# Applicability of Existing Objective Metrics of Perceptual Quality for Adaptive Video Streaming

Jacob Søgaard, Lukáš Krasula,
Muhammad Shahid, Dogancan Temel,
Kjell Brunnström, and Manzoor Razaak

DTU Technical University of Denmark
IRCCyN
Blekinge Tekniska Högskola · BTH ·
Mittuniversitetet MID SWEDEN UNIVERSITY
Czech Technical University in Prague
Kingston University London
Georgia Institute of Technology
PART OF RI. SE
SWEDISH ICT ACREO

# Introduction

- Recently video streaming services has grown in popularity
- Nowadays, most services use HTTP based adaptive streaming
- Degradations quite different from other type of distribution
- Slowly varying quality changes on a fairly long timescale (2 sec – 1 min)

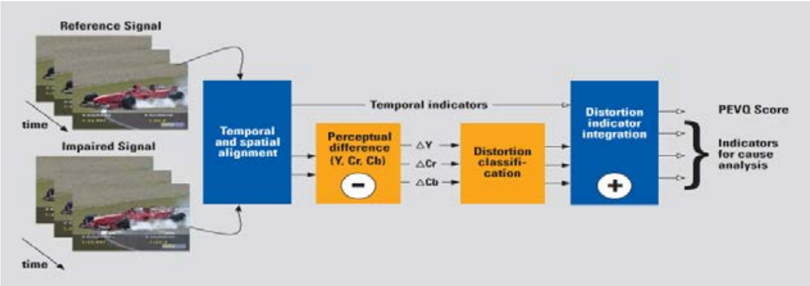
# Introduction

- Many objective video quality methods has been proposed
- Performance results have been published
- Their limitations not always clearly known
- This investigation: evaluated some published methods for adaptive video streaming

# Metrics

Perceptual Evaluation of Video Quality

*PEVQ:*




FR metric and part of ITU-T. Rec J.247 - objective metrics for multimedia

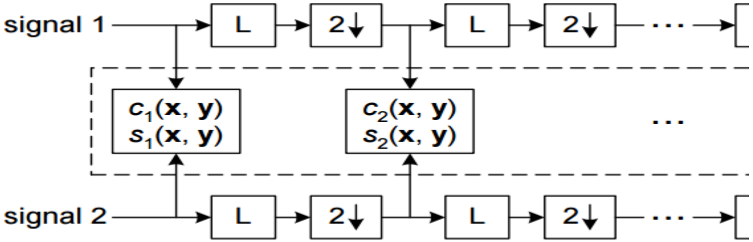
$$PSNR \quad = \quad 10 \cdot \log\left(\frac{MAX^2}{MSE}\right)$$

Luma component used

$$SSIM = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X + \mu_Y + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}$$

Analyzing structural information. Default parameters used.

*MS-SSIM:*




Multiscale version of SSIM. Default parameters used.

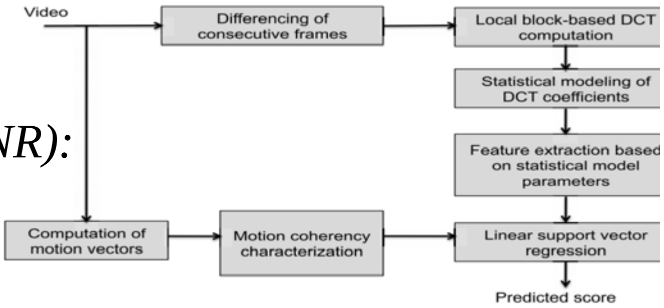
SWEDISH ICT ACREO

# Metrics

Standardized ITU-T Rec. J.144 RR-metric

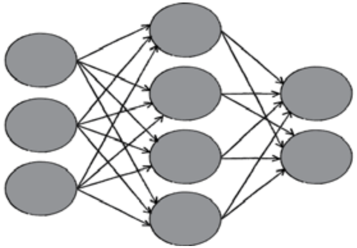
$$
\begin{aligned}
VQM = & -0.21 \cdot si\_loss \\
& +0.60 \cdot hv\_loss \\
& +0.25 \cdot hv\_gain \\
& +0.02 \cdot chroma\_spread \\
& -2.34 \cdot si\_gain \\
& +0.04 \cdot ct\_ati\_gain \\
& +0.01 \cdot chroma\_extreme
\end{aligned}
$$

