UNIVERSITY OF READING

Department of Mathematics and Statistics

Causal network approaches for the study of sub-seasonal to seasonal variability and predictability

Elena Saggioro

A thesis submitted for the degree of Doctor of Philosophy

September 2022

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Elena Saggioro

Statement of Original Contribution

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate to this work has been explicitly indicated below.

The work in Chapter 3 has appeared in publication as follows:

- **Elena Saggioro** and Theodore G. Shepherd (2019) *Quantifying the timescale and strength of Southern Hemisphere intra-seasonal stratosphere-troposphere coupling.* Geophysical Research Letters 46 (22), 13479–13487.

In this work, I contributed to the design of the research. I developed the code and performed the analysis of the results. I led the writing of the manuscript.

The work in Chapter 4 has appeared in publication as follows:

- **Elena Saggioro**, Jana de Wiljes, Marlene Kretschmer and Jakob Runge (2020) *Reconstructing regime-dependent causal relationships from observational time series.* Chaos 30 (11), 113115.

In this work, I contributed to the design of the research. I developed most of the code and I performed the analysis. I led the writing of the manuscript.

Other work conducted during this PhD that was not explicitly addressed in this thesis has appeared in publication as follows:

- Marlene Kretschmer, Samantha V. Adams, Alberto Arribas, Rachel Prudden, Nicholas Robinson, **Elena Saggioro** and Thodore G. Shepherd (2021) *Quantifying causal pathways of teleconnections.* Bulletin of the American Meteorological Society 102 (12), E2247–E2263.

Abstract

Statistics is fundamental for climate science. It helps making sense of its complex and multi-scale features by characterizing its aggregated behaviour. However, statistical methods used to extract causal information about this behaviour are often based on correlation and pattern detection, which lack a causal interpretation. Causal networks are emerging as a framework that can bridge statistics and causal meaning.

This thesis seeks to contribute in advancing the use of causal network-based methods for the understanding of sub-seasonal to seasonal variability and predictability. The system of interest is the Southern Hemisphere mid-to-high latitude large-scale circulation variability in spring-to-summer, which is characterised by a strong downward influence of the stratospheric polar vortex on the tropospheric eddy-driven jet and its seasonal latitudinal shifts. The coupling extends the predictability of the troposphere due to the usually more predictable stratospheric dynamics.

In this thesis, firstly the strength and timescale of the coupling are estimated from reanalysis with a time-series causal network, revealing the biasing effect of the vortex internal dynamics on all cross-correlations with the jet. The detected coupling can explain the enhanced jet autocorrelations and the effect of ozone depletion on its poleward trend in the late 20th century. Secondly, the predictability of the coupled variability is studied with a Bayesian causal network with a large ensemble hindcast. Marginal predictability is found given long-lead drivers, such as El Niño Southern Oscillation and Polar Night Jet oscillation. The jet is highly predictable given the vortex state, also for its poleward shift, confirming a hypothesis present in the literature.

Motivated by the presence of non-stationarity in this system, a causal discovery algorithm for regime-dependent non-stationarity is proposed. Its skill is shown for a suite of synthetic systems and one real-world example.

Acknowledgements

I am grateful to my supervisor, Ted Shepherd, for giving me the opportunity to conduct this research. Being guided by his wise advice and constantly stimulated by his pioneering ideas has been an extremely rewarding and inspiring experience. This thesis would not have been possible without him.

I am grateful to my co-supervisors, Sebastian Reich and Jeff Knight, both of whom have found ways to be supporting mentors during my PhD, despite being based at different institutions and working on different research topics.

I am grateful to the Mathematics of Planet Earth Centre for Doctoral Training for having taken me through four years of growth in this beautiful area of science. An enormous thank you to all the staff who put their hearts in making this CDT the best possible for us students. A special thanks to Jennifer Scott, Sam Williams and Janet Fillingham for the truly supportive and caring environment you created at Reading.

Thanks to my brilliant MPE cohort Delta and to cohorts Echo and Bravo for the companionship and fun time spent together. A special thanks to Ieva, Jennifer, Mariana, Giulia, Swinda, Manuel, Ben, Sebastiano and Leonardo for being such loving friends: you have made my time in Reading something that I will treasure forever.

Thanks to all the members of Ted's research group for all their help, guidance and companionship during my years in Reading.

Thanks to Jana and Marlene for being the mentors, sponsors and friends I had the luck to encounter while doing science. You made me believe in myself in hard times, and taught me more than you can imagine.

Thanks to my best friends Chiara and Bianca for being always by my side despite the physical distance between us.

Thanks to my family for being a constant source of love and support that keeps me going in all circumstances.

Thanks to my partner Mirko, for making my life the best it could be.

CONTENTS

I Background 1

1 Introduction 3

1.1 Motivation . . . 3

1.2 Stratosphere-troposphere coupling in the extratropics . . . 5

1.2.1 Stratosphere and troposphere in the extratropics . . . 7

1.2.2 Observational characteristics . . . 9

1.2.3 Elements of dynamical theory . . . 10

1.2.4 Role of S-T coupling in extended predictability . . . 14

1.3 Southern Hemisphere stratosphere-troposphere coupling . . . 16

1.3.1 Tropospheric dynamics . . . 16

1.3.2 Stratospheric dynamics . . . 18

1.4 Causal theory for climate science . . . 20

1.5 Outline and contributions . . . 24

2 Causal Theory 27

2.1 Graphs and Probabilities . . . 28

2.1.1 Graphs . . . 28

2.1.2 Probabilities . . . 29

2.2 Bayesian Networks . . . 31

2.2.1 Definition of Bayesian Network . . . 33

2.2.2 Bayesian Inference . . . 34

2.3 Causal Networks . . . 34

2.3.1 The concept of intervention . . . 36
2.3.2 Definition of causal network . . . 37
2.3.3 Structural causal models: from probabilities to functions . . . 38
2.4 Causal Inference . . . 40
2.4.1 Blocked and open pathways . . . 41
2.4.2 Estimation of a causal effect . . . 43
2.4.3 Linear causal effects . . . 44
2.5 Causal Discovery . . . 45
2.5.1 Time-series constraint-based causal discovery . . . 45
2.5.2 PCMCI algorithm . . . 47
2.5.3 Interpretation of discovered causal networks . . . 51
2.6 Can we prove causality using observational data alone? . . . 52

II Results **53**

3 SH intra-seasonal stratosphere-troposphere coupling **55**
3.1 Introduction . . . 55
3.1.1 Statistical quantification of the S-T coupling . . . 56
3.1.2 SAM autocorrelations . . . 56
3.2 Data . . . 57
3.3 Methods . . . 58
3.4 Results . . . 60
3.4.1 The Confounding Effect of Stratospheric Autocorrelations . . . 60
3.4.2 Parametrization of the stratosphere-troposphere coupling . . . 62
3.4.3 Connection Between Ozone-induced Observed Circulation Trends . . . 68
3.5 Discussion . . . 72
3.6 Conclusion . . . 74

Appendices **75**
3.A Tropospheric drivers of the SAM autocorrelation time scale . . . 75
3.B The role of stratosphere in SAM autocorrelations . . . 76

3.C Mathematical derivation of autocorrelation effects . . . 77
3.D Trend rescaling . . . 78
3.E Effect of jet auto-dependence in the causal network . . . 79

4 Reconstructing regime-dependent causal relationships from time series 81

4.1 Introduction . . . 81
4.1.1 Existing literature on non-stationary causal discovery . . . 83
4.2 Problem setting . . . 85
4.2.1 Causal graphs . . . 86
4.2.2 Persistence . . . 87
4.2.3 Optimisation problem . . . 88
4.2.4 Motivating example . . . 89
4.3 Method . . . 90
4.3.1 Causal discovery . . . 91
4.3.2 Regime-dependent causal discovery . . . 92
4.3.3 Reconstruction of time series . . . 95
4.3.4 Parameter selection . . . 95
4.4 Results from numerical investigation . . . 96
4.4.1 Low dimensional data with two underlying regimes . . . 97
4.4.2 Low dimensional data with three underlying regimes . . . 107
4.4.3 Regime parameter selection . . . 108
4.4.4 High dimensional linear network . . . 110
4.4.5 Computational complexity . . . 113
4.5 Results from a real-world example: the effect of El Niño Southern Oscillation on Indian rainfall . . . 114
4.6 Conclusions . . . 116

Appendices 119

4.A Heterogeneous noise . . . 119
4.A.1 Gaussian noise with variable-specific variances . . . 119
4.A.2 Different noise distributions . . . 120

4.B Definition of result statistics . . . 121
4.B.1 Regime assigning process . . . 121
4.B.2 Link detection . . . 121
4.B.3 Link coefficients . . . 122
4.B.4 Prediction error . . . 122
4.C Abbreviations . . . 123

5 Bayesian causal network modelling of the SH S-T coupling 125
5.1 Introduction . . . 125
5.1.1 Stratospheric and tropospheric drivers of Southern Hemisphere large-scale circulation variability from late winter to early summer . . . 125
5.1.2 Studying predictability with Bayesian causal networks (BCN) . . . 130
5.2 Data . . . 132
5.2.1 Representation of stratospheric ozone . . . 133
5.2.2 Definition of variables . . . 134
5.3 Methods . . . 137
5.3.1 Building a BCN . . . 139
5.3.2 Probabilistic forecasts . . . 140
5.4 Results . . . 142
5.4.1 The BCN . . . 142
5.4.2 Predictability arising from long-lead drivers of EDJ . . . 152
5.4.3 Skill of BCN in predicting observed variability . . . 155
5.4.4 Connection between BCN variables and surface climate . . . 158
5.4.5 BCN-guided selection of most skilful forecast ensemble members . . . 161
5.5 Discussion . . . 166
5.5.1 Implications of results . . . 167
5.5.2 Limitations . . . 168
5.6 Conclusions . . . 169

Appendices 171
5.A Biases in System 4 . . . 171

5.B Network Variables: time series and distribution . . . 173
5.C Conditional probabilities validated with ERA5 . . . 175
5.D Effect of QBO on [u] . . . 176
5.E Derivation of link strength formula . . . 177
5.F Conditional Probability Tables . . . 179
5.G Precursors probabilities of SPVlow, VB and EDJ-1 . . . 180
5.H PR AUC of the BCN forecast compared with the conditional hindcast . . . 181
5.I Hindcast two-variate composites of Z500, 2mT and MTPR . . . 182
5.J ERA5 vs hindcast univariate composites for Z500, 2mT and MTPR . . . 184
5.K Probability of true positives and negatives for a climatological forecast . . . 187

III Final Remarks **189**

6 Conclusion **191**
6.1 Summary . . . 191
6.1.1 Timescale and strength of the S-T coupling, and its role under external forcing . . . 192
6.1.2 Detection of regime-dependent dynamics using a causal network discovery algorithm . . . 192
6.1.3 Long-lead predictability of the S-T coupling interannual variability . . . 193
6.2 Discussion . . . 194
6.3 Future work . . . 196

References **200**

List of Figures

1.1 Monthly-mean climatologies of zonal-mean zonal wind 9
1.2 Schematic of stratosphere-troposphere coupling dynamics 11
1.3 Schematic of drivers of SH spring to summer circulation variability 21
1.4 Scenarios leading to a correlation between process X and Y without a direct causation 23

2.1 Comparison between a Bayesian network and a causal network under intervention 32
2.2 Schematic of a time series causal network 39
2.3 Elemental structures of a causal network and the implied dependencies 41
2.4 Different approaches to calculate the influence of X on V at lag 2: cross-correlation, Granger causality and PC-algorithm 46

3.1 Correlations and MIT partial correlations matrices for the monthly polar vortex and the jet in the spring to early summer months 61
3.2 Downward and upward MIT partial correlation and its significance level for varying time averaging and time windows 63
3.3 Graphical representation of the detected time-series causal network for the S-T coupling 67
3.4 Comparison between reanalysis and synthetic correlation structure of the vortex and jet 69
3.5 Graphical representation of the network describing the S-T coupling forced by ozone depletion 70
3.6 Comparison between observed and synthetic trends for vortex and jet due to ozone depletion 71

3.E.1 Synthetic trend of the jet when including jet auto-dependence 80

4.2.1 Schematic of the motivating example of a regime-dependent relationship between two variables 90

4.4.1 Results for example case *Sign* X^1X^2 99

4.4.2 Results for example case *Sign* X^1 101

4.4.3 Results for example case *Arrow direction* 102

4.4.4 Results for example case *Lag* 103

4.4.5 Results for example case *Causal effect* 104

4.4.6 Summary of Regime-PCMCI performance for low dimensional systems with two regime 105

4.4.7 Results for example case *Sign* X^1X^2 *and arrow direction* 109

4.4.8 Numerical investigation of AICc values for runs with different number of regimes 111

4.5.1 Climate example. (a) Prediction error for each annealing step, (b) Regime learning, (c) Network learning and (d) Seasonality of the regimes 117

5.1.1 Climatology of eddy-driven jet latitude and standard deviation between August and March 127

5.2.1 Daily vortex trend between 1981-2001 and 1995-2016 in the hindcast and ERA5 134

5.2.2 VB timeseries and EDJ-1 and EDJ-2 categories for the bias corrected hindcast . 138

5.4.1 Literature-informed initial BCN structure and selection of links that pass the conditional independence test procedure 143

5.4.2 Final BCN structure 148

5.4.3 Conditional probability ratio for VB, EDJ-1 and EDJ-2 149

5.4.4 Stratified 30-fold cross-validated ROC AUC 150

5.4.5 Probability distribution of EDJ-2 precursors 152

5.4.6 ROC AUC and PR AUC (%) for BCN predictions using as evidence the August, September, October-stratosphere or October variables 154

5.4.7 ROC AUC tables obtained by the BCN forecast, the conditional hindcast and the full hindcast assessed against ERA5 156

5.4.8 Hindcast Z500 anomalies composited according to EDJ-1(EDJ-2) and ENSO categories in SO (ND) . . . 158
5.4.9 Hindcast 2mT and MTPR anomalies composited according to EDJ-2 and ENSO categories in ND . . . 160
5.4.10 Z500, 2mT and MTPR SO correlation maps based on EDJ-1 filter . . . 163
5.4.11 Z500, 2mT and MTPR ND correlation maps based on EDJ-2 filter . . . 164
5.A.1 Bias correction of EDJ-1 and EDJ-2 latitude time series . . . 172
5.B.1 QBO, vT-flux, ENSO, PJO-1, PJO-2 and SPV-low time series for the bias corrected hindcast . . . 173
5.B.2 EDJ-1 and EDJ-2 time series for the bias corrected hindcast . . . 174
5.C.1 Conditional probabilities P(X | drivers) with X = vT-flux, PJO-2, SPV-low, VB . . . 175
5.C.2 Conditional probabilities P(X | drivers) with X = EDJ-1, EDJ-2 . . . 176
5.D.1 Hindcast composites of monthly zonal-mean u anomalies based on E or W QBO from August to November . . . 176
5.F.1 Conditional probability ratio for vT-flux, PJO-2 and SPV-low . . . 179
5.G.1 Probability distribution of SPV-low, VB and EDJ-1 precursors . . . 180
5.H.1 PR AUC (%) tables obtained by the BCN forecast, the conditional hindcast and the full hindcast assessed against ERA5 . . . 181
5.I.1 Hindcast Z500 anomalies composited according to VB and ENSO categories in ND 182
5.I.2 Hindcast 2mT and MTPR anomalies composited according to EDJ-1 and ENSO categories in SO . . . 183
5.J.1 Comparison between ERA5 and hindcast SO mean Z500, 2mT and MTPR composites based on EDJ-1 categories . . . 184
5.J.2 Comparison between ERA5 and hindcast ND mean Z500, 2mT and MTPR composites based on EDJ-2 categories . . . 185
5.J.3 Comparison between ERA5 and hindcast Z500 composites based on ENSO (in SO and ND) and on VB (in ND) . . . 186

LIST OF FIGURES

List of Tables

2.1 Graph notation. 30

3.1 PCMCI results for a 35 days time average and a Sept-Jan time window 66

4.2.1 Notation used throughout the chapter. 86
4.4.1 Artificial model configurations for low dimensional experiments with two regimes 100
4.4.2 Method parameters for low dimensional experiments with two regimes 100
4.4.3 Results for low dimensional systems with two regimes 106
4.4.4 Artificial model configuration for a low dimensional experiment with three regimes 108
4.4.5 Method parameters for low dimensional experiments with three regimes 108
4.4.6 Results for low dimensional system with three regimes 110
4.4.7 High dimensional network parameters 112
4.4.8 Method parameters for high dimensional experiments with two regimes 112
4.4.9 Results for high-dimensional examples 112
4.4.10 Summary performance statistics of all examples 113
4.A.1 Results for example *sign* X_1X_2 120
4.A.2 Results for example *sign* X_1X_2 for different noise distributions 120
4.C.1 Abbreviations 123

5.2.1 Definition of the variables used to construct the BCN 135
5.2.2 Definition of categories for each variable 137
5.4.1 Result of the Step 2 tests of conditional independence 144
5.4.2 BCN links, their χ^2, p-value and $LS_{blind}\%$ 147

5.4.3 EDJ-1 and EDJ-2 probability of correct category prediction and opposite category prediction for the hindcast and the refined hindcast 165
5.K.1 SPV-low and VB probability of correct category prediction and opposite category prediction for the hindcast and the refined hindcast 188

Part I

Background

Chapter 1

Introduction

1.1 Motivation

Statistical analysis is ubiquitous in atmospheric and climate science. Statistical tools are fundamental to aggregate, correlate, quantify and interpret observations, as well as the large amount of model data at our disposal. While first principles and fluid dynamics equations underpin weather and climate phenomena, and our deeper understanding thereof, they are often only a guiding star in interpreting the complexity of the world we see through data. In the real world, many of the simplifying assumptions of our theories do not apply, or only in part. In addition, many of the large-scale phenomena of Earth's climate are not apparent from the underlying microscopic equations, since they "emerge" from multiple-scale and multi-actor interactions, i.e. the collective behaviour of the atmosphere, land and ocean. For example, various large-scale oscillation modes in the ocean and atmosphere, including the Quasi-Biennial Oscillation (QBO), the El Niño Southern Oscillation (ENSO) or the Madden-Julian oscillation (MJO), and their influence on far away parts of Earth's climate (teleconnections), emerge as well recognizable phenomena only when zooming out of the detailed microscopic interactions into the appropriate time and space scale. The climate itself can be seen as an emergent property of the weather: looking at day-to-day weather one cannot see it, but only looking at its aggregate behaviour year after year. Statistical analysis has led to the discovery of crucial elements of climate variability, such as ENSO. The atmospheric Southern Oscillation, describing the relationship of the surface air pressure between the Indian and Pacific Oceans, was discovered via correlation analysis by Walker (1925) and Walker and Bliss (1932).

This discovery led eventually to its connection with El Niño and La Niña, ocean phenomena of unusual warming or cooling of surface waters in the eastern tropical Pacific Ocean (Bjerknes, 1966, 1969).

Statistics is thus fundamental for atmospheric and climate science, helping us making sense of its complex and multi-scale features by characterizing its aggregated behaviour. However, statistical methods used for the purpose of extracting causal information are often based on correlation and pattern detection, which lack a clear causal interpretation. The lagged cross correlation between the time series of two climate indices is commonly used as follows: the peak of the cross correlation is often interpreted as identifying the time scale of the influence of the leading variable onto the other. Another example is the use of multiple linear regression (MLR), where the coefficients are interpreted as representing the strength of the casual influence of the drivers (regressors) onto the target (regressand). In the case of MLR, the causal interpretation is given before the analysis is done, as the regressors are identified a priori as causes. For lagged correlation this may not be the case, as sometimes the analysis is performed to discover which variable leads the other. Yet, as we will see, also this supposedly data driven discovery can be very flawed in its causal interpretation.

Indeed, data never speaks for itself and a modelling framework (with its causal assumptions) is always needed for causal interpretation, whether this is presented explicitly or not (Pearl, 2009b; Kretschmer et al., 2021). To quote Harold Jeffreys in his book *Theory of Probability*, "It is sometimes considered a paradox that the answer depends not only on the observations but on the question; it should be a platitude." (Jeffreys, 1939). Yet the disconnect between statistical practice and physical reasoning in climate science is still present (Shepherd, 2021). This is particularly evident in the case of correlations, where their physical and causal interpretation of the coefficients is often accompanied with the caveat that "correlation does not necessarily imply causation". This statement generally calls for either further and different investigations (e.g. with targeted modelling experiments, where causality can be more clearly isolated). Or it requires a theory-guided interpretation of the correlation coefficients, grounded in casual reasoning.

Many scientists have been arguing that the statistical analysis used in climate science needs to be rooted on stronger grounds that introduce more explicitly the concept of causality (e.g. Ebert-Uphoff and Deng, 2012; Runge et al., 2019a; Shepherd, 2021; Kretschmer et al., 2021).

Causal understanding is after all the ultimate goal of the application of statistics in climate science (Shepherd, 2021). Causal analysis, and its main modelling tool of causal networks (Pearl, 2000; Spirtes et al., 2000), have indeed been increasingly adopted in the climate community over the past few decades. Successful applications range from testing hypotheses (e.g. Kretschmer et al., 2016; Di Capua et al., 2020) to statistical forecasting (e.g. Di Capua et al., 2019; Polkova et al., 2021) and model assessment (e.g. Hirt et al., 2020; Nowack et al., 2020), supporting the call for blending standard statistics with causal tools in climate science.

The goal of this thesis is to advance our understanding and use of causal network-based methods for the statistical analysis of sub-seasonal to seasonal variability and predictability. The main system studied is the Southern Hemisphere mid-to-high latitude large-scale circulation variability between late austral winter and early summer. In this time of the year, the interaction between the stratosphere and troposphere is particularly strong in the downward direction, and this gives rise to the potential for extended predictability of the troposphere. Yet quantification of the coupling as well as the predictability arising from long-lead driver is still uncertain. The coupling also offers a promising pathway to study the stratospherically-mediated effect of anthropogenic climate change on atmospheric circulation, including the effect of stratospheric ozone depletion and recovery in the late 20^{th} century and early 21^{th} century, respectively.

The remainder of this chapter consists of an introduction to the stratosphere-troposphere dynamical coupling and its role in sub-seasonal to seasonal predictability (Section 1.2). The focus is then placed on the specific features of the Southern Hemisphere coupling between late austral winter and early summer (Section 1.3). Causal networks and their application to climate science studies are briefly introduced (Section 1.4) but an extensive mathematical description and associated statistical methods for estimation and discovery are found in Chapter 2. Finally, the open scientific questions on the use of causal networks in climate science addressed by this thesis are presented, and the specific contributions of this thesis are outlined (Section 1.5).

1.2 Stratosphere-troposphere coupling in the extratropics

The coupling between stratospheric and tropospheric dynamics (S-T coupling) has been a topic of major interest in the past few decades. In recent years, new results are shedding light on the various

mechanisms involved and the differences and commonalities between hemispheres (e.g. Kidston et al., 2015; Baldwin et al., 2021). Despite this progress, this remains a very active area of research due to the several and crucial remaining open questions, such as a lack of complete theoretical explanation (Kidston et al., 2015) and its untapped potential for sub-seasonal to seasonal (S2S) predictions (Scaife et al., 2022).

The possibility of significant stratospheric effects on the troposphere has important implications for the understanding of climate variability on time scales from weeks to decades (Butchart, 2022). For the month-to-month and year-to-year variability, the coupling might help in bridging the tropospheric circulation with variations in the solar cycle (e.g. Kodera and Kuroda, 2002; Ineson et al., 2011), inputs of volcanic aerosol (e.g. Muthers, 2014), the equatorial Quasi-Biennial Oscillation (QBO) (e.g. Gray et al., 2014; Rao et al., 2021; Anstey et al., 2022) and the Madden-Julian Oscillation (MJO) (e.g. Haynes et al., 2021). This is opening new perspectives for the improvement of S2S forecasts of Numerical Weather Prediction models (NWPs), still to be fully exploited (as reviewed in Tripathi et al., 2014; Domeisen et al., 2020b; Scaife et al., 2022). S-T coupling is further emerging as a key driver of extreme weather events (as reviewed in Domeisen and Butler, 2020).

The role of S-T coupling could be important also to better constrain the atmospheric circulation response to climate change. The stratosphere can be directly influenced by increased greenhouse gas (GHG) concentration and ozone depletion, and thus propagate this information downward via S-T coupling (e.g. Sigmond and Scinocca, 2010; Manzini et al., 2014; Ceppi and Shepherd, 2019; Mindlin et al., 2021). Reducing uncertainty in the representation of the mechanisms and variables involved in S-T coupling could reduce uncertainty in the tropospheric circulation changes. This is particularly salient since today many aspects of circulation changes remain uncertain, largely due to a lack of agreement among GCMs and of poor understanding of the driving mechanisms (Shepherd, 2014). Yet this aspect is crucially important to anticipate potential impact on humans and ecosystems. Changes to atmospheric circulation are expected to influence considerably both extreme weather events and seasonal patterns affecting, among others, food and energy security, water resources, biodiversity and the quality of life in urban and rural environments (IPCC, 2022). One of piece of the puzzle to achieve more confidence in long-term climate projection is to enhance the skill of S2S predictions with NWPs through better understanding and representation of S-T

coupling (Tripathi et al., 2014; Kidston et al., 2015; Domeisen et al., 2020b; Scaife et al., 2022).

The rest of this section is devoted to introducing the main features of the stratospheric and tropospheric dynamics relevant to this thesis (Section 1.2.1). Then S-T coupling is described through its main phenomenological (Section 1.2.2) and theoretical (Section 1.2.3) aspects. Finally, its importance for enhanced seasonal predictability of the troposphere is outlined (Section 1.2.4).

1.2.1 Stratosphere and troposphere in the extratropics

The Earth's atmosphere is conventionally divided into layers due to changing characteristics of the composition and dynamics with height. The lowest layer is called the troposphere, reaching from the surface to the tropopause and accounting for the bottom 10-20 km of the atmosphere ($\sim$ from 1000 hPa to 100 hPa in standard atmosphere). The troposphere is characterized by decreasing temperature with height, appreciable vertical motion and water vapour, and is where most of the weather occurs. A prominent feature of the mid-to-high latitude tropospheric circulation is the presence of a westerly current, called the mid-latitude jet stream, also known as *polar-front jet*. Jet streams are fast flowing, narrow bands of wind in the upper troposphere that circle their way around the globe. The two major jet streams, the polar-front jet and the subtropical jet, form where air masses of different temperatures converge. The greater the difference in temperature, the stronger the winds. They are predominantly westerly winds due to the rotation of the Earth. To a first approximation, the polar-front jet stream is supported by convergence of eddy momentum flux by baroclinc eddies (Held, 1975; Rhines, 1975) hence it is also known by the term *eddy-driven jet*. Eddies being fluid currents whose flow direction differs from that of the general flow. The latitude and intensity of the eddy-driven jet are sensitive to the changes in tilting and stretching of eddies. Due to its variable nature this jet can sometimes merge with the subtropical jet. Unlike the subtropical jet, however, the eddy-driven jet is deep so it can be separated from the former jet by studying the flow at lower levels. Note that in this thesis, the tropospheric eddy-driven jet will be sometimes referred to as jet for short.

The stratosphere is the region of the atmosphere extending from the top of the tropopause to the base of the mesosphere (or stratopause), roughly from 10-20 to 50 km ($\sim$ from 100 hPa to 1 hPa). The stratosphere is characterized by constant or increasing temperatures with increasing

height and marked vertical stability. The temperature profile characterizing the stratosphere results from high ozone concentrations (O_3) which absorb ultraviolet radiation (Mohanakumar, 2008). Although ozone is foremost produced in the tropical stratosphere, highest concentrations are found in the polar region. The redistribution of ozone by the meridional stratospheric circulation was first proposed by Brewer in 1949 and Dobson in 1956 and was subsequently called the Brewer-Dobson circulation (Brewer, 1949; Dobson, 1956). This global-scale cell transports upwelling air from the tropics poleward where it sinks and warms due to compression (Mohanakumar, 2008). During boreal summer, the temperatures in the Arctic stratosphere are even higher than in the tropics and as a result, a weak easterly flow surrounds the polar stratosphere. This distinguishes the stratospheric from the tropospheric circulation, where the westerly polar jet circles the globe year-long. In the winter hemisphere during the polar night, the lack of incoming solar radiation leads to a drastic cooling of the stratosphere at high latitudes. A strong temperature gradient between Equator and Pole then induces the formation of a band of fast moving westerlies in the stratosphere, also known as the *polar night jet* or polar vortex, which act as a transport barrier and isolates the cold pole air inside from mixing with warmer lower-latitude air. The vortex breaks down in spring, when solar radiation becomes stronger and forms again in the next cold season (Mohanakumar, 2008). Note that in this thesis, the term polar vortex will be used interchangeably to indicate both the stratospheric vortex and the polar night jet.

The stratosphere and troposphere are dynamically very different. The much less dense, but more highly stratified, dry and stable stratosphere with its strong winds cannot support the dynamics that characterize weather systems. For a long time, it was assumed that the stratosphere exhibits only very little intra-seasonal variability (Labitzke, 1999). The discovery in 1952 by German meteorologist Richard Scherhag of a rapid increase of NH stratospheric temperatures by about 30 degrees Celsius in a few days marked a change in understanding (Scherhag, 1952; Labitzke, 1999). This phenomenon is today famously known as a sudden stratospheric warming (SSW) (Butler et al., 2015; Baldwin et al., 2021). It is now well understood that the stratosphere experiences large intraseasonal to interannual variability, due to a number of mechanisms (see Section 1.3.2), and that this not only affects stratospheric dynamics but also the troposphere.

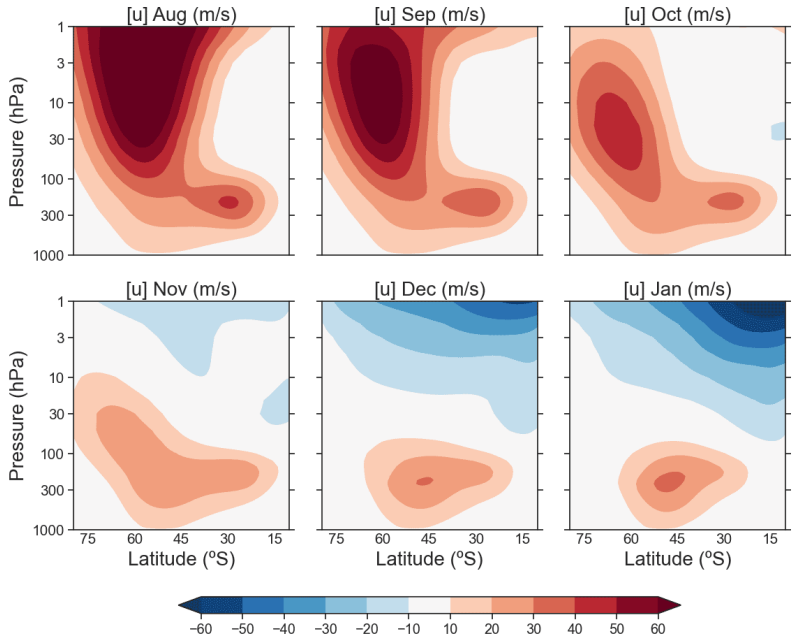

Figure 1.1: Monthly-mean climatologies of zonal-mean zonal wind. ERA-Interim, 1979-2017.

1.2.2 Observational characteristics

Investigations of coupled variability of the stratospheric and tropospheric circulations are a comparatively recent phenomenon in the meteorological literature. But today there is abundant evidence of the two-way coupling between the mid-to-high latitude tropospheric and stratospheric dynamics. Much of this work can be traced to a study by Kodera et al. (1990), which noted that lag-correlations between the averaged December zonal-mean zonal wind anomalies[1] at 1 hPa, 50° N and the averaged monthly zonal-mean zonal wind at each grid point over the vertical domain underneath appeared to migrate poleward and then downward over the course of 2-3 months. The dynamics of this poleward and downward migration in the NH are discussed in detail by Hitchcock (2012). A similar behaviour was documented for the SH by Kuroda and Kodera (1998) using a multiple empirical orthogonal function analysis (EOF) on zonal-mean zonal wind. The overall features of this progression were found to be similar to the NH, except that in the SH the entire sequence of evolution was closely locked to the annual cycle and that it required a longer period of 6 months (Baldwin and Dunkerton, 2001). It is now also established that the intensity of the SH coupling maximises between late spring and early summer, generally around the time of the SH stratospheric

[1] Anomalies here are deviations from a long-term monthly average.

polar vortex breakdown (Black and McDaniel, 2007; Byrne et al., 2017), as seen through both the coherent vertical structure of the zonal-mean zonal wind and geopotential height (Kuroda and Kodera, 1998; Thompson and Wallace, 2000).

The S-T coupling is clearly reflected in the correlated dynamics of the stratospheric polar vortex and the tropospheric eddy-driven jet. A colder stratosphere and strengthened vortex generally corresponds to a poleward shift of the jet, and conversely a weakened polar vortex corresponds to an equatorward shift (Kidston et al. (2015) and references therein). This effect is seen in both hemispheres and on time scales from seasonal (e.g. Ineson and Scaife, 2009) to decadal (e.g. Labitzke, 2005; Yamashita et al., 2015) to centennial (e.g. Scaife et al., 2012), thus the phenomenon seems to be rooted in some potentially fundamental mechanism. Whether in all these cases this arises from a downward, upward or mixed influence is not clear. However, in cases where forcing on the stratosphere is external to the troposphere (e.g. ozone hole, solar and with model experiments) the downward propagation is evident (e.g. McLandress et al., 2010; Tripathi et al., 2014). One example of the downward coupling that can be observed every year is around the time of the SH polar vortex breakdown, which is the focus of this thesis. More details on the dynamics of the Southern Hemisphere polar vortex and jet relevant to this thesis will be presented in Sections 1.3.1 and 1.3.2.

1.2.3 Elements of dynamical theory

Because the large-scale extratropical dynamics of the atmosphere is inherently vertically and horizontally non-local, changes in the stratosphere must affect the troposphere and vice versa, thus a two-way interaction seems to be inevitable. In fact there is abundant observational and modelling evidence for stratosphere-troposphere coupling in both directions. But while the physical mechanisms underpinning the coupling have been studied extensively, a comprehensive theory remains elusive, making the system of particular ongoing interest.

From a dynamical point of view, both the stratosphere and troposphere can be mathematically described via their mean flow and their wave (or eddy) components. The latter consists of planetary waves in the stratosphere, and also of synoptic scale eddies in the troposphere. Within each layer, the interaction between the mean flow and eddies is a two-way mechanism due to breaking and dissipation of the waves. The stratosphere and troposphere of course also interact, and this happens

via dynamics related to the mean circulation or thanks to downward and upward wave propagation. This fully coupled schematic is shown in Figure1.2, a reproduction from Haynes (2005a).

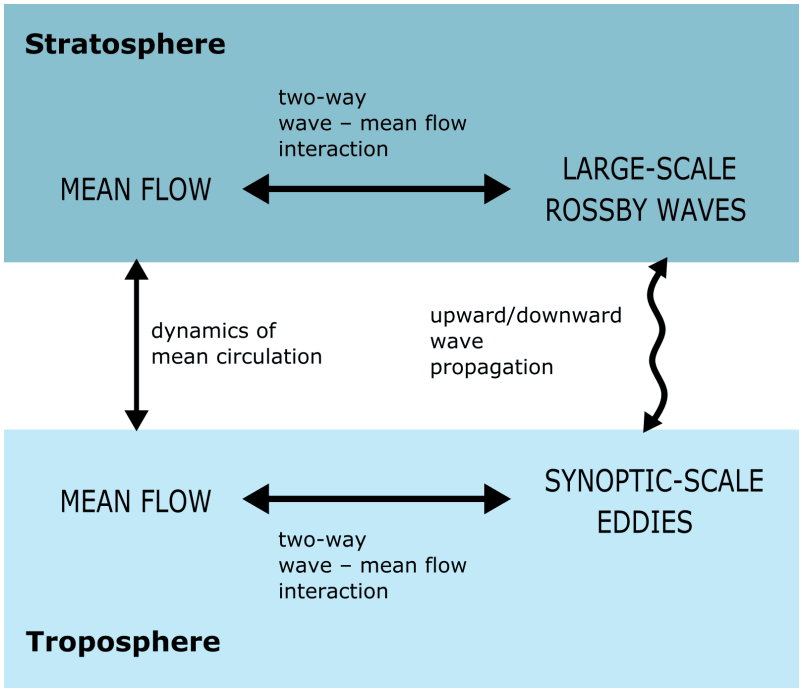

Figure 1.2: Schematic diagram indicating the role of different aspects of dynamical mechanisms in the stratosphere-troposphere coupling dynamics. A reproduction from Haynes (2005a). Note that 'dynamics of mean circulation' includes non-local PV inversion (or equivalently the short-time effect of the meridional circulation) and the effect of the meridional circulation on longer time scales, including the 'downward control' limit.

The upward dynamical influence of the troposphere on the stratosphere is the most understood aspect and the underlying theory is widely accepted. It happens through the upward propagation of waves, both low-frequency large-scale Rossby waves and high-frequency inertia-gravity waves (Haynes, 2005b), which can be generated from orography as well as tropical convection and extratropical weather systems. The waves can be either reflected back to the surface, or they can break and deposit easterly momentum, which rapidly warms the polar stratosphere and slows (and even reverses) the polar vortex winds. Simple theories of wave propagation, experiments in many different types of numerical models, and observational indicators all support this understanding (e.g. Charney and Drazin, 1961; Yoden et al., 2002; Haynes, 2005b). If the stratospheric westerly flow is sufficiently weak, this mechanism can give rise to (major) SSWs, impressive fluid dynamical events

in which large and rapid temperature increases in the winter polar stratosphere are associated with a complete reversal of the climatological wintertime westerly winds (Baldwin et al., 2021). SSWs are common in the NH, where they occur around every other year. SSWs can be termed *minor warming* if the stratospheric (e.g., at 10 hPa) temperature gradient between polar and midlatitudes reverses alone, and *major warming* if the polar night jet further reverses to easterly (Butler et al., 2015). The only observed SH SSW (major warming) was in 2002, and a near-SSW (minor warming) was seen in 2019. Studies of dynamically induced early breakdowns in the SH (major and minor warmings)(Hardiman et al., 2011; Hu et al., 2014; Jucker et al., 2021; Shen et al., 2022), found that events in this broad category have a frequency of about four events per decade (Shen et al., 2022), while actual SSWs is estimated to occur only once every 25 years or so (Jucker et al., 2021). The sharply contrasting behaviour between tropospheric wave sources in the NH and SH is the principal reason for such different winter evolutions of the extratropical stratospheres, with the SH having less topography and land-sea contrast which generates on average fewer waves perturbing the stratospheric polar vortex (Plumb, 1989; Scaife and James, 2000; Waugh and Polvani, 2010).

On the other hand, the downward influence of the stratosphere on the troposphere is still an open problem. Given the low mass of the stratosphere, its tropospheric impact was long regarded as minor. However, the impacts of events such as the SSW are hard to explain without introducing a downward influence (Butler et al., 2015; Domeisen and Butler, 2020; Baldwin et al., 2021).

There are a number of proposed mechanisms by which stratospheric variability might influence the troposphere. Following the classification in Tripathi et al. (2014), these can be broadly divided into three groups. The first is the large-scale adjustment in the troposphere to stratospheric potential vorticity anomalies; this mechanism describes the balanced geostrophic and hydrostatic response of the tropospheric flow to stratospheric potential vorticity (PV) anomalies, giving rise to changes in wind and temperature in the troposphere (e.g. Ambaum and Hoskins, 2002). If the restriction is made to zonal-mean fields, then the vertical nonlocality of PV inversion is precisely equivalent to a statement that a wave-induced force localised to the stratosphere will, through the instantaneous induced meridional circulation, give rise to an acceleration in the troposphere below and many of the papers on this topic have chosen the latter description (Haynes, 2005b). On longer timescales, in the presence of radiative damping, the meridional circulation tends to be narrower and deeper below (a ‘downward control’ response in the steady state limit) (e.g. Haynes

et al., 1991), potentially allowing an enhanced tropospheric response to a stratospheric wave force. The second is planetary wave–mean flow interaction, which involves what happens to upward propagating planetary-scale waves due to wave–mean flow interaction in the stratosphere. Not all the upward propagating planetary waves are absorbed or propagate in the stratosphere, but some can be reflected and propagate downward, cross the tropopause and continue to the troposphere, where they couple with the tropospheric planetary waves and subsequently modify the zonal mean flow. The potential reflection of upward propagating planetary waves occurs due to anomalous gradients in the stratospheric zonal wind when the stratospheric polar vortex is in certain states. The third mechanism accounts for the influences of stratospheric changes on the development or structure of tropospheric baroclinic systems, and for example include influences on eddy phase speed, influences on eddy length scales and more (see Tripathi et al., 2014 for a discussion).

At present there is no consensus in the literature on which, if any, of the above mechanisms can be considered as dominant to the downward influence. Nevertheless, overall it is believed that the coupling must involve al least two mechanisms: a forcing to communicate the stratospheric perturbation to the troposphere, and a tropospheric feedback mechanism to amplify the response. In addition, it is relatively well understood now that the wave–mean flow interactions associated with the critical layer mechanism for downward propagation (that originally proposed by Matsuno by which the zonal wind reversal migrates slowly downwards from the upper stratosphere) only reach the tropopause (e.g. Hitchcock and Haynes, 2016). Further, many studies have shown that the persistent lower-stratospheric anomalies are important for a continued tropospheric response (Hitchcock and Simpson, 2014; Maycock and Hitchcock, 2015; White et al., 2020), with it being mechanistically attributed to the induced meridional circulation by the lower stratospheric radiative recovery (Thompson et al., 2006; White et al., 2022).

Because the search for a mechanism has been so elusive while the phenomenon seems quite universal (see above), some works suggest that the tropospheric response to stratospheric perturbation may be just the intrinsic circulation response to any forcing (e.g. Gerber et al., 2008; Breul et al., 2022), reminiscent of the Fluctuation Dissipation Theorem (Gritsun and Branstator, 2007).

1.2.4 Role of S-T coupling in extended predictability

A relevant aspect of stratosphere-troposphere coupling is its potential to improve seasonal forecasts, including by increasing the lead-time of skilful forecasts both for anomalous conditions and for extreme weather events, and their impacts (Domeisen and Butler, 2020; Scaife et al., 2022, and references therein). Many numerical studies have clearly shown that an enhanced representation of the stratospheric state gives a more accurate forecast and longer time-window of predictability for the troposphere (e.g. as reviewed by Tripathi et al., 2014; Domeisen et al., 2020b; Scaife et al., 2022).

One modelling technique employed to assess the overall impact of stratospheric conditions on tropospheric forecasts is that of changing the representation of the stratosphere, by degrading (improving) it with lower (higher) stratospheric resolution or by raising/lowering the top level of the model. As an example for the SH, Roff et al. (2011) quantified the extended-range forecast skill that may be gained in spring by increasing stratospheric vertical resolution, being spring the time of the year when the downward S-T coupling is strong, and found increase skill in th stratosphere but also in the troposphere. Another experimental design is one of perturbation (i.e. adding additional artificial forcing to the stratosphere) and relaxation (i.e. damping the stratospheric state towards observations or other target states of interest) techniques, which allow to be more specific on the mechanism for different regions and at different times. These have been applied to extreme winter seasons in the NH and its S-T coupling (e.g. Douville, 2009; Hitchcock et al., 2013). Lastly, the approach of conditional hindcasting quantifies the stratospheric impact on tropospheric forecast skill by contrasting hindcasts with different stratospheric conditions, with the advantage of avoiding artificial perturbations (e.g. Sigmond et al., 2013; Byrne et al., 2019). A hindcast, also known as a historical re-forecast, is a forecast of a past period of time, which can therefore be compared with observations (or reanalysis). For example, Sigmond et al. (2013) use this approach to assess the model's ability to capture observed NAM indices, sea-level pressure, surface temperature and precipitation following SSWs in the NH. By comparing hindcast initialised at the onset date of SSW with an with an equivalent set of forecasts that are not initialized during SSWs, they show enhanced forecast skill for atmospheric circulation patterns, surface temperatures over northern Russia and eastern Canada and North Atlantic precipitation. In the SH and using a set of seasonal hindcast simulations initialised on the 1st August, Seviour et al. (2014) studied the variation of

hindcast skill with time and height of the Southern Annular Mode (SAM). They found that skillful predictions of the SAM are significantly influenced by stratospheric anomalies that descend with time and are coupled with the troposphere, with tropospheric skill re-emerging around mid October. This effect allows skilful statistical forecasts of the October mean SAM to be produced based only on midstratosphere anomalies on 1 August.

A common conclusion of all the above works is that a better representation of the stratosphere gives an enhanced predictability of the tropospheric behaviour. In particular, forecasts are generally found to be significantly better when the stratosphere is found in extreme vortex events, i.e. stratospheric events which represent a significant departure of the extratropical stratosphere from its climatological norm. This category mainly includes SSWs in the NH, final warmings in the SH and polar vortex intensification events (Tripathi et al., 2014). A number of recent modelling studies dedicated to the SH have provided robust evidence for extended-range predictability of the extratropical SH troposphere during austral spring and summer (Roff et al., 2011; Lim et al., 2013; Son et al., 2013; Seviour et al., 2014; Osman et al., 2015; Osman and Vera, 2016). The added skill arises in part from the stratosphere (e.g. Seviour et al., 2014; Byrne and Shepherd, 2018; Lim et al., 2018) and is robust to sampling uncertainty (Byrne et al., 2019).

The body of work developed on this topic in the last few decades has motivated operational forecasting centres to improve the resolution of the stratosphere in their numerical models to enhance tropospheric forecast skill (see summary in Domeisen et al. (2020a,b)). Better implementation of processes, such as gravity-wave drag (GWD) (Polichtchouk et al., 2018) and teleconnections, are avenues currently being explored to improve forecasts at S2S time scales. Note that better representation of GWD and its effect on the troposphere also require increased vertical resolution (Wicker et al., 2022), thus these strategies may need to work in synergy. Given the additional constraint of limited computational resources for NWP model runs in real time, another strategy under development is to prioritize model configurations that can most effectively reduce error growth. For example, in the case of Sigmond et al. (2013), the hindcasting suggests the occurrence of SSWs as the optimal re-initialization time for the NH. For the SH, the month of November is suggested as the one that might give the strongest predictability (Byrne and Shepherd, 2018), although there seem to be predictability from further ahead (e.g. Lim et al., 2018).

1.3 Southern Hemisphere stratosphere-troposphere coupling

The literature on the S-T coupling on the SH is less abundant than it is for the NH, although quite a few recent studies have focused on the austral summertime (e.g. Thompson et al., 2005; Son et al., 2013; Seviour et al., 2014; Byrne et al., 2017; Byrne and Shepherd, 2018; Polichtchouk et al., 2018; Byrne et al., 2019; Lim et al., 2018, 2019, 2021). Understanding the S-T coupling in the SH context can have several implications, such as providing better understanding of low-frequency variability in variations of Antarctic sea ice extent (e.g. Wang et al., 2021, 2020) and understanding the effect of tropical teleconnections on extra-tropical circulation variability, for example from ENSO (e.g. Domeisen et al., 2020b; Byrne et al., 2019) and the QBO (e.g. Anstey and Shepherd, 2014; Byrne and Shepherd, 2018; Yamashita et al., 2018).

1.3.1 Tropospheric dynamics

In the troposphere, the mid-latitude tropospheric large-scale circulation throughout the year is dominated by the eddy-driven jet (Section 1.2.1). Looking at the long term monthly averages, the main dynamical feature of the jet is that it undergoes a semi-annual oscillation in latitude (SAO, Van Loon, 1967). The strongest winds are located closest to the pole around the time of the equinoxes (Karoly and Vincent, 1998), corresponding to maxima in the tropospheric meridional temperature gradient due to different annual cycles of temperature over Antarctica and the surrounding open ocean (Van Loon, 1967; Hurrell and Loon, 1994; Meehl et al., 1998; Walland and Simmonds, 1999). This is in contrast with the NH, where winds exhibit more of an annual cycle with the strongest mid-latitude winds generally found in winter (Kallberg et al., 2005).

The jet vacillations strongly project on the Southern Annular Mode (SAM), also known as the Antartic Oscillation, which is the dominant mode of variability of the SH extratropical tropospheric circulation at timescales longer than a couple of weeks (e.g. Karoly, 1990; Kidson, 1988; Hartmann and Lo, 1998; Thompson and Wallace, 2000) and it impacts surface climate through the subtropics, extratropics, and polar regions of the SH on timescales from days to seasons (e.g. Thompson and Solomon, 2002; Silvestri and Vera, 2003; Sen Gupta and England, 2006; Lim et al., 2013; Swart et al., 2015). The SAM index is a time series, which describes the month-to-month variability in the SH extratropical circulation between the low pressure located over polar latitudes and the relatively high pressure at midlatitudes. Generally speaking, more positive indices of the SAM are

associated with both poleward shifts and strengthening of the jet; conversely more negative indices of the SAM are associated with both equatorward shifts and weakening of the jet (Hartmann and Lo, 1998), although the correspondence is not a one-to-one. A positive SAM also coincides with higher surface pressures and reduced precipitation in the midlatitudes and lower surface pressures and enhanced precipitation in high latitudes (Hartmann and Lo, 1998; Silvestri and Vera, 2003; Sen Gupta and England, 2006), and vice-versa for the negative SAM. The SAM is a widely adopted index to summarize jet variability and its phases are often used predict seasonal anomalies across the SH extra-tropics. However it is worth noting that the SAM is a very simplified yet useful description of the midlatitude circulation and its response to external forcing, but it does not constitute itself a driver (Mindlin et al., 2020).

Since the SH jet tends to be zonally symmetric, differently to its NH counterpart where localised jets are more prevalent (Hartmann and Lo, 1998), the zonal averages have been extensively used as an approximation for analysing SH mid-latitude circulation variability. Studies of the zonally-averaged flow of the SH also tend to also emphasise that the mid-latitude westerlies are approximately equivalent barotropic. Overall, the zonally-averaged and equivalent barotropic view of the jet is frequently used, and in this thesis this approximation will be adopted to represent the SH eddy-driven jet.

The variability of the SH eddy-driven jet (EDJ) can be characterised, to first approximation, as migrations in the meridional direction. These occur on all timescales and have been usually approximated as a Gaussian red-noise process with an e-folding timescale of about 10 days (Hartmann and Lo, 1998). No evidence for any preferred periodicities in this wandering of the jet has been found. More recently, its timing and amplitude between late spring and early summer have been shown to be strongly influenced by stratospheric variability, by means of the S-T coupling (Black and McDaniel, 2007; Byrne et al., 2017; Byrne and Shepherd, 2018). In particular, the variability appears to be dominated by the stratospheric signal, in contrast to unpredictable noisy behaviour, as a result of a consistent beginning of the shift of the jet towards the equator around the time of the vortex breakdown (thus influencing the timing of part of the SAO). In Byrne (2017), a similar suggestion was made for the variability of the jet around its poleward shift, which instead happens earlier in late winter/early spring. It was suggested that this shift is influenced by the lower stratospheric reach of the polar vortex in that time of the year. The role of the stratosphere in

early spring is supported also by the modelling study in Polichtchouk et al., 2018 where increasing the parametrized strength of nonorographic gravity wave drag is shown to remove the poleward transition of the SAO, as a result of a weakening of the vortex during the breakdown period. This hypothesis will be further explored and confirmed in this thesis (Chapter 5).

1.3.2 Stratospheric dynamics

The Southern Hemisphere winter-time high-latitude stratospheric westerlies (stratospheric polar vortex, SPV) are considerably stronger than the ones in the NH and exhibit much less daily variability (Karoly and Vincent, 1998), mostly due to less sources of upward wave fluxes in the SH due to less topography and land-sea contrast (as seen in Section 1.2.1). Another peculiar feature is that they shift downward as part of their seasonal cycle, with the strongest winds progressing from the upper stratosphere in August (late winter) to eventually reach the upper troposphere sometime in austral summer (Hartmann et al., 1984). The evolution of the zonal-mean zonal wind for the SH is shown in Figure 1.1. This systematic downward progression is unusual in the NH (Hardiman et al., 2011).

The distinct behaviour between the hemispheres also extends to the spring, due to contrasting evolutions in the vortex breakdown event. The SH vortex breakdown is in general driven by, and in lock-step with, the seasonal cycle, when springtime incoming of solar radiation warms the pole and destabilises the polar vortex. Thus in the SH the breakdown mostly coincides with what is called a *final warming* event. In the NH, the breakdown is more variable, with the event generally occurring between March and June, and less tied to the seasonal cycle. The larger interannual variability in the NH can in part be traced to larger vacillations in the vortex in the preceding winter (Hardiman et al., 2011). Despite its seasonal lock-step, the year-to-year variability of the SH vortex breakdown is also considerable: in ERA Interim the date spans from late October to early January (Byrne et al., 2017).

At least three separate mechanisms for interannual variability in the SH extratropical stratosphere have been suggested. These can be associated with a set of causal drivers of the SH stratospheric variability (and of the tropospheric variability via the S-T coupling), which are schematised in Figure 1.3.

The first mechanism is associated with interannual variability of the tropospheric circulation, in

particular of wave sources. Anomalously strong upward propagating planetary waves can advance the seasonal disruption of the SPV. The waves generally contribute to destabilise the state of the SPV with a lead time of one month. The wave perturbations are often quantified via the spatially weighted time-integrated upward wave activity flux at the lower stratosphere (vT-flux) (Rao et al., 2020; Lim et al., 2021; Reichler and Jucker, 2022). A major tropospheric source of upward propagating planetary waves is El-Niño Southern Oscillation (ENSO) (Van Loon et al., 1982; Lin et al., 2012; Byrne et al., 2019; Domeisen et al., 2020b; Stone et al., 2022).

A second important mechanism arises from the equatorial Quasi-Biennial Oscillation (QBO) (see a review in Butchart (2022)). Holton and Tan (1980) first found that the strength of the NH winter polar vortex correlated well with the phase of the equatorial QBO: a warmer, weaker vortex coincided with westward equatorial winds at 50 hPa (or easterly winds, hence QBO-E). In contrast, when the equatorial winds were eastward the vortex was colder and stronger. They also noted a QBO signal in the SH zonal wind, this time in spring, further confirmed in, e.g. Hio and Yoden (2005), Byrne and Shepherd (2018), Yamashita et al. (2018) and Byrne et al. (2019). Recently Garfinkel et al. (2018) showed predictability of the NH wintertime stratospheric polar vortex from the QBO in sub-seasonal forecast models. The mechanism underlying QBO–vortex coupling is generally believed to be that the prevailing winds in the tropical stratosphere influence the propagation and dissipation of equatorward-propagating extratropical planetary waves, leading to high-latitude changes in these waves and consequently in the polar vortex (Holton and Tan, 1980; Dunkerton and Baldwin, 1991), but the precise details of how this occurs remain unclear (Anstey and Shepherd, 2014; Anstey et al., 2022).

A third mechanism regards internal extratropical stratospheric variability. One element is the stratospheric ozone, with less ozone resulting in a cooling and strengthening of the vortex which leads to later a breakdown (Mohanakumar, 2008). Indeed studies have shown that anthropogenic ozone depletion during the last decades of the 20th century has resulted in a strengthened stratospheric polar vortex and the breakdown date was delayed by about two weeks (Randel and Wu, 1999; Waugh et al., 1999; McLandress et al., 2010; Young et al., 2013). The poleward and downward evolution of the so-called Polar night Jet Oscillation (PJO) in the high stratosphere, starting in late winter, also provides a source of variability to the vortex later in the spring (Kuroda and Kodera, 2001; Kuroda, 2002; Lim et al., 2018). Variability in the stratospheric circulation has

also been found to emerge as a result of a persistence of perturbations to the sub-tropical stratospheric circulation from one winter to the next (Scott and Haynes, 1998). This persistence arises because the smaller Coriolis parameter at low latitudes results in longer radiative damping times of perturbations compared to the extra-tropics. This so-called low-latitude 'flywheel' effect imprints a distinct two-year periodicity to variability in the mid-latitude stratosphere (Gerber et al., 2010; Byrne et al., 2016). Finally, a multiplicity of seasonal evolutions of the stratospheric polar vortex have been suggested, as a result of the inherently non-linear nature of primitive equation models of the stratosphere under forcing (e.g. Scott and Haynes, 2000, 2002). This view has been confirmed by Byrne et al. (2017) and Byrne and Shepherd (2018) which found that the variability of the SH SPV (and the downward coupling) can be seen as a non-stationary evolution of a deterministic seasonal cycle, rather than random variation around a climatological mean.

SPV variability can have significant impacts on the tropospheric weather. SH early vortex breakdown impacts mainly in the extratropics, and is associated with more likely hot and dry conditions in Australia and Antarctica and cold spells in south-eastern Africa and South America (Domeisen and Butler, 2020), and conversely for late vortex breakdown. To a large degree, the effects are associated with the earlier or later shift equatorward of the tropospheric jet due to earlier or later vortex breakdown (e.g. Byrne and Shepherd, 2018).

1.4 Causal theory for climate science

In the past few decades, causal networks have been used in numerous scientific fields. Early adopters were in econometrics (e.g. Granger, 1969), epidemiology (e.g. Greenland et al., 1999) and artificial intelligence (e.g. Pearl, 1988). In more recent years many examples have included behavioural sciences (e.g. Glymour, 2001), psychology (e.g. Rohrer, 2018), medical (e.g. Richens et al., 2020) and biological (e.g. Shipley, 2016) sciences.

Causal networks have been introduced to climate science more recently (Ebert-Uphoff and Deng, 2012). In the past decade, they have gained momentum and have been applied to a range of problems which stimulated the development of novel techniques to answer the requirements of the field. The first approach was to learn causal relationships from data and to test competing hypotheses involving multivariate teleconnections (e.g. Kretschmer et al., 2016; Samarasinghe et al., 2018; Di Capua et al., 2020). Causal networks have also been employed to study predictability

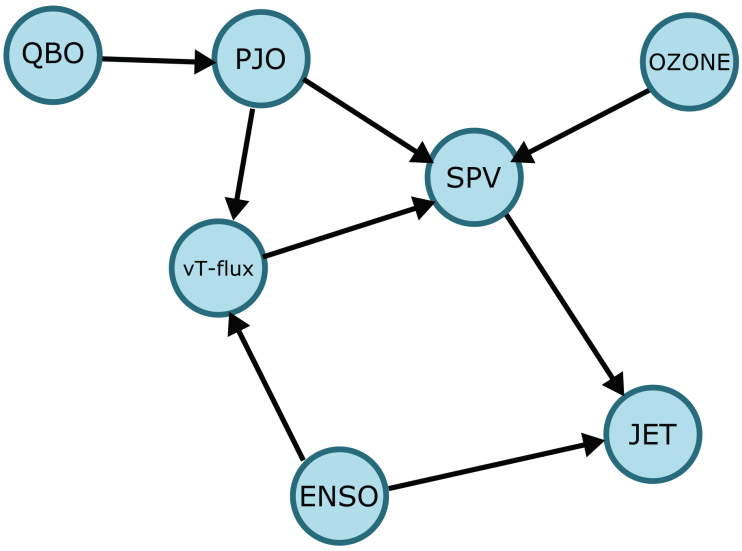

Figure 1.3: Schematic representing some of the drivers of the SH spring to summer large-scale zonal circulation variability.

on S2S time scales (e.g. Pfleiderer et al., 2020; Lehmann et al., 2020; Polkova et al., 2021), including for extreme polar vortex events in the NH (Kretschmer et al., 2017) or the onset and intensity of the Indian summer monsoon (Di Capua et al., 2019). They have been used also to quantify the causal effect of a specific mechanism, such as the role of Barents-Kara sea ice loss on projected polar vortex changes (Kretschmer et al., 2020). In the context of climate change, causal analysis has been applied to develop frameworks for attributing extreme weather events to climate change (e.g. Hannart et al., 2016), and to assess and compare climate models (e.g. Hirt et al., 2020; Nowack et al., 2020).

Despite these studies being increasingly more common and well accepted by the climate science community, it can be argued that causality still remains an elusive and controversial concept in the statistical analysis of climate data (Shepherd, 2021; Kretschmer et al., 2021), despite being in practice a concept widely implied in any such analysis. Climate scientists that work with numerical models, where synthetic experiments are possible, can more confidently connect their results with causal insight. That is, the causal effect can be directly seen as a result of interventions in a model, e.g. nudging of a variable, or the inactivation of a process. (Although, as will be described, due to common statistical pitfalls, also data analysis from model experiments can be prone to

causal misinterpretation.) The task of extracting causal information is harder for scientists that work with observational data alone, or that are interested in comparing mechanisms across models. In these cases, experiments are impossible and one is left with the task of correctly combining expert knowledge with mathematical and statistical tools to further the understanding. Statistical tools such as correlations, stratification (e.g. composite plots) and multiple linear regression are commonly used to study multi-variate relationships between elements of the climate system and are routinely interpreted as bearing causal meaning.

An informative example of a statistical method used in atmospheric science is lagged cross-correlation, which detects linear relationships at different time delays between two processes expressed via their time-series. Assuming the two processes are casually related, the lag at which the function peaks is often interpreted as capturing the time-scale and direction of the causal effect (Lorenz and Hartmann, 2001; Barnes and Hartmann, 2010). However, cross-correlations are known to be highly influenced by many effects: autocorrelations of the individual time-series, indirect connections via a third process, or a common driver (Runge et al., 2014). These biases are visualized with causal networks in Figure 1.4. Together with unobserved drivers, these are the reason behind the widely known statement that "correlation does not necessarily imply causation". The biasing effect of autocorrelations, for example, can be quantified in a simple two-variable linear system with one variable affecting the other and both being autocorrelated. The lagged-cross correlation (and its peak) can be easily shown to be strongly dependent on all the parameters of the system, including their respective autocorrelation coefficients (see the analysis in Runge et al., 2014). Thus strongly autocorrelated variables, which are very common in climate systems, are bound to inflate any cross-correlation if such effect is not controlled for (McGraw and Barnes, 2018). Yet, in absence of a better method, often-times the causal interpretation for lagged cross-correlations is taken, usually confirming rather than testing the already hypothesised plausible physical explanation.

While an understanding of the full Casual theory (Pearl, 2009b) is not straightforward, its fundamental object and modelling tool of a *causal network* is very intuitive. A causal network is a network that models the assumed causal relationships in a system: nodes are the variables and the directed links joining pairs of variables represent the causal relationships underpinning the functioning of the system. An example of a causal network of the SH S-T coupling variability

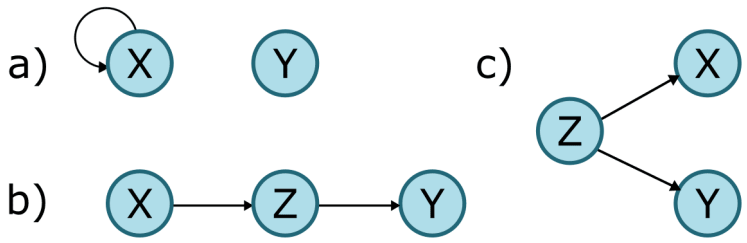

Figure 1.4: Possible scenarios leading to a correlation between process X and Y without a *direct* causation: (a) inflated correlation due to autocorrelation; (b) indirect chain via Z; (c) common driver Z. Note that in (b) the processes X and Y are causally related, but the effect is mediated by Z. While it can be argued that every (true) causal influence in climate has some mediator, from the causal network theory standpoint scenario (b) does not represent a direct causal effect from X to Y and thus has to be included in the possible biasing effects. Adapted from Figure 3 in Kretschmer et al., 2016.

in shown in Figure 1.3. Crucially, causal networks are the graphical scaffolding of mathematical objects with well defined properties. Such properties, rooted in basic logic and Bayesian statistics, allow to use data to test and quantify causal statements made about the system. Causal theory can be used to learn causal connections solely from data and a few modelling assumptions, in a process called *causal discovery*. Causal networks, either discovered from data or based on expert judgement, can be also used to disentangle the direct, indirect and multi-variable connections in a system and thus quantify the causal effect of individual links or pathways, an exercise called *causal inference*. In fact, the mathematical rules underpinning causal networks allow to identify unequivocally statistical confounders of any pair of variables, and thus to quantify the causal pathways of interest without suffering from statistical biases. It is worth noting that, in most systems, causal inference can be condensed in a few rules which lend themselves to relatively easy application. A causal approach to data analysis indeed often requires only minor modifications to standard practices in statistical analysis (Kretschmer et al., 2021), yet it bears substantial implications for the interpretation and robustness of the results. A detailed description of causal networks and the properties and methods used in this thesis is presented in Chapter 2.

1.5 Outline and contributions

The scope of causal network applications is growing as more opportunities are discovered for adapting them to the specific tasks of atmospheric and climate science. In this thesis, causal networks are used to address a few common, open challenges related to the quantification of causal effects from statistical relationships to understand sub-seasonal to seasonal variability and predictability. The following research questions are addressed:

- Can time-series causal networks be used to extract the timescale and strength of a specific physical link?
- Can causal network discovery be adapted to learn regime-dependent non-stationary causal relationships?
- Can probabilistic causal networks be used to quantify the predictive power of multiple drivers in the context of S2S forecasts?

The first and last questions are applied to the SH springtime S-T coupling to better quantitatively characterize it and the drivers of its variability. The second question is addressed with a proposed novel causal discovery algorithm for non-stationary time series.

In Chapter 3, time-series causal networks are used to extract the timescale and strength of an interaction in strongly autocorrelated time series. This is achieved by applying a well-known causal discovery algorithm (PCMCI, Runge (2018)) which enables to extract a robust signal from highly coupled and autocorrelated systems. The object of the study is SH S-T coupling, analysed via the zonal wind reanalysis time series. The results confirm from reanalysis that the variability of the stratospheric polar vortex is a predictor of the tropospheric eddy-driven jet between September and January. The vortex variability explains about 40% of monthly mean jet variability at a lead time of 1 month and can entirely account for the observed jet persistence. The derived statistical model can quantitatively connect the multidecadal trends observed in the vortex and jet during the satellite era, showing how short-term variability can help understand statistical links in long-term changes. The chapter is based on results published in Saggioro and Shepherd (2019).

Motivated by the non-stationarity of the SH downward coupling, which is organised every year around the variable timing of the stratospheric polar vortex breakdown, in Chapter 4 a novel

method for non-stationary time-series causal discovery is presented. Regime-dependence is a ubiquitous feature of climate systems, such as seasonal-dependent teleconnections. From a causal network perspective, this situation is the one of a causal link active only in part of the year (regime) and, crucially, the starting time can be variable from year to year and unknown a priori. Regime-PCMCI is introduced as an extension of the PCMCI causal discovery method (Runge et al., 2019a) to account for regime-dependent non-stationary causal relationships. The method is validated against a suite of synthetic time series data, and applied with success to detect the seasonally dependent relationship between ENSO and the intensity of the Indian Summer Monsoon. The chapter is based on results published in Saggioro et al. (2020).

