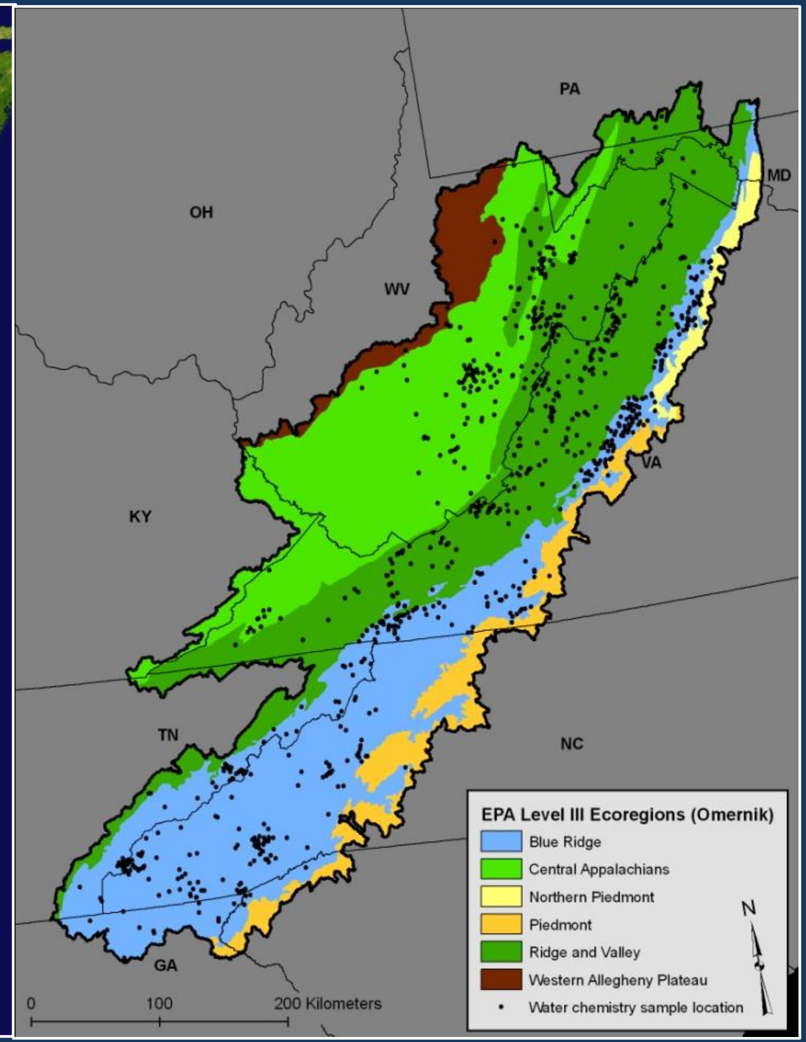
Implementing a DSS to assess CLs of atmospheric S deposition in the SE US

Paul Hessburg, Keith Reynolds,
Timothy Sullivan, Nick Povak, Brion Salter,
Todd McDonnell, Bill Jackson

Study area



- Study Area Size: 14.3 MM ha (35.4 MM ac)
- Domain: Ridge & Valley, Appalachian Plateau ecoregions in VA, WV; Blue Ridge ecoregion in VA, WV, NC, TN

