Technical Memo

922

Multi-decadal variability of tropical rainfall: reconciling ECMWF reanalyses and GPCP data

Franco Molteni and Anca Brookshaw

December 2024

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract

We present results from a statistical procedure aimed at finding temporally coherent signals in tropical rainfall from different datasets. Specifically, we compare rainfall from the ERA5 re-analysis in the period 1950-2022 with a dataset obtained by merging GPCP v3.2 data for 1983-2022 with 1950-1982 rainfall estimated by a regression of GPCP data against rainfall indices from the CERA20C re-analysis.

For each sub-sample of 3 consecutive months, a canonical correlation analysis (CCA) is performed on the first 21 principal components (PCs) of each dataset; from the CCA, two sets of rotated PCs (one for each dataset) are found, which are orthogonal within each set and highly correlated with the corresponding rotated PC (RPC) of the other dataset. The associated spatial patterns are computed taking the covariance of the RPCs with the original rainfall anomalies in each grid point. Filtered anomalies are then reconstructed from the first 16 RPCs and the associated spatial patterns (REOFs). The purpose of this procedure is to filter out variability components and trends which are poorly correlated across the two datasets, and which may therefore originate from changes in the assimilation methodology or in the availability of data sources across the 72-year period.

The two leading RPCs and REOFs are representative of rainfall variability associated with El Niño - Southern Oscillation (ENSO), and show a striking similarity across the two datasets. Subsequent modes, representative of rainfall variability over the Indian Ocean, are also well correlated but show differences in regional features. Looking at correlations of reconstructed grid-point data from the merged-GPCP and ERA5 data, it is found that the CCA-filtered anomalies provide a reliable estimate of rainfall variability in the ocean regions with large rainfall variability, but are unable to improve the poor correlation between the merged-GPCP and the ERA5 data in some dry regions such as the Sahara and upwelling regions on the western side of Africa and South America.

Finally, examples of tropical and extratropical teleconnection patterns computed from the CCA-filtered datasets in different period are presented, which show the potential for uncertainty estimation in diagnostics of inter-annual and inter-decadal variability of tropical rainfall and its remote impacts.

Plain Language Summary

The variability of tropical rainfall has a significant impact on both the tropical and extratropical circulation through the generation of anomalous planetary waves. These teleconnections determine the extra-tropical response to tropical phenomena with high predictability on seasonal to interannual scales, such as El Niño. It is therefore important to assess if such teleconnections are accurately reproduced in global numerical models used for long-range forecasting and climate simulations.

Estimates of tropical rainfall from satellite data extend back to the early 1980s. Climate re-analyses spanning several decades also provide tropical rainfall estimates covering longer periods (as in the ECMWF ERA5 re-analysis, which goes back to the 1940s). However, rainfall estimates in data-sparse regions from re-analyses may be affected by errors in the assimilating models. Here, we present results from a statistical methodology which extracts time-correlated signals from two datasets, namely ERA5 and GPCP rainfall data extended into the pre-satellite era using calibrated data from the CERA20C re-analysis. The leading modes of variability and their teleconnections from the two datasets are compared to assess their consistency and reliability.

1. Introduction

The importance of tropical rainfall variability as a source of diabatic heating anomalies which generate planetary-scale teleconnections has been recognized for many decades, since observational and dynamical studies in the early 1980s posed the foundations of our understanding of tropical-extratropical interactions (Horel and Wallace 1981, Hoskins and Karoly 1981 and many others).

The rainfall anomalies associated with the strongest teleconnections are mostly located over the tropical oceans, where surface measurements of precipitation are not available. However, since 1979 rainfall estimates have been obtained by processing radiances recorded by increasingly sophisticated satellite instruments, calibrated using data from tropical islands and maritime regions. Some of these satellite estimates have been merged with raingauge data from the continents to produce gridded precipitation products of global coverage, such as the GPCP (Adler et al. 2003) and CMAP (Xie and Arkin 1997) datasets. These datasets have been extensively used to investigate rainfall variability and the associated teleconnections on the interannual time scale.

An alternative source of information about tropical rainfall variability is provided by atmospheric and climate re-analyses. Here, rainfall estimates are produced by very-short-range forecasts started from subsequent atmospheric states. In addition to their global coverage, rainfall data from climate re-analyses may extend beyond the satellite era, potentially allowing an assessment of inter-decadal variability of tropical rainfall. The well-known drawback of using rainfall data from climate re-analyses is that such data are affected by model errors which often have a significant systematic component. This is particularly true for tropical regions, where the fast time scale of convective processes allows model error to grow even within the limited time scale of the assimilation cycle.

In this respect, it is useful to distinguish between century-long re-analyses using only surface and near-surface data available throughout the period, such as the NOAA-CIRES-DOE 20th Century Reanalysis (20CR; Compo et al. 2011, Slivinski et al. 2021) and the ECMWF ERA20C (Poli et al. 2016) and CERA20C (Laloyaux et al. 2018), and re-analyses using all available data in a given period, such as the NCEP-NCAR Reanalysis (Kalnay et al. 1996), the JRA-55 from the Japan Meteorological Agency (Kobayashi et al. 2015) or the recent ECMWF ERA5 (Hersbach et al 2020). In the former type, which for simplicity we call “homogeneous re-analyses”, the limited constraint provided by surface data allows model errors to affect rainfall data in a stronger way (on average) than in “full-data” re-analyses for the satellite era. On the other hand, in full-data re-analyses spanning several decades the substantial changes in the observing system through the decades produce a stronger constraint on model errors in recent periods (with abundant satellite data) than in earlier decades, especially before the satellite era. If a significant systematic model error is present in a region, its gradual reduction due to the increasing observational constraint may generate spurious inter-decadal variability and trends in full-data re-analyses.

If we assume that in homogeneous re-analyses the systematic model error is nearly uniform across the re-analysis period, a calibration against a different observational product can be performed using data from the common period, and then used to correct the re-analyses data in earlier periods. This approach was followed in the study by Molteni et al. (2020), where tropical rainfall variability from historical simulations from a range of European climate models was compared with a combined rainfall estimate from CERA20C and GPCP v2.3 data. Specifically, GPCP data were first regressed against local and large-scale predictors derived from CERA20C data in the 1979-2010 period. The regression was then used to estimate rainfall anomalies consistent with GPCP statistics from CERA20C data in the full 1950-2014 period covered by the historical simulations. A full description of the methodology is given in the Appendix of Molteni et al. (2020); in the following, we will refer to this rainfall dataset as the "calibrated CERA20C rainfall".

Possibly the most intriguing result obtained from the calibrated CERA20C rainfall dataset was that the difference in Indo-Pacific rainfall between the 1981-2010 and the 1951-1980 periods during boreal winter resembled the rainfall anomaly associated (during the most recent decades) with a positive-NAO response in the North Atlantic (see Fig. 14 in Molteni et al. 2020). This suggested that interdecadal variability in tropical Indo-Pacific rainfall may have played an important role in the wintertime shift towards positive North Atlantic Oscillation (NAO) states in the second part of the 20th century. A link between changes in Indo-Pacific SST and the wintertime North Atlantic circulation was already suggested by Hoerling et al. (2004) two decades ago; however, atmospheric model simulations driven by observed SST only managed to reproduce (at most) half of the amplitude of the observed extratropical trend (Hurrell et al. 2004; Sanchez-Gomez et al. 2008), and results appeared to be sensitive to the strength of the rainfall response in the Indian Ocean (Deser and Phillips 2006). A reliable observational estimate of tropical rainfall anomalies before the satellite era would be extremely useful to assess whether the sources of tropical-extratropical teleconnections are realistically reproduced in historical GCM simulations.

With the availability of ERA5 data starting from the 1940s (see https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5), it is easy to test whether the tropical interdecadal variability shown in the calibrated CERA20C rainfall dataset is reproduced in the most recent ECMWF re-analysis. As we discuss in detail in Sect. 4, taking a straightforward difference of ERA5 rainfall between consecutive multi-decadal periods shows a pattern with significant differences with respect to the earlier, calibrated re-analysis. Therefore, questions arise about the reliability of long-term tropical trends in the two products, whether coherent signals are present in both on the interdecadal scale, and what uncertainty is associated with such estimates.

In this study, we approach the problem by looking for patterns of tropical rainfall variability whose time evolution is coherently (albeit not identically) reproduced in different multi-decadal datasets, namely ERA5, the calibrated CERA20C product and the most recent GPCP version. The statistical methodology is described in Sect. 2, and variability modes derived from different datasets are compared in Sect.3. Sect. 4 investigates how extratropical teleconnections with tropical rainfall are reproduced using different rainfall products. Results are summarised and discussed in Sect. 5.

2. Methodology

The overall goal of this study is to compare tropical rainfall from ERA5 with other datasets where errors statistics (especially their biases) are expected to be relatively uniform in time. Here we used two such datasets:

- The calibrated CERA20C rainfall mentioned above (Molteni et al. 2020), which covers the period 1950-2010;

- The recent version 3.2 of the GPCP rainfall product, covering the 1983-2022 period (Huffman et al. 2022)

In practice, the calibrated CERA20C is a statistical estimate of GPCP (v.2.3) rainfall data based on CERA20C predictors, and by construction its large-scale modes of variability are very similar to those of the original GPCP data. Although the recent version 3.2 has a much higher spatial resolution (0.5x0.5 degrees) and calibration procedures are improved, we do not expect major differences in the large-scale modes of variability between the two versions. Therefore we decided to merge rainfall anomalies from GPCP v3.2 and the calibrated CERA20C (computed as deviations from the respective means in the overlapping 1983-2010 period) into a single dataset covering the 72-year period from Aug. 1950 to Oct. 2022, which was the latest month from which the GPCPv3.2 data were available when this study was started. For this purpose, data in the 40N-40S band from both datasets were re-gridded onto a 360x80 point, 1-degree regular grid starting at (0.5 E, 39.5 N). For brevity, we will refer to this dataset as the 'merged-GPCP'. For ERA5, rainfall anomalies were also computed on the same grid and for the same period.

For each month of the year, the following analyses were performed, using anomalies from that month, as well as from the previous and following month (for example, statistics for July were computed from June, July and August anomalies in 1951 to 2022):

- An Empirical Orthogonal Function (EOF) analysis was performed on both the ERA5 and the merged-GPCP rainfall, retaining the first 21 EOFs and principal components (PCs), which explain about 50% of the variance in each month;
- A canonical correlation analysis (CCA) was performed on the first 21 normalised PCs of the two datasets; this produces a rotation of the PCs which maintains their orthogonality in time (within each set), and makes each rotated PC in one set correlated only with one rotated PC of the other set: the rotated PCs are then ordered according to their correlation in time, starting from the highest correlation.
- For each month, only 16 rotated modes were retained; this further truncation ensured that a correlation greater than 0.7 (or 0.68 during boreal summer) existed between each couple of rotated PCs (see Figs. A1 and A2 in the Appendix for more detailed statistics).
- For each rotated PC, the associated spatial pattern was reconstructed by taking the covariance of the rotated PC with the original anomalies; although for brevity we will refer to these patterns as 'rotated EOFs', it should be noted that they are no longer spatially orthogonal (a PC/EOF rotation can maintain the orthogonality either in time or in space, but not both).
- For the central month of each 3-month sample, rainfall anomalies (with respect to the 72-year period) were then reconstructed using the 16 rotated PCs and EOFs.; these anomalies were then aggregated into two continuous datasets (for ERA5 and for the merged-GPCP data) spanning months from October 1950 to September 2022.

