An Iterative Algorithm for Automatic Fitting of Continuous Piecewise Linear Models

MIGUEL A. GARCÍA
University of Alicante
Department of Applied Mathematics
Apdo. 99, E-03080 Alicante
SPAIN
Miguel.Garcia@ua.es

FRANCISCO RODRÍGUEZ
University of Alicante
Department of Applied Mathematics
Apdo. 99, E-03080 Alicante
SPAIN
F.Rodriguez@ua.es

Abstract: Continuous piecewise linear models constitute useful tools to extract the basic features about the patterns of growth in complex time series data. In this work, we present an iterative algorithm for continuous piecewise regression with automatic change-points estimation. The algorithm requires an initial guess about the number and positions of the change-points or hinges, which can be obtained with different methods, and then proceeds by iteratively adjusting these hinges by displacements similar to those of Newton algorithm for function root finding. The algorithm can be applied to high volumes of data, with very fast convergence in most cases, and also allows for sufficiently close hinges to be identified, thus reducing the number of change-points, and so resulting in models of low complexity. Examples of applications to feature extraction from remote sensing vegetation indices time series data are presented.

Key–Words: Continuous piecewise regression, Segmented regression, Multiple change-point models, Remote sensing, NDVI, MODIS.

1 Introduction

Remote sensing of vegetation dynamics, soil properties, and other ecosystem variables and indicators constitutes a key tool in ecology, agriculture and environmental studies at several temporal and spatial scales (e.g., [1, 2]). Many different techniques can be used to analyze this kind of data (e.g., [3, 4, 5]), and the development of efficient methods to identify patterns and extract features from remote sensing derived spatio-temporal data series is a key point in the applications [6].

Time series of vegetation indices, such as the normalized difference vegetation index (NDVI), are derived products from data of Earth observing systems like the Moderate Resolution Imaging Spectroradiometers (MODIS) [7] on the Terra platform. MODIS derived NDVI data are available from the year 2000, every 16 days for a global grid of pixels with a maximum resolution of 250 m, and they are just an example, as paradigmatic as it could be, of many different fields where huge amounts of time series data are produced and need to be analyzed with efficient methods capable of extracting their main features, some of which may be readily noticeable to a human observer.

A common characteristic in NDVI time series is the presence of different regions with increasing and decreasing trends, which correspond to periods of growth and decline of vegetation. This pattern is well-defined in Fig. 1, which shows data -four years period- for a set of contiguous pixels in a semiarid area of the Valencia region (southeast Spain), although it may not be so clear in the individual curves, i.e., in the data series for each pixel.

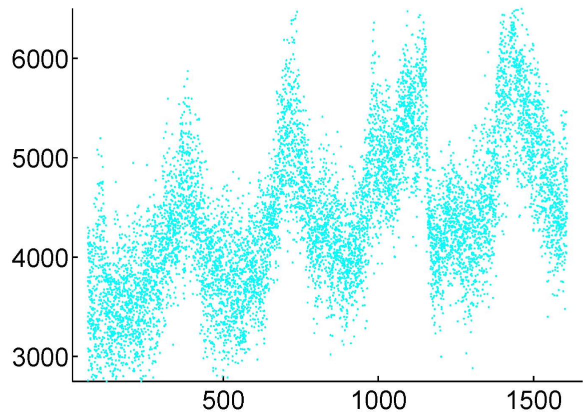

Figure 1: NDVI time series values ($\times 10000$) for an area of semiarid vegetation in southeast Spain. Time values (abscissas) are number of days, starting from 01/01/2000.

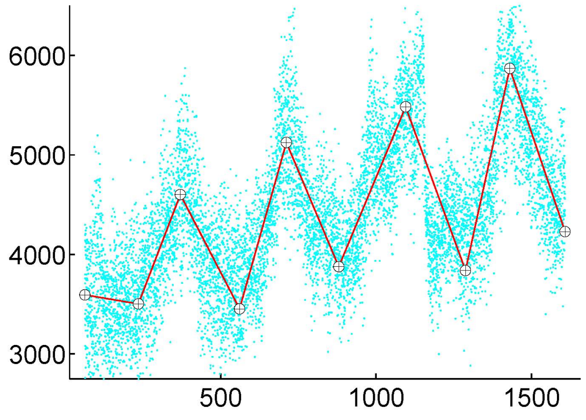

Figure 2: NDVI time series values from Fig. 1 with a continuous piecewise linear model fitted to the data.

A simplified model of the functional dependence suggested by this type of data is shown in Fig. 2, a continuous polygonal which characterizes the sequence of growth rate changes, providing the position of the change-points and the slopes of the linear segments.

The problem of fitting a continuous piecewise linear model to a series of data is referred to as piecewise [8] or segmented [9] regression, linear regression with multiple structural changes [10] or regimes, or in the case of two segments as broken-line or two-phase [11, 12] regression. In the classical statistical framework, this problem has been tackled as a particular case of nonlinear regression [13], or with specific approaches aimed at minimizing the sum of squares of errors, yielding least squares estimates of the parameters, or maximum likelihood estimates in the case of independent identically distributed normal errors [14, 15, 16] or under particular hypothesis on the error structure [9, 17].

When the number and positions of the change-points are known, the estimation of the model is straightforward. Segmented linear models with change-point estimation without the continuity requirement are special cases of model trees [18, 19], where induction methods such as Quinlan's M5 [20] are well-designed for predictive performance with many regressors. The restrictions imposed by the continuity condition, and the discontinuities in the derivative implied by a polygonal model, make the estimation of the model by minimizing some form of risk function much more difficult. Some authors use approximate smooth models to avoid these problems [21, 22], while the more direct algorithms are mainly based on grid search [16] or some form of greedy exploration of the possible change-points [15].

