SIGNALS AND SEDIMENT

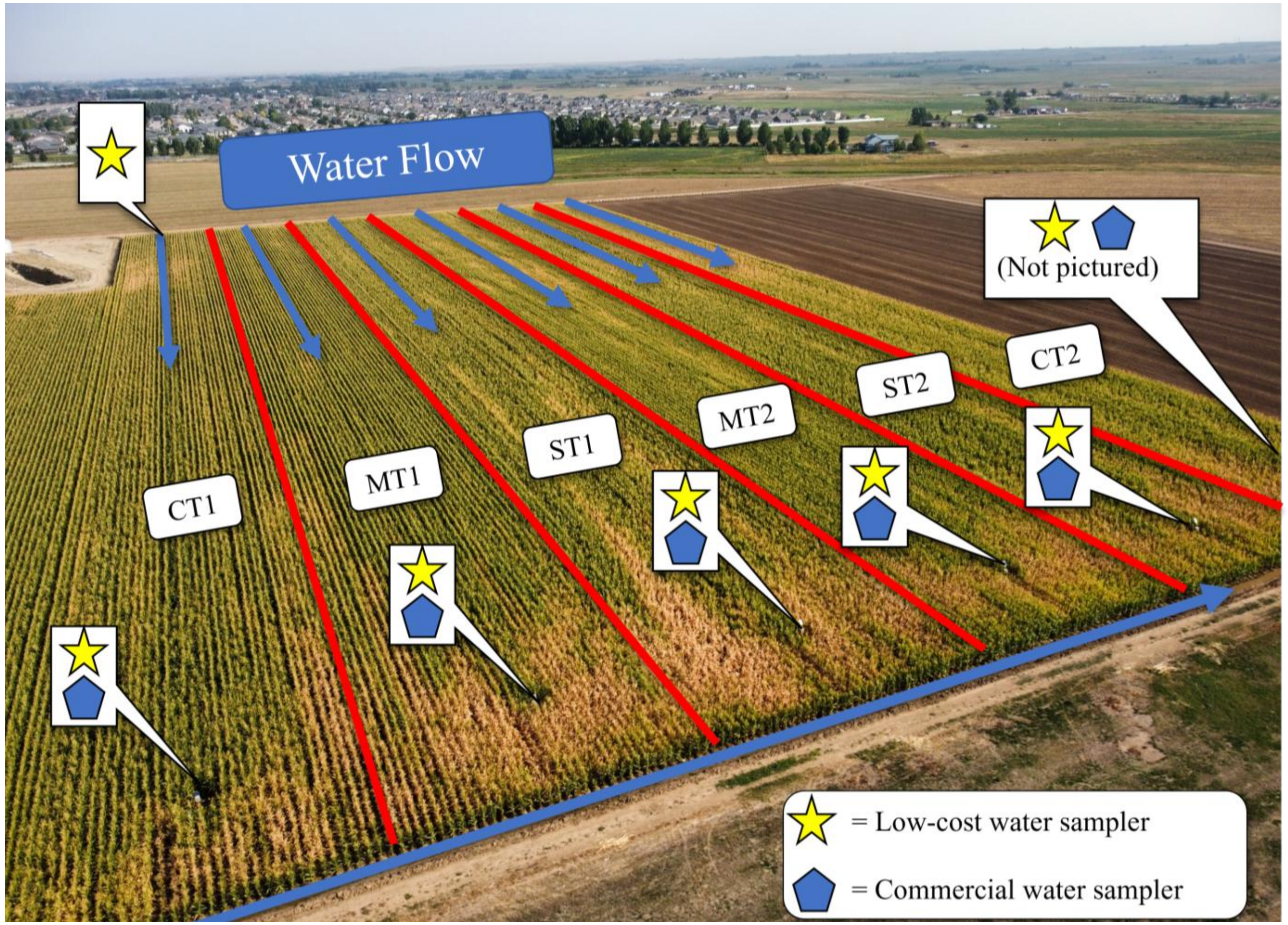
A BAYESIAN DECOMPOSITION OF TILLAGE IMPACTS AND LEGACY EFFECTS IN LONG-TERM RUNOFF WATER QUALITY

A.J. Brown
15 Dec 2025

Background

- Conservation tillage study initiated in 2010
- Planned with key farming stakeholders to ensure applicability to real-world scenarios
- Data have been collected over the 2011 – 2025 period using multiple methods for water quality and quantity
- Overarching goal of investigating the impacts of tillage on runoff water quality, soil health, and agronomic/economic productivity
- As of Oct 2025, all water quality, tillage, and crop data have been compiled and merged for research use

Livin' on a Layer (of Residue)



- Study site in Fort Collins, CO, USA
- 2011-2025
- Conservation Tillage Study started in 2011
  - `CT = Full tillage`
  - `MT = Less tillage`
  - `ST = Least tillage`
- Replicated treatment blocks (1 & 2)
- Each plot is about 1 hectare with a 300m irrigation run length
- Surface irrigated with siphon tube or gated pipe for 7-10 hours at a time

Diving deeper: Investigating tillage impacts on water quality

Trimarco et al., 2025 found:

- Reduced tillage cut particulate pollutant concentrations.
- Better soil structure and organic matter aligned with cleaner runoff.
- Strip and minimum till performed similarly and both outperformed conventional tillage.

What remains?

- Investigation into volume and loadings between treatments over time
- Quantifying the actual causal impact of tillage on various analyte concentration and load
- Investigating persistence effects of tillage on analytes, if any (unique aspect of these long-term data)

Proposed Research Objectives*

Characterize treatment differences in water-quality loads through time

- Use the long-term edge-of-field dataset to estimate how sediment and nutrient loads differed among conventional, strip, and minimum tillage, accounting for year-to-year variability and uncertainty in flow and concentration.

Estimate the causal effect of tillage on individual analytes

- Apply hierarchical Bayesian models to isolate the effect of tillage on concentrations and loads for each analyte, separating true treatment signals from measurement error, missing data, and shared temporal structure across analytes.

Identify whether tillage influences persist across years

- Leverage the multi-year record and the multi-output Gaussian process to determine whether tillage effects are short-lived or temporally persistent, evaluating the degree to which past disturbance continues to shape present-year water-quality outcomes.

**These objectives reflect ongoing analyses conducted as part of my Ph. D. research and will contribute directly to my dissertation.*

Problem: Tillage varies in reality

- Tillage treatments had fairly standard procedures
  - MT = Vertical Tillage + liquid N + banded P
  - ST = Strip Tillage + liquid N + banded P
  - CT = Moldboard Plow + broadcast N and P
- However, some years merit change due to practical considerations
  - E.g., VT whole field to prep for wheat planting the day following corn harvest in 2023

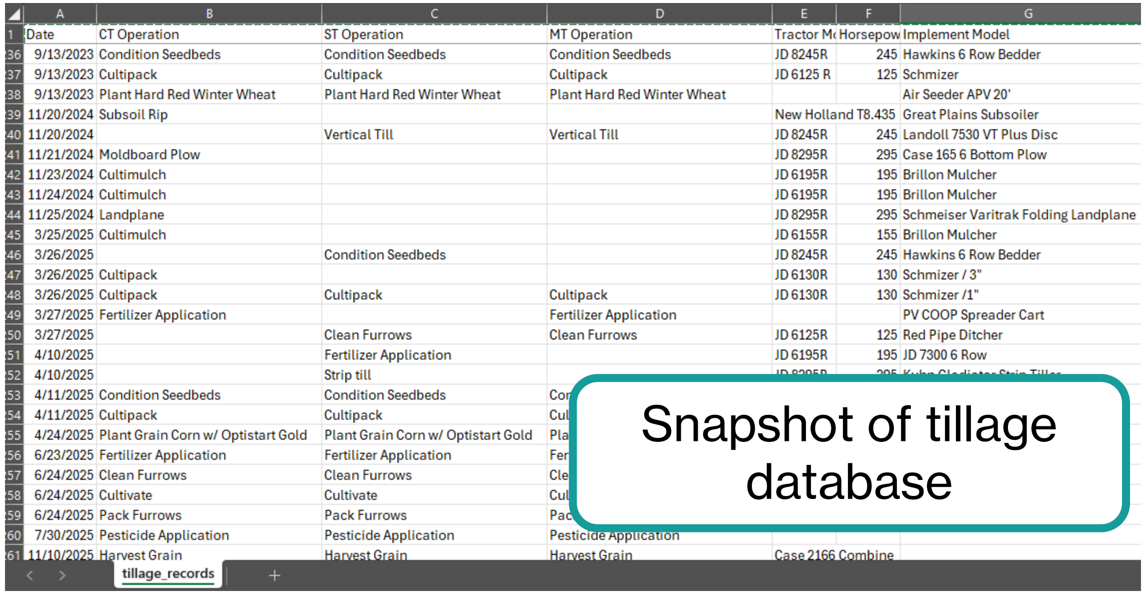
STIR Me Up Before You Go-Go

TERM	DESCRIPTION	CONVERSION	NOTES
Speed [km/h]	Average implement speed	÷ 1.609 → mph	Converts from km/h to miles per hour
TTM	Tillage Type Modifier [0–1]	none	Represents the aggressiveness of the implement (0 = no till, 1 = full inversion)
Depth [cm]	Average operating depth	÷ 2.54 → inches	Converts from centimeters to inches
Surf_Disturbance [%]	Percent of soil surface disturbed	÷ 100 → fraction	Captures the area of soil affected by the implement
STIR	Soil Tillage Intensity Rating	–	Dimensionless intensity index

