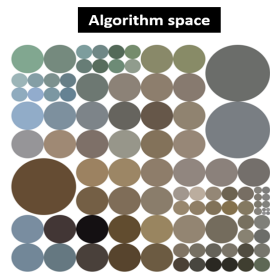
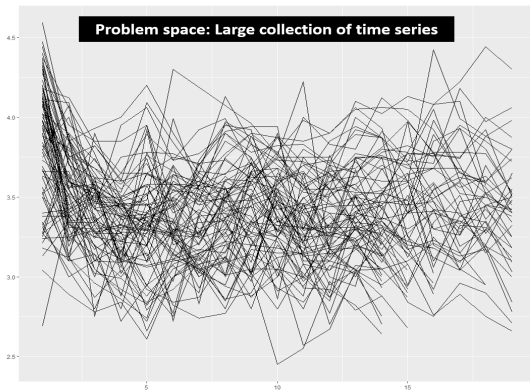


Peeking inside FFORMS: Feature-based FORecast-Model Selection

Thiyanga Talagala,
Rob J Hyndman, George Athanasopoulos

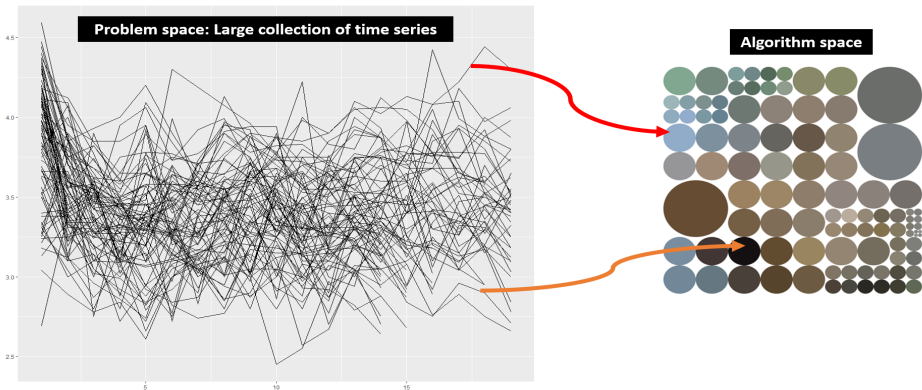
18 June 2019

Big picture



- What algorithm is likely to perform best?

Big picture



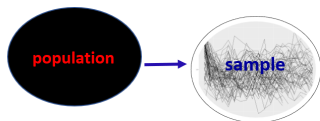
- What algorithm is likely to perform best?
- Algorithm selection problem, John Rice (1976)

FFORMS: Feature-based FOfRecast Model Selection

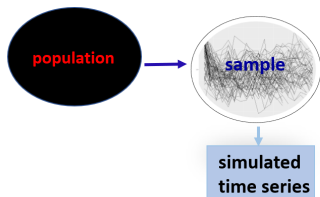


population

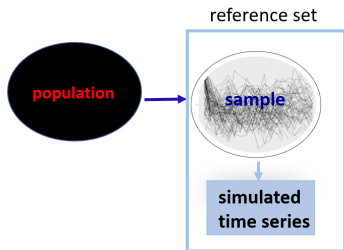
FFORMS: Feature-based FOfecast Model Selection



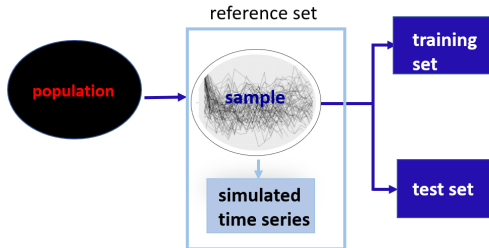
FFORMS: Feature-based FORecast Model Selection



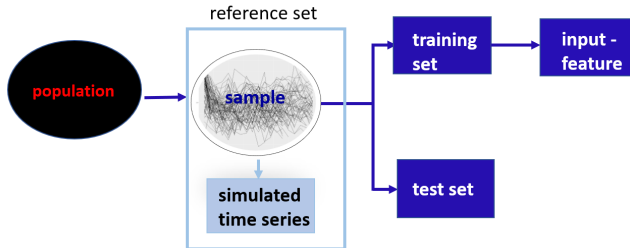
FFORMS: Feature-based FOfecast Model Selection



