

Chapter 5

Seismic Prediction of Reservoir Parameters

Problem presented by: Brian Russell (Hampson Russell Software)

Mentors: Len Bos (University of Calgary), Gemai Chen (University of Calgary), Lou Fishman (MDF International, University of Calgary)

Student Participants: Mahmud Akelbek (University of Regina), Mark Braverman (University of Toronto), Sandra Fital (University of Regina), Yaling Yin (University of Saskatchewan), Zhidong Zhang (University of Saskatchewan)

Report prepared by: Yaling Yin (yin@snoopy.usask.ca)

5.1 Introduction

In seismic analysis, our goal is to determine the properties of the subsurface of the earth using seismic measurements made on the surface of the earth. Seismic measurements are made by sending elastic waves into the earth (by setting off an artificial explosion at the earth's surface) and recording the reflected elastic waves at the earth's surface (using geophones on land and hydrophones offshore). This is illustrated in Figure 5.1. Although this technique has been practiced since the beginning of the twentieth century, the major advances in seismic analysis have occurred during the last fifty years, and were spurred by the joint development of digital recording and the digital computer. Most of the techniques that have been developed for the analysis of seismic data can be classified as deterministic. That is, we develop a deterministic physical model which relates the material properties of the subsurface of the earth to the physics of the transmitted and reflected seismic waves.

One such deterministic model is called the convolutional model. In the convolutional model, we assume that the amplitudes of the seismic reflections are directly related to the changes in impedance (the product of velocity and density) of the various geological formations below the surface, convolved with a "wavelet" that represents the oscillations caused by the seismic source. By "de-convolving" and inverting the seismic recording, we can therefore derive both the velocity (which can be either compressional or shear) and density distributions of the earth's subsurface. From these velocity and density values, we can then infer other properties of the subsurface, such as the porosity and fluid content of the seismic reservoir. This inversion is illustrated in Figure 5.2 for the Blackfoot case study.

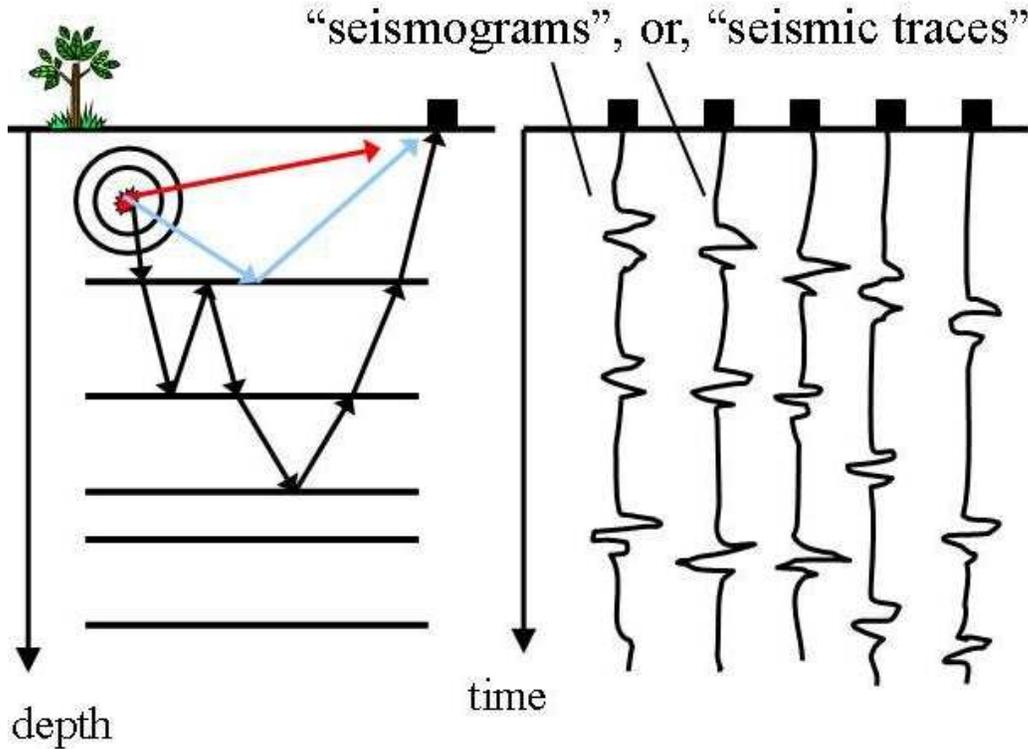


Figure 5.1: Elastic waves are sent into the earth and the reflected signals from the subsurface layers are recorded on the surface as a function of time.

More recently, researchers have started using a statistical, rather than deterministic, approach to the determination of the earth’s subsurface and its reservoir parameters. That is one of the main approaches that we will use here. Thus, instead of assuming an underlying model, we will use multivariate statistical techniques to determine the earth’s subsurface, using a set of derived “attributes” from the seismic data. This mathematical approach is described in Section 5.2. In Section 5.3, the previous statistical approaches to addressing this problem are reviewed and evaluated. Sections 5.4 and 5.5 describe and discuss two approaches that were developed in the IPSW. Section 5.4 considers a statistical approach, while Section 5.5 addresses spline approximation. Finally, Section 5.6 provides a general summary and discussion.

5.2 Mathematical Statement of the Problem

We assume that we have N , M -dimensional multivariate observations which can be written as $x_j = (x_{1j}, x_{2j}, \dots, x_{Mj})^T$, $j = 1, 2, \dots, N$, and that we have N scalar training values t_j , $j = 1, 2, \dots, N$. Our objective is to find some linear or nonlinear scalar function y such that

$$y(x_j) = t_j, \quad j = 1, 2, \dots, N. \quad (5.1)$$

In our case, the multivariate observations are a set of seismic attribute values at a given depth or



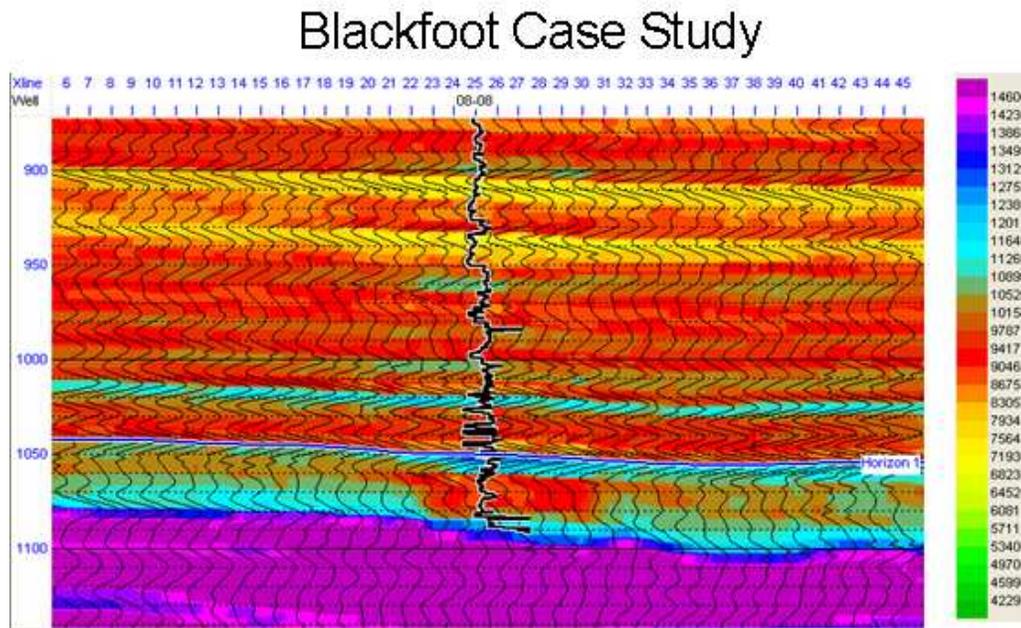


Figure 5.2: Here is the inversion of a seismic line from the survey, showing a channel sand at a time of 1075 ms and the intersection of well 08-08. The values are in impedance (velocity-density) and we wish to predict velocity.

