

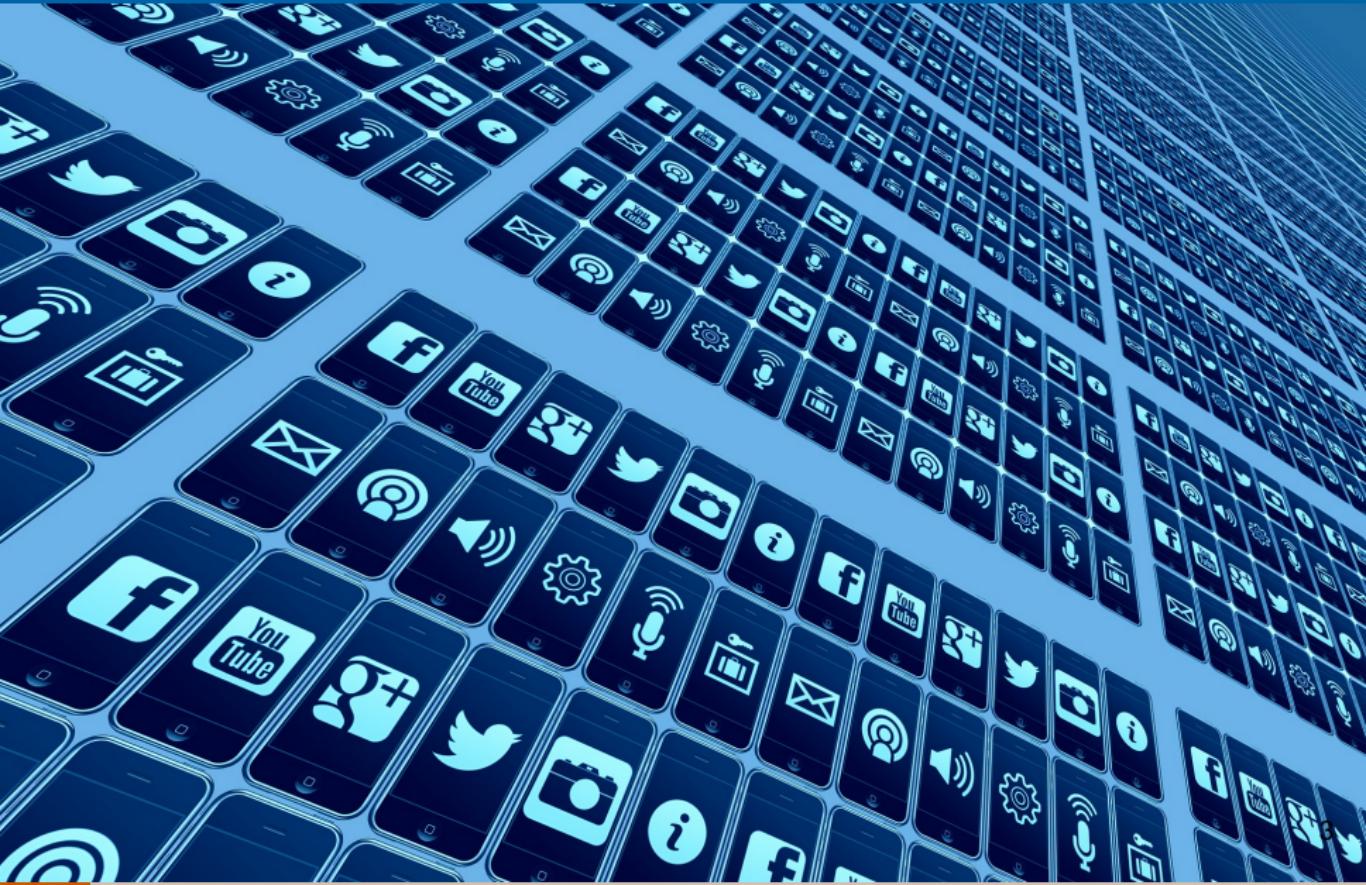
# Feature-based Time Series Forecasting

Thiyanga S. Talagala

13 March 2019

# Joint work with

# Introduction



# Big picture of the problem

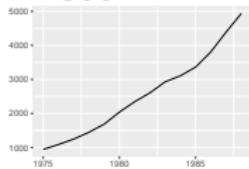
## Time series features

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

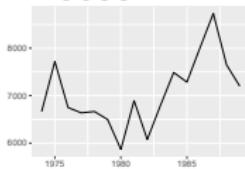
# Time series features

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

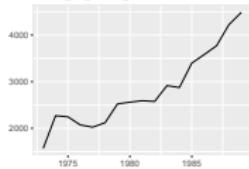
N0001



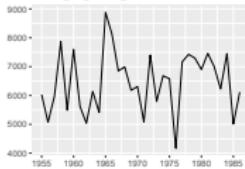
N0633



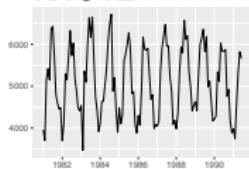
N0625



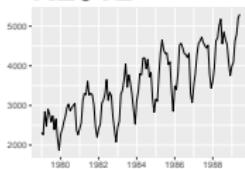
N0645



N1912



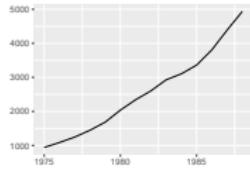
N2012



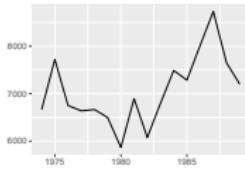
# Time series features

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

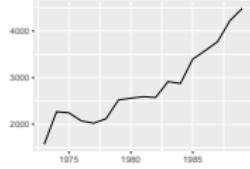
N0001



N0633



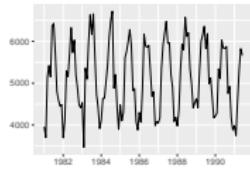
N0625



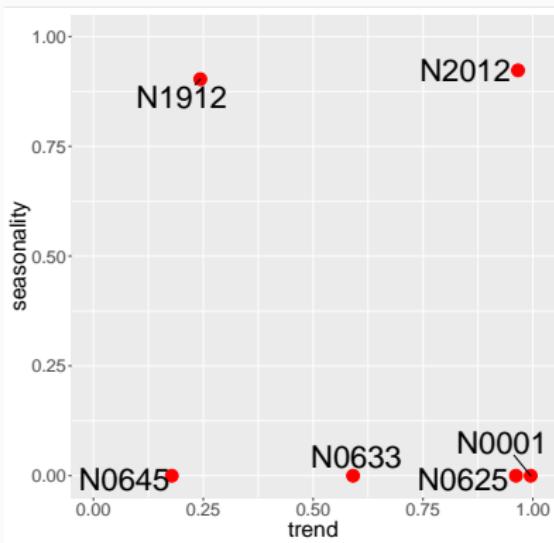
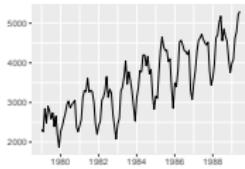
N0645



