A machine-learning approach for noise reduction in parameter estimation inverse problems, applied to characterization of oil reservoirs

To cite this article: C. Minutti et al 2018 J. Phys.: Conf. Ser. 1047 012010

View the article online for updates and enhancements.
A machine-learning approach for noise reduction in parameter estimation inverse problems, applied to characterization of oil reservoirs

C. Minutti, S. Gomez, G. Ramos
National University of Mexico, Institute of Applied Mathematics, Mexico City, Mexico
E-mail: carlos.minutti@iimas.unam.mx

Abstract. Reservoir characterization is an inverse problem where parameter values associated to the properties of the porous media, are estimated. In this problem, the pressure data and its log-derivative curves are used in the fitting process. However, the numerical differentiation needed to generate the log-derivative curve, is an ill-posed problem where the noise in the data can be largely propagated. This noise can produce several spurious local minima in the objective function and prevents to get a precise approximation to the parameters. Therefore, an efficient noise reduction method is required to achieve the desired accuracy in the parameter values.

In this work, we explore three noise reduction methods by applying them to well-test data. These methods are based on multi-step finite differences and splines, but using a machine-learning approach.

In addition, a new method is proposed which produce an optimal curve which is a linear combination of multiple data curves, that are approximations to the solution of the inverse problem.

We show how these methods which use information from the forward model are more efficient for noise reduction. The proposed methods can be used in many inverse problems where there is a mathematical model meant to describe the measured data.

1. Introduction
In many inverse problems, where the parameters of a forward model are identified through measured data, the noise in the data might be a major problem, especially in complex models (i.e. with a large number of parameters).

The characterization of oil reservoirs through well tests analysis is one of these problems, where the log-derivative curve (the derivative of the measured pressure in the well with respect to the logarithm of time), has also to be fitted as it indicates several characteristics of the reservoir. However, since the numerical derivation is an ill-posed problem, the noise of the data is highly propagated, making it difficult to achieve a good estimation of the parameters of the model.

For reservoir characterization, the forward model that we are using is a Triple Porosity-Double Permeability (TP-DP) model [2], in which it is necessary to identify up to 10 parameters for a real case (see [5], [6] and [8] where reservoir characterization using the TP-DP model, is described). The solution of the model for a given set of parameters, produces the curve of the pressure in the well under study. Due to the number of parameters to be identified, this model
can give more information of the porous media than the conventional double-porosity model of Warren and Root [11]. Thus, it is necessary a noise reduction method to be able to produce the so called type-curves of the pressure and the log-derivative, that have to be fitted to produce accurate estimate of the parameters of the model.

The Bourdet algorithm [1] based in a finite difference approach, has been the standard method to calculate the log-derivative curve of well-test data, but it has shown to be inefficient when the noise level is high.

Lane et al. [7] address the problem of noise reduction in well-test data for a simpler forward model, using splines in which the smoothing parameters are optimized. Later, Escobar et al. [3] also address the problem and compare different noise reduction methods, which are based on finite difference schemes (mainly the Bourdet algorithm with different values of $L$ and the derivative of interpolated polynomials) and splines, concluding that splines is the most effective method.

The approach of using multi-step finite differences and/or splines, to reduce the noise in the data, is well extended in many different areas. However, the finite differences based on the interpolation of polynomials at several points, considered by Escobar et al. actually increase the noise in the data.

In this study, we compare three methods of noise reduction: Multi-step finite differences,Splines and a Mixture of sub-optimal curves. In both, splines and multi-step finite differences, we propose a machine-learning approach, that uses information from the forward model to determine the optimal parameters of each of the methods. Being splines a method widely used in different areas to reduce noise in the data, we also study the effect of another type of splines: regression splines.

In the first part of the paper we present the methodology of how the classic algorithms of multi-step finite differences and splines are adapted to a machine-learning approach and the proposed method of mixture of sub-optimal curves is described. In the second part, we present the results obtained when applying these noise reduction methods to a large data set, measuring the error of the method and also its effect in the parameter estimation error of the inverse problem.

