Resolution Enhancement of GPR Data Using Bayesian Filtering Optimized by Machine Learning Algorithms

Alireza HAJIAN (dralirezahajian@gmail.com)
Islamic Azad University Najafabad Branch https://orcid.org/0000-0002-0931-9339
Rohollah Kimiaefar
Islamic Azad University Najafabad Branch

Research Letter

Keywords: GPR Data Filtering, Bayesian Filtering, ANFIS, FCM

Posted Date: June 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-38051/v1

Resolution Enhancement of GPR Data Using Bayesian Filtering Optimized by Machine Learning Algorithms

Rohollah Kimiaefar[1] ,Alireza Hajian[2,*]

1,2 Department of Physics, Najafabad Branch, Islamic Azad University, Najafabad, Iran.

*Email: dralirezahajian@gmail.com (Corresponding author)

ABSTRACT:

Blurring coherence events is the result of applying many spatial and temporal filtering algorithms when they are applied in order to suppress background random noise. Bayesian Filtering (BF) also suffers from mentioned problem. This paper develops a method for optimizing BF by Adaptive Neuro-Fuzzy Inference System (ANFIS) and Fuzzy C-Mean (FCM) clustering. First the structure of the GPR image is extracted using FCM. The structure and output of the BF for a random part of the data are used to produce output values for training ANFIS and after that, by generalizing the trained network to all data, filtered data would be achieved. The proposed method is applied on synthetic data-sets as well as two real 2-D GPR images gathered in an environmental study project. Performance of the method is evaluated by comparing the results of the proposed method to the output of BF. In synthetic data, the SNR value improved 63 percent more than of BF's output and the visual comparison of the results are suggesting better performance in noise cancellation and resolution enhancement, both in synthetic and real data-sets.

KEYWORDS: ***GPR Data Filtering, Bayesian Filtering, ANFIS, FCM***.

1. INTRODUCTION

GPR data are contaminated with random noise like all other kinds of geophysical data and this unwanted part of the raw records, negatively affects all further processing and interpreting steps. Furthermore the resolution of the data, in terms of continuity of events, is inversely proportional to the amount of the background random noise [1].

Characteristics of the transmitted signal, geology and the amount of the existing random noise, are the most important parameters that a supervising expert would consider and attenuate random noise by using conventional methods such as, filtering based on Hilbert transform [2], median and Wiener adaptive filtering [3] and methods based on multiresolution transforms [4, 5]. In recent years, methods based on artificial intelligence and soft computing are getting more and more popular in all aspects of geophysical methods [6]. Approaching to semi- and fully- automated methods is one of the main reasons of this commonness [7, ????????].

This paper intends to enhance the resolution of the GPR profile data by attenuating background random noise, utilizing powerful potential of the ANFIS and FCM in model discrimination and the ability of BF in random noise attenuation. The automation of the procedure and the performance of the method, in cases where signal to noise ratio is low, are also in the focus of attention.

2. BAYESIAN FILTERING

Bayes filters are a probabilistic tool for estimating the state of dynamic systems based on the bayesian formula where the state of the system refers to the package of dynamic variables which fully describe the system. The noise in the measurements are considered uncertainly even if the true system state is known. The measurements are not deterministic functions of the state, but are considered as distribution of possible values [50]. The time evolution is modeled dynamically that is perturbed by a certain process noise which is used for modeling the uncertainties in the system

dynamics. In most cases, the system is not truly stochastic, but is considered in such way to represent the model uncertainties.

Using a stochastic discrete-time state transition (Eq. 1) and an observation process (Eq. 2), any nonlinear stochastic system could be defined as [8]:

$$x_n = f_n(x_{n-1}, w_{n-1}) \quad (1)$$

$$y_n = h_n(x_n, v_n) \quad (2)$$

For the n-th sample (or time), x_n stands for system state vector that is not usually observable. w_n is the noise vector, y_n is observation vector and v_n is the observation noise vector. f_n and h_n relate the prior state to the current state and the current state to the observation vector.

The main problem in a typical bayesian context is to measure the posterior density, $p(x_n|y_{1:n})$, where the set of observations are defined by $y_{1:n} \equiv \{y_1, y_2, ..., y_n\}$.

The mentioned nonlinear and non-Gaussian state-space model (Eq. 1), specifies the predictive conditional transition density, $p(x_n|x_{n-1}, y_{1:n-1})$, of the current state, considering the previous state and all previous observations. Also, the observation process equation, Eq. 2, determines the likelihood function of the current observation given the current state, $p(y_n|x_n)$ [8].

The prior probability is defined by Bayes' rule as:

$$p(x_n|y_{1:n-1}) = \int p(x_n|x_{n-1}, y_{1:n-1}) p(x_{n-1}|y_{1:n-1}) dx_{n-1} \quad (3)$$

$p(x_{n-1}|y_{1:n-1})$ is the previous posterior density and The correction step generates the posterior probability density function from:

$$p(x_n|y_{1:n}) = c \times p(y_n|x_n) \times p(x_n|y_{1:n-1}) \quad (4)$$

Where c is a normalization constant.

The filtering problem is to estimate the first two moments of x_n given $y_{1:n}$. In a recursive manner and for a general distribution, $p(x)$, this consists of the recursive estimation of the expected value of any function of x, like $\langle g(x) \rangle_{p(x)}$, using Eq. 3 and Eq. 4 and requires calculation of integrals of the form [8]:

$$\langle g(x) \rangle_{p(x)} = \int g(x)\, p(x) \tag{5}$$

Usually this integral is solved by numerical approximation as the multivariate distribution of these kinds of integrals cannot be evaluated in closed form.

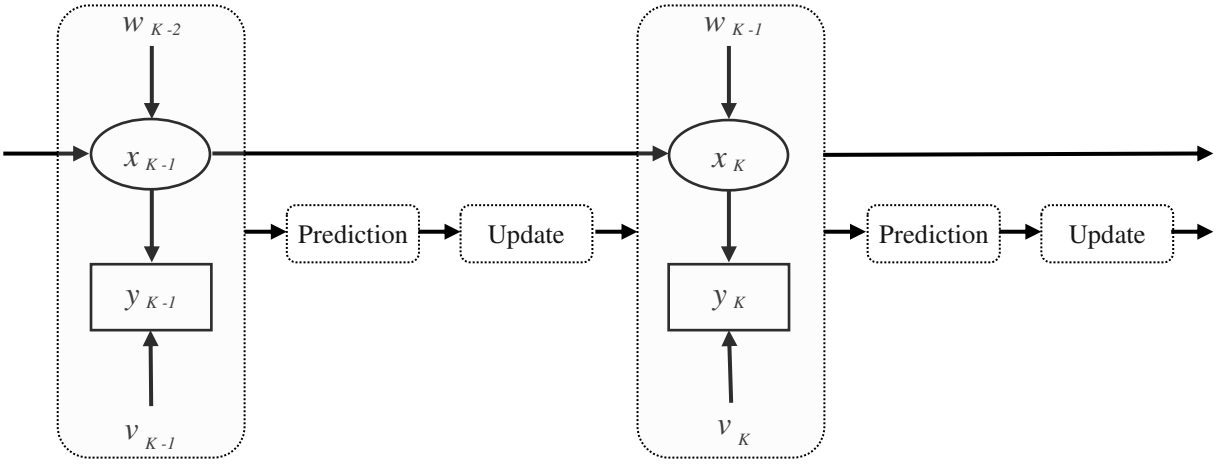

Figure 1: Schematic representation of bayesian filtering.

