Channel Prediction Using Deep Recurrent Neural Network With EVT-Based Adaptive Quantile Loss Function

Niloofar Mehrnia, Parmida Valiahdi, Sinem Coleri, *Fellow, IEEE*, and James Gross, *Senior Member, IEEE*

***Abstract*— Ultra-reliable low latency communication (URLLC) systems are pivotal for applications demanding high reliability and low latency, such as autonomous vehicles. In such contexts, channel prediction becomes essential to maintaining communication quality, as it allows the system to anticipate and mitigate the effects of fast-fading channels, thereby reducing the risk of packet loss and latency spikes. This letter presents a novel framework that integrates neural networks with extreme value theory (EVT) to enhance channel prediction, focusing on predicting extreme channel events that challenge URLLC performance. We propose an EVT-based adaptive quantile loss function that integrates EVT into the loss function of the deep recurrent neural networks (DRNNs) with gated recurrent units (GRUs) to predict extreme channel conditions efficiently. The numerical results indicate that the proposed GRU model, utilizing the EVT-based adaptive quantile loss function, significantly outperforms the traditional GRU. It predicts a tail portion of 7.26%, which closely aligns with the empirical 7.49%, while the traditional GRU model only predicts 2.4%. This demonstrates the superior capability of the proposed model in capturing tail values that are critical for URLLC systems.**

***Index Terms*— Channel prediction, deep recurrent neural network, extreme value theory, URLLC.**

I. Introduction

ACHIEVING ultra-reliable low latency communication (URLLC) in mission-critical applications necessitates effective real-time channel prediction, especially when dealing with extreme events, i.e., very low values happening infrequently, to mitigate the effects of fast-fading channels. Accurate channel predictions enable dynamic adjustments to modulation and coding scheme (MCS) selection, optimizing throughput and minimizing delays associated with channel quality indicators (CQI) for proactive tasks. This improves resource allocation efficiency in 5G and beyond [1]. URLLC ensures the delivery of critical information with a packet error rate ranging from 10^{-9} to 10^{-5}, and delays on the order of milliseconds. Neural networks are promising for achieving high prediction accuracy through the processing of extensive real-time data, thereby enhancing URLLC performance. To refine these predictions further, integrating extreme value theory (EVT) with neural network techniques can effectively analyze rare fading events that affect ultra-reliable communications. By concentrating on the lower tail of signal power distributions, EVT offers precise modeling of wireless channels under extreme conditions. Therefore, the combination of EVT and neural networks represents a robust strategy for improving channel prediction in URLLC systems.

Several advanced neural network architectures have been proposed to improve channel prediction accuracy [2], [3], [4]. Recurrent neural networks (RNNs) leverage temporal correlations in sequential data for channel prediction [4], unlike convolutional neural networks (CNNs) that use matrices as input, thereby ignoring temporal correlations [2]. However, RNNs face challenges like vanishing or exploding gradients, which have been mitigated by using long short-term memory (LSTM) or gated recurrent units (GRU) [3], [4]. In [5], deep RNNs (DRNNs) combined with LSTM or GRU have been proposed to predict CQI under URLLC latency constraints, improving reliability without reducing data rates. Despite these advancements, most studies focus on average channel prediction and fail to address extreme events critical for URLLC.

EVT provides a robust framework for modeling extreme events by leveraging the generalized Pareto distribution (GPD) to estimate tail statistics, crucial for ultra-reliability [6], [7], [8]. EVT has been applied to wireless channel modeling [6], including extensions like multivariate EVT (MEVT) for spatial diversity [7], and queue-aware resource allocation [8]. However, these studies focus on tail estimation using maximum likelihood estimation (MLE) and assume large data availability. Recently, [9] proposed a generative adversarial network (GAN)-based GPD parameter estimator to accurately estimate GPD parameters. GAN has consistently demonstrated superior performance compared to MLE in estimating the probability density function (PDF) of extreme events, especially with limited sample sizes. Nonetheless, none of these works consider real-time channel prediction of both average statistics and extreme values, crucial for URLLC performance.

