Bull. Hiroshima Inst. Tech. Research Vol.49（2015）89‒95

Article

Global Temperature-Quest for Models with Superior Fitting

Satyabrata PAL*, Arunava GHOSH**, Sanpei KAGEYAMA*** and Kazuki MATSUBARA****

（Received September 22, 2014）

Abstract

The record on the data on global annual average temperature over the last thirty two years presents a continuous increase (though very mild) in temperature over years and this warming of the globe is, indeed, a matter of deepest concern to all of us at the moment. It is required to develop appropriate measures to contain the temperature-rise and when such measures (which perceptibly and most satisfactorily address the problem) are created, it is our mandate to implement those holistically in order to save the human civilization from a perilous future. Global temperature models expose the inherent trend existing in the temperature-record. A literature-survey on modelling of global temperature reveals the following articles, Pal and Pal, 2011, Pal *et al*. (2013) and Pal *et al*. (2014), wherein the precision levels (in terms of R^2 values) of the models did not exceed 0.89.This paper presents superior global annual average temperature models (parametric based on mathematical functions) with precision levels (in terms of R^2 values) varying in the range, 0.875 to 0.940.

Key Words: Greenhouse gases, Global warming, Parametric models

1. Introduction

The phenomenon of global mean warming observed annually since the past 19thcentury is far outside the range of observational uncertainty in global temperature datasets, and there is therefore no shade of doubt that the world has warmed. A wide range of observed climate indicators continues to show changes that are consistent with a globally warming world, and our understanding of how the climate system responds to the rising greenhouse gas levels needs to be enriched and development of appropriate measures is the need of the hour. The analysis of the nature and causes of climate change is based on comprehensive observations of the climate system in combination with theoretical understanding of its physics and general circulation models themselves based on the fundamental laws of physics. Sophisticated statistics are used to demonstrate the significance of recent changes in the climate system. In understanding the climate change a faithful analysis of the record of global temperatures partially accounts for the impact of such changes.

This paper is devoted to a formulation of most superior models, the premise and caveat are that the data set on global average annual temperature available on NASA's site is regarded as the prime base on which an extensive modelling work has been under-

* Sampling and Official Statistics Unit, Indian Statistical Institute, Kolkata, 700108, India,
e-mail: satyabrata.pal@hotmail.com
** Uttar Banga Krishi Viswavidyalaya, Department of Agricultural Statistics, Pundibari, Coochbehar, West Bengal, 736165, India,
e-mail: arunava_ubkv@yahoo.co.in
*** Department of Environmental Design, Hiroshima Institute of Technology, Hiroshima, 731-5193, Japan,
e-mail: s.kageyama.4b@it-hiroshima.ac.jp
**** Matsunaga High School, Fukuyama, 729-0112, Japan
e-mail: additivestructure@gmail.com

taken. Recalling from the partitions (in entire time-trajectory of 132 years, 1880-2012, mentioned in Pal and Pal, 2011, Pal *et al*., 2013 and Pal *et al*., 2014), it is observed that the third phase (1981-2012) is the most critical period; reasons had been explained in the above papers appeared in 2011 and 2013 respectively. The study undertaken in this paper refers to this phase, and most superior models, which are structured by taking into account combinations of fourth-degree polynomial functions, exponential functions and trigonometric functions on the time variable '*t*', have been presented. Such models 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. The developed models (five in numbers and represented by the symbols, M_1, M_2, M_3, M_4 and M_5) have their precision levels (in terms of R^2 values) varying in the range, 0.875 to 0.940.The diagnostic test results concerning the built-up of the models confirm that the assumptions on independence and on normality hold unequivocally.

Section 2 is devoted to a description of the source of data, the sources being available on net as mentioned. Section 3 presents the method employed in the paper. Section 4 presents the results and subsequent discussions in details on the findings evolved in the paper. Extensive simulation works with respect to inclusion of the mathematical function in the model equation and also works with respect to choice of initial values have been made. Table 1 presents the expressions of the estimated model equations of the superior models. Table 2 contains the observed data-set (on NASA's site as mentioned above) the predicted values corresponding to the above five Models M_1 to M_5. Table 3 presents the standard errors of the estimates and 95% confidence limits. Table 4 includes the results on the diagnostic checks for each of the five models developed and Table 5 lists the values in respect of the alternative precision criteria (including R^2 values) for each of the five models. The last section, Annexure, displays the five graph-plots (with observed values and the corresponding fitted values) in respect of the five models. In fine, it can be mentioned that the developed models have affirmed their superiority and importance.