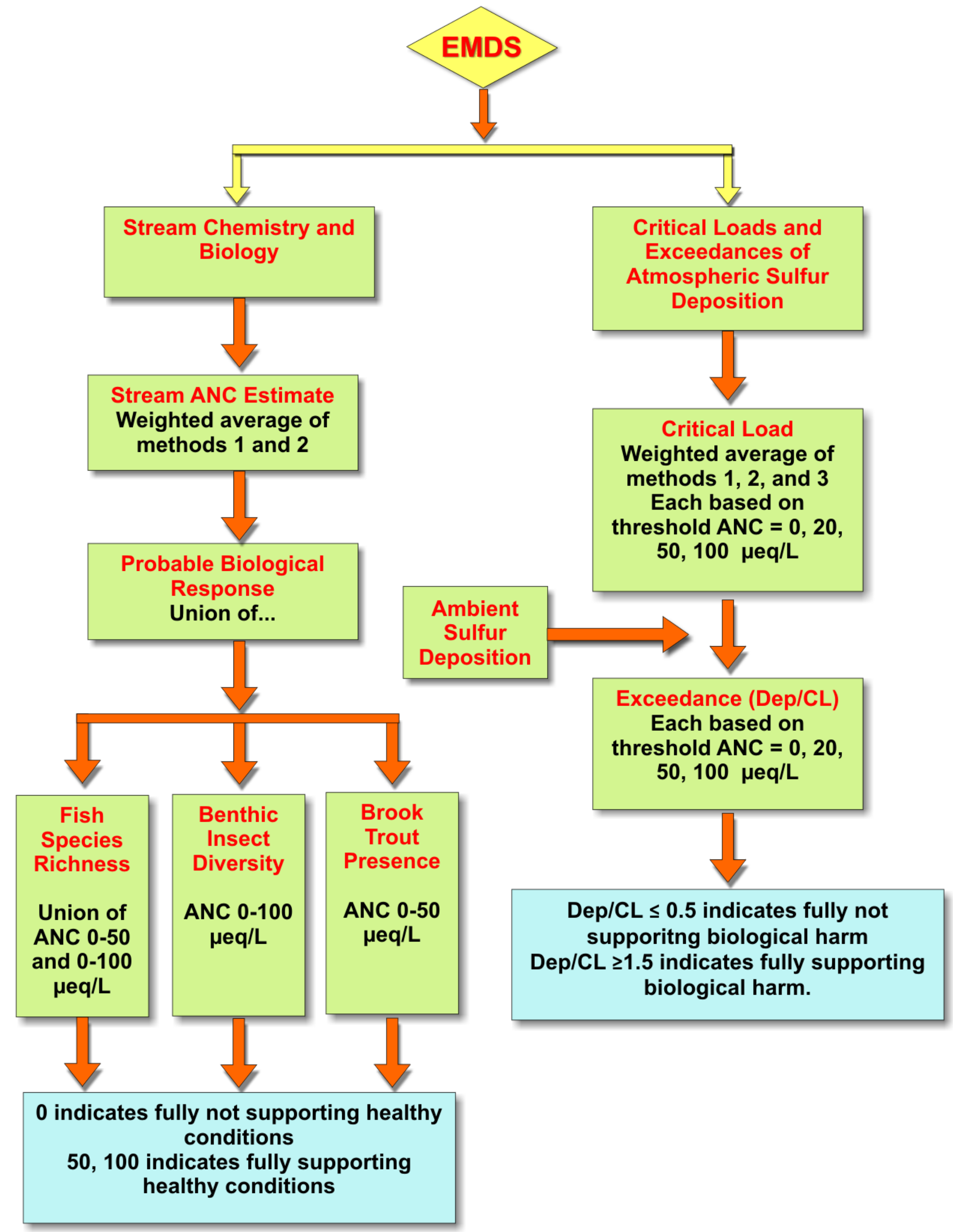
EMDS, logic, & decisions

- Spatial decision support for environmental analysis and planning in *ArcMap*
- Logic modeling
  - To assess environmental state(s)
- Decision modeling
  - To prioritize landscape elements based on environmental states & mgr. considerations

Logic for CL

CL methods:
1) MAGIC
2) Water chemistry
3) bgc ANC/BCw predictions

ANC methods:
1) Water chemistry
2) bgc ANC predictions



19 October 2010

Predicting ANC/BCw in SE US streams:

Status of current modeling efforts

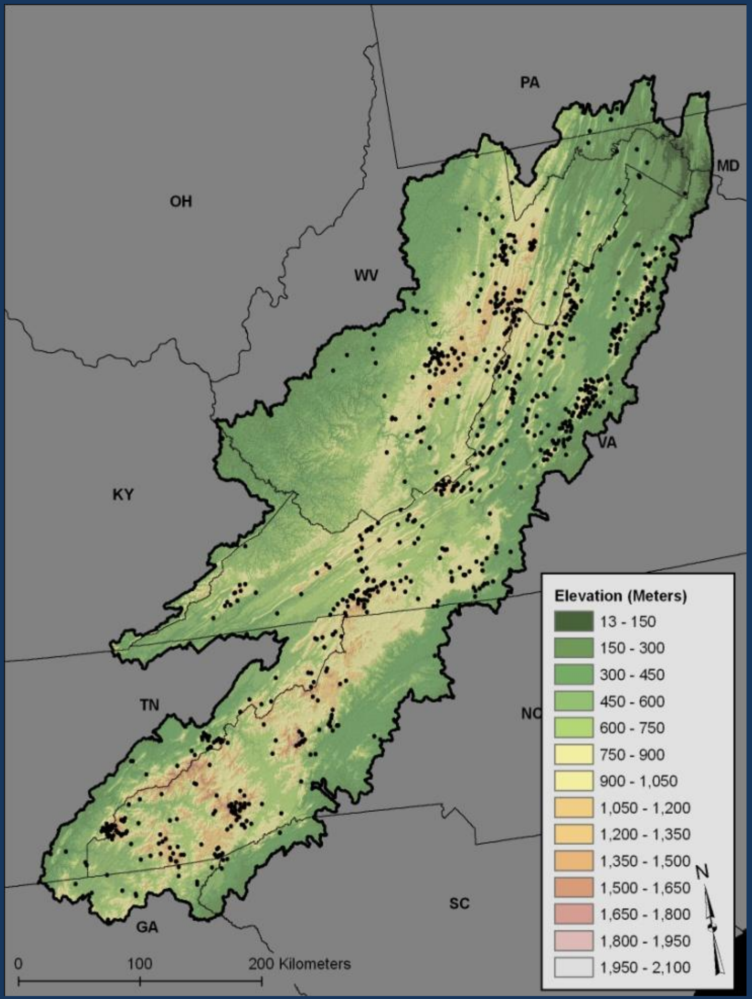
Paul Hessburg, Keith Reynolds, Nick Povak, Brion Salter
USDA-FS, PNW Res. Sta.
Tim Sullivan, Todd McDonnell, E&S Env. Chem., Corvallis, OR

Background

- Post-hoc modeling effort
- Predict continuous ANC/BCw from training data
- Water chemistry data, existing water quality databases
- Water chemistry sites represented by most recent spring sample
- 933 total sites with water chemistry, ANC values
- 140/933 sites have estimated BCw via MAGIC
- BCw + ANC used to estimate CL
- BCw results not shown, similar modeling approach

Study area


MN
ME
VT
NH
WI
NY
MA
CT
RI
MI
IA
PA
NJ
OH
MD
DE
IL
IN
WV
VA
KY
MO
NC
TN
AR
SC
MS
AL
GA
LA
FL
N
0
250
500 Kilometers


PA
MD
OH
WV
KY
VA
TN
NC
GA
SC
Elevation (Meters)
13 - 150
150 - 300
300 - 450
450 - 600
600 - 750
750 - 900
900 - 1,050
1,050 - 1,200
1,200 - 1,350
1,350 - 1,500
1,500 - 1,650
1,650 - 1,800
1,800 - 1,950
1,950 - 2,100
N
0
100
200 Kilometers

Methods

- Started with 56 potential predictors, incl. topo., soil, climatic, lithologic, vegetation, & SOx dep. vars (wet, dry, total)
- Eliminated sig. multi-collinearity → 33 vars
- Submitted the database of these 33 vars to a "Gatekeeper" modeling approach

Data Sources: Ameriflux, PRISM, STATSGO, EPA-CMAQ, NED, NLCD2000, E&S Env Chem

Gatekeeper model flow

Data

1. Threshold model[a] (RF & GWR)

Probability of low-ANC < cutoff[b] = Well-buffered[c] site [STOP]

Probability of low-ANC ≥ cutoff value; [ENTER CONT. MODEL]

2. Continuous model (RF & GWR)

Predicted continuous ANC value [STOP]

[a] **Threshold model**

1st uses a binary model to predict whether a cell is above/below a threshold ANC (e.g., </> 250)

2nd predicts a continuous ANC value for all observations predicted below that threshold.

