Structure and development of CalCAST

Isabella Bertani, Gopal Bhatt, Gary Shenk, Lewis Linker

Modeling Workgroup Quarterly Review
04/05/2022

What is CalCAST?

- Relatively parsimonious, spatially explicit, largely data-driven watershed modeling tool calibrated in a statistical framework
- Currently represents > 80,000 National Hydrography Dataset Plus (NHDPlus) catchments within the Bay watershed and leverages data from > 400 USGS monitoring stations for calibration
- Currently time-averaged (but may be extended to predict at the annual time step in the future)
- Currently predicts long-term average streamflow at NHDPlus catchments (but we plan to extend it to predict nutrient and sediment loads)

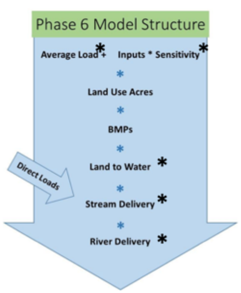
Why CalCAST?

- Primarily used as spatial calibration tool
- Main purpose: probabilistically test hypotheses on factors related to spatial variation in contaminant loads and quantify parameters that describe such relationships
- Spatial parameters estimated by CalCAST will inform the dynamic model simulation
- Incorporate data-driven line of evidence into modeling approach

We built the first CalCAST prototype last year

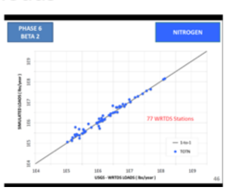
CalCAST Hydrology Model Development

- Average annual streamflow *(Q)* is the difference of Rainfall and Actual Evapotranspiration *(AET)*, where *AET* can be estimated from Potential Evapotranspiration *(PET)* and/or other watershed properties.



$$Q = \sum_{\substack{upstream \\ geography}} Precipitation - PET \times f_{LU}$$

Calibration of meta-parameters to spatial loads



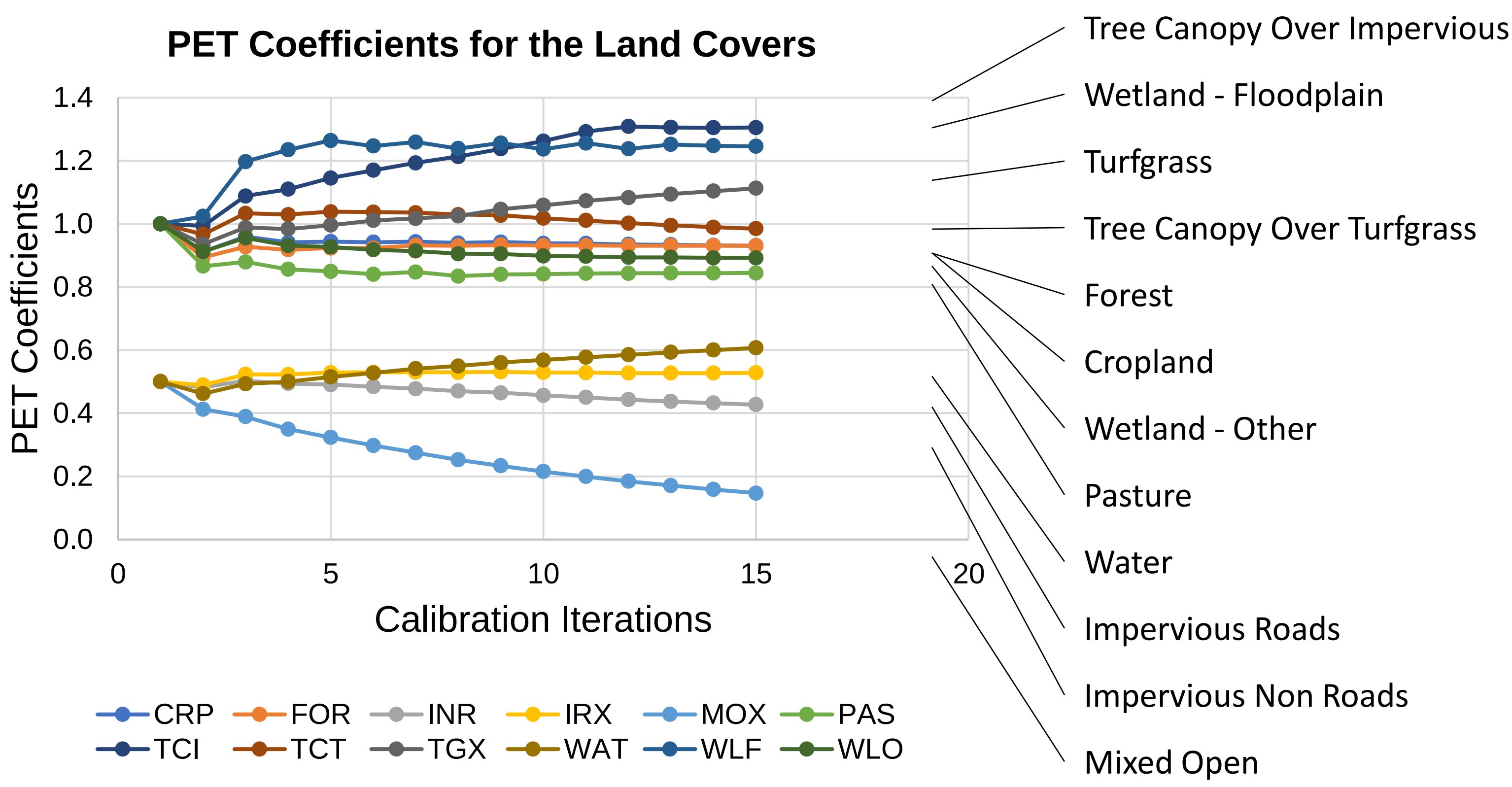
Currently 12 land cover classes, i.e., 12 PET coefficients to be estimated

Results (CalCAST Hydrology – an initial, simplified prototype)

From previous Quarterlies

$$Q = \sum_{\substack{upstream \\ geography}} Precipitation - PET \times f_{LU}$$

Calibration Iterations



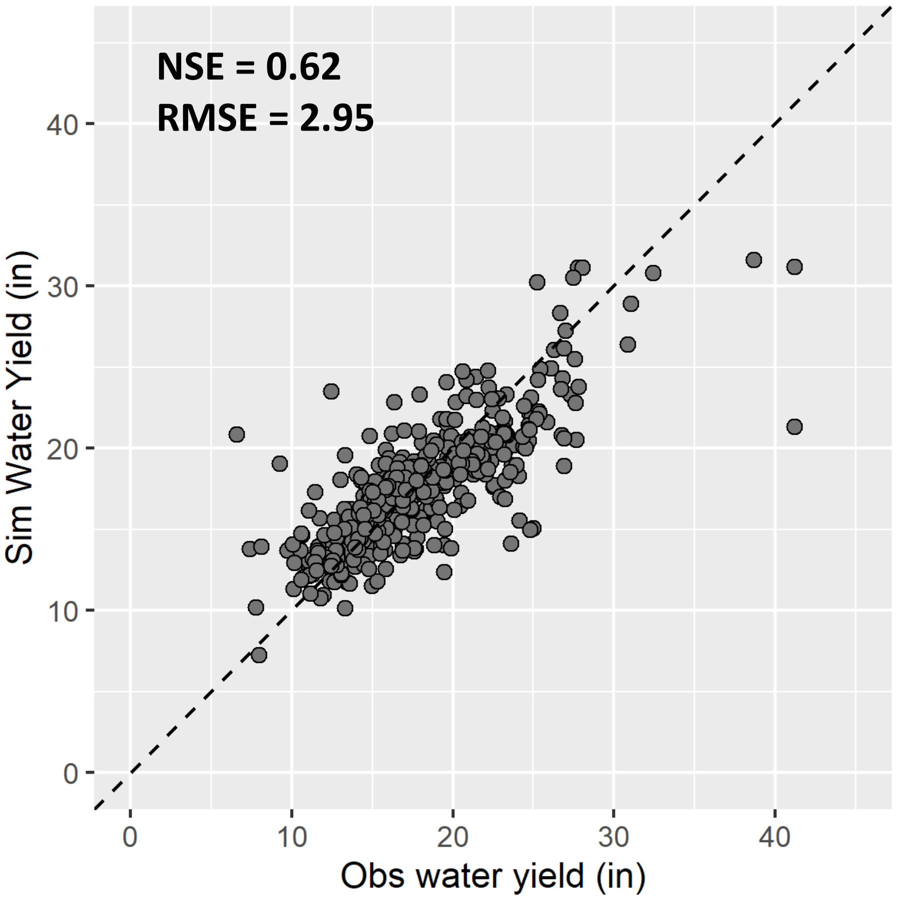
We used the prototype to test candidate predictors of streamflow

