Derivative analysis for layer selection of geophysical borehole logs

Aaron C. Davis a,*, Niels B. Christensen b

a CSIRO ARRC, 26 Dick Perry Ave., Kensington WA, 6151, Australia
b Department of Geoscience, Aarhus University, C F Moellers Allé 4, DK-8000 Aarhus C, Denmark

A R T I C L E   I N F O

Article history:
Received 18 April 2013
Received in revised form 18 June 2013
Accepted 19 June 2013
Available online 29 June 2013

Keywords:
Wavelet
Derivative analysis
Borehole
Blocking
Geophysical layer

A B S T R A C T

Analysis of geophysical borehole data can often be hampered by too much information and noise in the trace leading to subjective interpretation of layer boundaries. Wavelet analysis of borehole data has provided an effective way of mitigating noise and delineating relevant boundaries. We extend wavelet analysis by providing a complete set of code and functions that will objectively block a geophysical trace based on a derivative operator algorithm that searches for inflection points in the bore log. Layer boundaries detected from the operator output are traced back to a zero-width operator so that boundaries are consistently and objectively detected. Layers are then classified based on importance and analysis is completed by selecting either total number of layers, a portion of the total number of layers, selection of minimum layer thickness, or layers detected by a specified minimum operator width. We demonstrate the effectiveness of the layer blocking technique by applying it to a case study for alluvial aquifer detection in the Gascoyne River area of Western Australia.

1. Introduction

When examining information from boreholes, we often combine the use of drillers' lithological with geophysical borehole logs to determine relevant properties of the earth. However, the lithology record from the drilling process may be in error. This can be due to the type of drilling process used, such as a mud rig in a clay-rich environment, whereby discrimination of earth materials from drilling mud may be difficult or impossible, or it may be because drilling samples are only taken every metre to half-metre; and thereby information is lost during the drilling process. By comparison, geophysical logs are taken after the borehole has been drilled (and developed, in the case of water bores) and the tools are lowered to record information continuously. However, noise in the measurement of geophysical properties can be an issue, and the natural identification of layer boundaries may therefore be difficult or problematic. In addition, the actual resolution ability of the borehole device may be limited, so that layers from geophysical logs are smeared out. This can often be the case for natural gamma logging, which takes continuous recording over time intervals while being lowered; but it can also affect apparent conductivity logging. The separation of the transmitting coil from the receiving coil limits the minimum layer resolution in the well-log (e.g., Kaufman and Dashevsky, 2003).

Recent advances on the natural identification of borehole layers and boundaries have been made with the use of wavelet transforms of geophysical well-logs (see for example Cooper and Cowan, 2009; Webb et al., 2008; Choudhury et al., 2007; Cowan and Cooper, 2003). Wavelet analysis transforms the profile data into depth and transform information which, in the case of geophysical logs, shows the power of the log signal as a function of the depth under ground. The efforts of these authors, which rely mostly upon the Morlet and Mexican Hat wavelets, have shown that wavelet analysis is an effective tool for de-noising and blocking geophysical log data. Cowan and Cooper (2003) use Haar and Morlet wavelets to examine magnetic susceptibility data in Western Australia. The Haar wavelets effectively limited high-frequency noise from the downhole logs, and showed clear images of the long-wavelength variations in the subsurface. Continuous wavelet transforms were also applied to density and susceptibility data in drillcore measurement by Webb et al. (2008), revealing statistical patterns in the density data that were previously undiscovered. The work of Cowan and Cooper (2003) was expanded upon several years later, where Cooper and Cowan (2008) use the continuous transform Mexican Hat Wavelet to analyse magnetic susceptibility log data in Australia for banded-iron formations in Hammersley Basin. They compare wavelet analysis to median and mean filtering with favourable results, and show that the Mexican Hat continuous wavelet transform is as effective as, or more effective than, using a mean or median filter for automated layer selection of data.

