Technical Working Group: Meeting 6

April 4, 2023

Agenda: 4/4/23 Technical Working Group

- CalTRACK 2.1: Delivered Fuels Discussion
- CalTRACK 2.1 Daily Model

CalTRACK 2.1: Delivered Fuels

Initial Discussion

LF ENERGY

Items for OpenEE Meter WG Discussion 4/4/23

Why is CalTrack/Open EE Important to IRA Measured Savings?

- **It's the Law**: The IRA bill explicitly requires states to "use open-source advanced measurement and verification software… for purposes of measured performance home rebates."

- **Maximize Accessibility:** For states that offer measured savings programs, we want to ensure that as many homes as possible can be included. i.e., we want to avoid unnecessary "disqualifications"

Challenges in current methods that may limit eligibility

- Multiple items in *Section 2.2 - Data Constraints* exclude end uses with inconsistent meter readings (e.g., delivered fuels, certain gas utilities, etc.)
  - Estimated billing period requirements excludes delivered fuel customers due to gaps in deliveries/bills (2.2.3.1)
  - Off-Cycle reads < 25 days are required to be dropped. This will disqualify homes with smaller tanks that may get more frequent deliveries (2.2.3.4)
  - Billing periods periods spanning more than 35 days should be dropped from analysis. Often several months (e.g., 3 - 6 months is typical) between fill-ups and bills (2.2.3.5)
- Net metering exception language in 2.2.6 can be clarified to include sub-metered data, preventing sites from being dropped

Data Snapshot:

- n = 3330 (oil & propane homes)
- Avg length of billing history: **803 days**
- Avg duration between all deliveries: **77 days**
- Avg longest gap between deliveries per customer: **185 days**

Source: Sealed Inc.

Discussion & Next Steps

- Gather stakeholder input between now and May meeting
- Send proposed redline changes to group for consideration
- Discuss in May meeting

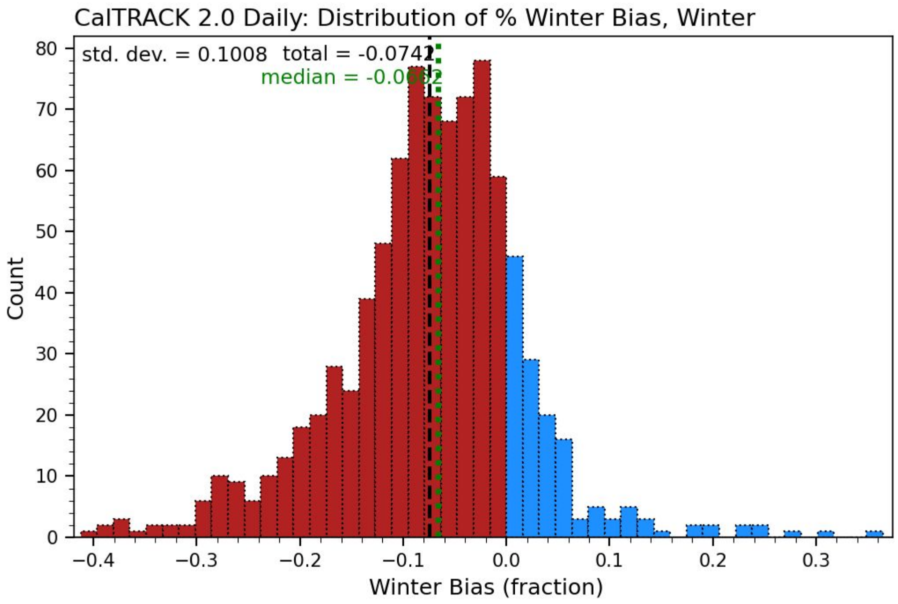
CalTRACK 2.1: Preliminary Results

LF ENERGY

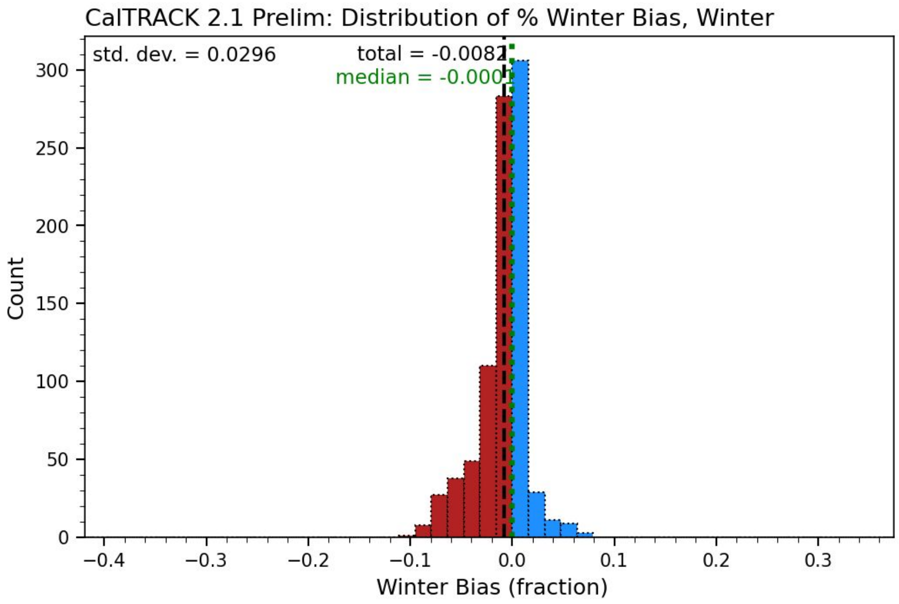
CalTRACK Winter Bias (Res Gas)