- Soil Tillage Intensity Rating (STIR)
- Continuous value that quantifies the overall physical disturbance to the soil caused by a tillage operation
- STIR is used in multiple mechanistic models for soil physical impacts (e.g., RUSLE2)
- For Kerbel: allows us to account for varying tillage in years where CT/MT/ST treatments were not consistent

$$STIR = \left(0.5 * \frac{Speed_{\left[\frac{km}{h}\right]}}{1.609}\right) * (3.25 * TTM) * \left(\frac{SurfDisturbance}{100}\right) * \left(\frac{Depth_{[cm]}}{2.54}\right)$$

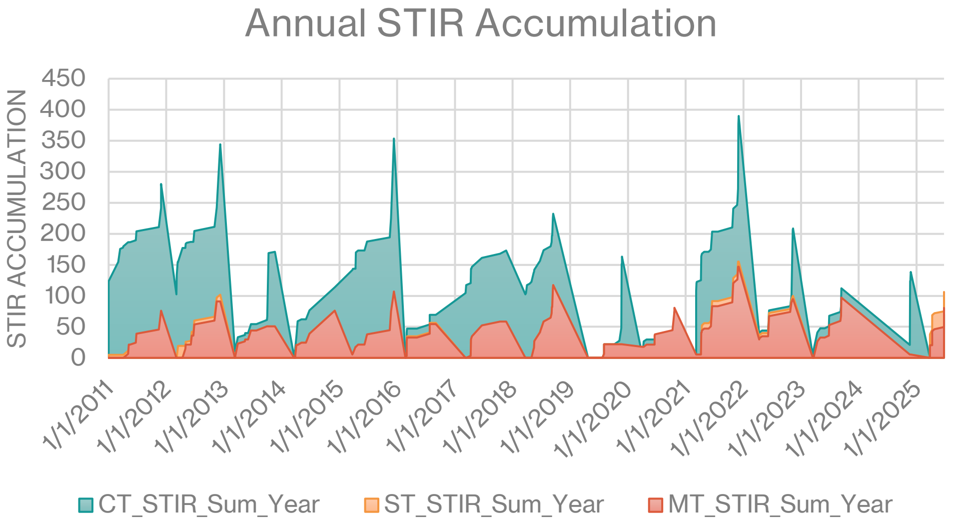
STIR Results for Kerbel 2011 - 2025

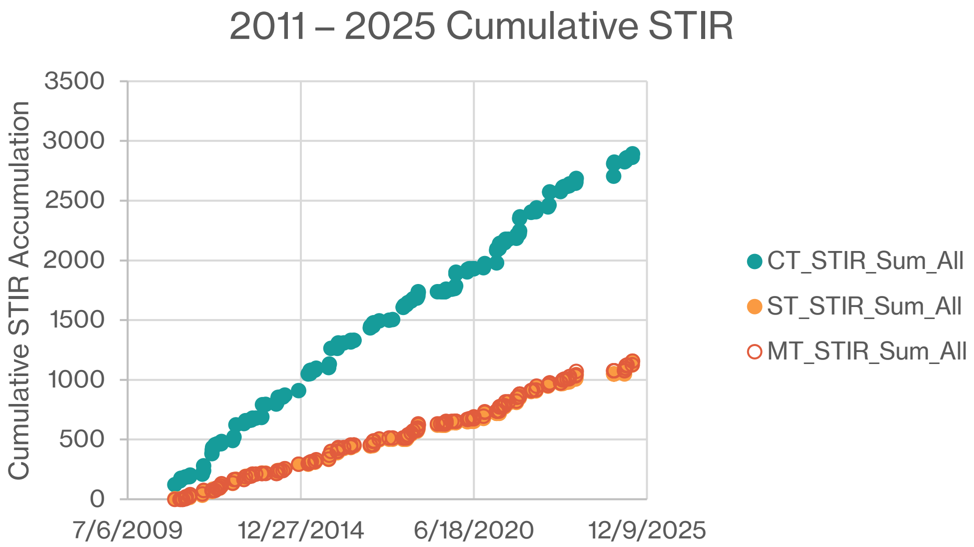


	A	B	C	D	E	F	G
1	Date	CT Operation	ST Operation	MT Operation	Tractor Mo	Horsepow	Implement Model
36	9/13/2023	Condition Seedbeds	Condition Seedbeds	Condition Seedbeds	JD 8245R	245	Hawkins 6 Row Bedder
37	9/13/2023	Cultipack	Cultipack	Cultipack	JD 6125 R	125	Schmizer
38	9/13/2023	Plant Hard Red Winter Wheat	Plant Hard Red Winter Wheat	Plant Hard Red Winter Wheat			Air Seeder APV 20'
39	11/20/2024	Subsoil Rip			New Holland T8.435		Great Plains Subsoiler
40	11/20/2024		Vertical Till	Vertical Till	JD 8245R	245	Landoll 7530 VT Plus Disc
41	11/21/2024	Moldboard Plow			JD 8295R	295	Case 165 6 Bottom Plow
42	11/23/2024	Cultimulch			JD 6195R	195	Brillon Mulcher
43	11/24/2024	Cultimulch			JD 6195R	195	Brillon Mulcher
44	11/25/2024	Landplane			JD 8295R	295	Schmeiser Varitrak Folding Landplane
45	3/25/2025	Cultimulch			JD 6155R	155	Brillon Mulcher
46	3/26/2025		Condition Seedbeds		JD 8245R	245	Hawkins 6 Row Bedder
47	3/26/2025	Cultipack			JD 6130R	130	Schmizer / 3"
48	3/26/2025	Cultipack	Cultipack	Cultipack	JD 6130R	130	Schmizer /1"
49	3/27/2025	Fertilizer Application		Fertilizer Application			PV COOP Spreader Cart
50	3/27/2025		Clean Furrows	Clean Furrows	JD 6125R	125	Red Pipe Ditcher
51	4/10/2025		Fertilizer Application		JD 6195R	195	JD 7300 6 Row
52	4/10/2025		Strip till				
53	4/11/2025	Condition Seedbeds	Condition Seedbeds	Con			
54	4/11/2025	Cultipack	Cultipack	Cul			
55	4/24/2025	Plant Grain Corn w/ Optistart Gold	Plant Grain Corn w/ Optistart Gold	Pla			
56	6/23/2025	Fertilizer Application	Fertilizer Application	Fer			
57	6/24/2025	Clean Furrows	Clean Furrows	Cle			
58	6/24/2025	Cultivate	Cultivate	Cul			
59	6/24/2025	Pack Furrows	Pack Furrows	Pac			
60	7/30/2025	Pesticide Application	Pesticide Application	Pesticide Application			
61	11/10/2025	Harvest Grain	Harvest Grain	Harvest Grain	Case 2166 Combine		

