Regional patterns and drivers of nitrogen trends in a human-impacted watershed and management implications

Qian Zhang[1], Joel Bostic[1], and Robert Sabo[2]

[1]University of Maryland Center for Environmental Science
[2]U.S. Environmental Protection Agency

November 21, 2022

Abstract

Nutrient enrichment is a major issue to many inland and coastal waterbodies worldwide, including Chesapeake Bay. River water quality integrates the spatial and temporal changes of watersheds and forms the foundation for disentangling the effects of anthropogenic inputs. However, many water-quality studies are focused on limited portions of the watershed or a subset of potential drivers. We demonstrate with the Chesapeake Bay Nontidal Monitoring Network (84 stations) that advanced machine learning approaches – i.e., hierarchical clustering and random forest – can be combined to better understand the regional patterns and drivers of total nitrogen (TN) trends in large monitoring networks. Cluster analysis revealed the regional patterns of short-term TN trends (2007-2018) and categorized the stations to three distinct clusters, namely, V-shape (n = 25), monotonic decline (n = 35), and monotonic increase (n = 26). Random forest models were developed to predict the clusters using watershed characteristics and major N sources, which provided information on regional drivers of TN trends. We show encouraging evidence that improved nutrient management has resulted in declines in agricultural nonpoint sources, which in turn contributed to water quality improvement. Additionally, water-quality improvements are more likely in watersheds underlain by carbonate rocks, reflecting the relatively quick groundwater transport of this terrain. However, TN trends are degrading in forested watersheds, suggesting new sources of N in forests. Finally, TN trends were predicted for the entire Chesapeake Bay watershed at the scale of 979 river segments, providing fine-level information that can facilitate targeted watershed management, especially in unmonitored areas. More generally, this combined use of clustering and classification approaches can be applied to other monitoring networks to address similar questions.

Regional patterns and drivers of nitrogen trends in a human-impacted watershed and management implications

Qian Zhang[1], Joel Bostic[2], Robert Sabo[3]

[1] University of Maryland Center for Environmental Science / EPA Chesapeake Bay Program

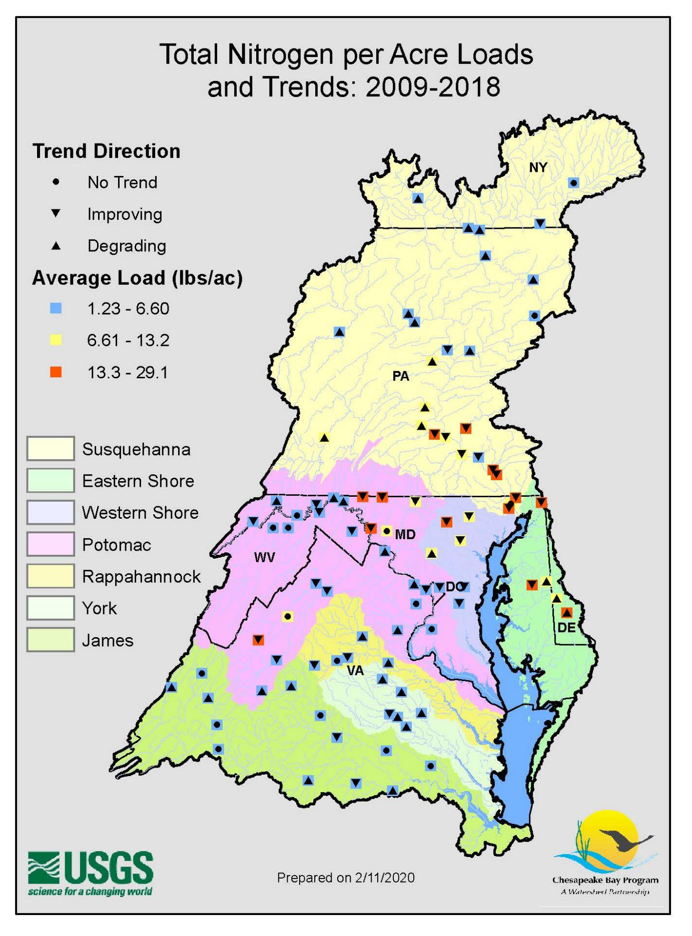
[2] University of Maryland Center for Environmental Science, Appalachian Laboratory

[3] U.S. Environmental Protection Agency

AGU FALL MEETING

New Orleans, LA & Online Everywhere

13–17 December 2021



Motivations

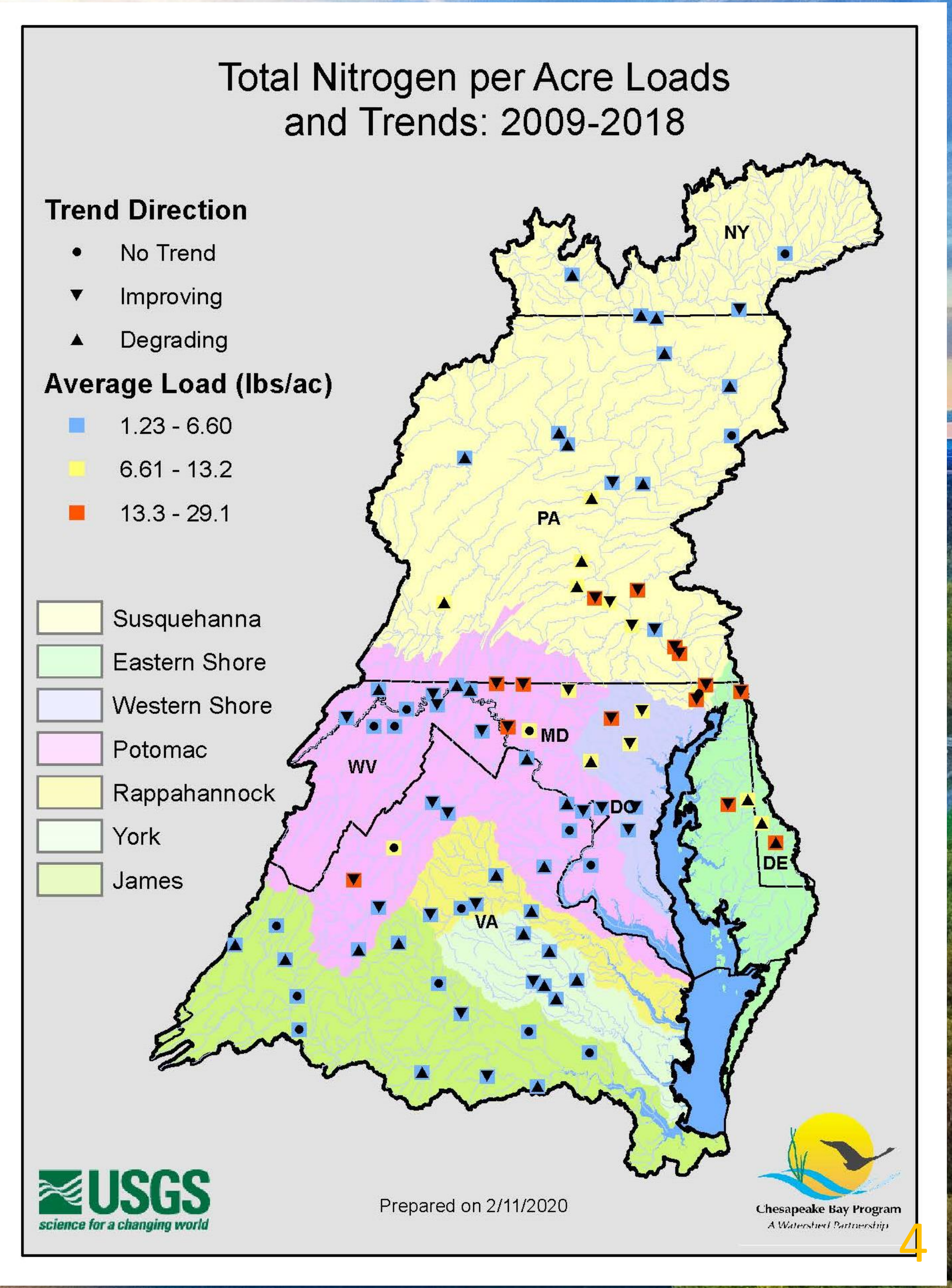
- River water-quality (WQ) trend studies often focus on one or a few monitoring locations, making conclusions difficult to generalize.
- Much can be learned from the similarity in WQ signals and the similarity in WQ responses to natural and anthropogenic drivers, which is made possible by data from regional monitoring networks.
- While many studies are aimed at the long-term scale (~30 years), short-term analysis can leverage data from newly established stations and provide relatively current information.
- Monitoring networks (i.e., CBNTN) do not often cover the entire watershed, leading to missing information in certain regions.
- Prior analyses of drivers do not always evaluate all major input sources, leading to potentially inaccurate or even contradicting inferences.

Objective

To reveal regional patterns and drivers of nitrogen trends using advanced machine learning approaches -- combined use of hierarchical clustering and random forest (RF).

