

# EPA Tools and Resources Webinar: Nutrient Explorer

An analytical framework to visualize and investigate drivers of surface water quality

Michael Pennino, Meridith Fry Robert Sabo and James Carleton US EPA Office of Research and Development

March 13, 2024

**\$EPA** 

#### Outline

- Background / Motivation
- Purpose
- Features
- Demonstration
  - Default dataset
  - New user dataset
- Conclusions & Impacts



#### **Background and Motivation**

 Excess nutrients (nitrogen and phosphorus) in lakes/rivers can lead to eutrophication, hypoxia, and algal blooms that may harm aquatic life and people.

**S**EPA

- Some U.S. states have established numeric water quality criteria for nutrients to protect surface waters.
- However, monitoring to determine if criteria are being met is limited by resources and time.





https://www.canr.msu.edu/



- Design an analytical framework to investigate potential drivers of nutrient concentrations in surface waters (lakes, streams, rivers)
- Visualize patterns in nutrient concentrations using a graphical user interface
- Predict areas at risk of exceeding certain criteria, especially for locations lacking observations



## **Set EPA**

#### **Nutrient Explorer – Features**

- Upload total nitrogen (TN) or total phosphorus (TP) concentration data
  - Select default dataset for lakes in northeastern U.S.
  - Alternatively, upload a new, formatted dataset for any surface waters (lakes, rivers, streams)
- Incorporate variables to predict nutrients in surface waters (e.g., land use, etc.)
- Create data subsets for further analysis, using characteristics of the variables, such as date, location, etc.

,	AutoSave (	• Off	89	• 6.	~& <del>~</del>		LAGO	S_Test3.c	sv 🗸	∕∕ Se	arch					L	+	Pennir	10, Mich	iael 🧃	•	<u> </u>	o	×
F	ile H	ome	Insert	Drav	w Page	Layou	it Fo	rmulas	Data	Review	View	Help	Ar	cGIS							C C C	omments	🖻 Shar	e
\$1		-	: ×		fr inve	ha																		
				*																				
-	A		В	С	D	E	E	F	G	Н	- I	J		К	L	M	N	0	P		Q	R	S	44
1	Name	LAG	OS_La HL	J8Zonel	HU8	prog	ramn pr	ogramty	Lat	Long	HU2	Year	N	Month	Day	Date	Elevation	MaxDept	h Mear	Depila	ake_area	lake_perir	iws_ha	1
2	ST. JAM	ES	156 HU	J8_365	7020010	MN_	MPC/St	ate Ager	43.97355	-94.6478		7 2	010	10	6	10/6/2010	99999	99999	NA		99999	99999	1599.226	
3	ST. JAM	ES	156 HU	J8_365	7020010	MN_	MPC/Sta	ate Ager	43.97355	-94.6478		/ 2	011	4	21	4/21/2011	. 999999	99999	NA		99999	99999	1599.226	<u>.</u>
4	PEARL	N	002 HL	18_339	7010202		MPC/St	ate Ager	45.39893	-94.3064		7 2	011	4	18	4/18/2011	99999	99999			99999	99999	3495.133	
6	MADISC	IN IN	902 HU	18 366	7020011	MN	MPC/St	ate Ager	44.10074	-93.801		7 2	010	10	29	4/29/2010	00000	00000			00000	00000	921.9220	4
7	MADISC	N	902 HI	18 366	7020011	MN	MPC/St	ate Ager	44.10074	-93.801		7 2	011	4	21	4/21/2011	99999	99999	NA		99999	99999	921.9226	
8	SOUTH	°F	1061 HI	18 350	7030005	MN	MPC/St	ate Ager	45.37856	-92.8214		7 2	010	5	24	5/24/2010	99999	99999	NA		99999	99999	1278.457	,
9	SOUTH	CE	1061 HU	J8 350	7030005	MN	MPC/St	ate Ager	45.37856	-92.8214		7 2	011	4	20	4/20/2011	99999	99999	NA		99999	99999	1278.457	7
10	BELLE		1822 HU	J8 342	7010205	MN	MPC/St	ate Ager	44.9813	-94.4266		7 2	010	4	29	4/29/2010	99999	99999	NA		99999	99999	542.246	ز
11	BELLE		1822 HU	J8_342	7010205	MN_	MPC/Sta	ate Ager	44.9813	-94.4266		7 2	011	4	18	4/18/2011	99999	99999	NA		99999	99999	542.246	5
12	PORTAG	E	54660 HU	J8_335	7010106	MN_	MPC/Sta	ate Ager	46.96561	-95.1219		7 2	011	5	9	5/9/2011	99999	99999	NA		99999	99999	845.5872	£
13	CARRIE		65948 HL	J8_342	7010205	MN_	MPC/Sta	ate Ager	45.08241	-94.7864		7 2	010	7	20	7/20/2010	99999	99999	NA		99999	99999	247.3733	\$
14	CARRIE		55948 HU	J8_342	7010205	MN_	MPC/Sta	ate Ager	45.08241	-94.7864		7 2	010	10	4	10/4/2010	99999	99999	NA		99999	99999	247.3733	1
15	CARRIE		55948 HU	J8_342	7010205	MN_	MPC/Sta	ate Ager	45.08241	-94.7864		7 2	011	8	3	8/3/2011	99999	99999	NA		99999	99999	247.3733	1
16	CARRIE		55948 HL	J8_342	7010205	MN_	MPC/St	ate Ager	45.08241	-94.7864		7 2	011	10	26	10/26/2011	. 999999	99999	NA		99999	99999	247.3733	J.
17	ST. JAM	ES	156 HU	J8_365	7020010	MN_	MPC/Sta	ate Ager	43.97355	-94.6478		7 2	010	4	29	4/29/2010	99999	99999	NA		99999	99999	1599.226	
18	WHITEI	RC	260 HL	J8_130	9030001	. MN_	MPC/Sta	ate Ager	47.87035	-91.7922		9 2	011	5	10	5/10/2011	. 99999	99999	NA		99999	99999	6933.989	1
19	CEDAR		557 HL	J8_338	7010201	. MN_	MPC/Sta	ate Ager	45.81413	-94.6362		7 2	011	4	18	4/18/2011	. 99999	99999		11.7	99999	99999	376.8463	4
20	PEARL		687 HL	18_339	7010202	IVIN_	MPC/Sta	ate Ager	45.39893	-94.3064		/ 2	010	4	29	4/29/2010	99999	99999	NA		99999	99999	3495.133	1
21	ARTICHO	JK	758 HU	18_362	7020002	IVIN_	MPC/Sta	ate Ager	45.35549	-96.13/3		/ 2	011	4	19	4/19/2011	. 999999	99999	NA NA		99999	99999	2581.156	4
22	AKTICH	ЛК	758 HU	J8_362	/020002	IVIN_	IVIPC/Sta	ate Ager	45.35549	-90.13/3		/ 2	011	10	26	10/26/2011	. 99999	99999	INA		2222	99999	2581.156	41

