



An evaluation of current capabilities of modelling low-frequency climate variability

Heikki Järvinen¹, Teija Seitola^{1,2}, Johan Silén², and Jouni Räisänen¹ ¹Department of Physics, University of Helsinki, Finland ²Finnish Meteorological Institute, Helsinki, Finland *Correspondence to:* Heikki Järvinen (heikki.j.jarvinen@helsinki.fi)

Abstract. A crucial performance test of Earth system models is their ability to simulate the climate mean state and variability. Here we concentrate on representation of inter-annual to multi-decadal variability in 12 CMIP5 climate model simulations. Reference climate is provided by two 20th century reanalysis data sets of monthly mean near-surface air temperature. The spectral decomposition is based on Randomised Multi-Channel Singular Spectrum Analysis (RMSSA). The main results are

- 5 as follows. First, the total spectra for the two reanalysis data sets are remarkably similar in all time scales, except that spectral power of decadal variability (10–30 yr) differ in these data by about 30 %. Second, none of the 12 coupled climate models closely reproduce all aspects of the reference spectra, although some models represent many aspects well. For instance, the IPSL-CM5B-LR model is close to reanalyses but has too little multi-decadal variability, and the HadGEM2-ES model is close to reanalyses except the notable over-activity at periods at and around 10 yr.
- 10 *Keywords:* climate model assessment, dimensionality reduction, random projection, 20th century reanalysis, significance testing, RMSSA algorithm

1 Introduction

20

The ultimate goal in developing Earth system models (ESM) is to exploit the inherent predictability of the Earth system to 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). Prediction tools are very complex and their testing goes hand-in-hand with their development. A crucial performance test of ESMs is related to their ability to simulate well the observed climate mean state and the variability around the mean.

Here we focus on ESMs of today and how they represent inter-annual to multi-decadal climate variability. This is a very broad range of temporal scales and it is associated with a multitude of spatial scales. Generally speaking, spectral misrepresentations appear either due to lack of variability in a model or over-activity of a model in some temporal scales. Conclusions about model deficiencies based on spectral differences are very scale dependent, and some general guidance can be obtained by

about model deficiencies based on spectral differences are very scale dependent, and some general guidance can be obtained by thinking about the mechanisms of natural climate variability (e.g., Ghil 2002). Essentially, short time scale variability (below 2 yr) in the model spectrum of near-surface air temperature is most likely related to the representation of internal variability of the





5

atmosphere. Associated model deficiencies, such as low resolution, can explain most of these weaknesses. Inter-annual variability (2–7 yr) is prominently related to the ENSO phenomenon, and simulation weaknesses point more towards deficiencies in atmosphere-ocean feedback processes and ocean model dynamics. Decadal-to-multi-decadal variability can be thought of being driven by ocean dynamics. There is however clear indication of multi-decadal variability, such as Atlantic multi-decadal oscillation (AMO), that may be driven by stochastic forcing of mid-latitude atmospheric circulation on ocean, and changes

in ocean circulation may rather be a response rather than driver of the variability (Clement et al., 2015). This interpretation widens the scope of possible root causes from model errors in ocean dynamics to coupling issues.

Representation of inter-annual to multi-decadal climate variability among models participating in climate model intercomparisons (such as CMIP5) has been studied by e.g. Bellenger et al. (2014), Knutson et al. (2013), Ba et al. (2014), and

- 10 Fredriksen and Rypdal (2016). We will add to this literature by applying a recently developed powerful spectral analysis tool in this field. We identify the spectral signatures by applying Randomised Multi-Channel Singular Spectrum Analysis (RMSSA; Seitola et al. 2014, 2015) which is an advanced time series analysis method for high-dimensional problems. The strength of RMSSA lies in the fact it is able to detect spatially evolving chains of events in high-dimensional systems by resolving eigenmodes of spatio-temporal covariance data. This is a significant advantage, say, over PCA which only resolves eigenmodes of
- 15 spatial covariances. This can lead to undesirable projection of temporal evolution of an event onto a number of different eigenmodes. Additional benefits of RMSSA are: (i) dimension reduction via random projections enable applications in extremely high dimensional problems, (ii) convergence properties of the eigenmode decomposition are very good allowing better physical interpretation of fewer components, and (iii) the resulting spectrum is straightforward to test for significance. The paper is organised as follows: data and methods are explained in Section 2, results in Section 3, followed by discussion and conclusions.

