Published in final edited form as:

Conf Rec Asilomar Conf Signals Syst Comput. 2014 November ; 2014: 420–422. doi:10.1109/ACSSC.2014.7094476.

Piecewise Linear Slope Estimation

A. N. Ingle[1,2], **W. A. Sethares**[1], **T. Varghese**[1,2], and **J. A. Bucklew**[1]

A. N. Ingle: ingle@wisc.edu; W. A. Sethares: sethares@ece.wisc.edu; T. Varghese: tvarghese@wisc.edu; J. A. Bucklew: bucklew@engr.wisc.edu

[1]Depts. of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI, USA

[2]Depts. of Medical Physics, University of Wisconsin-Madison, Madison, WI, USA

Abstract

This paper presents a method for directly estimating slope values in a noisy piecewise linear function. By imposing a Markov structure on the sequence of slopes, piecewise linear fitting is posed as a maximum *a posteriori* estimation problem. A dynamic program efficiently solves this by traversing a linearly growing trellis. The alternating maximization algorithm (a kind of pseudo-EM method) is used to estimate the model parameters from data and its convergence behavior is analyzed. Ultrasound shear wave imaging is presented as a primary application. The algorithm is general enough for applicability in other fields, as suggested by an application to the estimation of shifts in financial interest rate data.

I. Introduction

Piecewise linear functions are encountered in a variety of applications such as medical imaging, biological measurements, geology, economics and social sciences. The problem of fitting piecewise linear functions (and more general functions with regime changes) has been studied for many years in the statistics and signal processing literature [1], [2], [3], [4]. In some applications, the slopes of the constituent pieces of a piecewise linear response have a physical interpretation. Hence it would be expedient to design an algorithm that optimally estimates these slopes instead of producing a fit to the original function. The problem of slope estimation for continuous piecewise linear functions is addressed in this paper using a Bayesian maximum *a posteriori* (MAP) estimation approach. The strength of this method lies in the fact that the number and locations of slope changes is handled implicitly by the stochastic model and need not be specified in advance. Moreover, model parameters can be automatically estimated from data.

II. Data Model

It is assumed that the user has a reliable estimate of the minimum and maximum possible slopes that are present in the noisy data series. Hence, without loss of generality, the piecewise linear function is assumed to consist of slope values in [0, 1]. The data sequence

Correspondence to: A. N. Ingle, `ingle@wisc.edu`; T. Varghese, `tvarghese@wisc.edu`.

$\{Y_n\}_{n=1}^{N}$ is assumed to originate from observations of a hidden slope sequence $\{S_n\}_{n=1}^{N}$ through a "noisy accumulator" given by:

$$Y_n = \sum_{i=1}^{n} S_i + w_n,$$

where $w_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. For computational tractability, each slope value S_n is discretized into K equidistant bins $\{0, \frac{1}{K-1}, \ldots, 1\}$.

A Markov structure is imposed on the sequence of slope values as follows. The slope remains constant with a probability $1 - p$ and jumps to a new uniformly randomly chosen value with probability p, with the initial slope value S_1 chosen uniformly randomly from the discrete slope set. The log posterior probability mass function of the (unknown) slopes conditioned on observed data can be derived for this model and after some algebraic manipulations, the MAP estimation problem can be expressed as:

$$(s_1^*, \ldots, s_N^*) = \arg \max_{(s_1, \ldots, s_N)} - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{i} s_j \right)^2 + \sum_{i=2}^{N} \log \left((1-p) \mathbb{1}_{s_n = s_{n-1}} + \frac{p}{K-1} \mathbb{1}_{s_n \neq s_{n-1}} \right) \quad (1)$$

where $\mathbb{1}_C = 1$ if the condition C in the subscript is true and 0 otherwise. Note the the first term in this optimization problem is the "data fidelity" term which tries to minimize the squared error between the data and the model. The second term is a "structure preserving penalty" term which penalizes frequent changes in the slope sequence.