CalTRACK 2.0: -7.4%

CalTRACK 2.1 (Preliminary): -0.8%



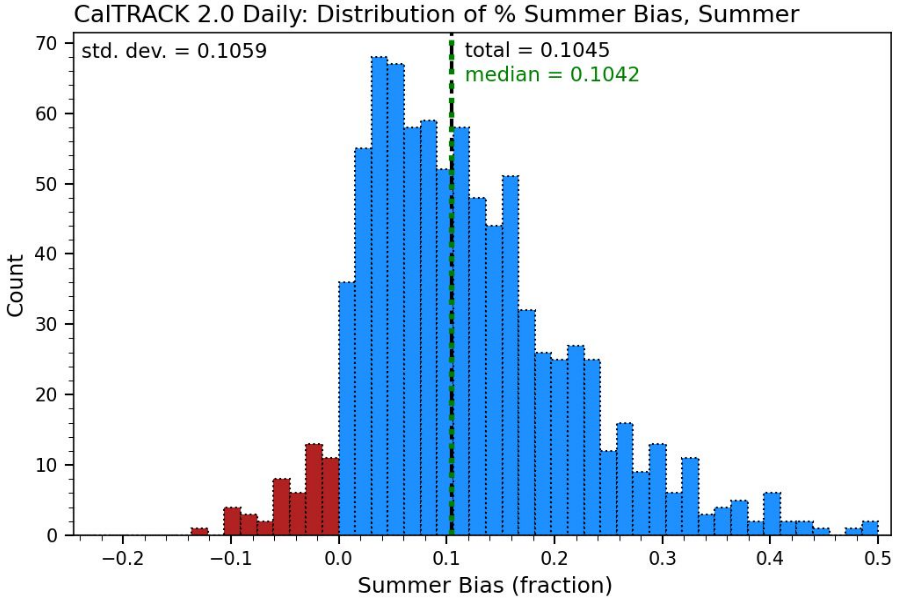
SSE = 329.8



SSE = 306.3

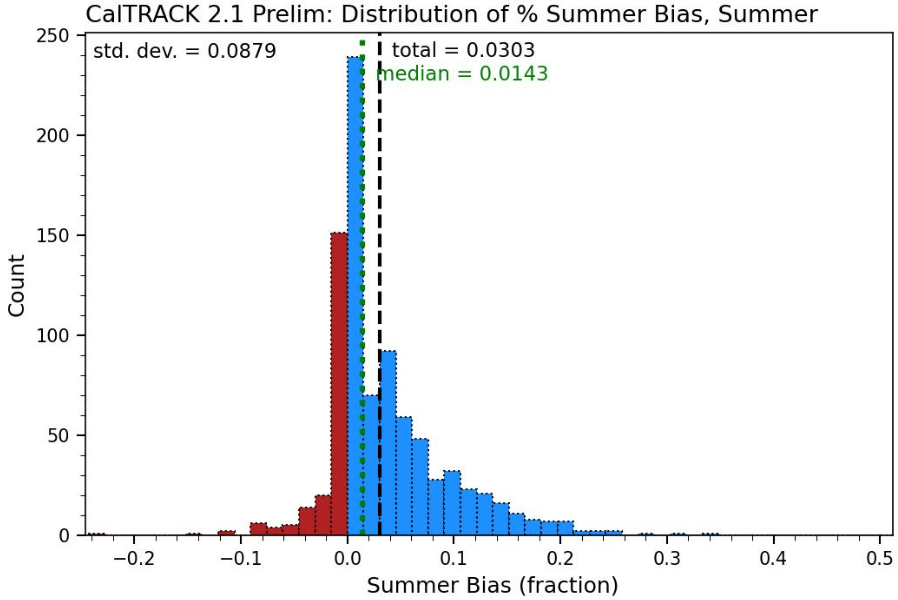
CalTRACK Summer Bias (Res Gas)

CalTRACK 2.0: 10.5%



SSE = 104.4

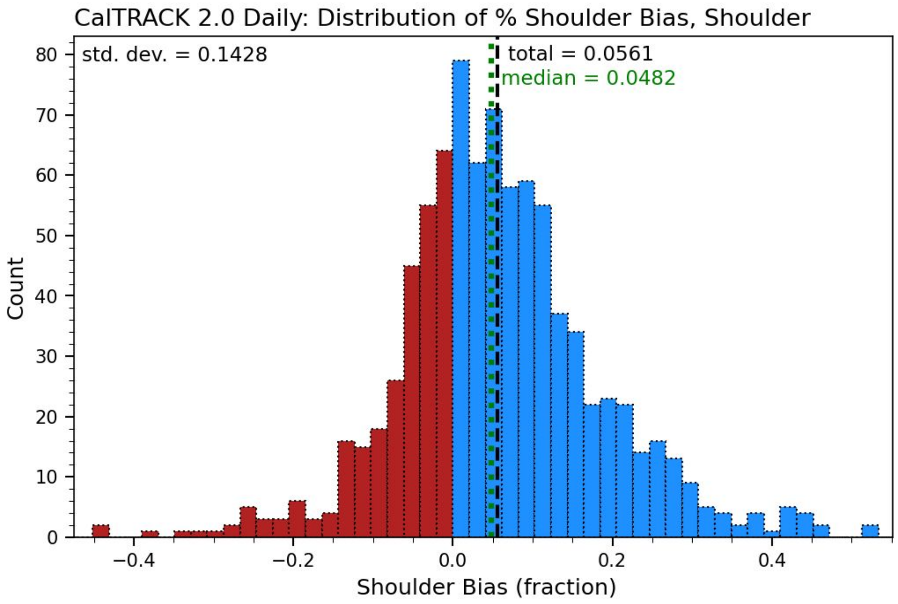
CalTRACK 2.1 (Preliminary):3.0%



SSE = 87.9

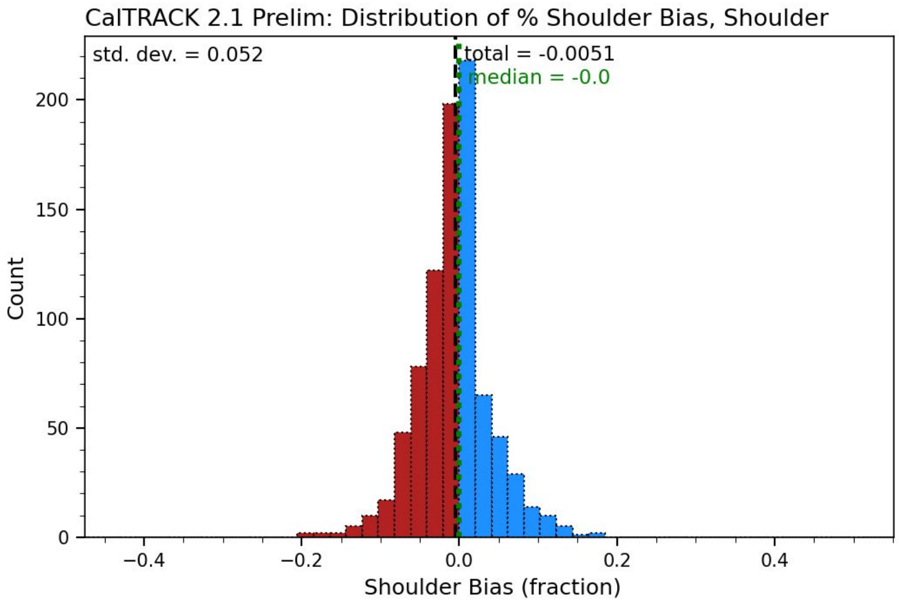
CalTRACK Shoulder Bias (Res Gas)