2. Methodology

Considering a machine-learning approach, we use the forward model (the TP-DP model) to generate a first group of 1,000 random synthetic data sets, for which Gaussian noise was subsequently added. In each of the methods to be studied, this data set is used to train the algorithms, that is, to determine the optimal value of the parameters that minimize the error between the noisy and noise-free data. Later a second group of 1,000 random synthetic data sets is generated, for which noise is also added. This data set is the cross-validation data set and we use this data to compare the performance of the proposed algorithms, with the classic approach and also with the optimized parameters. The purpose of determining the optimal parameters of each algorithm with a training data set, is to help the noise reduction method to reproduce better the kind of data created by the forward model in the inverse problem.

The error of the noise reduction algorithm at each data point is measured by:

\[
\text{Error} = \left| \frac{d'_{\text{algorithm}} - d'_{\text{noise-free}}}{d'_{\text{noise-free}}} \right|
\]

where $d'_{\text{noise-free}}$ is the log-derivative of the data without noise (in the synthetic examples) and $d'_{\text{algorithm}}$ is the log-derivative of the obtained data, after applying the noise reduction algorithm to the noisy data.
To compare the performance of each method we use the mean and median value of \( \log_{10}(\text{Error}) \) because, as we are measuring a relative error, we are more interested in the order of magnitude of the error. Also because the mean is more sensitive to the existence of outliers than the median, when relative errors are used, \( \log_{10}(\text{Error}) \) has a more symmetric distribution, so it is less affected by outliers.

2.1. Multi-step finite differences

In the case of noisy data, it does not make sense to use finite difference schemes that are based on interpolation of polynomials (as it is the case of the schemes presented by Forner 
[4]), since they increase the noise of the numerical derivative (this happens in Escobar 
[4].) Instead, we use the schemes presented by Savitzky and Golay [10], where the base polynomial, fits the data in the least squares sense, following an idea similar to the Lanczos derivative. Based on that, we choose a 9-point scheme because this is the maximum number of points showing a relevant improvement in the noise reduction for a set of data samples. The finite difference 9-point formula of Savitzky and Golay [10] is:

\[
f'_0 \approx \frac{(f_1 - f_{-1}) + 2(f_2 - f_{-2}) + 3(f_3 - f_{-3}) + 4(f_4 - f_{-4})}{60h}
\]

We propose here another 9-point finite difference method:

\[
f'_0 \approx \frac{1}{4} \left( \omega_1 \frac{f_1 - f_{-1}}{2h} + \omega_2 \frac{f_2 - f_{-2}}{4h} + \omega_3 \frac{f_3 - f_{-3}}{6h} + \omega_4 \frac{f_4 - f_{-4}}{8h} \right)
\]

where \( \sum_{i=1}^{4} \omega_i = 1 \). The weights values \( \omega_i \) are determined by minimizing the error (equation 1) obtained with the training data set. Subsequently, the errors of both methods are compared using the cross-validation data set.

2.2. Splines

For splines, we consider three cases. First, a classic approach where a cubic spline is used, and the smoothing parameter is determined by cross-validation, that is, for each one of the 1,000 data sets, a part of the data is used to determine the spline coefficients, and the other part to determine the smoothing parameter, selecting in each case the value that minimizes the mean squared error.

In the second approach, we use optimization to determine a single smoothing parameter value, which minimizes the error of the entire training data set and the same value will be used in the cross-validation data set.

In the third approach we use regression splines (see [9]) where, unlike the cubic splines, it is possible to use different types of base functions to approximate the data. In order to have a good result, the base functions must be similar or related to the functions that are generated by the forward model which are used to fit the measured data. In the case of the TP-DP model, the asymptotic approximations of the model at early and late times (which appear in Camacho 
[10]), are used as base functions. The smoothing parameter is determined in the same way as in the previous case, using optimization to get the value that minimizes the error of the entire training data set.

The performance of each algorithm is determined by measuring the error of the equation (1) in the cross-validation data set.
2.3. Mixture of sub-optimal curves

We propose here a method that we call mixture of sub-optimal curves, which consists of using the noisy (pressure) data to perform a series of quick optimizations to solve the inverse problem from different starting points, in order to obtain approximations to local optima, that is, to obtain different values of the parameters of the forward model, which can approximate the data in a rough way. These approximated data are the sub-optimal curves.

