Causal model evaluation of Arctic-midlatitude teleconnections in CMIP6

Evgenia Galytska[1,2], Katja Weigel[1,2], Dörthe Handorf[3], Ralf Jaiser[3], Raphael Köhler[3], Jakob Runge[4,5], and Veronika Eyring[2,1]

[1]University of Bremen, Institute of Environmental Physics, Bremen, Germany
[2]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[3]Alfred Wegener Institute, Potsdam, Germany
[4]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Datenwissenschaften, Jena, Germany
[5]Technische Universität Berlin, Berlin, Germany

Key Points:

- The robustness of Arctic-midlatitude teleconnections is evaluated by applying causal discovery to observations and CMIP6 simulations.
- Most of the observed links are found in the majority of historical and future model simulations.
- CMIP6 SSP5-8.5 simulations do not show drastic changes in the mechanism of Arctic-midlatitude links in the second half of the 21st century.

Corresponding author: Evgenia Galytska, egalytska@iup.physik.uni-bremen.de

Abstract

In this study, we apply causal discovery to analyse causal links among key processes that contribute to Arctic-midlatitude teleconnections. First, we calculate the causal dependencies from observations. We then evaluate climate models participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6) via a comparison of their causal graphs for the period of 1979-2019 with those derived from observations. Based on observations, we show that the increase (decline) of near-surface Arctic temperature is associated not only with the reduction (increase) of sea ice over the Barents and Kara seas, but also with the strengthening (weakening) of atmospheric blocking over central Asia. We show that the near-surface westerly winds are strongly associated with the phase of the North Atlantic Oscillation (NAO). Observations show that the phase of NAO is connected with the polar vortex (PV), which is affected by the strengthening of the poleward eddy heat flux at 100 hPa. The analysis of CMIP6 historical simulations is in good agreement with the observations but reveals a negative connection between near-surface Arctic temperature and sea ice over Barents and Kara seas, which was not found in observations during December-January-February 1979-2019. Moreover, climate models simulate a more robust link between Arctic temperature and Barents and Kara sea ice towards the end of the century. The analysis of future changes in the Arctic-midlatitude teleconnections during cold seasons 2059-2099 also reveals that the connection between the Aleutian Low and the poleward eddy heat flux is expected to become more robust than in the analysed past.

1 Introduction

The warming of the Arctic is generally considered to be about twice as fast as the global average. This robust phenomenon known as Arctic amplification, emerged relatively recently (England et al., 2021) and is one of the prominent indications of climate change (Koenigk et al., 2020; Previdi et al., 2021). However, recently Rantanen et al. (2021) showed that during the last 40 years the pace of Arctic warming is almost four times that of the globe as a whole, which is higher than previously reported. Arctic amplification occurs in all seasons except boreal summer, with the strongest warming in fall and winter (Cohen et al., 2020; Previdi et al., 2021). The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) concluded that "it is virtually certain that the Arctic will continue to warm more than global surface temperature, with high confidence above two times the rate of global warming" (IPCC, 2021). Different mechanisms have been suggested to contribute to the amplified Arctic warming, such as changes in the snow- and ice-albedo feedback (Serreze & Barry, 2011), lapse rate feedback (Pithan & Mauritsen, 2014), cloud cover and water vapor feedback (Graversen & Wang, 2009), increase of atmospheric CO_2 (Previdi et al., 2020), or atmospheric energy transport and ocean heat uptake (Previdi et al., 2021). The overviews of major Arctic climate feedbacks are for example provided by Goosse et al. (2018), Previdi et al. (2021), Vavrus (2018), and Wendisch et al. (2022).

Meanwhile the Northern Hemispheric (NH) midlatitudes, in particular Eurasia, have experienced an increase in severe winter weather events, which coincided with a cooling or lack of warming over this region (Overland et al., 2015). A number of studies link Arctic amplification to midlatitude weather variability (Francis & Vavrus, 2012; Francis, 2017; Screen et al., 2018; Shepherd, 2016; Walsh, 2014), which is not a one-way connection, but also works in reverse (Screen, 2017a). Despite of the progress made, the possible contribution of Arctic amplification to midlatitude weather remains an open question. This is associated with the inconclusive results from observational studies and model simulations that obscure a full understanding of the teleconnections between the Arctic and midlatitudes (Barnes & Screen, 2015; Cohen et al., 2020) and are related to the internal variability in the climate system with possible nonlinear connections, and biases of climate models.

In the following we discuss proposed plausible mechanisms for Arctic-midlatitude linkages. Amplified Arctic warming during boreal winters has been associated with the extensive reduction of sea ice (Cohen et al., 2014, 2020; Francis, 2017; Screen & Simmonds, 2010; Serreze et al., 2009). Arctic sea ice strongly modulates near-surface atmospheric conditions at high latitudes, which then influence regional and, potentially, remote climate (Cohen et al., 2014). In particular, the sea ice over the Barents and Kara seas has been identified as potentially being skillful in predicting midlatitude weather and climate (e.g. Hall et al., 2017; Scaife et al., 2014; Siew et al., 2020; Wang et al., 2017). In accordance, these studies provide evidence and are in agreement about the dynamical pathway linking sea ice changes over the Barents and Kara seas to the midlatitudes. This region experiences sea ice loss not only in summer and autumn, but also in winter, which is the strongest and is associated with positive sea surface temperature anomalies. Primarily, during autumn and winter the release of additional longwave radiation as well as sensible and latent heat from the open ocean waters causes a strong warming of the lower troposphere over the Barents and Kara seas. This contributes to the weakening of the westerly winds and to more frequent and persistent Ural blocking (Yao et al., 2018). The diabatic heating of the lower troposphere over the Barents and Kara seas results in direct forcing and constructive interference with existing planetary Rossby waves (Honda et al., 2009), contributing to a northwestward expansion and intensification of the Siberian High. These changes in tropospheric circulation (Ural blocking, strengthened Siberian High) are characterized by increased amplitudes of planetary waves and are favourable for an initiation of the stratospheric pathway for Arctic-midlatitude linkages (Nakamura et al., 2015; Peings, 2019; Siew et al., 2020). Enhanced upward propagating planetary waves lead to wave breaking in the polar stratosphere which in turn may weaken the polar vortex (e.g. Hoshi et al., 2017; Jaiser et al., 2016; Kim et al., 2014). This is then followed by downward propagation of stratospheric circulation anomalies into the troposphere. This mechanism favors the negative phase of the North Atlantic Oscillation (NAO, Baldwin & Dunkerton, 2001). Teleconnection patterns like the NAO are related to the zonal wind variability (Wallace & Gutzler, 1981), in particular the negative phase of the NAO is related to a southward shifted and weakened North-Atlantic jet stream (Athanasiadis et al., 2010). Moreover, the negative phase of the NAO is associated with cold winters and increased occurrence of cold temperature extremes in Northern Europe (Hurrell & Deser, 2010; Marshall et al., 2001; Riebold et al., 2022). Thus, the variability of the NAO has an impact on the occurrence probability of extreme weather events.

The link between Arctic sea ice cover and the NAO has been extensively analysed, for example, Pedersen et al. (2016) showed that the loss of sea ice in specific regions of the Arctic impacts the spatial structure of the NAO pattern, and Nakamura et al. (2015) discussed the contribution of the Barents-Kara sea ice changes to changes in the frequency of occurrence of positive or negative phases of the NAO. In turn, Ambaum et al. (2001) and Kolstad and Screen (2019) showed that the NAO strongly influences wintertime weather and climate in the North Atlantic region, therefore it can play a significant role in the linkage of Arctic and midlatitudes. Compared to the evidence found for the pathways for Arctic-midlatitude linkages related to Barents-Kara sea ice loss, the responses to sea ice loss and Arctic Amplification over other geographical regions are less robust. However, recent coupled ocean–atmosphere model experiments (Screen et al., 2018) provide evidence for a strengthened Aleutian low in response to Arctic sea ice loss. This is consistent with radical circulation changes detected from reanalysis data over 1958-2007 showing strong covariability between the strength of pressure anomalies over the Eurasian Arctic coast and the area of the Aleutian Low. The Aleutian Low is a semi-permanent low-pressure system over the midlatitude Northern Hemisphere centered near the Aleutian Islands, which acts to increase the Okhotsk sea ice cover by cold air advection (Ogi et al., 2015; Tachibana et al., 1996). Moreover, it is the dominant mode of the atmospheric circulation on interannual to decadal time-scales over the North Pacific in winter (Hwang et al., 2022; Trenberth & Hurrell, 1994).

To analyse the connection between the Arctic and midlatitudes, we apply causal discovery (Runge, Nowack, et al., 2019; Runge, Bathiany, et al., 2019; Spirtes et al., 2000) to identify and quantify significant causal interactions among various local and remote processes and utilize it for causal climate model evaluation (Eyring et al., 2019; Nowack et al., 2020). The goal of causal discovery is to estimate causal links including their time lags among a number of processes, often referred to as actors. Previous studies on Arctic-midlatitude linkages using causal discovery have focused on observational datasets (Kretschmer et al., 2016, 2018; Polkova et al., 2021; Siew et al., 2020). Our study, in turn, assesses causal Arctic-midlatitude links based on a number of climate model simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6, Eyring et al., 2016) operated under the auspices of the Working Group on Coupled Modelling (WGCM) World Climate Research Programme (WCRP). The detection of similar causal links both in observations and model simulations has been termed causal model evaluation (Nowack et al., 2020) and provides an opportunity to assess model performance, i.e. it indicates whether models are able to correctly reproduce local and remote processes in the climate system, and do not simulate expected links for the wrong or unknown reasons. In this study we compare causal graphs from CMIP6 historical simulations and reanalysis data to estimate how well these models reproduce the links in the current climate. Climate models provide the additional opportunity to test possible changes in the causal graphs in future projections. Thus, we also compare causal graphs from CMIP6 historical and Scenario Model Intercomparison Project (ScenarioMIP, O'Neill et al., 2016) simulations to estimate future changes in Arctic-midlatitude teleconnections. Based on this, those links that are found in ScenarioMIP simulations are analysed to estimate the impact of the increasing Arctic Amplification in a future climate and to investigate future changes in Arctic-midlatitude teleconnections. Furthermore, we apply the causal model evaluation framework developed by Nowack et al. (2020) to evaluate the overall performance of climate models in comparison to observations.

2 Data

2.1 Data sources

In this study we use CMIP6 historical simulations that are forced by natural (e.g. solar variability, volcanic eruptions) and anthropogenic forcings (e.g. greenhouse gas concentrations, land use, aerosols) and are available for the period 1850-2014 (Eyring et al., 2016). To extend the analysed period, we add data from ScenarioMIP (O'Neill et al., 2016) that provides climate projections based on different plausible scenarios of future emissions and land use changes during 2015-2100. In this study we use the Shared Socioeconomic Pathway (SSP)5-8.5 representing emissions that produce a radiative forcing of 8.5 W/m^2 in 2100, which is at the high end of the range of future pathways in the integrated assessment models (IAM) literature. We combine monthly mean CMIP6 historical simulations (1979-2014) and ScenarioMIP SSP5-8.5 (2015-2019) for the 41-year period of 1979-2019. We compare the CMIP6 historical simulations then with the monthly mean Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST, Rayner et al., 2003) and ERA5 reanalysis monthly mean data from the European Centre for Medium-Range Weather Forecasts (ECMWF, Hersbach et al., 2020). This comparison aims to identify the mechanism of Arctic-midlatitude teleconnections in the historical period. To address future changes, we analyse the ScenarioMIP SSP5-8.5 during 2059-2099, which is the same period length as historical simulations. In this study, we use data from 19 CMIP6 models listed in Table 1. We consider all available ensemble members (three or more per model) to account for the intrinsic variability of the analysed models. However, we show the analysis of only the first ensemble member (r1i1p1f1) for each of them. Following Gier et al. (2020), we do not use the ensemble mean of each model since it would reduce the intrinsic variability of the single realisations.

Table 1. CMIP6 models analysed in this study (both historical and ScenarioMIP SSP5-8.5 simulations).

Model	Origin	Reference
ACCESS-CM2	Commonwealth Scientific and Industrial Research Organisation, Australia and Australian Research Council Centre of Excellence for Climate System Science	Bi et al. (2020)
ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organisation, Australia	Ziehn et al. (2020)
BCC-CSM2-MR	Beijing Climate Center, China	Wu et al. (2019)
CAMS-CSM1-0	Chinese Academy of Meteorological Sciences, China	Rong et al. (2021)
CanESM5	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Canada	Swart et al. (2019)
CMCC-CM2-SR5 CMCC-ESM2	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	Cherchi et al. (2019) Lovato et al. (2022)
EC-Earth3 EC-Earth3-Veg EC-Earth3-Veg-LR	EC-Earth consortium, Europe	Döscher et al. (2022); Wyser et al. (2020)
GFDL-CM4 GFDL-ESM4	National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, USA	Held et al. (2019) Dunne et al. (2020)
INM-CM4-8 INM-CM5-0	Institute for Numerical Mathematics, Russian Academy of Science, Russia	Volodin et al. (2018) Vorobyeva and Volodin (2021)
IPSL-CM6A-LR	Institut Pierre Simon Laplace, France	Boucher et al. (2020)
MPI-ESM1-2-HR MPI-ESM1-2-LR	Max Planck Institute for Meteorology, Germany	Müller et al. (2018) Mauritsen et al. (2019)
MIROC6	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, The University of Tokyo, National Institute for Environmental Studies, and RIKEN Center for Computational Science, Japan	Tatebe et al. (2019)
MRI-ESM2-0	Meteorological Research Institute, Japan	Yukimoto et al. (2019)

Table 2. Detailed overview of the actors and their corresponding regions used in this study.