CalTRACK 2.0: 5.6%



SSE = 199.2

CalTRACK 2.1 (Preliminary): -0.5%

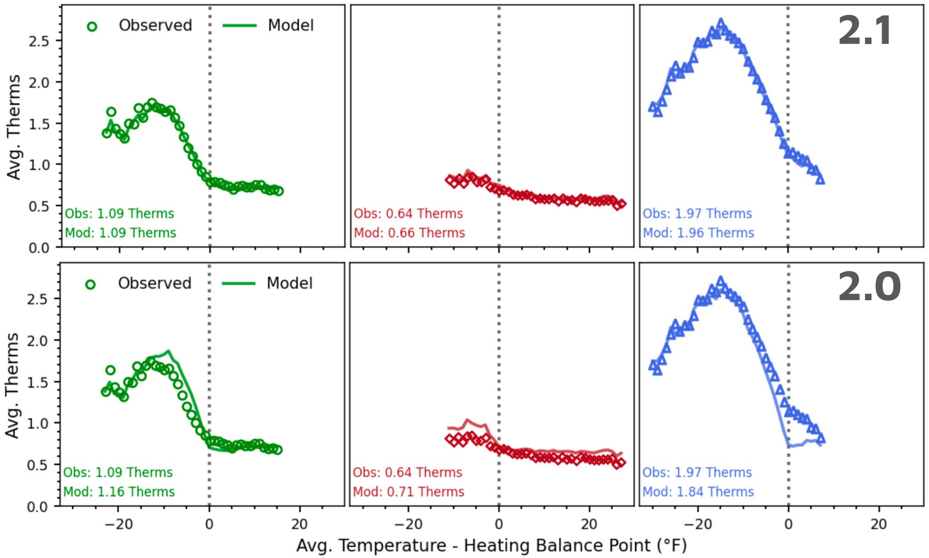


SSE = 164.2

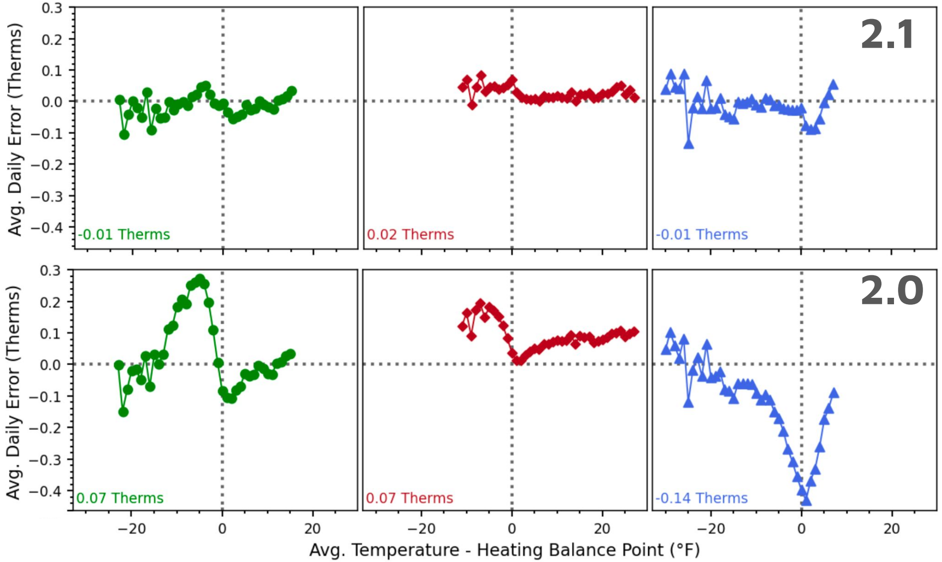
Comparison of Seasonal Error Profiles

Res Gas 880 Meter Sample

Observed vs. Model

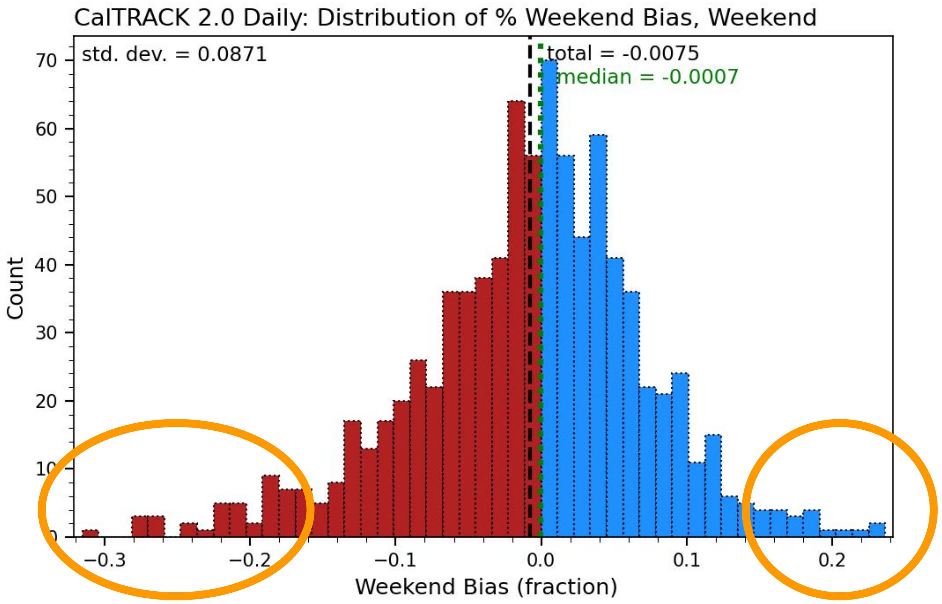


Error Profiles

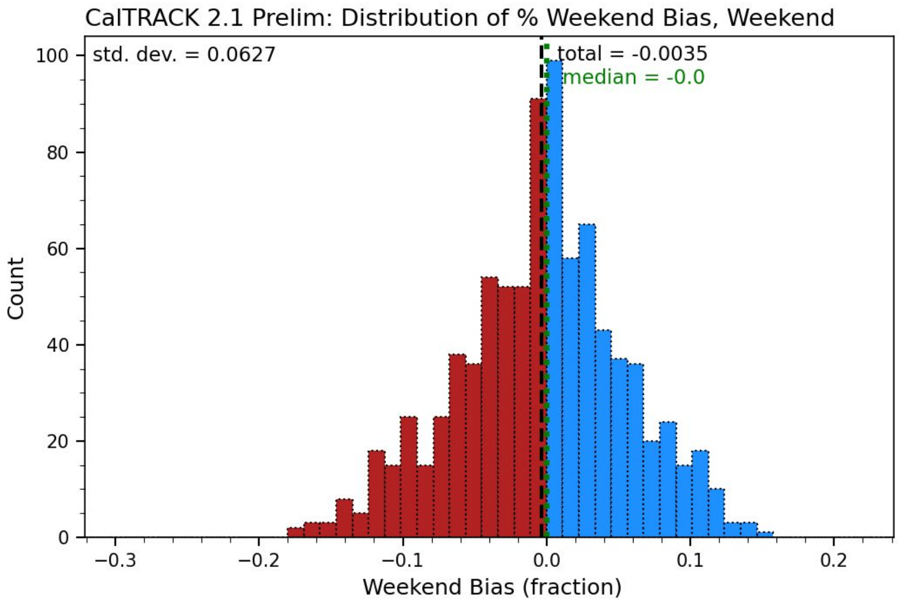


CalTRACK Weekend Bias (Res Gas)

CalTRACK 2.0



CalTRACK 2.1 (Preliminary)



CalTRACK 2.1 Prelim Splitting Behavior (Res Gas)

Number of Splits	Meters	SSE (2.0)	SSE (2.1)
0	208	378.2	375.6
1	558	818.7	722.7
2	210	360.8	291.1
3	23	936.4	668.2

Computational Efficiency

LF ENERGY

Computational Efficiency

Not viable to add more complexity to CalTRACK 2.0

CalTRACK 2.0 is Not a fast method

- CalTRACK specifies a 3°F grid search for HDD and CDD balance points
- OpenEEmeter uses 1°F increments between 30 - 90°F
- Checks all increments for HDD_TIDD_CDD, HDD_TIDD, TIDD_CDD models
- 1891 models are created and compared
- Approximately 20 - 60 seconds per meter

