

# Nutrient Criteria Database Quality Assurance and Statistical Analysis

November 16<sup>th</sup>, 2010

*Missouri Nutrient Criteria Technical Team*

# Compiling Datasets

## 1- Original Dataset from MDNR

- ~2,000 sites
- ~38,000 samples
- 44 agencies or labs

## 2- Dataset from USGS database (not all in original dataset)

- ~460 sites
- ~12,568 samples

## 3- Dataset from MDC (RAM)

- ~390 samples
- included 187 samples from EPA  
Wadeable Streams Assessment (WSA)

# Database Actions

- Merged MDNR file with location file
- Reconciled ambiguous/duplicate sites (MDNR)
- Removed all USGS data then merged MDNR, USGS (reloaded), and RAM datasets to a Master database
- Database Reviews (iterative manner)
  - Site level - (Reconciled ambiguous/duplicate sites), site name and location errors ("Maries R..." but plots on Moreau), remove site near point sources, "special sites"
  - Sample level - review of nutrient and field data (discharge, conductivity, temperature, etc.)
    - Other data (metals, major ions) retained but NOT QA'd

# Site Level

- Sites qualified:

- G: good
- R: rejected
- P: provisional
- L: large river
- LK: lake
- S: spring
- Retained perennial stream sites
- Rejected if < 1-mi below point source
- Rejected some sequential downstream sites  
(1.0 mi below, 2.0 mi below, 3.0 mi below)
- Rejected if location unconfirmed
- Rejected pre-1990 data (analytical "issues", too old to be relevant)
  - Also reduced noise from artificial "time trends" & sampling density and differing sites

# Sample level Review

- Selected nutrients of interest for QA
  - Kjeldahl Nitrogen
  - Ammonia
  - Nitrate
  - Total Nitrogen
  - Ortho-phosphate
  - Total Phosphorus
  - Field measurements (flow, T, SC, D.O., etc.)
- Focus on TN and TP

# The Big Issue -Censored Data

	Total Nitrogen (TN)		Total phosphorus (TP)		Orthophosphorus (PO4)		Total Kjeldahl Nitrogen (KJN)		Ammonia (NH3N)		Nitrite plus Nitrate (NO3N)	
	Censoring level	Number and (fraction)	Censoring level	Number and (fraction)	Censoring level	Number and (fraction)	Censoring level	Number and (fraction)	Censoring level	Number and (fraction)	Censoring level	Number and (fraction)
Total number of values		18,209		22,953		13,523		15,878		19,914		22,241
Total number of censored levels or values	76	210	18	3,889	22	4,109	14	2,840	20	10,459	30	3,034
Fraction of values censored	--	(0.012) =C5	--	(0.169)	--	(0.304)	--	(0.179)	--	(0.525)	--	(0.136)
Number of values rejected	--	182	--	1198	--	362	--	280	--	109	--	839
Fraction of values rejected	--	(0.010)	--	(0.052)	--	(0.027)	--	(0.018)	--	(0.005)	--	(0.038)
Censor level	0.01	8	0.004	55	0.001	28	0.005	5	0.003	1	0.004	1
	0.05	1	0.008	78	0.004	47	0.01	114	0.01	607	0.01	358
	0.1	1	0.01	466	0.005	35	0.02	5	0.015	13	0.02	255
	0.11	1	0.02	514	0.006	122	0.05	113	0.018	38	0.03	1
	0.12	2	0.03	6	0.008	54	0.08	34	0.02	2031	0.04	80
	0.15	1	0.04	934	0.01	1441	0.083	2	0.03	1518	0.05	963
	0.2	2	0.05	638	0.018	3	0.1	314	0.04	2715	0.06	430
	0.25	11	0.06	331	0.02	1755	0.11	11	0.041	1	0.067	1
	0.3	1	0.1	81	0.024	2	0.12	1	0.05	1665	0.1	106
	0.34	1	0.2	6	0.03	10	0.14	52	0.07	1	0.14	1
	0.4	1	0.4	5	0.031	7	0.2	1909	0.08	2	0.15	7
	0.52	1	0.5	764	0.04	37	0.5	29	0.09	1	0.2	42