No	Actor		Geographical area	Label
1	Near-Surface Air Temperature		65°-90°N, zm	TAS
2	Sea Ice Area Fraction	Barents-Kara	70°-80°N, 30-105°E	BK-SIC
3		Okhotsk	50°-60°N, 140-160°E	Ok-SIC
4	Sea Level Pressure	Ural blocking	45°-70°N, 40-85°E	Ural-SLP
5		Siberia high	40°-65°N, 85-120°E	Sib-SLP
6		Aleutian low	45°-80°N, 160-260°E	Aleut-SLP
7	Poleward Eddy Heat Flux		45-75°N, zm, 100hPa	vflux
8	Polar Vortex (geopotential height)		65°-90°N, zm, 100-10hPa	PV
9	North Atlantic Oscillation		20°-80°N, 500hPa	NAO
10	Near-Surface Zonal Wind		50°-70°N, zm	U

2.2 Potential actors

To understand the linkages within Arctic-midlatitude processes it is important to define a physically reasonable number of variables, so-called actors, that represent the investigated mechanisms. Table 2.2 provides a detailed overview of the actors used in this study, and Fig. 1 summarizes potential actors. To reconstruct an Arctic-midlatitude linkage that occurs in conditions of amplified Arctic warming we choose the near-surface air temperature (TAS) at high latitudes as one of the major actors. We include sea ice area fraction over the Barents and Kara seas (BK-SIC) similarly to Kim et al. (2014); Kretschmer et al. (2016); Siew et al. (2020). To understand the connections between the Arctic (and sub-Arctic) and Okhotsk sea ice variability, we include sea ice area fraction over Okhotsk Sea (Ok-SIC, Ogi et al., 2015). To represent the impact of changes in the high (Ural and Siberian) and low (Aleutian) pressure systems, we include sea level pressure over these regions (Ural-SLP, Sib-SLP, and Aleut-SLP correspondingly). The importance of Ural blocking and Siberian High for Arctic-midlatitude linkages has been extensively discussed by Cohen et al. (2014), and the effect of the Aleutian Low on the Okhotsk sea ice is analysed by Ogi et al. (2015).

To capture the linkage between troposphere and stratosphere through the upward propagation of planetary waves, we include the poleward eddy heat flux (vflux), which is proportional to the vertical component of the Eliassen-Palm flux in the transformed Eulerian mean framework (Andrews & Mcintyre, 1976). The poleward eddy heat flux has been calculated similarly to Kim et al. (2014); Kretschmer et al. (2016); Siew et al. (2020) as va^*Ta^* at 100 hPa, where va stands for the meridional wind velocity, Ta stands for the temperature, and the superscript * indicates the deviations from the zonal mean. To analyse the variability in stratospheric polar circulation patterns, we also include the Polar Vortex (PV), based on geopotential height anomalies. A recent study by Kolstad and Screen (2019) suggests a link between the reduction of Arctic sea ice and the negative phase of the NAO. Therefore, we also include NAO in our analysis, which is based on the study by Hurrell and Deser (2010). Moreover, the NAO is linked to the variability of midlatitude tropospheric circulation (Smith et al., 2022), thus we include NH near-surface zonal mean zonal winds (U) at midlatitudes into our analysis. To summarize, we use the following potential actors that represent Arctic and midlatitude processes: near-surface air temperature over the Arctic (TAS), sea ice over the Barents and Kara seas (BK-SIC) and Okhotsk Sea (Ok-SIC), sea level pressure over Ural (Ural-SLP), Siberia (Sib-SLP), and Aleutian Islands (Aleut-SLP), poleward eddy heat flux (vflux), polar vortex (PV), NAO, and zonal mean zonal wind (U).

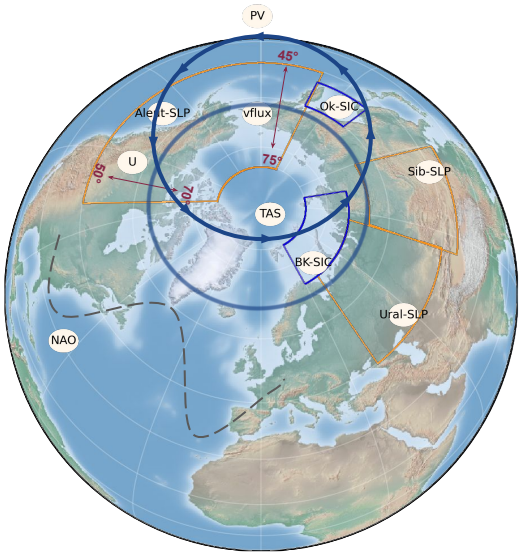

Figure 1. Summary of potential actors that represent Arctic and midlatitude processes in this study: near-surface air temperature over the Arctic (TAS), sea ice over the Barents and Kara seas (BK-SIC) and Okhotsk Sea (Ok-SIC), sea level pressure over Ural (Ural-SLP), Siberia (Sib-SLP), and Aleutian Islands (Aleut-SLP), poleward eddy heat flux (vflux), polar vortex (PV), North Atlantic Oscillation (NAO), and zonal mean zonal wind (U). The position of BK-SIC, Ok-SIC, Ural-SLP, Sib-SLP, and Aleut-SLP corresponds to their approximate geographical location defined in Table 2.

3 Methods and tools

3.1 ESMValTool and data preparation

In this study, we use the Earth System Model Evaluation Tool (ESMValTool) version 2 (Eyring et al., 2020; Lauer et al., 2020; Righi et al., 2020; Weigel et al., 2021). The development of ESMValTool is a community-based effort, and it provides well-documented source code and scientific background of developed diagnostics (see `https://github.com/ESMValGroup/ESMValTool`, last access: 05.10.2022). The output produced by ESMValTool comprises provenance information, which allows for traceability and reproducibility of the obtained results. ESMValTool provides for various simulations and experiments from CMIP6 models an additional recipe called `recipe_filler`, that offers the possibility to easily track a list of models that include necessary variables for the analysis. Since causal discovery requires the usage of timeseries, we use ESMValTool to obtain the timeseries that correspond to each actor. With the application of ESMValTool preprocessor function called `anomalies` we compute climatological monthly anomalies. Then, by applying the `area_statistics` preprocessor function we compute the area-weighted spatial average over the corresponding regions from Table 2. If the resulting timeseries has a linear trend, we remove it, since causal discovery requires a stationary timeseries (Runge, 2018). Similar to Kretschmer et al. (2016) we invert the sign of PV, so the positive values stand for the strong polar vortex, and negative values stand for the weak polar vortex. In this study we use a Climate Variability Diagnostics Package (CVDP, Phillips et al., 2014) implemented into ESMValTool (Eyring et al., 2020) to reproduce NAO for the analysed periods.

3.2 Causal discovery

Correlation techniques that are commonly adopted in climate research, such as Pearson correlation or linear regression models can hardly be used to understand causal relations that emerge from the physical mechanism behind the phenomena we observe (see e.g. Runge et al., 2014). Pearson correlation, also in its lagged form, is known to suffer from identifying spurious links due to confounders or even just autocorrelation (Runge et al., 2014). Granger causality (Granger, 1969), typically applied in a bivariate form, accounts for autocorrelation and has been recently suggested to detect and quantify climate system teleconnections (Silva et al., 2021). Based on the example of El Niño events and precipitation anomalies, the others discovered statistically robust relationships supported by physical mechanisms. However, it has been previously shown that Granger causality is limited to lagged causal dependencies and might detect misleading causal links in low-resolution data (see e.g. Runge, Bathiany, et al., 2019; Spirtes & Zhang, 2016). Moreover, the typically applied bivariate form of Granger causality cannot account for common drivers. Causal discovery methods (Runge, Bathiany, et al., 2019) utilize general assumptions about the underlying processes to fully account for common causes and reconstruct causal relations among multiple variables. Hence, they are a vital element for enhanced causal process understanding and can also be understood as an interpretable technique of machine learning (Xu et al., 2020). Here we employ the conditional independence-based causal discovery framework that utilizes the assumptions of time-order, causal sufficiency, the Causal Markov condition, and faithfulness (Runge, Nowack, et al., 2019). This method is based on iterative conditional independence testing and has already found its application in the analysis of various teleconnections based on observations (Di Capua et al., 2020; Ebert-Uphoff & Deng, 2012; Kretschmer et al., 2016, 2018; Runge et al., 2014; Siew et al., 2020), pathways of teleconnections (Karmouche et al., 2022; Kretschmer et al., 2021; Runge et al., 2015), marine cold-air outbreaks (Polkova et al., 2021), Walker circulation (Runge et al., 2014; Runge, Bathiany, et al., 2019), and process-oriented climate model evaluation (Nowack et al., 2020).

The PCMCI causal discovery framework (Runge, Bathiany, et al., 2019; Runge, Nowack, et al., 2019) used here is based on a combination of the PC algorithm (named after its inventors Peter and Clark, see Spirtes & Glymour, 1991) and the Momentary Conditional Independence (MCI) test, which is adapted to the typically ubiquitous autocorrelation in timeseries data. In this study, we use an extended version of the PCMCI algorithm, called PCMCI+ (Runge, 2020), which detects not only lagged (time lag $\tau > 0$), but also contemporaneous ($\tau = 0$) causal links. Like the PCMCI algorithm, PCMCI+ optimizes the choice of conditioning sets and therefore improves the reliability of the conditional independence tests. Additionally, Runge (2020) showed that PCMCI+ benefits from strong autocorrelated timeseries, which are typical in climate science, and leads to stronger adjacency detection power with better control of false positives and higher orientation recall for contemporaneous links compared to the standard PC algorithm or multivariate Granger causality.

In the causal graphical model framework one assumes an underlying discrete-time structural causal process $\mathbf{X_t} = (X_t^1, ..., X_t^N)$, where N stands for the different actors represented by timeseries. To reconstruct the causal graph including their time lags among these actors, the PCMCI+ algorithm encompasses three phases (Karmouche et al., 2022) and for a detailed description and pseudo-code see Runge (2020). In the first phase, the so-called lagged skeleton discovery phase, the PC1 Markov set discovery algorithm (based on the PC algorithm) is iteratively applied for every lagged pair of actors $(X_{t-\tau}^i, X_t^j)$ for $\tau > 0$. For each pair of actors PC1 tests whether they are conditionally independent on defined conditions of other lagged actors. If yes, then the algorithm removes the link between these actors. At this stage the algorithm estimates a set of lagged parents for each actor. However this step might still include remaining spurious connections due to contemporaneous links. In the second phase, the contemporaneous skeleton discovery phase, the contemporaneous conditions are iterated in MCI tests:

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t | \mathcal{S}, \widehat{\mathcal{B}}^-_t(X^j_t) \setminus \{X^i_{t-\tau}\}, \widehat{\mathcal{B}}^-_{t-\tau}(X^i_{t-\tau}) \quad (1)$$

where $\mathcal{S}$ stands for the subsets of contemporaneous adjacencies $\mathcal{S} \subset \mathbf{X_t}$, through which the algorithm is iterating to fully remove spurious links, $\widehat{\mathcal{B}}^-_t(X^j_t)$ are the lagged conditions of X^j_t, and $\widehat{\mathcal{B}}^-_{t-\tau}(X^i_{t-\tau})$ are the (time-shifted) lagged conditions of $X^i_{t-\tau}$ obtained in the skeleton discovery phase. During the third so-called orientation phase the contemporaneous links are oriented based on the collider rule on unshielded triples $X^i_{t-\tau}$- $- X^k_t - X^j_t$, where $\tau \geq 0$, and finally further orientation rules are applied that make sure that no cycles occur (for more details see Runge, 2020). Then, under the standard assumptions of causal sufficiency, faithfulness, and the Markov condition (Runge, Bathiany, et al., 2019; Runge, Nowack, et al., 2019), the outcome of the PCMCI+ algorithm is a causal graph with the following four types of links: (i) directed lagged causal links for $\tau > 0$, where τ stands for the time lag, (ii) directed contemporaneous causal links for $\tau = 0$, (iii) unoriented contemporaneous links indicating that the collider and orientation rules could not be applied due to Markov equivalence, and (iv) unoriented contemporaneous links where a direction is not defined due to conflicting orientation rules.

In this study we focus on linear dependencies and use linear partial correlation (ParCorr) as a conditional independence test. In addition, we set the main parameters of PCMCI+ as follows: the maximum time delay $\tau_{max} = 5$, which corresponds to 5 months, to account for a range of possible dependencies in the Arctic-midlatitude teleconnections on a monthly time scale and the significance level α_{pc}=0.01 for all tests. In the resulting causal graph the node color denotes the autocorrelation (auto-MCI) value at the lag with maximum absolute value and varies from 0 to 1. The strength of causal links is measured by the MCI partial correlation value (cross-MCI) and is, hence, normalized between -1 and 1. If links occur at multiple lags between two variables, the type of the link and its color shows the strongest link, but the label indicates all significant lags sorted by their strength. PCMCI+ is implemented in the python package Tigramite, freely available at `https://github.com/jakobrunge/tigramite`, last access: 05.10.2022.

In this study we focus on the NH cold season, when near-surface amplified Arctic warming is the strongest (Cohen et al., 2020). Thus, we apply causal discovery for the winter season consisting of the months December-January-February (DJF). In order to identify the differences in the mechanism of Arctic-midlatitude teleconnections within the whole cold season, we perform further analyses on two additional periods, i.e. early winter (October-November-December, OND) and late winter (January-February-March, JFM). The additionally analysed periods are constructed in a manner so that there is no overlap between the early and late winter period, thus, excluding possible contributions of a common month. The contribution of December from defined early winter period and January-February from late winter period is then included into the regularly analysed winter season DJF. To account for various months, we set `mask_type = y` within the PCMCI+ calculations. This implies that PCMCI+ will search for causal drivers of target actors restricted to defined seasons, e.g. DJF, but the potential causal drivers will be allowed outside of DJF, e.g. in autumn months. The resulting causal graphs contain the information on a direction and associated time lags of potential causal links, characterizing the pathways of the Arctic-midlatitude interaction network.

We apply the causal discovery algorithm to CMIP6 historical simulations and compare resulting causal graphs with observational data. We also analyse causal graphs based on SSP5-8.5 model simulations. We do not calculate causal networks for multi-model mean data, since it would reduce the amplitude of interannual variability and therefore the ability of the framework to detect connections.

