Mitigating Measurement Failures in Throughput Performance Forecasting

Maria C. Ferreira, Silvio E. Ribeiro, Francisco V. Nobre, Maria L. Linhares,
Thelmo P. Araújo, Rafael L. Gomes
State University of Ceará (UECE), Fortaleza-CE, Brazil
Email: {clara.mesquita,silvio.eduardo,valderlan.nobre,malu.linhares}@aluno.uece.br,
{thelmo.araujo,rafa.lopes}@uece.br

***Abstract*—Network monitoring services are performed by several companies and Internet Providers (ISP), which provide results of regular performance tests, such as throughput, loss, and delay, among others. These measurements help to understand the network's behavior and obtain information for strategic planning. However, when carrying out the measurements planned during network monitoring, failures may occur, which makes it difficult to carry out more complex activities, such as forecasting network performance. Within this context, this article presents a resilient and adaptive model for forecasting network performance, which includes the identification of measurement failures and applying data imputation techniques to adapt the data for the forecasting process (based on Neural Networks and Time Series Analysis). The experiments, using real data from the National Education and Research Network (RNP), show that the proposal can achieve high accuracy in forecasting with imputed data and outperform other existing forecasting approaches.**

***Index Terms*—Forecasting, Network Monitoring, Artificial Intelligence, Resilience.**

I. Introduction

Nowadays, almost all Internet Service Providers (ISPs) have network monitoring tools capable of providing relevant information for their administration. These tools perform regular performance tests, and the successful analysis of such tests constitutes useful metrics for better strategic planning. Corporations and institutions that operate on network infrastructures must have a solid analysis of this data, as it is essential for the maintenance and evolution of the network. This information covers key aspects of the network such as Bit Rate Control, Packet Loss, and Delay, among other important variables [1]–[3]. The analysis of network traffic through performance metrics primarily influences administrative decisions, including link capacity expansion, Quality of Service (QoS) and Quality of Experience (QoE) expectations, maintenance scheduling, Service Level Agreement (SLA) evaluations, and analysis of network resource requirements [4], [5].

Suitable throughput performance is critical for QoS and Internet service availability, as a low transfer rate can result in delays, bottlenecks, and network congestion, leading to an unsatisfactory user experience and reduced network efficiency [6]. On the other hand, a high bit rate ensures better data transmission, improving user experience, and increasing system productivity [7], [8]. Therefore, ensuring good network performance is crucial to providing high-quality services to users, requiring constant monitoring and optimization of key network metrics, fulfilling the SLA, and avoiding connection problems and high delays [9], [10].

Despite the evident importance of high network performance, it is crucial to recognize that the monitoring tools employed may encounter problems during testing (due to various aspects and limitations) resulting in lost measurement values and consequently low-quality data for analysis [11]. This can compromise data mining techniques that analyze the temporal and information, as well as hinder a good understanding and implementation of network management actions, in addition to planning and prospecting strategies [12]. Therefore, preprocessing techniques and synthetic data imputation help better understand network performance, bypassing potential limitations during measurements and allowing problems to be proactively anticipated [13]. Moreover, they prevent imminent failures and alterations, as they involve real data and imputation modeling techniques to estimate values and provide a better analysis of network performance under varied conditions [14].

Within this context, this paper presents a resilient network performance forecasting model. The proposal is considered resilient as it applies selected data imputation techniques to adjust the information for the forecast process when measurement failures occur (and consequently, gaps in the time series to be analyzed). Additionally, the proposed forecast model is considered adaptive as it adjusts the time series of the measurements performed through statistical analyses (such as decomposition, trend definition, and removal of cycle errors), creating a standardized time series that enables better training of forecast models. The processed time series is used as input for the Artificial Intelligence (AI) models that will perform the forecast process of a new series with the expected network performance, even when measurement failures occur.

To validate the solution and analyze its performance in terms of forecasting capability, experiments were conducted using real data from the Ipê Network Monitoring Service (Monipê)[1] of the National Research and Education Network (RNP). The results obtained show that the proposed model can achieve high levels of forecasting accuracy and outperforms the use of existing forecasting models about evaluation metrics within the context of network metrics management. Additionally, the

[1]monipe-central.rnp.br

978-3-903176-66-9

synthetic data imputation techniques improve the forecasting process compared to the original data, according to the RMSE evaluation.

The remainder of this paper is organized as follows: Section II presents existing solutions related to network monitoring and forecasting approaches. Section III describes the proposed forecasting model, while Section IV details the results of the experiments performed. Finally, Section VI concludes the paper and presents several future works.