[b] Cutoff value: probability required for entrance into a continuous model

[c] Well-buffered: ANC threshold value above which sites are considered well-buffered.

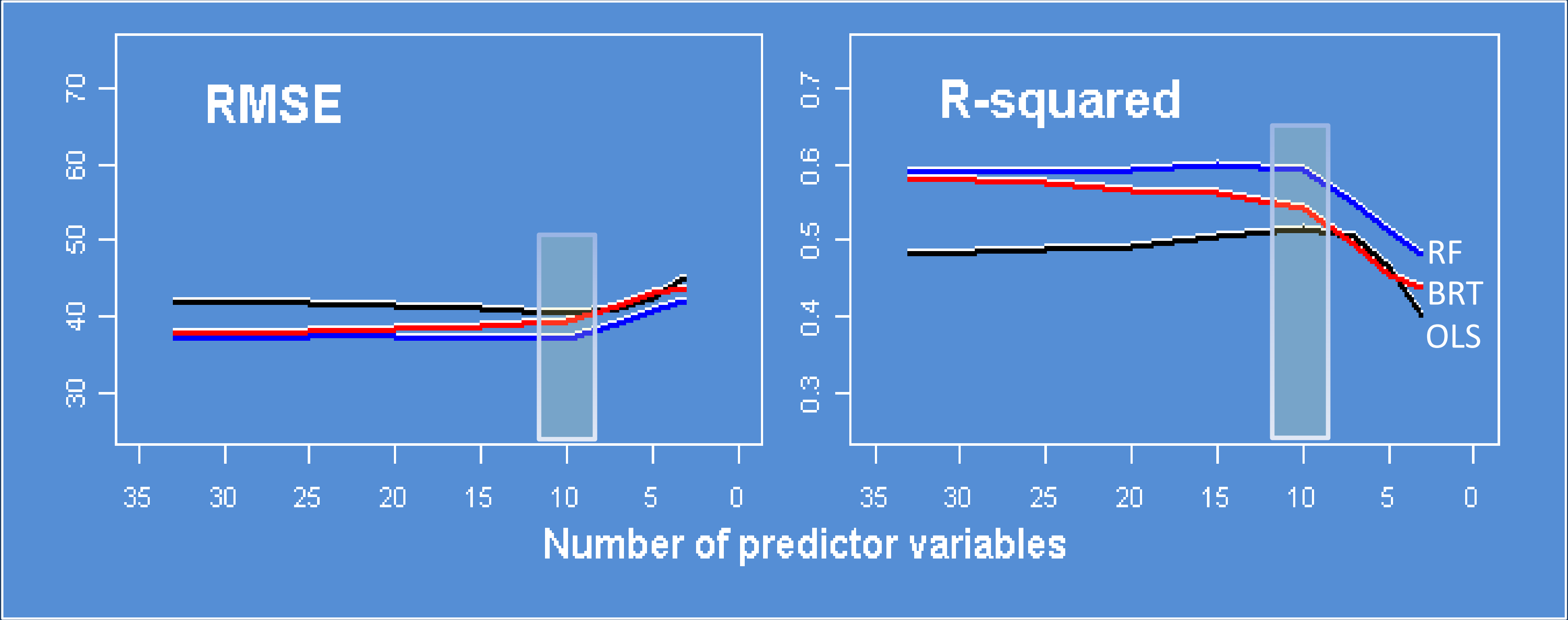
Results

Important predictors (% var. explained)

Threshold model--ANC </> 250 *μeq/L*		Continuous model--ANC < 250	
Carbonaceous lithology	(17.7)	Siliceous lithology	(24.3)
Percent public land	(17.6)	No. of GS Days w/ T > 32.2^{o} C	(18.2)
Percent forest cover	(15.2)	Dry sulfur deposition	(11.3)
Consec. day w/VPD>750pa,T>10^{o}C	(9.7)	Max # Days w/o PCP, T > 10^{o}C)	(10.1)
Soil pH	(8.7)	Soil pH	(8.3)
95 %-tile diurnal GS surface TMP	(7.5)	Percent soil clay	(7.6)
Non-GS precipitation	(6.8)	Topographic wetness index	(6.2)
Percent soil clay	(6.8)	Percent mixed-conifer cover	(5.8)
Flow accumulation	(5.4)	Percent forest cover	(4.9)
GS days >5.6^{o} C	(4.6)	Flow accumulation	(3.3)

