Validation of boreal summer tropical-extratropical causal links in seasonal forecasts

Giorgia Di Capua[1,2], Dim Coumou[2,3,4], Bart van den Hurk[3,5], Antje Weisheimer[6,7], Andrew G. Turner[8,9]
10 and Reik V. Donner[1,2]

[1]Department of Water, Environment, Construction and Safety, Magdeburg-Stendal University of Applied Sciences, Magdeburg, 39114, Germany
[2]Earth System Analysis, Potsdam Institute for Climate Impact Research (PIK) – Member of the Leibniz Association, Potsdam, 14473, Germany
20 [3]Water and Climate Risk Department, Institute for Environmental Studies (IVM), VU University of Amsterdam, Amsterdam, 1081 HV, Netherlands
[4]Royal Netherlands Meteorological Institute (KNMI), De Bilt, 3730 AE, Netherlands
[5]Department of Flood Risk Management, Deltares, Delft, 2629 HV, Netherlands
[6]European Centre for Medium-Range Weather Forecasts, Reading, RG2 9AX, UK
25 [7]National Centre for Atmospheric Science, University of Oxford, Oxford, OX1 4BH, United Kingdom
[8]National Centre for Atmospheric Science, University of Reading, Reading, RG6 6UR, United Kingdom
[9]Department of Meteorology, University of Reading, Reading, RG6 6UR, United Kingdom

35 *Correspondence to*: Giorgia Di Capua (dicapua@pik-potsdam.de)

Abstract. Much of the forecast skill in the mid-latitudes on seasonal timescales originates from deep convection in the tropical belt. For boreal summer, such tropical-extratropical teleconnections are less well understood compared to winter. Here we validate the representation of boreal tropical – extratropical teleconnections in a general circulation model in comparison with observational data. To characterise variability between tropical convective activity and mid-latitude circulation, we identify the South Asian monsoon (SAM) – circumglobal teleconnection (CGT) pattern and the western North Pacific summer monsoon (WNPSM) – North Pacific high (NPH) pairs as the leading modes of tropical-extratropical coupled variability in both reanalysis (ERA5) and seasonal forecast (SEAS5) data. We calculate causal maps, an application of the Peter and Clark momentary conditional independence (PCMCI) causal discovery algorithm, which identifies causal links in a 2D field, to show the causal effect of each of these patterns on circulation and convection in the Northern Hemisphere. The spatial patterns and signs of the causal links in SEAS5 closely resemble those seen in ERA5, independent of the initialization date of SEAS5. However, the strength of causal links in SEAS5 (β values ~ 0.1-03) is often too weak (about two thirds of those in ERA5, β values ~ 0.2-0.4). By performing a subsampling experiment (over time), we identify those regions for which SEAS5 data well reproduce ERA5 values, e.g., the south-eastern US, and highlight those where the bias is more prominent, e.g. North Africa. We demonstrate that different El Niño – Southern Oscillation phases have only a marginal effect on the strength of these links. Finally, we discuss the potential role of model mean-state biases in explaining differences between SEAS5 and ERA5 causal links.

1 Introduction

Seasonal forecasts provide a useful tool to study atmospheric dynamics and predict seasonal variations in wind, rainfall and temperature patterns across tropical and extratropical regions (Bauer et al., 2015; Palmer and Anderson, 1994). To a certain extent, seasonal forecasts can be used by stakeholders and governments to anticipate and mitigate extreme weather events, failures in crop yields, or water scarcity hazards for infrastructure such as electricity grids (Lazo et al., 2009; Challinor et al., 2003; Hagger et al., 2018; Meza et al., 2008). Tropical – extratropical interactions are linked to mid-latitude boreal surface weather conditions and represent a source of predictability at seasonal and subseasonal timescales (Shukla, 1998). Hence, improving the representation of these teleconnections in seasonal forecasts can help to improve our knowledge of atmospheric dynamics as well as helping to better forecast relevant weather patterns to support early warning.

Obtaining reliable seasonal forecasts is a challenging problem due to the intrinsic nonlinearity of processes governing atmospheric motions (Holton, 1973). While providing weather forecasts beyond a two-week threshold is a complex problem due to the chaotic nature of atmospheric processes (Tsonis and Eisner, 1989; Palmer and Anderson, 1994), slowly varying climatic fields such as sea surface temperatures (SST) and soil moisture can provide forecast skill beyond the weekly time scale (Charney and Shukla, 1981). The representation of the interaction of the atmosphere with other components of the climate system, e.g., SST, is an important requirement to achieve forecast skill (Roberts et al., 2021; Tietsche et al., 2020). Historically, both statistical (Gadgil et al., 2005; Kumar, 2012) and dynamical approaches (Jain et al., 2018; Scaife et al., 2018) have been used to provide seasonal forecasts, often with comparable skill (Seo et al., 2009; Barnston et al., 1999). However, when the focus is on the representation of physical processes rather than the forecast skill, dynamical forecasts, generated by general circulation models (GCMs), provide a more complete representation of the atmospheric physics that governs weather and climate behaviour (Shukla et al., 2000). Dynamical seasonal forecasts explicitly resolve dynamic and thermodynamic equations and are better suited for representing the dynamic and thermodynamic processes and emerging dynamical teleconnections within the climate system.

To produce accurate seasonal forecasts, GCMs need to represent the physical processes operating at those timescales truthfully in the current climate. Great progress in this field has been made in recent decades, leading to an improved representation of dynamic and thermodynamic processes and a steady increase in model resolution (Bauer et al., 2015; Palmer, 2017; Haarsma et al., 2016). Nevertheless, probabilistic reliability of seasonal forecasts remains limited (Weisheimer and Palmer, 2014). In this context, analysing seasonal forecasts such as the SEAS5 (Johnson et al., 2019) dataset provided by the European Centre for Medium-range Weather Forecasts (ECMWF), can help to identify, understand and improve biases between observations and model simulations.

Tropical–extratropical interactions in the Northern Hemisphere during boreal summer have been analysed in several recent studies (O'Reilly et al., 2018; Di Capua et al., 2020a; Ding et al., 2011). Heat generated by tropical convective activity provides a source of wave activity that can affect weather in the mid-latitudes (O'Reilly et al., 2018; Rodwell and Hoskins, 1996; Ding and Wang, 2005), while in turn mid-latitude wave activity can modulate rainfall events in the tropical belt (Ding

and Wang, 2007; Di Capua et al., 2020b). Here, we focus on the two main modes of covariability between tropical convection and mid-latitude circulation as identified in Di Capua et al. (2020a) and Ding et al. (2011). The first mode of covariability is represented by the circumglobal teleconnection (CGT) paired with the South Asian monsoon (SAM) convection (Ding et al., 2011; Di Capua et al., 2020a). Circumglobal wave trains such as the CGT are connected to temperature and precipitation anomalies at intraseasonal and interannual timescales in the northern mid-latitudes (Ding and Wang, 2005; Di Capua et al., 2020b). Recent work based on statistically evaluating causal relationships in reanalysis data has shown that the CGT pattern and the SAM circulation system are connected by a two-way causal interaction (Di Capua et al., 2020b). Moreover, the causal effect of each of these patterns on atmospheric circulation and surface conditions can be effectively represented on a 2D map (Di Capua et al., 2020a). The CGT has been studied in seasonal forecasts provided by ECMWF, and the corresponding results show that generally the model can reproduce this pattern (Beverley et al., 2019). However, the CGT pattern in seasonal forecasts is too weak, likely due to a misrepresentation of the SAM convective activity in the tropical belt.

The second mode of covariability between tropical convection and boreal summer circulation is represented by a pair of patterns consisting of the western North Pacific summer monsoon (WNPSM) and the North Pacific high (NPH) (Di Capua et al., 2020a). The NPH is the result of the northward displacement of the North Pacific sub-tropical high due to the onset of the WNPSM activity at the beginning of July (Di Capua et al., 2020a). In reanalyses, the influence of these two patterns on other atmospheric fields is weak and mostly confined to the Pacific Ocean compared to the SAM–CGT pair. Nevertheless, the WNPSM and NPH systems can affect typhoon cyclogenesis in the tropical Pacific (Briegel and Frank, 1997) and temperature and circulation patterns in East Asia and North America, respectively, potentially acting as a source of wave activity downstream (Di Capua et al., 2020a; Ding et al., 2011). Therefore, even though the direct area of influence of the WNPSM–NPH pair is found over the ocean, effects of changes in their intraseasonal variability are relevant to remote and highly populated areas (e.g., the US west coast or Japan).

Causal discovery algorithms, such as the Peter and Clark Momentary Conditional Independence (PCMCI) method, help overcome issues with commonly used statistical techniques, like correlation measures. When carefully applied, they allow one to identify and separate causal versus spurious links (Runge, 2018; Runge et al., 2014, 2019). PCMCI has been used to study stratospheric polar vortex variability (Kretschmer et al., 2017, 2016, 2018), the Silk Road pattern interdecadal variability (Stephan et al., 2019), Atlantic hurricane activity (Pfleiderer et al., 2020), and causal interactions between the Indian summer monsoon and mid-latitude wave trains (Di Capua et al., 2020a, b). Moreover, PCMCI has also proven useful in providing early forecasts of Moroccan crops (Lehmann et al., 2020), sub-seasonal statistical forecasts of US surface temperatures (Vijverberg and Coumou, 2022; Vijverberg et al., 2020) and statistical seasonal predictions of Indian summer monsoon rainfall (Di Capua et al., 2019).

Process-based validation can help us to understand and correct biases in seasonal forecasts (Eyring et al., 2019; Horak et al., 2021). Here, we propose to use causal discovery to perform a process-based validation (Nowack et al., 2020) of tropical – extratropical interactions in SEAS5 seasonal forecasts. We compare observed (i.e. reanalysis) causal interactions between tropical convective activity and mid-latitude wave trains in the Northern Hemisphere during boreal summer with those

provided by seasonal forecasts. The scope of this comparison is three-fold: (i) we validate causal links in a coupled general circulation model (GCM) in forecasting mode against those derived from observations and (ii) we gather information on missing or misrepresented links in the GCM in forecast mode. Finally, (iii) we analyse whether these differences can be attributed to model biases and what impact different phases of the El Niño–Southern Oscillation (ENSO), present in the initial conditions of the forecasts, have on the strength and representation of causal links. Thus, this work represents an initial step to improving forecast skill, and the representation of tropical–extratropical teleconnections in GCMs. The remainder of this paper is organized as follows: Section 2 presents the data and methods used. Section 3 describes the results obtained by applying causal maps first to ERA5 reanalysis and then SEAS5 data. Section 4 provides a discussion of the obtained results in the context of the existing literature and the final conclusions.

2 Data and Methods

2.1 Data

We analyse intraseasonal (weekly) tropical convective activity and mid-latitude circulation characteristics during the extended boreal summer period (May to September, MJJJAS) using gridded data (0.25°x0.25° upscaled to 2°x2°) from the ERA5 reanalysis dataset (Hersbach et al., 2020) and the SEAS5 seasonal retrospective forecast dataset (Johnson et al., 2019), both provided by the European Centre for Medium-range Weather Forecasts (ECMWF). From the ERA5 dataset, we use daily (temporally averaged to obtain weekly samples) geopotential height fields at 200 hPa (Z200), outgoing longwave radiation (OLR), sea surface temperature (SST) and zonal (U200) wind fields for the period 1979-2020 (ERA-L) and for the subset 1993-2016 (ERA-S) (to be consistent with the available SEAS5 dates). While Z200 is useful for representing the mid-latitude circulation, OLR can be used as proxy of tropical convective activity. Despite SEAS5 providing values for precipitation globally, we prefer using OLR instead of precipitation itself as precipitation is not assimilated in reanalysis (thus less reliable than assimilated fields such as SST), observational precipitation data coverage for tropical regions is sparse and to keep this analysis consistent with Di Capua et al. (2020a). For the general circulation model comparison, we analyse OLR, Z200, SST and U200 fields for the period 1993-2016 of SEAS5. As for ERA5, SEAS5 data are also regridded from the original 1°x1° onto a 2°x2° grid and daily data are temporally averaged to obtain weekly samples. The interannual variability, seasonal cycle and any long-term trend are removed. To do so, first the interannual variability, i.e. the average value of each May-to-September period is subtracted from the corresponding year (thus ensuring that the weekly signal does not include the interannual variability). Then, for each of the 21 time steps considered in each MJJAS season, the trend over the 24 (or 600) years is removed and anomalies around zero are calculated, thus removing both the trend and seasonal cycle.

While reanalysis data from ERA5 provide one realization per year, SEAS5 provides 25 ensemble members each year, thus a total of 25x24 (600) model years. A schematic of the SEAS5 ensemble is shown in Fig. 1. Considering only the summer season (June to September) plus May, as required by the causal discovery framework to ensure a proper handling of time lags (see Section 2.3), we have a total of 21 weeks per extended summer (MJJAS) for each year (Fig. 1a). Thus, for the common

period 1993-2016, we have a total of 21x24=504 weekly time samples for ERA5 and 21x24x25=12,600 time samples for SEAS5 (Fig. 1b). Note that for both ERA5 and SEAS5 datasets, the first week of the MJJAS period starts on the 7th of May (the first full week common to both datasets when for ERA5 the first week of the year starts on the 1st of January).

Treating the 25 ensemble members per year for 24 years as one unique time series composed of distinct sub-sequences is possible under the assumption that each member of each ensemble year is independent of the remaining members. While the ensemble is generated by varying the initial conditions, this assumption of uniqueness is not true in general, since each ensemble member for a given year has common lower boundary conditions for the atmosphere, inherited from slowly evolving features in the ocean, e.g., SST anomalies in the tropics and extratropics. However, for the purpose of this analysis we are interested in the *relative* effect of a certain (set of) variable(s) on the remaining atmospheric fields *inside* the intraseasonal variability, with a maximum lag of a few weeks (thus on a much shorter time scale than interannual). It must be noted that we do not intend to use SEAS5 data to assess or exploit its forecast skill, but instead assess the ability of a general circulation model in forecasting mode at reproducing observed tropical – extratropical teleconnections. In other words, whether there is shared information between two ensemble members for a certain year, e.g., whether a certain phenomenon in the analysed climate system is stronger or weaker, or whether a specific year shows better forecast skill than another, does not affect the relative effect of that phenomenon on some selected atmospheric fields. However, we cannot exclude that different initialisation dates (SEAS5 seasonal forecasts are initialized on the first day of each month and run for 7 months) and the vicinity of the target season to the beginning of the simulation may influence the outcome and the resulting causal links. Thus, we choose to analyse SEAS5 forecasts initialized on both the 1st of March and on the 1st of May, with a target season of June-September (see Fig. 1a). This way, the model has up to three (and at least one) months of spin-up to reduce the influence of the initial conditions. However, for the 1st of May initialized forecasts, although the month of May is outside the target season, May time steps enter the set of precursors (see below) for June time steps. Thus, we provide a sensitivity analysis to show which results depend on (or are independent of) the chosen initialization date. In the final step of this work, we will assess whether the effect of ENSO on wind fields and convective activity in the Northern Hemisphere is (i) sufficiently well reproduced in SEAS5 when compared to ERA5, and (ii) influences the identified tropical – extratropical causal links.

