MONASH University

MONASH BUSINESS SCHOOL

Feature-based forecasting algorithms for large collections of time series

Rob J Hyndman

25 January 2019

Outline

1. Makridakis competitions
2. Time series features
3. Feature-based forecasting
4. FFORMS: Feature-based forecast model selection
5. FFORMA: Feature-based forecast model averaging

Outline

1. Makridakis competitions
2. Time series features
3. Feature-based forecasting
4. FFORMS: Feature-based forecast model selection
5. FFORMA: Feature-based forecast model averaging

Makridakis and Hibon (1979)

J. R. Statist. Soc. A (1979),
142, *Part* 2, *pp.* 97–145

Accuracy of Forecasting: An Empirical Investigation

By Spyros Makridakis and Michèle Hibon

INSEAD—The European Institute of Business Administration

[Read before the Royal Statistical Society on Wednesday, December 13th, 1978, the President, Sir Claus Moser in the Chair]

Summary

In this study, the authors used 111 time series to examine the accuracy of various forecasting methods, particularly time-series methods. The study shows, at least for time series, why some methods achieve greater accuracy than others for different types of data. The authors offer some explanation of the seemingly conflicting conclusions of past empirical research on the accuracy of forecasting. One novel contribution of the paper is the development of regression equations expressing accuracy as a function of factors such as randomness, seasonality, trend-cycle and the number of data points describing the series. Surprisingly, the study shows that for these 111 series simpler methods perform well in comparison to the more complex and statistically sophisticated ARMA models.

Keywords: FORECASTING; TIME SERIES; FORECASTING ACCURACY

0. Introduction

The ultimate test of any forecast is whether or not it is capable of predicting future events accurately. Planners and decision makers have a wide choice of ways to forecast, ranging from purely intuitive or judgemental approaches to highly structured and complex quantitative methods. In between, there are innumerable possibilities that differ in their underlying philosophies, their cost, their complexity and their accuracy. Unfortunately, since information about these differences is not usually available, objective selection among forecasting methods

Makridakis and Hibon (1979)

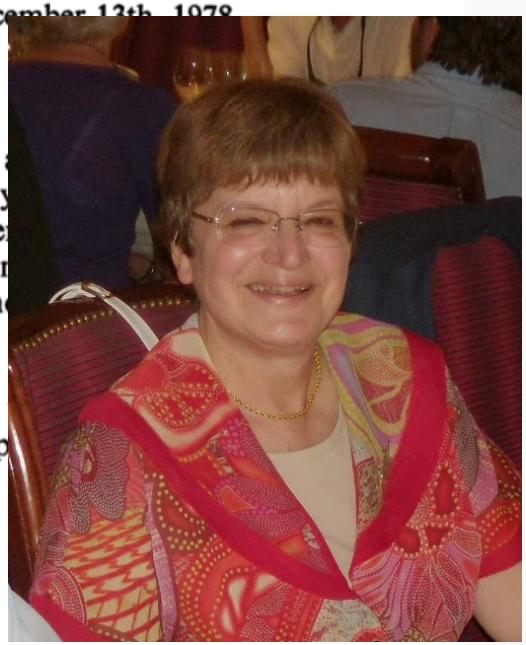
J. R. Statist. Soc. A (1979),
142, *Part* 2, *pp.* 97–145

Accuracy of Forecasting: An Empirical Investigation

By Spyros Makridakis and Michèle Hibon

INSEAD—The European Institute of Business Administration

[Read before the Royal Statistical Society on Wednesday, December 13th, 1978
President, Sir Claus Moser in the Chair]



Summary

s used 111 time series to examine the
icularly time-series methods. The study
hods achieve greater accuracy than othe
r some explanation of the seemingly co
on the accuracy of forecasting. One n
ent of regression equations expressing
ndomness, seasonality, trend-cycle and
es. Surprisingly, the study shows that
well in comparison to the more comp
ls.

SERIES; FORECASTING ACCURACY

0. Introduction

The ultimate test of any forecast is whether or not it is capable of predicting future events accurately. Planners and decision makers have a wide choice of ways to forecast, ranging from purely intuitive or judgemental approaches to highly structured and complex quantitative methods. In between, there are innumerable possibilities that differ in their underlying philosophies, their cost, their complexity and their accuracy. Unfortunately, since information about these differences is not usually available, objective selection among forecasting methods

Makridakis and Hibon (1979)

This was the first large-scale empirical evaluation of time series forecasting methods.

As a result of this paper, researchers started to:

- consider how to automate forecasting methods;
- study what methods give the best forecasts;
- be aware of the dangers of over-fitting;
- treat forecasting as a different problem from time series analysis.

Makridakis and Hibon (1979)

This was the first large-scale empirical evaluation of time series forecasting methods.

As a result of this paper, researchers started to:

- consider how to automate forecasting methods;
- study what methods give the best forecasts;
- be aware of the dangers of over-fitting;
- treat forecasting as a different problem from time series analysis.

Makridakis & Hibon followed up with a new competition in 1982.

M competition

Journal of Forecasting, Vol. 1, 111–153 (1982)

The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition

S. MAKRIDAKIS
INSEAD, Fontainebleau, France

A. ANDERSEN
University of Sydney, Australia

R. CARBONE
Université Laval, Quebec, Canada

R. FILDES
Manchester Business School, Manchester, England

M. HIBON
INSEAD, Fontainebleau, France

R. LEWANDOWSKI
Marketing Systems, Essen, Germany

J. NEWTON
E. PARZEN
Texas A & M University, Texas, U.S.A.

R. WINKLER
Indiana University, Bloomington, U.S.A.

ABSTRACT

In the last few decades many methods have become available for forecasting. As always, when alternatives exist, choices need to be made so that an appropriate forecasting method can be selected and used for the specific situation being considered. This paper reports the results of a forecasting competition that provides information to facilitate such choice. Seven experts in each of the 24 methods forecasted up to 1001 series for six up to eighteen time horizons. The results of the competition are presented in this paper whose purpose is to provide empirical evidence about *differences* found to exist among the various extrapolative (time series) methods used in the competition.

M competition

Journal of Forecasting, Vol. 1, 111–153 (1982)

The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition

S. MAKRIDAKIS
INSEAD, Fontainebleau, France

A. ANDERSEN
University of Sydney, Australia

R. CARBONE
Université Laval, Quebec, Canada

R. FILDES
Manchester Business School, Manchester, England

M. HIBON
INSEAD, Fontainebleau, France

R. LEWANDOWSKI
Marketing Systems, Essen, Germany

J. NEWTON
E. PARZEN
Texas A & M University, Texas, U.S.A.

R. WINKLER
Indiana University, Bloomington, U.S.A.

ABSTRACT

In the last few decades many methods have become available for forecasting. As always, when alternatives exist, choices need to be made so that an appropriate forecasting method can be selected and used for the specific situation being considered. This paper reports the results of a forecasting competition that provides information to facilitate such choice. Seven experts in each of the 24 methods forecasted up to 1001 series for six up to eighteen time horizons. The results of the competition are presented in this paper whose purpose is to provide empirical evidence about *differences* found to exist among the various extrapolative (time series) methods used in the competition.

M-competition

- 1001 series from demography, industry, economics. Annual, quarterly, monthly.
- Anyone could submit forecasts.
- Multiple forecast measures used.

M3 competition

ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

international journal of forecasting

www.elsevier.com/locate/ijforecast

The M3-Competition: results, conclusions and implications

Spyros Makridakis, Michèle Hibon*

INSEAD, Boulevard de Constance, 77305 Fontainebleau, France

Abstract

This paper describes the M3-Competition, the latest of the M-Competitions. It explains the reasons for conducting the competition and summarizes its results and conclusions. In addition, the paper compares such results/conclusions with those of the previous two M-Competitions as well as with those of other major empirical studies. Finally, the implications of these results and conclusions are considered, their consequences for both the theory and practice of forecasting are explored and directions for future research are contemplated.

Keywords: Comparative methods — time series: univariate; Forecasting competitions; M-Competition; Forecasting methods, Forecasting accuracy

M3 competition

"The M3-Competition is a final attempt by the authors to settle the accuracy issue of various time series methods... The extension involves the inclusion of more methods/ researchers (in particular in the areas of neural networks and expert systems) and more series."

- 3003 series
- All data from business, demography, finance and economics.
- Series length between 14 and 126.
- Either non-seasonal, monthly or quarterly.
- All time series positive.
- M&H claimed that the M3-competition supported the findings of their earlier work.

M4 competition

- January – May 2018
- 100,000 time series: yearly, quarterly, monthly, weekly, daily, hourly.
- Point forecast and prediction intervals assessed.
- Code must be public
- 248 registrations, 50 submissions.

M4 competition

- January – May 2018
- 100,000 time series: yearly, quarterly, monthly, weekly, daily, hourly.
- Point forecast and prediction intervals assessed.
- Code must be public
- 248 registrations, 50 submissions.

Winning methods

1. Hybrid of Recurrent Neural Network and Exponential Smoothing models
2. Forecast combination using xgboost to find weights

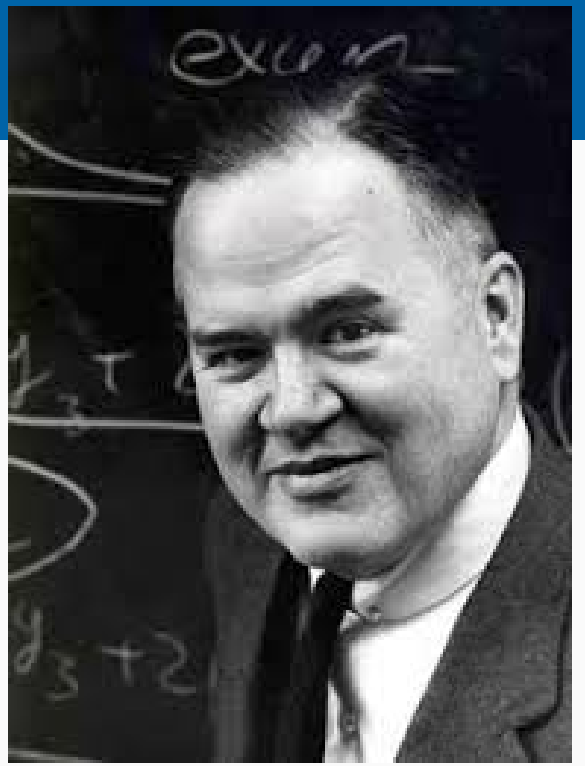
Outline

1. Makridakis competitions
2. Time series features
3. Feature-based forecasting
4. FFORMS: Feature-based forecast model selection
5. FFORMA: Feature-based forecast model averaging

Key idea

Cognostics

Computer-produced diagnostics (Tukey and Tukey, 1985).

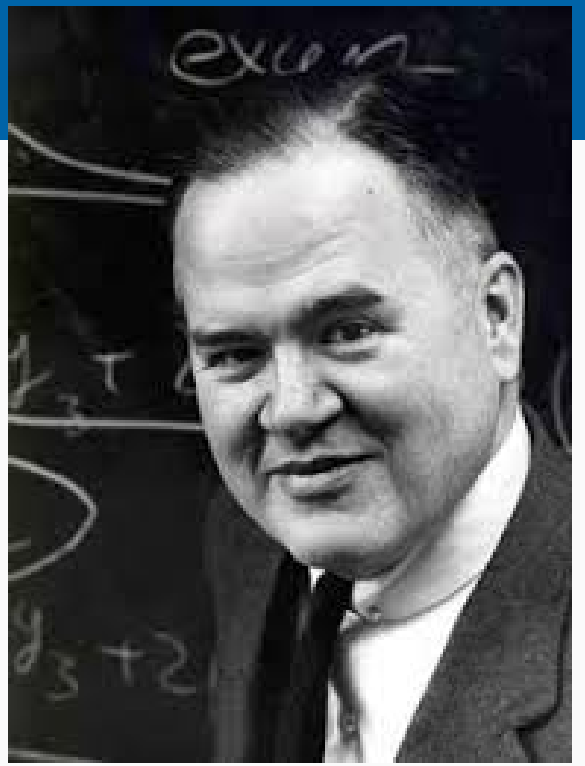


John W Tukey

Key idea

Cognostics

Computer-produced diagnostics (Tukey and Tukey, 1985).



John W Tukey

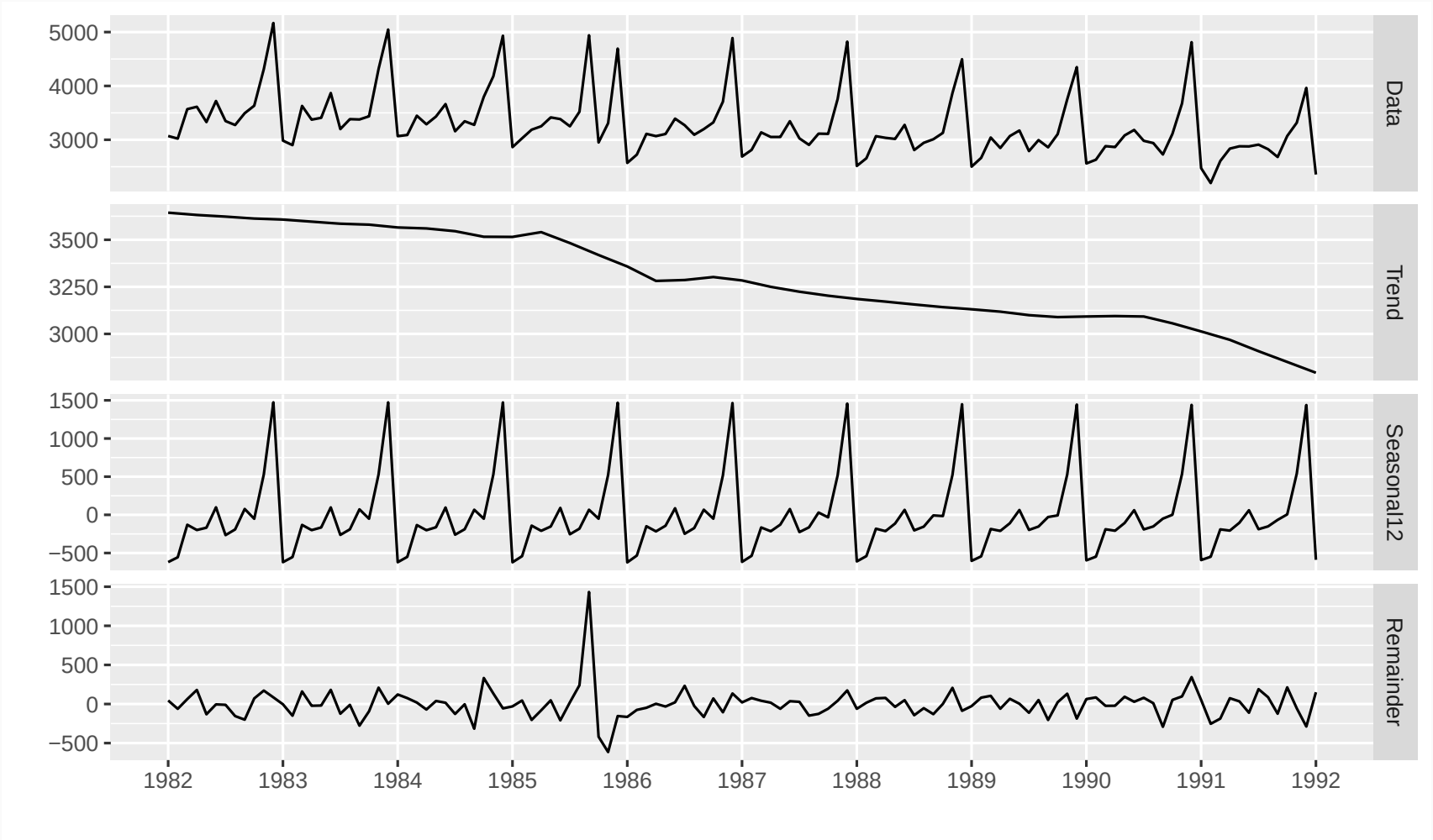
Examples for time series

- lag correlation
- size and direction of trend
- strength of seasonality
- timing of peak seasonality
- spectral entropy

Called “features” in the machine learning literature.

