Geoscientific
Model Development

Multi-annual modes in the 20th century temperature variability in reanalyses and CMIP5 models

Heikki Järvinen[1], **Teija Seitola**[1,2], **Johan Silén**[2], **and Jouni Räisänen**[1]

[1]Department of Physics, University of Helsinki, Helsinki, Finland
[2]Finnish Meteorological Institute, Helsinki, Finland

Correspondence to: Heikki Järvinen (heikki.j.jarvinen@helsinki.fi)

Abstract. A performance expectation is that Earth system models simulate well the climate mean state and the climate variability. To test this expectation, we decompose two 20th century reanalysis data sets and 12 CMIP5 model simulations for the years 1901–2005 of the monthly mean near-surface air temperature using randomised multi-channel singular spectrum analysis (RMSSA). Due to the relatively short time span, we concentrate on the representation of multi-annual variability which the RMSSA method effectively captures as separate and mutually orthogonal spatio-temporal components. This decomposition is a unique way to separate statistically significant quasi-periodic oscillations from one another in high-dimensional data sets.

The main results are as follows. First, the total spectra for the two reanalysis data sets are remarkably similar in all timescales, except that the spectral power in ERA-20C is systematically slightly higher than in 20CR. Apart from the slow components related to multi-decadal periodicities, ENSO oscillations with approximately 3.5- and 5-year periods are the most prominent forms of variability in both reanalyses. In 20CR, these are relatively slightly more pronounced than in ERA-20C. Since about the 1970s, the amplitudes of the 3.5- and 5-year oscillations have increased, presumably due to some combination of forced climate change, intrinsic low-frequency climate variability, or change in global observing network. Second, none of the 12 coupled climate models closely reproduce all aspects of the reanalysis spectra, although some models represent many aspects well. For instance, the GFDL-ESM2M model has two nicely separated ENSO periods although they are relatively too prominent as compared with the reanalyses. There is an extensive Supple-

ment and YouTube videos to illustrate the multi-annual variability of the data sets.

1 Introduction

The ultimate goal in developing Earth system models (ESM) is to enable exploitation of the inherent Earth system predictability, and hence reduce weather- and climate-related uncertainties in our daily life, and guide societies in making sustainable choices (e.g. Slingo and Palmer, 2011; Meehl et al., 2014). For the predictions to be useful and usable, the expectation is that the climate mean state and climate variability are well simulated by these tools. Due to the complexity of the models and the data they produce, testing the expectation poses a challenge: many aspects of the model performance are gathered under the variability concept and no single diagnostic alone is sufficient to exhaust its all facets. Yet, understanding the discrepancies between the observed and simulated variability is crucial feedback for model development.

Representation of climate variability among models participating in climate model inter-comparisons, such as CMIP5, has been studied by, for example, Bellenger et al. (2014), Knutson et al. (2013), Ba et al. (2014) and Fredriksen and Rypdal (2016). We will add to this literature by interfacing a representative set of contemporary coupled climate models with reanalysis data focusing on spatio-temporal modes of climate variability. One century of global reanalysis data is of course a very short period for this purpose and severely constrains inter-comparison studies (e.g. Wittenberg, 2009; Stevenson et al., 2010). First, time series should cover a sufficient number of recurring "events" for

obtaining significance for the findings. Therefore, decadal-to-multi-decadal variability is of interest but not as informative as focusing on shorter cycles of variability. Second, the applied methods have to be very effective in extracting information from the short but high-dimensional data sets. For these reasons, we concentrate on the representation of multi-annual variability in reanalyses and coupled climate models applying randomised multi-channel singular spectrum analysis (RMSSA; Seitola et al., 2014, 2015) which effectively separates mutually orthogonal spatio-temporal components from our high-dimensional data sets.

The aim of this study is to decompose the 20th century climate variability into its multi-annual modes, and to assess how these modes are represented by the contemporary climate models. We hope for this to provide guidance for model development due better understanding of the deficiencies in representing reanalysed modes of multi-annual climate variability. Ultimately, interpretation of the hints about model deficiencies as development topics is due for the development teams themselves. Our role is to point towards the potential error sources. For reassuring the teams that high-dimensional time series analysis is possible today, we emphasise the methodological aspect of this study. RMSSA can, under very weak assumptions about the data, decompose high-dimensional data sets in a unique way and separate statistically significant quasi-periodic spatio-temporal oscillations from one another. This is in contrast to many other approaches which either make assumptions about the oscillation structures, such as Fourier or spherical decomposition, or resolve only either spatial or temporal aspects of variability. RMSSA can detect spatially evolving "chains of events" through resolving eigenmodes of spatio-temporal covariance data. This is a significant advantage over, say, PCA, which only resolves eigenmodes of spatial covariances and often projects temporal evolution of an "event" onto a number of different eigenmodes. In addition, the novel data compression based on random projections enables here a vast increase in tractable problem size (i.e. data dimension) – even multivariate decomposition is now possible, although not included here.

2　Methods and data

2.1　Randomised multi-channel singular spectrum analysis

Multi-channel singular spectrum analysis (MSSA; Broomhead and King, 1986a, b) can be characterised as being a time series analysis method for high-dimensional problems. It effectively identifies spatially and temporally coherent patterns of a data set by decomposing a lag-covariance data matrix into its eigenvectors and eigenvalues (e.g. Ghil et al., 2002) using singular value decomposition (SVD). The lag window in MSSA is a user choice, recommended typically

to be shorter than approximately one third of the length of the time series (Vautard and Ghil, 1989). A long lag window enhances the spectral resolution, i.e. the number of frequencies that can be identified, but distributes the variance on a larger set of components. Here, MSSA eigenvectors are called space–time empirical orthogonal functions (ST-EOFs), and the projections of the data set onto those ST-EOFs space–time principal components (ST-PCs). Because of the lag window, ST-PCs have a reduced length and they cannot be located into the same index space with the original time series. However, they can be represented in the original coordinate system by the reconstructed components (RC; Plaut and Vautard, 1994).

MSSA is computationally expensive and practical limits are easily exceeded for large data sets and long lag windows. In order to overcome this limitation, the computationally more efficient variant called RMSSA is applied here. The RMSSA algorithm, in a nutshell, (1) reduces the dimension of the original data set by using so-called random projections (RP; Bingham and Mannila, 2001; Achlioptas, 2003), (2) decomposes the data set by calculating standard MSSA in the low-dimensional space and (3) reconstructs the components in the original high-dimensional space.

In RP, the original data set is projected onto a matrix of Gaussian-distributed random numbers (zero mean and unit variance) in order to construct a lower dimensional representation. In this study, we reduce the data volume to about 5 % of the original volume. Since the computational complexity of RP is low, involving only a matrix multiplication, it can be applied to very high-dimensional data sets. Although RP is not a lossless compression, it has the important property that the lower-dimensional data set has essentially the same structure as the original high-dimensional data set. This has been demonstrated for climate model data in Seitola et al. (2014). The RMSSA algorithm is briefly presented in the Appendix A.

