Time Series Analysis in Mobile Communications Systems

AMANDA DE FIGUEIREDO SOUSA
MÁRCIO JOSÉ TEIXEIRA
VARESE SALVADOR TIMÓTEO

Grupo de Óptica e Modelagem Numérica - GOMNI
Faculdade de Tecnologia - FT
Universidade Estadual de Campinas - UNICAMP
13484-332 , Limeira - *SP*, Brasil

Abstract. The evolution of mobile communication systems, marked by the exponential growth of users and the demand for efficient network management, underscores the importance of optimization to mitigate congestion, increase transmission rates, and reduce packet loss. In this context, the Kalman-Takens Filter (KTF) stands out for its analytical capability in prediction and optimization, offering a dynamic approach to spectrum allocation in mobile networks. This study, grounded in a comprehensive literature review and time series analysis, investigates the application of the KTF, with its effectiveness quantified by the root mean square error (RMSE). Computational simulations have shown significant improvements in network performance, demonstrating that the KTF can operate efficiently in real-time, optimizing resource allocation. This approach aims to enhance user experience in high-data-demand environments, contributing to the development of advanced management strategies in 5G networks. The meticulous analysis of the RMS error, aiming at its minimization, proved effective, providing crucial insights for resource management, essential to addressing the increasing traffic volume and complexity in future mobile communications networks.

Keywords: Kalman filtering, Mobile communication systems, Time series

(Received June 27th, 2025 / Accepted July 7th, 2025)

1 Introduction

The evolution of mobile communication systems from the 1970s to the present day is marked by technological innovations and advancements in telecommunications services. The first generation (1G) initiated mobile voice communication with the Advanced Mobile Phone System (AMPS) technology, operating on Frequency Division Multiple Access (FDMA) in the Ultra High Frequency (UHF) band, though limited in capacity and quality. In the 1990s, the second generation (2G) transitioned to digital technology. Technologies such as Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA), and Global System for Mobile Communications (GSM) evolved into Enhanced Data Rates for GSM Evolution (EDGE), improving connectivity and transmission rates. In the 2000s, the third generation (3G) facilitated voice calls, emails, Internet browsing, and video calls. The Universal Mobile Telecommunications System (UMTS) technology evolved into High-Speed Packet Access (HSPA) and Evolved High-Speed Packet Access (HSPA+), enhancing data transmission speeds [15].

The fourth generation (4G) in the 2010s transformed mobile devices into multifunctional tools, capable of high-quality video calls, high-definition streaming, online gaming, and support for cloud-based applications. Standards such as Worldwide Interoperability for Microwave Access (WiMAX) and Long-Term Evolution (LTE), enhanced to LTE Advanced-Pro (LTE-A), made 4G essential for industrial automation, the Internet of Things (IoT), and emerging fields of Artificial Intelli-

gence (AI) and Machine Learning (ML). The fifth generation (5G), currently being deployed, uses the NR standard to provide unprecedented Internet speeds, reduced latency, and expanded connectivity. 5G promotes advancements in IoT, AI, ML, cloud computing, Augmented Reality (AR), Virtual Reality (VR), enhanced security, and robotics, marking a leap in digital transformation [15]. The sixth generation (6G) is in the conceptual phase, with a launch expected around 2030, aiming to further extend wireless network capabilities, focusing on integrating Extended Reality (XR) and advanced ML services [8].

The motivation for this work arises from the urgent need for innovative spectrum management solutions in 5G networks, driven by the growing global demand for mobile network resources. In various countries, there are distinct efforts to allocate spectrum and expand 5G coverage, notably in the United States, which forecasts a 1400 MHz deficit by 2032 [21], and in Brazil, where the regulating agency recently allocated 120 MHz in the 4.9 GHz band [42]. China leads global implementation with over 3.38 million base stations and the pioneering allocation of spectrum in the 6 GHz band for 5G and 6G [43]. This study applies the Kalman-Takens Filter (KTF) in time series analysis to enhance spectrum use efficiency, aiming to improve user experience with higher transmission rates, lower packet loss, and reduced latency, while operating in real-time and conserving system resources.

Predictive resource allocation in mobile communication systems has been extensively studied. [11] classified predictive schemes into geographic, link, social, and traffic categories, highlighting methods based on user trajectories, channel quality, application preferences, and measurements from higher layers of the OSI model. Ref. [39] proposed using the Kalman Filter (KF) to estimate future transmission rates, enhancing throughput and reducing packet loss with micro predictions in 10-millisecond windows. In 2021, Ref. [38] extended this work by applying a black-box model based on the KTF to predict the signal-to-noise ratio over intervals of 10 and 100 seconds, utilizing traffic and link information to optimize modulation and resource allocation. These studies demonstrated the effectiveness of the KF and KTF in improving network performance based on predictive data [29, 40, 41].

Other relevant studies include [19], [20], who applied the KTF to systems such as Lorenz-96 and population dynamics, and [24] in hydrology. Ref. [34] used the Fourier transform to predict traffic in Shanghai, while [9] applied the KF for resource allocation with predictions up to 60 seconds. Ref. [25] compared schemes based on KF, neural networks, and Markov chains to predict the number of users in WLANs. Recent applications of the KF in mobile networks have been published in [14], [33], and [7]. Ref. [30] explored the use of ML in 5G networks, highlighting challenges related to latency and computational costs for real-time applications.