Data Sources: Ameriflux, PRISM, STATSGO, EPA-CMAQ, NED, NLCD2000, E&S Env Chem

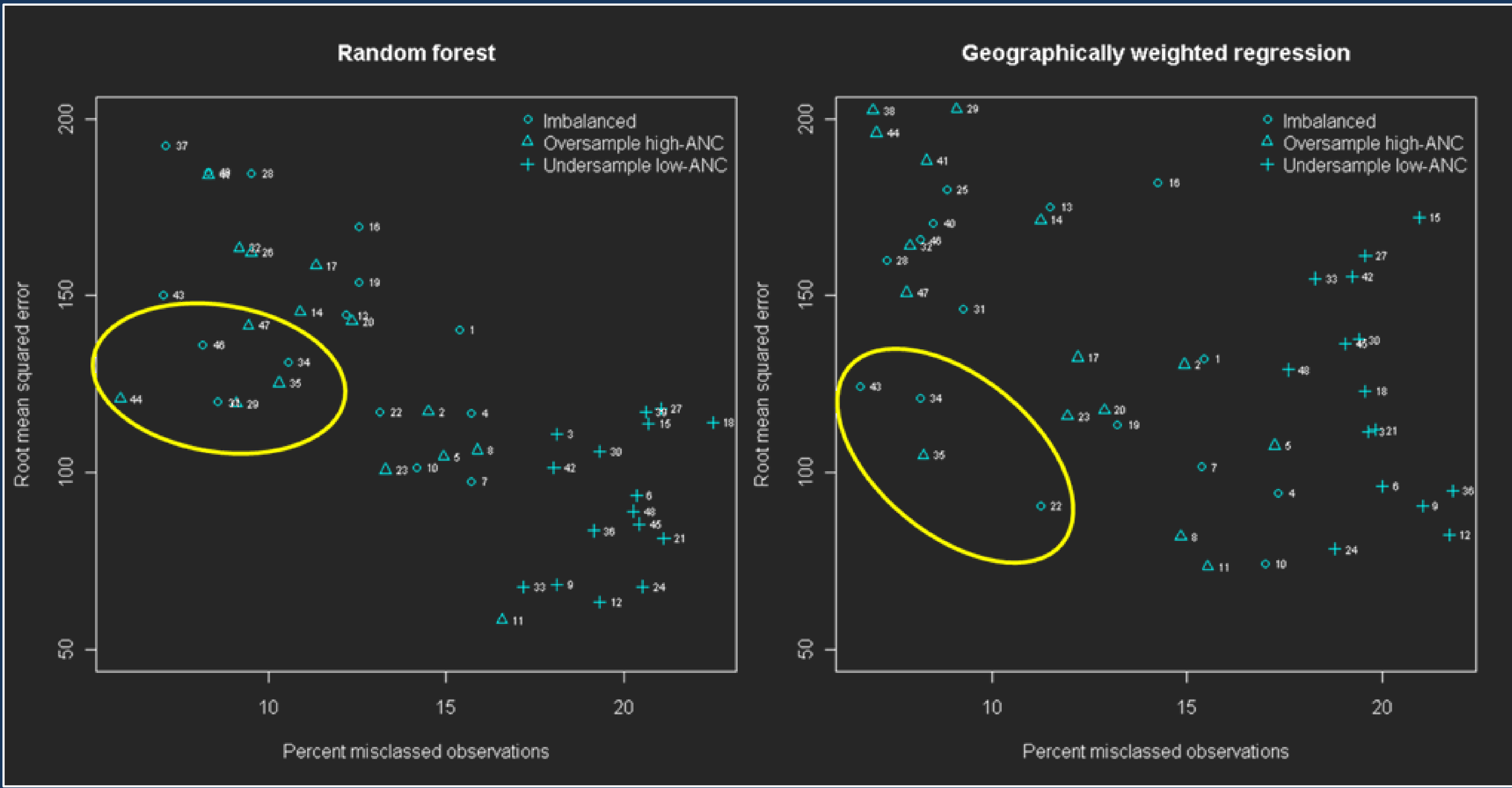
How we got to 10 predictors, continuous



Scatterplot of Gatekeeper model performance

48 unique combinations of the data, ea. RF & GWR, (3x4x4) below

1. Design balancing (3x): imbalanced$_{\text{as is}}$, oversampled ANC$_{\text{hi}}$, undersampled ANC$_{\text{low}}$)
2. Varying the ANC threshold (4x): [150, 200, 250, 300],
3. Varying the cutoff (4x): [0.4, 0.5, 0.6, 0.7]

