Objective QoE Models for Cloud-based First Person Shooter Game over Mobile Networks

Henrique Souza Rossi*, Karan Mitra*, Christer Åhlund*, Irina Cotanis†, Niclas Örgen†, Per Johansson†
*Pervasive and Mobile Computing, Department of Computer Science, Electrical and Space Engineering,
Luleå University of Technology.
†Infovista AB, Sweden.
henrique.souza.rossi@ltu.se

***Abstract*—Mobile cloud gaming (MCG) lets users play cloud games (CG) on mobile devices anywhere via mobile networks. However, the stochastic nature of network quality of service (QoS) can result in varying user quality of experience (QoE). Understanding, modeling, and predicting the impact of mobile networks' QoS on users' QoE is crucial. This helps stakeholders optimize networks, and game developers efficiently create cloud-hosted games provisioned over mobile networks. This paper investigates the impact of QoS on users' QoE and proposes, develops and validates novel models for predicting QoE for MCG in mobile networks using realistic subjective tests. In particular, we propose and develop three QoE models using multiple, polynomial, and non-linear regression. Our results validate that multiple regression (with R^2=0.79, RMSE=0.45) can model complex relationships between QoS factors that impact QoE. Multiple polynomial regression achieved the overall fit with (R^2=0.94, RMSE=0.24). Lastly, the non-linear model achieved a good RMSE of 0.24. To select the best model out of the three, we applied the F-test and determined that polynomial regression had the best statistical fit.**

***Index Terms*—Subjective tests, Modeling, Prediction, Quality of Experience, Quality of Service, Mobile Cloud Gaming**

I. Introduction

The game industry is shifting towards cloud computing, enabling users worldwide to play games anywhere and anytime, using a game-as-a-service (GaaS) model. As a result, a large catalog of games can be played on users' smartphones connected via mobile networks in public spaces, or while moving within a city. Mobility can result in heavy workloads on users' access points, resulting in traffic congestion and handoffs and leading to worse QoS metrics, including high RTT [1], RJ [1], infrequent bursty jitter (BJ) [2], [3], PL [1]–[3], and co-occurrence of RTT and PL [2]. Mobile cloud gaming (MCG) is a critical service to provide over mobile networks, and ensuring a good gaming experience or QoE requires stakeholders to learn how QoS metrics affect MCG so that they can optimize their infrastructure.

Objective QoE models for MCG are vital to predict user QoE [4] and provide the means for stakeholders to adapt to underlying QoS conditions in real time. Most of the work has focused on PC-CG [5]–[10]. However, due to hardware and input constraints, PC-CG does not cover factors related to the mobile field. Factors such as screen size [7], and input [11] have been linked to differences in how users perceive quality. On the contrary, MCG requires users to hold the device and look down on smaller screens while playing with different inputs, such as a touchscreen or a gamepad. For that, current research in MCG models [12], [13] is lacking, as they do not consider jitter and the combined effect of delay and packet loss, which are all common in real mobile networks.

This paper fills these gaps by focusing on the network side of MCG and studying an extensive set of QoS factors and ranges with the aim of answering the fundamental question: *"How can we model and predict users' QoE for MCG influenced by a variety of realistic mobile network conditions ?"* To address this question, this paper proposes, develops, and validates three novel QoE models for MCG provisioned over mobile networks. Major contributions are as follows:

- Perform modeling and prediction of QoE considering burst jitter (BJ), random jitter (RJ), and the combined effect of PL and RTT which were not yet explored by the state-of-the-art research in the context of MCG;
- develop novel statistical QoE models for MCG based on multiple linear, polynomial and nonlinear regression and determine models that are best based on F-tests;
- publicly release QoE dataset[1] for the research community, which was used to train our models.