CalTRACK 2.1 is much more efficient

- CalTRACK 2.1 has a legacy mode
- Results are nearly identical to CalTRACK 2.0
- ~0.5 seconds per meter
- 40 - 120 times faster

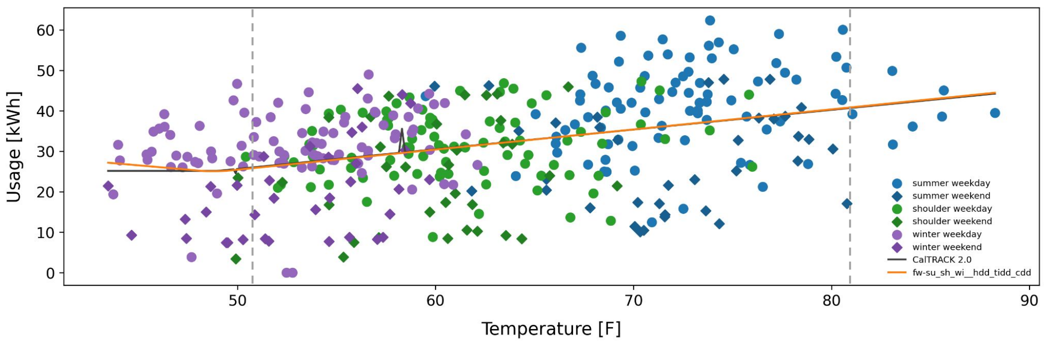
Computational Efficiency

Submodels: Grid search is dead. Long live optimization

Secret ingredient #1

- Breakpoint (balance point) optimization
  - Initial guess: BP at 10% and 90% of data
  - Use DIRECT global optimization method
    - 1 BP (hdd and cdd breakpoints are the same)
    - 2 BP

Initial guess



Computational Efficiency

Submodels: Elastic net regression is borderline magic

Secret ingredient #2

- Rather than fit multiple models, just fit one but penalize coefficients
- Technically only for linear models, but persuaded to work for smoothed models

$$objective = \underbrace{\frac{1}{N}\sum_{i}^{N} {\epsilon_i}^2}_{SSE} + \underbrace{\alpha\beta\sum_{j}^{M}|X_j|}_{lasso} + \underbrace{\alpha(1-\beta)\sqrt{\sum_{j}^{M} {X_j}^2}}_{ridge}$$

$$\epsilon \equiv residual$$

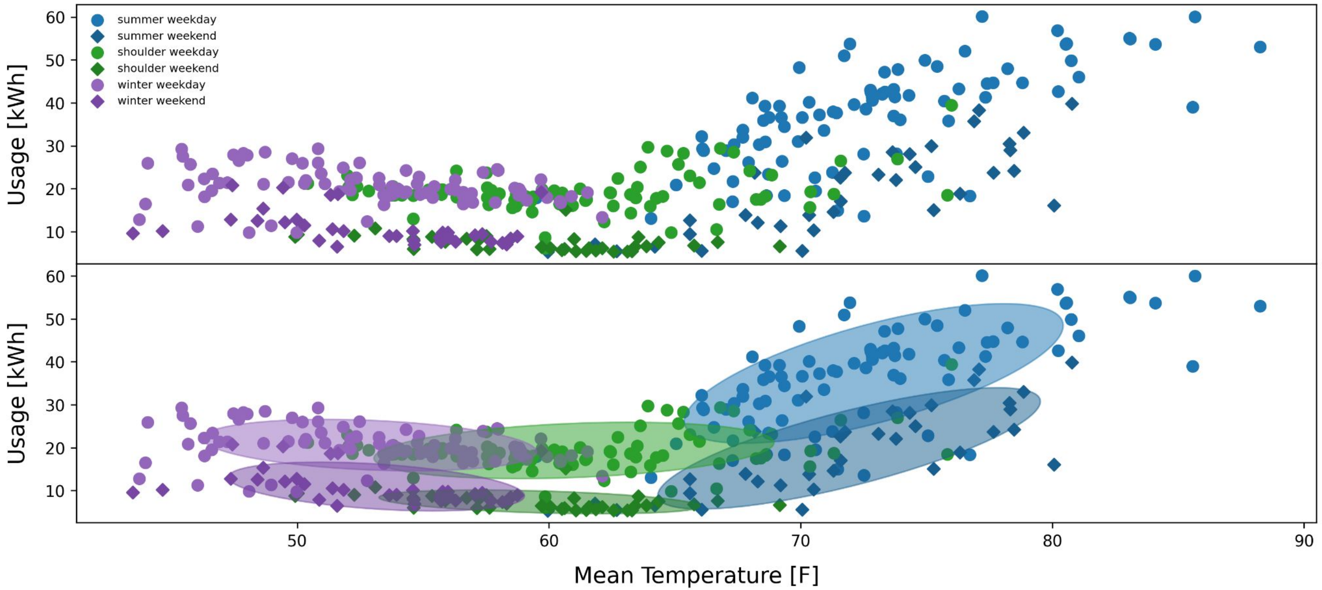
$$\alpha \equiv elastic\ net\ power$$

$$\beta \equiv percent\ lasso$$

$$X \equiv model\ coefficient$$

Computational Efficiency

Splits: Eliminate possibilities through overlapping clusters



Computational Efficiency

Splits: Only fit components once

- Standard fare: Better initial guesses, optimization algorithm etc.

- ~40-50 possible combinations of components
- Save intermediate fits and reuse

Component 1	Component 2	Component 3	Component 4	Component 5
fw-su_sh_wi				
fw-sh	fw-su_wi			
fw-sh	fw-su	fw-wi		
fw-sh	wd-su_wi	we-su_wi		
fw-sh	wd-su	wd-wi	we-su	we-wi

RMSE and Seasonal Bias Tradeoff

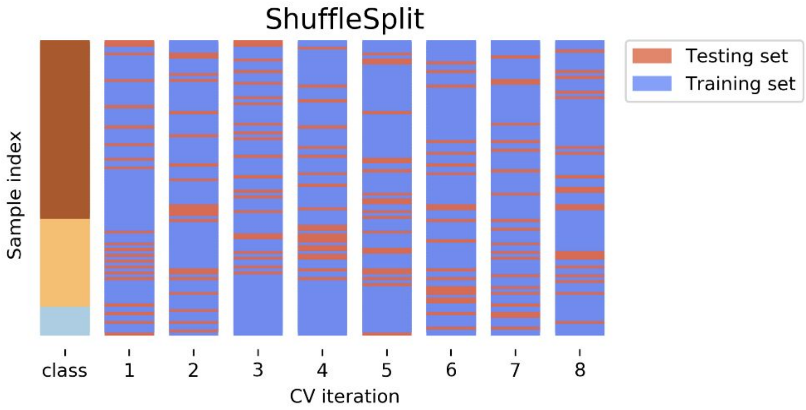
Splitting Evaluation

Using cross-validation to assess optimal splitting behavior

Cross-validation allows us to test within an existing dataset

By using CV we can estimate prediction error

- Critical to remember these models are for prediction
- Build on training set, evaluate error on test set



