Functional Linear Regression for the prediction of streaming video QoE

Jean Steve Tamo Tchomgui *†, Vincent Barriac †, Guillaume Fraysse†,
Julien Jacques* and Stéphane Chrétien*
*Univ Lyon, Univ Lyon 2, ERIC, France, †Orange Innovation, France
{guillaume.fraysse, vincent.barriac}@orange.com
{jean-steve.tamo-tchomgui, julien.jacques, stephane.chretien}@univ-lyon2.fr

***Abstract*—Streaming video is an ubiquitous service used by billions of people every day. Offering the best-in-class Quality of Experience (QoE) is a challenge to Mobile Network Operators (MNOs) which requires them to identify the obstacles to overcome. To properly achieve this, MNOs require efficient monitoring able to maintain this QoE at optimum levels. Machine Learning (ML) models play an essential role in this task. This paper introduces PenFFR, a novel approach based on Functional Data Analysis (FDA), to identify the underlying functions between the QoE and the features that characterize the streaming videos thus enabling to predict future values. The experimental results show that the performance of the prediction is on par for one metric with Deep Learning-based methods and even better for another metric. To facilitate the understanding of our method and enable hands-on application, the code is made available [1].**

***Index Terms*—Functional data analysis, function-on-function regression, Quality of Experience, QoE prediction, Video streaming**

I. Introduction

With the rise in popularity and usage of video streaming services and platforms triggered by larger available mobile network bandwidths and new devices with high resolution screens, we have been witnessing a huge increase in data traffic over mobile networks [2] and forecast confirm this trend for the coming years [3]. Surveys show that these services now represent a significant share of the time spent online with more than 33% of people above 18 spending at least an hour a day in average on an online video service [1]. In a context of competitive markets, consumers are naturally expecting a high-quality streaming experience. As a consequence, MNOs face the challenge of ensuring the best QoE possible for their users, to keep their promise on service experience and avoid churn. QoE is a measure of overall customer satisfaction with factors that encompass the whole service. In the case of streaming video, such factors include video playback quality (both objective and subjective) and buffering times. Delivering a good QoE for streaming video services requires therefore the possibility to know and master those factors. However MNOs generally do not have access to the factors that directly influence QoE, unless they are also application providers. These are located in service platforms or on users' devices, but MNOs only have access to network metrics like jitter, throughput, delay or packet loss rate. From these metrics and indicators representative of network performance, it becomes wishable for MNOs to infer the quality level of end-user perception that they cannot measure directly. MNOs are increasingly turning to ML techniques to model QoE from network data, thanks to the availability of massive amounts of data, as well as the increased maturity of ML tools and models and associated computing facilities. In the specific case of video streaming services, these models can predict QoE by analyzing vast amounts of data generated from network usage patterns, video streaming metrics from the devices and user feedback. The predictive capabilities of ML theoretically enable operators to proactively manage network resources, optimize streaming quality, and preemptively address potential issues before they affect the user experience. But there are challenges to overcome, related either to the high variability over time while streaming of factors such as network throughput, video bit rate and display size, or to different optimization strategies of streaming services (maximizing video quality versus minimizing rebuffering events). In this paper we apply our PenFFR (for Penalized Function-on-Function Regressor) method. It is a statistical method based on the FDA paradigm that enables to identify the underlying function in the data and allows, here, to compute the QoE from the input data. Section II provides a state of the art on QoE and methods to compute or predict it as well as on FDA. Section III describes PenFFR. Then, Section IV describes the dataset used for the evaluation of the method and the transformations performed on the data. Results of these experiments are then detailed, compared to several baselines and analyzed in Section V.

II. State of the Art

The integration of ML models for predicting QoE is now a relatively mature field of study [4]. Modelling of QoE was initially based on psycho-physic, by a replication of human perception (vision, hearing) and opinion building. Such models were very explanatory, with building blocks corresponding to well known physical or mental processes. With ML, the models become independent from psycho-physics and can rely therefore on more sources of information [5]. They also benefit of the flexibility, modularity and scalability of ML techniques. Most QoE models require access to the audio or video signals, which is most of the time very challenging, and cannot be

[1]https://www.statista.com/statistics/611750/millennial-time-spent-with-online-video/

978-3-903176-66-9 ©2024 IFIP

used for network-centric operations. This is why a few recent studies addressed alternative approaches, where information from the end-users' devices or the network could be envisaged as input for QoE prediction. In the field of voice quality, a good example of such models is the ITU-T Recommendation P.565.1 also know as sQLEAR [6]. It is using a combination of regression algorithms, namely a Random Forest (RF) regressor and support vector regressor. As far as we are aware, no similar model for video quality of streaming services exists, but some work on video QoE can be mentioned [7], [8] as basis for further development of models. The link between QoE and measurements by network operators is also a topic of study [9]. The development of such models relies on training with a large amount of data representative of network equipment behaviour and performance. The collection of such data is a huge task. Some solutions are proposed based on network simulators [10] or automated monitoring solutions [11]. There are several challenges faced by designers of such network-based QoE prediction models. The biggest one lies in the heterogeneity of data source in terms of data nature (radio, IP, etc.), source location (access, core, device) and time series structure (real time, average overt time, etc.).