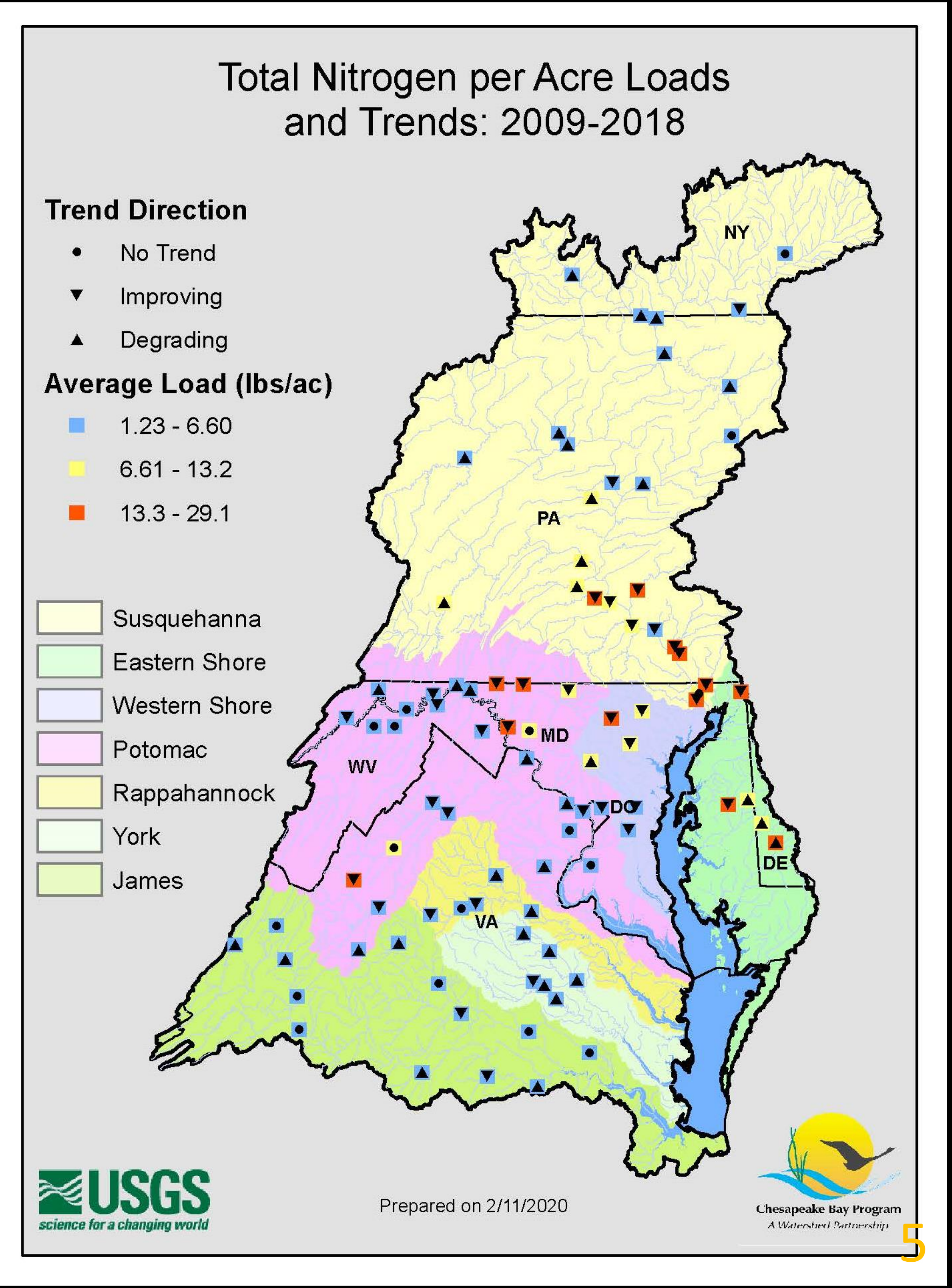
1. **Clustering**: Categorize the short-term (2007-2018) TN trends at the Chesapeake NTN stations (84) into distinct clusters,
2. **Classification**: Develop random forest (RF) models to identify the most influential drivers for the cluster assignment, and
3. **Prediction**: Use the RF model to predict short-term trend clusters for the entire watershed at a fine spatial resolution.

1. Regional patterns of nitrogen trends in the Bay watershed (Clustering)



CBNTN stations and TN data

- CBNTN watersheds (n = 84)
- 2007-2018 TN flow-normalized (FN) loads
- Standardized for each station (mean = 0, sd = 1)

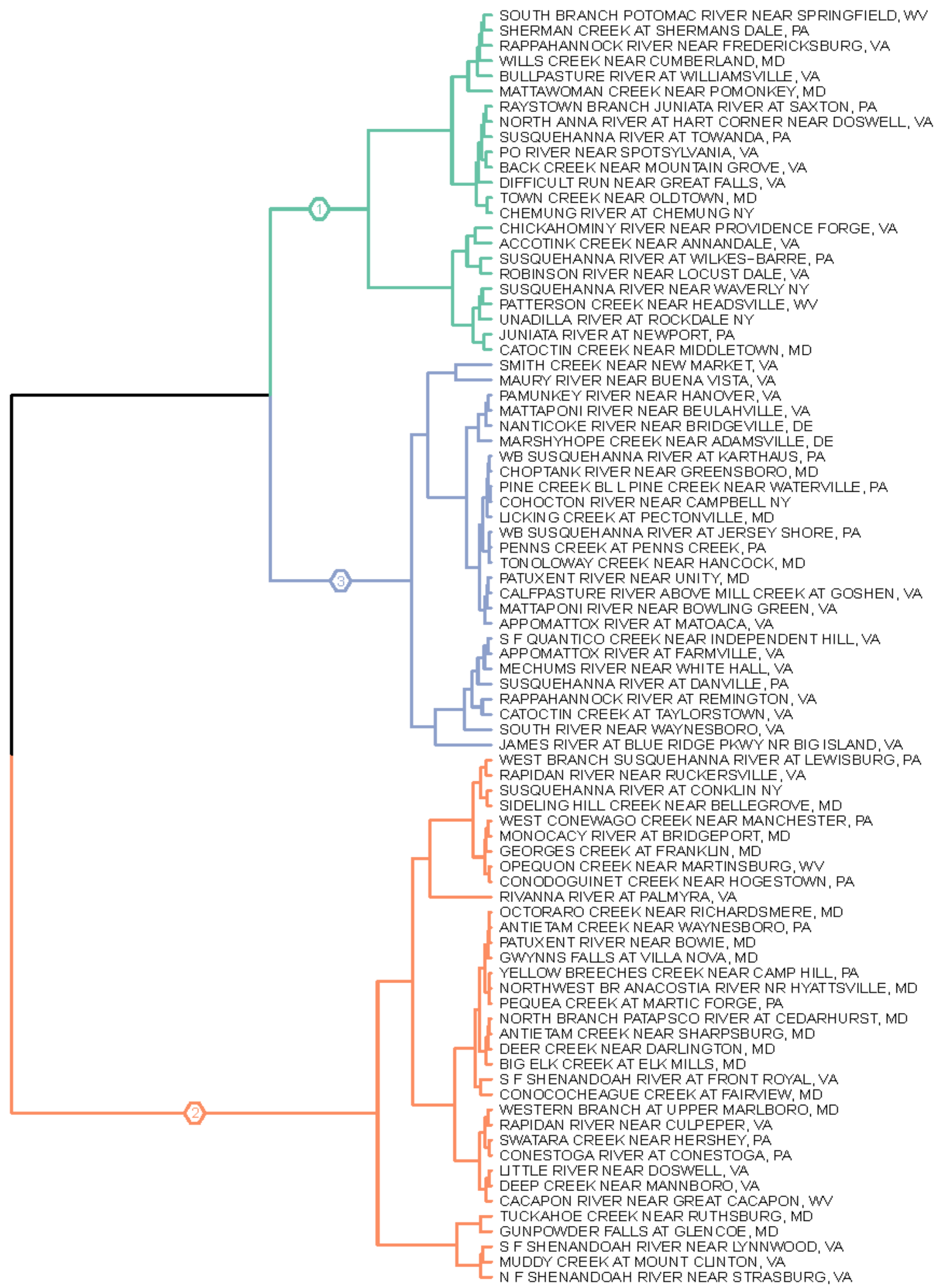


Hierarchical cluster analysis

Dissimilarity method:
Euclidean distance

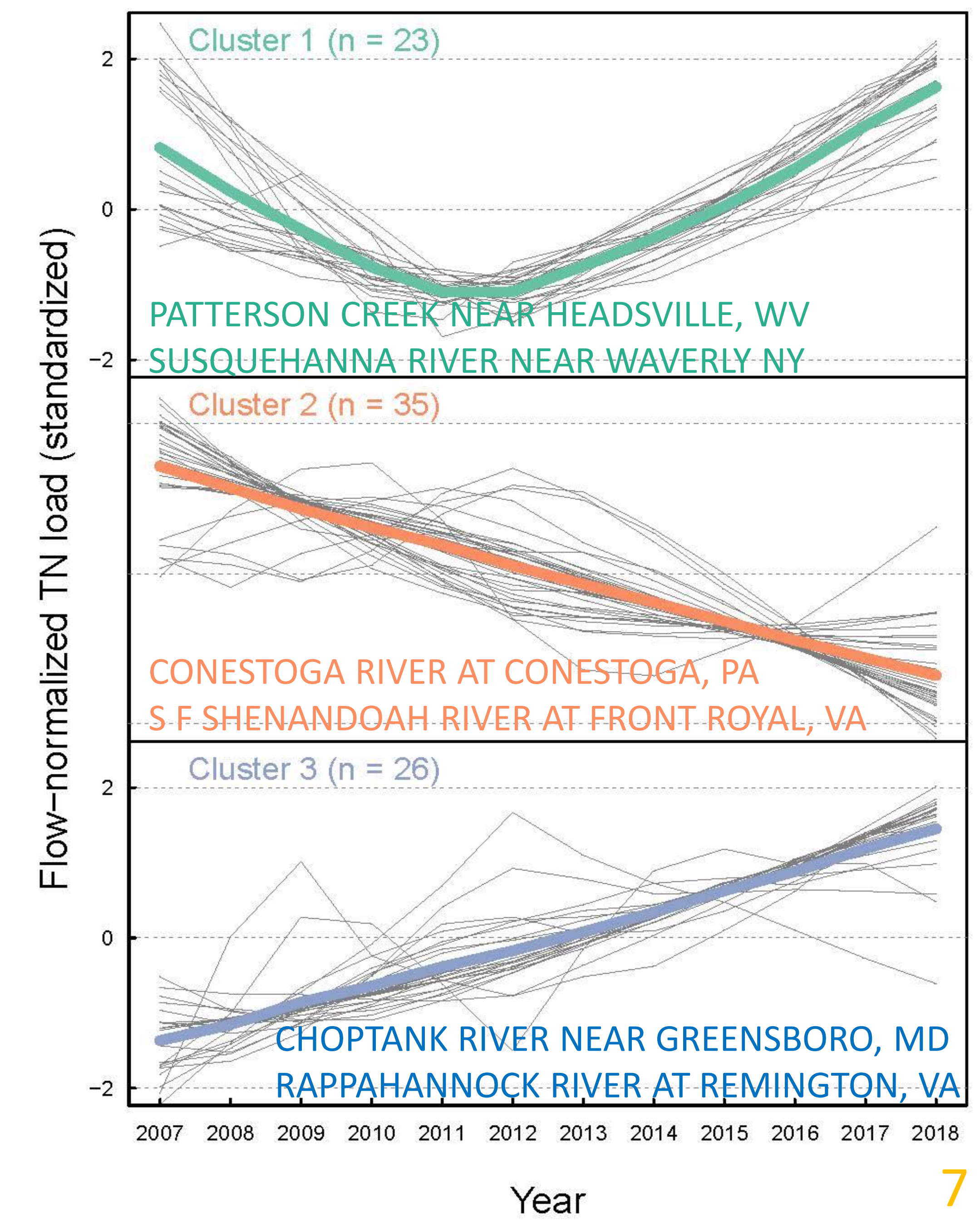
Linkage method:
Ward’s minimum variance method