The contributions of this work include the analysis and validation of a predictive model applying the KTF to optimize spectrum allocation in 5G mobile communication systems, improving network performance and user experience through efficient and real-time spectrum management. The contributions are as follows:

- Conceptualize the current challenges of spectrum allocation in 5G networks;
- Apply the KTF predictive model in 5G time series analysis;
- Analyze the model's performance in optimizing spectrum use and improving user experience.

The remainder of this article is organized as follows: Section 2 presents the fundamentals of the 5G mobile communication system; Section 3 contains the concepts of dynamic systems and time series; Section 4 discusses Kalman Filters; Section 5 reviews related works in the study area; Section 6 focuses on the case study with the description of the materials and methods used in the experiments; Section 7 provides an analysis of the experimental results; Section 8 concludes the work by discussing implications and suggesting future work.

2 5G Mobile Communications Systems

The architecture of the New Radio (NR) system is essential for 5G mobile communications, addressing requirements such as enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low-Latency Communications (URLLC), and massive Machine Type Communications (mMTC). It supports sub-1 GHz and millimeter-Wave (mmWave) frequencies, utilizing massive Multiple Input Multiple Output (MIMO) and beamforming to enhance efficiency. NR adopts a scalable Orthogonal Frequency-Division Multiplexing (OFDM) numerology, allows dual connectivity with 4G LTE and 5G, and supports network slicing to create virtualized logical networks. Divided into Access Network (NG-RAN) and Core Network (5GC), the 5GC architecture adopts a Service-Based Architecture (SBA), separating the control plane from the user plane to manage connection, sessions, routing, policies, and user data, aiming for flexibility and efficiency [5, 1].

The 5G access network protocols ensure fast, reliable, and efficient communication between user devices and the network, following a layered structure. The physical layer (PHY) handles the transmission of bits through the radio channel, using advanced modulation schemes and massive MIMO to increase data rates and signal quality. The Media Access Control (MAC) layer controls data transfer, implements Hybrid Automatic Repeat Request (HARQ) for reliability, and facilitates dynamic resource allocation. The Radio Link Control (RLC) layer ensures reliable data transfer, while the Packet Data Convergence Protocol (PDCP) handles header compression and data encryption. The Radio Resource Control (RRC) layer manages control plane signaling, mobility management, and session establishment. These protocols are divided between the Control Plane, managed by the RRC, and the User Plane, managed by PDCP and RLC, supporting high data rates, low latency, and massive connectivity, ensuring efficient use of radio spectrum and network resources [4].

The 5G air interface is the physical medium for wireless communications between User Equipment (UE) and next-generation Node B (gNBs), designed to meet the demands of eMBB, URLLC, and mMTC. It uses advanced modulation (256QAM) to increase data rates and spectral efficiency, massive MIMO to serve multiple users simultaneously, and beamforming to improve signal quality. Dynamic Spectrum Sharing (DSS) allows the coexistence of 5G NR and 4G LTE in the same frequency band, while Carrier Aggregation (CA) combines multiple bands for higher data rates. 5G supports URLLC with reduced transmission times and lower latency, operating in frequency bands from below 1 GHz to above 24 GHz for different coverage and capacity requirements. The flexible frame structure includes short Transmission Time Intervals (TTI) and supports Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD), ensuring efficient spectrum use [3], [2]. This design balances innovative technologies and practical adaptations to provide high speed, massive connectivity, and ultra-reliable low-latency communications.

In 5G, the basic unit of resource allocation is the Resource Block (RB), essential for efficient spectral resource management. An RB is the smallest allocable unit in the frequency domain, consisting of a number of subcarriers and a specific duration in the time domain. According to specification [6], the 5G frame structure consists of frames, subframes, slots, and mini-slots, with RBs dynamically allocated to users based on demand and channel conditions. An RB typically encompasses 12 subcarriers, with subcarrier spacings varying from 15 kHz to 240 kHz. The duration of a slot varies with subcarrier spacing. Flexible RB allocation allows support for diverse services, such as eMBB and URLLC, reflecting 5G's goal of providing a versatile, high-performance network [24].

Resource allocation in 5G is crucial to ensure efficient use of spectrum and network resources, meeting the diverse needs of users and applications. According to specification [3], allocation types include: in the frequency domain, where resources are allocated in terms of RBs across the spectrum, potentially using Carrier Aggregation (CA) to combine multiple bands; in the time domain, involving data scheduling in slots or mini-slots; and in the spatial domain, using massive MIMO and beamforming to serve multiple users simultaneously. The Signal-to-Interference-and-Noise Ratio (SINR) metric assesses the quality of the wireless communication link, essential for managing Quality of Service (QoS) and ensuring that high-priority services receive the necessary resources. 5G also supports network slicing, creating virtual networks with dedicated resources for specific service types. This dynamic process balances frequency, time, and space to optimize network performance and efficiency [3].