II. Related Work

França et al. [15] present two neural network-based regression models for imputing missing data in IoT gateways, considering runtime and memory usage. Similarly, Park et al. [16] propose an approach using a deep learning model to estimate missing values of a variable in multivariate time series data, focusing on filling long and continuous gaps (e.g., several months of missing daily observations) rather than individual random missing observations. In the same way, Ding et al. [17] evaluate a series of imputation methods for filling missing values in time series collected from IoT devices. In this study, interpolation techniques such as Radial Basis Functions, Moving Least Squares (MLS), and Adaptive Inverse Distance Weighted were compared with each other and with KNN to fill the gaps. It was found that in this case, the Lancaster MLS interpolation technique yielded the best results. However, this study does not assess the impact of these imputation methods on time series forecasting within this context.

From the literature review, it is noted that no article has focused on developing a network performance forecasting model adaptable to the context of the measurement to be performed, which is the focus of this article, while also considering various data imputation techniques in the process.

III. Proposal

This section describes the proposed adaptive resilient forecasting model, which aims to adjust the end-to-end flow measurement data between two points in the network through time series analysis techniques for situations with elastic demand, as well as the identification and correction of measurement failures, thus enabling greater efficiency in the performance forecasting process. The solution is generally executed in four stages: (i) Collection of data necessary for forecasting, (ii) Correction of measurement failures through data imputation techniques, (iii) Time series analysis (applying statistics techniques, such as decomposition and seasonality), and (iv) Application of AI-based forecasting techniques. An overview of the structure of the solution proposed in this project is presented in Figure 1.

Initially, the solution retrieves measurement data from existing services (red element in Figure 1, where any tool can be considered, such as Perfsonar, SolarWinds, Auvik, etc. Subsequently, with the data in hand, the solution checks for gaps in the measurements for the analyzed period (gray element), applying imputation techniques to correct such failures (blue element). Then, a proposed model (based on network measurements) is applied in the time series processing stage (orange elements in Figure 1) to ensure data adjustment, i.e., adapting the temporal analysis to the context of elastic demand and the unique characteristics of each communication scenario. Finally, the result of the time series processing and analysis is deployed to generate the forecasting model. The actions performed in each of the presented stages will be detailed below.

A. Mitigation of Measurement Failures

The mitigation of measurement failures in network monitoring tools is crucial for maintaining the integrity and reliability of network performance data since these failures in network monitoring tools can occur due to various reasons like hardware malfunctions, network congestion, or software bugs. This fact can lead to gaps or inaccuracies in the collected data, which compromises the usage of AI solutions.

Imputation techniques play a significant role in addressing the issues arising from measurement failures, where two approaches for missing data correction can be used [18], [19]: Interpolation and AI. Interpolation techniques are based on a mathematical method that fits a function to the data and uses this function to generate missing data in the analyzed series. Linear interpolation is the simplest, which calculates the average of the previous and subsequent values of the missing data. Additionally, several other interpolation methods can be applied, such as Time-Weighted, Moving Average, and Moving Median, among others found in the literature.

Regarding the use of AI, these can be applied to analyze the dataset as a whole, define the gaps as targets, and propose a model to address these gaps. Each AI technique has a distinct approach, with applicable methods in the context of this project including K-Nearest Neighbors (KNN), Deep Learning Models, Neural Networks, etc. Next, we describe the techniques applied in this work: Moving Average, Moving Median, Linear Interpolation and K-Nearest Neighbors (KNN).

By applying these techniques, network solutions can ensure that the imputed data aligns closely with the true values, minimizing the impact of measurement failures on the overall dataset. First, it helps maintain the historical data's integrity, which is crucial for long-term performance analysis and trend identification. Second, imputation enhances real-time monitoring capabilities by providing a more accurate and immediate picture of network performance, enabling quicker and more effective responses to emerging issues. Finally, reliable data through imputation techniques supports advanced analytics and machine learning applications that rely on high-quality datasets to develop forecast models, which is the solution proposed in this paper.

