KTH ROYAL INSTITUTE
OF TECHNOLOGY

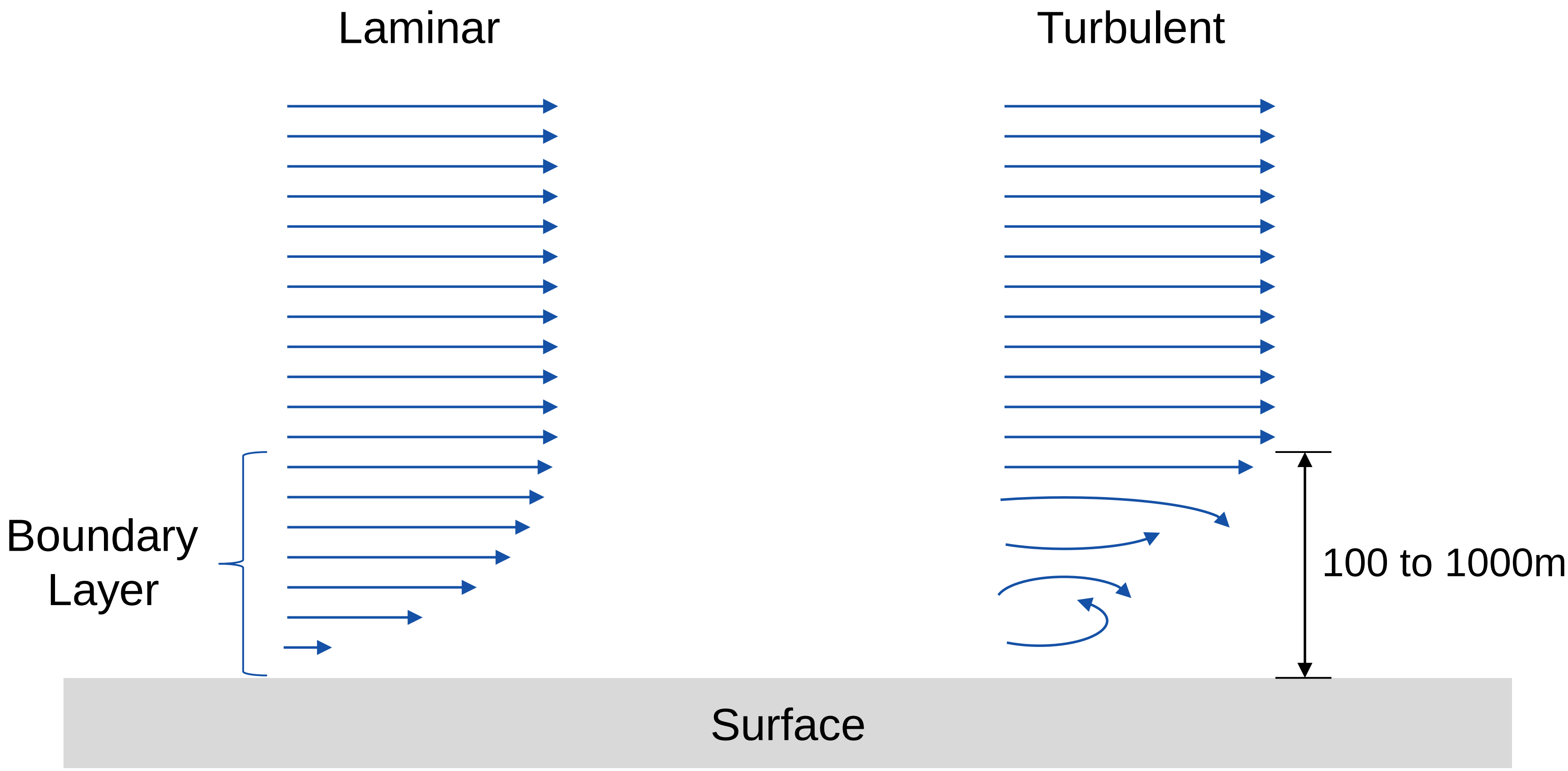
Predicting the Momentum Flux-Profile Relationship from Macro Weather Parameters