With the emergence of new generations of networks, we can collect information at very high frequencies in various places. It has become necessary to develop new tools for exploiting and analysing this ever increasing volume of data. This is one of reason why FDA has become very popular and useful in a constantly growing number of applications : medical [12], economics [13] and commerce [14], ... FDA is a branch of statistics that deals with data that can be represented as functions. Unlike traditional data analysis, which focuses on discrete observations, FDA involves analyzing data that is inherently continuous, such as curves, surfaces, and shapes. Extension of linear regression to the functional setting has therefore naturally become a major area of research in FDA. While the literature is too vast to cover here, the recommended references for this field are [15]–[18], which provide excellent introduction to FDA. A broad overview of Functional Linear Regression (FLR) methods is provide in [19] and [20]. In mobile network context, [21] use FDA to detect future malfunctions, capacity degradation, accessibility and call drops anomalies for Long-Term Evolution (LTE) networks. As it shown in [22], FDA can be use not only to handle the diversity both in terms of the situations that must be faced and the data used to reach conclusions, but also optimized and improve the scalability of solutions in network management task (*e.g.* anomaly detection).

III. The PenFFR Functional Linear model

This section introduces our PenFFR process for the estimation of the functional linear model. The code of its implementation in the R language has been made available open-source on GitHub [1].

A. Problem statement

Predicting the continuous QoE of video streaming is the problem of estimating a linear relationship between Key Performance Indicators (KPIs) and the subjective QoE perceived by the end-user of the video streaming service. This relationship has the following form:

$$\left\{ \mathrm{Y}_i(t),\, \mathrm{X}_i^1(t), \ldots, \mathrm{X}_i^p(t)\,,\ t \in [0, \mathrm{T}] \right\}, i = 1, \ldots, n\,,$$

where $\mathrm{Y}_i(t)$ is the QoE of the i^{th}-session at time t and the p input variables $(\mathrm{X}_i^l(t))_{1\le l\le p}$ are assumed to be the QoE's KPIs at the same time t. We assume that all these functions belong to the separable Hilbert space $\mathrm{L}^2\big([0;\mathrm{T}]\big)$.
We focus on the concurrent model defined by Model (1):

$$\begin{aligned} \mathrm{Y}_i(t) &= \beta_0(t) + \sum_{l=1}^{p} \beta_l(t)\, \mathrm{X}_i^l(t) + \varepsilon_i(t) \\ &= \mathrm{X}_i(t)^\top \beta(t) + \varepsilon_i(t)\,, \end{aligned} \quad (1)$$

where $\beta(t) = \big(\beta_0(t),\ \beta_1(t),\ \ldots,\ \beta_p(t)\big)^\top$ are the unknown functional parameters and are assumed to be square integrable; $\mathrm{X}_i(t) = \Big(1,\, \mathrm{X}_i^1(t),\, \ldots,\, \mathrm{X}_i^p(t)\Big)^\top$ the design matrix; $\varepsilon_i(t)$ is the model error and is a sample of centered random variables with variance σ_i^2, specific to the i^{th} individual (cf. [15], Chapter 13); $\varepsilon_i(t)$ and $\mathrm{X}_i(t)$ are assumed to be uncorrelated. The noise functions $\varepsilon_i(t)$ can be rigorously defined using white noise theory as presented in [23]. In our context, we will only use the fact that when sampled at various times from a finite set $\mathcal{T}$, the vector $(\varepsilon_i(t))_{t\in\mathcal{T}}$ can be expressed as a sum of a vector with i.i.d. components and a vector with prescribed covariance matrix, i.e. a vector with constant components in the simplest case.

B. Recovering the functional nature of data

In practice, we do not properly observe a continuous curve for each realization of both response variable $\mathrm{Y}_i(t)$ and covariate variables $\mathrm{X}_i^\ell(t)$. We only have access to a set of noisy observations at a finite number of points on a grid. As a result, the functional data can be presented as a numerical vector. In order to recover the continuous form, which generally belongs to an infinite dimensional space (e.g. Hilbert separable space $L^2([0,T])$), one efficient way to proceed is by expanding the considered functions in a functional basis. It can be a polynomial, splines, Fourier, wavelets ... basis depending on the shape and variations of the variable. The advantage of this approach is the fact that by truncating the series at a given level q_ℓ, we obtain an approximation of the function $\mathrm{X}_i^\ell(t)$ in a q_ℓ dimensional space. In mathematical terms, given a basis $\{\mathrm{B}_j^\ell(t)\}_{j\ge1}$, the function $\mathrm{X}_i^\ell(t)$ will be expressed as:

$$\mathrm{X}_i^\ell(t) = \sum_{j=1}^{q_\ell} x_{ij}^\ell\, \mathrm{B}_j^\ell(t),$$