In this letter, we propose a novel channel prediction framework incorporating EVT into DRNN to address the reliability and latency constraints of URLLC. Unlike previous DRNN applications focusing on frequent average statistic predictions, our approach leverages EVT-based adaptive quantile loss function to predict extreme values over the next few milliseconds. Specifically, we use DRNN with GRU, which is computationally more efficient than DRNN with LSTM for channel prediction and holds promise for future real-world implementations [5]. To show the superiority of the proposed scheme, we use real channel measurements to ensure extreme values are detected over the next few milliseconds.

Received 15 April 2025; accepted 13 May 2025. Date of publication 20 May 2025; date of current version 14 July 2025. Sinem Coleri acknowledges the support of the Scientific and Technological Research Council of Turkey 2247-A National Leaders Research Grant #121C314. The associate editor coordinating the review of this letter and approving it for publication was L. Fu. *(Corresponding author: Niloofar Mehrnia.)*

Niloofar Mehrnia and James Gross are with the Department of Electrical Engineering and Computer Sciences, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: nilome@kth.se; jamesgr@kth.se).

Parmida Valiahdi and Sinem Coleri are with the Department of Electrical and Electronics Engineering, Koc University, 34450 Istanbul, Türkiye (e-mail: pvaliahdi23@ku.edu.tr; scoleri@ku.edu.tr).

Digital Object Identifier 10.1109/LCOMM.2025.3571930

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

II. System Model

A. System Model Assumptions

We examine a real-time URLLC system designed to predict current and future received power samples based on historical data, particularly under conditions of severe fading that result in extremely low received power levels. In this context, the transmitter (Tx) transmits packets to a receiver (Rx) at a constant power over an unknown channel. As a result, the received signal power can be interpreted as the squared amplitude of the channel state information. Initially, the receiver focuses on estimating the channel, emphasizing extreme values. This estimation process can utilize either dedicated pilot signals and training sequences or the historical data from previous transmissions along with corresponding feedback. The Rx gathers channel measurements and subsequently estimates the channel tail distribution. This estimated distribution serves as prior knowledge, which is then relayed back to the Tx to facilitate channel prediction using a DRNN with GRU architecture.

B. Channel Prediction

The communication channel between Tx and Rx is characterized as a time-varying fading channel. Let $\boldsymbol{x}_t$ represent the transmitted signal at the t-th time instant, and denote the channel fade coefficient at the same instant by $\boldsymbol{h}_t$. The received signal at the receiver is denoted as $\boldsymbol{y}_t$, while $\boldsymbol{n}_t$ represents the additive noise. Consequently, the relationship can be expressed as follows $\boldsymbol{y}_t = \boldsymbol{h}_t\, \boldsymbol{x}_t + \boldsymbol{n}_t$. The channel coefficients $\boldsymbol{h}_t$ are statistically modeled using the pilot signals or previous data transmissions. Feedback and processing delays can render the channel coefficients obtained at the Tx outdated by the time they are utilized, leading to significant performance degradation in adaptive transmission systems. Channel prediction aims to determine a predicted value, $\hat{\boldsymbol{h}}_{t+\tau}$, at time t in Tx side that approximates its actual value at the upcoming time $t+\tau$ as close as possible, i.e., $\hat{\boldsymbol{h}}_{t+\tau} \approx \boldsymbol{h}_{t+\tau}$.

The time-varying channel coefficients exhibit temporal correlation. By modeling this correlation, the number of pilot signals required for estimating these coefficients can be reduced. The learned model can then be applied for channel prediction. However, since the correlation model may vary across different channel geometries, a purely statistical approach may lack robustness. To address this issue, we propose a deep learning-based solution to learn the channel correlation model and predict future channel coefficients, thereby minimizing the need for pilot transmissions. Specifically, we utilize a DRNN with GRU architecture to develop the proposed channel predictor. In the subsequent subsection, we provide a brief overview of DRNN, with a particular focus on the GRU component.

