JGR Atmospheres

RESEARCH ARTICLE
10.1029/2022JD037978

Evaluating Causal 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[1,2]

[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
- Observations, CMIP6 historical and SSP5-8.5 simulations do not robustly show Arctic-midlatitude links arising from sea ice variability
- CMIP6 SSP5-8.5 simulations predict a more robust link between poleward eddy heat flux and Aleutian Low toward the end of the 21st century

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

E. Galytska,
egalytska@iup.physik.uni-bremen.de

Citation:

Galytska, E., Weigel, K., Handorf, D., Jaiser, R., Köhler, R., Runge, J., & Eyring, V. (2023). Evaluating causal Arctic-midlatitude teleconnections in CMIP6. *Journal of Geophysical Research: Atmospheres*, *128*, e2022JD037978. https://doi.org/10.1029/2022JD037978

Received 5 OCT 2022
Accepted 24 AUG 2023

Author Contributions:

Conceptualization: Evgenia Galytska, Katja Weigel, Veronika Eyring
Formal analysis: Evgenia Galytska
Funding acquisition: Veronika Eyring
Methodology: Jakob Runge, Veronika Eyring
Software: Evgenia Galytska, Jakob Runge
Visualization: Evgenia Galytska
Writing – original draft: Evgenia Galytska, Katja Weigel, Dörthe Handorf, Ralf Jaiser, Raphael Köhler, Jakob Runge, Veronika Eyring

Abstract To analyze links among key processes that contribute to Arctic-midlatitude teleconnections we apply causal discovery based on graphical models known as causal graphs. First, we calculate the causal dependencies from observations during 1980–2021. Observations show several robust connections from early to late winter, such as atmospheric blocking within central Asia via the Ural blocking and Siberian High, the North Atlantic Oscillation phase and the polar vortex (PV). The PV is affected by poleward eddy heat flux at 100 hPa, which is also directly connected with the Aleutian Low. We then evaluate climate models participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6) by comparing their causal graphs with those derived from observations. Compared to observations, CMIP6 historical and future simulations do not robustly capture Arctic-midlatitude teleconnections arising from Arctic sea ice variability. This highlights the role of atmospheric internal variability in modulating the Arctic-midlatitude teleconnections. However, we find several distinct patterns that are simulated by most of the analyzed climate models. For example, both historical and future model simulations robustly capture observed atmospheric blocking in central Asia. But contrary to observations, model simulations show a robust link between the Arctic temperature and sea ice cover over Barents and Kara seas. The analysis of future changes also reveals that the connection between the Aleutian Low and the poleward eddy heat flux at 100 hPa is expected to become more robust toward the end of the 21st century than in the analyzed past.

Plain Language Summary The role of different mechanisms that link amplified Arctic warming and changes in midlatitude weather remains an open question. Observations and model simulations lead to different conclusions, making interpreting Arctic-midlatitude connections difficult. To improve the understanding of these processes, this study uses a novel method that goes beyond simple correlation analysis, known as causal discovery. The application of causal discovery in combination with physical reasoning provides a powerful tool to evaluate the performance of climate models and better understand the mechanisms of analyzed teleconnections. Causal discovery detects cause–effect relationships from analyzed data using graphical models known as causal graphs. As a first step, we calculate causal graphs for observations. Then we compare the causal graphs from observations with historical simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6). This comparison shows if the CMIP6 models reproduce the observed connections in the current climate. We also estimate future changes in Arctic-midlatitude teleconnections toward the end of the century by comparing causal graphs from the CMIP6 historical and Scenario Model Intercomparison Project (ScenarioMIP) simulations. This study demonstrates that the observations and CMIP6 model simulations do not robustly show Arctic-midlatitude links that arise from Arctic sea ice variability.

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. (2022) 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

Writing – review & editing: Evgenia Galytska, Katja Weigel, Dörthe Handorf, Ralf Jaiser, Raphael Köhler, Jakob Runge, Veronika Eyring

Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report 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), an increase of atmospheric CO_2 (Previdi et al., 2020), or poleward heat and moisture transport (Sang et al., 2022) 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 cooling or lack of warming over this region (Overland et al., 2015). A number of studies link Arctic amplification to midlatitude weather variability (Francis, 2017; Francis & Vavrus, 2012; Screen et al., 2018; Shepherd, 2016; Walsh, 2014), which is not a one-way connection, but also works in reverse (Screen, 2017). Despite 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 (Blackport et al., 2019; Blackport & Screen, 2020, 2021; Siew et al., 2021).

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., 2020; Cohen, Screen, et al., 2014; 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, Screen, 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 agree about the dynamical pathway linking sea ice changes over the Barents and Kara seas to the midlatitudes. Primarily, during autumn and winter, the release of additional longwave radiation as well as sensible and latent heat from the open ocean waters causes strong warming of the lower troposphere over the Barents and Kara seas. This leads to a weakening of the westerly wind related 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; Outten et al., 2023), contributing to a northwestward expansion and intensification of the Siberian High. These tropospheric circulation changes are characterized by increased amplitudes of planetary waves and are favorable for the initiation of the stratospheric pathway of Arctic-midlatitude linkages (e.g., Nakamura et al., 2015; Peings, 2019; Siew et al., 2020). As shown by Siew et al. (2020) the above described relation between sea ice reduction over the Barents and Kara seas and more frequent and persistent Ural blocking is of intermittent character. Furthermore, the occurrence frequency of Ural blockings is related to atmospheric internal variability and may be triggered by both stationary and transient Rossby waves from the tropics (Dunn-Sigouin et al., 2021; Woollings et al., 2023) with distinct roles of stationary waves and transient waves or eddies for the meridional energy transport into the Arctic (Graversen & Burtu, 2016; Sang et al., 2022).

Although the mechanisms leading to more frequent and persistent Ural blocking are manifold and of intermittent character, it has been identified as a precursor pattern (Cohen & Jones, 2011; Garfinkel et al., 2010; Köhler et al., 2023) for enhanced upward propagating planetary waves leading to wave breaking in the polar stratosphere, which in turn weakens the polar vortex (PV) (e.g., Hoshi et al., 2017; Jaiser et al., 2016; Kim et al., 2014). This is then followed by the 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., 2023). 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 analyzed, 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, the recent coupled ocean–atmosphere model experiments (Screen et al., 2018) demonstrate the strengthening of the wintertime Aleutian Low in response to Arctic sea ice loss. 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 atmospheric circulation on interannual to decadal time scales over the North Pacific in winter (Hwang et al., 2022; Trenberth & Hurrell, 1994).

To analyze the connection between the Arctic and midlatitudes, we apply causal discovery (Runge, Bathiany, et al., 2019; Runge, Nowack, 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. Previous studies on Arctic-midlatitude linkages using causal discovery have focused on observational data sets (Kretschmer et al., 2016, 2018, 2020; 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 Modeling (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 that 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 reanalyses 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 analyzed 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 analyzed 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, the Shared Socioeconomic Pathway (SSP)5–8.5 is used that represents 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 literature. Monthly mean CMIP6 historical simulations (1980–2014) and ScenarioMIP SSP5-8.5 (2015–2021) are combined for the 42-year period of 1980–2021. Then the CMIP6 historical simulations are compared 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). Apart from the ERA5 reanalysis data set, we also use the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2, Bosilovich, 2015) and the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) global reanalysis (R-1, Kalnay et al., 1996; Kistler et al., 2001). This

Table 1
CMIP6 Models Analyzed 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 Modeling and Analysis, Environment and Climate Change Canada, Canada	Swart et al. (2019)
CMCC-CM2-SR5	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	Cherchi et al. (2019)
CMCC-ESM2		Lovato et al. (2022)
EC-Earth3 EC-Earth3-CC EC-Earth3-Veg EC-Earth3-Veg-LR	EC-Earth consortium, Europe	Döscher et al. (2022); Wyser et al. (2020)
E3SM-1-1	E3SM-project	Golaz et al. (2019)
FGOALS-f3-L	Chinese Academy of Sciences	He et al. (2019)
FGOALS-g3		Pu et al. (2020)
GFDL-CM4	National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, USA	Held et al. (2019)
GFDL-ESM4		Dunne et al. (2020)
INM-CM4-8	Institute for Numerical Mathematics, Russian Academy of Science, Russia	Volodin et al. (2018)
INM-CM5-0		Vorobyeva and Volodin (2021)
IPSL-CM6A-LR	Institut Pierre Simon Laplace, France	Boucher et al. (2020)
MPI-ESM1-2-HR	Max Planck Institute for Meteorology, Germany	Müller et al. (2018)
MPI-ESM1-2-LR		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)

comparison aims to identify the mechanism of Arctic-midlatitude teleconnections in the historical period and to estimate the differences across different reanalyses. To address future changes, we analyze the ScenarioMIP SSP5-8.5 during 2058–2099, which is the same period length as historical simulations. This study uses data from 23 CMIP6 models listed in Table 1. All available ensemble members (three or more per model) are considered to account for the intrinsic variability of the analyzed models. However, the analysis is shown for 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 realizations.

2.2. Potentially Linked Variables

To understand the linkages within Arctic-midlatitude processes it is important to define a physically reasonable number of variables or processes representing the investigated mechanisms. Table 2 provides a detailed overview of the variables used in this study, and Figure 1 summarizes potential variables. To reconstruct an

Table 2
Detailed Overview of Variables and Their Corresponding Regions Used in This Study

No	Variable		Geographical area	Label
1	Near-surface air temperature		65°–90°N, zonal mean	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, zonal mean, 100 hPa	vflux
8	Polar vortex (geopotential height)		65°–90°N, zonal mean, 100-10 hPa	PV
9	North Atlantic oscillation		20°–80°N, 90°W–40°E, 500 hPa	NAO

Arctic-midlatitude linkage that occurs in conditions of amplified Arctic warming the near-surface air temperature (TAS) over the broad region at high latitudes from 65° to 90°N was chosen, based on Figure 1 of Cohen et al. (2020). The sea ice area fraction over the Barents and Kara seas (BK-SIC) was included similarly to Kim et al. (2014), Kretschmer et al. (2016, 2020), and Siew et al. (2020). To understand the connections between the Arctic (and sub-Arctic) and Okhotsk sea ice variability, the sea ice area fraction over Okhotsk Sea (Ok-SIC, Ogi et al., 2015) was included. 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), similar to Cohen, Furtado, et al. (2014) and Kretschmer et al. (2016). The importance of Ural blocking and Siberian High for Arctic-midlatitude linkages has been extensively discussed by Cohen, Screen, et al. (2014), and the effect of the Aleutian Low on the Okhotsk sea ice is analyzed 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, 2020), and Siew et al. (2020) as v^*T^* at 100 hPa, where v stands for the meridional wind velocity, T stands for the temperature, and the superscript * indicates the deviations from the zonal mean. To analyze the variability in stratospheric polar circulation patterns, we also include the PV, based on geopotential height averaged over pressure levels from 100 to 10 hPa. Similar to Kretschmer et al. (2016) we invert the sign of PV, so the positive values stand for the strong PV, and negative values stand for the weak PV. A recent study by Kolstad and Screen (2019) suggests a link between Arctic sea ice reduction and the negative phase of NAO. Therefore, the NAO is also included in the analysis, which is based on the study by Hurrell and Deser (2010). To summarize, the following variables are used in this study to 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), PV, and NAO.