tillage_records

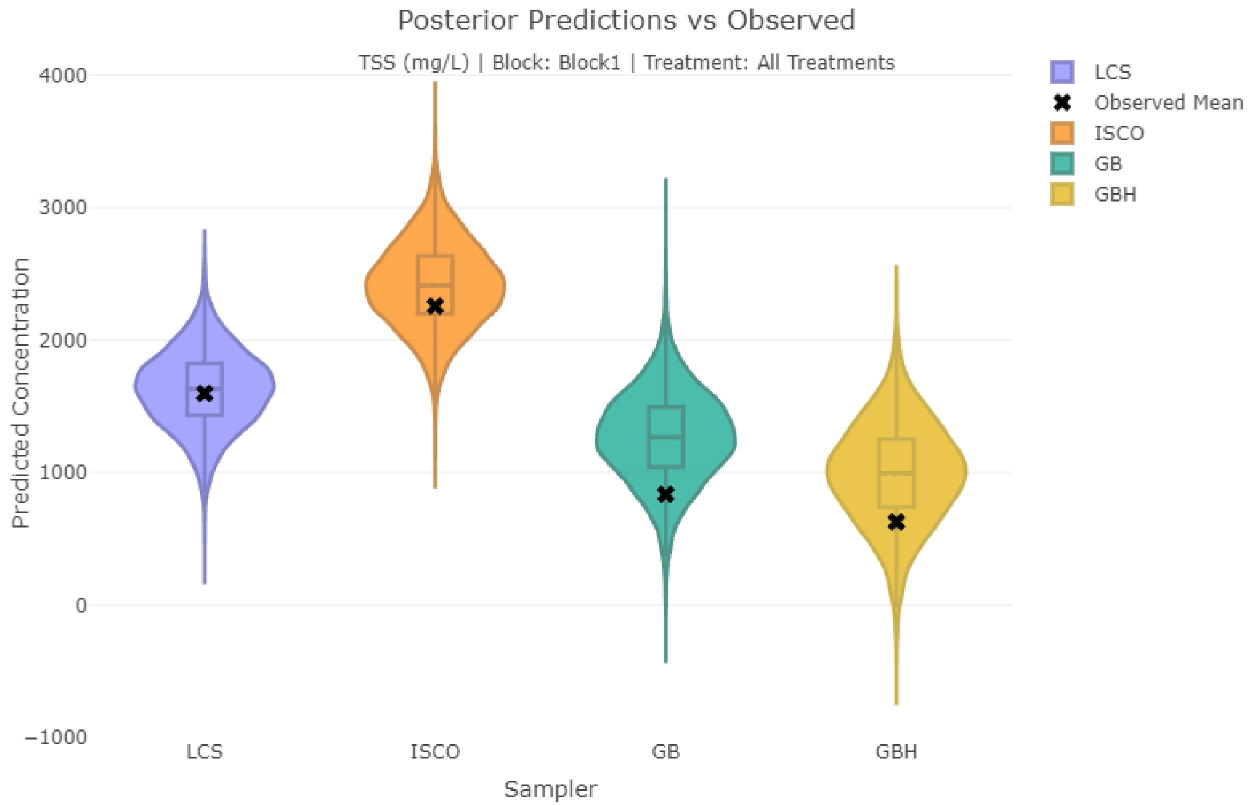
Snapshot of tillage database





Starting with Tad's compilation of tillage records as prepared for DayCent (a less robust version of the STIR), I matched these activities (and added 2024 and 2025) to STIR values with assumed tractor speeds from the MOSES (Management Operations and Soil Erosion Simulation) database from NRCS

Problem: Measurement is Hard



TSS comparison example: Our recent, Brown et al., 2025 paper is a great example of trying to account for this in sampler method

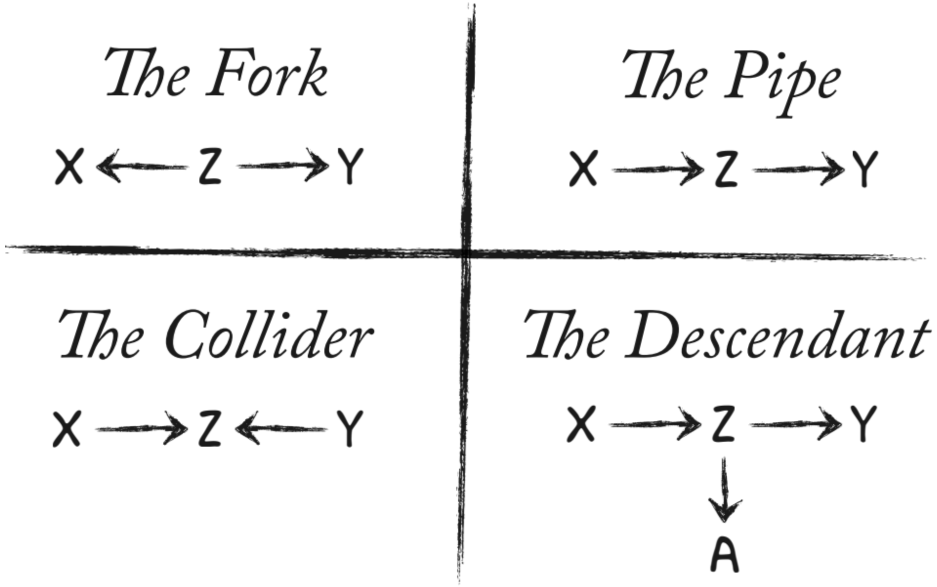
- Over the years, **various methods** have been used to collect water quality and quantity
- For example:
  - **Stage measurement:** Pressure Transducer, Ruler, Bubbler, and Etape
  - **Flume type:** Weir, 7” V-notch, 8” V-notch, 10” V-notch, and 60° V-Trapezoidal Flume (current)
  - **Sampler method:** GBW, ISCO, GB, LCS, GB3, and GBH
  - Analytes have also changed over years
- These various methods have varying certainties, which need to be accounted for to harmonize a long-term dataset like this.
- Additionally, there are tons of missing data for various reasons related to resource availability for any given year.

The Times They Are A-Changin', So Let's Use a Bayesian Causal Model

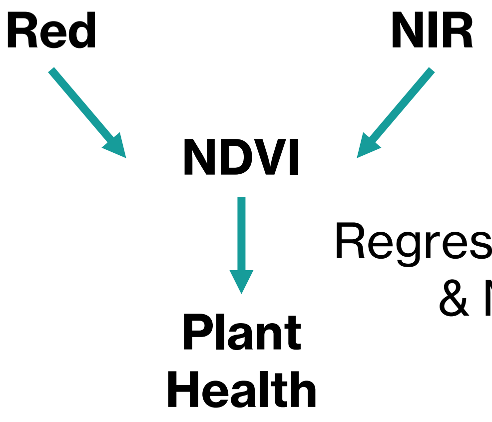
Thomas Bayes (1761), Presbyterian minister (and statistician) of the Mount Sion chapel in Kent, UK

- Causal models help define a statistical model with scientific assumptions
- Statistical models calibrated with Bayesian methods (e.g., Hamiltonian Monte Carlo) provide several advantages
  - Imputation of missing data
  - Handles small, irregular, and unbalanced data well with informed priors
  - Propagates measurement error explicitly
  - Models variables jointly, capturing shared correlations for extra 'horsepower'
  - And more!

Confounders from (McElreath, 2023)



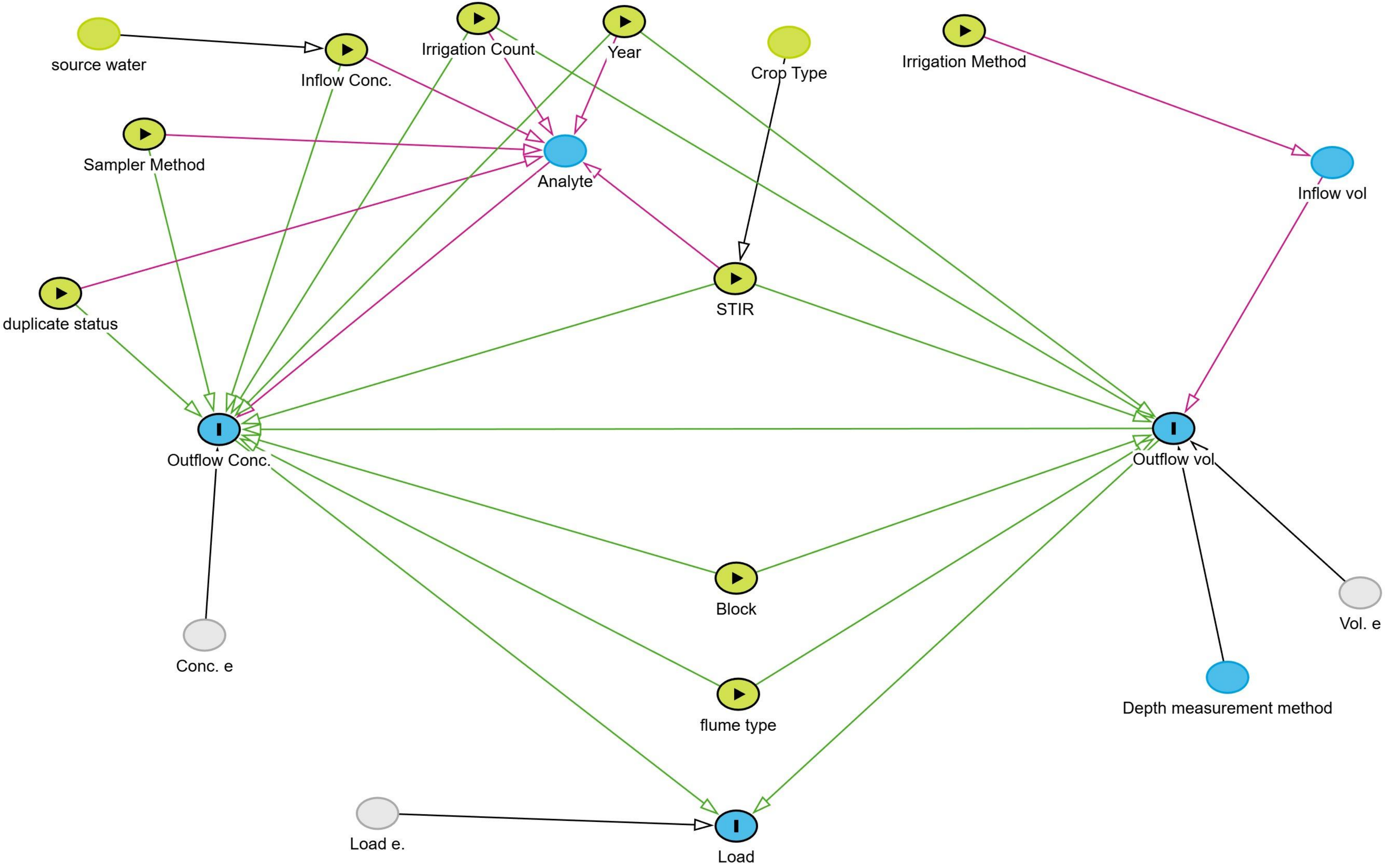
AJ Agronomy Example



Regress by NDVI OR Red & NIR NOT both

Under Pressure... From Confounders

- Directed Acyclical Graphs (DAGs) define rules for developing statistical models by following the rules of “do-calculus” i.e., P(y | do(x))
- This is where the scientist’s expertise comes in
- So, I made one of these for our Kerbel water dataset...



