischen Subsystemen des Klimasystems auf Basis von Modell- oder Beobachtungsdaten zu verstehen. Ein klassisches Beispiel hierfür war der Versuch von Sir Gilbert Walker den Niederschlag des indischen Sommer Monsuns durch Korrelations- und Regressionsanalyse vorherzusagen (Taylor 1962). Das Indian Meteorological Department benutzt zusätzlich zu dynamischen Wettervorhersagemodellen bis heute multivariate Regressionsanalyse (Rajeevan et al. 2007).

Obwohl die Korrelationsanalyse ein nützliches Werkzeug ist, um lineare Wechselwirkungen zu analysieren, ist sie nicht in der Lage die Richtung der Wechselwirkungen, also Ursache und Wirkung, zu bestimmen. Zeitversetzte Korrelationsanalyse hat den Nachteil sensitiv auf Autokorrelation zu reagieren, was zu unklaren Signalen führen kann (Runge et al. 2014). Wechselwirkungen zwischen Subsystemen können aus der Perspektive des Informationsaustuschs (engl. Information Exchange, IE) zwischen den Subsystemen betrachtet werden. Mithilfe dieser Methode sollte sich nicht nur der statistische Zusammenhang zwischen Subsystemen beschreiben lassen, sondern auch ein tieferes Verständnis für die physikalischen Prozesse und die Dynamik des Gesamtsystems erreicht werden. Kumar & Gupta (2020) erörtern, dass informationstheoretische Ansätze es ermöglichen, die Komplexität und das Verhalten von dynamischen Systemen besser charakterisieren, während physikalische Gesetze der Entwicklung des Systems wichtige Grenzen setzen. Diese Arbeit zielt darauf ab, Informationstheorie auf Klimadaten anzuwenden, um die zugrundeliegenden

Systemdynamik aufzudecken. Insbesondere versuchen wir unsere Methodik auf die Evaluierung von Klimamodellen auszuweiten, um so ihre Nutzbarkeit zu untersuchen. Nach unserem Wissen sind dies die ersten Versuche, diese Methodik umfangreich auf die Evaluierung von Klimamodellen anzuwenden.

Diese Arbeit untersucht die folgenden Fragen:

- Kann die zugrundeliegende Dynamik idealisierter und komplexer Abstraktionen des Klimasystems mithilfe informationstheoretischer Methoden robust abgeschätzt werden?
- Bieten diese Methoden zusätzliche Erkenntnisse bei der Evaluierung von regionalen Klimamodellen, vor allem in Bezug auf die Wechselwirkungen zwischen ENSO, IOD und dem indischen Monsunniederschlag (ISMR)?
- Wie verändert sich die Prozesskette zwischen Mittelmeer und Vb-Niederschlag in regionalen Erdsystemmodellen in Abhängigkeit von den Antriebsdaten?

Um die erste Frage zu beantworten, benutzen wir Methoden wie die axiomatische Transfer Entropie und den sogenannten "first principle-based information flow" zur Quantifizierung des Informationsaustauschs. Da die zugehörigen Schätzverfahren noch nicht umfassend untersucht wurden, haben wir sowohl nichtparametrische Schätzer wie "transfer-entropy (TE)-binning", "TE-kernel" und "TE k-nearest neighbor" als auch parametrische Schätzer wie "TE-linear" und "information flow (IF)-linear" in idealisierten, zweidimensionalen Testfällen zusammen mit ihrer Abhängigkeit von der Stichprobengröße getestet. Anschließend wurden diese Schätzverfahren auf zwei Klimaphänomene angewandt: die Kopplung des Indischen Ozeans mit dem pazifischen Ozean und die Kopplung der Nordatlantischen Oszillation mit der Lufttemperatur in Europa.

Die Ergebnisse zeigen, dass die parametrischen Schätzer, IF-linear und TE-linear, den einseitigen und den beidseitigen Informationsfluss im Falle eines idealisierten linearen Systems erkennen und verlässlich quantifizieren können, jedoch nicht in nicht-linearen Systemen. Unter den nichtparametrischen Schätzern liefern TE-kernel und TE-kraskov plausible Ergebnisse. Allerdings mussten dafür Parameter eingestellt werden, um konsistente numerische Ergebnisse zu erhalten. Außerdem wurde eine langsame Konvergenz des Informationsaustauschs mit dem TE-Kraskov Schätzer festgestellt. Aufgrund dessen kommen wir zu dem Schluss, dass die beiden plausiblen, nichtparametrischen Schätzer gemeinsam angewandt werden sollten und ihre Implementierung sorgfältiger Einstellungen bedarf, um quantitative Aussagen über die Wechselwirkungen in nicht-linearen Systemen zu ziehen.

Für realistische Klimaanwendungen zeigten die parametrischen und die zuverlässigen nichtparametrischen Schätzer einen erheblichen bidirektionalen Informationsaustausch zwischen dem Indischen und dem Pazifischen Ozean. Darüber hinaus stimmten die Zeitverzögerungen des signifikanten Informationsaustauschs zwischen dem Pazifik und dem Indischen Ozean mit der vorhandenen Literatur überein, was auf eine Verbindung zwischen physikalischen Prozessen und der Dynamik des Informationsaustauschs hinweist. Angesichts der Beschränkungen von TE und IF-linear kann jedoch die Möglichkeit eines versteckten Einflusses durch ein anderes System in der indo-pazifischen Kopplung nicht ausgeschlossen werden. Im zweiten Beispiel der Klimaanwendung, d. h. der NAO und der europäischen Lufttemperatur im Winter, zeigten diese Schätzer einen unrealistischen, bidirektionalen Informationsaustausch, was auf den Einfluss eines versteckten Treibers hindeutet.

Zur Evaluation regionaler Klimamodelle, insbesondere der Wechselwirkungen zwischen IOD, ENSO und dem indischen Sommermonsunregen, setzen wir Methoden wie den Informationsaustausch aus zwei Quellen ein. Zur Veranschaulichung der Konzepte und der Quantifizierung des Informationsaustauschs zwischen zwei Quellen und einem Ziel verwenden wir idealisierte Testfälle, die sowohl aus linearen als auch aus nichtlinearen, dynamischen Systemen bestehen. Unsere Ergebnisse zeigen, dass diese Systeme eine Netto-Synergie aufweisen (d. h. der kombinierte Einfluss von zwei Quellen auf ein Ziel ist größer als die Summe ihrer Beiträge), selbst bei unkorrelierten Quellen sowohl in linearen als auch in nichtlinearen Systemen.

Als nächsten Schritt untersuchten wir den Informationsaustausch zwischen ENSO und IOD zu ISMR IE in verfügbaren Beobachtungen, Reanalyse-Datensätzen und in drei Global Climate Model (GCM)-Simulationen, die auch mit dem Regional Climate Model (RCM) dynamisch verfeinert wurden. Die Ergebnisse der Beobachtungen und Reanalysedaten deuten darauf hin, dass sowohl IOD als auch ENSO die jährliche Variabilität der ISMR in den meisten Teilen des indischen Subkontinents beeinflussen. Insbesondere weisen IOD und ENSO über Zentralindien, der Kernregion des Monsuns, eine positive Nettosynergie auf, wohingegen sie über dem südlichen Teil Indiens redundante Informationen liefern. Darüber hinaus teilen sich ENSO und IOD redundante Informationen (negative Nettosynergie) im südlichen Teil des indischen Subkontinents.

Für die Modellevaluation haben wir drei CMIP5-GCMs - MPI-ESM-LR (Stevens et al. 2013) Nor-ESM-M (Bentsen et al. 2013) und EC-EARTH (Hazeleger et al. 2010) - mit dem nicht-hydrostatischen regionalen Klimamodell COSMO-crCLM Version v1-1 dynamisch verfeinert. Das COSMO-crCLIM ist eine beschleunigte Version des COSMO-Modells (Fuhrer et al. 2014) im Klimamodus (Leutwyler et al. 2016, Rockel et al. 2008), das für viele regionale

Klimasimulationen über Europa verwendet wurde. Die RCM-Simulation hat eine horizontale Auflösung von 0, 22° (d.h. 25km) und 57 vertikale Ebenen und verwendet einen Zeitschritt von 150s. Die Konfiguration der Modellsimulation folgte dem CORDEX-Rahmen, d.h. es wird ein historischer Zeitraum von 1950-2005 und das Business-as-usual-Szenario (RCP8.5) von 2006-2099 simuliert. In dieser Arbeit haben wir jedoch nur den historischen Zeitraum untersucht.

Die Ergebnisse zeigen, dass MPI-ESM-LR keinen synergetischen Informationsaustausch über dem indischen Subkontinent aufweist, während in Nor-ESM-M die IOD und ENSO über dem Westen Indiens gemeinsame Informationen liefern. EC-EARTH zeigt weniger Nettosynergie über dem indischen Subkontinent. Insgesamt weichen die Ergebnisse des IE-Austauschs von den Beobachtungen ab, was für alle drei GCM-Simulationen gilt. Die IE-Muster über dem indischen Subkontinent für die verfeinerten RCM-Simulationen zeigten eine Nettosynergie in Zentralindien und gemeinsame Informationen in Südindien in der verfeinerten Nor-ESM-M-Simulation. In der verfeinerten MPI-ESM-LR- und der verfeinerten EC-EARTH-Simulation sind solche Muster jedoch nicht vorhanden.

Dies stimmt mit den Ergebnissen der GCM-Simulationen überein, bei denen festgestellt wurde, dass die Nor-ESM-M-Simulation die ENSO- und IOD-induzierten anomalen Niederschlagsstrukturen besser nachbildet als die beiden anderen GCMs. Obwohl alle COSMO-crCLM-Simulationen dieselbe Physik und Dynamik aufweisen, konnte nur die verfeinerte Nor-ESM-M-Simulation realistische IE-Muster nachbilden.

Die besseren Ergebnisse in der regionalisierten Nor-ESM-M-Simulation können auf realistischere großräumige Informationen aus der GCM-Simulation zurückgeführt werden, wie z. B. den Feuchteflusstransport während verschiedener Phasen von ENSO- und IOD-Ereignissen. Eine genauere Untersuchung der Anomalien des Feuchteflusstransports ergab, dass sich die Muster der MPI-ESM-LR- und EC-EARTH-GCM-Simulationen stark von denen der Reanalyse unterscheiden und folglich falsch dargestellt werden. Eine bessere Replikation der Anomalie der Feuchtigkeitsflüsse in der Nor-ESM-M GCM-Simulation während ENSO und IOD könnte auf eine bessere Simulation der großräumigen Zirkulationsmuster, wie die Walker- und Hadley-Zirkulationen, zurückzuführen sein, da die SST besser dargestellt wird als in den beiden anderen GCM-Simulationen.

Was die RCM-Simulationen anbelangt, so zeigten die Ergebnisse ähnliche Anomalien der Feuchtigkeitsflüsse im Vergleich zu den treibenden GCM-Simulationen, wobei die regionalisierte Nor-ESM-M die regionalisierten MPI-ESM-LR und die regionalisierten EC-EARTH Simulation übertrifft. Diese Ergebnisse deuten darauf hin, dass ein realistisches großskaliges Signal aus den

GCM-Simulationen (z. B. der Feuchtetransport und die SST-Anomalien) für ein RCM von wesentlicher Bedeutung ist, um die GCM-Ergebnisse in Bezug auf die ISMR-Variabilität zu verbessern. Wenn das großskalige Signal aus dem GCM nicht korrekt ist und falsche Feuchtigkeitsflüsse an den seitlichen Grenzen des RCM aufgezwungen werden, werden die regionalisierten Ergebnisse beeinträchtigt.

Um die dritte und letzte Frage der Arbeit zu beantworten, wendeten wir die Informationsaustauschschätzer an, um die Prozessketten zu verstehen, die das nordwestliche Mittelmeer und die Vb-Ereignis-Niederschläge in den Donau-, Elbe- und Oder-Einzugsgebieten in den regionalen gekoppelten Atmosphäre-Ozean-Klimasimulationen verbinden. Das gekoppelte regionale Klimamodell besteht aus dem COSMO-CLM als Atmosphärenkomponente, dem Nucleus for European Modeling of the Ocean (NEMO) als Ozeankomponente über dem Mittelmeer (NEMOMED12) und TRIP für die Flussabflussberechnung. Diese einzelnen Komponenten sind durch einen Koppler namens OASIS miteinander verbunden.

In diesem Abschnitt der Arbeit wurden zwei gekoppelte regionale Klimamodell-Simulationen verwendet, nämlich das COSMO-CLM–NEMOMED12–TRIP für den Zeitraum 1951-2099 unter kontinuierlicher Verwendung des EC-EARTH GCM als Antriebsdaten mit dem RCP-8.5-Szenario für den zukünftigen Zeitraum (2006-2099) bei 0,11° (≈ 12 km) horizontal und die ECMWF-Reanalyse des zwanzigsten Jahrhunderts (ERA-20C) als gekoppelte regionale Klimasimulation mit demselben Aufbau. Die Leistung der ERA-20C-Simulation bei der realistischen Nachbildung der Vb-Zyklonereignisse und der damit verbundenen Niederschläge wurde bereits in der Studie von Krug et al. (2022) berichtet und analysiert. Daher ist die heruntergerechnete ERA-20C-Simulation eine Referenz für die Validierung der heruntergerechneten, von EC-EARTH simulierten Vb-Zyklonenereignisse und der damit verbundenen Niederschläge im historischen Zeitraum.

Bei der Häufigkeit der Vb-Zyklone wurde eine gute Übereinstimmung zwischen der GUF-Auswertung und den historischen GUF-Simulationen festgestellt. Darüber hinaus zeigten die Vb-Zyklonenzugdichte und -intensität in Bezug auf den minimalen Zyklonenzentraldruck sowie die Rangfolge der Vb-Zyklonenniederschlagsanomalien eine gute Übereinstimmung zwischen der GUF-Auswertung und den historischen Simulationen. Eine unbedeutende Zunahme der Vb-Zyklonenhäufigkeit um 1,8 % bis zum Ende des 21. Jahrhunderts wurde in der GUF-Zukunftssimulation festgestellt. Die Änderungen der zukünftigen Vb-Niederschlagsanomalien über den drei Einzugsgebieten waren ebenfalls nicht signifikant.

In der GUF-Bewertungssimulation entsprachen die SST-, Verdunstungs- und Windgeschwindigkeitsanomalien im nordwestlichen Mittelmeerraum

der Rangfolge der Vb-Zyklonenniederschlagsanomalien. Eine solche Übereinstimmung wurde in der EC-EARTH-gesteuerten historischen Simulation nicht festgestellt. Trotz Ähnlichkeiten im Modellaufbau (gleiche regionale Atmosphäre/Ozean-Modellkomponenten und -aufbauten über dem Mittelmeer) und guter Übereinstimmung in der Vb-Zyklonenhäufigkeit, -intensität und -niederschlag zwischen der GUF-Auswertung und der historischen GUF-Simulation unterscheiden sich die Zustands- und Prozessketten des nordwestlichen Mittelmeers im Hinblick auf den Informationsaustausch.

Die Unterschiede zwischen den Simulationen könnten auf das Auftreten von Simulationsverzerrungen zurückgeführt werden, die vom treibenden EC-EARTH GCM übernommen wurden, z.B. zu kalte Oberflächen- und Meeresoberflächentemperaturen über dem Mittelmeer im Vergleich zum ERA-20C Forcing. Das Downscaling von EC-EARTH3 (neueste Version von EC-EARTH), das eine geringere Verzerrung der Oberflächenlufttemperaturen und des SST (Döscher et al. 2021) aufweist, könnte zu einem besseren Verständnis der Zustands- und Prozessketten beitragen, die das nordwestliche Mittelmeer und die Vb-Zyklonenniederschläge in historischen und zukünftigen Zeiträumen verbinden.

Zusammenfassend lässt sich sagen, dass die in dieser Arbeit eingehend getesteten Schätzer für den Informationsaustausch dabei helfen, die regionalen Klimasimulationen prozessorientiert auszuwerten und die zugrunde liegende Systemdynamik aufzuzeigen. Diese Bewertungsmethode hilft der regionalen Klimagemeinschaft, die Modellleistung weiter zu verstehen und zu verbessern sowie den geeigneten globalen Modellantrieb zu wählen. Die Robustheit dieser Schätzer für den Informationsaustausch stellt jedoch aufgrund ihrer Empfindlichkeit gegenüber der verfügbaren Zeitreihenlänge und den freien Abstimmungsparametern noch eine Herausforderung dar und bedarf daher weiterer Forschung. Darüber hinaus wurden die Methoden genutzt und angewandt, um die Wechselwirkungen zwischen einer einzelnen oder zwei Quellen und einem Ziel zu quantifizieren. Das Klimaphänomen besteht jedoch aus hochdimensionalen Wechselwirkungen, so dass neue und effiziente Schätzer für den Informationsaustausch entwickelt werden müssen, um mehrdimensionale Wechselwirkungen aufzudecken.

List of Contributing Peer-Reviewed Publications

- ***Pothapakula, P.K.***, Primo, C., Ahrens, B. (2019). **Quantification of information exchange in idealized and climate System applications.** Entropy 2019, 21, 1094. https://doi.org/10.3390/e21111094

  (This article was featured as cover story and issue cover for Entropy volume 21, issue 11, 2019 and also appeared as one of the highly accessed article in Entropy website for Dec, 2019)

- ***Pothapakula, P.K.***, Primo, C., Soerland, S., Ahrens, B. (2020). **The synergistic impact of ENSO and IOD on the Indian Summer Monsoon Rainfall in observations and climate simulations - an information theory perspective**. Earth System Dynamics. https://doi.org/10.5194/esd-11-903-2020

- ***Pothapakula, P. K.***, Krug, A., Obermann-Hellhund, A., Keber, T., Ahrens, B. (2022). **Vb-cyclones and associated Mediterranean Sea state in regional coupled climate simulations: evaluation and projection** (submitted, Earth System Dynamics (ESD))

Part I

BACKGROUND

Chapter 1

Introduction

Earth is a complex dynamical system consisting of non-linear interactions among its sub-components namely the atmosphere, biosphere, cryosphere, hydrosphere, lithosphere, and the anthroposphere. These components interact at various spatial and temporal scales through mass, momentum, and energy exchange. Detecting and quantifying these interactions assists in understanding the processes which govern our earth's climate.

In addition to the large observational network, climate models based on the governing fluid equations are very useful tools in understanding the past and future climate state. With the recent advances in science and technology, an enormous amount of data is produced with the help of climate models on large computing facilities, ground-based, and satellite-based observational systems.

Methods such as correlation, regression analysis, empirical orthogonal functions, and time-lagged cross-correlations were often used on the observational and climate model data to understand the interactions among sub-components of earths climate. A classical example was the attempt by Sir Gilbert Walker to predict the Indian Summer Monsoon rainfall with correlations and multiple regression equations (Taylor 1962). The Indian Meteorological Department still uses multiple regression analysis for the long range prediction of the Indian Monsoon in addition to the dynamical weather prediction models (Rajeevan et al. 2007).

Though correlations are useful tools in analyzing the linear interactions, they do not detect the direction of interactions i.e., drive and response components. Often the time-lagged correlations were used for directionality, they are sensitive to the auto-correlations which result in obscure detection of directional interactions (Runge et al. 2014). The interactions among sub-components can be contemplated as information exchange (IE) between them. This IE should ideally reveal not just the statistical interdependence between these components, but a physical representation of the processes adding a new dimension to understanding the system dynamics. Kumar & Gupta (2020) argued that the information-theory-based approaches enable to better characterize the complexity and emergent behavior of the dynamical system, while the laws of physics provide important constraints on the evolution of the system behavior.

This thesis aims at exploring the information exchange methodology applied to the available climate data in an attempt to unravel the underlying system

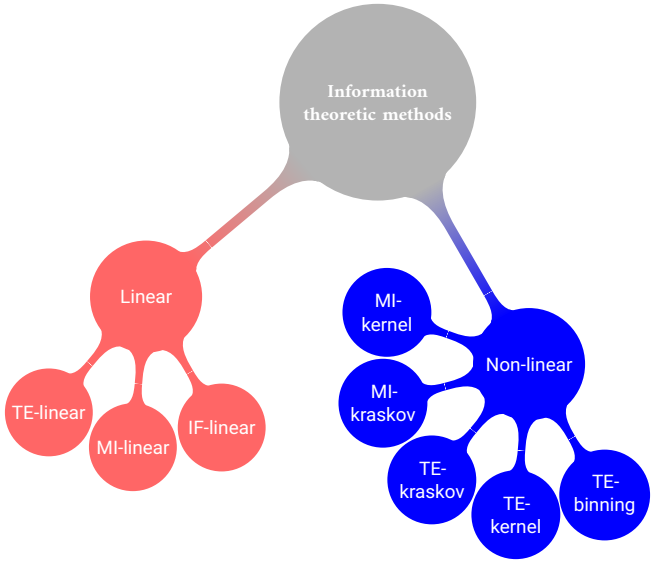


Figure 1.1. Various information theory methods and their estimators developed and used in this thesis

dynamics. Specifically, we aim to extend our methodology to climate model evaluation so as to investigate the usefulness of these methods. To the best of our knowledge these are the very first steps in applying this methodology extensively to climate model evaluation.

1.1 Introduction to Information Theory

In this sub-section, we discuss the fundamentals of the information theory, various methods derived from it, and their estimation techniques.

The concept of information entropy was first proposed by Shannon (2001), in his work "The Mathematical Theory of Communication" which connected with optimal coding and error-free communication. However, the information entropy was later adapted to many fields such as neurosciences, earth sciences, climate sciences, etc.

The mathematical formulation of Shannon Entropy for a random variable X is given as

$$H(X) = -\sum_{x} p(x) \log p(x),$$

where $p(x)$ represents the probability of an individual state for the random variable X. The summation goes through all the possible states of the random

variable X. The entropy is often expressed in bits if the logarithm base 2 is used, else in nats if a natural logarithm is used.

The information entropy can be understood as the average uncertainty or surprise about the outcome of the random variable X. For example, consider a fair coin with equal probability of heads and tails. Such a coin toss has 1 bit of entropy. Lets assume that the coin was not fair, lets say the probability of getting a head is $\frac{1}{4}$ and tail is $\frac{3}{4}$, the entropy of such a coin would be 0.81.

If an uncertainty about the random variable X is reduced by the knowledge of an other variable Y, then both the random variables X and Y share common information about each other. This can be mathematically represented as Mutual Information (MI),

$$MI_{xy} = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$$

The MI between the random variables X and Y is greater than zero when there exists any mutual dependence between each other regardless of their nonlinear dependence. Mutual information was used in the study by Knuth et al. (2013) to study the interactions between the cloud cover and the sea surface temperature anomalies around the globe. Their study observed high mutual information exchange values over the central Pacific indicating the dependence of the two variables over that particular location. The MI was also used in the climate model verification by the study of Ahrens & Walser (2008).

Though MI is a useful metric, it is a symmetric quantity, meaning, it does not contain any dynamical directional information. Though time-lagged mutual information can provide insights into the system dynamics, however, due to the common input from the history of the targeted random variable it could produce spurious unrealistic results. Hence, Schreiber (2000) introduced another important information theory metric called Transfer Entropy (TE) in an attempt to detect and quantify the asymmetric interactions among the subsystems.

The TE is based on the transitional probabilities explicitly incorporating the underlying dynamics of a complex system. The difference between the MI and TE is based on the notion that the former method relies on the static probabilities while the latter is based on the transitional probabilities. The TE measures the deviation between the transitional probabilities based on the generalized Markovian property of order k, $p(x_{n+1}|x_n, ..., x_{n-k+1})$ and $p(x_{n+1}|x_n, ..., x_{n-k+1}, y_n, ..., y_{n-l+1})$ of the subsystems X and Y.

In a detailed investigation, Schreiber (2000) experimentally applied TE to idealized spatiotemporal systems and then to a real-world application namely, to a bivariate physiological time series. The TE with kernel estimator for the first idealized system i.e., a one-dimensional lattice of a uni-directional coupled

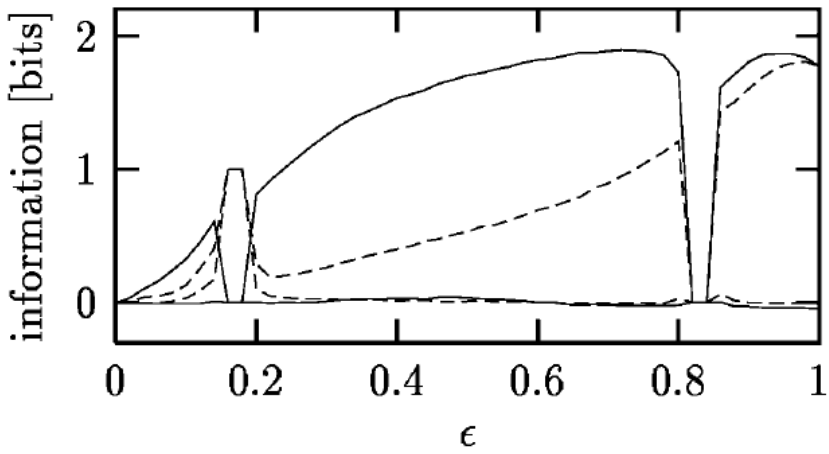


FIGURE 1.2. Transfer Entropy in Ulam maps from X_1 to X_2 and vice versa (in solid lines) with increasing coupling coefficient ϵ and time delayed Mutual Information (in dotted lines). Figure source: Schreiber (2000)

maps showed expected uni-directional information exchange showcasing the asymmetric relation.

With a more complicated idealized system, the Ulam map $f(x) = 2 - x^2$ with non-small coupling, the TE between the two systems X_1 and X_2 produced reliable asymmetric results compared to the time-lagged mutual information. Figure 1.2 shows the information exchange in bits along with the coupling coefficient ϵ for both the TE and MI.

The dynamics of the Ulam coupled map in Figure 1.2 undergo several bifurcations with changing coupling coefficient ϵ. At a coupling coefficient $\epsilon = 0.18$, the asymptotic state is of temporal and spatial period two. At this particular coupling coefficient information is neither exchanged nor produced between the sites. The values as derived from the time-lagged MI and TE are 1 and 0 respectively. The TE thus reveals the true system dynamics, unlike the time-lagged mutual information, as there is no information exchange. Moreover, applying the TE to the real world application revealed a dominant information exchange from heart rate to breath rate than vice versa unlike the mutual information revealing the dynamics between the heart and the breath rate of sleeping human suffering from sleep apnea. These results from Schreiber (2000) motivated many branches of science to adopt TE as a methodology for estimating the driving dynamics of complex systems.

The Transfer Entropy is widely applied in the field of neurosciences, especially in brain research. For example, Vicente et al. (2011) applied TE to test for effective connectivity to electrophysiological simulation data and magne-

toencephalography (MEG) observational recordings in a simple motor task. They concluded that the TE estimations proved reliable in quantifying the non-linear interactions where linear methods were hampered. In the field of Earth System Sciences, Ruddell & Kumar (2009) applied TE for quantifying the directionality and the time scale of information flow between pairs of ecohydrological variables using observed time series data. When the TE is applied to the ecohydrological system in a healthy peak growing season state and during a severe drought, they concluded that the process network during drought is substantially decoupled. These two studies were classical examples of the application of TE to practical applications.

The TE is also applied earlier in the field of climate science. For example, Runge et al. (2012) used the TE estimations to the daily mean sea level pressure anomalies in the winter months at four locations in eastern Europe. Their results suggested a south-eastward flow of entropy, which is physically consistent to the dynamics of lower and upper atmosphere. The information exchange between the four locations in the eastern Europe when estimated from the MI showed unrealistic physical links when compared to the TE estimations. Their results motivated many further studies in applying the TE to the climate data, e.g., the drivers of recent temperature variability (Bhaskar et al. 2017) possible relation between the Earth's magnetic field and climate (Campuzano et al. 2016) and, analyzing changes in the complexity of climate with radiation data (Delgado-Bonal et al. 2020).

Another information theory methodology was proposed by San Liang & Kleeman (2005) in an attempt to derive information exchange from fundamental principles, unlike TE which was axiomatically framed. From here after this method is referred an information flow. Considering two sub-systems X and Y, the information flow from Y to X is quantified by the difference between the marginal entropy evolution of X, i.e., $\frac{dH_X}{dt}$ and the entropy evolution of X excluding the influence of Y, i.e., $\frac{dH^*_X}{dt}$. To compute the information flow, the marginal evolution of the probability density functions needs to be estimated. However, without the knowledge of the system dynamics, estimating these quantities is a challenge. To overcome this challenge, San Liang (2014) linearized the governing equation for information flow and proposed a maximum likelihood estimator for estimating information flow.

The maximum likelihood information flow estimator is straightforward to apply to time-series data and does not involve knowing the system dynamical equations. This methodology is applied between the climate variables and the surface mass balance over the Antarctica region to investigate the dynamical processes affecting the surface mass balance in climate model simulations during the historical period (Vannitsem et al. 2019). The study by Tawia Hagan et al. (2019) applied the information flow to the soil moisture and near-surface tem-

peratures over China using the European Center for Medium-range weather forecast ERA-Interim dataset. Their results revealed an expected dominant information flow signal from the soil moisture to air temperature in the month of spring. Finally, the information flow is used in analyzing the causal interactions between the global radiative forcing and the global mean surface temperature anomalies. Various radiative forcing parameters such as aerosols, carbon dioxide concentrations, and natural variability such as solar cycle, and volcanoes are checked for causality with the global surface temperatures. Their results revealed dominating significant information exchange from the global CO_2 concentrations to air temperatures, indicating the role of anthropogenic forcing in changing the climate.

In this thesis, the information exchange methods described above are developed and tested rigorously on idealized test cases before applying them to climate application and climate models. While the information exchange methods are very useful, their estimation is very challenging. More details of the estimation techniques are described in chapter 2. However, Figure 1.1 provides an overview of various estimators which estimate the MI, TE and information flow in both linear and non-linear forms. In an attempt to explore various information exchange methods and furthermore apply them on climate data, we ask the following specific questions,

1.1.1 Research Question– 1

The information theory methods though useful in revealing the system dynamics, their estimation is a challenge. Especially, the non-parametric estimation of TE. Most of the literature mentioned in section 1.1 relied on a single estimation technique of TE while applying to the datasets, i.e., binning technique. Furthermore, the TE and information flow methodologies are derived from two different principles, hence their performance in unraveling the system dynamics on the idealized and climate data is yet to be investigated in detail. Moreover, the availability of climate data is limited and this poses a challenge in applying these methods to detect robust climate interactions.

The first scientific question of the thesis is the following,

- Do the IE estimators provide robust and reliable result's to quantify the system dynamics in idealized and climate system applications?

To answer this question, in chapter 3 a detailed investigation of various TE estimators shown in Figure 1.1 applied on linear and non-linear idealized systems and two large-scale climate teleconnections is discussed. Furthermore, the sensitivity of these estimators on time series length is also investigated.

1.2 Information theory as a tool for regional climate model evaluation

Climate models based on fundamental fluid equations of mass, momentum, and energy are widely used in simulating and understanding the current and future climate. Furthermore, they comprise of various components such as ocean, atmosphere, land, ice, etc. Global Climate Models (GCM) simulate the climate for the whole globe at various resolutions from hundreds of kilometers to tens of kilometers. However, with increased resolution comes increased computational and data storage costs. Hence there is a trade-off between the computational costs and the desired horizontal and vertical resolution.

Regional Climate Models (RCMs) simulate the regional climate to provide higher spatial and temporal characteristics of a specific region of interest by dynamically downscaling a GCM of course resolution (Giorgi 2019) as depicted in the Figure 1.3. The data from the RCMs are typically used for the long-term planning and adaptation by the impact studies modeling groups, national climate assessment reports, and strategies (Sørland et al. 2021). The GCMs provide large-scale information to the RCMs through the lateral and surface boundary conditions. The RCM then integrates the governing equations with its physics and dynamics to simulate the regional climate characteristics by accounting for the sub-GCM grid-scale forcings and processes eg., the orography, convection, etc.

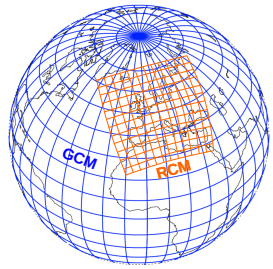


Figure 1.3. Schematic depiction of a global climate model grid and overlapping regional climate model grid explaining dynamical downscaling.

Typically, an RCM receives large-scale signals from the downscaling GCMs through the lateral boundaries every 6 hours. Generally, the wind components, temperature, water vapor, and surface pressure are provided to the lateral boundary area. To allow a smooth transition from the GCM at the RCM lateral boundaries relaxation techniques (Davies & Turner 1977) are used at the interface which is called a buffer zone. The information flows from the driving

GCM into the RCM through these buffer zone and a dynamical equilibrium is expected to reach after an initial spin-up between the large-scale GCM information and the internal dynamics within the RCM.

1.2.1 *Research Question–2*

Although the regional climate within the simulated domain is a result from the RCM physics and dynamics, the influence of driving GCM through the large scale forcing and surface boundary condition has an impact on the RCM simulation. A realistic lateral and surface boundary information flowing into the RCM is essential for an accurate regional climate simulation.

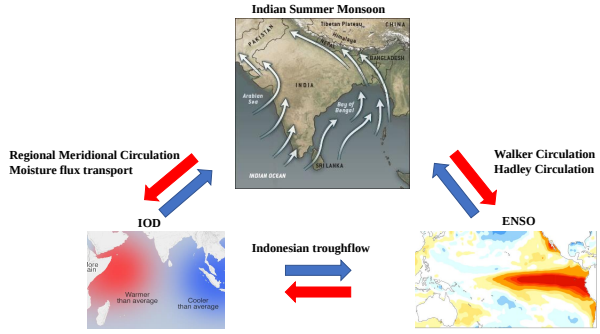


FIGURE 1.4. Schematic depiction of interactions between IOD, ENSO and the Indian Summer Monsoon.

The ENSO over the Pacific region is known to influence the ISMR through the changes in the Walker and Hadley Circulation while the IOD influences the monsoon rainfall through the changes in the regional circulation and moisture flux (Ashok et al. 2004). A representation of the interaction between ENSO, IOD and ISMR is shown in Figure 1.4. The example of the large scale signals, IOD, ENSO and their influence on the IMSR poses a good example for the research question–2 of the thesis.

- Could we quantify IE flowing from the GCM lateral boundaries into the RCM for a specific simulated climate phenomenon i.e., the dynamics of Indian Ocean Dipole (IOD), the El-Niño Southern Oscillations (ENSO) and the regionally simulated Indian Summer Monsoon rainfall (ISMR)?

An answer to this question can provide insights on the influence of large scale signals from the GCM on the RCM simulated climate, specifically the

influence of large scale signals, the IOD, ENSO and the regionally simulated ISMR. Chapter 4, discusses the quantification of the information exchange from the large scale phenomenon between the IOD, ENSO and ISMR from observational data and then in the GCM-RCM model chains.

1.3 Information exchange in coupled regional climate modeling systems

Most of the regional climate model simulations use stand-alone atmosphere, ocean, or land components. While such systems add value in capturing important regional climate information, most often they miss crucial feedback between the sub-components (ocean, atmosphere, or land). For example most of the coordinated regional downscaling experiments called CORDEX use stand-alone regional climate models (Sørland et al. 2021, Jacob et al. 2020). Recently there were efforts in coupling various components in the RCMs to incorporate various feed-backs for simulating a realistic climate (Somot et al. 2008). For example Akhtar et al. (2018) coupled the regional atmospheric model Consortium for Small-scale Modelling in Climate Mode (COSMO-CLM) to a regional climate model Nucleus for European Modeling of the Ocean (NEMO) over the Mediterranean Sea. This system is also coupled with a river runoff model named Total Runoff Integrating Pathways (TRIP) to close the water cycle over the Mediterranean region. A schematic figure describing such a regional climate system is shown in Figure 1.5. The sub-components in this system are coupled with each other through a coupler named OASIS which exchange fluxes between these components at regular intervals. More details of the simulations and the model set-ups are described in chapter 2.

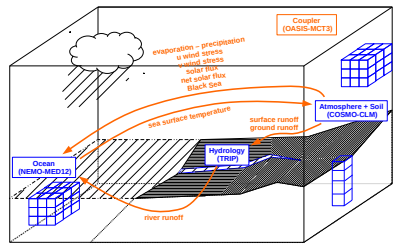


Figure 1.5. Schematic depiction of regional coupled COMSO-CLM + NEMO-MED + TRIP system along with various coupling variables.

The added value of such a coupled regional system in simulating the precipitation, air temperatures and extreme events such as Vb-cyclones were reported in several studies Akhtar et al. (2019), Kelemen et al. (2019), Primo et al. (2019),

Krug et al. (2022). However, a detailed investigation on the process chains connecting variables is yet to be investigated.

1.3.1 Research Question–3

To investigate the process chains and thereby evaluate the coupled regional climate model, we apply information exchange estimators to these coupled simulations. Specifically, we ask the following important third question of this thesis,

- Does the process chain connecting the coupled Mediterranean Sea and the Vb-cyclone precipitation change in the coupled RCM simulations when driven with different driving data?

Chapter 5 discusses answer to this specific question extensively.

1.4 Outline of the thesis and flow chart

The current chapter of the thesis consisted of a brief overview of the topic, motivation and three important scientific questions to be addressed. In chapter 2, a brief description of the methods, modeling systems, and various data sources used in this thesis are described. Thereafter, chapter 3 investigates various information exchange estimators on idealized and real-world applications such as the Indo-Pacific ocean coupling. In chapter 4 the IE estimators are applied for GCM-RCM model chain evaluation specifically on the interactions between IOD, ENSO and IMSR. In chapter 5, we apply the IE estimators to identify and quantify the process chain in coupled regional climate models, specifically the information exchange between the Mediterranean Sea and the Vb-cyclone precipitation. Finally, conclusions and outlook are given in chapter 6. A comprehensive overview of this thesis is shown in Figure 1.6.

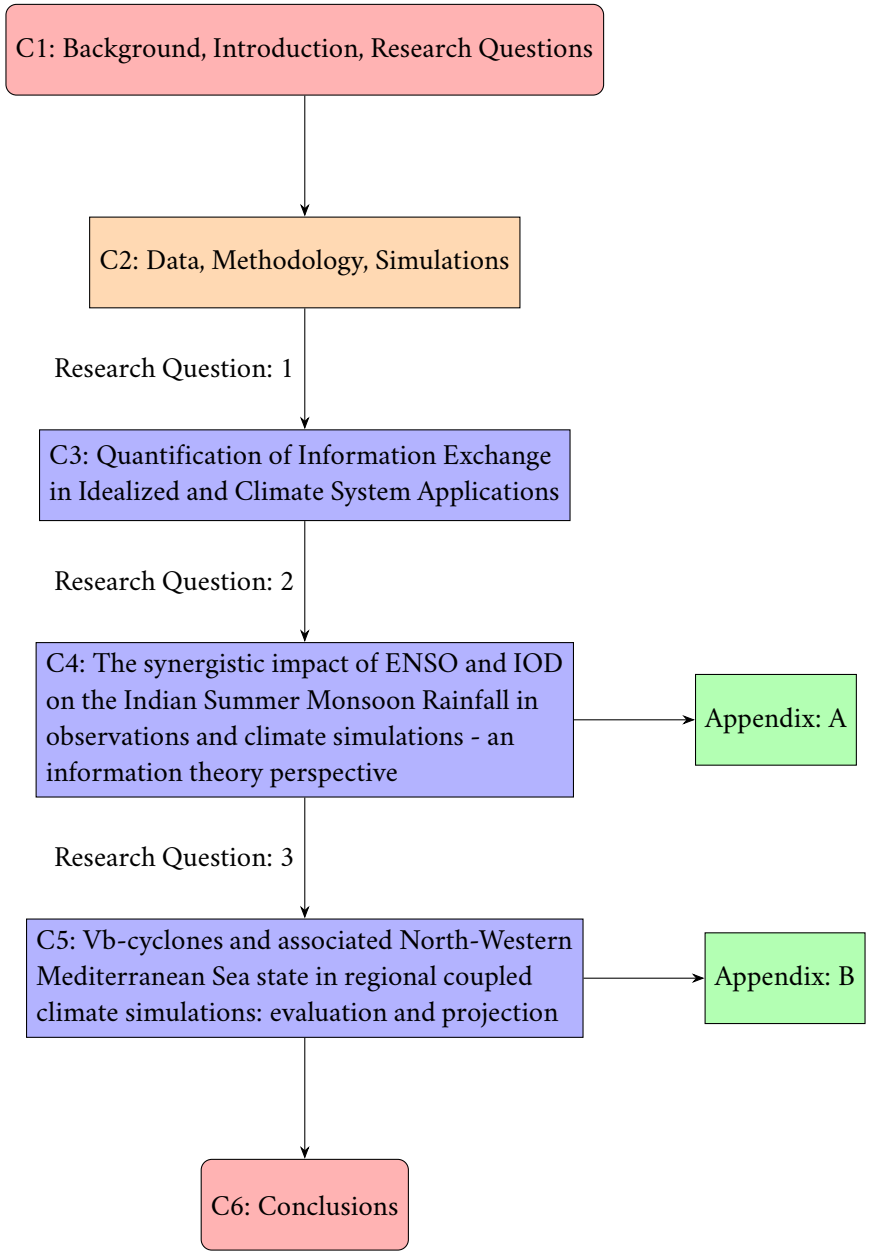


Figure 1.6. Flowchart for an overview of the thesis content. The characters (C1–C6) represents the chapters in this thesis.

Chapter 2
Methods, Modeling Systems and Data

This chapter provides an overview of various estimators derived from information theory, regional and global climate modeling systems, their model configuration and set-ups along with the data used for validation.

2.1 Information theory methods and estimators

One of the fundamental metrics used in this thesis is so-called mutual information (MI). The MI is used to detect the shared information between two independent processes (Shannon 2001, Thomas & Joy 2006). For random variables X and Y, MI measures the divergence between the joint distribution $p(x, y)$ and the product of the marginal distributions $p(x)$ and $p(y)$:

$$\mathrm{MI}_{xy} = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

Calculation of MI on real-valued data sets depends much on the estimation technique utilized for calculating the joint and marginal probability density functions (PDF). A simple technique for calculating the PDFs is to divide the continuous data into a random number of bins. This approach is commonly referred to as binning. Depending on the base of the logarithm, the units of MI are measured in either nats (natural logarithm) or in bits (logarithm to the base 2).

A more elegant way of constructing a PDF is to use a box step kernel for the estimation of the joint and marginal probabilities (Lizier 2014, Kantz & Schreiber 2004). For example, the joint probability distribution for $\hat{p}(x, y)$ is calculated as:

$$\hat{P}_r(x_n, y_n) = \frac{1}{N} \sum_{n'=1}^{N} \Theta(|(x_n - x_{n'}), (y_n - y_{n'})| - r),$$

where the norm corresponds to the maximum distance in the joint space of x, y. The r represents the kernel width.

Another method that uses the adaptive resolution technique for the calculation of MI uses k-nearest neighbors was proposed by Kraskov et al. (2004). Here the resolution of the effective PDF is higher in the regions where the data

are numerous, while the resolution is relaxed at less amount of data. Initially, for each point in the two dimensional space i.e., $z_i = (x_n, y_n)$, its neighbors' distance d=max$||z_i - zj||$ is calculated. While any norm could be used, in this thesis, the maximum norm is used. Next, the total number of points that fall within the range d in all the marginal spaces is counted. Thereafter, the number of points in each of the marginal spaces is substituted in the following equation:

$$\mathrm{MI}_{x,y} = \Psi(K) - < \Psi(n_x + 1) + \Psi(n_y + 1) > - \Psi(N),$$

where Ψ denotes the digamma function, K is the number of nearest neighbors, and the angle brackets indicate averaging over all the points, n_x, n_y that fall within the range d in the marginal spaces.

The linear version of the MI involves calculating the co-variance terms,

$$I(X;Y) = \frac{1}{2} \log \left[\frac{\det \Sigma(X)}{\det \Sigma(X|Y)} \right].$$

where $\Sigma(X)$ refers to the co-variance of X and $\Sigma(X|Y)$ refers to the partial co-variance of X with respect to Y.

An other important high dimensional metrics which estimates the asymmetric information exchange is the transfer entropy (TE) proposed by Schreiber (2000). Similar to the MI, the TE can be estimated with the above mentioned techniques, namely TE-binning, TE-kernel, TE-Kraskov and TE-linear.

The TE from subsystem Y to X through discrete binning approach is calculated by decomposing the individual joint entropies as:

$$\mathrm{TE}_{y->x} = H(x_n^k, y_n^l) - H(x_n^{(k+1)}, y_n^l) + H(x_n^{(k+1)}) - H(x_n^k), \tag{2.1}$$

where k and l correspond to the embedding dimensions of X and Y, respectively.

Similarly the MI could be estimated by box kernel by calculating the joint probabilities through kernel step function and then substituted in the equation (2.1).

Finally, the TE from the nearest neighbours is calculated with the following equation,

$$\mathrm{TE}_{y->x} = \Psi(K) + < \Psi(n_{x_n} + 1) - \Psi(n_{x_{n+1}}, n_{x_n}) - \Psi(n_{x_n}, n_{y_n}) >,$$

where K is the number of nearest neighbors, Ψ denotes the digamma function, while the angle brackets indicate averaging over all the points, $n_{x_{n+1}}, n_{x_n}$ and n_{y_n} are the number of points that fall within the range d in the marginal spaces. Similarly, this formula can be extended for higher embedding spaces.

So far, the methods proposed above only quantify the mutual information or the information exchange by transfer entropy between two subsystems. In

chapter 4, we compute the information exchange between two sources together with a target system. For example, the MI between the two sources Y and Z and a target X is expressed as

$$I(X;Y,Z) = \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)}{p(x)p(y,z)},$$

where $p(x,y,z)$ is the three dimensional joint probability distribution of variables X,Y and Z, while other probabilities represent the respective marginal probability density functions.

From the study by Williams & Beer (2010), the mutual information, $I(X;Y,Z)$, that the two sources Y, Z share with target X decomposes into four parts as

$$I(X;Y,Z) = U(X;Y|Z) + U(X;Z|Y) + R(X;Y,Z) + S(X;Y,Z), \tag{2.2}$$

where $U(X;Y|Z)$ is the unique information shared by Y to X, $U(X;Z|Y)$ is the unique information shared by Z to X, $R(X;Y,Z)$ redundant information shared by both sources Y and Z together with X, and $S(X;Y,Z)$ synergistic information about X while knowing the states of Y and Z together. One of the limitations of the above equation is that no complete solution is reached among the scientific community to calculate the individual contributions of the terms. However, Barrett (2015) proposed a quantity known as net synergy as,

$$\begin{aligned} \Delta I(X;Y,Z) &= I(X;Y,Z) - I(X;Y) - I(X;Z), \\ &= S(X;Y,Z) - R(X;Y,Z). \end{aligned} \tag{2.3}$$

When $\Delta I(X;Y,Z) > 0$, synergistic information from two sources is greater than redundant information and vice versa. The ΔI provides a lower-bound for synergistic/redundant information. From here on, if $\Delta I(X;Y,Z) > 0$ we refer as net synergistic information and if $\Delta I(X;Y,Z) < 0$ we refer to as net redundant information.

This quantity is a very useful metric in understanding the information dynamics of two sources. For example, if there exists a net synergy between the two sources implies that the knowledge of them together reduces the uncertainty of the target variable than the sum of individual contributions. This metric is applied to the IOD, ENSO and Indian Summer Monsoon interactions in chapter 4.

