

Estudo do artigo:

‘FFORMA: Feature-based forecast model averaging’
Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J.
e Talagala, T. S.

André F. B. Menezes | Dezembro, 2020

0 artigo



Contents lists available at [ScienceDirect](#)

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast



FFORMA: Feature-based forecast model averaging

Pablo Montero-Manso*, George Athanasopoulos, Rob J. Hyndman, Thiyanga S. Talagala

Department of Econometrics and Business Statistics, Monash University, Australia

ARTICLE INFO

Keywords:

Time series features
Forecast combination
XGBoost
M4 competition
Meta-learning

ABSTRACT

We propose an automated method for obtaining weighted forecast combinations using time series features. The proposed approach involves two phases. First, we use a collection of time series to train a meta-model for assigning weights to various possible forecasting methods with the goal of minimizing the average forecasting loss obtained from a weighted forecast combination. The inputs to the meta-model are features that are extracted from each series. Then, in the second phase, we forecast new series using a weighted forecast combination, where the weights are obtained from our previously trained meta-model. Our method outperforms a simple forecast combination, as well as all of the most popular individual methods in the time series forecasting literature. The approach achieved second position in the M4 competition.

© 2019 Published by Elsevier B.V. on behalf of International Institute of Forecasters.

Organização

- Contextualização e motivação
- Metodologia
- Implementação computacional
- Aplicação real

Contextualização e motivação

Competições M (Makridakis): Origem

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

Competições M (Makridakis): Origem

- 111 séries temporais foram avaliadas.
- Métodos mais simples (suavização exponencial, média móvel) apresentaram melhor performance do que métodos mais sofisticado (ARMA).
- Estatísticos da época criticaram os resultados:

Again, returning to the authors general approach, it is fair to say that they are comparing what people use in practice. It is amazing to me, however, that after all this exercise in identifying models, transforming and so on, that the autoregressive moving averages come out so badly. I wonder whether it might be partly due to the authors not using the backwards forecasting approach to obtain the initial errors.

-- Dr W. G. Gilchrist (Sheffield City Polytechnic)

Competição M1

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.

- Primeira versão organizada em 1983 pelo professor Spyros Makridakis.
- **Objetivo:** avaliar a acurácia preditiva de diferentes métodos.

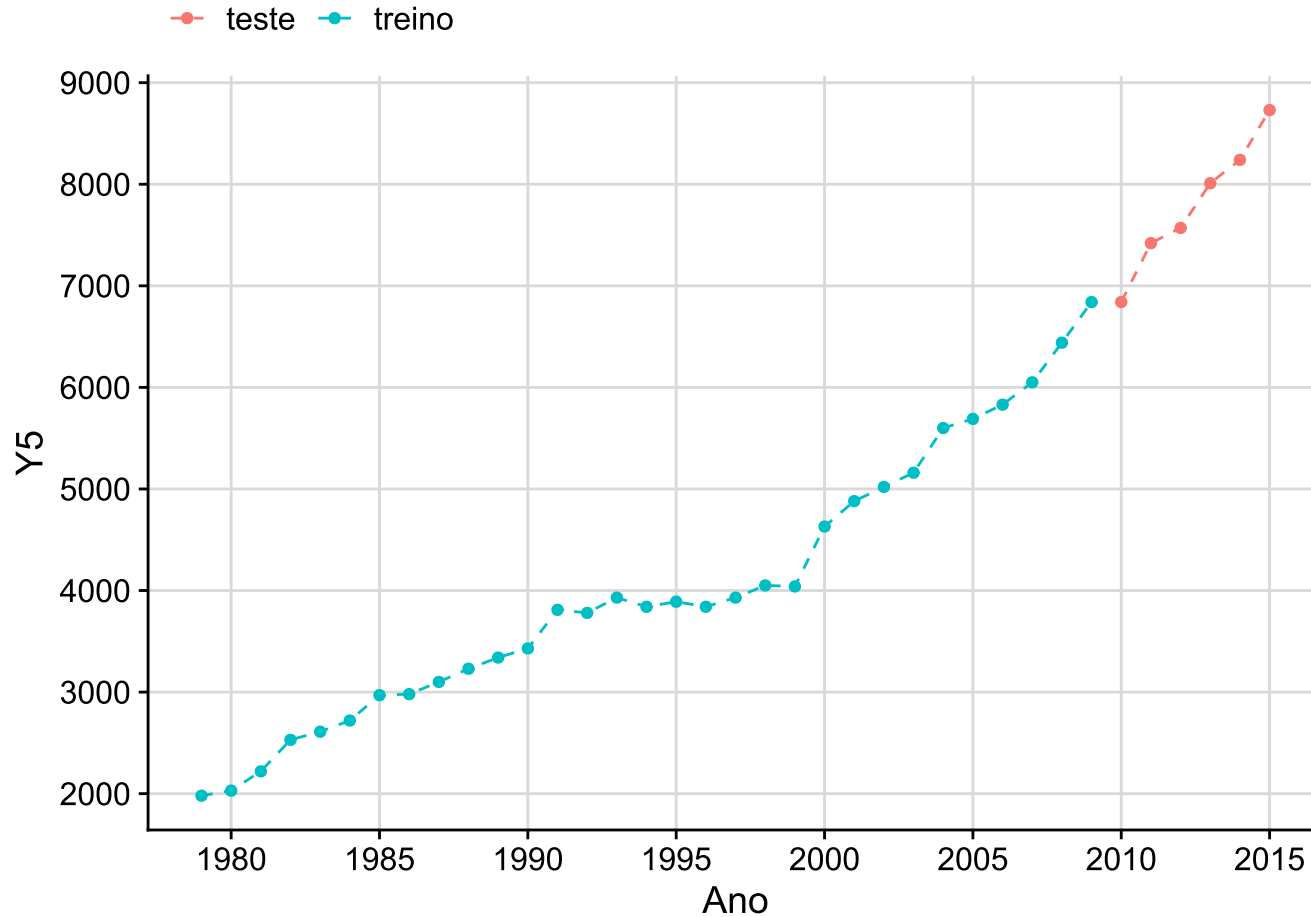
"...anyone could submit forecasts, making this the first true forecasting competition as far as I am aware"

-- Rob Hyndman

Competições M: Edições

- **M1** 1983:
 - 1001 séries: anual, trimestral e mensal.
 - 15 métodos mais 9 variantes foram avaliados.
- **M2** setembro a abril de 1987:
 - 29 séries.
 - Dados de quatro empresas dos EUA,
 - Conduzida em tempo real
- **M3** 2000:
 - 3003 séries.
 - Séries com tamanho entre 14 e 126.
- **M4** janeiro a março de 2018:
 - 100.000 séries: anual, trimestral, mensal, semanal, diária e por hora
 - Códigos devem ser públicos.
 - Performance avaliada pontualmente e intervalarmente.
- **M5** março a junho de 2020.
 - 42.840 séries de venda com estrutura hierárquica fornecida pelo Walmart.
 - Inclusão de variáveis explicativas.
 - Performance avaliada em termos da distribuição preditiva.
 - Maioria das séries exibem "intermitência" (vendas esporádicas, incluindo zeros).

Exemplo M4



Previsão de séries temporais

- Essencialmente, existem duas abordagens para previsão de séries temporais:
 - Previsão considerando um modelo.
 - Previsão combinando (em média) as previsões de vários modelos (*forecast model averaging*).
- A literatura estatística sobre *forecast model averaging* é vasta e teve origem no trabalho de **Bates and Granger (1969)**.
- **Clemen (1989)** e **Timmermann (2006)** realizam uma revisão dos principais métodos de previsão combinada e mostram vantagens empíricas sobre métodos considerando apenas um modelo.

Previsão de séries temporais

- Como combinar as previsões dos modelos?
- Problema conhecido na literatura como *forecast combination puzzle*. (Smith and Wallis, 2009)
- Algumas abordagens:
 1. **Média aritmética:** peso igual para todos os modelos.
 2. **Média ponderada:** protege contra risco de considerar modelos mal ajustados (previsões ruins).
- Como escolher os pesos para cada modelo?

Metodologia

FFORMA: visão geral

- **Montero-Manso, Athanasopoulos, Hyndman, and Talagala (2019)** propõem FFORMA: *Feature-based forecast model averaging*.
 - Modelo de meta-learning (xgboost) para estimar os pesos a partir das características (*features*) das séries.
- *reference set*: **conjunto** de séries temporais que será separado em período de **treino** e **teste**.



- *time series features*: características (*features*) das séries temporais.
 - 42 *features*, entre elas, tamanho da série, tendência, sazonalidade, autocorrelações, curvatura e linearidade.
 - Utiliza a base de treino para cálculo das *features*.
 - São as variáveis explicativas (*inputs*) do *meta-model*.

FFORMA: modelos

- Autores utilizaram um conjunto de 9 métodos/modelos:
 1. `naive` (`naive`).
 2. passeio aleatório com *drift* (`rwf`).
 3. `naive sazonal` (`snaive`).
 4. método theta (`thetaf`) (Assimakopoulos and Nikolopoulos, 2000; Hyndman and Billah, 2003).
 5. algoritmo automático para modelo arima (`auto.arima`).
 6. algoritmo automático para modelo de suavização exponencial (`ets`).
 7. modelo TBATS (`tbats`) (De Livera, Hyndman, and Snyder, 2011).
 8. modelo STL-AR (`stlm`).
 9. redes neurais para séries temporais (`nnetar`).
- Detalhes sobre seleção automática dos modelos arima e suavização exponencial são apresentados em Hyndman and Khandakar (2008).
- Detalhes sobre os métodos `stlm` e `nnetar` podem ser vistos em Hyndman and Athanasopoulos (2018).
- Os métodos são ajustados na base de **treino** e erros de previsão são obtidos na base de **teste**.

