Locally Optimized Random Forests, a Solution to Forecasting Severe Hurricane Power Outages

Tim Coleman

University of Pittsburgh Department of Statistics

June 5, 2020

Outline

Acknowledgements & Co-Authors

- This is work done during my studies at Los Alamos National Laboratory, under my mentors Kim Kaufeld and Mary Frances Dorn
- Work continued on this project at the University of Pittsburgh with my adviser, Dr. Lucas Mentch.

Los Alamos NATIONAL LABORATORY EST.1943

University of Pittsburgh

Introduction

Introduction

- High intensity hurricanes lead to severe power outages, but there are limited historical records of these
- Hurricane Irma was the strongest storm ever seen in the Gulf of Mexico (Cangialosi et al., 2018)
- Weather forecasts performed well, but human impact forecasts underestimate damage

Figure 1: Hurricane Irma - one big storm

Forecasting Power Outages

- For county i in a given hurricane, the ORNL (2018) EAGLE-I database provides 15 minute outage counts $\{O_{i,t}\}$, which we summarise using the "maximum 2-hour sustained outages", given by

$$Y_i = \log_{10}\left(\max_t \min_k \{O_{i,k} : k \in [t, t+8)\}\right).$$

- The fundamental task here is to forecast Y_i accurately, and if possible, provide prediction intervals.

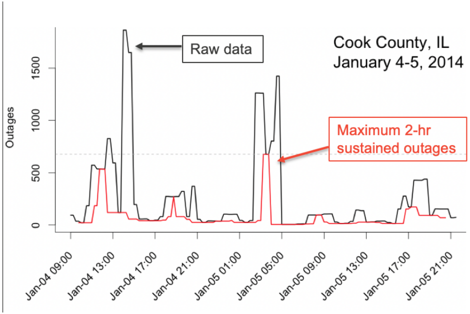


Figure 2: Time series $O_{i,k}$ with Y_i highligted in red.

Forecasting Power Outages

- Much of the existing work is local in nature, uses local power grid information to train random forests and other ML methods (Wanik et al., 2015; Liu et al., 2005; Nateghi et al., 2014).
- This information isn't standardized/regularly collected nationally, so any national forecasting method cannot rely on this information.

Forecasting Power Outages

- Much of the existing work is local in nature, uses local power grid information to train random forests and other ML methods (Wanik et al., 2015; Liu et al., 2005; Nateghi et al., 2014).
- This information isn't standardized/regularly collected nationally, so any national forecasting method cannot rely on this information.
- We use three sources of information:

Forecasting Power Outages

- Much of the existing work is local in nature, uses local power grid information to train random forests and other ML methods (Wanik et al., 2015; Liu et al., 2005; Nateghi et al., 2014).
- This information isn't standardized/regularly collected nationally, so any national forecasting method cannot rely on this information.
- We use three sources of information:
  1. Storm characteristics, such as wind speed, precipitation
  2. Environmental characteristics, such as drought index, land cover
  3. Socio-economic information about each county

Forecasting Power Outages

- Much of the existing work is local in nature, uses local power grid information to train random forests and other ML methods (Wanik et al., 2015; Liu et al., 2005; Nateghi et al., 2014).
- This information isn't standardized/regularly collected nationally, so any national forecasting method cannot rely on this information.
- We use three sources of information:
  1. Storm characteristics, such as wind speed, precipitation
  2. Environmental characteristics, such as drought index, land cover
  3. Socio-economic information about each county
- For each storm S, we get natural training/test splits

$$\mathcal{D}_S = \{(\boldsymbol{X}_i, Y_i) : \text{Observation } i \text{ is not from storm } S\}$$
$$\mathcal{T}_S = \{(\boldsymbol{X}_i, Y_i) : \text{Observation } i \text{ is from storm } S\}$$

Random Forest Definitions

- Random forest predictions are weighted averages of the training responses, i.e.

$$RF(\boldsymbol{x}; \mathcal{D}) = \sum_{i=1}^{n} \underbrace{\mathbb{E}_{\xi}\left[\frac{I(\boldsymbol{X}_i \in A_{\xi}^{*}(\boldsymbol{x}))}{\sum_{j=1}^{n} I(\boldsymbol{X}_j \in A_{\xi}^{*}(\boldsymbol{x}))}\right]}_{r_i(\boldsymbol{x};\mathcal{D})} Y_i.$$

- Meinshausen (2006) noted that the weights used for the empirical mean could be used instead in a local empirical distribution function, namely

$$\hat{F}(y|\boldsymbol{X} = \boldsymbol{x}) = \sum_{i=1}^{n} r_i(\boldsymbol{x}; \mathcal{D}) I(Y_i \leq y)$$

where the goal is to estimate $F(y|\boldsymbol{X} = \boldsymbol{x}) = P(Y \leq y|\boldsymbol{X} = \boldsymbol{x})$.