3.3 *F*1-score

To evaluate the similarity of the resulting causal graphs that consist of multiple causal and contemporaneous links from observations and climate models, we use the asym-

metric $F1$-score method from Nowack et al. (2020). Generally, the existence or non-existence of each link detected in the observations (reference causal graph) is pair-wise compared to the links from the climate models. This method depends on the statistical significance threshold α used in PCMCI+. The F1-score varies from 0 indicating no match, and 1 indicating a perfect match. The $F1$-score can be interpreted as a harmonic mean of the precision (P) and recall (R), which build the foundation of this method. The precision is defined as follows:

$$P = \frac{TP}{TP + FP} \tag{2}$$

where TP stands for the number of true positives, and FP stands for the number of false positive links in comparison to reference causal graph. The recall is defined as follows:

$$R = \frac{TP}{TP + FN} \tag{3}$$

where FN stands for not detected links. The relative contributions of precision and recall to the $F1$-score are equal. The $F1$-score is defined as follows:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{4}$$

Similarly to Nowack et al. (2020), the definition of TP in this work is slightly modified to take into account the sign of each identified link (positive or negative). If the sign is the same, then we check if the time lag in matching causal links between reference and modelled causal graphs is the same. If not, we allow a difference in the time delay of up to $\pm$ 2 time lags, i.e. two months for a monthly time scale. If the same-sign link in the modelled causal graph was found in a different time step than the reference graph, then the sign of the connection is used at the original time step. If the sign is the same, then we consider that the link exists in both datasets. Therefore, the F1-score depends on the causal graph, which is considered as a reference. For the analysis of CMIP6 historical simulations, we define observational causal graphs as the reference. To estimate how strong Arctic-midlatitude linkages are affected by climate change, for each model from a CMIP6 future simulation we choose historical simulations as a reference. In this case we do not allow a difference in the time delay between historical and future simulations and compare only identical matching links. To optimize the comparison among observational and number of modelled causal graphs, we relax the definition of contemporaneous links (with $\tau = 0$), and do not distinguish between directed, unoriented, and conflicting contemporaneous links. Instead, we combine them into one category of contemporaneous links to summarize the general outcome of modelled causal graphs.

3.4 Causal model evaluation

To characterize connections that occur in the framework of Arctic-midlatitude processes, Fig. 2 depicts the individual steps within the scope of causal model evaluation. First, based on the expert knowledge and literature review, we consider in the analysis potential actors (see Fig. 2a) that represent Arctic-midlatitude processes (see Sect. 2.2). In this step we involve ESMValTool to calculate the climatological anomalies of monthly mean data and the area-weighted spatial average over the defined region for observations and climate models (see Sect. 3.1). As a result, we receive a set of timeseries as proxies for different physical processes (Fig. 2b) for each data source (e.g. Observations, Model 1, Model 2), which we further detrend.

We calculate interactions among these processes in observations and model simulations by the application of PCMCI+ (see Sect. 3.2). The resulting networks (see Fig. 2c) contain the information on existence of associated links between actors at a given significance level. The color of the node of each actor indicates the autocorrelation and varies from 0 to 1 (predominantly 1-month lag, labeled auto-MCI). The color of the links stands for the sign and the strength of the connection (MCI), and varies from -1 to 1 (labeled cross-MCI). Straight lines stand for the contemporaneous links, which do not have

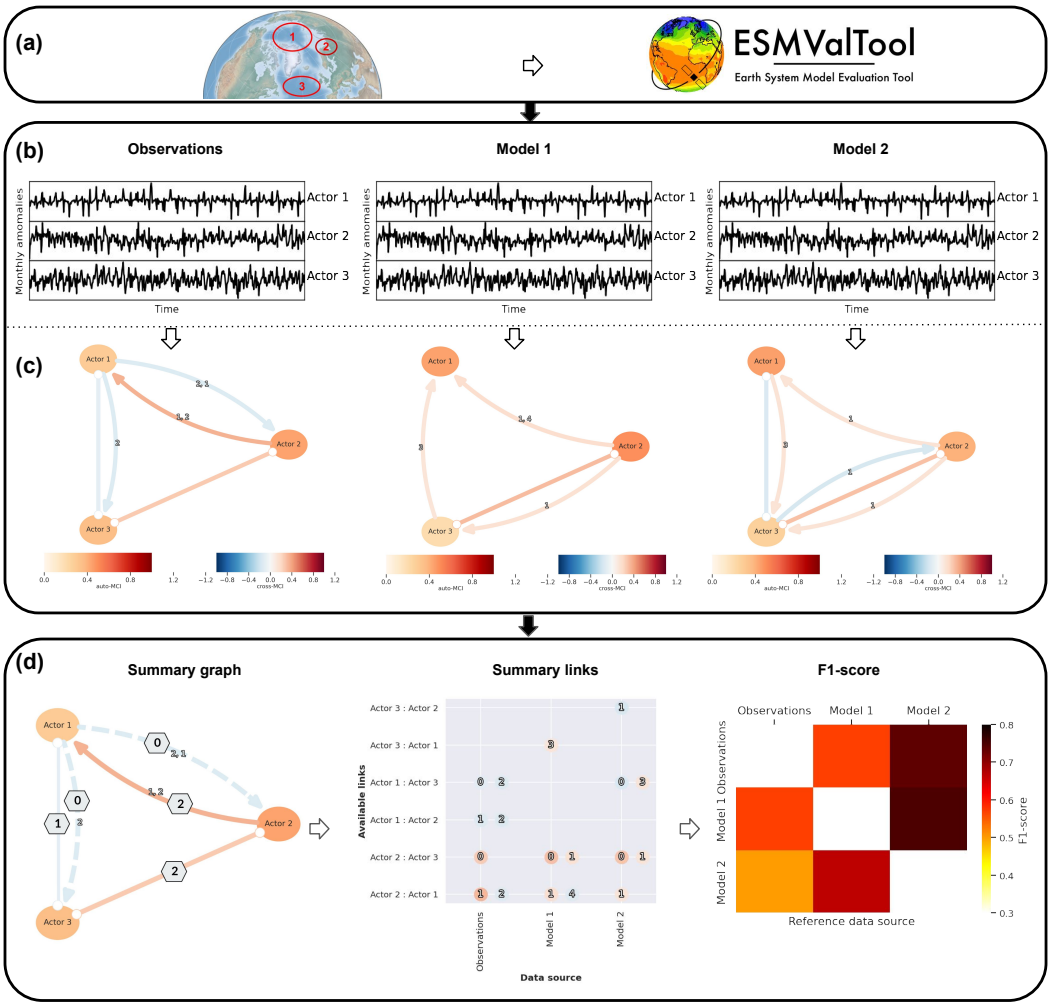

Figure 2. Schematic representation of the causal model evaluation framework using ESMValTool and causal discovery based on example data. For more explanation see Sect. 3.4.

a time lag (see e.g. positive contemporaneous link between Actor 2 and Actor 3 in all data sources in Fig. 2c). Curved lines with the arrow stand for the causal links, the arrow indicates the direction of the impact, and the number(s) stand for the time lag in months (see e.g. positive causal link from Actor 2 to Actor 1 in all data sources in Fig. 2c). If two processes are causally linked with each other for more than one time lag, then all lags are shown sorted by the strength of the impact. In this case the link color would be based on the stronger connection. For example, Fig. 2c shows the negative causal link in Observations from Actor 1 to Actor 2 with time lags 2 and 1, where 2 has stronger connection, thus is indicated first.

In the example given in Fig. 2b,c the analysis is based on three data sources with three actors. But the comparison of causal graphs can be a challenging task, especially if the number of actors and/or data sources are increased, since it can lead to complex causal structures. Therefore, Fig. 2d (left) summarizes the detection of various connections in climate models in comparison to observations. The hexagon above each link indicates the number of models that reproduce (contemporaneous or causal) observed links identical by sign and type. The width of the links shows the fraction of the models, that simulate the particular connections. If none of the models simulate observed connection, then the link is dashed. For example, both models (Model 1 and Model 2) similar to ob-

servations, simulate a contemporaneous positive link between Actor 2 and Actor 3 (Fig. 2c), which is indicated by the number "2" in the hexagon on top of this link (Fig. 2d left). Neither of models simulate observed negative causal link from Actor 1 to Actor 2, which is indicated by "0" in the hexagon on top of this link, and the corresponding link is dashed. This analysis is performed for all detected observational links and is done to provide an overview of how many models agree with observed links. While this analysis gives only an impression on how many models reproduce observed links, it does not include the information of (i) which models do not simulate the observed connections, and (ii) whether there are links that are found in model simulations, but not found in observations. Therefore, the matrix in Fig. 2d (middle) summarizes all discovered links (rows) versus data source (columns). Here the node color stands for the strength of the link, the number inside the node indicates the time lag in months, where 0 stands for a contemporaneous link, 1- causal link with one month delay, 2 - causal link with two months delay etc. And as a last step, to evaluate the general performance of climate models, we calculate the $F1$-score (see Fig. 2d, right, see also Sect. 3.3). Here, rows stand for the reference data, columns are the data sources that are compared to the references. The lower values represent worse, and the higher values (close to one) better matches with the reference data.

4 Results and discussion

4.1 Causal links in observations

To evaluate the performance of CMIP6 models and to understand the causal links in observations, we first reconstruct causal graphs of Arctic-midlatitude teleconnections based on ERA5 and HadISST datasets for the period 1979-2019. Figure 3 shows causal graphs for (a) winter (DJF), (b) early winter (OND), and (c) late winter (JFM) with a significance level of α_{pc}=0.01 and a maximum time lag of $\tau_{max} = 5$ for all tests in PCMCI+. The node color represents the value of the auto-MCI at the lag with the maximum absolute value. Directed lagged and contemporaneous causal links are depicted by straight and curved arrows respectively pointing the direction of the impact. The position of actors in Fig. 3 corresponds to their approximate geographical location defined in Table 2.

From the causal graph of Arctic-midlatitude linkages during DJF (Fig. 3a) among all actors we find the highest autoregressive values (auto-MCI) for Ok-SIC (0.6) and BK-SIC (0.58) both with a time lag of one month. Panel (a) shows a positive contemporaneous link (cross-MCI) between TAS and Sib-SLP (0.28), indicating the direct relationship between Arctic near-surface temperature and Siberian High. Since the link is not directed (no arrow), this connection only implies that (i) the increase (decline) of Arctic near-surface temperature is associated with the strengthening (weakening) of Siberian High, or (ii) the increase (decline) of sea level pressure over Siberia is associated with the increased (decreased) near-surface temperature over the Arctic. The relation between TAS and Sib-SLP also stands for a robust "warm Arctic- cold Siberia" pattern, since an anticyclonic anomaly around the Siberian coast induces cold air spells over NH midlatitudes (Tyrlis et al., 2020). In turn, the Siberian High is positively contemporaneously linked with the Ural blocking (Sib-SLP and Ural-SLP, 0.54), which agrees with findings of (Cohen & Jones, 2011; Cohen et al., 2014; Peings, 2019; Tyrlis et al., 2020) that Ural blocking is a key driver of "warm Arctic- cold Siberia" pattern". Also, the increased sea level pressure over Ural and Siberia is associated with the near-surface temperature increase over the Arctic (direct link between TAS and Sib-SLP, and indirect link between Ural-SLP and TAS via Sib-SLP), which is in agreement with Luo et al. (2016), who stated that the Arctic warming favors midlatitude blocking flows.

During DJF we find a contemporaneous positive link between the NAO and PV with a cross-MCI value 0.24, indicating direct tropospheric-stratospheric relationship between Arctic and midlatitudes. It implies that the negative (positive) phase of the NAO

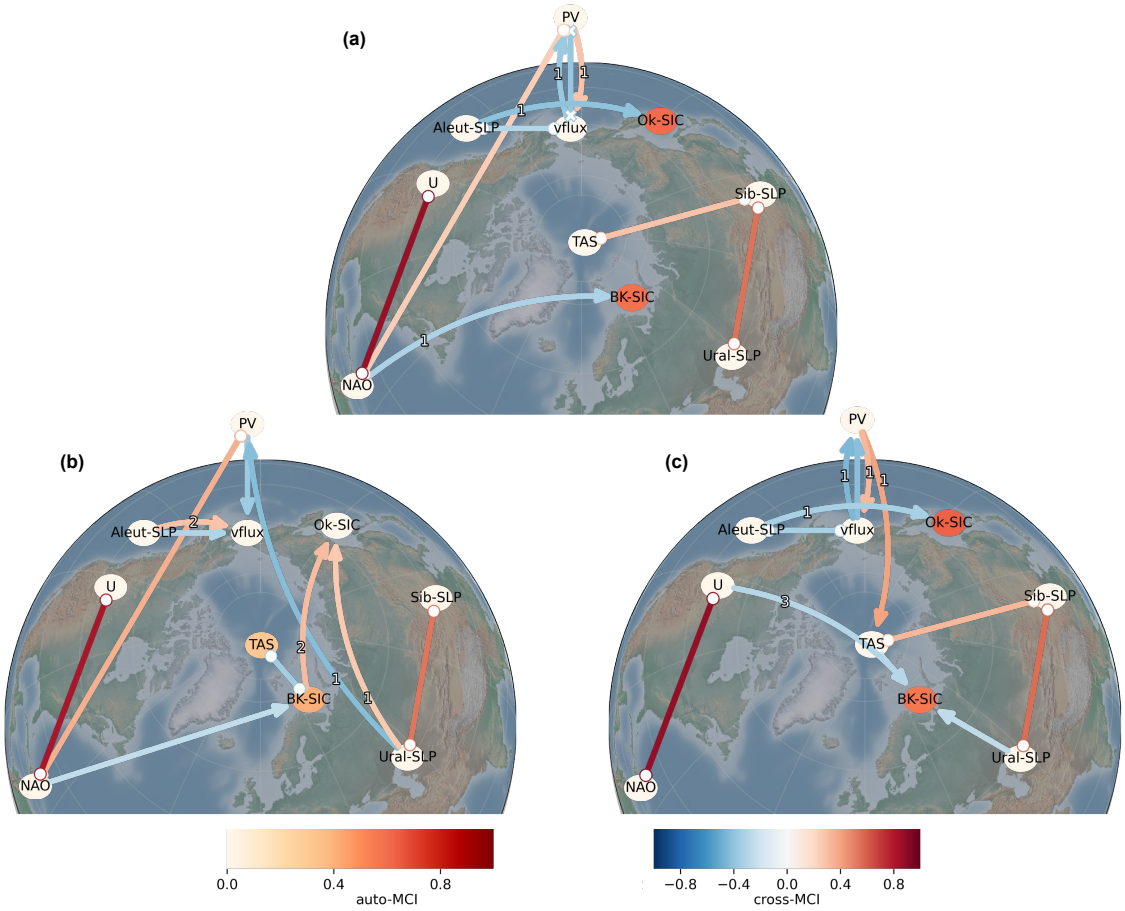

Figure 3. Causal graphs based on ERA5 and HadISST datasets for the period 1979-2019 during **(a)** winter (DJF), **(b)** early winter (OND), and **(c)** late winter (JFM). The significance level is α_{pc}=0.01 and the maximum time lag $\tau_{max} = 5$. Similar to Fig. 1, the position of actors corresponds to their approximate geographical location defined in Table 2.

is associated with a weakening (strengthening) of the PV. For example, a weaker PV produces then circulation anomalies that are associated with an increasingly meandering jet stream that resembles the negative phase of the NAO. The contemporaneous positive link between the NAO and U (0.86) indicates that the negative (positive) phase of the NAO results in a weakening (strengthening) of westerly winds, which is consistent with study of Wallace and Gutzler (1981) and more recent findings of Athanasiadis et al. (2010), Screen (2017b), and Smith et al. (2022). The link between NAO and U is the strongest among all detected DJF links in Fig. 3(a). Additionally, we find a causal negative impact from NAO to BK-SIC (-0.31) with one month lag, which represents a tropospheric connection between midlatitudes and the Arctic.