Schedulers in 5G networks use various metrics to efficiently allocate resources and meet user and application requirements. According to specification [3], metrics such as Channel Quality Indicator (CQI) help select the optimal modulation and coding scheme, while Buffer Status Reports (BSR) prioritize users with more data in the queue. QoS requirements, such as bandwidth and latency, guide resource allocation for different services (eMBB, URLLC, mMTC). Measures of signal strength and quality (Reference Signal Received Power - RSRP and Reference Signal Received Quality - RSRQ) influence allocation and handover decisions, and traffic load helps balance the network. Latency requirements are critical for low-latency services like URLLC, and resource utilization efficiency ensures effective network use. User mobility and historical data are also considered for predictive scheduling, optimizing resource allocation. These metrics balance CQI optimization, allocation history, and QoS requirements, continuously monitored for effective resource management [12].

3 Dynamical Systems and Time Series

Attractors and fractal dimensions are concepts from dynamical systems theory and chaos theory, applicable to the analysis of complex systems such as 5G telecommunications networks. An attractor is a set of values toward which a system tends to evolve, regardless of

initial conditions. In phase space, attractors represent stable states or repetitive cycles, as seen in dissipative and periodic systems [31]. Strange attractors, characteristic of chaotic systems, exhibit sensitivity to initial conditions, where small differences can lead to divergent trajectories and unpredictable behaviors. Classic examples include the Lorenz-63 model and the Tinkerbell map, systems that generate complex patterns dependent on initial conditions [18].

Attractors have fractal dimensions (DF), which are often non-integer, reflecting the self-similar and complex nature of the generated patterns. This attribute indicates the presence of detailed patterns that repeat at all scales within the system [27]. To determine the fractal dimension, the attractor's point set is overlaid with a grid of hypercubes, and the cubes containing at least one point are counted. The fractal dimension DF is calculated as the ratio of the logarithm of the number of cubes to the logarithm of the inverse of the cube's side length [28, 17]:

$$D_F = \lim_{\epsilon \to 0} \frac{\log[N(\epsilon)]}{\log(1/\epsilon)} \ . \tag{1}$$

Embedding dimensions and the method of time delays are techniques used in the analysis of dynamical systems and time series to reconstruct the state space of a system from observable data. Embedding dimensions refer to the number of dimensions necessary to fully represent a system's dynamics without overlaps. The method of time delays involves creating a multidimensional phase space using time-delayed copies of a time series, allowing the reconstruction of the original phase space of the dynamical system. Whitney's embedding theorem ensures that a curve in a D-dimensional space can be uniquely mapped into a $2D + 1$ dimensional space, guaranteeing a non-intersecting representation [46, 35]. Floris Takens, in 1981, demonstrated how to reconstruct a dynamical system from a single time series using time delay coordinates, creating a space topologically equivalent to the original phase space [37]. In Ref. [36] this approach was expanded for attractors with fractal dimensions, determining that the embedding dimension (d) must be greater than twice the fractal dimension of the attractor ($d > 2 \times D_F$).

The False Nearest Neighbours (FNN) method is used to determine the appropriate embedding dimension for reconstructing the state space of a dynamical system. This technique identifies whether the proximity between points in a reconstructed state space is real or results from projecting a higher-dimensional space into a lower-dimensional one. FNN analyzes how the relative distances between points change as the embedding dimension increases. If nearby points in a lower-dimensional space move significantly apart when the dimensionality is increased, they are considered false neighbours. The correct embedding dimension is found when the ratio of false neighbours becomes zero, ensuring an accurate and non-intersecting representation in the phase space [31, 10],

$$\sqrt{\frac{R_{m+1}^2(n) - R_m^2(n)}{R_m^2(n)}} > L_{\text{critical}} \ , \tag{2}$$

where R_m is the distance between a point and its nearest neighbour and m is the embedding dimension.

4 Kalman Filters

Developed in 1960 by Rudolf Emil Kalman, the KF is a set of mathematical equations that estimate the state of a linear dynamic system from incomplete and noisy measurements. This algorithm has become a fundamental tool in signal processing, control systems, and navigation, allowing precise real-time prediction and correction of system states. The KF provides estimates and predictions divided into two phases: prediction and update. In the prediction phase, state variables are estimated based on the system dynamics, while the update phase refines these estimates by incorporating new observational data. The state estimate is sequential, where the state at time k, denoted as x_k, is predicted from the previous state x_{k-1} using the system's state transition matrix A and the process noise w_{k-1} with covariance given by the matrix Q. The observational data are represented by the measurement equation, where H maps the true state space to the observed space, and v_k denotes the measurement noise [23]

$$x_k = A\, x_{k-1} + \omega_{k-1} \ , \tag{3}$$

$$z_k = H\, x_k + v_k \ . \tag{4}$$

The update phase begins with the previous state $k-1$, providing initial estimates for $\hat{x}_{k-1}$ and P_{k-1}. The a priori estimates $\hat{x}_{\underline{k}}$ and the a posteriori estimates $\hat{x}_k$, after the measurement update, are recalculated along with the error covariance $P_{\underline{k}}$ and P_k, respectively. The Kalman gain K is derived during the correction step to minimize the estimation error, leading to the updated state estimate and the adjusted error covariance [23]:

$$\begin{aligned} \hat{x}_{\underline{k}} &= A\, \hat{x}_{k-1} \ , \\ P_{\underline{k}} &= A\, P_{k-1}\, A^T + Q \ , \qquad (5) \\ K_k &= P_{\underline{k}}\, H^T (H\, P_{\underline{k}}\, H^T)^{-1} \\ \hat{x}_k &= \hat{x}_{\underline{k}} + K_k\, (z_k - H\, \hat{x}_{\underline{k}}) \\ P_k &= (I - K_k H)\, P_{\underline{k}} \ . \qquad (6) \end{aligned}$$