2.2 Climate Modelling Systems

In this thesis, a non-hydrostatic regional climate model COnsortium for Small-scale Modelling (COSMO-CLM) was used as the atmospheric component for

the regional climate simulations (Rockel et al. 2008). The COSMO-CLM is designed for applications ranging from meso-β to meso-γ scales (Steppeler et al. 2003). The model primitive dynamic and thermodynamical equations are solved with a Runge–Kutta numerical time-stepping scheme (Wicker & Skamarock 1998) on a three-dimensional Arakawa-C grid (Arakawa & Lamb 1977). This COSMO-CLM grid is based on rotated geographical coordinates and a generalized, terrain-following height coordinate (Doms et al. 2005).

2.2.1 *Regional Climate Simulations over South Asia*

The simulations conducted in chapter 4 used an accelerated version of the COSMO model called the COSMO-crCLIM in climate mode (Fuhrer et al. 2014, Schär et al. 2020). This accelerated version was developed to conduct simulations on heterogeneous hardware architectures including the graphics processing units (GPUs) and multicore central processing units (CPUs).

The domain used for conducting the simulations is depicted in Figure 2.1. The domain covers the South Asia Monsoon region which includes the Indian Sub-continent, the Indian Ocean, the Himalayas, etc. The COSMO-crCLIM simulation has a horizontal resolution of 0.22° (i.e., 25km) with 57 vertical levels and is using a time step of 150s. The domain is a standard set-up that follows the framework of coordinated regional downscaling experiments (CORDEX), an international collaborative effort for regional climate dynamical downscaling (Giorgi et al. 2009, Gutowski Jr et al. 2016).

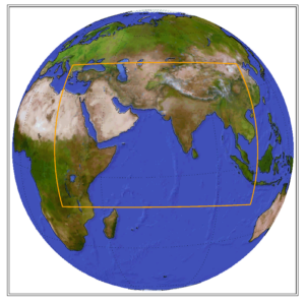


Figure 2.1. South Asia CORDEX domain used in the regional climate simulations with COSMO-crCLIM

The COSMO-crCLIM used a two-stream radiative transfer calculations based on (Ritter & Geleyn 1992). The convection is parameterized by (Tiedtke 1989), and the turbulent surface energy transfer and planetary boundary layer used the parametrization of (Raschendorfer 2001), with a four-category microphysics scheme which includes cloud, rainwater, snow, and ice (Doms et al.

2005). For the soil-vegetation-atmosphere-transfer TERRA-ML (Heise et al. 2006) was used with a modified groundwater formulation (Schlemmer et al. 2018).

The simulations were conducted continuously for the period 1950-2099. For the future period, i.e., 2006-2099 the business as usual future emission scenario (RCP8.5) was used for accessing the climate change signal.In chapter 4, only the information exchange between the IOD, ENSO, and the Indian Summer Monsoon Rainfall are investigated in the historical period. However, the results these simulations were validated and the climate change signal over the South Asia domain was discussed in the study by Sørland et al. (2021). A total of four simulations were conducted with similar settings, with one simulation driven by reanalysis data set and the remaining three simulations used three different GCM's as shown in Table 2.1

Table 2.1. Simulations conducted over South-Asia CORDEX domain @ 25km resolution

Driving Model	resolution	Time-period
ERA-Interim reanalysis (ERA-Interim)	$0.5° \times 0.5°$	1979-2005
Max Planck Institute ESM (MPI-ESM-LR)	$1.875 ° \times 1.875 °$	1950-2099
Norwegian Climate Centre ESM (Nor-ESM-M)	$2.5 ° \times 1.9 °$	1950-2099
SMHI, Sweden EM (EC-EARTH)	$1.125 ° \times 1.125$	1950-2099

2.2.2 *Coupled Regional Climate Simulations over Europe*

The simulations conducted in chapter 5 used coupled regional climate model with the COMSO-CLM 5-0-9 as the atmospheric component, a regional version of the Nucleus for European Modeling Ocean (NEMO) as the ocean component, the Total Runoff Integrating Pathway (TRIP) river runoff model connecting the rivers in the southern part of the domain to the Mediterranean Sea. The NEMO model was adapted over the regional Mediterranean Sea including a section of the Atlantic Ocean as a buffer zone for open boundary (Beuvier et al. 2012). The horizontal resolution of NEMO is about $1/12°$ ($\approx$ 6.5–8.0km in latitude and $\approx$ 5.5–7.5km in longitude) and 75 unevenly spaced z-levels in the vertical direction. The coupling of these components is done through a coupler OASIS3-MCT which exchange fluxes between these components at

regular intervals of time (Craig et al. 2017). Moreover, the coupler interpolates the coupling fields onto the respective grids before exchanging information to the sub-components. This system is extensively tested for varying resolutions and coupling intervals in the study by (Akhtar et al. 2018).

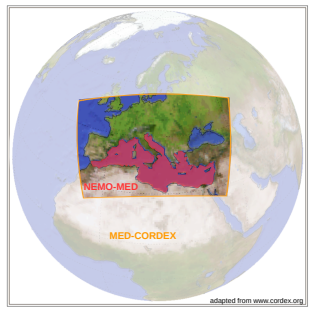


Figure 2.2. Med-CORDEX domain used in the coupled regional climate simulations with COSMO-CLM + NEMO-MED12 + TRIP

The horizontal resolution of COSMO-CLM was about 0.11° and used 40 vertical layers representing about 22.7 km of the atmospheric column. The domain used for conducting simulations is represented in Figure 2.2. The Mediterranean Sea represented in red color is actively coupled to the atmospheric component COSMO-CLM. The model configuration for COMSO-CLM follows similar settings as described in subsection 2.2.1. A total of three simulations were conducted over the Med-CORDEX domain and the details are provided in Table 2.2. The coupled regional climate simulations driven by EC-EARTH and ERA20CR were used in chapter 5.

2.3 Data used for validation

The following data sets were used as observation for real world climate applications or for validating the regional climate simulations in this thesis,

- **NOAA ESRL Niño 4 index** : The monthly time series of the Niño 4 index over the Pacific Ocean were obtained from NOAA ESRL for the time period of 1958–2010 corresponding to a total of 633 months (ESRL n.d.).
- **JAMSTEC IOD index** : The monthly time series of the Indian Ocean Dipole index over the Indian Ocean were obtained from JAMSTEC for the time period of 1958–2010 corresponding to a total of 633 months (JAMSTEC n.d.).

Table 2.2. Regional model simulations conducted over Med-CORDEX domain @ 12km resolution

RCM-Model	driving data	Time-period
Stand-alone COMSO-CLM	ERA-Interim	1979-2015
Coupled COMSO-CLM + NEMO-MED12 + TRIP	ERA-Interim	1979-2015
Coupled COMSO-CLM + NEMO-MED12 + TRIP	EC-EARTH	1950-2099
Coupled COMSO-CLM + NEMO-MED12 + TRIP + NEMO-BALTIC	ERA20CR	1901-2005

- **COBE-SST2** : The SST reanalysis data set consisting of Sea Surface Temperature and Ice was obtained from NOAA for the period 1958–2010 (Hirahara et al. 2014).
- **CRU air temperatures** : The near surface air-temperatures were obtained from the CRU database held at the British Atmospheric Data Centre, RAL, UK for the period of 1901 to 2016 (Harris et al. 2017).
- **HadISST 1.1** : The UK Met Office's Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST 1.1) was used as SST observational reference the Indian and the Pacific ocean (Rayner et al. 2003).
- **GPCC precipitation data set** : Monthly precipitation fields from Global Precipitation Climatology Centre (GPCC) was used for precipitation observational record for Indian Monsoon rainfall (Schneider et al. 2008).
- **APHRODITE precipitation data set** : A high-resolution data set, covering only the South Asia domain called Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE) at monthly intervals was used as a precipitation reference for the Indian Monsoon rainfall (Yatagai et al. 2012).
- **NCEP-NCAR reanalysis** : The rainfall, winds, and specific humidity reference were obtained from the National Center for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data set (Kalnay et al. 1996).

- **ERA-Interim reanalysis** : The rainfall reference for IMSR wass obtained from the ERA-Interim (Dee et al. 2011). This data set was also used as a driving GCM for south Asia regional climate simulations.
- **MERRA reanalysis** : MERRA reanalysis rainfall datasets (1980-2005) was used for one of the precipitation reference for Indian Monsoon (Rienecker et al. 2011).

Part II

RESULTS

Chapter 3

Quantification of Information Exchange in Idealized and Climate System Applications[1]

Abstract

Often in climate system studies, linear and symmetric statistical measures are applied to quantify interactions among subsystems or variables. However, they do not allow identification of the driving and responding subsystems. Therefore, in this study, we aimed to apply asymmetric measures from information theory: the axiomatically proposed transfer entropy and the first principle-based information flow to detect and quantify climate interactions. As their estimations are challenging, we initially tested nonparametric estimators like transfer entropy (TE)-binning, TE-kernel, and TE k-nearest neighbor and parametric estimators like TE-linear and information flow (IF)-linear with idealized two-dimensional test cases along with their sensitivity on sample size. Thereafter, we experimentally applied these methods to the Lorenz-96 model and to two real climate phenomena, i.e., (1) the Indo-Pacific Ocean coupling and (2) North Atlantic Oscillation (NAO)–European air temperature coupling. As expected, the linear estimators work for linear systems but fail for strongly nonlinear systems. The TE-kernel and TE k-nearest neighbor estimators are reliable for linear and nonlinear systems. Nevertheless, the nonparametric methods are sensitive to parameter selection and sample size. Thus, this work proposes a composite use of the TE-kernel and TE k-nearest neighbor estimators along with parameter testing for consistent results. The revealed information exchange in Lorenz-96 is dominated by the slow subsystem component. For real climate phenomena, expected bidirectional information exchange between the Indian and Pacific SSTs was detected. Furthermore, expected information exchange from NAO to European air temperature was detected, but also unexpected reversal information exchange. The latter might hint to a hidden process driving both the NAO and European temperatures. Hence, the limitations, availability of time series length and the system at hand must be taken into account before drawing any conclusions from TE and IF-linear estimations.

[1] Published as: *Pothapakula, P.K, Primo C, Ahrens B. (2019). Quantification of Information Exchange in Idealized and Climate System Applications. Entropy 2019, 21, 1094.* https://doi.org/10.3390/e21111094

3.1 Introduction

Complex dynamical systems consist of interacting subsystems. It is important to detect these interactions and quantify their strength to improve process understanding. Especially in the climate system, detecting the interactions could provide insights into system dynamics. These interactions can be contemplated as information exchanged or transferred among subsystems. Shannon (2001) introduced a mathematical theory for quantifying the information contained in the context of data compression and transmission. Recently, there has been a surge in applications of information theory in a wide range of fields, for example, uncertainty propagation (San Liang 2011, Tödter & Ahrens 2015, Kirchgessner et al. 2017), neurosciences (Schelter et al. 2006, Vicente et al. 2011), climate sciences (Stips et al. 2016, Bhaskar et al. 2017, Knuth et al. 2013), earth system sciences (Ruddell & Kumar 2009, Bennett et al. 2019, Gerken et al. 2019, Yu et al. 2019), turbulence research (Tissot et al. 2014), and networks and synchronization in dynamical systems (Boccaletti et al. 2002).

Often in climate studies, the relationship among subsystems is assessed with correlation analysis, empirical orthogonal functions, and linear regressions given time series. Among them, correlation analysis is a parametric method used in identifying linear interactions. For nonlinear interactions, mutual information, a nonparametric method (Ahrens & Walser 2008, Shannon 2001), is often used. Mutual information reveals any shared information between two subsystems. However, both correlation and mutual information are symmetric, i.e., they cannot distinguish between a drive and a response system. For a better understanding of the dynamics, detection of the directionality of interactions is essential. The time-lagged cross-correlation and time-lagged mutual information methods are frequently used for this purpose (Klein et al. 1999, Lanzante 1996). Runge et al. (2014) observed that these two methods are sensitive to autocorrelations which often obscure detection and quantification of the interaction mechanisms. For example, in his study, a spurious interaction between tropical east Pacific and the northern tropical Atlantic is detected at a time lag 3–6 months while cross-correlations are applied, while the information theoretic-based asymmetric measures detected an interaction between tropical east Pacific and the northern tropical Atlantic at a time lag of 1 month. This interaction time lag is physically consistent with the advection speed of Pacific–Atlantic Walker circulation.

For deeper insights into detecting drive and response linear interactions, Granger (1969) proposed a test based on the Wiener principle. According to the Granger test, X causes Y if the past of the system X assists in predicting the future of the system, Y. In a statistical sense, if the error variance of the optimal linear prediction of Y future state based on the past of X and Y has a

smaller error variance than considering the past of Y alone, then X causes, Y. A nonparametric method equivalent to Granger causality known as transfer entropy (TE) was proposed by Schreiber (2000) (for linear Gaussian systems, they are equivalent up to a factor of 2). The TE measures if any additional information is provided by the past of the source system, which assists in predicting the future state of the destination system. In other words, it measures the divergence or deviation between the entropy rates of the destination's own past and the past included from a source system. The TE, unlike mutual information, is an asymmetric quantity and hence, it can detect drive and response interactions. It is worth noting that, in nature, the interactions do not merely consist of driving and responding systems, but also systems which drive each other simultaneously with different interaction strength. Hence, any reliable asymmetric estimate, e.g., TE, should reveal this underlying behavior. Many information-theory-based methods were spawned based on a similar principle to that of TE, for example, momentary information transfer, Information Transfer to Y, and Information transfer to X (Runge 2015). These methods have their own applications and limitations.

The simplest estimation of TE uses a multivariate Gaussian model assuming linear interactions between the subsystems. This parametric estimation is hereafter referred to as TE-linear. While the parametric estimator TE-linear is straightforward to calculate, the nonparametric TE estimation is notoriously challenging. Some of the common nonparametric estimation techniques of TE in the literature include the binning, kernel density, and k-nearest neighbor. These estimators are sensitive to the parameter selection in their implementation, such as the bin width selection in the binning estimator, the kernel width in the kernel density estimator, and the number of nearest neighbors while applying the k-nearest neighbor estimator. Unfortunately, no clear consensus is reached among the scientific community on selecting these free parameters. As a result of this dependency, spurious detection of information exchange between the system components could arise. Regardless, TE has been widely used—for example, Bhaskar et al. (2017) applied TE in the identification of primary drivers of recent climate variability and quantified their influence on climate variability. Their results suggested that greenhouse gases are primary contributors to the recent climate variability. Campuzano et al. (2016) studied the information exchange between the South Atlantic anomaly and global sea level for the last 300 years using TE and concluded that larger information is exchanged from the south Atlantic anomaly to global sea-level rise than vice versa. However, these studies relied only on a single TE estimation technique (binning). It is still unclear if TE nonparametric estimations with the free parameters reliably produce numerically consistent estimations. Furthermore, these estimators are also sensitive to the length of the time series. For example,

robust kernel estimation asks for sufficient data. Hence, before applying TE to climate phenomena, we tested various estimators of TE along with their sensitivity on time series length with idealized systems where the system dynamics is expected or known.

San Liang & Kleeman (2005), realizing that information exchange could be derived rigorously rather than axiomatically, developed another method called it information flow (IF), which is derived from the first principles of information theory. In their framework, the information source and destination are abstracted as system components and thus derive the information flow between these dynamical components. The information flow between the source system Y and destination system X is equal to the difference between the time evolution of marginal entropy of X and entropy of X excluding the influence of, Y. To apply the information flow method, the time evolution of the marginal probabilities must be computed. This time evolution of the marginal probabilities, in turn, depends on the system dynamics. If the system dynamics is unknown, IF becomes difficult to apply. The information flow method was successfully applied to Heńon maps, the Rössler system, and truncated Burgers–Hopf with their respective system dynamics known (San Liang 2016). Unlike that of IF, the calculation of TE do not require system dynamics. Hence, given two climate time series, TE is straightforward to apply. San Liang (2014) proposed a simple and concise maximum likelihood estimator of IF for linear systems which is easy and straightforward to apply without system dynamics. This estimator is a very important result for the climate community as it bridges the gap between theory and real-world applications. From hereafter, this maximum likelihood estimator is referred to as information flow-linear (IF-linear). IF-linear has been successfully applied in detecting the causal structure between CO_2 and global temperatures (Stips et al. 2016), changing the relationship between the convection over the Western Tibetan Plateau and the sea surface temperature in the Northern Bay of Bengal (Vaid & Liang 2018) and forecasting the tropical cyclone genesis over the Northwest Pacific through identifying the causal factors in cyclone–climate interactions (Bai et al. 2018).

In this work, our aim was to apply information theory methods to detect interactions between climate phenomena. Moreover, we were provided with a limited amount of temporal series of climate data and with their dynamics unknown. Hence, we focused on IF-linear and TE methods. We wanted to find out if these methods are consistently able to detect the directionality of the interactions for climate phenomena. Having in mind that the nature of both methodologies differs, we first wanted to understand if this could lead to differences in detecting the information exchange from climate time series. On the other hand, TE methods are highly sensitive to the choice of free parameters and with time series length, which often might lead to brittle information

exchange detections. Thus, we initially checked if the different estimators (IF-linear, TE-linear, TE-binning, TE-kernel, and TE-kraskov) detect the directionality of the interactions using various temporal series lengths as realized by idealized systems, whose dynamics and directionality of the interactions is expected or known. These systems consist of uni- and bidirectional coupled linear and nonlinear systems. Thereafter, we applied these methods to the Lorenz-96 system (Lorenz 1996), which is known to mimic the mid-latitude atmosphere behavior. Finally, we experimentally applied these methods to (1) Indo-Pacific sea surface temperatures interbasin coupling and (2) the relation between North Atlantic Oscillation (NAO) and winter near-surface air temperatures over Europe. One of the limitations of this study is that we rigorously tested and applied various estimators to two-dimensional systems only. For a detailed and excellent review on the applications of TE on high dimensional interactions, refer to Runge (2018).

This paper is organized as follows. Section 3.2 comprises the background material for IF-linear, TE, and its estimation techniques such as TE-linear, TE-binning, TE-kernel, and TE k-nearest neighbor. In Section 3.3, the above mentioned methods are applied to uni- and bidirectional coupled linear and nonlinear systems, the Lorenz-96 system, and then to climate phenomena. Results are also discussed in this section. Finally, conclusions are drawn in Section 3.4.

3.2 METHODS

In this section, the basic concepts of information theory are discussed along with a brief introduction of information flow and transfer entropy. Estimation techniques of TE are also presented.

3.2.1 Transfer Entropy

Let $p(x)$ be the probability of a state for the random variable X. The Shannon entropy of X, $H(X)$, quantifies the amount of information needed to describe the random variable X (Shannon 2001):

$$H(X) = -\sum_{x} p(x) \log p(x),$$

where the summation goes through all states of the random variable X. The units of entropy are nats if a natural logarithm is applied; alternatively, it is often expressed in bits when the logarithm base is 2.

Let $q(x)$ be another probability distribution for the same random variable, and Kullback divergence measures the distance between the two probability distributions p and q. It is defined as:

$$D_{p||q} = \sum_x p(x) \log \frac{p(x)}{q(x)}.$$

Similarly, mutual information(MI) is the divergence between the joint distribution $p(x, y)$ of variable X and Y and the product of the marginal distributions $p(x)$ and $p(y)$:

$$\mathrm{MI}_{xy} = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

Mutual information quantifies the shared information between X and Y and thus is symmetric and lacks any direction of information exchange. Mutual information between X and Y while a third variable (Z) taken into account is given by conditional mutual information (CMI) as:

$$\mathrm{CMI}_{xy|z} = \sum_{x,y} p(x, y, z) \log \frac{p(x|y, z)}{p(x|z)}.$$

Transfer entropy is a special case of CMI, in which the conditioning is done on the past of Y inplace of, Z. Furthermore, the TE reflects the dependencies contained in the transitional probabilities which represent the dynamics of the system. The TE measures the deviation between the generalized Markov property, in other words, the deviation between the transitional probabilities $p(x_{n+1}|x_n, ..., x_{n-k+1})$ and $p(x_{n+1}|x_n, ..., x_{n-k+1}, y_n, ..., y_{n-l+1})$:

$$\mathrm{TE}_{y->x} = \sum_{x,y} p(x_{n+1}, x_n^k, y_n^l) \log \frac{p(x_{n+1}|x_n^k, y_n^l)}{p(x_{n+1}|x_n^k)}, \tag{3.1}$$

where k and l are the embedding dimensions of the destination and source variables, respectively. An important measure known as active information storage (AI) is given as:

$$\mathrm{AI} = \mathrm{MI}(x_n^k; x_{n+1})$$

which is used in approximating the embedding dimensions of the destination system. It measures how much of the information from the past of the X system (x_n^k) is observed to be in use in computing its next future observation (x_{n+1}). Here, MI refers to mutual information. If additional information from y_n assists in the reduction of the uncertainty in the future state of x_{n+1} given x_n, then there is an information transfer from Y to, X. Thus, TE quantifies the average information transfer from Y to X and similarly vice versa. It is important to note that the source Y can potentially influence the system X at various

interaction delays. A scanning approach for the largest information exchange values was employed by Wibral et al. (2013) to calculate the interaction delay between the source and destination systems. In this work, we adopted the same methodology to calculate the interaction delays for climate applications. Furthermore, the comparison of TE estimations only makes sense if the interaction delays in both directions are properly and independently reconstructed (Wollstadt et al. 2017). Hence, it is important to ensure the correct interaction delays are extracted. The TE, unlike MI and correlation, is an asymmetric measure. While the TE equations seem straightforward, its nonparametric estimation for continuous data is quite challenging. Coarse-graining the continuous data into discrete states is hard to interpret unless the measure converges when reducing the coarsening scale. The implementation of the binning estimator uses a bin width parameter upon which the estimates are highly sensitive. Similarly, the kernel density estimator relies on kernel width and the k-nearest neighbor estimator on the number of nearest neighbors. Unfortunately, no clear consensus is reached in optimally choosing these parameters. Hence, in the current study, we tested all the methods for numerical consistency before applying to climate applications. A brief introduction to these methods is given in the following sections.

Estimation of TE-Binning

For the estimation of TE, joint probability distributions are calculated from the underlying time series. Binning is one of the most straightforward approaches in the estimation of joint probabilities. Unfortunately, entropy estimations are highly sensitive to the number of bins chosen (i.e., bin width). In the literature, there exist numerous methods that describe the selection of an optimum number of bins for Gaussian distributions (Scott 1979, Freedman & Diaconis 1981) and suggestions for distributions without any underlying assumptions (Kang et al. 2017, Ruddell & Kumar 2009). In the current study, we used a more general method proposed by Knuth et al. (2013), known as generalized Knuth method with no underlying assumptions about the distribution. In this method, each of the N observed data points are placed into one of M-fixed width bins, where the number of bins is selected utilizing a Bayesian paradigm. The likelihood of the multidimensional data with volume V and probabilities π_i for the i^{th} bin is given by the multinomial distribution:

$$p(d|M,\pi) = (\frac{M}{V})^N \pi_1^{n_1} \pi_2^{n_2} ....\pi_M^{n_M},$$

where $d = [d_1, d_2..., d_N]$ denote N observed points, $n_1, n_2...n_N$ denotes the number of points in each sample, and $\pi = [\pi_1, \pi_2, .., \pi_M]$ denote the respective probabilities of the bin. The Dirichlet prior conjugate to the multinomial

likelihood function for the non-informative prior is given as:

$$p(\pi|M) = \frac{\Gamma(\frac{M}{2})}{\Gamma(\frac{1}{2})}[\pi_1, \pi_2, ..., \pi_{M-1}, (1 - \sum_{i=1}^{M-1} \pi_i)]^{-1/2}.$$

The posterior distribution of the bin probabilites from the Bayes theorem is:

$$p(\pi, M|d) \propto p(\pi|M)p(M)p(d|\pi, M).$$

Then, the optimum number of bins are chosen by the mode of the posterior distribution of:

$$\log p(\frac{M}{d}) = N \log M + \log \Gamma(\frac{M}{2}) - M \log \Gamma(\frac{1}{2}) - \log \Gamma(N + \frac{M}{2}) + \sum_{i=1}^{M} \log \Gamma(ni + \frac{M}{2}) + K,$$

where M is the number of bins, d is the observed data points, ni is the number of data sample in each bin, and K is a constant. For more details, refer to Knuth et al. (2013). After calculating the optimal number of bins M, TE is calculated by decomposing into individual joint entropies given as:

$$\mathrm{TE}_{y->x} = H(x_n^k, y_n^l) - H(x_n^{(k+1)}, y_n^l) + H(x_n^{(k+1)}) - H(x_n^k), \tag{3.2}$$

where k and l are the embedding dimensions of X and Y, respectively. An estimation bias for TE could arise due to the assumption of a uniform distribution within every single bin which corresponds to maximum entropy.

Estimation of TE-Kernel

This estimator uses the box step kernel Θ with $\Theta(x > 0) = 0$ and $\Theta(x < 0) = 1$ for the estimation of relevent joint probability distributions (e.g., $\hat{p}(x, y)$, $\hat{p}(x)$ and $\hat{p}(y)$). For example, the joint probability distribution $\hat{p}(x, y)$ is calculated as:

$$\hat{P}_r(x_n, y_n) = \frac{1}{N} \sum_{n'=1}^{N} \Theta(|(x_n - x_{n'}), (y_n - y_{n'})| - r),$$

where the norm corresponds to the maximum distance in the joint space and r is the kernel width. In simple terms, the resultant probability is the fraction of N values which fall within the kernel width r in the joint dimensional space. Here, r is the free parameter and the resultant probability is sensitive to the choice of r. Furthermore, the conditional probabilities are defined in terms of their respective component joint probabilities. These probabilities are substituted in Equation 3.1 to calculate TE. Kernel estimators are model-free (i.e., they do not assume parametric distribution). For more details about the estimator, refer to Kantz & Schreiber (2004), Goodwell & Kumar (2017) and information-theoretic toolkit from Lizier (2014).

Estimation of TE-K-Nearest Neighbor

Kraskov et al. (2004) introduced the k-nearest neighbor estimator which uses an adaptive binning strategy. This estimator uses the average distances to the k-nearest neighbor data points for the calculation of TE. This is a nonparametric estimation technique.

Transfer entropy from Y to X, TE_{yx} with embedding dimensions $k = 1$ and $l = 1$, is calculated as follows: For each point in the highest dimensional space given in Equation (3.2) (i.e., $z_i = (x_{n+1}, x_n, y_n)$), its neighbors' distance d=max$||z_i - zj||$ is calculated. While any norm could be used, in the current study, the maximum norm is utilized. The number of points that fall within the range d in all the marginal spaces is counted. Thereafter, the number of points in each of the marginal spaces are substituted in the equation below to calculate TE:

$$\mathrm{TE}_{y->x} = \Psi(K) + < \Psi(n_{x_n} + 1) - \Psi(n_{x_{n+1}}, n_{x_n}) - \Psi(n_{x_n}, n_{y_n}) >,$$

where K is the number of nearest neighbors, Ψ denotes the digamma function, while the angle brackets indicate averaging over all the points, $n_{x_{n+1}}$, n_{x_n} and n_{y_n} are the number of points that fall within the range d in the marginal spaces. Similarly, this formula can be extended for higher embedding spaces. Furthermore, this method enables for bias correction. From hereafter, this method is referred as TE-kraskov. For more details, refer to the information-theoretic toolkit of Lizier (2014).

Estimation of TE-Linear

The entropy for a continous random variable X is given as:

$$H(x) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx.$$

For the linear estimation of entropy, substituting the probability of a Gaussian distribution (Thomas & Joy 2006) in the above equation gives:

$$H(x) = \frac{1}{2} \log 2\pi e\sigma^2.$$

For a multivariate Gaussian model, the entropy is given as:

$$H(x) = \frac{1}{2} \log((2\pi e)^d |\Omega_x|),$$

where d is the number of dimensions, $|\Omega_x|$ is the determinant of the $d \times d$ covariance matrix $\Omega_x = \bar{x}x^T$, and the overbar indicates averaging. Furthermore, the TE is estimated as the sums and differences of the joint entropies given in Equation (3.2).

Assumptions in the Practical Estimation of TE

Before applying the TE estimators mentioned above, it is very important to observe their assumptions. While the TE estimations in the above sections involve the interactions between two subsystems X and Y, real-world climate applications often involve a tangle of higher dimensional interactions. In order to disentangle the sole interactions between X and Y, one needs to remove the influence of all the other interacting variables in the state phase. However, this can be computationally exhaustive and perhaps almost impossible. This unreal aspect was already mentioned in Granger (1969). Hence, while applying TE to retrieve the information exchange between two subsystems, there might be unobserved variables influencing the estimation. Hence, this limitation needs to be accounted for while applying TE. In the current study, we limited our detailed investigations to two-dimensional systems. Hence, a possible influence of unobserved variables on TE estimations cannot be ruled out. For a detailed review on the applications of TE on high dimensional interactions, refer to Runge (2018).

In the TE Equation (3.1), the deviation between the transitional probabilities is calculated with the embedding dimensions of the destination and source variables, k and l, respectively. The embedding dimension l of the source variable could be chosen at which the maximum information exchange takes place, while an ideal embedding k for the destination system should be as large as possible. Due to the computational complexity and a limited number of time series data available for climate applications, a minimum value for $k = 1$ could be chosen provided most of the information from the past of destination system is embedded within the minimum k. In the current study, this was verified through the calculation of active information of the destination system. Moreover, assumptions in the TE calculation involve stationarity, time aggregation, faithfulness, etc. An excellent review of these assumptions is available in Runge (2018).

3.2.2 Liang and Kleeman Information Flow

Consider a two-dimensional nonlinear system x_1 and x_2:

$$\frac{dx_1}{dt} = F_1(x_1, x_2, t),$$

$$\frac{dx_2}{dt} = F_2(x_1, x_2, t),$$

with randomness limited to its intial condition. The above equations follow the convention in physics which does not distinguish random and deterministic

variables. Let H_1 be the marginal entropy of x_1. The H_1 evolution of x_1 may be due internal mechanism of x_1 itself or subject to the influence of x_2. The later is the information flow from x_2 to x_1, which is of interest to us . If $\frac{dH_1^*}{dt}$ is the contribution of evolution of marginal entropy by x_1 itself, then:

$$\frac{dH_1}{dt} = \frac{dH_1^*}{dt} + T_{2->1},$$

where $T_{2->1}$ is the information flow from x_2 to x_1. Therefore, the information flow could be written as:

$$T_{2->1} = \frac{dH_1}{dt} - \frac{dH_1^*}{dt}.$$

The term $\frac{dH_1}{dt}$ can be calculated by Liouville equation. Thus, San Liang & Kleeman (2005) obtained the entropy evolution: $\frac{dH}{dt} = E(\nabla \cdot F)$, where E is the expectation operator. Based on this San Liang & Kleeman (2005) argued that the first term on the right hand side must be $\frac{dH_1^*}{dt} = E(\frac{\partial F_1}{\partial x_1})$, which was later proven by San Liang (2016). Substituting both terms in the above equation, San Liang & Kleeman (2005) argued that the information flow from x_2 to x_1 is equal to:

$$T_{2->1} = \frac{dH_1}{dt} - \frac{dH_1^*}{dt} = -E\Big(\frac{1}{\rho_1}\frac{\partial F_1 \rho_1}{\partial x_1}\Big),$$

where ρ_1 is the marginal probability density function of x_1. However, this heuristic argument was rigorously proven in San Liang & Kleeman (2007), San Liang (2008). The thus obtained information flow is asymmetric between x_1 and x_2. However, the above formalism is only for 2D deterministic systems. Moreover, for a 2D system with stochasticity involved, this formalism does not work. The Liang information flow rigorous formalism has undergone recent developments with higher dimensional systems and stochasticity involved. Consider a dynamical system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(t;\mathbf{x}) + \mathbf{B}(t;\mathbf{x})\dot{\mathbf{w}},$$

where $\mathbf{x}$ and $\mathbf{F}$ are n-dimensional vector, $\mathbf{B}$ is an $n \times m$ matrix, and $\mathbf{w}$ is an m-vector of standard Wiener process ($\dot{\mathbf{w}}$ is a vector of white noise). The rate of information flow from x_2 to x_1 for the above dynamical system is given by:

$$T_{2->1} = -\int_{R_n} \rho_{2/1}\frac{\partial(F_1\rho_2)}{\partial x_1}d\boldsymbol{x} + \frac{1}{2}\int_{R_n} \rho_{2/1}\frac{\partial^2(g_{11}\rho_2)}{\partial x_1^2}d\boldsymbol{x},$$

where $\rho_{2/1}$ is the conditional probability density function of x_2 on x_1, $\rho_2 = \int_R \rho dx_2$ and $g_{11} = \sum_{j=1}^m b_{1j}b_{1j}$. When $n = 2$, the equation reduces to:

$$T_{2->1} = -E\Big[\frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\Big] + \frac{1}{2}E\Big[\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\Big].$$

If the system dynamics F_1 and g_{11} are independent of x_2, then $T_{2->1} = 0$, which remarkably appears in the classical formalism. For systems with the dynamics unknown, the estimation of entropy evolution is a challenge. Hence, San Liang (2014), under the linear assumption, proposed a simple easy-to-use formula know as the maximum likelihood estimator of information flow. Given two series x_1 and x_2, for consistency with the formulae of TE mentioned above, we considered $x = x_1$ and $y = x_2$; the information flow maximum likelihood estimator or IF-linear from the system y to x is given by:

$$T_{y->x} = \frac{C_{xx}C_{xy}C_{y,dx} - C_{xy}^2 C_{x,dx}}{C_{xx}^2 C_{yy} - C_{xx}C_{xy}^2},$$

where C_{xx}, C_{yy} and C_{xy} are the covariances of x and y, while the subscript dx indicates time series derived from x which is formed as $\frac{x(n+k)-x(n)}{k.dt}$, with k some integers greater than or equal to 1. This easy-to-use formula bridges the gap between theory and real applications and has been successfully applied to real-world applications.

3.3 Results

In the current section, the above-discussed methods are tested for one-way and two-way coupled linear and nonlinear idealized systems. After testing various estimators, we applied them to the Lorenz-96 model which mimics midlatitude atmosphere behavior and finally to two important real-world climate phenomena.

3.3.1 *Applications to Idealized Systems*

Unidirectional Linearly-Coupled Autoregressive System

We considered a two-dimensional linear system x and y with the following governing (Knuth et al. 2013) :

$$\begin{aligned} y_{n+1} &= 0.5y_n + N(0,1), \\ x_{n+1} &= 0.6x_n + Cy_n + N(0,1), \end{aligned} \tag{3.3}$$

where $N(0,1)$ is Gaussian noise with zero mean and unit variance. The coupling coefficient C is varied from 0 to 1 with an increment of 0.1. The system was initialized with $(x_0, y_0) = (0,0)$. We integrated around 100,000 iterations and considered the last 5000 steps for detecting and quantifying the information exchange. Throughout this study, in all the idealized systems, a similar

number of iterations was followed. In real climate applications, the number of observations is limited and it is essential that the methods mentioned in Section 3.2 detect the direction of information exchange even with a limited number of data points. Hence, the total number of points was decreased from 5000 down to 200 time units to check the reliability of these methods. It is worth noting that this article does not aim at suggesting solutions for finite sample size effects; rather, we compared different estimators robustness against variations of the sample size. Equation (3.3) implies that the coupling coefficient C drives the information exchange from y to x. However, there is no exchange from x to y. Thus, any reliable method should reproduce this asymmetry. We also show the error bars representing two standard deviations representing the measure of uncertainty for the IF-linear and TE estimations. Furthermore, in order to choose the embedding dimensions for TE, we calculated the active information for the system x and y. For the source embedding dimensions for TE calculations, we chose $l = 1$, as the governing equations showed the past of source variable exchanges maximum information at time lag 1. The active information for the destination systems showed a minimal increase of AI from 0.2270 nats with time lag 1 to 0.2278 nats with 1 to 10 time lags. Hence, to reduce the computational complexity we chose the value of embedding dimension k to be 1. From the governing equations, the interaction time delay of 1 is chosen as the source transfers maximum information exchange at a time delay of 1. Before applying the TE-binning estimator, Knuth's method for detecting the optimal number of bins was applied to the data. However, as the data consist of various coupling strengths (C), the optimal number of bins varies for a particular coupling coefficient. For consistency, we kept the bin width constant throughout all the values of coupling coefficient C but allowed variations for different length of time series.

Figure 3.1 shows the information exchange from x to y for different coupling coefficients C and varying time series lengths. The IF-linear robustly measures a nonzero information exchange from y to x and zero information exchange from x to y for time series lengths $n \gtrsim 500$ time units. As expected from the governing equations, the information exchange increases from y to x with the coupling strength (Knuth et al. 2013). We also plotted the error bars which represent two standard deviations of IF-linear estimation for the respective time series length (see San Liang (2014) for details of the significance test). With time series of 200 time units, relying on the error bars, one can distinguish the asymmetry in the information exchange between systems x and y. The TE-linear estimator also shows the asymmetry in the interactions between x and y. The results are also stable with all the time series lengths. The error bar represents two standard deviations of 100 permuted surrogates for TE estimations (refer to Lizier (2014) for details into the significance test for

TE). As in the previous case, the strength of the association from y to x also increases with the increase in the coupling coefficient.

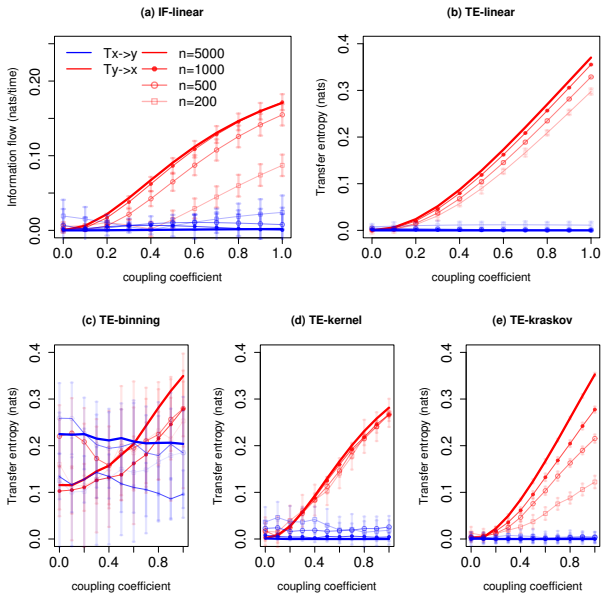


FIGURE 3.1. Information exchange in the unidirectional coupled linear system (Equation (3.3)) with various time series lengths (n) measured by (**a**) the IF-linear method (nats/time) and (**b**–**e**) with different variants of the TE measure (in nats). Error bars represent two standard deviations of the permuted surrogate samples.

Among the nonparametric estimators, TE-binning is able to reproduce the information exchange from y to x with a time length $n = 5000$ and $n = 3000$ time units (with a total number of 12 and 14 bins, respectively). However, it overestimates the TE from x to y. This overestimation may be attributed to the rough estimation of the joint pdf. Moreover, the assumption of a uniform bin width corresponds to biased entropy estimation. With the large overlapping error bars, it is difficult to distinguish the information exchange direction. The TE-kernel estimator (kernel width = 1) is also able to detect the asymmetry in the coupling between x and y and also the strength of the association. However, with a time series length $n \lesssim 500$ time units, spurious information exchange is detected from x to y. Nevertheless, the error bars from x to y with time series length $n \lesssim 500$ time units cross zero nats, and hence, no significant information exchange takes place. Furthermore, various kernel widths from 0.5 to 2 were tested, and the results are consistent between the kernel widths from 1 to 2. The better performance of TE-kernel might be attributed to the smoother estimation of the pdf when compared with the TE-binning. Similarly,

the TE-kraskov estimator (20 nearest neighbors) is able to show the one-way coupling and also the strength of the coupling with the time series $n \gtrsim 500$ time units. Below 500 time units, permutation surrogate improves to access the direction of information exchange. The TE-kraskov estimator ess tested between 4 to 60 neighbors, and the results are consistent from 20 to 60 nearest neighbors.

The estimations of information exchange are sensitive to the autocorrelation of the time series (Runge 2018). Hence, we calculated and compared the autocorrelations for various systems. For this system, the time lag-1 autocorrelation of the system x was around 0.5, and the autocorrelation of the system y was around 0.7. The presented results show that both parametric methods provide robust results along with the nonparametric TE-kernel and TE-kraskov estimators above the time series length of 500 time units. In addition, with permutation surrogates, even with time series length of 200 time units, the directionality of the information exchange is retrieved. The TE-binning method produced unreliable results.

Bidirectional Coupled Linear Autoregressive System

Often in climate systems, subsystems mutually exchange information. However, sometimes the interaction strength from one system might be stronger than the interaction in the opposite direction. Hence, it is very important that the methods mentioned in Section 3.2 not only detect the interactions but also quantify their relative strength. Therefore, as an example, we considered the system with governing equations:

$$\begin{aligned} x_{n+1} &= 0.1x_n + C_{yx}y_n + N(0,1), \\ y_{n+1} &= C_{xy}x_n + 0.1y_n + N(0,1), \end{aligned} \tag{3.4}$$

where $N(0,1)$ is Gaussian noise with zero mean and unit variance and $C_{yx} \in [0,1]$ and $C_{xy} = \frac{1}{2}C_{yx}$. We initialized the system with $(x_0, y_0) = (0,0)$. From the governing equations, a bidirectional information exchange exists and the relative information exchange from y to x is stronger than from x to y, since $C_{yx} = 2C_{xy}$. For the TE estimations, the AI values of this system are similar to those of the unidirectional coupled autoregressive system. Hence, for this system, we considered the embedding dimensions k and l to be 1. From the governing equations, the interaction time delay of 1 was chosen as the source transfers maximum information exchange at a time delay of 1.

The measured information exchange with different estimators for the system with governing Equation (3.4) is shown in Figure 3.2. The IF-linear shows a bidirectional information exchange. Further, it is able to quantify the strength of the information exchange between x and y realistically for all time series

length; in other words, as the coupling coefficients increases, the information exchange increases bidirectionally, and importantly, the strength of association from y to x is stronger than vice versa. This measured property would be very useful in climate applications to accurately quantify the strength of the associations among the subsystems. The TE-linear reproduced the bidirectional information exchange for all the time series lengths as well. Further, the strength of the information exchange is also accurately captured.

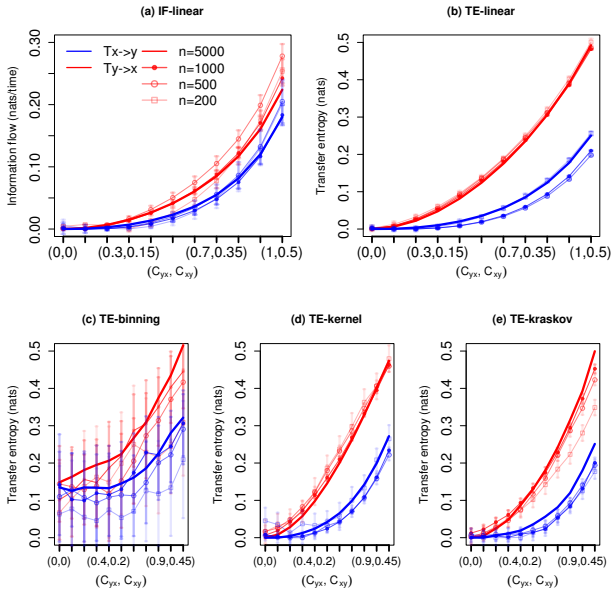


FIGURE 3.2. Information exchange in the bidirectional coupled linear system (Equation (3.4)) with various time series lengths (n) measured by (**a**) the IF-linear method (nats/time) and (**b**–**e**) with different variants of the TE measure (in nats). Error bars represent two standard deviations of the permuted surrogate samples.

The TE-binning estimator is able to reproduce the bidirectional information exchange with a time length of $n = 5000$ and $n = 3000$ time units (with a total number of 10 and 13 bins, respectively). However, with time series length less than 1000 time units, spurious detections are noted. Furthermore, with overlapping error bars, the detection of the directionality of information exchange from y to x and from x to y is difficult. The TE kernel estimator provides robust bidirectional information exchange (kernel width = 1) but shows spurious results for time series length of 200 time units, especially for weaker couplings. However, at weaker couplings, the error bars assist in detecting the direction of information exchange. Similarly, the TE-kraskov (10 nearest neighbors) estimator is able to show the bidirectional coupling and also the

strength of the information exchange for all time series lengths. The results are consistent between 10 to 60 nearest neighbors.

In summary, both the parametric methods produced stable results and also the nonparametric TE-kernel and TE-kraskov estimators showed consistent results for almost all time series lengths. The autocorrelations for this system were 0.2 and 0.4 with time lags of 1 and 2, respectively, for x and y. Unlike the previous example with unidirectional coupling, the results were relatively robust even with 200 points. This might be associated with the weaker autocorrelation magnitude in the bidirectional coupled system as compared to the unidirectional coupled linear system. This sensitivity of TE estimations on the autocorrelations was reported in Runge (2018).

Nonlinear Unidirectional Coupled Anticipatory System

Often in climate science, the relation among the subsystems is nonlinear, and it is very important to identify these associations and quantify their strength. Hence, in this section, we considered a special case of a nonlinear system called anticipatory system. Systems in which the response somehow predicts or anticipates the drive dynamics are known as anticipatory systems (Rosenblum et al. 1997, Voss 2000). For example, Scolozzi & Geneletti (2017) discussed the climate change paradigm illustrating on a systemic framework grounded in the concept of anticipatory system in which the anthroposphere acts as an anticipatory system anticipating and governing the climate dynamics. In the current study, we considered a one way coupled nonlinear anticipation system proposed by Hahs & Pethel (2011) with the following equations:

$$\begin{aligned} x_{n+1} &= f(x_n), \\ y_{n+1} &= (1-\epsilon) f(y_n) + \epsilon g_\alpha(x_n), \end{aligned} \tag{3.5}$$

where $f(x) = 4x(1-x)$ is the chaotic logistic map and $g_\alpha(x_n) = (1-\alpha)f(x) + \alpha f(f(x))$. For this system, $\epsilon = 0.3$ is chosen. For $\epsilon > 0.3$, the system is synchronized. The response y is strongly coupled to the driver, but it retains its independence. The function $g_\alpha(x_n) = (1-\alpha)f(x) + \alpha f(f(x))$ includes a tunable parameter α (coupling coefficient). When α =0, the y system is driven towards the x system and when $\alpha = 1$, the y system is driven towards the future of the system x. We initialized the system with $(x_0, y_0) = (0.4, 0.1)$. For the TE calculations, the source embedding dimension $l = 1$ is chosen based on the governing equations, while the AI for the destination system shows a maximum value at time lag 1 and then a decrement for higher lags. Hence, k is chosen to be 1. From the governing equations, the interaction time delay of 1 is chosen as the source transfers maximum information exchange at a time delay of 1.