In this paper, we further develop the concept of Cooper and Cowan (2009) by applying a derivative-type analysis based on a
piecewise-linear approximation to the Mexican Hat wavelet transform. Instead of focussing on the frequency and signal content of the geophysical log, we examine the other primary purpose of wavelets of this shape: derivative analysis. Our wavelet, which acts more as an operator, effectively takes the second derivative of any geophysical bore log. Layer boundaries are objectively detected by looking for the inflection points in the log, and boundaries of layers in the transformed space are then traced back to zero-width derivative operators. We briefly explain how the derivative operator transforms the borehole trace, and how the transformed data is analysed for blocking. We provide tools for examining the blocked trace, and hierarchically classify layers based on layer-importance, proportion of total layers, and layer thickness. We show that the derivative analysis can be used for a variety of post-processing analysis techniques, including forward modelling for other complementary geophysical measurements and improving lithology layers from drilling records. We present a freely available, complete, working and annotated Matlab code package for the geophysical community under the Creative Commons Attributions CC-BY licenise 3.0 (Creative Commons, 2013).

2. Method

In this section, we describe our method of obtaining a modified vector of the geophysical borehole trace in order to perform the derivative analysis. Here, we also describe how we construct the matrix of differential operators that are used to find inflection points in the data that represent layers or boundaries in the geophysical data. The geophysical borehole log is convolved with successively wider and wider differential operators in order to find inflection points in the data. Since convolution in the real domain is equivalent to multiplication in the Fourier domain, we convert all data to Fourier domain data for the calculations, and back to real space for the analysis.

2.1. Algorithm for calculating derivatives

We begin with a borehole trace of geophysical data that is sampled in discrete steps of increasing depth $\Delta d$. The geophysical trace data is $N$ points long, where $N$ is an even number, and we denote an individual sample of the total trace $x$ at depth $i$ as $x_i$ occurring at depth $d_i$. For the derivative analysis, an extended ensemble of the data is created by padding the original data and subtracting its mean:

$xe=x_0, -x'(x'-x_0), x_0, x_0-x', x_0, x_0-x', x_0-x', x_0-x', \ldots x_N, x_N-x', x_N, x_N-x', x_N, x_N-x', x_N-x', \ldots, x_N,$

where $x'$ is the reverse order of the borehole trace data. The new trace $x_e$, which is of length $2N+2$, is then transformed to the Fourier domain using the discrete Fourier transform $F(x_e)$.

Our method of layer analysis is similar to the use of the wavelet blocking technique in geophysical borehole logs (Cooper and Cowan, 2009; Cowan and Cooper, 2003). Instead of the ‘Mexican Hat’ continuous wavelet transform, we use a simple approximation of this that is linearly piecewise-continuous. For our process, we use even-numbered interpolations of the primary wavelet $w_p=[-1/2, 1, -1/2]$ which is, in its simplest interpretation, a double derivative operator. Our first differential operator, $w_1$, is formed as follows: beginning with a minimum operator of eight taps, we interpolate vector $w_p$ with eight evenly spaced samples, and pad the left and right sides of the operator $w_1$ with 0 s so that it is also of length $2N+2$. Each successive vector is constructed in the same way, stepping up by four taps at each new operator, until the final operator vector is $2N+2$ points long, and there are at most two zeros in the differential operator trace. For every operator, we normalise the area under the positive section of the curve by dividing by the total number of points that create the wavelet. As an example, Fig. 1 shows 10 differential operators used in the blocking case study from the Gascoyne River (Section 3).

The width of the positive section of each operator (i.e., the section of the operator window that is greater than 0 in the vector) is $2N/3$ points long, and the total number of operators is $M=\lfloor x\rfloor-1$. We define the width of the operator as the total number of points in the positive section of the operator window multiplied by the depth-step value in the geophysical bore-log. The resulting operator matrix $W$, which is $(2N+2 \times M)$ in size, also gets transformed into the Fourier domain, resulting in $F(W)$.

