



MONASH  
University

MONASH  
BUSINESS  
SCHOOL

# Feature-based Time Series Forecasting

Thiyanga S. Talagala

13 March 2019

# Joint work with

# Introduction

# Big picture of the problem

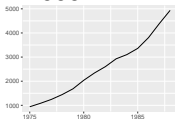
# Time series features

- Transform a given time series  $y = \{y_1, y_2, \dots, y_n\}$  to a feature vector  $F = (f_1(y), f_2(y), \dots, f_p(y))'$ .

# Time series features

- Transform a given time series  $y = \{y_1, y_2, \dots, y_n\}$  to a feature vector  $F = (f_1(y), f_2(y), \dots, f_p(y))'$ .

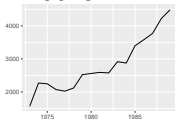
**N0001**



**N0633**



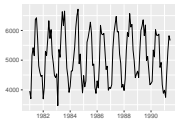
**N0625**



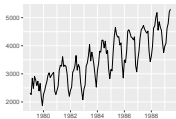
**N0645**



**N1912**



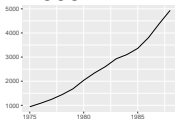
**N2012**



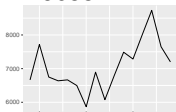
# Time series features

- Transform a given time series  $y = \{y_1, y_2, \dots, y_n\}$  to a feature vector  $F = (f_1(y), f_2(y), \dots, f_p(y))'$ .

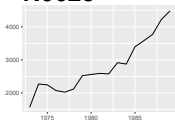
**N0001**



**N0633**



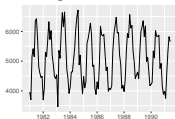
**N0625**



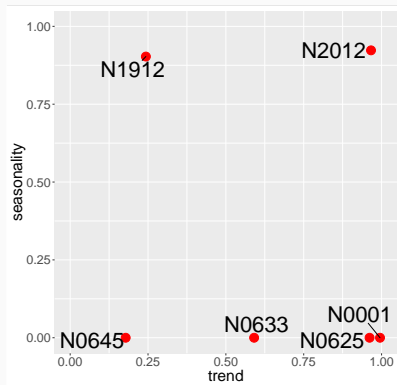
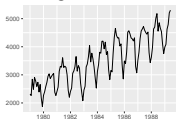
**N0645**



**N1912**



**N2012**

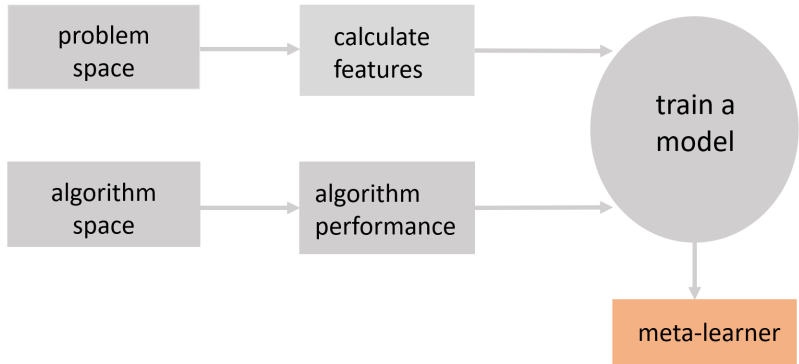


# Features used to select a forecasting model

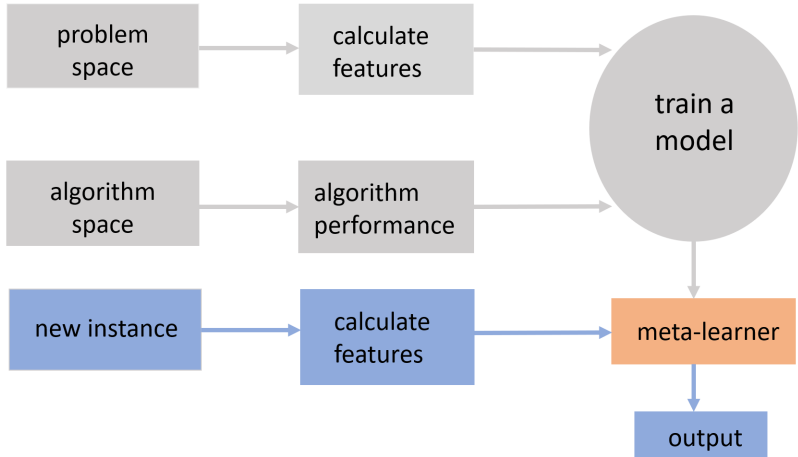
- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- parameter estimates of Holt's linear trend method
- spectral entropy
- Hurst exponent
- nonlinearity
- parameter estimates of Holt-Winters' additive method
- unit root test statistics
- crossing points, flat spots
- peaks, troughs
- ACF and PACF based features - calculated on raw, differenced, and remainder series.
- ARCH/GARCH statistics and ACF of squared series and residuals.



