Forecasting Network Exchange Time Series

Aleksandr Moshnikov[a], Aleksandr Syrov[b]

[a]*ITMO University, 49 Kronverksky Pr., St. Petersburg, Russian Federation*
[b]*ITMO University, 49 Kronverksky Pr., St. Petersburg, Russian Federation*

Abstract

This article deals with the problem of network traffic forecasting using the time series tool. A special feature of the proposed task is to consider the traffic of an individual network user. We consider 5 types of user traffic characterized by different parameters of the volume and number of transmitted packets. 13 main models for traffic forecasting are considered. To consider the effectiveness the parameters - RMSE, p-value, MAPE were evaluated. Leung-Box test is used to model assessment. To solve the problem the R software, the forecast package, is used. The results of an experiment using 13 different models are considered.

Keywords

Network traffic forecasting, Network modelling, R programming language, Time series

1. Introduction

Modeling network traffic is a notoriously difficult task. This is primarily due to the ever-increasing complexity of network traffic and the various ways in which the network can be excited by user activity. The ongoing development of new network applications, protocols, and usage profiles further necessitates models that can adapt to the specific networks in which they are deployed [1] demand for telecommunications network traffic continues to grow exponentially worldwide. Research has shown that the number of mobile cellular subscribers is growing worldwide, and the world's population is gradually equalizing with the level of its use [2].

Achieving this exponential growth requires effective planning and rapid expansion of telecommunications systems, as well as the introduction of modern equipment. One approach to the leadership industry players is the development and adoption of appropriate forecasting models for the implementation of this agenda. Forecasting methods can be classified as long-term and short-term. According to [3] the forecasting process based on a time interval in weeks, months, and years is a long-term forecast, while short-term forecasts are milliseconds, seconds, minutes, hours, and days. Time series modeling and forecasting are widely used for analyzing telecommunications network traffic [4]. It has been shown that ARIMA models are stable forecast for BitTorrent traffic [5]. In contrast, [6] indicated that the accuracy of prediction of ARIMA models has a limited time interval. Data sets of http network traffic grouped in different periods of the day [7] and activities [8].

Proceedings of the 12th Majorov International Conference on Software Engineering and Computer Systems, December 10–11, 2020, Online & Saint Petersburg, Russia
moshnikov.alex@gmail.com (A. Moshnikov)
0000-0002-3689-2472 (A. Moshnikov); 0000-0003-1475-4105 (A. Syrov)

In addition to the traffic itself, the task of analyzing the reliability of a network structure that provides information exchange is also important [9], as well as aspects related to the computational reliability of operations performed [10].

2. Problem statement

The problem of network traffic forecasting using the time series tool is considered. A special feature of the proposed task is to consider the traffic of a single network user. As part of the task under consideration, the following tasks were performed:

- typical traffic sections (time series) characteristic of a particular network user behavior model were formed;
- an overview of existing time series forecasting models was performed;
- a time series forecasting method has been selected that allows you to get acceptable forecast results for various models of network user behavior.

3. Methodology

3.1. Description of initial data

To generate time series, raw sockets are used to capture and analyze frames sent or received by the node in question. Time series are formed by measuring (at intervals of one second) the amount of data coming to the node under consideration over the TCP/IP Protocol stack.

Information about frames is saved in a table (the format of such a table is shown in the Fig. 1). This table is then converted to a time series data.

Typical traffic sections (time series) that are typical for various network user behavior models The following models of network user behavior are considered (based on personal experience):

- click on links (time series R1);
- download files (time series R2);
- listening to music (time series R3);
- view video (time series R4);
- the user's browser is running but not in use (time series R5).

The results are shown in Fig. 2. and Fig. 3.

3.2. Overview of existing time series forecasting models

This review is based on the following works [11] and [12]. The following designations are used: y_T - the observed (actual) value of the series at time T, $\hat{y}_{T+h|T}$ - the predicted value of the time series for time $T + h$.

Summary of time series forecasting models and fucntion parametres are presented in Table 2.