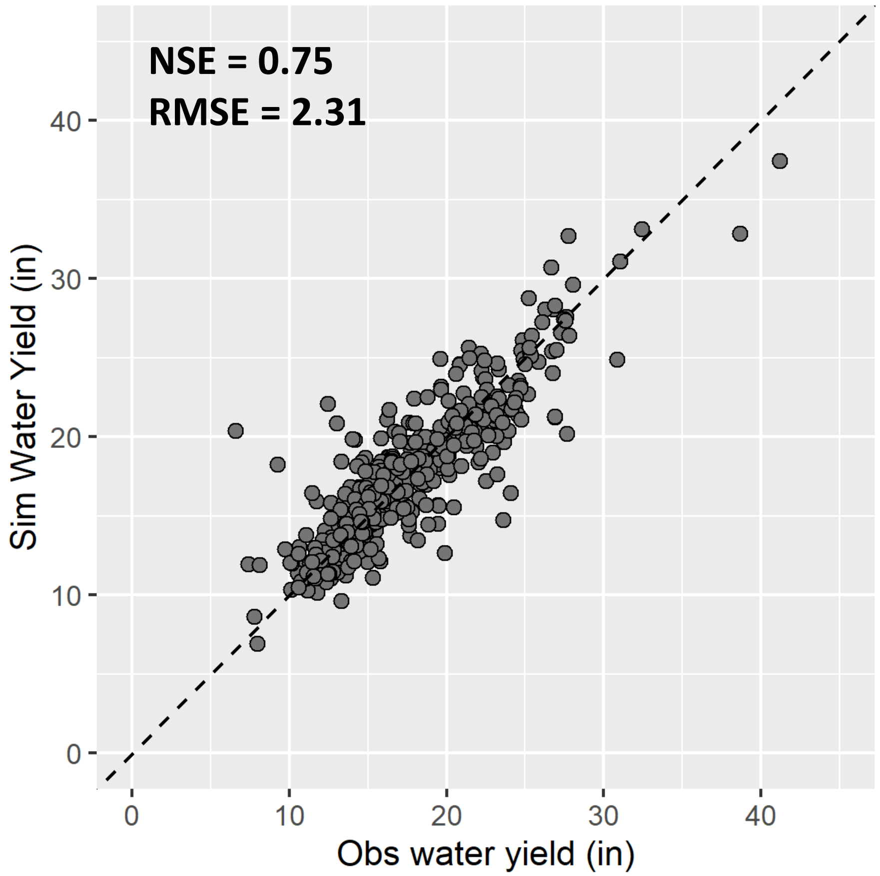
From previous Quarterlies

Candidate predictors of streamflow

$$Q = \sum_{\substack{upstream \\ geography}} Precipitation - PET \times f_{LU}$$

$$Q = \sum_{\substack{upstream \\ geography}} Precipitation - PET \times f_{LU} \times (T, GWRECH, AML, TWI)$$



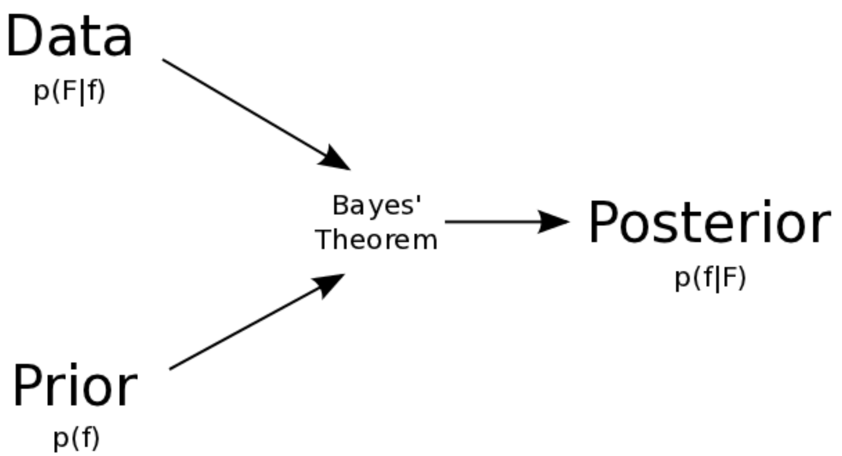


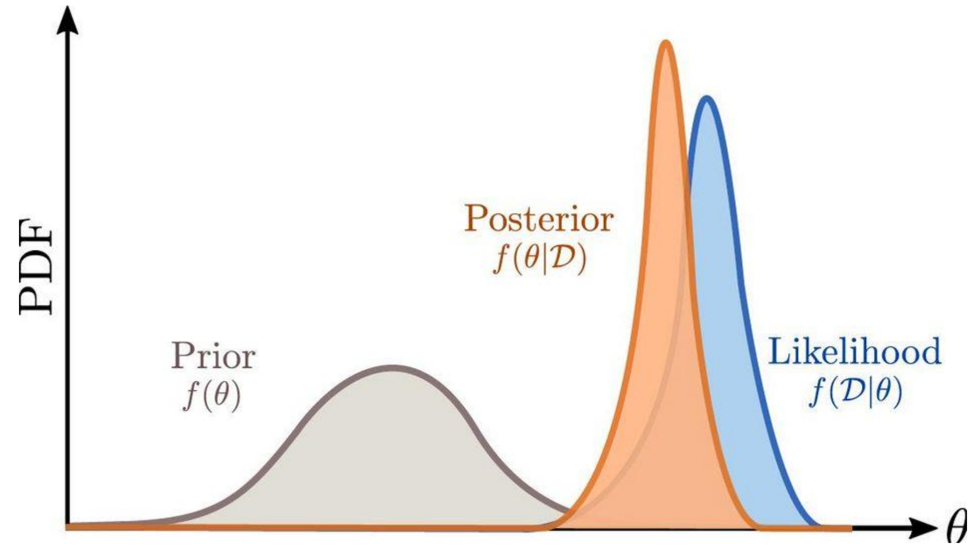
Plan for this year

- Move prototype to a Bayesian calibration framework
- Extend initial prototype to also predict sediment and nutrient loads
- Focus on getting the code structure up and running

Why Bayesian inference?

