Evaluation of Global Teleconnections in CMIP6 Climate Projections using Complex Networks

Clementine Dalelane[1], Kristina Winderlich[1], and Andreas Walter[1]

[1]Deutscher Wetterdienst Frankfurter Str. 135 6305 Offenbach Germany

Correspondence: Clementine Dalelane (clementin.dalelane@dwd.de)

Abstract. In climatological research, the evaluation of climate models is one of the central research subjects. As an expression of large-scale dynamical processes, global teleconnections play a major role in interannual to decadal climate variability. Their realistic representation is an indispensable requirement for the simulation of climate change, both natural and anthropogenic. Therefore, the evaluation of global teleconnections is of utmost importance when assessing the physical plausibility of climate projections.

We present an application of the graph-theoretical analysis tool δ-MAPS, which constructs complex networks on the basis of spatio-temporal gridded data sets, here sea surface temperature and geopotential height in 500 hPa. Complex networks complement more traditional methods in the analysis of climate variability, like the classification of circulation regimes or empirical orthogonal functions, assuming a new non-linear perspective. While doing so, a number of technical tools and metrics, borrowed from different fields of data science, are implemented into the δ-MAPS framework in order to overcome specific challenges posed by our target problem. Those are trend-EOFs, distance correlation and distance multicorrelation, and the Structural Similarity Index.

δ-MAPS is a two-stage algorithm. In the first place, it assembles grid cells with highly coherent temporal evolution into so-called domains. In a second step, the teleconnections between the domains are inferred by means of the non-linear distance correlation. We construct two unipartite and one bipartite network for 22 historical CMIP6 climate projections and two century-long coupled reanalyses (CERA-20C and 20CRv3). Potential non-stationarity is taken into account by the use of moving time windows. The networks derived from projection data are compared to those from reanalyses. Our results indicate that no single climate projection outperforms all others in every aspect of the evaluation. But there are indeed models, which tend to perform better/worse in many aspects. Differences in model performance are generally low within the geopotential height unipartite networks, but higher in sea surface temperature and most pronounced in the bipartite network representing the interaction between ocean and atmosphere.

1 Introduction

The evaluation of general circulation models (GCM) is one of the key topics of climate sciences. This evaluation is indispensable in the assessment of uncertainties in the projection of climate change. At the same time, it serves as a guideline for further model development.

Established methods of climate model evaluation include comparison of spatial and temporal means, and often also the variability, of important climate parameters such as air temperature, precipitation, wind speed, geopotential height, radiation, and energy fluxes between model output and observational/reanalysis data (Zhang et al., 2021). More elaborated evaluation techniques assess the temporal evolution of global mean/sea surface/hemispheric temperature (Papalexiou et al., 2020) with respect to increasing greenhouse gas concentration or regional trends (Duan et al., 2021).

Acknowledging its importance for consistent climate simulation, Simpson et al. (2020) evaluate the atmospheric circulation in terms of mean atmospheric fields, in combination with dynamical features like the jet stream, stationary waves and blocking. In contrast, Kristóf et al. (2020) evaluated the positions of potential action centers of atmospheric teleconnections as a proxy for circulation.

Another approach is taken by Brands (2022) and Cannon (2020), who both assess circulation biases in correspondence to the representation of circulation types. Whereas Brands (2022) uses Lamb weather types, the analysis in Cannon (2020) is based on Principal Component Analysis (PCA) derived modes of variability. Such modes of variability, extracted by eigentechniques from spatio-temporal gridded data, have been the objective of evaluation efforts in recent years as their spatial patterns are supposed to reflect large scale dynamical processes in the climate system. For example, Fasullo et al. (2020) and Coburn and Pryor (2021) have assessed the representation of six oceanic and atmospheric modes in terms of spatial and spectral accuracy, including an evaluation of the interaction between modes. Still, it has been recognized that eigenmethods suffer from a number of limitations, because geometric constraints as linearity and normality, orthogonality and simultaneity do not correspond to physical properties of the climate system (Monahan et al., 2009; Fulton and Hegerl, 2021; Hynčica and Huth, 2020; Lee et al., 2019) and hinder their interpretation.

Besides, the evaluation of climate modes, such as El Niño Southern Oscillation (ENSO) or North Atlantic Oscillation (NAO), is usually done at the component level. But it is the coupling among those components, which defines the large scale variability of climate at interannual and decadal time scales (Tsonis et al., 2008; Steinhäuser and Tsonis, 2014).

Complex network methods are able to account for non-linear, time-lagged, and high-order interactions in high-dimensional data, and have been introduced in climate sciences by the beginning of the 21st century (for an overview see Dijkstra et al. (2019)). Such networks investigate the interdependencies between all their constituent components, thereby unveiling dynamical features that could remain hidden to traditional analysis techniques. A rather fundamental property of climate networks is their organization in terms of communities—clusters of strongly connected nodes forming semi-autonomous subcomponents of the climate system with non-accidental similarity to many known modes of variability (Steinhäuser et al., 2009; Tsonis et al., 2011), (Tantet and Dijkstra, 2014), that interact dynamically in multiple ways. Such emergent property has been ascribed to the mismatch between spatial and temporal scales on a sphere, which allows only a finite number of degrees of freedom (Yang et al., 2021).

The comparison of such complex network-derived communities between climate simulations and observation/reanalysis data sets was used for evaluation purposes first by Steinhäuser and Tsonis (2014). They assessed the community structure in climatic fields finding rather low consistency between the model runs and the reference data set. Likewise, Fountalis et al. (2015) assessed the community structure of model simulations, but complemented it with an evaluation of the interaction

strength of the communities with ENSO. The idea was further developed by Fountalis et al. (2018) and Falasca et al. (2019) in their so-called δ-MAPS approach to comprise a whole network of all communities, which is evaluated with regards to the distribution and size of communities, the interaction strength and the distribution of the links.

Note that there is another line of research into the evaluation of causal networks (for instance Vázquez-Patiño et al. (2019) or Nowack et al. (2020)), which is somewhat different to the approach followed here.

In the present article, we will explain (Sect. 3) and apply (Sect. 4) δ-MAPS (Fountalis et al., 2018) to construct functional networks for sea surface temperature (SST) and geopotential height in 500 hPa (Z500) fields, as well as a cross-network between SST and Z500, using GCM output data from the Coupled Model Intercomparison Project Phase 6 (CMIP6). We compare the derived networks to analogous networks from reanalysis data, namely CERA-20C (Laloyaux et al., 2018) and 20CRv3 (Slivinski et al., 2019), to evaluate the capacity of the GCMs in reproducing complex non-linear processes in the atmosphere and the ocean.

This assessment is all the more instructive as it is not possible to tune the teleconnections directly. In nature and in models, teleconnections emerge from the interplay of the governing equations under the condition of the boundaries. A model gets them right, if and only if the model specifications are sufficiently well approximated and well balanced between model components.

2 Data

The objective of the present study is to compare the interaction networks derived from CMIP6 GCM output from historical simulations to reference networks derived from *two* century-long reanalyses in order to account for uncertainties in observations and differences in construction methods as recommended by Hynčica and Huth (2020); Lee et al. (2019) and others: (i) the Coupled Reanalysis for the 20$^{\text{th}}$ Century (CERA-20C) provided by the ECMWF (Laloyaux et al., 2018) (10 ensemble members and ensemble mean), and (ii) from the NOAA-CIRES-DOE Twentieth Century Reanalysis version 3 (20CRv3) provided by the NOAA/OAR/ESRL PSL (Slivinski et al., 2019) (best estimate).

The presented study is intended to help the selection of physically plausible GCM runs for further dynamical downscaling in the Coordinated Downscaling Experiment–European Domain (https://www.euro-cordex.net/). Therefore, the CMIP6 model ensemble evaluated here follows the list of model runs under consideration in EURO-CORDEX, for which all necessary forcing data had been provided at the time of writing, plus some extra models (Table 1).

We consider the parameters sea surface temperature (SST) and geopotential height at 500 hPa (Z500). These are relatively well-observed and smoothly varying fields suitable for the construction of networks. Steinhäuser et al. (2012) confirm good network properties for SST and Z500 with many proximity-based correlation links as well as a large number of teleconnections. In accordance, Donges et al. (2011) found the maximal link density for geopotential in about 4 to 6 km height, and Wiedermann et al. (2017) detected the highest transitivity between SST and geopotential height in 500–300 hPa.

From the coupled network perspective, it would be highly desirable to include further parameters into the analysis like sea surface salinity, or, still more interesting, variables from the stratosphere and the deep ocean. Unfortunately, the observations

Table 1. CMIP6 models

Model	hist. exp.	Reference	Model	hist. exp.	Reference
ACCESS-CM2	r1i1p1f1	Bi et al. (2020)	ACCESS-ESM1-5	r1i1p1f1	Ziehn et al. (2020)
BCC-CSM2-MR	r1i1p1f1	Wu et al. (2019)	CanESM5	r1i1p2f1	Swart et al. (2019)
CESM2	r2i1p1f1	Danabasoglu et al. (2020)	CMCC-CM2-SR5	r1i1p1f1	Cherchi et al. (2019)
CMCC-ESM2	r1i1p1f1	Cherchi et al. (2019)	CNRM-CM6-1	r1i1p1f2	Voldoire et al. (2019)
CNRM-ESM2-1	r1i1p1f2	Séférian et al. (2019)	EC-Earth3	r1i1p1f1	Döscher et al. (2022)
EC-Earth3-Veg	r1i1p1f1	Döscher et al. (2022)	HadGEM3-GC31-LL	r1i1p1f3	Roberts et al. (2019)
IPSL-CM6A-LR	r1i1p1f1	Boucher et al. (2020)	MIROC6	r1i1p1f1	Tatebe et al. (2019)
MIROC-ES2L	r1i1p1f2	Hajima et al. (2020)	MPI-ESM1-2-LR	r1i1p1f1	Gutjahr et al. (2019)
MPI-ESM1-2-HR	r1i1p1f1	Müller et al. (2018)	MRI-ESM2-0	r1i1p1f1	Yukimoto et al. (2019)
NorESM2-LM	r1i1p1f1	Seland et al. (2020)	NorESM2-MM	r1i1p1f1	Seland et al. (2020)
TaiESM1	r1i1p1f1	Lee et al. (2020)	UKESM1-0-LL	r1i1p1f2	Sellar et al. (2019)

of such parameters are only recently becoming more reliable and less sparse, such that the fidelity of their reanalyses fields is impossible to verify.

The SST (Z500) data was remapped to a common grid of $2.25° \times 2.25°$ ($2.5° \times 2.5°$) resolution. Regions with sea ice are avoided in SST as well as circles of $5°$ radius around the poles in Z500 because of possibly biased representation of the polar vortices. The analysis is carried out for seasonal anomalies on the overlapping time period from 1901 to 2010.

3 Methods

The procedure used to assign an assessment score to each model run comprises a number of algorithmic stages that build on each other. As they are not yet well-known in the climatological community, we present them in detail in the following subsections:

- Detrending with trend-EOF (Sect. 3.1)
- Network construction with δ-MAPS (Sect. 3.2)
  - Domain identification (Sect. 3.2.1)
  - Network of domains (Sect. 3.2.2)
- Distance covariance and distance correlation (Sect. 3.3)
  - Distance multivariance and distance multicorrelation (Sect. 3.3.1)
- Network comparison with structural similarity index and multivariate network quality score (Sect. 3.4)

3.1 Detrending with trend-EOF

Prior to the construction of the δ-MAPS networks, the data has to be detrended to avoid the correlations being distorted by long-term trends. Although it is still the most widely used technique, linear detrending has been shown little appropriate to remove the effects of external forcing (anthropogenic and natural) from climatic time series (Frankignoul et al., 2017), given its non-linear structure and the dynamical response mechanisms including long-range memory. Conventional Empirical Orthogonal Function (EOF) decomposition is not well suited for trend detection either for a number of reasons (Hannachi, 2007), which often cause the spreading of long-term trends between several modes of internal variability. Instead, we will apply a non-parametric technique, so-called trend-EOF (Hannachi, 2007), which identifies spatial patterns of trends defined as common non-linear, but monotone increase. The method is based on the Singular Value Decomposition (SVD) of the matrix of inverse ranks, instead of the direct observations as in conventional EOF-analysis. Since sequences of inverse ranks provide a robust measure of monotonicity, trend-EOFs are able to separate patterns associated with monotone (non-linear) trends, albeit small, from patterns not associated with trends.