Problems with this Approach

- Below shows results of a RF with hyper-parameters selected by cross-validation for each storm.

Storm	`mtry`	`nodesize`	MAE	RMSE	Covg	IntWidth
Matthew-2016	50	5	0.6247	0.7840	0.9580	3.3696
Nate-2017	40	5	0.6694	0.8069	0.9524	3.2204
Harvey-2017	50	5	0.7471	0.9020	0.8923	3.0574
Arthur-2014	45	5	0.8462	1.0325	**0.7728**	2.7799
Sandy-2012	40	10	0.9797	1.2199	**0.6562**	2.8495
Irma-2017	45	5	1.1871	1.4056	**0.5350**	3.0051

Problems with this Approach

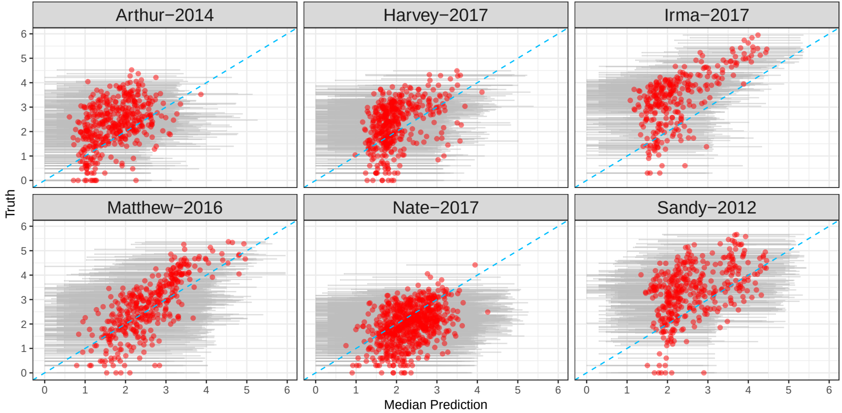
- Below shows results of a RF with hyper-parameters selected by cross-validation for each storm.

Storm	`mtry`	`nodesize`	MAE	RMSE	Covg	IntWidth
Matthew-2016	50	5	0.6247	0.7840	0.9580	3.3696
Nate-2017	40	5	0.6694	0.8069	0.9524	3.2204
Harvey-2017	50	5	0.7471	0.9020	0.8923	3.0574
Arthur-2014	45	5	0.8462	1.0325	**0.7728**	2.7799
Sandy-2012	40	10	0.9797	1.2199	**0.6562**	2.8495
Irma-2017	45	5	1.1871	1.4056	**0.5350**	3.0051

- Coverage is not close to nominal level for three of six storms - and error metrics are worst on most severe storms.

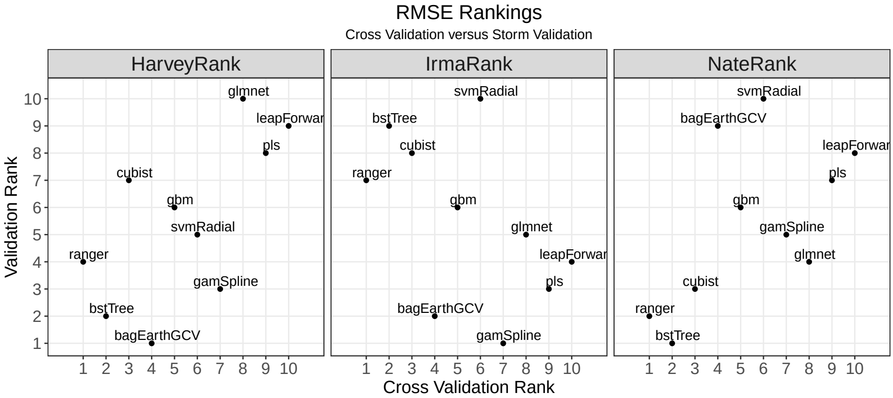
Problems with this Approach

- The regression forests also fail the eye test when plotting fitted vs predicted values.