If spurious trends existed in either dataset, these are likely to be uncorrelated between the two datasets; conversely, if a trend is present in highly correlated PCs of the two sets, we can be reasonably confident about the real nature of such a signal. In practice, our procedure amounts to a space-time filtering of anomalies in the ERA5 and merged-GPCP rainfall which only retains the most temporally coherent part of the signal.

3. Leading modes of variability from canonical correlation analysis

In this section, we show some examples of the leading modes of tropical rainfall obtained from the CCA of DJF data; as stated above, these modes were then used in the reconstruction of rainfall data for January.

Fig. 1 shows the time series of the first rotated PC (or RPC) from the merged-GPCP and ERA5 data (top panel) for the 72 January months in the 1951-2022 period, and the associated spatial patters (REOFs) in the bottom panels. Since the REOFs are obtained by computing the covariance of the rainfall anomalies in any grid-point with the RPCs, they can be defined over the whole globe; however, they only provide an "optimal" description of rainfall variability in the band 40 N-40 S, where the EOF analysis was carried out.

It is evident that this first mode represents the typical rainfall variability induced by the El Nino-Southern Oscillation (ENSO) phenomenon; the time series from the two datasets are hardly distinguishable, with a time correlation of 0.996, and their largest values in January 1983, 1998 and 2016. The associated spatial patterns are also very similar in both phase and amplitude, with a local maximum around the so-called NINO4 region, from 160 E to 150 W. We can therefore have a high confidence in the representation of this mode and its multi-decadal variability in both datasets.

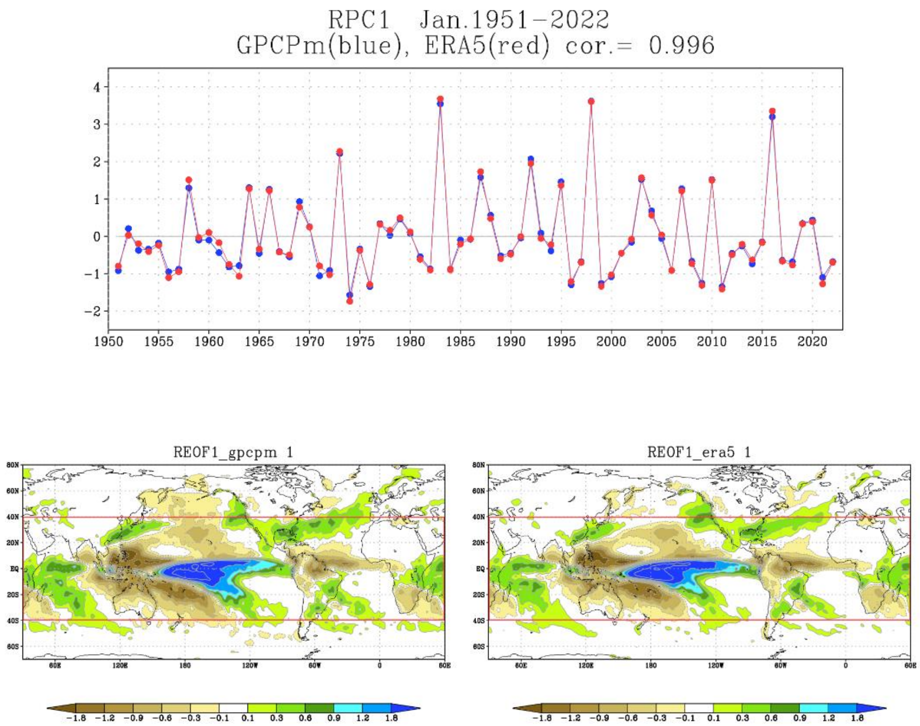

Fig. 1 *Top: Rotated PC1 from the merged-GPCP (blue line) and ERA5 (red line) rainfall data. Bottom: associated spatial patterns (REOF1) from the two datasets.*

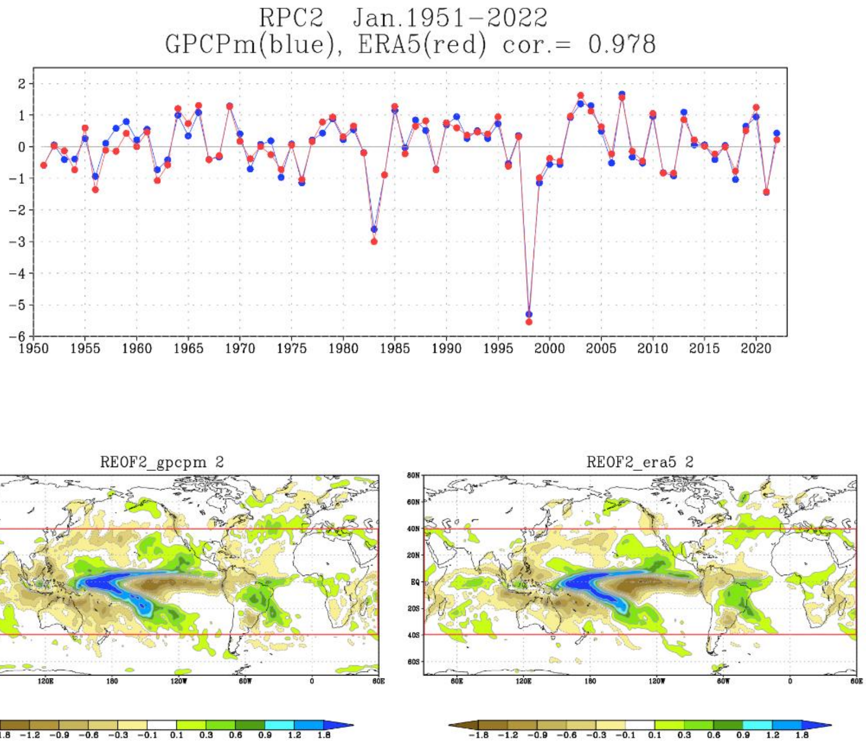

Fig. 2 *As in Fig. 1, for RPC2 and REOF2 from the merged-GPCP and ERA5 rainfall anomalies.*

A very good agreement is also found for the second mode (in Fig. 2), which is representative of the rainfall shift between the so-called ‘central Pacific’, or ‘Modoki’, and the ‘eastern Pacific’ ENSO episodes. The two RPC time series have a correlation of 0.978, with the largest negative values (indicating strong eastern Pacific events) in January 1998 and 1983.

For the third mode, the time correlation between the merged-GPCP and ERA5 time series remains very high (0.964), but some differences can be noticed in the amplitude of regional features in the two REOFs. Over the Indian-West Pacific sector, positive anomalies prevail over the tropical Indian Ocean, with negative anomalies further east, in both datasets. Although the local maxima are located off the Equator in both REOFs, the anomalies in the merged-GPCP dataset show a more uniform distribution in latitude, while the ERA5 pattern has stronger meridional gradients. Over the central and eastern Pacific, the third mode represents a northward shift of the ICTZ and a westward shift of the SPCZ; with good similarity between the features in the two REOFs.

Another important mode to describe the Indian Ocean rainfall variability is the RPC/REOF 5 (Fig. 4). Here, the dominant feature is a positive rainfall anomaly extending from the eastern Indian Ocean to the Maritime Continents, connected with below-average rainfall over eastern Brazil and equatorial eastern Africa. The peculiarity of the time series (again, well correlated between the two datasets) is the clear shift towards positive values after 2005.

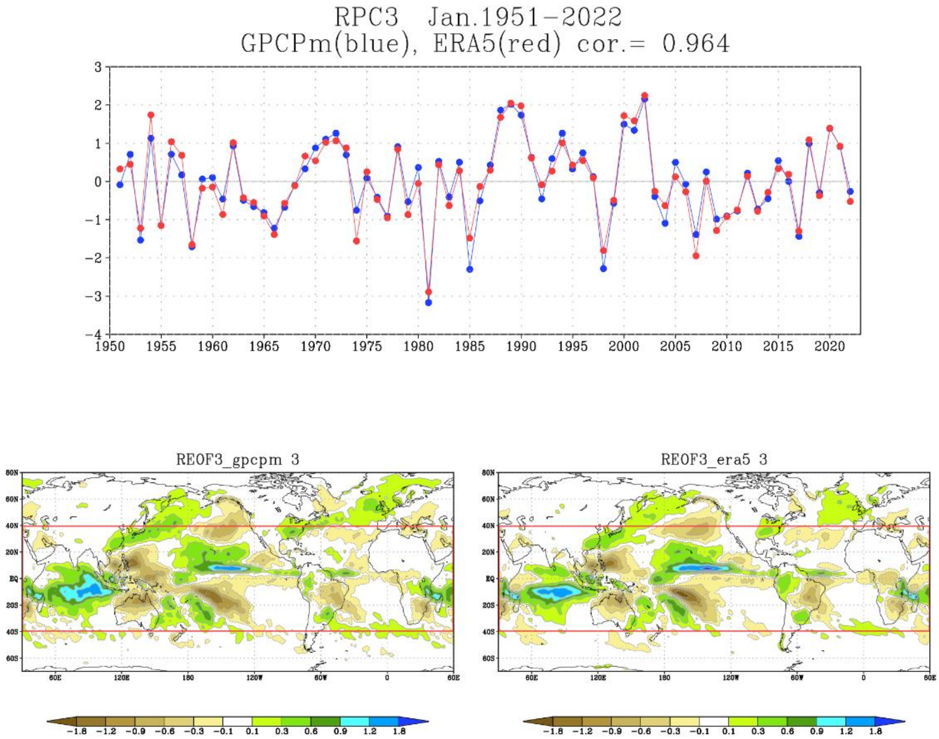

Fig. 3 *As in Fig. 1, for RPC3 and REOF3 from the merged-GPCP and ERA5 rainfall anomalies.*