2.2　Computation of spectra

The ST-PCs represent the different oscillatory modes extracted from the data set. In order to estimate the dominant frequencies associated with each ST-PC, the power spectrum is calculated with the multitaper spectral analysis method (MTM) (Thomson, 1982; Mann and Lees, 1996). To further compare the variability modes and their intensities in different data sets, the power spectrum of all the ST-PCs of each data set is summed up to obtain so-called total spectrum. The ST-PCs are already weighted by their respective explanatory power, i.e. multiplied by the corresponding eigenvalue. Therefore the components with more explanatory power also have higher spectral densities compared to the ones that explain only a small fraction of the variance. Therefore no extra weighting is needed in this step.

The uncertainty related to the explanatory power of each ST-PC (i.e. the confidence interval of the respective eigen-

value) is estimated using North's rule of thumb for sampling errors (North et al., 1982). The sampling error (e_k) is given by $e_k \sim \lambda_k(2/N)$, where λ_k is the eigenvalue associated with the kth ST-PC and N is the length of the time series. Thus, the confidence interval of the total spectrum describes the uncertainties related to the explanatory power of each ST-PC.

2.3 Statistical significance testing

In data sets of dynamical systems, ST-PCs/ST-EOFs of MSSA often appear as quadratic pairs that explain approximately the same variance and are $\pi/2$ out of phase with each other. However, the existence of such a pair does not guarantee any physical oscillation in the data set, and it may be due to some non-oscillatory processes, such as first-order autoregressive noise. Allen and Robertson (1996) formulated a test, where the oscillatory modes identified with MSSA are tested against a red noise null hypothesis through Monte Carlo simulation.

Significance testing in MSSA requires solving conventional PCs of the original data set. In case of very high-dimensional problems this easily exceeds practical computational limits. The RMSSA implementation in Seitola et al. (2015) contains the Allen-Robertson test such that the PCs are solved in the dimension-reduced space, and is thus affordable even in very high-dimensional problems. The Appendix A also includes a short description of the significance test.

2.4 Data sources

The data consist of the monthly mean near-surface air temperature from the historical 20th Century simulations of 12 different climate models (Table 1). The selected models originate from different modelling centres, and thus do not have close common-ancestor models. Furthermore, the selected models have undergone a long (generally several generations) history of development, suggesting that the chosen models collectively represent the state of the art. Near-surface temperature was chosen because many processes must be adequately represented in coupled models to realistically capture the observed temperature distribution (IPCC, 2013). These include processes in the Earth system component models (atmosphere, ocean, etc.) as well as in their mutual coupling models. Also, for the near-surface temperature, there are corresponding reanalysis data available.

The historical (1901–2005) simulations were extracted from the CMIP5 data archive and they follow the CMIP5 experimental protocol (Taylor et al., 2012). The 20th Century simulations use the historical record of climate forcing factors such as greenhouse gases, aerosols, solar variability, and volcanic eruptions. We used a single ensemble member of each model and the model data sets were interpolated into a common grid of 144×73 points.

Table 1. Climate models used in the study. For more details of the models, see Table 9.1. in IPCC (2013).

Model ID	Model name	Modelling centre	Country
a	CanESM2	CCCMA	Canada
b	CESM1(CAM5)	NSF-DOE-NCAR	USA
c	CNRM-CM5-2	CNRM-CERFACS	France
d	CSIRO-Mk3.6.0	CSIRO-QCCCE	Australia
e	GFDL-ESM2M	NOAA GFDL	USA
f	GISS-E2-R	NASA GISS	USA
g	HadGEM2-ES	MOHC	UK
h	INM-CM4	INM	Russia
i	IPSL-CM5B-LR	IPSL	France
j	MIROC-ESM	JAMSTEC/AORI/NIES	Japan
k	MPI-ESM-MR	MPI-M	Germany
l	MRI-CGCM3	MRI/JMA	Japan

As a reference, we used two reanalysis data sets: the 20th Century Reanalysis V2 data (hereafter 20CR) provided by the NOAA/OAR/ESRL PSD (Compo et al., 2011), and ERA-20C data provided by ECMWF (Poli et al., 2013). The data sets are produced using an ensemble of perturbed reanalyses, and the final data set corresponds to the ensemble mean. In 20CR, only surface pressure observations are assimilated, and the observed monthly sea-surface temperature and sea-ice distributions from HadISST1.1 (Rayner et al., 2003) are used as boundary conditions (Compo et al., 2011). In ERA-20C, observations of surface pressure and surface marine winds are assimilated (Poli et al., 2013). Unlike 20CR, it uses a more recent sea-surface temperature and sea-ice cover analysis from HadISST2 (Rayner et al., 2006). Both 20CR and ERA-20C are forced by historical record of changes in climate forcing factors (greenhouse gases, volcanic aerosols and solar variations). In order to be consistent with the climate model simulations, the same time period is used (1901–2005, i.e. 1260 monthly mean fields) and the reanalysis data sets were interpolated into the same grid as the model simulations (144×73 points).

2.5 Data processing

Some pre-processing of the data was needed before applying RMSSA. At each grid point the data sets were processed as follows:

- linear trend was fitted and removed,

- annual cycle was estimated using seasonal-trend decomposition (STL; Cleveland et al., 1990) and removed,

- resulting values were mean-centred and divided by the average standard deviation of all the data sets (see Fig. 1). Average standard deviation was obtained after removal of the trend and the annual cycle.

The reanalysis and climate model data sets have different temperature standard deviations, which would impact the
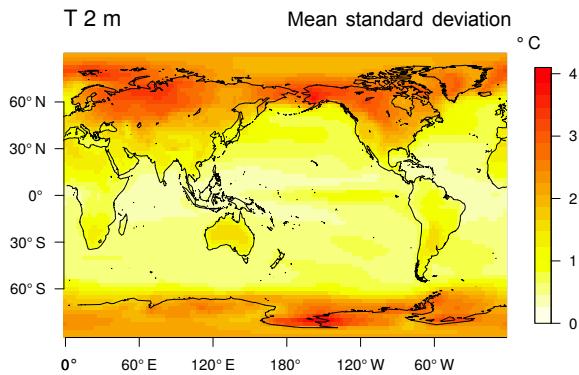
Figure 1. Map of the common normalisation factor. Shown is the mean standard deviation of 2 m temperature (°C) across all the data sets.

temperature variability from inter-annual to multi-decadal timescales (e.g. Thompson et al., 2015). To retain these differences, we have used a common normalisation factor (i.e. the average standard deviation of all the data sets). This procedure reduces the weight of grid points with high variance, typically at higher latitudes, and hence adds weight on the lower latitude features. After the pre-processing, the dimension reduction step of RMSSA was applied so that approximately 5 % of the original data dimensions were retained. The lag window in the analysis was 20 years (240 months). The total spectra were obtained from this analysis, and are comparable due to normalisation using the common standard deviation of the data sets.

