Credibility of climate predictions revisited

G. G. Anagnostopoulos, D. Koutsoyiannis, A. Efstratiadis, A. Christofides, and N. Mamassis
Department of Water Resources and Environmental Engineering, National Technical University of Athens (www.itia.ntua.gr)

1. Abstract

Predictions of future climate, obtained through climate models, are widely used in many scientific disciplines, and form the basis of important economic and social policies, but their reliability is rarely assessed.

In a recent paper*, the credibility of six climate models was assessed, based on comparisons with long (100 years or more) historical series of temperature and precipitation, obtained from 8 stations around the globe. The study showed that models perform poorly (worse than elementary predictions based on the time average), even at a climatic (30-year) scale, while none of the examined models proved to be systematically better than any other.

Extending this research**, we test the performance of climate models at over 50 additional stations. Furthermore, we make comparisons at a large sub-continental spatial scale after integrating modelled and observed series from 70 stations in the contiguous USA.

* Koutsoyiannis, D., Efstratiadis, A., Mamassis N. & Christofides, A. (2008) On the credibility of climate predictions. Hydrol. Sci. J.53(4), 671-684.

** See details in Anagnostopoulos, G. (2009) Assessment of the reliability of climate models, Diploma thesis supervised by D. Koutsoyiannis (in Greek), Department of Water Resources and Environmental Engineering – National Technical University of Athens, Athens.

2. Related questions

General questions

- Is climate deterministically predictable?
- Do global circulation models (GCMs) produce credible predictions of future climate for horizons of 50, 100 or even more years?
- Can such predictions serve as a basis to support decisions for important economic and social policies?
- Can the continental or global climatic projections be credible if the distributed information, from which the aggregated information is derived, is not?
- Are geographically distributed GCM predictions credible enough in order to be used in further studies at regional scales, e.g. to assess the freshwater future availability?

Specific questions addressed in this paper

- How well do current GCMs reproduce past climate (temperature and precipitation) at a local scale?
- Does integration of local predictions at a sub-continental scale improve the GCM performance in terms of reproducing past climate?

3. Methodological framework

1. **Selection of historical time series**: We selected monthly temperature and precipitation records from 55 stations worldwide (for point analysis) and 70 stations across the contiguous USA (for sub-continental analysis), according to the following criteria: (1) even geographical distribution of stations, (2) availability of data on the Internet, and (3) sample size at least 100 years, without (or with few) missing data.
2. **Selection of modelled time series**: We picked three TAR and three AR4 models and one simulation run for each model; the criterion for selecting the latter was to cover past periods rather than merely referring to future.
3. **Spatial adaptation of historical and modelled series**: For the point analysis, we extracted the monthly time series for the four grid points closest to each of the examined stations through an optimization approach, while for the sub-continental analysis we spatially integrated both the historical and the modelled series, using appropriate weighting techniques.
4. **Assessment of credibility of modelled time series**: To evaluate the performance of the GCM series against the observed ones, we made both graphical and numerical comparisons. For the latter, we used two typical statistical fitting criteria (correlation coefficient and efficiency) and also compared various statistical metrics of the two series (e.g. standard deviation, extremes, Hurst coefficient), using four time scales, i.e. monthly, seasonal, annual and climatic; the latter was assumed to be the 30-year centred moving average.

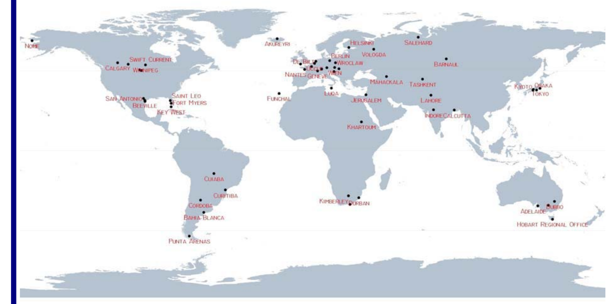
4. IPCC models and simulation runs (TAR & AR4)

- **IPCC model outputs**:
  - Three TAR and three AR4 general circulation models have been selected.
- **Simulation runs**:
  - For the TAR models we used the **SRES IS92a** scenario, most runs of which are based on historical GCM input data prior to 1989 and extended using scenarios for 1990 and beyond. For such runs, the choice of scenario is irrelevant for test periods up to 1989, while for later periods, there is no significant difference between different scenarios for the same model.
  - For the AR4 models we used the **20C3M** scenario (the only relevant with this study), generated from the outputs of late 19th and 20th century simulations from coupled ocean-atmosphere models, to help assess past climate change.

