Computers and Geosciences 112 (2018) 54–63

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

PolyWaTT: A polynomial water travel time estimator based on Derivative Dynamic Time Warping and Perceptually Important Points

Yuri Navarro Claure [a], Edson Takashi Matsubara [a,*], Carlos Padovani [b], Ronaldo Cristiano Prati [c]

[a] Federal University of Mato Grosso do Sul, Laboratory of Artificial Intelligence, Campo Grande, MS, Brazil
[b] Embrapa Pantanal, P.O. Box 109, 79320-900 Corumbá, MS, Brazil
[c] Federal University of ABC, Santo André, SP, Brazil

ARTICLE INFO

Keywords:
Timing parameters
Lag time
River level
Dynamic Time Warping
Time series

ABSTRACT

Traditional methods for estimating timing parameters in hydrological science require a rigorous study of the relations of flow resistance, slope, flow regime, watershed size, water velocity, and other local variables. These studies are mostly based on empirical observations, where the timing parameter is estimated using empirically derived formulas. The application of these studies to other locations is not always direct. The locations in which equations are used should have comparable characteristics to the locations from which such equations have been derived. To overcome this barrier, in this work, we developed a data-driven approach to estimate timing parameters such as travel time. Our proposal estimates timing parameters using historical data of the location without the need of adapting or using empirical formulas from other locations. The proposal only uses one variable measured at two different locations on the same river (for instance, two river-level measurements, one upstream and the other downstream on the same river). The recorded data from each location generates two time series. Our method aligns these two time series using derivative dynamic time warping (DDTW) and perceptually important points (PIP). Using data from timing parameters, a polynomial function generalizes the data by inducing a polynomial water travel time estimator, called PolyWaTT. To evaluate the potential of our proposal, we applied PolyWaTT to three different watersheds: a floodplain ecosystem located in the part of Brazil known as Pantanal, the world's largest tropical wetland area; and the Missouri River and the Pearl River, in United States of America. We compared our proposal with empirical formulas and a data-driven state-of-the-art method. The experimental results demonstrate that PolyWaTT showed a lower mean absolute error than all other methods tested in this study, and for longer distances the mean absolute error achieved by PolyWaTT is three times smaller than empirical formulas.

1. Introduction

Upstream water affects downstream locations after a certain amount of time. This amount of time, or simply this timing parameter, is a well-studied problem in hydrological science (Fang, 2005). A vast number of empirical formulas have been used to estimate travel time in the literature (Li and Chibber, 2008; Seyam and Othman, 2014; Demirel et al., 2013; Fang, 2005). These empirical formulas are derived for specific regions where local specifications, such as slope, watershed size, flow channel velocity, flow path length, are correlated with the measured travel time (Fang, 2005). The local specifications from one place can be very different from others, and the application of empirical formulas usually requires a parameter calibration, which can restrict the applications of empirical formulas.

In this work, we propose a data-driven method where the estimation is based on a single variable obtained from two different locations on the same river, regardless of their local specifications. Seyam and Othman (2014) have taken a step in the direction of using data-driven patterns to estimate travel time. Their method compares the peaks of two time series and estimates the travel time. Our proposal is a step forward in this direction. To the best of our knowledge, this is the first study where alignment of river travel time has been obtained from Derivative Dynamic Time Warping (DDTW, see Section 2) with Perceptually Important Points (PIP, see Section 3).

Timing parameters can have different names, such as travel time, time of concentration, excess-rainfall release time, wave travel time, time to

* Corresponding author.
E-mail addresses: yuri@lia.facom.ufms.br (Y.N. Claure), edsontm@facom.ufms.br (E.T. Matsubara), carlos.padovani@embrapa.br (C. Padovani), ronaldo.prati@ufabc.edu.br (R.C. Prati).

https://doi.org/10.1016/j.cageo.2017.12.002
Received 22 March 2016; Received in revised form 1 December 2017; Accepted 8 December 2017
Available online 13 December 2017

equilibrium, lag time, time base, and time to peak. The same timing parameter may have multiple meanings that create considerable confusion in hydrology (Fang, 2005). For the purposes of this paper, we have considered just the distance between peaks or valleys from an upstream to a downstream time series. As the time series dealt within this paper is river level, we named this time difference *Water Travel Time* (WaTT).

As a concrete example of how the proposed method works, consider measurements from two river level stations, Cáceres and Descalvados (129.7 km distant from each other) located on the Paraguay River, where the first is upstream from the second. Fig. 1 shows the water level during a year (365 Julian days), starting from July 1st, 1979 for these two stations. It is not obvious that the Cáceres's time series is a few days ahead (shifted to the left) of the Descalvados's time series.

Fig. 2 shows the alignment produced using our proposal. First, the time series in Fig. 1 were converted to standard z-scores ($\frac{x-\mu}{\sigma}$, where μ is the mean and σ the standard deviation of the series values). The proposed method applies DDTW (Keogh and Pazzani, 2001) (gray lines between time series and their respective number of shifted days to match the blue line time series with the green line time series); second, the peaks and valleys are detected in the series using PIP (red points) (Chung et al., 2001); third, the method matches the DDTW and PIP points. If DDTW and PIP agree that those points correspond, the method shows the correspondence (in red lines) along with the estimated delay, WaTT.

