



SNV

Status update and presentation of current
research activities

VQEG - JEG-Hybrid group

Outline

- Project Hodor
- Scope of the project
- Preliminary results

Project Hodor

- Joint collaboration between Sky (and affiliates) and JEG-Hybrid
- To investigate the behaviour of various VQ metrics, both open-source and commercial VQ metrics
- On industry-grade video signals of different types

Project Hodor: Scope currently under investigation

- Total 46 sequences, eight ABR profiles = 368 test conditions x 8 VQ metrics = 2,944 data points

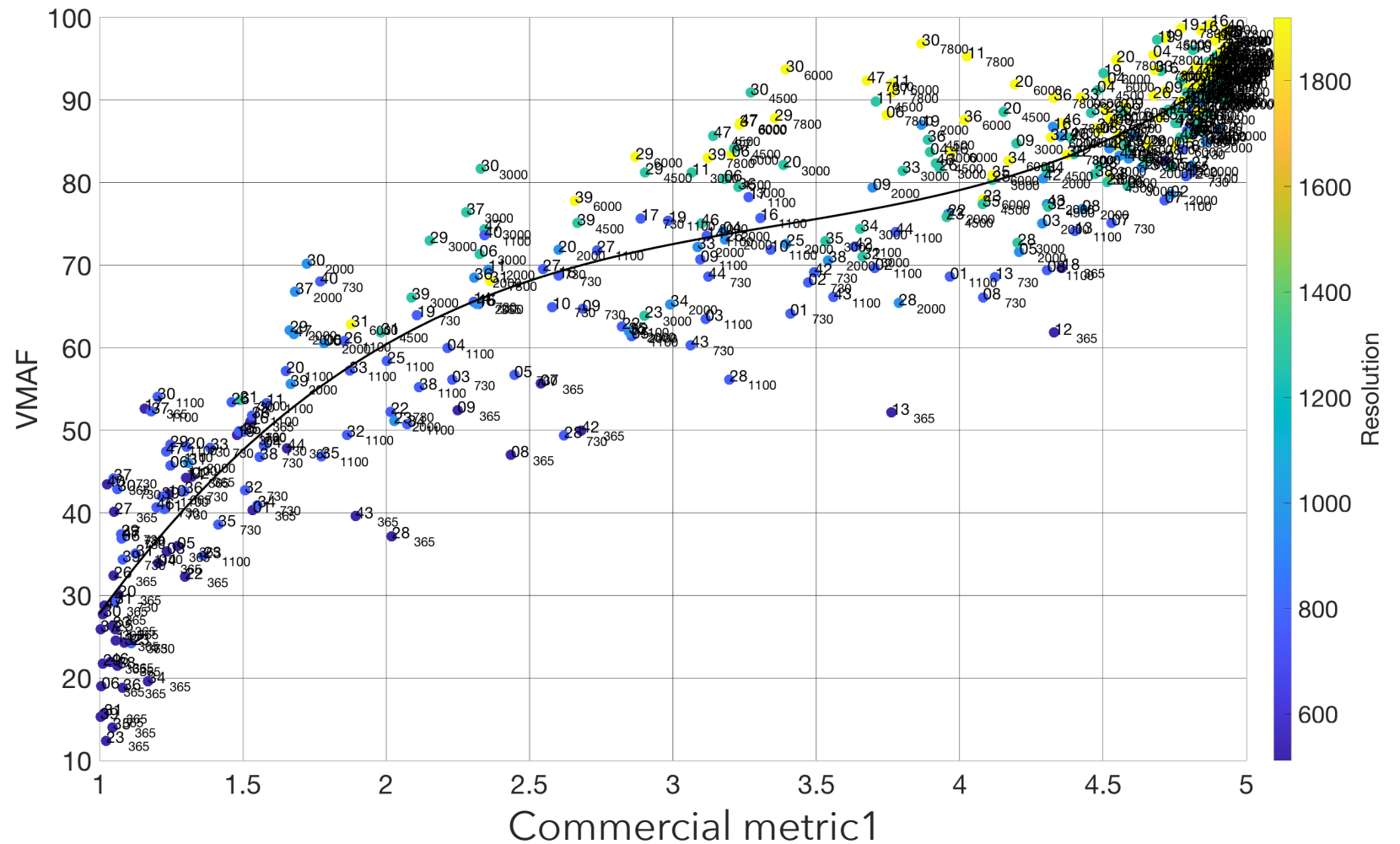
Coding	AVC	HEVC
Resolution	Adaptive streaming 1080	4K
Rate control	Constant bitrate	Constant quality
	Proprietary optimisations	Variable bit rate
Objective metric	VMAF	VQ metric 1
	VQ metric 2	MS-SSIM, SSIM
	VIF / PSNR	XPSNR _y

Scope of the project

- To understand what makes the quality of a PVS difficult to evaluate objectively
- How do the metrics rank in terms of accuracy
 - Current proposal: ranking based on disagreement analysis
- Finding conditions where metrics need to improve

