

# Telescope: A Hybrid Forecast Method for Univariate Time Series

## Abstract Paper

Marwin Züfle, André Bauer, Nikolas Herbst, and Samuel Kounev  
Julius-Maximilians-University Würzburg

Valentin Curtef  
MaxCon Data Science GmbH, Würzburg

**Abstract.** Forecasting is an important part of the decision-making process and used in many fields like business, economics, finance, science, and engineering. According to the “*No-Free-Lunch-Theorem*” from 1997, there is no general forecasting method, that performs best for all time series. Instead, expert knowledge is needed to decide which forecasting method to choose for a specific time series with its own characteristics. Since a trial and error approach is very inefficient and expert knowledge is useful but a time-consuming task that cannot be fully automated, we present a new hybrid multi-step-ahead forecasting approach based on time series decomposition. Initial evaluations show that this hybrid approach improves the forecast accuracy compared to six existing forecasting methods while maintaining a short runtime.

## 1 Introduction

Forecasting allows to predict the future by examining past observations. Classical forecasting methods have their benefits and drawbacks depending on the specific use cases. Thus, there is no globally best forecasting technique [19] and further respectively expert knowledge is required for determining the best forecast method. Typically, expert knowledge is needed for two domains, i.e., method selection and feature engineering. The serious problem of expert knowledge is that it can take a long time to deliver results and it cannot be completely automated. In the field of feature engineering, expert knowledge can be replaced by using deep learning [16,12] or random forests [8,2]. To overcome the need of expert knowledge in method selection, a more robust forecasting method compared to the classical forecaster is needed. In this field, robust means that the variance in forecasting results should be reduced, not necessarily improving the forecasting accuracy itself. By reducing the variance of the results, the risk when trusting a bad forecast is lowered. Hybrid forecasting is such a technique since the benefits of multiple forecasting methods can be combined to improve the overall performance. Thus, we introduce a new hybrid, multi-step-ahead forecasting approach for univariate time series. The approach is based on time series decomposition and makes use of existing forecasting methods, i.e., ARIMA, ANN, and XGBoost.

## 2 Related Work

In 1997, Wolpert and Macready presented the “*No Free Lunch Theorem*” for optimization algorithms [19]. It claims, that there is not a single algorithm that performs best for all scenarios since improving the performance of one aspect normally leads to a degradation in performance for some other aspect. This theorem can also be applied to forecasting methods as there is no single method that outperforms the others for all types of data. To address this issue, many hybrid forecasting approaches have been developed. A hybrid forecasting method makes use of at least two forecasting methods to compensate the limitations of individual forecasting methods. Hybrid methods can be categorized into three groups of approaches each sharing the same basic concept.

The first and historically oldest group is the concept of ensemble forecasting and is the technique of using at least two forecasting methods. While assigning a weight to each method, the forecast result is the weighted sum of each forecast method. This approach was introduced by Bates and Granger in 1969 [1]. The concept of this approach is rather simple, however, the assignment of weights is a crucial part. Thus, many methods for weight estimation have been investigated [4,7,11].

The second group of forecasting methods is based on the concept of forecast recommendation, where the goal is to build a rule set to guess the assumed best forecasting method based on certain time series features. There are two common ways to generate the rule set. One method is using an expert system. Collopy and Armstrong used this approach to create a rule set by hand in 1992 [6]. The other method is using machine learning techniques to automatically generate a rule set. In 2009, Wang et al. proposed clustering and algorithms for rule set learning based on a large variety of time series features [18].

The third group of forecasting methods is based on decomposition of the time series with the goal to leverage the advantages of each method to compensate for the drawbacks of the others. In literature, there are common approaches. The first approach is to apply a single forecasting method on the time series and then apply another method on the residuals of the first one [20,17]. The second forecasting method is intentionally chosen to have different characteristics than the first one, that is, it should have antithetic advantages and drawbacks compared to the first one. An alternative approach is to split the time series into its components trend, seasonality, and noise, applying a different forecasting method on each of them. Liu et al. introduced an approach like this targeted for short-term load forecasting of micro-grids [13]. They used empirical mode decomposition to split the time series and extended Kalman filter, extreme learning machine with kernel, and particle swarm optimization for the forecast.