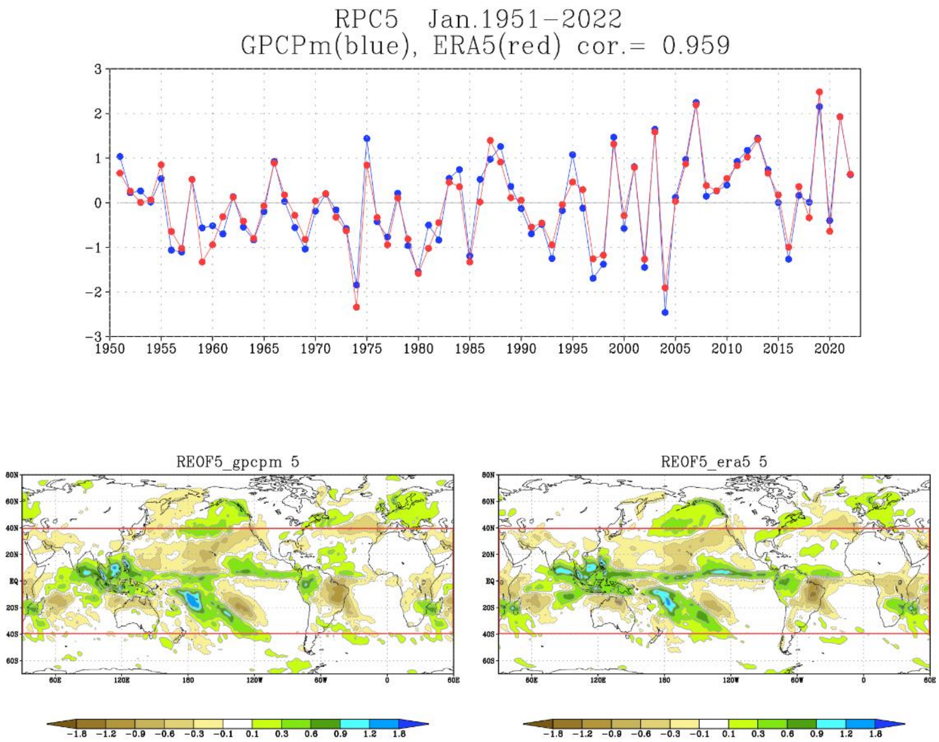

Fig. 4 *As in Fig. 1, for RPC5 and REOF5 from the merged-GPCP and ERA5 rainfall anomalies.*

On the basis of these leading modes, one may conclude that the tropical rainfall variability described by ERA5 is almost identical to the variability in the merged-GPCP dataset. However, differences in the time series (especially for the 1950s and 1960s) and some regional features in the REOFs become larger in higher-order modes, so that looking at the total reconstructed anomalies may lead to less optimistic conclusions. Fig. 5 shows the correlation of grid-point data in January from the two data sets, from both unfiltered anomalies and from those reconstructed from the CCA output; the corresponding maps for other months are shown in the Appendix. The correlations are shown for the 1951-1980 and 1981-2022 period separately. Although the correlations are high in regions of large rainfall, and are enhanced by the filtering procedure almost everywhere, one can see that areas of poor correlation still exist over some dry regions (Sahara, upwelling regions on the western side of Africa and South America) which are not improved by the filtering procedure.

In the next section, we use the anomalies reconstructed from the 16 selected RPCs to investigate inter-decadal variability in mean rainfall anomalies and teleconnections.

4. Inter-decadal variability and teleconnections.

We begin this section by checking whether the inter-decadal rainfall variability detected from the calibrated CERA20C data is also reproduced in the merged-GPCP and ERA5 anomalies. In Fig. 6 we show the difference between DJF tropical rainfall in the period 1981-2022 (fully residing in the so-called satellite era) and in the preceding 30 years 1951-1980 for the two datasets, comparing the results obtained from the original anomalies and after the CCA filtering.

Although both datasets show an increase in rainfall in the central Pacific after 1980, a number of regional features show significant differences. Over the tropical Indian Ocean, the merged-GPCP dataset shows a post-1980 increase in rainfall covering most of the western and central part of the basin, with a maximum around 10 S. In the original ERA5 data, the maximum increase is shifted over the eastern part of the basin, while in the western part a decrease in rainfall along the Equator compensates the increase around 10 S. However, the differences from the merged-GPCP results are reduced by the CCA filtering: the filtered ERA5 data show a post-1980 maximum increase around 90 E, the amplitude of the negative change along the Equator is reduced and a more uniform negative change is found east of 120 E. Other notable differences occur over the near-equatorial regions of South America and Africa, where the ERA5 data show much stronger inter-decadal changes than the merged-GPCP data (in both the original and the filtered versions), although there is an overall agreement about the sign of the change.

We now investigate to what extent the teleconnections of tropical Indo-Pacific rainfall have changed between the pre-1980 and post-1980 periods. As in Johnson et al. 2019 and Molteni et al. (2020), we focus on the western-central Indian Ocean (WCIO; 40E-90E, 10S-10N) and a meridionally-wider NINO4 region (160E-150W, 10S-10N) as two key sources of tropical and extratropical teleconnections. Because of our specific interest in teleconnections for the North Atlantic and Europe (NAE), we diagnose these teleconnections in the periods when the NAE signals are particularly strong and consistent: December-January (DJ) for WCIO, January-February (JF) for NINO4.

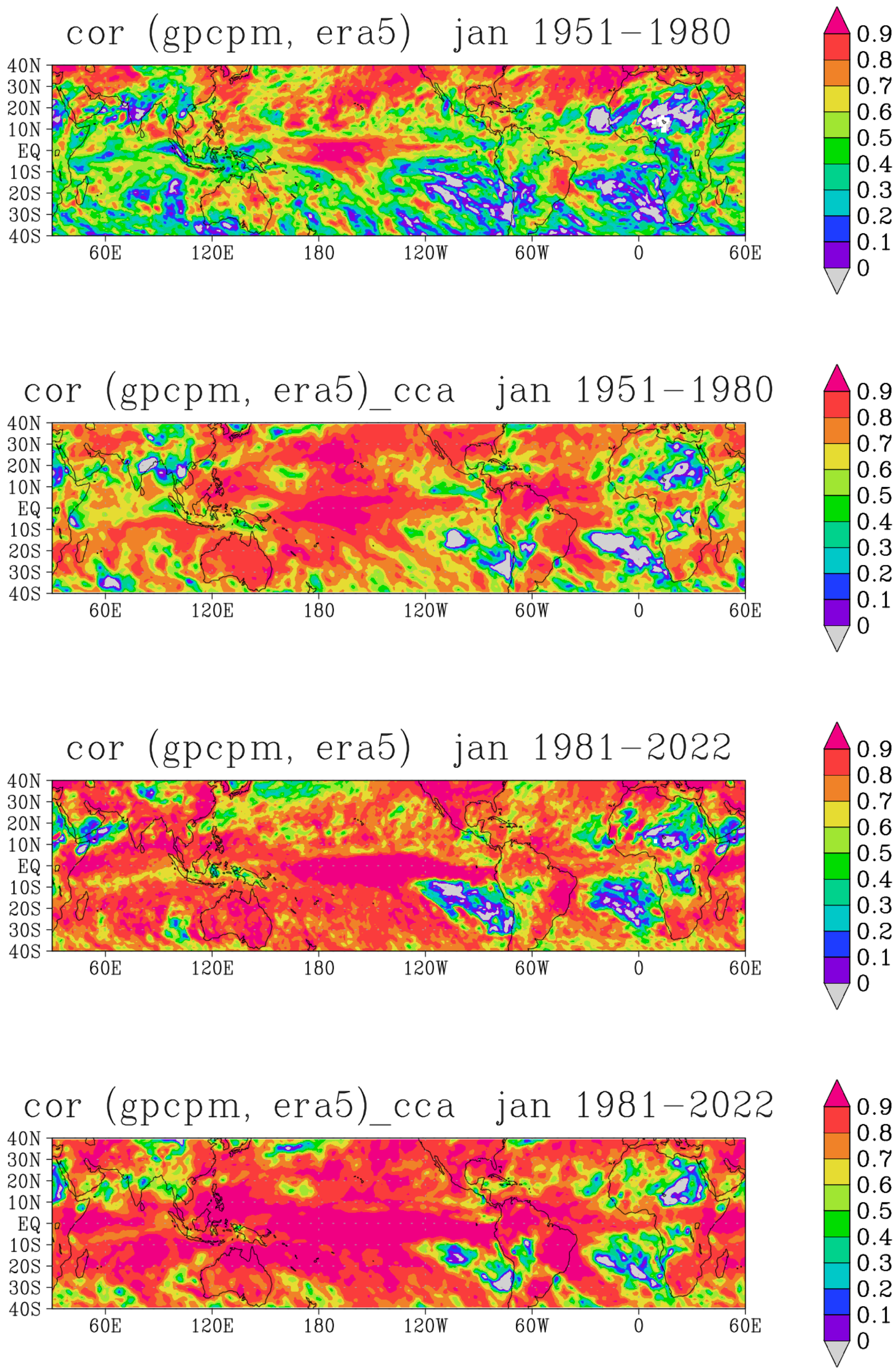

Fig. 5 *Correlation of unfiltered and CCA-filtered rainfall anomalies from the merged-GPCP and ERA5 datasets. From top to bottom: unfiltered data, Jan 1951-1980; CCA-filtered data, Jan 1951-1980; unfiltered data, Jan 1981-2022; CCA-filtered data, Jan 1981-2022.*

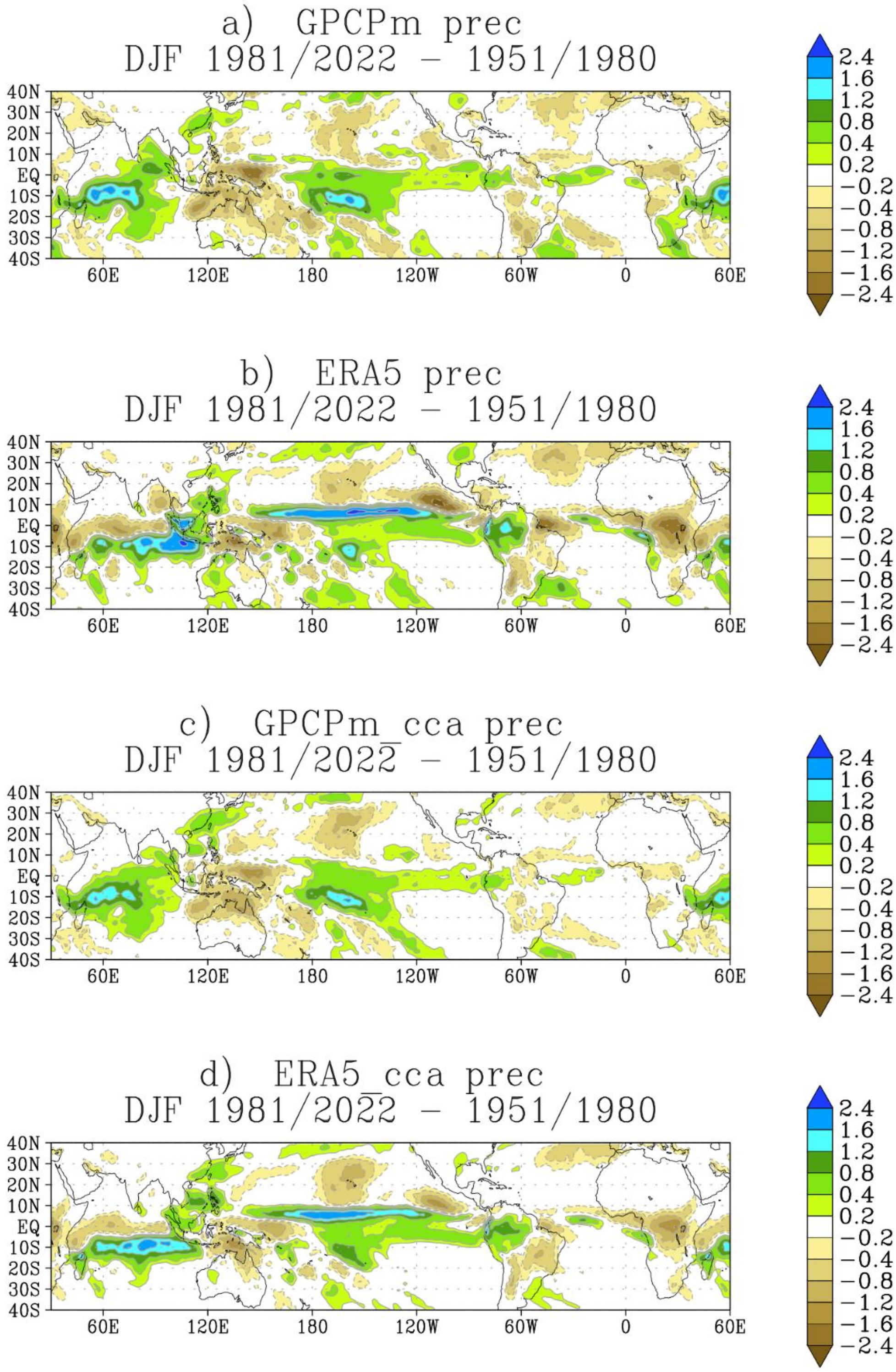

***Fig. 6** Tropical rainfall differences in DJF between the 1981-2022 and 1951-1980 periods, from different datasets: a) unfiltered merged-GPCP; b) unfiltered ERA5; c) CCA-filtered merged-GPCP; d) CCA-filtered ERA5.*