FFORMA: métrica de previsão

- Métricas do erro de previsão de **cada método** são calculadas para **cada série**.
- Os autores utilizam uma versão modificada do *Overall Weighted Average error* (OWA).
- Seja $e_t = y_t - \hat{y}_t$ o erro de previsão. A versão modificada do OWA utiliza as seguintes métricas:
 - *Mean Absolute Scaled Error* (MASE), proposto por **Hyndman and Koehler (2006)**:

$$\text{MASE} = \frac{1}{n} \sum_{t=1}^n |q_t|, \quad \text{em que} \quad q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|}$$

- *Symmetric Mean Absolute Percentage Error* (sMAPE):

$$\text{sMAPE} = \frac{1}{n} \sum_{t=1}^n \frac{200 |y_t - \hat{y}_t|}{(y_t + \hat{y}_t)}$$

FFORMA: métrica de previsão

- Seja m um método de previsão para cada série e modelo os autores calculam

$$\text{OWA} = \frac{1}{2} \left[\frac{\text{MASE}(m)}{\text{MASE}(\text{snaive2})} + \frac{\text{sMAPE}(m)}{\text{sMAPE}(\text{snaive2})} \right]$$

em que $\text{sMAPE}(\text{snaive2})$ e $\text{sMAPE}(\text{snaive2})$ indicam o MASE e sMAPE do modelo naive sazonal proposto na competição M4.

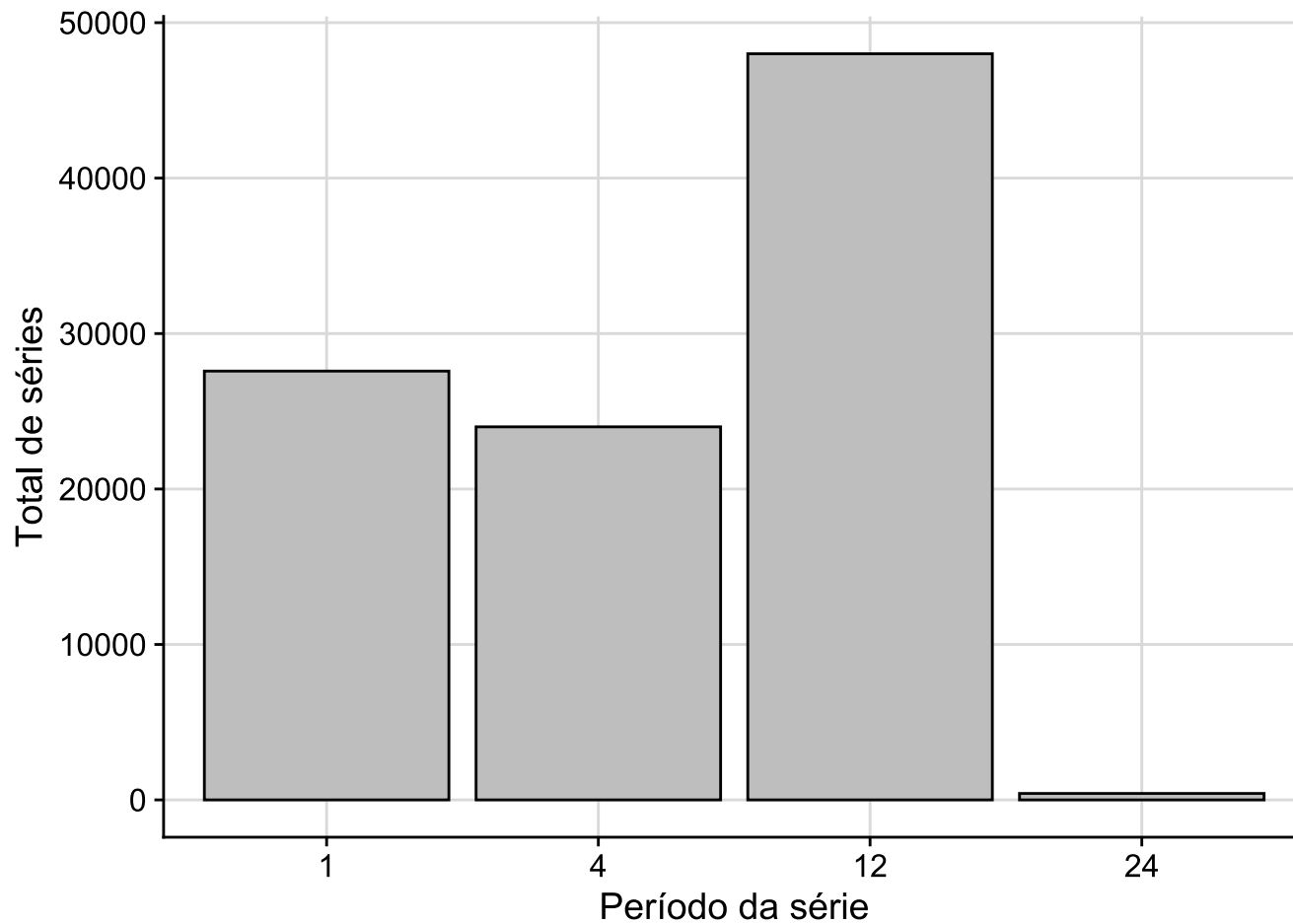
FFORMA: *features*

- *features*: estatísticas que caracterizam uma série temporal.
- Motivado pelos trabalhos:
 - [Hyndman, Wang, and Laptev \(2015\)](#): "Large-scale unusual time series detection".
 - [Kang, Hyndman, and Smith-Miles \(2017\)](#): "Visualising forecasting algorithm performance using time series instance spaces".
 - [Talagala, Hyndman, and Athanasopoulos \(2018\)](#): "Meta-learning how to forecast time series".
- Detalhes teóricos das *features* podem ser encontrados em [tsfeatures](#).