Main characteristics of the GCMs used in the study (identical to Koutsoyiannis et al.*, 2008).*

IPCC report	Name	Developed by	Resolution (°) in latitude and longitude	Grid points, latitudes × longitudes
TAR	ECHAM4/OPYC3	Max-Planck-Institute for Meteorology & Deutsches Klimarechenzentrum, Hamburg, Germany	2.8 × 2.8	64 × 128
TAR	CGCM2	Canadian Centre for Climate Modeling and Analysis	3.7 × 3.7	48 × 96
TAR	HADCM3	Hadley Centre for Climate Prediction and Research	2.5 × 3.7	73 × 96
AR4	CGCM3-T47	Canadian Centre for Climate (as above)	3.7 × 3.7	48 × 96
AR4	ECHAM5-OM	Max-Planck-Institute (as above)	1.9 × 1.9	96 × 192
AR4	PCM	National Centre for Atmospheric Research, USA	2.8 × 2.8	64 × 128

5. Meteorological stations for point analysis



Temperature time series

	Stations	Min. z (m)	Max. z (m)
Europe	18	15	569
N. America	12	6	1084
S. America	5	14	924
Asia	12	4	757
Africa	4	37	1250
Australia	4	4	275
Total	**55**	**4**	**1250**

Precipitation time series

	Stations	Min. z (m)	Max. z (m)
Europe	15	15	657
N. America	9	1	809
S. America	10	37	2408
Asia	9	6	1077
Africa	8	4	2556
Australia	4	4	432
Total	**55**	**1**	**2556**

Data source: http://climexp.knmi.nl/

6. Spatial adaptation of GCM series for point comparison

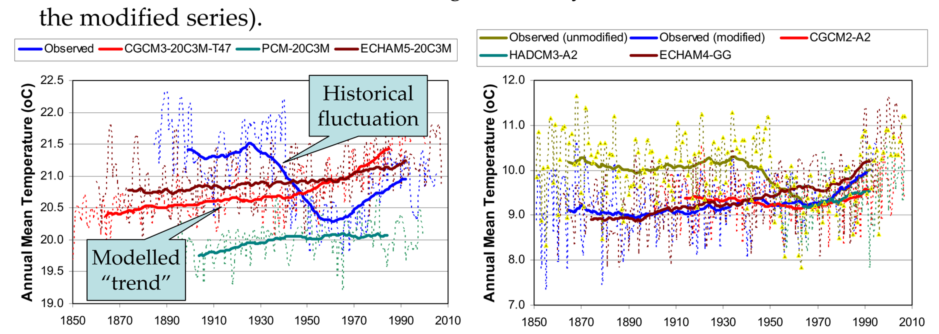
- To compare the climate model predictions with the observed time series, we interpolated GCM gridded outputs to the point of interest, using the four grid points nearest to each study location (the specific grid depends on the model).
- The generation of modelled time series for each location was based on the best linear unbiased estimation technique (BLUE), by optimizing the weighting coefficients λ_1, λ_2, λ_3, λ_4 in a linear relationship

$$\tilde{x} = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_4 x_4 \text{ (with } \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1),$$

  where $\tilde{x}$ is the best linear estimate of the monthly historical value x (i.e. $\tilde{x} - x$ is the prediction error), and x_1, x_2, x_3, x_4 are the climate model outputs for the four closest grid points.
- Optimization was implemented by maximizing the coefficient of efficiency, computed as $\mathrm{Eff} = 1 - e^2/\sigma^2$, where e^2 is the mean square error in prediction and σ^2 is the variance of the historical series. In that manner, we let the modelled time series fit the historical monthly ones as closely as possible. For physical consistency, we assumed non-negative values of the weights λ_1, λ_2, λ_3, λ_4.