time, and the training values are a set of well-log derived reservoir parameter values at the same depth or time. The well-logging process to obtain sets of high resolution values is illustrated in Figure 5.3. The physical situation is illustrated in Figure 5.4 for $M = 3$. Note from Figure 5.4 that each of the M attributes can be thought of as a N -dimensional vector function $a_i = (a_{1i}, a_{2i}, \dots, a_{Ni})^T$, $i = 1, 2, \dots, M$. The attributes were chosen from a set of over 50 attributes by multilinear regression.

Our objective function for success is that the answer y is as close to t as possible using a least-squares criterion. However, since we can make the error as small as we want by adding more model complexity, the least-squares error is computed using cross-validation. That is, the N training points are sub-divided into K separate classes (in our case, the K individual wells), of lengths N_1, N_2, \dots, N_K . We then leave out each of the K subsets of points in turn, and use the points in the other $K - 1$ wells to predict the removed wells. Next, we compute the least-squared error between the known values and the predicted values. The final error is the average error for the K individual cases.

5.3 Previous Approaches

The following approaches had already been applied to the solution of the problem:

1. The standard multilinear regression approach.
2. The Nadaraya-Watson estimator of multivariate statistics. This approach was re-discovered in the context of neural networks and named the generalized regression neural network (GRNN).

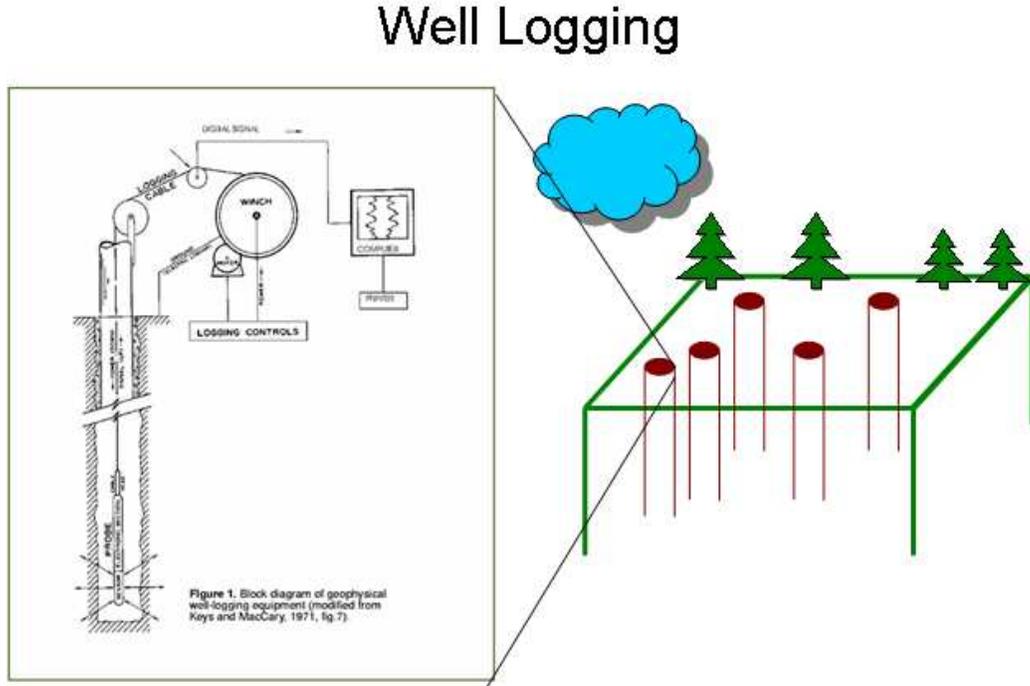


Figure 5.3: The well logging procedure directly obtains high resolution values of the desired parameters.

3. The radial basis function neural network (RBFN).

The multivariate dataset can be written $X = A^T$, where,

$$X = [\mathbf{x}_1 \cdots \mathbf{x}_N] = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MN}, \end{bmatrix}, \quad A = [\mathbf{a}_1 \cdots \mathbf{a}_M] = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix}. \quad (5.2)$$

5.3.1 Multilinear Regression

Multilinear regression of the attributes against the training values involves solving for the weights in the equation,

$$\mathbf{t} = w_0 \mathbf{a}_0 + w_1 \mathbf{a}_1 + \cdots + w_M \mathbf{a}_M = \hat{A} \mathbf{w}, \quad \hat{A} = \begin{bmatrix} 1 & a_{11} & \cdots & a_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{N1} & \cdots & a_{NM} \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_0 \\ \vdots \\ w_M \end{bmatrix}. \quad (5.3)$$

The solution is found by generalized least-squares to be

$$\mathbf{w} = (\hat{A}^T \hat{A} + \lambda I)^{-1} \hat{A}^T \mathbf{t}, \quad (5.4)$$

where λ is a prewhitening factor used to regularize the solution since $\hat{A}^T \hat{A}$ may not have a stable inverse. Also, too much prewhitening can create artifacts.