20 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

- 25 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). 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. MSSA eigenvectors are called here space-time EOFs (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 reduced length and they cannot be located into the same index
- 30

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, a computationally more efficient variant, called Randomised MSSA (RMSSA; Seitola et





al., 2015), 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) 5 in order to construct a lower dimensional representation. In this study, we reduce the data volume to about 0.8 to 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 lowerdimensional 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.

10 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 spectral properties of 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

15 explanatory power, i.e. multiplied by the corresponding eigenvalue. Therefore the components with more explanatory power have also 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.

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 π/2 out of phase with each other. However, 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 consists of the monthly mean near-surface air temperature from the historical 20th Century simulations of 12 different climate models (Table 1). Near-surface temperature was chosen, because many processes must be adequately represented in





5

coupled models to realistically capture the observed temperature distribution (Flato et al., 2013). These include processes in the Earth system component models (atmosphere, ocean, etc.) as well as in their mutual coupling models.

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.

As a reference, we used two reanalysis data sets: the 20th Century Reanalysis V2 data (hereafer 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,

- 10 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
- 15 with the climate model simulations, the same time period is used (1901–2005, i.e., 1260 monthly mean fields). 20CR has ~2.0 degree horizontal resolution and we have used gaussian gridded (192 × 94) data from 3-hour forecast values. The horizontal resolution of ERA-20C is approximately 125 km (T159) in a grid of 360 × 181 points.

2.5 Data processing

Some pre-processing of the data sets is needed before applying RMSSA and statistical significance testing. The data sets were standardised (i.e. the time series of each grid point was mean-centered and divided by its standard deviation) to avoid overweighting the grid points with higher variance. This adds weight on the lower latitude variability, where ENSO-type variability is pronounced. On the other hand, no-scaling would make the higher latitude variability dominant because of larger amplitude variations there. Furthermore, each data set was de-trended, and the dominating annual cycle was estimated using Seasonal-Trend Decomposition (STL; Cleveland et al., 1990) and removed from the original time series.

After the pre-processing, the dimension reduction step of RMSSA was applied so that approximately 3 % of the original dimensions of 20CR data, 0.8 % of ERA-20C, and about 5 % of climate model data were retained – the different percentages are due to different volumes of the original data. The lag window in the analysis was 20 yr (240 months). The total spectra was obtained from this analysis.

In the statistical significance test, the first 50 PCs of each data set were retained as input. Those PCs explain 80 % 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 intervals (90 % and 95 %) for the oscillatory modes were estimated. We note that transformation to PCs may intefere with the detection of weak signals, as demonstrated by Groth and Ghil (2015).





In the following section we will compare the spectral properties of the reanalysis and model data sets. Furthermore, we will test the spectra of each data set against a red noise null hypothesis in order to distinguish signal from noise. Finally, we will compare the spatial patterns of an oscillatory mode with a 3.5 yr period as represented by different data sets.

3 Results

20

5 3.1 Spectral similarity of the two reanalysis data sets

We first demonstrate the spectral similarity of the two reanalysis data sets. Figure 1 displays, in terms of explained variance, the leading 30 ST-PCs for 20CR and ERA-20C and the corresponding power spectra. The decomposition reveals that variance is distributed in a very similar way in 20CR and ERA-20C. The ordering of component pairs is not identical but there is a very clear correspondence of the spectral peaks. For instance, the 1.7 yr peak in the 20CR components 29–30 corresponds to the components 24, 25 in ERA 20C. In summary

10 components 24–25 in ERA-20C. In summary,

- the trend components with multi-decadal periods (components 1 and 2) explain 6 % and 5.3 % of the variance in 20CR and ERA-20C, respectively, with very similar spectra (the length of the time series 105 years restricts, of course, the correct identification of multi-decadal oscillations)
- the spectral peak at 3.5 yr and the broad one around 5 yr are very similar in components 3–6, although the associated 10–
- 15 20 yr variability in 20CR is separate in components 9–10 in ERA-20C (i.e. ENSO variability has a decadal component in 20CR)

The similarity of 20CR and ERA-20C is striking in the total spectra (Fig. 2a). Variability shorter than about 10 years is captured similarly by the two reanalyses. 20CR is somewhat more lively in decadal scale (10–30 yr) and has about 30 % higher spectral density there. One explanation for this difference may be the decadal variability of ENSO which is present in 20CR (Fig. 1, components 5–6) but is missing from ERA-20C.