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

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

where *a, b, c, d, p, f, h, q, u, w, r, v* and *z* are the coefficients involved, and ε_t's are errors and follow NID (0, σ^2).

The estimation of the parameters and development of the models are done using SAS (9.2) Software developing programmes on PROC NLIN module. PROC NLIN module (SAS) is a set of procedural codes for the parameter estimation in case of nonlinear model in Statistical Analysis System (SAS) Software. These Codes have been used to obtain the five model equations by applying SAS package.(It is needed to mention that extensive search in regard to model coefficients and also in regard to initial values have been made). The methods adopted here are distinctly different from the other papers cited in Section 1. This optimum choice has contributed undoubtedly to increasing the precision (in terms of R^2 values) of the models.

4. Results and Discussions

In Table 1, five models M_1, M_2, M_3, M_4 and M_5 (with model expressions) are given. These models are obtained by simulation (employing extensive trial and error procedure) keeping in mind the basic objectives (to minimise MSE and to maximise R^2 values) so as to generate superior models. It is evident that the generated models represent superior models.

Table 1: Estimated model equations

Models	Equations
M_1: (u=3, v=6)	$y = 13.91 + 0.105t - 0.011t^2 + 0.00057t^3 - 9.58\text{E-}6t^4 - 373(-0.00115)^t + 5.59\text{E-}15e^t + 0.106\sin^3(60430t) - 0.117\cos^6(-29514.8t) + \varepsilon$
M_2: (u=3, v=6)	$y = 14.10 + 0.015t - 0.00089t^2 + 0.000091t^3 - 2.03\text{E-}6t^4 - 0.6298(-0.3239)^t - 193\text{E-}18e^t - 0.107\sin^3(60433.3t) + 0.1013\cos^6(-19461.8t) + \varepsilon$
M_3: (u=7, v=2)	$y = 14.05 + 0.028t - 0.0026t^2 + 0.000175t^3 - 3.48\text{E-}6t^4 - 0.688(-0.3254)^t + 2.25\text{E-}15e^t + 0.1098\sin^7(619659t) + 0.093\cos^2(-150331t) + \varepsilon$
M_4: (u=3, v=6)	$y = 14.08 + 0.045t - 0.00435t^2 + 0.00025t^3 - 4.48\text{E-}6t^4 - 0.2208(-0.5397)^t + 6.97\text{E-}16e^t - 0.0674\sin^3(60429.9t) - 0.0973\cos^6(-29415.5t) + \varepsilon$
M_5: (u=7, v=3)	$y = 13.909 + 0.0447t + 0.00252t^2 - 0.00028t^3 + 5.705\text{E-}6t^4 - 0.2363(-0.6405)^t - 123\text{E-}17e^t - 0.0670\sin^7(619647t) - 0.1961\cos^3(-219465t) + \varepsilon$

For example, 123E-17 means 123.10^{-17}.