In Chapter 5 a probabilistic Bayesian causal network (BCN) is used to quantify the probabilistic predictive power of multiple drivers. Using data from a large ensemble hindcast, the predictability of the eddy-driven jet's SAO variability based on knowledge of some of its major and well-known long-lead drivers, such as ENSO, QBO, PJO and upward wave fluxes in late winter, is quantified and contrasted with predictability given also shorter-lead drivers, chiefly the SPV in late spring. The BCN reproduces the hindcast variability to a good degree, and is able to match the prediction skill of a conditional hindcast when tested against reanalysis. The SPV in late spring is found to be a strong predictor of the SAO variability, confirming the crucial role of the stratosphere in this time of the year. On the other hand, the skill of the identified long-lead drivers to predict the jet is found to be comparatively smaller. This is largely due to their inability to well predict stratospheric variability in late spring, which suggests a dominant role of shorter-lead internal stratospheric variability. This chapter is based on as yet unpublished work.

Chapter 6 concludes the thesis by summarising the results and discussing the implications of the work for the broader field. Potential for future work stemming from the results of the thesis is briefly outlined.

Chapter 2

Causal Theory

Understanding causal relationships among different processes is a ubiquitous task in many scientific disciplines. The concept of causality is intuitive in many fields where experiments and manipulation of the system of interest are possible, or ground-truth laws of nature available. In other fields where such experiments are infeasible on unethical, such in Earth system sciences, and the behaviour of the systems cannot (always) be traced directly from first principles, the concept of causality has remained often vaguely defined.

Historically, within the field of statistics the possibility to extract causal information from statistical methods of data analysis had been strongly opposed, chiefly by Karl Pearson, the famous and influential mathematician and biostatistician active in the late 19^{th} century (Pearl and Mackenzie, 2018). Pearson, an advocate of the philosophical school of Positivism, believed that causation outside the human brain has no meaning and that reality can be described with observations, which are merely correlations. "For Pearson, especially, the slippery old concepts of cause and effect seemed outdated and unscientific, compared to the mathematically clear and precise concept of a correlation coefficient." (Pearl and Mackenzie, 2018). A mathematical formulation of causality appeared to many an unreasonable pursuit. Causation seemed to imply an element of subjectivity, while correlation was viewed as an objective property and thus more reliable. Yet, while correlation does not necessarily imply causation, it is intuitive that a correlation generally must reflect some form of direct or indirect causation. There must be a causal flow of information in a potentially larger system that ends up manifesting in observed correlations. The task of extracting such causal information from data became the pursuit of scholars of causality (Pearl, 2000; Spirtes et al.,

2000; Pearl and Mackenzie, 2018). Pioneered by several mathematicians and computer scientists between the 1980s and 1990s (e.g. Pearl, 1988; Spirtes and Glymour, 1991; Shafer, 1993), new tools and concepts were developed to talk about causality in mathematical terms, since the ones of classical statistics were not fit for purpose and could not be interpreted with causal meaning. The theory of Causality was most famously formalised by Judea Pearl into what is today an established mathematical discipline (Pearl, 2009b).

In this chapter, following the historical and mathematical journey started with Bayesian networks (Section 2.2), key concepts of causal networks will be presented (Section 2.3). Whilst providing an overview of the general theory, specific focus will be devoted to the aspects that will be used in the remainder of the thesis for both causal inference (quantifying causal effects from data given a known network, Section 2.4) and causal discovery (learning causal networks from data, Section 2.5).

2.1 Graphs and Probabilities

To start with, it is helpful to introduce some notation relating to graphs and probability theory. These will be referred back to in the remainder of this chapter and are essential to work with Bayesian and causal networks.

2.1.1 Graphs

A graph (or network) G is defined by a pair (V,E), where V is the set of nodes and E is the set of edges (or links) between the nodes in V. A directed graph is a graph with directed edges (called arcs) from one node to another, e.g. $X \to Y$. An acyclic graph contains no cycles, i.e. no sequence of nodes starting and ending at the same node by following the direction of the arcs. It follows the definition of a directed and acyclic graph (DAG). Note that in physics, computer science and statistics, the term network is often used instead of graph; and link instead of edge. In this thesis, the terms networks and links will be mostly used.

Following the intuitive kinship notation, all the nodes X_j with links pointing at node X_i are called its *parents* $PA(X_i)$, that is $PA(X_i) = \{X_j \mid X_j \to X_i\}$. Note X_i is called the *child* of its parents $PA(X_i)$. Extending this concept, the *descendants* of a node X_i are all the nodes reachable from X_i

by repeatedly following the directed links. The *predecessors* of X_i (or non-descendants) are all the variables that are not descendants of node X_i (which include its parents). These concepts can be easily extended to a time-series extension of causal networks, defined later in Section 2.3.2.

A *path* is defined as a sequence of consecutive edges of any directionality in the graph. That is, both $X \to Y \to Z$ and $X \to Y \leftarrow Z$ are paths connecting X and Z. In a causal network, a *causal path* from X to Y is a path of the form $X \to \cdots \to Y$, i.e. consisting of a sequence of (possibly empty) intermediate variables along a path where all links follow in the same forward direction. All other paths in a causal network are non-causal, and may therefore induce "spurious" associations between X and Y. All the above terms are listed in Table 2.1.

2.1.2 Probabilities

Section 2.2 will introduce the concepts of Bayesian networks, which are a mathematical and graphical representation of probabilistic systems composed of a finite set of variables and their interactions. A few basic concepts and rules of probability theory will be used, and are summarised below.

Given two random variables X and Y that take values x and y, their *joint probability* is denoted by $p(x,y)$. Their *conditional probability* is defined as

$$p(x \mid y) = \frac{p(x,y)}{p(y)}, \qquad p(y) \neq 0. \tag{2.1}$$

Two variables X and Y are *independent*, denoted by $X \perp\!\!\!\perp Y$, if and only if

$$p(x,y) = p(x)p(y), \text{ or equivalently } p(x \mid y) = p(x) \tag{2.2}$$

Two variables X and Y are *conditionally independent* given another random variable Z, denoted by $X \perp\!\!\!\perp Y \mid Z$, if and only if

$$p(x|y,z) = p(x|z), \text{ or equivalently } p(x,y|z) = p(x|z)p(y|z). \tag{2.3}$$

A property of any joint probability distribution $p(X_1,\ldots,X_N)$ is that it can be always expressed with the following factorization, called the *chain rule*

$$p(X_1,\ldots,X_N) = p(X_1)\,p(X_2|X_1)\,p(X_3|X_1,X_2)\ldots p(X_N|X_1,\ldots,X_{N-1}), \tag{2.4}$$

which holds for any ordering of the variables.

Graph (or network)	A graph G is defined by a pair (V,E), where V is the set of nodes and E is the set of edges (links) between the nodes in V.
Directed graph	A graph with directed edges, e.g. $X \rightarrow Y$.
Acyclic graph	A graph that contains no cycles, i.e. no sequence of nodes starting and ending at the same node by following the direction of the arcs.
DAG	Directed and acyclic graph.
Parents and child	Following the intuitive kinship notation, all the nodes with links pointing at X_i are called its *parents* $PA(X_i)$ of the node, $PA_i = \{X_j \mid X_j \rightarrow X_i\}$.Similarly, X_i is their *child*.
Descendants	The *descendants* of a node X_i are all the nodes reachable from X_i by repeatedly following the directed links.
Predecessors	The *predecessors* of X_i (or non-descendants) are all the variables that are not descendants of node X_i and include its parents.
Path	A *path* connecting X and Y is any sequence of consecutive edges (of any directionality) in the graph connecting the two extremes.
Causal path	In a causal network, a *causal path* from X to Y consists of a sequence of variables along a path where all links follow in the same forward direction.

Table 2.1: Graph notation.

Bayes' Rule, named after Reverend Thomas Bayes (1702-1761), follows directly from the definition of conditional probability and reads

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)} \tag{2.5}$$

where A and B are events and $P(B) \neq 0$. The rule holds for the generalised expression for continuous random variables. One of the many applications of Bayes' rule is Bayesian inference, a particular approach to statistical inference. Consider the events representing our priori belief on a given object O and evidence acquired on it E. With Bayesian probability interpretation, the theorem expresses how our degree of belief in O, expressed as a probability, should rationally change to account for the availability of related evidence E via $P(O|E) = \frac{P(E|O)P(O)}{P(E)}$.

2.2 Bayesian Networks

In 1921, geneticist Sewall Wright introduced the use of DAGs to aid the statistical analysis of the genetics of populations. Recalling the definition in Section 2.1, a DAG is a graph whose edges have a direction (i.e. directed) and that do not include closed loops (i.e. acyclic). Wright developed a statistical method called *path analysis* to quantify the effect of a gene along each path in the DAG leading to a target of interest. Wright's approach is considered as marking the beginning of probabilistic graphical models (Pearl and Russell, 2002).

Probabilistic models based on DAGs have since appeared in many fields. In cognitive science and artificial intelligence they are known as Bayesian Networks. The term Bayesian network was coined in 1985 by Judea Pearl, computer scientist and mathematician and one of their first proponents (Pearl, 1985), in honour of Rev. Thomas Bayes. His famous *Bayes' rule* for updating probabilities in light of new evidence, constitutes the foundation of the approach (Equation (2.5)).

In a nutshell, a Bayesian network represents a probabilistic system by mapping each of its N variables onto the nodes of the graph, and the dependencies among the variables onto directed links. A classical example of a BN is given by the "sprinkler example" reproduced in Figure 2.1(a). The network represents the dependencies among five discrete-valued variables: the season of the year (X_1), whether rain is falling (X_2), whether a sprinkler is on (X_3), and whether the pavement is wet (X_4) or slippery (X_5). Quantitatively, the dependencies between the N variables are the conditional probabilities of each node given all its parents, for example $P(X_4 \,|X_2, X_3)$ (which in

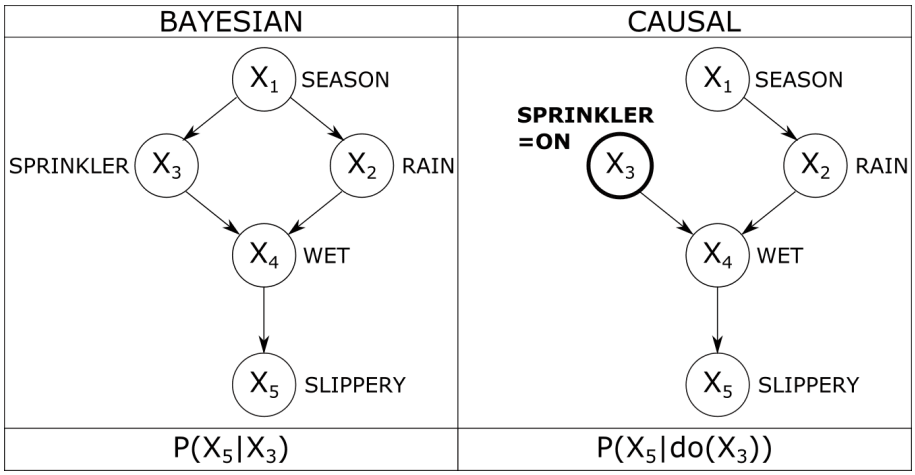

Figure 2.1: Network representing the dependencies among five variables determining if the pavement is slippery. a) Bayesian network and b) causal network under intervention on variable X_3. See text for discussion. Adapted from Pearl, 2009b, pg.15.

this case can be written as a probability table since the variables are discrete). Importantly, given these N conditional probabilities, probability rules can be used to answer any probabilistic query about any variable, e.g. what is $P(X_5 = \text{yes}\,|X_2 = \text{no})$? This is done by computing the (posterior) probabilities of any subset of variables X given evidence about any other subset in a process called probabilistic inference (or prediction given some evidence) (Pearl and Russell, 2002).

The initial development of Bayesian networks in the 1970s was sparked by the need to integrate both top-down and bottom-up combinations of empirical evidence in models of human reasoning, an emerging object of study in computational science at the time. In the late 1980s, the works of Pearl (Pearl, 1988) and Neapolitan (Neapolitan, 1990) summarized their properties establishing Bayesian networks (BNs) as a field of statistical study. The ability to perform bidirectional inferences and their rigorous (and not ad-hoc) probabilistic foundation led to the rapid establishment of BNs for modelling probabilistic reasoning in systems based on expert judgement (Spiegelhalter et al., 1993) and in Artificial Intelligence (AI) (Pearl and Russell, 2002). Today, Bayesian networks are used in a range of applications and have proved to be an incredibly efficient way to model and gain insight into the probabilistic connections between real-world variables. BNs have been widely used in neuroscience (e.g. reviewed in Bielza and Larrañaga, 2014), ecology (e.g. Marcot et al., 2006; Amstrup et al., 2008; Ramazi et al., 2020), environmental risk analysis (e.g. Marcot, 2012), energy modelling (e.g. Adedipe et al., 2020), medical meta-analyses (e.g. Greco et al., 2015; Klerk

et al., 2015; Dapeng et al., 2020) and diagnosis (e.g. Seixas et al., 2014), safety science (e.g. Zhang and Thai, 2016) and more. The wide ranging applications reflect the flexibility and practicality provided by BNs.

2.2.1 Definition of Bayesian Network

Mathematically, a Bayesian network (BN) is defined as a directed acyclic graph (DAG) with nodes associated to a set of random variables whose joint probability distribution (JPD) admits a specific factorization. Formally,

Definition 2.2.1 (Bayesian Network)**.** Let $\mathcal{G} = (V, E)$ be a DAG and let $\mathbf{X} = \{X_1, \dots X_N\}$ be set of ordered variables. Let $P(\mathbf{X})$ be the joint probability distribution on these variables. $\mathbf{X}$ is a Bayesian network with respect to $\mathcal{G}$ if the joint probability distribution $P(\mathbf{X})$ can be written as the product of the individual probability density functions, conditional on their parents as

$$P(\mathbf{X}) = \prod_{i=1}^{N} P(X_i \mid PA_i). \tag{2.6}$$

To understand the reason for this factorization, one needs to understand the mathematical definition of the parents in a BN (Pearl, 2009b).

Definition 2.2.2 (Parents in a Bayesian Network)**.** Let $\mathbf{X} = \{X_1, \dots X_N\}$ be an ordered set of variables, and let $P(\mathbf{X})$ be the joint probability distribution on these variables. A set of variables PA_i is said to be the parents of X_i if PA_i is the minimal set of predecessors of X_i that renders X_i independent of all its other predecessors. Equivalently, PA_i is the smallest subset of its predecessors $\{X_1, \dots, X_{i-1}\}$ satisfying

$$P(X_i \mid PA_i) = P(X_i \mid X_1, \dots X_{i-1}) \tag{2.7}$$

In other words, to express the conditional probability of X_i given its predecessors, only its parents are actually needed. As mentioned earlier, a crucial result following the definition above is that the JPD of a BN can be expressed as Equation (2.6). The proof follows from applying to $P(\mathbf{X})$ the chain rule in Eq. (2.4) for a suitable ordering of the variables (i.e. the topological ordering), followed by substituting the definition of parents in Eq. (2.7) to each factor of the chain rule.

A Bayesian network therefore can be seen as made of two interlinked components. The first component is called the BN structure (the DAG) and the second component is the set of BN parameters (the conditional probabilities $P(X_i \mid PA_i)$). When all the nodes are discrete variables these parameters are tabulated in what is usually referred to as a conditional probability tables (CPTs).

2.2.2 Bayesian Inference

BNs are useful for efficiently extracting information about parts of a system under changing evidence, a process called *Bayesian inference*. For instance, one could be interested in knowing the probabilistic state of variable X_i given known values e of another disjoint set of variables E called the evidence, $E = e$. Bayesian inference refers to finding the probability of variable X_i conditioned on E, i.e., $p(X_i|E)$. For example, in the case represented in Figure 2.1 (a), inference could be performed to quantify $P(X_5|X_3 = \mathrm{On})$.

All possible inference queries can be addressed by marginalization, i.e the sum over irrelevant variables, of the JPD. Given the factorization Equation (2.6), N conditional probabilities is all one needs to compute any inference query. However, even in that case, summing over the JPD takes exponential time due to its exponential size: in general BN exact inference is NP-hard (Cooper, 1990). More efficient methods have been developed, exploiting further factorizations and some applicable to special DAG structures, which cut down the computation time. For example, a first idea is to use the factored representation of the JPD for efficient marginalization. When summing (marginalizing) out irrelevant terms, the distributive law can be used to try to avoid doing actual calculation. This algorithm is called variable elimination. For a summary of other algorithms, both exact and approximated, for discrete and continuous variables we refer to the review paper by Bielza and Larrañaga, 2014.

2.3 Causal Networks

Despite the success of BNs, their theory cannot be directly used to describe causal mechanisms. The definition of a link in a BN only encodes assumptions about conditional dependence but says nothing about causality. The interpretation of a DAG as carrier of independence assumptions is in

fact valid for *any set of recursive independencies along any ordering of the variables, not necessarily causal or chronological* (Pearl, 2009b). A simple example of this fact is given by the two DAGs $X \to Y \to Z$ (also called a "chain" structure) and $Z \leftarrow Y \to X$ (a "fork" structure). Provided with the joint probability distribution $P(X,Y,Z)$, these are two fully equivalent Bayesian Networks because they represent the same conditional independence relations. In both, X and Z are dependent unconditionally ($X \not\perp\!\!\!\perp Z$), but are independent conditionally on Y ($X \perp\!\!\!\perp Z \,|\, Y$). Therefore the conditional probabilities alone cannot be used to distinguish between the two structures. However, it is clear that if we assign causal meaning to the arrows these two networks are not equivalent: in the chain, an intervention on X would affect Z, while in the fork, an intervention on X would not result in any changes to Z.

This simple example suggests that, although causality is very intuitive, its mathematical formulation requires a novel set of concepts that conditional probabilities alone cannot provide. Among the people to first recognize this fundamental limitation of BN was one of their first proponents, Judea Pearl (Pearl, 1985). In the 1990s, Pearl moved on from Bayesian networks to develop much of the language and theory for causal networks (e.g. Pearl, 1994), making fundamental progress in formalizing Causality as a mathematical discipline in its own right (Spirtes et al., 1993; Pearl, 2000). Pearl's work drew on the papers of Wermuth, 1980, Kiiveri and Speed, 1982, Wermuth and Laurizten, 1983 and Kiiveri et al., 1984, which in the early 1980s had already laid the foundations for a rigorous study of causal inference.

While its mathematical definition is non-trivial (see Section 2.3.2), a causal network is a very intuitive object to represent assumed causal relationships in a system. It has several advantages compared to other modelling choices. Firstly, thanks to its graphical representation, it makes very explicit the working assumptions of the analysts and thus facilitates others to follow (or question) the modelling choices. Secondly, no assumptions need to be made regarding the functional form of the causal relationships, nor about the distribution of variables. Thirdly, a causal network can be used to translate expert knowledge (even subjective statements) into a mathematical object with established properties and rules.The powerful interpretational advantages of causal networks, i.e. deriving causal statements from data, have made them attractive to many scientific fields. These include especially areas of research where experiments and physical interventions in the systems are impossible or unethical, and thus careful observational data analysis plays a crucial role in

advancing knowledge.

2.3.1 The concept of intervention

As seen above, despite the main intuition behind drawing links in a BN being to represent causal information, the definition of links in a BN is no guarantee for causality. The additional concept that allows to talk about causation in a network is the one of *intervention*.

Consider the example of the pressure measured by a barometer (B) and the actual pressure in a room (P). The physical (causal) law connecting the two variables can be summarised by the link $P \rightarrow B$ since the barometer measures the pressure in the room. If we were able to change the pressure in the room (intervention on P) we would see a change in the measured pressure by the barometer (change to B). However, if we were to move the needle of the barometer by hand to $B = b$ (intervention on B), no change would result in the surrounding room pressure P (no change to P). This is regardless of the fact that the observational probabilities $P(B|P)$ and $P(P|B)$ can be both computed.

Mathematically, interventions in the system are represented by the so-called do-operator, formally defined in Section 2.3.2. For intuition, the formula $do(X = x)$ represents the intervention of setting the variable X to take the value x, while keeping everything else (apart from the children of X) fixed. The causal effect of $do(X = x)$ onto another process Y is such that the distribution of Y is modified, i.e. the extent to which $P(Y \mid do(X)) \neq P(Y)$ (Pearl, 2000; Pearl et al., 2016). Furthermore, the so-called interventional conditional probability $P(Y \mid do(X))$, describing the effect of an intervention, generally does not coincide with the observational conditional probability, $P(Y \mid do(X)) \neq P(Y \mid X)$.

It is useful to note that an intervention in a causal network has also a graphical counterpart. The effect of a $do(X_i = x)$ on the network is that of a deletion of all the links from PA_i to X_i, reflecting the fact that the parents can no longer influence X_i once its value is set. An example of intervention and the resulting modified graph is shown in Figure 2.1(b) for the sprinkler example introduced in Section 2.2.1.

With this in mind, the barometer example can be revisited. The intervention $do(P)$ does not result in any change to the network because P has no parents, and the interventional probability is simply $P(B \mid do(P)) = P(B \mid P)$. However, an intervention $do(B)$ results in the removal of the link

from the node's parent, i.e. $P \not\rightarrow B$, and the interventional probability now is $P(P|do(B)) = P(P) (\neq P(P \mid B))$, because P has been disconnected from B. The latter is a correct inference of causal effect, regardless of the fact that $P(P|B)$ can be calculated. The fundamental difference between observations, $P(P|B)$, and interventions, $P(P|do(B))$, is at the heart of the difference between Bayesian and causal networks, and the additional level of information that can be represented with the latter.

2.3.2 Definition of causal network

In this section the formal definition of a causal network is provided, following mostly Chapters 2-3 in Pearl (2009b) and Pearl et al. (2016). As introduced, causality is fundamentally connected to the concept of *intervention* in a system, mathematically described by the *do*-operator.

Let us consider a system of N variables $V = \{X_1, \ldots, X_N\}$ whose causal relationships are represented by the DAG $\mathcal{G}$. The probability distribution of V after intervention on a set of variables X, $P(v \mid do(x))$, can be expressed in terms of a *sub-set* of the conditional probabilities defined on the original network $\mathcal{G}$:

$$P(v \mid do(x)) = \prod_{\{i|V_i \notin X\}} P(v_i \mid pa_i) \text{ for all } v \text{ consistent with } x. \tag{2.8}$$

Equation (2.8) reflects the removal of the term(s) $P(x_j|pa_j)$ (for all the intervened variables $X_j \in X$) from the original factorization of the BN shown in Equation (2.6), because PA_j no longer influences X_j. This is the mathematical equivalent to the graphical deletion of links to represent the effect of an intervention: $\mathcal{G} = (V,E) \rightarrow^{do(X)} \mathcal{G}' = (V,E')$ where $E' = E \setminus \{X_k \rightarrow X_j \text{ for each } X_k \in PA_j \text{ and for each } X_j \in X\}$.

Formally, a causal network is defined as follows (Pearl, 2009b, pg.23):

Definition 2.3.1 (Causal network)**.** Let $P(v)$ be a probability distribution on a set V of variables, and let $P(v \mid do(x))$ denote the distribution resulting from the intervention $do(X = x)$ that sets a subset X of variables to constants x. Denote by $\mathbf{P}_*$ the set of all interventional distributions $P(v \mid do(x))$, $X \subseteq V$. A DAG $\mathcal{G}$ is said to be a causal Bayesian network compatible with $\mathbf{P}_*$ if and only if the following three conditions hold for every $P(v \mid do(x)) \in \mathbf{P}_*$:

(i) $P(v|do(x))$ is Markov relative to $\mathcal{G}$, i.e. the probability can be factorised as in Eq. (2.6);

(ii) $P(v_i|do(x)) = 1$ for all $V_i \in X$ whenever v_i is consistent with $X = x$;

(iii) $P(v_i|pa_i, do(x)) = P(v_i|pa_i)$ for all $V_i \notin X$ whenever pa_i is consistent with $X = x$, i.e., each $P(v_i|pa_i)$ remains invariant to intervention not involving V_i.

The expression for $P(v|do(x))$ in Equation (2.8) therefore follows from the Definition 2.3.1 applied to the factorization of a JPD in a Bayesian network in Equation (2.6).

Time-series causal networks

The causal networks introduced so far do not include the dimension of time. But the variables represented by the nodes in the network could be time-dependent, X_t^i, and the causal relationships could be associated with a time lag τ, if the cause takes some time to impact the effect, $X_t^i \to X_{t+\tau}^j$. Indeed, this is how relationships in climate systems would be intuitively schematized. In the context of time evolving systems, *time-series causal networks* represent the natural extension of causal networks. A representation of time-series causal network is shown in Figure 2.2.

Definition 2.3.2 (Time-series causal network)**.** Let a multivariate time-dependent process be $\mathbf{X}_t = \{X_t^1, \ldots, X_t^N\}$ of dimension N with set of components V. The time series causal graph is given by a causal network with graph $\mathcal{G} = (V \times \mathbb{Z}, E)$ with nodes X_t^i consisting of the set of components V at each time step in $\mathbb{Z}$ and links in E being the lag-specific directed links between the variables, denoted by $X_{t-\tau}^i \to X_t^j$.

Note that the parents of a variable Y_t in a time-series causal network are $PA_{Y_t} = \{X_{t-\tau} : X \in \mathbf{X}, X_{t-\tau} \to Y_t\}$, and can include the variable Y itself at past lags if $Y_{t-\tau} \to Y_t$, called *auto links* (or auto-dependence). There is usually a maximum finite time lag considered, $\tau_{\max}$. Note that there can also be contemporaneous (but usually undirected) links called *neighbours* (not parents) of a node X_t: $N_{X_t} = \{X_t : X \in \mathbf{X}, X_t - X_t\}$.

2.3.3 Structural causal models: from probabilities to functions

In the previous sections, causal connections among variables were considered in fully general terms (e.g no functions specified) and only the conditional probabilities of each variable onto its parents were used to describe the networks. However, it is often the case that functional dependencies can be used to describe the links in a causal network and this formulation takes the name of *structural*

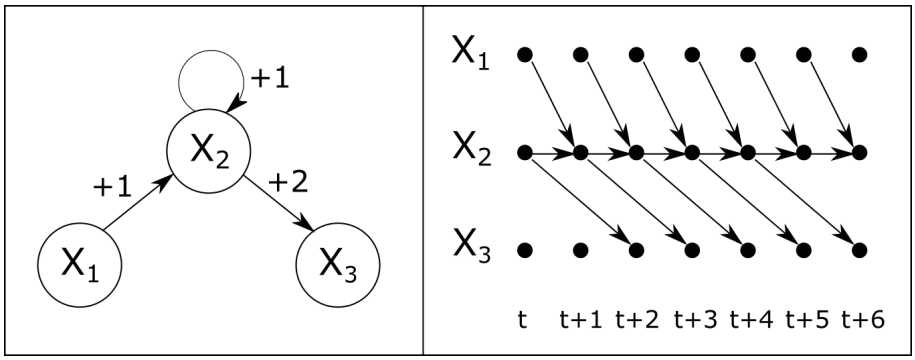

Figure 2.2: Time series causal network: static representation (left) and time dependent representation (right).

causal models (SCM). In other words, each child-parents relationship in a DAG can be represented by a deterministic function

$$X_i = f(PA_i, \varepsilon_i), \qquad i = 1, \ldots, n, \tag{2.9}$$

where PA_i are the parents of variable X_i and the ε_i are jointly independent arbitrarily distributed random disturbances due to omitted factors (Pearl, 2009b, Chapter 3.2). These disturbances represent independent background factors that are deemed not important for the analysis and can thus be considered as noise. However, if any of these factors is judged to be influencing two or more variables (thus violating the independence assumption), then they must be included as additional (possibly unmeasured) network variables. The functions f can be any, including nonlinear and nonparametric.

A crucial distinction between an SCM and a standard system of equations is in the interpretation of each functional relationship, which is not algebraic but causal. In an SCM such as Eq. (2.9) the value of the variable on the left-hand side (l.h.s) X_i is determined (caused) by the variables on the right-hand side (r.h.s.) via the function $f(PA_i, \varepsilon_i)$. This relationship cannot be read the other way round, because it represents the causal information codified in the directed links in the causal DAG which cannot be inverted or rearranged. Thus, while in general the equality symbol in mathematics is symmetric, $a = b$ implies $b = a$ and vice-versa, this is not the case in the context of a SCM. In this sense, SCM are similar to instructions in computer programming languages.

Assuming linearity, the SCM can be written as the multiple linear regression of each child onto its parents:

$$X_i = \sum_{k \neq i} \alpha_{ik} X_k + \varepsilon_i \quad \text{with } \alpha_{ik} \neq 0 \text{ iff } X_k \in PA_i, \qquad i = 1, \ldots, n, \tag{2.10}$$

A time-series causal network of process $\mathbf{X}_t$ can be associated with an SCM as follows,

$$X_t^j = g^j(PA_t^j, \eta_t^j) \quad \text{with } j = 1, \ldots, N. \tag{2.11}$$

where the sets $PA_t^j \subset (X_{t-1}, X_{t-2}, \ldots)$ define the parents of X_t^j. The noise variables η_t^j are jointly independent and, generally, are assumed to be stationary, i.e., $\eta_t^j \sim \mathcal{D}^j$ for all t for some distribution $\mathcal{D}^j$. Note that the form of the function g^j does not depend on time itself: at each time step, the functional dependence between child and parents is the same. This reflects the fact that the links E do not depend on time either. Thus the links in this time-series causal network are lag-dependent but stationary (the same at all times t), a property called *link stationarity* (Runge, 2018). Link stationarity is not a requirement of time-series causal networks, as shown and exploited in Chapter 4, but it is a common assumption. Link stationarity agrees with the idea that a true causal mechanism in a real-world physical system is usually persistent in time and not just instantaneous (to be distinguished from random noise).

In the remainder of the chapter, two main aspects of working with causal networks will be explored given their relevance to the rest of the thesis. The first is causal inference, which deals with extracting causal information from data given a network structure, and described in Section 2.4. The second is causal discovery, which deals with learning a causal network structure from data, and described in Section 2.5.

2.4 Causal Inference

Causal inference is the task of quantifying the effect of interventions in a system purely based on observed data, i.e. without doing any experiments. The underlying idea is that past and naturally occurring interventions in X that led to changes in Y are present in the data but, a priori, their estimation via simple pair-wise analysis could be biased by other processes that affect both X and Y. That is, the probability $P(Y|X)$ (or the correlation $\rho(X,Y)$) may be a biased estimate of $P(Y|do(X))$ (or the linear causal coefficient α_{YX}). To achieve a correct estimation of the causal effect, suitable controlling for confounding factors is needed.

A necessary condition for causal inference is to know the underlying causal network. This condition is crucial since no causal effect can be quantified from data alone. As seen in the pressure-

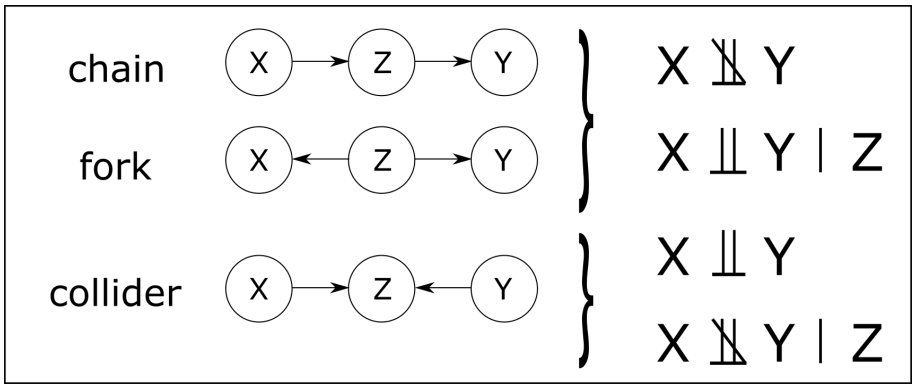

Figure 2.3: Three elemental causal structures (left) and the implied unconditional and conditional dependences (right).

barometer example, the structure of the causal network is decisive to quantify unbiased effect of the pressure on the barometer and viceversa.

Note that the causal network does not need to represent the full system of which the causal effect of interest is part of. For example, to study a climate mechanism it is not necessary (nor would it be possible) to represent the full climate system. Luckily, it can be proved that what is needed is a reduced model of the whole, tailored to the purpose of the study at hand. To study the causal effect of X on Y, only those processes that could confound the analysis, e.g., common drivers of X and Y, have to be included. Anything beyond that is irrelevant to the analysis and can thus be excluded from the causal network without loss of generality.

2.4.1 Blocked and open pathways

The most general definition of the causal effect of X on Y is simply the interventional probability $P(Y \,|\, do(X))$. Note that for a causal effect to be there it is not necessary that X is a parent of Y, but only that X can influence Y via a causal path. $P(Y \,|\, do(X))$ can be computed in various ways depending on the available data and the structure of the causal network. In an SCM, a causal effect can be represented with a single coefficient, derived from the expected value of Y after intervention $\mathbb{E}[Y|do(X)]$, as will be explored for linear SCM in Section 2.4.3.

In practice, the main work required to estimate the causal effect of X on Y is to identify the so-called *confounder* variables. These are the variables that need to be "controlled for" in the statistical estimation of the effect, e.g. included in a regression or in a conditional probability, in

order to obtain an unbiased quantification. The identification of confounders requires to understand *how information flows* along the paths of a causal network. When two variables at the extremes of a path are dependent, the path is said to be *open* for information to flow. Conversely, the path is said to be *closed* if the two variables are independent.

All the rules needed to assess if a path is *open* or *closed* are derived from the three elemental structures of which any path is made of. They are "chains" (X→Z→Y) which represent a consequential set of cause-effect, "forks" (X←Z→ Y) which represent a common cause influencing two effects, and "collider" structures (X→Z←Y) which represent two causes influencing a common effect (Figure 2.3). To complete the terminology, in a chain variable Z is also known as *mediator* between the other two, while in a fork Z is called the *common cause (or common driver)* and in the collider as the *common effect*. The following rules define the conditions under which the paths between X and Y are open or closed.

- Chains X→Z→Y: the path between X and Y is open because the two variables are dependent of each other, $X \not\perp\!\!\!\perp Y$. However, the path is blocked once we condition on Z because they become independent, $X \perp\!\!\!\perp Y|Z$. Figuratively, conditioning on Z may be seen to "block" the flow of information along the path: once Z is known, any further learning about X has no effect on the probability of Y.

- Forks X←Z→Y: follow the same rules as chains. Note that also here conditioning on Z blocks the flow of information along the path.

- Colliders X→Z← Y: the path between X and Y is closed because the two variables are independent of each other, $X \perp\!\!\!\perp Y$. However, the path is opened once we condition on the common cause, $X \not\perp\!\!\!\perp Y|Z$. Intuitively, this happens because the observation of a given outcome Z is consistently caused by some particular combination of X and Y, which thus appear correlated if stratified on Z. Failing to account for this effect can result in selection bias: stratifying a population based on a characteristic Z (e.g. reporting of illness via a phone app) may induce spurious correlation between two otherwise truly independent drivers of Z (e.g. age and infection status).

One final rule is that controlling for a descendant of a variable is equivalent to "partially" controlling for that variable, with all the consequences implied for that. As a corollary of the above rules,

any path p from X to Y (consisting of a sequence of mediators, common causes, or colliders, thus not necessarily causal) will be blocked conditionally on Z if, and only if, i) Z is a common cause or mediator along the path, or ii) if p contains a collider and Z is not that collider, nor any of its descendants.

These rules provide a criterion, called d-separation, to identify conditional dependencies between variables solely based on the structure of a causal graph (d- connotes "directional"). X and Y are d-separated by the set of variables Z if the path between them is closed by conditioning on Z.

Finally, note that the quantification of a causal effect is not always possible and the effect is called non identifiable. Causal theory provides the rules to establish this, and suggest additional observations or auxiliary experiments from which the desired inference can be obtained (Pearl, 2009b, Chapter 3).

2.4.2 Estimation of a causal effect

To quantify the unbiased causal effect of one variable X on another Y along a specific path p in a causal network, one has to "block" the effect along any other path that could bias the analysis. This means conditioning on a set Z of variables that (Pearl, 2009b, Chapter 11):

1. block all the spurious (i.e. non causal) paths from X to Y
2. block all other causal paths that are not p
3. do not open new spurious paths

In many cases, one can identify the correct adjustment set Z by looking at the graph and following the simple rules based on d-separation illustrated in the previous section. Cinelli et al. (2020) presents a set of simple but quite exhaustive network examples where "good" (and "bad") controls are identified and explained. For more complex networks, the comprehensive mathematical theory provides rules for if and how it is possible to extract a target causal effect from data (recall the concept of non identifiable effect) (Pearl, 2009b,a; Pearl et al., 2016). Furthermore, the problem of deciding which variables are "good" or "bad" controls has been automatized in a range of algorithms available via open-source software, for example *R* packages such as `pcalg` (Kalisch

et al., 2012), `dagitty`[1] (Textor et al., 2016), and `causaleffect` (Tikka and Karvanen, 2017).

Total and direct causal effect

Finally, note that if a variable X is connected to Y via both a link and a set of other (indirect) causal paths, its causal effect on Y can be computed both as its *total effect* or as its *direct effect*. The *total effect* of X on Y refers to an integration along all causal paths connecting X to Y, and is given by the formula introduced before

$$P(y|do(x)). \tag{2.12}$$

The *direct effect* of X on Y quantifies the effect only along the link $X \to Y$ only, and is given by

$$P(y|do(x), do(S_{XY})) \tag{2.13}$$

where S_{XY} are all the observed variables in the system except X and Y. The direct effect is non zero only if there is such a link. Due to the definition of parents, the direct effect can equivalently be described with $P(y|do(x), do(pa_y))$ (Pearl, 2009b, pg. 128).

In a linear system, the *indirect effect* is such that *indirect* = *total* - *direct*. Estimation of these effects in a linear system is exemplified in Section 2.4.3. For the general (nonlinear) definition, we refer to Pearl (2009b, Section 4.5).

2.4.3 Linear causal effects

In a linear SCM, the quantification of the causal effect α_p along a causal path p connecting X to Y can be simply estimated as the linear coefficient of X in a multiple linear regression, computed with ordinary least squares (OLS), which includes as regressors also all the confounders $\{Z_i\}$

$$y = \alpha_p x + \sum_i \beta_i z_i \tag{2.14}$$

where the confounders can be identified as described in Section 2.4.2. The direct effect α_d of X on Y (i.e. along the path $X \to Y$) is non-zero only if there is a link $X \to Y$ and can be quantified as

$$y = \sum_{i:pa_i} \gamma_i pa_i = \alpha_d x + \sum_{i:pa_i \neq x} \beta_i pa_i \tag{2.15}$$

[1] Also available online in `www.dagitty.net`.

which, recalling the multiple linear regression defining the linear casual model in Equation (2.10), is just the linear coefficient of the parent X on Y.

The general derivation of these formulae for the causal effects follows from the definition of the expected causal effects (total or direct) applied to linear systems and after differentiation. For example, $\frac{\partial}{\partial x}\mathbb{E}[Y|do(x)]$ for the total effect, where $\mathbb{E}(Y|do(X=x)) = \sum_y yP(Y|do(x))$. For a more formal derivation, see for example Pearl (2012a) (linear), Pearl (2012b) (nonlinear), Pearl (2009a) and Pearl (2009b)(Section 4.5).

2.5 Causal Discovery

Causal discovery is the task of reconstructing the graph of causal relationships between a set of variables from their observational data. Causal discovery has seen a steep rise with a plethora of novel approaches and methods in recent years, also thanks to the ever-growing abundance of dataset and computational resources. Each method comes with its particular set of assumptions about properties of the underlying processes and the observed data (Spirtes et al., 2000), and targets different real-world challenges (e.g. for Earth system sciences see Runge et al., 2019b). In general, causal discovery methods can be classified into classical Granger causality approaches (Granger, 1969; Barnett and Seth, 2015), constraint-based causal network learning algorithms (Spirtes et al., 2000), score-based Bayesian network learning methods (Chickering, 2002; Koller and Friedman, 2010), structural causal models (Peters et al., 2017; Porfiri and Marin, 2018), and state-space reconstruction methods (Arnhold et al., 1999; Sugihara et al., 2012). In this thesis we focus on constraint-based methods for time-series causal networks, which will be applied in Chapters 3 and 4.

2.5.1 Time-series constraint-based causal discovery

The constraint-based method for time-series causal network discovery is underpinned by the following definition of causal link (Runge et al., 2014; Runge, 2018):

Definition 2.5.1 (Lag-specific causal link)**.** Variables $X_{t-\tau}$ and Y_t of a multi-variate time-evolving process $\mathbf{X}_t$ are connected by a *lag-specific directed link* $X_{t-\tau} \to Y_t$ pointing forward in time if and

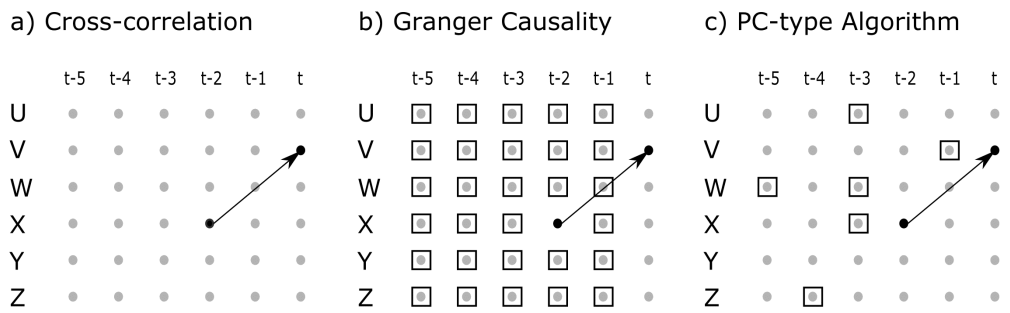

Figure 2.4: Different approaches to calculate the influence of X on V at lag 2. a) Cross-correlation calculates the correlation of X_{t-2} and V_t. b) Lag-specific Granger causality tests if X_{t-2} and V_t are conditionally independent given all other processes at all lags (black boxes). c) The adapted PC-algorithm iteratively tests for conditional independence by conditioning on different combinations of the considered processes (black boxes). It starts with conditioning on just one process and then gradually increases the number of conditions. Figure adapted from Kretschmer (2017).

only if $\tau > 0$ and

$$X_{t-\tau} \not\perp\!\!\!\perp Y_t \mid \mathbf{X}_t^- \setminus \{X_{t-\tau}\}, \tag{2.16}$$

i.e. the variables are not independent conditionally on the past of the whole process $\mathbf{X}_t^- = (\mathbf{X}_{t-1}, \mathbf{X}_{t-2} \ldots)$ excluding $X_{t-\tau}$.

In words, the link $X_{t-\tau} \to Y_t$ means that $X_{t-\tau}$ provides additional information to Y_t even once the past values of all variables are known. This definition is the lag-specific adaptation of Granger's causality (Runge et al., 2014). This concept had been introduced by Clive Granger in 1969 to investigate causal relations in econometric models (Granger, 1969). In a (multivariate) process $\mathbf{X}$, *X Granger causes Y* if X occurs before events in Y and if X improves forecasting Y, even when the past of the remaining process $\mathbf{X} \setminus \{X\}$ is known. This reasonably assumes that if Y depends on X, then its prediction will be improved by knowledge of X. Recasting the definition using conditional independence and for a specific lag τ, it can be stated as "X_t *Granger causes* $Y_{t+\tau}$ if X_t and $Y_{t+\tau}$ are conditionally dependent given the past of $\mathbf{X_{t+\tau}} \setminus \{X_t\}$". Figure 2.4(b) provides a visual representation of the Definition 2.5.1 (i.e. lag-specific Granger causality). It is contrasted with Figure 2.4(a) which represents unconditional dependence (e.g. cross-correlation).

In a constraint-based causal discovery method, links are learned from data by testing the conditional independences (CI) stated in Equation (2.16) with a smart selection of condition sets

and using a suitable CI measure I (Spirtes et al., 2000). Since computing $I(X_{t-\tau}, Y_t \mid Z)$ for the most general $Z = \mathbf{X}_t^- \backslash \{X_{t-\tau}\}$ is practically unfeasible already for relatively small systems and reduces detectability (Runge et al., 2012b), constraint-based methods are essentially strategies to select Z as a reduced set without loss of generality. A representation of the reduced sub-set of conditions is shown by the black boxes in Figure 2.4(c).

Compared to other techniques, a constraint-based method has the advantage that it can flexibly account for linear and nonlinear causal relations and different data-types (continuous and categorical, univariate and multivariate) by means of a suitable choice of the CI measure. One can draw from a large variety of conditional independence tests and measures, as discussed in detail in Runge (2018) and Runge et al. (2019a). If no hypothesis on the model can be made beforehand and data are continuous, a very general nearest-neighbour estimator can be used (Runge et al., 2012a), with the drawback of being computationally costly. If a linear model can be assumed, the conditional Pearson correlation (partial correlation) is extremely fast. It can be shown that $X \not\perp\!\!\!\perp Y \mid Z$ if the partial correlation between X and Y conditioned on Z, $\rho_{XY|Z}$, is significantly different from 0. A more detailed introduction to time-series causal discovery can be found in Runge et al. (2019a). Interestingly, because of the way links are detected in constraint-based discovery, the absent links tend to reflect more robust findings than the retained links (Runge, 2018).

2.5.2 PCMCI algorithm

In Chapters 3 and 4 of this thesis the *PCMCI algorithm* (Runge et al., 2019a) is adopted to reconstruct time-series causal networks. This method has been today widely applied in the climate science context, for example in studies by Kretschmer et al. (2016, 2017), Di Capua et al. (2019, 2020), Nowack et al. (2020), Kretschmer et al. (2020), Lehmann et al. (2020), Pfleiderer et al. (2020) and Polkova et al. (2021)

PCMCI is based on the constraint-based PC-stable algorithm, named after its inventors Peter Spirtes and Clark Glymour (Spirtes and Glymour, 1991; Spirtes et al., 1993, 2000), one of the oldest of such methods. PCMCI combines PC-stable with the momentary conditional independence (MCI) test that addresses the potential biasing effect of autocorrelation in time series. This aspect is very relevant to climate time series, where strong autocorrelation is a very common feature; for example of stratospheric dynamics and sea surface temperature. It has been shown (Runge,

2018) that the additional benefits of PCMCI compared to PC-stable (and other constraint-based methods) are of i) controlling more effectively for false positives, ii) maximising the detection power of the partial correlation, and iii) providing an estimate of link strength (in the MCI step, Runge et al. (2014)). PCMCI retains its high detection power and control for false positives also in high-dimensional and strongly autocorrelated time series settings (Runge et al., 2019a). An in depth discussion of these results can be found in Runge (2018).

The main steps performed by the PCMCI algorithm are:

1. Null hypothesis: Given a system of N variables $\mathbf{X}_t$ with $t = 1,\dots,T$, the starting (null) hypothesis is that all pairs of lagged variables $(X^i_{t-\tau}, X^j_t)$ are potentially connected $X^i_{t-\tau} \to X^j_t$ at each lag $\tau = 1,\dots,\tau_{max}$. Note that this includes $X^i_{t-\tau} \to X^i_t$, the so-called *auto-links*. Equivalently, for each X^j_t, the algorithm starts assuming all past variables (up to a user defined τ_{max}) are potential parents $\widehat{\mathcal{P}}^j_o$.

2. PC_1 step: At first the PC_1 condition pre-selection method is run to identify relevant conditions $\widehat{\mathcal{P}}^j$ for all time series variables X^j (also called potential parents). PC_1 is a Markov set discovery algorithm based on the PC-stable algorithm that removes irrelevant conditions for each of the N variables by iterative independence testing (Colombo and Maathuis, 2014). Specifically, all lagged pairs $\left(X^i_{t-\tau}, X^j_t\right)$ are tested for their independence under some low dimensional conditioning sets $\mathcal{S}$, $|\mathcal{S}| = 0, 1, .., q_{max}$ using the chosen CI test: $X^i_{t-\tau} \perp\!\!\!\perp X^j_t \mid \mathcal{S}$. q_{max} is user definable; the default value is 1. If the p-value of the CI test is larger than a user-defined significance level α_{PC}, the variable $X^i_{t-\tau}$ is removed from the hypothesized set of parents $\widehat{\mathcal{P}}^i_o$. After this pruning process, each variable X^i remains connected with a (generally) low dimensional set of potential parents $\widehat{\mathcal{P}}^i$.

3. MCI step: Then the MCI test is performed. In this step *all* pairs of variables $(X^i_{t-\tau}, X^j_t)$ with $i, j \in \{1,...,N\}$ and time-delay $0 < \tau \leq \tau_{max}$ are addressed again. The CI test is computed conditioning on the union of the newly found set of potential parents, i.e.

$$\text{MCI:} \quad X^i_{t-\tau} \perp\!\!\!\perp X^j_t \mid \widehat{\mathcal{P}}^j_t \setminus \{X^i_{t-\tau}\} \cup \widehat{\mathcal{P}}^i_{t-\tau}. \tag{2.17}$$

Thus, MCI conditions on both the (potential) parents of X^j_t and the time-shifted parents of $X^i_{t-\tau}$. Only the pairs that 'pass' the MCI test are retained as final links (i.e. the p-value

associated with the MCI test is smaller than a user-defined α significance level). The lag-specific connections that pass the MCI step constitute the discovered time-series causal network. Note that the condition set here can include the lagged-in-the-past version of both variables and thus removes biases induced by autocorrelations, together with other indirect effects (Runge, 2018).

The two stages of PCMCI serve the following purposes. PC_1 acts as a filter to remove irrelevant lagged conditions (up to some $\tau_{\max}$) for each variable. A liberal significance level α_{PC} in the tests lets PC_1 adaptively converge to typically only a few relevant conditions that include the causal parents with high probability, but might also include some false positives (usually $\alpha_{\text{PC}} = 0.2$). The MCI test then addresses false positive control for the highly-interdependent time series case, which are very common in weather and climate time series. More precisely, while the conditioning on the parents of X_t^j (the potential effect) is sufficient to establish conditional independence in the infinite sample limit (Markov property), the additional condition on the lagged parents of $X_{t-\tau}^i$ (the potential cause) leads to a test that removes biases for autocorrelated data (Runge et al., 2014).

In addition to the choice of CI test measure, the free parameters of PCMCI are three: the maximum time lag $\tau_{\max}$, and the significance levels α in MCI and α_{PC} in PC_1. The choice of CI test is a modelling assumption guided by the assumed (non)linearity of the underlying process and also finite sample size considerations. PCMCI can be used in combination with linear or nonlinear tests, and can therefore extract also nonlinear causal relationships. When assuming linear systems, partial correlations is the test of choice. α_{PC} should be regarded as a hyper-parameter and can be chosen based on model-selection criteria such as the Akaike Information Criterion (AIC) (Akaike, 1973) or cross-validation. $\tau_{\max}$ could be incorporated into this model selection. But since PCMCI is not very sensitive to this parameter (Runge et al., 2019a) (as opposed to, e.g., Granger causality), its choice can be based on lagged correlation functions, see Runge et al. (2019a) for a discussion. Finally, α is chosen based on the desired level of overall false positives.

After the discovery, the network can be further parametrized. For an assumed linear SCM, the MLR regression coefficients of each target (child) variable in Equation (2.10) provide the linear causal effect associated with each link. If the variables are standardised before the regression, the coefficients can be interpreted as the *fraction of change in the child's standard deviation* resulting from a change of one standard deviation of that parent.

Causal assumptions

A causal interpretation of the relationships estimated with PCMCI, and other discovery methods, comes from the standard assumptions in the constraint-based framework (Spirtes et al., 2000; Runge, 2018; Runge et al., 2019a), namely Causal Sufficiency, the Causal Markov Condition and Faithfulness.

- *Causal Sufficiency* is satisfied if and only if the observations $\mathbf{X}$ contain all the common parents of each pair of variables of $\mathbf{X}$. If this condition is not met (e.g. due to an unobserved confounder or also due sampling on a too coarse time interval compared to the actual causal lag), PCMCI may detect a spurious link due to the effect of a left out common influence.

- *Causal Markov Condition* states that any separation in the graph implies independence in the process (Pearl, 2009b). A potential pitfall comes with time aggregation, where a link that does not exist at the finer time scale appears in the coarser. Nonetheless, for many applications a time aggregation is usually performed specifically to remove sub-processes that are too noisy or non relevant to the analysis. A good practice is therefore to use an aggregation-step at least as small as the fastest potential causal link (Runge et al., 2014; Runge, 2018).

- *Faithfulness* is the inverse implication of the Markov condition, and can fail if there are perfectly counteracting mechanisms or some limit cases of non-pairwise dependencies (like a XOR-gate with binary variables having equal probability distributions).

For time-lagged causal discovery from observational time series, we also need the assumptions of no instantaneous effects and stationarity:

- *Absence of contemporaneous links*: today there exist causal discovery algorithms that can deal with such features (Runge, 2020) and also hidden confounders. In this thesis, lagged causal relationships are always assumed.

- *Link stationarity* is generally assumed, and reflects the concept that causal links persist in time, i.e. $X^i_{t-\tau} \to X^j_t$ *for all* t. In Chapter 4 of this thesis a modification of PCMCI is proposed that is able to account for regime-dependent non stationarity.

Another set of conditions relate not to the mathematical framework but to the implementation of causal discovery on a finite sample size of data. While it has been shown that PCMCI converges to the true graph in the limit of infinite sample size (Runge, 2014), it is also shown that causal discovery does not allow for such convergence with finite samples. This limit is quite common to any statistical technique, but has to be taken into account especially for short time series. Furthermore, the chosen CI test has to be chosen appropriately and determines the interpretation of the results. Note that this feature is also common to any statistical analysis method.

2.5.3 Interpretation of discovered causal networks

Even with a well-posed network discovery problem, the physical and/or statistical interpretation of the resulting data-driven casual network may not be straightforward. It is important to recall that the concept of *causation* in a learned constraint-based causal network context is the abbreviation for *pairwise conditional dependence under a set of conditioning variables*. Two specific applications of causal network, search of predictors and hypothesis testing, provide contexts where the causal network has a more intuitive interpretation.

As for the predictive interpretation, it can be mathematically shown that the parents of each variable in a causal network learned with PCMCI are its minimal and optimal[1] set of predictors (Runge, 2014). And that the network is robust in this regard even if extracted from partial and discrete-time observations of a larger and continuous Universe of events. In the context of optimal model set up for seasonal forecasting, the best predictors is what scientists are often after. However, if the goal is understanding the source of model error, than physical causality is key to the answer and the former reasoning may not be satisfactory.

Hypothesis testing involving multiple processes is a challenging and relevant practice in climate science. In this context, causal discovery is highly informative since the perimeter of the comparison is well defined. A good example is the one in Kretschmer et al. (2016), where the authors reconstruct a causal network using climatological reanalysis time series to test two proposed Arctic mechanisms driving observed circulation changes.

[1] Here optimal is defined as maximising a theoretical information measure.

2.6 Can we prove causality using observational data alone?

Causal models can be thought of as theories that describe the underlying mechanisms responsible for the generation of the data observed. Just as physical theories cannot be proved by observations, also causal models cannot ultimately be proved from data alone, be it causal discovery or inference. However, as with theories of physics, causal networks can be falsified if their underlying assumptions are not supported by the observational data. Therefore, data can be used to falsify a causal network, by identifying locally which parts of the network are in contradiction with observations, and thus need reformulation.

Applying this reasoning on falsification to causal discovery, a causal network discovered from the statistical dependencies present in the data is by definition consistent with this data. Such a network can be tested against unseen data to stress-test the links (and non-links).

On the other hand, in causal inference the causal network used is often prescribed, based on expert knowledge of the system. When data are used to quantify the causal effects, a preliminary and crucial step consists in testing that any such independence $X \perp\!\!\!\perp Y \mid Z$ implied by the network is in agreement with the data. These CI are also known as *testable implications* (or d-separations, see Section 2.4.1). This is just the falsification step mentioned above. For a linear causal model, where CI can be assessed via partial correlation, an example of a sufficient set of tests is $\{\rho_{x_i x_j | pa_i} = 0 \mid i > j\}$ where i ranges over all nodes and j ranges over all predecessors of i. Since not all the testable implications are independent from each other, their smallest relevant sub-set is sufficient. There are software packages that provide all the testable implications based on a given network, including the free online tool daggity[1]. It is worth stressing once more that this step is a "sanity check" rather than a proof of causality. Equally, the assessment of skill for a causal model (e.g. represented by an SCM) using statistical methods, such as accuracy of prediction, cannot prove the validity of the causal model.

[1] http://www.dagitty.net/dags.html [last accessed 25/09/2022]

Part II

Results

Chapter 3

Quantifying the timescale and strength of Southern Hemisphere intra-seasonal stratosphere-troposphere coupling

3.1 Introduction

As introduced in Section 1.3, during the spring-to-summer months the Southern Hemisphere (SH) stratospheric and tropospheric extratropical circulation variabilities are strongly coupled (S-T coupling). One of the most compelling observational fingerprint of the SH S-T coupling on the intra-seasonal timescales is a downward propagating structure of the composite geopotential heigh anomalies (Kuroda and Kodera, 2001; Thompson et al., 2005), seen also in the zonal winds (Kuroda and Kodera, 1998), around the time of the seasonal vortex breakdown event. These composites are commonly called "dripping paint plots" because of downward reaching "filaments" of the coherent signal of the anomalies which seem to drip down from the stratosphere.

Modelling studies of the SH have confirmed the stratospheric origin of this downward propagation of anomalies. Arguably the strongest evidence is provided by studies of the late 20th century ozone depletion. Multi-decadal trends of the SH tropospheric circulation, in particular of the eddy-driven jet (jet), have been unequivocally traced back to the stratospheric ozone decline. The effect of the ozone trends on the troposphere are mediated by a delay in the vortex breakdown

(McLandress et al., 2010, 2011). Such a tropospheric response to an exogenous stratospheric perturbation can be regarded as causal in the physical sense of the term.

3.1.1 Statistical quantification of the S-T coupling

Observational studies have shown that the SH stratospheric vortex breakdown (VB) plays an organising role for tropospheric circulation anomalies on weekly to monthly timescales (Hio and Yoden, 2005; Black and McDaniel, 2007; Byrne et al., 2017). This is largely due to the VB event corresponding to a shift down of the zonal wind anomalies which eventually leads to the early-summertime equatorward shift of the tropospheric jet (Kidston et al., 2015; Byrne et al., 2017; Byrne and Shepherd, 2018). Estimates of such potential for predictability of the troposphere due to the vortex, in terms of strength and timescale of the signal, have been quantified with lagged correlations (Graversen and Christiansen, 2003; Gerber et al., 2010; Byrne and Shepherd, 2018) or regression patterns (Thompson et al., 2005) (details in Appendix 3.B). These two methods are widely employed to identify and characterize statistical predictors (Wang et al., 2017). However, it is becoming increasingly understood that their results can be biased by the effect of autocorrelations (McGraw and Barnes, 2018). Further, these are not designed to distinguish direct from indirect effects (Ebert-Uphoff and Deng, 2012). More complex machine learning (ML) techniques have recently been adopted (Minokhin et al., 2017) to overcome those shortcomings, but such methods often lack physical interpretation (Runge et al., 2019b).

3.1.2 SAM autocorrelations

In these same spring-to-summer months, considerable interest has been devoted to the enhanced autocorrelation timescale of the tropospheric Southern Annular Mode (SAM), which is closely related to the dynamics of the jet. Most studies have interpreted the observed enhanced SAM autocorrelation timescale in terms of purely tropospheric eddy-feedback processes (Lorenz and Hartmann (2001), Barnes and Hartmann (2010) and Simpson et al. (2011, 2013), as discussed in further details in Appendix 3.A). Evidence for this was the observed non-zero correlation between the eddy momentum flux anomaly (m) and SAM index with SAM leading, interpreted as an eddy feedback. Interestingly, Simpson et al. (2011) also showed that suppressing stratospheric springtime variability in spring in a climate model reduces the autocorrelation timescale of the tro-

pospheric SAM by one-half (to about two weeks). No further exploration on this interesting result of the role of the stratosphere followed there, since the paper was also focusing on tropospheric feedbacks.

Several authors however have suggested that the longer tropospheric autocorrelation timescale reflects the stratospheric influence at this time of year (Baldwin et al., 2003; Graversen and Christiansen, 2003; Thompson et al., 2005; Gerber et al., 2010). For example, in an extensive statistical analysis of chemistry-climate models and reanalysis, Gerber et al. (2010) quantified SAM's variability, its decorrelation time scale, and its lagged correlation at various pressure levels, all showing a coherent vertical structure that seems to origin in the stratosphere. More recently, using a more mechanistic approach, Byrne et al. (2016) employed the zonally and vertically integrated quasi-geostrophic momentum equation to show that a synthetically generated external forcing (F) acting on the eddy momentum flux anomalies (m) can produce similar lagged cross-correlation between m and the SAM index to what is observed in austral spring, without invoking tropospheric SAM feedbacks. The authors argued that the stratosphere could play the role of the external forcing F using physical arguments of the S-T coupling. A theoretical model able to quantify the effect of the S-T downward connection on the SAM from reanalysis has however been lacking, arguably due to the difficulty of statistically isolating the downward effect from the internal (and independent) stratospheric and tropospheric dynamics.

In this chapter, a theoretically-based quantification is proposed for the intra-seasonal connection between the stratospheric polar vortex and tropospheric eddy-driven jet during the spring-to-summer transition, with a focus on the timescale and strength of the downward influence. Specifically, following Granger (1969) the problem is framed in terms of how much additional and unique information the stratosphere provides to the knowledge of the troposphere. The framework chosen for such analysis is that of time-series causal networks (introduced in Section 2.3.2) because they allow to distinguish direct from indirect effects by systematically accounting for multiple, interconnected and autocorrelated variables.

3.2 Data

The primary data used is the four-times-daily wind from the ERA-Interim reanalysis, available on an N128 Gaussian grid and 37 pressure levels (1000-1 hPa) (Dee et al., 2011). The period

considered is 1 June 1979 to 30 April 2018 (39 spring-to-summer transitions). The vortex strength (PoV) is defined using the daily-mean zonal-mean zonal wind ($[u]$) at 50 hPa and 60°S (Black and McDaniel, 2007; Byrne et al., 2017; Ceppi and Shepherd, 2019). This choice is considered to be robust due to the strong barotropic structure and temporal coherence of the vortex intra-seasonal variability. The eddy-driven jet strength at high latitudes (Jet) is defined as the sine-weighted integral of $[u]$ at 850 hPa between 50 and 65°S. By geostrophic balance, Jet is proportional to the difference in zonal-mean sea-level pressure between 50 and 65°S, which was used in Byrne and Shepherd (2018) as a measure of the Jet and thus it can be compared with their results. Anomalies are computed by removing the climatology and are used in all analysis. The n-days average of a daily time series is performed on non-overlapping blocks of length n. The season 2002-2003 is discarded due to its outlier nature, being the only SH stratospheric sudden warming on record. The date of the vortex breakdown (VB) is identified as the final time that the PoV drops below 10 ms^{-1}; this criterion is applied to running 5-day averages with the central day as reference (Black and McDaniel, 2007; Byrne et al., 2017). Only the first part of our two time series present a significant trend (see Figure 3.6), thus detrending is uncertain. Since the vortex trend is mostly due to a delayed VB in the earlier part of the record, conditioning the time-window of the analysis around that date acts effectively as a detrending. With this VB dependent analysis the results on the timescales of interest are found to be almost identical, as seen in Figure 3.2 (c,d), thus avoiding explicit detrending does not impact the conclusions.

3.3 Methods

The set of relationships explaining the behaviour of a system $\mathbf{X}_t$ can be represented with a time-series causal network, as formally introduced in Section 2.3.2. For linear systems, the definition of lag-specific causal link of Equation (2.16) translates into the partial correlation coefficient

$$\rho(X_t, Y_{t+\tau} | S_{t'<t+\tau}) \tag{3.1}$$

being significantly different from zero, where $S_{t'<t+\tau} = \{\mathbf{X}_{t+\tau-1}, \mathbf{X}_{t+\tau-2}, \dots \mathbf{X}_{t+\tau-\tau_{max}}\} \setminus \{X_t\}$ (up to a maximum lag in the past τ_{max}) (Spirtes and Glymour, 1991). Partial correlation is the correlation between the residuals of two separate linear regressions of X and Y onto S. Without loss of generality, causal network theory allows to replace the high-dimensional set of conditions $S_{t'<t+\tau}$

with a suitable subset based on the parents of X_t and $Y_{t+\tau}$ (Runge et al., 2012a). The algorithm here used to find these set is a time-series causal discovery algorithm called PCMCI (Runge, 2018), a constraint-based casual discovery method, described in Section 2.5.2.

The CI test chosen to run PCMCI in this analysis is a partial correlation, since the interest is in the linear dependence between the variables. Dependence (i.e. null hypothesis link) is rejected if the computed p-value associated with the test is larger than the desired significance level α. Here the p-value is computed according to a two-tailed Student's t-distribution (PCMCI assumes two-sided distribution; this is used everywhere in this analysis except for Figure 3.1 where is one-tailed to match the choice in Byrne and Shepherd (2018) and facilitate comparison). The maximum lag tested for connection, τ_{max}, and the significance level for link rejection, α, are the two free parameters in PCMCI, and their choices are specified in each part of the present analysis. All analysis with PCMCI was performed with the TiGraMITe Python package.

An important quantity that can be derived for each discovered link $X_t \to Y_{t+\tau}$ is the partial correlation

$$\rho^{MIT}(X_t \to Y_{t+\tau}) = \rho(X_t, Y_{t+\tau} | Z^{MIT}_{X_t, Y_{t+\tau}}) \tag{3.2}$$

where $Z^{MIT}_{X_t,Y_{t+\tau}} = \mathcal{P}_{X_t} \cup \mathcal{P}_{Y_{t+\tau}} \setminus \{X_t\}$ and $\mathcal{P}_{X_t}$ are the discovered parents of variable X_t. MIT stands for Momentary Information Transfer and refers to the property of the measure of being lag-specific (or momentary). In particular, Z^{MIT} can include the lagged-in-the-past version of both variables and thus removes biases induced by autocorrelations, together with other indirect effects (Runge, 2018). It can be proven that ρ^{MIT} can be used to reliably quantify the strength of the links (Runge et al., 2012a, pp.187 Runge, 2014).

An important assumption for unbiased causal discovery is to have included all potential common drivers of vortex and jet (Causal Sufficiency, pp.50). An obvious here is El Niño Southern Oscillation (ENSO) (e.g. Byrne et al., 2019). However, since the expected time scale of influence of ENSO should be much longer (1-3 months) than the ones used for the analysis (5-40 days) it can be instead considered as a background effect that should not influence the intra-seasonal dynamics.

