

Assessing the Spatio-temporal Structure of Annual and Seasonal Surface Temperature for CMIP5 and Reanalysis

Stefano Castruccio¹

February 20, 2016

Abstract

Comparison of climate models in an ensemble with reanalysis data is crucial for the climate model user community, as detection of discrepancies can convey information to improve models. Current comparison methodologies focus on statistical space-time properties of the climatological mean, and allow for a sensible model comparison only if the forcing scenario is identical or very similar. We analyse the annual and seasonal surface temperature of the CMIP5 ensemble and three reanalysis data products, and propose a scenario-independent, statistical-based classification relying on the space-time structure of the variability around the climatological mean. This approach exploits the gridded geometry of the atmospheric component of a global climate model, complements traditional criteria based on characteristics of the climatological trend and allows for a novel measure of similarity. For models with a similar physical scheme, we found a high degree of similarity for the same grid resolution, and a moderate similarity if the resolution is different. Further, we found that a considerable difference among reanalysis data products, thus indicating that different assimilation algorithms can significantly impact the space-time structure of the variability around the mean climate.

Key words: space-time models; CMIP5; spectral models; axially symmetric models

Short title: Evaluating CMIP5 and Reanalysis with Space-Time Models

¹School of Mathematics & Statistics, Newcastle University, Newcastle Upon Tyne, NE1 7RU United Kingdom. E-mail: stefano.castruccio@ncl.ac.uk

1 Introduction

Investigating if and to what extent Earth System Models (ESMs) can reproduce the state of the Earth's system, where discrepancies with observations occur, and how this information can be used to drive the development of the next generation of models is of paramount importance to the climate model user community. A vast literature discusses how climate model output should be compared with observational data (see e.g. Mauritsen et al. (2012); Bender (2008); Knutti et al. (2010); Sanderson and Knutti (2012); Buser et al. (2009); Furrer et al. (2007); Smith et al. (2009); Tebaldi et al. (2005); Jackson et al. (2008)), or with suitably post-processed gridded data product such as reanalysis (Kalnay et al., 1996; Kanamitsu et al., 2002; Saha et al., 2010), and which indices would be more suitable to quantify the agreement (Reichler and Kim, 2008). Most studies compare the same physical quantities from models and observations in terms of their empirical univariate distribution in time for every grid point or for spatially aggregated data. The comparison can be performed in terms of mean and variance (Taylor, 2001; Knutson et al., 2013; Xu et al., 2013), but also threshold exceedances indices such as growing-degree days (Moriondo and Bindi, 2006; Ye-Won et al., 2014), or quartile distributions (Braverman et al., 2011) and extreme events (Mannshardt-Shamseldin et al., 2010; Yao et al., 2013).

In this work we propose a novel comparison methodology based on the space-time dependence of the variability around the climatological mean between ESMs and reanalysis data. The approach allows for a comparison of models run under different inputs of carbon dioxide (CO_2) and other greenhouse gases concentration (*forcings*), a task which would not be easily performed with mean-driven criteria. The methodology relies on fitting a space-time statistical model to each ESM and comparing the estimated coefficients. If every climate model represents the true state of the system but deviates from it by a random error, the difference in estimated parameters between each ESM should not show any pattern (Buser et al.,

2009; Furrer et al., 2007; Smith et al., 2009; Tebaldi et al., 2005). Besides, since all reanalyses should represent the true state of the climate, we expect that different reanalysis data products should show small differences compared to those with the estimates of the ESMs. The analysis of spatial and temporal correlation of the climate output has been discussed in previous works (Koichi et al., 2012; Jun et al., 2008b,a; Lee et al., 2015), but always in terms of trend, although the necessity of pursuing an investigation on the variability about the mean climate was mentioned (but not implemented) in Jun et al. (2008a). This work is the first that addresses this issue by defining a space-time statistical model and comparing the estimated statistical parameters among climate models and reanalyses. This study focuses on annual and seasonal averages of surface temperatures for a collection of ESMs (an *ensemble*), the Coupled Model Intercomparison Project phase 5 (CMIP5) (Taylor et al., 2012), and three reanalysis data products, but the proposed approach can be also applied to finer temporal scales or to different physical quantities with nontrivial global patterns such as precipitation and wind. The new metric can be used to investigate if ESMs from the same institution display a similar space-time behavior, a feature widely observed in mean driven criteria (Tebaldi and Knutti, 2007; Knutti et al., 2010; Jun et al., 2008b,a; Masson and Knutti, 2011). This feature, caused by the inheritance of part of the code between different versions of the ESMs, will be investigated under this scenario-independent metric.

Climate model data are naturally evaluated over a sphere \times time domain, so a specific model for global data must be specified. Defining valid covariances on a spherical domain significantly restricts the flexibility of the models (Gneiting, 2013), and to date there is a significant dearth of literature in efficient and flexible models for global data. Jun and Stein (2007, 2008) proposed to generate a three-dimensional isotropic process and applying partial derivatives with respect to latitude and longitude to obtain processes that are longitudinally stationary (*axially symmetric*, see theoretical details in Hitczenko and Stein (2012); Huang

et al. (2012)), obtaining closed form expressions for the Matérn case.

More recently Castruccio and Stein (2013); Castruccio and Genton (2014, 2016); Castruccio and Guinness (2016) introduced a spectral based approach for axially symmetric processes on a regular grid. This methodology assumes a vector autoregressive process in time with axially symmetric innovations. The longitudinal dependence is formulated in the spectral domain with a Matérn-like covariance function, and a constant coherence across multiple bands. This model allows for efficient and parallelizable inference and is more flexible in reproducing local dependencies if global data are on a regular grid (Castruccio and Stein, 2013). In this work, we apply this spectral model for surface temperature, which is defined on a regular latitude \times longitude grid.

Section 2 introduces the multi-model ensemble and the reanalysis data that are used in this work. Section 3 illustrates the statistical setting for the ESMs that is assumed throughout the analysis. Section 4 describes the statistical model. Section 5 discusses the results of the comparison among different models in the ensemble and for reanalysis data. Section 6 draws some conclusions.

2 The CMIP5 ensemble and Reanalysis data

The CMIP5 ensemble (Taylor et al., 2012) is a global effort of many modeling groups to provide a set of coordinated experiments on climate sensitivity according to different ESMs. This ensemble comprises more than 60 ESMs run under different inputs. Here we focus only on the Representative Concentration Pathways (RCPs, Van Vuuren et al. (2011)), which comprise four possible scenarios (denoted as 2.6, 4.5, 6.0 and 8.5) of future greenhouse gases emissions. These represent four possible climate futures and are consistent with socio-economic assumptions regarding the anthropogenic emissions. RCP 2.6 assumes a peak of emissions in the current decade (2010-2020) followed by a decline, RCP 4.5 and 6.0 a peak

in 2040 and 2080 respectively, while RCP 8.5 assumes a continuous increase throughout the entire century. A plot of these four scenarios can be found in Figure S1 in the supplementary material.

We consider annual and seasonal (with respect to the northern hemisphere) surface temperature (at a standard height of 2 meters above ground level for all the CMIP5 data and Reanalysis 2. In Reanalysis 1 and CFSR the value at approximately 36 meters is given) from 18 models under the RCP scenarios (see Table 1). We select all the models with at least 2 realizations under at least one scenario of the RCP experiment family (FIO-ESM has been excluded because of numerical instabilities in the fit).

The data are all on a regular latitude \times longitude grid but the resolution varies substantially among models, and since the temperature is considered as an average over a grid cell, a straightforward comparison of the output is not possible without a common spatial scale. We discuss this issue in more detail in Section 5 and in the supplementary material. We discard the scenarios with a single realization in Table 1, since the statistical setting which is introduced in Section 3 requires at least two realizations for a given scenario. Therefore, we do not consider the 7 out of 210 runs where a scenario has a single realization. We believe this loss of information is acceptable for the purpose of this work.

