

Interpretation attributes derived from airborne electromagnetic inversion models using the continuous wavelet transform

Niels B. Christensen*

Department of Geosciences, University of Aarhus, Denmark

Received September 2017, revision accepted October 2018

ABSTRACT

Planning, contracting, data acquisition and processing plus the inverter's quality assessment and inversion of a regional airborne electromagnetic survey may take some months, while the interpretation of the results is a considerably more complex and comprehensive process. Most often an interpretation necessitates additional data that are time consuming to collect and complicated to integrate into an overall model, for example borehole logs, borehole core samples, water chemistry, surface vegetation, satellite imagery plus all existing geological background knowledge. Interpretation basically has to do with identifying categories and finding boundaries between them so that depths, thicknesses and a whole range of other model attributes can be estimated, qualitatively and/or quantitatively. I present two methods using the continuous wavelet transform of finding attributes intended to assist the interpreter: one finds layer boundaries in the smooth multi-layer models that are most often used in the inversion of large airborne electromagnetic data sets, and the other finds the natural categories of the model parameter. Naturally, being based on the subsurface conductivity distribution, the boundaries and categories suggested are useful only to the extent that they coincide with geological/hydrogeological boundaries and categories – which is for the interpreter to decide.

Key words: Airborne EM, Inversion, Interpretation, Wavelet transform.

INTRODUCTION

The planning and contracting of a regional airborne electromagnetic (AEM) survey may take of the order of 2–3 months. Data acquisition could take another month and data processing an additional couple of weeks. The inverter's assessment of data quality, reliability of the contractors information on system parameters and noise model and experiments with inversion settings would take from a couple of weeks to a month. Computation time for the final inversion is of the order of days. All up, a total of 4–6 months is needed before the inverter can put the results on the interpreter's table. However, interpreting the product delivered by the inverter, that is the distribution of subsurface conductivity (which is almost never

interesting in itself), in terms of geological formations and the most likely hydrogeological models is a much more complex and comprehensive process than collecting and inverting AEM data. As a rule, an interpretation necessitates additional data that can be difficult and time consuming to collect and considerably more complicated to integrate into an overall model, for example borehole logs, borehole core samples, water chemistry, surface vegetation, satellite imagery plus the generally accepted geological background knowledge. A complete picture with the required detail and reliability can only be obtained after considerable effort (Lawrie *et al.* 2015, 2016, 2017).

It is a basic tenet of this author that inversion and interpretation are two inseparable aspects of the same scientific process and that they must be done in a process of dialogue and mutual adjustment. However, they also differ in

*E-mail: nbc@geo.au.dk

important aspects. In the inversion process, a model discretization must be chosen that strikes a proper compromise between being complex enough to represent the details that we are interested in, and at the same time simple enough to be computable within a reasonable time. Once the model type is chosen, there is an explicit functional relationship between the model and the data that must be realized in some sufficiently accurate numerical implementation. In the interpretation process, physically and mathematically well-defined functional relationships are few. Most often the interpreter will have to make do with ad hoc relationships with limited validity in space and time containing intrinsically unknown parameters. A major part of the interpretation process is based on qualitative relationships and principles that, in the interpreters experience, have proven themselves valuable and valid. That is why, eventually, there is no substitute for the well-educated, experienced interpreter in making a final interpretation based on a hugely complex – and often apparently inconsistent – set of data of many kinds, representing a large span of spatial and temporal scales.

Basic processes of the interpretation effort have to do with identifying categories and finding boundaries between them so that depths, thicknesses and a whole range of other model attributes can be estimated, qualitatively and quantitatively. To assist in this process, it is desirable to have tools and procedures that provide suggestions, subsequently to be edited and corrected by the interpreter using interactive visualization and editing software. There are mainly two approaches to developing such tools. One is a statistical/machine learning/artificial intelligence approach that, given certain training information provided by interpreter input, provides predictions/suggestions about the parameters in question (Kotsiantis 2007; Cracknell and Reading 2014; Cordua *et al.* 2016; Guldbrandsen *et al.* 2017). Another approach is a deterministic one that from the inversion models extracts attributes that are more directly useful to the interpreter than the distribution of conductivity, that is attributes that more directly address the interpreters efforts of categorizing and delimiting.

In this paper, I develop some methods to derive attributes about layer boundaries and formation categories using the continuous wavelet transform (CWT). First I will present the CWT method using a traditional application as an example, namely finding layer boundaries in a log. Then I will develop a method of finding layer boundaries in the one-dimensional (1D) multi-layer models that are the most commonly used model discretization in standard inversion of large AEM data sets, and finally I will outline a procedure that will find the ‘natural’ categories of the model parameter: conductivity or

log(conductivity) for a whole survey or a part thereof. These methods must be categorized as deterministic, although conditions defined by the interpreter can – and should – be imposed in an interactive post-processing procedure. They are based on the subsurface conductivity distribution, and consequently their usefulness depends on the extent to which conductivity structure can be taken as a good proxy for geological/hydrogeological structure in a given environment.

THE CONTINUOUS WAVELET TRANSFORM METHOD

To introduce the methodology of the continuous wavelet transform (CWT), the basic principles will be outlined in this section illustrated with an example of using the CWT to find layer boundaries in an electrical log. More detailed presentations of the method can be found in Cowan and Cooper (2003), Cooper and Cowan (2009), and Davis and Christensen (2013).

One of the fundamental uses of log information is to identify boundaries between different formations and thus assist the interpreter in a lithological interpretation. The human brain is very good at recognizing patterns and structures, and when visually inspecting a log (see Fig. 1), it is quite straightforward to pick up a pencil and draw lines at the layer boundaries. However, even for a visual interpretation, it is not always obvious which boundaries are the most important, and it can be quite difficult to discern which boundaries can be relied upon as determined by the signal and which are more dubious because of the inevitable noise. An automatic procedure could be very useful in suggesting a proper interpretation, and if it could assist in discerning between signal and noise, between the more important and the less important boundaries, that would be even better. The CWT is an automatic procedure that does just that.

The CWT is defined as (Mallat 1998)

$$W[f(u, s)] = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \Psi^* \left(\frac{t-u}{s} \right) dt, \quad (1)$$

where s is the scale, u is the position, Ψ is the wavelet used and $*$ indicates complex conjugate.

In all of the following, the positions of boundaries are defined as the inflection points of the function (f), that is where the second derivative of the function is zero: $f'' = 0$. Using the so-called ‘Mexican Hat’ wavelet (also called the ‘Gauss2’ wavelet), it is given as the second derivative of a