Emilio Dorigatti

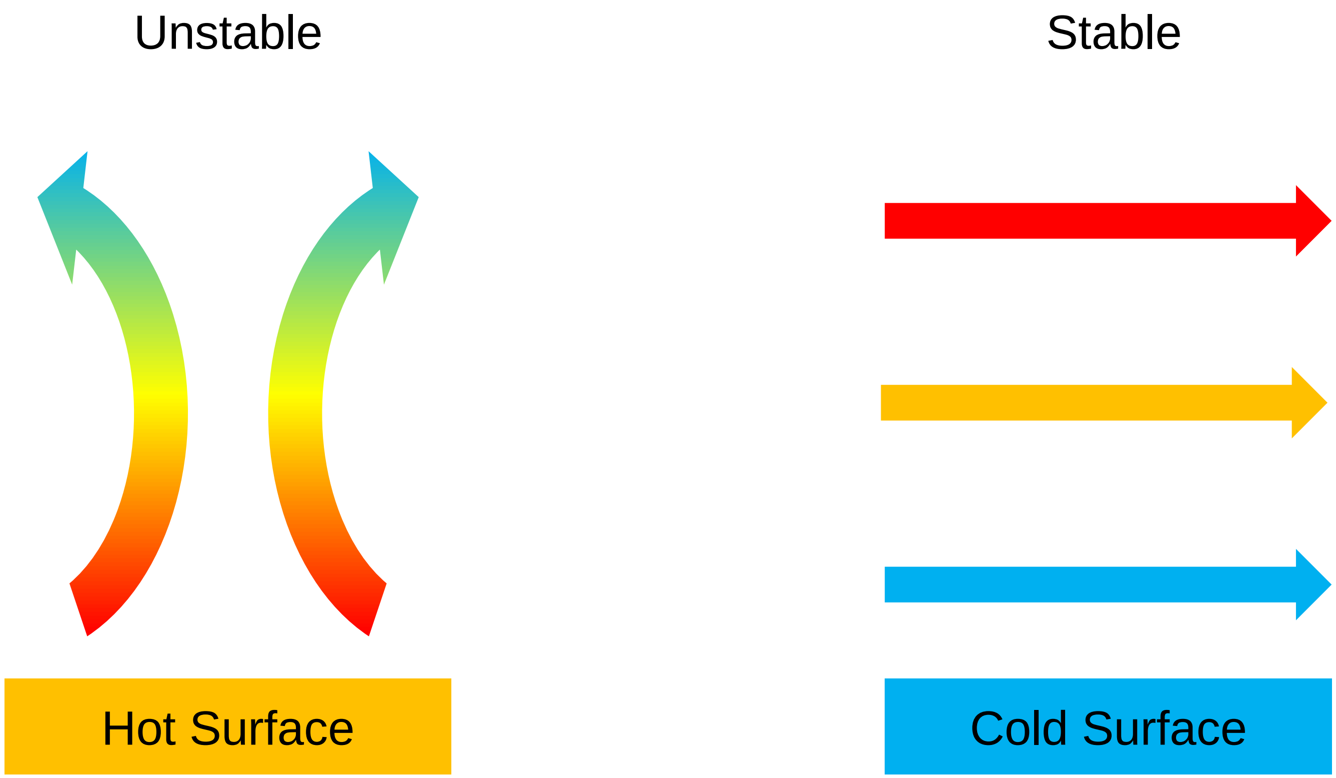
Motivation

1. Avg. surface temperature rose by 1°C in the last century
2. Paris agreement to hold future increase below 2°C
3. Adhering nations set their own goals
4. Understand climate to find effective and realistic goals
5. Climate Science is based on simulations
6. Simulations use many physical theories
7. Lots of efforts are made to improve its predictions
8. We target the Monin-Obukhov similarity theory

The (Atmospheric) Boundary Layer

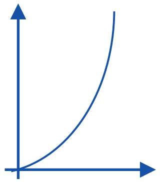


Stable, Neutral and Unstable Boundary Layers



Momentum? Flux? Profile?

Wind profile $\partial\bar{u}/\partial z$
change of horizontal wind
speed with altitude



Relationship between

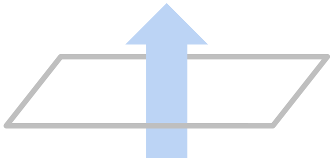
the momentum flux

and

the wind profile

$$\boxed{\frac{\partial\bar{u}}{\partial z}\frac{\kappa z}{u_*}}$$

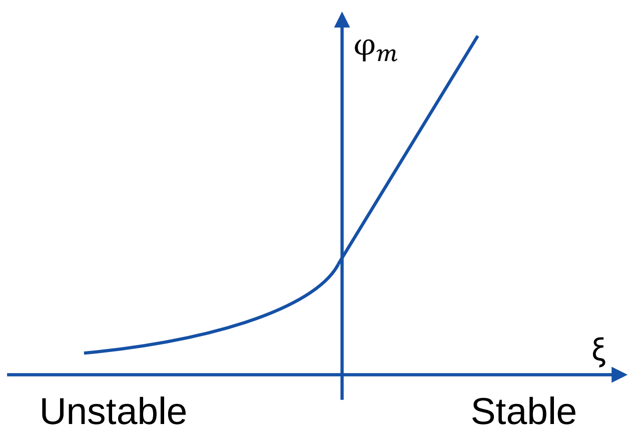
Momentum flux u_*
vertical movement of air
due to turbulence



Monin-Obukhov Similarity Theory

Obukhov Length $L = -\frac{u_*}{k\frac{g}{\theta_v}\frac{Q}{\rho c_\rho}}$ Instability Parameter $\xi = {}^{z}/_{L}$

$$\frac{\partial \bar{u}}{\partial z}\frac{\kappa z}{u_*} = \varphi_m(\xi) = \begin{cases} 1 + 6\xi & 0<\xi<1 \\ (1 - 19.3\xi)^{-0.25} & -2<\xi\leq 0 \end{cases}$$



Problem

MOS theory works, but

- in a limited set of conditions
- requiring many quantities that are hard to measure
- and is not very precise

RQ1: How can the data from the Cesar database be used to predict the momentum flux-profile relationship more accurately than the MOS theory, by using only macro-weather parameters?

RQ2: What impact do different macro-weather parameters have on the quality of the predictions obtained?

Method

1. Retrieve, process and clean the data
2. Compute the momentum flux-profile relationship
3. Reproduce the MOS theory (our baseline)
4. Predict the flux-profile relationship (RQ1)
5. Find the importance of the features (RQ2)

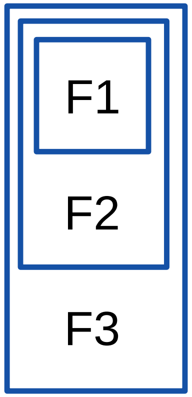
Data

Cabauw Observatory

- The Netherlands
- 213 meters high mast
- Measures several macro and micro weather parameters
- One data point every 10 minutes at 3 altitude levels
- We use data from 1/1/2001 to 31/12/2017
- 1.3 million data points after cleaning

Feature Importance

Compare the prediction error on 5 different feature sets



- Lower error iff feature set contains useful information
- Feature sets constructed cumulatively
  - Domain expertise
  - Second version with hourly trend (5x2=10)
- Traditional methods cannot deal with:
  - Multicollinearity (linear regression, trees, etc.)
  - Complex interactions (mutual information, etc.)
  - Large datasets (permutation importance, etc.)