As stated earlier, each application of the derivative transform operates on the geophysical borehole trace to pick out inflection points in the data. Since the operators act on the data as a filter, we can easily express the calculation of the double derivative as a convolution of data. This is most easily calculated in the Fourier domain; however, transformation of data to the Fourier domain implies that our data is cyclic in nature: something that is clearly not true of the original trace. It is for this reason that we have modified the borehole trace to $x'$, which includes a mean-subtracted, reverse trace of the borehole data coupled to a mean-subtracted forward trace of the data (with buffers of 0 between each sub-trace). The addition of the reversed discrete data ensures that the differential operators will discover boundary inflections at the beginning and end of the original geophysical trace, and that the circular convolution of the application of the differential operators in the Fourier space will not overlap layers near the beginning and ending of the original data set. The differentiation product matrix, which is the result of the application of each differential operator vector in $F(W)$ to the extended borehole trace $F(x')$, is then transformed back to real space and concatenated back to an $(N \times M)$ matrix which contains the derivative information for the geophysical borehole trace. This process is explained in the procedure below, where the symbol $\odot$ represents the Hadamard entry-wise product:

1. From $x$, construct $x_e$ and replicate it into matrix $X_e$ of size $(2N+2 \times M)$.
2. Create the $(2N+2 \times M)$ derivative matrix $W$. 

Fig. 1. Selection of 10 of the 224 differential operators used in the Gascoyne River case study (Section 3). Range is from $w_1$ (light blue) to $w_{224}$ (grey). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)
3. Compute the convolution of extended borehole trace with the derivative matrix: \( X \ast W = F^{-1}(F(X) \cdot F(W)) \).

4. Truncate the matrix back to \( T \), an \((N \times M)\) matrix that contains the derivative transforms.

It is the final transformed and truncated matrix \( T \) that is used for the analysis in boundary and layer detection.

### 2.2. Boundary detection

Once we have obtained the transformed \( T \), we can analyse this to detect boundaries. Starting from the widest differential operator, layer boundaries are defined by changes in sign in the last column of \( T \). The changes in sign along this column indicate inflection points in the original borehole log \( x \). The depths at which cross-overs in column \( M \) of the transformation matrix are detected are recorded, and we then turn our attention to the next smaller differential operator at \( M-1 \). Cross-over points at this operator width are also located and recorded, and this process continues until we have recorded all cross-over points for every differential operator in \( W \). We use the layer boundaries to divide the matrix \( T \) into regions of individual layers, and the contours of zero-crossings are traced onto it. Each individual region represents a layer detected by the differential operator method. Fig. 2 shows a colour image of \( T \) for the Gascoyne River case study (Section 3) with layer boundary contours marked with black lines. Sections in red denote positive deflections of the original trace curve, while blue sections denote negative deflections. This is clear from the left-hand panel of the figure, which shows the original apparent conductivity borehole log that we will be using as an example throughout much of this paper. We can see from this example that if we only looked to differential operator widths of 10 m, there would be six layers detected by the algorithm (five layer boundaries). Layers and boundaries of the original borehole data are then detected by tracing individual contour lines back to the left-hand side of the matrix. These lines indicate locations on the original trace where there is an inflection point on the data. Using the cutoff at operator width of 10 m, we see that the layer boundaries can be traced back to 5.9 m, 15.6 m, 27.4 m, 35.5 m and 43.2 m. For this reason, this method of layer and boundary detection is completely objective and independent of the bias of the geophysicist.