Since one of the goals of this work is to compare climate models with observational data, we consider annual and seasonal temperatures from the NCEP/NCAR Reanalysis 1 (see Table 1). This data set consists of a continuously updated record of climatological variables from many sources from 1948 to present, filtered and assimilated throughout numerical weather models and interpolated on a regular grid, and it is considered a valid representative of the actual state of the system (Kalnay et al. (1996)). We also use the NCEP-DOE Reanalysis II (Kanamitsu et al. (2002)), a more recent version of reanalysis (from 1979 to present), which incorporates new system components, and NCEP Climate Forecast System

Reanalysis (CFSR; Saha et al. (2010), from 1979 to 2010). Also, since the main goal of this work is to analyze the spatio-temporal covariance structure, we remove the mean of the reanalysis data by fitting a smoothing spline in time for every pixel with a very mild penalty term of 0.01 (see supplementary material for details on the parametrization). We show how it is possible to estimate the space-time dependence via restricted likelihood without specifying the mean (Section 3), provided a scenario has at least two realizations. It is in principle possible to also use the runs with a single realization via detrending, but we choose to avoid the subjective choice of the penalty term whenever possible.

Table 1: Models analyzed and scenario/realization/years availability. Scenarios are from the RCP experiment family. Years for the ESMs start from 2006.

no.	Model name	Modeling center	resolution	RCP2.6	RCP4.5	RCP6.0	RCP8.5	years
1	CanESM2	CCCma	64×128	5	5	0	5	30
2	CanCM4	CCCma	64×128	0	10	0	0	95
3	CCSM4	NCAR	192×288	5	6	6	5	95
4	CESM1-CAM5	NSF-DOE-NCAR	192×288	3	3	3	3	95
5	CESM1-WACCM	NSF-DOE-NCAR	96×144	3	3	0	3	45
6	CNRM-CM5	CNRM-CERFACS	128×256	1	1	0	5	95
7	CSIRO-Mk3-6-0	CSIRO-QCCCE	96×192	10	10	10	10	95
8	EC-EARTH	EC-EARTH	160×310	1	3	0	4	95
9	GFDL-CM2.1	NOAA GFDL	90×144	0	10	0	0	35
10	GFDL-HIRAM-C180	NOAA GFDL	360×576	0	3	0	0	10
11	GISS-E2-H	NASA GISS	90×144	1	3	0	0	95
12	GISS-E2-R	NASA GISS	90×144	1	3	0	0	95
13	HadCM3	MOHC	73×96	4	4	4	4	30
14	HADgem2-ES	MOHC	145×192	4	4	4	4	95
15	IPSL-CM5A-LR	IPSL	96×96	4	4	1	0	95
16	MIROC4h	MIROC	320×640	0	3	0	0	30
17	MIROC5	MIROC	128×256	3	3	1	3	95
18	MPI-ESM-LR	MPI-M	96×192	3	3	0	3	95
	Reanalysis 1	NCEP/NCAR	73×144	-	-	-	-	66
	Reanalysis 2	NCEP/DOE	94×192	-	-	-	-	34
	CFSR	NCEP	73×144	-	-	-	-	32

3 The statistical setting for the ESMs

Here we present the statistical setting that we use to model variability across the mean climate of the ESM without detrending, under the assumption that a scenario comprises at least two realizations. This setting does not apply to reanalysis data since there is only one realization: in that case a likelihood of the model in Section 4 is fit to the detrended data. The notation and the results in this section were first introduced by Castruccio and Stein (2013), and here we report some of them for completeness.

A key assumption of climate model runs is the statistical independence across different realizations. Every realization depends on initial conditions sampled from a spin-up run and in a short period of time the evolution of the model is effectively independent of initial condition due to its chaotic nature (Lorenz, 1963; Collins and Allen, 2002; Collins, 2002; Branstator and Teng, 2010). We denote with L_m for $m = 1, \dots, M$ the latitude, with ℓ_n for $n = 1, \dots, N$ the longitude, with $k = 1, \dots, T$ the time. Also, denote with $g = 1, \dots, G$ the ESM, with $s = 1, \dots, S_g$ the scenario and with $r = 1, \dots, R_{g,s}$ the realization.

We assume the following model

$$\mathbf{T}_{g,s,r} = \boldsymbol{\mu}_{g,s} + \boldsymbol{\varepsilon}_{g,s,r} \quad \boldsymbol{\varepsilon}_{g,s,r} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma}_g), \quad (1)$$

where

$$\mathbf{T}_{g,s,r} = \{\mathbf{T}_{g,s,r}(L_1, \ell_1, t_1), \dots, \mathbf{T}_{g,s,r}(L_M, \ell_1, t_1), \mathbf{T}_{g,s,r}(L_1, \ell_2, t_1), \dots, \mathbf{T}_{g,s,r}(L_M, \ell_N, t_T)\}$$

is the vector of temperature, $\mathbb{E}(\mathbf{T}_{g,s,r}) = \boldsymbol{\mu}_{g,s}$ is the mean across realizations, and $\boldsymbol{\varepsilon}_{g,s,r}$ is the random component of the statistical model, which is assumed to be normally distributed with covariance structure $\boldsymbol{\Sigma}_g$, independent across scenarios and realizations. The assumption that the spatio-temporal covariance structure does not depend on the scenario will be discussed more into detail in Section 4.3.

3.1 REML estimation of $\boldsymbol{\varepsilon}_{g,s,r}$

Suppose that $\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}(\boldsymbol{\theta}_g)$, that is the covariance structure depends on a vector of parameters $\boldsymbol{\theta}_g$. Also, denote by $\bar{\mathbf{T}}_{g,s}$ the average across realizations, $\mathbf{D}_{g,s} = (\mathbf{T}_{g,s,1} - \bar{\mathbf{T}}_{g,s}, \dots, \mathbf{T}_{g,s,R_s} - \bar{\mathbf{T}}_{g,s})$ and $\mathbf{D}_g = (\mathbf{D}_{g,1}, \dots, \mathbf{D}_{g,S_g})$.

We have

$$\mathbf{T}_{g,s} = (\mathbf{1}_{R_{g,s}} \otimes \mathbf{I}_{TNM}) (\mathbf{1}_{R_{g,s}} \otimes \boldsymbol{\mu}_{g,s}) + \boldsymbol{\varepsilon}_{g,s} \quad \boldsymbol{\varepsilon}_{g,s} \sim \mathcal{N}(0, \mathbf{I}_{R_{g,s}} \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta}_g)), \quad (2)$$

where $\mathbf{I}_{R_{g,s}}$ is the identity matrix of size $R_{g,s}$ and $\mathbf{1}_{R_{g,s}}$ is a column vector of size $R_{g,s}$ with all entries equal to 1. Since we have independence across scenarios and realizations for the stochastic component of the statistical model, and since realizations with the same scenario share the same mean, it is intuitive to estimate $\boldsymbol{\theta}_g$ by considering the difference of the realizations with the same scenarios, so to cancel the effect of the mean vector. The following result formalizes this:

Result 1 *The restricted loglikelihood for (2) is $l(\boldsymbol{\theta}_g; \mathbf{D}_g) = \sum_{s=1}^{S_g} l_{g,s}(\boldsymbol{\theta}_g; \mathbf{D}_{g,s})$ where*

$$\begin{aligned} l_{g,s}(\boldsymbol{\theta}_g; \mathbf{D}_{g,s}) &= -\frac{TNM(R_{g,s}-1)}{2} \log(2\pi) - \frac{1}{2}(R_{g,s} - 1) \log[\det\{\boldsymbol{\Sigma}(\boldsymbol{\theta}_g)\}] \\ &\quad - \frac{1}{2} TNM \log(R_{g,s}) - \frac{1}{2} \mathbf{D}_{\mathbf{g},s}' \{ \mathbf{I}_{R_{g,s}} \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta}_g) \}^{-1} \mathbf{D}_{\mathbf{g},s}. \end{aligned} \quad (3)$$

Also, the corresponding estimator for $\boldsymbol{\mu}_{g,s}$ obtained by generalized least squares is $\hat{\boldsymbol{\mu}}_{g,s} = \bar{\mathbf{T}}_{g,s}$.