All unknown quantities are treated as random variables





Advantages of Bayesian inference

- Leverage results of previous watershed modeling and monitoring studies through incorporation in the form of prior probability distributions, thereby effectively combining multiple lines of evidence within a formal statistical framework
- Rigorously test hypotheses about parameter values and predictor relative importance in a probabilistic framework
- Model spatial and temporal variability at different organizational levels (e.g., monitoring stations nested within sub-watersheds nested within larger regions)
- *Quantification of uncertainty*
- *Incorporate uncertainty in model inputs (e.g., WRTDS estimates)*

Progress so far

Model formulation

Total Flow:

$$\sum_{t} Q_{g,t} = \sum_{t}\sum_{c}\left(P_{c,t} - \sum_{l}\left(PET_{l,c,t} \times LUP_{l} \cancel{\times \prod_{f}\left(a_{f}^{*} \times V_{c,f}\right)}\right) - W_{c} + I_{c}\right)$$

- $Q_{g,t}$: Flow at gauge g in year t
- $P_{c,t}$: Precipitation in catchment c in year t
- $PET_{l,c,t}$: PET for land use l in catchment c in year t
- LUP_l: Land use-specific PET parameter
- a_f: Coefficient for factor f
- $V_{c,f}$: Value of factor f in catchment c
- W_c: Withdrawals in catchment c in year t
- I_c: Direct loads in catchment c in year t

* Ignoring watershed properties for now for simplicity in code building/testing and to be able to compare results to original CalCAST prototype

Model formulation

Stormflow/Total Flow:

$$\mathrm{SG_g} = \frac{1}{A_g^{a}} \times \frac{\sum_l \sum_c \sum_t \left(\mathrm{SL}_{l,c,t} \times R_{l,c,t}\right)}{\sum_l \sum_c \sum_t R_{l,c,t}}$$

- Fraction of stormflow at gauge g
- Drainage area of gauge g
- Drainage area coefficient
- Stormflow fraction for land use l, catchment c, year t
- Runoff for land use l, catchment c, year t

$$R_{l,c,t} = \sum_t \sum_c \left(P_{c,t} - \sum_l \left(PET_{l,c,t} \times LUP_l \cancel{\times \prod_f \left(a_f \times V_{c,f}\right)}\right)\right)$$

$$\mathrm{SL_{l,c,t}} = \mathrm{SLG}_l \cancel{\times \prod_f \left(b_f \times VS_{c,f}\right)}$$

- Stormflow fraction coefficient for land use l
- Coefficient for factor f
- Value of factor f in catchment c

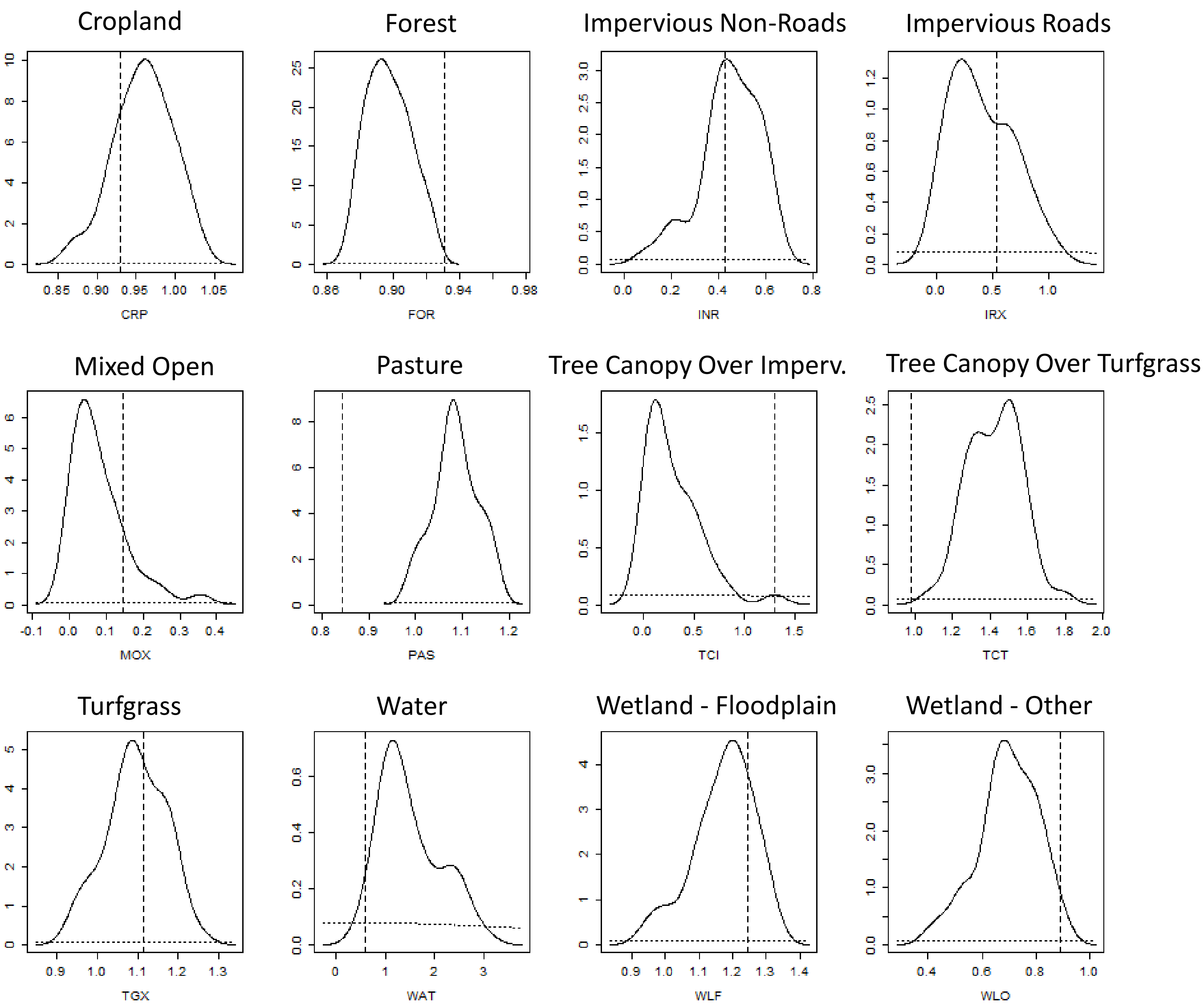
Total Flow results

At this point, we are mostly looking to make sure that the Bayesian code works as expected

To do so, we compare results of the Bayesian model to those of the initial CalCAST prototype

Land-use specific PET parameters for Total Flow

$$\sum_{t} Q_{g,t} = \sum_{t}\sum_{c}\left(P_{c,t} - \sum_{l}\left(PET_{l,c,t} \times \boldsymbol{LUP_l}\right) - W_c + I_c\right)$$



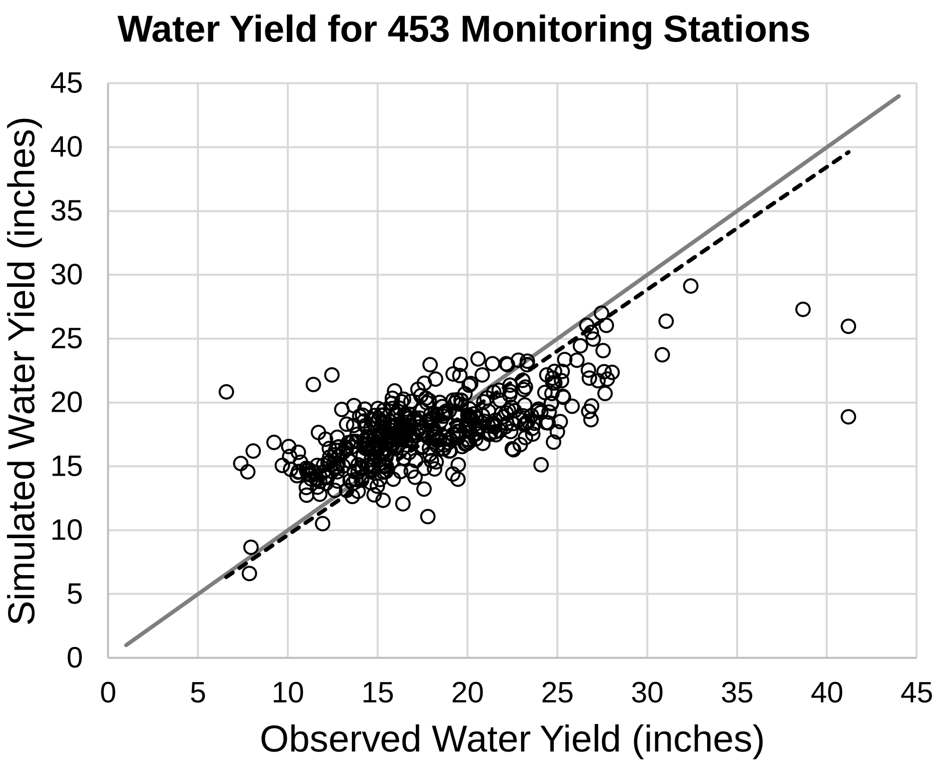
Solid lines: posterior probability distributions estimating land-use specific PET parameters ($\boldsymbol{LUP_l}$)

