Validation of Numerical Climate Models for the El Niño-Southern Oscillation System

Victor Privalsky[1] and Vladislav Yushkov[2]
[1]N.N. Zubov State Oceanographic Institute, Moscow, Russia
Email: vprivalsky@oceanography.ru
[2]Physics Department, Moscow State University, Moscow, Russia

Received: Dec. 16, 2013; Accepted: Feb. 14, 2014

Abstract
Statistical properties of the observed bi-variate ENSO time series (sea surface temperature within the Niño area 3.4 and the Southern Oscillation Index) from 1876 through 2005 are compared with respective properties of 46 CMIP5 models used in the historical experiment, one run per model. The models were found to exaggerate linear trend rates of SST; mean value and variance estimates have a large scatter, most probability densities are Gaussian, the shape of spectra is reproduced correctly in most cases though the spectra of simulated Southern Oscillation have a negative bias. Most estimates of coherence correctly reproduce the behavior of coherence between the observed SST and SOI that exceeds 0.9 at moderate frequencies. The average coherent spectrum of simulated SST is close to the "observed" coherent spectrum and has a negative bias in the SOI case. The results for the time domain require improvement; the frequency domain results are satisfactory.

Key words: climate model validation, coherence and coherent spectra, El Niño-Southern Oscillation

1. INTRODUCTION

This validation of climate models within the Coupled Model Intercomparison Project Phase 5 (CMIP5) historical experiment contains comparisons of major statistical properties of observed and simulated data at climatic time scales. The properties to be compared include linear trend rates, the first four statistical moments, probability and spectral densities, coherence functions, coherent spectra, and statistical predictability (persistence) properties. The choice of the bi-variate El Niño-Southern Oscillation (ENSO) system for validation purposes is based upon the fact that it presents a unique phenomenon in the Earth's climate; therefore, the results of such validation efforts can serve as a litmus test for further validations.

The ENSO system data have been used for validations of climate models before (e.g., Achuta Rao and Sperber 2006; Cibot et al 2005; van Oldenborgh et al 2005; Lin 2007) but in no case the analyses included comparisons between all major statistical properties of observed and simulated data. Quite recently, Guilyardi et al (2012) analyzed 18 CMIP5 models for the ENSO region and found no "quantum leap" over the CMIP3 results for the ENSO system. Our conclusions are more optimistic in some respects.

There are two well-known specific features that make the ENSO system unique. First, the spectra of both the sea surface temperature (SST) and the Southern Oscillation Index (SOI – a deseasonalized dimensionless difference of sea level pressure (SLP) between Tahiti and Darwin, Australia) behave in a special manner. They do not decrease with growing frequency as is common for spectra of climatic variability but are bell-shaped (e.g., von Storch 2001). The necessity for GCMs to generate time series

of SST and SLP with similar spectral densities will be called the spectrum requirement. Secondly, the two components of the system are closely related to each other: the magnitude of the linear correlation coefficient between them can be as high as 0.84 (e.g., Philander 1989) and coherence over 0.9 (Privalsky and Muzylev 2013) which, again, is very unusual at climatic time scales. The simulated ENSO data should possess a similar property which we will call the linear connection requirement.

ENSO is believed to affect other processes through teleconnections over the entire globe (e.g., Alexander et al 2002; Bartholomew and Jin 2013). Therefore, the ability of a GCM to generate time series of ENSO with the same statistical properties as the observed data can be regarded as a necessary condition for model's acceptability. If the model fails in these two respects, its further validation becomes questionable. In this study, we will test the ability of GCMs used in CMIP5 to correctly reproduce major statistical properties of the actual sea level pressure and sea surface temperature data as components of the bi-variate ENSO time series.

2. DATA AND METHODS

The initial observation data include monthly values of SST averaged over the Niño region 3.4 (bounded by 120° W - 170°W and 5°S - 5°N); the data is extracted from the HadCRUT4 archive at http://www.metoffice.gov.uk/hadobs/hadcrut4/. The SLP data at the Tahiti and Darwin stations are taken from the Australian Bureau of Meteorology site http://www.bom.gov.au/climate/current/soihtm1.shtml. The time interval for the observed and simulated data is from 1876 through 2005 (130 years), in accordance with the available observations of SOI and the span of the data simulated by the CMIP5 models within its historical experiment – normally, 1850-2005.

The simulated data for the sea surface temperature (averaged over the same region) within the CMIP5 historical experiment was downloaded in accordance with the list given at the ESGF portal http://pcmdi9.llnl.gov/esgf-web-fe/live#. The SLP data was taken from the same site. We analyzed only the first run of each model available for sea surface temperature and sea level pressure. The list of models analyzed here is given in Table 1.

Table 1. List of models

Name	Name	Name
ACCESS1.0 (1)	CSIRO-Mk3.6.0 (1)	HadGEM2-ES (5)
ACCESS1.3 (3)	CSIRO-Mk3L-1-2 (10)	INM-CM4 (1)
BCC-CSM1.1 (3)	EC-EARTH (3)	IPSL-CM5A-LR (6)
BCC-CSM1.1(m) (3)	FGOALS-g2 (10)	IPSL-CM5A-MR (3)
BNU-ESM (1)	FIO-ESM (5)	IPSL-CM5B-LR (1)
CanESM2 (5)	GFDL-CM2.1 (12)	MIROC5 (3)
CCSM4 (6)	GFDL-CM3 (1)	MIROC-ESM (3)
CESM1(BGC) (1)	GFDL-ESM2G (3)	MPI-ESM-LR (3)
CESM1(CAM5) (3)	GFDL-ESM2M (1)	MPI-ESM-MR (3)
CESM1(FASTCHE M) (4)	GISS-E2-H (1)	MPI-ESM-P (2)
CESM1(WCCM) (3)	GISS-E2-H-CC (17)	MRI-CGCM3 (5)
CMCC-CESM (4)	GISS-E2-R (1)	MRI-ESM1 (1)
CMCC-CM (1)	GISS-E2-R-CC (24)	NorESM1-M (3)
CMCC-CMS (1)	HadCM3 (10)	NorESM1-ME (1)
CNRM-CM5 (1)	HadGEM2-AO (1)	
CNRM-CM5-2 (10)	HadGEM2-CC (3)	

The number of available model runs is shown in parentheses.