Features

For RQ2:

1. Altitude z, wind speed at z, 10m, 20m and 40m, air temperature at z, 10m, 20m and 40m, soil temperature
2. … + Soil heat flux
3. … + Net radiation
4. … + Rain amount, dew point
5. … + Turbulent sensible and latent heat fluxes

For RQ1:
F1+F2+F3+F4+Specific humidity, relative humidity, air pressure, air density, virtual air temperature

Prediction

- Feature engineering: standardization and trend (RQ2 only)
- Random search for hyper-parameter optimization
- Nested CV for RQ1
  - Good for assessment (unbiased, but higher variance)
  - Ridge, k-Nearest Neighbors, Gradient Boosted Trees
- Repeated CV for RQ2
  - Good for comparison (biased, but lower variance)
  - For every feature set
  - Only Gradient Boosted Trees

Prediction (cont'd)

Two datasets:

1. MOS theory validity range
2. Full dataset

Inherent time dimension in the data

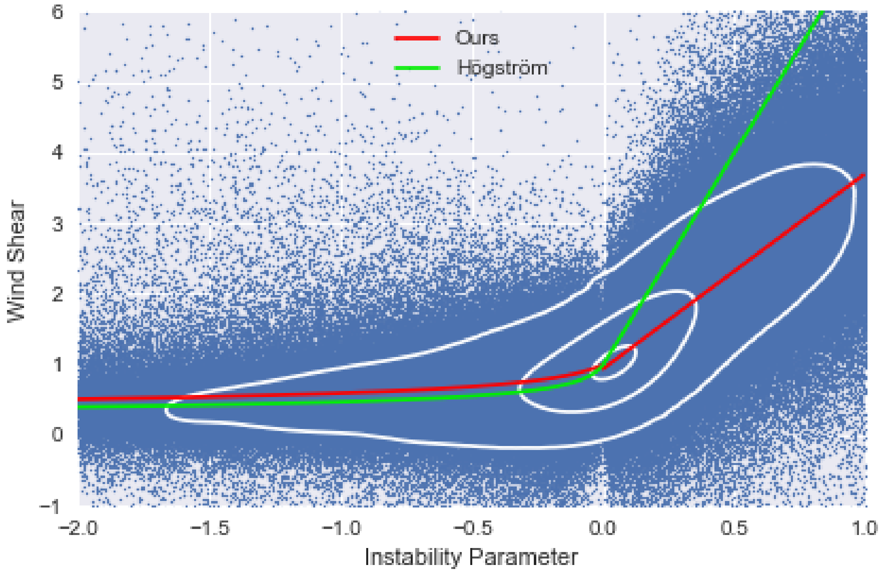
- Must be considered during cross validation
- Not considered in the models
- Train and test sets are created on whole months