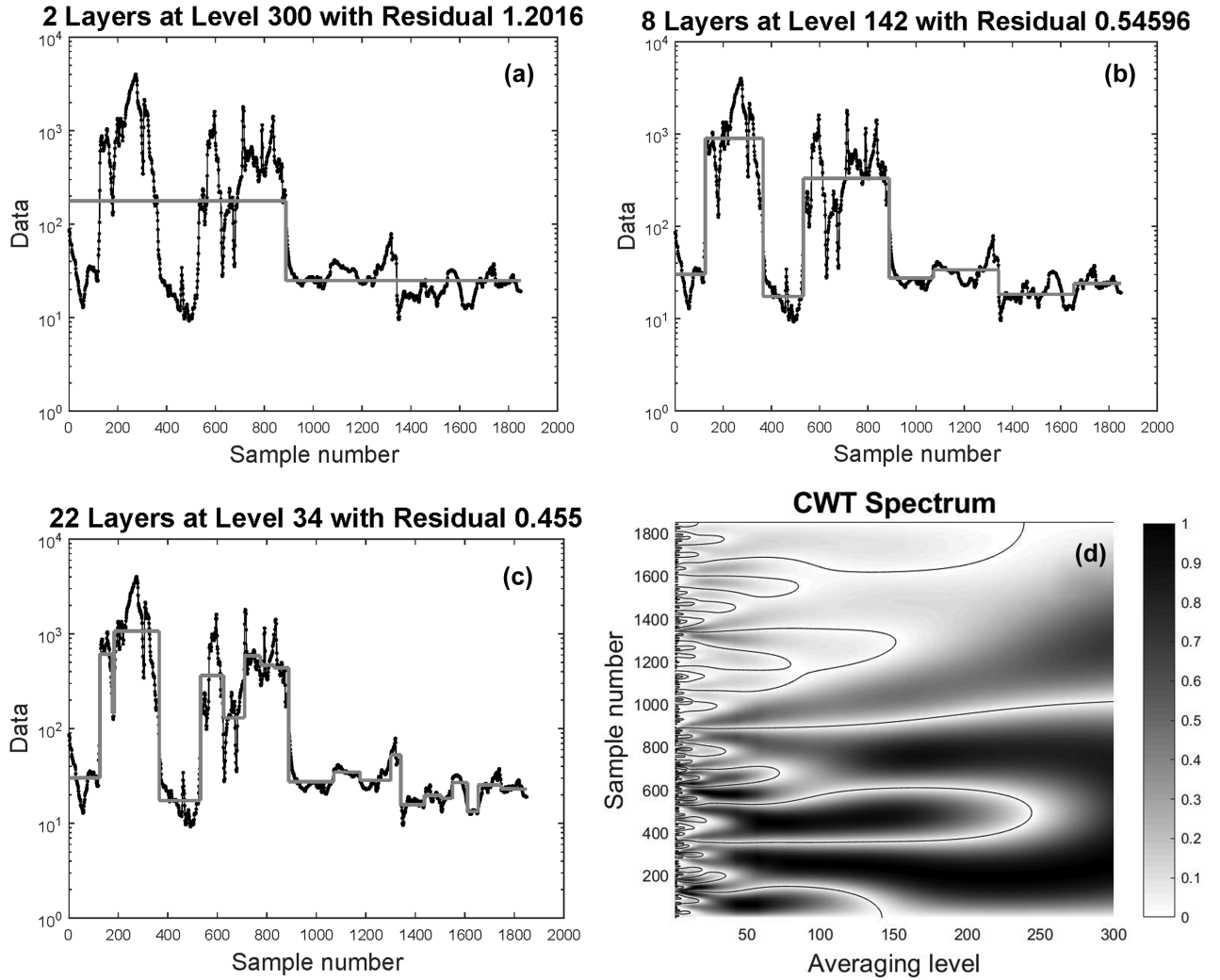


Figure 1 (a) The Ellog resistivity log Beder-F (black curve) with piecewise constant approximations (thick grey curves) obtained using the CWT with two layers, (b) eight layers and (c) 22 layers. The spectrum from the CWT analysis of the log is shown in (d).

Gaussian function:

$$\begin{aligned}\Psi^* &= \frac{d^2}{dt^2} \left\{ \frac{1}{\sigma \sqrt{2\pi}} \cdot \exp \left(-\frac{t^2}{2\sigma^2} \right) \right\} \\ &= \frac{t^2 - \sigma^2}{\sigma^5 \sqrt{2\pi}} \cdot \exp \left(-\frac{t^2}{2\sigma^2} \right),\end{aligned}\quad (2)$$

where σ is the standard deviation of the distribution, now playing the role of scale length, and the CWT will deliver the second derivative of f at different scales. In the presence of noise, the CWT of f at small scales will produce many more zeros than we are interested in. For increasing scale length, the CWT will provide an increasingly smooth version of the second derivative, the number of zeros will decrease and only the main boundaries will

survive. Eventually the user must decide the scale length, or equivalently the number of boundaries, relevant for the interpretation.

In the discrete numerical implementation of the CWT, I use an approximation to Ψ^* using discrete binomial filters. Making use of the fact that convolutions are associative, the sequence of differentiation and averaging can be freely chosen, and the procedure is as follows:

- (1) Convolve with the smallest binomial filter that will produce the second derivative, namely $[1, -2, 1]/4$. This gives the second derivative at the lowest averaging level.
- (2) Each step of the repeated averaging process is implemented by convolving the basic twice differentiated array with binomial averaging filters of increasing width, N , given

by $2^{-N} \binom{N}{n}$, $n = 0, \dots, N$, until all zeros have disappeared, or only the most resilient/important ones remain.

On every averaging level, the values of the second derivative are stored and the zeros found. The two-dimensional array spanned by the sample numbers of the original array and the averaging level constitute the CWT spectrum, also called a scalogram. When the averaging process is completed, the zeros of the CWT spectrum are contoured (see Fig. 1). Now it is up to the interpreter to choose the appropriate averaging level, that is choose the number of zeros and thereby the number of boundaries. This is the only parameter that needs to be provided by the user. Finally, the sample values between the boundaries are averaged to give a mean value of the original array between the boundaries, that is the average layer parameter.

The implementation with discrete binomial filters is attractive because of its simplicity and computational speed. It can be reduced to very simple elemental operations. Leaving out the normalization factor: 2^{-N} , the initial twice differentiating filter can be decomposed as

$$[1, -2, 1] = [-1, 1] * [-1, 1], \quad (3)$$

and is thus reduced to taking the difference between two consecutive elements of the array twice. The sequence of binomial averaging filters can likewise be decomposed:

$$\begin{aligned} N = 2 : [1, 2, 1] &= [1, 1] * [1, 1] \\ N = 4 : [1, 4, 6, 4, 1] &= [1, 2, 1] * [1, 2, 1] \\ &= [1, 1] * [1, 1] * [1, 1] * [1, 1] \\ N = 6 : [1, 6, 15, 20, 15, 6, 1] &= [1, 4, 6, 4, 1] * [1, 2, 1] \\ &= [1, 2, 1] * [1, 2, 1] * [1, 2, 1] \\ &= [1, 1] * [1, 1] * [1, 1] * [1, 1] * [1, 1] * [1, 1], \end{aligned} \quad (4)$$

and so forth, and thereby be reduced to adding two consecutive elements of the array. In this way, the averaging is quite simply implemented as an iterative process consisting of consecutive convolutions with $[1, 1]$.

The relation between the binomial filters and the Gaussian function is that, in the limit $N \rightarrow \infty$, we have (Lindeberg and Bretzner 2003):

$$\begin{aligned} \binom{N}{m} &\rightarrow G[m] = \frac{1}{\sigma \sqrt{2\pi}} \cdot \exp\left(-\frac{m^2}{2\sigma^2}\right), \quad m = n - N/2, \\ \sigma &= \sqrt{N}/2, \end{aligned} \quad (5)$$

where N is the number of coefficients minus 1. We can, thus, define a half-width of the binomial filters as $w_{1/2} = 2\sigma = \sqrt{N}$.

By keeping track of the equivalent cumulative number of binomial coefficients in the iterative process, the number of elemental convolutions in each iteration can be chosen so that the averaging level dimension of the CWT spectrum reflects the half-width in a simple way. In the CWT spectrum in Fig. 1 – and in all subsequent CWT spectra in this paper – the half-width is twice the index number of the averaging level.