# Maximum Censor Thresholds

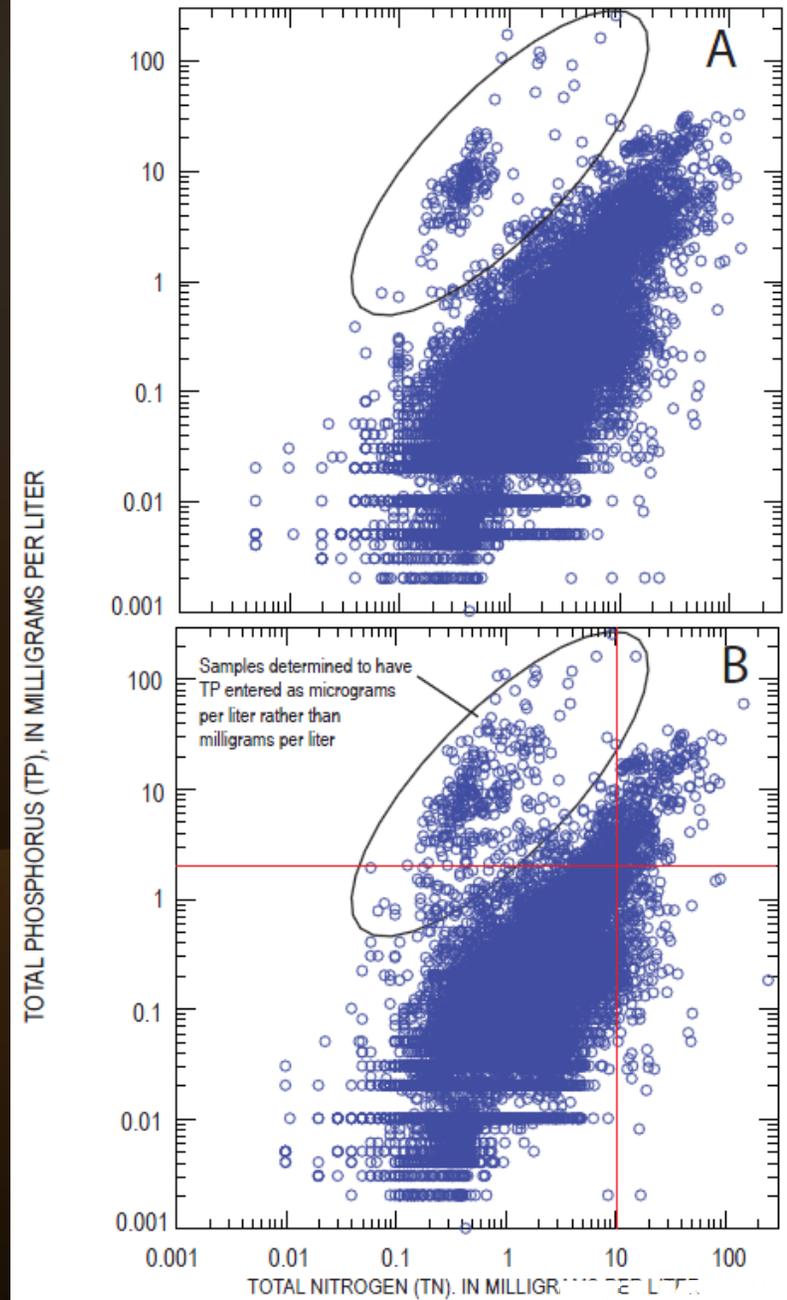
Constituent	Percent censored	Max. censor threshold (mg/L)	Censored values rejected
TN	1.2%	0.3 mg/L	182
TP	16.9%	0.05 mg/L	1,198
NO <sub>2</sub> +NO <sub>3</sub> (as N)	13.6%	0.1 mg/L	839
Ammonia	52.5%	0.1 mg/L	109
KJN	17.9%	0.2 mg/L	280
PO <sub>4</sub> (as P)	30.4%	0.05 mg/L	362

## Other "actions"

- Calculated TN where missing using unfiltered KJN and NO3N
- Removed some "blocks" of high density data (e.g. E. Fork Black River below Taum Sauk)
- Assigned drainage area and stream order
- Correct unit issues with some data
- Rejected high TN and TP data as "unreasonable" and not indicative of Mo streams ( $>10$  mg/L TN,  $> 2$  mg/L TP)
- Reference streams NOT necessarily "nutrient reference" picked for benthos habitat