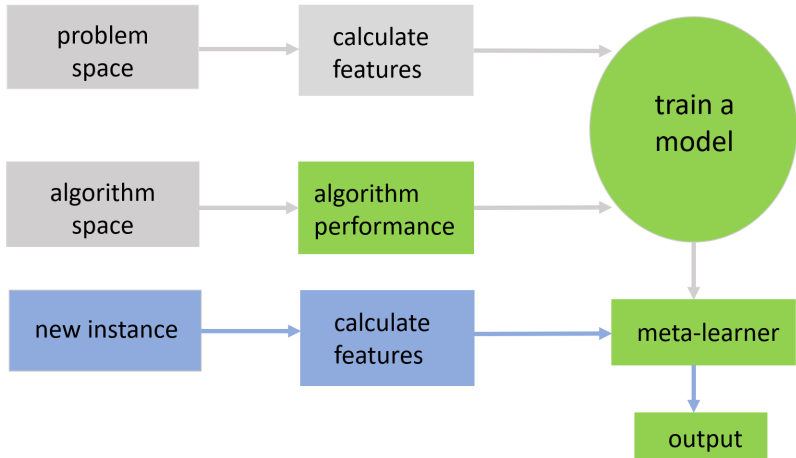
# Meta-learning



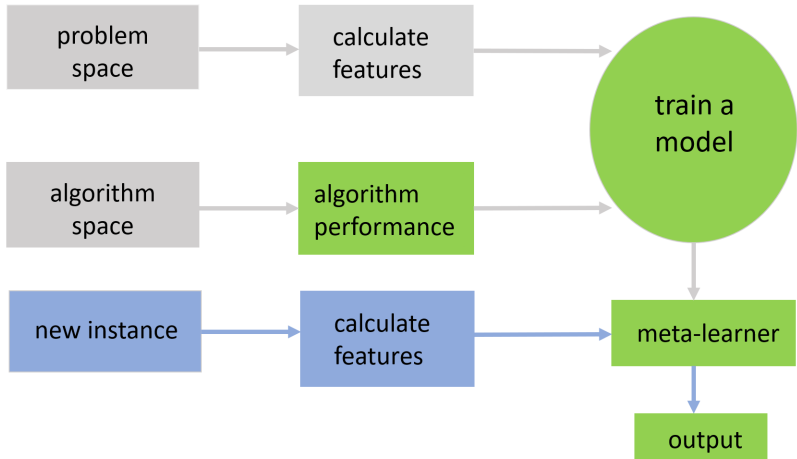
# Meta-learning



# Feature-based forecasting algorithms



# Feature-based forecasting algorithms



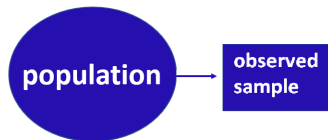
- three algorithms: FFORMS, FFORMA, FFORMPP