All calculations related to forecasting will be performed in the R software environment. R is a programming language for statistical data processing and graphics, as well as a free and open

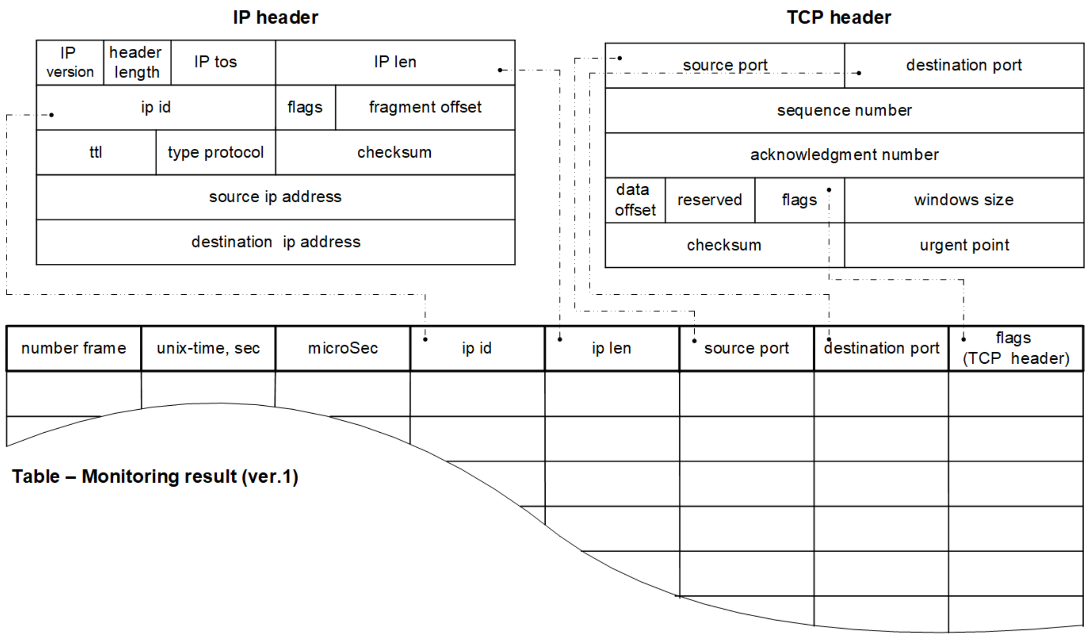

Figure 1: Network monitoring data

Table 1
Series forecasting models

Model name	Parameters	Function forecast library
Average	-	meanf(ts)
Last value	-	naive(ts)
Drift model	-	rwf(ts,drift=TRUE)
Simple exponential smoothing model	α	ses(ts)
The Holt Model	α, β	holt(ts)
Holt model with a fading trend	α, β, ϕ	holt(ts,damped=TRUE)
Holt-Winters model, with additive seasonality	α, β, γ	hw(ts,seasonal= additive)
Holt-Winters model, with multiplicative seasonality and fading trend	α, β, γ	hw(ts,seasonal= multiplicative)
Holt-winters model, with multiplicative seasonality and fading trend	$\alpha, \beta, \gamma, \phi$	hw(ts,seasonal= multiplicative, damped=TRUE)
An ARIMA model	-	
Forecasting the components of a series		stlf(ts, method)
Autoregression based on neural networks		nnetar(ts)

source computing environment within the GNU project. The R language contains tools that allow you to create several parallel threads of calculations (by simultaneously loading several processor cores) and reduce the time spent on modeling several times [13]. Series forecasting models provided in the forecast package are considered.