Optimal cluster number:
Total Within Sum of Square



Hierarchical cluster analysis

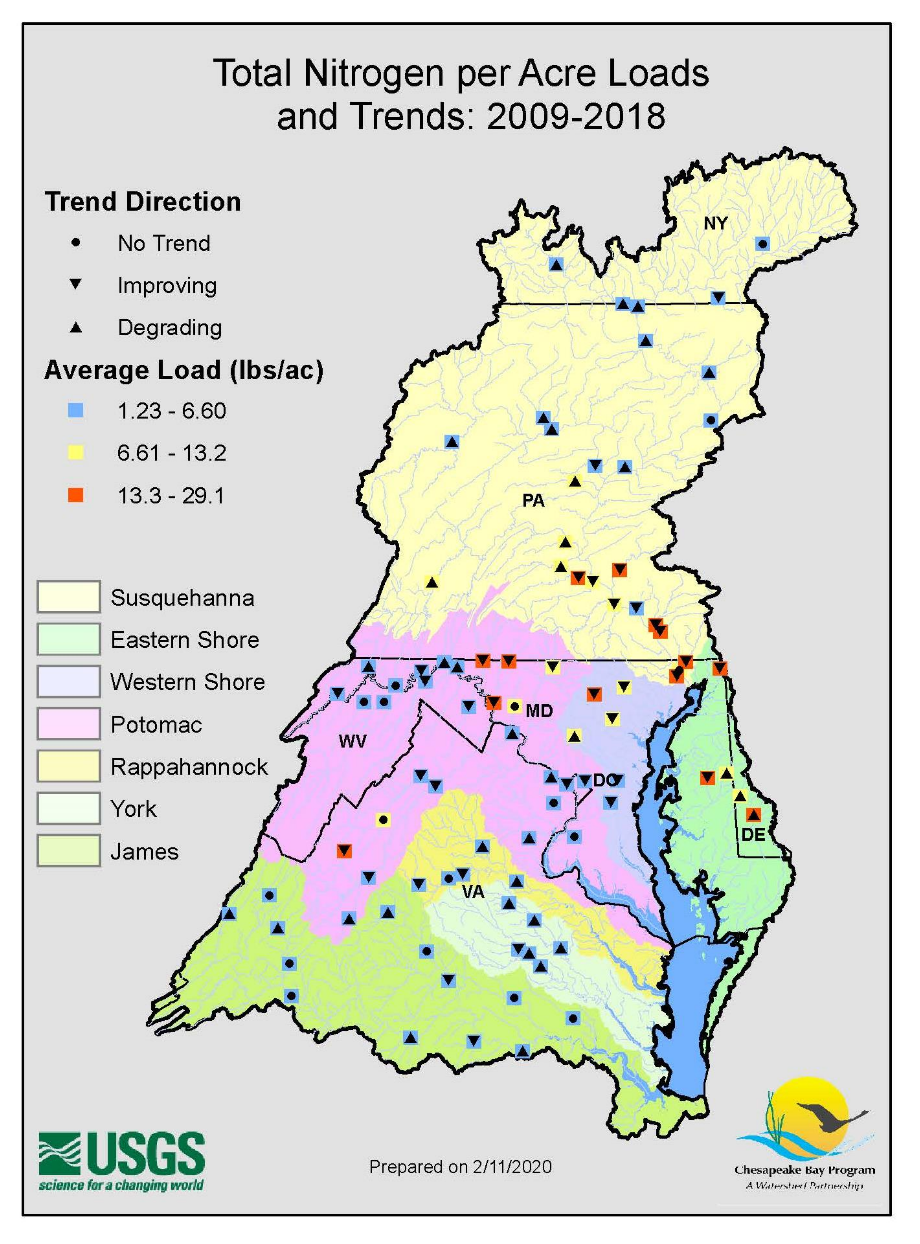
- Cluster 1 (n = 23):
  a V-shape trajectory.
- Cluster 2 (n = 35):
  a monotonic decline.
- Cluster 3 (n = 26):
  a monotonic increase.



Sensitivity Analysis

- 1/12 of the stations (n = 7) were removed without replacement.
- The remaining stations (n = 77) were reanalyzed using the same procedure.
- The number of clusters was set at three to be consistent.
- Cluster assignments are almost always consistent among the iterations.

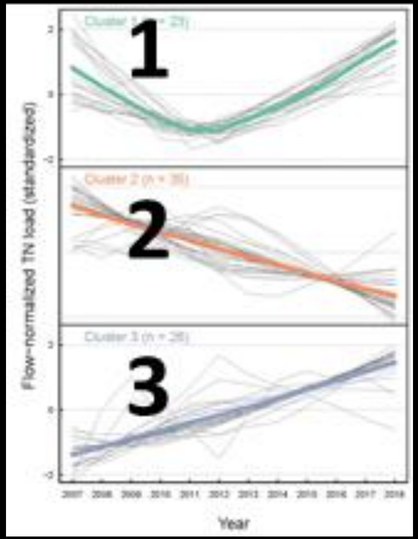
2. Regional drivers of nitrogen trend clusters (Classification)

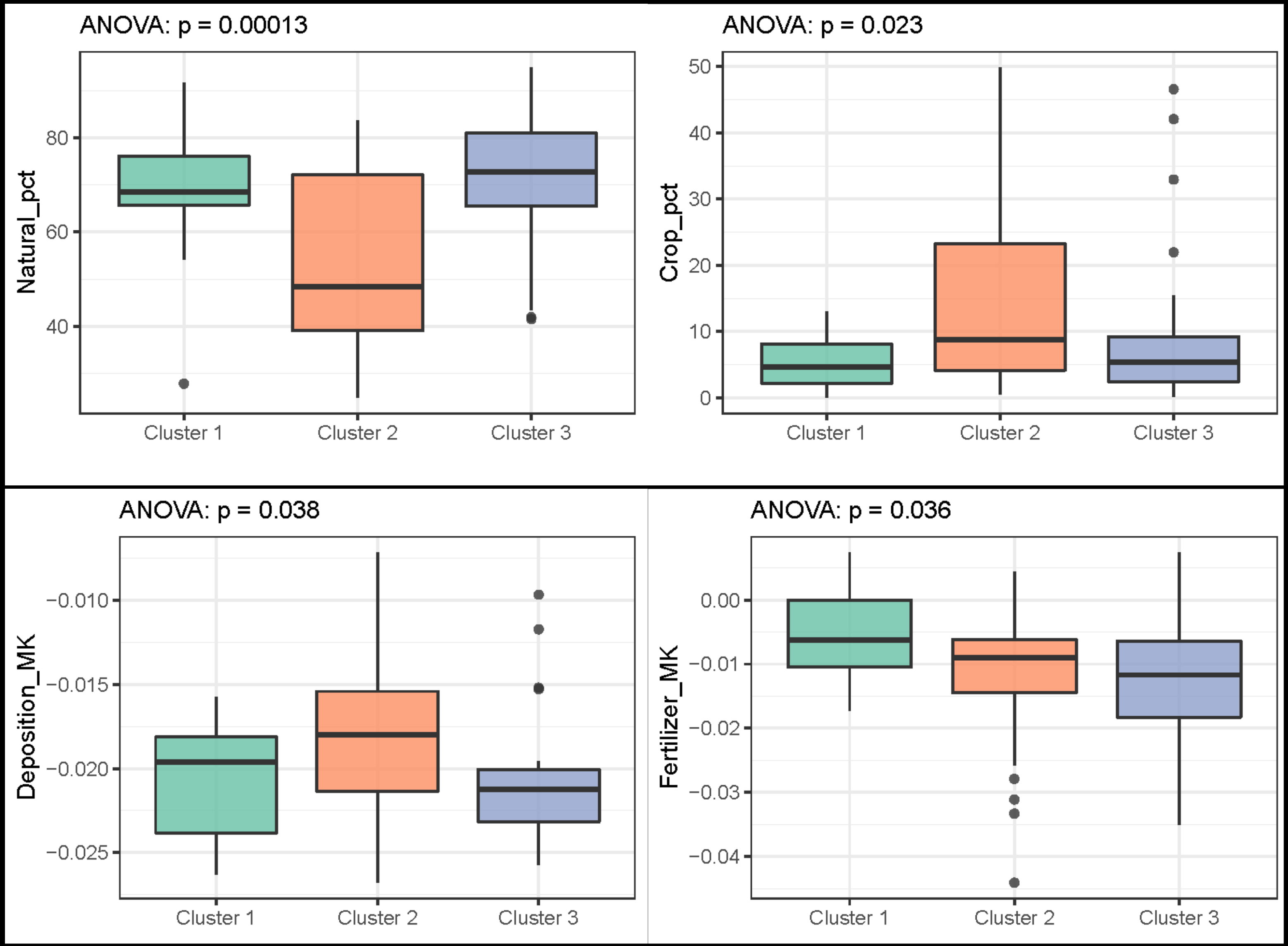


Explanatory Variables (Features)

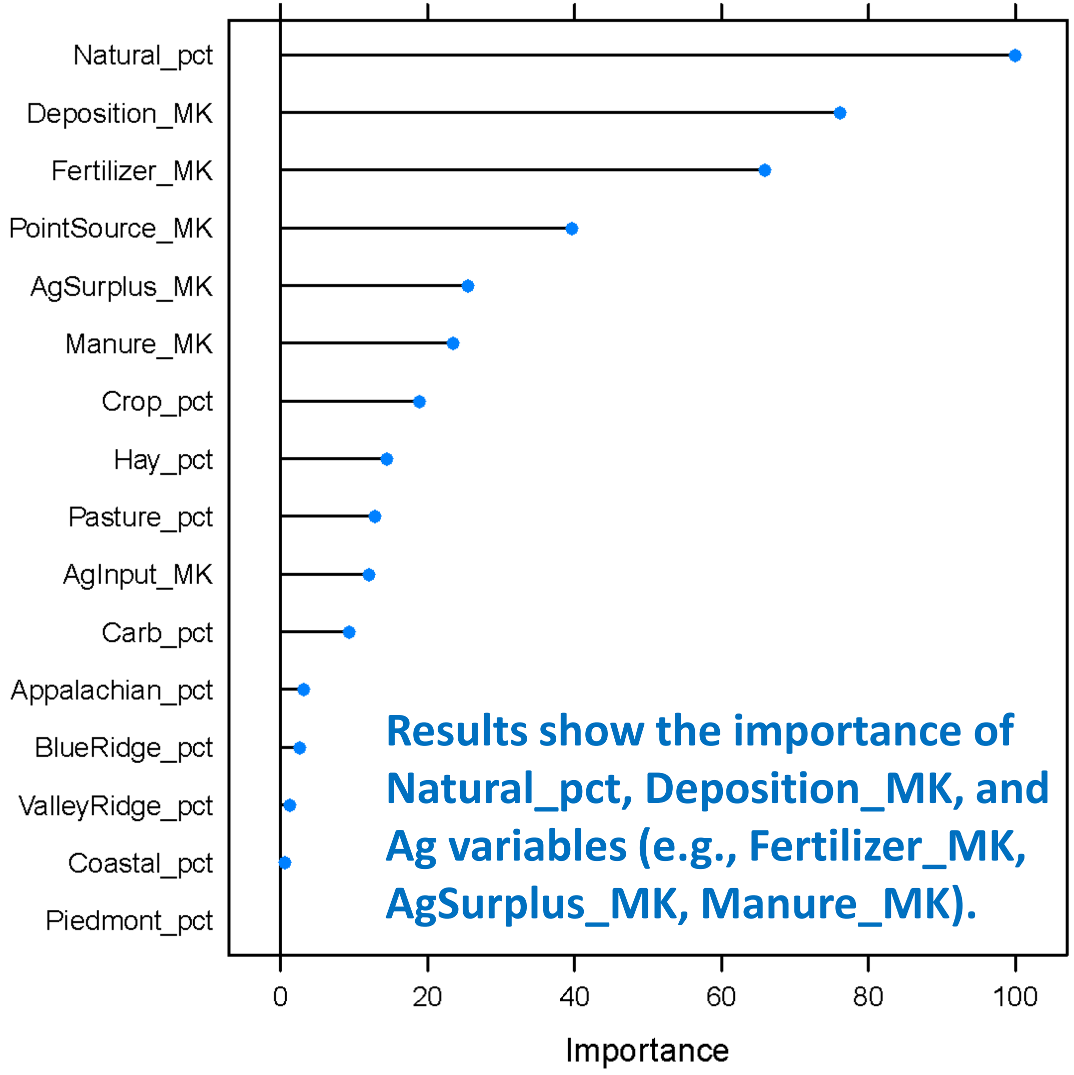
- Watershed size (n = 1) - Area_km2
- Land uses, in % (n = 4) - Natural_pct, Crop_pct, Pasture_pct, Hay_pct
- Geology, in % (n = 1) - Carb_pct
- Physiography, in % (n = 5) - Appalachian_pct, BlueRidge_pct, ValleyRidge_pct, Piedmont_pct, Coastal_pct
- N input source trends (n = 6) - PointSource_MK, Deposition_MK, Fertilizer_MK, Manure_MK, AgInput_MK, AgSurplus_MK
    1. CAST data aggregated for each NTN watershed – 2007-2018 for point sources; 1997-2018 for nonpoint sources.
    2. Annual time series scaled by respective period-of-record medians.
    3. Mann-Kendall trend and Sen's slopes computed.