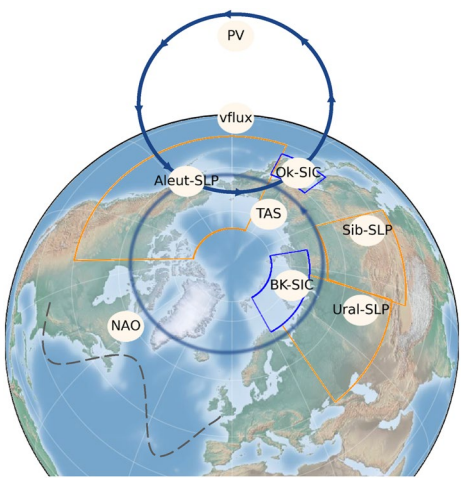

Figure 1. Summary of variables that represent Arctic and midlatitude processes in this study: near-surface air temperature over the Arctic (TAS), sea ice over the Barents and Kara seas (BK-SIC) and Okhotsk Sea (Ok-SIC), sea level pressure over Ural (Ural-SLP), Siberia (Sib-SLP), and Aleutian Islands (Aleut-SLP), poleward eddy heat flux (vflux), polar vortex (PV), and North Atlantic Oscillation (NAO). 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, providing well-documented source code and scientific background of developed diagnostics (see https://github.com/ESMValGroup/ESMValTool,

last access: 11.08.2023). The output produced by ESMValTool comprises provenance information, which allows for traceability and reproducibility of the obtained results. ESMValTool provides various simulations and experiments from CMIP6 models and an additional tool 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 variable. With the application of the ESMValTool preprocessor function called `anomalies`, we compute climatological monthly anomalies (observed value minus multiyear mean) similar to Kretschmer et al. (2016) for the potential variables defined in Sect 2.2. 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). In this study, a Climate Variability Diagnostics Package (CVDP, Phillips et al., 2014)is used, which is implemented into ESMValTool (Eyring et al., 2020) to reproduce NAO for the analyzed periods. To reproduce the calculations of potential variables described in Figure 1 and Table 2 we implemented the recipe `recipe_galytska23jgr.yml` and corresponding diagnostics in ESMValTool.

3.2. Causal Discovery

Statistical correlation and regression techniques commonly adopted in climate research might not always be useful to understand the causal interpretation of various 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 authors 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, Bathiany, et al., 2019; Runge et al., 2014), 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, improving 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 $\mathbf{X_t} = \left(X_t^1, \ldots, X_t^N\right)$, where N stands for the different variables represented by timeseries. To reconstruct the causal graph including their time lags among these variables, 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 variables $\left(X^i_{t-\tau}, X^j_t\right)$ for $\tau > 0$. For each pair of variables PC1 tests whether they are conditionally independent on defined conditions of other lagged variables. If yes, then the algorithm removes the link between these variables. At this stage, the algorithm estimates a set of lagged parents for each variable. 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:

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

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, $\widehat{\mathcal{B}}^-_t\left(X^j_t\right)$ are the lagged conditions of X^j_t, and $\widehat{\mathcal{B}}^-_{t-\tau}\left(X^i_{t-\tau}\right)$ 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 four types of links: (a) directed lagged causal links for $\tau > 0$, where τ stands for the time lag, (b) directed contemporaneous causal links for $\tau = 0$, (c) unoriented contemporaneous links indicating that the collider and orientation rules could not be applied due to Markov equivalence, and (d) unoriented contemporaneous links where a direction is not defined due to conflicting orientation rules. It is important to mention, that the definition of the links between PCMCI+ and PMCI algorithms differs and the interpretation of causal graphs should be done with caution.

This study focuses on linear dependencies and uses linear partial correlation (ParCorr) as a conditional independence test. In addition, we set the main parameters of PCMCI+ as follows: the minimum time delay $\tau_{\min} = 0$, which corresponds to contemporaneous connections, 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 $\alpha_{pc} = 0.01$ 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 normalized between −1 and 1. If links occur at multiple lags between two variables, the type of the link and its color show the strongest link, but the label indicates all significant lags sorted by their strength. To quantify the strength of detected connections, we additionally apply a multivariate regression on parents obtained with PCMCI + for different link coefficient strengths, similarly to Runge, Nowack, et al. (2019). Causal discovery with all supporting modules used in this study are implemented in the python package Tigramite and freely available at https://github.com/jakobrunge/tigramite, last access: 11.08.2023.

This study focuses on the NH cold season when near-surface amplified Arctic warming is the strongest (Cohen et al., 2020). Thus, the causal discovery was applied for the winter period consisting of the months December-January-February (DJF). In order to identify the differences in the mechanism of Arctic-midlatitude teleconnections within the entire cold season, further analysis was performed on two additional periods, that is, early winter (October-November-December, OND) and late winter (January-February-March, JFM). The additionally analyzed periods are constructed in a manner so that there is no overlap between the early and late winter periods, thus, excluding possible contributions of a common month. The contribution of December from the defined early winter period and January-February from the late winter period is then included in the regularly analyzed winter season DJF. We set `mask_type = y` to account for various months within the PCMCI+ calculations. This implies that PCMCI+ will search for causal drivers of target variables restricted to defined seasons, for example, DJF, but the potential causal drivers will be allowed outside of DJF, for example, in autumn months. The resulting causal graphs contain 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 analyze 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 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 asymmetric *F*1-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 *F*1-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 the 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 *F*1-score are equal. The *F*1-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 modeled causal graphs is the same. If not, we allow a difference in the time delay of up to two time lags, that is, two months for a monthly time scale. If the same-sign link in the modeled 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 it is considered that the link exists in both data sets. 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 modeled causal graphs, we relax the definition of contemporaneous links (with $\tau = 0$), and do not distinguish between directed, unoriented, and conflicting contemporaneous links. Instead, they are combined into one category of contemporaneous links to summarize the general outcome of modeled causal graphs.

3.4. Causal Model Evaluation

To characterize connections that occur in the framework of Arctic-midlatitude processes, Figure 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 variables (see Figure 2a) that represent Arctic-midlatitude processes (see Section 2.2). In this step ESMValTool is involved in calculating the climatological anomalies of monthly mean data and the area-weighted spatial average over the defined region for observations and climate models (see Section 3.1). As a result, a set of timeseries is received, which serves as proxies for different physical processes (noted as Var1, Var2, and Var3 in Figure 2b) for each data source (e.g., Observations, Model 1, Model 2), which is further detrended.

The interactions among these processes are calculated in observations and model simulations by the application of PCMCI+ (see Section 3.2). The resulting networks (see Figure 2c) contain the information on the existence of associated links between potential variables at a given significance level. The color of the node of each variable 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 a time lag. If the straight lines have an arrow, it denotes a directed contemporaneous causal link, where the arrow indicates the direction of the

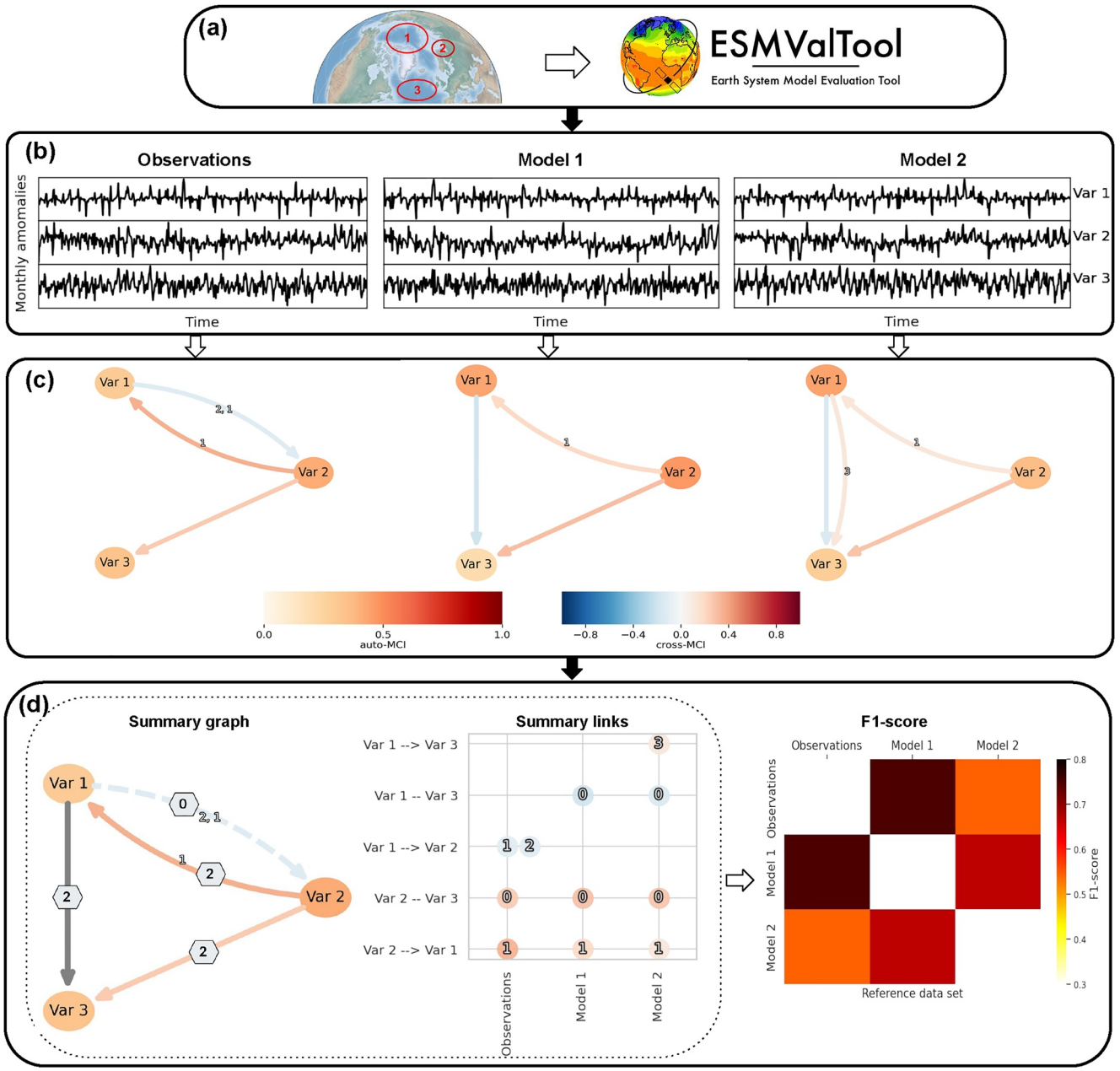

Figure 2. Schematic representation of the causal model evaluation framework using ESMValTool and causal discovery based on example data. For a detailed explanation see Section 3.4.

impact (see, e.g., the positive contemporaneous causal link from Var 2 to Var 3 in all data sources in Figure 2c and Runge (2020)). Curved lines with the arrow stand for the directed lagged causal links, where the number(s) indicates the time lag in months (see e.g., positive causal link from Var 2 to Var 1 in all data sources in Figure 2c). If two processes are causally linked with each other for more than one time lag, then all lags are shown and sorted by the strength of the impact. In such a case, the color of the link would be based on the stronger connection. For example, Figure 2c shows the negative causal link in Observations from Var 1 to Var 2 with time lags 2 and 1, where 2 has a stronger connection, thus is indicated first.

