H43I-2016

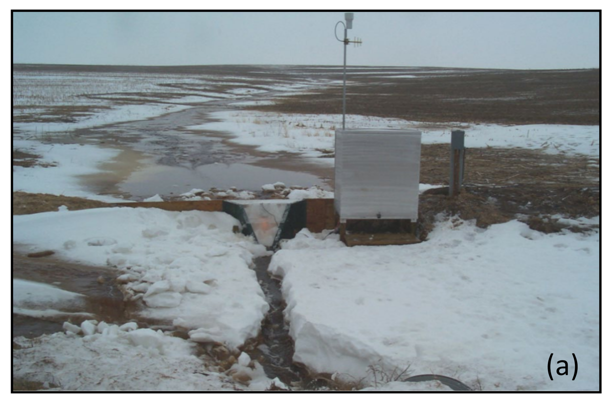
Assessing Runoff Risk to Support Nutrient Application Timing using a Hybrid of Physically-based and Statistical Model — an Application of National Water Model

Yao Hu[1], Lindsay Fitzpatrick[1], Dustin Goering[2], Lacey Mason[3], Lauren Fry[3,4], Laura Read[5], Andrea Thorstensen[2], Brent Lofgran[3]

[1]Cooperative Institute for Great Lakes Research, [2]NOAA National Weather Service, [3]NOAA Great Lakes Environmental Research Laboratory, [4]US Army Corps of Engineers, [5]National Center for Atmospheric Research

Introduction

Nutrient transport from runoff events has affected crop production and revenue due to loss of nutrients, and deteriorated water quality, leading to harmful algal blooms and hypoxia in receiving water bodies in the continental United States.



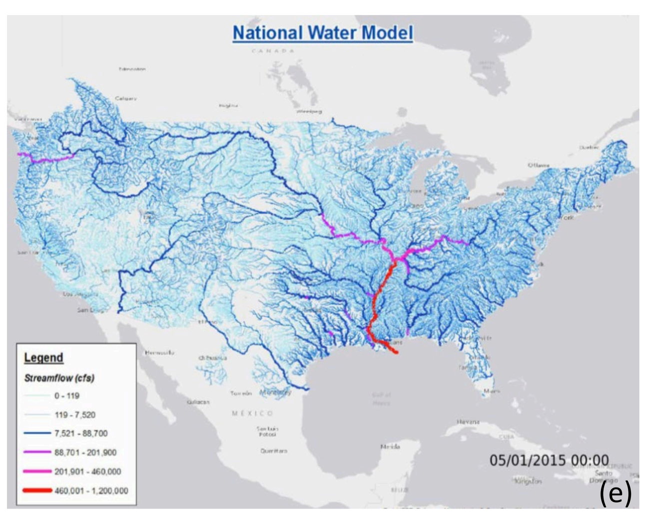
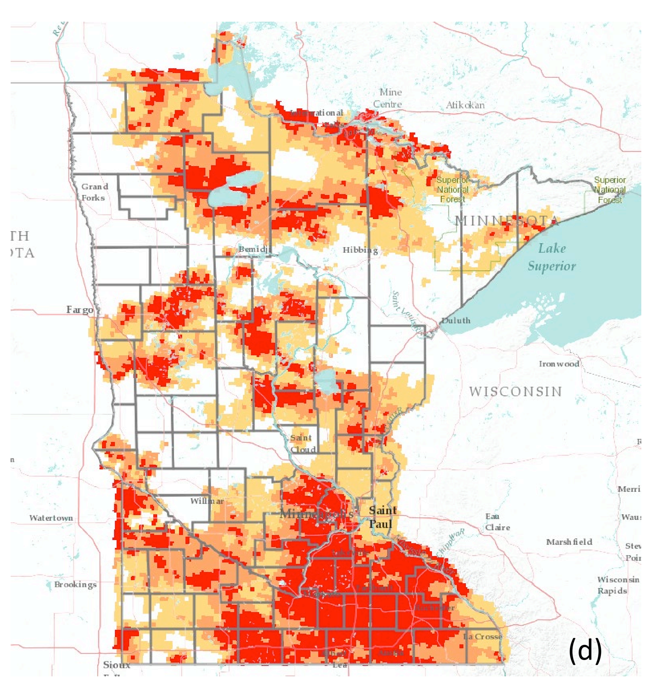


