RESEARCH REVIEW JOURNALS
Volume-03
Issue-06
June-2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Splines for annual temperature Data in India

[1]A. Srinivasulu, [*2]B. Sarojamma & [3]K. Anil kumar

[1]*Research Scholar, Department of Statistics, S.V.University, Tirupati, A.P. (India)*
[*2]*Assistant professor, Department of Statistics, S.V.University, Tirupati, A.P. (India)*
[3]*Assistant professor, Department of Mathematics, GITAM University, Hyderabad, Telangana (India)*

ARTICLE DETAILS

Article History
Published Online: 12 June 2018

Keywords
Annual Temperature data, Chow test, Cubic spline, Harmonic spline, Regression spline, Quadratic spline

[*]Corresponding Author
Email: saroja14397[at]rediffmail.com

ABSTRACT

Spline algorithms are the way to fit data points with a set of connecting curves (each one is called a Spline), such that the values between data points can be computed. They are various types/order of equations that can be used to specify the splines including Linear, Quadratic, Cubic, etc. Here annual temperature data is taken for 30 years from 1987 to 2016. In this data the highest temperature is in the year 1995, has structural break. So from 1987 to 1995 (9 years) we consider as before, and from 1996 to 2016 (21 years) consider as after. We applied four models such as Quadratic splines, Harmonic splines, Cubic splines and Regression splines for annual temperature data in India. In this paper We use Chow test for the presence of a structural break at a period. The four models are empirically tested using annual temperature data in India.

1. Introduction

One of the fundamental concepts of statistical modeling is the almost omnipresent balance between bias and precision. Models that derived from strong assumptions about the nature as a system to produce precise estimates, but the estimates will be biased if the assumptions are not correct. On the other hand, models that derived from weaker assumptions are less prone to bias, but the resulting estimates are less precise. Regression methods are used to model changes in a responsive variable as a function of changes in a predictor variable (or several predictor variables). Standard regression methods belong to the family of parametric models; they involve strong (parametric) assumptions about the nature of the system being modeled. At the other end of the spectrum, non-parametric regression models are an attempt to make no assumptions at all. Between these extremes lie the semi-parametric methods. Splines belong to the class of semi-parametric techniques.

According to Yuan Yuan et al. [1], Regression analysis, particularly least squares regression has been a common and useful tool to study the relationship between response and a set of covariates. Quantile regression as an alternative to conditional mean regression (i.e., least square regression) is widely used in many areas. It can be used to study the covariate effects on the entire response distribution by fitting quantile regression models at multiple different quantiles or even fitting the entire regression quantile process. However, estimating the regression quantile process is inherently difficult because the induced conditional quantile function needs to be monotone at all covariate values. The B-spline methods can easily ensure the validity of the regression quantile process.

Form Karin Meyer [2], Random regression (RR) analysis have become a standard procedure for the genetic analysis of "repeated records" on individuals, which are recorded along a continuous scale such as time. A large proportion of applications considered test day production of dairy cows, but RR analysis of growth records or data on feed intake of meat producing animals are finding increasing use. The RR analysis is the subject specific curves, which can be described as the weighted sum of a set of functions of the continuous co variable, the so-called basis functions. The majority RR analysis so far fitted polynomial of time regarded as basis functions. RR analysis fitting Cubic, Quartic or even higher order polynomials are used in "end-of range" problems.

An alternative to higher degree polynomials are "piece-wise polynomials", i.e., curves constructed from pieces of lower degree polynomials, joined at selected points so called knots. Such curves are commonly referred to as spline curves; and are widely used in particular in non-parametric analysis involving smoothing curves. Daniel B. Suits et al. [3], fitted piece-wise linear regression that suffers from two obvious short comings. First, although the function itself is continuous, but its derivatives are not. Discontinuity of derivatives at the kinks can prove a serious disadvantage in many economic applications where the result would be discontinuous and probably spurious-shifts is elasticities, marginals, or other relationships that would be cloud analysis. Second, a curvilinear relationship may provide a significantly better fit to the data than obtained from linear segments. This consideration is especially important when we confronted by a complicated curve without obvious critical positions to which linear segments could be fitted. Spline functions overcome these disadvantages by replacing the linear approximations by a system of piece-wise polynomial approximations. Any degree of polynomial could be employed, but the cubic is convenient for many purposes.