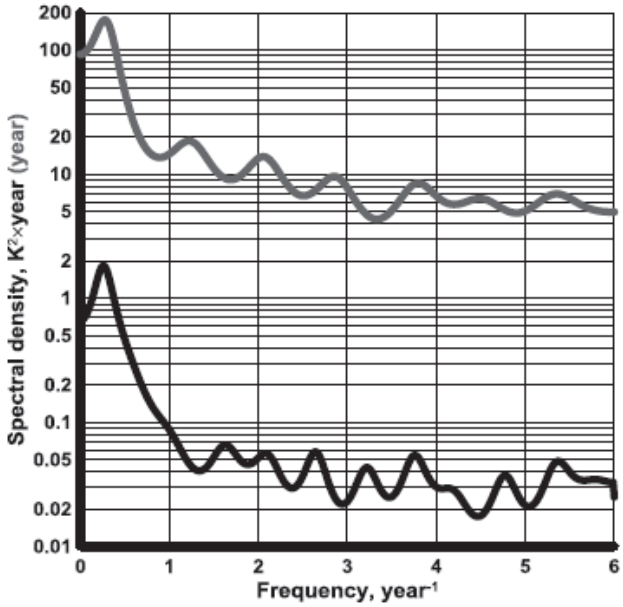

Figure 1. Spectral density of monthly values of SST 3.4 (black) and SOI (grey)

The validation experiment is conducted here for the annual data because we are interested in climatic time scales, which begin with year-to-year variability and do not include seasonal oscillations and other higher frequency variability. Besides, the spectral density of SST at frequencies exceeding 0.5 year^{-1} (with the mean seasonal trend removed) is one or two orders of magnitude below the spectral density at inter-annual scales (see Fig. 1). As mean annual values do not contain any seasonal trend, we substituted the differences of mean annual SLP values between Tahiti and Darwin for the atmospheric component of the ENSO time series and called it the annual Southern Oscillation Index (ASOI). At Δt = 1 year, ASOI is the result of a strictly linear transformation of SOI.

The statistical properties that are estimated here to compare observed and simulated data include

- linear trend rates,
- mean values,
- standard deviations,
- skewness and kurtosis,
- probability densities,
- order of stochastic difference equations that describe the time series,
- eigenfrequencies,
- statistical predictability (persistence),
- spectral densities,
- coherence functions,
- coherent spectra.

To the best of the authors' knowledge, this type of validation of GCM models that includes all major statistics has never been performed before with bi-variate simulated data at climatic time scales.

The historical experiment was conducted within CMIP5 under the assumption that the external forcing (including all anthropogenic effects) for simulations was as close as possible to the actual values that had occurred between 1876 (in our case) and 2005. As the anthropogenic influence is supposed to have caused, among other things, a higher temperature of the World Ocean, it can be assumed that the trend rate differences between the observed and simulated data reflect the differences between the sensitivities of the actual climate system and GCMs to the external forcing. If the trend in

the model data is significantly higher than in observations, the model is oversensitive; otherwise, it is not sensitive enough. (This is called the trend rate requirement here.)

The type of the probability density - Gaussian or non-Gaussian - was assumed in accordance with the values of the third and fourth statistical moments. If the magnitudes of standardized skewness and standardized kurtosis do not exceed 2, one can assume, at a 0.95 confidence level, that respective sample data have a Gaussian probability density (see D'Agostino and Pierson 1973). The estimation of the probability density type is important, in particular, because if the time series behave as samples of Gaussian random processes it simplifies further statistical analyses, including analysis of the time series predictability. It would also mean that the results obtained here within the linear Kolmogorov-Wiener theory of extrapolation cannot be improved through application of nonlinear methods (Yaglom 1962).

Other statistical properties of the observed and simulated data to be compared here characterize the behavior of SST and ASOI in the time and frequency domains and require some explanation. Let $x_{1,n}$ and $x_{2,n}$ be the time series of SST 3.4 and ASOI, respectively. Then the bi-variate time series $\mathbf{x}_n = [x_{1,n}, x_{2,n}]'$, $n = 1, 2, \ldots, N$, where $N = 130$ and the strike means matrix transposition, is regarded as a sample of a bi-variate linearly regular (specifically, autoregressive) random process, that is,

$$\mathbf{x}_n = \sum_{j=1}^{p} \mathbf{\Phi}_j \mathbf{x}_{n-j} + \mathbf{a}_n \,. \tag{1}$$

Here p is the order of autoregression,

$$\mathbf{\Phi}_j = \begin{bmatrix} \phi_{11}^{(j)} & \phi_{12}^{(j)} \\ \phi_{21}^{(j)} & \phi_{22}^{(j)} \end{bmatrix} \tag{2}$$

are matrix autoregressive (AR) coefficients, and $\mathbf{a}_n = [a_{1,n}, a_{2,n}]'$ is a bi-variate white noise with the covariance matrix

$$\mathbf{R}_\mathbf{a} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}. \tag{3}$$