*Video BLIINDS (NR):*




NR machine learning metric natural scene statistics

*VQM_VFD:*



Improved VQM with Variable Frame Delays and FR metric

SWEDISH ICT ACREO

# Test video sequences

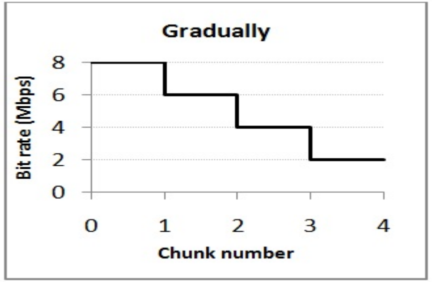


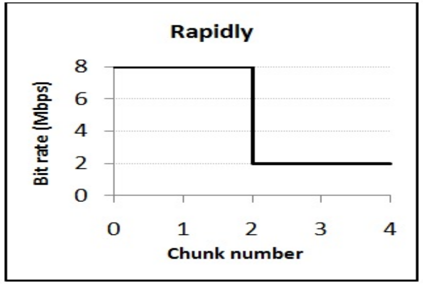



- 7 commercial content: Movie, Documentary, Sport, News, Music
- 6 min; originally 1080p Bluray video, 24/50 fps
- Different spatial and temporal characteristics

# Adaptation study parameters

- Switching behavior
  - Period (chunk length): 2 sec and 10 sec
  - Amplitude (difference between consequtive quality levels)
    gradual vs. rapid switching
  - Decreasing and increasing




- Adaptation dimension
  - Video quality (QP)
  - H.264/AVC
  - 4 streams: 5Mbps, 3Mbps, 1Mbps, 600kbps
  - 1280×720/ 25fps

→ **12 Test conditions in total**
→ **132 individual events (PVS) rated by test subjects**
→ **MOS based on > 60 test subjects**

SWEDISH ICT ACREO

# Performance metrics

Standardized ITU-T Rec. P.1401

$$PLCC = \frac{\sum_{i=1}^{N}(MOS_i - \overline{MOS}) \times (MOSp_i - \overline{MOSp})}{\sqrt{\sum_i (MOS_i - \overline{MOS})^2} \times \sqrt{\sum_i (MOSp_i - \overline{MOSp})^2}},$$

$$OR = \frac{n_{\text{outlier}}}{N},$$

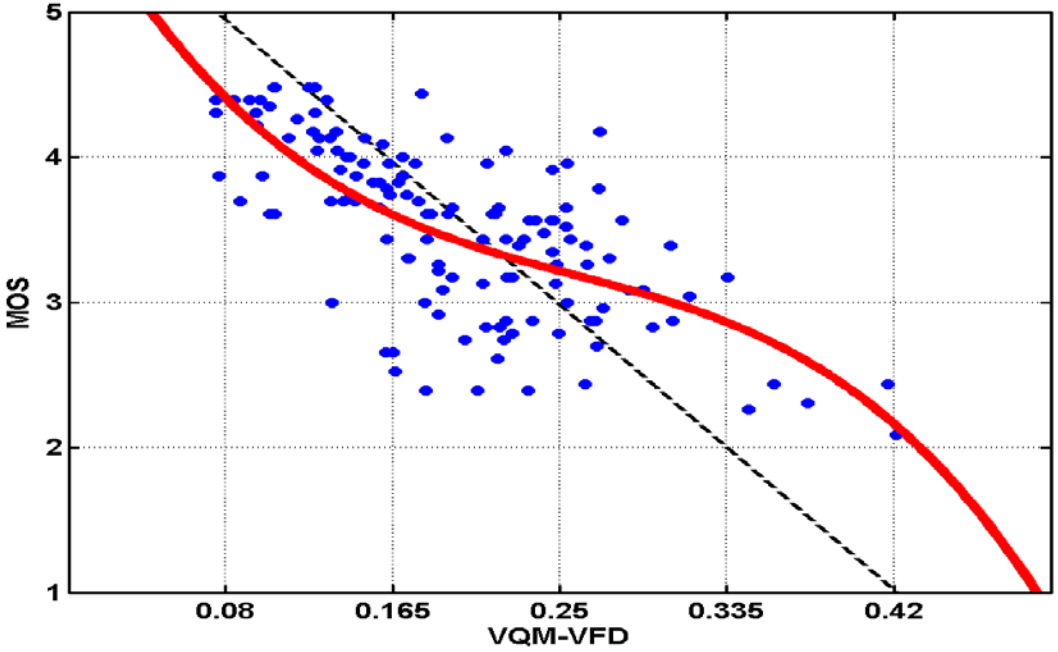
$$SROCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)},$$

$$RMSE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} MOS_i - MOSp_i}.$$

*Resolving Power (RP):* Significance of metric differences for sequences with statistically different MOS. Gives 95% confidence of a difference.

