International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:3, 2010

Peakwise Smoothing of Data Models using Wavelets

D Sudheer Reddy, N Gopal Reddy, P V Radhadevi, J Saibaba, and Geeta Varadan,

Abstract—Smoothing or filtering of data is first preprocessing step for noise suppression in many applications involving data analysis. Moving average is the most popular method of smoothing the data, generalization of this led to the development of Savitzky-Golay filter. Many window smoothing methods were developed by convolving the data with different window functions for different applications; most widely used window functions are Gaussian or Kaiser. Function approximation of the data by polynomial regression or Fourier expansion or wavelet expansion also gives a smoothed data. Wavelets also smooth the data to great extent by thresholding the wavelet coefficients. Almost all smoothing methods destroys the peaks and flatten them when the support of the window is increased. In certain applications it is desirable to retain peaks while smoothing the data as much as possible. In this paper we present a methodology called as peak-wise smoothing that will smooth the data to any desired level without losing the major peak features.

Keywords—smoothing, moving average, peakwise smoothing, spatial density models, planar shape models, wavelets.

I. Introduction

SMOOTHING and differentiation of signals corrupted by noise is a requirement in all the fields concerning analysis of data where obtaining derivatives by direct measurement is difficult or impossible. The evaluation of the derivatives of an experimentally measured data is an ill-posed problem, because the available values are finite in number and affected by errors. In fact, the measurement errors, which can never be avoided, complicate the differentiation, because this amplifies the noise present to such an extent that additional signal treatment such as smoothing becomes essential.

Moving average is the primitive method of smoothing the data widely used in many applications involving data analysis [1, 2]. Different moving average methods were proposed in literature [Jeffrey] with different weights/kernels [3], Savitzky and Golay generalized the moving average method by fitting a polynomial of degree p to a moving window of data points of length L(=2k+1) [4]. Savitzky and Golay method suffered from a limitation of truncating the data by k-points at each end. Peter A Gorry used general least squares method to overcome this limitation to accommodate end points [5]. Smoothing can be achieved by convolving with a kernel function that varies for different applications, most widely used weight functions are Gaussian or Kaiser. Function approximation by polynomial fitting also produces smoothing of the given data. Smoothing is also carried using Fourier series expansion but the kinks that are usually present in any data can never be reconstructed with any finite number (however large)of terms included in the series approximation, also, at the points of discontinuities there are ripple effects whose oscillations increased with increase in the terms included in partial sums famously called as Gibbs phenomenon. Many special weighted window designs are proposed in literature to handle Gibbs oscillations such as Lanczos and Fejer who used weighted Fourier coefficients to reduce these ripple effects [6]. Fredric Harris in his study compared many different window functions and suggested the use of Kaiser or Blackman-Harris windows as most optimal designs [7]. Recently, wavelets are playing an important role in data analysis; wavelets smooth the data to the required degree, the magnitude of smoothing achieved is dependent on the choice of basis function and a threshold applied on wavelet coefficients [8].

Almost all smoothing methods destroy the peaks and flatten them; the amount of flatness depends on the nature and support of the weight function. Wider the support greater is the smoothness and hence completely removing some of the important features like peaks. Many times it becomes difficult to retain required peaks and suppress the high frequency noise. Situations like these are not uncommon in areas of image processing such as restoration of certain important features, edge detection where the neighboring pixels are not important, filtering of SAR data, etc. In this paper we present a methodology called as peak-wise smoothing that will smooth the data to any desired level without losing the major peaks. The present methodology stems from the observation that selection of larger width weight functions/kernels will collapse the most important peak features of the data, hence the requirement of achieving enough smoothing without losing major features. The strength of this methodology is that it is effective with any kind of smoothing technique. This methodology is similar to piecewise smoothing assuming that the data between peaks are taken as pieces. This methodology was experimented on two data sets, one from spatial density models of space debris environment and the other on slope distance curves of regular shapes in a binary image, and the results are presented. The paper is organized in the following manner: in section 2 we present different methods of data smoothing, section 3 deals with peak identification using wavelets, the present methodology is described in section 4, description of data sets used for testing the present methodology is presented in section 5, results and discussions are elaborated in section 6.