with x_{ij}^{ℓ} the basis coefficients. Assume that we have n realizations $\left(\mathrm{X}_i^{\ell}(t)\right)_{1\leq i\leq n}$ whose values on an observation grid are given by:

$$\left\{(\mathrm{X}_{i1}^{\ell}, t_{i1}^{\ell}),\, (\mathrm{X}_{i2}^{\ell}, t_{i2}^{\ell}),\, \ldots, (\mathrm{X}_{im_i}^{\ell}, t_{im_i}^{\ell})\right\}_{1\leq i\leq n},$$

where X_{ij}^{ℓ} is the value of the curve $\mathrm{X}_i^{\ell}(t)$ at timestamp t_{ij}^{ℓ}. For each i, we want to recover the trajectory $\mathrm{X}_i^{\ell}(.)$ based on the information that:

$$\mathrm{x}_{ij}^{l} = \mathrm{X}_i^{\ell}(t_{ij}^{\ell}) + \varepsilon_{ij}, \quad \text{with } \varepsilon_{ij} \text{ an unobserved Gaussian noise.}$$

For example, in the case of a cubic B-spline basis [15], the above expression in functional basis $\{\mathrm{B}_j^{l}(t)\}_{j\geq 1}$ truncated at $q_{_l}$ becomes:

$$\begin{aligned}\mathrm{X}_i^{\ell}(t) \;&=\; \sum_{j=1}^{q_\ell} x_{ij}^{\ell}\,\mathrm{B}_j^{\ell}(t) \;=\; x_{i1}^{\ell} + x_{i2}^{\ell}\,t + x_{i3}^{\ell}\,t^2 + x_{i4}^{\ell}\,t^3 \\ &\quad + \sum_{j=1}^{q_\ell-4} x_{i(4+j)}^{\ell}\,(t-\tau_j)_+^3, \end{aligned} \tag{2}$$

where τ_j are the nodes, $f(t)_+ = \max(f(t), 0)$ the positive part of $f(t)$ and the parameters $(x_{ij}^{l})_{0\leq j\leq q_{_l}-1}$ with respect to the truncated cubic B-spline basis $\{1,\, t,\, t^2,\, t^3,\, (t-\tau_1)_+^3,\, (t-\tau_2)_+^3,\, \ldots,\, (t-\tau_{q-4})_+^3\}$. For any realization i the m_i observations on the discrete grid $t_{i1}^{\ell}, \ldots, t_{im_i}^{\ell}$, we can write the matrix/vector form as:

$$\mathrm{X}_i^{\ell} = \begin{pmatrix}\mathrm{X}_i^{\ell}(t_{i1}^{\ell}) & \ldots & \mathrm{X}_i^{\ell}(t_{im_i}^{\ell})\end{pmatrix}^{\top} := \mathrm{T}_i^{\ell}\, x_i^{\ell},$$

where T_i is the matrix of truncated B-spline basis in t_{ij}^{ℓ}. Therefore, the problem is to recover the functional basis coefficients which is equivalent to solving the minimization problem:

$$\min_{x_i^{\ell}} \|\mathrm{X}_i^{\ell} - \mathrm{T}_i^{\ell}\, x_i^{\ell}\|^2.$$

C. The model estimation

Given the shape and fluctuations of the studied data, described in Section IV, which does not have a high variability between successive observations or any periodicity, a B-splines basis is appropriate to process the data efficiently. So, parameters of Model (1) are estimated by expanding the functional covariates and parameters onto a common B-spline basis. So we decompose the covariates and the parameters as:

$$\mathrm{X}_i^{l}(t) \;=\; \sum_{j=1}^{q_{\mathrm{x}^l}} x_{ij}^{l}\,\mathrm{B}_j^{l}(t) = \mathrm{B}^{l}(t)^{\top} x_i^{l}\,, \tag{3}$$

$$\beta_l(t) \;=\; \sum_{j=1}^{q_{\beta^l}} b_j^{l}\,\phi_j^{l}(t) = \phi^{l}(t)^{\top} b^{l}\,, \tag{4}$$

where $\mathrm{B}^{l}(t) = \left(\mathrm{B}_j^{1}(t),\, \ldots,\, \mathrm{B}_j^{q_{\mathrm{x}^l}}(t)\right)^{\top}$ is the q_{x^l} -dimensional vector of basis function for the covariate $\mathrm{X}^{l}(t)$ and $x_i^{l} = (x_{i1}^{l}, \ldots, x_{iq_{\mathrm{x}^l}}^{l})$ the corresponding basis coefficients. Following the same pattern, $\{\phi^{l}(t),\, b^{l}\}$ are the basis functions and basis coefficient for $\beta_l(t)$.