Accurate, nonintrusive, and inexpensive techniques are needed to measure energy expenditure (EE) in free-living populations [4]. Especially in children and adolescents, Direct and indirect calorimetric methods measure EE that can be intrusive, confining and expensive and thus impractical for large-scale studies. The stable isotope method, doubly labeled water (DLW), is considered the "gold standard" for free-living measurements of 24-h total EE (TEE).

First we applied Cross-sectional time series (CSTS) modeling for the production of Energy Expenditure (EE) form Heart Rate (HR) and accelerometer counts (AC). This approach accounts for the independence of EE, HR, and AC over time. Second we applied Multivariate Adaptive Regression Splines (MARS) modeling, a non parametric regression method that approximate a complex non linear relationship by a series of spline functions on different intervals of the independent variable. Both models were developed and validated with the confines of a room respiration calorimeter using independent cohorts of children and adolescents.

Wenxin Mao et al. [5] explain the construct approximate confidence intervals for a nonparametric regression function using polynomial splines with free-knot locations. The number of knots is determined by generalized cross-validation. The estimates of knot locations and coefficients are obtained through a non linear least squares solution that corresponds to the maximum likelihood estimate. Confidence intervals are then constructed based on the asymptotic distribution of the maximum likelihood estimator. Average coverage probabilities and the accuracy of the estimate are examined via simulation. This includes comparisons between our method and some existing methods such as smoothing spline and variable knots selection as well as a Bayesian version of the variable knots method.

Leila D. Amorim et al. [6] explain, many Epidemiologic studies involve the occurrence of recurrent events and much attention has been given for the development of modeling techniques that take into account the dependence structure of multiple event data. The most commonly used model in survival analysis is the Cox's proportional Hazards model [7] which provides estimates of the relative risk associated with time-to-event occurrence. This method is applied those longitudinal studies in which the outcome can occur only once, for example, death or diagnosis of diabetes. Many studies involve the occurrence of recurrent events, such as times to opportunistic infections among AIDS patients or to lung exacerbations in cystic fibrosis patients, which has motivated methodological developments in survival analysis to handle complex recurrent event settings, including large number of recurrent events, presence of time-dependent covariates and time-dependent effects as well as potential dependent censoring among other features.

Rates models have been used to analyze multiple time-to-event data, where the rate of recurrence is modeled as a function of observed covariates and the effect of the covariate is assumed to be constant. The analysis of the data from the vitamin A, study using piecewise marginal rates model suggested that the effect of vitamin A supplementation on recurrent diarrhea may change over time. The use of regression spline based on time-varying coefficient rates model to examine changes in effects over time.

ZD Mulla et al. [8] explains, Dichotomous (binary) outcomes are common in clinical research. Clinical investigators, for example, may be interested in determining if there is a dose-response association between a continuous risk factor, such as body mass index (BMI) and pre-eclampsia; A traditional approach to dose-response analysis has been to convert the continuous independent variable into a categorical variable and then examine a risk of the outcome by category. A multivariate analysis could be performed in this situation. If BMI was broken down into four categories, then three indicator (dummy) variables could be entered into a regression model along with confounders. The author provides an instruction to spline regression and given an example using data from a study of patients who were hospitalized in Florida, USA, for Invasive Group A Streptococcal Disease (IGASD). The dose-response association between serum albumin and the risk of hospital mortality is evaluated using both the traditional categorical approach and the spline model.

Katerina M. Marcoulides et al. [9] explains, the analysis of longitudinal data holds a prominent role in the Behavioral, Educational, Medical and Social Sciences. Merrell [10] explains, individually fitted growth curves, which were necessary for the accurate analysis of longitudinal data. When fitting individual growth curves to longitudinal data, there are two broad challenges, the first concerns the estimation of a useful summary of the growth pattern, and the second involves making comparisons between the observed individuals. The author focuses mainly on the first of these two challenges. Due to potential complexities, a variety of different approaches have been proposed. They are logistic, exponential, Bezier, Catmull-Rom, Hermite, and Gompertz models. Fitting of these models requires either explicit a priori theoretical knowledge to specify the correct functional form or the comparative testing of alternative fitted models in order to identify the most plausible function. Piecewise linear spline models have also been frequently used to describe nonlinear trajectories, but they require the determination of a turning point (commonly referred to as a knot or change point) that indicates the occurrence of the change or shift in the studied process and denotes the point at which the linear segments should be joined. An alternative approach for fitting individual curves that can describe longitudinal data is the natural cubic smoothing spline modeling approach

2. Methodology

SPLINES:

Before computers were used, numerical calculations were done by hand. Although piecewise defined functions like the sign function or step function were used. Polynomials were generally preferred through the advent of computers, splines have gained importance. They were first used as a replacement for polynomials in interpolation, then as a tool to construct smooth and flexible shapes in computer graphics.