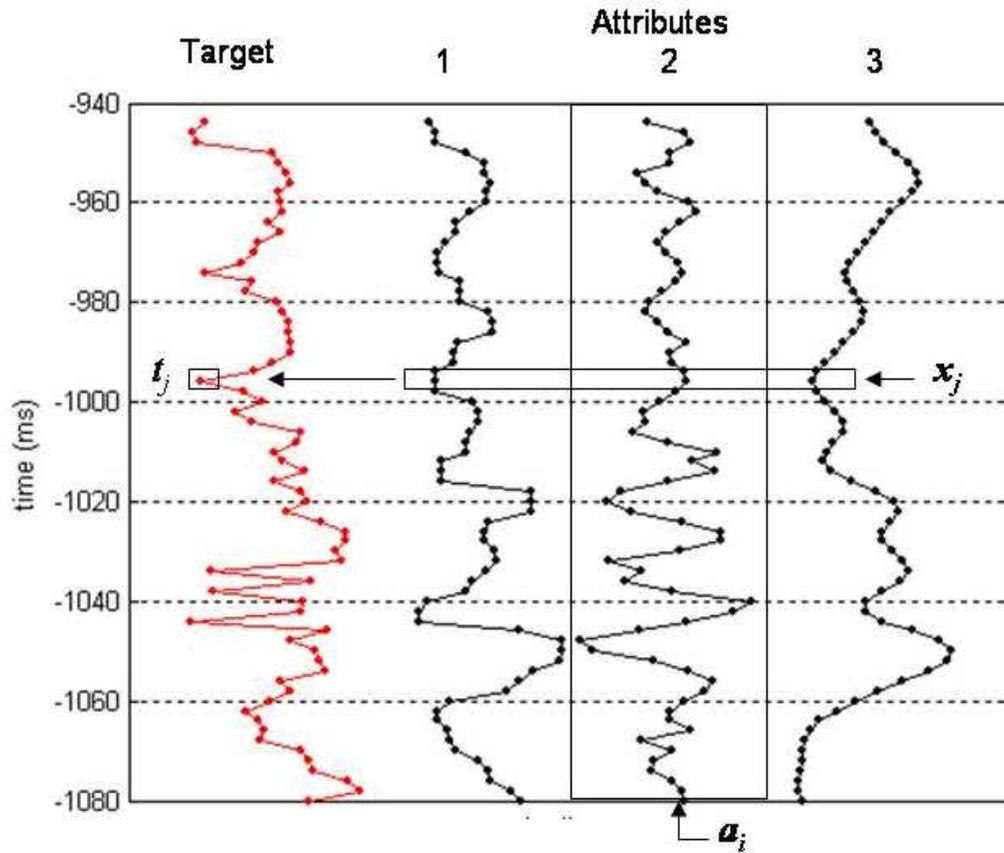


Figure 5.4: The basic prediction problem, where we want to predict t_j from x_j .

5.3.2 GRNN and RBFN

These two neural networks use basis functions. The basis functions are Gaussian functions of distance in attribute space, which can be written as

$$\phi_{ij} = \phi(d_{ij}) = \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), \quad d_{ij} = |x_i - x_j|. \quad (5.5)$$

where σ is a smoothness parameter. The GRNN computes the predicted values “on the fly” from the training values, using the basis functions defined below,

$$f(x_k) = \frac{\sum_{j=1}^N t_j \phi_{kj}}{\sum_{j=1}^N \phi_{kj}}, \quad k = 1, 2, \dots, M. \quad (5.6)$$

In the RBFN, the computation of the predicted values is similar,

$$f(x_k) = \sum_{j=1}^N w_j \phi_{kj}, \quad k = 1, 2, \dots, M. \quad (5.7)$$

However, the weights are computed from the training data using the following linear equations,

$$t(\mathbf{x}_k) = \sum_{j=1}^N w_j \phi_{kj}, \quad k = 1, 2, \dots, N. \quad (5.8)$$

In both GRNN and RBFN, the key parameter to optimize is the σ value, which controls the width of the basis functions. For RBFN, we estimate a global value of σ . For GRNN, we optimize σ so that it varies as a function of the number of parameters in the multivariate observation, or

$$y(\mathbf{x}) = \frac{\sum_{i=1}^N t_i \exp \left[-\frac{(x_1 - x_{i1})^2}{\sigma_1^2} - \dots - \frac{(x_M - x_{iM})^2}{\sigma_M^2} \right]}{\sum_{i=1}^N \exp \left[-\frac{(x_1 - x_{i1})^2}{\sigma_1^2} - \dots - \frac{(x_M - x_{iM})^2}{\sigma_M^2} \right]}. \quad (5.9)$$

A comparison of GRNN and RBFN is as follows. The GRNN has a simpler structure and computes its weights directly from the training data without having to pre-compute these weights. The RBFN pre-computes the weights by solving a $N \times N$ matrix, where N is the number of training points. The two methods give similar results for large numbers of training points, but RBFN is superior when the number of points is small.

5.3.3 Results

Amongst the three methods, GRNN gives the best results having the highest average correlation coefficient between the prediction and the validation target of 12 wells, 0.633. Figure 5.5 shows the best result obtained by GRNN, with the correlation coefficient 0.74. It shows the target (red) and predicted (blue) values of velocity versus time of well 11. The worst case has the correlation coefficient 0.47, which is obtained from well 5, and it is shown in Figure 5.6.

5.3.4 Challenges

Thus we have the following challenges:

1. Can we find other methods that may be more accurate?
2. Can we understand the nature and limitations of the methods from our results?
3. Can we devise other more local estimation procedures?

To address these challenges, we have tried the generalized additive model method and the spline method. In next two sections, we will describe the two methods and show our results. Finally we will give the summary of our work in the last section.

5.4 Generalized Additive Model

5.4.1 Introduction of Generalized Additive Model

The Generalized Additive Model (GAM), a generalization of the linear regression model, can be viewed as the combination of the additive model with the generalized linear model. Suppose that



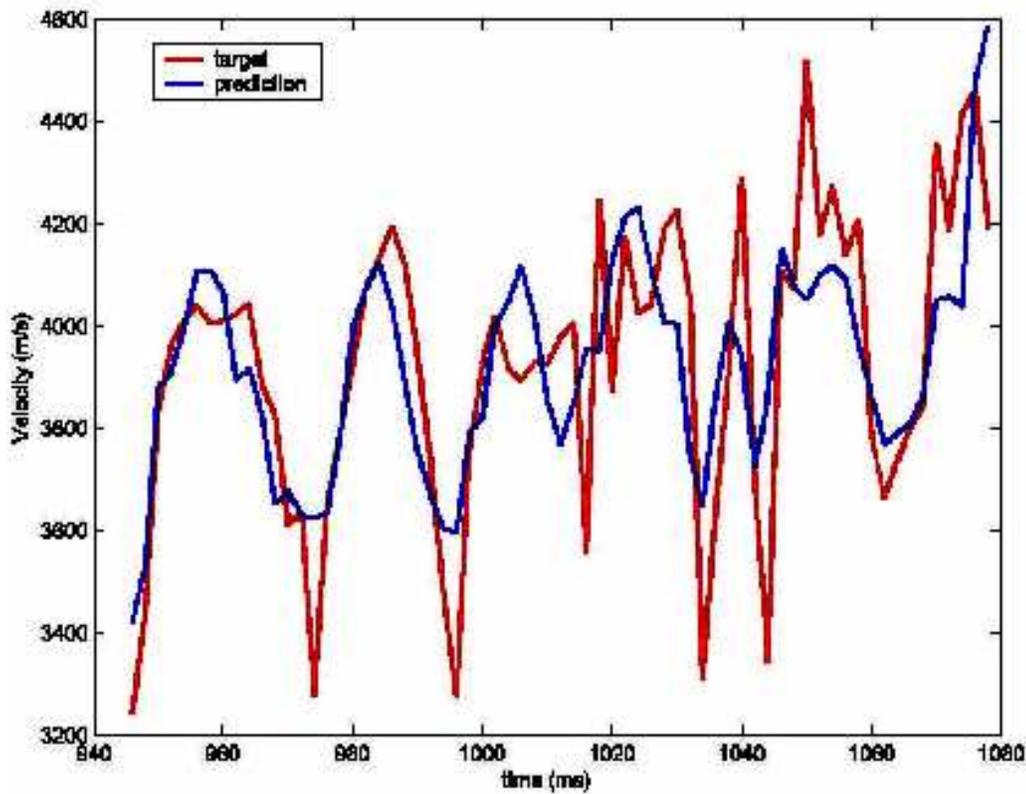


Figure 5.5: The true (red) and predicted (blue) values of velocity Y vs. time (Well 11—the best case).