### 2.3. Importance ordering and layer detection

Our method of layer detection allows us to characterise and categorise layers. By taking the average of the transformed data in each region, we can systematically arrange the detection of the layers according to layer importance. The relative importance of a layer is a function of its thickness and the deflection of the borehole trace relative to neighbouring values of the trace, as is determined by the transform algorithm in the last section. Consider, for example, Fig. 2: starting from the RHS of the image there are two layers defined, and the boundary separating those layers is just below 35 m depth. The next layer boundary is not detected until the operator width of 26 m, and it defines a new layer. More and more layers are detected as we move in from the right of the image to the left, and each layer then has a region that defines it in terms of its derivative width and depth. By averaging the values in each region, we are calculating an averaged differential transform value for that particular layer. This can be expressed mathematically as

\[
\overline{X_i} = \frac{\sum_{k \in L_i} |T|}{N_i},
\]

where \( \overline{X_i} \) is the average value of the \( k \) elements of the absolute value of the transformed matrix \( T \) that are contained in the subset \( L_i \) for layer \( i \) defined by the zero-crossing contours, and \( N_i \) is the number of elements in \( L_i \). Each value of \( \overline{X_i} \) is then divided by the maximum mean layer value.

Importance is ranked on an absolute value relative to 0 (least important) and normalised to 1 (most important). Fig. 3 shows a colour-ordering of the layers of \( T \) from the Gascoyne River case study borehole trace. More important layers are coloured in darker blue, while less important layers are coloured in lighter values. The first seven most important layers have been labelled in this figure; and we see, unsurprisingly, that the two most important layers are the ones corresponding with layer boundaries that have extended to greater wavelet widths.

By ordering layers according to importance of deflection or layer thickness, we can choose to block out a number of layers based on different criteria when using the algorithm. For each layer detected, we calculate the mean, median, and variance of the log data. Our program provides a function that selects layer boundaries based on minimum detected layer thickness, absolute number of layers, and a percentage of layers relative to the total number of layers detected. For example, Fig. 4 shows layers detected from three different selection criteria. In total, the derivative analysis picked about 30 separate layers from the apparent conductivity borehole trace. Panel (a) shows the organisation of layers depending on the total importance of each layer.
relative to other layers in the detection algorithm. We see that the importance of other layers drops off after about layer #8, which means that the most important layers are contained within the first 25% of the number detected. The solid grey vertical line shows a selection criterion based on the first three most important boundaries, and the dashed grey line shows the cut-off for the first 25% most important layers. Panel (b) shows the layers in a different way. Here, the thickness of each layer is shown. We see that layer 3 is thicker than layers 1 and 2, but it is deemed to be less important due to the fact that its overall rate of deflection was not as great as these layers. This is repeated for layers 7 and 11. The horizontal grey line in panel (b) shows the detection threshold for layers that are over 0.5 m thick. By selecting these layers, we are stipulating that any layers that are less than 0.5 m thick are noise, and can be ignored. Our final method of layer blocking is by using the one first discussed by Cooper and Cowan (2009). By selecting an operator width $w$, corresponding to some operator number $w_n$, all layers detected along column $n$ in matrix $T$ are detected and blocked. This is the example that we gave earlier in Section 2.2, where we showed that six layers were detected if we chose an operator width of $w = 10$ m.

The results from our four selection criteria are shown overlying the original apparent conductivity trace from a borehole in Gascoyne River, Western Australia. The black curve in Fig. 5 shows the original trace for apparent conductivity down the borehole, and the other curves show differing methods of blocking the trace. The red curve shows the first four most important layers (with layer boundaries at 5, 16 and 43 m), the blue curve shows the first 25% of the total number of layers (additional layers at 5, 7, and 10 m, etc.), the green dashed curve shows layers that are a minimum of 0.5 m thick and the purple dashed curve shows the layers detected if we choose an operator width of 10 m. The dashed green trace clearly has more layers than either of the other three, since our criterion for layer selection was much less stringent; but we also see that the layers reproduced by the third method fall inside the others. As well, the layer boundaries of the more important layers are preserved in the higher-fidelity layer blocking. This is to be expected, since we are simply choosing different criteria for layer display; and the detection of layers has already taken place in the earlier phase of the program (cf. Section 2.2). Finally, it is interesting to note that the purple curve, corresponding to a fixed wavelet width, detects layers that are, on average, ~10 m thick, as expected.