MUGHICK SALATED STATED	·	Nutrien	t Expl	lorer	
TENTAL PROTECTIO	Open			×	
.oad Data Explore Data		Kenter	C Search Nut	trient P	
	Organize - New fold	ler	≣ .	• 🛯 😗	
Dataset options	A Home	Name	Status	Date mo	
O LAGOS test	RI Gallery	2021	$\odot$	11/11/20	
O LAGOS	Michael - Enviro	2022	6	6/22/202	
	, Children Ennie	Science_Inventory_Version	3	7/28/202	
Upload new dataset	Duralizate A	HUC8_2007_7_Average_MaxDepth2.csv	$\odot$	5/4/2023	
Browse LAGOS	👱 Downloads 🖈	HUC8_2007_7_Average_MaxDepth3.csv	$\odot$	5/13/202	
Upload complete	Pictures 🖈	HUC8_newData.csv	$\odot$	5/3/2023	
	💾 Documents 🖈	LAGOS_Test.csv	$\odot$	5/3/2023	
If uploading new	Projects 🖈	I LAGOS TEST DATA	0	0/27/202	
dataset, wait until it says	Scrinte 🔺				



## Nutrient Explorer – Features (cont.)

- Visualize and summarize temporal and spatial patterns
- Apply random forest modeling or multiple linear regression to:
  - Assess which predictor variables best capture the spatial patterns in surface water nutrients
  - Predict nutrient concentrations for watersheds or locations lacking data
- Map which watersheds or locations may be predicted to exceed a certain threshold or water quality criteria







## Nutrient Explorer – Features (cont.)

- Visualize and summarize temporal and spatial patterns
- Apply random forest modeling or multiple linear regression to:
  - Assess which predictor variables best capture the spatial patterns in surface water nutrients
  - Predict nutrient concentrations for watersheds or locations lacking data
- Map which watersheds or locations may be predicted to exceed a certain threshold or water quality criteria



# **\$EPA**

### **Default Datasets in Nutrient Explorer**

- <u>Surface Water Nutrient Data</u>: The app comes with a default dataset from the Northeastern U.S. lakes based on the Lake Multi-Scaled Geospatial and Temporal Database of the Northeast (LAGOS-NE) data (<u>https://lagoslakes.org/lagos-ne/</u>).
  - Used to demonstrate Nutrient Explorer
  - Includes nutrient concentrations for nitrogen and phosphorus
  - For >51,000 lakes >4-hectare surface area
- <u>Landscape/Watershed Data</u>: Lake observations were matched with landscape and other watershed level data.
  - Includes nutrient input data, climatic, land use, etc.



# **EPA**

## **Publication & Tool Links**

- Published Article:
  - Describes application of Nutrient Explorer on LAGOS-NE Dataset for predicting total TP in northeastern U.S. lakes
  - Citation provided at end of presentation
- Where to Download the App:
  - EPA's Science Inventory: https://cfpub.epa.gov/si/si public record Re port.cfm?Lab=CPHEA&dirEntryId=358039
  - EPA's GitHub: https://github.com/USEPA/NutrientExplorer



Analytical Framework to Visualize and Investigate Drivers of Surface Water Quality (Version



## **Sections of the R Shiny App**

- 1. Load Data
- 2. Explore Data
- 3. Create a Subset
- 4. Explore Subset
- 5. Run Models (Random Forest or Multiple Linear Regression)
- 6. User Guide



# Start of Nutrient Explorer Demo



- Show Science Inventory & GitHub site for downloading
- Show how the User Interface opens from R Shiny application
- Walk through App



# End of Nutrient Explorer Demo



#### **Conclusions**

- This application demonstrates how surface water nutrient concentrations and landscape variables can be used to predict water quality in data-poor locations.
- The R code and Shiny user interface developed have a variety of useful features including:
  - 1. Visualization of temporal and spatial patterns of an endpoint of interest
  - 2. Analysis of datasets to identify correlations between independent and dependent variables
  - 3. Modeling and prediction of concentrations that may be exceeding levels of concern.







