Automating the Classification of Hysteresis in Event Concentration-Discharge Relationships

Scott D. Hamshaw, Post-doctoral Associate, Vermont EPSCoR, University of Vermont, Burlington, VT, Scott.Hamshaw@uvm.edu
Doug Denu, Research Assistant, Vermont EPSCoR, University of Vermont, Burlington, VT, Douglas.Denu@uvm.edu
Maike Holthuijzen, Research Assistant, Vermont EPSCoR, University of Vermont, Burlington, VT, Maike.Holthuijzen@uvm.edu
Safwan Wshah, Assistant Professor, Department of Computer Science, University of Vermont, Burlington, VT, Safwan.Wshah@uvm.edu
Donna M. Rizzo, Professor, Department of Civil & Environmental Engineering University of Vermont, Burlington, VT, Donna.Rizzo@uvm.edu

Abstract

The response of in-stream sediment concentration and discharge during rainfall-runoff events provides information about dominant watershed processes as it represents the amalgamation of the connectivity, erodibility, and the spatial location of sediment sources. A common way to collapse the sediment and streamflow response into a readily interpretable visualization is to utilize an event concentration-discharge (C-Q) plots which frequently exhibit patterns of hysteresis. However, challenges exist given the subjective nature of visual classifications and when scaling to large data sets. Hysteresis indices have been used to facilitate an automated and objective analysis method. In this study, we present an alternative method for automating hysteresis classification utilizing all the information present in the event C-Q plots. Thus, avoiding the loss of information that may occur when collapsing data into metrics and enabling the local sediment dynamics to be interpreted to a greater extent.

We developed an automated machine learning tool using images of event C-Q plots to classify storm events into pre-defined hysteresis pattern types. The classifier utilizes a convolutional neural network, a machine learning method that has achieved excellent predictive accuracy in image classification tasks. We then applied this tool using surrogate suspended sediment data from turbidity monitoring in eight watersheds within the Lake Champlain Basin in Vermont encompassing 760 individual storm events. The tool accurately and efficiently classifies events and represents an advancement over manual visual classification.

Background

Event Sediment Dynamics

The various mechanisms controlling suspended sediment transport in watersheds during hydrological events are complex. Efforts to gain insight into the watershed processes that result in soil erosion, sediment loading of rivers, and transport to downstream ecosystems has resulted in detailed study of watershed streamflow and sediment responses to rainfall events. The coupling between the streamflow and sediment responses to rainfall events is apparent in the relationship between streamflow and suspended sediment concentration (SSC) (Lefrançois et
This relationship can change as a function of the sediment source availability, sediment storage, and hydrological pathways (connectivity) present in the watershed (Asselman, 1999; Duvert et al., 2010; Sherriff et al., 2016; Hamshaw et al., 2018). A characteristic of this complex coupling is that the SSC response during a hydrological event is not in phase with the associated streamflow response (Gao and Josefson, 2012), which results in hysteretic behavior in the streamflow–SSC relationship. For decades, hydrological scientists have studied this hysteretic behavior to understand the origin and transport of sediments in watersheds using an approach commonly referred to as hysteresis analysis (Williams, 1989; Seeger et al., 2004; Lawler et al., 2006; Smith and Dragovich, 2009; Gellis, 2013; Sherriff et al., 2016; Hamshaw et al., 2018).

**Hysteresis in Event Concentration-Discharge Relationships**

The shape and direction of hysteresis loops highlight the temporal offset between SSC and streamflow. Hysteresis patterns can be used to understand the physical processes in watersheds. For example, clockwise loops have been recognized to be broadly characteristic of sediment sources being located near the watershed outlet, whereas counter-clockwise loops indicate that sediment sources are primarily located in the headwaters of the watershed (Gellis, 2013; Sherriff et al., 2016). Additional physical processes have been connected with categories of hysteresis resulting in a number of interpretations for each general type (Williams, 1989; Lefrançois et al., 2007; Smith and Dragovich, 2009; Gellis, 2013; Hamshaw et al., 2018).

The most common hysteresis analysis is to classify plots into a general set of five classes (i.e. linear/no hysteresis, clockwise, counter-clockwise, figure-eight, and complex). These general categories, first identified by Williams of USGS (1989), continue to be used to this day by researchers. The manual, visual categorization of hysteresis loops is human resource intensive and has critical computational drawbacks. There has been progress toward automatic categorization of hysteresis using hysteresis indices from event C-Q data (Lawler et al., 2006; Lloyd et al., 2016; Zuecco et al., 2016; Vaughan et al., 2017). This approach is computationally efficient, and to some extent, effective. However, the hysteresis indices are not unique (i.e., individual storm events with different hysteresis patterns can have the same index value), and, therefore, often require additional metrics such as loop area or direction to preserve information lost during data compression (Zuecco et al., 2016). An approach that utilizes the full information of the hysteresis plot is therefore necessary.

