MONASH University

MONASH BUSINESS SCHOOL

Feature-based time series analysis

Rob J Hyndman

25 November 2018

Outline

1 Visualization

2 Forecasting

3 Anomaly detection

Outline

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

ELSEVIER

International Journal of Forecasting 16 (2000) 451–476

international journal of forecasting

www.elsevier.com/locate/ijforecast

etition: results, conclusions a … s

Spyros Makridakis, Michèle Hibon*

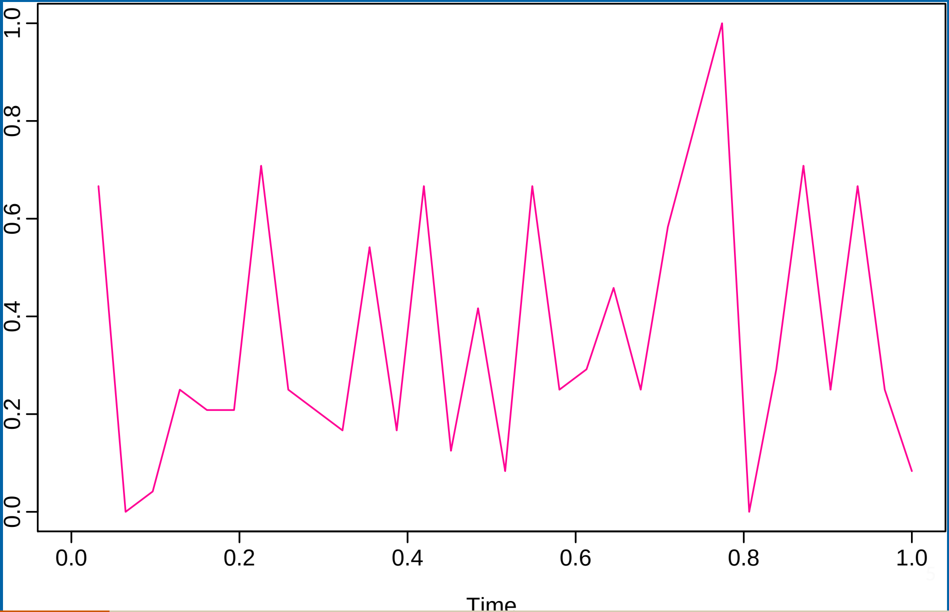
EAD, Boulevard de Constance, 77305 Fontainebleau, Fr

Abst

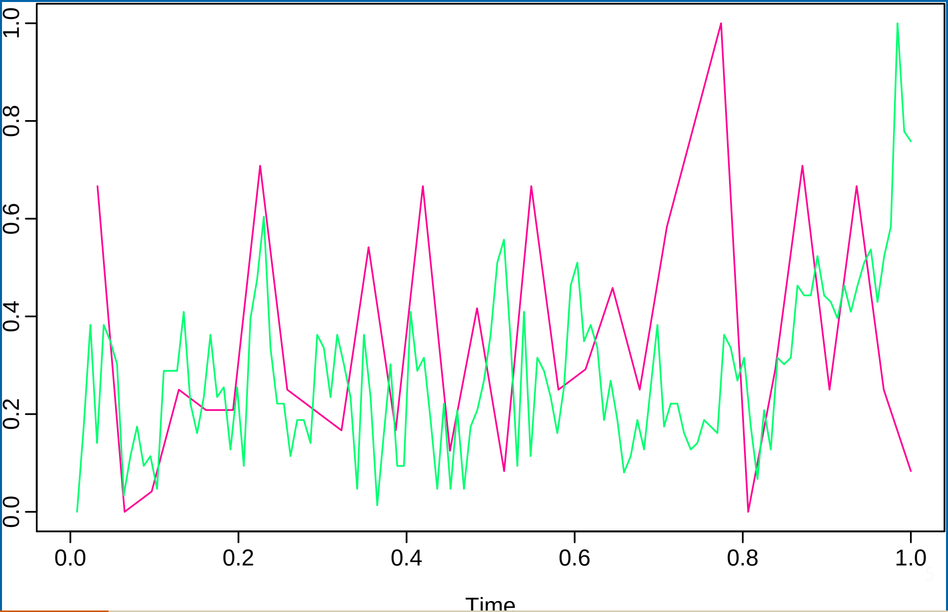
This paper describes the M3-Competition, the latest of the M-Competitions. It explains the reasons for conducting the competition and summarizes its results and conclusions. In addition, the paper compares such results/conclusions with those of the previous two M-Competitions as well as with those of 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

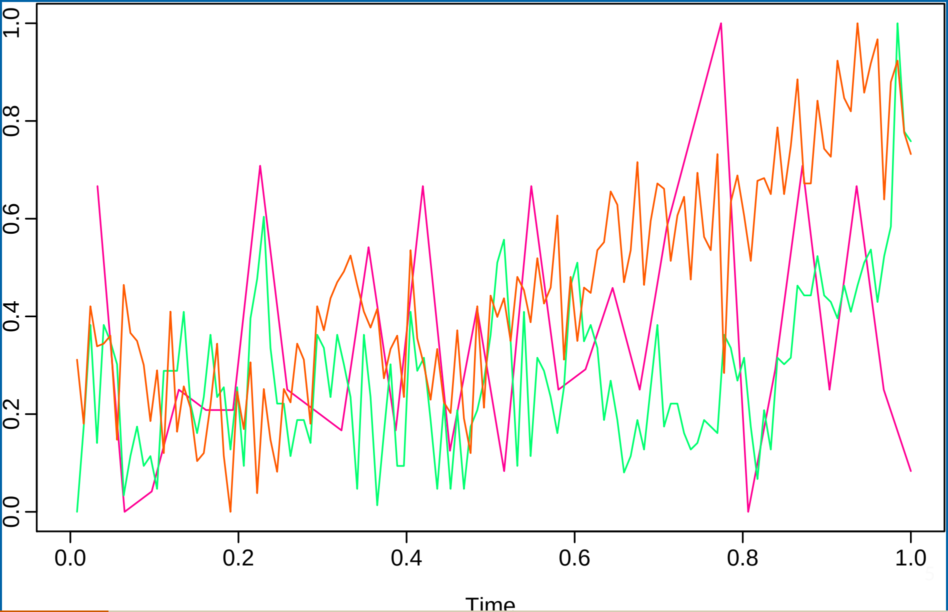
How to plot lots of time series?



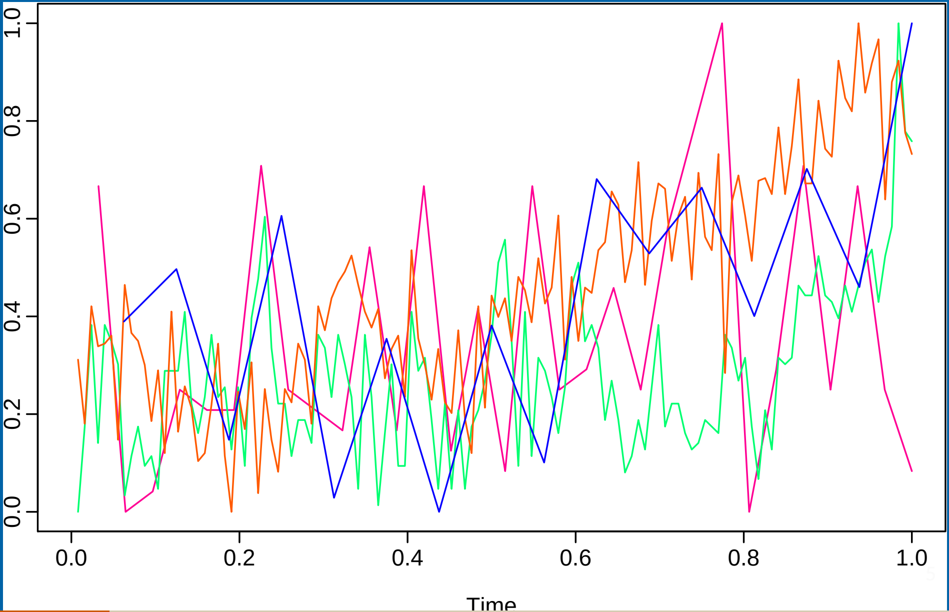
How to plot lots of time series?



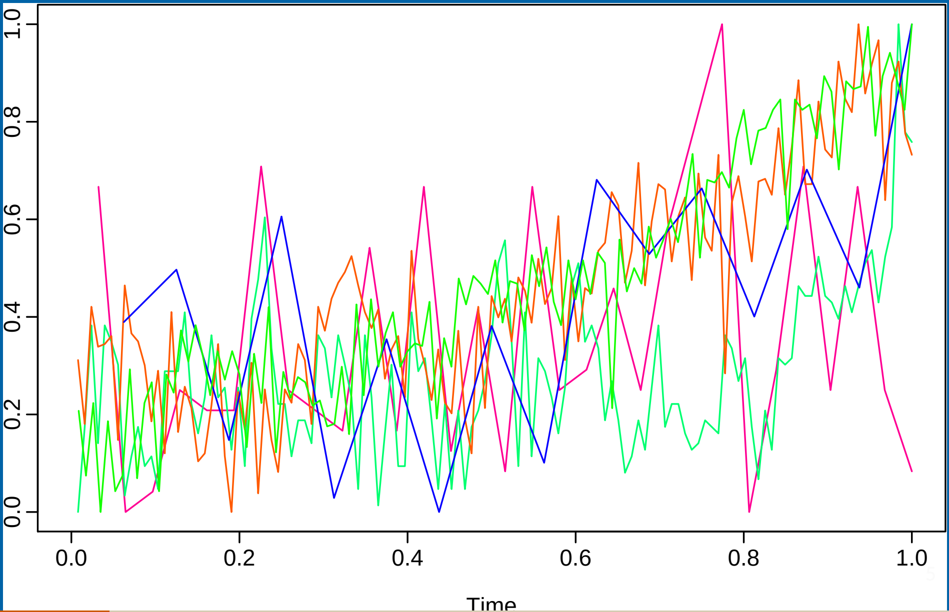
How to plot lots of time series?



