A Classification Framework for Forecast-Model Selection

Thiyanga S. Talagala

Professor Rob J Hyndman
A/Professor George Athanasopoulos

August 1, 2018

Large collections of time series

- Forecasting demand for thousands of products across multiple warehouses.

Freelance **Forecasting Multiple time series** 8 Mar 2016

Logistics Capital & Strategy – Posted by LogCapStrat – Anywhere

Job Description

Logistics Capital & Strategy is looking for a Data Scientist with expertise in Parallel computing to assist in code optimization and parallel processing of an under development forecasting model in R.

This is a contract position and we are expecting the project to completed over a period of 2 weeks.

Skills Required:
R Programming/Python/Scala (for code development)
MSSQL for data extraction into programming environment
Apache Spark or related big data processing frameworks to allow for high speed data processing

Project Scope:
The current forecasting model build on R needs to be scaled, and optimized to allow forecasting of millions of individual time series, ideally in a span of few hours.

Related

Statistician & R programmer	Quantitative Research Associate	R Shiny Developer
March 16, 2017	March 6, 2017	March 14, 2017
Similar post	Similar post	Similar post

How to Apply

Email at: smaity@logcapstrat.com

source:
https://www.r-users.com/jobs/forecasting-multiple-time-series/

Freelance **Forecasting Multiple time series** 8 Mar 2016

Logistics Capital & Strategy – Posted by LogCapStrat – Anywhere

Logistics CAPITAL & STRATEGY

Job Description

Logistics Capital & Strategy is looking for a Data Scientist with expertise in Parallel computing to assist in code optimization and parallel processing of an under development forecasting model in R.

This is a contract position and we are expecting the project to completed over a period of 2 weeks.

Skills Required:
R Programming/Python/Scala (for code development)
MSSQL for data extraction into programming environment
Apache Spark or related big data processing frameworks to allow for high speed data processing

Project Scope:
The current forecasting model build on R needs to be scaled, and optimized to allow forecasting of millions of individual time series, ideally in a span of few hours.

Related

forecasting of millions of individual time series

Statistician & R programmer
March 16, 2017
Similar post

Quantitative Research Associate
March 6, 2017
Similar post

R Shiny Developer
March 14, 2017
Similar post

How to Apply

Email at: smaity@logcapstrat.com

source:
https://www.r-users.com/jobs/forecasting-multiple-time-series/

Forecasting multiple time series

- Use a single method to provide forecasts across all time series.

Forecasting multiple time series

- Use a single method to provide forecasts across all time series.
    - No free lunch!

Forecasting multiple time series

- Use a single method to provide forecasts across all time series.
  - No free lunch!
- Develop a framework to select the most appropriate forecasting method.

Forecasting multiple time series

- Use a single method to provide forecasts across all time series.
    - No free lunch!
- Develop a framework to select the most appropriate forecasting method.

Motivation

- Reid(1972) pointed out that the **performance of various forecasting methods** changes according to the nature of data and if the reasons for these variations are explored they may be useful in selecting the most appropriate model.

Motivation

- Reid(1972) pointed out that the **performance of various forecasting methods** changes according to the nature of data and if the reasons for these variations are explored they may be useful in selecting the most appropriate model.
- "**What** are the specific time series characteristics for which each technique is generally best and also **what** are the time series characteristics for which it does not really matter which technique is chosen ?" (Lawrence, IJF, 2001)

Motivation

- Reid(1972) pointed out that the **performance of various forecasting methods** changes according to the nature of data and if the reasons for these variations are explored they may be useful in selecting the most appropriate model.
- "**What** are the specific time series characteristics for which each technique is generally best and also **what** are the time series characteristics for which it does not really matter which technique is chosen ?" (Lawrence, IJF, 2001)

Motivation

- Reid(1972) pointed out that the **performance of various forecasting methods** changes according to the nature of data and if the reasons for these variations are explored they may be useful in selecting the most appropriate model.
- "**What** are the specific time series characteristics for which each technique is generally best and also **what** are the time series characteristics for which it does not really matter which technique is chosen ?" (Lawrence, IJF, 2001)

Objective

Develop a framework that automates the selection of the most appropriate forecasting model for a given time series by using a large array of features computed from the time series.

Time series features

Cognostics: Computer-aided diagnostics
(John W. Tukey, 1985)

- Called “features” or “characteristics” in the machine learning literature.

Time series features

Cognostics: Computer-aided diagnostics
(John W. Tukey, 1985)

- Called “features” or “characteristics” in the machine learning literature.
- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.