When the user has chosen the number of relevant boundaries by choosing the proper averaging level, the position of the boundaries is chosen not as the actual position of the zeros at that averaging level, but by following the zero contours to the zero position at the lowest averaging level. This ensures that the position is found with the best possible precision and that the boundary thus defined does not change when/if more boundaries are added. This establishes a unique hierarchy between the boundaries, which is a prominent and very desirable property of the CWT analysis. The procedure of tracing the zero crossings to the lowest averaging level is equivalent to identifying the convergence of lines of maxima in other applications of CWT analyses (Mallat 1998).

Figure 1 illustrates a traditional application of the CWT in finding boundaries in a log. The electrical log, Beder-F, was measured with the Ellog measure-while-drilling auger tool (Sørensen and Larsen 1999) in eastern Jutland, Denmark. The figure shows an electrical log and the piecewise constant approximations to the log obtained by choosing one, seven and 21 boundaries (two, eight and 22 layers), corresponding to averaging levels number 250, 142 and 34, respectively, together with its CWT spectrum.

In the above example, and throughout this paper, I have used the Gauss2 wavelet. It is a well tried and understood wavelet, and numerically it is quite straightforward to identify the zero crossings to find the boundaries. Other wavelets have been used in the literature. Chandrasekhar and Rao (2012) compare the relative merits of finding maxima/minima of the Gauss1 wavelet with the zero crossings of the Gauss2 wavelet, and Hill and Uvarova (2018) use the zero crossings of the Gauss3 wavelet to estimate whether the boundaries found with the Gauss2 wavelet are sharp or gradational.

USING THE CONTINUOUS WAVELET TRANSFORM TO FIND LAYER BOUNDARIES IN A MULTI-LAYER MODEL

Once an airborne electromagnetic (AEM) survey has been conducted, a successful inversion process will provide the interpreter with proper inversion models that will indicate the distribution of subsurface conductivity.

Only very rarely is the conductivity interesting in itself; however, it carries information that the interpreter can use to make a geological and hydrogeological interpretation, thus making sense of the distribution of subsurface conductivity in terms of a quantitative delineation of depths to layer boundaries and a qualitative identification of different lithologies. There are mainly two different kinds of models used in AEM inversions: multi-layer models (MLMs) and few-layer models (FLMs). Their respective assets and drawbacks can be summarized as follows.

Multi-layer model inversion models

In the inversion, the layer boundaries of multi-layer models (MLMs) are fixed and only the layer conductivities are inversion parameters.

- MLM construction criteria: To avoid bias in the resulting inversion models from the discretization of the MLM, it must fulfil certain criteria: The top layer must be thin enough, the depth to the lowermost layer boundary deep enough and the number of model layers high enough that no individual layer is well determined/resolved in the inversion.
- MLM asset: With a simple initial model of a homogeneous half-space and with proper regularization control of the iterative inversion, convergence is stable and robust and a final model fitting the data (under the given constraints) is ensured.
- MLM asset: With a regularization of the correct strength, one that allows all model details supported by the data to appear without permitting the artefacts of fitting the data noise, MLMs will permit subtle conductivity changes to be seen in the model sections revealing features that are not necessarily well resolved, but nevertheless clearly indicated.
- MLM drawback: Because of the regularization, conductivity changes appear gradual and formation boundaries are not immediately evident.
- FLM drawback: FLM inversion often needs good initial models to ensure stable convergence, especially in the inversion of fairly sparse data sets, for example frequency domain data.
- FLM drawback: The number of model layers must be decided before inverting the data. However, the optimal number of layers is not always evident, meaning that often a trial and error procedure must be undertaken before the correct number is found. The choice of the number of layer boundaries is equivalent to a fairly strong regularization, the exact nature of which is difficult to quantify.
- FLM drawback: Layer boundaries are not necessarily formation boundaries, especially in present-day laterally correlated inversion, where the most prevalent practice is to use the same number of layers for all models along a profile no matter what the optimal number for the individual data set might be. This will sometimes have the consequence that a formation 'changes layer number' along the profile, in which case layer boundaries will obviously no longer be formation boundaries.
- FLM inversion model acceptance criteria: If no specific geological indications are available, an FLM must have the smallest possible number of layers that still makes it possible for the model response to fit the data.

Most often MLMs will be used for inversion of large data sets, often as the final inversion product, but at least as the first of a series of inversion products. If FLM inversion is required, very good initial models for FLM inversion that ensures quick and stable convergence can be derived from the MLMs.

Finding layer boundaries in multi-layer models

Finding layer boundaries in inversion models to compensate for the blurry definition of formation boundaries in multi-layer models (MLMs) has been a research effort in several instances, for example, reported by Chambers *et al.* (2010) and Hsu *et al.* (2010). In this section, I will develop a procedure that finds layer boundaries in MLMs using the continuous wavelet transform (CWT). The procedure consists of two steps: first the zero points of the second derivative of the MLM model are found and the few-layer model (FLM) boundaries are placed at these zero points. Optionally, if the user wishes to reduce the number of initial FLM layers and/or establish a hierarchy between the boundaries, a CWT analysis can be carried out by averaging a densely sampled version of the second derivative found in the first procedure.

Figure 2 shows a plot of a MLM with 30 layers from an airborne electromagnetic (AEM) survey. The inversion was

Few-layer model inversion models

In inversions with few-layer models (FLMs), both layer boundaries and layer conductivities are inversion parameters to be adjusted in the inversion.

- FLM asset: If sharp vertical conductivity contrasts are found in the subsurface conductivity structure, FLM inversion will deliver a better definition of the layer boundary than multi-layer model (MLM) inversion where such layer boundaries are blurred due to the regularization.

