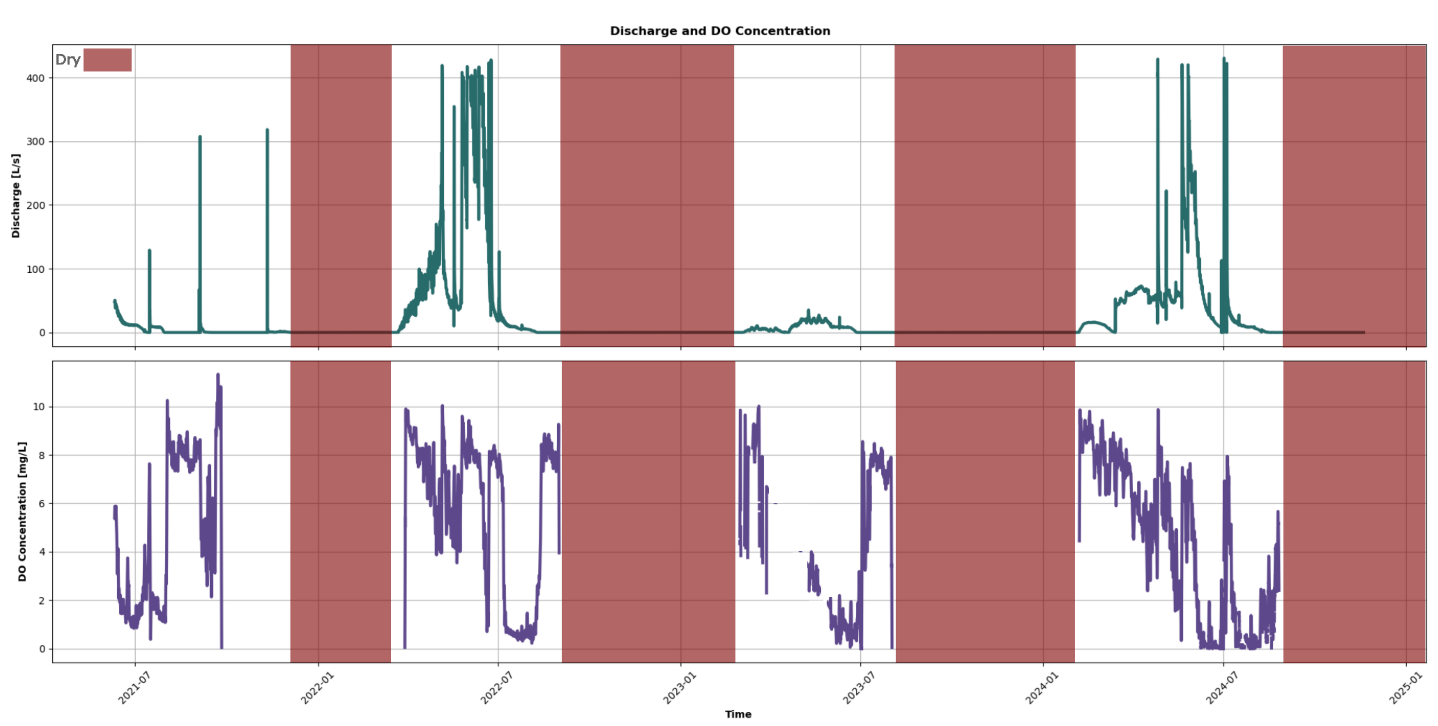
Rule Mining for Deoxygenation in Non-Perennial Kansas Catchments

Alexi I. Sommerville[1,2]; Erin C. Seybold[1,2], Kristin L. Underwood[3]; Donna Rizzo[3];

[1]University of Kansas Geology Department, Lawrence, KS; [2]Kansas Geological Survey, Lawrence, KS; [3]University of Vermont, Burlington, Vermont

Background

- Dissolved oxygen (DO) concentrations are a master control on biogeochemical rates in aquatic systems.
- DO behavior may vary as a function of flow intermittency in systems that experience drying and wetting.



- **Research Question**: What climatological and hydrologic variables and interactions drive low oxygen conditions in non-perennial streams?