During DJF we also find the negative links between vflux and PV, which are associated with the weakening of polar vortex due to increased poleward eddy heat flux not only instantaneously (-0.38), but also with one month delay (-0.37). In turn, the weakened PV induces further weakening of the poleward eddy heat flux, indicated via causal positive link with one month lag (0.28). In the North Pacific region during DJF we detect a causal negative link from the Aleutian Low to sea ice over the Sea of Okhotsk (from Aleut-SLP to Ok-SIC, -0.4) with one month delay. This is in agreement with Ogi et al. (2015), who linked the deepening of the Aleutian Low to the increase in the Okhotsk sea

ice cover. There is a negative contemporaneous link between Aleut-SLP and vflux (-0.3), so that a deepening of the Aleutian Low is associated with an increase in poleward eddy heat flux at 100 hPa.

During early winter (OND, Fig. 3b) the distribution of contemporaneous and causal links differs from DJF (Fig. 3a). Therefore, in the following we focus only on those links and changes in the connections that appeared newly in comparison to DJF. During OND the highest auto-MCI values in observations are found for BK-SIC (0.41, which is lower than during DJF, see panel (a) and TAS (0.31) both with one month time lag. Contrary to DJF, the positive link between TAS and Sib-SLP is not found during OND. However, we find a contemporaneous negative link between TAS and BK-SIC (-0.35), which was not found during DJF. This link indicates that the increase (decrease) of near-surface temperature over the Arctic is related to the decline (increase) of sea ice over Barents and Kara seas. This is consistent with findings of Screen et al. (2012), who showed that Arctic sea ice changes are one of the main drivers of near-surface Arctic temperature trends. However it is important to highlight that since observational causal graph is also estimated with errors, the absence of the link (for example between TAS and BK-SIC during DJF) could happen if the connection is detected as false negative in the causal graph.

We also find a causal positive link with two months lag from BK-SIC to Ok-SIC (0.31). Based on the causal graph, we find another positive causal link to the Ok-SIC from the Ural-SLP (0.26) with one month lag. Contrary to DJF (Fig. 3a), the link from NAO to BK-SIC is not lagged anymore, but is directed contemporaneous causal link (-0.25). There are no lagged causal links detected between PV and vflux, but only directed contemporaneous from PV to vflux (-0.36). And in the North Pacific region during OND we find a positive causal link from Aleut-SLP to vflux (0.29) with two months delay, which shows that the weakened Aleutian Low can lead to weaker poleward eddy heat flux at 100 hPa.

During late winter (JFM, Fig. 3c) the observations show pronounced auto-MCI values for both Ok-SIC (0.63) and BK-SIC (0.57) similarly to panel (a). We detect a directed contemporaneous negative link from TAS to BK-SIC (-0.24). We also find an unoriented, contemporaneous positive link between TAS and Sib-SLP (0.32), which was also found during DJF (panel a), but not during OND (panel b). Therefore, based on causal analysis of cold periods (panels a-c), the link between Arctic temperature and Siberian High is not occurring earlier than December and persists until late winter. We also find a directed contemporaneous link from Ural-SLP to BK-SIC (-0.25), which is in agreement with Yao et al. (2018), who reported that the reduction of sea ice cover over Barents and Kara seas is due to regional blocking on weekly time scales. Similarly, Siew et al. (2020) suggested that the connection between Ural-SLP and BK-SIC is a matter of days to a week. The latter conclusion explains why the link between Ural-SLP and BK-SIC is detected as contemporaneous on a monthly time scale. Based on Fig. 3a-b, we do not find the link between Ural-SLP and BK-SIC during DJF or OND. This suggests that the impact of Ural blocking to the decrease of sea ice over the Barents and Kara seas occurs only during JFM. We detect a further negative connection to BK-SIC, but from U (-0.29), which is a causal link with three months lag.

Similar to Fig. 3a, during JFM (panel c) in the North Pacific region we find a negative causal connection from Aleut-SLP to Ok-SIC (-0.33), whereas it does not take place during early winter. During JFM we also find two causal positive links with one month lag associated with the PV. Similar to DJF, the first link is from PV to vflux (0.33), indicating that the weakened PV causes weakening of the poleward eddy heat flux. The second link is from PV to TAS (0.4), indicating that the weakening of polar vortex leads to a decrease of Arctic near-surface temperature. This link is associated with the downward coupling of stratospheric polar vortex anomalies, and it is known to be particularly strong during sudden stratospheric warming (SSW) events that can effect tropospheric weather for up to 60 days after the event (Baldwin et al., 2021). The surface temperature impact of SSWs is particularly strong throughout northern Eurasia, where negative surface temperature anomalies often follow SSWs (see e.g. Baldwin et al., 2021;

Table 3. Summary of the Arctic-midlatitude connections based on the application of causal discovery on ERA5 and HadISST datasets for the period 1979-2019. Link type stands for either the lagged causal link (causal), where number in round brackets indicates the time lag in months, or the contemporaneous (contemp.) link. The sign indicates a positive (+) or negative (-) connection. The values in the last three columns show cross-MCI.

No	Link	Link type	Sign	OND	DJF	JFM
1	TAS and BK-SIC	contemp.	-	-0.35		-0.24
2	TAS and Sib-SLP	contemp.	+		0.28	0.32
3	Sib-SLP and Ural-SLP	contemp.	+	0.54	0.54	0.56
4	Ural-SLP and PV	causal (1)	-	-0.42		
5	Ural-SLP and BK-SIC	contemp.	-			-0.25
6	BK-SIC and Ok-SIC	causal (2)	+	0.31		
7	Ural-SLP and Ok-SIC	causal (1)	+	0.26		
8	NAO and BK-SIC	causal (1)	-		-0.31	
9		contemp.	-	-0.25		
10	NAO and PV	contemp.	+	0.35	0.24	
11	NAO and U	contemp.	+	0.82	0.86	0.87
12	Aleut-SLP and vflux	causal (2)	+	0.29		
13		contemp.	-	-0.3	-0.29	-0.32
14	vflux and PV	causal (1)	-		-0.38	-0.45
15		contemp.	-		-0.37	-0.42
16	PV and vflux	causal (1)	+		0.28	0.33
17		contemp.	-	-0.36		
18	Aleut-SLP and Ok-SIC	causal (1)	-		-0.40	-0.33
19	PV and TAS	causal (1)	+			0.4
20	U and BK-SIC	causal (3)	-			-0.29

Hall et al., 2022). These negative anomalies dominate the polar cap mean temperature response to SSWs. Consequently, this can explain the positive causal link between PV and Arctic TAS.

The summary of discussed links is provided in Table 3, where the second column indicates the linked actors, and the third column stands for the type of the link (causal or contemporaneous). If the link is causal then the number in round brackets shows the time delay in months. The forth column shows whether the link is positive (+) or negative (-). The last three columns show the cross-MCI values for those links, which were found in the observations. To summarize the changes in the connections within the cold period, we sort the periods from early (OND, column five) to late (JFM, column seven) winter. According to Table 3, there are four links that appear during all three analysed periods (OND, DJF, and JFM), namely contemporaneous links (sorted by the cross-MCI strength) between (i) NAO and U, (ii) Sib-SLP and Ural-SLP, (iii) PV and vflux, and (iv) Aleut-SLP and vflux. There are a number of connections that are found in only one of the analysed periods. For example, the tropospheric causal links associated with Ok-

SIC such as from BK-SIC to Ok-SIC and from Ural-SLP to Ok-SIC are found only during OND, the downward coupling from the stratosphere to the troposphere via PV to the TAS and link between U and BK-SIC are found only during JFM.

4.2 Evaluation of causal links in CMIP6 historical simulations

Similar to observations, we reconstruct causal graphs for each individual climate model from Table 1 as indicated in Fig. 2c. The representation of all causal graphs in DJF is shown in Fig. S1 in the supporting information. To generally estimate how many climate models simulate physical processes detected in observations, Fig. 4 shows causal graphs based on observations from Fig. 3, but in this case the number in the hexagon above each link indicates how many models out of 19 simulate the link identical to observations by sign (positive or negative) and type (contemporaneous or causal). Additionally, the width of the lines in Fig. 4 visualizes the fraction of the models, that simulate these particular connections. The dashed lines stand for the connections that were not found in the models (also denoted by the "0" in the hexagon). To optimize the comparison between observations and models, in the following we do not distinguish the three types of contemporaneous links, namely (i) directed contemporaneous causal links, (ii) unoriented, and (iii) conflicting contemporaneous adjacencies, but summarize them into one category and simply refer to these connections as contemporaneous connections. It is important to note, that Fig. 4 exhibits those connections, that were detected in observations and excludes those, which were detected by models, but not by the observations. We adhere to a criterion of more than half of analysed models (i.e. ≥ 10) to represent a robust model response.

Many links between Arctic and midlatitudes that were detected in observations are simulated by most of the 19 analysed models. However, there are some links that are detected by less than half of the considered models. For example, during DJF (Fig. 4a) the contemporaneous link between TAS and Sib-SLP is simulated by 8/19 models and during JFM (Fig. 4c) by 7/19 models. The contemporaneous link between NAO and PV during both DJF (panel a) and OND (panel b) is simulated by 6/19 models. Only 1/19 models simulates causal positive link from PV to vflux during DJF and 5/19 models detect this link during JFM (panel c). During JFM there is also 1/19 models that simulates contemporaneous link between BK-SIC and Ural-SLP. We found that 5/19 models simulate the causal link from Ural-SLP to PV, which was detected only during OND (panel b). In the North Pacific region, the causal link from Aleut-SLP to Ok-SIC is found in 2/19 models in DJF (panel a) and in 1/19 models during JFM (panel c). While during DJF the contemporaneous link between vflux and Aleut-SLP is simulated by most of the analysed models (12/19), this connections becomes less robust during OND (panel b, 6/19) and JFM (panel c, 9/19 models).

Additionally, we found several links that were found in observations, but were not simulated by models. During DJF (panel a) this is the causal link with one month lag from NAO to BK-SIC. During OND (panel b) there are three of such links: the contemporaneous link between NAO and BK-SIC, the causal links from Ural-SLP to OK-SIC and from Aleut-SLP to vflux. And during JFM (panel c) this is the causal link from U to BK-SIC.

While Fig. 4 shows general agreement of the analysed models based on the observational links, with Fig. 5 we summarize for observations and each out of 19 analysed CMIP6 models (x-axis) all detected contemporaneous and causal links (y-axis) during DJF 1979-2019. Similar to Fig. 4, here we do not separate three types of contemporaneous links, but summarize them into one group of contemporaneous connections. Figure 5 gives a detailed overview of every link that is found in observations and is simulated by each of the analysed models during DJF. Analogous to Fig. 5, the summary causal graphs for OND and JFM is shown in the supporting information Fig. S2 and Fig. S3 respectively.