Dotted lines: prior probability distributions

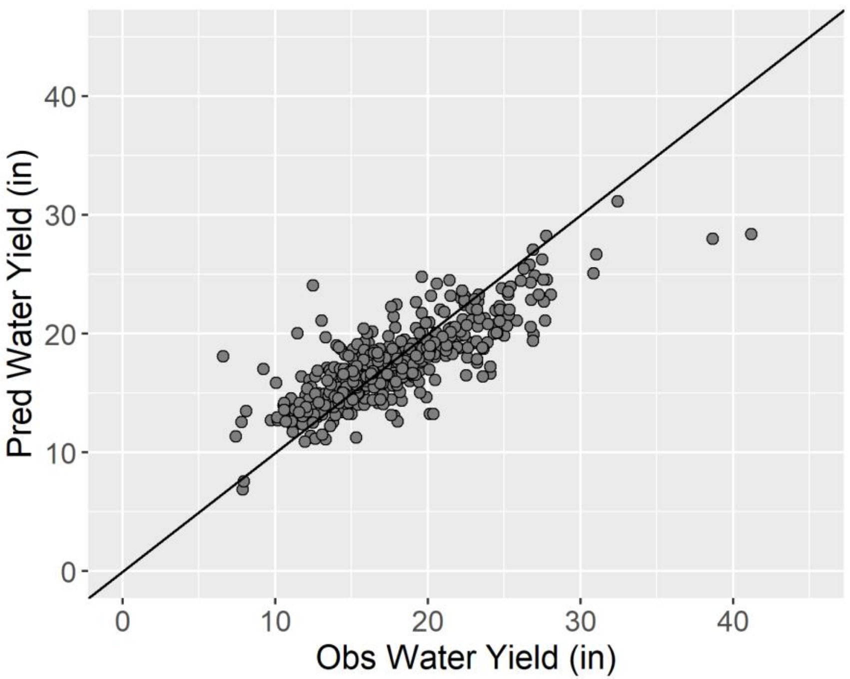
Dashed vertical lines: parameter estimates obtained by original CalCAST prototype