# FFORMS: Feature-based FORecast Model Selection

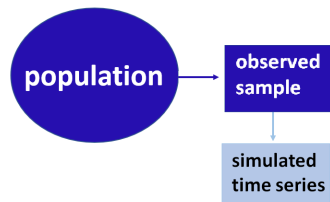


population

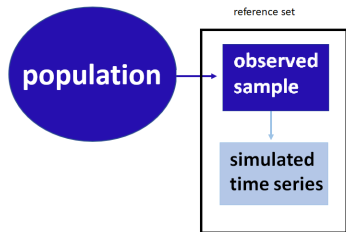
# FFORMS: observed sample



# FFORMS: simulated time series

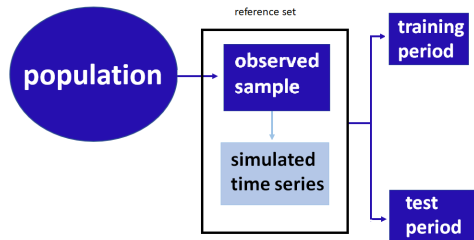


# FFORMS: reference set

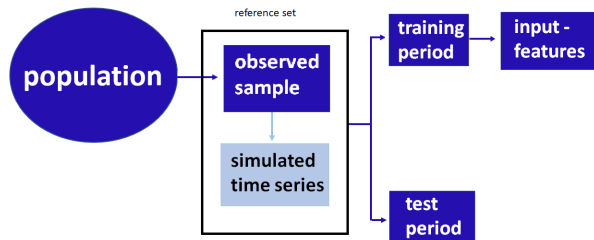




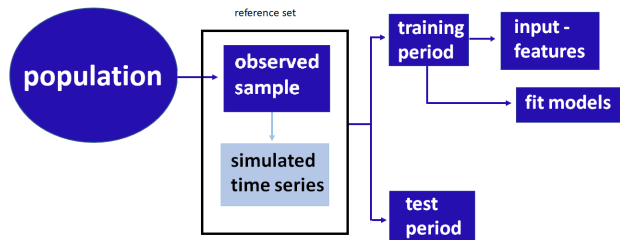
# FFORMS: Meta-data



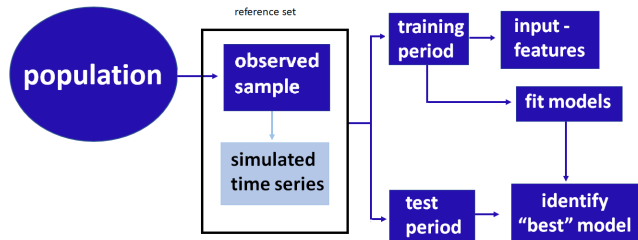
# FFORMS: Meta-data



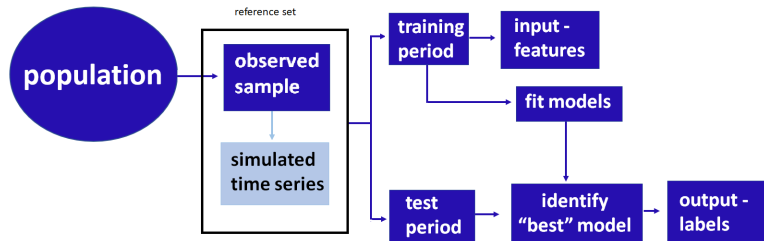
# FFORMS: Meta-data



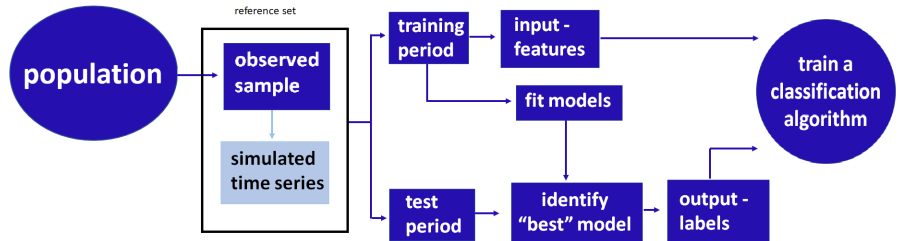