Finally, recall that the term causal link needs to be understood conditionally in the context of the variables analysed. The inevitable initial choice of variables relying on expert (i.e. physical) knowledge sets the boundary conditions for the interpretation of the analysis. It is likely that intermediate and (or) additional variables may be needed to explain the full causal mechanism at

play in the S-T downward coupling, and including additional variables could lead to a different set of linkages. However developing a full theory of the coupling goes beyond the goal of this chapter, which is to quantify the information flow between the stratosphere and the troposphere.

3.4 Results

3.4.1 The Confounding Effect of Stratospheric Autocorrelations

The S-T coupling is initially inspected by comparing monthly lagged correlations (ρ, as in Byrne and Shepherd, 2018) with the corresponding MIT partial lagged correlations (ρ^{MIT}), shown in Figure 3.1. The statistical coefficients are computed on monthly mean anomalies of a variable at a given month (m) and another variable at a later month ($m+\tau$), lagged by τ from +1 to +3 months and with m from September to January. Variables are correlated both with themselves (auto-correlation) and with the other (cross-correlation). The conditions used for ρ^{MIT} are extracted via PCMCI, with parameters $\tau_{max} = +3$, $\alpha = 0.05$ and a p-value computed from a one-sided Student's t-distribution with 38 samples (the number of seasons analysed excluding 2002-2003).

Figure 3.1(a) for ρ shows strong autocorrelation of the vortex, as well as a large downward cross-correlation and a relatively strong upward cross-correlation. Despite the slightly different proxies used for jet and vortex, these plots resemble closely those of Figures 7 and 14(a,b) in Byrne and Shepherd (2018). Having observed that the cross-correlation of vortex leading jet is significant and larger than the jet autocorrelation for the same months[1], Byrne and Shepherd, 2018 argued that this suggests that the vortex is better predictor of the jet than the jet itself between October and December and that this effect lasts for up to 3 months. This was also previously suggested by Gerber et al. (2010). However such a statement fails to account for the role of the strong vortex autocorrelation which acts to inflate the cross-correlations, already noticeable in the similarity between the plots for vortex autocorrelation and downward correlation. Furthermore, following a similar reasoning, the upward correlation $\rho(Jet_m, PoV_{m+\tau})$ could suggest there is an upward component to the S-T coupling too, since this is also statistically significant at lag +1 month from October to January (not computed in the above mentioned article).

A conditional analysis with ρ^{MIT} results in a different quantification of the coupling, as shown

[1] i.e. $\rho(PoV_m, Jet_{m+\tau}) > \rho(Jet_m, Jet_{m+\tau}) > 0$

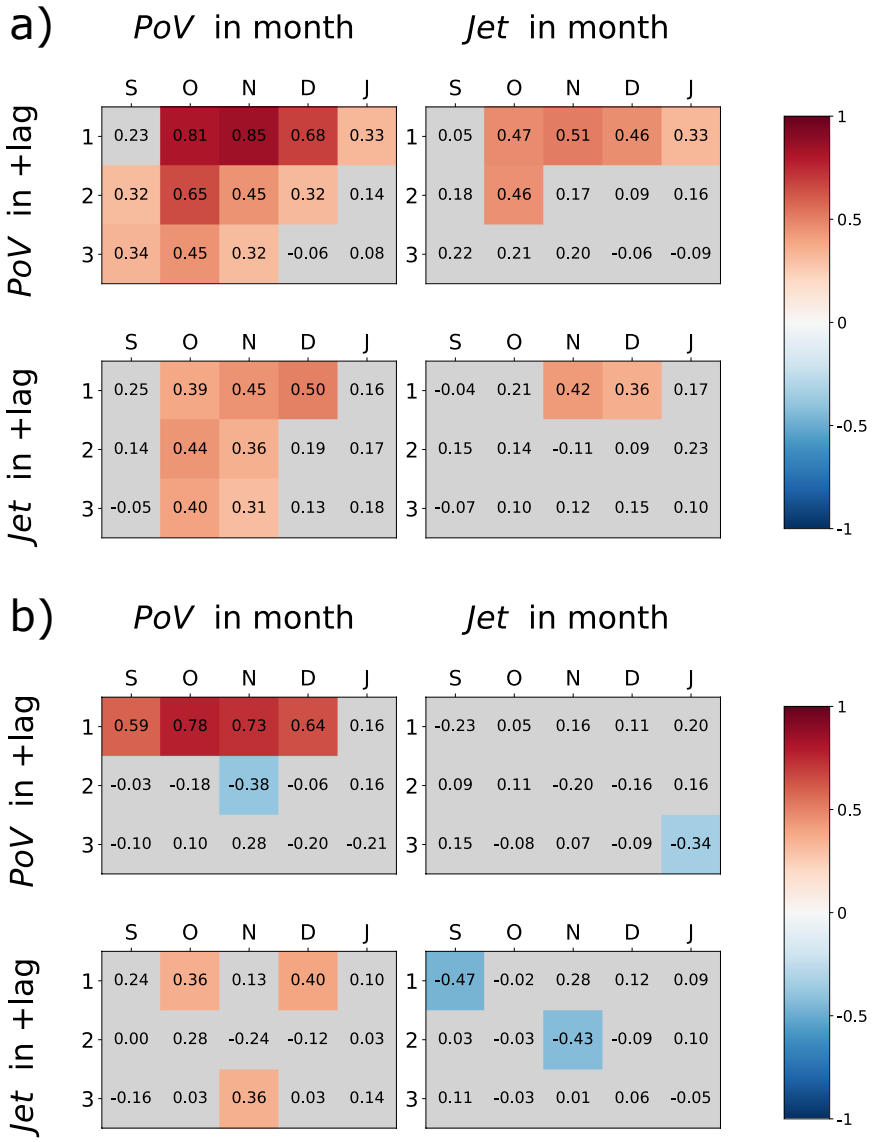

Figure 3.1: The correlation ρ (a) and partial correlation ρ^{MIT} (b) matrices for PoV and Jet. In each subfigure, left panels shows coefficients between monthly values of PoV and either (top) PoV or (bottom) Jet lagged by +1 to +3 months. Right panels show the same but for monthly values of Jet. If the p-value is larger than 0.05 in a one-tailed Student's t-distribution of 38 samples the colour of the entry is grey, otherwise the colour matches the correlation coefficient.

by Figure 3.1(b). Between September and December the vortex is found to provide significant additional information to its future state for a maximum of one month lag, compared with a lag of up to 3 months for the unconditional analysis. More strikingly, only a few positive downward links are found in the MIT panel, compared to the extended downward unconditional cross-correlation.Thus a confident conclusion on the vortex being a tropospheric predictor seems less obvious. A similar result holds for the MIT upward connection too, where the cross-correlation structure almost totally disappears after conditioning.

Why is the MIT downward coupling strength so much weaker than the unconditional estimate? The answer lies in the conditioning set used for the partial correlation. Inspecting all the pairs of variables and months, the conditioning set detected by PCMCI was always constituted by the one month-lagged vortex $Z^{MIT} = PoV_{m-1}$. This is due to the vortex being strongly auto-dependent, $PoV_{m-1} \rightarrow PoV_m$, and thus retained in the Z^{MIT} conditions for $\rho^{MIT}(PoV_m, Jet_{m+\tau})$. Given the extent of this weakening, it is evident that the stratospheric auto-dependence accounts for most of the signal of the downward correlation, including its persistence for multiple months. The vortex auto-dependence is found to also be causing the difference for the upward coupling, where the conditioning set $Z^{MIT} = PoV_{m+\tau-1}$ is also responsible for the disappearing cross-correlation with Jet leading PoV.

The mathematical explanation of the inflating effect of a variable's auto-dependence (here the vortex) in a cross-correlation is reported in Appendix 3.C, based on the analysis in Runge et al. (2014).

3.4.2 Parametrization of the stratosphere-troposphere coupling

The patchy structure of the downward partial correlations suggests that sample size issues might be at play and potentially hiding the coupling signal. Since modelling studies agree that the dynamical downward coupling persists from roughly September to January (e.g. Byrne et al., 2019), groups of months are analysed together rather than individually to increase the sample size.

Detecting timescale and time of the year

First, the timescale at which the upward and downward causal links come into play are explored. While a time-window spanning the months from October to January (ONDJ) for each season is

fixed, the time averaging applied to the time series is varied over n=5, ..., 55 days. The corresponding sample sizes are from 912 to 80 (compared to 38 for the previous analysis). Note that an alternative time-window SOND gives similar results (not shown). PCMCI parameters are set at $\alpha = 0.05$ and $\tau_{max} = +3$.

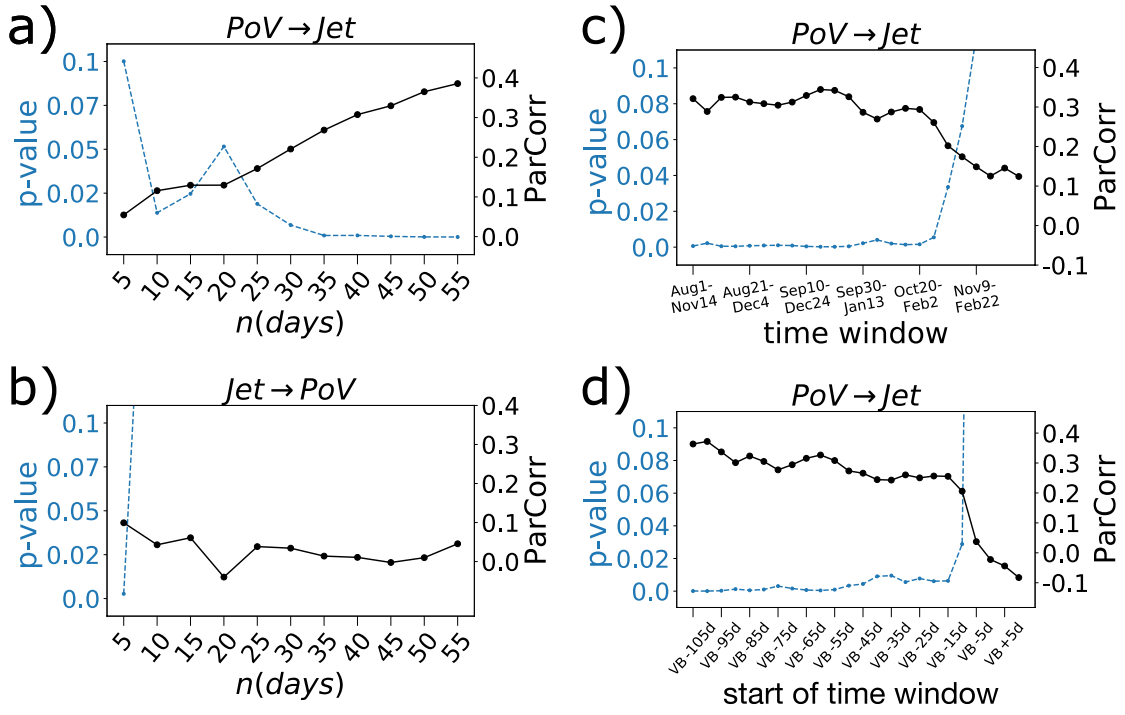

Figure 3.2: ρ^{MIT} (black) and p-value (blue) for increasing n-averaging and fixed ONDJ time-window: (a) downward link and (b) upward link. ρ^{MIT} (black) and p-value (blue) for the downward link at fixed 35 days aggregation and changing time-window: (c) a fixed period (see x-ticks) is analysed every year; (d) the period starting some days (x-ticks) before the yearly date of the vortex breakdown is analysed. The maximum lag for all the analysis is τ_{max}=3 and $\alpha = 0.05$. Also the value of the rejected links are shown to inspect the change in p-value and ρ^{MIT}.

Firstly, the downward and upward links at lag +1 are inspected. Figure 3.2(a) shows that the downward link at lag +1 becomes significant (and strong) for averaging larger than 30 days (p-val $\leq 10^{-3}$, $\rho^{MIT} \geq 0.25$). The known role of the stratosphere as a source of tropospheric predictability is recovered but the influence needs around one month to be manifest in the troposphere, which suggests a role for lower-stratospheric radiative processes. This timescale is very close to the one used in Figure 3.1, therefore the non-robust downward link in that analysis was most likely

a sample size issue (there samples were 38, here 152). Figure 3.2(b) shows that the upward link is detected only for time averaging shorter than 10 days, which agrees with planetary-wave propagation timescales.

In terms of auto-dependence at lag +1, the jet is significantly auto-dependent only up to a time averaging of 20 days (not shown). Note that in our framework any tropospheric feedback not forced by the stratosphere would show up in the jet auto-dependence, as no mediating eddy momentum flux variable is included. The vortex auto-dependence at lag +1 is significant up to 50 days (not shown). Finally, note that links at larger lags appear only for vortex auto-dependency (+2) and are non-robust to different averaging (not shown).

Overall, Figure 3.2(a,b) suggests that a reasonable intra-seasonal timescale to study the S-T coupling is around 35-40 days, and hereafter 35 days is considered. This choice that strikes a balance between signal detectability of the downward influence, which from physical considerations requires a large enough time lag (confirmed by this analysis, Figure 3.2(a)), and a large enough sample size for estimating partial correlations within the reanalysis dataset. The smooth increase of values in Figure 3.2(a) of $PoV_t \rightarrow Jet_{t+1}$ is expected due to the vortex auto-dependence, which happens at many smaller timescales and thus is baked into the downward link as soon as the averaging period becomes larger than the timescale of the direct connection. This consideration adds an independent motivation to choosing the smallest possible time average at which the coupling is manifest.

The persistence of the downward link from August to February is also studied (Figure 3.2(c)). The number of days T of the time-window is fixed (105 days, three times the n=35 days averaging) and a changing starting day t_0 is considered by shifting it every 5 days from 1 August to 1 December. Note that the 35 days average is performed after the time-window is selected. The downward link is found to persist from August to January but loses significance when the final part of the time-window enters into February. This confirms the well known feature that the vortex decouples again from the jet in the summer.

Finally, the effect of the variability in the vortex breakdown date is addressed with an alternative time-window selection (Figure 3.2(d)). In this case t_0 is chosen each year according to the VB date as $t_0 = VB - t$ days, therefore with t_0 iteratively increased by 5 days until it reaches $t_0 = VB + 10$ days . The breakdown-specific downward link is strong and robust when the time-window begins

well before the breakdown, and has a rapid drop in significance when t_0 is closer to VB than about 20 days. This provides some evidence that the breakdown date is an indicator of the culmination of the coupling period. Since PCMCI is here forced to constrain the driver variable (not the effect) in the time window specified above (a feature called x masking in PCMCI), this result implies that only the stratospheric dynamics *up to the vortex breakdown date* influences the tropospheric jet. After the breakdown, the stratospheric information does not transfer information downward as effectively (or at all).

Quantification of the linear S-T coupling strength

According to the above analysis, the best period to evaluate the intra-seasonal S-T downward coupling is between the beginning of September and mid-January (or alternatively well before the VB) and with a time averaging of about 35 days. Note that the sample size achieved for n=35 is of 152 = 4(samples/year)*38(years). With this choice of time parameters, PCMCI is re-applied testing all possible links with $\tau_{max} = +2$ and at a significance level $\alpha = 0.005$, thus strongly controlling for false positives.

The results from the MCI conditional dependence tests are shown in Table 3.1. The connections detected are two: the downward link and the vortex auto-dependence. This is shown graphically in Figure 3.3 and can be written as the following linear system:

$$\begin{cases} \overline{PoV}_t & = a\,\overline{PoV}_{t-1} + \varepsilon_{P,t} \\ \overline{Jet}_t & = c\,\overline{PoV}_{t-1} + \varepsilon_{J,t} \end{cases} \tag{3.3}$$

where $(\varepsilon_P, \varepsilon_J)_t$ is an independent and identically distributed Gaussian noise of covariance Σ and zero mean. The over-bar stands for standardised time series and lag +1 corresponds to 35 days. The linear regression coefficients of the standardized time series ($\pm$ their standard deviations) and the covariance matrix fitted to the two-dimensional residuals are:

$$a = 0.55 \pm 0.11, \qquad c = 0.37 \pm 0.15, \qquad \Sigma = \begin{pmatrix} 0.70 & 0.35 \\ 0.35 & 0.86 \end{pmatrix} \tag{3.4}$$

Note that Equation (3.3) recovers an interaction supported by previous literature but offers a more parsimonious description than the full set of cross-correlation parameters. The vortex explains around 40% of the monthly jet variability and can be used as its statistical predictor around 35 days

ahead. On this timescale, the jet does not provide a detectable source of information to itself nor to the vortex.

Link	lag	ρ^{MIT}	95% conf.	p-value
PoV → PoV	**+1**	**0.56**	**(0.472, 0.628)**	**1.5**$\cdot 10^{-5}$
PoV → Jet	**+1**	**0.29**	**(0.161, 0.399)**	**1.0**$\cdot 10^{-3}$
Jet → Jet	+1	0.063	(-0.072, 0.196)	0.44
Jet → PoV	+1	-0.064	(-0.197, 0.071)	0.43
PoV → PoV	+2	-0.14	(-0.272, -0.009)	0.07
PoV → Jet	+2	0.06	(-0.069, 0.200)	0.41
Jet → Jet	+2	0.06	(-0.071, 0.197)	0.43
Jet → PoV	+2	0.14	(0.008, 0.271)	0.08

Table 3.1: PCMCI results for $n = 35$ days and time window = Sept 1 to Jan 19. PCMCI parameters used are $\alpha = 0.005$, $\tau_{max} = +2$. The statistically significant links are indicated in bold font.

Reproducing the observed lagged correlation pattern

The correlation structure of reanalysis is compared with the one of data synthetically generated via Eq. (3.3), with coefficient as in Eq. (3.4). Since the causal drivers and their coefficients have not been chosen to optimize the resemblance of these features, this comparison provides an independent check on our result. Figure 3.4 shows that the synthetic (light blue) and reanalysis

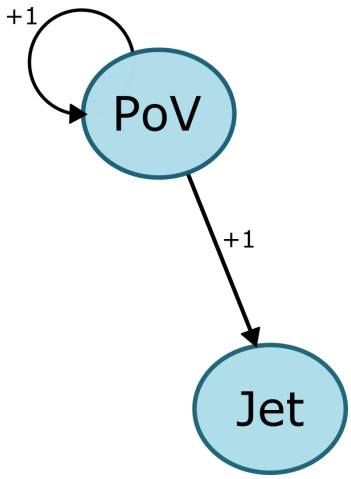

Figure 3.3: Graphical representation of the network described by Equation (3.3). Lag +1 corresponds to 35 days.

(red) correlations are in good agreement, meaning that this minimal model can recover the full reanalysis cross-correlation fingerprint. This suggests that this set of variables is capturing the first-order relationship between the observed vortex and the jet at this timescale, i.e. there is no major missing explanatory factor.

It is crucial to stress that the jet autocorrelation is also matching reanalysis (panel d), despite the absence of direct tropospheric auto-dependence or feedback (no $Jet_t \rightarrow Jet_{t+\tau}$). Here the jet "inherits" the vortex long autocorrelation timescale via the downward link. This can be also formally shown by the analytical formulae for the autocorrelations derived from Eq. (3.3):

$$\rho_{\overline{PoV}}(\tau) = a^{|\tau|} \tag{3.5}$$

$$\rho_{\overline{Jet}}(\tau) = \frac{1}{c^2\sigma_P^2 + \sigma_J^2(1-a^2)}\left[c^2\sigma_P^2\rho_{\overline{PoV}}(\tau) + \delta_{\tau,0}\sigma_J^2(1-a^2) + (1-\delta_{\tau,0})c(1-a^2)\sigma_{PJ}\,a^{|\tau|-1}\right] \tag{3.6}$$

where $\Sigma = \begin{pmatrix} \sigma_P^2 & \sigma_{PJ} \\ \sigma_{PJ} & \sigma_J^2 \end{pmatrix}$. Note that the jet autocorrelation function comprises of three terms. The first term is proportional to the vortex autocorrelation ($a^{|\tau|}$), thus it inherits the vortex autocorrelation timescale thanks to $c \neq 0$. The second term ensures $\rho_{\overline{Jet}}(0) = 1$. The third term includes the effects of a contemporaneous connection between the two time series ($\sigma_{JP} \neq 0$; physically this can represent processes influencing both vortex and jet on a timescale smaller than $\tau = 1$). and its decorrelation time scale is set by $a^{|\tau|-1}$, thus is also dependent on the vortex auto-dependence strength. The correlation structure of the time series generated using the standardised coefficient

can be directly compared with the reanalysis correlations because of the property of invariance under linear transformations.

Based on this analysis, the enhanced autocorrelation of the monthly mean tropospheric SAM observed during this period can be attributed in large part to the intra-seasonal stratospheric influence. The effect of a tropospheric eddy feedback could increase the jet autocorrelation timescale further, but is not needed to explain the observed values at the monthly timescale. This conclusion is consistent with the analysis by Breul et al. (2022) who found with a regression analysis that VB variability approximately doubles SAM autocorrelation timescale during austral summer (DJF), a result that holds in both re-analysis data and in models of CMIP5 and CMIP6 (although note that here SOND are considered). The causal role of the stratosphere was backed by simulations with a simple stochastically forced barotropic model, where the influence of the polar vortex on the jet was parameterised with a torque (Breul et al., 2022, for details).

The role of the stratosphere in inflating the jet autocorrelation may help account for the lack of any relationship between the jet response to climate change and the jet autocorrelation timescale in austral summer (Simpson and Polvani, 2016; Breul et al., 2022), which might otherwise be expected if such a timescale reflected mostly the strength of tropospheric feedback processes through Fluctuation-Dissipation relationship arguments (Gritsun and Branstator, 2007). See also Breul et al. (2022) for an updated discussion on this aspect.

3.4.3 Connection Between Ozone-induced Observed Circulation Trends

The strength of the late spring SH polar vortex is known to have increased in the late 20$^{\text{th}}$ century, mainly as a result of the cooling of the polar stratosphere due to anthropogenic ozone depletion (Waugh et al., 1999; McLandress et al., 2010). It is also established that the concurrent positive trend in the SAM seen in austral summer is a response to the same cause: a poleward trend of the tropospheric jet (McLandress et al., 2011) corresponding to an overall delayed seasonal equatorward shift (Sun et al., 2014; Byrne et al., 2017) due to an on average later VB caused by the colder and stronger stratosphere.

In an attempt to reproduce this dynamics, an exogenous forcing term, $\tilde{\gamma}t$, is added to $\overline{PoV}$ in

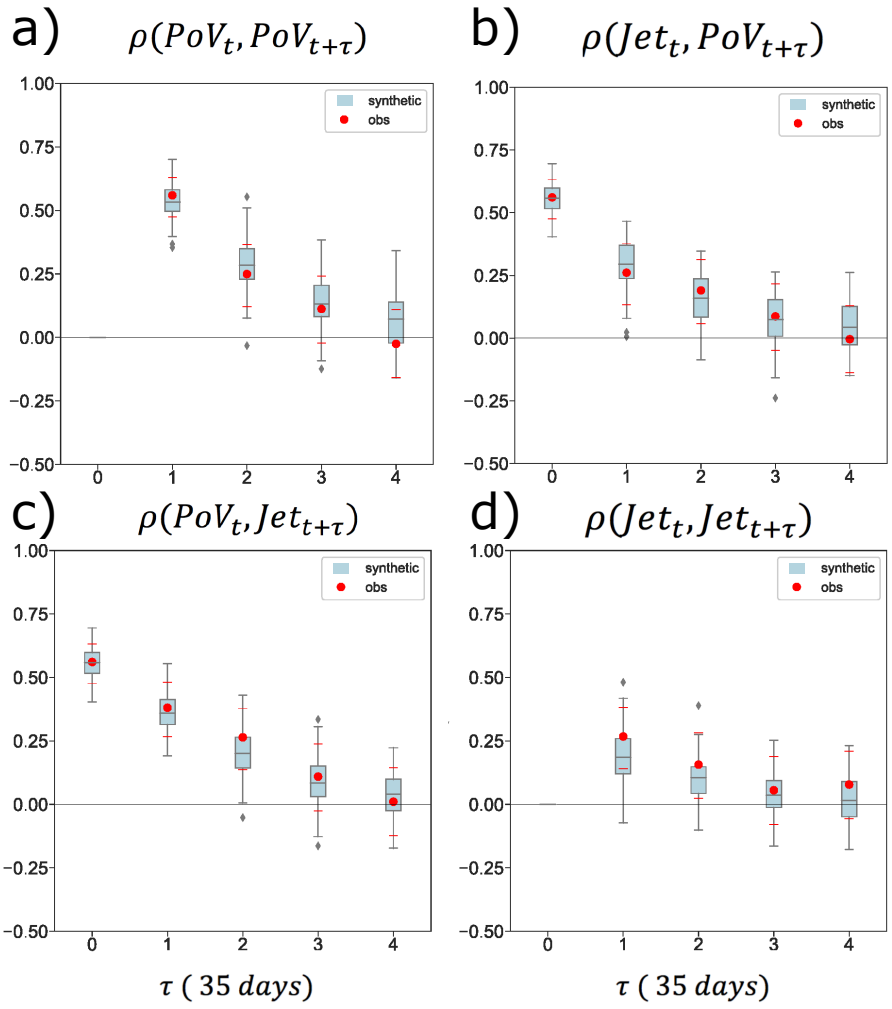

Figure 3.4: (a)-(d) Correlations of reanalysis (red dots) and synthetic (blue box plot) time series. 100 synthetic time series of length 152 are generated using Equation (3.3) and the coefficients in Equation (3.4). Blue boxes cover the 25-75 percentile ranges; the line is at the median; whiskers show 9-91 percentiles. Red dots are the observational data with bars indicating the 95% confidence range according to a two-tailed Student's t-distribution.

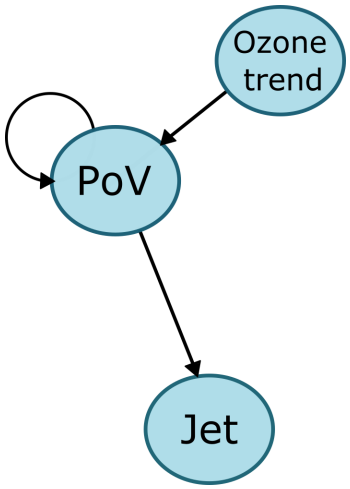

Figure 3.5: Graphical representation of the network described by Equation (3.7).

Equation (3.3). This leads to the following forced system of equations

$$\begin{cases} \overline{PoV}_t &= a\overline{PoV}_{t-1} + \tilde{\gamma}t + \varepsilon_{P,t} \\ \overline{Jet}_t &= c\overline{PoV}_{t-1} + \varepsilon_{J,t} \end{cases} \tag{3.7}$$

also represented graphically in Figure 3.5. Asymptotically, this equation leads to the following trends for both variables

$$\begin{aligned} \overline{PoV}_t &= \frac{\tilde{\gamma}}{1-a}t \\ \overline{Jet}_t &= c\frac{\tilde{\gamma}}{1-a}t \end{aligned} \tag{3.8}$$

where the coefficient c of the downward link $\overline{PoV}_t \rightarrow \overline{Jet}_{t+1}$ controls how effectively the externally induced stratospheric trend is transferred to the tropospheric variable.

The expected trend derived in Eq. (3.8) are here compared with reanalysis. The reanalysis trends are estimated as ${\gamma_P}^{obs}$ and ${\gamma_J}^{obs}$ by computing, for each calendar day between 1 September and 5 March, the increase per decade of the 30-day mean value of the time series around that calendar day across the years of interest. The first 22 years of the record corresponding to when ozone was declining strongly, 1979-2001 (Epoch 1), are analysed separately from the last 22 years when the ozone hole was approximately stabilized and started to recover, 1996-2018 (Epoch 2). The overlapping is chosen to avoid too strong an end-point inter-dependence, and to create reasonably long epochs of equal length. For each calendar day, the observed value for the PoV

trend is assigned to $\tilde{\gamma}$ in Equation (3.7), and 100 time series for the Jet of 30 time steps each are generated. From these, the synthetic Jet trend is estimated. Note that 100 members are used to estimate the uncertainty of the fit and 30 time steps are chosen to be close to the 22 data points used in reanalysis. Finally, since the lag of the downward link is 35 days, the synthetic Jet trend is associated to 35 calendar days later than the PoV forcing. For the synthetic vortex time series to be comparable with observed values, a rescaling is required as described in Appendix 3.D.

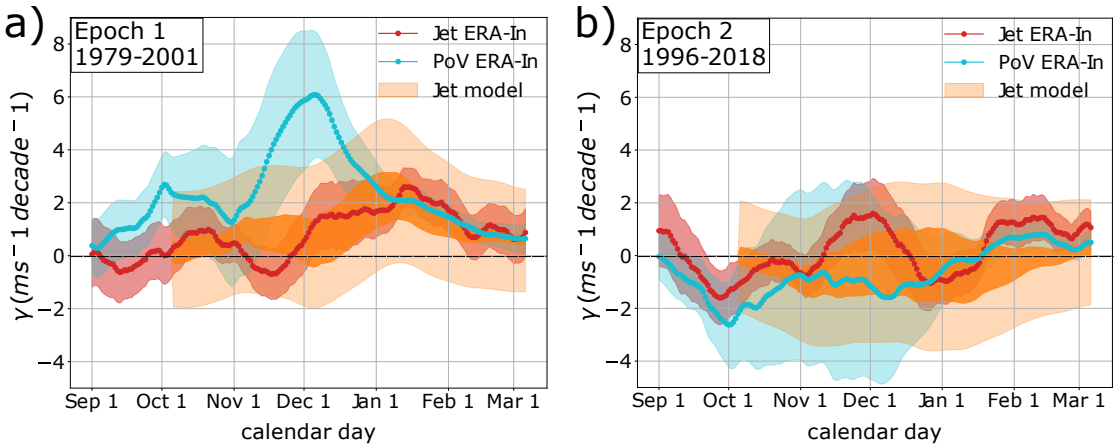

Figure 3.6: PoV and Jet observed trends (${\gamma_P}^{obs}$ in light blue and ${\gamma_J}^{obs}$ in red) between 1979-2001 (a) and 1996-2018 (b). Trends are computed as described in text and the uncertainty range is the standard error of the regression coefficient. The modelled Jet trend and its uncertainty (orange) are computed as the synthetic ensemble average trend and ensemble standard deviation. The spread is smoothed using a Gaussian window with a 15-days sigma width. Because the uncertainty in the best choice for the observed vortex trend is non negligible (blue error bar, $\sigma_{\gamma_P}^{obs}$), the associated uncertainty $c \cdot \sigma_{\gamma_P}^{obs}$ (dark orange) is added to the ensemble spread (light orange).

Figure 3.6(a) shows the result for Epoch 1. The observed month-specific vortex trend (light blue) is strong, as expected: it peaks around late November and early December, with a mean value of 6 m/s per decade. The induced synthetic jet trend (orange) overlaps with the observed trend (red): they are both smaller than the vortex trend, and peak in early January, consistent with the about one month of delay necessary for the stratospheric signal to reach the troposphere. On the other hand, in Epoch 2 of Figure 3.6(b) no significant stratospheric trend is seen (light blue) and therefore no synthetic jet trend can be induced either (orange), as it is the case in reanalysis too (red). Note that the results on the trends discussed here also hold if a small jet auto-dependence is

included, which could represent a tropospheric feedback that was too weak to be detected by our analysis. Appendix 3.E shows the effect of adding the term $b\overline{Jet}_{t-1}$ to the jet formula of Eq. (3.3) for a range of coefficients b.

Overall Figure 3.6 suggests that the linear model of the S-T coupling here derived has minimal but relevant ingredients to capture the statistical connection between long-term trends in vortex and jet in the late 20$^{\text{th}}$ century. This purely observational attribution of the tropospheric jet shift is the first of its kind, and complements previously published model-based attributions. This model agrees with the physical idea that a coupling relation based on internal variability can be used to understand the causal links between forced responses (Gritsun and Branstator, 2007). It is also consistent with the relation found by Ceppi and Shepherd (2019) for the SH S-T coupling, in that case the vortex breakdown delay is forced by greenhouse gases.

3.5 Discussion

The analysis presented in this chapter demonstrated that the autocorrelation timescale of the SH tropospheric eddy-driven jet between late spring and early summer is strongly influenced by the stratospheric polar vortex. The effect of the downward coupling in this time of the year results in the jet being to a large degree "controlled" by the vortex: a highly autocorrelated vortex determines the a highly autocorrelated jet. This implies that to understand present and future changes to the jet's persistence timescale the role of the stratosphere cannot be discarded (e.g. Ceppi and Shepherd, 2019; Mindlin et al., 2021; Breul et al., 2022).

Another finding is that the connection between the vortex and the jet happens on a timescale of about one month, confirming that the connection is not instantaneous but needs some time for the stratospheric information to impact on the troposphere. It further suggests a role for lower-stratospheric radiative processes. The connection is strongest in the time leading up to the vortex breakdown, more specifically the information travelling from the vortex to the jet originates mostly from the few months and weeks before the vortex breakdown. Afterwards it is very minimal. Hence the whole period of destabilization of the vortex can be a source of information for the troposphere (e.g. Byrne and Shepherd, 2018; Lim et al., 2018). Further, in this time of the year and at this timescale, the jet does not seem to have a detectable influence on the vortex.

The simple parametrised model of the coupling reproduced well the correlation fingerprint

present in ERA Interim. It was further found to be able to connect the observed trend in the stratospheric vortex strength (known to be driven by ozone depletion) with the observed trend in tropospheric jet strength, the latter resulting from a delayed equatorward shift instigated by a later vortex breakdown (Byrne et al., 2017). This further stresses the role of the stratosphere in understanding future changes in the SH large scale tropospheric circulation. For example, the underlying assumption in Ceppi and Shepherd (2019) is that, if the mechanism of the downward coupling (which explains intra-seasonal dynamics) holds also under forcing that affects the vortex (e.g. the effect greenhouse gases), then the downward connection should allow to understand the impact of such forcing onto the jet too.

Among the limitations of this analysis is the use of a dataset with small sample size (especially when time averaged at intervals longer than 20-30 days) which means that some of the results may be affected by observational noise. For example, the "true" values of the linear component of the coupling could be different from what is represented here in Equation (3.4). Also a weak but significant auto-dependence in the jet at this time scale could have been missed due to the small sample size; however the conclusions of this analysis have been shown to be robust to such a potentially missing link. The assumption of a linear relationship between the vortex and the jet may also be too simplistic, although it did seem to be enough for a statistical description of the zonally-averaged system. Finally, as explicitly formulated in causal network theory, the omission of variables that may affect both the vortex and the jet (also known as common drivers) may be biasing these results. In this case, a potential common driver is ENSO, however as already discussed its action is usually on longer timescales than the one month and below considered here. In Chapter 5 a study of this system with inclusion of ENSO suggests that the quantification of the connection between the vortex and the jet was robust to this effect.

Finally, based on our physical understanding of the coupling, as well as the statistical analysis carried out here, there is a suggestion that the best way to estimate the coupling (timescale and strength) is to condition the quantification to the days around the time of the vortex break down, which varies quite significantly from year to year. As in many other systems, being able to identify the months or seasons where a given physical process (a link in the network) is "active" in the year is not trivial. This identification is usually a result of expert knowledge, for example by selecting the SH spring months to study processes that dominate the spring time. But there could

be many cases where the start or end of the season is not well defined, or potentially shifting as anthropogenic factors interfere, such as ozone depleting substances, greenhouse gases or aerosols. In Chapter 4 a data-driven method is presented to identify the time points of change (regimes) in the behaviour of a system starting from the non-stationary time series of its variables.

3.6 Conclusion

In this chapter the direct influence of the SH stratospheric polar vortex on the tropospheric eddy-driven jet during the spring-to-summer transition has been estimated from reanalysis using a causal network approach employing partial correlations. While the causal network analysis confirms the expected connections (links), it allows further insight to be gained. For example, it identifies the most parsimonious set of connections in a complex autocorrelated system land is able to reproduce critical statistics, such as the jet autocorrelation. The process for link detection is rooted in causal network and information theory, and their formulation allows to interpret the links in terms of direct information flow between nodes, going beyond stating the significance of the regression coefficients. Although this difference seems subtle, it is significant in terms of interpretation and robustness of the results. More generally, the core idea behind the method seems suitable to identify robust connections that can be used for emergent constraints, and to perform succinct diagnosis of model error, and offers an alternative route to inter-model comparisons (Hammerling et al., 2018; Runge et al., 2019b; Nowack et al., 2020).

APPENDIX

3.A Tropospheric drivers of the SAM autocorrelation time scale

The enhanced autocorrelation time scale of the SH spring to summer tropospheric SAM has been extensively studied by previous works (e.g. Lorenz and Hartmann, 2001; Simpson et al., 2013; Barnes and Hartmann, 2010). The overall conclusion of these works has been that tropospheric feedbacks play a prominent role through the eddy momentum fluxes. Some relevant details to the present study are summarised below.

Lorenz and Hartmann (2001) studied tropospheric phenomena related to the persistence of the tropospheric SAM, inspecting the role of eddies in mediating a feedback on the SAM. They find that the zonal-mean wind influences the low-frequency component of the eddy momentum flux anomalies, which in turn feedbacks on the SAM and increases its persistence. They quantify the effect of this feedback as describing half of the low-frequency variance of the zonal-mean zonal wind (vertically integrated) EOF1. While this study inspects tropospheric correlations in detail, it does not account for the role of stratospheric effects.

Barnes and Hartmann (2010) studied the persistence of the tropospheric SAM using a three-dimensional vorticity budget that allows to account for non-zonally symmetric behaviour. This is the same feedback mechanism as in Lorenz and Hartmann (2001) but described with a vorticity budget. Seasonal and zonal asymmetries are found, and the authors understand them in terms of presence (absence) of local eddy feedbacks. For the SH summer, a main finding is that "positive feedback between the eddies and the jet anomalies extends the persistence of meridional shifts of the midlatitude eddy-driven jet". Quantification of the feedback is done via lagged cross correlation and spectral analysis (the latter employing a synthetic linear model as in Lorenz and Hartmann

(2001)).

Simpson et al. (2013) focused on the internal tropospheric mechanisms responsible for the excessive persistence of tropospheric SAM anomalies in GCMs. The authors quantify the eddy feedback strength for the vertically averaged zonal-mean zonal wind projected onto the meridional SAM pattern. Their method improves Lorenz and Hartmann (2001)'s estimation as it accounts for potential intra-seasonal variability. In Simpson et al. (2013) the effect of stratospheric forcing is not inspected as an earlier analysis found the persistence bias to remain in stratospherically nudged runs as well (Simpson et al., 2011) (although removing the effect of the stratosphere did reduce the persistence, as mentioned in Section 3.1).

3.B The role of stratosphere in SAM autocorrelations

The effect of stratospheric dynamics on SAM time scale has been suggested by others, for example Graversen and Christiansen (2003), Thompson et al. (2005), Baldwin et al. (2003), Gerber et al. (2010) and Byrne and Shepherd (2018), as referenced in the text. In Gerber et al. (2010) an extended study of the stratosphere-troposphere coupling in chemistry-climate models (CCMs) and reanalysis is presented. In the following the results that specifically investigate the time scale and magnitude of the coupling are commented.

The magnitude of the reanalysis intra-seasonal variability of the SAM (standard deviation of the index in each month) is computed separately at each pressure level (Figure 7b therein). A strength of the connection between variabilities at different pressure levels is however not easy to extrapolate and the authors do not comment on this.

The decorrelation of the variability is computed as e-folding time scale of the autocorrelation function of the reanalysis SAM index (see Figure 8b therein, constructed as Figure 7b). An increased timescale in the troposphere (NDJ) is clearly seen while the stratospheric timescale is maximising in SON. Still, inferring a time scale for the downward connection between these persistent periods is not possible from this analysis alone. For example, the connection could be almost instantaneously happening in November, or it could start in September in the stratosphere and take almost three-month time to impact at the tropospheric level.

The 'predictability' at various pressure levels is explored in Figure 9b therein. The index plotted in colour is the square of the correlation coefficient $\rho^2(X,Y)$ between the instantaneous

annular mode index at each season and pressure, X(day,z), and the 30 day mean 850 hPa annular mode index lagged by +10 days, Y(day+10). By linear regression theory, the correlation squared $\rho^2(X,Y)$ corresponds to the fraction of the variance explained by a linear least-squares fit between X and Y. Figure 9b shows that in September and October the stratosphere is more correlated with the future near surface annular mode index than the near surface itself. This is well known as to indicate potential predictability skill coming from the stratosphere. However, the values of this correlations cannot be interpreted as the strength of a direct connection between a stratospheric pressure level and 850hPa: correlation does not equate to predictability, and it can be biased by multiple effects such as autocorrelations, as demonstrated in this chapter. Further, this analysis does not answer to the timescale question either.

3.C Mathematical derivation for the inflating effect of autocorrelation on cross-correlations

The inflating effect of a variable's auto-dependence (here the vortex) in a cross-correlation can be demonstrated in general terms, as studied in Runge et al. (2014). Assuming a two-variable linear system, the lagged unconditional and partial cross-correlations can be explicitly derived as functions of the auto-dependence coefficients. The system

$$\begin{cases} X_t &= aX_{t-1} + \varepsilon_{X,t} \\ Y_t &= cX_{t-1} + bY_{t-1} + \varepsilon_{Y,t} \end{cases} \tag{3.9}$$

describes a stochastic process of two auto-dependent variables (through a, b, with $|a|, |b| < 1$ for stationarity) that are one-way coupled, since X contributes to the state of Y (through c) but not vice-versa. The noise terms are independent and Gaussian with covariance matrix Σ and zero mean. The analytical expression of the lagged cross-correlation

$$\rho(X_t, Y_{t+\tau}) = f(\tau; a, b, c, \Sigma) \tag{3.10}$$

depends on all coefficients of the system (see Table 2 in Runge et al. (2014) for the full derivation). Relevant to our analysis is that the amplitude and position of its peak – often used as a diagnostic for lagged interaction – can be modulated according to the values of the auto-dependent coefficients a and b while keeping the actual causal coefficient c fixed. Also, the function f does not markedly

decay for a specific sign of the lag, thus potentially (but wrongly) suggesting that Y directly influences X. This is because correlation measures the collinearity between lagged time series, which can be due to all sorts of effects.

On the other hand, the MIT partial correlation reads

$$\rho^{MIT}\left(X_t, Y_{t+\tau}\right) = \delta_{\tau,1}\sqrt{\frac{c^2\sigma_X^2}{c^2\sigma_X^2+\sigma_Y^2}} \tag{3.11}$$

where $\delta_{\tau,1}$ is the Kronecker delta function, $\sigma_X^2 = \Sigma_{XX}$ and $\sigma_Y^2 = \Sigma_{YY}$ (see Table 2 in Runge et al. (2014)). The effect of both autocorrelations has been removed (no a, b dependence) thanks to the conditioning set $Z^{MIT} = \{X_{t+\tau-1}, Y_{t+\tau-1}, X_{t-1}\}$. The difference between the correlation and partial correlation function is largely due to the auto-dependence coefficients.

In the present analysis for the vortex and jet, the large auto-dependence of the vortex can be seen as analogous to a large coefficient a (Figure 3.1(a), top-left) resulting in an inflated cross-correlation ρ. This analogy is confirmed to be accurate since conditioning on the set $Z^{MIT} = PoV_{m-1}$ ($PoV_{m+\tau-1}$) results in smaller coefficients for the downward (upward) ρ^{MIT} compared to the simple unconditional ρ, proving that the large vortex auto-dependence is responsible for the large and persistent cross-correlations (Fig. 3.1).

3.D Trend rescaling

For the synthetic vortex time series to be comparable with observed values, they require the same variability-to-trend ratio as the observed vortex, $R^{obs} = R^{synth}$, where the ratio is R is the vortex variability divided by the strength increase over time due to the trend. Specifically,

$$R^{obs} = \frac{\sqrt{Var[\mathrm{PoV}]}}{T^{obs}\gamma_P^{obs}} \tag{3.12}$$

with ${\gamma_P}^{obs}(ms^{-1}/month)$ obtained from reanalysis as described above, T^{obs} =22 (years in one Epoch) $\cdot$ 12 (months in a year) and $Var[\mathrm{PoV}]$ being the reanalysis variance of PoV monthly anomalies. The synthetic ratio is given by

$$R^{synth} = \frac{\sqrt{Var[\overline{PoV}]}}{T^{synth}\tilde{\gamma}} \tag{3.13}$$

with $Var[\overline{PoV}] = \sigma_P^2/(1-a^2)$ (standardized coefficient values as in Eq. (3.4)) and $T^{synth} = 30$ time steps. Imposing $R^{obs} = R^{synth}$, a rescaling factor K is obtained to determine the synthetic PoV trend

coefficient $\tilde{\gamma} = K \cdot {\gamma_P}^{obs}$, where

$$K = \left[\frac{T^{obs}}{T^{synth}} \frac{\sqrt{\sigma_P^2/(1-a^2)}}{\sqrt{Var[\mathrm{PoV}]}} \right]. \tag{3.14}$$

Note that K depends only on vortex parameters. In particular, the observed jet trend is not used in any way. Because the analytic prediction for the synthetic jet trend is $\tilde{\gamma}c/(1-a)$, its conversion to physical values is done by K^{-1} multiplication, which allows to compare it directly with reanalysis ${\gamma_J}^{obs}$.

3.E Effect of jet auto-dependence in the causal network

One of the results of this work is that both i) the long autocorrelation time scales of the jet and ii) the emergence of a more poleward position in the jet latitude in the late 20th century can be explained without any tropospheric mechanisms of eddy-feedbacks. In terms of causal network, this is an absence of jet auto-dependence, i.e no link $Jet_t \rightarrow Jet_{t+1}$ at the monthly timescale.

Let's assume that the equation describing the two variable system has a monthly jet auto-dependence component, which is expressed by coefficient $b \neq 0$:

$$\begin{cases} \overline{PoV}_t &= a\overline{PoV}_{t-1} + \varepsilon_{P,t} \\ \overline{Jet}_t &= b\overline{Jet}_{t-1} + c\overline{PoV}_{t-1} + \varepsilon_{J,t} \end{cases} \tag{3.15}$$

The auto-dependence coefficient b and the jet autocorrelation function are not trivially related. In the case where $b = 0$, there jet autocorrelation can still be non-zero for many lags as proven in Eq. (3.6). Be $Jet_t = b\,Jet_{t-1} + F_{t'<t}$ and let us consider its autocorrelation function. Although b will appear in the jet autocorrelation function, it does not define it completely as long as other terms appear in the equation, like the forcing F. Therefore, even if $b = 0$ this does not necessary imply a small or fast-decaying autocorrelation. As it happens in Eq. (3.6), the Jet time series can "inherit" persistence from the external influence of the vortex that has itself a long decorrelation time. In our case, this influence is provided by the vortex, $F = c\,PoV_{t-1}$. The rejection of the Jet auto-link (and thus b=0) could be a matter of sample size due to the relatively short period of reanalysis. However, the correlation analysis shown in Figure 3.1 suggests that even a truly non-zero b coefficient was missing, it would not be very large given the already good agreement with reanalysis.

As for the trend analysis in Section 3.4.3, the robustness of the results presented is tested with the introduction of a b coefficient. The numerical calculation of the trend are performed using a linear model with a small b coefficient included (0.1, 0.2, 0.3). The change in magnitude of the trend is not major as shown in Figure 3.E.1 and the formula presented in the paper (black line in figure, b=0) remains a valid approximation.

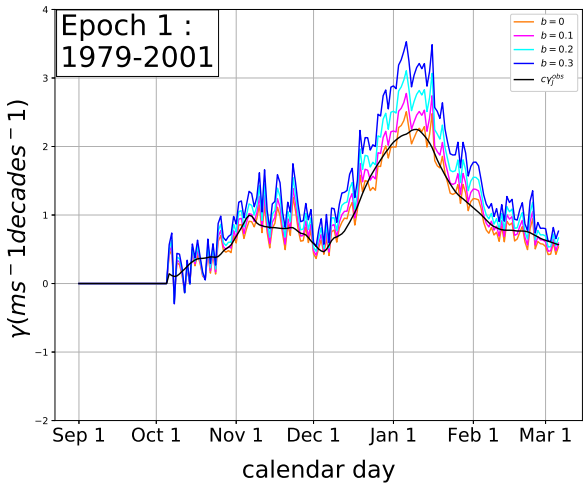

Figure 3.E.1: The modelled Jet trend for different values of coefficient b, and fixed c=0.37 and a=0.55 as in Eq. (3.3). The black line corresponds to the analytical prediction when b=0.

Chapter 4

Reconstructing regime-dependent causal relationships from observational time series

The causal network of the S-T coupling studied in Chapter 3 presented a feature of non-stationarity. It was shown that the downward coupling is organized around the time of the vortex breakdown event, whose interannually variable timing tends to happen between late October and early January. Indeed there are many climatic processes with "activation" that depends on time, for example teleconnections that are present only in some seasons of the year. Not considering the precise timing of such activation can confound interpretation of statistical analysis (e.g. Byrne et al., 2017). There can be cases where the timing is not know a priori, for example due to internal variability or changing external forcing. Motivated by this challenge, which is not unique to climate science and a general limitation of most causal discovery methods, in this chapter a time-series causal discovery method for regime-dependent non-stationarity is proposed.

4.1 Introduction

The interest in causal analysis for time series data is increasingly common in the context of climate research (Section 1.4), but also in many other disciplines such as econometrics (Matthieu et al., 2016), molecular (Gerber and Horenko, 2014) and animal group (Strandburg-Peshkin et al., 2018)

dynamics. In many of these fields conducting experiments is often infeasible or unethical and causal analysis is being considered as a way to make causal statements possible. What is often given is a set of time series describing the processes of interest with limited (or even no) specific knowledge about the direction and form of their causal relationships available. Thus the task at hand is to reconstruct the underlying graph of causal relationships from time series data, termed time-series causal discovery (introduced in detail in Section 2.5). Each causal discovery method comes with its particular set of assumptions about properties of the underlying processes and the observed data (Spirtes et al., 2000). Runge et al. (2019b) recently provided an overview of current methodological frameworks, their application scenarios and open challenges, with a focus on Earth system sciences.

Today's ever-growing abundance of time series datasets promises many new application scenarios for the now numerous causal discovery methods. But many challenges emerging from the dynamic nature of such datasets have not yet been met. One such issue is posed by time-varying causal relationships, a frequent feature of both natural and artificial systems. In Earth sciences, for example, the dominant causal relationship between soil moisture and air temperature periodically reverses due to land-atmosphere feedbacks (Seneviratne et al., 2010); in animal group dynamics, the leader-follower role of an individual often mutates in time (Nakayama et al., 2012); and in econometrics, the direction of influence between stock markets and macroeconomic variables is often dynamical (Muradoglu et al., 2000).

A feature commonly observed in non-stationary dynamical systems is regime-dependence. Regime-dependence means that the causal relationships between the considered processes vary depending on some prevailing background regime that may be modelled as switching between different states. Further, often such regimes have strong persistence, that is, they operate and affect causal relations on much longer time scales than the causal relations among the individual processes. In the climate system, for instance, several cases of such regime-dependencies exist. For example, rainfall in India in summer is known to be influenced by the El Niño Southern Oscillation (ENSO), an important mode of variability in the tropical Pacific affecting the large-scale atmospheric circulation and thereby weather patterns around the globe (Webster and Palmer, 1997; Shaman and Tziperman, 2007). It is, however, generally assumed that ENSO only marginally affects Indian rainfall in winter (Pal et al., 2015). Thus, the causal relationships between ENSO

and rainfall over India change depending on the season, which here defines the background regime operating on a longer time scale (several months) than the causal relations between ENSO and Indian rainfall (several weeks).

4.1.1 Existing literature on non-stationary causal discovery

Causal discovery has seen a steep rise in interest with a plethora of novel approaches and methods in recent years. As outlined in Section 2.5, in general causal (graph) discovery methods can be classified into classical Granger causality approaches (Granger, 1969; Barnett and Seth, 2015), constraint-based causal network learning algorithms (Spirtes et al., 2000), score-based Bayesian network learning methods (Chickering, 2002; Koller and Friedman, 2010), structural causal models (Peters et al., 2017; Porfiri and Marin, 2018), and state-space reconstruction methods (Arnhold et al., 1999; Sugihara et al., 2012). Here the focus is on the constraint-based framework which has the advantage that it can flexibly account for linear and nonlinear causal relations and different data-types (continuous and categorical, univariate and multivariate).

In particular, the PCMCI algorithm (Runge et al., 2019a) is adopted to reconstruct time-series causal graphs, and it is introduced in detail in Section 2.5.2. In summary, PCMCI is an adaptation of the constraint-based PC algorithm (Spirtes et al., 2000) that addresses autocorrelation of time series via the use of a momentary conditional independence (MCI) test. PCMCI yields high detection power also in high-dimensional and strongly autocorrelated time series settings (see Section 2.5.2 and Runge et al. (2019a) for more details). However, one of the general assumptions of PCMCI (as well as of other causal discovery algorithms) is *link stationarity*, i.e., that the existence or absence of a causal link does not change over the considered time series segment (Runge, 2018). While known changes in the background signal can be accounted for by restricting the time series to the stationary regimes, PCMCI cannot handle unknown background regimes which constitute a particular case of latent confounding.

Some recent work addresses causal discovery in the presence of non-stationarity. Malinsky and Spirtes (2019) model non-stationarity in the form of (continuous) stochastic trends in a linear autoregressive framework. Zhang et al. (2017) account for non-stationarity in the more general constraint-based framework. However, both address the case of a (smoothly) varying continuous background variable that continuously changes causal relations among the observed variables. This

means that these methods will not output regime-dependent causal graphs, but a "summary" graph that accounts for regimes modelled as latent drivers. In Peters et al. (2016) and Christiansen and Peters (2020) assumed known non-stationary regimes are exploited to estimate causal relations also in the presence of general latent confounders. Further, in the context of continuously varying causality, methods based on information transfer metrics have been proposed. In the field of animal group dynamics, for instance, detection of time-varying leader-follower relationships is achieved with the use of a time dependent transfer entropy [1] (Mwaffo et al., 2018; Porfiri and Marin, 2018; Butail and Porfiri, 2019). Applied to non-stationary climate systems, Hagan et al. (2019) proposed a Kalman filter estimate of the time-varying parameters for the Liang-Kleeman information flow (a formalism to compute transfer entropy). Benefits of these methods are the treatment of non-linearity (Mwaffo et al., 2018) and the identification of both timing and frequency of interactions (Hagan et al., 2019). However, in these approaches only bivariate influences are modelled, i.e. the effect of a third variable Z on the estimated effect of X onto Y cannot be accounted for. The practical extension to high dimensional systems, as well as to short time series, thus remains hard to address.

Currently few methods exist that address the case of a discrete regime variable leading to distinct causal regimes that may be physically interpreted. For example, in the climate science context, regime-dependent autoregressive models (RAM) were introduced already in 1990 (Zwiers and Storch, 1990). These can yield physically well interpretable results that, however, require well-chosen ancillary variables and a seasonal index which are not learned from data. Thus, RAM requires a priori knowledge of the regimes, which one often aims to learn rather than enforce. In the context of discrete state spaces, regime dependent causal discovery has been considered in Gerber and Horenko (2014) for boolean variables. Non-stationary Boolean network models have also been considered in Porfiri and Marin (2018), specifically the approach is to fit an appropriate parameterisation of associate transition probabilities. Another approach that has been proposed to model time dependent Granger (non-) causality is based on a Markov Switching VAR ansatz with an economics application in mind (Matthieu et al., 2016). Specifically, the regime assignments are computed by sampling from a Markov chain. Further methods have been proposed to obtain

[1] Transfer entropy from a process X to another process Y is the amount of uncertainty (entropy) reduced in future values of Y by knowing the past values of X given past values of Y. It is thus based on conditional mutual information.

time-step specific bivariate Granger Causality from partitioning the time series into regular time windows (Zanin and Papo, 2017; Jiang et al., 2017).

A more general framework to handle discrete regimes is the Markov-switching ansatz of Wiljes et al. (2014), which flexibly models regime-dependence utilising the assumption of a finite number of regimes and a level of persistence in the transitions between different regimes. This ansatz has been successfully realised in combination with many different model assumptions (e.g. see Horenko, 2010; Falkena et al., 2020). Here this is explored for causal networks by combining it with PCMCI. This method is called Regime-PCMCI.

The remainder of this chapter is structured as follows. In Section 4.2 the underlying mathematical problem, concepts, and key assumptions are formalised, and a motivating example is discussed to provide some intuition. The Regime-PCMCI method is then presented in Section 4.3. These theoretical and algorithmic parts are complemented by a thorough numerical investigation of the proposed method in various artificial settings in Section 4.4. Finally, in Section 4.5, Regime-PCMCI is applied to a real-world dataset from climate science, addressing the changing relationships of ENSO and rainfall over India.

4.2 Problem setting

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a sequence of real-valued N_X dimensional random variables $X_t \in \mathbb{R}^{N_X}$ where t is associated with time. A realisation over the time interval $[0,T]$ of this stochastic process is denoted $\{\mathbf{x}_t\}_{t\in[0,T]}$ and the assumption is that it is possible to obtain observations of these realisations. The underlying process is assumed to be modelled by a regime-stationary discrete-time structural causal model (SCM)

$$X_t^j = g_t^j(\mathcal{P}_t^j, \eta_t^j) \quad \text{with } j = 1,\dots,N_X \ . \tag{4.1}$$

Here the measurable functions g_t^j depend non-trivially on all their arguments, the noise variables η_t^j are jointly independent and are assumed to be stationary, i.e., $\eta_t^j \sim \mathcal{D}^j$ for all t for some distribution $\mathcal{D}^j$, and the sets $\mathcal{P}_t^j \subset (X_{t-1}, X_{t-2}, \dots)$ define the causal parents of X_t^j. Note that lagged causal relationships are assumed, but this is not a necessity since there exist causal discovery algorithms that can deal with contemporaneous causal links (Runge, 2020) and also hidden confounders. In contrast to approaches assuming stationarity, both g_t^j and $\mathcal{P}_t^j$ are allowed to depend on

List of notations			
Model		Parameters	
$\{X_t\}_{t\in[0,T]}$	stochastic process	N_X	dimension of stochastic process
η_t^j	noise variable of component X_t^j	T	time length of stochastic process
$\mathcal{D}^j$	stationary noise distribution	N_K	number of regimes
g_t^j	structural causal model function	N_C	max switches for each regime
$\mathcal{P}_t^j$	causal parents of X^j, time dependent	N_M	regime average persistence
$\mathbf{x}_t$	realization of X_t	$\tau_{\max}$	maximum causal time lag
$\widehat{\mathbf{G}}_t$	operator in inverse problem	CI test	conditional independence test
$\hat{g}_t^j$	components of operator $\widehat{\mathbf{G}}_t$	α_{PC}	significance level for PC_1 step
Θ_t	unknown parameter in inverse problem	α	link significance level
$\mathbf{L}(\Gamma,\mathcal{P},\Phi)$	cost functional	N_Q	number of optimisation iterations
$\Gamma(t)$	regime-assigning process	N_A	number of annealings
Φ_t	linear link coefficients, time dependent	N_R	number of realisations for a toy example
$\gamma_k(t)$	regime-assigning process for regime k	$\mathcal{N}(0,\{\sigma^2\}^{\text{ref}})$	ground-truth gaussian noise distribution
$\Phi_k^j(i,\tau)$	linear link coefficient in regime k	$\{\Phi_k^j(i,\tau)\}^{\text{ref}}$	ground-truth linear link coefficient
$\mathcal{P}_k^j$	causal parents of X^j in regime k	N_{para}	number of model parameters
Υ_k	collection of time steps associated with regime k	$\hat{\mathbf{x}}_{k,t}$	reconstructed time series for regime k

Table 4.2.1: Notation used throughout the chapter.

regimes in time as further formalised in Assumption 1 (Section 4.2.2).

The problem setting considered here is the following inverse problem

$$\mathbf{x}_t = \widehat{\mathbf{G}}_t\Big(\mathbf{x}_{t-1},\ldots,\mathbf{x}_{t-\tau_{\max}};\Theta_t\Big) \tag{4.2}$$

with $\widehat{\mathbf{G}}_t = [\hat{g}_t^1,\ldots,\hat{g}_t^{N_X}]$ where $\hat{g}_t^j$ belongs to an appropriate function space for each t and j. $\tau_{\max}$ is the maximum considered time lag. In other words, the aim is to fit a set of unknown parameters Θ_t on the basis of an observed time series $\{\mathbf{x}_t\}_{t\in[0,T]}$. In the next section the particular structure of the parameters Θ_t will be discussed.

The notation used in this chapter is summarised in Table 4.2.1 and abbreviations are reported in Appendix 4.C.

4.2.1 Causal graphs

The nodes in the time series casual network associated with Equation (4.1) are the individual time-dependent variables X_t^j with $j = 1,\ldots,N_X$ at each time $t \in \mathbb{Z}$. Variables $X_{t-\tau}^i$ and X_t^j for a time

lag $\tau > 0$ and a given t are connected by a lag-specific directed link, denoted $X^i_{t-\tau} \to X^j_t$, when $X^i_{t-\tau} \in \mathcal{P}^j_t$ for a particular t. The maximum ground truth time lag of any parent is $\tau^{\mathcal{P}}_{\max}$. Recall that, if a SCM is not given, a practical way to define links is that $X^i_{t-\tau}$ is not conditionally independent of X^j_t given the past of all variables, defined by $X^i_{t-\tau} \not\perp\!\!\!\perp X^j_t \mid \mathbf{X}^-_t \setminus \{X^i_{t-\tau}\}$, with $\not\perp\!\!\!\perp$ denoting the absence of conditional independence (see Section 2.5.1).

The collection of parent sets for all components at time t is denoted $\mathcal{P}_t = \{\mathcal{P}^1_t, \dots, \mathcal{P}^{N_X}_t\}$. This set of parents is part of the unknown parameters to be inferred. Note that their dimensionality is assumed finite, but not known a priori. The other quantity of interest is the functional form of the causal relations $g^j_t(\mathcal{P}^j_t, \eta^j_t)$ in Eq. (4.1) corresponding to these links. A known function class $\widehat{\mathbf{G}}_t(\dots;\Phi_t)$ is assumed, of unknown coefficients $\Phi_t = \{\Phi^1_t, \dots, \Phi^{N_X}_t\}$ that are going to be inferred via

$$\mathbf{x}_t = \widehat{\mathbf{G}}_t(\mathcal{P}_t;\Phi_t). \tag{4.3}$$

In other words for a given time series $\mathbf{x}_t \in \mathbb{R}^{N_X}$ with $t \in [0,T]$ and known function class $\widehat{\mathbf{G}}_t$ the aim is to find the unknown parameters $\Theta_t = [\mathcal{P}_t, \Phi_t]$. Note that the focus of this work is on linear function classes $\widehat{\mathbf{G}}_t$, as discussed in Section 4.3.

4.2.2 Persistence

As mentioned above, in many application areas non-stationarity may be modelled not in the form of abrupt or continuous changes, but via piece-wise stationary regimes (Gerber and Horenko, 2014; Risbey et al., 2015; Williams et al., 2017). These regimes will exhibit a certain persistent behaviour. In order to capture non-stationary systems with these properties the inference is restricted to regime-dependent persistent dynamics.

Assumption 1: *Denote the causal parents and functional dependency of a given variable j for a regime k as $\mathcal{P}^j_t = \mathcal{P}^j_k$ and $g^j_t(\mathcal{P}^j_t, \eta^j_t) = g^j_k(\mathcal{P}^j_k, \eta^j_t)$. A regime is called (N_M, N_K)-persistent if the parents and functional dependencies are stationary for an average of N_M consecutive time steps t, and that there is a finite number of regimes on the whole time domain, i.e., $k \in \{1, \dots, N_K\}$.*

The persistency enters here via the regime average persistence N_M, which also naturally implies a finite number of regimes $N_K \leq T/N_M$.

Under Assumption 1 the considered inverse problem of Equation (4.3) reduces to finding the

unknown parameters $\Theta_t = [\Gamma(t), \mathcal{P}, \Phi]$ comprising of (i) a set of regimes' network parameters

$$\mathcal{P}, \Phi = \{\mathcal{P}_1, \ldots, \mathcal{P}_{N_K}, \Phi_1, \ldots, \Phi_{N_K}\}$$

and, to encode their time-dependence, of (ii) the change points between the regimes given by the regime-assigning process

$$\Gamma(t) = [\gamma_1(t), \ldots, \gamma_{N_K}(t)]$$

with $\Gamma(t) \in [0,1]^{N_K \times T}$. For example, component k of the regime-assigning process can be of the form $\gamma_k = (0,1,1,\ldots 0,1) \in [0,1]^T$ indicating that regime k is active for all time steps for which $\gamma_k(t) = 1$.

4.2.3 Optimisation problem

Finally, in order to solve the inverse problem Equation (4.3) under the persistency Assumption 1, a cost functional can be defined

$$\mathbf{L}(\Gamma, \mathcal{P}, \Phi) = \sum_{t=0}^{T} \sum_{k=1}^{N_K} \gamma_k(t) d(\mathbf{x}_t - \widehat{\mathbf{G}}_t(\mathcal{P}_k; \Phi_k)) \tag{4.4}$$

subject to constraints

$$\sum_{k=1}^{N_K} \gamma_k(t) = 1 \quad \forall\, t, \text{ with } \gamma_k(t) \in [0,1] \tag{4.5}$$

and

$$\sum_{t=1}^{T-1} |\gamma_k(t+1) - \gamma_k(t)| \leq N_C \quad \forall k \tag{4.6}$$

where d is a distance measure, here the squared euclidean distance $\|\cdot\|_2^2$, and $\gamma_k(t)$ can be regarded as the weight of the k regime-specific network at each time t.

This learning approach is based on ideas first proposed in Horenko (2010) and later extended to many different models (Wiljes et al., 2014). The format of $\mathbf{L}(\Gamma, \mathcal{P}, \Phi)$ relies on the assumption that the system associated with the considered data exhibits metastability in time (see Assumption 1, which translates into the summation over k, controlled by the regime number N_K). The desired level of persistence enters the functional in form of a regularization (see Constraint Equation (4.6), controlled by parameter N_C). An alternative option is to add a regularisation term that enforces some form of smoothness of Γ (e.g. Tikhonov regularisation in Tikhonov et al., 1995).

The tuning parameter N_C is related to the *average* regime duration of N_M time steps of Assumption 1 as follows: an average regime duration of N_M in all N_K regimes is implemented by choosing $N_C \approx T/(N_M N_K)$. Importantly, note that the regularization Eq. (4.6) ensures an average persistence on each regime without imposing that *individual* regime durations are a constant; in fact they can be fully irregular within the bound of performing at maximum N_C switches. This regime learning method thus provides a simple, flexible and computationally tractable strategy to go beyond the assumption of fixed length for each regime duration often employed in previous methods (e.g. Jiang et al., 2017; Butail and Porfiri, 2019).