The statistical significance testing of the periodicities in 20CR (Fig. 2b) and ERA-20C (Fig. 2c) reveals that nearly the same periods rise above the red noise in the two data sets (the red dots above the area covered by the vertical bars). It is logical that the frequency corresponding to the annual cycle is present in the red noise surrogates while it is removed from the data sets (and therefore the red dots are far below the bars). The periods rising above the red noise at 5 % and 10 % significance level

are tabulated in Table 2, columns 1 and 2, and are nearly the same in 20CR and ERA-20C. Our conclusion is therefore that the low-frequency climate variability in the near-surface air temperature is very similar in the two reanalyses.

3.2 Evaluation of the simulated spectra

The climate model simulation data was processed exactly the same way as the reanalysis data. The total spectrum of each model is shown in Figure 3 with the reanalysis spectra on the background as a reference (dotted lines). The spectra are objective
30 measures of model performance. We evaluate *subjectively* how the models reproduce the reanalysis spectra, and adopt the





following terminology: a model can be either under-active (over-active) if the spectral density is lower (higher) than in the reanalysis spectra, or a model is on-target with respect to spectral density. This enables us to make an overall evaluation of the current capabilities of climate models to represent low-frequency variability.

The subjective evaluation is summarised in Table 3. In representing multi-decadal variability, three models are on-target 5 while the rest are under-active. In decadal variability, majority of the models are on target, while four are over-active. Only one model has both of these variabilities on-target (a). In ENSO variability, two models (i and g) seem to have both the 3.5 and 5 yr periods on-target. Five models have one of these periods on-target, and the rest of the models are either under- or over-active. With respect to the inter-annual variability, four models seem to be over-active, while the rest are on target.

We are not going to rank the models, but two models (i and g) seem to perform particularly well and have four out of five 10 key periodicities on-target. The IPSL-CM5B-LR model (i) is close to the reanalyses all the way from inter-annual and ENSO variability to the decadal scale of 20 years and only seem to have too little multi-decadal variability. The HadGEM2 model (g), on the other hand, is reasonably close to the reanalyses across the spectrum, except for the notable over-activity at periods at and around 10 yr. CanESM2 (a) is the only model that closely reproduces both the decadal and multi-decadal variability.

Overall, the clearest signal here is that the models generally seem to lack multi-decadal variability, and some models are over-active in representing decadal and inter-annual variability. In ENSO time scales, only two models are on-target.

3.3 Significance test by Monte-Carlo MSSA

Table 2 contains the periods rising above the red noise at 5 % and 10 % significance level. The periods for 20CR and ERA-20C are in the columns 1 and 2. Both data sets have five periodicities significant at 5 % level, four of which are common and one specific. Additionally, there are three common periodicities significant at 10 % level. In terms of statistical significance, 20CR and ERA-20C behave in a very similar manner.

20

Table 2 gives an impression that the number of statistically significant periodicities is very large in models, in general. In this respect, the climate model data is dissimilar to the reanalysis data. There is only one model (model k) which has fewer significant periodicities than the reanalyses, and three models (g, h, j) are somewhat close to reanalyses with 11-13 significant periodicities. The rest of the models have many more periodicities (up to 30). Interestingly, the HadGEM2 model (g) has the 10 ur peak discussed earlier significant at 5 % level.

25 10 yr peak, discussed earlier, significant at 5 % level.

The number of models that correctly detected either the 20CR or ERA-20C periodicities were seven models for the 1.7 yr period, five for 2.2 yr, four for 3.5 yr, seven for 3.6, four for 4.2 yr, and two for 5.2 yr. In addition to these, there were many "false alarms", as seen in Table 2.

3.4 Spatial patterns of the 3.5 yr mode

30 The oscillatory mode with a 3.5 yr period was significant in 20CR and ERA-20C and in seven climate models. The spectral peak at 3.5 yr has some power on both sides of the peak. Therefore, a 3–4 yr mode would perhaps be a more appropriate name for the peak. Oscillations at these periods are usually attributed to ENSO (e.g. Kleeman, 2008). We illustrate here how this





periodicity appears in different data sets. Since the 3.5 yr period is not reproduced by all of the climate models, we have chosen in these cases a periodicity that is close to 3.5 years.