C. Deep Recurrent Neural Network for Channel Prediction

In the training phase, the system model uses DRNN with GRU to effectively capture temporal dependencies in time series data. Then, in the operation phase, it uses the pre-trained model to predict the current and future received power values, especially the extreme ones. At each time step $t \in \{1, \ldots, T\}$, the output of GRU, $\hat{\boldsymbol{y}}_t$, is computed as

$$\hat{\boldsymbol{y}}_t = g(\boldsymbol{W}_y \boldsymbol{h}_t + \boldsymbol{b}_y), \tag{1}$$

where $\boldsymbol{b}_y$ is the bias vector in the output layer and g is the output activation function and is given by

$$g(\boldsymbol{k}) = \max\{\Re[\boldsymbol{k}], 0\} + j \max\{\Im[\boldsymbol{k}], o\}, \tag{2}$$

with $\Re[.]$ and $\Im[.]$ being the real and imaginary parts of the vectors, respectively. In (1), the hidden units, $\boldsymbol{h}_t$, is connected to the output layer, $\hat{\boldsymbol{y}}_t$, with the weighted connections $\boldsymbol{W}_y$. The GRU architecture consists of a total of two gates: a reset gate that combines the current input with previous memory and an update gate, that determines the percentage of previous state to remember, accompanied by a single hidden state. Within the GRU, each gate is equipped with a sigmoid activation function, while a separate "tanh" function is employed to generate the output, as depicted in Fig. 1. The hidden state $\boldsymbol{h}_t$ in (1) gets updated based on the current input $\boldsymbol{x}_t$ and the previous hidden state $\boldsymbol{h}_{t-1}$ as follows:

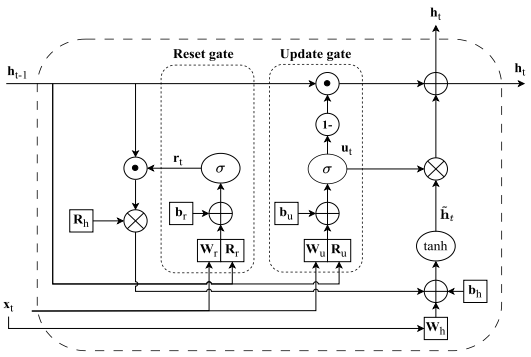


Fig. 1. The architecture of GRU.

$$\boldsymbol{r}_t = \sigma_g(\boldsymbol{W}_r \boldsymbol{x}_t + \boldsymbol{R}_r \boldsymbol{h}_{t-1} + \boldsymbol{b}_r), \tag{3a}$$

$$\boldsymbol{u}_t = \sigma_g(\boldsymbol{W}_u \boldsymbol{x}_t + \boldsymbol{R}_u \boldsymbol{h}_{t-1} + \boldsymbol{b}_u), \tag{3b}$$

$$\tilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{W}_h \boldsymbol{x}_t + \boldsymbol{R}_h(\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}) + \boldsymbol{b}_h), \tag{3c}$$

$$\boldsymbol{h}_t = (\boldsymbol{1} - \boldsymbol{u}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{u}_t \tilde{\boldsymbol{h}}_t, \tag{3d}$$

where $\boldsymbol{r}_t$ and $\boldsymbol{u}_t$ represent the reset gate and update gate for a time stamp t, respectively; $\boldsymbol{W}_r$ and $\boldsymbol{W}_u$ are the weight matrices for the input state at the reset gate and update gate, respectively; $\boldsymbol{R}_r$ and $\boldsymbol{R}_u$ represent the weight matrices for the hidden state at the reset gate and update gate, respectively; $\boldsymbol{b}_r$, $\boldsymbol{b}_u$, and $\boldsymbol{b}_h$ refer to the bias terms at the reset gate, update gate, and hidden state, respectively; and $\odot$ denotes Hadamard product of the matrix. $\boldsymbol{h}_{t-1}$ represents the previous hidden state of the GRU, and σ_g is the sigmoid activation function given by $\sigma_g(\boldsymbol{x}) = 1/(1+e^{-\boldsymbol{x}})$. To quantify the difference between the predicted values $\hat{\boldsymbol{y}}_t$ and the actual values $\boldsymbol{y}_t$, we need to define a loss function $\mathcal{L}(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)$. By employing a structured approach to weight updates based on feedback from the loss function, the model iteratively minimizes the loss such that

$$W^* = \underset{\boldsymbol{W},\boldsymbol{R},\boldsymbol{b}}{\arg\min}\ \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}). \tag{4}$$