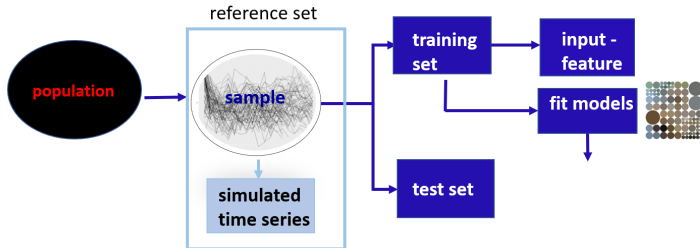
FFORMS: Feature-based FORecast Model Selection



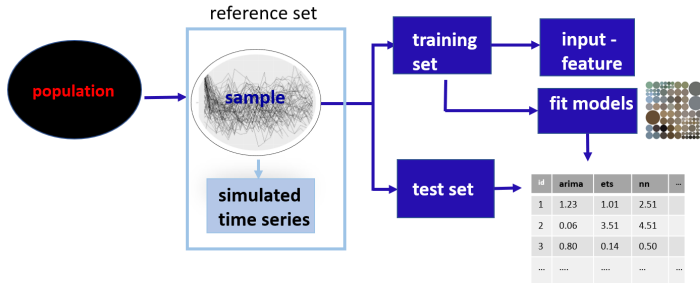
FFORMS: Feature-based FOfecast Model Selection



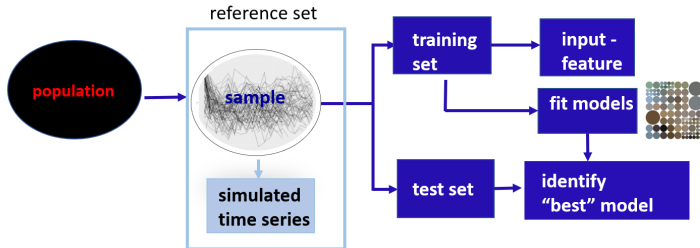
FFORMS: Feature-based FOREcast Model Selection



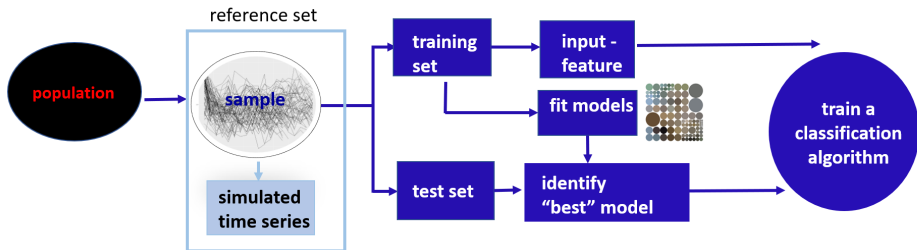
FFORMS: Feature-based FOREcast Model Selection



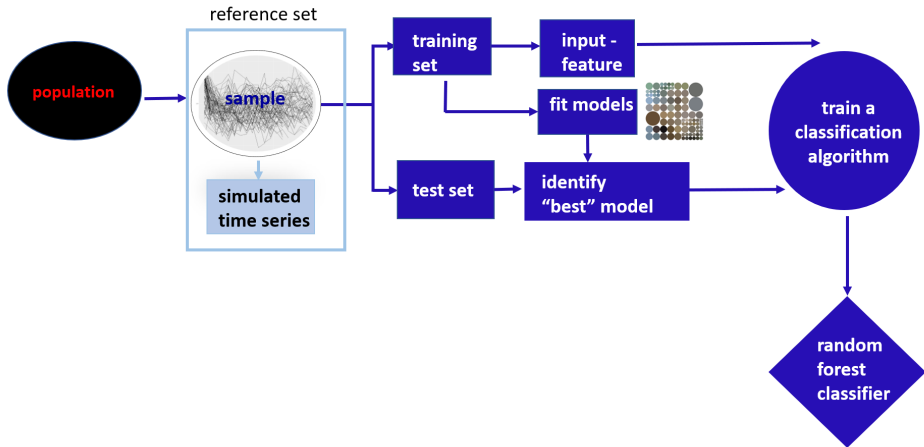
FFORMS: Feature-based FOREcast Model Selection



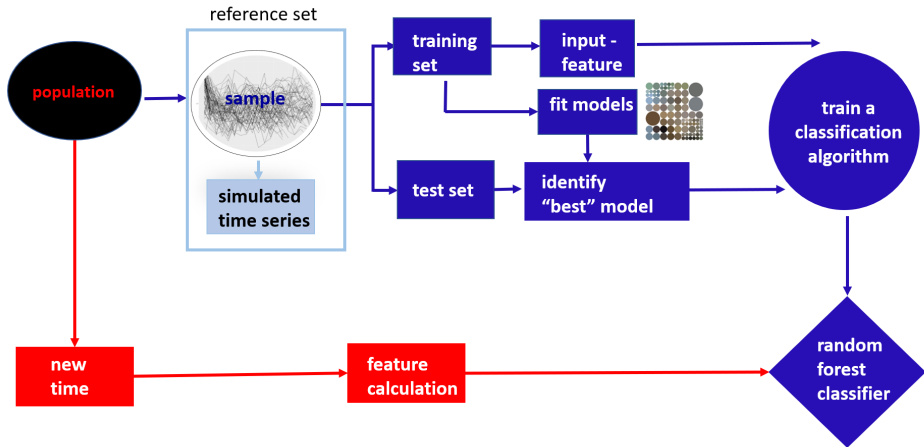
FFORMS: Feature-based FOREcast Model Selection