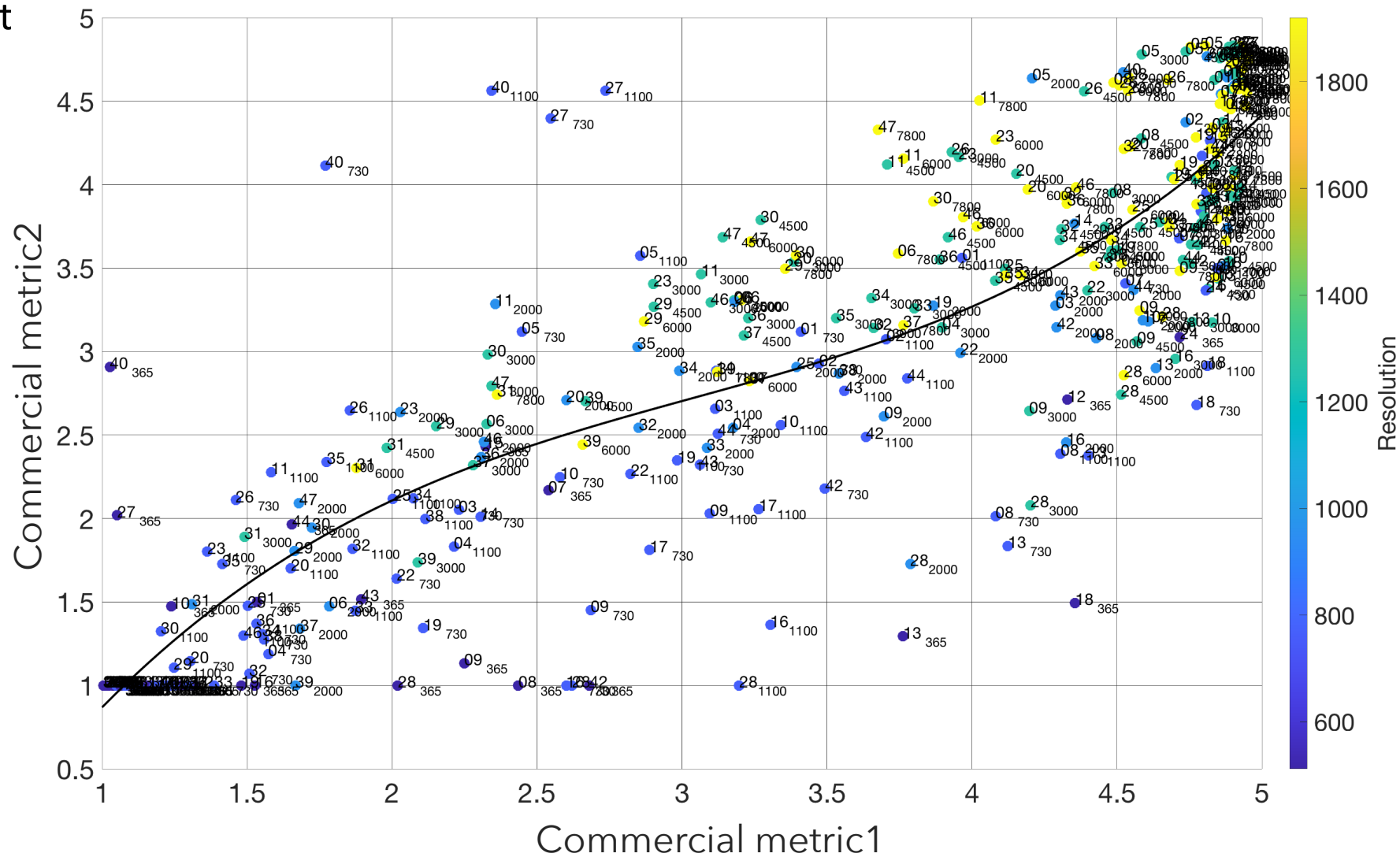
Results: Scatter plots comparing VMAF and commercial metric1

- Scatter plots comparing the VQ scores i.e. VMAF vs commercial metric1
- Fitting done with a third order polynomial function. Each point is a PVS, with SRC number and bitrate.
- Some outliers are present