The quantities R_{11} and R_{22} are the variances and $R_{12} = R_{21}$ covariances, of the white noise components $a_{1,n}$ and $a_{2,n}$. This parametric approach is quite reasonable and convenient in time series analysis, especially in situations when the number of observations is limited, as in our case. Note also that the AR models are capable of presenting random processes with practically any given spectrum. Estimates of the AR order p, matrix AR coefficients $\mathbf{\Phi}_j, j = 1, ..., p$, and the white noise covariance matrix $\mathbf{R}_\mathbf{a}$ are required in order to describe properties of the time series $\mathbf{x}_n$ in the time and frequency domains. For the sake of definiteness, we will be regarding the time series of SST $x_{1,n}$ and SLP $x_{2,n}$ as the output and input, respectively, of a linear system, which is described in the time domain with the above given stochastic difference equation (1). The matrices (2) and (3) are estimated here by using the multi-variate version of the Levinson's algorithm while the optimal order p of autoregression is selected on the basis of four criteria: Akaike's AIC, Parzen's BIC, Hennan-Quinn's CAT, and Schwarz-Rissanen's Ψ (Box et al 1994; Privalsky 1999).

The spectral matrix **s**(*f*) that corresponds to Eq. (1) is

$$\mathbf{s}(f)=\frac{2\left|\mathbf{R_a}\right|\Delta t}{\left|\mathbf{I}-\sum_{j=1}^{p}\Phi_j\exp(-i2\pi jf\Delta t)\right|^2},0\le f\le 1/2\Delta t, \tag{4}$$

where **I** is the identity 2×2 matrix, $|\mathbf{R_a}|$ the determinant of the matrix $\mathbf{R_a}$, Δt the unit time step (one year), *f* the cyclic frequency (year^{-1}), and $i=\sqrt{-1}$.

The elements of the spectral matrix

$$\mathbf{s}(f)=\begin{bmatrix} s_{11}(f)\ s_{12}(f) \\ s_{21}(f)\ s_{22}(f) \end{bmatrix} \tag{5}$$

are the spectral densities $s_{11}(f)$ and $s_{22}(f)$ of the output $x_{1,n}$ (SST 3.4) and the input $x_{2,n}$ (ASOI), respectively, while $s_{12}(f)=\bar{s}_{21}(f)$ are the complex-valued cross-spectra (the bar means complex conjugation). These quantities are used to calculate other functions which describe the response of the output component $x_{1,n}$ of the ENSO system to the input $x_{2,n}$ in the frequency domain but we will use here only the coherence function $Co(f)=\left|s_{12}(f)\right|\left[s_{11}(f)s_{22}(f)\right]^{-1/2}$ and the coherent spectrum $Cs_{12}(f)=Co^2(f)s_{11}(f)$. The former can be regarded as a frequency-dependent sequence of "correlation coefficients" while the latter describes the part of the spectrum $s_{11}(f)$ generated due to the linear dependence of $x_{1,n}$ upon $x_{2,n}$.

The confidence intervals for these spectral characteristics can be calculated in accordance with the number of degrees of freedom ν for the spectral estimates (see Bendat and Piersol 2010); for autoregressive analysis of bi-variate time series, ν can be set equal, as a first approximation, to $N/2p$ (Privalsky et al 1987). Note that estimates given without respective confidence intervals have no value.

3. RESULTS AND DISCUSSION

The goal here is to compare statistical properties of ENSO that are estimated from the observed data with respective properties deduced from simulated time series of SST and ASOI.

3.1. Linear trend rates

Sea Surface Temperature. The observed time series of SST contains a linear trend of 0.32×10^{-2} K/year, with the estimate's rms of 0.15×10^{-2} K/year. This means that the trend though statistically significant at the 95% confidence level, is close to being insignificant (and it does become insignificant if the SST time series is extended through 2012). The trend rate estimates in simulated time series vary between a statistically insignificant rate of 0.20×10^{-2} K/year (CSIRO-Mk3.6.0) and 0.93×10^{-2} K/year (IPSL-CM5A-LR) with the mean value of 0.52×10^{-2} K/year and the average rms of 0.16×10^{-2} K/year. This positive bias of more than 60% in the simulated data's trend estimate is statistically significant. In 36 out of the 46 cases of simulated data, the trend rates are faster than the trend in the observed data. On the whole, the models seem to be substantially more sensitive to the external forcing than the actual climatic system.

Sea Level Pressure Difference. The linear trend in the observed pressure differences (ASOI) is statistically insignificant. The simulated data reveal the same property though in several cases the trend was found to be significant (e.g., positive for EC-EARTH and negative for HadGEM2-ES). Yet, on the whole, the trend in simulated ASOI time series is not large and can be regarded as practically non-existent. In contrast to the SST case, the models seem to behave quite properly in this respect.

3.2. Mean values and standard deviations

Sea Surface Temperature. The mean observed SST is 298.7K with an rms error of slightly below 0.05K. The mean simulated SST is 299.5K, which does not differ much from the observed mean but is still significantly different from it at the 99% confidence level. The estimates of the mean SST in the simulated data vary between 297.2K (CSIRO-Mk3.6.0) and 301.3K (GISS-E2-R-CC) so that the range of the mean SST values as estimated by the models is 4.1K.

The differences between the mean SST of the observed and simulated data are statistically significant in all but three cases. The average difference with the observed mean SST is approximately 0.7K, with the maximum 2.6K (GISS-E2-R-CC) and from the minimum -1.5K (CSIRO-Mk3.6.0). This shows a tendency to a positive bias in the simulated mean annual SST.