Feb. 2011	Mar. 2011	Apr. 2011	Jun. 2011	Jul. 2011	Aug. 2011

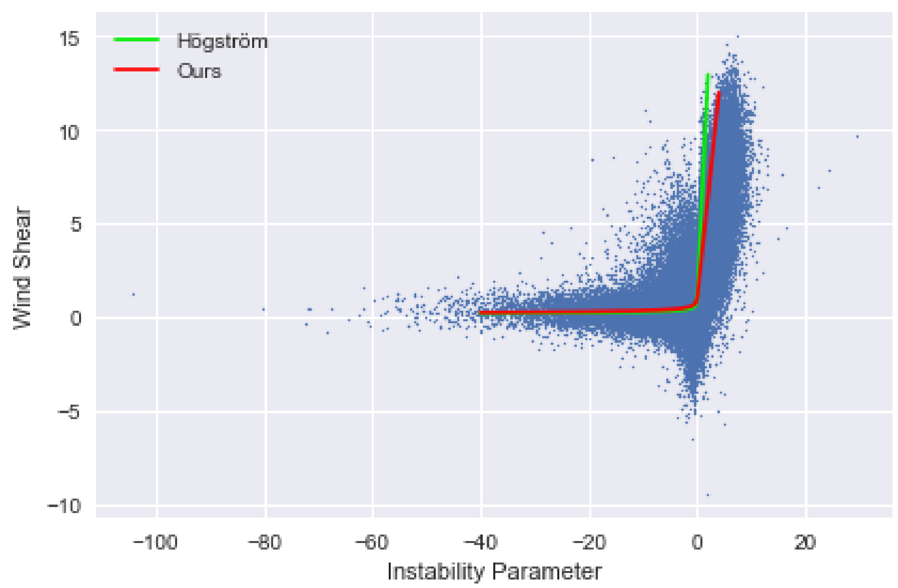
Results

Monin-Obukhov Similarity Theory

MOST Dataset
83% of the data

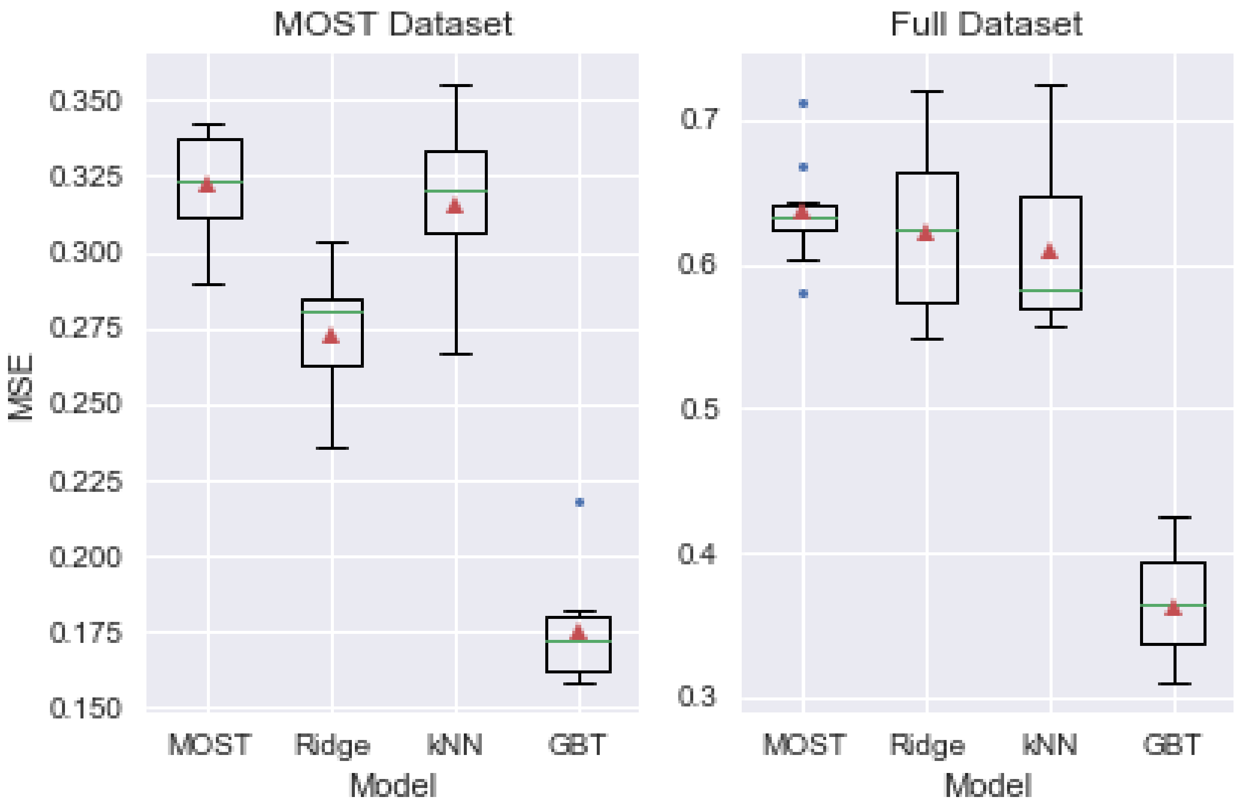


Full Dataset

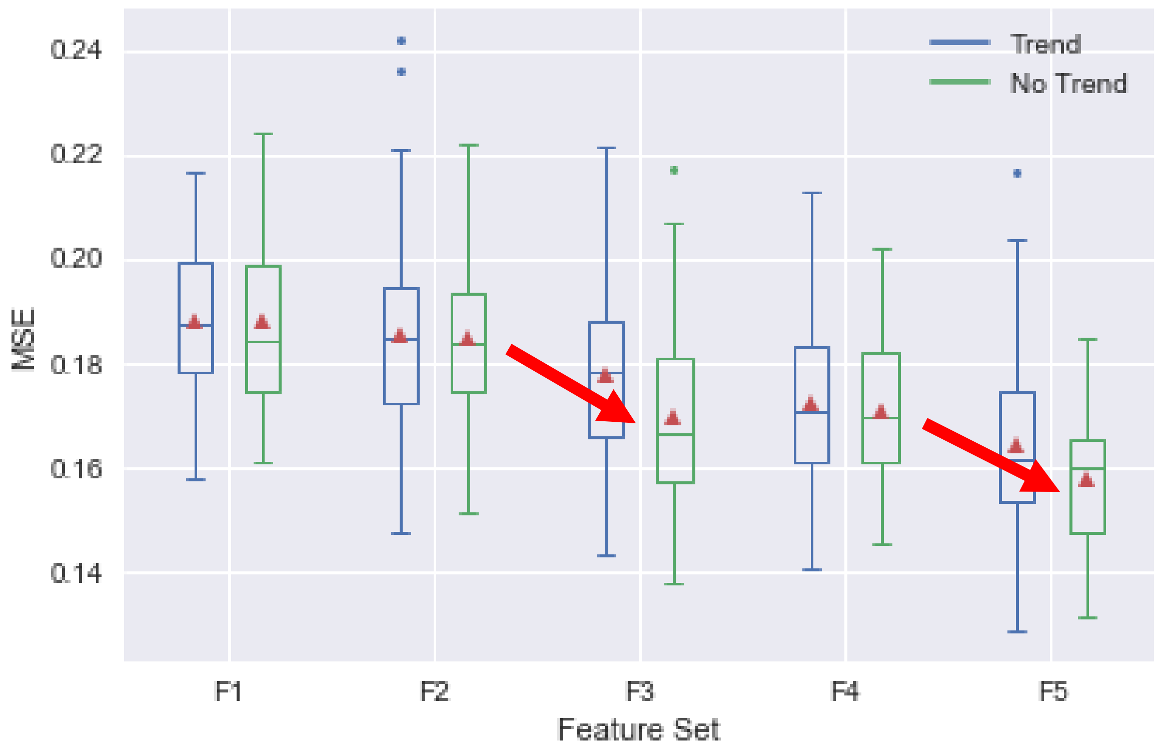


Results

First Research Question

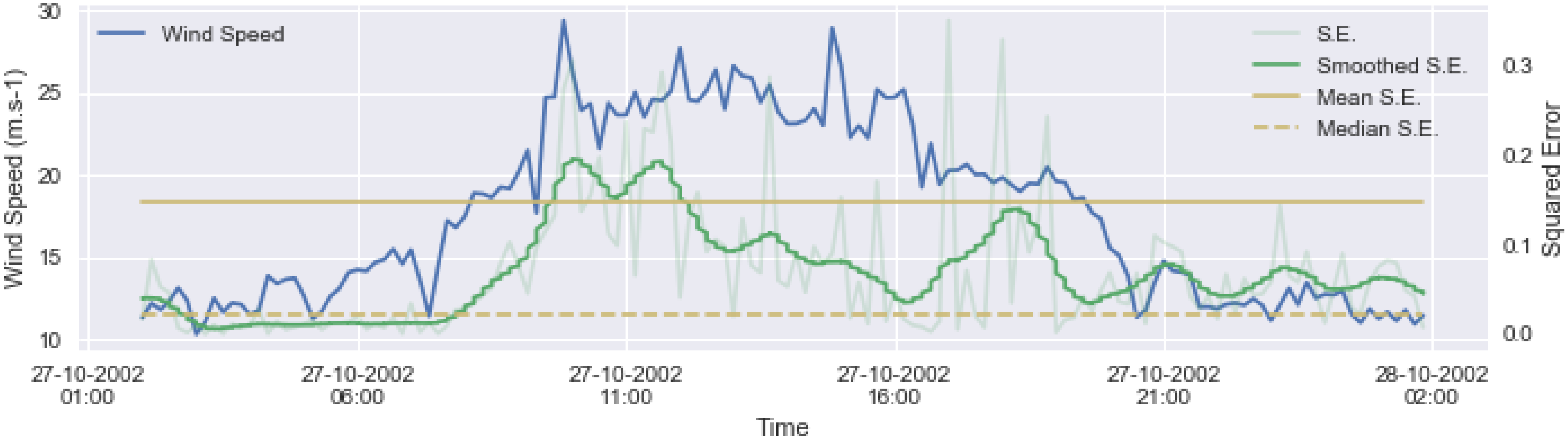


Results
Second Research Question

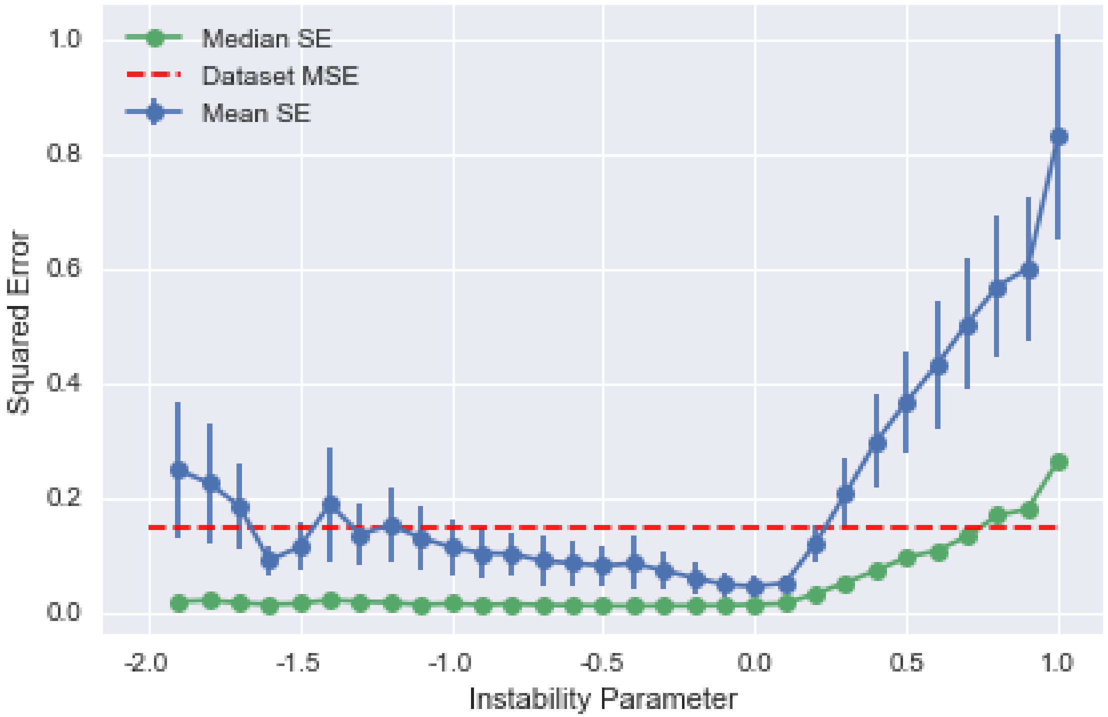


Improvement is small (5%), but significant (more than 0.5 std)

Prediction Error During a Cyclone



So Where is the Error?



Conclusions

- Yes, we can do better (error is halved)
- Wind gradient is hard to estimate
- Radiation is important, humidity and soil heat are not
- Quality affected by measurement errors

Future Work

- Deploy in actual simulations (error is halved… So what?)
- More locations to improve generalization
  - Turbulence is heavily affected by surface
  - Land, Ocean, Sea-ice
- Hybrid model (macro+micro)
- Predict other flux-profile relationships
  - Temperature and moisture
- Fancier Machine Learning

Deep Neural Networks

Fully Connected

- 128/64/32/16/8/4/2/1 neurons (~15k parameters)
- No regularization, PReLU, 1024 Batch Size
- Adam, Early Stopping

Validated on a random subset of 20 months

Dataset	MOST	GBT	F1	F2	F3	F4	F5
Full	0.64	0.36	0.31	0.29	0.26	0.24	0.21
MOST	0.32	0.18	-	-	-	-	0.13

