Probabilistic causal network modelling of Southern Hemisphere jet sub-seasonal to seasonal predictability

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Saggioro, E. ORCID: https://orcid.org/0000-0002-9543-6338, Shepherd, T. G. ORCID: https://orcid.org/0000-0002-6631-9968 and Knight, J. (2024) Probabilistic causal network modelling of Southern Hemisphere jet sub-seasonal to seasonal predictability. Journal of Climate, 37 (10). pp. 3055-3071. ISSN 1520-0442 doi: 10.1175/JCLI-D-23-0425.1 Available at https://centaur.reading.ac.uk/115586/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1175/JCLI-D-23-0425.1

www.reading.ac.uk/centaur

Probabilistic Causal Network Modeling of Southern Hemisphere Jet Subseasonal to Seasonal Predictability

Elena Saggioro[a], Theodore G. Shepherd,[b] and Jeff Knight[c]

[a] *Department of Mathematics and Statistics, University of Reading, Reading, United Kingdom*
[b] *Department of Meteorology, University of Reading, Reading, United Kingdom*
[c] *Met Office Hadley Centre, Exeter, United Kingdom*

ABSTRACT: Skillful prediction of the Southern Hemisphere (SH) eddy-driven jet is crucial for representation of mid-to-high-latitude SH climate variability. In the austral spring-to-summer months, the jet and the stratospheric polar vortex variabilities are strongly coupled. Since the vortex is more predictable and influenced by long-lead drivers 1 month or more ahead, the stratosphere is considered a promising pathway for improving forecasts in the region on subseasonal to seasonal (S2S) time scales. However, a quantification of this predictability has been lacking, as most modeling studies address only one of the several interacting drivers at a time, while statistical analyses quantify association but not skill. This methodological gap is addressed through a knowledge-driven probabilistic causal network approach, quantified with seasonal ensemble hindcast data. The approach enables to quantify the jet's long-range predictability arising from known late-winter drivers, namely, El Niño–Southern Oscillation (ENSO), Indian Ocean dipole (IOD), upward wave activity flux, and polar night jet oscillation, mediated by the vortex variability in spring. Network-based predictions confirm the vortex as determinant for skillful jet predictions, both for the jet's poleward shift in late spring and its equatorward shift in early summer. ENSO, IOD, late-winter wave activity flux, and polar night jet oscillation only provide moderate prediction skill to the vortex. This points to early spring submonthly variability as important for determining the vortex state leading up to its breakdown, creating a predictability bottleneck for the jet. The method developed here offers a new avenue to quantify the predictability provided by multiple, interacting drivers on S2S time scales.

SIGNIFICANCE STATEMENT: Predictions of the Southern Hemisphere midlatitude jet stream are crucial for skillful forecasts of the austral mid-to-high latitudes. Several oceanic and atmospheric phenomena could, if better represented in models, improve spring-to-summer jet predictions on subseasonal to seasonal time scales. However, the combined potential skill arising from the inclusion of such phenomena has not been quantified. This study does so by using a probabilistic causal network model, representing the connections between those drivers and the jet with conditional probabilities, trained on large sets of model data. The stratospheric polar vortex is confirmed as crucial predictor of jet variability but is itself hard to predict a month in advance due to submonthly variability, creating a predictability bottleneck for the jet.

---

1. Introduction

The midlatitude eddy-driven jet stream is a dominant feature of the Southern Hemisphere (SH) large-scale circulation. Via its association with storm tracks and low pressure systems, the jet influences surface climate in the mid- and high latitudes, including over Australia, New Zealand, South America, and Antarctica, even creating conditions for high-impact weather, such as wildfires in Australia (Lim et al. 2019). Jet anomalies strongly project onto the Southern Annular Mode, the dominant mode of SH extratropical tropospheric circulation variability at time scales longer than a couple of weeks (Thompson and Wallace 2000).

From late winter through early summer, the SH jet is often represented in terms of the zonally and height averaged wind, here referred to as the eddy-driven jet (EDJ), due to its presence across all longitudes and its equivalent barotropic structure (Lorenz and Hartmann 2001; Byrne et al. 2017; Saggioro and Shepherd 2019). The main feature of the EDJ dynamics is a twice-yearly vacillation in latitude and strength with a contraction and strengthening, and expansion and weakening of eddy activity, called the semiannual oscillation (Bracegirdle 2011). Between September and January, this vacillation can be described as two consecutive latitudinal migrations of the EDJ around roughly 50°S in opposite directions (Fig. 1a): first poleward between September and October (*EDJ-1*; abbreviations introduced for the first time in italics are for variables
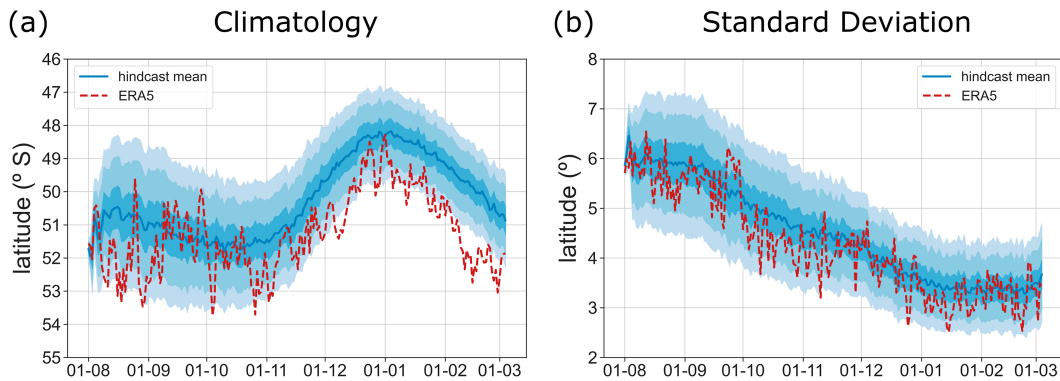

(a) Climatology

(b) Standard Deviation

FIG. 1. Climatology of (a) daily EDJ mean latitude and (b) daily interannual latitude standard deviation between 1 Aug and 1 Mar for the years 1981–2016, for ERA5 (red dashed line) and ECMWF System 4 hindcast initialized on 1 Aug (blue line for the mean and increasingly lighter blue bands for 25th–75th, 5th–95th, and 1st–99th percentile uncertainty ranges, generated with 10 000 bootstrapped time series where an ensemble member has been randomly selected from each year in the 36-yr period). Latitude is defined as the position of the maximum daily zonal-mean zonal wind at 850 hPa, between 35° and 70°S. Figure adapted from Byrne et al. (2019).

that will be included in the subsequent causal network analysis) and then equatorward between November and December stretching into January (*EDJ-2*). Year-to-year differences in timing and amplitude of EDJ-1 and EDJ-2 shifts range between 3° and 7° (Fig. 1b), with a significant impact on the surface climate (Bracegirdle 2011; Byrne and Shepherd 2018) because the EDJ position is associated with negative geopotential height anomalies and cooler and wetter conditions, as shown in Figs. S1 and S2 in the online supplemental material.

a. Stratospheric and tropospheric long-lead drivers of EDJ variability

This work is concerned with long-range predictability of the EDJ, defined as 1 month or more into the future. The key long-lead drivers of EDJ predictability, and their intermediate mediating variables, are identified through a literature review. Special attention is dedicated to the subseasonal to seasonal (S2S) time scales, from 2 weeks to a season, which overlaps with the long-range time scales.

Modeling studies have provided robust evidence for S2S predictability of the extratropical SH troposphere during spring and summer (Roff et al. 2011; Lim et al. 2013; Son et al. 2013; Seviour et al. 2014; Osman et al. 2015). Importantly, tropospheric prediction skill comes in part from the stratosphere (Seviour et al. 2014; Byrne and Shepherd 2018; Lim et al. 2018), a conclusion that is robust to sampling uncertainty (Byrne et al. 2019). This is due to a strong coupling between the stratospheric and EDJ variability around the timing of the stratospheric polar vortex (SPV) breakdown (Kuroda and Kodera 1998; Thompson et al. 2005; Byrne et al. 2017; Domeisen et al. 2020b). More specifically, the magnitude of the EDJ-1 shift in October is influenced by the strength of the lower stratospheric SPV (*SPV-low*) in early spring, a result obtained from both reanalysis (Bracegirdle 2011; Byrne 2017) and models (Seviour et al. 2014). And the timing of the EDJ-2 shift strongly follows the timing of the vortex breakdown (VB)

(Black and McDaniel 2007; Bracegirdle 2011; Byrne et al. 2017; Ceppi and Shepherd 2019). Westerly phases of the stratospheric quasi-biennial oscillation (*QBO*) are associated with stronger SPV-low and later VB, and vice versa for easterly QBO phases (Baldwin and Dunkerton 1988; Anstey and Shepherd 2014; Byrne and Shepherd 2018). The coupled stratosphere–troposphere variability in springtime can be traced back to the state of the vortex near the stratopause as early as June (Lim et al. 2018), with a progression that resembles the poleward and downward evolution of the polar night jet oscillation (PJO) (Kuroda and Kodera 2001; Kuroda 2002) (represented in two phases as *PJO-1* and *PJO-2* in the network), which could provide predictability for SPV-low in spring (Byrne et al. 2019). Stratospheric ozone concentrations also influence vortex variability on S2S time scales (Hendon et al. 2020; Oh et al. 2022). The development of the Antarctic ozone hole during the last decades of the twentieth century resulted in a strengthening of SPV-low and delayed VB by about 2 weeks (Waugh et al. 1999; McLandress et al. 2010), with a resulting downward effect on the jet (McLandress et al. 2011; Saggioro and Shepherd 2019).

There are also tropospheric long-lead drivers influencing EDJ variability, the dominant one being El Niño–Southern Oscillation (*ENSO*). The ENSO fingerprint in summer is of an approximately zonally symmetric tropospheric response, characterized by a shift in latitude of the EDJ (e.g., Seager et al. 2003; L'Heureux and Thompson 2006; Lim et al. 2013). In early spring, the fingerprint of ENSO is more zonally asymmetric, reflecting the modulation of the Pacific–South American pattern (Kidson 1999; Mo and Paegle 2001; Vera et al. 2004). The effect of ENSO on the EDJ is also due to stratospheric mechanisms, as El Niño events can force stronger upward propagating planetary waves, which can advance the seasonal disruption of the vortex with a lead time of about 1 month, affecting in turn the EDJ (van Loon et al. 1982; Hardiman et al. 2011; Hu et al. 2014; Lim et al. 2018; Byrne et al. 2019; Stone et al. 2022). Studies suggest a stronger effect of central Pacific El Niño

events on the SH vortex compared to eastern Pacific events, while a possibly differentiated effect of La Niña is less well understood (Domeisen et al. 2019, for a review). Upward wave activity fluxes are generally represented via the zonal-mean eddy heat flux (*vT-flux*), with conditions in August and September capturing the precursor signal leading to vortex weakening (Lim et al. 2018; Rao et al. 2020). The Indian Ocean dipole (*IOD*) can also influence the vortex through generating Rossby wave trains in the upper troposphere that propagate to the extratropical region (Rao et al. 2020; Huang et al. 2021; Jucker and Reichler 2023). The IOD can not only be internally produced by Indian Ocean–atmosphere coupling (Saji et al. 1999; Yang et al. 2015) but also externally forced by ENSO (Yu and Lau 2005), and ENSO can also be forced by IOD (Luo et al. 2010; Annamalai et al. 2005).

b. Causal network approach to quantify S2S predictability

While several long-lead drivers of EDJ variability are known, a quantification of their combined influence on its long-range predictability has been lacking. Modeling studies interrogate the effect of specific initializations or nudging (e.g., Seviour et al. 2014; Rao et al. 2020, 2021), which allows one to control for only one or a small number of variables, but not to easily and efficiently quantify the interaction of more causal elements at the same time. This can be problematic as drivers may work in concert or in opposition to favor a given EDJ state. On the other hand, statistical studies generally look at composites of long-lead drivers given EDJ states to infer predictability (Lim et al. 2018; Byrne and Shepherd 2018; Lim et al. 2019), looking effectively at the conditional probability P(driver|EDJ). While a signal can be detected in this way, this does not guarantee strong predictability of the EDJ state given those preconditions, that is, P(EDJ|driver) could be small (or not very different from climatology), because the two probabilities differ by a factor derived from Bayes's theorem: P(EDJ|driver) $= [P$(EDJ)$/P$(driver)$]P$(driver|EDJ). Therefore, while the presence of composite patterns of drivers given a target event can suggest a route for enhanced predictability, only a (statistical) prediction can confirm and quantify it. Based on this understanding, the aim of this work is to quantify the potential long-range predictability of EDJ variability via statistical modeling and attribute it to specific (combinations of) recognized long-lead drivers identified above, namely, ENSO, late-winter vT-flux, QBO, ozone, IOD, PJO-1, and PJO-2, mediated by the evolution of stratospheric variability represented by SPV-low and VB.