The standard deviation in the observed time series is 0.64K with the rms of 0.04K; the average over all 46 models is 0.65K. The two estimates are statistically equivalent. The difference between the standard deviation values of observed and simulated time series is statistically insignificant in 26 out of the 46 cases at the 95% confidence level. On the other hand, the minimum and maximum standard deviations amount to 0.4K (MIROC-ESM) and 1.2K (CMCC-CESM). In other words, the standard deviations of simulated data vary by 300%. This scatter in the standard deviation estimates will tell upon the variability of spectral estimates for the models and will make them less reliable.

Sea Level Pressure Difference. The mean difference of the observed annual sea level pressure values between Tahiti and Darwin amounts to 2.71 mb; respective statistics for the 46 simulated time series is 2.80 mb. The difference is statistically insignificant. The simulated mean values exhibit variability by an order of magnitude: from 0.6 mb (MRI-GCM3) to 6 mb (GISS-E2-H-CC).

The observed data have a standard deviation of 1.08 mb; the standard deviations of simulated data vary between 0.3 and 1.3 mb with the mean of 0.71 mb. The negative bias in the standard deviation estimates will cause a negative bias in the estimates of ASOI spectra.

3.3. Probability density

Sea Surface Temperature. The Gaussian approximation is acceptable for the probability density of the observed data as well as for 38 out of the 46 simulated time series. The time series whose probability distributions are significantly non-Gaussian were generated by the following models: CSIRO-MK3L-1-2, FIO-ESM, GFDL's CM2.1, GISS-E2-H, HadGEM2-ES, MIROC5, MPI-ESM-P, and MRI-CGCM3. But even in those seven cases, the deviations are not very spectacular. Having in mind the shortness of the data (130 annual values) and the sampling variability of estimates, it seems reasonable to assume that the simulated data can also be regarded as Gaussian.

Sea Level Pressure Difference. The observed and 39 out of the 46 simulated time series have probability densities which can be regarded as Gaussian. This is an important positive property of the simulated time series of SST and ASOI.

3.4. Time and frequency domain properties

Time Domain. The stochastic model for the observed data from 1876 through 2005 does not differ much from the first observation-based bi-variate ENSO time series model given in the recent article by Privalsky and Muzylev (2013). It is described in the time domain with a system of stochastic difference equations of the form (1) with a noise covariance matrix (2). The best autoregressive approximation for the observed time series SST/ASOI is a model of order $p = 2$; it has been chosen by all four order selection criteria. The simulated data have the same autoregressive order in 27 cases with the other 19 models following a bi-variate Markov model (that is, $p = 1$).

The correlation coefficient between the observed SST and ASOI is -0.84 and the simulated data agree with this, with the highest correlation coefficient of better than -0.9.We will not dwell on it here as the correlation coefficient is by no means the best measure of the degree of linear dependence between time series.

The observed data have two eigenfrequencies: at approximately 0.24 $year^{-1}$ and 0.05 $year^{-1}$; as will be shown later, the lower frequency oscillation is not even seen in the SST and pressure spectra. The

eigenfrequency of about 0.2 - 0.3 years^{-1} had been found for ENSO in earlier publications, both experimental (von Storch et al 2001; Mokhov et al 2004, Privalsky and Muzylev 2013) and theoretical (Jin, 1977; Kleeman, 2011). An eigenfrequency belonging to approximately the same interval is found in many time series of simulated data, though in five of the 19 cases when the optimal order of the model was $p = 1$, this oscillation is not found.

Actually, the role of the damped oscillation in the behavior of the ENSO system is quite meager because most of its energy is received through the bi-variate innovation sequence $\mathbf{a}_n$. The quantitative measure of its contribution to the variances of SST and ASOI (as well as a measure of the time series statistical predictability, or persistence) is given by the ratios R_{11}/σ_1^2, R_{22}/σ_2^2 where R_{11} and R_{22} are the variances of the white noise sequences $a_{1,n}$ and $a_{2,n}$ [see Eq. (3)] while σ_1^2 and σ_2^2 are the variances of SST and ASOI, respectively. For the observed temperature and pressure differences these ratios are close to 1: $R_{11}/\sigma_1^2 = 0.82$, $R_{22}/\sigma_2^2 = 0.84$, which means that the damped periodic oscillations do not carry much energy. Therefore, the annual values of SST and ASOI have low statistical predictability. Most simulated time series behave in the same manner but about a dozen models have this predictability parameter below 0.8 with the smallest value close to 0.6. This is still not a very high statistical predictability and it can probably be ascribed to the sampling variability of estimates. The share of the "deterministic" oscillation amounts to less than 20% in the observed time series and in most simulations. The average value of this parameter for the simulated SST time series is the same as for the observations. On the whole, the models' performance seems to be satisfactory in this respect.

Some deviations of individual models from observations, if they do not occur too often, could probably be treated as sampling variability. Essentially, we are dealing here with an ensemble of sample realizations of a regarded bi-variate random process, with each realization designed to simulate the same climatic phenomenon – natural and forced variability of the SST-ASOI system. Moreover, the simulated sample realizations of SST seem to be mutually independent and the same statement is true of the simulated SLP time series. This happens in spite of the fact that all simulated data contain the same contribution from the forcing function, which means that the predetermined forcing does not play an important role for the SST and SLP in the Niño area.

