



# Sampling Effect on Performance Prediction of Configurable Systems : A Case Study

Juliana Alves Pereira, Mathieu Acher, Hugo Martin, Jean-Marc Jezequel



# Configurable systems

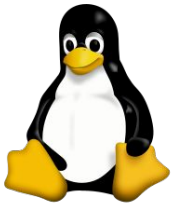

Pros

- Adaptive
- Lots of options

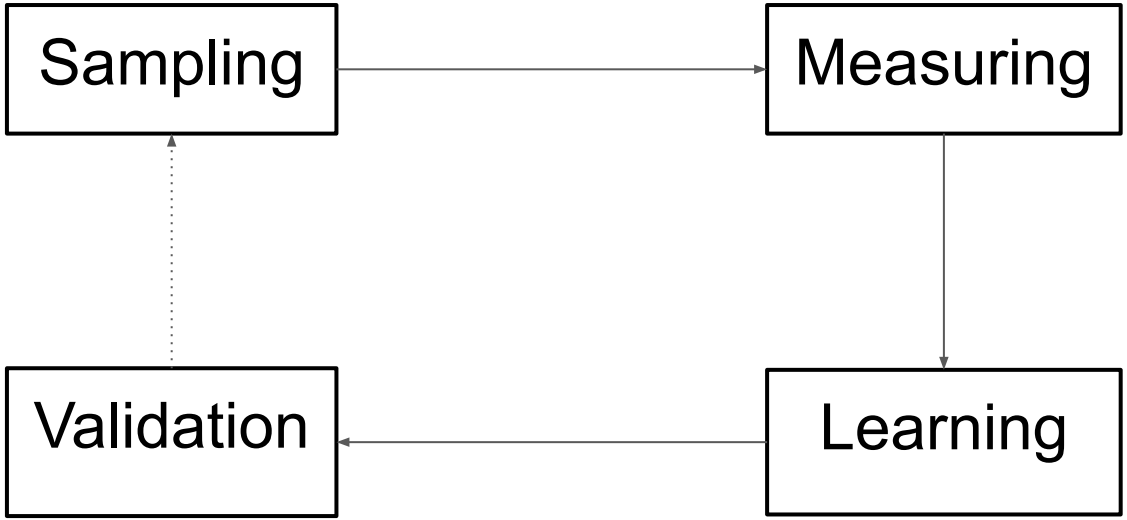

Cons

- Lots of options (and interactions)
- Increasingly complex


Machine learning to the rescue



# Machine Learning : Sampling, Measure, Learning, Validating



# Distance-Based Sampling of Software Configuration Spaces

- C. Kaltenecker, A. Grebhahn, N. Siegmund, J. Guo and S. Apel, "Distance-Based Sampling of Software Configuration Spaces," *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*, Montreal, QC, Canada, 2019, pp. 1084-1094.

- Proposing a new sampling solution : Distance-Based Sampling

- Empirical study on 10 subject systems and 6 sampling strategies

# Sampling strategies

- Coverage-based
- Solver-based
- Randomized solver-based

- Random

- Distance-based
- Diversified distance-based

# Subject systems

- 7z
- BerkeleyDB-C
- Dune MGS
- HIPAcc
- Java GC
- LLVM
- LRZIP
- Polly
- VPXENC
- x264

# Experiment setup

- Machine learning based on multiple linear regression and feature-forward selection
- Mean Relative Error (MRE)