Observed vs. predicted Total Flow

Original CalCAST prototype



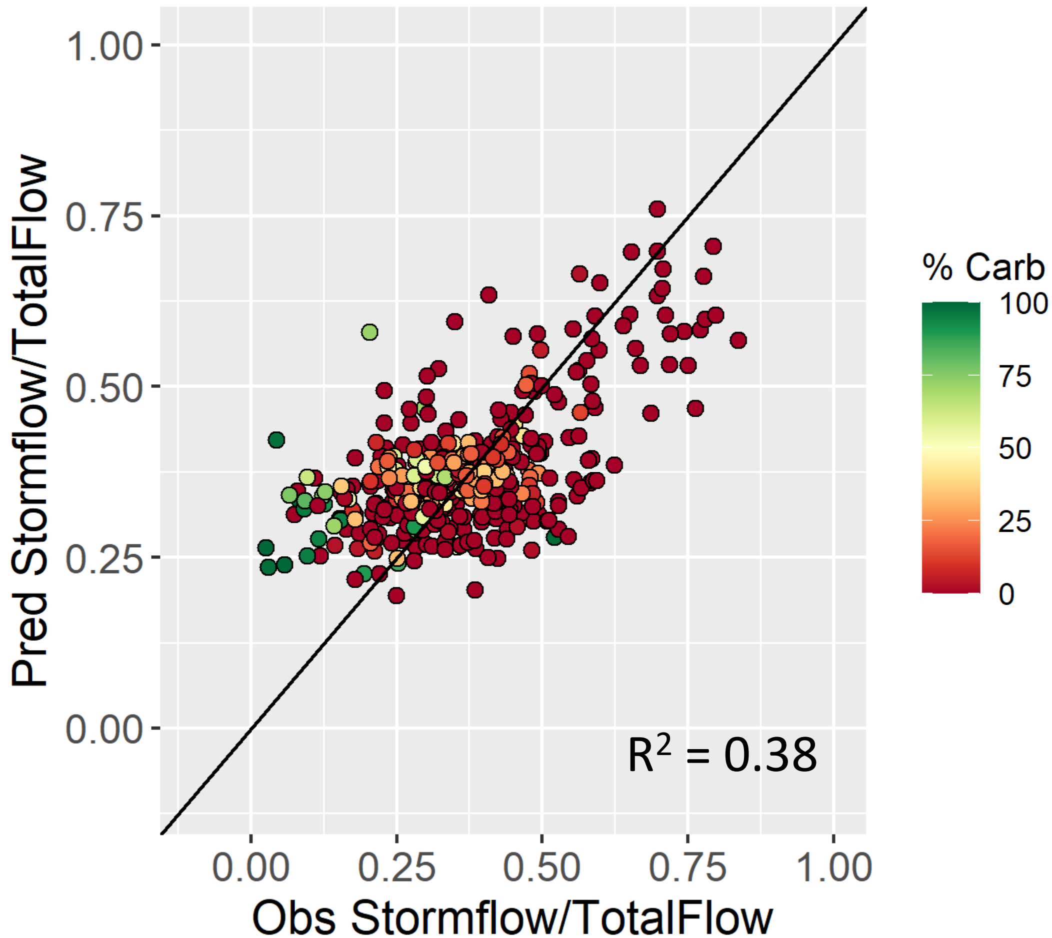
Bayesian CalCAST



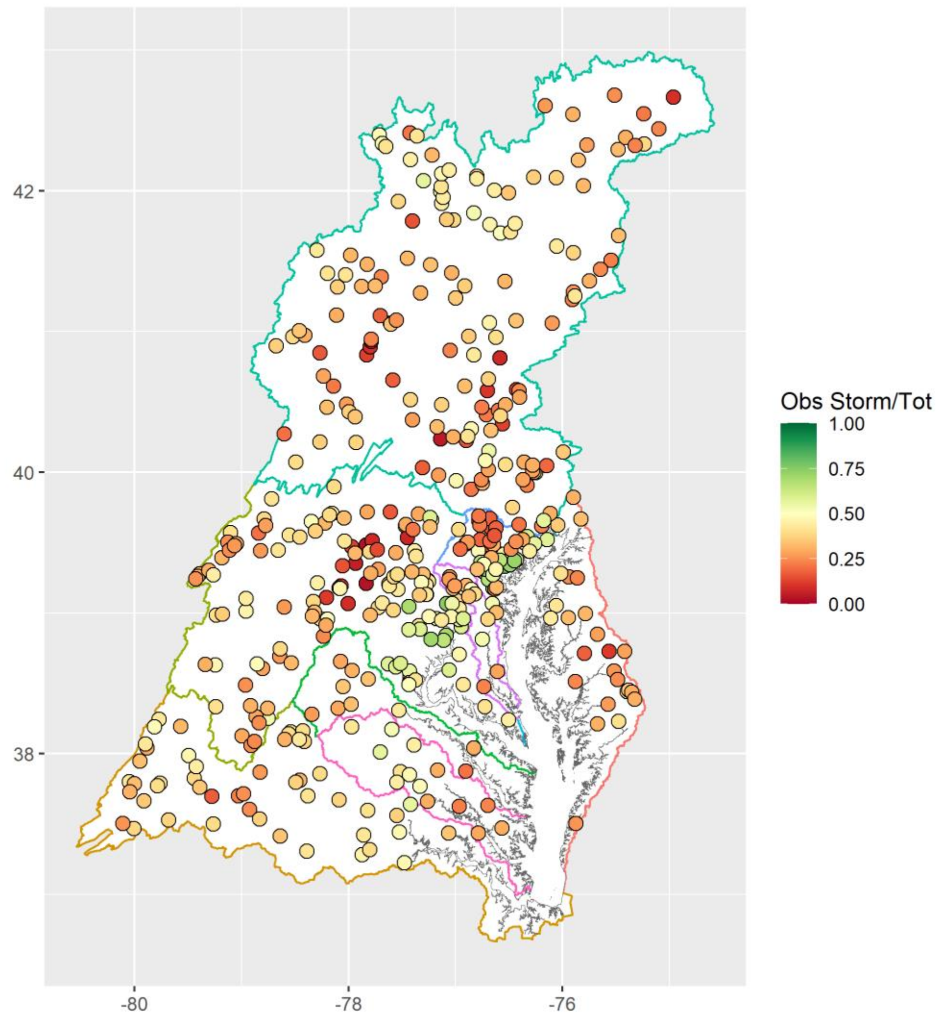
The Bayesian model roughly reproduces the results obtained with the initial prototype, suggesting that the code works as expected

Stormflow/Total Flow results

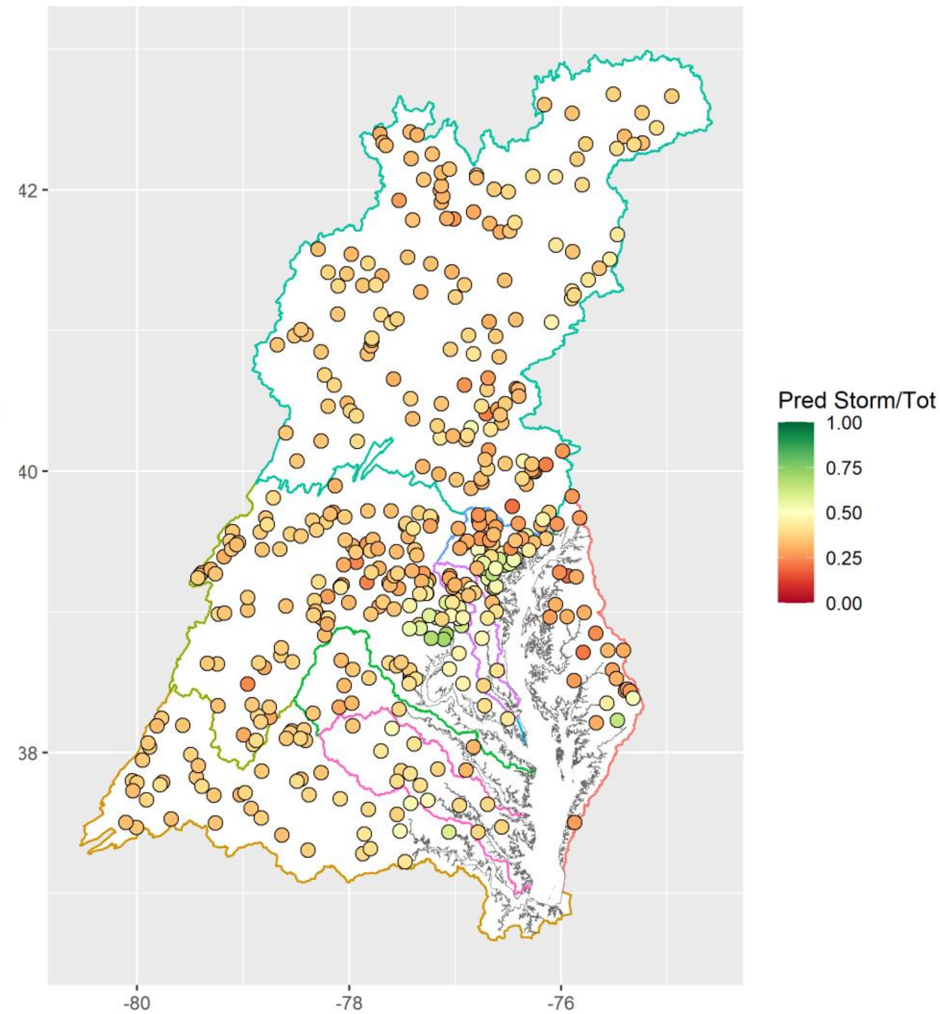
Observed vs. predicted Stormflow/Total Flow



Obs Stormflow/Total Flow



Pred Stormflow/Total Flow

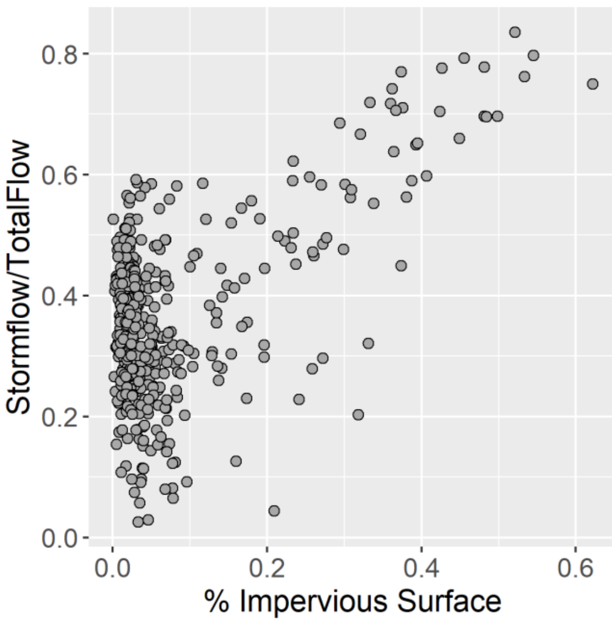


Can we improve Stormflow/Total Flow predictions?

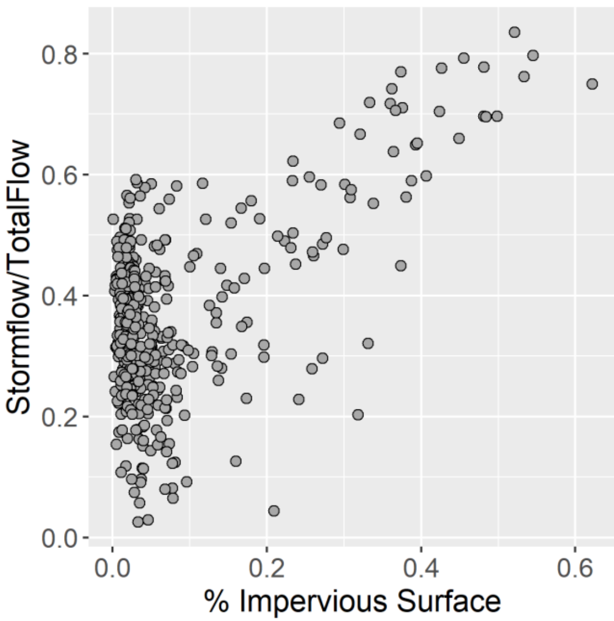
Formal analysis of watershed properties planned for next year