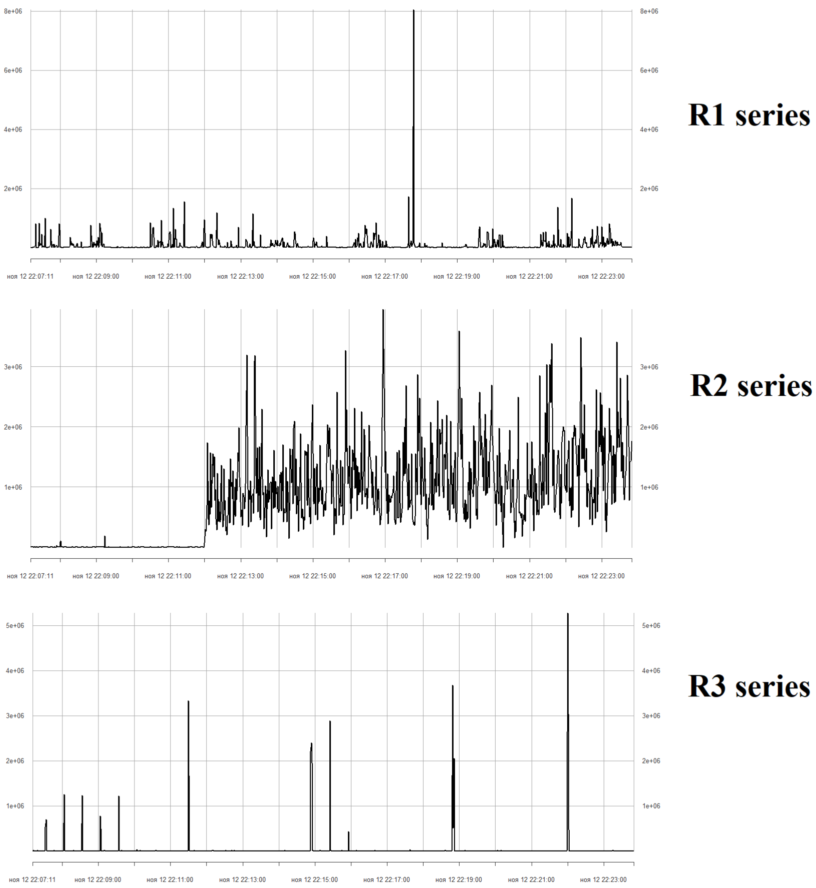

Figure 2: Timing diagram of the traffic R1, R2, R3 series

4. Estimation of forecast accuracy

4.1. Evaluation of prediction accuracy

Forecast error refers to the difference between the observed value and its forecast:

$$e_{T+h} = \hat{y}_{T+h} - \hat{y}_{T+h|T} \tag{1}$$

The root-mean-square error is used to estimate the prediction accuracy:

$$RMSE = \sqrt{mean(e_t^2)} \tag{2}$$

4.1.1. Schematic diagram of estimation of forecasting accuracy.

The considered time series is divided into training and test parts. The parameters of the forecasting model are determined by the training part. After determining the model parameters,

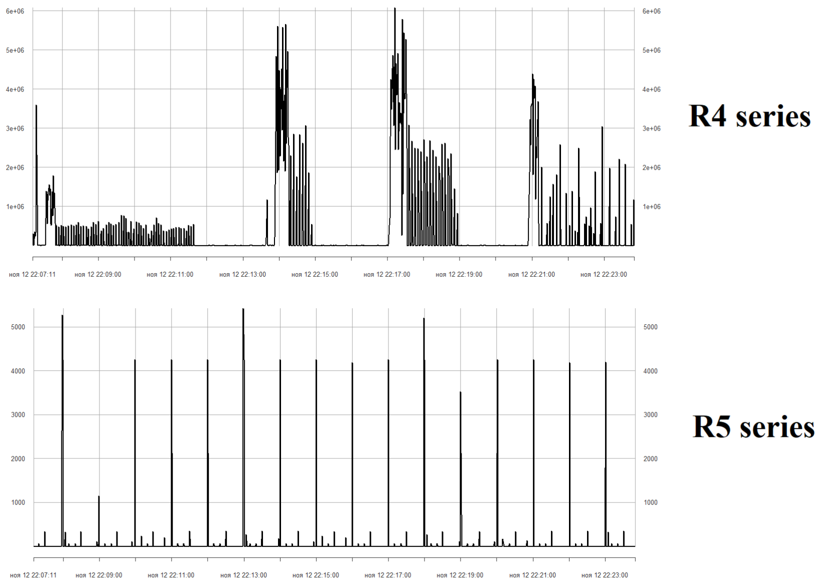

Figure 3: Timing diagram of the traffic R4, R5 series

the Leung-Box test is performed for the absence of auto-correlation in the model residuals. If the Leung-Box test is not passed (there is a strong auto-correlation in the remainder of the model), the model is rejected. If the test is passed, the accuracy of the forecast (RMSE value) is determined by the test part of the series. Algorithm presented at Fig. 4.