Y is a dependent random variable and X_1, X_2, \dots, X_p is a set of independent random variables. The standard linear regression model assumes the expected value of Y has a linear form,

$$E(Y) = \alpha + \sum_{j=1}^p \beta_j X_j. \quad (5.10)$$

Therefore the linear regression model assumes the effect of X_j has a linear form $\beta_j X_j$ as well. Given a sample of values for Y and X_1, X_2, \dots, X_p , the parameters β_j are often estimated by the least squares method.

The additive model relaxes the linear form $\beta_j X_j$ and assumes the effect of X_j has a general form $f_j(X_j)$, where $f_j(\cdot)$ is an unspecified (nonparametric) function. These functions are not given a parametric form but instead are estimated in a nonparametric fashion. However, the additive model still assumes the expected value of Y has an additive form,

$$E(Y) = \alpha + \sum_{j=1}^p f_j(X_j). \quad (5.11)$$

To extend the additive model to a wide range of distribution families, Hastie and Tibshirani [2] proposed the generalized additive model. This model assumes that the expected value of Y depends on

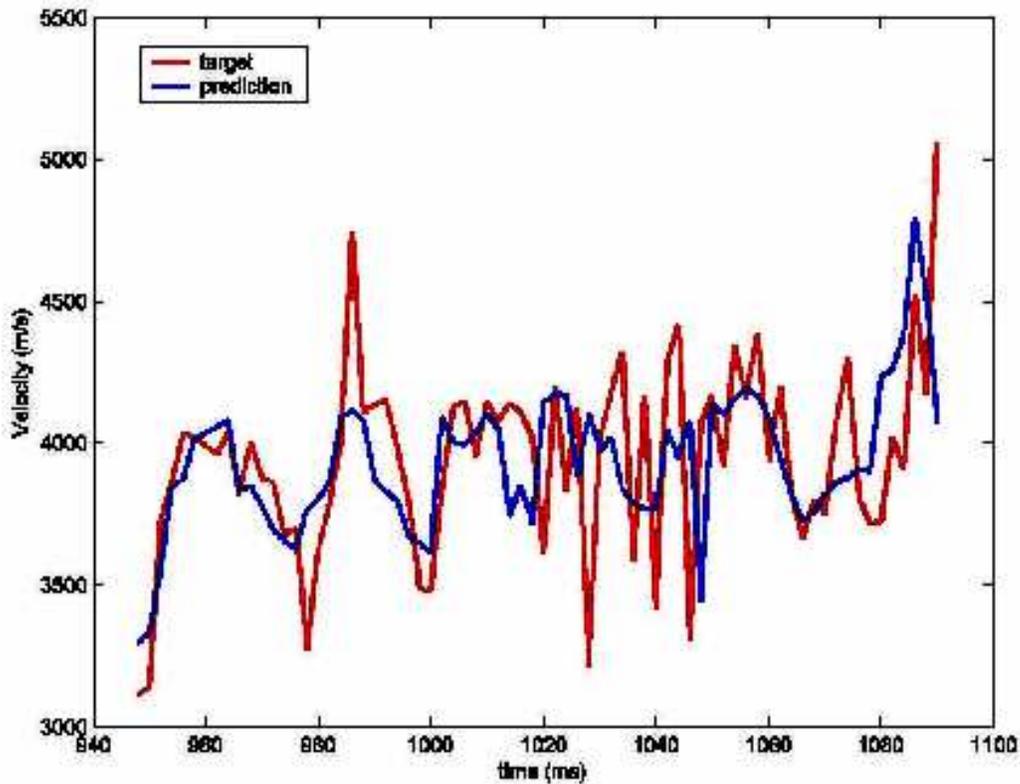


Figure 5.6: The true (red) and predicted (blue) values of velocity Y vs. time (Well 5—the worst case).

additive independent random variables X_j through a nonlinear link function $g(\cdot)$. GAM has the form,

$$g(E(Y)) = \alpha + \sum_{j=1}^p f_j(X_j). \quad (5.12)$$

5.4.2 Advantages of GAM

GAM has two main advantages. First, GAM is a kind of nonparametric regression and it relaxes the usual assumption of linearity, enabling users to reveal the hidden structure in the relationship between the independent variables and the dependent variable. Thus, GAM is more flexible than linear models. GAM includes the linear models as a special case, when g is the identity. Secondly, GAM allows for a link between $\sum_{j=1}^p f_j(X_j)$ and the expected value of Y . This amounts to allowing for an alternative distribution for Y besides the normal distribution. Actually it permits the dependent random variable Y to be any member of the exponential family of distributions.

5.4.3 Fitting GAM

In S-plus, the GAM is fit by the local scoring algorithm, which is based on the backfitting algorithm. The backfitting algorithm is a general algorithm that can estimate the smoothing terms $f_j(X_j)$ in the

additive model using any nonparametric smoothers. The GAM procedure in S-plus has two primary smoothers: splines and Loess (Locally weighted regression). The backfitting algorithm is iterative, starting with initial functions $f_j(X_j)^0$, and in each iteration, fitting the $f_j(X_j)$ to its partial residuals. Iteration proceeds until $f_j(X_j)$ does not change. The local scoring algorithm is used to estimate the smoothing terms $f_j(X_j)$ in the GAM. It is also an iterative algorithm and starts with initial estimates of $f_j(X_j)$. During each iteration, an adjusted dependent variable and a set weight are computed, and $f_j(X_j)$ are estimated using a weighted backfitting algorithm. The scoring algorithm stops when the deviance of the estimates ceases to decrease.

5.4.4 Results

The predicted values of Y are computed using a cross-validation procedure. That is, 855 training points are divided into 12 wells. At each time, we leave out the data points of a given well, and use the points in the other 11 wells to predict of the removed well. Then we calculate the correlation coefficient of the true values and the predicted values in each well which are used to determine the quality of prediction, and finally obtain the average correlation coefficient for 12 wells.

Figure 5.7 shows the true and predicted values of velocity Y versus time of well 6. The correlation coefficient of this well is 0.722, which is the best case among the 12 wells. From this figure, we can see that the predicted values have a similar trend as do the true values. The worst case, well 5 with correlation coefficient 0.364, is shown in Figure 5.8. The true and predicted values of the velocity Y of this well have no obvious correlation.