2.2 Modes of co-variability

To identify the dominant modes of intraseasonal co-variability between tropical convective activity and mid-latitude circulation in the Northern Hemisphere, we apply maximum covariance analysis (MCA) as described in Di Capua et al. (2020a). The first two MCA modes are calculated by applying MCA to OLR fields in the tropical belt (15°S-30°N, 0°-360°E) paired with Z200 fields in the northern mid-latitudes (25°N-75°N, 0°-360°E), thus using the same geographical borders as in Di Capua et al. (2020a).

By construction, the first MCA mode explains a maximum of *squared covariance between the selected two fields* (Ding et al., 2011; Wiedermann et al., 2017) and MCA modes are ranked according to their explained squared covariance fraction (SCF) (Wilks, 2011). With this method, it is possible to identify pairs of patterns that can explain shared covariance and (to

some extent) evolve simultaneously. However, shared covariance, as in general correlation-based techniques, does not imply causality. To check whether these patterns may be causally related (e.g., via dynamical mechanisms) we apply the PCMCI causal discovery algorithm (see Section 2.3).

Each MCA mode provides two coupled (2D) spatial patterns (one for tropical OLR and one for mid-latitude Z200) and two associated time series. Note that separate time series are created for the OLR and Z200 MCA patterns and for each MCA mode, for a total of four time series when MCA modes 1 and 2 are analysed. These time series are obtained by projecting each (2D) MCA spatial pattern on the corresponding time-varying atmospheric field, and represent the time-dependent MCA scores or pattern amplitudes for both fields. Each time series describes the magnitude (prominence) and phase (sign) of those patterns for each time step of the field's time series. The acronyms associated with the MCA patterns are shown in Table 1.

Here, we apply MCA both to the full 1979-2020 (ERA-L) and reduced 1993-2016 (ERA-S) periods and to SEAS5 for the period 1993-2016. On one hand, we attempt to be consistent with the available SEAS5 data (ERA-S). On the other hand, we want to provide a direct comparison with Di Capua et al. (2020a), who studied the previous ERA-Interim reanalysis product, by exploiting the full length of the time series (ERA-L). However, the first two MCA modes calculated with SEAS5 data do not provide a close enough representation of the patterns shown in ERA5 (see spatial correlation coefficients shown in Table 2 and Section 3.1). Thus, to provide a meaningful comparison between the two datasets, we define MCA patterns in SEAS5 by projecting the first two ERA5 MCA patterns onto SEAS5 data. This is done in the same way in which the ERA5 MCA scores/time series are calculated, i.e., by calculating the dot product of each ERA5 MCA mode with the corresponding OLR or Z200 field time series for each time step.

2.3 PCMCI and Causal Maps

The PCMCI algorithm is a causal discovery method using partial correlations to iteratively test for causality between two (or more) time series (*actors*) given a certain set of conditioning variables (Runge, 2018; Runge et al., 2014, 2019; Spirtes et al., 2000). The term *causal* builds upon a series of hypotheses, such as causal sufficiency (i.e. all relevant actors are included in the network) and stationarity of the detected causal chain (i.e. the causal links are stationary over time and actors show no trend). Hence, the detected causal links are valid in the set of analysed actors and also depend on the linear or non-linear framework applied (here, we employ linear partial correlations since they can be estimated in a more robust way from time series of limited length than their nonlinear counterparts and allow defining the direction of an effect relative to the cause) as well as a set of parameters, such as the significance threshold α. The causal links identified by PCMCI are represented in a so-called Causal Effect Network (CEN), a graph where each actor is represented by a node and causal links are shown as arrows connecting different nodes (Fig. 2a). The sign and strength of a certain causal link are given by the β coefficient, represented in the CEN by the colour of the arrow. For example, given the causal link $actor2_{\tau=-1} \rightarrow actor1_{\tau=0}$, a β coefficient of 0.25 represents a positive change of 0.25 standard deviation units (s.d.) of actor1 at lag 0 due to a positive change of 1 s.d. in actor2 at lag -1.

Here we apply the concept of *causal maps*, an extension of PCMCI to spatial fields of variables, to analyse the influence of a set of spatial patterns representing tropical – extratropical summer interactions (identified by applying MCA) on a 2D field (Di Capua et al., 2020a). Causal maps make use of the concept of CEN and the PCMCI algorithm. However, instead of showing the typical network-like shape with actors connected by arrows representing the direction, sign and strength of the causal links as shown in Fig. 2a, causal maps provide information in a similar conceptual manner to a classical correlation map (Fig. 2b). In a causal map, however, each grid point represents the sign and strength (given by the β coefficient) of a certain causal link, e.g. between actor1 and actor3, while the direction of the link and the set of actors involved is kept constant throughout each map (more maps are necessary to show different configurations of actors as shown in Fig. 2b).

To provide a meaningful comparison with previous work, in this analysis we apply the same framework as used in Di Capua et al. (2020a), i.e., each causal map is obtained by running at grid point level a CEN analysis with three actors. Of these three actors, two represent the pair of time series obtained for each MCA mode and are kept constant throughout the map (actor1 and actor2 in our example in Fig. 2a). The third time series is the time series for any individual grid point of the considered time-varying field (OLR or Z200) and thereby depends on longitude and latitude (actor3(lat, lon) in Fig. 2a). Thus, for each MCA mode we will have eight causal maps, as we have two target fields (OLR and Z200), two MCA modes and two MCA time series for each MCA mode (one for tropical OLR and one for mid-latitudes Z200). This can be summarized in Eq. (1):

$$MCA_i^k \rightarrow field_j | MCA_i^{k \neq l} \tag{1}$$

where $i \in \{trop.\ OLR, midlat.\ Z200\}$, $j \in \{Z200, OLR\}$ and $k, l \in \{1,2\}$ being the two MCA modes.

Here, we use the PCMCI version 4.1 from the Python package tigramite (https://github.com/jakobrunge/tigramite). We estimate causal maps with the parameters lag min = -1, lag max = -2 (units of the sampling period, i.e., weekly) and pc_alpha = 0.2 (unless otherwise indicated we use the default parameters of PCMCI). Note that only results for lag -1 are shown as almost no significant links are found for lag -2 in ERA5 causal maps. Note that pc_alpha is not the final significance threshold adopted to determine the significance of the identified causal links, instead it is a parameter used in the PC step of the PCMCI algorithm, which if taken too strictly, e.g., the usual α = 0.05, would prevent the algorithm from retaining some potentially meaningful links (for further details see https://jakobrunge.github.io/tigramite/).

The significance threshold adopted for plotting the results is α = 0.05 and we use corrected *p*-values by applying a false discovery rate (FDR) correction (Benjamini and Hochberg, 1995) to control for multiple testing among the multiple grid locations in causal maps. The false discovery rate is "the expected proportion of erroneous rejections among all rejections" (Benjamini and Yekutieli, 2001). Moreover, the robustness of a causal map for SEAS5 is assessed by calculating causal maps for a range of sub-periods. In 10 trials, we iteratively remove 10% of the record: 60 years out of 600; in the first iteration years 1-to-60 are removed, in the second years 61-to-120 are removed and so on. Finally, we plot β values for those grid points where a significant β value is found for at least 70% of the cases (similarly to what was done in Di Capua et al. 2020a). For ERA5 and SEAS5 subsamples we only apply the false discovery rate correction since this criterion is already strict and the *p*-values are affected by the length of the time series.

3 Results

The following results section is organized as follows: first we define and describe MCA patterns both in ERA5 and SEAS5 datasets (Section 3.1). Then we calculate the respective causal maps and compare those obtained in ERA5 with those obtained with SEAS5 (Section 3.2). In Section 3.3 we produce a 1000 member subsampling experiment to determine whether ERA5 β values fall in the range of realisations of SEAS5. Following those studies, in Section 3.4, we check whether ENSO may influence these tropical – extratropical teleconnections and what role model biases may play in explaining ERA5-SEAS5 differences.

3.1 MCA patterns in SEAS5 and ERA5

We first calculate the two leading maximum covariance analysis (MCA) coupled modes of tropical – extratropical co-variability between mid-latitude (25°-75°N) geopotential height at 200 hPa and tropical (15°S-30°N) outgoing longwave radiation in ERA5 reanalysis data for the period 1993-2016 (ERA-S, Fig. 3). MCA modes calculated for the extended period 1979-2020 (ERA-L) are shown in Fig. S1 in the Supplementary Material. The two pairs of patterns identified in this way are the South Asian Monsoon (SAM, Fig. 3d) paired with the circumglobal teleconnection (CGT, Fig. 3a) pattern for MCA mode 1, and the western North Pacific Summer Monsoon (WNPSM, Fig. 3j) paired with the North Pacific High (NPH, Fig. 3g) for MCA mode 2.

The SAM is characterized by a large rainfall band stretching from the Arabian Sea towards the western edge of the South China Sea, with a peak of negative OLR (relatively high rainfall) centred over the Indian peninsula (Fig. 3d). The CGT pattern shows five centres of positive Z200 anomalies over the Iberian Peninsula, central Asia on the eastern side of the Caspian Sea, East China, the North Pacific and the south-eastern US (Fig. 3a). The WNPSM pattern features a region of enhanced convective activity over the tropical western Pacific between 25° and 30°N accompanied by an area of suppressed convective activity on its western side centred over the South China Sea (Fig. 3j). The main feature of the NPH is a ridge in Z200 located at the western side of the North Pacific, however this pattern also shows a zonally oriented wave train like the CGT pattern, but with centres of action shifted in longitude (Fig. 3g). These four patterns can explain up to 25% of the variance in the Z200 and OLR fields depending on the region (not shown).

The key features described above for the ERA-S MCA modes 1 and 2 (Fig. 3a,d,g,j) are qualitatively very close to those found for the 1979-2020 period (ERA-L, Fig. S1a,d,g,j) or ERA-Interim (Figs. 3 and 4 in Di Capua et al. (2020a)). The spatial correlation between ERA-L and ERA-S MCA modes ranges between 0.8 and 0.9 (Table 2), showing that these patterns are robust across different versions of ERA reanalysis and quite insensitive to the chosen period. However, when MCA modes are calculated on SEAS5 seasonal forecasts initialized on the 1st of May and compared to ERA-S, a few key differences arise (Fig. S2). Overall each ERA-S MCA mode (Fig. 3a,d,g,j) shares common features with its corresponding SEAS5 MCA mode (Fig. S2a,d,g,j). This similarity can be quantified by calculating the spatial of between each ERA-S MCA mode shown in Fig. 3a,d,g,j with the

corresponding SEAS5 MCA mode in Fig. S2a,d,g,j, for both the tropical belt (15°S-30°N, 0°-360°E) and the mid-latitude region (25°-75°N, 0°-360°E). Results are shown in Table 2 (second column) and yield values ranging between 0.4 and 0.6.

285   However, some features that in ERA-S characterise MCA 1 are also found in SEAS5 MCA2, and vice versa for ERA-S MCA 2. Thus, features that in ERA-S are separated and characterise each of the first two MCA modes, appear mixed in SEAS5 (or vice versa). For example, in ERA-S the CGT-related high to the east of the Caspian Sea is only visible in MCA 1 (Fig. 3a), while in SEAS5, the same positive signal is found in both MCA 1 and MCA 2 (although stronger in MCA 1). Similarly, the convective activity over the Indian peninsula, which in ERA-S represents one of the characteristic features of MCA 1 (Fig. 3d) and is very

290   weak in MCA 2 (Fig. 3j), is found in both MCA 1 and MCA 2 of SEAS5 almost with similar magnitudes, though the negative anomalies are stronger in MCA 1 (Fig. S2d,j). In contrast, the wave pattern over Eurasia characterising ERA-S MCA 2 showing a high pressure region over Eastern Europe and a low over Central Asia (Fig. 3g) is very weak in SEAS5 MCA modes 2 (Fig. S2g) and not present in MCA mode 1 (Fig. S2a). This can be seen in spatial correlation values calculated between ERA-S MCA 1 (MCA 2) and SEAS5 MCA 2 (MCA 1) which also range between 0.4 and 0.6, although sometimes with reversed sign (opposite phase of the pattern) (Table 2, third column). Therefore, we conclude that despite the general resemblance of the first two SEAS5 MCA

295   modes to those calculated in ERA-S, working with a mixed signal would hinder the interpretability of parts of the results, since it would make it difficult to effectively separate the causal effect of the SAM – CGT patterns from that of the WNPSM–NPH pair.

To account for thisproblem, SEAS5 MCA modes have been re-calculated by projecting ERA-S MCA patterns onto SEAS5 Z200 and OLR 3D fields, as explained in the Data and Methods section. Then, to visualise the equivalent SEAS5-ERA5 MCA

300   modes, composites of time steps with the MCA time series values higher than 1 standard deviation (s.d.) are calculated (Fig. 4a,d,g,j). As a result, a much closer resemblance between SEAS5-ERA5 and ERA-S MCA patterns is obtained and the spatial correlation coefficients between SEAS-ERA5 and ERA-S MCA patterns reach values of +~0.8-0.9 (Table 2, fourth column).

3.2 Causal maps in SEAS5 and ERA5

Having defined ERA-S and SEAS5-ERA5 MCA modes, we apply the concept of causal maps to detect the causal effect of

305   each of these four patterns (two for each mode) to Z200 and OLR fields, in the range 15°S-75°N.