The statistical significance test uses a red noise null hypothesis. In the test we have used data sets that are normalised by their own standard deviations. Using a common normalisation interferes with generating the red noise surrogates corresponding to each data set. The first 50 PCs of each data set were retained as input. Those PCs explain 79 % of the variability in 20CR, 75 % in ERA-20C, and 70–80 % in the climate model data sets. A total of 1000 realisations of red noise surrogate data sets were generated, and confidence interval (95 %) for the oscillatory modes were estimated. We note that transformation to PCs may interfere with the detection of weak signals, as demonstrated by Groth and Ghil (2015).

2.6 Data visualisation

We used reconstructed components (RC; see Appendix A) for visualisation of the spatial patterns related to ST-PCs. For each grid point time series, we can calculate the RCs corresponding to the ST-PCs (or modes) of interest. These RC values, reflecting the contribution of each grid point to the mode, can be plotted on a map at each time step. We have used these maps to construct videos of the spatio-temporal modes. In Sect. 3.5, we have analysed RCs corresponding to 3–4 year variability. The result is a time series of the data correspond-

ing to the 3–4 year mode in each grid point and according to its variance after detrending and removing the annual cycle. In the analysis we have neglected 5 years in the beginning and the end of the time series, because the reconstruction procedure may be biased there (see the Appendix, Eq. A4). The videos can be found at our YouTube channel (https://www.youtube.com/channel/UCu1zJdwJfLaXvfvTqsKCLHw).

To summarise the animations, we have calculated composite maps of the modes. The compositing procedure follows the one described in Plaut and Vautard (1994). The idea is to choose the grid point time series (RC_1) for which the variance is largest, and calculate its time derivative (RC_1'). The phase of the mode at each time step is determined by calculating the angle between the vector (RC_1, RC_1') and the vector (0, 1). These phases, in the interval ($0, 2\pi$), are then classified into eight equally populated categories. Composite maps are constructed from these categories.

3 Results

3.1 Reanalysis decompositions

The main outcome of the RMSSA method, the space–time principal components (ST-PCs) characterise both the spatial and temporal structure of the modes of variability. Sections 3.1–3.4 focus on their temporal aspects. The leading 30 ST-PC time series and the corresponding power spectra are displayed in Fig. 2 for 20CR and ERA-20C, ordered according to the explained variance. We can see that

- components with predominantly multi-decadal periodicity (1, 2, 5, and 6) explain a total of 7.2 and 5.9 % of the variance in 20CR and ERA-20C, respectively, with clear similarities in their time series and spectra;

- multi-annual components (3, 4, 7, and 8) explain 4.2 and 3.2 % of the variance in 20CR and ERA-20C, respectively;

- there is a broad multi-annual peak centred at 5 years and a narrower peak at 3.5 years in both reanalyses; these are clearly separated in ERA-20C at the components 3 and 4 vs. 7 and 8. This separation in 20CR is less clear;

- there are many spectral peaks in the reanalyses at 2–3 year periods with little explained variance but some are well separated and distinct.

The conclusion based on Fig. 2 is that the leading sources of the near-surface air temperature variability at multi-decadal and multi-annual periods are well identifiable in the reanalysis data sets. 20CR and ERA-20C are composed of very similar components explaining the variance in the two data sets. This is of course expected but it is also reassuring from the methodological view point: despite its complexity, the RMSSA decomposition is consistent.
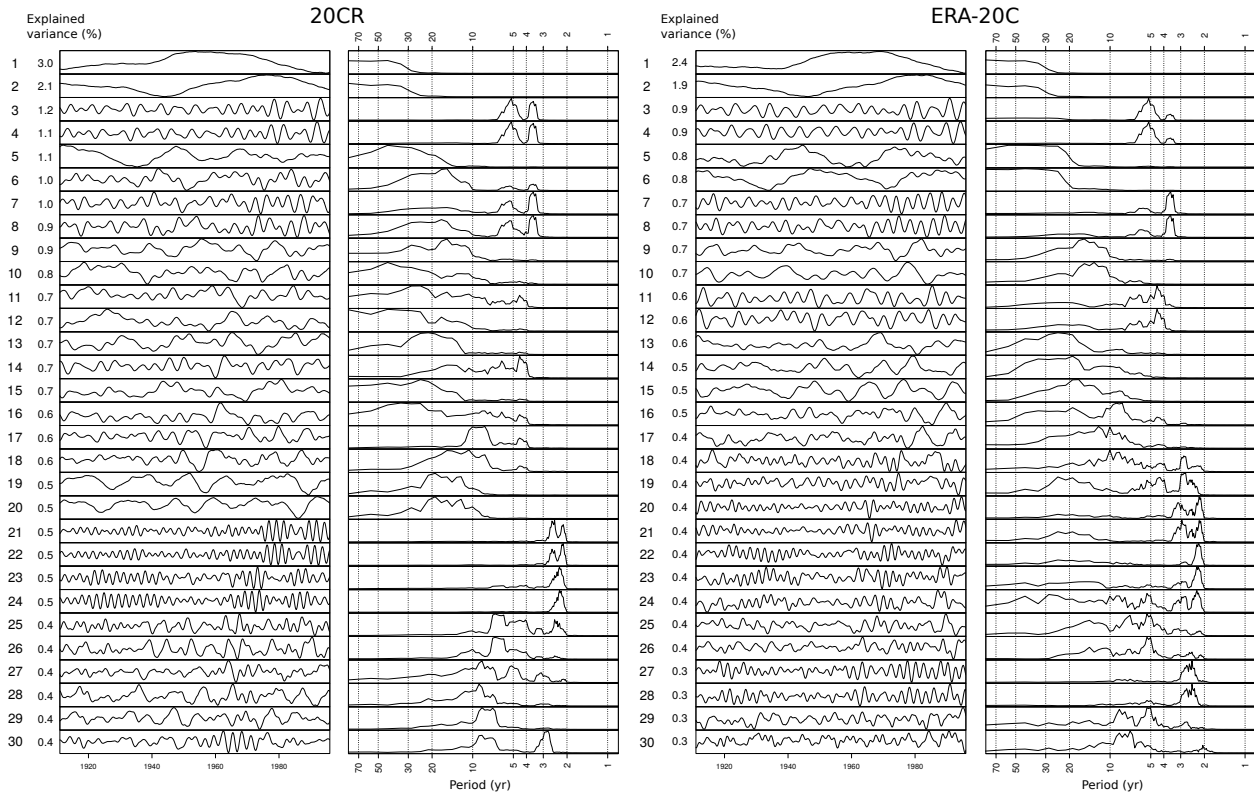
Figure 2. Reanalysis ST-PC time series (columns 1 and 3) of monthly near-surface temperature 1901–2005 and their spectra (columns 2 and 4) for 20CR and ERA-20C. The components are ordered according to the explained variance (%).