The average correlation coefficient between the prediction and the validation target of 12 wells is 0.5860 computed by using GAM. Compared to the 0.5857 obtained by the linear regression, we couldn't find an obvious improvement. Moreover, since the purpose of GAM is to maximize the quality of prediction of the dependent variable Y from various distributions, the correlation coefficients of all the wells are expected to be improved by using GAM instead of the linear model. However, we find there are four wells whose correlation coefficients decrease. The reason might be that only one distribution, Gaussian, was considered due to limited time. The smoother methods could also influence the results as well. We chose the spline smoother. One might try different smoothers in S-plus or various combinations of parametric linear functions and non-parametric smooth functions to get better results. Another insight might be provided by the scatter plots in Figure 5.9, which capture the information in the covariance matrix. Only the first attribute (impedance) displayed any significant correlation.

5.5 The Spline Method

5.5.1 Introduction of the Spline Method

Approximation of the target is computed using a function from a certain class. In the case of the spline method, the class of functions consists of the set of all linear combinations of centrally symmetric functions with centres at the training points. That is, $f(x)$ is chosen of the form,

$$f(x) = \sum_{i=1}^n w_i \varphi(|x - x_i|). \quad (5.13)$$



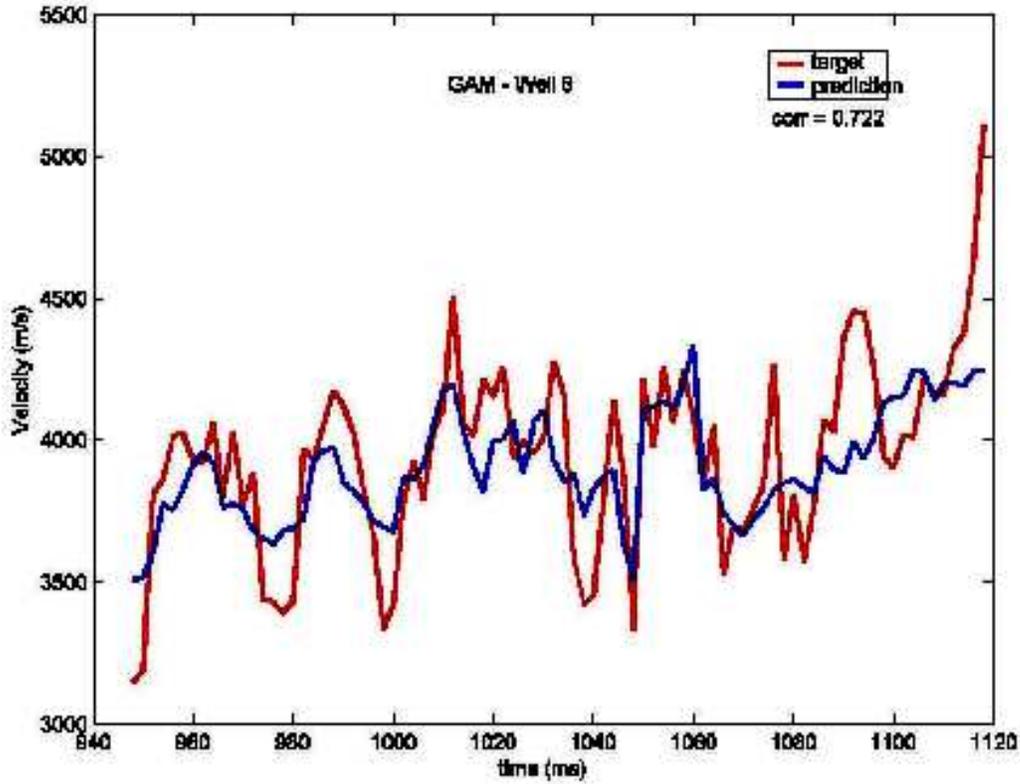


Figure 5.7: The true (red) and predicted (blue) values of velocity Y vs. time (Well 6 –the best case).

where $\varphi(z)$ is a one-variable function which can be experimentally chosen for optimized results, and x_i are the training points. The coefficients w_i are then selected to make the equality $f(x_i) = t_i$ hold exactly at the training points. This gives rise to the following set of n linear constraints on the coefficients w_i ,

$$t_i = f(x_i) = \sum_{j=1}^n w_j \varphi(|x_i - x_j|), \quad i = 1, 2, \dots, n. \quad (5.14)$$

This system of equations has a unique solution that can be easily found as long as the matrix,

$$M = \begin{bmatrix} \varphi(|x_1 - x_1|) & \varphi(|x_2 - x_1|) & \cdots & \varphi(|x_n - x_1|) \\ \varphi(|x_1 - x_2|) & \varphi(|x_2 - x_2|) & \cdots & \varphi(|x_n - x_2|) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(|x_1 - x_n|) & \varphi(|x_2 - x_n|) & \cdots & \varphi(|x_n - x_n|) \end{bmatrix}, \quad (5.15)$$

is invertible and well-conditioned. Solving the system for w_i gives us the predicting function $f(x)$ we are looking for.



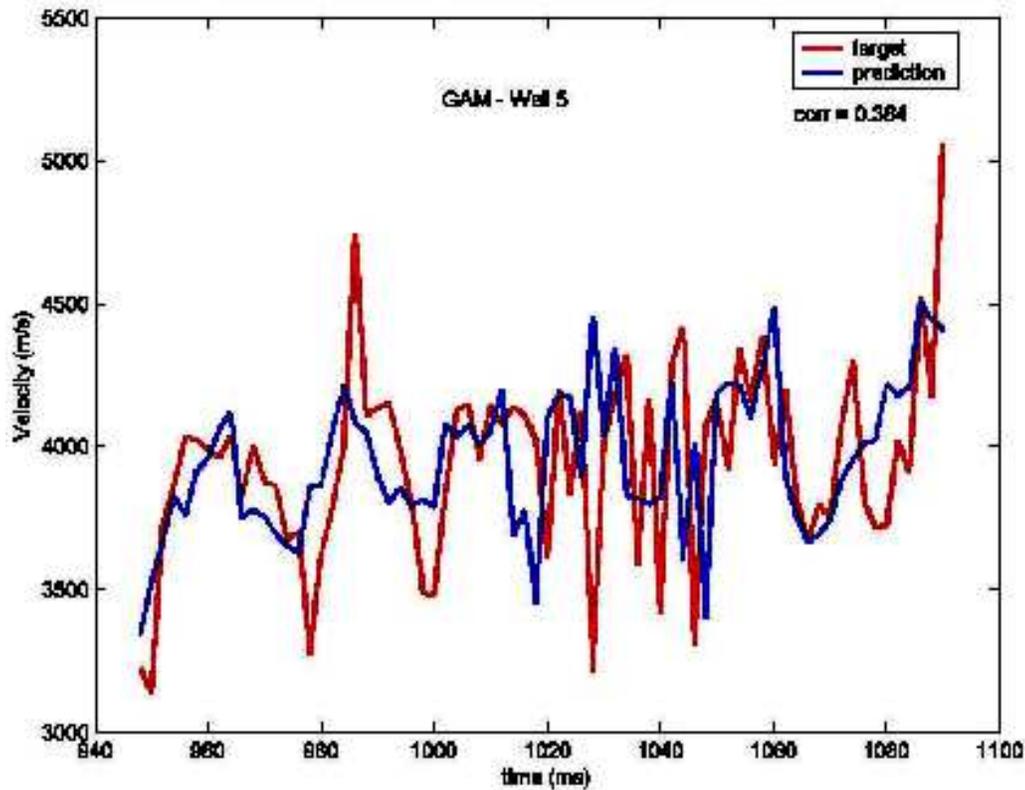


Figure 5.8: The true (red) and predicted (blue) values of velocity Y vs. time (Well 5—the worst case).

