ATMOSPHERIC SCIENCE LETTERS
Atmos. Sci. Let. **16**: 461–464 (2015)
Published online 14 July 2015 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/asl.582

Royal Meteorological Society

Validation of CMIP5 models for the contiguous United States

Victor Privalsky[1]* and Vladislav Yushkov[2]

[1] *Space Dynamics Laboratory (ret.), Logan, UT, USA*
[2] *Physics Department, Moscow State University, Russia*

Correspondence to: V. Privalsky, 1272 Eastridge Drive, Logan, UT 84321, USA. E-mail: vprivalsky@gmail.com

Received: 7 January 2015
Revised: 30 March 2015
Accepted: 31 March 2015

Abstract

Major statistical characteristics – trend rates, mean values, standard deviations, probability densities, autoregressive model orders, persistence criteria, and spectra – of annual surface temperature over the contiguous United States from 1889 through 2005 are compared with respective characteristics of 47 time series generated within the CMIP5 historical experiment. The observed and most simulated time series are Gaussian. Most autoregressive orders, persistence criteria, and spectra of simulated time series are close to what is found in nature. Although the multi-model mean value is not biased, individual models can err by almost 3.5 °C. In addition, the models exaggerate linear trend rates and temperature variance is overestimated.

Keywords: **climate models; statistical properties; verification; CONUS**

1. Introduction

Efforts to show that numerical models correctly reproduce climatic variability including its stochastic components present a traditional research area in climatology (e.g. Jones *et al.*, 2013). A full-scale validation of climate models constitutes a statistically unmanageable task because it requires quantitative comparisons between time-dependent sequences of observed and simulated multidimensional random fields of climatic data. However, the task becomes doable if the data are spatially averaged so that multidimensional random fields become scalar time series. Comparing time series to each other is relatively easy. If a simulated time series obtained in this manner has statistically the same basic properties – such as trend rates, mean values and variances, probability density functions, *etc.* – as respective observed time series, one may regard the climate model as reliable at the given scale of spatial averaging. This is an obvious necessary condition for recognizing models as being in agreement with observations. If, on the contrary, a model generates data whose major statistical properties differ significantly from those obtained from observations, the model is inadequate.

This spatial averaging approach, first suggested in Privalsky and Croley (1992), is used here to compare basic statistical properties of the observed annual surface temperature (AST) averaged over the 48 states of the contiguous United States (CONUS) with respective properties of AST generated with CMIP5 models within the framework of CMIP5 historical experiment at the same scale of spatial averaging. The choice of CONUS is by no means random because the observation data over the CONUS territory since the end of the 19th century are probably more reliable than respective data over any other region of similar or bigger size.

The statistical properties of AST to be analyzed here include:

1. linear trend rates,
2. mean values and standard deviations,
3. type of probability density functions (PDFs),
4. orders of optimal autoregressive approximations and time series persistence (statistical predictability),
5. spectral densities.

2. Data and methods

The initial monthly observation data were taken from the HadCRUT4 file from the website of the University of East Anglia (Morice *et al.*, 2012) and averaged over the CONUS and over 12 months to obtain a time series of CONUS annual surface temperature. The observation data are available for the entire time interval of the CMIP5 historical experiment from 1850 through 2005 but we selected a shorter interval from 1889 through 2005 during which the coverage with observations was, according to the data set, never below 94%. The simulation data for the same time interval were obtained from the CMIP5 project data site (see Taylor *et al.*, 2012) and averaged in the same manner to obtain 47 time series of simulated AST. With just a few exceptions, comparisons were conducted for run 1 of each model.

The first four statistical characteristics were calculated in the usual manner while the spectra were estimated through the autoregressive (AR) time series modeling; the time domain models with properly selected AR orders served as a basis for the maximum

Table 1. List of models and statistical properties of observed and simulated temperature.