Frequency Domain Properties - Spectra. As has been shown above, the one-dimensional probability densities of SST and ASOI are Gaussian. This means that their bi-variate probability densities are also Gaussian (Yaglom, 1962) so that our autoregressive estimates of spectra are the maximum entropy estimates.

According to the observation data, both SST and ASOI behave rather exotically for a climate process: their spectra do not decrease with growing frequency but have a single wide maximum centered at the eigenfrequency band (see Fig. 2). Again, this is a well-known result obtained earlier for scalar spectral estimates (e.g., Chu and Katz 1985; von Storch 2001; Mokhov et al 2004) and recently as bi-variate maximum entropy spectra for the ENSO system (Privalsky and Muzylev 2013). As seen from Fig. 2*a*, the spectra that correspond to models of order $p = 1$ behave in a different manner: most of them are monotonic and four – BCC-CSM1-1, BCC-CSM1-1(M), CNRM-CM5-2, GISS-E2-H – even grow with frequency, which is very unusual for climatic processes. The spectrum averaged over all estimates with $p = 1$ is very close to a white noise spectrum. Moreover, some of these spectra are monotonous functions of frequency, which means that in contrast to the SST observations, the random process that generated them does not contain a damped periodic oscillation. On the other hand, in criticizing the results shown in Fig. 2*a*, one should remember that the "true" spectrum estimated on the basis of observations is close to a white noise as well and that the contribution of the damped periodic oscillation to the SST variance is actually rather small. In other words, the differences between models and observations shown in Fig. 2*a* should hardly be regarded as critical.

The shapes of the spectra in Fig. 2*b* (models of order $p = 2$) are quite similar to that of the spectrum of the observed annual SST. The spectral maximum corresponding to the damped periodic oscillation mostly stays within the 0.2 year^{-1} to 0.3 year^{-1} frequency band and the weighted averages of the models of order $p = 2$ as well as of all 46 spectral estimates closely follows the "true" spectrum. In other words, the CMIP5 models studied here more or less successfully simulate the shape of the observed time series

spectrum and, on the average, its values. However, as the variance estimates for individual models vary by almost an order of magnitude, the spectral estimates of simulated data show substantial variability that makes them relatively unreliable. On the whole, it looks like more models have the correct shape of the SST spectrum than it used to be for the scalar case (SST only) at the CMIP3 stage of the IPCC program (Achuta Rao 2006).

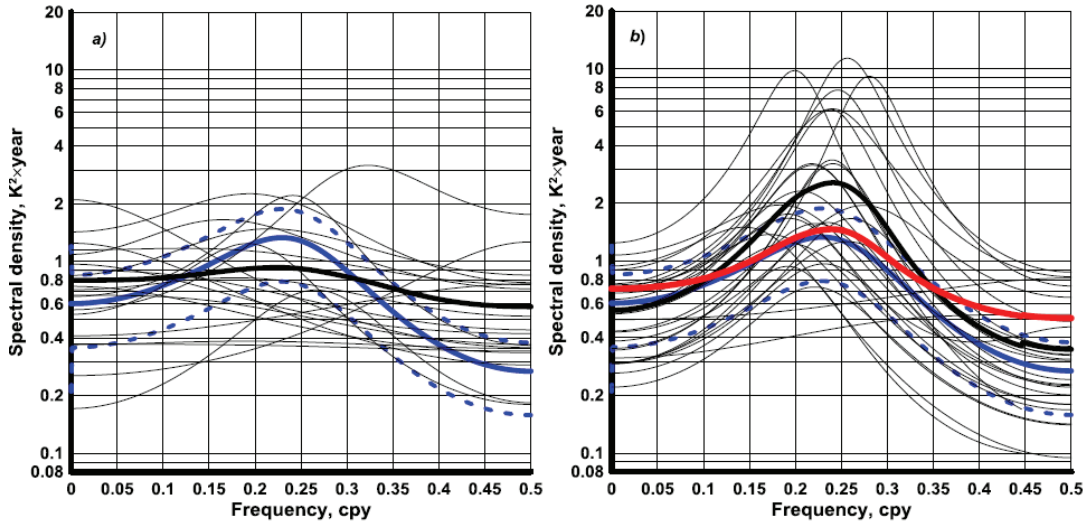

Figure 2. Maximum entropy spectral estimates of SST for p = 1 (*a*) and p = 2 (*b*): observations (blue) with 90% confidence limits (dashed), simulations (thin black lines), average for p = 1 or p = 2 (thick black lines), average for all 46 models (red)

The estimates of ASOI spectra behave in about the same manner as the SST spectra, except that the spectra of simulated ASOI for the case when p = 1 lie, as a rule, below the spectrum of the observed pressure differences (see Fig. 3*a*). For simulated time series with p = 1, the standard deviation estimates