**Machine Learning for Pattern Recognition**

Machine learning methods can help identify patterns in hydrological data. For example, feed-forward backpropagation algorithms have long been used in rainfall-runoff modeling and streamflow prediction (Abrahart et al., 2012). More recently, a new family of machine learning methods called deep learning that excel at classification applications such as hand-written character recognition (O’Connor et al., 2013) have been developed, sparking extensive research into deep learning over the last decade. One of the earliest and most successful deep networks for image classification is the convolutional neural network (CNN) (LeCun et al., 2015). These networks were introduced in the early 1990s (LeCun et al., 1989) and have exhibited excellent performance for tasks such as hand-written digit and face recognition, image classification competitions (Simonyan and Zisserman, 2014), and speech recognition (Sainath et al., 2015). CNNs can be conceptualized as a series of feature detectors (i.e. edges, corners, shape, color pattern, etc.) connected to a classifier.

The true potential of CNN networks was only fully understood when applied to large datasets (e.g. ImageNet dataset containing over 14 million labeled images) using high performance
computing. Several high-performance CNN architectures have been established including LeNet-5 (LeCun et al., 1990), AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan and Zisserman, 2014), GoogleNet (Szegedy et al., 2015), and ResNet50 (He et al., 2016). With the increase in computing power including GPU processing on standard desktop computers, CNNs are becoming more widespread in disciplines outside of computer vision, especially in biology and ecology (Dyrmann et al., 2016; Xu et al., 2017). Thus, due to the wide applicability of CNNs for a variety of tasks, they present a promising method for classifying hysteresis plots.

This project aims to demonstrate the performance of an advanced deep learning architecture (the ResNet50 CNN) to classifying hysteresis patterns in suspended sediment data. As efforts are made to collect and provide access to ever more high-frequency monitoring data, analysis methods such as those demonstrated here highlight the potential of continuous monitoring data and machine learning approaches to understand our watersheds.

**Methods**

**Study Area & Dataset**

We obtained streamflow and suspended sediment data from eight watersheds within the Lake Champlain Basin in Vermont between 2013 and 2016 (Figure 1). These included the Mad River watershed and five of its subwatersheds as well as Hungerford Brook and Wade Brook. Suspended sediment data was determined using surrogate monitoring of turbidity. In the case of the Mad River watersheds turbidity was measured using FTS DTS-12 sensors and YSI EXO2 sondes for Hungerford Brook and Wade Brook. Turbidity monitoring was paired with event sampling of total suspended solids (TSS) at each site for developing TSS-turbidity relationships. See Hamshaw et al. (2018) and Vaughan et al. (2017) for further details on streamflow and turbidity monitoring studies.

![Map of study watersheds in the Lake Champlain basin in Vermont.](image)

**Figure 1.** Map of study watersheds in the Lake Champlain basin in Vermont.
Event Concentration-Discharge Analysis

Hydrological events analyzed in this were previously extracted from continuous records of streamflow and turbidity as part of event-based analysis of nutrient dynamics in Hungerford Brook and Wade Brook (Vaughan et al., 2017) and sediment dynamics in the Mad River watershed (Hamshaw et al., 2018). Both studies utilized an approach of identifying an event start as a threshold increase in streamflow and concentration and event end as the inflection point in the falling limb of the hydrograph (Figure 2a). We constructed individual event concentration-discharge plots ("hysteresis plots") from the extracted time series segments (Figure 2b).

![Figure 2](image.png)

**Figure 2.** (a) Example segmentation of individual events in streamflow and sediment time series and (b) example (event C) concentration-discharge plot.