Note that in practice, it is reasonable to assume that an estimate of the average regime switching time N_M is available, consistent with the typical time scale of the application domain. The choice of parameters N_K and N_C (or N_M) will be discussed in Section 4.3.4.

4.2.4 Motivating example

Before introducing the regime detecting causal discovery algorithm, the underlying challenge of causal discovery in the face of regime-dependence is illustrated by giving a simple example. Consider the case of two background regimes and two time series X^1 and X^2 and the associated causal graphs as shown in Figure 4.2.1*a*. Variable X^1 linearly influences X^2 but the sign changes in time, alternating between a positive (during regime 1) and a negative (during regime 2) influence. Here the two regimes alternate regularly in time. The cross-correlation of X^1 and X^2 over the whole time-period is zero because the opposite sign effects cancel each other out in the linear regression. Thus, any linear causal discovery method would fail in detecting the influence of X^1 on X^2 when no a priori knowledge on the two background regimes exists. For example, applying a linear version of PCMCI on the whole time sample would give a network of disconnected variables (Figure 4.2.1*b*). In contrast, if the regimes are known and PCMCI is applied to samples from both regimes separately, the positive and negative links are correctly detected (Figure 4.2.1*a*). To deal with such problems automatically, our algorithm needs to learn both the regimes as well as the regime-dependent causal relations.

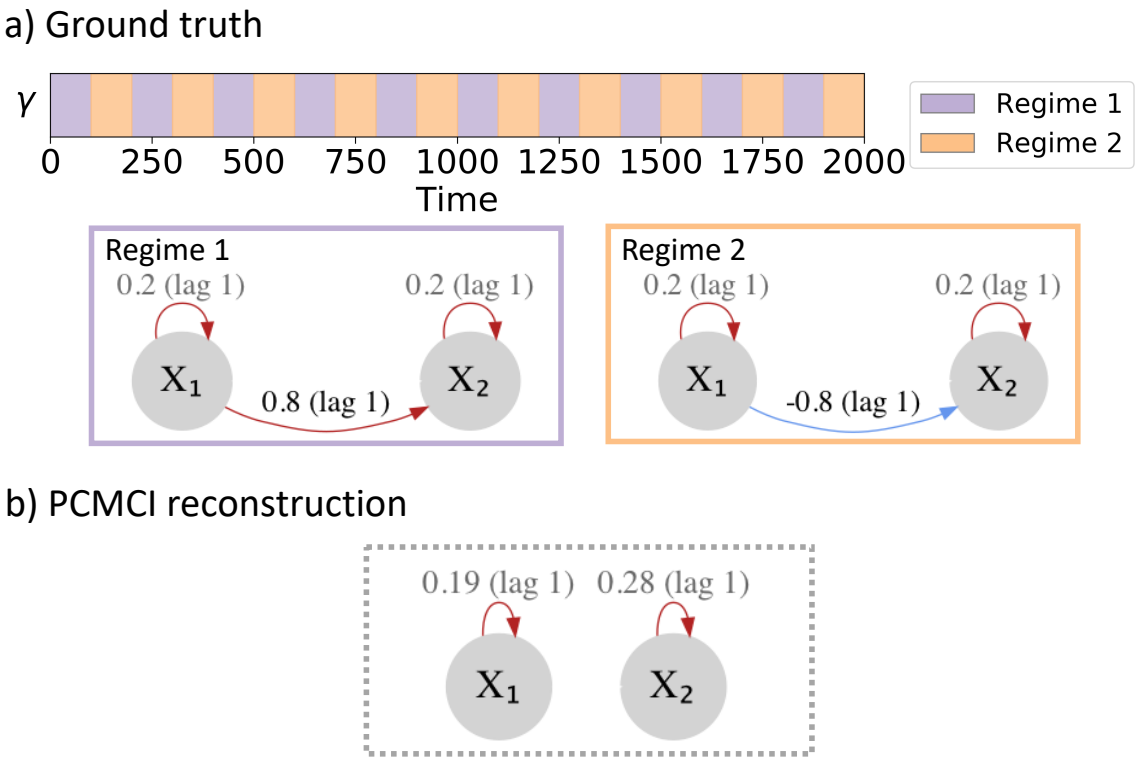

Figure 4.2.1: Motivating example. (a) Regime dependent ground truth: regime-assigning process and regime-dependent networks. The links are labelled with the associated linear coefficient $\Phi_k^j(i,\tau)$ and lag τ. The sign of the coefficient is highlighted by the color (red for positive, blue for negative). (b) Network reconstruction with PCMCI estimated from the whole time series, i.e. if links are wrongly assumed to be stationary.

4.3 Method

The proposed approach, Regime-PCMCI, is designed to solve the optimisation problem Eq. (4.4) by alternating between learning the regimes and learning the causal graphs for each regime in an iterative fashion. In principle, any causal discovery method that yields a causal graph ans associated SCM can be used. Here PCMCI is chosen as a well-tested method that adapts the constraint-based causal discovery framework to the time series case (Runge et al., 2019a).

In the following a linear setting is assumed, which is a reasonable assumption in many application areas (Montgomery et al., 2012; Seber and Lee, 2014). This implies that the function class $\widehat{\mathbf{G}}_t(\mathcal{P}_t;\Phi_t)$ in the inverse problem Eq. (4.3) is assumed to be linear in the parent variables with linear coefficients Φ_t.

4.3.1 Causal discovery

The constraint-based framework, of which PCMCI is a part, discovers causal links using tests of conditional independence (see link definition in Section 4.2.1). This can be done with an existing large variety of conditional independence tests (Runge, 2018; Runge et al., 2019a, for a discussion).

Recall that PCMCI is based on two steps (Section 2.5.2). First is the PC-step, which is a condition pre-selection method to identify relevant conditions $\widehat{\mathcal{P}}_t^j$ for all time series variables X_t^j; second is the MCI-step to confirm whether $X_{t-\tau}^i \to X_t^j$ by means of testing

$$\text{MCI:} \quad X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \widehat{\mathcal{P}}_t^j \setminus \{X_{t-\tau}^i\}, \widehat{\mathcal{P}}_{t-\tau}^i . \tag{4.7}$$

The MCI test then addresses false positive control for the highly-interdependent time series case, which is why it is chosen here. More precisely, while the conditioning on the parents of X_t^j (the potential effect) is sufficient to establish conditional independence in the infinite sample limit (Markov property), the additional condition on the lagged parents (parents of $X_{t-\tau}^i$, the potential cause) leads to a test that is better suited for autocorrelated data. As demonstrated in Runge et al. (2019a), PCMCI has high detection power and controlled false positives also in high-dimensional and strongly autocorrelated time series settings.

Note that a causal interpretation of the relationships estimated with PCMCI comes from the standard assumptions in the constraint-based framework (Spirtes et al., 2000; Runge, 2018; Runge et al., 2019a), namely causal sufficiency, the causal Markov condition, faithfulness, non-contemporaneous effects and stationarity within the regimes, the latter further discussed below (as detailed in Section 2.5.2).

The main free parameters of PCMCI are the chosen conditional independence test, the maximum time lag $\tau_{\max}$, and the significance levels α in MCI and α_{PC} in PC_1. The selection of these parameters is discussed in Section 4.3.4. In terms of the conditional independence test, note that PCMCI can be used in combination with linear or non-linear tests, and can therefore extract also non linear causal relationships. In this work linear systems are selected and we thus use PCMCI in conjunction with partial correlation. All analysis with PCMCI used code freely available in the TiGraMITe Python package[last accessed 20/09/2022].

In each iterative step of our approach, PCMCI is applied to the sample subset of the time series pertaining to the estimated regime k. Given a significance level α, the output of PCMCI is the set

of parents $\mathcal{P}_k = \{\mathcal{P}_k^1, \ldots, \mathcal{P}_k^{N_X}\}$ for all time series variables for that regime,

$$\mathcal{P}_k^j = \{X_{t-\tau}^i : \text{p-value}_{k,\text{MCI}}(X_{t-\tau}^i, X_t^j) \leq \alpha\} \quad \forall k, j\,. \tag{4.8}$$

Based on these parents and associated causal links, causal effects Φ_k that quantify the strength of a link can be estimated. Details of the regime-specific PCMCI fit are found in Section 4.3.2.

4.3.2 Regime-dependent causal discovery

The Regime-PCMCI algorithm iterates over two major estimation steps: (Step 1) causal discovery to obtain $\mathcal{P}_k$ and fit the coefficients Φ_k and (Step 2) regime learning to update the regime variable Γ. To find good estimates of the parameters and the regime variable optimising the cost functional in Eq. (4.4), this two-step procedure is necessary. In fact there are generally no analytic solutions to the problem available due to the complexity of the cost functional. Fixing one variable to estimate the other, allows in both cases to solve the individual optimisation step via linear programming. In Theorem 2.1 of Horenko (2010) and Section 2.b in Metzner et al. (2012) it is shown that these types of algorithms monotonously decrease the value of $\mathbf{L}$. It is important to note however that due to the non-convexity of the underlying problem the algorithm can be caught in regions of local minima. This issue is addressed via additional simulated annealing steps as discussed in more detail in Section 4.3.2. In the following, q indicates the current iteration. The superscript (q) is added combined with brackets to the variables updated in each loop. The details of the consecutive subroutines are laid out below.

Step 1: Causal discovery and model estimation

The first step is to estimate a set of parents $\{\mathcal{P}_k\}^{(q)}$ and coefficients $\{\Phi_k\}^{(q)}$ with $k \in \{1, \ldots, N_K\}$ on the basis of a fixed $\{\Gamma(t)\}^{(q)}$ obtained in step 2 of the previous iteration (first and second bullets of Algorithm 1). In the first iteration, the regimes are assigned randomly. $\{\mathcal{P}_k\}^{(q)}$ and $\{\Phi_k\}^{(q)}$ are estimated on the basis of a subset of the time series $\mathbf{x}_t$ with

$$t \in \{\Upsilon_k\}^{(q)} := \left\{t : \{\gamma_k(t)\}^{(q)} \geq 0.5\right\} \tag{4.9}$$

for each regime k. The regime-dependent parents $\{\mathcal{P}_k\}^{(q)}$ are estimated via PCMCI.

As stated at the beginning of Section 4.3, to solve Eq. (4.3) a linear functional relationship that relates each variable to its parents $\mathcal{P}_k$ is assumed. It implies that coefficients Φ_k can be estimated

from the following regression model for each fixed k:

$$x_t^j = \sum_{X_{t-\tau}^i \in \mathcal{P}_k^{j,(q)}} \{\Phi_k^j(i,\tau)\}^{(q)} x_{t-\tau}^i + \eta_t^j \tag{4.10}$$

for $t \in \{\Upsilon_k\}^{(q)}$. In other words for every $k \in \{1,\ldots,N_K\}$ the following optimisation has to be solved

$$\{\Phi_k^j(i,\tau)\}^{(q)} = \arg\min \left\| x_t^j - \sum_{X_{t-\tau}^i \in \mathcal{P}_k^{j,(q)}} \{\Phi_k^j(i,\tau)\} x_{t-\tau}^i \right\|_2^2 \tag{4.11}$$

for $t \in \{\Upsilon_k\}^{(q)}$. Note that the coefficients not indicated as relevant via the parent set are defined to be zero, i.e., $\Phi_k^j(i,\tau) = 0$ for $X_{t-\tau}^i \notin \mathcal{P}_k^{j,(q)}$.

Step 2: Regime learning

Step 2 is to determine an optimal regime assigning process $\{\Gamma_t\}^{(q+1)} \in [0,1]^{N_K \times T}$ given the current estimates $\{\mathcal{P}_k\}^{(q)}$ for the parents and $\{\Phi_k\}^{(q)}$ coefficients (see third bullet in q-loop of Algorithm 1). In agreement with the cost functional Eq. (4.4), the following optimisation problem needs to be solved: find

$$\{\Gamma_t\}^{(q+1)} = \arg\min \sum_{k=1}^{N_K} \sum_{t=1}^{T} \gamma_k(t) \left\| \mathbf{x}_t - \{\hat{\mathbf{x}}_{k,t}\}^{(q)} \right\|_2^2 \tag{4.12}$$

subject to the constraints Eq. (4.5) and Eq. (4.6), and where for each $k \in \{1,\ldots,N_K\}$

$$\hat{x}_{k,t}^j = \sum_{X_{t-\tau}^i \in \mathcal{P}_k^j} \Phi_k^j(i,\tau) x_{k,t-\tau}^i \quad \text{for } t \in \{1,\ldots,T\}. \tag{4.13}$$

Since the first $\tau_{\max}$ time steps cannot be predicted, the choice is made to set those to $\hat{x}_{k,t}^j = x_{k,t}^j$ and to not consider this portion of the time series in the algorithm evaluation. This step can be solved with standard optimisation linear programming routines.

In order to search for the global minimum of this non-convex problem, the algorithm is run for a number N_A of different initialisations of $\{\Gamma\}^{(0)}$ (annealing). The annealing run with the lowest cost functional objective is chosen as the optimal fit. Note that the individual annealing steps are *embarrassingly parallelizable*, i.e. the problem is already separated into tasks that can be readily solved in parallel.

Algorithm 1 Regime-PCMCI algorithm

Input:

- Time series $\mathbf{x}_t \in \mathbb{R}^{N_X}$ with $t \in \{1,\dots,T\}$
- Parameters:
  - Number of assumed regimes N_K
  - Maximum number of transitions within a single regime N_C
  - Maximum time lag $\tau_{\max}$
  - Functional model $\hat{\mathbf{G}}$, here linear
  - Conditional independence test according to $\hat{\mathbf{G}}$, here partial correlation
  - Significance level α (and α_{PC} for PC_1 step)
  - Annealing steps N_A
  - Number of optimisation iterations N_Q

for $a = 0 : N_A$ **do**

Initialize random $\{\Gamma\}^{(0)} \in [0,1]^{N_K \times T}$

for $q = 0 : N_Q$ **do**

Causal discovery and model estimation:

- Infer parents $\{\mathcal{P}_k\}^{(q)}$ by means of PCMCI run on subset $\left\{\mathbf{x}_t : t \in \{\Upsilon_k\}^{(q)}\right\}$ for each k
- Fit model coefficients $\{\Phi_k\}^{(q)}$ via Eq. (4.11) for each k, and use them to generate k reconstructed time series $\{\hat{\mathbf{x}}_{k,t}\}^{(q)}$ defined for every $t \in \{1,\dots,T\}$ according to Eq. (4.13).

Fit regime assigning process:

- Update $\{\Gamma\}^{(q+1)}$ solving Eq. (4.12).

Break if $\{\Gamma\}^{(q+1)} = \{\Gamma\}^{(q)}$ (a local or global minimum is reached)

end for

end for

Output:

- $\Gamma = [\gamma_1(t),\dots,\gamma_{N_K}(t)]^{\dagger} \in [0,1]^{N_K \times T}$
- Causal parents $\mathcal{P}_k$ and causal effects Φ_k for every $k \in \{1,\dots,N_K\}$

4.3.3 Reconstruction of time series

A single prediction from Eq. (4.13) can be derived as the weighted sum over k

$$\hat{x}_t^{*j} = \sum_{k=1}^{N_K} \lceil \gamma_k(t) \rceil \hat{x}_{k,t}^j \quad \text{for } t \in \{1, \dots, T\}. \tag{4.14}$$

Note however that this is never used in the code, but only Eq. (4.13) is used via its presence in Eq. (4.12).

4.3.4 Parameter selection

Regime-PCMCI involves a number of parameters that need to be chosen. They can be separated into parameters of the causal discovery method PCMCI and those of the regime learning part.

The main free parameters of PCMCI are the chosen conditional independence test, the maximum time lag $\tau_{\max}$, and the significance levels α in MCI and α_{PC} in PC_1. These are described in detail in Section 2.5.2. In summary, α_{PC} should be regarded as a hyper-parameter and can be chosen based on model-selection criteria such as the Akaike Information Criterion (AIC) (Akaike, 1973) or cross-validation. $\tau_{\max}$ could be incorporated into this model selection, but since PCMCI is not very sensitive to this parameter (as opposed to, e.g., Granger causality, Runge et al., 2019a), its choice can be based on lagged correlation functions, see Runge et al. (2019a) for a discussion. The choice of conditional independence test is a modelling assumption guided by the assumed nonlinearity of the underlying process and also finite sample considerations. Finally, α is chosen based on the desired level of false positives.

The two free parameters in the regime learning step are the bound on the number of switches N_C and the number of regimes N_K. Usually N_C can be reasonably inferred from the application and given the number of regimes, as explained in Section 4.2.3. Here N_C is assumed to be known. Note that, since N_C bounds the maximum number of switches between regimes, i.e. the optimal reconstructed number can be lower, and does not constrain the individual regime durations, its choice can account for a degree of error. Yet determining a suitable choice of the unknown number of regimes N_K is a difficult task. In particular, it is hard to find the right balance between avoiding to overfit and to choose appropriately complex models to describe a specific dataset and thus the underlying dynamics well. One way to assess this balance heuristically is to employ an information criterion (IC) (Burnham and Anderson, 2002) which has been derived in the context of regression

models and since been adapted to various other model scenarios including graphs (Shipley and Douma, 2020).

An IC is designed to capture the goodness of fit penalised by the number of parameters in order to prefer models with as few parameters as possible, to avoid overfitting (parsimony). Here the number of parameters is defined as

$$N_{\text{para}} = (N_K - 1)N_C + \sum_{k=1}^{N_K}\sum_{j=1}^{N_X} |\mathcal{P}_k^j|. \tag{4.15}$$

The first term in Eq. (4.15) relates to the number of parameters required to describe Γ which can be fully determined via the change points. The second term in Eq. (4.15) counts the number of relevant parents, or equivalently the non-zero coefficients $\Phi_k^j(i,\tau)$. Here the corrected Akaike Information criterion (AICc) is used, first proposed in Hurvich and Tsai (1989) to estimate N_K, assuming known N_C. Note that the corrected version of the original AIC (Akaike, 1973) is employed to correct for small sample sizes relative to the number of parameters

$$AICc = -2\log(\mathcal{L}) + 2N_{\text{para}} + \frac{2N_{\text{para}}(N_{\text{para}}+1)}{T - N_{\text{para}} - 1} \tag{4.16}$$

where $\mathcal{L}$ is the maximum value of the likelihood function for the model one assumes for the residuals (see Metzner et al. (2012) for a more detailed discussion). Note that the AICc also depends on N_C (as it enters the number of parameters N_{para}) and it is in general possible to simultaneously estimate N_K and N_C (Wiljes et al., 2014; Falkena et al., 2020). The choice of N_K is numerically investigated in Section 4.4.3.

The number of iteration steps N_Q should be chosen to ensure that the optimisation process converges. Via exploratory testing, N_Q was found to show convergence after about 10-20 iterations for all examples investigated. The number of annealing steps N_A should be chosen to ensure spanning a large number of local solutions to this non-convex optimisation problem (Eq. (4.4)). Computational time will set a limit to a too high parameter.

4.4 Results from numerical investigation

In the following the performance of Regime-PCMCI is investigated by means of several toy examples. The artificial data is designed to test the method's robustness and accuracy with respect to

various potential scenarios that could occur in real applications. At first low dimensional ($N_X = 2$) causal relations are studied as the results can be interpreted more easily. Next, higher dimensional settings are also considered ($N_X = 10$). The reference time series are generated with the following linear SCM time series model:

$$x_t^j = \sum_{k=1}^{N_K} \{\gamma_k(t)\}^{\text{ref}} \sum_{X_{t-\tau}^i \in \mathcal{P}_k^j} \{\Phi_k^j(i,\tau)\}^{\text{ref}} x_{t-\tau}^i + \eta_t^j, \qquad \eta_t^j \sim \mathcal{N}(0, \{\sigma^2\}^{\text{ref}}) \tag{4.17}$$

with predefined $\{\Gamma(t)\}^{\text{ref}}$, $\{\Phi_k\}^{\text{ref}}$, and $\{\sigma^2\}^{\text{ref}}$. Note that here equally distributed noises for all variables ($\eta_t^j \sim \mathcal{D}\ \forall j$) are numerically investigated. Appendix 4.A shows also a treatment of heterogeneous noise distributions. Note that the reference set of parents is specified by the non-zero coefficients $\{\Phi_k^j(i,\tau)\}^{\text{ref}}$.

4.4.1 Low dimensional data with two underlying regimes

Firstly, a simple setting of two regimes is considered, i.e. $\{N_K\}^{\text{ref}} = 2$, and a two dimensional underlying process, i.e. $N_X = 2$ or $X_t \in \mathbb{R}^2$. Our aim is to test the performance of Regime-PCMCI for different elemental features that can change between regimes. For brevity, links $X_{t-\tau}^i \to X_t^j$ will be called auto-links (or auto-dependencies) for $i = j$, and cross-links for $i \neq j$. The following scenarios are considered as summarised in Table 4.4.1: sign change of coefficient (in auto-link and cross-link), lag change (in cross-link), coefficient change (in auto-link) and child-parent inversion defined via an assortment of linear functions and associated coefficients. In all examples, each variable is also strongly auto-linked at lag 1 (linear coefficient 0.8), which is a realistic yet challenging assumption for many algorithms.

Experiment settings

Five toy models are defined, in network terms, corresponding to different sets of parents defined via the reference parameters $\{\Phi_k^j(i,\tau)\}^{\text{ref}}$ given in columns 4 to 5 of Table 4.4.1. Further, synthetic regime assigning processes $\{\Gamma(t)\}^{\text{ref}}$ are generated for all examples. More specifically, $\{\gamma_1(t)\}^{\text{ref}}$ is designed to consist of 41 alternating windows, i.e., $\{N_C\}^{\text{ref}} = 40$ regime transitions. The length of these windows is randomly selected to be between 70 and 100. The constraint Eq. (4.5) imposes

$\{\gamma_2(t)\}^{\text{ref}} = 1 - \{\gamma_1(t)\}^{\text{ref}}$. The final length of the time series is capped at $T = 3{,}000$ to ensure equally-long regime assignment time series.

Then an artificial time series $\mathbf{x}_t$ via Eq. (4.17) with $\{\sigma^2\}^{\text{ref}} = 1$ is generated. Note that the stochastic process Eq. (4.17) can be exactly reconstructed via the coefficients $\{\Phi_k^j(i,\tau)\}^{\text{ref}}$, their activation $\{\Gamma(t)\}^{\text{ref}}$ and a specific realisation of the innovation term η_t^j.

The PCMCI parameters are chosen as follows: partial correlation as a conditional independence test, $\alpha = 0.01$, $\alpha_{\text{PC}} = 0.2$ as recommended in Runge et al. (2012a), $\tau_{\max} = 3$, and masking type 'y' (see the documentation of the TiGraMITe Python package for PCMCI for the definition of masking types). The number of regimes was set to $N_K = 2$ and the maximum number of regime transitions is $N_C = 40$, i.e., correct guess on number of regimes and switches (model selection for N_K is investigated in Section 4.4.3). The number of iterations is $N_Q = 20$ and the number of annealings is $N_A = 50$. A summary of the parameters is shown in Table 4.4.2. $N_R = 100$ time series realisations are generated for each example.

Results

The ability of the proposed method to recover the networks and the regimes on the basis of the artificially designed time series are presented in the following. Figures 4.4.1-4.4.5 present results for each case in Table 4.4.1, focusing on one of the N_R synthetic data sets. Table 4.4.3 and Figure 4.4.6 show summary statistics over all N_R runs.

The case *sign* X^1X^2 is discussed in detail. The ground-truth regime evolution and networks are shown in the top part of panels *a* and *b* in Figure 4.4.1; in the middle part of both panels their Regime-PCMCI reconstruction is shown; in the bottom part the difference between reconstructed and true regimes are presented to visually inspect the accuracy. The reconstructed regime assigning process for each regime matches the truth in 99.6% of time steps (97% average value over N_R, see Table 4.4.3). The corresponding networks have all and only the correct links (TPR = 0.99 and FPR = 0.01 average value over N_R); their linear causal effect is also well estimated with each link correct up to ± 0.02 (N_R-averaged error per link is 0.028 (9%)).

The other four cases are presented in Figures 4.4.2-4.4.5. The case *causal effect*, and to a lesser extent *lag change*, are hardest to detect. This is because the difference between the individual regimes and a mixed state of the two is not very large. This adds to the general challenge of

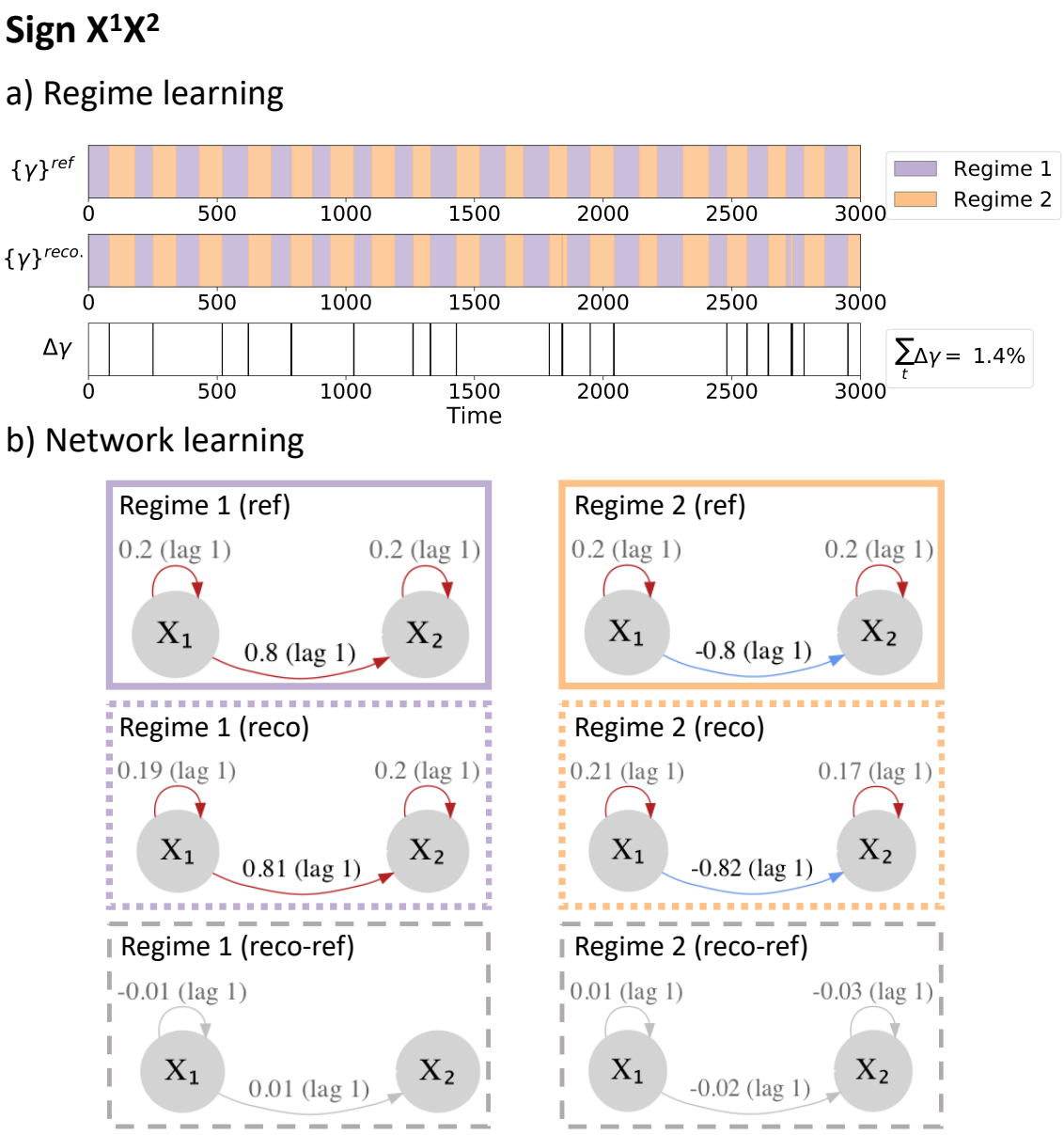

Figure 4.4.1: Example case *Sign* X^1X^2. (a) The ground-truth regime-assigning process, $\{\gamma\}^{ref}$ (top), the Regime-PCMCI reconstructed process, $\{\gamma\}^{reco.}$ (middle) and the difference between the two, $\Delta\gamma$ (bottom). (b) The ground-truth networks for each regime (top), the Regime-PCMCI reconstructed networks (middle) and the difference between the two (bottom). The links are labelled with the associated linear coefficient $\Phi_k^j(i,\tau)$ and the lag τ. The sign of the coefficient is highlighted by the color (red for positive, blue for negative).

Example	$k=1$	$k=2$	$\{\Phi_1^j(i,\tau)\}^{\text{ref}}$	$\{\Phi_2^j(i,\tau)\}^{\text{ref}}$
arrow direction	$X^1 \to X^2$	$X^1 \leftarrow X^2$	$\{\Phi_1^2(1,1)\}^{\text{ref}}=0.8$	$\{\Phi_2^1(2,1)\}^{\text{ref}}=0.8$
			$\{\Phi_1^1(1,1)\}^{\text{ref}}=0.2$	$\{\Phi_2^1(1,1)\}^{\text{ref}}=0.2$
			$\{\Phi_1^2(2,1)\}^{\text{ref}}=0.2$	$\{\Phi_2^2(2,1)\}^{\text{ref}}=0.2$
causal effect	$X^1 \xrightarrow{\lvert a\rvert} X^1$	$X^1 \xrightarrow{\lvert b\rvert} X^1$	$\{\Phi_1^1(1,1)\}^{\text{ref}}=0.8$	$\{\Phi_2^1(1,1)\}^{\text{ref}}=0.1$
			$\{\Phi_1^2(2,1)\}^{\text{ref}}=0.4$	$\{\Phi_2^2(2,1)\}^{\text{ref}}=0.4$
lag	$X^1 \xrightarrow{\tau=1} X^2$	$X^1 \xrightarrow{\tau=2} X^2$	$\{\Phi_1^2(1,1)\}^{\text{ref}}=0.8$	$\{\Phi_2^2(1,2)\}^{\text{ref}}=0.8$
			$\{\Phi_1^1(1,1)\}^{\text{ref}}=0.2$	$\{\Phi_2^1(1,1)\}^{\text{ref}}=0.2$
			$\{\Phi_1^2(2,1)\}^{\text{ref}}=0.2$	$\{\Phi_2^2(2,1)\}^{\text{ref}}=0.2$
sign X^1	$X^1 \xrightarrow{\lvert a\rvert} X^1$	$X^1 \xrightarrow{-\lvert a\rvert} X^1$	$\{\Phi_1^1(1,1)\}^{\text{ref}}=0.8$	$\{\Phi_2^1(1,1)\}^{\text{ref}}=-0.8$
			$\{\Phi_1^2(2,1)\}^{\text{ref}}=0.2$	$\{\Phi_2^2(2,1)\}^{\text{ref}}=0.2$
sign X^1X^2	$X^1 \xrightarrow{\lvert a\rvert} X^2$	$X^1 \xrightarrow{-\lvert a\rvert} X^2$	$\{\Phi_1^2(1,1)\}^{\text{ref}}=0.8$	$\{\Phi_2^2(1,1)\}^{\text{ref}}=-0.8$
			$\{\Phi_1^1(1,1)\}^{\text{ref}}=0.2$	$\{\Phi_2^1(1,1)\}^{\text{ref}}=0.2$
			$\{\Phi_1^2(2,1)\}^{\text{ref}}=0.2$	$\{\Phi_2^2(2,1)\}^{\text{ref}}=0.2$

Table 4.4.1: Artificial model configurations for different low dimensional experiments with $N_K=2$ underlying regimes.

non-convexity of the functional being optimised, which is here mitigated by the annealing steps as mentioned in Section 4.3.4. A similar challenge is found for some high dimensional runs as it is described in Section 4.4.4.

The average accuracy of Regime-PCMCI is estimated from N_R=100 synthetic data sets per each example, and is presented in Figure 4.4.6 and Table 4.4.3.

For a compact overview of the results and to facilitate the comparison between examples, Figure 4.4.6 focuses on two key statistics: the precision of the reconstructed regime-assigning

CI test	$\tau_{\max}$	α	α_{PC}	N_K	N_C	N_Q	N_A
ParCorr	3	0.01	0.2	2	40	20	50

Table 4.4.2: Method parameters for low dimensional examples with $N_K=2$ underlying regimes.

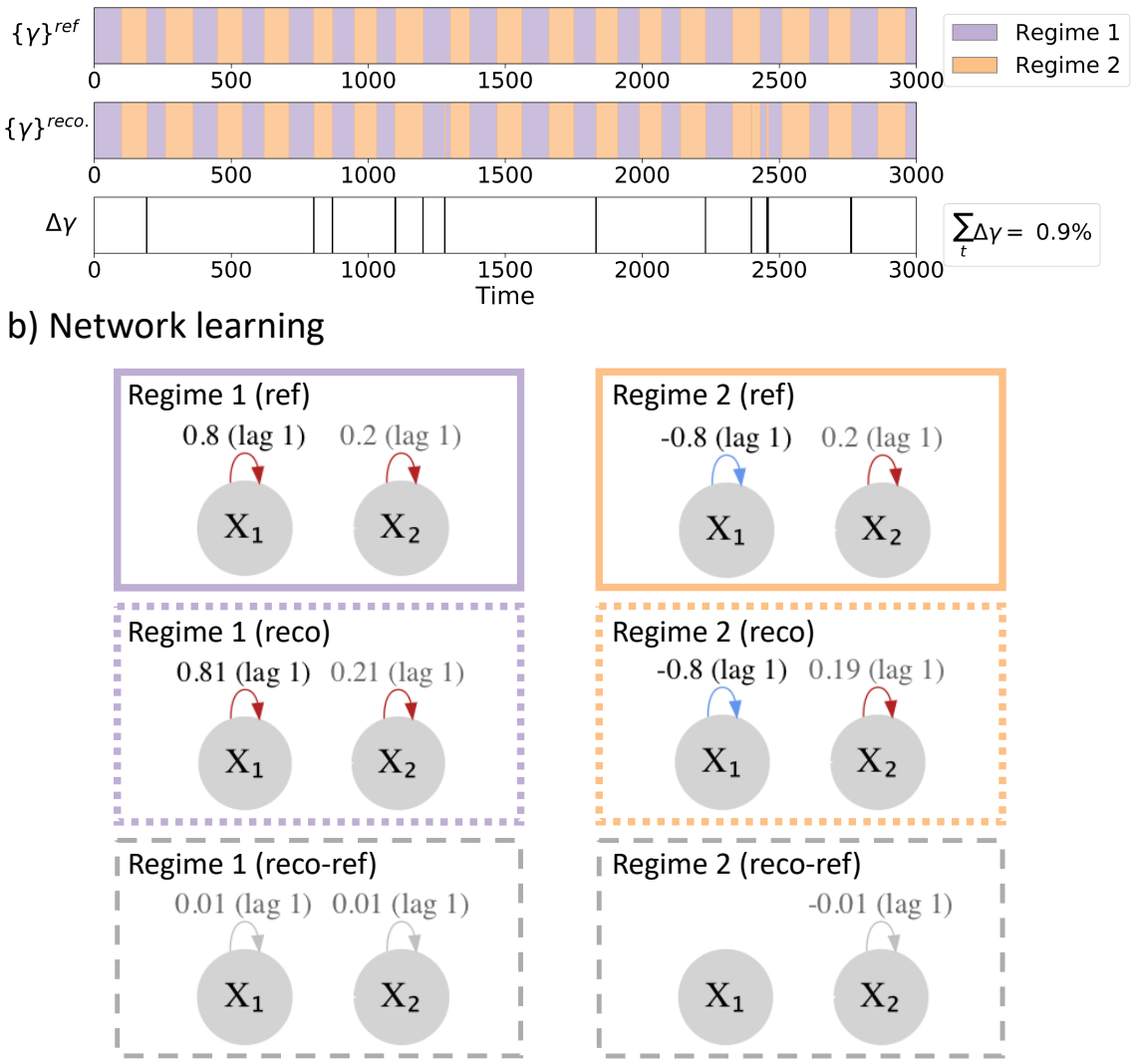

Figure 4.4.2: Example case *Sign* X^1. See description in Figure 4.4.1.

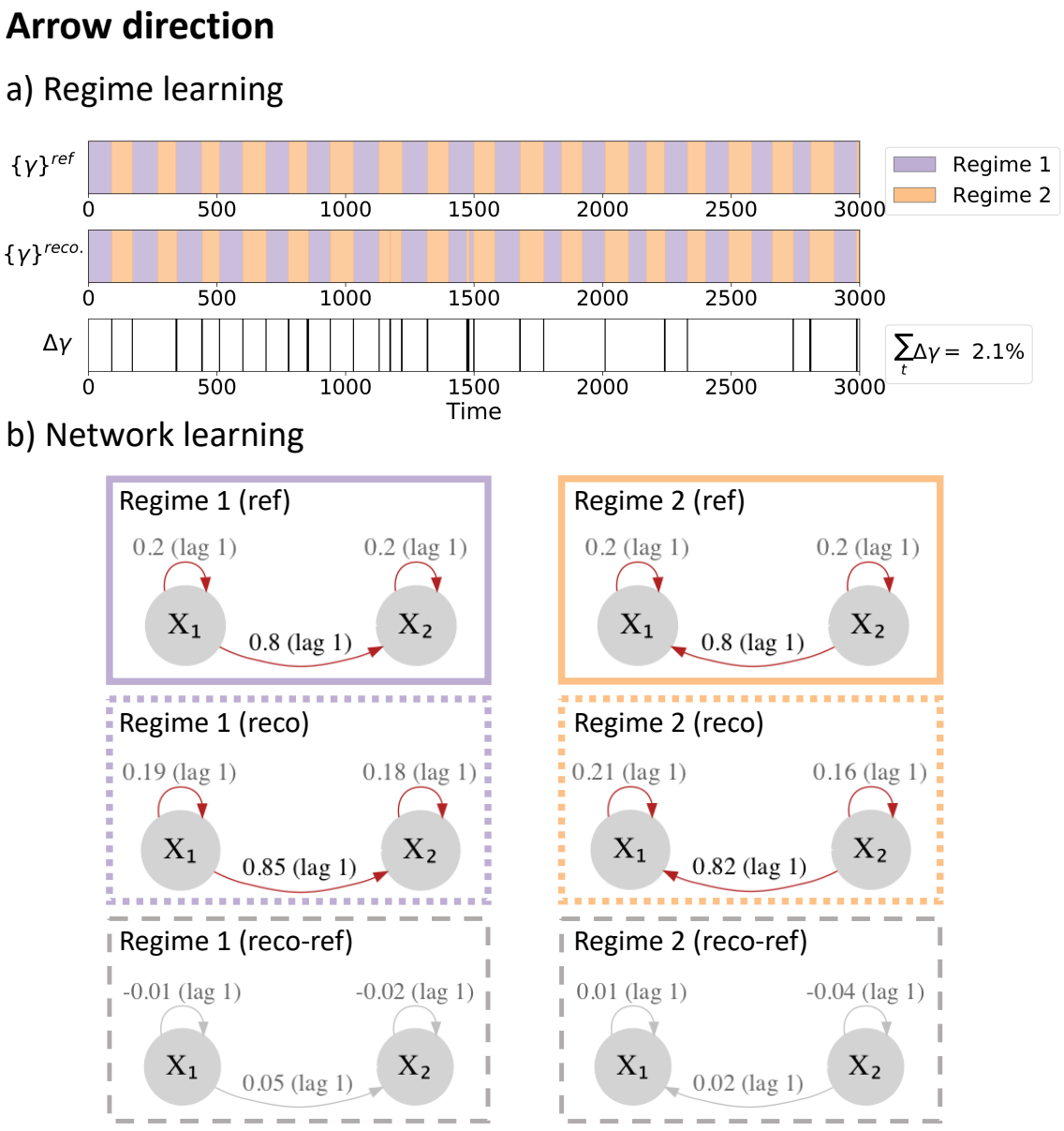

Figure 4.4.3: Example case *Arrow direction*. See description in Figure 4.4.1.

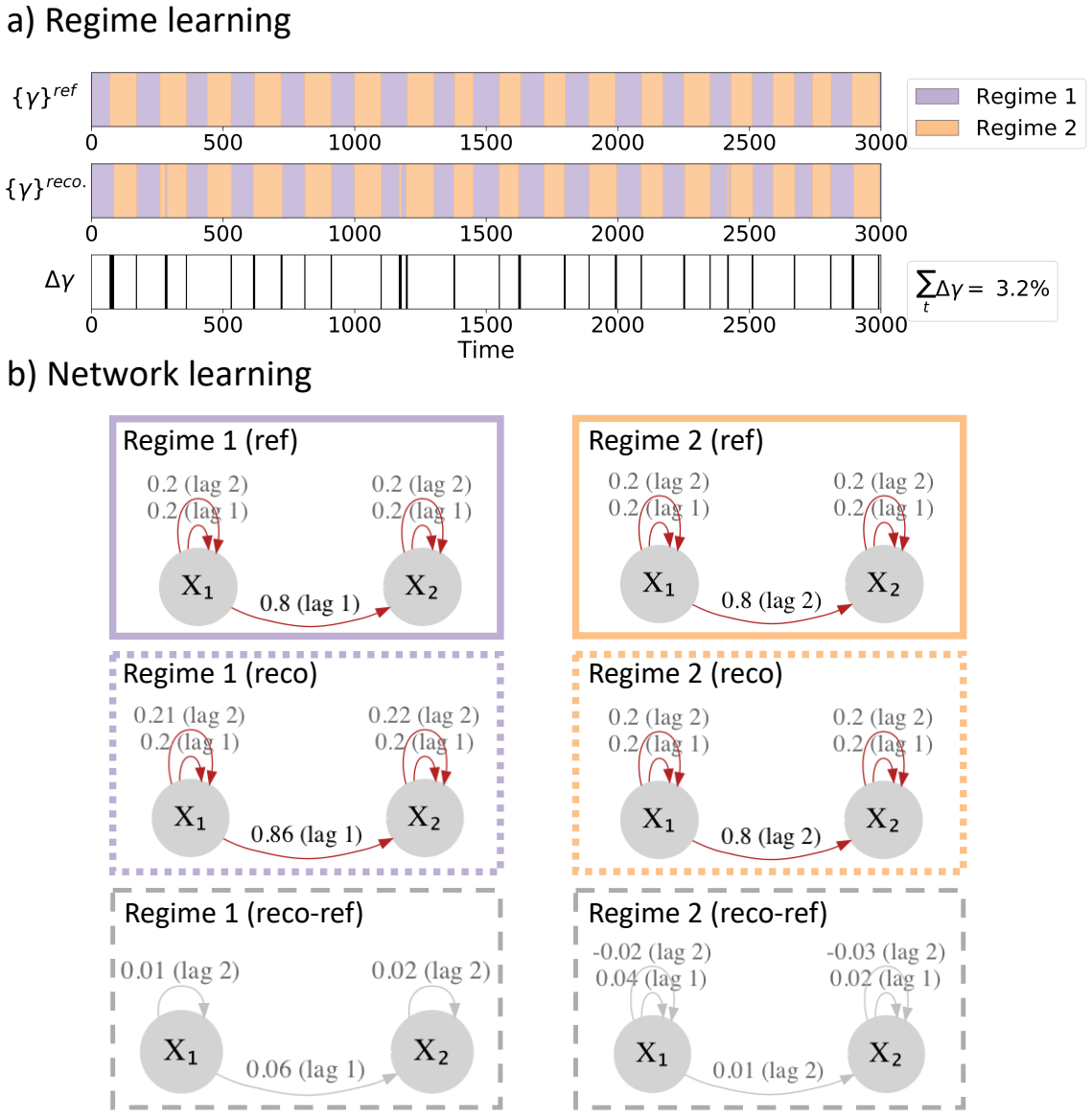

Figure 4.4.4: Example case *Lag*. See description in Figure 4.4.1.

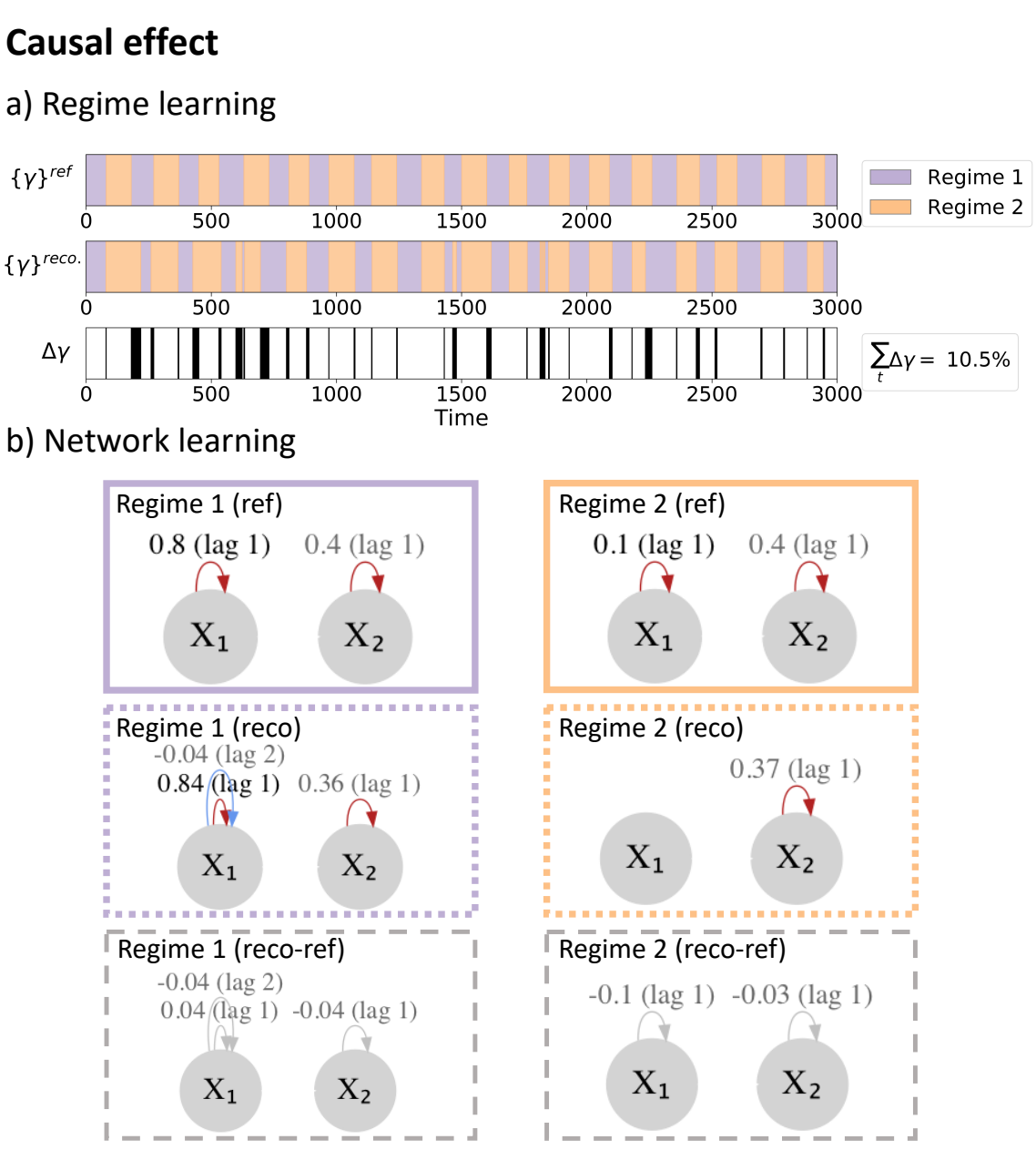

Figure 4.4.5: Example case *Causal effect*. See description in Figure 4.4.1.

process (light blue box plot, $\Delta\gamma\%$) and the precision of the reconstructed links' causal effects (pink box blot, $\Delta\Phi\%$). $\Delta\gamma\%$, is the average percentage of wrongly estimated time steps per regime (the lower the better, note that this value is the same for $k = 1, 2$, by construction). $\Delta\Phi\%$ is the average difference between the reconstructed linear coefficient and the reference values of the ground truth links expressed as a percentage, i.e. each difference is weighted by the inverse of the absolute value of the ground truth coefficient. The precise definition of the statistics can be found in Appendix 4.B. These quantities provide a summary of the reconstructed regimes' accuracy, previously shown in the bottom parts of Figs. 4.4.1-4.4.5, and are presented for all N_R runs. The examples are ranked in order of decreasing performance.

As already seen from inspection of single runs, change of *arrow direction* and of *causal effect* are the hardest to detect (highest error in Fig. 4.4.6). It is also clear that a worsening performance in regime detection results in higher network error. Except for *causal effect*, regime error $\Delta\gamma\%$ is between $1-7\%$ and network error $\Delta\Phi\%$ is between $1-10\%$. Note that the average $\Delta\Phi\%$ (dark pink cross) is very close to $\Delta\Phi^{\text{ref}}\%$ (black dot), the error for the PCMCI run with the ground-truth regime variable known (but causal structures unknown). This reference value sets the optimal baseline to which compare Regime-PCMCI performance, and is met by our algorithm in all but one example.

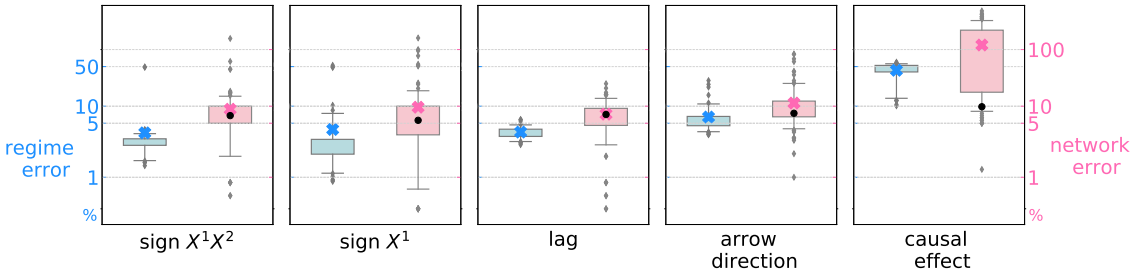

Figure 4.4.6: A summary of the general performance of the Regime-PCMCI for the examples described in Section 4.4.1. The performance skill is separated into *regime error* (defined as the average percentage of wrongly estimated time steps $\Delta\gamma\%$, light blue box plot) and *network error* (defined as relative percentage difference between reconstructed and reference links' coefficients $\Delta\Phi\%$, pink box plot). Box plots summarise the distribution of $N_R = 100$ runs (boxes between $25-75$ percentiles and whiskers between $5-95$ percentiles) and the mean is marked with a cross. The optimal baseline for the network fit, $\Delta\Phi^{\text{ref}}\%$, is marked with a black dot (see definition in the text). Note the symlog scale on the y-axes. Synthetic data sets are generated according to Table 4.4.1 for $N_R = 100$ random ground-truth regime-assigning processes.

Table 4.4.3 shows a more detailed summary of the results over the N_R realisations. The estimation errors are presented in terms of the regime assigning process (second column), the network structure (third to sixth column), the causal effects of links (seventh to tenth columns) and the overall reconstructed time series (last column). The second column is $\Delta\gamma\%$, the average percentage of wrongly estimated time steps per regime, introduced above. In terms of networks, the link detection performance is evaluated via the true positive (TPR) and false positive rates (FPR). Further, these are compared with the reference FPR and TPR (superscript *ref*) if PCMCI is run with the ground-truth regime variable known (but causal structure unknown). The accuracy in the links' causal effects is assessed via $\Delta\Phi$, the average difference between the reconstructed linear coefficient and the reference values of the ground truth links, and in the percentage version ($\Delta\Phi\%$, see above). The last column, $\hat{\varepsilon}$, is the expected prediction error per variable and per time step and is computed as $\hat{\varepsilon} = \sqrt{\mathbf{L}/(N_X T)}$ with $\mathbf{L}$ defined in Eq. (4.4) and with N_X and T referring to the number of variables, here two, and the length of the time series respectively. The precise definition of all the above statistics can be found in Appendix 4.B.

Example	$\Delta\gamma\%$	TPR_{all}	$\text{TPR}^{\text{ref}}_{\text{all}}$	FPR_{all}	$\text{FPR}^{\text{ref}}_{\text{all}}$	$\Delta\Phi$	$\Delta\Phi^{\text{ref}}$	$\Delta\Phi$ %	$\Delta\Phi^{\text{ref}}$ %	$\hat{\varepsilon}$
arrow direction	3.0	1.0	1.0	0.02	0.01	0.021	0.020	7.0	7.0	0.76
causal effect	43.0	0.81	0.98	0.11	0.01	0.286	0.020	120.0	10.0	0.68
lag	6.0	0.98	1.0	0.04	0.01	0.027	0.018	11.0	8.0	0.68
sign X^1	4.0	0.98	1.0	0.03	0.01	0.033	0.016	10.0	6.0	0.65
sign X^1X^2	3.0	0.99	1.0	0.01	0.01	0.028	0.019	9.0	7.0	0.75

Table 4.4.3: Results for $N_K = 2$ experiments averaged over $N_R = 100$ realisations generated for each example described in Table 4.4.1.

In summary, Table 4.4.3 shows that :

- $\Delta\gamma\%$: on average, the regime assigning process is reconstructed correctly in $\sim 94\%$ of the time steps for all cases except *causal effect*. The *causal effect* and *lag* examples are the hardest to infer, with causal effect being particularly deficient. In these examples a mixed-regime state (e.g., arising from assigning a considerable fraction of wrong time steps to a regime) is still quite close to any of the true regimes. Therefore the algorithm struggles to decide which time steps belong to which regime, since they could fit both to some degree. Yet,

also for *causal effect* there are 7 instances where $\Delta\gamma < 15\%$ (one presented in Figure 4.4.5) and those, as expected from PCMCI, give a very good network fit. Note that these runs do not correspond to the lowest objective values of the N_R set (i.e. better fit) which shows that runs that end up in mixed states can still fit the data quite well. Also, note that the causal effect setup reaches local minima in 16% of the 100 runs, thus in 84% of the runs the algorithm cannot easily find a stable solution, which points to a weaker confidence in the output.

- TPR: despite some errors in reconstructing the regime assigning process, the TPR is always very close to 1. This indicates that the true signals, dynamic wise, are strong enough to be detectable.

- FPR: Ideally the false positive rate should be upper-bounded by $\alpha = 0.01$. This is also the case if one assumes the correct regimes (see column FPR^{ref}). However, if the regimes are learned, in most of the examples the FPR value is higher due to errors in learning the regimes. If a wrong regime is learned, then both false positives and false negatives can occur. False negatives, i.e., missing links in the PC_1 step of PCMCI, can lead to false positives in the MCI step.

- $\Delta\Phi\%$: Errors in parents' detection (either due to false positives (FPR) or to false negatives (missed links, FNR = 1-TPR)) surely impact the estimation of link effects. Since the TPR and FPR are good, except for the causal effects case, the expectation is to obtain also good results for the linear coefficients. This is indeed the case, as the difference is of order 10^{-2} implying a relative error of about 10%. Also this matches very closely the optimal baseline for the network fit $\Delta\Phi^{\text{ref}}\%$.

4.4.2 Low dimensional data with three underlying regimes

To illustrate how Regime-PCMCI deals with more than two regimes, a toy time series is also considered based on three different causal regimes. It is of course possible to consider the case $N_K > 3$, yet in applications it is often desirable to infer a few prominent and relevant regimes rather than having too many that are not interpretable anymore. In other words, the aim is to avoid overfitting and to increase the information gain by reducing the complexity of the assumed model

(parsimony).

The artificial time series is generated via a regime dependent causal graph that is designed by combining two of the regime settings presented in Section 4.4.1, namely *sign* X^1X^2 change and *arrow inversion* (for details see Table 4.4.4). The regime assigning reference process $\{\Gamma\}^{\text{ref}}$ is generated by randomly choosing between different persistence lengths of 60, 70 and 80 time-steps and iterating for 20 times. The algorithm is run with free parameters in Table 4.4.5.

example	$k=1$	$k=2$	$k=3$	$\Phi_1^j(i,\tau)^{\text{ref}}$	$\{\Phi_2^j(i,\tau)\}^{\text{ref}}$	$\{\Phi_3^j(i,\tau)\}^{\text{ref}}$
sign X^1X^2	$X^1 \xrightarrow{\lvert a\rvert} X^2$	$X^1 \xrightarrow{-\lvert a\rvert} X^2$	$X^2 \xrightarrow{\lvert a\rvert} X^1$	$\{\Phi_1^2(1,1)\}^{\text{ref}} = 0.8$	$\{\Phi_2^2(1,1)\}^{\text{ref}} = -0.8$	$\{\Phi_3^1(2,1)\}^{\text{ref}} = 0.8$
and *arrow*				$\{\Phi_1^1(1,1)\}^{\text{ref}} = 0.2$	$\{\Phi_2^1(1,1)\}^{\text{ref}} = 0.2$	$\{\Phi_3^1(1,1)\}^{\text{ref}} = 0.2$
direction				$\{\Phi_1^2(2,1)\}^{\text{ref}} = 0.2$	$\{\Phi_2^2(2,1)\}^{\text{ref}} = 0.2$	$\{\Phi_3^2(2,1)\}^{\text{ref}} = 0.2$

Table 4.4.4: Artificial model configuration for a low dimensional example with $N_K = 3$ underlying regimes.

Figure 4.4.7 shows the results. There are only minimal deviations from the true reference values, which confirms that the proposed method is capable to deal with $N_K > 2$. This also holds for the summary results over $N_R = 100$ runs presented in Table 4.4.6. Yet, it is important to note that a combination of causal graphs is chosen that performed well for $N_K = 2$, i.e, causal effect changes would also be difficult to detect for $N_K = 3$.

4.4.3 Regime parameter selection

Parameter selection, i.e. the number of regimes, is here explored by means of the AICc scores defined in (Eq. (4.16)). Two test scenarios are investigated, $\{N_K\}^{\text{ref}} = 2,3$ for a selection of the examples defined in the previous Sections 4.4.1 and 4.4.2. The PCMCI parameters are as in those sections while $N_R = 29$, $N_Q = 20$ and $N_A = 20$. The resulting AICc values are displayed in Figure 4.4.8. The N_C value is changed adaptively for each N_K to ensure a similar N_M value for the

CI test	$\tau_{\max}$	α	α_{PC}	N_K	N_C	N_Q	N_A
ParCorr	3	0.01	0.2	3	40	20	50

Table 4.4.5: Method parameters for a low dimensional example with $N_K = 3$ underlying regimes.

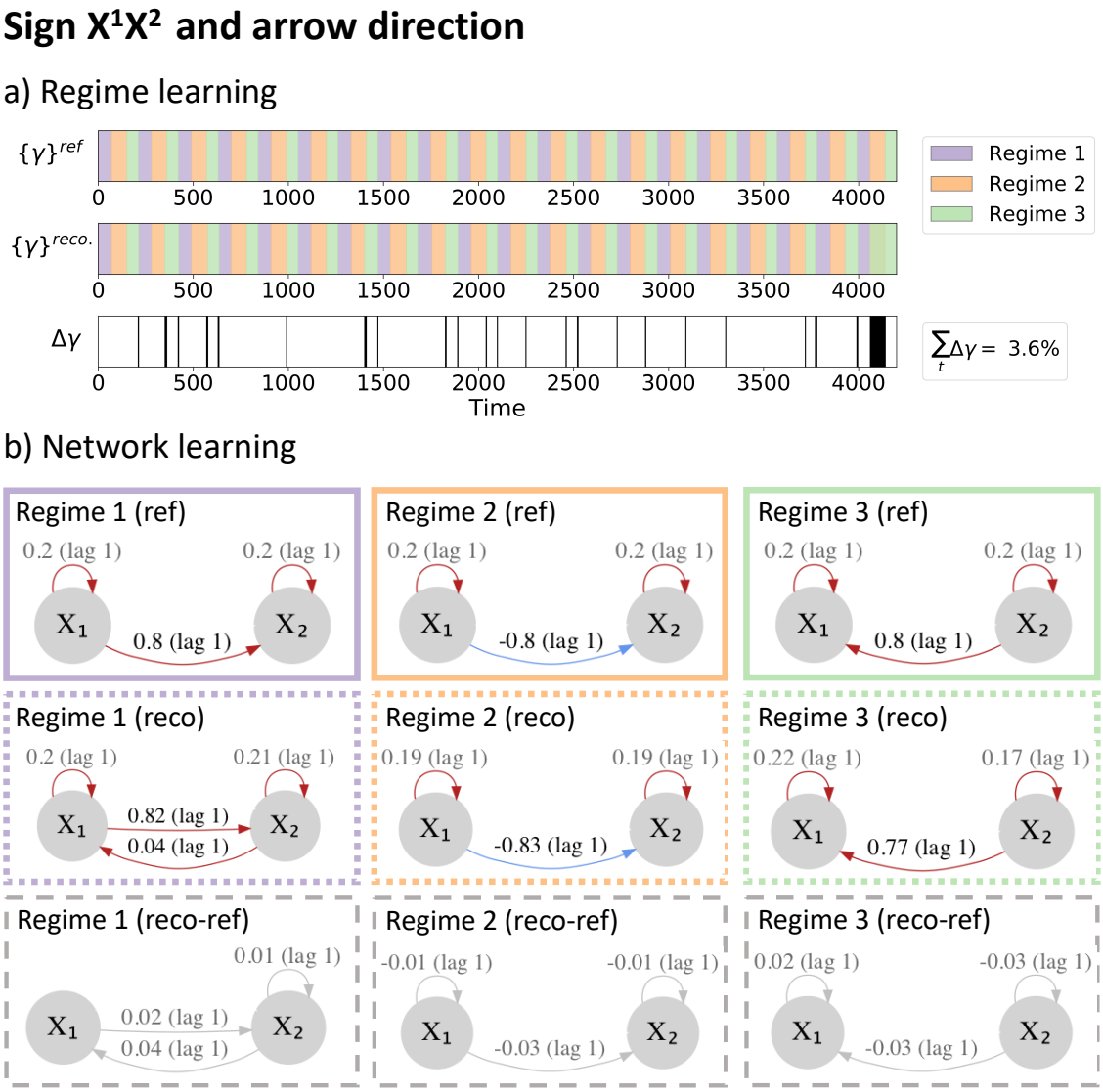

Figure 4.4.7: Example with $N_k = 3$ regimes for the case *Sign X^1X^2 and arrow direction*. See description in Figure 4.4.1 but with three regimes.

$\Delta\gamma\%$	TPR_{all}	$\text{TPR}_{\text{all}}^{\text{ref}}$	FPR_{all}	$\text{FPR}_{\text{all}}^{\text{ref}}$	$\Delta\Phi$	$\Delta\Phi^{\text{ref}}$	$\Delta\Phi$ %	$\Delta\Phi^{\text{ref}}$ %	$\hat{\varepsilon}$
4.0	0.98	1.0	0.05	0.01	0.033	0.020	10.0	7.0	0.5

Table 4.4.6: Results for $N_K = 3$ experiments averaged over $N_R = 100$ realisations generated for each example described in Table 4.4.4.

different number of regimes, i.e.,

$$N_C(N_K) = \{N_C^{\text{ref}}\}\{N_K\}^{\text{ref}}/N_K \tag{4.18}$$

for $N_K > \{N_K\}^{\text{ref}}$. The choice of N_C is based on N_M which in real life applications can be chosen according to the considered processes and data as a good estimate of the time scale of regime changes is often available. Nevertheless it is also possible to chose N_C via an Information criterion simultaneously with N_K (e.g., in the context of regime-dependent clustering in Falkena et al., 2020 or regime-dependent Markov regression in Wiljes et al., 2014). The reference value for the number of switches is on average (due to randomisation of $\{\Gamma\}^{\text{ref}}$) $\{N_C^{\text{ref}}\} = 40$ for both $\{N_K\}^{\text{ref}} = 2, 3$.

Note that the lowest N_K at which the AICc plateaus is the ground-truth one. The plateau itself occurs due to the fact that only the links with non-zero causal effect values are counted towards the number of parameters. Thus a higher number of regimes N_K does not necessarily result in an increase of the total number of parameters. In other words the penalisation is not becoming stronger with higher values of N_K. Concluding, it is clearly visible that no significant improvement is gained by increasing N_K beyond the reference number of regimes. Since the entry point to the plateau reveals the reference number of regimes, it seems possible to use AICc to address scenarios where the true number of regimes is unknown.

4.4.4 High dimensional linear network

In this section the algorithm is evaluated on high-dimensional datasets, with each dataset consisting of $N_X = 10$ interacting variables. The background regimes are generated with two regular alternating regimes of 300 time steps each, for a total length $T = 15,000$. The network structures are randomly generated from a family of linear networks defined via the parameters shown in Table 4.4.7, where L is the number of randomly drawn cross-variable links with random coefficients from the third column. Note that each variable is also auto-linked at lag 1 with a coefficient

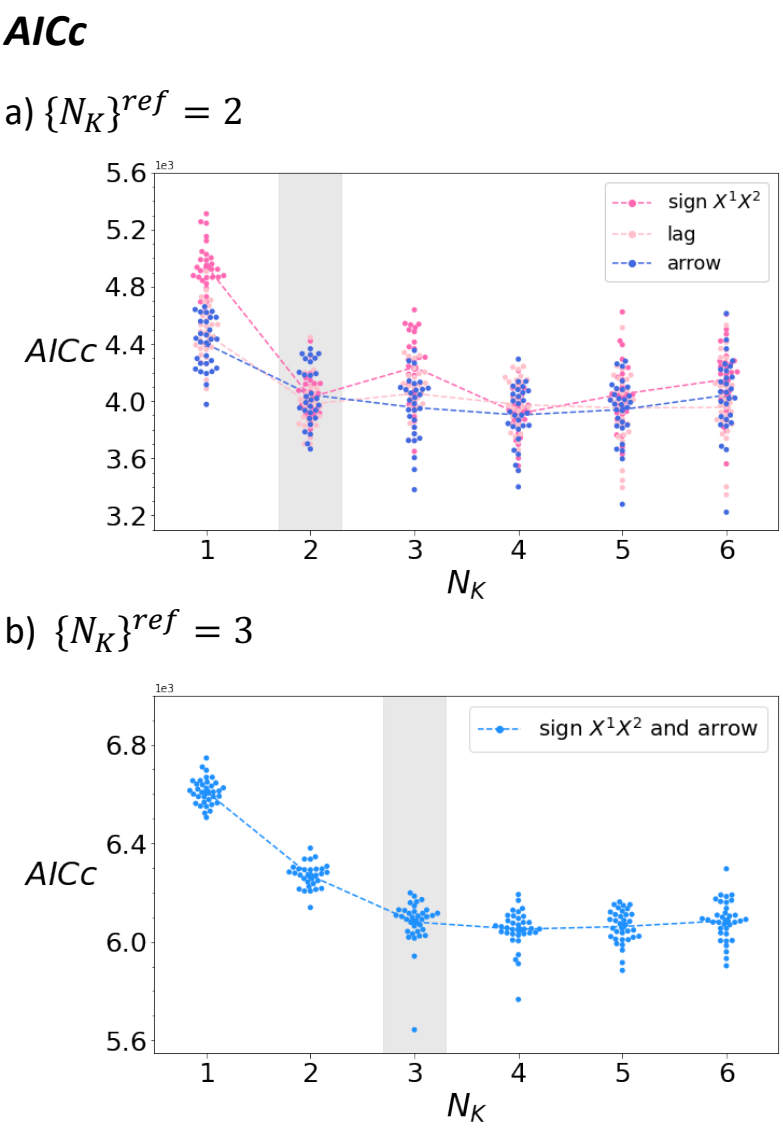

Figure 4.4.8: Numerical investigation of AICc values for runs with different N_K and (a) $\{N_K\}^{\text{ref}} = 2$ for three network examples (*sign* X^1X^2, *arrow* and *lag* change) and (b) $\{N_K\}^{\text{ref}} = 3$ for the *sign* X^1X^2 *and arrow* change examples. In each example, individual dots represent the value attained by the $N_R = 29$ runs, and the dashed line goes through the mean value of each set. The vertical grey bar highlights the ground-truth number of regimes $\{N_K\}^{\text{ref}}$.

randomly drawn from the fourth column. The time series $\mathbf{x}_t \in \mathrm{R}^{10}$ are generated with model Eq. (4.17) and for $N_R = 70$ realisations. Regime-PCMCI is then run with the settings shown in Table 4.4.8.