Sepa Impacts

- EPA's Nutrient Explorer is a framework for predicting nutrient concentrations in surface waters and can be used to help managers prioritize watersheds and waterbodies for mitigative activities.
  - Specifically, this application can help managers predict which regions are at higher risk for having nutrient concentrations above certain threshold or water quality criteria levels.
- The models can be used to assess the dominant factors (e.g., fertilizer amount) and their relative contribution to nutrient concentrations in surface waters.
- It is important to note that model predictions are not equivalent to observations and therefore cannot be used as sole determinants of waterbody impairment.



https://www.maine.gov/dep/water/lakes/algalbloom.html



15



#### **Future Iterations**

- The version associated with the journal article will be maintained and will get minor updates periodically.
  - More flexibility in the required variable naming convention for user provided datasets.
- Aspects of Nutrient Explorer will be part of a new application being developed called "Nutrient Inventory Explorer," headed by Robert Sabo.



National Nutrient Inventory

The nutrient inventory tool is a web tool to visualize nitrogen and phosphorus inventories as well as modeled changes annual export for drinking water acros the continuous United States.





# **SEPA** Links & Citation

- EPA's Science Inventory: <u>https://cfpub.epa.gov/si/si\_public\_record\_Report.cfm?Lab=CPHEA&dirEnt</u> <u>ryId=358039</u>
- EPA's Github: <a href="https://github.com/USEPA/NutrientExplorer">https://github.com/USEPA/NutrientExplorer</a>
- Citation:
  - Pennino, Michael J., Meridith M. Fry, Robert D. Sabo, and James N. Carleton. 2023. Nutrient Explorer: An analytical framework to visualize and investigate drivers of surface water quality. Environmental Modelling and Software. 170. https://doi.org/10.1016/j.envsoft.2023.105853.







#### **Michael Pennino**

Ecologist Center for Public Health and Environmental Assessment U.S. EPA Office of Research and Development <u>pennino.michael@epa.gov</u> 202-564-8489

#### Acknowledgements

- Environmental Modeling and Visualization Laboratory (EMVL) members Yadong Xu and Ray Burton and EPA OMS's Heidi Paulsen for developing the R Shiny UI framework for this paper.
- Ben Washington (formerly at ORD) who helped develop the dataset used in this analysis and some of the initial R code for the visualizations and modeling.
- EPA's Sylvia Lee (ORD) and Micah Bennett (R5) who provided helpful comments and review of the application and associated manuscript.

Disclaimer: The views expressed in these presentations are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Any mention of trade names or commercial products does not constitute EPA endorsement or recommendation for use.



# Screenshots of Nutrient Explorer (if issues during live demo)

#### How to Download and Open Application

Opening the Application:

8-P4

- 1. Download and unzip the Zip File containing the app and associated files (found here: <u>https://github.com/USEPA/NutrientExplorer</u>).
- 2. Open RStudio > Click on "Project" (blue box, upper right) > Select "Open Project."
- 3. Navigate to the file folder where you saved the Application.
- 4. Click on the "Nutrient\_Explorer.Rproj" file and select Open.
- 5. On bottom right panel of RStudio, under the Files Tab, click on "ui.R" to open this R code file.
- 6. On bottom right panel of RStudio, under the Files Tab, click on "server.R" to open this R code file.
- 7. Check to make sure you have all the necessary packages installed before the next step.
- 8. Near the Upper Right, click on the "Run App" button with the green arrow pointing to the right. This will open the User Interface for the application in a separate window.



### **Random Forest Modeling**

#### **Building the Model**





https://towardsdatascience.com/understanding-random-forest-58381e0602d2

- Train-Test Split
- Train Model on 90% of data
- Test model on remaining 10%
- Cross Validation

#### Build (train) the model with observed data

Create multiple Trees Based on different subsets of the data and variables



#### Applying the model





### **Random Forest Modeling: Classification**

#### Response Variables (1s & Os)

Observed catchments with Public Water Systems and nitrate violations



#### **Predictor Variables**

- Land use
- Soil Type
- Nitrogen inputs
- Climate

#### **Decision Tree**



Site	Nitrate Violation	% Cropland	Precip.	Fertilizer
1	1	10	12	70
2	0	9	33	5
3	1	22	41	50
4	1	31	51	60
5	0	7	20	10
		Penni	no et al. 20	020 STOTEN

#### Load Default Dataset

**SEPA**

127.0.0.1:3741	Open in Browser				😏 Repu
Sta Civingon Miles Tal	AGENCE		Nutrient Explorer		
oad Data	Explore Data	Create a Subset Explore Subset Run Models	User Guide		
				Endpoint options	
				LogTN	
Dataset opti	ions			LogTP	
LAGOS	est			© TN	
- DAG00				© TP	
Inload nou	dataaat			Color palette	
upioad new	dataset			options	
Browse	No file sele			Blues	
				Greens	
				viridis	
Load this d	Interest			magma	
Load this d	Jataset			plasma	
Display					
Head					
🛛 Tail					
Display file	contonic				
Display file	contents				