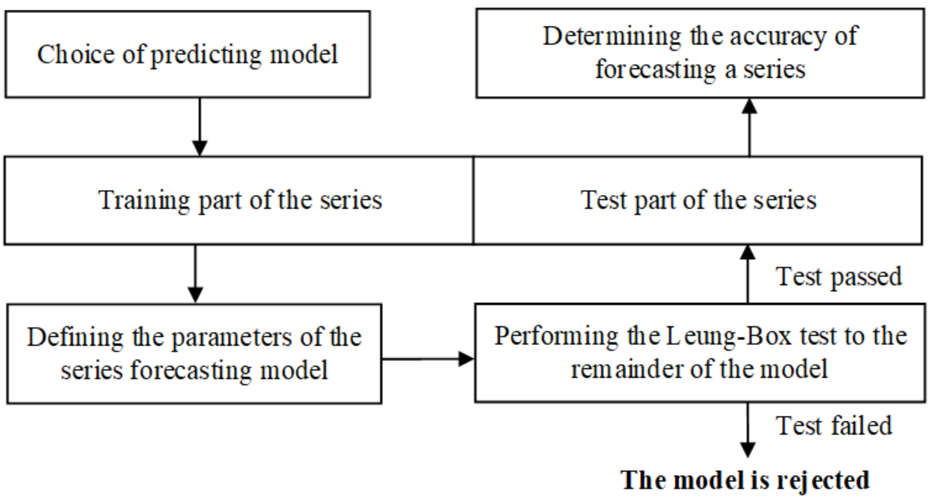

Figure 4: Schematic diagram of estimation of forecasting accuracy

4.1.2. Combination of forecasting models.

As an additional study, we consider the possibility of predicting time series based on combinations (combining) of several forecasting models. For a combination of forecasting models, the total forecast error is calculated using:

$$e_{T+h} = (1/N) \cdot \sum_{i=1}^{n} e_{T+h}^{(i)} \tag{3}$$

where N - number of prediction models in combination; $e_{T+h}^{(i)}$ prediction error of the i-th forecasting model.

5. The results of evaluation of prediction accuracy for different models

Results of estimating the accuracy of forecasting the R1 traffic type (as example): original series and the function of partial auto-correlation are shown in Fig. 5; and the residual plot and fitted model are shown in Fig. 6.

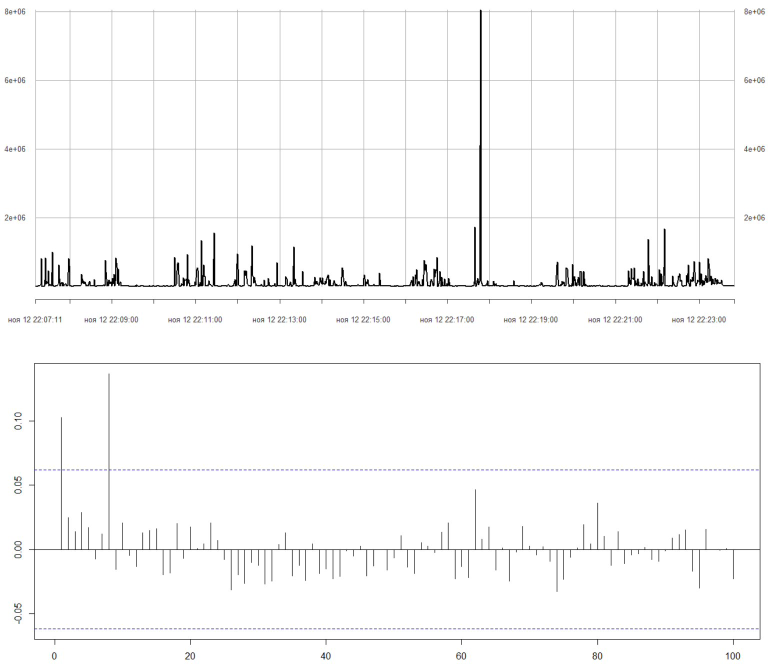

Figure 5: Results of estimating the accuracy of forecasting: original series, the function of partial auto-correlation