These sub-optimal curves are combined into one curve using a linear combination, where the weight of each curve is determined by minimizing the quadratic error between the combined curve and the noisy data. Based on the multiple linear regression model, we can get the vector of the weights \( b \) using the regression normal equation:

\[
b = (X'X)^{-1}X'y
\]  

where, in this model, \( X \) is a matrix whose columns are the different sub-optimal curves values and \( y \) is a vector which contains the noisy data.

A regularization parameter \( \lambda \) is added to the model in order to avoid problems of singular matrices (due to similar sub-optimal curves) and also to avoid overfitting, which can lead to reproduce the noise in the data. Thus, applying Tikhonov regularization, the final model to estimate the weights of each curve is:

\[
b = (X'X + I\lambda)^{-1}X'y
\]  

where \( I \) is the identity matrix.

The noise-reduced curve and its log-derivative are obtained with \( b'X \) and \( b'X_d \), where \( X_d \) is the matrix whose columns are the log-derivative of the columns of \( X \).

Both, the best number of sub-optimal curves to be used and the value of the regularization parameter, are obtained by optimizing (minimizing) the error of the training data set. For this case, 40 sub-optimal curves were obtained from the TP-DP model for each of the 1,000 cases in the training data set.

Once these parameters are determined, the process is applied to the cross-validation data set in order to estimate the error of the proposed method.

2.4. Training process and noise reduction comparative

The following is a summary of the training process for each algorithm and how the noise reduction comparison is performed.

(i) A training and a cross-validation data sets are created. Each set has 1,000 synthetics data sets which are generated by solving the forward model using random parameter values and adding random noise to the data.

(ii) For the cubic splines method, the regularization parameter is estimated dynamically for each data set by cross-validation, where part of the data is used to find the optimal value of the regularization parameter by minimizing the quadratic error.

(iii) For the trained algorithms, the training data set is used to obtain the (static) parameter values of each method by minimizing the error between the noisy and the noise-free data, using optimization.

- In the case of multi-step finite differences, the optimization is used to estimate the optimal weights of each finite difference.
- In the case of the spline based algorithms, the optimization is used to estimate the optimal value of the regularization parameters.
- In the case of the mixture of sub-optimal curves, the optimization is used to estimate the number of sub-optimal curves to be used as well as the regularization parameter value.
(iv) The cross-validation data set is used to estimate the error of each method applying equation (1), where the weights and regularization parameters (previously estimated with the training data set) are used. This data set is also used to estimate the error of the classic methods.

2.5. Error of the forward model parameters

In order to measure the effect of these noise reduction methods in the estimation of the parameters of the forward model, we solve the inverse problem using an optimization following the procedure described in Minutti et al. [8] to obtain an estimation of the parameters of the forward model for the synthetic data in the cross-validation data set.

Additionally, we provide as a starting point to the optimization algorithm, the true value of the parameters with which the noise-free data was generated. Because the optimization algorithm has as starting point the true value of the parameters, any error in the parameter estimates is due to the noise in the data.

To measure the error of each parameter we use the following formula:

$$err(\hat{\theta}) = \frac{|\theta - \hat{\theta}|}{\max_{\theta} - \min_{\theta}}$$ (6)

where $\theta$ is the true value of the parameter, $\hat{\theta}$ is the estimated value given by the optimization algorithm, and $\max_{\theta} - \min_{\theta}$ is the length of the interval of possible values that the parameter can take. Thus, equation (6) measure the absolute error expressed between 0 to 1.

3. Results

Using the cross-validation data set, the mean and median of $\log_{10}(Error)$ are determined for each of the methods described above, obtaining the values of Table 1, being the classic methods the 9-point scheme Savitzky and Golay [10] (FD-9_CLS) and the cubic splines (SPL_CLS).

| Multi-step finite differences | Smoothing Splines | Mixture of sub-optimal curves |
|-----------------------------|------------------|------------------------------|
| FD-9_CLS                    | FD-9             | SPL_CLS                      |
| Mean                        | -1.6257          | -1.7406                      |
| Median                      | -1.5021          | -1.6342                      |
| SPL_CLS                     | -2.1575          | -2.1846                      |
| SPL                         | -2.1716          | -2.1875                      |
| REG_SPL                     | -2.2604          | -2.2846                      |
| MIX_OPT_C                   | -2.5310          | -2.5974                      |

It can be observed that, as in multi-step finite differences and also in splines, the methods where the parameters of the algorithms were determined using a training data set (FD-9, SPL, REG_SPL), are more efficient than using a classic method (FD-9_CLS, SPL_CLS). It is also observed that the method of regression splines (REG_SPL) is more efficient than the other two methods of splines, and the method of mixture of sub-optimal curves is better than any other of the analyzed algorithms. Specifically, the mixture of sub-optimal curves (MIX_OPT_C) has only 53.6% of the noise of regression splines, the second best method ($10^{-2.5310}$ ≈ 0.5363), showing that the use of information of the forward model, produces a more efficient method for noise reduction.