Under this assumption, the matrix terms elements of Model (1) become:

$$\mathrm{X}_i(t) = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \mathrm{B}^1(t)^{\top} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathrm{B}^p(t)^{\top} \end{pmatrix} \begin{pmatrix} 1 \\ x_i^1 \\ \vdots \\ x_i^p \end{pmatrix} \;=\; \mathrm{B}(t)\, x_i.$$

$$\beta(t) = \begin{pmatrix} \phi^0(t)^{\top} & 0 & \ldots & 0 \\ 0 & \phi^1(t)^{\top} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \phi^p(t)^{\top} \end{pmatrix} \begin{pmatrix} b^0 \\ b^1 \\ \vdots \\ b^p \end{pmatrix} \;=\; \Phi(t)\, b\,.$$

Using the above expressions, Model (1) becomes:

$$\mathrm{Y}_i(t) \;=\; x_i^{\top}\mathrm{B}(t)^{\top}\Phi(t)\; b + \varepsilon_i(t) = \mathrm{R}_i(t)^{\top} b + \varepsilon_i(t) \tag{5}$$

The form of Model (5) is a classical linear regression expression with $\mathrm{R}_i(t) = \Phi(t)^{\top}\mathrm{B}(t)\, x_i$ the design matrix and b the parameters to be estimated. We can also include some non-functional KPIs to this model.

D. Confidence interval

Confidence interval found their origins in compilation works of Gosset and Fisher done in [24]. It is the most common tool to quantify uncertainty associated to predictions, i.e., the value predicted by a ML system can significantly differ from the actual value due to high variability. However, if the ML system can estimate a range that encompasses the actual value with a high probability, the method used to generate the confidence interval can determine an interval between the lowest and highest potential outcomes, enabling a more informed decision-making. Following this target, conformal prediction is a framework in ML and statistics that provides a principled way to assign confidence measures to predictions made by ML models. They provides prediction regions that come with a guarantee of validity. In conformal prediction, the goal is to construct prediction regions such that, with a specified confidence level, they contain the true value of the target variable for any new input instance. These prediction regions can take various forms depending on the type of prediction task and the underlying assumptions of the model. In regression tasks, they may be intervals or quantiles. For a given labelled dataset $\{(x_i, y_i), 1\leq i\leq n\}$ where x_i represent the covariates and y_i the response, the conformal prediction method performs the following steps [25]:

- Split the dataset into two disjoint subsets : a train subset noted $\mathcal{A}_1$ and a calibration subset $\mathcal{A}_2$;
- Fit a ML model f to learn the relationship between covariates and response. The specific model can vary widely, including linear regression, decision trees, neural networks, quantile regression, etc.
- Compute E_i the absolute residuals on calibration set:

$$E_i = |y_i - f(x_i)|,\; \forall i \in \mathcal{A}_2.$$

- Then for a given level τ, compute

$$Q_{1-\tau}(\mathcal{A}_2) := (1-\tau)(1+\frac{1}{|\mathcal{A}_2|})^{\text{th}} \text{ quantile of } (E_i)_i.$$

- Finally, the prediction interval $C(x_{n+1})$ of a new point x_{n+1} is given by:

$$C(x_{n+1}) = \Big[f(x_{n+1}) - Q_{1-\tau}(\mathcal{A}_2)\,;\, f(x_{n+1}) + Q_{1-\tau}(\mathcal{A}_2)\Big] \quad (6)$$

One of the key properties of conformal prediction is its validity guarantee. It ensures that the observed error rate of the prediction intervals matches the specified confidence level.

IV. Dataset and Data Analysis

This section introduces the dataset used for our experiments, then provides some insights on the data it contains. It is then used in the experiments described in Section V.

A. The LIVE-NFLX-II dataset

One challenge when using novel methods is to find appropriate data. There a few datasets for QoE prediction [26], [27] but almost all of them are built to predict a single scalar value, the Mean Opinion Score (MOS) score of the whole video. For example a dataset for the QoE prediction of video streaming will contains only a single scalar value for the QoE of each video, or each portion of video. To the best of our knowledge the only dataset that contains values for QoE at every timestep or frame of a video is the LIVE-NFLX-II [28]. LIVE-NFLX-II is a QoE dataset made at The University of Texas at Austin's LIVE subjective testing lab using realistic adaptive streaming pipeline model that contains four main modules: an encoding module, a video quality module, a network transmission module and a client-based video playout module. This streaming pipeline is designed to recreate a complete end-user QoE through 3 streaming dimensions: encoding, network throughput and the selected Adaptive Bitrate Streaming (ABR) algorithm. To have a look at every of those dimensions, LIVE-NFLX-II dataset is based on 15 video contents, 7 actual network traces and 4 adaptive algorithms, yielding 420 video streams in total. Each of the 15 videos belong to a content genre: action, documentary, sports, animation and video games. Each of these genres has different dynamics of the content of the video as well as different types of content of each frame. Concerning the network throughput, The 7 network traces were manually selected from the HSDPA dataset [29] which is widely used to compare adaptation algorithms, each of these traces allows to simulate specific network conditions. For the ABR algorithm, 4 representative adaptive algorithms are selected to cover the large design space of adaptation algorithms. Each sample in the dataset is a stream of one of the 15 videos, encoded one of the 4 adaptive algorithms and steamed over one of the 7 network conditions. The dataset also provide a subjective continous-time evaluation of each frame of each of these samples by a subset of at least 22 human subjects among a group of 65. In total there are 9750 QoE evaluation available in the dataset. We refers reader to [28] for more detailed description of the dataset and the protocol followed for the evaluation by human subject.