For a number of models, it was not possible to calculate the RMSE, MAPE, and p-value indicators this is due to the fact that the sample size exceeded the parameters of the input file

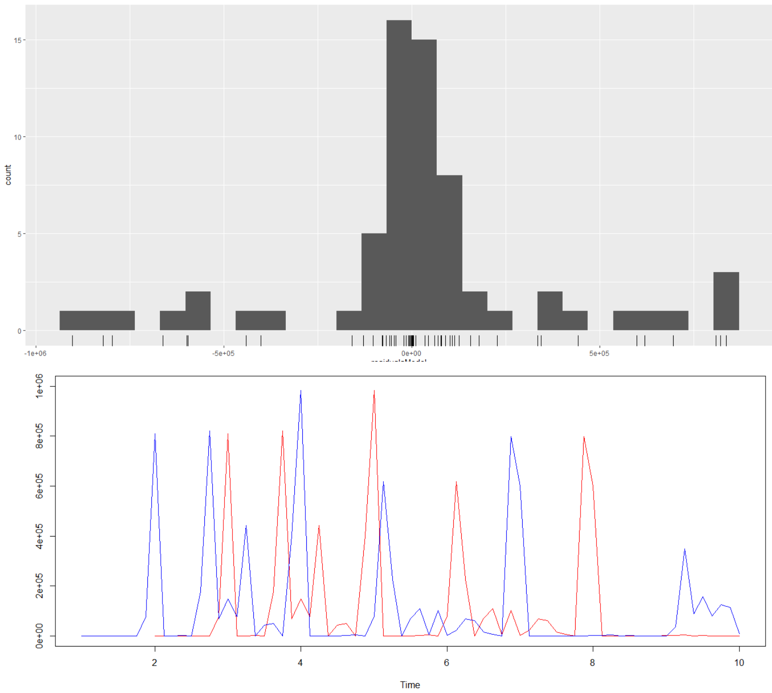

Figure 6: Results of estimating the accuracy of forecasting: the residual plot, fitted model

for the Foreach package procedure. Results of estimating the accuracy of forecasting time series R1-R5 shown in Table 2.

Time diagrams of traffic forecasting of the R1 type 1-6 are shown in Fig. 7 and type 7-13 are shown in Fig.8.

Based on the results of testing the forecasting models using the Leung-Box test, it was found:

- Average method, Simple exponential smoothing, Holt's linear trend method, Holt's linear trend method. Damped trend methods, Holt-Winters' additive method, Holt-Winters' multiplicative method. Damped method, ARIMA, Neural network models passed test for time series R1;
- Neural network models passed test for time series R2;
- ARIMA, Neural network models passed test for time series R3;
- ARIMA passed test for time series R4;
- Seasonal naive method, STL with multiple seasonal periods, Neural network models passed test for time series R5. The Neural network models showed the better flexibility by performing the Leung-Box test for the largest number of traffic types.

6. Conclusion

The article considers the problem of network traffic forecasting using the time series tool. A special feature of the proposed task is to consider the traffic of a single network user. The

Table 2
Results of estimating the accuracy of forecasting time series

Forecasting model	Metric	R1	R2	R3	R4	R5
Average	RMSE	$9 \cdot 10^4$	$6 \cdot 10^5$	$3 \cdot 10^4$	$2 \cdot 10^5$	$3 \cdot 10^4$
	MAPE	17548	32	$3 \cdot 10^6$	$2 \cdot 10^4$	$3 \cdot 10^4$
Naive method	RMSE	94519	951016	0	304040	20
	MAPE	17007	58	0	89	10
Seasonal naive method	RMSE	443164	642443	3416	390408	375832
	MAPE	7383279	45	1308	39881	5173244
Drift method	RMSE	95905	946121	0	304100	20
	MAPE	17407	57	0	1304	10
Simple exponential smoothing	RMSE	155610	788437	37169	269173	46863
	MAPE	32158	37	3482586	806723	4225309
Holt's linear trend method	RMSE	278414	788405	37206	279849	74122
	MAPE	58258	37	3720658	504311	6677233
Holt's linear trend method. Damped trend methods	RMSE	159568	788966	37206	269168	46912
	MAPE	33005	37	3720658	806727	4229920
Holt-Winters' additive method	RMSE	254612	875499	-	331179	-
	MAPE	1906904	42	-	1809142	
Holt-Winters' multiplicative method	RMSE	513598	856715	-	316473	-
	MAPE	3870810	43		3377	
Holt-Winters' multiplicative method. Damped method	RMSE	714634	729813	-	325126	-
	MAPE	13755727	48		1114028	
ARIMA	RMSE	104660	749206	28447	279456	35522
	MAPE	371876	42	2746378	4775078	3104142
STL with multiple seasonal periods	RMSE	179971	924319	135606	294120	112159
	MAPE	2208689	52	4851644	213482	3785941
Neural network models	RMSE	280028	778150	14116	204480	31622
	MAPE	2570521	44	1353843	2543793	2480219