Table 2: Predicted values under different models

Year	Observed values	Predicted values				
		M_1	M_2	M_3	M_4	M_5
1981	14.28	14.2780	14.2774	14.2783	14.2935	14.2995
1982	14.09	14.0802	14.0773	14.0819	14.0882	14.0966
1983	14.27	14.2447	14.2640	14.2668	14.1830	14.201
1984	14.12	14.1884	14.1449	14.1321	14.1318	14.1053
1985	14.08	14.1120	14.0809	14.0742	14.2271	14.1339
1986	14.15	14.1586	14.1963	14.2013	14.1754	14.1926
1987	14.28	14.2563	14.3211	14.3223	14.2568	14.2725
1988	14.35	14.2591	14.2896	14.2843	14.2344	14.3148
1989	14.24	14.2349	14.2382	14.2421	14.2601	14.3498
1990	14.39	14.3338	14.2860	14.2898	14.2839	14.292
1991	14.38	14.3785	14.2931	14.3110	14.3599	14.2711
1992	14.18	14.2572	14.2266	14.2621	14.2434	14.1884
1993	14.2	14.2003	14.2710	14.2532	14.2640	14.2641
1994	14.28	14.3598	14.3866	14.3222	14.3743	14.3023
1995	14.43	14.4186	14.3969	14.3876	14.4377	14.3994
1996	14.32	14.3626	14.3766	14.4168	14.3072	14.4431
1997	14.45	14.4262	14.4578	14.4528	14.4009	14.4637
1998	14.61	14.5199	14.5261	14.4879	14.5202	14.4769
1999	14.39	14.4107	14.4342	14.4690	14.4885	14.4196
2000	14.4	14.3479	14.3515	14.4312	14.3633	14.4742
2001	14.52	14.5333	14.4795	14.4665	14.5143	14.4776
2002	14.6	14.6786	14.6142	14.5418	14.6067	14.5541
2003	14.6	14.6101	14.5639	14.5488	14.5607	14.5548
2004	14.52	14.5224	14.5324	14.5399	14.5174	14.5889
2005	14.65	14.5834	14.6431	14.6334	14.6492	14.609
2006	14.59	14.6307	14.6916	14.7163	14.6354	14.5803
2007	14.62	14.5780	14.5746	14.6180	14.5752	14.6144
2008	14.49	14.5069	14.5013	14.4893	14.5627	14.5839
2009	14.59	14.5884	14.5908	14.5499	14.6150	14.6215
2010	14.66	14.6638	14.6337	14.6439	14.5718	14.5594
2011	14.55	14.5585	14.5586	14.5680	14.5792	14.573
2012	14.56	14.5564	14.5602	14.5573	14.5588	14.5619

Table 3: Standard errors of estimates and 95% confidence limits

Model coeffi-cients	M_1			M_2			M_3			M_4			M_5		
	SE	LCL	UCL	SE	LCL	UCL	SE	LCL	UCL	SE	LCL	UCL	SE	LCL	UCL
a	0.15	13.605	14.223	0.112	13.868	14.335	0.127	13.792	14.322	0.112	13.852	14.32	0.127	13.64	14.17
b	0.05	−0.005	0.216	0.044	−0.077	0.108	0.051	−0.077	0.133	0.046	−0.051	0.14	0.046	−0.051	0.14
c	0.005	−0.024	0.001	0.005	−0.012	0.01	0.006	−0.015	0.0100	0.006	−0.016	0.007	0.007	−0.012	0.016
d	0.0003	0.00001	0.001	0.0002	−0.0004	0.0006	0.0003	−0.0004	0.0007	0.0003	−0.0003	0.0008	0.0004	−0.001	0.0005
p	4.052E-6	−0.00002	−1.13E-6	3.877E-6	−0.00001	6.056E-6	4.361E-6	−0.00001	5.616E-6	4.265E-6	−0.00001	4.415E-6	6.674E-6	−8.22E-6	0.00002
f	65344.2	−136679	135933	0.906	−2.519	1.26	1.121	−3.026	1.65	0.332	−0.912	0.47	0.212	−0.678	0.206
g	0.202	−0.423	0.42	0.342	−1.038	0.39	0.379	−1.118	0.467	0.454	−1.487	0.408	0.262	−1.186	−0.095
h	2.09E-15	1.23E-15	9.95E-15	2.21E-15	−48E-16	4.42E-15	2.51E-15	−298E-17	7.47E-15	2.55E-15	−462E-17	6.01E-15	3.25E-15	−8E-15	5.55E-15
q	0.019	0.067	0.145	0.023	−0.155	−0.059	0.038	0.031	0.189	0.027	−0.123	−0.011	0.034	−0.1381	0.0045
r	0.031	0.056	0.182	0.037	−0.179	−0.024	0.039	−0.175	−0.011	0.041	0.011	0.184	0.09	−0.384	−0.008
w	0.009	60430.0	60430.1	0.008	60433.3	60433.3	0.008	619659	619659	0.015	60429.8	60429.9	0.127	13.65	14.17
z	0.006	−29514.8	−29514.8	0.008	−19461.8	−19461.8	0.011	−150331	−150331	0.009	−29415.5	−29415.5	0.046	−0.05	0.14

SE = standard error; LCL = lower confidence limit; UCL = upper confidence limit

[Note: Approximate 95% confidence limits (upper and lower) with respect to any estimator of a parameter may sometimes be equal. It happens when the precision level of the estimator (here any coefficient in the model equation) is extremely high. In such cases the length of the confidence interval may collapse to zero].