Splitting Evaluation

We see improvement regardless of using RMSE or seasonal bias

- If λ=0
  - RMSE improvement = 6%
  - MBE improvement = 56%
- If λ=1
  - RMSE improvement = 10%
  - MBE improvement = 46%
- What is optimal?

$$objective = \lambda \frac{RMSE_{split}}{RMSE_{unsplit}} + (1-\lambda) \frac{MBE_{split}}{MBE_{unsplit}}$$

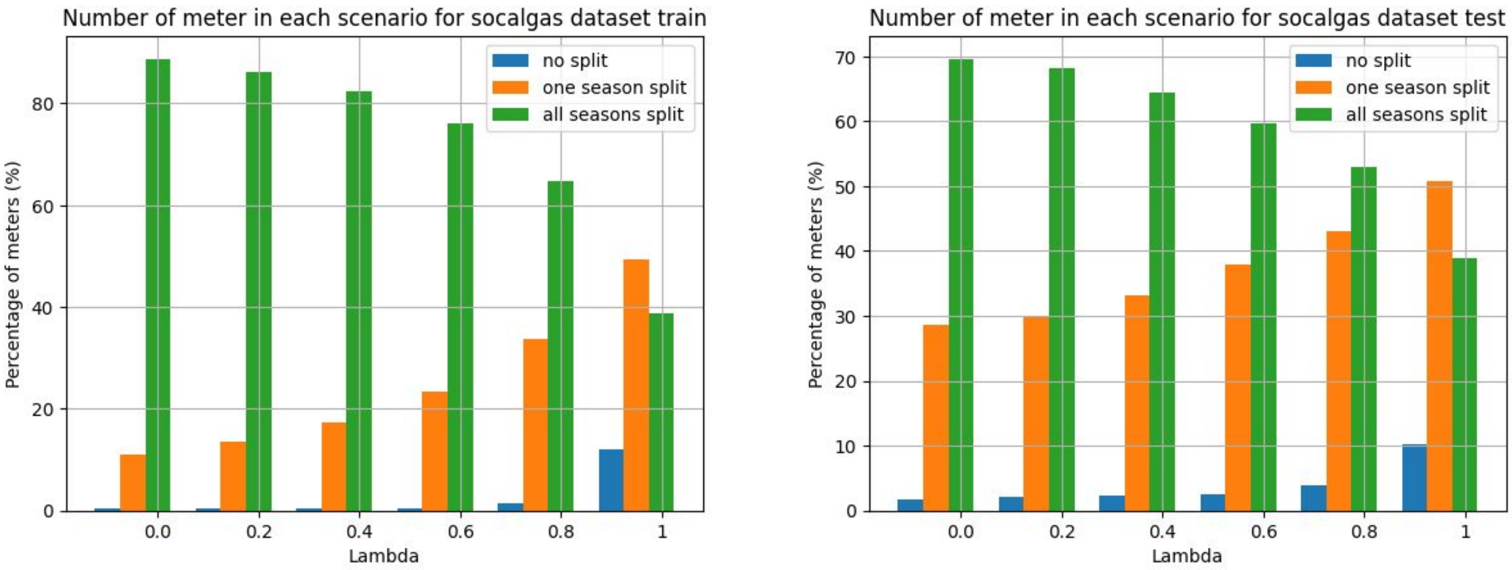
$$RMSE = \sqrt{\sum_{i}^{N} \epsilon^2}$$

$$MBE = \frac{1}{NM} \sum_{j}^{M} \sum_{i}^{N} |\epsilon|$$

$$\lambda \equiv percent\ RMSE$$

Splitting Evaluation

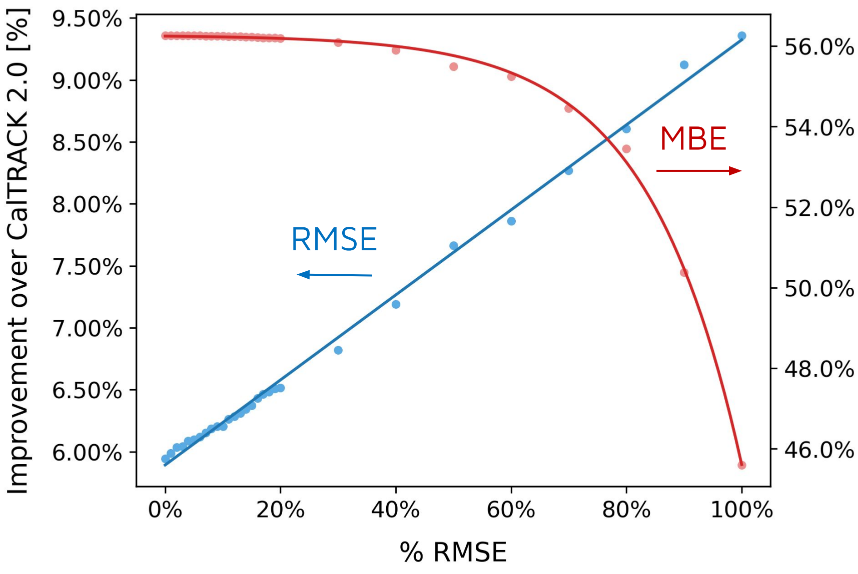
The information we have affects which splits we choose



Splitting Evaluation

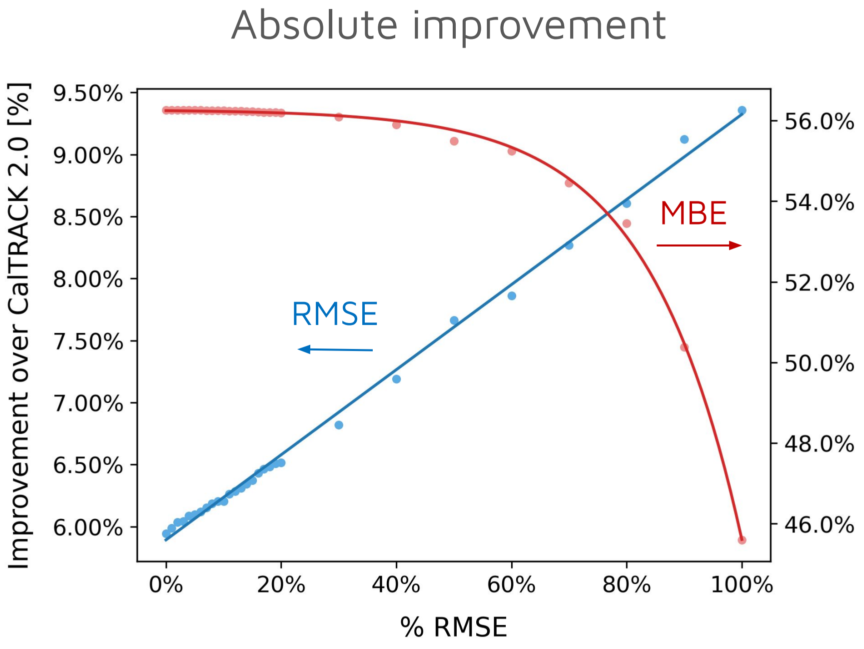
We see improvement regardless of using RMSE or seasonal bias