following activities were used as traffic types: clicking on links — time series R1; downloading files (for example, movies) - time series R2, and others.

Typical, generally accepted models were used for forecasting, such as the drift Model, the Holt-winters Model, the ARIMA Model, and others. The accuracy estimation was performed for the considered models. All implementation was carried out in the R software package and using the forecast package. Based on the results obtained, we can conclude the following - the most acceptable model for predicting network traffic of an individual user is a forecasting model based on the decomposition of a series into separate components and their independent forecasting. Further work will focus on the use of combinations of forecasting models for traffic forecasting.

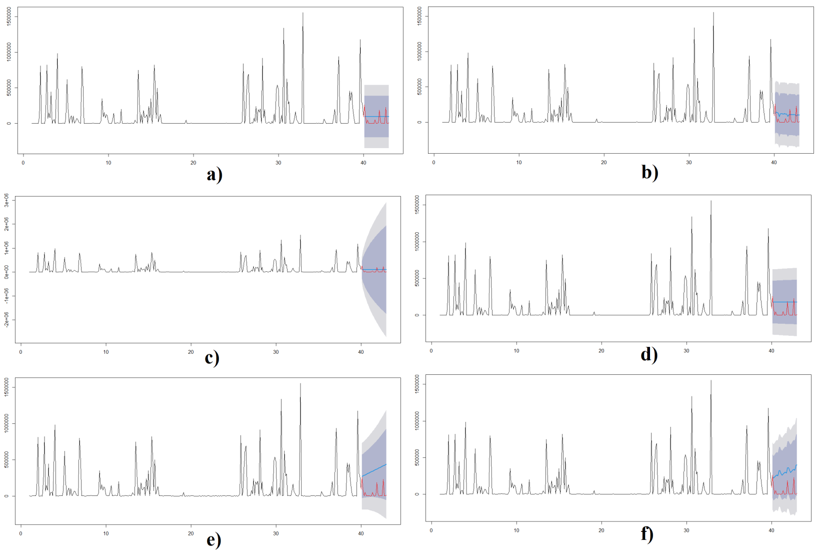

Figure 7: Forecast results for R1: a) average, b) an ARIMA model, c) Drift model, d) Holt model with a fading trend, e) Holt model, f) Holt-Winters model

References

[1] Ntlangu, M., Baghai-Wadji, A. (2017). Modelling Network Traffic Using Time Series Analysis: A Review. BDIOT2017. 10.1145/3175684.3175725

[2] International Telecommunication Union (ITU). (2014), The World in 2014: ICT Facts and Figures, Technical Report

[3] Gowrishankar, S. (2008), A Time Series Modelling and Prediction of Wireless Network Traffic, Georgian Electronic Scientific Journal: Computer Science and Telecommunications, Vol. 2, No. 16, pp. 40-52.

[4] Diaz-Aviles, E., Pinelli, F., Lynch, K., Nabi, Z., Gkoufas, Y., Bouillet, E. and Calabrese, F. (2015), Towards Real-time Customer Experience Prediction for Telecommunication Operators, IEEE International Conference on Big Data, pp. 1063-1072.

[5] KuanHoong, P., Tan, I.K.T. and YikKeong, C. (2012). Bit Torrent Network Traffic Forecasting with ARMA, International Journal of Computer Networks and Communications, Vol. 4, No. 4, pp. 143-156.