#### **Or Load User Provided Dataset**

ve 💽 Off	<b>田 ら</b> 、 (~ )	® ~ ≏	LAGOS_Test3.csv ∽	𝒫 Search			Pennino, Michael 🧃	• <b>E</b>	o x			
Home	Insert Draw	Page Layou	ut Formulas Data	Review View	Help ArcGIS				🖻 Share			
-	: × ×	<i>f</i> x iws_ha							v			
e LAG AMES AMES IL ISON ISON ISON ISON ISON FH CE E	B C OS_La HU8Zonel 156 HU8_365 156 HU8_365 687 HU8_339 902 HU8_366 902 HU8_366 902 HU8_366 1061 HU8_350 1061 HU8_350 1822 HU8_342	D 7020010 MN 7020010 MN 7020010 MN 7020011 MN 7020011 MN 7020011 MN 7020011 MN 7020005 MN 7030005 MN 7030005 MN	E         F         G           tramn programty Lat         MPC/State Ager         43.97355           MPC/State Ager         43.97355         MPC/State Ager         43.97355           MPC/State Ager         45.39895         MPC/State Ager         44.18874           MPC/State Ager         44.18874         MPC/State Ager         44.18874           MPC/State Ager         44.18874         MPC/State Ager         45.37856           MPC/State Ager         45.37856         MPC/State Ager         44.9813           MPC/State Ager         44.9813         44.9813         44.9813	H         I           Long         HU2           -94.6478         -           -94.3064         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.801         -           -93.814         -           -93.8214         -	J         K           Year         Month           7         2010         1           7         2011         -           7         2010         -           7         2010         -           7         2010         -           7         2010         -           7         2011         -           7         2010         -           7         2010         -           7         2010         -	L M N Day Date Elevation M 0 6 10/6/2010 99999 4 21 4/21/2011 99999 4 18 4/18/2011 99999	O P axDepth MeanDep la 99999 NA 99999 NA 99999 NA	Q R 99999 99999 99999 99999 99999 99999 99999 99999	S ws_ha 1599.226 1599.226 3495.133	Nutrier	nt Ex	plorer
E TAGE RIE RIE RIE	1822       HU8_342         54660       HU8_335         65948       HU8_342         65948       HU8_342         65948       HU8_342         65948       HU8_342	7010205 MN 7010106 MN 7010205 MN 7010205 MN 7010205 MN	MPC/ State Ager 44.9813 MPC/ State Ager 46.96563 MPC/ State Ager 45.08243 MPC/ State Ager 45.08243 MPC/ State Ager 45.08243	<ul> <li>-94.4266</li> <li>-95.1219</li> <li>-94.7864</li> <li>-94.7864</li> <li>-94.7864</li> <li>-94.7864</li> </ul>	7     2011       7     2011       7     2010       7     2010       7     2011	MARWIAL PROTECTION	8	Open				×
RIE AMES	65948 HU8_342 156 HU8_365	7010205 MN_ 7020010 MN_	MPC/State Ager 45.08242 MPC/State Ager 43.97355	L -94.7864 5 -94.6478	7 2011 7 2010	Load Data Explore Da	ta Cre	$\rightarrow$ $\checkmark$ $\uparrow$	<u> </u>	MVL > Nutrient_Explo > ~	C Searc	h Nutrient 🔎
TE IRC AR	260 HU8_130 557 HU8_338	9030001 MN_ 7010201 MN_	MPC/State Ager 47.8703 MPC/State Ager 45.8141	5 -91.7922 3 -94.6362	9 2011 7 2011		Or	rganize 👻 Nev	v folder			≣ - □ 3
l Chok Chok	687 HU8_339 758 HU8_362 758 HU8_362	7010202 MN 7020002 MN 7020002 MN	MPC/ State Ager 45.39893 MPC/ State Ager 45.35549 MPC/ State Ager 45.35549	3 -94.3064 9 -96.1373 9 -96.1373	7 2010 7 2011 7 2011	Dataset options		合 Home		Name	Status	; Date mo
						O LAGOS test		🔁 Gallery	1.1	2021	$\odot$	11/11/20
						O LAGOS	>	📥 Michael - Env	irc 📔 🖿	2022	0	6/22/202
										Science_Inventory_Version	C	7/28/202
						Upload new dataset		↓ Downloads	. 8	HUC8_2007_7_Average_MaxDepth2.csv	$\odot$	5/4/2023
						Browse LAGOS			<u> </u>	HUC8_2007_7_Average_MaxDepth3.csv	$\odot$	5/13/202
						Upload complete			<u> </u>	HUC8_newData.csv	$\odot$	5/3/2023
									Ĩ 🛛	LAGOS_Test.csv	$\odot$	5/3/2023
						If uploading new		Projects			0	0/27/202
						'Upload complete':		Scrinte	File name:	~	Custom Files (*	.csv;*.text;*.txt;* ∨
											Open	Cancel
						Load this dataset			-			