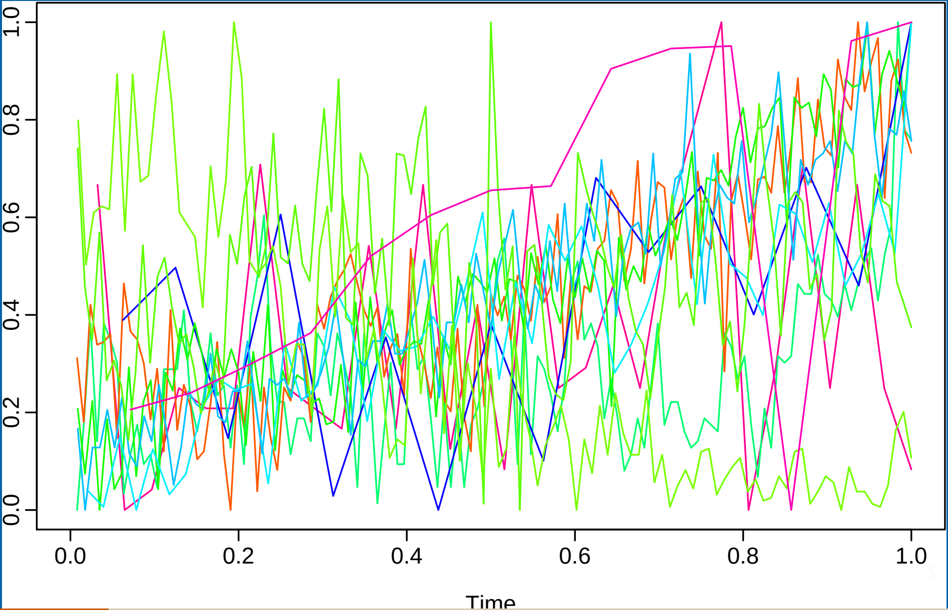
How to plot lots of time series?



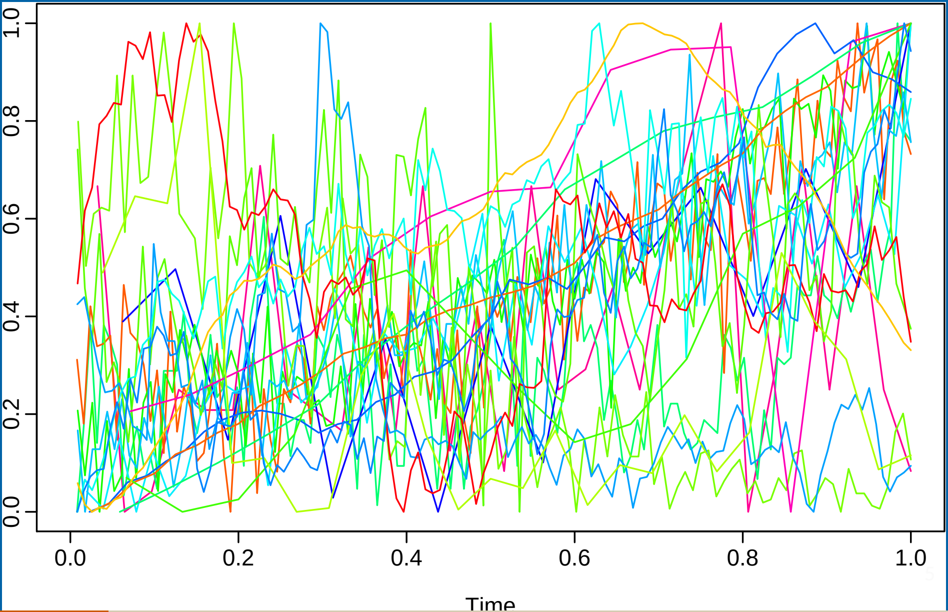
How to plot lots of time series?



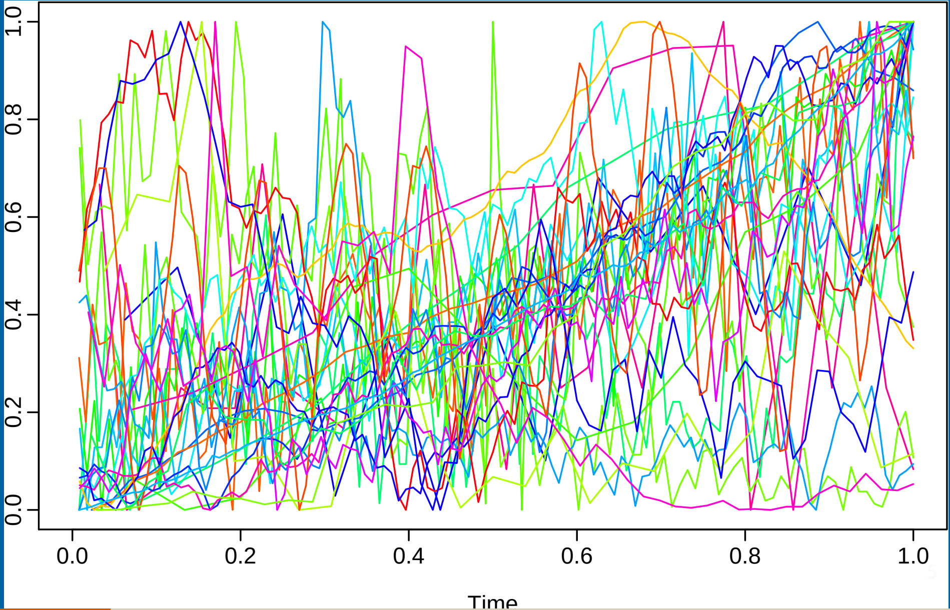
How to plot lots of time series?



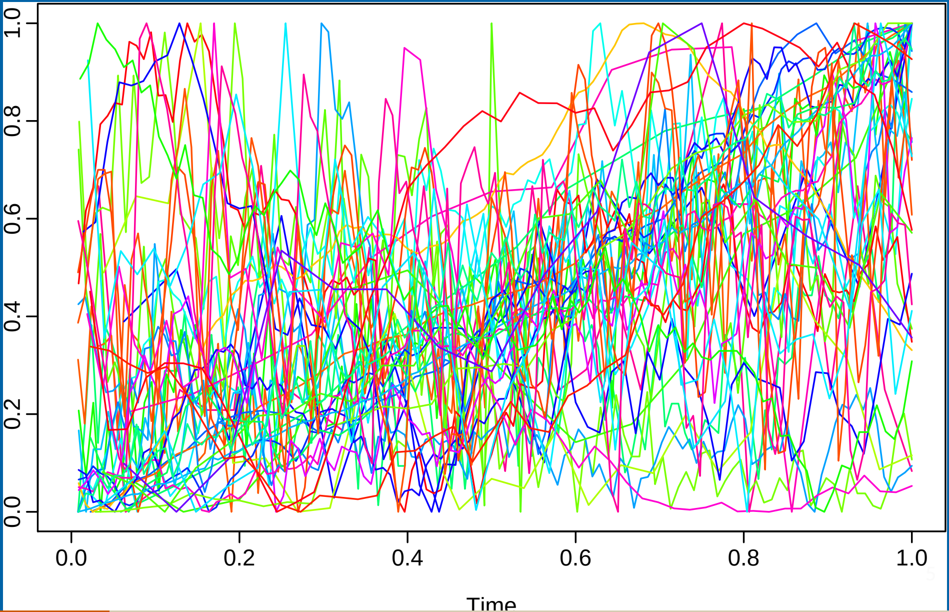
How to plot lots of time series?



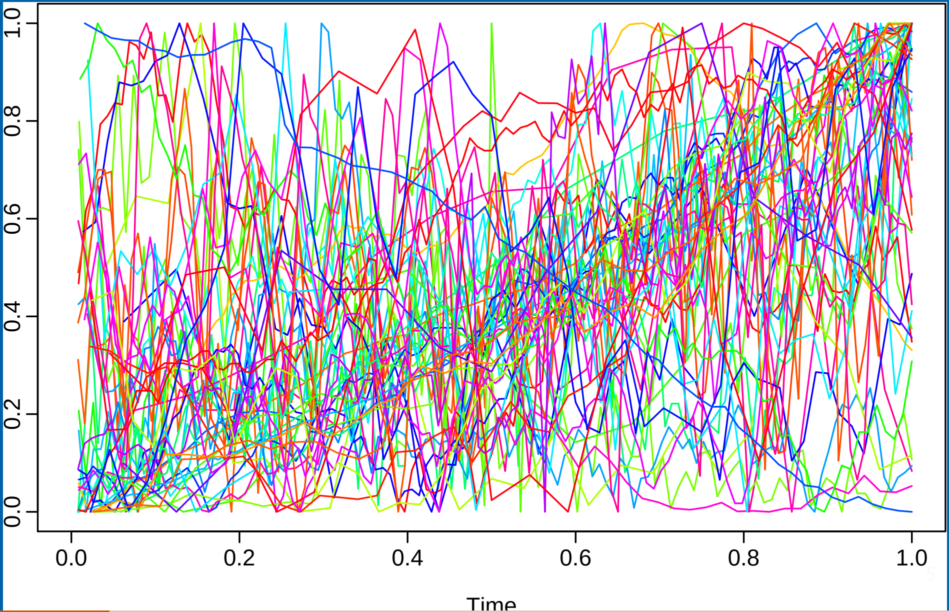
How to plot lots of time series?



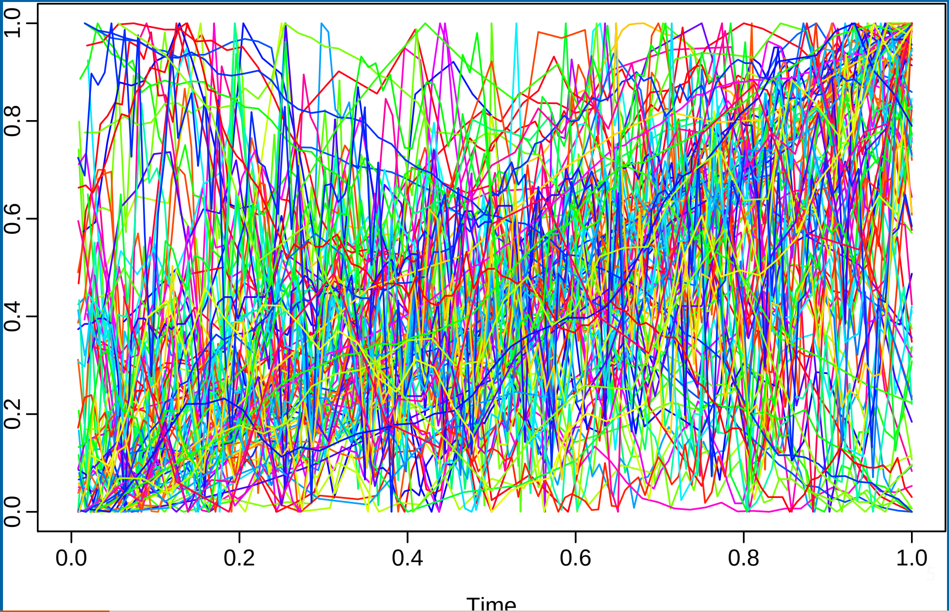
How to plot lots of time series?