B. Time Series Adjustment

In general, traditional forecasting models do not adequately address failures in datasets, which can impair the effectiveness of the results. This approach hinders the organization of regular time series and, consequently, precise analyses, directly affecting the quality of the forecasts, as evidenced in this

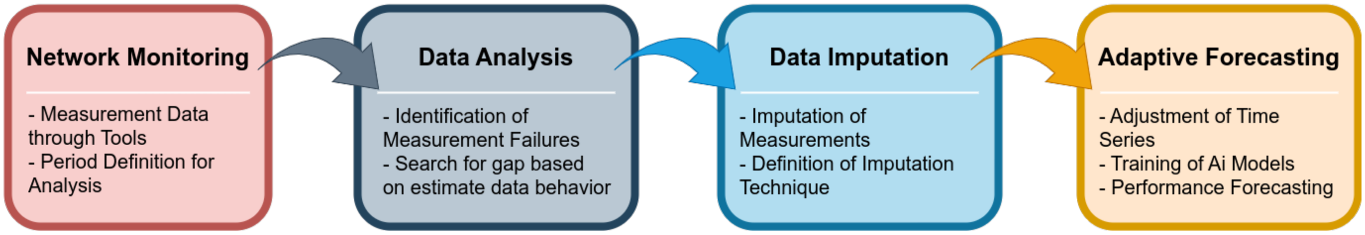

Fig. 1. Overview of the solution.

study. On the other hand, more robust models that include a data preprocessing step increase tolerance and resilience to measurement gaps. In this work, we combine this with Neural Network models, which are characterized by advanced computational capabilities and high tolerance to measurement errors.

A time series is a succession of observations recorded in chronological order at regular time intervals. The forecasting challenge involves fitting a model to anticipate future values of the series, considering past observations, i.e., its history. For this purpose, there are statistical techniques, Autoregressive or Smoothing models, as well as more robust techniques that use Neural Networks and Artificial Intelligence. Therefore, significant failures in time series datasets become a problem in forecasting, which relies on existing data to produce new observations. Moreover, the way this data is preprocessed directly interferes with the forecasting results.

The primary patterns observed in time series include (1) Trend (T), which manifests as a tendency for the temporal series to increase or decrease over time, though not necessarily in a linear fashion; (2) Seasonality (S), characterized by recurring patterns at fixed intervals within a period; and (3) Cycle (C), which also repeats but lacks a consistent periodicity, unlike seasonality. In time series analysis, the observed information Z_t at time t can be decomposed into its constituent components: the trend component T_t, the seasonal component S_t, and the error or residual component E_t, representing the part of the data not captured by the model.

Our solution employs a robust Seasonal Trend Decomposition using LOESS (STL) to highlight the components of seasonality, trend, and error. The robust variant of STL decomposition is resistant to outliers, enhancing the reliability of the decomposition results. Overall, STL decomposition offers a more detailed understanding of a time series' underlying patterns and trends, which can improve forecasting accuracy and help identify potential issues or anomalies in the data.

Initially, the STL method applies a low-pass filter to smooth the time series values, allowing low-frequency cycles to pass while attenuating cycles that are far from the cutoff measurement and removing high-frequency noise. This process produces a smooth trend component. Next, the trend component is subtracted from the original time series to obtain a detrended series, which is then subjected to a seasonal filter to estimate the seasonality. Subtracting this estimated seasonality yields a residual series. To reconstruct the original time series, the trend, seasonality, and residual components are added together. This makes STL decomposition particularly effective for analyzing and modeling data with seasonal patterns [20]. The method is versatile, accommodating a wide range of seasonal periods and applicable to both additive and multiplicative time series formats.

Recognizing the importance of minimizing the impacts of gaps during analysis and forecasting, a comprehensive preprocessing of the training data was conducted. This approach involved applying imputation techniques to correct and handle previously observed failures. The imputation step is crucial for enhancing the forecast's performance in the subsequent training phase. In environments where substantial gaps can undermine the reliability of analyses, imputation techniques emerge as a critical component of the process.

C. Forecasting Models

The LSTM model is a type of recurrent neural network designed specifically to overcome the issues of gradient explosion/vanishing, which typically arise when learning long-term dependencies, even across very long time intervals [21]. In general, this issue can be mitigated by using a Constant Error Carousel (CEC), which keeps the error signal within each unit's cell. These cells are recurrent networks with an enhanced architecture that extends the CEC with additional features, specifically the input and output gates that form the memory cell. The self-recurrent connections provide feedback with a time lag. LSTM is well-suited for classifying, processing, and forecasting time series with unknown durations. Its relative insensitivity to the length of gaps between relevant events gives LSTM an advantage over traditional RNN models and other sequence learning methods.

Similarly, the GRU is the next generation of recurrent neural networks and is quite similar to an LSTM. GRUs have eliminated the cell state and used the hidden state to transfer information. This architecture has only two gates, a reset gate, and an update gate, which are used to address the gradient vanishing problem of a standard RNN. Typically, these are two vectors that determine which information should be passed on to the output. What's special about them is their ability to be trained to retain information from a long time ago without dissipating it or discarding irrelevant information for forecasting. LSTMs have two distinct states passed between cells: the cell state, which carries long-term memory, and the hidden state, which carries short-term memory. In contrast, GRUs have only one hidden state transferred between time steps. This hidden state can maintain both long-term and short-term dependencies simultaneously due to the constraints and calculations applied to it and the input data.