5.5.2 Improvement of the Spline Method

The spline method generally works better where there is no linear “drift” in the target function. We improve on the spline method by “taking out” the multi-linear regression component before applying the spline method. Given the input we first compute a multi-linear regression $L(x)$. We then use the spline method on the training set of pairs $(x_i, t_i - L(x_i))$ to obtain a prediction function $g(x)$ as described above. We then output the prediction function for the original input: $f(x) = g(x) + L(x)$. Note that $f(x)$ fits perfectly on the training set, $f(x_i) = t_i - L(x_i) + L(x_i) = t_i$. Figure 5.10 is the schematic diagram of this process described here.

5.5.3 Results and Analysis

We have experimented with functions $\varphi(z)$ of the form $\varphi(z) = z^\alpha + \beta$. The best estimates for the given sample are achieved with,

$$\varphi(x) = x^{0.15}. \quad (5.16)$$

The average correlation coefficient between the prediction and the validation target for 12 wells is 0.631. The average error in the l_2 norm is 0.76σ and the average error in the l_1 norm is 0.57σ , where σ is the sample standard deviation of true target values. Sample performance of the method is illustrated on Figure 5.11 for well 6 (best case) and Figure 5.12 for well 9 (worst case).



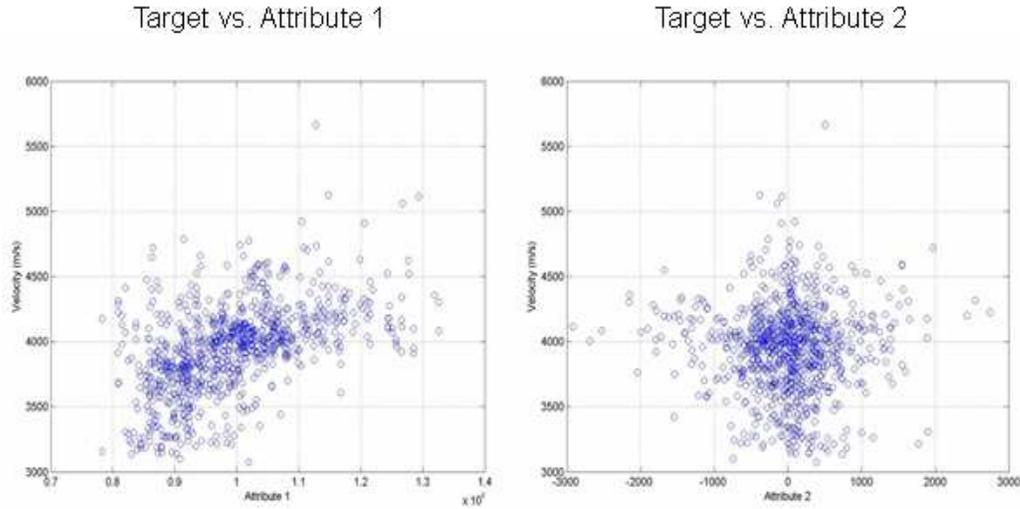


Figure 5.9: Scatter plots of the target versus the attributes. Attribute 1 is the impedance, and it is the only attribute that displays a significant correlation.

The correlations for well 6 and well 9 are 0.773 and 0.489, respectively. Visually, one can see that the prediction quality for well 6 is much better than that for well 9. The main place where the method fails is in predicting spikes. In particular, its performance is worse on well 9 because the target data for this well contains several sharp spikes which the method fails to predict.

5.6 Discussion

All of the methods have limited resolution. They cannot capture the rapid variations occurring on the high resolution scale of the well logging data. The supplemental information contained in the attributes derives from the (relatively) low resolution seismic survey data, thus the statistical attempt to combine the two sets of information into a single predictor will ultimately have a resolution limit as well. Figure 5.13 provides a typical illustration of this limitation.

Throughout the calculations, the number of attributes taken varied from 4 to 7, with the first attribute (impedance) common to all cases. There was relatively little sensitivity to the number of attributes, however, since only the first attribute exhibited any significant correlation (see Figure 5.9), this result is not surprising.

An alternative approach is to consider local estimation; that is to estimate properties in one layer based solely on the deterministic information from the log wells in that layer. This idea is illustrated in Figure 5.14. This is particularly appropriate for localized parameter estimation. Figure 5.15 illustrates a spline fit in the chosen depth plane. Figure 5.16 illustrates the predicted values in the chosen well based on the local spline estimation compared with the true values. This fit appears to be consistent with the best of the earlier spline fit results.

In conclusion, the spline method seemed to offer the best results. Given the limited duration of the IPSW, however, this is a tentative conclusion. There are other statistical approaches beside the GAM,

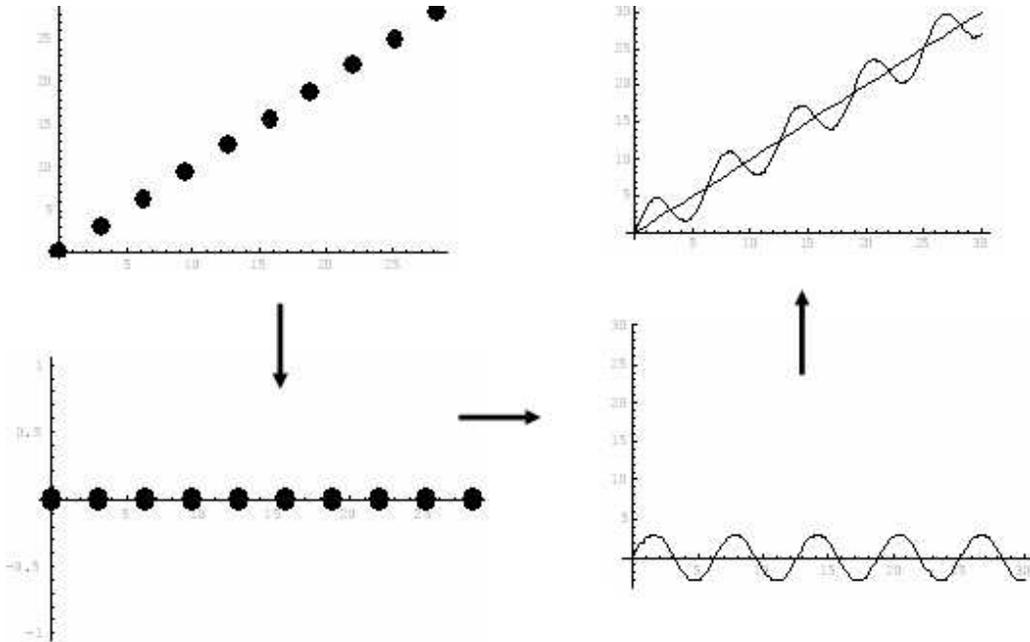


Figure 5.10: Improving the Spline Method.

and they should probably be considered. There is also the question of to what degree these examples were dominated by the first attribute.