It is noteworthy in Fig. 2 that the components 3, 4, 7 and 8 in both reanalyses have become more prominent with time. Since about the 1970's, the amplitudes of these 3.5- and 5-year oscillations have been at a higher level, presumably due to some combination of forced climate change, intrinsic low-frequency climate variability, or changes in global observing networks (the rather sudden increase in the amplitude seems to coincide with the onset of the modern era of satellite observations). This finding seems to be in support of e.g. Russell and Gnanadesikan (2014). In this connection it should be noted, however, that apparent low-frequency variations and changes in amplitude may simply arise from random fluctuations of the time series (Wunsch, 1999; Wittenberg, 2009). Back-projection of these components into the original grid representation (Fig. 3), reveals that the components are indeed associated with the ENSO phenomenon and are geographically similar in 20CR and ERA-20C. In the snapshots from January 1987 and January 1998 (Fig. 3), there is a typical El Niño pattern with positive anomalies in the equatorial Pacific Ocean, South America, and northwestern North America. These are associated with synchronous evolution of (i) a dipole structure in the western Antarctica with easterly motion, and (ii) a wave-train type pattern in the northernmost North America with northeasterly motion. The components 3, 4, 7 and 8

thus represent a global phenomenon, with an increased amplitude in recent decades. These features are nicely depicted on our YouTube channel (https://www.youtube.com/watch?v=vehbT8fOHeM, https://www.youtube.com/watch?v=xG–SiUqqAI).

3.2 Reanalysis total spectra

Figure 4a shows the total spectrum for the reanalyses constructed from the ST-PCs, and their confidence intervals (dashed lines). As in the ST-PCs, there is most power in the slow modes. At periods of about 3.5- and 5-years, there are the spectral peaks of the components 3, 4, 7 and 8. The dip at 1 year reflects the removed annual cycle.

As Fig. 2 already suggests, the shape of the two spectra is remarkably similar in all timescales (Fig. 4a). This leaves hardly any doubt that the data assimilation systems of 20CR and ERA-20C extract observed information in a very similar manner. There are some differences, however. The spectral power in ERA-20C is systematically slightly higher than in 20CR. This difference is statistically significant at almost all timescales. This is most likely due to generally higher temperature variance in ERA-20C compared to 20CR, especially in the Southern Ocean and Arctic Ocean. Also, in 20CR, the 3.5- and 5-year spectral peaks are relatively more pronounced than in ERA-20C.
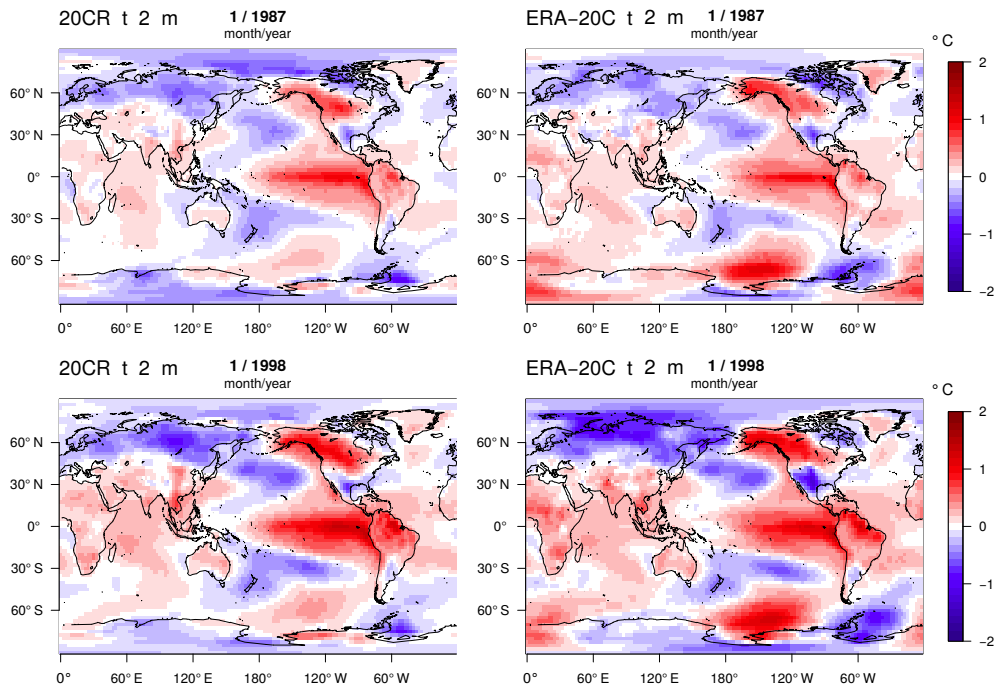
Figure 3. Global patterns of 2 m temperature for the components 3, 4, 7 and 8 in 20CR (left column) and ERA-20C (right column). Snapshots are taken from January 1987 (top row) and January 1998 (bottom row). Unit °C.

Statistical significance tests are presented in Figs. 4b and c for 20CR and ERA-20C, respectively. The multi-annual periods (less than 7 years) rising above the 95 % confidence interval (i.e. the red dots above the region covered by the vertical bars) are 3.5, 3.6 and 5.7 years in 20CR and 3.6, 5.2, 5.5 and 5.7 years in ERA-20C. Thus, nearly the same periodicities rise above the red noise in the two data sets. It is logical that the frequency corresponding to the annual cycle is present in the red noise surrogates while it is absent from the data, and therefore the red dots fall far below their expected values. Interestingly, the period of 2.9 years in 20CR and ERA-20C falls below the 95 % confidence interval. Our conclusion is therefore that the multi-annual climate variability in the near-surface air temperature is very similar in 20CR and ERA-20C.

3.3 CMIP5 model total spectra

The total spectra for the 12 CMIP5 models are shown in Fig. 5 (solid lines) with their 95 % confidence intervals (dashed envelopes) and the reanalysis spectra as a reference (thin lines). Statistically significant multi-annual modes (at 5 % level) are denoted by vertical dashed lines. As in the case of reanalyses, these spectra are unique expressions of the low-frequency variability present in the simulation data. A comparison between the simulated and the reanalysis spectra provides one means to assess the strengths and weaknesses of these models. However, one cannot simply rank the models based on how "far off" the model spectra are from the refer-

ence, because this comparison focuses on just one (although important) aspect of model performance and because seemingly good agreement with observations might occasionally result from compensating errors in model processes.

Here we will concentrate on the multi-annual aspects but note in passing that the level of multi-decadal variability (> 20 years) is close to reanalyses in models a, c, d, e, and g. In the rest of the models, the level seems too low. In the decadal scale ($\sim 10\ldots20$ years), the level of variance is close to reanalyses in a, b, c, f, i, j and l. Subjectively, the shape of the low-frequency end of the spectra appears most realistic in models a and c.

In multi-annual scales, the model performance varies a lot among the models. There is a group of models (a, b, d and e) with high spectral density at about 3–7 year periods. The models d and e have a bi-modal spectral structure, as in the reanalyses, while models a and b have a broad unimodal peak. Decompositions (available in the Supplement, Sect. S1) partly explain the reasons leading to these total spectra.