Here, $\mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \sum_{t=1}^{T} \mathcal{L}(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)$ is an overall summation of losses in each timestep $\mathcal{L}(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)$, and is differentiable to find the optimal parameters $\boldsymbol{W} \in \{\boldsymbol{W}_r, \boldsymbol{W}_u\}$, $\boldsymbol{R} \in \{\boldsymbol{R}_r, \boldsymbol{R}_u\}$, and $\boldsymbol{b} \in \{\boldsymbol{b}_r, \boldsymbol{b}_u, \boldsymbol{b}_h\}$, leading to improved predictions. An iterative procedure persists until the minimum achievable error between the estimated value $\hat{\boldsymbol{y}}$ and the actual value $\boldsymbol{y}$ is attained. Common loss functions like mean-squared error (MSE) and mean absolute error (MAE) are widely focused on average error metrics prediction. Although DRNNs, including GRU, have demonstrated strong capabilities in predicting the next received power of a wireless channel, they often struggle with predicting extreme values due to their infrequent occurrence.

III. EVT-Based Adaptive Quantile Loss Function

In this section, we define an EVT-based adaptive quantile loss function using the estimated GPD parameters fitted to the channel tail distribution. This loss function provides better feedback to the DRNN with the GRU model for predicting extreme events.

A. EVT-Based Channel Estimation

In the channel tail estimation stage, the receiver applies EVT-based channel estimation following [6, Theorem 1] and uses the received power values to estimate the channel tail distribution. The channel measurements are converted into a sequence of n independent and identically distributed (i.i.d.) samples with cumulative distribution function (CDF) F by removing dependencies using declustering [6]. This sequence is denoted by $X^n = \{x_1, \ldots, x_n\}$, where x_i denotes the i^{th} sample, $i \in \{1, \ldots, n\}$. The GPD is then fitted to the received power values exceeding a threshold to model the channel tail distribution as:

$$F(z) = 1 - [1 + \frac{\xi z}{\tilde{\sigma}_u}]^{-1/\xi}, \tag{5}$$

where z is a non-negative value denoting the exceedance below threshold u, i.e., $z = u - X$ for $X = \{x_i < u\}$; ξ and $\tilde{\sigma}_u = \sigma + \xi(u - \mu)$ are the shape and scale parameters of the GPD, respectively; and μ and σ are the location and scale parameters of the generalized extreme value (GEV) distribution fitted to the CDF of $M_n = \max\{x_1, \ldots, x_n\}$ [6, Theorem 1], [10]. The optimal threshold u is determined using two complementary methods: mean residual life (MRL) and parameter stability methods [6], [10]. This optimum threshold distinguishes extreme values from non-extreme ones and depends on the data distribution. The estimation of Pareto distribution parameters is performed using the GAN-EVT method introduced in [9], which has been shown to outperform the GPD parameters estimation using MLE. We assume the distribution of received powers is stationary, meaning the parameters defining the GPD distribution remain constant. If non-stationarity is detected using the Augmented Dickey-Fuller (ADF) test, external factors influencing the GPD parameters are identified. Based on these factors, the dataset is then divided into smaller, stationary groups for further analysis. The GPD parameters are then estimated for the sequence over which the channel can be considered stationary, as explained in [11].

B. GRU With Adaptive Quantile Loss Function for Channel Prediction

In our EVT-based channel prediction framework, initially, the channel tail distribution is estimated using GPD. The estimated parameters are incorporated into the loss function of the GRU. This allows the model to adaptively adjust the desired quantile value, improving its ability to predict extreme values over the next few milliseconds.

The quantile loss function is a promising candidate for integrating the prediction of the extreme quantiles into neural networks [12]. The quantile loss function $\mathcal{L}_q(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t) = q(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)$ if $\boldsymbol{y}_t > \hat{\boldsymbol{y}}_t$ and otherwise, $\mathcal{L}_q(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t) = (1 - q)(\hat{\boldsymbol{y}}_t - \boldsymbol{y}_t)$, for a given quantile q [12]. The loss is asymmetric, penalizing overestimates and underestimates differently depending on q. Traditional quantile regression targets a specific quantile, which can lead to an overestimation of extreme quantiles [12]. In contrast, the proposed EVT-based adaptive quantile loss function adjusts the q value using the GPD parameters in (5) fitted to the channel tail distribution, enabling predictions for all values in the tail rather than a single quantile.