# Results

| | Coverage-based | | | Solver-based | | | Randomized solver-based | | | Distance-based | | | Diversified distance-based | | | Random | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ |
| 7z | 51.2 % | 33.8 % | 22.6 % | 65.4 % | 58.2 % | 25.2 % | 55.1 % | 37.2 % | 16.7 % | 85.9 % | 27.3 % | 16.6 % | 74.3 % | **16.3 %** | 17.2 % | 58.2 % | 15.1 % | 9.9 % |
| BDB-C | 122.9 % | 29.0 % | 26.5 % | 49.5 % | 46.8 % | 42.0 % | **45.1 %** | 46.1 % | 18.1 % | 320.0 % | 75.1 % | 15.0 % | 237.0 % | **12.7 %** | **9.3 %** | 121.3 % | 39.1 % | 12.2 % |
| Dune | **15.5 %** | 12.5 % | 11.4 % | 23.6 % | 15.1 % | 11.8 % | 43.3 % | 16.8 % | 11.2 % | 24.4 % | 15.2 % | 11.4 % | 21.5 % | **11.8 %** | **11.0 %** | 17.6 % | 11.5 % | 11.3 % |
| Hipacc | **26.2 %** | 20.5 % | 20.5 % | 44.8 % | 17.2 % | 14.7 % | 31.9 % | 15.7 % | 14.2 % | 27.9 % | 19.0 % | 15.3 % | 31.5 % | **14.5 %** | **14.0 %** | 19.9 % | 13.9 % | 13.4 % |
| JavaGC | **36.7 %** | 32.1 % | 23.7 % | 54.2 % | 59.3 % | 35.8 % | 41.9 % | 37.8 % | 30.2 % | 72.9 % | 43.8 % | 28.2 % | 56.0 % | 29.9 % | **13.2 %** | 55.8 % | 13.9 % | 12.3 % |
| LLVM | 6.2 % | 6.2 % | 5.8 % | 9.5 % | 5.5 % | 5.2 % | 5.6 % | 5.2 % | 5.4 % | 5.8 % | **5.2 %** | 5.3 % | 5.9 % | 5.3 % | 5.2 % | 5.6 % | 5.2 % | 5.2 % |
| lrzip | **27.2 %** | 28.2 % | **13.4 %** | 47.3 % | 27.3 % | 23.9 % | 91.5 % | 36.0 % | 25.0 % | 162.5 % | 39.7 % | 21.9 % | 134.2 % | **25.1 %** | 18.2 % | 62.7 % | 18.3 % | 15.6 % |
| Polly | **19.7 %** | 12.7 % | **7.3 %** | 20.3 % | 16.1 % | 15.5 % | 20.0 % | 13.6 % | 14.0 % | 23.3 % | 14.2 % | 14.9 % | 25.8 % | **10.5 %** | 11.8 % | 25.1 % | 13.0 % | 10.3 % |
| VP9 | **100.3 %** | 96.3 % | 45.3 % | 413.0 % | 224.2 % | 80.8 % | 470.2 % | 389.1 % | 94.5 % | 721.9 % | 125.0 % | 84.5 % | 189.8 % | **66.5 %** | **32.0 %** | 80.6 % | 27.2 % | 23.3 % |
| x264 | 20.9 % | 11.9 % | 10.9 % | 26.2 % | 40.4 % | 42.2 % | 18.5 % | 22.2 % | 33.2 % | 14.7 % | 10.0 % | 9.4 % | **12.6 %** | **8.8 %** | **9.0 %** | 13.5 % | 9.2 % | 9.1 % |
| Mean | **42.7 %** | 28.3 % | 18.7 % | 75.4 % | 51.0 % | 29.7 % | 82.3 % | 62.0 % | 26.2 % | 145.9 % | 37.4 % | 22.2 % | 78.9 % | **20.1 %** | **14.1 %** | 46.0 % | 16.6 % | 12.3 % |

- Coverage-based is dominant at low sample size
- Diversified distance-based is dominant on higher sample size
- Diversified distance-based is close to random sampling accuracy, even better in some cases

# Is it true?

# Replicating the experiment

- Subject system : x264, video encoder

| | Coverage-based | | | Solver-based | | | Randomized solver-based | | | Distance-based | | | Diversified distance-based | | | Random | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ |
| x264 | 20.9 % | 11.9 % | 10.9 % | 26.2 % | 40.4 % | 42.2 % | 18.5 % | 22.2 % | 33.2 % | 14.7 % | 10.0 % | 9.4 % | **12.6 %** | **8.8 %** | **9.0 %** | 13.5 % | 9.2 % | 9.1 % |

- Changing the input video : 17 videos

- Changing the measured non-functional property

# Experimental setup

What does vary?