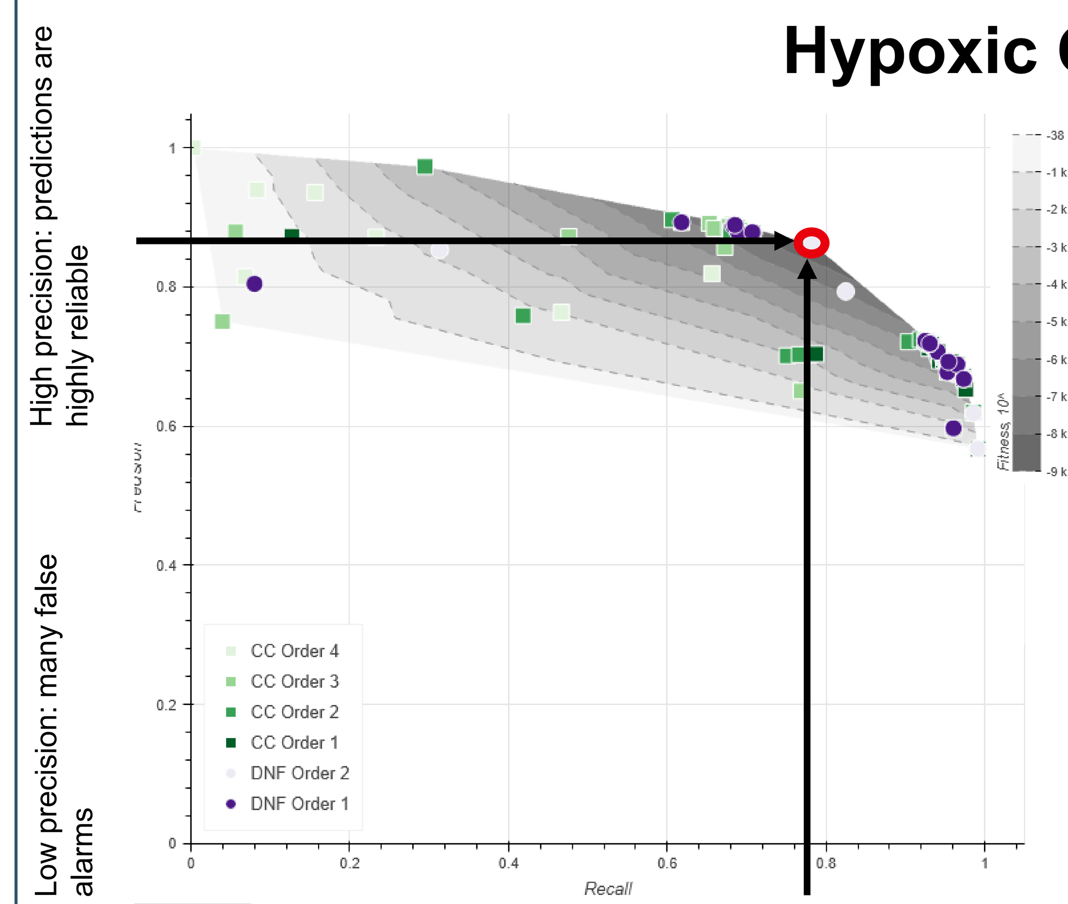
Modeling Approach

- **Objective**: Use machine learning-driven evolutionary computation algorithms to identify controlling features on non-perennial hypoxia
- **Tandem Evolutionary Algorithm** (TEVA):
  - Evolves predictive rules using conjunctive (AND) and disjunctive (OR) logic to identify combinations of features associated with hypoxic/oxic conditions
  - Manages equifinality, phase-dependent processes, high-dimensionality, and optimizes for both parsimony and performance

Input Variable	Units	Spatial Scale	Source	Temporal Resolution
Surface water temperature	C°	Outlet	Logger	15 minute
Surface water temperature	C°	Multiple upstream points	Logger	15 minute
Surface water temperature, time lagged (6hr, 12hr)	C°	Outlet	Logger	15 minute
Water depth	meters	Outlet	Logger	15 minute
Discharge	L/s	Outlet	Rating curve	15 minute
Discharge, rolling average (3h, 6h, 12h, 72h)	L/s	Outlet	Rating curve	15 minute
Velocity	m/s	Outlet	Derivation	15 minute
Velocity, rolling average (6h)	m/s	Outlet	Rating curve	15 minute
Groundwater gradient	meters	Outlet	Loggers	15 minute
Water source fraction (QBF)	L/s	Outlet	SC-EMMA	15 minute
Standardized Precipitation Evapotranspiration Index (SPEI)	unitless	grid, 800m	Derivation	Daily
Photosynthetically-active radiation (PAR)	$\mu mol \cdot m^{-2} \cdot s^{-1}$	Network	Derivation	Daily
Precipitation	mm/d	grid, 800m	PRISM	Daily
Precipitation, rolling average (72hr, 1wk)	mm/d	grid, 800m	PRISM	Daily
Alpha centrality	unitless	Network	streamDAG	Daily
Alpha centrality, weighted	unitless	Network	streamDAG	Daily
Alpha centrality, scaled	unitless	Network	streamDAG	Daily
Percent network wet	Percentage	Network	streamDAG	Daily
Harary index	unitless	Network	streamDAG	Daily
Active upstream length	km	Network	streamDAG	Daily