Explanatory Variables (Features)

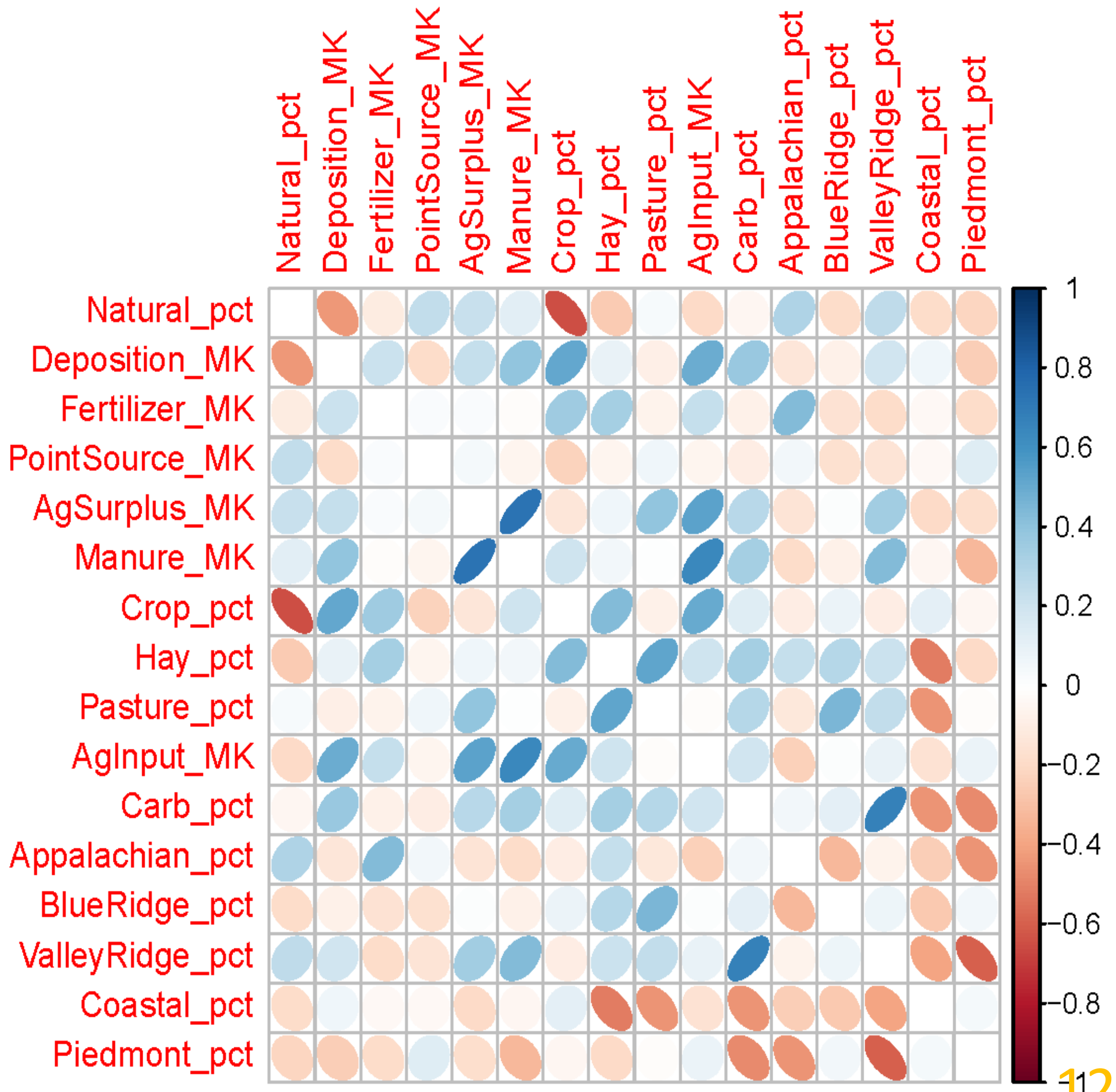




Random Forest (Base Model)



Results show the importance of Natural_pct, Deposition_MK, and Ag variables (e.g., Fertilizer_MK, AgSurplus_MK, Manure_MK).



Exhaustive Search for Optimal Models (n ≤ 6)

Model	Model form	OOB accuracy, percent			
		Overall	*Cluster1*	*Cluster2*	*Cluster3*
A	Class ~ Natural_pct + Fertilizer_MK + ValleyRidge_pct + Deposition_MK + Carb_pct	70.5	66.7	68.8	**76.0**
B	Class ~ AgSurplus_MK + Fertilizer_MK + Deposition_MK + Natural_pct	70.5	66.7	**75.0**	68.0
C	Class ~ BlueRidge_pct + Deposition_MK + Coastal_pct + Crop_pct + Fertilizer_MK + Natural_pct	69.2	**81.0**	65.6	64.0

The selected models have varying accuracies for each cluster, indicating that each model settled on a specific set of features that are most useful to explain a specific cluster. To make predictions, an ensemble model approach was adopted to combine the strengths of these three models – i.e., choosing the prediction with the highest probability from the three models.

Regional Drivers

Message 1 (AgSurplus_MK, Fertilizer_MK):

- Agricultural nutrient management contributed to water quality improvement.

Message 2 (Carb_pct, Coastal_pct):

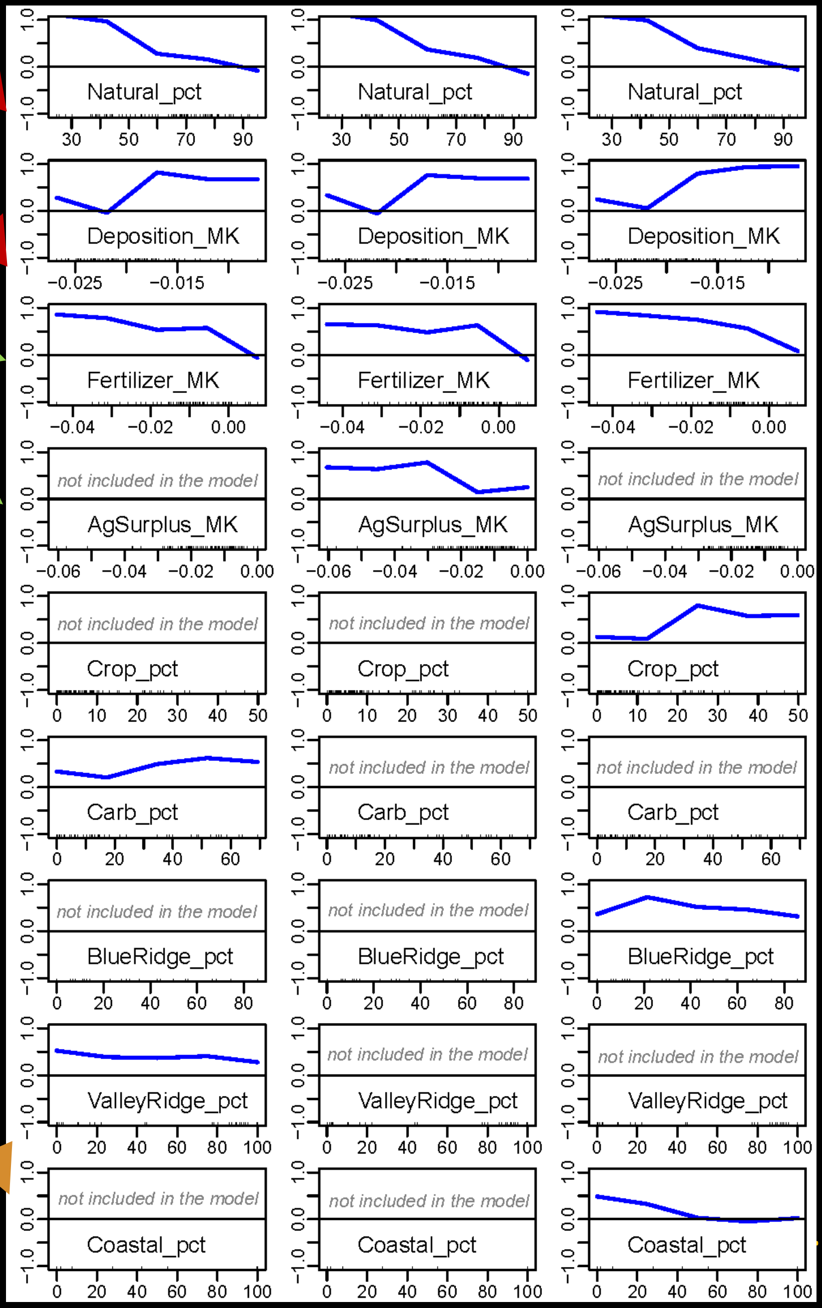
- Water-quality improvements are more likely in carbonate areas (relatively quick infiltration and faster groundwater transport) but less likely in Coastal Plain areas (accumulations of legacy N in the groundwater).