Time series features

Cognostics: Computer-aided diagnostics
(John W. Tukey, 1985)

- Called "features" or "characteristics" in the machine learning literature.
- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features

Time series features

Cognostics: Computer-aided diagnostics
(John W. Tukey, 1985)

- Called "features" or "characteristics" in the machine learning literature.
- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend

Time series features

Cognostics: Computer-aided diagnostics
(John W. Tukey, 1985)

- Called “features” or “characteristics” in the machine learning literature.
- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend
  - strength of seasonality

Time series features

Cognostics: Computer-aided diagnostics
(John W. Tukey, 1985)

- Called "features" or "characteristics" in the machine learning literature.
- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend
  - strength of seasonality
  - lag-1 autocorrelation

Time series features

Cognostics: Computer-aided diagnostics
(John W. Tukey, 1985)

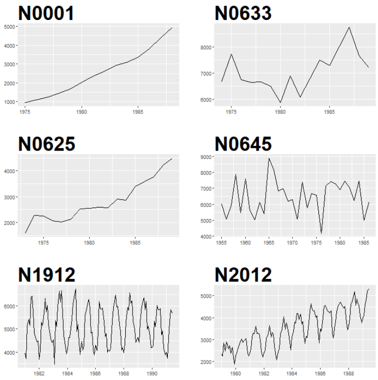
- Called "features" or "characteristics" in the machine learning literature.
- **Basic idea:**
  Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
- Examples for time series features
  - strength of trend
  - strength of seasonality
  - lag-1 autocorrelation
  - spectral entropy

Feature-space of time series

STL-decomposition

$$Y_t = T_t + S_t + R_t$$

- strength of trend: $1 - \frac{Var(R_t)}{Var(Y_t - S_t)}$
- strength of seasonality: $1 - \frac{Var(R_t)}{Va(Y_t - T_t)}$

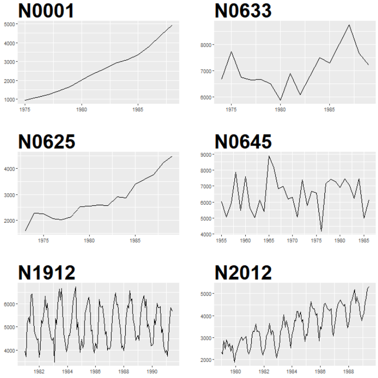


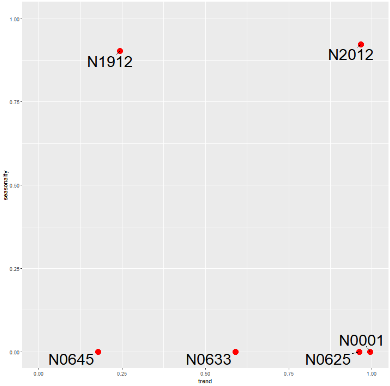
Feature-space of time series

STL-decomposition

$$Y_t = T_t + S_t + R_t$$

- strength of trend: $1 - \frac{Var(R_t)}{Var(Y_t - S_t)}$
- strength of seasonality: $1 - \frac{Var(R_t)}{Va(Y_t - T_t)}$





General framework for forecast-model selection using a meta-learning approach

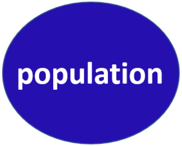
Offline

- Classification algorithm is trained

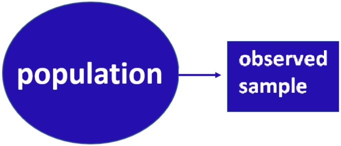
Online

- Use the classification algorithm to select appropriate forecast-models for new time series