7. Graphical comparisons and characteristic examples

- The majority of models (both TAR and AR4) are totally irrelevant to the observed temperature and precipitation time series, and they fail to predict the historical fluctuations at the annual and climatic time scale.
- One of the worst model performances is observed in Durban (left panel).
- One of the best model performances is observed in De Bilt (right panel) if the modified ("homogenized") time series is used in the comparison (the original observed time series, also shown in figure, is very different from both models and the modelled series).



Plots of observed and modelled annual (doted lines) and 30-year moving average (continuous lines) temperature time series at Durban, South Africa (left) and De Bilt, Netherlands (right).

8. Synoptic statistical comparisons at annual and climatic time scales – Average values for all models and locations

Annual time scale

Temperature	Correlation	Efficiency
Annual mean	0.12	-5.2
Maximum monthly	0.06	-5.3
Minimum monthly	0.03	-3.7
Annual amplitude	0.01	-4.1
Seasonal mean (DJF)	0.05	-3.9
Seasonal mean (JJA)	0.07	-7.5

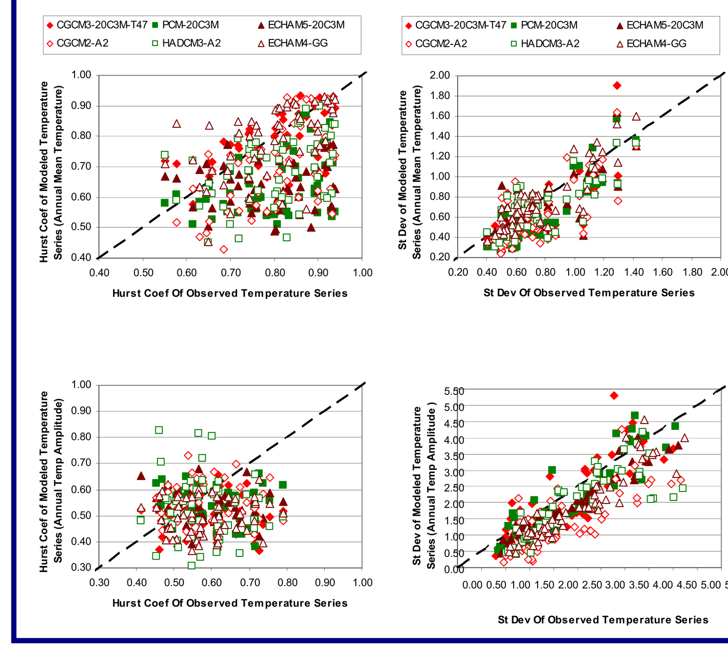
Precipitation	Correlation	Efficiency
Annual mean	0.00	-3.0
Maximum monthly	0.01	-1.3
Minimum monthly	0.00	-167.4
Seasonal mean (DJF)	0.00	-3.8
Seasonal mean (JJA)	0.00	-12.2

Climatic time scale

Temperature	Correlation	Efficiency
Annual mean	0.33	-89.0
Maximum monthly	0.21	-118.5
Minimum monthly	0.18	-117.4
Annual amplitude	0.03	-107.4
Seasonal mean (DJF)	0.24	-92.0
Seasonal mean (JJA)	0.21	-180.4

Precipitation	Correlation	Efficiency
Annual mean	0.02	-125.9
Maximum monthly	-0.02	-51.4
Minimum monthly	0.01	-5456.7
Seasonal mean (DJF)	0.05	-208.0
Seasonal mean (JJA)	-0.04	-1064.1

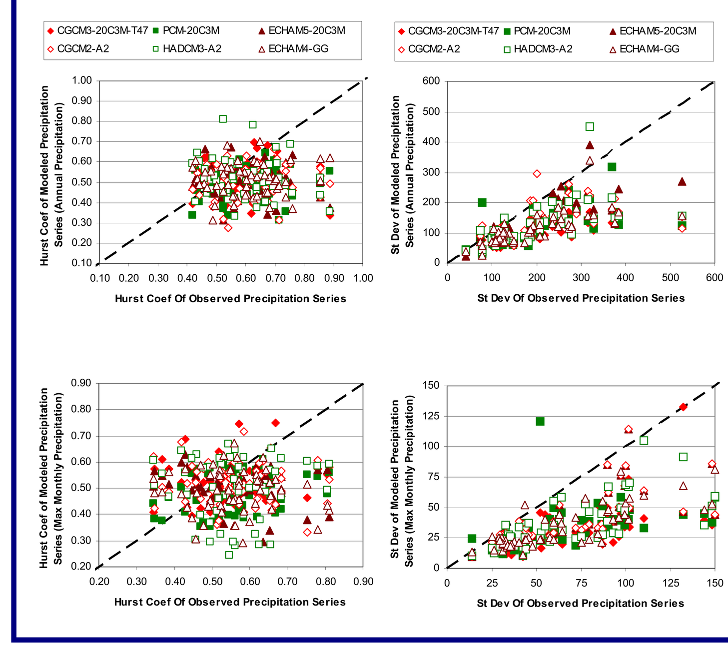
9. Comparison of variability metrics for temperature