N1912



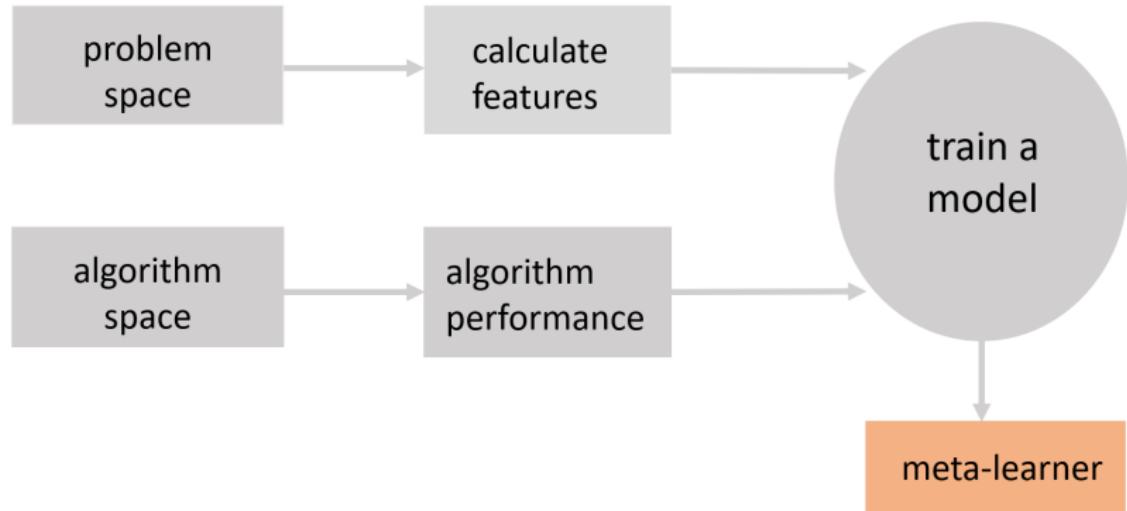
N2012



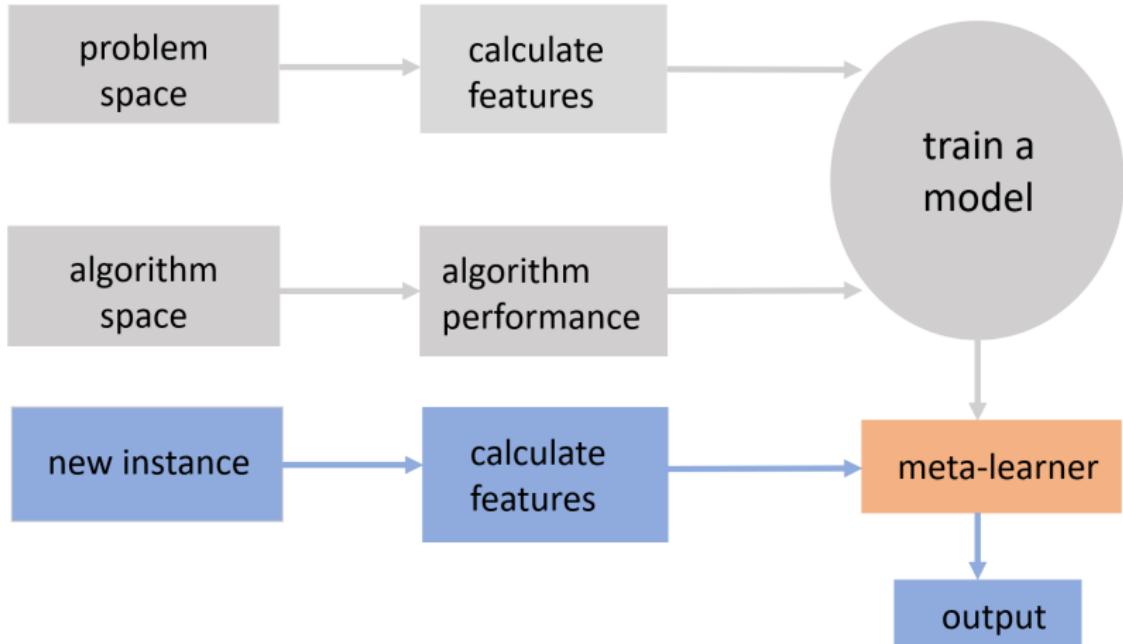
# Features used to select a forecasting model

- 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.

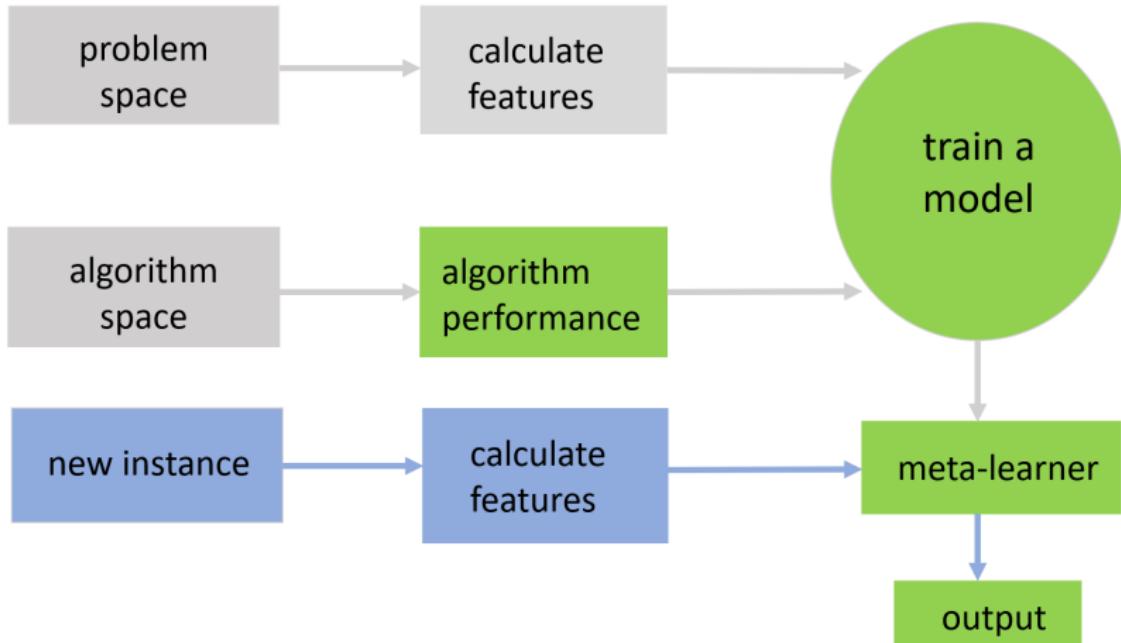
# Meta-learning



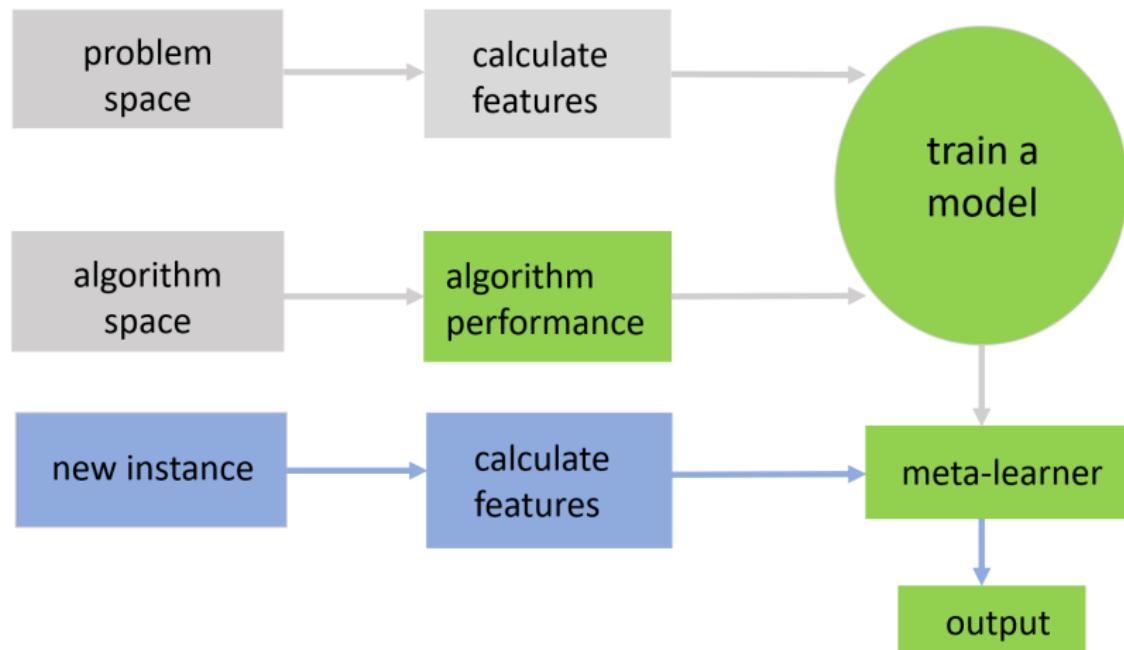
# Meta-learning



# Feature-based forecasting algorithms



# Feature-based forecasting algorithms

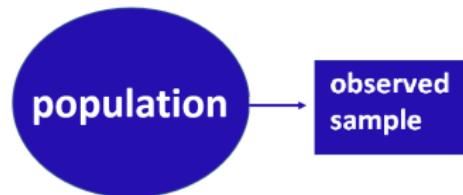


- three algorithms: **FFORMS**, **FFORMA**, **FFORMPP**

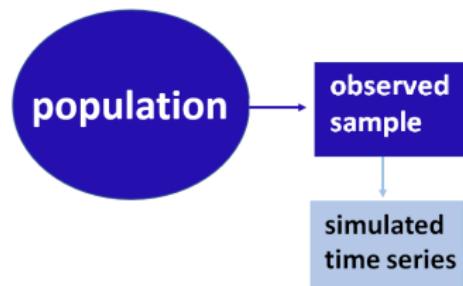
# FFORMS: Feature-based FOrecast Model Selection