Trend-EOFs have been applied since in a number of studies (e.g. Barbosa and Andersen (2009), Li et al. (2011), Meegan Kumar et al. (2021) among others). Fisher (2015) compared trend-EOFs, along with conventional EOFs, to a selection of other PCA-based techniques, which are designed to extract space-time patterns maximizing criteria like persistence, predictability or autocorrelation. In contrast to conventional EOFs, all the tested methods very robustly detect a leading EOF pattern with a respective principal component (PC) that presents a distinct non-linearly increasing trend. We consider trend-EOFs therefore an appropriate technique for identifying anthropogenic Greenhouse Gas (GHG)-forced trends.

Let $\mathbf{X} = ((x_{it}))$ be the matrix of anomaly data at grid cells $i = 1 \dots n$ (numbered consecutively) and times $t = 1 \dots T$. The time series $\boldsymbol{x}_i$ at grid cell i is transformed to the vector of inverse ranks $\boldsymbol{q}_i$ by setting q_{it} equal to the time position of the t^{th} largest value in $\boldsymbol{x}_i$. The sequence $\boldsymbol{q}_i$ indeed reflects the total monotonicity of $\boldsymbol{x}_i$: in monotone series the inverse ranks are ordered according to the trend. The stronger the trend in $\boldsymbol{x}_i$, the stronger the pattern in $\boldsymbol{q}_i$. By maximizing the correlation in $\mathbf{Q} = ((q_{it}))$, we find a common trend that is shared (to some extent) by all grid cells, which makes sense in the light of GHG-forced warming.

After centering and cosine weighting of $\mathbf{Q}$ w.r.t. the corresponding latitude, the principal components and the loading patterns are obtained by SDV: $\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}$. The trend is now concentrated in the first (few) principal component(s), strongly distinguished by high eigenvalue(s) outstanding over the remaining low and slowly descending spectrum. If second or third order outstanding eigenvalues should be detected, they indicate additional, regionally confined independent trends, which are generated by internal dynamical feedback processes. For our purpose of identifying regions with coherent time evolution, we would therefore want to retain such regional trends and eliminate only the trend associated to the first trend-PC. Likewise, regional trends caused by volcanic eruption are most probably not filtered either by the first trend-EOF. However, the impacts of 20$^{\text{th}}$ century eruptions lasted only for short time periods, and on the other hand they are not well represented in surface-input reanalyses like CERA-20C and 20CRv3 (Fujiwara et al., 2015). We therefore assume that our evaluations remain valid. The first trend-PC $\boldsymbol{u}_1$ is now transformed back to physical space by projection $\boldsymbol{w}_1 = \mathbf{X}\boldsymbol{u}_1$, and the corresponding spatial

pattern is composed of the regression coefficients between the trend-PC $\boldsymbol{w}_1$ and the anomaly time series of the original field $\boldsymbol{x}_i$, $i = 1 \dots n$.

To allow for an annual cycle in the trend patterns, we extend the trend-EOFs in analogy to season-reliant EOFs (Wang and An (2005), see also cyclo-stationary EOFs in Yeo et al. (2017)), $\mathbf{Q} = (\mathbf{Q}_{\text{MAM}}|\mathbf{Q}_{\text{JJA}}|\mathbf{Q}_{\text{SON}}|\mathbf{Q}_{\text{DJF}})$ (seasonally centered, inverse ranks calculated for each season individually), which extract a recurrent sequence of seasonal trend patterns with one associated trend-PC for the magnitude of the whole cycle as opposed to one common pattern for all seasons as in non-seasonal EOF analysis or four individual patterns with their associated individual PCs as in seasonal EOFs, respectively. At this stage it would be possible to apply a secondary SVD to the seasonal warming patterns to obtain a smoother annual cycle. While such procedure seems undue for seasonal data, it would be a reasonable approach in the case of monthly data. Instead of applying two sequential EOFs to $\mathbf{Q}$, a tensor decomposition like HOSVD (De Lathauwer et al., 2000) would serve this purpose more elegantly.

After having detrended the time series, we are able to standardize the seasonal variances without the interference of the seasonal trends, which would otherwise bias our estimates. On their part, seasonally varying variances could degrade the estimated correlations between grid cells in the first stage of the δ-MAPS algorithm, giving increased weight to seasons with higher variance. In turn, the *spatial* component of the variance will be important in the second stage of δ-MAPS, therefore we augment the deseasonalized time series again with their overall (non-seasonal) variance.

3.2 Network construction with δ-MAPS

3.2.1 Domain identification

To infer the functional interactions within and between spatio-temporal gridded datasets of climatological parameters, we adopt the δ-MAPS algorithm proposed by Fountalis et al. (2018). This algorithm is rooted in network sciences/graphical modelling, in which graphs are used to express the dependence structure between random variables. A graph or network consists of a set of nodes connected by a set of edges, which describe the interactions between the nodes. Networks can be classified depending on their topology: simple networks like lattices and fully-connected networks; complex networks like scale-free and small-world networks. Small-world networks are often observed in climate and other earth sciences, in the human brain and in social networks. Their nodes are strongly clustered into semi-autonomous components and the average shortest path length between any two nodes is small.

In contrast to structural networks or flow networks, where the edges are physically observable (like wired connections or trajectories of particles, respectively), functional networks are inferred from the behaviour of the nodes. We consider the grid cells of a selected climatological field as the nodes of the graph. The spatial embedding is naturally given by the locations of the grid cells. In Fountalis et al. (2018) the edges of a fully-connected grid cell-level network are defined using the unpruned Pearson correlation ϱ of the time series as association measure between any pair of nodes. Based on this weighted network, the δ-MAPS algorithm identifies semi-autonomous components $D_1 \dots D_K$, called domains. A domain is a spatially contiguous set of grid cells with highly correlated temporal activity. Fountalis et al. (2018) propose an iterative algorithm that alternately

expands and merges a preliminary set of domain-seeds S (neighborhoods with locally maximal correlation, 3×3 grid cells in our case) so as to find the maximum possible sets of grid cells that satisfy the homogeneity constraint δ: Let D be a spatially contiguous set of grid cells with cardinality $|D|$

$$\delta \leq \varrho_D := \frac{1}{|D|(|D|-1)} \sum_{i \neq j \in D} \varrho_{ij} \tag{1}$$

where ϱ_{ij} is the correlation between the time series at grid cells i and j and δ is a chosen parameter to regulate the number and size of the domains. The domains are expanded to neighboring grid cells (one at a time) as long as $\varrho_D \leq \delta$. Two domains D_i and D_j are merged if they contain at least one pair of adjacent grid cells and their union still satisfies the threshold δ. The algorithm stops, when no more domains can be merged or expanded.

The number of domains K generated by this algorithm is not predefined. Overlapping domains are allowed in δ-MAPS, because grid cells might be influenced by more than one physical process. If a grid cell does not satisfy the homogeneity constraint with any of its neighbors, it remains unassigned. Deviating from Fountalis et al. (2018), we use Spearman's Rank correlation to determine the similarity between grid cells to allow for monotone, yet non-linear association. Furthermore, we set the threshold δ for minimal average correlation within a domain to equal a selected high quantile of all pairwise correlations (our δ is not based on a significance test, therefore there is no need to correct for auto-correlation). Lower thresholds allow the domains to expand and merge further resulting in a smaller number of spatially larger domains, which means lower parcellation, and vice versa. In Sect. 4, we will choose δ so as to produce "intuitive" domains evocative of known teleconnection patterns.

In Falasca et al. (2020), the identification of domains was further refined: grid cells are assigned to a common domain if their time varying complexity (quantified by recurrence entropy) evolves coherently. Coherent evolution of complexity reflects coherent dynamical evolution and is thus an even stronger indicator for semi-autonomous component organisation than correlation between the original climatological time series. But for complexity time series to construct, the proposed recurrence measure has to be evaluated on moving time windows (100 year windows over 6000 years of monthly values in Falasca et al. (2020)). Unfortunately, our time series are not long enough to detect complexity changes by means of recurrence entropy (nor to actually occur in the real climatological fields), so we have to stick to the original definition of δ-MAPS in Fountalis et al. (2018).

The first stage of δ-MAPS is a local community detection algorithm, where the criterion to maximize is the number of grid cells assigned to a minimum number of communities under the conditions (i) $\varrho_D \geq \delta$, (ii) D is spatially contiguous, and (iii) D contains a seed $s \in S$ (Fortunato and Hric, 2016). As this problem is NP-hard (solvable in polynomial time, Fountalis et al. (2018)), the greedy algorithm of Fountalis et al. (2018) only approximates one possible solution. Even though, it is able to detect meaningful communities of any size (no preferred scale) and independently from the network structure in other spatial regions.

3.2.2 Network of domains

Subsequently, the domains identified above serve as super-nodes in the second stage of δ-MAPS. A functional weighted network is inferred between the domains on the basis of a dependence measure (in Fountalis et al. (2018) the lagged Pearson

correlation is used; we will use distance correlation, see Sect. 3.3). The time series of a domain is defined as

$$\boldsymbol{x}_D = (x_{D1} \dots x_{DT}), \;\; x_{Dt} = \frac{1}{\sum_{i \in D} \cos \varphi_i} \sum_{i \in D} x_{it} \cos \varphi_i \tag{2}$$

where φ_i is the latitude of grid cell i. In contrast to Falasca et al. (2019), we use the means instead of the sums of the grid cells for domain time series. We do so, because otherwise the variances of the domains would grow with their size, something that would hinder interpretation. On the other hand, the spatial correlation within the domains, the precondition for grid cells to form a domain, impedes the decrease of the variance of the domain mean following the Central Limit Theorem at the rate of $\sqrt{|D|}$. Instead, the variances of the domain means are of comparable magnitude regardless of the domain size.

Every possible link with every possible lag $-L \leq l \leq L$ is tested for significance, which constitutes a multiple-testing problem such that the cumulative probability of type I errors increases. One way to control the false discovery rate FDR to be smaller than a predefined level α was proposed by Benjamini (2010): the p-levels of the individual tests are ordered ascendingly, $p_{(1)} \leq \dots \leq p_{(\frac{1}{2}K(K-1)(2L+1))}$, and the hypothesis ($H_0$: link is insignificant) is rejected only for those tests, where $p_{(k)} < \frac{2k\alpha}{K(K-1)(2L+1)}$.

The network consists of two maps D and W. D : set of nodes (grid cells) $\longrightarrow$ power set of domains $\mathcal{P}(D_1 \dots D_K)$, which assigns one/several/no domains to every grid cell, and W : set of pairs of domains $\{D_1 \dots D_K\} \times \{D_1 \dots D_K\} \longrightarrow$ maximal (lagged) dependence $\in \mathbb{R}$, which assigns every pair of domains a link that equals the maximal (lagged) dependency between them (we allow lags up to 10 seasons).

The distinction between grid cells that are dependent within the same domain and grid cells that are dependent across two different domains allows δ-MAPS to differentiate between local diffusion phenomena and remote interactions as for instance an atmospheric bridge or an oceanic tunnel (Liu and Alexander, 2007).

Since the techniques to construct the δ-MAPS network are statistical, long time series are convenient in order to obtain robust estimates of the dependence measures. In the case of non-stationarity, such estimates would be biased and reflect only a temporal average connectivity between the components of the network. The time dependence can be addressed using evolving networks, which are constructed over sliding time windows (see for instance Kittel et al. (2021) and Novi et al. (2021)). The present study considers a time-constant network for the period 1901-2010, and a shorter period network for 1951-2010, where more observations are available for assimilation into the reanalyses. To investigate the temporal evolution, a third network is constructed for 1901-1955.

The complex networks framework offers a lot more approaches in order to exploit the richness of the data, as for instance multi-scale, causal, and multi-layer networks. Wavelet multi-scale networks were proposed for investigating interactions in the climate system simultaneously at different temporal scales, revealing features which usually remain hidden when looking at one particular time scale only (Agarwal et al., 2018, 2019). Interactions between processes evolving on different time scales are investigated by Jajcay et al. (2018). Moreover, as the number of identified domains within a climatological field is drastically smaller than the number of original grid cells, this also opens up the possibility of investigating the causal relationships between them (Nowack et al., 2020), although the basic assumption of causal networks inference, that the dependence structure can be represented by a directed acyclic graph, is questionable in the climate context. The construction of both dependence based and

causal networks can naturally be extended to cross-networks, which include multiple fields (Feng et al., 2012; Ekhtiari et al., 2021).