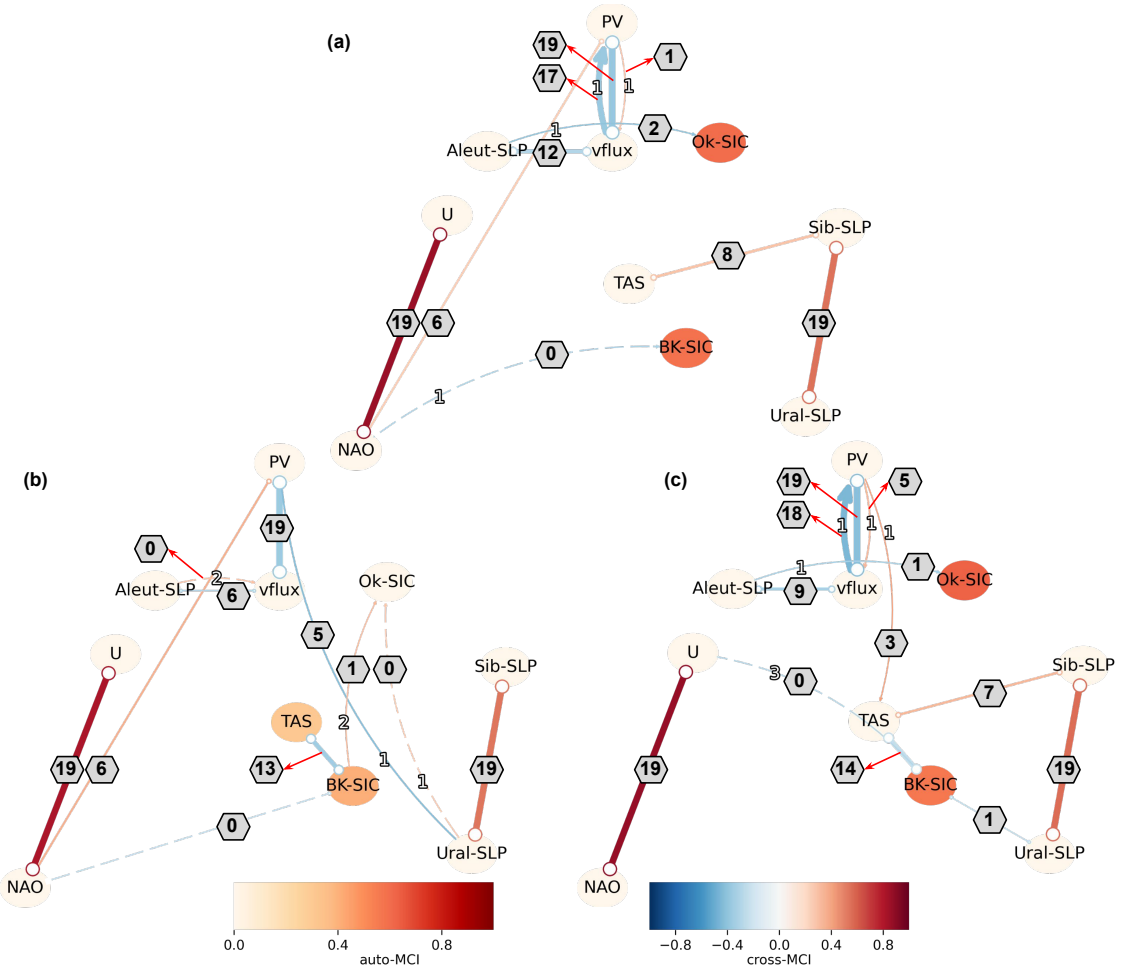

Figure 4. Summary causal graphs based based on 19 CMIP6 historical simulations for the period 1979-2019 during (a) winter (DJF), (b) early winter (OND), and (c) late winter (JFM). Similar to Fig. 3, α_{pc}=0.01 and τ_{max} = 5. The line width represents the fraction of the models that simulate the particular connection. The dashed line indicates those connections, that were not found in the models. The number in the hexagon above each link indicates how many out of 19 models simulate a (contemporaneous or causal) link identical by sign and type in comparison to observations. Here only those links are shown, which were detected in observations.

Based on Fig. 5 we detect that during DJF the only one model that reproduces the causal positive observational link from PV to flux is CMCC-ESM2. The contemporaneous link between TAS and Sib-SLP is simulated by less than half of the analysed models (8/19), namely BCC-CSM2-MR, CAMS-CSM1-0, EC-Earth-Veg, GFDL-CM4, GFDL-ESM4, INM-CM4-8, MPI-ESM1-2-HR, MRI-ESM2-0. While the analysed historical CMIP6 models do not simulate the causal negative connection from NAO to BK-SIC in DJF, which was found in observations, we find 4/19 models (CanESM5, INM-CM4-8, MPI-ESM1-2-LR, and MRI-ESM2-0) that show a contemporaneous negative link between these actors. In the North Pacific, the causal negative link with one month delay from Aleut-SLP to Ok-SIC is simulated by two models: BCC-CSM2-MR and MPI-ESM1-2-LR. We found that there are three other models (GFDL-CM4, GFDL-ESM4, and IPSL-CM6A-LR) that simulate contemporaneous negative link between Aleut-SLP and Ok-SIC. The negative contemporaneous link between vflux and Aleut-SLP is simulated by most of the

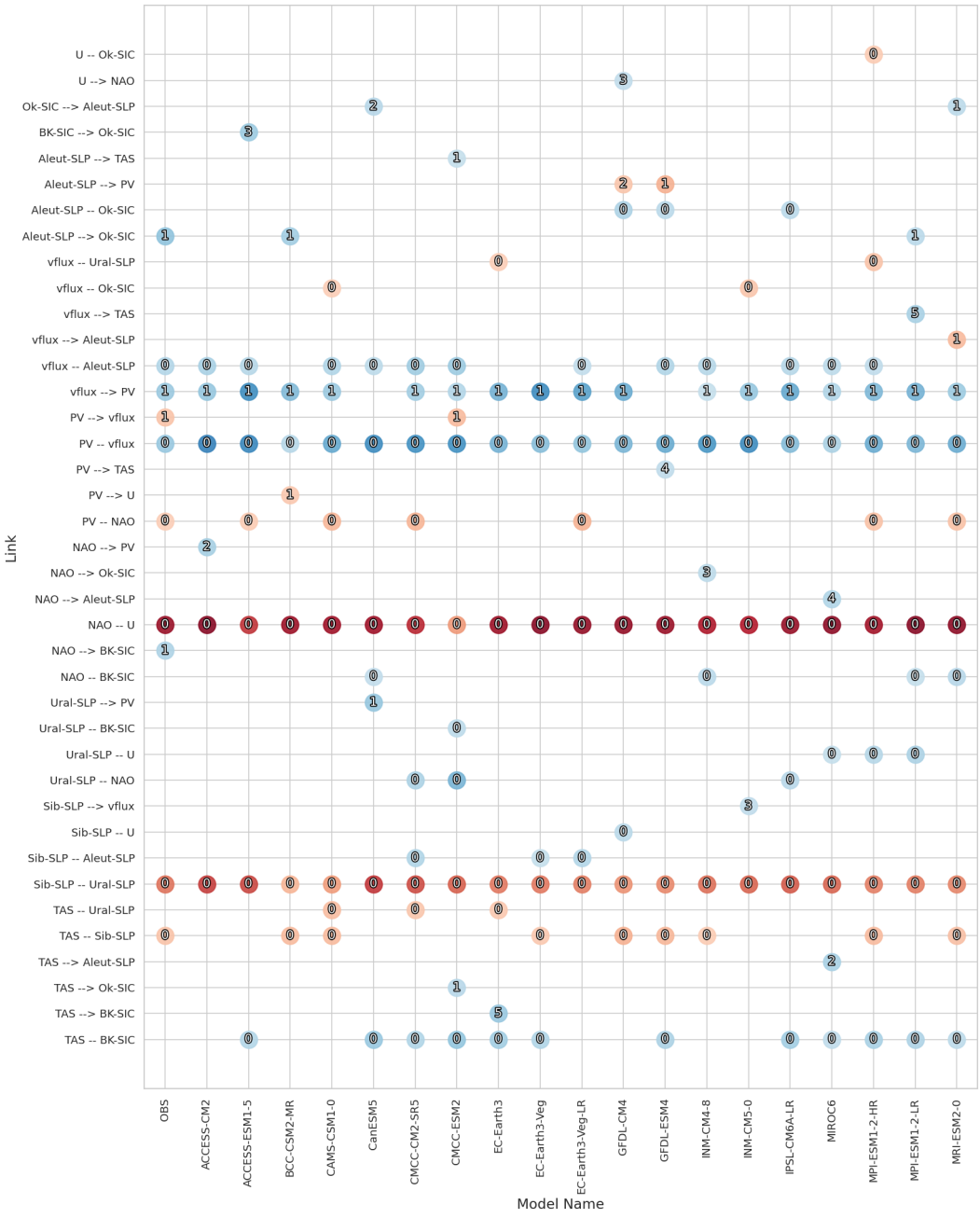

Figure 5. Summary of causal and contemporaneous links detected in observations (OBS) and 19 CMIP6 historical simulations during winters (DJF) for the period 1979-2019. Node color represents the strength of the link. The number within the node represents the time lag in months in causal links, zero represents contemporaneous links.

analysed models (12/19): ACCESS-CM2, ACCESS-ESM1-5, CAMS-CSM1-0, CanESM5, CMCC-CM2-SR5, CMCC-ESM2, EC-Earth-Veg-LR, GFDL-ESM4, INM-CM4-8, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2-HR.

In the following we also discuss the connections that were not found in observations, but were simulated by models. For example, as discussed in Fig. 4a, during DJF the observations do not exhibit a link between TAS and BK-SIC, however based on Fig. 5 we found that most of the analysed models (12/19) simulate this connection (see the lowermost row of Fig. 5). These models are: ACCESS-ESM1-5, CanESM5, CMCC-CM2-SR5, CMCC-ESM2, EC-Earth3, EC-Earth3-Veg, GFDL-ESM4, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2-HR, MPI-ESM1-2-LR, MRI-ESM2-0. Since most of analysed models agree on the connection between TAS and BK-SIC during DJF, the absence of this link in the observational causal graph (Fig. 3a) could occur if the algorithm defines it as false negative (also discussed in Sect. 4.1). We also find that during DJF the link between NAO and BK-SIC is simulated as the negative contemporaneous link (not lagged as in observations) by four models: CanESM5, INM-CM4-8, MPI-ESM1-2-LR, and MRI-ESM2-0. This is the reason why Fig. 4a shows no matches for negative lagged link from NAO to BK-SIC. We also find that during OND most on analysed models (14/19) simulate negative causal link from vflux to PV (see supporting information Fig. S2), which was not detected in observational causal graph on Fig. 4b.

To estimate the general performance of the 19 analysed CMIP6 historical simulations (only first ensemble member r1i1p1f1), we calculate the $F1$-score on the pair-wise comparison of causal graphs, which are based on monthly mean data of the ten analysed actors. Figure 6 shows the matrix of average $F1$-scores during DJF (a), OND (b), and JFM (c). Reference data is shown in rows and data sources that are compared to the references is shown in columns. Higher scores imply better agreement between causal graphs, i.e., that two models (or model and observations) are more similar in terms of their causal fingerprint; lower scores represent larger differences between reference and analysed causal graphs.

Figure 6 shows that generally there is a better agreement among observations and analysed models during DJF (panel a) and JFM (panel c) compared to OND (panel b). The F1-scores during OND (panel b) are notably lower in comparison to the two other periods, implying that the agreement of causal graphs among analysed models and observation is lower during OND. Below we revise, which models have the highest F1-scores during different seasons, thus indicating the best agreement of causal graphs among different models and observations. Below in the round bracket we denote the F1-score values. The highest F1-scores during DJF (Fig. 6a) is detected for EC-Earth3-Veg-LR model, when its causal graph is used as a reference in comparison to ACCESS-CM2 (0.83), ACCESS-ESM1-5, and CMCC-CM2-SR5 (both 0.81). Similar high F1-scores during DJF we found for CMCC-ESM2 and EC-Earth3-Veg-LR (0.81), when CMCC-ESM2 model is considered as a reference. The highest F1-score during OND (see Fig. 6b) is detected between GFDL-CM4 (reference) and GFDL-ESM4 (0.9); and during JFM (see Fig. 6c) between EC-Earth3-Veg (reference) and IPSL-CM6A-LR (0.9). The latter, for example, is one of the highest F1-scores detected among 19 analysed models and observations during analysed periods. Since both models IPSL-CM6A-LR (Boucher et al., 2020) and EC-Earth-Veg (Döscher et al., 2022) share the same ocean component, which is based on the Version 3.6 stable of Nucleus for European Models of the Ocean (NEMO, Madec & Team, 2015), this could explain the high F1-score, and thus the strong similarity in causal graphs. Based on the summary of causal and contemporaneous links shown in Fig. S3, both models simulate analogous links, except that EC-Earth-Veg also captures the observed causal positive links with one month lag from PV to TAS and from PV to vflux, and contemporaneous negative link between Sib- and Aleut-SLP, which are not simulated by IPSL-CM6A-LR.

In the following we also compare the highest F1-scores with observational datasets as a reference during different seasons. We highlight the top three-four models that have the highest F1-scores with observations, thus simulate Arctic-midlatitude links very sim-

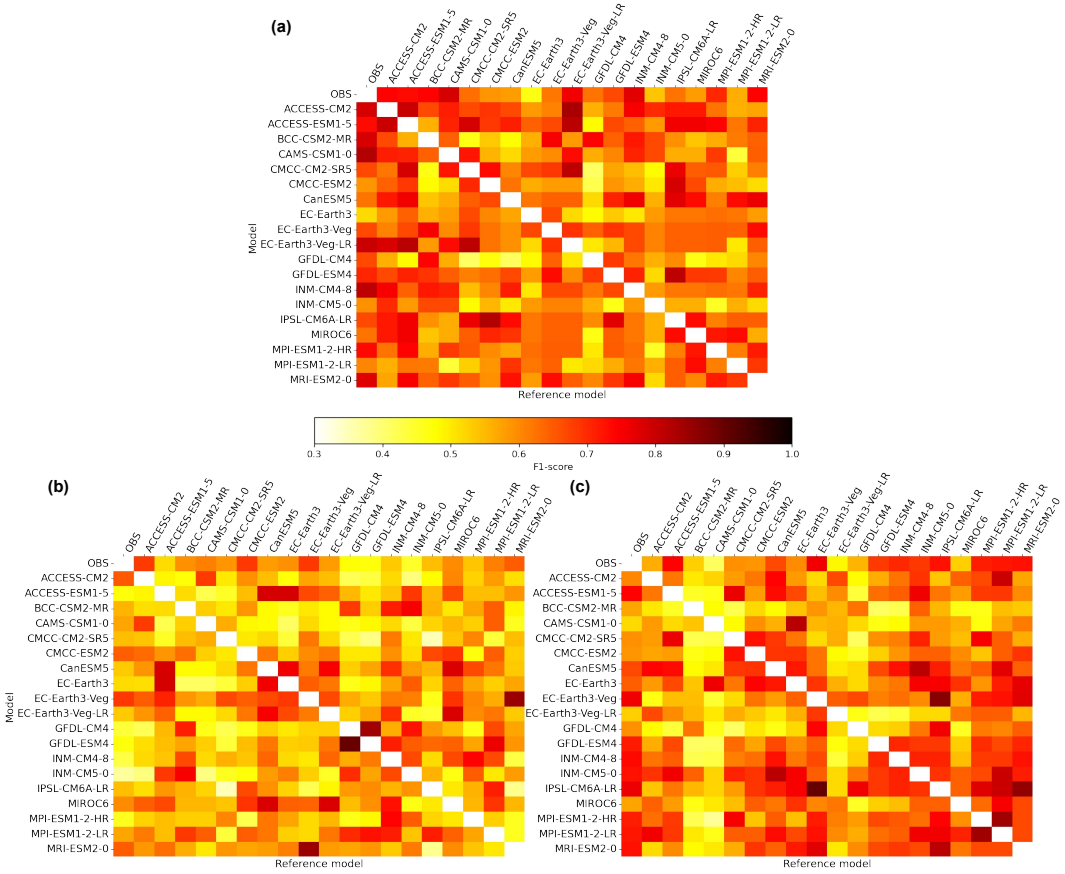

Figure 6. Matrix of average F1-scores for pair-wise comparisons of causal graphs among observations and the first ensemble member (r1i1p1f1) of 19 climate models during (a) DJF, (b) OND, and (c) JFM. Reference data is shown is rows, data that is compared to the reference is shown in columns. Higher (lower) scores stand for better (worse) agreement between causal graphs.