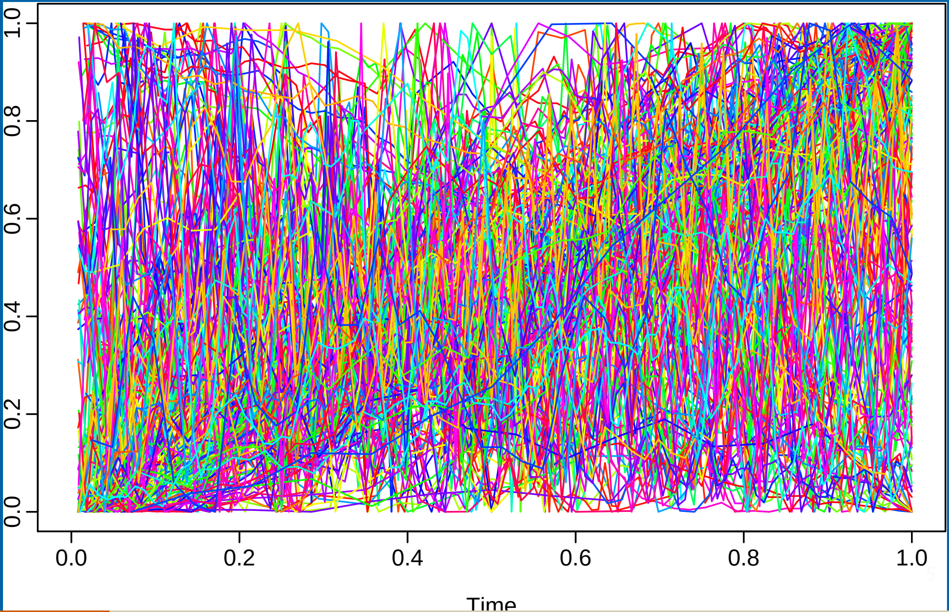
How to plot lots of time series?



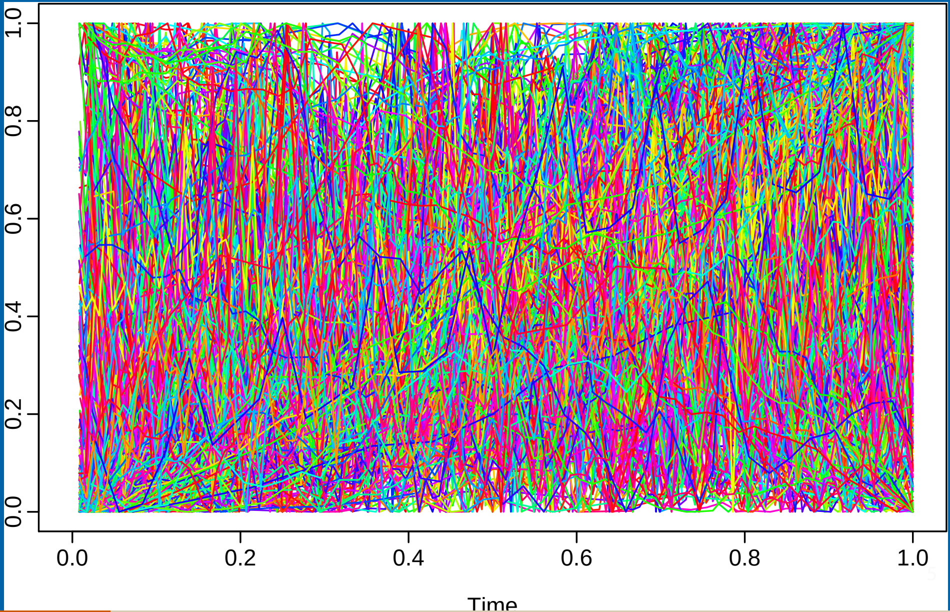
How to plot lots of time series?



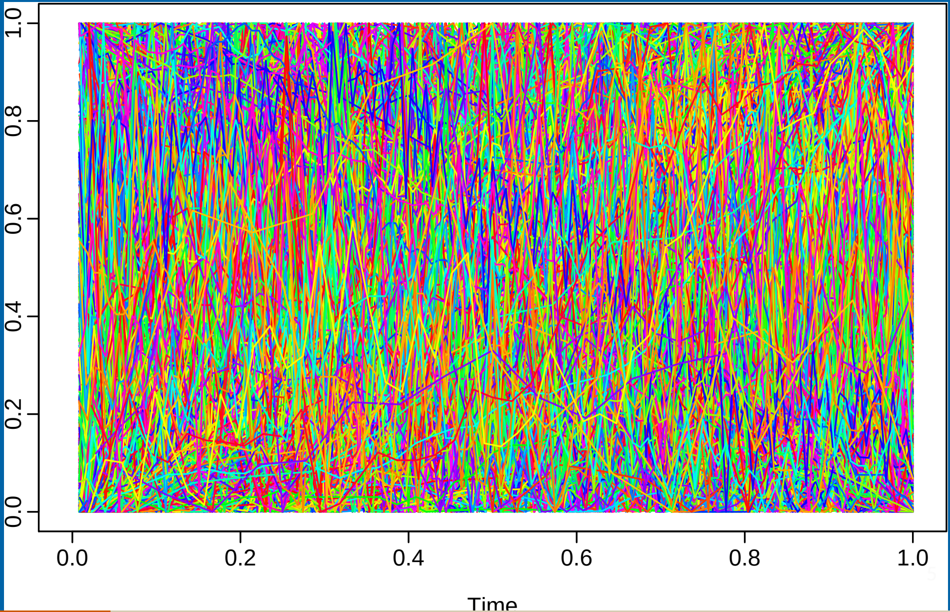
How to plot lots of time series?



How to plot lots of time series?



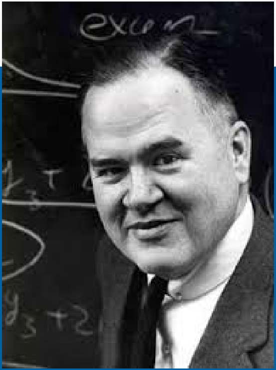
How to plot lots of time series?



Key idea

Cognostics

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

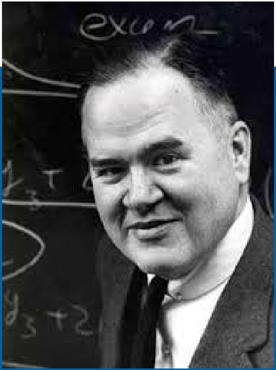


John W Tukey

Key idea

Cognostics

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



John W Tukey

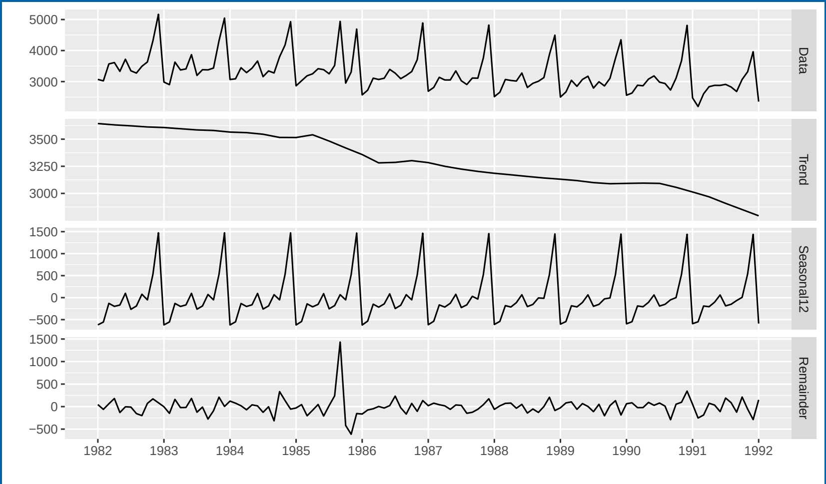
Examples for time series

- 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

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



Candidate features

STL decomposition

$$Y_t = S_t + T_t + R_t$$

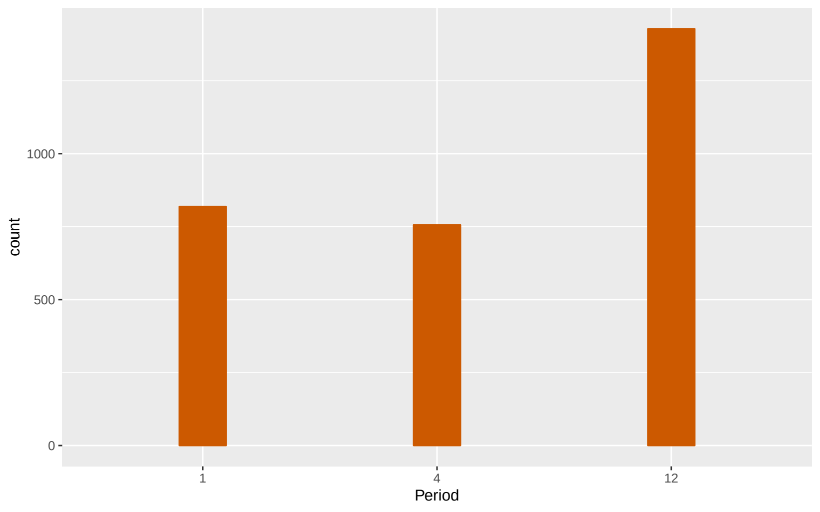
Candidate features

STL decomposition

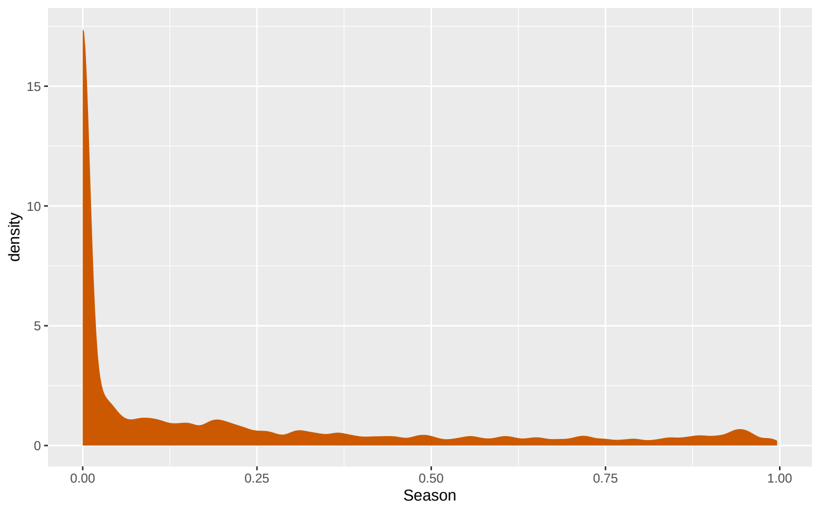
$$Y_t = S_t + T_t + R_t$$

- Seasonal period
- Autocorrelations of data $(Y_1, \dots, Y_T)$
- Autocorrelations of data $(R_1, \dots, R_T)$
- 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