3.3 Distance covariance and distance correlation

As physical processes in climate are highly dynamical and mostly non-linear (Donges et al., 2009), we decided to substitute the Pearson correlation in the second step of network inference by a non-linear dependence measure: distance correlation proposed by Székely et al. (2007). To begin with, distance covariance, calculated from the pairwise Euclidean distances within each sample, is an analogue to the product-moment covariance, but it is zero if and only if the random vectors are independent. The intuition of distance covariance is that if there exists a dependence between the random variables X and Y, then for two similar realizations of X, say x_s and x_t, the two corresponding realizations of Y, y_s and y_t, should be similar as well. Note that the opposite (x_s, x_t unsimilar $\implies$ y_s, y_t unsimilar) is true for linear dependence, but not true in general.

Unlike the widely used information measures, distance covariance has a compact representation, is computationally fast, and reliable in a statistical sense for sample sizes common in climatology, because it is not necessary to estimate the density of the samples. We use the unbiased version of distance covariance given in Székely and Rizzo (2014). Let $(x_t), (y_t)$, $t = 1 \dots T$ be a statistical sample from a pair of real or vector valued random variables X, Y. First, compute all pairwise Euclidean distances:

$$a_{st} = \|x_s - x_t\|_2 \quad \text{and} \quad b_{st} = \|y_s - y_t\|_2,$$

and perform a double centering for all $s \neq t$

$$A_{st} = a_{st} - \frac{1}{T-1}\sum_u a_{su} - \frac{1}{T-1}\sum_v a_{vt} + \frac{1}{(T-1)(T-2)}\sum_{uv} a_{uv}$$

$$B_{st} = b_{st} - \frac{1}{T-1}\sum_u b_{su} - \frac{1}{T-1}\sum_v b_{vt} + \frac{1}{(T-1)(T-2)}\sum_{uv} b_{uv}.$$

Then distance covariance dCov is definded as

$$\text{dCov}(X,Y) := \frac{1}{T(T-3)}\sum_{st} A_{st}B_{st} \tag{3}$$

Distance variance dVar and distance correlation dCor are defined analogously to moment variance and moment correlation, resepctively:

$$\text{dVar}(X) = \text{dCov}(X,X) \quad \text{and} \quad \text{dCor}(X,Y) := \frac{\text{dCov}(X,Y)}{\sqrt{\text{dVar}(X)\text{dVar}(Y)}} \tag{4}$$

Distance correlation has a number of desirable properties:

1. $0 \leq \text{dCor}(X,Y) \leq 1$
2. $\text{dCor}(X,Y) = 0 \iff X$, Y independent
3. $\text{dCor}(X,Y) = 1 \iff Y$ is a linear transformation of X

Distance correlation is furthermore robust against auto-dependence (K. Fokianos and M. Pitsillou, 2018), which eliminates the need to correct for autocorrelation, as it was done in Fountalis et al. (2018). The correction of autocorrelation involves the estimation of a rather large number of autocorrelation coefficients. This might add to statistical uncertainty and its expendebility is therefore statistically advantageous.

An efficient test of distance correlation based on the χ^2-distribution was proposed by Shen et al. (2022), which is universally consistent, and valid for $\alpha \leq 0.05$:

$$\Phi(X,Y) = \begin{cases} 1 & \text{if} \quad T\,\mathrm{dCor}(X,Y) \geq F^{-1}_{\chi^2_1-1}(1-\alpha) \\ 0 & \text{else} \end{cases} \tag{5}$$

Distance correlation is defined between vectors of arbitrary dimension. One way to take advantage of this property in the construction of networks would be to assign the measurement of more than one climatological variable to every node, e.g. sea surface temperature and salinity, or 500 hPA geopotential height and temperature.

We will apply distance correlation in the network inference between the domains, but not in the construction of the domains. The reason is that in domain construction we are looking for similar temporal behavior between grid cells. We choose Spearman's Rank correlation, because it accounts for non-linear, yet monotone association. In contrast, in network inference we are expressly interested in non-linear dependence including non-monotonicity.

3.3.1 Distance multivariance and distance multicorrelation

Distance correlation has also been generalized to distance multivariance/multicorrelation by Böttcher et al. (2019) to measure the dependence between an arbitrary number n of random variables in the sense of Lancaster interaction (Lancaster, 1969; Streitberg, 1990). The Lancaster interaction ΔF quantifies the fraction of dependence between them that is not explained by factorization, their *synergy*. For $n = 3$, let F_{123} be the 3-dimensional joint distribution function of X_1, X_2, X_3, F_{12}, F_{13} and F_{23} the pairwise joint and F_1, F_2 and F_3 the marginal distribution functions. Then the Lancaster interaction is defined as

$$\Delta F = F_{123} - F_1F_{23} - F_2F_{13} - F_3F_{12} + 2F_1F_2F_3$$

the fraction of F_{123} that is not explained by pair-wise dependence. Lancaster interaction excludes, in particular, linear dependence as this is indeed explained by pair-wise dependence.

The concept of higher-order dependence is related to joint cumulants and higher-order moments, in that $\kappa_n(X_1 \dots X_n) = \int x_1 \dots x_n d\Delta F$ (Streitberg, 1990). Joint cumulants are traditionally applied in multiple-point statistics and hyper-spectral analysis to describe non-linear interaction and non-gaussian multidimensional distributions. Climate science has seen only a small number of implementations, including the contributions of C. A. L. Pires related to teleconnections (e.g. Pires and Hannachi (2017, 2021)). As a feature of complex systems, higher-order interactions have already been recognized as critical for the emergence of complex behavior such as synchronization and bifurcation in scientific fields as diverse as social networks science, ecology, molecular biology, quantum physics, neurosciences, epidemics, geodesy, image processing and genetics (Battiston et al., 2020), and tools for the construction of hypergraphs (graphs with links that comprise more than two nodes) are increasingly available. To our knowledge, hypergraphs have not yet been introduced in climatology.

Distance multivariance is defined analogously to distance variance (Equation 3) and is a strongly consistent estimator of Lancaster interaction (Böttcher et al., 2019). For $n = 3$, with C_{st} the analogue to A_{st} and B_{st} for a third random variable Z:

$$\mathrm{dMvar}(X,Y,Z) = \frac{1}{T(T-3)} \sum_{st} A_{st} B_{st} C_{st} \tag{6}$$

and likewise distance multicorrelation with a slightly differing normalisation:

$$\mathrm{dVar}_3(X) := \mathrm{dMvar}(X,X,X) \quad \text{and} \quad \mathrm{dMcor}(X,Y,Z) = \frac{\mathrm{dMvar}(X,Y,Z)}{\Big(\mathrm{dVar}_3(X) \cdot \mathrm{dVar}_3(Y) \cdot \mathrm{dVar}_3(Z)\Big)^{1/3}} \tag{7}$$

Obviously, distance covariance between two random variables is covered by distance multivariance for $n = 2$. Significance tests for distance multivariance are also given in Böttcher et al. (2019). As the asymptotic test is conservative and furthermore, in the case of non-zero pairwise dependence, the test statistic is not guaranteed to diverge, it is convenient to choose a larger FDR level than the usually employed significance levels between 0.1 and 0.01.

3.4 Comparison of networks with structural similarity index and multivariate network quality score

This study aims at comparing the interaction networks derived from CMIP6 model output to the selected reference networks. Our metric of comparison netSSIM is a modification of the netCorr criterion for functional networks developed by Falasca et al. (2019). netCorr is a sophisticated metric, which evaluates the differences in topology and connectivity, combined in the adjacency matrix $\mathbf{M}$ of each network, simultaneously. Let $\mathbf{M} = ((M_{ij}))_{i,j=1}^{n}$ be a square matrix of dimension n (number of grid cells) with

$$M_{ij} := \begin{cases} 0 & \text{if} \quad D(\boldsymbol{x}_i) = \emptyset \ \text{ or } D(\boldsymbol{x}_j) = \emptyset \\ \frac{1}{|\{W(D(\boldsymbol{x}_i), D(\boldsymbol{x}_j)) > 0\}|} \sum\limits_{D_k \in D(\boldsymbol{x}_i), D_l \in D(\boldsymbol{x}_j)} W(D_k, D_l) & \text{if} \quad \emptyset \neq D(\boldsymbol{x}_i) \neq D(\boldsymbol{x}_j) \neq \emptyset \end{cases} \tag{8}$$

where we set $W(D_k, D_l) = \mathrm{dCor}(\boldsymbol{x}_{D_k}, \boldsymbol{x}_{D_l})$. Alternatively, M could be rearranged in a 4-modal hypermatrix or tensor made of the Kronecker product of the lat-lon field times itself containing the dependencies between the grid cells.

Apart from replacing Pearson by distance correlation, our definition of $\mathbf{M}$ differs from the one in Falasca et al. (2019) in three aspects. Firstly, our links are undirectional, because distance correlation is much less sensitive to temporal lag than Pearson correlation. The distance correlation coefficients for lags $-10 \leq L \leq 10$ differ only marginally from the value for $L = 0$. So although we do construct $\mathbf{M}$ using maximum lagged distance correlation, we do not venture to infer the direction of the interaction from it. Secondly, we have defined $W(D_k, D_k) = 1$, causing $M_{ij} = 1$ if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ pertain to the same domain (and no other) to emphasize that grid cells within one domain are more strongly linked to each other than to the grid cells of other domains. Thirdly, we set M_{ij} the average of the links between domains that $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ belong to instead of the maximum as a means to account for overlapping domains. We do not apply any weighting to this average, because the mean internal rank correlation within each domain, i.e. the bond of a grid cell to its domains, is equally $\approx \delta$ by construction.

netCorr between two networks measures the spatial correlation between the respective adjacency matrices, not considering the overall level and variability within the networks. We propose to augment netCorr to netSSIM. SSIM is the Structural Similarity Index, a measure very popular in image processing, which combines terms for brightness (mean), contrast (variance) and structure (pattern correlation) of images (Wang et al., 2004). It was introduced to the hydrological/meteorological community by Mo et al. (2014). Let X, Y be two gridded fields:

$$\text{SSIM}(X,Y) := \frac{2\mu_X\mu_Y + c_1}{\mu_X^2 + \mu_Y^2 + c_1} \cdot \frac{2\sigma_X\sigma_Y + c2}{\sigma_X^2 + \sigma_Y^2 + c2} \cdot \frac{\sigma_{XY} + c_3}{\sigma_X\sigma_Y + c_3} \tag{9}$$

where μ_X, μ_Y are the means, σ_X^2, σ_Y^2 are the variances and σ_{XY} is the Pearson covariance between X and Y, and small constants $c_1 = c_2 = c_3$ (we choose 0.00001) to ensure regularity. The SSIM ranges from -1 to 1, it equals 1 only in case of identity and -1 for an anti-analogue (equal mean and variance, but correlation $= -1$). SSIM $= 0$ means no similarity. Note that the SSIM is not invariant under translation and rotation, which corresponds to our requirements, because we want the teleconnections to sit in the right place. SSIM is not a distance metric, but a distance metric can be constructed from it (Brunet et al., 2011).

Falasca et al. (2019) recommend the use of their netCorr criterion always in combination with a criterion comparing the strength of the interaction, which they define as the sum of the links of a particular domain in terms of covariance. We argue that the strength is a criterion that intermingles the distribution of interactions between the domains with the variances of the domains, which, in turn, are determined by the size of the domains and the variances of the included nodes. We therefore prefer to evaluate the interactions on their own using the netSSIM. The evaluation of the variances (or standard deviations) of model output data is a task that is already routinely performed in conventional evaluation setups.

We apply the (latitude weighted) SSIM to two adjacency matrices $\mathbf{M}$ (Equation 8) constructed from the significant distance correlations in two reference and/or model networks. In this way, we calculate netSSIM indices for the unipartite networks for SST and Z500, and for the cross-networks between the SST and Z500 domains.

Alternatively, we could calculate the SSIM between adjacency matrices in a point-wise manner, comparing the slices of the 4-modal hyper-matrices that correspond to the links of one individual grid cell to all others and then taking the weighted mean of all point-wise SSIMs.