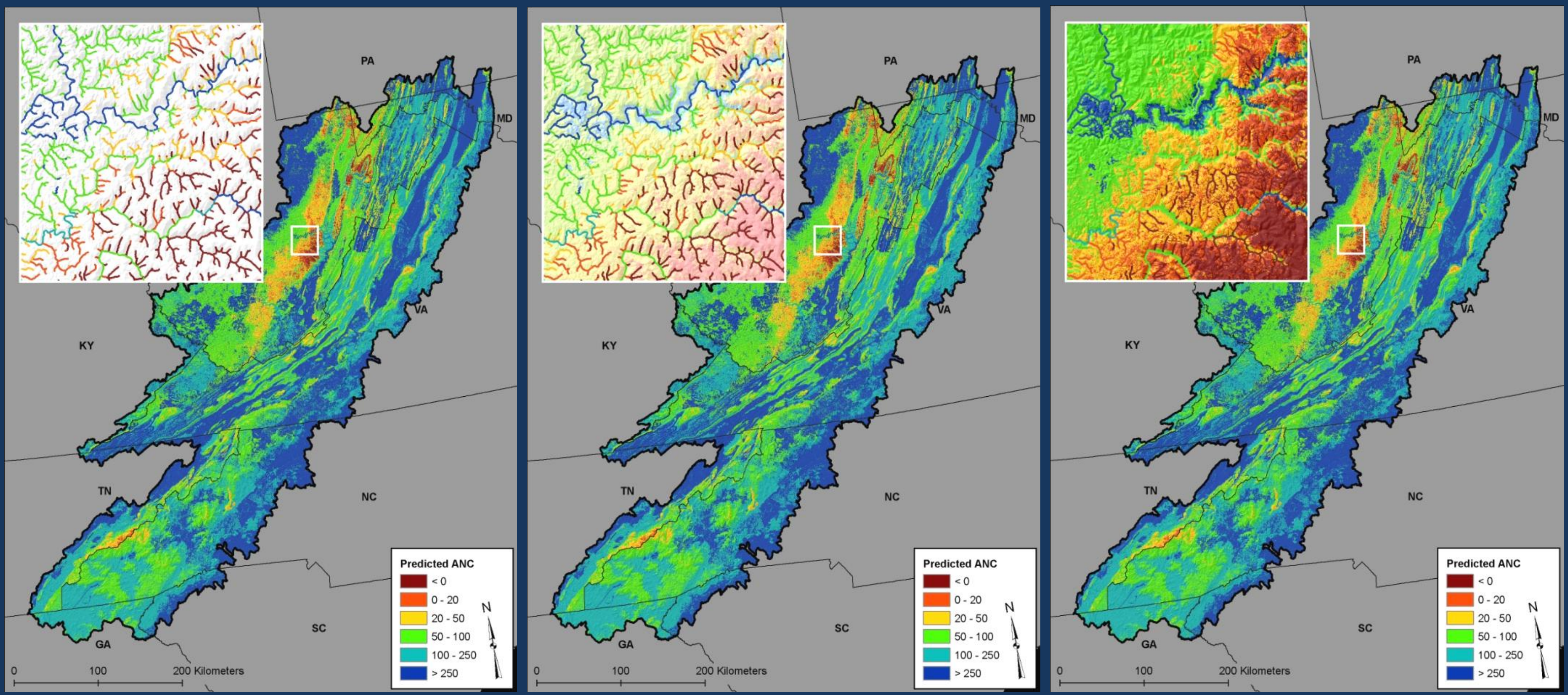


Best performing models

ID	Continuous model	Data sampling	Low ANC threshold	Prob. cutoff	Misclass. Rate (%)	RF RMSE	Gatekeeper RMSE
31	RF	Imbalanced	250	0.6	8.6	292.1	119.9
34	RF	Imbalanced	250	0.7	10.6	326.3	131.2
46	RF	Imbalanced	300	0.7	8.2	324.5	136.2
29	RF	Oversample	250	0.5	9.1	293.5	119.5
35	RF	Oversample	250	0.7	10.3	339.8	125.3
44	RF	Oversample	300	0.6	5.8	301.6	120.9
47	RF	Oversample	300	0.7	9.4	293.0	141.6
22	GWR	Imbalanced	200	0.7	11.2	341.5	90.7
34	GWR	Imbalanced	250	0.7	8.2	279.5	121.0
43	GWR	Imbalanced	300	0.6	6.6	338.0	124.3
35	GWR	Oversample	250	0.7	8.2	309.6	104.9