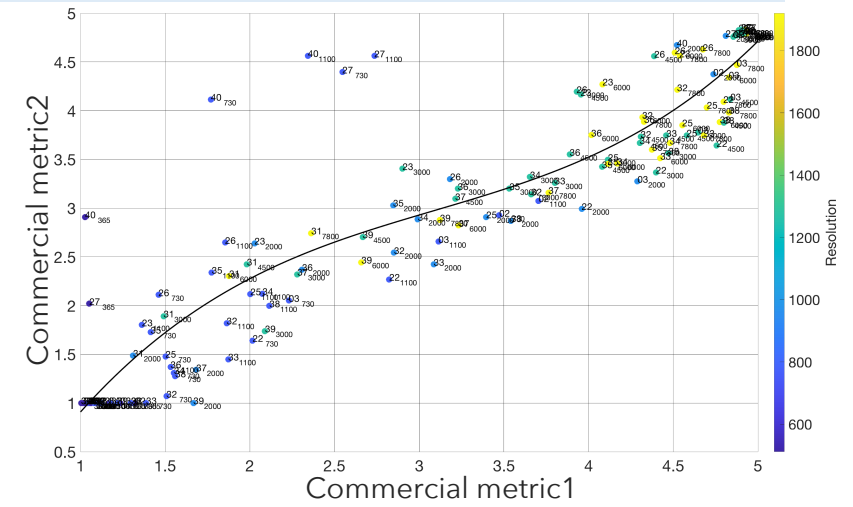
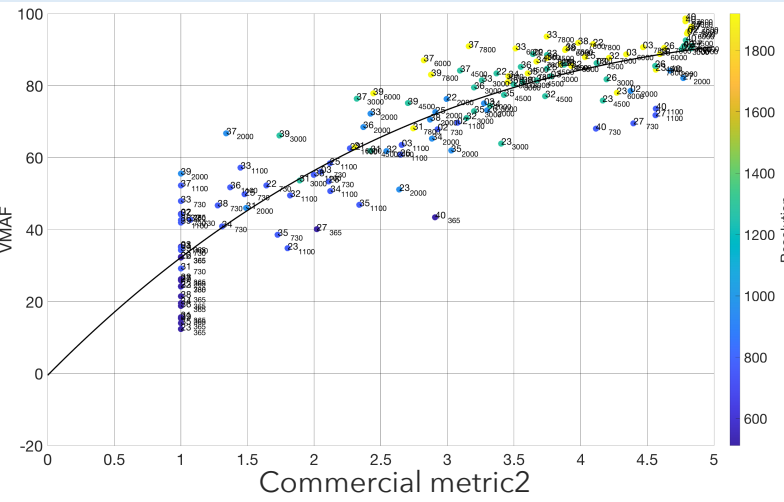
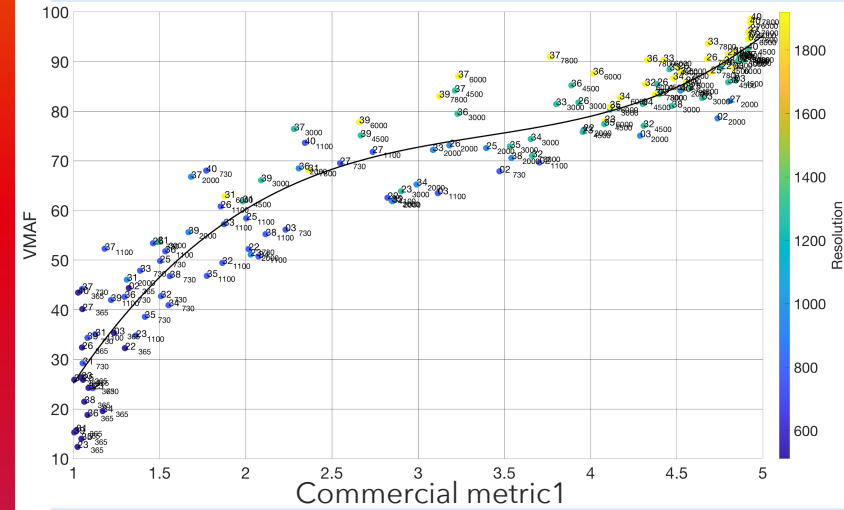
Results: Scatter plots comparing commercial metrics 1 & 2

- Scatter plots comparing the VQ scores i.e. commercial metric 1 and commercial metric 2
- Fitting done with a third order polynomial function. Each point is a PVS, with SRC number and bitrate.
- Some outliers are present