3. METHOD

As illustrated in Fig. 1, in the proposed method, at the very first stage, noise level of the input data is calculated based on the fact that, higher noise level cause more difference between original

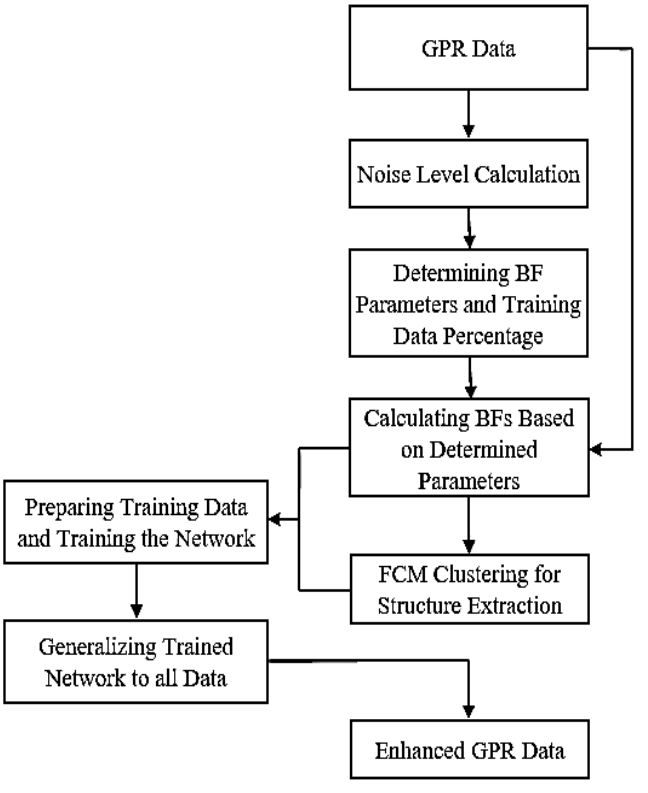

Figure 1: flowchart of the proposed method

(noisy) and filtered value and vice versa. Hence computing correlation of the outputs of BFs with different setups, logically could present an estimation of noise level [10].

By comparing known value of the noise variance for a synthetic data (Fig. 4a) with the calculated noise level, based on mentioned logic in an empirical manner, this method is evaluated and the result is illustrated in Fig. 2. The calculated noise level has could be considered as the noise variance with an acceptable approximation. It should be noticed that calculated noise level changes almost linearly and noise level increase (decrease) could be considered as noise variance increase (decrease). The calculated noise value has two usage in the method. It is used for determining

windows size range (for instance, 3 by 3, to, 17 by 17) and also for determination of training data percentage.

At the next step, outputs of BF with different window sizes are calculated. These values are suppliers of FCM clustering input data, for extracting structure of the data, which would be very useful in, especially, denoising data with low signal to noise ratio. Also mentioned values are used for ANFIS training as part of the input data. Among different fuzzy clustering methods, FCM was chosen mainly because of its performance in solving problems in diverse issues [11].

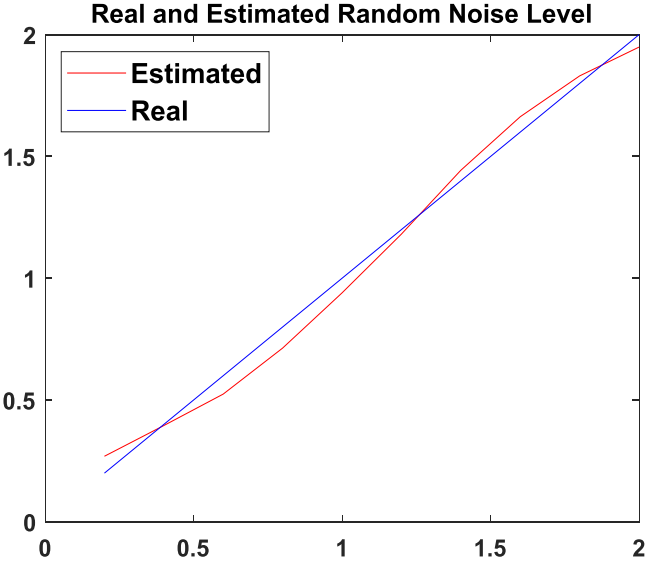

Figure 2: real and calculated noise level for a synthetic 2-D data with different noise variance from 0.2 to 2.

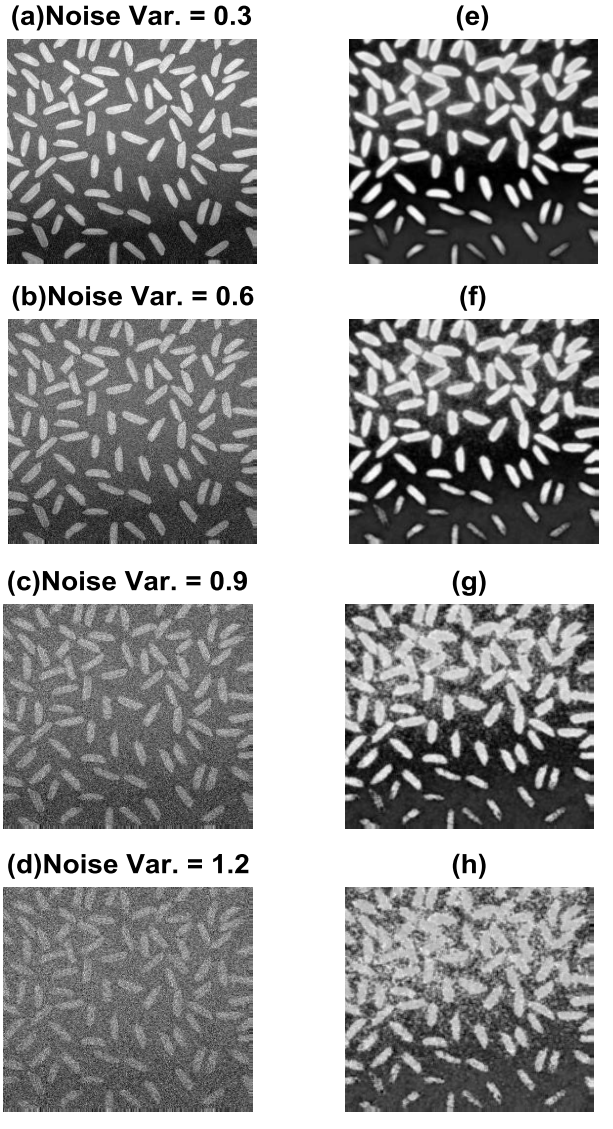

Figure 3: "rice" image with different level of random noise (a to d) and the structure extracted by FCM clustering (g to h respectively).