An STL decomposition: N2096

$Y_t = S_t + T_t + R_t$ S_t is periodic with mean 0



Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

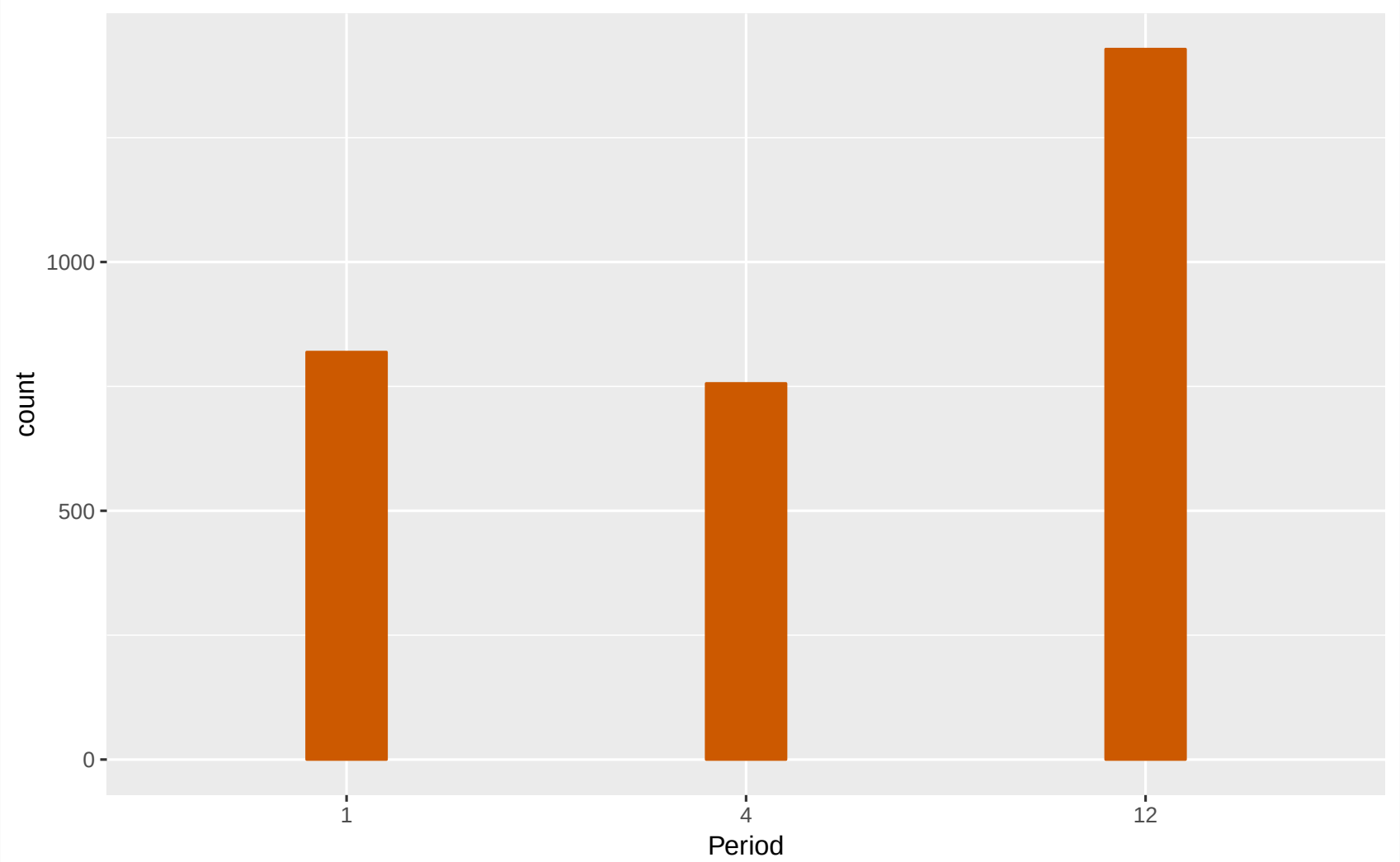
Candidate features

STL decomposition

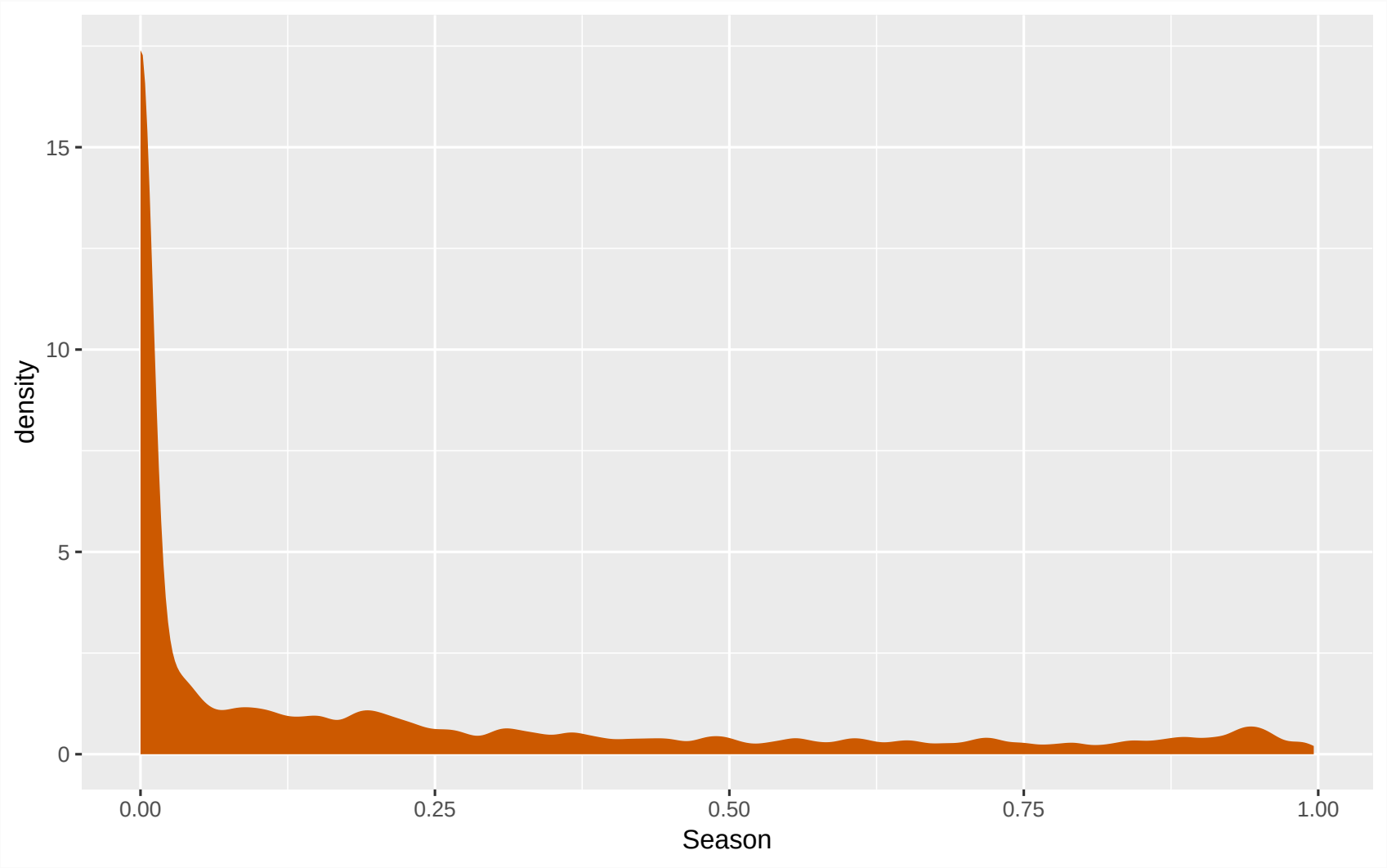
$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Autocorrelations of data $(Y_1, \ldots, Y_T)$
- Autocorrelations of data $(R_1, \ldots, R_T)$
- Strength of seasonality: $\max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - T_t)}\right)$
- Strength of trend: $\max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(Y_t - S_t)}\right)$
- Spectral entropy: $H = -\int_{-\pi}^{\pi} f_y(\lambda) \log f_y(\lambda) d\lambda$, where $f_y(\lambda)$ is spectral density of Y_t.
  Low values of H suggest a time series that is easier to forecast (more signal).
- Optimal Box-Cox transformation of data