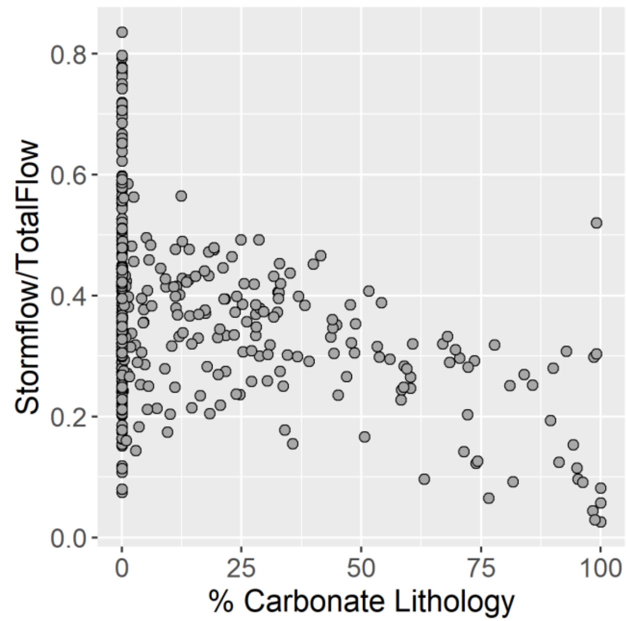
However, we did some preliminary testing of watershed properties as candidate predictors of Stormflow/Total Flow

Stormflow/Total Flow vs. watershed properties

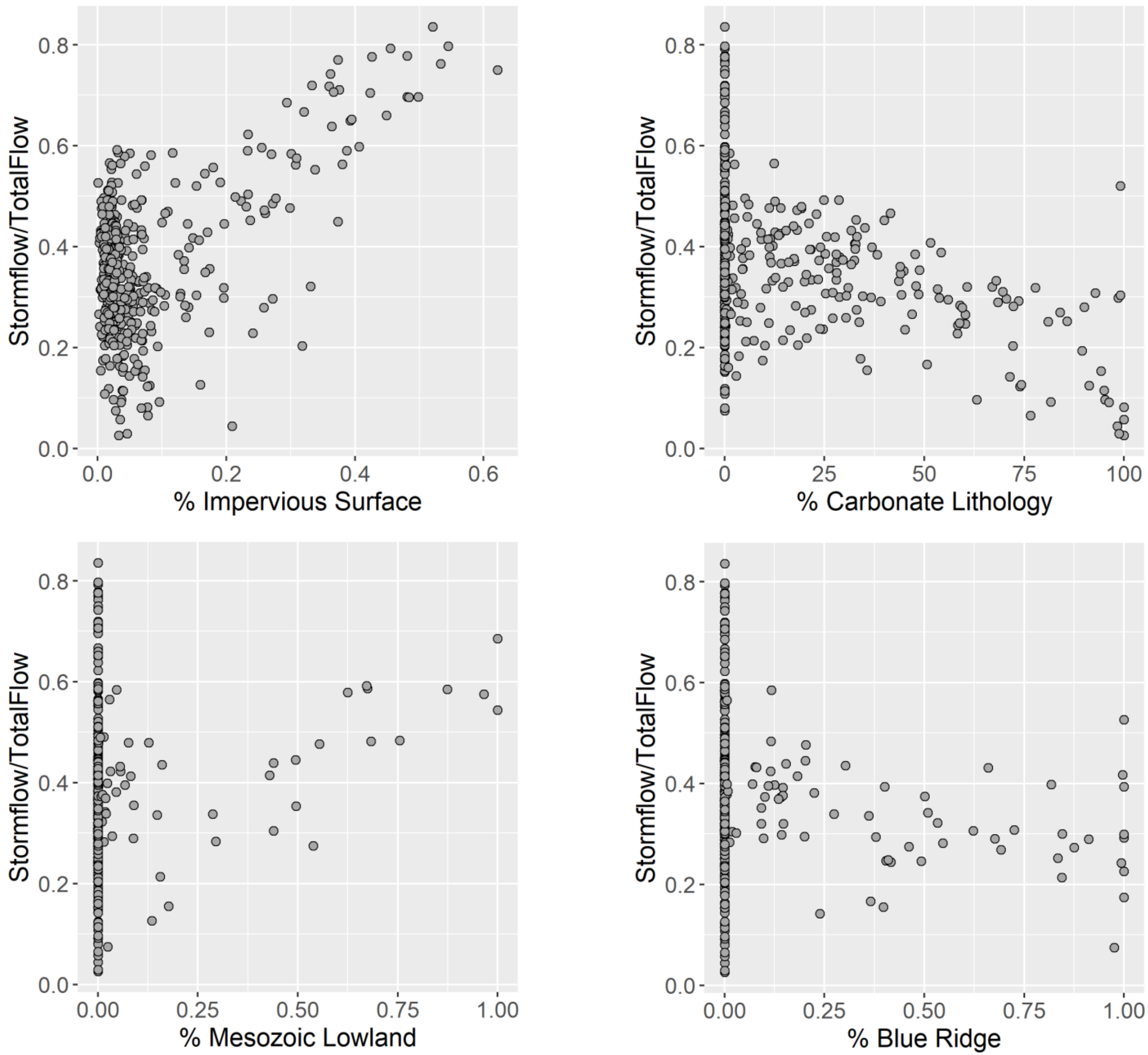


Stormflow/Total Flow vs. watershed properties





Stormflow/Total Flow vs. watershed properties



Model formulation

Stormflow/Total Flow:

$$\mathrm{SG_g} = \frac{1}{A_g^{a}} \times \frac{\sum_{\mathrm{l}} \sum_{c} \sum_{\mathrm{t}} \left(\mathrm{SL}_{l,c,t} \times R_{l,c,t} \right)}{\sum_{l} \sum_{c} \sum_{\mathrm{t}} R_{l,c,t}}$$

- Fraction of stormflow at gauge g
- Drainage area of gauge g
- Drainage area coefficient
- Stormflow fraction for land use l, catchment c, year t
- Runoff for land use l, catchment c, year t

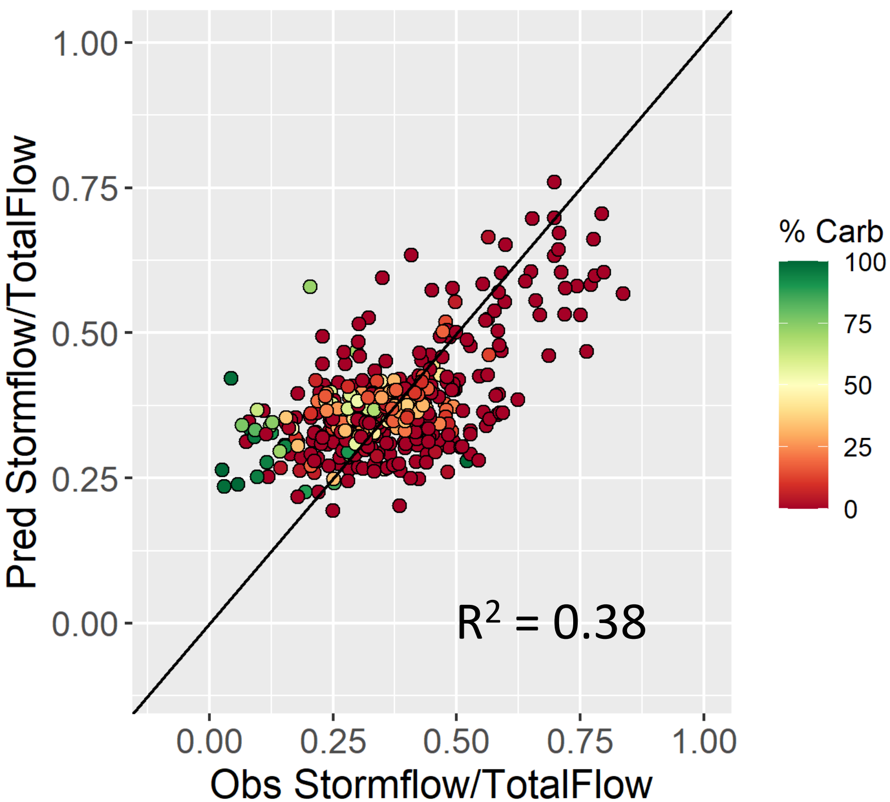
$$R_{l,c,t} = \sum_{\mathrm{t}} \sum_{c} \left(P_{c,t} - \sum_{l} \left(PET_{l,c,t} \times LUP_l \cancel{\times \prod_{f} \left(a_f \times V_{c,f} \right)} \right) \right)$$