II. Testbed and Dataset

In a laboratory environment, we conducted subjective tests following the guidelines of [8], [11], [13]. Users (n=31) were invited to play CS:GO matches (n=26) streamed via Steam HomePlay on a smartphone under various network conditions (see Table I). These values were based on [10]–[12] and by conducting drive-through tests in the city of Skellefteå, Sweden, for a period of several days. For each subjective test, the QoS parameters were varied using NetEM (controlled by the ATRUIST tool [14]) which sits between the smartphone and the game streaming servers. Based on users' ratings (likert like scale 1-5 where 1 = "poor" and 5 = "excellent"), for each match, we computed the mean opinion score (MOS) as the average of the scores. For the sake of brevity, we refer the careful reader to Rossi *et al.* [15] for further details regarding the lab setup and subjective tests. The next section focuses on building novel QoE models for MCG.

[1] https://github.com/hsr-research/MCG-CSGO-dataset

TABLE I: The conditions used for subjective tests [15].

Parameter	N. Conditions	Values
RTT	7	2,25,50,100,200,300,400 ms
PL	3	5%,25%,45% at RTT=2ms
(RTT,PL)	9	PL(0.2,1,5) %; RTT(25,50,100) ms
BJ	3	Jitter(50,200,1500) ms; Interval 15s
RJ	4	μ=50ms; Std(3,6,9,12)
Total	27	Final Count = 27

III. QoE Models for Mobile Cloud Games

For QoE modeling, we consider both *linear* and *non-linear* approaches to determine models that perform better but are lower in complexity (by performing F-tests), as these models should be used by stakeholders in real life in their products.

$$QoE = f_0 + f_1 \cdot RTT + f_2 \cdot PL + f_3 \cdot RJ + f_4 \cdot BJ + f_5 \cdot RTT \cdot PL \quad (1)$$

Linear Models: Our first model is described in Eq. 1, named **Linear Reg.**, is a multiple regression equation composed of single terms for each independent variable to model their relationship with QoE, along with an interaction between RTT and PL. All terms are statistically significant ($p < 0.05$), highlighting the importance of each tested network condition variable in the MCG modeling, including combined conditions (RTT, PL). This model is the most simple and should serve to verify the feasibility of modeling MCG using first-order only terms, similar to the work of [9]. Linear Reg. model was not enough to account for some of the conditions total variance (e.g., (RTT,PL) and RJ). Hence, we investigated interactions between model terms in pairs (e.g. RTT*RJ, PL*BJ) and triples (e.g. RTT*BJ*PL), following a brute-force approach. The term RTT * PL was the only combination that produced a valid coefficient. Thus, for our dataset, RTT and PL have dependency, while RJ and BJ are independent. Further, we identified that RJ and BJ terms are important terms, as their absence from the equation produces very low accuracy scores.

$$QoE = g_0 + g_1 \cdot RTT + g_2 \cdot PL + g_3 \cdot RJ + g_4 \cdot BJ + g_5 \cdot RTT^2 + g_6 \cdot RTT^2 \cdot PL + g_7 \cdot RTT \cdot PL^2 + g_8 \cdot RTT^2 \cdot PL^2 \quad (2)$$

Based on Eq. 1 we then investigated higher orders (i.e. second and third) for each term. As a result, we propose a second model (Eq. 2), named **Poly. Reg**, a multiple polynomial regression. It contains single terms for each network parameter as first order and the interaction between RTT and PL combined in different forms using both first and second degrees. To increase the explainability of the model, we considered the coefficient's p-value, the metrics R^2, and adjusted R^2, visually inspecting the distribution of residuals, and the model's complexity. To increase its accuracy, additional higher-order polynomial terms were necessary to account for all the RTT and PL conditions and their combined interaction. They were the majority of our dataset (that is, 19 out of 26 conditions). Both $RTT \cdot PL^2$, $RTT^2 \cdot PL$, and $RTT^2 \cdot PL^2$ support the PL conditions tested individually and together with RTT. All terms, except RTT^2, are statistically significant, suggesting that the combined conditions (RTT,PL) caused a different behavior in the network for MCG in addition to the RTT and PL conditions tested separately. We decided to keep RTT^2 since it helped to account for the slight nonlinear shape of the individual RTT conditions.

$$QoE = MaxQoE - I_{RP}(RTT, PacketLoss) - I_R(RandomJitter) - I_B(BusrtyJitter) \quad (3)$$