The proof is a straightforward generalization of Result 1 in Castruccio and Stein (2013). In this work, this quantity will be used to obtain an estimator of the covariance structure without specifying any model for $\boldsymbol{\mu}_{g,s}$.

4 A global space-time statistical model

Consider now $g = 1, \dots, G, G+1, G+2, G+3$, where the last three indices denote the reanalysis data. The ultimate goal of this work is to estimate $\boldsymbol{\theta}_g$ and to determine how this vector is changing with respect to g . The model is briefly described below.

4.1 Temporal component

We assume that the stochastic component of the statistical model is AR(1) in time, with different parameters for land and ocean. More specifically, if we call $\boldsymbol{\varepsilon}_t = \{\varepsilon(L_1, \ell_1, t), \dots, \varepsilon(L_N, \ell_M, t)\}$ the stochastic component at time t , we assume it has the following structure

$$\begin{aligned}\boldsymbol{\varepsilon}_t &= \boldsymbol{\Phi}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{C}), \\ \boldsymbol{\Phi} &= \text{diag}\{\varphi_{Q(L_1, \ell_1)}, \dots, \varphi_{Q(L_M, \ell_N)}\},\end{aligned}\tag{4}$$

where \mathbf{C} is the spatial covariance matrix, $Q(L, \ell) = 1$ if pixel (L, ℓ) is land (If the grid point is on the boundary we will consider it as land if its percentage is greater than 50%.) and $Q(L, \ell) = 0$ ocean.

4.2 Spatial component

In order to describe \mathbf{C} a spatial model for data on a sphere is needed, and for many applications such as the one in this work, an isotropic model is not a reasonable assumption. The main reason is that temperature at different latitudes is expected to have a different structure (e.g. we would expect different variances). Therefore, we work with the following class of processes:

Definition 1 *A Gaussian process Z on a sphere is axially symmetric (Jones, 1963) if the mean depends on latitude and*

$$\text{cov}\{Z(L_1, \ell_1), Z(L_2, \ell_2)\} = K(L_1, L_2, \ell_1 - \ell_2).\tag{5}$$

Jun and Stein (2007, 2008) described a flexible approach to generate axially symmetric processes by embedding the sphere in a three dimensional space, generating an isotropic process and applying partial derivatives to induce nonstationarity across latitudes. More recently, Castruccio and Stein (2013); Castruccio and Genton (2014, 2016); Castruccio and Guinness (2016) proposed a spectral based approach that proved more natural than the

previous methods for data on a regular grid. Since the geometry in this context is the same, we use this approach. The next two sections are devoted to explaining this method.

4.2.1 Longitudinal structure

If the process has covariance as in (5), then across a single latitude it is stationary. Therefore, since the data are evenly spaced on a circle, the covariance matrix is exactly circulant (Davis, 1979) and it is natural to model the process in the spectral domain. More precisely, we denote $\ell_n = 2\pi n/N$ and

$$f_L(c) = \sum_{n=1}^N e^{-ic\ell_n} K(L, L, \ell_n) \quad (6)$$

the spectral density of the process at (integer) wavenumber c . We choose the following parametrization for (6):

$$f_L(c; \phi_L, \alpha_L, \nu_L) = \frac{\phi_L}{\{\alpha_L^2 + 4 \sin^2(\frac{c}{N}\pi)\}^{\nu_L+1/2}}, \quad c = 0, \dots, N-1. \quad (7)$$

This functional form is similar to the Matérn class, but it avoids a loss in regularity for $c = \lfloor \frac{N}{2} \rfloor$ (Castruccio and Stein, 2013; Poppick and Stein, 2014).

The noticeable computational advantage of a spectral approach to axially symmetric processes is that every latitudinal band can be analyzed separately. Therefore, in our analysis, we divide the data set in latitudinal bands and estimate ϕ_L , α_L and ν_L using separate processors on a cluster. This process requires only a few minutes on a 16-cores workstation.

4.2.2 Latitudinal structure

Once the longitudinal parameters are estimated, we need to define a model for the latitudinal dependence. In other words, we need to define

$$f_{L_m, L_{m'}}(c) = \sum_{n=1}^N e^{-ic\ell_n} K(L_m, L_{m'}, \ell_n) \in \mathbb{C}, \quad c = 0, \dots, N-1.$$

If the process is axially symmetric, the covariance for different latitudes must be specified only for the same wavenumber, since otherwise is zero (Jun and Stein, 2008). This allows

the covariance matrix in the spectral domain to be block diagonal. Therefore we need to model the coherence

$$\rho_{L_m, L_{m'}}(c) = \frac{|f_{L_m, L_{m'}}(c)|}{\sqrt{f_{L_m}(c)f_{L_{m'}}(c)}}$$

and the phase. We assume the phase to be zero, a valid assumption with temperature data at this time resolution (see Castruccio and Genton (2014) for a complete discussion).

The coherence is assumed to depend on the latitudinal lag, but not on the particular latitude chosen:

$$\rho_{L_m, L_{m'}}(c) = \left[\frac{\xi}{\left\{1 + 4 \sin^2 \left(\frac{c}{N} \pi\right)\right\}^\tau} \right]^{|m-m'|}, \quad c = 0, \dots, N-1, \quad (8)$$

where ξ is a parameter that modulates the decay of the coherence across latitudes and τ modulates its decay across wavenumbers. This two step estimation procedure does not optimize the likelihood over the full parameter space, but it is computationally scalable and its local maximum is close to the global maximum (see (Castruccio and Genton, 2016) for a complete diagnostic and discussion on error and bias propagation).

4.3 Diagnostics

The model described in the previous sections relies on some assumptions about the ensemble. Firstly, with reference to the notation in Section 3.1, we assumed that Σ_g does not depend on the scenario. In other words, in order to pool the estimates from different scenarios, we assume that there is no change across them for a fixed ESM. A complete diagnostic would require a separate analysis for each of the 18 ESMs used in this work; here we present the results for CCSM4 but similar patterns were also observed for the other models. In Figures 1a-d we see how $\hat{\phi}_L$, $\hat{\alpha}_L$ and $\hat{\nu}_L$ and $\hat{\varphi}_{0,L}$ ($\hat{\varphi}_{1,L}$ is not shown but displays similar results) change across scenarios and in Table 2 we see the results for $\hat{\xi}$, $\hat{\tau}$, $\hat{\varphi}_0$ and $\hat{\varphi}_1$. There is evidence that the longitudinal parameters do not change across scenarios for most latitudes, as they fall within the 95% Bonferroni band. The latitude interval between -50° and -60°

seems to be more scenario dependent, as observed in Castruccio et al. (2014). For the latitudinal parameters, $\hat{\xi}$, $\hat{\varphi}_0$ do not change, $\hat{\varphi}_1$ show only a mild evidence of change between scenarios RCP2.6 and 8.5, and RCP2.6 and 4.5, and $\hat{\tau}$ show significant differences across all scenarios. Nevertheless, the differences occur in the second or (more often) third significant digit. We choose to ignore these differences in longitudinal and latitudinal parameters since such changes are small compared to the other factors involved in this analysis (Table 3).