The structure extraction procedure is evaluated by performing the method to the "rice" with different amount of additive white gaussian random noise. The results of FCM clustering, structures, are shown in Fig. 3. The noise level is expressed by noise variance here. Although the clustered data (right column of images is Fig. 3) are showing to have less random components, but the main reason for calculating them is the ability of providing data structure specially in case data has high level of additive noise, like Fig. 3c and 3d. In such mentioned circumstances, usually

96 denoising algorithms make the output's details smooth and unattainable. Averaging structure with
97 other components of the output could be an alternative for regular outputs and a way for
98 overcoming the problem.
99 Using original data and the set of BF values for each selected points (as ANFIS input data) and a
100 weighted average of mentioned values as well as extracted structure (as ANFIS output), the
101 training pairs for ANFIS network will be ready. The output of this stage will be automatically
102 achieved by generalizing the trained network to all data. This optimized output or as named here as
103 Bayesian-ANFIS Filtered (BAF) data, will be determined.

104 **4. EXPERMIENTS**

105 The method was tested on one synthetic and two real GPR data-sets which were extracted from a
106 data-set recorded during an environmental study handled by United States Geological Survey
107 (USGS) [12].

At the very first step, some zero mean white gaussian noise was added to the synthetic 250 by 250 pixel, layered image. After that, calculating noise level, suggested 10 percent for training data percentage and a maximum of 17 by 17 window size for BF calculation. A set of 36 inputs including BF values calculated in 3 to 17 square shape neighborhood, and the value of the pixel (in noisy data) were contributed in FCM. The number of clusters as mentioned before is two. The output of FCM provided structure of the data (this structure could be very beneficial in data with low SNR values). The procedure of filtering proceeded by randomly selecting 10 percent of the provided data and by weighting the BF values, the output for each row of input data was determined. After training data ANFIS network

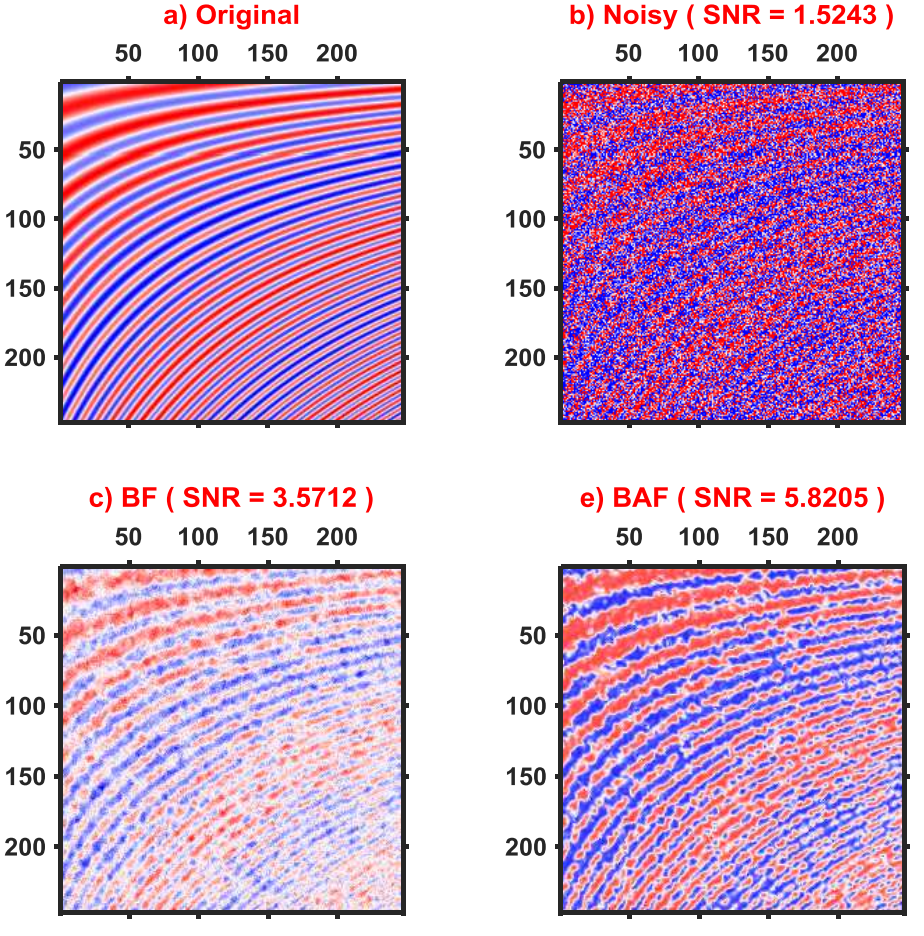

Fig. 4: (a) A synthetic GPR data with varying dips and layer thickness, (b) noisy version (Gaussian random noise added to original image), denoised images by (c) BF and (d) BAF. The output of proposed method is obviously brighter and more interpretable and the SNR has been improved 63 percent more than the output of

(Sugeno-type fuzzy inference system), using an initial grid partitioning fuzzy inference system and a hybrid optimization method (least-squares estimation with backpropagation), the trained network was generalized to all data resulting filtered data.

The original noise-free synthetic GPR data, noisy version and the data filtered by BF and BAF are illustrated in Fig. 2. The SNR values shown in this figure are calculate by Eq. 6 as:

$$SNR = \frac{P_s}{P_n} \quad (6)$$

Where P_s and P_n refer to average power of noise-free data and noise, respectively.

Confirmed by SNR values, obviously the performance of BAF is better than BF in eliminating random noise. Note to the high amount of the additive random noise in this data, as the method's ability could be evaluated better in such circumstances.

Beside visual and SNR confirmation and as another assessment tool, the cross-section of the original, noisy and filtered images (50th row and column of mentioned data) are plotted in Fig. 4. In terms of performance, efficiency of BAF in recovering the noisy signal could be visually corroborated. In deeper times and far offsets, (Fig. 4a and 4b and samples located between samples 150 to 250, where the frequency of layering is relatively higher) original trace has been recovered better by BAF.

It should be in the focus of attention that both BF and BAF outputs are structurally same although modified version of BF is more robust in separating noise from original data and resolution enhancement in an adaptive manner.

Fig. 5: Cross sections for data plotted in Fig. 4, at column (top) and row (bottom) No. 50. Higher performance could be visually confirmed for the BAF's output, as its plot is closer to original data plot.

As the results from synthetic experiment exhibited the robustness of the proposed method, in the following, the method is applied to real field data-sets. Illustrated in Fig. 6 and 7, two real GPR data-sets [12] which are covered with different level of random noise are selected. This first data-set (Fig. 6) is selected mainly because there are almost two zones with different amplitude characteristics. The two-phase structure of the data could result in data masking or causing artifacts on the output. The bottom–left corner of the data which is specified by yellow box, contains some faulting coherent events. It is expected that an ideal denoising algorithm not to perform much changes in the structure of this events. For the BF's output, these events are not clearly showing

mentioned faulting system especially in contrast with BAF's output.