The statistical model of choice is a probabilistic causal network, chosen as it removes potentially misleading causal interpretation of correlational signals by design. A causal network is a mathematical model that can be used to test and quantify causal statements made about a system of variables (Pearl 2009). In a causal network, the variables are mapped onto nodes and the causal relationships between variables onto directed links. As a result, each variable is causally influenced directly by a (possibly empty) set of variables, called its parents. Given the probabilistic nature of atmospheric systems

on S2S time scales, a probabilistic causal network (PCN) is preferred to a deterministic one, because uncertainty is represented explicitly and quantified through conditional probabilities. A probabilistic description also allows us to account for not only very general and nonlinear connections between variables, such as nonlinear functional dependences of one variable on another, but also state-dependent connections, where the effect of one variable on another depends on the state of a third variable. Indeed, to compute a PCN's conditional probabilities, it is not required to prescribe any functional dependencies between the variables.

It is important to note that the PCN is not built as an optimal predictive model, for example, by searching for the set of drivers that maximize the prediction skill of the target EDJ. Rather, the PCN is meant as a model to quantify the long-range predictability of EDJ provided by well-known drivers when considered in various combinations. Hence, this analysis is explicitly motivated by, and built on, previous work.

The remainder of the paper is organized as follows. The methods and data are introduced in sections 2 and 3. The PCN is derived in section 4. Experiments using the PCN as a statistical model for EDJ variability are performed to quantify the potential predictability (and observed skill) of various combinations of long-lead drivers in section 5. The effect of improved EDJ predictions on surface climate is assessed in section 6. Discussion and conclusions are presented in sections 7 and 8. Abbreviations and acronyms used in the text are reported in appendix A.

2. Methods

In a causal network, a causal link $X \rightarrow Y$ corresponds to X and Y being conditionally dependent given all other parents of Y. In other words, the dependence between X and Y is genuine and not a spurious consequence of indirect connections. A causal network needs to be acyclic, which means that starting from any variable and following the direction of links, no path can lead back to its starting variable. Feedbacks can be incorporated by adding time lags between variables. Conditional dependence and acyclicity allow for the joint probability distribution $P(X_1, X_2, \ldots, X_N)$ of the system of N variables comprising the causal network to be factorized as the product of N conditional probabilities $P(X_i|\text{PA}_i)$, where PA_i are the parents of variable X_i (the variables directly influencing X_i):

$$P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} P(X_i|\text{PA}_i). \quad (1)$$

If the network is sparse (that is, with only a few links compared to all the possible ones), the number of values needed to quantify all of the probabilities $P(X_i|\text{PA}_i)$ for $i = 1, \ldots, N$ is much smaller than the number of values needed for $P(X_1, X_2, \ldots, X_N)$, resulting in a much faster computation of any derived marginal probability of interest (e.g., for probabilistic forecasts).

a. Building and quantifying a PCN

Several techniques exist to build a causal network (e.g., Marcot et al. 2006; Chen and Pollino 2012; Runge et al. 2019; Young et al. 2020). Here, a hybrid knowledge- and data-driven

approach is used. To start, a hypothetical PCN structure is drawn by including all long-lead drivers of EDJ, the key mediating variables, and all plausible links between them, selected according to the literature reviewed in section 1. Any known common driver of any pair of variables also needs to be added, a requirement ensuring that conditional independence really reflects the absence of links in the network (causal sufficiency; here none were needed). Links cannot point backward in time, as in a causal network they need to be causal not just correlational. When variables are contemporaneous on the considered time scales, the direction is set by known physical processes.

To this hypothetical network, the PC-stable causal discovery algorithm is applied (Colombo and Maathuis 2014), using the Pearson chi-square independence test with significance level $\alpha = 0.05$. The goal of the algorithm is to reveal the links such that the joint probability of the system can be expressed as in Eq. (1). To start, the unconditional independence of all pairs of variables is tested and any link is removed if the corresponding p value of the test is larger than α. Then, the remaining links $X \rightarrow Y$ are assessed via conditional independence tests using as conditions a selection of the parents of Y in an iterative fashion. In the first iteration, each link is tested using each time as condition one variable from its parents (until all parents are tested), and at the end, the link is removed if any of these tests have $p > \alpha$. In the second iteration, the remaining links are retested with now two conditions using the same procedure. The algorithm continues with iterations using higher-dimensional condition sets until no more conditions are left, and the latest set of links retained constitutes the final network.

Given the final links, the conditional probabilities $P(X_i|\mathrm{PA}_i)$ of each variable X_i given its parents are computed. Because here the variables are discretized (see section 3), the probabilities are discrete too and are called conditional probability tables (CPTs). Assuming multinomial sampling, appropriate for large sample sizes, the CPT entries are estimated as relative frequencies.

b. Predictions with a PCN

The PCN with its CPTs can be used to perform predictions, which consist of calculating the probability distribution $P(X_i|E)$ of a target variable X_i updated given information about some of its drivers E (also referred to as evidence). This is computed by marginalizing Eq. (1) over the unobserved covariates and using the definition of conditional probability.

The probabilistic forecasts produced are assessed with two measures of skill: the area under the receiver operating characteristic curve (ROC AUC), which considers the trade-off between true-positive rate and false-positive rate, and the area under the precision–recall curve (PR AUC), which considers the trade-off between true-positive rate and precision. Considering ROC AUC and PR AUC together allows us to give a more rounded assessment of the forecast than just using one of them. A perfect forecast gives ROC AUC = PR AUC = 1, and a random forecast gives ROC AUC = 0.5 and PR AUC = b, the baseline fraction of positives in the observed data. Hence, to compare skill across different categories and variables, the percentage increase from the baseline is used: PR AUC% = $100 \times (\mathrm{PR\ AUC} - b)/(1 - b)$, where 100% is a

perfect forecast and 0% indicates no skill. Note that skill here is used to quantify how well the PCN approximates the hindcast (when all parents of a variable are used to predict that target variable) and also what is the amount of information provided by specific drivers (when a selection of drivers is used to predict a target). Skill values are not computed with the aim of removing all possible sources of bias in their estimate and thus are not intended as an absolute measure of skill.

3. Data

The European Centre for Medium-Range Weather Forecasts (ECMWF) 51-member ensemble hindcast is used to derive the PCN, as it provides a sample size suitable for probabilistic quantification. Although the latest operational system is ECMWF's fifth generation seasonal forecast system (SEAS5), the earlier version known as System 4 is chosen (Molteni et al. 2011) because it has a more realistic SH circulation variability at polar latitudes, possibly due to a better representation of SH stratospheric variability (Shepherd et al. 2018). The hindcast is initialized on 1 August for each year between 1981 and 2016 (36 seasons) and run for 7 months, of which the first six are analyzed here (up to 31 January). ERA5 reanalysis is used as verification data, which is independent of System 4 (Hersbach et al. 2020).

The basic data input for the network variables is the daily mean zonal wind u, meridional wind v, air temperature T, and sea surface temperature (SST). Stratospheric ozone is not included because the expected vortex strengthening over the late twentieth century is not found in System 4 (see Fig. S3 in the supplemental material). This is likely due to its simplified ozone chemistry, which does not allow for a realistic representation of the polar stratospheric cloud chemistry required to produce an ozone hole, in addition to being initialized with an ozone climatology (Monge-Sanz et al. 2022). For the analysis of surface climate impacts, monthly means of geopotential height at 500 hPa (Z500), 2-m temperature (2mT), and mean total precipitation rate (MTPR) are used.

Unlike previous analyses of this hindcast for the SH (Byrne et al. 2019; Osman et al. 2022), the 2002/03 season is retained despite it including the only SH stratospheric sudden warming (SSW) in the observational record. An analysis of the 50-hPa 60°S zonal-mean zonal wind did not reveal an outlier behavior of these ensemble members, in that there are several other members in other years that show a rapid deceleration and even breakdown around early spring (not shown). An SH vortex breakdown that breaks (too) early is a feature documented in several S2S systems including ECMWF (Lawrence et al. 2022).

Network variables

The definitions of the 10 indices associated with the network variables are in Table 1. Note that EDJ-1 and EDJ-2 are defined at different vertical levels, in accordance with the literature reviewed. Several works focusing on EDJ-2, and its links with the VB, define it at 850 hPa (e.g., Byrne et al. 2019; Ceppi and Shepherd 2019; Saggioro and Shepherd 2019). The fewer studies for the connection of EDJ-1 with the vortex point at a range of tropospheric vertical levels between 1000 and 500 hPa

TABLE 1. Network variables. Definitions column: Square brackets ([]) indicate the zonal mean and the prime symbol (′) indicates anomalies with respect to the zonal mean. For EDJ-1 and EDJ-2, a running mean is used to extract the latitudinal shift signal from the very noisy daily latitude index. Discretization column: In the SH, the vT-flux has a negative sign; therefore, lower percentiles mean stronger wave activity. Percentiles are computed from the hindcast distribution. For EDJ-1 and EDJ-2, the percentages refer to the proportion of ensemble members assigned to each cluster.