Figure 1: (a) an Edge-of-Field (EOF) measurement site where manure applied 5-6 days before a rain-on-snow runoff event; (b) Harmful algal blooms (HABs) on Lake Erie in July, 2019; (c) Water quality of Lake Erie affected by HABs. (d) The RRAF for Wisconsin on July 17, 2019 (e) The NOAA national water model (NWM).



Need: To assess these runoff risks, an enhancement of the existing runoff risk assessment tools (e.g. the Runoff Risk Advisory Forecast (RRAF) system) is needed to support agricultural producers to avoid nutrient application before significant runoff events.

Idea: Develop a statistical model to predict the occurrence/magnitude of EOF runoff events at a daily scale for the lower 48 using the outputs from the National Water Model (NWM).

Method

1. Observations and Model Outputs

Edge of field (EOF) observations were collected from over 50 locations across the upper Midwest and Great Lakes. Together with 72 out of 172 NWM outputs, these EOF measurements are used to train a statistical model for each watershed.

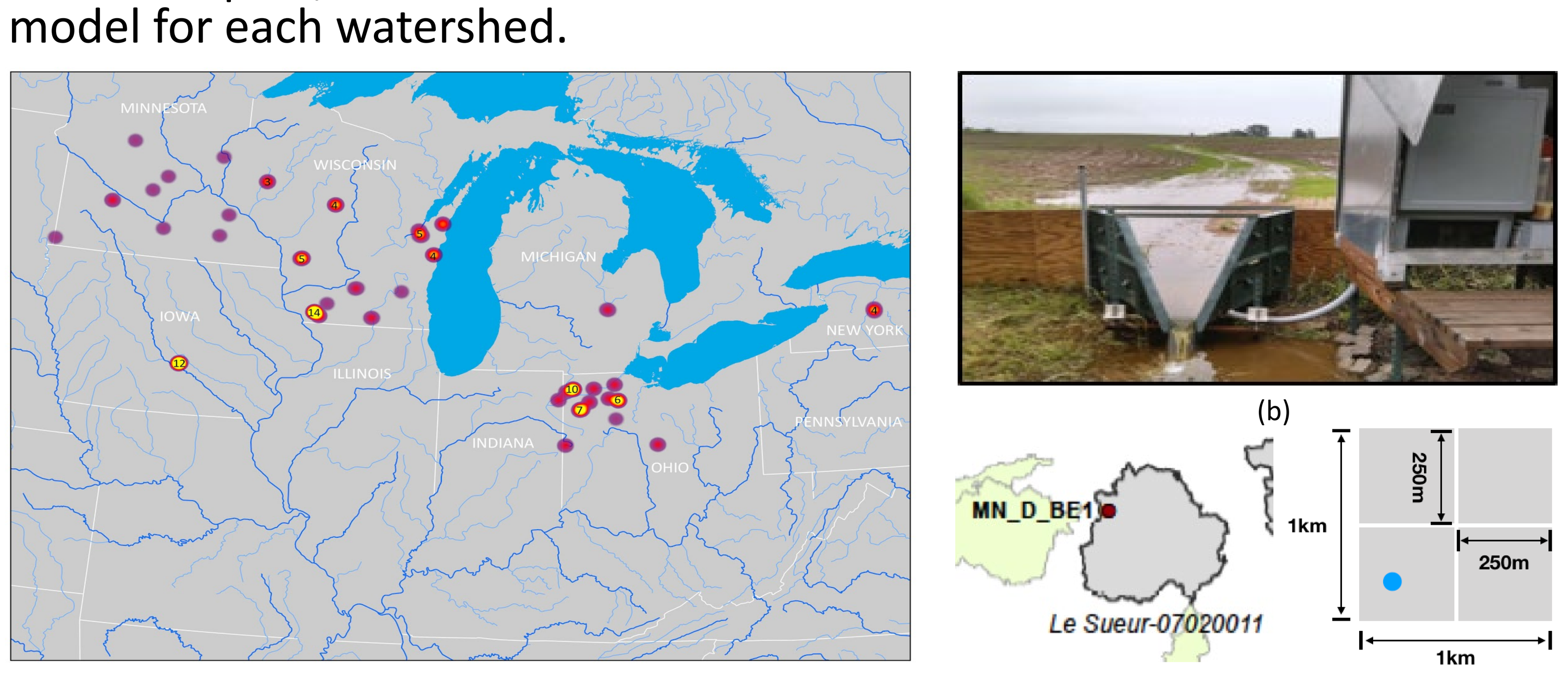
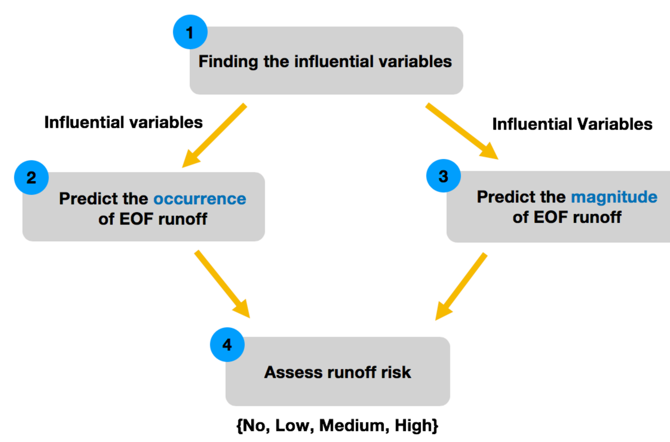