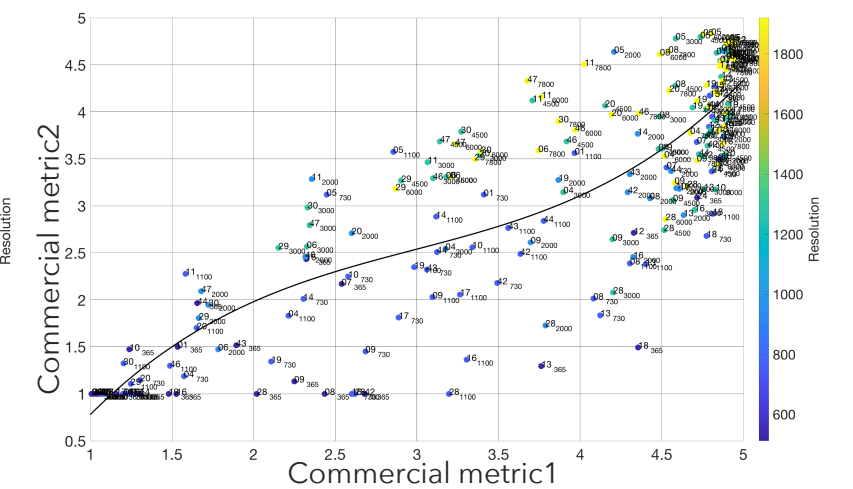
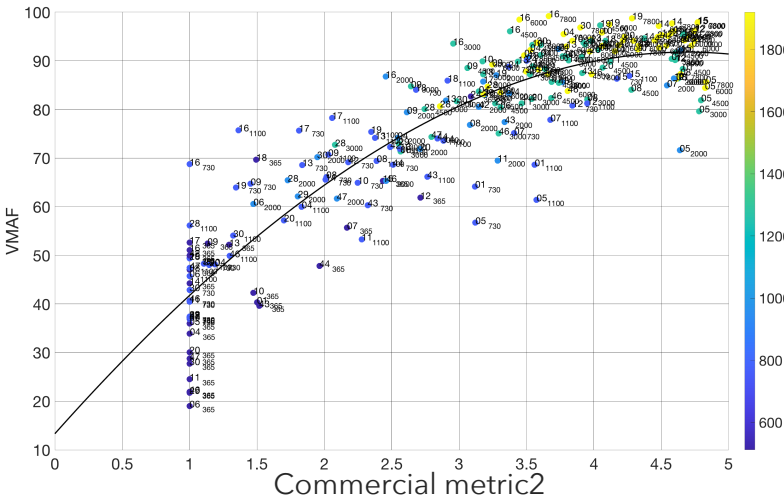
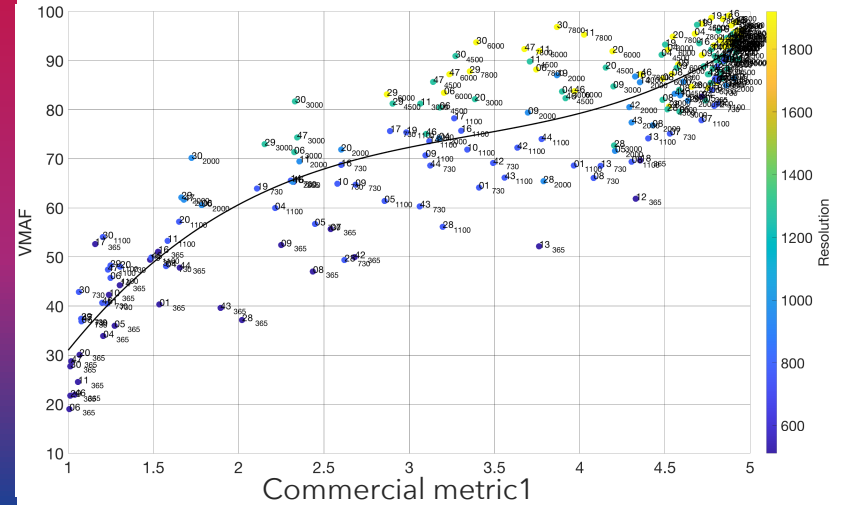