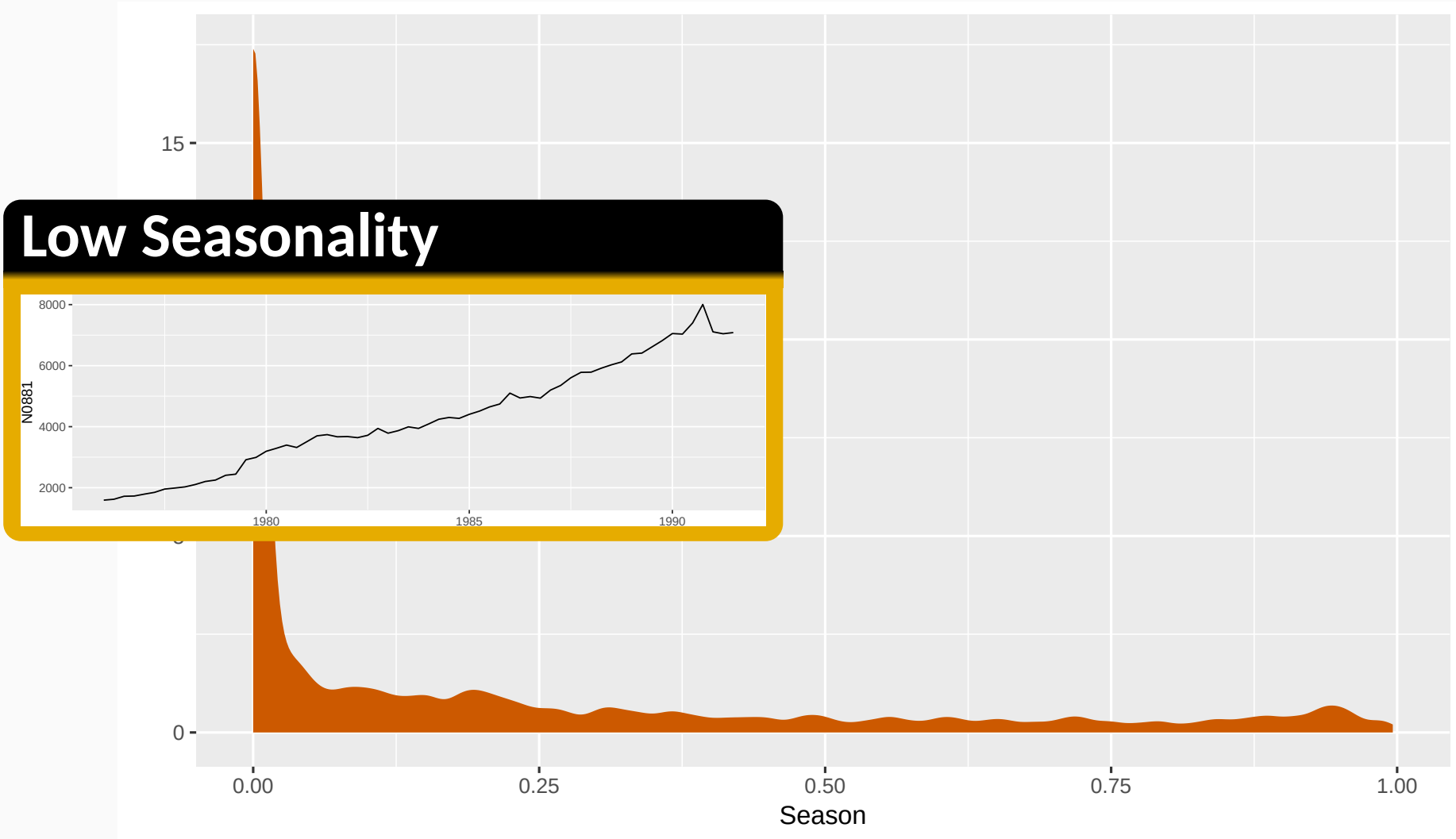
Distribution of Period for M3



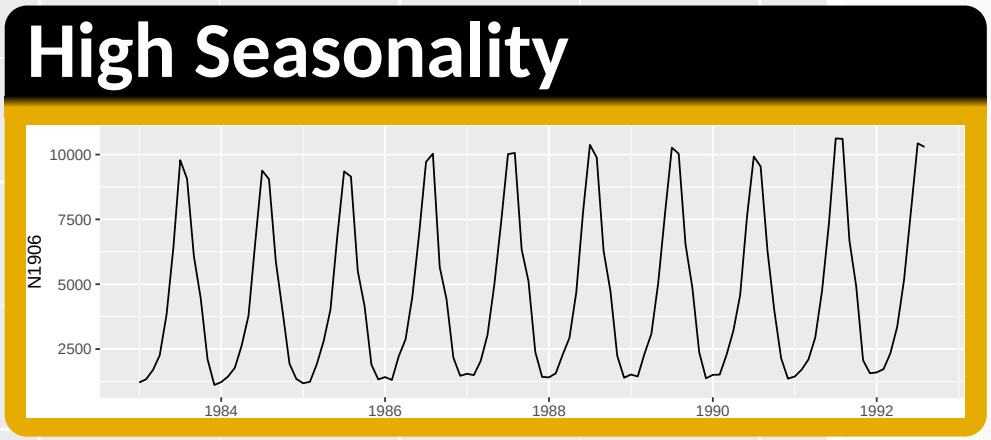
Distribution of Seasonality for M3



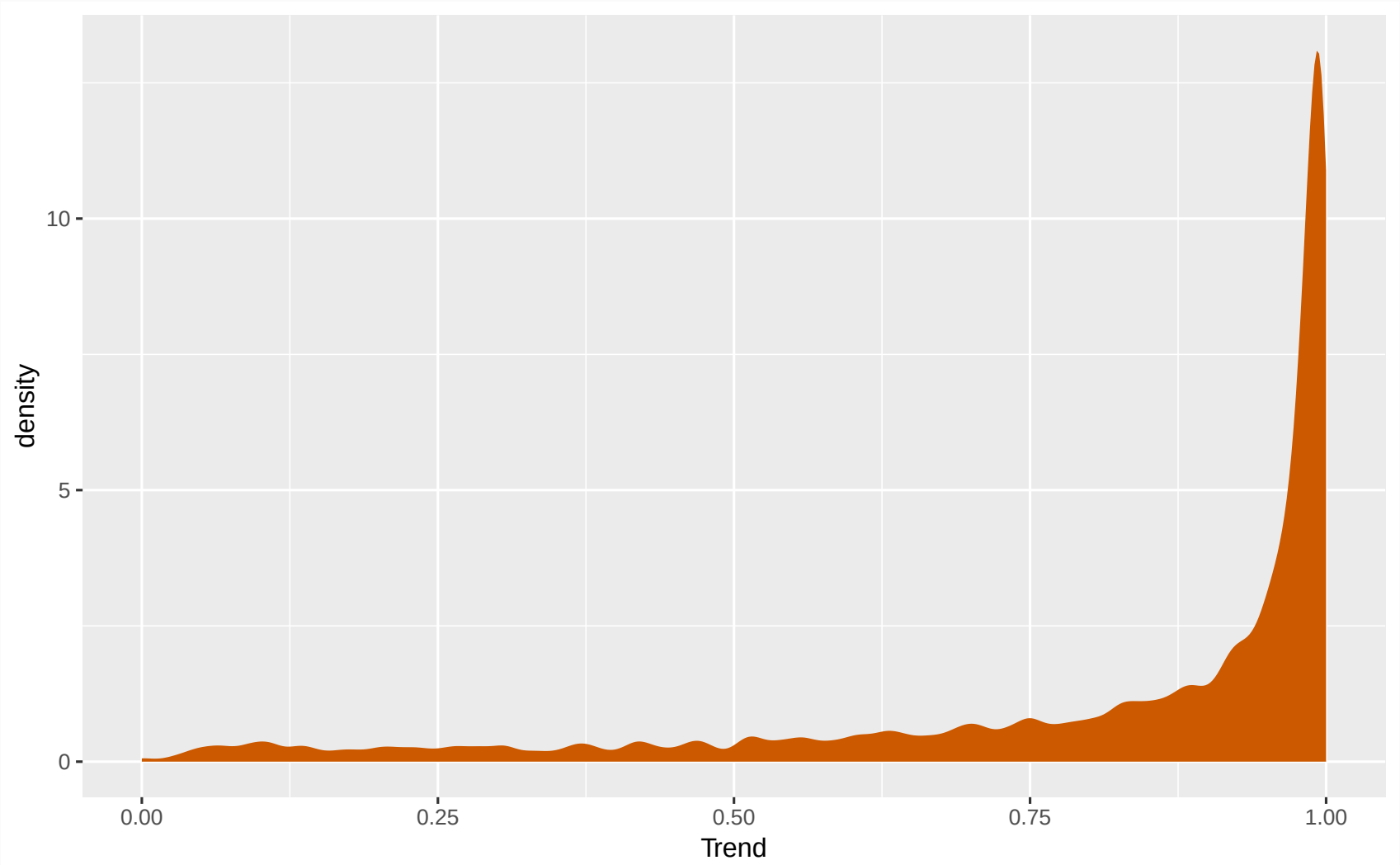
Distribution of Seasonality for M3



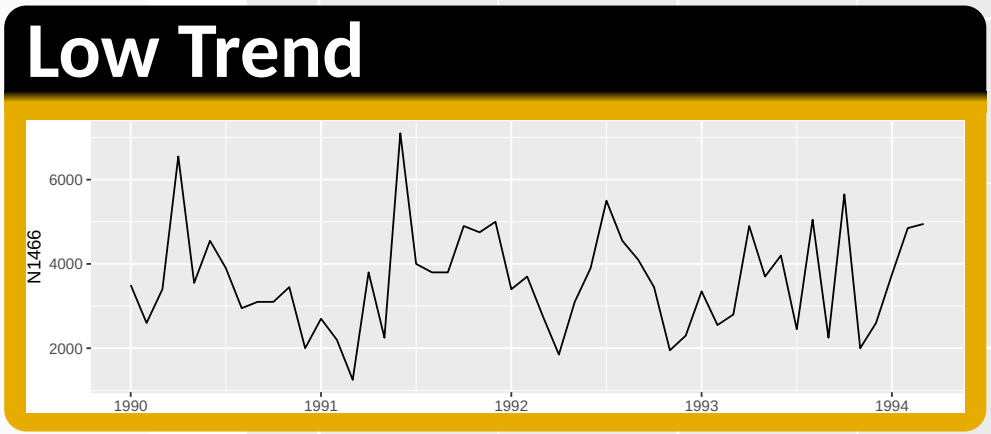
Distribution of Seasonality for M3



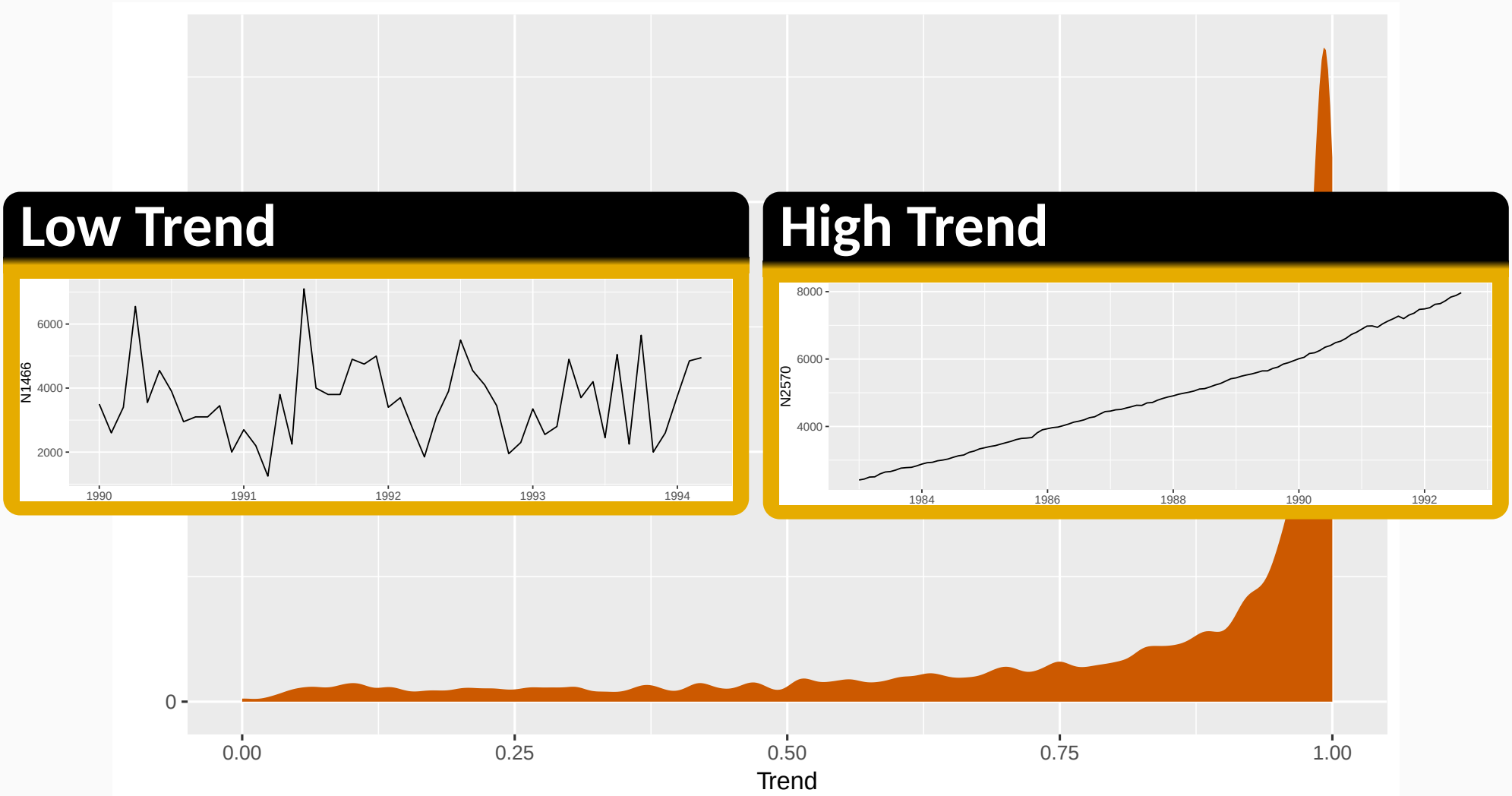
Distribution of Trend for M3



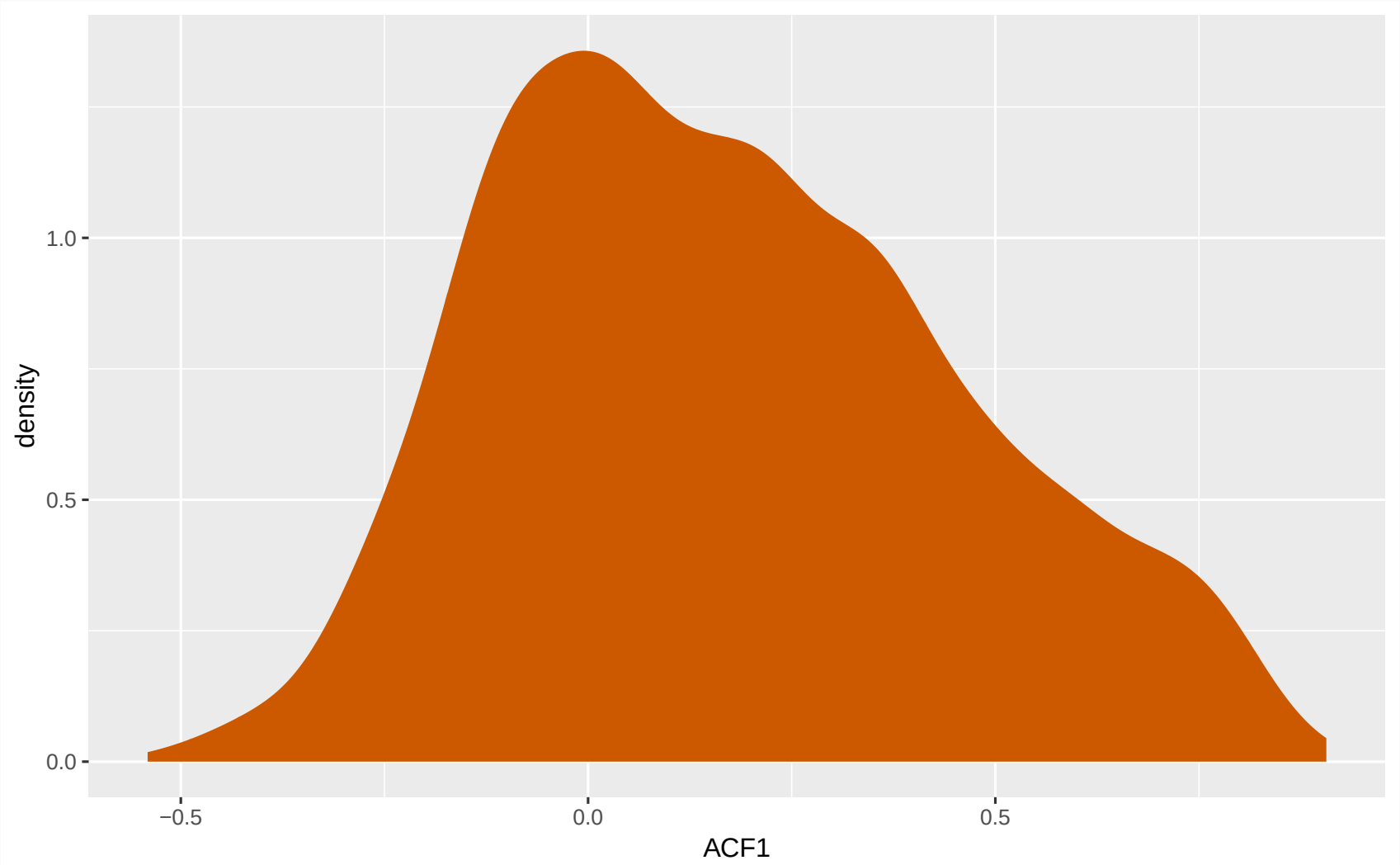
Distribution of Trend for M3



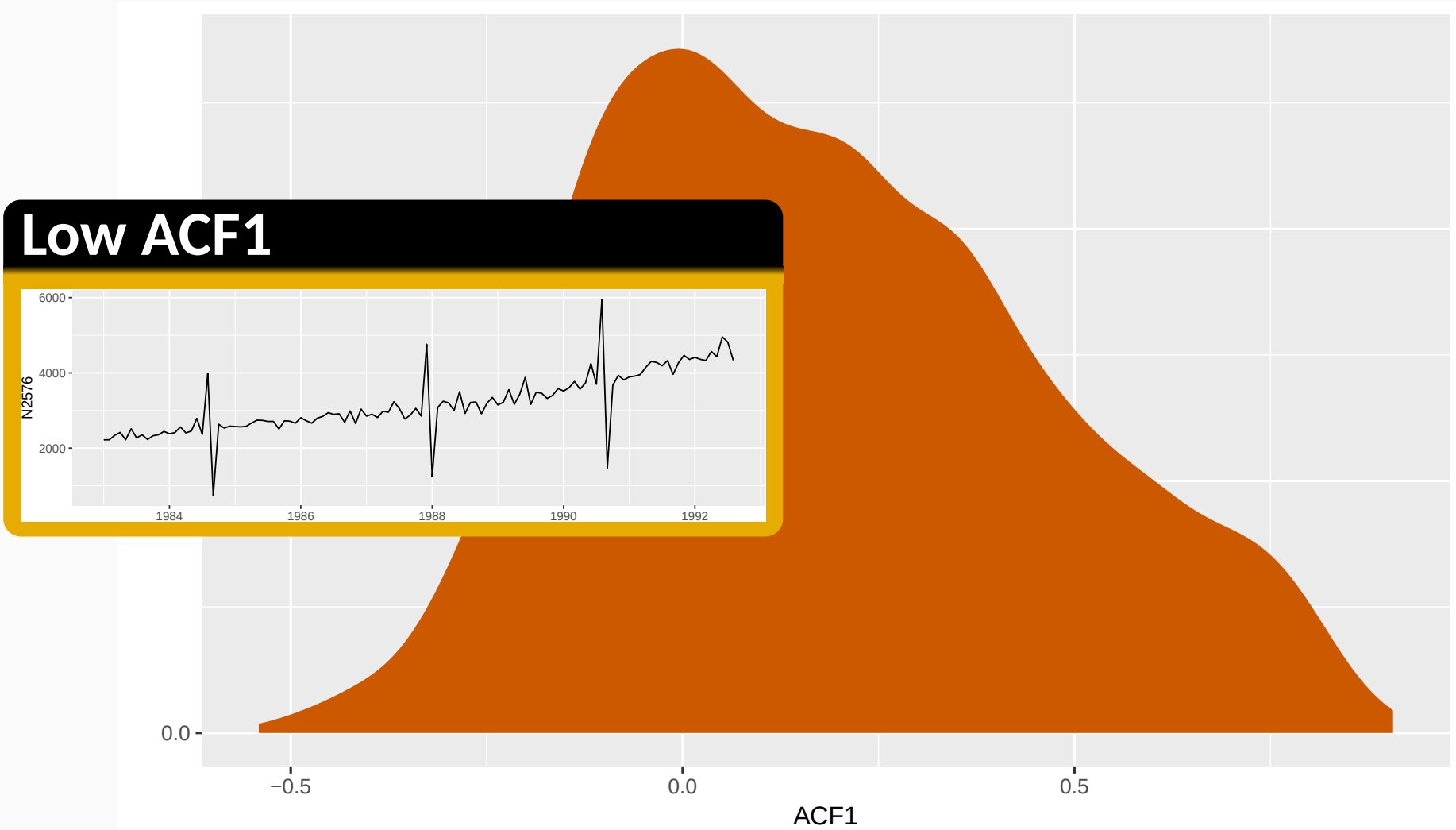
Distribution of Trend for M3



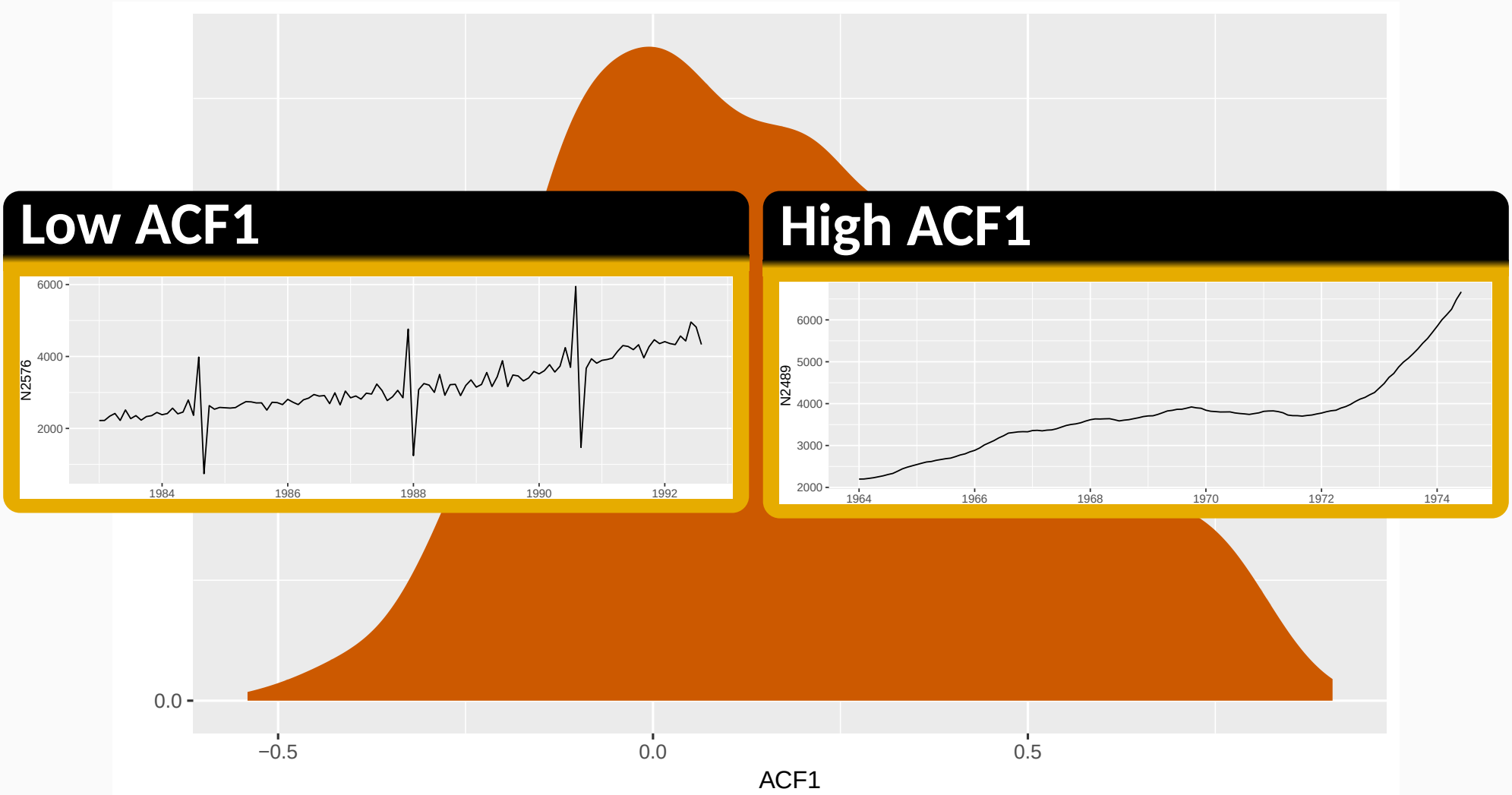
Distribution of Residual ACF1 for M3



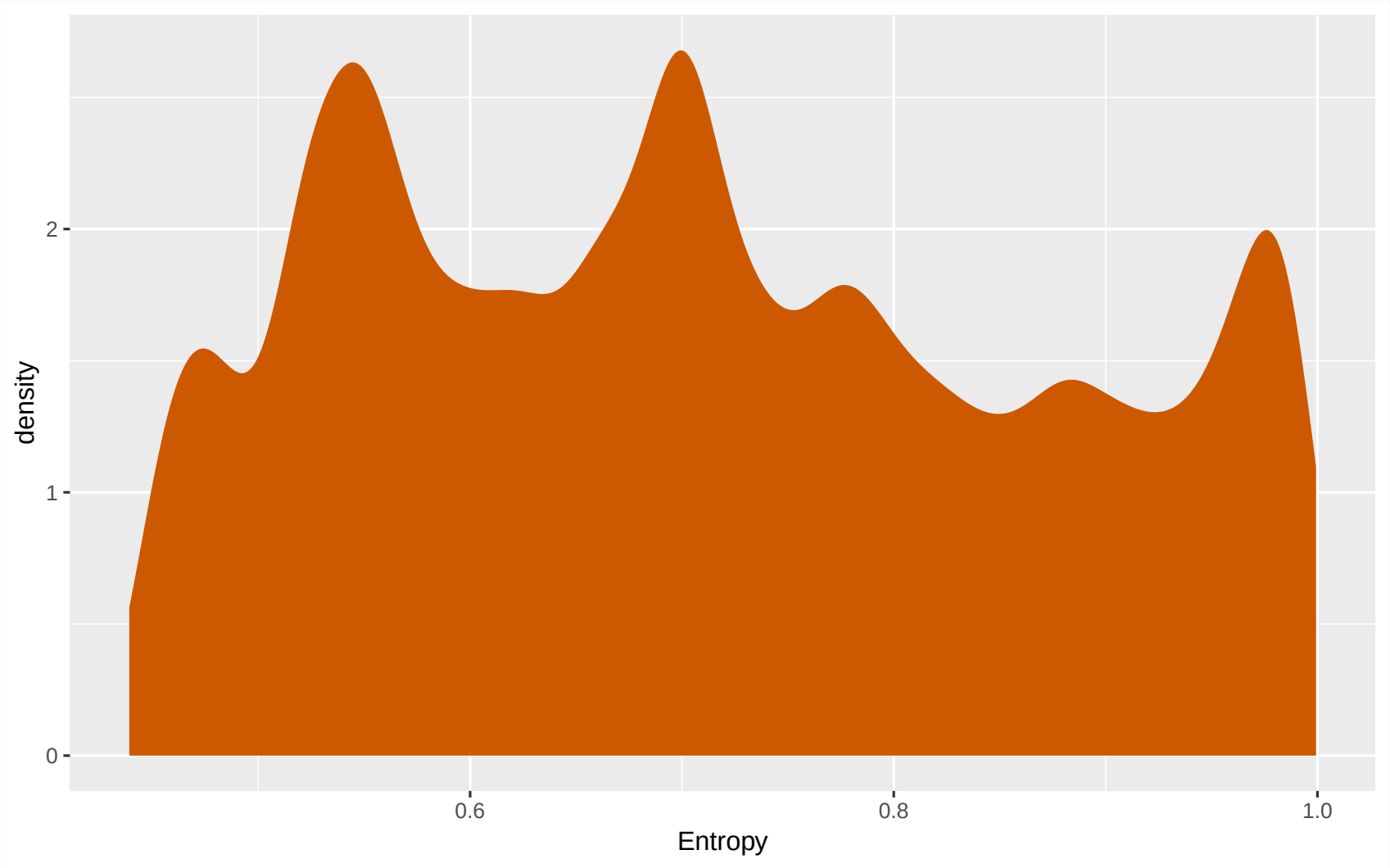
Distribution of Residual ACF1 for M3



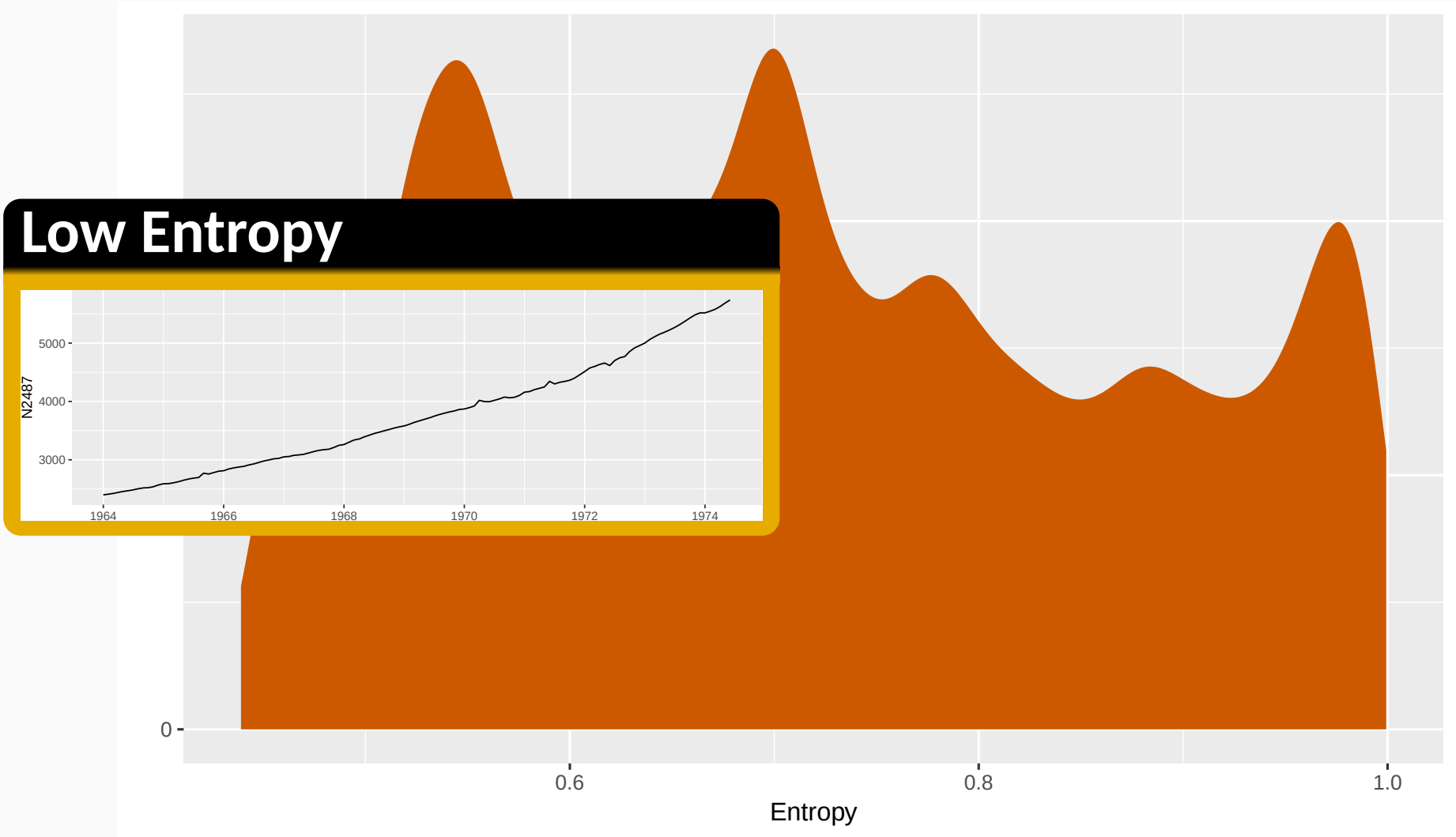
Distribution of Residual ACF1 for M3



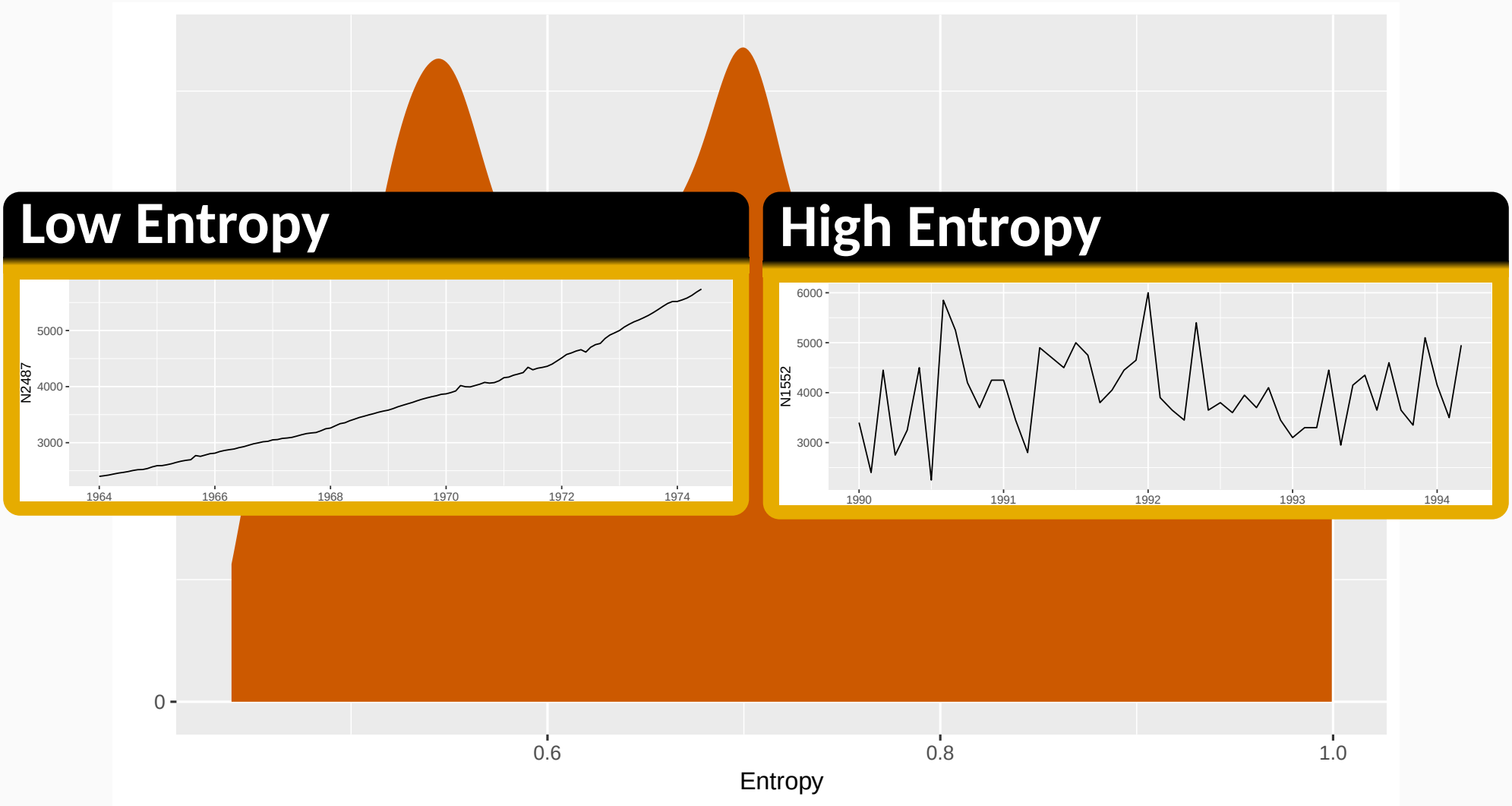
Distribution of Spectral Entropy for M3



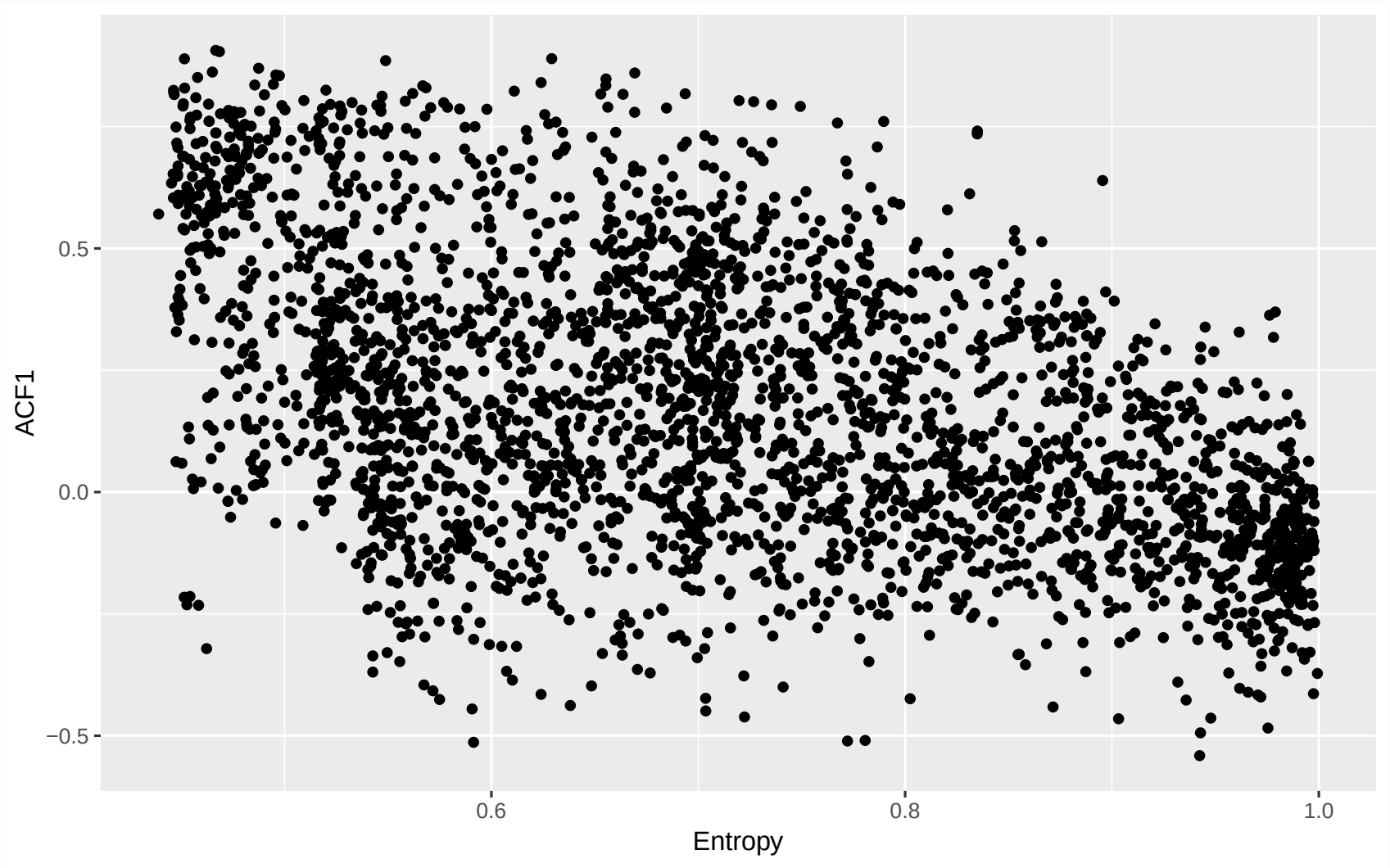
Distribution of Spectral Entropy for M3



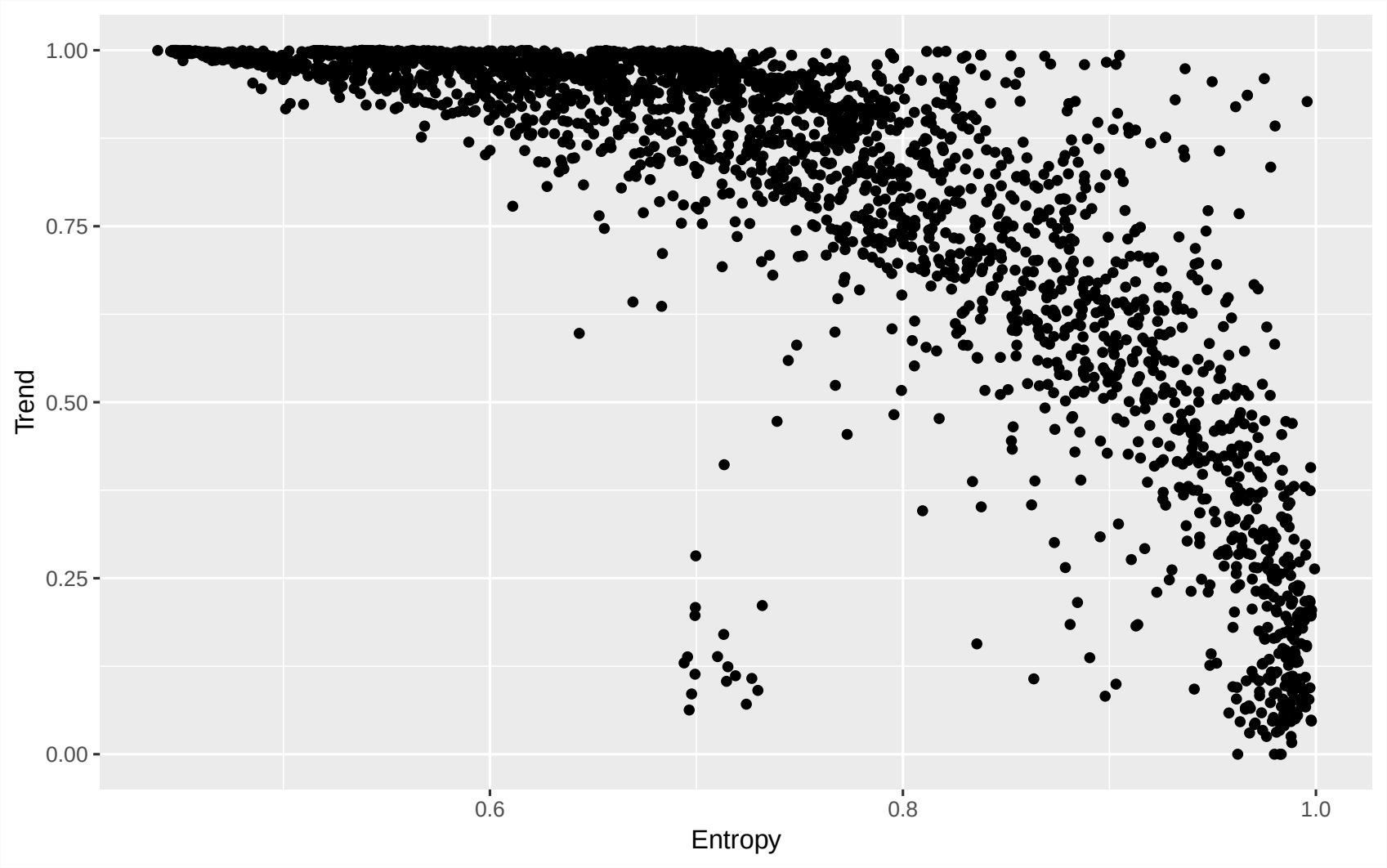
Distribution of Spectral Entropy for M3