Message 3 (Natural_pct, Deposition_MK):

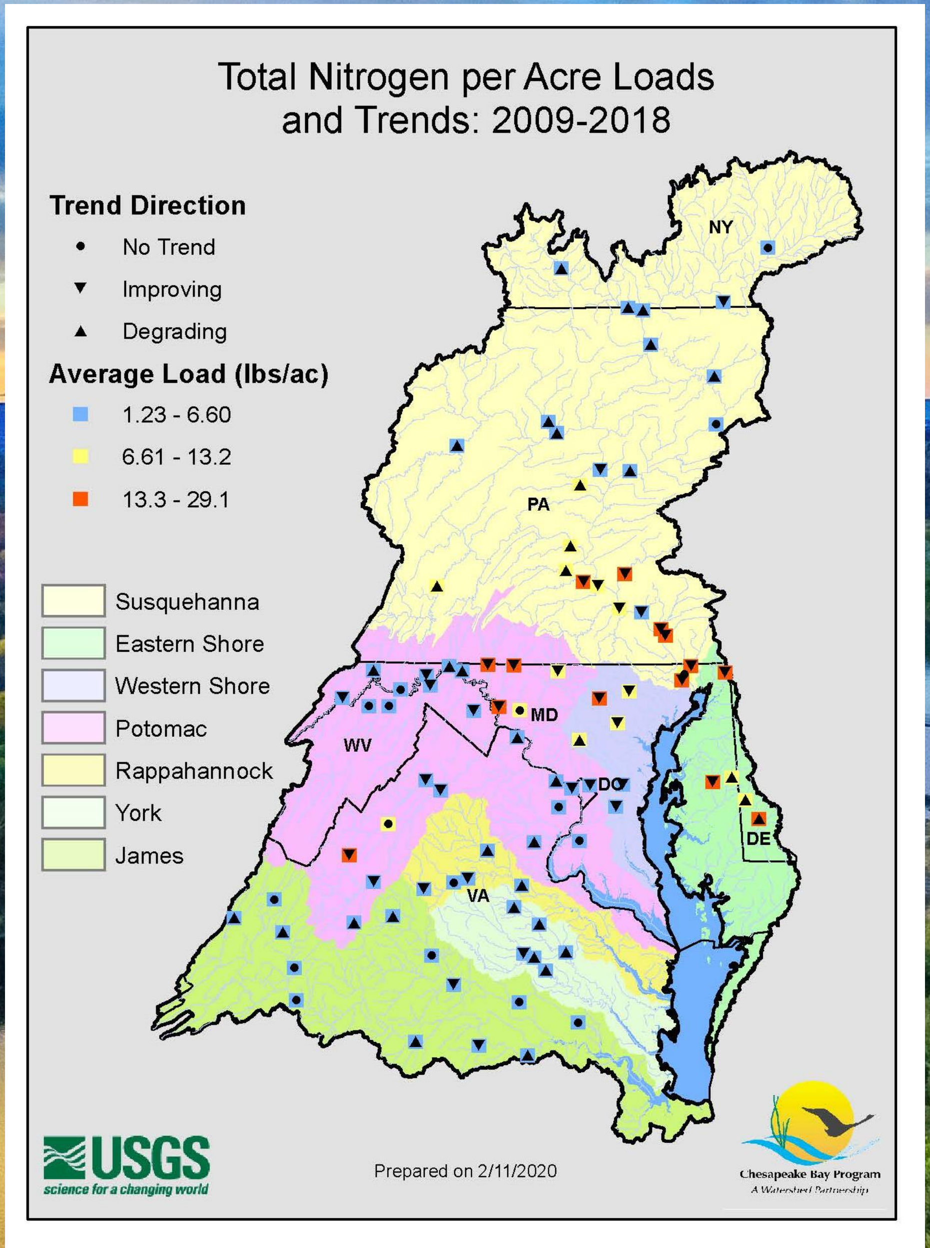
- We speculate that recent trends of increased TN in forested watersheds are attributed to: (1) increasing N inputs to non-forest regions and (2) mobilization of N from internal pools possibly due to deacidification and/or warming of forest soils.

Marginal Effects of Features on Cluster 2

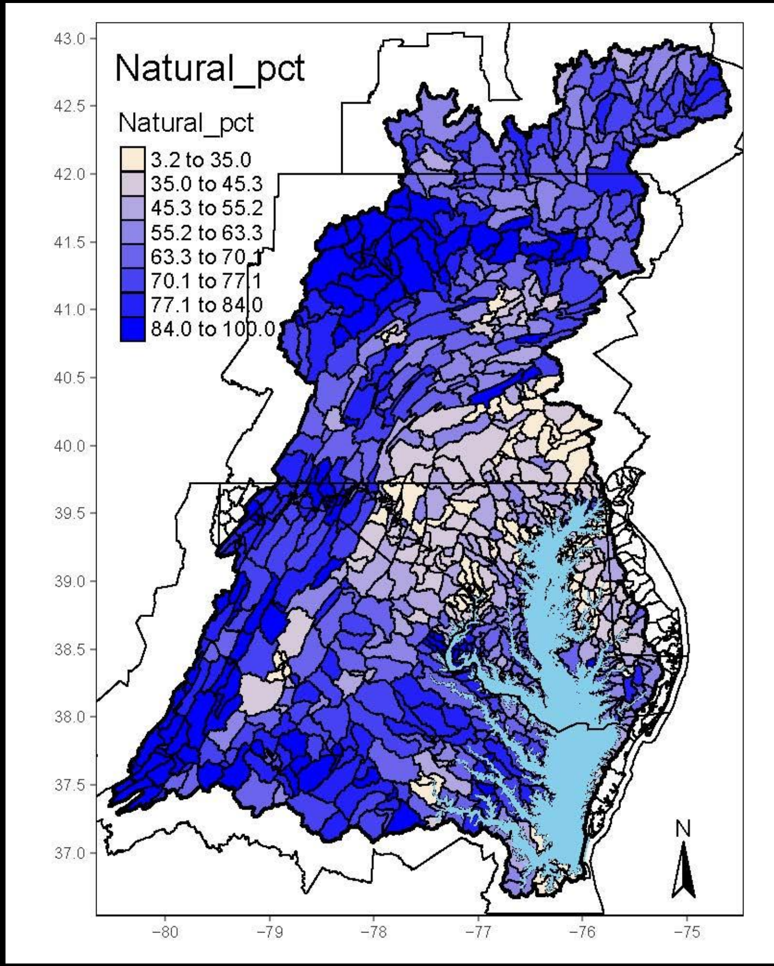
Model A *Model B* *Model C*

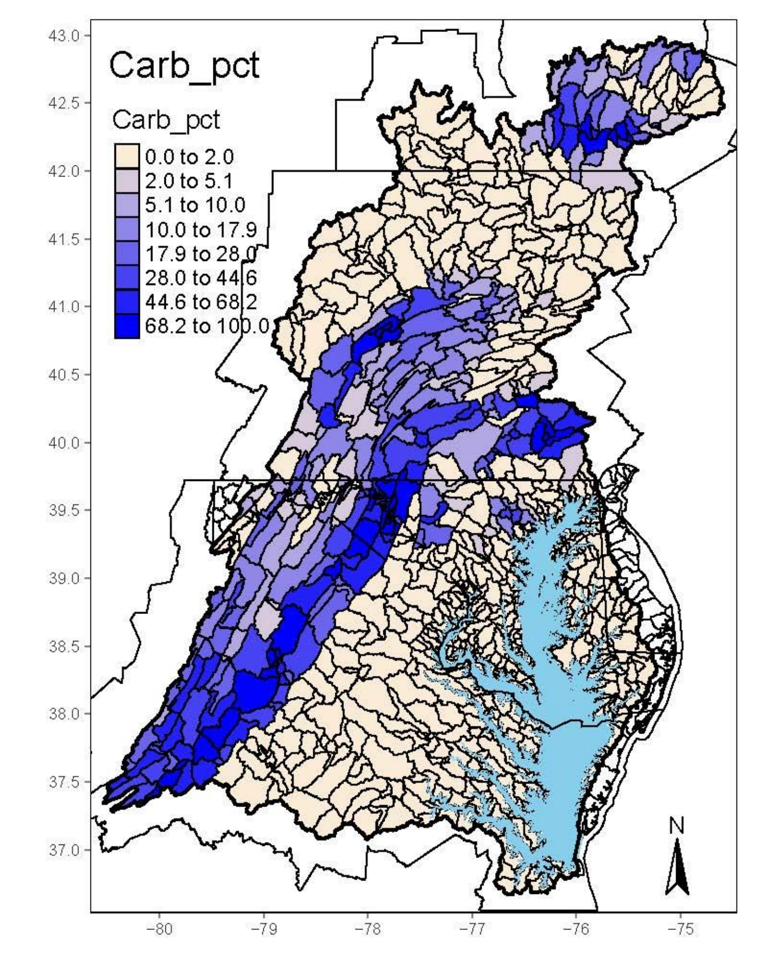


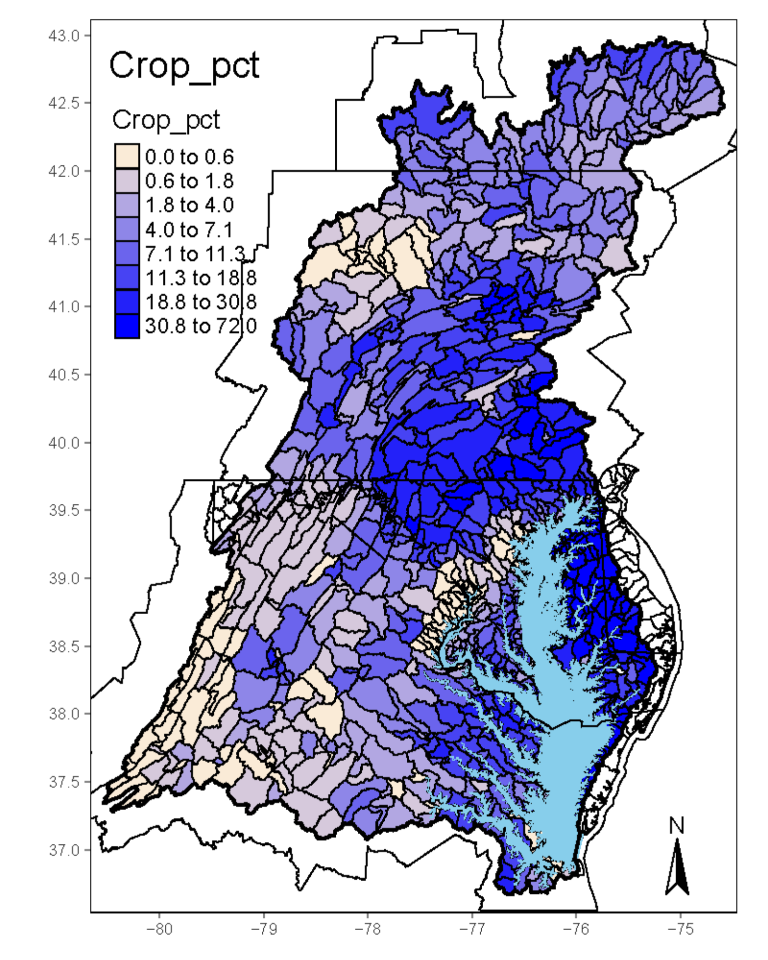
3. Prediction of nitrogen trend clusters for the entire watershed (Prediction)



