

Wavelet denoising algorithm to refine noisy geoelectrical data for versatile inversion

A. Stanley Raj¹ · D. Hudson Oliver² · Y. Srinivas³ · J. Viswanath⁴

Received: 14 January 2016 / Accepted: 4 February 2016 / Published online: 22 February 2016
© Springer International Publishing Switzerland 2016

Abstract This paper presents a denoising technique based on wavelet algorithm for inverting geoelectrical resistivity data. The presented work compares different denoising process by thresholding wavelet algorithm. Discrete wavelet transform is used to denoise the geoelectrical resistivity data. It is suitable for applying vertical electrical sounding data. The optimum performance is obtained and the result is investigated under several constraints. This method can be adopted to any geophysical data for pre-processing. Daubechies wavelet functions ('db') of different decomposition levels with four ("rigrsure", "universal thresholding", "minimax", "heursure") thresholds were attempted and the significant reduction of noise is effectively done. The data is initially subjected to synthetic noisy data with various levels of signal to noise ratio (SNR) and tested results with optimum condition is implemented to the noisy field data which is verified with the nearby ground truth information. Error measures reveal that is algorithm is best suited for denoising the geoelectrical resistivity data.

Keywords Geoelectrical resistivity inversion · Wavelet denoising · Discrete wavelet transform · Neuro Fuzzy inversion

Introduction

Geophysical studies show ample evidences of the successful use of the resistivity method in groundwater prospecting. Many geophysical techniques have been applied to groundwater investigations with some showing more success than others. Among all the geophysical techniques, the geoelectrical resistivity method is more successful in groundwater studies. Direct current resistivity methods of geophysical exploration are in extensive use globally for aquifer mapping and estimation of aquifer parameters viz., resistivity and thickness of the subsurface earth. While applying the current to the subsurface of the earth, due to its heterogeneity, and nonlinear characteristics of various soils, rock formations and aquifer zones, it will effectively distract the applied current and gave us information about the subsurface, as the resistance varies from various contents of the subsurface earth. Due to the non linearity noises were encapsulated with the information. A better technique need to eliminate such noises. Here experimented wavelet denoising algorithm with different thresholding functions and the results proved it would be the best tool for purging the noise.

The use of Schlumberger electrode configuration for conducting vertical electrical sounding is more common. Vertical Electrical Sounding (VES) method is one of the efficient methods in identifying the aquifers in a more reliable manner. In addition to the interpretation of VES data using curve matching procedure through theoretical curves prepared by different computational methods, the inversion has been used for many years to interpret electrical resistivity data. The theoretical interpretation of geoelectrical resistivity data using Adaptive Neuro Fuzzy Inference System (ANFIS) was adopted in the research work. The processing of each component of the algorithm

✉ A. Stanley Raj
stanleyraj_84@yahoo.co.in

¹ Department of Physics, Loyola College, Nungambakkam, Chennai 600 034, India

² Department of Physics, Senthamarai College of Arts and Science, Madurai, India

³ Centre for GeoTechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu 627012, India

⁴ Department of Mathematics, Vel Tech University, Chennai 600062, India

of the investigation essentially was qualitative followed by quantitative analysis. Thus it has been validated with different field data. The field data include, the VES data collected from Kanyakumari field areas. The results are compared with earlier results/litholog for the published data and with litholog for the Kanyakumari field areas.

Methodology

Direct current resistivity methods of geophysical exploration are in extensive use globally for aquifer mapping and estimation of aquifer parameters viz., resistivity and thickness (Kosinky and Kelly 1981; Niwas and Singhal 1981; Mazac et al. 1985; Yadav and Abolfazli 1998). The physical basis of the resistivity method is based on the relative distribution of impressed current in the earth controlled by subsurface resistivity distribution. Different computational methods were adapted by the various researchers to invert geoelectrical data. (Flathe 1955; Van Dam 1964; Mooney et al. 1966; Ghosh 1971; Srinivas et al. 2010, 2012).

In general, the characteristic sounding curves are represented in multiple layers Each of the four sets has particular properties that may be roughly classified. Types H and K have a definite minimum and maximum, indicating a bed, or beds, of anomalously low or high resistivity, respectively, at intermediate depth. Types A and Q show fairly uniform change in resistivity the first increasing, the second decreasing with depth. Obviously these curves also may be combined (Telford et al. 1990). Here in this research paper different combination of curves were attempted for interpretation.

Theoretical background

Wavelet transform plays major role in analysing the data in broad spectrum of time–frequency. Event related potential determination can be easily done through this transformation more clearly. The spatial and temporal information will provide the real time data strategy in eagle’s eye. Advantages over the conventional time–frequency analysis lies in decompose the signal of any time–frequency multiresolution functions. The admissibility condition is as follows.