# FFORMS: Meta-data



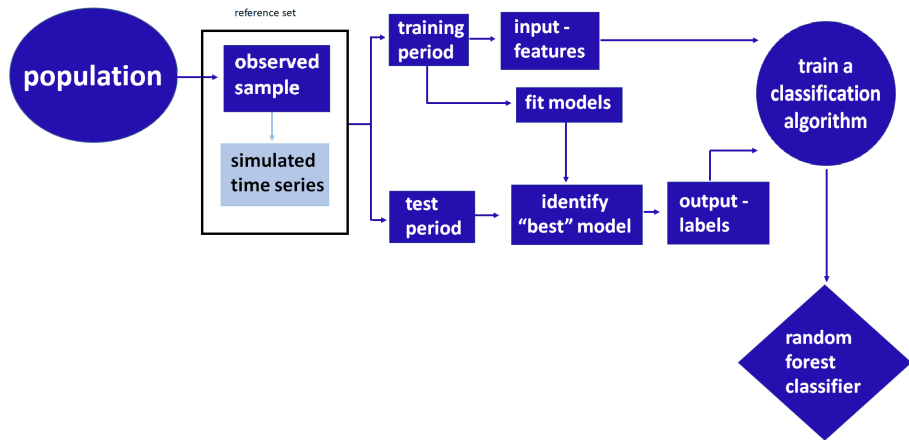
# FFORMS: Meta-data



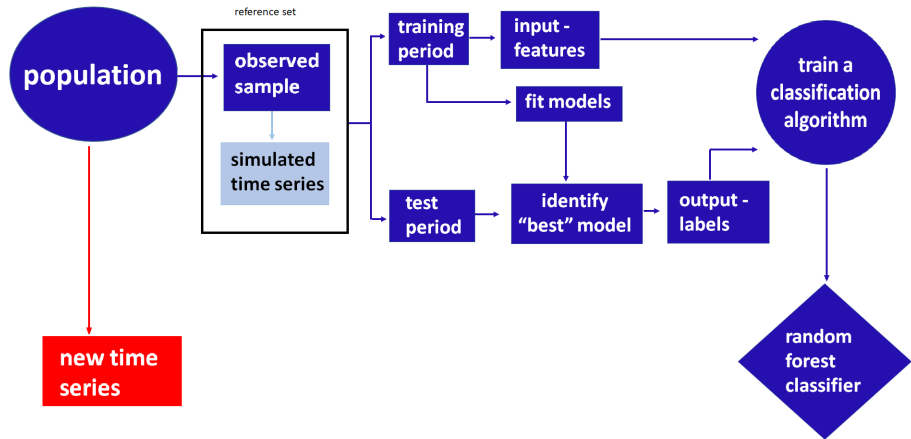
# FFORMS: Meta-data



# FFORMS: Random-forest classifier

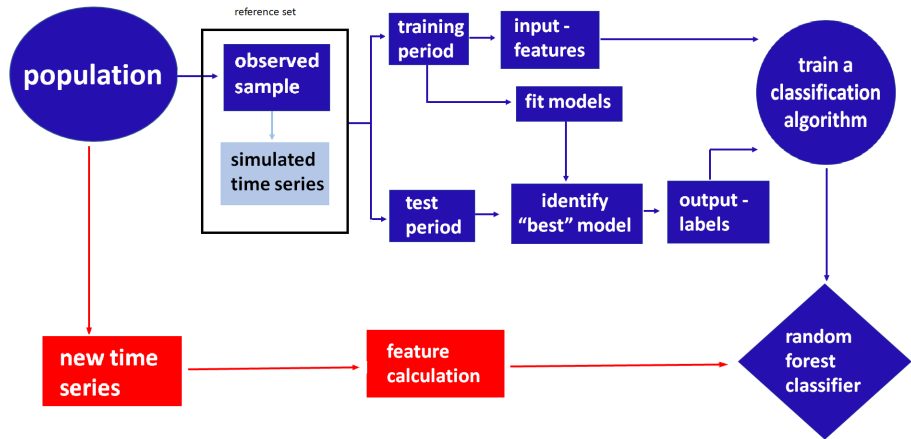


# FFORMS: “online” part of the algorithm

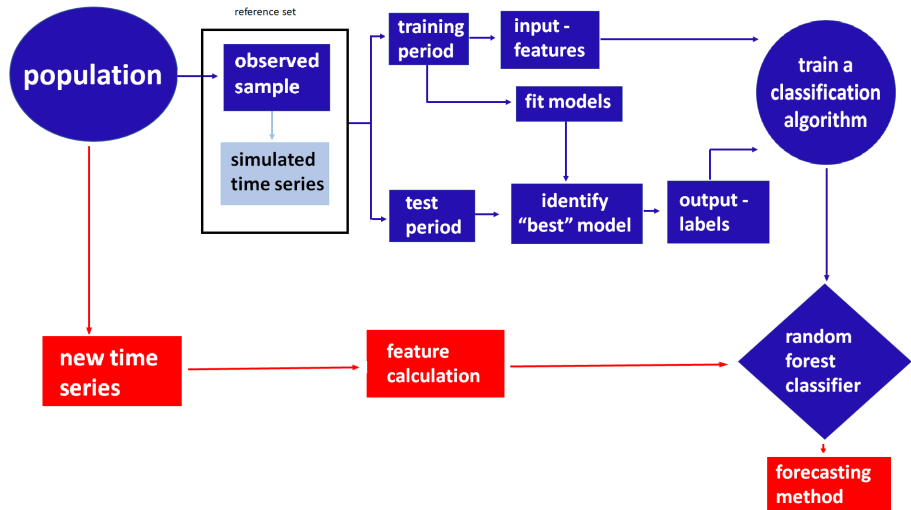




# FFORMS: “online” part of the algorithm



# FFORMS: “online” part of the algorithm



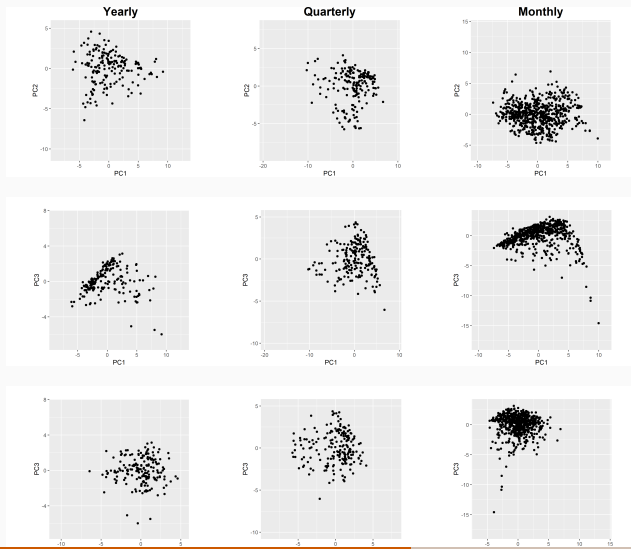
# Application to M competition data

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