III. Algorithm

The optimization problem in (1) is efficiently solved using dynamic programming, which is similar to the Viterbi algorithm. It operates on a trellis of partial sums $\sum_{j=1}^{i} s_j$. The "reward" for proceeding from any node at depth $n - 1$ of the trellis to a subsequent node at depth n is given by the constituent terms of the summations in (1) and does not depend on future slope values. It can be seen from Fig. 1 that the structure differs from a standard Viterbi trellis because of a linear increase in the number of nodes with depth. This is because each slope value can take on K different discrete values causing the trellis to be $n(K - 1) + 1$ nodes wide at a depth n.

The unknown model parameters p and σ are estimated from the data using an alternating maximization (AM) algorithm that alternates between maximizing over the slope sequence (using the trellis) and re-estimating the parameters from the most recent fit. Due to the linearly growing trellis structure, the usual EM algorithm becomes computationally intractable even with only a small amount of data.

A. Estimating Model Parameters

In any real application, the values of model parameters p and σ are seldom known in advance and these must be estimated from the data as well. Instead of using a

computationally intensive and exact expectation-maximization (EM) algorithm, an approximate AM method shown in Fig. 2 is used instead. Starting with an initial guess for model parameters, the iterative algorithm generates a MAP optimal slope sequence using the trellis and then refines the current guess for p and σ. This process repeats till a user defined convergence criterion is satisfied.

B. Theoretical Properties

It is possible to show that the sequence of complete data likelihood values generated from the algorithm in Fig. 2 is non-decreasing.

Theorem—Let $P(y_1,\ldots, y_N, s_1,\ldots, s_N; p, \sigma)$ denote the complete data likelihood function. Then the following holds for the AM algorithm of Fig. 2:

$$P(y_1,\ldots,y_N,s_1,\ldots,s_N;p,\sigma)$$

$$\le P(y_1,\ldots,y_N,s_1^*,\ldots,s_N^*;p,\sigma)$$

$$\le P(y_1,\ldots,y_N,s_1^*,\ldots,s_N^*;p^*,\sigma^*).$$

Under an additional mild assumption of noise variance strictly bounded away from zero, it can be shown that the AM algorithm converges [8].

IV. Results

This section presents the application of the AM algorithm of Fig. 2 for two different datasets. The value of K decides the amount of discretization of the continuous range of slope values and must be specified in advance. Empirically, $K = 15$ discrete slopes was found to be sufficient for the applications discussed here. Larger values of K require greater processing time and the gain in accuracy was negligible. The iterative procedure was allowed to proceed for a maximum of 10 iterations which was used as the termination criterion.

A. Ultrasound Shear Wave Imaging

Ultrasound shear wave imaging is an imaging modality that can be used for distinguising different types of tissues based on their mechanical stiffness moduli—specifically, their shear modulus values. Stiffness variations may not be easily visible on traditional B-mode ultrasound scans. Since shear waves travel faster in stiffer media, shear wave velocity images may be used to locate stiff masses in healthy tissue which is relatively softer.

Shear wave imaging can be used for monitoring of tumor ablations in the liver. This treatment procedure involves localized heating of cancerous tissue with the help of an ablation needle inserted into the tumor. The ability to monitor the extent of ablation is crucial because it can help prevent recurrence of the tumor due to untreated cancerous cells.

The algorithm is applied to analyze data from an electrode vibration shear wave velocity imaging experiment [6], [7]. The experiment was conducted using a tissue mimicking phantom. The needle is vibrating using an actuator to produce a shear wave pulse which travels away from the needle in the ultrasound image plane. By recording many snapshots of the image plane, it is possible to measure the time of arrival of the wavefront at different locations in the image plane. The arrival time plot as a function of distance away from the needle is ideally piecewise linear—slopes correspond to the reciprocal of the shear wave velocities in different media, and change points correspond to boundaries. A shear wave velocity image from data acquired on a tissue mimicking phantom is shown in Fig. 3. Some numerical results of estimated shear wave velocities are shown in Table I. The signal to noise (SNR) dB ratios in this table were calcuated by placing three different regions of interest in the image and calculating the ratio of the mean to the standard deviation of the measured shear wave velocities.