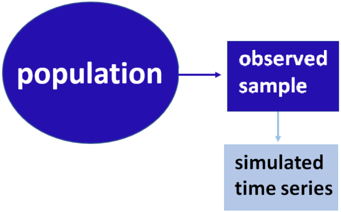
Methodology: population



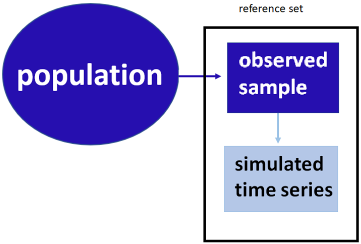
Methodology: observed sample



Methodology: simulated time series

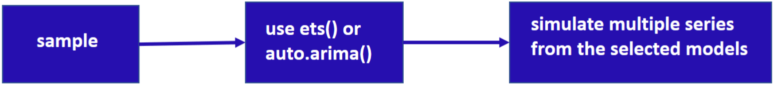


Methodology: reference set



Augmenting the observed sample with simulated series

- Time series simulation process



- when our sample is too small to build a reliable classifier

Augmenting the observed sample with simulated series

- Time series simulation process

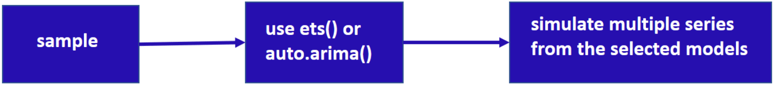
sample → **use ets() or auto.arima()** → **simulate multiple series from the selected models**

- when our sample is too small to build a reliable classifier

- to add more of some types of time series to the training set in order to get a more balanced sample

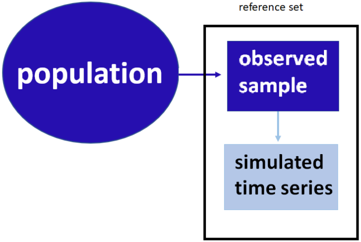
Augmenting the observed sample with simulated series

- Time series simulation process

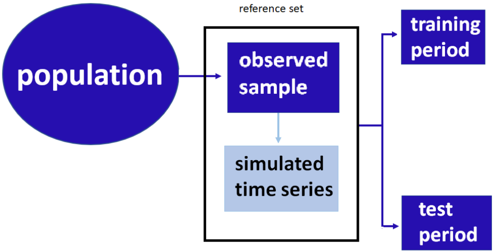


- when our sample is too small to build a reliable classifier
- to add more of some types of time series to the training set in order to get a more balanced sample
- to increase the diversity and evenness of the feature space