N_X	L	$\Phi_k^j(i,\tau)$	$\Phi_k^i(i,\tau)$	max lag
10	30	[-0.4, 0.4]	[0.2, 0.5, 0.9]	3

Table 4.4.7: High dimensional network parameters

CI test	$\tau_{\max}$	α	α_{PC}	N_K	N_C	N_Q	N_A
ParCorr	4	0.05	0.2	2	49	30	50

Table 4.4.8: Method parameters for high dimensional experiments with two underlying regimes.

The results are shown in Table 4.4.9, which is structured like Table 4.4.3 except for TPR and FPR being estimated for the cross-variable links thus focusing on the connections between variables. All links are considered in $\Delta\Phi$. Regime-PCMCI performs very well even in this challenging setting. Notably, individual runs can perform extremely well, with $\Delta\gamma$ reaching as low as 0.02%, and a total of 53 runs below total average of $\Delta\gamma = 11.7\%$ (second row in Table 4.4.9). The other 7 runs are responsible for most of the deviation of the average statistics from the reference values (first row).

As in the *causal effect* case, there is a mismatch between runs with the lowest prediction errors $\hat{\varepsilon}$ and the lowest error on the regime-assigning process $\Delta\gamma$, meaning that a filtering on $\hat{\varepsilon}$ cannot be used to find the best performing runs. This behaviour can be explained from the tendency of the algorithm to over-fit when too many degrees of freedom are available, as well as from the complexity of distinguishing different causal effects (a challenge already manifested in the *causal effect* case).

Selection	$\Delta\gamma$%	$\mathrm{TPR}_{\mathrm{cros}}$	$\mathrm{TPR}_{\mathrm{cros}}^{\mathrm{ref}}$	$\mathrm{FPR}_{\mathrm{cros}}$	$\mathrm{FPR}_{\mathrm{cros}}^{\mathrm{ref}}$	$\Delta\Phi$	$\Delta\Phi^{\mathrm{ref}}$	$\Delta\Phi$ %	$\Delta\Phi^{\mathrm{ref}}$ %	$\hat{\varepsilon}$	n. runs
all	11.7	0.94	1.0	0.18	0.08	0.059	0.005	16.0	1.5	0.85	70
$\Delta\gamma < 11.7$ %	0.19	1.0	1.0	0.08	0.07	0.006	0.005	1.8	1.5	0.70	53

Table 4.4.9: Results for high-dimensional experiments over $N_R = 70$ realisations generated for each example described in Table 4.4.7.

4.4.5 Computational complexity

Table 4.4.10 shows some indicators of performance of the method: the fraction of N_R runs that correspond to a (local) minimum, the average number of q-iterations needed to reach a local minima and the runtime for the whole N_R set of runs (the code run parallel over the N_A annealings and using 4 to 6 CPUs per job).

Most of the examples reach local minima in more than 50% of the N_R runs, while the percentage is very low for *causal effect* (second column). Note that examples with a high percentage of local minima correspond also to quick convergence in terms of iteration steps (third column). They are also associated with better regime reconstruction (see Tables 4.4.3, 4.4.6, 4.4.9), confirming that a clear cost functional minimum (as shown from the second and third column) is linked to better detection. Finally, the runtime is quite fast: the low dimensional examples take between 10 and 20 minutes for $N_K = 2$ and 45 minutes for $N_K = 3$ to complete 100 runs. The high dimensional example takes just below 3 hours for 70 runs.

Example	n. local minima/N_R %	iterations to minima	runtime (s)
arrow direction	92 %	7	600
causal effect	16 %	13	970
lag	60 %	11	1,130
sign X^1	52%	12	970
sign X^1X^2	70 %	9	700
sign X^1X^2 and *arrow*	56%	10	2,670
high dimensional	92%	6	10,780

Table 4.4.10: Summary performance statistics of all examples. The third column is the average value over the respective N_R.

4.5 Results from a real-world example: the effect of El Niño Southern Oscillation on Indian rainfall

Finally, the performance of Regime-PCMCI is tested on real-world data, to address the non-stationary relationship of El Niño Southern Oscillation (ENSO) and all-India rainfall (AIR) mentioned in the introduction. For given time series of ENSO and AIR, the method is tested in its ability to distinguish between the winter and summer months, i.e. the background-regimes, and to detect a reported link from ENSO to AIR during summer.

This example can be considered a difficult case since the expected signal from ENSO to AIR is likely small compared to natural variability (Webster and Palmer, 1997). The warming in the tropical Pacific Ocean because of El Niño weakens the southeast trade winds flowing to the intertropical convergence zone over India. Since these winds are the main driving force of the Indian summer monsoon, El Niño events are associated with weak monsoons and lower than average rainfall. The La Niña has the opposite effect to the El Niño and is responsible for stronger monsoons and above-average rainfall. Further, climate data is typically very noisy with causal relationships being diluted by other, often unknown processes given a complex coupled climate system (Williams et al., 2017).

The input data consist of monthly observations of ENSO and AIR, for the years 1871 to 2016, resulting in two time series consisting of 1740 monthly values each. More precisely, ENSO is represented by the so-called relative Nino3.4 index provided by the National Oceanic and Atmospheric Administration (NOAA) (Boyin et al., 2017)[1]. Data for AIR anomalies (with the climatology subtracted) are provided by the Indian Institute of Tropical Meteorology (IITM) (Indian Institute of Tropical Meteorology, 2016)[2].

The following parameters for Regime-PCMCI are chosen. For the regime part, $N_K = 2$ and $N_C = 292$ are set, which is equivalent to assuming two seasons per year. For the PCMCI settings, a significance level $\alpha = 0.01$ ($\alpha_{\text{PC}} = 0.2$) is chosen. Further, a maximum time-lag of two months is assumed, i.e. $\tau_{\max} = 2$. The optimisation is run $N_A = 100$ annealing times, to span many local

[1] https://climexp.knmi.nl/getindices.cgi?WMO=NCDCData/ersst_nino3.4a_rel&STATION=NINO3.4_rel&TYPE=i&id=someone@somewhere [last accessed 25/09/2022]

[2] https://climexp.knmi.nl/getindices.cgi?WMO=IITMData/ALLIN&STATION=All-India_Rainfall&TYPE=p&id=someone@somewhere [last accessed 25/09/2022]

minima, with each annealing allowed for up to $N_Q = 100$ iteration steps to converge.

Among the annealing steps, which correspond to different random initial guesses on the regime-assigning process Γ, some clearly performed better in terms of fitting the data. The average prediction error $\hat{\varepsilon}$ associated with each annealing is estimated (Eq. (4.19)) and it is shown in Figure 4.5.1*a* for all the annealings (ranked). A red box highlights the top performing cluster (13 runs). All of the top 13 annealings find a link from ENSO to AIR during one of their two regimes only (for simplicity hereafter called regime 1). In the following the results averaged over these annealings are presented, and links are shown if surpass a strength of 0.1.

The causal link from ENSO to AIR in regime 1 has an average standardized linear effect of -0.4, meaning that a one standard deviation increase in ENSO results in a reduction of 0.4 standard deviations in AIR (Figure 4.5.1*c*) . This negative dependence is well documented in the literature (Webster and Palmer, 1997). During regime 2, in contrast, ENSO and AIR are, on average, almost independent, with only a very weak link (-0.05, not shown) detected from AIR to ENSO. More importantly, our results indicate a clear seasonal dependence. Figure 4.5.1*d* shows the number of months assigned to each regime (normalised by the number one would expect on the hypothesis of no seasonality, see figure caption). A clear peak in summer months is found for regime 1. More precisely, most of the months between June to September are assigned to regime 1 (70%). These are the months in which the Indian summer Monsoon is active and for which a robust influence from ENSO has been shown. In contrast, months assigned to regime 2 are predominantly winter months (60% of all December to March months). Thus, despite the relatively weak mean causal effect of ENSO on AIR during summer, and the large inter-annual variability, our algorithm successfully reconstructed this well-documented relationship given all-year time series of ENSO and AIR.

A method to detect long-term changes of this summer teleconnection has recently been proposed by Bódai et al. (2020) using ensembles of climate models. The additional dimension provided by the ensemble members allows to compute year-dependent correlations to infer inter-annual changes. In contrast, our method uses a single realisation of the dynamics (e.g. observations) to still obtain time-dependent statistics (networks), although in a finite number (N_K). Note that long-term changes may still be detectable with Regime-PCMCI either as long term changes to the persistence and (or) the start of a regime each year or with the emergence of a regime in a specific

time period.

Overall, these results are promising and show the potential of Regime-PCMCI to detect regime-dependent causal structures in a system as complex as the climate system. On the other hand, it also shows that domain knowledge is required to assure a suitable choice of parameters (N_C and N_K) and interpretation of the results. This is a common caveat to many data-driven approaches, which nevertheless needs to be stressed.

4.6 Conclusions

Causal discovery is emerging as an important framework across many disciplines in science and engineering, but each discipline has particular challenges that novel methods need to address (Runge et al., 2014). This chapter introduced a method called Regime-PCMCI to learn regime-dependent causal relations, overcoming one of the key drawbacks of current causal discovery methods. The performance of Regime-PCMCI was analysed for many different artificially generated causal scenarios and for varying regimes showing that the method covers a wide range of settings (Figures 4.4.1-4.4.4,4.4.6 and Table 4.4.3). The performance of the algorithm is maintained also for high-dimensional settings with 10 variables (Table 4.4.9) as well as for more than two regimes (Figure 4.4.7 and Table 4.4.6). Limitations of the method were found for the case where only the causal effect strength of a link changes between regimes (Figure 4.4.5), which seems to be hard to detect with our optimisation scheme and requires further investigation. Further, the capability of Regime-PCMCI was verified by means of a well documented climate example using real data of ENSO and Indian rainfall (Figure 4.5.1). Overall, the proposed method presents itself as a promising approach in the context of non-stationary causal links manifested in regime changes in time.

Note that a causal interpretation of estimated links in our observational causal discovery framework still assumes causal sufficiency, that is, no unobserved common causes. However, estimated non-causality (zero coefficients) do not require this assumption (Runge, 2018) and can be interpreted as an absence of a causal relation already under the weaker Faithfulness assumption (Runge et al., 2019a). While for PCMCI asymptotic consistency was shown (Runge et al., 2019a), this is a more difficult task for Regime-PCMCI and deferred to further research.

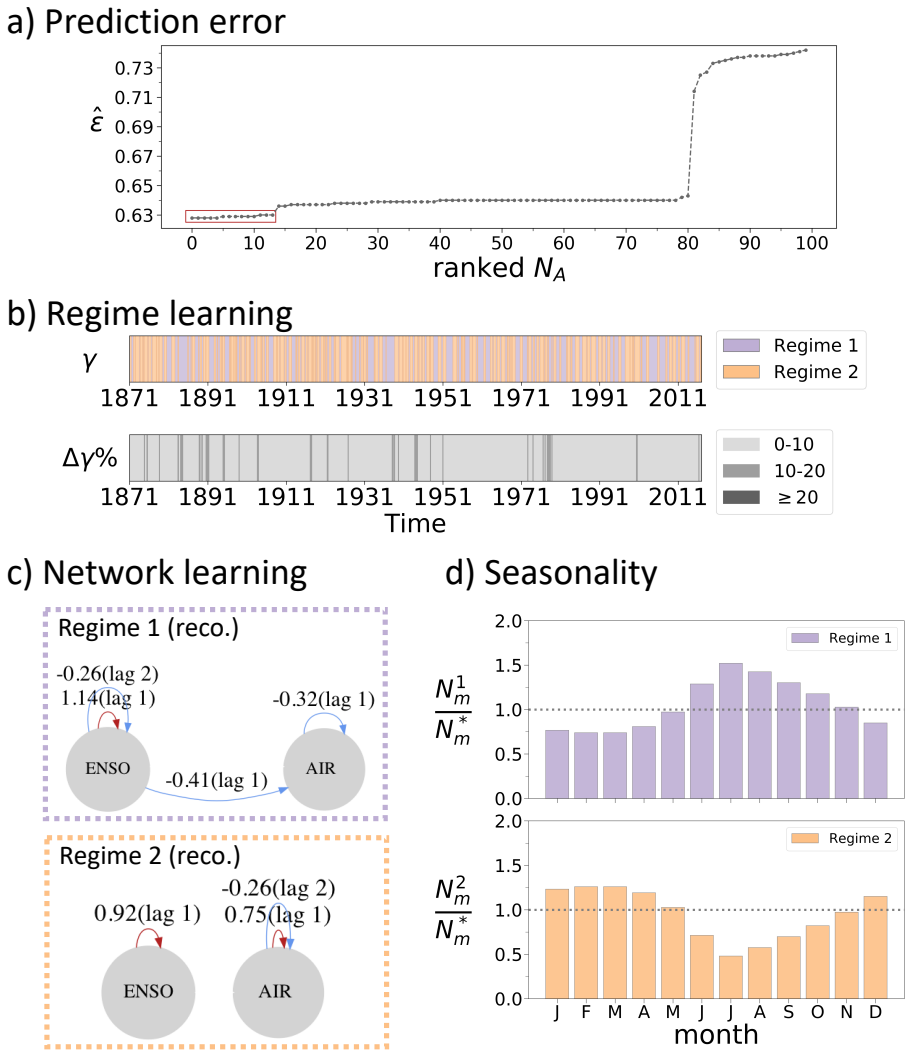

Figure 4.5.1: Climate example. (a) Prediction error for each annealing step in ascending order, lowest 13 annealings highlighted in red box. All the other panels refer to this selection. (b) Regime learning: regime-assigning process corresponding to the best annealing (rank 0) (top) and departure from this estimate of the remaining best 12 annealings (in percentage difference). (c) Network learning: mean networks per regime, each causal effect is the mean of the corresponding coefficient in the individual 13 annealings. (d) Seasonality of the regimes: Number of years per month m assigned to each regime (N_m^k), normalised by N_m^*, which refers to the expected number of months assigned to a given regime if one assumes equal probability $1/N_K$ of assigning a month to one of the two regimes. Thus, here, $N_m^* = 13 \cdot T/(12 * N_K)$).

APPENDIX

4.A Heterogeneous noise

In the general framework laid out in Equation (4.1) the noise variables η_t^j are only assumed to be jointly independent and stationary, each distributed according to a distribution $\mathcal{D}^j$. Given the primary focus of this work is to detect regime-dependent causal structures rather than noise structures, the effective choice for noise distributions used to generate the data is a Gaussian with unit variance $\eta_t^j \sim \mathcal{N}(0,1)$ for all variables j (Section 4.4).

Yet this simplification does not necessarily represent the variability of processes in real word scenarios. Here the performance of the proposed Regime-PCMCI is exemplified for Gaussian noises with variable-specific variances $\eta_t^j \sim \mathcal{N}(0,\sigma_j^2)$ (see Section 4.A.1) and noises from two different distributions, Gaussian and uniform (see Section 4.A.2).

4.A.1 Gaussian noise with variable-specific variances

The data is generated from model Eq. (4.17) with example *sign change* X^1X^2 coefficients. The noise terms η_j are Gaussian distributed with a fixed variance for variable X_1, $\mathcal{D}^1 = \mathcal{N}(0,1)$, and three different cases for variable X_2, $\mathcal{D}^2 = \mathcal{N}(0,\sigma_2^2)$ with $\sigma_2 = 0.25$, $\sigma_2 = 0.5$, and $\sigma_2 = 2.0$. The Regime-PCMCI results, averaged over 100 different realisations of the regime-assigning processes, are presented in Table 4.A.1. The algorithm performs very well in the first two cases (average regime detection error $\Delta\gamma \leq 1\%$). This is to be expected since a smaller noise in X^2 allows for a better fit of the data. The latter case is harder to infer, since the noise on X^2 is very large compared to the deterministic signal (here $\Delta\gamma \simeq 25\%$).

Case	$\Delta\gamma\%$	TPR_{all}	$\text{TPR}_{\text{all}}^{\text{ref}}$	FPR_{all}	$\text{FPR}_{\text{all}}^{\text{ref}}$	$\Delta\Phi$	$\Delta\Phi^{\text{ref}}$	$\Delta\Phi$ %	$\Delta\Phi^{\text{ref}}$ %
$\sigma_2 = 0.25$	0.3%	1.0	1.0	0.01	0.01	0.010	0.010	5.2%	5.2%
$\sigma_2 = 0.5$	0.8%	1.0	1.0	0.02	0.01	0.013	0.013	5.3%	5.2%
$\sigma_2 = 2.0$	24.0%	0.85	1.0	0.02	0.01	0.16	0.03	36.0%	9.0%

Table 4.A.1: Results for example *sign* X_1X_2 averaged over $N_R = 100$ realisations, for each noise variances' combination described in Section 4.A.1.

4.A.2 Different noise distributions

The data is generated from model Eq. (4.17) with example *sign change* X^1X^2 coefficients. The noise terms η_j are set to follow completely different distributions: variable X^1 is associated with a unit variance Gaussian noise, $\mathcal{D}^1 = \mathcal{N}(0,1)$, and variable X^2 with uniformly distributed noise between ± 1.5, $\mathcal{D}^2 = \mathcal{U}(-1.5, 1.5)$. The Regime-PCMCI results, averaged over 100 different realisations of the regime-assigning processes, are presented in Table 4.A.2. This scenario gives results comparable to the ones presented in the manuscript for the same example, i.e. $\Delta\gamma \simeq 3\%$.

Case	$\Delta\gamma\%$	TPR_{all}	$\text{TPR}_{\text{all}}^{\text{ref}}$	FPR_{all}	$\text{FPR}_{\text{all}}^{\text{ref}}$	$\Delta\Phi$	$\Delta\Phi^{\text{ref}}$	$\Delta\Phi$ %	$\Delta\Phi^{\text{ref}}$ %
Gauss, Unif	3%	1.0	1.0	0.01	0.01	0.025	0.019	8.0%	7.0%

Table 4.A.2: Results for example *sign* X_1X_2 averaged over $N_R = 100$ realisations, for different noise distributions described in Section 4.A.2.

To summarise, the results show that Regime-PCMCI can deal with specific heterogeneous noise distributions, even belonging to different families of distributions. Since the optimisation method acts on regression residuals, we can speculate that we expect good performance as long as the noise terms are not too large in their magnitude and are not too skewed. An elaborate study of these conclusions and an investigation of the potential for generalisation of the method to more extreme noise distributions is an interesting research aspect for the future.

4.B Definition of result statistics

The definitions for the statistics presented in Tables 4.4.3, 4.4.6 and 4.4.9 is outlined in the following.

4.B.1 Regime assigning process

$$\Delta\gamma(\%) = \frac{\sum_{t=\tau_{\max}} |\{\gamma_k(t)\}^{reco.} - \{\gamma_k(t)\}^{ref}|}{T - \tau_{max}} \times 100$$

4.B.2 Link detection

TPR

$$\mathrm{TPR} = \frac{\mathrm{TP_X}}{\mathrm{P_X}}$$

Over the cross-inks (in Table 4.4.9):

$$\mathrm{TP_{cros}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{reco.} \neq 0 \ \& \ \{\Phi_k^j(i,\tau)\}^{ref} \neq 0 \ \& \ i \neq j\}|$$

$$\mathrm{P_{cros}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{ref} \neq 0 \ \& \ i \neq j\}|$$

And over all links (in Tables 4.4.3 and 4.4.6):

$$\mathrm{TP_{all}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{reco.} \neq 0 \ \& \ \{\Phi_k^j(i,\tau)\}^{ref} \neq 0\}|$$

$$\mathrm{P_{all}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{ref} \neq 0\}|$$

FPR

$$\mathrm{FPR} = \frac{\mathrm{FP_X}}{\mathrm{N_X}}$$

Over the cross-links (in Table 4.4.9):

$$\mathrm{FP_{cros}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{reco.} \neq 0 \ \& \ \{\Phi_k^j(i,\tau)\}^{ref} = 0 \ \& \ i \neq j\}|$$

$$\mathrm{N_{cros}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{ref} = 0 \ \ \& \ \ i \neq j\}|$$

And over all links (in Tables 4.4.3 and 4.4.6):

$$\mathrm{FP_{all}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{reco.} \neq 0 \ \ \& \ \ \{\Phi_k^j(i,\tau)\}^{ref} = 0\}|$$

$$\mathrm{N_{all}} = |\{(i,j,\tau) : \{\Phi_k^j(i,\tau)\}^{ref} = 0\}|$$

4.B.3 Link coefficients

$$\Delta\Phi = \frac{1}{N_K}\sum_{k=1}^{N_K} \frac{\sum_j \sum_{X_{t-\tau}^i \in \mathcal{P}_k^j} | \{\Phi_k^j(i,\tau)\}^{reco.} - \{\Phi_k^j(i,\tau)\}^{ref} |}{\sum_j |\mathcal{P}_k^j|}$$

Can be also computed as average *percentage* error per regime:

$$\Delta\Phi(\%) = \frac{1}{N_K}\sum_{k=1}^{N_K} \frac{\sum_j \sum_{X_{t-\tau}^i \in \mathcal{P}_k^j} \frac{|\{\Phi_k^j(i,\tau)\}^{reco.} - \{\Phi_k^j(i,\tau)\}^{ref}|}{\{\Phi_k^j(i,\tau)\}^{ref}}}{\sum_j |\mathcal{P}_k^j|} \times 100$$

4.B.4 Prediction error

$$\hat{\varepsilon} \equiv \frac{1}{N_X T}\sum_t \sum_j |\{x^j(t)\}^{ref} - \{x^j(t)\}^{reco.}| \approx \sqrt{\frac{\mathbf{L}}{N_X \cdot T}} \tag{4.19}$$

with $\mathbf{L}$ defined in Eq. (4.4)).

4.C Abbreviations

Abbreviations	
AIC	Akaike Information criterion
AICc	Corrected Akaike Information criterion
ENSO	El Niño Southern Oscillation
FPR	False positive rate
MCI	Momentary conditional independence
PCMCI	Causal discovery method (Runge et al., 2019a)
RAM	Regime-dependent Autoregressive Model
SCM	Structural causal model
TPR	True positive rate

Table 4.C.1: Abbreviations used throughout the chapter.

Chapter 5

Bayesian causal network modelling of Southern Hemisphere large-scale circulation variability from late winter to early summer

5.1 Introduction

In Chapter 3 a deterministic time-series causal network of the intraseasonal S-T coupling is analysed. But the predictability of the timing of the coupling and of the effects of the eddy-driven jet on surface climate were not addressed. In this chapter, a probabilistic causal network is developed using hindcast data to study the long-lead predictability of the S-T coupled variability.

5.1.1 Stratospheric and tropospheric drivers of Southern Hemisphere large-scale circulation variability from late winter to early summer

The mid-latitude jet stream is a dominant feature of the Southern Hemisphere (SH) large-scale circulation variability. The jet is present almost uninterrupted in all seasons with the position of the strongest winds located somewhere between 50 and 65ºS. Via its association with storm tracks and low pressure systems, the jet stream is strongly linked to the surface climate in many regions

of the SH extra-tropics and high latitudes, including Australia, South American and Antarctica. For example, in summer the Australian east coast is cooler and wetter (hotter and drier) for a more poleward (equatorward) jet due to its contrasting effect against the moist air flowing westward from the ocean.

The main feature of the SH jet stream dynamics is a twice-yearly vacillation in latitude and strength with a contraction (and strengthening) and expansion (and weakening) of the storm track, reflected in the semi-annual oscillation (SAO) in sea-level pressure (Bracegirdle, 2011) (see also Section 1.3.1). Variability and trends of the jet strongly project onto the Southern Annular Mode (SAM). Generally speaking, more positive indices of the SAM are associated with both poleward shifts and strengthening of the jet, and viceversa for negative SAM, although the correspondence is not a one-to-one (Hartmann and Lo, 1998; Thompson and Wallace, 2000). Due to its presence across all longitudes from late winter through early summer and its equivalent barotropic structure, the SH mid-latitude jet can be viewed as a zonally-symmetric and height averaged wind, frequently referred to as the eddy-driven jet (EDJ) (Section 1.3.1).

As part of the SAO, the EDJ undergoes a latitudinal migration around its winter position, just south of 50ºS, between late winter and early summer. The EDJ first moves poleward between September and October (herein called phase 1, EDJ-1) and then equatorward between November and December (phase 2, EDJ-2) with a larger amplitude, as shown in Figure 5.1.1 (left panel). After mid January, the EDJ migrates poleward again but this second poleward shift is not considered in the present analysis. Year-to-year differences in both timing and amplitude of these shifts are seen in the large interannual variability: the EDJ standard deviation is about 4-6 degrees latitude between August and December, which decreases to about 3-4 degrees in January-February (Figure 5.1.1, right panel). Although the average meridional displacements as well as the variability of the EDJ during both phases are of just a few degrees, the associated impact on the SH climate can be large, and accurate forecast of the EDJ is therefore crucial for skilful representation of surface climate in many regions in the SH.

Improving the long-lead prediction of the EDJ migrations' variability has a number of potential benefits, including early warning of anomalous regional weather conditions which can have an impact on public health, agriculture and the shipping industry among other sectors (White et al., 2017). A number of recent modelling studies have provided robust evidence for S2S predictability

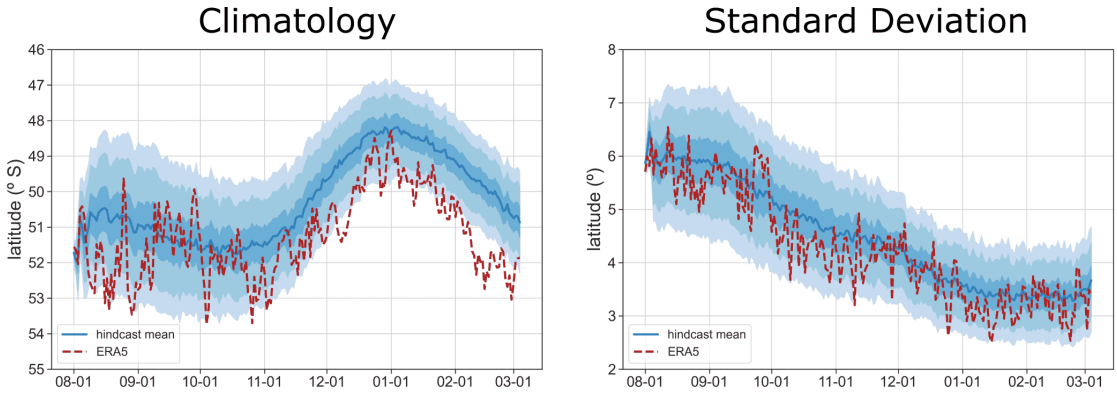

Figure 5.1.1: Climatology of the EDJ latitude (left) and standard deviation (right) between August and March, adapted from Byrne et al. (2019). Here the latitude is defined as the one corresponding to the maximum zonal-mean zonal wind at 850 hPa. The 1981-2016 ERA5 reanalysis climatology is shown as a dashed red line. The blue line and shaded bands are extracted from the ECMWF System 4 hindcast of 51 ensemble members initialised on the 1st of August for each year from 1981 to 2016 (see further description in Section 5.2). The ensemble mean climatology is the blue line. The increasingly lighter blue bands mark the 25-75, 5-95 and 1-99% percentiles of a 1,000 sample bootstrap procedure where each year 36 members are sampled from the total 36 (years) x 51 (members).

of the extratropical SH troposphere during austral spring and summer (Roff et al., 2011; Lim et al., 2013; Son et al., 2013; Osman et al., 2015; Osman and Vera, 2016). Importantly, the added tropospheric skill comes in part from the stratosphere (e.g. Seviour et al., 2014; Byrne and Shepherd, 2018; Lim et al., 2018) and this conclusion is robust to sampling uncertainty (Byrne et al., 2019). The SH extratropics are in fact particularly affected by early vortex weakening, resulting in extreme heat and droughts in Australia and Antarctica and cold spells in south-eastern Africa and South America (e.g. Lim et al., 2019, 2021; Mindlin et al., 2020). Negative stratospheric ozone anomalies are associated with a poleward shift of the storm tracks over the Southern Ocean, hot spells in southern Africa, Australia and South America, and increased UV radiation (e.g. Arblaster et al., 2011; McLandress et al., 2011; Thompson et al., 2011; Mindlin et al., 2021; Oh et al., 2022).

Stratospheric drivers of S-T coupled variability

Research in the past two decades has established that the tropospheric circulation variability is strongly coupled to the variability in the stratosphere in the SH spring-to-summer period, and it is

organised by the timing of the polar vortex breakdown (e.g. Kuroda and Kodera, 1998; Thompson et al., 2005; Byrne et al., 2017; Domeisen et al., 2020b). More in detail, the magnitude of the poleward shift of the EDJ in October (EDJ-1) is partially driven by the strength of the stratospheric vortex from August to October, a result obtained both from reanalysis and models (e.g. Seviour et al., 2014; Byrne, 2017). The timing of the equatorward swing of the EDJ (EDJ-2) between late October and December is strongly dependent on the timing of the VB event (Black and McDaniel, 2007; Byrne et al., 2017; Byrne and Shepherd, 2018; Ceppi and Shepherd, 2019). These findings confirm earlier analysis of the SAO by Bracegirdle (2011) for the 1979-2009 NCEP reanalysis, who found that the SAO's amplitude is maximally correlated with zonal-mean zonal wind $[u]$ at 60 ºS at 10 hPa between August and September; while the SAO phase[1] is maximally correlated with $[u]$ at 60 ºS in the lower stratosphere between September and November.

The stratospheric Quasi Biennial Oscillation (QBO) is also associated with SH extratropical stratospheric variability, and thus indirectly with the EDJ. Observed westerly phases (WQBO) are associated with stronger westerlies and a colder high-latitude stratosphere, and later final warming and vortex breakdown dates than for the easterly phase (EQBO) (Baldwin and Dunkerton, 1988; Anstey and Shepherd, 2014; Byrne and Shepherd, 2018).

Studies suggest that the EDJ springtime anomalies can be traced back to the state of the vortex near the stratopause as early as June. Lim et al. (2018) performed a height-time domain EOF analysis of polar $[u]$ monthly anomalies, finding that the leading EOF (40% of interannual variability) is characterized by a signal that develops in June-July near the stratopause, changes sign in August-September, and eventually propagates downward to the surface from October to January. The high-stratospheric shift of winds captures the poleward and downward evolution of the Polar Night Jet oscillation (PJO) (Kuroda and Kodera, 2001; Kuroda, 2002; Lim et al., 2018) which correlates with the year-to-year variability of $[u]$ at 55-65ºS, as noted already by Kuroda and Kodera (1998).

Observational (Randel and Wu, 1999; Young et al., 2013) and modelling (McLandress et al., 2010; Richard et al., 2010; Arblaster et al., 2011) studies have shown that during the last decades of the 20th century anthropogenic ozone depletion has resulted in a cooling of the Antarctic lower stratosphere in late-spring. As a consequence, the stratospheric polar vortex has strengthened and

[1] Phase marked as the date of the first peak of the SAO's second harmonic.

the breakdown date delayed by about 2 weeks (Waugh et al., 1999; McLandress et al., 2010), with downward effect on the jet too (McLandress et al., 2011).

Tropospheric drivers of S-T coupled variability

Tropospheric wave forcing of various sources can influence the EDJ indirectly via their effect on the stratospheric polar vortex strength and breakdown. Anomalously strong upward propagating planetary-waves can advance the seasonal disruption of the SH stratospheric vortex, with the waves generally contributing to destabilise the state of the vortex with a lead time of one month (e.g. Hardiman et al., 2011; Hu et al., 2014; Lim et al., 2018). Lim et al. (2018) found the strongest preconditioning of the S-T coupling from upward wave fluxes, represented via the zonal-mean eddy heat flux (vT-flux) in July-August around 45°S to 75°S at 100hPa, an index used also in Rao et al., 2020 and Lim et al., 2021. Because the eddy heat flux is proportional to the vertical component of the Eliassen-Palm flux, it is also used to represent upward propagation of planetary waves entering the stratosphere. Similarly Shen et al. (2022) found that observed Antarctic weak polar events (e.g. 2019) are primarily driven by enhanced wave-number 1 vT-flux when in the presence of a preconditioned polar stratosphere, i.e. a weaker and more contracted Antarctic stratospheric polar vortex.

Anomalies in the equatorial Pacific Ocean sea surface temperature, reflected in the El Niño Southern Oscillation (ENSO), are known to affect the SH extratropics both in the troposphere and the stratosphere. In summer, an approximatively zonally symmetric tropospheric response to ENSO is observed, characterized by a shift in latitude of the EDJ in the extratropics (e.g. Seager et al., 2003; L'Heureux and Thompson, 2006; Lim et al., 2013). This is due to a combination of tropospheric and stratospheric mechanisms, the latter acting on the EDJ through the influence of upward propagating planetary-waves (Van Loon et al., 1982; Byrne et al., 2019; Stone et al., 2022). The stratospheric pathway is more prominent during austral summer than in spring, and it is know to strongly project onto the zonally symmetric circulation (Hurwitz et al., 2011; Lin et al., 2012; Zubiaurre and Calvo, 2012; Domeisen et al., 2019). ENSO also affects the SH high-latitude zonally asymmetric circulation, especially visible in early spring (Kidson, 1999; Mo and Paegle, 2001; Vera et al., 2004; Ding and Li, 2012; Wilson et al., 2016), through the modulation of the Pacific-South American (PSA) pattern, an alternating wave-train of pressure anomalies emanating

from the tropics in response to convection in the tropical Pacific (Sardeshmukh and Hoskins, 1988; Mo and Paegle, 2001). In this work, the zonally asymmetric effect of ENSO is explored through the response of surface climate variables in Sections 5.4.4 and 5.4.5.

Other oceanic drivers can strongly affect tropospheric variability in the mid-to-high latitudes more regionally. For example, the Indian Ocean Dipole (IOD) is known to influence the Australian climate and in particular precipitation, playing a large role in the dry and windy conditions leading to the 2019/20 Australian 'black summer' of wildfires (Lim et al., 2021). There is also some modelling evidence that the Interdecadal Pacific Oscillation (IPO), another component of the tropical SST variability, played a role in the shifted poleward EDJ over recent decades over the South Indian and Southwest Pacific (Yang et al., 2020). For the sake of simplicity, these drivers are not included in the present analysis which focuses mostly on the zonally symmetric view of the EDJ migrations.

5.1.2 Studying predictability with Bayesian causal networks (BCN)

Despite a number of known long-lead drivers of EDJ variability described in Section 5.1.1, a quantification of which combinations can have a measurable impact on the predictability of EDJ variability has not been preformed yet. A lot of our understanding on predictability is based on modelling, for example with specific initializations or with variables nudged to a specific state and the effect on the EDJ interrogated (e.g. Seviour et al., 2014; Rao et al., 2020, 2021). However this allows to control only for a (few) nudged variable(s) and does not allow to easily and efficiently quantify the interaction of more elements, which may work in concert or in opposition to favour a given EDJ state. On the other hand, many of the statistical studies have looked at composites or correlations between a candidate driver (e.g. the vortex breakdown or ENSO) and the EDJ (Lim et al., 2018; Byrne and Shepherd, 2018; Lim et al., 2019). Assuming a truly causal connection (and that such a mechanism is not currently exploited to its maximum potential by forecast models), the presence of a correlation can indeed represent a route for enhanced predictability. However, to actually quantify this potential, a statistical prediction of the target variables given the identified drivers needs to be performed. This is rarely done, and it is going to be the main contribution of the present work.

To further exemplify the need for a statistical prediction, consider the composite analysis of the

S-T coupled dynamics, e.g. represented by $[u]$ anomalies, given the state of the stratospheric vortex. These composites of the conditional event "$[u]$ given vortex" not only show a strong downward propagating signal (e.g. Thompson et al., 2005), but also show stratospheric anomalies emerging a few weeks (to months) ahead of the SH vortex breakdown date (Byrne and Shepherd, 2018; Lim et al., 2018). This suggests that the vortex state may itself be predictable on multiple weeks lead-times. However, to study the predictability of the vortex one is actually interested in quantifying the likelihood of the opposite conditional event "vortex given $[u]$" in the preceding months, which can be achieved with a statistical prediction. Note that the inverse conditional probability may not lead to the same results.

A statistical model of the S-T coupled variability targeting the EDJ should include all relevant drivers and represent their multiple interactions and co-variabilities in a rigorous and manageable way. To avoid potentially misleading causal interpretation of correlational signals, here a causal network framework is used to guide the statistical modelling (Section 2.3). A causal network is also a parsimonious modelling choice can reproduce the statistical connections among variables by modelling just the few direct and relevant causal connections. This process of simplification accepts a loss of detail (e.g. compared to dynamical model) at the price of gaining insight into the connections between the variables. Another important aspect of modelling with causal networks is that the hypotheses are transparent and easy to communicate, since they are evident from the drawing of the network itself. When used to produce forecasts, the direct and indirect causes of a certain outcome can be easily read out from the network structure.

Given the probabilistic nature of the evolution of S-T dynamics on S2S time scales, a probabilistic Bayesian causal network (BCN) is preferred to a deterministic one (Section 2.2). In addition to explicitly representing uncertainty, another benefit of the probabilistic choice is that it allows for a fully non-linear description of the connections including state-dependence, which are thought to be potentially relevant in driving the S-T coupling variability. Most causal network studies have used deterministic functions to model the links between the variables, most commonly linear functions where links are characterised by MLR coefficients. (e.g. Kretschmer et al., 2016; Di Capua et al., 2020). Here instead links are parametrised with conditional probabilities (CP). To our knowledge, BCNs have rarely been applied to studies of climate variability or forecasting, probably due to their slightly more complicated modelling and especially the need for a large enough dataset to fit

the parametrization. The only studies found in our research were by Abramson et al., 1996; Boneh et al., 2015; Barnes et al., 2019; Harwood et al., 2021. Notably, however, BN have been routinely used in weather and climate risk assessment studies that wish to include expert knowledge and represent it and the relative uncertainties in a probabilistic way (e.g. Peter et al., 2009; Catenacci and Giupponi, 2013; Young et al., 2020), since the problem of data scarcity can be bypassed by parametrizing the networks via expert interviews.

5.2 Data

The ensemble hindcast (re-forecast) of the European Centre for Medium-Range Weather Forecasts (ECMWF) is used as training data, providing large enough sample size for probabilistic parametrization. Although the latest operational system is currently SEAS5, the earlier version known as System 4 (Molteni et al., 2011) is chosen because it has a more realistic SH circulation variability at polar latitudes than SEAS5, possibly due to a better representation of stratospheric variability in the SH (Shepherd et al., 2018). System 4 is based on the Integrated Forecasting System (IFS) atmospheric component coupled to the Nucleus for European Modelling of the Ocean (NEMO) ocean model. The atmospheric vertical resolution is 91 levels, with a model top in the mesosphere at 0.01hPa, and the horizontal spectral resolution is T255, which corresponds to approximately 80 km in the horizontal. The resolution of the ocean model is 1° in the horizontal and has 42 layers in the vertical. Hindcasts initialized on 1 August for each year between 1981 and 2016 included are considered (36 seasons). All hindcasts are issued as ensembles with 51 members and run for 7 months, of which the first six are analysed. The ECMWF ERA-5 reanalysis is used as verification data (Hersbach et al., 2020).

The basic data input for our study are the daily mean zonal wind (u), meridional wind (v), air temperature (T), and sea surface temperature (SST). Most of the network variables will be constructed from zonal averages, which will be denoted with square brackets $[]$ in what follows. For the analysis of surface climate impacts in Sections 5.4.4 and 5.4.5 we use the monthly mean geopotential height at 500hPa (Z500), 2 meter temperature (2mT) and mean total precipitation rate (MTPR).

Differently from Byrne et al. (2019) and Osman et al. (2022) which have analysed the same hindcast, the season 2002/03 is retained in the analysis despite it being the only SSW on obser-

vational record. Performing an analysis of [u] at 60ºS and 50 hPa (not shown) did not reveal an outlier behaviour of the 2002 ensemble members, which remained well within the large ensemble distribution (i.e. there are many other years where some members show a comparatively early vortex breakdown). Further, not all members in 2002 exhibit an extreme behaviour either. This is consistent with recent analysis of multimillennial simulations with a resolved stratosphere showed that the frequency of SSWs is about one in 22 years for 1990 conditions (Jucker et al., 2021).

5.2.1 Representation of stratospheric ozone

As mentioned in 5.1.1, good representation of ozone is important to capture the late 20th century trend and its impact on intra-seasonal variability. In System 4, ozone is activated as a prognostic variable and is radiatively interactive, thus the ozone field is free to evolve during the forecast and will develop anomalies physically consistent with e.g. temperature anomalies and specified CFC time history (Molteni et al., 2011). However its initial conditions are not time-dependent and are instead prescribed via a seasonally varying climatology computed from ERA interim ozone analyses (1996-2002, believed to be the most reliable years), missing the difference between the two epochs of ozone loss (1980s to early 2000s) and then recovery (2000s afterwards), due to errors in ERA Interim caused by changes in satellite instruments which were causing unrealistic temperature variations in the stratosphere. The simplified ozone chemistry in System 4 is likely not to have a very realistic representation of the polar stratospheric cloud chemistry required to produce an ozone hole, regardless of the issue with initial conditions. Indeed, Monge-Sanz et al. (2022) recently showed with their new stratospheric ozone model (implemented n their study in ECMWF SEAS5) that a realistic interactive prognostic ozone field is needed to reproduce the SH polar vortex behaviour (including VB delay due to lower ozone concentration), while an ozone climatology does not provide enough information for the model to reproduce all necessary feedbacks with dynamics.

An analysis of expected ozone-induced trend for Epoch 1 (1981-2001, ozone decline) and no trend in Epoch 2 (1995-2016, recovery) is performed the November to January $[u]$ 50 hPa 60ºS, following the same method used in Chapter 3. Figure 5.2.1 shows some sign of trend in the vortex in Epoch 1, hence the CFC historical trends are able to imprint a strengthening trend in the vortex (dark blue line). However the modelled trend is overall weaker than in observations (red line).

Further analysis would be needed to confirm the hypothesis of an issue with the representation of ozone chemistry, which is out of the scope of the present work. Although within sampling uncertainty, the weaker vortex compared to reanalysis led to the exclusion of an ozone variable in the network.

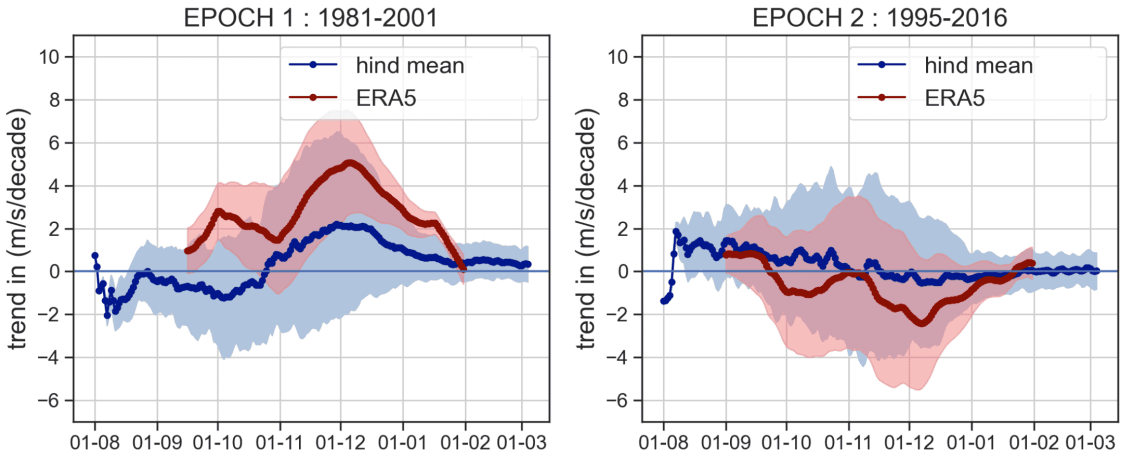

Figure 5.2.1: Daily hindcast and ERA5 vortex trend between 1981-2001 (left panel) and 1995-2016 (right panel). The ensemble mean (dark blue line) and the ensemble standard deviation (area filled in shaded blue) are shown. ERA5 trend (red line) and the regression coefficient's standard error (area filled in shaded red) in light red are shown. The ERA5 trend is calculated after a 30-days running average, centred around the day of year plotted, to remove noise. The peak of the hindcast ensemble mean trend in Epoch 1 is about 2 m/s/decade (standard deviation from −2 to 6 m/s/decade), while the ERA5 mean is about 5 m/s/decade.

5.2.2 Definition of variables

The fundamental elements of the zonal-mean S-T coupled seasonal variability introduced in Section 5.1.1 are translated into nine variables that constitute the nodes of the BCN. They are listed in Table 5.2.1 and briefly summarised in the following. EDJ-1 is defined following Byrne and Shepherd (2018) and the EDJ-2 following Byrne and Shepherd (2018), Byrne et al. (2019) and Ceppi and Shepherd (2019) (see also Chapter 3). Note they are defined over slightly different vertical levels for consistency with previous literature, but very similar results are obtained if both are defined at 850 hPa. ENSO is defined in the Nino 3.4 region and it can influence the EDJ both directly, and indirectly through the planetary wave activity entering the stratosphere in August and

Index name	Index definition	Time period
ENSO	SST anomalies averaged over the Niño 3.4 region	Aug-Oct mean
QBO	Daily zonal-mean zonal wind ($[u]$) at 30 hPa at 0ºS	Aug-Oct mean
vT-flux	Daily eddy heat flux $[v'T']$ at 100 hPa averaged over 45-75ºS	Aug-Sept mean
PJO-1	Daily $[u]$ at 10 hPa average over 30-45ºS	Aug mean
PJO-2	Daily $[u]$ at 30 hPa average over 50-60ºS	Sept mean
SPV-low	Daily $[u]$ mass-weighted between 100-50 hPa averaged over 60-70ºS	Oct mean
VB	Last date when the 5-days centred running mean of daily $[u]$ at 50 hPa and 60ºS falls below 15 m/s, in days after 1 Oct	Oct to Jan
EDJ-1	Latitude of the maximum of daily $[u]$ pressure-weighted between 1000-500 hPa, 10-days running mean	1 Oct to 8 Nov
EDJ-2	Latitude of the maximum of daily $[u]$ at 850 hPa, 10-days running mean	1 Nov to 19 Jan

Table 5.2.1: Definition of the variables used to construct the BCN. The square brackets represent the computation of zonal-mean and the prime symbol represents anomalies with respect to the zonal-mean. For EDJ-1 and EDJ-2, the running mean is used to extract the latitudinal shift signal from the noisy latitude.

September (vT-flux), defined as in Rao et al. (2020) and Lim et al. (2021). In the early spring stratosphere, variables include the QBO and the phases of the PJO, one in the mid-latitudes in August (PJO-1) and one in the high-latitudes in September (PJO-2). To define the PJOs, the results in Lim et al. (2018) are used to identify a PJO-1 as $[u]$ at 10 hPa averaged over 30-45ºS in August, and PJO-2 as $[u]$ at 30 hPa and averaged over 50-60ºS in September. To represent the stratospheric components of the S-T coupled variability, the strength of the lower stratospheric [u] (SPV-low) in October and the timing of the vortex breakdown (VB) (a date usually between late October and early January) are used.

Bias correction

The presence of mean biases in ENSO and EDJ-1 and EDJ-2 in this seasonal hindcast dataset has already been documented (Molteni et al., 2011; Byrne et al., 2019), and is described in detail in Appendix 5.A. This led to the choice of performing a mean bias correction, removing from each

ensemble member the difference between the hindcast multi-year ensemble mean and the ERA5 multi-year mean. Bias correction is performed for all variables to apply a consistent approach that avoids a too subjective choice of which level of bias is small enough to be ignored or not[1]. For the EDJ-1 and EDJ-2 time-series, for each day of the year and ensemble member, the difference between the hindcast and ERA5 latitude smoothed means (15-days running mean) are removed (shown in Figure 5.A.1).

Categorization

Finally, the variables are transformed from continuous to categorical. This is done to parametrize the network more simply, to produce an easy to interpret forecast and it is also in line with current practice of presenting seasonal forecasts (e.g. see Met Office seasonal outlooks [2]). The categories are defined in Table 5.2.2. They aim at best summarising the ensemble spread. They are defined based on physically meaningful thresholds (e.g. 0 m/s for E and W QBO) or on percentiles (e.g. 25-75th for VB), except for the EDJs. As an example, Figure 5.2.2(a) shows the split in categories of the ensemble distribution for VB in each year (left) and the overall ensemble distribution (right), also compared with ERA5 (red). Note that some variables have asymmetric categories due to the asymmetric shape of their distribution: for example the percentiles used for PJO-2 are the 35th, 75th (see Figure 5.B.1). Plots for all other variables are found in Appendix 5.B.

The EDJ variables are categorised differently via an agglomerative hierarchical clustering. This is a dendogram based approach that divides a set of time series into clusters of similar shape, thus can retain all the information present in the evolution of the EDJ (e.g. timing and magnitude of the shifts). The clustering is performed with Euclidean distance metric (measures the distance between two time series) and Ward's linkage criterion (measures the distance between two clusters). With these metrics, clusters have the minimal distance between their elements and the maximum

[1] Note that this bias correction is equivalent to computing anomalies for reanalysis and hindcast with respect to their own climatologies. However, not all variables considered here are anomalised, either for ease of interpretation (e.g. actual value of vT-flux rather than anomalous value) or due to necessity (e.g. to compute the variables QBO, EDJs and VB in the first place).

[2] https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/business/public-sector/civil-contingency/3moutlook-jfm-v2-secure.pdf [last accessed 23/04/2022].

Index name	Number of categories	Thresholds or method	Labelling of categories
ENSO	3	-0.4, +0.4 ºC	Niña , neutral, Niño
QBO	2	0 m/s	easterly, westerly
vT-flux	3	30th, 75th hindcast percentiles	strong, neutral, weak
PJO-1	2	60th	weak, strong
PJO-2	3	35th, 75th	weak, neutral, strong
SPV-low	3	25th, 75th	small, neutral, large
VB	3	25th, 75th	early, middle, late
EDJ-1	3	clustering (32.5%, 52%, 15.5%)	poleward, middle, equatorward
EDJ-2	3	clustering (26.5%, 51%, 22.5%)	poleward, middle, equatorward

Table 5.2.2: Definition of categories for each network variable. Note that in the SH the vT-flux has a negative sign, therefore lower percentiles mean stronger wave activity. For EDJ-1 and EDJ-2, the percentage refer to the proportion of ensemble members assigned to each cluster.

distance across different clusters. EDJ-1 and EDJ-2 categories are shown in Figure 5.2.2 (b,c). EDJ-1 categories can be summarised as an overall more poleward (blue) or equatorward (orange) position of the jet in October with respect to the climatological time series (black line). EDJ-2 categories are either a more poleward and early shift (blue) or a more equatorward and delayed shift (orange) between November-December. Both central states matches the climatologies. Note that the split of EDJ-2 (eq/pole) is very similar to the one detected by Byrne et al. (2017) in correspondence with the timing of the polar vortex breakdown (late/early).

ERA5 variables are categorised based on the hindcast thresholds. For EDJs, ERA5 season-specific time series are assigned to their closest hindcast category, identified as having the smallest Euclidian distance from the hindcast cluster-mean (shown as red lines in Figure 5.2.2(b,c)).

5.3 Methods

As introduced in Section 2.2, mathematically a Bayesian network (BN) is a DAG which exploits conditional independence relationships between the variables to summarise the rich information stored in their joint probability distribution $P(X)$ via the factorization of Equation (5.1), here

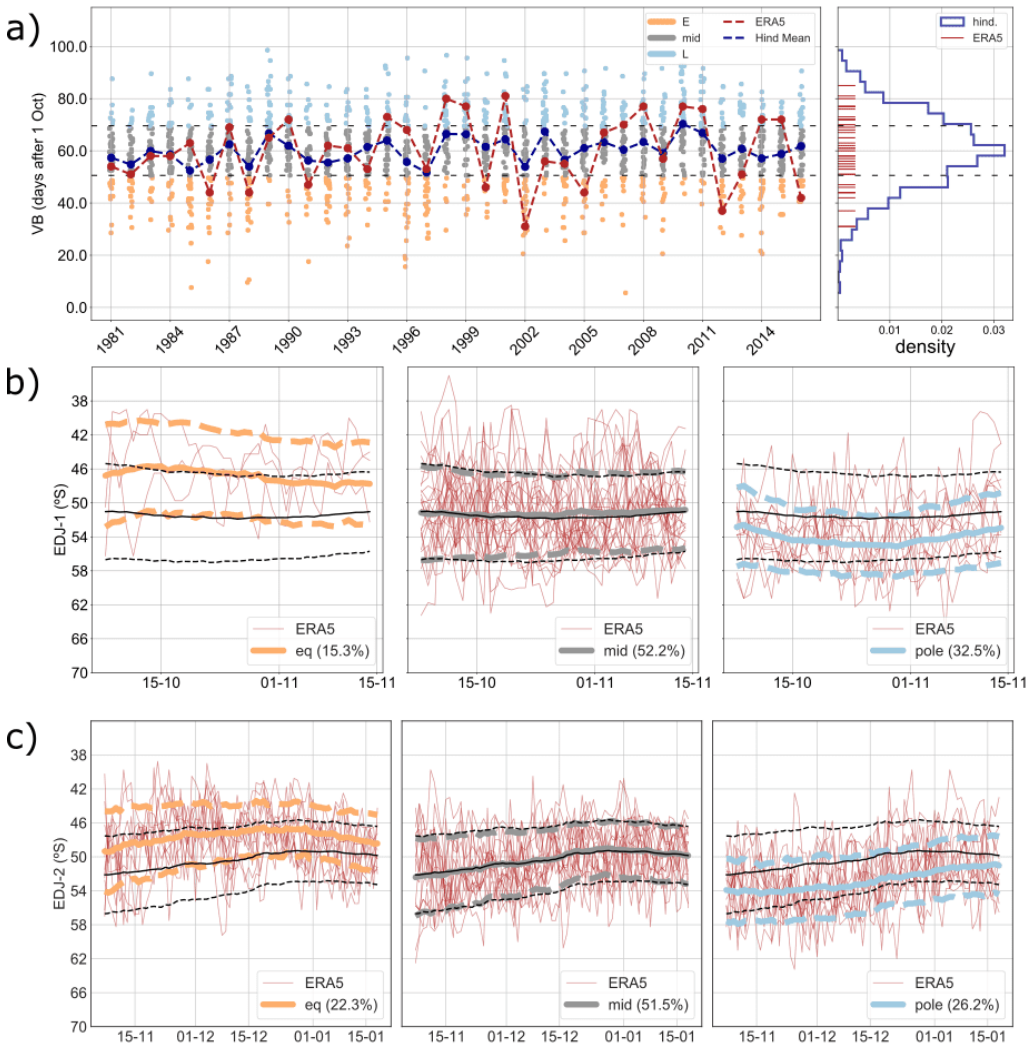

Figure 5.2.2: (a) VB distribution for the bias corrected hindcast (right) and the yearly time series (left) divided into strip-plots for each of the three categories (early, middle and late are in orange, grey and light blue respectively; ensemble mean in purple). ERA5 values are shown in red. (b) EDJ-1 and (c) EDJ-2 categories for the bias corrected hindcast, labelled as equatorward (orange), mid (grey) and poleward (blue). The percentage values in parentheses show the fraction of all members and years assigned to each category (sum is 100% by construction). The thick coloured lines represent the hindcast cluster mean (solid) and plus or minus one standard deviation (dashed). The thick black lines represent the ensemble mean climatology (solid) and plus or minus one standard deviation (dashed). The ERA5 yearly time-series assigned to each hindcast category, as the closest based on averaged Eucledian distance (see text), are shown in red.

restated for ease of reference:

$$P(X) = \prod_{i=1}^{N} P(X_i|PA_i) \tag{5.1}$$

where there are N random variables X_i, each being influenced directly by a set of variables PA_i, called the parents. Each variable X_i is therefore associated with a conditional probability distribution, $P(X_i|PA_i)$, which encodes the probability of observing specific values of X_i given the values of its parents. In a BCN links encode also the concept of causality and not just of conditional independence, as discussed in Section 2.3. The network developed here is placed within the causal network family because the linkages used are meant to represent physical and causal understanding of the system.

5.3.1 Building a BCN

Among the various techniques to build a causal network (e.g. Marcot et al., 2006; Chen and Pollino, 2012; Runge et al., 2019a; Young et al., 2020), a hybrid expert and data-driven approach is used, and described by the following two-step procedure.

In the first step (Step 1), the initial hypothetical structure of the BCN is drawn by including all possible links suggested by the literature presented in Section 5.1.1, resulting in a set of potential parents $\hat{P}(X)$ for each variable X. Note that the links are all defined so that they do not point backwards in time, but when variables are "contemporaneous" (on the considered time-scales) the direction is set by known physical processes. For example, the link is from SPV-low to EDJ-1 and not vice-versa, based on known downward direction of causal effect on these time scales. Any known common driver is included, to ensure Causal Sufficiency condition is met (see Section 2.5.2). Also as part of this step, the unconditional independence of all pairs of variables is tested for, to remove any trivial links. Note that Step 1 is similar to the first part of many constraint-based causal discovery methods (e.g. step PC of the PCMCI algorithm described in Section 2.5.2).

The second step (Step 2) is to test each potential link $X \to Y$ via $X \not\perp\!\!\!\perp Y \mid \hat{P}(Y) \setminus \{X\}$. This is crucial to ensure that the joint probability of the system can indeed be expressed as in Equation (5.1). If this test is "passed" the link is retained, otherwise the link is removed (similar to the MCI step in the PCMCI algorithm).

The conditional independence (CI) test chosen is the Pearson χ^2 which applies to categorical data. If the p-value p of the test is smaller than a significance level α, the independence is rejected

and the link is retained; otherwise the link is removed. Note that a different choice of CI test may lead to slightly different results, especially for the weaker or less significant links. The value of $\alpha = 0.05$ is chosen as it is relatively loose to allow for weak links to be included, but tight enough to remove the ones that cannot provide added skill to the prediction.

Given the links, the network is parametrized by computing the set of probabilities $P(X_i|PA_i)$, which for discrete variables are represented by conditional probability tables (CPT). Under a multinomial sampling assumption, the entries of a CPT can be estimated as relative frequencies $\pi_{ijk} = N_{ijk}/N_{ij}$, where N_{ijk} is the number of occurrences in the dataset where variable X_i has the value k and its parents have the (set of) values j. This is also known as maximum likelihood estimation and is a reasonable approximation for large enough sample sizes (as in this case) and for individual probabilities larger than 0[1]. This is the estimation choice of Barnes et al., 2019 too. Note that other probabilistic modelling choices are possible, for example Gaussian Process Regression which provides a non-linear fit of a variable onto the regressions and an envelope of uncertainty around it. It would be interesting to compare the insight gained via such alternative modelling choices but this remains out of scope of the current analysis.

5.3.2 Probabilistic forecasts

A network structure with associated CPTs is a fully parametrized model that can be used to perform predictions. Prediction in a BCN is the process of calculating a the distribution of target variables given information about some drivers (evidence). This is achieved by marginalizing Eq. (5.1) over the unobserved covariates (sums over all categories; express $P(X|E) = P(X,E)/P(E)$, then for each probability take the factorization and marginalize over the unobserved variables). In a recent study by Byrne et al. (2019) a categorical forecast for the EDJ based on knowledge of vortex breakdown state was shown to give an anomaly correlation coefficient comparable to that of a full dynamical seasonal hindcast (Byrne et al., 2019). The BCN model can be seen as a more formalised and multivariate extension of this approach.

[1] A multinomial sampling assumption treats counts n_{ijk} of a contingency table as the outcomes of a multinomial with probabilities π_{ijk} that sum up to one. An alternative is to use Bayesian estimation (with Dirichlet prior, with parameter α_{ijk}), which is important if occurrence of events close to $\pi_{ijk} \to 0$. Note that α_{ijk} are often hard to find, so with Laplace smoothing α_{ijk} is often considered constant and equal to 1. The two estimates converge for large sample sizes.

Forecast verification metrics

The probabilistic forecasts produced with the BCN are analysed with two metrics: the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) (ROC AUC) and the AUC of the Precision-Recall curve (PR AUC).

For the prediction of a given category x, a ROC curve is made up of points on the sensitivity and specificity plane, each computed for a different probability threshold (which sets the level for a prediction being a positive, i.e. that of the target category). Sensitivity (or recall) is the probability of a positive prediction given a positive observation; while specificity is the probability of a negative prediction given a negative observation. The area under the ROC curve (ROC AUC) summarizes the trade-off between sensitivity and specificity of the model in a single number. An AUC=0.8 means that the classifier ranks two test data points correctly with a probability of 80% (no skill being 50%), so the larger the AUC the better.

The Precision-Recall (PR) curve is defined similarly to ROC but in the recall and precision plane. Precision is the probability of a positive observation given a positive prediction. While ROC quantifies the trade-off between the true positive rate and the true negative rate, PR balances the true positive rate with the false discovery rate. The no skill value of PR AUC is the fraction of real positives in the data (here called baseline b), which decreases when the data become more imbalanced, and is 0.5 for perfectly balanced classes. Given the varying level of no-skill, the results here are presented as percentage increase compared to baseline PR AUC% = 100 · (PR AUC - b)/(1 -b), hence 100% is the perfect forecast and 0 indicates no skill.

PR AUC is introduced because ROC AUC alone can be misleading (too optimistic) in the case of imbalanced categories (like for some of the BCN variables) concurrently with a target easily predictable and the actual interest in low false alarms. For this, PR AUC is more informative than ROC AUC as it looks at precision as well. On the other hand, PR AUC alone can also be misleading or uninformative when there is class imbalance and the target variable is difficult to predict. Unfortunately, there is no way to unequivocally link PR AUC and ROC AUC, for example a higher PR AUC does not imply a higher ROC AUC. Here both are considered to give more a rounded assessment of the forecast.

5.4 Results

The following sections present the results of the BCN analysis. The BCN structure is discussed in terms of the strength of each link, its CPTs and the model cross-validation (Section 5.4.1). The predictability arising from long-lead drivers of EDJ is assessed by performing BCN forecasts, using each ensemble member to provide evidence and verification data (Section 5.4.2). Then, the BCN forecast skill is assessed against ERA5 and compared with a synthetic conditional hindcast (Section 5.4.3). Finally, the surface response associated with the tropospheric variables is quantified (Section 5.4.4) and the ability of the BCN to provide a filter for the most skilful hindcast ensemble members is assessed (Section 5.4.5).

5.4.1 The BCN

The literature-informed initial structure of the network is tested with the procedure described in Section 5.3.1, using a χ^2 test and α=0.05 significance level. The results of the conditional independence tests are reported in Table 5.4.1. The graphical result is shown in Figure 5.4.1, where the links that pass both Step 1 and 2 test are the black non-dashed ones.

The links removed with the initial unconditional independence test in Step 1 are all the links emanating from QBO (leaving it disconnected from the rest) and ENSO $\rightarrow$ EDJ-1. The absence of teleconnection from QBO to the stratospheric zonal wind can also be seen from hindcast composites of $[u]$ based on QBO (W minus E) which do not show the sign of teleconnection ususally found in reanalysis (Appendix 5.D). It is known that models struggle to represent this teleconnection, especially when initialised early in the season, losing information about the amplitude and even the sign of tropical winds within 1-2 weeks, for runs initialised with set QBO winds (Butler et al., 2016; Garfinkel et al., 2018; Lawrence et al., 2022). This is likely the reason for the absence of links from the QBO found here as well. Note however that the conditional probability $P(VB \mid QBO)$ in ERA5 shows only a marginally stronger signal, with E QBO associated with early VB (and vice-versa for W QBO with late VB) as shown in Figure 5.C.1 bottom panel in the Appendix. The link ENSO $\rightarrow$ EDJ-1 is also removed, a result that may be due to the zonally asymmetric pattern of ENSO's teleconnections in September-October, which is not captured by the zonally averaged EDJ-1 index.