Causal maps calculated for the effect of ERA-S CGT and SAM at lag -1 on Z200 and OLR fields are shown in Fig. 3b,c and 3e,f respectively. The main features are the effect of SAM on Sahel Z200, the tropical Pacific Ocean, a high-low pattern in the North Pacific and the negative effect on Z200 in Central Europe (Fig. 3e). A positive causal effect (positive β value) of SAM on Sahel Z200 means that enhanced rainfall over the Indian monsoon region tends to be followed one week later by high Z200 anomalies

310   over the Sahel region and the tropical Pacific. In contrast, increased SAM activity leads to negative β values over Central Europe Z200, positive β values southwest of Alaska and negative β value in the eastern North Pacific Z200 (Fig. 3e). A positive causal effect (positive β values) of the CGT pattern on Z200 is mostly concentrated in the subtropical North Atlantic, Southern Europe, Central, North and East Asia, North Pacific and Southeast US (Fig. 3b). Thus, these regions will experience positive Z200 anomalies one week after an enhanced CGT pattern. In contrast, the North Atlantic (around Iceland), South Asia and the Arabian Sea, the

315   eastern North Pacific, the Philippine Sea and Canada will experience negative Z200 anomalies (negative β values, Fig. 3b). In

general, β values for OLR causal maps (Fig. 3c and 3f) mirror those for the Z200 field in the mid-latitudes, while in the tropics they add the information on tropical convective activity which is not detected by Z200 anomalies. In Fig. 3c negative β values over Southern India indicate that an enhanced CGT pattern leads to lower OLR values and thus increased convective activity over the region. In Fig. 3f the tilted band of convective activity stretches from the Arabian Sea and central India towards the Maritime

320   Continent. This convective activity is related to the boreal summer intraseasonal oscillation (BSISO), as shown in Di Capua et al. (2020a).

Causal maps calculated for the effect of NPH and WNPSM on Z200 and OLR fields are shown in Fig. 3h,i and 3k,l respectively. Causal maps for the effect of NPH on Z200 and OLR field display a zonally oriented pattern that encircles the northern mid-latitudes (Fig. 3h,i). Positive β values in Fig. 3h indicate that an enhanced NPH pattern (Fig. 3g) leads to positive Z200 anomalies over the

325   North Atlantic (around Iceland), Southeast China, the NPH region and the north-western US. Negative β values over Central and East Asia and over the eastern North Atlantic indicate that an enhanced NPH pattern is followed one week later by negative Z200 anomalies over these regions. The effect of WNPSM is more localized to East Asia and the North Pacific, where an arch-shaped pattern stretches from the tropical western Pacific, reaching Alaska and the US west coast (Fig. 3k,l and S1k,l in the Supplementary Material). Suppressed convective activity over the South China Sea and the Philippine Sea and increased convective activity over

330   the WNPSM are followed one week later by negative Z200 anomalies over Southeast Asia and Central Asia and positive Z200 anomalies over Northeast Asia and the US west coast (Fig. 3k,l).

In general, a two-way causal link between tropical convection and mid-latitude circulation is shown for both MCA modes: the causal effect of SAM and WNPSM reaches the northern mid-latitudes, while the effect of the mid-latitude CGT pattern and NPH extends to subtropical latitudes. Consistent results are obtained for ERA-L (see Fig. S1) with more significant causal links (likely

335   due to the increased length of the 1979-2020 time series), improving the clarity of the spatial patterns. These patterns also show great resemblance to those produced using the ERA-Interim reanalysis in Di Capua et al. (2020a) (see their Figs. 3 and 4).

Causal maps produced with SEAS5-ERA5 MCA modes and SEAS5 OLR and Z200 fields (Fig. 4) show similar spatial patterns to those obtained with ERA-S (Fig. 3). In general, the sign and the geographical position of the causal links detected in SEAS5 are consistent with those found in ERA5. Thus, the main tropical – extratropical intraseasonal causal relationships in boreal summer in

340   the Northern Hemisphere are at least qualitatively well represented in the SEAS5 system. These causal maps also show that the two-way causal pathway between tropical convective activity and extratropical circulation is captured by the seasonal forecasts. Thus, on the one hand we gain confidence in the interpretation of the earlier ERA-Interim and ERA-S/L causal map analysis, which is reproduced by SEAS5, and on the other hand we show that, to a first approximation, seasonal forecasts can reproduce such causal links.

345   However, the strengths of the causal links detected in SEAS5 are generally weaker than those in ERA-S (or ERA-L). Note that in Figs. 3 and 4 the same colour bar for β values is used. To visualise the difference in strength and/or sign between SEAS5-ERA5 and ERA-S β values, we provide maps of SEAS5-ERA5 β coefficients (β_{SEAS5}) relative to the strength of ERA-S β coefficients (β_{ERA5}) (Fig. 5 and 6, left column). It should be noted here that a comparison between β_{SEAS5} and β_{ERA5} is only possible when both

causal maps show a significant causal link at a specific geographic location. Maps in Fig. 5 and 6 show values of Δ_β defined following:

$$\Delta_\beta = \frac{\beta_{SEAS5}}{\beta_{ERA5}}. \tag{2}$$

When $0 < \Delta_\beta < 1$, β_{SEAS5} values are smaller in magnitude than those shown by β_{ERA5} but the sign of β_{SEAS5} is correct; this means for example that increased SAM activity leads to positive Z200 anomalies over the Sahel both in SEAS5-ERA5 and ERA-S, however, β_{ERA5} is larger than those in SEAS5-ERA5 (Fig. 5c), meaning that the causal effect of SAM on this region is weaker in the model than in the reanalysis product. When $\Delta_\beta > 1$, the sign of β_{SEAS5} is correct but its magnitude is greater than β_{ERA5}; e.g. over the Tibetan Plateau in Fig. 5c. Finally, when $\Delta_\beta < 0$, the sign of β_{SEAS5} is wrong irrespective of the strength of the link; e.g. over the Philippine Sea in Fig. 5a. The colour bar in these plots has been chosen to visually distinguish between well represented β_{SEAS5} values (light yellow colours) and underestimated β_{SEAS5} values (dark blue colours). Δ_β maps for both MCA mode 1 and MCA mode 2 generally give values of Δ_β between 0 and 1, with mean values averaged over each map ranging between 0.4 for the causal effect of CGT on Z200 (Fig. 5a) up to 0.7 for the causal effect of SAM on OLR (Fig. 5g). ERA-S has a much smaller number of significant β, thus the spatial patterns in Figs. 5 and 6 shrink when compared to those shown in Fig. 4. For MCA mode 1, regions with especially poor β_{SEAS5} representation are the northern high-latitudes and the western North Pacific in Fig. 5a (CGT $\rightarrow$ Z200|SAM), the tropical belt and in particular the Sahel region in Fig. 5c (SAM $\rightarrow$ Z200|CGT), Siberia in Fig. 5e (CGT $\rightarrow$ OLR|SAM) and the Maritime Continent in Fig. 5g (SAM $\rightarrow$ OLR|CGT).

For MCA mode 2, regions with especially poor β_{SEAS5} representation are the western US in Fig. 6a (NPH $\rightarrow$ Z200|WNPSM), the western tropical Pacific in Fig. 6c (WNPSM $\rightarrow$ Z200|NPH), northern US in Fig. 6a (NPH $\rightarrow$ OLR|WNPSM) and the West Coast of the US and the western tropical Pacific in Fig. 6g (WNPSM $\rightarrow$ OLR|NPH). This difference may hint at potential biases or misrepresentations of these teleconnections in the GCM.

Together with these β-difference maps we also show the histograms indicating the percentage of grid points (in the domain 15°S-75°N, 0°-360°E) in which the SEAS5 β coefficient has the correct sign/magnitude compared to ERA5 both for MCA mode 1 (Fig. 5, right column) and MCA mode 2 (Fig. 6, right column). In general, these plots underline the results shown in the Δ_β maps: the largest amount (>90%) of β_{SEAS5} values have a correct sign with respect to β_{ERA5} but lower strength, with few exceptions, e.g. for CGT $\rightarrow$ Z200|SAM (Fig. 5b) where the number of β values with wrong sign is ~13%. Thus, in general the SAM–CGT and WNPSM–NPH patterns have the same causal influence on Z200 and OLR fields both in ERA5 and SEAS5 dataset. However, although the sign of β values is generally correct, in almost all cases the majority (more than half) of the Δ_β values are found between 0 and 0.5, indicating that SEAS5 strongly underestimates β values: β_{SEAS5} equal to half of β_{ERA5} means that the effect of an increase of 1 s.d. in, e.g., SAM on the Z200 field would also be reduces by half in the SEAS5 forecasts.

The weaker strength of the causal links is not dependent on the use of SEAS5-ERA5 MCA modes instead of SEAS5 MCA modes. When SEAS5 MCA modes, which capture the strongest co-variability signal in SEAS5 seasonal forecasts, are used to produce causal maps (see SI, Fig. S2), a similar strength of the causal links is detected.

The dependence of causal maps on the time of initialization of SEAS5 has also been analysed. The same plots produced with SEAS5 seasonal forecasts initialized on the 1st of March (rather than 1st of May) are shown in the Supplementary Material (see Figs. S3-S5). Figures 4-6 are very consistent with Figs. S3-S5, thus the initialization date of SEAS5 does not have a qualitative influence on causal maps calculated over the entire 1993-2016 period. The apparent independence on the initialization date is most likely

385 explained by the independence of the β coefficients on the absolute values of a certain variable in a specific year. As explained in the Section 2, the β coefficients represent the relative change in s.d. of one actor given a certain change in the values of its parents (expressed in s.d.).

3.3 Causal effect spread in the SEAS5 ensemble

We perform a sub-sampling experiment to better understand differences in the strength of causal links between ERA-S and

390 SEAS5-ERA5 and evaluate the spread inside the SEAS5 ensemble. We select 1000 samples of 24 years each (for each year one ensemble member is randomly selected out of the 25 available members), and for each sample we provide the corresponding causal map. In this way, the number of years used in each subsampling experiment (24 years) is the same as those available from ERA-S (24 years). Reducing the length of the time series in this way increases the variability and hence lowers the significance of the obtained β values. However, this should not by itself lower the strength of the β values themselves. Thus, a priori, we might expect

395 fewer regions to show a significant β value in a smaller dataset than in a larger one but not a difference in the strength of the β values. Hence, this 1000-ensemble member subsampling experiment allows us to evaluate the distribution of β values around their mean value and to compare it to the ERA-S (or ERA-L) values of reference. For each causal map, the p-values are corrected by applying the Benjamini-Hochberg false discovery rate correction and only β values with a corrected p-value < 0.05 are retained. The resulting 1000 causal maps are averaged and shown in the left column in Figs. 7 and 8. For each grid point, the mean β value

400 is calculated only if at least 100 β value results are significant at the $\alpha = 0.05$ threshold, however, non-significant values do not enter the mean. Applying this double threshold (which is not done in Fig. 3) shrinks notably the area of the spatial patterns compared to those in Figs. 3 and 4, however here we concentrate only on the β values contained in the regions highlighted by black boxes in Figs. 7 and 8.

Due to the complexity of the spatial patterns shown in the causal maps in Fig. 3, and to the smaller number of significant grid-

405 points available in ERA-S compared to SEAS5-ERA5 (as visualized in Figs. 3 and 4), calculating spatial correlations is not the most efficient way to compare the two sets of causal maps. A high (or low) spatial correlation may result from strong or weak agreement in different regions, but it would not be possible to discern from which specific area the signal is originating.

To provide a meaningful comparison, we choose four key regions for each MCA mode, and for those regions we analyse the characteristics of the causal effect in detail. We identify these regions based on (i) the prominence of the signal in Fig. 3 and 4 and

410 (ii) the misrepresentation of β values shown in Fig. 5 and 6. By applying these criteria, the chosen geographical regions (shown in Figs. 7 and 8) include: (a) the Sahel region, (b) southeast Asia, (c) India and (d) the Mediterranean for MCA mode 1 (Fig. 7), and (a) Japan, (b) central western US, (c) the Maritime Continent and (d) central eastern Europe for MCA mode 2 (Fig. 8).

The spatial domains used to define these regions are shown in Table 3. For each region and for each sample in the 1000-ensemble member subsampling experiment, the causal effect is spatially averaged (accounting only for significant values, i.e., zero

415   values are discarded as they are not significant) and the absolute value is taken after averaging. In this way, we obtain a distribution of 1000 β values for each region of interest. Note that we choose to consider spatially averaged *absolute* β values as indices for corresponding teleconnection strength to focus on the strength of the causal effect rather than the number of significant grid-points or the sign of the connection. Probability density functions (PDFs) for the subsampling experiments of each β index (one for each of the eight regions in Table 3) are calculated by standardizing each β index by its standard deviation (calculated over the 1000

420   samples) and centring it around its mean. The mean β values obtained by the SEAS5 causal maps calculated using all 600 years (Fig. 4) are represented by solid vertical lines (orange for Z200 and light purple for OLR PDFs) together with reference β values calculated in ERA-L and ERA-S (shown as purple and magenta vertical lines) in each PDF. All three β values are standardized by dividing by the standard deviation and mean value of the SEAS5-ERA5 subsample distribution. Despite causal maps for ERA-S and ERA-L showing good agreement, those for ERA-L show a greater number of significant grid-points (due to the length of the

425   time series); showing both standardized values in Figs. 7 and 8 helps to further put into perspective the importance of the choice of years and thus provides a more balanced interpretation of the results.

In general, the mean β values obtained in Fig. 4 tend to be lower than the mean β values obtained in the 1000 subsamples, indicating that taking all the 600 years together, despite spatial patterns showing good agreement, effectively increases the underestimation effect of the strength of the β values when compared to the average of the 1000 subsample. For MCA mode 1,

430   which is characterized by the SAM–CGT pair, we show that the link from the SAM to Sahel Z200 (SAM → Sahel Z200 | CGT) is the one with the largest bias between SEAS5-ERA5 and ERA-L (or ERA-S), with no subsample in our 1000 subsampling experiment being capable of reproducing the ERA-L (or ERA-S) causal link strength (Fig. 7b). By contrast, the causal effects of SAM on OLR over India (SAM → India OLR | CGT), of CGT on OLR over the Mediterranean (CGT → Medit. OLR | SAM) and of CGT towards Southeast Asia Z200 (CGT → SE-Asia Z200 | SAM) all fall in the respective range of values of the 1000 subsamples, with both

435   ERA-L and ERA-S β values falling between the 10th and the 90th quantile (Fig. 7d,f,h). Thus, in this region the β values show a good agreement with reanalysis data when the spread of SEAS5 β values is considered.