Finally, we define a Network Quality Score (NQS) by applying an exponential transform to the netSSIMs, which projects them to the interval $[0,1]$ (recall that the netSSIM lives on $[-1,1]$). The same transform was used in Sanderson et al. (2015) and Brunner et al. (2020) to construct quality scores from error measures, which are later fed into a model selection algorithm.

$$\text{NQS} := \exp\left\{-(1 - \text{netSSIM})^2\right\} \tag{10}$$

In order to combine the three NQSs wrt. SST, Z500 and SST–Z500, we take the geometric mean (equal to the exponential of the arithmetic mean of the squared differences $(1 - \text{netSSIM})^2$). This shall be the Multivariate Network Quality Score MNQS:

$$\text{MNQS} := \left(\text{NQS}_{\text{SST}} \cdot \text{NQS}_{\text{Z500}} \cdot \text{NQS}_{\text{SST-Z500}}\right)^{\frac{1}{3}} = \exp\left\{-\frac{1}{3}\left\|\mathbf{1} - \text{netSSIM}\right\|^2\right\} \tag{11}$$

The MNQS corresponds to the exponential transform of the squared Euclidean distance between the 3-dimensional vector-netSSIM to the ideal vector-netSSIM value $(1,1,1)$, which would be attained by a network identical to the reference, normalized with the distance between $(1,1,1)$ and the value $(0,0,0)$ indicating no similarity.

Any other vector norm could be utilized for the construction of MNQS, for instance an L_p-norm with $p \neq 2$ or some weighting of the directions. netSSIMs for additional parameters can be incorporated into the MNQS in a straight forward way. Finally, the considered models can be ranked with respect to these scores.

The netSSIM is also useful, when exploring the differences between networks in more detail. As mentioned above, the slices of $\mathbf{M}$ wrt. a single grid cell or domain can be compared one by one. It is further possible to calculate the netSSIM for all pairwise links in a certain region, excluding the rest of the globe, or for all links from one region to another. This way, differences across models or time periods can be tracked down directly to their origin.

4 Results and discussion

We will demonstrate the functioning of every sub-procedure considering the CERA-20C ensemble mean over the whole period 1901–2010 as an example. All procedures are furthermore applied to the periods 1901–1955 and 1951–2010. Individual runs of CERA-20C as well as 20CRv3 and CMIP6 model realisations will be discussed depending on special interest.

4.1 Detrending with trend-EOF

Trend-EOFs (Hannachi, 2007), as introduced in Sect. 3.1, produce time series of common change (in SST and Z500) generated from the trend-PCs in the inverse-rank space, and the respective trend-loading patterns (four seasonal trend-loading patterns per trend-PC in the case of season-reliant/cyclo-stationary trend-EOFs), indicating regions of stronger/weaker change. As expected, the increase in SST is concentrated in the first trend-PC (the leading eigenvalues are 30 to 50 times higher than the trailing ones), the other trend-PCs showing no secular trend. Figure 1(a) depicts the global mean sea surface temperature anomaly (GMSSTa) (wrt. the base period 1961–1990) in the CERA-20C ensemble mean, the forced temperature increase estimated by the first trend-EOF and the detrended anomalies. For comparison, we show the same plot for linearly detrended SSTs in Figure S1. The grid cell-wise detrended anomalies are deseasonalized with regard to variance.

The GMSSTa derived from trend-EOFs in all runs of CERA-20C (not shown) as well as in the ensemble mean show a very similar evolution among each other and to Zhu et al. (2018), the breakpoints in temperature increase postulated therein at 1942, 1975 and 2004 clearly discernible. Likewise, the physical space-loading patterns of the ensemble mean (Figure 1(b)) and all runs of CERA-20C are very similar to each other and resemble the leading modes extracted using slow feature analysis and dynamical mode decomposition in Fulton and Hegerl (2021), identified as warming trends.

Analogous plots for geopotential height anomalies in 500 hPa for the CERA-20C ensemble mean can be found in Figure 1(c) and (d). Unfortunately, we were not able to find any comparable study in the literature, where Z500 was analysed for trend over the 20$^{\text{th}}$ century. Gillett et al. (2013), Knutson and Ploshay (2021), Garreaud et al. (2021), and Raible et al. (2005) considered SLP trends over different time periods and regions. Although not fully comparable, there is a certain similarity.

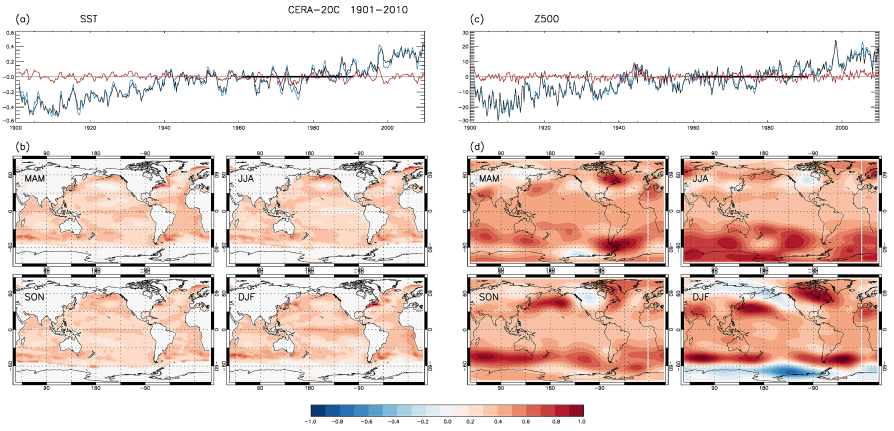

Figure 1. Season-reliant trend-EOF of the CERA-20C SST and Z500 fields over the time period 1901–2010. Global mean SST (a) and global mean Z500 (c) anomaly wrt. base period 1961–1990 (blue), forced component thereof (black), and residual (red). (b) and (d) respective seasonal trend-loading patterns in physical space (arbitrary units normalized to [-1,1] over all seasons).

The projected trends as well as the loading patterns in the 20CRv3 best estimate are somewhat different for the period 1901–2010 (Figure S2), but agree much better for 1951–2010 (not shown). This might well be related to low observational coverage during the first half of the century, we thus take this disagreement as a signal for caution.

When subject to the same procedure, the CMIP6 model output SST and Z500 anomalies produce trend-EOFs and loading patterns roughly similar to CERA-20C and 20CRv3 (not shown). Differences are more or less obvious, though, such that an evaluation of the GMSST anomaly time series in the spirit of Papalexiou et al. (2020) would be an obvious choice, but is out of the scope of this paper.

4.2 δ-MAPS for CERA-20C on 1901–2010

4.2.1 Domain identification

Our algorithm, presented in Sect. 3.2.1, combines grid cells with highly rank correlated time evolution into domains. Domains have to be contiguous, but may overlap, grid cells may remain unassigned. Average mutual rank correlation within a domain has to be higher than a selected threshold δ, we examined the quantiles $q_{0.9}(\varrho_{ij}|i \neq j = 1 \ldots n) \leq \delta \leq q_{0.99}(\varrho_{ij}|i \neq j = 1 \ldots n)$ of all pairwise rank correlations. The plots included in this paper refer to thresholds $q_{0.95}$ for SST, and $\delta = q_{0.93}$ for Z500, chosen for their intuitive parcellation of the fields evocative of known teleconnection patterns. As varying the threshold effects the networks for different data sets in a similar way, the choice of δ changes the results only marginally.

The domains constructed this way from the detrended, deseasonalized SST anomalies of the CERA-20C ensemble mean include all important SST teleconnection patterns with interannual to decadal time scales (see for example Messié and Chavez (2011)). The map of the CERA-20C SST domains (Figure 2(a)) resembles the corresponding maps for COBEv2 and HadISST

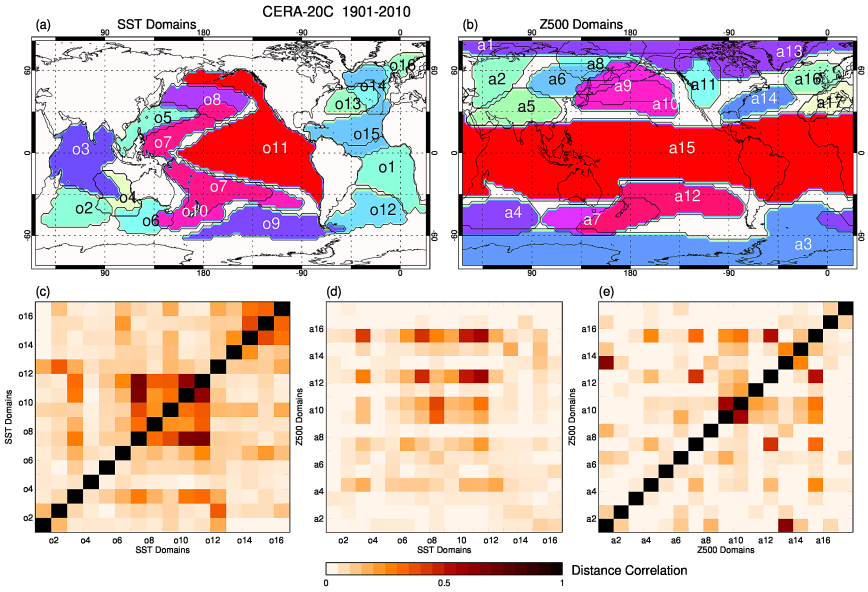

Figure 2. Domains of the CERA-20C ensemble mean (a) SST and (b) Z500 fields over the time period 1901–2010 (arbitrary colors). Maximum lagged distance correlation links between (c) SST and (e) Z500 domains and (d) cross-links.

in Falasca et al. (2019) reasonably well, taking into account the differing data sets and time periods. Their main domains are clearly identifiable: El Niño Southern Oscillation (ENSO, o11, for its broad extension also reminiscent to the region 2 of the Interdecadal Pacific Oscillation [IPO] tripol in Henley et al. (2015)); the Horse Shoe Pattern (o7); the South Pacific (o9); the Indian Ocean (o3); the North Tropical Atlantic (o15, with extension to the extratropics); the South Tropical Atlantic (o1). Furthermore there are domains in the extra-tropical southern (o2) and eastern Indian Ocean (o4), the extra-tropical southern (o12) and north-eastern (o14) Atlantic and the Norwegian Sea (o16), the Gulf Stream (o13), the North Pacific current (o8, region 1 of the IPO tripole), a domain corresponding to region 3 of the IPO tripole (o10), the Kuroshio Extension (o5), and a domain south of Australia including the Great Australian Bight (o6). Areas, where sea ice occurs, are omitted, because of the confounding effect on SST.

In the CERA-20C Z500 map of domains (Figure 2(b)), the seasonally migrating Tropical Belt (TB, a15) formed by the Hadley circulation and the two polar cells (Arctic a1/a13 largely overlapping, and Antarctic a3) stand out, stretching around the whole globe. The mid-latitudes are populated by numerous domains with more (over ocean) or less (over land) pronounced zonal extension (cyclone tracks). The missing segmentation of the tropical belt into several domains probably results from the seasonal time resolution.

4.2.2 Network of domains

The domains of SST and Z500 are now ready for network construction (see Sect. 3.2.2): Figure 2(c) illustrates the maximum lagged ($-10 \leq L \leq 10$) distance correlations (Sect. 3.3) for all pairs of SST domains in the CERA-20C ensemble mean, omitting the geographic information for enhanced clarity. Only significant links to the FDR level $\alpha = 0.05$ (Sect. 3.2.2) are shown. However, even weak links are assessed significant, because the time series are long enough (4 seasons $\times$ 110 years) to allow the distance correlation to be estimated accurately. The darkest shades (except of the self links) correspond to the links between ENSO (o11), the Horse Shoe (o7) and IPO3 (o10): o11↔o7, o11↔o10, o7↔o10. We see enhanced connectivity of o7, o10 and o11 to the northern and southern Pacific Ocean (o8, o9) and from the Pacific to the Indian Ocean (o7/o10/o11↔o3), corresponding to known ENSO-teleconnections, but not to the Kuroshio Extension extension (o5). The southern Indian Ocean domain is furthermore linked to the South Atlantic (o2↔o12).

The intra-Atlantic links are much weaker: o14, o15 and o16 are largely overlapping domains and together supposedly form the Atlantic Multidecadal Oscillation (AMO); the Gulf Stream is linked to the north-eastern Atlantic (o13↔o14) as well as the tropical to the extra-tropical South Atlantic (o1↔o12). The South Atlantic is also weakly connected to all North Atlantic domains (o12↔o13/o14/o15/o16), but there is no link between the North Atlantic and the South Tropical Atlantic (o1). It might be hypothesized that this bypass is related to the thermohaline circulation that tunnels the shallow subtropical cell (Liu and Alexander, 2007). According to the network, the Atlantic is connected to the other oceans only via the Southern Ocean, with links o12/o13/o14/o16↔o9, o14/o15/o16↔o6, o1/o12↔o2, and o16↔o2 that appear rather weak, although visible against their virtually zero background. A link between the South Tropical Atlantic (o1) and ENSO (o11) as proposed in Falasca et al. (2019) and Rodríguez-Fonseca et al. (2009) is not apparent in our network. This absence is likely caused by the non-stationarity of this link, which was not observed before 1970. Nevertheless, it does appear, when a network is constructed for the period 1971–2010 (not shown).