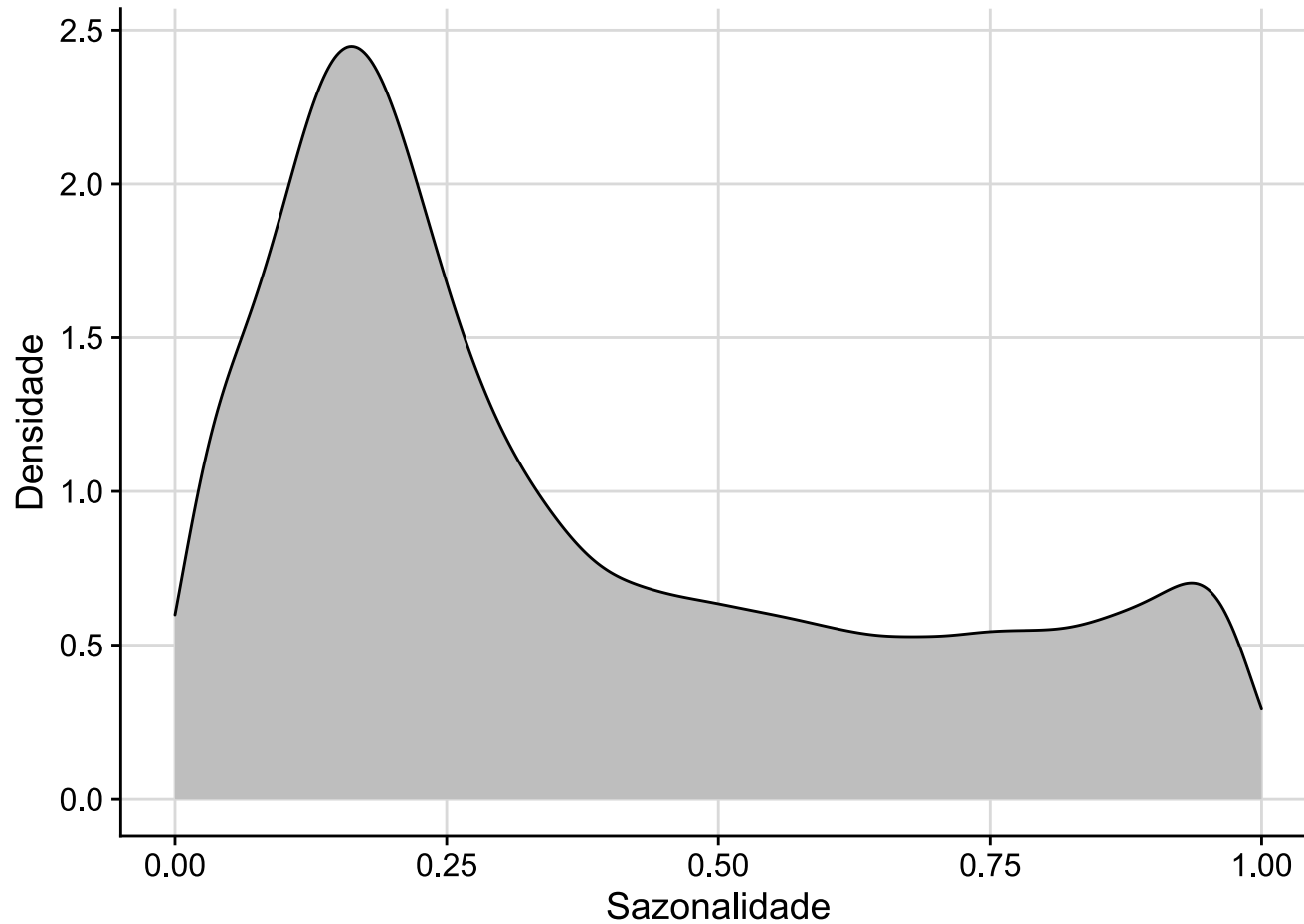
FFORMA: *features*

Feature	Description	Non-seasonal	Seasonal	
1	T	length of time series	✓	✓
2	trend	strength of trend	✓	✓
3	seasonality	strength of seasonality	–	✓
4	linearity	linearity	✓	✓
5	curvature	curvature	✓	✓
6	spikiness	spikiness	✓	✓
7	e_acf1	first ACF value of remainder series	✓	✓
8	e_acf10	sum of squares of first 10 ACF values of remainder series	✓	✓
9	stability	stability	✓	✓
10	lumpiness	lumpiness	✓	✓
11	entropy	spectral entropy	✓	✓
12	hurst	Hurst exponent	✓	✓
13	nonlinearity	nonlinearity	✓	✓
13	alpha	ETS(A,A,N) $\hat{\alpha}$	✓	✓
14	beta	ETS(A,A,N) $\hat{\beta}$	✓	✓
15	hwalpha	ETS(A,A,A) $\hat{\alpha}$	–	✓
16	hwbeta	ETS(A,A,A) $\hat{\beta}$	–	✓
17	hwgamma	ETS(A,A,A) $\hat{\gamma}$	–	✓
18	ur_pp	test statistic based on Phillips–Perron test	✓	✓
19	ur_kpss	test statistic based on KPSS test	✓	✓
20	y_acf1	first ACF value of the original series	✓	✓
21	diff1y_acf1	first ACF value of the differenced series	✓	✓
22	diff2y_acf1	first ACF value of the twice-differenced series	✓	✓
23	y_acf10	sum of squares of first 10 ACF values of original series	✓	✓
24	diff1y_acf10	sum of squares of first 10 ACF values of differenced series	✓	✓
25	diff2y_acf10	sum of squares of first 10 ACF values of twice-differenced series	✓	✓
26	seas_acf1	autocorrelation coefficient at first seasonal lag	–	✓
27	sediff_acf1	first ACF value of seasonally differenced series	–	✓
28	y_pacf5	sum of squares of first 5 PACF values of original series	✓	✓
29	diff1y_pacf5	sum of squares of first 5 PACF values of differenced series	✓	✓
30	diff2y_pacf5	sum of squares of first 5 PACF values of twice-differenced series	✓	✓
31	seas_pacf	partial autocorrelation coefficient at first seasonal lag	✓	✓
32	crossing_point	number of times the time series crosses the median	✓	✓
33	flat_spots	number of flat spots, calculated by discretizing the series into 10 equal-sized intervals and counting the maximum run length within any single interval	✓	✓
34	nperiods	number of seasonal periods in the series	–	✓
35	seasonal_period	length of seasonal period	–	✓
36	peak	strength of peak	✓	✓
37	trough	strength of trough	✓	✓
38	ARCH.LM	ARCH LM statistic	✓	✓
39	arch_acf	sum of squares of the first 12 autocorrelations of z^2	✓	✓
40	garch_acf	sum of squares of the first 12 autocorrelations of r^2	✓	✓
41	arch_r2	R^2 value of an AR model applied to z^2	✓	✓
42	garch_r2	R^2 value of an AR model applied to r^2	✓	✓