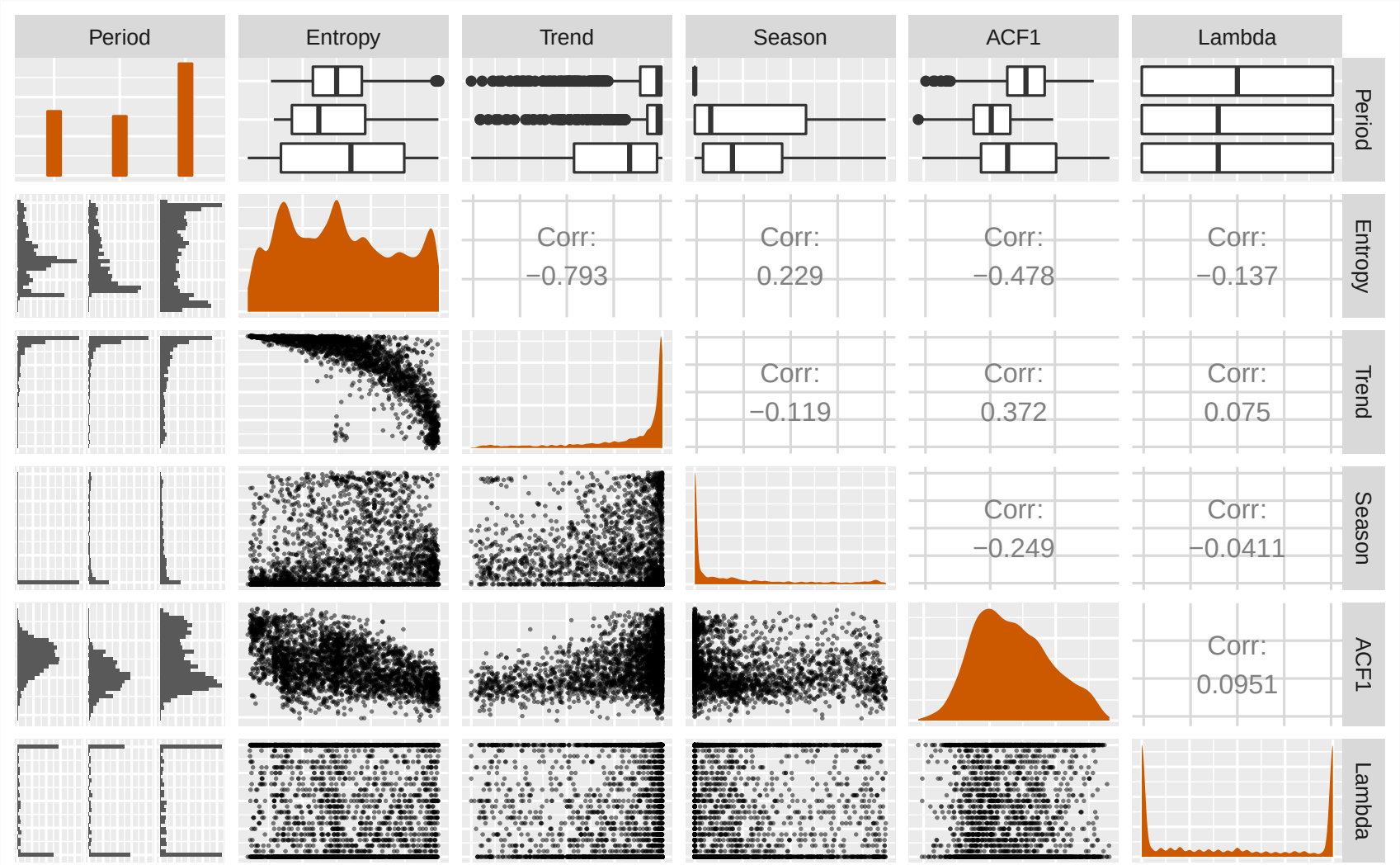
Feature distributions



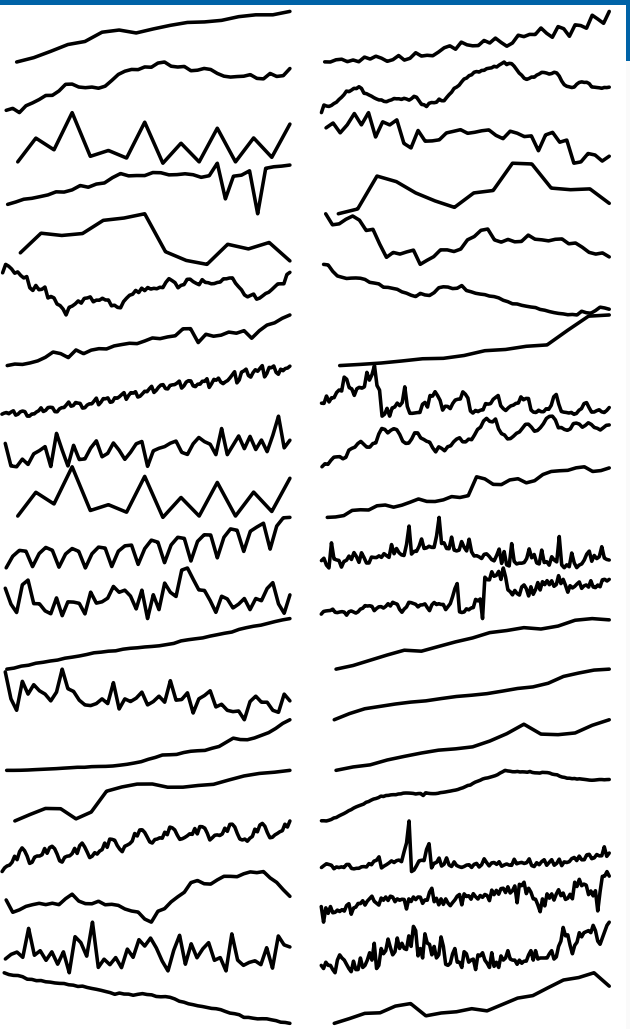
Feature distributions



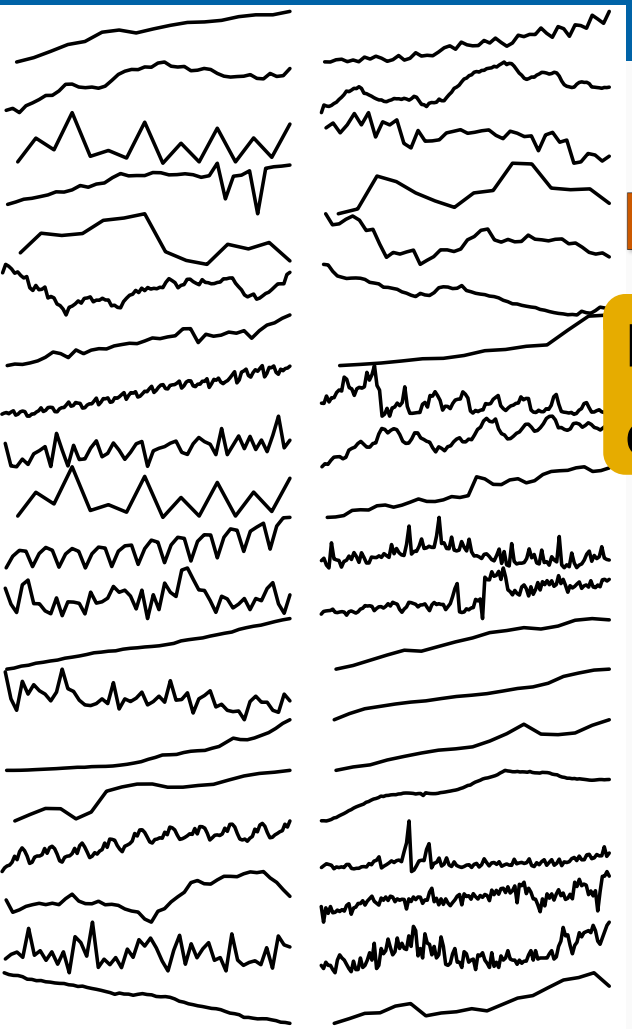
Feature distributions



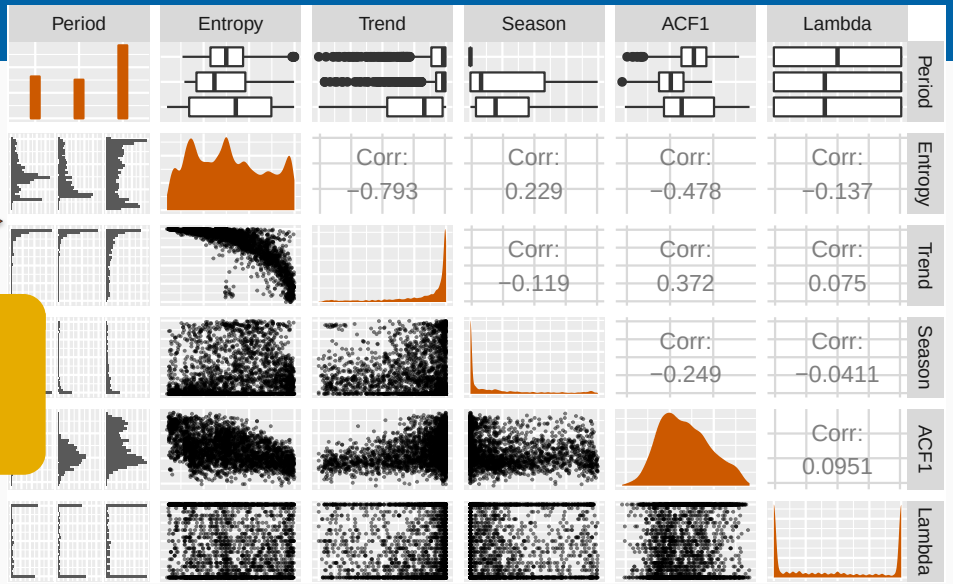
Dimension reduction for time series



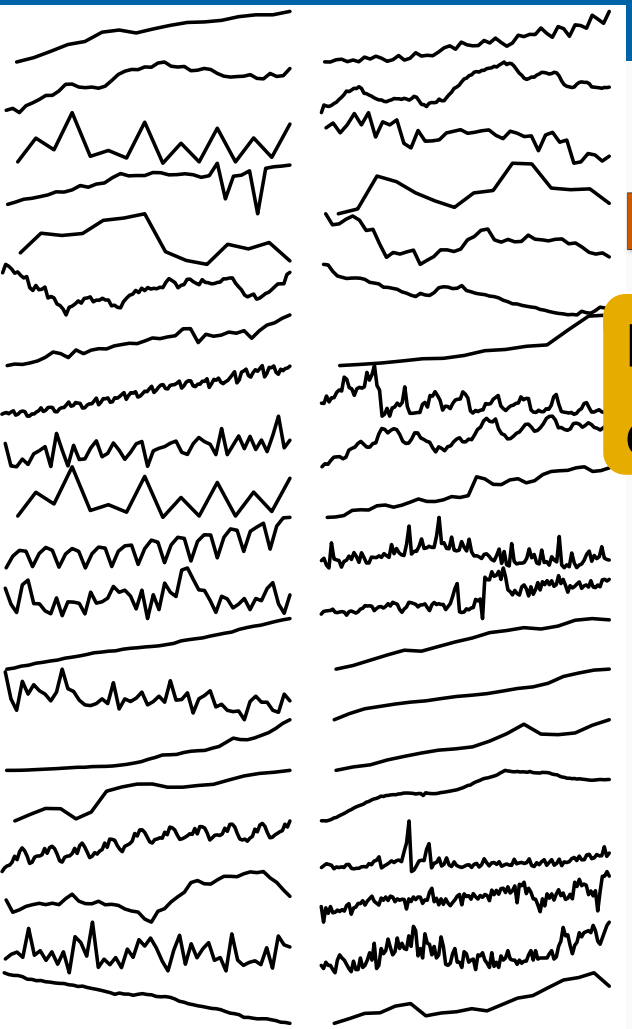
Dimension reduction for time series



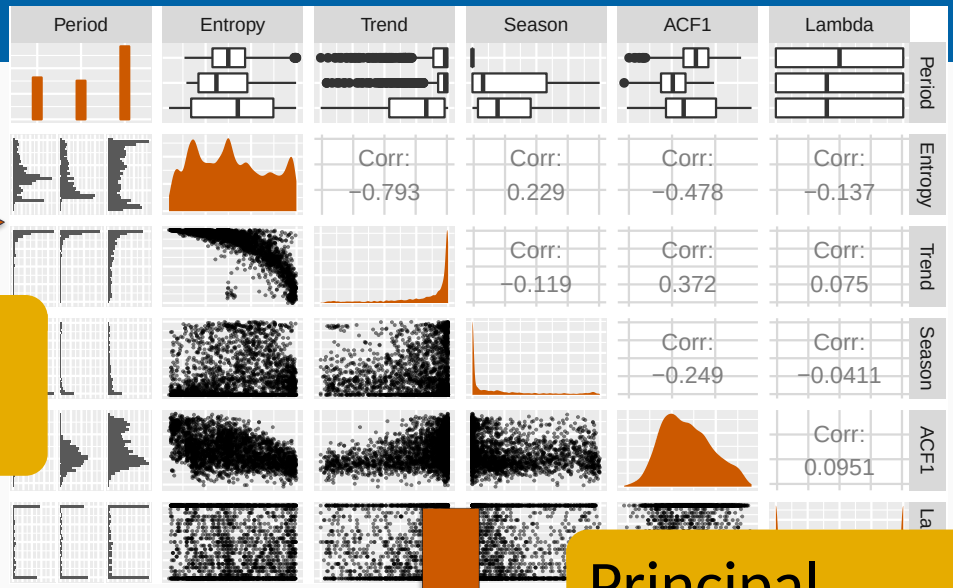
Feature calculation



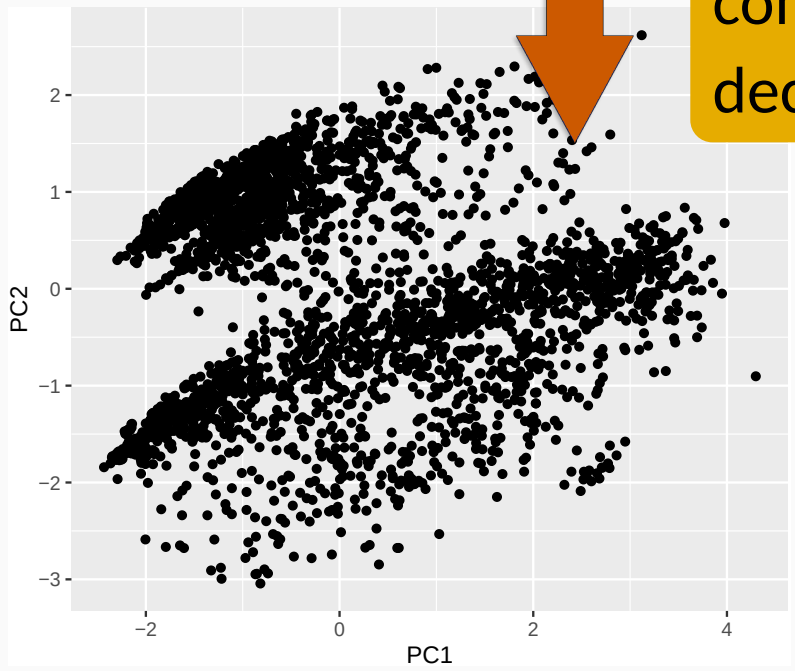
Dimension reduction for time series



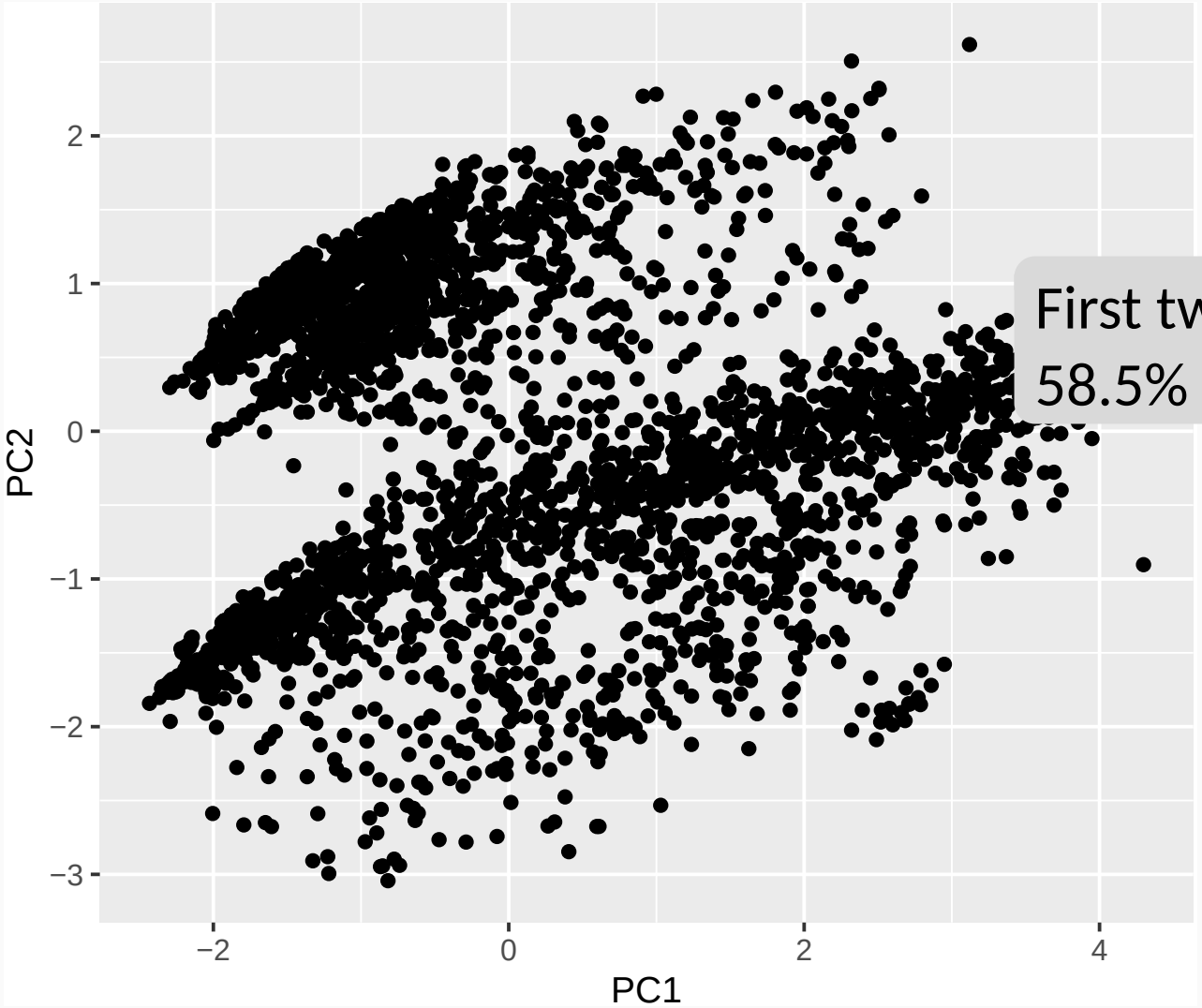
Feature calculation



Principal component decomposition

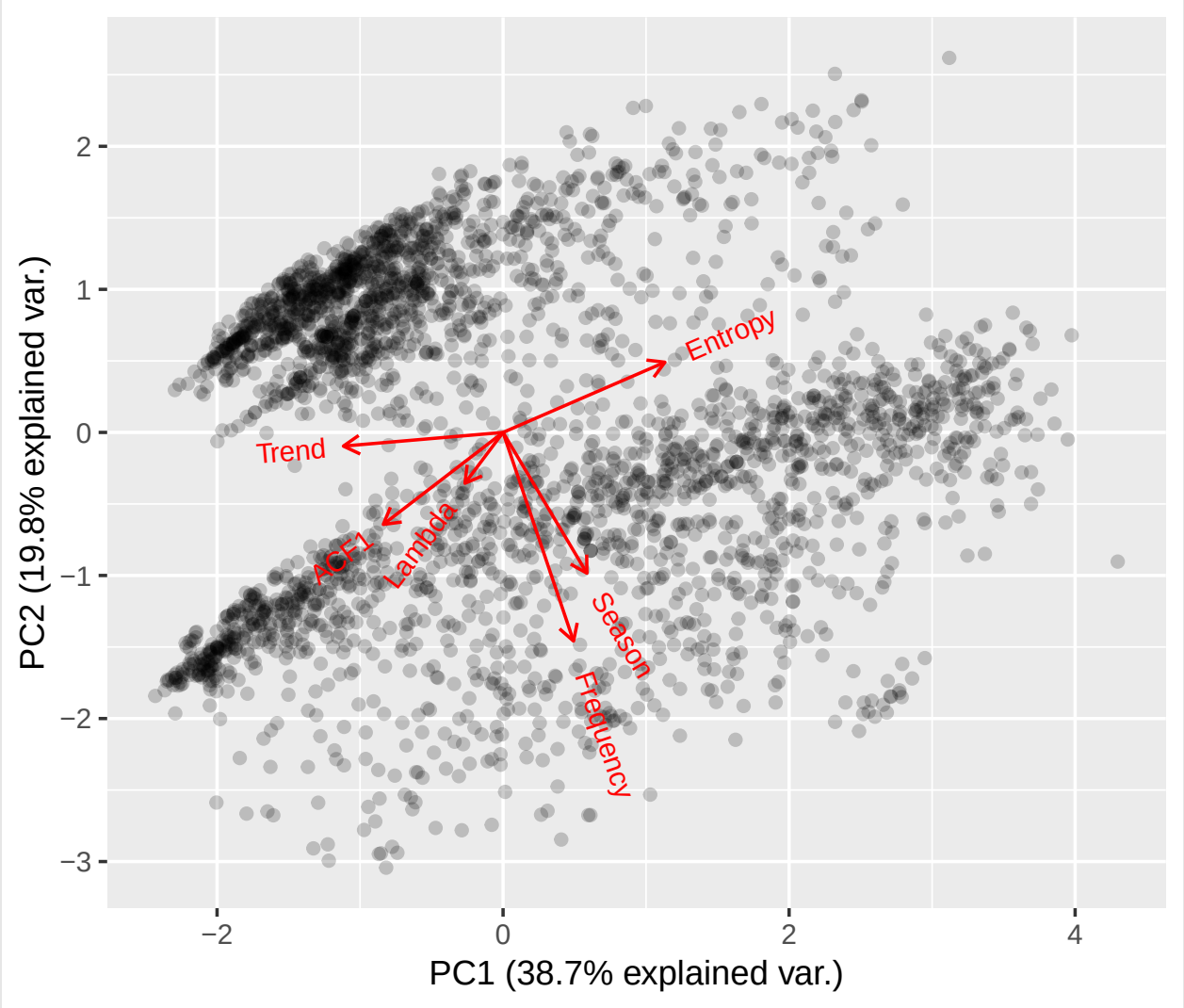


M3 feature space

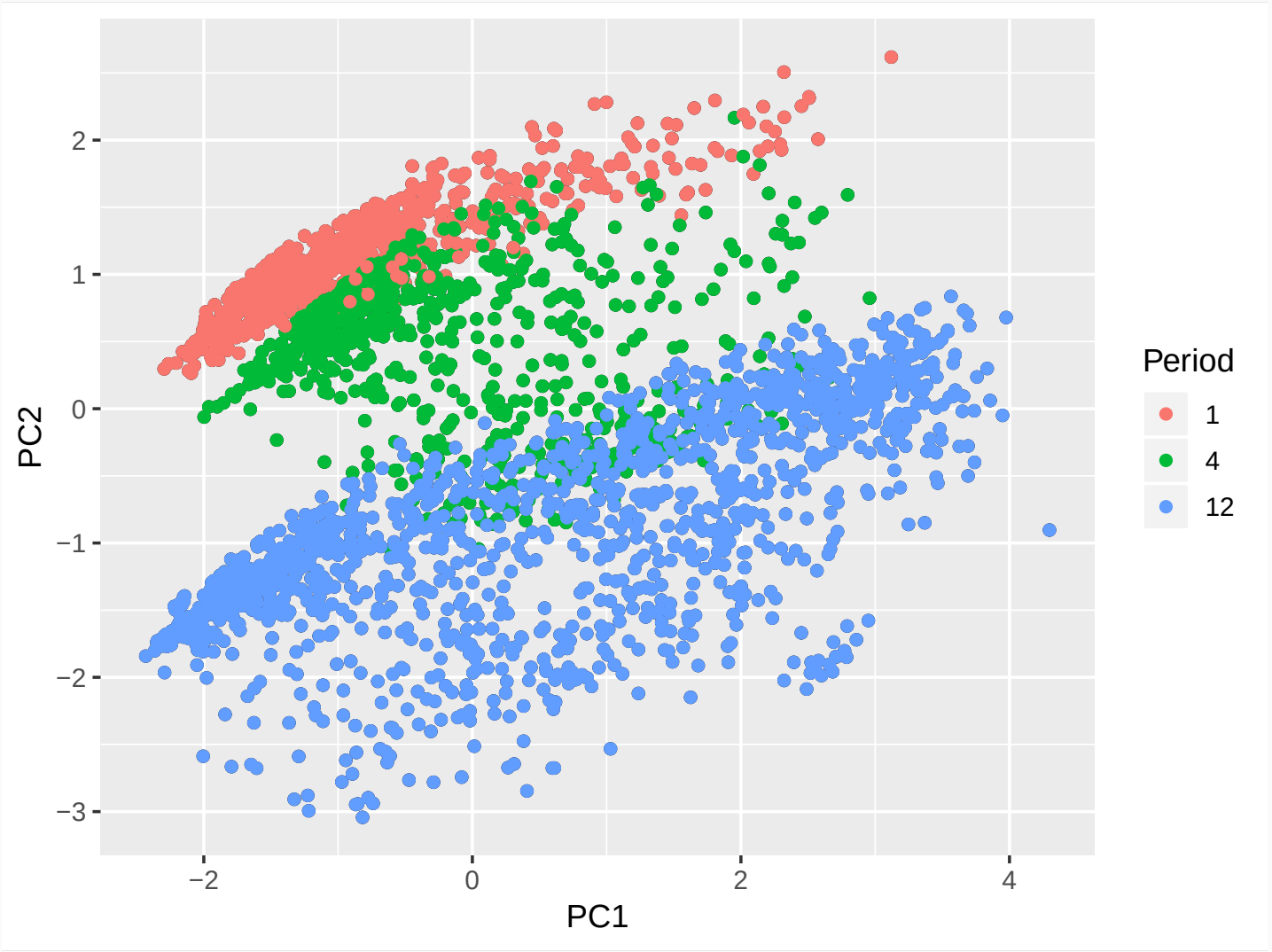


First two PCs explain 58.5% of the variance.

M3 feature space



M3 feature space



Feature properties

In this analysis, we have restricted features to be

- ergodic
- scale-independent

For other analyses, it may be appropriate to have different requirements.

R package: tsfeatures

github.com/robjhyndman/tsfeatures

```
library(tsfeatures)
library(tidyverse)