IV. Experiments

Initially, it is worth mentioning that the developed code and the data used in the experiments are available in the solution repository[2]. This initiative aims to allow the scientific community to reproduce the experiments and generate new results with other network data.

A. Configuration of Experiments

To conduct experiments using real-world data, data from RNP (Rede Nacional de Ensino e Pesquisa) were utilized through the Ipê Network Monitoring Service (MonIPÊ). MonIPÊ follows the international monitoring standard perfSONAR, measuring several network metrics, including Throughput. The throughput is measured between two endpoints in the network and it is executed using two different congestion control types for TCP (Cubic and BBR).

One important point is the difference between TCP BBR and Cubic, as they exhibit distinct behaviors [22], [23]. TCP Cubic uses a cubic function over time to increase the congestion window time responding effectively to packet loss, while TCP BBR, instead of directly responding to packet loss, also considers bandwidth, RTT, and other metrics to determine packet sending rates and congestion window sizes. Therefore, these two versions of TCP have distinct behaviors, that affect the forecasting process.

B. Communication Points

In the experiments, communication pairs covering Points of Presence (PoPs) across the network in the Brazilian states of Minas Gerais (MG), Rio Grande do Sul (RS), Pará (PA), Amazonas (AM), Bahia (BA), and Paraná (PR) were examined. These Ipê Network PoPs were selected due to their diverse geographical locations, which impact link usage and network load, varying infrastructure capacities (links ranging from 200 Gbps to 1 Gbps), leading to distinct communication behaviors.

It is worth noting that end-to-end communication between various points of the network exhibits distinct behaviors throughout the day and during the week, as the infrastructure usage follows a social pattern among its users. Therefore, these communication points were considered, covering a significant portion of the network traffic, interconnecting the North, Northeast, and South regions of Brazil. Specifically, these points were chosen due to: (i) Geographical heterogeneity affects the number of links used in the end-to-end path and the total load generated on the network infrastructure; (ii) Link Capacity since they exhibit high variability in load, causing communication originating from one PoP to behave differently compared to others with distinct end-to-end capacity variations.

[2]https://github.com/LarcesUece/Resilient-Performance-Forecasting

Before studying the impact of imputation methods on forecasting, a preliminary evaluation of the imputation techniques used was conducted. For this purpose, datasets from the same two communication pairs were selected. These pairs were chosen not only for the reasons already specified but also because they offer a greater contrast between the results generated by different imputation techniques. Additionally, it's important to note that the percentages of missing data in the datasets were taken into account during the selection process.

Such information was derived from an initial analysis that quantified the extent of missing data in each dataset. During the observed six-month period, the dataset revealed the following percentages of missing data: Pr-Am Bbr Flow had 39.43%, Pr-Am Cubic Flow had 39.37%, Mg-Rs Bbr Flow had 19.37%, Mg-Rs Cubic Flow had 20.26%, Pa-Ba Bbr Flow had 29.09%, and Pa-Ba Cubic Flow had 28.85%. To address this, the procedure involved identifying the longest contiguous sequence without measurement failures. From this sequence, values were randomly extracted to represent approximately the same percentage as the missing data. Subsequently, imputation techniques were applied to each series, followed by the calculation of RMSE for the generated values compared to the original values. This approach aimed to assess the quality of generating new observations using the imputation techniques compared to the existing original values.

C. Training of AI Models

Regarding the training of the forecasting model, time series representing end-to-end flow between the same two communication points were used. After the analysis and application of imputation techniques for each generated series, 80% of the series were set aside for training and 20% for testing and model validation, approximately 560 and 140 cycles, respectively. The series underwent four rounds of different training, varying the parameters of LSTM and GRU models to evaluate the performance of each forecasting model. Regarding the hyperparameters used in the models (available in the repository), a model selection process (GRU and LSTM) employed the grid search technique. This technique systematically explores a series of predefined hyperparameters to identify the optimal combination for each model.

Therefore, the four imputation techniques described in Section III-A were applied to each of the following time series: MG-RS Cubic Flow, MG-RS BBR Flow, PA-BA Cubic Flow, PA-BA BBR Flow, PR-AM Cubic Flow, and PR-AM BBR Flow, totaling 24 resulting time series with filled values. After this, all resulting series underwent the adaptive forecasting process focusing on LSTM and GRU models that showed better performance in forecasting. Finally, RMSE was calculated for each forecasting result compared to the actual values.