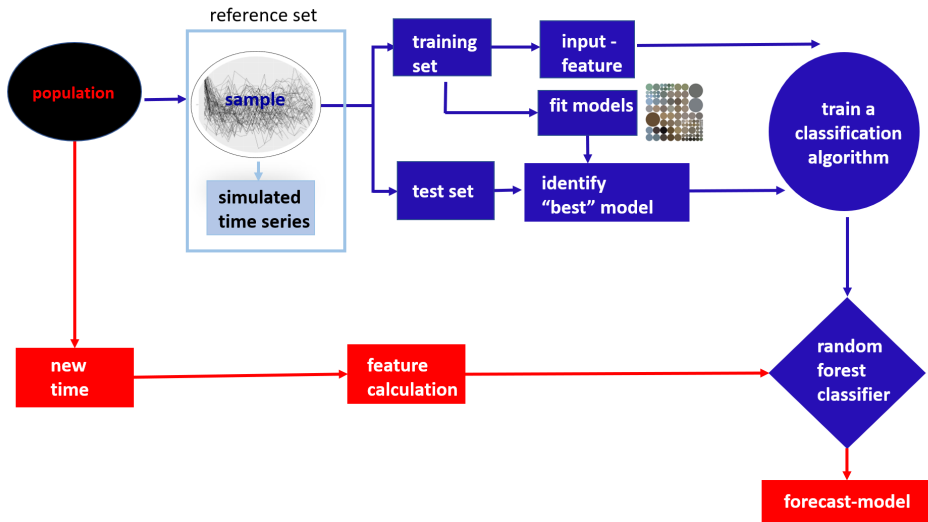
FFORMS: Feature-based FOREcast Model Selection



FFORMS: Feature-based FOREcast Model Selection



FFORMS: Feature-based FOREcast Model Selection



Forecast-models included

- White noise process
- ARMA/AR/MA
- ARIMA
- SARIMA
- Random walk with drift
- Random walk
- Seasonal naive
- TBATS
- neural network forecasts
- Theta method
- STL-AR
- ETS-without trend and seasonal
- ETS-trend
- ETS-damped trend
- ETS-trend and seasonal
- ETS-damped trend and seasonal
- ETS-seasonal
- MSTL-ETS
- MSTL-ARIMA

Time series features

- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- spectral entropy
- Hurst exponent
- nonlinearity
- unit root test statistics
- parameter estimates of Holt's linear trend method
- parameter estimates of Holt-Winters' additive method
- ACF and PACF based features - calculated on raw, differenced, seasonally-differenced series and remainder series.

Results: M4 Competition data

	Yearly	Quarterly	Monthly	Weekly	Daily	Hourly
FFORMS	3.17	1.20	0.98	2.31	3.57	0.84
auto.arima	3.40	1.17	0.93	2.55	-	-
ets	3.44	1.16	0.95	-	-	-
theta	3.37	1.24	0.97	2.64	3.33	1.59
rwd	3.07	1.33	1.18	2.68	3.25	11.45
rw	3.97	1.48	1.21	2.78	3.27	11.60
nn	4.06	1.55	1.14	4.04	3.90	1.09
stlar	-	2.02	1.33	3.15	4.49	1.49
snaive	-	1.66	1.26	2.78	24.46	2.86
tbats	-	1.19	1.05	2.49	3.27	1.30
wn	13.42	6.50	4.11	49.91	38.07	11.68
mstlarima	-	-	-	-	3.84	1.12
mstlets	-	-	-	-	3.73	1.23
combination (mean)	4.09	1.58	1.16	6.96	7.94	3.93
M4-1st	2.98	1.12	0.88	2.36	3.45	0.89
M4-2nd	3.06	1.11	0.89	2.11	3.34	0.81
M4-3rd	3.13	1.23	0.95	2.16	2.64	0.87