The word spline is used in connection with smooth, piece wise polynomial approximation. Generally splines are defined as curves which consist of individual segments. These segments are given by polynomials and the points at which they join are referred to as knots. The name spline originates thin, flexible, strip of wood traditionally used in drawing curves. The bivariate splines are commonly known as thin plate splines. Spline functions are used in number of fields such as Statistics, computer graphics, and programming computer aided design technology, numerical analysis and other areas of applied mathematics. Spline algorithms are a way to fit data points with a set of connecting curves (each one is called a Spline), such that the values between data points can be computed. They are various types/order of equations that can be used to specify the splines including Linear, Quadratic, Cubic etc. A set of N points requires N-1 splines to connect them. Each spline is described by an equation. The coefficients in those polynomials are initially unknown and the spline algorithm computes them.

2.1 Regression Splines:

One of the types of splines is Regression Splines. The time series modeling frequently involves economic data. These data tend to have abrupt changes in its trend. These changes may occur due to economic regression, epidemic outbreak, people power reduction or other political, social, economic or natural events. An important tool for evaluating these changes is a new method in regression analysis.

A linear spline regression model with N knots is given by

$Y = a + b_1x + b_2(x-c_1)_+ + b_3(x-c_2)_+ + \ldots\ldots\ldots\ldots + b_{N+1}(x-c_N)_+$
Where $(x-c_j)_+ = x-c_j$ If $x-c_j > 0$ and $(x-c_j)_+ = 0$ if $(x-c_j) \leq 0$.

This can be written as separate regression lines:

$Y = a + b_1X$ $\quad x \leq c_1$
$Y = a + b_1x + b_2(x-c_1)_+$ $\quad c_1 < x \leq c_2$
$Y = a + b_1x + b_2(x-c_1)_+ + b_3(x-c_2)_+$ $\quad c_2 < x \leq c_3$
$Y = a + b_1x + b_2(x-c_1)_+ + b_3(x-c_2)_+ + \ldots\ldots\ldots\ldots + b_{N+1}(x-c_N)_+$ $\quad c_N < x$

A polynomial of degree D is a function formed by linear combinations of the powers of its argument up to D:

$$Y = \beta_0 + \beta_1x + \beta_2x^2 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_D X^D$$

Specific polynomials are:

Linear	$Y = \beta_0 + \beta_1x$
Quadratic	$Y = \beta_0 + \beta_1x + \beta_2x^2$
Cubic	$Y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3$
Quartic	$Y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4$
Quintic	$Y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \beta_5x^5$

The linear regression spline equation is

$$Y_i = \beta_0 + \beta_i x + e_i \text{ for all } i = 1,2,\ldots n$$

Where y_i and x_i are the values of the response and predictor variables for the i^{th} observation; β_0 and β_1 are constants.

2.2 B -Splines:

Another type of splines is B-Splines. B standards for Basis. The number of knots is the minimum for the degree of the B-spline, which has a non zero value only in the range between the first and last knot. Each piece of the function is a polynomial of degree < n between and including adjacent knots. A B-Spline is a continuous function at the knots. The places where the pieces meet are known as knots. The number of interval knots must be equal to or > n-1. A fundamental theorem states that every spline function of a given degree, smoothness and domain partition can be uniquely represented as a linear combination of B-splines of that same degree and smoothness and over that same partition . B-splines yield the same fit as splines based on truncated power functions.

B-spline functions can be defined recursively basis functions of degree p=0, have values of units for all points in a given interval and zero otherwise. For the k^{th} interval given by knots T_K and T_{K+1} with $T_K \leq T_{K+1}$.

$$B_{k,0}(t) = \begin{cases} 1 \; if \, Tk \leq t \leq Tk \\ 0 \qquad\qquad otherv \end{cases}$$

Higher degree basis functions, $B_{K,P}$ for p>0 are then determined from the values of lower degree basis functions and the width of the adjoining intervals between knots.