Split by Genre: Scatter plots comparing VMAF, commercial metrics 1 & 2

Sports

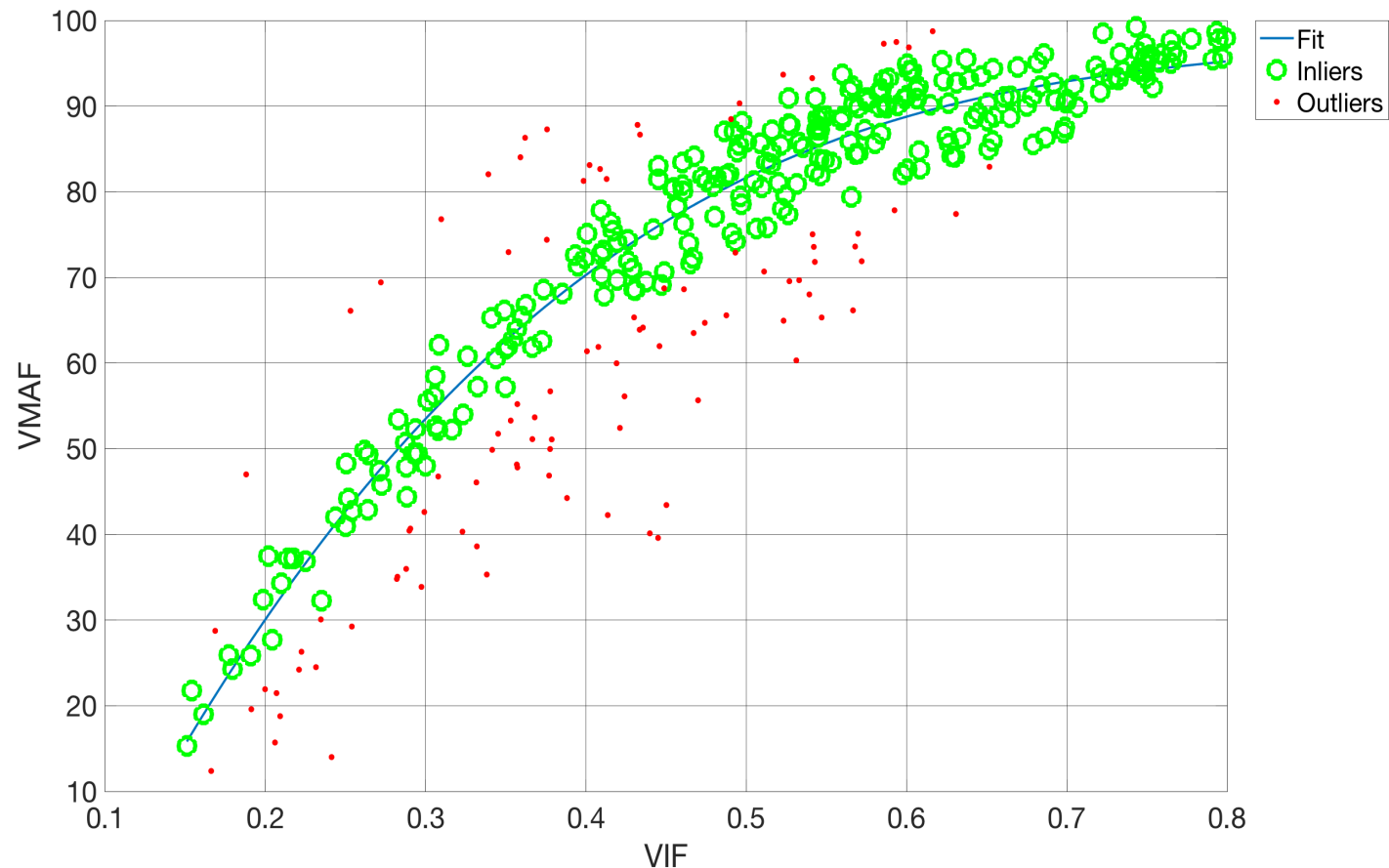


Movies

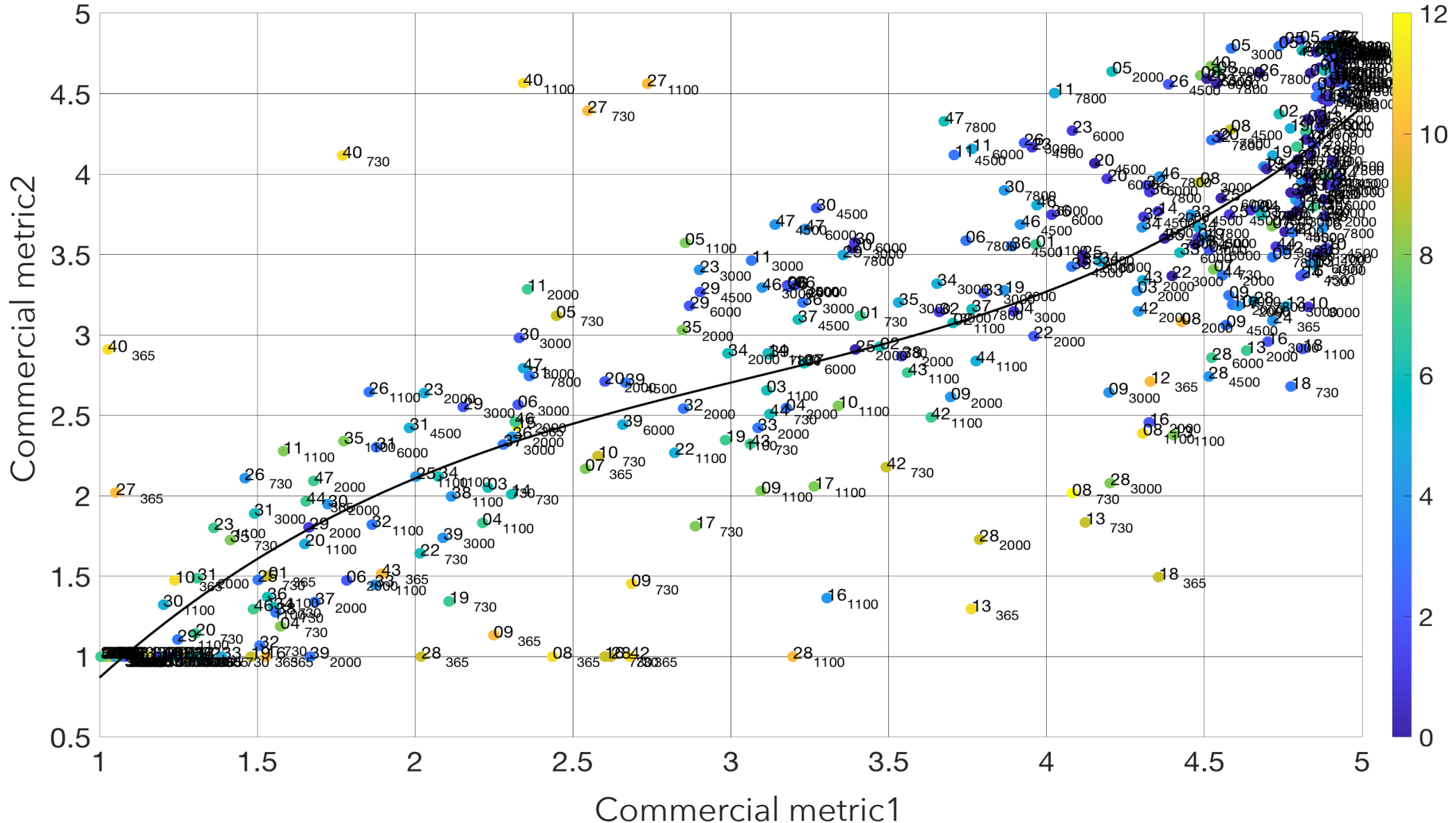


Disagreement analysis

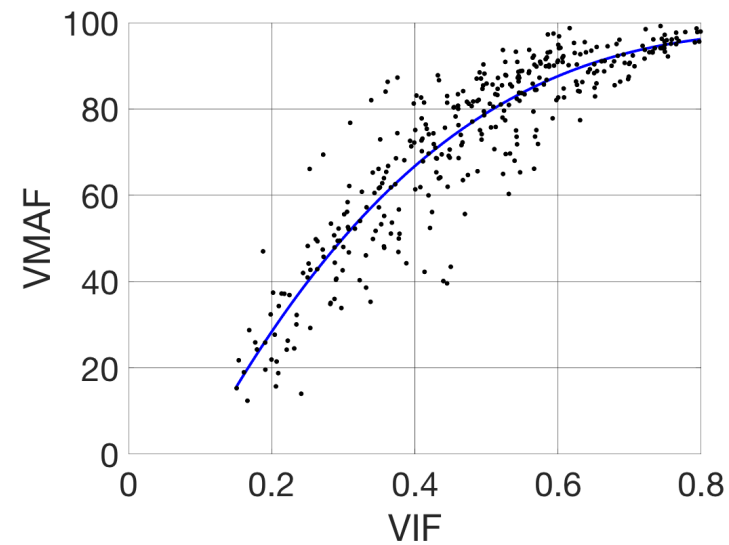
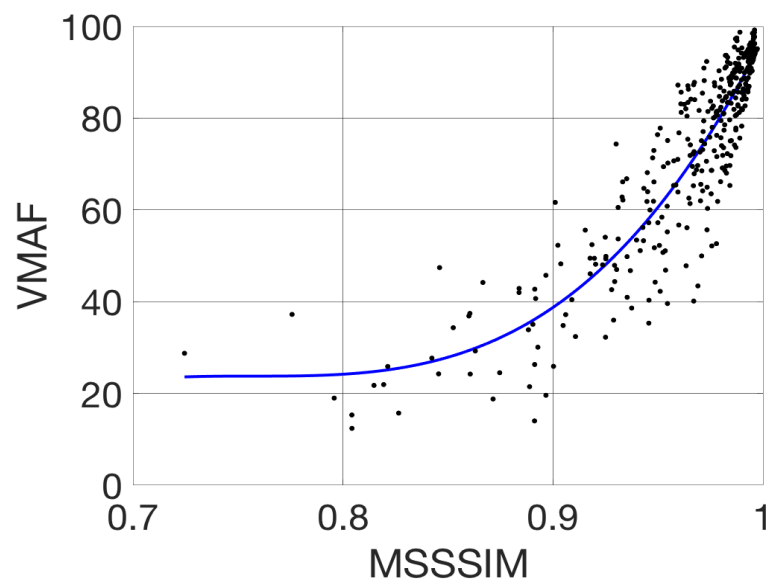