Figure 3.3 represents information exchange for the system with governing Equation (3.5). The IF-linear shows an information exchange from x to y at

all coupling coefficients except for the values 0.4 to 0.7 and no information exchange from y to x except at coupling coefficient 0.2 for the time series length of 5000 time units. Similarly, at lower time series lengths, this asymmetry is captured; however, at coupling coefficients values from 0.4 to 0.7, it is difficult to distinguish the information flow directionality even with the error bars. With longer time series, IF-linear successfully detects the unidirectional information exchange, and this result is reported in San Liang (2014). Even though one can argue that IF-linear has been developed strictly for linear systems, it qualitatively detects the asymmetry in the information exchange for this anticipatory system. San Liang (2014) presented a detailed description of how IF-linear is able to retrieve the properties of this anticipation system through linearization. This property of IF-linear is remarkable, as it is easy to apply given two time series. Moreover, unlike nonparametric TE, IF-linear does not depend on free tuning parameters, which is an added value for its application and also computationally efficient. The TE-linear fails to detect the direction of information exchange at all time series lengths for this system, in fact, the directionality is reversed. It is worth noting that the implementation of TE-linear does not involve any linearization, unlike IF-linear.

The TE-binning estimator (10 bins) shows information exchange from x to y at a time series length of 5000 time units. At higher coupling coefficients, nonzero information exchange is detected from y to x. With a time series length less than 1000 time units, unrealistic nonzero information exchange is detected from y to x at all coupling coefficients. The TE-kernel estimator (kernel width = 0.5) also detects the one-way coupling between x and y (consistently between 0.25 to 1 kernel widths) at all time series lengths. However, at greater coupling coefficients, spurious information exchange is detected from y to x at all-time series lengths with kernel width greater than 0.5. The TE-kraskov estimator (4 nearest neighbors) shows exactly the unidirectional coupling between x to y (consistently between 4 to 60 neighbors) at all time series lengths. The better estimation of TE-kraskov might be attributed to the adaptive data efficient discretization as well as bias correction (Kraskov et al. 2004). However, with TE-kraskov, a very slow convergence of information exchange with increasing time series length is noted. This behavior was reported in the study by Zhu et al. (2015). In summary, for this system, IF-linear is able to detect the asymmetry in the information exchange while TE-linear fails. The nonparametric TE-kraskov estimator provides reliable results for all time series lengths. While the asymmetry in coupling is captured by TE-kernel and TE-binning, unrealistic nonzero information exchange is detected at higher coupling coefficients from system y to x.

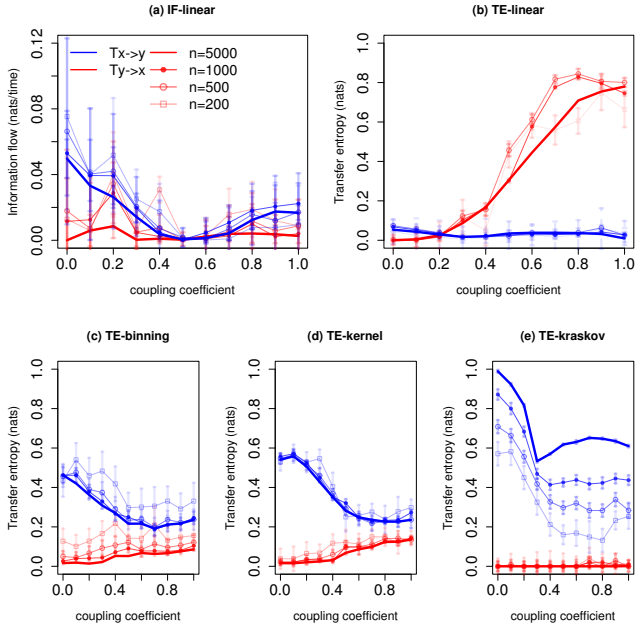


FIGURE 3.3. Information exchange in the unidirectional coupled nonlinear anticipatory system (Equation (3.5)) with various time series lengths (n) measured by (**a**) the IF-linear method (nats/time) and (**b**–**e**) with different variants of the TE measure (in nats). Error bars represent two standard deviations of the permuted surrogate samples.

Bidirectional Coupled Non-Linear System

In this section, we extend our analysis from a unidirectional nonlinear coupled system to a bidirectionally coupled nonlinear system. For this purpose, we considered Heńon maps which were motivated by the Lorenz equations. The Heńon map captures the stretching and folding dynamics of chaotic systems such as the Lorenz system which mimic the atmospheric behavior. We considered two identically coupled Heńon maps with the following governing equations (Wiesenfeldt et al. 2001):

$$\begin{aligned} x_{n+1} &= 1.4 - x_n^2 + 0.3x_{n-1} + C_{yx}(x_n^2 - y_n^2), \\ y_{n+1} &= 1.4 - y_n^2 + 0.3y_{n-1} + C_{xy}(y_n^2 - x_n^2), \end{aligned} \tag{3.6}$$

where the coupling coefficients C_{yx} and $C_{xy} \in [0,0.4]$. For the TE calculations, the source embedding dimension $l = 1$ was chosen based on the governing equations. The AI for the destination system shows a maximum value at first two time lags and then a decrement for higher time lags. Hence, we considered the embedding dimension $k = 1$. From the governing equations, the interaction time delay of 1 was chosen as the source transfers maximum information

exchange at a time delay of 1. For this system, the region outside the coupling interval [0.3,0.3] corresponds to synchronization.

Figure 3.4 represents information exchange for the system with governing Equation (3.6) for a time length of 5000 time units. The IF-linear shows an information exchange from x to y at coupling coefficient C_{yx} from 0.1 to 0.25 and C_{xy} from 0.1 to 0.2. Further, it shows an information exchange from system y to x at coupling coefficient C_{yx} from 0.1 to 0.3 and C_{xy} from 0.1 to 0.2. A similar behavior was also observed with the TE-linear estimator, but at a lower coupling coefficient below the synchronization, TE-linear fails. The nonparametric TE-binning (consistently between 10 to 20 bins) and TE-kernel (consistently between 0.25 to 1 kernel widths) shows an information exchange between x and y with couplings below [0.3,0.3]. Furthermore, the strength of the information exchange increases from x to y as the coupling strength C_{xy} increases and vice versa. This symmetric behavior is expected as the two Heńon systems are identically coupled. This was also shown by Lungarella et al. (2007). Similar patterns with the TE-kraskov estimator (4 nearest neighbors) are also seen (consistently between 4 to 60 neighbors). We also tested the information exchange for this system with a time series length of 500 time units (figure not shown). It was observed that the spatial patterns of the information exchange exhibit similar spatial patterns as those in Figure 3.4 but less clearly established for TE-binning, TE-kernel, and TE-kraskov estimators. However, the TE-binning estimation overestimates the information exchange when compared with TE-kernel and TE-kraskov. In summary, for this system, both the parametric methods failed to detect the bidirectional nonlinear interactions, while the nonparametric TE-kernel and TE-kraskov estimators showed consistent results for time series lengths $n = 5000$ and $n = 500$ time units.

From the above-discussed idealized systems, considering the dependencies of TE nonparametric estimations on the free parameters, instead of relying on any single estimator, we propose to use a composite of TE-kernel and TE-kraskov estimators for nonlinear systems. Furthermore, the free parameters, i.e., kernel width and the number of nearest neighbors are to be tuned until both the estimators consistently show a significant information exchange. For the linear systems in addition to TE-kernel and TE-kraskov along with the tuning of free parameters, linear estimators (IF-linear and TE-linear) shall be used simultaneously.

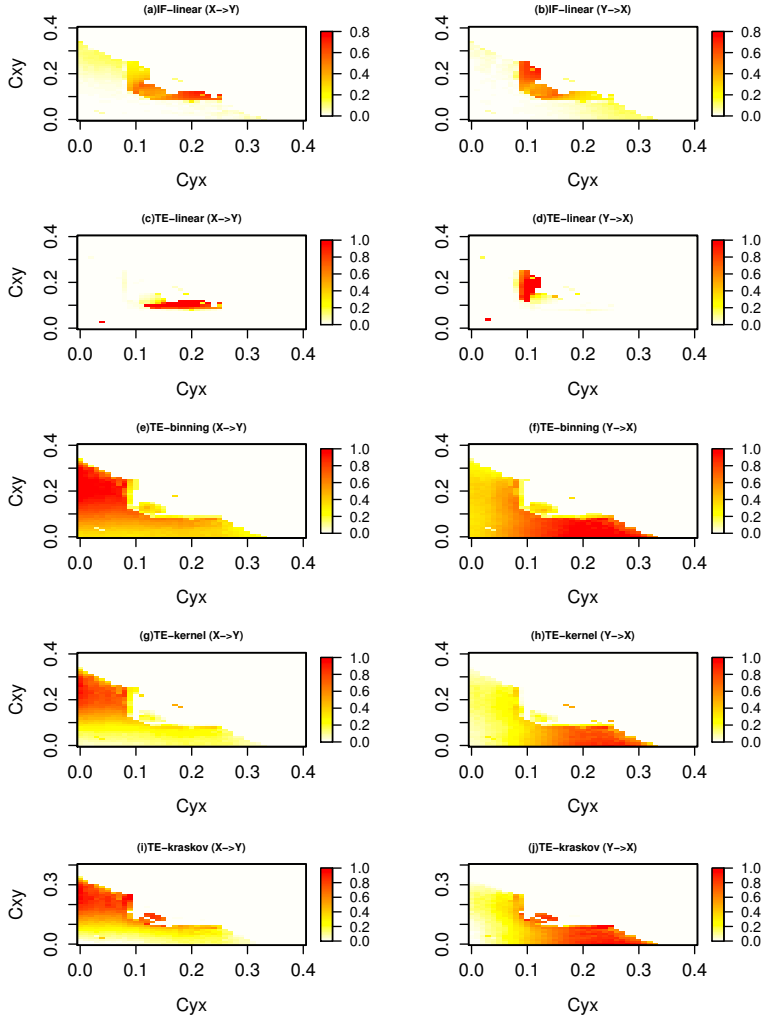


FIGURE 3.4. Information exchange in the bidirectional coupled nonlinear system (Equation (3.6)) with time series length of 500 time units measured by (**a**,**b**) the IF-linear method (nats/time) and (**c**–**j**) with different variants of the TE measure (in nats).

Two-Scale Lorenz-96 Model

So far, we had investigated information exchange in idealized linear and nonlinear coupled systems where the dynamics of the system is known or expected. In this section, we consider a simple conceptual model of atmosphere-like multiscale dynamics, namely, the Lorenz-96 system (Lorenz 1996) which consists of coupled fast and slow subsystems. This model was originally introduced to mimic multiscale midlatitude weather. Furthermore, it has been extensively

used in the study of the influence of multiple spatiotemporal scales on the predictability of atmospheric flows (Herrera et al. 2010, Tödter & Ahrens 2015). Although the equations in Lorenz-96 are known and identically coupled, the interaction behavior of Lorenz-96 is hard to expect, as its dynamics is dominated by the interplay between the fast and slow subsystems. Hence, to detect the direction of information exchange, we applied the methods discussed in Section 3.2 to the Lorenz-96 system. The Lorenz-96 system has the following governing equations:

$$\begin{aligned} \frac{dx_i}{dt} &= x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b}\sum_{j=1}^{n} y_{j,i}, \\ \frac{dy_{j,i}}{dt} &= cby_{j+1,i}(y_{j-1,i} - y_{j+2,i}) - cy_{j,i} + \frac{hc}{b}x_i. \end{aligned} \tag{3.7}$$

It consists of m slow variables x_i coupled to $m \times n$ fast variables $y_{j,i}$. Furthermore, the two systems are coupled by coupling constant h, scaling constants b and c, and F is a constant forcing. Here, $i = 1, ...m$ and $j = 1, ...n$. Both the x_i and $y_{j,i}$ have periodic boundary conditions, i.e., $x_{m+1} = x_1$; $x_0 = x_m$ and $y_{n+1,i} = y_{1,i+1}$; $y_{0,i} = y_{n,i-1}$. The conventional parameter values $F = 8$ and $m = 40$ are chosen. Furthermore, the values of n, b, and c are chosen to be 4, 10, and 10, respectively. This setup leads to a two-scale model where the fast variables fluctuate 10 times faster than the slow ones. The coupling parameter h is varied from 0 to 1 with an increment of 0.1. For more details of the system, refer to Herrera et al. (2010). For this system, we chose embedding dimensions $l = m = 1$ based on the peak AI values at time lag 1 for TE calculation. From the governing equations, the interaction time delay of 1 is chosen as the source transfers maximum information exchange at a time delay of 1.

Figure 3.5 represents information exchange for the system with governing Equation (3.7) for various time series lengths. The IF-linear indicates that there exists a bidirectional information exchange between the slow system and the fast system. Furthermore, the magnitude of information exchange from the fast system to the slow system is stronger than vice versa. The TE-linear shows that a greater amount of information is exchanged from the fast system to the slow system at low couplings. At higher couplings, the information converges to zero in both directions. However, as the IF-linear and TE-linear fail for strong nonlinear systems, we do not draw any conclusions from these results as the Lorenz-96 is a highly nonlinear system.

The TE-binning estimation (10 bins) shows information exchange from x to y, i.e., the slow system to the fast system. Moreover, weak information exchange is detected from y to x. The TE-kernel estimator (kernel width = 0.5) and TE-kraskov estimator (4 nearest neighbors) also show that the information exchange is dominant from the slow system to the fast system than vice versa.

From the composite use of TE-kraskov and TE-kernel along with the parameter tuning, we see a bidirectional information exchange between the systems x and y. Moreover, the system x leads the system y, i.e., the slow system leads the fast system. Herrera et al. (2010), in their study, also noted that there exists a bidirectional influence between the fast and slow system as revealed from the spatial patterns of the system trajectories. They also found that the fast system is conditioned by the slow system especially at lower couplings, i.e., the system dynamics is dominated by the slow system. From Figure 3.5, the nonlinear methods show that the information exchange decreases as the coupling coefficient increases; this requires a further detailed investigation.

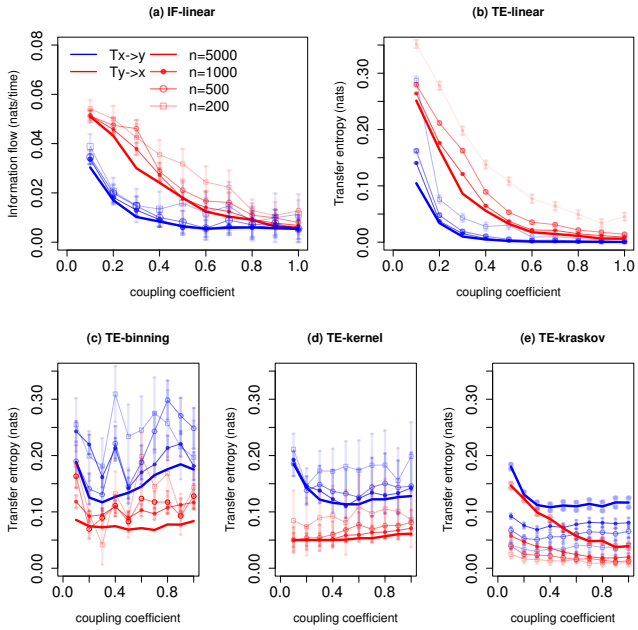


FIGURE 3.5. Information exchange in the Lorenz-96 system (Equation (3.7)) with various time series lengths (n) measured by (**a**,**b**) the IF-linear method (nats/time) and (**c**–**j**) with different variants of the transfer entropy (TE) measure (in nats). Error bars represent two standard deviations of the permuted surrogate samples.

3.3.2 *Application to Climate Phenomena*

Information Exchange between Indian and Pacific Ocean

The El Niño Southern Oscillation (ENSO) is an important large-scale coupled atmosphere–ocean phenomenon which has a remote influence on sea surface temperatures in other ocean basins (Klein et al. 1999). An atmospheric bridge is

one of the mechanisms through which this influence is mediated (Lau & Nath 1996, Ashok et al. 2003). ENSO oscillations are characterized by changes in the sea surface temperature (SST) patterns over the equatorial Pacific Ocean. Among many measures, the Niño 4 and Niño 3 indices measure the ENSO oscillation strength and phase. Another important climate variability called the Indian Ocean dipole (IOD) occurs in the Indian Ocean and is known to have an influence on Indian summer monsoon and on rainfall over Australia, Indonesia, and East Africa (Ummenhofer et al. 2009, Lau & Nath 1996, Cai & Cowan 2008, Hong et al. 2008). A positive phase of IOD represents a cooling in the eastern part of the Indian Ocean and warming in the western part of the Indian Ocean and vice versa. The strength and phase of the IOD are measured through the IOD index.

Earlier studies stated that IOD might arise due to the internal atmosphere–ocean coupling of the Indian Ocean (Yamagata et al. 2004). However, the ENSO forcing is one of the factors responsible for causing IOD, and furthermore, a co-evolution of ENSO and IOD events are noted (Luo et al. 2010, Meyers et al. 2007, Fischer et al. 2005, Saji et al. 1999). ENSO and IOD are known to interact with each other through the Walker circulation in the atmosphere (Ashok et al. 2001). It is also observed that the IOD could also contribute to the development phase of ENSO through feedbacks Annamalai et al. (2005), Ashok et al. (2001). Although the ENSO and IOD co-evolve during some years, there is still no consensus about their linkage. San Liang (2014) investigated the interaction of the Niño 4 index on the Indian Ocean SST and also the IOD index on the Pacific Ocean SST. In his study, the patterns of information flow from the Niño 4 index to Indian Ocean SST revealed an Indian Ocean dipole-like structure over the Indian Ocean basin. Furthermore, an ENSO-type pattern is also obtained over the Pacific Basin when the IOD index is applied to the Pacific Ocean SST. In our study, we applied the linear and nonlinear estimators to robustly measure the interaction mechanism between ENSO and IOD.

Here, the IOD and ENSO relationship was investigated with all the information exchange measures discussed above. The monthly time series of the Niño 4 index and Indian ocean dipole index were obtained from NOAA ESRL and JAMSTEC, respectively, for the period of 1958–2010 (i.e., 633 months). The lag-1 autocorrelation of Niño 4 is 0.8 and an IOD index of magnitude 0.9. Hence, a time series length longer than 500 time units is recommended to apply the methods mentioned in the previous examples. As the available length of the indices is 633 months, it is expected that information exchange detection and quantification could be robustly estimated by the information exchange methods mentioned in Section 3.2. However, detecting interaction delays might need a greater amount of time series while estimating TE (Wollstadt et al. 2014). Hence, we cross-checked if our interaction delays are physically consistent with

the existing literature. As data input, the SST reanalysis named COBE-SST2 Sea Surface Temperature and Ice (Hirahara et al. 2014) was obtained from NOAA for the same period. Before applying the information exchange methods, the indices were deseasonalized, and the linear trend was removed to fulfill the stationarity criteria for TE (Runge 2018) and IF-linear estimations.

Table 3.1 gives the information exchange between the two indices as quantified by the various methods. A permutation technique, under which the surrogates preserve the tuples $p(x_{n+1}, x_n^k, y_n^l)$ for TE estimation is used to determine the significant information exchange within 95% confidence interval (100 samples). For more details into the technique, refer to Lizier (2014). The significance test for IF-linear follows San Liang (2014). Moreover, based on the AI, we chose the destination embedding $k = 1$ due to the minimal increase from time lag 1 to time lag 10 months for both IOD and Niño 4 , while we scanned for the time interaction delay at which maximum information is exchanged from the source to the destination (Wibral et al. 2013). The parametric methods show significant information exchange from the Niño 4 index to IOD at a lag of 2 to 3 months, while significant information is exchanged from IOD to the Niño 4 index with a time lag of 0 and between 10 and 14 months. The instantaneous relation between the IOD and Niño 4 has been observed throughout the historical records. However, it is interesting to note that significant instantaneous information is exchanged from IOD to the Niño 4 index and not vice versa. This relation could be attributed to the feedback of the IOD on the Pacific Ocean due to the wind anomalies induced by ENSO over the Indian Ocean (Luo et al. 2010). The time lagged information exchange results suggest that ENSO events tend to influence the IOD (Ueda & Matsumoto 2000), and then the induced IOD tends to provide feedback to the ENSO (Behera & Yamagata 2003, Annamalai et al. 2005). The nonparametric TE-kernel and TE-kraskov estimators also exhibit a similar behavior except the TE-binning estimator. The free parameters are tuned and tested rigorously for numerical consistency.

TABLE 3.1. Information exchange between the Niño 4 (N4) and the Indian Ocean dipole (IOD) index (* refers to significant information exchange).

Method	N4 to IOD (lag=3)	IOD to N4 (lag=0)	IOD to N4 (lag=7)	Units	
IF-linear	1.0 *	1.2 *	1.1 *	nats/month	$\times 10^{-2}$
TE-linear	0.7 *	1.3 *	1.4 *	nats	$\times 10^{-2}$
TE-binning	0.5 *	0.9 *	0.7	nats	$\times 10^{-2}$
TE-kernel	0.3 *	1.5 *	1.6 *	nats	$\times 10^{-2}$
TE-kraskov	0.1 *	1.1 *	0.9 *	nats	$\times 10^{-2}$

We further investigated the patterns over the respective oceans at the time of maximum influence of the indices. Figure 3.6 shows significant information exchange from the Niño 4 index to the Indian Ocean SST at a lag of 3 months.

The IF-linear shows the influence of Niño 4 on the Indian Ocean, especially on the Southeastern Indian Ocean, suggesting an information exchange from the Pacific Ocean to the Indian Ocean. The TE-linear also shows a significant exchange of information from the Pacific Ocean to the Indian Ocean with a maximum value of TE-linear near the Southeast Indian Ocean. While the TE-kernel also replicated a similar pattern, the TE-binning estimator produced a spurious pattern. The results were also checked for consistency with various kernel widths ranging between 0.25 and 2. The TE-binning estimator could not reveal a similar pattern with various bin widths. The TE-kraskov estimation also revealed an information exchange from the Pacific Ocean to the Indian Ocean and was consistent within the range between 20 and 60 nearest neighbors.

The significant information exchange pattern from IOD to the Pacific Ocean at a lag of six is represented in Figure 3.7. The IF-linear shows an information exchange over the central Pacific Ocean, suggesting an information exchange from the Indian Ocean to the Pacific Ocean. The TE-linear also shows that information is exchanged from the Indian Ocean to the Pacific Ocean with maximum values of TE occurring near the central Pacific Ocean. The TE-kernel and TE-kraskov estimators show nonlinear interactions near the East Pacific with numerical consistency from 0.25 to 2 kernel widths and 20 to 60 nearest neighbors, respectively, while the TE-binning produce spurious patterns without any numerical consistency for all varying bin widths. These nonlinear interactions need further investigation.

As the current example is a large-scale process and the variables are expected to be near-Gaussian, and as all the estimators (except TE-binning) showed a bidirectional information exchange, it could be concluded that a bidirectional information exchange exists between the Pacific and Indian Ocean SST. However, as mentioned earlier, we limited our investigation to two-dimensional systems, here between ENSO and IOD. In real-world climate systems, there could always be an unobserved influence of another system on the information exchange estimations. Hence, care is to be taken before drawing any conclusions; furthermore, the information exchange magnitudes are quite smaller compared with the idealized test cases.

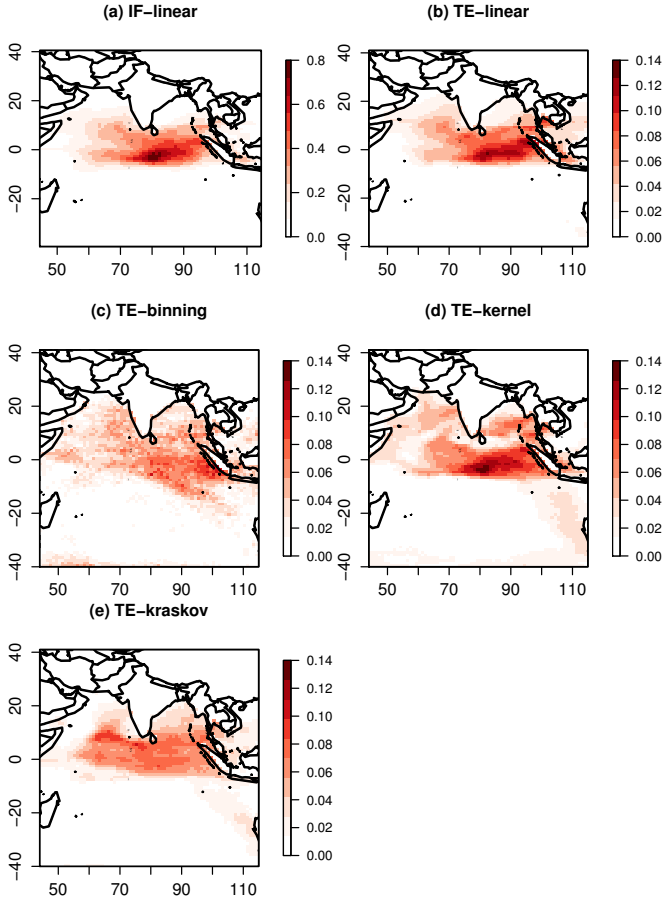


FIGURE 3.6. Information exchange from the Niño 4 Index to the Indian Ocean sea surface temperatures for the period of 1958–2010 measured by (**a**) the IF-linear method (nats/time) and (**b**–**e**) with different variants of the TE measure (in nats $\times 10^{-1}$).

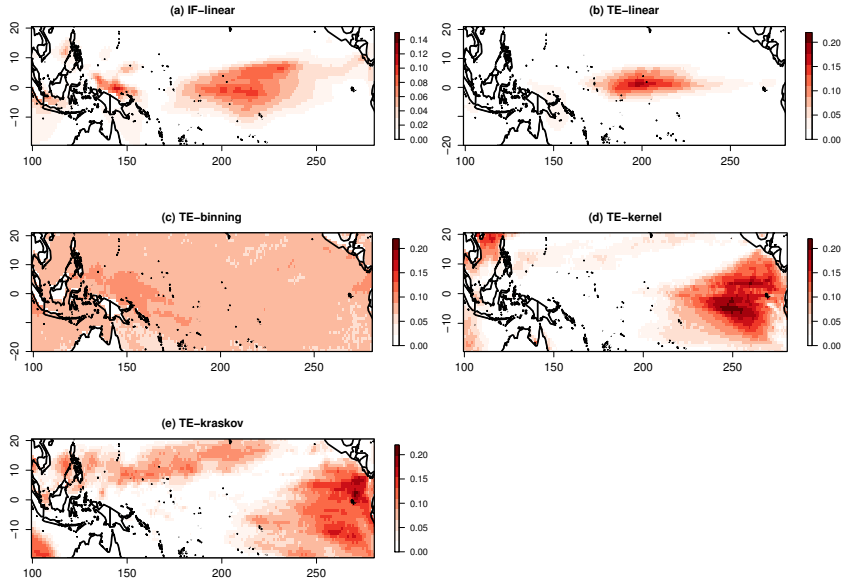


FIGURE 3.7. Information exchange from the Indian Ocean dipole Index to the Pacific Ocean sea surface temperatures for the period of 1958–2010 measured by (**a**) the IF-linear method (nats/time) and (**b**–**e**) with different variants of the TE measure (in nats $\times 10^{-1}$).

Information Exchange between Nao and European Near-Surface Temperatures

The North Atlantic Oscillation (NAO) is one of the dominant modes in the climate system which influences the European climate and many parts of the Northern Atlantic (Barnston & Livezey 1987). The NAO characterizes the strength of the subtropical high and polar low. The NAO is measured through the NAO index, which is obtained by the normalized pressure difference between the stations located in the Azores and Iceland (Papadimas et al. 2012). A positive NAO represents a deeper low over Iceland and stronger subtropical high than normal and vice versa.

There have been several studies suggesting the influence of the NAO on the winter temperatures over Europe (Trigo et al. 2002, Gámiz-Fortis et al. 2011). One of the physical mechanisms in which the NAO influences the temperatures is by the advection of the heat by anomalous mean flow (Trigo et al. 2002). Another mechanism through which the temperatures are affected is through modulation of radiation by the cloud cover (Gámiz-Fortis et al. 2011) through altering the storm track directions. In the current study, we investigated the

information exchange from NAO to winter temperatures over Europe and checked if all the methods reproduce this relation.

The NAO index and the near-surface air temperatures (T) were obtained from the CRU database (Harris et al. 2017) held at the British Atmospheric Data Centre, RAL, UK for the period of 1901 to 2016. December, January, and February temperatures were chosen as the winter months. The total time series length of 348 months was considered. The autocorrelation of NAO is of magnitude 0.1 with a time lag of 1 month, and surface air temperature autocorrelation is about 0.05 with a time lag of 1 month. Because of the low autocorrelations, we expected the methods mentioned in Section 3.2 to be robust with given time series length. Moreover, the AI for T and NAO showed a minimal increase from time lag 1 to time lag 10, and hence, $k = 1$ was chosen, while we scanned for an interaction delay at which the source exchanges maximum information exchange. The PRUDENCE regions (Christensen et al. 2002) over Europe were selected, and spatial mean air temperatures over different regions were used to detect the information exchange.

Table 3.2 shows the information exchange between the NAO and T over two PRUDENCE regions (the British Isles and Scandinavia). The parametric IF-linear shows significant information exchange from T over the British Isles to the NAO index at an interaction delay of zero days. The TE estimators show a significant information exchange bidirectionally between NAO and T over the British Isles except for the TE-binning estimator. For Scandinavia, IF-linear shows information exchange from NAO to T, while TE estimators show bidirectional information exchange. Over other PRUDENCE regions, similar bidirectional information exchange is observed except for the Mediterranean region. These results seem to be implausible, as one would not expect any information exchange from T over Europe to the NAO index. This unrealistic estimation of information exchange could have arisen from a common influence by a hidden third variable. Earlier studies have observed the sensitivity of TE estimation to a hidden variable (Runge et al. 2012). From previous literature, a possible influence on NAO might arise from the variations in sea surface temperatures, sea ice, volcanic activity, and solar activity (Wanner et al. 2001), which also influence the T over Europe.

Table 3.2. Information exchange between North Atlantic Oscillation (NAO) and winter near-surface temperatures (* refers to significant information exchange).

Method	NAO to TS	TS to NAO	Units	Region
IF-linear	0.02	0.09 *	nats/month	British Isles
TE-linear	0.318 *	0.314 *	nats	British Isles
TE-kernel	0.4 *	0.38 *	nats	British Isles
TE-kraskov	0.3 *	0.2 *	nats	British Isles
IF-linear	0.08 *	0.05	nats/month	Scandinavia
TE-linear	0.14 *	0.14 *	nats	Scandinavia
TE-kernel	0.24 *	0.2 *	nats	Scandinavia
TE-kraskov	0.16 *	0.17 *	nats	Scandinavia

3.4 Conclusions

This work targeted detecting and quantifying interactions in climate phenomena through asymmetric methods from information theory, IF, and TE. However, due to the difficulty in their estimations, we initially tested various estimators of these methods to idealized systems and then to two important climate phenomena. We limited our discussions only to two-dimensional systems.

The parametric estimators assuming linearity, such as the rigorously derived IF-linear and axiomatically proposed TE-linear, detected and reliably quantified the unidirectional and bidirectional information exchange in the idealized linear systems. IF-linear was able to detect the unidirectional information exchange for the tested unidirectional nonlinear system, whereas the TE-linear failed to do so. For the bidirectional nonlinear Heńon maps, both linear estimators failed to detect and quantify the information exchange. Hence, care has to be taken if linear information exchange measures is applied in climate system diagnosis, especially if the system variables have non-Gaussian distributions. However, these two estimators, IF-linear and TE-linear, were robust and reliable for the discussed linear systems and, in addition, IF-linear also for a weakly nonlinear system. For all the idealized systems discussed here, the nonlinear implementation of IF might reveal the interactions, but since we focused on climate applications given time series with an unknown dynamical model, we used IF-linear, which does not require system dynamics.

Among the nonparametric estimators, the TE-binning failed to be useful as a robust estimator. Even though the TE-kernel and TE-kraskov passed the idealized tests, their implementations had to be tuned to get consistent numerical results. Slow convergence of information exchange with the TE-kraskov estimator with increase in time series length was also noted. Therefore, we concluded that both reliable nonparametric estimators should be jointly applied and their implementation should be optimized for consistent results before any quantitative interpretation of the investigated nonlinear system is

drawn. This conclusion is conditioned on the availability of long enough data time series. The composite use of TE-kernel and TE-kraskov showed that the dynamics of the Lorenz-96 model is dominated by the slow subsystem.

For real climate applications, i.e., information exchange between the Indian and Pacific Oceans, the parametric and reliable nonparametric estimators showed a significant bidirectional information exchange. Moreover, the time lag of significant information exchange from Pacific to the Indian Ocean was about 2 to 3 months. An instantaneous information exchange from the Indian to Pacific Ocean was detected and also with a time lag of about 10 to 12 months. The respective spatial patterns over the Indian and Pacific Oceans revealed a significant bidirectional information exchange. Hence, given the consistent estimations, we conclude that a bidirectional information exchange exists between the Pacific and Indian Oceans, as expected from literature (Ueda & Matsumoto 2000, Behera & Yamagata 2003, Annamalai et al. 2005). However, given the limitations of TE and IF-linear, a possibility of a hidden influence by another system cannot be ruled out. This requires further analysis.

For the relation of NAO and European winter air temperatures, the estimators showed significant bidirectional information exchange. The process mechanism from NAO to European temperature is often discussed in the literature (Trigo et al. 2002, Gámiz-Fortis et al. 2011). However, the measured information exchange from European temperatures to the NAO cannot be explained by a straightforward process chain. This indicates an influence from a third hidden variable as a common driver.

Thus, even though TE and IF-linear are useful measures which allow for quantification of interactions and their directionality, their limitations and the system at hand need to be taken into account carefully before drawing any conclusions from their estimations. Hence, we propose a composite use of the information theory methods with parameter testing for various applications, for example, as a robust model evaluation framework. While this study was limited in investigating the relationship between two systems, in the future study, the authors plan to investigate interaction measures based on information theory in higher-dimensional climate system networks.

3.5 Acknowledgments

B.A. would like to acknowledge the support by the Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main. The authors also acknowledge support by the German Federal Ministry of Education and Research (BMBF) under Grant MiKlip: FKZ01LP1518C and the German Research Foundation ("Deutsche Forschungsgemeinschaft", DFG) in terms of the research

group FOR 2416 “Space-Time Dynamics of Extreme Floods (SPATE)”. The authors also thank Joseph T Lizier for providing the JDIT open source toolkit. Discussions with Stamen Dolaptchiev, Goethe University Frankfurt am Main are acknowledged.

Chapter 4

The synergistic impact of ENSO and IOD on the Indian Summer Monsoon Rainfall in observations and climate simulations - an information theory perspective[2]

[2] Published as: *Pothapakula, P.K., Primo, C., Soerland, S., Ahrens, B. (2020). The synergistic impact of ENSO and IOD on the Indian Summer Monsoon Rainfall in observations and climate simulations - an information theory perspective. Earth System Dynamics.* https://doi.org/10.5194/esd-11-903-2020

Abstract

El-Niño southern oscillation (ENSO) and Indian Ocean Dipole (IOD) are two well-know temporal oscillations in the sea surface temperature (SST), which both are thought to influence the interannual variability of the Indian Summer Monsoon Rainfall (ISMR). Until now, there has been no measure to assess the simultaneous information exchange (IE) from both ENSO and IOD to ISMR. This study explores the information exchange from two source variables (ENSO and IOD) to one target (ISMR). First, in order to illustrate the concepts and quantification of two-source IE to a target, we use idealized test cases consisting of linear as well as non-linear dynamical systems. Our results show that these systems exhibit net synergy (i.e., the combined influence of two sources on a target is greater than the sum of their individual contributions), even with uncorrelated sources in both the linear and non-linear systems. We test IE quantification with various estimators (the Linear, Kernel, and Kraskov estimators) for robustness. Next, the two-source IE from ENSO and IOD to the ISMR is investigated in observations, reanalysis, three global climate model (GCM) simulations, and three nested, higher-resolution simulations using a regional climate model (RCM). This (1) quantifies IE from ENSO and IOD to ISMR in the natural system, and (2) applies IE in the evaluation of the GCM and RCM simulations. The results show that both ENSO and IOD contribute to the ISMR interannual variability. Interestingly, significant net synergy is noted in the central parts of the Indian subcontinent, which is India's monsoon core region. This indicates that both ENSO and IOD are synergistic predictors in the monsoon core region. But, they share significant net redundant information in the southern part of Indian subcontinent. The IE patterns in the GCM simulations differ substantially from the patterns derived from observations and reanalyses. Only one nested RCM simulation IE pattern adds value to the corresponding GCM simulation pattern. Only in this case, the GCM simulation shows realistic SST patterns and moisture transport during the various ENSO and IOD phases. This confirms, once again, the importance of the choice of the GCM in driving a higher-resolution RCM. This study shows that two-source IE is a useful metric that helps in better understanding the climate system and in process-oriented climate model evaluation.

4.1 Introduction

The South Asian Monsoon is considered as a large-scale coupled air-sea-land interaction phenomenon that brings seasonal rainfall to the Indian subcontinent

and other near areas (Webster et al. 1998). Large parts of the Indian subcontinent receive rainfall from June to September known as the Indian Summer Monsoon Rainfall (ISMR). The ISMR contributes about 70–90% to the total annual precipitation amount in the Indian subcontinent (Shukla & Huang 2016). The agriculture in the Indian subcontinent depends substantially on the ISMR, and any variations on the interannual as well as intraseasonal variabilities of ISMR cause a significant impact on the country's economy. The interannual variation of the IMSR is only about 10% of the mean (Gadgil 2003), yet it has a large impact on crop production. The mean seasonal rainfall predictability significantly depends on the interannual variability of the ISMR (Goswami et al. 2006, Pillai & Chowdary 2016). The interannual variability of the ISMR is linked to many noted oscillations, the El Niño Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), Atlantic Multidecadal Oscillation (AMO), Atlantic Zonal Mode (AZM), Pacific Decadal Oscillation (PDO), etc., (Nair et al. 2018, Sabeerali et al. 2019, Hrudya et al. 2021). The oscillations thought to have the most significant impact on the ISMR are ENSO and IOD (Krishnaswamy et al. 2015). Hence, in this study, we majorly focus on the individual and combined influences of the two climate modes ENSO and IOD on the ISMR interannual variability in observations, reanalysis data sets, and climate models.

ENSO is an important large-scale coupled atmosphere-ocean aperiodic oscillation over the Pacific ocean that on average occurs every 2–7 years. The Sea Surface Temperature (SST) pattern over western (central-eastern) tropical Pacific ocean experience large cold (warm) anomalies during the El Niño phase. The normal patterns of SST over the Pacific ocean are enhanced during the La Niña phase. These variabilities in the SST are coupled to the atmospheric Walker circulation, and Sir Gilbert Walker in 1924 was the first to observe a relation between ENSO and ISMR (Walker 1924, Gadgil 2003, Goswami 1998, Yun & Timmermann 2018). He noticed that often the El Niño (La Niña) conditions over the Pacific ocean are linked to weak (strong) ISMR. During the El Niño conditions, the entire walker circulation is shifted eastwards by which the descending branch of the Walker cell on the western Indian ocean shifts eastward to overlie on the Indian subcontinent, thereby suppressing the convection (Walker 1924, Kumar et al. 2006, Palmer et al. 1992). In the La Niña years, the entire Walker circulation shifts slightly westward, which assists in enhancing the convection over the Indian subcontinent. Many other studies (Goswami 1998, Slingo & Annamalai 2000) argued that the El Niño conditions do not suppress the ISMR directly through the descending branch of the Walker circulation but rather, the changes in the Walker circulation enhances the meridional Hadley circulation decent over the Indian subcontinent. Hence, it could be that the ENSO affects the IMSR through interactions between the Walker and Hadley circulations.

Another important source that is linked to the ISMR interannual variability is a dipole like structure in the Indian ocean surface temperature known as IOD (Saji et al. 1999). During a positive (negative) IOD, the southeastern part of the Indian ocean is cooler (warmer) than normal while the western part of the Indian ocean is warmer (cooler). During the positive IOD event, the meridional circulation in the region is modulated through anomalous convergence patterns over the Bay of Bengal, thereby strengthening the monsoon with anomalous positive rainfall over the Indian subcontinent while the negative IOD events lead to the weakening of the rainfall (Ashok et al. 2001). Behera & Ratnam (2018) found that the opposite phases of IOD are associated with distinct regional asymmetries in IMSR anomalies over the Indian subcontinent contributing significantly to the interannual variability. Interestingly, Behera & Ratnam (2018) found that during the co-existence of El Niño and positive IOD, the IOD tends to compensate for the influence of El Niño leading to normal rainfall by inducing anomalous convergence over the Bay of Bengal. Similarly, the negative IOD events can reduce the impact of La Niña on ISM rainfall and cause deficit monsoon rainfall. However, the study of Chowdary et al. (2015) showed that the local air–sea interaction in the tropical Indian ocean opposes the Pacific ocean impact even in the absence of IOD. Hence, still there are uncertainties associated with the individual and combined influence of ENSO and IOD on the interannual variability of ISMR.

Motivated by these large uncertainties in the present knowledge about how ENSO and IOD influence the ISMR interannual variability, we are investigating these connections from a two-source information exchange (IE) perspective. The IE between two subsystems X and Y can be understood as the average uncertainty reduction about X in knowing Y or vice versa. The information theory, in its current form, provides a complete description of the IE relationship between a single-source and a target. However complex climate system often consist of multi-sources influencing a target such as the ENSO and IOD influencing the ISMR variability. The IE in a system composed of two-source systems Y and Z to the target variable X is decomposed into four parts (Figure 4.1) according to Williams & Beer (2010): (i) unique information shared by Y to X (ii) unique information shared by Z to X (iii) redundant information or overlapping information shared by both sources Y and Z together with X (iv) synergistic information about X while knowing Y and Z together but not either of them alone. An example of synergistic information from two sources is the classical binary exclusive-or (XOR) operation (Williams & Beer 2010, James et al. 2016), where the two sources Y and Z provide information that is not available from either of their states alone but by jointly knowing their states together. Since ENSO and IOD are known to simultaneously influence the ISMR variability, one could expect the component of synergy or redundant

information existing in this climate phenomenon. In the case of synergy, the target uncertainty of IMSR interannual variability is reduced only when the states of two sources, ENSO and IOD are known together but not individually. This decomposition of information is known as partial information decomposition (PID). It is very important to note that, though the methods from information theory are very useful in analyzing the complex system behavior, their estimations are quite challenging due to their sensitivity to free tuning parameters and sample size (Knuth et al. 2013, Smirnov 2013, Pothapakula et al. 2019). Hence, this study follows and uses various estimators we proposed in our earlier work (Pothapakula et al. 2019) for robustness in the results.

Here we are investigating the information exchange from ENSO and IOD to the IMSR interannual variability by using available observations, reanalysis data sets, and climate models. However, before exploring the two-source IE from the ENSO and IOD to IMSR variability, we first demonstrate the concept of two-source IE with results from a simple idealized linear and non-linear dynamical models for better understanding. We also use various estimators of IE, for example, Linear, Kraskov, and Kernel estimators for robustness. Then, the two-source IE concept is applied to observations and reanalysis data sets. This helps in understanding the IE dynamics of ENSO and IOD to the interannual variability of IMSR in the natural system. Thereafter, we investigate if the two-source information exchange dynamics of ENSO and IOD to ISMR interannual variability is replicated in three different global climate models (GCM) simulations from the 5th phase of the Coupled Model Intercomparison Project (CMIP5). Since it is well known that GCMs due to their low spatial resolution do not resolve all the subgrid-scale phenomena, we have used dynamical downscaling of the three GCM simulations with an RCM to obtain higher resolution details (Bhaskaran et al. 2012, Choudhary et al. 2018, Dobler & Ahrens 2011, Asharaf & Ahrens 2015, Lucas-Picher et al. 2011). The RCM simulations are performed with a horizontal resolution of 25km (~ 0.22) and follow the framework of coordinated regional downscaling experiments (CORDEX) (Giorgi et al. 2009, Gutowski Jr et al. 2016). By employing the two-source IE from the ENSO and IOD to the ISMR interannual variability on both the driving GCM simulations and the downscaled RCM simulations, we can evaluate the performance of the model chain. To our knowledge, this is a first of its kind evaluation study of GCM simulations and RCM simulations with information theory methods from the two-source IE viewpoint.

This paper is organized as follows. In Section 4.2 we explain briefly the information theory methods and estimators used in this study followed by a brief discussion about the idealized linear and non-linear dynamical systems. In Section 4.3 observational and reanalysis data, various GCMs in CMIP5 used in this study, and the RCM model used in dynamically downscaling the GCM

simulations are discussed. In Section 4.4, the results obtained from idealized systems and model evaluation are shown along with a detailed discussion. Finally, conclusions are drawn in Section 4.5.

4.2 The theory of information exchange

Shannon (2001) introduced the concept of information entropy, which quantifies the average uncertainty of a given random variable. Recently, various methods from information theory have been widely used in the fields of earth system sciences (Bennett et al. 2019, Gerken et al. 2019, Jiang & Kumar 2019, Ruddell et al. 2019), climate sciences (Nowack et al. 2020, Runge et al. 2019, Garland et al. 2019, Campuzano et al. 2018, Bhaskar et al. 2017) and in other interdisciplinary sciences (Wibral et al. 2017, Novelli et al. 2019, Ahmad 2018). This section comprises of the basic concepts of information theory along with a brief introduction of various estimators. Also, a description of the idealized systems used in this study is covered.

4.2.1 Concepts from Information Theory

The Shannon entropy (Shannon 2001) of a random variable X, quantifies the amount of uncertainty contained in it and is defined by

$$H(X) = -\sum_{x} p(x) \log p(x),$$

where $p(x)$ is the probability of a discrete state of the random variable X. The summation goes through all states of the random variable X. The units of entropy are expressed in nats if a natural logarithm is applied (in bits when the logarithm base is 2).

Mutual information (MI) quantifies the reduction in the uncertainty of one random variable given knowledge of another variable (Thomas & Joy 2006) and is defined by

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)},$$

where $p(x,y)$ is the joint distribution of variables X and Y, and $p(x)$, $p(y)$ are the marginal distributions of X and Y, respectively.

Mutual information between two sources Y and Z and a target X is given as

$$I(X;Y,Z) = \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)}{p(x)p(y,z)},$$

where $p(x, y, z)$ is the joint distribution of variables X,Y and Z, and $p(x)$, $p(y, z)$ are the marginal probabilities. Furthermore, the information $I(X; Y, Z)$ that the two sources share with target should decompose according to partial information decomposition by Williams & Beer (2010) into four parts (Figure 4.1) as

$$I(X; Y, Z) = U(X; Y|Z) + U(X; Z|Y) + R(X; Y, Z) + S(X; Y, Z), \quad (4.1)$$

where $U(X; Y|Z)$ is the unique information shared by Y to X, $U(X; Z|Y)$ is the unique information shared by Z to X, $R(X; Y, Z)$ redundant information shared by both sources Y and Z together with X, and $S(X; Y, Z)$ synergistic information about X while knowing the states of Y and Z together.