Non-linear Model: By learning about the most important non-linear interactions in the MCG modeling, and following the ITU modeling approach, we proposed a third model named **Non-lin. Reg.**, described in Eq. 3. It is composed of three individual functions presented in Eq. 4. Inspired by [8], we define the constant term MaxQoE equal to the best MOS possible that can be obtained during subjective tests. In our case, it was condition C0 (e.g. no degradation). The advantage of this approach is that if we preclude the terms from predicting MOS, for example, if RJ is 0, we can still predict the MOS as these factors are negatively additive.

$$\begin{aligned} I_{RP} &= h_1 \cdot (RTT \cdot PL)^{h_2} + h_3 \cdot PL^2 + h_4 \cdot PL + h_5 \cdot RTT \\ I_R &= h_6 + \frac{(h_7 - h_6)}{1 + e^{(\frac{RJ - h_8}{h_9})}} \\ I_B &= h_{10} \cdot BJ^{h_{11}} \end{aligned} \quad (4)$$

The first individual function (Eq. 4) I_{RP}, models RTT and PL and is a combination of first and second degree polynomials adjunct to a power function for the interaction term. We chose a power function, since it has the freedom of defining the curve shape from the exponent coefficient h_2, proving suitable to model the combined (PL,RTT) conditions. We also attempt to model this (RTT,PL) interaction in the same way as proposed in Linear Reg, but did not achieve a reasonable residual distribution and a lower R^2 value. The second individual function I_R, is a standard S-shaped logistic growth function composed of 4 coefficients to model RJ and matches its curve shape. Different S-shapes functions were also considered (e.g. Weibull, Gompertz equations) but producing very similar results or worse. The last function I_B, proposed for BJ cases is a power function with two coefficients necessary to control the curve's shape. In the next section, we present the statistical analyses for the three models.

IV. Results Analysis

The coefficients of the models are listed in Table II. Linear Reg. and Non-lin. Reg. models were trained using the same dataset that included 26 network conditions, while Poly. Reg. with a reduced dataset (25 network conditions) since RJ = 6std was removed (explanation in Section IV-B). The accuracy of the models is reported in Table III, in terms of degrees of freedom (DF) [16], root mean square error (RMSE), mean absolute error (MAE), Pearson's linear correlation (PLCC), coefficient of determination R^2 and adjusted R^2. Differences

TABLE II: Models' coefficients fitted values.

Model	Coefficients
Linear Reg.	f_0=4.0033, f_1=-0.0072, f_2=-0.0438, f_3=-0.1229, f_4=-0.0011, and f_5=-0.0043
Poly. Reg.	g_0=4.0157, g_1=-0.0091, g_2=0.0596, g_3=-0.1385, g_4=-0.0012, $g_5 = 1 \cdot 10^{-5}$, g_6=-0.0002, g_7=-0.0015, $g_8 = 4 \cdot 10^{-5}$
Non-Lin. Reg.	h_1=0.1404, h_2=0.4746, h_3=0.003, h_4=-0.0936, h_5=0.0074, $MaxQoE$=4 h_6=-0.1492, h_7=1.5913, h_8=6.2635, h_9=-0.1043,h_{10}=0.0445, h_{11}=0.5083

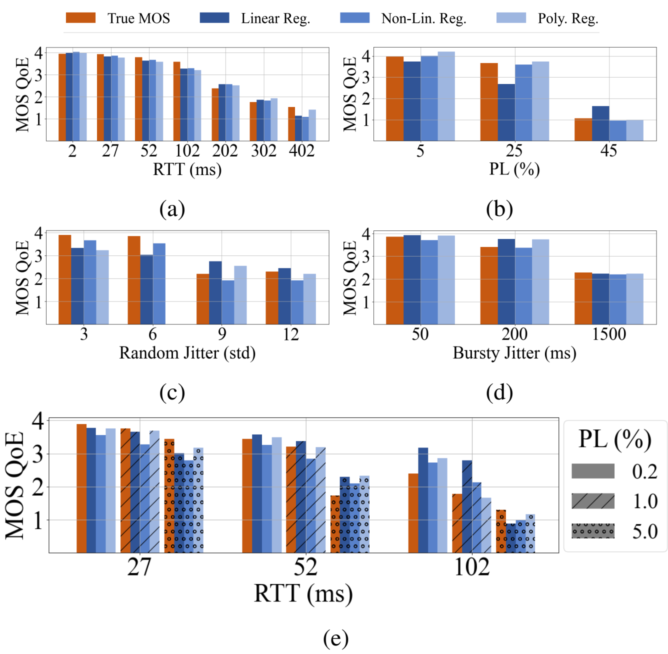

Fig. 1: Model's prediction in vertical bars per condition.