To assist with development and evaluation of automated hysteresis classification models all individual storm events were manually labeled based on the type of hysteresis observed in the event concentration-discharge plot. The categories (classes) of hysteresis patterns (Figure 3) used for classification of individual storm events are based on those identified in the Mad River watershed monitoring project (Hamshaw et al. 2018). These categories are an expansion of the typical categories (e.g. linear, clockwise, counter-clockwise, figure-eight).
Individual storm event hysteresis plots were converted to images for input to the deep learning model. This allowed for use of existing, open source deep learning architectures designed for image classification. We note this method of classification based on images of hysteresis plots was also previously developed in a proof-of-concept study by Hamshaw et al. (2018). Processing of events was automated in MATLAB (Version 9.2) where each hysteresis plot was converted to a grayscale image (256 pixels x 256 pixels) and time was preserved in the shading of the data with white representing storm start and darker grey, the end of event (Figure 4). The images were randomly divided into training and testing subsets based on their manually labeled hysteresis type (Figure 5). For the training data set, event sampling was done by hysteresis type in order to achieve as equal representation of hysteresis types as possible.

We used the model weights and architecture of ResNet50, a residual convolutional neural network (CNN), as our base model to classify the image data. Residual networks (ResNets), a type of CNN developed by (He et al., 2016), have achieved excellent predictive accuracies in image classification competitions. We decided to use the ResNet50 as our base model for hysteresis image classification because it is one of the more efficient (in terms of computing resources), yet highly accurate classification models for images. The ResNet architecture is a very deep neural network characterized by skipped connections and batch normalization (He et al., 2016). As the name suggests, ResNet50 is composed of 50 weight layers that are tuned during model training. Readers are referred to He et al. (2016) and Sze et al. (2017) for in depth discussion of model architecture and training methods. We used the Python (version 3.6) implementation available in the Keras deep learning package (Chollet and others, 2015), which utilizes TensorFlow (Abadi et al., 2015) as its computational engine. We utilized an NVIDIA DGX-1 deep learning server to train and test the model.
Figure 4. Model workflow for classification of hydrological events based on image of hysteresis plot using a deep learning classifier.

Figure 5. Training and testing data set by type of hysteresis.
Input data to the CNN was represented as 256x256 pixel grayscale image data of a hysteresis pattern. Due to the small size of the training dataset (223 images) compared to those typical of deep learning applications, we performed data augmentation prior to training the model to increase the training data set size. Data augmentation consisted of creating “new” storm events where hysteresis plots were slightly shifted and perturbed from the original sample. The built-in data augmentation tool in the Keras package was used to perform all data augmentation. After applying the augmentation methods, we obtained a total of 1115 hysteresis plot images. The CNN model was trained using 20% of the training data for validation. In Keras, when a validation split of the data is specified within the fit method, the validation set is chosen randomly prior to each epoch. Thus, for each epoch, a different validation set is used to estimate the validation loss.

We performed a parameter sweep over several values of learning rates and batch sizes to determine optimal values for these parameters. Models in the parameter sweep were run for 1000 epochs. We chose not to include the number of epochs in the parameter sweep and instead chose to set the number of epochs very high and assess the optimal number of epochs by inspecting graphs of loss vs. epochs. We used categorical cross-entropy as the loss function and the Adam optimizer (Kingma and Ba, 2014) for all models. After training the model, we chose the model in which the loss function decreased to the greatest degree over 1000 epochs and used that model to predict onto the test data (543 images). Accuracy was assessed as the percentage of 543 test events classified by the CNN model as the same as the manual labels.

**Results and Discussion**

**Hysteresis types of storm events**

The manual classification of the 765 storm events observed across the eight study watersheds showed variations in distribution of hysteresis types across watersheds. Across all watersheds Class II (Clockwise) patterns were most common (Figure 6). The exception to this was the larger Mad River watershed and low-gradient, agricultural land use dominated Hungerford Brook where Class III (counter-clockwise) and Class V (Figure-eight) patterns occurred frequently. Attributing to the predominance of smaller, steep forested watersheds in the dataset (six of eight watersheds) the representation of hysteresis types was not especially well balanced. A study area that features watersheds with more balanced variation of land use and drainage area would likely result in a more balanced distribution of observed hysteresis types. For data-driven modeling approaches such as machine learning, access to a wide variety of training and testing data can be significant for successful model development.
CNN Model Performance

The CNN classifier achieved an overall accuracy of 69% on the test dataset. This improves over automated classification using the more basic neural network classifier previously demonstrated by Hamshaw et al. (2018). The classification accuracy of our CNN model approaches manual, visual classification of hysteresis patterns (Romero et al., 2018). We found that when the CNN model mis-classified an event, it often was placed into one of the most visually similar categories. For example, while 78% of Type 2C were classified correctly, 21% of the misclassifications were into the similar Type 2D and 2E classes. The confusion matrix (Figure 7) shows classification accuracies for each hysteresis type. Values along the diagonal show the classification accuracy for each particular type of hysteresis whereas values on the off-diagonals are considered mis-classifications. While hysteresis type is not a true ordinal variable, in a general sense mis-classifications on the 1-off diagonal can be considered “near-misses.”
Dataset Challenges