GRU using all samples from previous hour

- Worse than MOST estimator

Ensemble of Linear Models

Average predictions of

1. Polynomial Features
2. Feature Standardization
3. PCA
4. Ridge Regression

MSE on 10-fold CV: 0.19 +/- 0.02

- Gradient Boosting was 0.17 +/- 0.02
- Good for deployment

Hyper-Parameter Optimization

Bayesian optimization with Gaussian processes

- Fit regression model, predict expected improvement

Hyperband

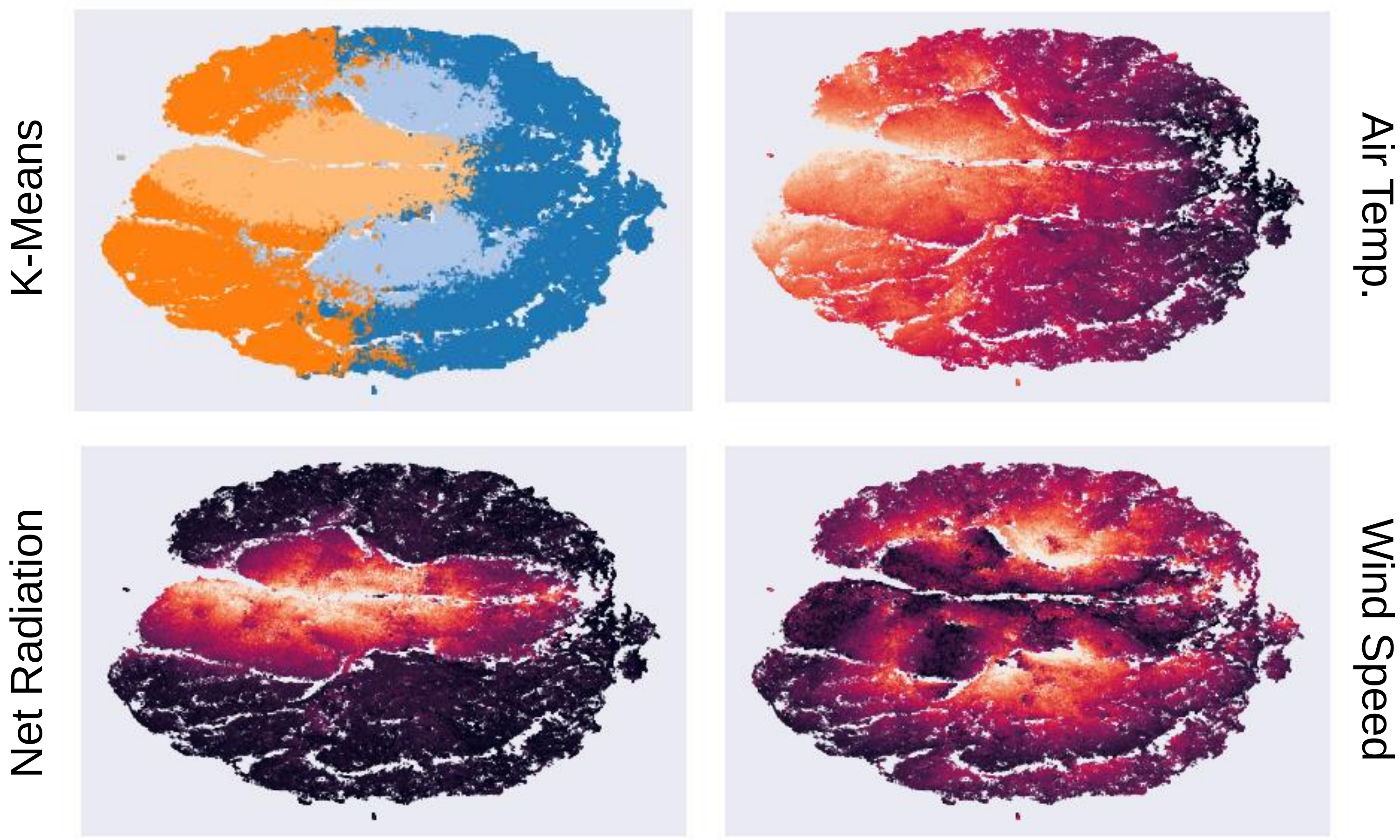
- Allocate more resources to promising combinations

1. Train xgboost on train set
2. Optimize score on validation set (20 months)
3. Evaluate on test set at the end (20 months)

Results (for both): 0.161 validation, 0.167 test

Thanks!

Seasons in the Data



$$L = -\frac{u_*^3}{\kappa \frac{g}{\theta_v} \frac{Q}{\rho c_\rho}} \tag{2.4}$$

where

- $g = 9.81\,\mathrm{m\,s^{-2}}$ the acceleration due to Earth's gravity
- θ_v virtual temperature[4] obtained as

$$\theta_v = \theta \frac{1 + r_v/\epsilon}{1 + r_v} = \theta(1 + 0.61 \cdot q) \tag{2.5}$$

  where θ is the air temperature, r_v is the mixing ratio, $q = r_v/(1 + r_v)$ the specific humidity, and ϵ is the ratio of the gas constants of dry air and water vapor, roughly 0.622.

- Q the buoyancy flux, approximated by $H + 0.07\lambda E$ and measured in $\mathrm{W\,m^{-2}}$
- ρ the air density, computed from the pressure P and the specific gas constant for dry air $R = 287.058\,\mathrm{J\,kg^{-1}\,K}$ as

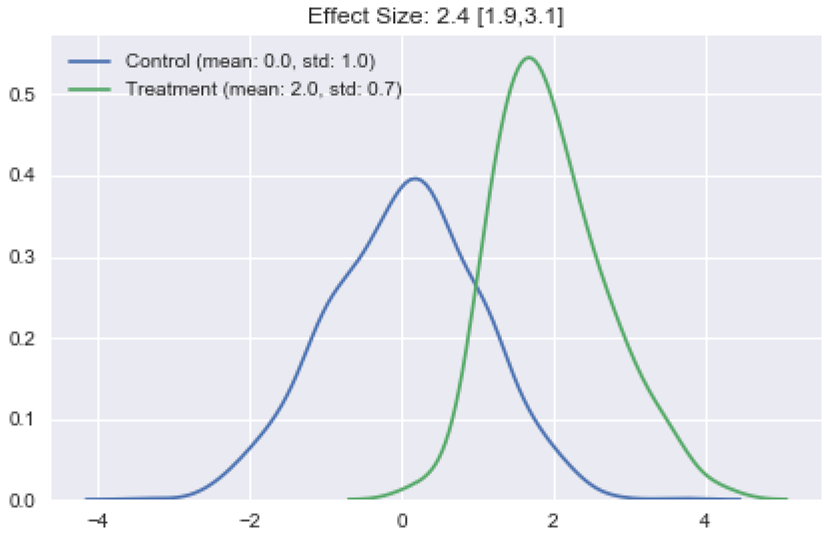
$$\rho = \frac{P_0}{RT_v}$$

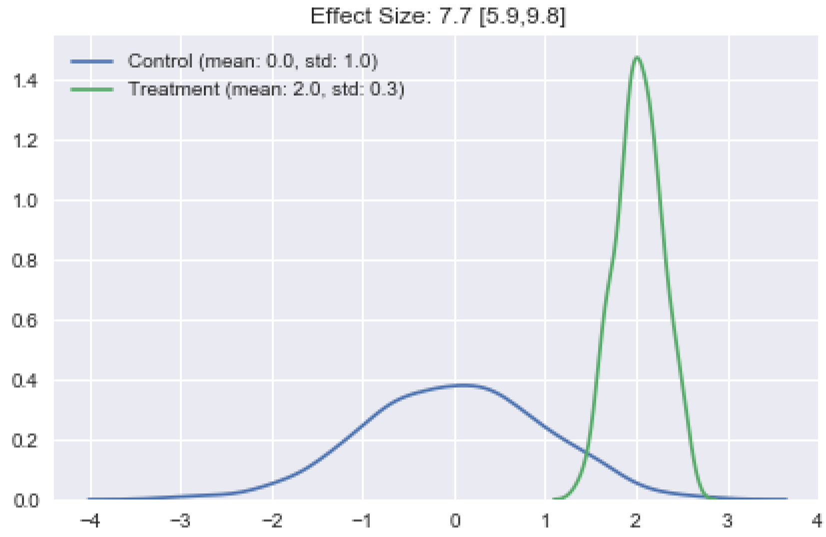
- c_ρ specific heat of dry air, $1005\,\mathrm{J\,kg^{-1}\,K^{-1}}$ at $300\,\mathrm{K}$