The time series of rainfall anomalies over the WCIO region and the covariance of the normalised time series with rainfall over the tropical Indo-Pacific region (i.e, the anomalies associated with one standard deviation of the time series) are shown in Fig. 7 for the pre-1980 period and in Fig. 8 for the post-1980 period. In both periods and for both datasets, the covariance pattern has a tri-polar structure with positive anomalies over the WCIO (by construction) and the central Pacific, and negative anomalies over most of the Maritime Continent (including the Philippines) and the south-western tropical Pacific. Both datasets indicate an increase in the overall amplitude of this signal after 1980, in particular over the central Pacific, and a westward extension of the negative anomaly over the Maritime Continent (note e.g. the negative anomaly south-west of Sumatra). Also, both datasets reveal a positive trend in the pre-1980 anomalies, although the correlation between the two time series is significantly lower in this period than in the satellite era. Finally, we note that the WCIO teleconnection looks very similar to the change in the mean rainfall between the pre-1980 and post-1980 periods detected in the merged-GPCP dataset, but shows a weaker resemblance to the ERA5 mean change (see again Fig. 6).

Fig. 9 and Fig. 10 show the corresponding time series and covariances for the extended NINO4 region. First of all, we note that the correlation between the merged-GPCP and the ERA5 time series remains very high even in the pre-1980 period, as suggested by the analysis of the leading modes in Sect. 3. Again, we note a post-1980 increase in the amplitude of the rainfall anomalies throughout the region, in a tri-polar pattern reminiscent of the typical ENSO response, and with a clear strengthening of the connected signal in the western Indian Ocean (indeed, the correlation between NINO4 and WCIO anomalies increased from 0.26 to 0.52 in the merged-GPCP dataset, from 0.34 to 0.65 in ERA5). Overall, the agreement between the two datasets suggests that a high level of confidence can be attributed to the tropical inter-decadal changes described above.

Having gained confidence in the tropical signals, we now explore what inter-decadal changes can be detected in the extra-tropical teleconnections of WCIO and NINO4 rainfall. This is done by computing covariances between the normalised rainfall time series from the two datasets with 500-hPa height from ERA5 over the northern extratropics. Again, we look at WCIO teleconnection in DJ and NINO4 teleconnections in JF, for the pre-1980 and post-1980 periods.

The extratropical WCIO teleconnections, from both the merged-GPCP and ERA5 filtered data and for both the selected periods, are shown in Fig. 11. There are significant differences between the teleconnections in the two periods, which show a very similar pattern whether we use the merged-GPCP or the ERA5 rainfall. In the pre-1980 period, we see a pattern with a predominant zonal wavenumber 3; two main negative centres are located over eastern Siberia and the north-west Atlantic, and a weaker negative one over the western USA, while the main positive centre is over northern Europe. Conversely, a strong zonal wavenumber 2 dominates the post-1980 teleconnection, with negative anomalies covering most of the northern Atlantic and Pacific Ocean. Over Europe, the positive anomaly has shifted to the south-west, with its centre just west of Portugal; this makes the North-Atlantic teleconnection very similar to a positive NAO pattern, as already noted in earlier studies (Molteni et al. 2015, 2020; Abid et al. 2023). Although the patterns obtained from the two rainfall datasets are very similar, the amplitude of the NAO signal is stronger in the merged-GPCP teleconnection (its rms amplitude in the area 70W-30E, 25N-80N is 18.2 m, against 13.3 m for ERA5 in the post-1980 period).

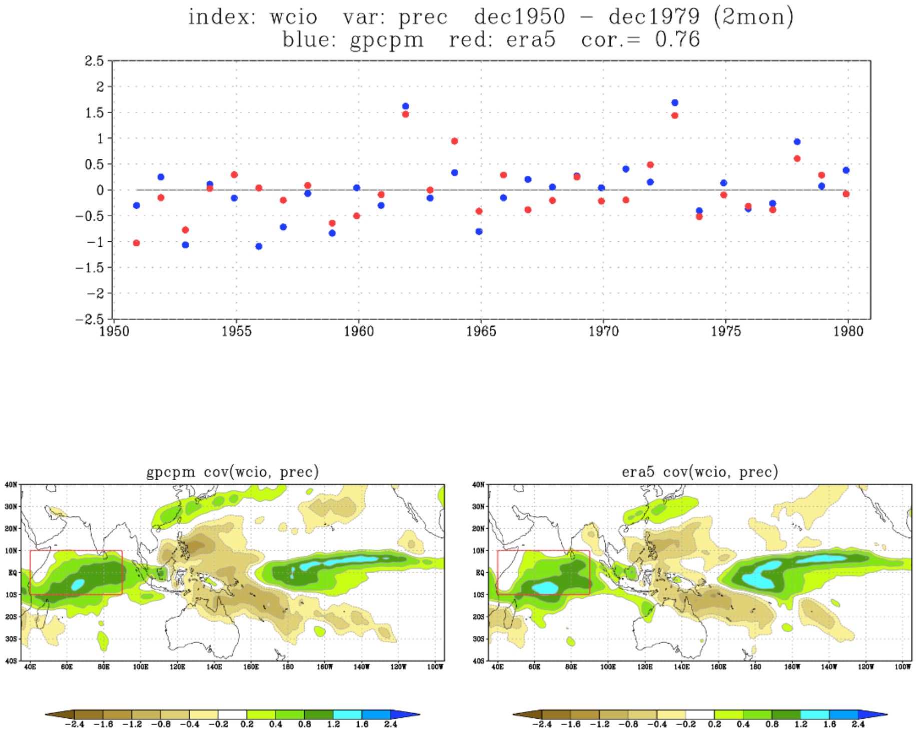

Fig. 7 *Top: time series of WCIO rainfall anomalies from the merged-GPCP (blue dots) and ERA5 (red dots) datasets, for DJ 1950/51 to 1979/80. Bottom: Covariance of normalised WCIO anomalies (red box) with rainfall across the Indo-Pacific region in the two datasets.*

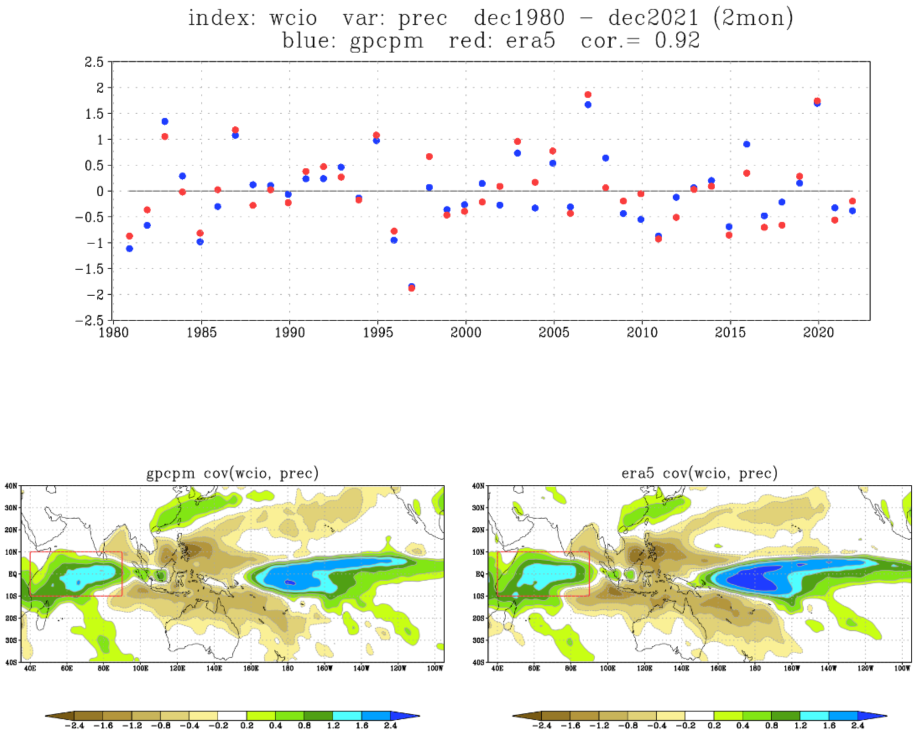

Fig. 8 *As in Fig. 7, for DJ 1980/81 to 2021/22.*

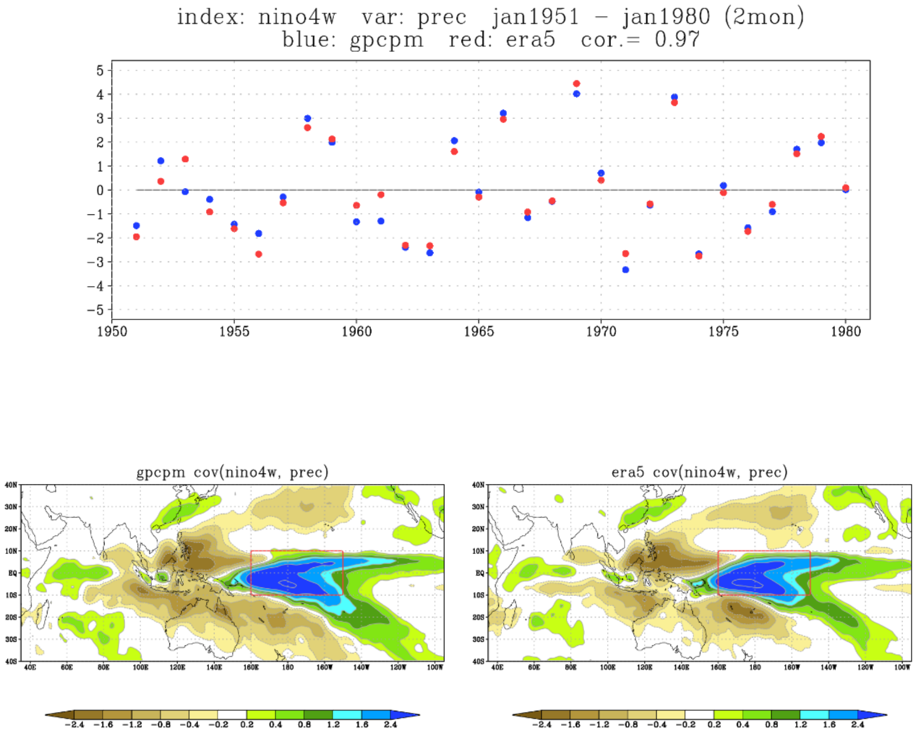

Fig. 9 *Top: time series of NINO4 rainfall anomalies from the merged-GPCP (blue dots) and ERA5 (red dots) datasets, for JF 1951 to 1980. Bottom: Covariance of normalised NINO4 anomalies (red box) with rainfall across the Indo-Pacific region in the two datasets.*