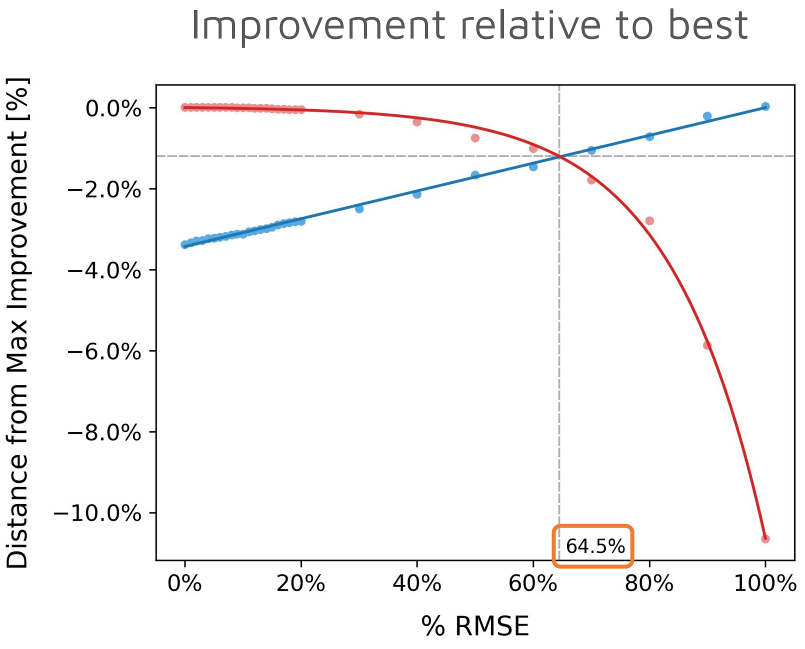
Absolute improvement



Splitting Evaluation

The optimal value of λ = 0.645 currently, this is being checked

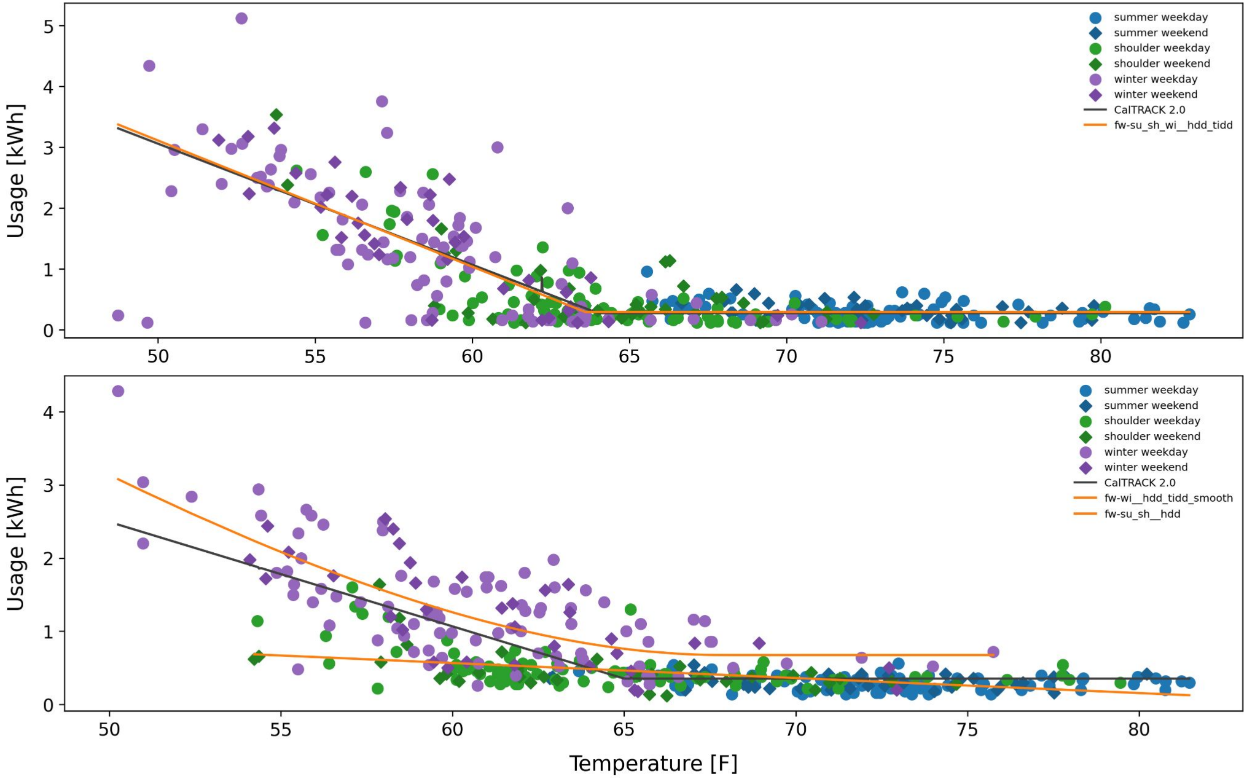




Selection Criteria Preliminary Results

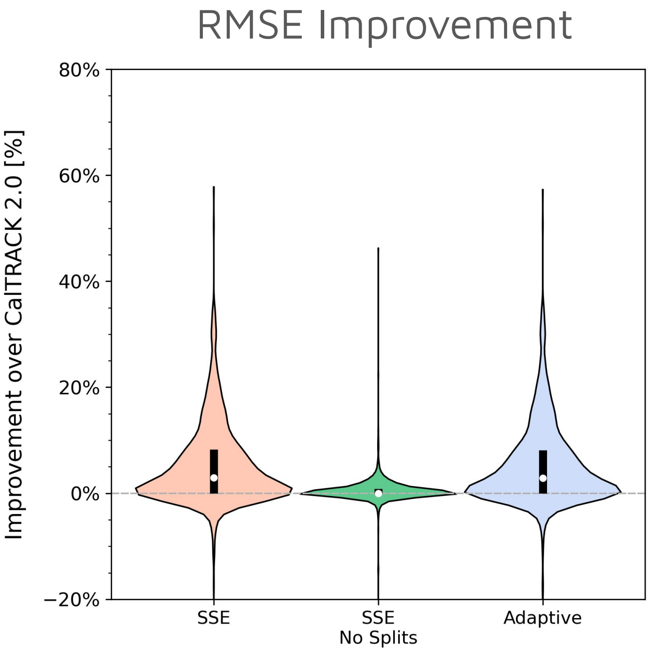
Sample Meters

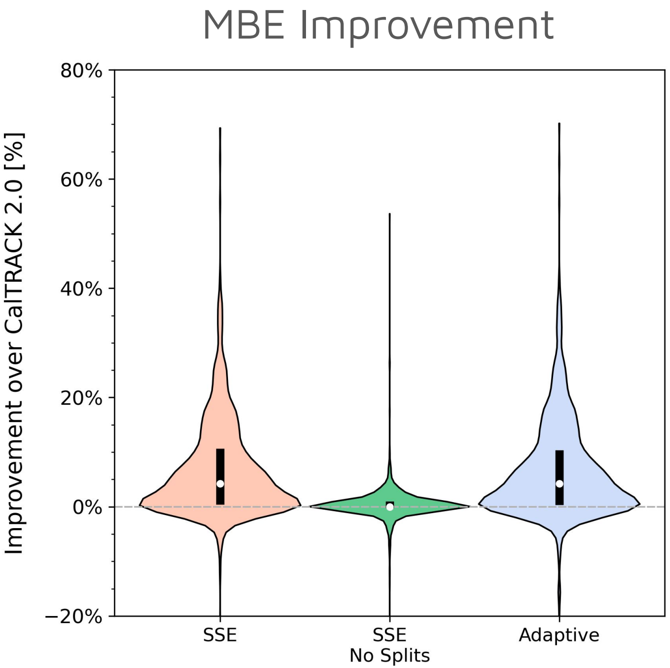
Same model produces both of these



RMSE: Gas-Residential (All Meters)