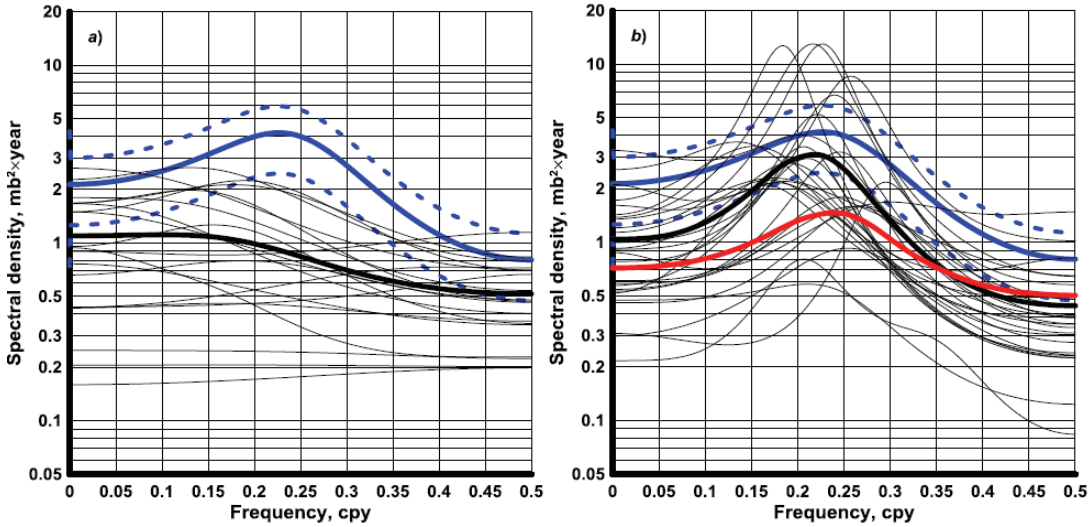

Figure 3. Maximum entropy spectral estimates of ASOI for p = 1 (*a*) and p = 2 (*b*): observations (blue) with 90% confidence limits (dashed), simulations (thin black lines), average for p = 1 or p = 2 (thick black lines), average for all 46 models (red)

remain below the estimate of the standard deviation of the observed ASOI data in most cases. Actually, the models that underestimate the "true" model's order also underestimate the ASOI variance. When the model's order $p = 2$, the general picture improves and again the spectra follow the behavior of the observed ASOI spectrum. On the whole, the average spectrum of simulated data for all 46 models follows the shape of the observed ASOI data but stays below the lower 90% confidence bound for it. This effect is again caused by the underestimation of the ASOI variance by many models. Also, the scatter in the estimates of ASOI variance causes substantial variability in the absolute values of ASOI spectra.

Frequency Domain Properties – Coherence between SST and ASOI. It has already been mentioned above that the correlation coefficient between time series may not contain the information required for describing the linear dependence between them. The quantity which should be used for this purpose is the coherence, which presents a frequency-dependent quantitative indicator of linear dependence between time series.

As seen from Fig. 4, the maximum entropy estimate of coherence between the observed time series of SST and ASOI stays above 0.8 everywhere from 0.11 year^{-1} to 0.40 year^{-1} and exceeds 0.90 between 0.17 year^{-1} and 0.32 year^{-1}. This means that SST and ASOI determine each other's variance and spectra by over 65% and over 80% within those two frequency bands, respectively. The simulated time series behave, generally, in about the same manner though the coherence estimates for simulated data with the AR order $p = 1$ are not bell-shaped, as for the $p = 2$ cases. The general conclusion is that with just one exception (GISS-E2-H) all models demonstrate strong linear dependence between the two components of the time series, quite similar to what is observed for the actual data.

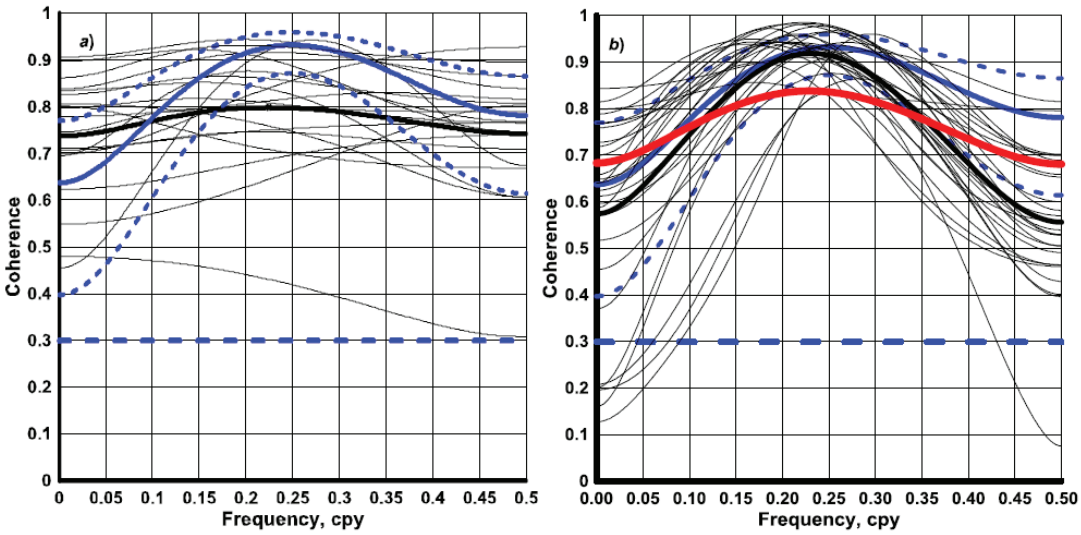

Figure 4. Coherence function estimates for $p = 1$ (*a*) and $p = 2$ (*b*): observations (blue) with 90% confidence bounds (dashed lines), simulation (thin black lines); average for $p = 1$ and $p = 2$ (thick black lines), average for all 46 models (red). The dashed horizontal line is the upper 90% confidence limit for the true zero coherence.

Another indicator of simulation quality is the coherent spectrum. If ENSO is a system with ASOI and SST as the input and output processes, respectively, then the coherent spectrum of SST is the part of the total SST spectrum generated by the linear contribution from the input. Similarly, by regarding SST and ASOI as the input and output processes, respectively, we will have the coherent spectrum of ASOI as the part of the total ASOI spectrum that is generated by the linear contribution from SST.