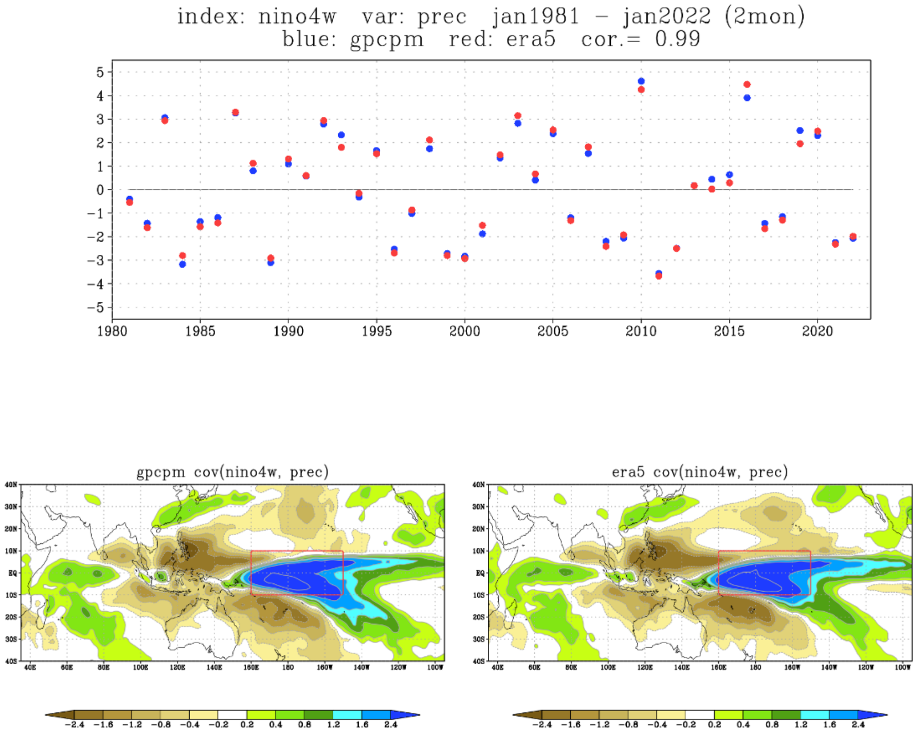

Fig. 10 *As in Fig. 9, for JF 1981 to 2022.*

The inter-decadal changes in the NINO4 teleconnections (in Fig. 12) are also significant, and coherently reproduced using both rainfall datasets. The strong negative anomaly covering most of the north-east Pacific in the satellite era is split into two separate centres in earlier decades, located close to Kamchatka and the western North American coast. A meridional dipole over the western North Atlantic, which projects on the negative phase of the NAO, is present in both periods, but its amplitude is much reduced in the post-1980 period. A possible explanation for this change is the increased correlation of NINO4 and WCIO rainfall discussed above; since the WCIO teleconnection shows a strong positive NAO signal, the stronger WCIO contribution during ENSO episodes in recent decades may partially cancel out the signal from the central Pacific. Indeed, the notion that teleconnections from the tropical Indian and Pacific Oceans may have a negative interference over the northern extratropics has been investigated in many earlier studies (Annamalai et al, 2007; Fletcher and Kushner 2011; Fletcher and Cassou 2015 among others).

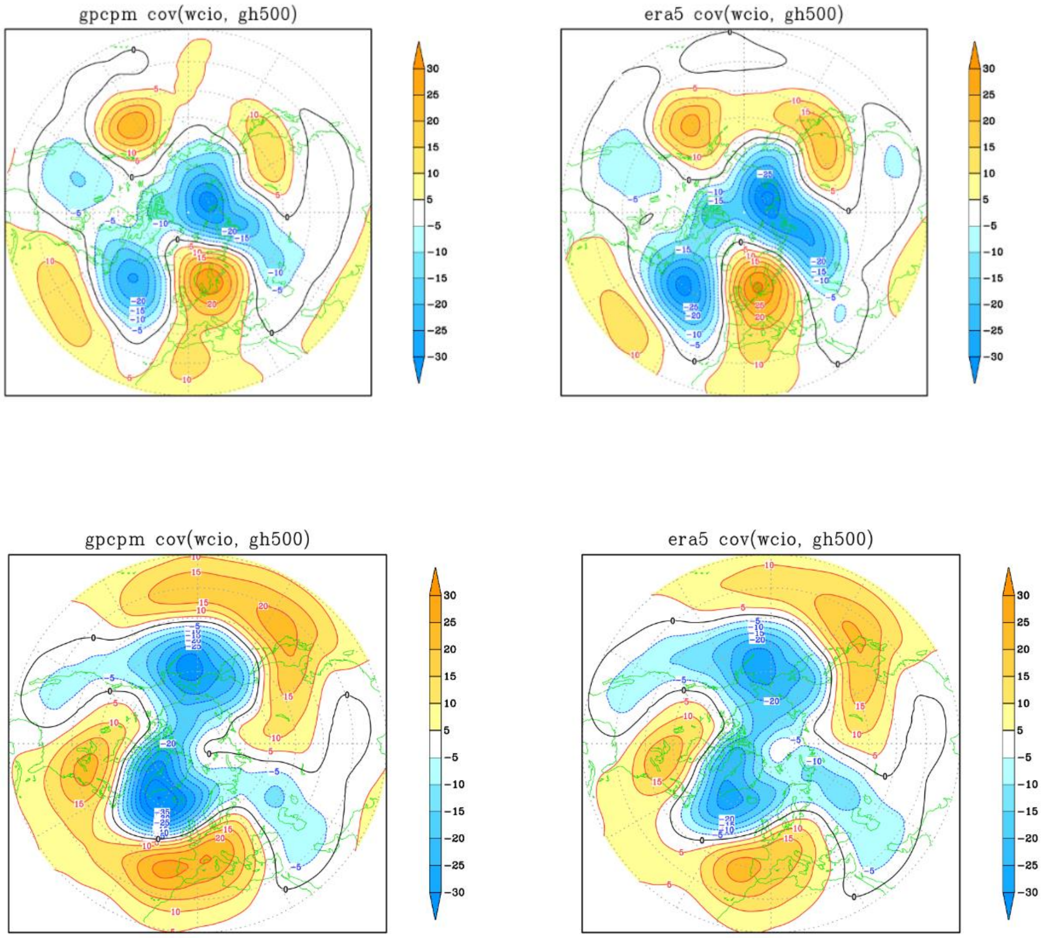

Fig. 11 *Extratropical teleconnections of the WCIO rainfall in DJ (from the CCA-filtered data), computed as covariance with ERA5 500-hPa height. Top-left: from merged-GPCP in DJ 1950/51 to 1979/80; top right: from ERA5 in DJ 1950/51 to 1979/80; bottom-left: from merged-GPCP in DJ 1980/81 to 2021/22; bottom right: from ERA5 in 1980/81 to 2021/22.*

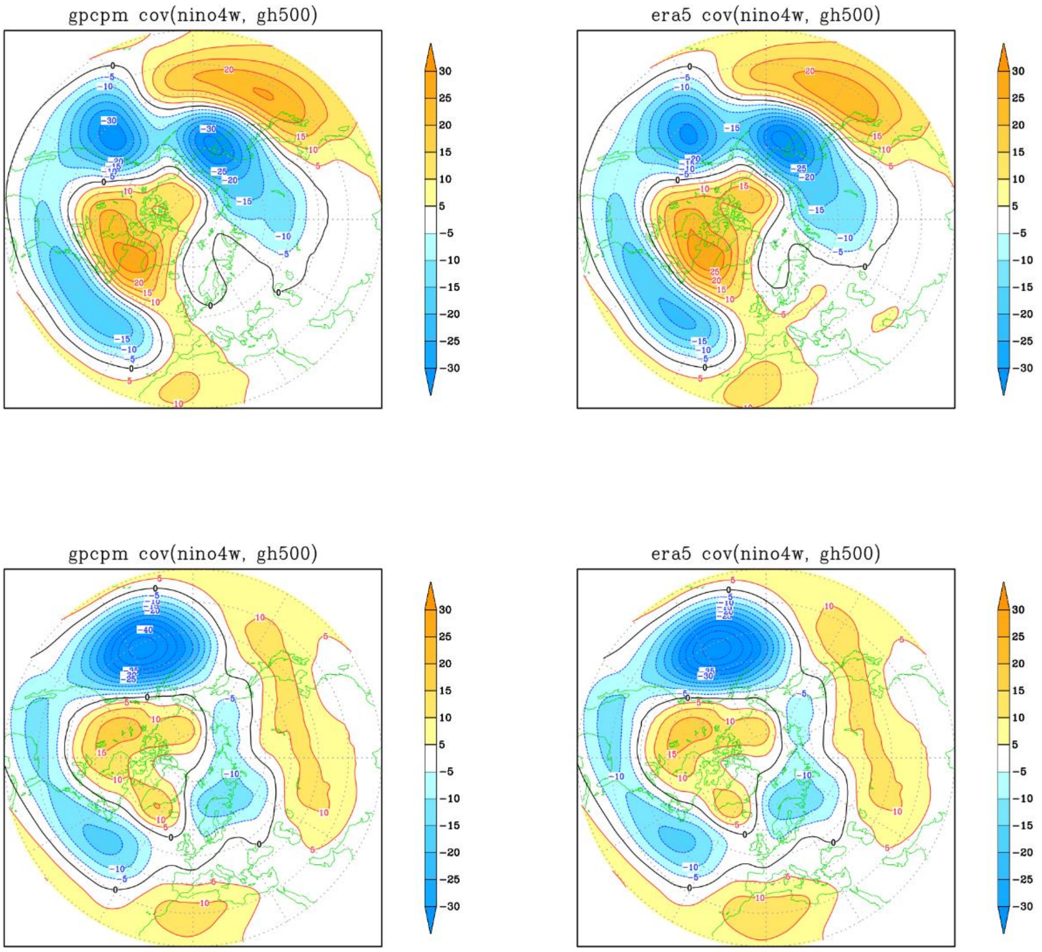

Fig. 12 *As in Fig. 11, for the extratropical teleconnections of the NINO4 rainfall in JF (from CCA-filtered merged-GPCP and ERA5 data).*

Understanding the causes of these teleconnection changes, however, is not the goal of this paper. Here, we simply want to provide some examples of how the robustness of some inter-decadal signals can be assessed by comparing results from the merged-GPCP and ERA5 data, after the CCA filtering has (hopefully) removed, or substantially weakened, any spurious variability related to changes in the observing systems.