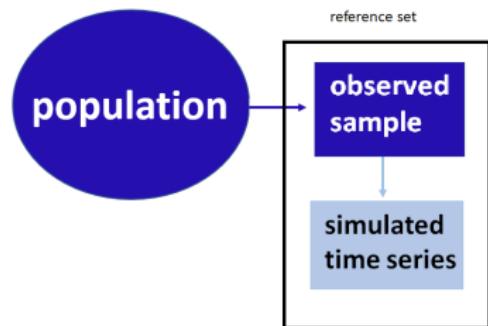
# FFORMS: observed sample



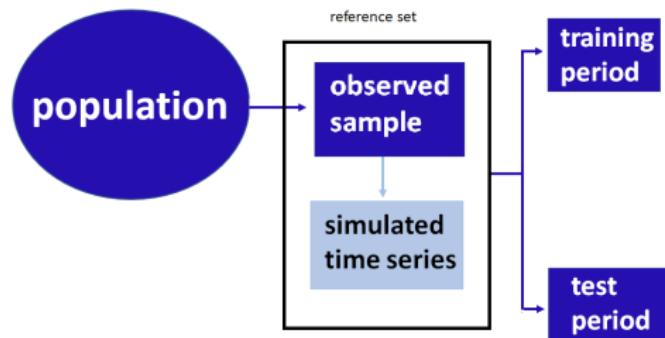
# FFORMS: simulated time series



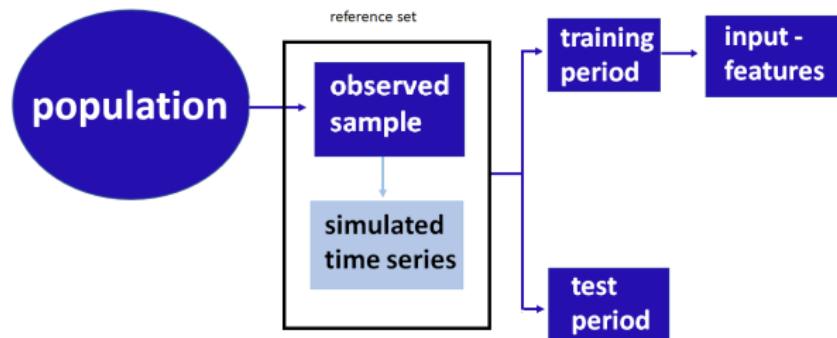
# FFORMS: reference set



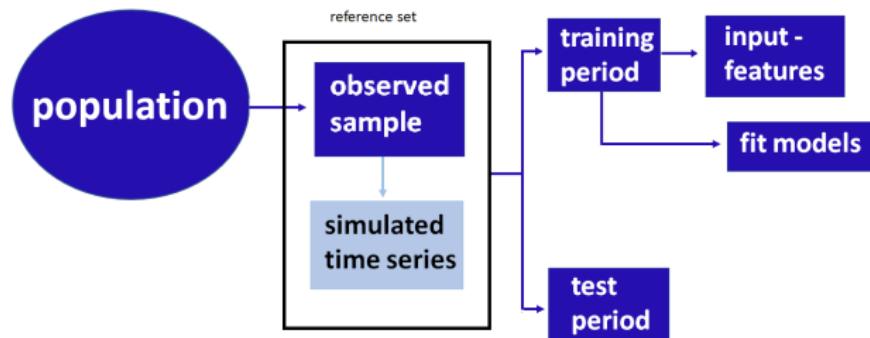
# FFORMS: Meta-data



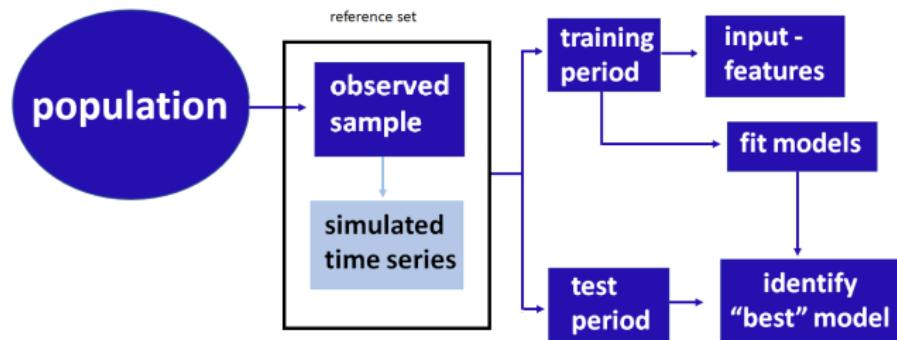
# FFORMS: Meta-data



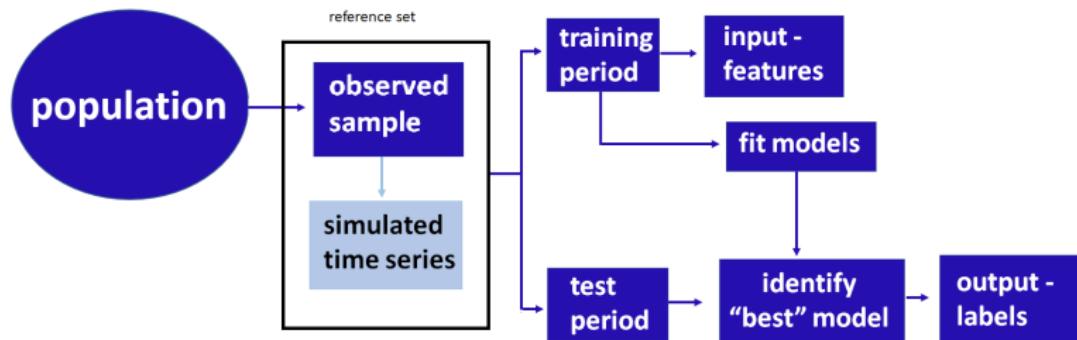
