Performance des méthodes de segmentation appliquées au continuum fluvial à l'échelle du réseau

Performance of Segmentation Methods Applied to River continuum at Network scale

T. De Almeida[1] ; L. Vaudor[1] ; S. Dray[2] and H. Piégay[1]

[1]Université de Lyon, ENS de Lyon, CNRS, UMR 5600, Lyon, France

[2]Université de Lyon, Université Lyon 1, CNRS, UMR 5558, Villeurbanne, France

RÉSUMÉ

La segmentation des rivières est une étape cruciale dans leur caractérisation pour détecter les discontinuités et délimiter des tronçons homogènes. Avec la disponibilité croissante de données longitudinales pour décrire les rivières, le besoin d'une segmentation univariée automatique en tant qu'outil de synthèse et d'analyse est devenu essentiel. De nombreuses méthodes existent, bien qu'elles ne soient généralement pas conçues spécifiquement pour segmenter une rivière. Nous avons mené une revue des méthodes existantes et sélectionné un ensemble de techniques potentiellement utilisables selon plusieurs critères pratiques. Nous avons comparé les méthodes sélectionnées à l'aide de données simulées et évalué leurs performances en fonction de critères de qualité, de robustesse et d'efficacité. Nous avons conclu qu'aucune technique de segmentation ne surpasse toutes les autres dans toutes les circonstances et avons décrit les conditions spécifiques dans lesquelles chaque méthode est la plus performante. Enfin, nous proposons des conseils pratiques pour choisir la technique la plus adaptée en fonction de la distribution des données et des objectifs de l'étude.

ABSTRACT

River segmentation is a crucial step in river characterization to detect discontinuities and delineate homogeneous reaches. With the increasing availability of longitudinal data to describe rivers, the need for automatic univariate segmentation as summary and analysis tool has become critical. Many methods exist, although they are not typically designed to segment a river. We conducted a review of existing methods and selected a set of potentially available techniques based on several practical criteria. We compared the selected methods using simulated data and evaluated the performances of methods based on quality, robustness and efficiency criteria. We concluded that no segmentation technique outperforms all others in all circumstances and described under which specific conditions each method best performs. Lastly, we provide practical advice for selecting the most suitable technique according to the data distribution and the objectives of the study.

Mots-clés : Discontinuité, Méthodes automatiques, Réseau fluvial, Segmentation, Tronçon de rivière

Keywords: Automatic methods, Discontinuity, Network River, River Reach, Segmentation

1 INTRODUCTION

The concept of longitudinal river continuum describes a downstream continuous gradient of the morphological characters, such as decreasing slope and grain size. River segmentation is thus an initial and crucial step in geomorphic analysis and river classification, enabling the interpretation of geomorphic processes through the identification of homogeneous river reaches (Brierley and Fryirs, 2005). Longitudinal discontinuities define reach boundaries and enhance understanding of fluvial dynamics (Fryirs and Brierley, 2005). These discontinuities can be abrupt, marking sudden changes in morphology like a rocky riffle threshold. They can also correspond to gradual transitions between two distinct geomorphic river reaches. Longitudinal discontinuities result from complex interactions among several structural or functional controlling factors (Benda et al., 2004).

In such a context, segmentation is based on single or multiple variables to delineate geomorphologically homogeneous river reaches (Alber and Piégay, 2011). Segmentation involves dividing a longitudinal signal into homogeneous segments separated by change points defining distributional changes (Basseville and Nikiforov, 1993). Many segmentation methods exist such as visual approaches using expert criteria, heuristic algorithms, homogeneity tests, clustering methods and algorithms. Each method has unique characteristics, and an effectiveness which varies according to data distribution and the type of information required (Basseville and Nikiforov, 1993).

With recent advancements in the longitudinal characterization of rivers, the need for segmentation, often univariate, is increasing, particularly favoring the automatic detection of homogeneous reaches. Reviews and comparisons of segmentation or change point methods exist for general applications and specifically in fluvial continuum studies (Parker et al., 2012 ; Leviandier et al., 2012 ; Martínez-Fernández et al., 2016). However, recent methods have since then been published and there is a need to evaluate them and assess their potential for any use in river continuum segmentation according to specific criteria. Moreover, some important criteria for large dataset analysis, such as computational efficiency, were not previously considered. There is a need to identify high-quality, robust and efficient segmentation methods suitable for application to large-scale data like river networks. We define the quality of a method based on its ability to accurately detect homogeneous segments, as discussed by Lemire (2007). We consider the robustness of methods subject to variations in data distribution and the efficiency depends on a minimum computational time.