5.7 Acknowledgements

The authors would like to acknowledge the Pacific Institute for the Mathematical Sciences (PIMS) for their sponsorship of the IPSW. The University of Calgary is also acknowledged for their generous hospitality in hosting the IPSW. Finally, specific thanks to Elena Braverman and Marian Miles for all of their hard work.

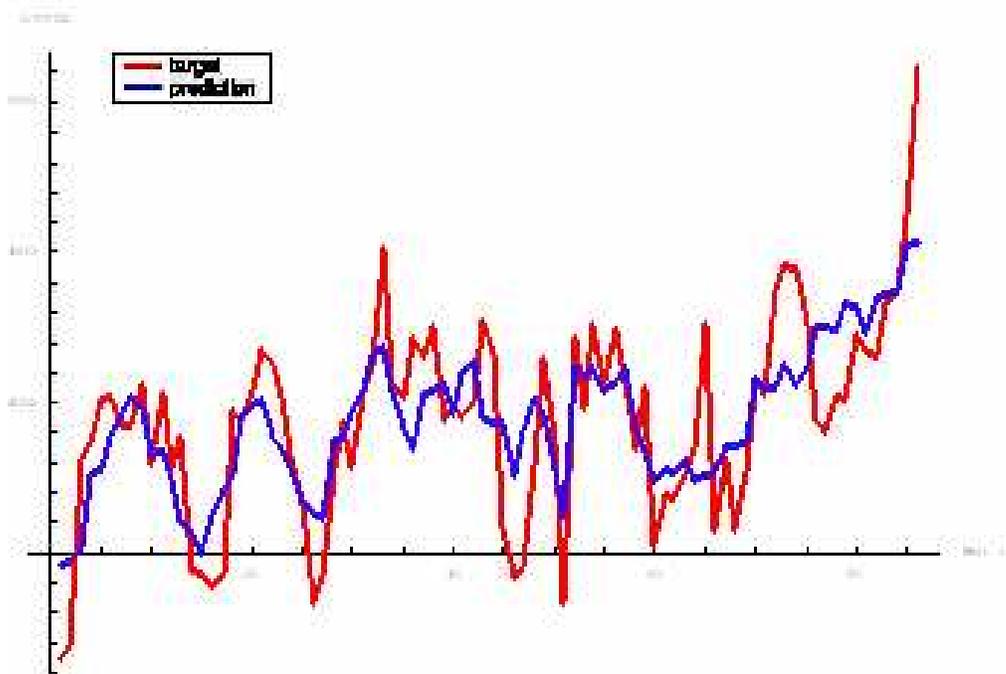


Figure 5.11: The true (red) and predicted (blue) values of velocity Y vs. time (Well 6—the best case).

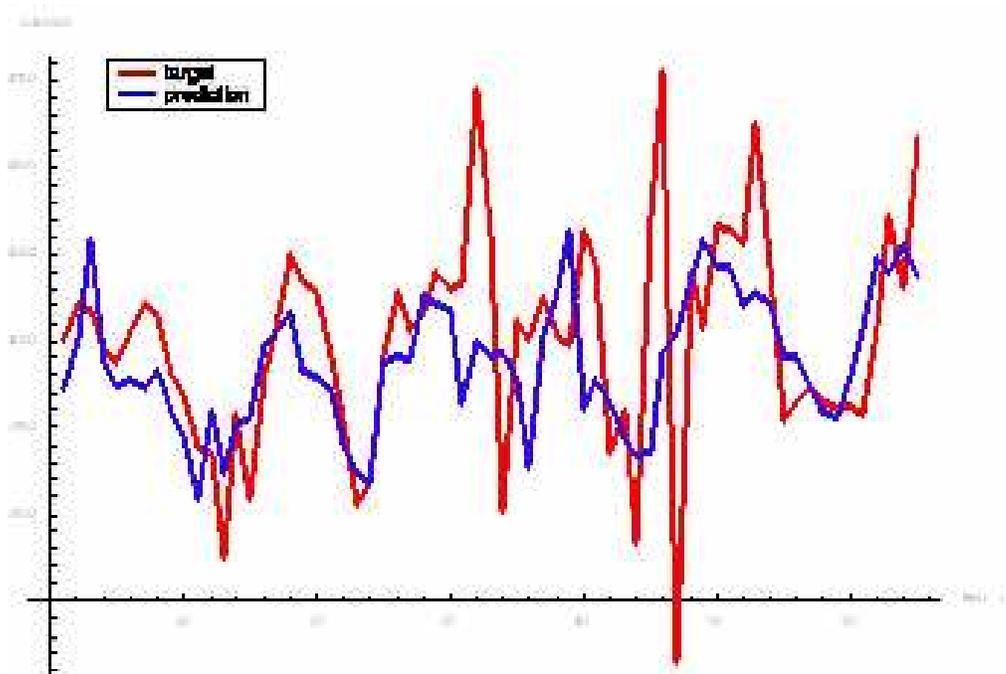


Figure 5.12: The true (red) and predicted (blue) values of velocity Y vs. time (Well 9—the worst case).