The general relationship is

$$B_{K,P}(t) = \frac{t-T_k}{T_{k+p}-T_K} B_{K,P-1}(t) + \frac{T_{K+P-1}-t}{T_{K+P-1}-T_{K+1}} B_{K+1,P-1}(t)$$

For each p, there are a limited number of non zero basis functions of lower order. For equally spaced knots, B-spline functions can be obtained as the difference between splines with a basis of truncated power functions. B-splines are the functions designed to generalize polynomials for the purpose of approximating arbitrary functions.

The polynomial of order k is

$$p(x) = a_0 + a_1x + a_2x^2 + \ldots\ldots + a_{k-1}x^{k-1}$$

Each control point is associated with unique basis function. Each point affects the shape of the curve over range of parameter values, where the basis function is non zero

$$X(u) = \sum_{i=1}^{n} N_{i,k}(u)x_i, 0 \leq u \leq n-k-2$$

Here n+1 control points and the order of the curve is k.

2.3 Smoothing Splines:

Another type of splines that is extensively used is the smoothing spline. Smoothing splines come close to the observed data points among all functions that satisfy a pre-specified "smootheness" criterion. Specifically, given set of data (x_i , y_i), i=1,2,…..,n, a smoothing spline is the function f that satisfies:

$$Min$$
$$f \in \left\{h:[a,b] \to \Re; h^{(2)} exists\right\} \left\{ \sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int_a^b [f^2(u)]^2 du \right\}$$

For some fixed λ, the integral quantity of the above equation is called the roughness penalty and λ is called the smoothing parameter. Larger values of λ produce functions that are smoother than those for smaller values of λ. The values of λ may be optimized using the mean squared error obtained from fitting the estimated function to an independent data set.

2.4 Cubic Splines:

A cubic spline is a spline constructed of piecewise third order polynomials, which pass through a set of m control points. The second derivative of each polynomial is commonly set to zero at the end points. Since this provides a boundary condition that complete the system of m-2 equations. This procedure also called natural cubic spline and leads to a simple tridiagonal system which can be solved easily to give the coefficients of the polynomials. However, this choice is not only one possible and other boundary conditions can be used instead.

The cubic spline equation is

$$y_i(t) = a_i + b_i t + c_i t^2 + d_i t^3$$

2.5 Harmonic Splines:

The equation for Harmonic splines is

$$y_t = a + b\sin\theta + c\cos\theta + d*t$$

Where t = time
Y_t = values
sinθ = Before intervention values
cosθ = after intervention values
a, b, c, and d are the constants.

2.6 Quadratic Splines:

The quadratic spline regression equation is

Y_t = a+b*t + c*t^2

Where t = time
Y_t = values, a,b and c are the coefficients.

These coefficients a,b,c in the above equation can be estimated using least squares.

2.7 CHOW TEST

Chow test is commonly used to test for structural change in some or all of the parameters of a model.
there is the assumption that the model errors are independent and identically distributed from a normal distribution with unknown variance.

Let Sc be the sum of squared residuals from the combined data, S1 be the sum of squared residuals from the first group and S2 be the sum of squared residuals from the second group. N1 and N2 are the number of observations in each group and K is the total number of parameters, then the chow test statistic is

$$chowtest = \frac{[S_C - (S_1 + S_2)]/K}{(S_1 + S_2)/(N_1 + N_2 - 2K)}$$

The test statistic follows the F distribution with K and $N_1 + N_2 - 2K$ degrees of freedom.

Here we consider four models for testing the yearly temperature data.

3. Results and Discussions

Here annual temperature data is taken for 30 years from 1987 to 2016. In this data the highest temperature is in the year 1995. So from 1987 to 1995 (9 years) we consider as before and from 1996 to 2016 (21 years) consider as after.

For testing the above data we consider four models.

3.1 Model I: *Quadratic Splines*

The quadratic spline regression equation is

$$y_t = a + b*t + c*t^2$$

H0: There is significant difference between before and after Quadratic spline models for data.