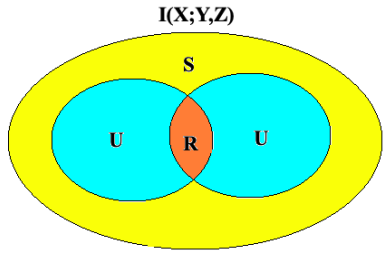


FIGURE 4.1. Information exchange from two sources Y, Z to the target X decomposed according to PID as unique information (U), redundant information (R) and synergistic information (S)

In the case of two sources influencing the target, the mutual information shared by a single source to the target is given by

$$\begin{aligned} I(X; Y) &= U(X; Y|Z) + R(X; Y, Z), \\ I(X; Z) &= U(X; Z|Y) + R(X; Y, Z). \end{aligned} \quad (4.2)$$

From the current information theory framework, the quantities $I(X; Y, Z)$, $I(X; Y)$, $I(X; Z)$ can be straightforwardly computed. Unfortunately, with the present standard methods available from information theory, one can not obtain the contributions of unique, synergy, and redundant information exchange metrics solely (Barrett 2015). Here, we would like to bring to the attention of the readers that many interesting studies have come up with various definitions of these metrics (Williams & Beer 2010, Griffith & Koch 2014, Bertschinger et al. 2014, Finn & Lizier 2018) and still, there has been no consensus among the scientific community for obtaining these metrics. A complete and consistent

framework on quantifying the individual contributions of various terms in PID would make information theory a complete framework for understanding the information dynamics of multi-source systems.

According to Barrett (2015), one can obtain a quantity known as net synergy from Eq.4.1 and Eq.4.2 as

$$\begin{aligned}\Delta I(X;Y,Z) &= I(X;Y,Z) - I(X;Y) - I(X;Z),\\ &= S(X;Y,Z) - R(X;Y,Z).\end{aligned} \tag{4.3}$$

When $\Delta I(X;Y,Z) > 0$, synergistic information from two sources is greater than redundant information and vice versa. The ΔI provides a lowerbound for synergistic/redundant information. From here on, if $\Delta I(X;Y,Z) > 0$ we refer as net synergistic information and if $\Delta I(X;Y,Z) < 0$ we refer to as net redundant information.

4.2.2 *Estimation techniques*

Though the information theory methods are very useful in assessing the behavior of dynamical systems, their estimation is challenging. Hence, in this study, we implemented various estimators for robustness in our results.

Estimation under linear approximation (Linear estimator)

Here we will briefly introduce the basic concepts for estimation of the two-source IE under linear approximation. For a detailed explanation of the concept, we are referring the reader to Barrett (2015).

The entropy for a continuous random variable X under linear approximation is given as

$$H(X) = \frac{1}{2}\log[\det \Sigma(X)] + \frac{1}{2}m\log(2\pi e),$$

where m is the dimension of random variable X, $\Sigma(X)$ is the $m \times m$ matrix covariances i.e., $\text{cov}(X^i, X^j)$.

Following Barrett (2015), the partial covariance of X with respect to Y is given as

$$\Sigma(X|Y) = \Sigma(X) - \Sigma(X,Y)\Sigma(Y)^{-1}\Sigma(Y,X).$$

From then the conditional entropy can be derived as

$$H(X|Y) = \frac{1}{2}\log[\det \Sigma(X|Y)] + \frac{1}{2}m\log(2\pi e).$$

The mutual information $I(X;Y)$ is the difference between $H(X)$ and $H(X|Y)$,

$$I(X;Y) = \frac{1}{2}\log\left[\frac{\det\Sigma(X)}{\det\Sigma(X|Y)}\right].$$

For a general three dimensional jointly Gaussian system $(X,Y,Z)^T$, and by setting zero mean and unit variance, the covariance matrix is given by,

$$\Sigma = \begin{bmatrix} 1 & a & c \\ a & 1 & b \\ c & b & 1 \end{bmatrix}$$
.

Thus, from the above matrix, the mutual information is given as

$$I(X;Y) = \frac{1}{2}\log\left(\frac{1}{1-a^2}\right),$$

$$I(X;Z) = \frac{1}{2}\log\left(\frac{1}{1-c^2}\right),$$

$$I(X;Y,Z) = \frac{1}{2}\log\left(\frac{1-b^2}{1-(a^2+b^2+c^2)+2abc}\right).$$

The net synergy can be obtained by $I(X;Y,Z) - I(X;Y) - I(X;Z)$, given as

$$\Delta I(X;Y,Z) = \frac{1}{2}\log\left(\frac{(1-a^2)(1-b^2)(1-c^2)}{1-(a^2+b^2+c^2)+2abc}\right).$$

Estimation through box step kernel (Kernel estimator)

The estimation of non-linear entropy and mutual information estimators contains Probability Density Functions (PDFs). The uni-variate and bi-variate PDFs for continuous data can be estimated through various available discretization methods (e.g., binning, kernel etc). Here we use a simple box step kernel Θ with $\Theta(x > 0) = 0$ and $\Theta(x < 0) = 1$ for the estimation of relevant joint probability distributions (e.g., $\hat{p}(x,y)$, $\hat{p}(x)$ and $\hat{p}(y)$). For example, the joint probability distribution $\hat{p}(x,y)$ is calculated as

$$\hat{p}_r(x_n, y_n) = \frac{1}{N}\sum_{n'=1}^{N}\Theta(|(x_n - x_{n'}), (y_n - y_{n'})| - r),$$

where the norm corresponds to the maximum distance in the joint space and r is the kernel width. Similarly one can estimate the PDF for high dimensional systems for the estimation of MI. For more details into the estimator, refer to Kantz & Schreiber (2004), Goodwell & Kumar (2017) and information-theoretic toolkit from Lizier (2014).

Estimation through k-nearest neighbor (Kraskov estimator)

The k-nearest neighbor estimator uses an adaptive binning strategy by estimating the average distances to the k-nearest neighbor data points. For example, the MI can be computed as

$$I(X;Y) = \Psi(k) - < \Psi(n_x + 1) + \Psi(n_y + 1) > + \Psi(N),$$

where N is total number of points, n_x and n_y are the number of points that fall in the marginal spaces of X and Y respectively within the distance taken as d= max($||x - x'||, |y - y'||$) and Ψ denotes the digamma function. For more details refer to Kraskov et al. (2004). Similarly, the equation mentioned above can be extended to higher dimensional estimation of MI. From hereafter, the estimation through k-nearest neighbor is called as Kraskov estimator.

4.2.3 *Idealized systems for demonstration*

Before we apply information theory estimators to two-source information exchange in climate applications, we consider idealized linear systems as given in the following sub-section to demonstrate the concept of two-source IE.

Linear autoregressive systems

Often in climate systems, the future state prediction of a variable relies on the past of its own state (persistence) or from past of another variable (Runge et al. 2014), or from the linear/non-linear combination of both (possible case of net synergy/redundancy). Hence, as a first case of demonstration, we considered a two-dimensional linear system (Barrett 2015) x and y, with x receiving information from its immediate past and from the immediate past of y with the following governing equations:

$$\begin{aligned} x_t &= \alpha x_{t-1} + \alpha y_{t-1} + \mathcal{N}_x(0,1), \\ y_t &= \mathcal{N}_y(0,1), \end{aligned} \tag{4.4}$$

where α is the coupling coefficient varied from 0 to 0.8 with an increment of 0.1 and $\mathcal{N}(0,1)$ is Gaussian noise with zero mean and unit variance. The system was initialized with ($x_0 = 0$) and is integrated around 100,000 iterations. For the analysis of two-source IE with various estimators, we use the last 5000 time units from the available time series.

In the first example, we considered IE from two sources (one source being the persistence) contributing to the target prediction, however not all predictions of target depend on two sources simultaneously (i.e., net synergy/redundancy do not exist), hence as a second case , we considered a system consisting

of two subsystems which are coupled with each other but only having a single source with the governing equations

$$\begin{aligned} x_t &= \alpha y_{t-1} + \mathcal{N}_x(0,1), \\ y_t &= \alpha x_{t-1} + \mathcal{N}_y(0,1), \end{aligned} \tag{4.5}$$

with α being the coupling coefficient. We followed similar steps for integration as in the previous linear system.

Finally as third example, we test a three-dimensional system in which two individual sub-systems contribute to the evolution of third system such as the ENSO and IOD, as two individual systems contributing to the interannual variability of the ISMR. This system has the governing equations

$$\begin{aligned} x_t &= \alpha y_{t-1} + \alpha z_{t-1} + \mathcal{N}_x(0,1), \\ y_t &= \mathcal{N}_y(0,1), \\ z_t &= \mathcal{N}_z(0,1), \end{aligned} \tag{4.6}$$

where system y and z are two individual sub-systems exchanging information to the target system x.

We also extended our analysis to a non-linear Heńon system described in the Appendix section.

4.3 DATA AND CLIMATE MODELS

In this section, we will discuss various observational and reanalysis data sets used to quantify the two-source IE from ENSO and IOD to ISMR interannual variability in the natural system. Furthermore, the details of various GCM and RCM simulations used in this study are also covered.

4.3.1 Observational, reanalysis data sets and climate simulations

We are focusing on the South Asian Summer Monsoon seasons, starting from June and ending in September (June- July-August-September: JJAS), thus monthly data sets for JJAS for the time period 1951-2005 from observations and model simulations are used in this study. Various observational, reanalysis data sets and model simulations used to quantify the two-source IE from the ENSO and IOD to the ISMR interannual variability are listed in Table 4.1 and are also described here.

Observational, reanalysis data sets and indices

The UK Met Office's Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST 1.1) (Rayner et al. 2003) is used to retrieve SST information for the Indian and the Pacific ocean. Monthly precipitation fields from Global Precipitation Climatology Centre (GPCC) (Schneider et al. 2008) is used as precipitation observational record together with a high-resolution data set, covering only the monsoon south Asia domain, namely the Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE) monthly accumulated precipitation (Yatagai et al. 2012). The rainfall, winds, and specific humidity are taken from the National Center for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data set (Kalnay et al. 1996). The ENSO and IOD indices are obtained from the National Oceanic and Atmospheric Administration Earth System Research Laboratories(NOAA ESRL) and Japan Agency for Marine-Earth Science and Technology(JAMSTEC) for validation of PCs derived from the observational SST data sets, i.e., the HadISST, and NCEP reanalysis SST. In addition to the above-mentioned data sets, we also used ERA-Interim (Dee et al. 2011) and MERRA (Rienecker et al. 2011) reanalysis rainfall datasets (1980-2005) as additional resources.

Global and regional climate simulations

The three CMIP5 GCMs (details in Table 4.1), the MPI-ESM-LR (Stevens et al. 2013), Nor-ESM-M (Bentsen et al. 2013) and EC-EARTH (Hazeleger et al. 2010) were dynamical downscaled with the non-hydrostatic regional climate model COSMO-crCLM version v1-1. The COSMO-crCLIM is an accelerated version of the COSMO model (Fuhrer et al. 2014) in climate mode (Leutwyler et al. 2016, Rockel et al. 2008). A two-stream radiative transfer calculations are based on Ritter & Geleyn (1992), the convection is parameterized by Tiedtke (1989), the turbulent surface energy transfer and planetary boundary layer are using the parametrization of (Raschendorfer 2001), and precipitation is based on a four-category microphysics scheme that includes cloud, rainwater, snow, and ice (Doms et al. 2005). The soil-vegetation-atmosphere-transfer is using the TERRA-ML (Heise et al. 2006), however, this current version is employing a modified groundwater formulation (Schlemmer et al. 2018). The RCM simulation has a horizontal resolution of 0.22° (i.e., 25km) and with 57 vertical levels and is using a time step of 150s. The model simulation configuration is following the CORDEX framework, meaning that a historical period is simulated from 1950-2005, and the business as usual future emission scenario (RCP8.5) is simulated from 2006-2099. However, here we are only looking into the historical period. It is to be noted that for the analysis of

rainfall anomaly composites, moisture anomalies, and IE plots, the GCM and RCM simulations are interpolated to a common observational grid (a grid with 0.25°). Our interpretation of results does not change much with the original resolution of the datasets.

Table 4.1. CMIP5–GCMs/RCM/observations descriptions used in the current study.

GCM Modeling center	Acronym	Ensemble member	Atm.Resolution
Max Planck Institute for Meteorology	MPI-ESM-LR	r1i1p1	1.875 ° × 1.875 °
Norwegian Climate Centre	Nor-ESM-M	r1i1p1	2.5 ° × 1.9 °
SMHI, Sweden	EC-EARTH	r12i1p1	1.125 ° × 1.125°
RCM Modeling center			
CLMCom-ETH	COSMO-crCLIM		0.22° × 0.22°
Observations and Reanalysis data sets			
APHRODITE	–	–	0.25° × 0.25°
GPCC	–	–	0.5° × 0.5°
HadISST	–	–	1° × 1°
NCEP Reanalysis	–	–	1.875° × 1.875°
ERA-Interim Reanalysis	–	–	0.5° × 0.5°
MERRA Reanalysis	–	–	0.5° × 0.65°

4.4 Results and discussion

In the current section, first, we discuss the results of two-source IE obtained from various idealized linear dynamical systems mentioned in Section 4.2. Thereafter, we present results of two-source IE in the climate system with the observations, reanalysis data sets, GCM simulations, and the RCM simulations.

4.4.1 Applications to idealized systems

First, we will start with the discussion of results obtained from idealized systems with various IE estimators.

Linear autoregressive system

Figure 4.2 shows the information exchange (in nats) from y_{t-1} (immediate past of y) to x_t (present of x) and also from x immediate past to present of x (i.e., x_{t-1} to x_t), for the system with Equation 4.4. The two-source mutual information linear estimator shows that as the coupling coefficient increases, the IE from $I(x_t; y_{t-1}, x_{t-1})$ increases, indicating that the immediate pasts of x_{t-1} and y_{t-1} exchange information to the future state of x as expected from the system dynamics. Also, as expected the $I(x_t; y_{t-1}, x_{t-1}) > I(x_t; y_{t-1})$ or $I(x_t; x_{t-1})$, indicating that the two-source IE dominates the dynamics of this system. The IE from the immediate past of x i.e., x_{t-1} is a stronger source of information

to the target x_t due to self feedback/large persistence and y_{t-1} is a weaker source to the target x_t (this behavior is often observed in the climate system where persistence/self feedback plays an important role (Runge et al. 2014). The error bars represents two standard deviations of the 100 permuted surrogates showing the measure of uncertainty for the IE estimations. Furthermore there exists a significant positive net synergy (ΔI) indicating that the two sources at higher couplings exchange synergistic information to the target even though the two sources y_{t-1} and x_{t-1} are uncorrelated with each other, in other words, a certain degree of uncertainty about the system x_t is reduced by knowing the state of x_{t-1} and y_{t-1} together. Here in this system, the synergy between the two sources (y_{t-1} and x_{t-1}) to the prediction of target (x_t) might be arising from their linear combination. This shows that linear systems can exhibit synergies, which is also shown analytically in the work by Barrett (2015). The non-linear estimators, i.e., Kraskov estimator (40 k-nearest neighbors) and Kernel estimator (1.5 kernel width) also show the similar system behavior. The free parameters i.e., kernel width (1–2 kernel widths) and number of k-nearest neighbors (20–60 neighbors) are tested and tuned for consistent and robust results.

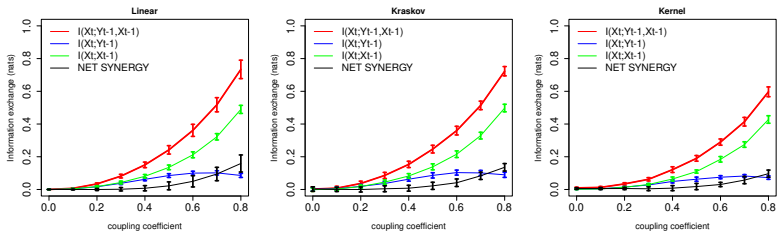


FIGURE 4.2. Information exchange in nats from two-source (red line), single source (green and blue lines), and net synergy (black line) to the target with Linear, Kraskov and Kernel estimators. The error bars represents two standard deviations of the 100 permuted samples.

Next, we tested another system consisting of two subsystems, coupled with each other but only having a single source as in Equation 4.5. From Figure A.1 (in appendix A), the MI linear metrics shows that $I(x_t; y_{t-1}) = I(x_t; y_{t-1}, x_{t-1})$ indicating that the immediate pasts of x_{t-1} does not contribute to IE for the target x_t. The net synergy from y_{t-1}, x_{t-1} to the target x_t is as expected zero. The IE from y_{t-1} to x_t increases as the coupling coefficient increases, which is also expected. This is also seen in Kraskov estimator (40 k-nearest neighbors) and Kernel estimator (1.5 kernel width). The free tuning parameters are tested and tuned for consistent results. Finally, among the linear systems, we tested a three-dimensional system (similar to the situation

of ENSO, IOD influencing ISMR variability) with the Equation 4.6. Figure A.2 shows that the information exchange from $I(x_t; y_{t-1}) = I(x_t; z_{t-1})$ indicating that the two sources contribute to the target system equally and moreover the IE increases with increase with coupling coefficient. This behavior is expected as observed from the governing equations. Even though the two sources are uncorrelated with each other, they exhibit positive net synergy. The similar behavior in the system is seen with non-linear Kernel estimator (1.5 kernel width) and Kraskov estimator (40 k-nearest neighbors). The free parameters are tested and tuned for consistent results. The results for non-linear system are discussed in Appendix section.

The results from idealized linear and non-linear examples show that some systems do exhibit positive net synergy from two-sources to target for both linear as well as non-linear systems, even when the two sources are uncorrelated. Furthermore, all the three estimators mentioned above i.e., Linear, Kernel and Kraskov estimators are able to detect consistently the two-source information exchange.

4.4.2 *Application of dual-source IE to climate phenomenon*

In this section, we examine the two-source IE from ENSO, IOD to the inter-annual variability of ISMR. Foremost, we present results obtained from the observational, reanalysis data sets and then extend our analysis of two-source IE to three GCM simulations as mentioned in Table 4.1 . Thereafter, we present results from our dynamically downscaled simulations with COSMO-crCLM with the three GCMs as driving models.

Observation and reanalysis data

In the observations and reanalysis data sets, empirical orthogonal function (EOF) analysis of the detrended SST anomalies is performed over the tropical Indian ocean (25°S–20°N,50–120°E) and the tropical Pacific ocean (25°S–25°N,120°E–80°W) to obtain the major oscillations and their respective PCs. The ENSO and IOD indices are taken as the time series associated with their respective PCs obtained from the EOF spatial patterns replicating them. Figure 4.3 shows the second EOF patterns of the SST anomalies over the Indian ocean and first EOF patterns over the Pacific ocean for HadISST and from NCEP reanalysis. From the two SST data sets, it is observed that both ENSO and IOD like structures are captured with the second EOF and the first EOF patterns i.e., a zonal dipole like structure in the Indian ocean and the Pacific ocean respectively. We use EOF analysis as opposed to standard indices such as the dipole mode index known as DMI (Saji et al. 1999) and Niño-3.4 to allow each

model to exhibit their own patterns as opposed to an imposed structure (Saji et al. 2006, Cai, Cowan & Sullivan 2009, Cai, Sullivan & Cowan 2009, Liu et al. 2011).

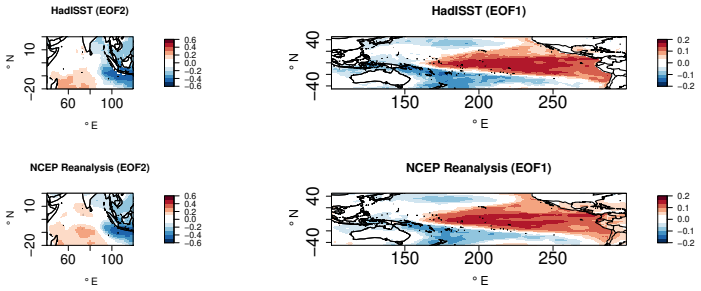


Figure 4.3. EOF2 patterns of SST anomalies (JJAS) in the Indian ocean and EOF1 patterns in the Pacific ocean for observed HadISST and NCEP reanalysis.

To ensure that the EOF patterns in the observed SST data sets replicate the ENSO and IOD modes, the obtained PCs are compared against the corresponding Niño 3.4 and IOD index obtained from the NOAA ESRL Physical Sciences Division, and JAMSTEC observations (shown in Figure A.3). These indices are widely used in several studies concerning the IOD and Niño 3.4 teleconnections. The percentage of the total variance contributed by the first 20 EOFs from the Indian and Pacific ocean SST anomalies for the seasons JJAS are also shown in Figure A.3. The linear fit between the Indian ocean PCs of EOF-2 obtained from the HadISST against the observed IOD index has a correlation of about 0.78, and the correlation of NCEP reanalysis SST with the observed IOD index is 0.77. These results are significant at a 99 % confidence level. This indicates that the EOF2 replicates the IOD like variability for the two mentioned datasets. The percentage of the total variability contributed by the EOF1 of the Indian ocean is about 30% which is associated with the basin-scale anomalies of uniform polarity in the Indian ocean associated with the ENSO events. The dipole mode (EOF2) explains about 15% of the total variance which is associated with the IOD. Our results for the Indian ocean EOF patterns and their respective contribution to the total variance are consistent with the study by Saji et al. (1999). Similarly, the PCs associated with the first EOF over the Pacific ocean are highly correlated against the observed Niño 3.4 index with a correlation value greater than 0.8 for both data sets indicating that the EOF1 captures the ENSO like variability. The percentage of total variance contributed by the first EOF $\approx$ 20% is also consistent with the ENSO literature.

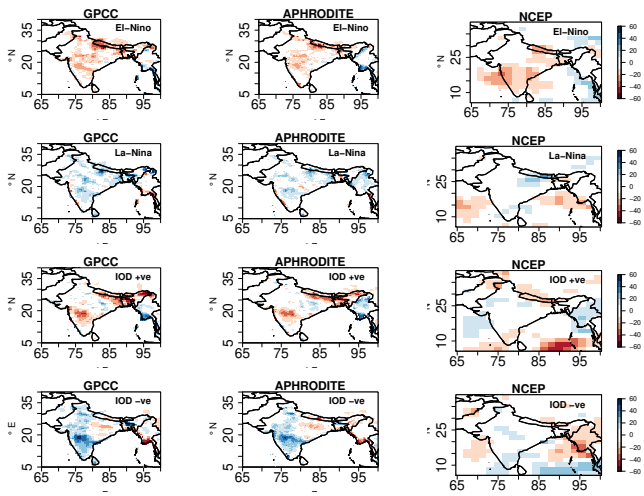


FIGURE 4.4. Total precipitation anomaly (mm/month) composites (JJAS) over the Indian subcontinent for El-Niño, La-Niña, positive IOD and negative IOD events observed in GPCC, APHRODITE and NCEP reanalysis data sets for the period of 1951-2005

The ENSO and IOD are known to influence the ISMR distribution across the Indian subcontinent. Hence to investigate the rainfall anomaly distribution during various phases of ENSO and IOD (i.e., El-Niño, La-Niña, IOD+ve, and IOD-ve), we plotted the anomaly composite figures (Figure 4.4) for the ISMR during these events. The anomalies are constructed by subtracting the Indian subcontinent climatology mean JJAS rainfall with the rainfall months associated with various phases of IOD and ENSO. The anomaly composites with El-Niño (La-Niña) events show that most parts of Indian subcontinent receive less (more) rainfall during the El-Niño (La-Niña) phases. This behavior can be attributed to the suppression of convection over the Indian subcontinent during the El-Niño phase through the zonal and meridional circulation and vice-versa during La-Niña phase. The rainfall anomaly composites associated with the positive and negative phases of the IOD represent distinct regional asymmetric rainfall anomalies i.e., a meridional tripolar pattern, with above than normal rainfall in central parts of India and below than normal rainfall to the north and south of it. Conversely, the negative IOD is associated with a zonal dipole having above (below) normal rainfall on the western (eastern) half of the Indian subcontinent. These results with rainfall composites during IOD phases are consistent with Behera & Ratnam (2018), where it was concluded that these rainfall anomaly patterns are due to the differences in the atmospheric responses and the associated differences in moisture transports to the region during contrasting phases of the IOD. Hence, Figure 4.4 indicates that both ENSO and IOD contribute to the interannual variability of the IMSR.

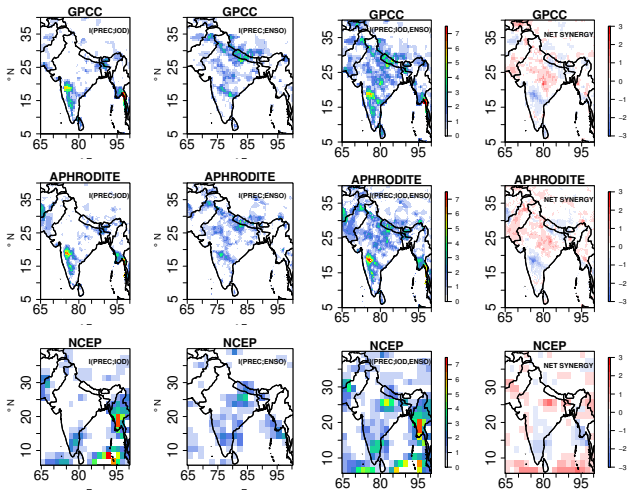


FIGURE 4.5. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$, two-source information exchange $I(PREC;ENSO,IOD)$ and NET SYNERGY $\times 10^{-2}$ nats for observational data sets GPCC, APHRODITE and NCEP reanalysis. Only significant values at 95% confidence intervals are plotted.

Figure 4.5 represents the IE from the IOD to precipitation i.e., $I(PREC;IOD)$, ENSO to precipitation i.e., $I(PREC;ENSO)$, the two-source IE i.e., $I(PREC;IOD,ENSO)$ together with the NET SYNERGY for the observations GPCC, APHRODITE, and the NCEP reanalysis data sets under linear approximation. We chose various precipitation data sets to accommodate uncertainties due to the sparse data networks, especially in regions with complex topography. The observed IE from IOD to total precipitation i.e., $I(PREC;IOD)$ shows that the IOD transmits information to the southwest sector of the Indian subcontinent especially the lee-ward side of the western ghat regions in GPCC and APHRODITE data sets. This feature is slightly shifted to the east in the NCEP reanalysis data sets. All the IE plotted values are significant at 95% confidence level obtained from 100 surrogate samples. Some regions in the northeast sector also are influenced by the IE from IOD which is replicated in all three observational data sets. It is interesting to note that the location at which the IE from IOD to the precipitation over the Indian subcontinent matches the significant rainfall anomalies shown in Figure 4.4. The $I(PREC;ENSO)$ shows that the northern parts of the Himalayas, central India receive information from the Pacific ocean in all the three data sets, this also matches the anomaly locations shown in Figure 4.4. The two-source information exchange covers most parts of the Indian subcontinent indicating that both ENSO and IOD contribute to the ISMR during JJAS seasons. Also, interestingly from the NET SYNERGY plot, a positive net synergy over certain parts of central India also known as monsoon

core region is observed, indicating that both ENSO and IOD synergistically contribute to the interannual variability of ISMR. Furthermore, the ENSO and IOD share net redundant information (negative net synergy) in the southern sector of the Indian sub-continent. The Kraskov estimator (Figure A.4) and the Kernel estimator (Figure A.5) also show similar IE patterns over the Indian subcontinent with 40 k-nearest neighbors for Kraskov and 0.5 kernel width for Kernel estimators (free parameters are tested and tuned for consistent results). In addition, we also checked the two-source IE patterns in the two reanalysis datasets, MERRA and ERA-Interim (1980-2005), shown in Figure A.6 and Figure A.7. It is found that in both the data sets, similar IE patterns are replicated i.e., positive net synergy in central India and net redundant information in southern part of the Indian subcontinent. We also did a similar analysis for the months of DJFM, our results show that the net synergy from IOD and ENSO to the rainfall is absent (Figure A.8–Figure A.13). This is expected as the IOD mode during these months is dissipated and absent.

The net synergy between the ENSO and IOD to the ISMR interannual variability in JJAS indicates that the central India monsoon rainfall predictability lies in knowing the states of ENSO and IOD together than by knowing the states of ENSO and IOD individually (similar to the idealized test case example 3). This is also exactly similar to the XOR logic gate, where the uncertainty of the output is known only with the simultaneous knowledge of the two input states. To understand the information synergy physically, we show the moisture transport figures from the NCEP reanalysis datasets for various phases of ENSO and IOD during the JJAS. From Figure 4.6 it is observed that the anomalous negative moisture flux during the El-Niño is compensated with the positive moisture flux anomaly by IOD +ve especially in central India, and vice-versa during the La-Niña and IOD-ve events. It is known that the El-Niño events are often associated with IOD+ve events (Behera & Ratnam 2018) and vice versa (the ENSO and IOD are positively correlated in our data sets). From the precipitation composites (Figure 4.3), in central India, an anomalous negative (positive) rainfall during the El Niño (La Niña) is observed, and during the IOD+ve (IOD-ve) a positive (partly negative) anomalous rainfall is observed. This could explain why both the IOD and ENSO states should be known together to explain the variability of the central Indian subcontinent rainfall as the IOD and ENSO are having compensating effects. This compensating behavior is not seen in the southern or northern part of the Indian subcontinent, hence this could explain the net redundant information between ENSO and IOD to the precipitation to the southern region. The readers are referred to Fig.3 by Barrett (2015) to further explore the relation of synergy dependence on the compensating influence from both sources, i.e., the correlation between two sources and to their targets respectively.

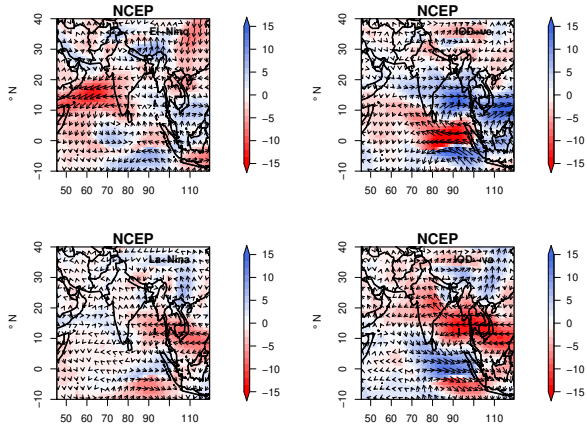


Figure 4.6. Moisture flux anomalies (g/kg m/sec) over the Indian subcontinent (JJAS) for El-Niño, La-Niña, IOD+ve and IOD-ve events observed in NCEP reanalysis data sets for the period of 1951-2005.

Global and regional climate model simulations

Next, we are performing the same analysis, starting with the EOF patterns from the SST fields obtained from the three GCM simulations listed in Table 4.1, to investigate how the ENSO and IOD associated variability in the Indian and Pacific oceans are represented. Figure 4.7 shows the second EOF pattern of the SST anomalies over the Indian ocean and the first EOF pattern in the Pacific ocean, for the GCM simulations of MPI-ESM-LR, Nor-ESM-M, and EC-EARTH. It is found that all the GCM simulations replicate the zonal dipole like patterns over the Indian ocean and Pacific ocean similarly as the observations. The percentage of the total variability contributed by EOF1 of the Indian ocean is about 30% in all the GCM simulations (Figure A.14) which is comparable to the observations. The EOF2, which is associated with the IOD, explains about 15% of the total variance in all the GCMs, also similar to observations. The percentage of total variance contributed by the first EOF is between $20 - 25\%$ in all the GCM simulations in the Pacific ocean, which is similar to variance in the observations. Thus, these results indicate that the variability associated with the SST anomalies over the Indian and the Pacific ocean is represented in the three GCM simulations. The SST anomaly composites during various phases of IOD and ENSO events (Figure A.15 and Figure A.16) show that most of the GCM simulations can replicate the SST anomaly composite patterns found during the IOD+ve events in HadISST (Figure A.15). On the contrary, during IOD-ve events, the MPI-ESM-LR portrays unrealistic warm anomalies throughout the Indian ocean. Over the Pacific ocean, the MPI-ESM-LR and

EC-EARTH have an unrealistic westward extension of the warm (cold) pool during El Niño (La Niña) events. The patterns from Nor-ESM-M are closer to the observation, shown in Figure A.16. The unrealistic westward extension of the SSTs in EC-EARTH and MPI-ESM-LR simulations might influence the walker circulation through unrealistic large scale teleconnections patterns.

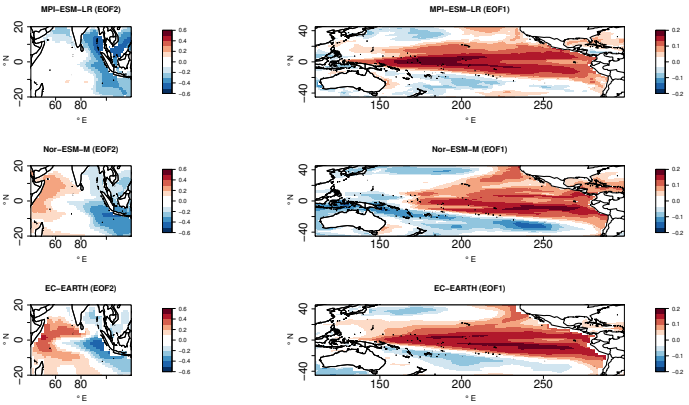


FIGURE 4.7. EOF2 patterns of SST anomalies for (JJAS) in the Indian ocean and EOF1 patterns for (JJAS) in the Pacific Ocean for three GCM simulations, i.e., MPI-ESM-LR, Nor-ESM-M and EC-EARTH for the period of 1951-2005.

Figure 4.8 represents the ISMR anomaly composites during the El-Niño, La-Niña, IOD+ve and, IOD-ve events for the three GCM simulations, the MPI-ESM-LR, Nor-ESM-M, and EC-EARTH, when selecting the associated years given by the respective PCs. The rainfall anomaly composites associated with the positive phase of ENSO show dry conditions over the northern/northwest parts of the Indian subcontinent in the MPI-ESM-LR, dry conditions throughout the Indian sub-continent in Nor-ESM-M. The EC-EARTH simulation does not show a clear rainfall anomaly signal. Similar opposite polarity of rainfall anomalies are observed in the La-Niña conditions in the MPI-ESM-LR and Nor-ESM-M simulations, while slight wet conditions in north-east India in EC-EARTH. For the IOD+ve events, MPI-ESM-LR shows dry conditions in the southwest, while the Nor-ESM-M simulation shows dry conditions in the northwest and the Himalayan region, the EC-EARTH does not show any variability. The Nor-ESM-M during the IOD-ve phase shows overall positive anomaly, while no clear signal is observed in MPI-ESM-LR and EC-EARTH. Overall, the ENSO phase signal is better replicated in Nor-ESM-M simulation and partly in MPI-ESM-LR as in the observations, while most of the GCM simulations failed to replicate the regional rainfall asymmetric response in IOD events as in observations (except Nor-ESM-M, which partly can replicate the

dipole patterns). This might be due to the coarse resolution of GCMs which may not be able to replicate the fine-scale precipitation response to the IOD.

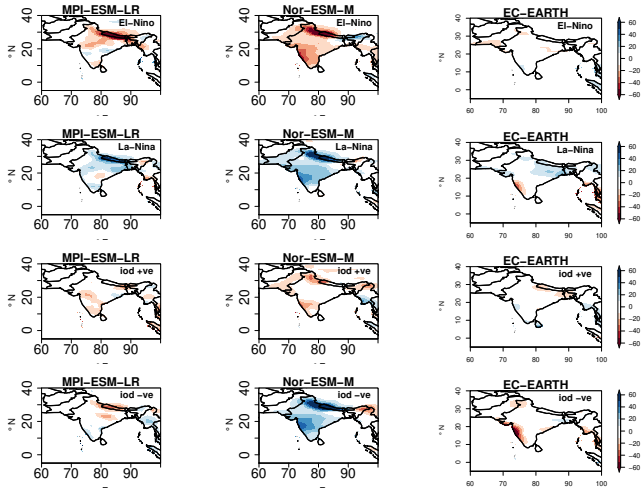


Figure 4.8. Total precipitation anomaly composites over the Indian subcontinent (JJAS) for El-Niño, La-Niña, positive IOD and negative IOD events in MPI-ESM-LR, Nor-ESM and EC-EARTH simulations(1951-2005)

Figure 4.9 represents the IE spatial patterns from the IOD and ENSO i.e., $I(PREC;IOD)$, $I(PREC;ENSO)$, the two-source IE, $I(PREC;IOD,ENSO)$ together with the NET SYNERGY over the Indian subcontinent in the three GCM simulations i.e., MPI-ESM-LR, Nor-ESM-M, and EC-EARTH with the linear estimator. The information exchange from IOD to total precipitation in MPI-ESM-LR shows that the information from the IOD is exchanged to the southeastern part of the Indian Subcontinent. This is contrary to what is seen in the results from the observations, where most of the IE takes place to the leeward side of the western ghats and the northeastern sector of India. The Nor-ESM-M simulation shows that IE from IOD is transmitted to the western side of the Indian subcontinent, where the observed significant anomalies are noted in Figure 4.9. The EC-EARTH does not show any information exchange from IOD to the land points over the Indian sub-continent. The $I(PREC;ENSO)$ show that the northern parts of the Himalayas and north west-central India receive information from the Pacific ocean in MPI-ESM-LR. For Nor-ESM-M, the western ghats and its leeward side are influenced by ENSO. The EC-EARTH does not show as much IE as the Nor-ESM-M or MPI-ESM-LR over the Indian continent, with an exception for some scattered locations over the Himalayas.

The two-source information exchange $I(PREC;ENSO,IOD)$ covers the northwest part of the Indian subcontinent for MPI-ESM-LR and the extreme

southeast. For Nor-ESM-M the information exchange covers mostly the western part of India. The EC-EARTH show IE over isolated places of northeast India. These results indicate that the three GCMs exhibit a IE pattern which is different from the observed patterns. Moreover, the results of the NET SYNERGY show that MPI-ESM-LR does not show any net synergistic IE over the Indian subcontinent, while in Nor-ESM-M the IOD and ENSO share common information over the west of India. EC-EARTH show less net synergy over the Indian sub-continent. Overall, the results from the IE exchange differ from the observations, seen for all the three GCM simulations. These results are consistent with Kernel and Kraskov estimators (Figure A.17).

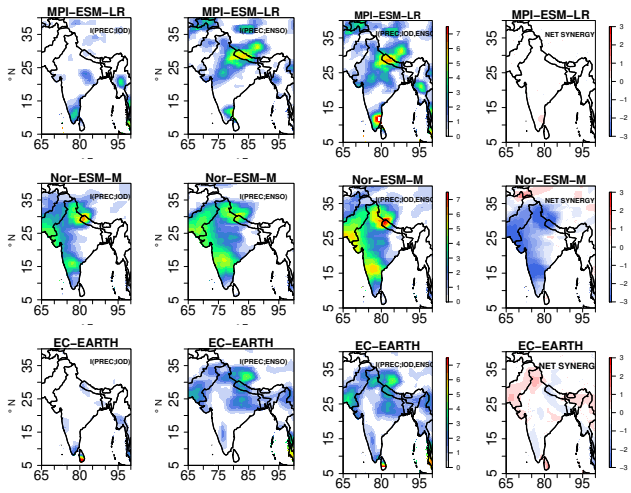


FIGURE 4.9. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$, two-source information exchange $I(PREC;ENSO,IOD)$ and NET SYNERGY $\times 10^{-2}$ nats for the GCM simulations MPI-ESM-LR, Nor-ESM-M and EC-EARTH for JJAS (1951-2005). Only significant values at 95% confidence intervals are plotted.

Next, we are investigating how the two-source information exchange is represented when we dynamically downscale the three GCM simulations (MPI-ESM-LR, Nor-ESM-M, and EC-EARTH) with the regional model COSMO-crCLIM (0.22°). We are applying the same two-source information exchange method on the RCM fields as we have done for the GCM simulations. However, since the RCM simulations are only covering a limited area, namely the South Asian CORDEX domain, we had to combine the RCM results with the GCM simulations, in particular for the EOF-analysis over Indian and Pacific oceans. Figure 4.10 represents the ISMR anomaly composites during the positive IOD+ve, IOD-ve, El-Niño, and La-Niña events for the COSMO-crCLM RCM simulation driven with three GCM simulations, the MPI-ESM-LR, Nor-ESM-M, and EC-EARTH. Here we are selecting the same years as given by the

principal components from the driving GCM simulations. The rainfall anomaly composites associated with the El-Niño events show dry conditions over the northern parts of Himalayas for the downscaled MPI-ESM-LR and wet conditions in western ghats and isolated parts in central India. During the La-Niña phase, dry conditions in the central Indian subcontinent, western ghats and wet conditions elsewhere are observed. In the downscaled Nor-ESM-M, dry (wet) signal is observed throughout Indian subcontinent during El-Niño (La-Niña) phases. In the downscaled EC-EARTH, dry regions are noted throughout most parts of Indian subcontinent during El-Niño, while dry conditions are seen in central India and wet conditions elsewhere in La-Niña phase. The rainfall anomalies composites associated with the positive IOD in the observations, i.e., a meridional tripolar pattern with above than normal rainfall in central parts of India and below than normal rainfall to north and south of it is only observed in the downscaled Nor-ESM-M. Similarly, the negative IOD in downscaled Nor-ESM-M is associated with a zonal dipole having above (below) normal rainfall on the western (eastern) half of India similar to that of the observations as seen in Figure 4.5. Overall, these results suggest that the downscaled results from Nor-ESM-M better reproduces the spatial patterns of precipitation anomalies associated with ENSO and IOD, when comparing to the observations, than the downscaled results from EC-EARTH and MPI-ESM-LR.

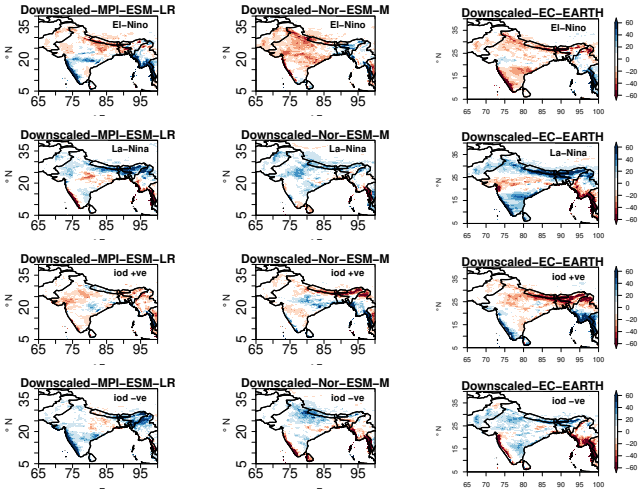


Figure 4.10. Total precipitation anomaly composites over the Indian subcontinent for El-Niño, La-Niña, positive IOD and negative IOD events for the downscaled COSMO-crCLM simulations driven by MPI-ESM-LR, Nor-ESM-M and EC-EARTH GCM simulations for JJAS (1951-2005)

Figure 4.11 represents the IE patterns over the Indian subcontinent for the downscaled RCM simulations with the Linear estimator (these patterns are

also consistent with Kraskov (Figure A.18) and Kernel (Figure A.19) estimators). The net synergy in central India, and shared information in southern India is better represented in the downscaled Nor-ESM-M simulation, compared to the downscaled MPI-ESM-LR and downscaled EC-EARTH. This is in agreement with the results from the GCM simulation, where it was found that Nor-ESM-M simulation had a better replication of ENSO and IOD induced anomalous precipitation structures than the two other GCMs (see Figure 4.10). These results are interesting, even though all the COSMO-crCLM simulations have the same physics and dynamics, only downscaled Nor-ESM-M replicated realistic patterns of IE. The improvement in results in downscaled Nor-ESM-M can be attributed to a more realistic large-scale information coming from the GCM simulation, such as the moisture flux transport during various phases of ENSO and IOD events (see Figure A.20 – Figure A.24 and Figure 4.6). For the MPI-ESM-LR and EC-EARTH GCM simulations, the moisture flux anomalies are very different from the reanalysis fluxes and thus seem misrepresented. A better replication of the moisture flux anomaly in Nor-ESM-M GCM simulation during ENSO and IOD might be from a better simulation of the large scale circulation patterns, like the Walker and Hadley circulations, due to the better representation of the SST than the two other GCM simulations (Figure A.8 and Figure A.9). The RCM simulation results exhibit similar moisture flux anomalies compared to the driving GCM simulations, in which the downscaled Nor-ESM-M outperforms the downscaled MPI-ESM-LR and downscaled EC-EARTH. These results indicate that a realistic large-scale signal from the GCM simulations (e.g., the moisture transport and SST anomalies) is essential for an RCM to properly improve the GCM results in terms of IMSR variability. When the large-scale signal from the GCM is incorrect, and wrong moisture fluxes are imposed on the lateral boundaries of the RCM, the downscaled results are hampered.

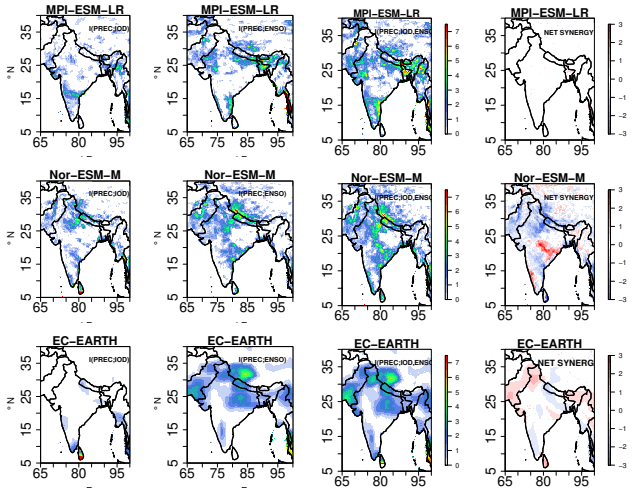


FIGURE 4.11. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$ and two- source information exchange $I(PREC;ENSO,IOD)$, NET SYNERGY $\times 10^{-2}$ nats for the downscaled COSMO-crCLM simulations for JJAS (1951-2005). Only significant values at 95% confidence intervals are plotted.

4.5 CONCLUSIONS

In this article, we explored two-source information exchange (IE) from ENSO and IOD (quantified by SST variabilities in the Pacific and Indian oceans) to the Indian Monsoon Summer Rainfall (IMSR) interannual variability. But, first, we used simple idealized linear and non-linear dynamical systems to demonstrate the concepts of two-source IE. Results showed that both the linear and the non-linear idealized systems can exhibit positive net synergy (i.e., the combined influence of two sources is greater than their individual contributions). Interestingly, two uncorrelated sources can show positive net synergistic IE to a target.

The two-source ENSO and IOD to IMSR IE was explored in observations, reanalysis data sets, and in three GCM simulations which were also further dynamically downscaled with the RCM. The results from the observations and reanalysis data suggest that both IOD and ENSO influence the interannual variability of the ISMR throughout most parts of Indian subcontinent. Interestingly, we found that IOD and ENSO exhibit positive net synergy over central India, which is the monsoon core region, and net redundant information over the southern part of India.

The IE patterns in the three GCM simulations differ from that in the observations. However, the GCM Nor-ESM-M better captured the precipitation anomalies from ENSO and partly from IOD than the other two GCMs. Previ-

ous studies also showed that Nor-ESM-M outperforms other CMIP5 GCM simulations in terms of rainfall climatology, and most aspects of the climatological annual cycle and interannual variability in the Indian subcontinent (Sperber et al. 2013, McSweeney et al. 2015).

Downscaling Nor-ESM-M simulation with the RCM COSMO-crCLM better replicated the observed IE patterns than downscaling the MPI-ESM-LR and EC-EARTH simulations. Importantly, the downscaled Nor-ESM-M IE results are in better agreement with the observations than the Nor-ESM-M results. Downscaling Nor-ESM-M adds value to the GCM simulation. This can not be concluded here for downscaling of MPI-ESM-LR and EC-EARTH simulations. Downscaling the latter simulations did not add value because of a missing realism in their large-scale SST patterns and horizontal moisture flux variability, which are important RCM boundary conditions and which were better represented in the Nor-ESM-M simulation. Downscaling did not compensate errors in the large-scale driving simulations. These results highlight the importance of the choice of GCM simulations when performing dynamically downscaling for high-resolution regional climate projections.