Name	Definition (index; time span/average)	Discretization [No. of categories; threshold(s); labels]
ENSO	Monthly mean SST anomalies (wrt 1981–2016) area averaged over the Niño-3.4 region (5°N–5°S, 170°–120°W); August–October mean	3; ±0.4°C; Niña/neutral/Niño
QBO	Daily zonal-mean zonal wind $[u]$ at 30 hPa at 0°; August–October mean	2; 0 m s^{-1}; easterly/westerly
IOD	Difference in monthly mean SST anomalies (wrt 1981–2016) between the tropical western Indian Ocean (area averaged over 10°N–10°S, 50°–70°E) and the tropical southeastern Indian Ocean (area averaged over 0°–10°S, 90°–110°E); August–October mean	3; ±0.4°C; negative/neutral/positive
vT-flux	Daily eddy heat flux $[v'T']$ at 100 hPa averaged over 45°–75°S; August–September mean	3; 30th–75th percentile; strong/neutral/weak
PJO-1	Daily $[u]$ at 10 hPa averaged over 30°–45°S; August mean	2; 60th percentile; weak/strong
PJO-2	Daily $[u]$ at 30 hPa averaged over 50°–60°S; September mean	3; 35th–75th percentile; weak/neutral/strong
SPV-low	Daily $[u]$ mass-weighted between 100 and 50 hPa averaged over 60°–70°S; October mean	3; 25th–75th percentile; small/neutral/large
VB	Last date when the 5-day centered running mean of daily $[u]$ at 50 hPa and 60°S falls below 15 m s^{-1}, in days after 1 Oct; note that depending on the year VB spans October–January	3; 25th–75th percentile; early/middle/late
EDJ-1	Latitude of the maximum of daily $[u]$ pressure-weighted between 1000 and 500 hPa, 10-day running mean; categories clusters built using the time series between 1 Oct and 8 Nov	3; hierarchical clustering (32.5%, 52%, 15.5%); poleward/middle/equatorward
EDJ-2	Latitude of the maximum of daily $[u]$ at 850 hPa, 10-day running mean; categories clusters built using the time series between 1 Nov and 19 Jan	3; hierarchical clustering (26.5%, 51%, 22.5%); poleward/middle/equatorward

as relevant to capture this connection (Bracegirdle 2011; Seviour et al. 2014; Byrne 2017). A comparison of the different EDJ-1 categories using the 850-hPa level was performed, resulting in minor differences, a result largely expected given the approximately equivalent barotropic structure of the EDJ.

All the indices are transformed from continuous to discrete to allow for the generation of probabilistic forecasts of above, normal, or below climatological conditions (Table 1). For each variable, categories are defined to best summarize the ensemble spread. Thus, depending on the shape of its distribution and the physical meaning of the specific variable's values, the categories may not comprise the same number of data points. The categories are identified either based on physically meaningful thresholds (e.g., 0 m s^{-1} for QBO), based on percentiles (e.g., 25th–75th percentiles for VB), or using hierarchical clustering (for EDJ-1 and EDJ-2). Hierarchical clustering is a dendogram-based approach that divides a set of time series into clusters of a similar shape. The clustering is performed with Euclidean distance metric and Ward's linkage criterion so that clusters have the minimum distance between their elements and the maximum distance across clusters (Ward 1963). The EDJ categories defined via hierarchical clustering retain all the information relating to the timing and magnitude of the shifts: EDJ-1 categories can be summarized as an overall more poleward or equatorward position of the jet in October with respect to the climatological time series, and EDJ-2 categories are either a more poleward

and delayed shift or a more equatorward and early shift between November and December (see Figs. S5b,c).

Mean bias correction is performed so that categorization of variables in the hindcast is meaningful when compared with ERA5. Mean biases in ENSO, EDJ-1, and EDJ-2 in System 4 are already documented (Molteni et al. 2011; Byrne et al. 2019), but bias correction is performed to all variables for consistency. In practice, from each ensemble member, the difference between the hindcast multiyear ensemble mean and the ERA5 multiyear mean between 1981 and 2016 is subtracted.

The time series and distribution plots, including categorization, of the 10 indices can be found in Figs. S5–S7 and the EDJ bias correction in Fig. S4.

4. Probabilistic causal network of EDJ variability

In this section, the PCN derived from the hindcast data and its CPTs are presented.

a. Links

The causal network structure resulting from the procedure described in section 2a is shown in Fig. 2, where solid links are the retained links and dashed are those removed by conditional independence tests, based on the results reported in appendix B (chi-square and p values).
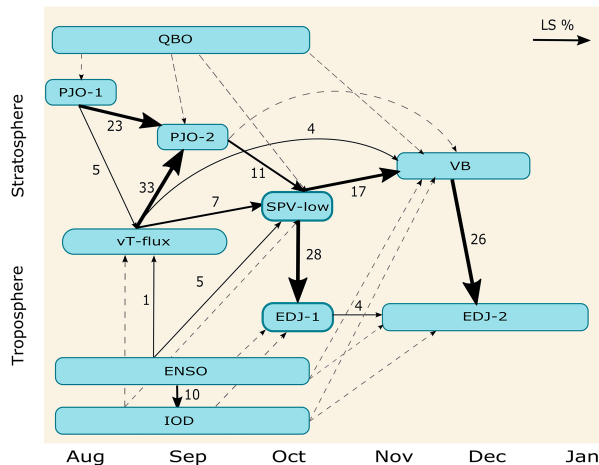
FIG. 2. The PCN variables and their connections, represented in a schematic time–height plane. The dashed gray links are the ones removed via the PC-stable algorithm, using a chi-squared test with a 0.05 significance level. The black solid links are the ones retained and together constitute the PCN structure used in the rest of the analysis. Links are labeled with the value of their link strength LS% defined in section 4c, and widths are proportional to it.

The links removed are all the ones emanating from QBO, PJO-2 → VB, ENSO → VB, ENSO → EDJ-1, ENSO → EDJ-2 and all the ones emanating from IOD.

All links from QBO are removed at the unconditional test stage, with all tests giving p values largely above the 0.05 threshold, except for the QBO → PJO-2 link which has a small $p = 0.051$ but still above the significance threshold. The absence of the known QBO teleconnection with the extratropical stratosphere in the network is likely due to modeling deficiencies, as even high-top models with a well-resolved stratosphere like System 4 tend to lose information about the amplitude and even the sign of tropical winds within 10–50 days and tend to model extratropical teleconnections rather weakly (Butler et al. 2016; Garfinkel et al. 2018; Lawrence et al. 2022). This is confirmed by an absence of the QBO fingerprint on extratropical zonal-mean zonal wind anomalies after September (Fig. S8).

All but one of the links emanating from IOD are removed at the unconditional test stage, and IOD → SPV-low is removed when ENSO is used as the condition, showing that the dependence between IOD and SPV-low is a result of their common driver ENSO. While it has been shown that IOD can influence the vortex and SSWs (Rao et al. 2020; Huang et al. 2021; Jucker and Reichler 2023), an independent contribution from that of ENSO is not detectable in this analysis.

PJO-2 → VB is removed after SPV-low conditioning, and ENSO → VB is removed after conditioning on both vT-flux and SPV-low, showing the role of SPV-low and vT-flux as mediators for those pathways.

ENSO → EDJ-1 is removed after conditioning on SPV-low, as is ENSO → EDJ-2 after accounting for the information given by VB, with however a much smaller $p = 0.06$ in the latter case compared with $p = 0.44$ in the former. This confirms

findings in Byrne et al. (2019) where no direct connections between ENSO and EDJ-2 were found after conditioning on VB, also in an analysis of ECMWF System 4.

Note that the link ENSO → IOD is included in the initial network, but not the link IOD → ENSO which, however plausible according to the literature, would have created a causal cycle. To test the effect of this choice on the network, the detection algorithm was repeated including IOD → ENSO (and excluding ENSO → IOD), which gave an identical network to Fig. 2 except for the link between ENSO and IOD. Given that the IOD is disconnected from the rest of the variables, the link between ENSO and IOD could be substituted with one without orientation and none of our conclusions would be affected.

Sensitivity analysis using a stricter significance level of 0.01 results in additionally removing the link ENSO → SPV-low, which however does not change the conclusions drawn in the following analysis. Sensitivity to the definition of ENSO is tested using the central Pacific Niño-4 index, as it has been found to better capture the SH stratospheric connections with El Niño events than the east Pacific (Domeisen et al. 2019). This results in identical links as when using the Niño-3.4 index, which straddles the east and central Pacific. Hence, the ENSO index choice does not affect the analysis.

b. Conditional probability tables

Inspection of the CPTs in Fig. 3, expressed as a ratio with their unconditional probability, quantifies the effect of combined states of the parents on each variable, in terms of making each state more or less likely than climatology. Increased (decreased) probability compared to climatology corresponds to values larger (smaller) than 1, further indicated with a blue (red) color scale. SPV-low is shown in Fig. S9, as the large number of entries due to three parents does not allow an intuitive visualization.

In general, the effect of one parent often strongly depends on the state of the other parent(s) (specially for target variables PJO-2, VB, and EDJ-2). A weak (strong) vT-flux is caused by a strong (weak) PJO-1, with generally little effect of ENSO. The combination of PJO-1 weak and El Niño is the one that increases the probability of a strong vT-flux the most: a warmer tropical Pacific SST increases the chances of a larger vT-flux, provided the subtropical stratospheric winds are weak enough (Fig. 3a). PJO-2 is very strongly influenced by combinations of PJO-1 and vT-flux, and none of the parents looks dominant (Fig. 3b). VB is strongly influenced by the strength of SPV-low, with small (large) SPV-low leading to earlier (later) VB. Independently, vT-flux acts on the vortex as well, with a strong vT-flux favoring earlier VB (and vice versa), and suggests that vT-flux anomalies may persist late in the season to affect VB (e.g., Hardiman et al. 2011; Hu et al. 2014; Lim et al. 2018); this is best seen when SPV-low is neutral (Fig. 3c). EDJ-1 is more poleward when a large lower stratospheric signal is seen in October (large SPV-low), and vice versa (Bracegirdle 2011) (Fig. 3d). Finally, an early (late) VB typically precedes an earlier (later) EDJ-2 equatorward shift in November and December (Black and McDaniel 2007; Byrne et al. 2017; Byrne and Shepherd 2018) (Fig. 3e). The
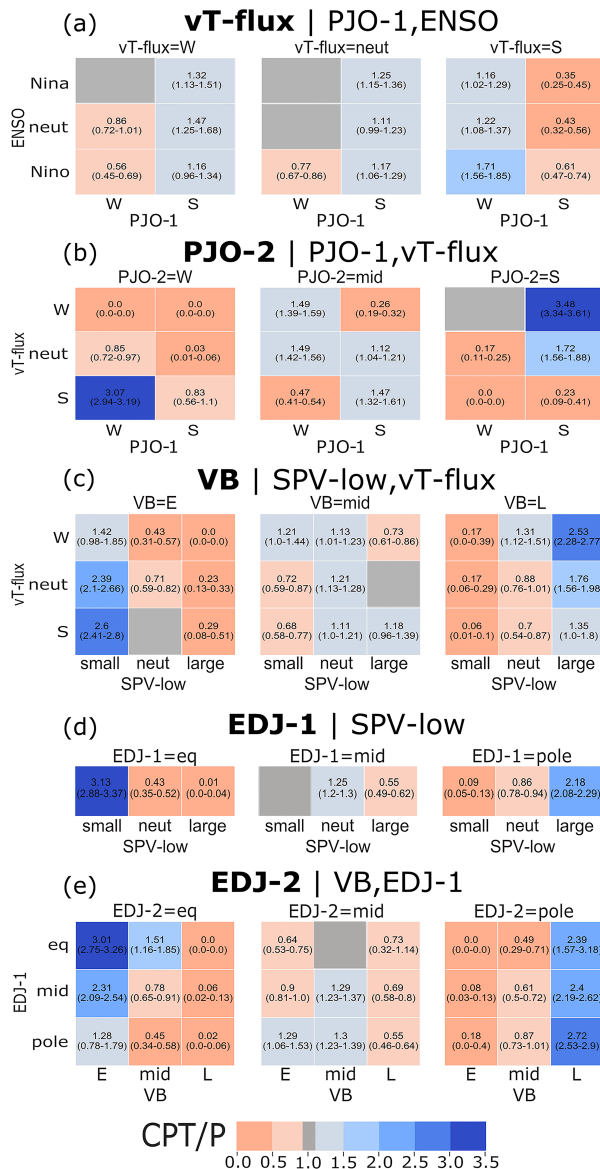
(a) vT-flux | PJO-1,ENSO

vT-flux=W

ENSO \ PJO-1	W	S
Nina		1.32 (1.13-1.51)
neut	0.86 (0.72-1.01)	1.47 (1.25-1.68)
Nino	0.56 (0.45-0.69)	1.16 (0.96-1.34)

vT-flux=neut

ENSO \ PJO-1	W	S
Nina		1.25 (1.15-1.36)
neut		1.11 (0.99-1.23)
Nino	0.77 (0.67-0.86)	1.17 (1.06-1.29)

vT-flux=S

ENSO \ PJO-1	W	S
Nina	1.16 (1.02-1.29)	0.35 (0.25-0.45)
neut	1.22 (1.08-1.37)	0.43 (0.32-0.56)
Nino	1.71 (1.56-1.85)	0.61 (0.47-0.74)