Results

Hypoxic Outcome



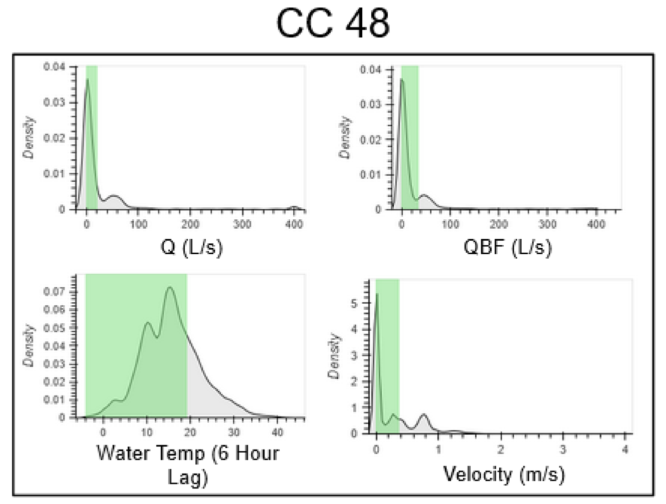
Low recall: many true events missed

High recall: many true events detected

1. Conjuctive and disjunctive clauses are both seeing both **high recall and high precision** (high coverage and high accuracy)
2. DCs outperforming CCs
3. CCs maintain high precision but miss some true hypoxic instances, suggesting single-rule interactions capture core drivers yet overlook alternative hypoxia pathways