source water
Inflow Conc.
Irrigation Count
Year
Crop Type
Irrigation Method
Sampler Method
Analyte
Inflow vol
duplicate status
STIR
Outflow Conc.
Outflow vol
Block
Conc. e
Vol. e
flume type
Depth measurement method
Load e.
Load

Model You Like a Hurricane

- 2 simultaneous models are needed to quantify the direct impact of STIR on load:
  1. STIR → Volume
  2. STIR → Concentration
- What needs to be included in these regression models?
  1. STIR Only
  2. STIR, Analyte, Block, Inflow Conc., Irrigation Count, Outflow vol, Sampler Method, Year, duplicate status, flume type
- In Bayes, we can run two models simultaneously to let them learn from each other, and it's often needed in cases like these, termed, "Full Luxury Bayes"

"Full Luxury Bayes"

Model 1: STIR → Runoff Volume (9845 missing data, i.e., NAs)

$$V_i^{\text{obs}} \sim \text{Normal}(a_V + b_V\, S_i, \sigma_V).$$

Model 2: STIR → Concentration (5599 NAs)

$$\mu_i = \alpha_a + \beta_a^{(STIR)}\, S_i + \beta_a^{(CIN)}\, \text{CIN}_i + \beta^{(VOL)}\, V_i + \beta^{(IRR)}\, \text{IRR}_i + \beta_a^{(DUP)}\, \text{DUP}_i + \gamma_{a,B_i}^{(B)} + \gamma_{a,S_i}^{(S)} + \gamma_{a,F_i}^{(F)} + f_{a,Y_i}.$$

f_{a,Y_i} is a special parameter called a multi-output Gaussian Process (**MOGP**).

The MOGP is some fancy matrix algebra that:

- Learns a year-to-year trend that all analytes share to some degree (i.e., the “Kernel function”).
- Allows each analyte to respond differently to that trend, so they can move together or independently.
- Borrows “strength” across analytes and years, improving estimates when data are sparse, missing, or noisy.
- Provides realistic uncertainty by modeling temporal variation directly rather than treating it as random noise.

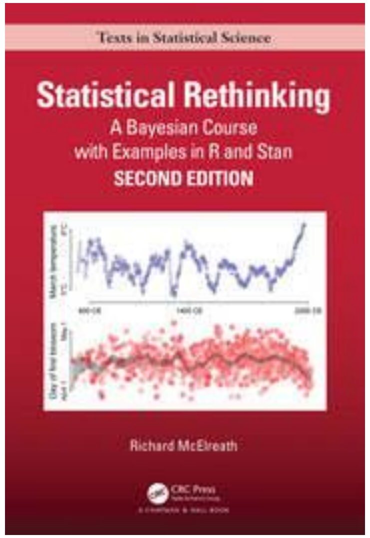
Bonus model: Inflow NA values (9999 NAs)

$$\text{CIN}_i^{\text{obs}} \sim \text{Normal}(0, 1),$$

In Bayes, we can (and should) model missing data as another parameter, resulting in many many parameters like we saw here

“Stan” by Me

- Run model in R
- 50 minutes, 0.7 GBs of result
- 27, 415 parameters fit on 15,366 data points??
  - Normal in Bayes ☺
- Model “converges” well!



Rethinking package

Stan Software

```
# 3. Hierarchical Bayesian model

```{r fit-model, eval=FALSE}
# ulam cannot handle multi-output GP directly, so we write a custom Stan program in next block
m_stir_nc <- ulam(
  alist(
    ##################################################
    # 1. Likelihood and linear predictor
    ##################################################
    C ~ normal(mu_C, sigma_analyte[A]),

    mu_C <- alpha[A] +
            beta_stir[A] * STIR +
            beta_cin[A]  * CIN +
            beta_vol     * VOL +
            beta_irr     * IRR +
            beta_dup[A]  * DUP +
            gamma_B[A, B] +
            gamma_S[A, S] +
            gamma_F[A, Fu] +
            tau_year[A] * g_year[Y],

    ##################################################
    # 2. Volume model (with missing data)
    ##################################################
    VOL ~ normal(mu_V, sigma_V),
    mu_V <- a_V + b_V * STIR,

    ##################################################
    # 2b. Inflow concentration model
    ##################################################
    CIN ~ normal(0, 1),

    ##################################################
    # 3. Analyte-level MVN (non-centered)
    ##################################################
    transpars> vector[A_n]:alpha     <<- mu_A[1] + v_A[,1],
    transpars> vector[A_n]:beta_stir <<- mu_A[2] + v_A[,2],
    transpars> vector[A_n]:beta_cin  <<- mu_A[3] + v_A[,3],
    transpars> vector[A_n]:beta_dup  <<- mu_A[4] + v_A[,4],

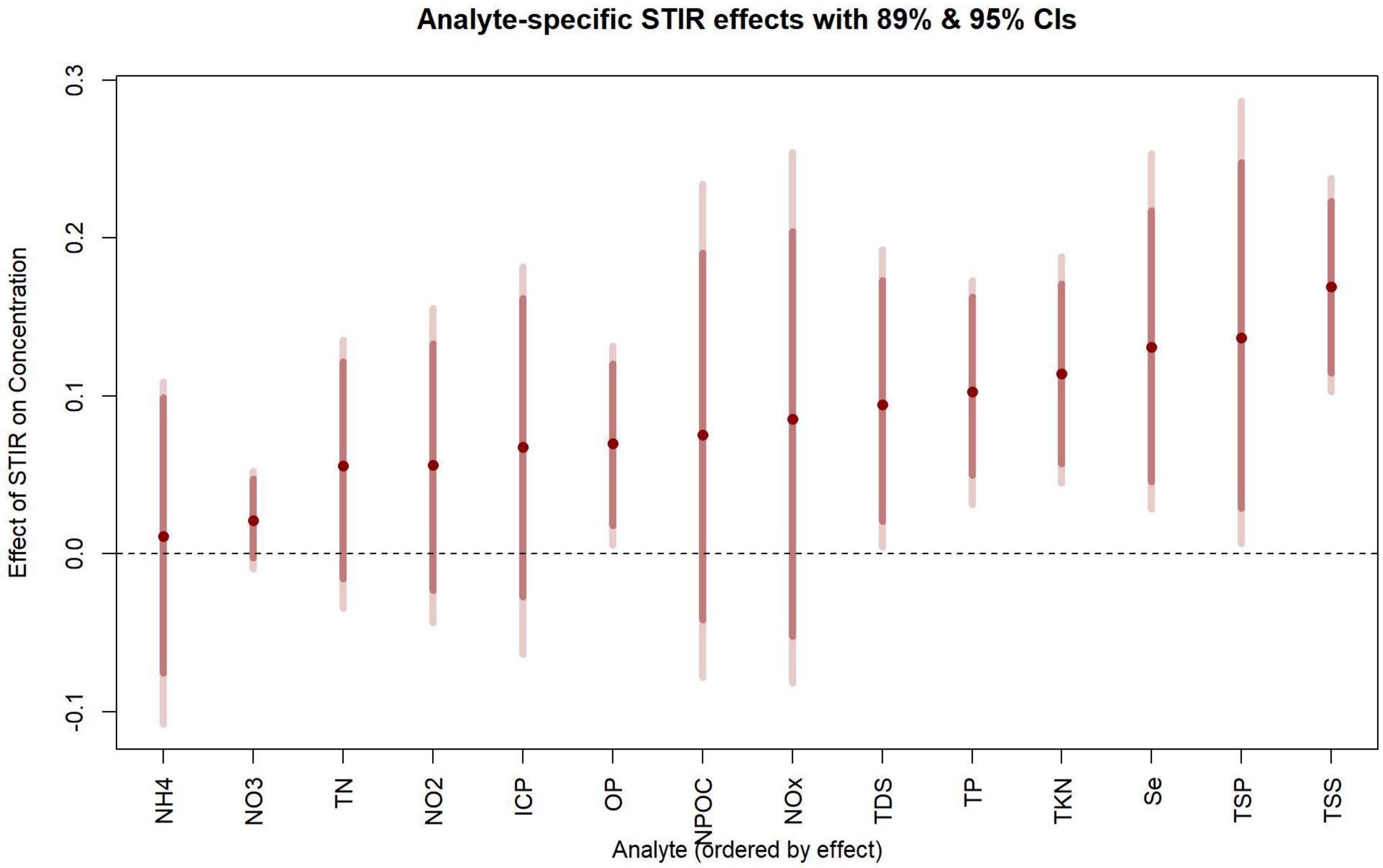
    transpars> matrix[A_n,4]:v_A <- compose_noncentered(sigma_A, L_A, Z_A),

    matrix[4, A_n]:Z_A ~ normal(0, 1),
    vector[4]:mu_A ~ normal(0, 1),
    cholesky_factor_corr[4]:L_A ~ lkj_corr_cholesky(2),
    vector[4]:sigma_A ~ exponential(1),