Removed large TN and  
TP values  
( $>10$  TN,  $>2$  TP)

Final database:  
22,632 samples  
1,766 sites  
18,027 TN  
21,755 TP



\*Note -- red lines indicate final upper limit for TN (10 mg/L) and TP (2 mg/L) data retained in project database and calculated TN values not included in figure A

# Statistical Methods

- **Reference stream approach**: use 75<sup>th</sup> percentile of reference site samples and 25<sup>th</sup> percentile of all sample data
- **Geography**: calculate statistics by EDUs
- **Sample-based method**: calculate statistics using all values from a site
- **Site-based method**: Calculate a single value for each site (mean, geo-mean, median)

# Calculations: How to treat remaining censored data

- **Single substitution**: substitute all censored values with the maximum censored value, abandoned because of large amount of data removed
- • **Multiple substitution**: substitute each censored value with  $\frac{1}{2}$  of that value
- • **Kaplan-Meier**: a stat package that deals with multiple detection levels (distribution independent, initially "right censored data")
- **ROS**: (regression on ordered statistics) abandoned because of assumed distribution and similarity to multiple substitution method

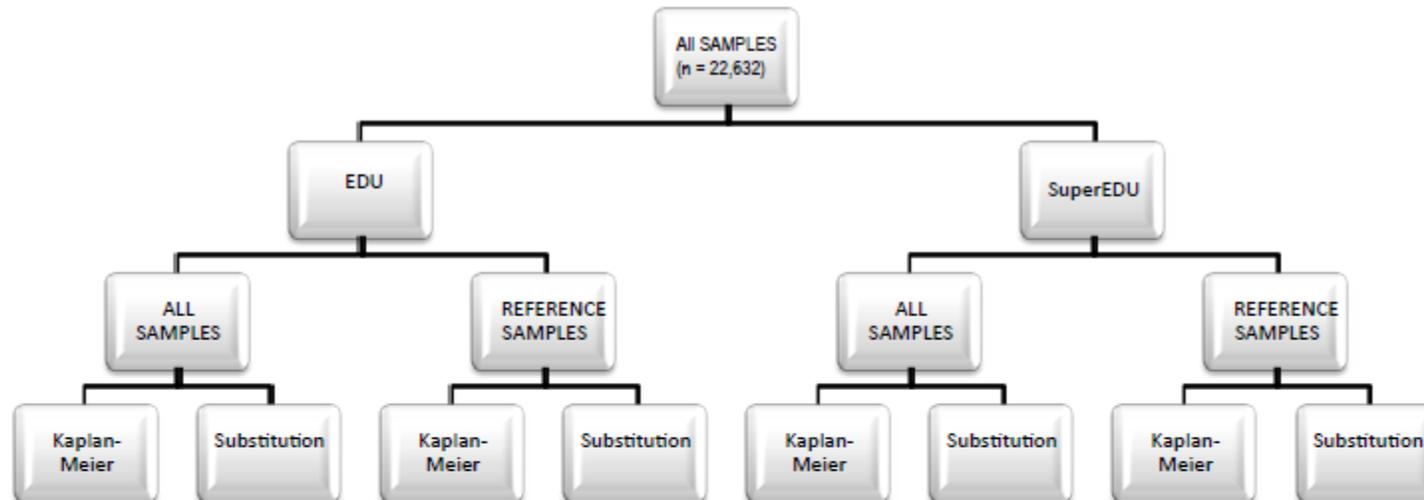
Used Kruskal-Wallis and Tukey's multiple comparison tests to investigate difference in TN and TP between EDU's within each EDU Region