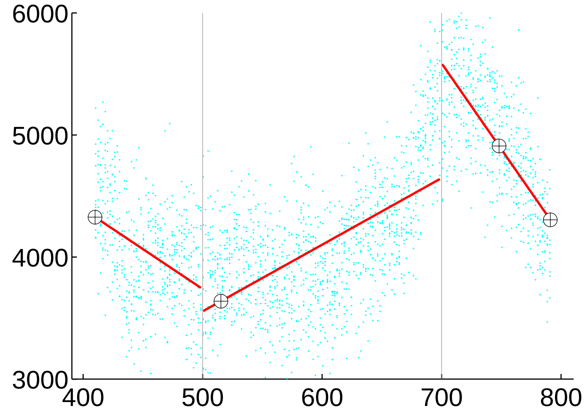

Figure 3: A section of the NDVI data from Fig. 1, showing a step of the fitting process, where for the current selection of change-points (indicated by vertical lines) a line is fitted in each segment (bars) and their intersections obtained as new change-points (crossed circles).

Other computationally intensive approaches include bayesian [23] and fuzzy methods [24].

Piecewise linear models are usually approximations to complex real phenomena, that allow to extract the basic features of the data, and so find application in many different fields, as in economy [10], ecology [25] or cancer research [26]. The objective of this work is to efficiently fit a continuous polygonal model to large datasets, with a computational approach that does not intend to yield a global optimum for some measure of adjustment, but to capture in an objective manner the main trends of the data, providing estimates for the relevant parameters in the problem considered.

2 Description of the method

The method proposed, which we denote with the acronym HANDFIT, standing for Hinges Adjustment by Newton-like Displacements FIT, consists of two phases. First, an initial guess about the number and positions of the change-points or hinges is made, for which various alternatives suited for different particular problem are considered. Then, an iterative process to displace these hinges, analogous to Newton method for root finding, is applied. Several parameters of the algorithm can be adjusted so that the type of features that are of main interest be extracted, although a completely automated functioning is also possible.

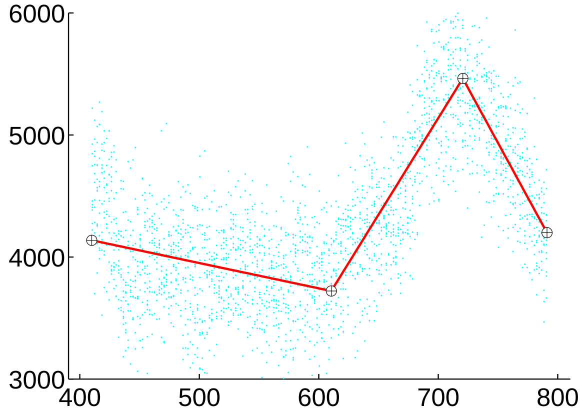

Figure 4: Continuous piecewise linear model fitted to the data from Fig. 3, given by the final convergent solution of the algorithm.

2.1 Iterative Adjustment of the Hinges

Assume we have an initial estimation of the abscissas of the increasing sequence of change-points, $\{x_i\}_{i=1\ldots n}$, in a fixed domain with endpoints $a = x_0$ and $b = x_{n+1}$. The iterative algorithm proceeds in two steps. Firstly, for each interval $[x_i, x_{i+1}]$ a line segment is adjusted to the data by ordinary least squares, although any other fitting method could be used as well (Fig. 3). Secondly, the intersection points between consecutive segments are computed, so that their abscissas define the new change-points, and the process is repeated until convergence is achieved (Fig. 4). The stopping criterion defining convergence is simply that a fixed threshold for the displacements of the hinges not be exceeded, and for sound choices of the threshold it is usually reached after only a few iterations.

The second step in the algorithm is essentially a Newton iteration, as the new points are the intersection of affine varieties, which in the simplest Newton method, the Newton-Raphson algorithm to compute zeros of real functions, are the tangent line and the X-axis. Newton-like algorithms are in most cases very fast, but it is well-known that these type of methods may produce abrupt jumps, which in our problem could yield non-admissible values when the ordering of the hinges is not preserved (Fig. 5).

To tackle this eventuality, if $\widehat{x}_i$ is to be the abscissa of the new i-hinge, the relative increments,

$$\frac{\widehat{x}_i - x_i}{x_{i+1} - x_i}, \quad \widehat{x}_i > x_i; \qquad \frac{\widehat{x}_i - x_i}{x_i - x_{i-1}}, \quad \widehat{x}_i < x_i;$$

are always transformed using a sigmoid function (Fig. 6), with a dampening coefficient that can be adjusted to successfully avoid any problems. This cautionary safeguard has the cost of increasing the number of iterations, but it results in a more robust algorithm, with still a fast speed of convergence for most data. It should be clear, however, that convergence can not be guaranteed for any arbitrary dataset, as no sound continuous linear model can be expected to fit a data series resulting from a process with intrinsic discontinuities.