D. Sudheer Reddy is with Digital Mapping and Modeling Division, Advanced Data Processing Research Institute, Department of Space, 203, Manovikas Nagar, Secunderabad-500009, India. e-mail: (donsudhir@gmail.com).

N Gopal Reddy is currently Principal, Lalitha PG College, Affiliated to Osmania University, Hyderabad. e-mail: (dr_ngr@yahoo.com).

P V Radhadevi, J Saibaba, Geeta Varadan are with Advanced Data Processing Research Institute, Department of Space.

Manuscript received Decl 2, 2010;

II. Data Smoothing Methods

In this section we quickly review some of the smoothing methodologies that are frequently used in many smoothing and filtering applications (1) Moving window methods (n-point moving average, n-point moving median Savitzky-Golay method, Gaussian, Kaiser) (2) Function approximation method (Polynomial, Fourier, wavelet)

1) Moving window methods: All moving window smoothing methods are based on convolution. Different window functions are designed to extract a particular aspect of the data, such as desired frequency, while eliminating other aspects. These methods falls in two categories, spatial domain methods that directly work on the data itself, and frequency domain, which are based on Fourier transforms. We will not delve much on frequency domain methods as there are abundant resources available to describe them in a separate subject called Filter Design. However we will present one optimally designed window function given by Kaiser.

a. Moving average method: In many real time applications, the observations x[n], n=1,2,..N, are smoothed by a weight vector/filter w_j. Smoothing creates a new sequence y_j according as, $y[n] = x[n] * w[n] = \frac{1}{2k+1}\sum_{i=-k}^{k} x[n-i]$ for instance, $y[j] = \frac{1}{3}(x[j-1] + x[j] + x[j+1])$ replaces x[j] by a moving average of x[j] and its two nearest neighbors. Note that the moving-average systems significantly reduce the short-term fluctuations in the data, and the system with the larger value of k produces a smoother output. The challenge with smoothing applications of moving-average methods is how to choose the window length N so as to identify the underlying trend of the input data in the most informative manner. The general form of moving-average methods with unequal weights is given by $y[n] = x[n] * w[n] = \sum_{i=-k}^{k} w[i] * x[n-i]$. In many situations it is required that weights must satisfy $w[j] \geq 0$ and $\sum_{i=-k}^{k} w[j] = 1$.

b. Savitzky-Golay method: This is a generalization of moving average method given by Savitzky & Golay [4] by performing a least squares fit to a small set of L(=2k+1) consecutive data points to a polynomial and taking the central point of the fitted polynomial curve as output. The smoothed data point is given by the following

$y[n] = x[n] * w[n] = \frac{\sum_{i=-k}^{k} A_i * x[n-i]}{\sum_{i=-k}^{k} A_i}$, where,

$w[n] = \frac{A_n}{\sum_{i=-k}^{k} A_i}$, $-k \leq n \leq k$ here A_i controls the polynomial order.

c. Moving median method: The smoothed sequence y_j from the data x_j is given by $y[n] = median\{x[n-i], i = -k\, to\, k\}, n \geq k$, where 2k+1 is the window size.This method essentially removes peaks present in the data, usually not required for the kind of applications mentioned above.

d. Smoothing by Gaussian Window: The given function/data x(t) is convolved with a Gaussian kernel of the form $w[t] = \frac{1}{\sqrt{2\pi\sigma}} exp(\frac{-t^2}{2\sigma^2})$, the degree of smoothing is determined by the standard deviation . In fact, we can view Gaussian filter also as a weighted moving average filter. This kernel function sets larger weight factors for points in the center and smaller weight factors for points away from the center.

e. Smoothing by Kaiser Window: Kaiser designed a two parameter window function by using weights based on Fourier coefficients, given by

$$w[n] = \begin{cases} \frac{1}{I_0(\alpha)} I_0(\alpha\sqrt{1-(\frac{n}{N})^2}), & \text{for } -N \leq n \leq N \\ 0, & else. \end{cases}$$

where, $I_0(x) = \sum_{k=0}^{\infty}[\frac{1}{k!}(\frac{x}{2})^k]^2$ is the zeroth order modified Bessel function. The parameter controls the shape of the window and N is half width of the window size.