We have calculated the reconstructed components (RC) corresponding to the 3.5 yr mode in order to visualise the associated spatial temperature anomaly patterns. The result is a time series of the original (centered) data corresponding to the 3.5 yr mode in each grid point and according to its original variance. We therefore have an animation of the global 3.5 yr mode for

5

each data set for the whole time period (1901–2005).

To synthesise the animations, we have calculated composite maps of the 3.5 yr mode for each data set (Fig. 4). The patterns are composites of eight instances when the mode is in its maximum positive phase in the Niño3.4 region (120° W– 170° W, 5° N– 5° S). Positive events are defined as an average of winter months (November–March).

- 10 The top row of Fig. 4 displays the patterns for the reanalyses. One can see typical El Niño related temperature anomalies, such as positive anomalies in the equatorial Pacific Ocean and South-America, and northwestern North-America. There is also a cold anomaly over the northern parts of the Eurasian continent, a dipole structure in the western Antarctica, and warm Africa. The patterns are remarkably similar in 20CR and ERA-20C with somewhat larger amplitudes in 20CR, except the stronger dipole in ERA-20C. Next, we will try to *subjectively* assess model performance in reproducing the spatial patterns of the 3.5
- 15 yr oscillation.

All models produce a warm pool to the equatorial Pacific Ocean and South-America. The amplitude is larger or the area extends too far to the west in five models (a, b, d, e, h). Unlike in the reanalyses, the pool extends to the Atlantic in four models (a, d, e, h). The warm anomaly in the South-American land area is too weak in five models (c, f, g, j, k).

The warm anomaly in the northwestern North-America is present in all the models. In the reanalyses, the anomaly is strictly confined to land areas but in most models, it is either misplaced or extends to the adjacent sea areas and the Eurasian continent.

The cold anomaly over the Eurasian continent is not well represented in the model data. There is a weak cold anomaly in three models (f, i, l), a weak warm anomaly in one model (c), and a mixture of cold and warm areas in the rest.

The warm pool in the Amundsen sea related to the warm-cold dipole around the western Antarctica is present in nine models (a, e, f, g, h, i, j, k, l). The cold pool in the Weddell sea is present in five models (c, d, f, g, k). Three models (f, g, k) represent both pools, and thus the dipole structure in well represented. In Africa, the anomaly is on the positive side in all models.

In Table 3, there are three over-active models with respect to the 3.5 yr period (a, d, e). These are associated with large amplitudes in Fig. 4. Additionally, two models also seem to have large amplitudes in Fig. 4 (b, h). Both of these have higher spectral density than the reanalyses (Table 3), but were anyhow assessed as "on-target" in Table 3. Four models were underactive in Table 3 (c, f, j, k). These correspond to the low-amplitude maps in Fig. 4. In summary, the patterns of Fig. 4 are in support of the subjective analysis of the total spectra of Table 3.

30

25

4 Discussion

Table 2 shows that there is a much larger number of significant periodicities in the model data than in the reanalyses. This seems to imply that "nature" tends to produce only a few but statistically significant periodicities, and the other potential periodicities





5

are somehow "worn out" via non-liner interactions and feedbacks so that they cannot be distinguished from red noise. We can think of two possible causes for this. First, it may simply be because the reanalysis data represent ensemble mean while the model data are individual simulations. This difference may appear as a difference in the number of significant periodicities. Second, it may be that something in models prohibits the wearing process from occurring, and we can observe the excessive number of periodicities above the noise level – as if the models were missing some non-linear processes. However, MPI-ESM (model k) has fewer significant periodicities than observed, and to our knowledge, this model does not fundamentally differ from the other models. Therefore, it is unclear what exactly lies behind this result.

We have not discussed the interpretation of the 1.7 yr interannual variability. Based on the related spatial patterns, it may be that it is a harmonic of the ENSO variability $(2 \times 1.7 \text{ yr} = 3.4 \text{ yr})$.