Note that allowing for lagged dependence changes the network only marginally compared to a net with only instantaneous links. Few connections are increased in strength of distance correlation by more than 0.05, and none by more than 0.1. All links already exist in the instantaneous network, the structure of the network remains unchanged.

The network between CERA-20C Z500 domains (Figure 2(e)) is considerably weaker than the SST network, possibly a consequence of the stronger high-frequency variability of the Z500 time series in response to seasonally varying solar forcing combined with weaker low-frequency variability caused by stronger mixing of the freely flowing air masses. Moreover, many of the known atmospheric teleconnections vary considerably throughout the year, which weakens the all-season dependence between the involved domains. Apart from the overlapping domains a1/a13, a9/a10, and a7/a12, the Tropical Belt (a15) is the most strongly connected domain with links to the mid-latitudinal Ferrel cell domains, enveloping the cyclone tracks, over all oceans (a15↔a4/a7/a9/a10/a12/a14), to which the undisturbed Hadley circulation releases a substantial amount of energy. Domains over land have fewer and weaker links. Known atmospheric teleconnections are clearly identifiable: the Pacific North America Pattern (PNA) with links a10↔a11, a10↔a14, but interestingly not a11↔a14, and the North Atlantic Oscillation (NAO) with a link a13↔a14 (and much weaker a1↔a14). Other complex teleconnections also seem to involve the

Arctic domains: a1↔a2/a6/a8/a16 and a13↔a8. In contrast, the Antarctic domain (a3), is largely autonomous as discussed in Spensberger et al. (2020). Lagged dependence is irrelevant in the Z500 network.

We notice that many known atmospheric teleconnections are defined as higher-order modes of some EOF decomposition. As such they exist only as additive modulations of their corresponding leading modes. We would therefore not expect to find many of them in our networks.

Network methods allow the investigation of interactions between different climatological fields in a straight forward way, constructing cross-networks between (in our case) SST and Z500 domains that describe the coupled ocean–atmosphere variability (Liu and Alexander, 2007). We notice that the inference of links between the domains of two unipartite networks is different from the construction of bipartite communities in multi-layer networks as in Ekhtiari et al. (2021). Here, we just calculate the distance correlations between pairs of one SST and one Z500 domain. The inferred CERA-20C SST–Z500 cross-links are shown in Figure 2(d). The connectivity is mostly quite weak, except for the cross-links from the Tropical Belt (a15) and the northern and southern Pacific Z500 domains (a9, a10, a12) to the ENSO-related SST domains (o7, o8, o9, o10, o11), the tropical Indian Ocean (o3), but also the Great Australian Bight (o6). This feature was also observed by Feng et al. (2012), who related it to the Walker circulation. Z500 domains a4 and a7 participate in this pattern, but to a lesser extent. Z500 domains over oceans are usually connected to their underlying SST counterparts (a4↔o2, a7↔o6, a9/a10↔o8, a14↔o13 [SST modulating the NAO], a15↔o3/o11), although in the Atlantic this dependence is exceptionally weak (a15↔o1/o15, a16↔o14, a3↔o12). But teleconnections to more distant SST domains are, in some instances, as strong as or even stronger than those proximate cross-links (a4↔o3/o10/o11/o12, a7↔o12, a14↔o15). Interestingly, the Arctic Z500 domain (a13) is weakly linked to the AMO-domain (o15), but not to the North Pacific. Except for slightly increased overall connectivity levels, supposedly mediated by the SST, allowing for lagged dependence does not change the network.

The analogous plot for 20CRv3 can be found in Figure S3.

4.2.3 3rd-order interactions

As before, this subsection presents only results for CERA-20C over the time period 1901–2010. The overall high level of connectivity between SST domains motivated us to take a deeper look into the dependence structure of the climate system. In a modest first attempt, we search for interacting triples in the sense of Lancaster, in graph theory termed as 2-hyperedges, taking all combinations of three SST domains and calculating their distance 3^{rd}-order multicorrelation as introduced in Sect. 3.3.1 Equation 7. As discussed there, we choose a large FDR level $\alpha = 0.2$ in order not to suppress too many distance multicorrelations. To avoid cumbersome evaluations with different lag combinations, we stick to instantaneous networks.

Only a small number (13) of significant 3^{rd}-order dependencies is detected (we list them in Table 2 instead of plotting them), all somehow related to the ENSO phenomenon, one of them the IPO tripole itself. The hyperedges also include the tropical Indian Ocean (o3) and the Great Australian Bight (o6). As the nature of Lancaster Interaction is inherently non-linear, this concentration on ENSO corresponds to the findings in Hlinka et al. (2014), who detect substantial non-linear contributions to mutual information in SST (apart from trends and seasonal variance) mainly in the central tropical Pacific. Likewise, Pires and Hannachi (2017) find synchronized extremes of uncorrelated PCs of SST in the Pacific that cannot be explained by linear

Table 2. Significant interaction strength between domain triples in CERA-20C over 1901–2010.
dMcor=distance multicorrelation, ∑dCor=sum of pairwise distance correlations

SST	SST	Z500	dMcor	∑dCor	SST	SST	SST	dMcor	∑dCor	SST	Z500	Z500	dMcor	∑dCor
o3	o7	a15	0.188	1.080	o3	o7	o10	0.183	1.179	o7	a12	a15	0.188	1.349
o3	o10	a15	0.200	1.140	o3	o7	o11	0.150	1.159	o8	a9	a10	0.236	1.268
o3	o11	a15	0.193	1.156	o3	o10	o11	0.160	1.099	o10	a12	a15	0.246	1.442
o7	o11	a10	0.174	1.108	o6	o7	o11	0.119	0.869	o11	a10	a15	0.191	1.081
o7	o10	a12	0.242	1.552	o6	o10	o11	0.130	0.867	o11	a12	a15	0.274	1.580
o7	o11	a12	0.283	1.655	o7	o8	o10	0.162	1.108					
o7	o8	a15	0.183	0.935	o7	o8	o11	0.199	1.209					
o7	o10	a15	0.281	1.529	o7	o9	o10	0.150	1.301					
o7	o11	a15	0.300	1.601	o7	o9	o11	0.195	1.359					
o8	o10	a15	0.165	0.899	o7	o10	o11	0.403	1.902					
o8	o11	a15	0.186	1.037	o8	o10	o11	0.173	1.054					
o9	o11	a15	0.146	1.023	o9	o10	o11	0.165	1.211					
o10	o11	a10	0.184	1.028	o14	o15	o16	0.118	0.968					
o10	o11	a12	0.313	1.627										
o10	o11	a15	0.327	1.585										

interaction. Even though, one distance multicorrelation is also detected in the North Atlantic: the SST triple (o14,o15,o16), which corresponds to the AMO.

Note that not every triple with strong pairwise dependencies also has a significant 3rd-order dependence. Table 2 shows the hyperedges along with their distance multicorrelation and the sum of their pairwise distance correlations. As distance multicorrelation is symmetric, every significant hyperedge is listed only once in the table. Note also that the sum of pairwise distance correlations is not bounded by 1, because the pairwise dependencies are not mutually exclusive. Although the detected distance multicorrelations are significant, they are at most 20% of the sum of the respective pairwise distance correlations. That means 3rd-order interactions complement, but not outweigh pairwise dependence in the 3-dimensional joint dependence.

The same comments essentially apply to cross-hyperedges consisting of two SST and one Z500 domains or one SST and two Z500 domains. We detected 15 and 5 significant cross-hyperedges, respectively, in the Pacific, which all resemble some ENSO interaction. The Z500 domains a13 and a14 (NAO) have no notable distance multicorrelation with North Atlantic SST domains, indicating that the North Altlantic is linked to the NAO-domains on a pairwise basis (o15↔a13/a14), but no higher-order interaction is taking place. There is no hyperedge of three Z500 domains with significant multicorrelation. Known atmospheric tripoles like the Arctic Oscillation (a9, a13, a14) and the Pacific North America Pattern (a10, a11, a14) apparently lack significant 3rd-order dependence.

We believe that the construction of higher-order networks including hyperedges by means of distance multicorrelation might well be one step towards understanding the synergies emerging from multivariate coupling of large-scale oceanic/atmospheric teleconnections.

4.3 Comparison of networks

4.3.1 Reference networks

We turn to the comparison of reference networks in terms of the NQS and MNQS criteria (see Sect. 3.4), calculated from the adjacency matrices **M** containing the regionally distributed distance correlation links between all pairs of domains (Equation 8). As CERA-20C was produced as a 10-member ensemble representing the inevitable sampling and modelling uncertainty inherent in the production process, we take this opportunity to construct the δ-MAPS networks individually for each member. The results are matched to the networks derived for the CERA-20C ensemble mean.

The CERA-20C individual networks for the complete time period 1901–2010 are very similar to each other, with average NQSs close to 1 for all three parameters (average $\mathrm{NQS}_{\mathrm{SST}} = 0.98$, average $\mathrm{NQS}_{\mathrm{Z500}} = 0.94$, average $\mathrm{NQS}_{\mathrm{SST\text{–}Z500}} = 0.96$), such that the MNQSs have mean 0.96 with only a small spread. The average MNQS between the individual CERA-20C runs and the CERA-20C ensemble mean is 0.96. The small differences are brought about by the pattern correlation factor in netSSIM, the mean and variance factor being virtually equal to 1.

The networks for the shorter periods 1901–1955 and 1951–2010 are equally similar with average $\mathrm{MNQS} = 0.95/0.95$ between runs and $0.95/0.96$ to the ensemble mean, respectively. Because the networks for individual CERA-20C runs and the CERA-20C ensemble mean are nearly indistinguishable, we will only take the CERA-20C ensemble mean networks for reference in the following comparisons.

When analysing the temporal evolution of the connectivity in the CERA-20C ensemble mean, we find good agreement between the first and second half of the century ($\mathrm{MNQS} = 0.87$, Table 3), resulting from comparable differences in the SST and SST–Z500 networks, and higher similarity in Z500 ($\mathrm{NQS}_{\mathrm{SST}} = 0.84$, $\mathrm{NQS}_{\mathrm{Z500}} = 0.93$, $\mathrm{NQS}_{\mathrm{SST\text{–}Z500}} = 0.84$). In contrast, the full period is more similar to the first half in all networks ($\mathrm{MNQ} = 0.96$, $\mathrm{NQS}_{\mathrm{SST}} = 0.95$, $\mathrm{NQS}_{\mathrm{Z500}} = 0.96$, $\mathrm{NQS}_{\mathrm{SST\text{–}Z500}} = 0.95$), than to the second half, because especially the SST–Z500 networks bear more differences ($\mathrm{MNQS} = 0.92$, $\mathrm{NQS}_{\mathrm{SST}} = 0.92$, $\mathrm{NQS}_{\mathrm{Z500}} = 0.94$, $\mathrm{NQS}_{\mathrm{SST\text{–}Z500}} = 0.89$). We emphasize that the networks contain only information about the strength of the dependencies between the domains and not about their functional form.

Because of deviating domain extension and numbering, comparing the networks by means of the rectangular network plots (like in Figure 2(c)–(e)) is cumbersome. In Figure 3 we have plotted 2-modal slices of the spatially distributed adjacency hypermatrices **M** wrt. grid cells in the ENSO-domain, in the AMO-domain and in the Tropical Belt, respectively. The comparison of these slices is evidently not exhaustive, but may give a hint on the nature of the differences between the networks.