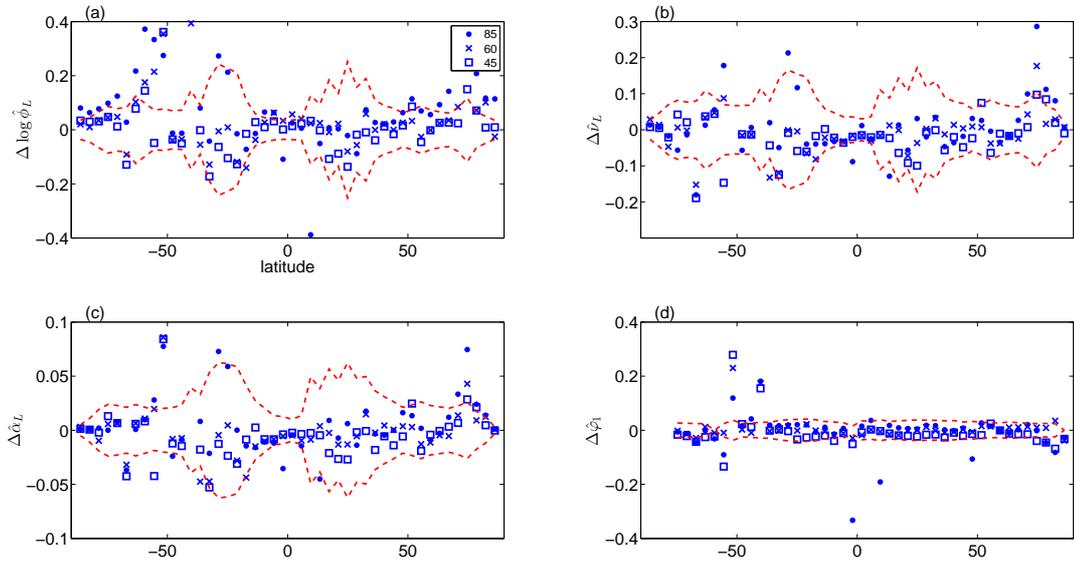


Figure 1: Comparison of the change in estimated parameters for all scenarios of CCSM4. The baseline estimated parameters are from RCP2.6, and the dotted red lines are the 95% Bonferroni confidence bands.

Table 2: CCSM4 95% confidence intervals for the difference in latitudinal parameter estimates for different scenarios. Reference is RCP26. All numbers are multiplied by 10^2 .

parameter	RCP45	RCP60	RCP85
$\Delta\xi$	(-0.13, 0.10)	(-0.17, 0.05)	(-0.03, 0.21)
$\Delta\tau$	(0.51, 1.72)	(0.19, 1.39)	(0.06, 1.33)
$\Delta\varphi_1$	(-1.28, -0.88)	(-0.04, 0.36)	(0.05, 0.47)
$\Delta\varphi_0$	(-1.27, -0.87)	(-0.03, 0.37)	(0.06, 0.48)

Another model assumption is that the temporal component has an AR(1) structure. To evaluate this assumption, for every model we consider $\mathbf{D}_1 = \mathbf{T}_1 - \bar{\mathbf{T}}$ of an RCP85 scenario and fit AR(2) model separately for every grid point:

$$\mathbf{D}_1(t) = \psi_1 \mathbf{D}_1(t-1) + \psi_2 \mathbf{D}_1(t-2) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2). \quad (9)$$

We then test if $\psi_1 = 0$ or if $\psi_2 = 0$, and plot their p-values for every grid point. The results are in Figure 2 and 3. Across a subset of 6 models chosen there is evidence that $\psi_1 \neq 0$ but this is not true for Reanalysis 1, Reanalysis 2 and CFSR. Figures 2 and 3 show the p-values for $\psi_1 = 0$ of an AR(1) process and $\psi_2 = 0$ of an AR(2) process, respectively. Most of the models and the reanalysis data show some evidence of second order temporal structure in the tropical region, likely due to El Niño Southern Oscillation, but its spatial distribution differs across models. In this work, we will make the assumption of AR(1) structure, and we do not describe the spatial distribution of ψ_2 , to avoid specifying different statistical models for different ESMs and to have results which are as homogeneous as possible. If the focus is on a single ESM with a clear indication of a higher order temporal dependence, then a stronger temporal dependence can be accounted for in the statistical model (Castruccio and Genton, 2014).

5 Results

As mentioned in Section 2, a direct comparison of different ESMs and reanalysis data is problematic, as different models have different spatial resolutions. Since the grid values are averages over a cell, we expect variances to change with the size of the grid. Therefore, since in (7) ϕ depends on the variance, its interpretation will change according to the resolution. Similarly, α and ν are dependent on N , the number of points per latitudinal band.

Even if some models share the same grid, for a uniform assessment of their features, an upscaling to a common grid is needed (see supplementary material for details about

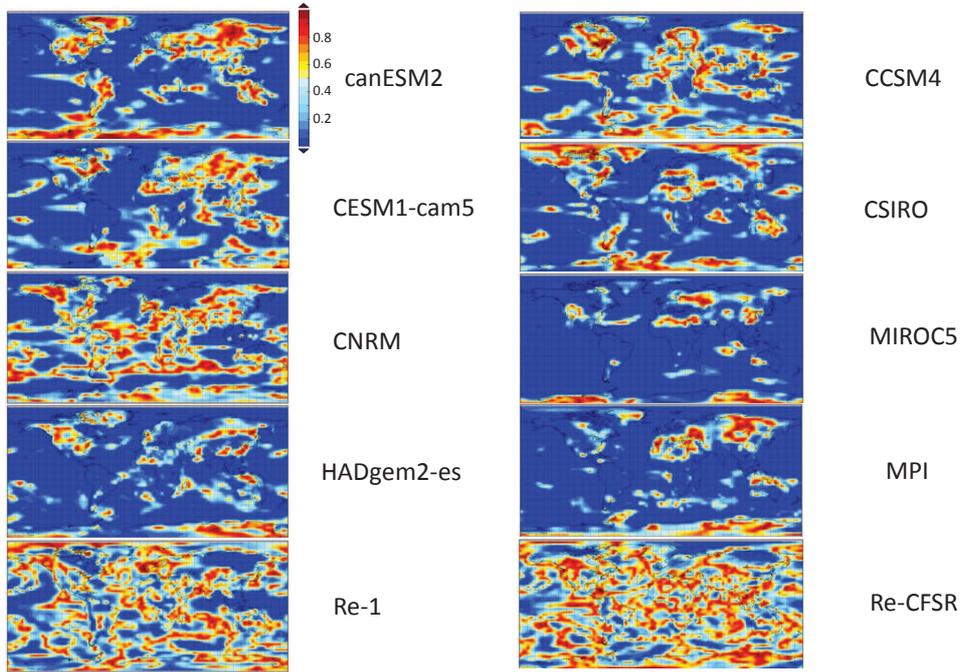


Figure 2: Diagnostic of the temporal structure of the data set for the CMIP5 ensemble: p-values of the significance of a test for $\psi_1 = 0$.

the algorithm). We choose a coarse 48×96 resolution grid with equally spaced latitudinal bands, corresponding to approximately a T31 ($3.75^\circ \times 3.75^\circ$) resolution. In Sections 5.1, 5.2 and 5.3 we compare all ESMs and reanalyses with respect to their temporal, longitudinal and latitudinal structure respectively. In Section 5.4 we show an example of a comparison without upscaling for two models at the same resolution.