Scatter plots of **Hurst coefficient** (left) and **standard deviation** (right) of observed vs. modelled **mean annual temperature** (underestimation in 74% of cases for Hurst and 70% for standard deviation).

Scatter plots of **Hurst coefficient** (left) and **standard deviation** (right) of observed vs. modelled **annual temperature amplitude** (underestimation in 69% of cases for Hurst and 68% for standard deviation).

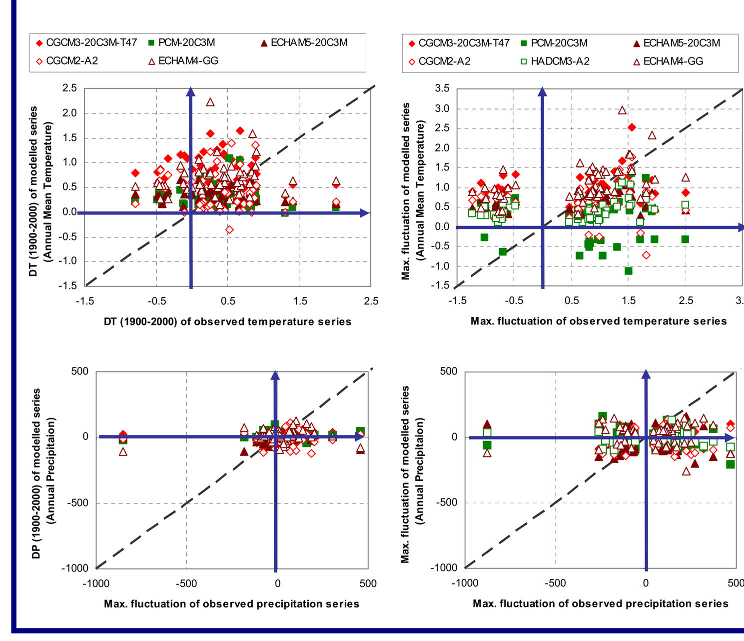
10. Comparison of variability metrics for precipitation



Scatter plots of **Hurst coefficient** (left) and **standard deviation** (right) of observed vs. modelled **annual precipitation** (underestimation in 79% of cases for Hurst and 89% for standard deviation).

Scatter plots of **Hurst coefficient** (left) and **standard deviation** (right) of observed vs. modelled **maximum monthly precipitation** (underestimation in 67% of cases for Hurst and 95% for standard deviation).

11. Do models reproduce the observed "climate changes"?

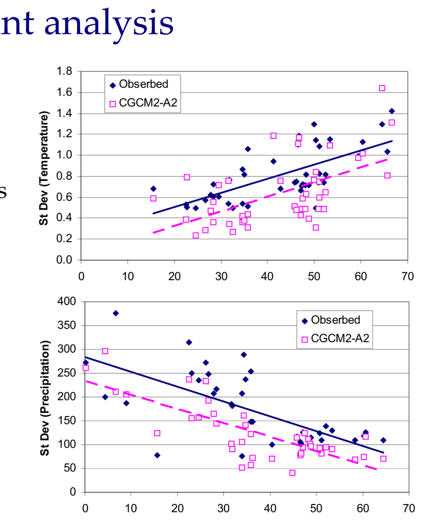


Scatter plots of **over-year difference** (left) and **maximum fluctuation** (right) during the 20th century of observed vs. modelled **30-year moving average temperature** series.

