

# DISTANCE FUNCTIONS AND ATTRIBUTE WEIGHTING IN A $k$ -NEAREST NEIGHBORS CLASSIFIER WITH AN ECOLOGICAL APPLICATION\*

Alyssa C. Frazee<sup>1</sup>, Matthew A. Hathcock<sup>2</sup> and Samantha C. Bates Prins<sup>3</sup>

<sup>1</sup> St. Olaf College

Department of Mathematics, Statistics, and Computer Science

Northfield, MN 55057 USA

e-mail: frazee@stolaf.edu

<sup>2</sup> Winona State University

Department of Mathematics and Statistics

Winona, WI 55987

email: MAHathco4262@winona.edu

<sup>3</sup> James Madison University

Department of Mathematics and Statistics

Harrisonburg, VA 22807

email: prinssc@jmu.edu

## Abstract

To assess environmental health of a stream, field, or other ecological object, characteristics of that object should be compared to a set of reference objects known to be healthy. Using streams as objects, we propose a  $k$ -nearest neighbors algorithm (Bates Prins and Smith, 2006) to find the appropriate set of reference streams to use as a comparison set for any given test stream. Previously, investigations of the  $k$ -nearest neighbors algorithm have utilized a variety of distance functions, the best of which has been the Interpolated Value Difference Metric (IVDM), proposed by Wilson and Martinez (1997). We propose two alternatives to the IVDM: Wilson and Martinez's Windowed Value Difference Metric (WVDM) and the Density-Based Value Difference Metric (DBVDM) developed by Wojna (2005). We extend the WVDM and DBVDM to handle continuous response variables and compare these distance measures to the IVDM within the ecological  $k$ -nearest neighbors context. Additionally, we compare two existing attribute weighting schemes (Wojna 2005) when applied to the IVDM, WVDM, and DBVDM, and we propose a new attribute weighting method for use with these distance functions as well. In assessing environmental impairment, the WVDM and DBVDM were slight improvements over the IVDM. Attribute weighting also increased the effectiveness of the  $k$ -nearest neighbors algorithm in this ecological setting.

---

\*This research was supported by NSF grant NSF-DMS 0552577 and was conducted during an 8-week summer research experience for undergraduates (REU).