- Sampling strategy (6 strategies)
- Sample size (3 sample size)
- Encoded video (17 videos) 🔴
- System configuration (1152 configurations)
- Measured property (Encoding time, encoding size) 🔴

What doesn’t vary?

- Learning algorithm (Multiple Linear Regression)
- Learning algorithm hyperparameters
- Configurable Software (x264) 🔴
- Version 🔵
- Hardware 🔵

# Results

- High variation between videos, between non-functional properties

- Encoding time :
  - Similar results
  - Random sampling dominant over Diversified Distance-based sampling

- Encoding size :
  - Random sampling and randomized solver-based sampling overall dominant
  - Most strategies present good and similar accuracy for higher sample size

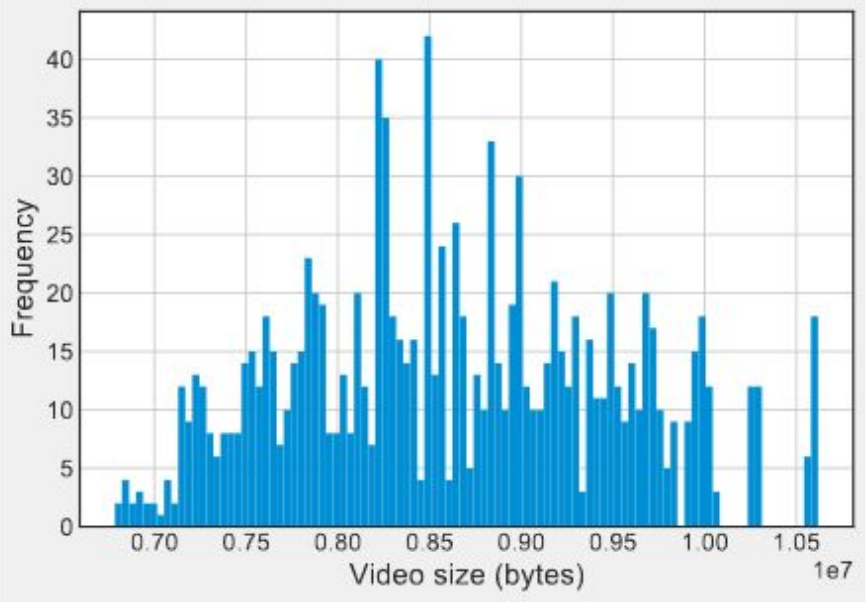
| Video | Coverage-based | | | Solver-based | | | Randomized solver-based | | | Distance-based | | | Diversified distance-based | | | Random | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ |
| $x264_0$ | 18.2 % | 13.9 % | 13.4 % | 24.0 % | 27.0 % | 27.5 % | 22.3 % | 19.9 % | 24.3 % | 16.5 % | 12.7 % | 10.6 % | 16.3 % | 8.8 % | **8.2 %** | 16.7 % | 9.2 % | 8.2 % |
| $x264_1$ | 15.4 % | 13.2 % | 12.1 % | 26.9 % | 23.7 % | 24.9 % | 21.4 % | 21.5 % | 23.2 % | 17.3 % | 14.2 % | 9.5 % | 17.4 % | 9.8 % | **8.7 %** | 16.1 % | **9.2 %** | 8.7 % |
| $x264_2$ | 29.3 % | 10.3 % | 9.7 % | 21.4 % | 19.4 % | 16.4 % | 19.1 % | 19.6 % | 19.4 % | 17.4 % | 11.4 % | 9.8 % | 17.6 % | 9.6 % | 9.3 % | **15.3 %** | 9.5 % | 9.3 % |
| $x264_3$ | 21.4 % | 13.7 % | 10.1 % | 25.2 % | 25.3 % | 26.4 % | 16.4 % | 22.3 % | 24.8 % | 13.6 % | 10.7 % | 10.2 % | **12.8 %** | **9.8 %** | 9.7 % | 14.5 % | 9.8 % | **9.2 %** |
| $x264_4$ | 21.8 % | 12.3 % | 14.4 % | 23.9 % | 21.2 % | 22.0 % | 18.3 % | 21.1 % | 22.5 % | 14.2 % | 11.7 % | 9.7 % | 13.9 % | 10.1 % | 8.9 % | 13.9 % | **9.4 %** | **8.8 %** |
| $x264_5$ | 26.1 % | 14.1 % | 13.2 % | 28.8 % | 23.2 % | 24.1 % | 21.8 % | 22.5 % | 23.3 % | 16.4 % | 13.4 % | 11.4 % | 16.8 % | 10.7 % | 9.5 % | **15.7 %** | **10.0 %** | **9.3 %** |