Model	Trend, °C (100 years)$^{-1}$	Mean, °C	Standard deviation, °C	p*	$d(1)$*	$\tau_{0.9}$, yrs*
HadCRUT4 (observed)	0.56	11.16	0.46	1	0.97	0
ACCESS1-0	0.48	12.97	0.54	2	0.95	0
ACCESS1-3	0.29	14.59	0.46	1	0.93	1
BCC-CSM1-1	0.94	10.68	0.57	1	0.96	0
BCC-CSM1-1-M	0.93	11.65	0.58	0	1	0
BNU-ESM	1.24	9.86	0.64	0	0.95	1
CanESM2	0.53	10.26	0.62	1	0.92	1
CCSM4	0.23	11.46	0.57	0	1	0
CESM1-BGC	0.78	11.83	0.60	1	0.94	1
CESM1-CAM5	0.75	11.80	0.49	1	0.96	0
CESM1-FASTCHEM	1.04	12.11	0.57	1	0.94	1
CESM1-WACCM	0.97	11.61	0.55	0	1	0
CMCC-CESM	0.29	10.83	0.53	2	0.95	0
CMCC-CM	0.56	9.83	0.48	2	0.95	0
CMCC-CMS	0.42	10.39	0.54	2	0.95	0
CNRM-CM5	0.46	10.72	0.64	0	1	0
CNRM-CM5-2	0.40	10.52	0.53	0	1	0
CSIRO-MK3-6-0	0.35	12.20	0.54	1	0.95	1
CSIRO-MK3L-1-2	0.81	12.94	0.46	0	1	0
EC-EARTH	1.19	11.73	0.43	1	0.97	0
FGOALS-G2	0.90	7.80	0.52	1	0.91	1
FIO-ESM	0.84	11.63	0.43	2	0.97	0
GFDL-CM21	1.09	10.10	0.74	1	0.98	0
GFDL-CM3	0.38	10.64	0.52	1	0.91	1
GFDL-ESM2G	0.63	10.06	0.56	0	1	0
GFDL-ESM2M	0.67	10.62	0.63	2	0.84	2
GISS-E2-H	0.54	9.90	0.41	1	0.97	0
GISS-E2-H-CC	0.49	10.51	0.38	0	1	0
GISS-E2-R	0.27	9.66	0.45	1	0.97	0
GISS-E2-R-CC	0.27	9.40	0.44	1	0.92	1
HADCM3	0.68	10.00	0.71	1	0.94	1
HADGEM2-AO	0.34	12.49	0.64	3	0.84	3
HADGEM2-CC	−0.03	11.87	0.61	0	1	1
HADGEM2-ES	0.67	11.83	0.63	3	0.78	4
INM-CM4	0.96	10.02	0.42	0	1	0
IPSL-CM5A-LR	1.49	9.92	0.47	1	0.96	0
IPSL-CM5A-MR	0.88	10.97	0.47	0	1	0
IPSL-CM5B-LR	0.89	9.67	0.52	1	0.89	1
MIROC5	0.58	13.00	0.54	1	0.92	1
MIROC-ESM	0.74	12.24	0.48	3	0.88	2
MIROC-ESM-CHEM	0.42	11.92	0.45	1	0.94	1
MPI-ESM-LR	1.29	10.90	0.55	1	0.96	0
MPI-ESM-MR	1.37	11.33	0.58	1	0.98	0
MPI-ESM-P	1.21	11.65	0.65	1	0.95	1
MRI-CGCM3	0.58	9.66	0.44	0	1	0
MRI-ESM1	0.78	9.87	0.39	0	1	0
NORESM1-M	0.65	10.84	0.50	1	0.96	0
NORESM1-ME	0.62	10.48	0.55	1	0.95	0
AVERAGE	0.70	11.00	0.53	1	0.95	1

*p, optimal autoregressive order; $d(1)$, relative prediction error variance within the Kolmogorov–Wiener theory at lead time $\tau = 1$ year; $\tau_{0.9}$, lead time at which the relative prediction error variance attains 0.9 or more.

entropy spectral estimates and estimates of the time series' persistence.

3. Results

The list of models analyzed here is given in the first column of Table 1 with graphs of observed and simulated time series shown in Figure 1.

3.1. Linear trend

As seen from Figure 1(a), both observed and simulated temperature anomalies reveal a positive trend. The trend rates have been estimated under the assumption that every time series x_n, $n = 1, 2, \ldots, N$, presents a sum of a zero mean white noise sequence ε_n with a linear function of time: $x_n = \alpha n + \varepsilon_n$. As will be seen later, the statistical models of the observed and simulated time series of AST do not differ much from such a model. Estimates of the coefficient α are shown in Table 1 for the entire interval from 1889 through 2005. As seen from the table, the trend rate in the observed data amounts to 0.56 °C (100 years)$^{-1}$ with the estimate's standard error 0.13 °C (100 years)$^{-1}$. The average trend rate estimate for the simulated data is 0.70 °C (100 years)$^{-1}$. Only 22 trend rate estimates for the simulated