Figure 2: (a) all the EOF sites (112 observations) across the upper Midwest and Great Lakes; (b) an EOF measurement site; (c) A small watershed (grey) where the NWM simulation is executed; (d) land grid(1kmx1km) and routing grid(250kmx250km).

2. Workflow

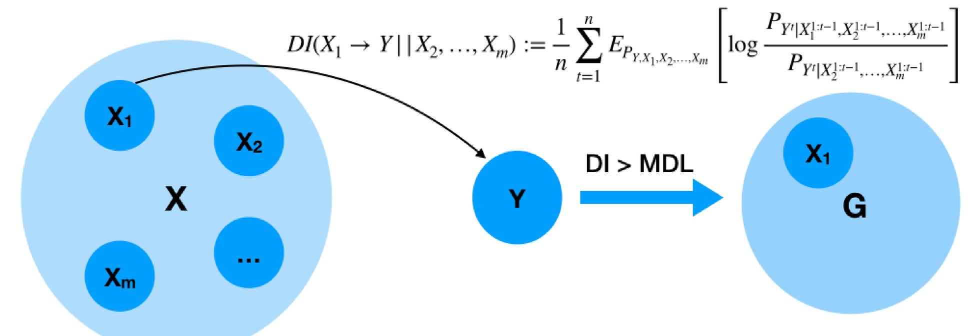


Step 1: Identify the influential variables using Directed Information (DI) (Hu et al., 2018)