Scatter plots of **over-year difference** (left) and **maximum fluctuation** (right) during the 20th century of observed vs. modelled **30-year moving average precipitation** series.

12. General observations on point analysis

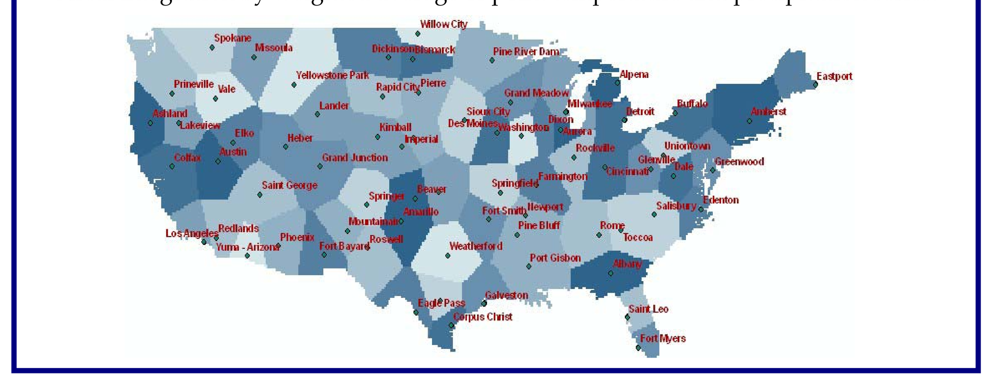
- All examined long historical records exhibit large over-year variability (i.e. long-term fluctuations), with no systematic signatures across the different locations/climates.
- At the monthly scale, although significant bias may be present, GCMs generally reproduce the broad climatic behaviours at the different locations and the sequence of wet/dry or warm/cold periods. This is expected, since models represent the seasonal variations of climatic variables, and also account for key factors such as latitude (see figure) and proximity to the sea.
- Yet, the performance of GCMs remains poor, regarding all statistical indicators at the seasonal, annual and climatic time scales; in most cases the observed variability metrics (standard deviation and Hurst coefficient), extremes (annual minima and maxima) and long-term fluctuations during the 20th century are underestimated.



Standard deviation of the annual mean temperature (up) and annual precipitation (low) with respect to the **latitude** for the observed time series and the CGCM-A2 series (northern hemisphere stations).

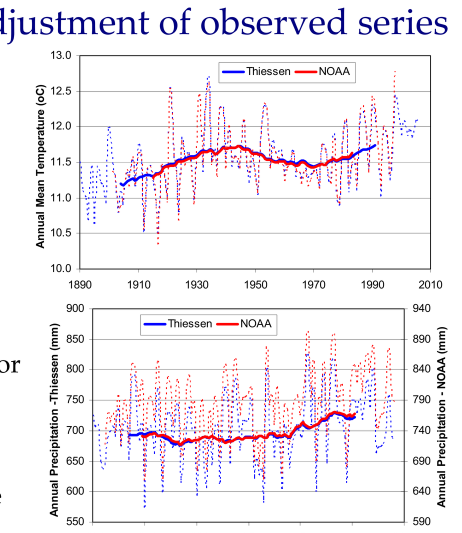
13. Sub-continental analysis: stations across USA and spatial computations

- Collection of monthly temperature and precipitation records, from 70 stations distributed across the contiguous USA (obtained from http://climexp.knmi.nl/).
- Older stations: Amherst and Detroit (continuous operation from 1836).
- Elevations: minimum +4 m (Fort Myers); maximum +2013 m (Austin).
- Mapping of stations, construction of Thiessen (Voronoi) polygons and computation of the distribution area A_i for each station i, using GIS.
- Estimation of weighting coefficients $w_i = A_i / \sum A_i$ for all stations $i = 1, \dots, 70$.
- Areal integration by weighted average of point temperature and precipitation series.