**SEPA**

AutoSave 💽 🗄 🏷 🗸

File **S1** 

Α

3 ST. JAMES

5 MADISON

6 MADISON

7 MADISON

8 SOUTH CE

9 SOUTH CE

12 PORTAGE

13 CARRIE 14 CARRIE

15 CARRIE

16 CARRIE

17 ST. JAMES

18 WHITE IRC

21 ARTICHOK

22 ARTICHOK

19 CEDAR

20 PEARL

10 BELLE

11 BELLE

1 Name 2 ST. JAMES

4 PEARL

#### Summary Table

**SEPA**

vnload this dataset		México	La Hab	ena I cuba	6	Leaflet   © OpenStree	etMap contributors, CC-BY-SA	
			Summary table	]				
	Not Not	te 1: There are 314 variables te 2: You can use the radio bu	in total for this d ttons on the right	ataset,they are to select a var	e divided int riable group	to 11 groups as li to see their summa	sted under 'Variabl ary statistics.	
		variable	+ minimum	mean	median	maximum	≑ sd ≑	Variable annua antiana
	1	LAGOS_Lake_ID	16.00	18,306.25	5,320.00	140,866.00	32,736.77	Default Group
	2	Month	1.00	7.00	7.00	12.00	1.74	<ul> <li>NLCD Land Use</li> </ul>
	3	Day	1.00	16.72	17.00	31.00	8.66	<ul> <li>Weather related</li> </ul>
	4	iws_ha	0.70	20,746.06	753.08	1,482,385.00	109,505.79	O Vegetation related
	5	iws_perimkm	1.96	110.26	24.30	3,826.51	336.97	Aerosol related
	6	iws_lakeareaha	4.01	1,199.27	80.59	113,496.40	7,266.71	N Inventory Raw
	7	lake_area_ha	4.01	1,199.27	80.59	113,496.40	7,266.71	N Inventory Scaled     P Inventory Raw
	8	lake_perim_meters	797.62	37,171.26	5,958.78	1,326,792.00	137,004.80	<ul> <li>P Inventory Scaled</li> </ul>
	9	MeanDepth	0.30	5.77	4.27	88.60	8.33	Deposition
	10	MaxDepth	0.30	13.83	10.10	198.40	17.21	Sampling Information
	11	iws_tri_mean	0.04	0.63	0.49	5.35	0.48	
	12	iws_slope_mean	0.29	4.50	3.56	33.19	3.33	
	13	iws_road_density_mperha	0.00	33.15	25.89	174.85	28.15	
	14	iws_canopy2001_mean	0.00	51.29	58.09	96.13	28.15	
	15	iws_damdensity_pointsperha	0.00	0.00	0.00	0.54	0.00	
	16	Lat	36.49	43.80	44.19	48.19	2.29	
	17	Long	-96.36	-85.80	-89.42	-67.27	8.30	

+

### Explore Data: Summary Info

€PA



#### **Explore Data: Time Series**

EPA



#### **Explore Data: Time Series**

*€* EPA



#### **Explore Data: HUC2 Summaries**

€PA



#### **Explore Data: HUC2 Summaries**



**€EPA** 



#### Explore Data: Predictor Variable HUC2 Summaries

http://127.0.0.1:3741	Den in Browser				😏 Republish 🕞
Load Data	Explore Data	Create a Subset Explore Subset Run Models Use	er Guide		•
Summary	Info Time Se	ies Endpoint HUC2 Summaries Predictor Variables HUC	2 Summaries Correla	ations Maps Misc	
Select one predictor variable P_Ag_Surp <del>i</del> u	40 IS. 30-	•			
Adjust Y axis scale 0.05 0.65 0.05 0.45 0.85					
	0 -				
	-	1 2 4 5 7 HUC2 Z	one ID	9 10	11

#### **Explore Data: Correlations - Bivariate**



**€PA**

#### **Explore Data: Correlations - Multivariate**



**€**EPA

#### **Explore Data: Maps SEPA** Sepublish \* http://127.0.0.1:3741 Copen in Browser VVIRG Nutrient Explorer User Guide Explore Subset Run Models Load Data Explore Data Create a Subset Endpoint HUC2 Summaries Predictor Variables HUC2 Summaries Maps Misc Summary Info **Time Series** Correlations Endpoint options LogTP O TN ◎ TP ◎ LogTN LogTP Summarized by HUC2 LogTP Summarized by HUC8



### Explore Data: Misc.

**€PA** 

C:/Users/mpen	nino/OneDrive - Er	vironmental Protection	Agency (EPA)/R/Shi	ny/EMVL/Nutrient_Expl	orer/2022/Final/Nutrient_Explorer_	final_paper_version - Shiny			- 0 >
http://127.0.0.1:3741	Open in Browser								😎 Republist
SNURONINEW AL P	STATES - LINDS				Nutrient E	xplorer			
Load Data	Explore Data	Create a Subset	Explore Subset	Run Models U	Jser Guide				
Summary	Info Time Se	ries Endpoint HU	C2 Summaries	Predictor Variables HL	JC2 Summaries Correlation	s Maps Misc			
rvalid LogTP data records				© LogTN	Endpoint options • LogTP O TN (	D TP			District Agency Citizen Monitoring Federal Agency LTER Non-Profit Agency State Agency Tribal Agency
Number of	i	ż	á	5	7 HUC2 Zone ID	8 9	10	11	University

#### Create a Subset of the Data

**\$EPA** 

C:/Users/mpennino/OneDrive - Environmental Protection Ag	ency (EPA)/R/Shiny/EMVL/Nutrient_Explorer/2022/Final/Nutrient_Explorer_final_paper_version - Shiny	- O × S Republish *
STATES STATES - JUNITED STATES - JUNITED STATES - JUNITED STATES	Nutrient Explorer	
Load Data Explore Data Create a Subset	Explore Subset Run Models User Guide	
Year Range:	Create my subset spatially	
2010 2010 2011 2011 2012 2012 2012 2013 2013		
Month Range:		
IWS ha Range(ha):		
0 148,23996,478 592,956 889,434 1,185,912 1,482,385		
IWS perimeter Range(km):		
1 384 767 1,150 1,533 1,917 2,300 2,663 3,066 3,449 3,827		
Elevation Range(m):		
0 91 182 273 364 455 546 637 728 619 906		
I aka Araa Danaa/ha):		•