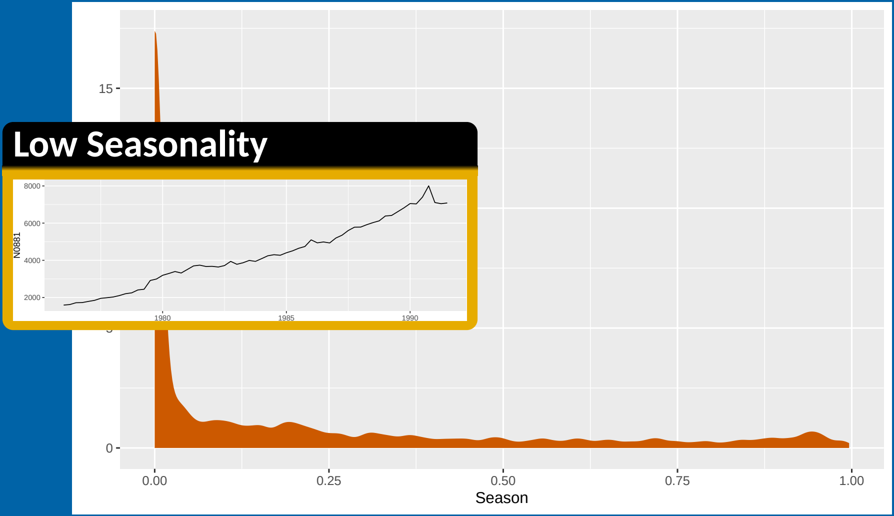
Distribution of Period for M3



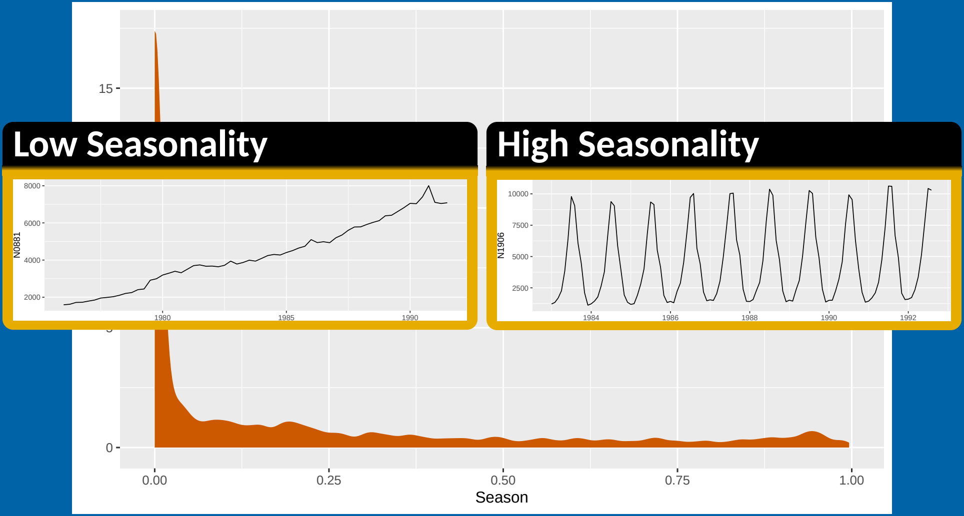
Distribution of Seasonality for M3



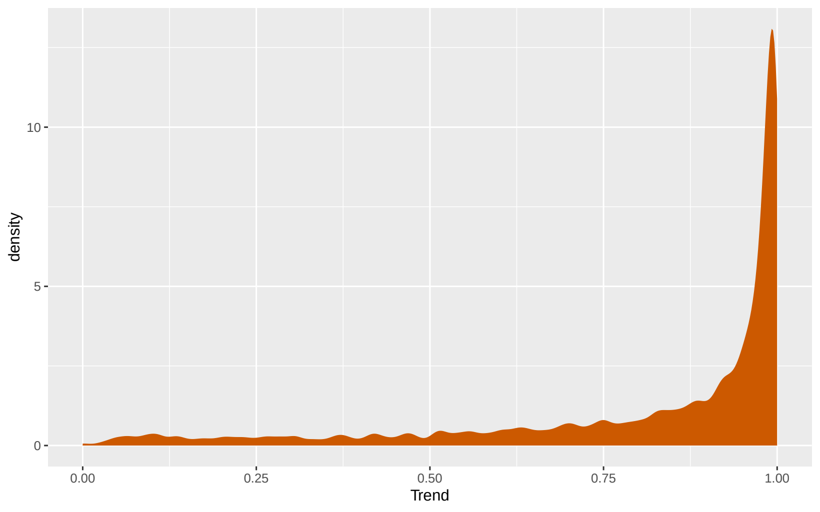
Distribution of Seasonality for M3



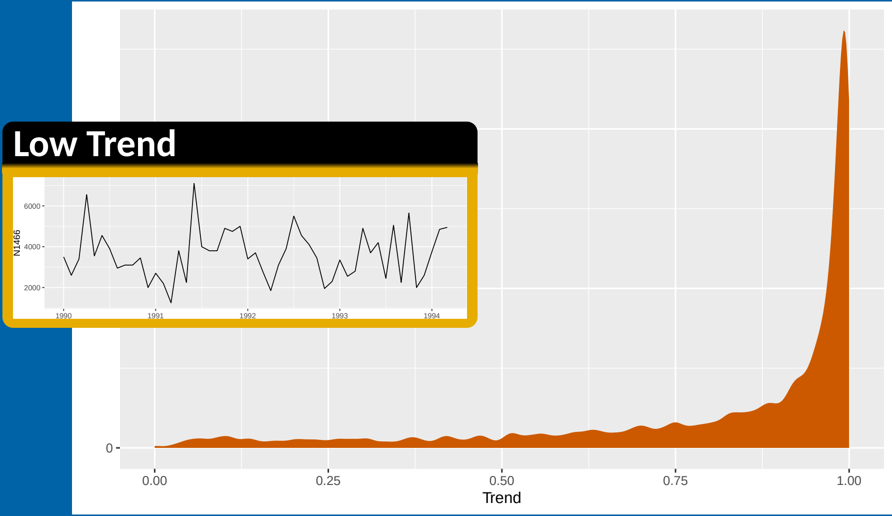
Distribution of Seasonality for M3



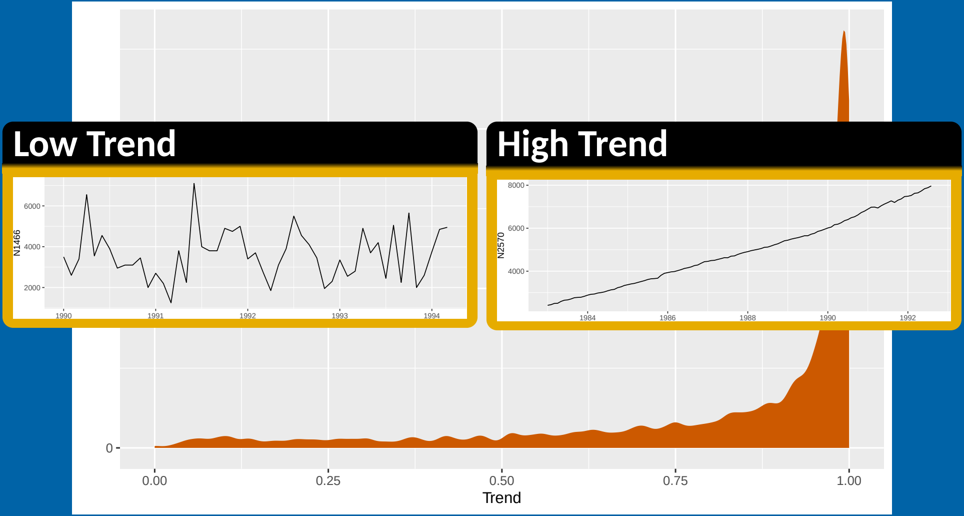
Distribution of Trend for M3



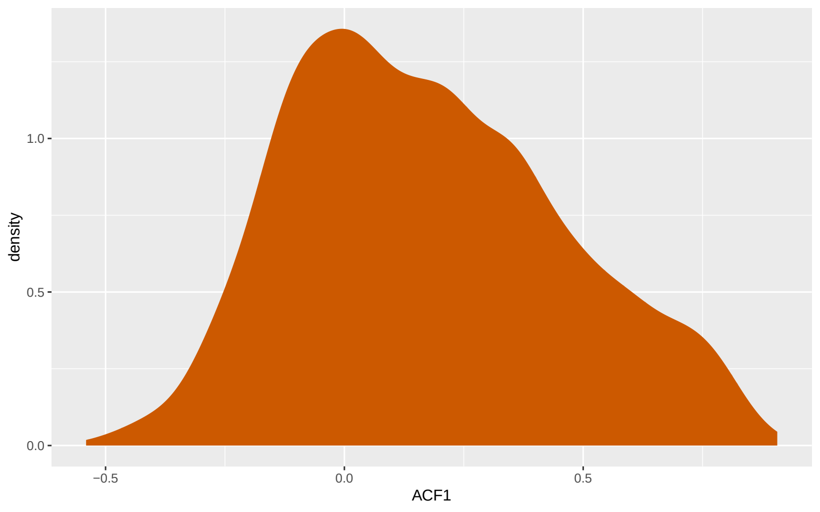
Distribution of Trend for M3



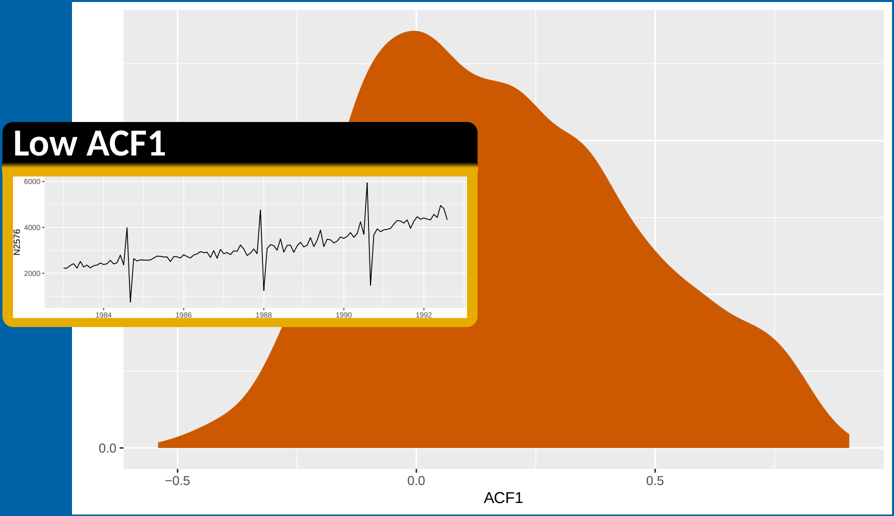
Distribution of Trend for M3



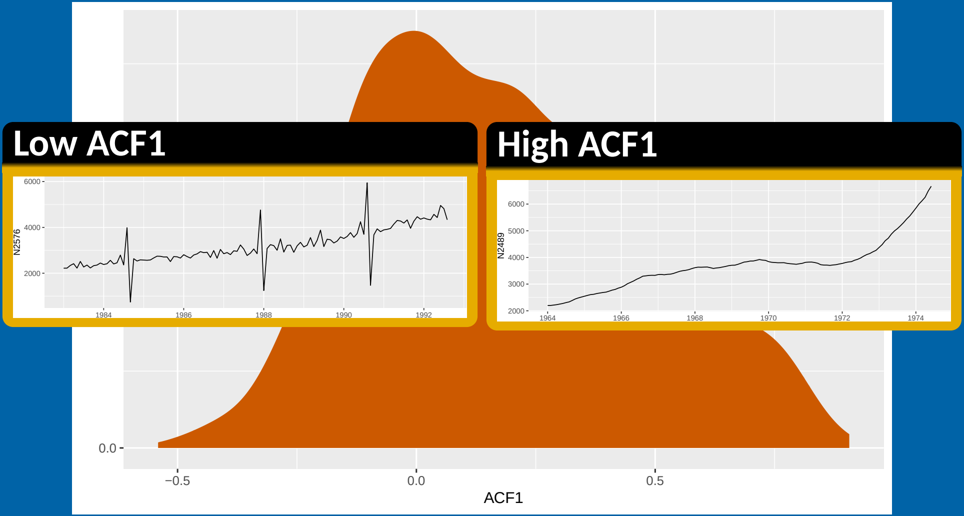
Distribution of Residual ACF1 for M3