Results: M4 Competition data

	Yearly	Quarterly	Monthly	Weekly	Daily	Hourly
FFORMS	3.17	1.20	0.98	2.31	3.57	0.84
auto.arima	3.40	1.17	0.93	2.55	-	-
ets	3.44	1.16	0.95	-	-	-
theta	3.37	1.24	0.97	2.64	3.33	1.59
rwd	3.07	1.33	1.18	2.68	3.25	11.45
rw	3.97	1.48	1.21	2.78	3.27	11.60
nn	4.06	1.55	1.14	4.04	3.90	1.09
stlar	-	2.02	1.33	3.15	4.49	1.49
snaive	-	1.66	1.26	2.78	24.46	2.86
tbats	-	1.19	1.05	2.49	3.27	1.30
wn	13.42	6.50	4.11	49.91	38.07	11.68
mstlarima	-	-	-	-	3.84	1.12
mstlets	-	-	-	-	3.73	1.23
combination (mean)	4.09	1.58	1.16	6.96	7.94	3.93
M4-1st	2.98	1.12	0.88	2.36	3.45	0.89
M4-2nd	3.06	1.11	0.89	2.11	3.34	0.81
M4-3rd	3.13	1.23	0.95	2.16	2.64	0.87

■ Can we trust ML-algorithms if we don't know how it works?



Peeking inside FFORMS!!!

- Which features are the most important?
- Where they get important?
- How they get important?
- When and how these features linked with the prediction outcome?
- When and how strongly each feature interacts with other features?

Global explanation of feature contribution

Overall role of features in the choice of different forecast-model selection.


- contribution to predictive accuracy
 - ▶ Permutation-based variable importance
 - ▶ Mean decrease in Gini coefficient
- causality: change in the value of Y for an increase or decrease in the value of x
 - ▶ Partial dependence plots (Jerome H. Friedman, 2001)
 - ▶ Individual Conditional Expectation (ICE) curves (Goldstein et al., 2015; Zhao and Hastie, 2017)

Partial dependence plots and ICE curves

x1	x2	x3
11	4	5
12	6	7

Partial dependence plots and ICE curves

x1	x2	x3
11	4	5
12	6	7



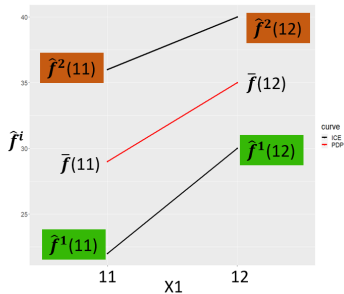
x1	x2	x3
11	4	5
11	6	7
12	4	5
12	6	7

Partial dependence plots and ICE curves

x1	x2	x3		x1	x2	x3		x1	x2	x3	$\hat{f}^i(\mathbf{x1})$	$\overline{\hat{f}(\mathbf{x1})}$
11	4	5	→	11	4	5	→	11	4	5	$\hat{f}^1(11)$	$\frac{\sum \hat{f}^i(11)}{2}$
12	6	7		11	6	7		11	6	7	$\hat{f}^2(11)$	
				12	4	5		12	4	5	$\hat{f}^1(12)$	$\frac{\sum \hat{f}^i(12)}{2}$
				12	6	7		12	6	7	$\hat{f}^2(12)$	

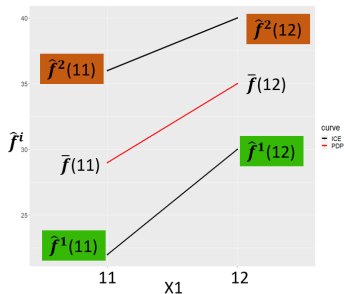
Partial dependence plots and ICE curves

x1	x2	x3		x1	x2	x3		x1	x2	x3	$\hat{f}^i(\mathbf{x1})$	$\overline{\hat{f}(\mathbf{x1})}$
11	4	5	→	11	4	5	→	11	4	5	$\hat{f}^1(11)$	$\frac{\sum \hat{f}^i(11)}{2}$
12	6	7		11	6	7		11	6	7	$\hat{f}^2(11)$	
				12	4	5		12	4	5	$\hat{f}^1(12)$	$\frac{\sum \hat{f}^i(12)}{2}$
				12	6	7		12	6	7	$\hat{f}^2(12)$	