2) Smoothing by Function approximation:

a. Smoothing by Polynomials: Least squares polynomial fitting to the given data also gives a smoothed output of the given data but in the case where the data are widely spread with higher value of standard deviation then polynomial fitting gives very poor performance. The polynomial functional form is given by $x(t) = a_0 + a_1 t + a_2 t^2 + \;\; + a_k t^k$. Also another limitation of polynomial fitting is the choice of polynomial order and its dependence on the number of data points.

b. Fourier smoothing: Fourier approximation of the given function/data using a sum of weighted sine and cosine terms of increasing frequency also produces smoothing to the data. The data must be equi-spaced and discrete smoothed data points are returned. The major advantage of this method is that it achieves very good smoothing of the data depending on truncating the number of terms taken for expansion (called as low pass filtering in Signal processing literature) or thresholding the series up to a tolerable error. Truncation of Fourier series leads to Gibbs phenomenon [6]. Truncated Fourier approximation is given by $x(t) = \frac{a_0}{2} + \sum_{k=0}^{N}[a_k cos(kt) + b_k sin(kt)]$

c. Smoothing by thresholding wavelet coefficients: The wavelet transform with a wavelet function having n vanishing moments is a multi-scale differential operator, i.e., it has both the properties of smoothing and differentiation [8]. Signals x(t) are represented by a series such as $x(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} d_k^j \psi_k^j(u)$, where,$\psi_k^j(u) = \psi(2^j u - k)$ are wavelet functions and d_j^k are wavelet coefficients $d_j^k = \int x(u)\psi_k^j(u)du$. An important property of wavelets is the possibility to show that the amplitude of wavelet coefficients is associated with abrupt signal variations or details of higher frequency. Effective thresholding techniques are available to denoise the high frequency content of the signal, here we use the soft universal thresholding $\tau = \sigma\sqrt{2log(m)}$, if σ is unknown a robust datadriven estimate $\widehat{\sigma}$ can be used. Smoothing is achieved by setting all those coefficients d_j^k that are less than the threshold τ to zero.

III. Peak identification with wavelets

The presence of abrupt variations/peaks in any data carry important information which gives an indication of some transient behavior of the underlying physical phenomenon. Classification, detection and measurement of such singularities using wavelets was given by Mallat [9]. The advantage of wavelet transform is that they can characterize the local regularity of the data. The presence of singularities or sharp variations in the data is sensed in detail coefficients of almost all levels of decomposition which acts as a key in many wavelet based peak estimation methods. At higher level of

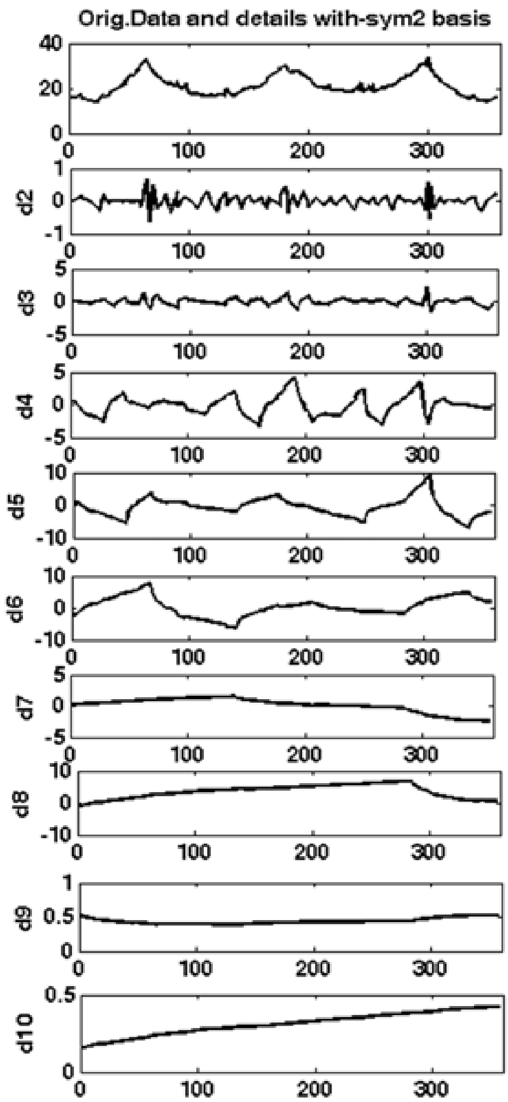

Fig. 1. Figure(1) Boundary shape data and its decomposition up to 10 levels using symlets basis of order 2