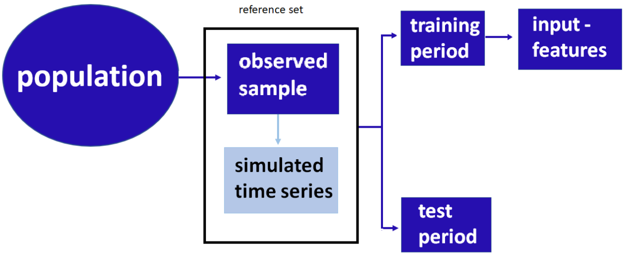
Methodology: reference set



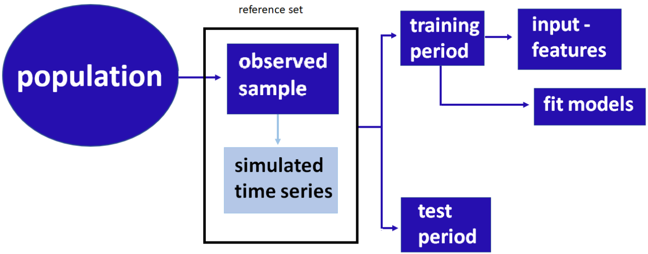
Methodology: Meta-data



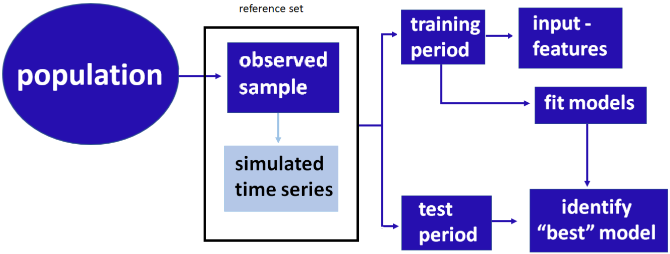
Methodology: Meta-data



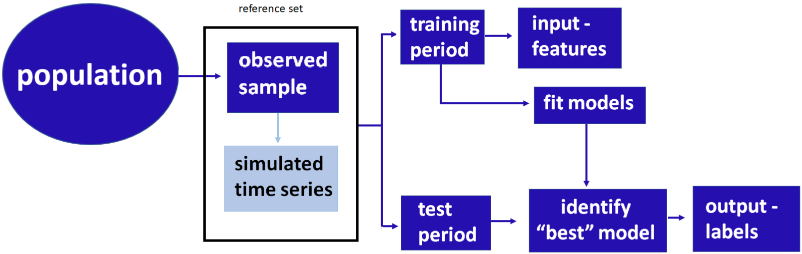
Methodology: Meta-data



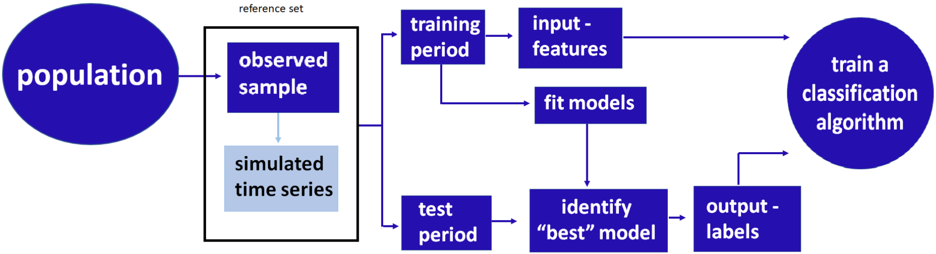
Methodology: Meta-data



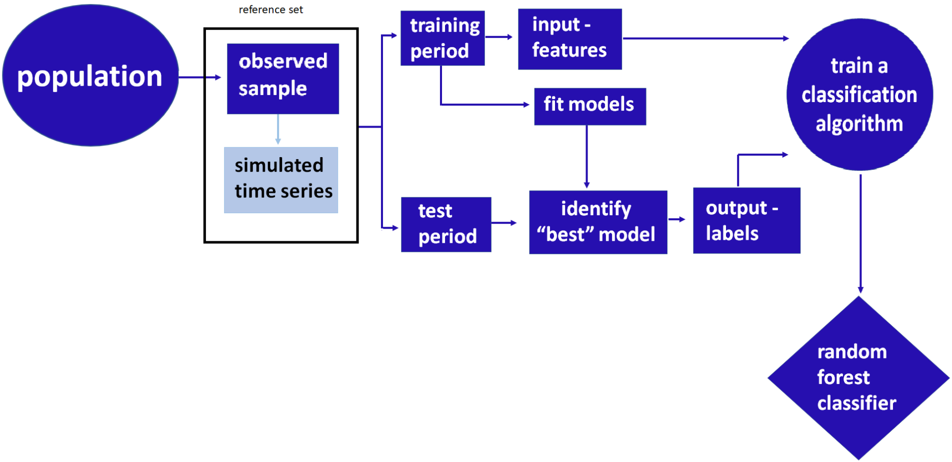
Methodology: Meta-data



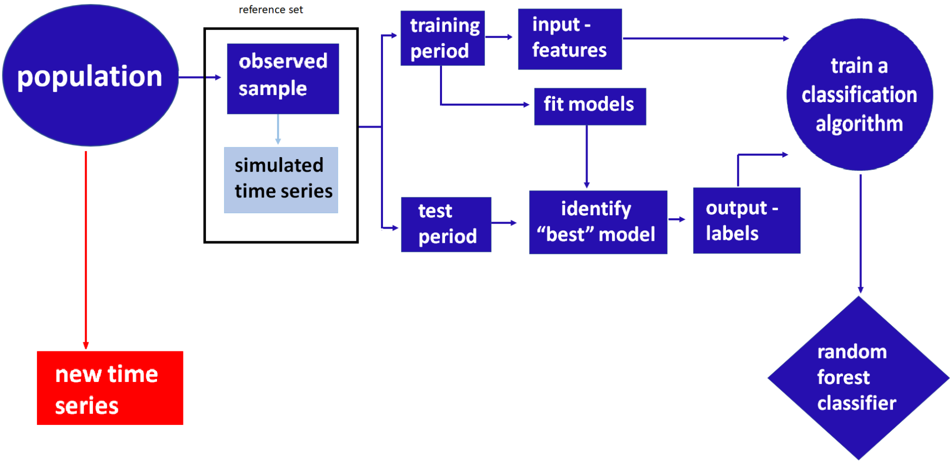
Methodology: Meta-data



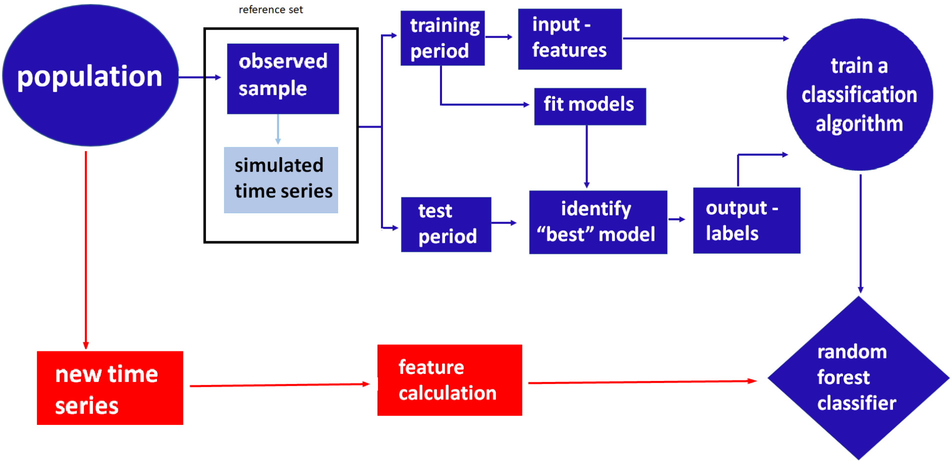
Methodology: Random-forest classifier



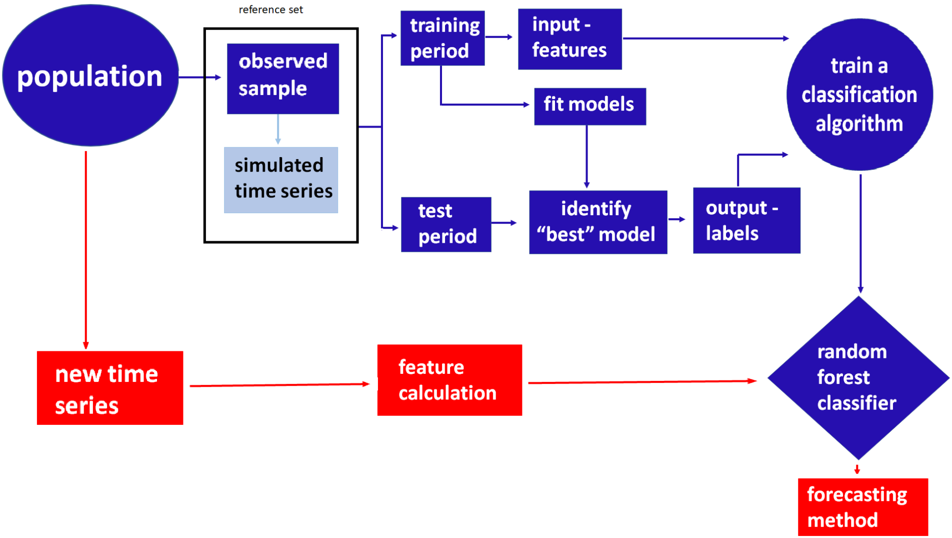
Methodology: “online” part of the algorithm



Methodology: "online" part of the algorithm



Methodology: “online” part of the algorithm



Application to M competition data

- Proposed algorithm is applied to yearly, quarterly and monthly series separately
- We run two experiments for each case.

	Experiment 1				Experiment 2			
	Source	Y	Q	M	Source	Y	Q	M
Observed series	M1	181	203	617	M3	645	756	1428
New series	M3	645	756	1428	M1	181	203	617

Application to M competition data

- Simulate time series based on `auto.arima` and `ets`

	Source	Experiment 1 Y	Q	M	Source	Experiment 2 Y	Q	M
Observed series	M1	181	203	617	M3	645	756	1428
Simulated series		362000	406000	123400		1290000	1512000	285600

Example:

Experiment 1 - Yearly

simulated time series (‘ets‘)	181000
simulated time series (‘auto.arima‘)	181000
Total simulated	362000

output-labels

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

Model calibration

We consider three approaches to address the class imbalance in the data

- incorporate class priors into the random forest classifier (Chen, Liaw & Breiman, 2004)

Model calibration

We consider three approaches to address the class imbalance in the data

- incorporate class priors into the random forest classifier (Chen, Liaw & Breiman, 2004)
- Balanced random forest

Model calibration

We consider three approaches to address the class imbalance in the data

- incorporate class priors into the random forest classifier (Chen, Liaw & Breiman, 2004)
- Balanced random forest
- re-balance the reference set with down-sampling

Model calibration

We consider three approaches to address the class imbalance in the data

- incorporate class priors into the random forest classifier (Chen, Liaw & Breiman, 2004)
- Balanced random forest
- re-balance the reference set with down-sampling

Model calibration

We consider three approaches to address the class imbalance in the data

- incorporate class priors into the random forest classifier (Chen, Liaw & Breiman, 2004)
- Balanced random forest
- re-balance the reference set with down-sampling

