Seabed Loging Data Curve Fitting Using Cubic Splines

Samsul Ariffin Bin Abdul Karim¹,º and Noorhana Yahya²

¹,² Department of Fundamental and Applied Sciences
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750 Tronoh, Perak Darul Ridzuan, Malaysia.
{¹samsul_ariffin, ²noorhana}@petronas.com.my

Abstract

Deep water and deep target environment for Seabed Loging (SBL) has been a challenge, and the better delineation method is needed. This paper describes the curve fitting methods namely smoothing spline and piecewise cubic Hermite interpolating polynomial (PCHIP) to give better delineation. In this work we simulated five geophysical layers namely air (500m), sea water (2000m), over burden (1000m), hydrocarbon (100m) and under burden (1000m) respectively. Target depth was increased gradually by 500m from 1000m to 3000m at constant frequency of 0.125Hz and 1250A current. Due to the weak electromagnetic signal from deep target of SBL we had chosen curve fitting methods to describe hydrocarbon reservoir. The results show that smoothing spline is quantified as compare to the PCHIP due to the existence of various noises in the SBL data. Several numerical results will be presented.

Keywords: Sea Bed Loging (SBL) data, smoothing spline, PCHIP, interpolation, noise

1 Introduction

Data fitting is an important tool in various science and engineering disciplines. For examples, data fitting is used to remove signal noise received from a source, from chemical experiments etc. Data fitting is also a main research in Computer Aided Geometric Design (CAGD) and Computer Aided Design (CAD) software’s. Piegl and Tiller gives a very good discussion on data fitting by using Non-Uniform B-
spline (NURBS) [20]. Data fitting techniques can be used either for 1 Dimensional (1D) – curves or 2 Dimensional (2D) – surfaces problems. Data fitting can be done in two ways either approximation or interpolation. In approximation method, resultant curves/surfaces will not passes through all the data points and it will approximate the data in the sense of some error measurement for example Least Square Method (LSM), Weighted Least Square (WLS) etc. In the interpolation method, the curve will pass through all the data points. In case of experimental data, it will be expected that the data will consists of few error in measurements resulting in the existence of noise and approximation methods are more preferred than interpolation. Spline function was one of the method being used for data fitting purpose [5, 6, 16-18, 20, 25, 27-33]. Cubic spline method is used to get the good results with second degree of continuity [2, 3, 5-7, 19-21, 25]. Smoothness is also one of the important requirements in scientific visualization, to obtain a better delineation of data. Ordinary spline schemes and smoother are not extensively used to preserves the geometric shape of the data (i.e. positivity, monotonicity or convexity) due to the unwanted oscillations which may completely destroy the original data features. Thus, shape preserving interpolation is used to solve this problem [13, 14]. Preserving the shape of the SBL data is very important since it may give us a better delineation. Therefore, PCHIP technique is used to preserve the shape of SBL data.

This work premise deals with the application of smoothing spline and monotonicity preserving interpolation of seabed logging data. Normally data fitting in Oil and Gas (O&G) areas utilized linear fitting which is subject to the least square errors [26]. But, linear fitting are the easiest method for data fitting. It may be worthy to apply any polynomial or possibly piecewise polynomial with higher degree but with a lower residual (error) and greater smoothness. This is why spline function is a really good option to the existing methods. For better delineation of SBL data, spline function (smoothing and interpolation) is used which provide good results as compared to the liner fitting. This paper is organized by four sections which are described as follows. First Section is about the introduction of curve fitting method used for smoothing the SBL data and the problems facing in this research area. Second section describes the detail of the theory of smoothing spline and PCHIP. Third section describes the smoothing of sea bed logging (SBL) data by using smoothing spline and PCHIP. Numerical results and discussion will be presented in Section Four. Conclusions drawn from this curve fitting techniques are discussed in the final section of this paper.

2. Smoothing Spline and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)

2.1 Smoothing Spline

We begin with the definition of cubic spline function:
Seabed logging data curve fitting using cubic splines

\[ S_i(x) = a_i(x-x_i)^3 + b_i(x-x_i)^2 + c_i(x-x_i) + d_i \]

(1)

Where \(a_i, b_i, c_i,\) and \(d_i\) are the spline coefficients at data point \(i\) and the abscissa intervals \(x_0 < x_1 < x_2 < \ldots < x_{i-1} < x_i < \ldots < x_n\) or \(x \in [x_0, x_n]\). The cubic spline function \(S_i(x)\) satisfies the following conditions:

(C1) \(S_i(x)\) is a cubic function in each subinterval \([x_{i-1}, x_i]\) for \(i = 1, 2, \ldots, n-1\).