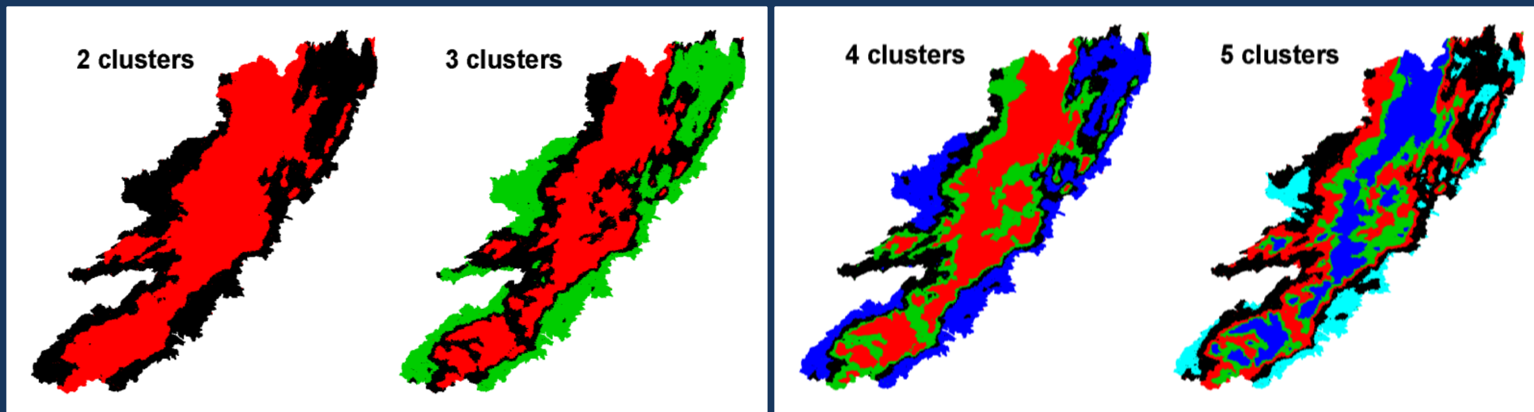
Predicted ANC, 250, 0.5, RF

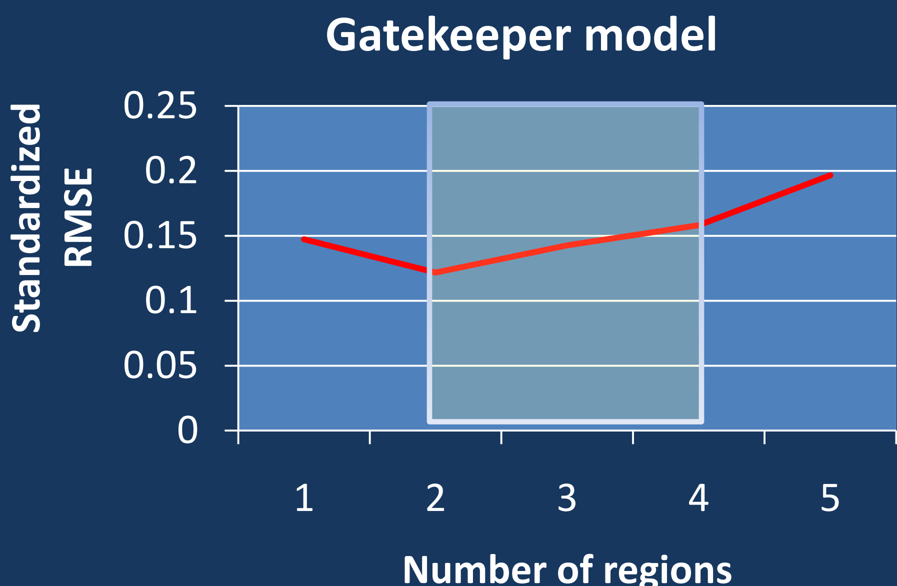
- Can be specified w/ geographic variants to the model

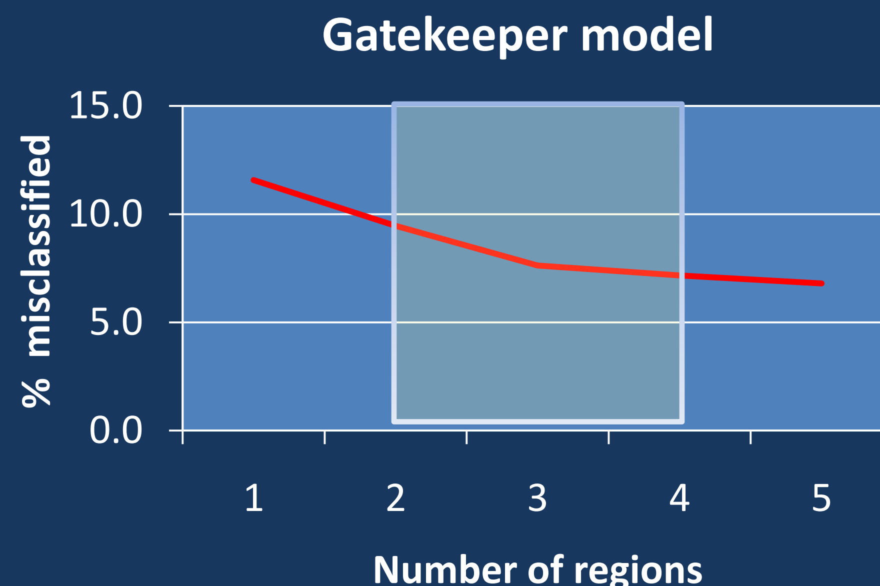


Note: We are predicting for 159 MM grid cells using 933 sampled cells; Stream networks represented by 4MM cells, more tractable, 1/40th

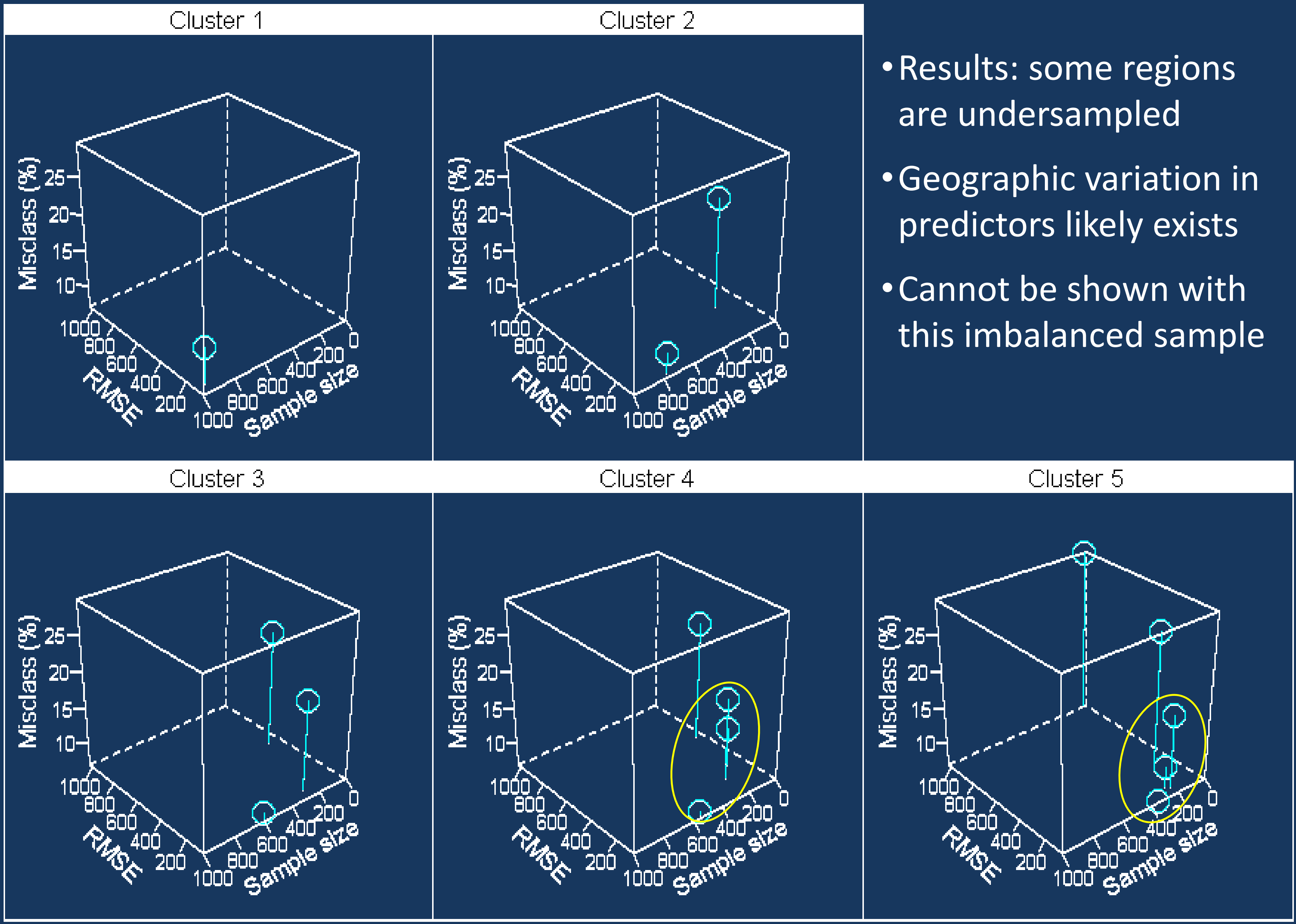
K-means clustered top 10 predictors into 2-5 regions