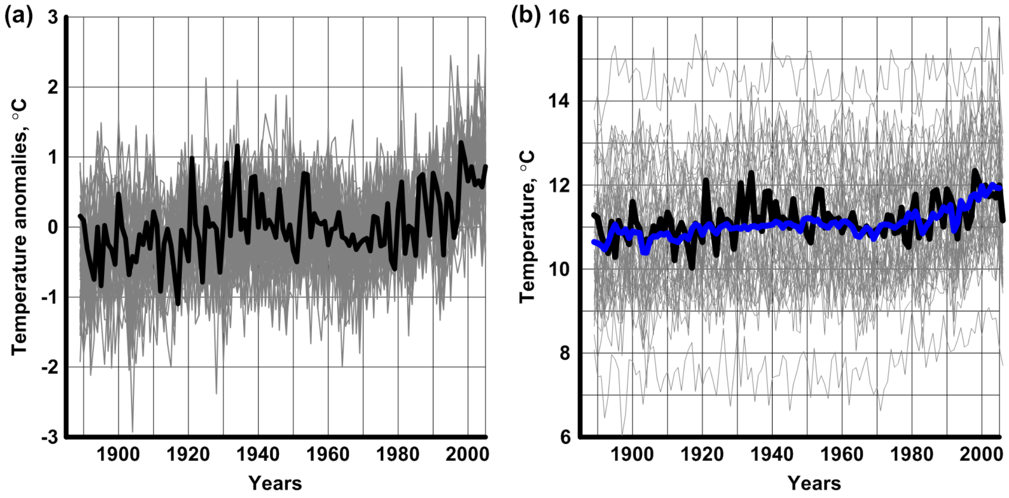

Figure 1. Observed (black) and simulated (grey) AST over CONUS: (a) anomalies; (b) temperature; the blue line shows the average simulated temperature.

data lie within the 95% confidence interval for the estimate of the trend rate in the observed time series. The range of trend rate estimates in simulated data is 1.5°C (100 years)$^{-1}$ [from 0 to 1.5°C (100 years)$^{-1}$]. In other words, the trend rate in the simulated data is seriously overestimated. This result is unsatisfactory. Note also that the positive bias in trend rate estimates can hardly be caused by observation errors because, according to Menne *et al.* (2010), there is 'no evidence that the CONUS average temperature trends are inflated due to poor station siting'.

As the linear trend is supposed to have been caused by external forcing, it has been removed from all time series prior to further analysis.

3.2. Mean values and standard deviations

The observed mean value for 1889–2005 is 11.16 °C. The multi-model average mean value estimate for the simulated data is 11.00 °C (Figure 1(b)). However, the mean value estimates for the simulated data lie between 7.8 and 14.6 °C (a 6.8 °C range) and everyone of them differs statistically significantly from the observed value. These results are unsatisfactory.

As seen from Figure 1(a), the simulated time series have a higher variance than observations. The observed standard deviation (s.d.) of AST is 0.42 °C with a 95% confidence interval between 0.37 and 0.48 °C. The average simulated s.d. is 0.53 °C – a statistically significant difference with observations. The difference between the standard deviation estimates of the observed and simulated data is statistically significant for 30 models at a 95% confidence level. On the whole, the standard deviations estimates of the CMIP5 data are positively biased. These results are unsatisfactory.

3.3. Probability density

The observed data have a probability density function that can be regarded as Gaussian according to several criteria, including Kolmogorov–Smirnov's and chi-square. The simulated time series have the same Gaussian type of PDF with just two exceptions (HADGEM2-AO and IPSL-CM5B-LR). This is definitely a positive result.

3.4. AR model orders and persistence

The optimal model in this study was chosen for each time series with four order-selection criteria: Akaike's AIC, Parzen's CAT, Schwarz-Rissanen's BIC, and Hennan-Quinn's Ψ (see Bhansali, 1986; Broersen, 2000). After the linear trend removal, the observed time series x_n, $n = 1, \ldots, N$, where $N = 117$, is best approximated with an AR model of order $p = 1$ (a Gaussian Markov chain):

$$x_n \approx 0.21 x_{n-1} + a_n \tag{1}$$

where a_n is a zero mean innovation sequence (white noise) with the variance $\sigma_a^2 = 0.17$ (°C)^2. The RMS error of the coefficient estimate in Equation (1) is 0.09.

The persistence of the stochastic model (1) is very low: the relative error of a one-year lead time prediction within the Kolmogorov–Wiener theory of extrapolation is $d(1) = \sigma_a^2/\sigma_x^2 \approx 0.97$. In other words, the observed time series is very close to a white noise. Most simulated time series behave in the same manner.

The limit of statistical predictability is defined here as the lead time of a Kolmogorov–Wiener linear extrapolation at which the error variance becomes equal to 0.90 or higher. Obviously, the observed time series has a zero limit of statistical predictability (zero persistence). Most simulated time series behave in the same manner (see Table 1). This is again a positive result.

