Journal of Hydrology 530 (2015) 751–761

Contents lists available at ScienceDirect

Journal of Hydrology

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

Event-based prediction of stream turbidity using a combined cluster analysis and classification tree approach

Amanda L. Mather, Richard L. Johnson *

Oregon Health & Science University, Institute of Environmental Health, 3181 SW Sam Jackson Park Road, Mail Code: GH230, Portland, OR 97239, United States

ARTICLE INFO

Article history:
Received 23 January 2015
Received in revised form 29 September 2015
Accepted 10 October 2015
Available online 22 October 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:
Turbidity
Event model
Prediction
Forecasting
Stream water quality

SUMMARY

Stream turbidity typically increases during streamflow events; however, similar event hydrographs can produce markedly different event turbidity behaviors because many factors influence turbidity in addition to streamflow, including antecedent moisture conditions, season, and supply of turbidity-causing materials. Modeling of sub-hourly turbidity as a function of streamflow shows that event model parameters vary on an event-by-event basis. Here we examine the extent to which stream turbidity can be predicted through the prediction of event model parameters. Using three mid-sized streams from the Mid-Atlantic region of the U.S., we show the model parameter set for each event can be predicted based on the event characteristics (e.g., hydrologic, meteorologic and antecedent moisture conditions) using a combined cluster analysis and classification tree approach. The results suggest that the ratio of beginning event discharge to peak event discharge (an estimate of the event baseflow index), as well as catchment antecedent moisture, are important factors in the prediction of event turbidity. Indicators of antecedent moisture, particularly those derived from antecedent discharge, account for the majority of the splitting nodes in the classification trees for all three streams. For this study, prediction of turbidity during streamflow events is based upon observed data (e.g., measured streamflow, precipitation and air temperature). However, the results also suggest that the methods presented here can, in future work, be used in conjunction with forecasts of streamflow, precipitation and air temperature to forecast stream turbidity.

1. Introduction

Turbidity is an important physical and visual characteristic of surface waters, and it serves as a useful indicator of overall stream health. Stream turbidity is relatively simple and inexpensive to monitor continuously (Davies-Colley and Smith, 2001) and is an important constituent in water quality monitoring and regulation (Gray and Glysson, 2003). Turbidity is also a versatile stream water quality surrogate, often enabling the estimation of suspended sediment (e.g., Gray and Glysson, 2003), agricultural chemicals (e.g., Hickman, 2004), pathogens (e.g., Christensen et al., 2000) and heavy metals (e.g., Miller, 1997).

High levels of turbidity can have negative consequences for both the natural and societal services and functions a stream provides. Streams are an important source of drinking water in the United States (U.S.) (Wickham et al., 2011), and turbidity is of particular interest to drinking water providers. Turbidity-causing materials can harbor pathogens and interfere with disinfection (U.S. EPA, 1999). Drinking water systems are subject to increased treatment cost with increases in source water turbidity (Dearmont et al., 1998), and episodes of extreme turbidity can surpass the operational limits of treatment systems (Duncan and Grant, 2003; Portland Water Bureau, 2011).

Stream turbidity typically increases during streamflow events and is influenced by many factors, including hydrograph shape (e.g., baseflow contribution; Bača, 2008), antecedent moisture conditions (Seeger et al., 2004; Giménez et al., 2012), season (Lana-Renault and Regüés, 2009; Mather and Johnson, 2014) and supply of turbidity-causing materials (Brasington and Richards, 2000; Doomen et al., 2008; Rodríguez-Blanco et al., 2010). Many of these factors vary on an event-by-event basis, and similar event hydrographs may be accompanied by notably different event turbidity behaviors (Fig. 1).

Recent modeling of sub-hourly stream turbidity as a function of streamflow has shown that model parameters vary on an event-by-event basis (Eder et al., 2010; Mather and Johnson, 2014). A logical next step is to examine the extent to which those model parameters can be predicted from event characteristics (e.g., antecedent moisture, hydrologic and meteorologic characteristics), which would allow the prediction of the stream turbidity time

* Corresponding author. Tel.: +1 5033463432.
E-mail address: rick.johnson.phd@gmail.com (R.L. Johnson).