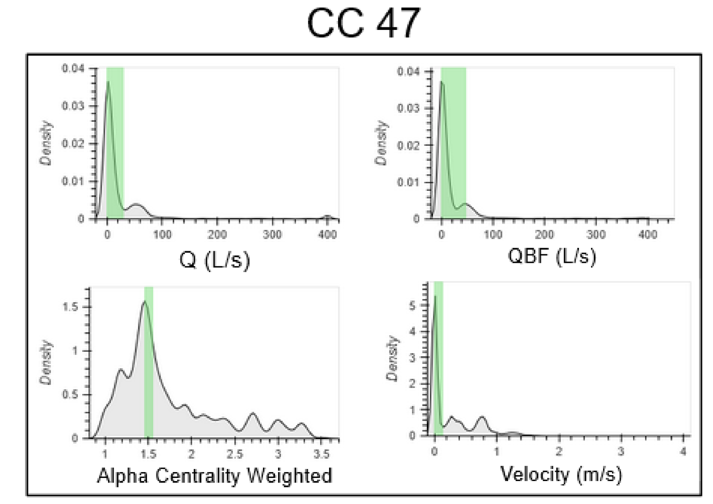
Influential Feature Ranges

CC 48



OR

CC 47

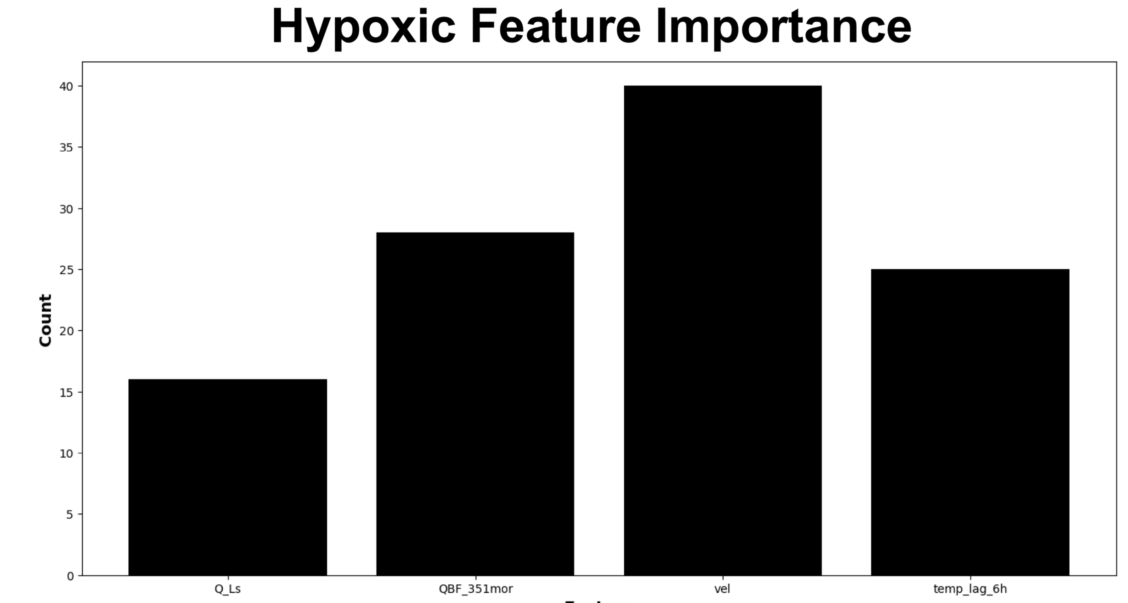


Lagged water temperature:
- Cooler temps capturing nighttime diurnal signal (photosynthesis low, respiration continues, reaeration low, temp low)
- Cooler temps capturing greater nighttime subsurface contribution

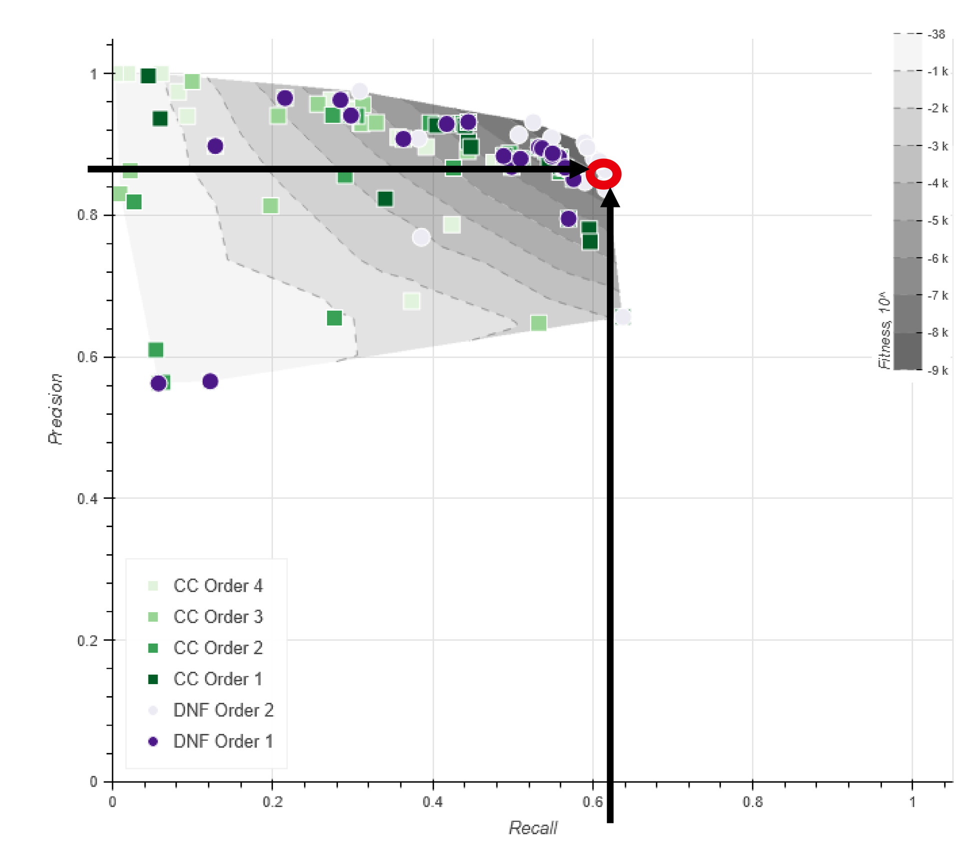
Hydrologic controls dominant in both feature sets
1. Hypoxia occurring in low discharge, low velocity, and high groundwater input scenarios

Network connectivity:
- Lower alpha centrality limiting reaeration, transport dynamics, and flow

Hypoxic Feature Importance



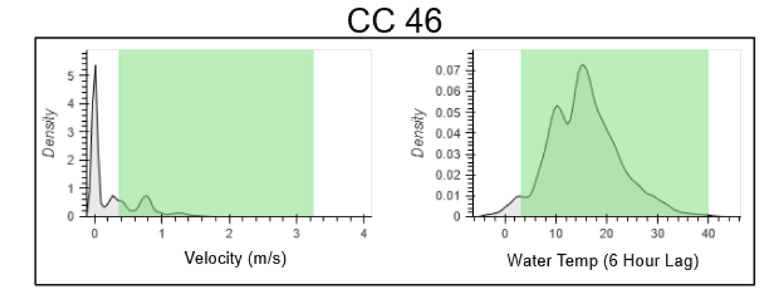
Oxic Outcome



1. High precision in CCs and DCs, but **both seeing low recall** (CC recall < .41, DC recall <0.46) –alternative pathways to high oxygen states missed in solution space
2. **DCs significantly outperforming** CCs: non-hypoxic conditions arise from a wider array of independent processes – single combinations can't capture diverse routes to high-oxygenation states