In the example given in Figures 2b and 2c the analysis is based on three data sources with three variables. But the comparison of causal graphs can be a challenging task, especially if the number of variables and/or data sources is increased since it can lead to complex causal structures. Therefore, Figure 2d (left figure) 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 simulates the observed connection, then the link is dashed. If more than half of the analyzed models simulate the particular connection, which was not found in observations, then the link is gray. For example, both models (Model 1 and Model 2) similar to observations, simulate a contemporaneous positive link from Var 2 to Var 3 (Figure 2c), which is indicated by the number "2" in the hexagon on top of this link (Figure 2d left). Neither of the models simulates the observed negative causal link from Var 1 to Var 2, which is indicated by "0" in the hexagon on top of this link, and the corresponding link is dashed. Both of the analyzed models simulate a negative contemporaneous link from Var 1 to Var 3, which was not detected in observations, thus the link between these two variables is gray with "2" in the hexagon on top of it. This analysis is performed for all detected and simulated links and is done to provide an overview of how many models agree with observed links. However, this analysis does not have information on particular models that do or do not simulate the observed connections. Therefore, the matrix in Figure 2d (middle) summarizes all discovered links (rows) versus the 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—a causal link with 1 month delay, 2—a causal link with 2 months delay, etc. And as a last step, to evaluate the general performance of climate models, the $F1$-scores are calculated (see Figure 2d, right, see also Section 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 match 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 for nine potential variables defined in Table 2 based on ERA5 and HadISST data sets for the period 1980–2021. Figure 3 shows causal graphs for (a) winter (DJF), (b) early winter (OND), and (c) late winter (JFM) with a significance level of $\alpha_{pc} = 0.01$ and a maximum time lag of $\tau_{\max} = 5$ for all tests in PCMCI+. The node color represents the strength of the autocorrelation. The link color stands for the MCI partial correlation value. Directed lagged and contemporaneous causal links are depicted by curved and straight arrows, respectively, pointing to the direction of the impact. The position of potential variables in Figure 3 corresponds to their approximate geographical location defined in Table 2. Similarly to Figure 3, the representation of causal graphs based on MERRA-2 and NCEP-NCAR-R-1 reanalyses is shown in Figures S1 and S2 in Supporting Information S1, respectively. Sea ice coverage from the HadISST data set is used in combination with each of these reanalyses.

From the causal graph of Arctic-midlatitude linkages during DJF (Figure 3a) among all potential variables the strongest autoregressive values are found for Ok-SIC and BK-SIC with a time lag of 1 month. Panel (a) shows a positive contemporaneous link between TAS and Sib-SLP, indicating the direct relationship between near-surface air temperature over a broad Arctic region and Siberian High. Since the link is not directed (no arrow), this connection implies that (a) the increase (decline) of Arctic near-surface temperature is associated with the strengthening (weakening) of Siberian High, or (b) 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). The increased sea level pressure over both Ural and Siberia is associated with the near-surface temperature increase over the Arctic (direct link between TAS and Sib-SLP, and an indirect link between Ural-SLP and TAS via Sib-SLP).

In Figure 3a we do not detect a connection between TAS and BK-SIC. It could be associated with the fact that the TAS is averaged over a broad domain, while BK-SIC is only over a small region (see Table 2). The reader should also remember that we combine two different data sets in the observational causal graphs, namely HadISST for BK-SIC and Ok-SIC and ERA5 for the rest of the variables. Since the observational causal graph is estimated with errors, the absence of the link (e.g., between TAS and BK-SIC) could happen if the connection is detected as a false negative in the causal graph. In the particular case shown in Figure 3a, the link from TAS to BK-SIC was detected as non-significant and thus was rejected in the conditional independence tests.

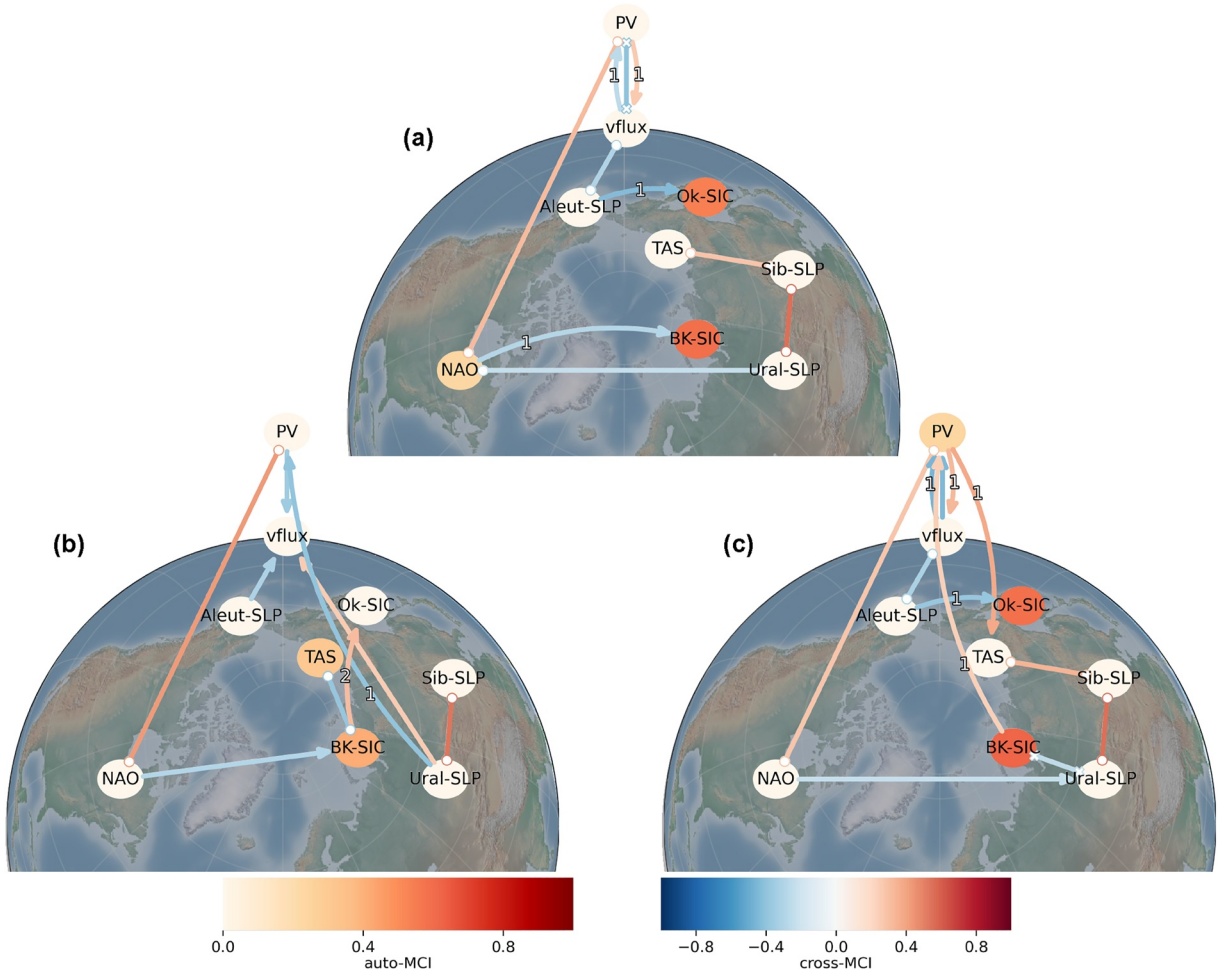

Figure 3. Causal graphs based on ERA5 and HadISST data sets for the period 1980–2021 during (a) winter (DJF), (b) early winter (OND), and (c) late winter (JFM). The significance level is $\alpha_{pc} = 0.01$ and the maximum time lag $\tau_{max} = 5$. The definition of variables is provided in Table 2. Similar to Figure 1, the position of variables corresponds to their approximate geographical location defined in Table 2.

During DJF a negative contemporaneous connection is found between Ural-SLP and NAO, indicating that the increased (declined) sea level pressure over Ural is associated with a negative (positive) phase of NAO, which is in agreement with the findings of Kretschmer et al. (2016). We also find a causal negative link from NAO to BK-SIC with 1 month lag, representing a tropospheric connection between midlatitudes and the Arctic, which is in agreement with Deser et al. (2000). Additionally, there is a contemporaneous positive link between the NAO and PV, which indicates a direct tropospheric-stratospheric relationship between the Arctic and midlatitudes (similar link was also detected in Kretschmer et al., 2016). It implies that the negative (positive) phase of the NAO is associated with a weakening (strengthening) of the PV. For example, a weaker PV produces circulation anomalies that are associated with an increasingly meandering jet stream that resembles the negative phase of the NAO.

During DJF we also find the negative links between vflux and PV, which are associated with the weakening of PV due to increased poleward eddy heat flux not only contemporaneously but also causally with 1 month lag. In turn, the weakened PV induces further weakening of the poleward eddy heat flux, indicated via a causal lagged positive link with 1 month lag. These connections were also reported by Kretschmer et al. (2016). In the North Pacific region we find another direct pathway of tropospheric-stratospheric connection via a negative contemporaneous link between Aleut-SLP and vflux, so for example, a deepening of the Aleutian Low is related to an increase in poleward eddy heat flux at 100 hPa. The Aleutian Low, in turn, is also connected via a negative causal link to sea ice over the Sea of Okhotsk (from Aleut-SLP to Ok-SIC) with 1 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.

During early winter (OND, Figure 3b) the distribution of contemporaneous and lagged causal links differs from DJF (Figure 3a). Therefore, in the following, we focus only on those changes in the connections that appeared newly in comparison to DJF. During OND the strongest autoregressive values are found for BK-SIC and TAS, both with 1 month delay. Contrary to DJF, the positive link between TAS and Sib-SLP is not detected during OND. However, we find a contemporaneous negative link between TAS and BK-SIC, which was not detected 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 the Barents and Kara seas.

Contrary to DJF (Figure 3a), the link from NAO to BK-SIC is not lagged but is a directed contemporaneous causal link. The causal graph from early winter does not detect the connection between NAO and Ural-SLP, which indicates that this particular link occurs starting in December. The following interesting patterns are found during OND for Ural-SLP: (a) positive contemporaneous connection to vflux and (b) direct causal negative link with 1 month lag to PV. However, these particular connections were reported by Kretschmer et al. (2016) during DJF.

During late winter (JFM, Figure 3c) the observations show pronounced autoregressive values for BK-SIC, Ok-SIC, and PV. We find a contemporaneous positive link between TAS and Sib-SLP, 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, which is in agreement with Li et al. (2021), Ma et al. (2022), Siew et al. (2020), and Yao et al. (2018). However, the observations do not detect this particular link during OND or DJF. This could be associated with the complexity of the interactions between sea ice and atmospheric internal variability that occurs across a range of timescales (Siew et al., 2020). The link between NAO and BK-SIC is not observed during JFM, which could be associated with the intermittency of this connection (Siew et al., 2020). However, during JFM the causal discovery detects a direct positive causal link from BK-SIC to PV with 1 month lag, which was not found in other periods.

Similar to Figure 3a, during JFM (panel c) in the North Pacific region we find a negative causal connection with 1 month delay from Aleut-SLP to Ok-SIC, whereas it does not take place during early winter (panel b). During JFM we also find two causal positive links with 1 month lag associated with the PV. Similar to DJF, the first link is from PV to vflux, indicating that the weakened PV causes the weakening of the poleward eddy heat flux. The second link is from PV to TAS, indicating that the weakening of the PV leads to a decrease in Arctic near-surface temperature. This link is associated with the downward coupling of stratospheric PV anomalies, and it is known to be particularly strong during sudden stratospheric warming (SSW) events that can affect 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; 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.

Overall we found that the following observed teleconnections were detected similarly to Kretschmer et al. (2016), such as links that connect the NAO (Arctic Oscillation, AO in case of Kretschmer et al., 2016) and (i) Ural-SLP, (ii) BK-SIC, and (iii) PV; (iv) as well as links between vflux and PV. Reported by Kretschmer et al. (2016) connection between BK-SIC and Ural-SLP we detected during late winter (panel c) and reported connection from Ural-SLP to vflux and PV we detected only during early winter (panel b). These particular differences may arise from a number of potential factors, that is, different reanalyses, different analyzed periods of time, the application of another version of the causal discovery with different settings on a slightly different set of analyzed variables. The estimation of the sensitivity of detected connections to the parameter settings of causal discovery is shown in Figure S3 in Supporting Information S1. Therefore, the differences in Arctic-midlatitude connections detected by causal discovery during different analyzed periods with different significance levels may indicate an important contribution of atmospheric internal variability to the observed processes.