(b) PJO-2 | PJO-1,vT-flux

PJO-2=W

vT-flux \ PJO-1	W	S
W	0.0 (0.0-0.0)	0.0 (0.0-0.0)
neut	0.85 (0.72-0.97)	0.03 (0.01-0.06)
S	3.07 (2.94-3.19)	0.83 (0.56-1.1)

PJO-2=mid

vT-flux \ PJO-1	W	S
W	1.49 (1.39-1.59)	0.26 (0.19-0.32)
neut	1.49 (1.42-1.56)	1.12 (1.04-1.21)
S	0.47 (0.41-0.54)	1.47 (1.32-1.61)

PJO-2=S

vT-flux \ PJO-1	W	S
W		3.48 (3.34-3.61)
neut	0.17 (0.11-0.25)	1.72 (1.56-1.88)
S	0.0 (0.0-0.0)	0.23 (0.09-0.41)

(c) VB | SPV-low,vT-flux

VB=E

vT-flux \ SPV-low	small	neut	large
W	1.42 (0.98-1.85)	0.43 (0.31-0.57)	0.0 (0.0-0.0)
neut	2.39 (2.1-2.66)	0.71 (0.59-0.82)	0.23 (0.13-0.33)
S	2.6 (2.41-2.8)		0.29 (0.08-0.51)

VB=mid

vT-flux \ SPV-low	small	neut	large
W	1.21 (1.0-1.44)	1.13 (1.01-1.23)	0.73 (0.61-0.86)
neut	0.72 (0.59-0.87)	1.21 (1.13-1.28)	
S	0.68 (0.58-0.77)	1.11 (1.0-1.21)	1.18 (0.96-1.39)

VB=L

vT-flux \ SPV-low	small	neut	large
W	0.17 (0.0-0.39)	1.31 (1.12-1.51)	2.53 (2.28-2.77)
neut	0.17 (0.06-0.29)	0.88 (0.76-1.0)	1.76 (1.56-1.98)
S	0.06 (0.01-0.1)	0.7 (0.54-0.87)	1.35 (1.0-1.8)

(d) EDJ-1 | SPV-low

EDJ-1=eq

SPV-low	small	neut	large
	3.13 (2.88-3.37)	0.43 (0.35-0.52)	0.01 (0.0-0.04)

EDJ-1=mid

SPV-low	small	neut	large
		1.25 (1.2-1.3)	0.55 (0.49-0.62)

EDJ-1=pole

SPV-low	small	neut	large
	0.09 (0.05-0.13)	0.86 (0.78-0.94)	2.18 (2.08-2.29)

(e) EDJ-2 | VB,EDJ-1

EDJ-2=eq

EDJ-1 \ VB	E	mid	L
eq	3.01 (2.75-3.26)	1.51 (1.16-1.85)	0.0 (0.0-0.0)
mid	2.31 (2.09-2.54)	0.78 (0.65-0.91)	0.06 (0.02-0.13)
pole	1.28 (0.78-1.79)	0.45 (0.34-0.58)	0.02 (0.0-0.06)

EDJ-2=mid

EDJ-1 \ VB	E	mid	L
eq	0.64 (0.53-0.75)		0.73 (0.32-1.14)
mid	0.9 (0.81-1.0)	1.29 (1.23-1.37)	0.69 (0.58-0.8)
pole	1.29 (1.06-1.53)	1.3 (1.23-1.39)	0.55 (0.46-0.64)

EDJ-2=pole

EDJ-1 \ VB	E	mid	L
eq	0.0 (0.0-0.0)	0.49 (0.29-0.71)	2.39 (1.57-3.19)
mid	0.08 (0.03-0.13)	0.61 (0.5-0.72)	2.4 (2.19-2.62)
pole	0.18 (0.0-0.4)	0.87 (0.73-1.01)	2.72 (2.53-2.9)

CPT/P: 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5

FIG. 3. CPT values for each variable, expressed as the ratio with their unconditional probability *P*. Variables considered are (a) vT-flux, (b) PJO-2, (c) VB, (d) EDJ-1, and (e) EDJ-2. The color blue (red) indicates if the mean ratio is higher (lower) than 1, and gray values are masked if between 0.9 and 1.1. Each category of the variable is associated with one subtable, and each of the possible combinations of categories of its parents is associated with one column of each subtable (or column and row, if the parents are two). The values written on each table's entry show the mean and 5th–95th percentile range of a 1000 sample bootstrap obtained by resampling the ensemble members and years.

state of EDJ-1 also influences EDJ-2, although much less than VB does: EDJ-1 more equatorward (poleward) generally favors an earlier (later) shift of the EDJ-2.

Note that almost identical CPT values (difference below 5%) are found using the central Pacific Niño-4 index for ENSO.

c. Link strength

Links in a causal network are often associated with a measure of strength, used to compare their relative importance. For a PCN, measures of link strength are not straightforward because no specific functional dependence is assumed (Barnes et al. 2019; Harwood et al. 2021). An information-theoretic metric called blind average link strength is used here, based on the conditional mutual information between X and Y given Z where the link considered is $X \rightarrow Y$ and Z are the parents of Y other than X (Ebert-Uphoff 2007). Expressed as a percentage of conditional entropy, LS% is a measure of how close the link is to deterministic and therefore how influential the parent X alone is in determining the state of Y. The mathematical derivation is in appendix C.

The values LS% are used to label the links and scale their width in Fig. 2. The highest LS% values are for PJO-1 → PJO-2, vT-flux → PJO-2, SPV-low → EDJ-1, and VB → EDJ-2 (all around 20%–30%). Because both parents of PJO-2 have high LS%, this suggests that PJO-2's variability should be skillfully predicted by its parents. The two links pointing at vT-flux are among the weakest (<5%), and thus its predictability is expected to be very small. Links pointing at SPV-low and VB have medium strength (~5%–20%) with the ones emanating from vT-flux the weakest. Note however that link-specific metrics are built by marginalization of the effect of all the other parents; therefore, they may mask the importance of combined nonlinear interactions (see the higher predictability of VT-flux strong compared to vT-flux weak in Fig. 4). Despite this, the overall connection between LS% and predictability in this system is confirmed by the analysis in section 5.

5. EDJ predictability

The PCN derived in section 4 is used to quantify the long-range predictability of EDJ-1 and EDJ-2 shifts. After assessing how much hindcast variability is reproduced by the PCN parents (section 5a), the potential predictability of EDJ given the long-lead drivers is quantified by comparing PCN predictions of EDJ with its evolution in the hindcast (section 5b) and its real-world predictability assessed against reanalysis as compared with that of the hindcast (section 5c).

a. PCN parents' ability to reproduce the hindcast variability

First, we assess how much of each hindcast variable's variability can be reproduced by its PCN parents, estimated via cross-validated prediction skill. A good level of predictability given the parents means the PCN captures the dominant mechanisms of this system on the time scales simulated by the PCN. As the prediction skill of PCN predictions using the parents as evidence is by design higher than the one obtainable using the long-lead, indirect drivers as evidence, the skill values computed in this section provide an upper bound for the long-lead predictions of section 5b.

The ROC AUC and PR AUC% for predictions of each variable given its parents as evidence are quantified with a stratified 30-fold cross validation. Cross validation provides
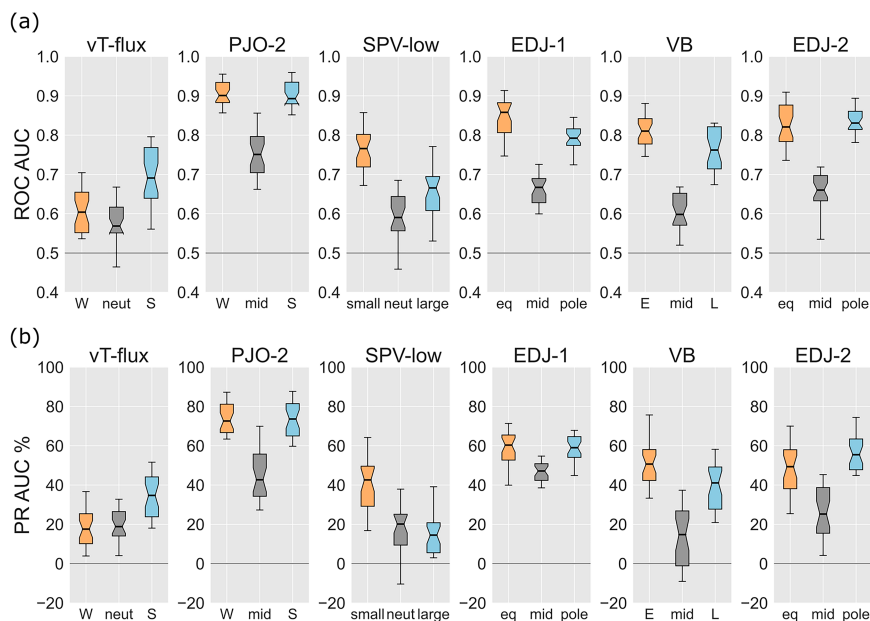
FIG. 4. Cross-validated (a) ROC AUC and (b) PR AUC% of the predictions of each variable vT-flux, PJO-2, SPV-low, EDJ-1, VB, and EDJ-2 (from left to right) using the respective parents as evidence. Thirty train–test sets of length 29/30 and 1/30 of all data points are considered, and a stratification approach to the selection of train–test sets is applied. The distributions of the 30 predictions are shown as boxplots, which cover the 25th–75th percentiles (boxes) and the 5th–95th percentiles (whiskers). Different colors highlight the different categories of each target variable.

an estimate of skill not inflated by overfitting, and stratification accounts for unbalanced categorical data by selecting the training and the test sets so that they have the same proportion of the target variable's categories (Fig. 4).

The following considerations apply to the extreme categories, but not to the central category which is discussed separately. A very high cross-validation skill for PJO-2, VB, EDJ-1, and EDJ-2 establishes that their PCN parents represent most of their variability (ROC AUC between 0.75 and 0.95 and PR AUC% between 50% and 75%). The SPV-low can only be partially predicted by its parents, pointing already at its limited long-range predictability given the variables used here (ROC AUC between 0.65 and 0.75 and PR AUC% between 25% and 45%; SPV-low large has the worst PR AUC% of all variables and much worse than its weak state). The overall least reproducible variable is vT-flux, which is only weakly influenced by PJO-1 and ENSO, confirming the indication provided by the link strengths. Despite different degrees of skill, it is notable that the few mechanisms modeled by the PCN still reproduce the hindcast variability significantly better than a no-skill climatological forecast for all variables, at least for the extreme categories (no skill means ROC AUC = 0.5 and PR AUC% = 0).