2 METHODS

2.1 Selecting a set of segmentation techniques

We have selected a set of techniques based on these criteria:

- Univariate, fully automated, and implemented in a programming language
- Compatibility with continuous spatial data without requiring prior knowledge of the number of breakpoints
- Efficiency i.e. computational time suitable for large dataset

In this study, we focus exclusively on seven univariate methods: BEAST, Binary Segmentation, CPM, Jumpoint, Hubert, PELT, and Segment Neighbourhood.

2.2 Comparison on artificial signal

We simulated univariate data series based on reference models which record the actual locations of the breakpoints. The reference models are a series of plateaus with varying value and length, each representing a homogeneous segment. The simulated series are the reference models plus some noise. We vary several signal generation parameters to decrease the « segmentability » of the signal i.e. the ability to segment correctly. The key parameters are:

- Number of segments (K)
- Mean Segment length (S)
- Ratio between inter-segment and intra-segment standard variation (r). It influences the ability to segment correctly by adjusting the noise levels and mean values of each segment. A low ratio results in a simpler-to-segment series with clearer distinctions between segments, while a high ratio leads to a series with increased overlap between segments.
- Distance coefficient between segment (δ)

For each combination of parameters, we perform 1,000 simulation iterations to obtain statistically significant estimates of performance. We apply the techniques using two types of signal distributions: normal distributions with homoscedasticity and gamma distributions with heteroscedasticity, which reflect frequent characteristics of real longitudinal data. This protocol allows to generate data while controlling the signal-to-noise ratio for different distribution types. To measure the performance of segmentation, we calculate two indices specifically adapted to evaluate the accuracy of signal segmentation: the Rand index and Jaccard index.

3 RESULTS

3.1 Applications on simulation data

3.1.1 Efficient computation time

Different segmentation techniques exhibit varying computation times based on the algorithms employed and the complexity of the calculations involved. The distribution of computation times required by each technique to perform segmentation across all simulations is analyzed based on signal length. An increase in computation time is observed with longer signal lengths, particularly for the Hubert technique. However, the computational times are relatively low for all the techniques, particularly for PELT and Binary Segmentation, with signal length having minimal impact.

3.1.2 Robustness

The different methods have specific characteristics that vary depending on the signal properties. Robustness evaluation helps identify methods which perform well in a variety of situations, including those with non-normal distributions, high noise, or gradual transitions. The analysis is based on the distribution of the mean lengths of detected segments compared to their actual lengths. BEAST, PELT and Segment Neighbourhood algorithms perform well in a variety of scenarios and are relatively unaffected by changes in signal parameters for normal distributions with shift changes. However, for gamma distributions, Binary Segmentation, PELT and CPM are the most effective techniques.

The introduction of gradual transitions disrupts the performance of most techniques, which are primarily designed to detect shift changes in signals. PELT, Segment Neighbourhood and Hubert work well for normal distributions, while Binary Segmentation performs best for gamma distributions.

3.1.3 Quality of the accuracy

The analysis of the Rand Index and Jaccard Index highlighted the accuracy of the techniques. The two parameters, r and S, influence the similarity between the reference model and the segmentation results. The accuracy of the techniques decreases as the r ratio increases. For normal distribution, all techniques perform well and particularly PELT and Segment Neighbourhood at high ratio in contrast to Jumpoint. CPM achieves reliable performance across a wide range of conditions for gamma distributions.

From the perspective of breakpoint accuracy, the Jaccard Index highlights Hubert as the most effective technique for normally distributed signals and CPM and Hubert for gamma distributed signals. These techniques perform particularly well at low ratio values. For signals with higher ratios, PELT and Segment Neighbourhood demonstrate good accuracy.

3.2 Applications on real data

3.2.1 Drac River

We demonstrate the application of segmentation techniques on real-world data to highlight their practical use in fluvial geomorphology. The data were obtained using the Fluvial Corridor Toolbox. More specifically, the data was generated using the Python version (Dunesme et al., 2024), where the metrics are calculated per 200-meter DGOs (see Alber et Piégay, 2011). We applied the segmentation techniques to the metric of active channel width (Figure 1). Segmenting the active channel width provides critical insights into the spatial variability and underlying geomorphological processes. BEAST and CPM demonstrate greater sensitivity compared to other techniques, while the remaining methods exhibit similar behavior, as confirmed by the Rand Index applied to each segmentation technique. A consensus algorithm was used to combine the outputs of the previously applied segmentation techniques, with the breakpoints identified through consensus clustering highlighting major variations in width, such as those caused by dams on the Drac river.