decomposition, the magnitude of most of the detail coefficients at higher levels of decomposition that corresponding to local and small variations are zeros or close to zero and the rest corresponds to large or sharp variation which occurs at peaks. Using this fact as a criteria and fixing a threshold to filter out the detail coefficients at one particular level of decomposition and there by sorting the remaining coefficients will leads us to find the exact peak location and thus the number of peaks. Figure(1) shows a signal and its detail coefficients decomposed up to 10 levels using Symlets basis function of support 2. Observe that choosing an appropriate threshold at level 2 detail coefficients gives exact peak locations.

IV. Present Methodology

This method is mainly dependent on two techniques, estimation of peaks in the data, and smoothing method. The present procedure employs the following steps

1. Read the data (x_i, y_i) , i=1,2,3,....,N, and choose a smoothing method $movingaverage, Savitzky-Golay, wavelets, etc,..$
2. Locate the peaks that are to be retained using wavelets (as described in *III*), say $x_{p_1}, x_{p_2}, x_{p_3}, ...x_{p_k}$ are k peaks.
3. Partition the data into k+1 pieces at the peak locations obtained in 1 $[x_1, x_{p_1}], [x_{p_1}, x_{p_2}], [x_{p_2}, x_{p_3}], ...[x_{p_k}, x_N]$.
4. Smoothing the data that lie in these partitions with the selected smoothing method (if window smoothing method is chosen in step (1) then leave those partitions without any manipulation if the number of data points in that partition is less than window size)
5. Join the smoothed pieces obtained in (4), that results in peak wise smoothing of the data.

V. Description of Experimental data

In this section we describe data sets used for testing the present methodology, one data set is about spatial density models of space debris data and the other regards slope distance curves of a binary image.

Space Debris Data : Any man-made object in orbit around the Earth that no longer serves a useful purpose is defined as space debris. As of Jun 2010, approximately 15,550 debris objects larger than 10 cm are being tracked in Earth orbit [10]. Because of their high orbital velocities, collisions with even small pieces of debris can involve considerable energy, and therefore pose a significant danger to spacecraft and astronauts. In this context the term spatial density refers to number of objects per km^3 around Earth [11]. Space debris researchers use variety of methods to characterize the space debris environment in terms of number, altitude, and inclination distributions to improve the understanding and knowledge of the risks debris poses to operational satellites, as well as determine sources of debris for future mitigation. One such characterization was developed by Ananthasayanam etal, to model spatial density in different inclination regions [12] by a mixture of Laplace distributions. Anilkumar & Sudheer Reddy [13] used similar characterization and identified three inclination regions of high risk of conjunctions. These models are generated using the two line elements (TLE) data sets which are obtained from space-track web site (www.space-track.com). Figure (2) shows the spatial density calculated for June 2009 TLE data from 200 km to 2000 km from the earth surface taking 25 km as bin size in all inclinations from 0-180 deg. Observe that there is high concentration of debris objects in the orbits whose altitude is around 600-700 km and 1300-1400 km, these are the areas of high risk for the operational satellites. The peaks around 700-750 km and 800-850 km are temporary and not due to noise, but these concentrations will be moving towards a more stabilizing orbits under the influence of gravity field and third body perturbations as time progresses.