5.1 Temporal dependence

In Figure 4 we present a comparison of the ensemble with reanalysis data in terms of the estimated temporal parameters $\hat{\varphi}_0$ and $\hat{\varphi}_1$ (land and ocean autoregressive parameters in (4)). There seems not to be a noticeable difference in the land and ocean temporal component, uniformly across models. Both autoregressive parameters for all ESMs are smaller in absolute

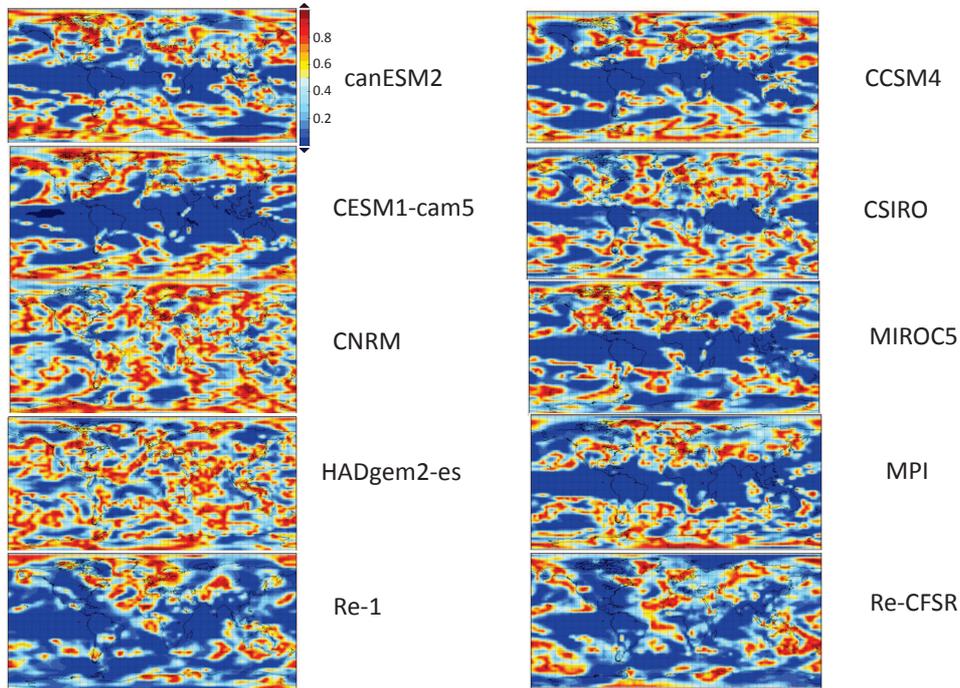


Figure 3: Diagnostic of the temporal structure of the data set for the CMIP5 ensemble: p-values of the significance of a test for $\psi_2 = 0$.

value than the ones for reanalysis data, indicating a weaker temporal dependence in the simulated data, uniformly across seasons. Further, all reanalyses display a negative lag one correlation, while for all the models this is positive.

CCSM4, CESM1-CAM5 and to some extent CESM1-WACCM (models 3, 4 and 5) are ESM from the same modeling group (NCAR) and have similar estimates. Since CESM1 is an evolution of CCSM4, the observed similarities are likely due to the inheritance of some parts of the schemes between the models. A similar feature can be observed for the two CCCma models (models 1 and 2), and the NOAA GFDL models (9 and 10), which also share the same native grid. Similarity among models from the same group has been reported frequently in literature (Tebaldi and Knutti, 2007; Knutti et al., 2010; Jun et al., 2008b,a; Masson and Knutti, 2011), but never in the context of space-time dependence. Instead, the models with

different native grid (13 and 14 for MOHC, 11 and 12 for NASA GISS and most noticeably 16 and 17 the MIROC) display very different parameters. All these remarks are consistent across seasons (figure 4) and for annual averages as well (plot not shown).

The estimated uncertainty of φ_0 and φ_1 in (4) is very small for all models but GFDL-HIRAM-C180, where only 10 years and 3 realizations were available. Reanalysis 1 and especially Reanalysis 2 and CFSR have wider confidence intervals, since there is only one realization and considerably fewer time points than most of the ESMs.

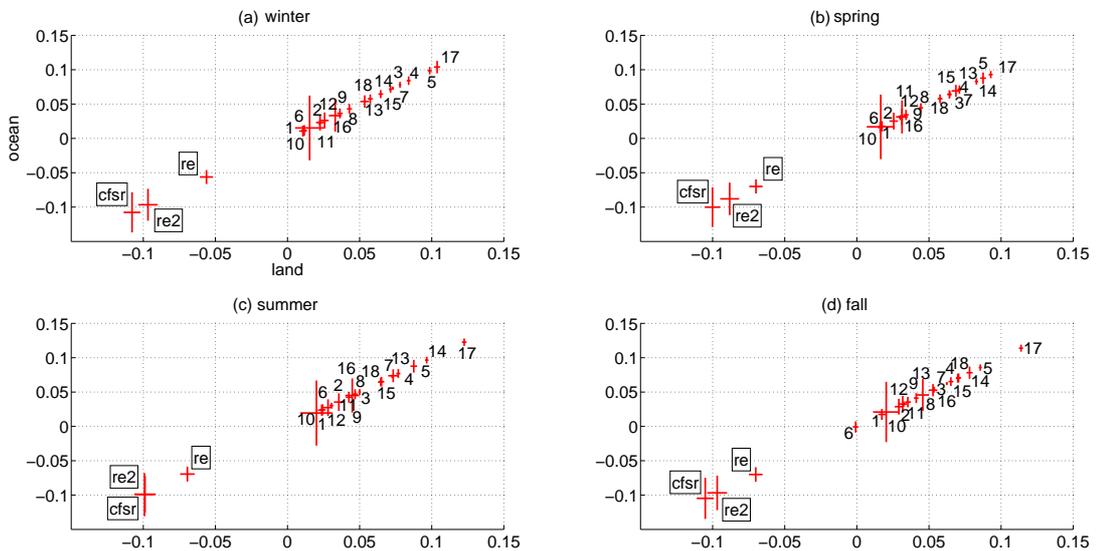


Figure 4: Comparison of the temporal component of the models and reanalysis in Table 1: φ_0 vs φ_1 as in (4) plotted for all the models and the reanalysis data (in square) for the four seasons. The number code of the climate model is with reference to Table 1 and the red lines represent the 99.99% confidence intervals.

5.2 Longitudinal dependence

To compare the spatial structure among multiple models, we first consider CCSM4 as a reference and compute the L^1 distance of the estimated parameters with respect to this model across latitudes, and the L^1 distance of the estimated spectra (across wavenumbers) averaged across latitudes. The results in Table 3 show that for the inverse range $\hat{\alpha}_L$ and

most importantly for f_L , the two closest models are CESM1-CAM5 and CESM1-WACCM, developed from the same modeling group. For $\log(\hat{\phi}_L)$ and $\hat{\nu}_L$ the closest model is CESM1-CAM5, but the second closest is not the third NCAR model, albeit the distance is still among the shortest. According to the same metric, it can also be shown how the CCCma models (1 and 2), the NASA GISS (11 and 12), the MOHC (13 and 14) display a very similar spatial structure, but for NOAA GFDL (9 and 10) and the MIROC (16 and 17) this is not the case. The patterns observed in Table 3 are also consistent across seasons (tables are not shown). As in Section 5.1, the pairs of models from the same modeling group that show the largest dissimilarities are the ones that do not share the same native grid. If Reanalysis 2 is considered as a reference, HADgem2-ES is among the closest two models according to all the four metrics, MPI-ESM-LR has similar parameters but CNRM-CM5 has a closer spectra (and therefore a similar spatial structure). It must be noticed however that the distance among reanalysis data product (especially Reanalysis 1) is larger than the distance between Reanalysis 2 and the models in the ensemble. This implies that different assimilation algorithms in reanalysis drastically change the longitudinal structure of the variability across the climatological mean.