B. Finance

In some applications, although the original data generation process may not be piecewise linear, a simple transformation leads to piecewise linear response. Quarterly real interest rate data for the US (1961-Q1 to 1968-Q3) was analyzed for an apparent piecewise constant trend to summarize different interest rate regimes over this 7 year period. The data was pre-processed by calculating the cumulative sum which was analyzed for a piecewise linear trend. The slopes of constituent segments of the resulting piecewise linear function were estimated. The result is shown in Fig. 4.

This method was compared to an existing algorithm in the `strucchange` package in R [2] and the numerical comparison is shown in Table II. The AM method detects the same number of change points at almost the same locations as the `strucchange` routine.

V. CONCLUSION

This paper presented a novel method for estimating slope values from noisy observations of a piecewise linear function with unknown locations and number of slope change points. A computationally tractable algorithm that solves a Bayesian MAP optimization problem jointly with model parameter estimation through AM iterations was analyzed. Performance of this method was evaluated through two real world applications.

Acknowledgments

This work was supported in part by NIH-NCI grants R01CA112192-S103 and R01CA112192-06.

References

1. Vieth E. Fitting piecewise linear regression functions to biological responses. Journal of Applied Physiology. 1989 Jul.67(1):390–396. [PubMed: 2759968]
2. Bai J, Perron P. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica. 1998; 66(1):47–78.
3. Kolaczyk E, Nowak R. Multiscale Generalized Linear Models for Nonparametric Function Estimation. Biometrica. 2005; 92(1):119–133.

4. Fearnhead P. Exact Bayesian curve fitting and signal segmentation. IEEE Trans. Sig. Proc. 2005; 53(6):2160–2166.
5. Rabiner L, Juang B-H. An introduction to hidden Markov models. IEEE ASSP Magazine. 1986; 3(1):4–16.
6. DeWall RJ, Varghese T, Madsen EL. Shear wave velocity imaging using transient electrode perturbation: phantom and ex vivo validation. IEEE Trans. Med. Imag. 2011; 30(3):666–678.
7. Ingle AN, Varghese T. Three-Dimensional Sheaf of Ultrasound Planes Reconstruction (SOUPR) of Ablated Volumes. IEEE Trans. Med. Imag. 2014 Aug.33(8):1677–1688.
8. Ingle AN, Bucklew JA, Sethares WA, Varghese T. Slope Estimation in Noisy Piecewise Linear Functions. Elsevier Signal Processing. 2015 Mar.108:576–588.