$\psi(t) \in L^2(\mathbb{R})$ satisfies certain admissibility conditions as

$$C\psi = \int_{\mathbb{R}} \frac{|\widehat{\Psi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{1}$$

$\Psi(t)$ is called wavelet; $\widehat{\Psi}(\omega)$ is fourier transform of $\Psi(t)$. The Continuous Wavelet Transform (CWT) is defined as

the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function,

$$\Psi_{a,b}(t) = |a|^{-\frac{1}{2}} \Psi\left(\frac{t-b}{a}\right) \tag{2}$$

$$W_f(a,b) = \left\langle f(t), \Psi_{a,b}(t) \right\rangle = \Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-a}{a}\right) \tag{3}$$

$a, b \in \mathbb{R}, a > 0$

Wavelet denoising

This application of wavelet denoising is a nonparametric regression analysis and it works well with added gaussian noise in the true resistivity data. The noisy data have the form

$$y_i = g(x_i) + e_i \tag{4}$$

Here $g(x_i)$ is the true function to be experimented and e_i which represents the noise added to the signal. This true data which additive noise is transformed to a denoised data which is of the form

$$d^* = d + \epsilon \tag{5}$$

where d is the transform off the true data and ϵ is the transform of the noise.

Hard thresholding

$$D(U, \lambda) = U \quad \text{for all } |U| > \lambda \\ = 0 \quad \text{Otherwise} \tag{6}$$

Soft thresholding

$$D(U, \lambda) = \text{sgn}(U) \max(0, |U| - \lambda) \tag{7}$$

Universal threshold

$$\lambda^u = \sigma \sqrt{2 \log n} \tag{8}$$

$$\sigma = \text{MAD}(d_{J-1}) = \text{median}_i(|d_{J-1,i} - \text{median}_k(d_{J-1,k})|) \tag{9}$$

Minimax threshold

It is better to know some prior information about the data, but it is difficult to understand the distribution of complex data. Explicitly, the obtained geoelectrical apparent resistivity data from different geologic conditions will suffer different kind of noises which could not be easily discriminated. Since there is no homogeneous model to fix the distribution, the data is never in uniform behaviour. For the data it is good to find a best estimator whose maximal risk is minimal among all estimators. The prior information of the signal set is Θ . But this set doesnot specify the

Table 1 Experimental result of Daubechies wavelet functions with different levels of decomposition for VES 1 data

Signal to noise ratio (SNR)	Rigrsure	Heursure	Sqtwolog	Minimax
db4 (Daubechies wavelet functions with level 4 decomposition)				
0	16.1764	15.2690	14.7549	15.7102
2	16.1762	15.2693	14.7728	15.6839
5	16.2468	15.2692	14.7780	15.7107
10	16.1761	15.2693	14.7594	16.2024
15	16.2151	15.2711	14.7698	15.7104
db8 (Daubechies wavelet functions with decomposition level 8)				
0	14.6184	14.6477	14.6376	14.6162
2	14.5999	14.6281	14.6369	14.5237
5	14.6153	14.6303	14.6376	14.6162
10	15.4245	14.6588	14.6374	14.464
15	14.5784	14.6333	14.7037	14.6157
db12 (Daubechies wavelet functions with level 12 decomposition)				
0	14.8393	15.3933	15.3169	15.3536
2	14.8392	15.3934	15.6494	14.3652
5	14.8393	15.3933	15.7449	14.4096
10	14.8393	15.7185	15.3971	15.3498
15	14.8386	15.3548	15.4761	15.3503
db16 (Daubechies wavelet functions with level 16 decomposition)				
0	16.3749	23.4714	23.4719	16.4533
2	16.5397	23.4728	23.4726	16.5117
5	16.5072	23.4728	23.4735	16.5117
10	16.5547	23.4895	23.472	16.5118
15	16.5396	23.4721	23.8899	16.5118

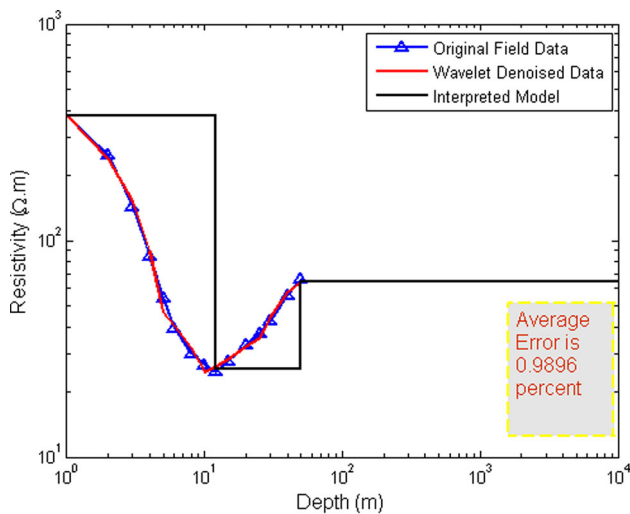


Fig. 1 VES 1 inverted layer model with average error

probability distribution of data in Θ . The more prior information smaller the set Θ (Mallat 2009). For the data $f \in \Theta$ with noise as given in Eq. (4). The risk estimation $\tilde{F} = DX$ is