Distribuição do Período

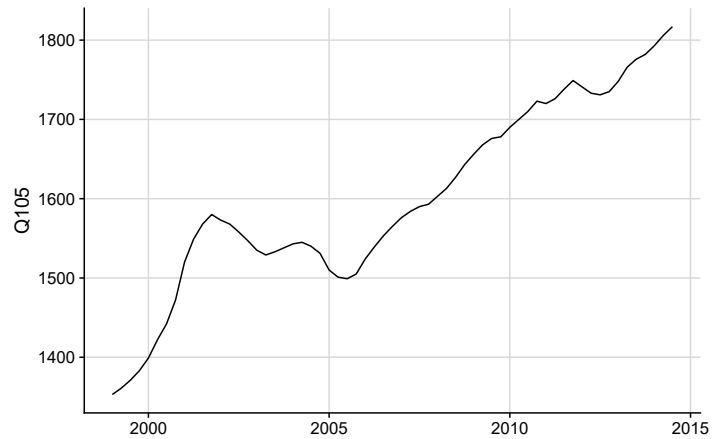


Distribuição da Sazonalidade

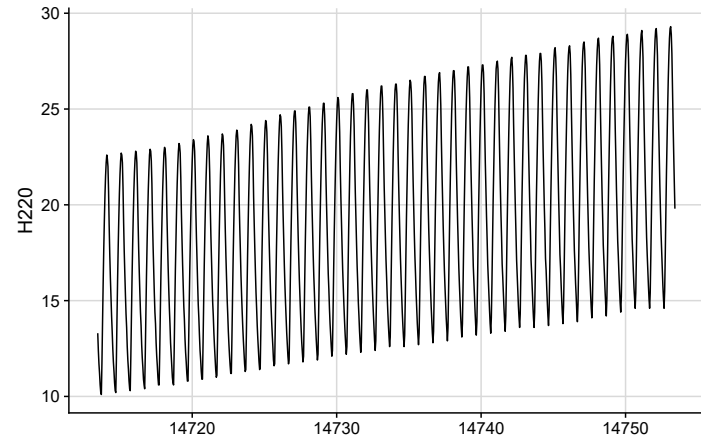


Distribuição da Sazonalidade

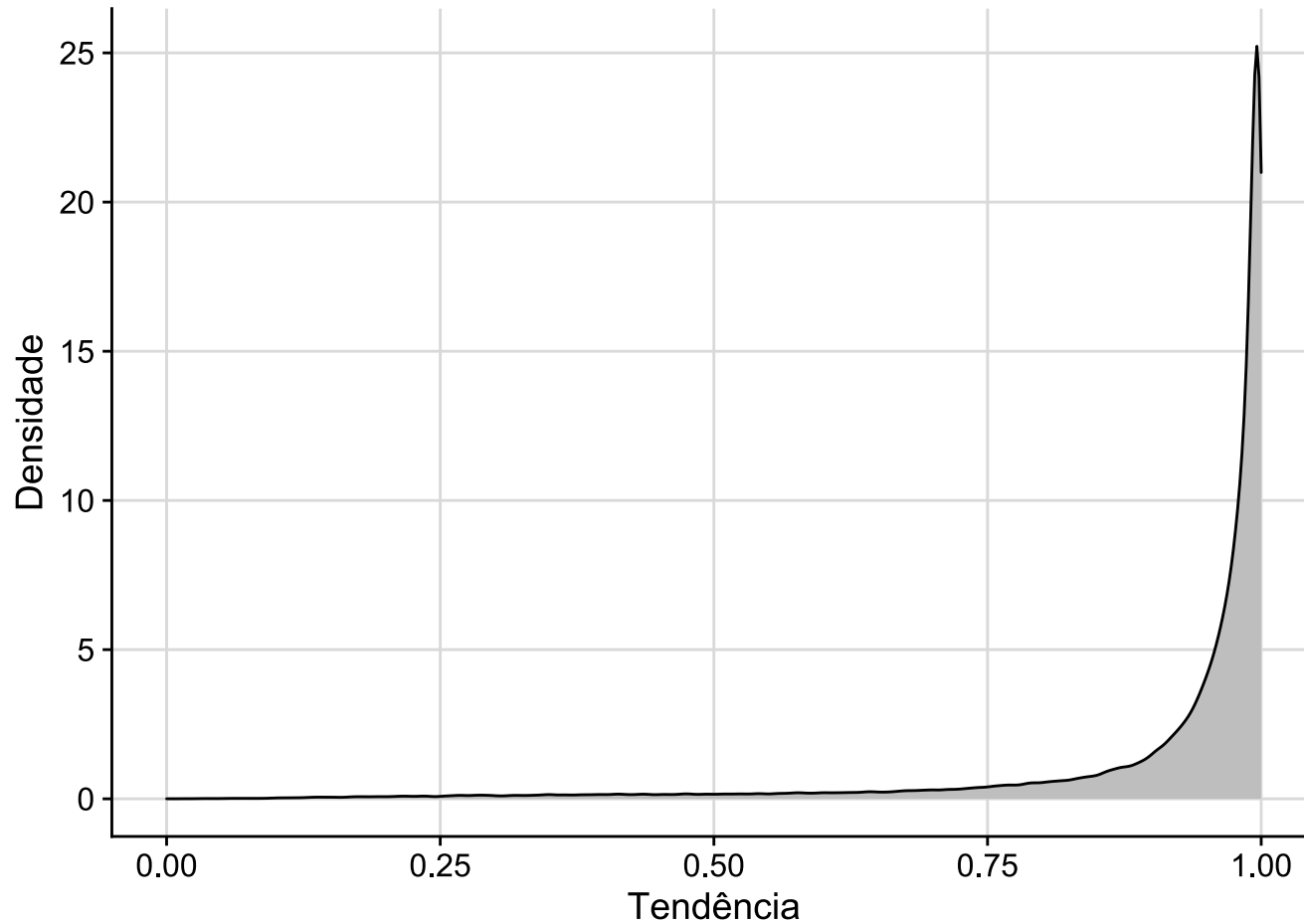
- Baixa sazonalidade



- Alta sazonalidade

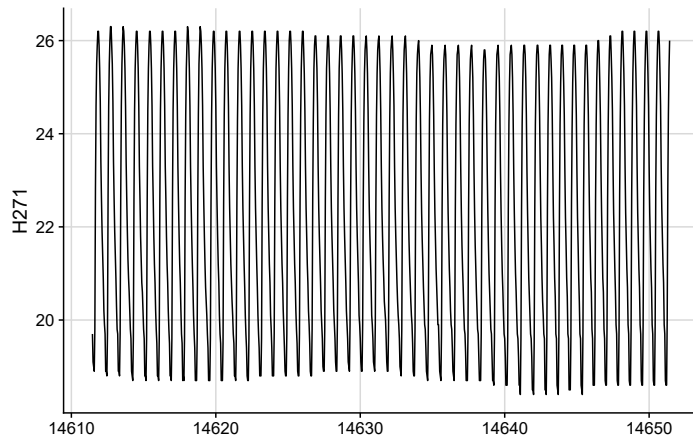


Distribuição da Tendência

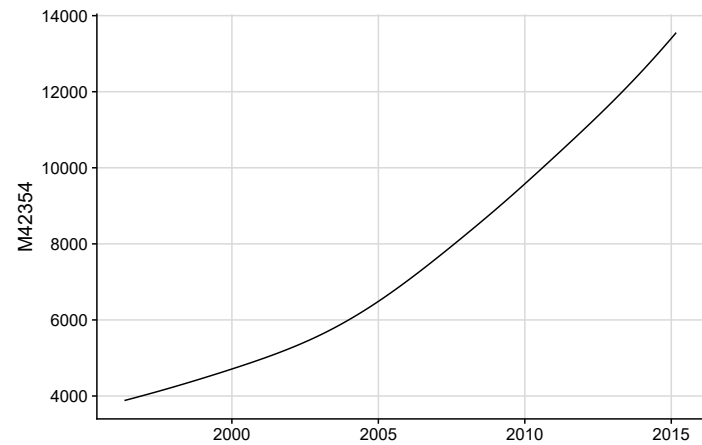


Distribuição da Tendência

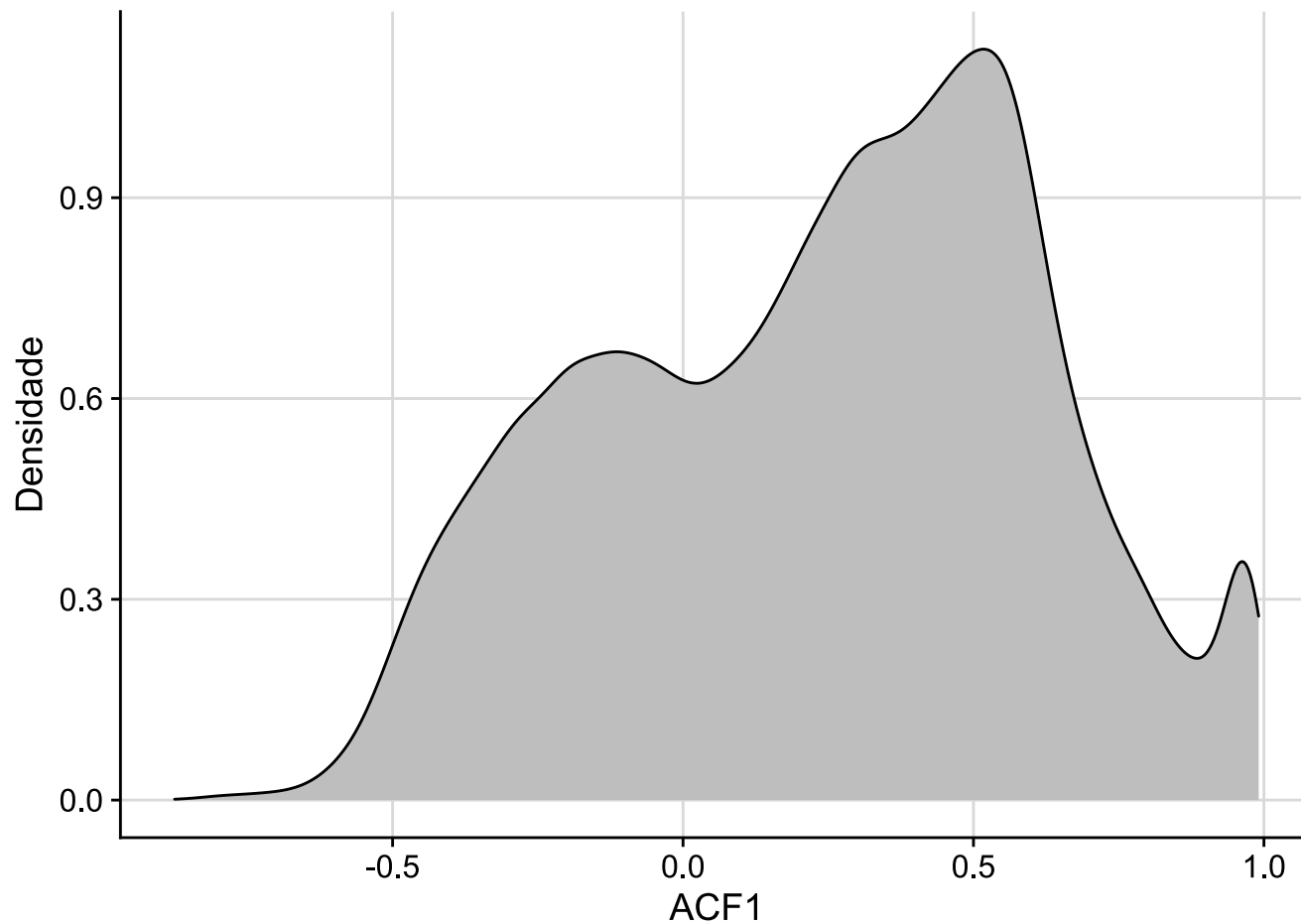
- Baixa tendência



- Alta tendência

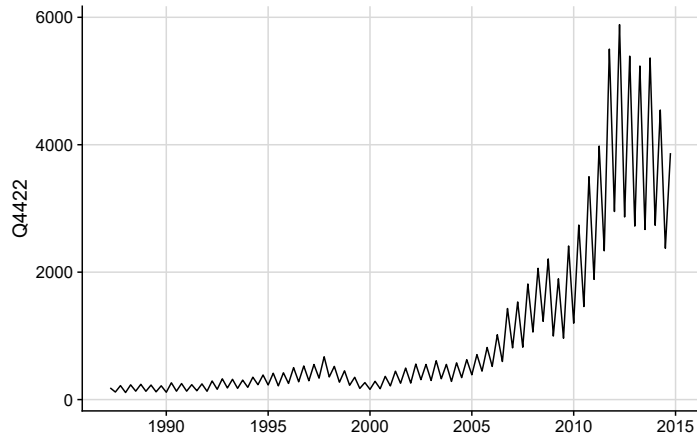


Distribuição da ACF1



Distribuição da ACF1

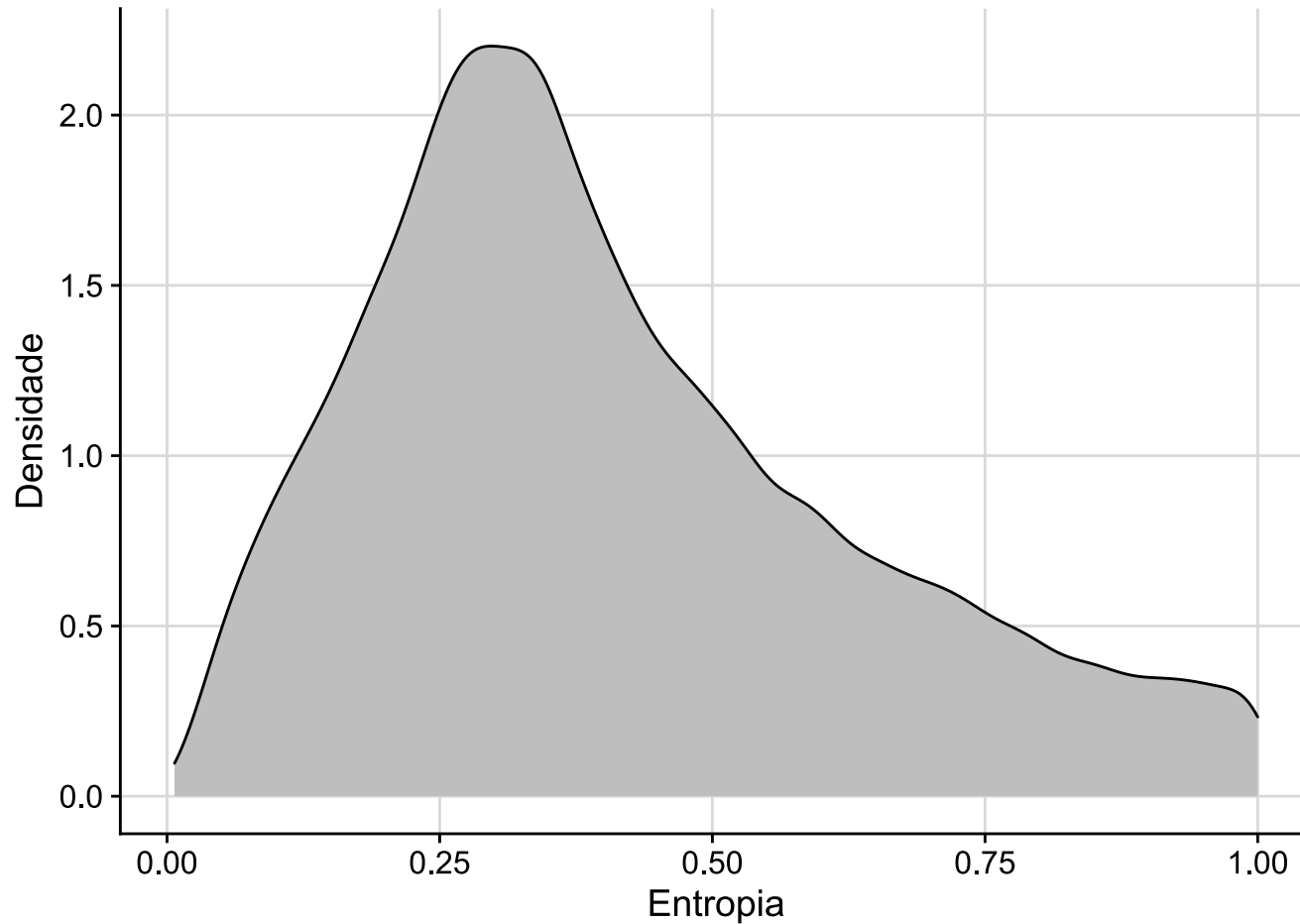
- Baixa ACF1



- Alta ACF1

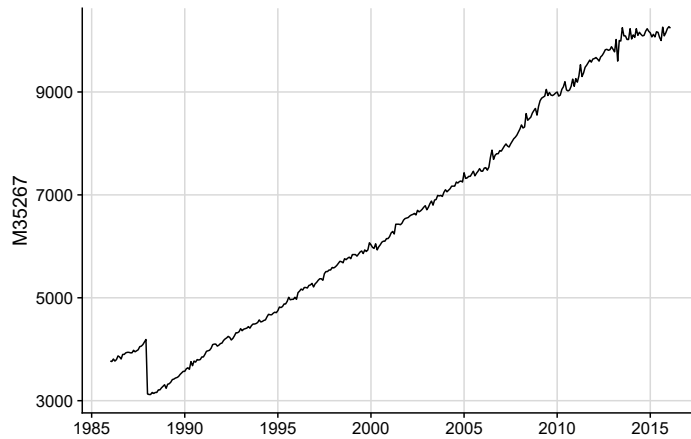


Distribuição da Entropia

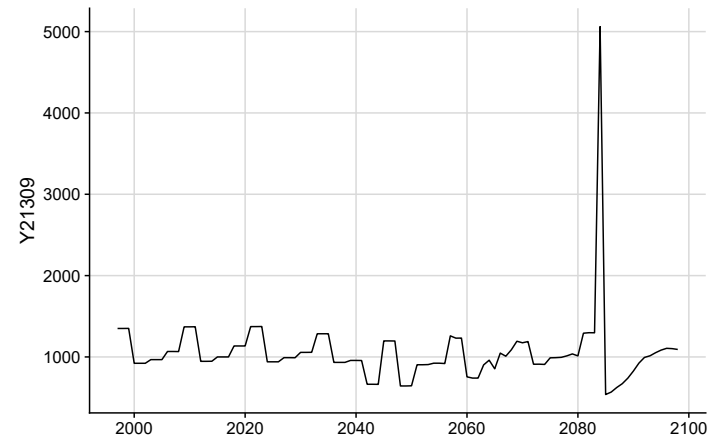


Distribuição da Entropia

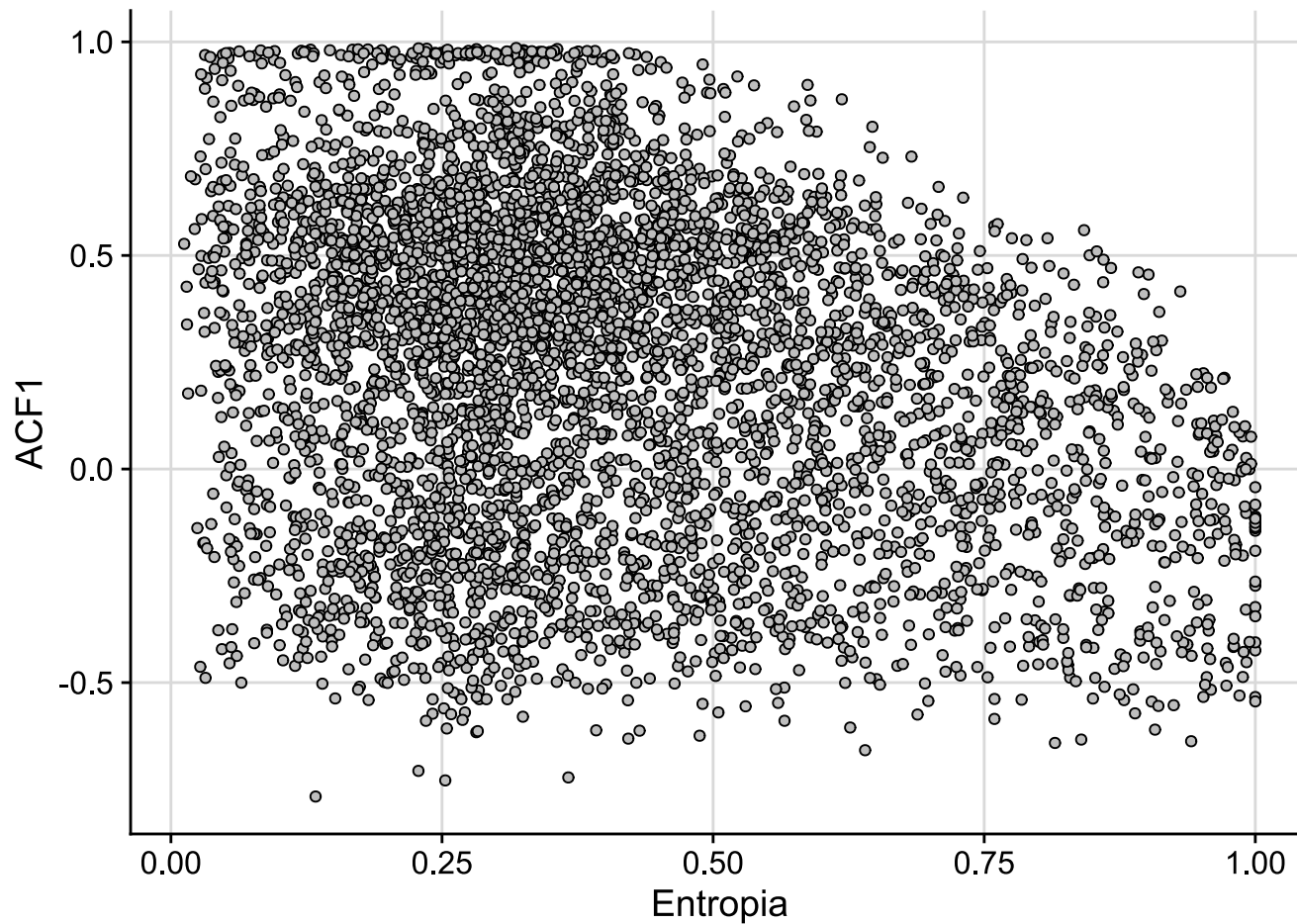
- Baixa entropia



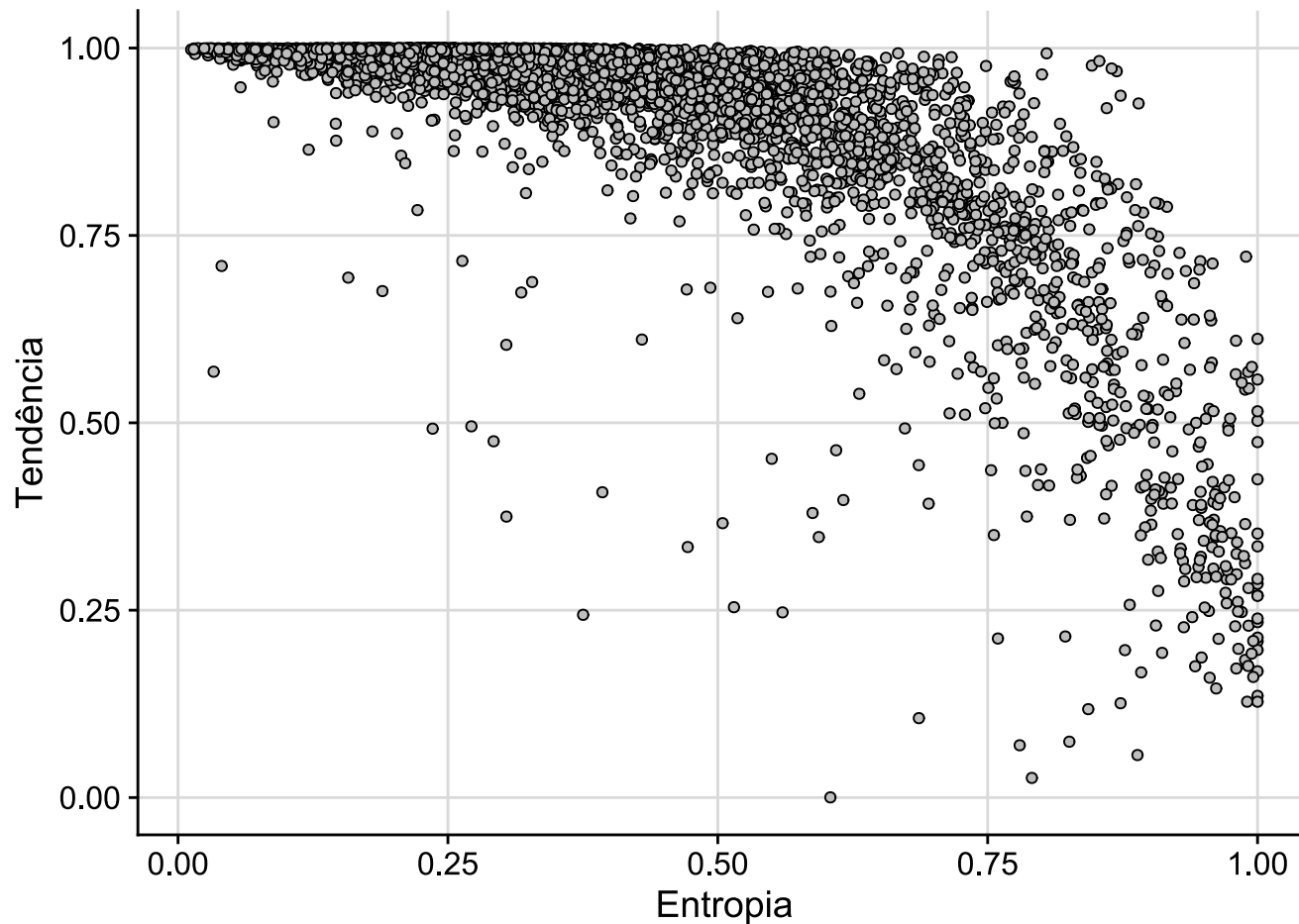
- Alta entropia



Distribuição das *features*



Distribuição das *features*



FFORMA: meta-model

- Seja $\mathbf{f}_n \in \mathcal{F}$ o conjunto de *features* calculadas na base de treino da série n
- Seja $p(\mathbf{f}_n)_m$ o *output* do modelo **XGBoost** em função das *features*.
 - **Importante:** esse *output* é estimado pelo (XGBoost) conforme as árvores de decisão ajustadas sequencialmente.
- Esse *output* é transformado em um **peso** do método m na previsão da série n pela transformação *softmax*:

$$w(\mathbf{f}_n)_m = \frac{\exp[p(\mathbf{f}_n)_m]}{\sum_{m=1}^M \exp[p(\mathbf{f}_n)_m]}$$

- **Obs.:** essa transformação é a função de ligação (natural) de um modelo multinomial.
- Detalhes teóricos sobre XGBoost são encontrados em **Chen and Guestrin (2016)**.

FFORMA: meta-model