The central category is generally less skillfully predicted than the extreme ones. This is because the CPT values of the central category tend to have some similarities with the CPT values of one or the other extreme (low signal), making it harder for the PCN to discriminate between the central

category and another. The CPT values for the extreme categories are instead very different from one another (high signal); hence, they allow for an easier discrimination and thus better predictions. Given that the tails of the variables' distributions are usually the ones associated with stronger impacts, it is encouraging to see that the performance for the extreme categories is generally good.

b. Potential long-range EDJ predictability

After having found the upper bound for the predictability of each variable using the PCN parents, the degree to which this skill can be realized using indirect, long-lead drivers is quantified. Each of the 51 ensemble members in each year is used to provide evidence and verification data. Forecasts are performed for different evidence sets consisting of variables that can be jointly observed by the end of each month [excluding the ones that would be redundant given Eq. (1)]. They are ENSO and PJO-1 for August, ENSO, PJO-2, and vT-flux for September, vT-flux and SPV-low for October stratosphere only, and vT-flux, SPV-low, and EDJ-1 for October. The prediction given the parents is shown for comparison.

The results are shown in Fig. 5 for prediction of SPV-low (Figs. 5a,e), VB (Figs. 5b,f), EDJ-1 (Figs. 5c,g), and EDJ-2 (Figs. 5d,h). Given August and September evidence, the extreme categories are predicted with low skill but still better than climatology (median values between 0.6 and 0.7 for ROC and 20%–40% for PR). Recalling that VB and EDJ-2 are defined in the months of November and December, the
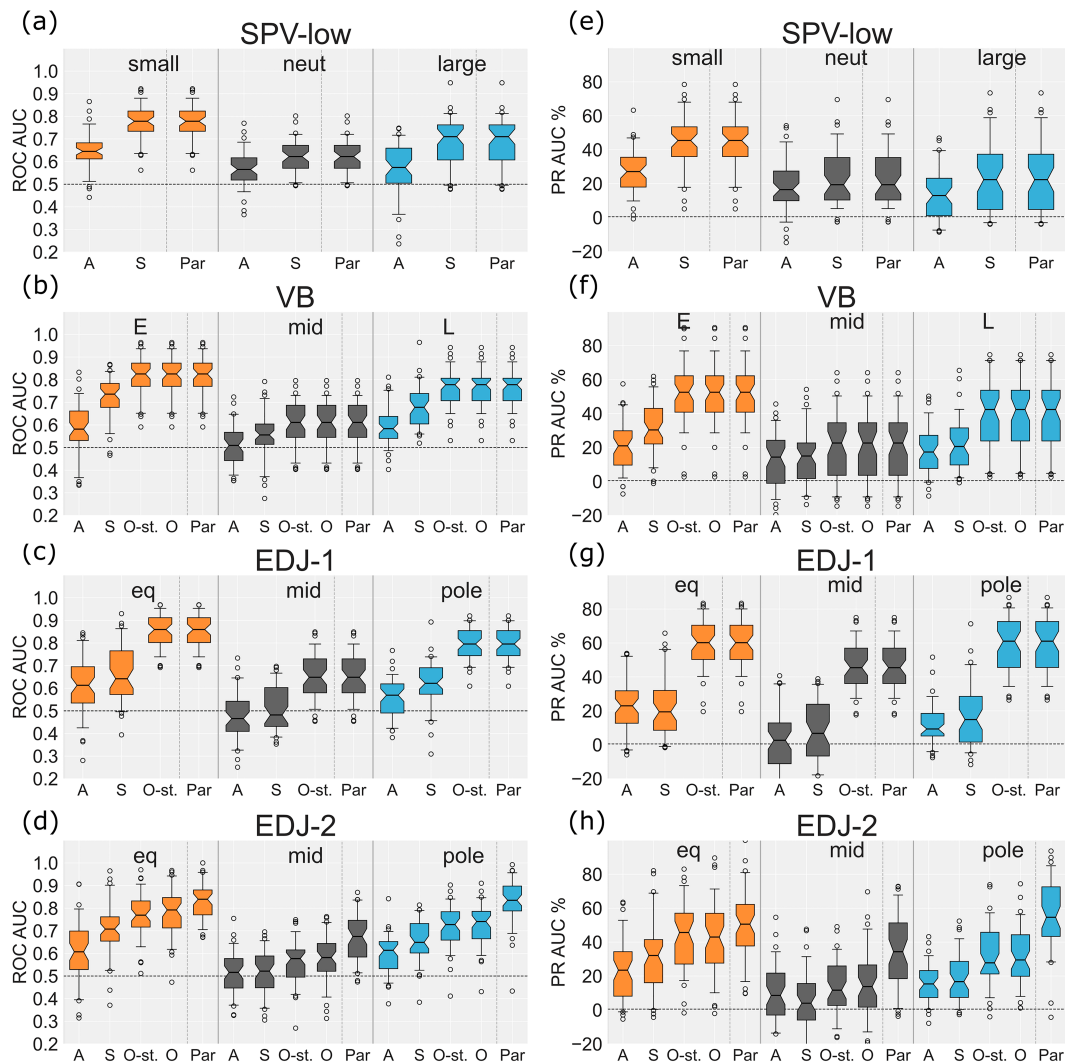
FIG. 5. (left) ROC AUC and (right) PR AUC (%) of (a),(e) SPV-low, (b),(f) VB, (c),(g) EDJ-1, and (d),(h) EDJ-2 predictions using as evidence variables observable in August (A), September (S), October-stratosphere (O-st.), October (O), and parents (Par) (specific variables listed in text). The network is trained on all hindcast data and tested on each of the 51 ensemble members separately (described in text). The boxplots represent the distributions of the 51 predictions, with box edges (whiskers) covering the 25th–75th (5th–95th) percentiles, and outliers also indicated. Different colors correspond to the three different categories of each target variable, also labeled at the top of each plot.

PCN can thus reproduce with some skill the variability of these with a 2–3-month lead.

Importantly, both ROC and PR metrics show a marked improvement when evidence is provided for months closer to the target, especially for 0–1-month lead, reaching median values between 0.7 and 0.9 for ROC and 40%–60% for PR. Thus, each intermediate linkage results in a loss of information, which builds up in time. EDJ-1 and EDJ-2 have the largest increase in skill given precursors in the October stratosphere, confirming the importance of the stratosphere for a skillful tropospheric forecast. Interestingly, EDJ-2 can be predicted knowing SPV-low (Oct-st.) almost as skillfully as when knowing its parents (VB and EDJ-1); thus, its predictability can be high with 1-month lead. Among the four

target variables analyzed in the figure, SPV-low has the lowest skill, which indicates a predictability bottleneck in our network from September variables. Considering vT-flux in September as a parent, rather than the August–September mean, does not improve SPV-low skill (see Fig. S10), further supporting this conclusion. To represent SPV-low, elements not present in the network are clearly needed. Given that, to our knowledge, all long-lead drivers of SPV-low variability have been included, these missing elements should be in midspring, developing on submonthly time scales. An analysis including vT-flux in October as an additional variable in the network improves SPV-low predictions (but not VB predictions), supporting this hypothesis (see Fig. S11).
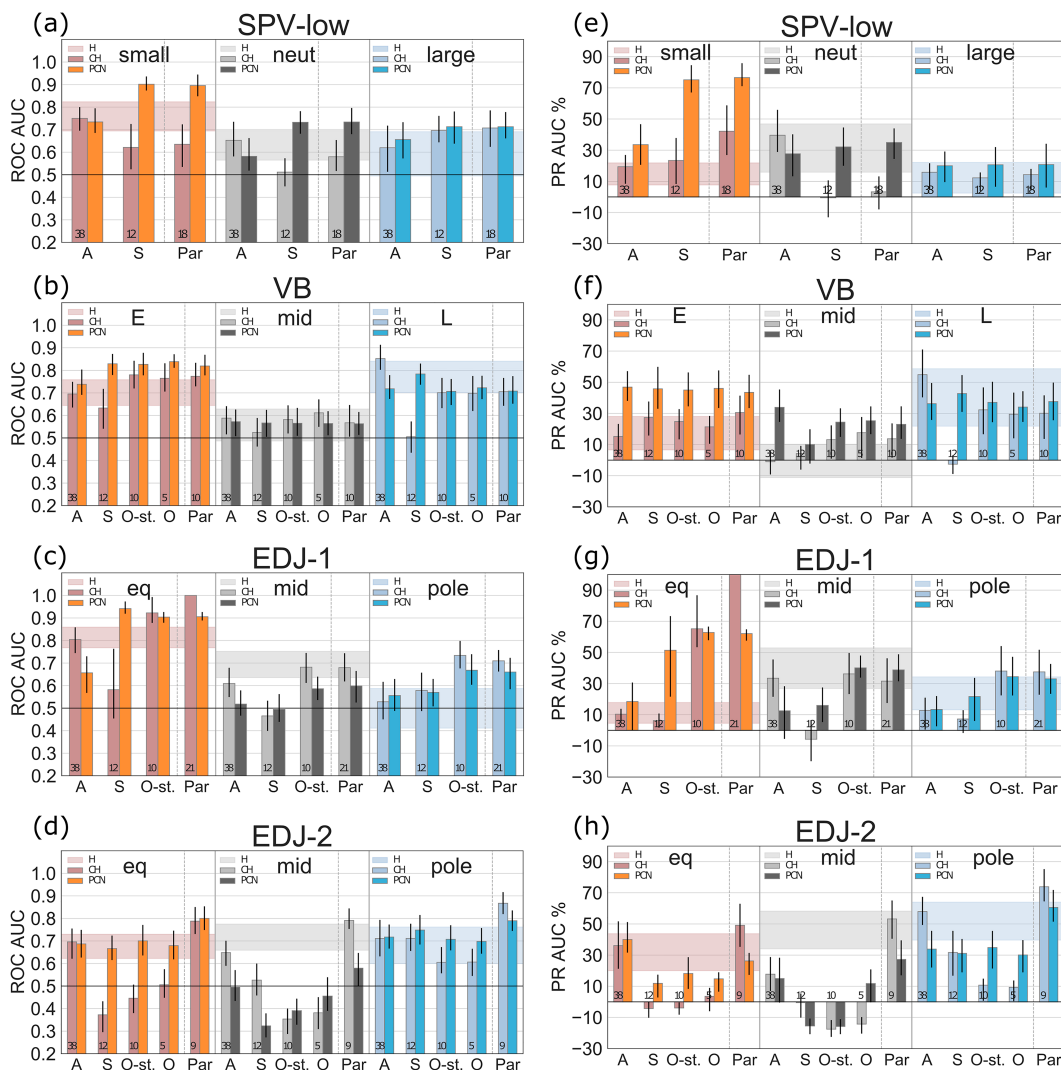
FIG. 6. ROC AUC and PR AUC% of (a),(e) SPV-low, (b),(f) VB, (c),(g) EDJ-1, and (d),(h) EDJ-2 predictions using as evidence variables observable in August (A), September (S), October-stratosphere (O-st.), October (O), and parents (Par) (as in Fig. 5). Three forecasts assessed against ERA5 are compared: the PCN forecast (PCN; rightmost vertical bar), the conditional hindcast (CH described in text; leftmost vertical bar), and the full hindcast (H; horizontal shaded bar). The height of the vertical CH and PCN bars (error bars) shows the mean value (25th–75th percentiles) of a distribution obtained from a 1000 sample bootstrap procedure. The numbers on the CH bar are the average number of members selected each year for a category. The horizontal H skill shading spans the 25th–75th percentiles across all years and members.

c. Relevance of PCN variables for real-world predictions

Long-range predictability of EDJ-1 and EDJ-2, although small, has been quantified and attributed to specific combinations of network variables in section 5b. Yet the system assessed was the hindcast ensemble, not the real world. To test if the variables and links used in the PCN correspond to mechanisms useful to predict the real world, a conditional hindcast (CH) is constructed and its skill assessed against ERA5 reanalysis.