# FFORMS: Meta-data



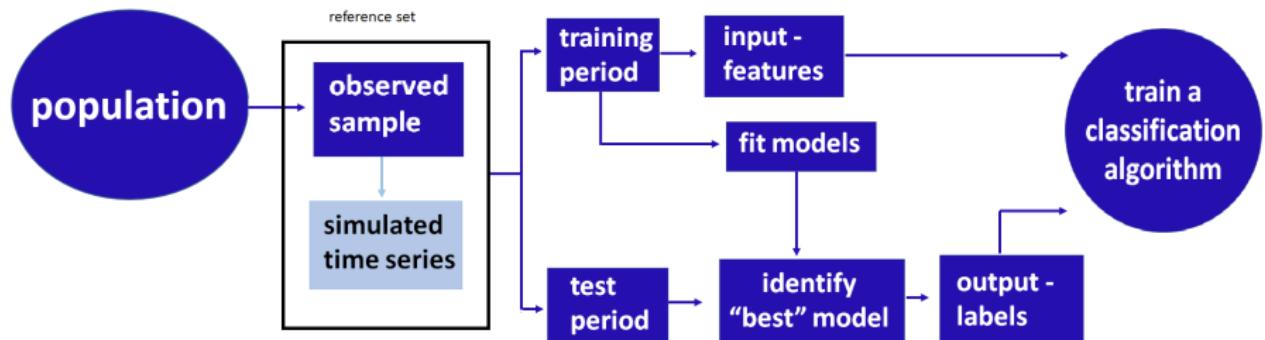
# FFORMS: Meta-data



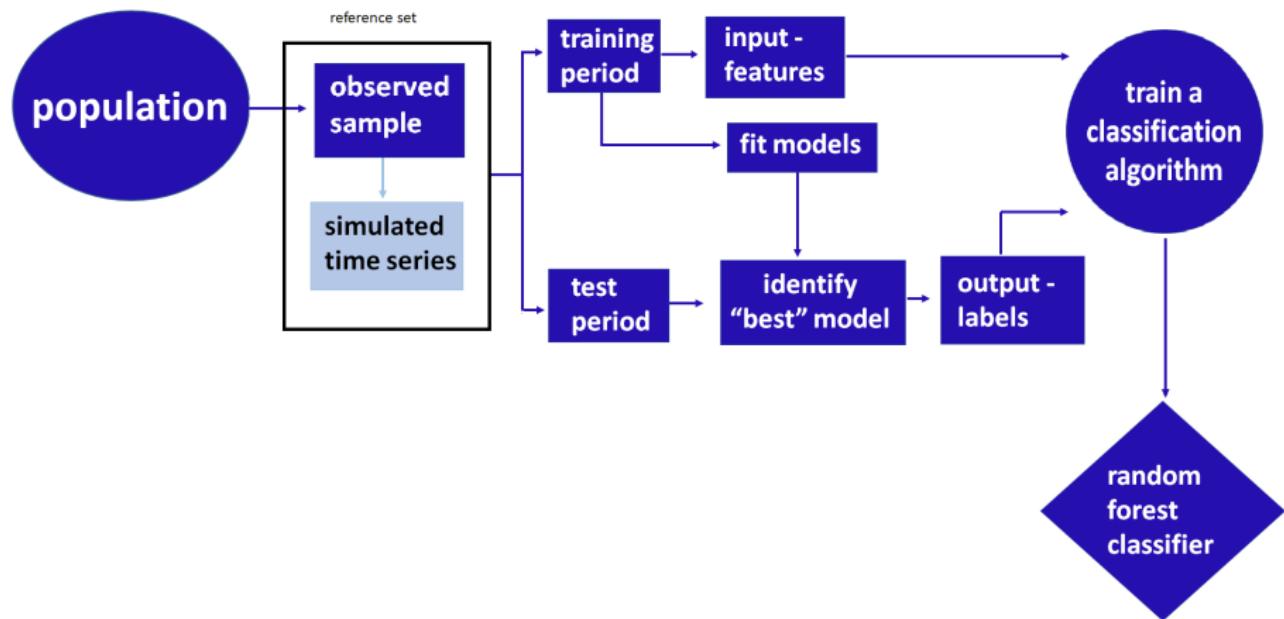
# FFORMS: Meta-data



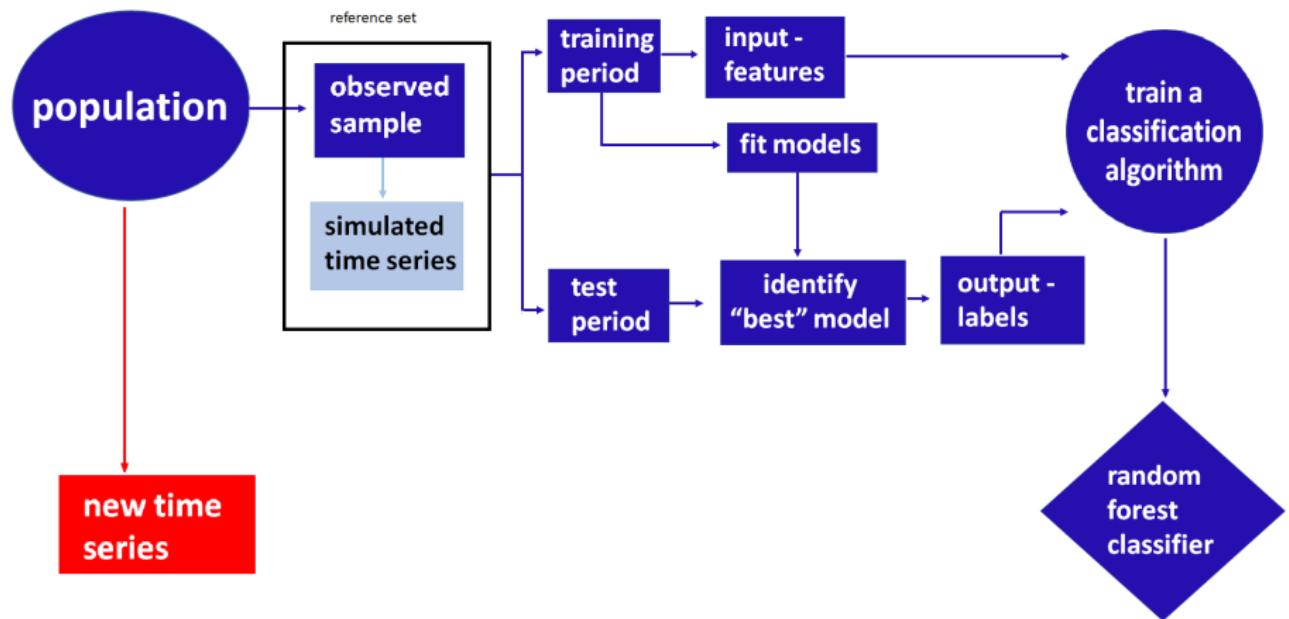
# FFORMS: Meta-data



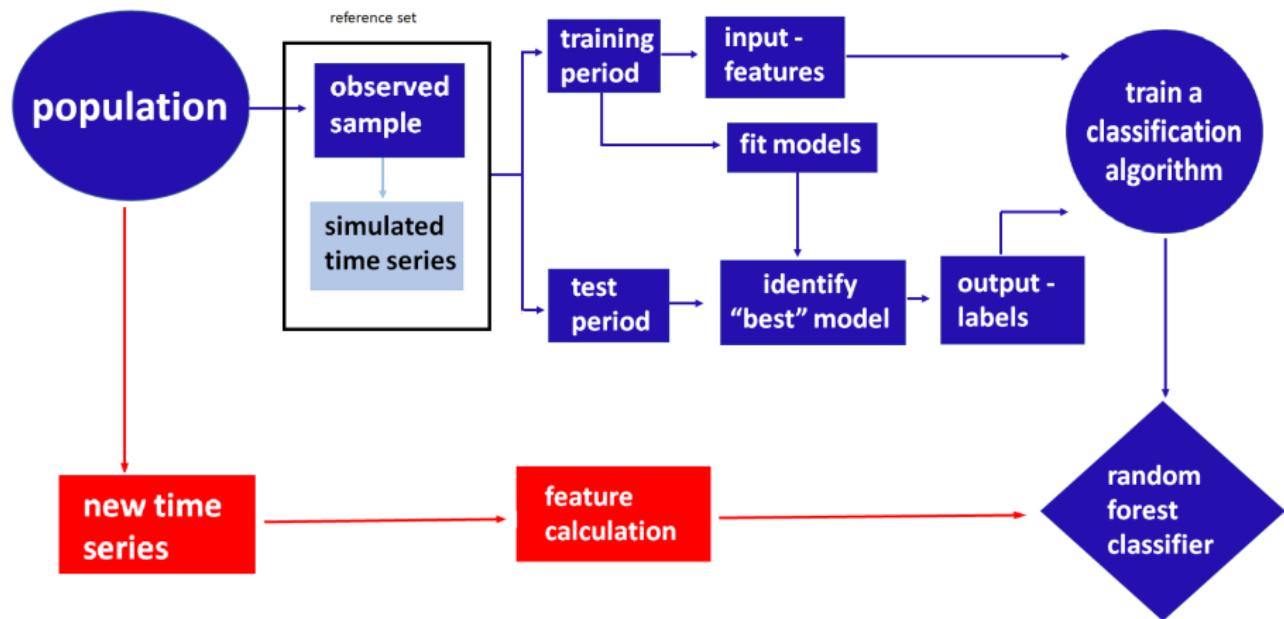
# FFORMS: Random-forest classifier



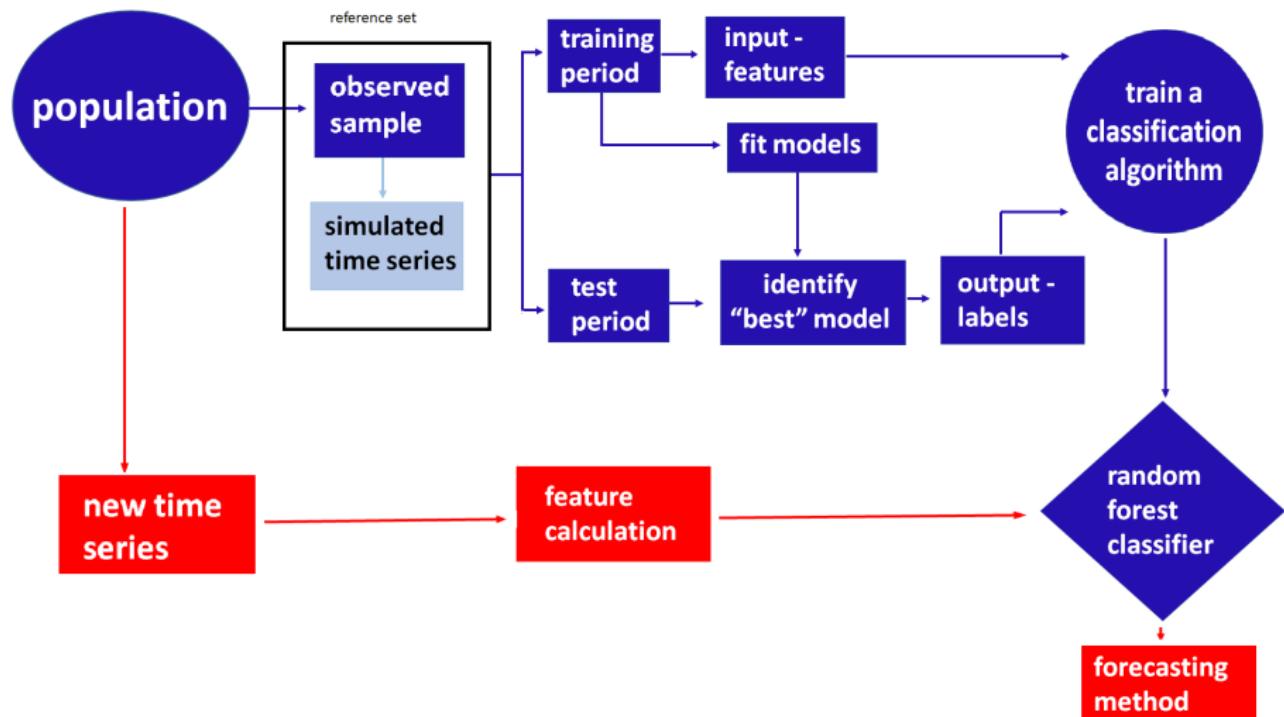
# FFORMS: “online” part of the algorithm