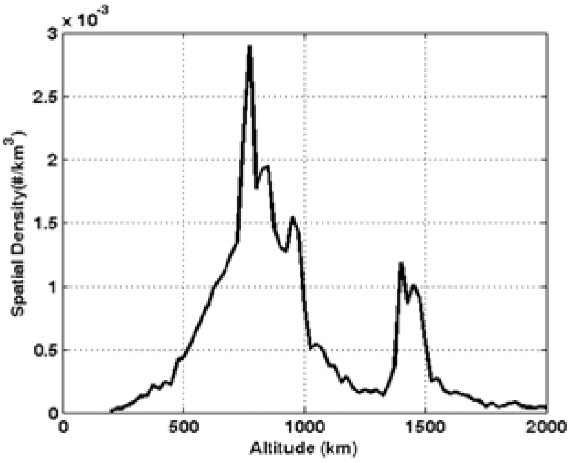

Fig. 2. Spatial density distribution

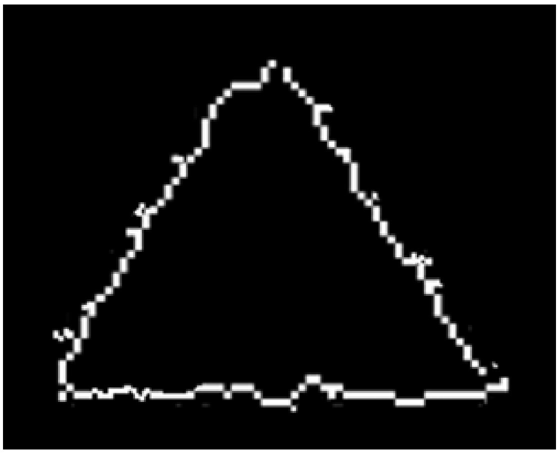

Fig. 3. a triangle shape in binary image

Boundary Shape data in binary images : Shape identification is an important and the most fundamental aspect of robot vision where the emphasis is to design algorithms that enable a machine to recognize and interpret shapes of various objects as perceived by humans. This process involves capturing a scene from a robotic eye (camera) in the form of images and then extraction of shapes from these images. Many shapes are not that obvious to be identified; hence recognition will be based on certain features of the shape and its boundaries. Yang Mingqiang etal [14]., survey explores around 40 different techniques for extraction of shape features, one such description is through slope distance curves i.e., distance expressed as a function of slope when a radius vector sweeps from some point inside the boundary to all points of the boundary. Slope distance curves characterize the shape of convex boundary of an object [15]. The slope distance curve of a circle from its center is a straight line. The slope distance curve of the triangle shape shown in figure (3) is given in figure (4). Here the peaks correspond to corners of the triangle that forms the most important features for shape identification.

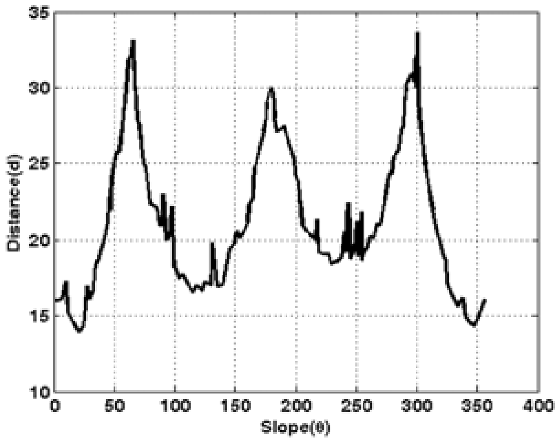

Fig. 4. slope distance curve of triangle shape in binary image shown above

TABLE I
PARAMETERS CONSIDERED FOR DIFFERENT SMOOTHING METHODS IN SPATIAL DENSITY DATA

Smoothing method	Parameters
Moving Average	Window size = 5
Savitzky Golay	Window size = 5, degree of polynomial =4.
Gaussian window	Window size=5, std=3.0
Kaiser Window	Window size=5, alpha = 6.5
Fourier	Number of terms = 25
Wavelet	Symlets2 basis, 5 levels of decomposition=5

TABLE II
PARAMETERS CONSIDERED FOR DIFFERENT SMOOTHING METHODS IN SHAPE MODEL DATA

Smoothing method	Parameters
Moving Average	Window size = 7
Savitzky Golay	Window size = 7, degree of polynomial =4.
Gaussian window	Window size=7, std=2.5
Kaiser Window	Window size=7, alpha =6.5
Fourier	Number of terms = 25
Wavelet	Symlets8 basis, 5 levels of decomposition=5

VI. RESULTS AND DISCUSSIONS

We have experimented both data sets with different smoothing methods described above viz., as Moving average method, Savitzky-Golay method, Gaussian window method, Kaiser window method, Fourier approximation method and wavelet approximation method. Different parameters considered for different smoothing methods are tabulated in Table-I &Table-II. As the number of data points is less in spatial density data we have taken smaller windows than compared with shape model data.