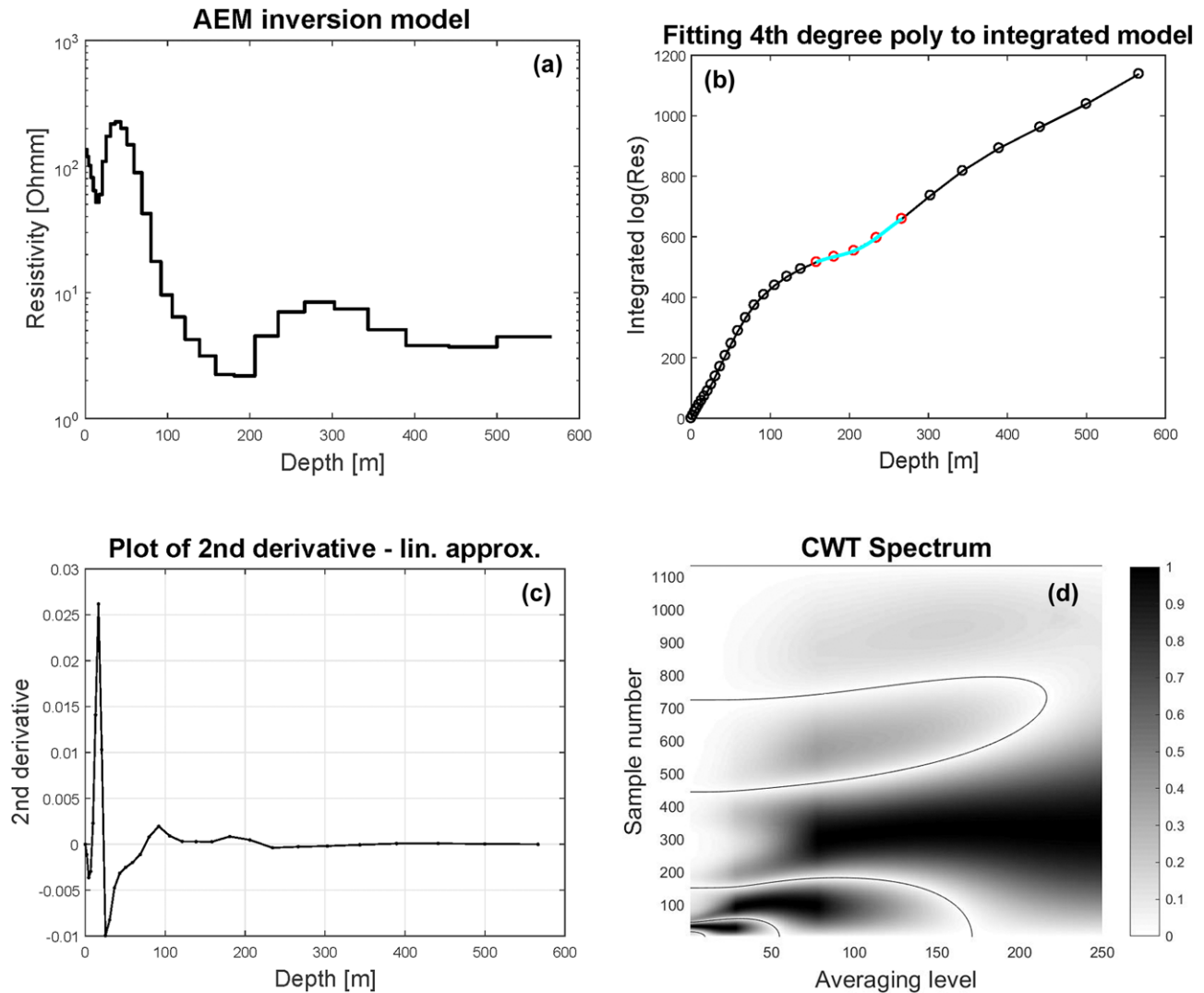


Figure 2 (a) Multi-layer model with 30 layers. (b) The integrated model with a fourth-degree polynomial (cyan curve) fitting the five points indicated with red circles. (c) The piecewise linear approximation of the second derivative of the model. (d) CWT spectrum of the model log (conductivity).

carried out using the methodologies outlined in Christensen (2016a, 2016b). Again, as was the case with the log example, a visual inspection would enable the interpreter to define the major layer boundaries, but with large surveys with millions of inversion models a manual procedure is out of the question and there is a need for an automatic algorithm.

Keeping the definition that a boundary is situated where the model has an inflection point, the question is how to find the second derivative of the model. With the piecewise constant MLMs, the second derivative is everywhere zero except at the MLM boundaries where the first derivative is a Dirac delta function and the second derivative is the derivative of a

delta function – which in a generalized sense is also zero at the boundaries. To get around this situation, what is needed is a smooth version of the model, but then the problem arises as to the position of the samples that define the model; the layer conductivity is the same all through a layer. To get around these difficulties, instead of working with the model parameter itself – conductivity or log(conductivity) – I consider the integrated model parameter also shown in Fig. 2 and seek a way to find its third derivative. For the integrated parameter, the values are well defined at the MLM layer boundary positions and by fitting the unique fourth-degree polynomial that fits five consecutive points and taking its third derivative

for the central of the five points, the second derivative of the model is defined at that central point. The third derivative of the fourth-degree polynomial is a non-degenerate linear function, but I will use its value only for the central point.

By first using points 1–5 to find the second derivative of the model at point 3, I proceed with using points 2–6 to find the derivative at point 4, and then points 3–7, and so on all through the integrated model parameter array to find the second derivative of the model at all the MLM layer boundary positions (except the end points, but that is not a great loss). The procedure is illustrated in Fig. 2. Linear interpolation between these values to find the zeros will explicitly produce all layer boundaries implicitly present in the MLM. If this result is acceptable – possibly after some post-processing (see below) – a CWT analysis is not necessary.

If a hierarchy between the FLM boundaries is desired by the interpreter, a dense equidistant sampling of the piecewise linear function that passes through the values thus defined will then provide the input array to the CWT.

Figure 2 also shows the piecewise linear approximation to the second derivative of the MLM and the pertinent CWT spectrum. Compared with the CWT spectrum of the Ellog, it is seen that because the input array is quite simple and without short-wavelength noise, the number of zeros is small and the CWT spectrum is simple. Figure 3 shows the CWT models resulting from choosing four, five and six layers, and it seems fair to choose a model with six layers. The root-mean-square (RMS) errors of the four, five and six layer models are 0.504, 0.477, and 0.469, respectively. The reason why the RMS errors are quite similar is of course that the CWT algorithm ascribes the mean values in the depth intervals resulting from the CWT analysis of the respective model, and that the lower and well-conducting parts of the models are the same.

In some interpretation situations, the most interesting layer boundary is not necessarily the one that is highest in the hierarchy defined by the CWT analysis. In this case, the best procedure might be to choose all the zeros of the second derivative found without applying the CWT analysis and supplement with a series of post-processing steps demanding that certain criteria must be fulfilled by the FLMs. I have used two criteria: (1) all layers must be thicker than a certain user-defined limit and (2) no two neighbouring layers can have a relative parameter contrast smaller than a user-defined limit. When using criterion (1), the thin layer is merged with the neighbouring layer for which the parameter difference is the smallest. When using criterion (2), the boundary is just deleted from the array of boundaries and a new mean value of the model parameter for the joined layer is calculated. Other

criteria can be defined to highlight a certain boundary demanding, for example that it lies in a certain depth interval or that the parameter contrast between the lower and upper layers of a boundary lies in a certain interval. These post-processing steps are interpretation oriented and can preferably be implemented in an interactive program where boundaries can be displayed and edited.

Figure 3a shows the MLM model section from the inversion, and Fig. 3d shows the FLM model section found by using the above procedure of retaining all the layer boundaries without a CWT analysis. It is seen that it is identical to the one obtained using the CWT with the same number of layers – as it should be. The two top plot frames of Fig. 4 show the MLM model section and the FLM model section produced using the layer boundary routine, respectively, and Fig. 5 shows that the FLMs derived in this manner fit the data at least as well as the MLMs.

USING A CONTINUOUS WAVELET TRANSFORM ANALYSIS TO IDENTIFY PARAMETER CATEGORIES

The procedures of the previous section are focused on finding boundaries and works on one multi-layer model (MLM) at a time. The number of layers of the derived few-layer models (FLMs) varies along the profile, but this must be seen as an advantage: the procedure adapts to the complexity of the models. However, in this approach there is no focus on the lateral correlation of the model parameters or on what is actually delineated with the boundaries. It would be desirable to have a method of identifying the basic formations as defined by the parameter values by taking a whole survey, or a part thereof, into account. In this section, I develop a method based on a continuous wavelet transform (CWT) analysis of finding the ‘natural’ bins for the parameter values of a survey.

In the previous CWT analysis examples, focus has been on the boundary positions along the index dimension of the input array and the resulting average values of the layers defined by the boundaries followed as a consequence of their position. In this section, the main interest is on the value dimension of the input array and the aim is to develop a routine that will indicate the ‘natural’ bins of values of the input array. The suggested procedure is as follows:

- (1) collect all layer parameters, that is conductivities or log(conductivities), from all models of (a subarea of) a survey into one array,
- (2) sort the array,
- (3) use the sorted array as input array to the CWT analysis.