# FFORMS: “online” part of the algorithm



# FFORMS: “online” part of the algorithm



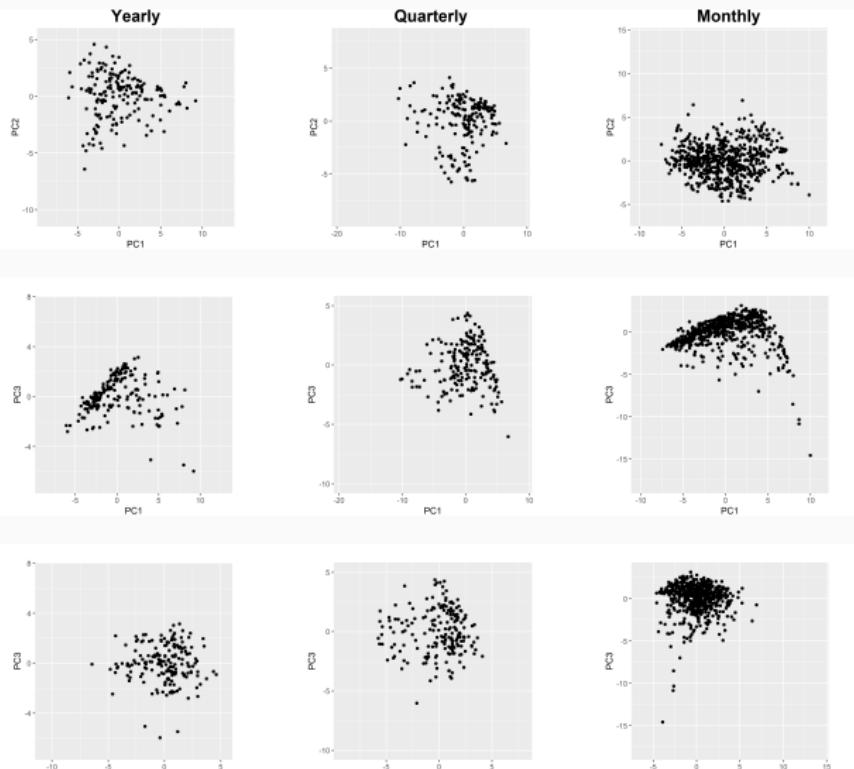
# Application to M competition data

- 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

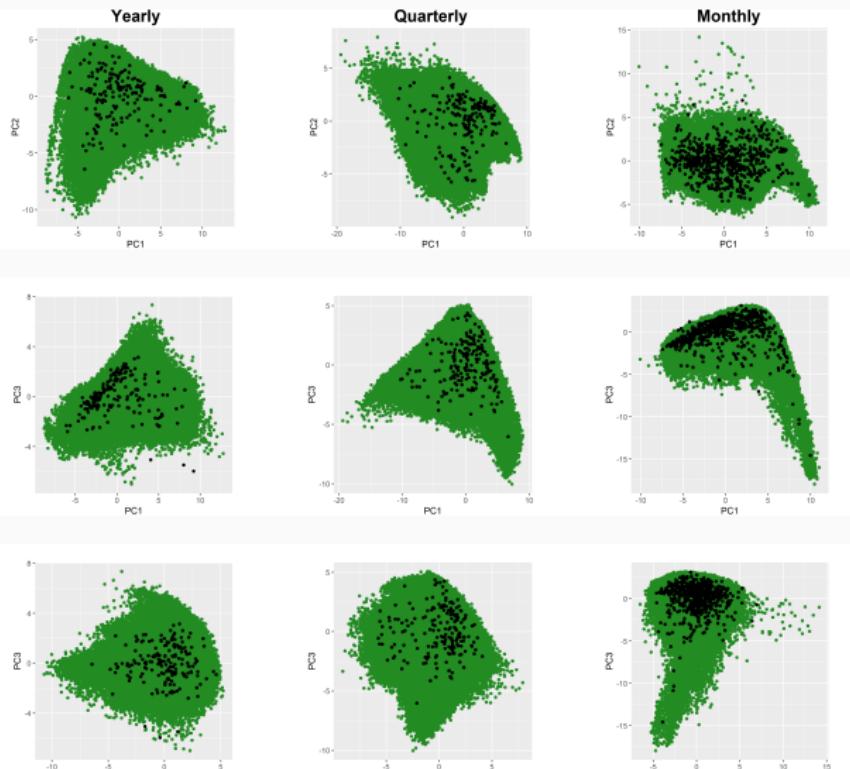
# Experiment 1: Distribution of time series in the PCA space

observed - M1



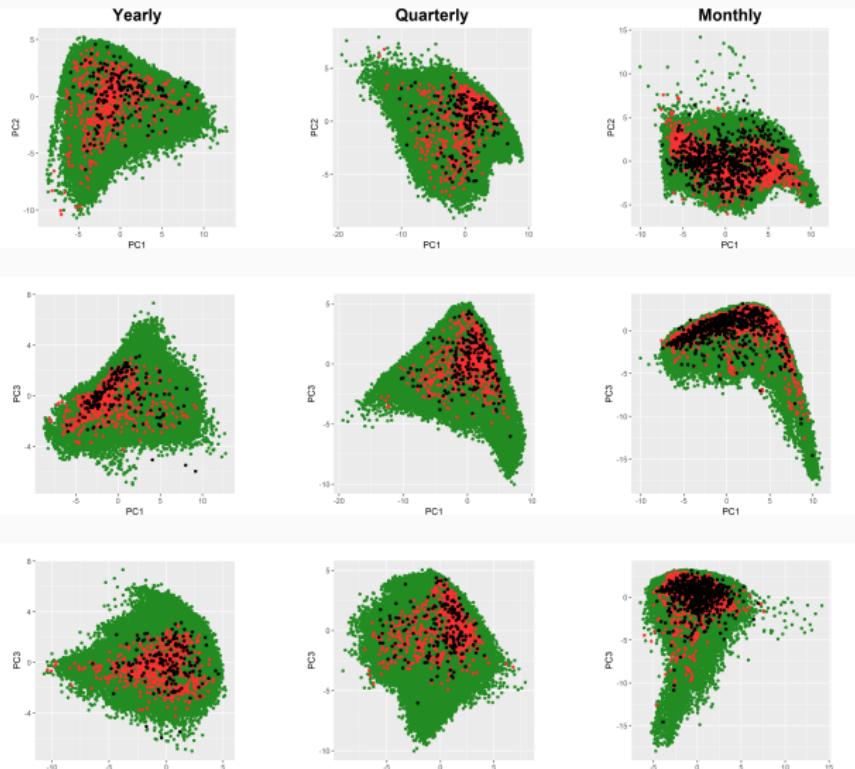
# Experiment 1: Distribution of time series in the PCA space

observed - M1 simulated



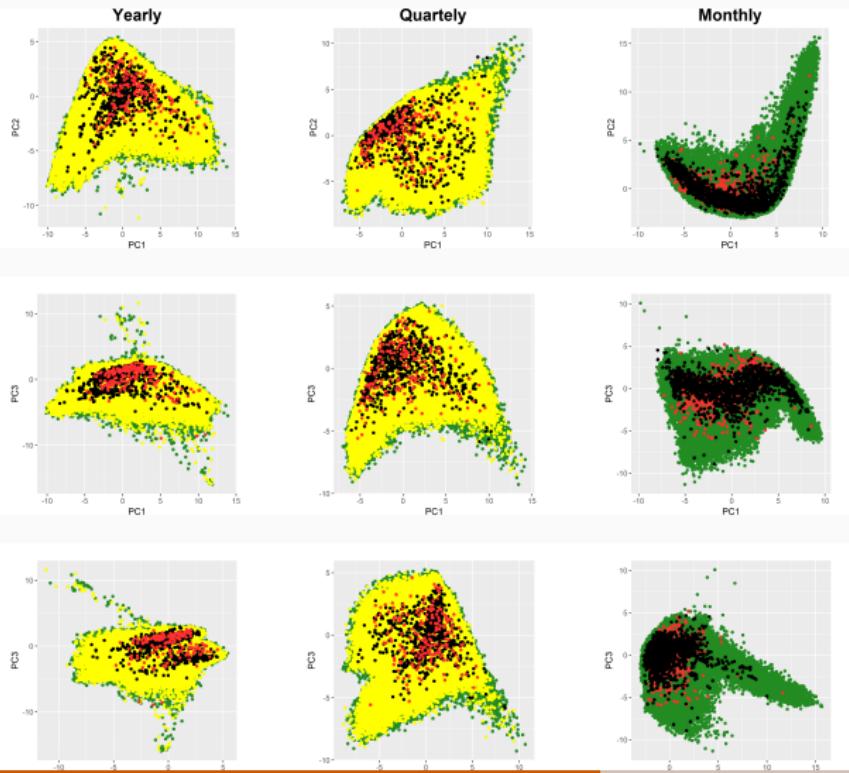
# Experiment 1: Distribution of time series in the PCA space