The hybrid forecasting approach, we propose in this paper, focuses on forecasts based on decomposition. In contrast to Zhang or Pain and Lin [20,17], we use explicit time series decomposition, that is, forecast methods are applied to the individual components of the time series as opposed to the residuals of previous forecasts. Liu et al. introduced a short-term hybrid forecasting method based on an intrinsic mode function. In our approach, the time series is split into trend, seasonality, and noise

components. Additionally, completely different forecasting methods are used as a basis. Furthermore, the hybrid approach of this paper is designed to perform multi-step-ahead forecasting with low overhead and short runtime.

### 3 Telescope Approach

We call the proposed hybrid forecasting approach *Telescope* according to the analogy with the vision on far-distanced objects. Telescope is developed in *R* to perform multi-step-ahead forecasting while maintaining a short runtime. To this end, only fast and efficient single forecasting methods are used as components of Telescope. A diagram of the forecasting procedure is shown in Fig. 1.

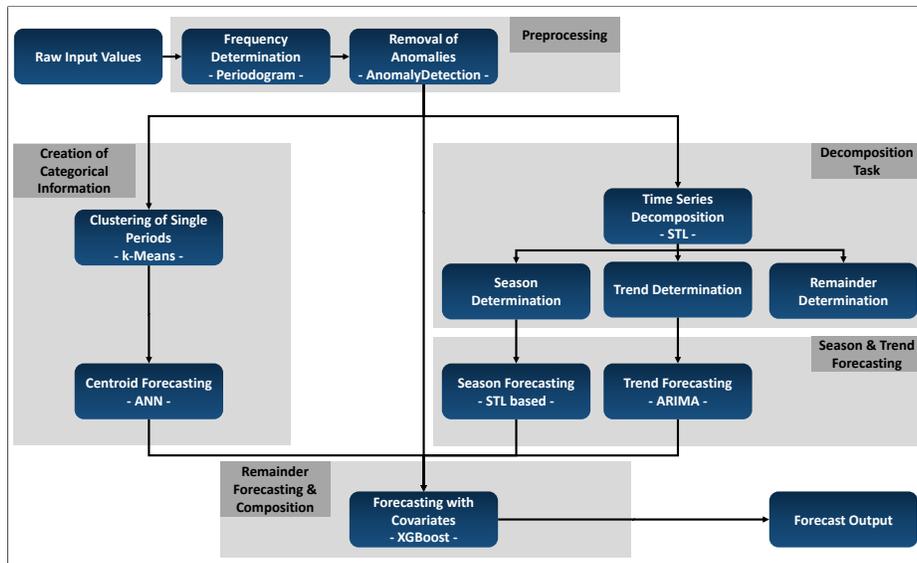


Fig. 1: A simplified illustration of the Telescope approach.

First, a preprocessing step is executed. The frequency of the time series is estimated using periodograms, i.e., applying the *R* function `spec.pgram`. This function uses fast Fourier transformation to estimate the spectral density. The estimated frequency is needed to remove anomalies in the time series by applying the *Anomaly Detection R* package [9]. This package uses a modified version of the seasonal and trend decomposition using Loess (STL) [5]. The STL decomposition splits the time series into the three components season, trend, and remainder. After the decomposition, *Anomaly Detection* applies generalized extreme studentized deviate test (ESD) with median instead of mean and median absolute deviation instead of standard deviation on the remainder to identify outliers. Furthermore, we use STL for an additive decomposition of the revised time series. If the amplitude of the seasonal

pattern increases as the trend increases and vice versa, we assume multiplicative decomposition. Thus, a heuristic testing for such a behavior is implemented. If a multiplicative decomposition is detected, the logarithmized time series is used for the STL decomposition and the components are back-transformed after the decomposition. We apply the STL package because of its short runtime compared to other *R* decomposition functions like `bfast`. Afterwards, the season and trend forecasting is executed. The seasonality determined by STL is simply continued, whereas the trend is forecast using the `auto.arima` function from the `forecast` *R* package by Hyndman [10]. Since there is no seasonal pattern left, seasonality is disabled in `auto.arima` for this purpose. Moreover, this seasonality disabling decreases the runtime of the algorithm. Additionally, the time series with removed anomalies is used to create categorical information. For this purpose, the time series is cut into single periods. Then, the single periods are clustered into two classes using the `kmeans` *R* function. Each class is represented by its centroid. Next, this history of centroids is forecast using an artificial neural network (ANN), i.e., the `nnetar` function of the `forecast` *R* package [10]. If a specific time series is forecast several times, this clustering task does not need to be performed every time. Finally, the last step is the remainder forecast and composition. XGBoost is used [3], which is an implementation of gradient boosted decision trees and it works best when obtaining covariates. XGBoost is applied using the trend, seasonality, and centroid forecasts as covariates and the raw time series history as labels. In addition, 10% of the history data are used for validation in the training phase to prevent XGBoost from overfitting. The sources of the Telescope approach are currently under publication as open-source<sup>1</sup>.