In model a, for instance, there is a unimodal broad peak at 3.5–4 year periods (Fig. 5a). The decomposition reveals that there are, in fact, two well separated component pairs at 3.5 and 4 years generating one merged peak to the total spectrum (Fig. S1a in the Supplement). A development hint is thus to investigate these modes which can help to better understand some underlying modelling deficiencies, and to keep monitoring how this aspect of model performance evolves in the future model upgrades. An additional concern in model a is
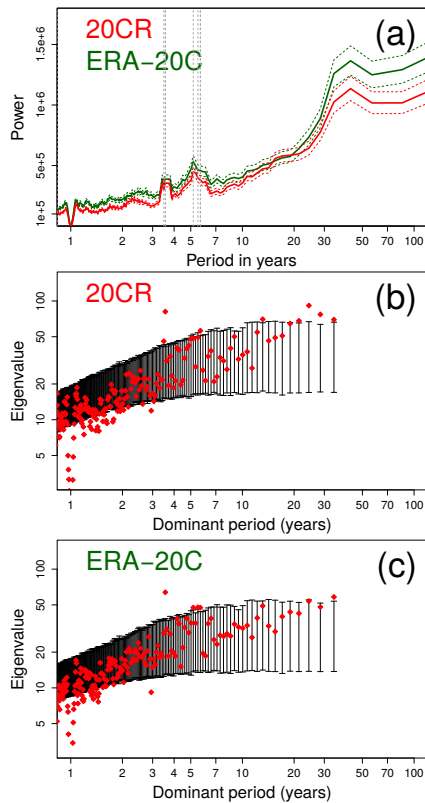
Figure 4. (a) Total spectrum of 20CR (red line) and ERA-20C (green line) with their min.–max. confidence intervals. The unit of the spectral density is arbitrary. **(b)** Significance of the 20CR periodicities against red noise null hypothesis. Shown are the data eigenvalues (red squares) and the 2.5th and 97.5th percentiles of the eigenvalue distribution of the red noise surrogates (vertical bars). **(c)** Same as **(b)**, but for the ERA-20C data set.

the excessive spectral density at about 2 and 7–10 year periods.

In model e, there is a bimodal total spectrum (Fig. 5e), although far too pronounced as compared with the reanalyses. The decomposition (Fig. S1e in the Supplement) reveals that the ST-PC components 1–10 (except 7–8) are all multi-annual and peak strongly and well in isolation at 3, 3.5, 4 and 5 years, explaining together no less than 13.9 % of the total variance. The development hint for model e is thus to investigate the mechanisms behind the components 1–10 and thereby obtain guidance for improving the realism of simulations.

In most other models, the multi-annual variability is less prominent than in the reanalyses. In model c (Fig. 5c), on the one hand, the decomposition points out (Fig. S1c in the Supplement) that there are about 12 ST-PC components with periods between 1.5 and 3 years leading to a total spectrum with a broad peak of 2–3 year periods. These components tend to have very regular cycles, remotely resembling a coupled harmonic oscillator and seemingly missing the "offbeats" or

true quasi-periodicity of the reanalyses. The task seems to be to find out reasons why model c produces too-rapid and regular multi-annual variability. In model g (Fig. 5g), on the other hand, the leading ST-PC components 1–9 are on either decadal or multi-decadal periods and these overwhelm the total spectrum. It should be important to find out the causes for this accentuated variability, especially on the decadal scale.

Finally, Fig. 5 casts light on models' overall level of variability compared to reanalyses. Clearly, this level in model h (Fig. 5h) is low. Curiously enough, the leading ST-PC component pair in model h explains only 1.4 % of variance and peaks at 3.2 year. This corresponds to the isolated peak in the total spectrum.

3.4 Significance of multi-annual modes in CMIP5 models

In the reanalyses (Fig. 4), only a few multi-annual periods rise above the red noise (three in 20CR and four in ERA-20C). They are at approximately 3.5- and 5-year periods. For the CMIP5 models, the test results are available in the Supplement (Sect. S2). In Fig. 5, the multi-annual modes with periods less than 7 years at the 5 % significance level are denoted by dashed vertical lines.

In summary, there are 5–15 statistically significant periods in the models, except for model k (Fig. 5k) with three and model g (Fig. 5g) with zero periods. The large number of significant periods (models d and e, for instance) can be explained, at least partly, by the fact that the modes are quasi-periodic and the spectral density therefore appears on a range of frequencies. This manifests as an excursion of the red noise threshold on several adjacent frequencies. This is typical for models with large spectral power on certain timescales. In model l (Fig. 5l), for instance, there are two broad and distinct spectral peaks at about 3.5 and 6 year periods, and many significant periods are gathered at these and nearby frequencies. In contrast, models f and h (and to some extent model c) have several significant and distinct periods between 2 and 7 years. In terms of number of significant modes, models a, i, j and k seem to be closest to the reanalyses.

3.5 Spatial patterns of the 3–4 year mode

ST-PC components can be represented in the original coordinate system as so-called reconstructed components that can be visualised. In this section, some visualisation results are presented and discussed.

In ERA-20C, there is a spectral peak at the 3.5-year period, which is significant at 5 % level (Fig. 4). This peak is due to the ST-PC components 7 and 8 with spectral density closely concentrated on this frequency (Fig. 2). Figure 6 depicts composite maps of each of the eight phases of the 3.5-year mode in ERA-20C. Firstly, the mode is global with the largest temperature anomalies in the Pacific and North Amer-
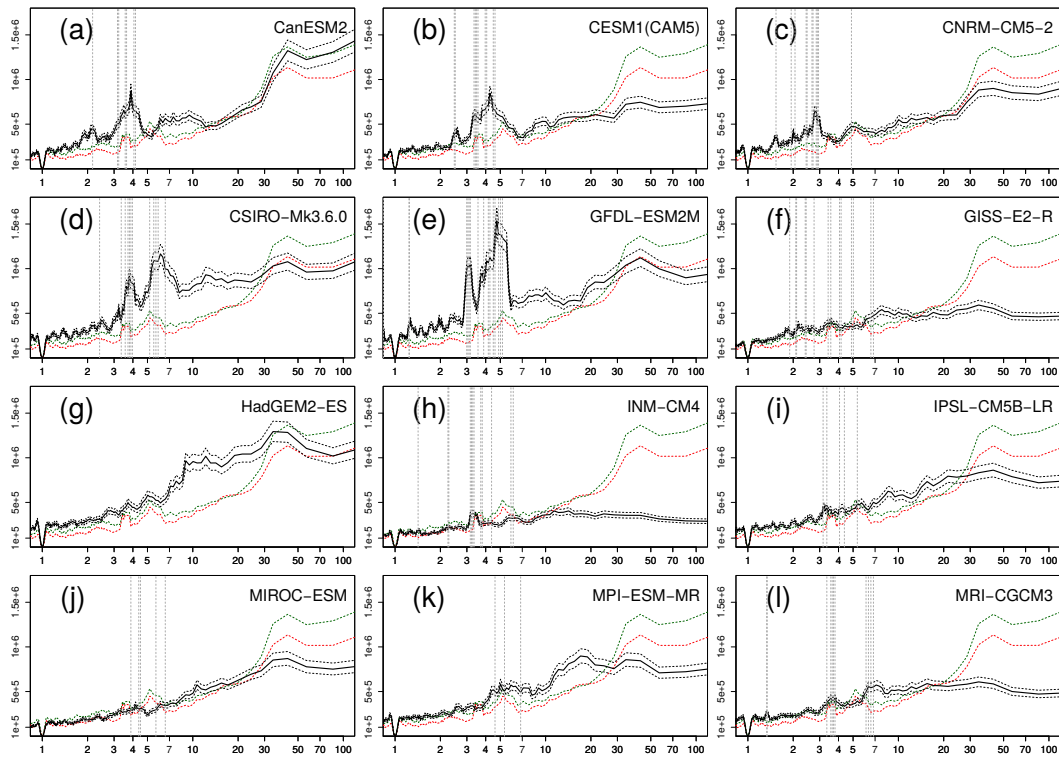
Figure 5. As Fig. 4a but now for each climate model (black line). The reanalysis spectra are shown as a reference (dashed green and red lines). The dashed vertical lines indicate the climate model multi-annual periods significant at 5 % level.