Although DDTWcomputes the match between two time series, we need to couple the DDTW with PIP to filter out possible non representative matchings. In Fig. 2 the black number between the series shows the estimate delay calculated by DDTW. The DDTW can capture minor variations in the series and the use of DDTWalone can mislead the timing parameter method. For instance, the number fourteen, close to the global peaks of the two time series, is detected by DDTW because of visually imperceptible tiny peaks that coincidently occurs in both series. The PIP helps to clean the possible matchings of these non-representative matchings.

The use of peak and valley detectors alone can also lead to non representative matchings, specially when the time series have plateau shapes, consecutive peaks, or consecutive valleys. Time series with these shapes, very common in hydrological science, can confuse timing parameter estimation methods by creating one to many possible matchings of peaks and valleys or incorrect matchings of peaks and valleys.

To generalize the timing parameter for unseen data (test data), WaTT are generalized by learning a polynomial regression using pairs of river level and WaTT named PolyWaTT. PolyWaTT is a function that maps an upstream river level measurement and the amount of time to reach the downstream location (WaTT). We performed a use case scenario analysis of our proposal on data of river levels at the Pantanal region in Brazil.

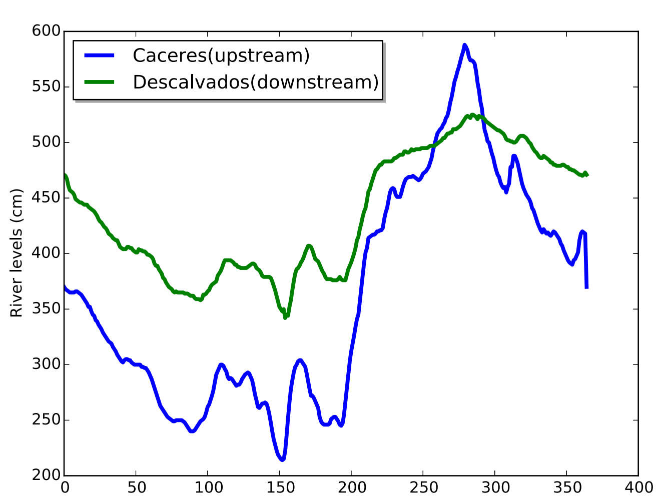

Fig. 1. River level measurements at the Paraguay River. Cáceres upstream of Descalvados.

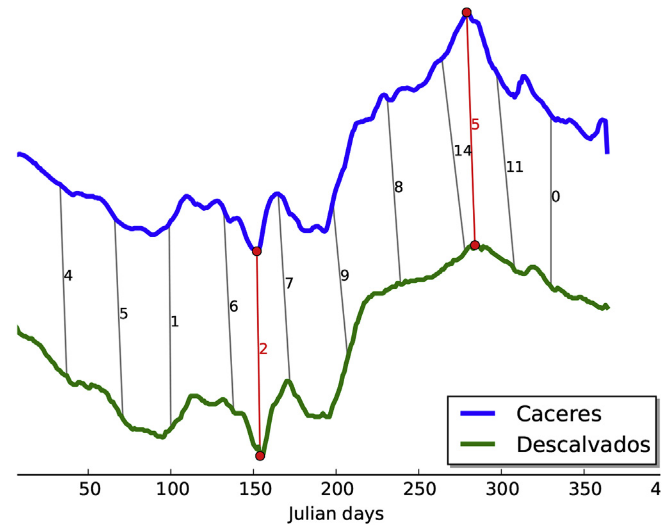

Fig. 2. The red dots represent peaks and valleys detected by PIP. The gray line between time series represents the aligned days using DDTW and the respective number of days to align the blue line with the green line. When we couple red dots and the gray lines we obtain the red line. The red line maps the correspondent days at Cáceres in Descalvados. For left most red line, PIP indicates day 152 in Descalvados (river level of 214 cm) and day 154 in Cáceres (river level of 350 cm), two days ahead, therefore the valley at Cáceres takes two days to represent a valley in Descalvados. This value is used in this work to better align two time series. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

To summarize, in this paper we make the following contributions:

- We demonstrate how to extract river level travel time from DDTW matrix. Although this strategy is well known in the time series literature, to the best of our knowledge, it is the first time that it is applied to estimate timing parameters in hydrological sciences.
- We propose an algorithm to define a polynomial function, PolyWaTT, that estimates the water level travel time by combining PIP with DDTW. This is a novel approach that can substitute the use of empirical formulas. The empirical formulas requires a extensive study to find formulas obtained from regions similar to our problem. PolyWaTT requires just training data of a single variable obtained from two different locations of the same river. In this work we used river level, but it can used for stream flow, precipitation and many other variables.
- We demonstrate that PolyWaTT is a robust method for water travel time estimation, specially for longer distances.
- The experimental results show that our proposal is significantly better, with 95% confidence, than the data-driven state of the art method;

2. Derivative dynamic time warping (DDTW)

To estimate the travel time between two river level stations, we need to be able to align two time series. The alignment of time series is a problem well studied in time series literature (Fu, 2011). One of the most used algorithms to address this challenge is Dynamic Time Warping (DTW) (Berndt and Clifford, 1994), described in Section 2.1. In this work, we use a variation of DTW named DDTW (for Derivative Dynamic Time Warping), described in Section 2.2.

2.1. Dynamic time warping

The Dynamic Time Warping (DTW) is widely used technique for finding an optimal alignment between two time series under certain conditions. Roughly speaking, a non-linear warping procedure is used to match the two sequences. DTW was developed in the decade of 1970 for speech recognition (Velichko and Zagoruyko, 1970; Sakoe and Chiba,