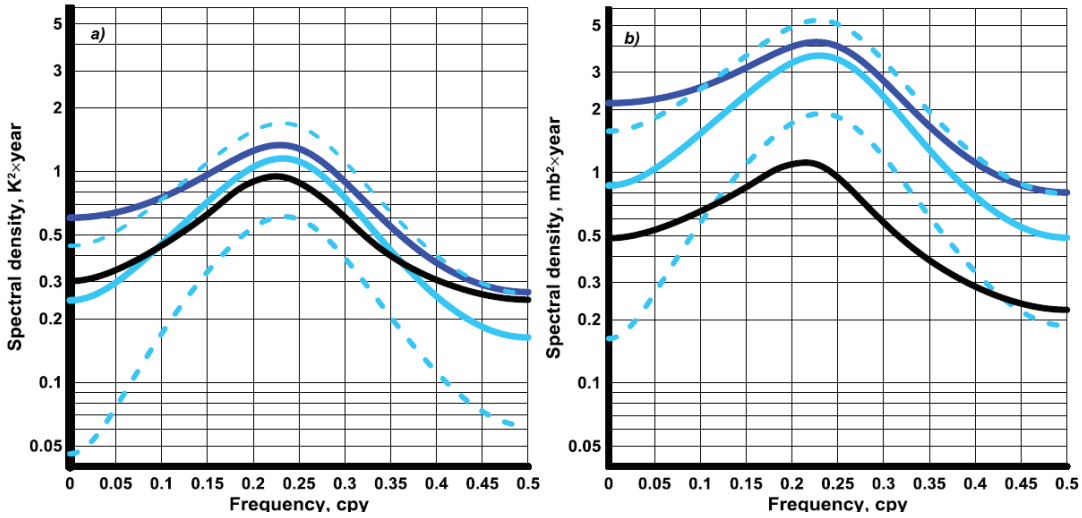

Figure 5. Coherent spectra of SST (*a*) and ASOI (*b*): spectra of observed data (blue), coherent spectra of observed data with 90% confidence limits (light blue), average coherent spectra for simulated data (black)

The coherent spectra of SOI and SST are given in Fig. 5. The solid blue lines in the figure show the coherent spectra of SST and ASOI (Fig. 5*a* and Fig. 5*b*, respectively) with their 90% confidence bounds as determined from observations. Due to the high coherence values, the coherent spectra are rather close to the total spectra (dark blue lines), especially at moderate frequencies. Approximate estimates of coherent spectra for simulated data were determined as products of the ensemble averaged spectra of simulated SST and simulated ASOI by the squared coherence averaged over the ensemble of simulated data. The resulting average coherent spectra are shown with black lines. It is easy to see that the average coherent spectrum of simulated SST is closer to the coherent spectrum of observed SST (Fig. 5*a*) than in the case of ASOI (Fig. 5*b*). As the coherence function is the same for the systems SST-ASOI and ASOI- SST, this behavior is explained by the fact that the average spectrum of simulated SST is closer to the spectrum of observed SST than in the case of ASOI (cf. Figs. 2 and 3). In other words, the quality of simulation of SST is better in this respect than for simulations of ASOI and the drop in the quality of simulation is due to the negative bias in simulated ASOI spectra.

4. CONCLUSIONS

The following conclusions have been arrived at on the basis of our analysis of the quality of SST and SLP data simulated by CMIP5 models for the ENSO system within the Niño area 3.4.

1. Most models overestimate the linear trend rate in the SST; the positive bias amounts to about 60% of the trend rate in the observed time series. In 16 cases out of the 46, the bias amounts to at least 100%. Thus, the climate models seem to be more sensitive to the external forcing than the actual climate in the region. In other words, most models fail to meet the trend rate requirement for the annual SST variations.
   The trend in the time series of observed annual pressure difference between Tahiti and Darwin is statistically insignificant. The model results agree with this completely.
2. The mean values of simulated time series differ statistically significantly from the mean of the observed time series in practically all cases. The average difference between the mean value of the observed SST data and those of the simulated time series amounts to 2.2K, with the maximum and minimum values of 2.6K and -1.5K, respectively. Thus, the climate models seem to have a tendency to overestimate the mean annual SST in the region.

The differences between the standard deviations of SST are statistically significant in 60% of cases and the estimates show a large (about 300%) variability which also affects estimates of the SST spectra thus diminishing their reliability.

The mean value of the observed pressure differences between Tahiti and Darwin is 2.71 mb. The average mean value of simulated data is 2.81 mb. The ensemble-averaged standard deviation is 0.71 mb, while the standard deviation of the observed data is 1.08 mb. Most estimates differ significantly from the observed value. The estimates of standard deviation are negatively biased.

3. The probability density functions (pdf) of simulated time series of SST and ASOI coincide with the pdf of observation (which is Gaussian) in most cases.
4. The average value of the persistence (predictability) parameter for the simulated SST is the same as for the observed data and amounts to about 0.8. Though a few models have a better predictability, the overall performance in this respect is satisfactory.
5. The maximum entropy spectral estimates of simulated time series of SST show substantial variability with most of them having the same Gaussian curve shape as the spectrum of observations. The average spectrum of simulated SST stays within the 90% confidence bounds for the spectrum of observed SST at frequencies up to about 0.35 $year^{-1}$. Having in mind that the behavior of the actual SST spectrum in the Niño area is unique for the climate system, this result is quite positive. On the other hand, the spectral estimates show large variability that makes the average estimate less reliable. On the whole, the spectrum requirement for the quality of SST simulation in the area seems to have been fulfilled only partially. The same conclusion refers to the spectra of ASOI.
6. The coherence within the observed bi-variate time series SST-ASOI has the same form as the bell-shaped SST and ASOI spectra. It attains values over 0.90 at frequencies between 0.18 $year^{-1}$ and 0.33 $year^{-1}$. This remote connection seems to be absolutely unique for the climate system. The coherence between simulated time series behaves in the same manner and the majority of coherence estimates imitate the shape of the "observed" coherence function. The weighted average of simulated coherence estimates has the right shape. It stays below the lower 90% confidence limit for the observed coherence within the frequency range from 0.20 through 0.30 $year^{-1}$. This probably happens because of the relatively low coherence values at those frequencies for models with the AR order $p = 1$. On the whole, it can be said that the linear connection requirement has been fulfilled.
7. The quality of simulations of coherent spectra is lower for ASOI due to the negative bias in the estimates of simulated ASOI spectra.