To test the above H0, we do the following calculations:

The fitted Quadratic equation for before data is

$$y_t = 0.046*t^2 - 0.401*t + 24.96 \qquad 1 \le t \le 9$$

The fitted Quadratic equation for after data is

$$y_t = 0.002*t^2 - 0.013*t + 24.60 \qquad 10 \le t \le 30$$

The Error Sum of Squares (ESS) is
ESS for Total = 6.68018
ESS for before = 0.176933
ESS for after = 2.655155
Chow test value = 10.86998
F (3, 24) = 4.72 for α = 0.01

Comparison: Chow test statistic is greater than F table value at α = 0.01. So we reject the null hypothesis.

3.2 Model II: *Harmonic Splines*

The harmonic spline equation is

$$y_t = a + b\sin\theta + c\cos\theta + d*t$$

H0: There is significant difference between before and after Harmonic spline models for data.

To test the above H0, we do the following calculations,

The fitted Harmonic spline equation for before data is

$$y_t = 24.60461 + 51.92799*\sin\hat{\theta} - 23.8518*\cos\hat{\theta} + 2.655227*t \qquad 1 \le t \le 9$$

The fitted Harmonic spline equation for after data is

$$y_t = 25.20468 + 51.67683*\sin\hat{\theta} - 24.3968*\cos\hat{\theta} + 6.19166*t \qquad 10 \le t \le 30$$

The Error Sum of Squares (ESS) is
ESS for Total = 44481.2
ESS for before = 2009.298
ESS for after = 126932.6

Chow test value = 3.6027
F (4, 22) = 4.31 for α = 0.01

Comparison: Chow test statistic is less than table value at α = 0.01. So we accept the null hypothesis.

3.3 Model III: Cubic Splines

The cubic spline equation is

$$y_t = a + b*t + c*t^2 + d*t^3$$

H0: There is significant difference between before and after Cubic spline models for data.

To test the above H0, we do the following calculations,

The fitted Cubic spline equation for before data is

$$y_t = 0.004*t^3 - 0.018*t^2 - 0.127*t + 24.67$$

1≤ t ≤ 9

The fitted Cubic spline equation for after data is

$$y_t = 0.001*t^3 - 0.032*t^2 + 0.299*t + 23.97$$

10 ≤ t ≤ 30

The Error Sum of Squares (ESS) is
ESS for Total = 99.63574
ESS for before = 0.217149
ESS for after = 2.171636
Chow test value = 223.904
F (4, 22) = 4.31 for α = 0.01

Comparison: Chow test statistic is greater than table value at α = 0.01. So we reject the null hypothesis.

3.4 Model IV: Regression Splines

The equation for regression spline is

$$y_t = a + b*t$$

H0: There is significant difference between before and after Regression spline models for data.

To test the above H0, we do the following calculations,

The fitted Regression spline equation for before data is

$$y_t = 0.062*t + 24.11$$ 1 ≤ t ≤ 9

The fitted Regression spline equation for after data is

$$y_t = 0.038*t + 24.41$$ 10 ≤ t ≤ 30

The Error Sum of Squares (ESS) is
ESS for Total = 3.634875
ESS for before = 0.83668
ESS for after = 2.667964
Chow test value = 0.483074
F (2, 26) = 5.53 for α = 0.01

Comparison: Chow test statistic is less than table value at α = 0.01. So we accept the null hypothesis.

4. Conclusions

In this paper we fitted four splines, they are Quadratic splines, Harmonic splines, Cubic splines and Regression splines.

4.1 Quadratic Splines

The fitted Quadratic equation for before data is

$$y = 0.046*t^2 - 0.401*t + 24.96$$

The fitted Quadratic equation for after data is

$$y = 0.002*t^2 - 0.013*t + 24.60$$

and Chow test for structural break is

Chow test value = 10.86998
F (3, 24) = 4.72 for α = 0.01

Comparison: Chow test statistic is greater than F table value. So we reject the null hypothesis.

4.2 Harmonic Splines

The fitted Harmonic spline equation for before data is

$$y_t = 24.60461 + 51.92799*\sin\hat{\theta} - 23.8518*\cos\hat{\theta} + 2.655227*t$$

The fitted Harmonic spline equation for after data is

$$y_t = 25.20468 + 51.67683*\sin\hat{\theta} - 24.3968*\cos\hat{\theta} + 6.19166*t$$

Chow test value = 3.6027
F (4, 22) = 4.31 for α = 0.01

Comparison: Chow test statistic is less than F table value at α = 0.01. So we accept the null hypothesis.