Although we achieved a reasonable overall accuracy of 69%, we speculate that our inability to achieve a greater overall classification accuracy was primarily due to inherent complexity of hysteresis patterns present in event C-Q relationships. In a related study using a subset of the Mad River storm events, manual visual classification at best was 85% accurate (Romero et al., 2018). However, we speculate three additional factors also contributed to inability to achieve higher classification accuracies: 1) the sparsity of data, 2) imbalanced class representation, 3) and similarity among classes. Due to the substantial number of parameters that need to be estimated during training, deep learning techniques are best suited to large datasets (Najafabadi et al., 2015), and deep learning techniques tend to overfit on smaller datasets (LeCun et al., 2015) like ours. We attempted to resolve the issues associated with small datasets by conducting a simple data augmentation. Furthermore, a small training dataset size and class imbalance are two attributes that, when combined, can have a substantial, negative impact on the performance of deep learning methods. Since our data exhibited these characteristics, it is not surprising that our classification accuracy plateaued at 69%. Expansion of the training data set to encompass additional storm events from more watersheds and monitoring periods would likely improve CNN model performance.

Finally, the hysteresis data were characterized by poor distinction among several of the 14 different hysteresis types. We chose to test the methodology on the expanded hysteresis classification scheme developed by Hamshaw et al. (2018). However, the model could be just as easily trained and tested on storm events using a simplified classification scheme, such as the more traditional clockwise, counter-clockwise, and figure eight patterns (Williams, 1989). For example, if we aggregated our results to the four commonly used categories of no hysteresis, clockwise, counter-clockwise, and figure-eight hysteresis, then the classification accuracy would improve to 89.7%. Our results would suggest that approach may be more appropriate on datasets smaller than ours. However, we believe our initial results show that a CNN classifier is a promising method to automate the classification of hysteresis types.

Modeling Environment

In this study, the pre-processing routines were developed in MATLAB (version 9.2) and the CNN model implementation in Python. This approach clearly requires programming knowledge, and therefore, presents a challenge over simpler methods such as hysteresis indices. However,
we believe the availability of deep learning models are more accessible than ever; given the variety of algorithms available off the shelf ready to use or to provide a starting point. The CNN used as proof-of-concept here (ResNet50) is now available in a variety of common programming languages (e.g., MATLAB, R, Python) and can be implemented in very few lines of code as a result of wrapper packages (e.g. Keras) or included functions (MATLAB). While efficient training of a CNN often requires high-performance GPU hardware and therefore may not be practical for all applications, once trained CNNs operate fast and may be used directly on more basic computing hardware (e.g. typical desktop and laptop computers) without requiring additional network training required. Finally, because this approach is analogous to machine learning applications used for handwritten character recognition, a number of tutorials are now available based on the similar MNIST benchmark data set (e.g. Pedregosa et al., 2011; Eclipse Deeplearning4j Development Team, 2017).

**Summary**

The deep learning tool for automating visual pattern recognition from hydrological data presented here represents a novel application of machine learning in hydrology. As illustrated in this study, the automation of analysis of hysteresis patterns allows for rapid application to data sets containing hundreds of storm events. While further research is warranted on developing consistent, automated methods for extracting storm events from continuous streamflow and sediment monitoring data, the potential to generate datasets containing thousands of storm events exists. As these datasets begin to reflect a variety of watersheds with different land use, climate, geology, topography, and drainage area, the application of deep learning methods offer an opportunity for building a greater understanding of drivers of sediment loading during storms across both time and space.

**Acknowledgements**

Support provided by Vermont EPSCoR, with funds from the National Science Foundation (NSF) Grant OIA-1556770, is acknowledged. We would also like to acknowledge the contributions of Scott Turnbull, Ali Javed, and Kevin Andrew.

**References**


Chollet F, others. 2015. *Keras*. Available at: https://keras.io

10.1016/j.biosystemseng.2016.08.024


Seeger M, Errea M-P, Begueri a S, Arnáez J, Marti c, Garcia a-Ruiz J. 2004. Catchment soil moisture and rainfall characteristics as determinant factors for discharge/suspended
sediment hysteretic loops in a small headwater catchment in the Spanish pyrenees.