5. Summary and discussion

In this study, we presented resulted results from a statistical procedure aimed at finding temporally coherent signals in tropical rainfall from different datasets. Specifically, we compared rainfall from the ERA5 re-analysis in the period 1950-2022 with a dataset obtained by merging GPCP v3.2 data for 1983-2022 with 1950-1982 rainfall values estimated by a regression of GPCP data against local and large-scale rainfall indices from the CERA20C re-analysis (as in Molteni et al. 2020).

For each sub-sample of 3 consecutive months, an EOF/PC analysis was first applied to data of both datasets in the 40S-40N domain; a canonical correlation analysis (CCA) was then performed on the first 21 PC of the two datasets, to find two sets of rotated PCs (one for the merged-GPCP and one for ERA5) which are orthogonal within each set and highly correlated with the corresponding RPC of the other dataset. The associated spatial patterns were computed taking the covariance of the RPCs with the original rainfall anomalies in each grid point. The first 16 RPCs, which showed a correlation exceeding 0.7, were selected and filtered anomalies were reconstructed from these RPCs and the associated spatial patterns (REOFs). The purpose of this procedure is to filter out variability components and trends which are poorly correlated across the two datasets, and which may therefore originate from changes in the assimilation methodology or in the availability of data sources across the 72-year period.

As any filtering procedure, ours has its own drawbacks. In order for the CCA analysis to produce meaningful results, the number of input PCs has to be at least one order of magnitude smaller than the number of time points. With monthly data in 3 months and 72 years, 21 is the maximum number of PCs which satisfies this criterion. Due to the substantial spatial variability of the rainfall field across the tropical band, 21 PCs only explain about 50% of the total variance, and areas of low rainfall tend to be represented less accurately than the main areas of tropical convection. As a consequence, the filtered anomalies provide a good estimate of rainfall variability in the ocean regions which act as the main sources of teleconnections, but are not able to correct the poor correlation between the merged-GPCP and the ERA5 data in some dry regions such as the Sahara and upwelling regions on the western side of Africa and South America.

The two leading RPCs in both datasets represent rainfall variability associated with ENSO, and are highly correlated between the two datasets even before the satellite era. Indian Ocean rainfall variability is well described by RPCs 3 to 5, but here differences can be found in some local features of the REOFs, and the correlation of the RPCs decreases in the pre-1980 period. This is also reflected in the inter-decadal changes between the pre-1980 and post-1980 periods as represented in the two datasets. In particular, ERA5 data tend to display a north-south shift of rainfall in the western part of the equatorial Indian Ocean, while in this region the merged-GPCP dataset shows rainfall variations with consistent sign on both sides of the Equator.

Using the rainfall anomalies reconstructed from the set of RPCs and REOFs, we have shown that changes in tropical and extratropical teleconnections from regions in the Indo-Pacific basin are consistently represented when estimated from the two datasets. However, some regional differences are worth noting, particularly with regard to the connection between Indian Ocean rainfall and NAO variability, which appears to be stronger when the merged-GPCP data are used.

In conclusion, our analysis does not completely remove the inconsistencies between rainfall datasets from different re-analyses and/or observational sources, but allows an estimate of the uncertainty associated with tropical rainfall variability and its teleconnection on inter-annual and inter-decadal scale. As a practical application, it provides a useful reference for the assessment of the fidelity of tropical variability and teleconnections as represented in seasonal forecasts and historical multi-decadal simulations. In particular, it allows an estimation of the uncertainty in decadal variability and multi-decadal trends of tropical precipitation, which is important for the assessment not only of CMIP-type historical simulations, but also for multi-decadal samples of seasonal reforecasts.

Acknowledgements

The authors are grateful to Chris Roberts and Tim Stockdale for their comments on the manuscript.

Appendix: additional statistics

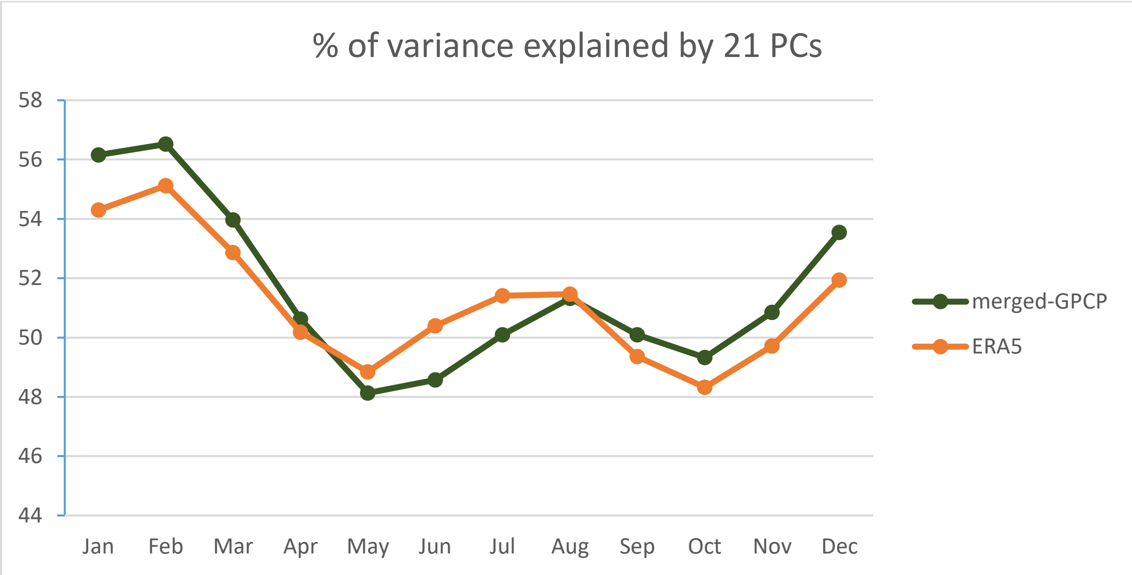

Fig A1 *Percentage of variance explained by the first 21 PCs in each month, for merged-GPCP (green) and ERA5 (orange) data.*

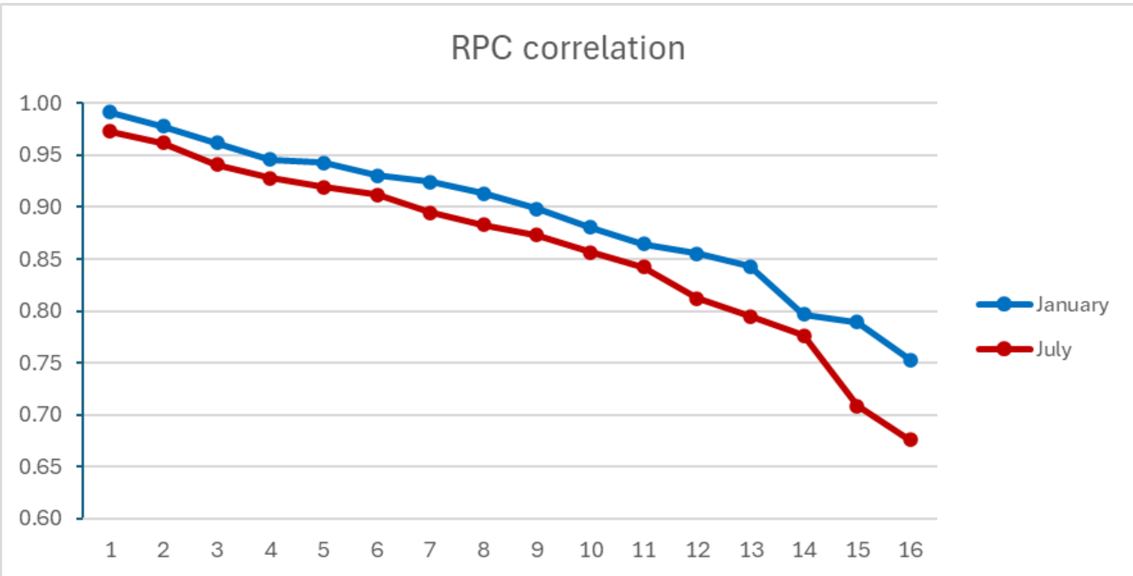

Fig A2 *Correlation of pairs of rotated PCs from the CCA of merged-GPCP and ERA5 data, for January (blue) and July (red) data.*