Model Selection Fails

- Unfortunately, standard cross validation fails, especially for severe storms.



- The CV rankings are not predictive of the validation rankings for Irma.

A Statistical Solution

Violations in Assumptions

- Essentially, the historical record (training data) and incoming storms (validation data) have different distributions, P_1 and P_2, so that

$$\mathbb{E}_{(\boldsymbol{X},Y)\sim P_1} L(\hat{f}(\boldsymbol{X}), Y) \neq \mathbb{E}_{(\boldsymbol{X},Y)\sim P_2} L(\hat{f}(\boldsymbol{X}), Y).$$

  where $\hat{f}$ is an estimated regression function and $L(\cdot)$ is a loss function
- In general, we will observe the covariates associated with a particular storm before the impacts on the power grid are realized. Thus, $\boldsymbol{X}_{\text{test}}$ is available at training time.

Modeling the Violation

- Thus, if we assume that there exists a common collection of conditional distributions $P(Y|\boldsymbol{X})$ between P_1 and P_2, we have a hope of making our training procedure adapt to the distributional shift.
- Formally, we assume that

$$\begin{aligned} P_1(\boldsymbol{X}, Y) &= P(Y|\boldsymbol{X})P_1^*(\boldsymbol{X}) \\ P_2(\boldsymbol{X}, Y) &= P(Y|\boldsymbol{X})P_2^*(\boldsymbol{X}) \end{aligned} \tag{1}$$

- Equation 1 is commonly referred to as the *covariate shift* model (Shimodaira, 2000; Sugiyama and Müller, 2005; Sugiyama et al., 2007).

Estimating w

- Kanamori et al. (2009) developed a means of estimating $w(\boldsymbol{x})$ via kernel regression - assume that $w(\boldsymbol{x}) = \sum_{l=1}^{n_{\text{test}}} \alpha_l K_\sigma(\boldsymbol{x}, \boldsymbol{x}_l)$ where $K_\sigma(\cdot, \cdot)$ is a Gaussian kernel with bandwidth σ, and $\boldsymbol{x}_l$ is a point in the test set.
- The bandwidth σ is fit through cross-validation, and then regularized regression is performed to estimate α_l.
- Other methods include minimizing the KL divergence between P_{test} and $\hat{P}_{\text{test}} = P_{\text{train}}\hat{w}$ (Sugiyama et al., 2008).
- We additionally regularize the weights, by setting $w_\gamma(\boldsymbol{x}) = w(\boldsymbol{x})^\gamma$ for some $\gamma \in (0, 1)$.

Using Weights in a Random Forest

- Regression forest makes splits by recursively minimizing empirical variances, and makes predictions according to localized empirical expectations.
- Let A_0 be a parent node, and let $A_L = \{\boldsymbol{X} \in A : \; X^{(j)} \leq z\}$, $A_R = A \setminus A_L$. Then, the CART criterion for a candidate split dimension j and point z is given as

$$\mathrm{CART}(j,z) = \frac{1}{N_n(A)} \sum_{i=1}^{n} (Y_i - \bar{Y}_A)^2 I(\boldsymbol{X}_i \in A) - \frac{1}{N_n(A)} \sum_{i=1}^{n} \left(Y_i - \bar{Y}_{A_L} I(X_i^{(j)} \leq z) - \bar{Y}_{A_R} I(X_i^{(j)} > z)\right)^2 I(\boldsymbol{X}_i \in A) \quad (2)$$

where $\bar{Y}_A = \frac{1}{N_n(A)} \sum Y_i I(\boldsymbol{X}_i \in A)$.