The domains in the three CERA-20C SST network slices for the ENSO-domain (Figure 3 (a),(b),(c)) are very similar in shape and size, but the links between the domains are differently distributed. The networks most obviously disagree in link

Table 3. Multivariate Network Quality Scores between reanalyses in various time periods

	CERA-20C			20CRv3		
	1901–1955	1951–2010	1901–2010	1901–1955	1951–2010	1901-2010
CERA-20C						
1901–1955		0.87	0.96	0.88		
1951–2010	0.87		0.92		0.89	
1901–2010	0.96	0.92				0.82
20CRv3						
1901–1955	0.88				0.81	0.84
1951–2010		0.89		0.81		0.90
1901–2010			0.82	0.84	0.90	

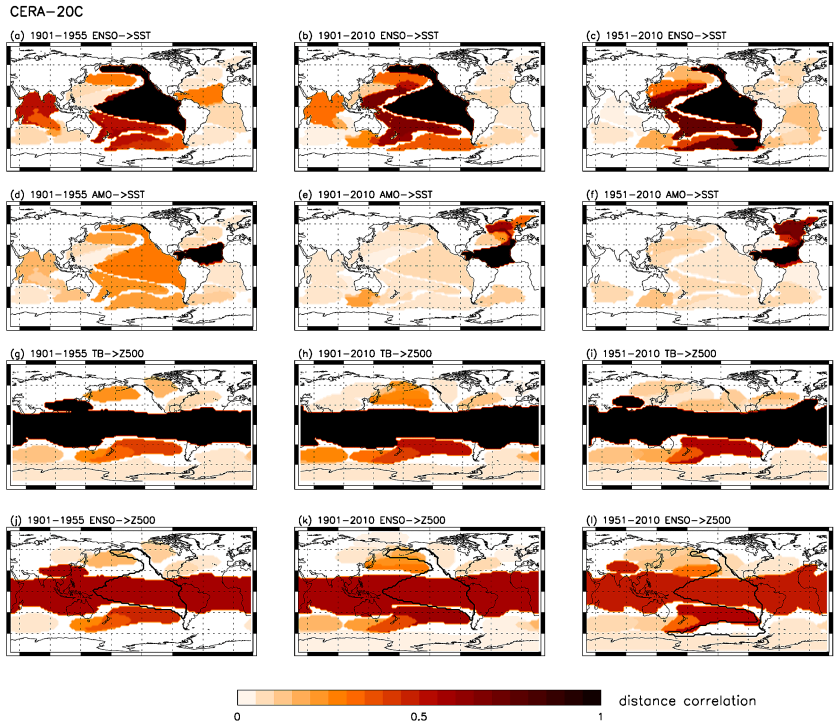

Figure 3. Spatially distributed maximum lagged distance correlation links and cross-links between SST and/or Z500 domains in CERA-20C. (a)–(c) ENSO (black) to SST domains; (d)–(f) North Tropical Atlantic (AMO, black) to SST domains; (g)–(i) Tropical Belt (TB, black) to Z500 domains; (j)–(l) ENSO (contoured) to Z500 domains. (a),(d),(g),(j) time period 1901–1955, (b),(e),(h),(k) time period 1901–2010, (c),(f),(i),(l) time period 1951–2010.

strength from ENSO to the tropical Indian Ocean, but also from ENSO to the North Tropical Atlantic, to the North Pacific, and to the Southern Ocean. The same is visible in the network slices for the AMO-domain (Figure 3(d),(e),(f)).

In contrast, the CERA-20C Z500 network slices for the Tropical Belt (Figure 3(g),(h),(i)) bear more apparent similarity than the SST network slices, which was already apparent in the Network Quality Scores above. Although the shape of the Tropical Belt differs slightly more than the shape of the ENSO-domain, the links to the rest of the globe resemble each other more strongly. However, the domains over the North Pacific and the southern Indian Ocean seem somewhat ambiguous.

The cross-links from ENSO to the Z500 domains (Figure 3(j),(k),(l)) and from the Tropical Belt to the SST domains (not shown) show similar differences like the unipartite networks. Yet, the stabilizing effect of the self-links (large patches with distance correlation 1) does not apply to the SST–Z500 cross-networks, such that the network scores may turn out a little lower.

As regards the second reanalysis 20CRv3, we observe strong similarity to the CERA-20C ensemble mean on the two shorter time periods 1951–2010 and 1901–1955 (MNQS$= 0.89$ and MNQS$= 0.88$, Table 3 and Figure S4), where disagreement within the same time period is mainly restricted to higher southern latitudes (remember that the SSIM includes an area weighting). But dissimilarities between the first and the second half of the century are stronger in 20CRv3 than in CERA-20C (MNQS$= 0.81$ and MNQS$= 0.87$, Table 3). Notably, in 1901–1955 20CRv3 shows the same strong connection between SST domains around the whole tropics as CERA-20C, which is lost in 1951–2010 in both reanalyses. In contrast, the similarity between 20CRv3 and CERA-20C is slightly reduced in 1901–2010 (MNQS$= 0.82$, Table 3 and Figure S4) mainly due to differing atmospheric interactions and the weaker cross-links in 20CRv3 compared to CERA-20C ($\text{NQS}_{\text{SST}}= 0.94$, $\text{NQS}_{\text{Z500}}= 0.81$, $\text{NQS}_{\text{SST–Z500}}= 0.72$). In all three networks (SST, Z500, SST–Z500) we observe that regional unsimilarity increases with latitude. Table S1 shows pairs of most similar domains between CERA-20C and 20CRv3 along with their domain-wise network quality score.

Besides, the similarity between the different time periods in 20CRv3 is not the same as in CERA-20C, with 1901–2010 more similar to 1951–2010 than to 1901–1955 (MNQS$= 0.84$ and MNQS$= 0.90$, Table 3 and Figure S4). For example, in contrast to CERA-20C, the link between ENSO and the South Pacific vanishes after 1950 in 20CRv3. This might be a consequence of sparse observations in the first half of the century and thus a stronger dynamical heritage from the models used to produce the reanalyses. On the other hand, there might have been changes in connectivity driven by increasing GHG levels, which are not equally reflected in CERA-20C and 20CRv3 (they are model results after all). Caution leads us therefore to restrict the comparison of CMIP6 data sets to reanalyses to the period 1951–2010.

4.3.2 CMIP6 networks

The networks belonging to the CMIP6 historical projections (listed in Table 1) are compared in Figure 4 to the CERA-20C ensemble mean (bold black cross marks) and to the 20CRv3 best estimate (bold red cross marks) on the time period 1951–2010 in terms of individual network NQSs (for SST networks (a), for Z500 networks (b) and for the cross-networks (c)) and in terms of MNQSs for each reference respectively (d). Finally we take the average of both MNQSs to account for the uncertainty inherent in the reanalyses: $1/2(\text{MNQS}(\text{CERA-20C}) + \text{MNQS}(\text{20CRv3}))$ ((e), bold crossmarks). As expected, the similarity between models and references is generally weaker than between references, although in the Z500 networks some

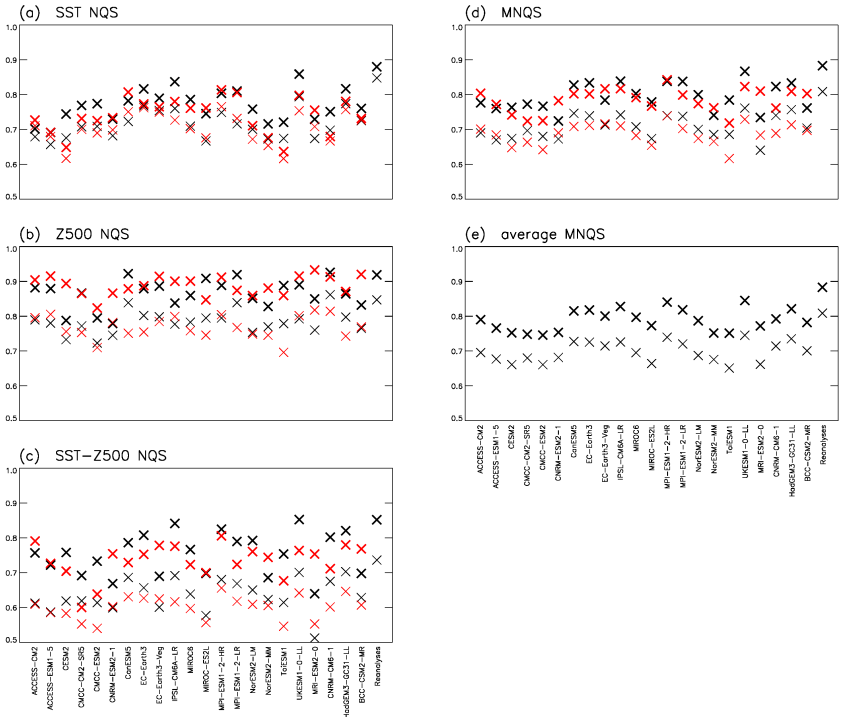

Figure 4. Network Quality Scores (bold) and point-wise Network Quality Scores (thin) of CMIP6 Models wrt. CERA-20C (black) and 20CRv3 (red) over the time period 1951–2010. (a) networks for SST fields, (b) networks for Z500 fields, (c) cross-networks between SST and Z500 fields. (d) Multivariate Networks Quality Scores, (e) average of the two Multivariate Network Quality Scores for CERA-20C and for 20CRv3.

models reach a comparable level. Network quality scores are highest for Z500, followed by SST and SST–Z500. SST–Z500 cross-networks show the greatest deviations across models as well as across references. The seemingly contradictory scores for Z500 wrt. CERA-20C and 20CRv3 have to be put into perspective with their very high values and can be traced back to the differences between the reanalyses.

When applying the alternative, point-wise SSIM calculation (Figure 4 black and red thin cross marks), the final average MNQS values are somewhat lower in their overall level, but similar in spread, and the model ranking suffers only minor changes.

The differences between the reanalyses are also reflected in the MNQSs of the models, where the reanalyses agree very well upon some models (HadGEM3-GC31-LL, IPSL-CM6A-LR, MPI-ESM11-2-HR, MIROC-ES2L, MIROC6), but less upon others (MRI-ESM2-0, TaiESM1, CNRM-CM6-1, CNRM-ESM2-1). But altogether, a tendency to differentiate between more/less similar models wrt. reanayses is clearly visible. We conclude that, when combining several references from independent sources, the average MNQS over these references is a valid evaluation instrument for assessing, whether the teleconnec-

tions between large climate components in a general circulation model are realistically represented. Still, as our evaluation is restricted to a single run per model, we are not able to differentiate between good runs and good models as such.

Using the example of four of the highest-ranking GCM runs wrt. MNQS, we illustrate in short the opportunities offered by the δ-MAPS approach to detect model deficiencies. We examine some of the point-wise adjacency maps of EC-Earth3, UKESM1-0-LL, MPI-ESM1-2-HR and IPSL-CM6A-LR in comparison to CERA-20C and 20CRv3 over 1951–2010 (Figures S5–S9). In the SST networks, we notice that differences are not restricted to higher latitudes, as was the case for the two reanalyses. Even in the main feature of interannual variability, ENSO, spatial connectivity deviates significantly. In all models the tropical Indian Ocean depends much more strongly, although to varying degrees, on ENSO than in both reanalyses (Figure S5). EC-Earth3 and IPSL-CM6A-LR not at all reproduce the northern extension of the ENSO domain seen in both reanalyses (Figure S5(a),(d),(e) and (f)), which reflects the widely recognized low-frequency interdependency between ENSO and the Pacific Decadal Oscillation (PDO) (Henley et al., 2015). The links to the southern Indian Ocean and the South Atlantic differ considerably across models, but no model shows better performance in all domains. In MPI-ESM1-2-HR the dependence between AMO and ENSO is exaggerated, whereas in IPSL-CM6A-LR the Norwegian Sea is nearly disconnected from the Tropical North Atlantic, which is not consistent with AMO (Figure S6(c) and (d)). As regards Z500, UKESM1-0-LL shows an unrealistic link between the Tropical Belt and the Antarctic domain (Figure S7(b)). At the same time, the dependence of the Arctic domain is matched well *only* in UKESM1-0-LL (Figure S8(b)). In contrast, the cross-links from ENSO to Z500 are well-represented in all four models (Figure S9).

Continuing the analysis of all point-wise adjacency maps, it would be possible to identify regions/climate phenomena of higher and lower confidence in any model. An exercise that might be instructive for both modelling groups and downstream users of climate projections.