The second real data-set (Fig. 7) is covered highly with random components. Many of coherent energy is masked and detecting layers could not easily be handled. Having applied BF and BAF

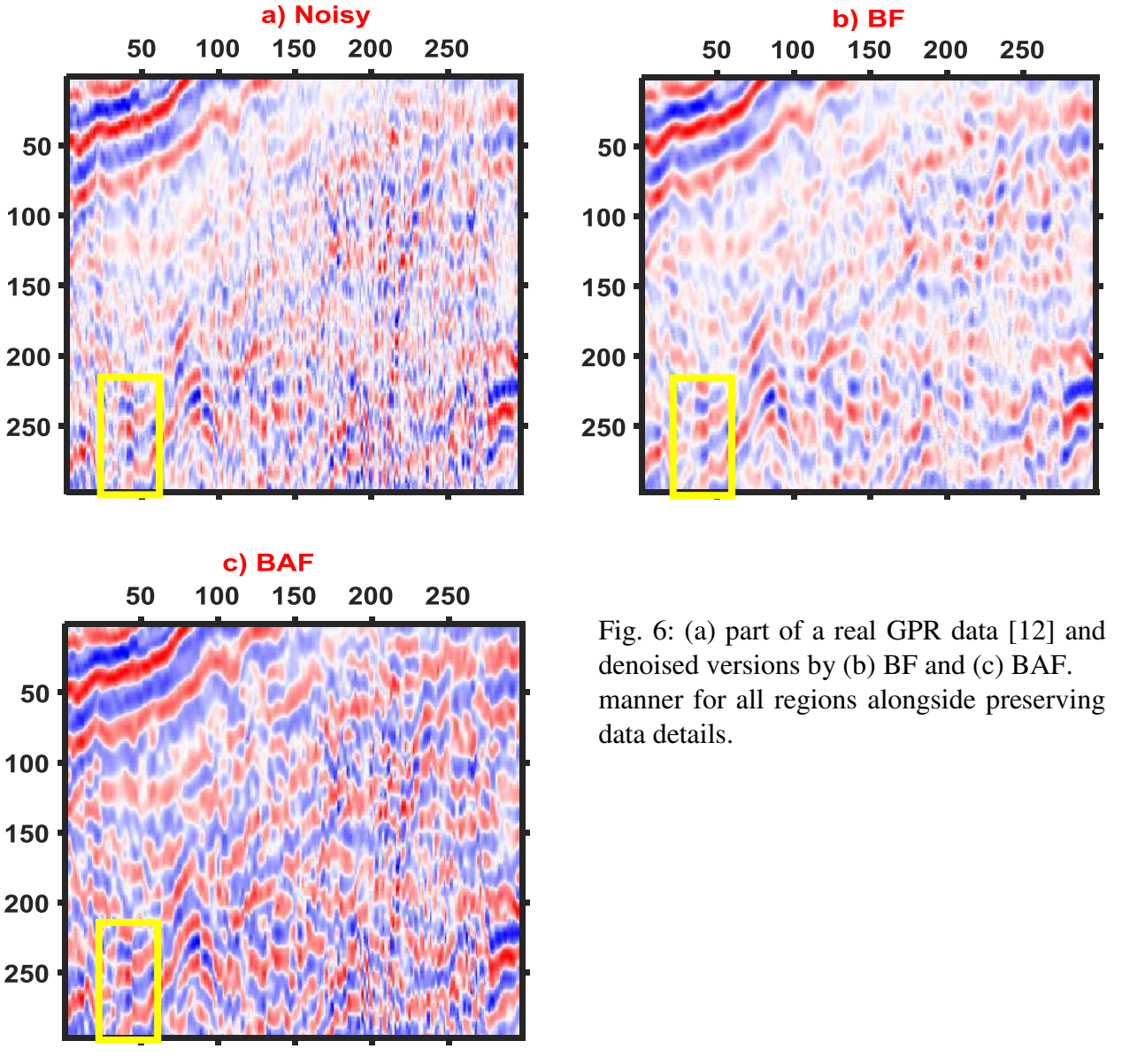

Fig. 6: (a) part of a real GPR data [12] and denoised versions by (b) BF and (c) BAF. manner for all regions alongside preserving data details.

(parts b and c in Fig 7) and comparing the resolution of the outputs, confirms the claim that the proposed, ANFIS modified version of BF, can be considered as a powerful alternative for the original BF. In BAF's output, random energy is attenuated strongly in all part of the data in an even manner while the details are preserved in the data.