Let GPD(σ,ξ) be the Pareto model fitted to the training samples $X^n = \{x_1, x_2, \ldots, x_n\}$, $x_i < u$ for $i \in \{1, \ldots, n\}$, where σ and ξ are the scale and shape parameters of the GPD model fitted to the tail samples, respectively. Then, the EVT-based adaptive loss function of the GRU is defined as:

$$\mathcal{L}_\alpha(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t) = \begin{cases} \alpha(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t) & \text{if } \boldsymbol{y}_t > \hat{\boldsymbol{y}}_t, \ \ \boldsymbol{y}_t < u \\ (1 - \alpha)(\hat{\boldsymbol{y}}_t - \boldsymbol{y}_t) & \text{if } \boldsymbol{y}_t \leq \hat{\boldsymbol{y}}_t, \ \ \boldsymbol{y}_t < u \end{cases} \tag{6}$$

where $\alpha = F(u - \boldsymbol{y}_t) \Pr(\boldsymbol{y}_t < u)$. Here, $F(.)$ is the CDF of the GPD fitted to the channel tail distribution in (5), u is the optimum threshold separating extreme values from the near-average values, and $\Pr(\boldsymbol{y}_t < u)$ denotes the probability of observing $\boldsymbol{y}_t$ less than u. In the case of $\boldsymbol{y}_t > u$, the loss function is the traditional MSE-based loss. The objective is to find a prediction $\hat{\boldsymbol{y}}_t$ that minimizes the expected adaptive quantile loss, $\min \mathbb{E}[\mathcal{L}_\alpha(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)]$, where $\alpha = F(u - \boldsymbol{y}_t) \Pr(\boldsymbol{y}_t < u)$ for $\boldsymbol{y}_t < u$. The expected loss is given by

$$\begin{aligned} \mathbb{E}[\mathcal{L}_\alpha(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)] = &\int_{-\infty}^{\hat{\boldsymbol{y}}_t} (1 - \alpha)(\hat{\boldsymbol{y}}_t - \boldsymbol{y}_t) f(\boldsymbol{y}_t)\, d\boldsymbol{y}_t \\ &+ \int_{\hat{\boldsymbol{y}}_t}^{\infty} \alpha(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t) f(\boldsymbol{y}_t)\, d\boldsymbol{y}_t, \end{aligned} \tag{7}$$

where $\mathbb{E}[.]$ denotes the expectation operator and $f(\boldsymbol{y}_t)$ is the probability density function of $\boldsymbol{y}_t$. To proceed further, we need to show that (7) is convex. Therefore, we differentiate the function with respect to $\hat{\boldsymbol{y}}_t$ as follows:

$$\begin{aligned} \frac{d}{d\hat{\boldsymbol{y}}_t} \mathbb{E}[\mathcal{L}_\alpha(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)] = (1 - \alpha) &\int_{-\infty}^{\hat{\boldsymbol{y}}_t} f(\boldsymbol{y}_t)\, d\boldsymbol{y}_t \\ &- \alpha \int_{\hat{\boldsymbol{y}}_t}^{\infty} f(\boldsymbol{y}_t)\, d\boldsymbol{y}_t. \end{aligned} \tag{8}$$

Taking the second derivative of (7), we obtain

$$\frac{d^2}{d\hat{\boldsymbol{y}}_t^2} \mathbb{E}[\mathcal{L}_\alpha(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)] = (1 - \alpha) f(\hat{\boldsymbol{y}}_t) + \alpha f(\hat{\boldsymbol{y}}_t) = f(\hat{\boldsymbol{y}}_t). \tag{9}$$