$$r(D, f) = E\{DX - f^2\} \tag{10}$$

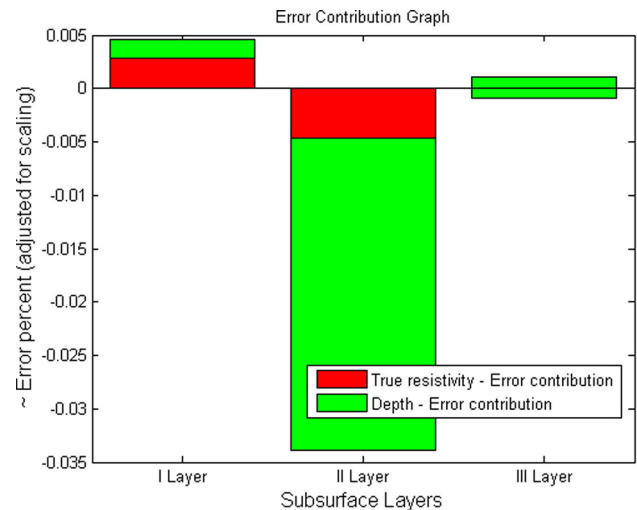


Fig. 2 Error contribution map for true resistivity and thickness in comparison with actual litholog of VES 1

The probability distribution of data in set Θ is unknown; the precise risk cannot be estimated. Only the range can be determined. The maximum risk of this range is given by (Donoho and Johnstone 1998)

$$r(D, \Theta) = \sup E\{DX - f^2\} \quad \text{with } f \in \Theta \tag{11}$$

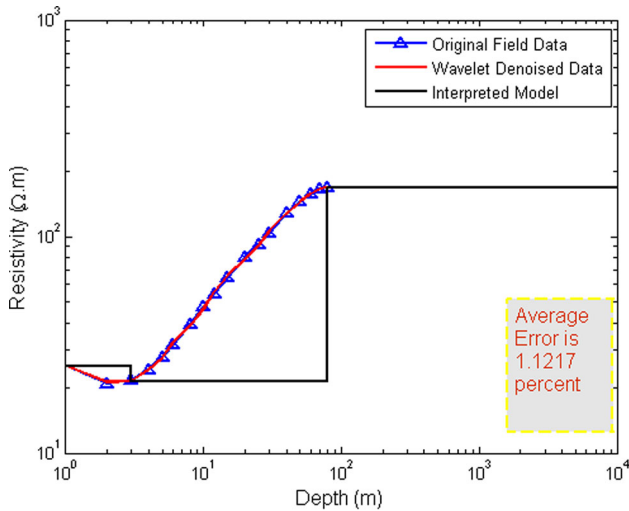


Fig. 3 VES 2 inverted layer model with average error

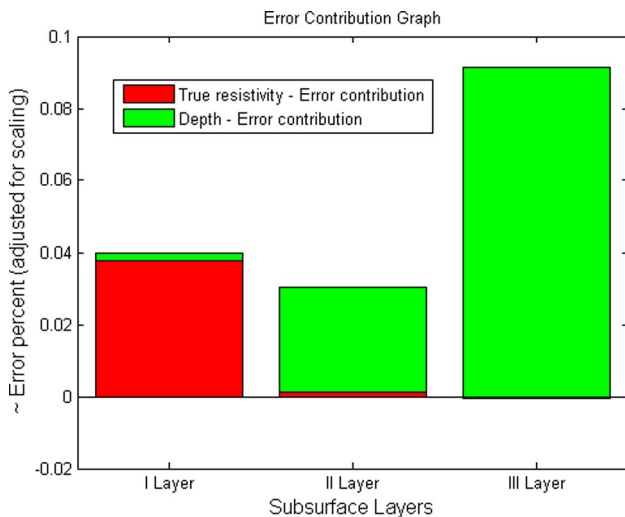


Fig. 4 Error contribution map for true resistivity and thickness in comparison with actual litholog of VES 2

In this estimation of minimax function, the lower bound risk of above equation with all possible with operators D is represented as

$$r_n(D, \Theta) = \text{Infr}(D \in \Theta) \quad \text{with} \quad D \in O_n \quad (12)$$

where O_n represents the sets of all operators

Rigrsure thresholding

It is a soft threshold evaluator of unbiased risk. $W = [\omega_1, \omega_2, \dots, \omega_N]$ is a vector consists of the square of wavelet coefficients from small to large. Select the minimum value $r_b(b^{\text{th}}r)$ from risk vector, which is given as,

$$R = \{r_i\}_{i=1,2,\dots,N} = \frac{[N - 2i + (N - i)\omega_i + \sum_{k=1}^i \omega_k]}{N} \quad (13)$$

as the risk value. The selected threshold is

$$\lambda = \sigma \sqrt{\omega_b} \quad (14)$$

where, ω_b is the b^{th} squared wavelet coefficient (coefficient at minimum risk) chosen from the vector and σ is the standard deviation of the noisy signal.