Results for MCA mode 2, analysing the North Pacific high (NPH) paired together with the western North Pacific summer monsoon (WNPSM), are shown in Fig. 8. The links from the NPH towards the north-western US Z200 (NPH → NW-US Z200 | WNPSM) (Fig. 8d) and from the WNPSM towards the Maritime Continent OLR (WNPSM → M.Cont. OLR | NPH) (Fig. 8f) are

440   those with the largest bias between SEAS5-ERA5 and ERA-L (or ERA-S) β values, with ERA-L (or ERA-S) β values falling at the upper edge of the PDF (above the 90th quantile). In contrast, the links from the WNPSM towards Japan Z200 (WNPSM → Japan Z200 | NPH) (Fig. 8b) and from the NPH towards central Europe OLR (NPH → CE-EU OLR | WNPSM) (Fig. 8h) fall well in the centre of the distribution, thus showing the smallest bias among the analysed β values. That is, for those cases in which ERA5 values fall in the middle of the distribution, the particular modelled field is more likely to have a low bias.

445   Again, we test our results for the dependence on the initialization date of the SEAS5 dataset. Figures S6 and S7 in the Supplementary Material show the same as Figs. 7 and 8 produced using SEAS5 initialized on the 1st of March. The results obtained

are qualitatively and quantitatively very similar, with β values for 5 out of the 8 analysed regions falling between the 10th and the 90th quantile consistently between SEAS5 init05 and SEAS5 init03. Hence, the initialisation date does not affect the spread of the subsampling experiment.

3.4 The effect of the ERA5 – SEAS5 mean state bias and ENSO on tropical – extratropical causal links

We investigate how the bias in convective activity between ERA5 and SEAS5 may affect the monsoon-desert mechanisms and find inconclusive results. SEAS5 shows enhanced convective activity with respect to ERA5 around the equator (negative OLR anomalies in Fig. S8m,o) and a drier tendency over central India and the Arabian Sea (positive OLR anomalies in Fig. S8m,o). Rodwell and Hoskins (1996) have shown that the heat source provided by the convective activity in the Indian Ocean/Bay of Bengal region generates Rossby waves that reach the Sahara Desert. However, the latitudinal position of the heat source is critical: a heat source located in the south (10°N) does not act as a source of Rossby waves capable of reaching the Sahara Desert, while a heat source located around 25°N does. Thus, we investigate whether the dry bias over central India may explain low causal effect values over the Sahel and North African region and calculate causal maps for years with enhanced convective activity over central India and for those with enhanced convective activity over the tropical Indian Ocean. Despite a slight tendency towards higher β values over North Africa being detected during years with enhanced convection over central India in SEAS5 initialized at 1st May (40% higher compared to years with enhanced convection over the Indian Ocean, Fig. S9e), this result is not found in SEAS5 initialized at 1st March and thus remains inconclusive (Fig. S9j).

Finally, we investigate the effect of ENSO states **on** the sign and strength of tropical-extratropical causal interactions shown in Fig. 4 and show that the effect of ENSO positive and negative phases is mostly marginal with a few exceptions. We define El Niño and La Niña years based on seasonal (JJAS) SST anomalies averaged over the Niño3.4 region (5°S-5°N, 190°-240°E) and calculate causal maps for the effect of MCA mode 1 and 2 on Z200 field separately for the 102 Niño3.4 positive and 142 Niño3.4 negative years (those years that exceed the +0.5°C/-0.5°C thresholds are defined as El Niño/La Niña years, respectively). The results for MCA mode 1 are shown in Fig. 9 for both Niño3.4 positive (left column) and Niño3.4 negative years (middle column) separately and for the difference Niño3.4 positive minus Niño3.4 negative (right column) and for different initialization dates (1st of March and 1st of May). Comparing the causal maps in Fig. 9 left and middle column with those in Fig. 4 shows that, in general, the spatial patterns and the sign of the causal links are not affected by the sign of the ENSO anomalies.

Changes in the strength of the links are shown in the right column in Fig. 9. For each grid point, the difference Δ_β between the β value for Niño3.4 positive years ($\beta_{niño}$) and the β value for Niño3.4 negative years ($\beta_{niña}$) is calculated and then divided by $\beta_{niño}$, following Eq. (3):

$$\Delta_\beta = \frac{\beta_{niño} - \beta_{niña}}{\beta_{niño}} \times 100\% \tag{3}$$

Hence, Δ_β is expressed as a percentage, where a zero value represents perfect agreement between $\beta_{niño}$ and $\beta_{niña}$, a positive value of, e.g., 50% means that $\beta_{niño}$ is 50% larger than $\beta_{niña}$ and vice versa. In general, β values in the tropical Pacific and over eastern Africa and the western tropical Indian Ocean are 40 to 80% larger during El Niño years and this result is consistent for both

initialization dates. However, in other regions, results differ depending on the initialization date, e.g. north-western Africa or the
480   North Pacific region for the effect of SAM on the Z200 field. ENSO causal maps for MCA mode 2 and for the ENSO versus neutral
years are shown in Fig. S10-S12 in the Supplementary Material and show similar results. Thus, we conclude that in general ENSO
does not alter the sign and spatial patterns of tropical extratropical teleconnection but can however modify the strength of these
connections in some specific areas, especially close to the equator.

485   **4 Discussion**

In this work, we provide an process-guided statistical analysis, built on causal discovery, of the representation of boreal
summer tropical–extratropical intraseasonal teleconnections in the Northern Hemisphere in SEAS5 seasonal forecasts by
ECMWF. We have analysed the two first modes of covariability identified using maximum covariance analysis (MCA)
between weekly geopotential heights (Z200) and convective activity (OLR) fields in reanalysis data from the ERA5 dataset
490   for the period 1993-2016. The first MCA mode shows the South Asian monsoon (SAM) paired with the circumglobal
teleconnection pattern (CGT). The second MCA mode shows the western North Pacific summer monsoon (WNPSM) paired
with the North Pacific High (NPH). Causal maps showing the causal effect of these four patterns on Northern Hemisphere
Z200 and OLR fields at 1-week lead for periods 1979-2020 (Fig. S1, see Supplementary Material) and 1993-2016 (Fig. 3) are
largely consistent with results obtained with ERA-Interim data for the period 1979-2018 (Di Capua et al., 2020).

495   Here, we focus on assessing the ability of SEAS5 seasonal forecasts in reproducing those results. To achieve this goal
and provide a meaningful comparison we project the first two MCA modes calculated in ERA5 onto SEAS5 data and we
calculate the corresponding causal maps (Fig. 4). In general, causal maps obtained with SEAS5 correctly reproduce the sign
and the spatial patterns of ERA5 causal maps, though with weaker magnitudes (Figs. 5 and 6). Thus, spatial patterns shown in
SEAS5 seasonal forecast causal maps are validated by those extracted from ERA5: while the SEAS5 forecasting system can
500   reproduce the patterns seen in ERA5 reanalysis, we gain confidence that observed causal maps represent actual physical
mechanisms.

We analyse the negative bias found in SEAS5, i.e., a general underestimation of the causal effect, which may arise for
different reasons: (i) ERA5 causal maps are subject to multidecadal variability or affected by the limited time span considered
or (ii) the SEAS5 forecasting system is missing or misrepresenting a key mechanism for a correct representation of the strength
505   of the analysed causal links. To test the first hypothesis, we perform a subsampling experiment, providing a thousand causal
maps obtained using 24 years randomly selecting one member for each forecast year (out of the 25 members available for each
year). Random year selection may overestimate the range for potential β coefficients, thus further emphasizing the
underestimation of the strength of these tropical – extratropical teleconnections in SEAS5. We identify a few key regions for
which we compared the observed ERA5 causal link strength with the range of SEAS5 values obtained from the subsampling
510   (Figs. 7 and 8). Through this experiment we show that, although ERA5 values are generally higher than the mean of the SEAS5

distribution, approximately half of them fall below the 90th quantile of the SEAS5 distribution (3 out of 4 for MCA 1 and 2 out of 4 for MCA2). Thus, SEAS5 has difficulty generating high values of the teleconnection strength especially over North Africa, North America and the Maritime Continent (when the ERA5 reference values exceed the SEAS5 90th percentile). In the other analysed regions, we show that for a correct estimation of the strength of the causal links, using time series of the same length is crucial to avoid underestimation effects due to the length of the time series.

We calculate the biases between ERA5 and SEAS5 for SST and U200 JJAS climatologies and show that biases are present both in tropical and extratropical SST and in the mid-latitude jet over Eurasia (Fig. S8 in the Supplementary Material). The U200 mean-state biases in the mid-latitudes suggest a systematic northward shift of the subtropical westerly jet all across Eurasia, possibly affecting the waveguide for the CGT and the Silk Road patterns. Future work may employ nudging experiments to test the sensitivity to this northward shift, despite previous work had shown little effect of this phenomenon on the CGT pattern (Beverley et al., 2019). Changes in the initialization dates of the SEAS5 simulations (here 1st of March and 1st of May) also impact the magnitude of both SST and U200 biases, with reduced SST biases in SEAS5 data initialized on the 1st of May and seemingly increased biases for the Eurasian jet when compared to SEAS5 data initialized on the 1st of March (see Fig. S8).

In boreal summer, the CGT pattern arises even without the heat source provided by SAM (Ding et al., 2011), as it represents a preferred mode of variability of boreal summer circulation that can be ignited by different forcings (Kornhuber et al., 2020; Teng and Branstator, 2019). Recent work has shown that there is a positive causal link from the SAM to the CGT (Di Capua et al., 2020b, a). In general, climate models struggle to reproduce the climatology of SAM rainfall patterns, both in magnitude and spatial pattern (Menon et al., 2013) and SEAS5 seasonal forecasts underestimate the strength of the SAM convective activity and rainfall over the Indian peninsula and the Bay of Bengal (see Fig. S8 or Chevuturi et al. (2021)). The CGT pattern has been shown to be too weak in SEAS4 (ECMWF's previous operational seasonal forecasting system), likely due to a dry bias in precipitation in SEAS4: weaker convective activity over the Indian subcontinent does not provide the heat source that reinforces the CGT pattern (Beverley et al., 2019; Ding and Wang, 2005; Di Capua et al., 2020b). If the forcing (SAM) is too weak, the response (CGT) will be too weak, but this does not necessarily affect the strength of the causal link. Here, our analysis shows a negative bias in β coefficients over North Africa in the first MCA mode, thus there is not only a negative bias in the precipitation over the Indian peninsula, but also the causal link strength is too weak. In general, our results also show good agreement with what was shown in Beverley et al. (2021), in which the interaction between the CGT and the SAM has been explored by applying a heating source over the Indian subcontinent in ECMWF System 4. Their results show that the heating source induced by SAM convective activity is effective at driving a CGT-like wave train in northern mid-latitudes, however the response in the model is weak compared to the observed patterns.

Despite consistent underestimation of causal link strength in certain regions (Figs. 5 and 6), these results imply the ability of the SEAS5 forecast system to reproduce the sign and the spatial distribution of the observed causal patterns for boreal summer intraseasonal variability in the Northern Hemisphere (Figs. 7 and 8). Although this analysis does not rely on nor imply a skilful forecast, the causal effect of tropical and mid-latitude patterns on circulation and convection in the Northern

545   Hemisphere in SEAS5 seasonal forecasts is qualitatively well comparable with that shown in ERA5 reanalyses. Here we have shown for which regions the agreement between SEAS5 and ERA5 is good (or for which ERA5 values at least fall in the range of values shown in the 1000 subsample experiments) and those for which no subsample of SEAS5 can reproduce values comparable to ERA5. The region with the strongest bias, which cannot be reproduced in SEAS5, is the Sahel region. This may be explained by the southward shift of OLR and rainfall activity towards the equatorial Indian Ocean in SEAS5 , disrupting

550   the Rossby-forced teleconnection pattern (Rodwell and Hoskins, 1996). However, our analysis to determine the importance of the latitudinal position of convective activity in the Indian Ocean basin is inconclusive (Fig. S9).

    We further analyse the effect of the El Niño–Southern Oscillation (ENSO) on these boreal summer tropical–extratropical links in SEAS5 and we find that, in general, different ENSO phases do not affect the spatial patterns and sign of the causal links substantially (see Fig. 9 and Fig. S10). However, depending on the specific region, some effect on the strength of the

555   causal links is detected. This effect is generally relatively weak (~20-40% of the total strength of the link). These findings are generally in agreement with Di Capua et al. (2020), who showed that the spatial patterns and the sign of the causal links were mainly unaffected by ENSO; however, they noticed a regional dependence of the strength of the causal links on ENSO (see their Fig. 6). Although the effect of ENSO on the strength of the causal links seems to be marginal, we highlight those regions where a change in the strength is identified consistently both for 1st of March and 1st of May initialization dates. These regions

560   are mainly located in the tropical belt, in particular the western tropical Pacific, eastern Africa and Southeast Asia. Similar results are found for the effect of neutral versus ENSO active years (see Figs. S11 and S12 in the Supplementary Material). In general, the spatial pattern and sign of the causal links do not change, however the effect of the identified MCA patterns seems to increase in the tropical Pacific during neutral years, likely due to the absence of a strong ENSO signal. These results align with the findings of Goddard and Dilley (2005) who showed that the active phase of ENSO does not affect the numbers of

565   extremes but their overall predictability. Nevertheless, as discussed above, a change in the forcing (e.g., stronger convective activity induced by ENSO) would still result in a stronger effect even if the (relative) causal effect does not change.