   We believe that the partial success achieved by climate modelers in simulating the ENSO system is satisfactory enough to continue with validations of numerical climate models for other spatial scales.

5. ACKNOWLEGEMENT

The authors wish to express their sincere appreciation to the N.N. Zubov Institute of Oceanography (Moscow, Russia) for the support of the study.

6. REFERENCES

Achuta Rao, K. and Sperber, K. R. (2006) "ENSO simulation in coupled ocean-atmosphere models: Are the current models better?" *Climate Dynamics*, Vol. 27(1), pp. 1 – 15.

Alexander, M.A., Bladé, I., Newman, M., Lanzante, J.R., Lau, N.-C. and Scott, J.D., 2002. "The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air–Sea Interaction over the Global Oceans." *J. Climate*, Vol. 15(16), pp. 2205-2231.

Bendat, J. and Piersol, A. 2010. "Random Data. Analysis and Measurements Procedures", John Wiley &Sons, Hoboken, New Jersey.

Box, G.E.P., Jenkins, G.W. and Reinsel, G.C., 1994. "Time Series Analysis. Forecasting and Control", Third Edition. Prentice-Hall, New Jersey.

Bartholomew, H. and Jin, M.S., 2013. "ENSO Effects on Land Skin Temperature Variations: A Global Study from Satellite Remote Sensing and NCEP/NCAR Reanalysis", *Climate*, Vol. 1(2), pp. 53-73.

Cibot, C., Maisonnave, E., Terray, L., Dewitte, B., 2005. "Mechanisms of tropical Pacific interannual-to- decadal variability in the ARPEGE/ORCA global coupled model." *Climate Dynamics*, Vol. 24(7-8), pp. 823–842

Chu, P.-S. and Katz, R.W., 1985. "Spectral estimation from time series models with relevance to the Southern Oscillation". *J. Climate*, Vol. 2(1), pp. 86-90.

D'Agostino, R.D. and Pearson, E.S., 1973. "Tests for departure from normality. Empirical results for the distribution of b_2 and $\div b_1$." *Biometrika*, Vol. 60(3), pp. 613-622.

Guilyardi E., Bellenger, H., Collins, M., Ferrett, S., Cai, W. and Wittenberg, A. 2012. "A first look at ENSO in CMIP5." *CLIVAR Exchanges* No. 58, Vol. 17(1), pp. 29-32.

Jin, F.-F. 1997. "An equatorial ocean recharge paradigm. Part I: Conceptual model", *J. Atmos. Sci*., Vol. 54(7), pp. 811-829.

Kleeman, R. 2011. "Spectral analysis of multidimensional stochastic geophysical models with an application to decadal ENSO variability", *J. Atmos. Sci*., Vol. 68(1), pp. 13-25.

Lin, J-L. 2007. "Interdecadal variability of ENSO in 21 IPCC AR4 coupled GCMs", *Geophys. Res. Letters*, Vol. 34(12), L12702, doi:10.1029/2006GL028937.

Mokhov, I., Khvorostyanov, D., Eliseev, A. 2004. "Decadal and longer term changes in El Niño – Southern Oscillation characteristics", *Intern. J. Climatology*, 24(4), pp. 401-414.

Philander, S. G. 1989. "El Niño, La Niña, and the Southern Oscillation", *Academic Press*, CA.

Privalsky, V., Protsenko, I. and Fogel G. 1987. "The sampling variability of autoregressive spectral estimates for bi-variate hydrometeorological processes", *Proc. 1st World Congress of the Bernoulli Society on Mathematical Stat. Theory and Applications*, Tashkent, USSR, Vol. 2, pp. 651-654 VNU Science Press, Utrecht.

Privalsky, V. 1999. "Parametric model; order selection", *Probability and Mathematical Statistics Encyclopedia*. Yu. V. Prokhorov (ed.), The Great Russian Encyclopedia Publishers, Moscow, Russia, 442-443 (in Russian).

Privalsky, V., Muzylev, S. 2013. "An Experimental Stochastic Model of the El Niño – Southern Oscillation System at Climatic Time Scales", *Universal Journal of Geoscience*, Vol. 1(1), pp. 28-36

Van Oldenborgh, G.J., Philip, S.Y. and Collins, M. 2005. "El Niño in a changing climate: a multi-model study", *Ocean Science* Vol. 1(2), pp. 81-95.

Von Storch, J.-S. Wang B. and Wang, Z. 2001. "The ENSO-spectrum: a result of deterministic chaos or stochastic forcing?", *Dynamics of Atmospheric and Oceanic Circulation and Climate*, China Meteorological Press, Beijing, pp. 579-600.

Yaglom, A.M. 1962. "An Introduction to the Theory of Stationary Random Functions", *Dover Publications*, NY.