- Seja L_{nm} a contribuição no OWA do método m para a série n .
- O XGBoost aproxima de forma additiva a função $p(\mathbf{f}_n)_m$ de tal forma que a função perda média ponderada seja minimizada:

$$\arg \min_w \sum_{n=1}^N \sum_{m=1}^M w(\mathbf{f}_n)_m L_{nm}$$

- Os hiperparâmetros do XGBoost foram estimados via validação cruzada utilizando o procedimento de otimização Bayesiana implementado no pacote `rBayesianOptimization` (Snoek, Larochelle, and Adams, 2012).

FFORMA: previsão

- Portanto, o valor previsto h passos a frente é dado por:

$$\hat{y}(h) = \sum_{m=1}^M \hat{w}_m \hat{y}_m(h)$$

em que

- $\hat{y}_m(h)$ é o valor previsto h passos a frente do método m .
- \hat{w}_m é o peso estimado pelo *meta-model* do método m .

FFORMA: *workflow*

FFORMA: fase *offline*

OFFLINE PHASE: TRAIN THE LEARNING MODEL

Inputs:

$\{x_1, x_2, \dots, x_N\}$: N observed time series that form the reference set.

F : a set of functions for calculating time series features.

M : a set of forecasting methods in the pool, e.g., naïve, ETS, ARIMA, etc.

Output:

FFORMA meta-learner: A function from the extracted features to a set of M weights, one for each forecasting method.

Prepare the meta-data

for $n = 1$ to N : **do**

Split x_n into a training period and test period.

Calculate the set of features $f_n \in F$ over the training period.

Fit each forecasting method $m \in M$ over the training period and generate forecasts over the test period.

Calculate forecast losses L_{nm} over the test period.

end for

Train the meta-learner, w

Train a learning model based on the meta-data and errors, by minimizing:

$$\operatorname{argmin}_w \sum_{n=1}^N \sum_{m=1}^M w(f_n)_m L_{nm}.$$

FFORMA: fase *online*