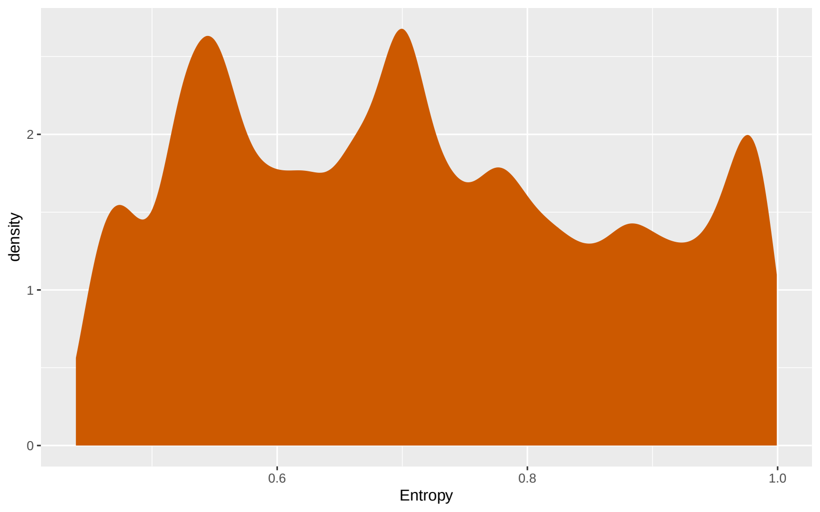
Distribution of Residual ACF1 for M3



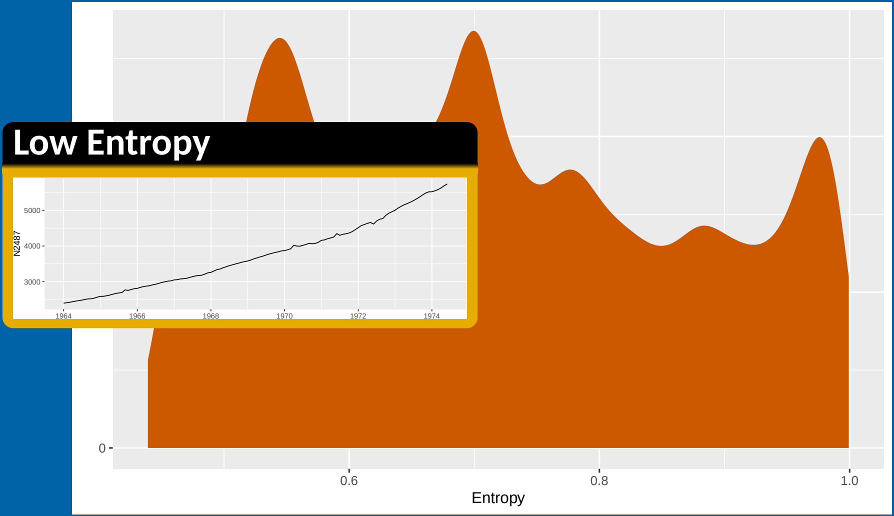
Distribution of Residual ACF1 for M3



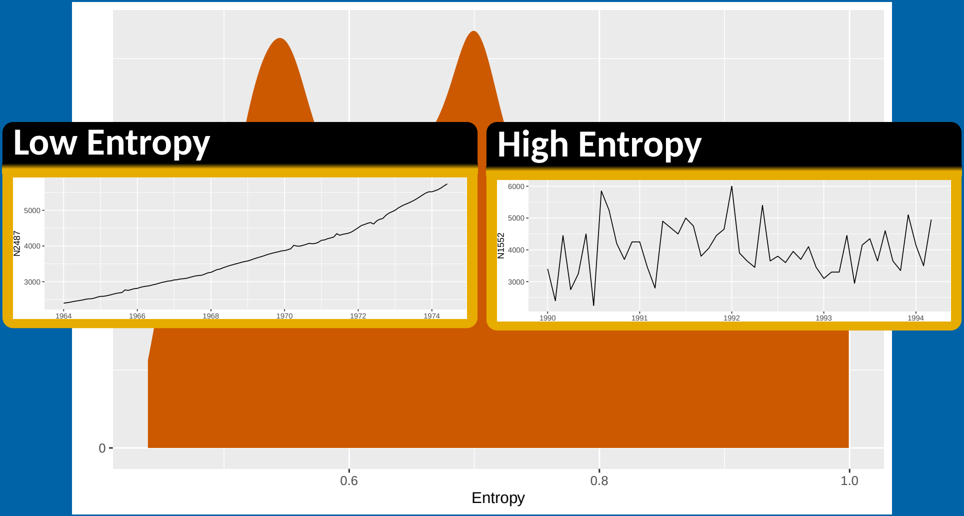
Distribution of Spectral Entropy for M3



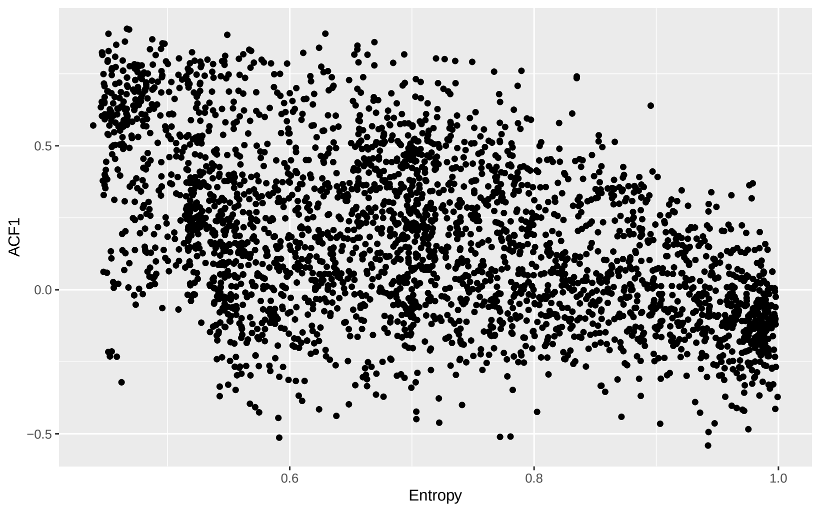
Distribution of Spectral Entropy for M3



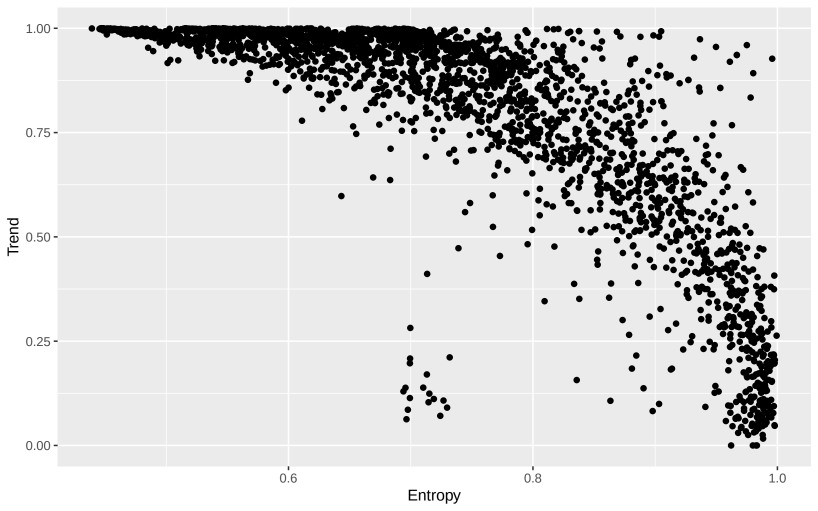
Distribution of Spectral Entropy for M3



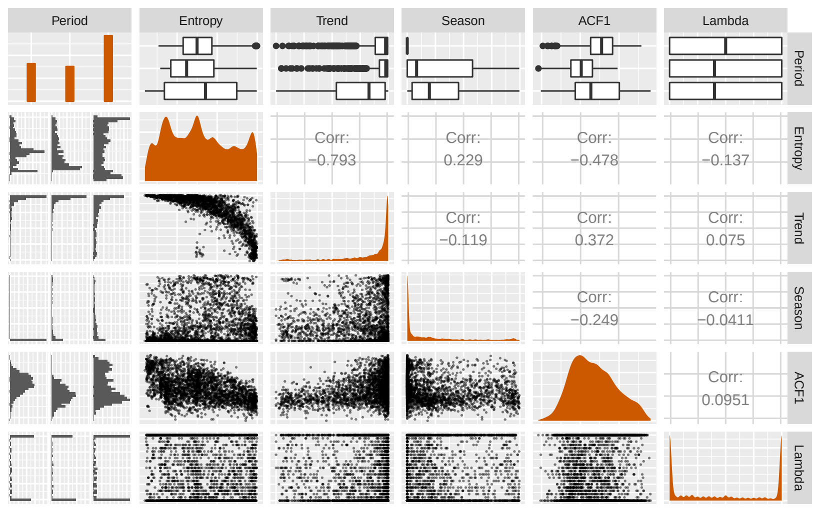
Feature distributions



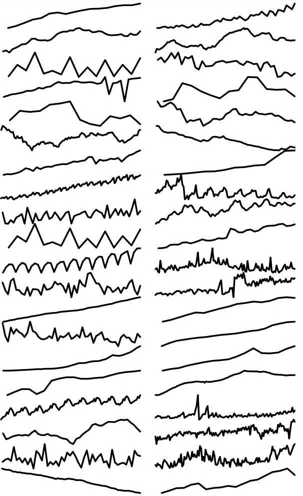
Feature distributions



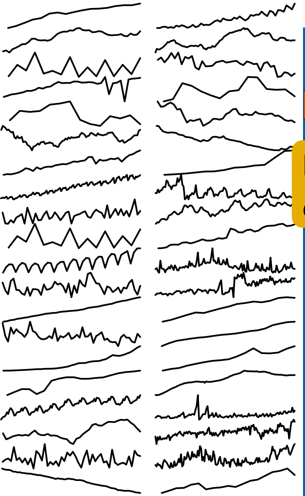