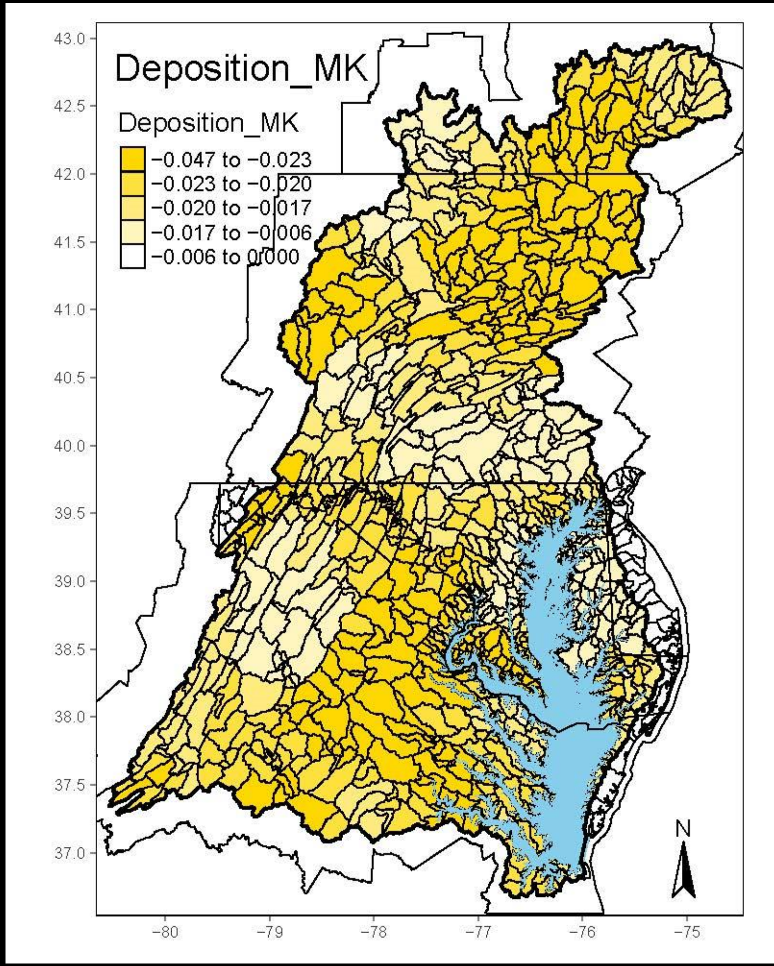
Explanatory Variables for River Segments

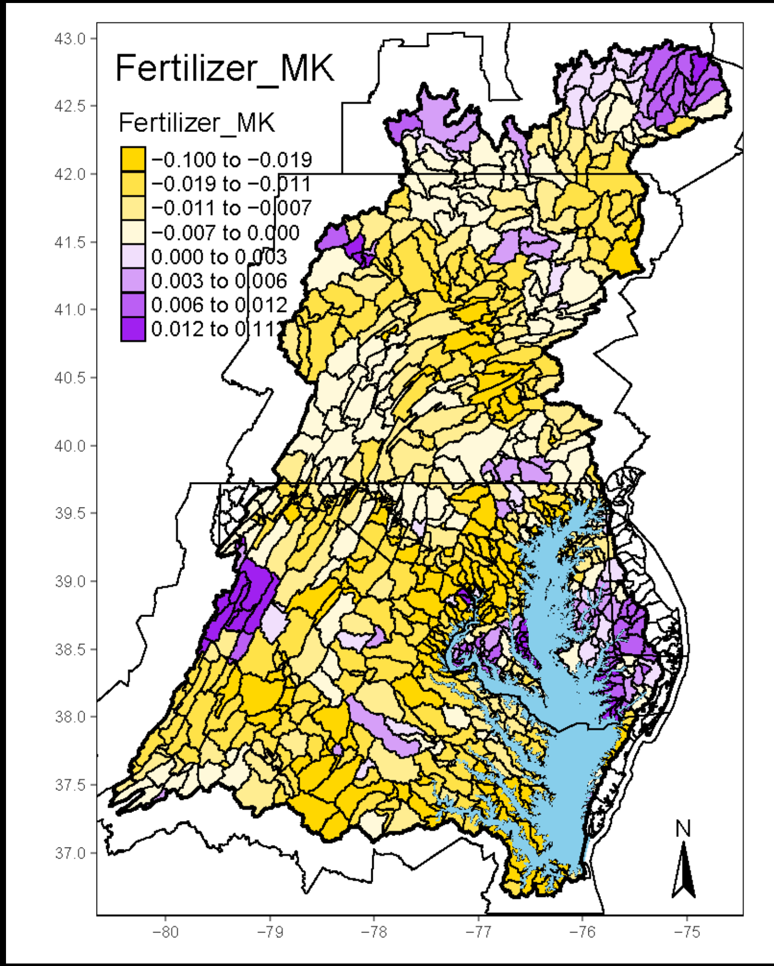


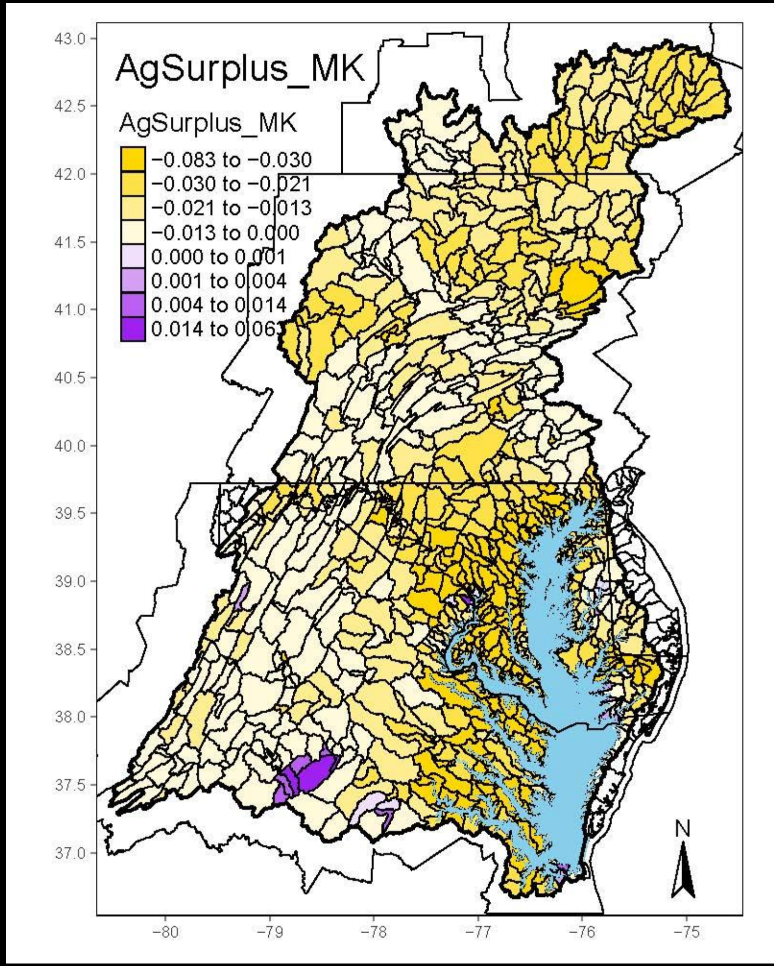




Explanatory Variables for River Segments



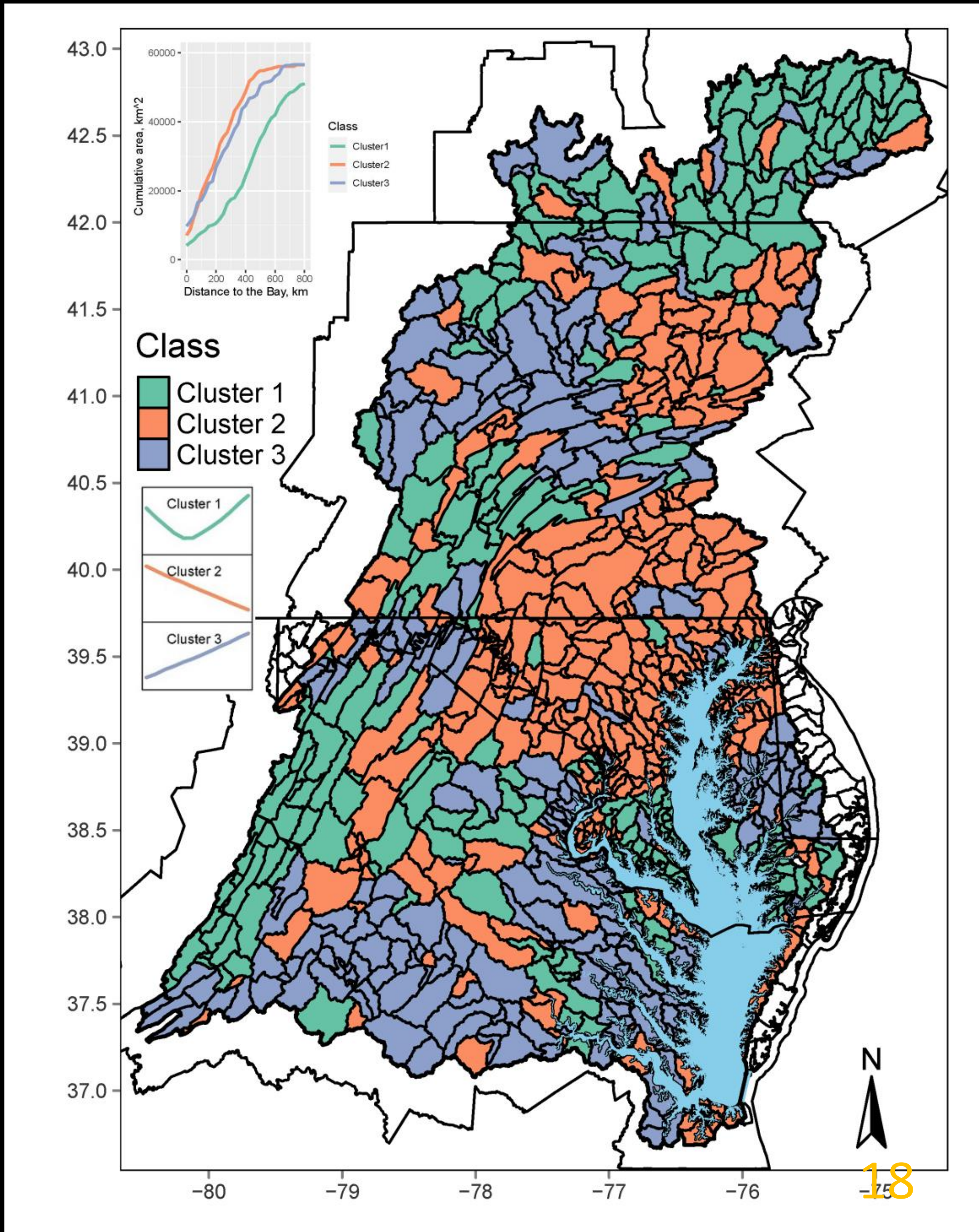




Predictions for River Segments

Cluster	No. of Segments	High Likelihood	Medium Likelihood	Low Likelihood
Cluster 1	293 (30%)	103	138	51
Cluster 2	392 (40%)	227	122	43
Cluster 3	295 (30%)	128	117	50

- These predictions are useful for watershed managers to understand trends across the watershed, including unmonitored areas.
- Combined with the effects of the model features, these predictions may inform managers on choosing priority watersheds toward water-quality improvement.