The Extended-Kalman Filter (EKF) adapts Kalman's original theory for nonlinear systems by using Taylor series approximations to linearize models around the current estimate. This linearization allows the EKF to apply the iterative prediction and correction process of the KF to nonlinear systems. In the prediction phase, the nonlinear state dynamics and measurement models are approximated using the first-order Taylor series expansion, resulting in linearized models that approximate the system's behavior near the current state estimate. The state transition and measurement equations are rewritten for nonlinear processes, with noise components following normal distributions with zero mean and covariances Q and R. While effective for systems that can be well-approximated by linearization, the EKF may exhibit inaccuracies in highly nonlinear systems and requires numerical methods to compute the Jacobian matrices [44].

The Unscented-Kalman Filter (UKF) was developed to overcome the limitations of the EKF in nonlinear systems by avoiding the complexities of linearization through the Unscented Transformation (UT). The UKF uses strategically chosen sigma points to capture the mean and covariance of the probability distribution more accurately, offering a more reliable estimate of the system's statistical characteristics [22, 45].

The KTF is a hybrid method based on the UKF, replacing model equations with a local model constructed using time-delay vectors. It appeared originally in Ref. [19] to denoise a set of Lorentz-96 equations [26]. This model relates the observations z_k to the states x_k at time k, incorporating the noise ν, calculating the value of x_{K+1} using a function $\tilde{f}(x_k,\ t_k)$ as follows:

$$\begin{aligned} x_{k+1} &= \tilde{f}(x_k,\ t_k)\ +\ w_k\ , \\ z_k &= h(x_k,\ t_k)\ +\ v_k\ . \end{aligned} \tag{7}$$

The function $\tilde{f}(x_k,\ t_k)$ is built using the past information of the system via its time-series. Using an embedding dimension d, the time-delay vector is constructed with a step of 1 for the observed times

$$\xi_k(T) = [\ z_k,\ z_{k-1}, \cdots,\ z_{k-(d-1)}\ ]\ . \tag{8}$$

For non-parametric prediction, a sequence of data collected over the interval $[1,\ 2, \cdots,\ Tt]$ is used to form a training sequence and predict $[T\ +\ 1, T\ +\ 2, ..., T_f]$, where T_f is the number of observations to be predicted, i.e.,

$$\begin{aligned} \xi_k(T') &= [\ z_k(T'),\ z_{k-1}(T'), \cdots,\ z_{k-(d-1)}(T')\ ] \\ \xi_k(T'') &= [\ z_k(T''),\ z_{k-1}(T''), \cdots,\ z_{k-(d-1)}(T'')\ ] \\ &\ \ \vdots \\ \xi_k(T^k) &= [\ z_k(T^k),\ z_{k-1}(T^k), \cdots,\ z_{k-(d-1)}(T^k)\ ] \end{aligned} \tag{9}$$

Once the vectors are obtained, they are advanced in time $z(T) \rightarrow z(T+i)$ for each calculated neighbor. The average of the time-advanced vectors gives the non-parametric model for the system's evolution

$$\begin{aligned} \tilde{f}(\xi_k\ +\ i) &= w'\xi_k(T'+i)\ +\ w''\xi_k(T''+i) \\ &+\ \cdots\ +\ w_k\xi_k(T^k+i)\ , \end{aligned} \tag{10}$$

where the time-delay vector's weight, or the average of the distances is given by

$$w^{(i)} = \frac{e^{-d^{(i)}/\sigma}}{\sum_{j=1}^{k} e^{-d^{(i)}/\sigma}}\ . \tag{11}$$

Therefore. the KTF does not rely in a determined function or a transition matrix to estimate the x_{k+1}. Therefore, it is a "model free" predictor built using techniques from chaos and dynamic systems.

5 Case Study

5.1 Data

The numerical data used in this work were collected from a mobile network operator in South Korea, using a Samsung Galaxy A90 5G smartphone equipped with a Qualcomm Snapdragon X50 5G modem. For packet capture, the PCAPdroid application was used, allowing direct data collection from mobile terminals without specialized equipment. Data collection was conducted sequentially for each application on fixed terminals, with no background traffic, ensuring data accuracy [13].

The dataset covers various types of traffic in 5G networks, with a special focus on intensive video traffic, which is crucial for network planning and management due to its high bandwidth consumption. Video traffic constitutes about 73% of mobile data traffic in 2023 and is projected to grow with the expansion of 5G networks [16]. The data, collected from May to October 2022, totals 328 hours and is in CSV format, including records mapped by timestamp with details such as source and destination addresses [13].