Feature distributions



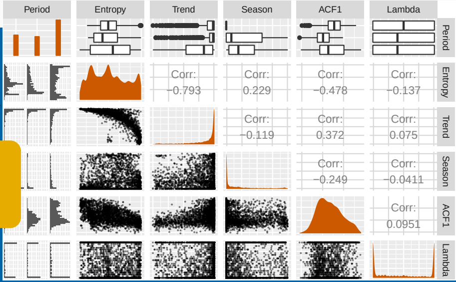
Dimension reduction for time series



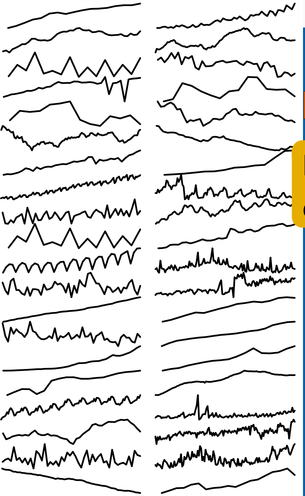
Dimension reduction for time series



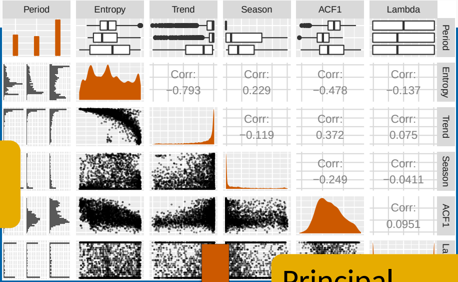
Feature calculation



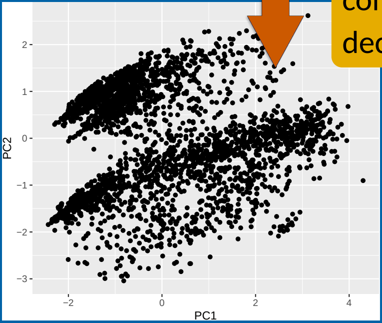
Dimension reduction for time series



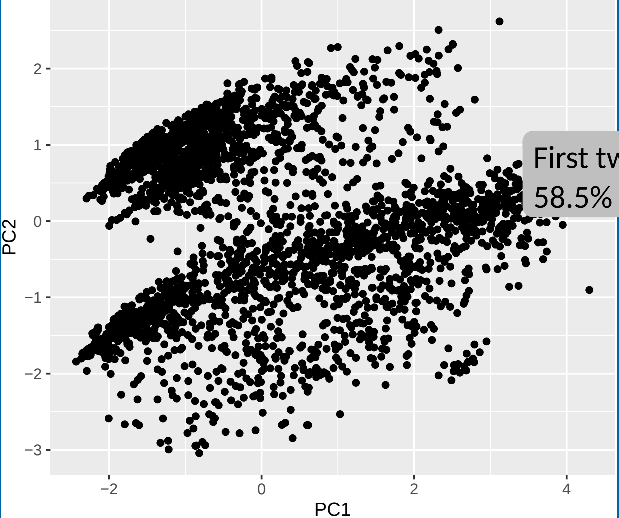
Feature calculation



Principal component decomposition

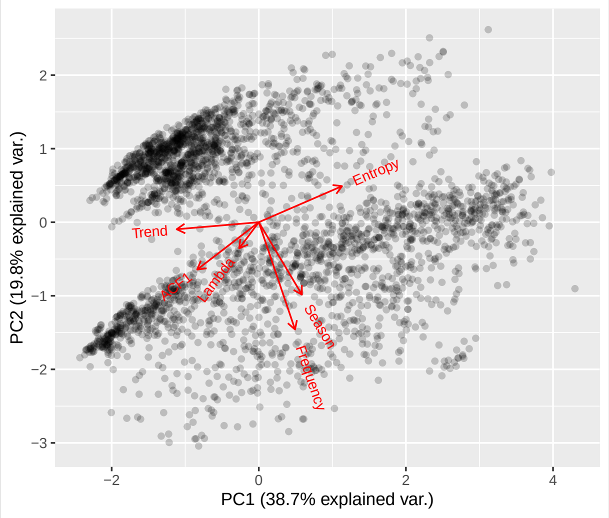


M3 feature space

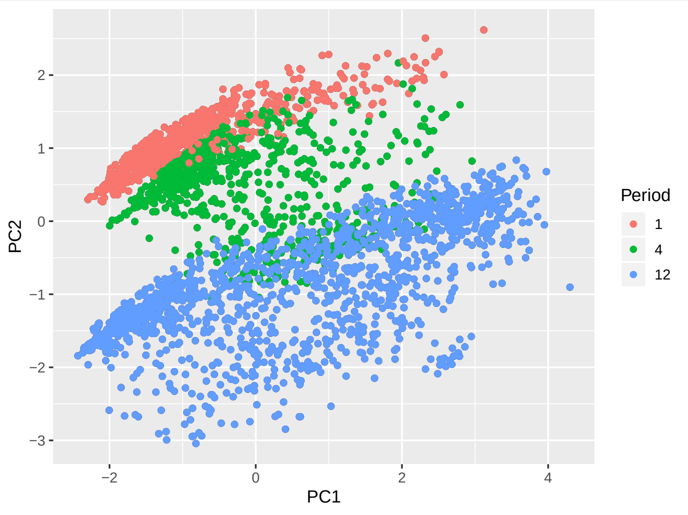


First two PCs explain 58.5% of the variance.

