



Global annual average temperature – a precise modelling

Satyabrata Pal¹, Arunava Ghosh²

¹Former Dean, Post Graduate Studies and Professor of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Faculty of Agriculture, Mohanpur, Nadia, West Bengal, 741252, India, e-mail: satyabrata.pal@hotmail.com

²Uttar Banga Krishi Viswavidyalaya, Department of Agricultural Statistics, Pundibari, Coochbehar, West Bengal, 736165, India, e-mail: arunava_ubkv@yahoo.co.in

SUMMARY

Global annual average temperature (GAAT) is regarded as a precise indicator of the warming of the globe over the centuries, and its spectre is looming large with the passage of time and with the advancement of civilization. Global warming, caused by the accumulation of greenhouse gases in the atmosphere, has become the worst environmental threat to mankind. The phase 1981 to 2012 was the most crucial phase, and the impact of global warming in that phase indeed points to a disaster if not controlled now. Work on the building of appropriate models to represent the GAAT data can be found in the literature, although the precision levels (in terms of R^2 values) of such models do not exceed 0.86. In this paper, six models are developed by using different combinations of mathematical functions. The developed models are superior to existing models in terms of their precision. In fact, to generate such models, extensive simulation work has been carried out not only with respect to the types of mathematical functions, but also with respect to the choices of initial values of the coefficients involved in each model. The models developed here have attained R^2 values as high as 0.896.

Key words: Global warming, Greenhouse gases, Parametric modelling

1. Introduction

Global warming is a potentially catastrophic environmental phenomenon, whose impact threatens the very existence of human beings on the globe as we go forward to the future. The concern and tragedy are that human activities are the principal factors behind the alarming state which the earth is almost bound to reach in the future unless appropriate measures are devised to eliminate the underlying effects of global warming. A look at the literature on the modelling

of global/sectorial temperature reveals the existence of a few papers (so far surveyed by us), a brief overview of which follows in the next paragraph.

One of the environmental parameters that can be used to track climate change is sea surface temperature (SST) (Houghton et al. 2001). It is important to note that time series data on SST anomalies exhibit much short-term natural variability, and thus the goal should be to make an assessment of the longer-term changes in temperature over the period of data coverage. Trend models are generally applied to surface or air temperature data, and the finding is that linear trend and breakpoint models are appropriate to analyse the temperature data on a global scale; see for example Zheng and Basher (1999), Siedel and Lanzante (2004). Woodward and Gray (1993) suggest the potential applicability of random walk models with no deterministic trend. Siedel and Lanzante (2004) have performed a detailed study employing flat-step models, piecewise linear models, sloped-step models, and linear models with fits to a number of datasets. It is mentioned in the literature that in the case of the 1979–2001 datasets, a linear trend with the first-order autoregressive [AR(1)] model has been found to best explain data records. The method employed by Good et al. (2007) is equivalent to that of Lawrence et al. (2004) in that the seasonal cycle and the influence of El Niño are removed from the SST datasets, allowing the trend due to other natural processes and anthropogenic influences to be determined. However, rather than performing these steps separately as in Lawrence et al. (2004), the authors adapt the formalism of Weatherhead et al. (1998) to find the seasonal cycle, the influence of El Niño and the trend simultaneously, and then to calculate the error on the trend. Their method of analysis consists of a number of steps. In the process of detecting climate change, Casey and Cornillon (2001) have obtained the existence of a warming trend of between 0.09° and 0.14°C per decade (depending on dataset and averaging approach). Houghton et al. (2001) quote a global increase of 0.10°C to 0.14° per decade between 1976 and 2000, from Jones et al. (2001). In their paper, Good et al. (2007) modelled the Global Sea Surface Temperature and found a warming trend of between 0.09° and 0.14°C per decade (depending on dataset and

averaging approach). Fomby and Vogelsang (2001) presented an application of Size-Robust Trend Statistics to global-warming temperature series.