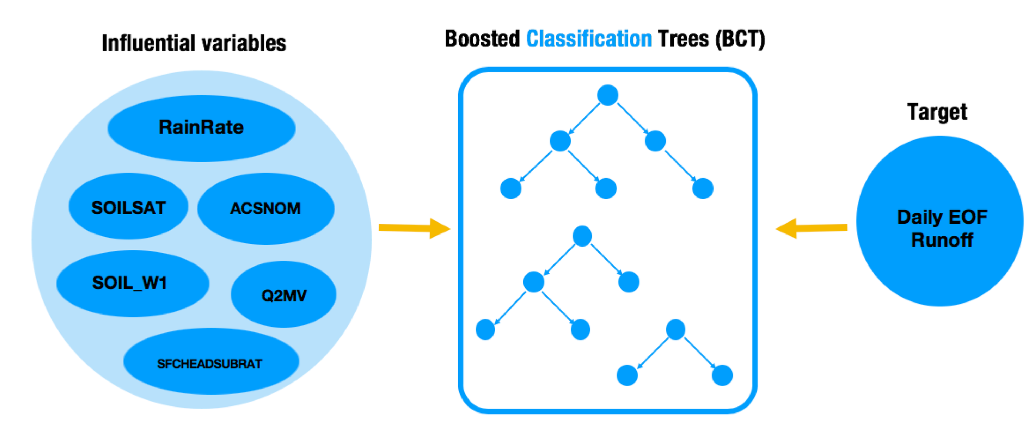
Input: X = { $X_1, X_2, \ldots, X_m$ } and Y

$$DI(X_1 \to Y \,||\, X_2, \ldots, X_m) := \frac{1}{n}\sum_{i=1}^{n} E_{P_{Y,X_1,X_2,\ldots,X_m}}\left[\log \frac{P_{Y|X_1^{i-1},\ldots,X_m^{i-1}}}{P_{Y|X_2^{i-1},\ldots,X_m^{i-1}}}\right]$$

Date	Runoff (mm)	Event
09/01/2012	0	0
…	…	…
04/10/2013	36.0	1
…	…	…

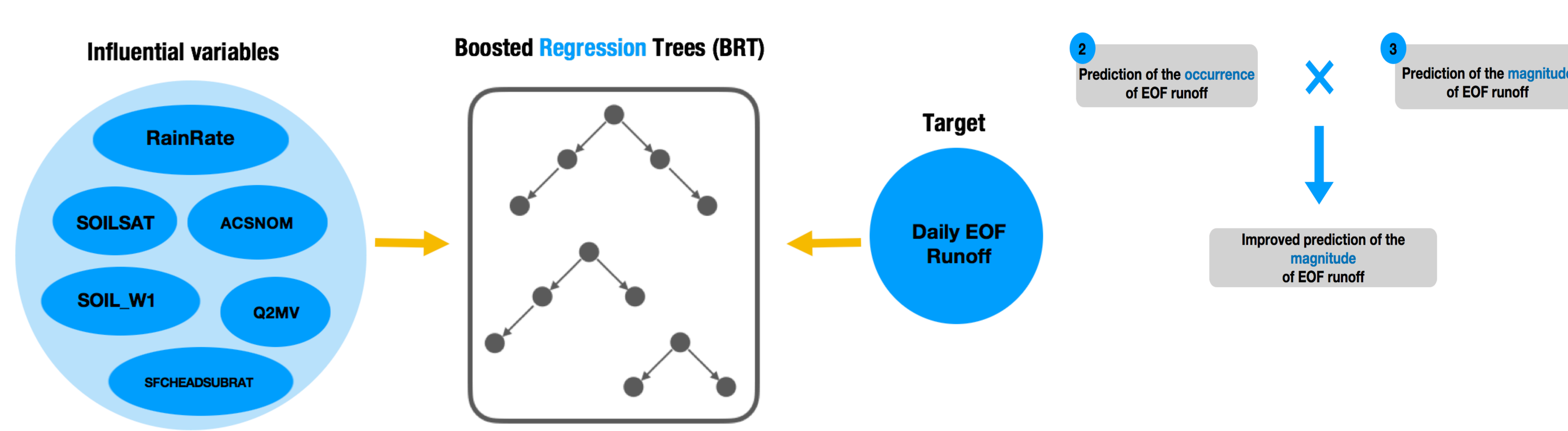


Step 2: Predict the occurrence of an EOF event using boosted classification Trees (BCT)



Daily EOF Runoff = BCT(set of influential variables)
Training : Validation = 70% : 30%