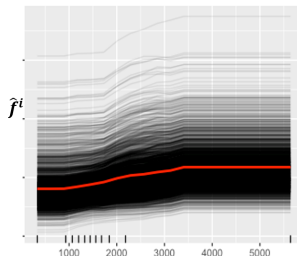


Partial dependence plots and ICE curves

x1	x2	x3		x1	x2	x3		x1	x2	x3	$\hat{f}^i(\mathbf{x1})$	$\overline{\hat{f}(\mathbf{x1})}$
11	4	5	→	11	4	5	→	11	4	5	$\hat{f}^1(11)$	$\frac{\sum \hat{f}^i(11)}{2}$
12	6	7		11	6	7		11	6	7	$\hat{f}^2(11)$	
				12	4	5		12	4	5	$\hat{f}^1(12)$	$\frac{\sum \hat{f}^i(12)}{2}$
				12	6	7		12	6	7	$\hat{f}^2(12)$	

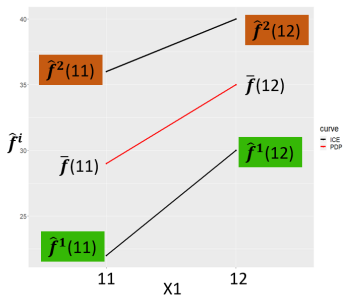


More than 2 observations
and more than 2 levels

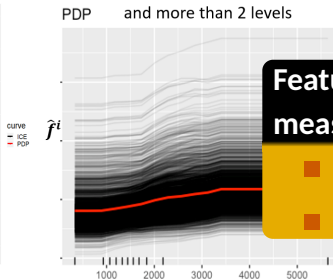


Partial dependence curve and ICE curves

x1	x2	x3		x1	x2	x3		x1	x2	x3	$\hat{f}^i(x1)$	$\overline{f(x1)}$
11	4	5	→	11	4	5	→	11	4	5	$\hat{f}^1(11)$	$\frac{\sum \hat{f}^i(11)}{2}$
12	6	7		11	6	7		11	6	7	$\hat{f}^2(11)$	$\frac{\sum \hat{f}^i(11)}{2}$
				12	4	5		12	4	5	$\hat{f}^1(12)$	$\frac{\sum \hat{f}^i(12)}{2}$
				12	6	7		12	6	7	$\hat{f}^2(12)$	$\frac{\sum \hat{f}^i(12)}{2}$



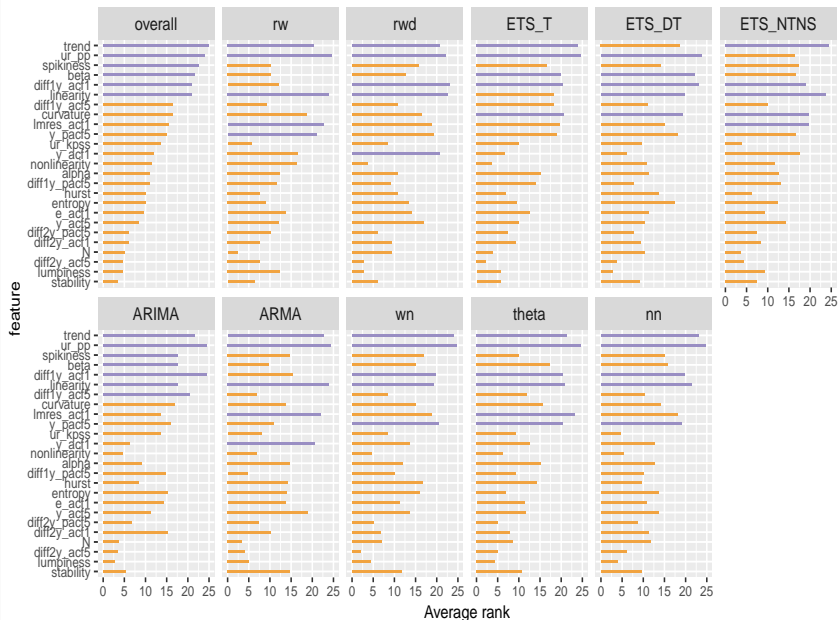
More than 2 observations
and more than 2 levels