between model prediction output and the true MOS value per condition can be visualized in Fig. 1. The residual analyzes were performed by visually inspecting their distribution in the histogram and the normal probability plots in Fig. 2.

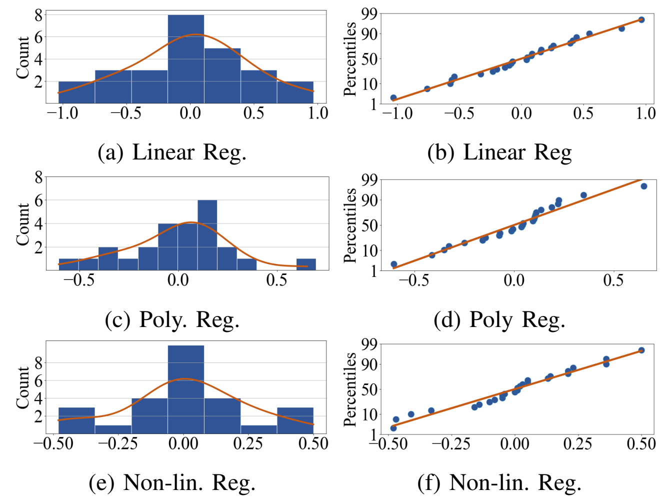

Fig. 2: Residuals in histogram and normal probability plot.

A. Linear Regression Model

The first mode, Linear Reg. (Eq. 1), has the lowest accuracy compared to other models (see Table III); however, it still proves to be capable of producing a fairly good model explainability (with almost 80% R^2 value) for most cases. This result is surprising, since this model contains only single-order terms and one interaction. The residual plot in Fig. 2a resembles a normal distribution, and the normal probability graph in 2b follows a straight line, clearly showing a good fit of the model. When it comes to the residual sizes (in Fig. 1), the most significant residuals belong to the following conditions:

- Condition (RTT=102ms,PL=1%) with rsid = -1 and (RTT=102ms,PL=0.2%) with rsid = -0.75 (in Fig. 1e).
- Condition (PL=24%) with rsid = 0.97 (in Fig. 1b).
- Condition (RJ=6std) with rsid = 0.75 (in Fig. 1c).

These results were expected and show that Linear Reg. did not model the non-linearity of combined PL and RTT very well. However, for most of the conditions (n=22) the residual (rsid) $=<$ $|0.55|$ which is reasonably good.

B. Polynomial Regression Model

Poly. Reg. model's (Eq. 2 has the smaller RMSE and MAE along with Non-Lin. Reg. (see Table III). The model showed good performance with fewer coefficients compared to Non-Lin. Reg. However, Poly. Reg. was trained with a reduced version of our dataset, motivated from our initial analyses. At first, we identified an outlier in the normal probability plot for the condition RJ = 6std with a considerable deviation from the line. Since in the RJ tests (see Fig. 1c), condition RJ = 6std (MOS = 3.82) had a score similar to condition RJ = 3std (MOS = 3.87), we decided to remove RJ= 6std without compromising the dataset and the modeling of the RJ variable significantly. After this change, we had a better fit of the model with R^2 of 0.94% (it was 0.91% before outlier management). Afterwards, the residual histogram for Poly. Reg. (Fig. 2c) and the normal probability plot (Fig. 2d) were reasonably well distributed.

Regarding the size of the residuals, the most significant was in (RTT=52ms,PL=5%) with rsid=-0.6, which represents the combined (RTT,PL) conditions; this residual is not as large as Linear Reg., but still demonstrates that the polynomials were not able to fully account for all the RTT and PL conditions, which is expected. The remaining conditions had rsid $<$ $|0.51|$. Still Figs. 1e and 1c show that Poly. Reg. performs better than Linear Reg. in cases where PL and RTT have interaction, and regarding jitter conditions.