Since the results of (9) is always non-negative, $\mathbb{E}[\mathcal{L}_\alpha(\boldsymbol{y}_t, \hat{\boldsymbol{y}}_t)]$ is convex and we can take its derivative with respect to $\hat{\boldsymbol{y}}_t$ and set it to zero, $(1-\alpha)F(\hat{\boldsymbol{y}}_t) - \alpha(1 - F(\hat{\boldsymbol{y}}_t)) = 0$, where $F(\hat{\boldsymbol{y}}_t) = \int_{-\infty}^{\hat{\boldsymbol{y}}_t} f(\boldsymbol{y}_t)\, d\boldsymbol{y}_t$ is the CDF of $\boldsymbol{y}_t$. Since $\alpha = F(u - \boldsymbol{y}_t)\Pr(\boldsymbol{y}_t < u)$, simplifying this equation results in:

$$F(\hat{\boldsymbol{y}}_t) = F(u - \boldsymbol{y}_t)\Pr(\boldsymbol{y}_t < u). \tag{10}$$

Therefore, minimizing the EVT-based adaptive quantile loss function results in a prediction $\hat{\boldsymbol{y}}_t$ that corresponds to the $F(u - \boldsymbol{y}_t)\Pr(\boldsymbol{y}_t < u)$-th quantile of the distribution. The intuition is that, for the tail values, if the prediction $\hat{\boldsymbol{y}}_t$ underestimates the observed value $\boldsymbol{y}_t$, the penalty should be proportional to $F(u - \boldsymbol{y}_t)\Pr(\boldsymbol{y}_t < u)$. Otherwise, if $\hat{\boldsymbol{y}}$ overestimates $\boldsymbol{y}_t$, the penalty should be proportional to $1 - F(u - \boldsymbol{y}_t)\Pr(\boldsymbol{y}_t < u)$. Please note that the proposed EVT-based adaptive loss function assigns the same weight to the calculated loss for $\boldsymbol{y}_t > u$ and the loss calculated for $\boldsymbol{y}_t < u$.

IV. Numerical Results

This section aims to evaluate the effectiveness of the proposed methodology in predicting received power values for a system operating under URLLC conditions. We focus on assessing the ability of the proposed scheme to predict the extreme received powers in the next few milliseconds (msec). We also compare these results with those obtained using the traditional GRU model for channel prediction. To the best of our knowledge, DRNN models, especially GRU, with the traditional MSE loss function are the optimum learning approach for predicting the time series [5].

A. Data Characterization

The channel measurements were collected from the engine compartment of a Fiat Linea operating at 60 GHz, a millimeter-wave frequency that holds significant potential for both automotive and indoor applications. The position of the transmitter and receiver antennas are selected to capture the effects of engine vibration on the received power. A vector network analyzer (VNA) (R&S® ZVA67) was connected to the antennas via R&S® ZV-Z196 port cables, which have a maximum transmission loss of 4.8 dB. The horn antennas used for transmission and reception feature a nominal gain of 24 dBi and a 12° vertical beamwidth. The transmitter and receiver antenna locations were chosen from the available options for wireless sensors within the engine compartment, specifically at positions 5 and 13 as shown in [13]-Fig. 1. Approximately 10^6 consecutive samples were recorded with a time resolution of 2 msec. This resolution is sufficiently short to capture rapid fluctuations in the received signal strength caused by the engine vibrations. The proposed algorithm was implemented using Python.

B. Model Characterization

The proposed method employs a single GRU layer with 50 units to capture temporal dependencies in the input data. We conducted ablation studies to explore different GRU configurations and determined that a single GRU layer with 50 units achieved the best balance between performance and computational efficiency. Since the primary objective of this study is to compare the performance of GRU models using different loss functions, we selected this simple and efficient architecture. This choice ensures fast predictions while allowing us to focus on evaluating the impact of the loss functions on extreme value prediction. Compiled with the Adam optimizer at a learning rate of 0.0008, it utilizes an EVT-based adaptive quantile loss function to enhance extreme value predictions. The model is trained for 20 epochs with a batch size of 256, using a validation set to ensure generalization for predicting the next $\{1, 2, 3, 4, 5\}$ samples, corresponding to a maximum prediction horizon of 10 ms. Hyperparameters are tuned experimentally for optimal performance. To prepare the dataset, we convert raw time-series data into a supervised format using a sliding window technique, extracting sequences of past observations. We selected look-back (LB) periods of 5 or 15 data points, with 5 addressing the strict 10 ms latency requirement of URLLC, while 15 allows for greater prediction accuracy. The GRU with EVT-based adaptive quantile loss accurately predicts tail samples by utilizing GPD parameters estimated from excess values below a threshold of -49 dBm, derived from the GAN-EVT model [9]. During prediction, the loss function for the values above the threshold is the MSE loss, while those below are predicted with the EVT-based adaptive quantile loss, guided by pre-assigned parameters.