5.3 Latitudinal dependence

In Figure 5, a comparison of the models with respect to the latitudinal parameters in (8) is shown. As noticed in Section 5.2, Reanalysis 1 shows a remarkably different behavior from Reanalysis 2 and CFSR, as the latter have estimated parameters closer to the ones in the model ensemble. Reanalysis 1 instead shows a spatial structure which is less regular, both with respect to latitudinal lag and wavenumbers. The two models from CCCma, labeled 1 and 2 are almost indistinguishable, despite the very small parameter uncertainty. The same is observed between the two NASA GISS (11 and 12) models and between CCSM4 (3) and

Table 3: L^1 distance of all models with respect to the CCSM4 estimates (columns 2-5) and Reanalysis 2 (columns 6-9) of the longitudinal parameters and corresponding spectra. Annual temperatures are considered and the distance is computed by averaging across latitudes. For the spectra, the distance is computed across wavenumbers and then averaged across latitude. In bold the two closest models. Similar patterns are evident for seasonal averages.

no	model	$\Delta_1 \log(\hat{\phi}_L)$	$\Delta_1 \hat{\alpha}_L$	$\Delta_1 \hat{\nu}_L$	$\overline{\Delta_1 f_L}$	$\Delta_2 \log(\hat{\phi}_L)$	$\Delta_2 \hat{\alpha}_L$	$\Delta_2 \hat{\nu}_L$	$\overline{\Delta_2 f_L}$
1	CanESM2	53.08	8.84	31.74	11.03	41.81	7.55	31.95	10.68
2	CanCM4	51.15	8.49	30.51	8.62	47.93	8.15	35.84	7.90
3	CCSM4	0.00	0.00	0.00	0.00	42.28	8.79	31.24	11.31
4	CESM1-CAM5	12.59	1.49	6.50	5.46	39.89	9.35	34.12	11.09
5	CESM1-WACCM	15.86	3.06	15.95	5.93	40.90	7.25	28.52	12.60
6	CNRM-CM5	17.69	4.43	17.93	9.85	41.43	7.79	27.48	6.65
7	CSIRO-Mk3-6-0	64.67	10.94	47.63	12.22	49.55	7.39	36.11	7.95
8	EC-EARTH	20.03	5.12	23.38	8.51	38.11	7.30	31.38	6.92
9	GFDL-CM2.1	30.39	5.35	27.84	9.33	39.05	6.32	26.52	11.18
10	GFDL-HIRAM-C180	40.08	11.75	20.44	15.07	67.84	14.12	41.45	11.16
11	GISS-E2-H	30.93	4.92	21.41	11.55	34.95	6.50	33.34	8.14
12	GISS-E2-R	32.16	5.11	20.35	11.72	36.23	6.44	32.43	8.88
13	HadCM3	54.99	6.88	38.53	14.44	41.28	15.33	63.45	16.55
14	HADgem2-ES	29.18	4.95	14.88	10.71	26.71	6.07	26.25	6.30
15	IPSL-CM5A-LR	14.72	4.55	17.46	7.44	42.78	11.57	41.12	9.72
16	MIROC4h	20.71	4.42	15.83	10.98	50.23	10.54	38.86	7.05
17	MIROC5	105.45	11.59	68.50	12.81	120.52	14.80	74.77	10.92
18	MPI-ESM-LR	50.81	7.89	26.90	10.33	35.40	5.00	24.05	9.41
	Reanalysis 1	552.96	29.59	285.80	14.38	549.53	26.67	270.34	10.80
	Reanalysis 2	42.28	8.79	31.24	11.31	0.00	0.00	0.00	0.00
	CFSR	17.04	3.95	18.88	9.52	43.70	10.15	39.82	8.02

CESM1-CAM5 (4), but not when compared with the CESM1-WACCM (5), a model that was not run at the same grid resolution. Similarly to the previous comparisons, models with different native grid such as MOHC (13 and 14), MIROC (16 and 17) and also NOAA GFDL (9 and 10) have significantly different parameter estimates.

According to the proposed metric, we observed a strong degree of similarity among ESMs from the same group with common native grid, while different models on the same grid have significantly different statistical behavior. We offer two possible explanations. If, as observed in previous literature for the climatological trend (Tebaldi and Knutti, 2007; Knutti et al., 2010; Jun et al., 2008b,a; Masson and Knutti, 2011), the similarity should be more evident also for models on a different grid, then these results are dependent upon the upscaling algorithm and a simple areal average could alter some of the statistical properties. If instead these differences are not caused by the upscaling algorithm, then the patterns of similarity among some ESMs in terms of their spatial structure are not always as strong as reported.

It is also noticeable how the differences in parameter estimates between an ensemble member and a reanalysis are often smaller than the differences among two reanalyses, as in the case of the longitudinal parameters. This is another indication that different data assimilation algorithms have a strong impact on the space-time structure of the variability across the climatological trend.

The estimation uncertainty is small for the ESMs and is considerably larger for the reanalysis data, and as in Section 5.1 this is because there is only one realization and a shorter temporal scale.

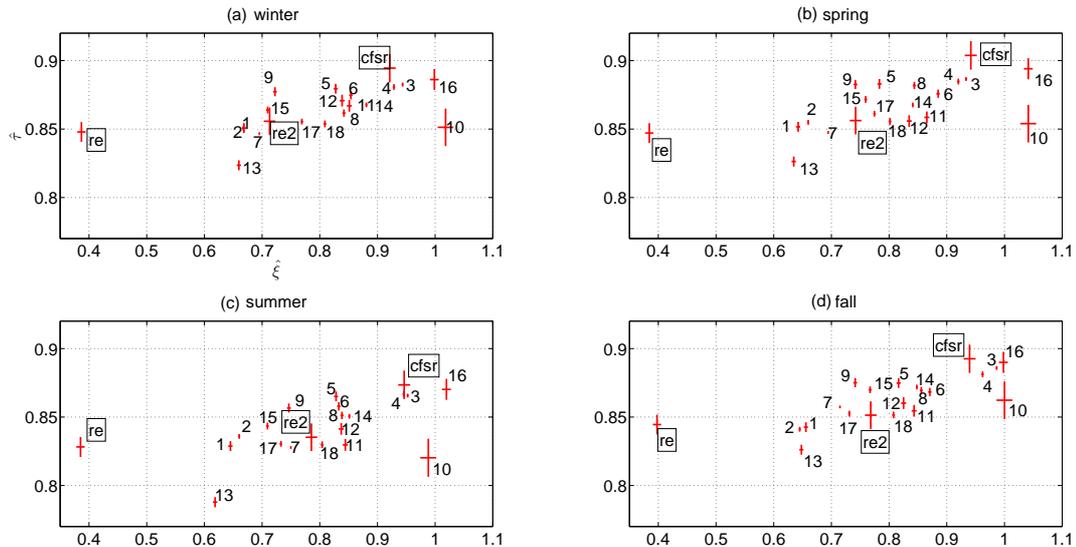


Figure 5: Latitudinal parameter comparison of the models in Table 1 with reanalysis: $\hat{\xi}$ vs $\hat{\tau}$ as in (8) plotted for all the models and the reanalysis data (in square) for all four seasons. The number code of the climate model is with reference to Table 1 and the red lines represent the 99.99% confidence intervals.