    This information becomes even more relevant in the context of climate change. If EC-Earth (Döscher et al., 2022) (the Earth system model built by the ECMWF which shares the same atmosphere model as SEAS5) behaves similarly to ERA5, we can have some confidence that at least the sign and spatial patterns of these tropical – extratropical teleconnections are well

570   represented, though the strength of the links shows a large spread (Figs. 7 and 8). Future work will analyse how these teleconnections change in future projections under global warming scenarios. The analysed regions (from the South Asian monsoon to the North American and Eurasian continents) are prone to suffering effects of ongoing anthropogenic climate change (Pfleiderer et al., 2018; Mann et al., 2018; Coumou et al., 2018, 2017; Huntingford et al., 2019; Turner and Annamalai, 2012). Therefore, it is critical to evaluate a model's ability at reproducing key seasonal modes of variability and in doing so,

575   identify key targets for model development and motivate the improvement of parameterisation schemes. Ultimately, this could lead to increased reliability of seasonal and sub-seasonal forecasts; helping in improving warning systems and taking sensible early action to protect economy and society in the most vulnerable regions, especially for boreal summer, when the effect of

climate change is felt the most (Christidis et al., 2014; Teng and Branstator, 2019; Kornhuber et al., 2020; Coumou et al., 2015, 2017).

580   Finally, this process-based validation analysis represents a step towards the implementation of hybrid forecasts that combine statistical with dynamical models to increase the skill of seasonal and sub-seasonal weather forecasts (Schepen et al., 2012). By identifying the regions where a certain pattern exerts a significant influence and/or deriving information on which regions have a bias in the model, we provide useful information on the regions where the model representation of these mechanisms should be improved and work towards targeted forecasts. Moreover, general circulation models show higher skill

585 at forecasting tropical atmospheric dynamics than in the mid- or high-latitudes (Shukla, 1998; Chen et al., 2015), thus, knowing which regions in the Northern Hemisphere are more affected by tropical precipitation (as shown in Fig. 4) provides valuable information to improve seasonal forecast skill.

5 Conclusions

      In summary, this analysis has shown that ECMWF's seasonal forecasts have good ability at reproducing the sign and the spatial

590 patterns of the causal effect of the two main modes of covariability between tropical convection and mid-latitude circulation in boreal summer in the Northern Hemisphere. Despite a general underestimation of the causal link strength, our subsampling experiment shows that in most of the analysed regions, this negative bias is actually contained in the spread of the SEAS5 seasonal forecasts (Figs. 7 and 8). Thus, our confidence in both the ability of the PCMCI causal discovery tool to identify meaningful patterns in observation and the ability of the SEAS5 forecasting system to correctly represent those causal links is increased. The effect of

595 different ENSO phases (or ENSO versus neutral years) on tropical-extratropical links seems to be marginal, although biases in the SEAS5 model (e.g., reduced convective activity in the SAM region) may explain this discrepancy with observations. Further work is needed to confirm these results and their implications. Finally, the causal links represented in these causal maps represent a starting point to produce a new family of hybrid statistical model-based forecasts. In conclusion, this analysis has shown the usefulness of causal discovery algorithms as a tool for providing process-based validation of general circulation models and has led

600 to increased knowledge on the effects of tropical–extratropical teleconnections in boreal summer in the Northern Hemisphere.

Acknowledgments

This work has been financially supported by the German Federal Ministry for Education and Research of Germany (BMBF) via the JPI Climate/JPI Oceans project ROADMAP (grant no. 01LP2002B) (G.D.C. and R.V.D.). AGT's contribution to the

605 work was supported by the National Centre for Atmospheric Science.

Data availability. The data used in this article can be accessed by contacting the corresponding author.

Author contributions. All authors contributed to the design of the analysis. GDC performed the analysis and wrote the first draft of the paper. All authors contributed to the interpretation of the results and to the writing of the paper.

Competing interests. The authors declare that they have no conflict of interest.

610

615

References

Barnston, A. G., Glantz, M. H., and He, Y.: Predictive Skill of Statistical and Dynamical Climate Models in SST Forecasts during the 1997-98 El Nino Episode and the 1998 La Nina Onset, Bull. Am. Meteorol. Soc., 80, 217–244, 1999.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, https://doi.org/10.1038/nature14956, 2 September 2015.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing, J. R. Stat. Soc. Ser. B, 57, 289–300, https://doi.org/10.2307/2346101, 1995.

Benjamini, Y. and Yekutieli, D.: The control of the false discovery rate in multiple testing under dependency, Ann. Stat., 29, 1165–1188, https://doi.org/10.1214/aos/1013699998, 2001.

Beverley, J. D., Woolnough, S. J., Baker, L. H., Johnson, S. J., and Weisheimer, A.: The northern hemisphere circumglobal teleconnection in a seasonal forecast model and its relationship to European summer forecast skill, Clim. Dyn., 52, 3759–3771, https://doi.org/10.1007/s00382-018-4371-4, 2019.

Beverley, J. D., Woolnough, S. J., Baker, L. H., Johnson, S. J., Weisheimer, A., and O'Reilly, C. H.: Dynamical mechanisms linking Indian monsoon precipitation and the circumglobal teleconnection, Clim. Dyn., 57, 2615–2636, https://doi.org/10.1007/s00382-021-05825-6, 2021.

Briegel, L. M. and Frank, W. M.: Large-scale influences on tropical cyclogenesis in the western North Pacific, Mon. Weather Rev., 125, 1397–1413, https://doi.org/10.1175/1520-0493(1997)125<1397:LSIOTC>2.0.CO;2, 1997.

Di Capua, G., Kretschmer, M., Runge, J., Alessandri, A., Donner, R. V., van den Hurk, B., Vellore, R., Krishnan, R., and Coumou, D.: Long-Lead Statistical Forecasts of the Indian Summer Monsoon Rainfall Based on Causal Precursors, Weather Forecast., 34, 1377–1394, https://doi.org/10.1175/waf-d-19-0002.1, 2019.

Di Capua, G., Runge, J., Donner, R. V., van den Hurk, B., Turner, A. G., Vellore, R., Krishnan, R., and Coumou, D.: Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: causal relationships and the role of timescales, Weather Clim. Dyn., 1, 519–539, https://doi.org/10.5194/wcd-1-519-2020, 2020a.

Di Capua, G., Kretschmer, M., Donner, R. V., Van Den Hurk, B., Vellore, R., Krishnan, R., and Coumou, D.: Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: A theory-guided causal effect network approach, Earth Syst. Dyn., 11, 17–34, https://doi.org/10.5194/esd-11-17-2020, 2020b.

Challinor, A. J., Slingo, J. M., Wheeler, T. R., Craufurd, P. Q., and Grimes, D. I. F.: Toward a Combined Seasonal Weather and Crop Productivity Forecasting System: Determination of the Working Spatial Scale, 2003.

Charney, J. G. and Shukla, J.: Predictability of monsoons, in: Monsoon dynamics, 99–109, 1981.

Chen, M., Kumar, A., and Wang, W.: A study of the predictability of sea surface temperature over the tropics, Clim. Dyn., 44, 1767–1776, https://doi.org/10.1007/s00382-014-2187-4, 2015.

Chevuturi, A., Turner, A. G., Johnson, S., Weisheimer, A., Shonk, J. K. P., Stockdale, T. N., and Senan, R.: Forecast skill of the Indian monsoon and its onset in the ECMWF seasonal forecasting system 5 (SEAS5), Clim. Dyn., 56, 2941–2957,

https://doi.org/10.1007/s00382-020-05624-5, 2021.

Christidis, N., Jones, G. S., and Stott, P. a.: Dramatically increasing chance of extremely hot summers since the 2003 European heatwave, Nat. Clim. Chang., 5, 46–50, https://doi.org/10.1038/nclimate2468, 2014.

Coumou, D., Lehmann, J., and Beckmann, J.: The weakening summer circulation in the Northern Hemisphere mid-latitudes,

655   Science (80-.)., 348, 324–327, https://doi.org/10.1126/science.1261768, 2015.

Coumou, D., Kornhuber, K., Lehmann, J., and Petoukhov, V.: Weakened Flow Persistent Circulation, and Prolonged Weather Extremes in Boreal Summer, in: Climate Etremes: Patterns and Mechanisms, Geoph Monogr 226, 61–73, 2017.

Coumou, D., Di Capua, G., Vavrus, S., Wang, L., and Wang, S.: The influence of Arctic amplification on mid-latitude summer circulation, Nat. Commun., 9, 2959, https://doi.org/10.1038/s41467-018-05256-8, 2018.

660   Ding, Q. and Wang, B.: Circumglobal teleconnection in the Northern Hemisphere summer, J. Clim., 18, 3483–3505, https://doi.org/10.1175/JCLI3473.1, 2005.

Ding, Q. and Wang, B.: Intraseasonal teleconnection between the summer Eurasian wave train and the Indian Monsoon, J. Clim., 20, 3751–3767, https://doi.org/10.1175/JCLI4221.1, 2007.

Ding, Q., Wang, B., Wallace, J. M., and Branstator, G.: Tropical-extratropical teleconnections in boreal summer: Observed

665   interannual variability, J. Clim., 24, 1878–1896, https://doi.org/10.1175/2011JCLI3621.1, 2011.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, 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., Hardenberg, J. V., 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.,

670   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 Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

675   Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C., Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C., Simpson, I. R., Stouffer, R. J., and Williamson, M. S.: Taking climate model evaluation to the next level, Nat. Clim. Chang., 9, 102–110, https://doi.org/10.1038/s41558-018-0355-y, 2019.

680   Gadgil, S., Rajeevan, M., and Nanjundiah, R.: Monsoon prediction - Why yet another failure?, Curr. Sci., 88, 1389–1400, 2005.

Goddard, L. and Dilley, M.: El Niño: Catastrophe or opportunity, J. Clim., 18, 651–665, https://doi.org/10.1175/JCLI-3277.1, 2005.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Catherine, A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas,

685   V., Von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J. J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and Von Storch, J. S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hagger, V., Dwyer, J., Shoo, L., and Wilson, K.: Use of seasonal forecasting to manage weather risk in ecological restoration,
690   2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P.,
695   Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holton, J. R.: An introduction to dynamic meteorology, Fourth Edition., edited by: Cynar, F., Elsevier Academic Press, 1973.

Horak, J., Hofer, M., Gutmann, E., Gohm, A., and Rotach, M. W.: A process-based evaluation of the Intermediate Complexity Atmospheric Research Model (ICAR) 1.0.1, Geosci. Model Dev., 14, 1657–1680, https://doi.org/10.5194/gmd-14-1657-2021,
700   2021.

Huntingford, C., Mitchell, D., Kornhuber, K., Coumou, D., Osprey, S., and Allen, M.: Assessing changes in risk of amplified planetary waves in a warming world, Atmos. Sci. Lett., 20, 1–11, https://doi.org/10.1002/asl.929, 2019.

Jain, S., Scaife, A. A., and Mitra, A. K.: Skill of Indian summer monsoon rainfall prediction in multiple seasonal prediction systems, Clim. Dyn., https://doi.org/10.1007/s00382-018-4449-z, 2018.

705   Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: The new ECMWF seasonal forecast system, Geosci. Model Dev., 12, 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019, 2019.

Kornhuber, K., Coumou, D., Vogel, E., Lesk, C., Donges, J. F., Lehmann, J., and Horton, R. M.: Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions, Nat. Clim. Chang., 20, 48–53,
710   https://doi.org/10.1038/s41558-019-0637-z, 2020.

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J.: Using Causal Effect Networks to analyze different Arctic drivers of mid-latitude winter circulation, J. Clim., 29, 4069–4081, https://doi.org/10.1175/JCLI-D-15-0654.1, 2016.

Kretschmer, M., Runge, J., and Coumou, D.: Early prediction of extreme stratospheric polar vortex states based on causal precursors, Geophys. Res. Lett., 44, 8592–8600, https://doi.org/10.1002/2017GL074696, 2017.

715   Kretschmer, M., Cohen, J., Matthias, V., Runge, J., and Coumou, D.: The different stratospheric influence on cold-extremes in Eurasia and North America, npj Clim. Atmos. Sci., 1, 1–10, https://doi.org/10.1038/s41612-018-0054-4, 2018.

Kumar, A.: Statistical Models for Long-range Forecasting of Southwest Monsoon Rainfall over India Using Step Wise Regression and Neural Network, Atmos. Clim. Sci., 02, 322–336, https://doi.org/10.4236/acs.2012.23029, 2012.

Lazo, J. K., Morss, R. E., and Demuth, J. L.: 300 billion served, Bull. Am. Meteorol. Soc., 90, 785–798,
720 https://doi.org/10.1175/2008BAMS2604.1, 2009.

Lehmann, J., Kretschmer, M., Schauberger, B., and Wechsung, F.: Potential for early forecast of Moroccan wheat yields based on climatic drivers, Geophys. Res. Lett., 1–10, https://doi.org/10.1029/2020gl087516, 2020.

Mann, M. E., Rahmstorf, S., Kornhuber, K., Steinman, B. A., Miller, S. K., Petri, S., and Coumou, D.: Projected changes in persistent extreme summer weather events: The role of quasi-resonant amplification, Sci. Adv., 4, 1–10,
725 https://doi.org/10.1126/sciadv.aat3272, 2018.

Menon, A., Levermann, A., Schewe, J., Lehmann, J., and Frieler, K.: Consistent increase in Indian monsoon rainfall and its variability across CMIP-5 models, Earth Syst. Dyn., 4, 287–300, https://doi.org/10.5194/esd-4-287-2013, 2013.

Meza, F. J., Hansen, J. W., and Osgood, D.: Economic value of seasonal climate forecasts for agriculture: Review of ex-ante assessments and recommendations for future research, https://doi.org/10.1175/2007JAMC1540.1, 2008.

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

O'Reilly, C. H., Woollings, T., Zanna, L., and Weisheimer, A.: The impact of tropical precipitation on summertime euro-Atlantic circulation via a circumglobal wave train, J. Clim., 31, 6481–6504, https://doi.org/10.1175/JCLI-D-17-0451.1, 2018.

Palmer, T.: The primacy of doubt: Evolution of numerical weather prediction from determinism to probability,
735 https://doi.org/10.1002/2017MS000999, 1 June 2017.

Palmer, T. N. and Anderson, D. L. T.: The prospects for seasonal forecasting-A review paper, Q. 1. R. Meteorol. SOC, 755–793 pp., 1994.

Pfleiderer, P., Schleussner, C., and Coumou, D.: Boreal summer weather becomes more persistent in a warmer world, Nat. Clim. Chang., https://doi.org/10.1038/s41558-019-0555-0, 2018.