ilar to observations. During DJF, the following models have the best agreement with the observational causal graphs: CAMS-CSM1-0 (0.82), INM-CM4-8 (0.81), and EC-Earth3-Veg-LR (0.8). During OND these models are: EC-Earth3-Veg (0.69), ACCESS-CM2, CMCC-ESM2, and MRI-ESM2-0 (all 0.65). The highest F1-score in comparison to observations during JFM show IPSL-CM6A-LR and EC-Earth3-Veg (both 0.76), and INM-CM4-8 and ACCESS-ESM1-5 (both 0.74). With this here we highlight two points. First, generally high F1-scores suggest that analysed climate models adequately represent observed mechanism behind Arctic-midlatitude connections. In particular, the higher F1-scores during DJF and JFM suggest that climate models better capture observed links (or agree on connections across different models) during these seasons. Thus, second, the high F1-scores between models and/or observations during one season do not imply that the agreement of causal graphs between particular models and/or observations is expected to be good during other seasons. For example, while the F1-score between INM-CM4-8 and observations as reference is one of the highest during DJF (Fig. 6a; 0.81), during OND the similarity in causal graphs is much lower (0.43, panel b), and during JFM it is higher again (0.74, panel c). This shows that the agreement between observations and models (and among models) is highly dependent on the analysed period. Since F1-score does not simply indicate a similarity of individual parameters between different data sources, but is rather a process-based evaluation of model performance (in our case of Arctic-midlatitude connections), we find that these similarities of Arctic-midlatitude processes vary for different analysed periods (in our case OND, DJF, or JFM) across analysed CMIP6 models.

4.3 Causal links in CMIP6 SSP5-8.5 future simulations

To address future changes in Arctic-midlatitude teleconnections, Fig. 7 displays causal graphs similar to Fig. 4, but for the ScenarioMIP SSP5-8.5 simulations for the period 2059-2099. The detailed summary of causal and contemporaneous links detected in CMIP6 SSP5-8.5 simulations for winters 2059-2099 is provided in supporting information Fig. S4. The links detected in the observations and most of the CMIP6 historical simulations are also seen in most of the SSP5-8.5 scenario simulations. Therefore, we first analyse the links that are expected to become more robust based on SSP5-8.5 simulations. The most prominent future changes are found between poleward eddy heat flux and Aleutian Low. The historical simulations showed robustness of this link only during DJF (see Fig. 4a, 12/19 models), while during OND only 6/19 models and during JFM 9/19 models simulated this connection. During 2059-2099 it is expected that this connection will become more robust for all analysed periods: for DJF and JFM it is simulated by 16/19 models and for OND by 13/19 models(see Fig. 7a-c). This pronounced change from past to future could be associated with future intensification of extreme Aleutian low reported by Giamalaki et al. (2021), since the Aleutian low has previously experienced a weakening trend during the past two decades (Hu et al., 2018).

We also find that the contemporaneous connection between TAS and BK-SIC becomes more robust from historical to SSP5-8.5 simulations. Figure 5 (lowermost row) shows that during 1979-2019 12/19 models simulate this connection in DJF, although it was not found in the observations. In the future this link is simulated by 15/19 models during DJF (see supporting information Fig. S5). In historical simulations the link between TAS and BK-SIC is found in 13/19 models during OND and in 14/19 models during JFM. Future simulations show more robust link between TAS and BK-SIC, i.e. in 16/19 models both during OND and JFM (see Fig. 7b,c).

Several links become less robust in the future SSP5-8.5 simulations. For example, during DJF the contemporaneous positive link between TAS and Sib-SLP is simulated by 8/19 models in historical (BCC-CSM2-MR, CAMS-CSM1-0, EC-Earth3-Veg, GFDL-CM4, GFDL-ESM4, INM-CM4-8, MPI-ESM1-2-HR, MRI-ESM2-0, Fig. 4a and 5) and by 5/19 in SSP5-8.5 simulations (CAMS-CSM1-0, CMCC-ESM2, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2-0, see Fig. 7a and S5). Note that they are not identical mod-

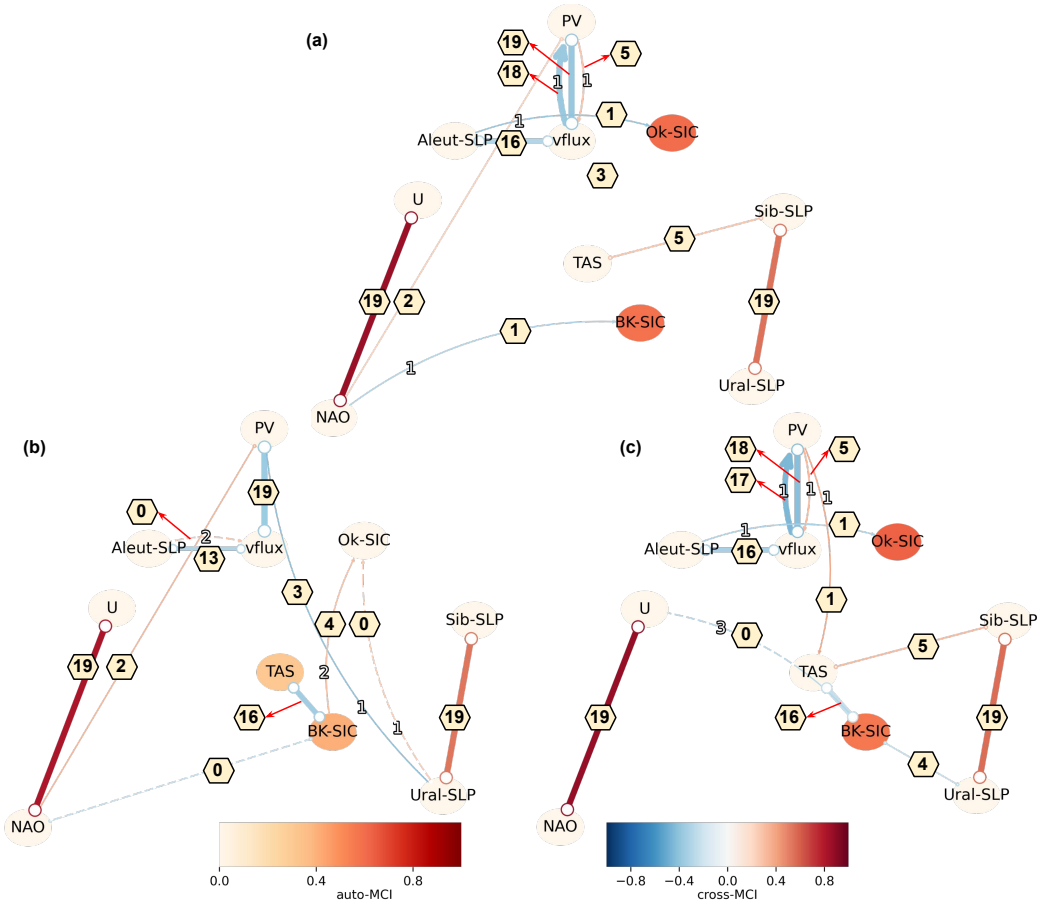

Figure 7. Same as Fig. 4 but for the ScenarioMIP SSP5-8.5 simulations for the period 2059-2099.

els that detect this link by historical and future simulations. During OND there are two connections, that become less robust from historical to future simulations, and both are related to PV. The first one is the link between NAO and PV, which is simulated by 6/19 models in historical simulations (ACCESS-CM2, CAMS-CSM1-0, CMCC-CM2-SR5, EC-Earth-Veg, MPI-ESM1-2-HR, and MRI-ESM2-0, see Fig. 4b and S2) and by only 2 models in SSP5-8.5 (CMCC-ESM2, EC-Earth3-Veg-LR, see Fig. 7b and S6). The second link is from Ural-SLP to PV, which is simulated by 5/19 in historical (BCC-CSM2-MR, CMCC-CM2-SR5, CMCC-ESM2, EC-Earth3-Veg, MRI-ESM2-0) and 3/19 in SSP5-8.5 (CAMS-CSM1-0, CMCC-CM2-SR5, EC-Earth3). During JFM it is the causal link from PV to TAS, which is simulated by 3/19 models in historical (BCC-CSM2-MR, EC-Earth3-Veg, EC-Earth3-Veg-LR, see Fig. 4c and S3) and by 1/19 models in future simulations (EC-Earth3-Veg-LR, see Fig. 7 and Fig. S7).

To summarize the difference between observed and modelled causal graphs in past and future simulations, Fig. 8 illustrates the F1-score for historical (circles) and SSP5-8.5 (triangles) simulations for DJF (panel a, blue), OND (panel b, green), and JFM (panel c, orange). Similar to Fig. 6, we use observational causal graphs as reference for the analysis of CMIP6 historical simulations. To estimate the impact of a changing climate on Arctic-midlatitude links, we choose the historical simulation as a reference for each CMIP6 model for the future scenarios. The lower the F1-score for SSP5-8.5 simulations, the larger is the difference between the causal graphs for the historical and future simulations. We find that the F1-scores for the future simulations are generally higher during DJF and JFM (see Fig. 8a, c), and lower during OND (Fig. 8b), which is similar to the findings for historical F1-scores shown in Fig. 6a-c. Most of SSP5-8.5 F1-scores are similar to the historical scores, implying that causal graphs between historical and SSP5-8.5 simulations do not differ much. This is associated with the small future changes in the context of Arctic-midlatitude links (or absence of drastic complex future changes) as simulated under the SSP5-8.5 scenario.

Those models that have the highest F1-score for the SSP5-8.5 simulations (triangles in Fig. 8) predict only small future changes in comparison with historical simulations. During DJF (panel a) these models are CMCC-CM2-SR5 (0.83), EC-Earth3-Veg-LR (0.79), and IPSL-CM6A-LR (0.76); during OND (panel b) these are CanESM5 and MIROC6 (both 0.73), and during JFM (panel c) these are MPI-ESM1-2-LR (0.85), INM-CM5-0 and CanESM5 (both 0.79), and MPI-ESM1-2-HR (0.78). Therefore, these are the models that predict the smallest change in Arctic-midlatitude connections in comparison to their historical simulations during analysed periods.

The lowest F1-scores in SSP5-8.5 simulations, thus the largest differences between past and future Arctic-midlatitude teleconnections are simulated during DJF by GFDL-CM4 (0.41), BCC-CSM2-MR (0.47), MIROC6 and INM-CM5-0 (both 0.5). During OND these models are CAMS-CSM1-0 (0.41), CMCC-CM2-SR5 EC-Earth3-Veg-LR (both 0.42), and GFDL-CM4 (0.43). And during JFM these models are BCC-CSM2-MR (0.4), GFDL-CM4 (0.45), and CAMS-CSM1-0 (0.46). Thus, GFDL-CM4 is the model that simulates the largest future changes in comparison to historical Arctic-midlatitude links for all three analysed periods. For example, based on the DJF causal graphs from GFDL-CM4 historical and SSP5-8.5 simulations (see supporting information Fig. S1 and S4), in the future this model simulates a contemporaneous impact of Ural-SLP on TAS via BK-SIC; a causal negative link from BK-SIC to Ok-SIC with five months lag; a causal negative link from Ural-SLP to PV with one month lag. None of these links were found in historical simulations of GFDL-CM4 model.

Additionally, the F1-score is useful to estimate how different versions of the same model reproduce observational physical mechanisms. For example, based on the mechanisms of Arctic-midlatitude teleconnections, we find that the MPI-ESM1-2 models (HR and LR) during JFM have similar F1-scores with each other (see Fig. 6c, 0.84), moreover both models agree well with observational causal graph (Fig. 8c, around 0.72). During DJF MPI-ESM1-2-HR better captures observed connections, thus MPI-ESM1-2-HR F1-score is higher (0.73) than MPI-ESM1-2-LR (0.6); and during OND MPI-ESM1-2-

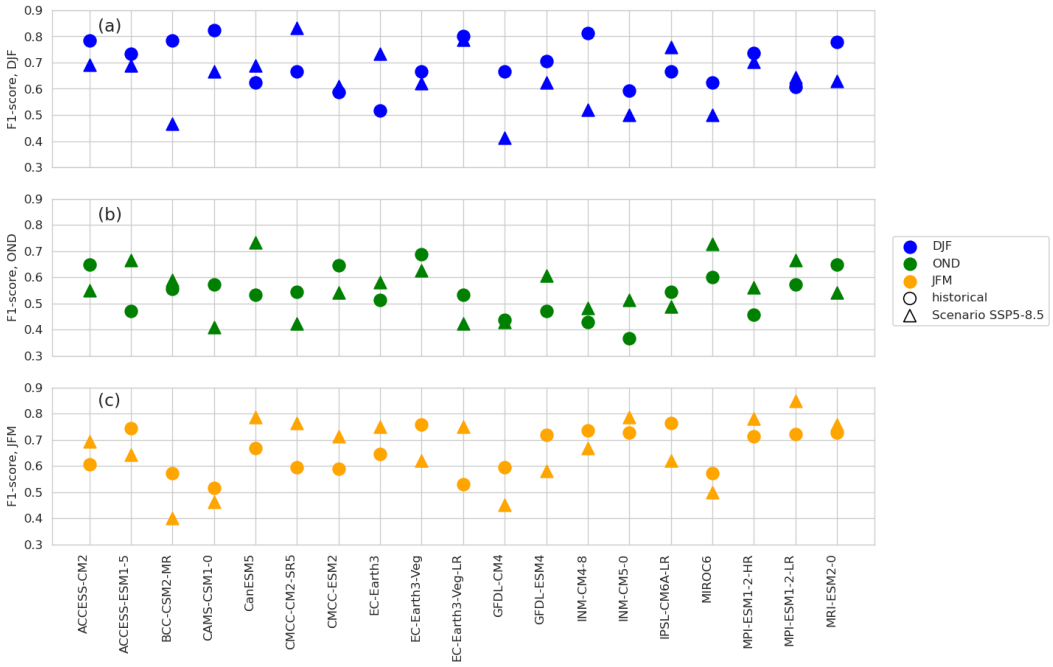

Figure 8. *F*1-score scatter plot of individual causal network comparison with observations taken as reference for the first ensemble member (r1i1p1f1) of 19 CMIP6 historical (circles) and SSP5-8.5 (triangles) simulations for winter DJF (panel a, blue), early winter OND (panel b, green), and late winter JFM (panel c, orange).