Conclusions

- Machine learning approaches – i.e., hierarchical clustering and random forest – can be combined to better understand the regional patterns and drivers of TN trends in large river monitoring networks.
- We explicitly incorporated temporal trends in agricultural fertilizer, manure, and agricultural input as well as agricultural surplus, providing evidence that improved nutrient management has resulted in declines in agricultural nonpoint sources, which in turn contributed to water quality improvement.
- Water-quality improvements are more likely in watersheds underlain by carbonate rocks but less likely in watersheds in the Coastal Plain.
- Results show degrading trends in forested watersheds, suggesting new and/or remobilized sources of N.
- Although we aimed for parsimony, models may be improved with additional features, e.g., management practice, legacy N, and riparian buffers.

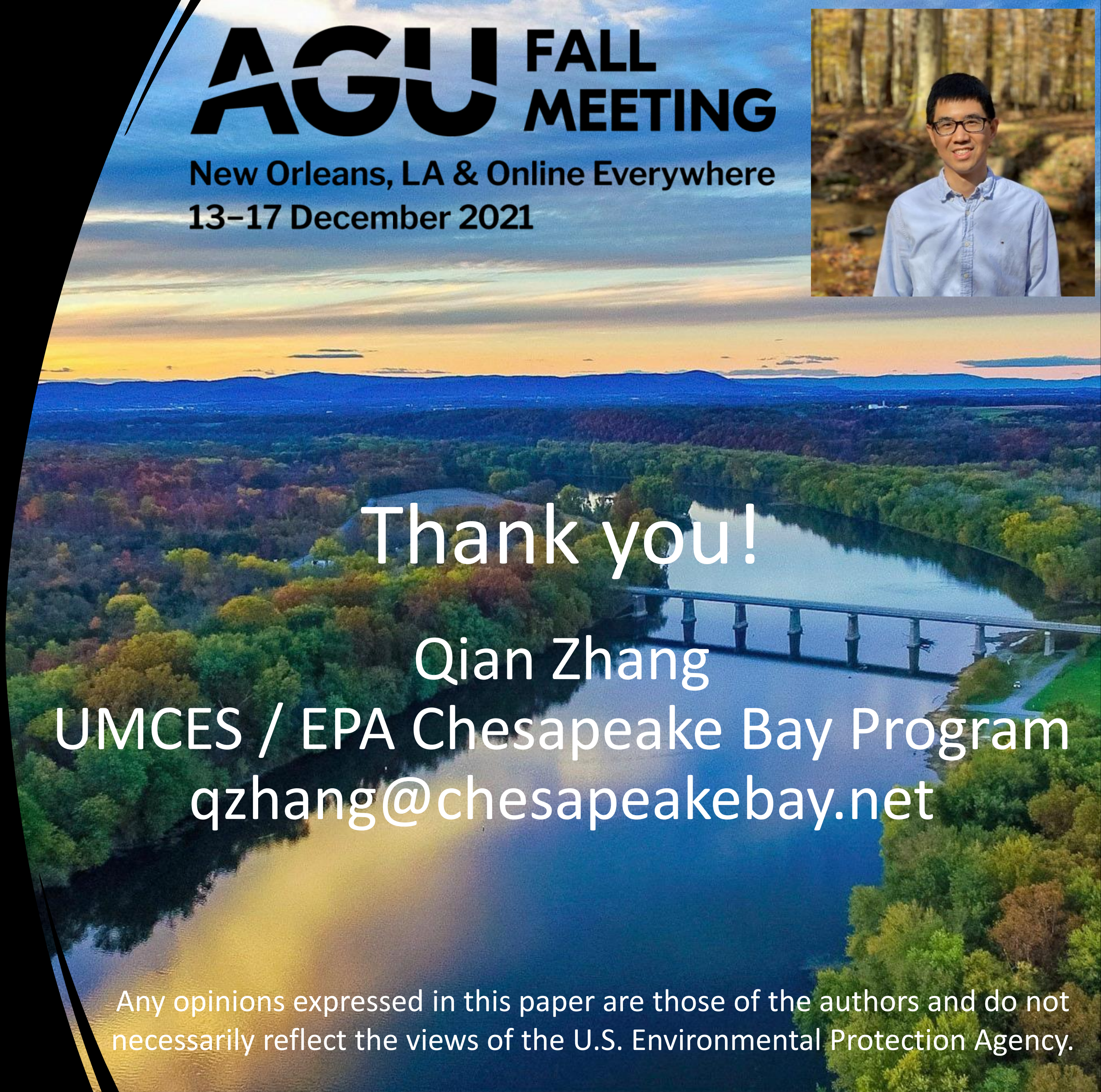
Regional patterns and drivers of nitrogen trends in a human-impacted watershed and management implications

Qian Zhang[1], Joel Bostic[2], Robert Sabo[3]

[1] University of Maryland Center for Environmental Science / EPA Chesapeake Bay Program

[2] University of Maryland Center for Environmental Science, Appalachian Laboratory

[3] U.S. Environmental Protection Agency



REGIONAL PATTERNS AND DRIVERS OF NITROGEN TRENDS IN A HUMAN-IMPACTED WATERSHED AND MANAGEMENT IMPLICATIONS

QIAN ZHANG[1,2], JOEL BOSTIC[1], ROBERT SABO[3]

[1] UNIVERSITY OF MARYLAND CENTER FOR ENVIRONMENTAL SCIENCE

[2] CHESAPEAKE BAY PROGRAM

[3] U.S. ENVIRONMENTAL PROTECTION AGENCY

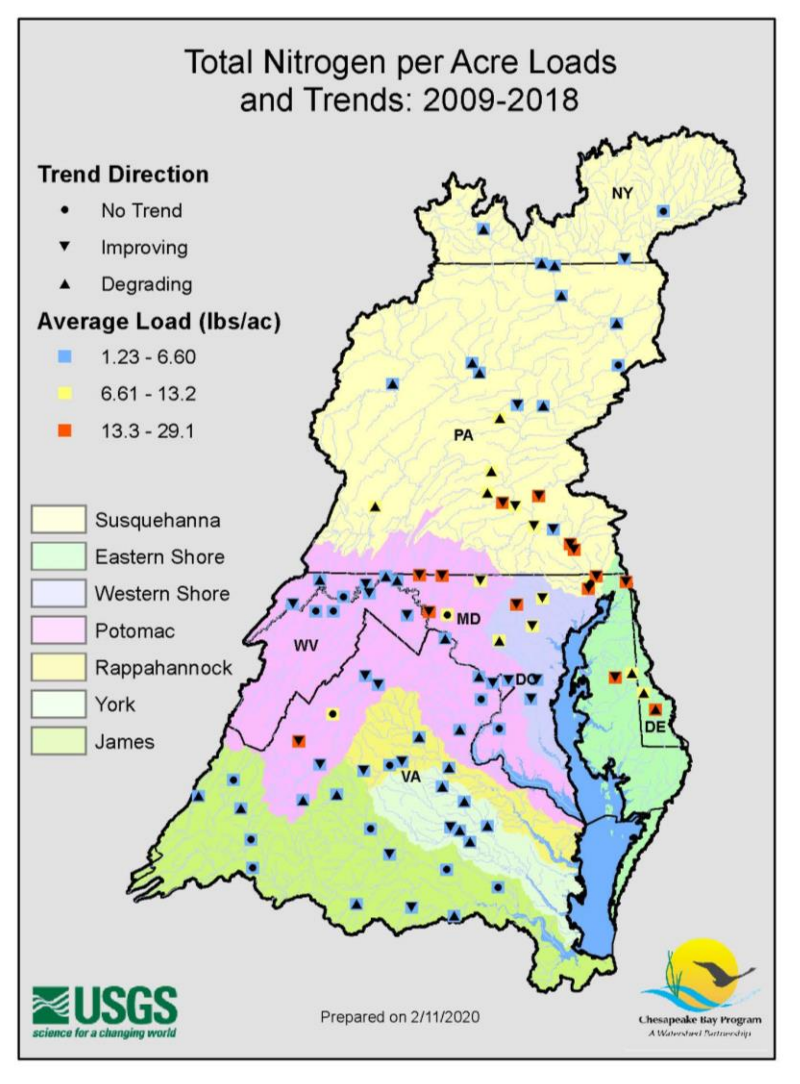
AGU FALL MEETING

Disclaimer: Any opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the U.S. Environmental Protection Agency.

AGU FALL MEETING

OBJECTIVE AND MOTIVATIONS

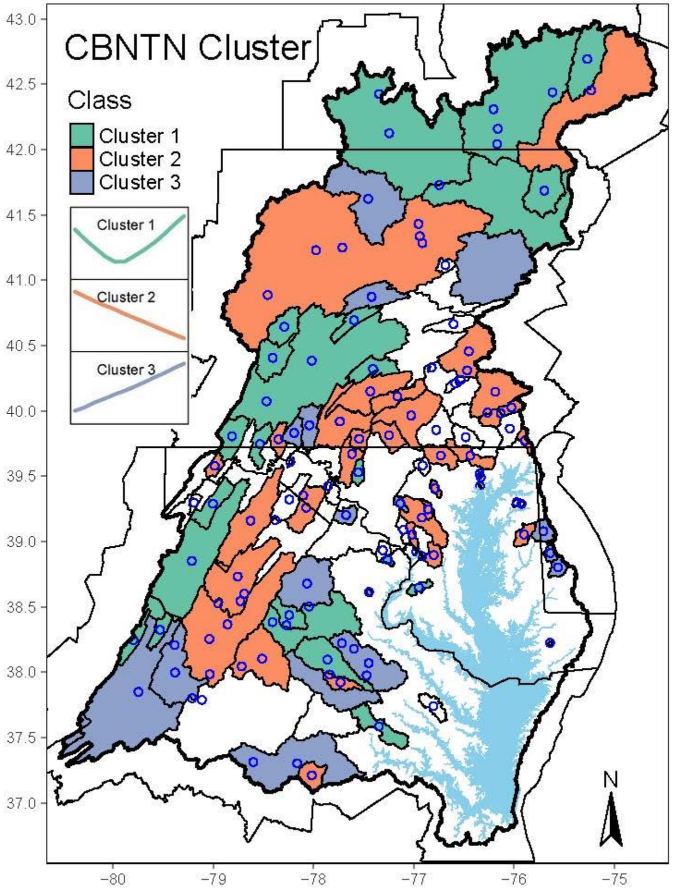
- To reveal regional patterns and drivers of total nitrogen (TN) trends using advanced machine learning approaches -- combined use of hierarchical clustering and random forest (RF).
  - ❑ Cover the Nontidal Monitoring Network (NTN).
  - ❑ Examine the similarity in TN trend signals and responses to natural and anthropogenic drivers.
  - ❑ Analyze short-term trends in order to incorporate newly established stations.
  - ❑ Incorporate important Agricultural variables.
  - ❑ Provide predictions for unmonitored areas.