observed - M1 simulated new - M3

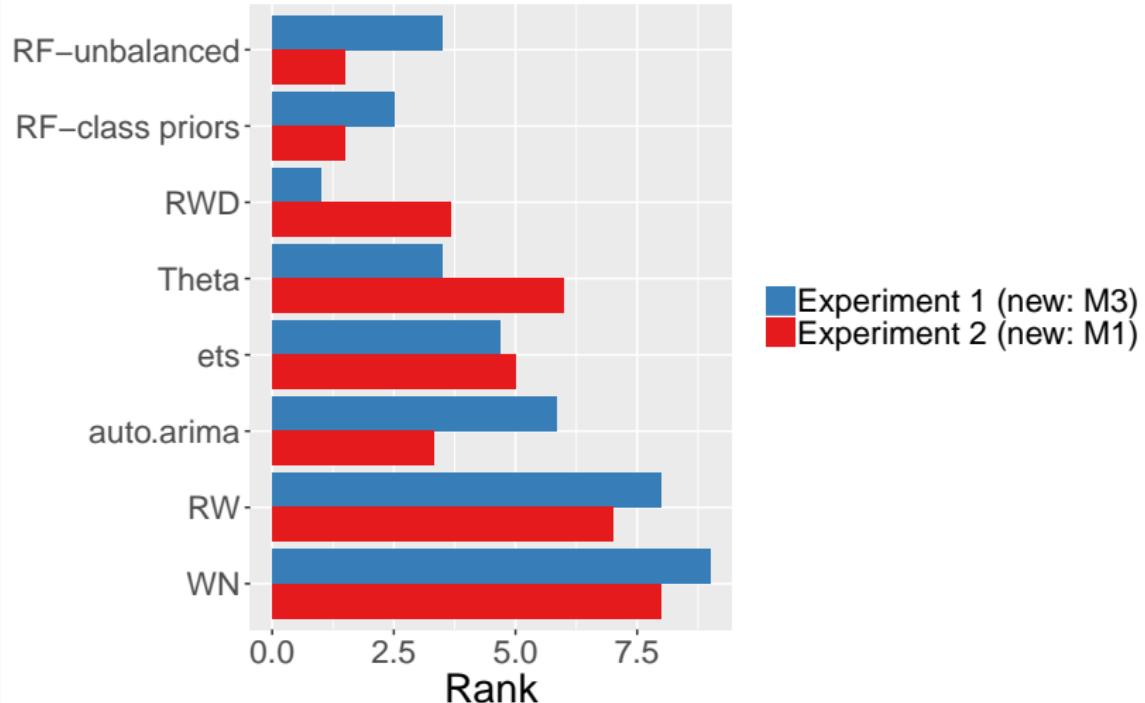


## Experiment 2: Distribution of time series in the PCA space

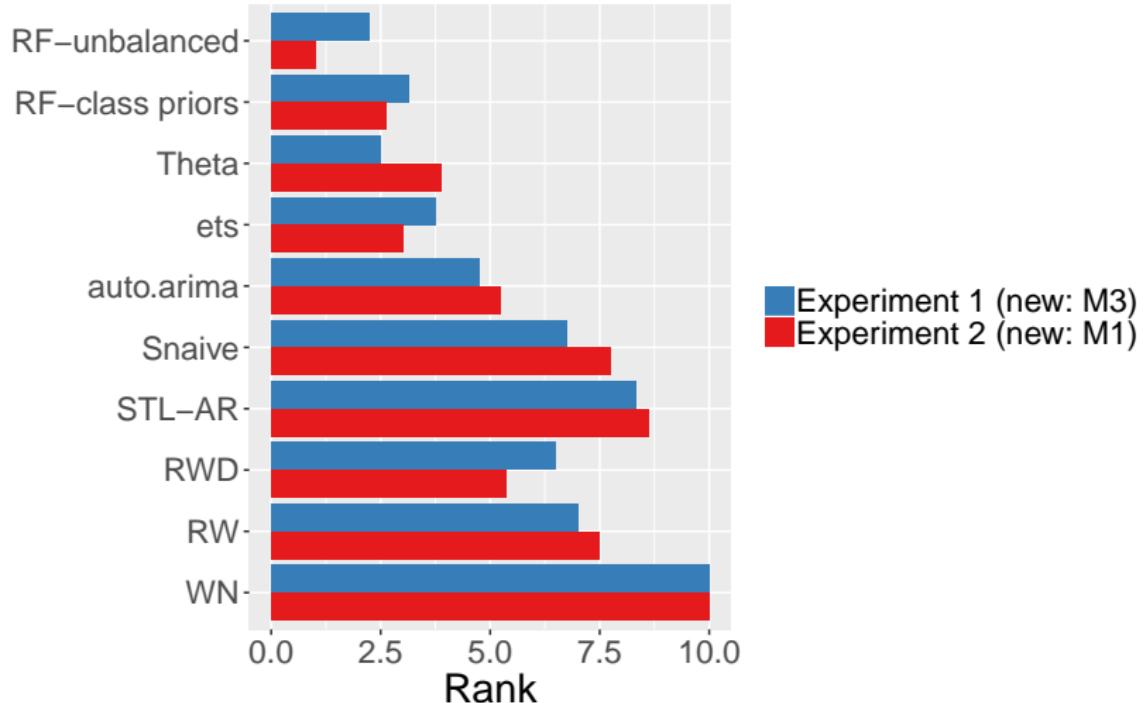
observed - M3   simulated   subset   new - M1