library(forecast)

myfeatures <- function(x,...) {
  lambda <- BoxCox.lambda(x, lower=0, upper=1, method='loglik')
  y <- BoxCox(x, lambda)
  c(stl_features(y,s.window='periodic', robust=TRUE, ...),
    lambda=lambda)
}
M3Features <- bind_cols(
  tsfeatures(M3data, c("frequency", "entropy")),
  tsfeatures(M3data, "myfeatures", scale=FALSE))
```

Outline

Forecast model selection

Features used to select a forecasting model

- 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

Outline

FFORMS: Feature-based FORecast Model Selection

Features used to select a forecasting model

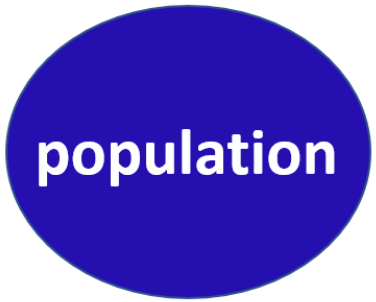
- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- first ACF value of remainder series
- parameter estimates of Holt's linear trend method
- spectral entropy
- Hurst exponent
- nonlinearity
- parameter estimates of Holt-Winters' additive 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

FFORMS: Feature-based FORecast Model Selection

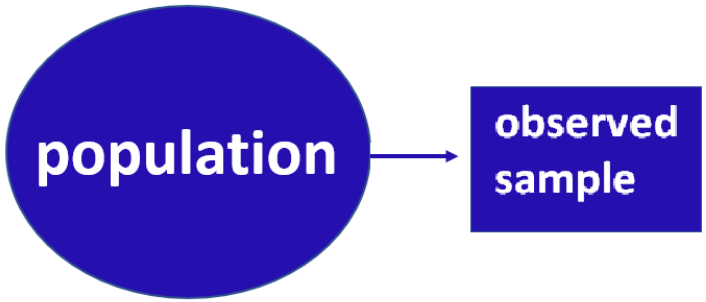
Why these features?