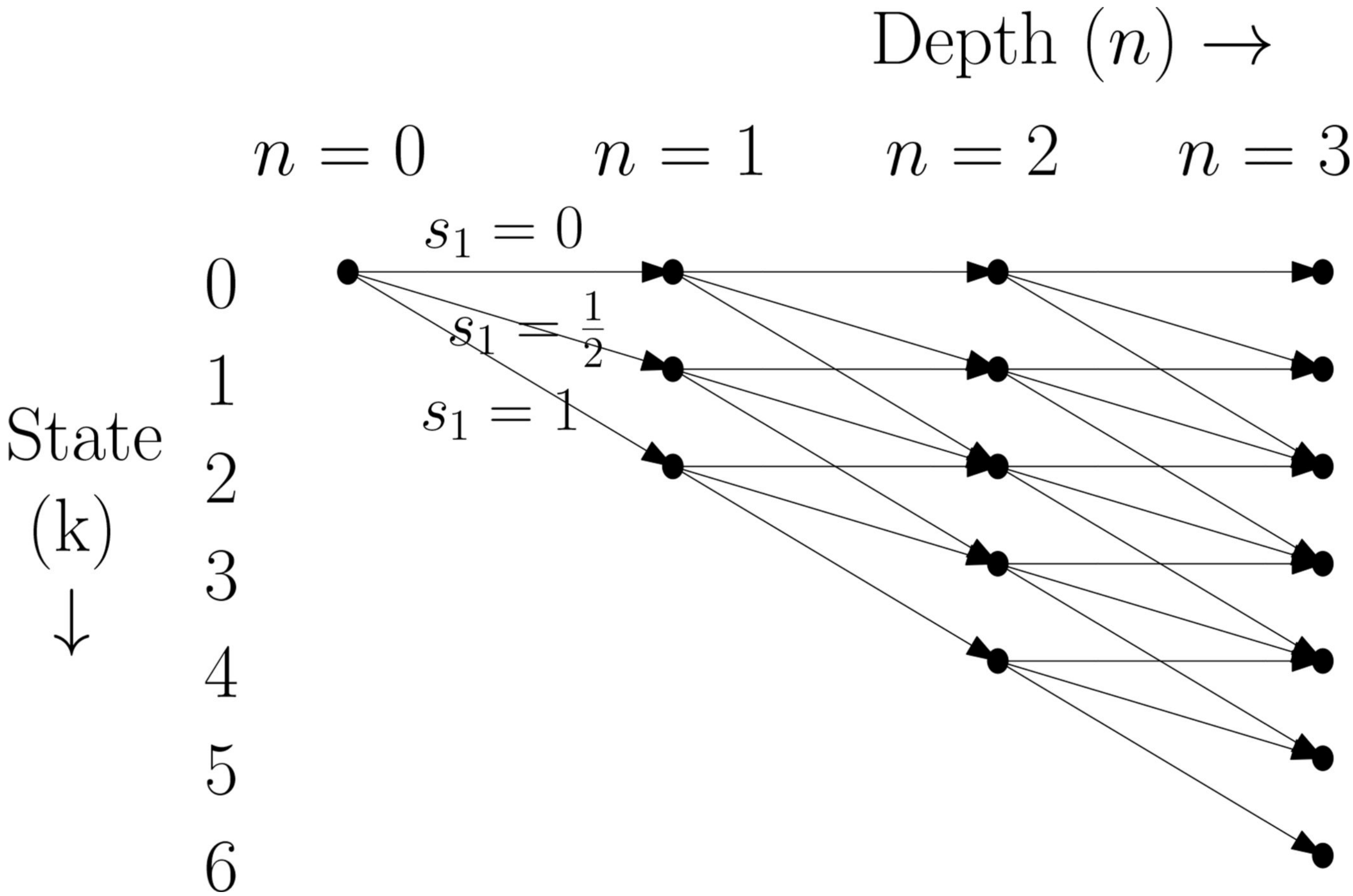
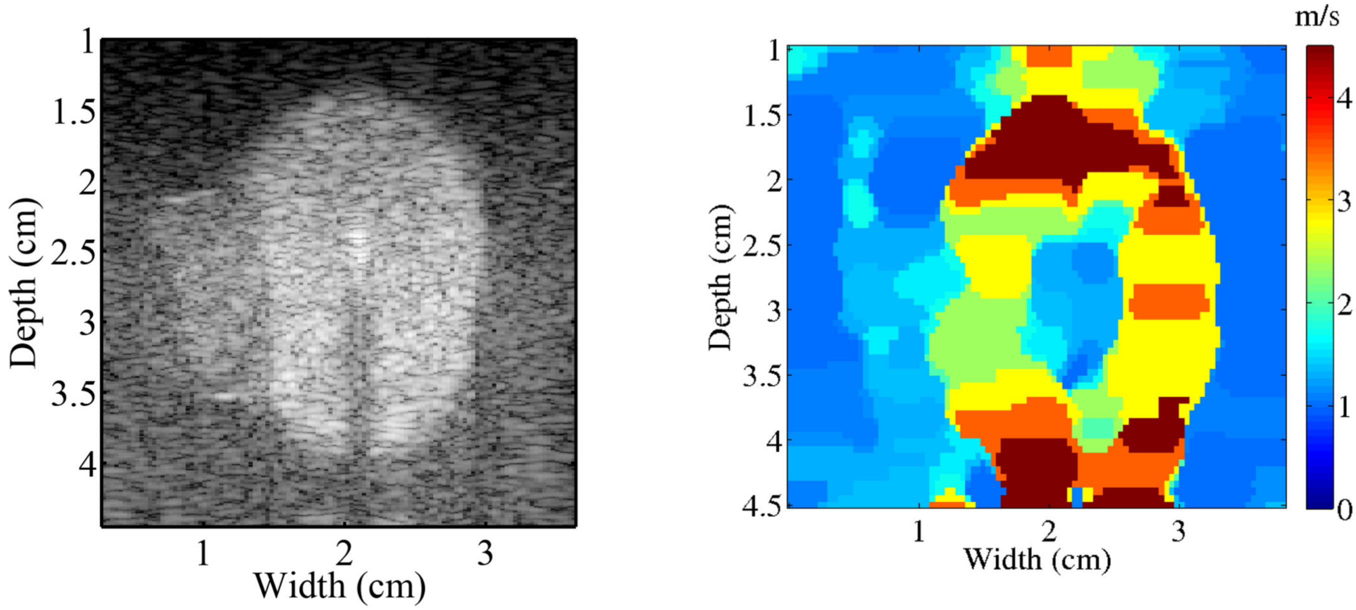