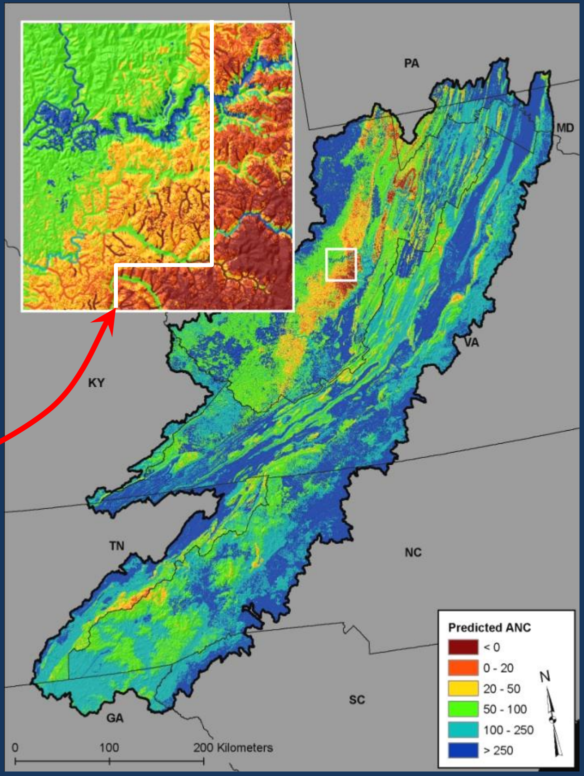
*RF model performance statistics for modeled regions



- Results: some regions are undersampled
- Geographic variation in predictors likely exists
- Cannot be shown with this imbalanced sample

Summary

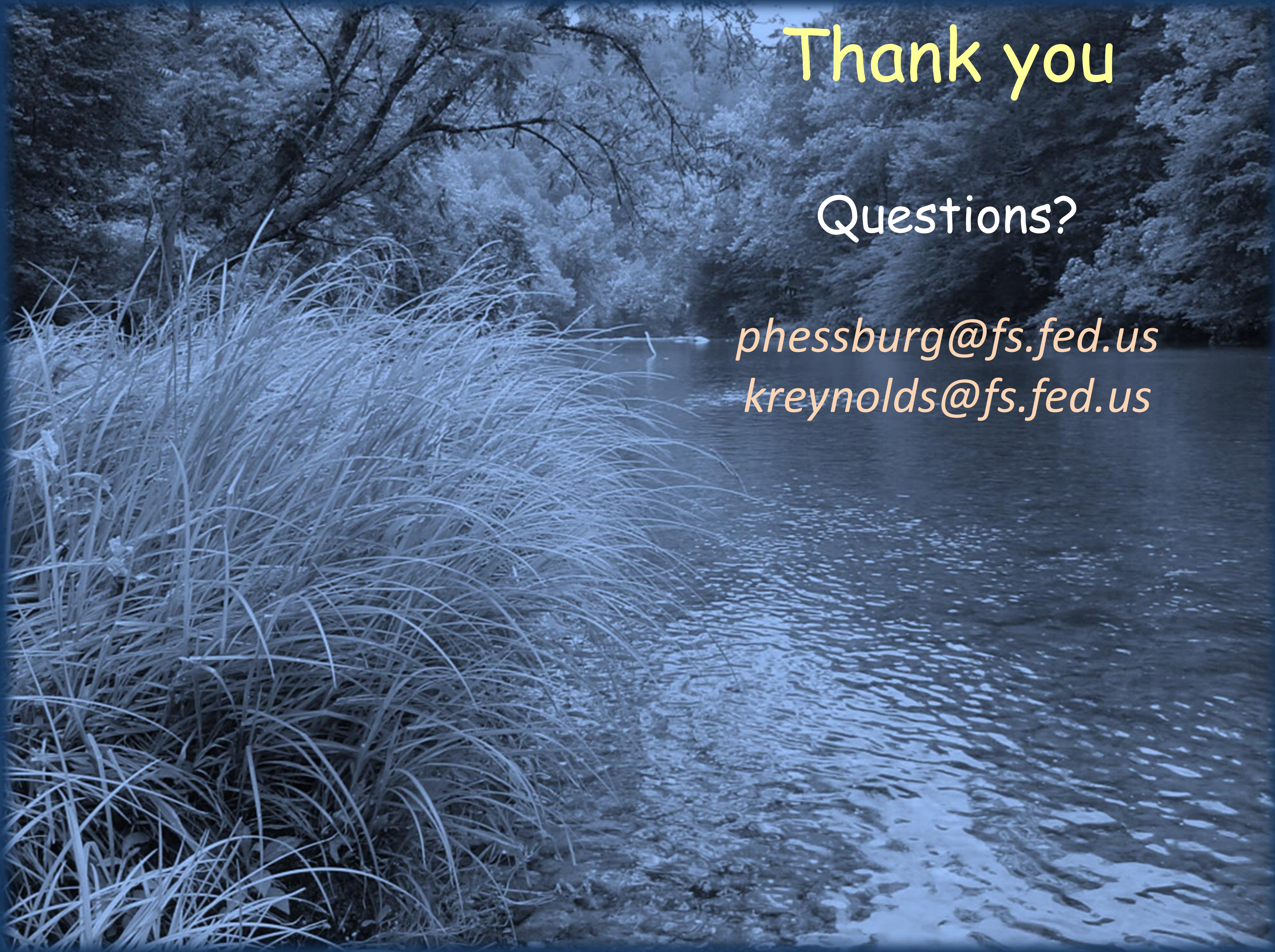
- ANC predictive model (imbalanced design) is reasonably robust to prediction & stable:
  - Gatekeeper model with RF explains ~60% of the multi-variance
  - Balancing design would:
    - Improve variance explained
    - Expose geographic variants
  - Highest uncertainty – what drives high ANC
    - Affects overall model predictions, low + high
- Lack of QC & evenness in data scaling w/ soils & geology data reduces model sensitivity
  - Better data → better model specificity
- Dry S deposition was leading predictor
  - expected to see wet S-deposition a stronger predictor in continuous model