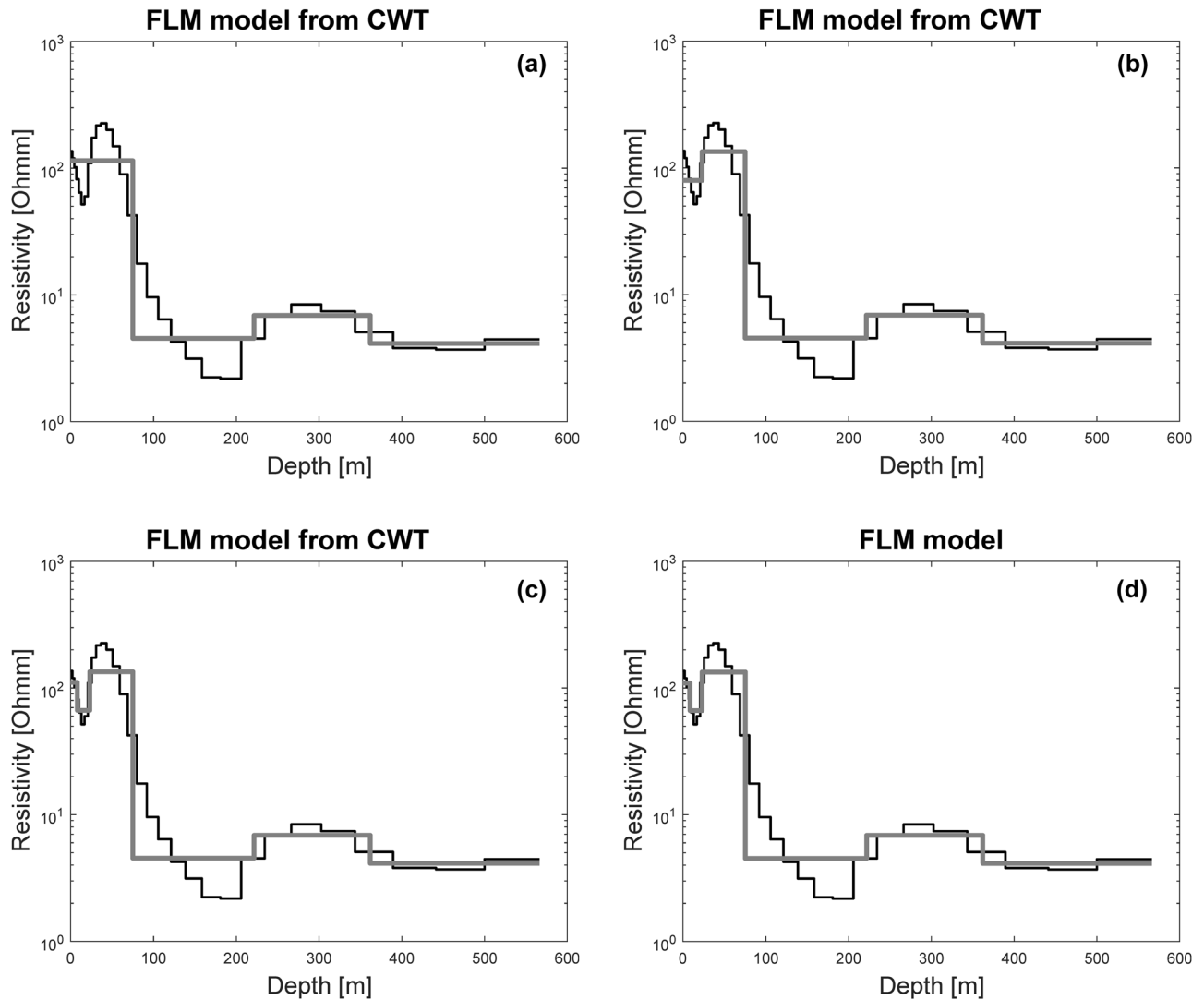


Figure 3 (a) A few-layer approximation with four layers. (b) A few-layer approximation with five layers. (c) A few-layer approximation with six layers. (d) Using all layer boundaries without CWT analysis. In all plot frames, the black curves are the multi-layer models and the grey curves are the few-layer models.

This procedure is motivated by the observation that if all model parameters are collected in a one-dimensional array and sorted, a ‘plateau’ in the sorted array will indicate that a large number of model parameters are of approximately the same size, and the plateau will therefore reflect a natural category of the model parameter. Neighbouring plateaus will be separated by inflection points, and the boundary between bins can thus be found in a CWT analysis. As pointed out several times through this paper, it must be emphasized that the binning of the model parameters of a survey (sub)area makes geological/hydrogeological sense only to the extent that the categories

of interest to the interpreter are characterized by differences in conductivity.

The boundaries found in the CWT analysis are now used to define the end points of the ‘natural’ intervals of the sorted array values, that is the values that ‘naturally’ belong together. As seen in the previous sections, the CWT analysis defines a hierarchy between the number of boundaries and thereby the number of bins. Furthermore, due to the fact that boundaries on a higher averaging level do not change when more boundaries are added, it has the very valuable property that bin interval end points do not change when more bins are

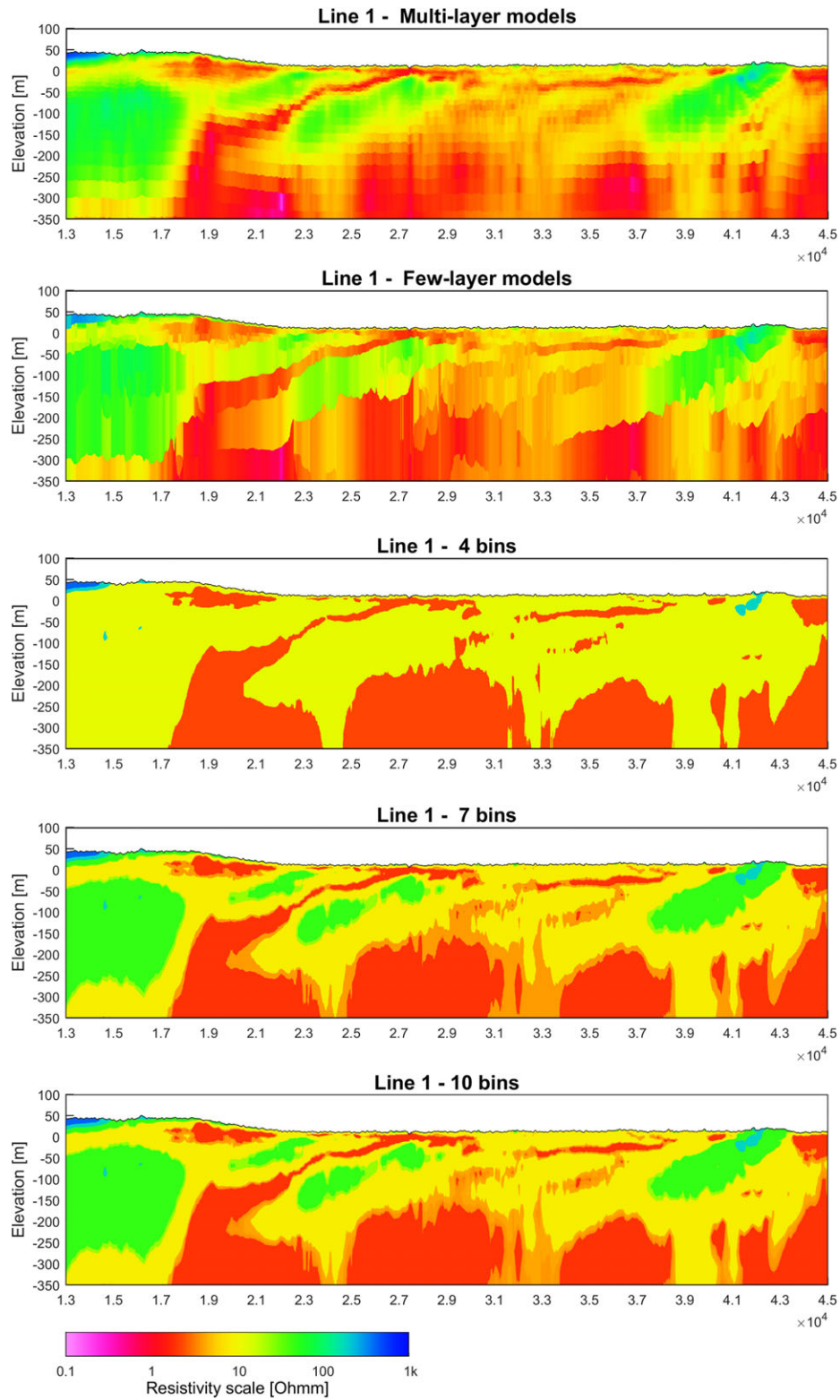


Figure 4 Model sections of an AEM line from an Australian Survey. From the top: multi-layer models; few-layer models produced with the layer boundary routine; model sections showing four, seven and 10 bins produced with the CWT binning routine.