14. Spatial integration and adjustment of observed series

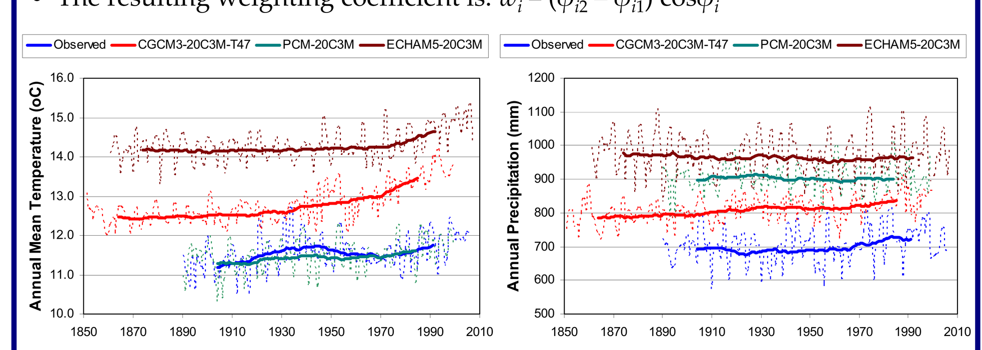
- The Thiessen integrated temperature series is adjusted for elevation by estimating (via linear regression of mean annual temperature against elevation) a temperature gradient of $\theta = -0.0038$°C/m and using the mean elevation of the contiguous USA, H = 745 m, and the weighted average of station elevations, H_m = 670 m.
- A similar correction was impossible for precipitation, since no correlation between precipitation and elevation was found.
- The areal estimations for temperature are very close to those of NOAA. Precipitation slightly differs (by 40 mm).



Comparison between areal time series of NOAA (http://www.ncdc.noaa.gov/oa/climate /research/cag3/cag3.html) and the areal time series derived by the Thiessen method (and adjusted for elevation for T).

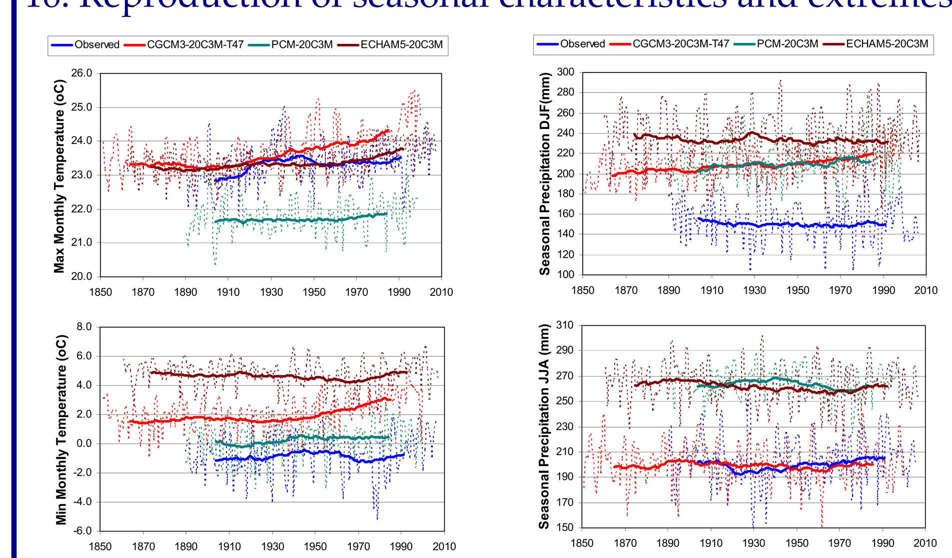
15. Spatial integration of GCM outputs and comparison with observed series

- The weights w_i were estimated on the basis of the influence area of each grid point i.
- The influence area of each grid point is a rectangle whose "vertical" (perpendicular to the equator) side is proportional to $(\phi_2-\phi_1)/2$ and its "horizontal" side is proportional to $\cos\phi$, where ϕ is the latitude of each grid point, and ϕ_2 and ϕ_1 are the latitudes of the adjacent "horizontal" grid lines.
- The resulting weighting coefficient is: $w_i = (\phi_{i2} - \phi_{i1}) \cos\phi_i$



Observed and modelled areal temperature and precipitation, on annual and climatic time scales.

16. Reproduction of seasonal characteristics and extremes



Observed vs. modelled **temperature maxima** (up) and **minima** (down) on annual and climatic time scales.

Observed vs. modelled **seasonal precipitation** (up winter; down summer) on annual and climatic time scales.