Fig. 1.
Example of a trellis used for calculating the MAP optimal sequence of slope values.

Set initial guess for (p, σ).
CONVERGED $\leftarrow$ False
while !CONVERGED **do**
    Find MAP optimal slopes $(s_1^*, \ldots, s_N^*)$
    Re-estimate $p^* = \dfrac{\#\text{slope changes}}{\#\text{data points}}$
    Re-estimate $\sigma^* = \text{stdev}(\text{noisy data} - \text{current fit})$
    $p \leftarrow p^*$, $\sigma \leftarrow \sigma^*$
    if Convergence criterion met **then**
        CONVERGED $\leftarrow$ True
    end if
end while

Fig. 2.
An alternating maximization algorithm to jointly estimate the unknown slope sequence and model parameters from noisy data.



(a) B-mode

(b) Velocity (m/s)

Fig. 3.

Results from shear wave velocity reconstruction using a piecewise linear function fitting model for time of arrival of shear wave pulse at different locations away from the needle. An ultrasound "B-mode" image of the phantom is shown in (a), and the shear wave velocity image is shown in (b). Note that the stiff structure (inclusion) can be easily visualized in the shear wave velocity image. The structure is easily discernible in B-mode too because the phantom material was specially designed with different acoustic echogenicities. In actual tissues, these stiffness variations are not easily visible in B-mode.