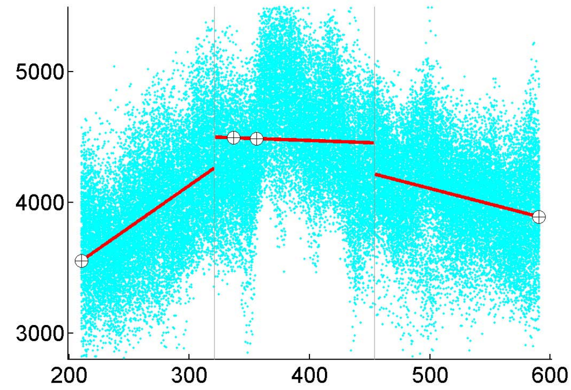

Figure 5: Example of data where non-admissible values would be produced at an iteration step.

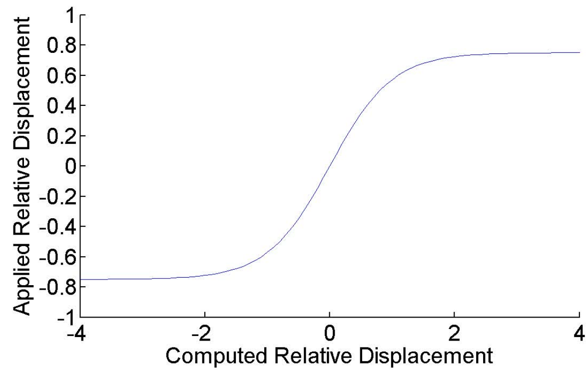

Figure 6: Example of sigmoid function used to correct for possible jumps during the iteration process.

Depending on the data and the number and positions of the starting points, two consecutive hinges might get close enough to consider that they should be identified, and this is what the algorithm does when a proximity threshold is crossed, which can also be automatically defined in terms of the minimum number of different data points in a segment for the least square adjustment be considered sound. Thus, the algorithm can automatically correct to a certain extent an excess of hinges in the initial set, and hence results

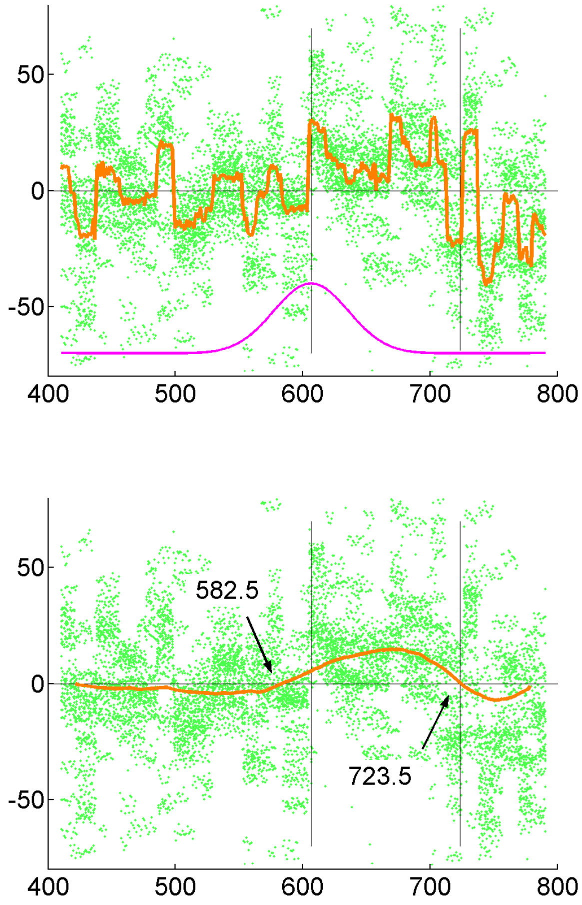

Figure 7: Up: Local slopes for the NDVI data series in Fig. 3, showing their medians and the gaussian kernel used to filter them. Down: Smoothed curve of the medians. The values of the abscissas corresponding to zeros of the filtered medians are displayed, and the change-points of the final solution indicated by vertical lines.

in models that are not of much higher complexity than that suggested by the data.

2.2 Initial Selection of the Change-Points

Different strategies can be applied to select the initial set of change-points, according to the type of data and the particular features of interest, although the iterative step leads from many distinct reasonable elections of the starting points to the same final convergent solution, as will be discussed in the next section.

For the data in Fig. 1, where the pattern is essentially a sequence of periods with alternate growing and decay behavior, a plausible election would be those points where the mean slope of the curves changes in sign, i.e., where it passes trough zero.

In Fig. 7 (up), the cloud of slopes for 120 similar curves is displayed, showing their medians as robust estimations of the slopes. Although there are many points where the curve of the medians changes in sign, in the three sections marked in the graph, which correspond to the segments of a final convergent fitted model, the second one consists essentially in positive values, whereas in the other two the values of the medians are mostly negative. Considering all the zeros in this curve as initial change-points would produce a model of very high complexity, despite the limited reduction in the number of hinges that the iteration step of the algorithm is capable of perform. A more reasonable election is obtained filtering the medians using a gaussian kernel filter, as the one shown in the same figure, and working with the curve of the smoothed medians (Fig. 7 down). The zeros of the filtered median give values close to the final iterated solution obtained from a subjective selection of the starting points, and very similar results are obtained if the data are averaged and the smoothed slope of the mean curve is used instead.