Figure(5) shows the result of different smoothing methods (here (a), (b), (c), (d), (e), (f) corresponds to moving average, Savitzky-Golay, Gaussian, Kaiser, Fourier, and Wavelet methods respectively) taking full data at a stretch, and figure(6) shows the corresponding result when smoothed peak wise. In the prior case a flat peak gives a misplaced position with decreased spatial density value. Observe that methods (Moving average, Gaussian, Kaiser, Fourier) that have produced enough smoothing has suppressed noise to a greater extent and flattened peaks, on the other hand methods (Savitzky-Golay, Wavelet) that have retained peaks to some extent have failed to suppress noise. Clearly these two demands seem to be contradicting. Wavelet smoothing has taken care of smoothing and preserved desired peaks to a better extent than other methods, polynomial smoothing can be achieved with other wavelet basis functions that have larger support and high regularity; however such basis will result in similar effects as with moving average method. Peak-wise smoothing has obtained better performance meeting the contradicting demands, the individual pieces from peak to peak are smoothed with the said methods and joined resulting in suppression of noise to good extent. The truncation error at end points in Fourier series is more compared to any other method. Since peak wise smoothing involves smoothing of data partitions between peaks, the truncation error at the end points in each partition has to be addressed separately. Figure (7) gives the smoothing of slope distance curve with different smoothing methods, here a flat peak corresponds to round corner triangle and noise makes sides of the triangle untidy. Figure (8) shows peak wise smoothing of slope distance curve with different smoothing methods. We also remark from figure (6) and figure (8) that peak wise smoothing can be made more effective by properly tuning the parameters of each of the smoothing methods. Peak wise smoothing with moving average method is the simple and fastest of all, however, peak wise smoothing with wavelet method is more promising even with larger support basis functions. Depending on the nature of the application suitable smoothing method can be adopted to smooth the data partitions. From both the data sets we observe that peak wise smoothing has achieved better smoothing without sacrificing major features. The performance of this method is dependent on peak identification algorithm and effective smoothing method.

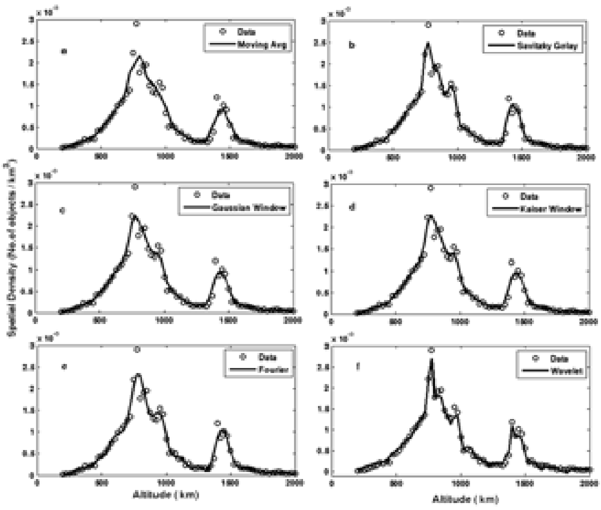

Fig. 5. smoothing of spatial density data for June 2009 TLE data set with different methods. (a), (b), (c), (d), (e), (f) are correspondingly smoothing with Moving average, Savitzky-Golay, Gaussian, Kaiser, Fourier, Wavelet methods respectively.