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))
```

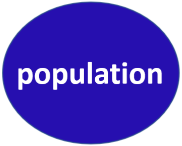
Outline

Forecast model selection

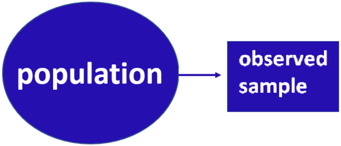
Features used to select a forecasting model

- 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

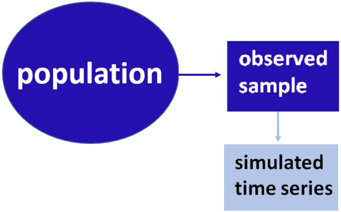
FFORMS: Feature-based FORecast Model Selection



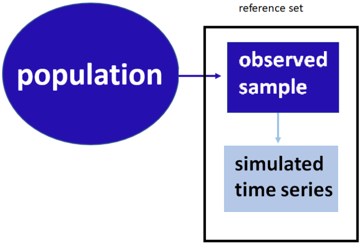
FFORMS: Feature-based FORecast Model Selection



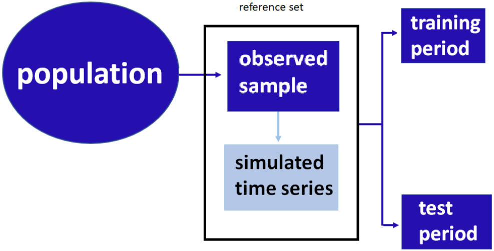
FFORMS: Feature-based FORecast Model Selection



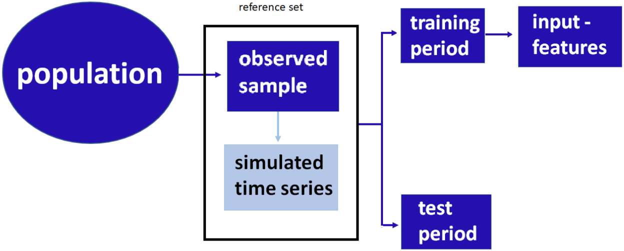
FFORMS: Feature-based FORecast Model Selection



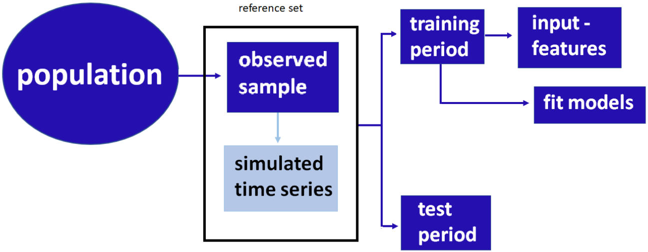
FFORMS: Feature-based FORecast Model Selection



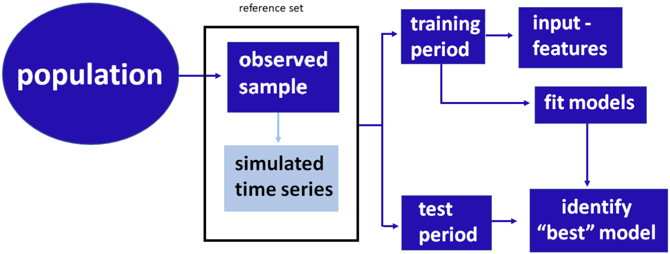
FFORMS: Feature-based FORecast Model Selection



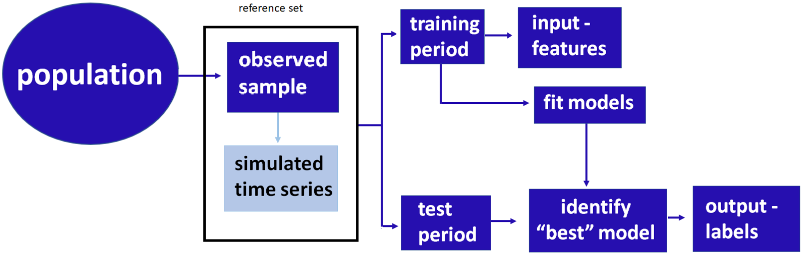
FFORMS: Feature-based FORecast Model Selection



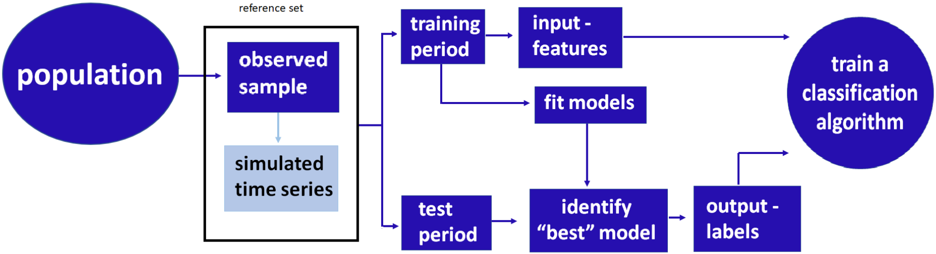
FFORMS: Feature-based FORecast Model Selection



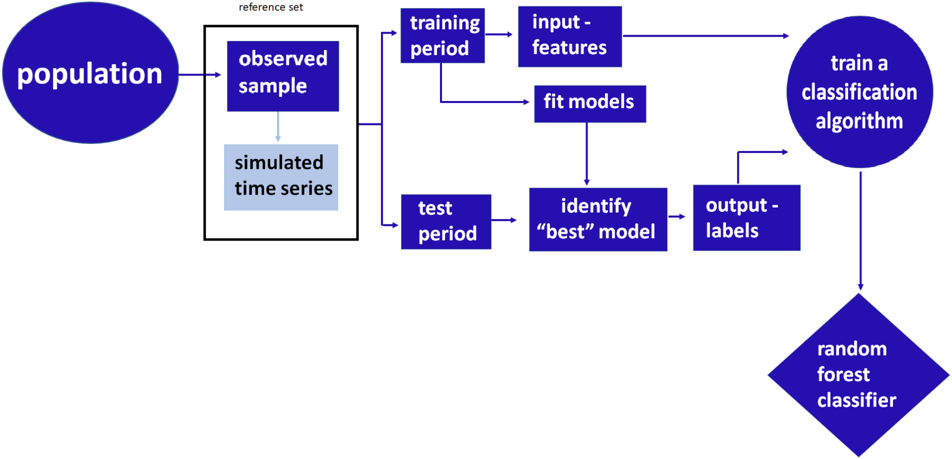
FFORMS: Feature-based FORecast Model Selection



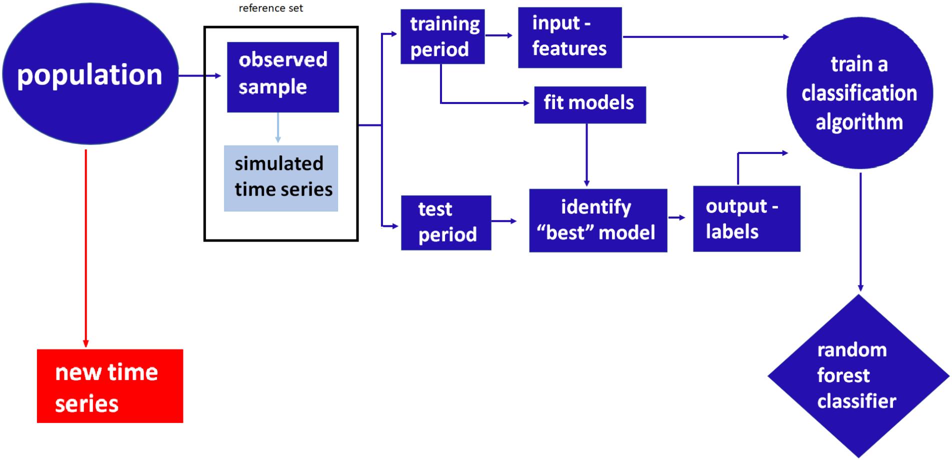
FFORMS: Feature-based FORecast Model Selection



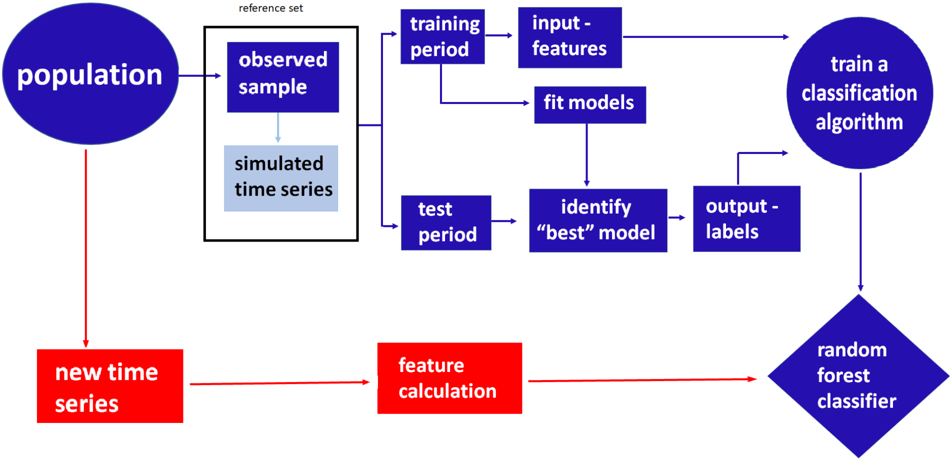
FFORMS: Feature-based FORecast Model Selection



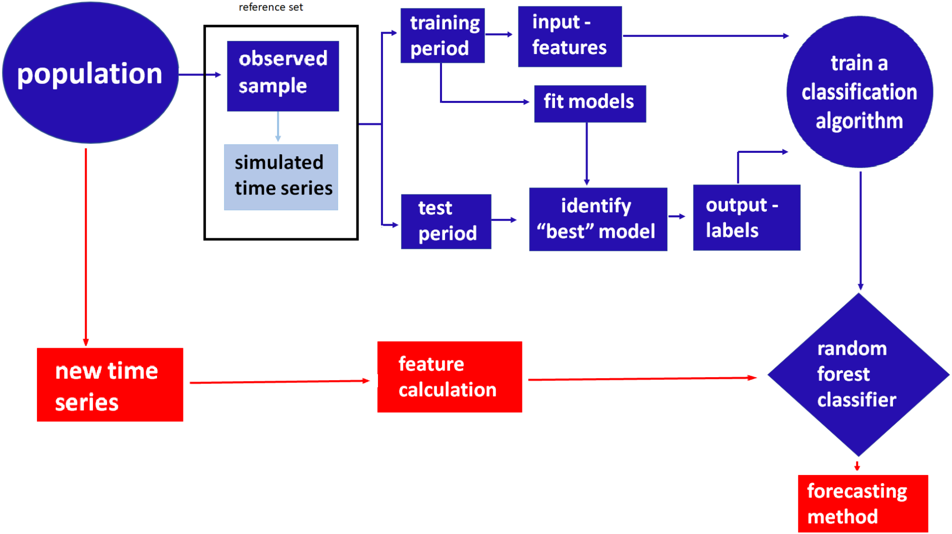
FFORMS: Feature-based FORecast Model Selection



FFORMS: Feature-based FORecast Model Selection



FFORMS: Feature-based FORecast Model Selection



Application to M competition data

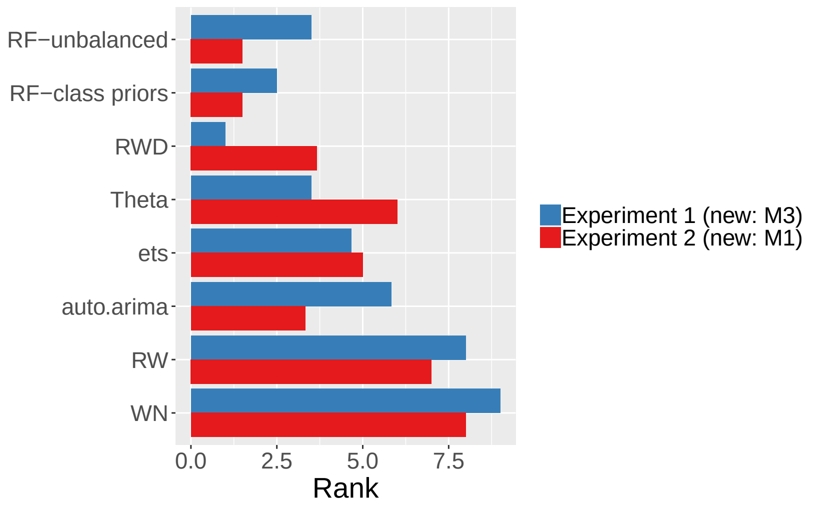
Experiment 1

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

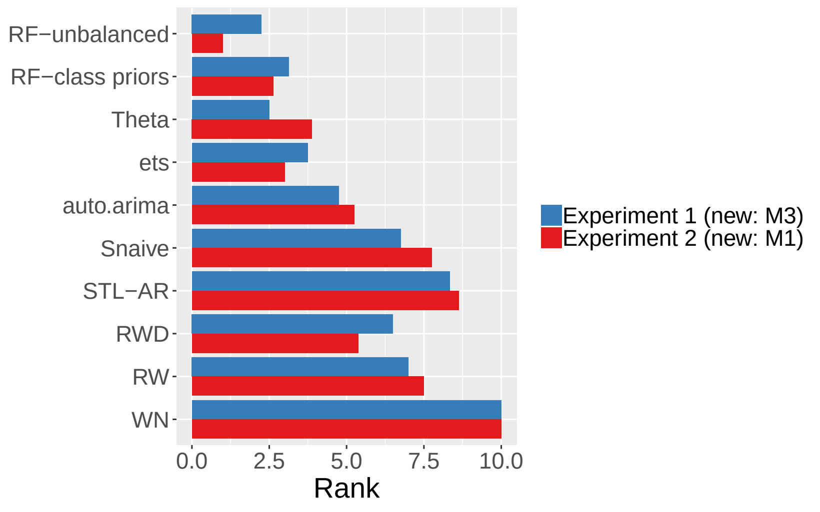
Experiment 2

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

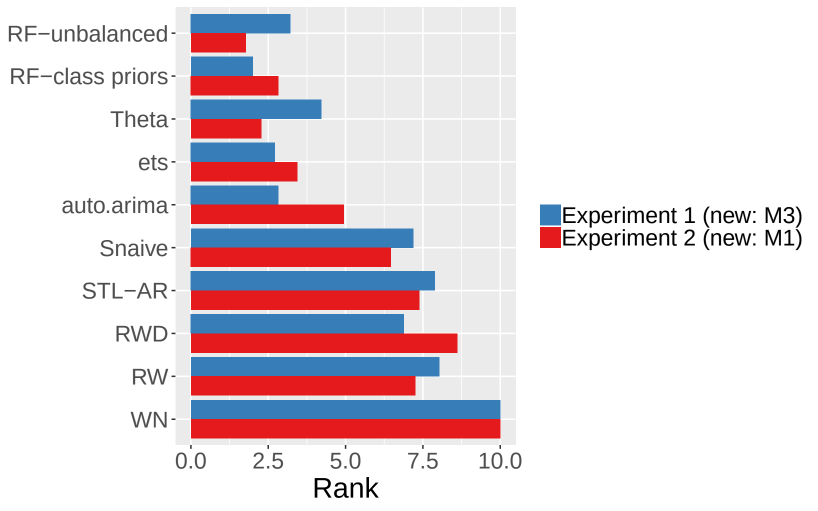
Results: Yearly



Results: Quarterly



Results: Monthly



FFORMA: Feature-based FORecast Model Averaging

- Like FFORMS but we use xgboost 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

Models included

1. Naive
2. Seasonal naive
3. Random walk with drift
4. Theta method
5. ARIMA
6. ETS
7. TBATS
8. STLM-AR

R Packages

- **seer**: FFORMS — selecting forecasting model using features.
  `github.com/thiyangt/seer`

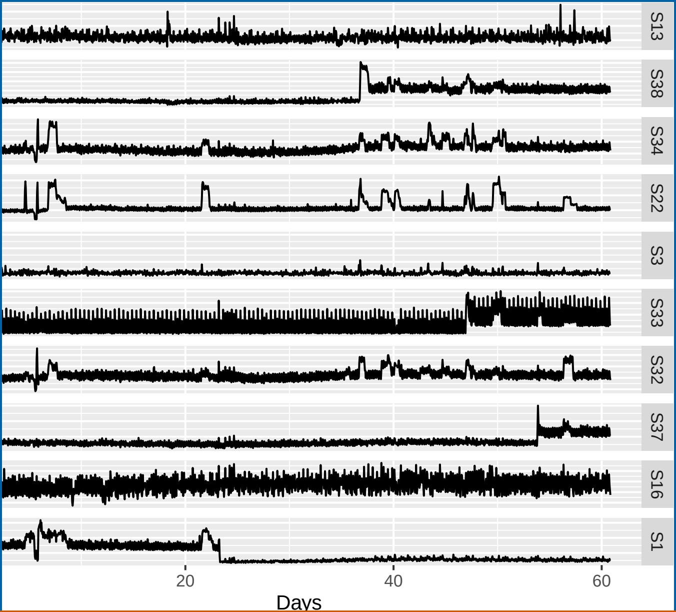
- **M4metalearning**: FFORMA – forecast combinations using features to choose weights.