ITU-T Rec. J.149

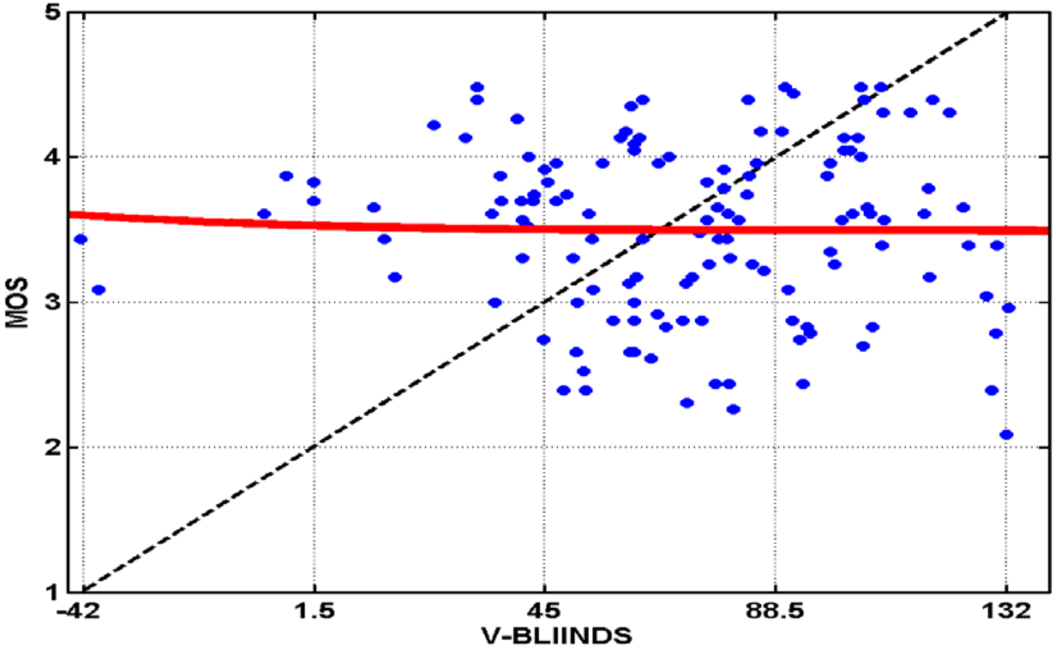
# SCATTER PLOT: VQM-VFD




monotonic regression with 3rd order polynomial function

SWEDISH ICT ACREO

# SCATTER PLOT: VIDEO BLIINDS




monotonic regression with 3rd order polynomial function

# Performance

Table 1: Measures of performance.

| | PLCC | SROCC | OR | RMSE | RP |
|---|---|---|---|---|---|
| **PSNR** | 0.46 | 0.39 | 0.50 | 0.53 | 0.31 |
| **SSIM** | 0.55 | 0.54 | 0.47 | 0.49 | 0.25 |
| **MSSSIM** | 0.64 | 0.64 | 0.39 | 0.45 | 0.28 |
| **VQM** | 0.56 | 0.54 | 0.40 | 0.49 | 0.26 |
| **VQM-VFD** | **0.69** | **0.67** | **0.30** | **0.43** | **0.23** |
| **PEVQ** | 0.33 | 0.19 | 0.51 | 0.56 | **0.23** |
| **V-BLIINDS** | 0.02 | 0.02 | 0.54 | 0.59 | 1.00 |

# Performance

**Table 2: Test of significant differences for SROCC.**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| PSNR (1) | | | | | - | | + |
| SSIM (2) | | | | | | + | + |
| MS-SSIM (3) | | | | | | + | + |
| VQM (4) | | | | | | + | + |
| VQM-VFD (5) | + | | | | | + | + |
| PEVQ (6) | | - | - | - | - | | |
| V-BLIINDS (7) | - | - | - | - | - | | |

Significance test ITU-T Rec P.1401 and
Bonferroni multiple-comparison compensation