Specific subsets were selected for analysis: AfreecaTV for live streaming traffic, MS Teams for

video conferencing, and YouTube for video streaming. These subsets are relevant for daily activities and traffic behavior analysis. The initial data processing involved filtering and cleaning to ensure analytical accuracy, including the elimination of incomplete records and error correction. The analysis focused on the time and length columns of the collected data, providing a detailed view of bandwidth usage and traffic dynamics in the 5G network. These insights are essential for applying the KTF to analyze traffic dynamics [16].

5.2 Applying KTF to data

The KTF was applied to the time series data to model and predict network traffic fluctuations, identifying patterns, trends, and irregularities. Before applying the KTF, the CSV files underwent preprocessing to ensure temporal uniformity, adjusting the time series to regular intervals. Specific retime intervals were applied: 5 seconds for AfreecaTV, 0.5 seconds for MS Teams, and 10 seconds for YouTube, due to the variability of the collected data.

The KTF was parameterized for 5G data analysis, including the definition of variables, state and observation parameters, and fine-tuning the filter parameters to optimize temporal analysis accuracy. The application of the KTF used the Root Mean Square (RMS) Error (RMSE) as the performance metric. RMSE is a standard statistical measure used to assess the accuracy of model predictions, quantifying the difference between predicted values and actual observed values. RMSE is calculated as the square root of the mean of the squared differences between predicted and observed values, penalizing larger errors more severely

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_{oi} - z_{p_i})^2}\,, \qquad (12)$$

where N is the number of observations, z_o is the actual observed value and z_p is the predicted value [32].

The Normalized Root Mean Square Error (nRMSE),

$$nRMSE = \frac{RMSE}{|\max(z_o) - \min(z_o)|}\,, \qquad (13)$$

where the denominator is the range of the observed data [32], assesses model prediction accuracy, normalized by the range or standard deviation of the observed data, facilitating performance comparison across different datasets or scales.

The KTF parameters, such as k-Nearest Neighbors (kNN), Transient Filter, Delay, Number of State Variables (N), Filter-Row-From, and Filter-Row-To, impact predictive accuracy. Each parameter in the KTF configuration affects the RMSE differently. The Transient Filter controls the number of initial data points ignored in the RMSE calculation, excluding initial fluctuations or noise. The delay is the time interval used in Takens' embedding to reconstruct the phase space from a single time series, influencing the state estimate accuracy. The kNN determines how many neighbors are used in the prediction, balancing bias and variance in the predictions. N represents the state dimensionality, and an appropriate N captures all relevant variables without redundancy. Filter-Row-From and Filter-Row-To are also parameters that define the segment of data used in the analysis, focusing on the most relevant parts of the time series. Correctly adjusting these parameters can minimize RMSE and optimize KTF performance for a specific dataset, ensuring accurate analysis of traffic behavior in the 5G network. In the next section we show how we used the KTF to predict the packet sizes.

6 Numerical Results

The initial evaluation of the KTF involved a meticulous analysis of parameters to assess the predictive efficiency of the filter in the dynamics of 5G network traffic. A MATLAB script was developed to automate the execution of the KTF under various parameter configurations, aiming to minimize the RMSE and enhance model accuracy. This script explored different settings for the Transient Filter, Delay, kNN, N, Filter-Row-From, and Filter-Row-To parameters, ensuring a thorough evaluation of their influence on the predictive accuracy of the KTF.

The data analysis identified parameter configurations that resulted in the lowest RMSE, providing deeper insights into the model's effectiveness. Parameters such as a higher Transient Filter value, lower Delay values, and a moderate kNN number were found to be most suitable for the analyzed data. The relationships between parameters and RMSE were highlighted: higher Transient Filter values were associated with RMSE reduction, indicating that ignoring initial observations can improve prediction accuracy. Delay values showed no clear linear pattern, suggesting that optimal values should be empirically determined; the impact of kNN did not follow a uniform linear pattern, indicating the need to balance capturing relevant details and managing data variability.

These results demonstrate that the ideal set of parameters is specific to the data context and dependent on the system dynamics being studied. Therefore, careful calibration of the Transient Filter, Delay, and kNN

parameters is crucial to optimize KTF accuracy. Experimenting with a range of parameter values and subsequently evaluating their impact on RMSE proved to be an effective strategy for determining the most efficient KTF configuration.

In the following subsections we show the performance of the KTF. We compare the observations, i.e., measurements with the predictions obtained using the KTF.

6.1 AfreecaTV Series

The analysis of KTF predictions applied to the AfreecaTV dataset, shown in Fig. 1, focused on the model's effectiveness in capturing the dynamics of 5G network traffic. Various parameter configurations were explored to minimize the RMS error and maximize prediction accuracy. It was observed that increases in the Transient Filter value generally reduce the RMS error, indicating that eliminating initial observations can discard noise and stabilize predictions, especially when kNN is set between 10 and 50. A Delay configuration of 2 resulted in a substantially lower average RMS error compared to Delays of 5 and 10, suggesting that a smaller Delay is preferable for this dataset.