Further,

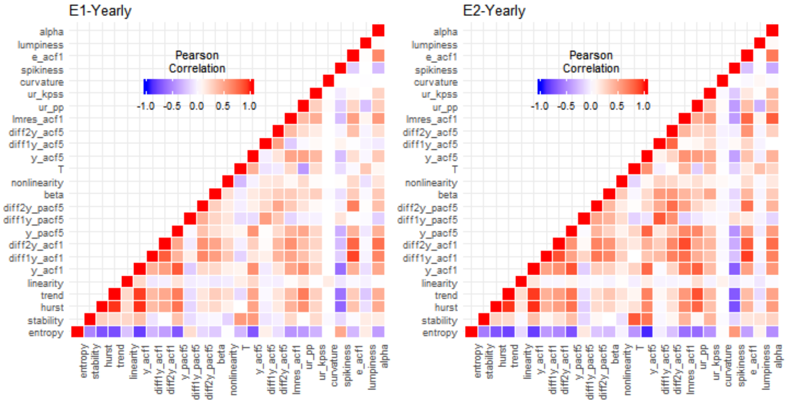
- Random forest classifier built on unbalanced data

Results: Yearly series

25 features:

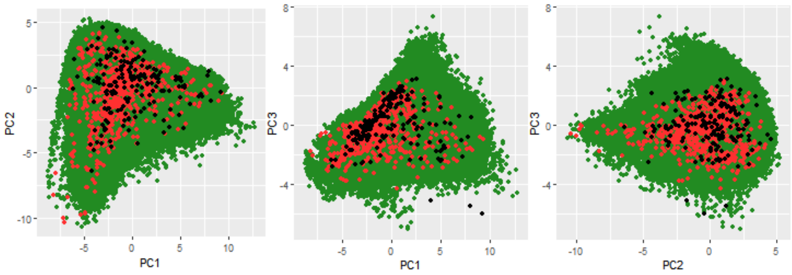
- length
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- first ACF value of remainder series
- parameter estimates of Holt's linear trend method
- spectral entropy
- Hurst exponent
- nonlinearity
- unit root test statistics
- first ACF value of residual series of linear trend model
- ACF and PACF based features - calculated on both the raw and differenced series

Yearly series: correlation matrix plots for the reference sets



Yearly - Experiment 1: Distribution of time series in the PCA space

- observed-M1, simulated, new-M3

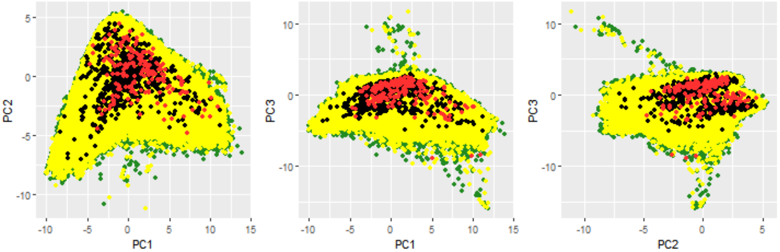


Yearly - Experiment 1: Observed-M1, new-M3

	Yearly: 645 M3 time series				
	h= 1	1–2	1–3	1–4	1–5
RF-unbalanced	1.06	1.40	2.17	2.82	3.50
RF-class priors	1.04	1.38	2.15	2.79	2.50
auto.arima	1.11	1.48	2.28	2.96	5.83
ets	1.09	1.44	2.20	2.86	4.67
WN	6.54	6.91	7.48	8.07	9.00
RW	1.24	1.68	2.48	3.17	8.00
RWD	1.03	1.36	2.05	2.63	1.00
Theta	1.12	1.47	2.18	2.77	3.50

Yearly - Experiment 2: Distribution of time series in the PCA space

- simulated, subset, observed-M3, new-M1



Yearly - Experiment 2: observed-M3, new-M1

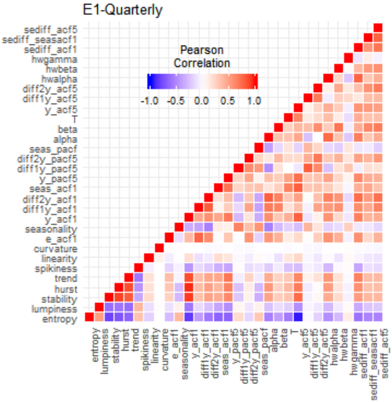
	Yearly: 181 M1 time series				
	h=1	1–2	1–4	1–6	Average Rank
RF-unbalanced	0.98	1.40	2.43	3.39	1.50
RF-class priors	1.01	1.40	2.43	3.38	1.50
auto.arima	1.06	1.47	2.51	3.47	3.33
ets	1.12	1.59	2.72	3.77	5.00
WN	6.38	7.08	8.59	10.01	8.00
RW	1.35	2.00	3.50	4.89	7.00
RWD	1.04	1.44	2.51	3.49	3.67
Theta	1.15	1.70	3.00	4.19	6.00