D. Evaluation Methods

Regarding evaluation methods, two metrics were analyzed during the experiments:

- **Accuracy in Performance Ranges**: Internet Providers typically establish performance ranges to categorize measurements, aiming to mitigate fluctuations in performance during evaluations and to emphasize thresholds based on service requirements that utilize the network [24]. We defined five throughput value ranges based on the framework used in RNP's MonIpe: Red for values less than 0.2 Gbit/s; Orange for values in the range [0.2; 0.5] Gbit/s; Yellow for values in the range [0.5; 0.8] Gbit/s; Blue for values in the range [0.8; 1.0] Gbit/s; and Green for values greater than or equal to 1.0 Gbit/s. These ranges establish an accuracy criterion: If the forecasted value falls within the same range as the actual value, it is considered correct forecasting; otherwise, it is a deviation.

- **Root Mean Square Error (RMSE)**: To more accurately evaluate the imputation techniques and the performance of forecasting models against real values, RMSE was used. This allowed validation of the values generated by the techniques and assessment of the forecasting model using the same imputation methods. RMSE for a given period T is defined by Equation 1, where $\hat{y}_t$ represents the forecast value and y_t denotes the actual performance value at time t. In the context of the RMSE metric, higher values indicate less accurate imputation performance from the approach.

$$\text{RMSE}(T) = \frac{1}{\sqrt{T}} \left(\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 \right)^{\frac{1}{2}} \tag{1}$$

Accuracy and RMSE are essential evaluation metrics in forecasting experiments due to their distinct ways of measuring model performance, providing complementary insights into model performance. While accuracy gives a broad sense of the model's reliability in predicting categorical outcomes, RMSE offers a detailed measure of prediction quality for continuous variables, highlighting larger errors that may need attention. Together, these metrics enable a comprehensive evaluation of forecasting models, helping researchers and practitioners understand both the overall success rate and the magnitude of errors, ultimately guiding the refinement and improvement of forecasting methods.

V. Results

This section presents the results obtained from the analysis of experiments conducted with a real dataset, where Subsections V-A and V-B discuss the main points regarding data imputation actions and performance forecasting, respectively.

A. Analysis of Data Imputation Methods

Regarding data imputation, Figure 2 illustrates the comparison of each technique with the original data using RMSE. For example, Cubic traffic was used due to its high variability and greater difficulty in forecasting exact values. Thirty percent of values were randomly removed (representing the general order of missing data in collected measurements) within a narrow and regular interval of 18 and 35 values for MG-RS, PR-AM, and PA-BA, respectively. It is worth noting that data imperfections are common, resulting in irregularities that especially affect temporal analysis aspects.

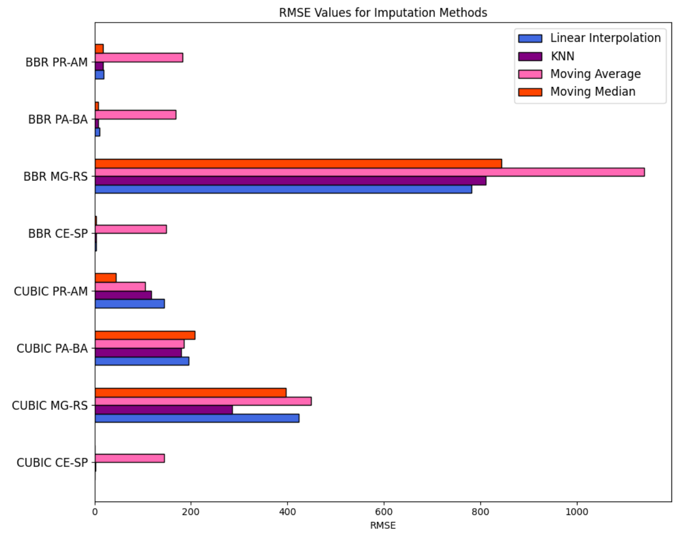

Fig. 2. Results of RMSE for Data Imputation.

It is observed that the dense data distribution in the cases of PA-BA and PR-AM (for Cubic traffic) reduces the need for imputation within narrow intervals, making accuracy calculations based on specific intervals less informative as a result. Consequently, the RMSE metric tends to be higher in such situations due to limited room for substantial improvements through imputation techniques. Therefore, regarding data imputation efficiency, the Moving Average and KNN stand out with lower RMSE values. However, it is important to note that these methods still result in imputation errors, even if reduced. Thus, it is necessary to understand the impact of these imputations on the forecasting process.