[4] Potential temperature of dry air if it had the same density as moist air. It allows to use formulas for dry air when the air is not dry.

Effect Size

Treatment vs. Control





Descriptive Statistics

Table 4.1: Basic descriptive statistics of the features in the Cesar database. H, λE, the net radiation and the soil heat flux are measured in $\mathrm{W\,m^{-2}}$, u_* and the wind speed in $\mathrm{m\,s^{-1}}$, the CO_2 flux in $\mathrm{mg\,m^{-2}\,s^{-1}}$, the temperatures in K, the air pressure in hPa, the air density in $\mathrm{kg\,m^{-3}}$, the rain amount in $6\,\mathrm{mm\,h^{-1}}$, and the specific humidity is a pure number.

	Mean	**Std.**	**Min.**	**25%**	**50%**	**75%**	**Max.**
H	3.7	47.3	-460.0	-25.8	-14.7	28.2	479.6
λE	74.7	114.9	-534.8	-4.9	23.8	132.9	6,940.8
u_*	0.3	0.1	0.1	0.2	0.3	0.4	1.8
CO_2 **Flux**	-0.1	0.3	-7.5	-0.2	0.0	0.1	6.1
Wind Speed	5.9	2.6	1.0	4.0	5.4	7.2	29.4
Air Temp.	284.7	6.4	259.6	280.1	284.8	289.4	306.9
Soil Temp.	284.4	5.7	271.9	279.6	284.3	289.1	302.6
Dew Point	280.3	6.1	188.0	276.6	280.6	284.7	296.7
Spec. Hum.	6.6	2.5	-0.9	4.8	6.4	8.3	17.8
Rel. Hum.	76.5	15.2	0.0	66.9	79.2	88.3	133.7
Air Press.	1,009.6	69.0	960.9	1,008.3	1,014.8	1,020.7	1,046.4
Rain Amount	0.0	0.1	0.0	0.0	0.0	0.0	15.1
Net Rad.	81.8	167.2	-158.5	-35.4	-3.0	172.5	7,775.4
Soil Heat Flux	2.9	15.6	-77.8	-6.5	-0.7	8.9	1,139.0
Air Dens.	0.3	0.1	0.1	0.2	0.3	0.3	2.8
Virt. Air T.	1,443.8	454.9	124.8	1,104.2	1,391.0	1,753.2	3,506.0

Gap-Filling Statistics

Table 4.2: Number of measurements by gap filling method. Due to inconsistencies in the Cabauw Documentation (Bosveld 2014), we cannot be sure of the meaning of the columns, but this is our interpretation: 0 - Unknown, 2 - Cabauw In situ, 3 - Automatic Weather Station in Cabauw, 4 - Computed from profiles, 5 - Interpolated, 6 - Cabauw-based model, 7 - De Bilt-based model.

	0	**2**	**3**	**4**	**5**	**6**	**7**
H	-	1,192,346	-	-	17,291	97,720	1,005
λE	-	1,180,434	-	-	18,503	97,846	11,579
u_*	-	1,174,593	-	132,461	465	618	225
CO_2 **Flux**	-	964,749	-	-	48,983	287,855	6,775
Wind Sp.	-	1,302,308	2,338	-	879	331	2,506
Air Temp.	-	1,298,075	-	-	2,419	5,959	1,909
Spec. H.	-	1,292,626	-	-	3,149	10,693	1,894
Air Press.	7,096	-	1,298,205	-	654	-	2,407
Rain Am.	7,096	1,181,167	117,844	-	346	-	1,909
Net Rad.	-	1,274,106	-	20,962	1,715	-	11,579
Soil H. Fl.	-	1,232,822	-	-	-	72,256	3,284

MOST Estimator Results

Table 4.3: Evaluation metrics for the MOST estimator on both datasets obtained with nested cross-validation. The first number is the average over the 10 outer folds, followed by the standard deviation in parentheses.

Metric	Full Dataset	MOST Dataset
MSE	0.64 (0.04)	0.32 (0.02)
R^2	0.71 (0.01)	0.61 (0.01)
MAE	0.53 (0.01)	0.39 (0.01)
mAE	0.37 (0.01)	0.28 (0.01)
MAPE	257.80 (207.39)	178.30 (68.99)
mAPE	28.48 (0.81)	24.44 (0.81)

Table 4.4: Values of the coefficients of the MOST estimator of equation 3.7 fitted on the MOST dataset. The minimum values for c and d seem to be outliers, as well as the next smallest value, but the other 8 values are closely clustered together within an interval of about 0.15.

	Mean	Std.	Min.	25%	50%	75%	Max.
a	0.94	0.00	0.94	0.94	0.94	0.95	0.95
b	2.77	0.01	2.76	2.77	2.77	2.78	2.78
c	2.45	0.47	1.25	2.61	2.65	2.67	2.71
d	-0.30	0.11	-0.60	-0.26	-0.26	-0.26	-0.25

Comparison with Baseline

Table 4.5: Descriptive statistics of the MSE achieved by each estimator with nested CV, and effect size comparing it with the MOST baseline (control). Gradient boosted trees perform better in both datasets, and Ridge regression beat the MOST estimator on the MOST dataset.