5.4 Comparison at higher resolution for CCSM4 and CESM1-CAM5

As mentioned in the last Section, the upscaling resolution plays a role in the data interpretation. Therefore, whenever some models share a common grid, it is preferable to compare them without any upscaling. This is the case, for example, of the CCSM4 and CESM1-CAM5, which are on a common 190×288 grid (the real resolution was 192×288 , but the first and last band were removed since they corresponded to latitudes -90° and 90°). In Figure 6 we can see such a comparison for annual averages (seasonal averages show similar results) in terms of the single latitudinal parameters and in Table 4 for the latitudinal parameters. The similarity of the parameter profiles in Figure 6 indicates a strong similarity between the two ESMS, and it is not an artifact of higher resolution. Indeed, a similar comparison can be done with CSIRO and MPI, which share the same grid, but have noticeably

different parameter estimates across latitudes (see supplementary material). The latitudinal parameters in Table 4 are estimated with very high precision: the confidence intervals for ξ , φ_0 and φ_1 have the same three decimal digits in the upper and lower bound, while τ has two.

Besides its interest in the value of the parameters, this analysis shows how the spectral model introduced in this work is scalable and can fit in a reasonable time models with a state-of-the-art resolution. In this case, the CCSM4 data set consists of approximately 114 million data and the latitudinal parameters with their uncertainties were obtained in less than two days, and the CESM1-CAM5 data set, with 62 million data, in approximately the same time. The longitudinal analysis requires 190 optimizations that are done in parallel on a 16-cores workstation and require less than one hour.

Table 4: 95% confidence intervals for the latitudinal parameter estimates for CCSM4 and CESM1-CAM5 at high resolution.

parameter	CCSM4	CESM1-CAM5
ξ	(0.9740, 0.9744)	(0.9733, 0.9739)
τ	(1.0880, 1.0883)	(1.1278, 1.1284)
φ_1	(0.2721, 0.2722)	(0.2909, 0.2910)
φ_0	(0.2716, 0.2727)	(0.2901, 0.2918)

6 Concluding remarks

In this work we propose a new methodology for comparing climate models and reanalysis data products. This approach focuses on the space-time variability around the climatological mean for different ESMs, and between ESMs and reanalysis. The metric developed is scenario independent, and thus overcomes major limitations associated with existing approaches related to the climatological mean, which is scenario dependent. Climate model

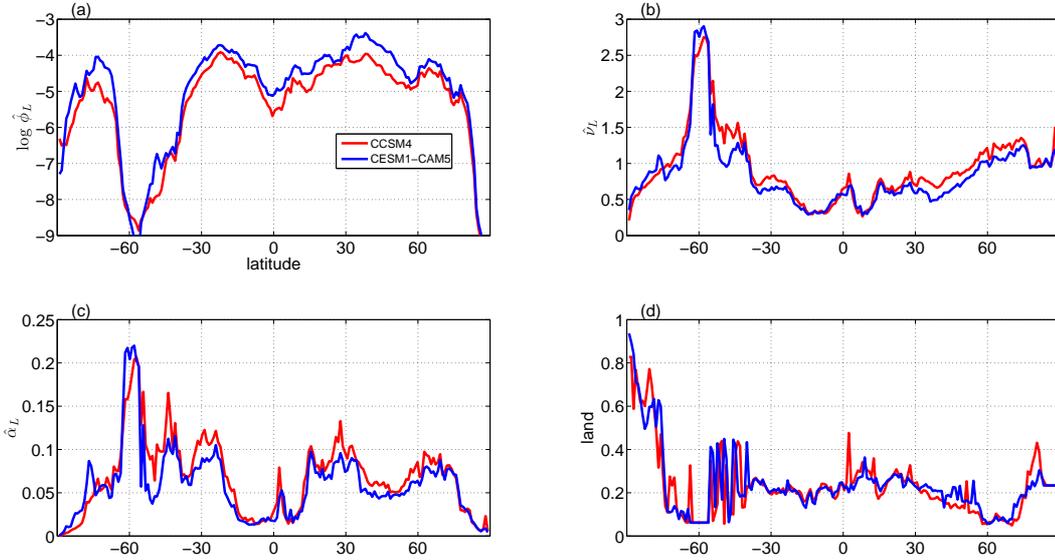


Figure 6: Longitudinal parameter comparison at high resolution of CCSM4 and CESM1-CAM5. With reference to (7), a plot of (a) $\log(\hat{\phi}_L)$, (b) \hat{v}_L , (c) $\hat{\alpha}_L$ and (d) $\hat{\phi}_{1;L}$ for different latitudes is shown.

output shows a weaker temporal correlation than all three reanalysis data products, and the sign of temporal dependence is reversed. For some latitudes, Reanalysis 1 shows considerably different results with respect to the CMIP5 models and the other reanalyses. Reanalysis 2, CFSR and the CMIP5 models instead have a similar structure, but the difference between reanalyses is larger than the difference with the ensemble members, and this has been observed for both annual and seasonal temperatures. This suggests that while different assimilation algorithms could approximately reproduce the same climatological mean, they significantly modify the spatio-temporal dependence of the observations.

The proposed methodology further shows interesting patterns of dependence across some models developed by the same institution, although this is true almost exclusively for climate models sharing the same native grid. This is an indication that models at different resolutions do also display a different space-time behavior, although an alternative explanation could be that the upscaling algorithm is significantly impacting on the space-time properties of the

ESMs. While the CMIP5 provides a wide array of simulations for a variety of investigations it is not, unfortunately, suitable to conclusively answer this question. Indeed, to explain if the spatio-temporal structure changes are due to upscaling or to an actual change in the physics of the model, an ensemble of runs from the same climate model at different nested resolutions is needed.

The choice of analyzing and comparing different climate models based on the variability across the climatological mean instead of the mean itself is a novel perspective in the context of comparing climate models and, although proposed before, it was never implemented. The methodology introduced here implies the definition of a space-time statistical model, and consequently relies on some assumptions. The simplification of the complex climate model output dynamics to a simple class of statistical models (uniformly in the ensemble and reanalyses) is an assumption that allows to summarize the features of an ESM with a small number of statistical parameters. While a statistical model hinges on assumptions such as that of a stationary Gaussian process, which is not a perfect description of a climate model output, it has been shown in this work how it can reveal some important differences among ESMs, and among ESMs and Reanalysis.

Since the proposed approach does not compare temperature trends, a physical interpretation of the results is more challenging compared to the existing literature, which compares temperature fields of CMIP5 and Reanalysis, observational data, and CMIP3 (Liu et al., 2012; Koichi et al., 2012; Xu et al., 2013; Yao et al., 2013; Ye-Won et al., 2014). Nevertheless, this metric is more flexible as it allows to estimate and compare the space-time structure among climate models run under different scenarios.

Acknowledgements

This work is supported by STATMOS, an NSF funded Network (NSF-DMS awards 1106862, 1106974 and 1107046).

I acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP (<http://cmip-pcmdi.llnl.gov/cmip5/>), and I thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

References

- Bender, F. (2008), "A Note on the Effect of GCM Tuning on Climate Sensitivity," *Environmental Research Letters*, 3, 014001.
- Branstator, G. and Teng, H. (2010), "Two Limits of Initial-Value Decadal Predictability in a GCGM," *Journal of Climate*, 23, 6292–6311.
- Braverman, A., Cressie, N., and Teixeira, J. (2011), "A Likelihood-Based Comparison of Temporal Models for Physical Processes," *Statistical Analysis and Data Mining*, 4, 247–258.
- Buser, C., Künsch, H., Lüthi, D., Wild, M., and Schär, C. (2009), "Bayesian Multi-model Projection of Climate: Bias Assumptions and Interannual Variability," *Climate Dynamics*, 33, 849–868.
- Castruccio, S. and Genton, M. G. (2014), "Beyond Axial Symmetry: An Improved Class of Models for Global Data," *Stat*, 3, 48–55.