For each of the five month-specific evidence sets used in the previous section, a CH is obtained as a subensemble whose members have the evidence variables' categories matching the categories observed in ERA5. Therefore, the CH has

the "correct" categories of those evidence variables deemed relevant for prediction of subsequent targets.

Further, a PCN prediction is also performed with corresponding evidence sets initialized with ERA5 categories for each year. This PCN prediction can be compared with the CH meaningfully, thanks to the filtering on the evidence variables. Note that while the network only uses the values for those evidence set variables as input, the CH has access to all the information of all other parts of the atmosphere–ocean system for the selected ensemble members.

The ROC AUC and PR AUC% for PCN and CH against ERA5 are shown in Fig. 6 (vertical bars), with the skill for the

full hindcast (H) also shown (horizontal shaded range). Note that the (conditional) hindcast skill score may be an overestimate of actual skill in a forecast setting due to the use of the entire hindcast period for the bias correction (Risbey et al. 2021). However, this does not affect the conclusions based on a relative comparison between the PCN, CH, and H skill.

The following comments apply to the extreme categories and not the central category. First, when parents are selected, the CH skill is generally better than that of the full hindcast. This suggests that the mechanisms schematized here provide meaningful rules to identify ensemble members that have a higher likelihood of evolving in time toward the correct ERA5 target category. For earlier conditions, however, the improvement of the CH is lower and at times nonexistent, showing that indirect variables in the network are not as effective as the parents in selecting more skillful members. Other sources of variability evidently come into play, steering the evolution of the system in other directions. Note also that the August filters for CH result in a selection of 38 out of 51 members on average across the target categories, because the ensemble spread is still small, and therefore a similar skill to the full hindcast is expected.

Second, PCN skill is generally equal to or higher than that of the CH. The PCN, with its simple representation of the connections via CPTs, is therefore as good as the CH at predicting the observed probabilistic evolution of the ERA5 target variables. When PCN outperforms the CH, this can be an indication that the background information used to quantify the conditional probabilities reduces the noise that arises in finite ensembles.

Last, note that the EDJ-2 in the CH forecast shows a peculiar feature. Its skill is high given August conditions, then deteriorates given September to October conditions, and improves drastically again given the parents. This suggests that the ensemble members filtered based on the September (ENSO, PJO-2, and vT-flux) and October conditions (vT-flux, SPV-low, and EDJ-1) do not tend to evolve toward a consistent and correct EDJ-2 anomaly, and therefore some of them constitute bad deterministic filters. Namely, given that the August conditions and parents are good, the bad filters are vT-flux, PJO-2, and possibly also SPV-low. On the other hand, members filtered based on the parents VB and EDJ-1 improve the skill.

6. Effect of improved EDJ prediction on surface climate

Finally, the effect of predicting skillful EDJ on SH surface climate is inspected. Gridpoint anomaly correlation coefficients (AC) between the hindcast and ERA5 across the modeled years are computed for Z500, 2mT, and MTPR fields taken as November–December mean. Analysis for September–October AC given EDJ-1 gives very similar results (see Fig. S12).

The AC between the full hindcast ensemble mean and ERA5 shows no or low skill in the mid-to-high latitudes (leftmost column in Fig. 7). However, a considerable AC improvement is achieved with a correct representation of EDJ-2 variability (central column), computed using only the ensemble

members that, each year, predict the correct EDJ-2 category (i.e., as in ERA5). The increase in AC is seen specifically over much of Antarctica, southern South America, and the Southern Ocean south of Australia, including in regions where the ensemble mean had no skill at all. These regions are the ones just north and south of the climatological position of the EDJ. Notably, a good part of this AC improvement can be achieved by using knowledge of the EDJ-2 parents only (rightmost column). The ensemble members used for this latter AC computation are selected with a statistical postprocessing, similar to recent methods to select the most skillful members of S2S ensemble forecasts (Dobrynin et al. 2018; Bouchet et al. 2019; Polkova et al. 2021). To be specific, ensemble members are selected each year based on the EDJ-2 category predicted by the PCN, where the PCN uses the parents as evidence by setting them to the ERA5 categories for that year. More details on the computation are given in the figure caption.

7. Discussion

The main goal of this analysis was to quantify the long-range predictability, meaning one month or more in the future, of the SH EDJ across the spring to summer transition given the knowledge of long-lead drivers regarded as important in the literature, namely, QBO, ENSO, IOD, vT-flux, and PJO in August and September, and establishing the importance of the mediating role of SPV-low and VB between October and December.

Assuming the hindcast represents the true system, ENSO, PJO, and vT-flux generally provide nonnegligible but low predictability to the state of EDJ-1 and EDJ-2, with the median ROC AUC of 0.6–0.75 and PR AUC% of 10%–30%. This shows that known correlations between a target and a driver do not necessarily correspond to a strong added prediction skill: either because other drivers can counteract that effect or because while detectable the effect is itself not very strong.

SPV-low, a crucial mediator for the effect of long-lead drivers on EDJ, cannot be skillfully predicted given preconditions such as ENSO, vT-flux, or PJO, even when tested against hindcast data on which the network was trained. This represents a long-range predictability bottleneck for the jet. The limited predictability of SPV-low given its parents suggests that important mechanisms that lead to the destabilization or persistence of the springtime vortex come from additional variables not represented in the PCN. Considering vT-flux in October significantly improves SPV-low predictions (but not VB predictions), suggesting the role of tropospheric forcing on submonthly time scales in midspring. Weekly average variables between the end of September and early October may be needed to improve the network representation of SPV-low and thus also of VB. For example, Domeisen et al. (2020a) found that while around 30% of ensemble members in S2S forecasts predict the final warming date within ±3 days for a 30–20-day lead time, the number increases to 75%–100% for a 15–5-day lead time. Although the allowed error of ±3 days is a much higher accuracy than required by the PCN forecast (where VB categories are about 20 days wide), this is still significant in showing the importance of shorter time scales.
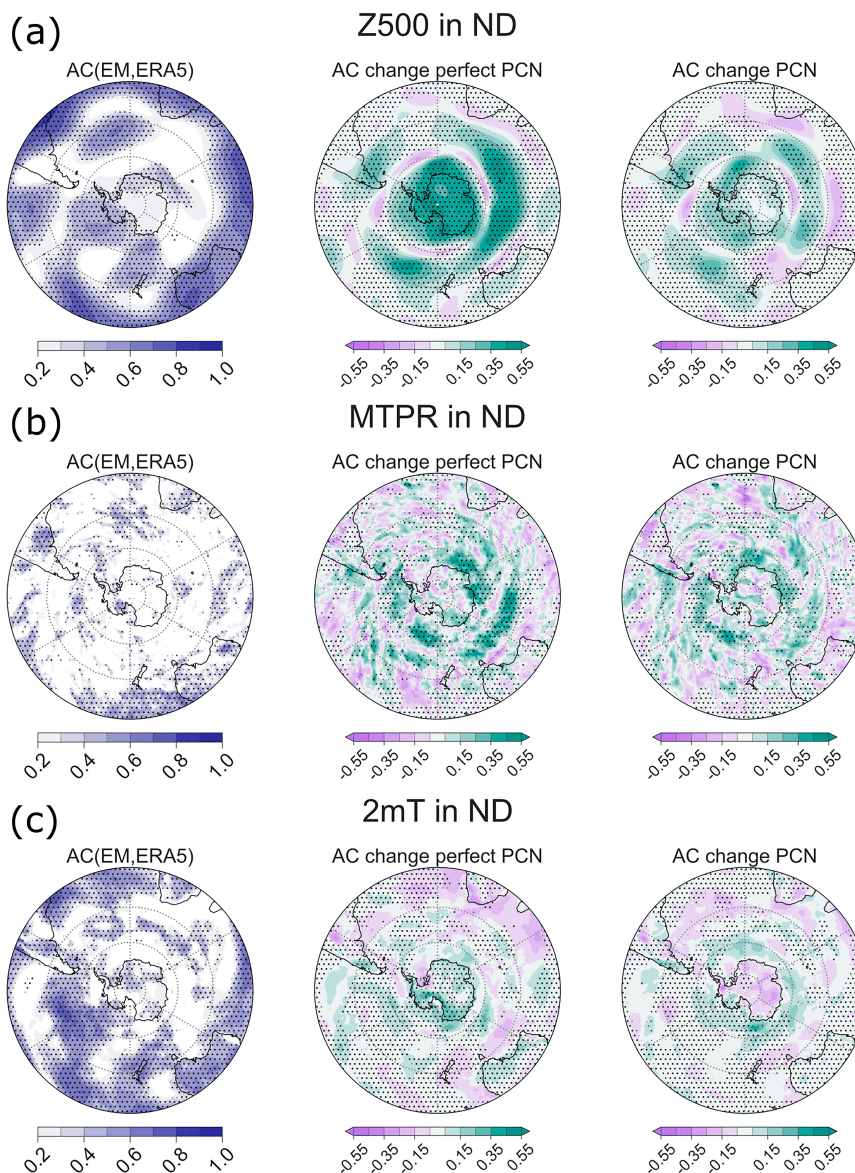
FIG. 7. November–December mean (ND) (a) Z500, (b) MTPR, and (c) 2mT anomaly correlation (AC) maps. AC between hindcast ensemble mean and ERA-5 is in the leftmost subpanels (darker blue for higher AC). Changes in AC for a selection of members based on having EDJ-2 categories matching ERA5 are in the middle subpanels (purple and green colors for negative or positive changes). Changes in AC for a selection of members based on having EDJ-2 categories matching the predictions of the PCN given the parents set to ERA5 categories each year are in the rightmost subpanels (colors as for the middle). The PCN probabilistic forecast is translated into a single predicted category by selecting the one with the largest relative increase in probability with respect to its climatological occurrence. Black dots show AC significantly different from 0 according to a one-sample two-sided t test at the 0.1 significance level. A stereographic projection is used, and latitudes between 20° and 90°S are shown at intervals of 20°.

However, introducing variables defined on shorter time averages would change the purpose of the network, which is of quantifying the possibility of long-range EDJ predictability.

Note that the PCN long-range prediction of SPV-low and VB against ERA5 could be further improved if the connections from the QBO and from late-spring stratospheric ozone were included.

This has not been possible due to their insufficient representation in System 4 (as in many other forecast systems) but could be achieved with another dataset used as a basis for the PCN detection and quantification (Oh et al. 2022; Monge-Sanz et al. 2022).

A strong predictability of the poleward migration of the jet (EDJ-1) is found to be associated with the strength of SPV-low,

and the timing of VB is a strong predictor of the timing of the jet's subsequent equatorward migration (EDJ-2), which agrees with what was initially proposed by Bracegirdle (2011). The EDJ-1 and EDJ-2 PCN skill against the hindcast given SPV-low has median ROC AUC of 0.8–0.9 and PR AUC% of 50%–60%.

Tested against ERA5, the PCN skill is found to be comparable to that of a conditional hindcast filtered on parents used for PCN predictions. This shows that the few linkages modeled by the PCN are a good source of information that applies to the real world too, although for some variables they provide more information than for others. The PCN is therefore a useful tool to quantify predictability and to attribute it to specific causal pathways.

Finally, an improved representation of EDJ location is found to improve the anomaly correlation of the hindcast with surface climate variables, Z500, 2mT, and MTPR. The improvement is especially strong over Antarctica and over the Southern Ocean, in regions just north and south of the climatological EDJ location, and part of this improvement is attributable to stratospheric information.