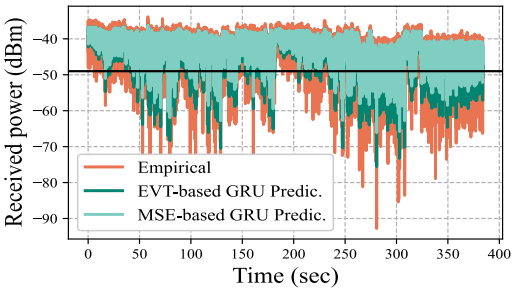


Fig. 2. GRU prediction using MSE-based and EVT-based adaptive quantile losses. The horizontal black line shows the optimum threshold of -49 dBm.

C. Discussion

Fig. 2 depicts the power values for the validation dataset, representing 20% of the experimental data and using the last 5 samples (LB = 5) for prediction. The proposed GRU model with EVT-based adaptive quantile loss outperforms the GRU model trained with MSE loss in predicting extreme values, particularly below the -49 dBm threshold. Specifically, the EVT-based model correctly predicts 21% of values below -49 dBm, compared to 7% for the MSE-based model. Both models have identical computational complexity, each containing 7950 trainable parameters and requiring approximately 30 μsec per prediction. Thus, the proposed adaptive quantile loss function does not increase model complexity. When the LB is increased to 15, the prediction delay grows by less than a factor of 2, contrary to the expected linear scaling. This deviation can be attributed to TensorFlow's ability to parallelize certain operations, which mitigates the scaling

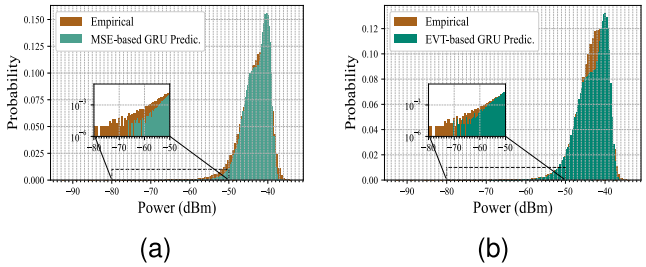


Fig. 3. PDF comparison between the empirical values and GRU predicted values: (a) GRU with the MSE-based loss function, and (b) GRU with the EVT-based adaptive quantile loss function.

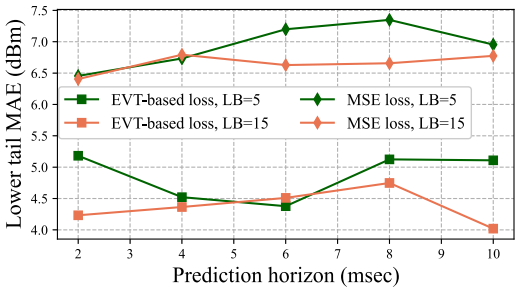


Fig. 4. MAE results for GRU using MSE-based and EVT-based adaptive quantile losses for tail values across different prediction horizons.

effects. Additionally, fixed overhead costs, such as model loading, remain constant regardless of input length, further contributing to the sub-linear increase in delay.

Fig. 3 compares the PDFs of the empirical data with those predicted by the GRU using both EVT-based adaptive quantile loss and MSE loss functions. Both models effectively estimate average values. However, the GRU with the EVT-based adaptive quantile loss function outperforms in predicting tail distributions. It also provides more accurate predictions near the average values, with better scaling around the peak of the distribution. This is particularly evident in the zoomed-in sections of the tail, where the proposed model successfully captures the tail values that significantly influence the performance of URLLC systems. The results show that the GRU with MSE loss predicts the tail portion of the distribution as 2.4%. In comparison, the GRU with the EVT-based adaptive quantile loss function predicts 7.26%, which closely matches the empirical tail portion of 7.49%.