Heursure thresholding

If the signal to noise ratio is very small, the sure method estimation is not good to choose. The Heuristic threshold is given by,

$$\lambda = \begin{cases} \lambda_1 & A > B \\ \min(\lambda_1, \lambda_2) & A \geq B \end{cases} \quad (15)$$

where $A = \frac{S-N}{N}$ and $B = (\log_2 N)^{3/2} \sqrt{N}$. The N is length of wavelet coefficient vector and S is the sum of squared wavelet coefficients given as

$$S = \sum_{i=1}^N \omega_i^2 \quad (16)$$

A small threshold which gives noisy output and large threshold can cut significant part of signal thus the information contained in the signal may be lost.

Adaptive Neuro Fuzzy inversion

The role model for Soft Computing is the Human Mind. Unlike the conventional or hard computing, it is tolerant of imprecision, uncertainty and partial truth. It is also tractable, robust, efficient and inexpensive. Apart from neural networks and fuzzy logic, ANFIS is hybrid approach network architecture integrates the concepts of both ANN and FL. Thus the fundamental knowledge was supplied by the ANNs and the fuzzified membership function rules thus evaluate the real structure of the earth more clearly.

In order to make full use of soft computing for intelligent subsurface characterization, it is important to note that the knowledge based system has been framed carefully and the implementation of the hybrid systems should aim to improve prediction and its reliability. At the same time, the improved systems should contain small number of sensitive user-definable model parameters and use less CPU time. The future development of hybrid systems should incorporate various inter disciplinary knowledge and maximize the amount of useful information extracted between data types so that reliable extrapolation could be obtained. In

Table 2 Vertical electrical sounding locations with lithology

S.no	Location	Sounding Nos.
1.	8°6'31.79" N 77°30'50.29" E	VES 1
2.	8°8'48" N 77°38'39.8" E	VES 2
3.	8°11'30.89" N 77°29'47.6" E	VES 3

Legend
Kottaram

Lithology

- Topsoil
- Gneiss highly weathered
- Gneiss weathered

Table 2 continued

S.no	Location	Sounding Nos.
4.	8°8'44.8" N 77°29'0.38" E	VES 4
5.	8°10'14.4" N 77°30'31.39" E	VES 5
6.	8°9'8.3" N 77°30'1.99 E	VES 6
7.	8°10'12.5" N 77°29' 37.67" E	VES 7

Legend

Varioor

Lithology

- Gneiss highly weathered
- Gneiss weathered
- Gneiss fractured
- Clay
- Charnockite fractured