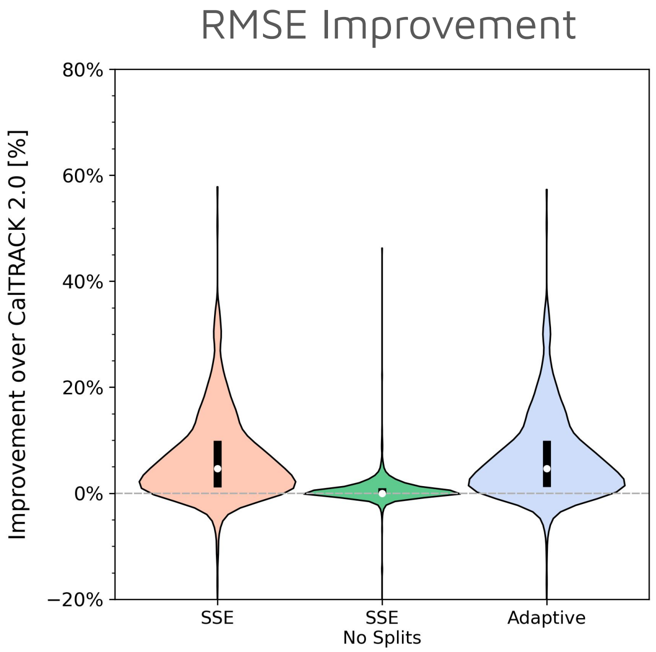
Seeing RMSE and MBE improvement in all meters (999 total)

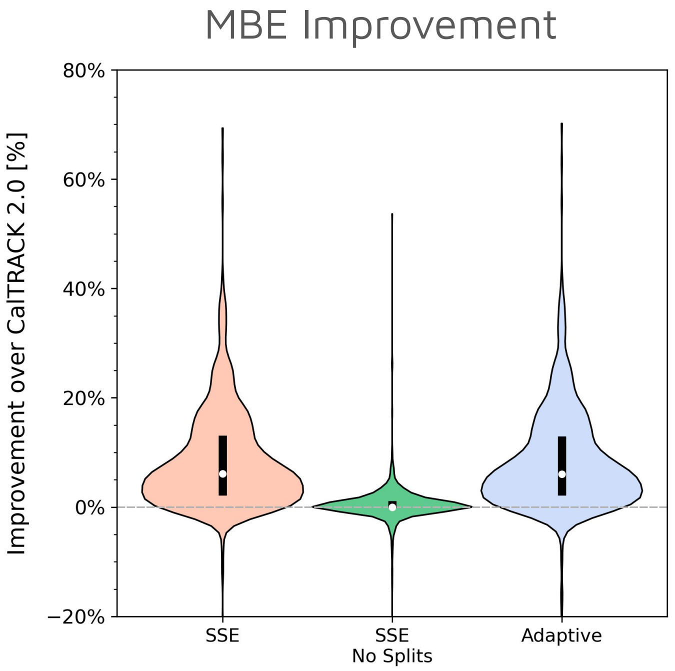




RMSE: Gas-Residential (Split Meters)

Improvement coming from split models (~80%)





Hyper Parameter Optimization

What is the best penalization/selection criterion for splits

- Based on the work of Liu, We, Zidek (1997)
- Hyperparameters to optimize: α, β, ω, and η

$$log\ likelihood \equiv \mathcal{L} = -\frac{N}{2}\left(\ln(2\pi) + \ln\left(\frac{loss}{N}\right) + 1\right)$$

$$BIC_{mod} = -2\mathcal{L} + \omega^* K^* \log(N)^{\eta}$$

$$N = \text{Number of data points (days)}$$

$$loss = \text{mean SSE}$$

$$K = \text{Number of models (splits + 1)}$$

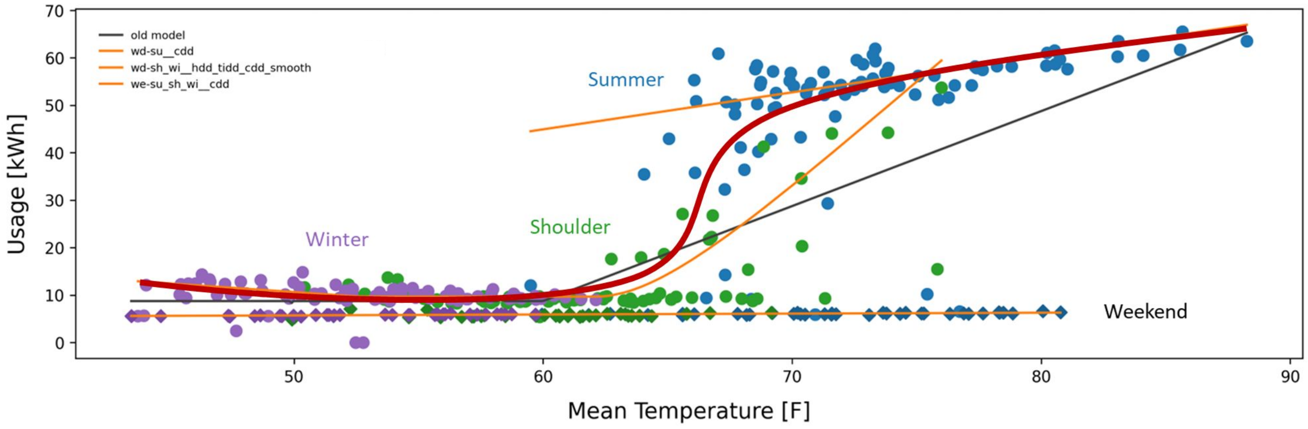
$$\alpha = \text{Regularization penalty term}$$