## 4 Preliminary Evaluation

To evaluate the performance of this hybrid forecasting approach, we have conducted some initial experiments presented in this section. As example time series, a trace of completed transactions on an IBM z196 Mainframe during February 2011 and a trace of monthly international airline passengers from 1949 to 1960 are used.

Each observation in the IBM trace contains the quarter-hourly amount of transactions (e.g., bookings or money transfers). The trace contains 2670 observations and is depicted in Fig. 2a. It shows a typical seasonal pattern with a daily and weekly cycle, where the amount of transactions differs completely during weekdays and weekends. The trace exists of about 28 daily periods and 4 weekly periods. Since the approach is designed to perform multi-step-ahead forecasts, the last 20% observations of the time series are chosen as forecast horizon. Thus, the history of the IBM trace incorporates 2136 observations and the forecast horizon is set to 534 observations. The border between history and horizon is shown as vertical purple line in Fig. 2a. The forecast of the IBM trace is shown in Fig. 2b. The original time series is depicted in black, whereas the forecast of Telescope is represented by the red line. As a reference, the second best forecast produced by the tBATS approach [14] is shown as

---

<sup>1</sup> Telescope: <http://descartes.tools/telescope>

dashed blue line. Besides the good fitting of the observed history, the hybrid approach succeeds in capturing the weekdays and weekends. In contrast to Telescope, tBATS only repeats a single pattern for the whole horizon. For weekdays, the forecast of Telescope and tBATS are very close to each other. However, tBATS misses capturing the weekends.

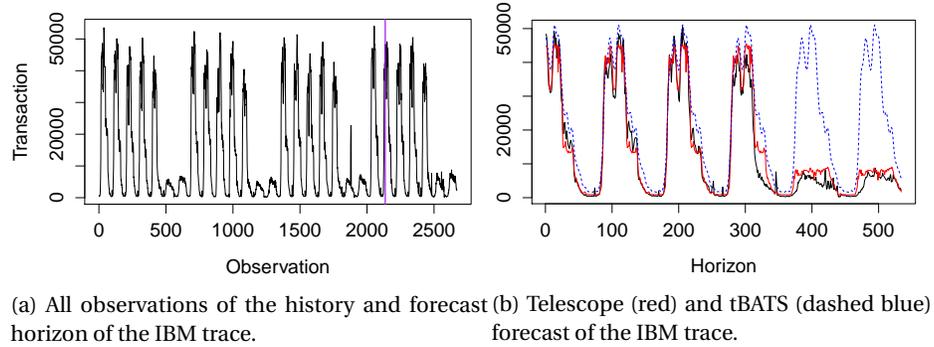


Fig. 2: Observations and forecast of the IBM trace.

The airline passengers trace consists of 144 observations and shows an exponential trend pattern as well as a seasonal pattern with yearly cycle. Furthermore, the amplitude of the seasonal pattern increases as the trend rises. The airline passengers trace is shown in Fig. 3a. Since the forecast horizon is set to 20%, the history contains 115 observations and the forecast horizon consists of 29 observations. Again, the border between history and horizon is shown as vertical purple line in Fig. 3a.