Table 4: Diagnostic checking for five models

Models	Test for Normality				Test for Independence	
	Shapiro-Wilk test		Kolmogorov-Smirnov test		Run test	
	Statistic	p-value	Statistic	p-value	z-value	Asymp.Sig. (2-tailed)
M_1: (u=3, v=6)	0.957389	0.2326	0.11688	> 0.1500	0.000	1.000
M_2: (u=3, v=6)	0.979969	0.7987	0.10691	> 0.1500	0.203	0.839
M_3: (u=7, v=2)	0.984552	0.9156	0.10493	> 0.1500	0.639	0.523
M_4: (u=3, v=6)	0.977162	0.7139	0.09738	> 0.1500	0.203	0.839
M_5: (u=7, v=3)	0.985271	0.9299	0.07678	> 0.1500	0.000	1.000

Table 5: Values of different precision criteria for four models

Models (*edf* = 20)	MSE (mean square error)	R^2 values	MAE (mean absolute error)	F_o	Pr ($F>F_0$)
M_1: (u=3, v=6)	0.00597	0.940	0.03155	28.39	< 0.0001
M_2: (u=3, v=6)	0.00388	0.921	0.037197	21.29	< 0.0001
M_3: (u=7, v=2)	0.00481	0.902	0.041204	16.83	< 0.0001
M_4: (u=3, v=6)	0.00576	0.883	0.045301	13.75	< 0.0001
M_5: (u=7, v=3)	0.00615	0.875	0.049199	12.77	< 0.0001

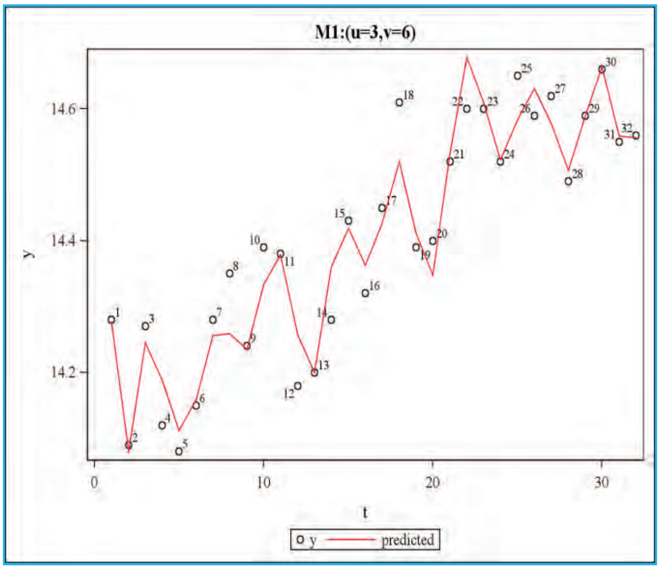
edf = error degrees of freedom

This section contains the results obtained after fitting the models to the data. These are given in five tables, Tables 1 to 5 respectively. Tables 1 to 3 display the model equations, observed and predicted values, standard errors of the estimates and confidence limits respectively. Tables 4 and 5 present the results on the diagnostic checks (Draper and Smith, 1998) and values of the alternative precision criteria respectively in respect of each model. The models developed in this communication are the most precise models (each of four models has its precision level more than 0.88, each of three models has its precision level more than 0.90, and each of two models has its precision level more than 0.92 and there is 1 (one) model which has its precision level as high as 0.94, the precision levels are determined on the basis of R^2 values) existing in the literature (so far surveyed by us).