740 Pfleiderer, P., Schleussner, C.-F., Geiger, T., and Kretschmer, M.: Robust predictors for seasonal Atlantic hurricane activity identified with causal effect networks, Weather Clim. Dyn., 1, 313–324, https://doi.org/10.5194/wcd-1-313-2020, 2020.

Roberts, C. D., Vitart, F., and Balmaseda, M. A.: Hemispheric Impact of North Atlantic SSTs in Subseasonal Forecasts, https://doi.org/10.1029/2020GL091446, 28 February 2021.

Rodwell, M. J. and Hoskins, B.: Monsoons and the dynamics of deserts, Q. J. R. Meteorol. Soc., 122, 1385–1404, 1996.

745 Runge, J.: Causal network reconstruction from time series: From theoretical assumptions to practical estimation, 28, 075310, https://doi.org/10.1063/1.5025050, 2018.

Runge, J., Petoukhov, V., and Kurths, J.: Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models, J. Clim., 27, 720–739, https://doi.org/10.1175/JCLI-D-13-00159.1, 2014.

750 Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D.: Detecting causal associations in large nonlinear time series datasets, Sci. Adv., 5, eaau4996, 2019.

Scaife, A. A., Ferranti, L., Alves, O., Athanasiadis, P., Baehr, J., Dequé, M., Dippe, T., Dunstone, N., Fereday, D., Gudgel, R.

G., Greatbatch, R. J., Hermanson, L., Imada, Y., Jain, S., Kumar, A., MacLachlan, C., Merryfield, W., Müller, W. A., Ren, H. L., Smith, D., Takaya, Y., Vecchi, G., and Yang, X.: Tropical rainfall predictions from multiple seasonal forecast systems, Int. J. Climatol., 1–15, https://doi.org/10.1002/joc.5855, 2018.

Schepen, A., Wang, Q. J., and Robertson, D. E.: Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall, J. Geophys. Res. Atmos., 117, https://doi.org/10.1029/2012JD018011, 2012.

Seo, K. H., Wang, W., Gottschalck, J., Zhang, Q., Schemm, J. K. E., Higgins, W. R., and Kumar, A.: Evaluation of MJO forecast skill from several statistical and dynamical forecast models, J. Clim., 22, 2372–2388, https://doi.org/10.1175/2008JCLI2421.1, 2009.

Shukla, J.: Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting, Science (80-.)., 282, 728–731, 1998.

Shukla, J., Anderson, J., Baumhefner, / D, Brankovic, # C, Chang, @ Y, Kalnay, E., Marx, L., Palmer, T., Paolino, @ D, Ploshay, J., Schubert, / S, Straus, & D., Suarez, M., and Tribbia, J.: Dynamical Seasonal Prediction, Bull. Am. Meteorol. Soc., 82, 2593–2606, 2000.

Spirtes, P., Glymour, C., and Scheines, R.: Causation, prediction, and search, Boston: The MIT Press., 2000.

Stephan, C. C., Klingaman, N. P., and Turner, A. G.: A mechanism for the interdecadal variability of the Silk Road Pattern, J. Clim., 32, 717–736, https://doi.org/10.1175/JCLI-D-18-0405.1, 2019.

Teng, H. and Branstator, G.: Amplification of Waveguide Teleconnections in the Boreal Summer, Curr. Clim. Chang. Reports, 5, 421–432, https://doi.org/10.1007/s40641-019-00150-x, 2019.

Tietsche, S., Balmaseda, M., Zuo, H., Roberts, C., Mayer, M., and Ferranti, L.: The importance of North Atlantic Ocean transports for seasonal forecasts, Clim. Dyn., 55, 1995–2011, https://doi.org/10.1007/s00382-020-05364-6, 2020.

Tsonis, A. A. and Eisner, J. B.: Chaos, Strange Attractors, and Weather, 1989.

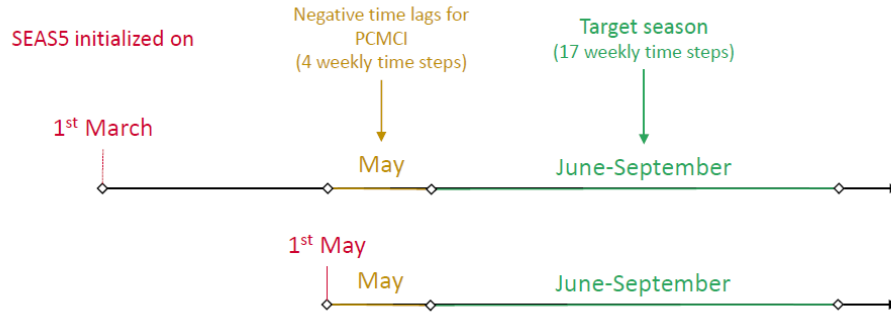
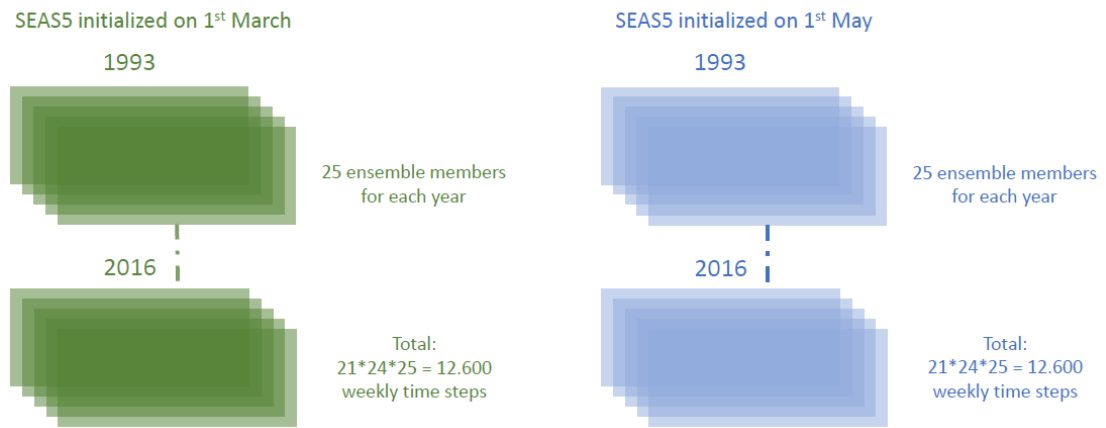
Turner, A. G. and Annamalai, H.: Climate change and the south Asian summer monsoon, Nat. Clim. Chang., 2, https://doi.org/10.1038/NCLIMATE1495, 2012.

Vijverberg, S. and Coumou, D.: The role of the Pacific Decadal Oscillation and ocean-atmosphere interactions in driving US temperature variability, npj Clim. Atmos. Sci., 5, 1–11, https://doi.org/10.1038/s41612-022-00237-7, 2022.

Vijverberg, S., Schmeits, M., van der Wiel, K., and Coumou, D.: Subseasonal Statistical Forecasts of Eastern U.S. Hot Temperature Events, Mon. Weather Rev., 148, 4799–4822, https://doi.org/10.1175/MWR-D-19-0409.1, 2020.

Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, J. R. Soc. Interface, 11, 20131162, https://doi.org/10.1098/rsif.2013.1162, 2014.

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.

Wilks, D. S.: Canonical Correlation Analysis (CCA), Int. Geophys., 100, 563–582, https://doi.org/10.1016/B978-0-12-385022-5.00013-0, 2011.

Figures

790   **Figure 1.** Schematic representation of the SEAS5 forecasting setup. Panel (a) shows the time line for SEAS5 initialization and target period. Panel (b) shows a schematic of the SEAS5 ensemble members.

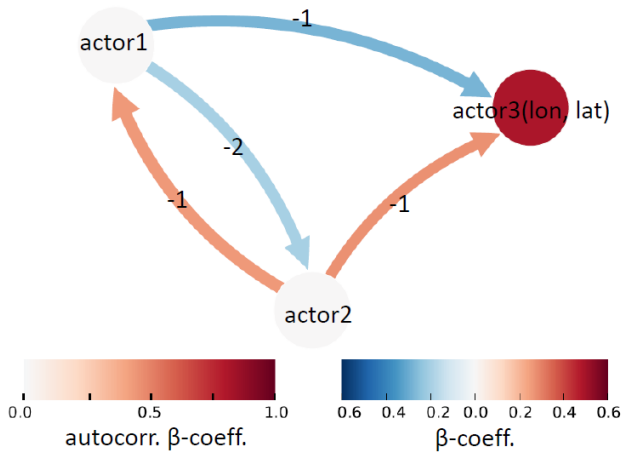
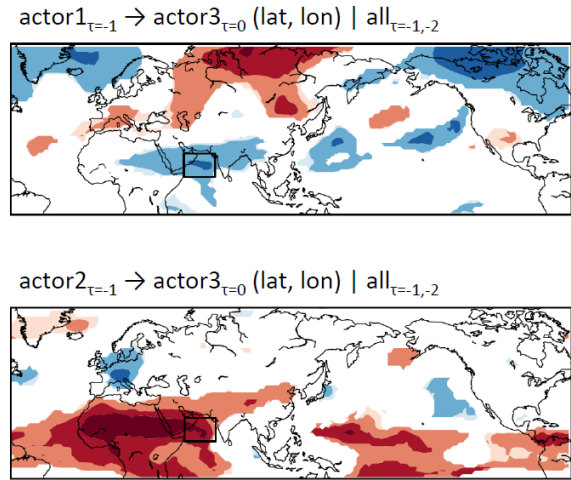



Figure 2. Schematic representation of CEN and causal map. Panel (a) an example of a CEN built with three actors and lag$_{max}$ -2. Panel (b) shows an example of a causal map: the β value for the causal effect of actor1 and actor2 on actor3 (a time-varying field) vary with the latitude and longitude in the map.

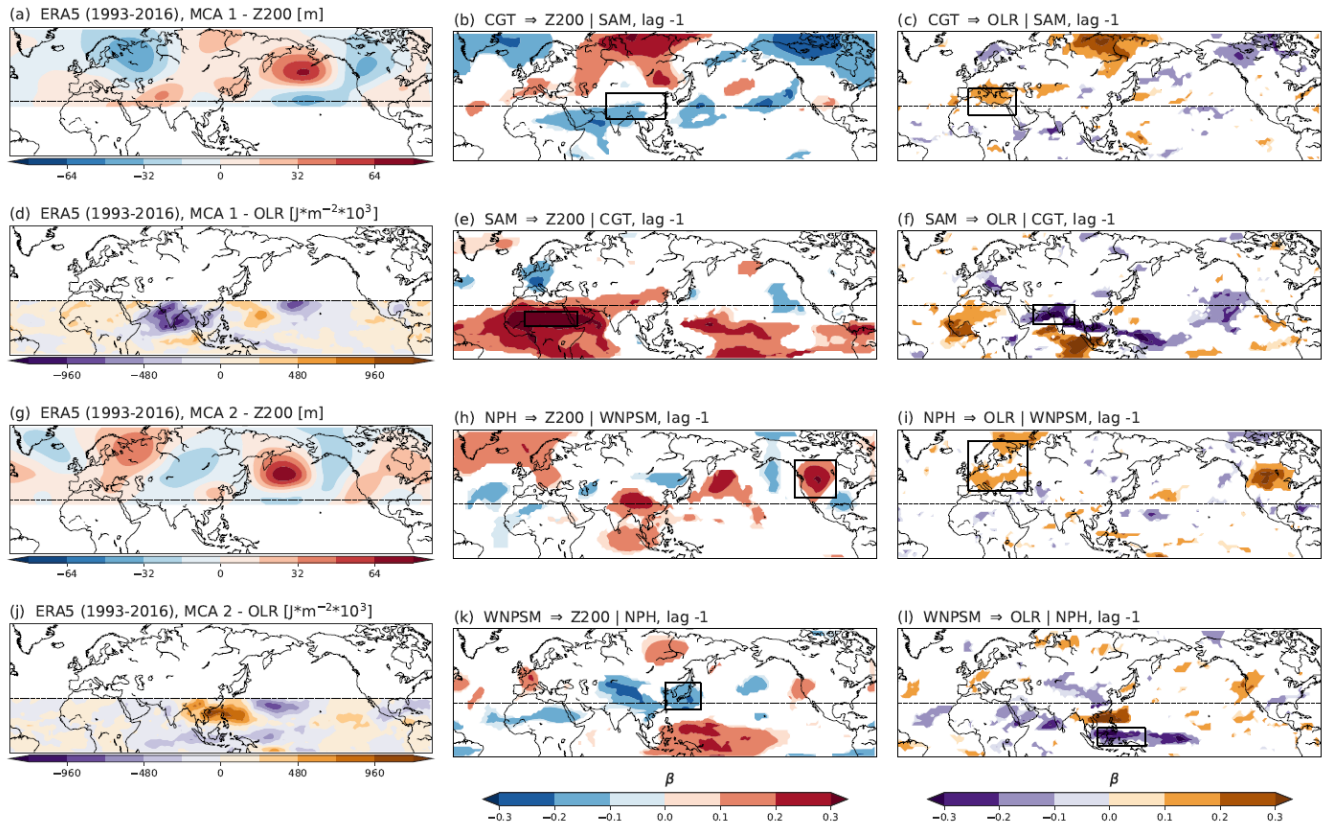



Figure 3. Causal maps for ERA5. Left column: MCA patterns for Z200 and OLR fields for ERA-S. MCA mode 1 Z200 (panel a) shows the CGT pattern. MCA mode 1 OLR shows the SAM (panel d). MCA mode 2 Z200 shows the NPH (panel g). MCA mode 2 OLR shows the WNPSM (panel j). Central column (Panels b,e,h,k): causal maps for the effect of each MCA pattern on Z200 fields. Right column (Panels c,f,i,l): causal maps for the effect of each MCA pattern on OLR fields. The black boxes highlight Southeast Asia (Panel b), the Mediterranean (Panel c), the Sahel region (Panel e) and India (Panel f) for MCA mode 1 and central western US (Panel h), eastern Europe (Panel i), Japan (Panel k) and the Maritime Continent (Panel l). Refer to the text in Section 3.2 for an interpretation of the results. Only grid points with corrected p-values significant at $\alpha = 0.1$ are shown.