A further three links are removed with the conditional independence tests in Step 2: PJO-2 $\rightarrow$ VB, ENSO $\rightarrow$ VB and ENSO $\rightarrow$ EDJ-2. The unconditional connection between ENSO and VB

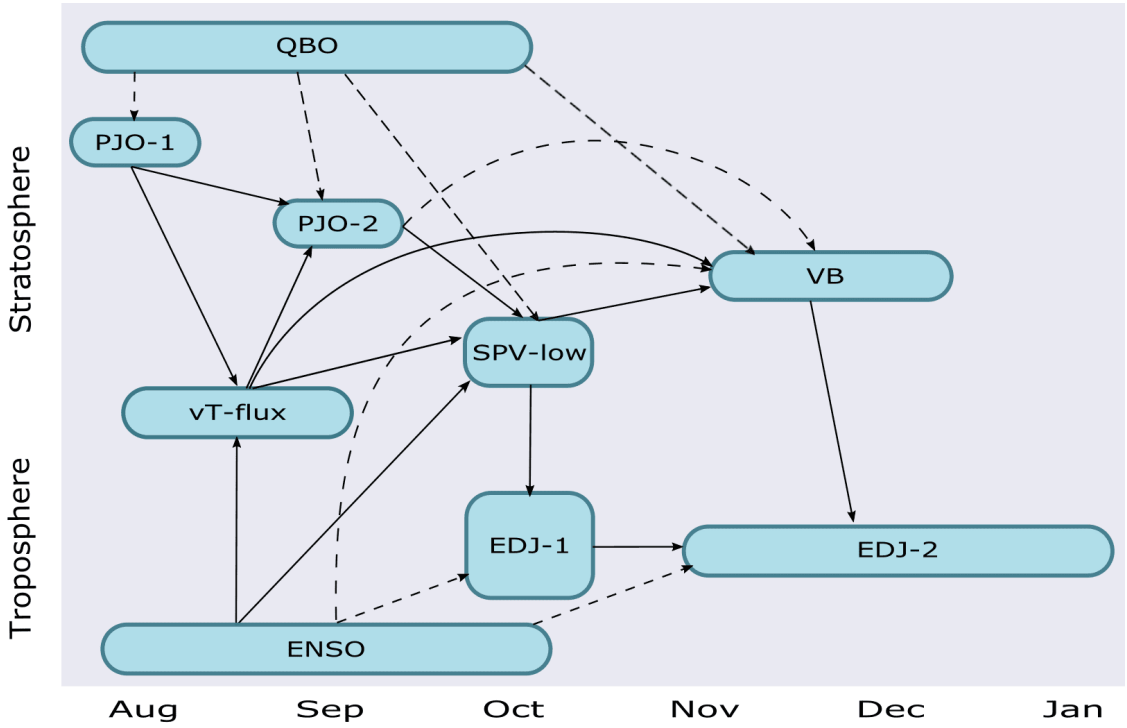

Figure 5.4.1: The literature-informed initial structure of the network (all links). The ones that do (do not) not pass the conditional independence test procedure described in Section 5.3.1 with χ^2 and 0.05 significance level are the black (dashed) links.

X	Y	given Z	χ^2	p-value	remove link $\alpha = 0.05$
PJO-1	vT	ENSO	160	10^{-9}	No
ENSO	vT	PJO-1	34	$2.9 \cdot 10^{-5}$	No
PJO-1	PJO-2	vT	529	10^{-9}	No
vT	PJO-2	PJO-1	851	10^{-9}	No
PJO-2	SPV-low	vT, ENSO	131	10^{-9}	No
vT	SPV-low	PJO-2, ENSO	105	10^{-9}	No
ENSO	SPV-low	vT, PJO-2	55	0.02	No
PJO-2	VB	vT, SPV-low	26	0.88	**Yes**
vT	VB	PJO-2, SPV-low	64	0.002	No
SPV-low	VB	PJO-2, vT	415	10^{-9}	No
SPV-low	EDJ-1	ENSO	845	10^{-9}	No
ENSO	EDJ-1	SPV-low	11	0.44	**Yes**
EDJ-1	EDJ-2	ENSO, VB	81	$2.3 \cdot 10^{-5}$	No
ENSO	EDJ-2	VB, EDJ-1	31	0.64	**Yes**
VB	EDJ-2	ENSO, EDJ-1	645	10^{-9}	No

Table 5.4.1: Result of the Step 2 tests of conditional independence $X, Y \mid Z = \hat{P}_u(Y) \setminus \{X\}$. Z are the potential parents that survive Step 1 (all links except for the ones emanating from QBO, and ENSO $\rightarrow$ EDJ-1).

was found to be significant ($\chi^2(ENSO, VB)$ = 37, p-value = 3.7 10^{-7}) although weaker than most other links (not shown). However, once conditioned on vT-flux and SPV-low, the link ENSO $\rightarrow$ VB is removed (p-value=0.09). This suggest that much of the information from ENSO to VB is mediated by the upward propagating waves in August-September and the stratospheric pathway. Similarly happens for PJO-2 $\rightarrow$ VB, suggesting that SPV-low mediates most of this effect.

The link ENSO $\rightarrow$ EDJ-2 is removed after accounting for the information given by VB and EDJ-1. This is in line with the analysis done in Byrne et al. (2017) on reanalysis and Byrne et al., 2019 on this same hindcast data. There the authors argued that the correlation between ENSO and the SAM between late spring and early summer is mainly the result of sampling uncertainty, and that the summertime SAM variations are mainly driven by variations in the breakdown date of the stratospheric polar vortex, which happen to be correlated with ENSO in the limited historical record. In further support of this thesis, in Appendix 5.I the hindcast Z500 anomaly composites based on ENSO and EDJ-2 (Fig. 5.J.2) is found to be very similar to the composites for ENSO and VB (Fig. 5.J.3).

The missing direct link from ENSO to EDJ-2 may however be a feature of the forecast model and not of the real world. For example, using NCEP reanalysis and with a MLR Kretschmer et al. (2021) find a link from ENSO to EDJ-2 even after controlling for VB. In the current probabilistic framework, a way to test this potential inconsistency with ERA5 is to compare the probability $P(\text{EDJ-2} \mid \text{ENSO})$ (Figure 5.C.2): a larger probability is found for ERA5, yet within the 5-95% confidence interval of the hindcast estimate. This may suggests that the direct connection between ENSO and EDJ-2 is slightly under-represented in System 4. However, $P(\text{VB} \mid \text{ENSO})$ in ERA5 is also a bit stronger than in the hindcast. Given the pathway from ENSO to VB via vT-flux, it remains hard to tell which of the indirect or direct connections between ENSO and EDJ-2 is responsible for the stronger correlation in ERA5.

Link strength

Links in a causal network are often associated with a measure of strength, which is useful to compare the relative importance of the parents for their child. For a categorical BCN, measures of link strength are however not straightforward, as also discussed in Barnes et al. (2019). For continuous BCN, Harwood et al. (2021) performed a linear analysis of variance (ANOVA) to

estimate the relative importance of each relationship. The non linear interactions are however interesting to consider in this context, and furthermore the EDJ variables cannot be easily converted into continuous variables. An alternative route is to use information theoretic estimates. Given $X \to Y$ and Z the only parents of Y other than X, Ebert-Uphoff (2007) showed that the conditional mutual information between X and Y given Z can be interpreted as a measure of link strength, and called it the Blind Average Link Strength LS_{blind} (see derivation in Appendix 5.E). Expressed as a percentage of conditional entropy, $LS_{blind}\%(X \to Y)$, LS% for short, is a measure of *how close the link is to deterministic* and therefore how influential the parent X is in determining the state of the child Y. The values LS% are used to label the links and scale their width in Figure 5.4.2, and are also listed in Table 5.4.2 (together with the link's χ^2 and p-value, for completeness).

The highest LS% values are for the links PJO-1 $\to$ PJO-2, vT-flux $\to$ PJO-2, SPV-low $\to$ EDJ-1 and VB $\to$ EDJ-2 ($\sim 20-30\%$). Because both parents of PJO-2 have large LS%, this suggests that PJO-2's variability is well represented by its parents, and thus should be skilfully predicted. On the other hand, only one of the two parents of EDJ-2 have large link strength (namely VB) which suggests EDJ-2 is less predictable than PJO-2 given the parents. The two links pointing at vT-flux are among the the weakest ($< 5\%$) and thus its predictability is likely very small. Links pointing at SPV-low and VB have medium strength ($\sim 5-20\%$), but the ones emanating from vT-flux are weak ($< 5\%$). This may be due to the fact that strong vT-fluxes in the lower stratosphere may not drive a weaker vortex and the wave activity could just propagate and break farther equatorward. Indeed, it is the derivative of the wave flux that determines how much the mean flow is affected, as this represents the convergence (divergence) of the said wave activity. Also, this metric summarises the effect of both weak, neutral and strong vT-flux, and a weak connection in just one of the states may lead to this result too. Finally, note that the link-specific metrics are built by marginalization of the effect of all the other parents, therefore this analysis may mask the importance of combined non-linear interactions. Despite this, the overall connection between LS% and predictability is confirmed by the analysis in Section 5.4.1.

Conditional probability tables

The final BCN, shown in Figure 5.4.2, is parametrised by computing the CPTs associated with each variables. The ratio with the unconditional probability CPT/P are shown for the variables VB,

X	Y	given Z	χ^2	p-value	$LS_{blind}\%$
PJO-1	vT-flux	ENSO	160	10^{-9}	4.7
ENSO	vT-flux	PJO-1	34	$2.9 \cdot 10^{-5}$	0.9
PJO-1	PJO-2	vT-flux	529	10^{-9}	23.0
vT-flux	PJO-2	PJO-1	851	10^{-9}	33.3
PJO-2	SPV-low	vT-flux, ENSO	131	10^{-9}	11.3
vT	SPV-low	PJO-2, ENSO	105	10^{-9}	7.3
ENSO	SPV-low	vT-flux, PJO-2	55	0.02	5.3
vT-flux	VB	SPV-low	74	10^{-9}	3.7
SPV-low	VB	vT-flux	426	10^{-9}	17.5
SPV-low	EDJ-1		847	10^{-9}	27.7
EDJ-1	EDJ-2	VB	65	10^{-9}	4.0
VB	EDJ-2	EDJ-1	655	10^{-9}	25.9

Table 5.4.2: All BCN links $X \to Y$ as shown in Figure 5.4.2. χ^2 and p-values are obtained from $\chi^2(X, Y \mid Z = PA(Y) \setminus \{X\})$, thus using the final parents shown in Fig. 5.4.2. $LS_{blind}\%$ is defined in the text.

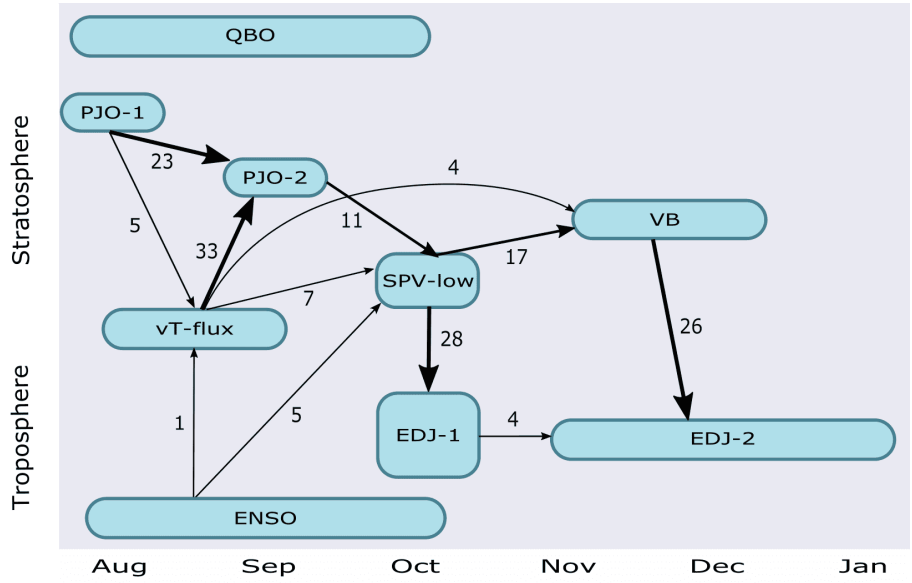

Figure 5.4.2: The BCN structure that emerged from the conditional independence tests (Step 1 and Step 2). Links' width and labels indicate the associate percentage Blind Average Link Strength, $LS_{blind}\%$, defined in the text.

EDJ-1 and EDJ-2 in Figure 5.4.3 (others are found in Appendix 5.F). A ratio larger (smaller) than 1 indicates a more (less) likely state of the target compared to climatology, indicated with a blue (pink) colourscale respectively.

The CPT ratios shown in confirm, but also quantify, known underlying physical relationships, as described below.

- VB: The strength of SPV-low in the period leading up to the vortex breakdown has a strong influence on the timing of the event: a small/large SPV-low tends to be associated with an early/late vortex breakdown date since the strength of SPV-low is correlated with the strength of the polar vortex higher up at 50 hPa where the VB is defined. Independently, the vT-flux forcing acts on the vortex as well, with a strong vT-flux favours and early VB (and vice-versa); this is best seen when SPV-low is neutral.

- EDJ-1. A large lower stratospheric signal in October (large SPV-low) is associated with a more poleward reach of the poleward shift of the EDJ (poleward EDJ-1), and vice-versa. This downward effect reflects evidence of model skill in forecasting the poleward transition of the EDJ based on a model initialisation in early August (Seviour et al., 2014) and is also

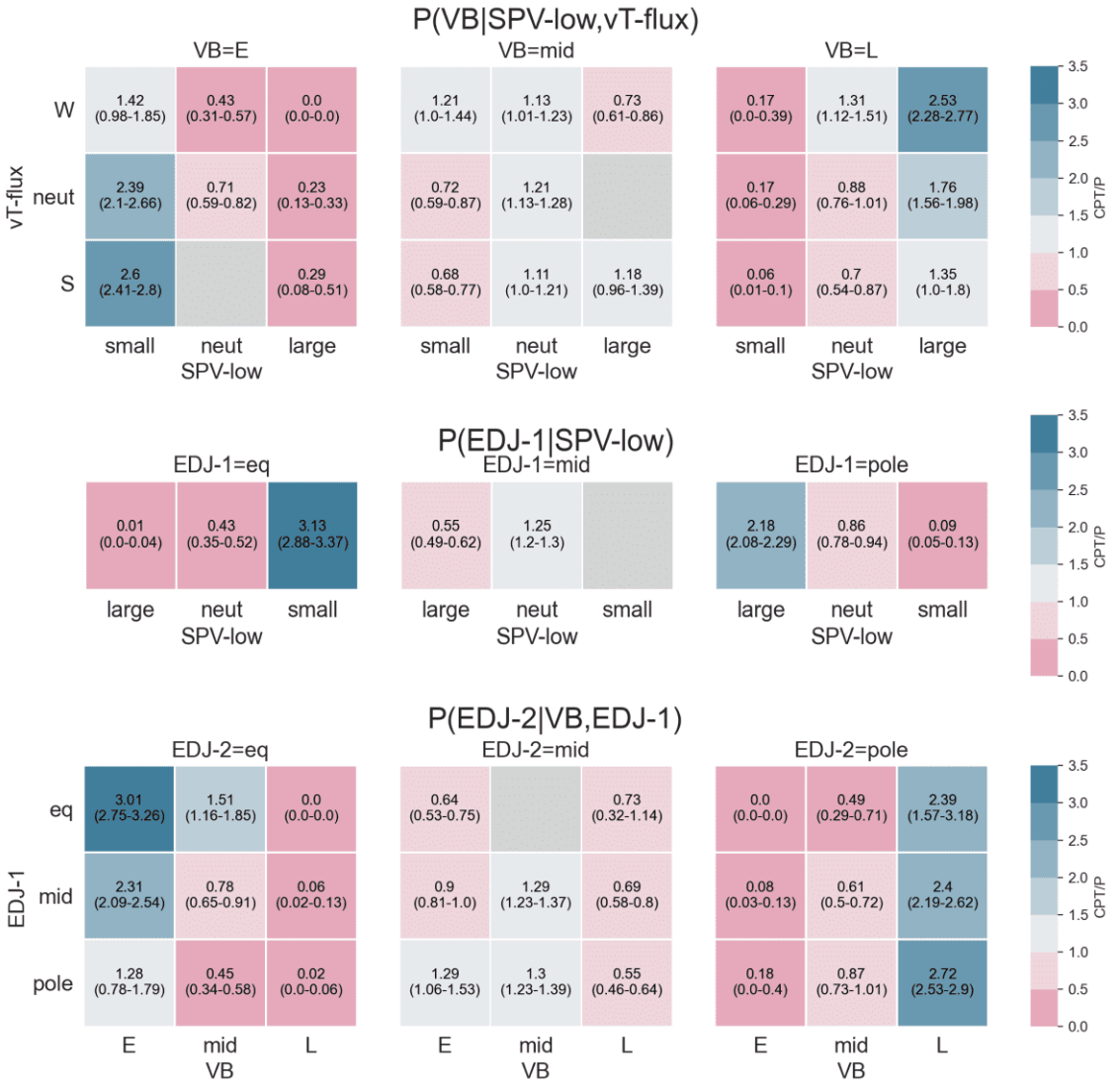

Figure 5.4.3: CPT/P for VB (top), EDJ-1 (mid) and EDJ-2 (bottom). Each category of the child variable X is associated with one table, and each of the possible combinations of categories of its parents is associated with one column (or column and row, if the parents are two) of each table. The values written on each table's entry show the mean and the 5-95 percentile ranges (in parenthesis) of a 1,000 samples bootstrapped estimate. Each bootstrap is obtained by re-sampling the ensemble members and years. The color blue (pink) of each entry indicates if the mean ratio is higher (lower) than 1, that is if the target category is more (less) likely than climatology given the conditions of its parents. The minimum possible value is 0 and the maximum here is around 3. Entries are masked with grey if the mean ratio falls between 0.9-1.1.

observed in reanalysis (Byrne and Shepherd, 2018).

- EDJ-2. An early/late vortex breakdown event typically precedes an earlier/later equatorward shift of the EDJ in November-December (Black and McDaniel, 2007; Byrne et al., 2017; Byrne and Shepherd, 2018), as captured by the EDJ-2 categories (recall Fig. 5.2.2(c)). The state of EDJ-1 also plays a role, although minor in terms of shift in probability, with a EDJ-1 equatorward/poleward generally favouring an EDJ-2 equatorward/poleward.

Cross-validation

The ability of the BCN to reproduce the hindcast variability is assessed computing its predictive accuracy on out-of-sample hindcast data (cross-validation). In the absence of a simple metric such as the mean squared error, the cross-validated accuracy is an alternative way to assess the "goodness" of the BCN fit to the data (Harwood et al., 2021). A stratified 30-fold cross-validation is performed. Stratification allows to consider equal numbers of points falling in a category across the various test-train set (needed for imbalanced categorical data). The parents are used as evidence for the prediction of each target variable. For each iteration, the accuracy is quantified with the ROC AUC.

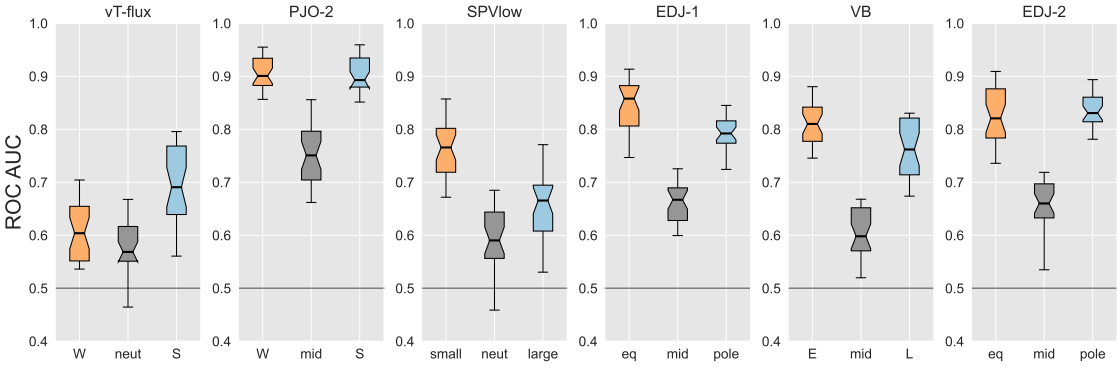

Figure 5.4.4: Stratified 30-fold cross-validated ROC AUC, computed from predictions of each variable using its parents as deterministic evidence. The 30 values, one per each cross-validation iteration, are shown in a box plot where the box extends from the 25-75 percentiles and whiskers show the 5-95 percentiles. Different colours highlight the different categories of each target variable.

The results of the cross-validation are shown in Figure 5.4.4. Despite the few mechanisms modelled by the network, the BCN reproduces the hindcast dynamics significantly better than a

climatological forecast (AUC much higher than 0.5), especially for the the extreme categories. The performance is especially good for PJO-2, EDJ-1 and EDJ-2 (median between 0.8 to 0.95), followed by VB, SPV-low (0.65-0.8) and finally vT-flux (0.6-0.7). A small SPV-low, corresponding to a weaker and less deep lower stratospheric vortex in October, is better predicted than a large SPV-low. Similarly, the early VB seems slightly more predictable than the late VB. Note that vT-flux is predicted with some skill in the strong state (i.e. when forced by the tropospheric sources of waves, such as El Niño) but not for weak or neutral states.

The central category (grey) is generally the less skilfully predicted of the three. A possible reason for this is that the central category includes values close to both extremities and thus may mix instances where the actual state was better represented by one of the extreme categories. Another reason could be that the central category is associated with a smaller signal to noise ratio: a prediction of an extreme is more likely to happen due to an actual signal, while prediction of a central category may more often happen due to noise. Given that the tails of the variables distributions are usually the ones associated with stronger impacts, it is encouraging to see that their performance is generally good.

Note that the variables with higher ROC AUC, PJO-2 and the two EDJs, are also the ones with the highest *LS%* incoming links (Table 5.4.2). This confirms that here LS% can be used to highlight which links have a more consistent signal, meaning that the state of the parent is very influential in determining the state of the child.

The results for SPV-low suggest that there are potentially missing mechanisms that the BCN is not capturing. The reasons for a worse skill for SPV-low (and to a lesser extent for VB) can be threefold. First, the wrong long-lead precursors may have been identified through a bad choice of August-September variables; however, given the extensive literature on the role of PJO and vT-flux this seems unlikely. However it is also known that strong vT-fluxes in the lower stratosphere may not drive a weaker vortex and the wave activity could just propagate and break farther equatorward, thus not providing enough predictability to the SPV-low and VB. Second, relevant additional stratospheric precursors may currently be missing, e.g. QBO and ozone. Third, shorter time scales than the ones represented in the network may be responsible for the stratospheric evolution in October-November, for example weekly dynamics may be needed to represent the shift in probability of SPV-low. The last hypothesis agrees with the fact that NH and SH stratospheric extreme events are

difficult to predict weeks in advance, at least deterministically speaking, and this includes early and late final warming events represented by VB. This challenge remains despite significant progress in stratospheric representation including higher model lids and increased stratospheric resolution (e.g. see Butler et al., 2016; Domeisen et al., 2020a). The poor prediction skill for SPV-low may represent a "bottleneck" for the network, as the selected variables do not allow for information of the September variables (PJO-2 and vT-flux) to effectively flow to the October and November stratosphere, and ultimately to the troposphere.

5.4.2 Predictability arising from long-lead drivers of EDJ

At the heart of the present work is the question of how much of the interannual variability of the EDJ can be predicted with some skill given its long-lead precursors identified in the literature. A variable X (driver) will be considered a source of predictability for variable Y (target) if having information on X results in a reduced uncertainty in the outcome of Y, here quantified with improved forecast verification score.

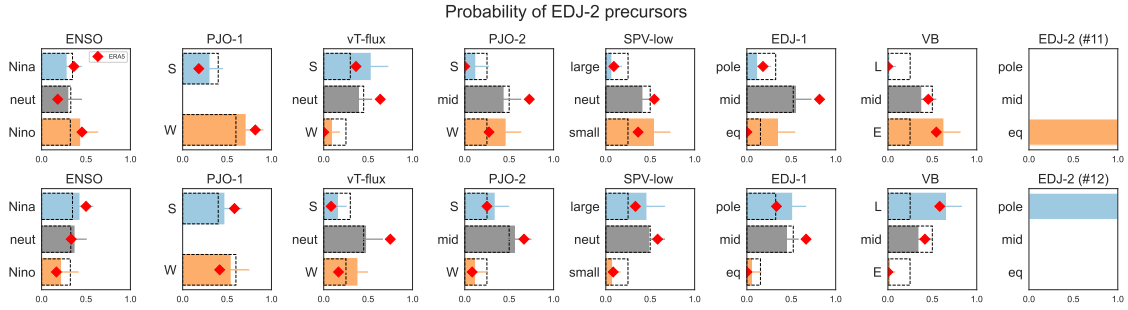

Figure 5.4.5: Probability distribution of EDJ-2's precursors. The hindcast's distributions are estimated with a bootstrap procedure of 1,000 subsamples of M random members/year, with M the number of the specific target's category observed in ERA5 between 1981-2016. The length of the horizontal bar is the median value, the error bar covers the 10-90 percentiles and the black dashed bar is the climatological probability value. The red diamonds shows the observed frequencies of the precursors, estimated from ERA5 data in the hindcast period 1981-2016, shown for validation. Different colours highlight the different categories of each target variable.

Before looking at predictions, it is interesting to inspect the conditional probability distribution of the precursors (X) of the target variable EDJ-2 given the latter, namely $P(X|\text{EDJ-2})$ when EDJ-2 is either equatorward (top) or poleward (bottom) (Figure 5.4.5). Since X precedes EDJ-2 in

time, this means looking "backward in time". The conditional probabilities of EDJ-2 precursors are significantly different from the climatological distribution, even for ENSO, PJO-1 and vT-flux. These results suggest there is *potential* for predictability of EDJ-2 based on these long-lead precursors, up to 3 months in advance. The same analysis is presented for the precursors for SPV-low, VB and EDJ-1 in Appendix 5.G, and similar conclusions apply.

This potential predictability is put to the test by computing predictions of EDJ-2, as well as of SPV-low, VB, EDJ-1 given various combinations of long-lead precursors. This time therefore looking at predictions forward in time. Each ensemble member is used to provide evidence and verification data, for a total of 51 predictions. Each forecast is performed for different evidence sets: ones that could be observed by August (ENSO, PJO-1), by September (ENSO, PJO-2, vT-flux), by October in the stratosphere only (vT-flux, SPV-low) and October in both stratosphere and troposphere (vT-flux, SPV-low, EDJ-1). These are all the variables in the network up to and including the stated month, except the ones that would be redundant for the BCN prediction given Eq. (5.1). The forecasts are assessed with two metrics: the ROC AUC and the PR AUC %.

The results are shown in Figure 5.4.6, with box plots showing the distribution of the 51 predictions. Generally speaking, both ROC AUC and PR AUC given evidence set in August and September are higher than the no-skill values, with median values generally between 0.6-0.7 for ROC, and 20-40% for PR. This is true also for the target variables VB and EDJ-2 which are defined in the months of November-December, therefore with a 2-3 months lead. Part of the behaviour of the hindcast variability can be reproduced using solely long-lead predictors, suggesting there is indeed some long-lead predictability based on these variables. A second point to note is that both ROC and PR metrics show a marked improvement when evidence is provided for months closer to the target, especially for 0-1 months lead, reaching median values between 0.7-0.9 for ROC, and 40-60% for PR. EDJ-1 and EDJ-2 have the largest increase in AUC given stratospheric precursors in October-stratosphere, confirming the importance of a skilful stratospheric forecast for a skilful tropospheric forecast.

Yet the metrics are far from their maximum values, especially for PR, and the large variability across different predictions suggest that the precursors included in the network explain only part of the target's dynamics. As noted in the cross-validation, SPV-low is the target that is harder to predict given the identified precursors. Also here the central categories have lower AUC compared

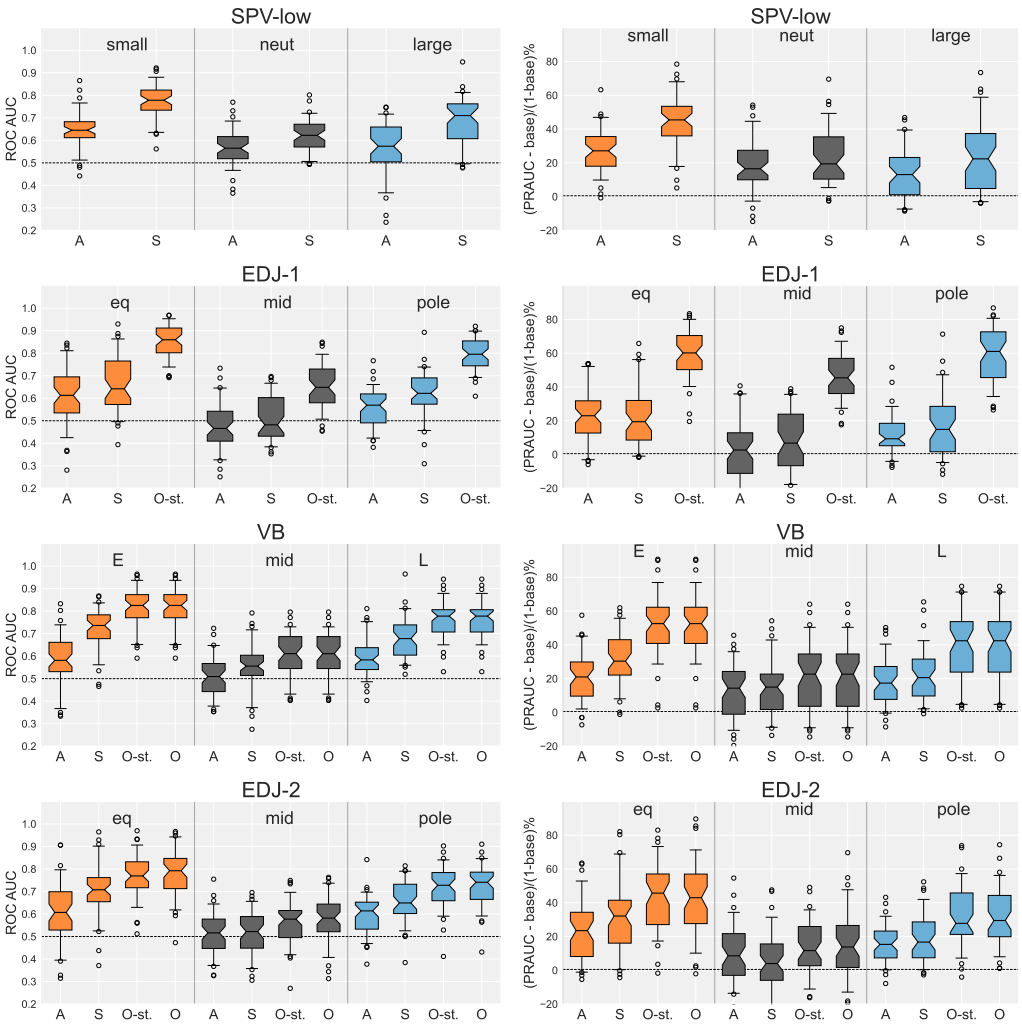

Figure 5.4.6: ROC AUC (left) and PR AUC (%)(right) for each category of the targets SPV-low, EDJ-1, VB and EDJ-2 (from top to bottom). BCN predictions are performed using different sets of variables given as evidence (August, September, October-stratosphere and October, as described in the text). The distribution of the 51 predictions (one for each ensemble member) are shown as box plots, which cover the 25-75 percentiles (boxes) and the 5-95 percentiles (whiskers). Different colours highlight the different categories of each target variable.

to the "extreme" categories. Finally, the spread across the 51 predictions is very large hence individual members are predicted much better or worse than others.

Overall, this analysis shows that the long-lead precursors included in the network do provide non-negligible information to both stratospheric and tropospheric late-spring targets: their inform-

ation is sufficient to shift the probability distribution of the targets away from climatology in the right direction (i.e the one modelled by the hindcast). However, the information provided by long-lead drivers is much less effective compared to knowledge of the short-lead precursors, especially for the stratospheric variables used to predict the EDJs anomalies.

5.4.3 Skill of BCN in predicting observed variability

The BCN is here tested against observed variability, as represented by the ERA5 reanalysis, and compared with the skill of the ECMWF forecast model. The experimental set up is as follows. BCN predictions are produced by setting the evidence variables S to the categories observed in ERA5, and producing a probabilistic forecast for each year between 1981 and 2016. Four groups of variables S are considered as in Section 5.4.2, corresponding to the ones observable in August (ENSO, PJO-1), September (ENSO, PJO-2, vT-flux), October stratosphere (vT-flux, SPV-low), October stratosphere-troposphere (vT-flux, SPV-low, EDJ-1) and Parents (target-specific). A corresponding conditional hindcast (CH) is built by selection of those of the 51 members whose variables S match their corresponding ERA5 categories in the specific year. The conditional hindcast can be seen as a filtered version of the full hindcast made of those members whose trajectories "pass through" the observed states of all the S variables. Therefore, the conditional hindcast has a comparable set of "initial conditions" to the BCN's. However, while the network only uses the values for those S variables as input, the conditional hindcast has access to all the information of all other parts of the atmosphere-ocean system for the selected ensemble members. Both predictions are evaluated against ERA5 with the ROC AUC (Figure 5.4.7) and PR AUC % metrics (in Appendix 5.H).

Figure 5.4.7 shows the comparison between the BCN forecast (BCN), the conditional hindcast (CH) and the full hindcast (H) in terms of ROC AUC for each variable SPV-low, EDJ-1, VB and EDJ-2. Each panel shows the skill obtained for the three categories of the target variable (sub-panels). The skill of the H, CH and BCN are the top, middle and bottom row of each sub-panel respectively. The evidence set used constitute the columns of each sub-panel (August, September, October-stratosphere, October, Parents). The colour of the heatmap scales with the mean AUC value, with darker red corresponding to higher skill, and no skill coloured in grey. The results for PR AUC % are shown in Appendix 5.H and they present similar features to the ROC AUC.

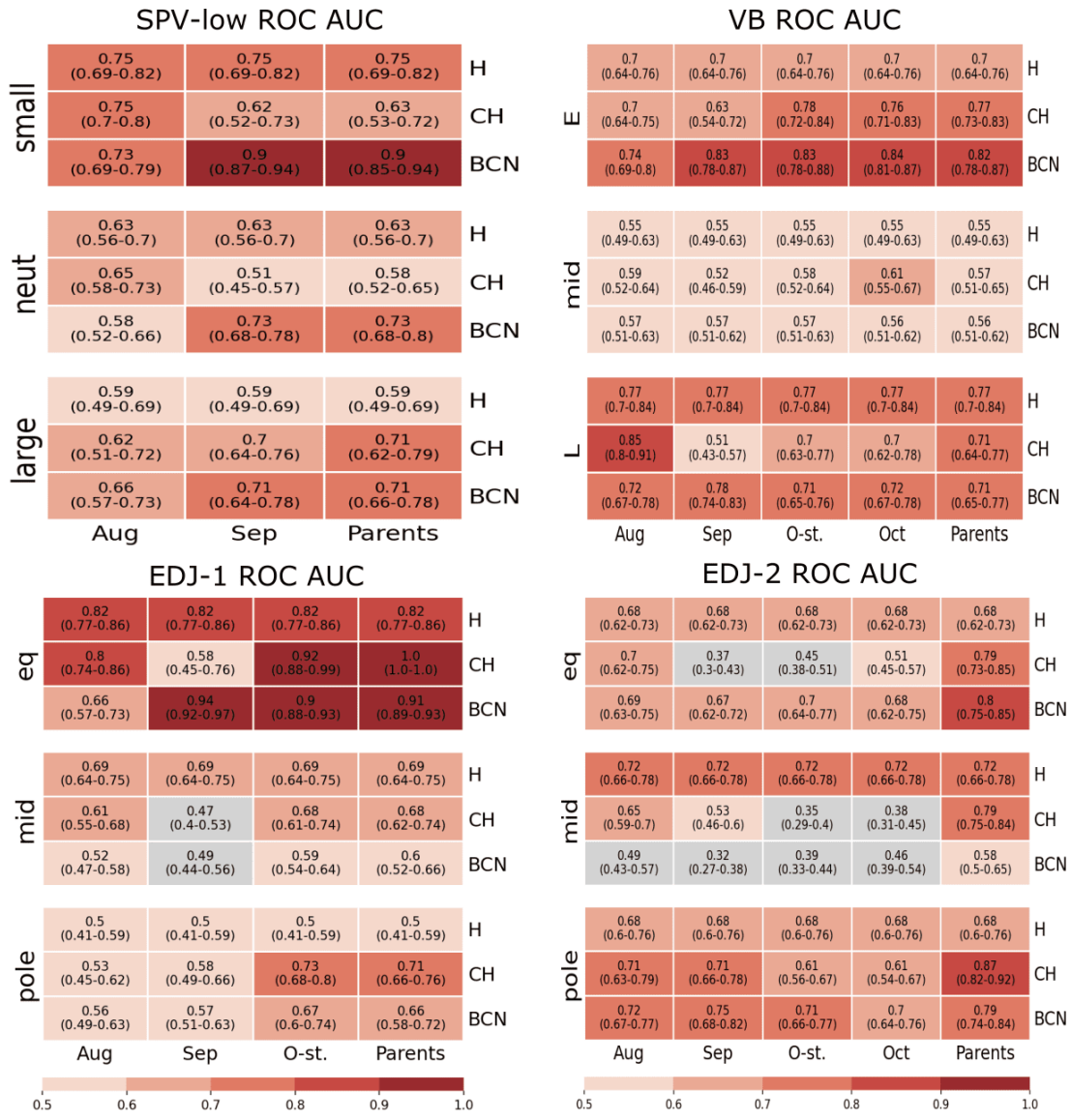

Figure 5.4.7: ROC AUC tables obtained by the BCN forecast (BCN), the conditional hindcast (CH) and the full hindcast (H) assessed against ERA5. Variables SPV-low, EDJ-1, VB and EDJ-2 are the targets (panels). Each table shows the three categories's skill (sub-panels). The skill of the hincast is the top row, for the conditional hindcast is the middle row and for the BCN is the bottom row of each sub-panel. The evidence set used (August, September, October-stratosphere, October, Parents) constitute the columns of each sub-panel. The colour scales with the mean AUC value, with darker red corresponding to higher skill, and no skill coloured in grey. The numbers in each table entry shows the mean and 25-75 percentiles (in parenthesis) obtained via a 1,000 samples bootstrap procedure.

Firstly, note that both CH and BCN tend to produce skilful forecasts (few grey colors), especially for the extreme categories. Secondly, the BCN skill is generally equal to or higher than the CH (darker colours) meaning that the BCN, with its simple parametrization of the connections via CPTs, is as good as the conditional hindcast at representing the observed dynamics. Such a similar skill is to be expected on average for the Parents evidence set, by construction of the CPTs, but is not guaranteed for the others. In some cases, for example SPV-low small and VB early, the BCN skill is higher than that of the CH. This suggests that the BCN is likely encapsulating the predictable signal more effectively. If the hindcast comprises a signal term (predictable component) and a noise term (unpredictable component), then the BCN is mostly signal since it is estimated from a large dataset that filters out the noise : for given values of the parents, the BCN always gives the same probabilistic prediction for the child variable. The third result is that the CH is generally more skilful than the H (darker colour; note H has the same value for each evidence set since it has not been sub-sampled as the CH). This suggests that the mechanisms schematised here provide meaningful rules to identify ensemble members that have a higher likelihood of evolving in time towards the correct target category. As expected, the largest increase happens when the conditioning variables are after September (i.e. one month after initialization), since by that time a large spread across ensemble members develops and thus the conditioning can be more effective. Finally, the equatorward EDJ-2 in the CH forecast shows a peculiar feature. Its skill is high given August conditions, then slightly deteriorates given September to October conditions (grey color), and improve drastically again given the Parents. This is also seen for the other two categories, but less markedly. This suggests that the ensemble members filtered based on the September (ENSO, PJO-2 and vT-flux) to October conditions (vT-flux, SPV-low and EDJ-1; note no VB included) do not tend to evolve towards a consistent and correct EDJ-2 anomaly, and therefore constitute "bad filters". On the other hand, members filtered based on the Parents (VB and EDJ-1) improve the skill. The fact that the CH given August filter is relatively high points at the role of ENSO in offering an effective pre-condition to EDJ-2, which may persist in practice but which signal may get lost through looking only at the September and October variables.

Overall, the analysis presented in Figure 5.4.7 demonstrates that the BCN can forecast ERA5 variability as well as the conditional hindcast, and generally better than the hindcast initialised on 1 August. Note that the BCN long-lead prediction of observed SPV-low and VB could be

further improved if the connection to the QBO and to the late spring stratospheric ozone were to be included. This has not been possible due to their insufficient representation in System 4 (an in many other forecast systems), but could be achieved with another dataset used as basis for the BCN parametrization (e.g. a model with improved ozone representation as in Oh et al., 2022; Monge-Sanz et al., 2022).

5.4.4 Connection between BCN variables and surface climate

EDJ-1, EDJ-2 and ENSO variables provide a route for connecting the probabilistic BCN forecasts with the expected surface climate response. This is inspected here with Z500, 2mT and MTPR. The location of the EDJ is known to have a large effect on the mid-to-high latitudes surface climate, given its association with storm tracks and low pressure systems, and ENSO further and independently modulates the large scale surface dynamics and regional climate.

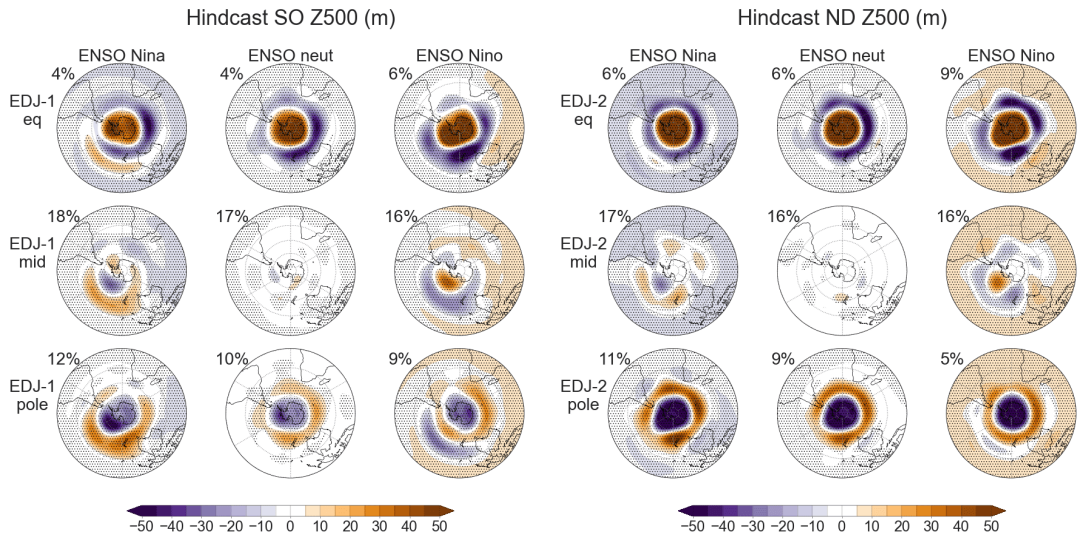

Figure 5.4.8: Southern Hemisphere hindcast geopotential height anomalies at 500 hPa (Z500). SO mean composites with respect to EDJ-1 and ENSO (left), and ND composites with respect to EDJ-2 and ENSO (right). The dots show areas where anomaly values are significantly different from 0 according to a one-sample two-sided t-test at 0.1 significance level. A stereographic projection is used, and latitudes are shown at intervals of 20º starting at the equator.

Confirming this, Figure 5.4.8 shows the Z500 composites of the combined states of ENSO and EDJ-1 in September-October (SO; left) and of EDJ-2 and ENSO in November-December (ND;

right). The SO Z500 composites show how ENSO modulates the otherwise very zonally symmetric effect of EDJ-1. The isolated effect of one variable can be more clearly seen when the other is in the 'central' state. When EDJ-1 is neutral, the PSA-like wave train pattern associated with ENSO anomalies emerges clearly as a zonally asymmetric pattern of anomalies (Section 5.1.1). Conversely, when ENSO is neutral the zonally symmetric pressure anomalies associated with the EDJ-1 westerlies clearly stand out. In all the other combinations, the EDJ-1 pattern is modulated to various extents by the PSA-like wave train, resulting in a slightly asymmetric geopotential anomaly pattern.

The ND Z500 composites based on ENSO and EDJ-2 show a very similar pattern to the above. When EDJ-2 is neutral, the PSA-like pattern in the South Pacific is still clearly visible as in SO, but now in ND the anomalies extend into the South Atlantic and Indian Ocean. The ENSO-driven pattern is still zonally asymmetric even if heading into summer, a time when ENSO's response is generally considered to be zonally symmetric. When ENSO is neutral, once again the zonally symmetric pattern due to the latitudinal position of the jet is seen, and its signature is very similar to EDJ-1 in SO but the anomalies are larger, especially in the south Pacific sector. Notably, the EDJ-2 driven annular pattern is very similar to the one associated with early/late vortex breakdown, as shown in Figure 5.I.1 in Appendix 5.I. Given the strong causal association VB $\rightarrow$ EDJ-2, documented also in this analysis, the Z500 annular pattern in ND can be considered as mostly driven by VB.

Overall, ENSO's fingerprint from September to December emerges as a zonally asymmetric pattern that can be, to a first approximation, summarised as the propagation of a PSA-like wave-2 train, which is overall the dominating asymmetric response due to ENSO (Osman et al., 2022). The combined effect of ENSO and EDJs on Z500 is close to a linear superimposition. For example, Z500 (eq, Niña) $\simeq$ Z500 (eq, neut) + Z500 (mid, Niña). Namely the patterns at the 'corners' of the panels can be seen as the sum of the anomalies in their immediate horizontal and vertical neighbouring plots.

The ND MTPR patterns associated with the EDJ-2 and ENSO composites are shown in Figure 5.4.9 (left) (for EDJ-1 composites see Figure 5.I.2 in Appendix 5.I). Focusing on the ENSO neutral maps, the MTPR composites show a clear pattern of wetting around the location of the jet and drying south/north of it when EDJ-2 is more equatorward/poleward, giving rise to a two-band

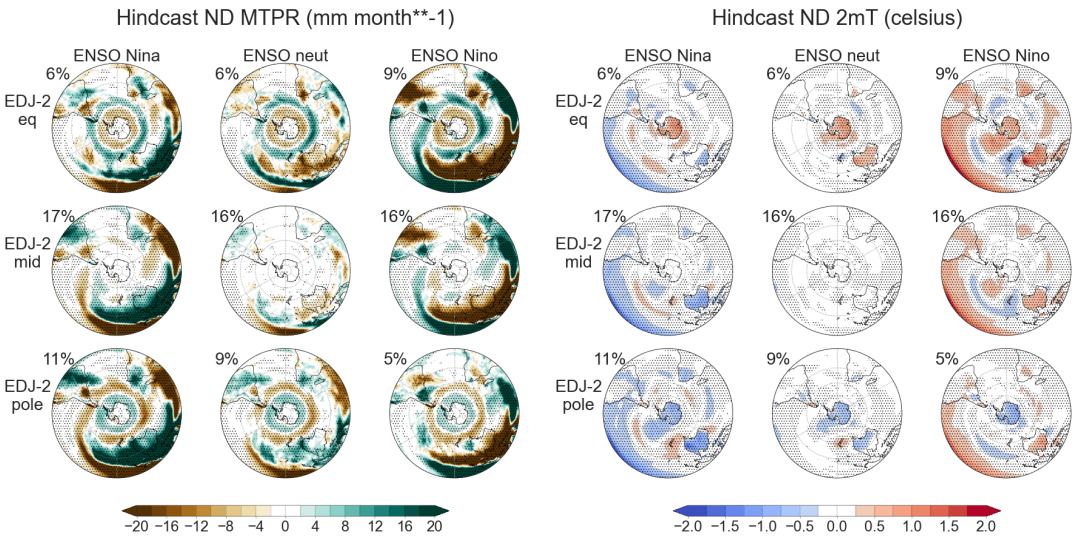

Figure 5.4.9: Southern Hemisphere hindcast precipitation (MTPR, left) and 2 meters air temperature (2mT, right) anomalies in ND. Mean composites with respect to EDJ-2 and ENSO. The dots show areas where anomaly values are significantly different from 0 according to a one-sample two-sided t-test at at 0.1 significance level. A stereographic projection is used, and latitudes are shown at intervals of 20º starting at the equator.

pattern of opposite anomalies around Antarctica. The effect of EDJ-2 seems to extend to Brazil, South Africa and south-east Australia, changing the sign of the MTPR anomalies depending on its position. To a large degree, the pattern due to EDJ-2 reflect the effects of the SAM. When EDJ-2 is in the central state (close to climatology), ENSO's teleconnections can be seen to dominate the strong wet/dry signal in the tropical Indian and tropical-subtropical west Pacific Ocean, as well as in South America, South Africa and south-east Australia and New Zealand. Focusing on these land regions, the combined effect of ENSO and EDJ-2 is 'in phase' and reinforcing the signal due to ENSO when EDJ-2 is equatorward and there is an El Niño, which brings drier conditions to south-east Australia and South Africa (or when EDJ-2 is poleward and La Niña, which brings wetter conditions). In the 'out-of-phase' combination, ENSO's signal still tends to dominate the effect over these land regions, notably except for south-eastern Australia where the EDJ-2 signal prevails.

The ND 2mT composites clearly show ENSO's fingerprint in the Pacific and east Indian Ocean and in south Africa and South America (Fig. 5.4.9, right). The position of the EDJ-2 sets the sign

of the temperature anomalies over and around Antarctica and influences Australia, New Zealand and South Africa. As for the MTPR, the EDJ-2 influence over the mid-latitude land areas can be reversed if ENSO is out-of-phase (Niño with poleward, or Niña with equatorward) or strengthened when in-phase (opposite combination).

To establish our confidence in the hindcast's representation of the surface signals, the composites are compared with ERA5 in Figures 5.J.1 to 5.J.3 in Appendix 5.J. Composites are computed for one variable at a time, due to sample size limitations in ERA5. Overall there is a good agreement and no major biases are found; the otherwise occurring EDJ bias in ND is removed by means of bias-correction of the hindcast EDJs variables. The teleconnection patterns of ENSO and the local effect of EDJs are however much clearer in the hindcast composites, thanks to the much larger sample size.

5.4.5 BCN-guided selection of most skilful forecast ensemble members

The analysis in Section 5.4.4 suggests that skilful predictions of EDJs could be be used to enhance the representation of surface climate. Methods for selecting the most skilful members from an ensemble, sometimes called a 'refined ensemble', have been recently developed for the study of seasonal forecasts. Dobrynin et al. (2018) used a refined ensemble, selected through an NAO filter, to improve dynamical seasonal predictions of boreal winter; a similar analysis has been performed by Polkova et al. (2021). A similar concept of targeted (re-)sampling is increasingly used to study rare climate events (Bouchet et al., 2019), such as heatwaves, by means of filtering ensemble forecasts to privilege members that exhibit extreme statistics (Ragone et al., 2017; Webber et al., 2019).

Given the improved BCN skill of EDJs compared to the hindcast (Section 5.4.3) and given that combinations of ENSO and EDJ categories clearly influence the surface climate, especially in the mid-to-high latitudes (Section 5.4.4), it is reasonable to expect that BCN predictions of EDJs can lead to an improved representation of the surface climate compared to the hindcast ensemble mean.

In this section the EDJ-1 and EDJ-2, as predicted by the BCN, are used to select a refined ensemble to improve the hindcast representation of surface climate conditions. First, for each year between 1981 and 2016 the BCN is initialised with ERA5 evidence for the EDJs parents, and a probabilistic forecast of EDJ-1 and EDJ-2 is derived. Then, each probabilistic forecast is translated

into a single category selecting the one with the largest relative increase in probability with respect to its climatological prevalence, that is $k^* = \max_k\{\frac{p_i(k)-p(k)_{climat}}{p(k)_{climat}}\}$. For each year only the members with EDJ category matching k^* are selected. Finally, the Z500, 2mT and MTPR composites of this refined ensemble are inspected with grid-point anomaly correlation with ERA5 across the modelled years.

The anomaly correlation for the SO (ND) mean surface variables filtered via EDJ-1 (EDJ-2) are shown in Figure 5.4.10 (Figure 5.4.11). In each figure, the Z500, 2mT and MTPR correlation are shown (top to bottom) for latitudes from 20º to 90ºS. The leftmost maps show the correlation between the full hindcast ensemble mean and ERA5 (darker blue for higher correlations). The central maps show the change in correlation between the BCN-guided (BCN) refined ensemble and the full ensemble (green for increased and purple for a decreased correlation). The leftmost maps show the change in correlation between a refined ensemble based on the best possible BCN forecast (perfect BCN) and the full ensemble. The perfect BCN selection takes only those ensemble members whose EDJ-1 (or EDJ-2) categories match ERA5, hence showing the best possible improvement based on the BCN approach.

Figure 5.4.10 shows a marked increase in the SO Z500 correlation for the BCN-filtered ensemble: especially over Antarctica, in a zonal band over the southern Indian Ocean sector and over New Zealand. Note that the ensemble mean correlation over Antarctica and Indian Ocean is originally not significant, but becomes so thanks to the filter. The BCN is however also associated with a decrease in correlation over the South America and Atlantic sector. A perfect BCN improves further the correlation over the Indian Ocean and Antarctica, and does not decrease much the correlation anywhere. The SO 2mT and MTPR correlation are improved by the BCN as well, although less clearly than for Z500: the signal is stronger over Antarctica for 2mT, and in the southern Indian Ocean for MTPR where a zonally symmetric pattern of improved correlation emerges following the EDJ-1 shape. Figure 5.4.11 shows also a marked increase in the Z500 ND correlation of the BCN-guided ensemble, this time also over southern South America and in the south Pacific Ocean, in addition to Antarctica and the southern Indian Ocean. The perfect BCN improves further and almost in the whole area analysed, confirming the dominant role of EDJ-2 on Z500 in this time of the year. The 2mT ND BCN correlation increases in the ocean around Antarctica and in southern South America, and MTPR improves in the same region as well. 2mT and MTPR filtered also

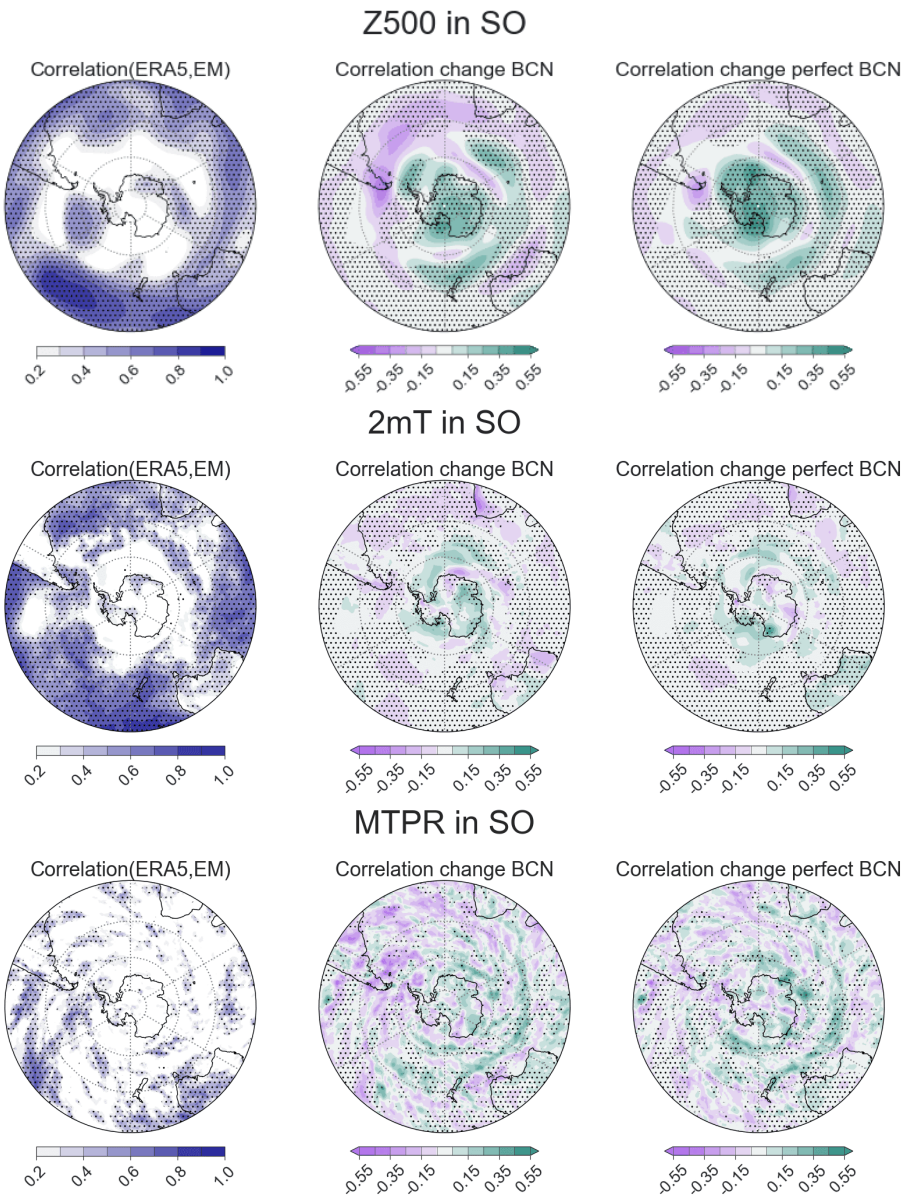

Figure 5.4.10: Z500 (top), 2mT (middle) and MTPR (bottom) correlation maps in September-October based on EDJ-1 filter between 20º and 90ºS. Correlation between hindcast ensemble mean and ERA5 (left). Change in correlation for the BCN refined ensemble compared to the full ensemble (middle), with green color marking an improvement and purple a decrease in correlation. Change in correlation for the perfect BCN refined ensemble compared to the full ensemble (middle). Black dots show anomaly correlation significantly different from 0 according to a one-sample two-sided t-test at 0.1 significance level. A stereographic projection is used, and latitudes are shown at intervals of 5º starting at 20ºS.

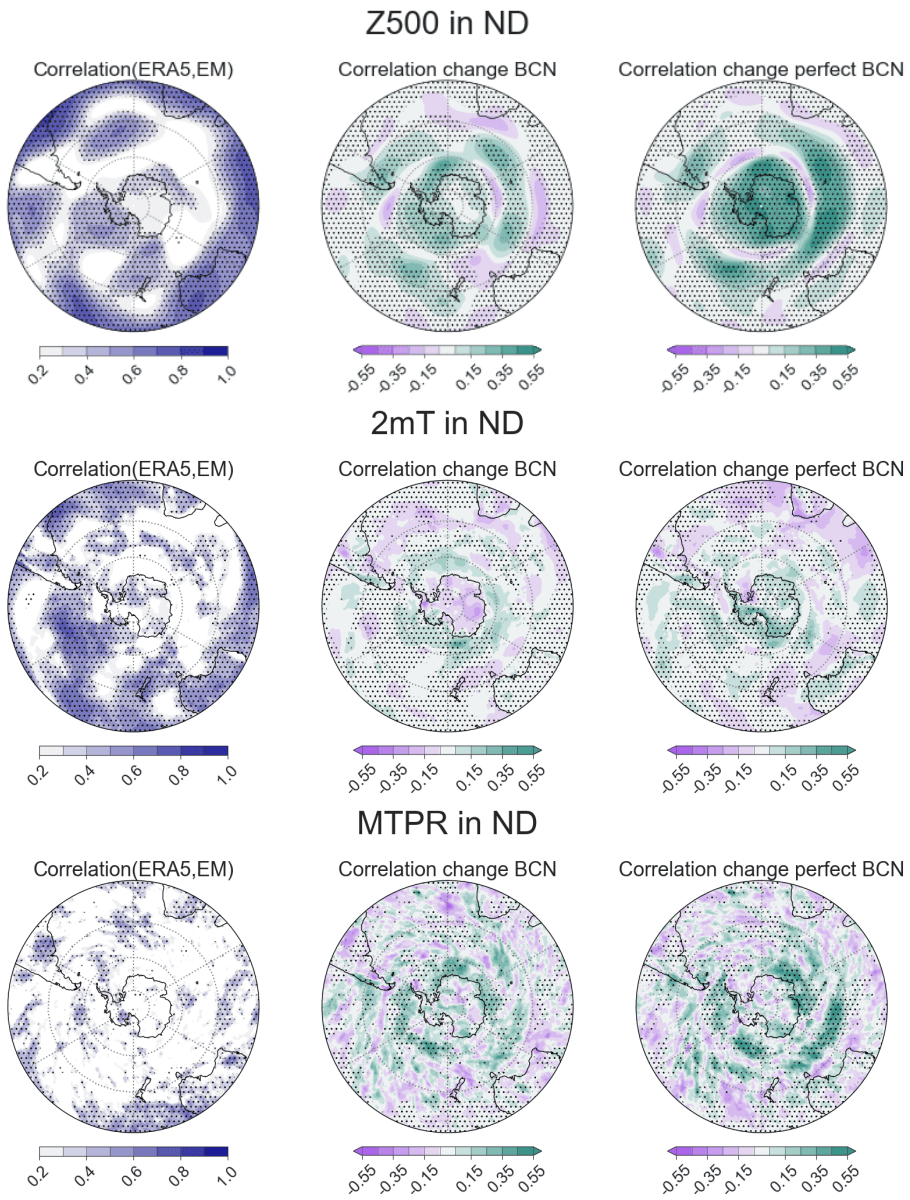

Figure 5.4.11: Z500 (top), 2mT (middle) and MTPR (bottom) anomaly correlation maps in November-December based on EDJ-2 filter between 20º and 90ºS. Correlation between hindcast ensemble mean and ERA5 (left). Change in correlation for the BCN refined ensemble compared to the full ensemble (middle), with green color marking an improvement and purple a decrease in correlation. Change in correlation for the perfect BCN refined ensemble compared to the full ensemble (middle). Black dots show anomaly correlation significantly different from 0 according to a one-sample two-sided t-test at 0.1 significance level. A stereographic projection is used, and latitudes are shown at intervals of 5º starting at 20ºS.

improve the significance of the correlation in those areas. The perfect BCN shows an even stronger increase in correlation, particularly in the Southern Ocean south of Australia, where the ensemble mean has no skill at all.

The effect of filtering the hindcast ensemble based on BCN forecast is of an overall improvement, and especially in the areas where the EDJ anomalous signal is most influential, that is over and around Antarctica, and for some variables also in the southern parts of South Africa, South America and east Australia.

An improved representation of the surface climate by the refined ensemble means that the selected members have a larger probability of EDJ being true positives, as well as a smaller probability of false positives and false negatives (especially for categories at the opposite side of the variability, i.e. equatorward vs poleward). Table 5.4.3 summarises the *Probability of Correct Prediction* and *Probability of Opposite and Wrong Prediction* for EDJ-1 and EDJ-2. The probability of correct prediction for EDJ-1 and EDJ-2 of the refined ensemble (4th column) is indeed higher than for the full ensemble (3rd column). Possibly more importantly, the probability of wrongly predicting EDJ as equatorward when an opposite poleward anomaly is observed (or vice-versa) decreases drastically for the refined ensemble (7th column) compared to the full hindcast (6th column). Note that the table shows also the expected values of these probabilities for a climatological prediction (i.e. due to chance; 2nd and 5th columns) and the BCN performs better than those as well; note this is not for the full hindcast Probability of Opposite and Wrong Prediction. For completeness, the same results for SPV-low and VB are shown in Table 5.K.1.

	Probability of Correct Prediction			Probability of "Opposite" Prediction		
target	climatology	full hindcast	refined with BCN	climatology	full hindcast	refined with BCN
EDJ-1	39%	36%	**56%**	10%	14%	**3%**
EDJ-2	39%	50%	**56%**	11%	14%	**0%**

Table 5.4.3: Probability expressed in percentage of correct predictions (true positives averaged over all categories) and of wrong predictions (only considering the two categories at the opposite side of the distribution; for example prediction EDJ-2=eq when observed EDJ-2=pole). The forecasts considered are of climatology, full hindcast and refined hindcast with BCN given evidence for the parents. Bold probabilities indicate if BCN performs better than both the full hindcast and the climatology. Calculations for the climatology are in Appendix 5.K.

5.5 Discussion

A Bayesian Causal Network has been used to schematise and quantitatively summarise the probabilistic progression of major components of the SH late-winter to early-summer atmospheric variability as represented by the ECMWF System 4 hindcast initialised on the 1 August.

Despite its simplicity, the BCN can reproduce a significant part of the variability of the system. It does so especially well for the connection between the stratospheric drivers and the two phases of the tropospheric eddy-driven jet (EDJ) semi-annual oscillation (Figure 5.4.4). The analysis supports the hypothesis of a dominant role for the strength of the lower stratospheric polar vortex in influencing the amplitude of the poleward migration of the jet (EDJ-1), while the timing of the vortex breakdown influences the timing and latitudinal reach of the following equatorward migration (EDJ-2), as initially proposed by Bracegirdle (2011). The strength of the two downward links furthermore supports the conclusion of Byrne and Shepherd (2018) and Byrne et al. (2019) that the evolution of the S-T coupling in this time of the year is close to a deterministic system whose timing and strength are determined by the stratospheric state.

The lower predictability of SVP-low and VB given the drivers included in the BCN suggests it may be missing potentially important mechanisms that lead to the destabilisation or persistence of the springtime vortex. Domeisen et al. (2020a) quantified the prediction skill of SH final warmings for a number of S2S forecast models (including ECMWF). They found around 30% of ensemble members predict the final warming date within an error of $\pm$3 days for 30-20 days lead time, and increases to almost 75-100% for 15-5 days lead time. This suggests that weekly average variables between September and November may be needed to improve the skill in the BCN too.

The main goal of this analysis was to quantify the predictability of the EDJ given the knowledge of well known long-lead precursors, such as ENSO, upward propagating wave activity flux (vT-flux) and the Polar nigh Jet Oscillation (PJO) in August and September. It may be that only specific combinations of these drivers provide predictability, as suggested by the CPTs. The results in Figure 5.4.6 show that on average the identified long-lead precursors provide non negligible predictability to the state of EDJ-1 and EDJ-2, with median ROC AUC of 0.6-0.75 and PR AUC of 10-30%. The skill increases considerably when the lower stratospheric drivers in October are known, with median ROC AUC of 0.8-0.9 and PR AUC of 50-60%. These results highlight the dominant role of the S-T coupling in the EDJ evolution, as well as the high degree of uncertainty in

the evolution of the stratospheric zonal wind anomalies even given influential preconditions, such as ENSO, vT-flux or PJOs (as discussed above). Tested against ERA5, the BCN skill is found to be comparable to a conditional hindcast with similar sets of initial conditions, showing that the few linkages modelled are a good source of information that applies outside the hindcast world into the real world.

The BCN tropospheric variables of EDJ-1, EDJ-2 and ENSO provide a pathway to connect the BCN forecasts with surface climate response. When ensemble members are selected to match the BCN forecast for the EDJ, the refined ensemble improves the anomaly correlation for Z500, 2mT and MTPR, especially over and around Antarctica, compared to the full ensemble mean. The network allows to identify a good part of the most skilful members among the 51 of the hindcast, reducing the large uncertainty for late spring and early summer conditions for the 1 August hindcast.

5.5.1 Implications of results

The analysis presented here suggest there can be scope for using simple statistical forecasts like the BCN in the S2S context, which provide a still largely unexplored path to study the predictability of climate systems on these timescales. The atmospheric and oceanic systems relevant at weekly to monthly timescales lend themselves to be described statistically with multiple variables interacting in probabilistic terms, which can be well summarised by BCNs. A Bayesian causal network model for S2S forecasts has multiple benefits, some of which have been explored in this work:

- causal network theory provides a rigorous set of rules that allow to separate direct from indirect effects, which can reveal the most effective pathways of enhanced predictability;
- the probabilistic parametrization allows to introduce uncertainty as well as non-linearities in a manageable form, especially if variables are discretized;
- the Bayesian framework allows to quantify the added predictive power of various combinations of remote drivers by means of (very fast) probabilistic forecasts that make use only of the desired inputs;
- it can be presented visually, which enhances transparency and helps communicating sources of predictability to non-specialist audiences.