Finally, we propose to use the two-source IE metric as a complementary tool to gain additional insight into the climate system and to perform process-oriented climate model evaluation.

4.6 Acknowledgments

The authors acknowledge the support by the German Research Foundation ("Deutsche Forschungsgemeinschaft", DFG) in terms of the research group FOR 2416 "Space-Time Dynamics of Extreme Floods (SPATE)". The authors also thank Joseph T Lizer for providing the JDIT open source toolkit. CLMcom-ETH-COSMO-crCLIM-v1-1 simulations were run on Piz Daint at CSCS (Switzerland), and we acknowledge PRACE for awarding us access and computing time to Piz Daint. We thank two reviewers for their constructive comments and insights.

Chapter 5

Vb-cyclones and associated North-Western Mediterranean Sea state in regional coupled climate simulations: evaluation and projection[3]

Abstract

Vb-cyclones propagating from the North-Western Mediterranean Sea (NWMS) into central Europe are often associated with extreme precipitation. This study explores the state and process chain linking the NWMS state and the Vb-cyclone precipitation in the Danube, Elbe, and Odra catchments in regional coupled atmosphere-ocean climate simulations with COSMO-CLM+NEMO. Two high-resolution simulations, an evaluation simulation (1951-2005) downscaling the centennial ERA-20C reanalysis and a continuous simulation (historical 1951-2005 + RCP8.5 future scenario 2006-2099) downscaling the EC-EARTH global climate data set are used for this purpose. The results show a good agreement in mean annual Vb-cyclone frequency between the evaluation (9.7 events/year) and the historical (10.1 events/year) simulations. But, there are significant discrepancies in the seasonal cycle. The mean cyclone intensity measured with minimum central pressure, track density, and precipitation rankings in the three catchments also show good agreement. The simulations for the future period show a basin-average SST warming of $\approx 2.5 - 3$ K by the end of 21^{st} century, but insignificant changes in Vb-cyclone frequency, mean intensity, and precipitation in the selected catchments. The NWMS sea surface temperature, evaporation, and wind speed anomalies corresponding to the Vb-cyclone precipitation rankings differ between the evaluation and historical simulations. In the evaluation simulation, Vb-cyclone precipitation rankings correspond with sea surface temperature, evaporation, and wind speed anomalies, while in the historical and the future simulation no such correspondence is seen. Especially the Adriatic and Ionian basins in the simulation driven by EC-EARTH show no sensitivity to the Vb-cyclone precipitation over the catchments. The change in the processes between evaluation and historical simulations might be due to the emergence of biases inherited from the driving EC-EARTH global simulation. The future simulation shows no significant process changes compared to the historical simulation.

[3] Submitted as: *Pothapakula, P. K., Krug, A., Anika, O.H., Timo, K., Ahrens, B. (2022). Vb-cyclones and associated Mediterranean Sea state in regional coupled climate simulations: evaluation and projection* (Earth System Dynamics)

5.1 Introduction

Observational and modeling studies relating to the global temperature and precipitation changes provide confidence in the current ongoing global warming (Stocker et al. 2013). Changes in the extreme weather and climate events such as warm/cold days and nights, heat waves, droughts, heavy precipitation events induced by the anthropogenic global warming were observed in the last century (Fischer et al. 2013, Wilcox & Donner 2007, Trenberth 1999, Nishant & Sherwood 2021, Beniston et al. 2007, Seneviratne et al. 2012). Specifically, short-term precipitation extremes often result in heavy damage to infrastructure and life, and hence are in need of further investigations (Hochman et al. 2022, Mathias et al. 2021).

Over central Europe, extra-tropical cyclones named Vb-cyclones are often associated with extreme precipitation, especially in the summer season (Hofstätter et al. 2018, Blöschl et al. 2013). These Vb-cyclones develop over the North-Western Mediterranean Sea (NWMS) typically over the Gulf of Lions and travel northeastward through the eastern Alps to central Europe (van Bebber 1891, Messmer et al. 2015). Often, extreme precipitation occurring in the catchments of Danube, Elbe and Odra is linked to the Vb-cyclones (Krug et al. 2022).

Though the occurrence of Vb-cyclones is rare (typically about 4-10 events on average per year) they are of considerable importance due to the extreme precipitation they bring to central and eastern Europe (Hofstätter et al. 2016, Messmer et al. 2015). These events occur throughout the year with a peak frequency in spring (Hofstätter et al. 2016, 2018). The Vb-cyclones are typically fed from the evaporation over continental land and nearby oceans. For example, enhanced evaporation over the Mediterranean Sea and subsequent increase of available total water content in the atmosphere during Vb-cyclones was studied by Hofstätter & Chimani (2012) and Messmer et al. (2017). Based on a model sensitivity study, Volosciuk et al. (2016) reported an increase in precipitation by 17% over central Europe with warmer sea surface temperatures (SST) over the Mediterranean Sea compared with a simulation run by average Mediterranean SST for the period 1970-1999. Results from Volosciuk et al. (2016) relied on a stand-alone coarse-resolution global atmosphere model without dynamic coupling of the ocean, missing crucial air-sea feedback processes. The role

of the Mediterranean Sea in enhancing the August 2002 flood in the early stages was shown in Sodemann et al. (2009), James et al. (2004), Gangoiti et al. (2011). Furthermore, a sensitivity study by Messmer et al. (2017) confirmed the Mediterranean Sea role in supplying moisture to Vb-cyclones. However, other studies concluded that the moisture transport from the North Atlantic, Black Sea, continental moisture are major sources contributing to the precipitable water for Vb-cyclones (James et al. 2004, Gangoiti et al. 2011, Ho-Hagemann et al. 2015, 2017, Krug et al. 2022).

Krug et al. (2022) analyzed about 1107 Vb-cyclone events simulated by a regional coupled climate model during the period 1901-2010. Their study concluded that the NWMS played an active role in the early stage intensification of the Vb-cyclones and also in pre-moistening the continental land. Furthermore, high precipitation Vb-cyclone events were associated with anomalously high dynamically driven evaporation. However, with Lagrangian moisture source diagnostics on selected 16 Vb-cyclones, their study revealed that continental moisture recycling, the north Sea, the Baltic Sea, the north Atlantic, and the Black Sea were major sources of moisture supply to the Vb-cyclones. Krug et al. (2021) reported a significant information exchange between the evaporation over the NWMS and Vb-cyclone precipitation over the Odra catchment. Above mentioned studies highlight the NWMS's role in modulating Vb-cyclone intensity.

As the role of the Mediterranean Sea in modulating the Vb-cyclone events is well established in the literature for the historical period, investigating the Vb-cyclone's future projections and the role of the Mediterranean Sea in the warming climate becomes extremely important. Especially, the Mediterranean Sea being a hot spot of climate change, investigating its role in the future Vb-cyclones is also of extreme importance. Using a Global Climate Model (GCM), ECHAM5/OM1, Nissen et al. (2013) reported a decrease in the number of Vb-cyclones by the end of the 21^{st} century but an increase in precipitation amount by 16% compared to the present. Messmer et al. (2020), investigated the climate change impacts on Vb-cyclone characteristics using a global climate model, i.e., Community Earth System Model ensemble simulations. Their results confirmed a minor decrease in the frequency of Vb-cyclones from 2.9 to 2.6 Vb-cyclones per year by the end of 21^{st} century. They also found a subtle eastward shift in the Vb-cyclone frequency pattern. Furthermore, by downscaling the 10 heaviest precipitation Vb-cyclone events with the Weather Research and Forecasting model in future and historical periods, they reported insignificant changes in the total precipitation amount. It is to be noted that the study by Messmer et al. (2020) downscaled only 10 Vb-cyclone heavy precipitation events in the past and future periods.

To simulate mesoscale systems such as the Vb-cyclones and analyze their

climatic characteristics, high-resolution regional climate model simulations in long centennial periods are desirable. Mittermeier et al. (2019) using a Canadian Regional Climate Model Large Ensemble (Leduc et al. 2019) at a resolution of about 12 km (0.11°) studied future Vb-cyclone frequency and precipitation changes over Bavaria from 1950-2099. They reported a non-significant increase in the absolute number of Vb-cyclones per year in the future period. Also, a significant decrease in future summer Vb-cyclone frequency and increase in spring. In terms of Vb-cyclone daily precipitation intensity, a significant increase was reported over Bavaria.

The Coordinated Regional Climate Downscaling Experiment (CORDEX) is an initiative that coordinates scientific groups for high-resolution regional climate data sets (Giorgi 2006). The regional climate model, Consortium for Small-scale Modelling in Climate Mode (COSMO-CLM, Rockel & Geyer (2008)) is used for dynamically downscaling GCMs over various CORDEX domains (Asharaf & Ahrens 2015, Russo et al. 2020, Drobinski et al. 2020, Evans et al. 2021). The added value of such high-resolution simulations was well documented in the studies by Schlemmer et al. (2018), Imamovic et al. (2019), Panosetti et al. (2019), Hentgen et al. (2019), Brogli et al. (2019), Sørland et al. (2021). However, the COSMO-CLM often uses prescribed SSTs from the driving GCMs which are handicapped by their coarse resolution and unrealistic air-sea dynamic interactions (Akhtar et al. 2018). Especially given the importance of the Mediterranean Sea in the evolution of Vb-cyclone events, a realistic and dynamically interactive ocean model is thus necessary.

The COSMO-CLM is coupled to Nucleus for European Modeling of the Ocean (NEMO, Madec (2008)) over the Mediterranean sea (NEMOMED12, Brossier et al. (2011, 2012)) along with a river run-off model, Total Runoff Integrating Pathways (TRIP) to make the regional system dynamically interactive (Akhtar et al. 2018). The added value of such a coupled regional system was reported by Kelemen et al. (2019) on the representation of European continental precipitation. Furthermore, Primo et al. (2019) reported the added value in terms of extreme air temperatures. This coupled system was earlier used to study Vb-cyclones in the historical period by Akhtar et al. (2019) and Krug et al. (2022). Their study demonstrated the ability of the coupled system in representing the past Vb-cyclone events realistically. Furthermore, Krug et al. (2022) analyzed the total precipitation sums of a few selected Vb-cyclone events (1901–2010) simulated by the coupled system driven by the ERA-20C reanalysis. They reported that the coupled system precipitation patterns and magnitudes agree well with the CRU (Harris et al. 2020) and the E-OBS (Cornes et al. 2018) precipitation observational data sets.

In the current study, we apply the regional climate coupled model (COSMO-CLM–NEMOMED12–TRIP) for the period 1951-2099 continuously using the

EC-EARTH GCM (Hazeleger et al. 2012) as driving data. The simulation used historical greenhouse gas emissions for the historical period 1951-2005 and RCP8.5 forcing scenario for the future period (2006-2099) at 0.11° (≈ 12 km) horizontal resolution. This simulation was a part of the coordinated activity by various institutions within the Med-CORDEX phase-II framework. The Med-CORDEX focuses on coordinated multi-model and multi-scenario studies covering the Mediterranean region with high resolution coupled regional climate models (Ruti et al. 2016).

We evaluate the EC-EARTH driven regional climate simulation with the ECMWF twentieth century reanalysis (ERA-20C) driven coupled regional climate simulation for Vb-cyclone frequency and their characteristics in the historical period before we proceed to investigate the future Vb-cyclone characteristics. Thereafter the NWMS state in terms of SST, evaporation and wind speed corresponding to the Vb-cyclone precipitation over the three catchments, the Danube, Elbe and Odra in the two simulations is analyzed. Finally to quantify the process chain linking the NWMS and the Vb-cyclone precipitation over the three catchments we use information theory methods similar to the studies by Pothapakula et al. (2019, 2020), Krug et al. (2021). These studies used information exchange to quantify the Indo-Pacific coupling, the interplay between the Indian Ocean dipole and El-Niño Southern Oscillation with the Indian Monsoon precipitation, and, the role of NWMS evaporation during the Vb-cyclone precipitation over Odra catchment. More details about these methods and the simulations are explained in the data and methodology section.

Specifically, in this study we ask the following questions:

1. Does the EC-EARTH driven coupled regional simulation produce Vb-cyclones comparable to the ERA-20C reanalysis driven coupled regional simulation in the historical period.

2. Does the state of the NWMS and the process chain differ between the two simulations in the historical period?

3. Do the characteristics of Vb-cyclones and their associated process chain linking the NWMS change in the future period?

This paper is organized as follows. Section 5.2 consists of data and methodology describing the climate models, the Vb-cyclone tracking, and a brief introduction to information theory methods. Thereafter, we present the results and discussion in Section 5.3 which includes validation of the Vb-cyclones in historical periods, future changes compared with historical period, results representing the state of NWMS and quantification of the process chain linking NWMS and Vb-cyclone precipitation in-terms of information exchange over the three catchments. Finally, some conclusions and outlook are given in Section 5.4.

5.2 Data and Methods

5.2.1 Regional Coupled Climate Model Setup

The dynamical downscaling was performed with a regional climate coupled atmosphere and ocean system consisting of atmospheric component COMSO-CLM 5-0-9 and the NEMOMED12 ocean component. The COSMO-CLM is a non-hydrostatic regional model designed for applications across various spatial and temporal scales. The governing equations were numerically solved by the Runga Kutta time-stepping scheme (Wicker & Skamarock 2002). It used the Arakawa-C grid in rotated geographical coordinates and follows terrain sigma vertical coordinates. The horizontal resolution of COSMO-CLM was about 0.11° and used 40 vertical layers representing about 22.7 km of the atmospheric column. The applied physical parameterizations included the Ritter & Geleyn (1992) radiative scheme, the Tiedtke convection scheme (Tiedtke 1989), and a four-category microphysics scheme (Doms & Baldauf 2011, Doms et al. 2011). The soil-vegetation-atmosphere-transfer sub-model TERRA provided the lower boundary conditions over land (Schrodin & Heise 2001, Schulz et al. 2016). The current simulation used the AeroCom Global AOD data (Kinne et al. 2006) to represent the aerosol properties. The initial and the lateral boundary files were taken from the EC-EARTH available through the SMHI Sweden (Hazeleger et al. 2012). The lateral boundary files were updated every 6 hours for the entire simulation period (1951-2099).

The NEMOMED12 is the ocean component of the regional climate coupled atmosphere and ocean system used in this study. The regional version of the NEMO-V3.6 was adapted over the Mediterranean region named NEMOMED12 in the current study. The domain of the NEMOMED12 covers the entire Mediterranean Sea at a horizontal resolution of ≈ 7.5 km along with a buffer zone nearby the Atlantic Ocean. A 3D relaxation of the temperature and salinity was performed in the buffer zone so as to realistically simulate the circulation from the Atlantic through the Gibraltar Strait and into the Mediterranean Sea (Sevault et al. 2009). The Black sea was parameterized such that the resultant net balance of the water budget is added into the Mediterranean Sea. The water budget was closed through the Total Runoff Integrating Pathways (TRIP) model (Oki & Sud 1998) which supplies freshwater influx at the Mediterranean river mouths. For more information on the NEMOMED12 readers are advised to refer Somot et al. (2008), Sevault et al. (2014). The coupling of the sub-components in the regional coupled system was done by the OASIS-MCT3 coupler (Craig et al. 2017). The coupling fields between the ocean and atmosphere are interpolated and exchanged every three hourly by the coupler.

In addition to the NEMOMED12, the NEMO configuration adapted over the North and Baltic marginal seas was coupled to the atmospheric component COSMO-CLM in a simulation driven by ECMWF twentieth century reanalysis, ERA-20C (Poli et al. 2016). The performance of ERA-20C downscaled simulation in realistically replicating the Vb-cyclone events and their associated precipitation was already reported and analyzed in the study by Krug et al. (2022). Hence in this study, we used the ERA-20C downscaled simulation as a reference for validating the downscaled EC-EARTH simulated Vb-cyclone events and their associated precipitation in the historical period. It is to be noted that though the EC-EARTH driven simulation was only coupled to NEMOMED12, both simulations used same coupling frequency and set-ups over the Mediterranean Sea region. From hereafter the downscaled simulation driven by ERA-20C is referred to as Goethe University Frankfurt (GUF) evaluation simulation, the downscaled EC-EARTH simulation in the historical period (1951-2005) as GUF historical simulation, and finally the future period (2045-2099) of downscaled EC-EARTH simulation as a GUF future simulation, or simply the evaluation, historical and future respectively. As the Vb-cyclone are rare events, we considered 55 years in the historical and future periods to account for sufficient Vb-cyclone cases.

5.2.2 *Vb-cyclone tracking*

For detecting and tracking Vb-cyclones in the two simulations a tracking algorithm developed by Wernli & Schwierz (2006) which was later modified by Sprenger et al. (2017) was used. The mean sea level pressure was used as input for the tracking algorithm (at a 6-hourly interval). Within the domain, 25° W–45° E and 25° N–75° N, closed isobars were tracked and the deepest pressure within the closed isobar was considered to be the cyclone center. Thereafter, the next following track point cyclone center was selected by a guess on the past displacement vector within a search of radius 1000 km. The tracking algorithm considers all the cyclones crossing the 47° N latitude and between the longitudes, 12° E and 22° E with a lifetime greater than 24 hours (Hofstätter et al. 2016, Wernli & Schwierz 2006). For more details regarding the tracking algorithm, the readers are directed to refer Krug et al. (2022).

5.2.3 *Vb-cyclones and North-Western Mediterranean Sea state*

After the Vb-cyclone tracking, their frequency of occurrence, track density, minimum central pressure, and precipitation was analyzed. Hence the historical period was taken from 1951-2005 and the future period from 2045-2099 in

this study. The number of Vb-cyclone events per year along with their linear trends and respective 95% confidence intervals was analyzed. The linear trend line is computed with the least square regression approximation through R software package (R Core Team 2013). The track density of the Vb-cyclones represent the probability with which Vb-cyclone centers cross a grid point in the given time period similar to the study by Messmer et al. (2015). The minimum central core pressures obtained from the simulations were plotted using box whiskers. The box whiskers ends at the quartiles, the horizontal line at the center represents the median, and the points/circles are the values more than 1.5 times the interquartile range from the end of the box.

For the Vb-cyclone precipitation analysis, three important catchments i.e., the Danube, Elbe, and Odra were selected. The daily anomalies of precipitation, SST, evaporation and wind speeds were calculated by subtracting the daily value from the daily climatological mean during the respective analysis period to remove the seasonal cycle and to account for possible systematic biases. There after, the spatial and temporal averages of the precipitation over the Danube, Elbe, and Odra were calculated for each Vb-cyclone life time as detected from the tracking algorithm. Similarly, the spatial and temporal averages of the SST, evaporation and wind speeds over the pre-defined uptake region over the NWMS were calculated. The spatially averaged precipitation sum anomaly accumulated during the Vb-cyclone life time over the respective catchments was further ranked according to the intensity indicating the rank of Vb-cyclone events. In other words, the precipitation anomaly during the entire Vb-cyclone life time were ranked and a Vb-cyclone event with highest positive precipitation anomaly was ranked as 1. In addition to the Vb-cyclone precipitation anomaly rankings, the absolute precipitation amounts were also showed.

Corresponding to the Vb-cyclone precipitation anomaly rankings we analyzed the state of the NWMS (7° E, 22° E, 35° N, 46° N) similar to the study by Krug et al. (2022). The corresponding spatially and temporally averaged SST, evaporation, and wind speeds anomalies over this pre-defined region were plotted corresponding to the Vb-cyclone precipitation anomaly rankings. We also showed the moving averages of SST, evaporation, and wind speed anomalies for 10 Vb-cyclones with the Local Polynomial Regression Fitting (LOESS) lines corresponding to the precipitation anomalies over the Danube, Elbe, and Odra catchments. We adapted the methodology as in Krug et al. (2022) for analyzing the precipitation rankings and corresponding NWMS state.

5.2.4 *Quantifying process chain between North-Western Mediterranean Sea and Vb-cyclone precipitation*

Methods from information theory were recently used in quantifying interactions among sub-systems, especially in climate system applications (Pothapakula et al. 2019, 2020, Krug et al. 2021, Ruddell et al. 2019). Transfer entropy (TE) was especially used in detecting and quantifying the direction of information exchange between two or more sub-systems. Unlike correlation, TE is an asymmetrical measure. Generally, the estimations from TE are free from any underlying assumptions of the probability distributions. However, the estimation of TE in non-parametric form is still a challenge and requires rigorous parameter tuning and testing (Kaiser & Schreiber 2002, Pothapakula et al. 2019).

A study by Pothapakula et al. (2019) tested various TE estimators on idealized and real climate test cases along with the sensitivity of these estimators on time series length. Their results showed that the TE-linear which assumes Gaussianity is robust in revealing the system dynamics. While the non-linear estimations like TE kraskov, kernel gave reliable results, their free-tuning parameters such as the number of nearest neighbors, kernel width were tested and tuned for reliable estimations.

In this study, we used the robust TE-linear estimation to quantify the process chain in terms of information exchange between the NWMS and the spatio–temporal averaged precipitation over three catchments during all the detected Vb-cyclone days. Here we bring to the attention of readers that unlike the precipitation rankings where the precipitation anomaly for the entire Vb-cyclone event life time was considered, the information exchange calculations used individual days of all the detected Vb-cyclone events. By considering all the Vb-cyclone days, we aim for robust estimation of information exchange with long time series as proposed in (Pothapakula et al. 2019). Furthermore, it is to be noted that the Granger causality is equivalent to the linear approximation of TE by a factor of 2 (Barnett et al. 2009). Krug et al. (2021) applied the same methodology to quantify the information exchange between the NWMS and the precipitation over Odra catchments during Vb-cyclones.

At the heart of the information theory lies the concept of Entropy (H). The Entropy quantifies the uncertainty of a random variable X (Shannon 1948) and is defined as,

$$H(X) = -\sum_{x} p(x) \log p(x), \tag{5.1}$$

where $p(x)$ represents the probability of a state of the random variable X. The summation goes through all the states of the random variable quantifying

the average uncertainty of X. The units of entropy are generally expressed in nats when natural logarithm is used, whereas in the units of bits if the logarithm to the base of 10 is used. In this study, all the results quantifying information exchange were expressed in the units of nats.

Mutual information (MI) is defined as the average uncertainty reduction in the random variable X provided by the knowledge of random variable Y or vice-versa.

$$MI_{XY} = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{5.2}$$

Where the $p(x,y)$ represents the joint probability of a state corresponding to the random variables X and Y. The MI is a symmetric quantity and thus can not detect the direction of information exchange.

The TE builds upon the MI measure and is defined as mutual information between the future target variable X and the whole past of the source Y^- conditioned on the whole past of the target variable X^-. The TE is an asymmetric measure giving directional information exchange.

$$TE_{Y->X} = MI(X; Y^-|X^-). \tag{5.3}$$

Due to computational complexity in the estimation of joint probability densities, the whole past of the source and target random variables are reduced as follows,

$$TE_{Y->X} = MI(X; Y_{t-\tau}|X_{t-\omega}), \tag{5.4}$$

where τ and ω represents the time lags of the history of source and target variables. The values of the τ and ω are generally chosen depending on the system dynamics. For more detailed review on TE and its estimation refer to Pothapakula et al. (2019).

In this study, we chose the target variable to be the spatial averaged daily precipitation anomaly over the respective catchments during a Vb-cyclone event and the source being the simultaneous state of SST, evaporation or wind speed anomalies over the NWMS. The value of τ was taken to be zero and ω as one consistent to the study of Krug et al. (2021). The TE measure in this study quantifies the reduction in uncertainty about the present state of precipitation in the respective catchment while knowing the state of NWMS (SST, evaporation or wind speed) during the same day given the knowledge of one day precipitation persistence in the catchment region. Significance tests with permuted surrogates were conducted for information exchange values (Lizier 2014, Pothapakula et al. 2019). Though the measure of TE is highly useful, its limitation interms of common drivers influencing the source and

target needs to carefully taken into account. For example, while investigating the information exchange between Vb-cyclone daily precipitation and NWMS daily evaporation anomaly, it is assumed that no common driver is influencing the source and target variables (Runge et al. 2019, Pothapakula et al. 2019).

5.3 Results and discussion

In this section, first, the Vb-cyclone characteristics interms of occurrence, precipitation, minimum sea level pressure are discussed in evaluation, historical and future simulations. Thereafter, we analyze the NWMS state and associated process chain interms of information exchange during these events. Additional analysis which includes the evaluation of the SST's obtained from the two simulations were performed and discussed in the Appendix B section.

5.3.1 Vb-cyclones in the historical and future periods

In this sub-section, we present and discuss the results obtained from Vb-cyclone tracking in various simulations. For the evaluation simulation, a total of 531 Vb-cyclone events were detected for the period 1951-2005 corresponding to 9.7 Vb-cyclone events per year (standard deviation is 2.1 events per year). In the historical simulation, a total of 557 Vb-cyclone events were detected for the period 1951-2005 corresponding to 10.1 Vb-cyclones per year (standard deviation is 1.6 events per year). The historical simulation slightly overestimated the number of Vb-cyclones by a statistically insignificant amount of 4.8% compared to the evaluation simulation. The blue line in Figure 5.1 represents the fitted linear regression line and the grey bands represent the 95% confidence interval bands for all the tracked Vb-cyclones each year in various simulations. Overall, the trends revealed to be statistically insignificant in evaluation, historical and future simulations. With respect to the seasonal differences, on average the historical simulation overestimated the Vb-cyclone occurrence by 49% per year in summer (significant at 95% confidence) while in winter it underestimated Vb-cyclones by 41% per year (significant at 95% confidence) compared to the evaluation simulation (see Figure B.2) in supplementary material). Furthermore, the SST's in the historical simulation during summer season are colder than in the evaluation simulation (see Appendix B) indicating that the Vb-cyclone occurrence's in the historical simulation during summer are determined by some large scale dynamics rather than by the thermodynamic instability created by the Mediterranean SSTs.

In total 567 Vb-cyclones were detected in the future simulation for the period 2045-2099 corresponding to 10.3 Vb-cyclones per year (standard devi-

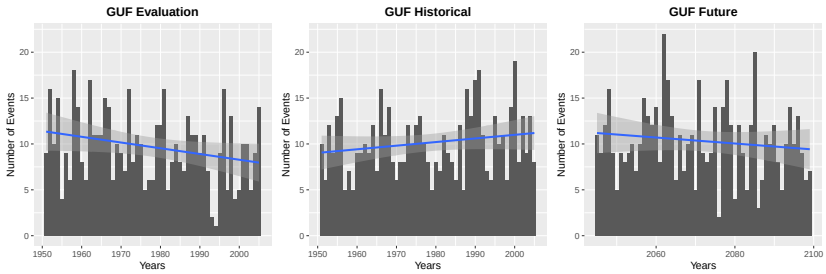
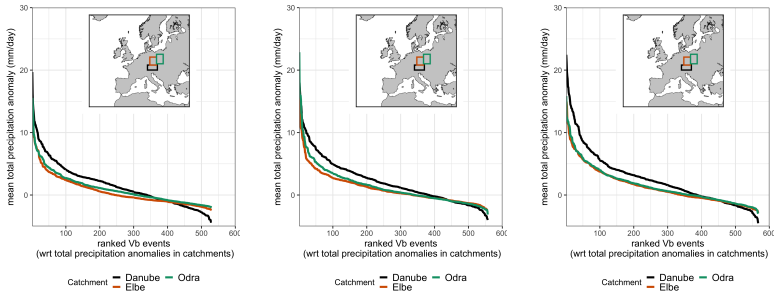


FIGURE 5.1. Time series of annual Vb-cyclone event number and their associated linear trends for the evaluation, historical, and future simulations. The shaded intervals correspond to the 95% confidence intervals for the Vb-event trend line.

ation is 2.3 events per year). This indicates an increase by 1.8% Vb-cyclones events per year in the future period compared to the historical period. Standard student's t-test analysis revealed that this percentage increase was insignificant. This result is consistent with the findings by Mittermeier et al. (2019) where an insignificant percentage increase of Vb-cyclone frequency in the far future was reported with 0.11° resolution stand-alone regional climate model. No significant changes in the Vb-cyclone seasonal frequency and trends were revealed in the future simulation compared to historical simulation (Figure B.2 in appendix). This result is contrary to the findings of Mittermeier et al. (2019) , where they reported significant changes in the Vb-cyclone occurrences in the future spring and summer seasons.

A good agreement in the Vb-cyclone track density was also detected between the historical and evaluation simulations (Figure B.3 in appendix). However, a minor underestimation of ≈ 1% of the Vb-cyclone centers over the eastern flanks of the Alps and a very slight overestimation over Italy was noted in the historical simulation. In the simulated future, the Vb-cyclones travelled further north-eastwards compared to the historical period (also reported in Messmer et al. (2020). The Vb-cyclones intensity in terms of minimum cyclone central pressure also revealed good agreement between the historical and evaluation simulation with respect to the median values (the center line of the box plots in Figure B.4 in appendix), however a discrepancy exists in the outliers, i.e., the extreme values, especially values with greater minimum cyclone central pressures (values > 1020 hPa). This difference might be attributed to the low number of Vb-cyclones detected in the winter season in historical simulation compared to the evaluation simulation (Figure B.2, winter season box plot). The future simulation indicated no significant changes in the Vb-cyclone minimum central pressures compared to historical simulation, especially with the median values in the box whisker plots. This is also true in all the seasons where no significant changes are observed between the median values of historical

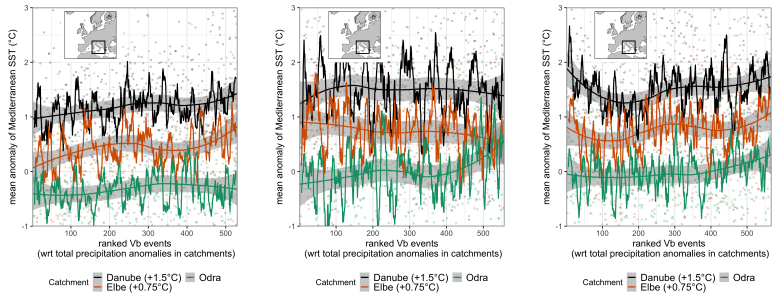


GUF Evaluation GUF Historical GUF Future

FIGURE 5.2. Ranked Vb-cyclone total precipitation anomalies in the Danube, Elbe, and Odra catchments obtained from various simulations.

and future simulations, however differences are noted in the outliers in all the seasons (Figure B.5 in appendix).

Figure 5.2 shows the Vb-cyclone precipitation anomalies ranked according to their magnitudes (lowest rank for maximum anomaly) in the Danube, Elbe, and Odra river catchments for various simulations. The rankings over the Elbe and Odra catchments in evaluation showed similar anomaly magnitudes due to their close spatial proximity. The Danube catchment showed higher precipitation anomalies in high and medium ranks ($\approx >$ 400 ranks) while a higher variability in lower ranks compared to Elbe and Odra in both historical and evaluation simulations. This behavior of Danube was attributed to the presence of complex orography and also to the typical Vb-cyclone pathways (Krug et al. 2022). The similarity between the Elbe and Odra catchments precipitation rankings, high precipitation anomalies in the Danube catchment is also seen from the absolute precipitation amounts (Figure B.6 in supplementary notes). Over all, there is a good agreement between the evaluation and historical simulated precipitation rankings over all the three catchments. However the precipitation magnitudes for a few high-ranked events ($\approx$ 1-20 ranks) were greater ($\approx$ 0.5, 0.45, 0.25 mm/day) in the Danube, Elbe, and Odra catchments respectively in evaluation simulation compared to the historical simulation (Figure B.7). The spatial precipitation patterns showed a zonal difference between the historical and evaluation simulation. The precipitation in the historical simulation was slightly greater in magnitude ($\approx$ 1mm/day) in the western part of the domain, while in the eastern domain, there was an under estimation ($\approx$ 1mm/day) (Figure B.8). In the future, all the three catchments show similar precipitation anomaly magnitudes as the historical simulation, but with very slight higher precipitation magnitudes for the top 10 high ranking Vb-cyclone events (Figure B.7). With Vb-cyclone events rankings greater than 10, no significant differences were found in the future precipitation anomalies



GUF Evaluation GUF Historical GUF Future

FIGURE 5.3. Sea surface temperature anomalies corresponding to the Vb-cyclone precipitation anomaly rankings in various simulations for Danube, Elbe and Odra catchments. The lines show the moving average and the LOESS regression. The data for Danube and Elbe catchments were shifted by constant values for improved representation.

and magnitudes over the three catchments. The spatial plots of the precipitation show that the future simulation on average shows an insignificant increase of total precipitation by about (≈ 1mm/day) for all the detected Vb-cyclones compared to the historical simulation (Figure B.8).

5.3.2 North-Western Mediterranean Sea state during the Vb-cyclones and associated process chains

This sub-section presents the state of the SST, evaporation, and wind speed anomalies over the NWMS and associated process chains interms of information exchange.

Figure 5.3 shows the spatially averaged SST anomalies of the NWMS (domain shown in black rectangle box) with respect to the Vb-cyclone precipitation anomaly rankings. In the evaluation simulation the high precipitation anomalies tend to be realized for low SST anomalies, especially for the Danube and Elbe catchments. This might be attributed to the strong upper sea mixing and evaporative cooling during the Vb-cyclone life time. These cooler anomalies were also partially replicated for the Danube and Odra basins in the historical simulation (≈ 1-100 ranks). In the future simulation the SST cooling was not noticed. Figure 5.4 presents the spatial distribution of the mean SST anomalies over the NWMS during all the Vb-cyclones. The evaluation simulation on average showed negative SST anomalies in the NWMS. This was expected as the Vb-cyclones usually originate from the NWMS. Though the cooling in the historical simulation was noticed with less magnitude in the north-western domain, no such cooling was seen over the Adriatic sea and Ionian region. This means that the SST's in these regions were not responsive in the historical

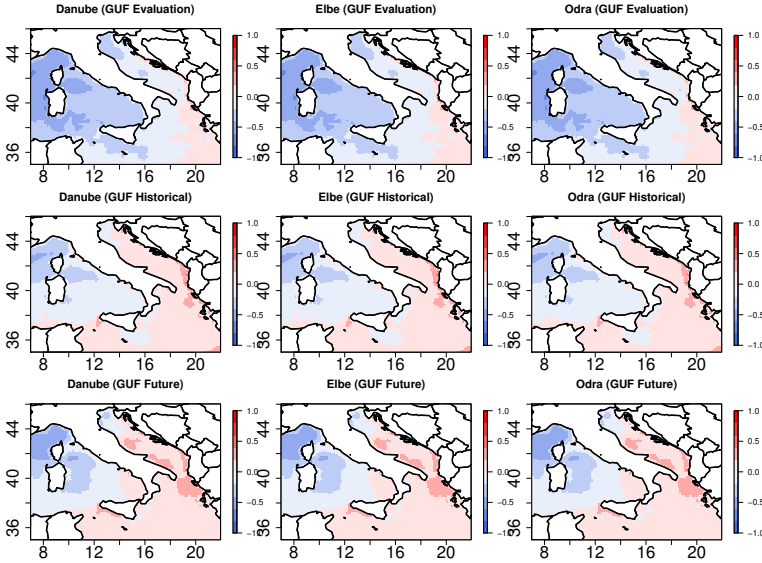


FIGURE 5.4. Mean sea surface temperatures anomalies (K) during all the Vb-cyclones in various simulations corresponding to precipitation over Danube, Elbe and Odra catchments.

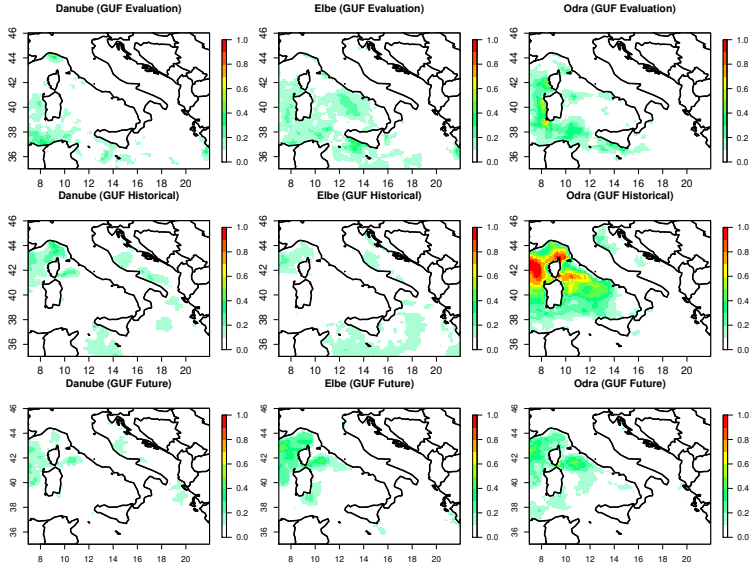
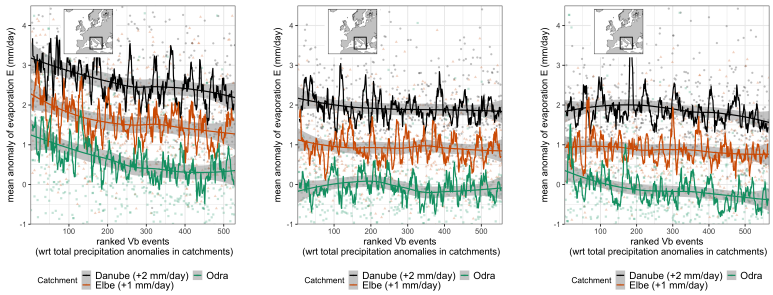


FIGURE 5.5. Information exchange ($\times 10^{-2}$ nats) between the SST's and the total precipitation anomalies over the Danube, Elbe and Odra catchments for various simulations. Only 95% significant range is plotted.



GUF Evaluation GUF Historical GUF Future

FIGURE 5.6. Evaporation anomalies corresponding to the Vb-cyclone precipitation anomaly ranking in various simulations for Danube, Odra and Elbe catchments. The lines show the moving average and the LOESS regression. The data for Danube and Elbe catchments were shifted by constant values for improved representation.

simulation. This behaviour was also seen in the mean bias plots, where the cooling of the SST's in the historical was underestimated compared to evaluation (Figure B.9 in appendix). The difference between the future and historical simulations on average showed no major differences in the magnitude of SST cooling during the tracked Vb-cyclones indicating no significant changes in the response of SST to the future Vb-cyclone events over the NWMS.

Thereafter, we investigated the information exchange between the SST's and the Vb-cyclone induced precipitation over the three catchments to diagnose the process chains linking the NWMS and Vb-cyclone precipitation (Figure 5.5). We noted significant differences in the information exchange spatial locations between the evaluation and historical simulations, especially in the Elbe and Odra catchments. In the historical simulation an underestimation of information exchange between the NWMS and the precipitation over the Elbe, and an overestimation in the information exchange over the Odra catchments was noted indicating that the process linking the SST's and the Vb-cyclone differ between the evaluation and historical simulation. The spatial locations of the information exchange in the future simulation remained the same as in historical simulation, but with minor changes in the magnitude of information exchange, especially over the Odra catchment, a significant reduction in the amount of information exchange is noted.

Figure 5.6 shows the evaporation anomalies over the NWMS with respect to the Vb-cyclone precipitation anomaly rankings. In the evaluation simulation the evaporation anomalies corresponded to the precipitation anomaly rankings in all the catchments indicating the dependence of Vb-cyclones on the NWMS moisture. High evaporation anomalies over the NWMS are linked with high ranked precipitation Vb-cyclone events overall the catchments, re-

iterating the importance of NWMS moisture feeding the Vb-cyclones. This result is consistent with the findings of Messmer et al. (2017) concluding that the moisture from the Mediterranean Sea leads to high atmospheric moisture availability causing an increased precipitation amount over central Europe. The historical simulation however showed no such correspondence between the NWMS evaporation anomalies and the Vb-cyclone precipitation except for only a minor increase in the anomalies of evaporation over the Danube and Elbe catchment (high ranks, $\approx >$ 100 ranks). The spatial plots in Figure 5.7 show that on average the magnitude of evaporation over the NWMS was greater in the evaluation simulation compared to the historical simulation during the detected Vb-cyclones. The historical simulation underestimated the evaporation from the NWMS, especially over the Adriatic and Ionian regions. This was further evident from the mean bias plots where a large negative bias is noted between the historical and evaluation simulation (Figure B.10 in supplement). The spatial patterns corresponding to the future simulations on average showed no significant changes in the magnitude of evaporation anomalies compared to the historical simulation indicating that the Mediterranean response with respect to the evaporation does not change in the future compared with the historical period.

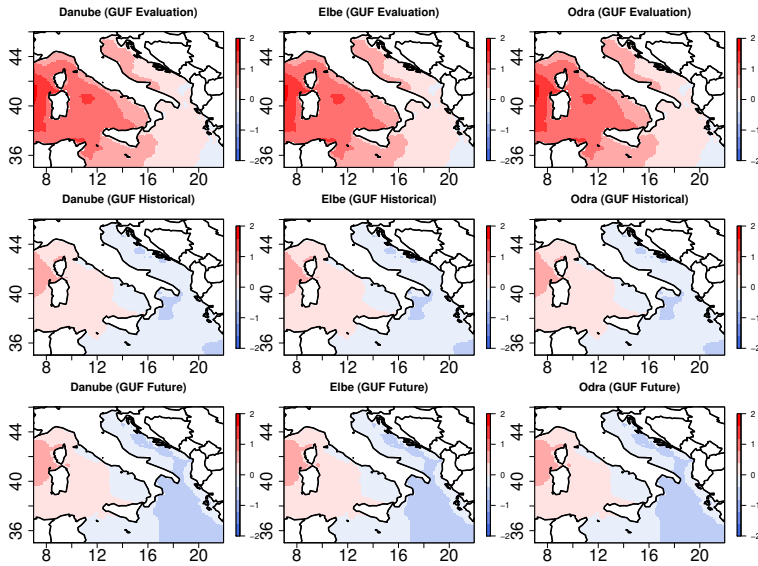


FIGURE 5.7. Mean anomaly patterns of evaporation (mm/day) over the Mediterranean Sea from various simulations for all Vb-cyclone events.

The information exchange between the evaporation over the NWMS and

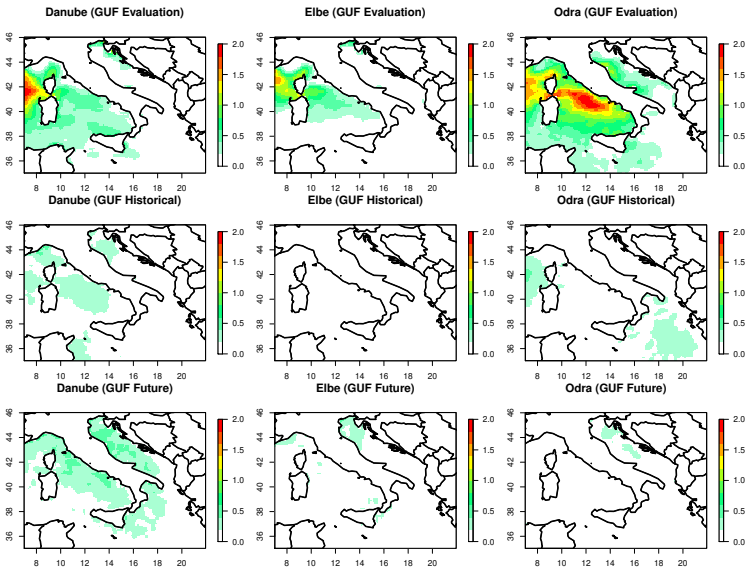
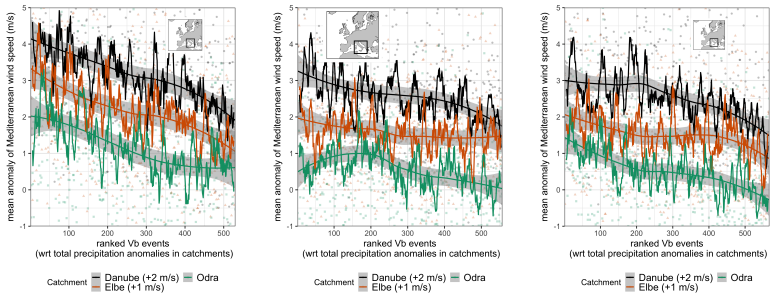


FIGURE 5.8. Information exchange ($\times 10^{-2}$ nats) between the evaporation over the Mediterranean Sea and the total precipitation anomalies over the Danube, Elbe and Odra catchments for various simulations. Only 95% significant range is plotted.

the Vb-cyclone precipitation over the three catchment's is presented in Figure 5.8. We noticed a significant amount of information exchange between the NWMS's evaporation (and also Adriatic Sea for Odra catchment) and the Vb-cyclone precipitation in the evaluation simulation for all the catchments. This result is in line with the findings of Krug et al. (2021) concluding significant information exchange between the NWMS and the Odra catchments for all the detected Vb-cyclones from 1901-2010. However, the historical simulation does not show significant information exchange linking the NWMS and Vb-cyclone precipitation in all the catchments. This indicates that some crucial physical processes linking the evaporation over the NWMS and Vb-cyclone precipitation were missing in the historical simulation. Nevertheless, the precipitation magnitudes over the three catchments are comparable between the evaluation and historical simulation implying that the underestimation of the Mediterranean moisture in the historical simulation is compensated by other moisture sources feeding the Vb-cyclone precipitation. Messmer et al. (2017) found a non-linear response of the Vb-cyclone precipitation with increment and decrement of moisture flux over the Mediterranean Sea. An increase in the Mediterranean moisture flux contributed to about 24% significant increase of Vb-cyclone precipitation, while a decrease in the moisture flux resulted in insignificant decrease in Vb-cyclone precipitation amount. This implies a non-

linear compensation effects of atmospheric moisture feeding the Vb-cyclone's, and perhaps this is also the reason for historical simulation compensation effects. However, this needs a detailed investigation. The information exchange spatial locations do vary between the historical and future simulations, but the difference in the magnitudes of information exchange is less compared to the differences between the historical and evaluation simulations. This indicates that the processes linking the NWMS evaporation and the Vb-cyclones are not likely to change in the future.



GUF Evaluation GUF Historical GUF Future

FIGURE 5.9. Wind speed anomalies corresponding to the precipitation anomaly rankings in various simulations for all Vb-cyclones. The lines show the moving average and the LOESS regression. The data for Danube and Elbe catchments were shifted by constant values for improved representation.

Figure 5.9 shows the wind-speed anomalies with respect to the Vb-cyclone precipitation anomaly rankings. The high wind speed anomalies tend to be realized for high precipitation rankings in evaluation simulation. Krug et al. (2022) showed that these strong winds result in the dynamic evaporative forcing over the NWMS fueling the Vb-cyclone precipitation especially during their initial phase. This phenomena also leads to the SST's cooling over the NWMS, which is evident from the results of our evaluation simulation. The increasing trends in wind-speed are also replicated in the historical and future simulations, however the wind speed anomaly magnitudes were slightly lower in historical simulations compared to the evaluation simulation. The future simulation also show a good correspondence in the wind speed anomalies, i.e, high ranked Vb-cyclone events linked with high wind speed anomalies.