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

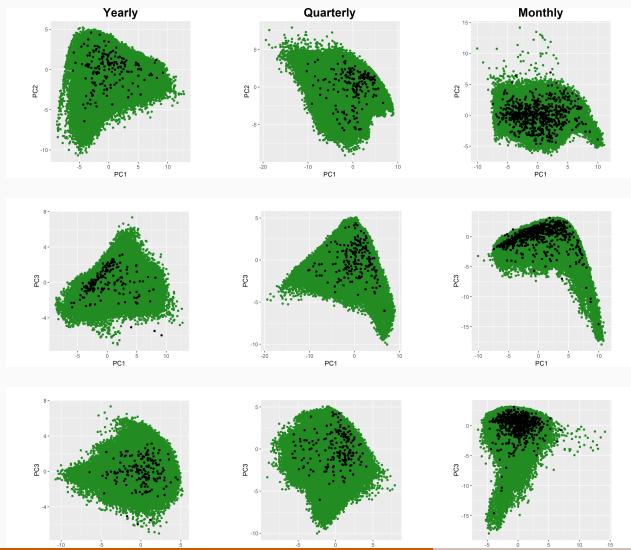
# Experiment 1: Distribution of time series in the PCA space

observed - M1



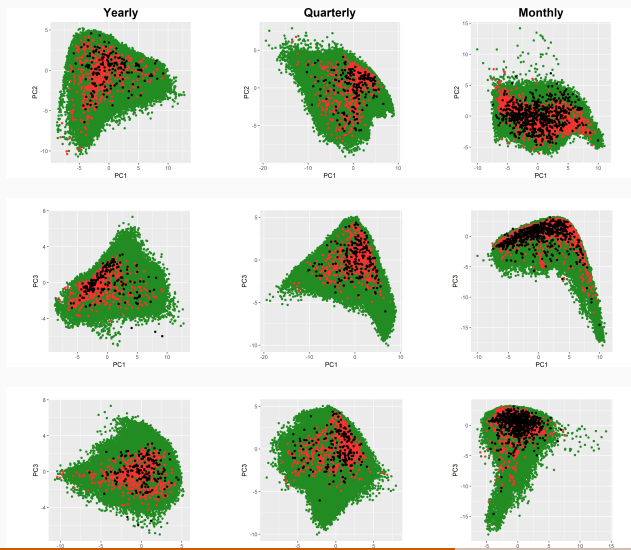
# Experiment 1: Distribution of time series in the PCA space

observed - M1 simulated



# Experiment 1: Distribution of time series in the PCA space

observed - M1 simulated new - M3



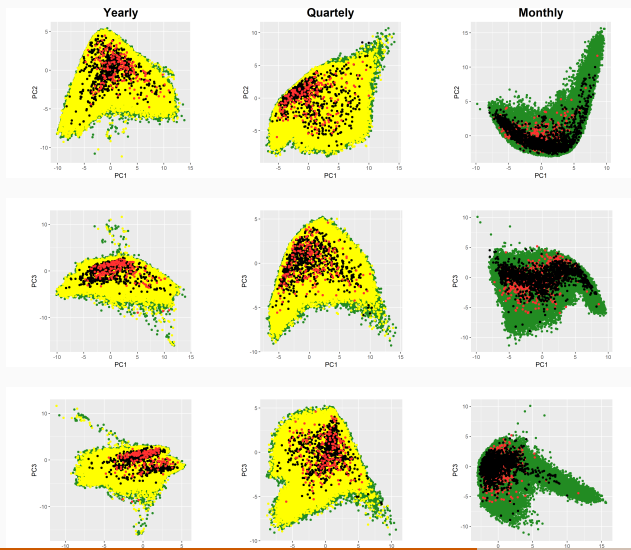
## Experiment 2: Distribution of time series in the PCA space