| $x264_6$ | 25.9 % | 18.1 % | 8.6 % | 23.6 % | 28.5 % | 29.1 % | 18.2 % | 21.6 % | 24.9 % | 13.7 % | 9.9 % | 9.0 % | 13.2 % | 8.8 % | 7.8 % | **12.6 %** | **8.0 %** | **7.3 %** |
| $x264_7$ | 23.3 % | 14.2 % | 12.0 % | 20.2 % | 25.3 % | 26.1 % | 15.3 % | 23.0 % | 23.8 % | 12.2 % | 9.2 % | 7.2 % | 10.8 % | 8.5 % | **7.2 %** | 11.4 % | 8.2 % | 7.3 % |
| $x264_8$ | 20.8 % | 13.1 % | 11.5 % | 20.3 % | 22.7 % | 23.6 % | 16.7 % | 23.4 % | 23.4 % | 12.6 % | 10.4 % | 9.6 % | 11.1 % | 9.3 % | 8.3 % | 12.0 % | 8.7 % | **7.6 %** |
| $x264_9$ | 23.4 % | 13.2 % | 5.6 % | 22.1 % | 28.6 % | 29.7 % | 16.8 % | 24.2 % | 25.3 % | 11.4 % | 6.5 % | 6.5 % | 9.2 % | **5.8 %** | 5.4 % | 10.9 % | 6.6 % | **5.4 %** |
| $x264_{10}$ | 21.9 % | 12.3 % | 9.3 % | 22.6 % | 23.2 % | 24.0 % | 17.9 % | 22.4 % | 24.3 % | 14.0 % | 10.2 % | 9.7 % | 13.5 % | 9.4 % | 8.9 % | 14.0 % | **9.0 %** | **8.8 %** |
| $x264_{11}$ | 21.1 % | 12.6 % | 10.3 % | 25.7 % | 23.5 % | 23.8 % | 20.0 % | 21.1 % | 24.7 % | 13.3 % | 10.8 % | 10.4 % | 13.0 % | 10.1 % | 9.7 % | 13.9 % | **9.4 %** | **9.1 %** |
| $x264_{12}$ | 25.4 % | 13.4 % | 10.4 % | 26.2 % | 21.2 % | 21.6 % | 19.8 % | 20.6 % | 20.9 % | 16.2 % | 13.7 % | 10.9 % | 16.3 % | 11.4 % | 9.1 % | **15.0 %** | **9.7 %** | **8.5 %** |
| $x264_{13}$ | 16.4 % | 10.5 % | 10.0 % | 20.6 % | 18.8 % | 19.1 % | 18.3 % | 19.4 % | 19.8 % | 16.0 % | 13.9 % | 10.0 % | 16.2 % | 10.5 % | 9.6 % | **15.5 %** | **9.7 %** | **9.0 %** |
| $x264_{14}$ | 20.7 % | 16.9 % | 15.8 % | 34.3 % | 39.5 % | 40.6 % | 28.5 % | 29.7 % | 32.4 % | 18.1 % | 11.1 % | 9.6 % | 18.4 % | 7.8 % | 7.3 % | 17.4 % | **7.5 %** | **7.2 %** |
| $x264_{15}$ | 26.2 % | 12.7 % | 11.1 % | 23.2 % | 26.5 % | 27.2 % | 20.3 % | 22.7 % | 25.1 % | 15.1 % | 11.9 % | 10.7 % | 14.8 % | 10.6 % | 9.5 % | **13.9 %** | **9.1 %** | **8.9 %** |
| $x264_{16}$ | 22.9 % | 12.3 % | 8.4 % | 22.1 % | 24.5 % | 25.2 % | 18.0 % | 22.2 % | 23.6 % | 13.4 % | 9.4 % | 8.9 % | 12.6 % | 8.5 % | 7.8 % | 12.5 % | **8.1 %** | **7.4 %** |
| Mean | 22.4 % | 13.3 % | 10.9 % | 24.2 % | 24.8 % | 25.4 % | 19.4 % | 22.2 % | 23.9 % | 14.8 % | 11.3 % | 9.6 % | 14.3 % | 9.4 % | 8.5 % | **14.2 %** | **8.9 %** | **8.2 %** |