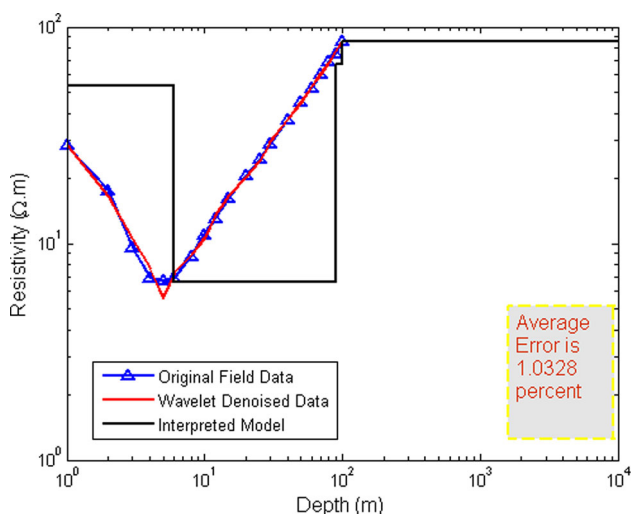


Fig. 5 VES 3 inverted layer model with average error

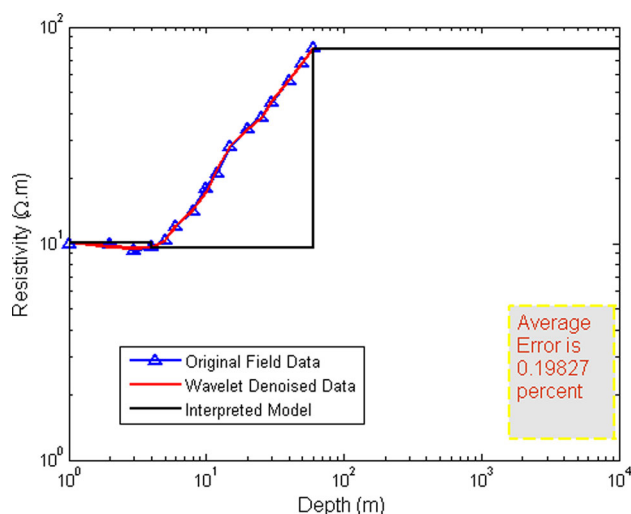


Fig. 7 VES 4 inverted layer model with average error

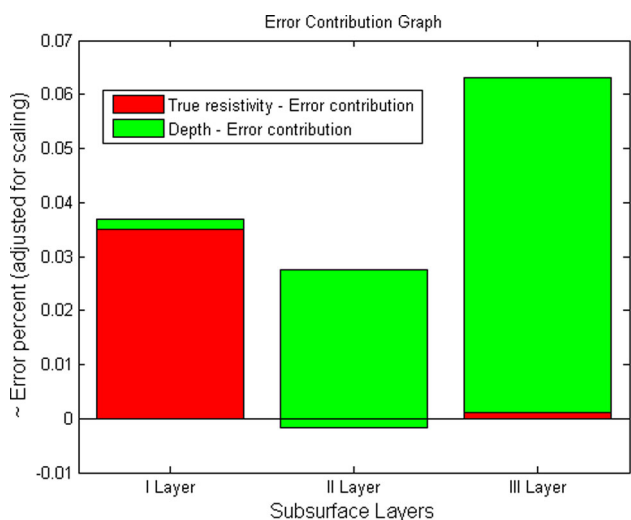


Fig. 6 Error contribution map for true resistivity and thickness in comparison with actual litholog of VES 3

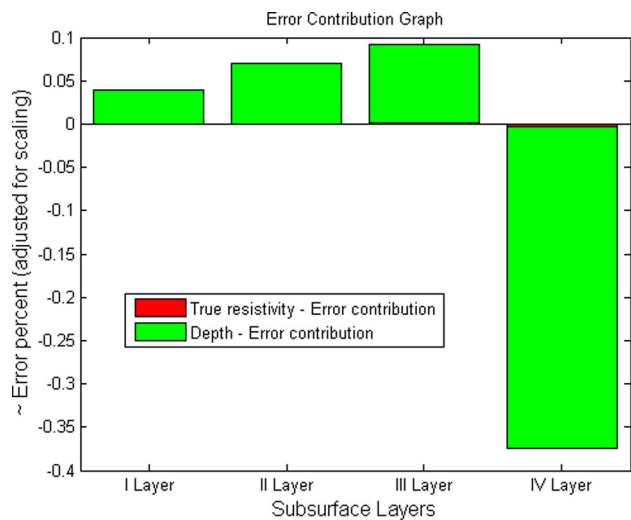


Fig. 8 Error contribution map for true resistivity and thickness in comparison with actual litholog of VES 4

recent years, considerable attention has been devoted to the use of hybrid neural-network/fuzzy-logic approaches as an alternative for pattern recognition, statistical and mathematical modeling. It has been shown that neural network models can be used to construct internal models that recognize fuzzy rules.

Neuro-fuzzy modeling is a technique for describing the behavior of a system using fuzzy inference rules within a neural network structure. The model has a unique feature in that it can express linguistically the characteristics of a complex nonlinear system. Geoelectrical resistivity method interpretation can be done with various tools. ANFIS network was developed by Jang 1993. An adaptive network is a multilayer feed forward network in which each node performs a particular function on giving the input to the network.

Electrical resistivity data collected using the vertical electrical sounding (VES) method is used for training the data set. The data collected from the field will be the apparent resistivity data, while on interpretation the subsurface parameters viz., resistivity and thickness of the individual layers are obtained. Trained data set will be the reference data for interpreting the subsurface layer parameters of the earth. This dataset will be used to train ANFIS by adjusting the membership function parameters that best models this data.

ANFIS architecture

The basic structure of the fuzzy inference system that maps input characteristics to input membership functions, input membership function to rules, rules to a set of output