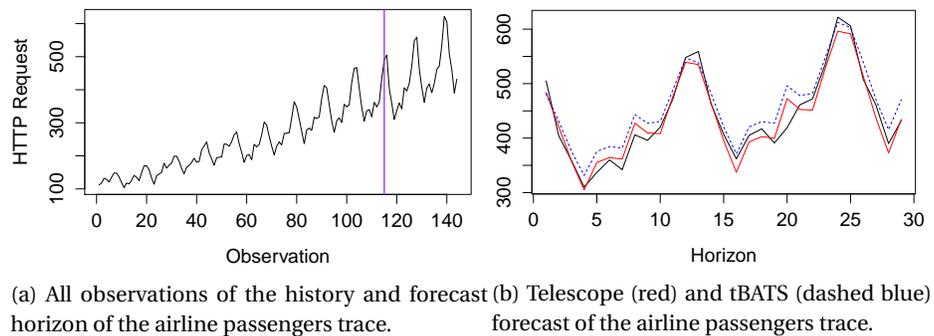


Fig. 3: Observations and forecast of the airline passengers trace.

Fig. 3b shows the forecast of the airline passengers trace. Again, the original time series is depicted as black line, the forecast of Telescope is shown as red line, and the tBATS forecast is depicted as dashed blue line. Both, tBATS and Telescope succeed in capturing the trend and season pattern. Though, besides the first value of the horizon, the tBATS forecast is always greater than the Telescope forecast.

To evaluate the forecasting accuracy in a mathematical way the mean absolute percentage error (MAPE) and mean absolute scaled error (MASE) measures are used. The MAPE is a widely-used measure to assess forecasting accuracy based on the forecasting error normalized with the observations. However, MAPE has some serious limitations, i.e., it cannot be used for time series with zeros in the forecasting horizon and it punishes positive errors harder than negative errors. Thus, we additionally use a second measure called MASE. Both measures are independent of the data scale but in contrast to MAPE, MASE is suitable for almost all situations and the error is based on the in-sample MAE from the random walk forecast. For a 20% forecast, the random walk forecast would forecast the last value of the history for the entire horizon. Thus, the investigated forecast is better than the random walk forecast if the MASE value is less than 1 and worse if the MASE value is greater than 1. The MAPE and MASE values can be calculated as follows:

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{Y_i} \right| \quad (1)$$

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |e_i|}{\frac{1}{n-1} \times \sum_{i=2}^n |Y_i - Y_l|} \quad (2)$$

Where  $Y_l$  is the observation at time  $l$  with  $l$  being the index of the last observation of the history.  $Y_i$  is the observation at time  $l+i$ . Thus,  $Y_i$  is the value of the  $i$ -th observation in the forecast horizon. The forecast error at time  $l+i$  is calculated as  $e_i = Y_i - F_i$  where  $F_i$  is the forecast at time  $l+i$ . The amount of observations in the forecast horizon is represented by  $n$ . Another important measure to evaluate the performance of the forecasting approach is the elapsed time for the forecasting process. The total time elapsed for the forecast is measured in seconds. Table 1 shows the MASE and MAPE values and runtime of the hybrid approach for the IBM and airline passengers traces compared to six state-of-the-art forecasting methods:

- **ARIMA**: auto-regressive integrated moving averages (`auto.arima` with seasonality in package `forecast` [10]),
- **ANN**: artificial neural nets (`nnetar` in package `forecast` [10]),
- **ETS**: extended exponential smoothing (`ets` in package `forecast` [10]),
- **tBATS**: trigonometric, Box-Cox transformed, ARMA errors using trend and seasonal components (`tbats` in package `forecast` [10,14]),
- **SVM**: support vector machine (`svm` in package `e1071` [15]),
- **XGBoost**: scalable tree boosting (`xgboost` in package `xgboost` [3]) using only the index of the observation as covariate.

On the one hand, the experiment shows that the hybrid approach reaches the lowest MASE and MAPE values for both time series. The Telescope forecast reaches a MASE value of about 0.064 for the IBM trace and 0.179 for the airline passengers trace. The MAPE values are about 51.628% and 3.382%. The second best MASE values are achieved by tBATS with about 0.191 for the IBM trace and 0.276 for the airline passengers trace. Furthermore, tBATS reaches the second best MAPE value for the airline passengers trace with about 5.472%. However, ANN outperforms tBATS in matters of the MAPE value for the IBM trace, i.e., ANN achieves a MAPE value of about 179.537. On the other hand, the experiment shows that Telescope has a very short runtime with about 8.557 and 2.679 seconds. On the IBM trace, ETS, SVM, and XGBoost itself achieve shorter runtimes compared to Telescope. Though, each of these forecasting methods delivers a bad accuracy. On the airline passengers trace, only tBATS has a longer runtime than Telescope. Since the IBM trace is about 15 times as long as the airline passengers trace, this experiment implies that the runtime of the Telescope approach does not depend as much on the time series length as some state-of-the-art forecasting methods, i.e., ARIMA, ANN, and tBATS, do.