observed - M3

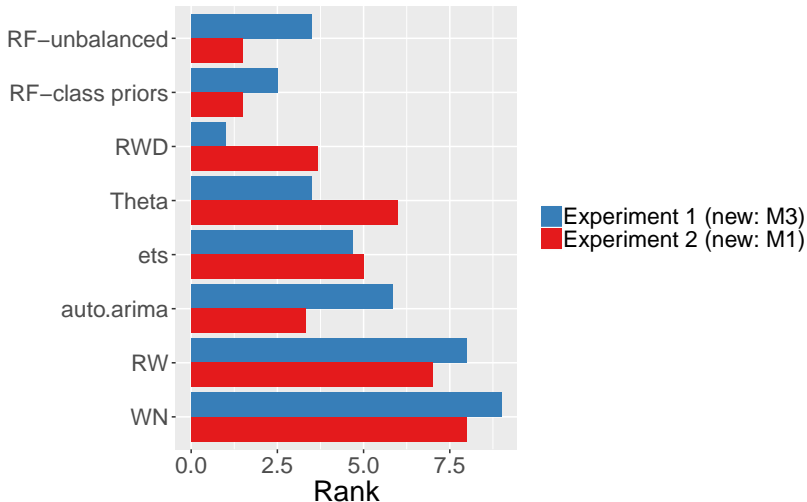
simulated

subset

new - M1

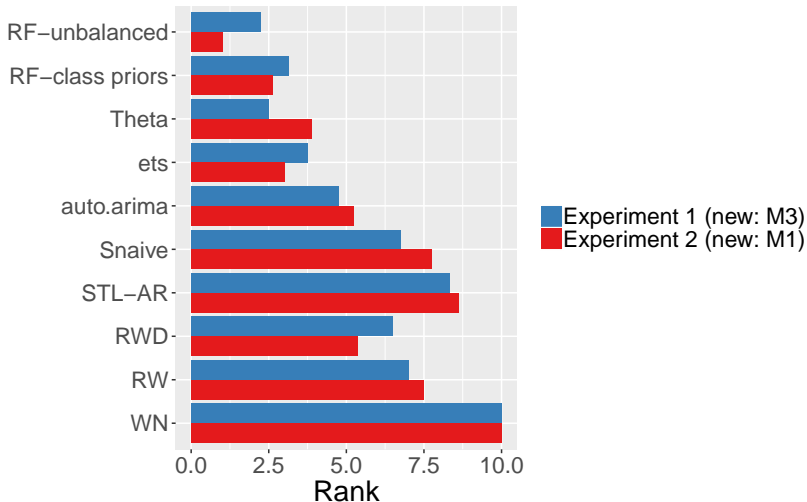


# Results: Yearly

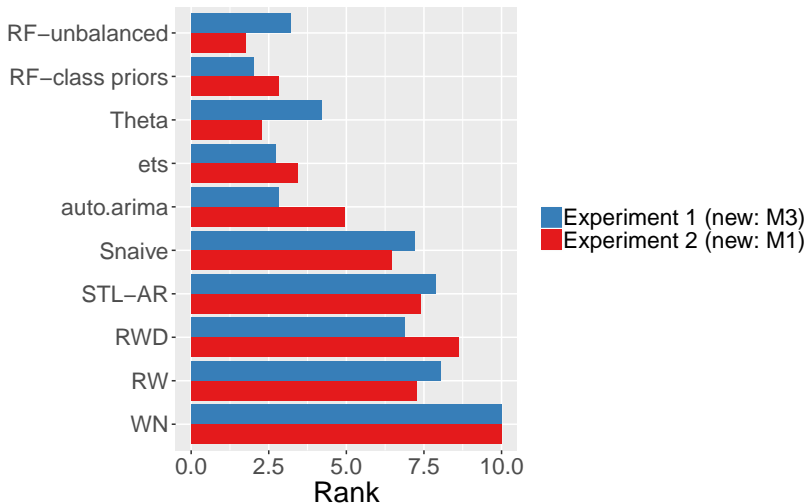




# Results: Quarterly



# Results: Monthly



# M4 Competition: 2018

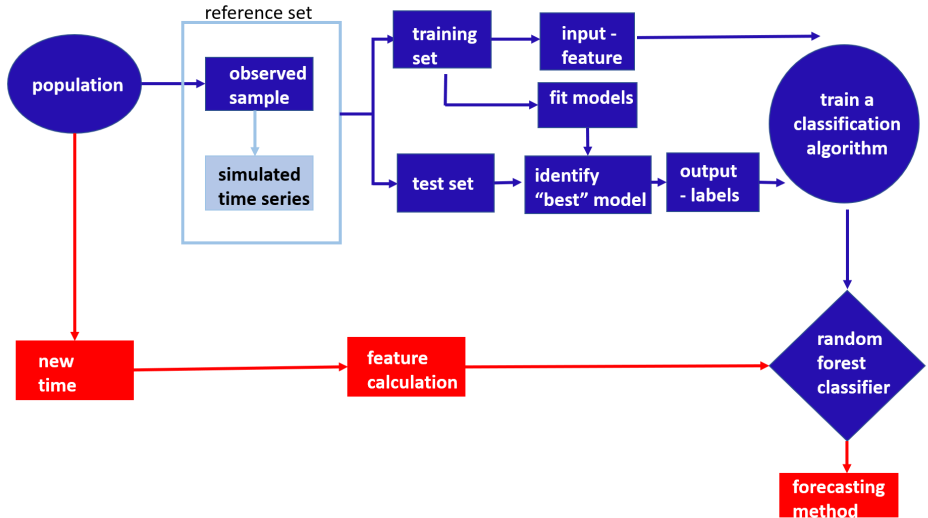
**M4COMPETITION** Forecast. Compete. Excel.

The background of the slide features a stylized orange grid. Overlaid on this grid is a bar chart with approximately 20 vertical bars of varying heights. A dashed white line, representing a time series, meanders across the chart, showing peaks and troughs. In the center of the slide, the text 'M4' is prominently displayed. The 'M' is a large, bold, white sans-serif letter. The '4' is a smaller, bold, black sans-serif digit, positioned to the right of the 'M' and slightly lower than its top edge.