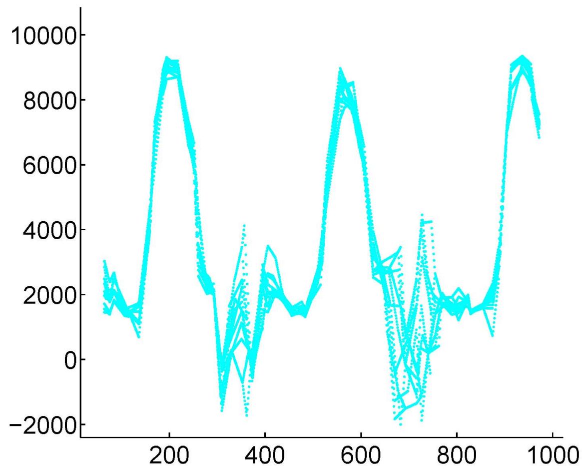

Figure 8: NDVI time series values ($\times 10000$) for an area of rice crops in southeast Spain. Time values (abscissas) are number of days, starting from 01/01/2000.

Consider, however, the NDVI data presented in Fig. 8, corresponding to rice crops, also in the Valencia region. Here the dynamics of the vegetation is more complex, as besides the clearly defined evolution of the crop, there are also other periods that correspond to phases of harvesting, growing of natural vegetation and preparation of the fields for the new season. All the pixels show similar and synchronized behaviors, since they are subjected to the same labors at specific moments in time, and so the curves are much more better defined than those in Fig. 1, allowing for a more detailed description than just the incresing/decreasing pattern.

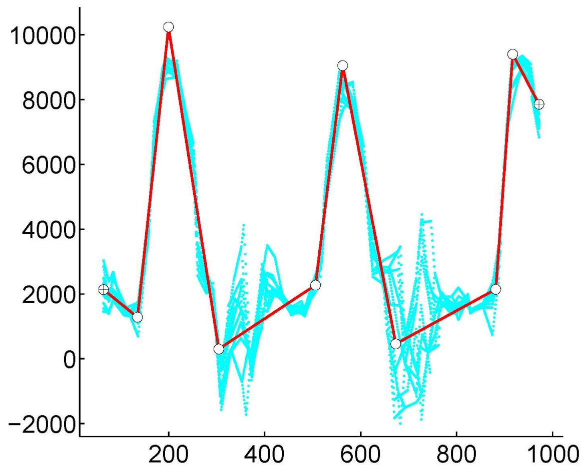

Figure 9: NDVI time series values from Fig. 8, with a continuous piecewise linear model fitted to the data. Note the variations in the slopes of contiguous segments, which do not restrict to a positive/negative sequence.

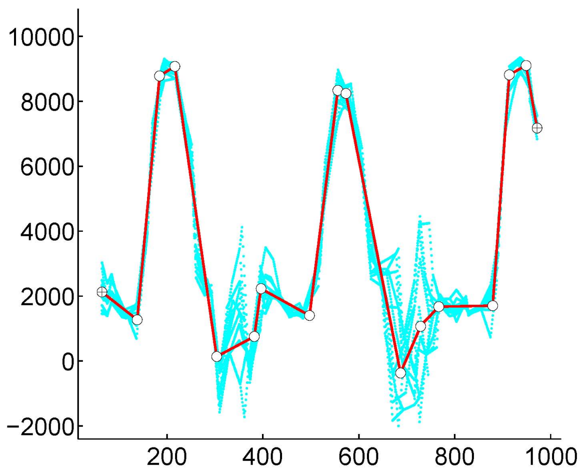

Figure 10: NDVI time series values from Fig. 8, with a continuous piecewise linear models of higher complexity than that of Fig. 9 fitted to the data.

In Fig. 9, a simple model that represents fairly well the periods of growth and decay of the crops is fitted to the data. In Fig. 10, a different model of higher level of complexity, in terms of the number of change-points considered, is presented. This last model reflects better than the previous one the transitions between the periods of growth and decay, as well as the phases between consecutive cropping seasons. Both models are final equilibrium solutions of the iterative algorithm for different elections of the number and positions of the initial change-points. The rationale for deciding between these two models depends on the kind of features that we are interested in, either the basic characterization of the crop dynamics, as in Fig. 9, or a more detailed description of the whole vegetation dynamics as in Fig. 10. Hence, this decision must be set by the analyst according to the objectives of the study, although it can be incorporated into the algorithm, either in an explicit or implicit way, trough the the method used for the selection of the initial points and with the setting of the different parameters modulating the outcome of the algorithm, as window sizes of the smoothing filters or thresholds levels.

In any case, in this more general context the previously discussed method relying on the change in sign of the slopes is clearly inappropriate, and the selection of the starting points could instead be based on the distribution of the curvatures (Fig. 11). In Fig. 11 (up), the cloud of curvatures for the data in Fig. 8 is shown, where it has been obtained using a moving window of suitable amplitude, to compute the curvatures for each curve and abscissa fitting a second degree poly-