(C2) At each knots \(i\), \(S_i(x)\) has second degree continuity (smoothness), \(C^2\) which is important for certain applications.

The derivatives of (1) w.r.t. \(x\) are

\[ S_i'(x) = 3a_i(x-x_i)^2 + 2b_i(x-x_i) + c_i \]

\[ S_i''(x) = 6a_i(x-x_i) + 2b_i \]

\[ S_i'''(x) = 6a_i \]

For more details the reader is encouraged to refer [2, 6, 19].

For smoothing spline, Eq. (1) can be rewrite as

\[ y_i = f(x_i) + \varepsilon_i \]

(2)

where \(\varepsilon_i; i = 0, 1, 2, \ldots, n\) form a sequence of independent distributed random (i.i.d.) variables with variances, \(V(\varepsilon_i) = \sigma_i^2\). Smoothing spline \(S(x)\) can be constructed to minimize the value of

\[ L = p \sum_{i=0}^{n} \left( \frac{y_i-S_i}{\sigma_i} \right)^2 + (1-p) \int_{0}^{1} \left| S'(x) \right|^2 dx \]

(3)

where \(S_i = S(x_i)\). the parameter, \(p \in [0, 1]\) reflects the resultant smoothing spline. For example when \(p = 0\), the smoothing spline corresponds to the least-square straight-line fit to the data while \(p = 1\) the smoothing spline is a natural interpolating cubic spline (or variational cubic spline interpolation) with \(C^2\) continuity. So the choices of smoothing parameter, \(p\), is really an important task and it must be chosen properly. This can be done by doing various data simulation until the user obtains the required results.

2.2 Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)

The cubic interpolant, PCHIP has been proposed by Fristch and Carlson [10] to preserve the monotonicity, positivity and convexity of the data. For certain
applications, it is important for the user to preserve the shape of data. For example, if
data for curve fitting that is monotonic, then the interpolant (either interpolation or
smoothing) must be monotonic too. Certain conditions to the interpolant at each
interval are imposed to preserve the data. PCHIP is briefly described. For more
details, the reader is encouraged to refer \[9-10\]. Let \( a = x_1 < x_2 \ldots < x_i \ldots < x_n = b \)
be a partition of the interval \([a,b]\). Let \( \{(x_i, f_i) : i = 1,2,\ldots,n\} \) be the given data where
\( f_i \leq f_{i+1} \) (monotonic increasing) or \( f_i \geq f_{i+1} \) (monotonic
decreasing). For \( x \in [x_i, x_{i+1}] \), \( i = 1,2,\ldots,n-1 \), \( p(x) \) is a cubic polynomial which can be represented as follows:
\[
p(x) = f_i H_i(x) + f_{i+1} H_{i+1}(x) + d_i H_3(x) + d_{i+1} H_4(x) \tag{4}
\]
Where \( d_j = p'(x_j) \), \( j = i,i+1 \) and \( H_k(x) \) are the usual cubic Hermite basis functions
given by 
\[
H_i(x) = \phi \left( \frac{x_{i+1} - x_i}{h_i} \right), H_{i+1}(x) = \phi \left( \frac{x - x_i}{h_i} \right), H_3(x) = -h_i \psi \left( \frac{x_{i+1} - x_i}{h_i} \right)
\]
and \( H_4(x) = h_i \psi \left( \frac{x - x_i}{h_i} \right) \) and \( h_i = x_{i+1} - x_i \), \( \phi(t) = 3t^2 - 2t^3 \), \( \psi(t) = t^3 - t^2 \).
Note that the cubic Hermite will not preserve the monotonicity of the data \[10, 16\].
Fritsch and Carlson \[10\] has derived the necessary and sufficient conditions for
PCHIP to be monotonic on each interval \( x \in [x_i, x_{i+1}] \), \( i = 1,2,\ldots,n-1 \), they also
proposed an interactive algorithms in order to determined the derivatives values,
\( d_i, d_2, \ldots, d_n \), that will make the piecewise cubic Hermite polynomial in (4) will be
monotonic. But, the main drawback of their method is that, the method is global. To
overcome this weakness, Fritsch and Butland \[9\] has proposed a new approach to
determined the derivative values that will ensure the method is local which is very
important since any changing of data points, for example at point \( x_i \), the
interpolating curves may affect only at two segments of curve in the interval \([x_0, x_i]\)
and \([x_i, x_2]\). The reader is encouraged to refer to \[4-5, 8-14, 16, 22-25\] for more
details on shape preserving interpolation and approximation theory as well as its
applications.