$$\mathrm{SL_{l,c,t}} = \mathrm{SLG}_l \times \prod_{f} \left(\boldsymbol{b_f} \times \boldsymbol{VS_{c,f}} \right)$$

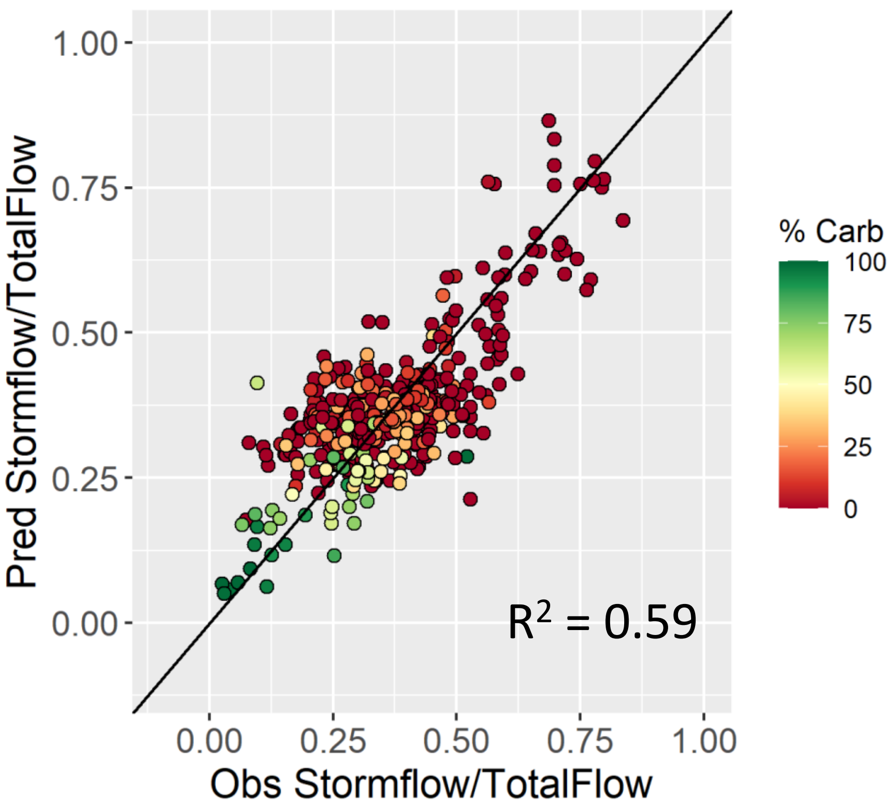
$VS_{c,f}$ = % Carbonate Lithology, %, Mesozoic Lowland, % Blue Ridge, % Impervious Surface

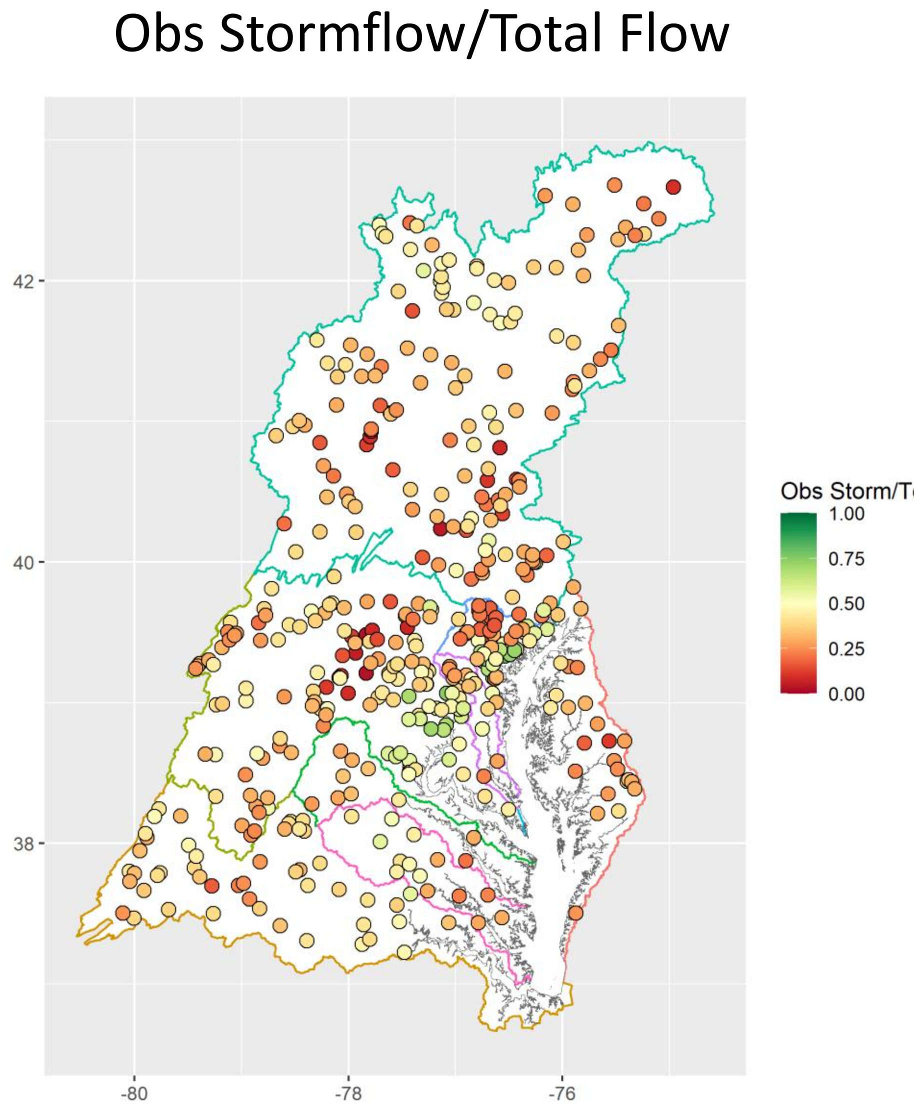
Observed vs. Predicted Stormflow/Total Flow