In this paper the premise and caveat are that the dataset on global average annual temperature available on NASA's site is regarded as the prime base, on which extensive modelling work has been undertaken. Considering the partitions (over the entire time trajectory of 132 years, 1880–2012, mentioned in Pal and Pal (2011), Pal et al. (2013), and Pal (2014)), it is observed that the third phase (1981–2012) is the most critical period; reasons are explained in the aforementioned papers from 2011 and 2013. The present study is referenced to this phase, and efforts have been undertaken to develop models (combinations of polynomials, cubic or bi-quadratic, trigonometric, exponential, and with or without power function) which are superior (in terms of having greater precision levels, as measured by R^2 values) to the models available in the literature so far surveyed. After an intensive quest for superior models, six models have been developed, each of which has an R-square value equal to or greater than 0.84. The diagnostic test results concerning the building of the models confirm that the assumptions on independence and normality hold unequivocally. Table 1 presents the forms of the estimated model equations of the precise models. Table 2 includes the results on diagnostic checks for each of the models. Table 3 contains values of the alternative precision criteria in respect of each model. Graph plots are given of the observed and predicted values (also shown in a table) with respect to each of the six models. Each model expression is identified by being assigned a distinct symbol.

Interested readers may consult the book by Draper and Smith (1998) for gaining a moderately advanced knowledge of the different aspects of statistical modelling.

It is our experience as modellers that very fine (precise) parametric models do not always provide very precise nearest - future forecasts-that is why forecasts have not been included. Moreover, it is observed that the third phase (1981–2012) is the most critical and crucial period (Pal and Pal, 2011), as this phase imbibes maximum temperature variation and occurrence of maximum

peak temperature. Thus the 32- year period deserves serious attention for the humanity to be concerned about the disastrous effects of global warming.

Section 2 contains a description of the source of data, the sources being available online. Section 3 presents the method employed in the paper. Section 4 presents the results and subsequent detailed discussion of the findings obtained. Extensive simulation is carried out with respect to the inclusion of particular mathematical functions in the model equation, in addition to work on the choice of initial values. Among the developed models a parsimonious model is identified. Finally the appendix contains the six graph plots (with observed values and the corresponding fitted values) for the six models. To conclude, it can be noted that the models developed have confirmed their precision and significance.

2. Materials

The data sources are:

IPCC Report 2007 –

http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml#1

Temperature Data Source –

http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts+dSST.txt
www.nature.com/news/earth-summit-rio-report-card-1.10764

3. Methods

The general structure of the model equation is:

$$Y(t) = a + bt + ct^2 + dt^3 + pt^4 + he^t + q.\sin^u(wt) + r.\cos^v(zt),$$

where a, b, c, d, p, h, q, u, v are the coefficients involved. Estimation of the parameters and development of the models are performed using SAS (9.2) software, PROC NLIN module. Theoretical details are not included. The

methods adopted here are distinctly different from those used in the papers cited in the introduction.

4. Results and discussion

Table 1 presents the estimated model equations. After an extensive permutation with regard to the types of mathematical functions to be included and their subsequent mixing to generate the models, the above six models were obtained. Two types of distinct combinations (one type with 9 coefficients and the other type involving 10 coefficients) were developed.

Table 1. Estimated model equations

Model identity	Model Expression	R ² value
M1	$Y=14.17-0.0058t+0.0019t^2-0.00004t^3-101E-17e^t+0.15\sin^5(-48.45t)+0.099\cos^3(1.875t)+\varepsilon$	0.896
M2	$Y=14.29-0.058t+0.0075t^2-0.00029t^3+3.771E-6t^4-239E-17e^t-0.065\sin^3(1311465t)+0.122\cos^6(29368-5t)+\varepsilon$	0.867
M3	$Y=14.44-0.107t+0.015t^2-0.00065t^3+9.716E-6t^4+387E-17e^t+0.08\sin^3(-163031t)-0.118\cos^4(23009.2t)+\varepsilon$	0.863
M4	$Y=-3.39-0.258t+0.156t^2-0.007t^3+0.00009t^4-452E-17e^t-0.06\sin^3(431219t)+17.86\cos^6(9600.7t)+\varepsilon$	0.853
M5	$Y=14.48-0.181t+0.035t^2-0.0018t^3+0.000027t^4-221E-17e^t-0.704\sin^3(-923.5t)-0.145\cos^5(0.314t)+\varepsilon$	0.842
M6	$Y=14.2713-0.0258t+0.0039t^2-0.00013t^3+1.27E-6t^4-61E-17e^t-0.1323\sin^2(2.2841t)+0.1037\cos^4(0.3692t)+\varepsilon$	0.892

In what follows, the results obtained (as given in the above three tables) are commented on at length.