- Central Plains
- Ozarks
- Mississippi River Alluvial Valley

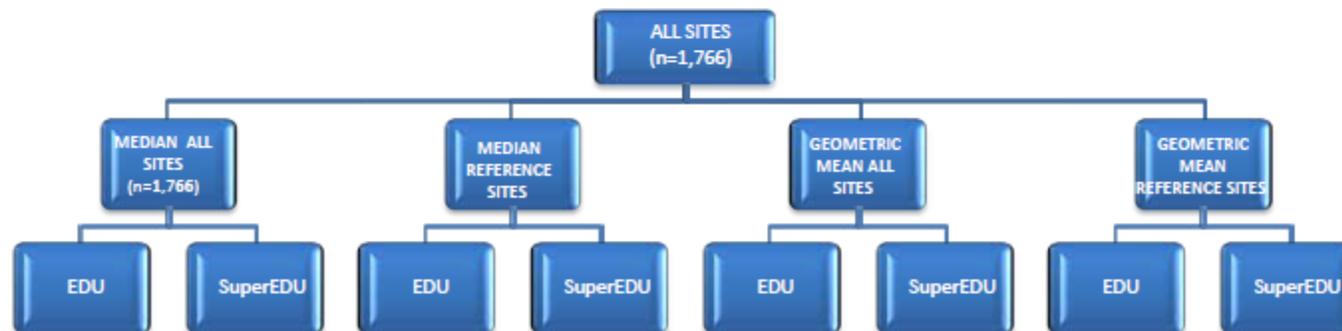
Combined adjacent EDUs where tests indicated no significant differences into "Super EDUs"