ONLINE PHASE: FORECAST A NEW TIME SERIES

Input:

FFORMA meta-learner from offline phase.

Output:

Forecast the new time series x_{new} .

for each x_{new} : **do**

Calculate features f_{new} by applying F .

Use the meta-learner to produce $w(f_{new})$, an M -vector of weights.

Compute the individual forecasts of the M forecasting methods in the pool.

Combine the individual forecasts using w to generate final forecasts.

end for

Implementação computacional

Implementação computacional

- Toda implementação computacional foi realizada no software R usando os seguintes pacotes:
 - `forecast`: métodos de previsão.
 - `tsfeatures`: cálculo das *features*.
 - `xgboost`: modelo XGBoost.
 - `rBayesianOptimization`: otimização bayesiana para encontrar hiperparâmetros.
 - `fforma`: implementação da metodologia proposta.
 - `M4metalearning`: códigos apresentados na competição M4.
- Os códigos são abertos e podem ser encontrados no Github do [robjhyndman](#) e [pmontman](#).

Aplicação

Aplicação: competição M4

- Uma amostra de 500 séries foi selecionada aleatoriamente das 100.000 séries da competição M4.
- O modelo FFORMA foi ajustado utilizando o pacote `fforma`:

```
library(fforma)
library(M4comp2018)

# Conjunto de dados com 500 séries escolhidas ao acaso
set.seed(6969)
ts_dataset <- M4[sample(length(M4), 500)]

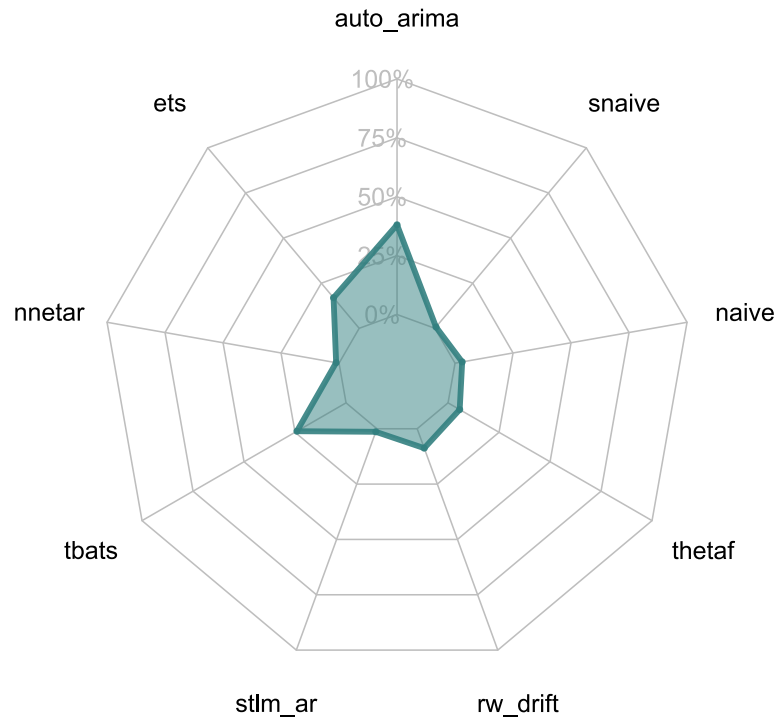
# Treinamento do meta-model
fforma_fit <- train_metalearning(ts_dataset)

# Previsão na base de teste
fforma_forec <- forecast_metalearning(fforma_fit, ts_dataset)
```

- O ajuste do meta-model (`fforma_fit`) demorou em torno de 1h e 50 minutos.
- A previsão na base de teste (`fforma_forec`) demorou em torno de 1h e 58 minutos.