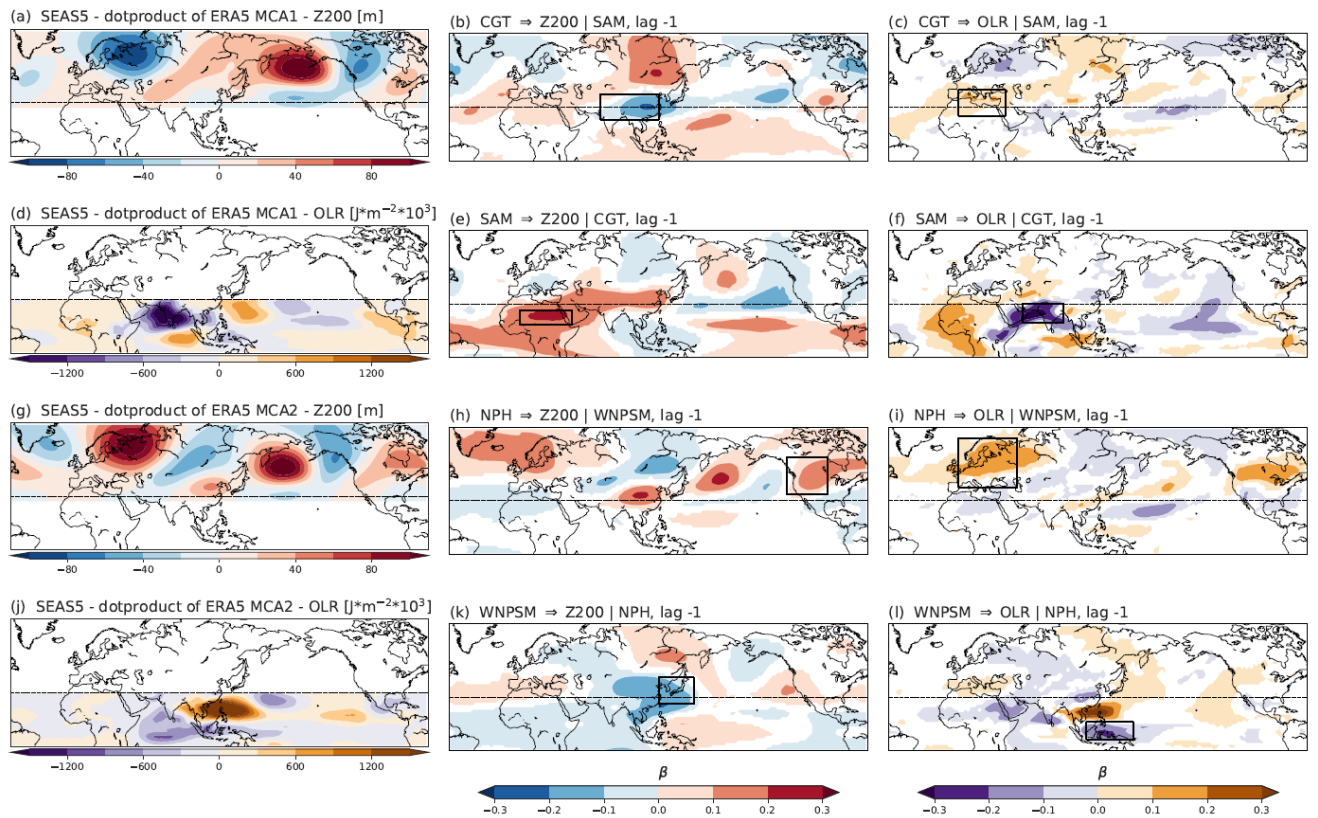



Figure 4. Causal maps for SEAS5. Same as for Fig. 3 but for SEAS5-ERA5 MCA (calculated as the projection of ERA-S on SEAS5 Z200 and OLR fields) and related causal maps. The black boxes highlight Southeast Asia (Panel b), the Mediterranean (Panel c), the Sahel region (Panel e) and India (Panel f) for MCA mode 1 and central western US (Panel h), eastern Europe (Panel i), Japan (Panel k) and the Maritime Continent (Panel l). Refer to the text in Section 3.2 for an interpretation of the results. Only grid points with corrected p-values significant at $\alpha = 0.05$ are shown (see Method section for further details).

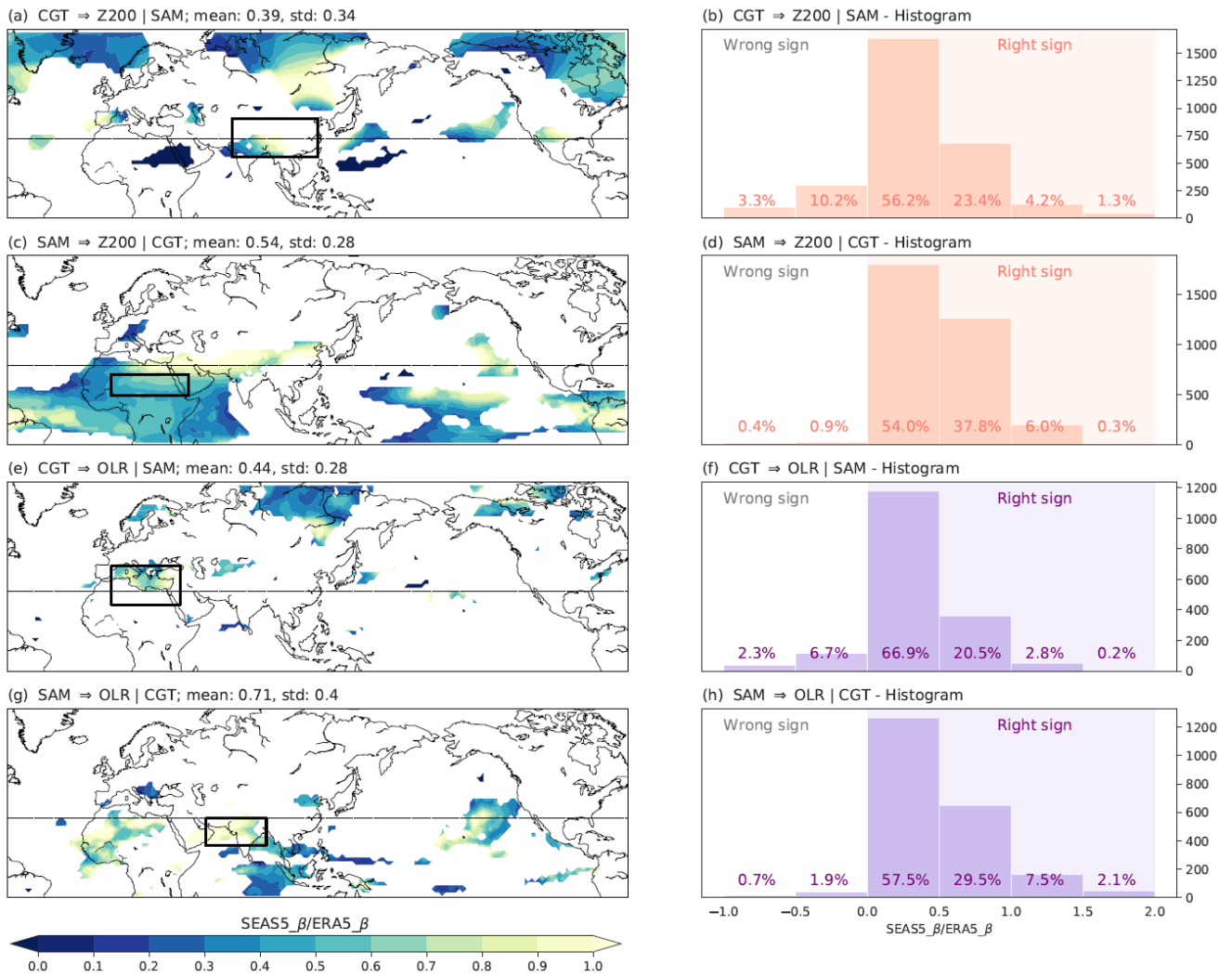



Figure 5. Differences between ERA5 and SEAS5 for MCA mode 1. Left column: Δ_β maps, where $\Delta_\beta = \frac{\beta_{SEAS5}}{\beta_{ERA5}}$ for CGT $\rightarrow$ Z200|SAM (panel a), SAM $\rightarrow$ Z200|CGT (panel c), CGT $\rightarrow$ OLR|SAM (panel e) and SAM $\rightarrow$ OLR|CGT (panel g). Right column: Histograms for SEAS5-ERA5 Δ_β for CGT $\rightarrow$ Z200|SAM (panel b), SAM $\rightarrow$ Z200|CGT (panel d), CGT $\rightarrow$ OLR|SAM (panel f) and SAM $\rightarrow$ OLR (panel h). All grid-points in the domain 15°S-75°N, 0°-360°E are used. The black boxes highlight Southeast Asia (Panel a), the Sahel region (Panel c), the Mediterranean (Panel e) and India (Panel g). The mean and the standard deviation for Δ_β values for each plot are shown in the plot title. Only grid points with corrected p-values significant at $\alpha = 0.05$ in both ERA5 and SEAS5-ERA5 causal maps are shown.

815

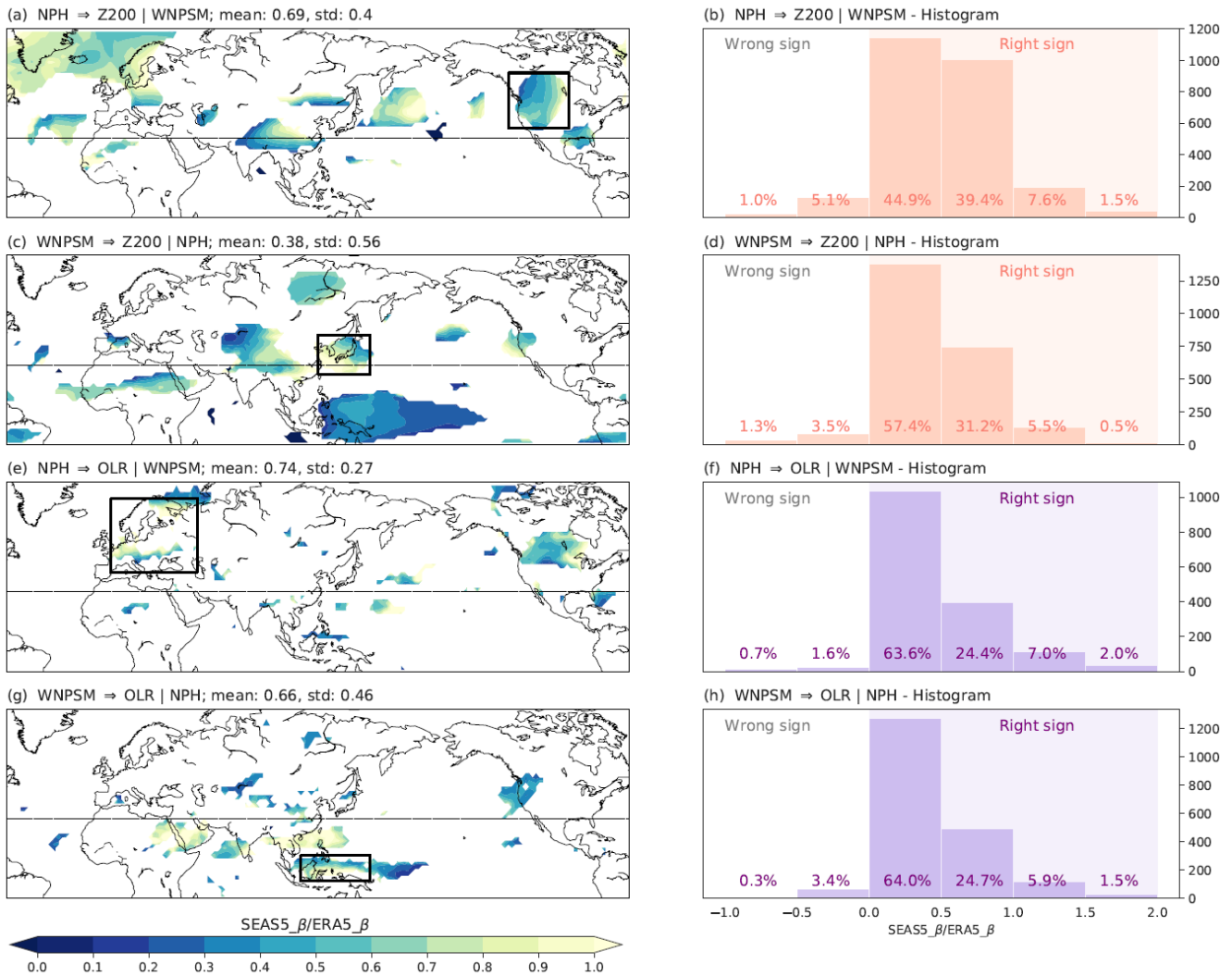
820

825

Figure 6. Differences between ERA5 and SEAS5 for MCA mode 2. Left column: Δ_β maps where $\Delta_\beta = \frac{\beta_{SEAS5}}{\beta_{ERA5}}$ for NPH → Z200|WNPSM (panel a), WNPSM → Z200|NPH (panel c), NPH → OLR|WNPSM (panel e) and WNPSM → OLR|NPH (panel g). Right column: Histogram for SEAS5-ERA5 Δ_β for NPH → Z200|WNPSM (panel b), WNPSM → Z200|NPH (panel d), NPH → OLR (panel f) and WNPSM → OLR|NPH (panel h). All grid-points in the domain 15°S-75°N, 0°-360°E are used. The black boxes highlight central western US (Panel a), Japan (Panel c), Eastern Europe (Panel e), and the Maritime Continent (Panel g). The mean and the standard deviation for Δ_β values for each plot are shown in the plot title. Only grid points with corrected *p*-values significant at $\alpha = 0.05$ in both ERA5 and SEAS5-ERA5 causal maps are shown.