LR simulates less links, which leads to lower number of false positive and false negative connections, thus its F1-score is higher (0.57) than MPI-ESM1-2-HR (0.46).

5 Conclusions

Our study applies causal discovery to compare Arctic-midlatitude teleconnections across a number of CMIP6 simulations to the causal links derived from observations. Causal discovery is recognized as an interpretable machine learning tool that goes beyond correlation analysis to study complex systems. By analysing detrended timeseries from observations (ERA5 and HadISST) and 19 CMIP6 models, this study shows the variety of causal and contemporaneous links that were identified in the Arctic-midlatitude teleconnections in conditions of amplified Arctic warming by CMIP6 historical and SSP5-8.5 simulations. To identify the differences in the mechanism of Arctic-midlatitude teleconnections during the NH cold season, we apply causal discovery to the following periods: winter (DJF), early winter (OND), and late winter (JFM).

Based on the application of causal discovery on observational data for 1979-2019, we detect links between Arctic and midlatitude processes that occur not only in the troposphere, but also in the stratosphere. For example, in observations OND and JFM we find a negative contemporaneous link between near-surface Arctic temperature and sea ice over Barents and Kara seas, implying that an increase (decrease) in temperature is connected to a decline (increase) of sea ice over Barents and Kara seas. Furthermore, the increase of Arctic temperature is also associated with atmospheric blocking at midlatitudes, such as a strengthening of the Siberian High, which is found during DJF and JFM. Observational causal graphs also detect a positive connection between near-surface westerlies and NAO: weak (strong) westerly winds are associated with the negative (pos-

itive) phase of the NAO. The latter is related to the weakened stratospheric PV during OND and DJF. Further analysis of observations has revealed that there are several teleconnections that occur only during specific periods. For example, the direct link of increased Ural blocking with weakened stratospheric PV is detected only during OND; the causal impact of a weakened PV on the decline of Arctic near-surface temperature, which is associated with sudden stratospheric warming events was found only during JFM.

The causal analysis of CMIP6 historical simulations showed good agreement with observations from early to late winters, since most of the analysed models simulate the observed links. Based on historical simulations, we found several links that were simulated by most of the models, but were not detected in the observations. For example, during OND most of models simulate that the increased poleward eddy heat flux induces weakening of the polar vortex with one month lag. During DJF most of analysed models simulate a negative contemporaneous link between Arctic temperature and Barents-Kara sea ice. However these links were not detected in the observations. However, the observational causal graph could be estimated with errors, thus the observational link between these actors might be detected as false negative. We also found a link that was detected in the observations, but was not simulated by any of analysed models, namely during DJF the negative lagged connection from NAO to the sea ice over Barents and Kara seas. During OND models do not simulate observed contemporaneous link between NAO and sea ice changes over Barents and Kara seas, the causal positive connection from Ural blocking to the sea ice over the Sea of Okhotsk, and the causal positive link from the Aleutian Low to the poleward eddy heat flux. And during JFM analysed models do not simulate the causal negative link from the near-surface westerly winds to the Barents-Kara sea ice cover.

The analysis of CMIP6 SSP5-8.5 simulations during 2059-2099 does not show drastic changes in future Arctic-midlatitude teleconnections, since most of the analysed models reproduce the detected links. Additionally, the analysis based on the F1-score supports the conclusion that the expected future changes will be small, since the differences between the historical and SSP5-8.5 F1-scores are not large.

However, we find several connections that are clearly expected to become more robust in the future. For example, observational datasets detected a negative link between near-surface Arctic temperature and sea ice over the Barents and Kara seas during OND and JFM, and most of the analysed historical models simulated this connection during all considered periods. Furthermore, future simulations suggest that this connection becomes more robust during 2059-2099: 15 out of 19 analysed models show this connection in the future simulations during DJF, and 16 out of 19 models during OND and JFM. While in the historical simulations the link between Arctic temperature and sea ice over the Barents and Kara seas was simulated by 12 out 19 models during DJF, 13 out of 19 models during OND, and 14 out of 19 during JFM.

We also find a pronounced change between past and future simulations associated with a robust detection of a link between the Aleutian Low and poleward eddy heat flux at 100 hPa, which is detected in most of the historical simulations only during DJF. Most of the analysed models simulate this connection in the future: 13 out of 19 models during OND, and 16 out of 19 during DJF and JFM. This change could be associated with predicted future intensification of the Aleutian Low (Giamalaki et al., 2021).

Overall, this study shows that causal model evaluation of Arctic-midlatitude teleconnections provides not only a powerful tool to assess the physical mechanisms and their interconnections in climate models, but also yields robust information for the research of observed atmospheric processes and their changes.

6 Open Research

6.1 Availability statement

The CMIP6 model data is publicly available upon registration via the Earth System Grid Federation (ESGF, `https://esgf.llnl.gov`, last access: 05.10.2022, Williams et al., 2009). ERA5 data is available via Climate Data Store (CDS, `https://cds.climate.copernicus.eu/`, last access: 05.10.2022.

ESMValTool v2.0 is released under the Apache License, VERSION 2.0 (Eyring et al., 2020; Lauer et al., 2020; Righi et al., 2020). The latest release of ESMValTool v2 is publicly available on Zenodo at `https://doi.org/10.5281/zenodo.3401363` (Andela et al., 2022b). The source code of ESMValCore package, which is installed as a dependency of ESMValTool v2, is also publicly available on Zenodo at `https://doi.org/10.5281/zenodo.3387139` (Andela et al., 2022a). Both ESMValTool and ESMValCore are developed on the GitHub repositories available at `https://github.com/ESMValGroup`, last access: 05.10.2022. At the time of publication of this paper, we will implement a new recipe to ESMValTool that can be used to reproduce the calculations of potential actors provided in this study.

The causal discovery algorithm is freely available via `https://github.com/jakobrunge/tigramite`, last access: 05.10.2022. The code used to reproduce summary of causal and contemporaneous links and F1-scores and to produce figures of this manuscript will be accessible at the time of publication of the manuscript in the following Github repository: `https://github.com/EyringMLClimateGroup/galytska22jgr_CME_arctic-midlat_teleconnections`.

Acknowledgments

Funding for this study was provided by the European Research Council (ERC) Synergy Grant "Understanding and Modelling the Earth System with Machine Learning (USMILE)" under the Horizon 2020 research and innovation programme (Grant agreement No. 855187) and the "Advanced Earth System Model Evaluation for CMIP (EVal4CMIP)" project funded by the Helmholtz Society. J.R. has received funding from the European Research Council (ERC) Starting Grant CausalEarth under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 948112). We acknowledge the World Climate Research Program's (WCRP's) Working Group on Coupled Modelling (WGCM), which is responsible for CMIP, and we thank the climate modelling groups listed in Table 1 for producing and making available their model output. This work used the computational resources of the Deutsches Klimarechenzentrum (DKRZ, Germany) granted granted by its Scientific Steering Committee (WLA) under project ID bd1083.

Author contribution. EG performed the analysis, prepared all figures, and lead the writing of the manuscript. All co-authors commented on the initial and revised drafts of the manuscripts. KW contributed to the scientific interpretation of the results. JR provided methodology and the expertise on application of the method. DH, RJ, and RK contributed to the discussions of the topic. VE initiated and coordinated the study and contributed to the scientific interpretation of the results.

References

Ambaum, M. H., Hoskins, B. J., & Stephenson, D. B. (2001). Arctic oscillation or north atlantic oscillation? *Journal of Climate*, *14*(16), 3495–3507.

Andela, B., Broetz, B., de Mora, L., Drost, N., Eyring, V., Koldunov, N., ... Kazeroni, R. (2022a, June). *Esmvalcore.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.6778167` doi: 10.5281/zenodo.6778167

Andela, B., Broetz, B., de Mora, L., Drost, N., Eyring, V., Koldunov, N., ... Kazeroni, R. (2022b, March). *Esmvaltool.* Zenodo. Retrieved from

https://doi.org/10.5281/zenodo.6359405 doi: 10.5281/zenodo.6359405

Andrews, D., & Mcintyre, M. E. (1976). Planetary waves in horizontal and vertical shear: The generalized eliassen-palm relation and the mean zonal acceleration. *Journal of the Atmospheric Sciences*, *33*(11), 2031–2048.

Athanasiadis, P. J., Wallace, J. M., & Wettstein, J. J. (2010). Patterns of wintertime jet stream variability and their relation to the storm tracks. *Journal of the Atmospheric Sciences*, *67*(5), 1361–1381.

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., ... others (2021). Sudden stratospheric warmings. *Reviews of Geophysics*, *59*(1), e2020RG000708.

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584.

Barnes, E. A., & Screen, J. A. (2015). The impact of arctic warming on the midlatitude jet-stream: Can it? has it? will it? *Wiley Interdisciplinary Reviews: Climate Change*, *6*(3), 277–286.

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Sullivan, A., Bodman, R., ... others (2020). Configuration and spin-up of access-cm2, the new generation australian community climate and earth system simulator coupled model. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 225–251.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... others (2020). Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., ... others (2019). Global mean climate and main patterns of variability in the cmcc-cm2 coupled model. *Journal of Advances in Modeling Earth Systems*, *11*(1), 185–209.

Cohen, J., & Jones, J. (2011). A new index for more accurate winter predictions. *Geophysical Research Letters*, *38*(21).

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., ... others (2014). Recent arctic amplification and extreme mid-latitude weather. *Nature geoscience*, *7*(9), 627–637.

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., ... others (2020). Divergent consensuses on arctic amplification influence on midlatitude severe winter weather. *Nature Climate Change*, *10*(1), 20–29.

Di Capua, G., Runge, J., Donner, R. V., van den Hurk, B., Turner, A. G., Vellore, R., ... Coumou, D. (2020). Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: causal relationships and the role of timescales. *Weather and Climate Dynamics*, *1*(2), 519–539.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., ... Zhang, Q. (2022). The ec-earth3 earth system model for the coupled model intercomparison project 6. *Geoscientific Model Development*, *15*(7), 2973–3020. Retrieved from https://gmd.copernicus.org/articles/15/2973/2022/ doi: 10.5194/gmd-15-2973-2022

Dunne, J., Horowitz, L., Adcroft, A., Ginoux, P., Held, I., John, J., ... others (2020). The gfdl earth system model version 4.1 (gfdl-esm 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS002015.

Ebert-Uphoff, I., & Deng, Y. (2012). Causal discovery for climate research using graphical models. *Journal of Climate*, *25*(17), 5648–5665.

England, M. R., Eisenman, I., Lutsko, N. J., & Wagner, T. J. (2021). The recent emergence of arctic amplification. *Geophysical Research Letters*, *48*(15), e2021GL094086.

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., ... others (2020). Earth system model evaluation tool (esmvaltool) v2. 0–an extended

set of large-scale diagnostics for quasi-operational and comprehensive evaluation of earth system models in cmip. *Geoscientific Model Development*, *13*(7), 3383–3438.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., ... others (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110.

Francis, J. A. (2017). Why are arctic linkages to extreme weather still up in the air? *Bulletin of the American Meteorological Society*, *98*(12), 2551–2557.

Francis, J. A., & Vavrus, S. J. (2012). Evidence linking arctic amplification to extreme weather in mid-latitudes. *Geophysical research letters*, *39*(6).

Giamalaki, K., Beaulieu, C., Henson, S., Martin, A., Kassem, H., & Faranda, D. (2021). Future intensification of extreme aleutian low events and their climate impacts. *Scientific reports*, *11*(1), 1–12.

Gier, B. K., Buchwitz, M., Reuter, M., Cox, P. M., Friedlingstein, P., & Eyring, V. (2020). Spatially resolved evaluation of earth system models with satellite column-averaged co 2. *Biogeosciences*, *17*(23), 6115–6144.

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., ... others (2018). Quantifying climate feedbacks in polar regions. *Nature communications*, *9*(1), 1–13.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Graversen, R. G., & Wang, M. (2009). Polar amplification in a coupled climate model with locked albedo. *Climate Dynamics*, *33*(5), 629–643.

Hall, R. J., Mitchell, D. M., Seviour, W. J., & Wright, C. J. (2022). How well are sudden stratospheric warming surface impacts captured in cmip6 climate models? *Journal of Geophysical Research: Atmospheres*, e2021JD035725.

Hall, R. J., Scaife, A. A., Hanna, E., Jones, J. M., & Erdélyi, R. (2017). Simple statistical probabilistic forecasts of the winter nao. *Weather and Forecasting*, *32*(4), 1585–1601.

Held, I., Guo, H., Adcroft, A., Dunne, J., Horowitz, L., Krasting, J., ... others (2019). Structure and performance of gfdl's cm4. 0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... others (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Honda, M., Inoue, J., & Yamane, S. (2009). Influence of low arctic sea-ice minima on anomalously cold eurasian winters. *Geophysical Research Letters*, *36*(8).