10 5 Conclusions

In this study, we decomposed the 20th Century near-surface temperature variability into its inter-annual to multi-decadal eigenmodes. We used two state-of-the-art 20th Century reanalysis data sets and 12 historical climate model simulations extracted from the CMIP5 data archive. The analysis was performed using the randomised multi-channel singular spectrum analysis (RMSSA), which is particularly well suited to high-dimensional time series analysis. The statistical significance of the identified eigenmodes was estimated with Monte-Carlo simulations. The main conclusions are as follows:

- The spectral properties of the two reanalyses (20CR and ERA-20C) are remarkably similar, the only notable difference being the different spectral density in the decadal scale variability (10–30 yr). Also, nearly the same periodicities rise above the red noise at the 5 % and 10 % significance levels.
- 20

Majority of the climate models are under-active in representing the multi-decadal climate variability (> 30 yr), some models are over-active in decadal (10–30 yr) or inter-annual (< 2 yr) variability, and only two models are on-target in both the 3.5 yr and 5 yr ENSO variabilities.

- The IPSL-CM5B-LR and the HadGEM2-ES are the closest to the total spectra of the reanalyses.

Relaxation of the uni-variate nature of the present study remains as a subject for future research. It would be very interesting to extend the set of variables since the use of random projections allows efficient data compression. Of special interest would

25 be to study behaviour of variables directly linked with coupling processes, such as heat, momentum, and moisture fluxes over oceans.

6 Data and code availability

All data used in this study was 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 (www.r-project.org). Our





15

implementation is available on request. The animations of the 3.5 yr periodicity are available for all data sets at https://www.youtube.com/channel/UCu1zJdwJfLaXvfvTqsKCLHw.

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 5 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 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} :

10
$$\mathbf{P}_{N \times k} = \mathbf{X}_{N \times L} \mathbf{R}_{L \times k}.$$
 (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 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 D are the eigenvalues of C or 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 A_o is the augmented matrix of the original data matrix **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 reconstructed components, RCs (Plaut and Vautard, 1994; Ghil et al., 2002). This transformation is given by

$$rc_{le}(n) = \frac{1}{M_n} \sum_{m=I_n}^{J_n} u_e(n-m+1)v_{le}(m),$$
(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 ($M \le n \le N - M + 1$) they are (M, 1, M), respectively.





5

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 **P** is calculated to obtain the principal components (PCs, calculated as $UD^{1/2}$). PCs fullfil the orthogonality constraint exactly. PCs, that explain large part of the variance of the data set (e.g. 50 first), are retained to obtain matrix **T**, where the columns are the PCs. Next, the augmented matrix A_{PC} is constructed from **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 signicance test (Monte-Carlo MSSA) is described in detail in Allen and Robertson (1996).

10 *Author contributions*. HJ suggested the study and mostly wrote the article. TS implemented the methods, performed all computations, and wrote data and method descriptions, JS supported the method development, and JR the climate model data analysis.

Acknowledgements. This research has been funded by the Academy of Finland (project number 140771), Academy of Finland Centre of Excellence Programme (project number 272041), and the Fortum Foundation (grant number 201500127).





References

- Achlioptas, D.: Database-friendly random projections: Johnson-Lindenstrauss with binary coins, Journal of computer and System Sciences, 66(4), 671–687, 2003.
- 5 Allen, M. R. and Robertson, A. W.: Distinguishing modulated oscillations from coloured noise in multivariate datasets, Climate Dynamics, 12(11), 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, Climate Dynamics, 9, 2333–2348, 2014.
- 10 Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M. and Vialard, J.: ENSO representation in climate models: from CMIP3 to CMIP5, Climate Dynamics, 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.
- 15 Broomhead, D. S., and King, G. P.: On the qualitative analysis of experimental dynamical systems, in: Nonlinear Phenomena and Chaos, S. Sarkar (Ed.), Adam Hilger, Bristol, England, 113–144, 1986b.
 - Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Rädel, G. and Stevens B.: The Atlantic Multidecadal Oscillation without a role for ocean circulation, Science, 350, 320–324, doi: 10.1126/science.aab3980, 2015.