B. Data analysis and data engineering

The collected data of LIVE-NFLX-II consists of 420 distorted videos, each of them viewed by at least 22 subjects. Each of the 65 subjects made 150 evaluation. Overall they have $65 \times 150 = 9750$ continuous z-normalized scores to study the subjective QoE. Fig. 1 shows the distribution of the MOS scores of the overall 420 distorted videos which has a non surprisingly normal form (after a z-normalized transformation). Figure 2 shows for each frame (along the x-axis, from the first frame of the video to the last) the perceived MOS on the y-axis and there is one line for each of the 420 samples, the figure does not show a specific trend.

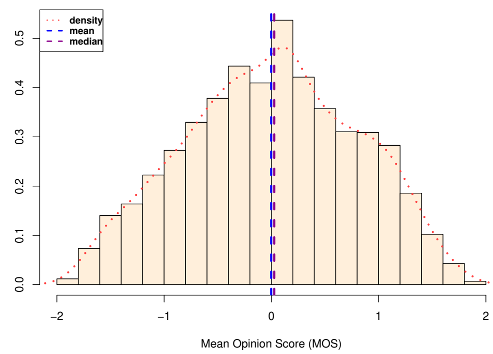


Fig. 1: Mean Opinion Score (MOS) distribution

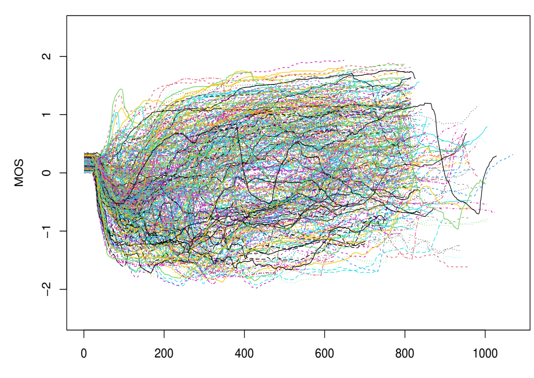


Fig. 2: Mean Opinion Score (MOS) curves

Figure 3 shows for each frame (along the x-axis) the distribution of the value of the QoE in a boxplot. The darker section in the middle represent the 50% around the median value. The green line show the median values of QoE for each frame (the y-axis on the left). The red line shows the total number of videos that have that many frame (the y-axis on the right), all videos are composed of at least 625 frames,

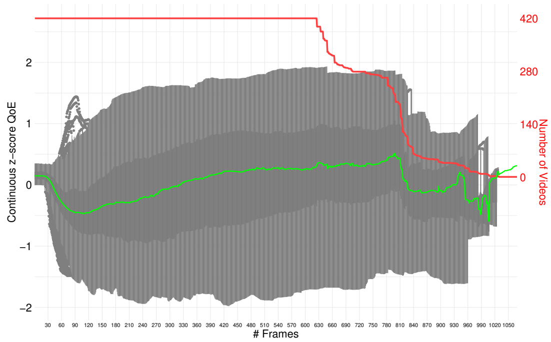
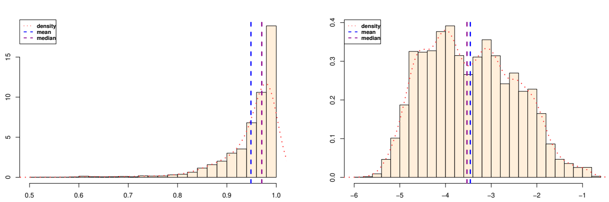


Fig. 3: Boxplots of MOS for each frame, red line is the total number of videos

but not all of them have the same number of frames. This figure, like Figure 2, does not show a general trend except for the very first frames where the perceived quality for all videos is within a small interval. For further frames the distribution of the perceived QoE is similar.

The rest of this section details the features used and transformations applied to them. To understand the behavior of end-user quality, key metrics are collected on top of playout bitrate, number of rebuffers, rebuffering time across adaptive algorithms. Others several well-known Quality of Service (QoS) metrics are used for our model for QoE prediction. The main metrics in the dataset are the Peak Signal to Noise Ratio (PSNR), the Spatio Temporal Reduced Reference Entropic Differencing scores (ST-RRED), the Structural SIMilarity index measure score (SSIM), the MultiScale SSIM (MS-SSIM), the Video Multi-Method Assessment Fusion (VMAF) and the throughput traces. Since we are trying to fit a linear Gaussian model whose variables are assumed to follow this distribution we applied $x \mapsto \log(x)$ and $x \mapsto \log(1-x)$ transformations respectively for metrics whose distribution was either skewed to the left or to the right, to obtain a more or less normal distribution. Fig. 4 shows an example, for the MS-SSIM metric, of the histograms before and after this transformation.