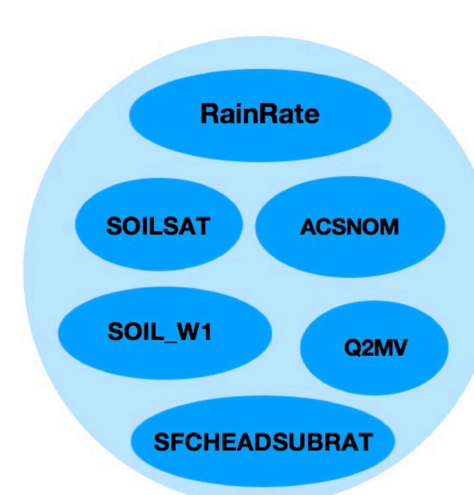
Step 3: Predict the magnitude of an EOF event using boosted regression Trees (BRT)



Daily EOF Runoff = BRT(set of influential variables)
Training : Validation = 70% : 30%

Results and Discussion

1. Influential Variables



Six out of 72 variables are selected as influential variables for the two EOF sites in MN.

2. Prediction of the occurrence of EOF runoff

For all 1530 daily measurements

Measurement \ Prediction	Y	N
Y	82	11
N	193	1244

Accuracy: 86.7%
Miss: 0.7%
False Alarm: 12.6%

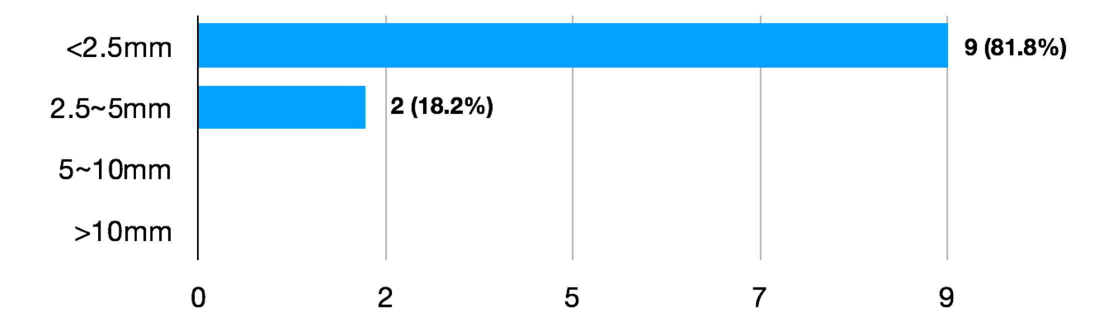
True Positive Rate (TPR): 82/93 = 88.2%

False Positive Rate (FPR): 193/(1530 - 93) = 13.4%

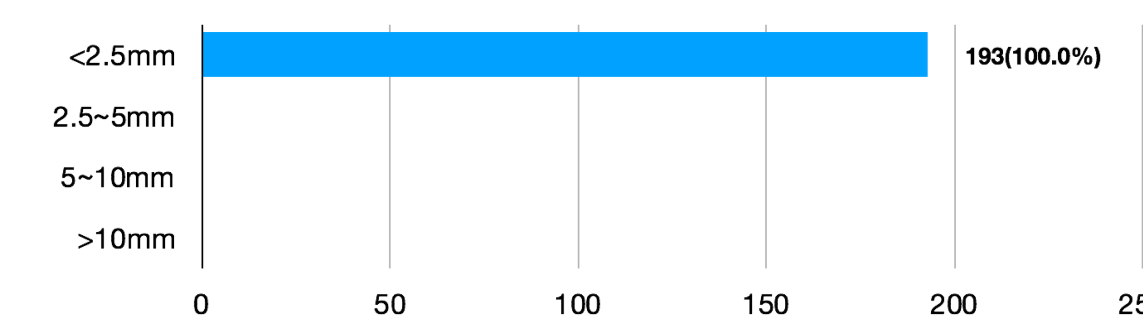
2. Prediction of the occurrence of EOF runoff