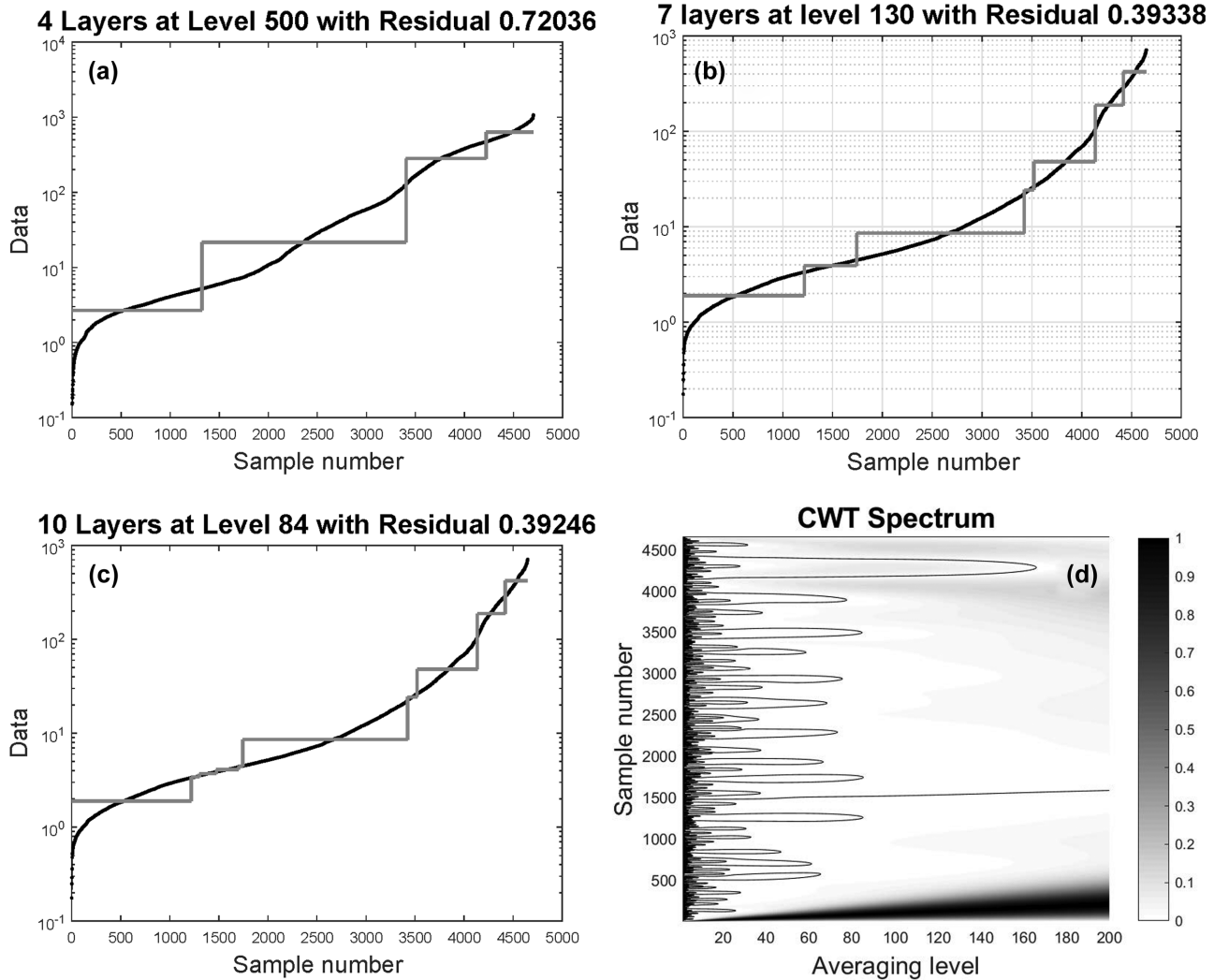


Figure 5 The sorted array (black curve) of the AEM line shown in Fig. 4 with the piecewise constant approximations illustrating the binning with (a) four bins, (b) seven bins and (c) 10 bins. The CWT spectrum is shown in (d).

considered. This means that when more complex structures are permitted by introducing more bins, the complexity is always expressed as a subdivision of already existing bins. In this respect, it differs from traditional techniques of clustering (Jain, 2010; Wu, 2012).

If a survey area consists of several different geological provinces, it is obviously a good idea to subdivide the area to ensure that the CWT analysis does not lump parameters of very different origin together. In this way, the bins are more likely to be indicative of the specific geological setting.

Inversion of large airborne electromagnetic (AEM) surveys will often produce millions of inversion models with tens of millions of parameter values. Computation time for the CWT analysis is approximately N^2 , where N is the number

of elements in the input array. This means that though the sorting of millions of values is not a problem these days, the CWT analysis would become too time consuming. I have, therefore, developed a staged approach to the CWT analysis for large arrays:

- (1) If there are less than 10,000 elements in the input array, use the array as it is as input to the CWT analysis.
- (2) If there are more than 10,000 elements in the input array, choose the first 5000 elements and perform the iterative CWT analysis on those. Only average until at least 100 intervals remain after the averaging, select the interval end points of these and store the values. Continue this way until the whole original array is exhausted.

(3) Collecting the stored values from the previous round of CWT analyses, sort the array and subject it to the final CWT analysis. If there are more than 10,000 elements of this array, repeat the selection process as in the previous step.

(4) When the staged selection process results in an array of less than 10,000 values, perform the final CWT analysis.

The averaging procedure of an array with 5000 elements to produce 100 samples is very fast, of the order of 0.1–0.2 s, so it is not a problem to perform thousands of those. It is the final analysis that takes up most of the computation time, but for an array of 5–10,000 elements, the computation time is less than 1 min. Thus, it is not a computational problem to handle hundreds of millions of parameter values in the analysis. Other approaches could be used to select the final array, but with the procedure outlined above, the same CWT analysis is used to select the elements in the staged process as the one used in the final analysis. This reduces the chance of introducing an unwanted bias in the selection and the final analysis. This is an application of the superposition principle, which is generally valid for CWT analyses provided that the scales involved are similar and that the data lengths of the segments are similar.

As before, the only user-selected parameter is the number of bins to be used, but once the CWT averaging has been continued until only a small number of bins remain, the larger numbers follow naturally and without further computation time from the inherent hierarchy of the CWT analysis. Thus, the analysis only needs to be performed once to make various numbers of bins available to the user for a subsequent selection according to which one best suits the purpose.