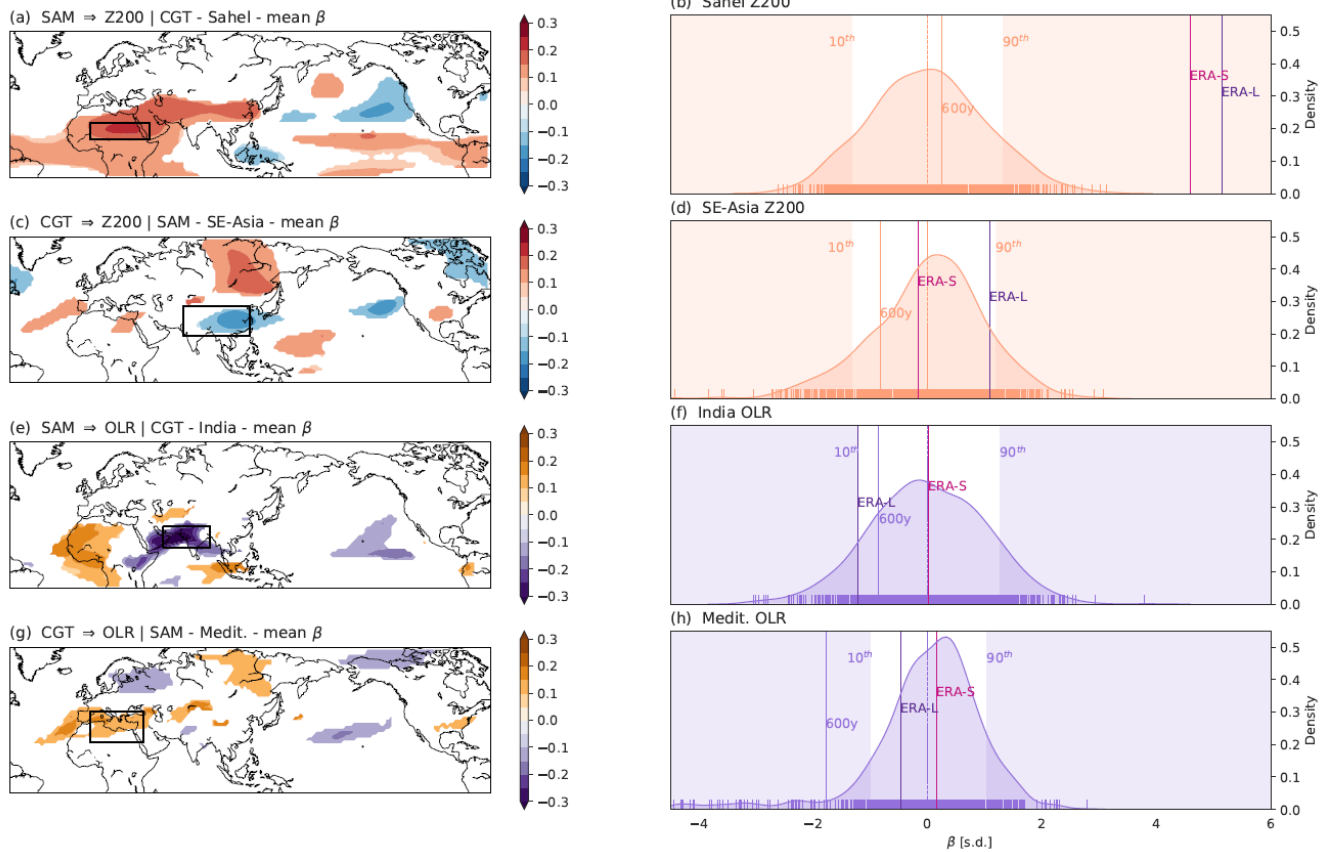



Figure 7. Ensemble spread for MCA mode 1 SEAS5-ERA5. Causal effect averaged over four key regions. Mean causal effect averaged over the 1000 subsamples for SAM → Z200|CGT (panel a), CGT → Z200| SAM (panel c), SAM → OLR| CGT (panel e) and CGT → OLR| SAM (panel g). PDF (standardized) of the distribution with highlighted 10th and 90th quantiles for the Sahel region (Panel b), Southeast Asia (Panel d), India (Panel f) and the Mediterranean (Panel h). The value for ERA-L (ERA-S) is shown in the PDFs for comparison by a vertical solid purple (magenta) line. The black boxes highlight the Sahel region (Panel a), Southeast Asia (Panel c), India (Panel e) and the Mediterranean (Panel g). The significance of the β values shown in the causal maps in panels a,c,e,g is described in Section 3.3.

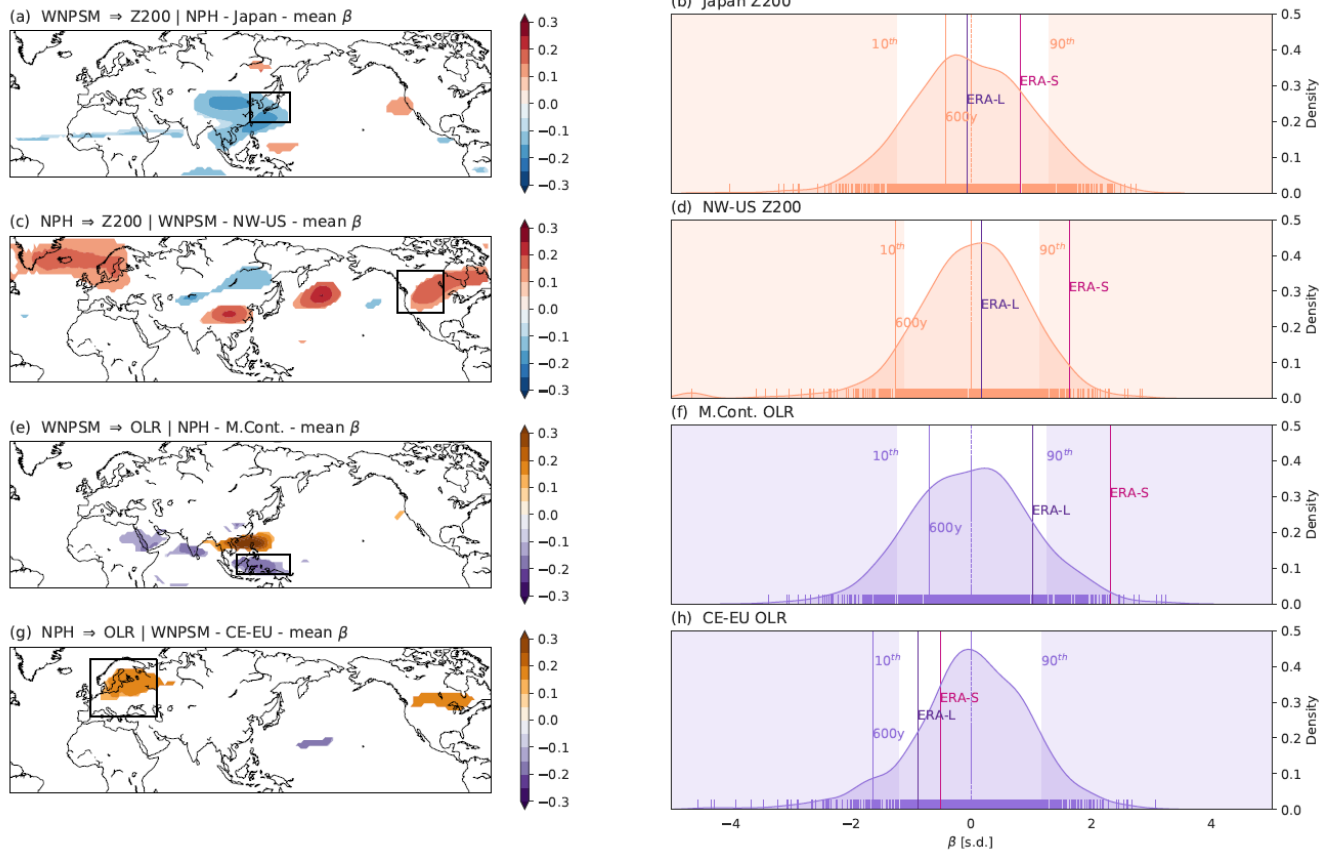



Figure 8. Ensemble spread for MCA mode 2 SEAS5-ERA5. Causal effect averaged over four key regions. Mean causal effect averaged over the 1000 subsamples for WNPSM → Z200|NPH (panel a), NPH → Z200|WNPSM (panel c), WNPSM → OLR|NPH (panel e) and NPH → OLR|WNPSM (panel g). PDF (standardized) of the distribution with highlighted 10th and 90th quantiles for Japan (Panel b), north-western US (Panel d), Maritime Continent (Panel f) and central eastern Europe (Panel h). The value for ERA-L (ERA-S) is shown in the PDFs for comparison by a vertical solid purple (magenta) line. The black boxes highlight Japan (Panel a), central western US (Panel c), the Maritime Continent (Panel e) and Eastern Europe (Panel g). The significance of the β values shown in the causal maps in panels a,c,e,g is described in Section 3.3.

850

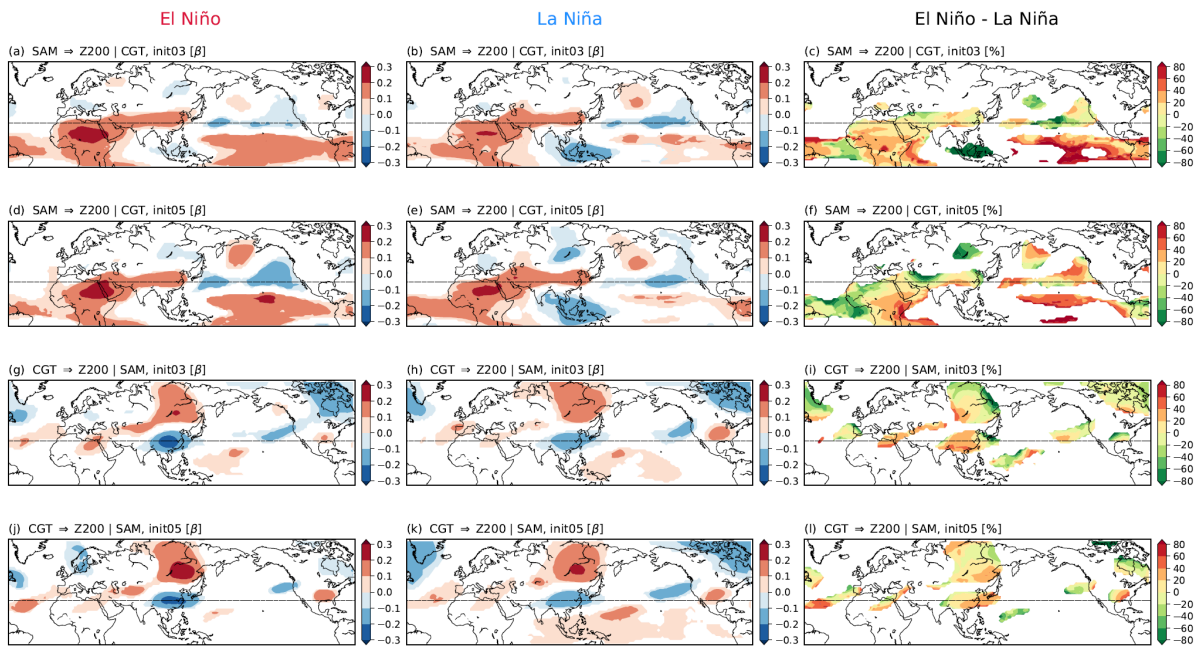



Figure 9. ENSO effect on tropical-extratropical links: MCA mode 1. Panels (a) and (b) show the causal maps for SEAS5 data initialized on the 1st of March for the SAM → Z200|CGT link respectively during Nino 3.4 positive and negative years. Panel (c) shows the differences between Nino 3.4 positive and negative years. Panel (d), (e) and (f): same as for panels (a), (b) and (c) but for SEAS5 data initialized on the 1st of May. Panels (g), (h) and (i): same as for panels (a), (b) and (c) but for the link CGT → Z200|SAM. Panels (j), (k) and (l): same as for panels (g), (h) and (i) but for SEAS5 data initialized on the 1st of May. Only grid points with corrected p-values significant at α = 0.05 are shown.

860

 Tables

ACRONYM	EXTENDED	METHOD – FIELD
SAM	South Asian Monsoon	MCA mode 1, OLR
WNPSM	Western North Pacific summer monsoon	MCA mode 2, OLR
CGT	Circumglobal teleconnection	MCA mode 1, Z200
NPH	North Pacific High	MCA mode 2, Z200
ERA-S	ERA5 short (1993-2016)	
ERA-L	ERA5 long (1979-2020)	
SEAS5-ERA5	SEAS5 (1993-2016), MCA as projection of ERA5-S MCA on SEAS5 data	
SEAS5	SEAS5 (1993-2016), MCA calculated directly on SEAS5	

Table 1: List of the main acronyms used in the manuscript.

	ERA-S – ERA-L	ERA-S – SEAS5		ERA-S – SEAS5-ERA
	$MCA_i - MCA_i$ (i=j)	$MCA_i - MCA_i$ (i=j)	$MCA_i - MCA_j$ (i≠j)	$MCA_i - MCA_i$ (i=j)
MCA1 Z200 (CGT)	0.89	0.61	-0.40	0.93
MCA2 Z200 (NPH)	0.84	0.60	0.57	0.93
MCA1 OLR (SAM)	0.85	0.58	-0.47	0.86
MCA2 Z200 (WNPSM)	0.77	0.40	0.44	0.83

Table 2: Spatial pattern correlation between MCA modes obtained from ERA5 data over the periods 1979-2020 and 1993-2016,
 between ERA5 and SEAS5 data over the common period 1993-2016 and for the same period but as the projection of ERA5 MCA (see main text for more details). Numbers highlighted in bold are significant at p-value = 0.05.

	REGION	SPATIAL DOMAIN	CAUSAL LINK
MCA MODE 1	Sahel	13°-45°N, 0°-45°E	SAM → Z200|CGT
	Southeast Asia	20°-40°N, 70°-120°E	CGT → Z200|SAM
	India	15°-30°N, 55°-90°E	SAM → OLR|CGT
	Mediterranean	23°-43°N, 0°-40°E	CGT → OLR|SAM
	Japan	25°-45°N, 120°-150°E	WNPSM → Z200|NPH
MCA MODE 2	Central western US	35°-60°N, 230°-265°E	NPH → Z200|WNPSM
	Maritime Continent	5°S-10°N, 110°-150°E	WNPSM → OLR|NPH
	Central eastern Europe	40°-70°N, 0°-50°E	NPH → OLR|WNPSM

Table 3: Spatial domains of selected β regions.

875