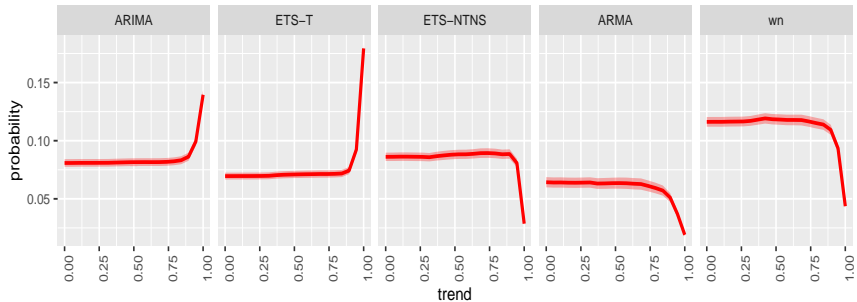
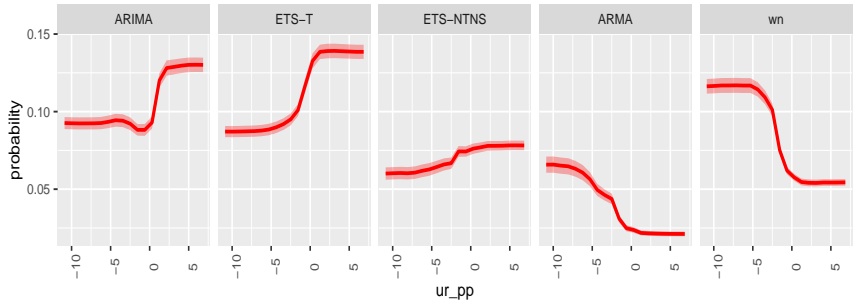
**Feature importance
measures:**

- "flatness" of PD curve
- "flatness" of ICE curves

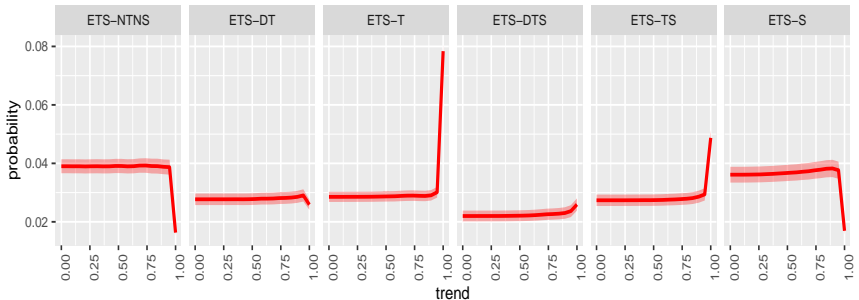
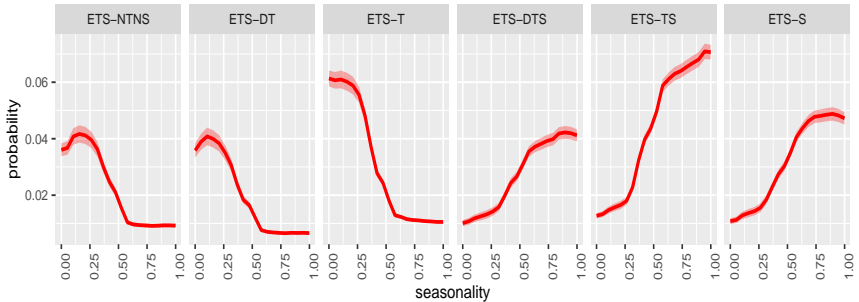
Feature importance plot for yearly data



Partial dependency plots for yearly data

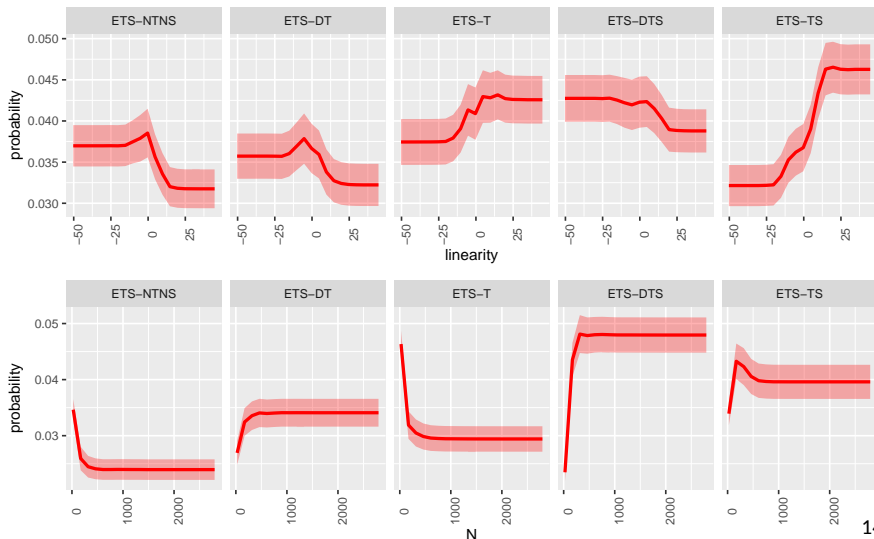


Partial dependency plots for quarterly data



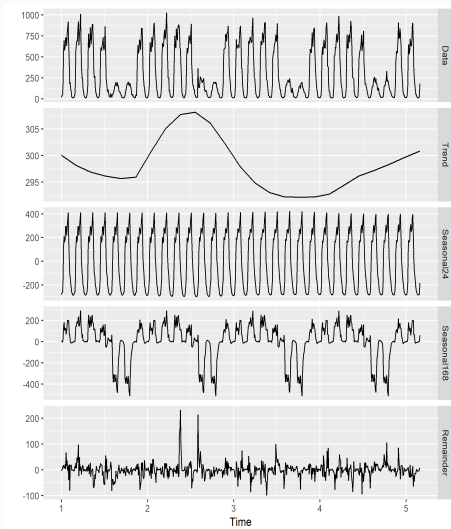
Partial dependency plots for monthly data