ica. Secondly, the mode contains tropical Pacific temperature anomalies, like in the ENSO phenomenon (e.g. Kleeman, 2008). The cold (warm) maximum is in phase 1 (5) with the anomalies extending to the South American continent. Thirdly, there are travelling temperature anomalies at high latitudes on both hemispheres. These are described next.

In phase 1 (Fig. 6), there is a small warm temperature anomaly in the North Pacific (long 160° W, lat 30° N). This pattern slowly moves northeast, reaching Alaska in phase 5 and then gradually dissipating over northernmost North America in phase 8 (and being visible still in phase 1). There is a very similar evolution of a cold anomaly starting in phase 5. At the same time, there is an oscillating temperature anomaly over the Eurasian continent in the opposite phase. In Fig. 6, there is also a travelling temperature anomaly in the Southern Hemisphere. In phase 8 (Fig. 6), there is a cold anomaly over the Southern Ocean (long 160° W). This strengthens, moves east, weakens, crosses the Antarctic Peninsula in phase 4 and remains in the Weddell Sea until phase 7. Similarly, there is a warm anomaly in phase 4 (long 160° W) with similar evolution as the cold one.

Next, 20CR and the CMIP5 model behaviour is studied. The 3.5-year mode is significant in 20CR and ERA-20C. For the illustration, we have chosen component pairs from the model decompositions (Fig. S1 in the Supplement) that have spectral peaks between 3 and 4 years and do not express sub-

stantial variability on other timescales. In most climate models, such a corresponding mode exists, except in models g and k. In model c this mode is not significant at 5 % level, but it is illustrated anyway. The Supplement reveals how these modes are represented in different data sets (Figs. S3–S14). The format is the same as in Fig. 6. A short summary is presented next.

In 20CR (Fig. S3), the anomalies are weaker compared to ERA-20C (Fig. S4). This is mainly because the 3–4 year mode is distributed on two component pairs in 20CR whereas in ERA-20C it is concentrated on one pair. Nevertheless, a similar although weaker signal is evident in 20CR, such as the northeast propagation of the North Pacific temperature anomaly. (Note that in Fig. 3, the combination of components 3, 4, 7 and 8 produce highly similar global patterns for 20CR and ERA-20C.) A prominent feature is also the opposite temperature anomalies in the northern Eurasia vs. North America. All models (Figs. S5–S14) produce a temperature anomaly to the equatorial Pacific Ocean (and South America). The amplitude is larger and/or the area extends further to the west than in ERA-20C in six models (a, b, d, e, h, l). The anomaly pattern in the northwestern North America is present in all the models to some extent. In the reanalyses, the anomaly is strictly confined to land areas but in most models, it is either somewhat misplaced or extends to the adjacent sea areas and the Eurasian continent. Models c, e, and f
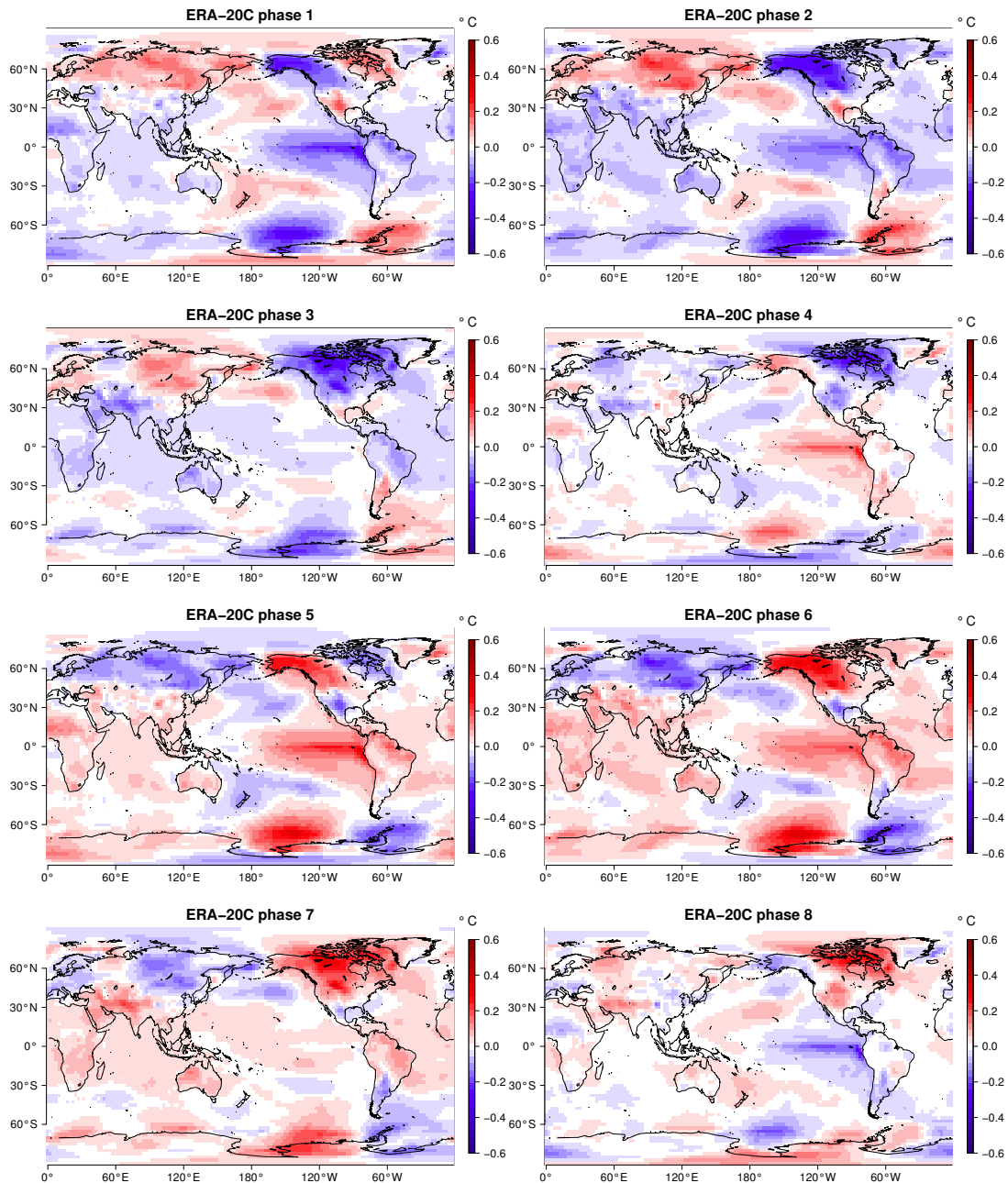
Figure 6. ERA-20C phase (1–8) composites of the 3–4 year variability mode. Unit °C.

produce the North American pattern quite similar to reanalyses, and the northeast propagation is captured to some extent by models b, c, f, i and l.