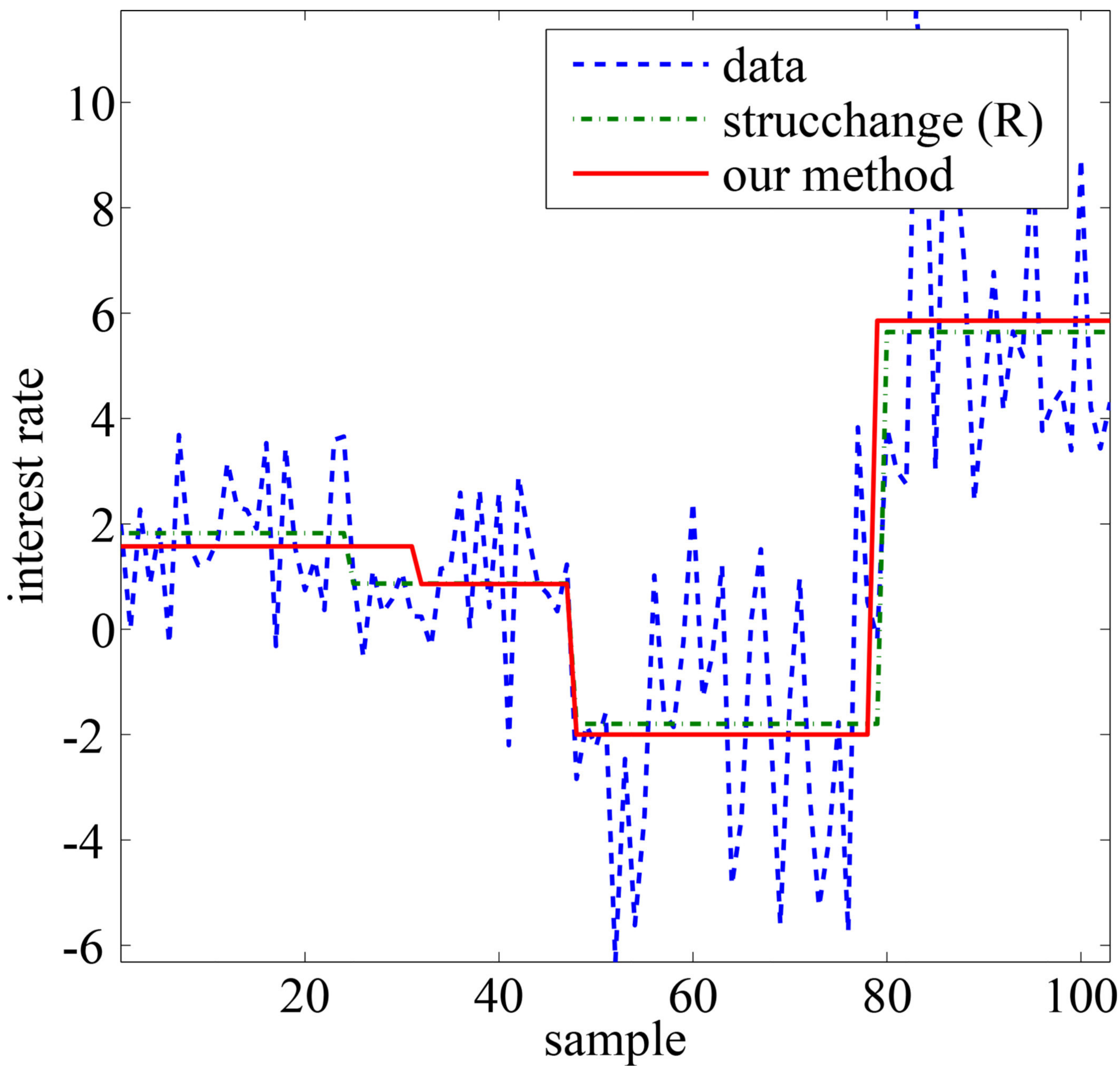


Fig. 4.

Result of piecewise constant function estimation from real interest rate data for USA between 1961–1968. For comparison, the plot also shows the result of using a routine from the `strucchange` package in R. Both algorithms show four different interest rate regimes, with only slight difference in locations of change points.

TABLE I

Quantitative estimates of shear wave slowness and velocity and signal to noise ratios for different regions of the phantom.

	SNR (dB)	Slowness (s/m)	Velocity (m/s)
background	12.7 ± 1.9	0.86 ± 0.18	1.21 ± 0.29
irregular region	17.6 ± 3.4	0.54 ± 0.11	2.03 ± 0.35
ellipsoid	12.7 ± 3.7	0.36 ± 0.10	3.09 ± 0.90

TABLE II

Numerical evaluation of the AM algorithm for financial interest rate data.

	`strucchange`	**AM algorithm**
Mean squared error	0	0.64
Squared residual	4.32	4.74
Average # breaks	3	3

The mean squared error values are obtained by treating the result of the `strucchange` method as the ground truth. Residual sum squared values are calculated with respect to the noisy data.