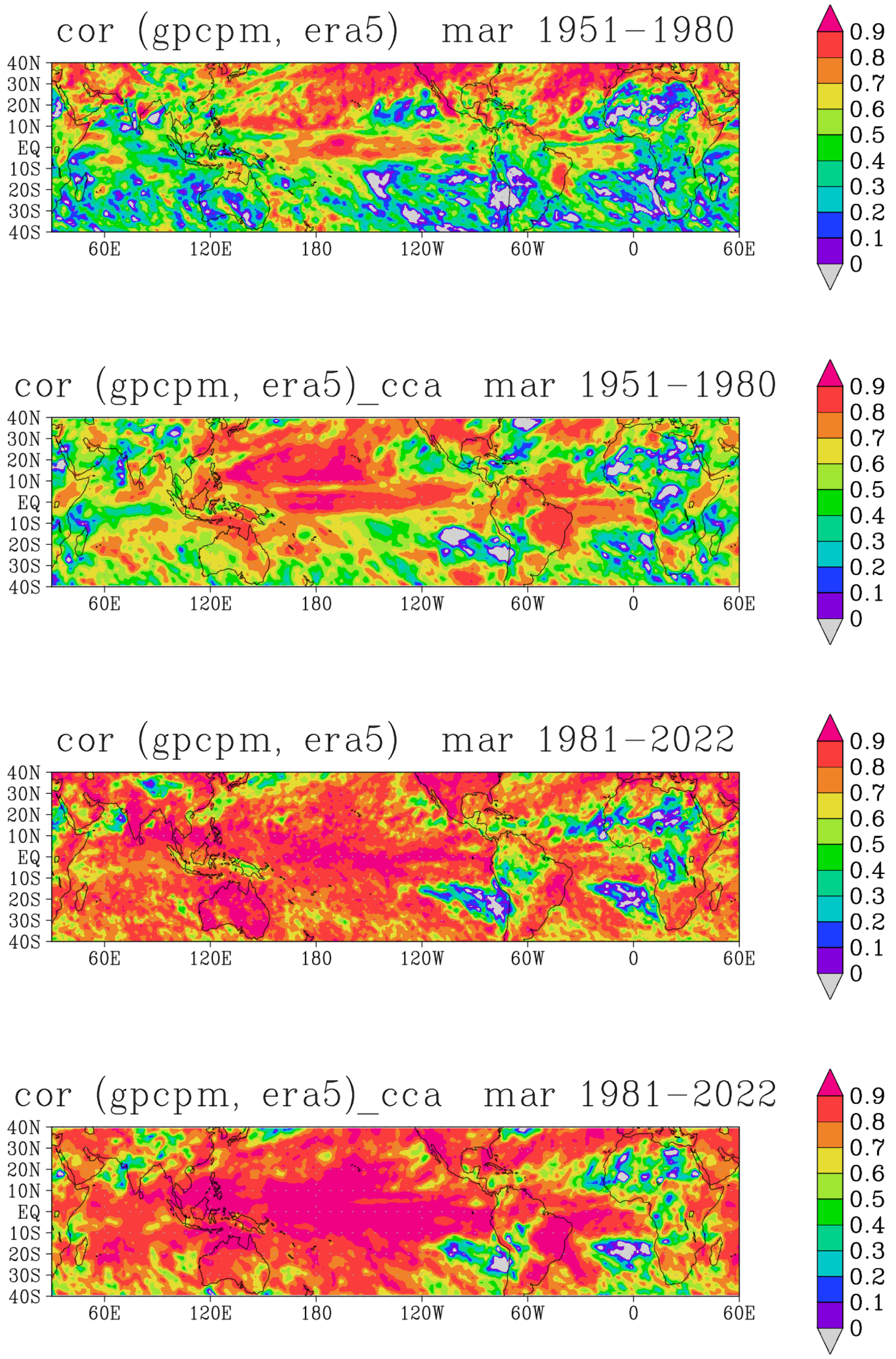

Fig. A3 *Correlation of unfiltered and CCA-filtered rainfall anomalies from the merged-GPCP and ERA5 datasets. From top to bottom: unfiltered data, March 1951-1980; CCA-filtered data, March 1951-1980; unfiltered data, March 1981-2022; CCA-filtered data, March 1981-2022.*