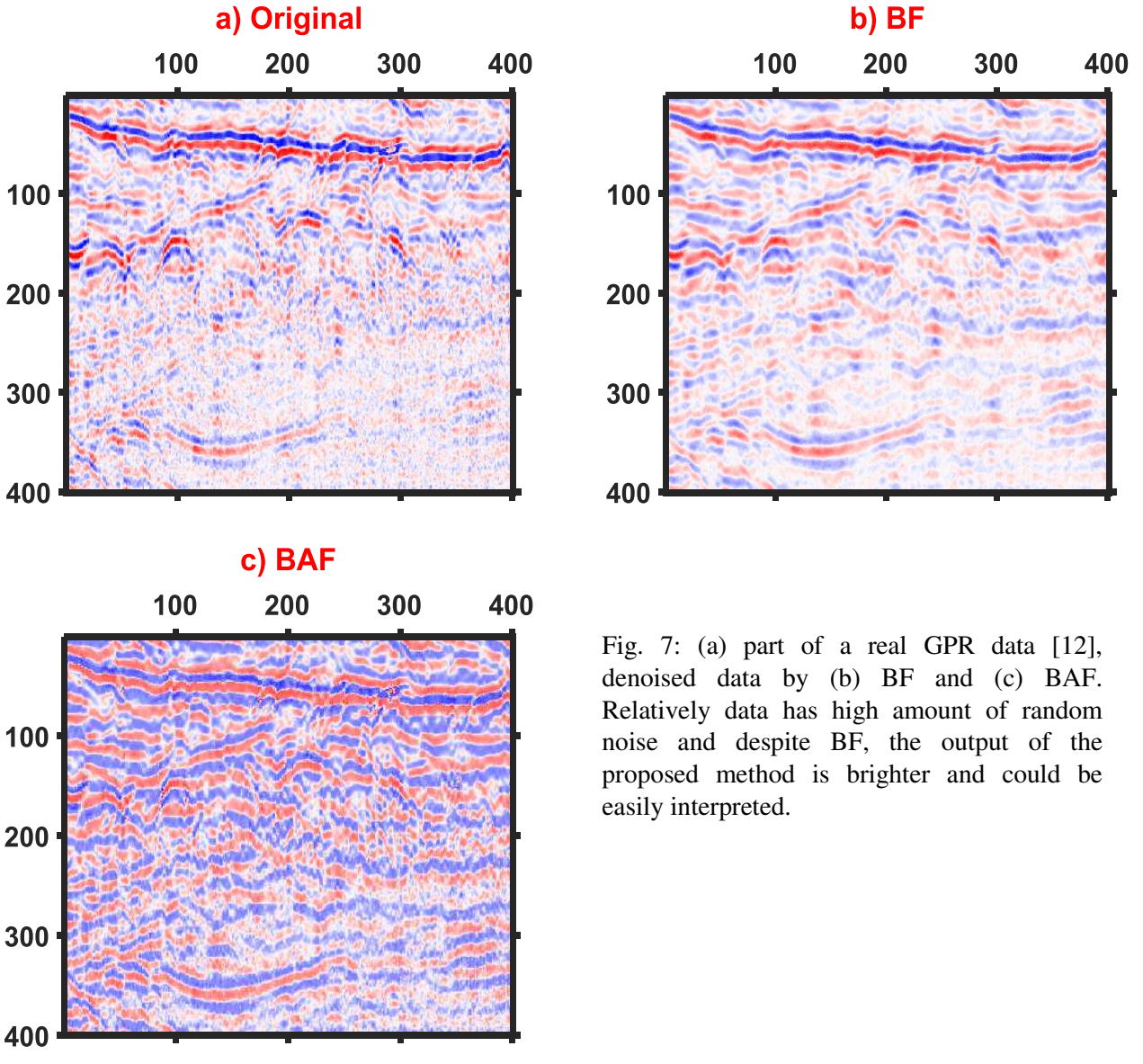

Fig. 7: (a) part of a real GPR data [12], denoised data by (b) BF and (c) BAF. Relatively data has high amount of random noise and despite BF, the output of the proposed method is brighter and could be easily interpreted.

5. CONCLUSIONS

In the method introduced in this paper, the ability of ANFIS and Fuzzy Clustering in model discrimination and problem solving was used for attenuating random noise in GPR data-sets. BF calculated with different setups, was used in noise-signal separation, structure extraction and automation of algorithm. Noise level estimation measure was also used here for determination of optimized output. The method was applied on one synthetic and two real data-sets and the results

showed that BAF has better performance in SNR improvement (more than 2 times) that of BF. In real data-sets, resolution enhanced by proposed method definitely better and event tracing was easier in BAF output.

REFERENCES

[1] R. E. **Sheriff**, Seismic resolution a key element. **1997**, AAPG Explorer: **18(10)** pp. 44-51.

[2] F. **Guangyou**, M. **Pipan**, Synthetic and field examples of ground-penetrating radar (GPR) using two-phase detection techniques. **2003**, *Geophysics*: **68 (2)**: pp. 554–558.

[3] Y. **Jeng**, Y. **Li**, C. **Chen**, H. **Chien**, Adaptive filtering of random noise in near-surface seismic and ground-penetrating radar data, **2009**, Journal of Applied Geophysics: **68(1)**, pp. 36-46.

[4] Y. **Jeng**, C. **Lin**, Y. **Li**, C. **Chen**, H. **Yu**, Application of sub-image multiresolution analysis of Ground-penetrating radar data in a study of shallow structures, **2011**, Journal of Applied Geophysics, **73(3)**, pp. 251-260.

[5] A. **Tzanis**, The Curvelet Transform in the analysis of 2-D GPR data: Signal enhancement and extraction of orientation-and-scale-dependent information, **2015,** Journal of Applied Geophysics: **115**, pp. 145-170.

[6] R. **Kimiaefar**, H. **Siahkoohi**, A. **Hajian**, A. **Kalhor**, Seismic random noise attenuation using artificial neural network and wavelet packet analysis, **2016**, Arabian Journal of Geosciences, **9**: 234.

[7] R. **Kimiaefar**, H. **Siahkoohi**, A. **Hajian**, A. **Kalhor**, Random noise attenuation by Wiener-ANFIS filtering, **2018**, Journal of Applied Geophysics: **159** pp. 453-459.

[8] A. J. Haug, A Tutorial on Bayesian Estimation and Tracking Techniques Applicable to Nonlinear and Non-Gaussian Processes, 2005, MITRE Technical Report. MTR 05W0000004, The MITRE Corporation.

[9] X. **Liu**, M. **Tanaka**, M, **Okutomi**, Noise Level Estimation Using Weak Textured Patches of a Single Noisy Image, **2012**, IEEE International Conference on Image Processing.

[10] R. **Fani**, H. **Hashemi,** Random noise attenuation by application of GK clustering on relevant seismic attributes. **2011**, 124th SEG Conference, Tokyo, Japan.

[11] R. **Suganya**, R. **Shanthi**, Fuzzy C- Means Algorithm- A Review, **2012**, International Journal of Scientific and Research Publications, **2(11)**: pp. 2250-3153.

[12] A.S. **Forde**, C.G. **Smith**, B.J. **Reynolds**, 2016, Archive of ground penetrating radar data collected during USGS field activity 13BIM01—Dauphin Island, Alabama, April 2013: U.S. Geological Survey Data Series 982, http://dx.doi.org/10.3133/ds982.

[50] S. **Sarkka**, Bayesian Filtering and Smoothing, **2013**, Cambridge University Press.

Figures

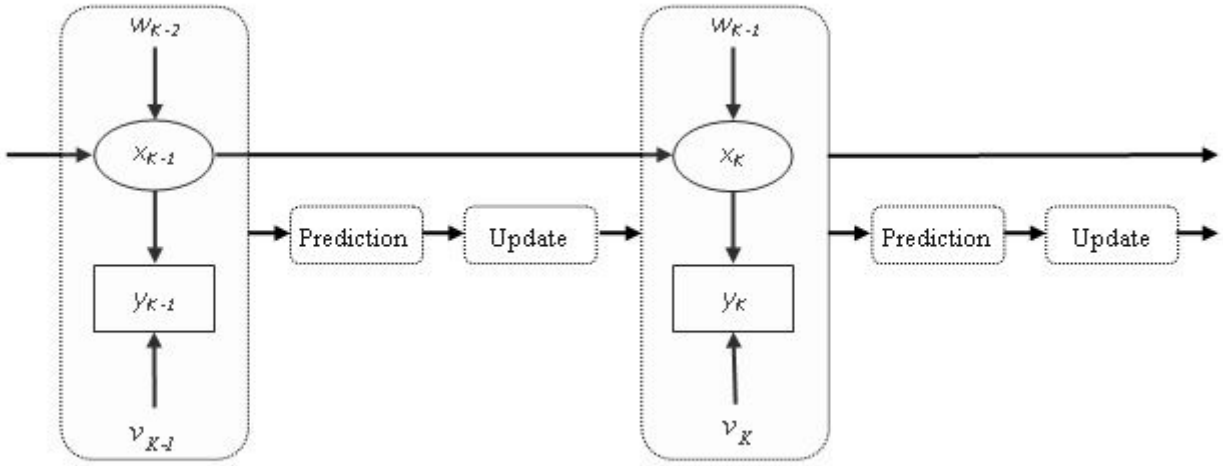

Figure 1 : Schematic representation ofbayesian filtering.

Figure 1

Schematic representation ofbayesian filtering.