(a) Original data (b) log Transformation

Fig. 4: MS-SSIM scores and its logarithm transformation

V. Experiments and results

This section first details the experimental protocol followed to apply the model detailed in Section III on the dataset described in Section IV. It then gives the results and compare them to three types of baseline, one from the original work that introduced the dataset, a more basic baseline based on a RF model and another FDA method called pffr [30].

A. Baseline

We compare the results of our PenFFR model, described in Section III, to the two prediction algorithms presented in [31]. The first of these models is based on AutoRegressive (AR) Neural Networks (NNs) (G-NARX) and the other based on recurrent NNs (G-RNN). These 2 models were trained using VMAF measurements with frame as the continuous-time Video Quality Assessment (VQA) feature. They also included two other continuous-time features: a per-frame boolean variable indicating the presence of rebuffering and another indicating the time elapsed since the last rebuffer. G-NARX used 8 input delays and 8 feedback delays. G-RNN used 5 layer delays. Those results from [28] are used here as a baseline to compare with our methods. The experience were not reproduced and we compare our results to the result of this paper.

We also compare our FDA-based PenFFR method and these two algorithms to a FDA-based pffr method and a RF model. To train the RF model, the per-frame metrics described in Section IV-B are used. For each of them, statistical features are computed : *maximum, minimum, total, quartiles, standard deviation, mean, skewness* and *kurtosis*. The output of the RF model is the mean value of continuous QoE z-scores of distorted videos. When all the features are computed, a grid search is performed for the hyper-parameters of the RF model: the maximum depth of trees, and the number of trees. Once the best values for the hyper-parameters are found, an initial RF model is trained using all the features.

Figure 5 shows the pipeline of the 5 different models: PenFFR, pffr, G-NARX and G-RNN and RF. All models rely of course on the same input data. Additional statitical feature are added in the case of the RF model as detailed above. It is important to understand that each of these methods has a different format of output. The RF model in the baseline outputs a single scalar value for each of the n videos (the mean MOS). Both G-NARX and G-RNN output n vectors (one for each video) of values, i.e. the MOS for each frame. Finally the proposed PenFFR method computes a function for each of the videos that allow to compute the MOS for any frame f.

For all the 5 algorithms, the train/test split process is performed based on which of the 7 network conditions was used (cf. description in Section III) : choosing 5 types of network conditions for training and 2 for testing each time, which yields 300 videos (15 contents, 4 adaptors and 5 traces) for training and 120 videos (15 contents, 4 adaptors and 2 traces) for testing. After describing implementations details of the methods, selecting appropriate evaluation metrics predictive outputs to ground truth is crucial. We can use:

1) The **Root Mean Square Error** (RMSE) is a metric commonly used for the evaluation of the accuracy of predictions. It provide a quantitative measure of how

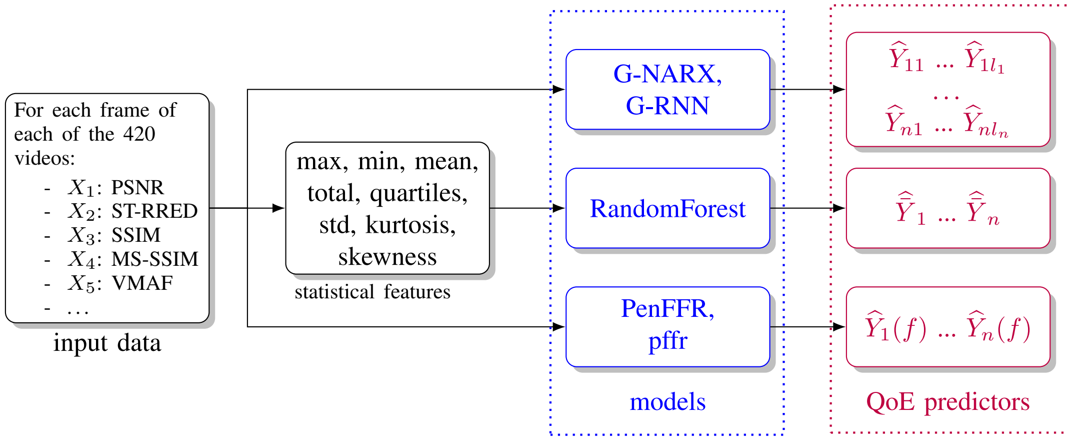


Fig. 5: Pipelines of the models

well predictions p_i align with ground truth values g_i. His formula is:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - g_i)^2}$$

where N is the number of frames. RMSE is intuitive and easy to interpret but sensitive to outliers.

2) The **Outage Rate** (OR) assess the quality of prediction by quantifying the proportion of instances (time) where the predicted value falls outside a predefined tolerance range or threshold. Here we consider the 95% confidence interval of the ground truth g_i at frame i across all videos. OR is calculated as :

$$\text{OR} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\{p_i-\frac{\varepsilon}{2}\leq g_i \leq p_i+\frac{\varepsilon}{2}\}}$$

where ε is the length of the confident interval.
OR should be interpreted in conjunction with other performance metrics to provide a comprehensive assessment of prediction quality.