Using Weights in a Random Forest

- We use a weighted CART criterion, given by

$$\mathrm{CART}^w(j,z) = \frac{1}{\sum_{\mathbf{X}_j \in A} w_j} \sum_{i=1}^{n} w_i (Y_i - \tilde{Y}_A)^2 I(\mathbf{X}_i \in A) - \frac{1}{\sum_{\mathbf{X}_j \in A} w_j} \sum_{i=1}^{n} w_i \left(Y_i - \tilde{Y}_{A_L} I(X_i^{(j)} < z) - \tilde{Y}_{A_R} I(X_i^{(j)} \geq z)\right)^2 I(\mathbf{X}_i \in A) \quad (3)$$

and $\tilde{Y}_A = \frac{1}{\sum_{\mathbf{X}_j \in A} w_j} \sum w_i Y_i I(\mathbf{X}_i \in A)$.

- Iteratively minimize this new quantity and then use the weighted means as predictions, generating trees $\{T_w(\mathbf{X}_i, \xi_k)\}_{k=1}^{B}$.
- Similar in spirit to *case specific* random forests, proposed by Xu et al. (2016), but they rely on a weighted bootstrap, and only works for a single test point.

Weighted Random Forests

- An advantage of bagged models is the ability to use the *out-of-bag* error as an estimate of generalization error - no need to perform exhaustive data splitting
- For each $(\boldsymbol{X}_i, Y_i)$ in the training data, let $\mathcal{B}_i \subset \{1, \ldots, B\}$ be the indices where $(\boldsymbol{X}_i, Y_i)$ was not included in the resample. Then,

$$\mathrm{OOB}_{m,B} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{|\mathcal{B}_i|} \sum_{k \in \mathcal{B}_i} T(\boldsymbol{X}_i; \xi_k, \mathcal{D}_{-i}) - Y_i \right)^2 .$$

- Because $\lim_{B \to \infty} |\mathcal{B}_i| = \infty$ a.s., the infinite forest version of $\mathrm{OOB}_{m,B}$ is almost surely equal to the leave one out CV (LOOCV) error, which is defined as

$$\mathrm{LOOCV}_{RF} = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{E}_{\xi} T(\boldsymbol{X}_i; \xi, \mathcal{D}_{-i}) - Y_i \right)^2 .$$

Weighted Loss Estimates

- We can compute a weighted version, for the weighted tree model

$$\mathrm{OOB}^{w}_{m,B} = \frac{1}{\sum_{j=1}^{n} w_j} \sum_{i=1}^{n} w_i \left(\frac{1}{|\mathcal{B}_i|} \sum_{k \in \mathcal{B}_i} T_{w}(\mathbf{X}_i; \xi_k, \mathcal{D}_{-i}) - Y_i \right)^2 .$$

- Sugiyama et al. (2007) showed that the importance-weighted LOOCV is approximately unbiased for the generalization error under P_2.
- Thus, we see that

$$\mathbb{E}_{P_1} OOB^{w}_{m,B} = \mathbb{E}_{P_1} LOOCV^{w}_{RF} = \mathbb{E}_{\mathcal{D}_{n-1},(\mathbf{X},Y)\sim P_2} \left(\mathbb{E}_{\xi}(\mathbf{X}; \mathcal{D}_{n-1}) - Y\right)^2 ,$$

and so we can use the weighted out of bag error as a measure of model performance.

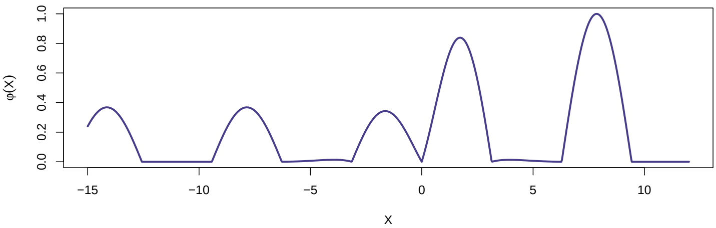
Simulations

A Toy Regression Example

- Consider the univariate data model

$$Y|X \sim \mathcal{N}\left(\varphi(X), 0.5\right)$$
$$\varphi(X) = \max\left\{\frac{e^X}{1+e^X}\sin(X), \frac{e^{-X}}{1+e^{-X}}\sin(-X)\right\}$$

- $\varphi(X)$ has a lot of "local" features - intrinsically difficult to extrapolate here