The summary of all detected links based on observations with the significance level $\alpha_{pc} = 0.01$ and the maximum time lag $\tau_{\max} = 5$ is provided in Table 3, where the second column indicates the linked variables, and the third column stands for the type of the link (lagged or contemporaneous). If the link is lagged then the number in round brackets shows the time delay in months. The fourth column shows whether the link is positive (+) or negative (−). The last three columns show multivariate regression values based on the parents obtained with PCMCI+. To summarize the changes in the connections, we sort the periods from early (OND, column five) to late (JFM, column seven) winter. According to Table 3, four links appear during all three analyzed periods (OND, DJF, and

Table 3
Summary of the Links Based on the Application of Causal Discovery on ERA5 and HadISST Data Sets for the Period 1980–2021 Based on the Significance Level $\alpha_{pc} = 0.01$ and the Maximum Time Lag $\tau_{max} = 5$

No	Link	Link type	Sign	OND	DJF	JFM
1	TAS and BK-SIC	contemp.	−	−0.39		
2	TAS and Sib-SLP	contemp.	+		0.50	0.48
3	Sib-SLP and Ural-SLP	contemp.	+	0.60	0.64	0.60
4	Ural-SLP and PV	lagged (1)	−	−0.49		
5	Ural-SLP and vflux	contemp.	+	0.24		
6	Ural-SLP and BK-SIC	contemp.	−			−0.22
7	BK-SIC and Ok-SIC	lagged (2)	+	0.30		
8	BK-SIC and PV	lagged (1)	+			0.14
9	NAO and Ural-SLP	contemp.	−		−0.38	−0.26
10	NAO and BK-SIC	lagged (1)	−		−0.28	
11		contemp.	−	−0.26		
12	NAO and PV	contemp.	+	0.44	0.46	0.20
13	vflux and PV	lagged (1)	−		−0.6	−0.46
14	PV and vflux	lagged (1)	+		0.25	0.28
15		contemp.	−	−0.50	−0.39	−0.34
16	PV and TAS	lagged (1)	+			0.47
17	Aleut-SLP and vflux	contemp.	−	−0.38	−0.31	−0.31
18	Aleut-SLP and Ok-SIC	lagged (1)	−		−0.34	−0.26

Note. Link type stands for either the lagged causal link (lagged), where the number in round brackets indicates the time lag in months or the directed contemporaneous causal link or unoriented contemporaneous adjacency (contemp.). The sign indicates a positive (+) or negative (−) connection. The values in the last three columns show multivariate regression values on the parents obtained with PCMCI+.

JFM), namely contemporaneous links between (i) Sib-SLP and Ural-SLP, (ii) NAO and PV, (iii) PV and vflux, and (iv) Aleut-SLP and vflux. There are a number of connections that are found in only one of the analyzed periods. For example, the contemporaneous connection between TAS and BK-SIC and the causal link from BK-SIC to Ok-SIC were detected only during early winter OND; the downward coupling from the stratosphere to the troposphere via PV to TAS or contemporaneous connection between Ural-SLP and BK-SIC were detected 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 Figure 2c. The representation of all simulated causal graphs in DJF from analyzed climate models is shown in Figure S4 in Supporting Information S1. To generally estimate how many climate models simulate physical processes detected in observations, Figure 4 shows causal graphs based on observations from Figure 3, but in this case the number in the hexagon above each link indicates how many models out of 23 simulate the link identical to observations by sign (positive or negative) and type (lagged or contemporaneous). Additionally, the width of the lines in Figure 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). The gray links indicate the connections that were not detected in observations but were simulated by more than half of the analyzed models. To optimize the comparison between observations and models, in the following we do not distinguish the three types of contemporaneous links, namely (a) directed contemporaneous causal links, (b) unoriented, and (c) conflicting contemporaneous adjacencies, but summarize them into one category and simply refer to these connections as contemporaneous connections. We adhere to a criterion of more than half of the analyzed models (i.e., ≥12) to represent a robust model response.

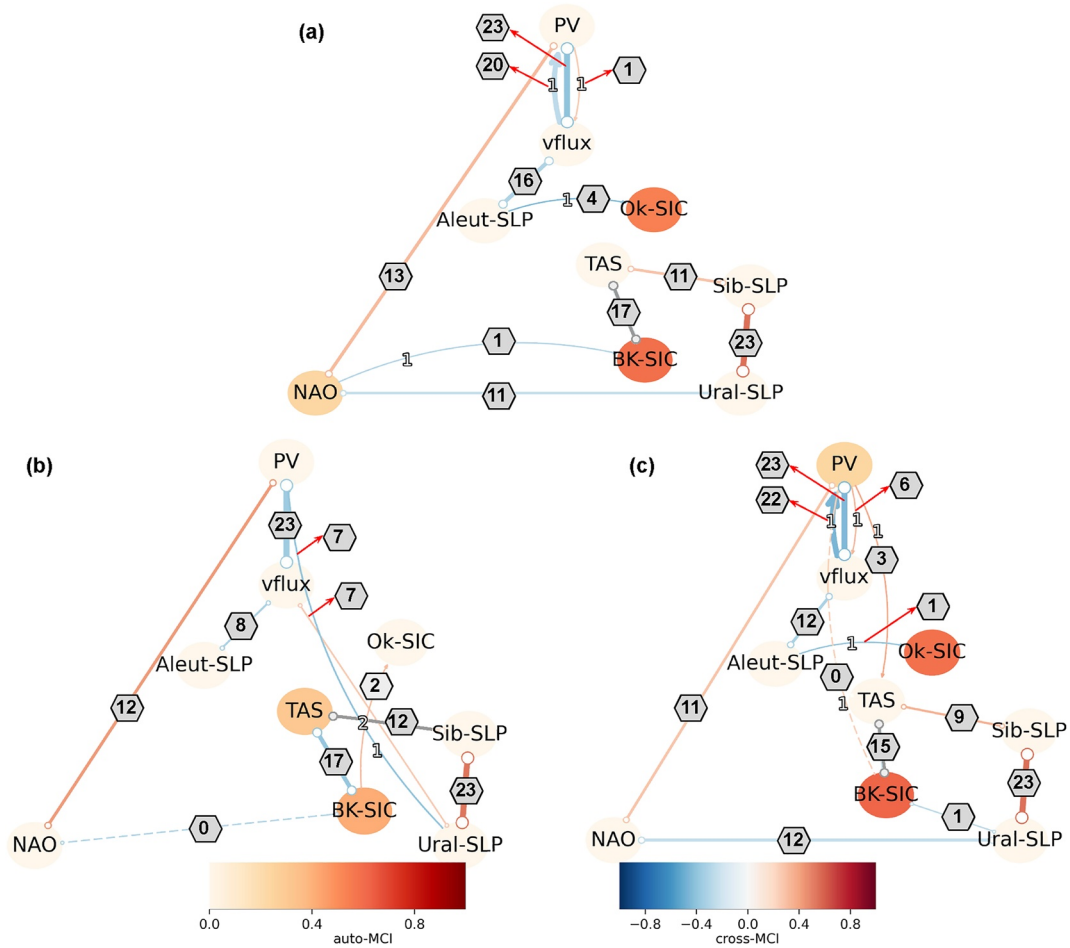

Figure 4. Summary causal graphs based on 23 CMIP6 historical simulations for the period 1980–2021 during (a) winter (DJF), (b) early winter (OND), and (c) late winter (JFM). Similar to Figure 3, $\alpha_{pc} = 0.01$ and $\tau_{\max} = 5$. The line width represents the fraction of the models that simulate 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 the number of models that simulate a (contemporaneous or causal) link identical by sign and type to the observations. The gray links indicate the connections that were not detected in observations but were simulated by more than half of the analyzed models.

There are a few observational links that were simulated by most of the 23 analyzed models in all analyzed periods, such as the negative contemporaneous and lagged connections from vflux to PV and the positive contemporaneous connection between Sib-SLP and Ural-SLP (Figures 4a–4c). A robust connection between vflux and PV is in agreement with Kretschmer et al. (2020), who made their conclusions based on the analysis of 35 CMIP5 models. The rest of the simulated links exhibit seasonal differences in model simulations. For example, the contemporaneous link between TAS and Sib-SLP during OND (Figure 4b) was simulated by 12/23 models (but not detected in observations), but during DJF (Figure 4a) was simulated by 11/23 models and during JFM (Figure 4c) by 9/23 models. There is no robust modeled connection between NAO and BK-SIC, which was detected in observations during OND and DJF. The causal lagged link from Ural-SLP to PV, which was detected in observations only during OND (Figure 3b) is also simulated by only 7/23 models (Figure 4b). During JFM (Figure 4c) the direct link between BK-SIC and Ural-SLP was not robustly detected by the models, which is in agreement with the findings of Kretschmer et al. (2020), who demonstrated a weak mediated signal on a monthly timescale.

Interestingly, the link between TAS and BK-SIC, which was not detected in observations during DJF and JFM, is simulated by most of the analyzed models in all considered periods, namely 17/23 in OND and DJF, and 15/23 in JFM. This connection is in agreement with, for example, Screen and Simmonds (2010) and Screen et al. (2012). Notably, the recent study of Docquier et al. (2022) showed the two-way influence of Arctic sea ice and near-surface air temperature.

In the North Pacific region, the causal link from Aleut-SLP to Ok-SIC is found in 4/23 models in DJF (panel a) and in 1/23 models during JFM (panel c). The contemporaneous link between vflux and Aleut-SLP is simulated by most of the analyzed models (16/23) during DJF, less robust during OND (panel b, 8/23) and JFM (panel c, 12/23 models), which is in agreement with findings of Kretschmer et al. (2020) regarding the robust representation of this particular link by climate models.

While Figure 4 shows a general overview of the analyzed models based on the observational links, Figure 5 summarizes all detected contemporaneous and lagged links (rows) for observations and 23 analyzed CMIP6 models (columns) during DJF 1980–2021. Similar to Figure 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 found in observations and simulated (or not simulated) by CMIP6 models during DJF. Analogous to Figure 5, the summary causal graphs for OND and JFM are shown in the Figures S5 and S6 in Supporting Information S1, respectively. Based on Figure 5 we detect that during DJF apart from the observations, the only model that reproduces the causal positive observational link from PV to vflux is CMCC-ESM2. While only one model simulates the observed causal negative connection from NAO to BK-SIC in DJF (MRI-ESM2-0), we find 5/23 models that show a contemporaneous negative link between these variables (CanESM5, GFDL-ESM4, INM-CM4-8, MPI-ESM1-2-LR, and MRI-ESM2-0). In the North Pacific, the causal negative link with 1 month delay from Aleut-SLP to Ok-SIC is simulated by four models: BCC-CSM2-MR, CMCC-CM2-SR5, E3SM-1-1, and MPI-ESM1-2-LR. We found that there are five other models (FGOALS-g3, GFDL-CM4, GFDL-ESM4, INM-CM4-8, and IPSL-CM6A-LR) that simulate contemporaneous negative link between Aleut-SLP and Ok-SIC.

To estimate the general performance of the 23 analyzed CMIP6 historical simulations (only first ensemble member r1i1p1f1), we calculate the F1-score on the pair-wise comparison of causal graphs that are based on monthly mean data of the nine analyzed variables. 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 are shown in columns. Higher scores imply a better agreement between causal graphs, that is, that two models (or model and observations) are more similar in terms of their causal fingerprint; lower scores represent larger differences between reference and analyzed causal graphs.