In addition, the model can be easily extended to include additional dynamics by including new relevant variables, with the benefit that all the parts of the network that remain unmodified do not need recalibrating. New observations can be used to update the probabilities used to describe the model.

5.5.2 Limitations

The drivers used here were identified through a literature review, but they may be sub-optimally defined for the task of achieving the best prediction skill. Methods for optimal search of drivers, such as Response-Guided Causal Precursor Detection (Kretschmer et al., 2017), could be used to find the best variables, latitudes, longitudes and altitude over which to define the drivers. The definition of percentile-based categories could also be optimized by searching for the ones resulting in the highest BCN skill, for example using information criteria like AIC or BIC (Ramazi et al., 2020).

The lack of a detectable link in the hindcast data between the QBO and the vortex variables of SPV-low and VB means that this analysis cannot account for the expected added predictability coming from QBO, which is generally accepted to be present in reality (Baldwin and Dunkerton, 1988; Anstey and Shepherd, 2014; Byrne and Shepherd, 2018) but is lacking in seasonal forecasting systems initialised more than a few weeks ahead of the expected effect (Butler et al., 2016; Garfinkel et al., 2018; Domeisen et al., 2020b; Lawrence et al., 2022). Similarly, the inclusion of a stratospheric ozone variable could improve predictions for VB, but was not included here given the documented deficiencies of the hindcast in representing the effect of the late 20th century ozone depletion on the vortex (Oh et al., 2022; Monge-Sanz et al., 2022).

The present BCN represented the eddy-driven jet dynamics through its zonal average. The work could be extended to the zonally asymmetric component, where the inclusion of another tropospheric circulation variable would probably be needed. Other drivers, for example IOD and IPO, could be included to provide additional predictors of the asymmetries. This could be an important extension, which would likely improve the representation of and connection with regional impacts.

5.6 Conclusions

In this work a BCN is used to quantify how much potential predictability for the spring to summer EDJ derives from combinations of variables considered important to its dynamics on S2S timescales. The network was tailored to this purpose by using a categorical and probabilistic parametrisation, which allowed to describe the evolution of the system and thus the predictability of its parts in probabilistic terms. The BCN allowed to summarise, however imperfectly, a much more complex system with relatively few linkages. This results in a clear visualization of the assumed data-generating process, which provides a statistical model that can reproduce the hindcast behaviour to a good extent. It has been used to perform prediction experiments to better understand the combined effect of multiple drivers on the S-T coupled dynamics. The BCN predictions have been also used to filter the hindcast, resulting in a more skilful representation of the mid-to-high latitude surface climate.

There are a number of interesting features to the BCN modelling choice that can be used in other contexts or systems. For example, inspecting the conditional probability tables entries with largest shift from climatology allows to easily identify so-called "windows of opportunities" for S2S forecasts. The conditional probability tables show one-link-ahead connections, but a similar analysis could be done by combining chains of links. In a changing climate, patterns of teleconnections may also change (e.g. Yu et al., 2015). If a BCN is build to capture physical relationships that are not modified themselves by a changing environment, the forcing due to climate change can be added to the model as a remote driver (see e.g. Chapter 3). If the changing climate affects the physical connections (e.g. by strengthening them), Bayesian methods provide a way to include new information and update their parametrization to account for it.

APPENDIX

5.A Biases in System 4

Bias correction is needed here in order to correctly assign ERA5 and hindcast variables that behave in the same anomalous way to consistent categories.

Most of the present analysis relies on the zonal-mean zonal wind $[u]$. Its monthly mean biases in System 4 as compared to ERA Interim have been already studied in Byrne et al. (2019). They found a good agreement in the stratosphere (no statistically significant bias), except for the tropics in November-January and 70-90ºS in September. Of all the stratospheric variables used here, none should be directly affected by these biases. In the troposphere, a small negative bias around 60ºS emerges in December-January (around 1-4 m/s, Fig. 2 in the paper). This small equatorward bias can be seen also in Figure 5.1.1 (left). The bias correction operated in this analysis is shown in Figure 5.A.1, with a procedure described in 5.2.

In terms of tropical SST, System 4 is known to develop a cold tongue bias which originates from too strong trade winds simulated by the atmospheric model in the central and western Pacific. In our ENSO variable, August mean values are a few decimals of degrees biased but growing to reach almost a degree of bias for the October mean. In terms of teleconnections, Molteni et al. (2011) report that various diagnostics show that the atmospheric response to a 1 degree SST anomaly in System 4 typically has a realistic spatial structure, but the amplitude is too weak, which overall appear to balance the too large SST anomalies. As confirmed in Byrne et al. (2019), there is also observational uncertainty on the links between ENSO and the SH zonally-symmetric response, therefore it is hard to conclude how and if System 4 is biased in this regard.

From inspection ERA5 and hindcast yearly time series, ENSO and the zonal-mean zonal wind

eddy-driven jets latitude position (EDJ-1 and EDJ-2) stand out as variables that must be bias corrected. In order to use a systematic method across all variables and avoid subjective choices of what a significant bias is, the mean bias correction is applied to all variables as described in Section 5.2.2.

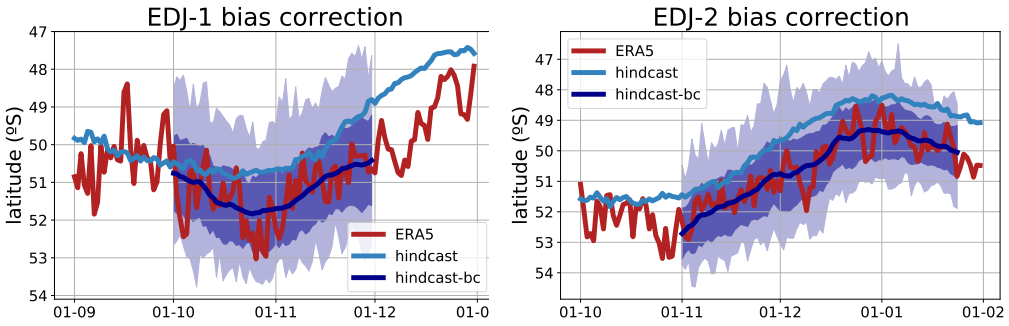

Figure 5.A.1: Bias correction of EDJ-1 (left) and EDJ-2 (right) latitude time series. ERA5 climatology (red), hindcast ensemble mean (blue) and bias corrected hindcast (ensemble mean in dark purple line, standard deviation range in light purple, min-max range in very light purple).

5.B Network Variables: time series and distribution

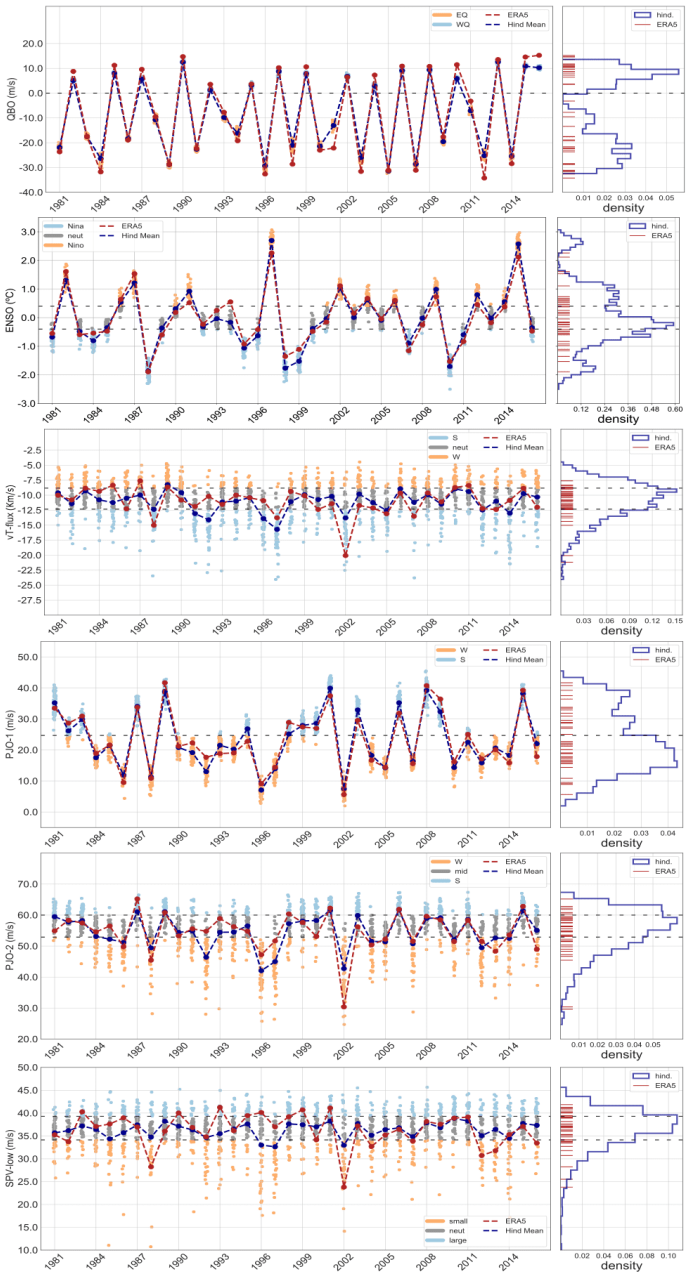

Figure 5.B.1: QBO, ENSO, vT-flux, PJO-1, PJO-2, SPV-low (top to bottom) distribution for the bias corrected hindcast (right) and the yearly time series (left) divided into strip-plots for each of the three categories (early, middle and late; ensemble mean in purple). ERA5 values are shown in red.

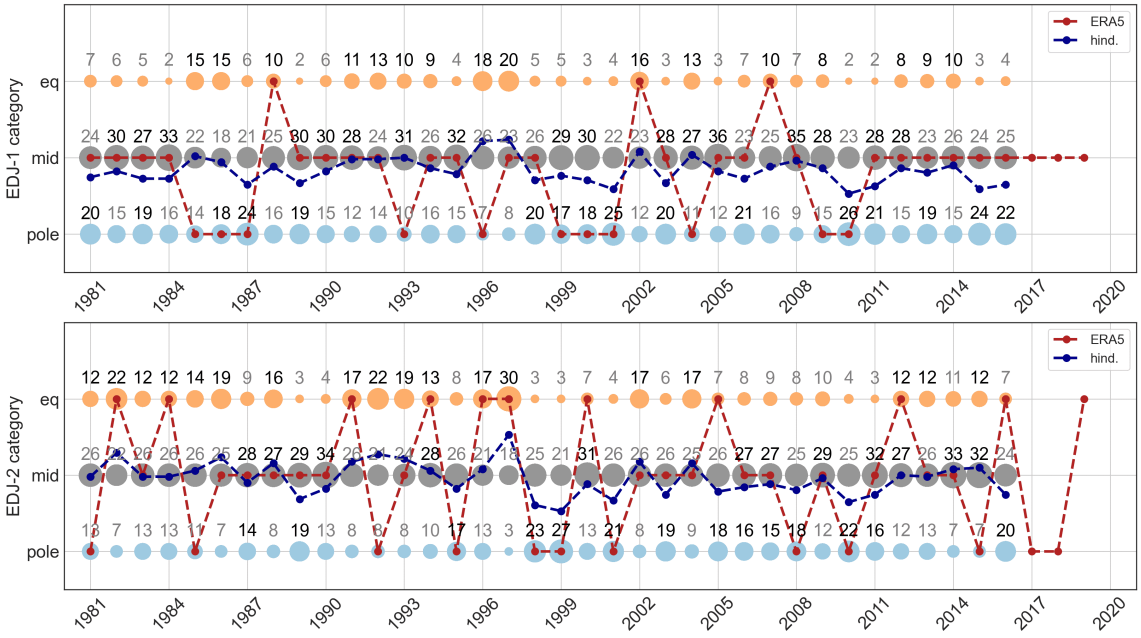

Figure 5.B.2: EDJ-1 (top) and EDJ-2 (bottom) yearly time series with bias corrected hindcast members, grouped by categories (eq, mid and pole). The ensemble mean, here computed assigning values 0, 1, 2 to pole, mid and eq respectively, is shown in purple. The numbers on top of dot show the number of members assigned to the category. ERA5 categories are shown in red.

5.C Conditional probabilities validated with ERA5

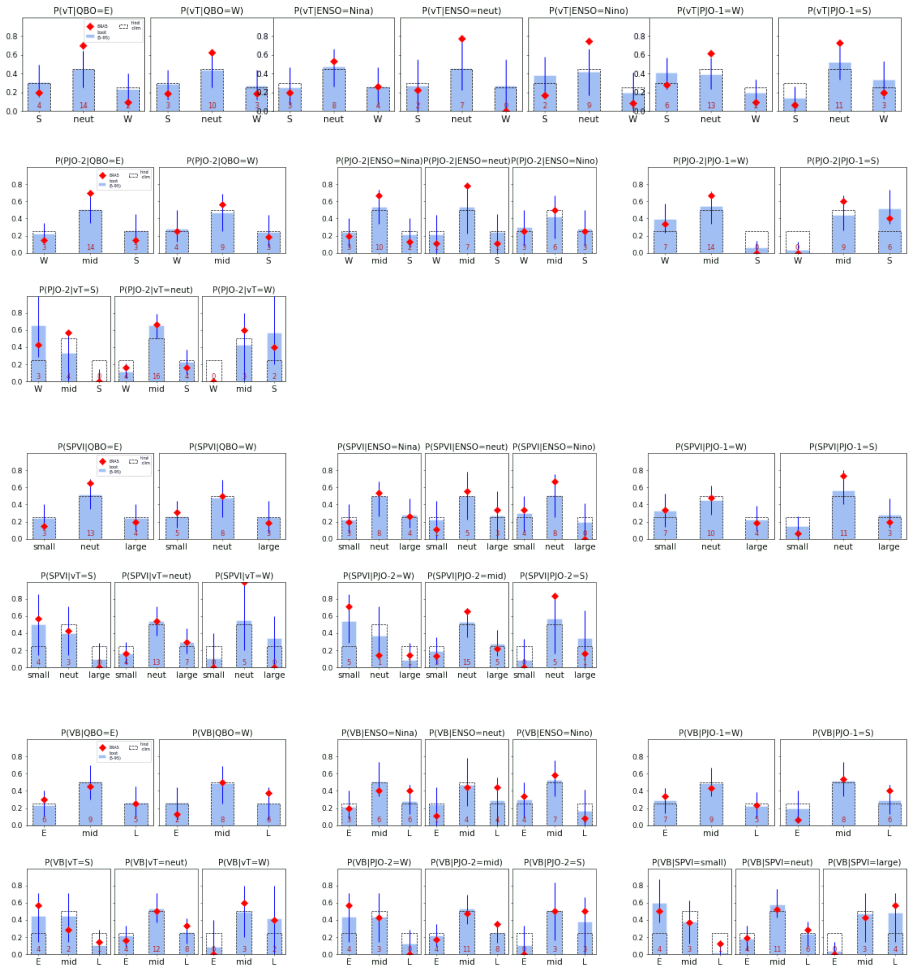

Figure 5.C.1: Conditional probabilities P(X | driver) with X being vT-flux, PJO-2, SPV-low, VB (from top to bottom).

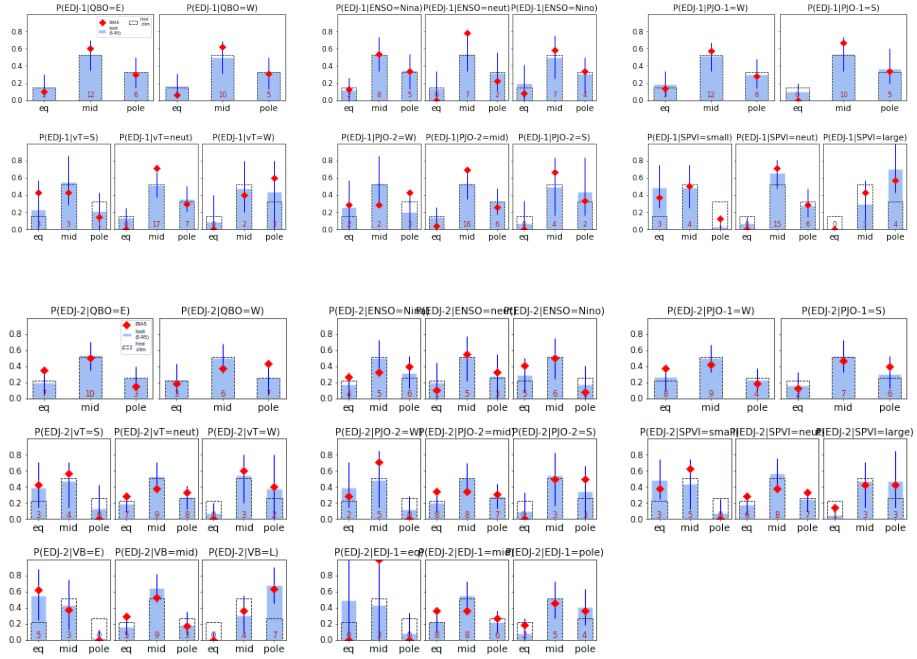

Figure 5.C.2: Conditional probabilities P(X | driver) with X being EDJ-1, EDJ-2 (from top to bottom).

5.D Effect of QBO on [u]

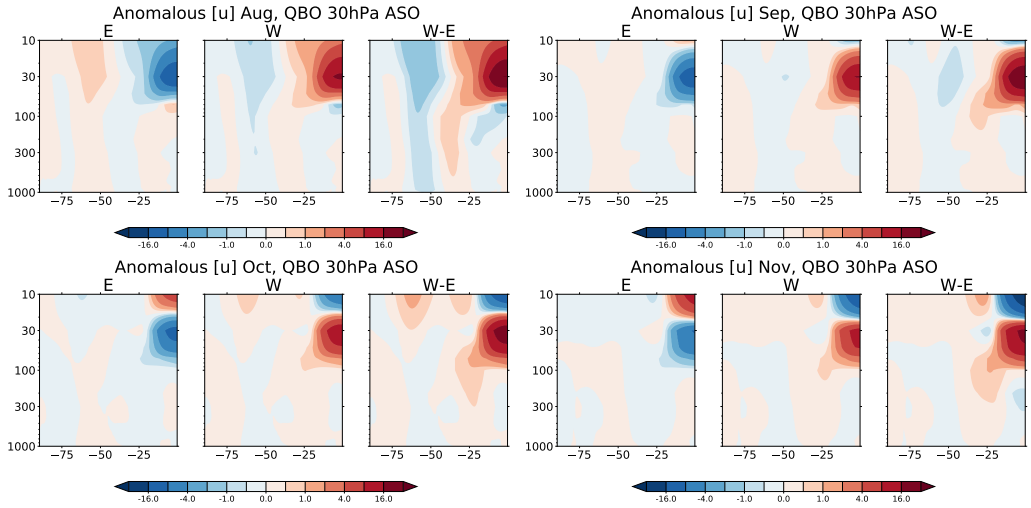

Figure 5.D.1: Hindcast composites of [u] anomalies in August, September, October and November based on E or W QBO. Units in m/s.

5.E Derivation of link strength formula

For non-autocorrelated variables, Ebert-Uphoff (2007) showed that the conditional mutual information between the linked variables given the other parents of the child can be interpreted as a measure of link strength[1]. Ebert-Uphoff (2007) define two measures of link strength, LS_{true} and LS_{blind}. The True Average Link Strength, LS_{true}, of link $X \to Y$ is defined as the conditional mutual information of the pair (X,Y) conditioned on all parents of Y other than X, namely

$$LS_{true}(X \to Y) = CMI(X,Y \mid Z) = U(Y \mid Z) - U(Y \mid X,Z) \tag{5.2}$$

$$U(Y \mid X,Z) = \sum_{x,z} P(x,z)U(Y \mid x,z) \tag{5.3}$$

$$Z = \text{parents of Y that are not X} \tag{5.4}$$

$U(Y \mid X,Z)$ is a conditional entropy. By "controlling" the influence of all other parents Z of Y, $LS_{true}(X \to Y)$ essentially blocks all information flow through the other parents and only allows for the information directly shared by X and Y to be quantified.

Note that here the parents variables are averaged over their actual joint probability, $P(x,z)$ (hence the term *true* (Ebert-Uphoff, 2007)). However, one could be interested in measures the effect of X on Y, regardless of the actual observed frequency of values of X. This is estimated by the Blind Average Link Strength, LS_{blind}, where X and Z are assumed to be independent and uniformly distributed $P(x,z) = P(x)P(z)$, $P(x) = 1/N_X, P(z) = 1/N_Z$ (blind assumption). This leads to

$$LS_{blind}(X \to Y) = \frac{1}{N_X N_Z} \sum_{x,y,z} P(y \mid x,z) \log_2 \left(\frac{P(y \mid x,z)}{1/N_X \sum_x P(y \mid x,z)} \right) \tag{5.5}$$

where $P(y \mid x,z)$ are the networks' conditional probability tables, as in Eq (5.1). LS_{blind} is a local measure because it depends only on the child node and its conditional probability table, but nothing else in the network. The interpretation of this quantity is "by how much is the uncertainty in Y reduced by knowing the state of X, if the states of all other parent variables are known?". To ease the interpretation of the link strength measure, it is useful to look at the percentage of uncertainty reduction they contribute to. Namely,

$$LS_{blind}\%(X \to Y) = \frac{LS_{blind}(X \to Y)}{\hat{U}(Y \mid Z)} \cdot 100 \tag{5.6}$$

[1] For strongly autocorrelated variables, Runge et al. (2012a) showed that for $X \to Y$ the conditional mutual information given *both* the parents of X and Y is an unbiased measure of link strength (called MIT therein)

where $\hat{U}$ is the conditional entropy computed under the "blind" assumption $\hat{U}(Y \mid X,Z) = \sum_{x,z} 1/N_X \cdot 1/N_Z \cdot U(Y \mid x,z)$ where $U(y \mid x,z)$ is the conditional entropy.$LS_{blind}\%(X \rightarrow Y)$ can be seen as a measure of *how close the link is to deterministic* and therefore how influential parent X can be in determining the state of the child Y. From a probabilistic forecast point of view, this information is very relevant. This measure of strength is different and cannot be directly related to the concept represented by, for example, a standardized linear regression coefficient.

5.F Conditional Probability Tables

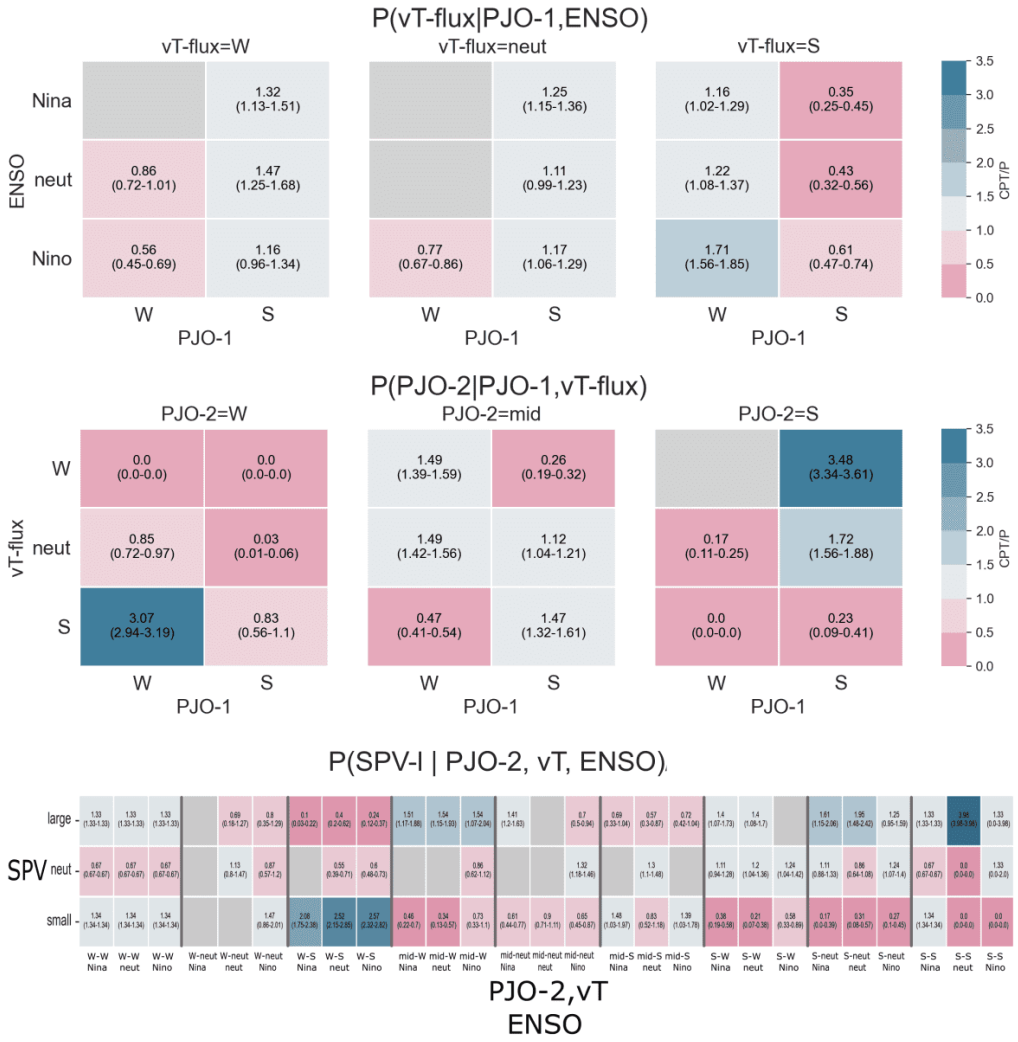

Figure 5.F.1: Conditional probability ratio CPT/P for vT-flux (top), PJO-2 (mid) and SPV-low (bottom). The values written on each table's entry show the mean and, in parenthesis, the 5-95 percentile ranges of a 1,000 sample bootstrap estimate of the CPT. Each bootstrap is obtained by re-sampling the ensemble members/years. The color blue (pink) of each entry indicates if the mean ratio is higher(lower) than 1, that is if the category of X is more(less) likely than climatology given the conditions of its parents. Entries are masked with grey color if the mean values of the ratio is between 0.9-1.1.

5.G Precursors probabilities of SPVlow, VB and EDJ-1

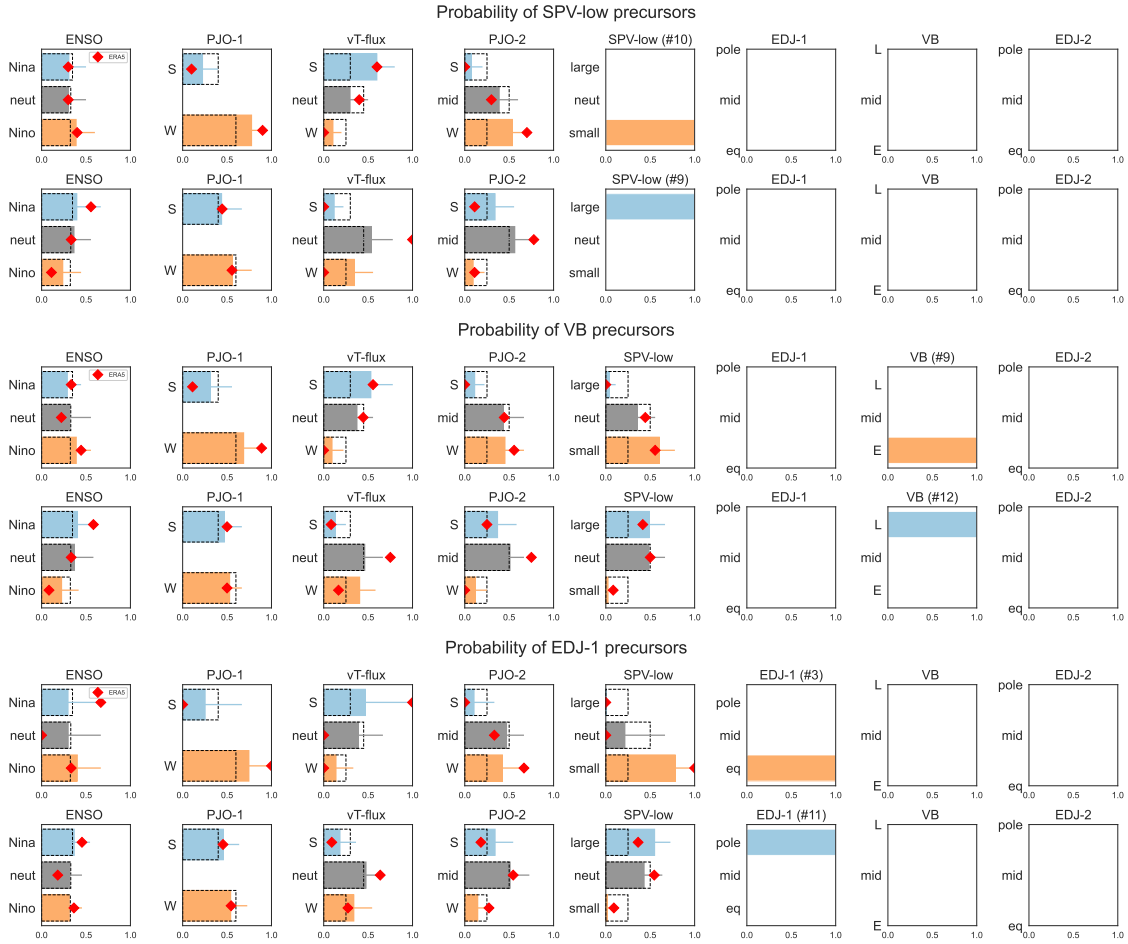

Figure 5.G.1: Probability distribution of SPV-low, VB and EDJ-1 precursors. The hindcast's distributions are estimated with a bootstrap procedure of 1,000 subsamples (see text): the hight of the horizontal bar is the median value, the error bar covers the 10-90 percentiles and the black dashed bar is the climatological probability value. The red diamonds shows the observed frequencies of the precursors, estimated from ERA5 data in the hindcast period 1981-2016.

5.H PR AUC of the BCN forecast compared with the conditional hindcast

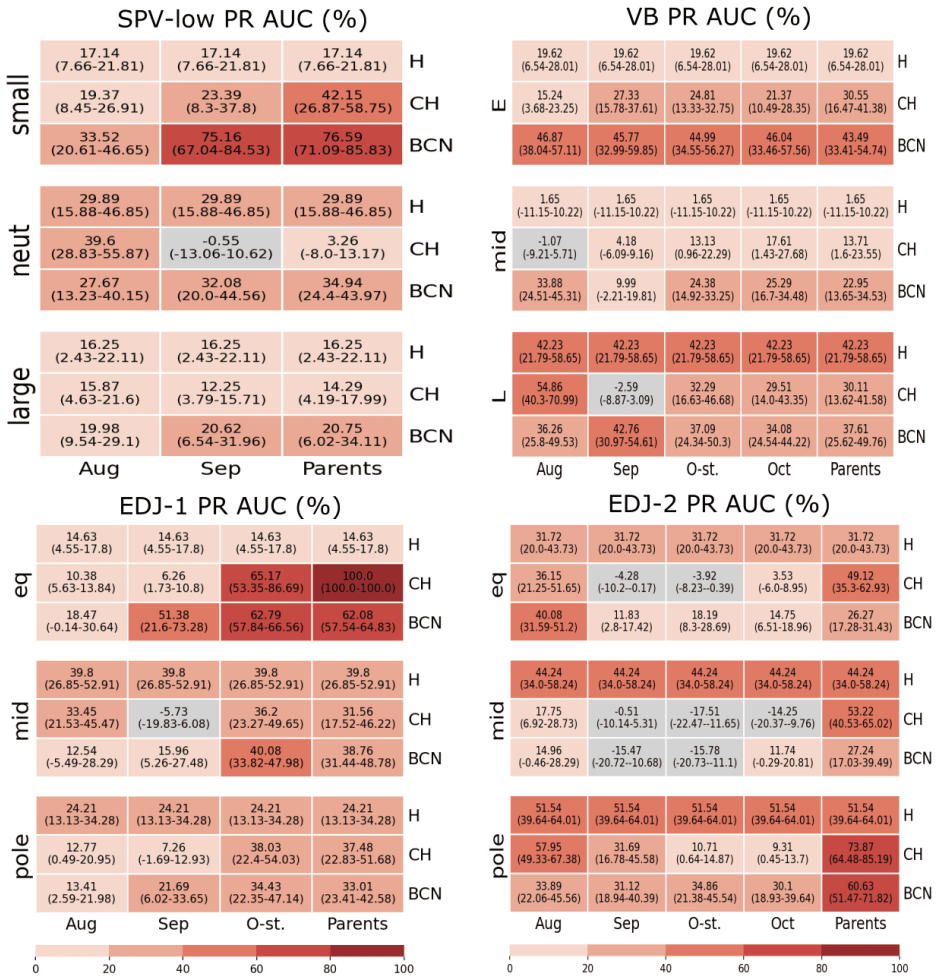

Figure 5.H.1: PR AUC (%) tables obtained by the BCN forecast (BCN), the conditional hindcast (CH) and the full hindcast (H) assessed against ERA5. Variables SPV-low, EDJ-1, VB and EDJ-2 are the targets (panels). Each table shows the three categories's skill (sub-panels). The skill of the hincast is the top row, for the conditional hindcast is the middle row and for the BCN is the bottom row of each sub-panel. The evidence set used (August, September, October-stratosphere, October, Parents) constitute the columns of each sub-panel. The colour scales with the mean PR AUC value, with darker green corresponding to higher skill, and no skill coloured in grey. Note that value is scaled, namely the percentage (PR AUC - base)/(1-base) % is shown, as described in the text. The numbers in each table entry shows the mean and 25-75 percentiles (in parenthesis) obtained via a 1,000 samples bootstrap procedure.

5.I Hindcast two-variate composites of Z500, 2mT and MTPR

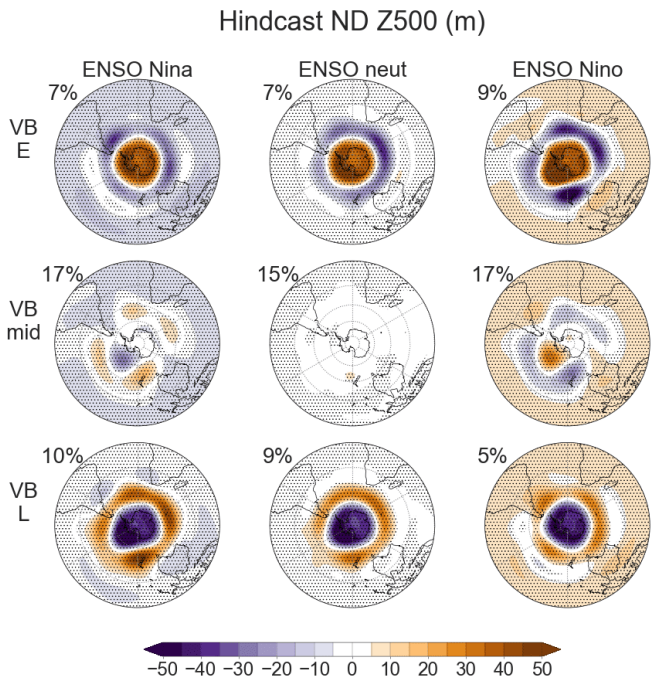

Figure 5.I.1: Hindcast geopotential height anomalies at 500 hPa, ND mean composites with respect to VB and ENSO. The dots show areas where anomalies values are significantly different from 0 according to one sample two-sided t-test at 0.1 significance level.

It is worth noting that the Z500 response to VB is not fully symmetric, but there is also a zonally asymmetric component which has been recently studied in detail by Osman et al. (2022) for this same dataset. Their analysis shows it to be a wave-1 pattern with positive (negative) anomalies occupying most of the eastern (western) Hemisphere in September and October; while in November, at polar latitudes the anomalies reduce their extension and in midlatitudes they shift westward. In this analysis the anomalies are not separated into zonally symmetric and asymmetric but look at the full (anomalous) field. Since the asymmetric component of VB response is smaller than the symmetric, this effect is harder to see in our composites, and especially in ND when that is further reduced.

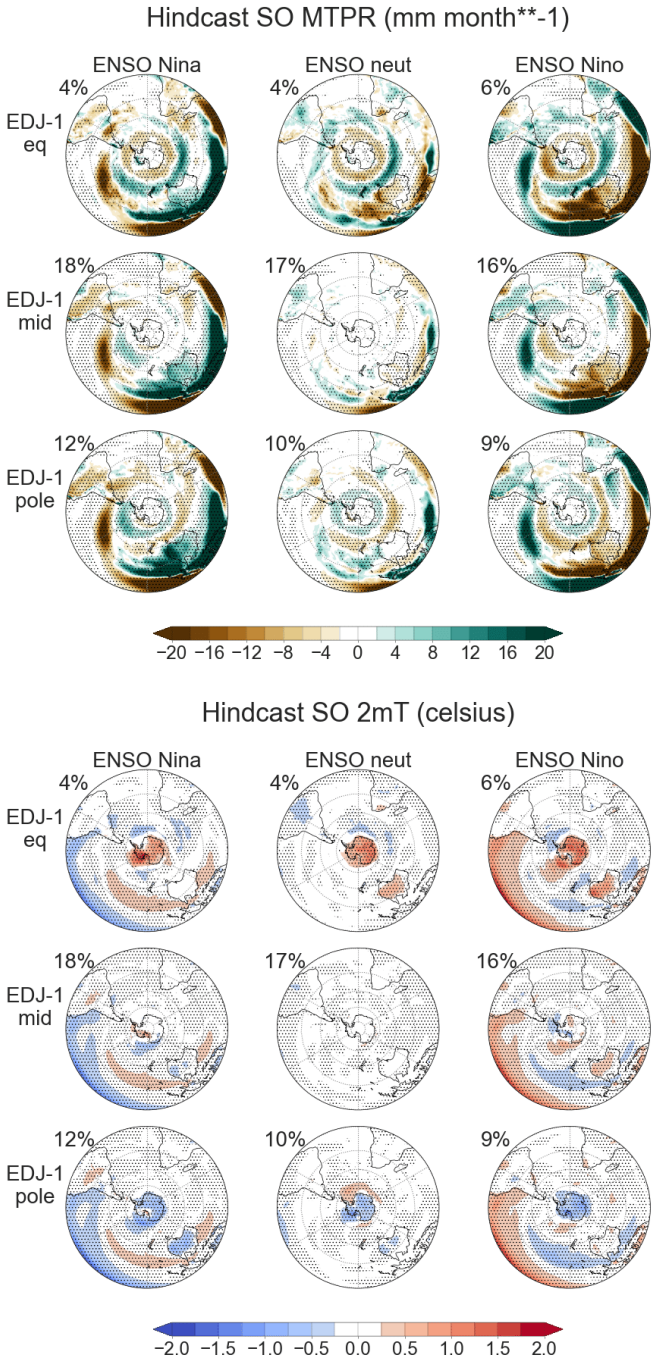

Figure 5.I.2: MTPR (left) and 2mT (right) anomalies in SO. Mean composites with respect to EDJ-1 and ENSO. The dots show areas where anomalies values are significantly different from 0 according to one sample two-sided t-test at 90% p-value.

5.J ERA5 vs hindcast univariate composites for Z500, 2mT and MTPR

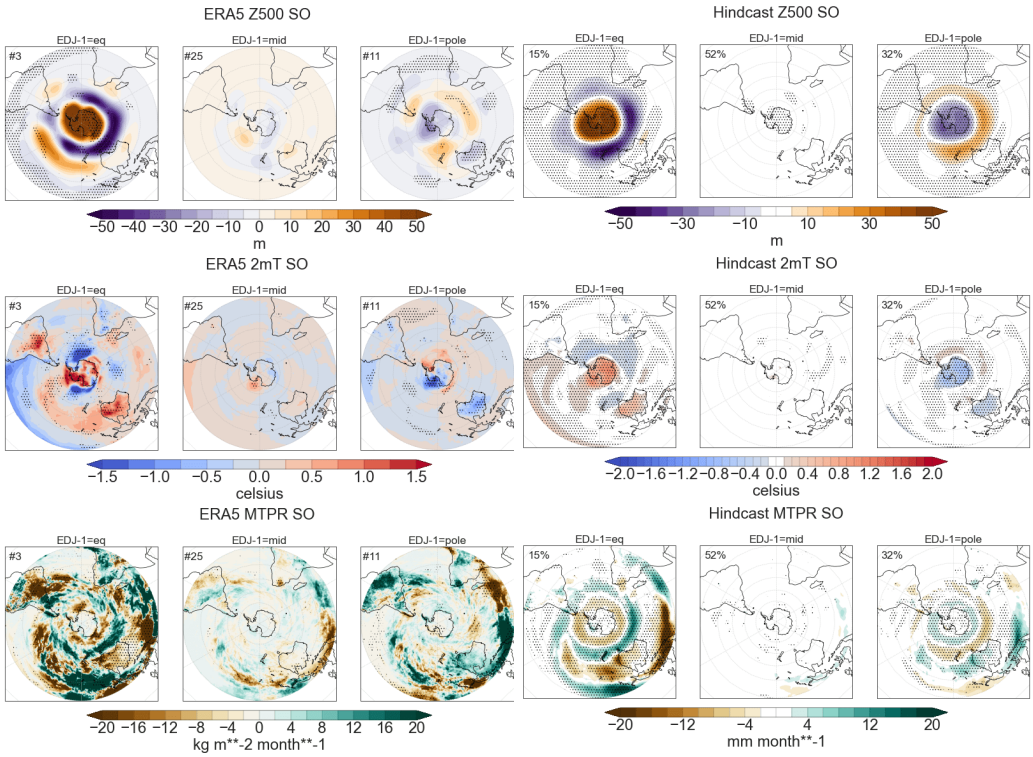

Figure 5.J.1: SO mean composites based on EDJ-1 categories, in ERA5 (left) and hindcast (right). Geopotential height anomalies at 500 hPa (top), 2mT (middle) and MTPR (bottom). The dots show areas where anomalies values are significantly different from 0 according to one sample two-sided t-test at 90% p-value.

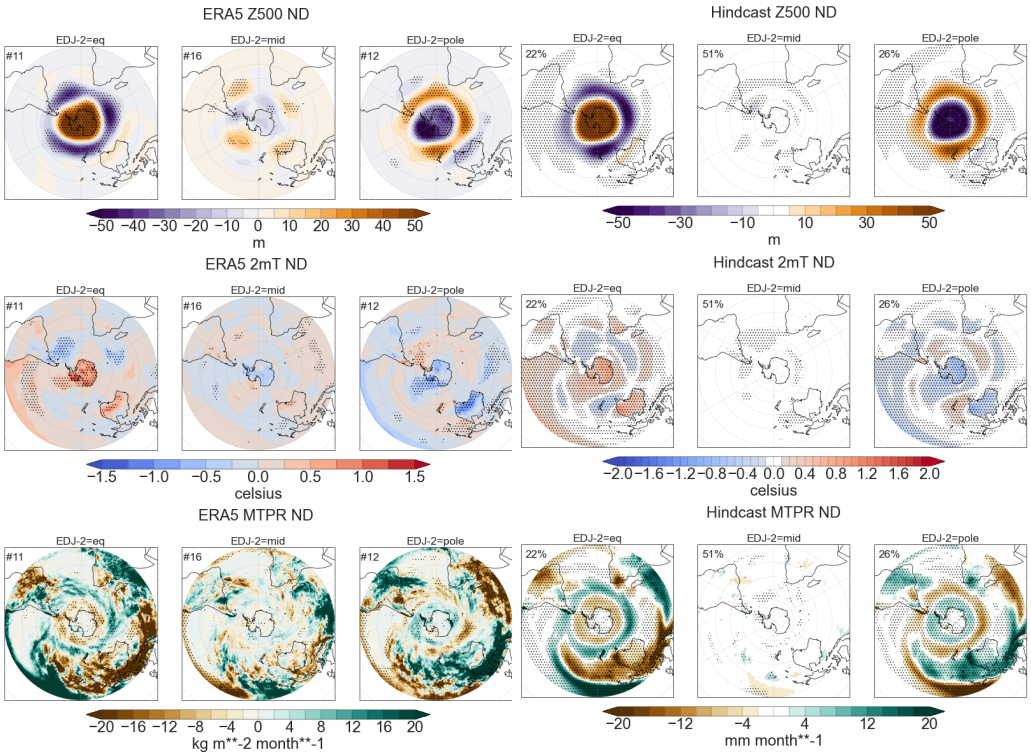

Figure 5.J.2: ND mean composites based on EDJ-2 categories, in ERA5 (left) and hindcast (right). Geopotential height anomalies at 500 hPa (top), 2mT (middle) and MTPR (bottom). The dots show areas where anomalies values are significantly different from 0 according to one sample two-sided t-test at 90% p-value.

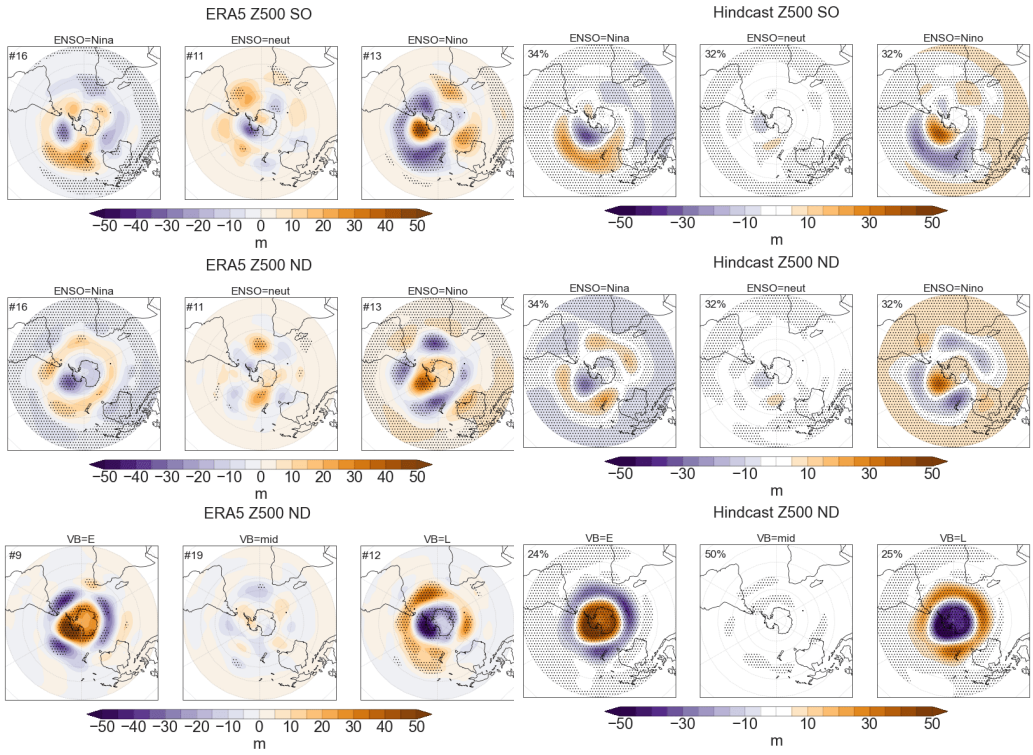

Figure 5.J.3: Z500 SO and ND mean composites based on ENSO categories (top and middle rows) and ND mean composites based on VB (bottom). In ERA5 (left) and hindcast (right). The dots show areas where anomalies values are significantly different from 0 according to one sample two-sided t-test at 90% p-value.

5.K Probability of true positives and negatives for a climatological forecast

This sections shows the calculations for the 2nd and 5th columns of Table 5.4.3. A climatological forecast can be defined as $P(r = \text{x}) = P(c = \text{x})$ for all realizations of forecast r, where P(c) is the climatological probability and x is one of the categories admitted for variable X. Note that also the unconditional probability of a given observation $o = x$ coincides with climatology, $P(o = \text{x}) = P(c = \text{x})$. The climatological forecast (random) and observations are independent, therefore $P(r = y \,\&\, o = x) = P(r = y) \cdot P(o = \text{x})$. Be x-opp the opposite category to x, for example for SPV-low they are small and large. The expected cumulated probability of true positives for a climatological forecast is the sum of joint probabilities $\sum_x P(r = x\text{x} \,\&\, o = \text{x})$ for x spanning all the categories of variable X. The expected cumulated probability of negatives for opposite categories for a climatological forecast is the sum of joint probabilities $\sum_x P(r = \text{x-opp} \,\&\, o = \text{x})$ for x spanning only the categories at the opposite sides of variable X's distribution.

For SPV-low, these quantities give the following:

$$\sum_x P(r = \text{x} \,\&\, o = \text{x}) = \sum_x P(r = \text{x}) \cdot P(o = \text{x}) = \sum_x P(o = \text{x})^2 = (1/4)^2 + (1/2)^2 + (1/4)^2 = 2/8 + 1/4 = 3/8 = 37\%$$

$$\sum_x P(r = \text{x-opp} \,\&\, o = \text{x}) = \sum_x P(r = \text{x-opp}) \cdot P(o = \text{x}) = \sum_x P(o = \text{x-opp}) \cdot P(o = \text{x}) = (1/4)(1/4) + (1/4)(1/4) = 1/8 = 12\%$$

Table 5.K.1 shows the results for SPV-low and VB.

	Probability of Correct Prediction			Probability of "Opposite" Prediciton		
target	climatology	full hindcast	refined with BCN	climatology	full hindcast	refined with BCN
SPV-low	37%	50%	48%	12%	11%	**3%**
VB	37%	50%	48%	12%	7%	**3%**

Table 5.K.1: Probability expressed in percentage of correct predictions (true positives averaged over all categories) and of wrong predictions (only considering the two categories at the opposite side of the distribution, for example prediction VB=E when observed VB=L). The forecasts considered are of climatology, full hindcast, refined hindcast with BCN given evidence for the parents. Bold probabilities if BCN performs better than both the full hindcast and the climatology.

Part III

Final Remarks

Chapter 6

Conclusion

6.1 Summary

This thesis has considered the application of both deterministic and probabilistic causal network-based methods for the statistical analysis of sub-seasonal to seasonal (S2S) variability and predictability. The main system studied is the Southern Hemisphere mid-to-high latitude large-scale circulation variability between late austral winter and early summer. This is a time of the year when the downward coupling between the stratosphere and troposphere is particularly strong, giving rise to the potential for extended predictability of the troposphere. The main body of work has concerned itself with the following three research focuses. Firstly, the timescale and amount of information that propagates downward from the polar vortex to the eddy-driven jet via the S-T coupling has been quantified using deterministic causal network methods, including a time-series causal discovery algorithm (Chapter 3). Motivated by the non-stationarity of the SH downward coupling, a novel method has been developed for non-stationary regime-dependent time-series causal discovery (Chapter 4). Lastly, the predictability of the interannual S-T coupled variability arising from well-known long-lead drivers has been quantified with a probabilistic causal network approach (Chapter 5). The main results are summarized in the next subsections, followed by a discussion of their theoretical and practical implications. The thesis concludes with some suggestions for potential future directions.

6.1.1 Timescale and strength of the S-T coupling, and its role under external forcing

Chapter 3 studied the direct influence of the SH stratospheric polar vortex on the tropospheric eddy-driven jet during the spring-to-summer transition. The intra-seasonal downward S-T coupling has been estimated from reanalysis using a causal network approach (PCMCI, Runge, 2018) employing partial correlations. The causal method ensures proper control for confounding effects, such as the strong stratospheric autocorrelation. This allows to estimate a timescale of about 35 days for the direct information flow from the vortex to the jet. Using a linear approximation of the S-T coupling, the vortex explains around 40% of monthly jet variability (more exactly, using the 35 days time aggregation). At this timescale, the jet does not provide a detectable source of information to itself. This is a particularly relevant finding since the enhanced monthly autocorrelation of the jet in this time of the year is usually attributed to tropospheric feedbacks. Here, however, evidence is presented for the dominant role of the stratospheric forcing (with its strong autocorrelation) in inflating the jet's autocorrelation via the S-T coupling. Finally, the estimated linear model for the S-T coupling is used to connect stratospheric polar vortex strengthening due to ozone depletion during the late 20th century with the concurrently observed poleward shift of the eddy-driven jet in the troposphere. The analysis presented suggests that the S-T coupling can propagate the vortex strengthening downward to induce a jet shifting. Based on the knowledge of the author, this purely observational attribution of the tropospheric jet shift is the first of its kind, and complements previously published model-based attributions. The tropospheric trend, corresponding to on average a more poleward position, is argued to be due to a later equatorward shift of the jet in late spring, which is itself a result of a later vortex breakdown in a stronger and cooler stratosphere.

6.1.2 Detection of regime-dependent dynamics using a causal network discovery algorithm

Chapter 4 presented one of the first methods for non-stationary time-series causal discovery. The special case of regime-dependent non-stationarity is addressed. Regime-dependence means that the causal relationships between the considered processes vary depending on some prevailing background regime that may be modelled as switching between different states. Further, often such regimes have strong persistence, that is, they operate and affect causal relations on much

longer time scales than the causal relations among the individual processes. Although a special case of non-stationarity, this is an observed behaviour in many weather and climate systems, such as seasonally-varying teleconnections. To detect such regime-dependent causal relations, the conditional independence-based causal discovery method, PCMCI, is combined with a regime learning optimization approach. PCMCI is chosen because it allows for causal discovery from high-dimensional and highly correlated time series, and has been shown to efficiently control for false positives. The proposed method, Regime-PCMCI, is evaluated on a number of numerical experiments demonstrating that it can distinguish regimes with different causal directions, time lags, sign of causal links and change in the variables' autocorrelation, although the latter is achieved less skilfully. The number of regimes can also be inferred from data using a corrected Akaike information criterion. Regime-PCMCI is employed to observations of El Niño Southern Oscillation (ENSO) and All-India rainfall (AIR), demonstrating skill also in real-world datasets. The algorithm is able to detect the seasonal dependence of the teleconnection, with an influence of ENSO on AIR detected predominantly in the summer months and no such influence in the rest of the year.

6.1.3 Long-lead predictability of the S-T coupling interannual variability

Chapter 5 presented the use of a probabilistic Bayesian causal networks (BCN) for the study of S2S predictability of the S-T coupling variability. In particular, the spring-to-summer migrations of the eddy-driven jet (EDJ), both in the poleward (September-October or EDJ-1) and equatorward (November-December or EDJ-2) directions, are inspected. A BCN is built with a literature-based selection of long-lead drivers of the S-T variability from late winter, including ENSO, QBO, PJO and upward propagating wave activity flux (vT-flux). Crucial late spring mediators, such as the stratospheric polar vortex variability (SPV-low and VB), are included to build a BCN that seamlessly connects late winter to early summer. The BCN links are tested and quantified using a large hindcast ensemble dataset, revealing a much stronger influence (link strength) of the vortex on the jet compared to the effect of long lead-drivers. To study how this translates into predictability, a number of synthetic predictions of the EDJ and vortex anomalies based on various combinations of long-lead drivers are performed. The predictability of the extreme states of SPV-low and VB is relatively low given information on ENSO, PJO and vT-flux in August and September (ROC AUC = 0.6-0.8), suggesting a dominant role of internal variability at shorter time

lags. QBO and ozone could not be assessed due to model deficiencies. The predictability of the extreme states of EDJs variability is very high given knowledge of the strength of the stratospheric vortex winds in October (SPV-low) for EDJ-1 and of the timing of the vortex breakdown (VB) for EDJ-2, respectively (ROC AUC = 0.75-0.9). The weak links between long-lead drivers and vortex variability result in a relatively small long-lead predictability for the EDJ, as a consequence of the dominant S-T coupling. The role of ENSO alone as a predictor of EDJs is found to be minor. Evidence is presented for a zonally asymmetric lower-troposphere ENSO fingerprint from September to December which could explain this somewhat unexpected result, given that the EDJ is a zonally averaged quantity in this study. Finally, the BCN prediction of EDJ based on its (stratospheric) parents is used to filter the hindcast ensemble members, retaining only the ones matching the predicted category for EDJ. The obtained refined ensemble shows a better anomaly correlation with reanalysis than the ensemble mean for geopotential height at 500 hPa, surface temperature and precipitation, especially at high latitudes. The result shows the importance of a correct EDJ position for enhanced surface forecasts, as well as the ability of a simple statistical model to predict which ensemble members are the most skilful.

6.2 Discussion

This thesis has explored a few ways in which causal networks can help advancing our understanding of climate variability and change. In Chapter 3, time-series causal networks are shown to be an effective framework to identify the timescale and strength of pairwise couplings. This is a common and important task in climate science and it is often addressed with standard methods of statistical analysis, such as lagged cross-correlation, EOF and multiple linear regression analysis. Time-series causal networks are suggested to be a preferable framework to such approaches because they allow by design to avoid estimation biases due to autocorrelation, common drivers and intermediate variables (or mediators), which such standard approaches cannot identify and remove. An unbiased description of intraseasonal connections is also important to better understand the timescale and magnitude of the associated predictability as well as the effect of climate change on such dynamics. Disentangling the change in variables under forcing (e.g. trends) and the change in their linkages (e.g. strength and timescale, if any) under climate change scenarios could help in breaking down and eventually constraining the response. A proof of concept of such an approach was presented by

applying a forcing to the detected causal network of the S-T coupling, adding an ozone depletion forcing to the polar vortex (trend in strength), and seeing the resulting changes in jet, transmitted through the S-T coupling link. The latter coupling link was estimated from intraseasonal, unforced variability only.

In Chapter 4, time-series causal discovery has been extended to regime-dependent non-stationarity by designing one of the first algorithms of this kind, called Regime-PCMCI. This constitutes an important extension because it allows to account for unknown beginning and end of certain dynamics (e.g. connections characterising one season of the year versus the others). This could be used to detect changes in duration of seasons and changes in strengths of the intraseasonal connections overt time due to climate change, a particularly impact-relevant research area. In addition, this method could be used to discover time-varying dynamics of multi-variate climate systems that are not yet known and that could be hard to detect due to the complex nature of the problem. A caveat of Regime-PCMCI is that it requires a large enough dataset to test and fit multiple regimes, thus ensemble hindcasts or climate runs could be ideal datasets. Furthermore, the algorithm does not provide one single answer but a range of possible solutions (i.e. networks associated with each regime and activation timing for each of them), due to the existence of multiple local minima in the functional that is optimized in the discovery process. The output of the algorithm thus needs some care in its inspection, and potentially requires a selection among the various minima, which needs to be rooted in the physical understanding of the system. Finally, Regime-PCMCI has been implemented for linear systems, but it could be adapted to nonlinear functional ones with some further work.

Chapter 5 showed how a Bayesian causal network can be used to test and probabilistically quantify claims of predictability based on certain drivers. The analysis presented here showed that an association between a target and a driver, detected with bivariate conditional probability (but also potentially with composite analysis or regressions), does not necessarily correspond to a strong added prediction skill. This is because other drivers can counteract that effect (both linearly and nonlinearly), or because while detectable the effect is itself not very strong. Note that the BCN presented here was built to assess already well-known drivers, but if the purpose is that of maximizing prediction skill then the selection of variables and their definitions could be optimized further. If good predictors are identified, a BCN can become a simple and quick-to-run model

for probabilistic S2S predictions. Such a model can be easily extended to include other drivers (e.g. climate change by adding forcing variables) and it can also be coupled with other models (e.g. impact models). A BCN could be a basis for "intermediate technologies", i.e. simple but not elementary statistical models, which can bring together physical and statistical understandings in a way that can create intelligibility and also democratize the production of climate information (Rodrigues and Shepherd, 2022).

6.3 Future work

A number of interesting questions remain to be explored to understand the SH S-T coupled variability, its predictability and its change under global warming. While the intraseasonal effect of the vortex on the jet is increasingly well established, also thanks to the analysis presented in Chapter 3, many aspects remain to be understood about the specific processes affecting the stratospheric dynamics and stratosphere-troposphere dynamical coupling in both hemispheres, and how well they are modelled in numerical weather and climate prediction models. One of such elements is nonorographic gravity wave drag (NOGWD), which parametrization strength has recently been shown to have a substantial impact on EDJ in both hemispheres, but with opposite effects on the EDJ due to the different (and stratospherically mediated) causes of the variability in the two hemispheres. A study of the model runs used by Polichtchouk et al., 2018 with time-series causal networks could help further illuminate the role of NOGWD on the S-T coupling and would provide an interesting case study for expressing "interventions" with causal networks.

In terms of the predictability of the S-T coupled variability, the results presented in Chapter 5 suggest that long-range predictability of the S-T coupling seasonal variability is somewhat limited by internal, short-range variability of the mid-to-high latitude stratosphere. It would be of interest to further explore this hypothesis by identifying shorter-range predictors and understanding how they are linked to the presently studied long-lead drivers. Studying the dynamics of some variables at multiple individual months, and their auto-dependence (e.g. vT-flux in August, September and October), could also help to improve the predictability of the vortex variability, assuming their variations at monthly time scale are important to this goal.

In order to further link the tropospheric component of the BCN with regional surface climate, and thus better study the predictability thereof, an asymmetric component of jet variability could

be included. Given the detected zonally asymmetric PSA-like pattern associated with ENSO, it is likely that such an asymmetric component of the jet could be strongly influenced by ENSO and thus being more predictable than the symmetric one. Additional regional drivers, such as the Indian Ocean Dipole and the Interdecadal Pacific Oscillation, could be added as additional drivers of regional variability. The asymmetric component of the jet may need to be represented by different variables for different longitudes, e.g. for the Pacific, Indian and Atlantic sectors.

Another interesting extension of this work would be to include relevant climate change forcings. For example, polar lower-tropospheric warming tends to shift the midlatitude westerlies equatorward, and tropical upper-tropospheric warming tends to shift them poleward (Butler et al., 2010; Mindlin et al., 2020). Increased GHGs is also known to lead to a delayed breakdown (for DJF) and strengthening (for JJA) of the stratospheric polar vortex, which itself leads to a poleward shift of the westerly winds and the storm tracks (Mindlin et al., 2020, 2021). An inclusion of variables representing GHG forcing in the BCN could illustrate the change in the S-T coupling variability as driven by these different factors. If the physical relationships captured by the network are not modified themselves by a changing environment, the forcing due to climate change can be added to the model as a remote driver. If the changing climate affects the physical connections (e.g. by strengthening them), Bayesian methods could provide a way to include new information and update their parametrization to account for it.

As for the time-series regime-dependent causal discovery method presented here, there are several interesting aspects that could be explored in the future, building on the present work. A useful extension would be to nonlinear systems. Systems where each variable depends nonlinearly on its parents and where the combined effect is a linear superposition (sum) could be a first target, as it would require minimal variations to the current Regime-PCMCI. For these systems, however, the nonlinear functions would have to be known (e.g. a sine function) in order to extract residuals, which the algorithm uses to minimize the cost functional. More general nonlinear dependencies, e.g. ones where the conditional independence can be estimated with nearest neighbours tests (Runge, 2018), would be harder to implement as these are not easily translated to a fit that bears residuals. Recent extensions of PCMCI to the case of contemporaneous causal relations could also be incorporated (Runge, 2020).

Applying the presented Regime-PCMCI to an ensemble hindcast or to climate model runs

could allow to detect changes in seasonality and/or strength of known teleconnection patterns. Regime-PCMCI could be a useful tool to discover such changes, or test the emergence of known ones with an approach that can find more and more rigorously such nuances in the data. A first application could be to detect the breakdown-dependent S-T coupling using the large ensemble hindcast dataset used in Chapter 5. This could also shed light in the differences between early and late breakdown years. The method could be also used to quantify the likelihood that an alleged non-stationarity is actually real. As an initial idea to be tested, the likelihood of non-stationarity could be estimated from statistics of the various local minima solution to the optimization problem. Beyond climate, regime shifts are a feature of many other systems including ecosystems. Causal networks could be used to represent coupled climate-biological processes that undergo regime changes, for example driven by the underlying climate non-stationarity, and Regime-PCMCI could be used to explore their time variability and trends.

Taking a broader perspective, both deterministic and probabilistic causal networks provide interesting modelling features that could allow to study complex aspects of climate variability and change, as well as climate impacts and adaptation options. As argued in Kretschmer et al. (2021), causal networks provide a robust framework to study pathways of teleconnections, and could be applied to systematically compare the representation of teleconnections across climate models in present and future climate change scenarios, and their expected impact on regional climate.

Causal networks could provide a complementary modelling tool to better understand, and potentially help to bridge, the gap in predictability at the S2S timescale, which sits between the medium to long-range forecasting timescales. This is currently a topic of great interest because of the large and mostly still untapped societally-relevant range of applications of S2S products, including disaster risk reduction (White et al., 2017). At the S2S timescale, inspecting the BCN conditional probability tables of a system looking at the entries with the largest shift from climatology could allow to identify "windows of opportunity" for S2S predictions in multivariate systems. Better understanding and modelling the mechanisms dominating the S2S timescales could also improve our understanding of climate variability at the seasonal to decadal timescale, as well as our understanding of climate change in the next few decades. The study of the S-T coupling at the S2S time scale in Chapter 5 and the analysis of the connection between decadal stratospheric trends with the intraseasonal coupling in Chapter 3 may be considered two simple but instructive

examples of how causal networks could contribute to this area of research.

Finally, thanks to the modularity and ability to incorporate different sources of knowledge and data into a consistent modelling framework, causal networks seem well placed to complement other types of climate impact modelling. A causal network for the climate system could be coupled with one representing ecosystems or social dynamics and their interaction with the former. Given the complexity of designing impact models, a causal network approach able to summarize (and test) the key interactions could be of great importance in complementing other impact models (e.g. physics-based or agent-based). An example of such an attempt is the BCN developed by Young et al. (2020) where the local weather is coupled with a socio-economic network affecting household food security in a region of Namibia. The network is used to study the storyline leading to household impacts in 2013/14, a consumption year affected by flooding, and the effects of a range of interventions across wealth groups. A crucial advantage of a causal impact model compared to most of the alternatives is its graphical representation, which makes its basic assumptions and interactions visual and more easy to understand and to question. These are two key elements for a model that would be adopted by policy-makers and other, potentially non-specialist, stakeholders. Secondly, by looking at the causal network as a reduced-order version of a more complex underlying impact model, it could help advancing our understanding of the key components of the system, for example by testing its ability to reproduce observations or expected behaviour. Simulations made with causal models generally take very little runtime, allowing for fast large ensemble simulations of a range of scenarios. Finally, and very importantly, causal networks are a natural framework to describe and simulate interventions, such as climate adaptation options, and assess their effect.