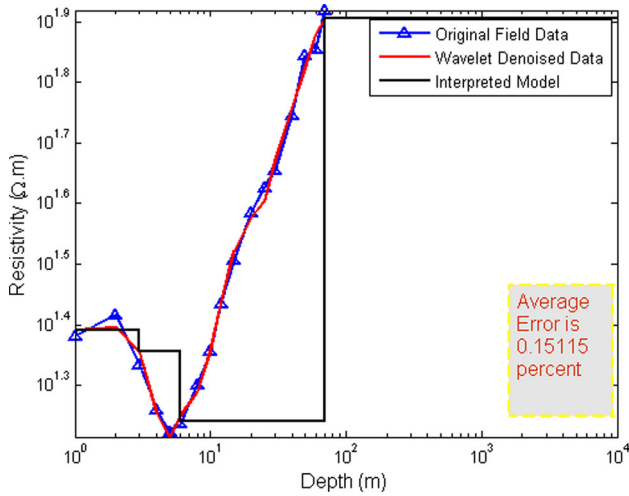


Fig. 9 VES 5 inverted layer model with average error

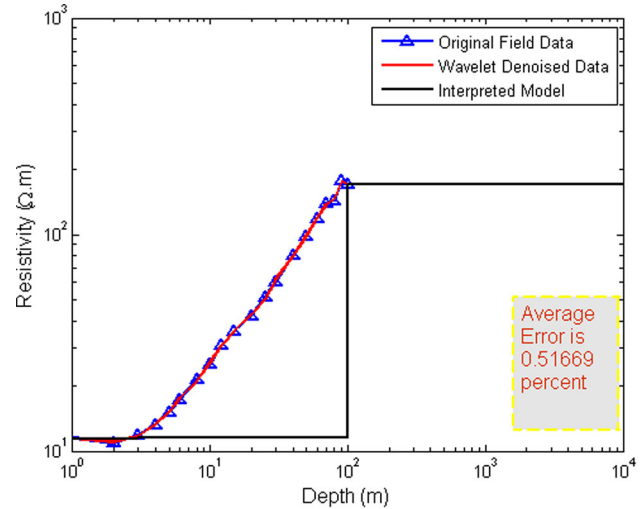


Fig. 11 VES 6 inverted layer model with average error

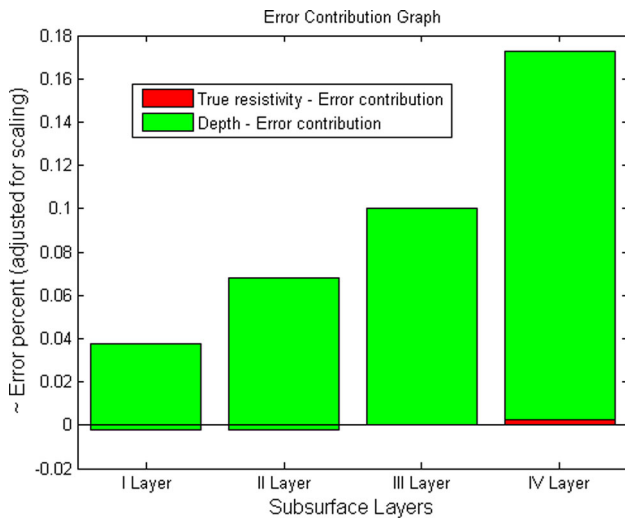


Fig. 10 Error contribution map for true resistivity and thickness in comparison with actual litholog of VES 5

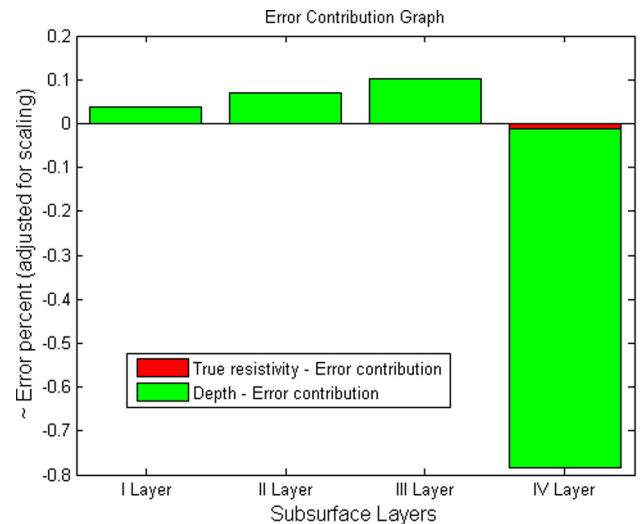


Fig. 12 Error contribution map for true resistivity and thickness in comparison with actual litholog of VES 6

characteristics, output characteristics to output membership functions, and the output membership function to a single-valued output or a decision associated with the output.

ANFIS system consists of 5 layers, layer symbolized by the box is a layer that is adaptive. Meanwhile, symbolized by the circle is fixed. Each output of each layer is symbolized by $O_{l,i}$ with i is a sequence of nodes and l is the sequence showing the lining. Here is an explanation for each layer, namely:

Layer 1

Serves to raise the degree of membership

$$O_{1,i} = \mu_A(x), i = 1, 2. \tag{17}$$

and

$$O_{1,i} = \mu_B(y), i = 1, 2. \tag{18}$$

with x and y are the input for the i -th node

$$\mu_A(x) = 1/[1 + (\det(x - ci)/ai)^{2bi}],$$

by $\{a_i, b_i$ and $c_i\}$ are the parameters of membership function or called as a parameter premise.

Layer 2

Serves to evoke *firing-strength* by multiplying each input signal.