Figure 6 shows that generally there is a better agreement among observations and analyzed models during DJF (panel a) compared to OND (panel b) and JFM (panel c). The lower F1-scores during OND and JFM imply that the agreement of causal graphs among analyzed models and observations is lower during these periods. Below we discuss the highest F1-scores during different seasons, and in the round brackets we denote the F1-score values. The highest F1-scores during OND and DJF are detected for the GFDL-CM4 model when its causal graph is used as a reference compared to INM-CM4-8 (0.88 in OND and 0.92 in DJF). During JFM the highest F1-scores (all 0.87) were found when EC-Earth3-Veg-LR and CMCC-CM2-SR5 (both reference) are compared to ACCESS-ESM1-5 and when ACCESS-ESM1-5 (reference) is compared to CMCC-CM2-SR5. The application of F1-score metric does not simply provide the similarity of simulated causal graphs but also provides a marker for a possible similar background of model development. For example, during JFM one of the highest F1-scores detected in IPSL-CM6A-LR (Boucher et al., 2020) is in comparison with EC-Earth-Veg (Döscher et al., 2022). Since both models 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.

In the following, we highlight a few models with the highest F1-scores with observations, thus simulating the processes closest to the observations. During DJF, the following models agree best with the observational causal graphs: BCC-CSM2-MR (0.84), CAMS-CSM1-0 and MPI-ESM1-2-LR (both 0.8). During OND these models exhibit the highest scores: ACCESS-CM2 (0.69), FGOALS-f3-L, GFDL-CM4 and IPSL-CM6A-LR (all 0.67). The highest F1-score in comparison to observations during JFM are detected in CAMS-CSM1-0 and EC-Earth3-Veg-LR (both 0.72), and ACCESS-ESM1-5 (0.71). In this context, it is important to highlight the following two points. First, while generally relatively high F1-scores suggest that analyzed climate models adequately represent the observed connections, the average agreement among models and observations with a F1-score of around 0.5–0.6 might indicate an important contribution of internal variability to the observed processes. In particular, the higher F1-scores during DJF suggest that climate models better capture observed links (or agree on connections across different models) during this season. Thus, second, the high F1-scores between

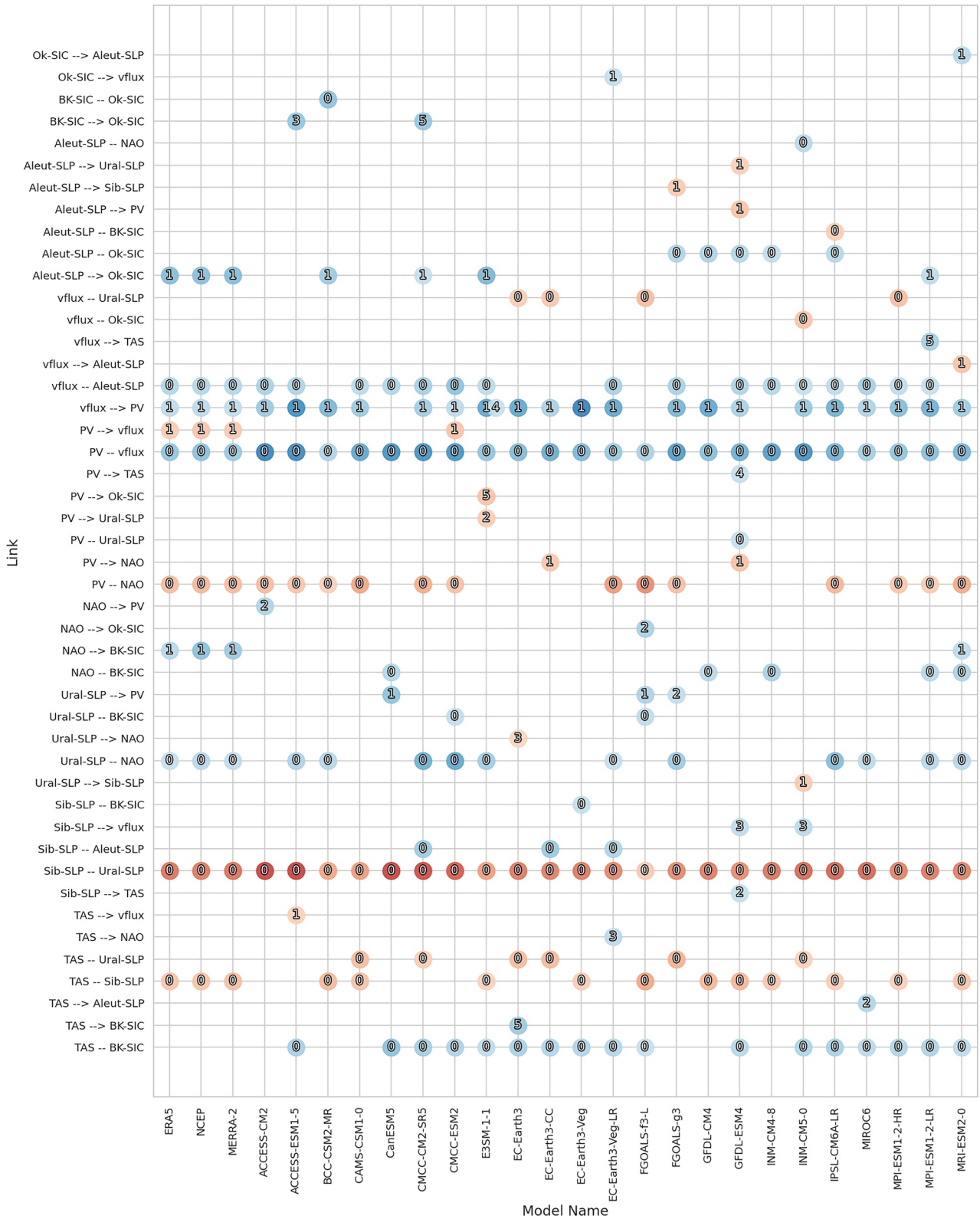

Figure 5. Summary of lagged and contemporaneous connections detected in reanalyses ERA5, NCEP (stands for NCEP-NCAR-R-1) and MERRA-2 and 23 CMIP6 historical simulations during winters (DJF) for the period 1980–2021. The node color represents the strength of the link. The number within the node represents the time lag in months in causal links, and zero represents contemporaneous links.

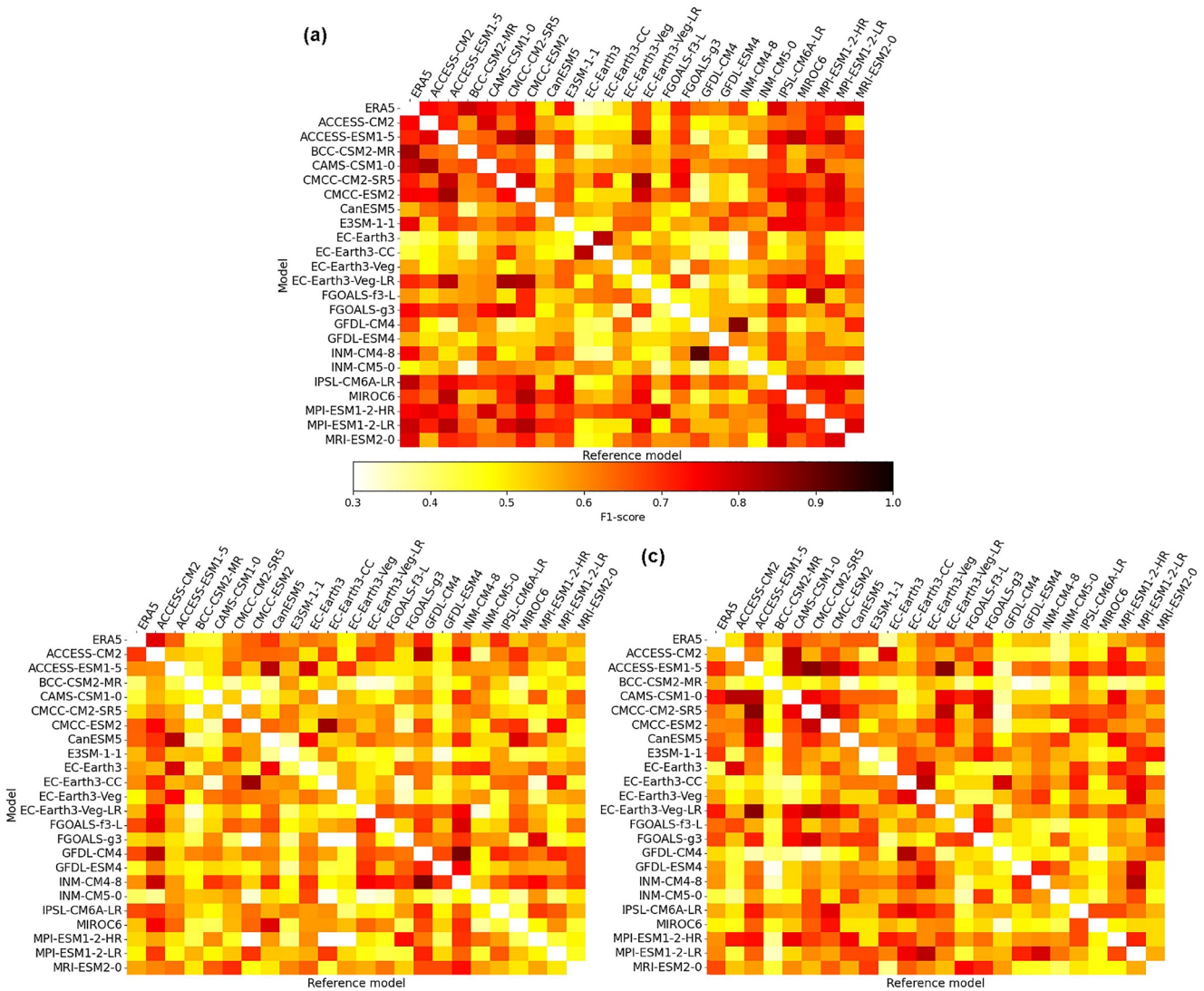

Figure 6. Matrix of average F1-scores for pair-wise comparisons of causal graphs among observations and the first ensemble member (r1i1p1f1) of 23 climate models during (a) DJF, (b) OND, and (c) JFM. Reference data is shown in rows, and data that is compared to the reference is shown in columns. Higher (lower) scores stand for better (worse) agreement between causal graphs.

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 BCC-CSM2-MR and observations as reference is the highest during DJF (Figure 6a; 0.84), during OND the similarity in causal graphs is much lower (0.4, panel b), and during JFM it is higher again (0.57, panel c). This shows that the agreement between observations and models (and among models) highly depends on the analyzed period. Since the 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 analyzed processes vary for different analyzed periods (in our case OND, DJF, or JFM) across the analyzed CMIP6 models.