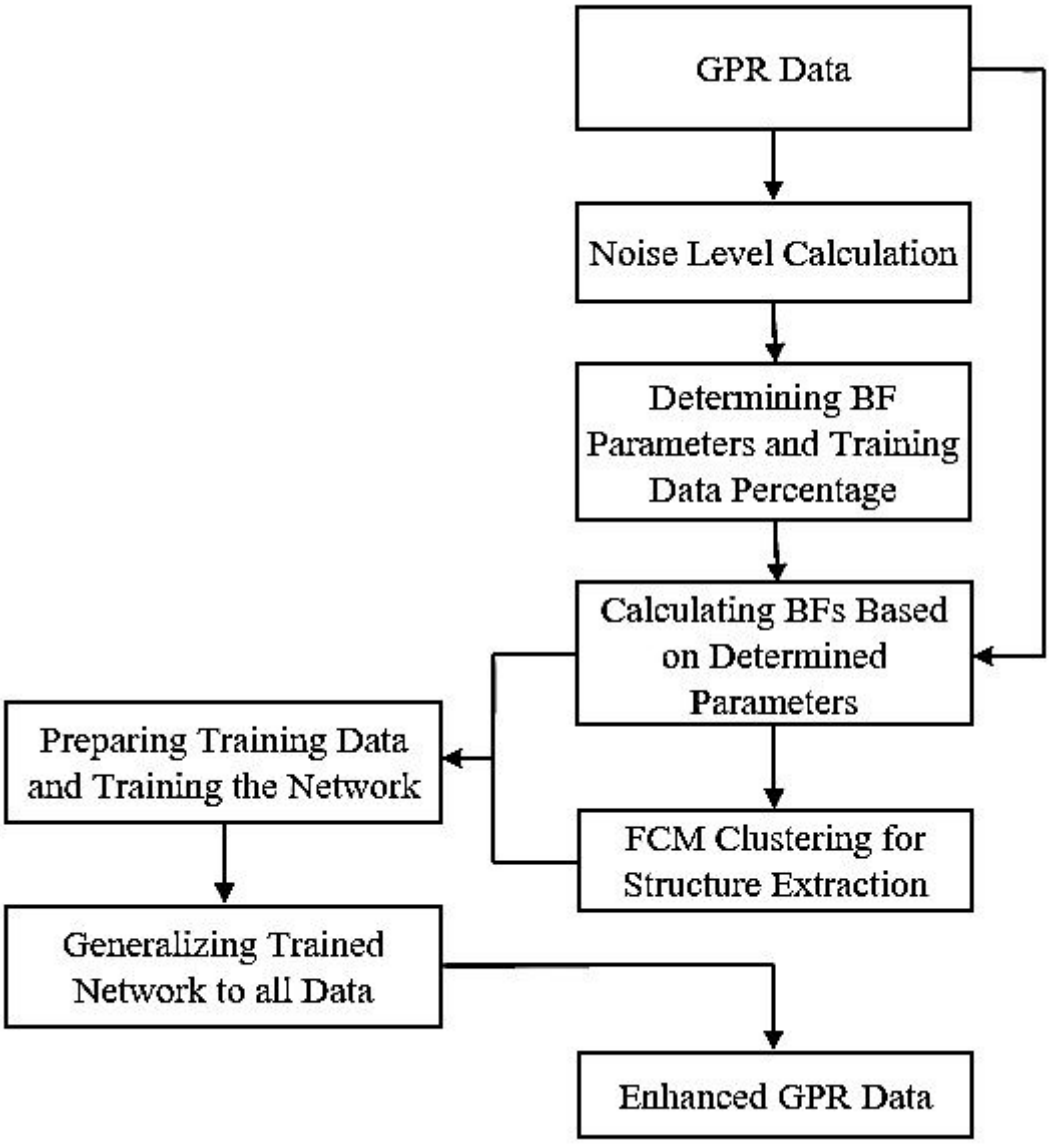

Figure 2

flowchart of the proposed method

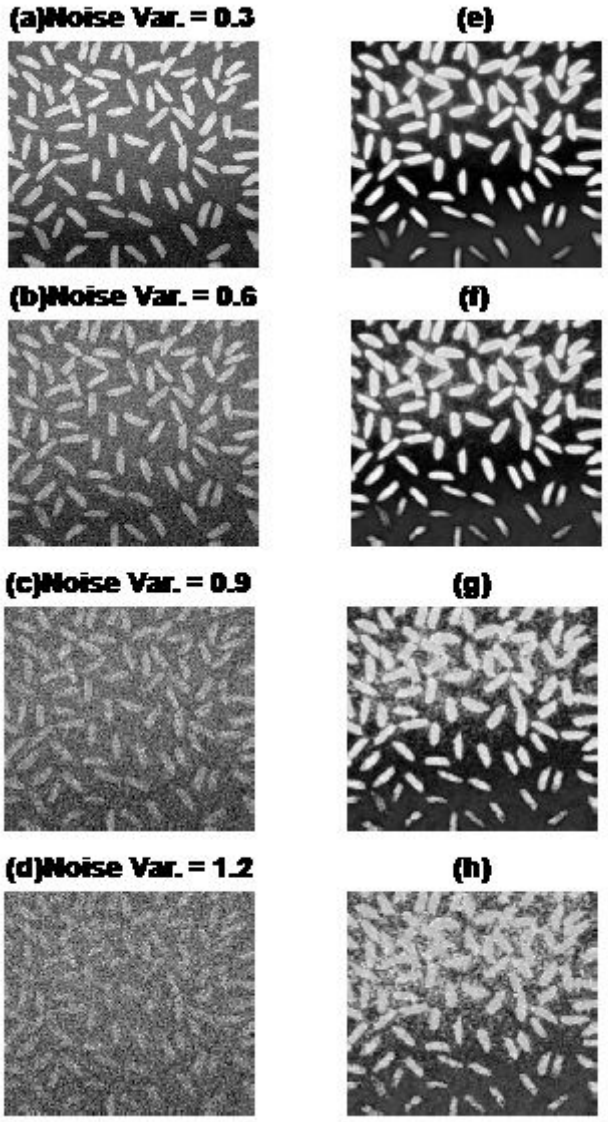

Figure 3

“rice” image with different level of random noise (a to d) and the structure extracted by FCM clustering (g to h respectively(.