B. Evaluation of Forecasting Process

This section presents the results obtained from the analysis of experiments conducted with real datasets from the imputation techniques described earlier.

Analyzing Figure 3, it is possible to notice that, overall, for the communication points PA-BA, the forecasting models achieved satisfactory levels of error, with RMSE values ranging between 50 and 250 Mbits/s. The TCP Cubic, due to its "aggressive" congestion window behavior, which complicates model adjustment for forecasting, leads to increased errors in all cases, as this behavior causes significant variation in the collected measurements. It is important to note that for TCP BBR, the forecasting models achieved very satisfactory RMSE results, outperforming TCP Cubic in all cases, as both forecasting models reached the lowest error values, with the GRU showing a slight advantage over LSTM in all cases.

It can be observed in Figure 3 that the factor that most influences the forecasting outcome is exactly the imputation technique applied to the data. The variation in some techniques produces similar results, especially Linear Interpolation and KNN, which proved efficient for both cases.

TABLE I
ACCURACY OF FORECASTING PROCESS FOR PERFORMANCE RANGE DEFINITION

AI Model \ Comm. Points	PA-BA BBR	PA-BA Cubic	PR-AM BBR	PR-AM Cubic	MG-RS BBR	MG-RS Cubic
GRU - Linear Interpolation	94.9%	62.7%	93.9%	86.2%	99.8%	54.1%
GRU - KNN	94.2%	64.9%	97.7%	90.0%	99.4%	54.3%
GRU - Moving Average	19.4%	24.4%	18.9%	18.9%	99.2%	52.6%
GRU - Moving Median	92.0%	39.4%	98.4%	91.6%	99.2%	52.6%
LTSM - Linear Interpolation	94.9%	59.8%	93.1%	83.9%	99.2%	54.1%
LTSM - KNN	94.2%	69.3%	96.2%	90.1%	99.2%	54.1%
LTSM - Moving Average	25.8%	25.1%	19.6%	19.6%	99.4%	52.3%
LTSM - Moving Median	92.0%	44.5%	95.4%	90.0%	99.8%	53.2%

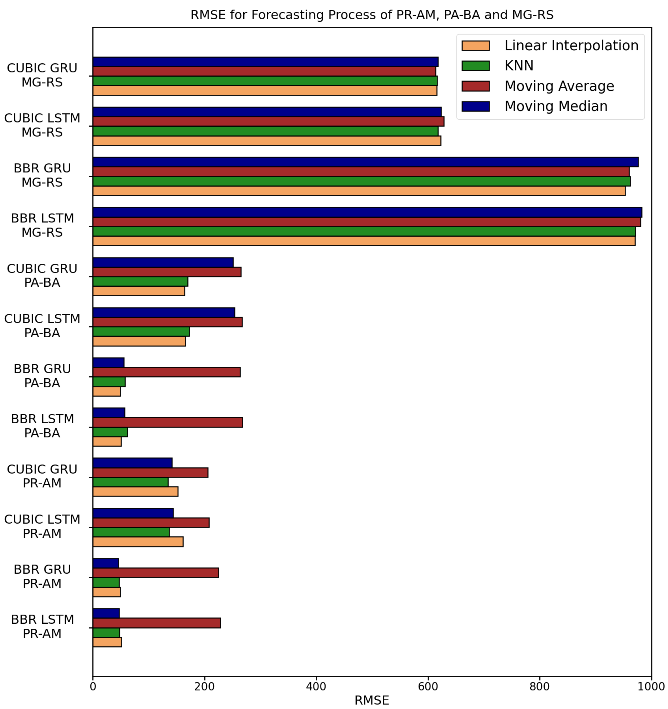

Fig. 3. RSME of Forecasting Process.

It can be observed that there is a similarity in the performance of the model using Linear Interpolation, KNN, and Moving Median for the PR-AM communication points. Although KNN is a recognized Artificial Intelligence technique, it performs similarly when compared to Linear Interpolation, for example, which is a less sophisticated technique. In both cases, the Moving Average shows the worst performance when considering the forecasting process in all scenarios. This is a relevant observation, considering that it was noted in Section V-A that, in general, the same technique was the closest to the actual measurement values. This reinforces that a narrow range of occurrences cannot effectively represent the efficacy of techniques.