Using these metrics allows us to compare the performance of our model and of pffr method, the RF baseline to the G-NARX and G-RNN models since these metrics were used in [28].

B. Results and discussion

This section introduces the results of the 5 different algorithms described above. For G-NARX and G-RNN the results are extracted from [28]. For RF, pffr and PenFFR the results come from the experiments described in Section IV.

The RF baseline has the worst RMSE value, 150% of the G-NARX algorithm which gives the best RMSE. The functional PenFFR method has a RMSE of 0.289, 108% of G-NARX and the other FDA-based method pffr has a RMSE of 0.344.

For the OR metric, the FDA-based methods offers the best results: PenFFR with 4.72% and pffr with 4.84% of predicted values of all frames within the 95% confidence interval, when G-NARX reaches 7.14% and G-RNN 5.96%. RF reaches 5.6%. The results are summed up in Table I.

The proposed PenFFR method allows for at least a 25% improvement of the OR metric over the G-NARX and G-RNN methods while getting a 8% worse RMSE value. PenFFR also improves upon pffr on both metrics. This method does not only predict the values of the QoE metric but also outputs a function that describes how the QoE metric can be computed from the input features.

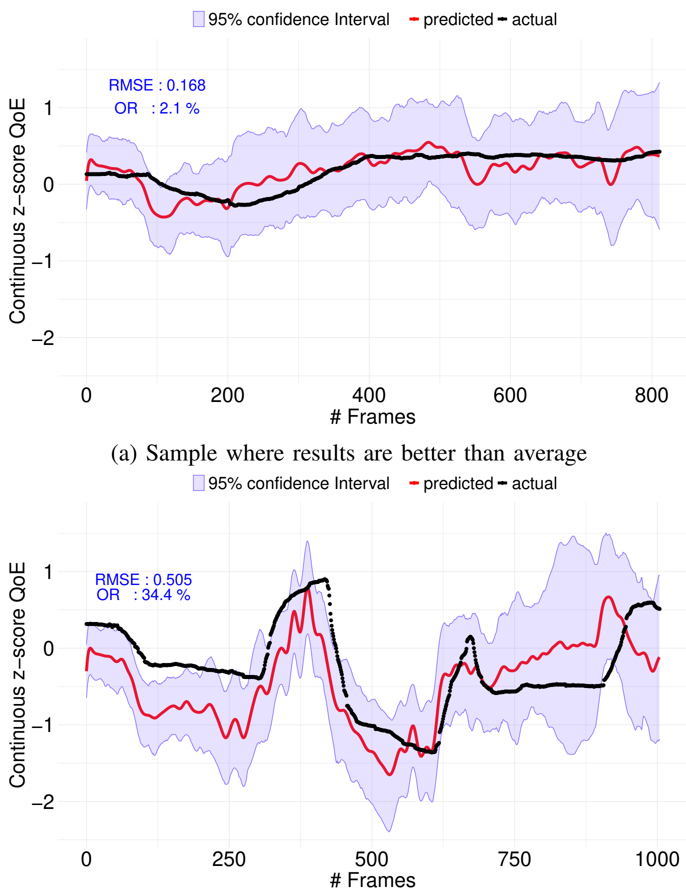
Methods	RMSE	OR
G-NARX [28]	**0.267**	7.14%
G-RNN [28]	0.276	5.96%
RF	0.402	5.6%
pffr	0.344	4.84%
PenFFR	0.289	**4.72%**

TABLE I: Results of RMSE and OR metrics for the 4 methods

Figure 6 shows the prediction (red line) compared to the actual values (black dots) for two examples taken from the test set. The example in Figure 6a reaches better values for both the RMSE and OR metrics, while the second example in Figure 6b reaches worst values. The purple area shows the 95% confidence interval as defined in Section III-D.

VI. Conclusion and Future Works

This paper details PenFFR, a new model based on the FDA framework, and how to apply it for the prediction of the QoE of streaming videos. This model allows to identify the underlying function between an output variable (here the QoE) and the input variable (here the characteristics of the video) when both input and output variables are functions. In the dataset used for the experiment, both the input and output variables are functions of the frame within a video stream. The experimental results show that this model can outperform current state of the art Deep Learning models for the Outage Rate metric with a slight degradation of the RMSE.
This initial application of FDA in a Network Management use case could be further pursued on other datasets with comprehensive evaluations and on the development of advanced methodologies that leverage the inherent strengths of this expanding framework. Possible other FDA-based models include another of our works [32] that presents, in a theoretical



(a) Sample where results are better than average

(b) Sample where results are worse than average

Fig. 6: Examples of prediction vs ground truth

aspect, the functional Mixture-of-Experts (MoE) dedicated to handle heterogeneous data. We are confident that by applying this MoE model in the task of QoE prediction for streaming video, we will able, in a single model, to model data coming from multiple sources of streaming applications. Ultimately, this work could also help steering the standard on QoE evaluation towards new continuous metrics.