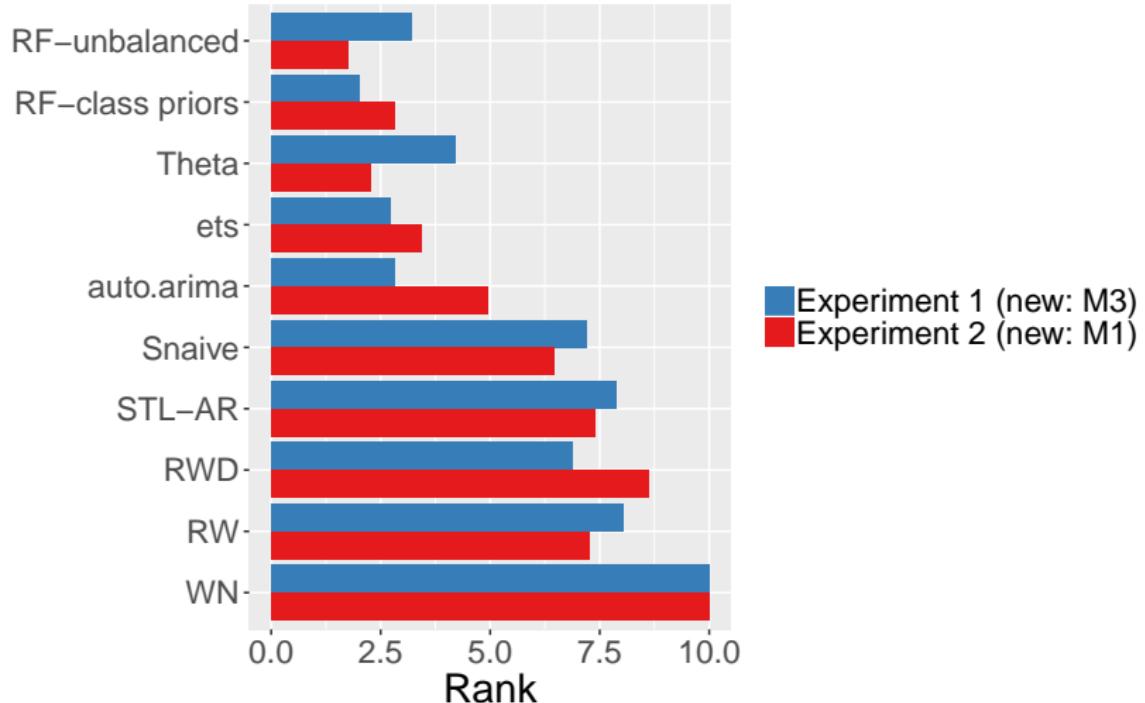
# Results: Yearly



# Results: Quarterly



# Results: Monthly



# M4 Competition: 2018

**M4COMPETITION** Forecast. Compete. Excel.

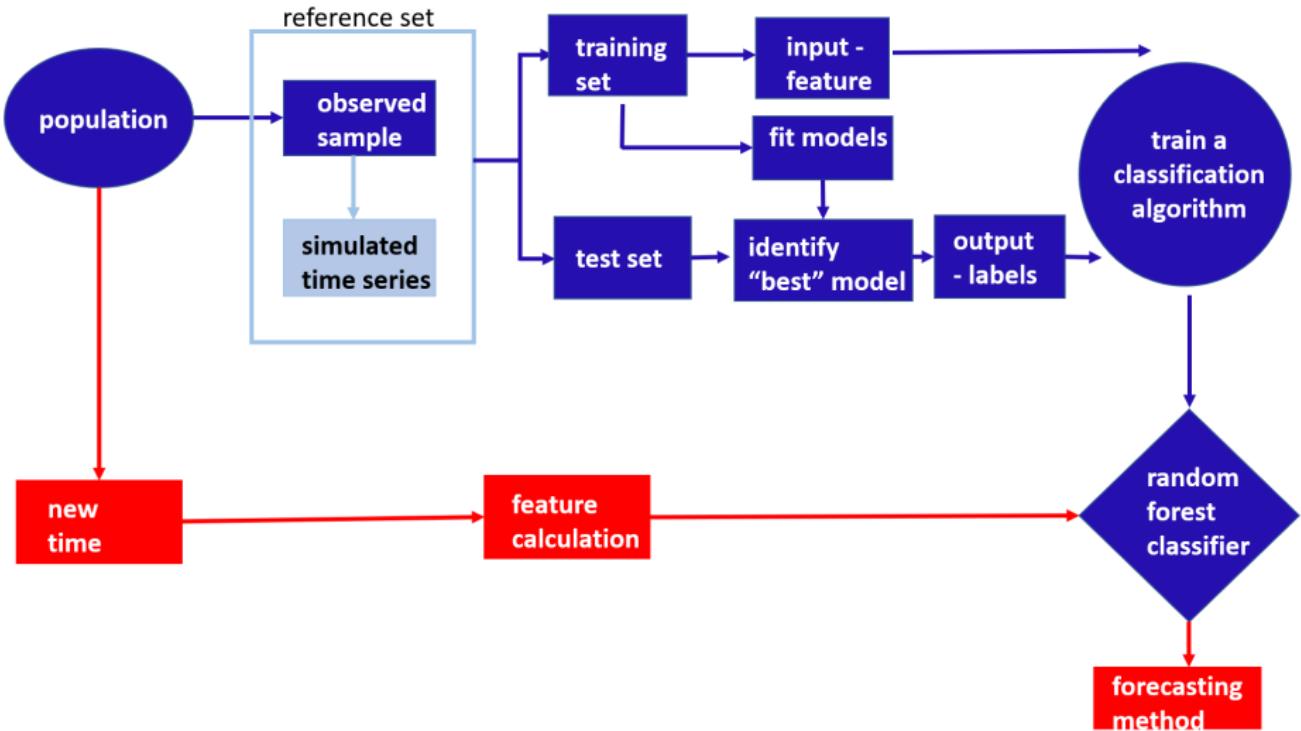


The M4 logo is centered on a grid background. The letters 'M' and '4' are large, with 'M' being white and '4' being black. Behind the logo, there are several faint, overlapping time series plots showing various trends and seasonal patterns.