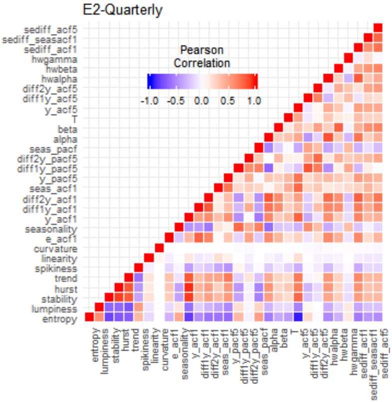
Results: quarterly and monthly series

30 features:

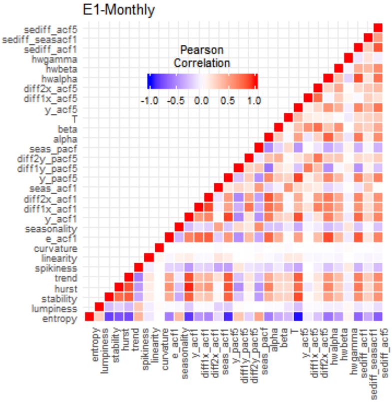
- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- first ACF value of remainder series
- parameter estimates of Holt's linear trend method
- spectral entropy
- Hurst exponent
- nonlinearity
- parameter estimates of Holt-Winters' additive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

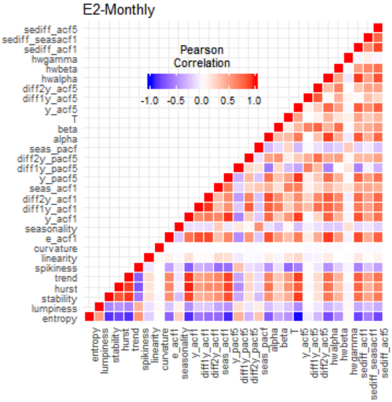
Quarterly series: correlation matrix plots for the reference sets





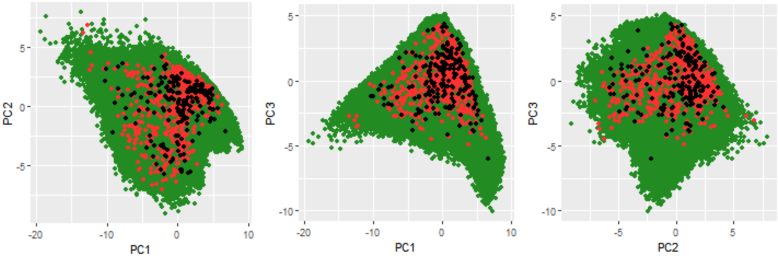
Monthly series: correlation matrix plots for the reference sets





Quarterly - Experiment 1: Distribution of time series in the PCA space

- observed-M1, simulated, new-M3

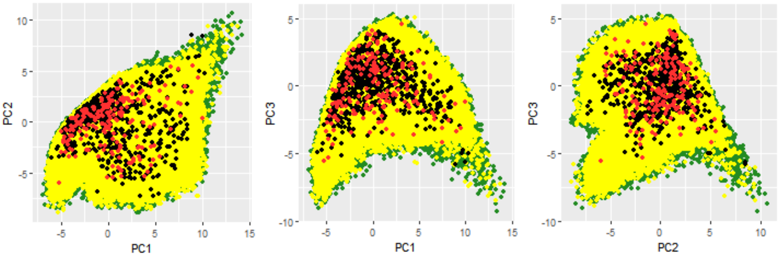


Quarterly - Experiment 1: observed-M1, new-M3

	Quarterly: 756 M3 time series				
	h=1	1–4	1–6	1–8	Average Rank
RF-unbalanced	0.59	0.81	0.97	1.12	2.25
RF-class priors	0.59	0.82	0.97	1.13	3.13
auto.arima	0.59	0.85	1.02	1.19	4.75
ets	0.56	0.82	0.99	1.17	3.75
WN	3.25	3.59	3.70	3.87	10.00
RW	1.14	1.16	1.32	1.46	7.00
RWD	1.20	1.17	1.36	1.47	6.50
STL-AR	0.70	1.27	1.60	1.91	8.34
Theta	0.62	0.83	0.97	1.11	2.50
Snaive	1.11	1.09	1.30	1.43	6.75