    ##################################################
```
```


Hit Me With Your Best Plot
(Results so far)
KINCAID
C6

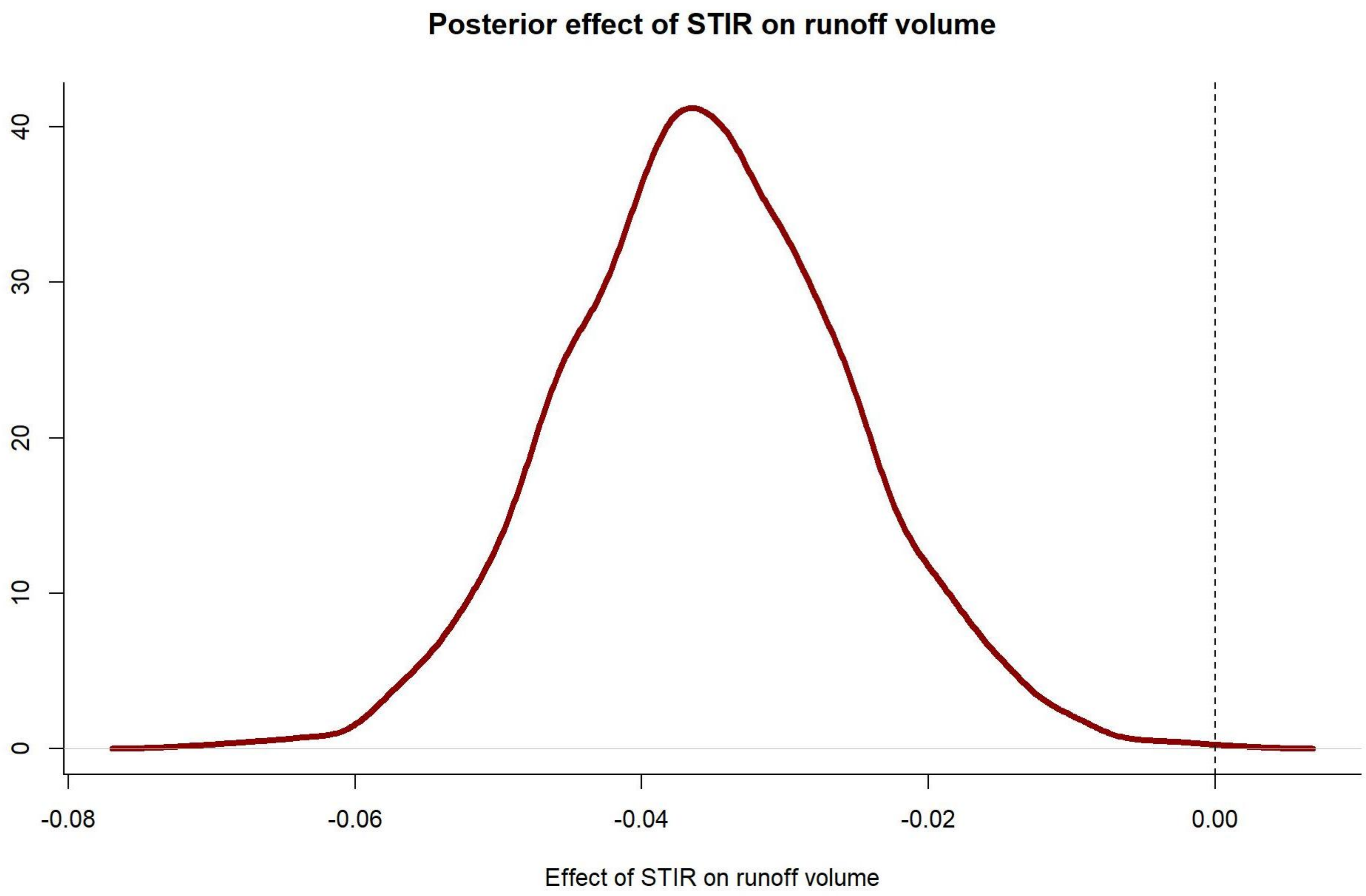
Signed, Sealed, Delivered (It’s STIR)

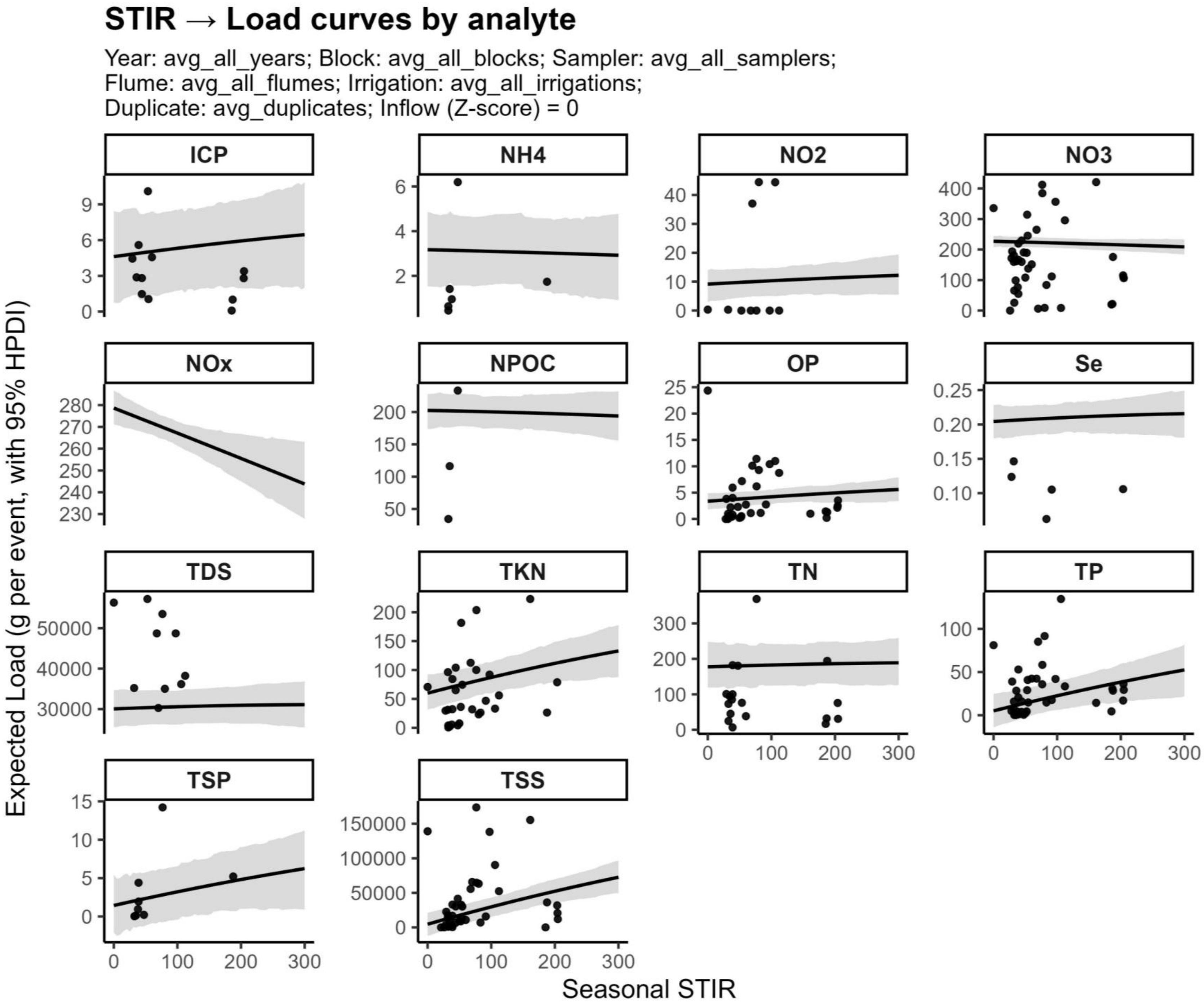


- STIR effects ranked from least to most by analyte type
- TDS, TP, TKN, Se, TSP, OP, and TSS were significantly affected
- Wider CIs are indicative of fewer data
- Results seem to follow theory that insoluble particulates are impacted most, minus OP