The six models are identified as M1, M2, M3, M4, M5 and M6 respectively. The R-square values corresponding to the different models range from 0.842 to 0.896 (approx. 0.84 to 0.90).

Table 2 contains information on the diagnostic checking (with respect to independence and normality) of the six models. For the test of independence, run test and test for normality, the KS test and SW test were used. The results corresponding to the above tests performed on the model residuals confirm the validity of the above two assumptions (independence and normality) as revealed from the corresponding p-values.

M1 and M6 are the two best models (M1 is also parsimonious), M2 and M3 are the next best models, and M4 and M5 are the third best models obtained by us. All models have high levels of precision.

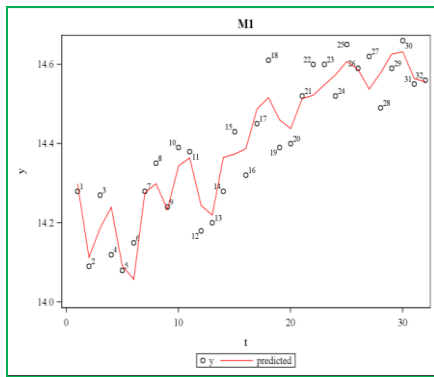
Table 3 presents values corresponding to alternative precision criteria (MSE, MAE, etc.), and the results obtained are of a similar nature.

Table 2. Diagnostic checks for six models

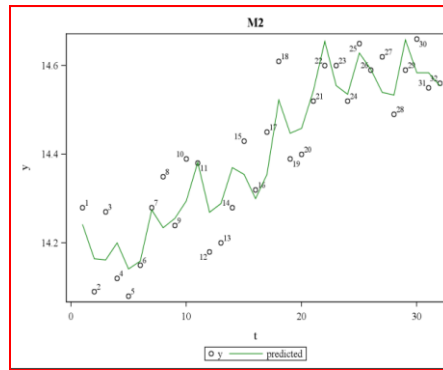
Model	Test for normality				Test for Independence	
	Shapiro-Wilk test		Kolmogorov-Smirnov		Run test	
	Statistic (W)	p-value	Statistic (D)	p-value	Z-value	Asymp.Sig.(2-tailed)
M1	0.971427	0.5397	0.082423	>0.1500	-0.158	0.875
M2	0.932289	0.0454	0.12636	>0.1500	0.639	0.523
M3	0.973912	0.6135	0.091597	>0.1500	-0.519	0.604
M4	0.966218	0.4021	0.123531	>0.1500	0.639	0.523
M5	0.973877	0.6125	0.092581	>0.1500	0.564	0.573
M6	0.9830	0.877	0.069	>0.1500	0.274	0.784

Table 3. Values of different precision criteria for six models

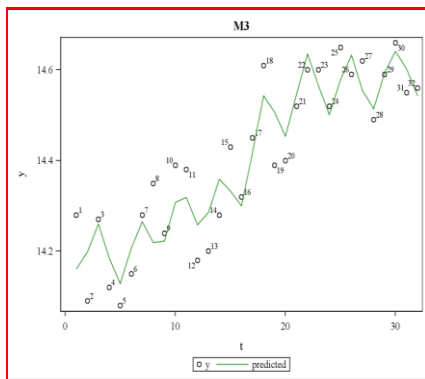
Model Identity	MSE (Mean square error)	R-square	MAE (Mean absolute error)	F ₀	Pr (F ₀ >F)
M1	0.00444	0.896	0.046543	24.92	<.0001
M2	0.00597	0.867	0.054267	15.91	<.0001
M3	0.00612	0.863	0.054906	15.46	<.0001
M4	0.00660	0.853	0.049915	14.17	<.0001
M5	0.00706	0.842	0.058452	13.08	<.0001
M6	0.00482	0.892	0.047418	20.27	<.0001