A FIELD EXAMPLE

Figure 4 shows a part of the original multi-layer model (MLM) inversion model section of a line from an Australian survey together with the model sections resulting from applying the layer boundary and the binning procedures. Data were recorded with the SkyTEM312 system and the inversion was implemented with a hybrid approach using both the approximate inversion (Christensen 2016a) and full accuracy calculations. Laterally correlated models were produced using a strictly horizontal lateral parameter correlation procedure (Christensen 2016b). The MLM profile consists of 3696 soundings inverted with 30-layer models, which gives in total 110,880 parameter values. The continuous wavelet transform (CWT) was performed on $\log(\text{conductivity})$ and because the number of parameters exceeds 10,000, 22 initial runs

with 5040 samples in each were conducted, each extracting 227 samples. Typically the CWT only had to run to an averaging level of 8–9 and the computation time for the 22 runs was approximately 3 s. The final CWT analysis took 19 s.

In Fig. 4, the top plot frame shows the MLM inversion model section, the next frame shows the model section resulting from using the layer boundary routine of the previous section, including all boundaries, and the bottom three frames show the model sections resulting from selecting four, seven and 10 bins. As can be seen, the simplifications obtained are close to what one would expect from a manual categorization of the model section, and it is quite obvious that increasing the number of bins results in a subdivision of already existing bins. Looking at the results, it seems quite obvious that the few-layer model (FLM) section produced with the layer boundary routine reflects all of the characteristics of the MLM section. Looking at the binned sections, it is seen that four bins are not quite enough to represent the original MLM inversion section; seven bins clearly do a better job than four bins, while the additional complexities of the 10 bin model section only serve to introduce transitional layers between already existing bins. In the next section, it is shown that the seven-bin models are the simplest ones that will actually fit the data and thereby account for the same amount of information as the MLM model section, making seven bins a good choice if no other issues must be taken into consideration. It must be emphasized that the binning procedure is an algorithmic activity resulting in a suite of attributes; choosing which of these are the relevant ones is part of a more complex interpretation situation. Figure 5 shows the final sorted array with the binning intervals indicated as a piecewise constant function approximating the sorted array for the above mentioned cases of four, seven and 10 bins together with the CWT spectrum of the sorted array.

In a similar way to the layer boundary routine, an interpreter-oriented post-processing of the CWT results can be implemented to enhance the characteristics that the interpreter is most interested in. In Fig. 5, it is seen that when choosing 10 bins in the CWT analysis, four fairly small bins with average parameters very close to each other around a value of 40 Ωm are produced. This could be perceived as an unwanted complication within a small interval, and a post-processing procedure can therefore be applied that permits the interpreter to remove bin boundaries separating bins with average parameter contrast smaller than a user-defined limit. The seven-bin case shown in Figs. 4 and 5 has actually been created from the 10-bin situation by requiring that the relative

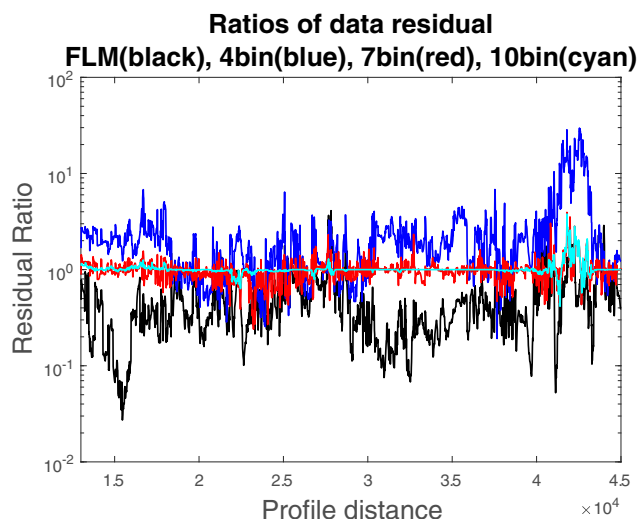


Figure 6 The data residuals of the few-layer models normalized with the data residuals of the multi-layer models using the layer boundary routine (black) and the binning routine, retaining four (blue), seven (red) and 10 (cyan) bins.

difference between neighbouring bin average values is larger than 0.2.

The choice of the limiting value must be seen as part of the interpretation situation and must be made based on the conductivities of the lithologies present in the survey. If it is known that the situation is simple with large differences between the conductivities of the lithologies, a larger value might be preferable. Generally, in hydrogeophysical surveys, small differences in conductivity can be important and often the conductivity ranges of various geological units overlap considerably. Therefore, in the above example, I have chosen the fairly small number of 0.2 as the limiting value.

Data fits of the simplified models

A relevant question is to which extent the models derived with the layer boundary routine and the binning routine fit the data. To answer that question, I have calculated the one-dimensional (1D) forward responses of the few-layer models (FLMs) arising from the layer boundary routine, including all boundaries, and the binning routine for the models shown in Fig. 4. Figure 6 shows the ratio between the data residuals of the FLMs and the data residual of the MLM. For the FLM model section derived using the layer boundary routine, the mean of the data residual ratio is 0.65. For the binned sections with four, seven and 10 bins, the mean of the data residual ratios is 2.24, 0.99 and 1.02, respectively. Not surprisingly, the

residual ratio decreases when the number of bins increases, and for seven- and 10-bin cases, the data residuals are essentially the same as for the MLM, meaning that they fit the data just as well. The FLMs derived using the layer boundary routine actually fit the data better than the MLMs on average, which of course is due to the vertical smoothness constraint of the MLMs.

DISCUSSION

The main focus of this paper is to present the method of finding boundaries in multi-layer model (MLM) inversion models with the option of qualifying the relative importance of the boundaries, and the method of finding the ‘natural’ bins of the parameter values. The aim of both methods is to provide interpreters with good suggestions that might be useful in their interpretation efforts of quantifying depths to boundaries and identifying various formations. Both methods offer a simplification of the MLM model structures and outline the boundaries between different formations, but they are also different in several aspects.

The first method is boundary oriented and finds boundaries in one-dimensional (1D) conductivity structures. Parameter values for the resulting layers are subsequently found as the mean between the boundaries. The analysis of the previous section shows that the few-layer models (FLMs) constructed this way fit the data at least as well as the MLMs from which they were derived. An analysis of how well the FLM layer parameters, that is layer thicknesses and depths to layer boundaries, are determined could be performed in the traditional way with a minimal computational effort. The FLMs would also be excellent initial models – and could also be used as prior constraints – in a subsequent FLM inversion to adjust the model parameters to improve the fit to the data – if that is deemed necessary at all. Because of the high-quality initial models, the FLM inversion would be very fast.

The second method is parameter oriented and finds the interval endpoints and subsequently the mean value within these intervals of the parameter values of a (subarea of a) whole survey. The mean of the parameter values within a bin is then ascribed to the original model parameters lying within the bin limits. By plotting the resulting model sections, boundaries between formations appear and structures become clear. As expected, the 1D FLMs derived this way fit the data increasingly well for an increasing number of bins, and the field example suggests that even a small number of bins is adequate to make the FLMs fit the data as well as the original MLMs. As was the case for the first method, the local 1D

FLMs thus defined could be analyzed to see how well their model parameters are determined or serve as initial model for a 1D FLM inversion. However, the main purpose is to automatically define the formations of a model section or a whole survey and the simplified models may serve as the starting point for geological/hydrogeological model building.

An advantage of the first method is that layer boundaries are found for each individual inversion model and that the number of layers can change along a model section to properly reflect the varying model complexity. A weakness of the approach is that it looks at only one model at a time, and the lateral continuity observed in the continuous wavelet transform (CWT)-derived FLM model sections is only present because of the lateral continuity of the original MLM model section ensured by the lateral constraints of the inversion.