#### Create a Subset of the Data



### Create a Subset Spatially

**SEPA**



#### **Create a Subset Spatially**

€PA

Lake Mean Depth Range(m)

(B) C:/Users/MPennino/OneDrive - Environmental Protection Agency (EPA)/R/Shiny/EMVL/P\_Explorer/P\_Explorer\_testing - Shiny đ X http://127.0.0.1:6924 🔊 Open in Browser 😏 Republish 💌  $\sim$ 2,010 2,011 2,011 2,012 2,012 2,013 Choose states Choose HUC2 Month Range: You selected these HUC2s: 5 11 12 3 01(New England Region) 07(Upper Mississippi Region) IWS ha Range(ha): 1.482.385 Golfe Clear all selected HUC2s 0 148,23296,478 592,956 889,434 1,185,912 1,482,385 IWS perimeter Range(km): 17 Ottaw 3.827 10 1 384 767 1.150 1.533 2.300 3.066 3.827 Elevation Range(m): 11 0 91 182 273 364 455 546 637 728 819 908 hoe**fli**5 Lake Area Range(ha): LOS ATTL 03 113.497 12 4 11,35322,703 45,401 68,100 90,798 113,497 Lake Perimeter Range(m) The Bahamas 797 1.326.792 HUC2(clickable) México La Habana 👳 Cuba Ciudad 265,998 531,196 798,395 1,081,595 1,328,792 797 lakes de México República

Dominicana-

Leaflet | @ OpenStreetMap contributors, CC-BY-SA



 This section produces the same figures as the Explore Data section, but specifically for the subsetted dataset

<b>\$EPA</b>	Rι

#### **Run Models**

http://127.0.0.1:3741 Open in Browser		Republish •
SNURROWINE AL PROTECTION	Nutrient Explorer	
Load Data Explore Data Create a Subset	Explore Subset Run Models User Guide	
Model Types Random forest (on data subset) Multilinear regression (on data subset) Choose this model		
۰		



#### **Random Forest Model**

**SEPA**

//127.0.0.1:3741 🖉 Open in Browser	G				🧐 Repub
Load Data Explore Data	Create a Subset Ex	xplore Subset	Run Models Use	er Guide	
Model Types	Select all default va	Clear all	Link variables from	multilinear regression mo	del
<ul> <li>Random forest (on data subset)</li> </ul>	Select all in Group	Select all in Group	Select all in Group	Select all in Group 4	
<ul> <li>Multilinear regression (on data subset)</li> </ul>	Group 1:NLCD Land	Group 2:P Inventory	Group 3:N Inventory	Group 4:Aerosol related	
please select end points		Scaled			
LogTP	Select all in Group	Select all in Group	Select all in Group	Select all in Group 8	
ntree	Group	Group	Group	Group 8:Lake	
500	5:Weather related	6:Vegetation related	7:Deposition	characteristics	
mtry					

#### **Random Forest Model**

http://127.0.0.1:3741 Dopen in Browser

**SEPA**

Model Types	Select all default variables	Clear all	LINK variables from multilin	near regression model
<ul> <li>Random forest (on data subset)</li> <li>Multilinear regression (on data subset)</li> </ul>	Select all in Group 1	Select all in Group 2	Select all in Group 3	Select all in Group 4
Choose this model	Group 1:NLCD Land Use	Group 2:P Inventory Scaled	Group 3:N Inventory Scaled	Group 4:Aerosol related
please select end points	nlcd_pct_0 ×	P_Crop_removal.1 ×		
LogTP 🝷	nlcd_pct_11 × nlcd pct 31 ×	P_f_fertilizer.1 × Legacy P.1 ×		
ntree	nlcd_pct_52 ×	P_livestock_Waste.1		
500	nlcd_pct_71 ×	P_nf_fertilizer.1 ×		
500	NLCD_pct_developed	P_Deposition.1 ×		
mtry	NLCD_pct_forest ×	P_human_waste_kg.1		
9	NLCD_pct_farm ×	P_kg_ww.1 ×		
3	NLCD_pct_wetlands =	Recovered_P.1 ×		
estingset percentage		P_Ag_Surplus.1 ×		
0.1		NAPI.1 ×		
0.1		P_Ag_Inputs.1 ×		
ninimum node		P_Anthro_Inputs.1 ×		
5				
set seed	Select all in Group 5	Select all in Group 6	Select all in Group 7	Select all in Group 8
10	Group 5:Weather	Group 6:Vegetation	Group 7:Deposition	Group 8:Lake characteristics
	related	related	Atmo_Pdep ×	HU2 × Year × Month ×
Run random forest	SNOW ×	Vegetation ×	Atmo_Pdep_Year.Lag1	iws_ha × Elevation ×
The rendom forest	SNOW_Lag1 ×	Vegetation_Lag1 ×	Tot_Ndep ×	lake_area_ha ×
	SNOW_Lag2 ×	Vegetation_Lag2 ×	Tot_Ndep_Year.Lag1	lake_perim_meters ×
	SNOW_Lag3 ×	Vegetation_Lag3 ×	Tot_Sdep ×	MaxDepth × iws_slope_mean ×
	SNOW YrMean ×	Vegetation YrMean ×	Tet Edan Vand ant	hus road density moreha w