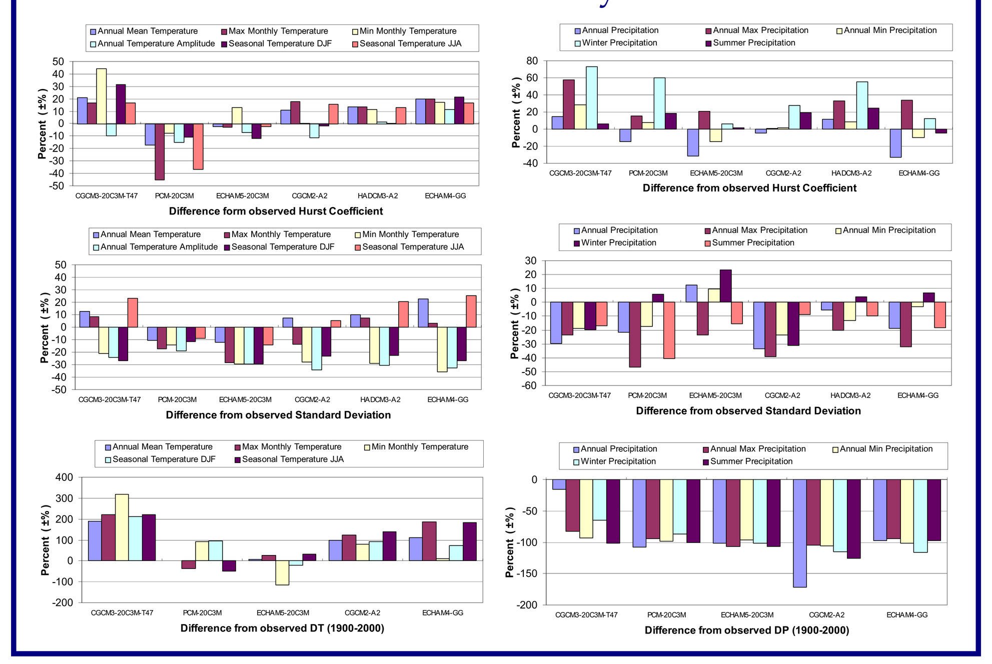
17. Detailed results for temperature series

	Period	Average (°C)	Standard deviation (°C)	Autocor-relation	Hurst coefficient	Evaluation indices against observations: Correlation	Efficiency
Annual Mean Temperature							
Observed	1890-2006	11.5	0.4	0.30	0.77		
CGCM3-20C3M-T47	1890-2006	12.9	0.5	0.67	0.93	0.19	-11.0
PCM-20C3M	1890-2006	11.4	0.4	0.21	0.63	0.30	-0.3
ECHAM5-20C3M	1890-2006	14.3	0.49	0.31	0.75	0.31	-41.0
CGCM2-A2	1900-2006	12.9	0.5	0.53	0.85	0.19	-10.5
HADCM3-A2	1950-2006	12.7	0.5	0.45	0.87	0.37	-9.4
ECHAM4-GG	1890-2006	15.1	0.5	0.69	0.92	0.40	-70.1
Max Monthly Temperature							
Observed	1890-2006	23.2	0.6	0.22	0.76		
CGCM3-20C3M-T47	1890-2006	23.7	0.7	0.48	0.89	0.02	-2.2
PCM-20C3M	1890-2006	21.7	0.5	0.02	0.42	0.17	-5.9
ECHAM5-20C3M	1890-2006	23.4	0.4	0.32	0.74	0.09	-0.5
CGCM2-A2	1900-2006	23.0	0.5	0.59	0.90	0.17	-0.6
HADCM3-A2	1950-2006	23.1	0.7	0.51	0.87	0.12	-1.4
ECHAM4-GG	1890-2006	25.1	0.6	0.60	0.91	0.08	-10.5
Min Monthly Temperature							
Observed	1890-2006	-0.7	1.4	0.04	0.52		
CGCM3-20C3M-T47	1890-2006	2.1	1.2	0.15	0.74	-0.05	-4.9
PCM-20C3M	1890-2006	0.3	1.2	-0.02	0.47	-0.13	-1.5
ECHAM5-20C3M	1890-2006	4.6	1.0	0.16	0.58	0.09	-14.4
CGCM2-A2	1900-2006	3.1	1.0	0.03	0.52	-0.09	-7.3
HADCM3-A2	1950-2006	1.9	1.0	0.01	0.57	0.11	-3.0
ECHAM4-GG	1890-2006	5.1	0.9	0.12	0.60	0.06	-16.8