For the parameter binning method, the advantage is that it considers a whole survey (sub)area and that the CWT analysis provides a natural hierarchy between the number of bins. A consequence of this hierarchy is that when more bins are introduced the previous bin boundaries remain and the effect consists of subdivisions of already existing bins.

Computation time is not a problem. With the layer boundary method, thousands of models can be processed per second. With the parameter binning, hundreds of millions of parameter values can be processed within minutes. If a hydrogeological interpreter intends to use the electrical conductivity values to estimate the hydraulic conductivity by using the simple assumption that the logarithm of the hydraulic conductivity is a linear function of the logarithm of the electrical conductivity: $\log \sigma_{hy} = a \cdot \log \sigma_{el}$, using $\log \sigma_{el}$ as the parameter of the binning process is a perfect proxy for binning $\log \sigma_{hy}$.

Although it is generally acknowledged that the presence of the inevitable uncertainty in geophysical data and the resulting model ambiguity should be taken into account in a subsequent interpretation, it is still common practice to deliver only one final model (at a time) to the interpreter. The full potential of the attributes developed in this paper can only be realized if the interpreter has some interactive display/editing software at his/her disposal in which qualified questions can be asked of the inversion product and the result be presented rapidly. Ideally, the inversion MLM, the FLM model sections created by the layer boundary method and the formation binning should all be available to the interpreter as input for the geological/hydrogeological model building, and it should be possible for the interpreter to select the number of layers and/or the number of bins in an interactive way. A further perspective and a more powerful option would be

to have the attributes included in a machine learning/AI engine embedded in the interpretation software. Therefore, it is this author's belief that the focused attributes developed here would improve the predictive power of a machine learning approach and/or reduce the uncertainty of the predictions – but to verify this is the subject of another paper.

CONCLUSIONS

This paper presents the continuous wavelet transform (CWT) analysis method as a tool for finding attributes in the volume of inversion models from, for example, an airborne electromagnetic (AEM) survey that can be of assistance in the geological and hydrogeological interpretation. Two methods have been developed. One focuses on finding the depths to boundaries of the MLM inversion models and ascribes the mean value of the original MLM to the model layers of the resulting few-layer model (FLM). In the other method, all layer parameters of (parts of) a survey are binned into 'natural' intervals as defined by the hierarchical structure of the CWT analysis. In both methods, the CWT analysis offers a hierarchy of boundaries/bins, and presented with these attributes the interpreter must decide, based on his/her experience and understanding of the EM inversion process and the CWT methodology, to which extent the categories delivered can be identified as formation boundaries.

The CWT analysis is implemented as an iterative averaging process using binomial filters and the simplicity of the processing means that the computation cost is low. Furthermore, a staged approach to the binning process has been devised that ensures that hundreds of millions of parameters can be processed in a few minutes.

In this paper, the CWT analysis has been demonstrated on logs and on the inversion models of an AEM survey, but the principles are applicable to the inversion models of any survey from a variety of geophysical methods. In addition, the method offers itself to be integrated into machine learning approaches, providing a first deterministic suggestion of formation boundaries which, together with the user picks, can be used in a machine learning prediction process.

ACKNOWLEDGEMENTS

This work was inspired through collaboration with colleagues in the Groundwater Branch at Geoscience Australia (GA), and is funded by GA as part of a collaboration project between GA and Aarhus University. The results presented here represent some of the findings of a broader inter-disciplinary

approach to developing new methods to automate hydrogeological mapping and interpretation procedures incorporating AEM data.

REFERENCES

- Chambers J.E., Wilkinson P.B., Penn S., Weller A.L., Ogilvy R.D., Kuras O., *et al.* 2010. Bedrock interface detection for sand and gravel mineral reserve assessment using 3D ERT. Near Surface 2010 – 16th European Meeting of Environmental and Engineering Geophysics, September 6–8, 2010, Zurich, Switzerland.
- Chandrasekhar E. and Rao V.E. 2012. Wavelet analysis of geophysical well-log data of Bombay offshore basin, India. *Mathematical Geoscience* **44**, 901–928.
- Christensen N.B. 2016a. Fast approximate 1D modelling and inversion of transient electromagnetic data. *Geophysical Prospecting* **64**, 1620–1631.
- Christensen N.B. 2016b. Strictly horizontal lateral parameter correlation for 1D inverse modelling of large data sets. *Near Surface Geophysics* **14**, 391–399.
- Cooper G.R.J. and Cowan D. 2009. Blocking geophysical borehole log data using the continuous wavelet transform. *Exploration Geophysics* **40**, 233–236.
- Cordua K.S., Hansen T.M., Gulbrandsen M.L., Barnes C. and Mosegaard K. 2016. Mixed-point geostatistical simulation: a combination of two- and multiple-point geostatistics. *Geophysical Research Letters* **43**, 9030–9037.
- Cowan D.R. and Cooper G.R.J. 2003. Wavelet analysis of detailed drillhole magnetic susceptibility data, Brockman Iron Formation, Hamersley Basin, Western Australia. *Exploration Geophysics* **34**, 87–92.
- Cracknell M.J. and Reading A.M. 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences* **63**, 22–33.
- Davis A.C. and Christensen N.B. 2013. Derivative analysis for layer selection of geophysical borehole logs. *Computers & Geosciences* **60**, 34–40.
- Gulbrandsen M.L., Cordua K.S., Bach T. and Hansen T.M. 2017. Smart interpretation—automatic geological interpretations based on supervised statistical models. *Computational Geoscience* **21**, 427–440.
- Hill E.J. and Uvarova Y. 2018. Identifying the nature of lithogeochemical boundaries in drill holes. *Journal of Geochemical Exploration*, **184**, 167–178.
- Hsu H.-L., Yanites B.J., Chen C.-C. and Chen Y.-G. 2010. Bedrock detection using 2D electrical resistivity imaging along the Peikang River, central Taiwan. *Geomorphology* **114**, 406–414.
- Jain A.K. 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* **31**, 651–666.
- Kotsiantis S.B. 2007. Supervised machine learning: a review of classification techniques. *Informatica* **31**, 249–268.
- Lawrie K., Brodie R.S., Magee J., Tan K., Halas L., Mueller N., *et al.* 2016. An inter-disciplinary approach to airborne electromagnetic (AEM) survey design for groundwater exploration using the Australian geoscience data cube and morphotectonics. ASEG-PESA-AIG 2016, 25th Conference and Exhibition, Adelaide, Australia.
- Lawrie K., Christensen N.B., Brodie R.S., Abraham J., Halas L., Tan K., *et al.* 2015. Optimizing airborne electromagnetic (AEM) inversions for hydrogeological investigations using a transdisciplinary approach. ASEG_PESA, 24th International Conference and Exhibition, 2015, Perth Western Australia.
- Lawrie K., Murray T., Brodie R.S., Gibson D., Halas L., Symington N., *et al.* 2017. An inter-disciplinary, multi-physics approach for rapid mapping and hydrogeological characterisation of neogene intra-plate fault systems in depositional landscapes. *Symposium on the Application of Geophysics to Engineering and Environmental Problems* 2017. <http://doi.org/10.4133/SAGEEP.30-050>.
- Lindeberg T. and Bretzner L. 2003. Real-time scale selection in hybrid multi-scale representations. *International Conference on Scale-Space Theories in Computer Vision*, 148–163.
- Mallat S. 1998. *A Wavelet Tour of Signal Processing*. Academic Press.
- Sørensen K.I. and Larsen F. 1999. Ellog auger drilling: 3-in-one method for hydrogeological data collection. *Ground Water Monitoring & Remediation* **19**, 97–101.
- Wu J. 2012. *Advances in K-Means Clustering—A Data Mining Thinking*. Springer.