Pump Up the Volume

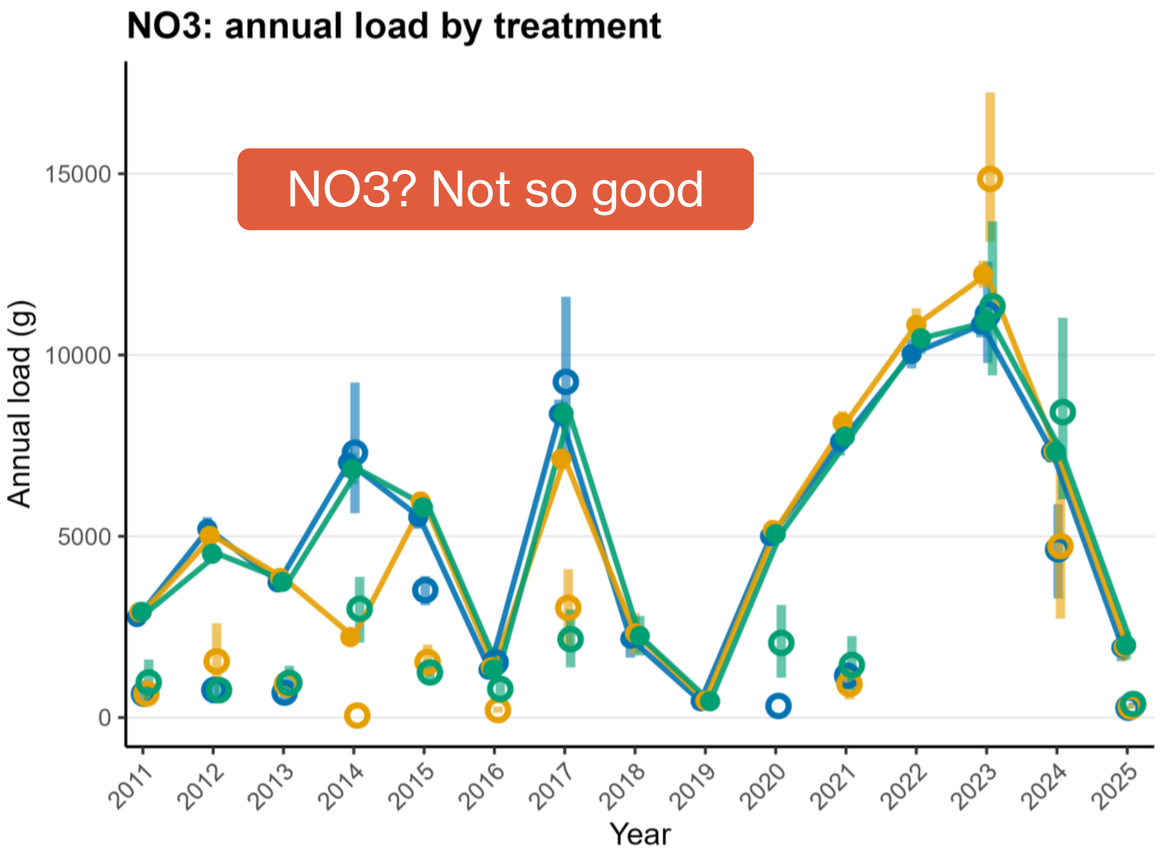
- STIR had a negative impact on volume
- This seems contrary to observations and theory. Need to look into this.
- Impact is small in magnitude; could be biased by recent MT/ST data

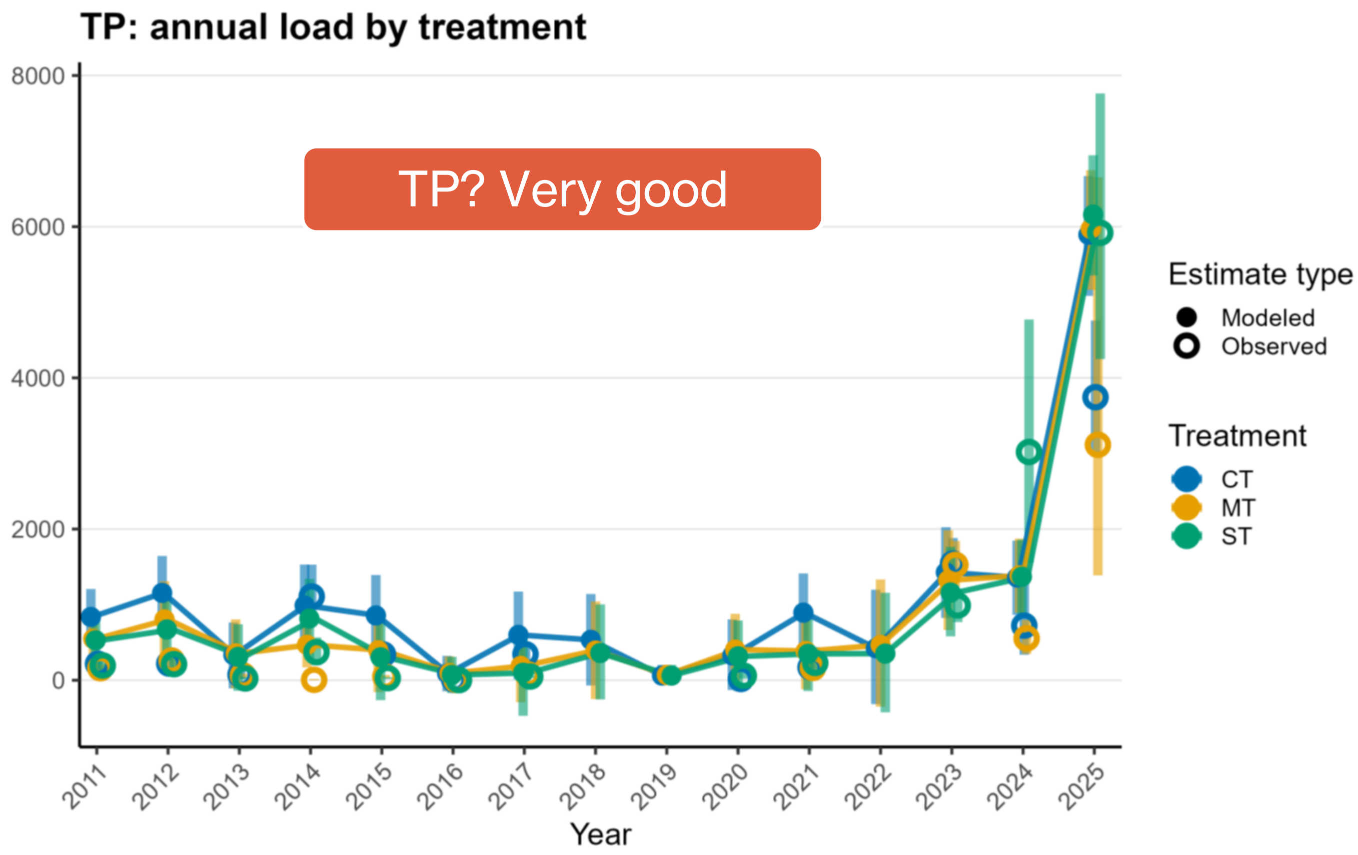




Load on the Water

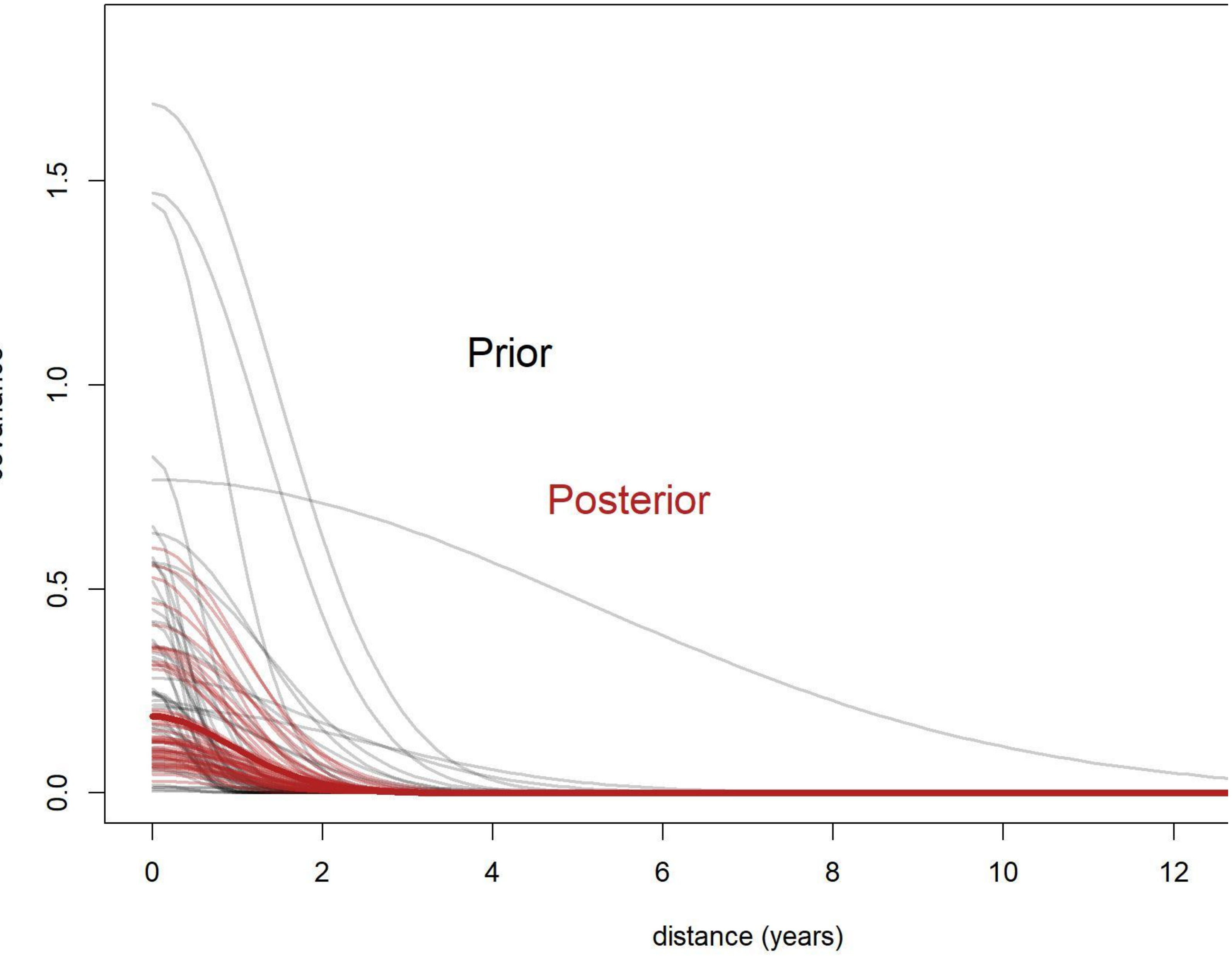
- Here are the results we want; STIR impacts on LOADS
- Fit looks poor b/c it's averaging over a lot of stuff
- Checks out with theory
  - TP, TSS, TSP, TKN most affected