The spatial plots of the mean daily wind speed anomalies for all the Vb-cyclones are shown in Figure 5.10. On average the wind speed anomaly magnitudes in the evaluation simulation over the NWMS were slightly higher in magnitude compared to the historical simulation (Figure B.11 in appendix). It is to be noted that the differences between the evaluation and historical simulation were greater in magnitudes for the SST anomalies (Figure 5.4)

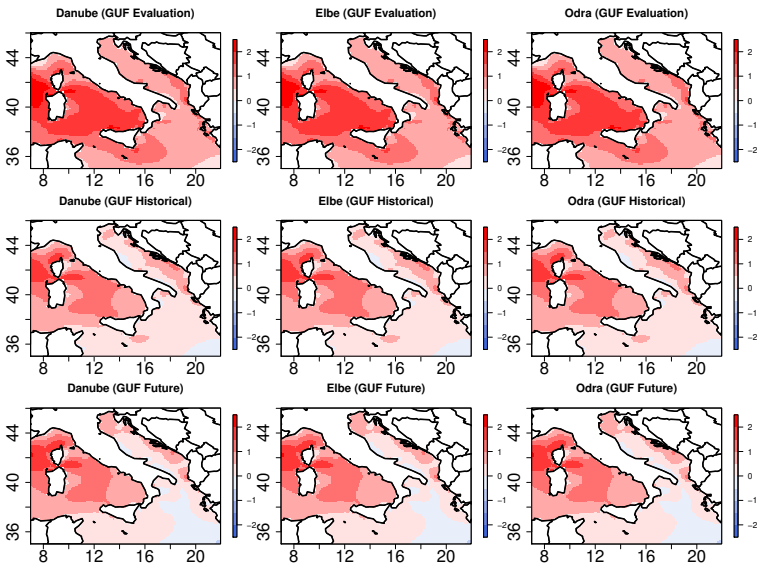


FIGURE 5.10. Mean anomalies of wind speed (m/s) over the Mediterranean Sea in the evaluation, historical, and future simulations over Danube, Elbe and Odra catchments for all Vb-cyclone events.

and the evaporation anomalies (Figure 5.7) over the NWMS compared to the wind speed anomalies differences. Similar to the higher significant SST and evaporation anomaly differences over the Ionian basin and the Adriatic Sea, the wind speed biases are also observed in the historical simulation over these regions compared to evaluation simulation. This implies that some of the crucial feed-backs interms of SST's, evaporation and, wind speed linking the Adriatic Sea and some parts of the Ionian basin are entirely missing in the historical simulation. The future simulation on average showed no significant differences from the historical simulation.

The information exchange spatial locations linking the NWMS wind speed anomalies and the Vb-cyclone precipitation over the three catchments is shown in Figure 5.11. Unlike the information exchange patterns linking the NWMS's SST and the evaporation anomalies to the Vb-cyclone precipitation where significant differences were found between the evaluation and historical simulations, the information exchange patterns linking the wind speed anomalies and Vb-cyclone precipitation showed very similar patterns and magnitudes. This indicates that though the historical simulation captured the process linking wind speed anomalies over the NWMS and Vb-cyclone precipitation, errors in the dynamical evaporative forcing still exist. We found a cold bias in SSTs simulated in the historical simulation which might be causing the unrealistic

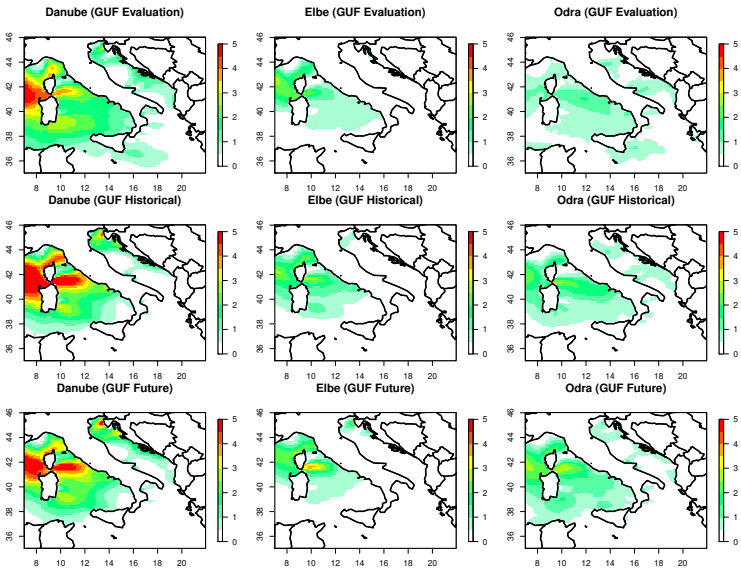


FIGURE 5.11. Information exchange ($\times 10^{-2}$ nats) between the wind speed and total precipitation anomalies for various simulations. Only 95% significant range is plotted.

dynamical evaporative forcing response over the NWMS. Though the same coupled RCM model setup was used for both the historical and evaluation simulation, we see differences in the SST magnitudes. We further investigated plausible reasons for such a behaviour of coupled RCM and a detailed explanation is provided in the Appendix A section. The investigation showed that the coupled RCMs closely follow the SSTs of their driving GCMs. The EC-EARTH GCM used in the historical simulation had a cold bias in the Mediterranean SST, which seemed to be further inherited by the coupled RCM. Furthermore, (Hazeleger et al. 2012) reported biases in surface parameters such as cold surface temperature and surface fluxes in the EC-EARTH GCM simulation. This indicates that the historical simulation must have inherited the simulation biases from its driving GCM, the EC-EARTH, resulting in cold bias of SST and surface air temperatures which influence the evaporation over the NWMS. A study by Pothapakula et al. (2020) reported that the biases from the GCM large scale signals were inherited by the downscaled COSMO-CLM simulations over the South Asia domain resulting in unrealistic process chain linking the Indian Summer Monsoon Rainfall, the Indian Ocean dipole and the El-Niño Southern Oscillation. However, study by Sørland et al. (2018) reported that the RCMs tend to improve the driving GCM biases over Europe, however the RCMs were reported to do a better job in correcting the GCMs when the

GCM has a warm bias, compared to a cold bias. Since the historical simulation forced by EC-EARTH is reported to have a cold bias on the surface temperature and sea surface temperatures, it is plausible that the cold biases were inherited by the coupled RCM historical simulation. The information exchange patterns and magnitudes corresponding to the future simulation in Figure 5.11 showed no significant differences compared to the historical simulation. This indicates that the future process linking the wind speeds and the Vb-cyclone induced precipitation over the three catchments do not significantly change in the future.

5.4 Conclusions

This work focused on the evaluation and projection of Vb-cyclones, the corresponding state and process chains linking the North-West Mediterranean Sea (NWMS) and Vb-cyclone related precipitation in two high resolution coupled regional climate model simulations. One regional simulation was driven by ERA-20C reanalysis (1951-2005) called evaluation, and the other simulation was driven by EC-EARTH for the period 1951-2099. The simulation for the period 1951-2005 was referred as historical, and from the period 2006-2099 as future simulation.

The results revealed a good agreement in the Vb-cyclone frequency between the evaluation (9.7 events per year) and historical simulations (10.1 events per year) but with significant seasonal differences. This discrepancy in the seasonal cycle of Vb-cyclone occurrences between the evaluation and historical simulation needs a further detailed investigation. The Vb-cyclone track density and intensity in terms of minimum cyclone central pressure showed good agreement between the evaluation and historical simulation. An insignificant increase by 1.8 % in the Vb-cyclone frequency by the end of 21^{st} century was revealed from future simulation. The Vb-cyclone precipitation anomaly magnitude rankings also showed good agreement between the evaluation and historical simulations. Changes in the future Vb-cyclone precipitation anomalies over the three catchments were insignificant.

In investigating the state of NWMS and the Vb-cyclone related precipitation in the Danube, Odra and Elbe catchments, the evaluation simulation showed a correspondence of NWMS SST, evaporation, and windspeed anomalies to the Vb-cyclone precipitation anomaly rankings. Such a correspondence was not detected in the EC-EARTH driven historical simulation. Despite good agreement in the Vb-cyclone frequency, intensity, and precipitation between the evaluation and historical simulation, the state and process chains differ. These differences might be attributed to the emergence of simulation biases

inherited from the driving EC-EARTH GCM. Furthermore, (Hazeleger et al. 2012) reported biases in surface parameters such as cold surface temperatures and fluxes, which might be inherited by the coupled RCM resulting in cold SSTs and unrealistic dynamical evaporative response in the historical simulation. Downscaling the EC-EARTH3 (latest version of EC-EARTH) which has smaller bias in the surface air temperatures, SST and surface fluxes (Döscher et al. 2021) might assist in further understanding the state and process chains linking the NWMS and the Vb-cyclone precipitation in historical and future periods.

5.5 Acknowledgements

The financial support of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in terms of the research group FOR 2416 Space-Time Dynamics of Extreme Floods (SPATE) is gratefully acknowledged. B.A and A.O would like to thank the European Union's Horizon 2020 research and innovation program, SOCLIMPACT (DownScaling CLImate imPACTs and decarbonisation pathways in EU islands and enhancing socioeconomic and non-market evaluation of Climate Change for Europe, for 2050 and beyond) for the support. The authors acknowledge the Deutsches Klimarechenzentrum (DKRZ) and the computing time granted by the NHR4CES Resource Allocation Board on the supercomputer Lichtenberg at TU Darmstadt as part of the NHR4CES infrastructure. The calculations for this research were conducted with computing resources under the project 967 at Lichtenberg supercomputer. The simulations were conducted as a part of Med-CORDEX initiative. The authors would also like to thank all the institutions which provided SST data through the Med-CORDEX website.

Part III

EPILOGUE

Chapter 6

Conclusion

This thesis investigated various methods derived from information theory and applied them to idealized dynamical systems, real-world climate applications including evaluation of regional climate model simulations. These methods were chosen to reveal and quantify the underlying system dynamics of complex systems such as climate. The first part of the thesis focused on methods such as mutual information, transfer entropy, information flow and their estimation techniques. Furthermore, the sensitivity of these estimators to free tuning parameters, available time-series length were explored on various idealized dynamical systems and few climate applications with known dynamics. Thereafter, these methods were applied to evaluate the dynamics underlying ENSO, IOD, and the Indian Summer Monsoon Rainfall in the observations, global climate models, and regional climate models. Finally, these methods were applied in evaluating the regional coupled climate modeling system and the interactions between the Mediterranean Sea and Vb-Cyclones. The following are the broad conclusions of the thesis,

6.1 Quantifying information exchange in idealized and climate system application

Applying various information theory methods and their estimators on two dimensional idealized test cases and real world climate applications and their sensitivity on the times series length and free tuning parameters, the following conclusions were drawn:

- The parametric estimators IF-linear and TE-linear detected and reliably quantified information exchange in the idealized linear systems however, they failed for nonlinear systems.
- Among the nonparametric estimators the TE-kernel and TE-kraskov passed the idealized tests, however, their implementations had to be tuned for consistent numerical results.
- For real world applications, these estimators revealed expected dynamics between IOD and ENSO interactions, while unrealistic bi-directional

information exchange between the NAO and European air temperatures indicating the influence of hidden drivers.

- The study proposed a composite use of the tested estimators with careful consideration of available time series length, parameter testing, limitations of the methods specifically involving high dimensional interactions and common drivers.

6.2 The synergistic impact of ENSO and IOD on the Indian Summer Monsoon Rainfall in observations and climate simulations - an information theory perspective

Exploring the two source information exchange dynamics on a single target in idealized experiments and the underlying dynamics of ENSO, IOD on the Indian Summer Monsoon Rainfall in observations, GCM's and RCM's revealed the following conclusions:

- Testing information exchange from two sources to a single target on idealized test cases revealed that both linear and the non-linear idealized systems can exhibit positive net synergy (i.e., the combined influence of two sources is greater than their individual contributions).
- Exploring the information exchange from two source variables (ENSO and IOD) to one target (ISMR) in the observational and reanalysis data revealed a synergistic role by the two sources in contributing towards the interannual variability over the Monsoon core region.
- Only one RCM simulation driven by a GCM with realistic ENSO, IOD moisture transport signals replicated the synergy between the two sources (ENSO and IOD) contributing to the interannual variability of Indian Monsoon Core rainfall.
- This study reiterates the importance of the driving GCM's large scale signals while dynamical downscaling with RCM's and the usefulness of IE methods in process based climate model evaluation.

6.3 Vb-cyclones and associated North-Western Mediterranean Sea state in regional coupled climate simulations: Evaluation and projection

Evaluating the coupled regional climate simulations in terms of the Vb-cyclones and the state of the North-West Mediterranean Sea in evaluation, historical

and projection simulations with information exchange methods revealed the following conclusions:

- There is a good agreement in the mean annual Vb-cyclone frequency, mean cyclone intensity, and precipitation rankings in the Danube, Elbe, and Odra catchments between the evaluation and the historical simulations.
- The projection simulation under the RCP-8.5 pathway revealed a basin-average SST warming of $\approx 2.5 - 3$ K, but insignificant changes in Vb-frequency, mean intensity, and precipitation in the catchments by the end of the 21^{st} century.
- Significant differences are found between the North-Western Mediterranean SST, evaporation, and wind speed anomalies corresponding to the precipitation rankings over the three catchments along with the process chains between the historical and evaluation simulation. No significant differences were found between the historical and future simulations.
- The discrepancies in the state of the North-Western Mediterranean Sea and the process chains in terms of information exchange might be attributed to the inheritance of the simulation biases from the driving GCM. This result reiterates the usefulness of IE methods in process based climate model evaluation.

6.4 Future Prospects

The present thesis tested and applied various methods derived from information theory to reveal the underlying system dynamics. After testing these methods and their estimation techniques on various idealized test cases, they were applied to real-world climate applications and regional climate model evaluation. After some findings and conclusions mentioned in the above sections, the following future directions in exploring these methods are recommended.

- While the estimation techniques such as TE-linear, IF-linear, TE-kernel, and TE-kraskov provided reliable results, they were still sensitive to the available time-series length and moreover, the non-linear estimators were sensitive to the free tuning parameters. Hence investigating the robustness of these estimators is still a challenge and needs further research. For example, the study by Zhu et al. (2015) proposed a new novel estimation technique extending the kraskov estimator for calculating TE.

- All the methods used in this thesis assume stationary time series in the real world climate applications and regional model evaluation. While most of the climate time series data is non-stationary, a method providing insights into the time-varying information exchange providing robust estimation for non-stationary time series needs further research. One such study in this direction was done by Tawia Hagan et al. (2019). Their study showed that the IF-linear can capture the time-varying causality structure within soil moisture–air temperature coupling.

- While this thesis investigated the synergy between the ENSO, IOD on the Indian Monsoon Summer Rainfall over the Monsoon core region, more recent studies focused on the influence of El Niño Modoki, a tripolar pattern of the SST and sea level pressure anomalies over the tropical Pacific on the Indian Summer Monsoon Rainfall (Feba et al. 2021). Hence future research can apply the information exchange methods to investigate the influence of El Niño Modoki and IOD on the Indian Summer Monsoon Rainfall.

- The information exchange between the Mediterranean Sea and the Vb-cyclone precipitation is investigated on two regional coupled climate model simulations. However, these methods could be applied to multi-model ensemble simulations for more robust estimations (Zittis et al. 2019).

- The methods in this thesis were applied between a single source and a target (eg., the interactions between the Mediterranean Sea and Vb-cyclone precipitation), dual-source and a target (eg., the ENSO, IOD influence on the Indian Summer Monsoon Rainfall). However, real-world climate applications involve high dimensional interactions involving multiple sources influencing a target. Revealing such multi-directional information exchange between the sub-components in the climate system requires robust measures by overcoming the curse of dimensionality. Further research is required in this direction (Runge et al. 2019).

==========

Appendix A

Supporting figures for chapter 4

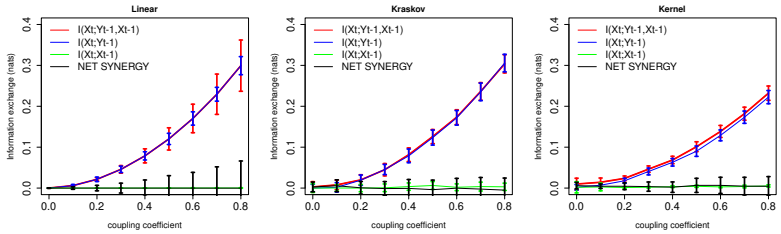


Figure A.1. Information exchange in nats from two-source (red line), single source (green and blue lines), and net synergy (black line) to target for Linear, Kraskov and Kernel estimators. The error bars represents two standard deviations of the 100 permuted samples.

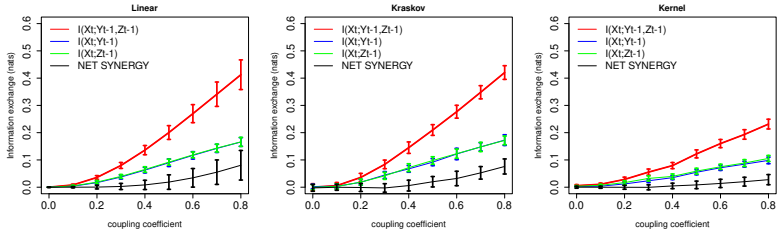


Figure A.2. Information exchange in nats from two-source (red line), single source (green and blue lines), net synergy (black line) to target for Linear, Kraskov and Kernel estimators. The error bars represents two standard deviations of the 100 permuted samples.

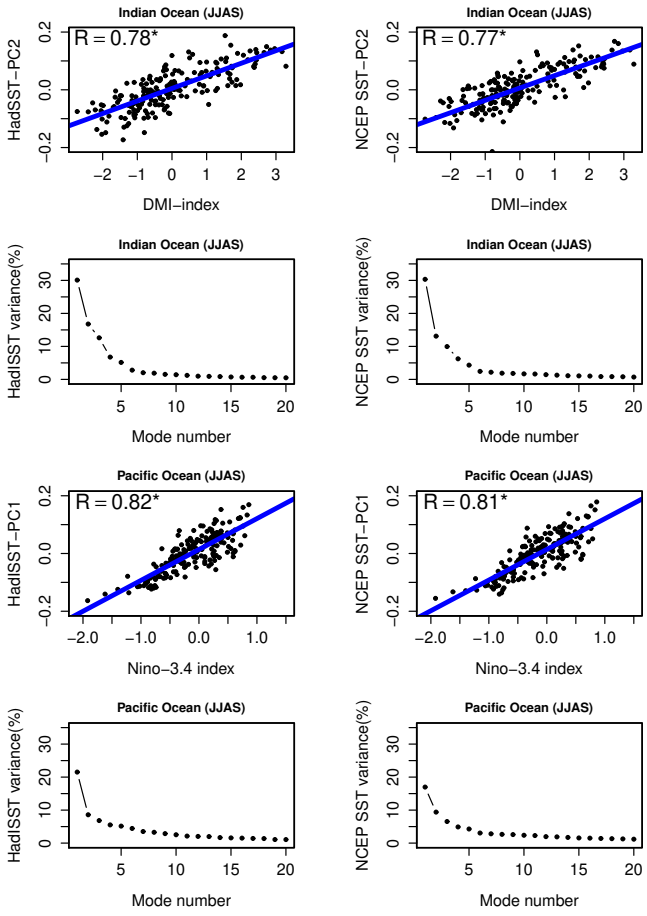


FIGURE A.3. Regressions of PCs obtained from their respective EOFs over the Indian and Pacific Oceans with the observed IOD and Niño 3.4 Index and their associated percentage contribution to the total variance for HadISST and NCEP reanalysis SST data sets for JJAS.

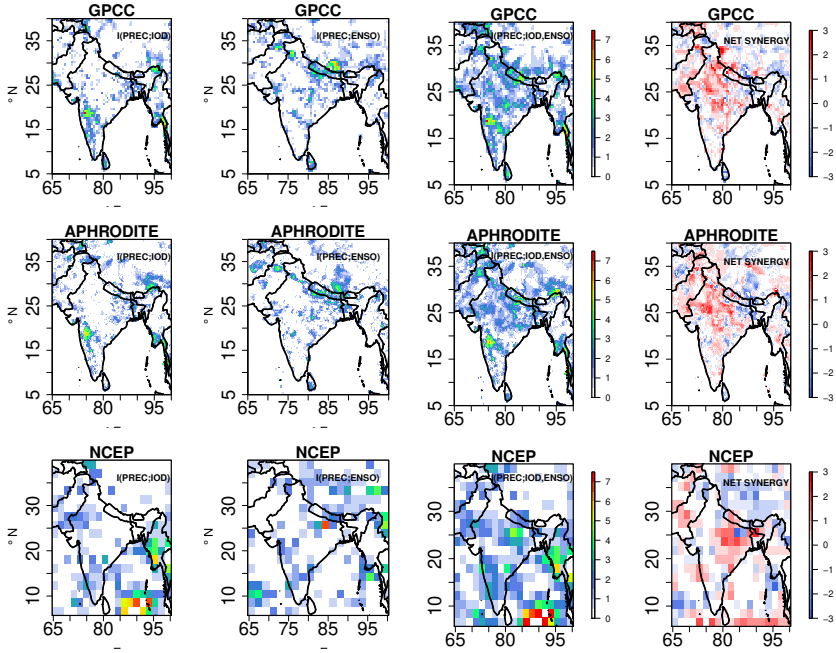


FIGURE A.4. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$, two-source information exchange $I(PREC;ENSO,IOD)$ and NET SYNERGY $\times 10^{-2}$ nats for observational data sets GPCC, APHRODITE and NCEP reanalysis with Kraskov estimator for JJAS. Only significant values at 95% confidence intervals are plotted.

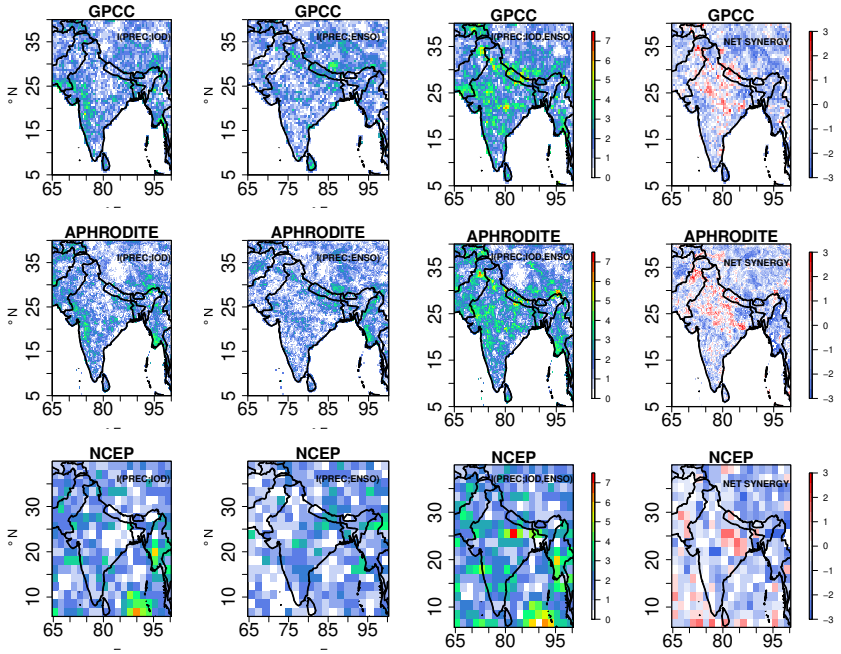


FIGURE A.5. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$, two-source information exchange $I(PREC;ENSO,IOD)$ and NET SYNERGY $\times 10^{-2}$ nats for observational data sets GPCC, APHRODITE and NCEP reanalysis with Kernel estimator for JJAS. Only significant values at 95% confidence intervals are plotted.

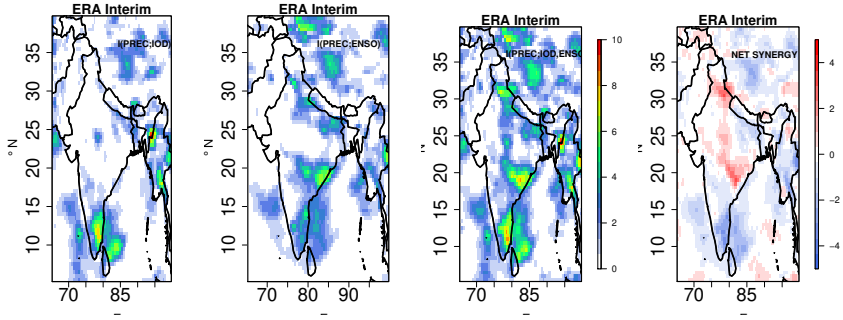


FIGURE A.6. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$, two-source information exchange $I(PREC;ENSO,IOD)$ and NET SYNERGY $\times 10^{-2}$ nats for observational data set ERA Interim reanalysis (1980-2005) for JJAS. Only significant values at 95% confidence intervals are plotted.

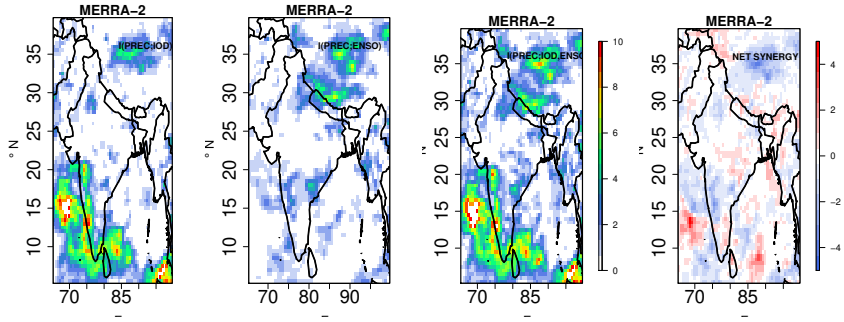


FIGURE A.7. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$, two-source information exchange $I(PREC;ENSO,IOD)$ and NET SYNERGY $\times 10^{-2}$ nats for observational data set MERRA-2 reanalysis (1980-2005) for JJAS. Only significant values at 95% confidence intervals are plotted.

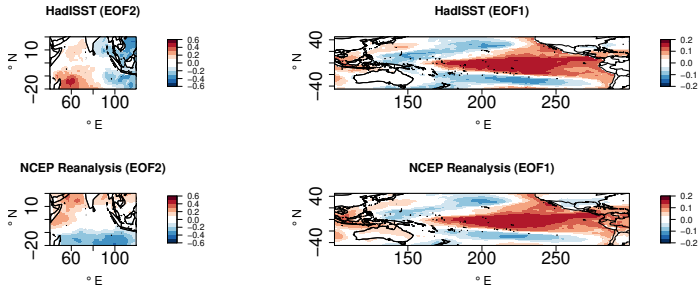


FIGURE A.8. EOF2 patterns of SST anomalies (DJFM) in the Indian ocean and EOF1 patterns in the Pacific ocean for observed HadISST and NCEP reanalysis.

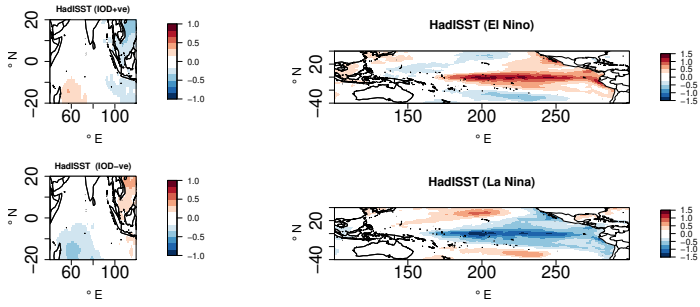


FIGURE A.9. SST composites (DJFM) in the Indian ocean and the Pacific ocean for observed HadISST.

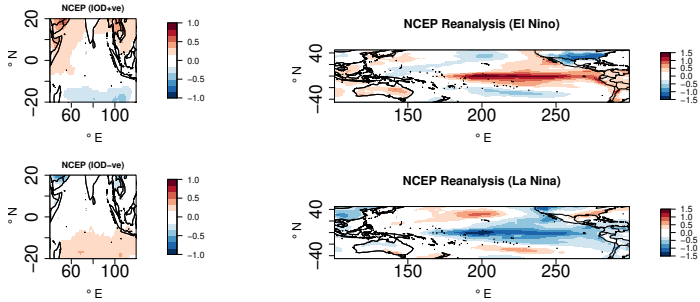


FIGURE A.10. SST composites (DJFM) in the Indian ocean and the Pacific ocean for observed NCEP reanalysis.

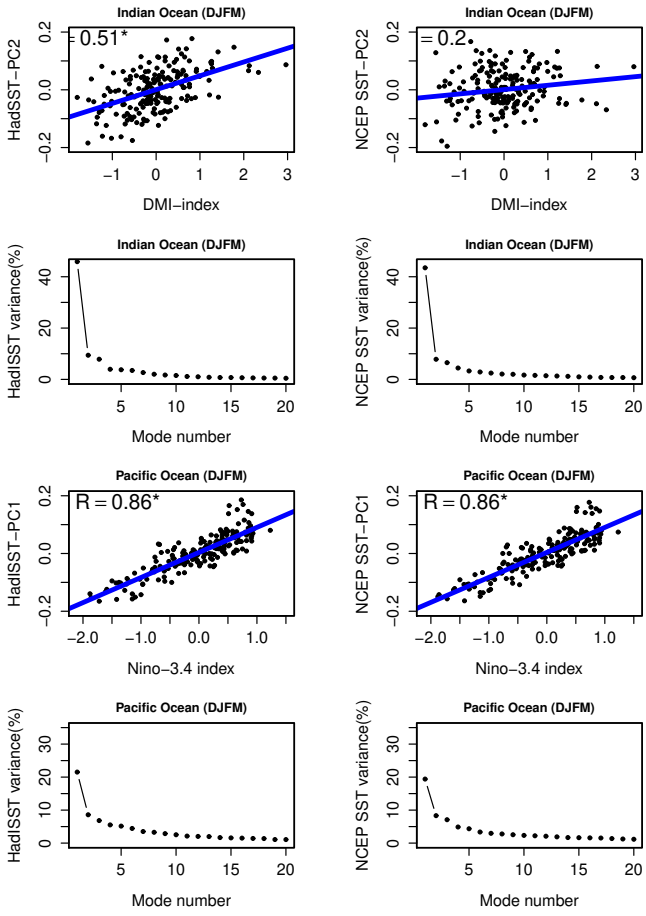


FIGURE A.11. Regressions of PCs obtained from their respective EOFs over the Indian and Pacific Oceans with the observed IOD and Niño 3.4 Index and their associated percentage contribution to the total variance for HadISST and NCEP reanalysis SST data sets for DJFM.

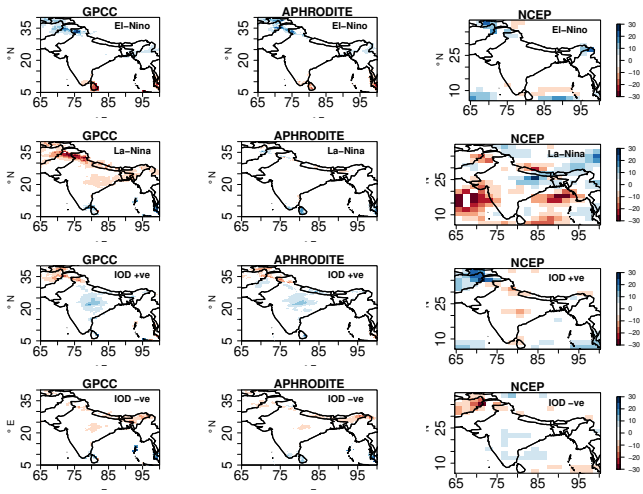


FIGURE A.12. Total precipitation anomaly (mm/month) composites (DJFM) over the Indian subcontinent for El-Niño, La-Niña, positive IOD and negative IOD events observed in GPCC, APHRODITE and NCEP reanalysis data sets for the period of 1951-2005

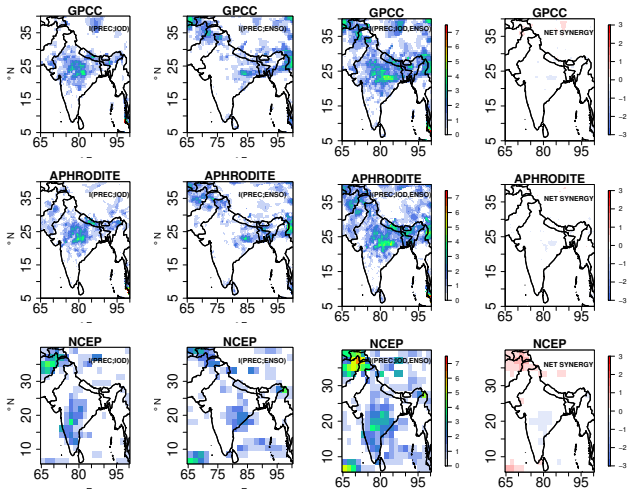


FIGURE A.13. Information exchange from $I(PREC;IOD)$, $I(PREC;ENSO)$, two-source information exchange $I(PREC;ENSO,IOD)$ and NET SYNERGY $\times 10^{-2}$ nats for observational data sets GPCC, APHRODITE and NCEP reanalysis for DJFM with Linear estimator. Only significant values at 95% confidence intervals are plotted.

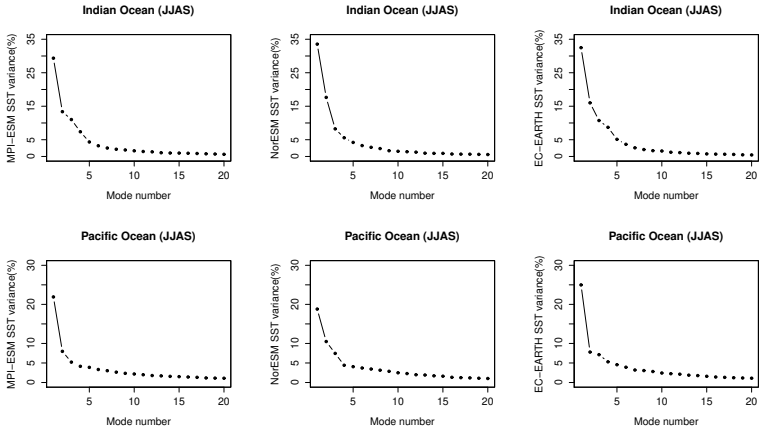


FIGURE A.14. Percentage of the total variance contributed by the first 20 EOFs to the total variability in Indian and Pacific Ocean SST for MPI-ESM-LR, Nor-ESM-M and EC-EARTH models for the month of JJAS (1951-2005)

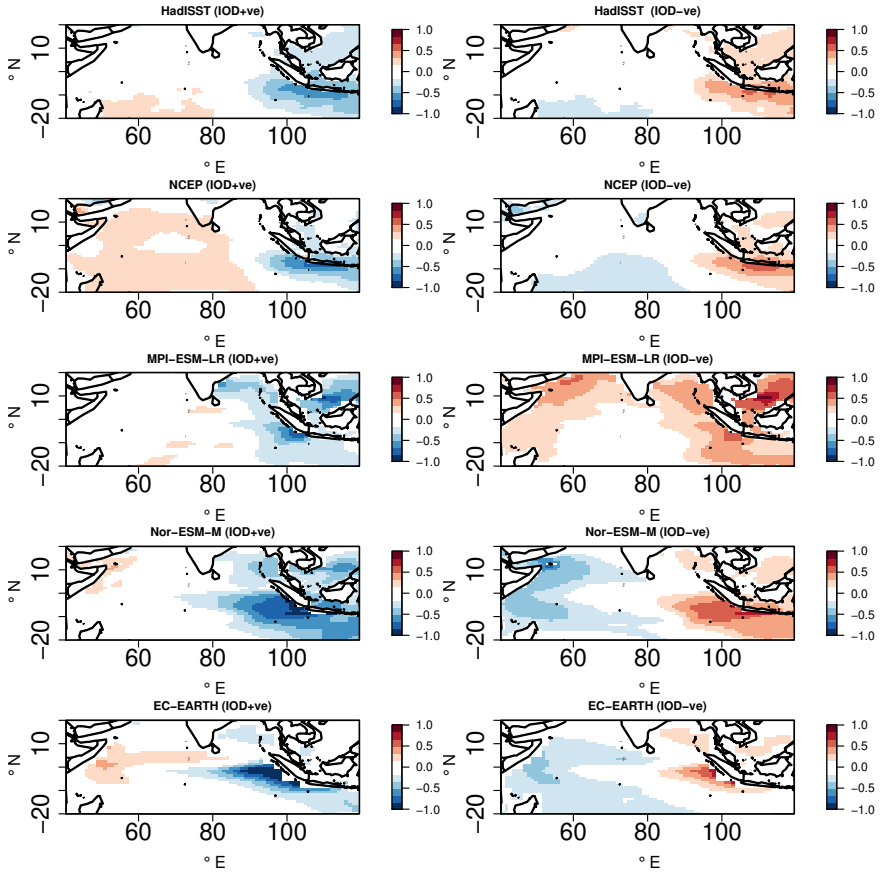


FIGURE A.15. SST composites for observations and GCMs for various phases of IOD events over the Indian ocean for JJAS.

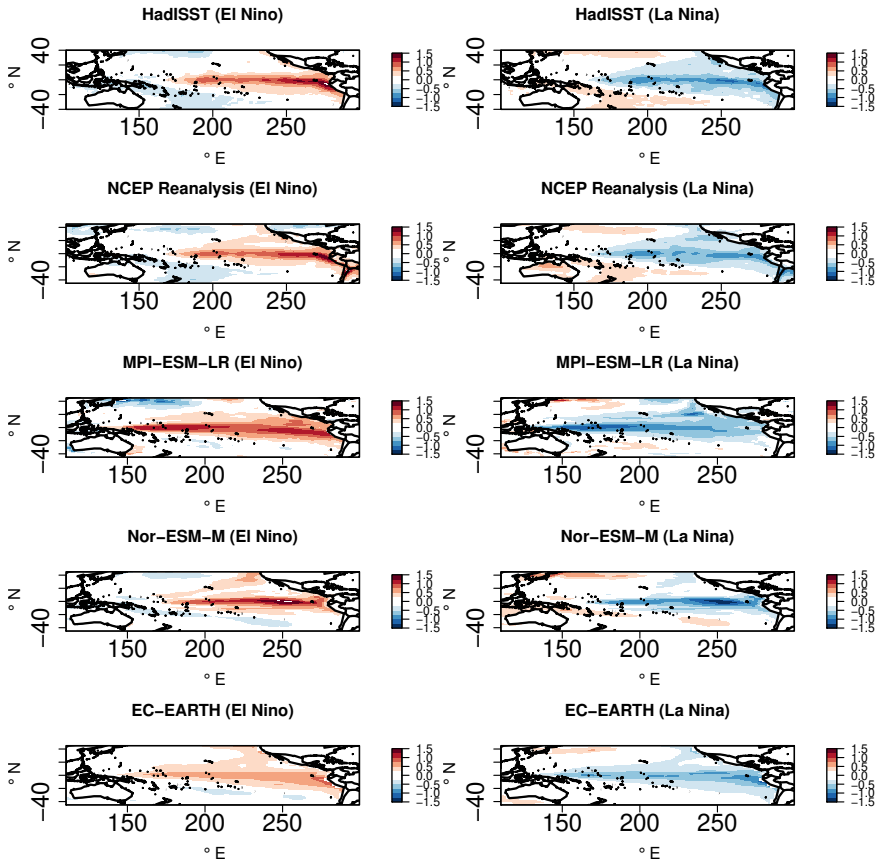


Figure A.16. SST composites for observations and GCMs for various phases of ENSO events over the Pacific ocean for JJAS.

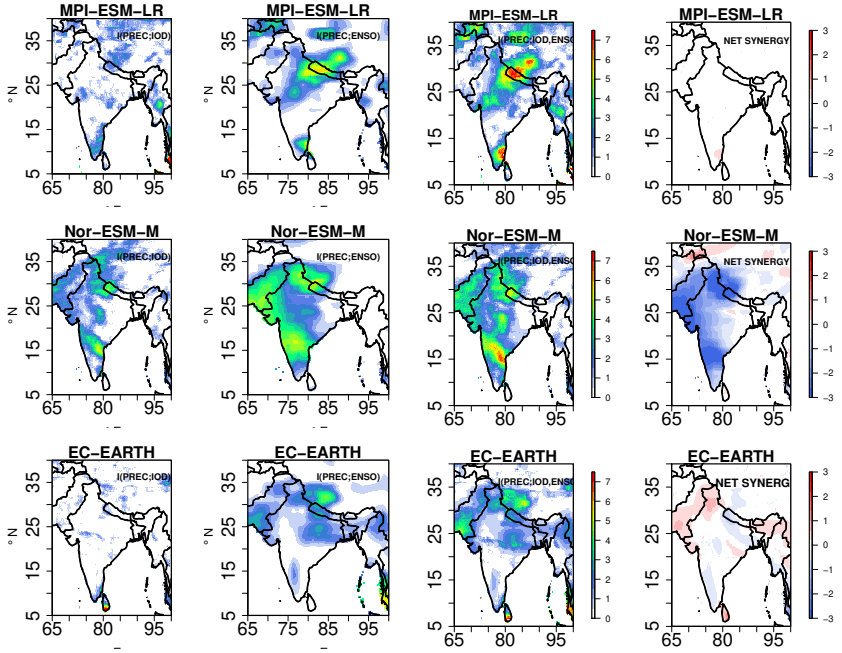


FIGURE A.17. Information exchange from I(PREC;IOD), I(PREC;ENSO), two-source information exchange I(PREC; ENSO,IOD) and NET SYNERGY $\times 10^{-2}$ nats for MPI-ESM-LR, Nor-ESM-M and EC-EARTH GCM models with Kraskov estimator for JJAS. Only significant values at 95% confidence intervals are plotted.

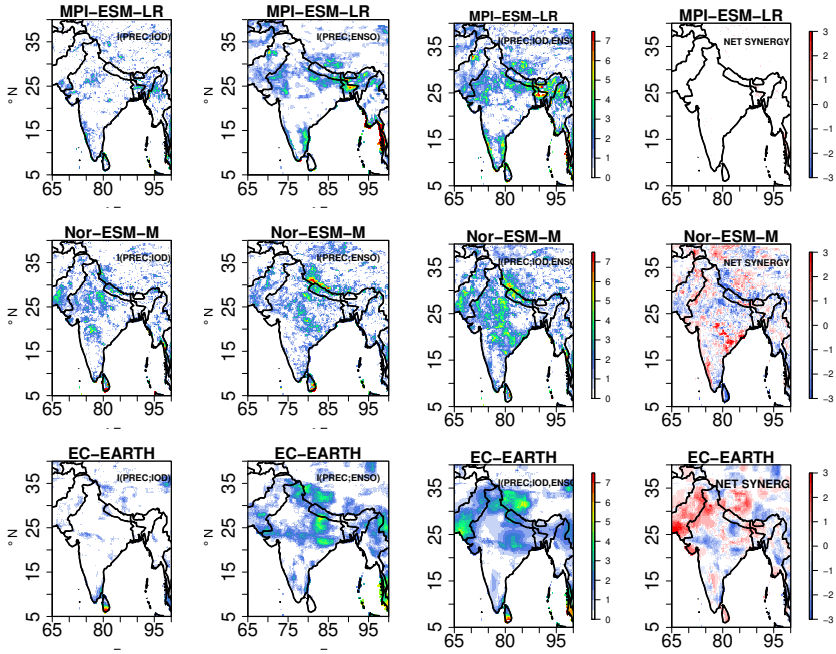


FIGURE A.18. Information exchange from I(PREC;IOD), I(PREC;ENSO), two-source information exchange I(PREC; ENSO,IOD) and NET SYNERGY $\times 10^{-2}$ nats for downscaled COSMO-crCLM simulations for JJAS (1951-2005) with Kraskov estimator. Only significant values at 95% confidence intervals are plotted.

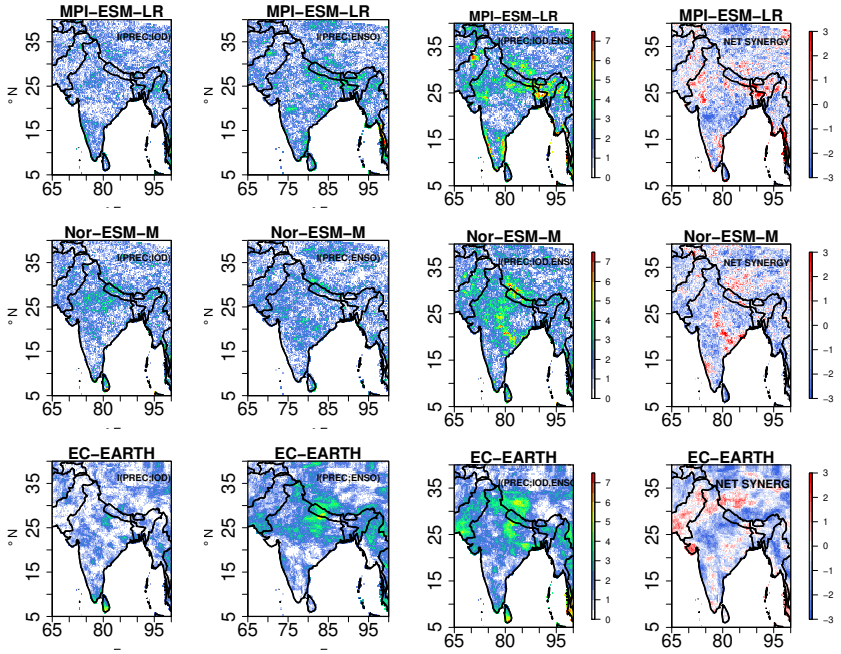


FIGURE A.19. Information exchange from I(PREC;IOD), I(PREC;ENSO), two-source information exchange I(PREC; ENSO,IOD) and NET SYNERGY $\times 10^{-2}$ nats for downscaled COSMO-crCLM simulations for JJAS (1951-2005) with Kernel estimator. Only significant values at 95% confidence intervals are plotted.