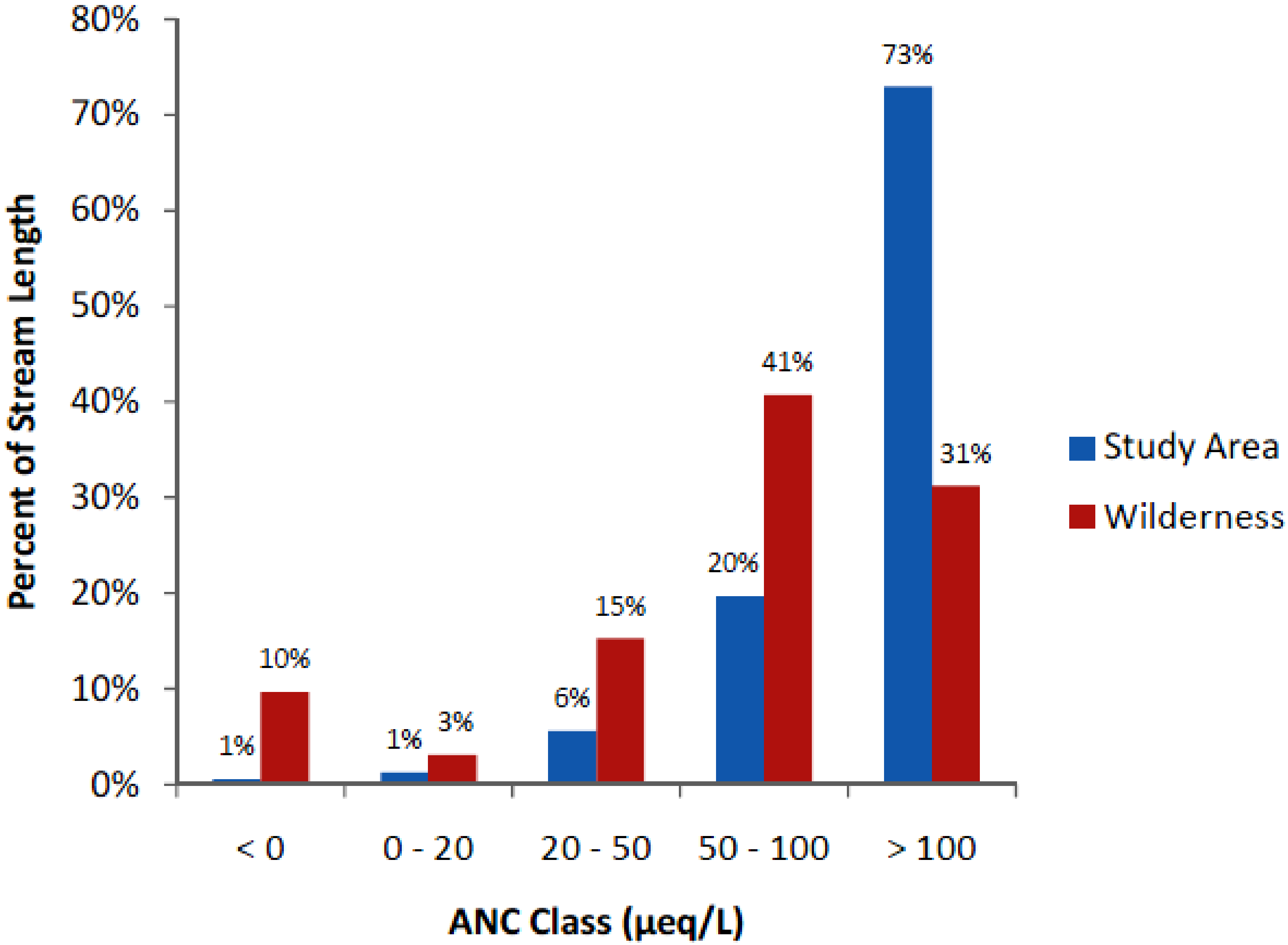


Thank you

Questions?

phessburg@fs.fed.us
kreynolds@fs.fed.us





Percent of Stream Length
80%
70%
60%
50%
40%
30%
20%
10%
0%
1%
10%
1%
3%
6%
15%
20%
41%
73%
31%
< 0
0 - 20
20 - 50
50 - 100
> 100
ANC Class (μeq/L)
Study Area
Wilderness