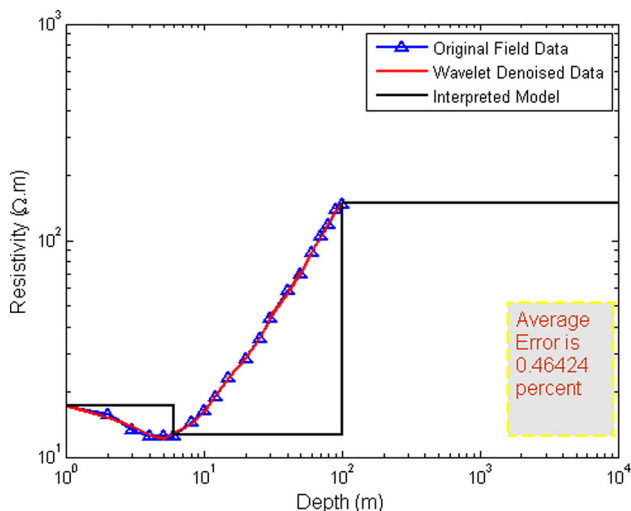


Fig. 13 VES 7 inverted layer model with average error

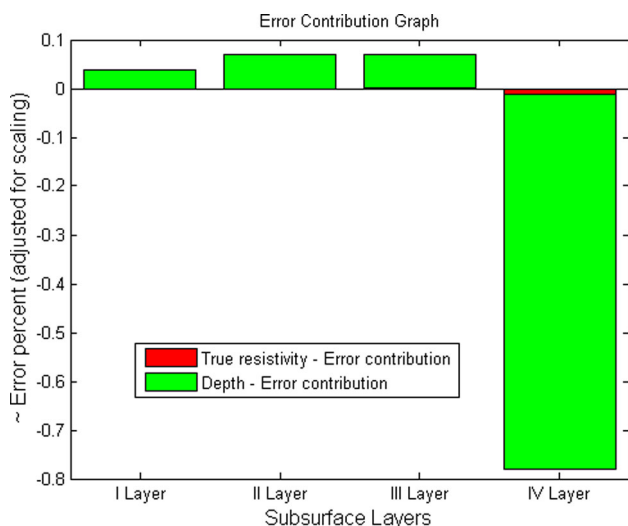


Fig. 14 Error contribution map for true resistivity and thickness in comparison with actual litholog of VES 7

$$O_{2,i} = w_i = \mu_A(x)\mu_B(y), i = 1, 2. \tag{19}$$

Layer 3

Normalize the firing strength

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2. \tag{20}$$

Layer 4

Calculating the output based on the parameters of the rule consequent $\{p_i, q_i \text{ and } r_i\}$

$$O_{4,i} = \bar{w}_i f_i = (p_i x + q_i y + r_i). \tag{21}$$

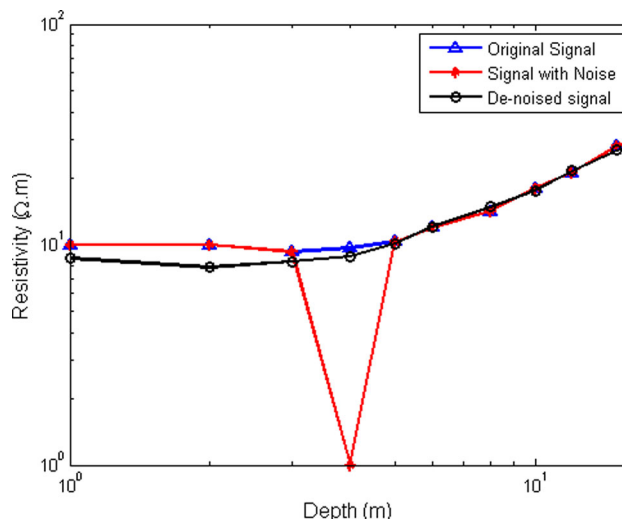


Fig. 15 Wavelet denoising application of noisy data using Daubechies wavelet functions with level 4 decomposition

Layer 5

Counting the ANFIS output signal by summing all incoming signals will produce

$$\sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \tag{22}$$

Functionally, there are almost no constants on the node functions of an adaptive network except piecewise differentiability.

Results and discussion

Resistivity sounding data were obtained from certain regions of Kanyakumari district (8°6'31.79" N; 77°30'50.29" E). Synthetic noise was added to the data with 2, 5, 10 and 15 % for which Rigrsure, Heursure, Sqtwolog, Minimax wavelet thresholding functions were applied and the results were experimented. The experimental results of VES 1 with certain noise percentage using Daubechies decomposition of various levels was tabulated (Table 1). The experimental analysis were done by using MATLAB software (MATLAB R 2013b). Besides noting down the performance analysis of various methods, the effect of decomposition levels also has been checked. For comparing and analysing the performance the true resistivity and depth data were compared with the litholog data available near to the data location.

Average error in true Resistivity

$$= \text{mean} \left(\frac{1}{N} \sum_{i=1}^N \text{Actual resistivity} - \text{Obtained resistivity} \right)$$

Table 3 Noise level estimation with respect to the correlation coefficient

S.No	Number of data points with field noise	Wavelet function (Daubechies)	Correlation coefficient	Percentage of noise	Comments
1.	1	db4	0.9903	1–5 %	Information sustains
2.	2	db4	0.9512	6–13 %	Information sustains
3.	3	db4	0.8004	14–20 %	Information distracted
4.	4	db4	0.4771	21–27 %	Information lost

Average error in depth

$$= \text{mean} \left(\frac{1}{N} \sum_{i=1}^N \text{Actual depth} - \text{Obtained depth} \right)$$

Where N represents the number of layers.