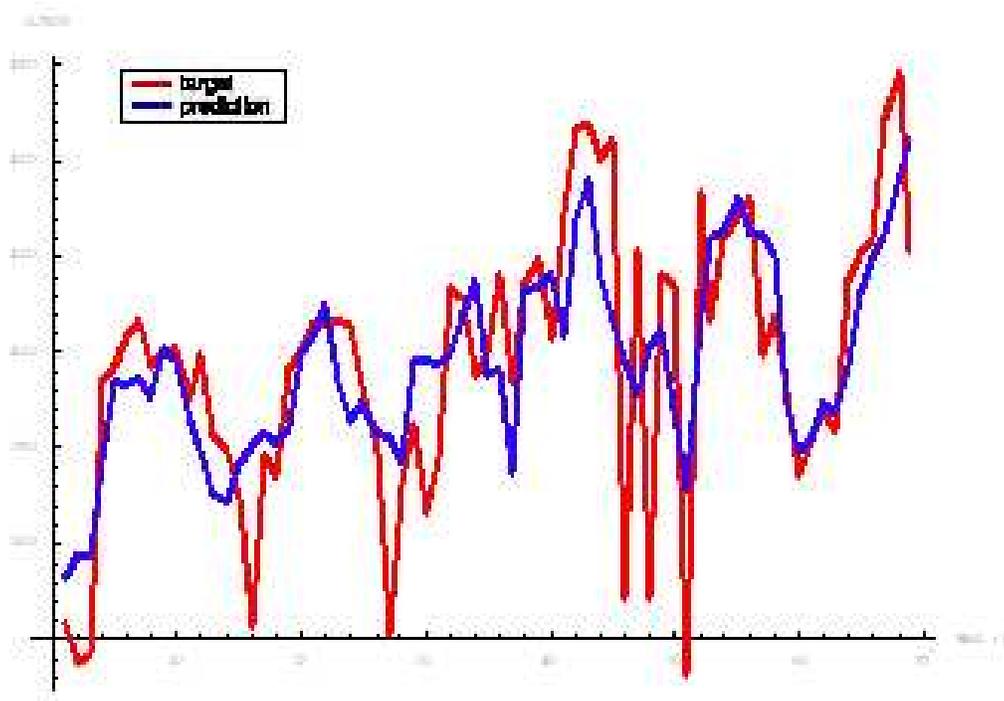


Figure 5.13: The true (red) and predicted (blue) values of the velocity Y vs. time for a typical example in the study. Notice that the rapidly changing features are very hard to predict correctly.

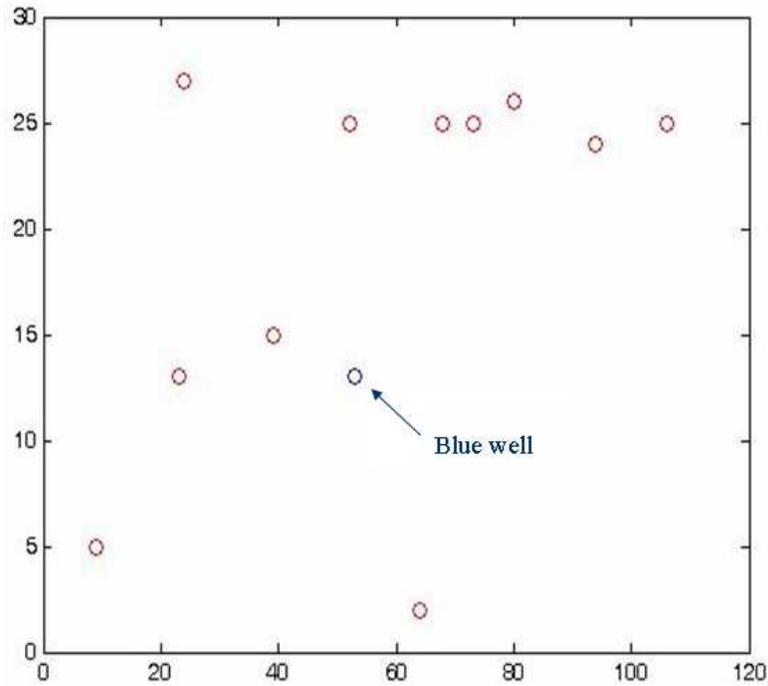


Figure 5.14: The location of the 12 wells in a particular depth plane, with the blue well signifying the well to be predicted.

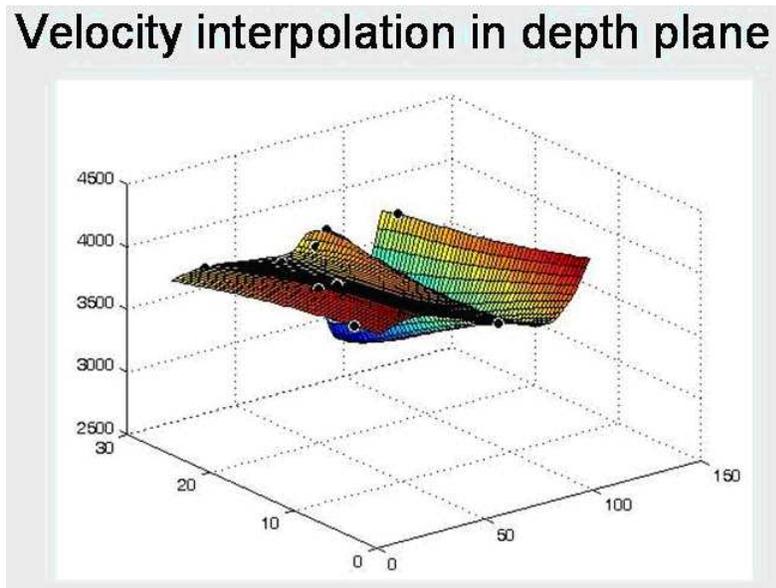


Figure 5.15: A spline fit in a given depth plane.

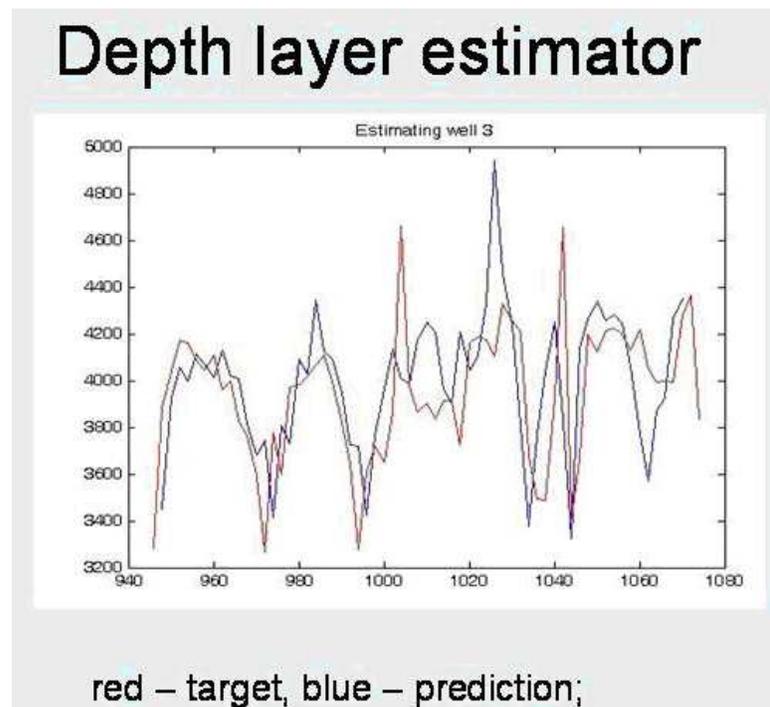


Figure 5.16: The true (red) and the predicted (blue) values for the velocity Y vs. time for the local depth spline estimator.



Bibliography

- [1] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- [2] Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall: New York.
- [3] Johnson, R. A. & Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*. Prentice Hall.
- [4] Masters, T. (1995). *Advanced Algorithms for Neural Networks*. John Wiley & Sons, Inc.
- [5] Russell, B. (2005). *The Application of Multivariate Statistics and Neural Networks to the Prediction of Reservoir Parameters from Seismic Attributes*: Ph.D. dissertation, University of Calgary.
- [6] Russell, B. (2005). *Seismic Prediction of Reservoir Parameters (An IPSW Problem)*.