Cleveland, R.b., Cleveland, W.S., McRae, J.E. 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, Quarterly J. Roy. Meteorol. Soc., 137, 1–28, 2011.
- 25 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, [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. (Eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- 30 Fredriksen, H.-B. and Rypdal, K.: Spectral Characteristics of Instrumental and Climate Model Surface Temperatures, J. Climate, 29, 1253– 1268, doi: http://dx.doi.org/10.1175/JCLI-D-15-0457.1, 2016.
 - Ghil, M.: Natural climate variability, in: Encyclopedia of Global Environmental Change, T. Munn (Ed.), Vol. 1, J. Wiley & Sons, Chichester/New York, pp. 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, Reviews of geophysics, 40(1), 1–41, 2002.
- Groth, A. and Ghil, M.: Monte Carlo Singular Spectrum Analysis (SSA) Revisited: Detecting Oscillator Clusters in Multivariate Datasets, Journal of Climate, 28(19), 7873–7893, 2015.





25

- Kleeman, R.: Stochastic theories for the irregularity of ENSO, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366(1875), 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.
- 5 Mann, M. E. and Lees J. M.: Robust estimation of background noise and signal detection in climatic time series, Clim. 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
- 10 from the Trenches, Bull. Amer. Meteor. Soc., 95, 243–267, 2014.
 - Plaut, G. and Vautard, R.: Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere, Journal of the atmospheric sciences, 51(2), 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
- 15 the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), ERA Report Series, ECMWF, 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, Journal of Geophysical Research: Atmospheres, 108(D14), 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
- 20 changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset, Journal of Climate, 19(3), 446–469, 2006.
 - Seitola, T., Mikkola, V., Silen, J. and Järvinen, H.: Random projections in reducing the dimensionality of climate simulation data, Tellus A, 66, doi:http://dx.doi.org/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, doi:http://dx.doi.org/10.3402/tellusa.v67.28876, 2015.
 - Slingo, J. and Palmer, T.: Uncertainty in weather and climate prediction, Phil. Trans. R. Soc. A., 369, 4751–4767, 2011.
 - Taylor, K. E., Stouffer, R. J. and Meehl, G. A.: An Overview of CMIP5 and the experiment design, Bull. Amer. Meteor. Soc., 93, 485–498, 2012.
 - Thomson, D. J.: Spectrum estimation and harmonic analysis, Proc. IEEE, 70, 1055–1096, 1982.
- 30 Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D: Nonlinear Phenomena, 35(3), 395–424, 1989.





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

Model ID	Model name	Modeling center	Country	
а	CanESM2	CCCMA	Canada	
b	CESM1(CAM5)	NSF-DOE-NCAR	USA	
с	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	
1	MRI-CGCM3	MRI/JMA	Japan	





Table 2. Approximate periodicities (in years) detected by the Monte-Carlo significance test with a 20 year lag window length. Similar periodicities among the data sets are aligned. Numbers in the table are in bold if the significance level is at 5 % and in italics if at 10 %. Dominant periodicities of the oscillations are estimated with MTM.