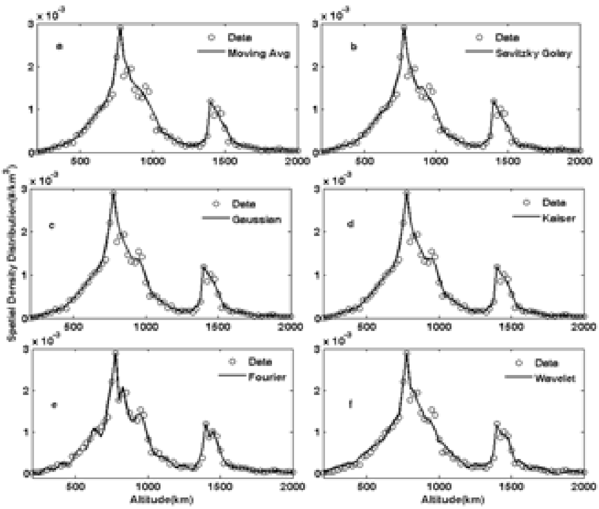

Fig. 6. peak wise smoothing of spatial density data for June 2009 TLE data set with different methods. (a), (b), (c), (d), (e), (f) are correspondingly smoothing with Moving average, Savitzky-Golay, Gaussian, Kaiser, Fourier, Wavelet methods respectively.

VII. Conclusion

We observe that peak wise smoothing can result in any degree of smoothing compared to smoothing of the data at full length. This is most useful in the applications which demand for better smoothing without sacrificing major features of the data. Peak wise smoothing with moving average is fastest and is more effective with wavelet basis.

Acknowledgment

The author is greatly indebted to the support and encouragement received from Geeta Varadan, Director, Advanced Data Processing Research Institute , J. Saibaba, Deputy Director, Satellite Data Analysis and Applications, P.V. Radhadevi, Head, Digital Mapping and Modeling Division and N. Gopal Reddy, Professor, Department of Mathematics, Osmania University.

References

[1] Edward. A. Bender, An Introduction to mathematical modeling, 1978, John Wiley and Sons Inc.
[2] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery,Numerical Recipes, 3ed. Cambridge University Press, 2007.
[3] Jeffrey. S.Simonoff, Smoothing methods in statistics, Springer Verlag, New York, 1996.
[4] Savitzky A, Golay MJE, Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 1964, 36:1627-1639.
[5] Peter A Gorry, General Least-Squares Smoothing and Differentiation by the Convolution (Savitzky-Golay) Method. Analytical Chemistry, 1990, 62, pp 570-573.
[6] Hamming. R.W., Digital Filters, Prentice Hall Signal Processing Series, 3ed, 1989.
[7] Fredric J Harris, On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform, Proceedings of the IEEE, Vol 66, No. 1. January 1978.
[8] Mallat.S, A Wavelet Tour of Signal Processing, 2 ed. San Diego, CA: Academic, 1999.
[9] Mallat. S., Wen Liang Hwang, Singularity Detection and processing with

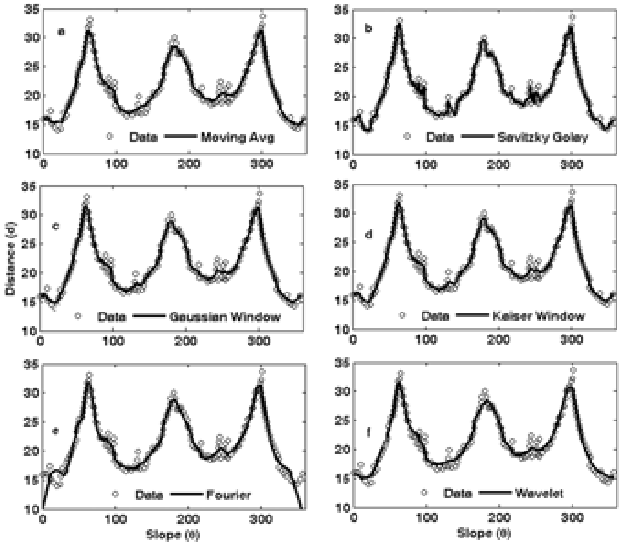

Fig. 7. Effect of different smoothing methods with the slope distance curve of triangular feature. (a), (b), (c), (d), (e), (f) are correspondingly smoothing with Moving average, Savitzky-Golay, Gaussian, Kaiser, Fourier, Wavelet methods respectively.