Based on the result universal threshold (Sqtwolog) shows the best performing threshold function of wavelet denoising for geoelectrical data inversion. Thus the VES data 1, 2, 3, 4, 5, 6, 7 were interpreted and the results were shown in figures. Figures 1 and 2 represents the VES 1 inverted model and error contribution of resistivity and thickness respectively. Figure 3 shows the interpreted model for VES 3 data and the corresponding error contribution is shown in Fig. 4. Table 2 shows the litholog for corresponding sounding data. Figures 5 and 6 shows the model with error contribution of VES 3. Figures 7, 9, 11 and 13 shows the interpreted layer model for sounding data 4, 5, 6 and 7 respectively. Figures 8, 10, 12 and 14 represents the corresponding error contribution of each VES data respectively. Error contribution graph shows the average error of each layer for true resistivity and depth taken into account. Approximate sharing of errors for each layer is shown in graphs.

Figure 15 shows the sample denoising application of wavelet denoising response to the noisy data. The correlation coefficient of wavelet denoising with respect to the information losing phenomena for VES 1 is shown in Table 3. The results obtained from Table 3 clearly portray the denoising capability of wavelet algorithm. Above 10–15 % the information gets distracted and gets lost finally when it reaches the noise level above 15 %.

Conclusions

In this paper, performance of various thresholding methods of wavelet viz., Sqtwolog, Rigrsure, Heusure, Minimax is applied for denoising geoelectrical sounding dataset. Vertical Electrical Sounding data obtained from Kanyakumari district was tested and verified with actual litholog section. The effect of various decomposition

levels were investigated on the performance of various thresholding functions. The results show that as level of decomposition increases from 2 to 15 % in correspondence to the information losing criteria during the wavelet denoising algorithm. As from the above studies it reveals that sqtwolog universal thresholding proves to be giving better result than the other algorithm which has been taken under consideration. This research work accompanies the wavelet denoising procedure which has been one of the best pre-processing algorithms for attempting to any signal decomposition or inversion techniques.

References

- Donoho DL, Johnstone IM (1998) Minimax estimation via wavelet shrinkage. *Ann Stat* 26(3):879–921
- Flathe H (1955) A practical method of calculating geoelectrical model graphs for horizontally stratified media. *Geophys Prospect* 3:268–294
- Ghosh DP (1971) Inverse filter coefficients for the computation of the apparent resistivity standard curves for horizontally stratified earth. *Geophys Prospect* 19:769–775
- Jang JSR (1993) Adaptive-network based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23:665–685
- Kosinky WK, Kelly WE (1981) Geoelectrical sounding for predicting aquifer properties. *Groundwater* 19:163–171
- Mallat SG (2009) Denoising. In: Mallat SG (ed) *A wavelet tour of signal processing: the Sparse way*. Elsevier/Academic Press, Amsterdam, Boston, pp. 535–606. ISBN 13:978-0-12-374370-1
- MATLAB R (2013b) The Mathworks, Inc., Natick, MA
- Mazac O, Kelly WE, Landa I (1985) A hydrophysical model forrelation between electrical and hydraulic properties of aquifer. *J Hydrol* 79:1–19
- Mooney HM, Orellana E, Pickett H, Tornheim L (1966) A resistivity computation method for layered earth model. *Geophysics* 31:192–203
- Niwas S, Singhal DC (1981) Estimation of aquifer transmissivity from Dar-Zarrouk parameters in porous media. *J Hydrol* 50:393–399
- Srinivas Y, Stanley Raj A, Hudson Oliver D, Muthuraj D, Chandrasekar N (2010) An application of artificial neural network for the interpretation of three layer electrical resistivity data using feed forward back propagation algorithm. *Curr Dev Artif Intell* 1:1–11

- Srinivas Y, Stanley Raj A, Hudson Oliver D, Muthuraj D, Chandrasekar N (2012) A robust behaviour of feed forward back propagation algorithm of artificial neural networks in the application of vertical electrical sounding data inversion. *Geosci Front* 3(5):729–736
- Telford WM, Geldart LP, Sheriff RE (1990) *Applied geophysics*, 2nd edn. Cambridge University Press, Cambridge, pp 661–663
- Van Dam JC (1964) A simple method for the calculation of standard graphs to be used in geoelectrical prospecting, Ph.D. Thesis, Delft Technological University, The Netherland
- Yadav GS, Abolfazli H (1998) Geoelectrical soundings and their relationships to hydraulic parameters in semi arid regions of Jalore, North West India. *J Appl Geophys* 39:35–51