4 Discussion

We note that a substantial portion of variance at interannual to inter-decadal timescales can be attributed to "climate noise" associated with processes with timescales much shorter than the inter-annual scale (Wunsch, 1999; Feldstein,

2000). If the amplitude of the variability mode exceeds some noise threshold (such as red noise), then the variability mode is also likely driven by some process external to the atmosphere, in addition to the climate noise. For example, a large part of the inter-annual atmospheric ENSO pattern is presumably driven by anomalies of tropical diabatic heating associated with sea surface temperature anomalies (Feldstein, 2000). We assume that for this reason the multi-annual patterns related to ENSO clearly exceed the noise threshold in the results of this study.

5　Conclusions

The aim of this study is to decompose the 20th century climate variability into its multi-annual modes, and to assess how these modes are represented by the contemporary climate models. To this end, two 20th century reanalysis data sets and 12 CMIP5 model simulations for the years 1901–2005 of the monthly mean near-surface air temperature have been decomposed using RMSSA. The statistical significance of the identified modes has been estimated with Monte Carlo simulations. The main conclusions are as follows.

Spectral properties of the 20CR and ERA-20C reanalysis data appear remarkably similar. The most prominent forms of variability in both data sets are related to approximately 3.5- and 5-year modes which are significant at 5 % level. The spectral power in ERA-20C is systematically slightly higher than in 20CR. The 3.5-year mode is illustrated in more detail. In ERA-20C, the mode is associated with a typical ENSO pattern of temperature anomalies in the equatorial Pacific Ocean, South America, and northwestern North America. On top of these, the mode also contains a northeast-propagating temperature anomaly over northernmost North America, and another eastward-propagating anomaly in the vicinity of western Antarctica. Since about the 1970s, the amplitude of this 3.5-year global mode have increased.

None of the 12 coupled climate models closely reproduce all aspects of the reanalysis spectra, although some models represent many aspects well. For instance, the GFDL-ESM2M model has two nicely separated ENSO-related periods although they are relatively too prominent as compared to the reanalyses. Also, a number of models represent the propagating temperature anomalies at 3–4 year time frame. Some suggestions are provided in the text for potential model development aspects.

There is an extensive Supplement available presenting the results in visual format for each reanalysis and model data set. In the future, relaxation of the uni-variate nature of the present study would seem a natural extension. This is now possible since the use of random projections allows efficient data structures, preserving compression. Of special interest would be to study behaviour of variables directly linked with atmosphere–ocean coupling processes, such as heat, momentum and moisture fluxes over oceans.

6　Data and code availability

All data used in this study were downloaded from open sources. The RMSSA algorithm and the statistical significance testing are implemented using GNU licensed free software from the R Project for Statistical Computing (http://www.r-project.org). Our implementation is available on request. The animations of the 3–4 year mode are available for all data sets at https://www.youtube.com/channel/UCu1zJdwJfLaXvfvTqsKCLHw (Järvinen et al., 2016). See also Compo et al. (2011, 20CR), Taylor et al. (2012, CMIP5) and Poli et al. (2013, ERA-20C).

Appendix A: Randomised multi-channel singular spectrum analysis (RMSSA)

The RMSSA algorithm and the significance test is briefly presented here. The original data matrix is $\mathbf{X}_{N \times L}$, where the columns are called *channels*. In case of gridded data set, N represents the time steps and L is the number of grid points. It is useful to think of N as the time steps when the *sample* of dimension L is collected. The dimension reduction is a projection $\mathbf{X}_{N \times L} \rightarrow \mathbf{P}_{N \times k}$, where $L \gg k$. In other words, we preserve all samples but reduce the sample dimension from L to k. The dimension reduction is performed in two steps: (1) generate a random matrix $\mathbf{R}_{L \times k}$, where the matrix elements are $r_{ij} \sim N(0, 1)$ and column vectors of $\mathbf{R}$ are normalised to unit length, and (2) project $\mathbf{X}$ onto $\mathbf{R}$:

$$\mathbf{P}_{N \times k} = \mathbf{X}_{N \times L} \mathbf{R}_{L \times k}. \tag{A1}$$

The next step is to construct an augmented data matrix $\mathbf{A}$, which contains M lagged copies of each channel in $\mathbf{P}$. In RMSSA, M represents the lag window. $\mathbf{A}$ now has Mk columns and $N' = N - M + 1$ rows. The singular value decomposition of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{U}\mathbf{D}^{1/2}\mathbf{V}^T \tag{A2}$$

The vectors of $\mathbf{U}$ are the eigenvectors of $\mathbf{Z} = \frac{1}{Mk}\mathbf{A}\mathbf{A}^T$ and $\mathbf{V}^T$ contains the eigenvectors of $\mathbf{C} = \frac{1}{N'}\mathbf{A}^T\mathbf{A}$. These vectors are orthogonal and often called space–time principal components (ST-PCs) and space–time empirical orthogonal functions (ST-EOFs), respectively. Note that the ST-EOFs are now in reduced space k. Diagonal elements of $\mathbf{D}$ are the eigenvalues of $\mathbf{C}$ or $\mathbf{Z}$. Finally, the eigenvectors (ST-EOFs) are calculated in the original L dimensional space by

$$\mathbf{V} \approx \mathbf{A}_o^T \mathbf{U}(\mathbf{D}^{1/2})^{-1}, \tag{A3}$$

where $\mathbf{A}_o$ is the augmented matrix of the original data matrix $\mathbf{X}$. Note that the calculation of ST-EOFs in Eq. (A3) can be limited only to the eigenmodes of interest.

The ST-PCs can be represented in the original coordinate system by the RCs (Plaut and Vautard, 1994; Ghil et al., 2002). This transformation is given by

$$rc_{le}(n) = \frac{1}{M_n} \sum_{m=I_n}^{J_n} u_e(n - m + 1)v_{le}(m), \tag{A4}$$

where u_e are the ST-PCs and v_{le} are the ST-EOFs calculated in Eq. (A3) (the part of ST-EOF corresponding to channel l). e is the index of the eigenmode that is calculated. The normalisation factor M_n and the summation bounds I_n and J_n are given in Ghil et al. (2002) and for the central part of the time series ($M \leq n \leq N - M + 1$) they are $(M, 1, M)$, respectively.