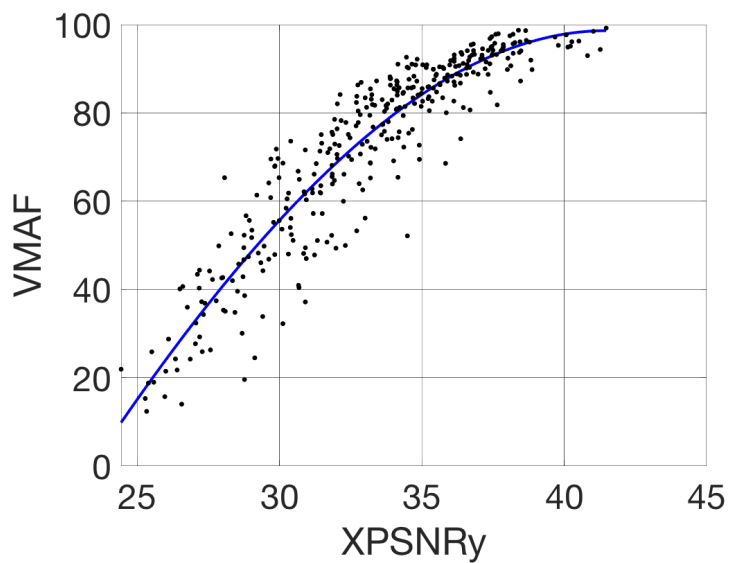
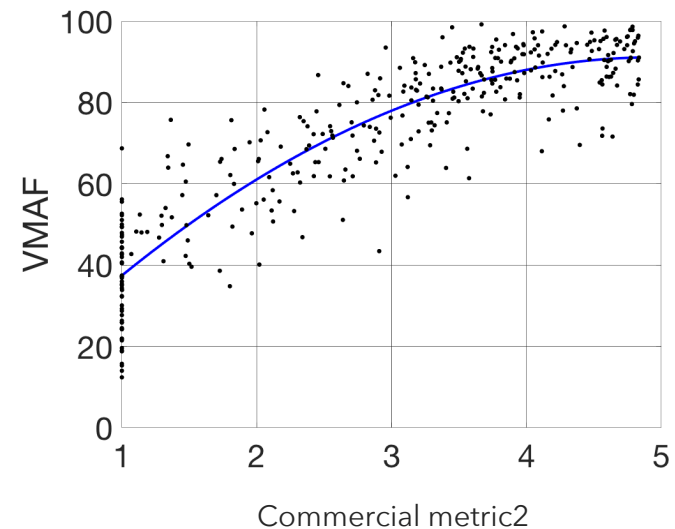
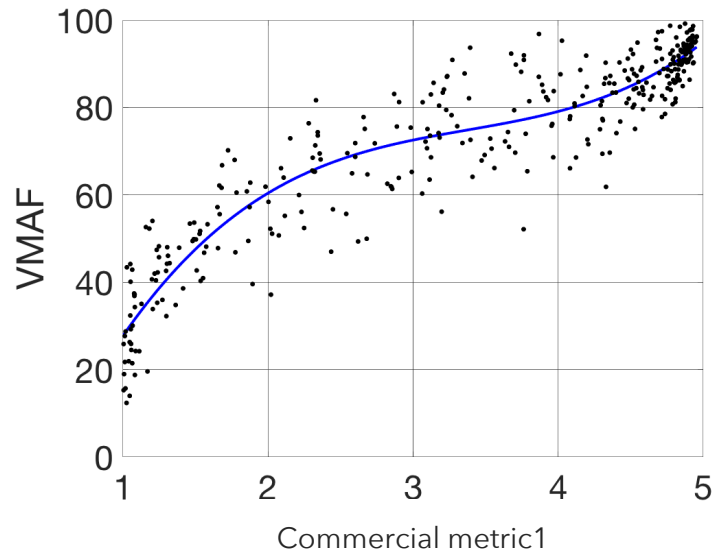
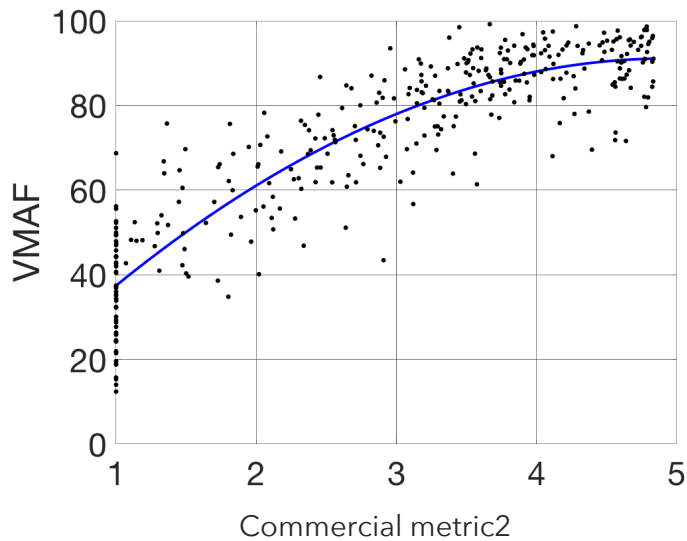
- The disagreement for each sequence is measured as the number of couples of metrics that do not agree on the quality of such sequence.
 - Two metrics VQ1 and VQ2 are in disagreement (far from each other) if $|VQ2 - P(VQ1)| > \text{threshold_VQ2}$ where $P(.)$ is the fitting polynomial function obtained using the RANSAC algorithm