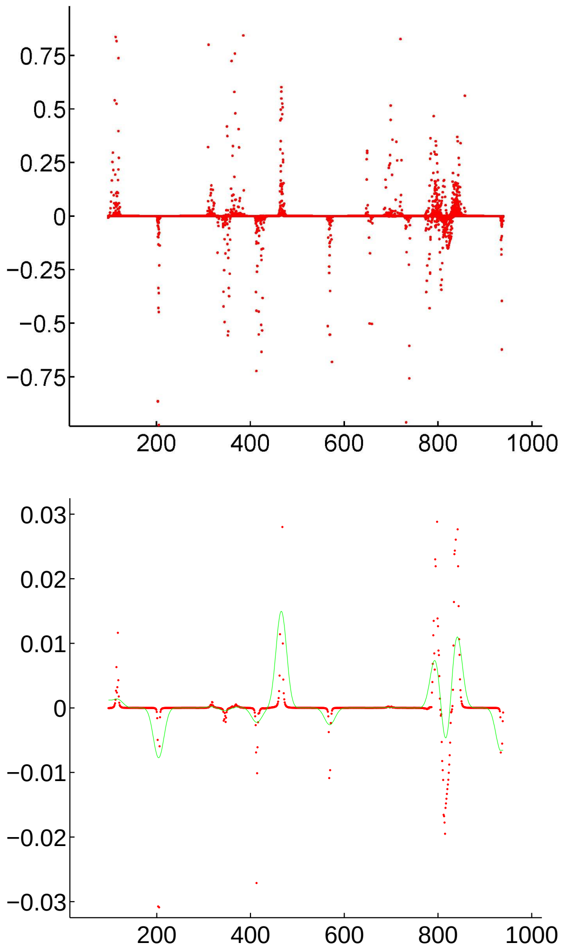

Figure 11: Up: Local curvatures for the data in Fig. 8. Down: Medians of the distribution of local curvatures (dots), and continuous smoothed median after filtering with a gaussian kernel.

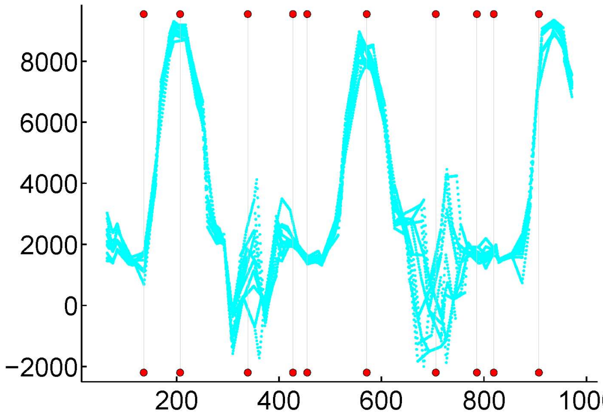

Figure 12: Positions of the ten most extreme values of the local curvatures for the NDVI data in Fig. 8.

nomial to the points inside the window. The medians of the curvatures have been smoothed with a gaussian kernel, as shown in Fig. 11 (down), and the zeros of the derivative of the filtered curvature function, corresponding to the most extreme values above some threshold, have been selected (Fig. 12).

For equally spaced data, curvatures can be computed in an very efficient way using a Savitzki-Golay type method [27] to adjust the quadratic polynomials for each position of the moving window. In case that for the specific data quadratic polynomials were not flexible enough to detect the zones of interest, polynomials of higher degree could be used, and the computational effort would be comparable if a Savitzki-Golay strategy could be employed, i.e., if abscissas were uniformly spaced.

A robust and computationally efficient alternative, to avoid computing curvatures through fitting of second or higher degree polynomials, is to consider the angles between lines adjusted to contiguous sets of points (Fig. 13). Using a double-sized moving window, two lines are fitted to the points at the left and the right of each abscissa, and the angles between these lines are computed (Fig. 13 up). Then, they can be post-processed as in the previous method (Fig. 13 down), and the most extreme values selected with some threshold criteria (Fig. 14).

2.3 Sensitivity to initial conditions

Although a variety of methods can be used to determine the starting set of hinges, we note that many different selections of the initial points lead to the same final result, which is one of a very restricted set, that of the fixed points for the Newton-like iteration algorithm. To illustrate this behavior, for the data presented in Fig. 3, where two change-points seem to pro-

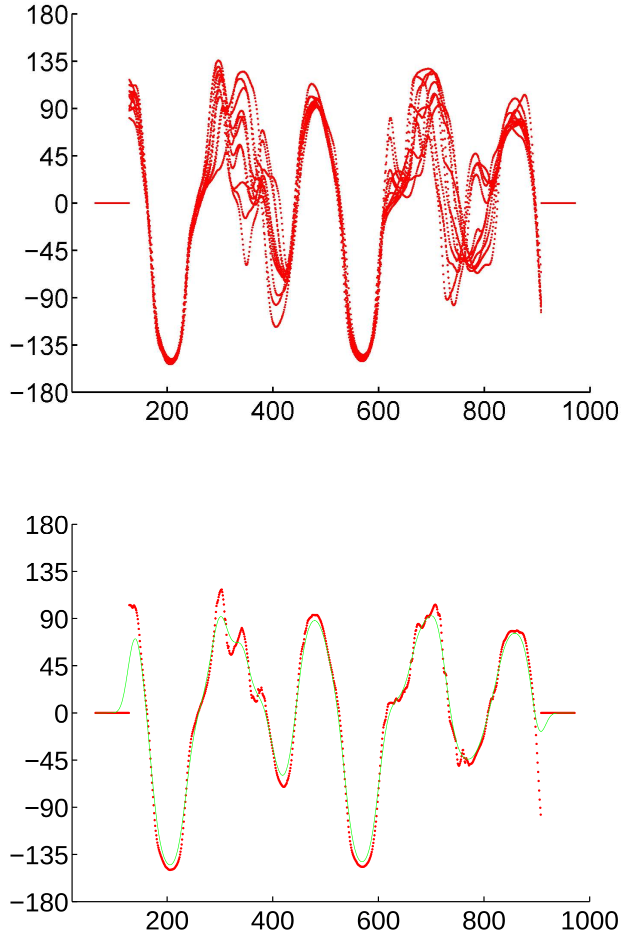

Figure 13: Up: Local angles for the data in Fig. 8. Down: Medians of the distribution of local angles (dots), and continuous smoothed median after filtering with a gaussian kernel.