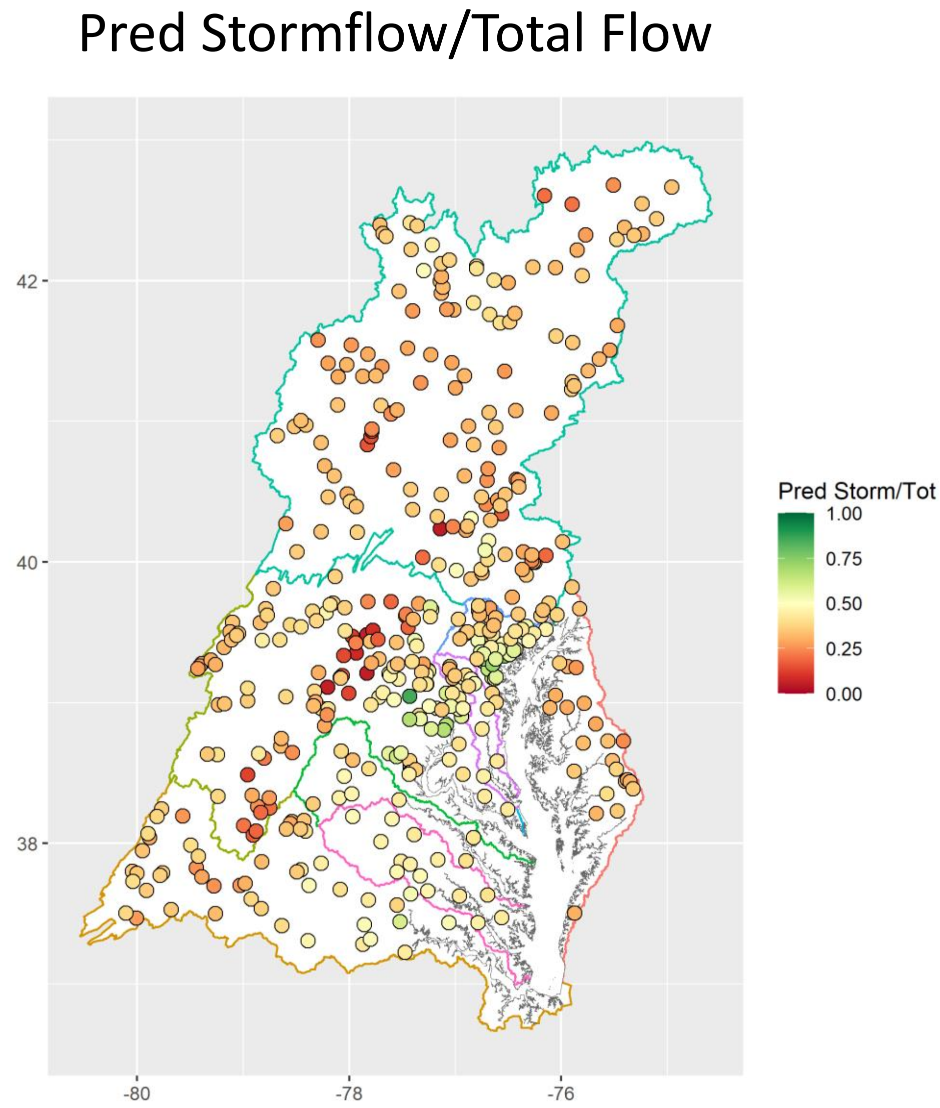
Without watershed properties



With watershed properties




Obs Stormflow/Total Flow
42
40
38
-80
-78
-76
Obs Storm/T
1.00
0.75
0.50
0.25
0.00


Pred Stormflow/Total Flow
42
40
38
-80
-78
-76
Pred Storm/Tot
1.00
0.75
0.50
0.25
0.00

Next steps for this year

- Extend the model to also predict sediment and nutrient loads
- Focus on getting the code structure up and running

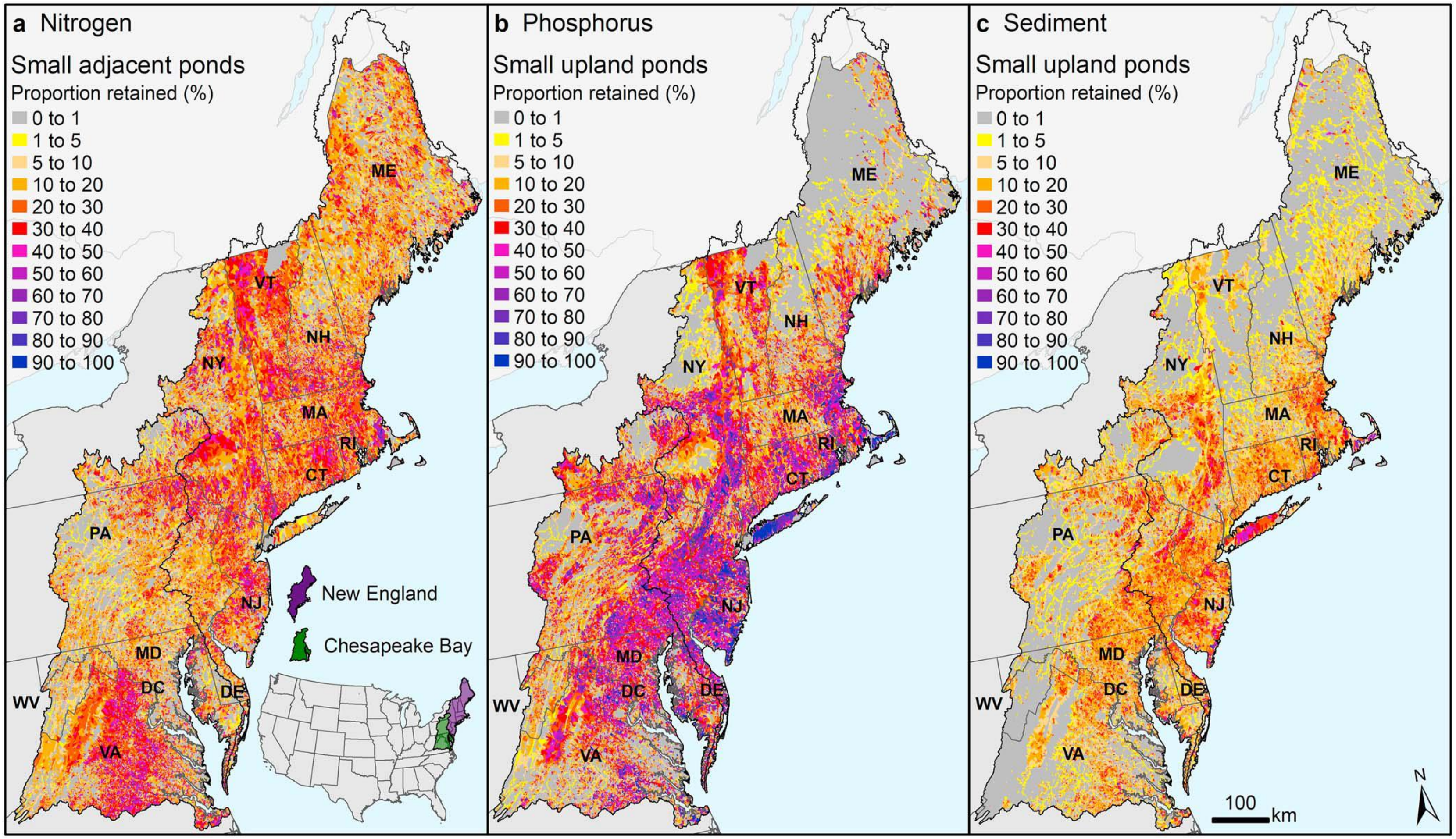
Next steps for this year

- Extend the model to also predict sediment and nutrient loads
- Focus on getting the code structure up and running

Plan for next year

- Use CalCAST as primary calibration tool
- Investigate spatial drivers of loads and estimate parameters that best match observations
- Work closely with Land Use Team to leverage their upcoming data products

Example of how we could leverage datasets from Land Use Team in CalCAST



Schmadel et al., 2019

Example of how we could leverage datasets from Land Use Team in CalCAST

- Small ponds were found to significantly impact nutrient and sediment budgets in the Northeast US

- Labeeb Ahmed and Peter Claggett have presented on the ability to map small ponds as part of the upcoming high-res land use data

- Test inclusion of small pond density (or other relevant pond characteristics) as predictors within CalCAST?