$$\beta = \text{Percent of elastic net which is lasso}$$

What's Left

Next time

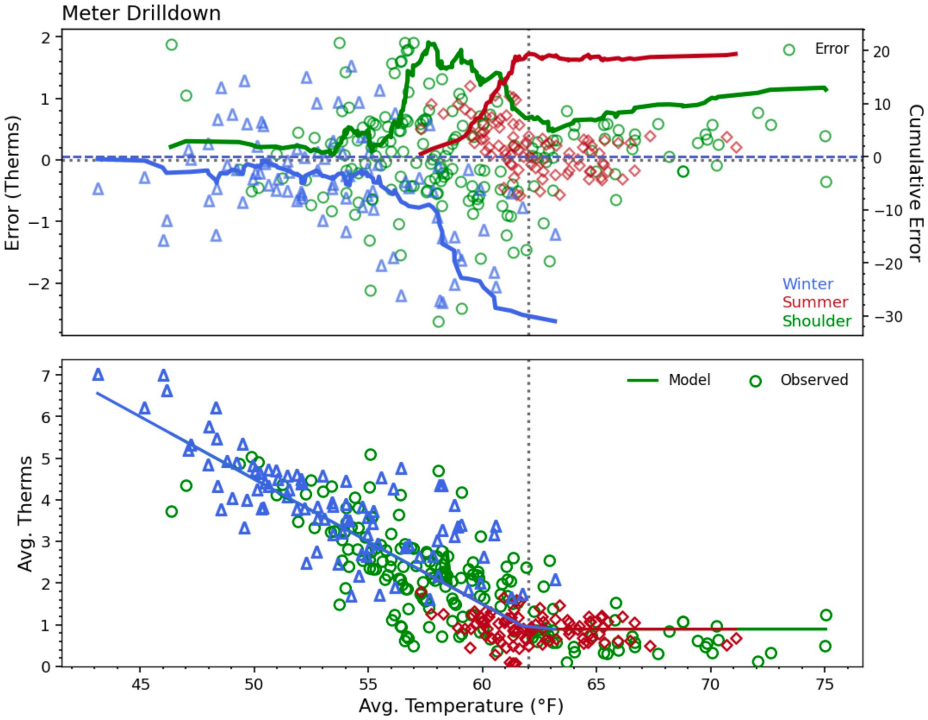
- New models? Push to next CalTRACK daily version?



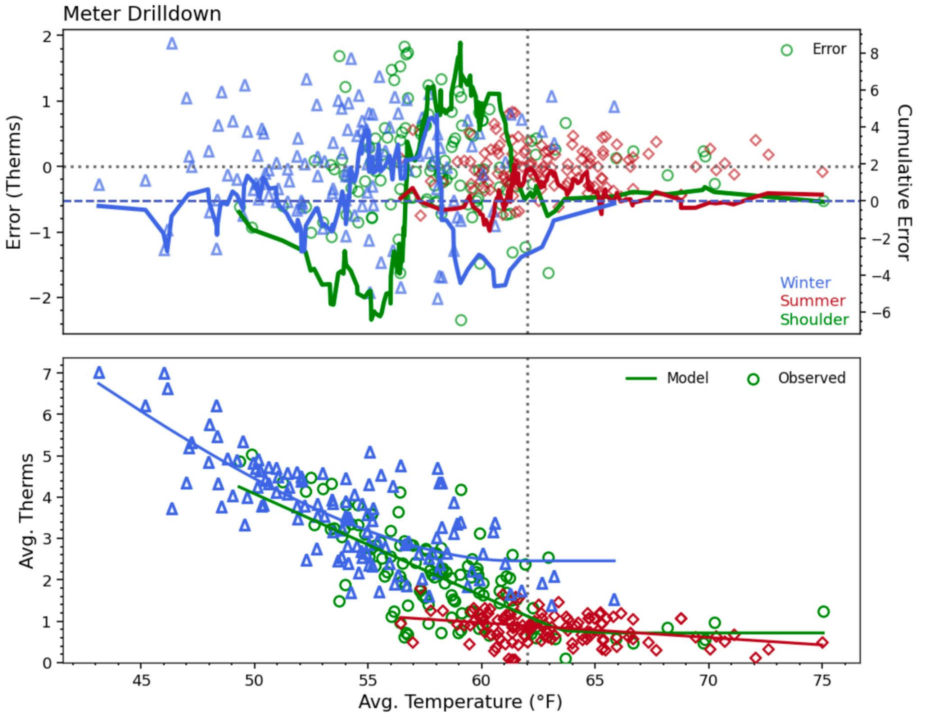
- Hyperparameter optimization (expensive)
- Final CalTRACK 2.1 Daily recommended model formulation (ambitious)

Meter Example

CalTRACK 2.0



CalTRACK 2.1 (Preliminary)



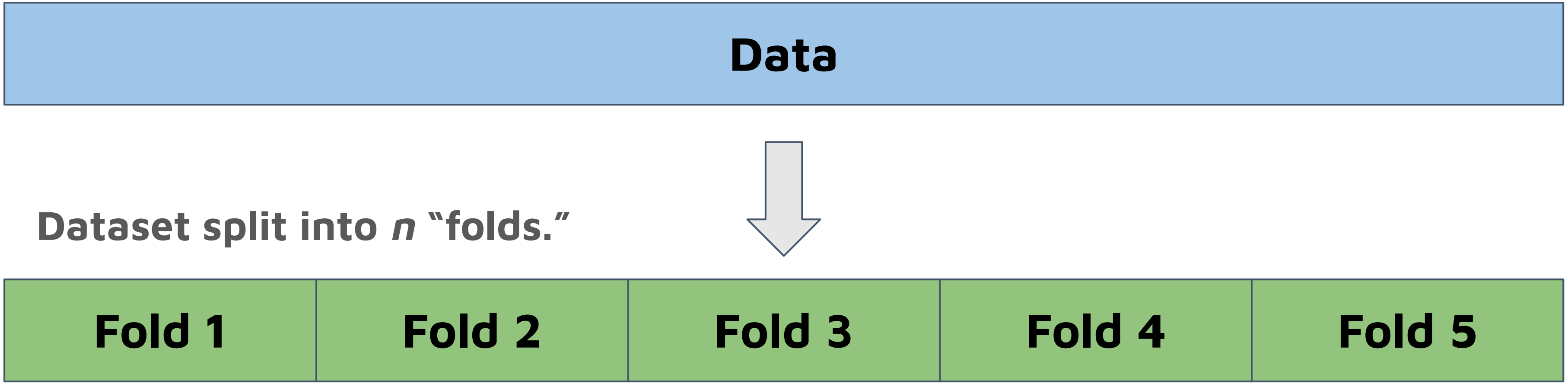
Appendix

LF ENERGY

Open Questions

1. What is the best way to penalize additional parameters?
2. Is Cross Validation a viable penalization option itself or is it too computationally intensive?
3. Is there a better formulation for thermal lag?
4. What is ultimately the balance we should be striving for between **solving seasonal bias** vs. **remedying other sources of bias** vs. **overfitting** vs. **model complexity** vs. **computational cost**?

Cross Validation: A Rigorous Approach to Test/Train Splitting



Each fold takes a turn as the test data with remaining folds serving as training data. Each iteration is a "split."

n **models developed. Model parameters and performance determined by averaging results on test samples.**

How to Avoid Overfitting: Penalization

Balancing model error/performance with number of parameters

How (Preliminary)?

- Selection Criterion (AIC for example): Introduce penalty that increases with model complexity
- Additional penalty term
- Cost/Benefit test on adding parameters
- Empirical estimate of Cross Validation

Open Questions

What metric to use?

- RMSE/weighted RMSE/MAE
- Savings uncertainty of non-participants in reporting year

Thermal lag seems largely detrimental

- Is this agnostic to implementation?
- Bugs?

Optimization!

- Many parameters here:
    - Optimization Algorithm (Subplex, SLSQP, COBYLA, ...)
    - Selection criteria (AIC, BIC, ...)
    - Segmentation penalty amount
    - SSE, L1, Adaptive