Sepublish \*

#### **RF Model: Results**

**SEPA**

Dahua Drofila Toole Hala				
C:/Users/mpennino/OneDrive - E	nvironmental Protection Agency (EPA)/R/Shiny/EMVL/Nut	trient_Explorer/Science_Inve	ntory_Version/Nutri — 🗆 🗙	
p://127.0.0.1:6510 Dpen in Browser			😏 Publish 👻	
2	related related			
	LST_Lag2 :			
testingset percentage				
0.1			Model Uncertainty Metrics	Values
-1.1			r.squared (reported by Ranger function)	0.7569
minimum node				0.04.40
5	Model Uncertainty Metrics	Values	prediction.error (reported by Ranger function)	0.2149
set seed	r.squared (reported by Ranger function)	0.7569	R squared (testing set)	0.7840
Set Seeu	prediction.error (reported by Ranger function)	0.2149		
10	R squared (testing set)	0.7840	R squared (training set)	0.8859
	R squared (training set)	0.8859	Root Mean Squared Error	0.4366
Run random forest	Root Mean Squared Error	0.4366		0.0400
	Mean Bias	0.0166	Mean Blas	0.0166
	Standard deviation of the error	0.4365	Standard deviation of the error	0.4365
	Nash-Sutcliffe efficiency coefficient	0.7837	Nash-Sutcliffe efficiency coefficient	0.7837

#### Random Forest Observed vs Predicted



## **\$**EPA

#### **RF Model: Variable Importance Plot**



### **RF Model: Partial Dependence Plots**

mental Protection Agency (EPA)/R/Shiny/EMVL/Nutrient\_Explorer/Science\_Inventory\_Version/Nutrient\_Explorer\_v.1.1 - Shiny

**€EPA** 



😏 Republish 💌

#### **RF Model Predictions**

**SEPA**



#### **RF Prediction Map for specified Threshold**



**SEPA**





#### **MLR Steps**

http://127.0.0.1:3741	Open in Browser

c//127.0.0.1:3741 🖉 Open in Browser 🕓					😏 Republish 💌
Random forest (on data subset)					A
<ul> <li>Multilinear regression (on data subset)</li> </ul>	Select all default variables	Clear all	Link variables from random forest	N important variables 10	
Choose this model					
	Select all in Group 1	Select all in Group 2	Select all in Group 3	Select all in Group 4	
please select end points	Come ANII CD Long I	Correct 20 Different correct	Correct 2 Million of a million		
LogTP -	Group 1:NLCD Land Use	Group 2:P Inventory Scaled	Group 3:N Inventory Scaled	Group 4:Aerosol related	
	NLCD_pct_forest ×	P_Crop_removal.1 ×			
correlation oritoria	NLCD_pct_farm ×	P_f_fertilizer.1 ×			
correlation criteria		P_Anthro_Inputs.1 ×			
0.9					
Step 1: Run Correlation Analysis	Select all in Group 5	Select all in Group 6	Select all in Group 7	Select all in Group 8	
nvmax	Group 5:Weather	Group 6:Vegetation	Group 7:Deposition	Group 8:1 ake characteristics	
20	related	related		Group O.Lake characteristics	
		NPP YrMean ×	lot_Ndep ×	MaxDepth ×	
nbest				iwo concerv2001 moon *	
1				Iws_carropy2001_mean ×	
Step 2: Run Subset Selection					
otop 2. Hun oubset ooledion					
number of best final models to display					
5					
Step 3: Finalize Regression Model					

#### **MLR: Correlation Analysis**

**SEPA**

0.9	Select all in Group 5	Select all in Group 6	Select all in Grou	up 7	Select all in Group 8			
Step 1: Run Correlation Analysis	Group 5:Weather related	Group 6:Vegetation	Group 7:Deposi	ition Grou	p 8:Lake characteristics	5		
nvmax		NPP YrMean x	Tot_Ndep ×		MaxDepth ×			
20				IWS_	road_density_mperha ×			
nbest								
1								
	Copy Print Downloa	d 🔻				Search:		
Step 2: Run Subset Selection		NLCD pct forest	NLCD pct farm	P Crop removal.1	P f fertilizer.1	P Anthro Inputs.1	NPP YrMean	
number of best final models to display	NLCD not forest					·	-	
5	NLCD_pct_farm	-0.68						
	P Crop removal.1	-0.33	0.39					
Step 3: Finalize Regression Model	P_f_fertilizer.1	-0.34	0.41	0.98				
Step 4: Make Model Predictions	P_Anthro_Inputs.1	-0.42	0.55	0.94	0.94			
	NPP_YrMean	0.73	-0.67	-0.31	-0.34	-0.42		
4	Tot_Ndep	-0.68	0.47	0.22	0.26	0.3	-0.77	
	MaxDepth	0.03	0.03	-0.03	-0.05	-0.02	0	
				0.00	-0.04	-0.05	-0.25	
	iws_road_density_mperha	-0.42	-0.1	-0.03				
	iws_road_density_mperha iws_canopy2001_mean	-0.42 0.92	-0.1 -0.8	-0.03	-0.39	-0.49	0.81	

#### **MLR: Subset Selection**



Please click on one row in the above table to select the final model subset, we recommend to select the row with the lowe st delta BIC('bic\_delta').

Then click on 'Step 3: Finalize Regression Model'to proceed to the next step.