Hoshi, K., Ukita, J., Honda, M., Iwamoto, K., Nakamura, T., Yamazaki, K., ... Handorf, D. (2017). Poleward eddy heat flux anomalies associated with recent arctic sea ice loss. *Geophysical Research Letters*, *44*(1), 446–454.

Hu, D., Guan, Z., Tian, W., & Ren, R. (2018). Recent strengthening of the stratospheric arctic vortex response to warming in the central north pacific. *Nature Communications*, *9*(1), 1–10.

Hurrell, J. W., & Deser, C. (2010). North atlantic climate variability: the role of the north atlantic oscillation. *Journal of marine systems*, *79*(3-4), 231–244.

Hwang, S.-O., Yeh, S.-W., Oh, S.-Y., & Lee, J. (2022). Recent weakening linkage between arctic oscillation and aleutian low during boreal winter and its impact on surface temperature over eastern eurasia. *Atmospheric Science Letters*, e1089.

IPCC. (2021). Summary for policymakers [Book Section]. In V. Masson-Delmotte

et al. (Eds.), *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (p. 332). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/9781009157896.001

Jaiser, R., Nakamura, T., Handorf, D., Dethloff, K., Ukita, J., & Yamazaki, K. (2016). Atmospheric winter response to arctic sea ice changes in reanalysis data and model simulations. *Journal of Geophysical Research: Atmospheres*, *121*(13), 7564–7577.

Karmouche, S., Galytska, E., Runge, J., Meehl, G. A., Phillips, A. S., Weigel, K., & Eyring, V. (2022). Regime-oriented causal model evaluation of atlantic-pacific teleconnections in cmip6. *EGUsphere*, *2022*, 1–42. Retrieved from `https://egusphere.copernicus.org/preprints/egusphere-2022-1013/` doi: 10.5194/egusphere-2022-1013

Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., ... Yoon, J.-H. (2014). Weakening of the stratospheric polar vortex by arctic sea-ice loss. *Nature communications*, *5*(1), 1–8.

Koenigk, T., Key, J., & Vihma, T. (2020). Climate change in the arctic. In *Physics and chemistry of the arctic atmosphere* (pp. 673–705). Springer.

Kolstad, E., & Screen, J. (2019). Nonstationary relationship between autumn arctic sea ice and the winter north atlantic oscillation. *Geophysical Research Letters*, *46*(13), 7583–7591.

Kretschmer, M., Adams, S. V., Arribas, A., Prudden, R., Robinson, N., Saggioro, E., & Shepherd, T. G. (2021). Quantifying causal pathways of teleconnections. *Bulletin of the American Meteorological Society*, *102*(12), E2247–E2263.

Kretschmer, M., Cohen, J., Matthias, V., Runge, J., & Coumou, D. (2018). The different stratospheric influence on cold-extremes in eurasia and north america. *npj Climate and Atmospheric Science*, *1*(1), 1–10.

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *Journal of climate*, *29*(11), 4069–4081.

Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., ... others (2020). Earth system model evaluation tool (esmvaltool) v2. 0–diagnostics for emergent constraints and future projections from earth system models in cmip. *Geoscientific Model Development*, *13*(9), 4205–4228.

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., ... others (2022). Cmip6 simulations with the cmcc earth system model (cmcc-esm2). *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002814.

Luo, D., Xiao, Y., Yao, Y., Dai, A., Simmonds, I., & Franzke, C. L. (2016). Impact of ural blocking on winter warm arctic–cold eurasian anomalies. part i: Blocking-induced amplification. *Journal of Climate*, *29*(11), 3925–3947.

Madec, G., & Team, N. S. (2015). *Nemo ocean engine* (Tech. Rep. No. 27). Institut Pierre-Simon Laplace (IPSL). doi: 10.5281/zenodo.1464816

Marshall, J., Kushnir, Y., Battisti, D., Chang, P., Czaja, A., Dickson, R., ... Visbeck, M. (2001). North atlantic climate variability: phenomena, impacts and mechanisms. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *21*(15), 1863–1898.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... others (2019). Developments in the mpi-m earth system model version 1.2 (mpi-esm1. 2) and its response to increasing co2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038.

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... Marotzke, J. (2018). A higher-resolution version of the max planck institute earth system model (mpi-esm1.2-hr). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383-1413. Retrieved from `https://`

agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017MS001217 doi: https://doi.org/10.1029/2017MS001217

Nakamura, T., Yamazaki, K., Iwamoto, K., Honda, M., Miyoshi, Y., Ogawa, Y., & Ukita, J. (2015). A negative phase shift of the winter ao/nao due to the recent arctic sea-ice reduction in late autumn. *Journal of Geophysical Research: Atmospheres*, *120*(8), 3209–3227.

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

Ogi, M., Taguchi, B., Honda, M., Barber, D. G., & Rysgaard, S. (2015). Summer-to-winter sea-ice linkage between the arctic ocean and the okhotsk sea through atmospheric circulation. *Journal of Climate*, *28*(12), 4971–4979.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... others (2016). The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*, *9*(9), 3461–3482.

Overland, J., Francis, J. A., Hall, R., Hanna, E., Kim, S.-J., & Vihma, T. (2015). The melting arctic and midlatitude weather patterns: Are they connected? *Journal of Climate*, *28*(20), 7917–7932.

Pedersen, R. A., Cvijanovic, I., Langen, P. L., & Vinther, B. M. (2016). The impact of regional arctic sea ice loss on atmospheric circulation and the nao. *Journal of Climate*, *29*(2), 889–902.

Peings, Y. (2019). Ural blocking as a driver of early-winter stratospheric warmings. *Geophysical Research Letters*, *46*(10), 5460–5468.

Phillips, A. S., Deser, C., & Fasullo, J. (2014). Evaluating modes of variability in climate models. *Eos, Transactions American Geophysical Union*, *95*(49), 453–455.

Pithan, F., & Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature geoscience*, *7*(3), 181–184.

Polkova, I., Afargan-Gerstman, H., Domeisen, D. I., King, M. P., Ruggieri, P., Athanasiadis, P., ... Baehr, J. (2021). Predictors and prediction skill for marine cold-air outbreaks over the barents sea. *Quarterly Journal of the Royal Meteorological Society*, *147*(738), 2638–2656.

Previdi, M., Janoski, T. P., Chiodo, G., Smith, K. L., & Polvani, L. M. (2020). Arctic amplification: A rapid response to radiative forcing. *Geophysical Research Letters*, *47*(17), e2020GL089933.

Previdi, M., Smith, K. L., & Polvani, L. M. (2021). Arctic amplification of climate change: a review of underlying mechanisms. *Environmental Research Letters*.

Rantanen, M., Karpechko, A., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., ... Laaksonen, A. (2021). The arctic has warmed four times faster than the globe since 1980.

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

Riebold, J., Richling, A., Ulbrich, U., Rust, H., Semmler, T., & Handorf, D. (2022). On the linkage between future arctic sea ice retreat, euro-atlantic circulation regimes and temperature extremes over europe. *EGUsphere*, *2022*, 1–32. Retrieved from https://egusphere.copernicus.org/preprints/egusphere-2022-953/ doi: 10.5194/egusphere-2022-953

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., ... others (2020). Earth system model evaluation tool (esmvaltool) v2. 0–technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199.

Rong, X., Li, J., Chen, H., Su, J., Hua, L., Zhang, Z., & Xin, Y. (2021). The cmip6 historical simulation datasets produced by the climate system model cams-csm. *Advances in Atmospheric Sciences*, *38*(2), 285–295.

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.

Runge, J. (2020). Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on uncertainty in artificial intelligence* (pp. 1388–1397).

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., ... others (2019). Inferring causation from time series in earth system sciences. *Nature communications*, *10*(1), 1–13.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances*, *5*(11), eaau4996.

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., ... Kurths, J. (2015). Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature communications*, *6*(1), 1–10.

Runge, J., Petoukhov, V., & Kurths, J. (2014). Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *Journal of climate*, *27*(2), 720–739.

Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., ... others (2014). Skillful long-range prediction of european and north american winters. *Geophysical Research Letters*, *41*(7), 2514–2519.

Screen, J. A. (2017a). Far-flung effects of arctic warming. *Nature Geoscience*, *10*(4), 253–254.

Screen, J. A. (2017b). The missing northern european winter cooling response to arctic sea ice loss. *Nature communications*, *8*(1), 1–9.

Screen, J. A., Deser, C., & Simmonds, I. (2012). Local and remote controls on observed arctic warming. *Geophysical Research Letters*, *39*(10).

Screen, J. A., Deser, C., Smith, D. M., Zhang, X., Blackport, R., Kushner, P. J., ... Sun, L. (2018). Consistency and discrepancy in the atmospheric response to arctic sea-ice loss across climate models. *Nature Geoscience*, *11*(3), 155–163.

Screen, J. A., & Simmonds, I. (2010). The central role of diminishing sea ice in recent arctic temperature amplification. *Nature*, *464*(7293), 1334–1337.

Serreze, M. C., Barrett, A., Stroeve, J., Kindig, D., & Holland, M. (2009). The emergence of surface-based arctic amplification. *The Cryosphere*, *3*(1), 11–19.

Serreze, M. C., & Barry, R. G. (2011). Processes and impacts of arctic amplification: A research synthesis. *Global and planetary change*, *77*(1-2), 85–96.

Shepherd, T. G. (2016). Effects of a warming arctic. *Science*, *353*(6303), 989–990.

Siew, P. Y. F., Li, C., Sobolowski, S. P., & King, M. P. (2020). Intermittency of arctic–mid-latitude teleconnections: stratospheric pathway between autumn sea ice and the winter north atlantic oscillation. *Weather and Climate Dynamics*, *1*(1), 261–275.

Silva, F. N., Vega-Oliveros, D. A., Yan, X., Flammini, A., Menczer, F., Radicchi, F., ... Fortunato, S. (2021). Detecting climate teleconnections with granger causality. *Geophysical Research Letters*, *48*(18), e2021GL094707.

Smith, D. M., Eade, R., Andrews, M., Ayres, H., Clark, A., Chripko, S., ... others (2022). Robust but weak winter atmospheric circulation response to future arctic sea ice loss. *Nature communications*, *13*(1), 1–15.

Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, *9*(1), 62–72.

Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.

Spirtes, P., & Zhang, K. (2016). Causal discovery and inference: concepts and recent methodological advances. In *Applied informatics* (Vol. 3, pp. 1–28).

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... others (2019). The canadian earth system model version 5 (canesm5. 0.3).

Geoscientific Model Development, *12*(11), 4823–4873.

Tachibana, Y., Honda, M., & Takeuchi, K. (1996). The abrupt decrease of the sea ice over the southern part of the sea of okhotsk in 1989 and its relation to the recent weakening of the aleutian low. *Journal of the Meteorological Society of Japan. Ser. II*, *74*(4), 579–584.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... others (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in miroc6. *Geoscientific Model Development*, *12*(7), 2727–2765.

Trenberth, K. E., & Hurrell, J. W. (1994). Decadal atmosphere-ocean variations in the pacific. *Climate Dynamics*, *9*(6), 303–319.

Tyrlis, E., Bader, J., Manzini, E., Ukita, J., Nakamura, H., & Matei, D. (2020). On the role of ural blocking in driving the warm arctic–cold siberia pattern. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 2138–2153.

Vavrus, S. J. (2018). The influence of arctic amplification on mid-latitude weather and climate. *Current Climate Change Reports*, *4*(3), 238–249.

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., ... others (2018). Simulation of the modern climate using the inm-cm48 climate model. *Russian Journal of Numerical Analysis and Mathematical Modelling*, *33*(6), 367–374.

Vorobyeva, V., & Volodin, E. (2021). Evaluation of the inm ras climate model skill in climate indices and stratospheric anomalies on seasonal timescale. *Tellus A: Dynamic Meteorology and Oceanography*, *73*(1), 1–12.

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly weather review*, *109*(4), 784–812.

Walsh, J. E. (2014). Intensified warming of the arctic: Causes and impacts on middle latitudes. *Global and Planetary Change*, *117*, 52–63.

Wang, L., Ting, M., & Kushner, P. (2017). A robust empirical seasonal prediction of winter nao and surface climate. *Scientific reports*, *7*(1), 1–9.

Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., ... others (2021). Earth system model evaluation tool (esmvaltool) v2. 0–diagnostics for extreme events, regional and impact evaluation, and analysis of earth system models in cmip. *Geoscientific Model Development*, *14*(6), 3159–3184.

Wendisch, M., Brückner, M., Crewell, S., Ehrlich, A., Notholt, J., Lüpkes, C., ... others (2022). Atmospheric and surface processes, and feedback mechanisms determining arctic amplification: A review of first results and prospects of the (ac) 3 project. *Bulletin of the American Meteorological Society*.

Williams, D. N., Ananthakrishnan, R., Bernholdt, D., Bharathi, S., Brown, D., Chen, M., ... others (2009). The earth system grid: Enabling access to multimodel climate simulation data. *Bulletin of the American Meteorological Society*, *90*(2), 195–206.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., ... Liu, X. (2019). The beijing climate center climate system model (bcc-csm): the main progress from cmip5 to cmip6. *Geoscientific Model Development*, *12*(4), 1573–1600. Retrieved from `https://gmd.copernicus.org/articles/12/1573/2019/` doi: 10.5194/gmd-12-1573-2019

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., & Döscher, R. (2020). On the increased climate sensitivity in the ec-earth model from cmip5 to cmip6. *Geoscientific Model Development*, *13*(8), 3465–3474. Retrieved from `https://gmd.copernicus.org/articles/13/3465/2020/` doi: 10.5194/gmd-13-3465-2020

Xu, G., Duong, T. D., Li, Q., Liu, S., & Wang, X. (2020). Causality learning: a new perspective for interpretable machine learning. *arXiv preprint arXiv:2006.16789*.

Yao, Y., Luo, D., & Zhong, L. (2018). Effects of northern hemisphere atmospheric blocking on arctic sea ice decline in winter at weekly time scales. *Atmosphere*, *9*(9), 331.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., ... others (2019). The meteorological research institute earth system model version 2.0, mri-esm2. 0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Ser. II*.

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., ... Srbinovsky, J. (2020). The australian earth system model: Access-esm1. 5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 193–214.