4.3. Causal Links in CMIP6 SSP5-8.5 Future Simulations

To address potential future changes in Arctic-midlatitude teleconnections, Figure 7 displays causal graphs similar to Figure 4, but for the ScenarioMIP SSP5-8.5 simulations for the period 2058–2099. The detailed summary of causal and contemporaneous links detected in CMIP6 SSP5-8.5 simulations for winters 2058–2099 is provided in

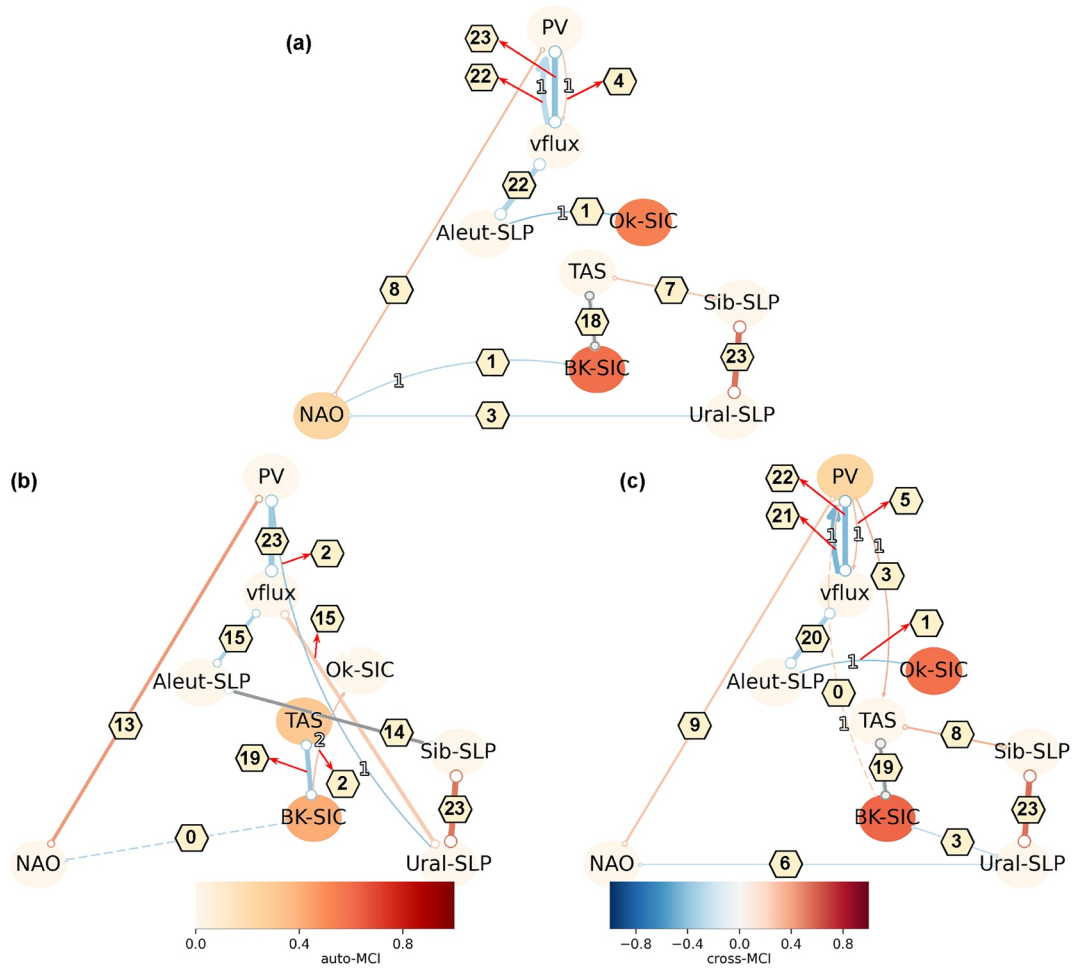

Figure 7. Same as Figure 4 but for the ScenarioMIP SSP5-8.5 simulations for the period 2059–2099.

supporting information Figure S7 in Supporting Information S1. First, we analyze those 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 the robustness of this link only during DJF (see Figure 4a, 16/23 models) and JFM (see Figure 4c, 12/23 models), while during OND only 8/23 models simulated this connection. During 2058–2099 it is expected that this connection will become more robust for all analyzed periods: 22/23 models for DJF, 20/23 models for JFM, and 15/23 models for OND (see Figures 7a–7c). 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 slightly more robust from historical to SSP5-8.5 simulations. For example, Figure 4 shows that during 1980–2021 17/23 models in DJF and 15/23 models in JFM simulated this connection, although it was not found in the observations; also 17/23 models simulated this link in OND. In the future this link is simulated by more models: 18/23 models during DJF, 19/23 in OND and JFM (also see Figures S8–S10 in Supporting Information S1).

Several links become less robust in the future SSP5-8.5 simulations. For example, the contemporaneous negative connection between NAO and PV, which was found in 13/23 models in the historical simulations (Figure 4a), is simulated in the future by only 8/23 models. Or the connection between NAO and Ural-SLP in JFM that was found in more than half of analyzed models 12/23 in historical simulations (Figure 4c), is simulated in the future by only 6/23 models. A new robust connection is found in future OND (Figure 7b), which was not detected before as robust in observations and CMIP6 historical simulations. This is a contemporaneous negative connection between Sib-SLP and Aleut-SLP (also seen in Figure S9 in Supporting Information S1).

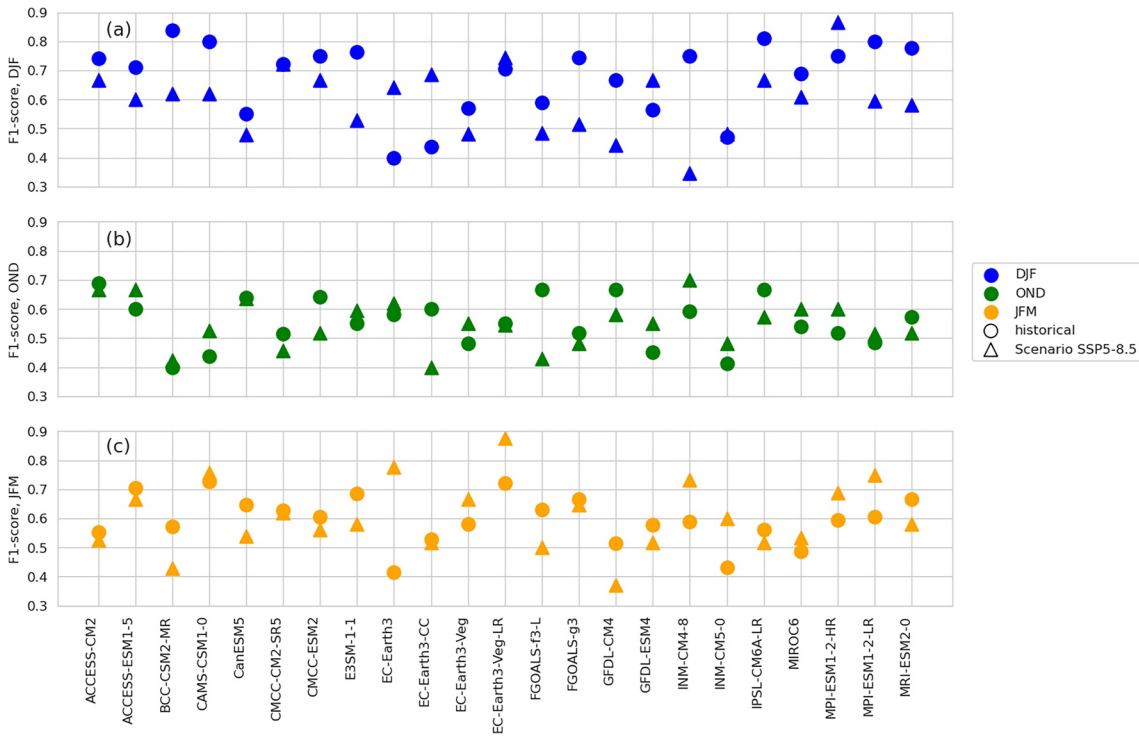

Figure 8. *F*1-score scatter plot of individual causal network comparison for the first ensemble member (r1i1p1f1) of 23 CMIP6 historical (circles) and SSP5-8.5 (triangles) simulations for winter DJF (panel a, blue), early winter OND (panel b, green), and late winter JFM (panel c, orange). Similar to Figure 6, the observational causal graphs from ERA5 and HadISST data sets are used as a reference for the analysis of CMIP6 historical simulations. And the historical simulations are chosen as a reference for the future scenario for each analyzed model.

To summarize the difference between observed and modeled causal graphs in past and future simulations, Figure 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 Figure 6, the observational causal graphs from ERA5 and HadISST data sets are used as a reference for the analysis of CMIP6 historical simulations. To estimate the impact of a changing climate on observed and simulated links, the historical simulations are chosen as a reference for each CMIP6 model future scenario. The lower the F1-score for SSP5-8.5 simulations, the larger the difference between the causal graphs of the historical and future simulations. We find that the F1-scores for the future simulations are generally higher during DJF and JFM (see Figures 8a and 8c), and lower during OND (Figure 8b). 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 small 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 Figure 8) predict only small future changes in comparison with historical simulations. During DJF (panel a) these models are MPI-ESM1-2-HR (0.86) and EC-Earth3-Veg-LR (0.74); during OND (panel b) these are INM-CM4-8 (0.7), ACCESS-CM2 and ACCESS-ESM1-5 (both 0.67), and during JFM (panel c) these are EC-Earth3-Veg-LR (0.88) and EC-Earth3 (0.78). Therefore, these models predict the smallest future change in the analyzed connections compared to their historical simulations.

The lowest F1-scores in SSP5-8.5 simulations, thus the largest differences between past and future processes are simulated during DJF by INM-CM4-8 (0.35) and GFDL-CM4 (0.44). During OND these models are EC-Earth3-CC (0.4), BCC-CSM2-MR and FGOALS-f3-L (both 0.42). And during JFM these models are GFDL-CM4 (0.37) and BCC-CSM2-MR (0.43).

5. Conclusions

Our study applies causal discovery to compare Arctic-midlatitude teleconnections across a number of CMIP6 simulations to observational data sets. Causal discovery is recognized as an interpretable machine learning tool

that goes beyond correlation analysis to study complex systems. By analyzing detrended timeseries from observations and 23 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. To reveal possible differences in the observed and simulated connections 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 1980–2021 across the analyzed periods, we detect robust connections related to the variability of tropospheric circulation and troposphere-stratosphere coupling. These are the positive contemporaneous links between (i) Ural blocking and Siberian high, (ii) NAO and stratospheric PV, and negative contemporaneous links between (iii) stratospheric PV and poleward eddy heat flux at 100 hPa, and (iv) Aleutian Low and poleward eddy heat flux at 100 hPa. Contrary to the expectations, based on observations, the causal discovery did not detect a robust connection between sea ice over Barents and Kara seas and Arctic temperature: a negative contemporaneous link was only detected during OND. This connection might not be found during DJF and JFM if the observational causal graph was estimated with errors and the causal discovery detects this connection as a false negative.

The causal analysis of CMIP6 historical simulations, similar to observations, does not robustly capture Arctic-midlatitude teleconnections that arise from Arctic sea ice variability in the analyzed periods during 1980–2021. We also show that only a few observational links are reproduced by most of the analyzed models. This is mainly (i) the negative connection between poleward eddy heat flux at 100 hPa and stratospheric PV and (ii) the positive link between Ural blocking and Siberian high. Contrary to observations, the negative connection between sea ice over Barents and Kara seas and the Arctic temperature is robustly simulated by most of the analyzed models during all seasons.

The analysis of CMIP6 SSP5-8.5 simulations during 2058–2099 shows similar results as the observations and historical simulations and does not show the new robust links associated with Arctic-midlatitude teleconnections. However, future simulations show several connections that are clearly expected to become more robust. The most prominent future changes are found between poleward eddy heat flux at 100 hPa and Aleutian Low for all analyzed periods based on CMIP6 SSP5-8.5 in comparison to historical simulations, which could be associated with the predicted future intensification of Aleutian Low (Giamalaki et al., 2021). Additionally, the connection between Arctic temperature and sea ice over Barents and Kara seas is expected to become more robust in the future during all analyzed periods since more climate models simulated this connection. This could be associated with a thinner ice cover in the future and its pronounced vulnerability to thermodynamic forcing.

This study combines a statistical method with physical reasoning via the application of the causal discovery on the topic of Arctic-midlatitude teleconnections. This approach provides a powerful tool to evaluate the performance of climate models and assesses the physical mechanisms of these teleconnections. However, there is a number of issues that the user should be familiar with. A careful selection of the potential variables as well as the length of the timeseries plays a crucial part in the further application of causal discovery. Similarly to Siew et al. (2020), we also experienced an issue when the algorithm rejected some of the physically meaningful links when some of the variables were changed (or extended/reduced) or by prolonging or shortening the analyzed period by a few years. In this regard, Kretschmer et al. (2021) highlighted that in order to make plausible estimates of causal discovery, it is important to explicitly define the causal assumptions and account for the correct confounding variables, which is done in this study (see Sect. 2.2).