BIBLIOGRAPHY

Abramson, B., J. Brown, W. Edwards, A. Murphy and R. L. Winkler (1996). 'Hailfinder: A Bayesian system for forecasting severe weather'. *International Journal of Forecasting* 12 (1). Probability Judgmental Forecasting, 57–71. DOI: 0.1016/0169-2070(95)00664-8.

Adedipe, T., S. M. and E. Zio (2020). 'Bayesian Network Modelling for the Wind Energy Industry: An Overview'. *Reliability Engineering and System Safety* 202, 107053. DOI: 10.1016/j.ress.2020.107053.

Akaike, H. (1973). 'Information theory and an extension of the maximum likelihood principle '. *2nd International Symposium on Information Theory*.

Ambaum, M. H. P. and B. J. Hoskins (2002). 'The NAO Troposphere-Stratosphere Connection'. *Journal of Climate* 15 (14), 1969–1978. DOI: 10.1175/1520-0442(2002)015<1969:tntsc>2.0.co;2.

Amstrup, S. C., B. G. Marcot and D. C. Douglas (2008). 'A Bayesian network modeling approach to forecasting the 21st century worldwide status of polar bears.' *Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications* 180, 213–268.

Anstey, J. A. et al. (2022). 'Teleconnections of the quasi-biennial oscillation in a multi-model ensemble of QBO -resolving models'. *Quarterly Journal of the Royal Meteorological Society* 148, 1568–1592. DOI: 10.1002/qj.4048.

Anstey, J. and T. G. Shepherd (2014). 'High-latitude influence of the Quasi-Biennial Oscillation'. *Quarterly Journal of the Royal Meteorological Society* 140, 1–21. DOI: 10.1002/qj.2132.

Arblaster, J., M. Meehl, G. A. Karoly and D. John (2011). 'Future climate change in the Southern Hemisphere: Competing effects of ozone and greenhouse gases'. *Geophysical Research Letters* 38 (2). DOI: https://doi.org/10.1029/2010GL045384.

Arnhold, J., P. Grassberger, K. Lehnertz and C. Elger (1999). 'A robust method for detecting interdependences: application to intracranially recorded EEG'. *Physica D: Nonlinear Phenomena* 134 (4), 419–430. DOI: 10.1016/S0167-2789(99)00140-2.

Baldwin, M. P. and T. J. Dunkerton (1988). 'Quasi-biennial modulation of the Southern Hemisphere stratospheric polar vortex'. *Geophysical Research Letters* 25 (17), 3343–3346. DOI: 10.1029/98GL02445.

Baldwin, M. P. and T. J. Dunkerton (2001). 'Stratospheric Harbingers of Anomalous Weather Regimes'. *Science* 294 (5542), 581–584. DOI: 10.1126/science.1063315.

Baldwin, M. P. et al. (2021). 'Sudden Stratospheric Warmings'. *Reviews of Geophysics* 59 (1), e2020RG000708. DOI: 10.1029/2020RG000708.

Baldwin, M., D. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton and A. O'Neill (2003). 'Stratospheric memory and skill of extended-range weather forecasts'. *Science* 301 (5633), 636–640. DOI: 10.1126/science.1087143.

Barnes, E. A., S. Samarasinghe, I. Ebert-Uphoff and J. C. Furtado (2019). 'Tropospheric and Stratospheric Causal Pathways Between the MJO and NAO'. *Journal of Geophysical Research: Atmospheres* 124, 9356–9371. DOI: 10.1029/2019JD031024.

Barnes, E. A. and D. L. Hartmann (July 2010). 'Dynamical Feedbacks of the Southern Annular Mode in Winter and Summer'. *Journal of the Atmospheric Sciences* 67 (7), 2320–2330. DOI: 10.1175/2010jas3385.1.

Barnett, L. and A. K. Seth (2015). 'Granger causality for state space models'. *Physical Review E* 91 (040101). DOI: 10.1103/PhysRevE.91.040101.

Bielza, C. and P. Larrañaga (2014). 'Bayesian networks in neuroscience: a survey'. *Frontiers in Computational Neuroscience* 8. DOI: 10.3389/fncom.2014.00131.

Bjerknes, J. (1966). 'A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature'. *Tellus* 18 (5), 820–829. DOI: 10.1111/j.2153-3490.1966.tb00303.x.

Bjerknes, J. (1969). 'Atmospheric Teleconnections from the Equatorial Pacific'. *Monthly Weather Review* 97 (2), 163–172. DOI: 10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2.

Black, R. X. and B. A. McDaniel (2007). 'Interannual Variability in the Southern Hemisphere Circulation Organized by Stratospheric Final Warming Events'. *Journal of the Atmospheric Sciences* 64 (8), 2968–2974. DOI: 10.1175/jas3979.1.

Bódai, T., G. Drótos, M. Herein, F. Lunkeit and V. Lucarini (Feb. 2020). 'The Forced Response of the El Niño–Southern Oscillation–Indian Monsoon Teleconnection in Ensembles of Earth System Models'. *Journal of Climate* 33 (6), 2163–2182. DOI: 10.1175/JCLI-D-19-0341.1.

Boneh, T., G. T. Weymouth, P. Newham, R. Potts, J. Bally, A. E. Nicholson and K. B. Korb (Oct. 2015). 'Fog Forecasting for Melbourne Airport Using a Bayesian Decision Network'. *Weather and Forecasting* 30 (5), 1218–1233. DOI: 10.1175/waf-d-15-0005.1.

Bouchet, F., J. Rolland and J. Wouters (2019). 'Rare Event Sampling Methods'. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29 (8), 080402. DOI: 10.1063/1.5120509.

Boyin, H. et al. (2017). 'Extended Reconstructed Sea Surface Temperatures Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons'. *Journal of Climate* 20 (30), 8179–8205. DOI: doi.org/10.1175/JCLI-D-16-0836.1.

Bracegirdle, T. J. (2011). 'The seasonal cycle of stratosphere-troposphere coupling at southern high latitudes associated with the semi-annual oscillation in sea-level pressure'. *Climate Dynamics* 37 (11), 2323–2333. DOI: 10.1007/s00382-011-1014-4.

Breul, P., P. Ceppi and T. G. Shepherd (2022). 'Relationship between southern hemispheric jet variability and forced response: the role of the stratosphere'. *Weather and Climate Dynamics Discussions* 3, 645–658. DOI: 10.5194/wcd-3-645-2022.

Brewer, A. W. (1949). 'Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere'. *Quarterly Journal of the Royal Meteorological Society* 75 (326), 351–363. DOI: 10.1002/qj.49707532603.

Burnham, K. and D. Anderson (2002). *Model Selection and Multimodel Inference*. Springer.

Butail, S. and M. Porfiri (2019). 'Detecting switching leadership in collective motion'. *Chaos* 29 (1), 011102. DOI: 10.1063/1.5079869.

Butchart, N. (2022). 'The stratosphere: dynamics and variability'. *Weather and Climate Dynamics Discussions*, 1–53. DOI: 10.5194/wcd-2022-34.

Butler, A. H., D. J. Seidel, S. C. Hardiman, N. Butchart, T. Birner and A. Match (2015). 'Defining Sudden Stratospheric Warmings'. *Bulletin of the American Meteorological Society* 96 (11), 1913–1928. DOI: 10.1175/BAMS-D-13-00173.1.

Butler, A., D. W. Thompson and R. Heikes (2010). 'The steady-state atmospheric circulation response to climate change–like thermal forcings in a simple general circulation model'. *Journal of Climate* 23 (13), 3474–3496. DOI: 10.1175/2010JCLI3228.1.

Butler, A. H. et al. (Mar. 2016). 'The Climate-system Historical Forecast Project: do stratosphere-resolving models make better seasonal climate predictions in boreal winter?' *Quarterly Journal of the Royal Meteorological Society* 142 (696), 1413–1427. DOI: 10.1002/qj.2743.

Byrne, N. J. (2017). 'Deterministic models of Southern Hemisphere circulation variability'. PhD thesis. University of Reading, Department of Meteorology.

Byrne, N. J. and T. G. Shepherd (2018). 'Seasonal Persistence of Circulation Anomalies in the Southern Hemisphere Stratosphere and Its Implications for the Troposphere'. *Journal of Climate* 31 (9), 3467–3483. DOI: 10.1175/jcli-d-17-0557.1.

Byrne, N. J., T. G. Shepherd and I. Polichtchouk (2019). 'Subseasonal-to-seasonal predictability of the Southern Hemisphere eddy-driven jet during austral spring and early summer.' *Journal of Geophysical Research: Atmospheres.* 124, 6841–6855. DOI: 10.1029/2018JD030173.

Byrne, N. J., T. G. Shepherd, T. Woollings and R. A. Plumb (2016). 'Annular modes and apparent eddy feedbacks in the Southern Hemisphere'. *Geophysical Research Letters* 43 (8), 3897–3902. DOI: 10.1002/2016gl068851.

Byrne, N. J., T. G. Shepherd, T. Woollings and R. A. Plumb (2017). 'Nonstationarity in Southern Hemisphere Climate Variability Associated with the Seasonal Breakdown of the Stratospheric Polar Vortex'. *Journal of Climate* 30 (18), 7125–7139. DOI: 10.1175/jcli-d-17-0097.1.

Catenacci, M. and C. Giupponi (2013). 'Integrated assessment of sea-level rise adaptation strategies using a Bayesian decision network approach'. *Environmental Modelling and Software* 44. Thematic Issue on Innovative Approaches to Global Change Modelling, 87–100. DOI: https://doi.org/10.1016/j.envsoft.2012.10.010.

Ceppi, P. and T. G. Shepherd (2019). 'The role of the stratospheric polar vortex for the austral jet response to greenhouse gas forcing'. *Geophysical Research Letters* 46, 6972–6979. DOI: 10.1029/2019GL082883.

Charney, J. G. and P. G. Drazin (1961). 'Propagation of planetary-scale disturbances from the lower into the upper atmosphere'. *Journal of Geophysical Research (1896-1977)* 66, 83–109. DOI: 10.1029/JZ066i001p00083.

Chen, S. H. and C. A. Pollino (2012). 'Good practice in Bayesian network modelling'. *Environmental Modelling and Software* 37, 134–145. DOI: 10.1016/j.envsoft.2012.03.012.

Chickering, D. M. (2002). 'Learning Equivalence Classes of Bayesian-Network Structures'. *Journal of Machine Learning Research* 2, 445–498. DOI: 10.1162/153244302760200696.

Christiansen, R. and J. Peters (2020). 'Switching regression models and causal inference in the presence of discrete latent variables'. *Journal of Machine Learning Research* 41 (21), 1–46.

Cinelli, C., A. Forney and J. Pearl (Sept. 2020). *A Crash Course in Good and Bad Controls*. Available at SSRN: https://ssrn.com/abstract=3689437. DOI: 10.2139/ssrn.3689437.

Colombo, D. and M. H. Maathuis (2014). 'Order-Independent Constraint-Based Causal Structure Learning'. *Journal of Machine Learning Research* 15, 3921–3962.

Dapeng, H., A. M. O'Connor, C. Wang, J. M. Sargeant and C. B. Winder (2020). 'How to Conduct a Bayesian Network Meta-Analysis'. *Frontiers in Veterinary Science* 7. DOI: {10.3389/fvets.2020.00271}.

Dee, 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. DOI: 10.1002/qj.828.

Di Capua, G., M. Kretschmer, J. Runge, A. Alessandri, R. Donner, B. van den Hurk, R. Vellore, R. Krishnan and D. Coumou (2019). 'Long-Lead Statistical Forecasts of the Indian Summer Monsoon Rainfall Based on Causal Precursors'. *Weather and Forecasting* 34 (5), 1377–1394. DOI: 10.1175/WAF-D-19-0002.1.

Di Capua, G., J. Runge, R. V. Donner, B. van den Hurk, A. G. Turner, R. Vellore, R. Krishnan and D. Coumou (2020). 'Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: Causal relationships and the role of time-scales'. *Weather and Climate Dynamics Discussions* 2020, 1–28. DOI: 10.5194/wcd-2020-14.

Ding, R. and J. Li (Dec. 2012). 'Influences of ENSO Teleconnection on the Persistence of Sea Surface Temperature in the Tropical Indian Ocean'. *Journal of Climate* 25 (23), 8177–8195. DOI: 10.1175/jcli-d-11-00739.1.

Dobrynin, M. et al. (2018). 'Improved Teleconnection-Based Dynamical Seasonal Predictions of Boreal Winter'. *Geophysical Research Letters* 45 (8), 3605–3614. DOI: 10.1002/2018GL077209.

Dobson, G. M. B. (1956). 'Origin and Distribution of the Polyatomic Molecules in the Atmosphere'. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 236, 187–193. DOI: 10.2307/100027.

Domeisen, D. et al. (2020a). 'The role of the stratosphere in subseasonal to seasonal prediction. Part 1: Predictability of the stratosphere'. *Journal of Geophysical Research* 125 (2), e2019JD030920. DOI: 10.1029/2019JD030920.

Domeisen, D. et al. (2020b). 'The role of the stratosphere in subseasonal to seasonal prediction. Part 2: Predictability arising from stratosphere - troposphere coupling'. *Journal of Geophysical Research* 125 (2), e2019JD030923. DOI: 10.1029/2019JD030923.

Domeisen, D. I., C. I. Garfinkel and A. H. Butler (2019). 'The teleconnection of El Niño Southern Oscillation to the stratosphere'. *Reviews of Geophysics* 57, 5–47. DOI: 10.1029/2018RG000596.

Domeisen, D. and A. Butler (2020). 'Stratospheric drivers of extreme events at the Earth's surface'. *Communications Earth & Environment* 1 (1), 59. DOI: 10.1038/s43247-020-00060-z.

Douville, H. (2009). 'Stratospheric polar vortex influence on Northern Hemisphere winter climate variability'. *Geophysical Research Letters* 36 (18). DOI: 10.1029/2009GL039334.

Dunkerton, T. and M. Baldwin (1991). 'Quasi-biennial modulation of planetary-wave fluxes in the Northern Hemisphere winter'. *Journal of the Atmospheric Sciences* 48 (8), 1043–1061. DOI: 10.1175/1520-0469(1991)048<1043:QBMOPW>2.0.CO;2.

Ebert-Uphoff, I. (2007). *Measuring Connection Strengths and Link Strengths in Discrete Bayesian Networks*. Georgia Tech Research Report GT-IIC-07-01. Atlanta, Georgia 30308, USA: Georgia Institute of Technology.

Ebert-Uphoff, I. and Y. Deng (2012). 'Causal Discovery for Climate Research Using Graphical Models'. *Journal of Climate* 25 (17), 5648–5665. DOI: 10.1175/jcli-d-11-00387.1.

Falkena, S., J. de Wiljes, A. Weisheimer and T. Shepherd (2020). 'Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector'. *Quarterly Journal of the Royal Meteorological Society* 146 (731), 2801–2814. DOI: 10.1002/qj.3818.

Garfinkel, C. I., C. Schwartz, D. I. V. Domeisen, S.-W. Son, A. H. Butler and I. P. White (2018). 'Extratropical Atmospheric Predictability From the Quasi-Biennial Oscillation in Subseasonal

Forecast Models'. *Journal of Geophysical Research: Atmospheres* 123 (15), 7855–7866. DOI: 10.1029/2018JD028724.

Gerber, E. P., S. Voronin and L. M. Polvani (2008). 'Testing the Annular Mode Autocorrelation Time Scale in Simple Atmospheric General Circulation Models'. *Monthly Weather Review* 136, 1523–1536. DOI: 10.1175/2007MWR2211.1.

Gerber, E. P. et al. (2010). 'Stratosphere-troposphere coupling and annular mode variability in chemistry-climate models'. *Journal of Geophysical Research: Atmospheres* 115 (D3). DOI: 10.1029/2009JD013770.

Gerber, S. and I. Horenko (2014). 'On Inference of causality for discrete state models in a multiscale context'. *Proceedings of the National Academy of Sciences of the United States of America* 111 (41), 14651–14656.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Granger, C. W. J. (1969). 'Investigating causal relations by econometric models and cross-spectral methods'. *Econometrica: Journal of the Econometric Society* 37 (3), 424–438.

Graversen, R. and B. Christiansen (2003). 'Downward propagation from the stratosphere to the tropopshere: A comparison of the two hemispheres'. *Journal of Geophysical Research D: Atmospheres* 108, 9–1.

Gray, L. J., S. Crooks, C. Pascoe, S. Sparrow and M. Palmer (2014). 'Solar and QBO influences on the timing of stratospheric sudden warmings'. *Journal of the Atmospheric Sciences* 61 (23), 2777–2796.

Greco, T., M. G. Calabrò, R. D. Covello, M. Greco, L. Pasin, A. Morelli, G. Landoni and A. Zangrillo (2015). 'A Bayesian network meta-analysis on the effect of inodilatory agents on mortality'. *BJA: British Journal of Anaesthesia* 114 (5), 746–756. DOI: 10.1093/bja/aeu446.

Greenland, S., J. Pearl and J. M. Robins (1999). 'Causal diagrams for epidemiologic research'. *Epidemiology* 10 (1), 37–48. DOI: http://www.jstor.org/stable/3702180.

Gritsun, A. and G. Branstator (2007). 'Climate Response Using a Three-Dimensional Operator Based on the Fluctuation–Dissipation Theorem.' *Journal of the Atmospheric Sciences* 64 (7), 2558–2575. DOI: 10.1175/JAS3943.1.

Hagan, T., D. Fiifi, W. Guojie, S. X. Liang and D. A. J. Han (Oct. 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. DOI: 10.1175/JCLI-D-18-0881.1.

Hammerling, D., A. H. Baker and I. Ebert-Uphoff (2018). 'Ensemble Consistency Testing Using Causal Connectivity'. *Proceedings of the Eighth International Workshop on Climate Informatics (CI 2018)*. DOI: https://www.engr.colostate.edu/~iebert/PAPERS/CI2018_paper_23.pdf.

Hannart, A., J. Pearl, F. E. L. Otto, P. Naveau and M. Ghil (2016). 'Causal counterfactual theory for the attribution of wether and climate-related events'. *Bulletin of the American Meteorological Society* 97 (1), 99–110.

Hardiman, S. C. et al. (Sept. 2011). 'Improved predictability of the troposphere using stratospheric final warmings'. *Journal of Geophysical Research* 116 (D18). DOI: 10.1029/2011jd015914.

Hartmann, D. L. and F. Lo (1998). 'Wave-driven zonal flow vacillation in the Southern Hemisphere'. *Journal of the Atmospheric Sciences* 55 (8), 1303–1315.

Hartmann, D. L., C. R. Mechoso and K. Yamazaki (1984). 'Observations of wave-mean flow interaction in the Southern Hemisphere'. *Journal of the Atmospheric Sciences* 41 (3), 351–362.

Harwood, N., R. Hall, G. Di Capua, A. Russell and A. Tucker (2021). 'Using Bayesian networks to investigate the influence of subseasonal arctic variability on midlatitude North Atlantic circulation'. *Journal of Climate* 34 (6), 2319–2335. DOI: 10.1175/JCLI-D-20-0369.1.

Haynes, P., P. Hitchcock, M. Hitchman, S. Yoden, H. Hendon, G. Kiladis, K. Kodera and I. R. Simpson (2021). 'The influence of the stratosphere on the tropical troposphere'. *Journal of the Meteorological Society of Japan, Ser. II* II (99), 803–845. DOI: 10.2151/jmsj.2021-040,2021.

Haynes, P. H. (2005a). 'Stratosphere-Tropophere coupling'. *SPARC Newsletter* 25, 27–30.

Haynes, P. H. (2005b). 'Stratospheric dynamics'. *Annual Review of Fluid Mechanics* 37 (1), 263–293. DOI: 10.1146/annurev.fluid.37.061903.175710.

Haynes, P. H., M. E. McIntyre, T. G. Shepherd, C. J. Marks and K. P. Shine (1991). 'On the "Downward Control" of Extratropical Diabatic Circulations by Eddy-Induced Mean Zonal Forces'.

Journal of the Atmospheric Sciences 48 (4), 651–678. DOI: 10.1175/1520-0469(1991)048<0651:otcoed>2.0.co;2.

Held, I. M. (1975). 'Momentum transport by quasi-geostrophic eddies'. *Journal of the Atmospheric Sciences* 32 (7), 1494–1497. DOI: DOI:10.1175/1520-0469(1975)032<1494:MTBQGE>2.0.CO;2.

Hersbach, H. et al. (2020). 'The ERA5 global reanalysis'. *Quarterly Journal of the Royal Meteorological Society* 146 (730), 1999–2049. DOI: 10.1002/qj.3803.

Hio, Y. and S. Yoden (2005). 'Interannual Variations of the Seasonal March in the Southern Hemisphere Stratosphere for 1979–2002 and Characterization of the Unprecedented Year 2002'. *Journal of the Atmospheric Sciences* 62 (3), 567–580. DOI: 10.1175/jas-3333.1.

Hirt, M., G. Craig, S. Schäfer, J. Savre and R. Heinze (Mar. 2020). 'Cold pool driven convective initiation: using causal graph analysis to determine what convection permitting models are missing'. *Quarterly Journal of the Royal Meteorological Society*. DOI: 10.1002/qj.3788.

Hitchcock, P. (2012). 'The Arctic polar-night jet oscillation'. PhD thesis. University of Toronto.

Hitchcock, P. and P. H. Haynes (2016). 'Stratospheric control of planetary waves'. *Geophysical Research Letters* 43 (22), 884–892. DOI: 10.1002/2016GL071372.

Hitchcock, P., T. G. Shepherd, M. Taguchi, S. Yoden and S. Noguchi (2013). 'Lower-Stratospheric Radiative Damping and Polar-Night Jet Oscillation Events'. *Journal of the Atmospheric Sciences* 20, 1391–1408. DOI: 10.1175/JAS-D-12-0193.1.

Hitchcock, P. and I. Simpson (2014). 'The Downward Influence of Stratospheric Sudden Warmings.' *Journal of the Atmospheric Sciences* 71, 3856–3876. DOI: 10.1175/JAS-D-14-0012.1.

Holton, J. R. and H. C. Tan (1980). 'The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb'. *Journal of the Atmospheric Sciences* 37, 2200–2207.

Horenko, I. (2010). 'Finite Element Approach to Clustering of Multidimensional Time Series'. *SIAM Journal on Scientific Computing* 32 (1), 62–83.

Hu, J., R. Ren and H. Xu (June 2014). 'Occurrence of Winter Stratospheric Sudden Warming Events and the Seasonal Timing of Spring Stratospheric Final Warming'. *Journal of the Atmospheric Sciences* 71 (7), 2319–2334. DOI: 10.1175/jas-d-13-0349.1.

Hurrell, J. and H. van Loon (1994). 'A modulation of the atmospheric annual cycle in the Southern Hemisphere'. *Tellus* 46A, 325–338.

Hurvich, M. and C.-L. Tsai (1989). 'Regression and time series model selection in small samples'. *Biometrika* 76 (2), 297–307.

Hurwitz, M., P. A. Newman, L. D. Oman and A. Molod (2011). 'Response of the Antarctic stratosphere to two types of El Nino events'. *Journal of the Atmospheric Sciences* 68 (4), 812–822.

Indian Institute of Tropical Meteorology (2016). *IITM Indian regional/subdivisional Monthly Rainfall data set (IITM-IMR)*. DOI: https://tropmet.res.in/static_pages.php?page_id=53.

Ineson, S. and A. A. Scaife (2009). 'The role of the stratosphere in the European climate response to El Niño'. *Nature Geoscience* 2, 32–36.

Ineson, S., A. A. Scaife, J. R. Knight, J. C. Manners, N. J. Dunstone, L. J. Gray and J. D. Haigh (2011). 'Solar forcing of winter climate variability in the Northern Hemisphere'. *Nature Geoscience* 4, 753–757. DOI: 10.1038/ngeo1282.

IPCC (2022). 'Climate Change 2022: Impacts, Adaptation, and Vulnerability'. In: *IPCC Sixth Assessment Report*. Ed. by H. Portner et al. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)] Cambridge University Press. In Press.

Jeffreys, H. (1939). *The theory of probability*. 1st. Oxford University Press.

Jiang, M., X. Gao, H. An, H. Li and B. Sun (2017). 'Reconstructing complex network for characterizing the time-varying causality evolution behavior of multivariate time series'. *Scientific Reports* 7 (1), 104863. DOI: 10.1038/s41598-017-10759-3.

Jucker, M., T. Reichler and D. W. Waugh (2021). 'How Frequent Are Antarctic Sudden Stratospheric Warmings in Present and Future Climate?' *Geophysical Research Letters* 48 (11), e2021GL093215. DOI: 10.1029/2021GL093215.

Kalisch, M., M. Mächler, D. Colombo, M. H. Maathuis and P. Bühlmann (2012). 'Causal Inference Using Graphical Models with the R Package pcalg'. *Journal of Statistical Software* 47 (11), 1–26. DOI: 10.18637/jss.v047.i11.

Kallberg, P., P. Berrisford, B. J. Hoskins, A. Simmons, S. Uppala, S. Lamy-Thepaut and R. Hine (2005). *ERA-40 Atlas*. Project Report Series 19. ECMWF, p. 199.

Karoly, D. J. (1990). 'The role of transient eddies in low-frequency zonal variations of the Southern Hemisphere circulation'. *Tellus A: Dynamic Meteorology and Oceanography* 42 (1), 41–50. DOI: 10.3402/tellusa.v42i1.11858.

Karoly, D. J. and D. G. Vincent (1998). *Meteorology of the Southern Hemisphere*. American Meteorological Society.

Kidson, J. W. (1988). 'Indices of the Southern Hemisphere Zonal Wind'. *Journal of Climate* 1 (2), 183–194. DOI: 10.1175/1520-0442(1988)001<0183:IOTSHZ>2.0.CO;2.

Kidson, J. W. (Sept. 1999). 'Principal Modes of Southern Hemisphere Low-Frequency Variability Obtained from NCEP–NCAR Reanalyses'. *Journal of Climate* 12 (9), 2808–2830. DOI: 10.1175/1520-0442(1999)012<2808:pmoshl>2.0.co;2.

Kidston, J., A. A. Scaife, S. C. Hardiman, . M. Mitchell, N. Butchart, M. P. Baldwin and L. J. Gray (2015). 'Stratospheric influence on tropospheric jet streams, storm tracks and surface weather'. *Nature Geoscience* 8 (6), 433–440. DOI: 10.1038/ngeo2424.

Kiiveri, H. and T. Speed (1982). 'Structural analysis of multivariate data: A review'. *Sociological methodology* 13, 209–289.

Kiiveri, H., T. Speed and J. Carlin (1984). 'Recursive causal models'. *Journal of the Australian Mathematical Society. Series A.* 36 (1), 30–52. DOI: 10.1017/S1446788700027312.

Klerk, S. de, B. P. Veldkamp and T. J. H. Eggen (2015). 'Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example'. *Computers and Education* 85, 23–34.

Kodera, K. and Y. Kuroda (2002). 'Dynamical response to the solar cycle'. *Journal of Geophysical Research* 107 (4749). DOI: 10.1029/2002JD002224.

Kodera, K., K. Yamazaki, M. Chiba and K. Shibata (1990). 'Downward propagation of upper stratospheric mean zonal wind perturbation to the troposphere'. *Geophysical Research Letters* 17 (9), 1263–1266. DOI: 10.1029/GL017i009p01263.

Koller, D. and N. Friedman (2010). *Probabilistic Graphical Models: Principles and Techniques*. Cambridge: MIT Press. DOI: 10.1017/CBO9781107415324.004.

Kretschmer, M. (Oct. 2017). 'Disentangling Causal Pathways of the Stratospheric Polar Vortex. A Machine Learning Approach'. PhD thesis. University of Potsdam.

Kretschmer, M., S. V. Adams, A. Arribas, R. Prudden, N. Robinson, E. Saggioro and T. G. Shepherd (2021). 'Quantifying causal pathways of teleconnections'. *Bulletin of the American Meteorological Society* 102 (12), E2247–E2263. DOI: 10.1175/BAMS-D-20-0117.1.

Kretschmer, M., D. Coumou, J. F. Donges and J. Runge (2016). 'Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation'. *Journal of Climate* 29 (11), 4069–4081. DOI: 10.1175/jcli-d-15-0654.1.

Kretschmer, M., J. Runge and D. Coumou (2017). 'Early prediction of extreme stratospheric polar vortex states based on causal precursors'. *Geophysical Research Letters* 44 (16), 8592–8600. DOI: 10.1002/2017GL074696.

Kretschmer, M., G. Zappa and T. G. Shepherd (2020). 'The role of Barents–Kara sea ice loss in projected polar vortex changes'. *Weather and Climate Dynamics* 1 (2), 715–730. DOI: 10.5194/wcd-1-715-2020.

Kuroda, Y. (2002). 'Relationship between the Polar-Night Jet Oscillation and the Annular Mode'. *Geophysical Research Letters* 29 (8), 132-1-132–4. DOI: 10.1029/2001GL013933.

Kuroda, Y. and K. Kodera (1998). 'Interannual variability in the troposphere and stratosphere of the southern hemisphere winter'. *Journal of Geophysical Research* 103 (12), 787–799. DOI: 10.1029/98JD01042.

Kuroda, Y. and K. Kodera (2001). 'Variability of the polar night jet in the Northern and Southern Hemisphere'. *Journal of Geophysical Research* 106 (20), 20703–20713. DOI: 10.1029/2001JD900226.

L'Heureux, M. L. and D. W. J. Thompson (2006). 'Observed Relationships between the El Niño–Southern Oscillation and the Extratropical Zonal-Mean Circulation'. *Journal of Climate* 19 (2), 276–287. DOI: https://www.jstor.org/stable/10.2307/26253818.

Labitzke, K. (1999). *Die Stratosphäre : Phänomene, Geschichte, Relevanz*. Springer.

Labitzke, K. (2005). 'On the solar cycle–QBO relationship: a summary'. *Journal of Atmospheric and Solar-Terrestrial Physics* 67 (1-2), 45–54.

Lawrence, Z. D. et al. (2022). 'Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems'. *Weather and Climate Dynamics Discussions*. DOI: 10.5194/wcd-2022-12.

Lehmann, J., M. Kretschmer, B. Schauberger and F. Wechsung (June 2020). 'Potential for Early Forecast of Moroccan Wheat Yields Based on Climatic Drivers'. *Geophysical Research Letters* 47 (12). DOI: 10.1029/2020gl087516.

Lim, E. P., H. H. Hendon, G. Boschat et al. (2021). 'The 2019 Southern Hemisphere Stratospheric Polar Vortex Weakening and Its Impacts'. *Bulletin of the American Meteorological Society* 102 (6), E1150–E1171. DOI: 10.1175/BAMS-D-20-0112.1.

Lim, E. P., H. Hendon, G. Boschat, D. Hudson, A. Thompson D.W.J. abd Dowdy and J. M. Arblaster (2019). 'Australian hot and dry extremes induced by weakenings of the stratospheric polar vortex'. *Nature Geoscience* 12 (11), 896–901. DOI: 10.1038/s41561-456-x.

Lim, E., H. H. Hendon and H. Rashid (2013). 'Seasonal predictability of the Southern Annular Mode due to its association with ENSO'. *Journal of Climate* 26 (20), 8037–8054.

Lim, E., H. H. Hendon and D. W. J. Thompson (2018). 'Seasonal Evolution of Stratosphere-Troposphere Coupling in the Southern Hemisphere and Implications for the Predictability of Surface Climate'. *Journal of Geophysical Research: Atmospheres* 123 (21), 12, 002–12, 016. DOI: 10.1029/2018JD029321.

Lin, P., Q. Fu and D. L. Hartmann (2012). 'Impact of tropical SST on stratospheric planetary waves in the Southern Hemisphere'. *Journal of Climate* 25 (4), 5030–5046.

Lorenz, D. J. and D. L. Hartmann (2001). 'Eddy–Zonal Flow Feedback in the Southern Hemisphere'. *Journal of the Atmospheric Sciences* 58, 3312–3327. DOI: 10.1175/1520-0469(2001)058<3312:EZFFIT>2.0.CO;2.

Malinsky, D. and P. Spirtes (2019). 'Learning the Structure of a Nonstationary Vector Autoregression'. In: *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2986–2994.

Manzini, E. et al. (2014). 'Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling'. *Journal of Geophysical Research: Atmospheres* 119 (13), 7979–7998. DOI: 10.1002/2013JD021403.

Marcot, B. G. (2012). 'Metrics for evaluating performance and uncertainty of Bayesian network models'. *Ecological Modelling* 230, 50–62.

Marcot, B. G., Steventon, S. J. D., G. D. and R. K. McCann (2006). 'Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation'. *Journal of Forest Research* 36 (12), 3063–3074. DOI: 10.1139/x06-135.

Matthieu, D., W. Anders and W. Tomasz (2016). 'Granger Causality and Regime Inference in Markov Switching VAR Models with Bayesian Methods'. *Journal of applied Econometrics* 32 (4), 802–818.

Maycock, A. and P. Hitchcock (2015). 'Do split and displacement sudden stratospheric warmings have different annular mode signatures?' *Geophysical Research Letters* 42 (24), 943–951. DOI: 10.1002/2015GL066754.

McGraw, M. and E. A. Barnes (2018). 'Memory matters: A case for Granger causality in climate variability studies'. *Journal of Climate* 31, 3289–3300. DOI: 10.1175/JCLI-D-17-0334.1.

McLandress, C., A. J. Jonsson, D. A. Plummer, M. C. Reader, J. Scinocca and T. G. Shepherd (2010). 'Separating the Dynamical Effects of Climate Change and Ozone Depletion. Part I: Southern Hemisphere Stratosphere'. *Journal of Climate* 23 (18), 5002–5020. DOI: 10.1175/2010JCLI3586.1.

McLandress, C., T. G. Shepherd, J. Scinocca, D. A. Plummer, M. Sigmond, A. J. Jonsson and M. C. Reader (2011). 'Separating the Dynamical Effects of Climate Change and Ozone Depletion. Part II: Southern Hemisphere Troposphere'. *Journal of Climate* 24 (6), 1850–1868. DOI: 10.1175/2010JCLI3958.1.

Meehl, G. A., J. W. Hurrell and H. V. Loon (1998). 'A modulation of the mechanism of the semiannual oscillation in the Southern Hemisphere'. *Tellus A: Dynamic Meteorology and Oceanography* 50 (4), 442–450.

Metzner, P., L. Putzig and I. Horenko (2012). 'Analysis of persistent non-stationary time series and applications'. *Communications in Applied Mathematics and Computational Science* 7 (2), 175–229.

Mindlin, J., T. G. Shepherd, C. Vera and M. Osman (2021). 'Combined Effects of Global Warming and Ozone Depletion/Recovery on Southern Hemisphere Atmospheric Circulation and Regional Precipitation'. *Geophysical Research Letters* 48 (12), e2021GL092568. DOI: 10.1029/2021GL092568.

Mindlin, J., T. G. Shepherd, C. S. Vera, . Osman, G. Zappa, R. W. Lee and K. I. Hodges (2020). 'Storyline description of Southern Hemisphere midlatitude circulation and precipitation response to greenhouse gas forcing'. *Climate Dynamics* 54 (9), 4399–4421. DOI: 10.1007/s00382-020-05234-1.

Minokhin, I., C. Fletcher and A. Brenning (2017). 'Forecasting northern polar stratospheric variability with competing statistical learning models'. *Q. J. R. Meteorol. Soc.* 143, 1816–1827. DOI: 10.1002/qj.3043.

Mo, K. C. and J. N. Paegle (2001). 'The Pacific-South American modes and their downstream effects'. *International Journal of Climatology* 21 (10), 1211–1229. DOI: 10.1002/joc.685.

Mohanakumar, K. (2008). *Stratosphere troposphere interactions : an introduction*. Springer.

Molteni, F., T. Stockdale, M. A. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti et al. (2011). *The new ECMWF seasonal forecast system (System 4).* UK: 656 ECMWF. Reading.

Monge-Sanz, B. M. et al. (2022). 'A stratospheric prognostic ozone for seamless Earth system models: performance, impacts and future'. *Atmospheric Chemistry and Physics* 22 (7), 4277–4302. DOI: 10.5194/acp-22-4277-2022.

Montgomery, D. C., E. A. Peck and G. G. Vining (2012). *Introduction to Linear Regression Analysis*. Wiley, 5nd Edition.

Muradoglu, G., F. Taskin and I. Biga (2000). 'Causality between Stock Returns and Macroeconomic Variables in Emerging Markets'. *Russian and East European Finance and Trade* 36 (6), 33–53.

Muthers, S. (2014). 'Northern Hemispheric winter warming pattern after tropical volcanic eruptions: Sensitivity to the ozone climatology'. *Journal of Geophysical Research: Atmospheres* 119, 1340–1355. DOI: 10.1002/2013JD020138.

Mwaffo, V., J. Keshavan, T. Hedrick and S. Humbert (2018). 'Detecting intermittent switching leadership in coupled dynamical systems'. *Scientific Reports* 8 (1), 10338. DOI: 10.1038/s41598-018-28285-1.

Nakayama, S., J. L. Harcourt, R. A. Johnstone and A. Manica (May 2012). 'Initiative, Personality and Leadership in Pairs of Foraging Fish'. *PLOS ONE* 7 (5), 1–7. DOI: 10.1371/journal.pone.0036606.

Neapolitan, R. (1990). *Probabilistic reasoning in expert systems - theory and algorithms.* John Wiley and Sons, Inc.: John Wiley and Sons, Inc.

Nowack, P., J. Runge, V. Eyring and J. D. Haigh (Mar. 2020). 'Causal networks for climate model evaluation and constrained projections'. *Nature Communications* 11 (1), 1415. DOI: 10.1038/s41467-020-15195-y.

Oh, J., S.-W. Son, J. Choi, E.-P. Lim, C. Garfinkel, H. Hendon, Y. Kim and H.-S. Kang (2022). 'Impact of stratospheric ozone on the subseasonal prediction in the Southern Hemisphere spring'. *Progress in Earth and Planetary Science* 9 (1), 25. DOI: 10.1186/s40645-022-00485-4.

Osman, M., T. G. Shepherd and C. Vera (2022). 'The combined influence of the stratospheric polar vortex and ENSO on zonal asymmetries in the southern hemisphere upper tropospheric circulation during austral spring and summer'. *Climate Dynamics*. DOI: 10.1007/s00382-022-06225-0.

Osman, M. and C. S. Vera (Nov. 2016). 'Climate predictability and prediction skill on seasonal time scales over South America from CHFP models'. *Climate Dynamics* 49 (7-8), 2365–2383. DOI: 10.1007/s00382-016-3444-5.

Osman, M., C. S. Vera and F. J. Doblas-Reyes (June 2015). 'Predictability of the tropospheric circulation in the Southern Hemisphere from CHFP models'. *Climate Dynamics* 46 (7-8), 2423–2434. DOI: 10.1007/s00382-015-2710-2.

Pal, I., A. W. Robertson, U. Lall and M. A. Cane (2015). 'Modeling winter rainfall in Northwest India using a hidden Markov model: understanding occurrence of different states and their dynamical connections'. *Climate Dynamics* 44, 1003–1015.

Pearl, J. (1985). 'Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning'. In: *Proceedings of the 7th Conference of the Cognitive Science Society*. University of California, Irvine, CA, pp. 15–17.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Revised Second Printing. San Francisco (CA): Morgan Kaufman.

Pearl, J. (1994). 'Bayesian networks and probabilistic reasoning'. In: ed. by A. Gammerman. Alfred Walter, London. Chap. From Bayesian networks to causal networks, pp. 1–31.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pearl, J. (2009a). 'Causal inference in statistics: An overview'. *Statistics Surveys* 3, 96–146. DOI: 10.1214/09-SS057.

Pearl, J. (2009b). *Causality: Models, Reasoning, and Inference*. 2nd Edition. Cambridge University Press.

Pearl, J. (2012a). 'The Causal Mediation Formula: A Guide to the Assessment of Pathways and Mechanisms'. *Prevention Science* 13, 426–436. DOI: 10.1007/s11121-011-0270-1.

Pearl, J. (2012b). 'The Mediation Formula: A guide to the assessment of causal pathways in nonlinear models'. In: *Causality: Statistical Perspectives and Applications*. Ed. by C. Berzuini, A. Dawid and L. Bernardinelli. Chichester, UK: John Wiley and Sons, pp. 151–179.

Pearl, J., M. Glymous and N. Jewell (2016). *Causal Inference in Statistics: A Primer*. Wiley.

Pearl, J. and D. Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Pearl, J. and S. J. Russell (2002). 'Bayesian Networks'. In: *The Handbook of Brain Theory and Neural Networks*. Ed. by M. A. Arbib. 2nd. Cambridge, Massachusetts: Bradford Books (MIT Press), pp. 157–160. DOI: ISBN978-0-262-01197-6..

Peter, C., W. de Lange, J. Musango, K. April and A. Potgieter (Dec. 2009). 'Applying Bayesian modelling to assess climate change effects on biofuel production'. *Climate Research* 40, 249–260. DOI: 10.3354/cr00833.

Peters, J., P. Bühlmann and N. Meinshausen (Nov. 2016). 'Causal inference by using invariant prediction: identification and confidence intervals'. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 78 (5), 947–1012. DOI: 10.1111/rssb.12167.

Peters, J., D. Janzing and B. Schölkopf (2017). *Elements of causal inference: foundations and learning algorithms*. Cambridge, MA: MIT Press, pp. 1214–1216.

Pfleiderer, P., C.-F. Schleussner, T. Geiger and M. Kretschmer (2020). 'Robust predictors for seasonal Atlantic hurricane activity identified with causal effect networks'. *Weather and Climate Dynamics* 1 (2), 313–324. DOI: 10.5194/wcd-1-313-2020.

Plumb, R. A. (1989). 'On the seasonal cycle of stratospheric planetary waves'. *Pure and Applied Geophysics* 130, 233–242. DOI: 10.1007/BF00874457.

Polichtchouk, I., T. G. Shepherd and N. J. Byrne (2018). 'Impact of Parametrized Nonorographic Gravity Wave Drag on Stratosphere-Troposphere Coupling in the Northern and Southern Hemispheres'. *Geophysical Research Letters* 45. DOI: 10.1029/2018GL078981.

Polkova, I. et al. (2021). 'Predictors and prediction skill for marine cold-air outbreaks over the Barents Sea'. *Quarterly Journal of the Royal Meteorological Society* 147 (738), 1–19. DOI: {10.1002/qj.4038}.

Porfiri, M. and M. R. Marin (2018). 'Inference of time-varying networks through transfer entropy, the case of a Boolean network model'. *Chaos* 28, 103123.

Ragone, F., J. Wouters and F. Bouchet (2017). 'Computation of extreme heat waves in climate models using a large deviation algorithm'. *Proceedings of the National Academy of Sciences* 115 (1), 24–29. DOI: 10.1073/pnas.1712645115.

Ramazi, P., M. Kunegel-Lion, R. Greiner and M. A. Lewis (Oct. 2020). 'Exploiting the full potential of Bayesian networks in predictive ecology'. *Methods in Ecology and Evolution* 12 (1). Ed. by A. Ellison, 135–149. DOI: 10.1111/2041-210x.13509.

Randel, W. and F. Wu (1999). 'A stratospheric ozone trends data set for global modeling studies'. *Geophysical Research Letters* 26 (20), 3089–3092. DOI: 10.1029/1999GL900615.

Rao, J., C. I. Garfinkel and I. P. White (2021). 'Development of the Extratropical Response to the Stratospheric Quasi-Biennial Oscillation'. *Journal of Climate* 34 (17), 7239–7255. DOI: 10.1175/JCLI-D-20-0960.1.

Rao, J., C. I. Garfinkel, I. P. White and C. Schwartz (2020). 'The Southern Hemisphere minor sudden stratospheric warming in September 2019 and its predictions in S2S models'. *Journal of Geophysical Research D: Atmospheres* 125, e2020JD032723. DOI: https://doi.org/10.1029/2020JD032723.

Reichler, T. and M. Jucker (2022). 'Stratospheric wave driving events as an alternative to sudden stratospheric warmings'. *Weather and Climate Dynamics* 3 (2), 659–677. DOI: 10.5194/wcd-3-659-2022.

Rhines, P. B. (1975). 'Waves and turbulence on a beta-plane'. *Journal of Fluid Mechanics* 69 (3), 417–443. DOI: 10.1017/S0022112075001504.

Richard, S. S., R. D. Anne, A. N. Paul, P. Steven and R. S. Mark (Jan. 2010). 'Relative Contribution of Greenhouse Gases and Ozone-Depleting Substances to Temperature Trends in the

Stratosphere: A Chemistry–Climate Model Study'. *Journal of Climate* 23 (1), 28–42. DOI: 10.1175/2009jcli2955.1.

Richens, J. G., C. Lee and S. Johri (2020). 'Improving the accuracy of medical diagnosis with causal machine learning'. *Nature Communications* 11 (1). DOI: 10.1038/s41467-020-17419-7.

Risbey, J., T. O'Kane, D. Monselesan, C. Franzke and I. Horenko (2015). 'Metastability of Northern Hemisphere teleconnection modes'. *Journal of the Atmospheric Sciences* 72 (1), 35–54.

Rodrigues, R. R. and T. G. Shepherd (2022). 'Small is beautiful: climate-change science as if people mattered'. *PNAS Nexus* 1 (1), pgac009. DOI: 10.1093/pnasnexus/pgac009.

Roff, G., D. W. J. Thompson and H. Hendon (Mar. 2011). 'Does increasing model stratospheric resolution improve extended-range forecast skill?' *Geophysical Research Letters* 38 (5). DOI: 10.1029/2010gl046515.

Rohrer, J. M. (2018). 'Thinking clearly about correlations and causation: Graphical causal models for observational data.' *Advances in Methods and Practices in Psychological Science* 1, 27–42. DOI: 10.1177/2515245917745629.

Runge, J. (Aug. 2014). 'Detecting and quantifying causality from time series of complex systems'. PhD thesis. University of Humboldt Berlin, Faculty of Mathematics and Natural Sciences.

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. DOI: 10.1063/1.5025050.

Runge, J., J. Heitzig, N. Marwan and J. Kurths (2012a). 'Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy'. *Physical Review E* 86 (6), 061121. DOI: 10.1103/PhysRevE.86.061121.

Runge, J., J. Heitzig, V. Petoukhov and J. Kurths (2012b). 'Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy'. *Physical Review Letters* 108 (25). DOI: 10.1103/PhysRevLett.108.258701.

Runge, J., P. Nowack, M. Kretschmer, S. Flaxman and D. Sejdinovic (2019a). 'Detecting and quantifying causal associations in large nonlinear time series datasets'. *Science Advances* 5 (11). DOI: 10.1126/sciadv.aau4996.

Runge, J., V. Petoukhov and J. Kurths (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. DOI: 10.1175/jcli-d-13-00159.1.

Runge, J. et al. (2019b). 'Inferring causation from time series in Earth system sciences'. *Nature Communications* 10 (1), 2553. DOI: 10.1038/s41467-019-10105-3.

Runge, J. (2020). 'Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets'. In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*. Ed. by D. Sontag and J. Peters. AUAI Press.

Saggioro, E. and T. G. Shepherd (2019). 'Quantifying the timescale and strength of Southern Hemisphere intra-seasonal stratosphere-troposphere coupling'. *Geophysical Research Letters* 46 (22), 13479–13487. DOI: 10.1029/2019gl084763.

Saggioro, E., J. de Wiljes, M. Kretschmer and J. Runge (2020). 'Reconstructing regime-dependent causal relationships from observational time series'. *Chaos* 30 (11), 113115. DOI: 10.1063/5.0020538.

Samarasinghe, S., M. McGraw, E. Barnes and I. Ebert-Uphoff (2018). 'A study of links between the Arctic and the midlatitude jet stream using Granger and Pearl causality.' *Environmetrics* 30 (4), e2540. DOI: 10.1002/env.2540.

Sardeshmukh, P. D. and B. J. Hoskins (Apr. 1988). 'The Generation of Global Rotational Flow by Steady Idealized Tropical Divergence'. *Journal of the Atmospheric Sciences* 45 (7), 1228–1251. DOI: 10.1175/1520-0469(1988)045<1228:tgogrf>2.0.co;2.

Scaife, A. A. and I. N. James (2000). 'Response of the stratosphere to interannual variability of tropospheric planetary waves.' *Quarterly Journal of the Royal Meteorological Society* 126, 275–297. DOI: 10.1002/qj.49712656214..

Scaife, A. A. et al. (2022). 'Long-range prediction and the stratosphere'. *Atmospheric Chemistry and Physics* 22 (4), 2601–2623. DOI: 10.5194/acp-22-2601-2022.

Scaife, A. A. et al. (May 2012). 'Climate change projections and stratosphere–troposphere interaction'. *Climate Dynamics* 38 (9-10), 2089–2097. DOI: 10.1007/s00382-011-1080-7.

Scherhag, R. (1952). 'Die explosionsartigen Stratosphärenerwärmungen des Spätwinters'. *Berichte des Dtsch. Wetterdienstes aus der US-Zone* 6 (38), 51–63.

Scott, R. K. and P. H. Haynes (2000). 'Internal vacillations in stratosphere-only models'. *Journal of the Atmospheric Sciences* 57 (19), 3233–3250. DOI: 10.1175/1520-0469(2000)057<3233:IVISOM>2.0.CO;2.

Scott, R. K. and P. Haynes (2002). 'The seasonal cycle of planetary waves in the winter stratosphere'. *Journal of the Atmospheric Sciences* 59 (4), 803–822. DOI: 10.1175/1520-0469(2002)059<0803:TSCOPW>2.0.CO;2.

Scott, R. and P. Haynes (1998). 'Internal interannual variability of the extratropical stratospheric circulation: The low-latitude flywheel.' *Quarterly Journal of the Royal Meteorological Society* 124 (550), 2149–2173.

Seager, R., N. Harnik, Y. Kushnir, W. Robinson and J. Miller (2003). 'Mechanisms of hemispherically symmetric climate variability'. *Journal of Climate* 16 (18), 2960–2978.

Seber, G. A. F. and A. J. Lee (2014). *Linear Regression Analysis*. Wiley, 2nd Edition.

Seixas, F. L., B. Zadrozny, J. Laks, A. Conci and D. Muchaluat Saade (2014). 'A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment'. *Computers in Biology and Medicine* 51, 140–158. DOI: {10.1016/j.compbiomed.2014.04.010}.

Sen Gupta, A. and M. England (2006). 'Coupled Ocean-Atmosphere-Ice response to variations in the Southern Annular Mode'. *Journal of Climate* 19, 4457–4458. DOI: 10.1175/JCLI3843.1.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky and A. J. Teuling (2010). 'Investigating soil moisture-climate interactions in a changing climate: A review'. *Earth-Science Reviews* 99 (3), 125–161. DOI: 10.1016/j.earscirev.2010.02.004.

Seviour, W. J. M., S. C. Hardiman, L. J. Gray, N. Butchart, C. MacLachlan and A. A. Scaife (2014). 'Skillful Seasonal Prediction of the Southern Annular Mode and Antarctic Ozone'. *Journal of Climate* 27 (19), 7462–7474. DOI: 10.1175/jcli-d-14-00264.1.

Shafer, G. (1993). *The Art of Causal Conjecture*. Cambridge, MA: MIT Press.

Shaman, J. and E. Tziperman (2007). 'Summertime ENSO–North African–Asian Jet teleconnection and implications for the Indian monsoons'. *Geophysical Research Letters* 34 (11), L11702. DOI: 10.1029/2006GL029143.

Shen, X., L. Wang, S. Osprey, S. C. Hardiman, A. A. Scaife and J. Ma (2022). 'The Life Cycle and Variability of Antarctic Weak Polar Vortex Events'. *Journal of Climate* 35 (6), 2075–2092. DOI: 10.1175/JCLI-D-21-0500.1.

Shepherd, T. G. (2014). 'Atmospheric circulation as a source of uncertainty in climate change projections.' *Nature Geoscience* 7, 703–708. DOI: 10.1038/ngeo2253.

Shepherd, T. G., I. Polichtchouk, R. Hogan and A. J. Simmons (2018). *Report on the stratosphere task force*. Technical Report 824. European Center for Medium-Range Weather Forecasts.

Shepherd, T. (2021). 'Bringing physical reasoning into statistical practice in climate-change science'. *Climatic Change* 169 (2). DOI: 10.1007/s10584-021-03226-6.

Shipley, B. (2016). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R*. 2nd. Cambridge University Press.

Shipley, B. and J. C. Douma (2020). 'Generalized AIC and chi-squared statistics for path models consistent with directed acyclic graphs'. *Ecology* 101 (3), e02960.

Sigmond, M., J. F. Scinocca, V. V. Kharin and T. G. Shepherd (2013). 'Enhanced seasonal forecast skill following stratospheric sudden warmings'. *Nature Geoscience* 6 (2), 98–102. DOI: 10.1038/ngeo1698.

Sigmond, M. and J. F. Scinocca (Mar. 2010). 'The Influence of the Basic State on the Northern Hemisphere Circulation Response to Climate Change'. *Journal of Climate* 23 (6), 1434–1446. DOI: 10.1175/2009jcli3167.1.

Silvestri, G. and C. Vera (2003). 'Antarctic oscillation signal on precipitation anomalies over southeastern South America'. *Geophysical Research Letters* 30 (21), 2115–2118. DOI: 10.1029/2003GL018277.

Simpson, I. R., P. Hitchcock, T. G. Shepherd and J. F. Scinocca (2011). 'Stratospheric variability and tropospheric annular mode timescales.' *Geophysical Research Letters* 38 (L20806). DOI: 10.1029/2011GL049304.

Simpson, I. R. and L. M. Polvani (2016). 'Revisiting the relationship between jet position, forced response, and annular mode variability in the southern midlatitudes.' *Geophysical Research Letters* 43, 2896–2903. DOI: 10.1002/2016GL067989.

Simpson, I. R., T. G. Shepherd, P. Hitchcock and J. F. Scinocca (2013). 'Southern Annular Mode Dynamics in Observations and Models. Part II: Eddy Feedbacks.' *Journal of Climate* 26, 5220–5241. DOI: 10.1175/JCLI-D-12-00495.1.

Son, S., A. Purich, H. H. Hendon, B. Kim and L. M. Polvani (2013). 'Improved seasonal forecast using ozone hole variability?' *Geophysical Research Letters* 40 (23), 6231–6235. DOI: 10.1002/2013gl057731.

Spiegelhalter, D., S. Laurizten, A. Dawid and R. Cowell (1993). 'Bayesian analysis in expert systems (with Discussion)'. *Statistical Sciences* 8 (3), 219–274. DOI: https://www.jstor.org/stable/2245959.

Spirtes, P. and C. Glymour (1991). 'An Algorithm for Fast Recovery of Sparse Causal Graphs.' *Social Science Computer Review* 9 (1), 62–72. DOI: 10.1177/089443939100900106.

Spirtes, P., C. Glymour and R. Scheines (1993). *Causation, prediction, and search*. New York, NY: Springer.

Spirtes, P., C. Glymour and R. Scheines (2000). *Causation, prediction, and search*. 2nd. Vol. 81. Cambridge, MA: MIT Press.

Stone, K. A., S. Solomon, D. W. J. Thompson, D. E. Kinnison and J. C. Fyfe (2022). 'On the Southern Hemisphere Stratospheric Response to ENSO and Its Impacts on Tropospheric Circulation'. *Journal of Climate* 35 (6), 1963–1981. DOI: 10.1175/jcli-d-21-0250.1.

Strandburg-Peshkin, A., D. Papageorgiou, M. C. Crofoot and D. R. Farine (2018). 'Inferring influence and leadership in moving animal groups'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373 (1746), 20170006. DOI: 10.1098/rstb.2017.0006.

Sugihara, G., R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty and S. Munch (Oct. 2012). 'Detecting causality in complex ecosystems.' *Science* 338 (6106), 496–500. DOI: 10.1126/science.1227079.

Sun, L., G. Chen and W. Robinson (2014). 'The Role of Stratospheric Polar Vortex Breakdown in Southern Hemisphere Climate Trends'. *Journal of the Atmospheric Sciences* 71, 2335–2353. DOI: 10.1175/JAS-D-13-0290.1.

Swart, N. C., J. C. Fyfe, E. Hawkins, J. E. Kay and A. Jahn (2015). 'Influence of internal variability on Arctic sea-ice trends'. *Nature Climate Change* 5 (2), 86–89. DOI: 10.1038/nclimate2483.

Textor, J., B. van der Zander, M. S. Gilthorpe, M. Liśkiewicz and G. T. Ellison (2016). 'Robust causal inference using directed acyclic graphs: the R package dagitty'. *International Journal of Epidemiology* 45 (6), 1887–1894. DOI: 10.1093/ije/dyw341.

Thompson, D. W. J., M. P. Baldwin and S. Solomon (2005). 'Stratosphere-Troposphere Coupling in the Southern Hemisphere'. *Journal of the Atmospheric Sciences* 62 (3), 708–715. DOI: 10.1175/jas-3321.1.

Thompson, D. W. J., J. C. Furtado and T. G. Shepherd (2006). 'On the Tropospheric Response to Anomalous Stratospheric Wave Drag and Radiative Heating'. *Journal of the Atmospheric Sciences* 63 (10), 2616–2629. DOI: 10.1175/JAS3771.1.

Thompson, D. W. J. and S. Solomon (2002). 'Interpretation of recent Southern Hemisphere climate change'. *Science* 296, 895–899. DOI: 10.1126/science.1069270.

Thompson, D. W. J., S. Solomon, P. J. Kushner, M. H. England, K. M. Grise and D. J. Karoly (2011). 'Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change'. *Nature Geoscience* 4. DOI: 10.1038/ngeo1296.

Thompson, D. W. J. and J. M. Wallace (2000). 'Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability'. *Journal of Climate* 13 (5), 1000–1016. DOI: 10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.

Tikhonov, A. N., A. Goncharsky, V. V. Stepanov and A. G. Yagola (1995). *Numerical Methods for the Solution of Ill-Posed Problems*. Springer.

Tikka, S. and J. Karvanen (2017). 'Identifying Causal Effects with the R Package causaleffect'. *Journal of Statistical Software* 76 (12), 1–30. DOI: 0.18637/jss.v076.i12.

Tripathi, O. P. et al. (2014). 'The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts'. *Quarterly Journal of the Royal Meteorological Society* 141 (689), 987–1003. DOI: 10.1002/qj.2432.

Van Loon, H. (Sept. 1967). 'The Half-Yearly Oscillations in Middle and High Southern Latitudes and the Coreless Winter'. *Journal of the Atmospheric Sciences* 24 (5), 472–486. DOI: 10.1175/1520-0469(1967)024<0472:thyoim>2.0.co;2.

Van Loon, H., C. Zerefos and C. Repapis (1982). 'The Southern Oscillation in the Stratosphere'. *Monthly Weather Review* 110 (3), 225–229. DOI: 10.1175/1520-0493(1982)110<0225:TSOITS>2.0.CO;2.

Vera, C., G. Silvestri, V. Barros and A. Carril (May 2004). 'Differences in El Niño Response over the Southern Hemisphere'. *Journal of Climate* 17 (9), 1741–1753. DOI: 10.1175/1520-0442(2004)017<1741:dienro>2.0.co;2.

Walker, G. T. (1925). 'Correlation in seasonal variations of weather—A further study of world weather.' *Monthly Weather Review* 53, 252–254. DOI: 10.1175/1520-0493(1925)53<252:CISVOW>2.0.CO;2.

Walker, G. T. and E. W. Bliss (1932). 'World Weather V'. *Memoirs of the Royal Meteorological Society* 4 (36), 53–84.

Walland, D. and I. Simmonds (1999). 'Baroclinicity, Meridional Temperature Gradients, and the Southern Semiannual Oscillation'. *Journal of Climate* 12 (12), 3376–3382. DOI: 10.1175/1520-0442(1999)012<3376:bmtgat>2.0.co;2.

Wang, L., M. Ting and P. Kushner (2017). 'A robust empirical seasonal prediction of winter NAO and surface climate'. *Nature Scientific Reports* 7 (279). DOI: 10.1038/s41598-017-00353-y.

Wang, S., J. Liu, X. Cheng, T. Kerzenmacher and P. Braesicke (2020). 'Is enhanced predictability of the Amundsen Sea low in subseasonal to seasonal hindcasts linked to stratosphere-troposphere coupling?' *Geophysical Research Letters* 47 (e2020GL089700). DOI: 10.1029/2020GL089700.

Wang, S., J. Liu, X. Cheng, T. Kerzenmacher, Y. Hu, F. Hui and P. Braesicke (2021). 'How do weakening of the stratospheric polar vortex in the Southern Hemisphere affect regional Antarctic sea ice extent?' *Geophysical Research Letters* 48, e2021GL092582. DOI: 10.1029/2020gl089700.

Waugh, D. and L. M. Polvani (2010). 'Stratospheric Polar Vortices'. In: *The Stratosphere: Dynamics, Transport, and Chemistry*. American Geophysical Union. Chap., 43–57. DOI: 10.1002/9781118666630.ch3.

Waugh, D., W. Randel, S. Pawson, P. Newman and E. Nash (1999). 'Persistence of the Lower Stratospheric Polar Vortices'. *Journal of Geophysical Research: Atmospheres* 104 (D22), 27191–27201. DOI: 10.1029/1999JD900795..

Webber, R. J., D. A. Plotkin, M. E. O. Neill, D. S. Abbot and J. Weare (2019). 'Practical rare event sampling for extreme mesoscale weather'. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29 (5), 053109. DOI: 10.1063/1.5081461.

Webster, P. J. and T. N. Palmer (1997). 'The past and the future of El Niño'. *Nature* 390, 562–564.

Wermuth, N. (1980). 'Linear recursive equations, covariance selection, and path analysis.' *Journal of the American Statistical Association* 75, 963–972.

Wermuth, N. and S. Laurizten (1983). 'Graphical and recursive models for contingency tables'. *Biometrika* 70 (3), 537–552.

White, C. J. et al. (2017). 'Potential applications of subseasonal-to-seasonal (S2S) predictions'. *Meteorological Applications* 24 (3), 315–325. DOI: 10.1002/met.1654.

White, I. P., C. I. Garfinkel, E. P. Gerber, M. Jucker, P. Hitchcock and J. Rao (2020). 'The Generic Nature of the Tropospheric Response to Sudden Stratospheric Warmings'. *Journal of Climate* 33 (13), 558–5610. DOI: 10.1175/JCLI-D-19-0697.1.

White, I. P., C. I. Garfinkel and P. Hitchcock (2022). 'On the Tropospheric Response to Transient Stratospheric Momentum Torques'. *Journal of the Atmospheric Sciences* 79 (8), 2041–2058. DOI: 10.1175/JAS-D-21-0237.1.

Wicker, W., I. Polichtchouk and D. I. V. Domeisen (2022). 'Increased vertical resolution in the stratosphere reveals role of gravity waves after sudden stratospheric warmings'. *Weather and Climate Dynamics Discussions* Under review, 1–24. DOI: 10.5194/wcd-2022-41.

Wiljes, J. de, L. Putzig and I. Horenko (2014). 'Discrete nonhomogeneous and nonstationary logistic and Markov regression models for spatiotemporal data with unresolved external influences'. *Communications in Applied Mathematics and Computational Science* 284, 184–193.

Williams, P. D. et al. (2017). 'A Census of Atmospheric Variability From Seconds to Decades'. *Geophysical Research Letters* 44 (21), 201–211.

Wilson, A. B., D. H. Bromwich and K. M. Hines (2016). 'Simulating the Mutual Forcing of Anomalous High Southern Latitude Atmospheric Circulation by El Niño Flavors and the Southern Annular Mode'. *Journal of Climate* 29 (6), 2291–2309. DOI: 10.1175/jcli-d-15-0361.1.

Yamashita, Y., H. Akiyoshi, T. G. Shepherd and M. Takahashi (2015). 'The Combined Influences of Westerly Phase of the Quasi-Biennial Oscillation and 11-year Solar Maximum Conditions on the Northern Hemisphere Extratropical Winter Circulation'. *Journal of Meteorological Society of Japan* 93 (6), 629–644. DOI: 10.2151/jmsj.2015-054.

Yamashita, Y., H. Naoe, M. Inoue and M. Takahashi (2018). 'Response of the Southern Hemisphere Atmosphere to the Stratospheric Equatorial Quasi-Biennial Oscillation (QBO) from Winter to Early Summer'. *Journal of the Meteorological Society of Japan, Ser. II* 96 (6), 587–600. DOI: 10.2151/jmsj.2018-057.

Yang, D., J. M. Arblaster, G. A. Meehl, M. H. England, E.-P. Lim, S. Bates and N. Rosenbloom (July 2020). 'Role of Tropical Variability in Driving Decadal Shifts in the Southern Hemisphere Summertime Eddy-Driven Jet'. *Journal of Climate* 33 (13), 5445–5463. DOI: 10.1175/jcli-d-19-0604.1.

Yoden, S., M. Taguchi and Y. Naito (2002). 'Numerical Studies on Time Variations of the Troposphere-Stratosphere Coupled System.' *Journal of the Meteorological Society of Japan* 80 (4B), 811–830. DOI: 10.2151/jmsj.80.811.

Young, H., T. G. Shepherd, J. Acidri, R. Cornforth, C. Petty, J. Seaman and L. Todman (2020). 'Storylines for decision-making: climate and food security in Namibia'. *Climate and Development* 13 (6), 515–528. DOI: 10.1080/17565529.2020.1808438.

Young, P. et al. (2013). 'Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP)'. *Atmospheric Chemistry and Physics* 13 (4), 2063–2090. DOI: https://doi.org/10.5194/acp-13-2063-2013.

Yu, J.-Y., H. Paek, E. S. Saltzman and T. Lee (Dec. 2015). 'The Early 1990s Change in ENSO–PSA–SAM Relationships and Its Impact on Southern Hemisphere Climate'. *Journal of Climate* 28 (23), 9393–9408. DOI: 10.1175/jcli-d-15-0335.1.

Zanin, M. and D. Papo (2017). 'Detecting switching and intermittent causalities in time series'. *Chaos* 27 (4), 047403. DOI: 10.1063/1.4979046.

Zhang, G. and V. V. Thai (2016). 'Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review'. *Safety Science* 87, 53–62. DOI: 10.1016/j.ssci.2016.03.019.

Zhang, K., B. Huangy, J. Zhang, C. Glymour and B. Schölkopf (2017). 'Causal discovery from Nonstationary/heterogeneous data: Skeleton estimation and orientation determination'. In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 1347–1353.

Zubiaurre, I. and N. Calvo (2012). 'The El Niño–Southern Oscillation (ENSO) Modoki signal in the stratosphere'. *Journal of Geophysical Research: Atmospheres* 117 (D4), DOI: 10.1029/2011JD016690.

Zwiers, F. and H. V. Storch (1990). 'Regime-Dependent Autoregressive Time Series Modeling of the Southern Oscillation'. *Journal of Climate* 3, 1347–1363.