Dataset	Model	Mean	Std.	Min.	50%	Max.	Effect Size
Full	**MOST**	0.64	0.04	0.58	0.63	0.71	-
	GBT	0.36	0.04	0.31	0.36	0.42	[**4.2, 7.2**
	kNN	0.61	0.05	0.56	0.58	0.73	[-0.1, 0.6
	Ridge	0.62	0.06	0.55	0.62	0.72	[-0.5, 0.2
MOST	**MOST**	0.32	0.02	0.29	0.32	0.34	-
	GBT	0.17	0.02	0.16	0.17	0.22	[**4.2, 11.6**
	kNN	0.32	0.03	0.27	0.32	0.36	[-0.9, 0.2
	Ridge	0.27	0.02	0.24	0.28	0.30	[**1.3, 3.2**

Gradient Boosted Trees Metrics

Table 4.6: Descriptive statistics of all the metrics obtained by the gradient boosted trees estimator in the nested CV procedure, for both datasets.

Dataset	Metric	Mean	Std.	Min.	50%	Max.
Full	R^2	0.84	0.01	0.82	0.84	0.85
	MAE	0.35	0.02	0.31	0.35	0.38
	MAPE	105.34	77.54	60.41	69.33	296.47
	MSE	0.36	0.04	0.31	0.36	0.42
	Train MSE	0.32	0.03	0.28	0.33	0.36
	mAE	0.19	0.01	0.16	0.19	0.20
	mAPE	16.74	0.90	15.18	16.68	18.23
MOST	R^2	0.79	0.02	0.73	0.79	0.80
	MAE	0.26	0.02	0.24	0.26	0.30
	MAPE	87.62	57.80	50.59	63.66	228.67
	MSE	0.17	0.02	0.16	0.17	0.22
	Train MSE	0.15	0.03	0.12	0.15	0.22
	mAE	0.16	0.02	0.14	0.15	0.20
	mAPE	15.10	1.10	13.87	14.78	17.49

Features Importance

Table 4.8: Effect sizes comparing the MSE scores of each feature set with and without trend. Using the trend does not improve the performance, and sometimes not using the trend is better.

Treatment	**F1**	**F2**	**F3**	**F4**	**F5**
Not Trend	[-0.2, 0.1	[-0.2, 0.0	**[0.2, 0.5**	[-0.3, 0.1	**[0.0, 0.3**
Trend	[-0.5, -0.1	[-0.3, -0.0	[-1.0, -0.5	[-0.4, -0.1	[-0.6, -0.3

Table 4.9: Effect sizes comparing the MSE scores of all pairs of feature sets without trend. The control is on rows, and the treatment is on columns. The only feature sets that do not bring an improvement are F2 over F1 and F4 over F3.

Features	**F1**	**F2**	**F3**	**F4**	**F5**
F1	-	[-0.2, 0.2	**[0.6, 1.0**	**[0.6, 1.0**	**[1.4, 2.0**
F2	[-0.4, -0.1	-	**[0.5, 0.8**	**[0.4, 0.8**	**[1.3, 1.8**
F3	[-1.3, -0.9	[-1.4, -0.9	-	[-0.6, -0.2	**[0.3, 0.7**
F4	[-1.2, -0.8	[-1.2, -0.8	[-0.2, 0.2	-	**[0.5, 0.9**
F5	[-1.9, -1.4	[-2.0, -1.5	[-0.8, -0.5	[-1.2, -0.8	-

Features Importance – Raw Results

Table 4.7: Descriptive statistics of the MSE used to produce figure 4.7, table 4.8 and table 4.9.

Features	F1		F2		F3		F4		F5	
Trend	N	Y	N	Y	N	Y	N	Y	N	Y
Mean	0.19	0.19	0.19	0.19	0.17	0.18	0.17	0.17	0.16	0.16
Std.	0.02	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.02
Min.	0.16	0.16	0.15	0.15	0.14	0.14	0.15	0.14	0.13	0.13
50%	0.18	0.19	0.18	0.18	0.17	0.18	0.17	0.17	0.16	0.16
Max.	0.22	0.22	0.22	0.24	0.22	0.22	0.20	0.21	0.18	0.22

Ensemble of Linear Models

Table C.2: Distribution of the hyper-parameters used in the random search and optimal value found.

	Hyper-P.	Distribution	Values	Optimal
Base Est.	Polynomial Degree	Uniform	1 or 2	2
	Polynomial Interactions	Uniform	Yes or No	Yes
	PCA Target Dimension	Uniform	2 to 19	19
	Ridge Regularization	Log_{10}-uniform	10^{-7} to 10^{1}	2×10^{-5}
Bag Est.	Number of B.E.	Uniform	5 to 100	13
	Samples per B.E.	Uniform	5% to 25%	11%
	Features per E.E.	Uniform	4 to 20	20

Validation MSE | 0.187(0.022)

Hyper-Parameter Optimization

Table C.3: Distribution of the hyper-parameters used and best values found by the two optimization procedures. NoT is the number of trees, LR the learning rate, MCW the minimum child weight, γ the minimum decrease in loss for adding a new split, and NoF the number of features.

Hyper-P.	**Distribution**	**Range**	**Best (Hyperband)**	**Best (GPyOpt)**
NoT	Log_{10}-Uniform	10^1 to 10^3	1000	1000
LR	Log_{10}-Uniform	10^{-4} to 10^0	8.86×10^{-3}	4.34×10^{-2}
Max. Depth	Uniform	2 to 12	12	12
MCW	Uniform	1 to 10	9	10
Subsample	Uniform	10% to 100%	64.6%	49.3%
γ	Uniform	0 to 1	0.194	0
α (L1 reg.)	Log_{10}-Uniform	10^{-20} to 10^0	1.87×10^{-11}	1.70×10^{-7}
λ (L2 reg.)	Log_{10}-Uniform	10^{-20} to 10^0	2.46×10^{-10}	5.96×10^{-5}
NoF	Uniform	10% to 100%	22.7%	12.5%
Loss	Uniform	MSE or MAE	MAE	MAE
		Test MSE	0.1676	0.1670