C. Non-linear Regression Model

Our Non-Lin. Reg. model provided the highest accuracy (see Table III), with slightly better MAE compared to Poly. Reg. In Figs. 2e and 2f, the residual and normal probability plots are depicted, suggesting that the residuals are normally distributed. As Non-Lin. Reg. provided the best fit, we observed minimal residual values below $|0.5|$, the lowest among the proposed models. The largest residual was for the condition (RTT = 27 ms, PL = 5%) and (RTT = 52 ms, PL = 5%) with rsid = 0.49, which belong to the combined conditions

TABLE III: Models' performance from various metrics.

Model	DF	RMSE	MAE	PLCC	R2	R2-Adj
Linear Reg.	20	0.45	0.35	0.89	0.79	0.73
Poly. Reg.	17	0.24	0.19	0.97	0.94	0.9
Non-Lin. Reg.	15	0.24	0.18	0.97	* [2].	*

TABLE IV: F-Test to assess pair-wise models comparison.

Test ID	Model	SSR	F Statistics	P >|T|	DF	Hypothesis
1	Poly. Reg.	1,495	–	–	17	–
1	Non-Lin. Reg.	1,480	0,14931	0,70462	15	Reject Non-Lin. Reg.
2	Linear Reg.	5,199	–	–	20	–
2	Non-Lin. Reg.	1,480	753,954	0,00102	15	Reject Linear Reg.
3	Linear Reg.	5,199	–	–	20	–
3	Poly. Reg.	1,495	9.91422	0,00031	17	Reject Linear Reg.

(RTT, PL), see (Fig. 1e). Regarding the RJ conditions (in Fig. 1c), Non-Lin. Reg. could nearly match their true value. The modeling of RJ as an S-shaped function (Eq. 3) proved to be the best technique rather than first or second degree polynomials. Therefore, the results of our experiment suggest that the Linear Reg. could not account well for all the RJ and combined conditions (RTT,PL), while Poly. Reg and Non-Lin. Reg. resulted in better fits.

D. Best Model Selection

Since Poly. Reg. and Non-lin. Reg. models have very similar accuracy, although Poly. Reg. slightly worse regarding MAE but with far fewer terms, it becomes hard to decide which one has a better fit. To answer this question, we performed a pairwise F-test [16]. This test considers the trade-off between the models' complexity in degrees of freedom (DF) and the models' prediction performance based on the sum of squared residuals (SSR). The models are tested in pairs (identified by the test ID), and the null hypothesis states that each pair's second row has the best statistical fit (see Table IV).

From the hypothesis column, it can be seen that Poly. Reg. has the best fit, followed by Non-Lin. Reg. and Linear Reg. Non-Linear Reg. and Linear Reg. were rejected when compared with Poly. Reg. This means that although Non-Lin. Reg. has a better prediction accuracy, Poly. Reg. is statistically better due to its lower number of terms. In contrast, between Linear Reg. and Non-Lin Reg., the non-linear model had a statistically better fit. Therefore, we conclude that the model Poly. Reg. is the best model for MCG.

V. Conclusion

This paper proposes three different models to predict QoE for MCG, for static and mobility aspects never considered before; and for the first time, we release the dataset used in model training. Further, we found that 1) Surprisingly, simple multiple linear regression can model and predict different network behaviors for MCG fairly well $R^2 = 0.79$. However, polynomial regression and non-linear (NL) models offer highest accuracy with $R^2 = 0.94$ and RMSE of 0.24, respectively; 2) The network conditions for RJ were better modeled using NL functions rather than linear/polynomial regression; 3) Individual and combined (RTT, PL) conditions were difficult to model, as their combination caused a different MOS degradation pattern than their individual tests. Future studies are advised to consider both cases; 4) Our extended polynomial regression model proved to be statistically the best fit.