Figures 1 and 2 show the boxplots and histograms for the methods where the algorithms were trained: Multi-step finite differences (FD-9), cubic splines with optimized smoothing parameters
Figure 1. Boxplot of $\log_{10}(Error)$ for the trained algorithms.

Figure 2. Histogram of $\log_{10}(Error)$ for the trained algorithms.
(SPL), regression splines with optimized smoothing parameters (REG_SPL) and the mixture of sub-optimal curves (MIX_OPT_C).

In Figures 3-5 we present three examples taken from the cross-validation data set, where the pressure and its derivative (log-derivative) are plotted on a log-log scale with the result of the three best noise reduction methods: trained splines (cubic and regression splines), and mixture of sub-optimal curves. The noisy data and its derivative, calculated with a finite difference method (Bourdet), is also presented.

The major noise reduction is obtained with the mixture of sub-optimal curves. Cubic splines and regression splines have problems of oscillations, especially in the valleys of the log-derivative curves.

![Derivative with Central Differences](image1)

![Noise reduction with Splines](image2)

![Noise reduction with Regression Splines](image3)

![Noise reduction with a Mixture of Sub-optimal Curves](image4)

**Figure 3.** Example 1: Noise reduction with the three best methods.

In Table 2 we show the mean estimation error (using equation 6) of the parameters of the forward model, when we solve the inverse problem using the noisy data in the cross-validation data set and also when we use the noise-reduced data given by the trained algorithms. As we provide to the optimization algorithm with the true value of the parameters as a starting point, the estimation error is due to the noise in the data.

For each parameter a significant reduction in the estimation error can be observed, compared with the noisy data. For parameters like $\omega$, $\lambda_{mf}$, $\lambda_{mv}$ and $\kappa$, the mean estimation error is one-third of the error of the noisy data when the mixture of sub-optimal curves was used.

Although regression splines was a better noise reduction method than cubic splines, there was not much difference in the mean estimation error of the parameters.

Results in Table 2 show that:
Figure 4. Example 2: Noise reduction with the three best methods.

Table 2. Mean estimation error of the model parameters with different noise reduction methods

| Method          | $\omega_f$ | $\omega_v$ | $\lambda_{mf}$ | $\lambda_{me}$ | $\lambda_{ef}$ | $\kappa$ | $s$  | $C_D$ |
|-----------------|------------|------------|-----------------|-----------------|-----------------|----------|------|-------|
| NOISY DATA      | 0.1862     | 0.1844     | 0.1008          | 0.0966          | 0.1615          | 0.1567   | 0.0148| 0.0011 |
| SPL             | 0.0955     | 0.0962     | 0.0403          | 0.0355          | 0.0775          | 0.0670   | 0.0012| 0.0003 |
| REG_SPL         | 0.0923     | 0.0945     | 0.0395          | 0.0365          | 0.0800          | 0.0639   | 0.0011| 0.0003 |
| MIX_OPT_C       | 0.0666     | 0.0623     | 0.0360          | 0.0340          | 0.0719          | 0.0488   | 0.0005| 0.0001 |

- If a suitable noise reduction method is not used, even with a good initial guess of the parameters (or a good optimization algorithm), due to the noisy data we can only expect a limited accuracy in the estimation of the parameters of the inverse problem.
- Even if a noise reduction method performs better (i.e. smaller error between the noisy data and the noise-free data), it may not have a significant effect on the inverse problem, this is the case for SPL and REG_SPL, where REG_SPL has a smaller error than SPL (Table 1), but the improvement in the mean estimation error of the parameters is almost the same. This may mean that part of the improvement done by the algorithm could be in an area (a time interval in our pressure curve) where the data may have little sensitivity to the parameters.
- The MIX_OPT_C method does not have the problem mentioned above, because we are performing quick optimizations, solving the inverse problem many times and it is because of this that the shape of the curve changes mainly in areas that are sensitive to the parameters.
4. Conclusions

By training the algorithms with a synthetic data set it can be seen a smaller error between the noisy and the noise-free data compared with a classic approach. Especially when a regression spline (with base functions derived from the forward model) is used, we obtain better results than using a classic cubic spline.