SWEDISH ICT ACREO

# Performance

Table 3: Test of significant differences for RMSE.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| PSNR (1) | | | | | | | |
| SSIM (2) | | | | | | | |
| MS-SSIM (3) | | | | | | | + |
| VQM (4) | | | | | | | |
| VQM-VFD (5) | | | | | | + | + |
| PEVQ (6) | | | | | - | | |
| V-BLIINDS (7) | | | - | | - | | |

# Discussion

- No method performs well
- SSIM, MS-SSIM perform slightly better than PSNR – but not significantly
- VQM-VFD performs overall best

# Discussion

- V-BLIINDS are NR models based on machine learning not trained for this case
- PEVQ – dataset definitely out of scope trained on much severe transmission errors
- Videos varying length up to 40 sec – additional challenge for the methods

# Conclusion

- No method performed well
- Caution should be used when using metrics not designed for a particular scope

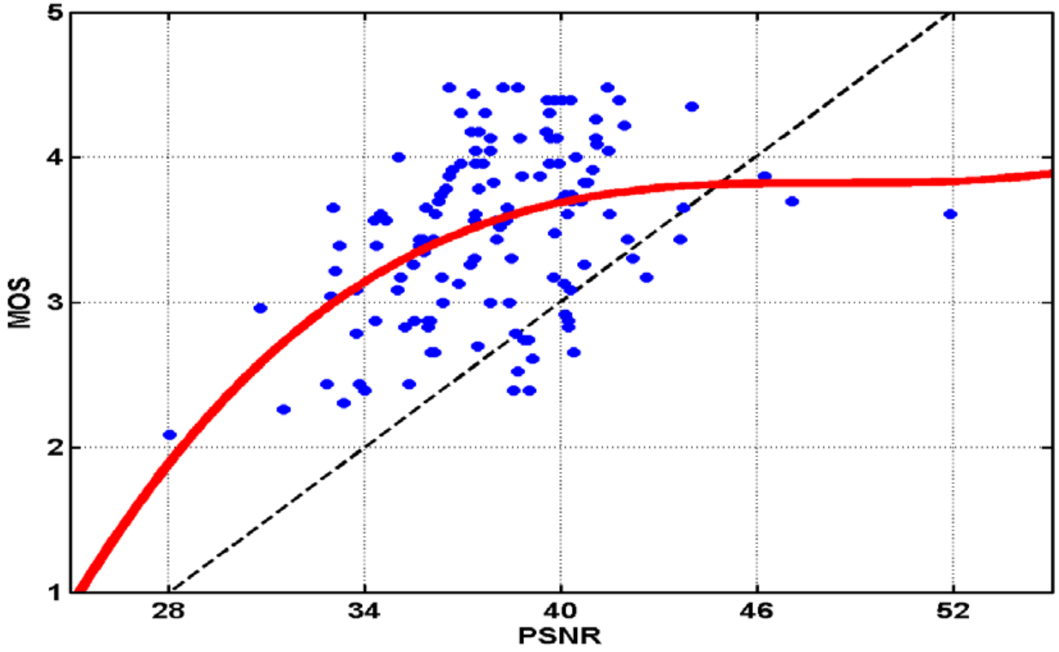
# More on this topic

- Subjective Analysis and Objective Characterization of Adaptive Bitrate Videos
- HVEI@4.30pm later today