http://dx.doi.org/10.1016/j.jhydrol.2015.10.032
0022-1694/

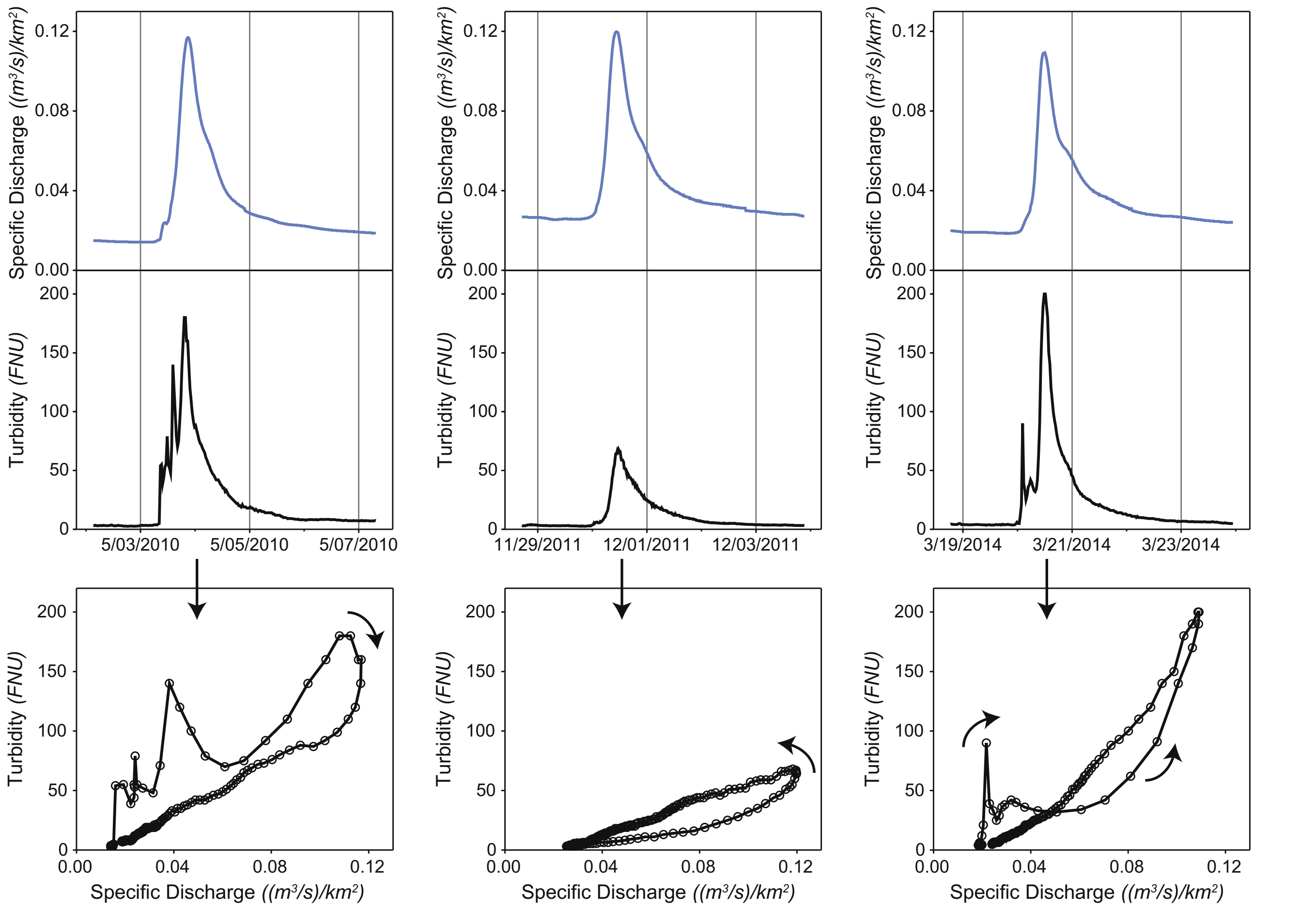

Fig. 1. Three example streamflow events from the Raritan River with similar hydrographs (blue) and different event turbidity (black). The resulting turbidity–discharge loops are shown below the time series plots. Data points are only shown on the loop plots and the loop direction is shown by the curved arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

series during streamflow events. Prior work predicting stream turbidity has examined event turbidity load/yield (Mukundan et al., 2013b), mean event turbidity (Mukundan et al., 2013b) or prediction of daily turbidity (Mukundan et al., 2013a), often in support of reservoir turbidity models (e.g., Samal et al., 2013). While useful in many contexts, these studies do not address the prediction of the turbidity time series during streamflow events, which could provide additional information, such as the event peak turbidity, the timing of the peak relative to the streamflow peak and the recession of turbidity following the peak.

The prediction approach used here is a combination of hierarchical cluster analysis and classification trees. Cluster analysis is a method of grouping objects to maximize group member similarity and has, for example, been used to group streams (Sawicz et al., 2011; Reidy Liermann et al., 2012), streamflow (Laaha and Blöschl, 2006; Olden et al., 2012) and streamflow events (Giménez et al., 2012; Mukundan et al., 2013b). Classification trees are commonly used in conjunction with cluster analysis and have been used to predict the cluster membership for "new" objects (e.g., ungauged streams (Reidy Liermann et al., 2012)), to examine how cluster membership may have changed over time (e.g., streams shifting from water- to energy-limited (Sawicz et al., 2014)), or to examine the connection between physical processes and tree design (e.g., seasonality and low flow discharges (Laaha and Blöschl, 2006)). In this study, cluster analysis is used to group streamflow events that display similar turbidity behavior. Classification trees are then used to predict both cluster membership and model parameters for "new" events that are not part of the original training dataset. For this study, prediction of turbidity during streamflow events is based upon observed data (e.g., streamflow, precipitation and air temperature). However, the methods presented here may, in future work, allow the forecasting of stream turbidity based on forecasts of streamflow, precipitation and air temperature.

2. Methods

2.1. Study catchments

This study uses data from three catchments within the Mid-Atlantic hydrologic region of the U.S. (Seaber et al., 1987), as shown in Fig. 2. This region receives an annual rainfall of approximately 1 m and precipitation has low seasonality; however, significant seasonality of streamflow is observed due to increases in evapotranspiration during the growing season (Neff et al., 2000). The mean annual temperature over the last century was approximately 11 °C (Polsky et al., 2000). The Raritan River (U.S. Geological Survey (USGS) 01400500) catchment has an area of 1269 km^2 and rises from 6 m elevation at the gage to a maximum elevation of 378 m with a mean slope of 5.5% (Falcone et al., 2010). This catchment is approximately half within the Piedmont physiographic province and half in the New England province (Fenneman and Johnson, 1946). The dominant land-cover classes for the Raritan River catchment are forest (40%, mostly deciduous), agriculture (28%, about equal parts pasture/hay and cultivated crops) and urban (22%) (Fry et al., 2011).