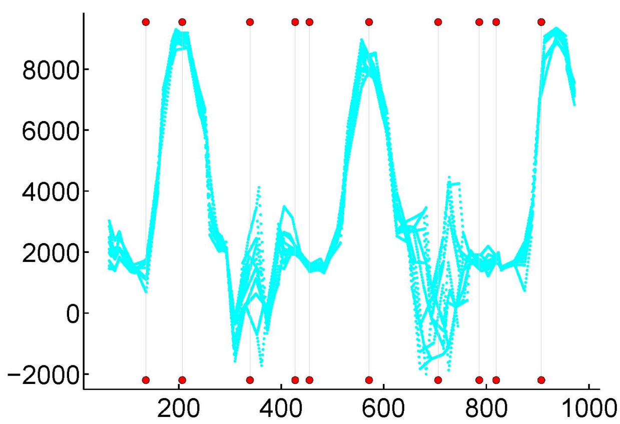

Figure 14: Positions of the ten most extreme values of the local angles for the NDVI data in Fig. 8.

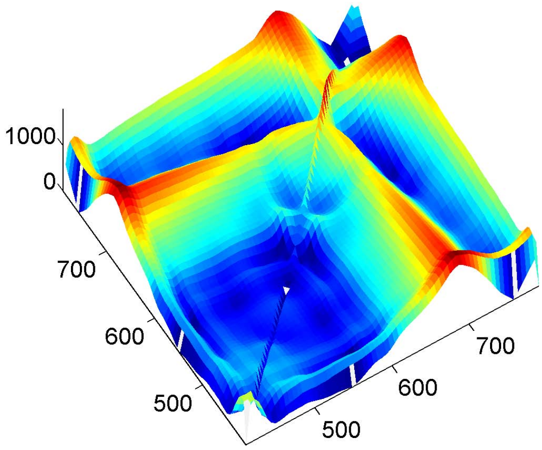

Figure 15: Index of discontinuity (root mean squares of jumps between contiguous segments) as a function of the positions of the two change-points when fitting a piecewise linear model to the data in Fig. 3.

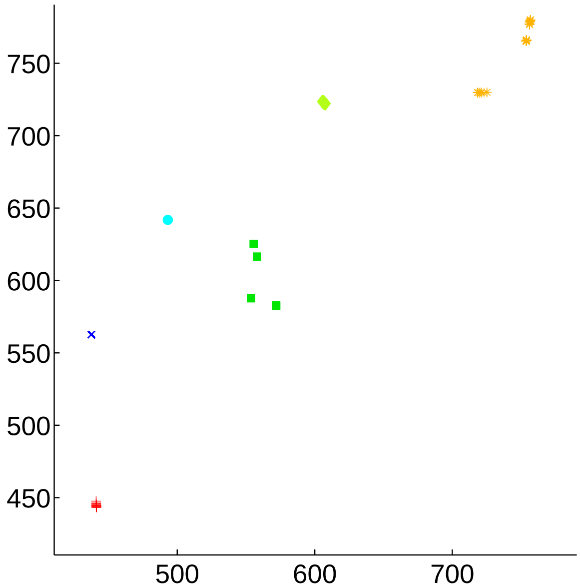

Figure 16: Abscissas of the two change-points in the set of the final iterated solutions for the data in Fig. 3, resulting from an exhaustive search of the initial positions for the change-points.

vide a sound model, an exhaustive search on the initial positions of the two change-points was performed.

For each position of the two change-points, when a simple piecewise linear model is fitted, i.e., when each segment is optimally fitted to their data by ordinary least squares without requiring the continuity condition, a measure of the magnitude of the jumps across contiguous segments gives an idea of the regions that can sustain a continuous model (Fig. 15), the zeros of this function being the points sought by the algorithm. If we apply the iterative algorithm from any pair of these initial points, a convergent solution is eventually reached, and the set of the positions of the change-points in the possible final solutions is shown in Fig. 16. The regions defined by the set of initial points that result in the same final solution are presented in Fig. 17. The larger central region in this figure, comprising more than half of the possible elections for the initial change-points, result in the solution shown in Fig. 4, which intuitively provides a sound model for the data. In fact, this is also the two change-points model with the minimum global error of fit (Fig. 18).

A different question that can be raised is the sensitivity of the algorithm to small variations in the original data. In real applications, data are measured with a certain degree of error, and any feature extraction algorithm should not give much different outcomes

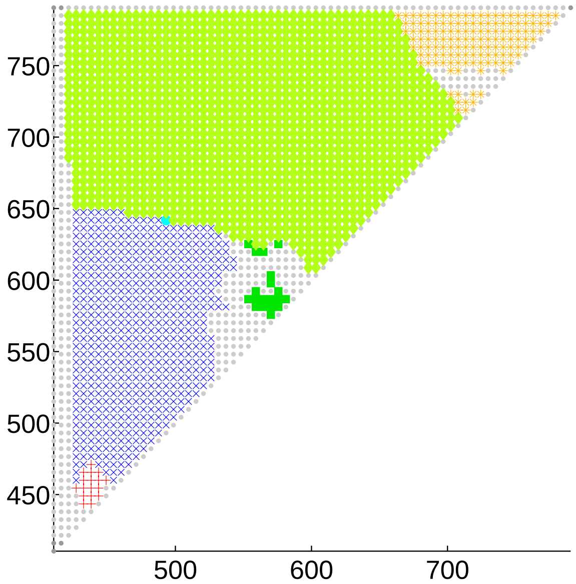

Figure 17: For the data in Fig. 3, regions of initial positions for the change-points that result in the same final solution.