5 Conclusion

In order to evaluate the physical plausibility of CMIP6 GCM output, we have constructed functional interaction networks within and between the SST and Z500 multivariate time series of two reanalyses (CERA-20C and 20CRv3) and 22 GCM output data sets using the δ-MAPS procedure. In response to several theoretical challenges related to the nature of long term climate data, a number of innovations were introduced into δ-MAPS:

- Detrending with season-reliant trend-EOFs
- Network construction using distance correlation
- Distance multicorrelation for higher-order interactions
- Network comparison with the structural similarity index
- Construction of a multi-reference multivariate network quality score

First of all, the two reanalyses were compared to one another in considerable detail, including the temporal evolution of the interactions in the course of the 20th century. It could not be excluded that inconsistencies between the first and second half of the century arise at least partly from data uncertainty. The evaluation of CMIP6 model output against the references revealed a very high general similarity of the atmospheric connectivity, though with gradual differences. Oceanic teleconnections are less accurately reflected and the model differences more pronounced. The strongest deviations are found in the cross-networks between Z500 and SST, which co-occur sometimes, but not always, with lower network quality scores in the unipartite networks. We combined the three network quality scores for each CMIP6 model on an equal basis, emphasizing the equivalent importance of all considered geophysical subsystems in the generation of the earth's climate. Taking into account the uncertainty inherent in any reference, the average multivariate network quality score over several, preferably independent, references can certainly be considered a suitable criterion to assess the similarity of physical interactions between climate components in a model to those in observations.

In addition, the proposed complex networks framework combined with the distance correlation measure offers many promising multivariate extensions of δ-MAPS as, for example, node definition based on multivariate time series, consideration of higher-order dependence, interactions on multiple time scales and time-evolving networks. Such comparisons could be very useful to investigate subtle differences between various reanalyses. Besides, the characterization of network evolution from past to future could add a new facet to the understanding of climate change.

Code and data availability. The δ-MAPS software can be obtained in https://github.com/FabriFalasca/delta-MAPS. The GCM data used in this study is part of the World Climate Research Programme's (WCRP) 6th Coupled Model Intercomparison Project (CMIP6) open access data. It was accessed through the Earth System Grid Federation (ESGF, https://esgf-node.llnl.gov/search/cmip6/). CERA-20C data is available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/cera-20c. 20CRv3 data is available at https://psl.noaa.gov/data/gridded/data.20thC_ReanV3.html.

Author contributions. CD developed the concept, processed the data, prepared the manuscript, and produced all figures; KW and AW contributed with in-depth discussions, interpretation and review.

Competing interests. The contact author has declared that neither she nor her co-authors have any competing interests.

References

Agarwal, A., Maheswaran, R., Marwan, N., Caesar, L., and Kurths, J.: Wavelet-based multiscale similarity measure for complex networks, Eur. Phys. J. B, 91, 296, https://doi.org/10.1140/epjb/e2018-90460-6, 2018.

Agarwal, A., Caesar, L., Marwan, N., Maheswaran, R., Merz, B., and Kurths, J.: Network-based identification and characterization of teleconnections on different scales, Sci. Rep., 9, 8808, https://doi.org/10.1038/s41598-019-45423-5, 2019.

Barbosa, S. M. and Andersen, O. B.: Trend patterns in global sea surface temperature, Int. J. Climatol., 29, 2049–2055, https://doi.org/10.1002/joc.1855, 2009.

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., M., L., Patania, A., Young, J.-G., and Petri, G.: Networks beyond pairwise interactions: Structure and dynamics, Phys. Rep., 874, 1–92, https://doi.org/10.1016/j.physrep.2020.05.004, 2020.

Benjamini, Y.: Discovering the false discovery rate, J. R. Statist. Soc. B, 72(4), 405–416, https://doi.org/10.1111/j.1467-9868.2010.00746.x, 2010.

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Sullivan, A., Bodman, R., Law, R., Harman, I., Srbinovsky, J., Rashid, H. A., Dobrohotoff, P., Mackallah, C., Yan, H., Hirst, A., Savita, A., Boeira Dias, F., Woodhouse, M., Fiedler, R., and Heerdegen, A.: 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, 2020.

Böttcher, B., Keller-Ressel, M., and Schilling, R.: Distance multivariance: New dependence measures for random vectors, Ann. Stat., 47(5), 2757–2789, https://doi.org/10.1214/18-AOS1764, 2019.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, L. E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, J. Adv. Model. Earth Sy., 12(7), e2019MS002 010, https://doi.org/https://doi.org/10.1029/2019MS002010, 2020.

Brands, S.: A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the northern hemisphere, Geophys. Model Dev., 15(4), 1375–1411, https://doi.org/10.5194/gmd-15-1375-2022, 2022.

Brunet, D., Vrscay, E. R., and Wang, Z.: A Class of Image Metrics Based on the Structural Similarity Quality Index, in: Image Analysis and Recognition, edited by Kamel, M. and Campilho, A., vol. 6753 Part 1, pp. 100–110, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-21593-3_11, 2011.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R.: Reduced global warming from CMIP6 projections when weighting models by performance and independence, Earth Syst. Dynam., 11(4), 995–1012, https://doi.org/10.5194/esd-11-995-2020, 2020.

Cannon, A. J.: Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: CMIP5 to CMIP6, Environ. Res. Lett., 15(6), 064 006, https://doi.org/10.1088/1748-9326/ab7e4f, 2020.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model, J. Adv. Model. Earth Sy., 11(1), 185–209, https://doi.org/https://doi.org/10.1029/2018MS001369, 2019.

Coburn, J. and Pryor, S. C.: Differential Credibility of Climate Modes in CMIP6, J. Climate, 34(20), 8145 – 8164, https://doi.org/10.1175/JCLI-D-21-0359.1, 2021.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Sy., 12(2), e2019MS001 916, https://doi.org/10.1029/2019MS001916, 2020.

De Lathauwer, L., De Moor, B., and Vandewalle, J.: A Multilinear Singular Value Decomposition, SIAM J. Matrix Anal. A., 21(4), 1253–1278, https://doi.org/10.1137/S0895479896305696, 2000.

Dijkstra, H. A., Hernández-García, E., Masoller, C., and Barreiro, M.: Networks in Climate, Cambridge University Press, Cambridge, https://doi.org/10.1017/9781316275757, 2019.

Donges, F. J., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods, Eur. Phys. J. Spec. Top., 174, 157–179, https://doi.org/10.1140/epjst/e2009-01098-2, 2009.

Donges, J., Schultz, H., Marwan, N., Zou, Y., and Kurths, J.: Investigating the topology of interacting networks: Theory and application to coupled climate subnetworks, Eur. Phys. J. B, 84(4), 635–651, https://doi.org/10.1140/epjb/e2011-10795-8, 2011.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6, Geosci. Model Dev., 15(7), 2973–3020, https://doi.org/10.5194/gmd-2020-446, 2022.

Duan, Y., Kumar, S., and Kinter, J. L.: Evaluation of Long-Term Temperature Trend and Variability in CMIP6 Multimodel Ensemble, Geophys. Res. Lett., 48(10), e2021GL093 227, https://doi.org/10.1029/2021GL093227, 2021.

Ekhtiari, N., Ciemer, C., Kirsch, C., and Donner, R.: Coupled network analysis revealing global monthly scale co-variability patterns between sea-surface temperatures and precipitation in dependence on the ENSO state, Eur. Phys. J.-Spec. Top., 230(14-15), 3019–3032, https://doi.org/10.1140/epjs/s11734-021-00168-z, 2021.

Falasca, F., Bracco, A., Nenes, A., and Fountalis, I.: Dimensionality reduction and network inference for climate data using δ-MAPS: Application to the CESM Large Ensemble sea surface temperature, J. Adv. Model. Earth Sy., 11, 1479–1515, https://doi.org/10.1029/2019MS001654, 2019.

Falasca, F., Crétat, J., and Braconnot, P. Bracco, A.: Spatiotemporal complexity and time-dependent networks in sea surface temperature from mid- to late Holocene, Eur. Phys. J. Plus, 135(5), 392, https://doi.org/10.1140/epjp/s13360-020-00403-x, 2020.

Fasullo, J. T., Phillips, A. S., and Deser, C.: Evaluation of Leading Modes of Climate Variability in the CMIP Archives, J. Climate, 33(13), 5527 – 5545, https://doi.org/10.1175/JCLI-D-19-1024.1, 2020.

Feng, A., Gong, Z., Wang, Q., and Feng, G.: Three-dimensional air–sea interactions investigated with bilayer networks, Theor. Appl. Climatol., 109(3-4), 635–643, https://doi.org/10.1007/s00704-012-0600-7, 2012.

Fisher, M. J.: Predictable Components in Australian Daily Temperature Data, J. Climate, 28(15), 5969–5984, https://doi.org/10.1175/JCLI-D-14-00713.1, 2015.

Fortunato, S. and Hric, D.: Community detection in networks: A user guide, Phys. Rep., 659, 1–44, https://doi.org/10.1016/j.physrep.2016.09.002, 2016.

Fountalis, I., Bracco, A., and Dovrolis, C.: ENSO in CMIP5 simulations: network connectivity from the recent past to the twenty-third century, Clim. Dynam., 45, 511–538, https://doi.org/10.1007/s00382-014-2412-1, 2015.

Fountalis, I., Dovrolis, C., Bracco, A., Dilkina, B., and Keilholz, S.: δ-MAPS: from spatio-temporal data to a weighted and lagged network between functional domains, Appl. Netw. Sci., 3, 21, https://doi.org/10.1007/s41109-018-0078-z, 2018.

Frankignoul, C., Gastineau, G., and Kwon, Y.-O.: Estimation of the SST Response to Anthropogenic and External Forcing and Its Impact on the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation, J. Climate, 30(24), 9871 – 9895, https://doi.org/10.1175/JCLI-D-17-0009.1, 2017.

Fujiwara, M., Hibino, T., Mehta, S., Gray, L., Mitchell, D., and Anstey, J.: Global temperature response to the major volcanic eruptions in multiple reanalysis data sets, Atmos. Chem. Phys., 15(23), 13 507–13 518, https://doi.org/10.5194/acp-15-13507-2015, 2015.

Fulton, D. J. and Hegerl, G. C.: Testing Methods of Pattern Extraction for Climate Data Using Synthetic Modes, J. Climate, 34(18), 7645 – 7660, https://doi.org/10.1175/JCLI-D-20-0871.1, 2021.

Garreaud, R., Clem, K., and Veloso, J.: The south pacific pressure trend dipole and the southern blob, J. Climate, 34(18), 7661–7676, https://doi.org/10.1175/JCLI-D-20-0886.1, 2021.

Gillett, N. P., Fyfe, J. C., and Parker, D.: Attribution of observed sea level pressure trends to greenhouse gas, aerosol, and ozone changes, Geophys. Res. Lett., 40(10), 2302–2306, https://doi.org/10.1002/grl.50500, 2013.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., Haak, H., and Stössel, A.: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP), Geosci. Model Dev., 12(7), 3241–3281, https://doi.org/10.5194/gmd-12-3241-2019, 2019.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks, Geosci. Model Dev., 13(5), 2197–2244, https://doi.org/10.5194/gmd-13-2197-2020, 2020.

Hannachi, A.: Pattern hunting in climate: a new method for finding trends in gridded climate data, Int. J. Climatol., 27, https://doi.org/10.1002/joc.1375, 2007.

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., and Folland, C. K.: A Tripole Index for the Interdecadal Pacific Oscillation, Clim. Dynam., 45(11-12), 3077–3090, https://doi.org/10.1007/s00382-015-2525-1, 2015.

Hlinka, J., Hartman, D., Vejmelka, M., Novotná, D., and Paluš, M.: Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity, Clim. Dynam., 42(7-8), 1873–1886, https://doi.org/10.1007/s00382-013-1780-2, 2014.

Hynčica, M. and Huth, R.: Modes of atmospheric circulation variability in the Northern Extratropics: A comparison of five reanalyses, J. Climate, 33(24), 10 707–10 726, https://doi.org/10.1175/JCLI-D-19-0904.1, 2020.

Jajcay, N., Kravtsov, S., Sugihara, G., Tsonis, A. A., and Paluš, M.: Synchronization and causality across time scales in El Niño Southern Oscillation, npj Clim. Atmos. Sci., 1, 33, https://doi.org/10.1038/s41612-018-0043-7, 2018.

K. Fokianos, K. and M. Pitsillou, M.: Testing independence for multivariate time series via the auto-distance correlation matrix, Biometrika, 105(2), 337–352, https://doi.org/10.1093/biomet/asx082, 2018.

Kittel, T., Ciemer, C., Lotfi, N., Peron, T., Rodrigues, F., Kurths, J., and Donner, R.: Evolving climate network perspectives on global surface air temperature effects of ENSO and strong volcanic eruptions, Eur. Phys. J.-Spec Top., 230(14-15), 3075–3100, https://doi.org/10.1140/epjs/s11734-021-00269-9, 2021.