Quarterly - Experiment 2: Distribution of time series in the PCA space

- simulated, subset, observed-M3, new-M1

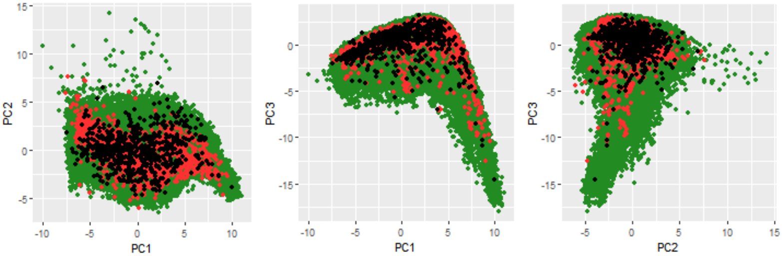


Quarterly - Experiment 2: observed-M3, new-M1

	Quarterly: 203 M1 time series				
	h=1	1–4	1–6	1–8	Average Rank
RF-unbalanced	0.74	1.08	1.35	1.57	1.00
RF-class priors	0.76	1.12	1.40	1.62	2.63
auto.arima	0.78	1.17	1.50	1.74	5.25
ets	0.78	1.11	1.42	1.66	3.00
WN	3.97	4.27	4.45	4.64	10.00
RW	0.97	1.35	1.67	1.95	7.50
RWD	0.95	1.26	1.56	1.81	5.38
STL-AR	0.96	1.63	2.05	2.43	8.63
Theta	0.79	1.13	1.42	1.67	3.88
Snaive	1.52	1.56	1.87	2.08	7.75

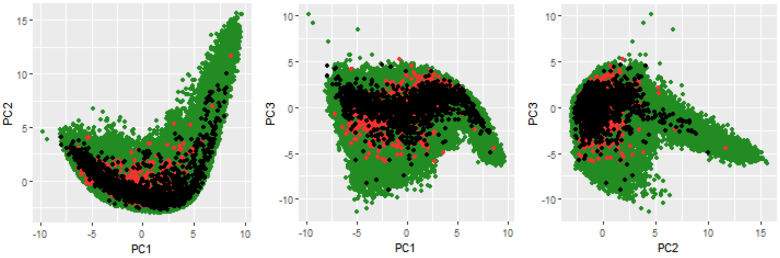
Monthly - Experiment 1: Distribution of time series in the PCA space

- observed-M1, simulated, new-M3



Monthly - Experiment 2: Distribution of time series in the PCA space

- observed-M3, simulated, new-M1



Monthly - Experiment 2: observed-M3, new-M1

	Monthly: 617 M1 time series				
	h=1	1–6	1–12	1–18	Average rank
RF-unbalanced	0.61	0.76	0.90	1.03	1.77
RF-class priors	0.60	0.75	0.92	1.06	2.83
auto.arima	0.60	0.76	0.96	1.12	4.94
ets	0.59	0.76	0.93	1.07	3.44
WN	1.93	2.09	2.18	2.28	10.00
RW	1.05	1.24	1.33	1.47	7.25
RWD	1.06	1.27	1.39	1.40	1.39
STL-AR	0.63	0.91	1.17	1.39	7.38
Theta	0.61	0.75	0.92	1.04	2.27
Snaive	1.06	1.11	1.14	1.31	6.47

Discussion and Conclusions

- Proposed a novel framework for forecast-model selection using meta-learning based on time series features.

Discussion and Conclusions

- Proposed a novel framework for forecast-model selection using meta-learning based on time series features.
- Our method almost always performs better than common benchmark methods, and better than the best-performing methods from the M3 competition.

Discussion and Conclusions

- Proposed a novel framework for forecast-model selection using meta-learning based on time series features.
- Our method almost always performs better than common benchmark methods, and better than the best-performing methods from the M3 competition.
- The framework is general and can be applied to any large collection of time series.

Discussion and Conclusions

- Proposed a novel framework for forecast-model selection using meta-learning based on time series features.
- Our method almost always performs better than common benchmark methods, and better than the best-performing methods from the M3 competition.
- The framework is general and can be applied to any large collection of time series.
- Advantage: Not necessary to estimate several different models for the data and undertake an empirical evaluation of their forecast accuracy on a given time series.