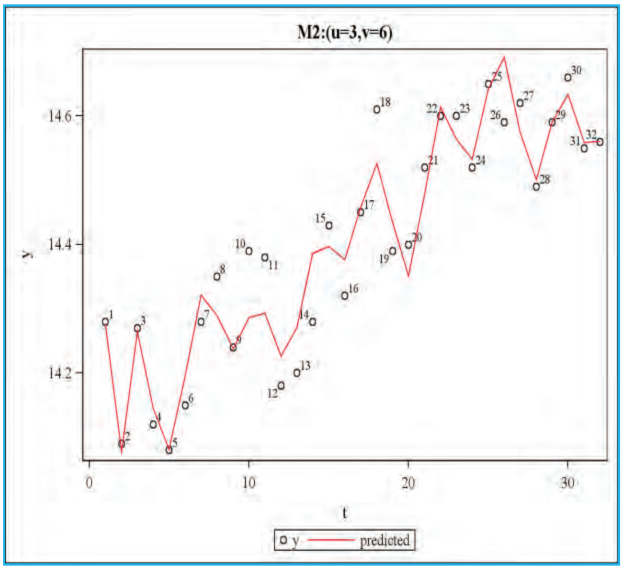
Annexure

Here, the subscript, "t", denotes year, e.g., t = 1,2,3, ... mean first, second, third, ... years, respectively. O^i represents the observed value of temperature at the i-th point of time, i.e., y_i.