Knutson, T. R. and Ploshay, J.: Sea Level Pressure Trends: Model-Based Assessment of Detection, Attribution, and Consistency with CMIP5 Historical Simulations, J. Climate, 34(1), 327–346, https://doi.org/10.1175/JCLI-D-19-0997.1, 2021.

Kristóf, E., Barcza, Z., Hollós, R., Bartholy, J., and Pongrácz, R.: Evaluation of Historical CMIP5 GCM Simulation Results Based on Detected Atmospheric Teleconnections, Atmosphere, 11(7), 723, https://doi.org/10.3390/atmos11070723, 2020.

Laloyaux, P., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Boisseson, E., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H., Kosaka, Y., Martin, M., Poli, P., Rayner, N., Rustemeier, E., and Schepers, D.: CERA-20C: A coupled reanalysis of the twentieth century, J. Adv. Model. Earth Sy., 10, 1172–1195, https://doi.org/10.1029/2018MS001273, 2018.

Lancaster, H. O.: The Chi-squared Distribution, Wiley & Sons, Inc., New York, 1969.

Lee, J., Sperber, K., Gleckler, P., Bonfils, C., and Taylor, K.: Quantifying the agreement between observed and simulated extratropical modes of interannual variability, Clim. Dynam., 52(7-8), 4057–4089, https://doi.org/10.1007/s00382-018-4355-4, 2019.

Lee, W.-L., Wang, Y.-C., Shiu, C.-J., Tsai, I., Tu, C.-Y., Lan, Y.-Y., Chen, J.-P., Pan, H.-L., and Hsu, H.-H.: Taiwan Earth System Model Version 1: description and evaluation of mean state, Geosci. Model Dev., 13(9), 3887–3904, https://doi.org/10.5194/gmd-13-3887-2020, 2020.

Li, G., Ren, B., Zheng, J., and Yang, C.: Trend Singular Value Decomposition Analysis and Its Application to the Global Ocean Surface Latent Heat Flux and SST Anomalies, J. Climate, 24(12), 2931–2948, https://doi.org/10.1175/2010JCLI3743.1, 2011.

Liu, Z. and Alexander, M.: Atmospheric bridge, oceanic tunnel, and global climatic teleconnections, Rev. Geophys., 45(2), RG2005, https://doi.org/10.1029/2005RG000172, 2007.

Meegan Kumar, D., Tierney, J. E., Bhattacharya, T., Zhu, J., McCarty, L., and Murray, J. W.: Climatic drivers of deglacial SST variability in the eastern Pacific, Paleoceanography and Paleoclimatology, 36, e2021PA004 264, https://doi.org/10.1029/2021PA004264, 2021.

Messié, M. and Chavez, F.: Global Modes of Sea Surface Temperature Variability in Relation to Regional Climate Indices, J. Climate, 24(16), 4314–4331, https://doi.org/10.1175/2011JCLI3941.1, 2011.

Mo, R., Ye, C., and Whitfield, P. H.: Application Potential of Four Nontraditional Similarity Metrics in Hydrometeorology, J. Hydrometeorol., 15(5), 1862–1880, https://doi.org/10.1175/JHM-D-13-0140.1, 2014.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H. P., B., S. D., and North, G. R.: Empirical Orthogonal Functions: The Medium is the Message, J. Climate, 22(24), 6501 – 6514, https://doi.org/10.1175/2009JCLI3062.1, 2009.

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Esch, F. B. M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., Tian, F., and Marotzke, J.: A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR), J. Adv. Model. Earth Sy., 10(7), 1383–1413, https://doi.org/10.1029/2017MS001217, 2018.

Novi, L., Bracco, A., and Falasca, F.: Uncovering marine connectivity through sea surface temperature, Sci. Rep., 11(1), 8839, https://doi.org/10.1038/s41598-021-87711-z, 2021.

Nowack, P., Runge, J., Eyring, V., and Haigh, J. D.: Causal networks for climate model evaluation and constrained projections, Nat. Commun., 11(1), 1415, https://doi.org/10.1038/s41467-020-15195-y, 2020.

Papalexiou, S. M., Rajulapati, C. R., Clark, M. P., and F., L.: Robustness of CMIP6 Historical Global Mean Temperature Simulations: Trends, Long-Term Persistence, Autocorrelation, and Distributional Shape, Earth's Future, 8(10), e2020EF001 667, https://doi.org/10.1029/2020EF001667, 2020.

Pires, C. A. L. and Hannachi, A.: Independent subspace analysis of the sea surface temperature variability: Non-Gaussian sources and sensitivity to sampling and dimensionality, Complexity, 3076810, https://doi.org/10.1155/2017/3076810, 2017.

Pires, C. A. L. and Hannachi, A.: Bispectral analysis of nonlinear interaction, predictability and stochastic modelling with application to ENSO, Tellus A, 73(1), 1–30, https://doi.org/10.1080/16000870.2020.1866393, 2021.

Raible, C., Stocker, T., Yoshimori, M., Renold, M., Beyerle, U., Casty, C., and Luterbacher, J.: Northern Hemispheric trends of pressure indices and atmospheric circulation patterns in observations, reconstructions, and coupled GCM simulations, J. Climate, 18(19), 3968–3982, https://doi.org/10.1175/JCLI3511.1, 2005.

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., Jackson, L. C., Kuhlbrodt, T., Mathiot, P., Roberts, C. D., Schiemann, R., Seddon, J., Vannière, B., and Vidale, P. L.: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments, Geosci. Model Dev., 12(12), 4999–5028, https://doi.org/10.5194/gmd-12-4999-2019, 2019.

Rodríguez-Fonseca, B., Polo, I., García-Serrano, J., Losada, T., Mohino, E., Mechoso, C. R., and Kucharski, F.: Are Atlantic Ninños enhancing Pacific ENSO events in recent decades?, Geophys. Res. Lett., 36(20), L20 705, https://doi.org/10.1029/2009GL040048, 2009.

Sanderson, B. M., Knutti, R., and Caldwell, P.: A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble, J. Climate, 28(13), 5171–5194, https://doi.org/10.1175/JCLI-D-14-00362.1, 2015.

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J.-F., Moine, M.-P., Msadek, R., Ribes, A., Rocher, M., Roehrig, R., Salas-y Mélia, D., Sanchez, E., Terray, L., Valcke, S., Waldman, R., Aumont, O., Bopp, L., Deshayes, J., Éthé, C., and Madec, G.: Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate, J. Adv. Model. Earth Sy., 11(12), 4182–4227, https://doi.org/https://doi.org/10.1029/2019MS001791, 2019.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y.-C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations, Geosci. Model Dev., 13(12), 6165–6200, https://doi.org/10.5194/gmd-13-6165-2020, 2020.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, J. Adv. Model. Earth Sy., 11(12), 4513–4558, https://doi.org/10.1029/2019MS001739, 2019.

Shen, C., Panda, S., and Vogelstein, J.: The Chi-Square Test of Distance Correlation, J. Comput. Graph. Stat., 31(1), 254–262, https://doi.org/10.1080/10618600.2021.1938585, 2022.

Simpson, I. R., Bacmeister, J., Neale, R. B., Hannay, C., Gettelman, A., Garcia, R. R., Lauritzen, P. H., Marsh, D. R., Mills, M. J., Medeiros, B., and Richter, J. B.: An Evaluation of the Large-Scale Atmospheric Circulation and Its Variability in CESM2 and Other CMIP Models, J. Geophys. Res.-Atmos., 125, e2020JD032 835, https://doi.org/10.1029/2020JD032835, 2020.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C. nd Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Q. J. Roy. Meteor. Soc., 145(724), 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Spensberger, C., Reeder, M. J., Spengler, T., and Patterson, M.: The Connection between the Southern Annular Mode and a Feature-Based Perspective on Southern Hemisphere Midlatitude Winter Variability, J. Climate, 33(1), 115–129, https://doi.org/10.1175/JCLI-D-19-0224.1, 2020.

Steinhäuser, K. and Tsonis, A.: A climate model intercomparison at the dynamics level, Clim. Dynam., 42, 1665–1670, https://doi.org/10.1007/s00382-013-1761-5, 2014.

Steinhäuser, K., Chawla, N. V., and Ganguly, A. R.: An Exploration of Climate Data Using Complex Networks, in: Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data, SensorKDD'09, pp. 23–31, Association for Computing Machinery, Paris, https://doi.org/10.1145/1601966.1601973, 2009.

Steinhäuser, K., Ganguly, A. R., and Chawla, N. V.: Multivariate and multiscale dependence in the global climate system revealed through complex networks, Clim. Dynam., 39, 889–895, https://doi.org/10.1007/s00382-011-1135-9, 2012.

Streitberg, B.: Lancaster Interactions Revisited, Ann. Stat., 18, 1878–1885, https://doi.org/10.1214/aos/1176347885, 1990.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12(11), 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019.

Székely, G. J. and Rizzo, M. L.: Partial distance correlation with methods for dissimilarities, Ann. Stat., 42(6), 2382–2412, https://doi.org/10.1214/14-AOS1255, 2014.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K.: Measuring and Testing Dependence by Correlation of Distances, Ann. Stat., 35(6), 2769–2794, https://doi.org/10.1214/009053607000000505, 2007.

Tantet, A. and Dijkstra, H. A.: An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature, Earth Syst. Dynam., 5, 1–14, https://doi.org/10.5194/esd-5-1-2014, 2014.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M.: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, Geosci. Model Dev., 12(7), 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019, 2019.

Tsonis, A. A., Swanson, K. L., and Wang, G.: On the Role of Atmospheric Teleconnections in Climate, J. Climate, 21(12), 2990 – 3001, https://doi.org/10.1175/2007JCLI1907.1, 2008.

Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A., and da Fontura Costa, L.: Community structure and dynamics in climate networks, Clim. Dynam., 37, 933–940, https://doi.org/10.1007/s00382-010-0874-3, 2011.

Vázquez-Patiño, A., Campozano, L., Mendoza, D., and Samaniego, E.: A causal flow approach for the evaluation of global climate models, Int. J. Climatol., 40(10), 4497–4517, https://doi.org/10.1002/joc.6470, 2019.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J., Michou, M., Moine, M., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, J. Adv. Model. Earth Sy., 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Wang, B. and An, S.-I.: A method for detecting season-dependent modes of climate variability: S-EOF analysis, Geophys. Res. Lett., 32(15), L15 710, https://doi.org/10.1029/2005GL022709, 2005.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: From error visibility to structural similarity, IEEE T Image Pprocess., 13(4), 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.

Wiedermann, M., Donges, J. F., Handorf, D., Kurths, J., and Donner, R. V.: Hierarchical structures in Northern Hemispheric extratropical winter ocean–atmosphere interactions, Int. J. Climatol., 37, 3821–3836, https://doi.org/10.1002/joc.4956, 2017.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y., and Liu, X.: The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6, Geosci. Model Dev., 12(4), 1573–1600, https://doi.org/10.5194/gmd-12-1573-2019, 2019.

Yang, P., Wang, G., Xiao, Z., Tsonis, A. A., Feng, G., Liu, S., and Zhou, X.: Climate: a dynamical system with mismatched space and time domains, Clim. Dynam., 56, 3305–3311, https://doi.org/10.1007/s00382-021-05646-7, 2021.

Yeo, S.-R., Yeh, S.-W., Kim, K.-Y., and Kim, W.-M.: The role of low-frequency variation in the manifestation of warming trend and ENSO amplitude, Clim. Dynam., https://doi.org/10.1007/s00382-016-3376-0, 2017.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M.: 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, 97(5), 931–965, https://doi.org/10.2151/jmsj.2019-051, 2019.

Zhang, M., Xu, Z., Han, Y., and Guo, W.: An improved multivariable integrated evaluation method and tool (MVIETool) v1.0 for multimodel intercomparison, Geosci. Model Dev., 14(5), 3079–3094, https://doi.org/10.5194/gmd-14-3079-2021, 2021.

Zhu, X., Dong, W., Wei, Z., Guo, Y., Gao, X., Wen, X., Yang, S., Zheng, Z., Yan, D., Zhu, Y., and Chen, J.: Multi-decadal evolution characteristics of global surface temperature anomaly data shown by observation and CMIP5 models, Int. J. Climatol., 38(3), 1533–1542, https://doi.org/10.1002/joc.5264, 2018.

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y.-P., and Srbinovsky, J.: 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, 2020.