linearity: estimated value of β_1 based on $T_t = \beta_0 + \beta_1 \phi_1(t) + \beta_2 \phi_2(t) + \varepsilon_t$



Hourly series

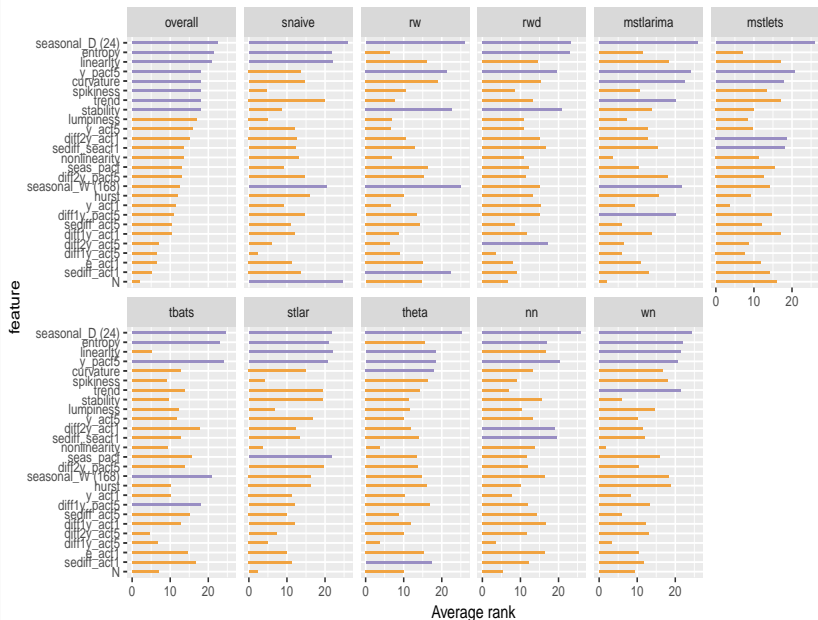
■ multiple seasonality



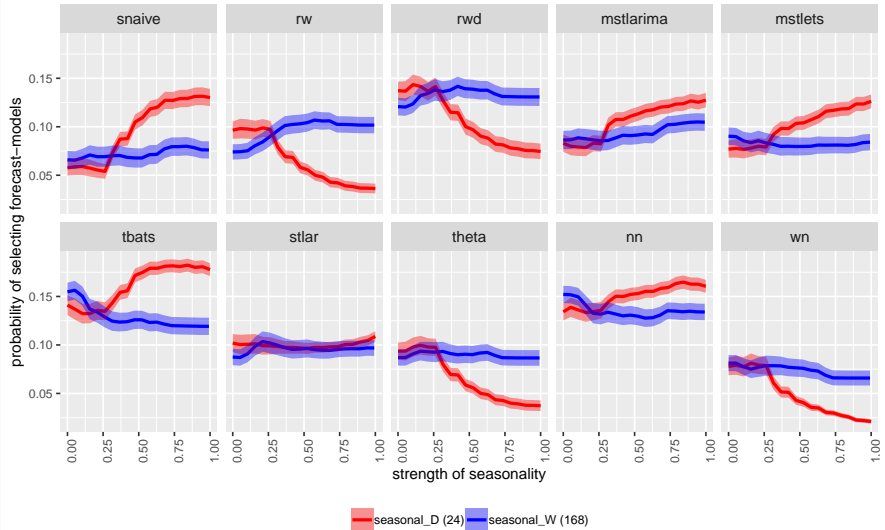
Hourly data

- ▶ daily - 24
- ▶ weekly - 168

Feature importance plot for hourly data



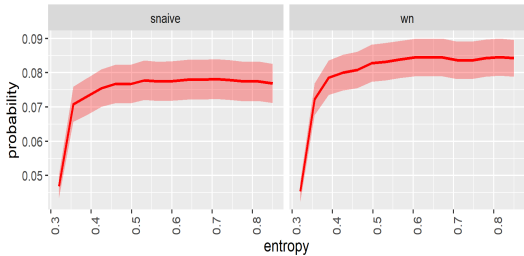
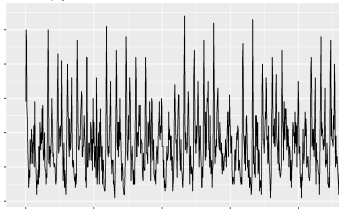
Partial dependency plots for hourly data: Seasonality