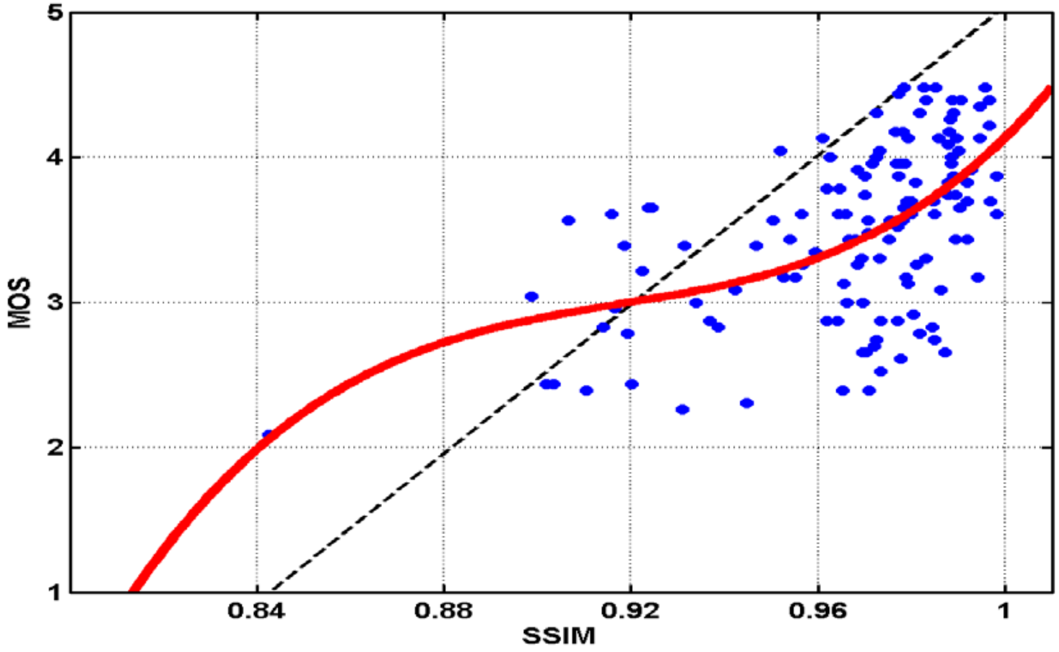
# THANK YOU

- Acreo’s work was funded by Vinnova (Sweden’s innovation agency) and EIT ICT Labs, which is hereby gratefully acknowledged.
- The support of Multimedia and Sensors lab (MSL) of Center for Signal and Information processing group at Georgia Tech is also gratefully acknowleded.