	Y	N
Y	82	11
N	193	1244

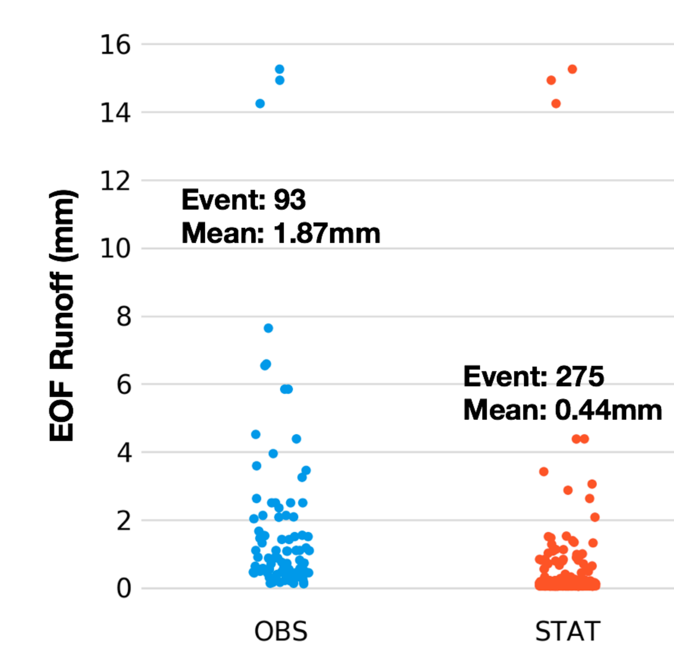
For all 11 missing events

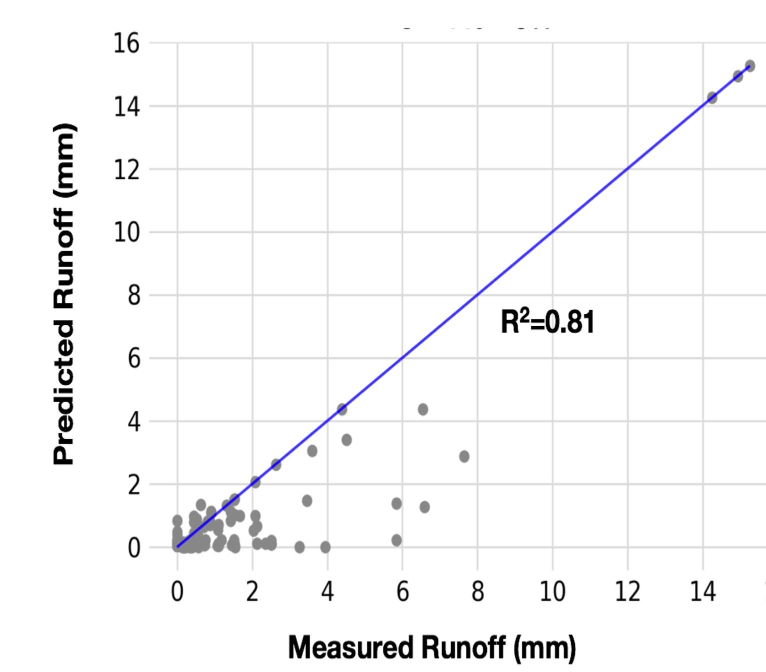


For all 193 false alarms



3. Prediction of the magnitude of EOF runoff





- The statistical models (BCT+BRT) have a good prediction of the occurrence and magnitude of EOF runoff events.
- Statistical models tend to miss small events (less than 2.5mm). To capture these small events can result in overfitting the statistical models. To avoid overfitting, we thus need to balance the miss and false alarm rate.
- Different sites can have different sets of influential variables, arising from different physical characteristics of watersheds. When regionalizing the statistical models, physical parameters need to be taken into account.

References & Acknowledgments

Hu, Y., Scavia, D., & Kerkez, B. (2018). Are all data useful? Inferring causality to predict flows across sewer and drainage systems using directed information and boosted regression trees. *Water research*, *145*, 697-706.

NCAR
Great Lakes RESTORATION
National Weather Service
USGS science for a changing world
Discovery Farms Minnesota
Discovery Farms Wisconsin
US Army Corps of Engineers