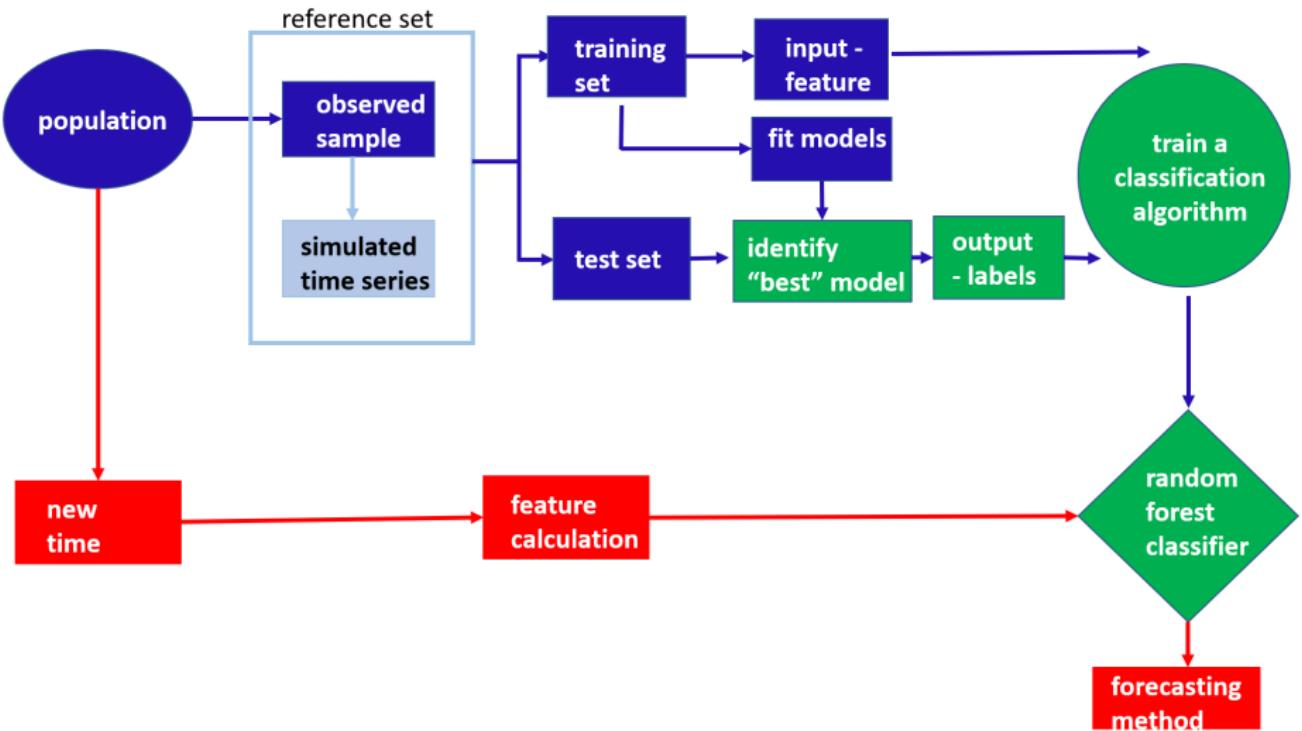
**M4**

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

# FFORMS: Feature-based FORecast Model Selection

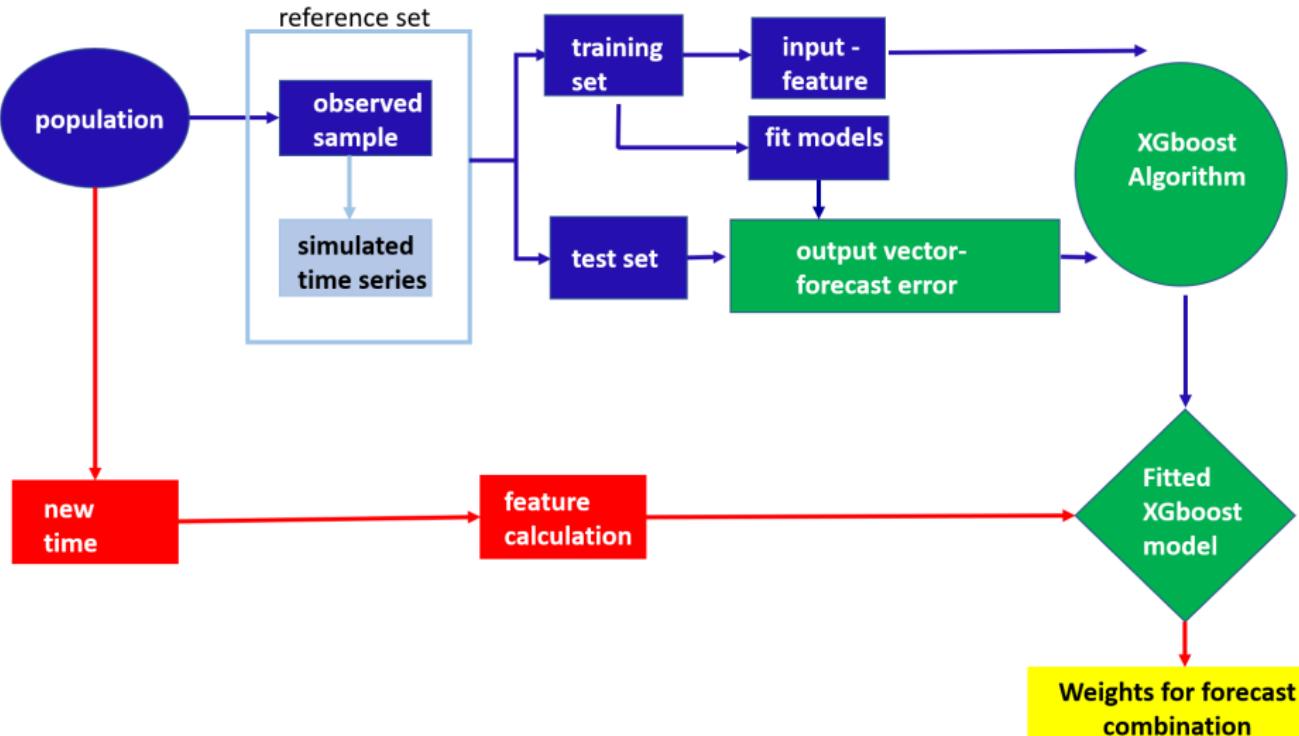


# FFORMS: Feature-based FORecast Model Selection



■ optimization criterion: classification accuracy

# FFORMA: Feature-based FORecast 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.

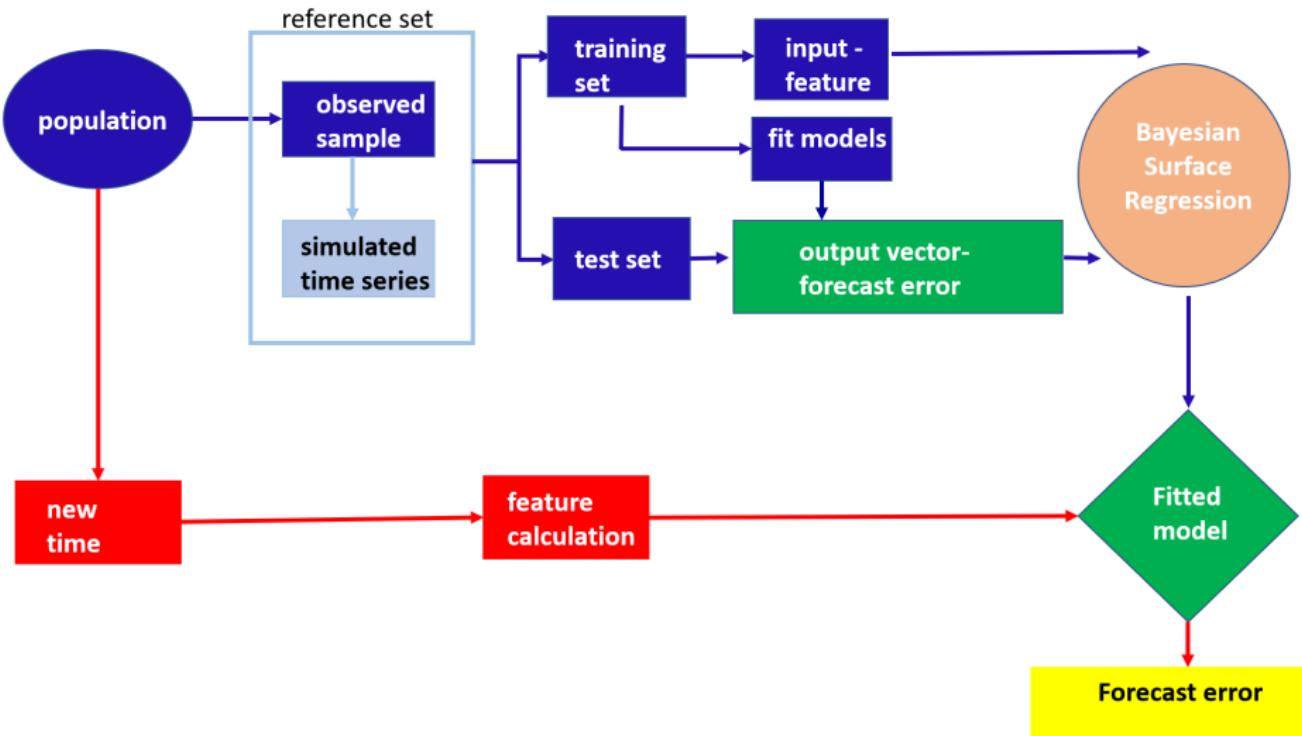
## FFORMA: Feature-based FOrecast Model Averaging

- 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 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
- Came second in the M4 competition

# FFORMPP: Feature-based FOrecast Model Performance Prediction



- use the minimum predicted MASE to select

## FFORMPP: Feature-based FOrecast Model Performance Prediction

- We use Efficient Bayesian Multivariate Surface Regression

## FFORMPP: Feature-based FOrecast Model Performance Prediction

- We use Efficient Bayesian Multivariate Surface Regression
- Y: forecast error of each method, we take the correlation structure between the errors into account.

- We use Efficient Bayesian Multivariate Surface Regression
- Y: forecast error of each method, we take the correlation structure between the errors into account.
- Why Efficient Bayesian Multivariate Surface Regression?
  - ▶ handles interactions and nonlinear relationships between features
  - ▶ allows the knot locations to move freely in the feature space, thus a less number of knots are usually used

# Application to M4 Competition

- Composition of the time series in the reference set and collection of new time series

Frequency	Reference set			New series M4
	M1	M3	Simulated	
Yearly	181	645	10000	23000
Quarterly	203	756	10000	24000
Monthly	617	1428	10000	48000
Weekly	-	-	10000	359
Daily	-	-	10000	4227
Hourly	-	-	10000	414

# Application to M4 Competition

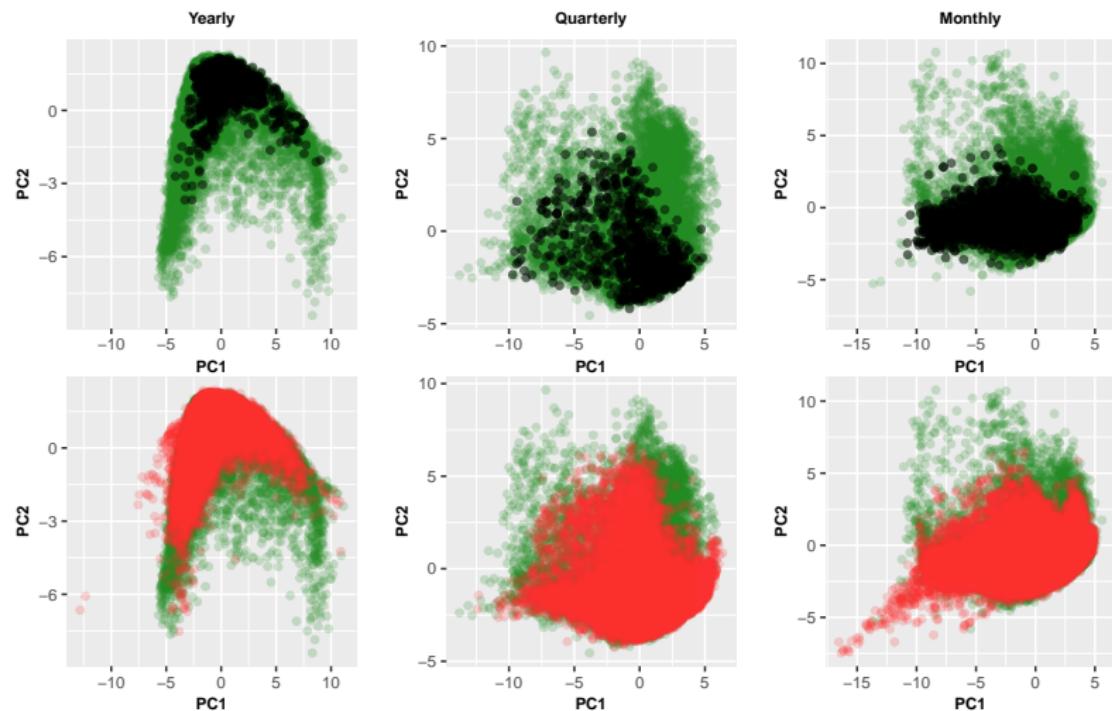
- Composition of the time series in the reference set and collection of new time series

Frequency	Reference set			New series M4
	M1	M3	Simulated	
Yearly	181	645	10000	23000
Quarterly	203	756	10000	24000
Monthly	617	1428	10000	48000
Weekly	-	-	10000	359
Daily	-	-	10000	4227
Hourly	-	-	10000	414

## Simulated data

- Augmenting the observed sample with simulated time series from mixture autoregressive(MAR) models
- Proposed by Kang, Hyndman and Li (2018), Efficient generation of time series with diverse and controllable characteristics
- R package:  
<https://github.com/ykang/tsgeneration>

# Distribution of time series in the PCA space



■ black: observed, green-simulated, orange-M4

# R packages

# References