References

[1] Penffr. [Online]. Available: https://github.com/Stevens05/PenFFR

[2] Cisco, "Global mobile data traffic forecast update 2017-2022 - cisco visual networking index white paper," 2019. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html

[3] Mobile data traffic forecast – Mobility Report. ericsson.com. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast

[4] G. Mittag, *Deep learning based speech quality prediction*. Springer, 2022.

[5] G. Kougioumtzidis, V. Poulkov *et al.*, "A survey on multimedia services qoe assessment and machine learning-based prediction," *IEEE Access*, vol. 10, pp. 19 507–19 538, 2022.

[6] I.-T. R. P.565.1, "Machine learning model for the assessment of transmission network impact on speech quality for mobile packet-switched voice services," 2021.

[7] N. Wehner, A. Seufert *et al.*, "Explainable data-driven qoe modelling with xai," in *2023 15th international conference on quality of multimedia experience (QoMEX)*. IEEE, 2023, pp. 7–12.

[8] J. Schleicher, N. Wehner *et al.*, "(not) the sum of its parts: Relating individual video and browsing stimuli to web session qoe," in *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2024, pp. 104–110.

[9] T. Hoßfeld and P. Pérez, "A theoretical framework for provider's qoe assessment using individual and objective qoe monitoring," in *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2024, pp. 235–241.

[10] S. Schwarzmann, C. C. Marquezan *et al.*, "Ml-based qoe estimation in 5g networks using different regression techniques," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 3516–3532, 2022.

[11] G. Dobreff, M. Szalay *et al.*, "Data collection framework for end-to-end radio and transport network quality monitoring," in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2023, pp. 127–130.

[12] S. Ullah and C. F. Finch, "Applications of functional data analysis: A systematic review," *BMC medical research methodology*, vol. 13, pp. 1–12, 2013.

[13] S. Das, R. Demirer *et al.*, "The effect of global crises on stock market correlations: Evidence from scalar regressions via functional data analysis," *Structural Change and Economic Dynamics*, vol. 50, pp. 132–147, 2019.

[14] W. Jank and G. Shmueli, "Functional Data Analysis in Electronic Commerce Research," *Statistical Science*, vol. 21, no. 2, pp. 155–166, 2006.

[15] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd ed., ser. Springer Series in Statistics. New York: Springer-Verlag, 2005.

[16] J. O. Ramsay, G. Hooker *et al.*, *Functional Data Analysis with R and MATLAB*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[17] L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications*, ser. Springer Series in Statistics. New York: Springer-Verlag, 2012.

[18] P. Kokoszka and M. Reimherr, *Introduction to Functional Data Analysis*. Chapman and Hall/CRC, 09 2017.

[19] J. Goldsmith, J. Bobb *et al.*, "Penalized functional regression," *Journal of Computational and Graphical Statistics*, vol. 20, no. 4, pp. 830–851, 2011, pMID: 22368438.

[20] J. Morris, "Functional regression," *Annual Review of Statistics and Its Application*, vol. 2, 06 2014.

[21] Y. Ben Slimen, S. Allio *et al.*, "Anomaly prevision in radio access networks using functional data analysis," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[22] D. Muelas, J. López de Vergara Méndez *et al.*, "Functional data analysis: A step forward in network management," in *Functional Data Analysis: A step forward in Network Management*, 06 2015, pp. 882–885.

[23] T. Hida, K. Hui-Hsiung *et al.*, *White Noise: An Infinite Dimensional Calculus*, ser. Mathematics and its applications. Kluwer Academic Publishers, 1993.

[24] J. F. Box, "Gosset, fisher, and the t distribution," *The American Statistician*, vol. 35, no. 2, pp. 61–66, 1981.

[25] Y. Romano, E. Patterson *et al.*, "Conformalized quantile regression," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle *et al.*, Eds., vol. 32. Curran Associates, Inc., 2019.

[26] G. Mittag, B. Naderi *et al.*, "LSTM-Based Video Quality Prediction Accounting for Temporal Distortions in Videoconferencing Calls," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[27] Z. Li, "The Waterloo Streaming Quality-of-Experience Database-IV." [Online]. Available: https://ieee-dataport.org/open-access/waterloo-streaming-quality-experience-database-iv

[28] C. G. Bampis, Z. Li *et al.*, "Towards perceptually optimized end-to-end adaptive video streaming," *arXiv preprint arXiv:1808.03898*, 2018.

[29] "Dataset: Hsdpa-bandwidth logs for mobile http streaming scenarios," https://skuld.cs.umass.edu/traces/mmsys/2013/pathbandwidth/.

[30] A. Ivanescu, A.-M. Staicu *et al.*, "Penalized function-on-function regression," *Computational Statistics*, vol. 30, no. 2, pp. 539–568, June 2015.

[31] C. G. Bampis, Z. Li *et al.*, "Recurrent and dynamic models for predicting streaming video quality of experience," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3316–3331, March 2018.

[32] J. S. Tamo Tchomgui, J. Jacques *et al.*, "A mixture of experts regression model for functional response with functional covariates," *Statistics and Computing*, vol. 34, no. 5, p. 154, 2024.