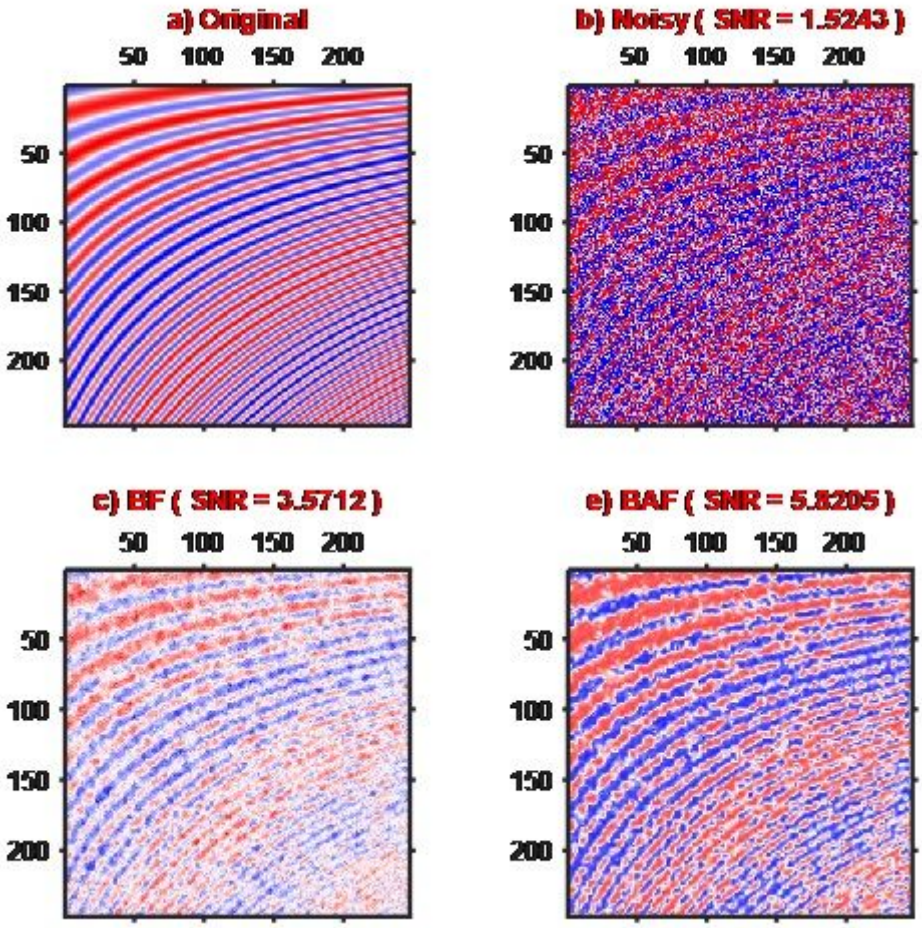

Figure 4

(a) A synthetic GPR data with varying dips and layer thickness, (b) noisy version (Gaussian random noise added to original image), denoised images by (c) BF and (d) BAF. The output of proposed method is obviously brighter and more interpretable and the SNR has been improved 63 percent more than the output of BF.

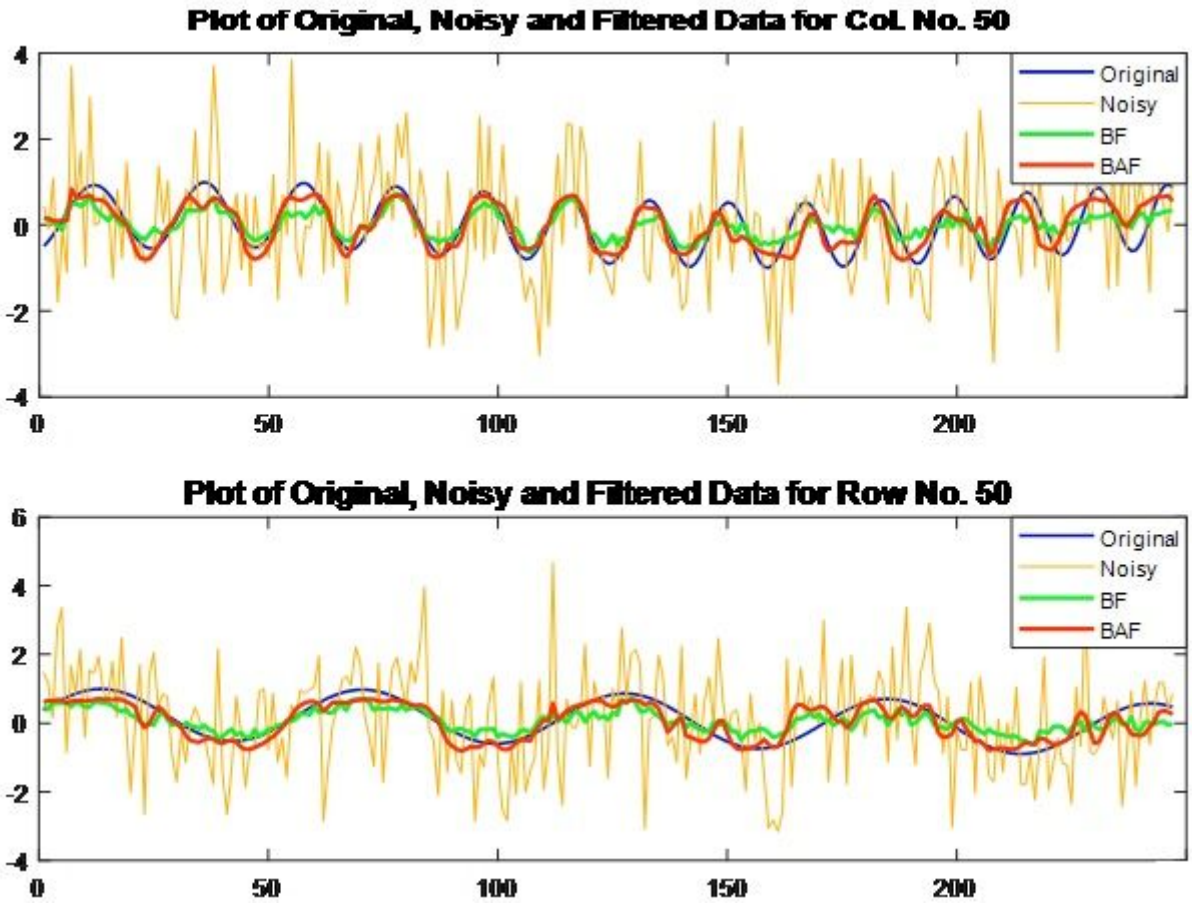

Figure 5

Cross sections for data plotted in Fig. 4, at column (top) and row (bottom) No. 50. Higher performance could be visually confirmed for the BAF's output, as its plot is closer to original data plot.