RMSSA with significance testing is briefly presented in the following. Testing the MSSA components against a red noise null hypothesis requires orthogonal input vectors, which are obtained by calculating first a conventional PCA and retaining a set of dominant PCs. Therefore some additional calculation steps are included in the RMSSA-algorithm:

SVD of lower dimensional matrix $\mathbf{P}$ is calculated to obtain the principal components (PCs, calculated as $\mathbf{U}\mathbf{D}^{1/2}$). PCs fulfil the orthogonality constraint exactly. PCs, which explain a large part of the variance of the data set (e.g. 50 first), are retained to obtain matrix $\mathbf{T}$, where the columns are the PCs. Next, the augmented matrix $\mathbf{A}_{PC}$ is constructed from $\mathbf{T}$ and SVD is calculated as in Eq. (A2).

Finally, a large number of red noise processes (i.e. surrogate data sets) are generated, and the confidence limits for the MSSA eigenmodes are determined. This significance test (Monte Carlo MSSA) is described in detail in Allen and Robertson (1996).

References

Achlioptas, D.: Database-friendly random projections: Johnson-Lindenstrauss with binary coins, J. Comput. Syst. Sci., 66, 671–687, 2003.

Allen, M. R. and Robertson, A. W.: Distinguishing modulated oscillations from coloured noise in multivariate datasets, Clim. Dynam., 12, 775–784, 1996.

Ba, J., Keenlyside, N. S., Latif, M., Park, W., Ding, H., Lohmann, K., Mignot, J., Menary, M., Otterå, O. H., Wouters, B., Salas y Melia, D., Oka, A., Bellucci, A., and Volodin, E.: A multi-model comparison of Atlantic multidecadal variability, Clim. Dynam., 9, 2333–2348, 2014.

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., and Vialard, J.: ENSO representation in climate models: from CMIP3 to CMIP5, Clim. Dynam., 42, 1999–2018, 2014.

Bingham, E. and Mannila, H.: Random projection in dimensionality reduction: applications to image and text data, in: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'01, ACM, New York, 245–250, 2001.

Broomhead, D. S. and King, G. P.: Extracting qualitative dynamics from experimental data, Physica D, 20, 217–236, 1986a.

Broomhead, D. S. and King, G. P.: On the qualitative analysis of experimental dynamical systems, in: Nonlinear Phenomena and Chaos, edited by: Sarkar, S., Adam Hilger, Bristol, UK, 113–144, 1986b.

Cleveland, R. B., Cleveland, W. S., McRae, J. E.b and Terpenning, I.: STL: A Seasonal-Trend Decomposition Procedure Based on Loess, Journal of Official Statistics, 6, 3–73, 1990.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Q. J. Roy. Meteor. Soc., 137, 1–28, doi:10.1002/qj.776, 2011.

Feldstein, S. B.: The timescale, power spectra, and climate noise properties of teleconnection patterns, J. Climate, 13, 4430–4440, 2000.

Fredriksen, H.-B. and Rypdal, K.: Spectral Characteristics of Instrumental and Climate Model Surface Temperatures, J. Climate, 29, 1253–1268, doi:10.1175/JCLI-D-15-0457.1, 2016.

Ghil, M.: Natural climate variability, in: Encyclopedia of Global Environmental Change, edited by: Munn, T., Vol. 1, J. Wiley & Sons, Chichester/New York, 544–549, ISBN 0-471-97796-9, 2002.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climatic time series, Rev. Geophys., 40, 1–41, 2002.

Groth, A. and Ghil, M.: Monte Carlo Singular Spectrum Analysis (SSA) Revisited: Detecting Oscillator Clusters in Multivariate Datasets, J. Climate, 28, 7873–7893, 2015.

IPCC (Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.): Evaluation of Climate Models, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Järvinen, H., Seitola, T., Silén, J., and Räisänen, J.: Animations of the 3–4 year mode for all data sets, available at: https://www.youtube.com/channel/UCu1zJdwJfLaXvfvTqsKCLHw, last access: 16 November 2016.

Kleeman, R.: Stochastic theories for the irregularity of ENSO, Philos. T. R. Soc. A, 366, 2509–2524, 2008.

Knutson, T. R., Zeng, F., and Wittenberg, A. T.: Multimodel Assessment of Regional Surface Temperature Trends: CMIP3 and CMIP5 Twentieth-Century Simulations, J. Climate, 26, 870–8743, 2013.

Mann, M. E. and Lees, J. M.: Robust estimation of background noise and signal detection in climatic time series, Climatic Change, 33, 409–445, 1996.

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., van Oldenborgh, G., Vecchi, G., and Yeager, S.: Decadal Climate Prediction: An Update from the Trenches, B. Am. Meteorol. Soc., 95, 243–267, 2014.

North, G. R., Bell, T. L., Cahalan, R. F., and Moeng, F. J.: Sampling errors in the estimation of empirical orthogonal functions, Mon. Weather Rev., 110, 699–706, 1982.

Plaut, G. and Vautard, R.: Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere, J. Atmos. Sci., 51, 210–236, 1994.

Poli, P., Hersbach, H., Tan, D., Dee, D., Thepaut, J. N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P., Dragani, R., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: The data assimilation system

and initial performance evaluation of the ECMWF pilot re-analysis of the 20th-century assimilating surface observations only (ERA-20C), ERA Report Series, ECMWF, available at: http://www.ecmwf.int/en/research/climate-reanalysis/era-20c and http://www.ecmwf.int/en/elibrary/11699-data-assimilation-system-and-initial-performance-evaluation-ecmwf-pilot-reanalysis (last access: 14 November 2016), 2013.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res.-Atmos., 108, 446–469, 2003.

Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., Ansell, T. J., and Tett, S. F. B.: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset, J. Climate, 19, 446–469, 2006.

Russell, A. M. and Gnanadesikan, A.: Understanding Multidecadal Variability in ENSO Amplitude, J. Climate, 27, 4037–4051, doi:10.1175/JCLI-D-13-00147.1, 2014.

Seitola, T., Mikkola, V., Silen, J., and Järvinen, H.: Random projections in reducing the dimensionality of climate simulation data, Tellus A, 66, 25274, doi:10.3402/tellusa.v66.25274, 2014.

Seitola, T., Silén, J., and Järvinen, H.: Randomised multichannel singular spectrum analysis of the 20th century climate data, Tellus A, 67, 28876, doi:10.3402/tellusa.v67.28876, 2015.

Slingo, J. and Palmer, T.: Uncertainty in weather and climate prediction, Philos. T. R. Soc. A., 369, 4751–4767, 2011.

Stevenson, S., Fox-Kemper, B., Jochum, M., Rajagopalan, B., and Yeager, S. G.: ENSO model validation using wavelet probability analysis, J. Climate, 23, 5540–5547, 2010.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Thomson, D. J.: Spectrum estimation and harmonic analysis, Proc. IEEE, 70, 1055–1096, 1982.

Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E., and Phillips, A. S.: Quantifying the role of internal climate variability in future climate trends, J. Climate, 28, 6443–6456, 2015.

Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D, 35, 395–424, 1989.

Wittenberg, A. T.: Are historical records sufficient to constrain ENSO simulations?, Geophys. Res. Lett., 36, L12702, doi:10.1029/2009GL038710, 2009.

Wunsch, C.: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations, B. Am. Meteorol. Soc., 80, 245–255, 1999.