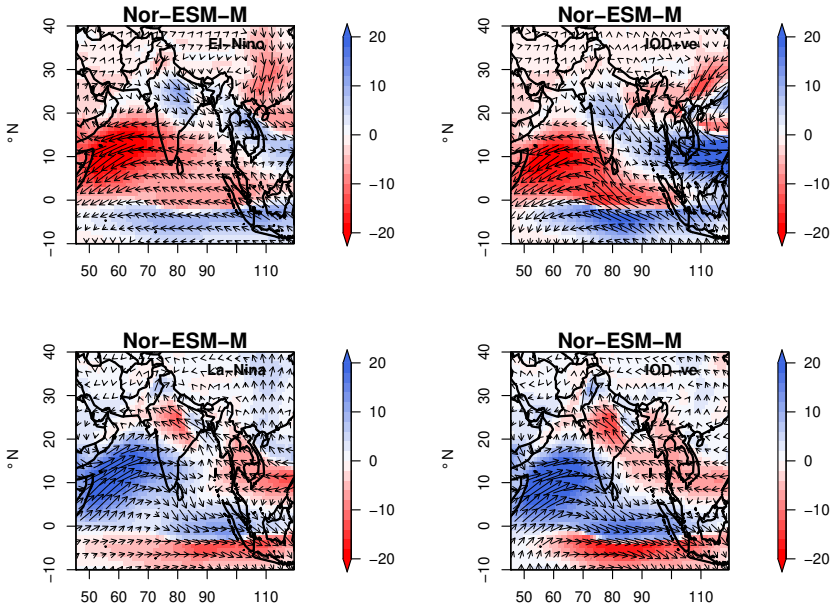


FIGURE A.20. Moisture flux anomalies (g/kg m/sec) over the Indian subcontinent (JJAS) for El-Niño, La-Niña, positive IOD and negative IOD events observed in Nor-ESM-M GCM for the period of 1951-2005

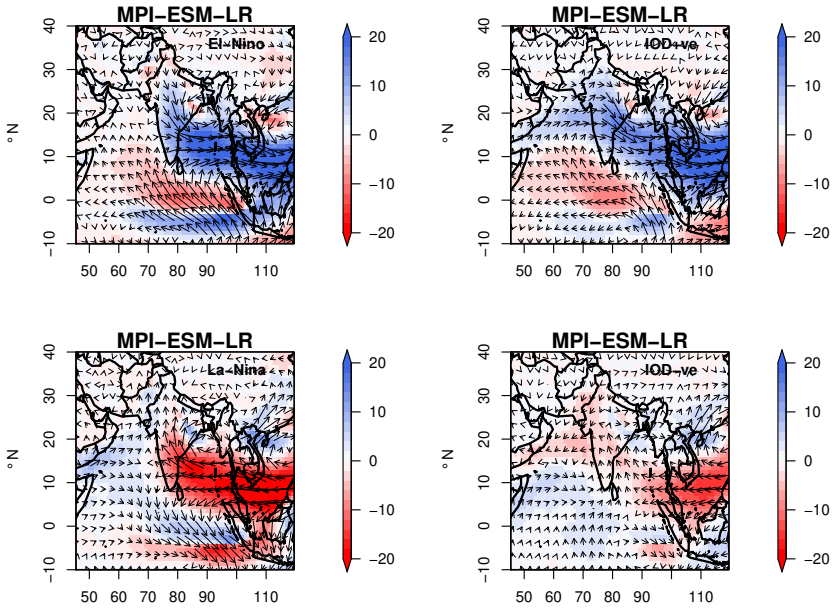


Figure A.21. Moisture flux anomalies (g/kg m/sec) over the Indian subcontinent (JJAS) for El-Niño, La-Niña, positive IOD and negative IOD events observed in MPI-ESM-LR GCM for the period of 1951-2005

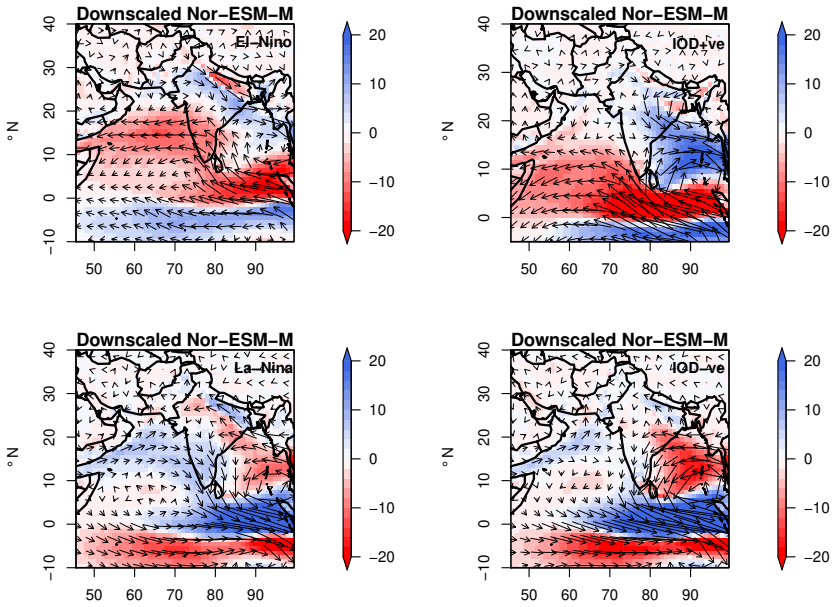


FIGURE A.22. Moisture flux anomalies (g/kg m/sec) over the Indian subcontinent (JJAS) for El-Niño, La-Niña, positive IOD and negative IOD events observed in downscaled Nor-ESM-M for the period of 1951-2005

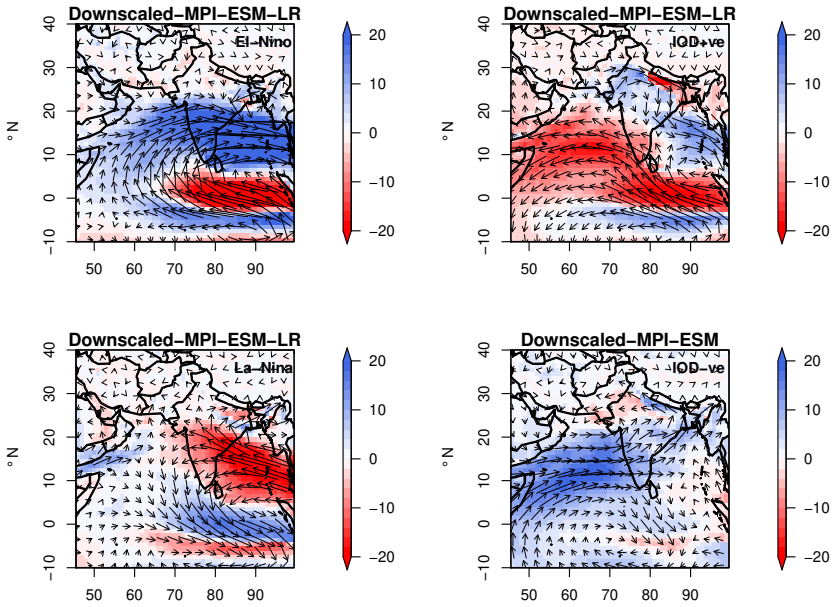


Figure A.23. Moisture flux anomalies (g/kg m/sec) over the Indian subcontinent (JJAS) for El-Niño, La-Niña, positive IOD and negative IOD events observed in downscaled MPI-ESM-LR for the period of 1960-1990

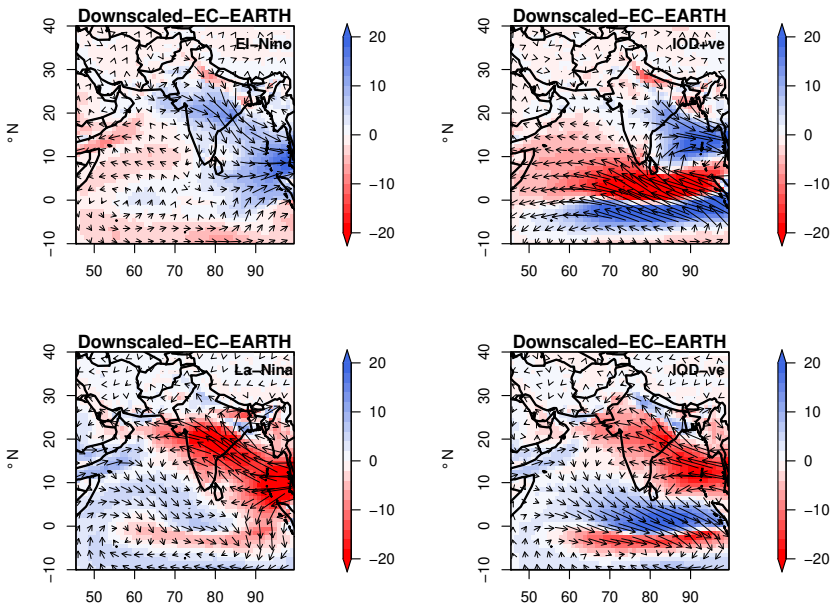


FIGURE A.24. Moisture flux anomalies (g/kg m/sec) over the Indian subcontinent (JJAS) for El-Niño, La-Niña, positive IOD and negative IOD events observed in downscaled EC-EARTH for the period of 1951-2005

Appendix B

Supplementary for chapter 5

B.1 Mediterranean Sea Surface Temperatures in the GCM-RCM chains

To understand the difference in the responses of the Mediterranean Sea in the historical and evaluation simulations, we compared the SST's obtained from the evaluation and historical regional coupled simulations with the UK Met Office's Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST 1.1) with a resolution about $1^\circ \times 1^\circ$ (Rayner et al. 2003) in the historical period. In addition, we also used the NOAA Optimum Interpolation (OI) SST V2 (Reynolds et al. 2002) with a horizontal resolution of about $0.25^\circ \times 0.25^\circ$ as another source of observational data set. Furthermore, we investigated the SST's as replicated in the GCM-RCM chain for various available coupled simulations in the Med-CORDEX phase II experiments as the simulation used in the current study were a part of Med-CORDEX phase II initiative. The data sets available in the Med-CORDEX data base used in this analysis is shown in Table B.1.

The SST data sets from the models in Table B.1 are available through the Med-CORDEX website (https://www.medcordex.eu). The University of Belgrade used the Princeton Ocean Model (POM) as the regional ocean component and the limited area model Eta/NCEP for the atmospheric component (Djurdjevic & Rajkovic 2008), the Centre National de Recherches Meteorologiques (CNRM), Meteo France used the NEMOMED8 as the ocean model and the ALADIN-Climate model as the atmospheric component (Sevault et al. 2014). The GERICS-AWI Helmholtz-Zentrum Hereon Geesthacht, Climate Service Center Germany used the MPIOM developed at the Max Planck Institute for Meteorology (Hamburg, Germany) as the ocean component and REMO as the atmospheric component. All the SST data sets were linearly interpolated onto a common grid prescribed by the Med-CORDEX community named OMED-11i which is approximately 12 km in resolution. Note that simulations used in this study are refered to as Goethe-University Frankfurt (GUF) simulations in this Appendix section.

B.1.1 Evaluation and future projections of Mediterranean SST

Figure B.1(a) shows the temporal evolution of the basin averaged annual Mediterranean SST for various simulations. Comparing the GUF historical and GUF

evaluation simulated SST's to the HadISST and OISST observational data sets, we noticed a cold bias (≈ 2 K). This cold bias was more pronounced in the GUF historical compared to the GUF evaluation simulation. This may be attributed to a more realistic atmospheric forcing by COSMO-CLM on the ocean model NEMOMED12 in the evaluation simulation compared to historical simulation. The Med-CORDEX phase-II ensemble also simulated a cold bias, but a smaller one compared with GUF simulations. A closer look into the seasonal cycle revealed that almost all the simulations had a cold bias in the spring and summer seasons (Figure B.12). It was interesting to note the close correspondence of the driving GCM's and the downscaled simulated SST time evolution in Figure B.1(a) indicating the inheritance of the GCM SST magnitudes by the respective RCMs. The global model, EC-EARTH's SST was colder than the other considered CMIP5 GCM SST's, hence, this explains the comparably larger cold bias of the GUF historical simulation which appears to be inherited by the coupled RCM. Furthermore, a narrow spread in the Med-CORDEX ensemble and CMIP5 GCM ensemble was identified. Selection of only two GCM's for downscaling, i.e., the MPI-ESM-LR and CNRM from the CMIP5 simulations so far might the reason for such a narrow spread.

A closer look into the spatial SST and the bias plots in historical period revealed that the cold bias was present throughout the Mediterranean Sea (Figure B.13 and Figure B.14). Especially the south-eastern warm pool was not very well captured by the GUF simulations and also by the Med-CORDEX ensemble members. However, overall important SST patterns (e.g., the warm eastern pool in the Levantine compared to the western cold pool over the North-Western Mediterranean) of the Mediterranean Sea were well captured by the GUF evaluation and GUF historical simulations.

The Mediterranean SST climate change signal is presented in Figure B.1(b). The SST anomaly was calculated with respect to the reference period 1951-2005. Almost all the simulations agreed very well that the basin averaged Mediterranean SST will warm ≈ 2.5 K – 3 K under the RCP8.5 scenario by the end of 21^{st} century. This warming of the Mediterranean Sea is consistent with the findings by Soto-Navarro et al. (2020). Spatial climate change SST patterns in GUF and the Med-CORDEX ensemble simulations reveal a homogeneous warming throughout the Mediterranean Sea (Figure B.15). This results indicate that the GUF simulation captured the future warming signal of the Mediterranean SST inline with the Med-CORDEX ensemble.

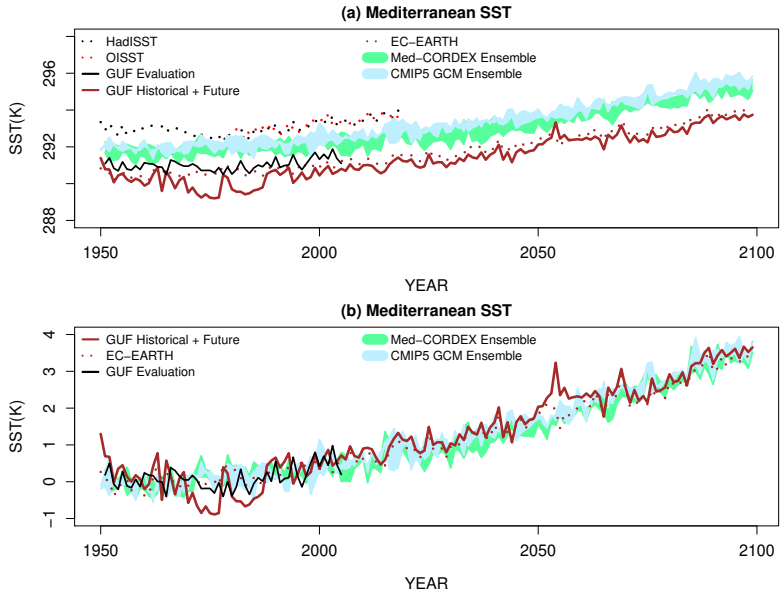


FIGURE B.1. (a) Mediterranean Sea basin averaged annual SST (K) evolution and (b) SST anomalies for the time period 1951-2099 (with reference to historical period 1951-2005) obtained from various simulations along with observational data sets, the HadISST and OISST.

Table B.1. RCMs/observations descriptions for SST evaluation over the Mediterranean Sea.

RCM Modeling Institution	**Acronym**	**driving model**
University of Belgrade	EBUPOM2c	MPI-ESM-LR
CNRM Meteo-France	CNRM-RCSM4	CNRM-CM5
Helmholtz-Zentrum Hereon	GERICS-AWI-ROM44	MPI-ESM-LR
Helmholtz-Zentrum Hereon	GERICS-AWI-ROM22	MPI-ESM-LR
Goethe University Frankfurt	CLMcom-GUF	EC-EARTH
Observations and Reanalysis data sets		
HadISST	–	–
OISST	–	–
Goethe-University Frankfurt (GUF)	CLMcom-GUF	ERA-20C

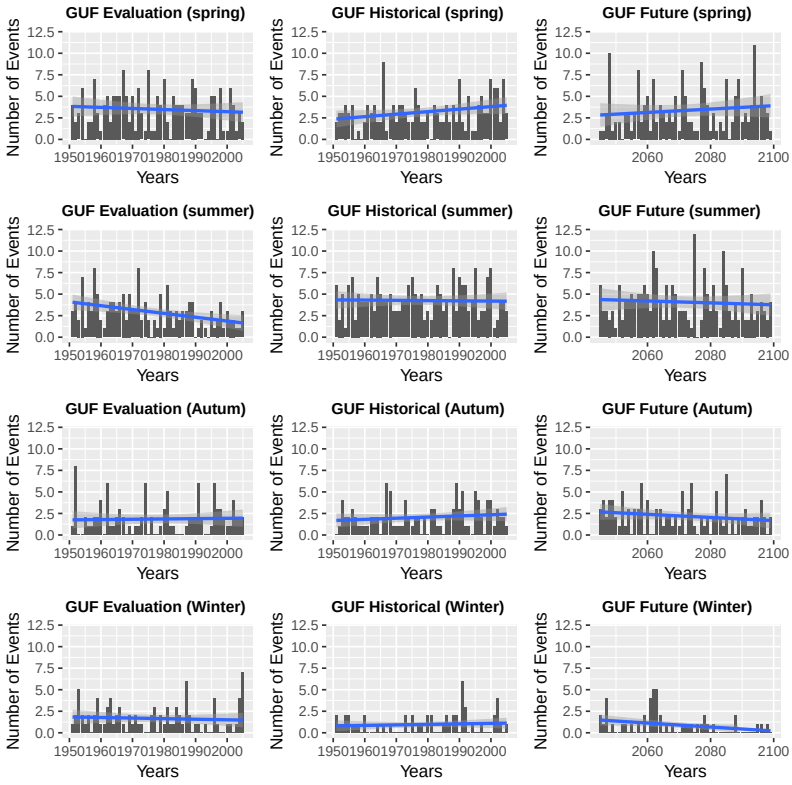


FIGURE B.2. Total number of Vb-cyclone events occured during Spring, Summer, Autumn and Winter and their associated trends in various GUF simulations

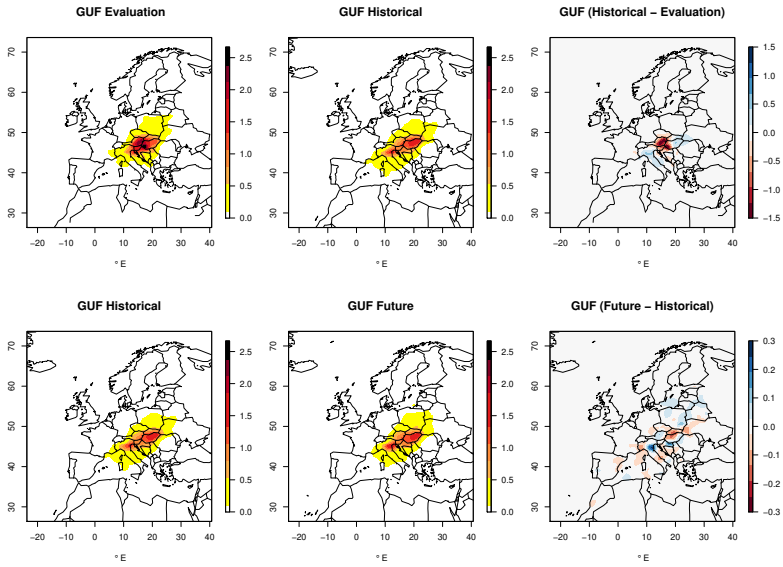


FIGURE B.3. Probability density field of all detected Vb-cyclone centres various GUF simulations and their respective differences.

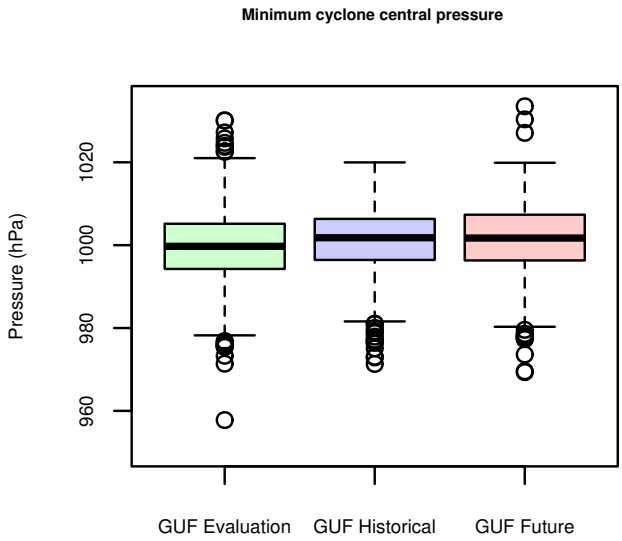


FIGURE B.4. Box plots representing minimum central core pressure (hPa) for all the Vb-cyclones in various GUF simulations.

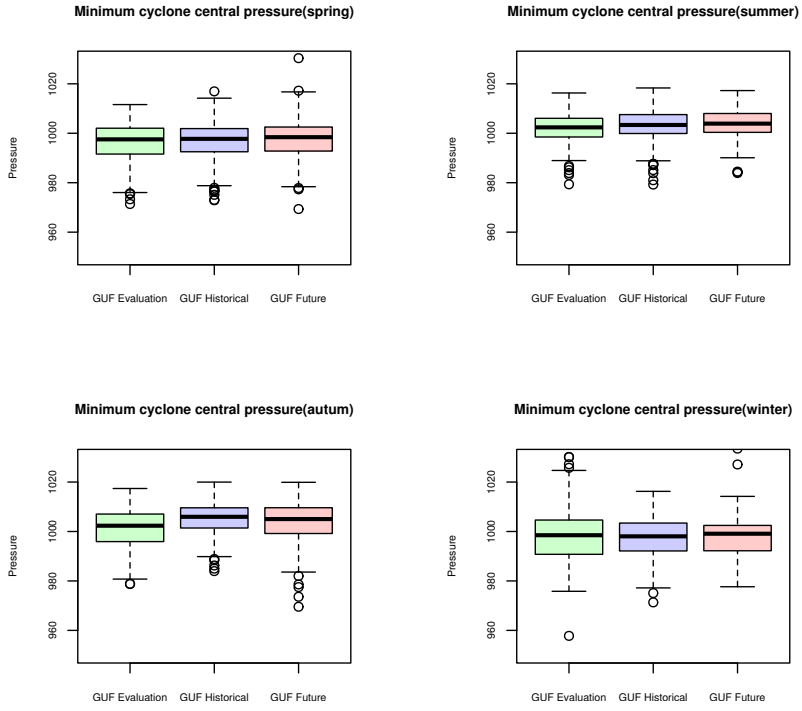


FIGURE B.5. Minimum central core pressure for all the Vb-cyclone detected tracks in Spring, Summer, Autumn and Winter seasons as simulated in various GUF simulations.

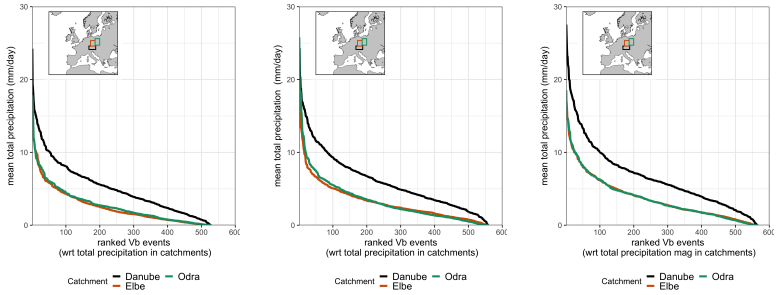
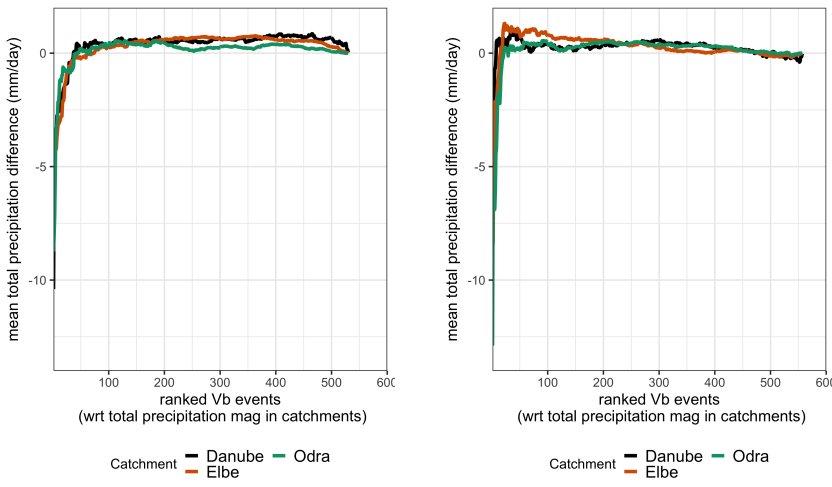


FIGURE B.6. Ranked total absolute Vb-precipitation amounts over the Danube, Elbe, and Odra catchments in (a) GUF evaluation (b) GUF historical (c) GUF future simulations.



GUF Historical - Evaluation GUF Future - Historical

FIGURE B.7. Difference between the total absolute precipitation amounts in the Danube, Elbe, and Odra catchment (a) GUF historical -GUF evaluation (b) GUF future - GUF historical

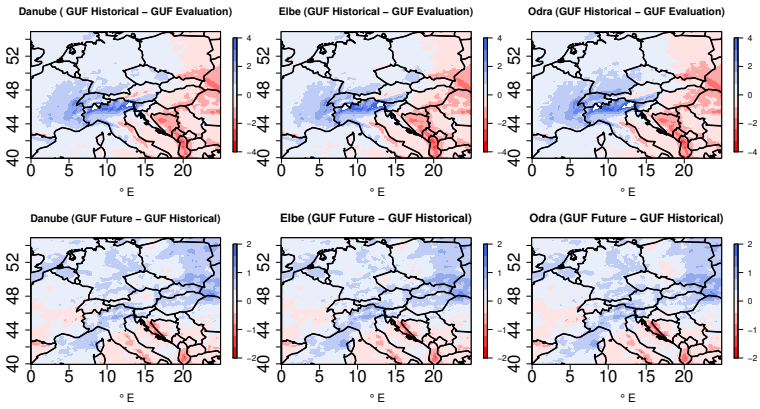


FIGURE B.8. Difference in the total absolute precipitation amounts (mm/day) between the GUF historical and evaluation (upper panel), GUF Future and historical (lower panel) for all the catchments

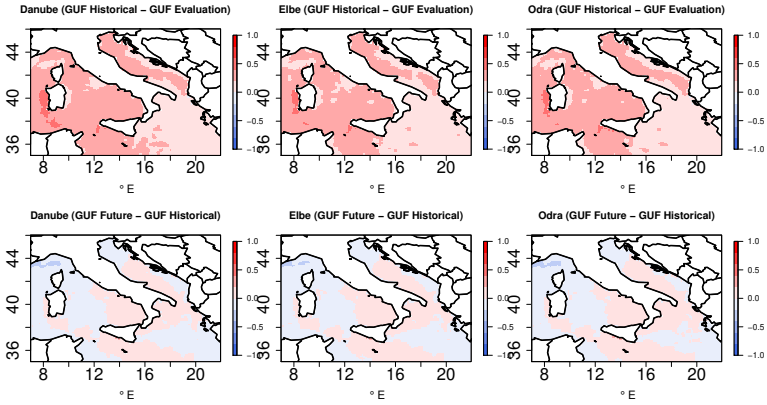


FIGURE B.9. Difference in the sea surface temperatures mean anomalies between GUF historical and evaluation (upper panel), GUF future and historical (lower panel) corresponding to Vb-precipitation precipitation over all the catchments.

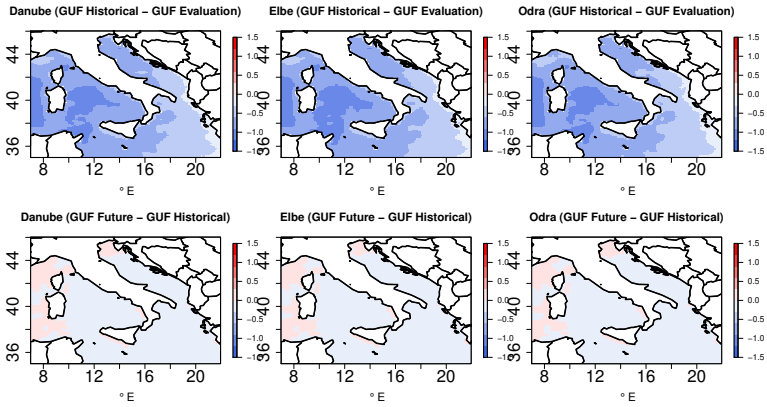


FIGURE B.10. Difference in the evaporation (mm/day) mean anomalies between GUF historical and evaluation (upper panel), GUF future and historical (lower panel) corresponding to Vb-precipitation over all the catchments.

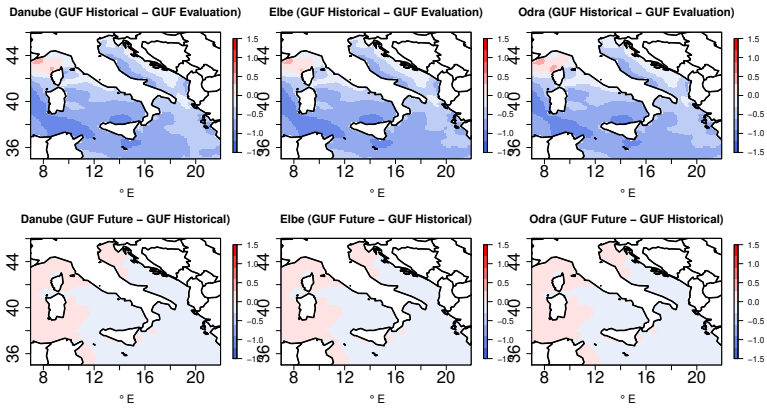


FIGURE B.11. Difference in the wind speed (m/s) mean anomalies between GUF historical and evaluation (upper panel), GUF future and historical (lower panel) corresponding to Vb-precipitation over all the catchments.

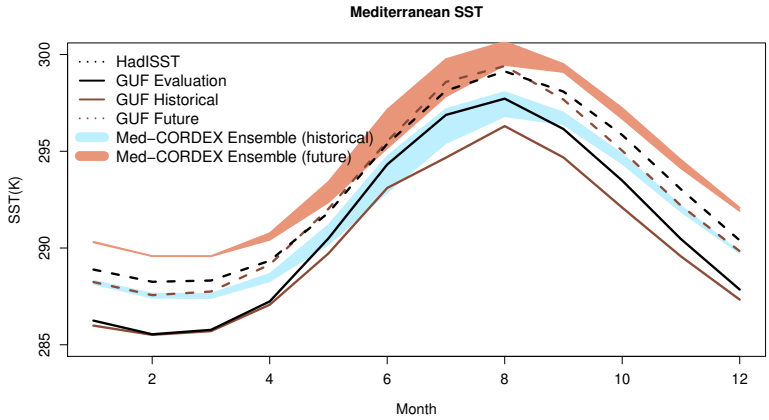


FIGURE B.12. Annual cycle of the averaged Mediterranean SST (K) as observed by HadISST observational data set and for various simulations.

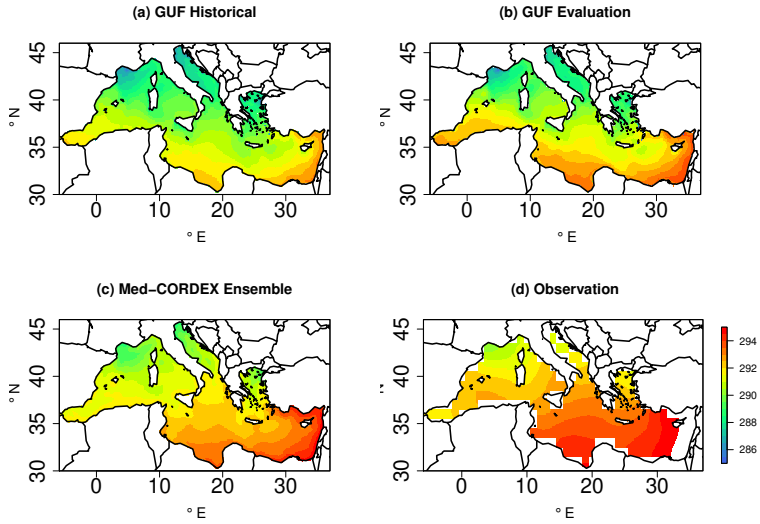


FIGURE B.13. The averaged Mediterranean sea surface temperatures (K) for the period 1951-2005 as simulated in (a) GUF historical (b) GUF evaluation (c) Med-CORDEX ensemble (d) Observations (HadISST).

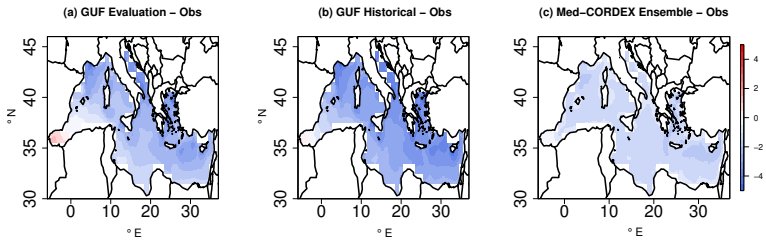


FIGURE B.14. Bias of the Mediterranean averaged sea surface temperatures (K) with respect to the observations (HadISST) for the period 1951-2005. (a) GUF evaluation - observation (b) GUF historical - observation (c) Med-CORDEX ensemble - observation.

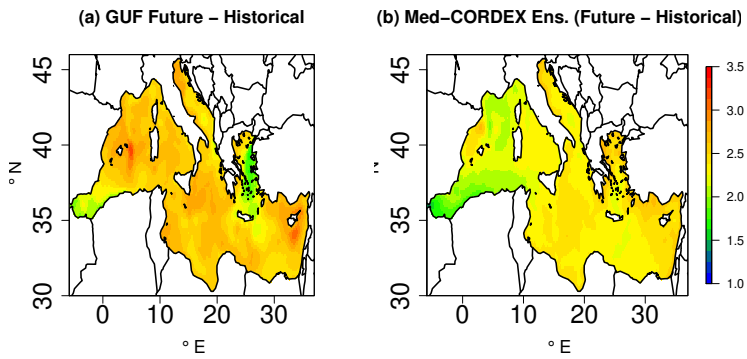


FIGURE B.15. The Mediterranean sea surface temperature change (K) in the future compared to the historical period for RCP 8.5 scenario in (a) GUF simulation (b) Med-CORDEX ensemble

Biblography

Bibliography

Ahmad, S. (2018), 'Information-theoretic model of self-organizing fullerenes and the emergence of c60', *Chemical Physics Letters* **713**, 52–57. Cited on page/s 62.

Ahrens, B. & Walser, A. (2008), 'Information-based skill scores for probabilistic forecasts', *Monthly Weather Review* **136**(1), 352–363. Cited on page/s 5, 26.

Akhtar, N., Brauch, J. & Ahrens, B. (2018), 'Climate modeling over the mediterranean sea: impact of resolution and ocean coupling', *Climate Dynamics* **51**(3), 933–948. Cited on page/s 11, 20, 88.

Akhtar, N., Krug, A., Brauch, J., Arsouze, T., Dieterich, C. & Ahrens, B. (2019), 'European marginal seas in a regional atmosphere–ocean coupled model and their impact on vb-cyclones and associated precipitation', *Climate Dynamics* **53**(9), 5967–5984. Cited on page/s 11, 88.

Annamalai, H., Xie, S., McCreary, J. & Murtugudde, R. (2005), 'Impact of indian ocean sea surface temperature on developing el niño', *Journal of Climate* **18**(2), 302–319. Cited on page/s 48, 49, 55.

Arakawa, A. & Lamb, V. R. (1977), 'Computational design of the basic dynamical processes of the ucla general circulation model', *General circulation models of the atmosphere* **17**(Supplement C), 173–265. Cited on page/s 18.

Asharaf, S. & Ahrens, B. (2015), 'Indian summer monsoon rainfall processes in climate change scenarios', *Journal of Climate* **28**(13), 5414–5429. Cited on page/s 61, 88.

Ashok, K., Guan, Z., Saji, N. & Yamagata, T. (2004), 'Individual and combined influences of enso and the indian ocean dipole on the indian summer monsoon', *Journal of Climate* **17**(16), 3141–3155. Cited on page/s 10.

Ashok, K., Guan, Z. & Yamagata, T. (2001), 'Impact of the indian ocean dipole on the relationship between the indian monsoon rainfall and enso', *Geophysical research letters* **28**(23), 4499–4502. Cited on page/s 48, 60.

Ashok, K., Guan, Z. & Yamagata, T. (2003), 'A look at the relationship between the enso and the indian ocean dipole', *Journal of the Meteorological Society of Japan. Ser. II* **81**(1), 41–56. Cited on page/s 48.

Bai, C., Zhang, R., Bao, S., San Liang, X. & Guo, W. (2018), 'Forecasting the tropical cyclone genesis over the northwest pacific through identifying the causal factors in cyclone–climate interactions', *Journal of Atmospheric and Oceanic Technology* **35**(2), 247–259. Cited on page/s 28.

Barnett, L., Barrett, A. B. & Seth, A. K. (2009), 'Granger causality and transfer entropy are equivalent for gaussian variables', *Physical review letters* **103**(23), 238701. Cited on page/s 93.

Barnston, A. G. & Livezey, R. E. (1987), 'Classification, seasonality and persistence of low-frequency atmospheric circulation patterns', *Monthly weather review* **115**(6), 1083–1126. Cited on page/s 52.

Barrett, A. B. (2015), 'Exploration of synergistic and redundant information sharing in static and dynamical gaussian systems', *Physical Review E* **91**(5), 052802. Cited on page/s 17, 63, 64, 66, 70, 75.

Behera, S. K. & Ratnam, J. (2018), 'Quasi-asymmetric response of the indian summer monsoon rainfall to opposite phases of the iod', *Scientific reports* **8**(1), 1–8. Cited on page/s 60, 73, 75.

Behera, S. K. & Yamagata, T. (2003), 'Influence of the indian ocean dipole on the southern oscillation', *Journal of the Meteorological Society of Japan. Ser. II* **81**(1), 169–177. Cited on page/s 49, 55.

Beniston, M., Stephenson, D. B., Christensen, O. B., Ferro, C. A., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylhä, K., Koffi, B. et al. (2007), 'Future extreme events in european climate: an exploration of regional climate model projections', *Climatic change* **81**(1), 71–95. Cited on page/s 86.

Bennett, A., Nijssen, B., Ou, G., Clark, M. & Nearing, G. (2019), 'Quantifying process connectivity with transfer entropy in hydrologic models', *Water Resources Research* **55**(6), 4613–4629. Cited on page/s 26, 62.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C. et al. (2013), 'The norwegian earth system model, noresm1-m–part 1: description and basic evaluation of the physical climate', *Geoscientific Model Development* **6**(3), 687–720. Cited on page/s xxix, 68.

Bertschinger, N., Rauh, J., Olbrich, E., Jost, J. & Ay, N. (2014), 'Quantifying unique information', *Entropy* **16**(4), 2161–2183. Cited on page/s 63.

Beuvier, J., Lebeaupin Brossier, C., Béranger, K., Arsouze, T., Bourdallé-Badie, R., Deltel, C., Drillet, Y., Drobinski, P., Lyard, F., Ferry, N. et al. (2012), 'Med12,

oceanic component for the modeling of the regional mediterranean earth system', *Mercator Ocean Quarterly Newsletter* **46**, 60–66. Cited on page/s 19.

Bhaskar, A., Ramesh, D. S., Vichare, G., Koganti, T. & Gurubaran, S. (2017), 'Quantitative assessment of drivers of recent global temperature variability: an information theoretic approach', *Climate Dynamics* **49**(11), 3877–3886. Cited on page/s 7, 26, 27, 62.

Bhaskaran, B., Ramachandran, A., Jones, R. & Moufouma-Okia, W. (2012), 'Regional climate model applications on sub-regional scales over the indian monsoon region: The role of domain size on downscaling uncertainty', *Journal of Geophysical Research: Atmospheres* **117**(D10). Cited on page/s 61.

Blöschl, G., Nester, T., Komma, J., Parajka, J. & Perdigão, R. A. (2013), 'The june 2013 flood in the upper danube basin, and comparisons with the 2002, 1954 and 1899 floods', *Hydrology and Earth System Sciences* **17**(12), 5197–5212. Cited on page/s 86.

Boccaletti, S., Kurths, J., Osipov, G., Valladares, D. & Zhou, C. (2002), 'The synchronization of chaotic systems', *Physics reports* **366**(1-2), 1–101. Cited on page/s 26.

Brogli, R., Kröner, N., Sørland, S. L., Lüthi, D. & Schär, C. (2019), 'The role of hadley circulation and lapse-rate changes for the future european summer climate', *Journal of Climate* **32**(2), 385–404. Cited on page/s 88.

Brossier, C. L., Béranger, K., Deltel, C. & Drobinski, P. (2011), 'The mediterranean response to different space–time resolution atmospheric forcings using perpetual mode sensitivity simulations', *Ocean Modelling* **36**(1-2), 1–25. Cited on page/s 88.

Brossier, C. L., Béranger, K. & Drobinski, P. (2012), 'Sensitivity of the northwestern mediterranean sea coastal and thermohaline circulations simulated by the 1/12-resolution ocean model nemo-med12 to the spatial and temporal resolution of atmospheric forcing', *Ocean Modelling* **43**, 94–107. Cited on page/s 88.

Cai, W. & Cowan, T. (2008), 'Dynamics of late autumn rainfall reduction over southeastern australia', *Geophysical Research Letters* **35**(9). Cited on page/s 48.

Cai, W., Cowan, T. & Sullivan, A. (2009), 'Recent unprecedented skewness towards positive indian ocean dipole occurrences and its impact on australian rainfall', *Geophysical Research Letters* **36**(11). Cited on page/s 72.

Cai, W., Sullivan, A. & Cowan, T. (2009), 'Rainfall teleconnections with indo-pacific variability in the wcrp cmip3 models', *Journal of climate* **22**(19), 5046–5071. Cited on page/s 72.

Campuzano, S. A., De Santis, A., Pavón-Carrasco, F. J., Osete, M. L. & Qamili, E. (2016), 'Transfer entropy between south atlantic anomaly and global sea level for the last 300 years', *Natural Hazards and Earth System Sciences Discussions* pp. 1–17. Cited on page/s 7, 27.

Campuzano, S., De Santis, A., Pavón-Carrasco, F. J., Osete, M. L. & Qamili, E. (2018), 'New perspectives in the study of the earth's magnetic field and climate connection: The use of transfer entropy', *PloS One* **13**(11), e0207270. Cited on page/s 62.

Choudhary, A., Dimri, A. & Maharana, P. (2018), 'Assessment of cordex-sa experiments in representing precipitation climatology of summer monsoon over india', *Theoretical and Applied Climatology* **134**(1), 283–307. Cited on page/s 61.

Chowdary, J. S., Bandgar, A. B., Gnanaseelan, C. & Luo, J.-J. (2015), 'Role of tropical indian ocean air–sea interactions in modulating indian summer monsoon in a coupled model', *Atmospheric Science Letters* **16**(2), 170–176. Cited on page/s 60.

Christensen, J. H., Carter, T. R. & Giorgi, F. (2002), 'Prudence employs new methods to assess european climate change', *EOS, Transactions American Geophysical Union* **83**(13), 147–147. Cited on page/s 53.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. & Jones, P. D. (2018), 'An ensemble version of the e-obs temperature and precipitation data sets', *Journal of Geophysical Research: Atmospheres* **123**(17), 9391–9409. Cited on page/s 88.

Craig, A., Valcke, S. & Coquart, L. (2017), 'Development and performance of a new version of the oasis coupler, oasis3-mct_3. 0', *Geoscientific Model Development* **10**(9), 3297–3308. Cited on page/s 20, 90.

Davies, H. & Turner, R. (1977), 'Updating prediction models by dynamical relaxation: An examination of the technique', *Quarterly Journal of the Royal Meteorological Society* **103**(436), 225–245. Cited on page/s 9.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P. et al. (2011), 'The era-interim reanalysis: Configuration and performance of the data assimilation

system', *Quarterly Journal of the royal meteorological society* **137**(656), 553–597. Cited on page/s 22, 68.

Delgado-Bonal, A., Marshak, A., Yang, Y. & Holdaway, D. (2020), 'Analyzing changes in the complexity of climate in the last four decades using merra-2 radiation data', *Scientific reports* **10**(1), 1–8. Cited on page/s 7.

Djurdjevic, V. & Rajkovic, B. (2008), Verification of a coupled atmosphere-ocean model using satellite observations over the adriatic sea, *in* 'Annales Geophysicae', Vol. 26, Copernicus GmbH, pp. 1935–1954. Cited on page/s 137.

Dobler, A. & Ahrens, B. (2011), 'Four climate change scenarios for the indian summer monsoon by the regional climate model cosmo-clm', *Journal of Geophysical Research: Atmospheres* **116**(D24). Cited on page/s 61.

Doms, G. & Baldauf, M. (2011), 'A description of the nonhydrostatic regional cosmo-model–part i: Dynamics and numerics consortium for small-scale modelling', *Deutscher Wetterdienst, Offenbach, Germany* . Cited on page/s 90.

Doms, G., Förstner, J., Heise, E., Herzog, H., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P. et al. (2011), 'A description of the nonhydrostatic regional cosmo model, part ii: Physical parameterization', *Deutscher Wetterdienst, Offenbach, Germany* . Cited on page/s 90.

Doms, G., Förstner, J., Heise, E., Herzog, H., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J. et al. (2005), 'A description of the nonhydrostatic regional model lm. part ii: Physical parameterization.–deutscher wetterdienst'. Cited on page/s 18, 68.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P. et al. (2021), 'The ec-earth3 earth system model for the climate model intercomparison project 6', *Geoscientific Model Development Discussions* pp. 1–90. Cited on page/s xxxii, 107.

Drobinski, P., Da Silva, N., Bastin, S., Mailler, S., Muller, C., Ahrens, B., Christensen, O. B. & Lionello, P. (2020), 'How warmer and drier will the mediterranean region be at the end of the twenty-first century?', *Regional Environmental Change* **20**(3), 1–12. Cited on page/s 88.

ESRL, N. (n.d.), 'El niño southern oscillation (enso)'.
URL: *https://psl.noaa.gov/enso/* Cited on page/s 20.

Evans, J. P., Di Virgilio, G., Hirsch, A. L., Hoffmann, P., Remedio, A. R., Ji, F., Rockel, B. & Coppola, E. (2021), 'The cordex-australasia ensemble:

evaluation and future projections', *Climate Dynamics* **57**(5), 1385–1401. Cited on page/s 88.

Feba, F., Govardhan, D., Tejavath, C. & Ashok, K. (2021), 'Enso modoki teleconnections to indian summer monsoon rainfall—a review', *Indian Summer Monsoon Variability* pp. 69–90. Cited on page/s 114.

Finn, C. & Lizier, J. T. (2018), 'Pointwise partial information decomposition using the specificity and ambiguity lattices', *Entropy* **20**(4), 297. Cited on page/s 63.

Fischer, A. S., Terray, P., Guilyardi, E., Gualdi, S. & Delecluse, P. (2005), 'Two independent triggers for the indian ocean dipole/zonal mode in a coupled gcm', *Journal of climate* **18**(17), 3428–3449. Cited on page/s 48.

Fischer, E. M., Beyerle, U. & Knutti, R. (2013), 'Robust spatially aggregated projections of climate extremes', *Nature Climate Change* **3**(12), 1033–1038. Cited on page/s 86.

Freedman, D. & Diaconis, P. (1981), 'On the histogram as a density estimator: L 2 theory', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **57**(4), 453–476. Cited on page/s 31.

Fuhrer, O., Osuna, C., Lapillonne, X., Gysi, T., Cumming, B., Bianco, M., Arteaga, A. & Schulthess, T. C. (2014), 'Towards a performance portable, architecture agnostic implementation strategy for weather and climate models', *Supercomputing frontiers and innovations* **1**(1), 45–62. Cited on page/s xxix, 18, 68.

Gadgil, S. (2003), 'The indian monsoon and its variability', *Annual Review of Earth and Planetary Sciences* **31**(1), 429–467. Cited on page/s 59.

Gámiz-Fortis, S., Esteban-Parra, M., Pozo-Vázquez, D. & Castro-Díez, Y. (2011), 'Variability of the monthly european temperature and its association with the atlantic sea-surface temperature from interannual to multidecadal scales', *International Journal of Climatology* **31**(14), 2115–2140. Cited on page/s 52, 55.

Gangoiti, G., Sáez de Cámara, E., Alonso, L., Navazo, M., Gómez, M., Iza, J., García, J., Ilardia, J. & Millán, M. (2011), 'Origin of the water vapor responsible for the european extreme rainfalls of august 2002: 1. high-resolution simulations and tracking of air masses', *Journal of Geophysical Research: Atmospheres* **116**(D21). Cited on page/s 87.