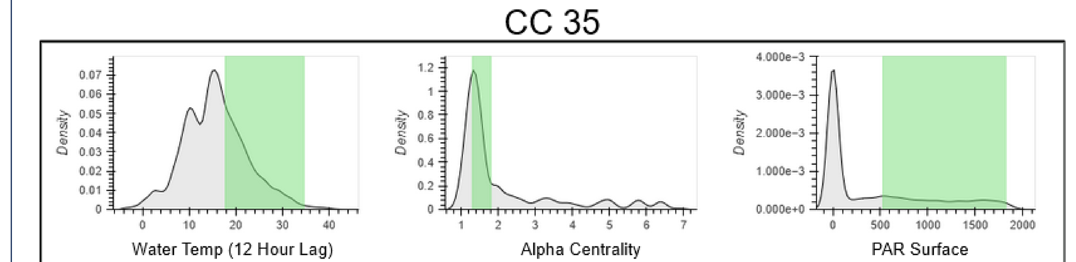
Influential Feature Ranges

CC 46



OR

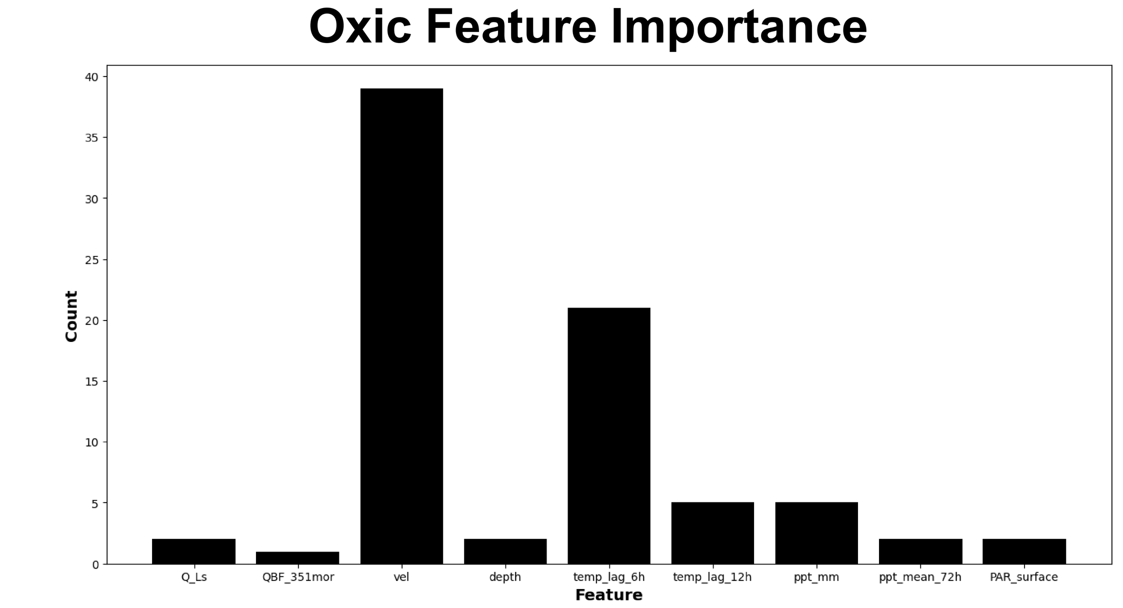
CC 35



High flow feature set:
- High water velocity forcing reaeration and reoxygenation
- Variable lagged water temperature representing diurnal solar forcings (low at night, high during day) – not about absolute temperature, but variability

Controls shift to local biogeochemistry:
- Higher light availability (PAR enabling oxygen input via photosynthetic activity
- Warmer temps proxy for past high ecosystem productivity
- Low network connectivity preserves local DO production rather than mixing/transporting away

Oxic Feature Importance



Discussion and Next Steps

Global Conclusions:
- Clear equifinality in drivers of hypoxic and oxic states – processes and variable categories shift according to season + hydrograph
- Hypoxic states: **driven largely by hydrologic controls** – low discharge, velocity, high groundwater sourcing are ubiquitous across solutions
- Oxic states: drivers vary significantly – **either high velocity** enabling constant reaeration, **or high PAR** enabling photosynthesis-based replenishment

Next Steps:
- Evaluate feature selection capabilities of TEVA by using selected features in DO forecasting model