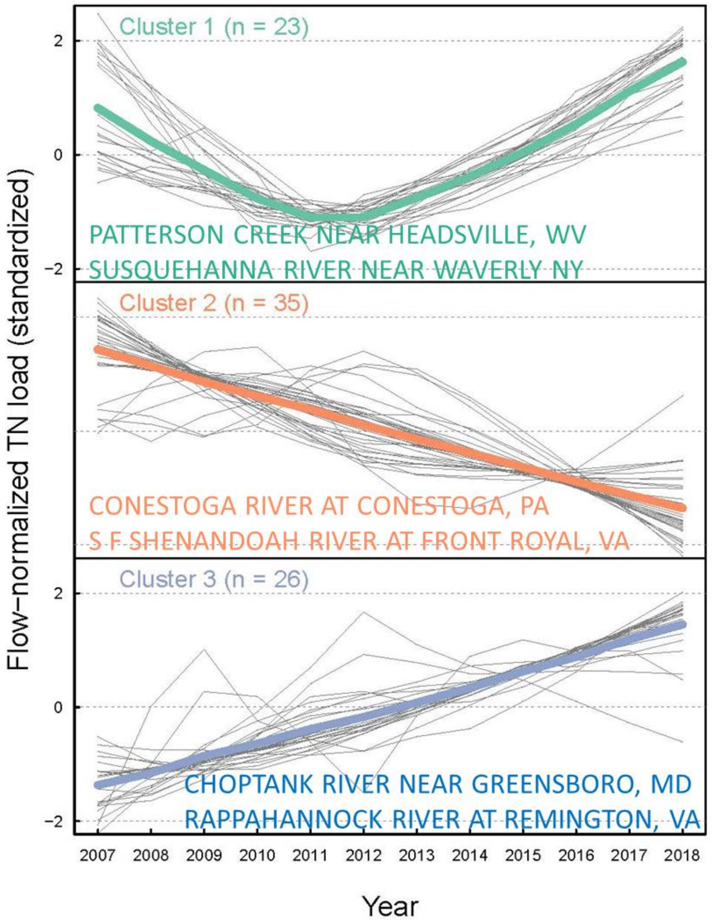


AGU FALL MEETING

1. REGIONAL PATTERNS OF TN LOAD TRAJECTORY (CLUSTERING)

- We used hierarchical cluster analysis to categorize the short-term (2007-2018) TN trends at the Chesapeake NTN stations (84) into three distinct clusters.
- Cluster 2 (n = 35) represents a trajectory of long-term decline in TN.





2. REGIONAL DRIVERS OF TN TREND CLUSTERS (RANDOM FOREST)

- We developed an exhaustive search algorithm to identify random forest (RF) models that can explain the TN cluster assignment.
- Three RF models selected by the search algorithm each settled on a specific set of features that are most useful to explain a specific cluster.

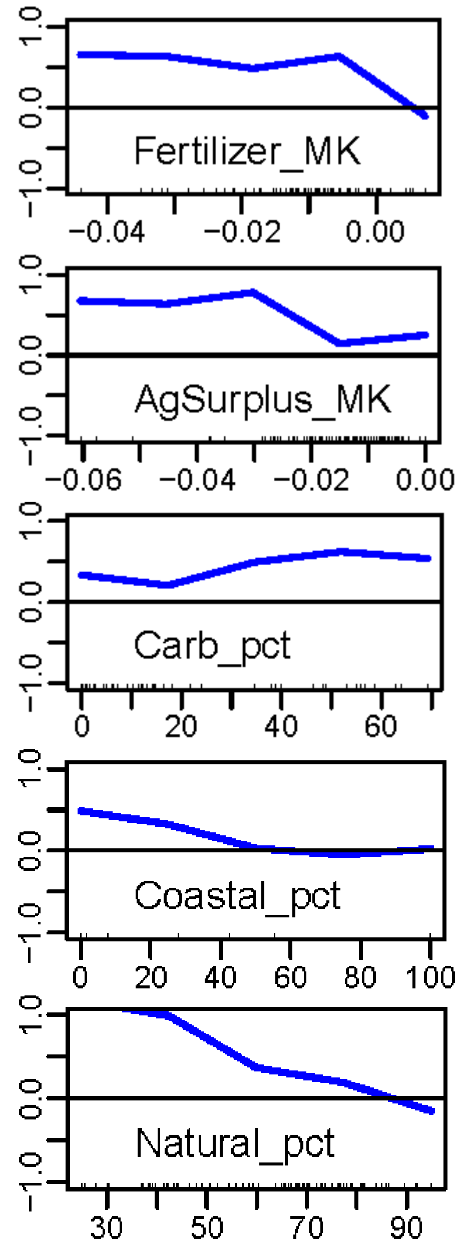
Model	Model form	OOB accuracy, percent			
		Overall	*Cluster1*	*Cluster2*	*Cluster3*
A	Class ~ Natural_pct + Fertilizer_MK + ValleyRidge_pct + Deposition_MK + Carb_pct	70.5	66.7	68.8	**76.0**
B	Class ~ AgSurplus_MK + Fertilizer_MK + Deposition_MK + Natural_pct	70.5	66.7	**75.0**	68.0
C	Class ~ BlueRidge_pct + Deposition_MK + Coastal_pct + Crop_pct + Fertilizer_MK + Natural_pct	69.2	**81.0**	65.6	64.0

AGU FALL MEETING

2. REGIONAL DRIVERS OF TN TREND CLUSTERS (RF)

- Improved nutrient management has resulted in declines in agricultural nonpoint sources, which in turn contributed to water quality improvement.
- Water-quality improvements are more likely to occur in watersheds underlain by carbonate rocks, reflecting the relatively quick groundwater transport of this terrain.
- By contrast, water-quality improvements are less likely to occur in watersheds in the Coastal Plain, reflecting the effect of legacy N in groundwater.
- Results show degrading trends in forested watersheds, suggesting new and/or remobilized sources of N that may compromise downstream watershed restoration plans more focused on agricultural and urban areas.

MARGINAL EFFECTS OF SELECTED FEATURES ON CLUSTER 2

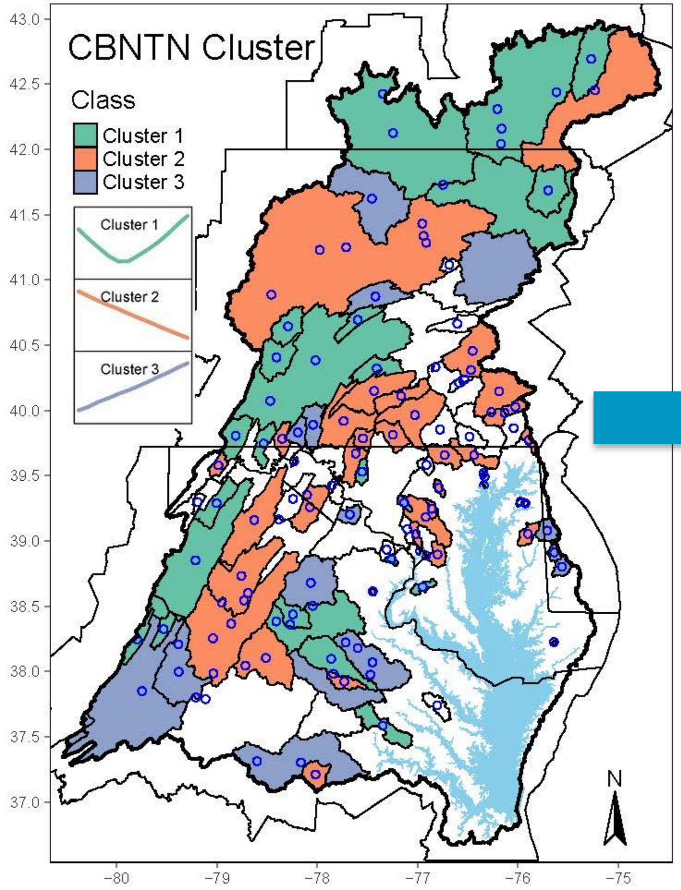


AGU FALL MEETING

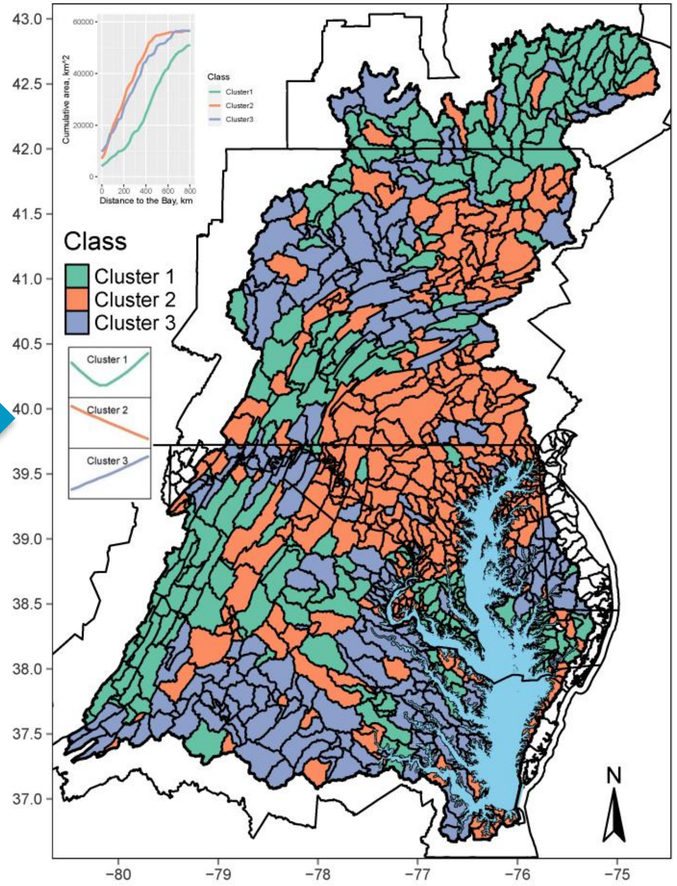
3. PREDICTIONS OF TN TREND CLUSTERS FOR THE ENTIRE WATERSHED

- We applied the RF models to predict short-term trend clusters for the entire Bay watershed at a fine spatial scale (i.e., river segments).
- These predictions are useful for managers to understand trends across the watershed, including unmonitored areas, and to choose priority watersheds toward water-quality improvement.

Observations



Predictions



THANK YOU

Qian Zhang
UMCES / EPA Chesapeake Bay Program
qzhang@chesapeakebay.net
qzhang@umces.edu

AGU FALL MEETING