# Summary of Statistical Methods

## SAMPLE-BASED APPROACH

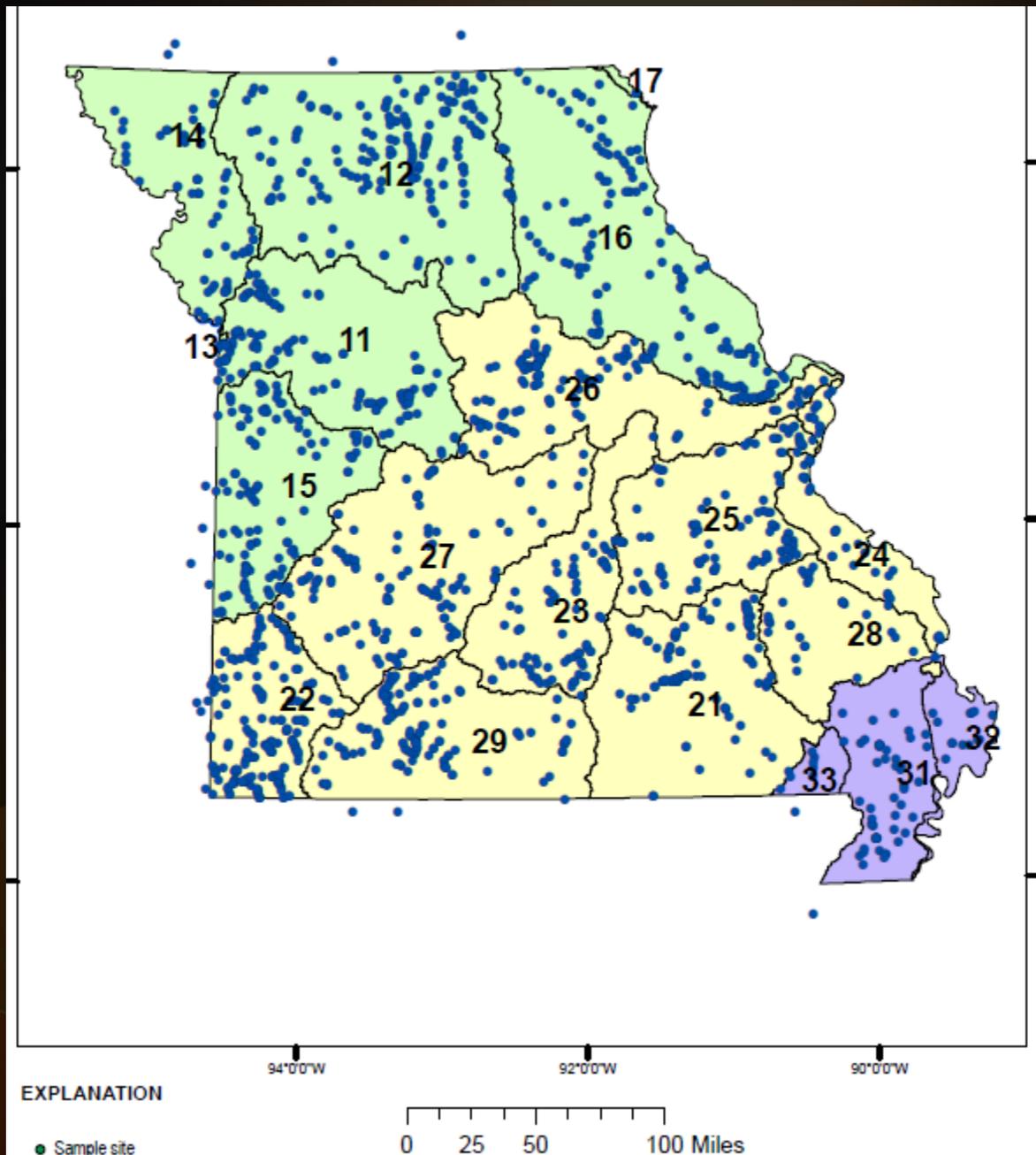


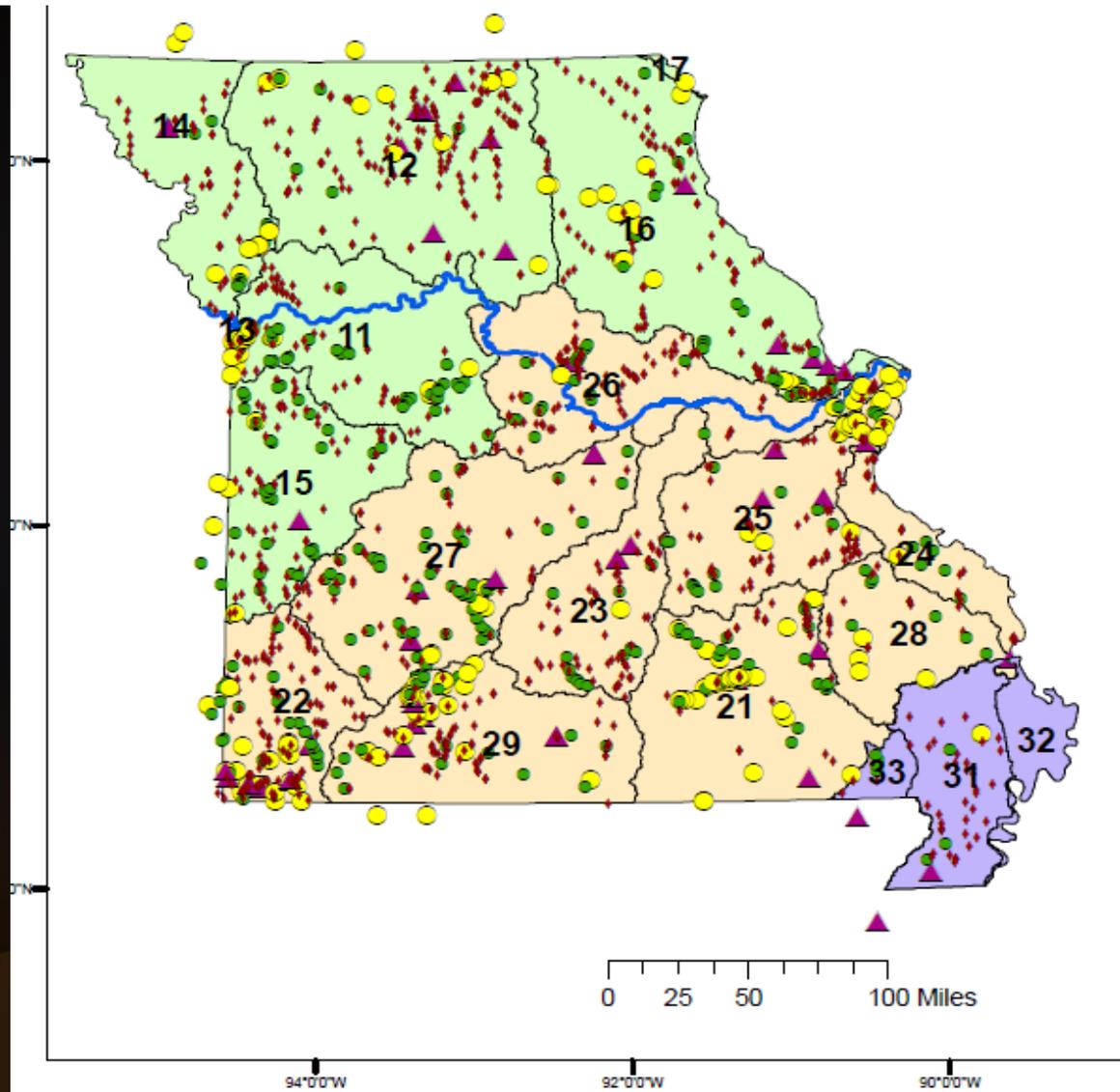
## SITE-BASED APPROACH



\* Note -- Constituent concentration for each station is the median value of all samples from that station.