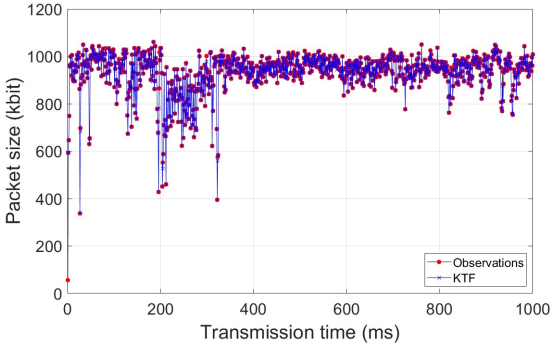

Figure 1: Packet size vs transmission time in the case of the AfreecaTV time series. Observations in red and Kalman-Takens filtering in blue.

The variation in RMSE with different kNN configurations revealed a non-linear behavior, where an intermediate kNN, specifically 10, was ideal for minimizing RMSE. Additionally, RMSE was significantly higher when the analysis started at the first line compared to starting at line zero, suggesting that including initial data can be beneficial. Changes in Filter-Row-To from 100 to 1000 did not result in significant RMSE changes, indicating that extending the analysis to more lines does not drastically affect accuracy.

The optimal configuration identified, with a Transient Filter of 1000, Delay of 2, kNN of 1, N of 1, Filter-Row-From of 1, and Filter-Row-To of 1000, resulted in a significantly low RMSE of approximately 0.061, demonstrating its effectiveness in predictive modeling for live streaming traffic on AfreecaTV in a 5G network. This configuration showed that a high Transient Filter eliminates distorting initial data, a reduced Delay ensures predictions influenced by recent observations, and a single kNN highlights the importance of precise historical analysis.

6.2 MS Teams Series

The analysis of MS Teams videoconference data, displayed in Fig. 2, highlights significant parameter influences on RMSE, demonstrating how appropriate configuration can enhance predictive modeling. RMSE consistently decreases as the Transient Filter increases from 1 to 1000, indicating the effectiveness of excluding initial observations to improve prediction accuracy. However, exceeding a Transient Filter of 1000 does not result in significant additional RMSE improvement, suggesting a saturation point. The kNN parameter shows a non-linear relationship with RMSE, with a kNN of 10 being ideal, achieving an RMSE of 39.50, while extremes in the number of neighbors are counterproductive.

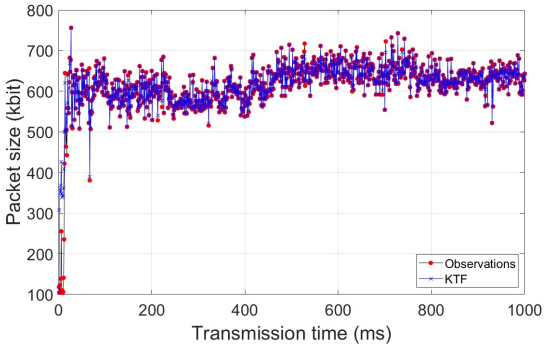

Figure 2: Packet size vs transmission time in the case of the MS Teams time series. Observations in red and Kalman-Takens filtering in blue.

The influence of Delay is evidenced by the significant RMSE reduction, reaching the lowest value of 40.29 with a Delay of 2. Larger Delays, such as 5 and 10, increase RMSE to 59.48 and 67.39, respectively, demonstrating that smaller Delays are preferable as they capture essential temporal dependencies without redundancy. The Filter-Row-From and Filter-Row-To configuration shows an impact, where starting the

analysis from zero and extending up to 1000 lines improves RMSE. Analysis starting from zero presents a lower RMSE of 32.38, compared to starting from one with 87.71.

For the MS Teams series, the best configuration achieved was with a Transient Filter of 1000, Delay of 2, kNN of 10, N of 1, Filter-Row-From of 1, and Filter-Row-To of 1000, reaching an RMSE of 0.128, the lowest observed. Implementing a Transient Filter of 1000 removes initial observations that may distort modeling, promoting a more stable and representative model. A reduced Delay ensures predictions are based on recent observations, and a kNN of 10 proves to be most effective, balancing noise excess and information scarcity. The careful selection of the data interval for analysis, ensuring the inclusion of the most representative data, is essential to optimize prediction accuracy in 5G network traffic.

6.3 YouTube Series

The analysis of YouTube data reveals that RMSE decreases with the increment of the Transient Filter as shown in Fig. 3, suggesting that increasing the filter value helps improve model accuracy by removing initial observations. It was observed that the kNN value has a non-linear relationship with RMSE, with kNN of 10 resulting in lower RMSE, unlike kNN of 1 and 100, which showed higher RMSE. The Delay parameter also significantly affects RMSE, with a Delay of 2 resulting in a lower average RMSE, while Delays of 5 and 10 increase RMSE.

The Filter-Row-From and Filter-Row-To parameters influence RMSE, where starting the analysis from point 0 results in a lower RMSE compared to starting from point 1, suggesting that including initial data can be beneficial. The prediction error gradation and the RMS error histogram displays the variability and distribution of RMSE results, reflecting the extensive experimentation necessary to capture the nuances of YouTube traffic.