- Hyndman, Wang and Laptev. “Large scale unusual time series detection” (ICDM 2015).
- Kang, Hyndman & Smith-Miles. “Visualising forecasting algorithm performance using time series instance spaces” (IJF 2017).
- Talagala, Hyndman and Athanasopoulos. “Meta-learning how to forecast time series” (2018).
- Implemented in the tsfeatures R package

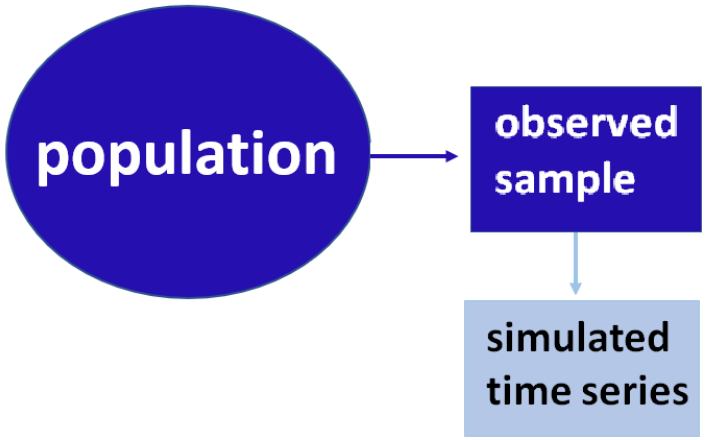
FFORMS: Feature-based FORecast Model Selection



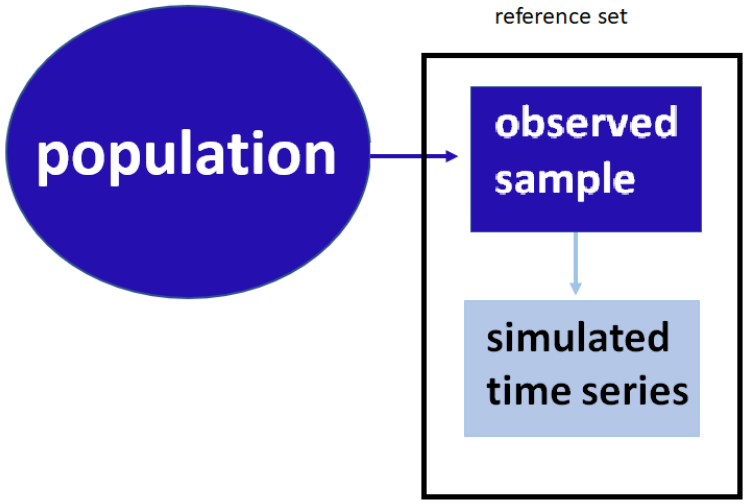
FFORMS: Feature-based FORecast Model Selection



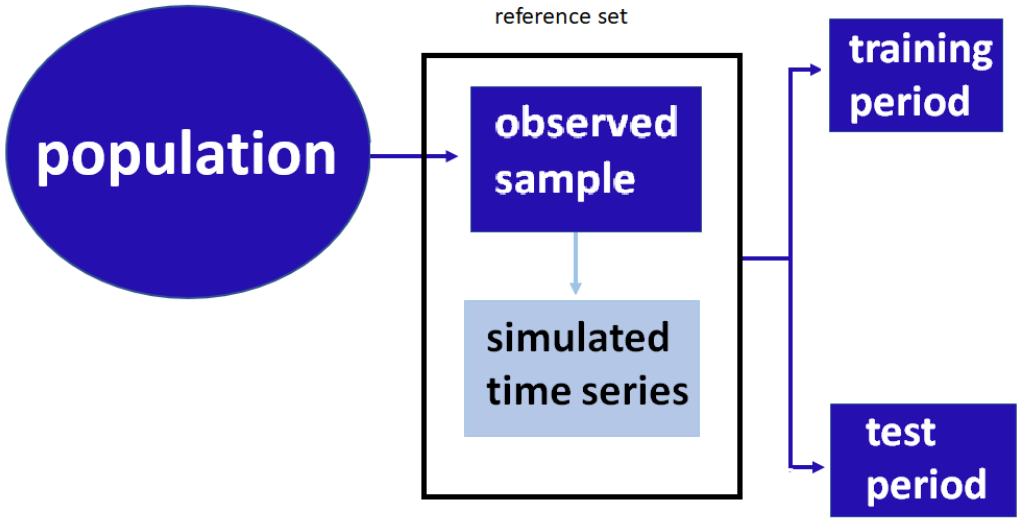
FFORMS: Feature-based FORecast Model Selection



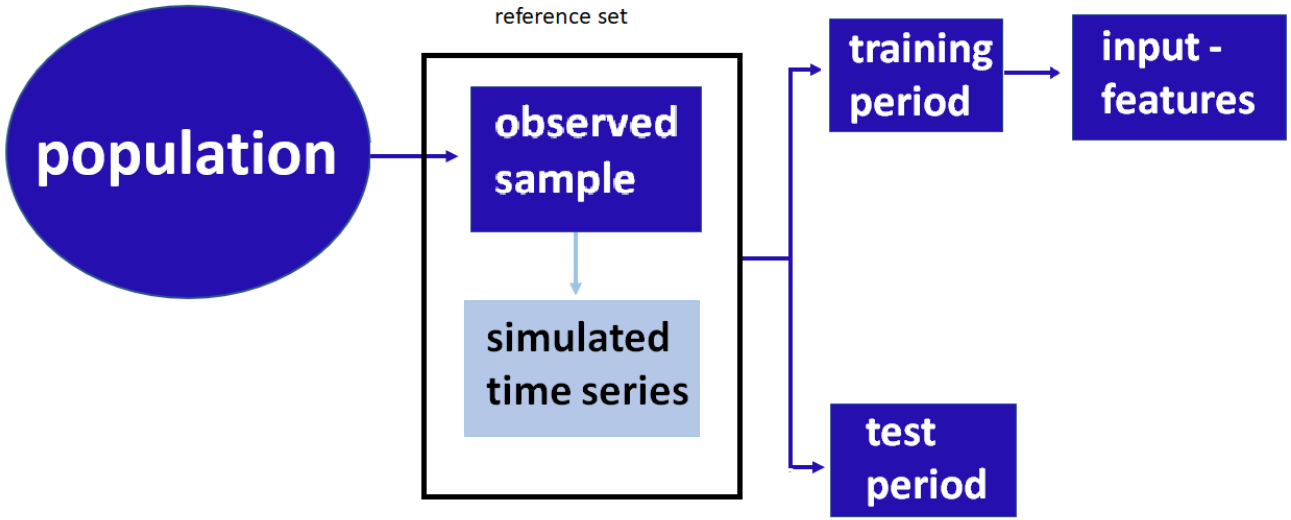
FFORMS: Feature-based FORecast Model Selection



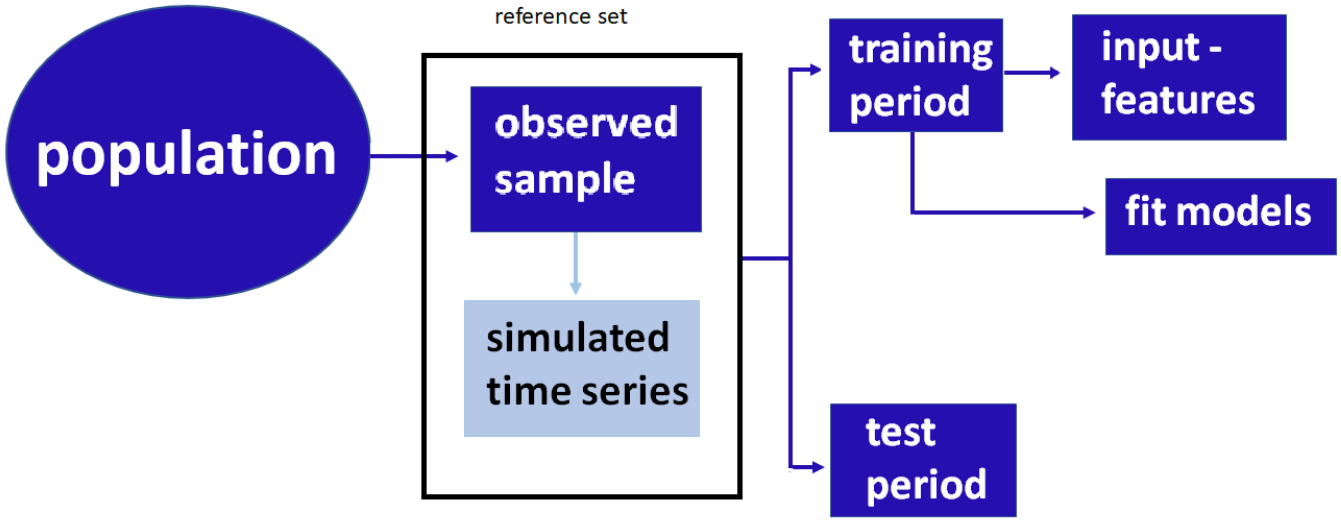
FFORMS: Feature-based FORecast Model Selection



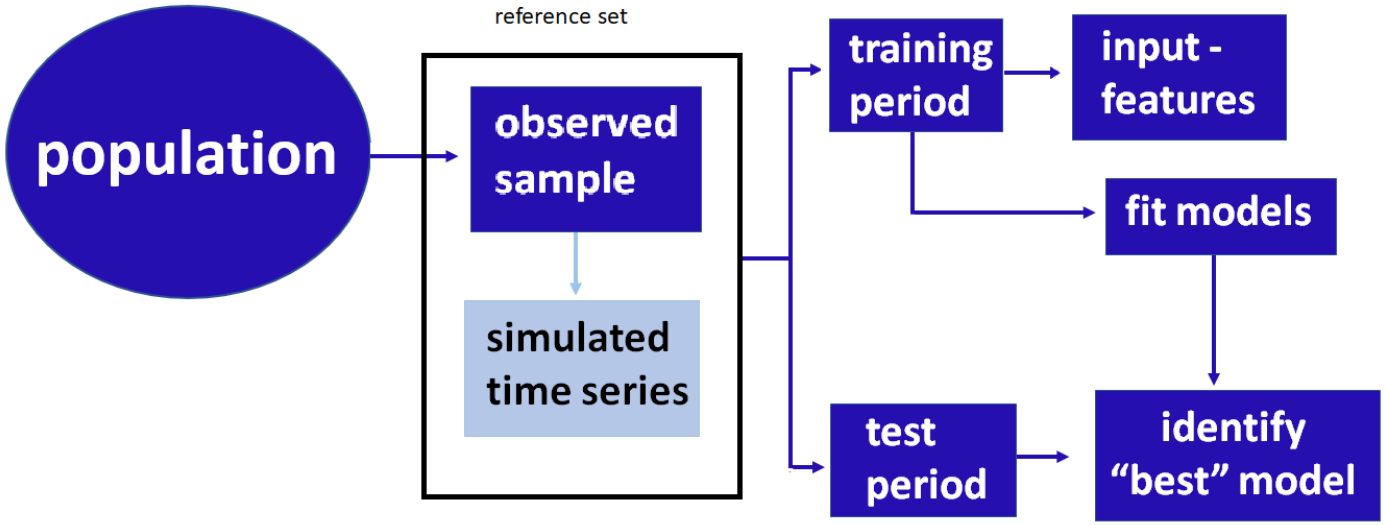
FFORMS: Feature-based FORecast Model Selection



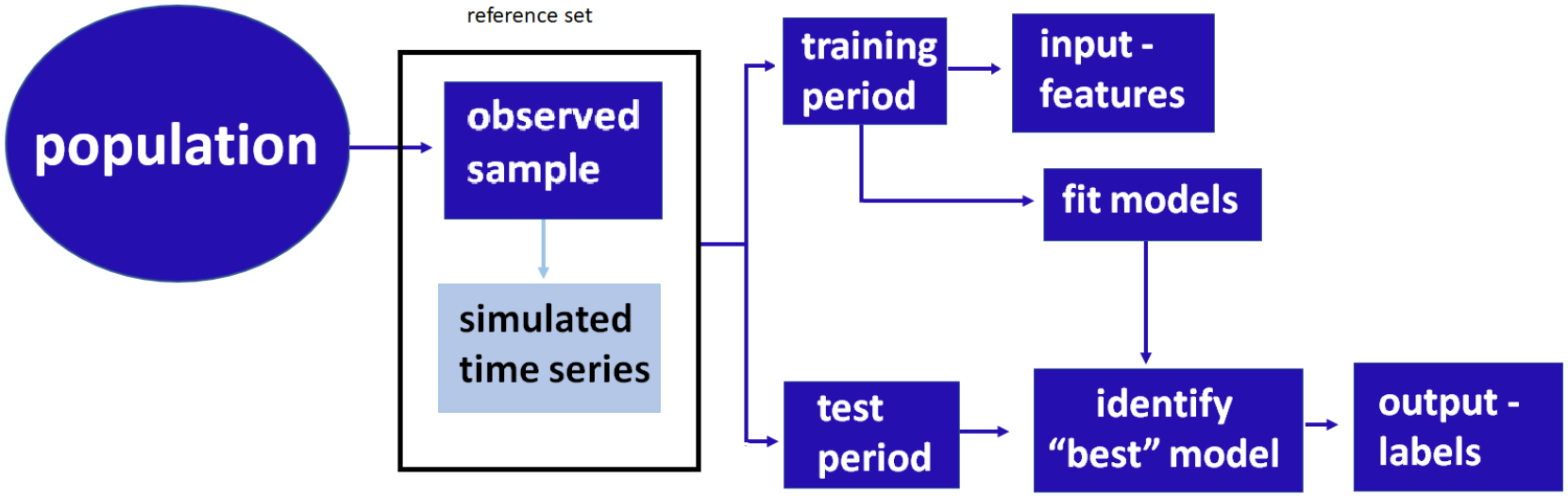
FFORMS: Feature-based FORecast Model Selection



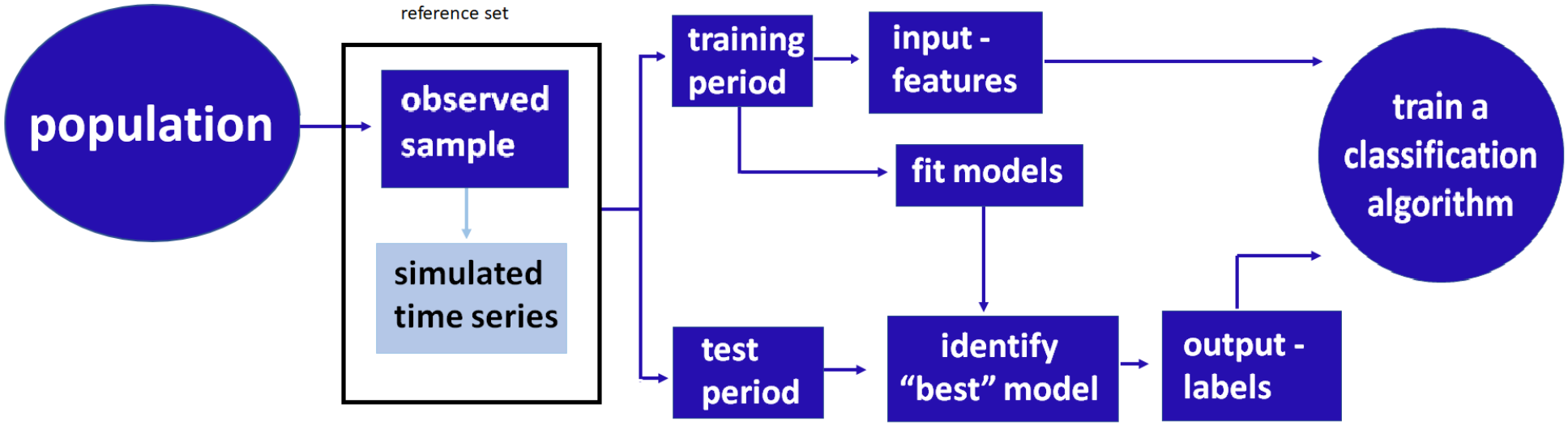
FFORMS: Feature-based FORecast Model Selection



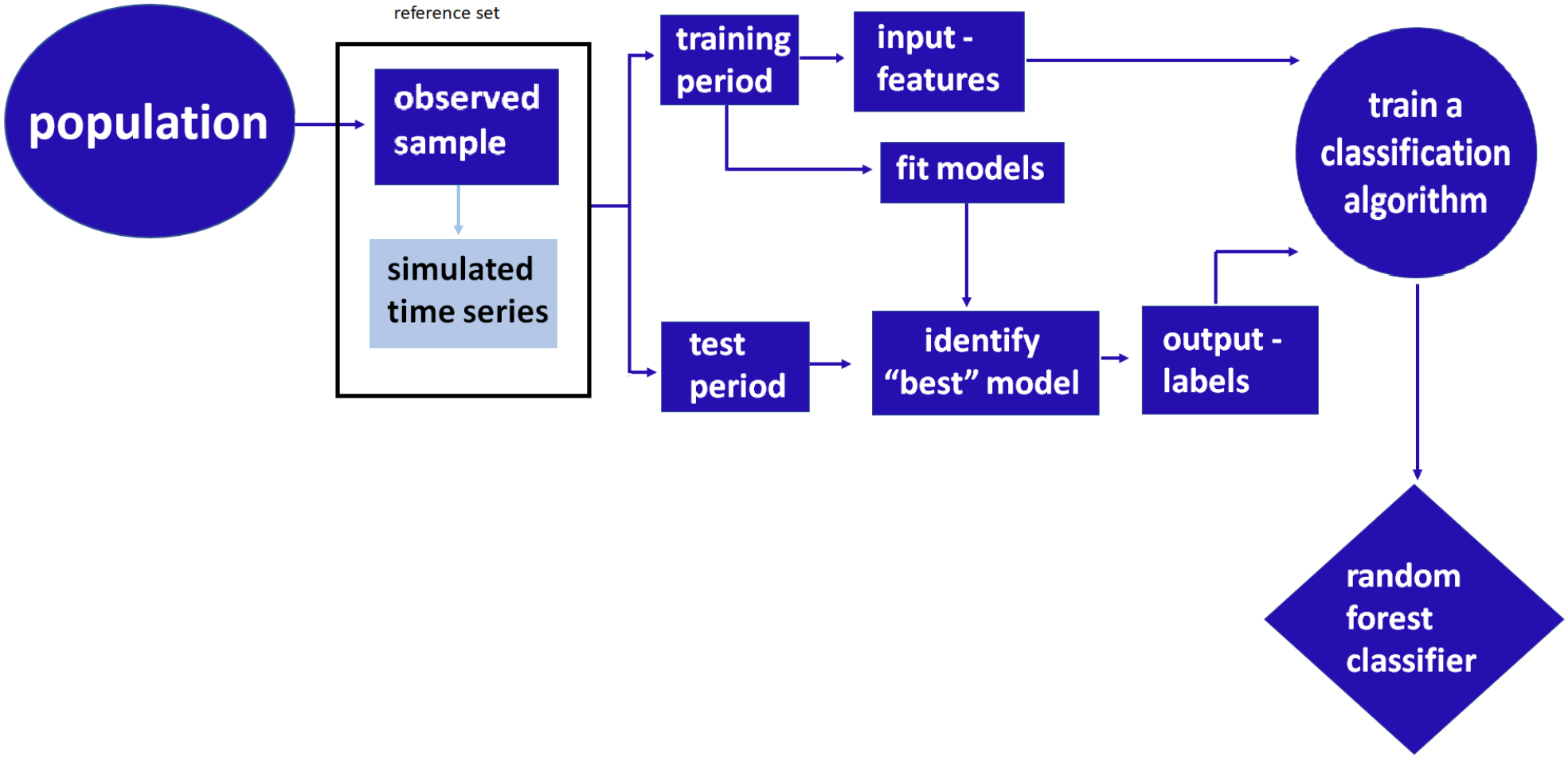
FFORMS: Feature-based FORecast Model Selection