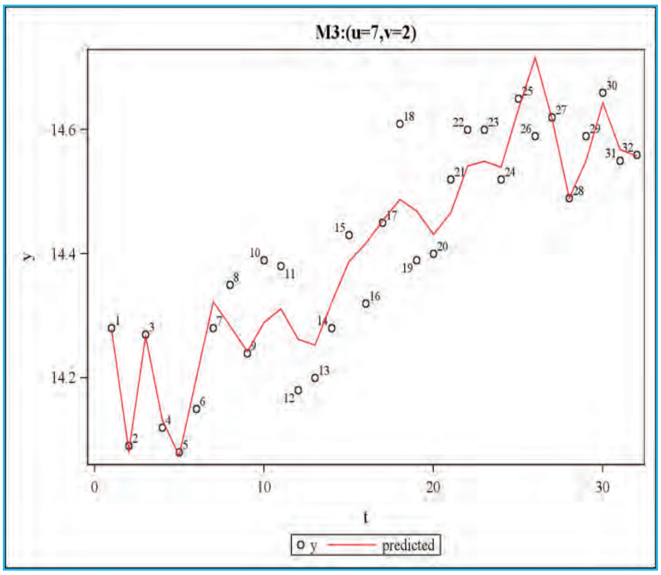
M_1



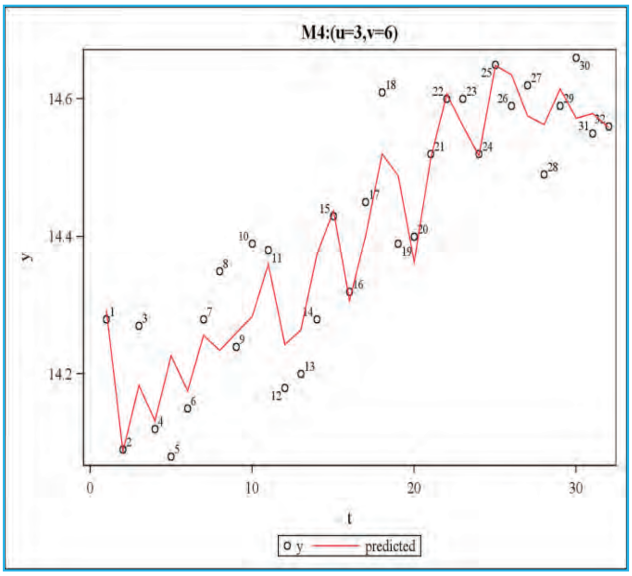
M_2



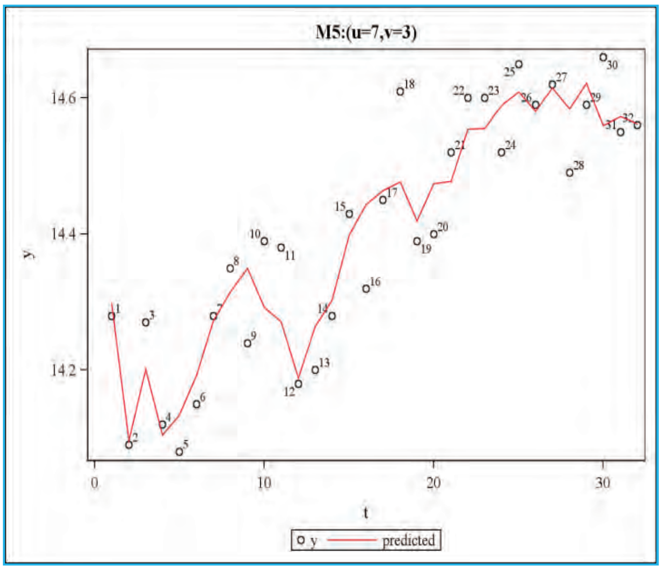
M_3



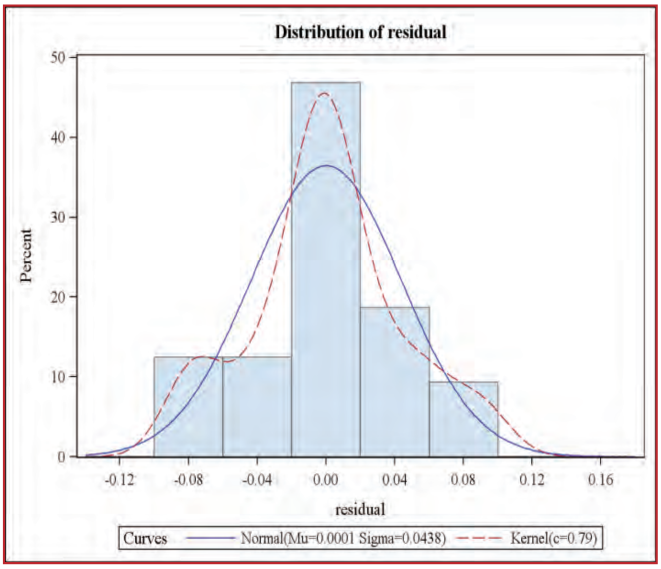
M_4



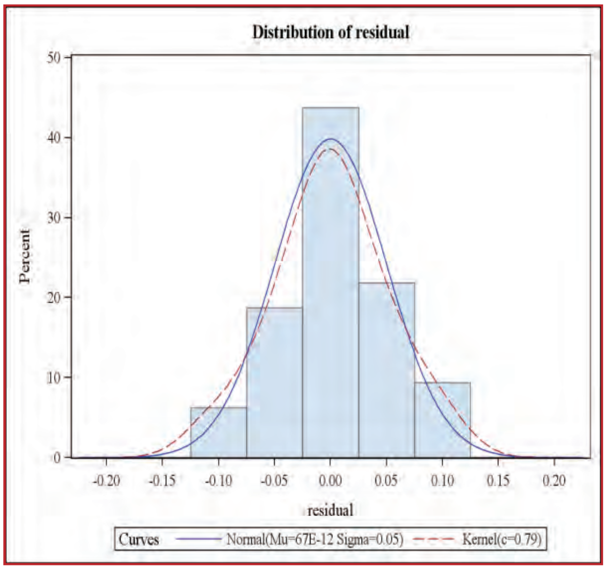
M_5


M5:(u=7,v=3)
y
t
y
predicted

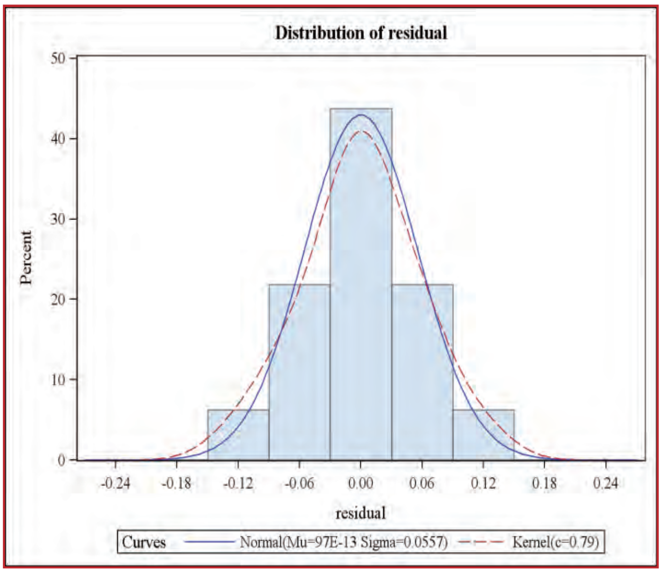
M_1


Distribution of residual
Percent
residual
Curves
Normal(Mu=0.0001 Sigma=0.0438)
Kernel(c=0.79)

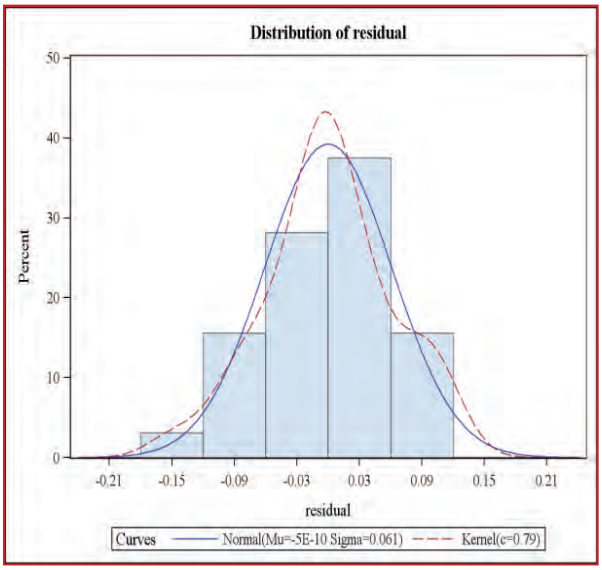
M_2


Distribution of residual
Percent
residual
Curves
Normal(Mu=67E-12 Sigma=0.05)
Kernel(c=0.79)

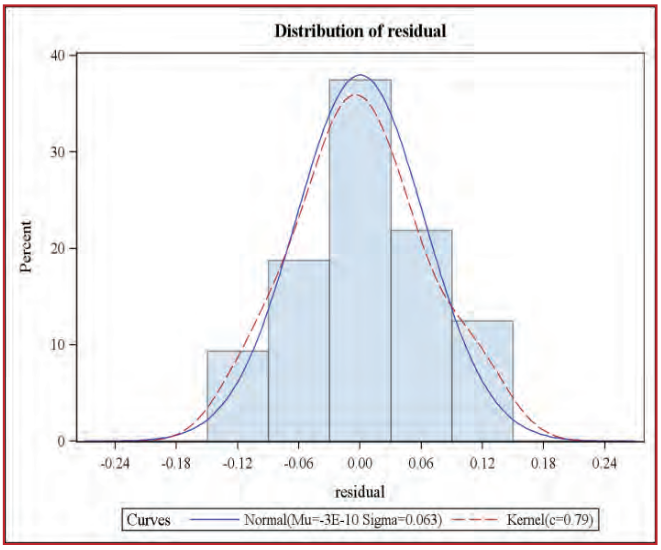
M_3


Distribution of residual
Percent
residual
Curves
Normal(Mu=97E-13 Sigma=0.0557)
Kernel(c=0.79)

M_4


Distribution of residual
Percent
residual
Curves
Normal(Mu=-5E-10 Sigma=0.061)
Kernel(c=0.79)

M_5



References

Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*, John Wiley and Sons.

Pal, S. and 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. and Kageyama, S. (2013). Modelling the Global Mean Temperature. *Bull. Hiroshima Inst. Tech. Research* 47, 149–152.

Pal, S., Ghosh, A., Pal, S. and Kageyama, S. (2014). Revisit to Modelling Global Annual Average Temperature - A Parametric Approach. *Bull. Hiroshima Inst. Tech. Research* 48, 87–91.