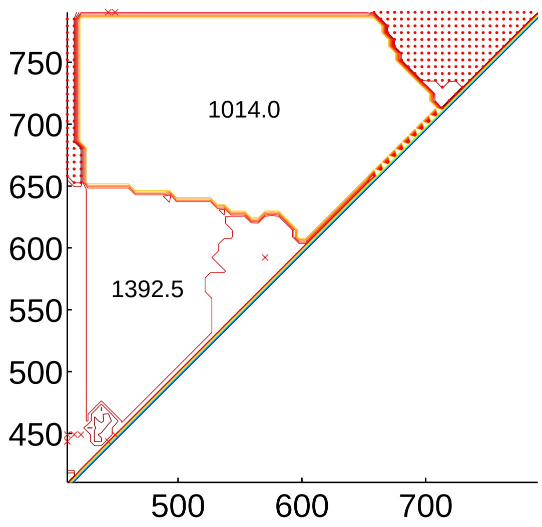

Figure 18: Global errors of fit for the models corresponding to the different final solutions in Fig. 16. The values displayed correspond to the models with the two lowest global errors.

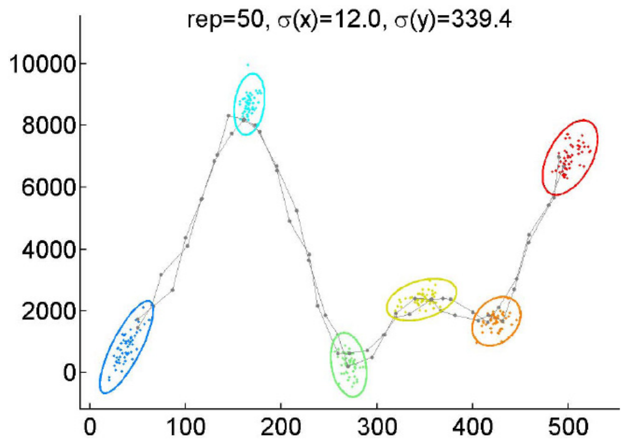

Figure 19: Sensitivity of the final positions of the hinges to random perturbation of the data. Two sections of data series from Fig. 8 were perturbed, adding independent gaussian noise to the x and y coordinates of the points, with standard deviations as indicated in the figure, producing 50 replications. For each cluster of the final change-points, elipsoids determined by three standard deviations of their distributions are shown.

for close data inputs. To explore the robustness of the algorithm to random perturbations of the data, we selected a section of two curves from data in Fig. 8 and added gaussian noise to the positions of the data points. The final positions of the change-points given by the algorithm were consistent, determining models exhibiting similar behaviors (Fig. 19).

3 Discussion

The algorithms presented in this work provide a computationally efficient method to fit continuous piecewise linear models to data series when the number of points is high and many change-points have to be considered, and can be an alternative to methods based on exhaustive or grid searches aimed at minimizing some global risk function.

Our objective in fitting these kinds of models is to extract the main features, in terms of different growth regimes, present in the data, and in this context it is clear that some kind of *a priori* information, either explicit or implicit, has to be used to define what a trait of interest is. Consider, for instance, the data presented in Fig. 8. As shown in Fig. 9 and Fig. 10, models of different levels of complexity can be fitted, depending wether the interest lies essentially in the succession of grothw/decline seasons or a more detailed description is sought.

Our primary envisaged application for these type of models is in the analysis of remote sensing vegetation data, as exemplified along the paper, although there are many other fields where continuous piecewise linear models are sound models and can provide basic description of the patterns of growth exhibited by the data, and where efficient algorithms are needed to cope with high volumes of data. However, it should be kept in mind that no algorithm for continuous piecewise regression can be successful when the data does not reflect the continuity properties needed in these models (Fig. 20).

There are several options and parameters in the algorithms that can be set to fine tune the method, and obtain the type of model more adequate for the data in consideration. Besides the different options for the selection of the initial points, the size of the moving windows used to compute local curvatures or angles, the shape of the smoothing kernels and the values of the thresholds to select the most extreme values determine the number and positions of the initial change-points. Nevertheless, any sound choice for these parameters would lead to very similar sets of starting points, as exemplified by comparing Fig. 12 and Fig. 14. Moreover, as discussed in the previous section, the final set of change-points are the points of equilibrium of the

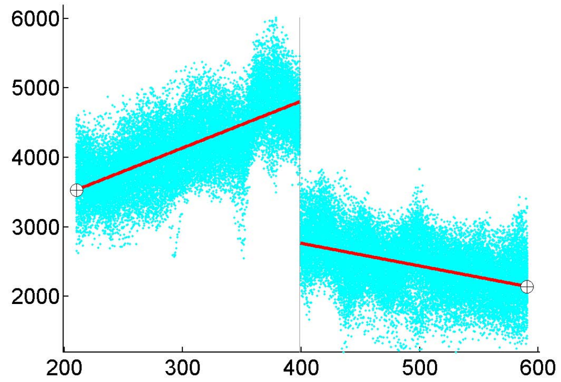

Figure 20: Example of simulated data produced by a discontinuous model.

iteration process, and so there is no need for an intensive effort to optimally determine the positions of the initial points, as most of them will usually lead to the same final solution.

Finally, it should be clear that the method can be run in a completely automated way. The choice between different models with the same number of change-points can be based on the global error of fit, while some suitable model selection criteria [28] taking into account the number of parameters of the model, such as AIC [29] or BIC [30], can be employed to select between models with different number of change-points.