Garland, J., Jones, T. R., Neuder, M., White, J. W. & Bradley, E. (2019), 'An information-theoretic approach to extracting climate signals from deep polar ice cores', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29**(10), 101105. Cited on page/s 62.

Gerken, T., Ruddell, B. L., Yu, R., Stoy, P. C. & Drewry, D. T. (2019), 'Robust observations of land-to-atmosphere feedbacks using the information flows of fluxnet', *Npj Climate and Atmospheric Science* **2**(1), 1–10. Cited on page/s 26, 62.

Giorgi, F. (2006), 'Climate change hot-spots', *Geophysical research letters* **33**(8). Cited on page/s 88.

Giorgi, F. (2019), 'Thirty years of regional climate modeling: where are we and where are we going next?', *Journal of Geophysical Research: Atmospheres* **124**(11), 5696–5723. Cited on page/s 9.

Giorgi, F., Jones, C., Asrar, G. R. et al. (2009), 'Addressing climate information needs at the regional level: the cordex framework', *World Meteorological Organization (WMO) Bulletin* **58**(3), 175. Cited on page/s 18, 61.

Goodwell, A. E. & Kumar, P. (2017), 'Temporal information partitioning networks (tipnets): A process network approach to infer ecohydrologic shifts', *Water Resources Research* **53**(7), 5899–5919. Cited on page/s 32, 65.

Goswami, B. (1998), 'Interannual variations of indian summer monsoon in a gcm: External conditions versus internal feedbacks', *Journal of Climate* **11**(4), 501–522. Cited on page/s 59.

Goswami, B. N., Madhusoodanan, M., Neema, C. & Sengupta, D. (2006), 'A physical mechanism for north atlantic sst influence on the indian summer monsoon', *Geophysical Research Letters* **33**(2). Cited on page/s 59.

Granger, C. W. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica: journal of the Econometric Society* pp. 424–438. Cited on page/s 26, 34.

Griffith, V. & Koch, C. (2014), Quantifying synergistic mutual information, *in* 'Guided self-organization: inception', Springer, pp. 159–190. Cited on page/s 63.

Gutowski Jr, W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G. et al. (2016), 'Wcrp coordinated regional downscaling experiment (cordex): a diagnostic mip for cmip6', *Geoscientific Model Development* **9**(11), 4087–4095. Cited on page/s 18, 61.

Hahs, D. W. & Pethel, S. D. (2011), 'Distinguishing anticipation from causality: Anticipatory bias in the estimation of information flow', *Physical review letters* **107**(12), 128701. Cited on page/s 41.

Harris, I., Jones, P. et al. (2017), 'Cru ts4. 01: Climatic research unit (cru) time-series (ts) version 4.01 of high-resolution gridded data of month-by-month variation in climate (jan. 1901–dec. 2016)', *Centre for Environmental Data Analysis* **25**. Cited on page/s 21, 53.

Harris, I., Osborn, T. J., Jones, P. & Lister, D. (2020), 'Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset', *Scientific data* **7**(1), 1–18. Cited on page/s 88.

Hazeleger, W., Severijns, C., Semmler, T., Ştefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R. et al. (2010), 'Ec-earth: a seamless earth-system prediction approach in action', *Bulletin of the American Meteorological Society* **91**(10), 1357–1364. Cited on page/s xxix, 68.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B. et al. (2012), 'Ec-earth v2. 2: description and validation of a new seamless earth system prediction model', *Climate dynamics* **39**(11), 2611–2629. Cited on page/s 89, 90, 105, 107.

Heise, E., Ritter, B., Schrodin, R. & Wetterdienst, D. (2006), *Operational implementation of the multilayer soil model*, Citeseer. Cited on page/s 19, 68.

Hentgen, L., Ban, N., Kröner, N., Leutwyler, D. & Schär, C. (2019), 'Clouds in convection-resolving climate simulations over europe', *Journal of Geophysical Research: Atmospheres* **124**(7), 3849–3870. Cited on page/s 88.

Herrera, S., Fernández, J., Rodríguez, M. A. & Gutiérrez, J. M. (2010), 'Spatio-temporal error growth in the multi-scale lorenz'96 model', *Nonlinear Processes in Geophysics* **17**(4), 329–337. Cited on page/s 46, 47.

Hirahara, S., Ishii, M. & Fukuda, Y. (2014), 'Centennial-scale sea surface temperature analysis and its uncertainty', *Journal of Climate* **27**(1), 57–75. Cited on page/s 21, 49.

Ho-Hagemann, H. T. M., Gröger, M., Rockel, B., Zahn, M., Geyer, B. & Meier, H. (2017), 'Effects of air-sea coupling over the north sea and the baltic sea on simulated summer precipitation over central europe', *Climate Dynamics* **49**(11), 3851–3876. Cited on page/s 87.

Ho-Hagemann, H. T. M., Hagemann, S. & Rockel, B. (2015), 'On the role of soil moisture in the generation of heavy rainfall during the oder flood event in july 1997', *Tellus A: Dynamic Meteorology and Oceanography* **67**(1), 28661. Cited on page/s 87.

Hochman, A., Marra, F., Messori, G., Pinto, J. G., Raveh-Rubin, S., Yosef, Y. & Zittis, G. (2022), 'Extreme weather and societal impacts in the eastern mediterranean', *Earth System Dynamics* **13**(2), 749–777. Cited on page/s 86.

Hofstätter, M. & Chimani, B. (2012), 'Van bebber's cyclone tracks at 700 hpa in the eastern alps for 1961-2002 and their comparison to circulation type classifications', *Meteorologische Zeitschrift* pp. 459–473. Cited on page/s 86.

Hofstätter, M., Chimani, B., Lexer, A. & Blöschl, G. (2016), 'A new classification scheme of european cyclone tracks with relevance to precipitation', *Water Resources Research* **52**(9), 7086–7104. Cited on page/s 86, 91.

Hofstätter, M., Lexer, A., Homann, M. & Blöschl, G. (2018), 'Large-scale heavy precipitation over central europe and the role of atmospheric cyclone track types', *International Journal of Climatology* **38**, e497–e517. Cited on page/s 86.

Hong, C.-C., Li, T. & Kug, J.-S. (2008), 'Asymmetry of the indian ocean dipole. part i: observational analysis', *Journal of climate* **21**(18), 4834–4848. Cited on page/s 48.

Hrudya, P., Varikoden, H. & Vishnu, R. (2021), 'A review on the indian summer monsoon rainfall, variability and its association with enso and iod', *Meteorology and Atmospheric Physics* **133**(1), 1–14. Cited on page/s 59.

Imamovic, A., Schlemmer, L. & Schär, C. (2019), 'Mountain volume control on deep-convective rain amount during episodes of weak synoptic forcing', *Journal of the Atmospheric Sciences* **76**(2), 605–626. Cited on page/s 88.

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., Benestad, R., Boberg, F., Buonomo, E., Cardoso, R. M. et al. (2020), 'Regional climate downscaling over europe: perspectives from the euro-cordex community', *Regional Environmental Change* **20**(2), 1–20. Cited on page/s 11.

James, P., Stohl, A., Spichtinger, N., Eckhardt, S. & Forster, C. (2004), 'Climatological aspects of the extreme european rainfall of august 2002 and a trajectory method for estimating the associated evaporative source regions', *Natural Hazards and Earth System Sciences* **4**(5/6), 733–746. Cited on page/s 87.

James, R. G., Barnett, N. & Crutchfield, J. P. (2016), 'Information flows? a critique of transfer entropies', *Physical review letters* **116**(23), 238701. Cited on page/s 60.

JAMSTEC (n.d.), 'The indian ocean dipole'.
URL: *https://www.jamstec.go.jp/aplinfo/sintexf/e/iod/about_iod.html* Cited on page/s 20.

Jiang, P. & Kumar, P. (2019), 'Using information flow for whole system understanding from component dynamics', *Water Resources Research* **55**(11), 8305–8329. Cited on page/s 62.

Kaiser, A. & Schreiber, T. (2002), 'Information transfer in continuous processes', *Physica D: Nonlinear Phenomena* **166**(1-2), 43–62. Cited on page/s 93.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J. et al. (1996), 'The ncep/ncar 40-year reanalysis project', *Bulletin of the American meteorological Society* **77**(3), 437–472. Cited on page/s 21, 68.

Kang, M., Ruddell, B. L., Cho, C., Chun, J. & Kim, J. (2017), 'Identifying co2 advection on a hill slope using information flow', *Agricultural and forest meteorology* **232**, 265–278. Cited on page/s 31.

Kantz, H. & Schreiber, T. (2004), *Nonlinear time series analysis*, Vol. 7, Cambridge university press. Cited on page/s 15, 32, 65.

Kelemen, F. D., Primo, C., Feldmann, H. & Ahrens, B. (2019), 'Added value of atmosphere-ocean coupling in a century-long regional climate simulation', *Atmosphere* **10**(9), 537. Cited on page/s 11, 88.

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., Berntsen, T., Berglen, T., Boucher, O., Chin, M. et al. (2006), 'An aerocom initial assessment–optical properties in aerosol component modules of global models', *Atmospheric Chemistry and Physics* **6**(7), 1815–1834. Cited on page/s 90.

Kirchgessner, P., Tödter, J., Ahrens, B. & Nerger, L. (2017), 'The smoother extension of the nonlinear ensemble transform filter', *Tellus A: Dynamic Meteorology and Oceanography* **69**(1), 1327766. Cited on page/s 26.

Klein, S. A., Soden, B. J. & Lau, N.-C. (1999), 'Remote sea surface temperature variations during enso: Evidence for a tropical atmospheric bridge', *Journal of climate* **12**(4), 917–932. Cited on page/s 26, 47.

Knuth, K. H., Gotera, A., Curry, C. T., Huyser, K. A., Wheeler, K. R. & Rossow, W. B. (2013), 'Revealing relationships among relevant climate variables with information theory', *arXiv preprint arXiv:1311.4632* . Cited on page/s 5, 26, 31, 32, 36, 37, 61.

Kraskov, A., Stögbauer, H. & Grassberger, P. (2004), 'Estimating mutual information', *Physical review E* **69**(6), 066138. Cited on page/s 15, 33, 42, 66.

Krishnaswamy, J., Vaidyanathan, S., Rajagopalan, B., Bonell, M., Sankaran, M., Bhalla, R. & Badiger, S. (2015), 'Non-stationary and non-linear influence of enso and indian ocean dipole on the variability of indian monsoon rainfall and extreme rain events', *Climate Dynamics* **45**(1), 175–184. Cited on page/s 59.

Krug, A., Aemisegger, F., Sprenger, M. & Ahrens, B. (2022), 'Moisture sources of heavy precipitation in central europe in synoptic situations with vb-cyclones', *Climate Dynamics* pp. 1–19. Cited on page/s xxxi, 12, 86, 87, 88, 91, 92, 97, 103.

Krug, A., Pothapakula, P. K., Primo, C. & Ahrens, B. (2021), 'Heavy vb-cyclone precipitation: a transfer entropy application showcase', *Meteorologische Zeitschrift* . Cited on page/s 87, 89, 93, 94, 102.

Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G. & Cane, M. (2006), 'Unraveling the mystery of indian monsoon failure during el niño', *Science* **314**(5796), 115–119. Cited on page/s 59.

Kumar, P. & Gupta, H. V. (2020), 'Debates—does information theory provide a new paradigm for earth science?'. Cited on page/s xxvii, 3.

Lanzante, J. R. (1996), 'Lag relationships involving tropical sea surface temperatures', *Journal of Climate* **9**(10), 2568–2578. Cited on page/s 26.

Lau, N.-C. & Nath, M. J. (1996), 'The role of the "atmospheric bridge" in linking tropical pacific enso events to extratropical sst anomalies', *Journal of Climate* **9**(9), 2036–2057. Cited on page/s 48.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M. et al. (2019), 'The climex project: a 50-member ensemble of climate change projections at 12-km resolution over europe and northeastern north america with the canadian regional climate model (crcm5)', *Journal of Applied Meteorology and Climatology* **58**(4), 663–693. Cited on page/s 88.

Leutwyler, D., Fuhrer, O., Lapillonne, X., Lüthi, D. & Schär, C. (2016), 'Towards european-scale convection-resolving climate simulations with gpus: A study with cosmo 4.19', *Geoscientific Model Development* **9**(9), 3393–3412. Cited on page/s xxix, 68.

Liu, L., Yu, W. & Li, T. (2011), 'Dynamic and thermodynamic air–sea coupling associated with the indian ocean dipole diagnosed from 23 wcrp cmip3 models', *Journal of Climate* **24**(18), 4941–4958. Cited on page/s 72.

Lizier, J. T. (2014), 'Jidt: An information-theoretic toolkit for studying the dynamics of complex systems', *Frontiers in Robotics and AI* **1**, 11. Cited on page/s 15, 32, 33, 37, 49, 65, 94.

Lorenz, E. (1996), 'Predictability: a problem partly solved. in: Proceedings of seminar on predictability, 4–8 september 1995'. Cited on page/s 29, 45.

Lucas-Picher, P., Christensen, J. H., Saeed, F., Kumar, P., Asharaf, S., Ahrens, B., Wiltshire, A. J., Jacob, D. & Hagemann, S. (2011), 'Can regional climate models represent the indian monsoon?', *Journal of Hydrometeorology* **12**(5), 849–868. Cited on page/s 61.

Lungarella, M., Ishiguro, K., Kuniyoshi, Y. & Otsu, N. (2007), 'Methods for quantifying the causal structure of bivariate time series', *International journal of bifurcation and chaos* **17**(03), 903–921. Cited on page/s 44.

Luo, J.-J., Zhang, R., Behera, S. K., Masumoto, Y., Jin, F.-F., Lukas, R. & Yamagata, T. (2010), 'Interaction between el nino and extreme indian ocean dipole', *Journal of Climate* **23**(3), 726–742. Cited on page/s 48, 49.

Madec, G. (2008), 'Nemo ocean engine, note pôle modél, 27', *Inst Pierre-Simon Laplace, Paris* . Cited on page/s 88.

Mathias, L., Ludwig, P. & Pinto, J. G. (2021), 'The damaging tornado in luxembourg on 9 august 2019: towards better operational forecasts', *Weather* **76**(8), 264–271. Cited on page/s 86.

McSweeney, C., Jones, R., Lee, R. W. & Rowell, D. (2015), 'Selecting cmip5 gcms for downscaling over multiple regions', *Climate Dynamics* **44**(11), 3237–3260. Cited on page/s 83.

Messmer, M., Gómez-Navarro, J. J. & Raible, C. C. (2015), 'Climatology of vb cyclones, physical mechanisms and their impact on extreme precipitation over central europe', *Earth system dynamics* **6**(2), 541–553. Cited on page/s 86, 92.

Messmer, M., Gómez-Navarro, J. J. & Raible, C. C. (2017), 'Sensitivity experiments on the response of vb cyclones to sea surface temperature and soil moisture changes', *Earth system dynamics* **8**(3), 477–493. Cited on page/s 86, 87, 101, 102.

Messmer, M., Raible, C. C. & Gómez-Navarro, J. J. (2020), 'Impact of climate change on the climatology of vb cyclones', *Tellus A: Dynamic Meteorology and Oceanography* **72**(1), 1–18. Cited on page/s 87, 96.

Meyers, G., McIntosh, P., Pigot, L. & Pook, M. (2007), 'The years of el niño, la niña, and interactions with the tropical indian ocean', *Journal of Climate* **20**(13), 2872–2880. Cited on page/s 48.

Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y. & Ludwig, R. (2019), 'Detecting climate change effects on vb cyclones in a 50-member single-model ensemble using machine learning', *Geophysical Research Letters* **46**(24), 14653–14661. Cited on page/s 88, 96.

Nair, P., Chakraborty, A., Varikoden, H., Francis, P. & Kuttippurath, J. (2018), 'The local and global climate forcings induced inhomogeneity of indian rainfall', *Scientific Reports* **8**(1), 1–12. Cited on page/s 59.

Nishant, N. & Sherwood, S. C. (2021), 'How strongly are mean and extreme precipitation coupled?', *Geophysical Research Letters* **48**(10), e2020GL092075. Cited on page/s 86.

Nissen, K. M., Ulbrich, U. & Leckebusch, G. C. (2013), 'Vb cyclones and associated rainfall extremes over central europe under present day and climate change conditions', *Meteorologische Zeitschrift* **22**(6), 649–660. Cited on page/s 87.

Novelli, L., Wollstadt, P., Mediano, P., Wibral, M. & Lizier, J. T. (2019), 'Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing', *Network Neuroscience* **3**(3), 827–847. Cited on page/s 62.

Nowack, P., Runge, J., Eyring, V. & Haigh, J. D. (2020), 'Causal networks for climate model evaluation and constrained projections', *Nature communications* **11**(1), 1–11. Cited on page/s 62.

Oki, T. & Sud, Y. (1998), 'Design of total runoff integrating pathways (trip)—a global river channel network', *Earth interactions* **2**(1), 1–37. Cited on page/s 90.

Palmer, T., Branković, Č., Viterbo, P. & Miller, M. (1992), 'Modeling interannual variations of summer monsoons', *Journal of Climate* **5**(5), 399–417. Cited on page/s 59.

Panosetti, D., Schlemmer, L. & Schär, C. (2019), 'Bulk and structural convergence at convection-resolving scales in real-case simulations of summertime moist convection over land', *Quarterly Journal of the Royal Meteorological Society* **145**(721), 1427–1443. Cited on page/s 88.

Papadimas, C., Bartzokas, A., Lolis, C. & Hatzianastassiou, N. (2012), 'Sea-level pressure–air temperature teleconnections during northern hemisphere winter', *Theoretical and Applied Climatology* **108**(1), 173–189. Cited on page/s 52.

Pillai, P. A. & Chowdary, J. S. (2016), 'Indian summer monsoon intra-seasonal oscillation associated with the developing and decaying phase of el niño', *International Journal of Climatology* **36**(4), 1846–1862. Cited on page/s 59.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., Thépaut, J.-N. et al. (2016), 'Era-20c: An atmospheric reanalysis of the twentieth century', *Journal of Climate* **29**(11), 4083–4097. Cited on page/s 91.

Pothapakula, P. K., Primo, C. & Ahrens, B. (2019), 'Quantification of information exchange in idealized and climate system applications', *Entropy* **21**(11), 1094. Cited on page/s 61, 89, 93, 94, 95.

Pothapakula, P. K., Primo, C., Sørland, S. & Ahrens, B. (2020), 'The synergistic impact of enso and iod on indian summer monsoon rainfall in observations and climate simulations–an information theory perspective', *Earth System Dynamics* **11**(4), 903–923. Cited on page/s 89, 93, 105.

Primo, C., Kelemen, F. D., Feldmann, H., Akhtar, N. & Ahrens, B. (2019), 'A regional atmosphere–ocean climate system model (cclmv5. 0clm7-nemov3. 3-nemov3. 6) over europe including three marginal seas: on its stability and performance', *Geoscientific Model Development* **12**(12), 5077–5095. Cited on page/s 11, 88.

R Core Team, R. C. (2013), 'R: A language and environment for statistical computing'. Cited on page/s 92.

Rajeevan, M., Pai, D., Anil Kumar, R. & Lal, B. (2007), 'New statistical models for long-range forecasting of southwest monsoon rainfall over india', *Climate Dynamics* **28**(7), 813–828. Cited on page/s xxvii, 3.

Raschendorfer, M. (2001), 'The new turbulence parameterization of lm. cosmo newsletter, no. 1, consortium for small-scale modeling, offenbach, germany, 89–97'. Cited on page/s 18, 68.

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., Kent, E. C. & Kaplan, A. (2003), 'Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century', *Journal of Geophysical Research: Atmospheres* **108**(D14). Cited on page/s 21, 68, 137.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. (2002), 'An improved in situ and satellite sst analysis for climate', *Journal of climate* **15**(13), 1609–1625. Cited on page/s 137.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K. et al. (2011), 'Merra: Nasa's modern-era retrospective analysis for research and applications', *Journal of climate* **24**(14), 3624–3648. Cited on page/s 22, 68.

Ritter, B. & Geleyn, J.-F. (1992), 'A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations', *Monthly weather review* **120**(2), 303–325. Cited on page/s 18, 68, 90.

Rockel, B. & Geyer, B. (2008), 'The performance of the regional climate model clm in different climate regions, based on the example of precipitation', *Meteorologische Zeitschrift* **17**(4), 487–498. Cited on page/s 88.

Rockel, B., Will, A. & Hense, A. (2008), 'The regional climate model cosmo-clm (cclm)', *Meteorologische Zeitschrift* **17**(4), 347–348. Cited on page/s xxix, 18, 68.

Rosenblum, M. G., Pikovsky, A. S. & Kurths, J. (1997), 'From phase to lag synchronization in coupled chaotic oscillators', *Physical Review Letters* **78**(22), 4193. Cited on page/s 41.

Ruddell, B. L., Drewry, D. T. & Nearing, G. S. (2019), 'Information theory for model diagnostics: Structural error is indicated by trade-off between functional and predictive performance', *Water Resources Research* **55**(8), 6534–6554. Cited on page/s 62, 93.

Ruddell, B. L. & Kumar, P. (2009), 'Ecohydrologic process networks: 1. identification', *Water Resources Research* **45**(3). Cited on page/s 7, 26, 31.

Runge, J. (2015), 'Quantifying information transfer and mediation along causal pathways in complex systems', *Physical Review E* **92**(6), 062829. Cited on page/s 27.

Runge, J. (2018), 'Causal network reconstruction from time series: From theoretical assumptions to practical estimation', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**(7), 075310. Cited on page/s 29, 34, 39, 41, 49.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J. et al. (2019), 'Inferring causation from time series in earth system sciences', *Nature communications* **10**(1), 1–13. Cited on page/s 62, 95, 114.

Runge, J., Heitzig, J., Petoukhov, V. & Kurths, J. (2012), 'Escaping the curse of dimensionality in estimating multivariate transfer entropy', *Physical review letters* **108**(25), 258701. Cited on page/s 7, 53.

Runge, J., Petoukhov, V. & Kurths, J. (2014), 'Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models', *Journal of climate* **27**(2), 720–739. Cited on page/s xxvii, 3, 26, 66, 70.

Russo, E., Sørland, S. L., Kirchner, I., Schaap, M., Raible, C. C. & Cubasch, U. (2020), 'Exploring the parameters space of the regional climate model cosmo-clm 5.0 for the cordex central asia domain', *Geoscientific Model Development Discussions* **2020**, 1–33. Cited on page/s 88.

Ruti, P. M., Somot, S., Giorgi, F., Dubois, C., Flaounas, E., Obermann, A., Dell'Aquila, A., Pisacane, G., Harzallah, A., Lombardi, E. et al. (2016), 'Med-cordex initiative for mediterranean climate studies', *Bulletin of the American Meteorological Society* **97**(7), 1187–1208. Cited on page/s 89.

Sabeerali, C., Ajayamohan, R., Bangalath, H. K. & Chen, N. (2019), 'Atlantic zonal mode: an emerging source of indian summer monsoon variability in a warming world', *Geophysical Research Letters* **46**(8), 4460–4467. Cited on page/s 59.

Saji, N., Goswami, B. N., Vinayachandran, P. & Yamagata, T. (1999), 'A dipole mode in the tropical indian ocean', *Nature* **401**(6751), 360–363. Cited on page/s 48, 60, 71, 72.

Saji, N., Xie, S. & Yamagata, T. (2006), 'Tropical indian ocean variability in the ipcc twentieth-century climate simulations', *Journal of Climate* **19**(17), 4397–4417. Cited on page/s 72.

San Liang, X. (2008), 'Information flow within stochastic dynamical systems', *Physical Review E* **78**(3), 031113. Cited on page/s 35.

San Liang, X. (2011), 'Uncertainty generation in deterministic flows: Theory and application with an atmospheric jet stream model', *Dynamics of atmospheres and oceans* **52**(1-2), 51–79. Cited on page/s 26.

San Liang, X. (2014), 'Unraveling the cause-effect relation between time series', *Physical Review E* **90**(5), 052150. Cited on page/s 7, 28, 36, 37, 42, 48, 49.

San Liang, X. (2016), 'Information flow and causality as rigorous notions ab initio', *Physical Review E* **94**(5), 052201. Cited on page/s 28, 35.

San Liang, X. & Kleeman, R. (2005), 'Information transfer between dynamical system components', *Physical review letters* **95**(24), 244101. Cited on page/s 7, 28, 35.

San Liang, X. & Kleeman, R. (2007), 'A rigorous formalism of information transfer between dynamical system components. i. discrete mapping', *Physica D: Nonlinear Phenomena* **231**(1), 1–9. Cited on page/s 35.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D. et al. (2020), 'Kilometer-scale climate models: Prospects and challenges', *Bulletin of the American Meteorological Society* **101**(5), E567–E587. Cited on page/s 18.

Schelter, B., Winterhalder, M., Eichler, M., Peifer, M., Hellwig, B., Guschlbauer, B., Lücking, C. H., Dahlhaus, R. & Timmer, J. (2006), 'Testing for directed influences among neural signals using partial directed coherence', *Journal of neuroscience methods* **152**(1-2), 210–219. Cited on page/s 26.

Schlemmer, L., Schär, C., Lüthi, D. & Strebel, L. (2018), 'A groundwater and runoff formulation for weather and climate models', *Journal of Advances in Modeling Earth Systems* **10**(8), 1809–1832. Cited on page/s 19, 68, 88.

Schneider, U., Fuchs, T., Meyer-Christoffer, A. & Rudolf, B. (2008), 'Global precipitation analysis products of the gpcc', *Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation* **112**. Cited on page/s 21, 68.

Schreiber, T. (2000), 'Measuring information transfer', *Physical review letters* **85**(2), 461. Cited on page/s xiii, 5, 6, 16, 27.

Schrodin, R. & Heise, E. (2001), 'The multi layer version of the dwd soil model terra_lm–cosmo technical report no. 2'. Cited on page/s 90.

Schulz, J.-P., Vogel, G., Becker, C., Kothe, S., Rummel, U. & Ahrens, B. (2016), 'Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil', *Meteorologische Zeitschrift* **25**(5), 607–620.
URL: *http://dx.doi.org/10.1127/metz/2016/0537* Cited on page/s 90.

Scolozzi, R. & Geneletti, D. (2017), 'The anthroposphere as an anticipatory system: Open questions on steering the climate', *Science of The Total Environment* **579**, 957–965. Cited on page/s 41.

Scott, D. W. (1979), 'On optimal and data-based histograms', *Biometrika* **66**(3), 605–610. Cited on page/s 31.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M. et al. (2012), 'Changes in climate extremes and their impacts on the natural physical environment'. Cited on page/s 86.

Sevault, F., Somot, S., Alias, A., Dubois, C., Lebeaupin-Brossier, C., Nabat, P., Adloff, F., Deque, M. & Decharme, B. (2014), 'A fully coupled mediterranean regional climate system model: design and evaluation of the ocean component for the 1980–2012 period', *Tellus A: Dynamic Meteorology and Oceanography* **66**(1), 23967. Cited on page/s 90, 137.

Sevault, F., Somot, S. & Beuvier, J. (2009), 'A regional version of the nemo ocean engine on the mediterranean sea: Nemomed8 user's guide', *Note de centre* **107**, 2009. Cited on page/s 90.

Shannon, C. E. (1948), 'A mathematical theory of communications', *Bell Syst. Tech. J.* **27**, 379–423. Cited on page/s 93.

Shannon, C. E. (2001), 'A mathematical theory of communication', *ACM SIGMOBILE mobile computing and communications review* **5**(1), 3–55. Cited on page/s 4, 15, 26, 29, 62.

Shukla, R. P. & Huang, B. (2016), 'Interannual variability of the indian summer monsoon associated with the air–sea feedback in the northern indian ocean', *Climate Dynamics* **46**(5), 1977–1990. Cited on page/s 59.

Slingo, J. & Annamalai, H. (2000), '1997: The el niño of the century and the response of the indian summer monsoon', *Monthly Weather Review* **128**(6), 1778–1797. Cited on page/s 59.

Smirnov, D. A. (2013), 'Spurious causalities with transfer entropy', *Physical Review E* **87**(4), 042917. Cited on page/s 61.

Sodemann, H., Wernli, H. & Schwierz, C. (2009), 'Sources of water vapour contributing to the elbe flood in august 2002—a tagging study in a mesoscale model', *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* **135**(638), 205–223. Cited on page/s 87.

Somot, S., Sevault, F., Déqué, M. & Crépon, M. (2008), '21st century climate change scenario for the mediterranean using a coupled atmosphere–ocean regional climate model', *Global and Planetary Change* **63**(2-3), 112–126. Cited on page/s 11, 90.

Sørland, S. L., Brogli, R., Pothapakula, P. K., Russo, E., Van de Walle, J., Ahrens, B., Anders, I., Bucchignani, E., Davin, E. L., Demory, M.-E. et al. (2021), 'Cosmo-clm regional climate simulations in the coordinated regional climate downscaling experiment (cordex) framework: a review', *Geoscientific Model Development* **14**(8), 5125–5154. Cited on page/s 9, 11, 19, 88.

Sørland, S. L., Schär, C., Lüthi, D. & Kjellström, E. (2018), 'Bias patterns and climate change signals in gcm-rcm model chains', *Environmental Research Letters* **13**(7), 074017. Cited on page/s 105.

Soto-Navarro, J., Jordá, G., Amores, A., Cabos, W., Somot, S., Sevault, F., Macías, D., Djurdjevic, V., Sannino, G., Li, L. et al. (2020), 'Evolution of mediterranean sea water properties under climate change scenarios in the med-cordex ensemble', *Climate Dynamics* **54**(3), 2135–2165. Cited on page/s 138.

Sperber, K., Annamalai, H., Kang, I.-S., Kitoh, A., Moise, A., Turner, A., Wang, B. & Zhou, T. (2013), 'The asian summer monsoon: an intercomparison of cmip5 vs. cmip3 simulations of the late 20th century', *Climate dynamics* **41**(9), 2711–2744. Cited on page/s 83.

Sprenger, M., Fragkoulidis, G., Binder, H., Croci-Maspoli, M., Graf, P., Grams, C. M., Knippertz, P., Madonna, E., Schemm, S., Škerlak, B. et al. (2017), 'Global climatologies of eulerian and lagrangian flow features based on era-interim', *Bulletin of the American Meteorological Society* **98**(8), 1739–1748. Cited on page/s 91.

Steppeler, J., Doms, G., Schättler, U., Bitzer, H., Gassmann, A., Damrath, U. & Gregoric, G. (2003), 'Meso-gamma scale forecasts using the non-hydrostatic model lm', *Meteorology and atmospheric Physics* **82**(1), 75–96. Cited on page/s 18.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K. et al. (2013), 'Atmospheric component of the mpi-m earth system model: Echam6', *Journal of Advances in Modeling Earth Systems* **5**(2), 146–172. Cited on page/s xxix, 68.

Stips, A., Macias, D., Coughlan, C., Garcia-Gorriz, E. & Liang, X. S. (2016), 'On the causal structure between co2 and global temperature', *Scientific reports* **6**(1), 1–9. Cited on page/s 26, 28.

Stocker, T. F., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. et al. (2013), 'Contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change', *Climate change* **5**, 1–1552. Cited on page/s 86.

Tawia Hagan, D. F., Wang, G., San Liang, X. & Dolman, H. A. (2019), 'A time-varying causality formalism based on the liang–kleeman information flow for analyzing directed interactions in nonstationary climate systems', *Journal of Climate* **32**(21), 7521–7537. Cited on page/s 7, 114.

Taylor, G. I. (1962), 'Gilbert thomas walker, 1868-1958'. Cited on page/s xxvii, 3.

Thomas, M. & Joy, A. T. (2006), *Elements of information theory*, Wiley-Interscience. Cited on page/s 15, 33, 62.

Tiedtke, M. (1989), 'A comprehensive mass flux scheme for cumulus parameterization in large-scale models', *Monthly weather review* **117**(8), 1779–1800. Cited on page/s 18, 68, 90.

Tissot, G., Lozano-Durán, A., Cordier, L., Jiménez, J. & Noack, B. R. (2014), Granger causality in wall-bounded turbulence, *in* 'Journal of Physics: Conference Series', Vol. 506, IOP Publishing, p. 012006. Cited on page/s 26.

Tödter, J. & Ahrens, B. (2015), 'A second-order exact ensemble square root filter for nonlinear data assimilation', *Monthly Weather Review* **143**(4), 1347–1367. Cited on page/s 26, 46.

Trenberth, K. E. (1999), Conceptual framework for changes of extremes of the hydrological cycle with climate change, *in* 'Weather and climate extremes', Springer, pp. 327–339. Cited on page/s 86.

Trigo, R. M., Osborn, T. J. & Corte-Real, J. M. (2002), 'The north atlantic oscillation influence on europe: climate impacts and associated physical mechanisms', *Climate research* **20**(1), 9–17. Cited on page/s 52, 55.

Ueda, H. & Matsumoto, J. (2000), 'A possible triggering process of east-west asymmetric anomalies over the indian ocean in relation to 1997/98 el niño', *Journal of the Meteorological Society of Japan. Ser. II* **78**(6), 803–818. Cited on page/s 49, 55.

Ummenhofer, C. C., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J., Risbey, J. S., Gupta, A. S. & Taschetto, A. S. (2009), 'What causes southeast australia's worst droughts?', *Geophysical Research Letters* **36**(4). Cited on page/s 48.

Vaid, B. & Liang, X. S. (2018), 'The changing relationship between the convection over the western tibetan plateau and the sea surface temperature in the northern bay of bengal', *Tellus A: Dynamic Meteorology and Oceanography* **70**(1), 1–9. Cited on page/s 28.

van Bebber, W. (1891), *Die Zugstrassen der barometrischen Minima nach den Bahnenkarten der deutschen Seewarte für den Zeitraum 1875-1890*. Cited on page/s 86.

Vannitsem, S., Dalaiden, Q. & Goosse, H. (2019), 'Testing for dynamical dependence: Application to the surface mass balance over antarctica', *Geophysical Research Letters* **46**(21), 12125–12135. Cited on page/s 7.

Vicente, R., Wibral, M., Lindner, M. & Pipa, G. (2011), 'Transfer entropy—a model-free measure of effective connectivity for the neurosciences', *Journal of computational neuroscience* **30**(1), 45–67. Cited on page/s 6, 26.

Volosciuk, C., Maraun, D., Semenov, V. A., Tilinina, N., Gulev, S. K. & Latif, M. (2016), 'Rising mediterranean sea surface temperatures amplify extreme summer precipitation in central europe', *Scientific reports* **6**(1), 1–7. Cited on page/s 86.

Voss, H. U. (2000), 'Anticipating chaotic synchronization', *Physical review E* **61**(5), 5115. Cited on page/s 41.

Walker, G. T. (1924), 'Correlations in seasonal variations of weather. i. a further study of world weather', *Mem. Indian Meteorol. Dep.* **24**, 275–332. Cited on page/s 59.

Wanner, H., Brönnimann, S., Casty, C., Gyalistras, D., Luterbacher, J., Schmutz, C., Stephenson, D. B. & Xoplaki, E. (2001), 'North atlantic oscillation–concepts and studies', *Surveys in geophysics* **22**(4), 321–381. Cited on page/s 53.

Webster, P. J., Magana, V. O., Palmer, T., Shukla, J., Tomas, R., Yanai, M. & Yasunari, T. (1998), 'Monsoons: Processes, predictability, and the prospects for prediction', *Journal of Geophysical Research: Oceans* **103**(C7), 14451–14510. Cited on page/s 59.

Wernli, H. & Schwierz, C. (2006), 'Surface cyclones in the era-40 dataset (1958–2001). part i: Novel identification method and global climatology', *Journal of the atmospheric sciences* **63**(10), 2486–2507. Cited on page/s 91.

Wibral, M., Finn, C., Wollstadt, P., Lizier, J. T. & Priesemann, V. (2017), 'Quantifying information modification in developing neural networks via partial information decomposition', *Entropy* **19**(9), 494. Cited on page/s 62.

Wibral, M., Pampu, N., Priesemann, V., Siebenhühner, F., Seiwert, H., Lindner, M., Lizier, J. T. & Vicente, R. (2013), 'Measuring information-transfer delays', *PloS one* **8**(2), e55809. Cited on page/s 31, 49.

Wicker, L. J. & Skamarock, W. C. (1998), 'A time-splitting scheme for the elastic equations incorporating second-order runge–kutta time differencing', *Monthly Weather Review* **126**(7), 1992–1999. Cited on page/s 18.

Wicker, L. J. & Skamarock, W. C. (2002), 'Time-splitting methods for elastic models using forward time schemes', *Monthly weather review* **130**(8), 2088–2097. Cited on page/s 90.

Wiesenfeldt, M., Parlitz, U. & Lauterborn, W. (2001), 'Mixed state analysis of multivariate time series', *International Journal of Bifurcation and Chaos* **11**(08), 2217–2226. Cited on page/s 43.

Wilcox, E. M. & Donner, L. J. (2007), 'The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model', *Journal of Climate* **20**(1), 53–69. Cited on page/s 86.

Williams, P. L. & Beer, R. D. (2010), 'Nonnegative decomposition of multivariate information', *arXiv preprint arXiv:1004.2515* . Cited on page/s 17, 60, 63.

Wollstadt, P., Martínez-Zarzuela, M., Vicente, R., Díaz-Pernas, F. J. & Wibral, M. (2014), 'Efficient transfer entropy analysis of non-stationary neural time series', *PloS one* **9**(7), e102833. Cited on page/s 48.

Wollstadt, P., Sellers, K. K., Rudelt, L., Priesemann, V., Hutt, A., Fröhlich, F. & Wibral, M. (2017), 'Breakdown of local information processing may underlie isoflurane anesthesia effects', *PLoS computational biology* **13**(6), e1005511. Cited on page/s 31.

Yamagata, T., Behera, S. K., Luo, J.-J., Masson, S., Jury, M. R. & Rao, S. A. (2004), 'Coupled ocean-atmosphere variability in the tropical indian ocean', *Earth's Climate: The Ocean–Atmosphere Interaction, Geophys. Monogr* **147**, 189–212. Cited on page/s 48.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N. & Kitoh, A. (2012), 'Aphrodite: Constructing a long-term daily gridded precipitation dataset for asia based on a dense network of rain gauges', *Bulletin of the American Meteorological Society* **93**(9), 1401–1415. Cited on page/s 21, 68.

Yu, R., Ruddell, B. L., Kang, M., Kim, J. & Childers, D. (2019), 'Anticipating global terrestrial ecosystem state change using fluxnet', *Global change biology* **25**(7), 2352–2367. Cited on page/s 26.

Yun, K.-S. & Timmermann, A. (2018), 'Decadal monsoon-enso relationships reexamined', *Geophysical Research Letters* **45**(4), 2014–2021. Cited on page/s 59.

Zhu, J., Bellanger, J.-J., Shu, H. & Le Bouquin Jeannès, R. (2015), 'Contribution to transfer entropy estimation via the k-nearest-neighbors approach', *Entropy* **17**(6), 4173–4201. Cited on page/s 42, 113.

Zittis, G., Hadjinicolaou, P., Klangidou, M., Proestos, Y. & Lelieveld, J. (2019), 'A multi-model, multi-scenario, and multi-domain analysis of regional climate projections for the mediterranean', *Regional Environmental Change* **19**(8), 2621–2635. Cited on page/s 114.

Dissemination of research

Publications in peer reviewed journals

[1] ***Pothapakula, P.K*** , C. Primo, S. Soerland, B. Ahrens (2020). **The synergistic impact of ENSO and IOD on the Indian Summer Monsoon Rainfall in observations and climate simulations - an information theory perspective**. Earth System Dynamics. https://doi.org/10.5194/esd-11-903-2020

[2] ***Pothapakula, P.K***, Primo C, Ahrens B. (2019). **Quantification of Information Exchange in Idealized and Climate System Applications.** Entropy 2019, 21, 1094. https://doi.org/10.3390/e21111094

(This article was featured as cover story and issue cover for Entropy volume 21, issue 11, 2019 and also appeared as one of the highly accessed article in Entropy website for Dec, 2019)

[3] ***Pothapakula P.K***, Primo C, Ahrens B (2019). **Information exchange in the high dimensional idealized systems and in climate inter-model comparison**. Conference proceedings on Climate Informatics-2019. http://dx.doi.org/10.5065/y82j-f154

(A travel grant to participate in Climate Informatics-2019, Paris was offered based on this article)

[4] ***Pothapakula,* P.K**, Osuri, K.K, Pattanayak S, Mohanty U.C, Sil S, Nadimpalli R (2017). **Observational perspective of SST changes during life cycle of tropical cyclones over Bay of Bengal**. Natural Hazards. doi:10.1007/s11069-017-2945-9

[5] ***Pothapakula P.K***, Sil S, Mohanty U.C (2016). **Impact of High Resolution ROMS-SST on the Simulation of Tropical Cyclone Phailin Over Bay of Bengal using ARW Modeling System**, The Indian Ocean Bubble, Issue 5, Page 10 -11.

[6] Krug, A., ***Pothapakula P.K.***, C. Primo, B. Ahrens (2021). **Heavy Vb-cyclone precipitation: a transfer entropy application showcase**. MetZ. DOI: 10.1127/metz/2021/1071

[7] Soerland, S. L., Brogli, R., ***Pothapakula, P. K.***, Russo, E., Van de Walle, J., Ahrens, B., Anders, I., Bucchignani, E., Davin, E. L., Demory, M.-E., Dosio,

A., Feldmann, H., Früh, B., Geyer, B., Keuler, K., Lee, D., Li, D., van Lipzig, N. P. M., Min, S.-K., Panitz, H.-J., Rockel, B., Schar, C., Steger, C., and Thiery, W. (2021). **COSMO-CLM regional climate simulations in the Coordinated Regional Climate Downscaling Experiment (CORDEX) framework: a review**, Geosci. Model Dev. https://doi.org/10.5194/gmd-14-5125-2021

Manuscript under review

[1] ***Pothapakula, P. K.***, Krug, A., Anika, O.H., Timo, K., Ahrens, B. (2022). **Vb-cyclones and associated Mediterranean Sea state in regional coupled climate simulations: evaluation and projection** (under review Earth System Dynamics)

[2] Prein, A.F., Ban, N., Ou, T.M., Tang, J., Sakaguchi, K., Collier, E., Jayanarayanan, S., Li, L., Sobolowski, S., Chen, X., Zhou X., Sugimoto, S., Zou, L., Hasson, S., Ekstrom, M., ***Pothapakula, P. K.,*** Stuart, R., Steen-Larsen, H.C., Leung, R., Belusic, D., Kukulies, J., Curio J., Chen, D. (2022). **Convection-Permitting Third Pole Experiment - Towards Ensemble-Based Kilometer-Scale Climate Simulations over the Third Pole Region.** (under review climate dynamics)

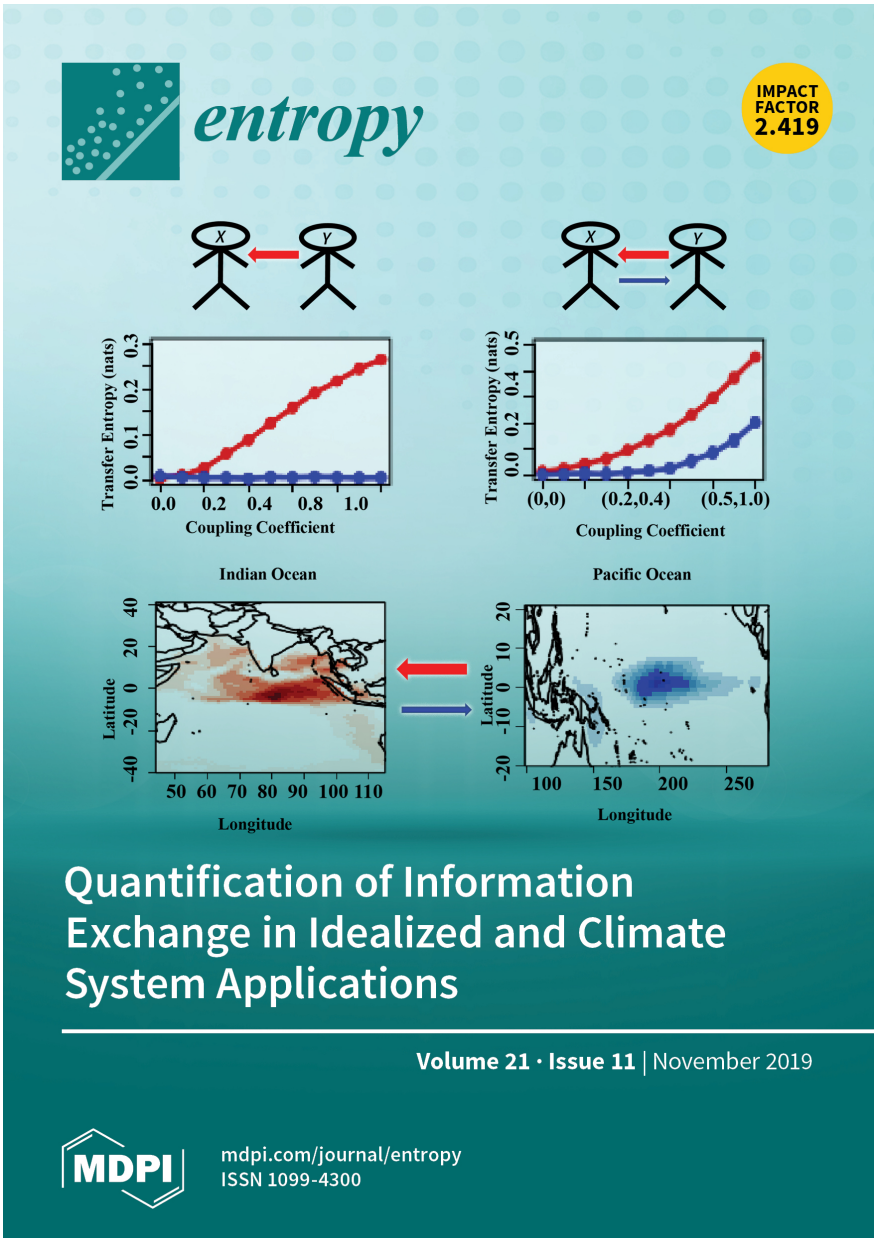


FIGURE D.1. Our work was selected as a cover image/featured article for the Entropy Journal first issue of 2019 (November).

Selected Conference participation (presentations/posters)

[1] **Pothapakula P K.**, Primo C, Ahrens B. (2021). **Exploring Information Exchange in Climate System Applications**. THE FIRST JOINT WCRP-WWRP SYMPOSIUM ON DATA ASSIMILATION AND REANALYSIS, BONN.

[2] **Pothapakula P. K.**, Primo C., Ahrens B. (2019). **Quantification of information exchange in idealized and climate system applications** . EGU

[3] **Pothapakula P.K.** and Ahrens B. (2018). **Comparison of transfer entropy and information flow methods in idealised simulations and real world applications**. Second work shop on information theory and Earth sciences, Santander, Spain.

[4] Ahrens B., **Pothapakula P K.**, Leps N., (2017). **Coupled Regional Climate Modeling Systems: Pros and Cons**. AGU

[5] **Pothapakula P K.**., Ahrens B. (2017). **Information flow associated with precipitation in coupled and uncoupled regional climate simulations**. EGU