The PCN was built to assess the importance of well-known long-lead drivers of EDJ, but if the purpose is that of maximizing prediction skill, then the selection of variables could be optimized further (with care of avoiding overfitting, e.g., with cross validation), with the benefit that all the parts of the network that remain unmodified do not need recalibrating.

The zonally asymmetric components of the EDJ have not been addressed in this analysis; however, their analysis would be important to connect the stratospheric variability with more localized impacts (e.g., Osman et al. 2022). This would require detailed regional analyses for different longitude sectors and associated drivers and an inspection of the modulation of the zonal-mean response in each sector. Such analysis goes beyond the scope of the current study but would be an interesting avenue to pursue.

8. Conclusions

A probabilistic causal network (PCN) has been used to quantify the long-range predictability of the spring-to-summer variability of the SH eddy-driven jet (EDJ) migrations given known long-lead predictors, and their intermediate, mediating pathways via the stratospheric polar vortex. The ECMWF System 4 seasonal forecast system was used as dataset to build and quantify the PCN. Despite its simplicity, this network can reproduce a significant part of the variability of the system. Network-based predictions confirm the stratospheric polar vortex as determinant for skillful jet predictions, for both its poleward shift in late spring and its equatorward shift in early summer. The long-lead drivers ENSO, IOD, late-winter wave activity flux, and PJO are found to only provide moderate prediction skill to the vortex and hence to the jet. It is argued that the reason is the role of the submonthly variability in early spring, not represented in the network, creating a bottleneck for the long-range predictability of the jet.

The analysis has also shown that using simple statistical forecast models, like a PCN, offers a still largely unexplored approach to study the predictability provided by multiple, interacting drivers on S2S time scales. The evolution of the systems relevant at these time scales lends itself to be described statistically with multiple variables interacting in probabilistic terms. The present work demonstrates the following practical benefits of a PCN approach:

- Causal network theory provides a rigorous set of rules that allow us to distinguish direct from indirect effects, which can reveal the most effective pathways of enhanced predictability.
- The probabilistic framework represents uncertainty as well as nonlinearities in a manageable form, especially if variables are discretized.
- The added predictive power of various combinations of remote drivers can be quantified by means of fast PCN forecasts that make use only of the desired inputs.

These networks can also identify "trigger points" for prediction of particular outcomes, i.e., what set of prior events one needs to observe to be confident (at some given level) of an outcome. Moreover, the visual graphical representation of the model enhances transparency and could help communicate sources of predictability to nonspecialist audiences.

Although beyond the scope of this work, it is worth noting that probabilistic causal networks could be used as simple and quick-to-run models for probabilistic S2S predictions of specific target variables. The statistical model used here could be potentially enhanced using machine learning techniques, such as for detection of additional (and possibly unknown) sources of predictability from a large number of spatiotemporal fields. A PCN can be easily extended to include more drivers (e.g., climate change by adding forcing variables) and can be coupled with other models (e.g., impact models). They could be the basis for "intermediate technologies," i.e., simple but not elementary statistical models, which can bring together physical and statistical understanding in a way that can create intelligibility and also democratize the production of climate information (Rodrigues and Shepherd 2022).

APPENDIX A

Abbreviations and Acronyms

2mT	Two-meter temperature
CH	Conditional hindcast
CPT	Conditional probability table
EDJ	Zonal-mean eddy-driven jet
H	Hindcast
LS	Link strength
MTPR	Mean total precipitation rate
PCN	Probabilistic causal network
PJO	Polar night jet oscillation
PR AUC	Area under the precision–recall curve

ROC AUC	Area under the receiver operating characteristic curve
SPV	Stratospheric polar vortex
VB	Vortex breakdown
vT-flux	Eddy heat flux

APPENDIX B

Conditional Independence Tests

Table B1 shows the results of each iteration of the PC-stable conditional independence tests carried out to reveal the final probabilistic causal network.

TABLE B1. Results of each iteration step of the PC-stable conditional independence tests on the link $X \to Y$ via chi-square $(X, Y|Z)$ at significance level $\alpha = 0.05$; Z is a subset of the parents of the target variable (other than X), which is of increasing size at each iteration of PC-stable. Bold entries highlight the links removed via the PC-stable procedure.

X	Y	Z (iteration)	Chi-square	p value	Remove link
QBO	PJO-1	None (0)	1.6	0.20	**Yes**
QBO	vT	None (0)	3.6	0.16	**Yes**
QBO	PJO-2	None (0)	5.9	0.051	**Yes**
QBO	SPV-low	None (0)	1.7	0.42	**Yes**
QBO	VB	None (0)	1.1	0.59	**Yes**
IOD	vT	None (0)	7.1	0.13	**Yes**
IOD	VB	None (0)	7.7	0.10	**Yes**
IOD	EDJ-1	None (0)	6.4	0.17	**Yes**
IOD	EDJ-2	VB (1)	19.7	0.07	**Yes**
IOD	SPV-low	ENSO (1)	13.6	0.32	**Yes**
ENSO	IOD	None (0)	358.7	10^{-9}	No
ENSO	SPV-low	vT, PJO-2 (2)	55	0.02	No
ENSO	vT	PJO-1 (1)	34	10^{-5}	No
ENSO	EDJ-1	SPV-low (1)	11	0.44	**Yes**
ENSO	EDJ-2	VB (1)	20.2	0.06	**Yes**
ENSO	VB	vT, SPV-low (2)	47.7	0.09	**Yes**
vT	PJO-2	PJO-1 (1)	851	10^{-9}	No
vT	SPV-low	PJO-2, ENSO (2)	105	10^{-9}	No
vT	VB	SPV-low (1)	74	10^{-9}	No
PJO-1	vT	ENSO (1)	160	10^{-9}	No
PJO-1	PJO-2	vT (1)	529	10^{-9}	No
PJO-2	SPV-low	vT, ENSO (2)	131	10^{-9}	No
PJO-2	VB	vT, SPV-low (2)	26	0.88	**Yes**
SPV-low	VB	vT (1)	426	10^{-9}	No
SPV-low	EDJ-1	None (0)	847	10^{-9}	No
VB	EDJ-2	EDJ-1 (1)	655	10^{-9}	No
EDJ-1	EDJ-2	VB (1)	65	10^{-5}	No

APPENDIX C

Link Strength Formula

In section 4c, Ebert-Uphoff (2007) defines blind average link strength of link $X \rightarrow Y$ as the conditional mutual information of the pair (X, Y) conditioned on Z, where Z is defined as the set of all parents of Y other than X, and under the "blind" assumption that that X and Z are independent and uniformly distributed $P(x, z) = P(x)P(z)$, $P(x) = 1/N_X$, and $P(z) = 1/N_Z$:

$$\text{LS}(X \rightarrow Y) = \frac{1}{N_X N_Z} \sum_{x,y,z} P(y|x, z)\log_2\left[\frac{P(y|x, z)}{1/N_X \sum_x P(y|x, z)}\right],$$
(C1)

where $P(y|x, z)$ is the CPT of Y. The blind average link strength is then expressed as percentage of uncertainty reduction, following the formula:

$$\text{LS\%}(X \rightarrow Y) = \frac{\text{LS}(X \rightarrow Y)}{\hat{U}(Y|Z)} \times 100,$$
(C2)

where the conditional entropy computed also under the blind assumption is $\hat{U}(Y|Z) = [1/(N_X N_Z)]\sum_{x,y,z} P(y|x, z)\log_2[N_X/\sum_x P(y|x, z)]$. LS% is the value used for the link strengths of section 4c.

REFERENCES

Annamalai, H., S. P. Xie, J. P. McCreary, and R. Murtugudde, 2005: Impact of Indian Ocean sea surface temperature on developing El Niño. *J. Climate*, **18**, 302–319, https://doi.org/10.1175/JCLI-3268.1.

Anstey, J. A., and T. G. Shepherd, 2014: High-latitude influence of the quasi-biennial oscillation. *Quart. J. Roy. Meteor. Soc.*, **140**, 1–21, https://doi.org/10.1002/qj.2132.

Baldwin, M. P., and T. J. Dunkerton, 1988: Quasi-biennial modulation of the Southern Hemisphere stratospheric polar vortex. *Geophys. Res. Lett.*, **25**, 3343–3346, https://doi.org/10.1029/98GL02445.

Barnes, E. A., S. M. Samarasinghe, I. Ebert-Uphoff, and J. C. Furtado, 2019: Tropospheric and stratospheric causal pathways between the MJO and NAO. *J. Geophys. Res. Atmos.*, **124**, 9356–9371, https://doi.org/10.1029/2019JD031024.

Black, R. X., and B. A. McDaniel, 2007: Interannual variability in the Southern Hemisphere circulation organized by stratospheric final warming events. *J. Atmos. Sci.*, **64**, 2968–2974, https://doi.org/10.1175/JAS3979.1.

Bouchet, F., J. Rolland, and J. Wouters, 2019: Rare event sampling methods. *Chaos*, **29**, 080402, https://doi.org/10.1063/1.5120509.

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 Dyn.*, **37**, 2323–2333, https://doi.org/10.1007/s00382-011-1014-4.

Butler, A. H., and Coauthors, 2016: The climate-system historical forecast project: Do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Quart. J. Roy. Meteor. Soc.*, **142**, 1413–1427, https://doi.org/10.1002/qj.2743.

Byrne, N. J., 2017: Deterministic models of Southern Hemisphere circulation variability. Ph.D. thesis, University of Reading, 105 pp.

——, and T. G. Shepherd, 2018: Seasonal persistence of circulation anomalies in the Southern Hemisphere stratosphere and its implications for the troposphere. *J. Climate*, **31**, 3467–3483, https://doi.org/10.1175/JCLI-D-17-0557.1.

——, ——, T. Woollings, and R. A. Plumb, 2017: Nonstationarity in Southern Hemisphere climate variability associated with the seasonal breakdown of the stratospheric polar vortex. *J. Climate*, **30**, 7125–7139, https://doi.org/10.1175/JCLI-D-17-0097.1.

——, ——, and I. Polichtchouk, 2019: Subseasonal-to-seasonal predictability of the Southern Hemisphere eddy-driven jet during austral spring and early summer. *J. Geophys. Res. Atmos.*, **124**, 6841–6855, https://doi.org/10.1029/2018JD030173.

Ceppi, P., and T. G. Shepherd, 2019: The role of the stratospheric polar vortex for the austral jet response to greenhouse gas forcing. *Geophys. Res. Lett.*, **46**, 6972–6979, https://doi.org/10.1029/2019GL082883.

Chen, S. H., and C. A. Pollino, 2012: Good practice in Bayesian network modelling. *Environ. Modell. Software*, **37**, 134–145, https://doi.org/10.1016/j.envsoft.2012.03.012.

Colombo, D., and M. H. Maathuis, 2014: Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.*, **15**, 3921–3962.

Dobrynin, M., and Coauthors, 2018: Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophys. Res. Lett.*, **45**, 3605–3614, https://doi.org/10.1002/2018GL077209.

Domeisen, D. I. V., C. I. Garfinkel, and A. H. Butler, 2019: The teleconnection of El Niño Southern Oscillation to the stratosphere. *Rev. Geophys.*, **57**, 5–47, https://doi.org/10.1029/2018RG000596.

——, and Coauthors, 2020a: The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *J. Geophys. Res. Atmos.*, **125**, e2019JD030920, https://doi.org/10.1029/2019JD030920.

——, and Coauthors, 2020b: The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *J. Geophys. Res. Atmos.*, **125**, e2019JD030923, https://doi.org/10.1029/2019JD030923.