#### MLR: Finalize Regression Model

http://127.0.0.1:4889 🔊 Open in Browser 🕝 😏 Republish 💌 **Observed vs Predicted** Standardized Coefficients Plot Coefficients 0.0 0.0 0.2 Standardized Predicted Log TP -0.1 -0.2 d Has Yeal and Atmo PdeP Plan Yearland Reconser P.1 To Step Yeat and P. Deposition.1 Unen-wester 491 P.M.Billiter.1 Flevaion HUZ \*e Pain naers astod Water WHS BORE IREA Sec. and for 2 6 Observed LogTP **Predictor Variables** Estimate Std..Error t.value Pr...t.. std.Beta VIF Atmo\_Pdep 0.032260 0.009526 3.386571 0.000710852778370727 0.052513 4.272007 Atmo Pdep Year.Lag1 0.047697 0.008955 5.326549 1.02597525095346e-7 0.094113 5.546679 Elevation -0.000550 0.000063 -8.794254 1.7157902653858e-18 -0.090479 1.880740 HU2 0.046731 0.005404 8.647773 6.18735566021186e-18 0.125179 3.722895 iws\_canopy2001\_mean -0.009341 0.000462 -20.210369 8.47625757585889e-89 -0.275243 3.295421 iwe roaddoneity doneity mnorha 0 003255 0.000/115 7 8/0106 5 0202101672152/0 15 0 078/03 1 720270

€,FPA

#### MLR: Finalize Regression Model



http://127.0.0.1:6510 🥥 Open in Browser 🕝							😏 Publish	
	Atmo_Pdep	0.129590	0.004296	30.163190	5.36391533447769e-194	0.247000	1.206854	
	nlcd_pct_52	-0.011738	0.001306	-8.989966	2.76929100920758e-19	-0.068041	1.030959	
	NPP	-0.018336	0.003940	-4.653443	0.00000329279617213373	-0.037126	1.145584	
	P_human_waste_kg							
	P_livestock_Waste.	P_livestock_Waste. Model Uncertainty Metrics						
	The cells marke ted in the mode		Mul	tiple R-	squared	0	.4550	
	predictor varia Then rerun the	predictor varia Then rerun the		Adju	sted R-	-squared	0	.4543
	p 4: Make Model		Resid	ual stan	idard error	0	.6405	
	Model Uncertair							
	Multiple R-se		0	.6401				
	Adjusted R-s Residual stand	stand Mean Bias						
	Root Mean Squ							
	Mean Bi	Mean Bi Standard deviation of the error						
	Standard deviation	Standard deviation					.0101	
	Nash-Sutcliffe efficie	Nasł	n-Sutcli	ffe effic	iency coefficier	nt 0	.4550	

#### **MLR: Make Model Predictions**



**SEPA**

**SEPA**

#### **Download Dataset**

AutoSave 🤅		<b>७</b> • ९ ·	£ ~ ÷	Log	gTP_linear_r	egression_m	odel_predict	tion_results_2	2023-08 ·	~	⊖ Sear	ch			F	<sup>9</sup> ennino, M	ichael 🌗	Ē	
File Ho	me Ins	ert Draw	Page La	ayout	Formulas	Data	Review	View	Help									🖓 Comm	ients 🖻
Paste 🖋	Calibri B I	 U ~   ⊞ ~	11 - A^	A* 3	≡≡ ≡≡≡	&?~ <u>€</u> ≣ <u>⇒</u> ≣	같 Wrap Te 탄 Merge &	xt & Center ~	General \$ ~ 9	% 🤊 🔚	~ 0 .00 0 →0	Conditional F Formatting ~	Format as Table ~ St	Cell In yles ~	sert Delete	Format	$\begin{array}{ccc} \Sigma & \bullet & A_{\nabla} \\ \hline & & Z \\ \hline & & \bullet & \\ \hline & & & \\ \hline \end{array}$	Find & r Y Select Y	Sensitivity
Clipboard	2	Font		۲ <u>م</u>		Alignm	ent	5	ā N	lumber	E1	St	tyles		Cells		Edit	ing	Sensitivity
A1	• :	× 🗸 fs	LogTP	<b>)</b>															
A	В	с	D	Е	F	G	н	1	J	к	L	м	N	0	Р	Q	R	S	т
1 LogTP	ТР	huc8																	
2 4.075155	58.85958	1030001																	
3 4.116161	61.32339	2050301																	
4 3.884194	48.62772	2050305																	
5 3.907735	49.78607	2050306																	
6 3.888807	48.85259	2070009																	
7 3.864697	47.68884	2070002																	
8 3.585509	36.07173	5020006																	
9 3.637168	37.98413	4130003																	
10 3.80531	44.93918	4140201																	
11 3.758407	42.88007	2050106																	
12 3.131621	22.9111	1020004																	
13 3.681808	39.71812	2050201																	
14 3.651802	38.54406	2050202																	
15 3.653076	38.5932	2050203																	
16 3.685423	39.86197	2050204																	
17 4.067766	58.42631	2050205																	
18 2.880276	17.81919	2050206																	
19 3.702811	40.56116	2050302																	
20 3.82896	46.01468	2050303																	
21 3.90031	49.41778	2050304																	
22 3.860111	47.4706	5010005																	
23 3.55736	35.0705	1020005																	
	LogTP	_linear_regres	ssion_mode	el_p	+							: •							

#### SEPA MLR: Prediction Map for specified Threshold