The proposed method, a mixture of sub-optimal curves, combines the idea of using base functions from the forward model and the training of the algorithm in order to get a set of optimized parameters to the problem. This method was the most efficient for noise reduction, having only 53.6% of the noise that has the regression splines, being the former the second best method.

When the inverse problem was solved for the cross-validation data set, we observe the impact of the different noise reduction methods in the parameter estimation problem, being the mixture of sub-optimal curves also the best method, having for many parameters, up to one-third of the estimation error, as compared with the error of the noisy data.

Because in many inverse problems there is a mathematical model which is used to approximate or reproduce a real phenomena (forward model), the methods presented in this work can be applied to solve them, as the forward model can be used to train the algorithms to reduce the noise in the observed data. Especially mixing different sub-optimal solutions of the inverse problem is a good approach to get noise-reduced data which is more consistent with the real phenomena, and because of that it can be better exploited in an inverse problem.

These methods, based in the forward model, are especially important when there are many parameters to be identified, because it increases the likelihood of adjusting the solution of the
inverse problem to noise in the data. So, a noise reduction method which is consistent with the forward model helps to improve the accuracy of the estimated parameters.

In our application (characterization of oil reservoirs), these parameter estimates can be used to take decisions to forecast production. It is because of this that a more accurate estimation of the parameters can have a large impact in the production of a well.

References
[1] Bourdet D, Ayoub J A and Pirard Y M 1989 Use of Pressure Derivative in Well Test Interpretation. em Society of Petroleum Engineers. doi:10.2118/12777-PA
[2] Camacho-Velazquez R, Vasquez-Cruz M A, Castrejon-Aivar R and Arana-Ortiz V 2005 Pressure Transient and Decline Curve Behaviors in Naturally Fractured Vuggy Carbonate Reservoirs. Society of Petroleum Engineers. doi:10.2118/77689-PA.
[3] Escobar F H, Navarrete J M and Losada H D 2004 Evaluation of Pressure Derivative Algorithms for Well-Test Analysis. Society of Petroleum Engineers. doi:10.2118/86936-PA.
[4] Fornber B 1988 Generation of Finite Difference Formulas on Arbitrarily Spaced Grids. Mathematics of Computation, Vol. 51, No. 184. pp. 699 – 706, doi:10.2307/2008770
[5] Gomez S, Camacho R, Vasquez M, del Castillo N and Mesejo J A 2014 Well Test Characterization of Naturally Fractured Vuggy Reservoirs, with a Global Optimization Method. Offshore Technology Conference. doi:10.4043/24762-MS
[6] Gomez S, Ramos G, Mesejo A, Camacho R, Vasquez M and del Castillo N 2014 Well test analysis of naturally fractured vuggy reservoirs with an analytical triple porosity - double permeability model and a global optimization method, In Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles. pp. 653 – 671
[7] Lane H S, Lee W J and Watson A T 1991 An Algorithm for Determining Smooth, Continuous Pressure Derivatives From Well-Test Data. Society of Petroleum Engineers. doi:10.2118/20112-PA
[8] Minutti C, Ramos G, Gomez S, Camacho R, Vasquez M and del Castillo N del 2016 Robust Characterization of Naturally Fractured Carbonate Reservoirs Through Sensitivity Analysis and Noise Propagation Reduction. Society of Petroleum Engineers. doi:10.2118/181182-MS
[9] Monahan J F 2011 Numerical Methods of Statistics. Cambridge University Press. chapter 7 pp 159–163
[10] Savitzky A and Golay M J E 1964 Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. In Analytical Chemistry, Vol. 36, No. 8, pp. 1627-1639, doi:10.1021/ac60214a047
[11] Warren J E and Root P J 1963 The Behavior of Naturally Fractured Reservoirs. Society of Petroleum Engineers. doi:10.2118/426-PA