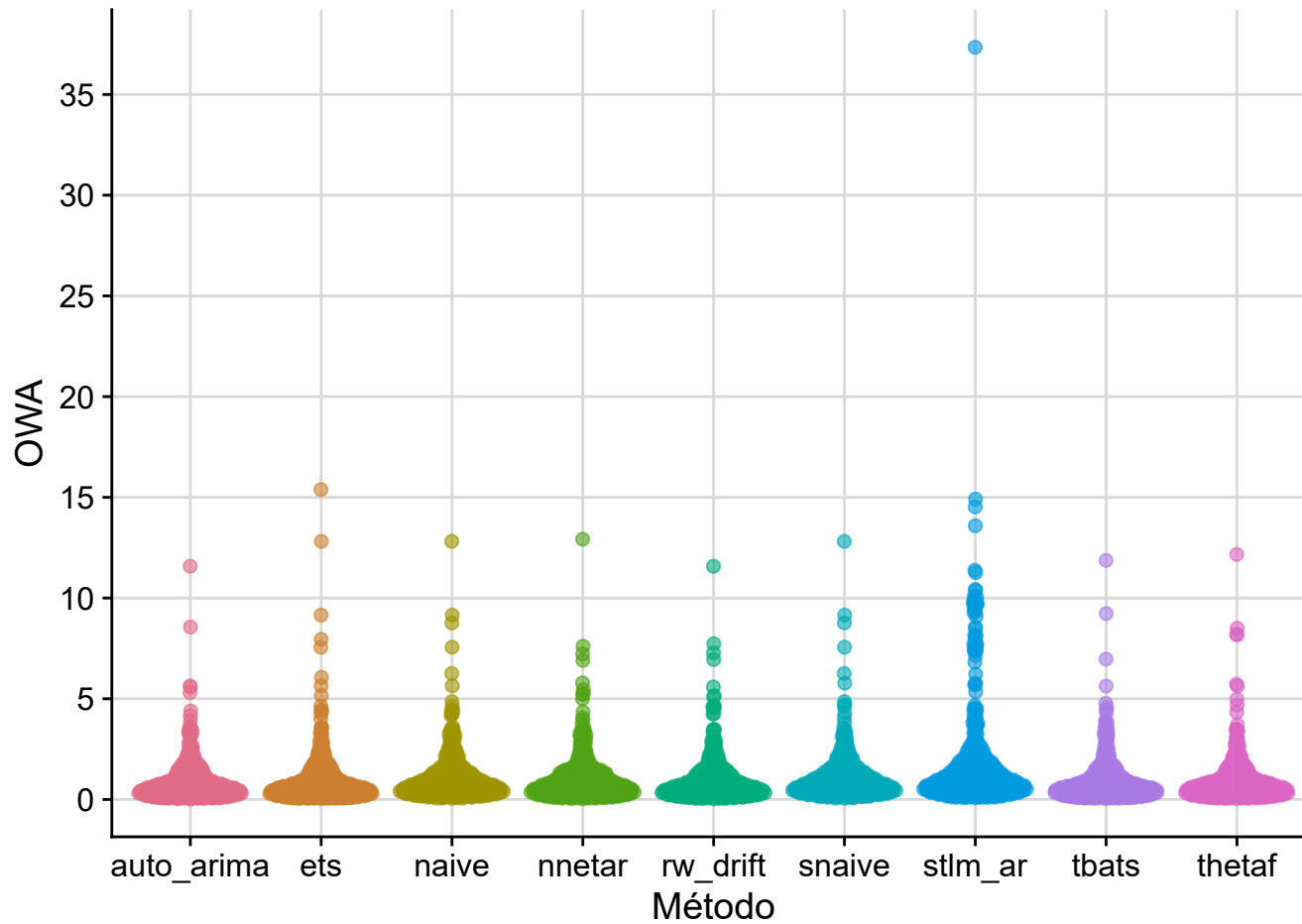
Aplicação

- Peso médio estimado por método.



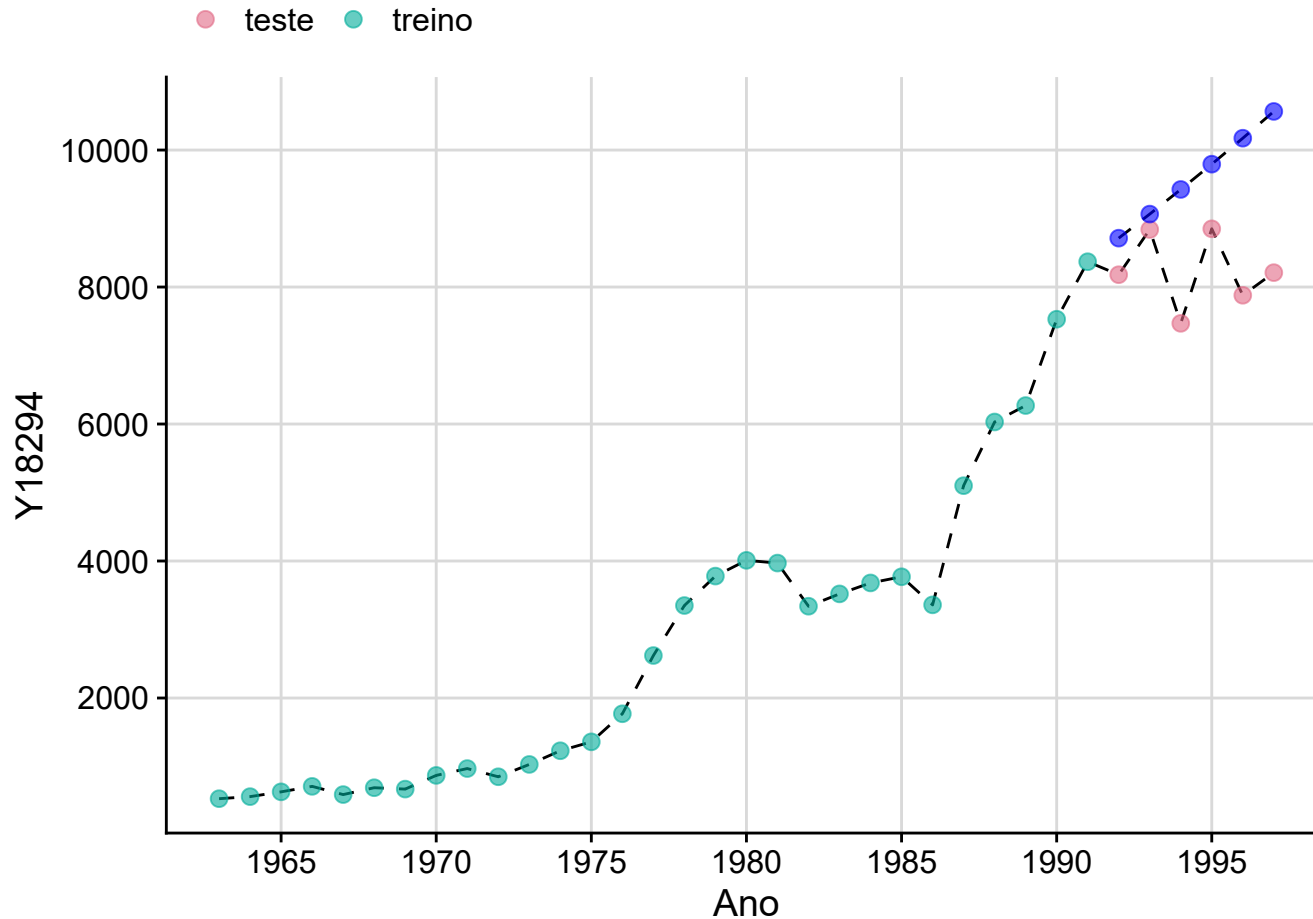
Aplicação

- Distribuição do OWA por método.