A Toy Regression Example

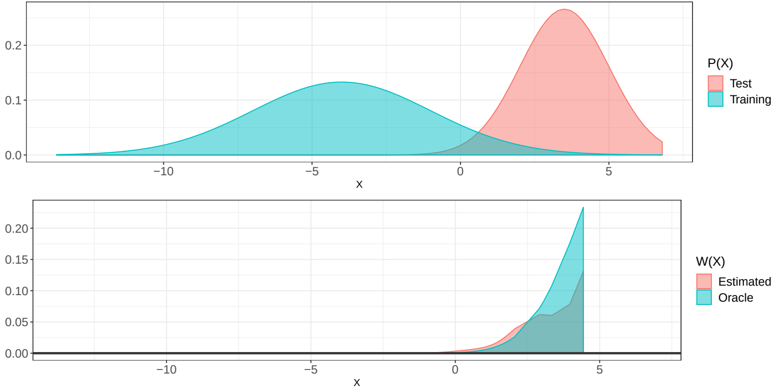
- We simulate a covariate shift by introducing two different training and testing distributions for X

$$P_{\text{train}} = \mathcal{N}\left(-4, 3.5^2\right)$$
$$P_{\text{test}} = \mathcal{N}\left(3.5, 1.5^2\right)$$

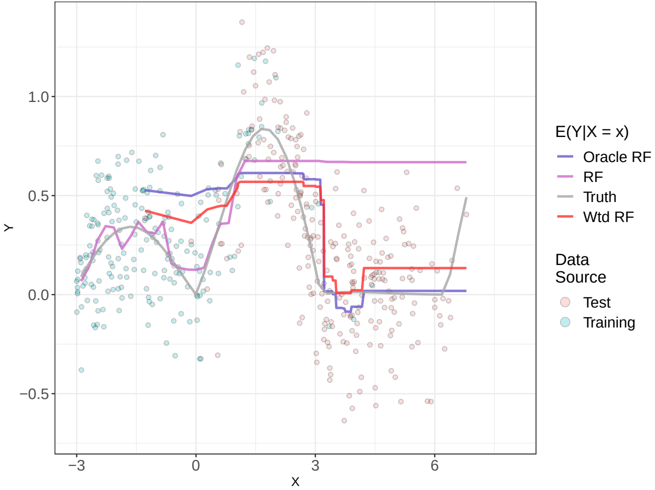
- Training distribution is dispersed, but test data is centered around a particular region
- We compare 3 models:
  - A random forest from the `ranger` package as a baseline
  - A weighted forest, where the weights are learned with the method of Kanamori et al. (2009)
  - A weighted forest, where the weights are the oracle weights - i.e. $w(\boldsymbol{x}) \propto \frac{\phi\left(\frac{x-3.5}{1.5}\right)}{\phi\left(\frac{x+4}{3.5}\right)}$

Weights Estimation

- Outputs look like



Regression Estimation



- `ranger` model fails to pick up on the local behavior around which P_{test} is centered
- Error metrics for this run are shown below

Model	RMSE
`ranger`	0.3407
Weighted	0.0976
Oracle Weights	0.0939

A More Severe Shift

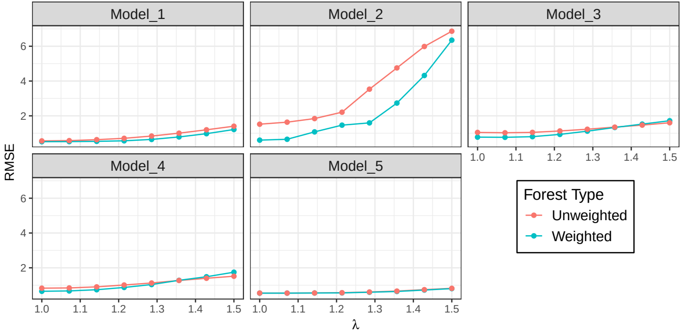
- Now we consider a collection of regression functions that have medium dimensional ($p = 31$) inputs:

Model #	Data Generating Model
1	$Y = 5X^{(1)} + \epsilon$
2	$Y = 5\sin(\pi X^{(1)}) + \epsilon$
3	$Y = \underbrace{10\sin(\pi X^{(1)}X^{(2)}) + 20(X^{(3)} - 0.5)^2 + 10X^{(4)} + 5X^{(5)} + \epsilon}_{\text{MARS model (Friedman, 1991)}}$
4	$Y = 5e^{2\sqrt{X^{(1)}X^{(2)}}+X^{(6)}} + \epsilon$
5	$Y = 5\sum_{j=1}^{5}\left(X^{(j)}\right)^2 + \epsilon$

- In each case, ϵ is mean 0, Gaussian noise with $\mathbb{E}(\epsilon^2) = 0.25$.
- We set $\boldsymbol{\alpha}_1 = [\lambda^1, ..., \lambda^6]$ (the training distribution) and $\boldsymbol{\alpha}_2 = [\lambda^6, ..., \lambda^1]$ (the test distribution), so that λ controls the magnitude of the covariate shift.