4.3 Cubic Splines:

The fitted Cubic spline equation for before data is

$$y = 0.004*t^3 - 0.018*t^2 - 0.127*t + 24.67$$

The fitted Cubic spline equation for after data is

$$y = 0.001*t^3 - 0.032*t^2 + 0.299*t + 23.97$$

Chow test value = 223.904
F (4, 22) = 4.31 for α = 0.01

Comparison: Chow test statistic is greater than F table value at α = 0.01. So we reject the null hypothesis.

4.4 Regression Splines

The fitted Regression spline equation for before data is

$$y_t = 0.062 * t + 24.11$$

The fitted Regression spline equation for after data is

$$y_t = 0.038 * t + 24.41$$

Chow test value = 0.483074

F (2, 26) = 5.53 for α = 0.01

Comparison: Chow test statistic is less than F table value at α = 0.01. So we accept the null hypothesis.

So from the above results, the two models: Harmonic splines and Regression splines are suited for the annual temperature data in India.

References

1. Nan Chen, Shiyu Zhou, Yuan Yuan,(2016): 'Modelling Regression Quantile Process using Monotone B-splines'.
2. Karin Meyer.(2005): 'Random Regression Analyses using B-splines to model growth of Australian Angus Cattle'.,Genet. Sel. Evol. 37 , 473-500.
3. Andrew Mason, Daniel B. Suits, Louis Chan.(1977): 'Spline Functions Fitted by Standard Regression Methods'.
4. Anne L.Adolph, Firoz A. Vohra Nancy F.Butte, Maurice R. Puyau, William W. Wong,(2010):'Validation of Cross-Sectional Time Series and Multivariate Adaptive Regression Splines Models for the Prediction of Energy Expendituere in Chlidren and Adolescents Using Doubly Labled Water'.The Journal of Nutrion Methodology and Mathematical Modelling, doi:10.3945/jn.109.120162.
5. Linda H.Zhao Wenxin Mao. (2003): 'Free-Knot Polynomial Splines with Confidence Intervals'., J.R.Statist. Soc.B 65, part4, 901-919.
6. Donglin Zeng, Jianwen Cai,Leila D. Amorim, Mauricio L. Barreto.(2008): 'Regression Splines in the Time-Dependent C oefficient Rates Model for Recurrent Event Data'. Stat Med.27(28):5890-5906. Doi:10.1002/sim.3400.
7. Cox DR.(1972): 'Regression Models and Life-tables(with discussion)'. Jounal of Royal Statistics Society series B. 34:182-220.
8. ZD Mulla.(2007): 'Spline Regression in Clinical Research',West Indian Med J, 56(1):77.
9. Jam Khojasteh, Katerina M. Marcoulides.(2018): 'Analyzing Longitudinal Data Using Natural Cubic Splines. Structural Equation Modelling' A Multidisciplinary Journal ,DOI: 10.1080/10705511.2018.1449113.
10. Merrell, M. (1931): 'The relationship of individual growth to average growth'. Human Bioligy, 3, 37-70
11. Lars k. Nielsen, Lake-Ee Quek, Maria Buchsteiner, Peter Gray Veronica S. Martinez.(2015): 'Dynamic Metabolic Flux Analysis using B-Splines to study the effects of Temperature shift on CHO Cell Metabolism', Metabolic Engineering Communications 2 , 46-57.
12. Carolina B. Baguio, Rex L. Sarvida.(2010): 'On the Efficiency of Spline Regression for Time Series Data'. 11th National Convention on Statistics (NCS),Oct-2010
13. G. G. S. Pegram, M.C.Peel, T.A.McMahon.(2008): ' Empirical Mode Decomposition using Rational Splines: An Application to Rainfall Time Series' Proc. R. Soc. A , 464, 1483-1501
14. A. Baccini, P.Besse, P.G. P.Martin, S. Dejean.(2007): 'Clustering Time Series Gene Expression Data Using Smoothing Spline Derivatives'. EURASIP Journal on Bioinformatics and Systems Biology, doi: 10.1155/2007/70561.
15. http//:www.data.gov.in.