Acknowledgements: This work has been partly funded by grants from Valencia Regional Government (GVPRE/2008/310, Conselleria de Educación, Generalitat Valenciana) and University of Alicante (VIGROB-162).

References:

[1] M.H. Ismail and K. Jusoff, Estimating Forest Area using Remote Sensing and Regression Estimator, *WSEAS Trans. Signal Processing* 3(1), 2007, pp. 88–94.

[2] H. Diofantos, T. Leonidas, H. Marinos, V. Photos, P. Papadopoulos, C. Chris and L. Athina, Satellite Remote Sensing for Water Quality Assessment and Monitoring - an Overview on Current Concepts, Deficits and Future Tasks, *WSEAS Trans. Signal Processing* 3(1), 2007, pp. 67–73.

[3] A. Kulkarni and S. Mccaslin, Knowledge Discovery from Satellite Images, *WSEAS Trans. Signal Processing* 2(11), 2006, pp. 1523–1530.

[4] V. Barrile and G. Bilotta, Object-oriented analysis applied to high resolution satellite data, *WSEAS Trans. Signal Processing* 4(3), 2008, pp. 68–75.

[5] D. Hadjimitsis, I. Evangelou, A. Retalis, A. Lazakidou and C. Clayton, Classification of Satellite Images for Land-Cover Changes using an Unsupervised Neural Network Algorithm, *WSEAS Trans. Signal Processing* 1(2), 2005, pp. 155–162.

[6] L. Bruzzone, P.C. Smits and J.C. Tilton, Foreword special issue on analysis of multitemporal remote sensing images, *IEEE Trans. Geosc. Rem. Sens.* 41, 2003, pp. 2419–2420.

[7] V. V. Salomonson, W. L. Barnes, W. P. Maymon, H. Montgomery and H. Ostrow, MODIS: Advanced facility instrument for studies of the Earth as a system, *IEEE Trans. Geosc. Rem. Sens.* 35, 1989, pp. 145–153.

[8] V. E. McGee and W. T. Carleton, Piecewise Regression, *J. Amer. Statist. Assoc.* 65, 1970, pp. 1109–1124.

[9] H. P. Piepho and J. O. Ogutu, Inference for the Break Point in Segmented Regression with Application to Longitudinal Data, *Biom. J.* 5, 2003, pp. 591–601.

[10] J. Bai and P. Perron, Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 1998, pp. 47–78.

[11] D. V. Hinkley, Inference about the intersection in two-phase regression, *Biometrika* 56, 1969, pp. 495–504.

[12] D. V. Hinkley, Inference in Two-Phase Regression, *J. Amer. Statist. Assoc.* 66, 1971, pp. 736–743.

[13] G. A. F. Sebert and C. J. Wild, *Nonlinear Regression*, Wiley, New York 1989.

[14] D. M. Hawkins, Point estimation of the parameters of piecewise regression models, *Appl. Statist.* 25, 1976, pp. 51–57.

[15] D. J. Hudson, Fitting segmented curves whose join points have to be estimated, *J. Amer. Statist. Assoc.* 61, 1966, pp. 1097–1129.

[16] P. M. Lerman, Fitting Segmented Regression Models by Grid Search, *Appl. Statist.* 29, 1980, pp. 77–84.

[17] T. -S. Lee, Estimating coefficients of two-phase linear regression model with autocorrelated errors, *Statist. Probab. Lett.* 18, 1993, pp. 113–120.

[18] A. Ciampi, Generalized regression trees, *Comput. Statist. Data Anal.* 12, 1991, pp. 57–78.

[19] I. H. Witten and E. Frank, *Data Mining*, Elsevier, San Francisco 2005.

[20] J. R. Quinlan, in *Proceedings AI'92*, p. 343, World Scientific, Singapore 1992.

[21] A. Tishler and I. Zang, A New Maximum Likelihood Algorithm for Piecewise Regression, *J. Amer. Statist. Assoc.* 76, 1981, pp. 980–987.

[22] G. Chiu, R. Lockhart and R. Routledge, Bent-Cable Regression Theory and Applications, *J. Amer. Statist. Assoc.* 101, 2006, pp. 542–553.

[23] B. P. Carlin, A. E. Gelfand and A. F. M. Smith, Hierarchical Bayesian analysis of change point problems, *Appl. Statist.* 41, 1992, pp. 389–405.

[24] J. -R. Yu, G.- H. Tzeng and H.- L. Li, General fuzzy piecewise regression analysis with automatic change-point detection, *Fuzzy Sets and Systems* 119, 2001, pp. 247–257.

[25] J. D. Toms and M. L. Lesperance, Piecewise regression: A tool for identifying ecological thresholds, *Ecology* 84, 2003, pp. 2034–2041.

[26] B. Yu, M. J. Barrett, H. -J. Kim and E. J. Feuer, Estimating joinpoints in continuous time scale for multiple change-point models, *Comput. Statist. Data Anal.* 51, 2007, pp. 2420–2427.

[27] A. Savitzky and M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.* 36, 1964, pp. 1627–1639.

[28] K.P. Burham and D.R. Anderson, *Model Selection and Inference. A Practical Information-Theoretic Approach*, Springer-Verlag, New York, 1998.

[29] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov and F. Csáki (eds.), *2nd International Symposium on Information Theory*, Akadémia Kiado, Budapest, 1973, pp. 267–281.

[30] G. Schwarz, Estimating the dimension of a model, *Ann. Statist.* 6, 1978, pp. 461464.