		/		/	Mb	6	260	N2M	/	4		and a	~ ~ /	MR	~/
		ç Ç		SM2	im1-CA	am-CMD	ROMMES	N.E.SMI	, Elif	GENRI	CMA .	CM56	CHESTY LE	SW.M. CCM	A.J
DOCE	(RAI)	»/、	alcan	NOCE NO	S CH	r MCS	n jeft	NGIST	(a) Has	in hu	an in hes	i) MIR	N MPT	II MRI	/
<u> </u>	/ v	<u> </u>	<u> </u>	<u> </u>	~ ~	~ ~		<u>/ 🤄</u>	/ @	/ @	<u> </u>	12.2	<u> </u>	<u> </u>	
									10						
												9.5			
						8.1									
						7.8			7.8						
			6.8			6.9			7.1				6.8	6.9	
			0.0			6.1							0.0	0.0	
						5.9				5.9		5.9			
						5.7									
5.5	5.5					5.5					5.5				
						5.3	5.3				5.3		5.3		
	5.2						5.2	5.2	5.2						
							5	5	5						
							5	5	4.9						
			-						+.0						
				4.7		4.7	4.7	4.7							
			4.6	4.6		4.6	4.6					4.6	4.6		
				4.5			4.5				4.5	4.5			
				4.4		4.4	4.4			4.4	4.4	4.4			
4.2	1.2		4.3	4.3			4.3	4.3		4.3	12		4.3		
4.2	4.2		4.2	4.2		41	41	4.2			4.2			41	
			4	4		4	4					4			
			3.9			3.9	3.9				3.9	3.9		3.9	
			3.8			3.8		3.8		3.8	3.8			3.8	
			3.7			3.7				3.7				3.7	
3.6	3.6		3.6	3.6		3.6	3.6	3.6						3.6	
3.5	3.5		3.5	3.5	2.4	2.4	2.4	3.5	3.5		2.4	2.4		2.4	
			3.4	3.4	3.4	3.4	3.4	3.4		33	3.4	3.4		3.4	
			3.2		0.0	3.2	3.2	3.2		3.2	3.2	3.2		0.0	
			-			3.1	3.1	3.1		3.1	3.1	3.1			
			3	3			3	3			3			3	
			2.9		2.9										
			2.8	2.8	2.8			2.8	2.8	2.8	0.7				
			2.1		2.7		2.7	2.7			2.7		2.6	2.7	
2.5	25		2.5	2.5	2.0	2.5	2.0	2.5			2.5		2.0	2.0	
	2.0			2.4	2.4	2.4	2.4	2.4			2.4			2.4	
2.3	2.3				2.3		2.3	2.3		2.3	2.3	2.3		2.3	
2.2	2		2.2		2.2	2.2				2.2				2.2	
			2.1		2.1	2.1	2.1	2.1		2.1					
			2	1.0	2	4.0	2	2			2				
			1.9	1.9	1.9	1.9	1.9	1.9						1.8	
1.7	1.7		1.7	1.7	1.7		1.0	1.0			1.7			1.0	
			1.6	1.6	1.6	1.6		1.6	1.6						
				1.5	1.5		1.5	1.5							
			1.4	1.4		1.4	1.4		1.4	1.4	1.4			1.4	
				1.3			1.3							1.3	
				1.1		1.2	1.2								
	1			1.1		1.1	1.1		1.1						





 Table 3. Evaluation summary.

Under-active	On-target	Over-active
	Multi-decadal (> 30 yr)	
b, c, d, e, f	a, g, j	
h, i, k, l		
	Decadal (10–30 yr)	
	a, b, c, e, f,	d, g, j, k
	h, i, l	
	ENSO at \sim 5 yr	
a, f, h, j, l	c, g, i, k	b, d, e
	ENSO at \sim 3.5 yr	
c, f, j, k	b, g, h, i, l	a, d, e
	Inter-annual (< 2 yr)	
	d, f, g, h, i,	a, b, c, e
	j, k, l	







Figure 1. ST-PCs 1–30 of 20CR and ERA-20C monthly near-surface temperature 1901–2005 and their spectra. The lag window length Mused in RMSSA is 20 years (240 months). The annual cycle and linear trend are removed from the original data set before applying RMSSA. The proportion of the remaining variance explained by each component is also presented in the figure.







Figure 2. (a) Total spectrum of the reanalysis data sets, 20CR and ERA-20C. The total spectrum is a sum of the spectral density of all components (ST-PCs) related to each data set. The unit of the spectra is arbitrary. (b) Significance test of the monthly near-surface temperature variability in the 20CR data set 1901–2005. PCs 1–50 are used as input channels and the lag window length M is 20 years (240 months). In the test the red-noise basis is used. Red squares show the data eigenvalues plotted against the dominant frequency of the ST-PC corresponding to each eigenvalue. The vertical bars show the 2.5th and 97.5th percentiles of the eigenvalue distribution calculated from 1000 realisations of the red-noise surrogates. The ST-PCs that correspond to eigenvalues rising above the 97.5th percentiles are considered significant at the 5 % level. There is missing power at 1 yr due to the removal of the annual cycle. (c) Same as (b), but for the ERA-20C data set.







Figure 3. Total spectrum of each climate model. The total spectrum is the sum of the spectral density of all components (ST-PCs) related to each data set. For comparison, the total spectra of the reanalysis data sets are plotted in each subfigure. The unit of the spectra is arbitrary.







Figure 4. Global patterns of 3–4 yr oscillation of the near-surface temperature anomaly (°C) in the reanalysis and climate model data sets 1901–2005. The patterns are composites of 8 cases, when the oscillation is in its maximum positive phase in the Niño3.4 region (120° W– 170° W, 5° N– 5° S). The positive events are defined as an average of November-March. The identified patterns have similarities to the El Niño -phenomenon.