As an example of shape preserving interpolating polynomial (PCHIP) and cubic
spline functions, we consider Akima \[1\] data sets.

<table>
<thead>
<tr>
<th>Table 1. Akima Data sets [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
</tr>
<tr>
<td>( f )</td>
</tr>
</tbody>
</table>
Figure 1 shows the examples when we apply cubic spline interpolating polynomial (dashed) and PCHIP (solid) to the Akima [1] data set. We can see clearly that, cubic spline does not preserve the monotonicity in the interval [9, 11] and [12, 14]. But PCHIP preserve the shape of the data very well even though the degree of smoothness attained is only $C^1$ compared to cubic spline with the degree of smoothness attained is $C^2$.

3 Experimental Setup for SBL Data Collection

3.1 Modelling of Seabed logging by Computer simulated software (CST)

CST software was used to detect deep target hydrocarbon between 1000 m to 3000 m underneath seabed by using finite integration method (FIM). Model area was assigned as 20×20 km to replicate the real seabed environment with various target positions. Environment with and without hydrocarbon were also prepared for comparison purpose later. There were few steps involved in generating the CST simulated model. First step was to set parameters for aluminium antenna. In this case we used length of 270 m, frequency of 0.125Hz and current of 1250A. Second step was to set parameters for the model. Air thickness was set as 500 m, sea water depth of 1000 m, overburden thickness of 1000 m, hydrocarbon thickness of 100 m and under burden thickness of 1000 m. Table III gives the detail of simulation set up. Schematic diagram of proposed sea bed model is given (Figure 2).
Figure 2 (a) Schematic diagram of proposed model and (b) CST simulated model

Table II: Different layers material properties used for proposed sea bed model

<table>
<thead>
<tr>
<th>Material parameters</th>
<th>Air</th>
<th>Sea water</th>
<th>Under burden/Overburden</th>
<th>Hydrocarbon (HC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative permittivity ($\varepsilon_r$)</td>
<td>1.006</td>
<td>81</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Conductivity (S/m)</td>
<td>$1.0e^{-11}$</td>
<td>4</td>
<td>1.5</td>
<td>0.001</td>
</tr>
<tr>
<td>Thermal conductivity (W/mK)</td>
<td>0.024</td>
<td>0.6</td>
<td>2</td>
<td>0.492</td>
</tr>
<tr>
<td>Density (Kg/m$^3$)</td>
<td>1.293</td>
<td>1025</td>
<td>2400</td>
<td>824</td>
</tr>
</tbody>
</table>
Thickness of the overburden was increased as the target depth varies gradually (every 500m) from 1000m to 3000m. Third step was to apply electric boundary conditions with material parameters given (Table II). Fourth step was to run low frequency full wave solver to simulate sea bed model. The final step was post processing to generate the simulated data for results analysis at different target depths. Maxwell's equations for magnetic and electric fields are used as a code in the software to get electric and magnetic field response with and without hydrocarbon reservoir (HC). We collect all data sets at various target depths from 1000 m until 3000 m. But, in this paper, for the numerical testing, we used data sets at 3000 m both with HC and without HC.

### 4 Numerical Results and Discussion

In this section we will discussed the applications of smoothing spline and PCHIP to fit the SBL data. We used two data sets: one at 3000 m with hydrocarbon (HC) and at 3000 m without hydrocarbon (HC). Both data sets have 295 data points. We decided to used two different values of smoothing parameter, \( p \). More comprehensive numerical results will be reported later in Karim [15].
Figure 3. (a) Cubic spline interpolation for 3000m target depth with Hydrocarbon HC (b) PCHIP for 3000m target depth with HC

Figure 4. (a) Smoothing spline with $p = 0.9981$ for 3000 m with HC (b) Smoothing spline with $p = 0.5$ for 3000 m with HC
Seabed logging data curve fitting using cubic splines

Figure 5. (a) Cubic spline interpolation for 3000m target depth without HC  
(b) PCHIP for 3000m target depth without HC

Figure 6. (a) Smoothing spline with $p=0.9981$ for 3000 m without HC  
(b) Smoothing spline with $p=0.1$ for 3000 m without HC