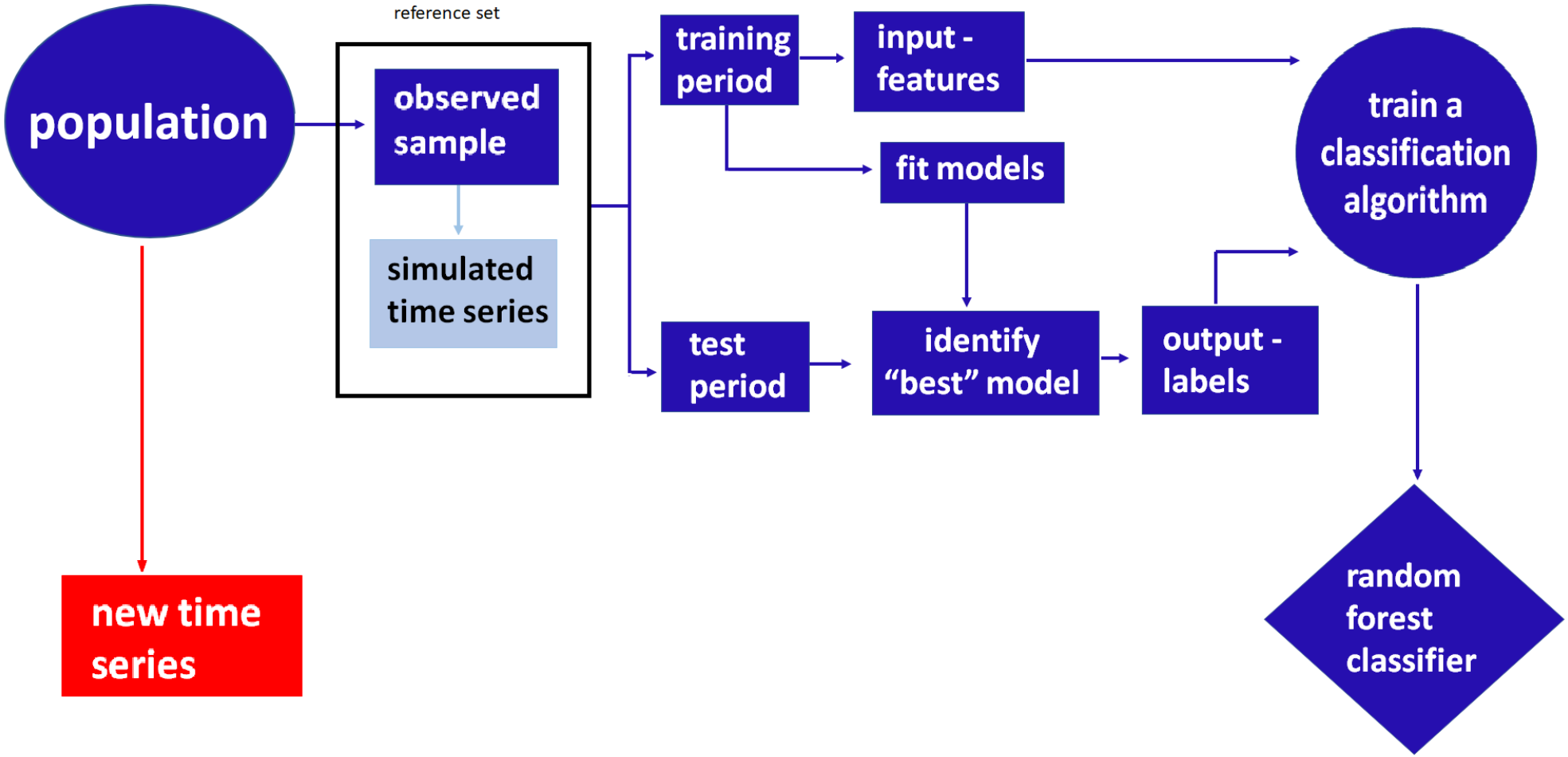
FFORMS: Feature-based FORecast Model Selection



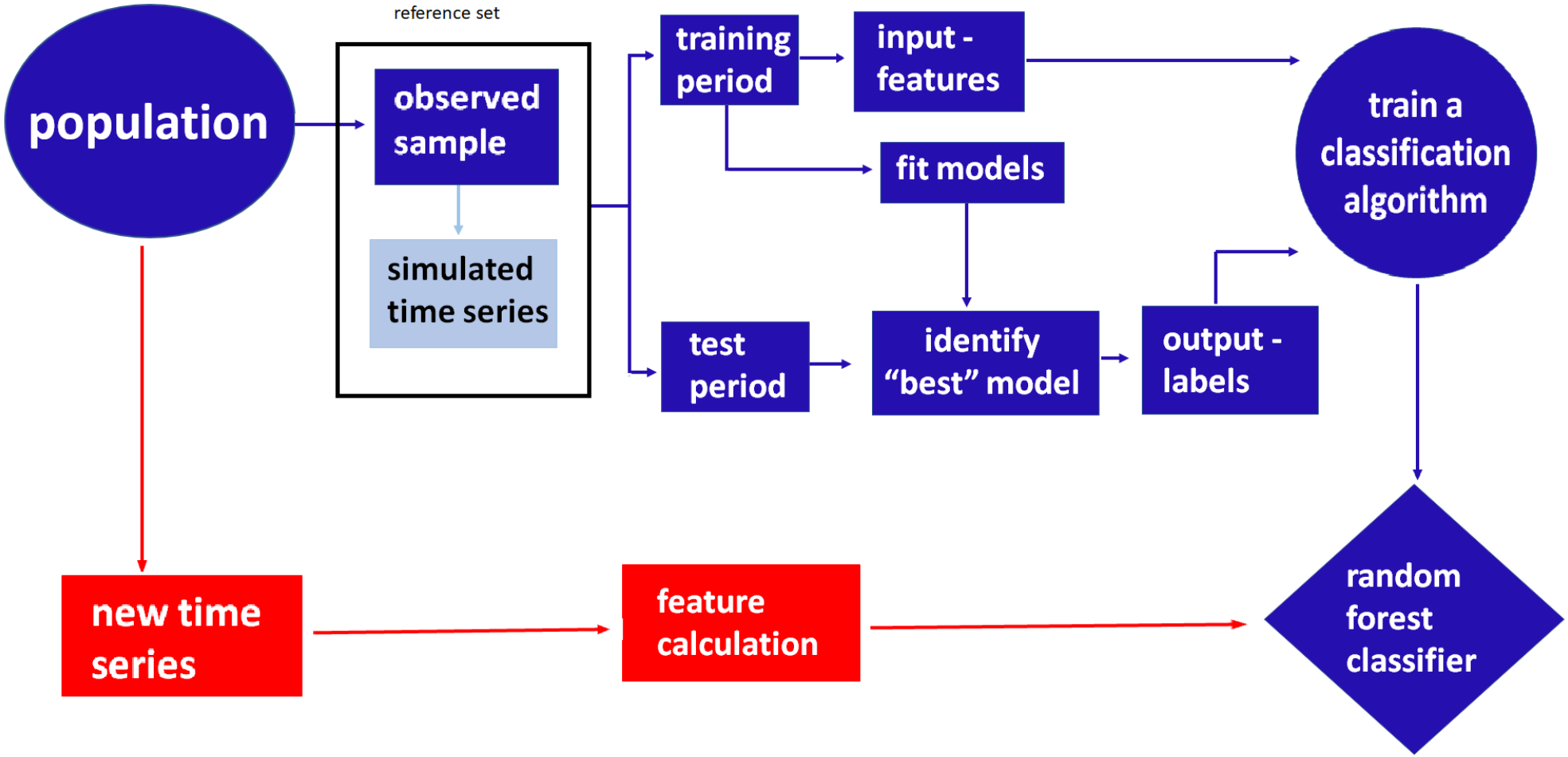
FFORMS: Feature-based FORecast Model Selection



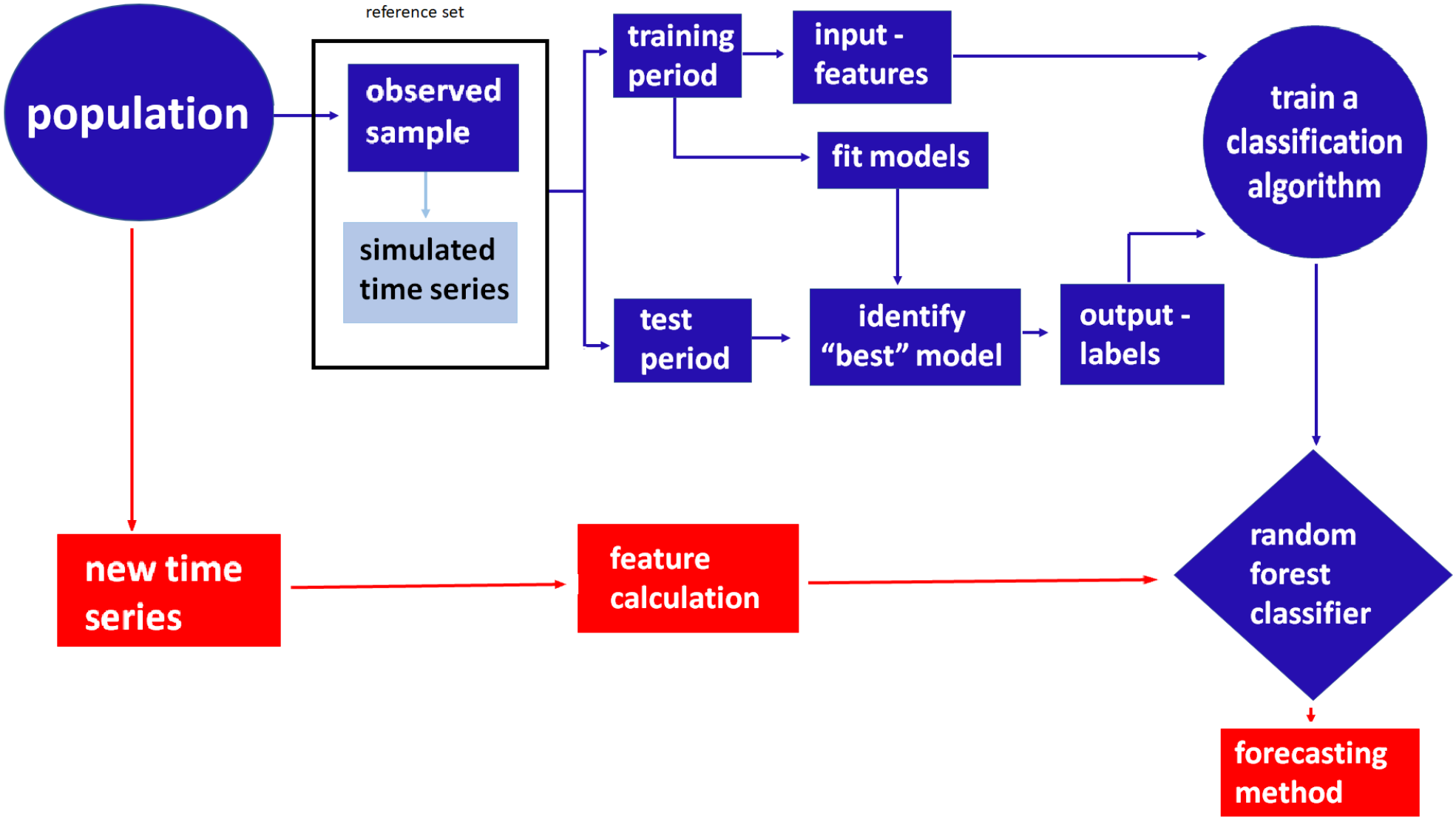
FFORMS: Feature-based FORecast Model Selection



FFORMS: Feature-based FORecast Model Selection



FFORMS: Feature-based FORecast Model Selection



Application to M competition data

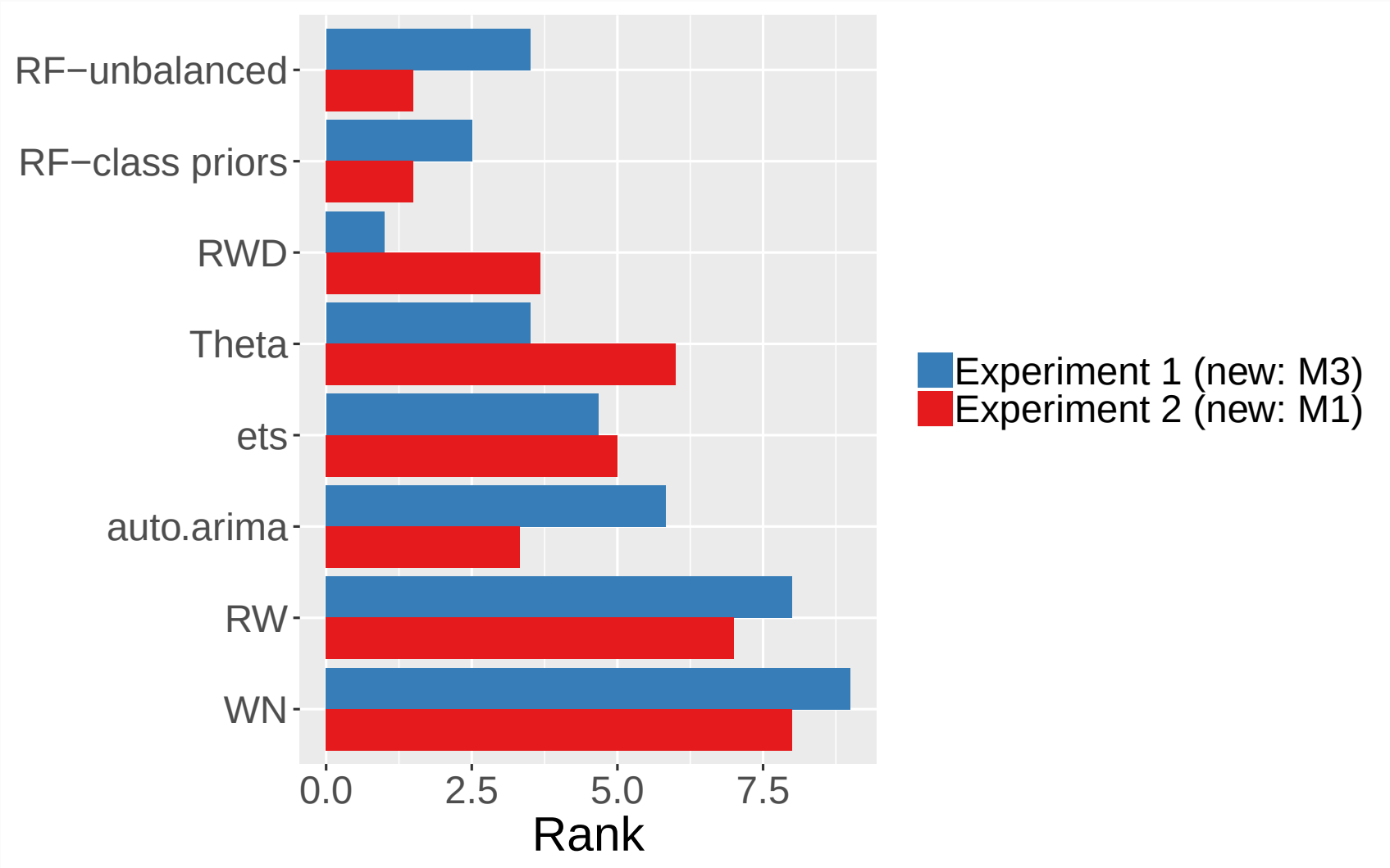
Experiment 1

	Source	Y	Q	M
Observed series	M1	181	203	617
Simulated series		362000	406000	123400
New series	M3	645	756	1428

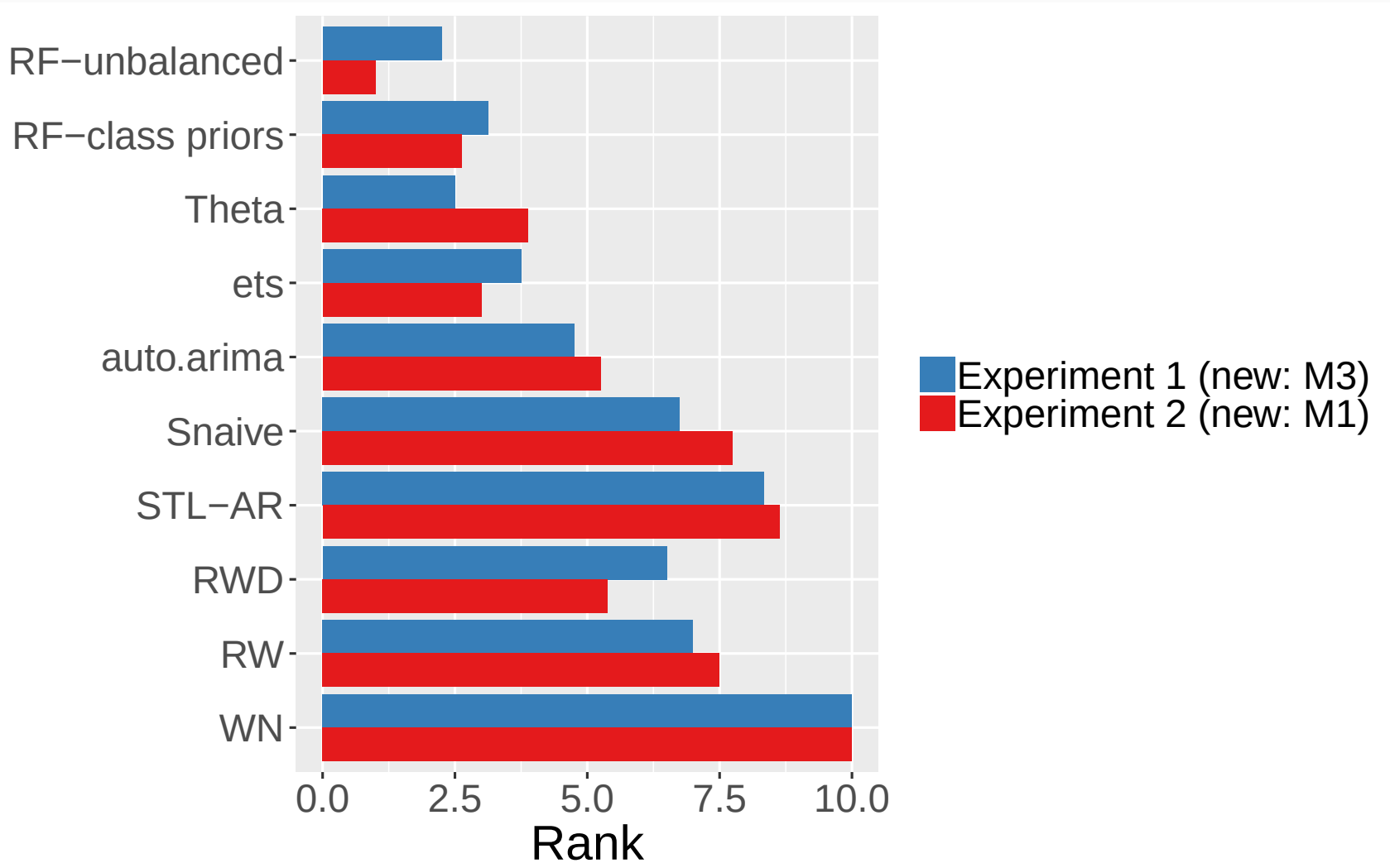
Experiment 2

	Source	Y	Q	M
Observed series	M3	645	756	1428
Simulated series		1290000	1512000	285600
New series	M1	181	203	617