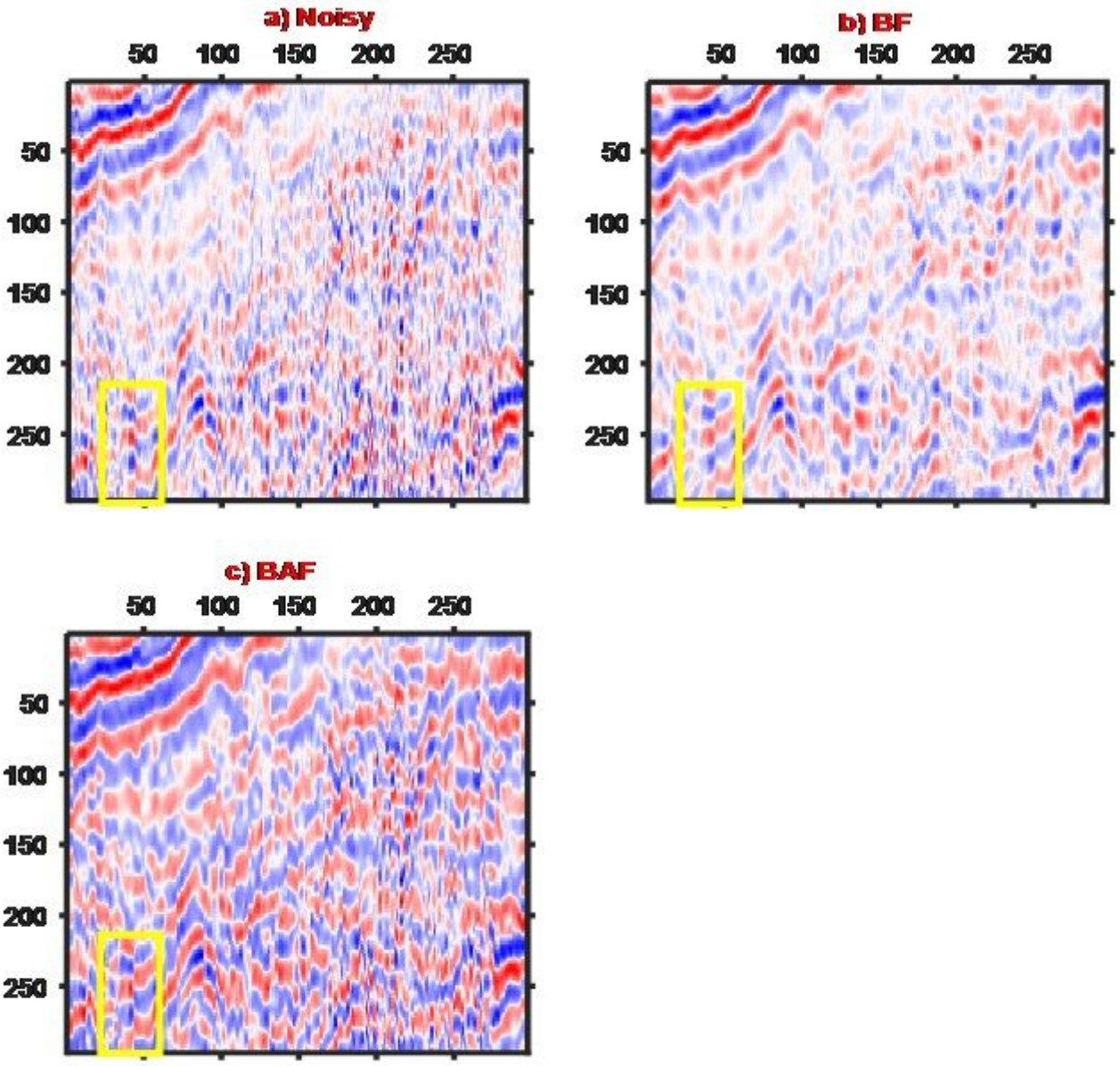

Figure 6

(a) part of a real GPR data [12] and denoised versions by (b) BF and (c) BAF. manner for all regions alongside preserving data details.

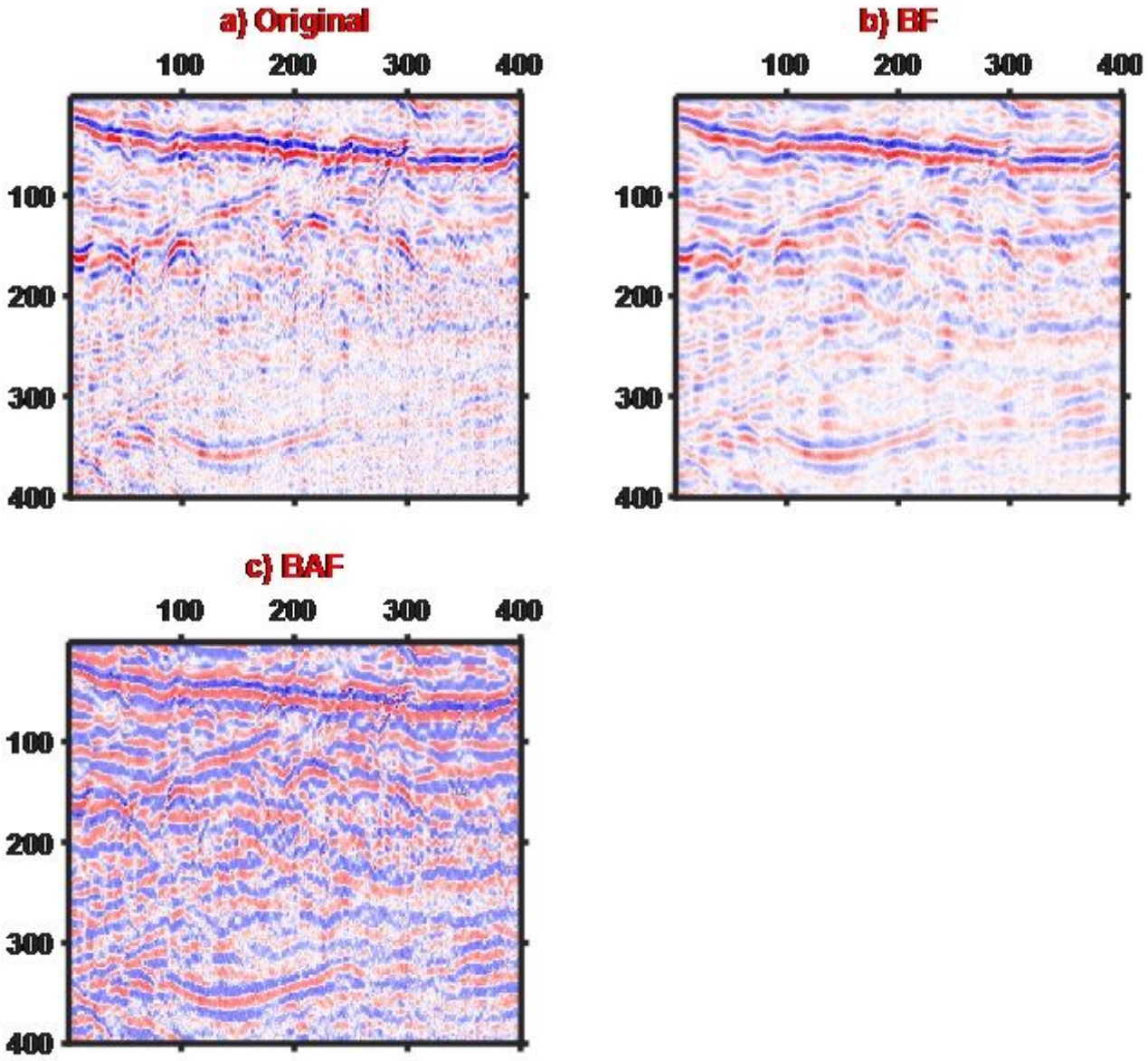

Figure 7

(a) part of a real GPR data [12], denoised data by (b) BF and (c) BAF. Relatively data has high amount of random noise and despite BF, the output of the proposed method is brighter and could be easily interpreted.