18. Detailed results for precipitation series

	Period	Average (°C)	Standard deviation (°C)	Autocor-relation	Hurst coefficient	Evaluation indices against observations: Correlation	Efficiency
Annual Precipitation							
Observed	1890-2006	698.3	52.2	0.20	0.63		
CGCM3-20C3M-T47	1890-2006	815.2	36.7	0.31	0.72	0.17	-5.1
PCM-20C3M	1890-2006	901.2	41.0	0.08	0.54	0.11	-15.0
ECHAM5-20C3M	1890-2006	961.9	58.7	0.17	0.43	-0.04	-27.1
CGCM2-A2	1900-2006	971.1	34.8	0.03	0.60	-0.10	-26.4
HADCM3-A2	1950-2006	966.8	49.2	0.21	0.70	0.02	-23.3
ECHAM4-GG	1890-2006	894.2	42.4	-0.06	0.42	-0.03	-14.9
Max Monthly Precipitation							
Observed	1890-2006	81.1	8.9	0.11	0.42		
CGCM3-20C3M-T47	1890-2006	83.2	6.8	0.19	0.66	0.04	-0.6
PCM-20C3M	1890-2006	94.6	4.7	0.04	0.48	-0.05	-2.6
ECHAM5-20C3M	1890-2006	101.7	6.8	0.17	0.51	-0.03	-6.0
CGCM2-A2	1900-2006	96.9	5.4	-0.05	0.42	-0.05	-3.6
HADCM3-A2	1950-2006	97.9	7.1	0.10	0.56	-0.12	-4.8
ECHAM4-GG	1890-2006	95.6	6.1	-0.00	0.56	0.08	-3.0
Min Monthly Precipitation							
Observed	1890-2006	34.3	7.2	-0.06	0.47		
CGCM3-20C3M-T47	1890-2006	52.9	5.9	0.16	0.60	0.11	-7.3
PCM-20C3M	1890-2006	56.2	6.0	0.13	0.51	0.07	-9.9
ECHAM5-20C3M	1890-2006	59.5	7.9	-0.05	0.40	0.11	-13.1
CGCM2-A2	1900-2006	65.2	5.5	0.01	0.48	-0.08	-18.1
HADCM3-A2	1950-2006	63.7	6.3	0.01	0.51	0.02	-15.0
ECHAM4-GG	1890-2006	53.4	7.0	-0.13	0.42	-0.17	-8.3

19. Observed vs. modeled variability metrics



20. Conclusions

- The performance of the models at local scale at 55 stations worldwide (in addition to the 8 stations used in Koutsoyiannis et al., 2008) is poor regarding all statistical indicators at the seasonal, annual and climatic time scales. In most cases the observed variability metrics (standard deviation and Hurst coefficient) are underestimated.
- The performance of the models (both the TAR and AR4 ones) at a large spatial scale, i.e. the contiguous USA, is even worse.
- None of the examined models reproduces the over-year fluctuations of the areal temperature of USA (gradual increase before 1940, falling trend until the early 1970's, slight upward trend thereafter); most overestimate the annual mean (by up to 4°C) and predict a rise more intense than reality during the later 20th century.
- On the climatic scale, the model whose results for temperature are closest to reality (PCM-20C3M) has an efficiency of 0.05, virtually equivalent to an elementary prediction based on the historical mean; its predictive capacity against other indicators (e.g. maximum and minimum monthly temperature) is worse.
- The predictive capacity of GCMs against the areal precipitation is even poorer (overestimation by about 100 to 300 mm). All efficiency values at all time scales are strongly negative, while correlations vary from negative to slightly positive.
- Contrary to the common practice of climate modellers and IPCC, here comparisons are made in terms of actual values and not departures from means ("anomalies"). The enormous differences from reality (up to 6°C in minimum temperature and 300 mm in annual precipitation) would have been concealed if departures from mean had been taken.

Could models, which consistently err by several degrees in the 20th century, be trusted for their future predictions of decadal trends that are much lower than this error?