[2] R^2 and R^2-Adj was not calculated for NL models since [17]

References

[1] S. Neumeier, E. A. Walelgne, V. Bajpai, J. Ott, and C. Facchi, "Measuring the Feasibility of Teleoperated Driving in Mobile Networks," in *2019 Network Traffic Measurement and Analysis Conference (TMA)*, Jun. 2019, pp. 113–120.

[2] A. N. Mian, T. Liaqat, and A. Hameed, "A Fresh Look into the Handoff Mechanism of IEEE 802.11s under Mobility," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–5.

[3] J. Prokkola, P. H. J. Perala, M. Hanski, and E. Piri, "3G/HSPA Performance in Live Networks from the End User Perspective," in *2009 IEEE International Conference on Communications*, Jun. 2009, pp. 1–6, iSSN: 1938-1883.

[4] K. Mitra, A. Zaslavsky, and C. Ahlund, "Context-Aware QoE Modelling, Measurement, and Prediction in Mobile Computing Systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 920–936, May 2015.

[5] I. Slivar, L. Skorin-Kapov, and M. Suznjevic, "Cloud gaming QoE models for deriving video encoding adaptation strategies," in *Proceedings of the 7th International Conference on Multimedia Systems*. Klagenfurt Austria: ACM, May 2016, pp. 1–12.

[6] I. Slivar, M. Suznjevic, and L. Skorin-Kapov, "The impact of video encoding parameters and game type on QoE for cloud gaming: A case study using the steam platform," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015, pp. 1–6.

[7] H.-J. Hong, C.-F. Hsu, T.-H. Tsai, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "Enabling Adaptive Cloud Gaming in an Open-Source Cloud Gaming Platform," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 12, pp. 2078–2091, Dec. 2015.

[8] S. Zadtootaghaj, S. Schmidt, and S. Möller, "Modeling Gaming QoE: Towards the Impact of Frame Rate and Bit Rate on Cloud Gaming," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2018, pp. 1–6, iSSN: 2472-7814.

[9] I. Slivar, M. Suznjevic, L. Skorin-Kapov, and M. Matijasevic, "Empirical QoE study of in-home streaming of online games," in *2014 13th Annual Workshop on Network and Systems Support for Games*, Dec. 2014, pp. 1–6, iSSN: 2156-8146.

[10] M. Claypool and D. Finkel, "The effects of latency on player performance in cloud-based games," in *2014 13th Annual Workshop on Network and Systems Support for Games*. Nagoya, Japan: IEEE, Dec. 2014, pp. 1–6.

[11] S. Schmidt, S. Zadtootaghaj, S. S. Sabet, and S. Möller, "Modeling and Understanding the Quality of Experience of Online Mobile Gaming Services," in *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*, Jun. 2021, pp. 157–162, iSSN: 2472-7814.

[12] S. Wang and S. Dey, "Modeling and Characterizing User Experience in a Cloud Server Based Mobile Gaming Approach," in *GLOBECOM 2009-2009 IEEE Global Communications Conference*, 2009, pp. 1–7.

[13] Y. Liu, S. Wang, and S. Dey, "Modeling, characterizing, and enhancing user experience in Cloud Mobile Rendering," in *2012 International Conference on Computing, Networking and Communications (ICNC)*, Jan. 2012, pp. 739–745.

[14] H. S. Rossi, K. Mitra, C. Åhlund, I. Cotanis, N. Ögren, and P. Johansson, "ALTRUIST: A Multi-platform Tool for Conducting QoE Subjective Tests," in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, Jun. 2023, pp. 99–102, iSSN: 2472-7814.

[15] H. S. Rossi, N. Ögren, K. Mitra, I. Cotanis, C. Åhlund, and P. Johansson, "Subjective Quality of Experience Assessment in Mobile Cloud Games," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Dec. 2022, pp. 1918–1923.

[16] H. J. Motulsky and L. A. Ransnas, "Fitting curves to data using nonlinear regression: a practical and nonmathematical review," *The FASEB Journal*, vol. 1, no. 5, pp. 365–374, 1987.

[17] A.-N. Spiess and N. Neumeyer, "An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research," *BMC Pharmacology*, vol. 10, no. 1, p. 6, Jun. 2010.