Ebert-Uphoff, I., 2007: Measuring connection strengths and link strengths in discrete Bayesian networks. Georgia Tech. Research Rep. GT-IIC-07-01, 10 pp.

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. *J. Geophys. Res. Atmos.*, **123**, 7855–7866, https://doi.org/10.1029/2018JD028724.

Hardiman, S. C., and Coauthors, 2011: Improved predictability of the troposphere using stratospheric final warmings. *J. Geophys. Res.*, **116**, D18113, https://doi.org/10.1029/2011JD015914.

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. *J. Climate*, **34**, 2319–2335, https://doi.org/10.1175/JCLI-D-20-0369.1.

Hendon, H. H., E.-P. Lim, and S. Abhik, 2020: Impact of interannual ozone variations on the downward coupling of the 2002 Southern Hemisphere stratospheric warming. *J. Geophys. Res. Atmos.*, **125**, e2020JD032952, https://doi.org/10.1029/2020JD032952.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hu, J., R. Ren, and H. Xu, 2014: Occurrence of winter stratospheric sudden warming events and the seasonal timing of spring stratospheric final warming. *J. Atmos. Sci.*, **71**, 2319–2334, https://doi.org/10.1175/JAS-D-13-0349.1.

Huang, R., W. Tian, K. Qie, F. Xie, S. Zhang, H. Tian, and J. Luo, 2021: Contrasting effects of Indian Ocean basin and dipole modes on the stratosphere. *J. Geophys. Res. Atmos.*, **126**, e2021JD035156, https://doi.org/10.1029/2021JD035156.

Jucker, M., and T. Reichler, 2023: Life cycle of major sudden stratospheric warmings in the Southern Hemisphere from a multimillennial GCM simulation. *J. Climate*, **36**, 643–661, https://doi.org/10.1175/JCLI-D-22-0425.1.

Kidson, J. W., 1999: Principal modes of Southern Hemisphere low-frequency variability obtained from NCEP–NCAR reanalyses. *J. Climate*, **12**, 2808–2830, https://doi.org/10.1175/1520-0442(1999)012<2808:PMOSHL>2.0.CO;2.

Kuroda, Y., 2002: Relationship between the polar-night jet oscillation and the annular mode. *Geophys. Res. Lett.*, **29**, 1240, https://doi.org/10.1029/2001GL013933.

——, and K. Kodera, 1998: Interannual variability in the troposphere and stratosphere of the Southern Hemisphere winter. *J. Geophys. Res.*, **103**, 13 787–13 799, https://doi.org/10.1029/98JD01042.

——, and ——, 2001: Variability of the polar night jet in the Northern and Southern Hemispheres. *J. Geophys. Res.*, **106**, 20 703–20 713, https://doi.org/10.1029/2001JD900226.

Lawrence, Z. D., and Coauthors, 2022: Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems. *Wea. Climate Dyn.*, **3**, 977–1001, https://doi.org/10.5194/wcd-3-977-2022.

L'Heureux, M. L., and D. W. J. Thompson, 2006: Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. *J. Climate*, **19**, 276–287, https://doi.org/10.1175/JCLI3617.1.

Lim, E.-P., H. H. Hendon, and H. Rashid, 2013: Seasonal predictability of the Southern Annular Mode due to its association with ENSO. *J. Climate*, **26**, 8037–8054, https://doi.org/10.1175/JCLI-D-13-00006.1.

——, ——, and D. W. J. Thompson, 2018: Seasonal evolution of stratosphere-troposphere coupling in the Southern Hemisphere and implications for the predictability of surface climate. *J. Geophys. Res. Atmos.*, **123**, 12 002–12 016, https://doi.org/10.1029/2018JD029321.

——, ——, G. Boschat, D. Hudson, D. W. J. Thompson, A. J. Dowdy, and J. M. Arblaster, 2019: Australian hot and dry extremes induced by weakenings of the stratospheric polar vortex. *Nat. Geosci.*, **12**, 896–901, https://doi.org/10.1038/s41561-019-0456-x.

Lorenz, D. J., and D. L. Hartmann, 2001: Eddy–zonal flow feedback in the Southern Hemisphere. *J. Atmos. Sci.*, **58**, 3312–3327, https://doi.org/10.1175/1520-0469(2001)058<3312:EZFFIT>2.0.CO;2.

Luo, J.-J., R. Zhang, S. K. Behera, Y. Masumoto, F.-F. Jin, R. Lukas, and T. Yamagata, 2010: Interaction between El Niño and extreme Indian Ocean dipole. *J. Climate*, **23**, 726–742, https://doi.org/10.1175/2009JCLI3104.1.

Marcot, B. G., J. D. Steventon, G. D. Sutherland, and R. K. McCann, 2006: Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Can. J. For. Res.*, **36**, 3063–3074, https://doi.org/10.1139/x06-135.

McLandress, C., A. I. Jonsson, D. A. Plummer, M. C. Reader, J. F. Scinocca, and T. G. Shepherd, 2010: Separating the dynamical effects of climate change and ozone depletion. Part I: Southern Hemisphere stratosphere. *J. Climate*, **23**, 5002–5020, https://doi.org/10.1175/2010JCLI3586.1.

——, T. G. Shepherd, J. F. Scinocca, D. A. Plummer, M. Sigmond, A. I. Jonsson, and M. C. Reader, 2011: Separating the dynamical effects of climate change and ozone depletion. Part II: Southern Hemisphere troposphere. *J. Climate*, **24**, 1850–1868, https://doi.org/10.1175/2010JCLI3958.1.

Mo, K. C., and J. N. Paegle, 2001: The Pacific–South American modes and their downstream effects. *Int. J. Climatol.*, **21**, 1211–1229, https://doi.org/10.1002/joc.685.

Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (System 4). ECMWF Tech. Rep. 656, 51 pp.

Monge-Sanz, B. M., and Coauthors, 2022: A stratospheric prognostic ozone for seamless Earth system models: Performance, impacts and future. *Atmos. Chem. Phys.*, **22**, 4277–4302, https://doi.org/10.5194/acp-22-4277-2022.

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. *Prog. Earth Planet. Sci.*, **9**, 25, https://doi.org/10.1186/s40645-022-00485-4.

Osman, M., C. S. Vera, and F. J. Doblas-Reyes, 2015: Predictability of the tropospheric circulation in the Southern Hemisphere from CHFP models. *Climate Dyn.*, **46**, 2423–2434, https://doi.org/10.1007/s00382-015-2710-2.

——, T. G. Shepherd, and C. S. 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 Dyn.*, **59**, 2949–2964, https://doi.org/10.1007/s00382-022-06225-0.

Pearl, J., 2009: *Causality: Models, Reasoning, and Inference.* 2nd ed. Cambridge University Press, 464 pp.

Polkova, I., and Coauthors, 2021: Predictors and prediction skill for marine cold-air outbreaks over the Barents Sea. *Quart. J. Roy. Meteor. Soc.*, **147**, 2638–2656, https://doi.org/10.1002/qj.4038.

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. *J. Geophys. Res. Atmos.*, **125**, e2020JD032723, https://doi.org/10.1029/2020JD032723.

——, ——, and ——, 2021: Development of the extratropical response to the stratospheric quasi-biennial oscillation. *J. Climate*, **34**, 7239–7255, https://doi.org/10.1175/JCLI-D-20-0960.1.

Risbey, J. S., and Coauthors, 2021: Standard assessments of climate forecast skill can be misleading. *Nat. Commun.*, **12**, 4346, https://doi.org/10.1038/s41467-021-23771-z.

Rodrigues, R. R., and T. G. Shepherd, 2022: Small is beautiful: Climate-change science as if people mattered. *PNAS Nexus*, **1**, pgac009, https://doi.org/10.1093/pnasnexus/pgac009.

Roff, G., D. W. J. Thompson, and H. Hendon, 2011: Does increasing model stratospheric resolution improve extended-range forecast skill? *Geophys. Res. Lett.*, **38**, L05809, https://doi.org/10.1029/2010GL046515.

Runge, J., P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, 2019: Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.*, **5**, eaau4996, https://doi.org/10.1126/sciadv.aau4996.

Saggioro, E., 2023: Causal network approaches for the study of sub-seasonal to seasonal variability and predictability. Ph.D. thesis, University of Reading, 250 pp.

——, and T. G. Shepherd, 2019: Quantifying the timescale and strength of Southern Hemisphere intraseasonal stratosphere-troposphere coupling. *Geophys. Res. Lett.*, **46**, 13 479–13 487, https://doi.org/10.1029/2019GL084763.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363, https://doi.org/10.1038/43854.

Seager, R., N. Harnik, Y. Kushnir, W. Robinson, and J. Miller, 2003: Mechanisms of hemispherically symmetric climate variability. *J. Climate*, **16**, 2960–2978, https://doi.org/10.1175/1520-0442(2003)016<2960:MOHSCV>2.0.CO;2.

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. *J. Climate*, **27**, 7462–7474, https://doi.org/10.1175/JCLI-D-14-00264.1.

Shepherd, T. G., I. Polichtchouk, R. Hogan, and A. J. Simmons, 2018: Report on stratosphere task force. ECMWF Tech. Rep. 824, 34 pp.

Son, S.-W., A. Purich, H. H. Hendon, B.-M. Kim, and L. M. Polvani, 2013: Improved seasonal forecast using ozone hole variability? *Geophys. Res. Lett.*, **40**, 6231–6235, https://doi.org/10.1002/2013GL057731.

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. *J. Climate*, **35**, 1963–1981, https://doi.org/10.1175/JCLI-D-21-0250.1.

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J.*

Climate, **13**, 1000–1016, https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.

——, M. P. Baldwin, and S. Solomon, 2005: Stratosphere–troposphere coupling in the Southern Hemisphere. *J. Atmos. Sci.*, **62**, 708–715, https://doi.org/10.1175/JAS-3321.1.

van Loon, H., C. S. Zerefos, and C. C. Repapis, 1982: The Southern Oscillation in the stratosphere. *Mon. Wea. Rev.*, **110**, 225–229, https://doi.org/10.1175/1520-0493(1982)110<0225:TSOITS>2.0.CO;2.

Vera, C., G. Silvestri, V. Barros, and A. Carril, 2004: Differences in El Niño response over the Southern Hemisphere. *J. Climate*, **17**, 1741–1753, https://doi.org/10.1175/1520-0442(2004)017<1741:DIENRO>2.0.CO;2.

Ward, J. H., Jr., 1963: Hierarchical grouping to optimize an objective function. *J. Amer. Stat. Assoc.*, **58**, 236–244, https://doi.org/10.1080/01621459.1963.10500845.

Waugh, D. W., W. J. Randel, S. Pawson, P. A. Newman, and E. R. Nash, 1999: Persistence of the lower stratospheric polar vortices. *J. Geophys. Res.*, **104**, 27 191–27 201, https://doi.org/10.1029/1999JD900795.

Yang, Y., S.-P. Xie, L. Wu, Y. Kosaka, N.-C. Lau, and G. Vecchi, 2015: Seasonality and predictability of the Indian Ocean dipole mode: ENSO forcing and internal variability. *J. Climate*, **28**, 8021–8036, https://doi.org/10.1175/JCLI-D-15-0078.1.

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 Dev.*, **13**, 515–528, https://doi.org/10.1080/17565529.2020.1808438.

Yu, J.-Y., and K. M. Lau, 2005: Contrasting Indian Ocean SST variability with and without ENSO influence: A coupled atmosphere-ocean GCM study. *Meteor. Atmos. Phys.*, **90**, 179–191, https://doi.org/10.1007/s00703-004-0094-7.