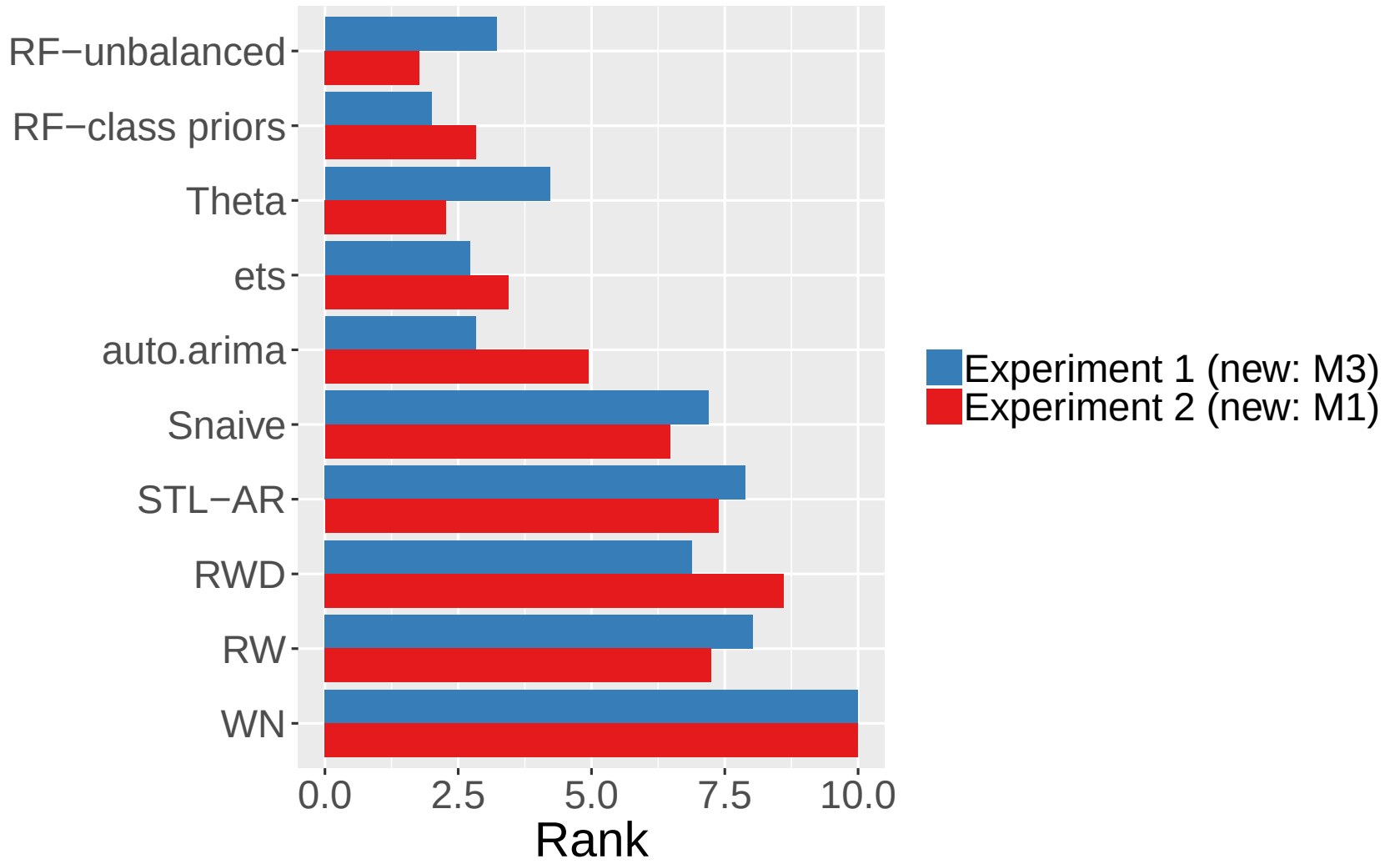
Results: Yearly



Results: Quarterly



Results: Monthly



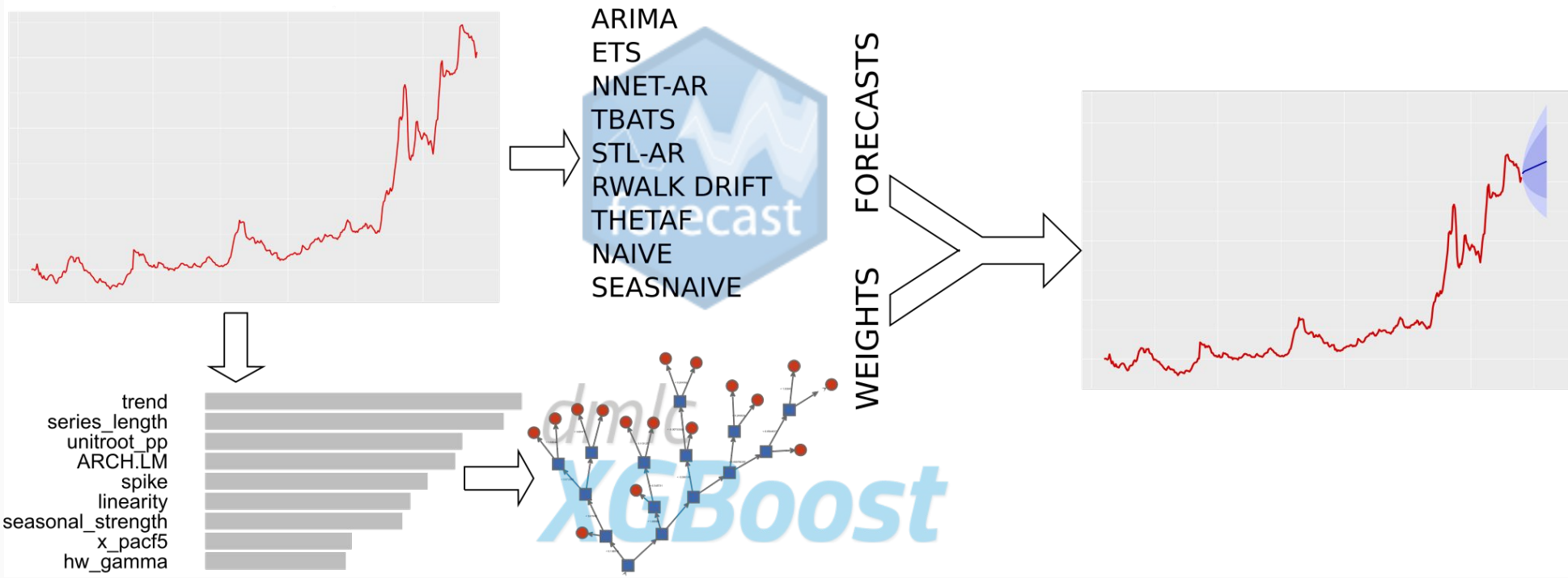
Outline

1. Makridakis competitions
2. Time series features
3. Feature-based forecasting
4. FFORMS: Feature-based forecast model selection
5. FFORMA: Feature-based forecast model averaging

FFORMA: Feature-based FORecast Model Averaging

- Like FFORMS but we use gradient boosted trees rather than a random forest.
- The optimization criterion is forecast accuracy not classification accuracy.
- The probability of each model being best is used to construct a model weight.
- A combination forecast is produced using these weights.
- **Came second in the M4 competition**

FFORMA: Feature-based FORecast Model Averaging



Results

2nd according to average OWA: 0.838

1st: 0.821

3rd: 0.841

FFORMA: Feature-based FORecast Model Averaging

Models included

1. Naive
2. Seasonal naive
3. Random walk with drift
4. Theta method
5. ARIMA
6. ETS
7. TBATS
8. STL decomposition with AR for seasonally adjusted series
9. Neural network autoregression

FFORMA: Feature-based FORecast Model Averaging

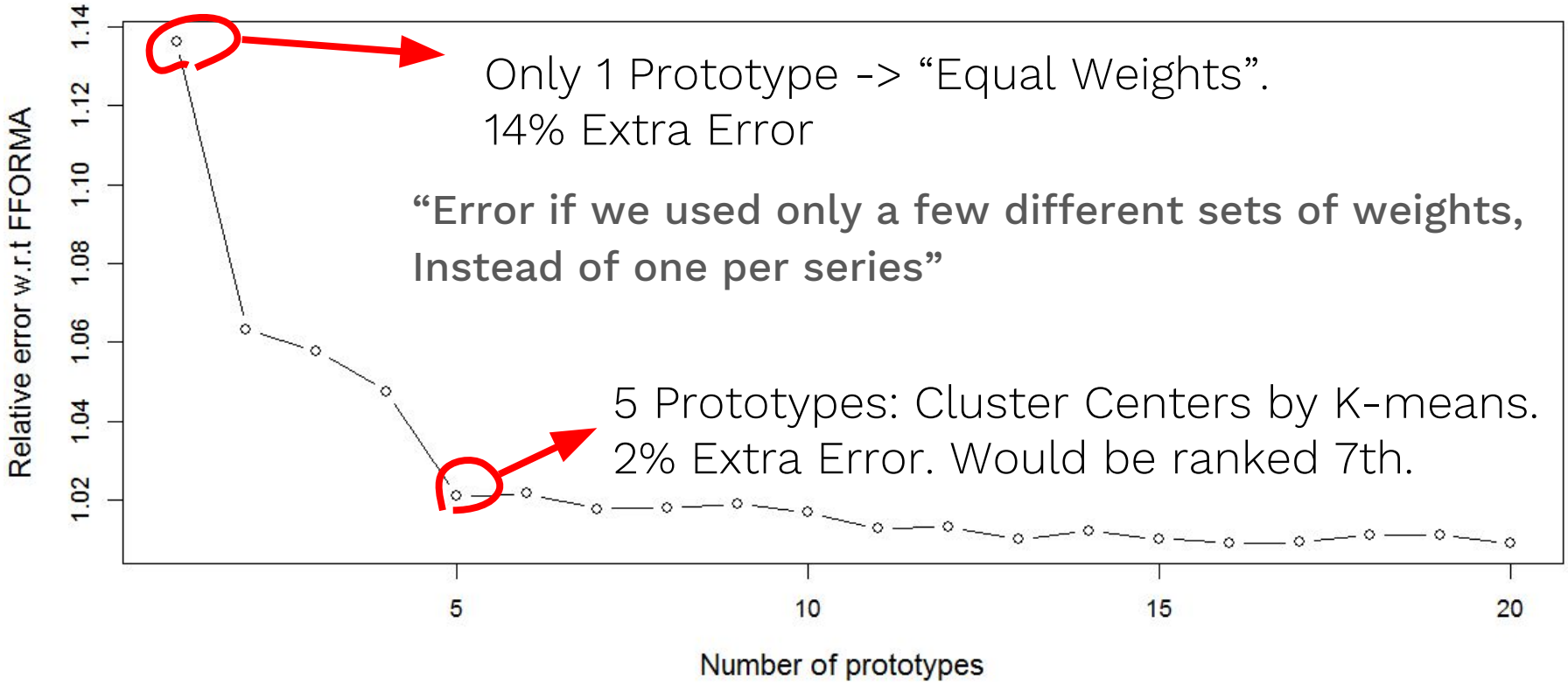
Weights for the combination:

- A **decision tree model** generates the weights using xgboost.
- The model is trained on a temporal holdout version of the M4 dataset, where the test sets are equal to the required forecast horizons.

FFORMA: Feature-based FORecast Model Averaging

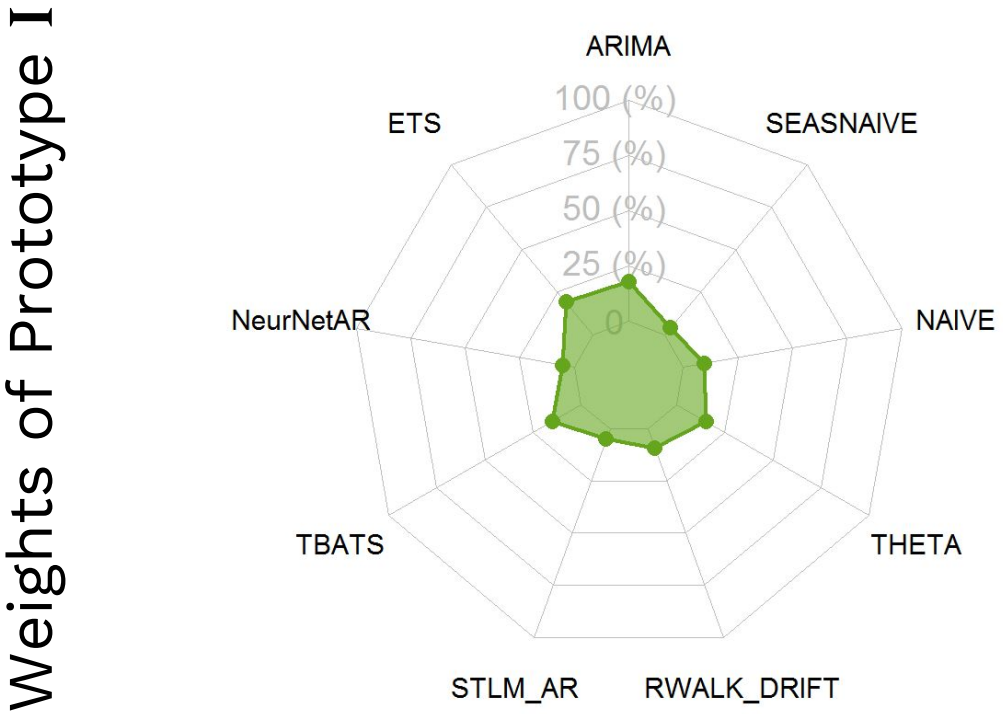
FFORMA: Feature-based FORecast Model Averaging

Looking for **Prototypes** in the weights



FFORMA: Feature-based FORecast Model Averaging

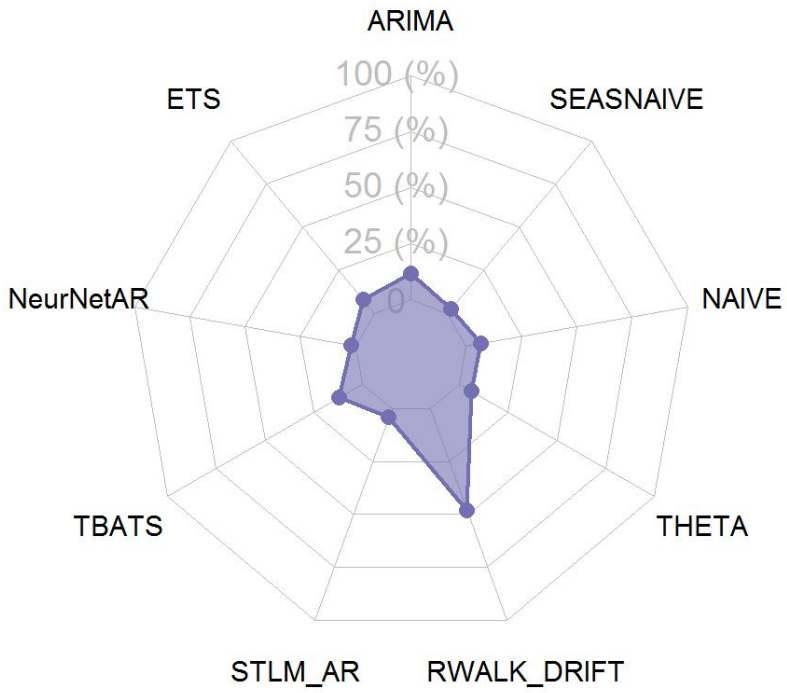
“Roughly Equal Weights”. 40000 Series in M4



FFORMA: Feature-based FORecast Model Averaging

“Mostly RandomWalk Drift”. 20000 Series in M4

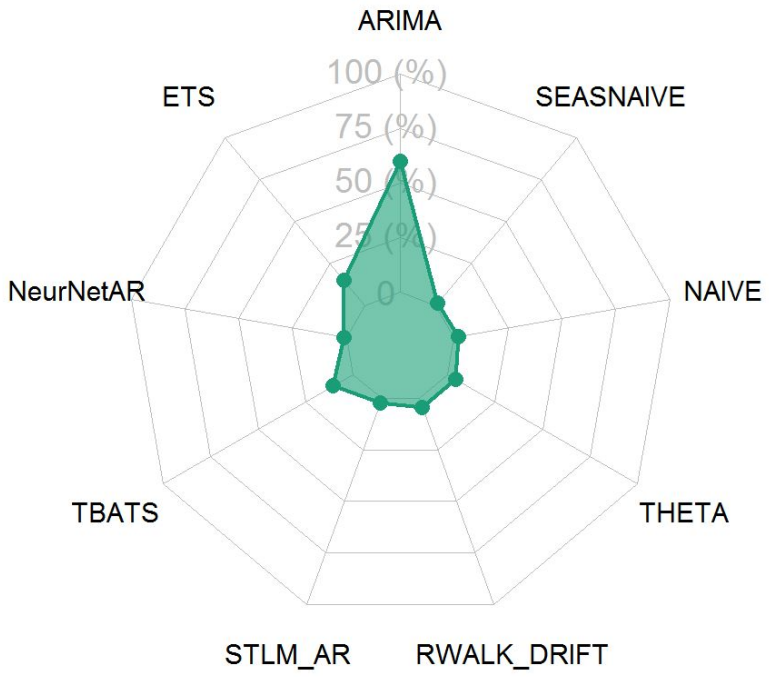
Weights of Prototype II



FFORMA: Feature-based FORecast Model Averaging

“Mostly ARIMA”. 16000 Series in M4

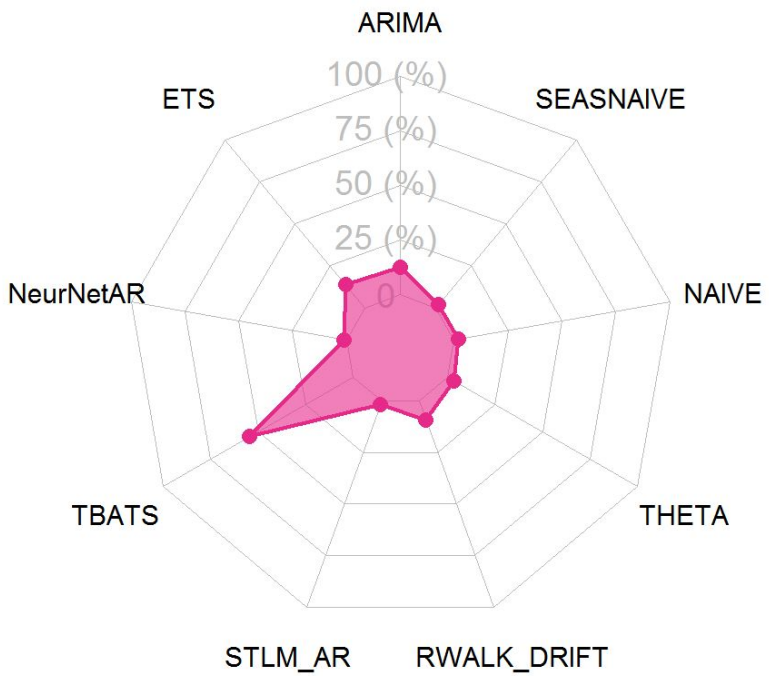
Weights of Prototype III



FFORMA: Feature-based FORecast Model Averaging

“Mostly TBATS”. 13000 Series in M4

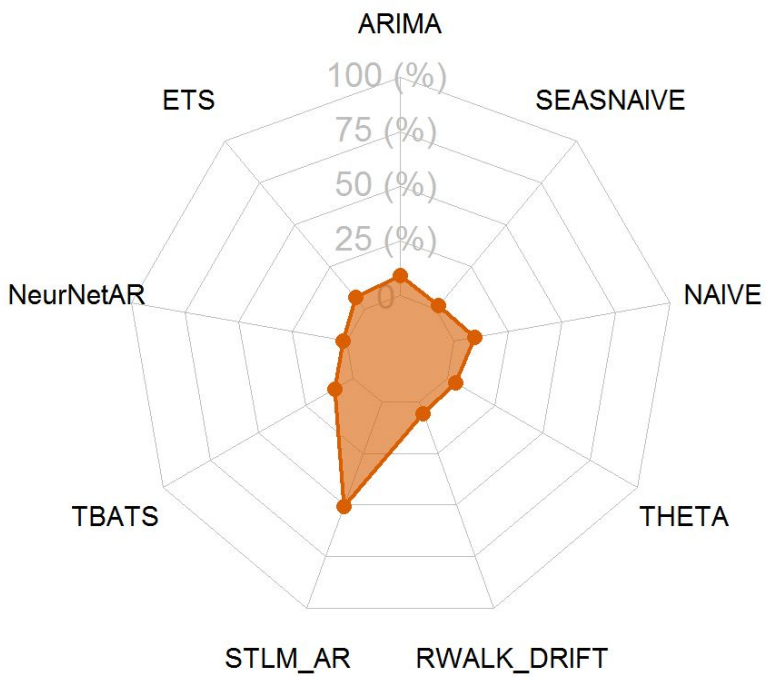
Weights of Prototype IV



FFORMA: Feature-based FORecast Model Averaging

“Mostly STLM-AR”. 8000 Series in M4

Weights of Prototype V



R Packages

- **seer**: FFORMS — selecting forecasting model using features.

  `github.com/thiyangt/seer`

- **M4metalearning**: FFORMA – forecast combinations using features to choose weights.

  `github.com/robjhyndman/M4metalearning`

Acknowledgements

Kate Smith-Miles

Yanfei Kang

Earo Wang

Thiyanga Talagala

George Athanasopoulos

Pablo Montero-Manso