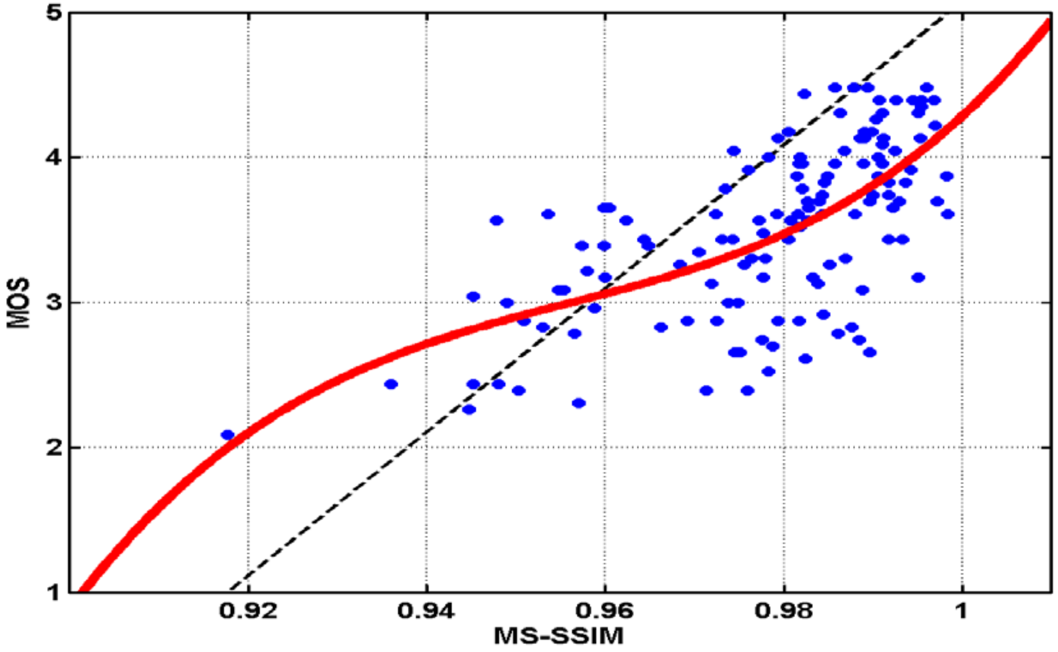
# SCATTER PLOT: PSNR



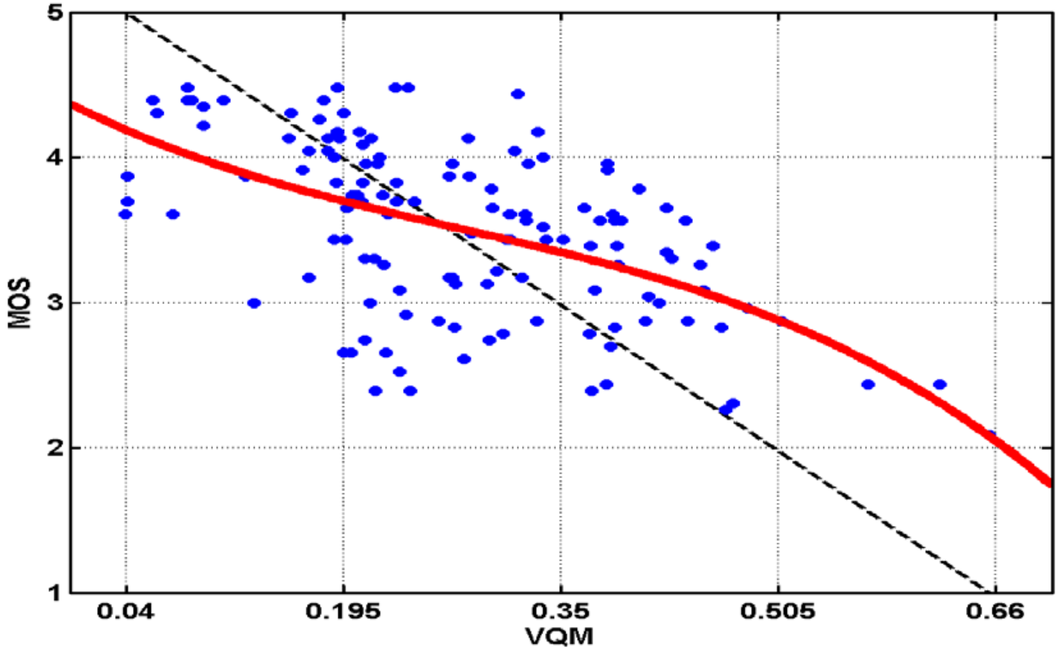
# SCATTER PLOT: SSIM



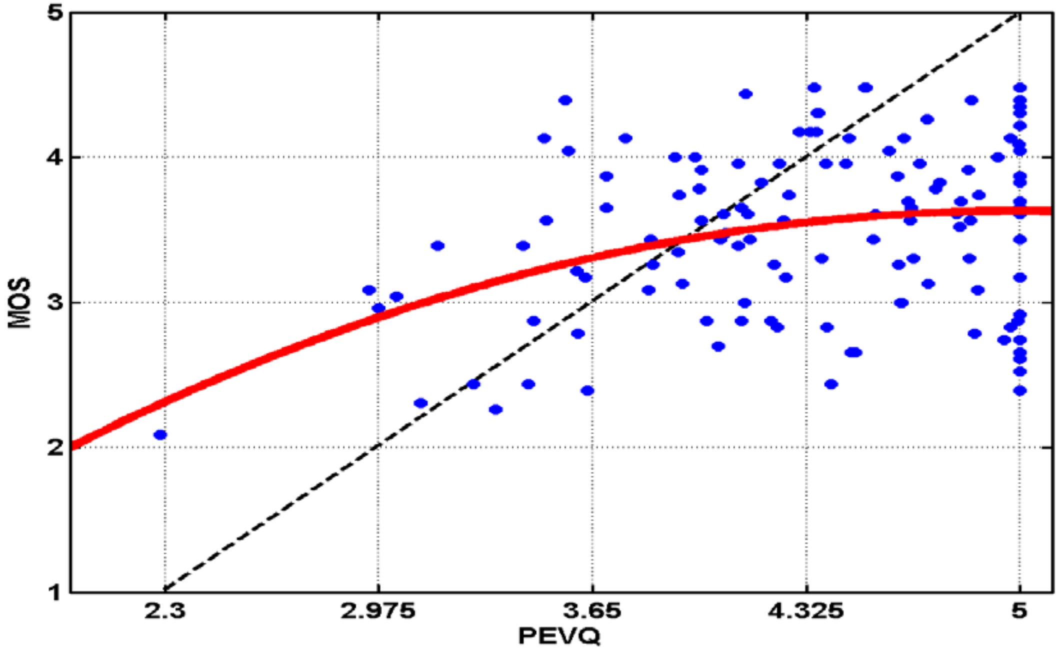
# SCATTER PLOT: MS-SSIM



# SCATTER PLOT: VQM



# SCATTER PLOT: PEVQ



# WWW.ACREO.SE

**Read more about our cases at 2014.swedishict.se**

**Follow us @AcreoSwedishICT**

WWW.ACREO.SE