3. Case study: Gascoyne River, Western Australia

We examine a case to illustrate the utility of our boundary detection algorithm for both forward modelling of airborne electromagnetic (AEM) systems from borehole data, and to improve lithological classification of strata in conductivity and gamma geophysical borehole logs. The case study along the Gascoyne River area, outside Carnarvon, Western Australia, is part of the Gascoyne Foodbowl Initiative (Royalties for Regions Program)
administered by the Department of Food and Agriculture, Western Australia (DAFWA). The Department has an interest in determining the full extent of the groundwater resources along the Gascoyne River. High yielding zones within the Quaternary Gascoyne alluvial aquifer along the River are predominantly gravel and sand beds interspersed between clay and silt layers. These aquifers lie above the Cainozoic calcilutite sequences that act as aquitards. High yielding zones, with good quality groundwater, can occur between the surface and about 50 m depth and have a lateral extent of several tens to several hundreds of metres. In order to establish whether the Gascoyne groundwater resource can support a proposed increase in agricultural production in this important food-producing area of Western Australia, CSIRO and DAFWA are investigating the potential use of an airborne electromagnetic (AEM) survey. Our purpose is to provide better exploration planning through a targeted approach by identifying potential high-yield zones of good quality water in advance of drilling using an AEM platform. As part of the AEM system selection process for the proposed survey, we used electromagnetic apparent conductivity logs from boreholes near existing production wells as ground-truth models of the conductivity-depth structure of desirable targets. One of the data sets used in the AEM selection process is the borehole apparent-conductivity log that we have been examining in this paper (Fig. 5). To aid in computation speed and provide a simplified conductivity-depth model, we use the derivative analysis algorithm to select layer boundaries consistent with the borehole log data. In our example, we choose the 25% of the total number of layers threshold to select the most important quarter of the total number of layers. This has been shown by the dashed orange line in Fig. 5, but is repeated again for clarity in Fig. 6 (black line). To determine the ability of an AEM system to resolve the conductivity structure of the ground, we use the blocked borehole trace as the ground truth model for the system. Appropriate noise levels are added to the forward model data of the AEM system, and we re-invert this data.
for a 30-layer smooth earth-model. Fig. 6 shows inversions for three separate AEM systems. The red curve is the resulting jointly inverted conductivity for the low and high moment SkyTEM\textsuperscript{101} forward model (Schamper et al., 2012), orange shows the inversion result for the SkyTEM\textsuperscript{304}, while the blue curve shows the inversion result for the Resolve frequency-domain AEM system. It is beyond the scope of this paper to discuss the results of each AEM system, however, we see that there is a good general correlation between the blocked apparent conductivity and the inverted conductivity profiles for each system in the near-surface. Conductivity values below 50 m are inconsistent, and the AEM systems cannot resolve the thin conductive layer in the near surface (7 m). The main point we wish to illustrate here is that the ground-truth conductivity model was generated by our derivative analysis algorithm; and it is used directly as the truth information for the AEM modelling exercise. Thus comparisons of AEM inversion can be made against the blocked conductivity models which was, in turn, derived from an objective analysis of the logged borehole trace.