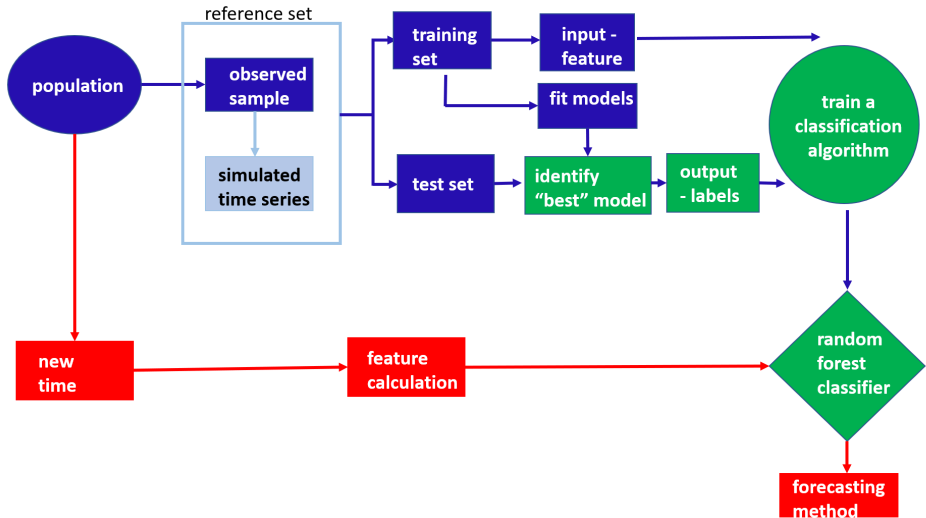
**M4**

- 100,000 time series: yearly, quarterly, monthly, weekly, daily, hourly

# FFORMS: Feature-based FOREcast Model Selection

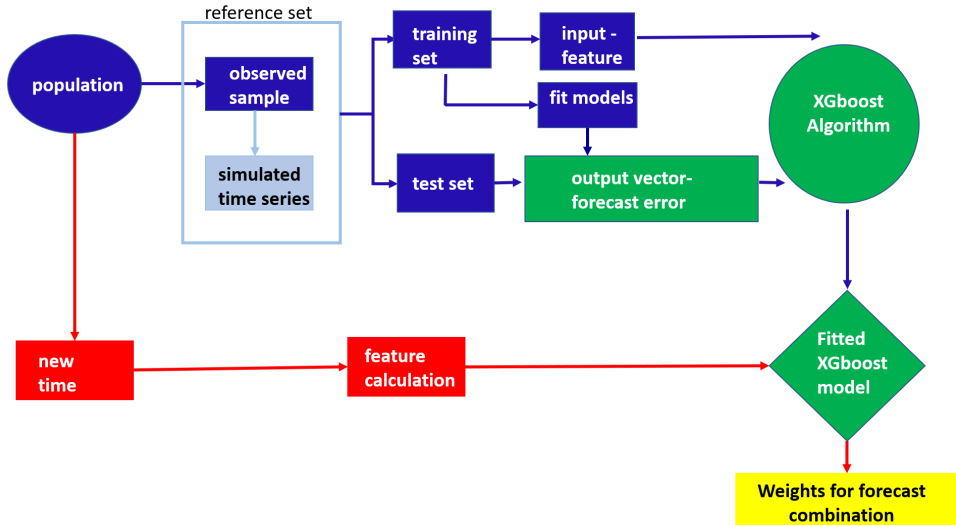


# FFORMS: Feature-based FOREcast Model Selection



■ optimization criterion: classification accuracy

# FFORMA: Feature-based FOfrecast Model Averaging



# FFORMA: Models included

- naive
- random walk with drift
- seasonal naive
- theta method
- automated ARIMA algorithm
- automated exponential smoothing algorithm
- TBATS model
- STLM-AR Seasonal and Trend decomposition  
using Loess with AR modeling of the seasonally  
adjusted series
- neural network time series forecasts

## FFORMA: Feature-based FOrecast Model Averaging

- Like FFORMS but we use xgboost rather than a random forest.



# FFORMA: Feature-based FOrecast Model Averaging

- Like FFORMS but we use xgboost rather than a random forest.
- Optimization criterion: forecast accuracy

## FFORMA: Feature-based FOrecast Model Averaging

- Like FFORMS but we use xgboost rather than a random forest.
- Optimization criterion: forecast accuracy
- The probability of each model being best is used to construct a model weight.