3.5. Spectral densities

The spectra of the observed and simulated time series of annual surface temperature over the CONUS were

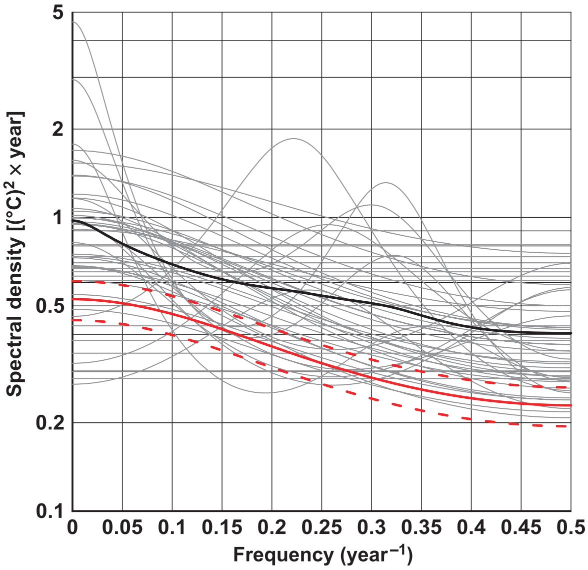

Figure 2. Maximum entropy estimates of observed (red, with 90% confidence bounds) and simulated (grey) AST spectra. The average simulated spectrum is shown with a thick black line.

obtained with the maximum entropy method (Burg, 1967; Jaynes, 1982). The algorithm includes fitting an optimal AR model to the time series and then calculating the spectral density on the basis of the selected time domain model. The approximate number of degrees of freedom ν for the spectral estimates is determined as $\nu = N/p$, where $p > 0$ is the order of the selected AR model (Ulrich and Bishop, 1975).

The spectral estimates obtained for the 47 simulated time series are shown in Figure 2 along with the spectrum of observations (the red curves) and the average of the simulated spectra (the black curve). The higher position of the average simulated spectrum happens because of the positive bias in the variance estimates. Only nine spectra of simulated data have AR orders $p = 2$ or 3 and therefore can have one or two extrema or inflection points. All other spectra are constant (white noise, 13 cases) or monotonic (Markov chain, 25 cases). With few exceptions (BNU-ESM, CMCC-CMS, GFDL-ESM2M, HADGEM2-AO, MIROC-ESM), all spectra are close to the spectrum of observations. These results should be regarded as positive.

4. Conclusions

An agreement between basic statistical moments of the observed and simulated climates is a necessary condition for recognizing the validity of numerical models of climate. The CONUS cover a sizable part of the Earth's territory – over 5% – and present a region with a practically complete coverage with instrumental observations of surface temperature since the end of the 19th century. These features make CONUS an object particularly suitable for trustworthy validations of numerical models of climate. Our analysis of time series of the average over the CONUS annual surface temperature generated with CMIP5 models showed that the mean value averaged over the ensemble is unbiased, most simulated time series have the same Gaussian PDF as the observed time series. Most CMIP5 models correctly reproduce the time and frequency domains behavior of the observed time series (including probability distribution type, persistence parameters and spectrum), which is close to the behavior of a white noise sequence. These results are satisfactory.

Yet, the linear trend rates and standard deviations of simulated annual surface temperature differ significantly from respective characteristics of the observed time series while the mean value estimates vary by as much as 6.8 °C. The trend rates are positively biased and can exceed the observed rate by as much as 220%. All mean value estimates for simulated data are statistically different from the observed mean annual temperature; they can differ from the observed mean value by more than 3 °C. The standard deviation estimates are positively biased and can exceed the observed value by 60%. These results are unsatisfactory.

References

Bhansali RJ. 1986. The criterion autoregressive transfer function of Parzen. *Journal of Time Series Analysis* **7**: 79–104.

Broersen MT. 2000. Finite sample criteria for autoregressive order selection. *IEEE Transactions of Signal Processing* **48**: 3550–3558.

Burg JP. 1967. Maximum entropy spectral analysis. In Paper presented at the 37th Meeting of Society for Exploration Geophysics, Oklahoma City, OK, Oct. 31, 1967.

Jaynes ET. 1982. On the rational of the maximum-entropy methods. *Proceedings of IEEE* **70**: 939–952.

Jones GS, Stott PA, Christidis N. 2013. Attribution of observed historical near–surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research, [Atmospheres]* **118**: 4001–4024.

Menne MJ, Williams CN Jr, Palecki MA. 2010. On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research* **115**: D11108, doi: 10.1029/2009JD013094.

Morice CP, Kennedy JJ, Rayner AA, Jones PD. 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *Journal of Geophysical Research* **117**: D08101, doi: 10.1029/2011JD017187.

Privalsky V, Croley TE. 1992. Statistical validation of GCN-simulated climates for the U.S. Great Lakes and C.I.S. Emba and Ural River basins. *Stochastic Hydrology and Hydraulics* **6**: 69–80.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* **93**: 485–498.

Ulrich TJ, Bishop TN. 1975. Maximum entropy spectral analysis and autoregressive decomposition. *Reviews of Geophysics and Space Physics* **13**: 183–200.