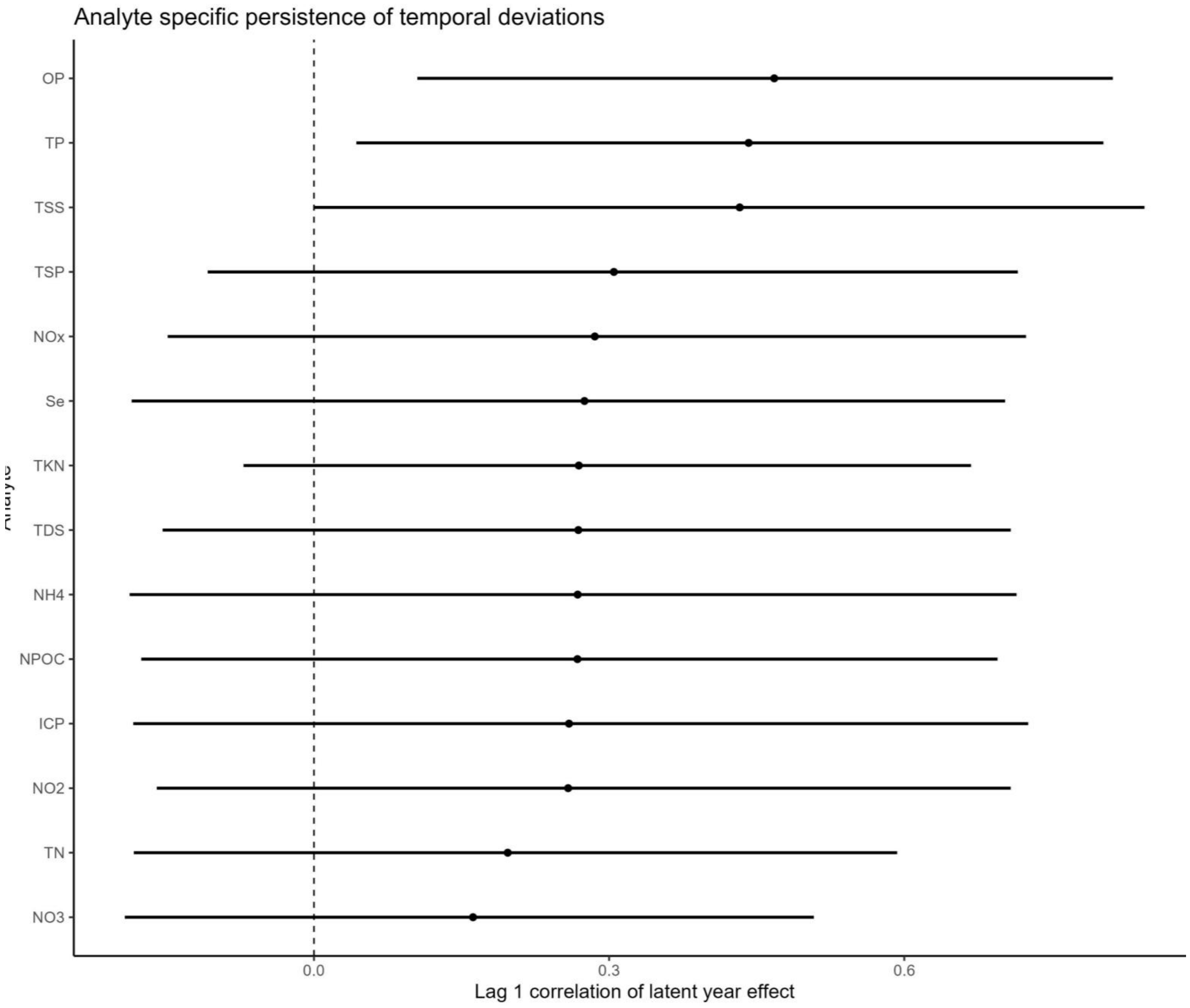
Bridge Over Troubled Data

- Even though the model doesn't utilize treatment (CT/ST/MT), I thought it would be useful to calculate loads on a trt-level
- Graph shows modeled and observed data with 95% CIs
- This was done for all analytes
- Model does well on variables that have stronger STIR relationships!



Don’t Stop Decaying

- This is the temporal kernel function from the MOGP
- Data did change the prior (good)
- Posterior shows that there is very little covariance after a single year over all analytes

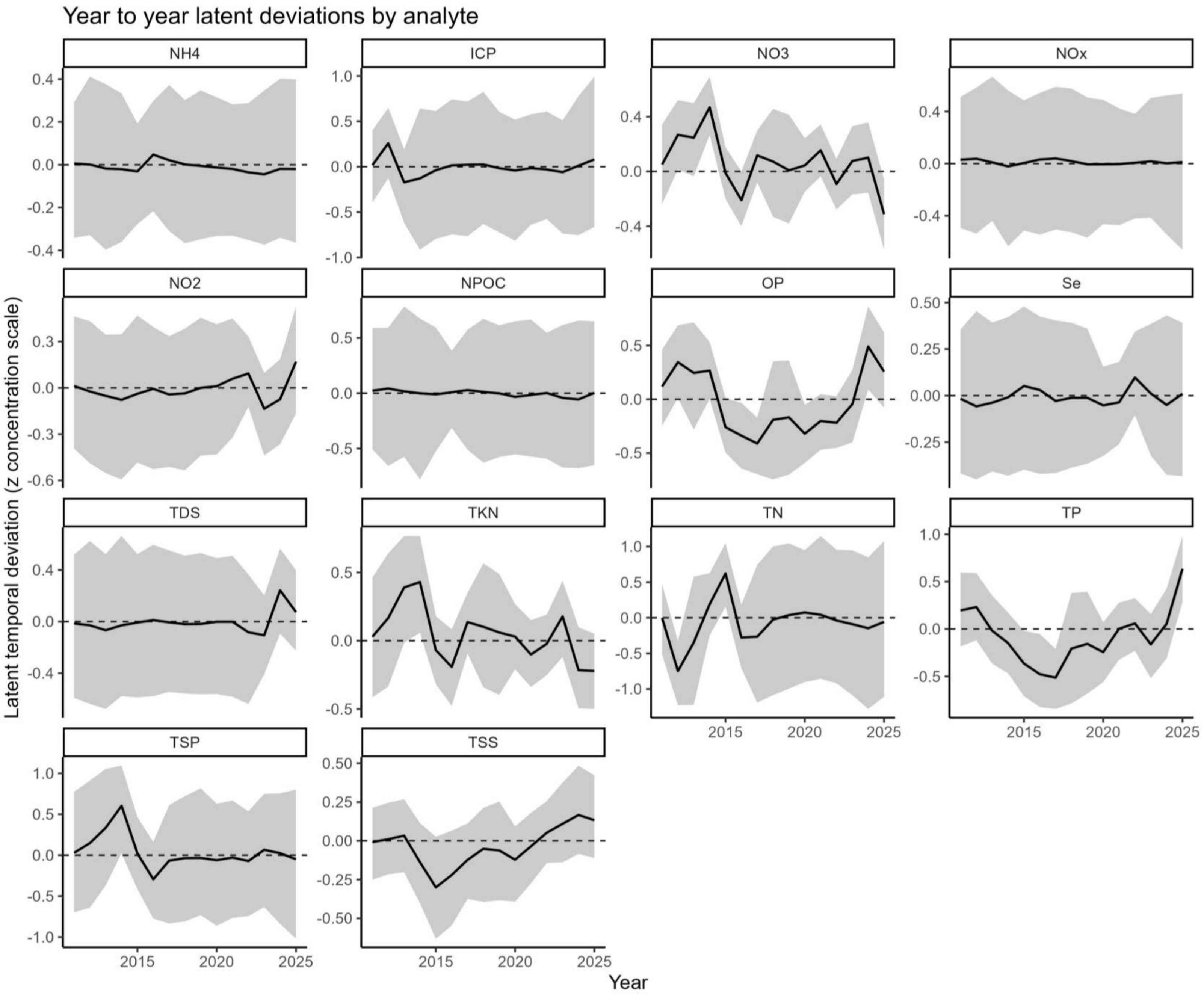


Time After Time (But Only for Some Analytes)

- AR(1) regression effect from MOGP
- So which analytes had temporal effects?
  - OP, TP, TSS <- clearly
  - TSP, TKN <- sort of
- This graph represent she ability of a previous year to predict this year's load

Yesterday (All My Deviations Seemed So Far Away)

- Visualize the actual **latent AR(1) trajectories for each analyte**
- These trajectories **show how much real-world year-to-year variability remains** *after accounting for tillage, sampler, block, and irrigation covariates*.
- **NO3, TKN:** A sequence of above-average years around mid-study followed by reversion towards baseline, consistent with the moderate ρ.
- **TP, TSS:** More pronounced multi-year deviations, which is biogeochemically plausible because P and sediment export are tightly linked to cumulative soil surface conditions and disturbance history.