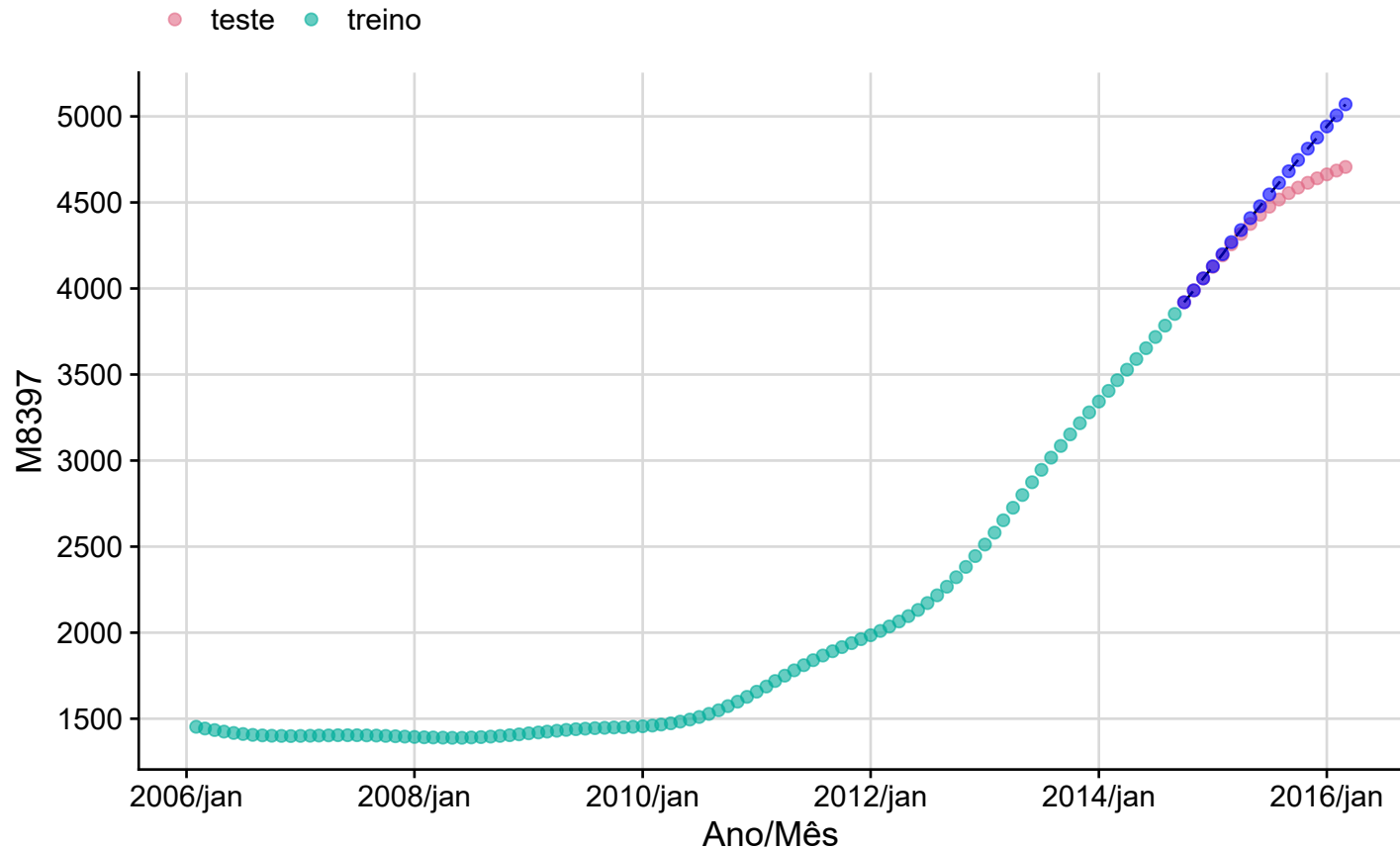
Aplicação: Exemplos de previsão

- Série Y18294 **anual** com 29 observações para prever 6 anos a frente.



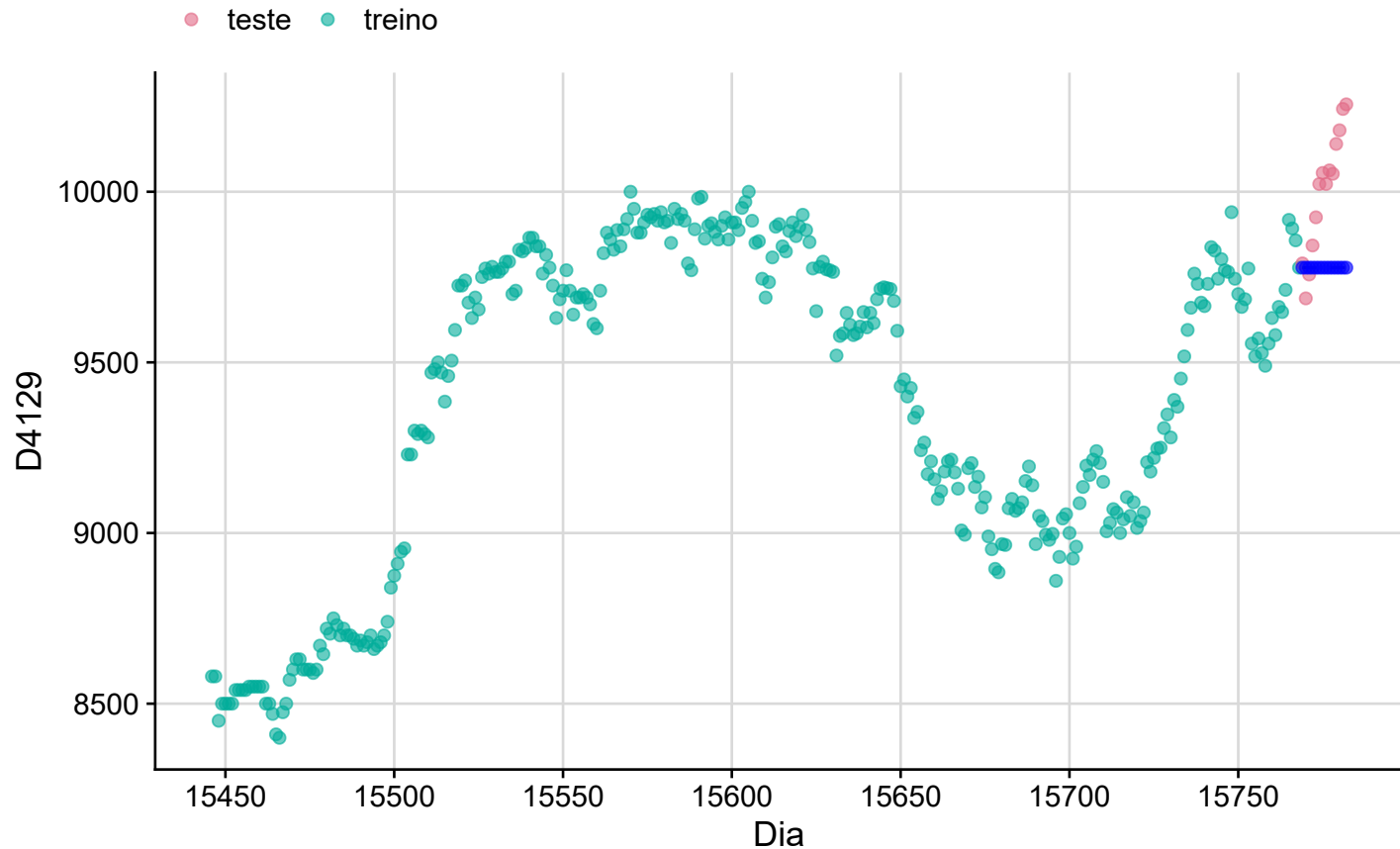
Aplicação: Exemplos de previsão

- Série M8397 **mensal** com 104 observações para prever 18 meses a frente.



Aplicação: Exemplos de previsão

- Série D4129 **diária** com 323 observações para prever 14 dias a frente.



Aplicação: Considerações gerais

- O OWA geral foi de 0.8528.
- O modelo auto ARIMA teve maior peso médio na previsão das 500 séries.
- Embora não seja o objetivo principal das competições M é importante inspecionar as previsões para cada série graficamente, pois permite maior entendimento sobre o erro do modelo.
- O custo computacional foi muito alto.
 - Inviabiliza implementar este modelo num servidor para deixar rodando online.
- É necessário melhor implementação computacional do f forma uma vez que:
 - Código pouco otimizado: Segundo os autores o modelo demorou 5 dias para realizar o ajuste nas 100.000 séries da competição M4.
 - Código pouco estruturado: Repetições de conta e funções não foram escritas em classes.

Referências

- Assimakopoulos, V. and K. Nikolopoulos** (2000). "The theta model: a decomposition approach to forecasting". In: *International Journal of Forecasting* 16, pp. 521-530.
- Bates, J. M. and C. W. J. Granger** (1969). "The Combination of Forecasts". In: *Journal of the Operational Research Society* 20.4, pp. 451-468.
- Chen, T. and C. Guestrin** (2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Clemen, R.** (1989). "Combining forecasts: a review and annotated bibliography with discussion". In: *International Journal of Forecasting* 5, pp. 559-608.
- De Livera, A, R. Hyndman, and R. D. Snyder** (2011). "Forecasting time series with complex seasonal patterns using exponential smoothing,". In: *Journal of the American Statistical Association* 106.496, pp. 1513-1527.
- Hyndman, R. J. and A. B. Koehler** (2006). "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4, pp. 679-688.

Referências

Hyndman, R. J., E. Wang, and N. Laptev (2015). "Large-Scale Unusual Time Series Detection". In: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. , pp. 1616-1619.

Hyndman, R. and G. Athanasopoulos (2018). *Forecasting: principles and practice*. 2nd. OTexts: Melbourne.

Hyndman, R. and B. Billah (2003). "Unmasking the Theta method". In: *International Journal of Forecasting* 19, pp. 287-290.

Hyndman, R. and Y. Khandakar (2008). "Automatic Time Series Forecasting: The forecast Package for R". In: *Journal of Statistical Software, Articles* 27.3, pp. 1-22.

Kang, Y, R. J. Hyndman, and K. Smith-Miles (2017). "Visualising forecasting algorithm performance using time series instance spaces". In: *International Journal of Forecasting* 33.2, pp. 345-358.

Montero-Manso, P, G. Athanasopoulos, R. J. Hyndman, et al. (2019). "FFORMA: Feature-based forecast model averaging". In: *International Journal of Forecasting* 36.1, pp. 86-92.

Referências

Smith, J. and K. F. Wallis (2009). "A simple explanation of the forecast combination puzzle". In: *Oxford Bulletin of Economics and Statistics* 71.3, pp. 331-355.

Snoek, J., H. Larochelle, and R. P. Adams (2012). *Practical Bayesian Optimization of Machine Learning Algorithms*.

Talagala, T. S., R. J. Hyndman, and G. Athanasopoulos (2018). "Meta-learning how to forecast time series". Working Paper 6/18. Department of Econometrics & Business Statistics, Monash University..

Timmermann, A. (2006). "Forecast Combinations". In: *Handbook of economic forecasting*. Ed. by G. Elliot, C. W. J. Granger and A. Timmermann. Amsterdam, North-Holland. Chap. 4, pp. 135-196.

Obrigado!