We may increase the degree of smoothness, but for our applications, $C^1$ seem to be good. If we apply rational interpolate, we may increased the computational aspect. Therefore, we will state with the non-rational interpolate. As an error measurement for spline smoothing, we used (1) Sum Square Error (SSE), (2) $R^2$ and (3) Root Mean Square Error (RMSE). Figure 3 (a) and Figure 5 (a), show the SBL data interpolation by using cubic spline meanwhile Figure 3 (b) and Figure 5 (b) show the SBL data interpolation using PCHIP. Both methods seem to works well for both data
sets. But PCHIP will preserve the shape of the data. From Figure 4 (a), Figure 4 (b), Figure 6 (a) and Figure 6 (b), we can see that both smoothing spline when $p = 0.9981$ and $p = 0.10$, looks almost similar. Therefore we need an error measurement in order to justify which one better in smoothing the SBL data (in our case we apply the goodness of fit). Hence, from Table IV, we can see clearly that when $p = 0.9981$ the smoothing splines gives better results as compared with $p = 0.10$. This can be well understand since when $p$ is closed to 0, then the smoothing spline closed to become least-square straight-line fit and when $p$ is closed to 1, then the smoothing splines closed to become natural or variational cubic spline interpolation. This is why the SSE and RMSE when $p = 0.9981$ are smaller as compared to SSE and RMSE when $p = 0.10$. Table III gives statistical results for smoothing spline data fitting. For fitting the SBL data by using PCHIP and cubic spline, the $R^2 = 1$ (this is due to the fact that all data points will be interpolate). For smoothing spline, the smaller the value of SSE and RMSE, then the better the fitting obtained and the fitted data will be useful for further prediction and analysis. This will be our main subject for future research.

The closer $R^2$ to 1, indicated that a greater proportion of variance is accounted by the proposed model. In our case all the method that being considered (PCHIP, Cubic spline and smoothing splines) are capable to fit the seabed logging data (SBL). But, we need to choose suitable smoothing parameter, $p$ when we want to used smoothing spline for data fitting purposed. For further numerical comparison with various smoothing parameter, $p$ please refer to [15].

<table>
<thead>
<tr>
<th>Error Measurements (Goodness of Fit)</th>
<th>3000 m with HC ($p = 0.9981$)</th>
<th>3000 m with HC ($p = 0.5$)</th>
<th>3000 m without HC ($p = 0.9981$)</th>
<th>3000 m without HC ($p = 0.1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum Square Error (SSE)</td>
<td>$4.307 \times 10^{-9}$</td>
<td>$7.447 \times 10^{-1}$</td>
<td>$1.130 \times 10^{-9}$</td>
<td>$1.841 \times 10^{-2}$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Root Mean Square Error (RMSE)</td>
<td>$1.243 \times 10^{-3}$</td>
<td>$8.113 \times 10^{-2}$</td>
<td>$6.369 \times 10^{-4}$</td>
<td>$4.304 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

5 Conclusions and Recommendation

Two different approaches for fitting the seabed logging (SBL) data namely smoothing spline and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) are discussed. Comparison between these two fitting methods was done. The numerical results indicate that smoothing spline shows better delineation for sea bed
logging data (SBL) data fitting with error measurement, SSE is $4.307 \times 10^{-9}$ and RMSE is $1.243 \times 10^{-3}$ and $R^2 = 1$ for $p = 0.9981$ (3000 m with HC) and SSE is $1.130 \times 10^{-9}$ and RMSE is $6.369 \times 10^{-4}$ and $R^2 = 1$ for $p = 0.9981$ (3000 m without HC). Based on the numerical results given in Section 4, we would like suggest that, the smoothing technique is a good choice when we deal thousands of data sets rather than we just interpolate the data by using cubic spline interpolation or any other techniques. This is due to the fact that, the unwanted noise might be exists in the SBL data taken from all experiments. Therefore, we conclude that smoothing splines are a better choice for SBL data pre-processing before the data are used for further analysis. Finally, currently we are studying the use of cubic spline and wavelets to fitting and to recover the weak signal that being detected in SBL applications. These findings will be reported in our forthcoming papers.

Acknowledgement. The first author gratefully acknowledges Universiti Teknologi PETRONAS (UTP) for the financial support received in the form of a research grant: Short Term Internal Research Funding (STIRF) No. 76/10.11 and No. 35/2012 and computing facilities including MATLAB software.

References


Received: March 15, 2013