  `github.com/robjhyndman/M4metalearning`

Outline

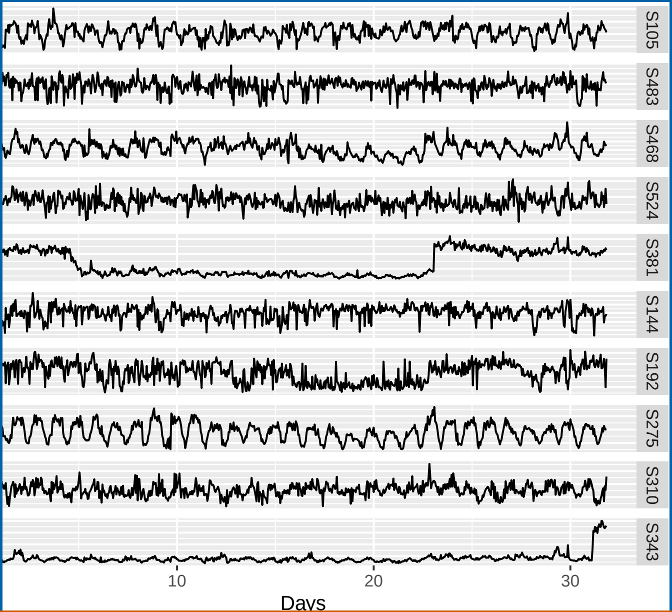
Yahoo server metrics

- Tens of thousands of time series collected at one-hour intervals over 1–2 months.
- Consisting of several server metrics (e.g. CPU usage and paging views) from many server farms globally.
- Aim: find unusual (anomalous) time series.

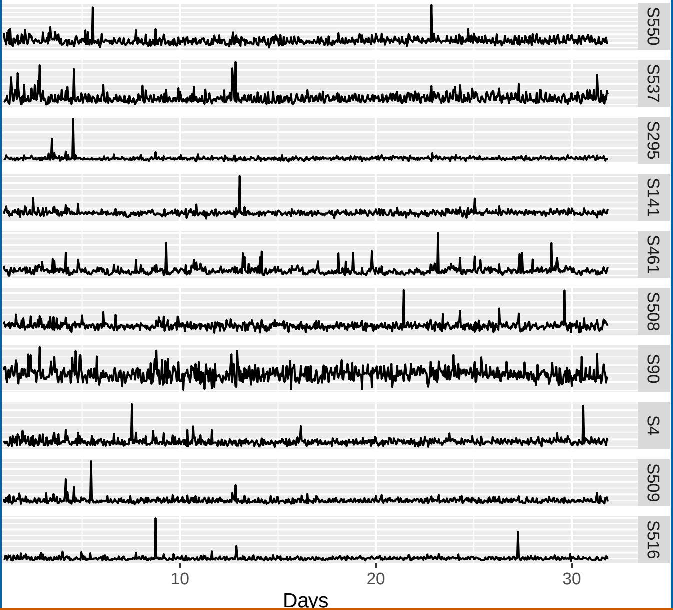
Yahoo server metrics



Yahoo server metrics



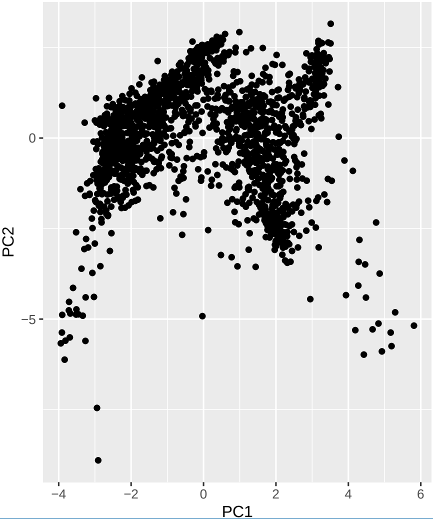
Yahoo server metrics



Yahoo server metrics

- **ACF1**: first order autocorrelation = $\text{Corr}(Y_t, Y_{t-1})$
- Strength of **trend** and **seasonality** based on STL
- Size of seasonal **peak** and **trough**
- Spectral **entropy**
- **Lumpiness**: variance of block variances (block size 24).
- **Spikiness**: variances of leave-one-out variances of STL remainders.
- **Level shift**: Maximum difference in trimmed means of consecutive moving windows of size 24.
- **Variance change**: Max difference in variances of consecutive moving windows of size 24.
- **Flat spots**: Discretize sample space into 10 equal-sized intervals. Find max run length in any interval.
- Number of **crossing points** of mean line.
- **Kullback-Leibler score**: Maximum of $D_{KL}(P\|Q) = \int P(x) \ln P(x)/Q(x) dx$ where P and Q are estimated by kernel density estimators applied to consecutive windows

Feature space



Feature space

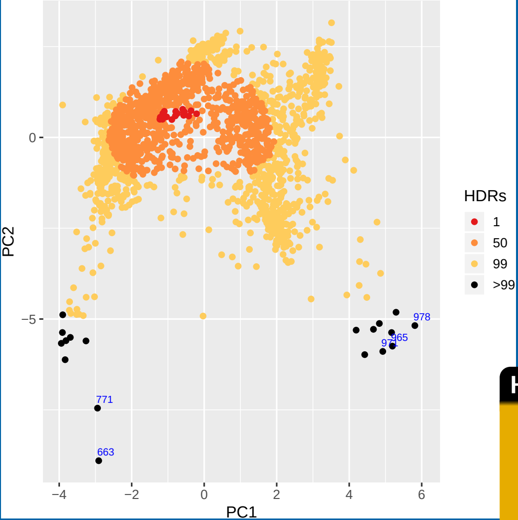
Feature space

What is "anomalous"?

- We need a measure of the "anomalousness" of a time series.
- Rank points based on their local density using a bivariate kernel density estimate.

Finding weird time series

```
hdrcde::hdrscatterplot(pc[,1], pc[,2], noutliers=5)
```



Highest Density Regions

- Estimate using `hdrcde` package
- Highlight outlying points as those with lowest density.

Packages

- **hdrcde**: scatterplots with bivariate HDRs.
  CRAN | `github.com/robjhyndman/hdrcde`
- **stray**: finding outliers in high dimensions.
  `github.com/pridiltal/stray`
- **oddstream**: finding outliers in streaming data.
  `github.com/pridiltal/oddstream`
- **anomalous**: yahoo data.
  `github.com/robjhyndman/anomalous`

Acknowledgments


Earo Wang


Yanfei Kang


Dilini Talagala


Thiyanga Talagala


Pablo Montero-Manso


Mitchell O'Hara-Wild