It is possible to note that for all communication points, according to Table I, the accuracy was suitable for most AI techniques with data imputation. The exception of this scenario was the moving average method. From a strategic perspective, accurate forecasts support long-term planning and strategic decision-making by providing reliable data on performance trends and future demands. It also aids in identifying network weak points, managing risks, and ensuring compliance with service level agreements. Moreover, accurate forecasting supports deploying new technologies and services, ensuring the network can scale efficiently in response to increasing demands and fostering business growth and innovation.

To provide visual feedback, Figure 4 illustrates one prediction case compared to the measurements, illustrating the model's fit to the original data for the PA-BA throughput, using one of the forecasting cases as an example. It can be observed that the LSTM and GRU forecasting models achieve good results against the selected test values, accurately capturing fluctuations and long-term dependencies. The robustness of using Neural Networks shows that the forecasting models still recognize patterns with high precision even though they occur over imputed data and adapted time series. They can extrapolate future values with quality, indicating an adequate generalization capability for this context.

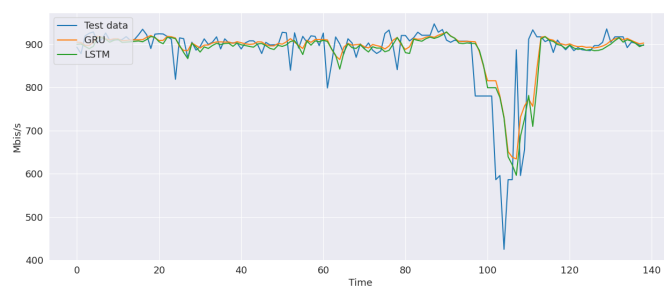

Fig. 4. Example of Forecasting Model.

From the experimental analysis conducted in this paper, it is possible to conclude that the forecasting models provide a significant advantage for ISPs since it is possible to understand future network performance (that affects issues like SLA, QoS, and QoE). It is also noted that despite its efficiency, the step of removing cycle errors could be revisited to maximize its effectiveness on datasets with numerous outliers.

VI. Conclusion

The evaluation of network performance is an action aimed at acquiring relevant data for the strategic planning of companies and Internet service providers. The ability to forecast this performance becomes an essential functionality to ensure the effectiveness of services operating on the network, directly contributing to the user experience in this context. However, the process of forecasting performance is complex, especially given the current reality of measurement failures, which lead to gaps in data and consequently hinder the execution of more complex tasks.

To address this situation, this article presented an adaptive and resilient model for network performance forecasting. This model identifies measurement failures and utilizes data imputation techniques to prepare the data for the forecasting process, based on Neural Networks and Time Series Analysis. The experiments conducted using real data from the National Research and Education Network (RNP) demonstrate that the proposed solution achieves high levels of accuracy in forecasting with imputed data. Thus, the proposed solution enables network administrators and ISPs to plan network infrastructure and to perform actions to improve QoS delivered from the network communication.

As future work, the aim is to evolve the solution to consider multivariate time series, including other types of network measurements such as latency and packet loss. Additionally, there is interest in exploring the behavior of the solution using combined machine-learning models.

Acknowledgement

The authors would like to thank the CNPq (N^o 303877/2021-9 and N^o 405940/2022-0) and CAPES (N^o 88887.954253/2024-00) of Brazil for the financial support.

References

[1] M. M. Silveira, D. S. Silva, S. J. R. Rodriguez, and R. L. Gomes, "Searchable symmetric encryption for private data protection in cloud environments," in *Proceedings of the 11th Latin-American Symposium on Dependable Computing*, ser. LADC '22. New York, NY, USA: Association for Computing Machinery, 2023, p. 95–98. [Online]. Available: https://doi.org/10.1145/3569902.3570171

[2] P. Vasileva, S. McKee, A. Penev, and I. Vukotic, "Ps-dash –analysis, monitoring and visualization of network measurements," in *2021 International Conference Automatics and Informatics (ICAI)*, 2021, pp. 93–96.

[3] H. Yang, X. Li, W. Qiang, Y. Zhao, W. Zhang, and C. Tang, "A network traffic forecasting method based on sa optimized arima–bp neural network," *Computer Networks*, vol. 193, p. 108102, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621001821

[4] R. K. P. Mok, H. Zou, R. Yang, T. Koch, E. Katz-Bassett, and K. C. Claffy, "Measuring the network performance of google cloud platform," in *Proceedings of the 21st ACM Internet Measurement Conference*, ser. IMC '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 54–61. [Online]. Available: https://doi.org/10.1145/3487552.3487862

[5] R. Gomes, W. Junior, E. Cerqueira, and A. Abelem, "A qoe fuzzy routing protocol for wireless mesh networks," in *Future Multimedia Networking*, S. Zeadally, E. Cerqueira, M. Curado, and M. Leszczuk, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–12.