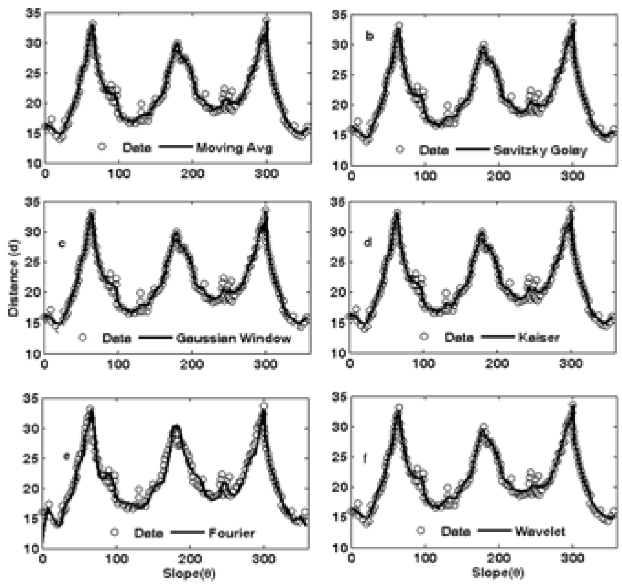

Fig. 8. Effect of peak wise smoothing using same parameters that were chosen for full length smoothing. (a), (b), (c), (d), (e), (f) are correspondingly smoothing with Moving average, Savitzky-Golay, Gaussian, Kaiser, Fourier, Wavelet methods respectively

wavelets, IEEE Trans. on Information Theory, vol. 38, No. 2, March 1992.
[10] Orbital Debris Quarterly News, Vol. 14, Issue-3.
[11] Klinkrad, H., Space Debris, Models and Risk Analysis, Springer-Praxis Publishing, 2006.
[12] Ananthasayanam, M. R., Anilkumar, A. K., and SubbaRao, P. V., 2006, Stochastic impressionistic Low Earth Model of the space debris scenario, ACTA Astronautica, Vol. 59, pp. 547-559.
[13] Anilkumar, A. K., Sudheer Reddy. D., Statistical conjunctiona analysis of LEO catalogued objects, Journal of Spacecraft and Rockets, AIAA, pp. 160-167, Jan-Feb 2009.

D Sudheer Reddy has received his M.Sc(Tech), and M.Phil both in Mathematics and M.Tech in Information Technology from Jawaharlal Nehru Technological University, Osmania University, and Guru Gobind Singh Institute respectively. His research interestes are in the areas of application of Wavelets, Image processing, computational geometry, orbital mechanics, space debris, Mathematical and Stochastic modeling.

N Gopal Reddy has received his Ph.D in Mathematics from Osmania University, and has joined the same Department and served over 30 years in various capacities as Professor, Head, Chairman Board of studies, etc. his research interests are in the areas of Functional Analysis, Operator theory, Frame operators, wavelets and its applications, Banach and Hilbert space operators. He has guided several students at doctoral level in these areas.

P V Radhadevi has received her Ph.D. in applied Mathematics from Bangalore University in 1994. she has joined Advanced Data processing Research Institute, a division of Department of Space, as a scientist. Since then she has significantly contributed to in the areas of modeling of satellite data and remote sensing. She is also a receipent of young scientiest award in 1989 from Indian Science Congress(ISCA) and Space Gold Medal in 2000 from Astronautical Society of India(ASI). She has published over 30 research articles in National and International journals.

J Saibaba has received his Ph.D. from IISC, Bangalore in 1985. His research interestes are in the areas of GIS applications and Remote sensing. He has significatnly contributed to various projects of Deartment of space involving analysis of satellite imagery.

Geeta Varadan has received her M.E. in Electronics from School of Automation, IISC, Bangalore. she has been very instrumental in design, development and setting up earth stations in different parts of the country. Her research interests are in the areas of satellite data processing, Signal analysis and Digital communications. She has served ISRO at various levels and capacities, currently she is Director, Advanced Data processing Research Institute.