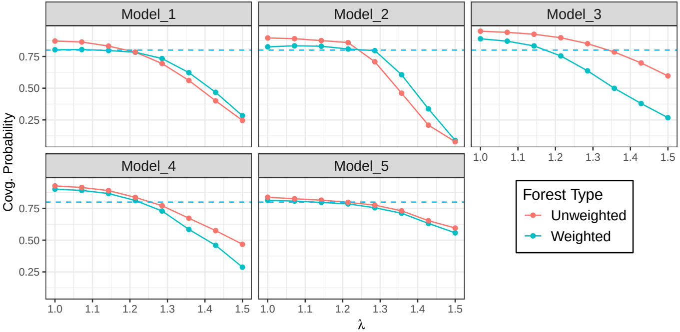
RMSE Results

#	Data Generating Model
1	$Y = 5X^{(1)} + \epsilon$
2	$Y = 5\sin(\pi X^{(1)}) + \epsilon$
3	MARS model
4	$Y = 5e^{2\sqrt{X^{(1)}X^{(2)}}+X^{(6)}} + \epsilon$
5	$Y = 5\sum_{j=1}^{5}\left(X^{(j)}\right)^2 + \epsilon$



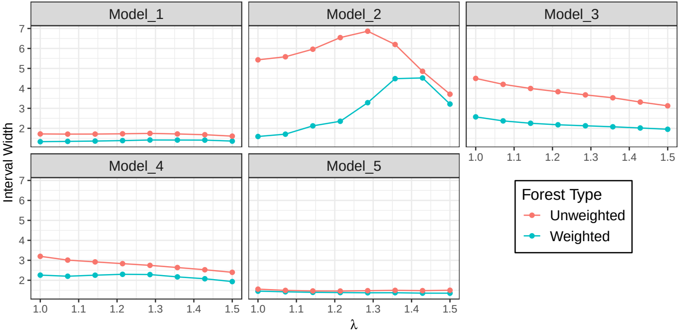
Coverage Percentage Results

#	Data Generating Model
1	$Y = 5X^{(1)} + \epsilon$
2	$Y = 5\sin(\pi X^{(1)}) + \epsilon$
3	MARS model
4	$Y = 5e^{2\sqrt{X^{(1)}X^{(2)}}+X^{(6)}} + \epsilon$
5	$Y = 5\sum_{j=1}^{5}\left(X^{(j)}\right)^2 + \epsilon$



Interval Width

#	Data Generating Model
1	$Y = 5X^{(1)} + \epsilon$
2	$Y = 5\sin(\pi X^{(1)}) + \epsilon$
3	MARS model
4	$Y = 5e^{2\sqrt{X^{(1)}X^{(2)}}+X^{(6)}} + \epsilon$
5	$Y = 5\sum_{j=1}^{5}\left(X^{(j)}\right)^2 + \epsilon$



Application to Hurricane Outage Forecasting

Returning to the Original Problem

- Recall the goal is to generate more accurate forecasts for the hurricane outages.
- Now we apply the weighted procedure for the six storms presented in the introduction.
- For comparison, we also again train an unweighted random forest. For the unweighted forest, we tune over the `mtry` parameter, and for the weighted forest we tune over `mtry` and the γ weight regularizer.
- To assess performance across regression/quantile regression, we introduce the following "score"

$$\text{Score} = \left(\frac{1}{\text{MAE}} + \frac{1}{\text{RMSE}} + \frac{4}{\text{IntWidth}} \right) \frac{\text{Covg}}{1 - \alpha} \tag{4}$$