We finish the study with an examination of the lithological logs accompanying the information of borehole 10/10. This hole was constructed in 2010 as part of the Northern Gascoyne River Borefield program and is owned by DAFWA. Borehole 10/10 was drilled to a depth of 60 m with a 140 mm diameter air-core bit reverse circulation with full face discharge (e.g., Jönsson, 2013), offering good returns off the cutting. Fig. 7 shows the blocked apparent conductivity and natural gamma (black lines) logs over top of the actual measured geophysical logs (grey lines). Coloured boxes under the traces show the lithological layers recorded by the hydrogeologist during drilling. Inspection of the traces shows remarkably good correlations between recorded lithological layers and the blocked layers detected by the differential blocking algorithm. As an example, we see complete agreement with the sand at 5 m, the clayey sand directly beneath it, as well as the boundaries of the calcrete layer underneath. The beginning of the Cainozoic aquitard layer of calcite calcilutite at about 55 m depth is detected by both geophysical tools as well. However, we also see additional details in the geophysical logs that has not been recorded from the drilling. For example, the layer of calcite from 7 m to 13 m, which is classified as ‘…pale red, orange, white, subrounded, poorly sorted, moderately carbonate cemented silty sand…’, can be clearly differentiated into two separate layers with the apparent conductivity and into about five layers using the gamma tool. This also occurs in the sandstone layer at 20 m depth,
and again in the gravelly sand layer at 39 m depth. We also believe that the interface between the sandstone (at 25.5 m) and the gravelly sand (at 26.5 m depth) is incorrectly recorded and that the interface between these two layers in more likely to be at 27.5 m as shown by the layers detected by both the gamma and the apparent conductivity tools. There is an advantage in using the blocking algorithm in the geophysical detection of layer boundaries in lithology logging: not only does the algorithm detect more layers than the drilling record (if they are present) but also it provides depth estimates for the layer boundaries based on the inflection point of the geophysical traces.

4. Conclusions

Following on the previous work of other authors, we present a package of open-source code that is freely available under the Creative Commons Attributions CC-BY licence 3.0. Our software revisits the use of wavelet transforms to de-noise and block geophysical borehole data. Instead of using the Mexican Hat wavelet, we approximate it by piecewise linear curves that essentially act as a second-derivative operator at increasing widths. The transformed matrix is then examined for areas and points of inflection (which are seen as zero-crossings in the data). The inflection points are traced back to zero-width operators which define the layer boundaries in the geophysical log. Layer regions are then organised into closed regions and indexed according to a hierarchical order based on layer thickness and total deflection. Layer blocking is then determined based on operator input so that the number of blocked layers is determined by three selection criteria: number of layers detected, % portion of total layers detected, or layers that are above a certain minimum thickness. The output data is then a Matlab structure that contains layer thickness, layer depth, and the mean, median and variance of each layer.

In this paper, we demonstrate the usefulness of our geophysical data analyser by applying it to apparent conductivity and natural gamma data of a borehole selected from the Gascoyne Foodbowl Initiative operated by the Department of Food and Agriculture, Western Australia. In a joint project, we use the blocked apparent conductivity trace as a tool to assist in the selection of an electromagnetic airborne system for the Royalties for Regions Project in the Gascoyne River area outside Carnarvon, Western Australia. We show how we used the blocked conductivity data as ground truth information for generating the forward and inverse models of the AEM systems. Part of our selection process for this project is based on an analysis such as this. We have also shown that by combining apparent conductivity and natural gamma logs, we can compare to the lithology records obtained by hydrogeologists during drilling. Our blocked layers match extremely well with the lithology boundaries recorded by the hydrogeologist, however, we detect additional layers in the geophysical logs that were only hinted at in the drilling notes. We assert that our technique can be used in routine improvement of lithology data through geophysical borehole analysis. Derivative analysis of borehole data can objectively and consistently discriminate geophysically meaningful boundaries. The degree to which the trace is blocked out for subsequent analysis is entirely controlled by the operator, so that a good balance of signal to noise can easily be achieved.

Acknowledgements

The authors would like to thank CSIRO Water for a Healthy Country and the Department of Food and Agriculture Western Australia for support in the production of this paper. We also acknowledge the use of the function ilabel.m written by Peter Corke (2011). All software presented here (other than the function ilabel.m) is under the Creative Commons Attribution.

References

Creative Commons, 2013. Creative commons attribution 3.0 unported CC BY 3.0. URL (http://creativecommons.org/licenses/by/3.0/).