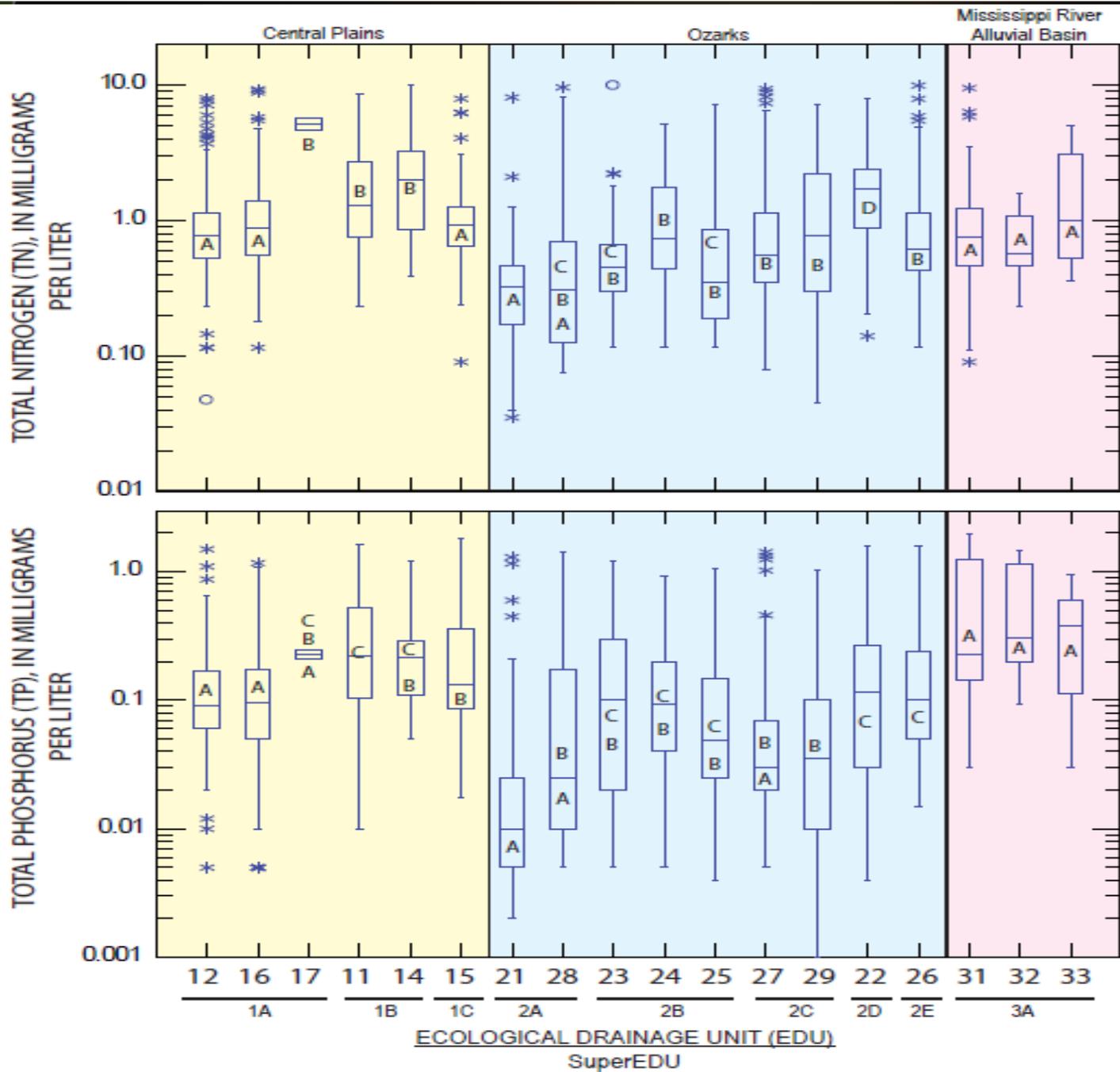
EDU indicates Ecological Drainage Unit



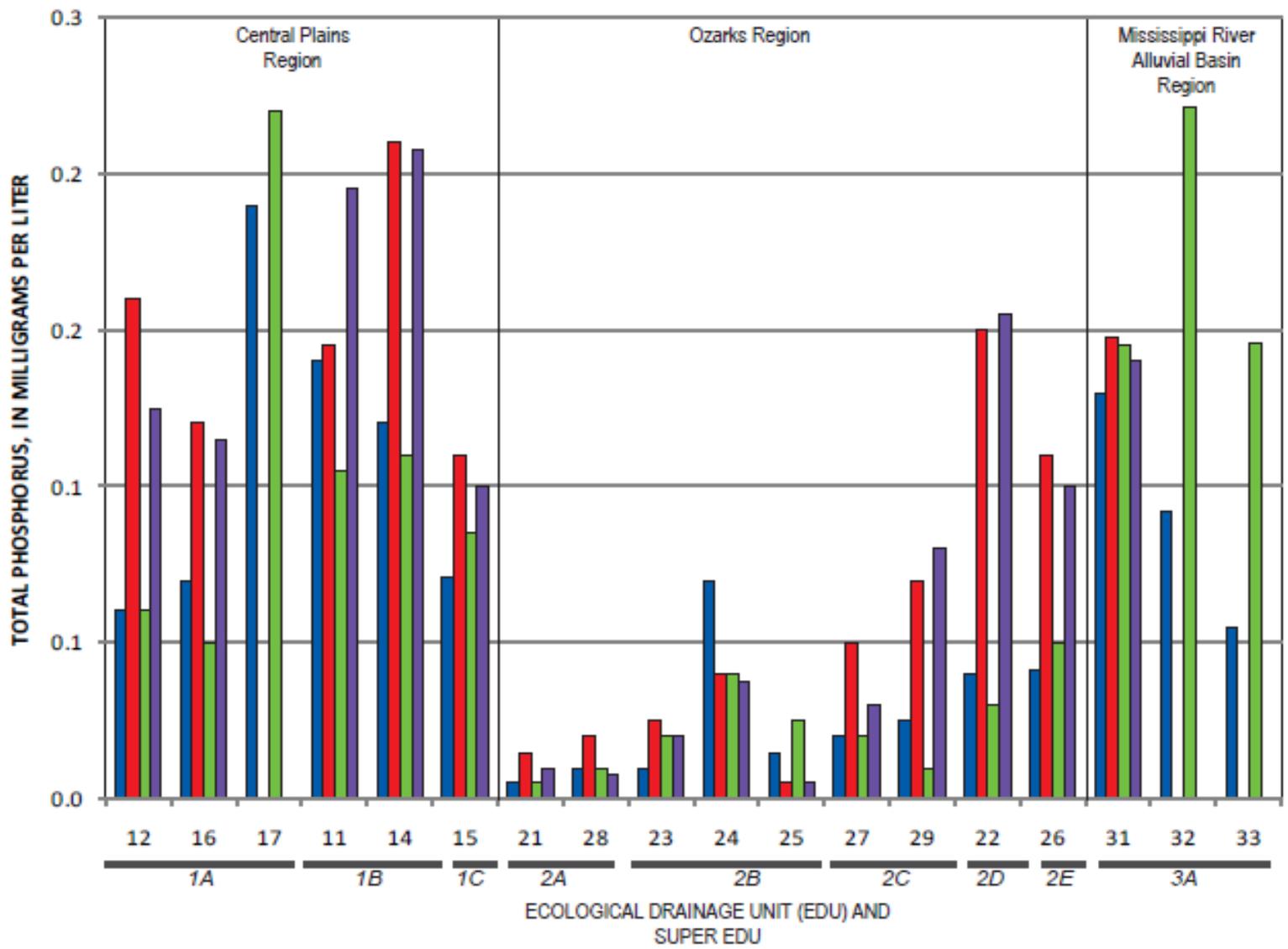


**EXPLANATION**

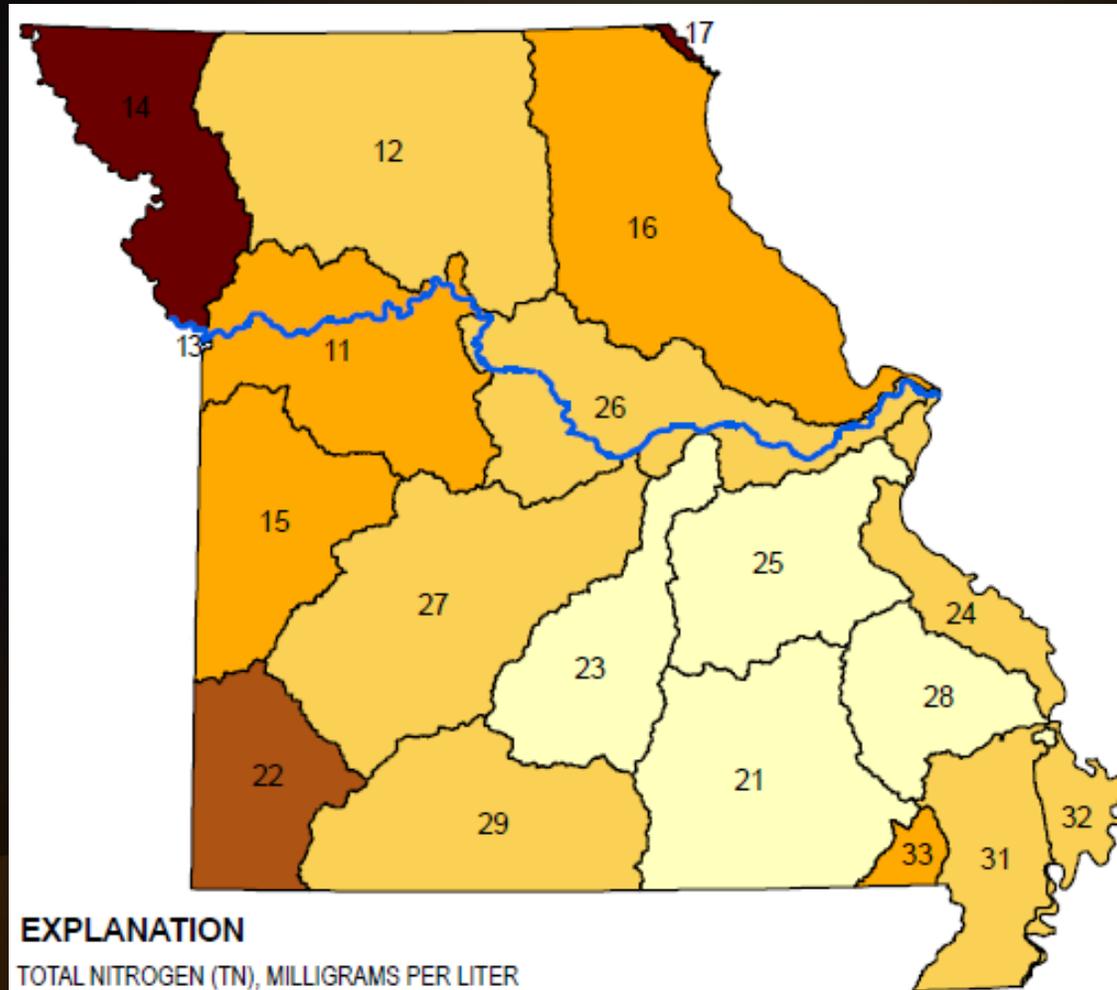
- |   |  |
|---|--|
|  Missouri River                        | Number of samples per site   |
|  Ecological Drainage Unit (EDU) Region |  0 to 5     |
|  Central Plains                        |  6 to 25    |
|  Mississippi Alluvial Basin            |  26 to 100  |
|  Ozarks                                |  101 to 364 |



\*note Boxes with the same letter are not significantly different at an alpha level of 0.05, and letter designation based on ranked data



■ 25th Percentile, sample-based approach, substitution method     
 ■ 25th Percentile, site-based approach, substitution method  
■ 75th Percentile of reference samples, sample-based approach, substitution method     
 ■ 75th Percentile of reference sites, site-based approach, substitution method



**EXPLANATION**

TOTAL NITROGEN (TN), MILLIGRAMS PER LITER

Light yellow	0.305000 - 0.500000
Yellow-orange	0.500001 - 0.800000
Orange	0.800001 - 1.300000
Brown	1.300001 - 1.750000
Dark red	1.750001 - 5.150000