Moreover, since the observations and climate models provide inconsistent responses, the scientific findings are influenced by the prior knowledge and findings (Shepherd, 2021). A way to overcome these issues is by applying the Bayes factor (Kass & Raftery, 1995), which quantifies the evidence in favor of a scientific theory. The Bayes factor does not depend on the prior odds, and the alternative hypothesis can be precisely specified. In a recent study, Kretschmer et al. (2021) applied the Bayes factor to test the hypothesis of the influence of BK-SIC on the non-linearity of the PV. As a result, their analysis neither proved nor provided evidence against their assumption, which highlighted the complexity of the researched teleconnections. Combining the extensive analysis of the present study with the complex approach of Kretschmer et al. (2021) is a useful direction for further research.

From the perspective of causal model evaluation of Arctic-midlatitude processes, there are a few additional challenges associated with the non-stationarity of the Arctic-stratospheric pathways and their intermittency (Siew et al., 2020), unrealistic sea ice distribution and variability in the models in comparison to the observations

(Kretschmer et al., 2020). Additionally, for future studies to further estimate the role of the poleward eddy heat flux, it is potentially beneficial to disentangle the impact of stationary waves, synoptic and low-frequency transient eddies that may have distinct roles in Arctic-midlatitude connections (Sang et al., 2022). Also, to better understand the links behind Arctic-midlatitude teleconnections and evaluate detected connections, further research should also focus on the multi-model large ensembles to address the complexity of the connections between atmospheric internal variability and Arctic sea ice changes.

Data Availability Statement

The CMIP6 model data is publicly available upon registration via the Earth System Grid Federation (ESGF, https://esgf.llnl.gov, registration is required, last access: 11.08.2023, Williams et al., 2009). ERA5 data is available via Climate Data Store (CDS, https://cds.climate.copernicus.eu/, registration is required, last access: 11.08.2023, Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate., 2017). The HadISST data set was obtained from the UK Met Office Hadley Centre and is available online at http://www.metoffice.gov.uk/hadobs/hadisst/data/download.html, no registration is required, last access: 11.08.2023 (HadISST, 2023).

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 release of ESMValTool is publicly available on Zenodo at https://doi.org/10.5281/zenodo.6900341 (Andela et al., 2022). ESMValTool is developed on the GitHub repository available at https://github.com/ESMValGroup, last access: 11.08.2023. The new recipe recipe_galytska23jgr.yml and corresponding diagnostics are implemented in ESMValTool, which can be used to reproduce the calculations of potential variables provided in Table 2 of this study.

The causal discovery algorithm is implemented in the python package Tigramite is released under GNU General Public License v3.0. Tigramite v5 is publicly available on Zenodo https://doi.org/10.5281/zenodo.6247837 (Runge, 2022) or via https://github.com/jakobrunge/tigramite, last access: 11.08.2023. The code used to reproduce the summary of causal and contemporaneous links and F1-scores and to produce figures for this manuscript is accessible in the following GitHub repository https://github.com/EyringMLClimateGroup/galytska-23jgr_EvaluatingCausalArcticMidlatTelec, last access: 29.08.2023, publicly available on Zenodo at https://doi.org/10.5281/zenodo.8268180 (Galytska, 2023).

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 program (Grant agreement 855187), the European Union's Horizon 2020 research and innovation programme under Grant Agreement 101003536 (ESM2025—Earth System Models for the Future), and the "Advanced Earth System Model Evaluation for CMIP (EVal4CMIP)" project funded by the Helmholtz Society. EG is partly supported by Central Research Development Fund at the University of Bremen, Funding ZF04A/2023/FB1/Galytska Evgenia. JR has received funding from the ERC Starting Grant CausalEarth under the European Union's Horizon 2020 research and innovation program (Grant Agreement 948112). DH is partly supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) Transregional Collaborative Research Center SFB/TRR 172 "Arctic Amplification: Climate Relevant Atmospheric and Surface Processes, and Feedback Mechanisms (AC)3" (Project-ID 268020496). RK is supported by the European Union's Horizon 2020 research and innovation framework programme under Grant agreement 101003590 (PolarRES). RJ is primarily supported by "Synoptic events during MOSAiC and their Forecast Reliability in the Troposphere-Stratosphere System" (SynopSys) funded by the German Federal Ministry for Education and Research (Grant/Award Number: 03F0872A). We acknowledge the World Climate Research Program's (WCRP's) 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 by its Scientific Steering Committee (WLA) under project ID bd1083. The authors thank Wiebke Günther for the help with Tigramite, Kevin Debeire for the help with the F1-score, and Rémi Kazeroni for the technical support. Open Access funding enabled and organized by Projekt DEAL.

References

Ambaum, M. H., Hoskins, B. J., & Stephenson, D. B. (2001). Arctic oscillation or North Atlantic Oscillation? *Journal of Climate*, *14*(16), 3495–3507. https://doi.org/10.1175/1520-0442(2001)014<3495:aoonao>2.0.co;2

Andela, B., Broetz, B., de Mora, L., Drost, N., Eyring, V., Koldunov, N., et al. (2022). Esmvaltool [Software]. Zenodo. https://doi.org/10.5281/zenodo.6900341

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. https://doi.org/10.1175/1520-0469(1976)033<2031:pwihav>2.0.co;2

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. https://doi.org/10.1175/2009jas3270.1

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., et al. (2021). Sudden stratospheric warmings. *Reviews of Geophysics*, *59*(1), e2020RG000708. https://doi.org/10.1029/2020rg000708

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. https://doi.org/10.1126/science.1063315

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. https://doi.org/10.1002/wcc.337

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Sullivan, A., Bodman, R., et al. (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. https://doi.org/10.1071/es19040

Blackport, R., & Screen, J. A. (2020). Insignificant effect of arctic amplification on the amplitude of midlatitude atmospheric waves. *Science Advances*, *6*(8), eaay2880. https://doi.org/10.1126/sciadv.aay2880

Blackport, R., & Screen, J. A. (2021). Observed statistical connections overestimate the causal effects of arctic sea ice changes on midlatitude winter climate. *Journal of Climate*, *34*(8), 3021–3038. https://doi.org/10.1175/jcli-d-20-0293.1

Blackport, R., Screen, J. A., van der Wiel, K., & Bintanja, R. (2019). Minimal influence of reduced arctic sea ice on coincident cold winters in mid-latitudes. *Nature Climate Change*, *9*(9), 697–704. https://doi.org/10.1038/s41558-019-0551-4

Bosilovich, M. G. (2015). *Merra-2: Initial evaluation of the climate*. National Aeronautics and Space Administration, Goddard Space Flight Center.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. https://doi.org/10.1029/2019MS002010

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., et al. (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. https://doi.org/10.1029/2018MS001369

Cohen, J., Furtado, J. C., Jones, J., Barlow, M., Whittleston, D., & Entekhabi, D. (2014). Linking Siberian snow cover to precursors of stratospheric variability. *Journal of Climate*, *27*(14), 5422–5432. https://doi.org/10.1175/jcli-d-13-00779.1

Cohen, J., & Jones, J. (2011). A new index for more accurate winter predictions. *Geophysical Research Letters*, *38*(21). https://doi.org/10.1029/2011gl049626

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., et al. (2014). Recent arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, *7*(9), 627–637. https://doi.org/10.1038/ngeo2234

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., et al. (2020). Divergent consensuses on arctic amplification influence on midlatitude severe winter weather. *Nature Climate Change*, *10*(1), 20–29. https://doi.org/10.1038/s41558-019-0662-y

Copernicus climate change service (c3s) (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. [Dataset]. Retrieved from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview and https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels-monthly-means?tab=overview

Deser, C., Walsh, J. E., & Timlin, M. S. (2000). Arctic sea ice variability in the context of recent atmospheric circulation trends. *Journal of Climate*, *13*(3), 617–633. https://doi.org/10.1175/1520-0442(2000)013<0617:asivit>2.0.co;2

Di Capua, G., Runge, J., Donner, R. V., van den Hurk, B., Turner, A. G., Vellore, R., et al. (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. https://doi.org/10.5194/wcd-1-519-2020

Docquier, D., Vannitsem, S., Ragone, F., Wyser, K., & Liang, X. S. (2022). Causal links between arctic sea ice and its potential drivers based on the rate of information transfer. *Geophysical Research Letters*, *49*(9), e2021GL095892. https://doi.org/10.1029/2021GL095892

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

Dunne, J., Horowitz, L., Adcroft, A., Ginoux, P., Held, I., John, J., et al. (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. https://doi.org/10.1029/2019ms002015

Dunn-Sigouin, E., Li, C., & Kushner, P. J. (2021). Limited influence of localized tropical sea-surface temperatures on moisture transport into the arctic. *Geophysical Research Letters*, *48*(8), e2020GL091540. https://doi.org/10.1029/2020GL091540

Ebert-Uphoff, I., & Deng, Y. (2012). Causal discovery for climate research using graphical models. *Journal of Climate*, *25*(17), 5648–5665. https://doi.org/10.1175/jcli-d-11-00387.1

England, M. R., Eisenman, I., Lutsko, N. J., & Wagner, T. J. (2021). The recent emergence of arctic amplification. *Geophysical Research Letters*, *48*(15), e2021GL094086. https://doi.org/10.1029/2021gl094086

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., et al. (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. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

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. https://doi.org/10.1175/bams-d-17-0006.1

Francis, J. A., & Vavrus, S. J. (2012). Evidence linking arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, *39*(6). https://doi.org/10.1029/2012gl051000

Galytska, E. (2023). *EyringMLClimateGroup/galytska23jgr_EvaluatingCausa lArcticMidlatTelec: Evaluating causal Arctic-midlatitude teleconnections in CMIP6*. Zenodo. https://doi.org/10.5281/zenodo.8268180

Garfinkel, C. I., Hartmann, D. L., & Sassi, F. (2010). Tropospheric precursors of anomalous northern hemisphere stratospheric polar vortices. *Journal of Climate*, *23*(12), 3282–3299. https://doi.org/10.1175/2010jcli3010.1

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. https://doi.org/10.1038/s41598-021-97615-7

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. https://doi.org/10.5194/bg-17-6115-2020

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129. https://doi.org/10.1029/2018MS001603

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., et al. (2018). Quantifying climate feedbacks in polar regions. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-04173-0

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Graversen, R. G., & Burtu, M. (2016). Arctic amplification enhanced by latent energy transport of atmospheric planetary waves. *Quarterly Journal of the Royal Meteorological Society*, *142*(698), 2046–2054. https://doi.org/10.1002/qj.2802

Graversen, R. G., & Wang, M. (2009). Polar amplification in a coupled climate model with locked albedo. *Climate Dynamics*, *33*(5), 629–643. https://doi.org/10.1007/s00382-009-0535-6

Hadisst (2023). Met Office Hadley Centre for Climate Change [Datatset]. Retrieved from http://www.metoffice.gov.uk/hadobs/hadisst/data/download.html

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*, *127*(7), e2021JD035725. https://doi.org/10.1029/2021jd035725

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. https://doi.org/10.1175/waf-d-16-0124.1

He, B., Bao, Q., Wang, X., Zhou, L., Wu, X., Liu, Y., et al. (2019). CAS FGOALS-F3-l model data sets for CMIP6 historical atmospheric model intercomparison project simulation. *Advances in Atmospheric Sciences*, *36*(8), 771–778. https://doi.org/10.1007/s00376-019-9027-8

Held, I., Guo, H., Adcroft, A., Dunne, J., Horowitz, L., Krasting, J., et al. (2019). Structure and performance of gfdl's cm4. 0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. https://doi.org/10.1029/2019ms001829

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Honda, M., Inoue, J., & Yamane, S. (2009). Influence of low arctic sea-ice minima on anomalously cold Eurasian winters. *Geophysical Research Letters*, *36*(8), L08707. https://doi.org/10.1029/2008gl037079

Hoshi, K., Ukita, J., Honda, M., Iwamoto, K., Nakamura, T., Yamazaki, K., et al. (2017). Poleward eddy heat flux anomalies associated with recent arctic sea ice loss. *Geophysical Research Letters*, *44*(1), 446–454. https://doi.org/10.1002/2016gl071893

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. https://doi.org/10.1038/s41467-018-04138-3

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. https://doi.org/10.1016/j.jmarsys.2009.11.002

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*, *23*(7), e1089. https://doi.org/10.1002/asl.1089

IPCC. (2021). Summary for policymakers [Book Section], In V. Masson-Delmotte, et al. (Eds.). In *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 University Press. https://doi.org/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. https://doi.org/10.1002/2015jd024679

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2

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*, 1–42. https://doi.org/10.5194/egusphere-2022-1013

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., et al. (2014). Weakening of the stratospheric polar vortex by arctic sea-ice loss. *Nature Communications*, *5*(1), 1–8. https://doi.org/10.1038/ncomms5646

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., et al. (2001). The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, *82*(2), 247–268. https://doi.org/10.1175/1520-0477(2001)082<0247:tnnyrm>2.3.co;2

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

Köhler, R. H., Jaiser, R., & Handorf, D. (2023). How do different pathways connect the stratospheric polar vortex to its tropospheric precursors? *Weather and Climate Dynamics Discussions*, 1–20.