Hurricane Results

Storm	Model	RMSE	MAE	Covg	Interval Width	Score
Harvey-2017	Weighted	0.9376	0.7307	0.8038	2.5425	**3.5801**
Harvey-2017	Unweighted	0.9089	0.7541	0.7847	2.4619	3.5321
Irma-2017	Weighted	1.3855	1.1651	0.3916	2.3906	**1.4155**
Irma-2017	Unweighted	1.4052	1.1775	0.3706	2.4064	1.3273
Sandy-2012	Weighted	1.2320	1.0159	0.5677	2.2663	**2.2462**
Sandy-2012	Unweighted	1.2148	0.9933	0.5521	2.1930	2.2414
Nate-2017	Weighted	0.7859	0.6839	0.8860	2.7199	4.1400
Nate-2017	Unweighted	0.8133	0.6682	0.8687	2.4809	**4.1874**
Matthew-2016	Weighted	0.7994	0.6292	0.8635	2.5532	4.2282
Matthew-2016	Unweighted	0.7848	0.6275	0.8950	2.6813	**4.3355**
Arthur-2014	Weighted	0.9966	0.7942	0.7213	2.4303	**3.1325**
Arthur-2014	Unweighted	1.0637	0.8580	0.6698	2.2712	2.8776

Table 1: Model performance by storm, with weighted and unweighted storms fitted. Bolded values represent which model attained the higher "Score", as defined earlier.

Tuning the Models

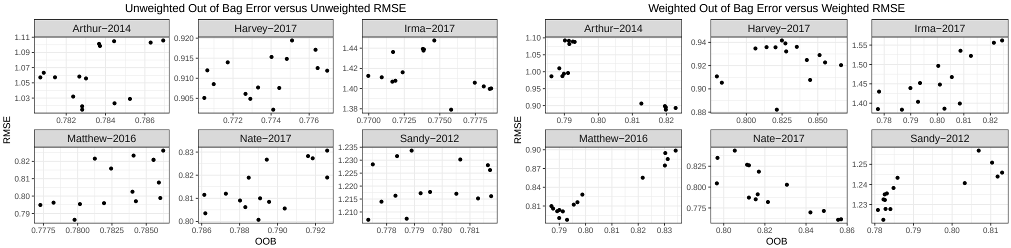


Figure 3: Out of bag error versus holdout RMSE. **Left:** Results for the unweighted forest (repeated five times per `mtry` value, for visual consistency with the right figure). **Right:** Results for the weighted forest.

Challenges in High Dimensions

- In simulations, we show definite model improvements in ”medium” sized shifts.
- In our motivating problem, we demonstrate modest improvements on several storms, including severe storms such as Irma. However, the improvements do not reach the threshold of no shift.
- It’s possible that the covariate shift assumption is violated, that is, $P(Y|X)$ is also different for particularly extreme storms.

Thanks!

References I

Cangialosi, J. P., Latto, A. S., and Berg, R. (2018). Hurricane irma. In *National Hurricane Center Tropical Cyclone Report*.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67.

Kanamori, T., Hido, S., and Sugiyama, M. (2009). A least-squares approach to direct importance estimation. *Journal of Machine Learning Research*, 10(Jul):1391–1445.

Liu, H., Davidson, R. A., Rosowsky, D. V., and Stedinger, J. R. (2005). Negative binomial regression of electric power outages in hurricanes. *Journal of infrastructure systems*, 11(4):258–267.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun):983–999.

Nateghi, R., Guikema, S. D., and Quiring, S. M. (2014). Forecasting hurricane-induced power outage durations. *Natural hazards*, 74(3):1795–1811.

ORNL (2018). Eagle-i data. data retrieved from DoE website, `https://eagle-i.doe.gov/`.

Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244.

Sugiyama, M., Krauledat, M., and MÃžller, K.-R. (2007). Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8(May):985–1005.

Sugiyama, M. and Müller, K.-R. (2005). Input-dependent estimation of generalization error under covariate shift. *Statistics & Decisions*, 23(4/2005):249–279.

Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P. V., and Kawanabe, M. (2008). Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in neural information processing systems*, pages 1433–1440.

Wanik, D., Anagnostou, E., Hartman, B., Frediani, M., and Astitha, M. (2015). Storm outage modeling for an electric distribution network in northeastern usa. *Natural Hazards*, 79(2):1359–1384.

Xu, R., Nettleton, D., and Nordman, D. J. (2016). Case-specific random forests. *Journal of Computational and Graphical Statistics*, 25(1):49–65.