Let It Be... Quantified (Summary)

- Certain analytes seem to be very impacted by tillage (STIR)
  - TP, TSS, OP, TKN primarily
- There are temporal effects that persist between years
  - For TP, TSS, OP, and TKN
- Key theme: Solubility!
- The Bayes model performed best to worst in the same order as STIR analyte effect
- Bayes also let us predict annual loads, even in years with no data, with honest uncertainty

You Can’t Always Get the Effect You Want (Limitations)

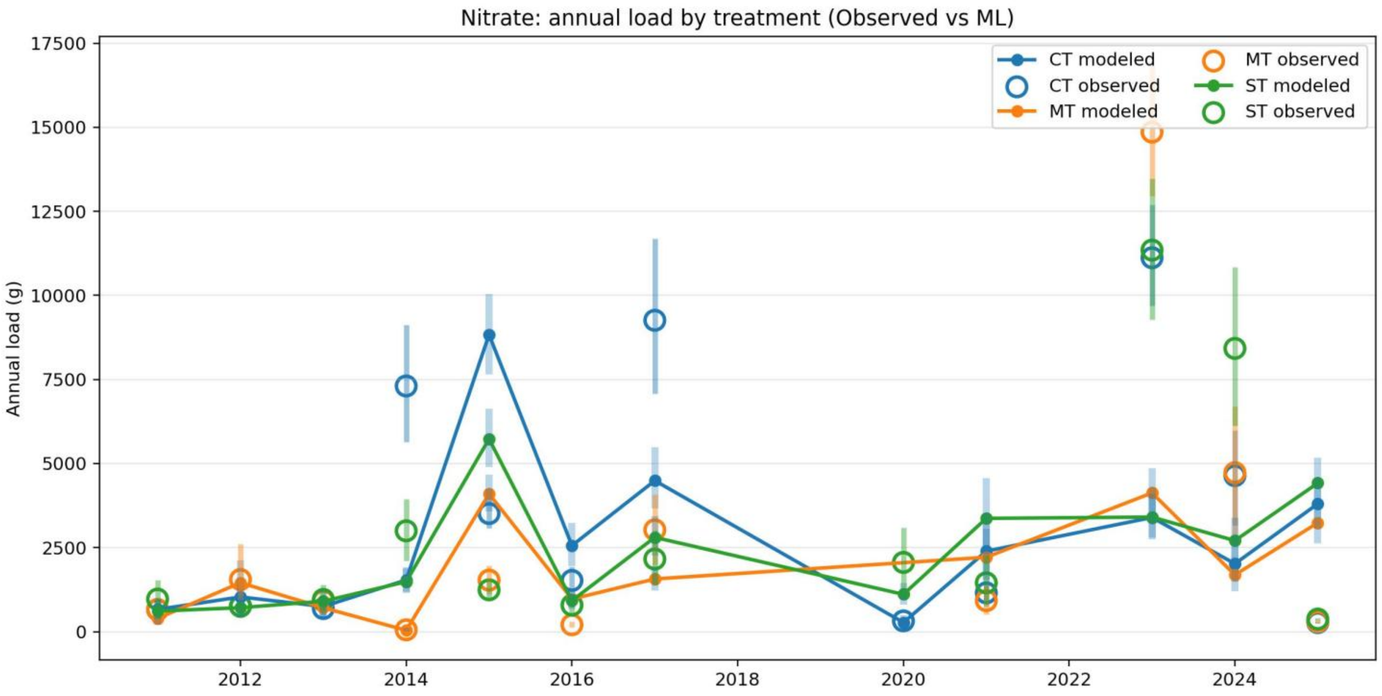
- The model is not designed to capture analytes dominated by soluble, subsurface, or legacy transport processes; nitrate (NO_3) performance illustrates this limitation clearly.
- STIR does not encode fertilizer timing, nitrogen application rate, crop uptake, or subsurface hydrology, all of which strongly influence nitrate dynamics.
- Annual load estimates for observed data reflect aggregation uncertainty and monitoring heterogeneity rather than true event-level completeness, particularly in early or sparsely sampled years.
- Treatment is not included as a causal predictor, so treatment-level comparisons should be interpreted descriptively, not as direct treatment effects.
- The framework assumes event-scale surface runoff as the primary export pathway; analytes governed by longer temporal integration or delayed pathways require additional structure to be modeled mechanistically.

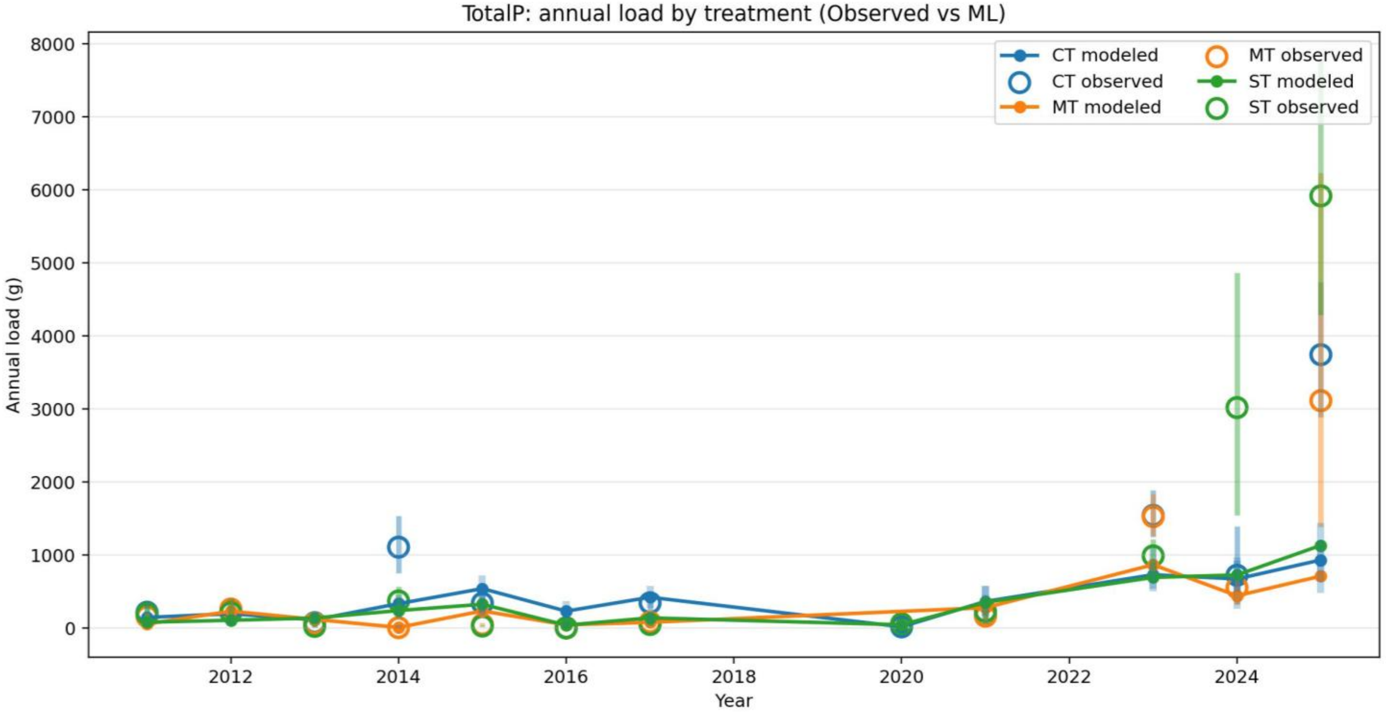
What's Left?

- Investigate STIR → Volume relationship
- Decide on which analytes to present, if not all of them
- Forecast 2026 and compare?
- Try machine learning?
- Potential papers currently
  - **STIR impacts on long term WQ and legacy effects**
  - "Completing" a dataset with Bayesian techniques
    - Will include handling NDs better, right now they're zero
  - Validate model using 2026 data?
    - Not really how Bayes and ML models work but okay
  - Compare a machine learning model (next slide sneak peak)

Sneak Peek

Using CatBoost machine learning model to predict concentrations and volumes to derive loads





CatBoost

The End

...and in the end, the model you take is equal to the assumptions you make.