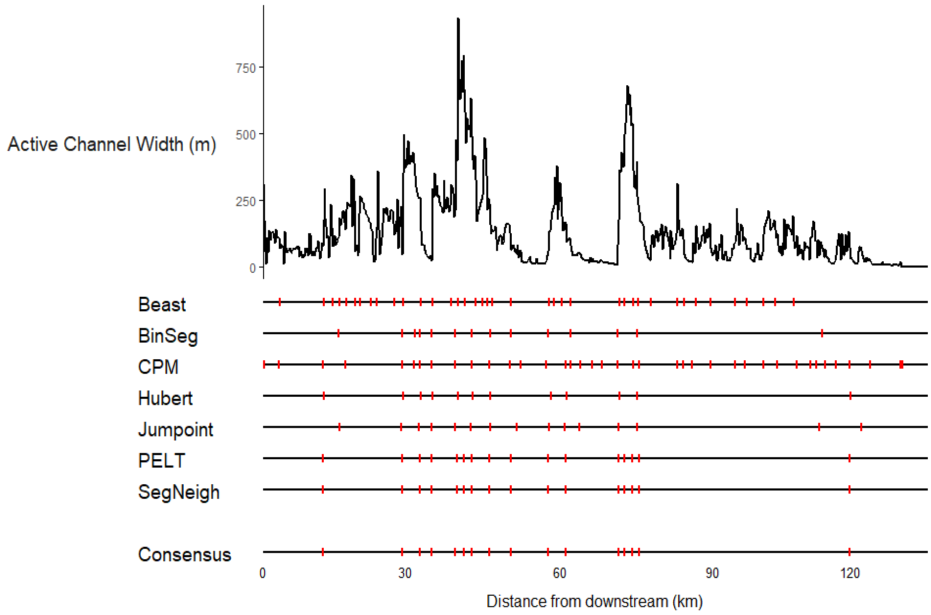

Figure 1 : The longitudinal evolution of the active channel width for the Drac River with the localization of discontinuities found by all the techniques

4 LIMITATION, STRENGTH AND APPLICATIONS

Our study aims to compare various signal segmentation methodologies to deepen understanding and assess their applicability within the fluvial continuum. Although no single technique proves to be universally superior, it is essential to develop a summary table highlighting the strengths and limitations of each method based on different criteria and signal parameters. The selection criteria for existing techniques and the definition of a performant approach to identify the most suitable technique remain important topics for discussion. Segmentation is often a fundamental step in river characterization, and refining these methods provides many opportunities for future research. One potential direction is to identify geomorphologically homogeneous segments by combining key metrics to analyze discontinuities in terms of transition types, drivers, or transition patterns. This could then be followed by multivariate classification to develop specific typologies such as fluvial styles (Brierley and Fryirs, 2005). An R Shiny application named RiverSeg was created to provide several segmentation techniques and simplify both graphical and cartographic visualization.

BIBLIOGRAPHIE

Alber, A., Piégay, H., 2011. Spatial disaggregation and aggregation procedures for characterizing fluvial features at the network-scale: Application to the Rhône basin (France). Geomorphology 125, 343–360

Basseville, M., Nikiforov, I.V., 1993. Detection of Abrupt Changes: Theory and Application.

Benda, L., Poff, N.L., Miller, D., Dunne, T., Reeves, G., Pess, G., Pollock, M., 2004. The Network Dynamics Hypothesis: How Channel Networks Structure Riverine Habitats. BioScience 54, 413–427

Brierley, G.J., Fryirs, K.A., 2005. Geomorphology and River Management: Applications of the River Styles Framework. John Wiley & Sons.

Dunesme, S., Manière, L., Rousson, C., 2024. Database of Metrics on River Corridors in France at the Regional Scale.

Lemire, D., 2007. A Better Alternative to Piecewise Linear Time Series Segmentation, in: Proceedings of the 2007 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, pp. 545–550

Leviandier, T., Alber, A., Le Ber, F., Piégay, H., 2012. Comparison of statistical algorithms for detecting homogeneous river reaches along a longitudinal continuum. Geomorphology 138, 130–144

Martínez-Fernández, V., Solana-Gutiérrez, J., González del Tánago, M., García de Jalón, D., 2016. Automatic procedures for river reach delineation: Univariate and multivariate approaches in a fluvial context. Geomorphology 253, 38–47

Parker, C., Clifford, N.J., Thorne, C.R., 2012. Automatic Delineation of Functional River Reach Boundaries for River Research and Applications. River Research and Applications 28, 1708–1725