## FFORMA: Feature-based FOrecast Model Averaging

- Like FFORMS but we use xgboost rather than a random forest.
- Optimization criterion: forecast accuracy
- The probability of each model being best is used to construct a model weight.
- A combination forecast is produced using these weights.

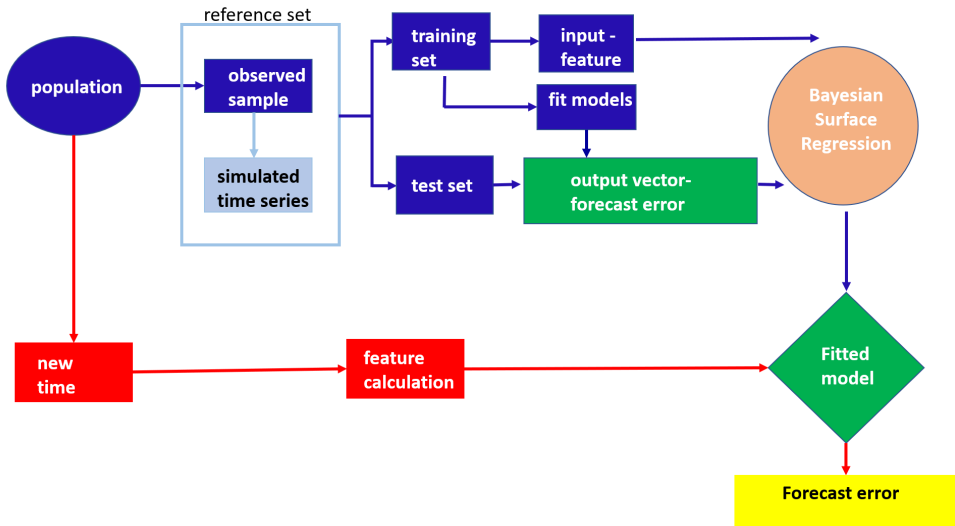
## FFORMA: Feature-based FOrecast Model Averaging

- Like FFORMS but we use xgboost rather than a random forest.
- Optimization criterion: forecast accuracy
- The probability of each model being best is used to construct a model weight.
- A combination forecast is produced using these weights.
- 248 registrations, 50 submissions

## FFORMA: Feature-based FOfecast Model Averaging

- Like FFORMS but we use xgboost rather than a random forest.
- Optimization criterion: forecast accuracy
- The probability of each model being best is used to construct a model weight.
- A combination forecast is produced using these weights.
- 248 registrations, 50 submissions
- Came second in the M4 competition

# FFORMPP: Feature-based FOREcast Model Performance Prediction



# Peeking inside FFORMS

# Feature-based FOfRecast Model Averaging



# Feature-based FOfecast Model Performance Prediction

# R packages

# References