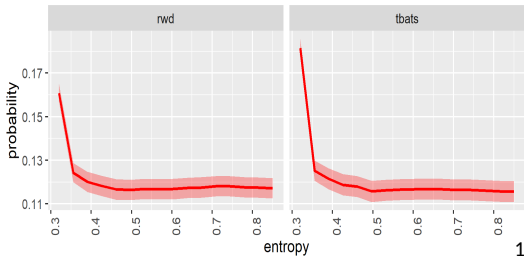
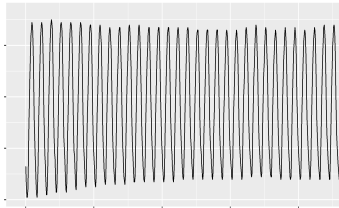
Partial dependency plots for hourly data: entropy

■ forecastability of a time series

entropy: 0.85



entropy: 0.44



- Friedman's H-statistic

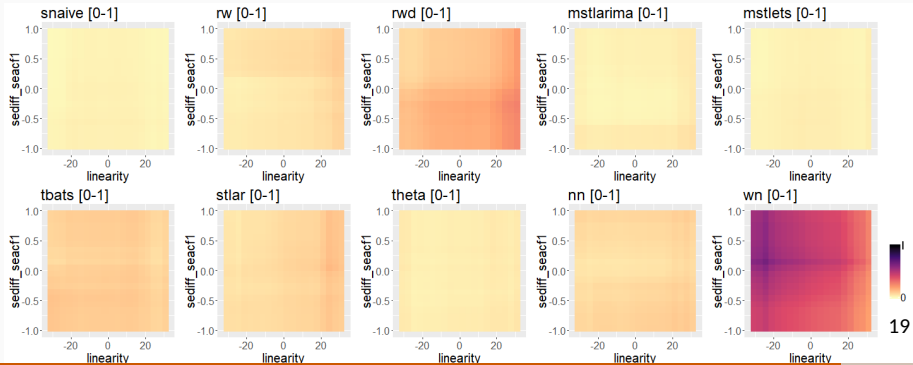
fraction of variance of two-variable partial dependency not captured by sum of the respective individual partial dependencies.

Interaction effect

■ Friedman's H-statistic

fraction of variance of two-variable partial dependency not captured by sum of the respective individual partial dependencies.

Hourly: interaction between linearity and seasonal lag at seasonally-differenced series

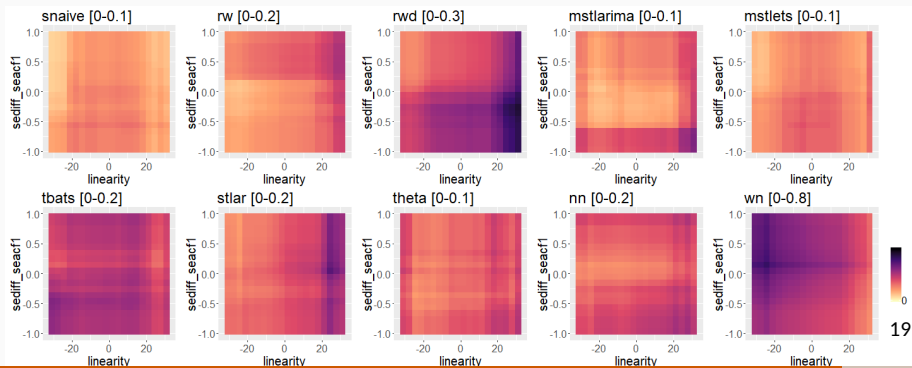


Interaction effect

■ Friedman's H-statistic

fraction of variance of two-variable partial dependency not captured by sum of the respective individual partial dependencies.

Hourly: interaction between linearity and seasonal lag at seasonally-differenced series



- Global perspective of feature contribution: the overall role of features in the choice of different forecast-models.

- Global perspective of feature contribution: the overall role of features in the choice of different forecast-models.
- **What next?** Local perspective of feature contribution: zoom into local regions of the data to identify which features contribute most to classify a specific instance.



available at: <https://github.com/thiyanagt/seer>

```
devtools::install_github("thiyanagt/seer")  
library(seer)
```