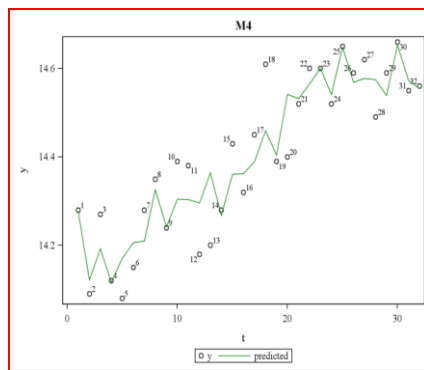
Graph I



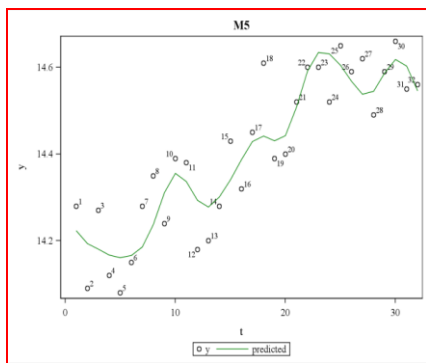
Graph II



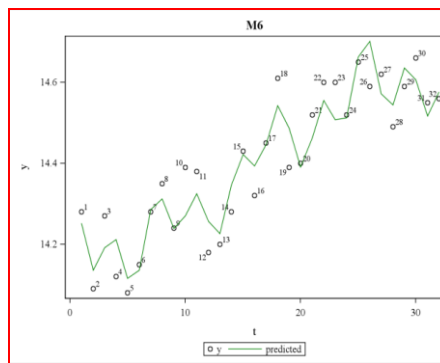
Graph III



Graph IV



Graph V



Graph VI

Figure 1. Graph plots corresponding to the six models

Graph plots corresponding to the six models are presented in Figure 1. The predicted values (as obtained using the models) are found to be closest to the observed values in the cases of M1 and M6. Typically, other models also displayed high degrees of closeness between the observed and fitted values.

Acknowledgement

The authors are thankful to the Referees for their comments.

REFERENCES

- Casey K.S., Cornillon P. (2001): Global and regional sea surface temperature trends. *J. Climate* 14: 3801–3818.
- Draper N.R., Smith H. (1998): *Applied Regression Analysis*, John Wiley and Sons.
- Fomby T.B., Vogelsang T.J. (2001): The Application of Size-Robust Trend Statistics to Global-Warming Temperature Series. *J. Climate* 14.
- Houghton J.T., Ding Y., Griggs D.J., Noguer M., van der Linden P.J., Dai X., Maskell K., Johnson C.A., Eds. (2001): *Climate Change: The Scientific Basis*. Cambridge University.
- Jones P.D., Osborn T.J., Briffa K.R., Folland C.K., Horton E.B., Alexander L.V., Parker D.E., Rayner N.A. (2001): Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.* 106: 3371–3380.
- Lawrence S.P., Llewellyn-Jones D.T., Smith S.J. (2004): The measurement of climate change using data from the Advanced Very High Resolution and Along Track Scanning Radiometers. *J. Geophys. Res.* 109.
- Pal S., Pal S. (2011): A Revisit to the Global Warming Phenomenon. Proceedings (published online) of the 58th World Statistics Congress organised by the International Statistical Institute, Hague, held at Dublin, Ireland.
- Pal S., Pal S., Kageyama S. (2013): Modeling the Global Mean Temperature. *Bull. Hiroshima Inst. Tech. Research* 47: 149–152.
- Pal S., Ghosh A., Kageyama S. (2014): Revisit to Modelling Global Annual Average Temperature – A Parametric Approach. *Bull. Hiroshima Inst. Tech. Research* 48 (to appear).
- Woodward W.A., Gray H.L. (1993): Global warming and the problem of testing for trend in time series data. *J. Climate* 6: 953–962.
- Zheng X., Basher R.E. (1999): Structural time series models and trend detection in global and regional temperature series. *J. Climate* 12: 2347–2358.