The configuration with a Transient Filter of 1000, Delay of 2, kNN of 1, N of 1, Filter-Row-From of 1, and Filter-Row-To of 1000 resulted in the lowest observed RMSE, standing out as the most efficient for minimizing the RMS error. A Transient Filter of 1000, a Delay of 2, and a kNN of 1 shows that an immediate approximation of temporal data and using the most similar previous state as a reference are essential to achieving maximum prediction accuracy. Starting the analysis from line 1 to line 1000 allowed for considering the most relevant data, optimizing prediction accuracy and indicating that careful parameter adjustments can significantly improve data traffic modeling in 5G networks.

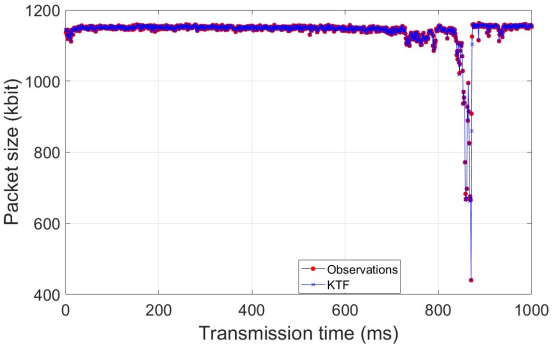

Figure 3: Packet size vs transmission time in the case of the YouTube time series. Observations in red and Kalman-Takens filtering in blue.

7 Conclusions

The aim of this work was to analyze and validate a predictive model by applying the KTF to optimize spectrum allocation in 5G mobile communication systems, in order to improve network performance and user experience. The literature set the basis for the application of the KTF in modeling and forecasting 5G time series data. Related works indicated the relevance of predictive resource allocation and the application of the KF to estimate future transmission rates for scheduler use.

The developed methods enabled the application of the KTF in modeling and forecasting traffic variations, using metrics such as the RMSE to evaluate the model's accuracy. The results demonstrated the effectiveness of the KTF in capturing network traffic dynamics, highlighting the importance of careful parameter calibration to achieve highly accurate predictions. The analysis revealed that parameters such as Transient Filter, Delay, and kNN directly influence RMSE, evidencing the need for specific adjustments for each dataset.

In conclusion, this work showed that fine-tuning the KTF involves a large number of variables, requiring a detailed understanding of their collective dynamics. The optimal parameter configuration achieved the lowest RMS error, underscoring the importance of a holistic approach to predictive modeling. This knowledge is crucial for creating accurate and reliable predictive models, essential for an efficient distribution of network resources amid increasing data consumption.

References

[1] 3GPP. Nr and ng-ran overall description. Technical Specification (TS) 38.300, 3rd Generation

Partnership Project (3GPP), 03 2023. Version 17.6.0.

[2] 3GPP. Physical channels and modulation. Technical Specification (TS) 38.211, 3rd Generation Partnership Project (3GPP), 03 2023. Version 18.0.0.

[3] 3GPP. Physical layer procedures for data. Technical Specification (TS) 38.214, 3rd Generation Partnership Project (3GPP), 03 2023. Version 18.0.0.

[4] 3GPP. Radio resource control (rrc) - protocol specification. Technical Specification (TS) 38.331, 3rd Generation Partnership Project (3GPP), 03 2023. Version 17.6.0.

[5] 3GPP. System architecture for the 5g system (5gs). Technical Specification (TS) 23.501, 3rd Generation Partnership Project (3GPP), 03 2023. Version 18.4.0.

[6] 3GPP. User equipment (ue) radio transmission and reception - part 1: Range 1 standalone. Technical Specification (TS) 38.101, 3rd Generation Partnership Project (3GPP), 03 2023. Version 18.3.0.

[7] Alamu, O., Iyaomolere, B., and Abdulrahman, A. An overview of massive mimo localization techniques in wireless cellular networks: Recent advances and outlook. *Ad Hoc Networks*, 111:102353, 2021.

[8] Alraih, S., Shayea, I., Behjati, M., Nordin, R., Abdullah, N. F., Abu-Samah, A., and Nandi, D. Revolution or evolution? technical requirements and considerations towards 6g mobile communications. *Sensors*, 22(3), 2022.

[9] Atawia, R., Abou-zeid, H., Hassanein, H. S., and Noureldin, A. Joint chance-constrained predictive resource allocation for energy-efficient video streaming. *IEEE Journal on Selected Areas in Communications*, 34(5):1389–1404, 2016.

[10] Bezruchko, B. P. and Smirnov, D. A. *Extracting Knowledge From Time Series: An Introduction to Nonlinear Empirical Modeling*. Springer, Heidelberg, 2010.

[11] Bui, N., Cesana, M., Hosseini, S. A., Liao, Q., Malanchini, I., and Widmer, J. A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques. *IEEE Communications Surveys and Tutorials*, 19(3):1790–1821, 2017.

[12] Capozzi, F., Piro, G., Grieco, L., Boggia, G., and Camarda, P. Downlink packet scheduling in lte cellular networks: Key design issues and a survey. *IEEE Communications Surveys and Tutorials*, 15(2):678–700, 2013.