Fig. 4 illustrates the MAE for the tail values across different prediction horizons. As shown in this figure, the MAE for the GRU model with the EVT-based adaptive quantile loss function is approximately 3 dBm lower than that for the GRU with the MSE-based loss function. This indicates that the primary challenge of predicting tail values within a few milliseconds has been effectively addressed. Additionally, for a LB period of 15, the loss is slightly lower than that for a LB of 5. However, this negligible difference suggests the outperformance of the GRU with the proposed EVT-based adaptive quantile loss function to estimate tail values using the samples collected only within the last 10 msec. On the other hand, our results show that the GRU model with the MSE loss function outperforms the GRU model with the EVT-based adaptive quantile loss function by less than 0.5 dBm in terms of MAE for predicting average statistic channel, i.e., received power values above the threshold of -49 dBm.

V. Conclusion

In this letter, we propose a DRNN with GRU model using an EVT-based adaptive quantile loss function for the channel prediction in URLLC. By incorporating EVT to estimate the GPD parameters fitted to the channel tail distribution, we formulate an adaptive quantile loss function for the DRNN with the GRU model, enhancing its ability to predict extreme values. Our approach demonstrates improved predictive performance for extreme values as validated through quantitative metrics of MAE and probability density, and visual comparisons with traditional GRU model using the MSE-based loss function. In the future, to further improve the channel prediction performance of the proposed scheme, we plan to increase the number of training samples using GAN-based data augmentation, and optimize the weights assigned to the loss of near-average samples and the tail samples. We also plan to account for non-stationarity in the dataset by estimating GPD parameters in real-time and integrating them into the EVT-based adaptive quantile loss function. Moreover, we plan to utilize the channel prediction results for the link adaptation to achieve URLLC.

References

[1] Y. Huang, Y. T. Hou, and W. Lou, "DELUXE: A DL-based link adaptation for URLLC/eMBB multiplexing in 5G NR," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 143–162, Jan. 2022.

[2] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive MIMO with channel aging," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2960–2973, May 2020.

[3] W. Jiang and H. D. Schotten, "Deep learning for fading channel prediction," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 320–332, 2020.

[4] W. Jiang and H. D. Schotten, "Recurrent neural networks with long short-term memory for fading channel prediction," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.

[5] G. Bartoli and D. Marabissi, "CQI prediction through recurrent neural network for UAV control information exchange under URLLC regime," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5101–5110, May 2022.

[6] N. Mehrnia and S. Coleri, "Wireless channel modeling based on extreme value theory for ultra-reliable communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1064–1076, Feb. 2022.

[7] N. Mehrnia and S. Coleri, "Multivariate extreme value theory based channel modeling for ultra-reliable communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4964–4975, May 2024.

[8] M. K. Abdel-Aziz, S. Samarakoon, C.-F. Liu, M. Bennis, and W. Saad, "Optimized age of information tail for ultra-reliable low-latency communications in vehicular networks," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1911–1924, Mar. 2020.

[9] P. Valiahdi and S. Coleri, "GANs for EVT based model parameter estimation in real-time ultra-reliable communication," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, Jun. 2024, pp. 1187–1191.

[10] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An Introduction to Statistical Modeling of Extreme Values*, vol. 208. Cham, Switzerland: Springer, 2001.

[11] N. Mehrnia and S. Coleri, "Non-stationary wireless channel modeling approach based on extreme value theory for ultra-reliable communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8264–8268, Aug. 2021.

[12] Q. Xu, K. Deng, C. Jiang, F. Sun, and X. Huang, "Composite quantile regression neural network with applications," *Expert Syst. Appl.*, vol. 76, pp. 129–139, Jun. 2017.

[13] U. Demir, C. U. Bas, and S. Coleri Ergen, "Engine compartment UWB channel model for intravehicular wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2497–2505, Jul. 2014.