Table 1: Accuracy and runtime of state-of-the-art forecasting methods and Telescope.

Forecasting Method	IBM Trace			Passengers Trace		
	MASE	MAPE [%]	Time [s]	MASE	MAPE [%]	Time [s]
Telescope	0.064	51.628	8.557	0.179	3.382	2.679
ARIMA	0.343	813.570	12.301	0.358	6.255	1.065
ANN	0.788	179.537	12.172	0.400	7.473	0.801
ETS	0.986	2992.701	0.531	0.358	6.361	2.371
tBATS	0.191	253.243	38.078	0.276	5.472	4.538
SVM	0.276	574.624	2.312	3.711	67.909	0.233
XGBoost	0.736	545.469	0.484	0.692	11.936	0.278

## 5 Conclusion

In this paper, Telescope, a new hybrid approach for multi-step-ahead forecasting, is introduced. The approach uses clustering for creating categorical information like weekdays and weekends. STL decomposition is used to split the time series into trend, seasonality, and noise. ARIMA without seasonality is used to forecast the trend with low overhead. The seasonal pattern delivered by STL is simply continued. Finally, XGBoost is applied using the trend, season, and cluster forecasts as covariates. Initial evaluations show that the approach achieves good accuracy while maintaining short runtime. As future work, we plan to perform more evaluations and include several extensions to the algorithm like denoising based on wavelet transformations and identification of break points in the trends.

## References

1. Bates, J.M., Granger, C.W.: The combination of forecasts. *Journal of the Operational Research Society* 20(4), 451–468 (1969)
2. Cano, G., Garcia-Rodriguez, J., Garcia-Garcia, A., Perez-Sanchez, H., Benediktsson, J.A., Thapa, A., Barr, A.: Automatic selection of molecular descriptors using random forest: Application to drug discovery. *Expert Systems with Applications* 72, 151–159 (2017)
3. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: *ACM SIGKDD 2016*. pp. 785–794. ACM (2016)
4. Clemen, R.T.: Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5(4), 559–583 (1989)
5. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* 6(1), 3–73 (1990)
6. Collopy, F., Armstrong, J.S.: Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science* 38(10), 1394–1414 (1992)
7. De Menezes, L.M., Bunn, D.W., Taylor, J.W.: Review of guidelines for the use of combined forecasts. *European Journal of Operational Research* 120(1), 190–204 (2000)
8. El Haouij, N., Poggi, J.M., Ghozi, R., more: Random forest-based approach for physiological functional variable selection: Towards driver's stress level classification (2017)
9. Hoehenbaum, J., Vallis, O.S., Kejariwal, A.: Automatic anomaly detection in the cloud via statistical learning. *arXiv preprint arXiv:1704.07706* (2017)
10. Hyndman, R.J., Khandakar, Y., et al.: Automatic time series for forecasting: the forecast package for r. Tech. rep., Monash University (2007)
11. Krishnamurti, T.N., Kishitawal, C., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., Surendran, S.: Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate* 13(23), 4196–4216 (2000)
12. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *nature* 521(7553), 436 (2015)
13. Liu, N., Tang, Q., Zhang, J., more: A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Applied Energy* 129, 336–345 (2014)
14. Livera, A.M.D., Hyndman, R.J., Snyder, R.D.: Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association* 106(496), 1513–1527 (2011)
15. Meyer, D.: Support vector machines (2017), <https://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf>
16. Ngiam, J., Khosla, A., Kim, M., Nam, J., more: Multimodal deep learning. In: *ICML 2011*. pp. 689–696 (2011)
17. Pai, P.E., Lin, C.S.: A hybrid arima and support vector machines model in stock price forecasting. *Omega* 33(6), 497–505 (2005)
18. Wang, X., Smith-Miles, K., Hyndman, R.: Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series. *Neurocomputing* 72(10 - 12), 2581 – 2594 (2009)
19. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1(1), 67–82 (Apr 1997)
20. Zhang, G.P.: Time series forecasting using a hybrid arima and neural network model. *Neurocomputing* 50, 159–175 (2003)