[13] Choi, Y.-H., Kim, D., Ko, M., Cheon, K.-y., Park, S., Kim, Y., and Yoon, H. Ml-based 5g traffic generation for practical simulations using open datasets. *IEEE Communications Magazine*, 61(9):130–136, 2023.

[14] Costa, A., Pacheco, L. P., Rosário, D., Villas, L., Loureiro, A., Sargento, S., and Cerqueira, E. Skipping-based handover algorithm for video distribution over ultra-dense vanet. *Computer Networks*, 176(1):107252–107252, July 2020.

[15] Dangi, R., Lalwani, P., Choudhary, G., You, I., and Pau, G. Study and investigation on 5g technology: A systematic review. *Sensors*, 22(1), 2022.

[16] Ericsson. *Ericsson Mobility Report November 2023*.

[17] Falconer, K. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, Chichester, 2014.

[18] H. E. Nusse, J. A. Y. *Dynamics: Numerical Explorations: Accompanying Computer Program Dynamics*. Springer, New York, 2012.

[19] Hamilton, F., Berry, T., and Sauer, T. Ensemble kalman filtering without a model. *Phys. Rev. X*, 6:011021, Mar 2016.

[20] Hamilton, F., Lloyd, A. L., and Flores, K. B. Hybrid modeling and prediction of dynamical systems. *PLOS Comput. Biol.*, 13:1005655, 2017.

[21] Johnson, C. The national security benefits of reallocating federal spectrum for 5g. 2023.

[22] Julier, S. and Uhlmann, J. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[23] Kalman, R. E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960.

[24] Khaki, M., Hoteit, I., Kuhn, M., Forootan, E., and Awange, J. Assessing data assimilation frameworks for using multi-mission satellite products in a hydrological context. *Sci. Total Environ.*, 647:1031–1043, 2019.

[25] Leopoldseder, M., Svoboda, P., Eller, L., and Rupp, M. Benchmarking lightweight user mobility predictors on operational wlan data. In *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pages 1–5, 2019.

[26] Lorenz, E. N. Predictability: a problem partly solved. *Semin. Predict. Shinfield Park Read.*, 1, 1996.

[27] Mandelbrot, B. B. *The fractal geometry of nature - Revised edition*. W. H. Freeman, San Francisco, 1982.

[28] McCauley, J. L. An introduction to nonlinear dynamics and chaos theory. *Phys. Scr. Vol. T*, 20:1–57, 1988.

[29] Moreira, M. A., Teixeira, M. J., and Salvador Timóteo, V. Kalman-takens filtering in communication systems. *INFOCOMP Journal of Computer Science*, 23(1), July 2024.

[30] Morocho-Cayamcela, M. E., Lee, H., and Lim, W. Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions. *IEEE Access*, 7:137184–137206, 2019.

[31] N. Fiedler-Ferrara, C. P. C. d. P. *Caos: uma introdução*. Edgar Bl'ucher, São Paulo, 1994.

[32] R. J. Hyndman, G. A. *Forecasting: principles and practice*.

[33] Sharma, A., Vanjani, P., Paliwal, N., Basnayaka, C. M., Jayakody, D. N. K., Wang, H.-C., and Muthuchidambaranathan, P. Communication and networking technologies for uavs: A survey. *Journal of Network and Computer Applications*, 168:102739, 2020.

[34] Shi, H. and Li, Y. Discovering periodic patterns for large scale mobile traffic data: Method and applications. *IEEE Transactions on Mobile Computing*, 17(10):2266–2278, 2018.

[35] Sprott, J. C. *Chaos and Time-series Analysis*. Oxford University Press, Oxford, 2003.

[36] T. Sauer, M. C., J. A. York. Embedology. *Journal of Statistical Physics*, 65(3), November 1991.

[37] Takens, F. Detecting strange attractors in turbulence. In Rand, D. and Young, L.-S., editors, *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.

[38] Teixeira, M. J. *Alocação Preditiva de Recursos em Sistemas de Comunicações Móveis*. Phd thesis, State University of Campinas, Limeira, SP, November 2021. Available at `https://repositorio.unicamp.br/Acervo/Detalhe/1231904`.

[39] Teixeira, M. J. and Salvador Timóteo, V. Using a kalman filter to improve schedulers performance in mobile networks. In *2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 853–858, 2019.

[40] Teixeira, T. V. S., M. J. A predictive resource allocation for wireless communications systems. *SN Computer Science*, 2(6), September 2021.

[41] Teixeira, T. V. S., M. J. Model-free predictor of signal-to-noise ratios for mobile communications systems. *SN Computer Science*, 4(4), April 2023.

[42] Tomás, J. P. Brazilian regulator allocates 2 mhz in the 4 ghz band for 5g. 2023.

[43] Tomás, J. P. China ends 2023 with 3.38 million 5g base stations. 2023.

[44] Urrea, C. and Agramonte, R. Kalman filter: Historical overview and review of its use in robotics 60 years after its creation. *Journal of Sensors*, 2021(1):9674015, 2021.

[45] Wan, E. and Van Der Merwe, R. The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pages 153–158, 2000.

[46] Whitney, H. Differentiable manifolds. *Annals of Mathematics*, 37(3):645–680, 1936.