[6] Moussas, V. C., Daglis, M. and Kolega, E. (2005), Network Traffic Modelling and Prediction Using Multiplicative Seasonal ARIMA Models, 1st International Conference on Experiments/Process/System/ Modelling/Simulation/Optimisation, 1st IC-EpsMsO, 6-9 July, Athens, pp. 1-7

[7] Santos A. (2011). Network traffic characterization based on Time Series Analysis and Computational Intelligence. Journal of Computational Interdisciplinary Sciences. 2. 10.6062/jcis.2011.02.03.0046.

[8] Oduro-Gyimah F., Boateng, K. Analysis and modelling of telecommunications network traffic: a time series approach (2018).

[9] A. S. Moshnikov and V. S. Kolomoitcev, "Reliability Assessment of Distributed Control

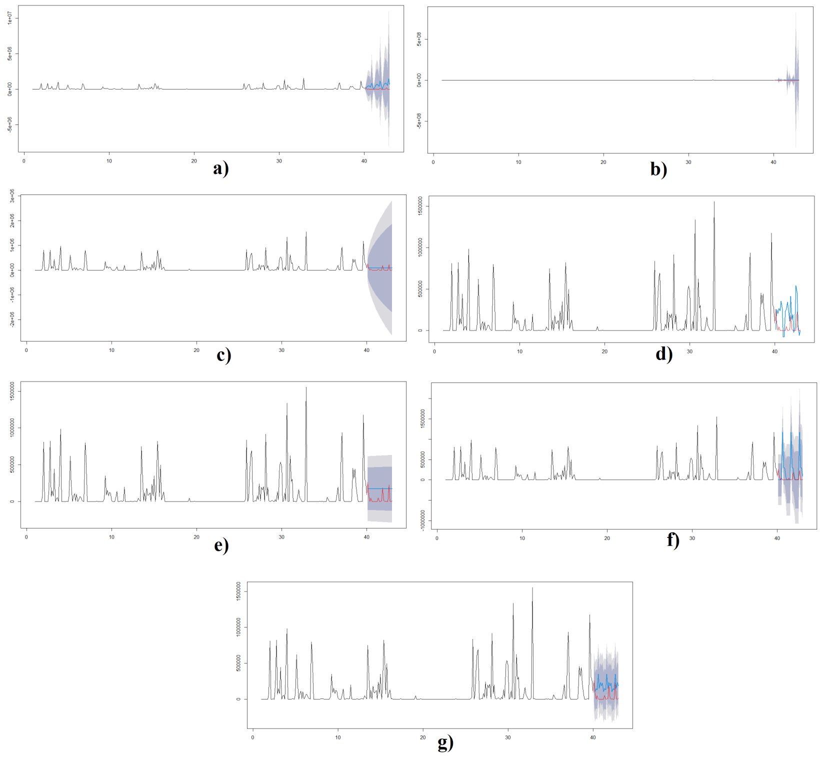

Figure 8: Forecast results for R1: a) Holt-Winters model, with multiplicative seasonality, b) Holt-Winters model, with multiplicative seasonality and fading trend, c) Last value, d) Auto-regression based on neural networks, e) Simple exponential smoothing model, f) The last value is seasonally adjusted (seasonal naive method), g) Neural network models

Systems with Network Structure," 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), Saint-Petersburg, Russia, 2020, pp. 1-4, doi: 10.1109/WECONF48837.2020.9131490.

[10] V. A. Bogatyrev, S. V. Bogatyrev and A. V. Bogatyrev, "Reliability and Probability of Timely Servicing in a Cluster of Heterogeneous Flow of Query Functionality," 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), Saint-Petersburg, Russia, 2020, pp. 1-4, doi: 10.1109/WECONF48837.2020.9131165.

[11] Ruey S. Tsay.: Analysis of Financial Time Series. 2nd edn. A JOHN WILEY and SONS, New Jersey (2005)

[12] Robert H. Shumway, David S. Stoffer.: Time series analysis and it Applications, With R Examples. Third edition, Springer, (2017)

[13] Crawley MJ.: The R Book. 2nd ed. Wiley Publishing; 2012