- (2016), “Compressing an Ensemble with Statistical Models: An Algorithm for Global 3D Spatio-Temporal Temperature,” *Technometrics*, in press.
- Castruccio, S. and Guinness, J. (2016), “An Evolutionary Spectrum Approach to Incorporate Large-scale Geographical Descriptors on Global Processes,” ArXiv:1507.03401v1.
- Castruccio, S., McInerney, D. J., Stein, M. L., Liu, F., Jacob, R. J., and Moyer, E. J. (2014), “Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs,” *Journal of Climate*, 27, 1829–1844.
- Castruccio, S. and Stein, M. L. (2013), “Global Space-Time Models for Climate Ensembles,” *Annals of Applied Statistics*, 7, 1593–1611.
- Collins, M. (2002), “Climate Predictability on Interannual to Decadal Time Scales: The Initial Value Problem,” *Climate Dynamics*, 19, 671–692.
- Collins, M. and Allen, M. R. (2002), “Assessing the Relative Roles of Initial and Boundary Conditions in Interannual to Decadal Climate Predictability,” *Journal of Climate*, 15, 3104–3109.
- Davis, P. (1979), *Circulant Matrices*, New York: Wiley.
- Furrer, R., Knutti, R., Sain, S. R., Nychka, D. W., and Meehl, G. A. (2007), “Spatial Patterns of Probabilistic Temperature Change Projections from a Multivariate Bayesian Analysis,” *Geophysical Research Letters*, 34.
- Gneiting, T. (2013), “Strictly and Non-Strictly Positive Definite Functions on Spheres,” *Bernoulli*, 19, 1327–1349.
- Hitczenko, M. and Stein, M. (2012), “Some theory for anisotropic processes on the sphere,” *Statistical Methodology*, 9, 211 – 227, special Issue on Astrostatistics + Special Issue on Spatial Statistics.

- Huang, C., Zhang, H., and Robeson, S. M. (2012), “A simplified representation of the covariance structure of axially symmetric processes on the sphere,” *Statistics and Probability Letters*, 82, 1346–1351.
- Jackson, C. S., Sen, M. K., G., H., Deng, Y., and Bowman, K. (2008), “Error Reduction and Convergence in Climate Prediction,” *Journal of Climate*, 21, 6698–6709.
- Jones, R. (1963), “Stochastic Processes on a Sphere,” *The Annals of Mathematical Statistics*, 34, 213–218.
- Jun, M., Knutti, R., and Nychka, D. (2008a), “Local Eigenvalue Analysis of CMIP3 Climate Model Errors,” *Tellus A*, 60, 992–1000.
- (2008b), “Spatial Analysis to Quantify Numerical Model Bias and Dependence: How Many Climate Models Are There?” *Journal of the American Statistical Association*, 103, 934–947.
- Jun, M. and Stein, M. (2007), “An Approach to Producing Space-Time Covariance Functions on Spheres,” *Technometrics*, 49, 468–479.
- (2008), “Nonstationary Covariance Models for Global Data,” *Annals of Applied Statistics*, 2, 1271–1289.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., R., J., and Joseph, D. (1996), “The NCEP/NCAR 40-year Reanalysis Project,” *Bull. Amer. Meteor. Soc.*, 77, 437–470.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J. J., Fiorino, M., and Potter, G. L. (2002), “NCEP/DOE AMIP-II Reanalysis (R-2),” *Bull. Amer. Meteor. Soc.*, 83, 1631–1643.

- Knutson, T., Zeng, F., and Wittenberg, A. (2013), “Multimodel Assessment of Regional Surface Temperature Trends: CMIP3 and CMIP5 Twentieth-Century Simulations,” *Journal of Climate*, 26, 8709–8743.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. (2010), “Challenges in Combining Projections from Multiple Climate Models,” *Journal of Climate*, 23, 2739–2758.
- Koichi, X., Brunke, M., and Wittenberg, A. (2012), “Temporal- and Spatial-Scale Dependence of Three CMIP3 Climate Models in Simulating the Surface Temperature Trend in the Twentieth Century,” *Journal of Climate*, 25, 2456–2470.
- Lee, M., Jun, M., and Genton, M. G. (2015), “Validation of CMIP5 Multimodel Ensembles through the Smoothness of Climate Variables,” *Tellus A*, 67, 23880.
- Liu, C., Allan, R., and Huffman, G. (2012), “Co-variation of Temperature and Precipitation in CMIP5 Models and Satellite Observations,” *Geophysical Research Letters*, 39.
- Lorenz, E. (1963), “Deterministic Nonperiodic Flow,” *J. Atmos. Sci.*, 20, 130–141.
- Mannshardt-Shamseldin, E., Smith, R. L., Sain, S. R., Mearns, L. O., and Cooley, D. (2010), “Downscaling Extremes: A Comparison of Extreme Value Distributions in Point-Source and Gridded Precipitation Data,” *Annals of Applied Statistics*, 1, 484–502.
- Masson, D. and Knutti, R. (2011), “Climate Model Genealogy,” *Geophysical Research Letters*, 38.
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L. (2012), “Tuning the Climate of a Global Model,” *Journal of Advances in Modeling Earth Systems*, 4.
- Moriondo, M. and Bindi, M. (2006), “Comparison of Temperatures Simulated by GCMs, RCMs and Statistical Downscaling: Potential Application in Studies of Future Crop Development,” *Climate Research*, 30, 149–160.

- Poppick, A. and Stein, M. L. (2014), “Using Covariates to Model Dependence in Nonstationary, High-Frequency Meteorological Processes,” *Environmetrics*, 25, 293–305.
- Reichler, T. and Kim, J. (2008), “How Well do Coupled Models Simulate Today’s Climate,” *Bull. Amer. Meteor. Soc.*, 89, 303–311.
- Saha, S. et al. (2010), “The NCEP Climate Forecast System Reanalysis,” *Bull. Amer. Meteor. Soc.*, 91, 1015–1057.
- Sanderson, B. M. and Knutti, R. (2012), “On the Interpretation of Constrained Climate Model Ensembles,” *Geophysical Research Letters*, 39.
- Smith, R. L., Tebaldi, C., Nychka, D., and Mearns, L. O. (2009), “Bayesian Modeling of Uncertainty in Ensembles of Climate Models,” *Journal of the American Statistical Association*, 104, 97–116.
- Taylor, K. (2001), “Summarizing Multiple Aspects of Model Performance in a Single Diagram,” *Journal of Geophysical research*, 106, 7183–7192.
- Taylor, K., Stouffer, R., and Meehl, G. (2012), “An Overview of CMIP5 and the Experiment Design,” *Bull. Amer. Meteor. Soc.*, 93, 485–498.
- Tebaldi, C. and Knutti, R. (2007), “The Use of Multi-Model Ensemble in Probabilistic Model Projections,” *Philos. Trans. R. Soc. A*, 365, 2053–2075.
- Tebaldi, C., Richard, L. S., Nychka, D., and Mearns, L. O. (2005), “Quantifying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multimodel Ensembles,” *Journal of Climate*, 18, 1524–1540.
- Van Vuuren, D. et al. (2011), “The Representative Concentration Pathways: An Overview,” *Climatic Change*, 109, 5–31.
- Xu, J., Powell, A., and Zhao, L. (2013), “Intercomparison of Temperature Trends in IPCC CMIP5 Simulations with Observations, Reanalyses and CMIP3 Models,” *Geoscientific Model Development*, 6, 1705–1714.

- Yao, Y., Luo, Y. and Huang, J., and Zhao, Z. (2013), “Comparison of Monthly Temperature Extremes Simulated by CMIP3 and CMIP5 Models,” *Journal of Climate*, 26, 7692–7707.
- Ye-Won, S., K., H., Kyung-Sook, Y., June-Yi, L., Kyung-Ja, H., and Ja-Yeon, M. (2014), “Future Change of Extreme Temperature Climate Indices over East Asia with Uncertainties Estimation in the CMIP5,” *Asia-Pacific Journal of Atmospheric Sciences*, 1–16.