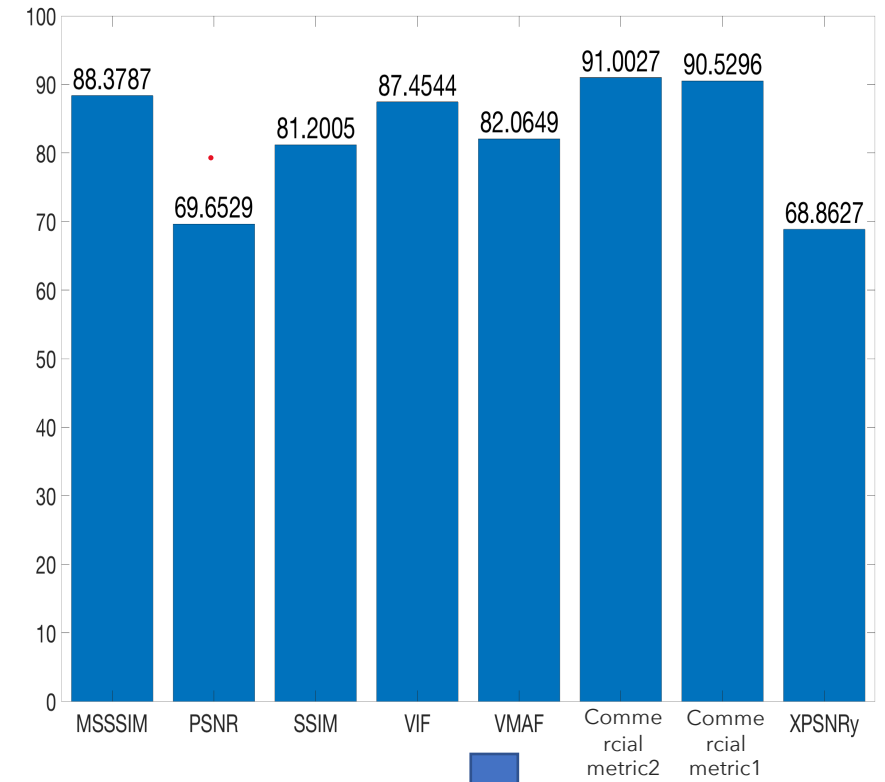
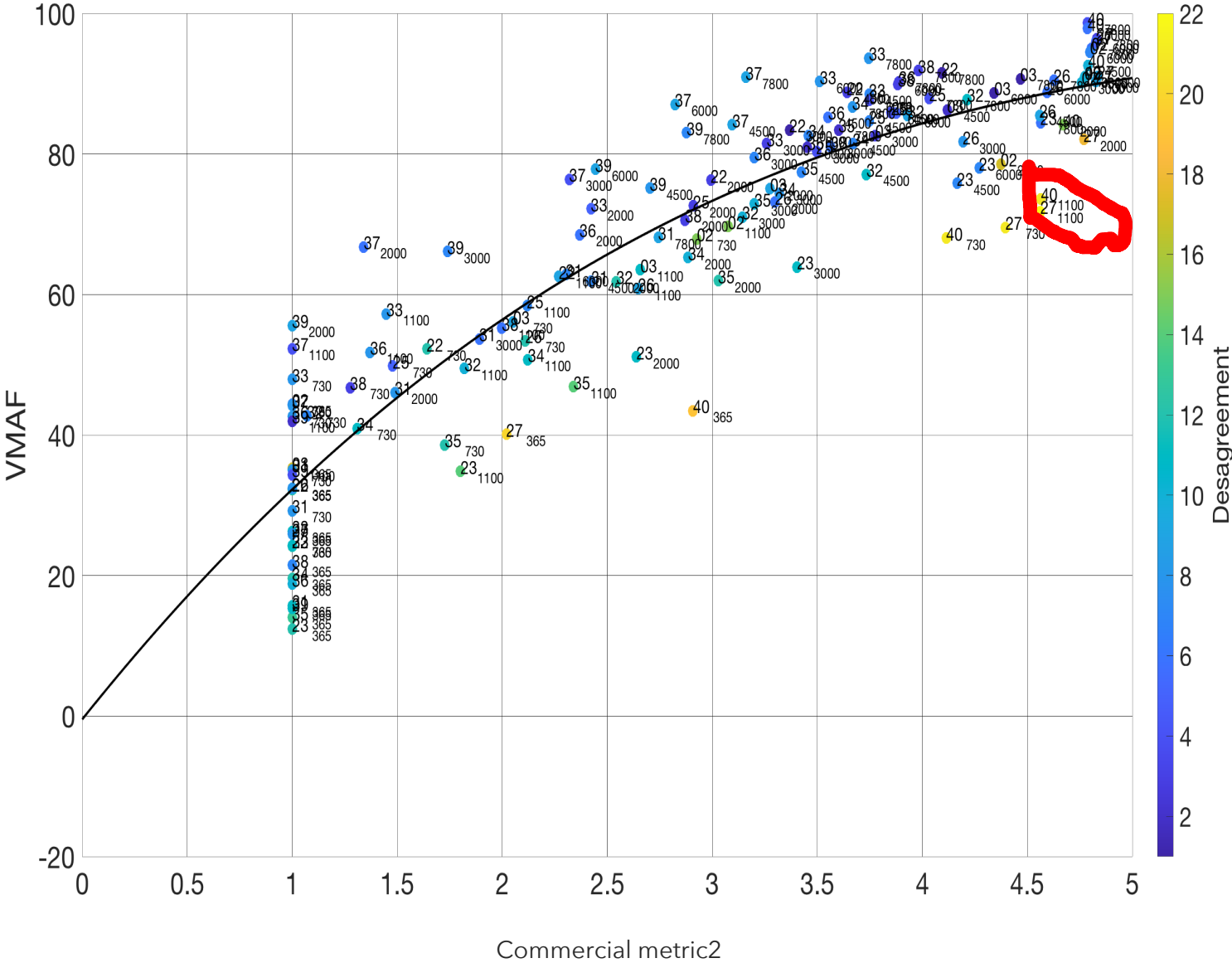
Disagreement analysis



Disagreement Analysis– Fitting VQMs to VMAF



Disagreement Analysis



Single frame of Src 27_1100



Next steps

- To understand what characteristics are causing the VQ metrics to disagree. There are different options to explore such as clustering, subjective tests, bitstream characteristics etc
- How do the metrics rank in terms of accuracy i.e. rank metrics based on the disagreement analysis
- To understand the causes of disagreement, i.e., what makes the video quality of a transcoded sequence difficult to evaluate objectively

Thank you

Appendix

Clustering approach

OPTICS 14

Silhouette Coefficient (<-1, 1>, higher = better): -0.04

Calinski-Harabasz (higher = better): 14.88

Davies-Bouldin (closer to zero = better): 1.16