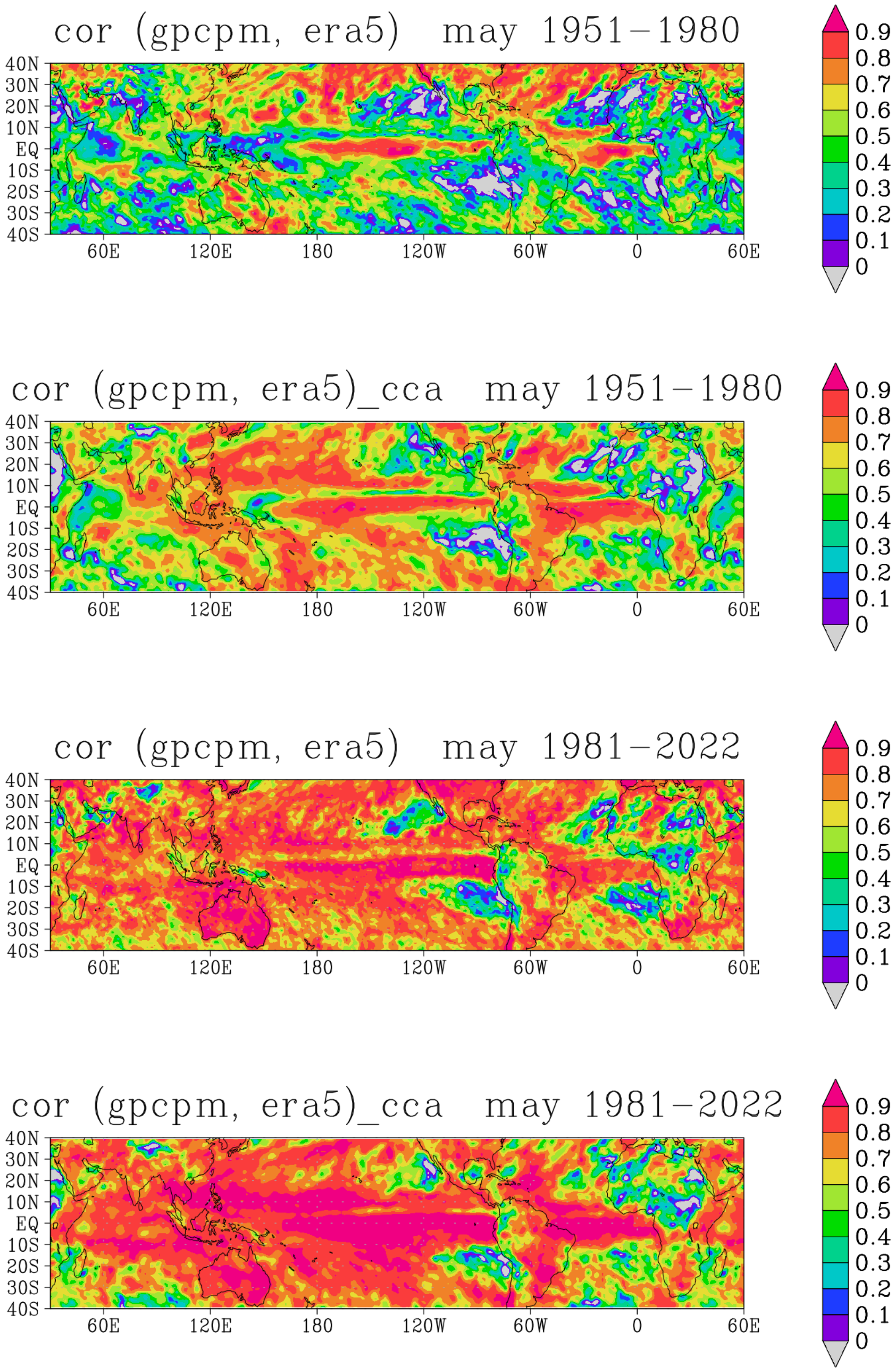

Fig. A4 *As in Fig. A3, but for data in May 1951-1980 and May 1981-2022.*

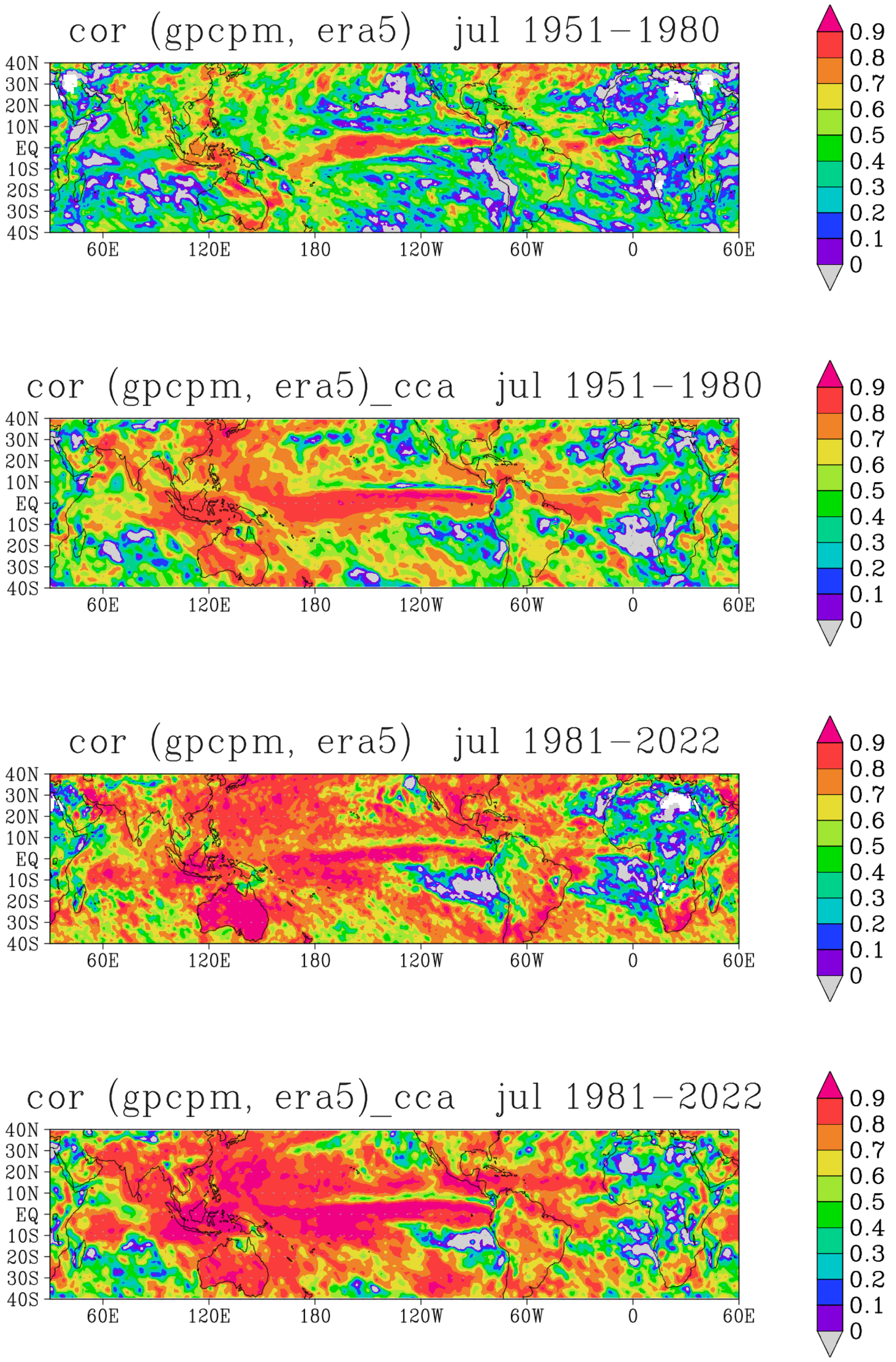

Fig. A5 *As in Fig. A3, but for data in July 1951-1980 and July 1981-2022.*

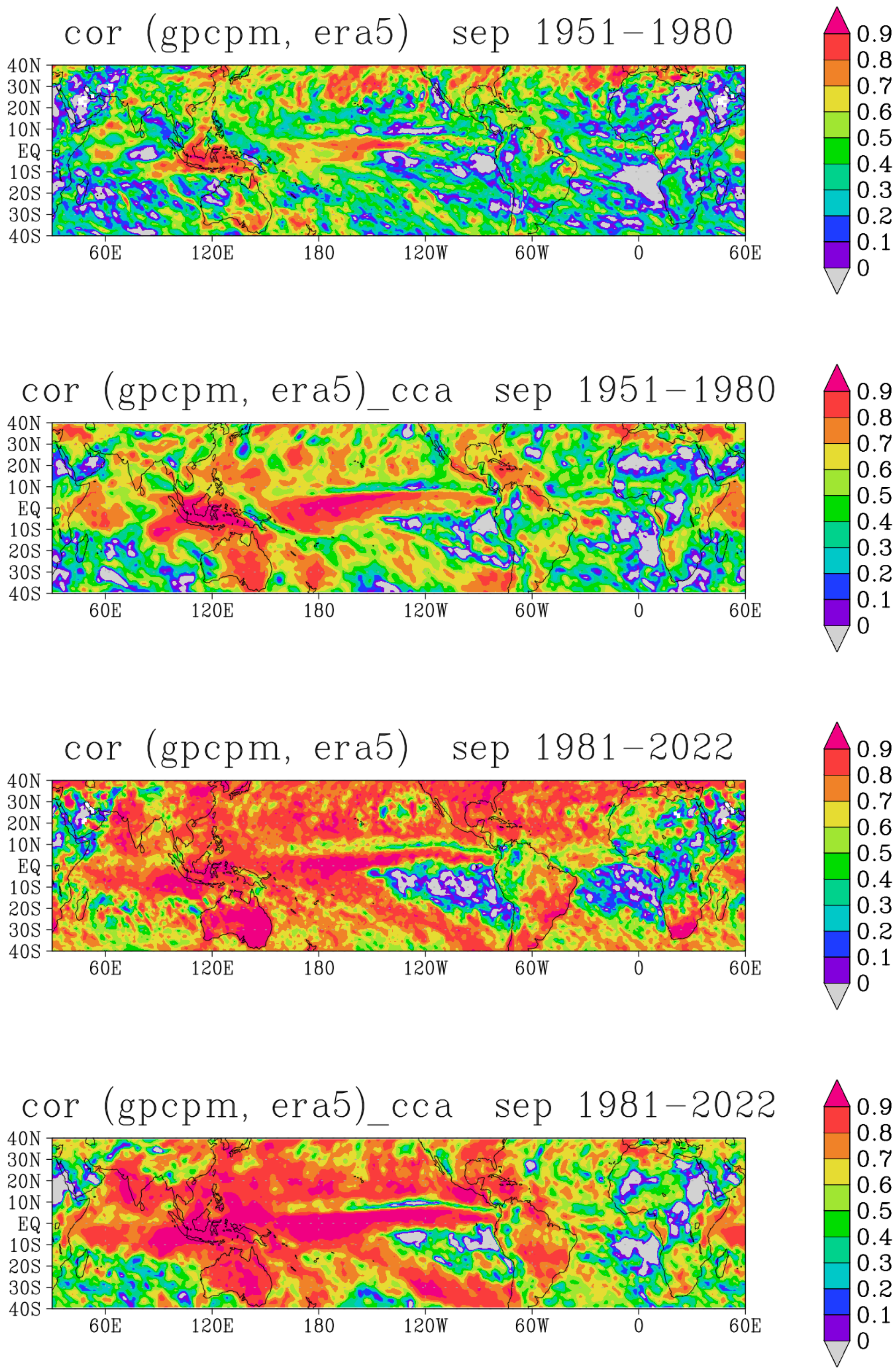

Fig. A6 *As in Fig. A3, but for data in September 1951-1980 and September 1981-2022.*

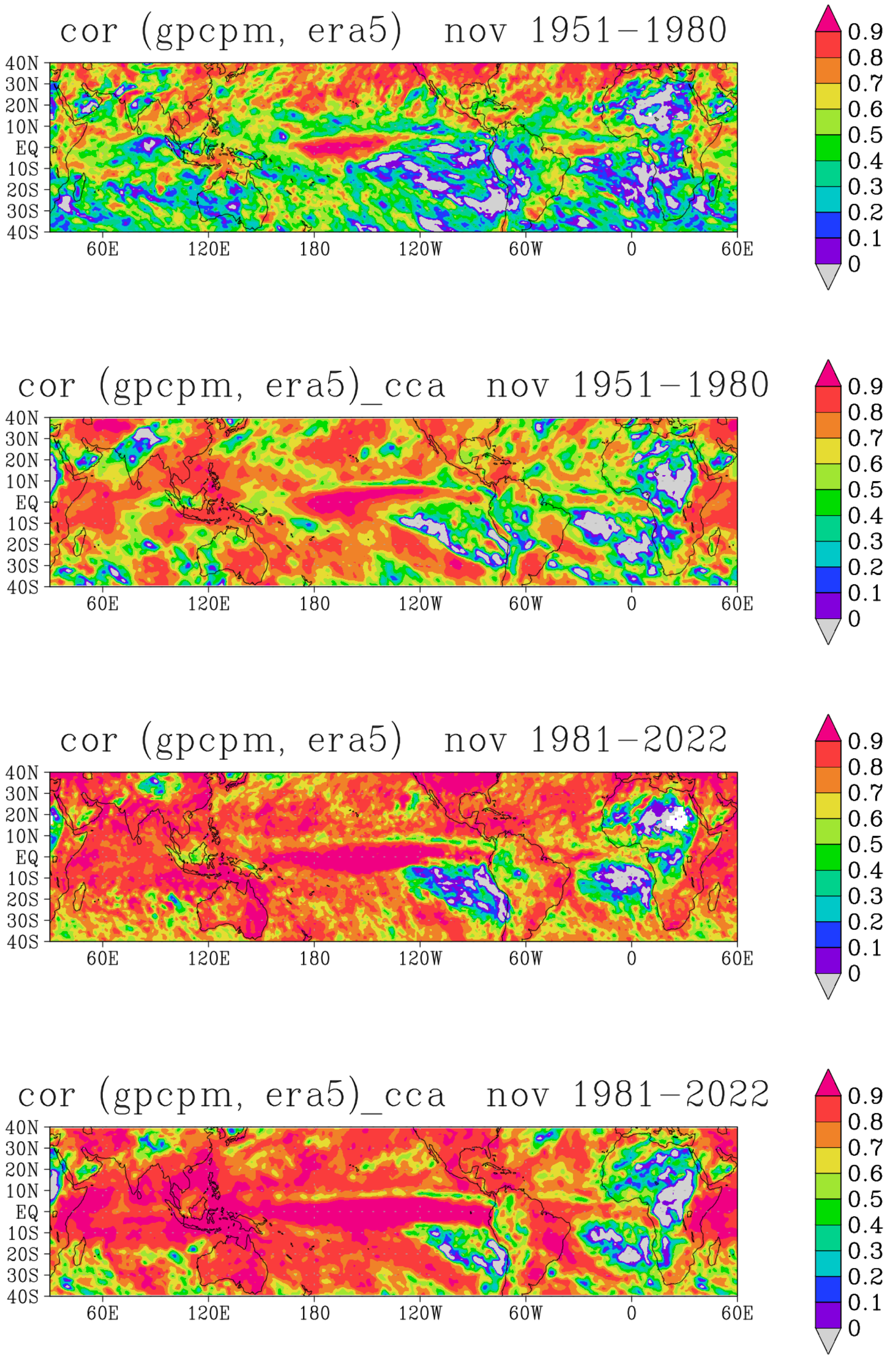

Fig. A7 *As in Fig. A3, but for data in November 1951-1980 and November 1981-2022.*

References

- Adler, R.F., and Coauthors (2003) The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). J. Hydrometeor., 4, 1147-1167, 2003.
- Compo, G. P., and Co-authors, 2011: The Twentieth Century Reanalysis Project. Quart. J. Roy. Meteor. Soc. 137, 1–28. https://doi.org/10.1002/qj.776.
- Deser, C., and A. S. Phillips, 2006: Simulation of the 1976/77 Climate Transition over the North Pacific: Sensitivity to Tropical Forcing. J. Clim., 19, 6170–6180.
- Fletcher, C. G. and Cassou C. (2015) The Dynamical Influence of Separate Teleconnections from the Pacific and Indian Oceans on the Northern Annular Mode. J. Clim. 28, 7985–8002.
- Fletcher, C. G. and Kushner P. J. (2011) The Role of Linear Interference in the Annular Mode Response to Tropical SST Forcing. J. Clim. 24, 778–794
- Hersbach H. and Coauthors (2020) The ERA5 global reanalysis. Q J R Meteorol Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803
- Hoerling M.P., Hurrell J.W., Xu T., Bates G.T., Phillips. A (2004) Twentieth century North Atlantic climate change. Part II: understanding the effect of Indian Ocean warming. Clim Dyn 23, 391–405
- Horel, J.D. and Wallace J.M. (1981) Planetary Scale Atmospheric Phenomena Associated with the Southern Oscillation. Monthly Weather Review, 109, 813-829.
- Hoskins, B.J. and Karoly, D.J. (1981) The Steady Linear Response of a Spherical Atmosphere to Thermal and Orographic Forcing. Journal of the Atmospheric Sciences, 38, 1179-1196.
- Huffman, G.J., A. Behrangi, D.T. Bolvin, E.J. Nelkin (2022) GPCP Version 3.2 Satellite-Gauge (SG) Combined Precipitation Data Set, Edited by Huffman, G.J., A. Behrangi, D.T. Bolvin, E.J. Nelkin, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center [Accessed: 03/12/2024]. https://doi.org/10.5067/MEASURES/GPCP/DATA304
- Hurrell J.W., Hoerling M.P., Philliphs A.S., Xu T. (2004) Twentieth century North Atlantic climate change. Part I: assessing determinism. Clim Dyn 23:371–389
- Johnson S.J., and Coauthors (2019) SEAS5: The new ECMWF seasonal forecast system. Geosci. Model Dev. 12: 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019
- Kalnay E and Coauthors (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77, 437–471.
- Kobayashi, S., and Coauthors (2015): The JRA-55 Reanalysis: General Specifications and Basic Characteristics. J. Meteor. Soc. Japan, 93, 5–48.
- Laloyaux, P., and Coauthors (2018). CERA-20C: A coupled reanalysis of the twentieth century. J. Adv. Mod. Earth Sys. 10, 1172–1195. https://doi.org/10.1029/2018MS001273
- Molteni F., Stockdale T. N., Vitart F. (2015) Understanding and modelling extra-tropical teleconnections with the Indo-Pacific region during the northern winter. Clim. Dyn. 45, 3119–3140. https://doi.org/10.1007/s00382-015-2528-y
- Molteni F., Roberts C.D., Senan R., Keeley S.P.E., Bellucci A., Corti S., Fuentes Franco R., Haarsma R., Levine X., Putrasahan D., Roberts M.J. and Terray L. (2020) Boreal-winter teleconnections with tropical Indo-Pacific rainfall in HighResMIP historical simulations from the PRIMAVERA project. Clim. Dyn. 55, 1843–1873. https://doi.org/10.1007/s00382-020-05358-4
- Poli, P., and Coauthors (2016) ERA-20C: An atmospheric reanalysis of the twentieth century. J. Clim. 29, 4083–4097.

- Sanchez-Gomez E, Cassou C, Hodson DLR, Keenlyside N, Okumura Y, Zhou T (2008) North Atlantic weather regimes response to Indian-western Pacific Ocean warming: a multi-model study. Geophys Res Lett 35, L15706. doi:10.1029/2008GL034345
- Slivinski L.C, and Coauthors (2021) An evaluation of the performance of the Twentieth Century Reanalysis Version 3. J. Clim. 34, 1417-1438. https://doi.org/10.1175/JCLI-D-20-0505.1
- Xie, P. and Arkin, P.A. (1997) Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. Bull. Amer. Meteorol. Soc., 78, 2539-2558.

Data availability

ERA5 : https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5

GPCP v3.2: https://disc.gsfc.nasa.gov/datasets/GPCPMON_3.2/summary?keywords=GPCPMON_3.2

The filtered ERA5 and merged-GPCP datasets reconstructed from the output of the CCA analysis are freely available from the authors upon request.