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. https://doi.org/10.1029/2019gl083059

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. https://doi.org/10.1175/bams-d-20-0117.1

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. https://doi.org/10.1038/s41612-018-0054-4

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. https://doi.org/10.1175/jcli-d-15-0654.1

Kretschmer, M., Zappa, G., & Shepherd, T. G. (2020). The role of Barents–Kara sea ice loss in projected polar vortex changes. *Weather and Climate Dynamics*, *1*(2), 715–730. https://doi.org/10.5194/wcd-1-715-2020

Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., et al. (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. https://doi.org/10.5194/gmd-13-4205-2020

Li, M., Luo, D., Simmonds, I., Dai, A., Zhong, L., & Yao, Y. (2021). Anchoring of atmospheric teleconnection patterns by arctic sea ice loss and its link to winter cold anomalies in East Asia. *International Journal of Climatology*, *41*(1), 547–558. https://doi.org/10.1002/joc.6637

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., et al. (2022). CMIP6 simulations with the CMCC earth system model (CMCC-ESM2). *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002814. https://doi.org/10.1029/2021ms002814

Ma, X., Mu, M., Dai, G., Han, Z., Li, C., & Jiang, Z. (2022). Influence of arctic sea ice concentration on extended-range prediction of strong and long-lasting ural blocking events in winter. *Journal of Geophysical Research: Atmospheres*, *127*(5), e2021JD036282. https://doi.org/10.1029/2021JD036282

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

Marshall, J., Kushnir, Y., Battisti, D., Chang, P., Czaja, A., Dickson, R., et al. (2001). North Atlantic climate variability: Phenomena, impacts and mechanisms. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *21*(15), 1863–1898. https://doi.org/10.1002/joc.693

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M earth system model version 1.2 (MPI-ESM1. 2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018ms001400

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (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. 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. https://doi.org/10.1002/2014jd022848

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. https://doi.org/10.1038/s41467-020-15195-y

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. https://doi.org/10.1175/jcli-d-14-00297.1

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (SCENARIOMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Outten, S., Li, C., King, M. P., Suo, L., Siew, P. Y., Cheung, H., et al. (2023). Reconciling conflicting evidence for the cause of the observed early 21st century Eurasian cooling. *Weather and Climate Dynamics*, *4*(1), 95–114. https://doi.org/10.5194/wcd-4-95-2023

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. https://doi.org/10.1175/jcli-d-14-00822.1

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. https://doi.org/10.1175/jcli-d-15-0315.1

Peings, Y. (2019). Ural blocking as a driver of early-winter stratospheric warmings. *Geophysical Research Letters*, *46*(10), 5460–5468. https://doi.org/10.1029/2019gl082097

Phillips, A. S., Deser, C., & Fasullo, J. (2014). Evaluating modes of variability in climate models. *Eos, Transactions American Geophysical Union*, *95*(49), 453–455. https://doi.org/10.1002/2014eo490002

Pithan, F., & Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, *7*(3), 181–184. https://doi.org/10.1038/ngeo2071

Polkova, I., Afargan-Gerstman, H., Domeisen, D. I., King, M. P., Ruggieri, P., Athanasiadis, P., et al. (2021). Predictors and prediction skill for marine cold-air outbreaks over the Barents Sea. *Quarterly Journal of the Royal Meteorological Society*, *147*(738), 2638–2656. https://doi.org/10.1002/qj.4038

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. https://doi.org/10.1029/2020gl089933

Previdi, M., Smith, K. L., & Polvani, L. M. (2021). Arctic amplification of climate change: A review of underlying mechanisms. *Environmental Research Letters*, *16*(9), 093003. https://doi.org/10.1088/1748-9326/ac1c29

Pu, Y., Liu, H., Yan, R., Yang, H., Xia, K., Li, Y., et al. (2020). CAS FGOALS-G3 model data sets for the CMIP6 scenario model intercomparison project (scenariomip). *Advances in Atmospheric Sciences*, *37*(10), 1081–1092. https://doi.org/10.1007/s00376-020-2032-0

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., et al. (2022). The arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, *3*(1), 168. https://doi.org/10.1038/s43247-022-00498-3

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*, *108*(D14), 4407. https://doi.org/10.1029/2002jd002670

Riebold, J., Richling, A., Ulbrich, U., Rust, H., Semmler, T., & Handorf, D. (2023). On the linkage between future arctic sea ice retreat, Euro-Atlantic circulation regimes and temperature extremes over Europe. *Weather and Climate Dynamics*, *4*(3), 663–682. https://doi.org/10.5194/wcd-4-663-2023

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., et al. (2020). Earth system model evaluation tool (ESMVALTOOL) v2. 0–technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199. https://doi.org/10.5194/gmd-13-1179-2020

Rong, X., Li, J., Chen, H., Su, J., Hua, L., Zhang, Z., & Xin, Y. (2021). The CMIP6 historical simulation data sets produced by the climate system model CAMS-CSM. *Advances in Atmospheric Sciences*, *38*(2), 285–295. https://doi.org/10.1007/s00376-020-0171-y

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. https://doi.org/10.1063/1.5025050

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

Runge, J. (2022). jakobrunge/tigramite: Tigramite 5.0 [Software]. Zenodo. https://doi.org/10.5281/zenodo.6247837

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., et al. (2019). Inferring causation from time series in earth system sciences. *Nature Communications*, *10*(1), 1–13. https://doi.org/10.1038/s41467-019-10105-3

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series data sets. *Science Advances*, *5*(11), eaau4996. https://doi.org/10.1126/sciadv.aau4996

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., et al. (2015). Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, *6*(1), 1–10. https://doi.org/10.1038/ncomms9502

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. https://doi.org/10.1175/jcli-d-13-00159.1

Sang, X., Yang, X.-Q., Tao, L., Fang, J., & Sun, X. (2022). JAN) Decadal changes of wintertime poleward heat and moisture transport associated with the amplified arctic warming. *Climate Dynamics*, *58*(1–2), 137–159. https://doi.org/10.1007/s00382-021-05894-7

Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*(7), 2514–2519. https://doi.org/10.1002/2014gl059637

Screen, J. A., & Simmonds, I. (2010). The central role of diminishing sea ice in recent arctic temperature amplification. *Nature*, *464*(7293), 1334–1337. https://doi.org/10.1038/nature09051

Screen, J. A. (2017). Far-flung effects of arctic warming. *Nature Geoscience*, *10*(4), 253–254. https://doi.org/10.1038/ngeo2924

Screen, J. A., Deser, C., & Simmonds, I. (2012). Local and remote controls on observed arctic warming. *Geophysical Research Letters*, *39*(10). https://doi.org/10.1029/2012gl051598

Screen, J. A., Deser, C., Smith, D. M., Zhang, X., Blackport, R., Kushner, P. J., et al. (2018). Consistency and discrepancy in the atmospheric response to arctic sea-ice loss across climate models. *Nature Geoscience*, *11*(3), 155–163. https://doi.org/10.1038/s41561-018-0059-y

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. https://doi.org/10.5194/tc-3-11-2009

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. https://doi.org/10.1016/j.gloplacha.2011.03.004

Shepherd, T. G. (2016). Effects of a warming arctic. *Science*, *353*(6303), 989–990. https://doi.org/10.1126/science.aag2349

Shepherd, T. G. (2021). Bringing physical reasoning into statistical practice in climate-change science. *Climatic Change*, *169*(1–2), 2. https://doi.org/10.1007/s10584-021-03226-6

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. https://doi.org/10.5194/wcd-1-261-2020

Siew, P. Y. F., Li, C., Ting, M., Sobolowski, S. P., Wu, Y., & Chen, X. (2021). North Atlantic oscillation in winter is largely insensitive to autumn barents-kara sea ice variability. *Science Advances*, *7*(31), eabg4893. https://doi.org/10.1126/sciadv.abg4893

Silva, F. N., Vega-Oliveros, D. A., Yan, X., Flammini, A., Menczer, F., Radicchi, F., et al. (2021). Detecting climate teleconnections with granger causality. *Geophysical Research Letters*, *48*(18), e2021GL094707. https://doi.org/10.1029/2021gl094707

Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, *9*(1), 62–72. https://doi.org/10.1177/089443939100900106

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 *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., et al. (2019). The Canadian earth system model version 5 (CANESM5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

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. https://doi.org/10.2151/jmsj1965.74.4_579

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Trenberth, K. E., & Hurrell, J. W. (1994). Decadal atmosphere-ocean variations in the pacific. *Climate Dynamics*, *9*(6), 303–319. https://doi.org/10.1007/bf00204745

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. https://doi.org/10.1002/qj.3784

Vavrus, S. J. (2018). The influence of arctic amplification on mid-latitude weather and climate. *Current Climate Change Reports*, *4*(3), 238–249. https://doi.org/10.1007/s40641-018-0105-2

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., et al. (2018). Simulation of the modern climate using the inm-cm48 climate model. *Russian Journal of Numerical Analysis and Mathematical Modelling*, *33*(6), 367–374. https://doi.org/10.1515/rnam-2018-0032

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. https://doi.org/10.1080/16000870.2021.1892435

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. https://doi.org/10.1175/1520-0493(1981)109<0784:titghf>2.0.co;2

Walsh, J. E. (2014). Intensified warming of the arctic: Causes and impacts on middle latitudes. *Global and Planetary Change*, *117*, 52–63. https://doi.org/10.1016/j.gloplacha.2014.03.003

Wang, L., Ting, M., & Kushner, P. (2017). A robust empirical seasonal prediction of winter NAO and surface climate. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-00353-y

Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., et al. (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. https://doi.org/10.5194/gmd-14-3159-2021

Wendisch, M., Brückner, M., Crewell, S., Ehrlich, A., Notholt, J., & Lüpkes, C. (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., et al. (2009). The earth system grid: Enabling access to multimodel climate simulation data. *Bulletin of the American Meteorological Society*, *90*(2), 195–206. https://doi.org/10.1175/2008bams2459.1

Woollings, T., Li, C., Drouard, M., Dunn-Sigouin, E., Elmestekawy, K. A., Hell, M., et al. (2023). The role of rossby waves in polar weather and climate. *Weather and Climate Dynamics*, *4*(1), 61–80. https://doi.org/10.5194/wcd-4-61-2023

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., et al. (2019). The Beijing climate center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*(4), 1573–1600. https://doi.org/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. https://doi.org/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. https://doi.org/10.3390/atmos9090331

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (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., et al. (2020). The Australian earth system model: Access-ESM1. 5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 193–214. https://doi.org/10.1071/es19035