[6] D. A. Moreira, H. P. Marques, W. L. Costa, J. Celestino, R. L. Gomes, and M. Nogueira, "Anomaly detection in smart environments using ai over fog and cloud computing," in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2021, pp. 1–2.

[7] A. L. C. Portela, S. E. S. B. Ribeiro, R. A. Menezes, T. de Araujo, and R. L. Gomes, "T-for: An adaptable forecasting model for throughput performance," *IEEE Transactions on Network and Service Management*, pp. 1–1, 2024.

[8] R. L. Gomes, L. F. Bittencourt, E. R. Madeira, E. Cerqueira, and M. Gerla, "A combined energy-bandwidth approach to allocate resilient virtual software defined networks," *Journal of Network and Computer Applications*, vol. 69, pp. 98–106, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804516300728

[9] R. L. Gomes, L. F. Bittencourt, and E. R. M. Madeira, "Reliability-aware network slicing in elastic demand scenarios," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 29–34, 2020.

[10] A. L. Portela, R. A. Menezes, W. L. Costa, M. M. Silveira, L. F. Bittecnourt, and R. L. Gomes, "Detection of iot devices and network anomalies based on anonymized network traffic," in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–6.

[11] W. L. Costa, A. L. Portela, and R. L. Gomes, "Features-aware ddos detection in heterogeneous smart environments based on fog and cloud computing," *International Journal of Communication Networks and Information Security*, vol. 13, no. 3, pp. 491–498, 2021.

[12] M. M. Silveira, A. L. Portela, R. A. Menezes, M. S. Souza, D. S. Silva, M. C. Mesquita, and R. L. Gomes, "Data protection based on searchable encryption and anonymization techniques," in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023, pp. 1–5.

[13] J. Naf, M.-L. Spohn, L. Michel, and N. Meinshausen, "Imputation scores," 2022.

[14] G. da Silva, D. Oliveira, R. L. Gomes, L. F. Bittencourt, and E. R. M. Madeira, "Reliable network slices based on elastic network resource demand," in *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, 2020, pp. 1–9.

[15] C. M. França, R. S. Couto, and P. B. Velloso, "Missing data imputation in internet of things gateways," *Information*, vol. 12, no. 10, 2021. [Online]. Available: https://www.mdpi.com/2078-2489/12/10/425

[16] J. Park, J. Müller, B. Arora, B. Faybishenko, G. Pastorello, C. Varadharajan, R. Sahu, and D. Agarwal, "Long-term missing value imputation for time series data using deep neural networks," *Neural Computing and Applications*, vol. 35, no. 12, pp. 9071–9091, 2023.

[17] Z. Ding, G. Mei, S. Cuomo, Y. Li, and N. Xu, "Comparison of estimating missing values in iot time series data using different interpolation algorithms," *International Journal of Parallel Programming*, vol. 48, no. 3, pp. 534–548, June 2020. [Online]. Available: https://doi.org/10.1007/s10766-018-0595-5

[18] S. Thakur, J. Choudhary, and D. P. Singh, "A survey on missing values handling methods for time series data," in *Intelligent Systems*, A. Sheth, A. Sinhal, A. Shrivastava, and A. K. Pandey, Eds. Singapore: Springer Singapore, 2021, pp. 435–443.

[19] S.-F. Wu, C.-Y. Chang, and S.-J. Lee, "Time series forecasting with missing values," in *2015 1st International Conference on Industrial Networks and Intelligent Systems (INISCom)*, 2015, pp. 151–156.

[20] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition," *J. Off. Stat*, vol. 6, no. 1, pp. 3–73, 1990.

[21] S. H. Rafi, Nahid-Al-Masood, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated cnn and lstm network," *IEEE Access*, vol. 9, pp. 32 436–32 448, 2021.

[22] K. Miyazawa, S. Yamaguchi, and A. Kobayashi, "Mechanism of cyclic performance fluctuation of tcp bbr and cubic tcp communications," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2020, pp. 1139–1144.

[23] T. Kanaya, N. Tabata, and S. Yamaguchi, "A study on performance of cubic tcp and tcp bbr in 5g environment," in *2020 IEEE 3rd 5G World Forum (5GWF)*, 2020, pp. 508–513.

[24] L. Bayne, D. Schepis, and S. Purchase, "A framework for understanding strategic network performance: Exploring efficiency and effectiveness at the network level," *Industrial Marketing Management*, vol. 67, pp. 134–147, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019850117305655