Results table for encoding time

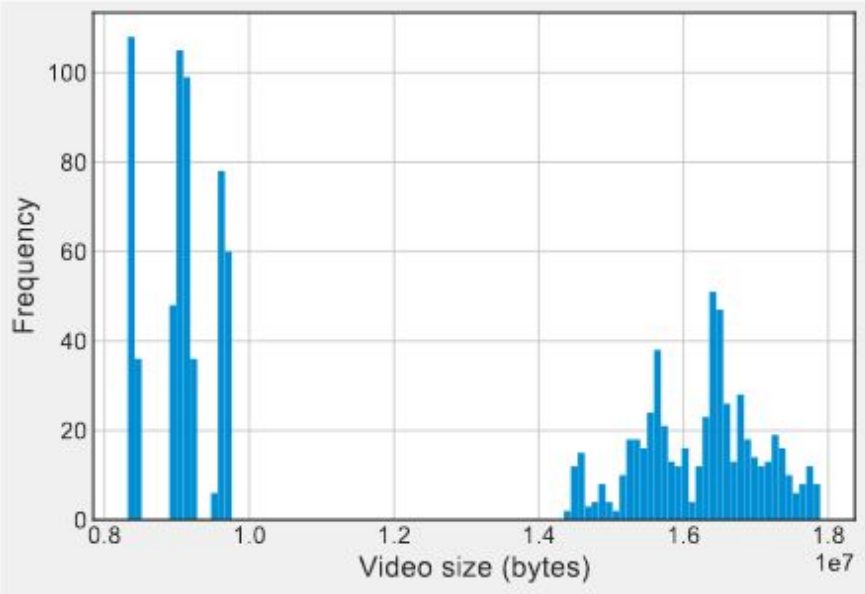
| Video | Coverage-based | | | Solver-based | | | Randomized solver-based | | | Distance-based | | | Diversified distance-based | | | Random | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ | $t=1$ | $t=2$ | $t=3$ |
| $x264_0$ | **12.3 %** | 11.6 % | 11.1 % | 12.3 % | 11.4 % | 11.3 % | 25.1 % | 12.7 % | 13.3 % | 25.3 % | 12.5 % | 10.6 % | 23.3 % | 10.6 % | 9.2 % | 13.1 % | **9.8 %** | **9.1 %** |
| $x264_1$ | 4.0 % | 3.9 % | 3.8 % | 3.1 % | 3.8 % | 3.8 % | **1.7 %** | **3.8 %** | **3.8 %** | 4.0 % | 4.0 % | 3.8 % | 3.9 % | 3.8 % | 3.8 % | 3.9 % | 3.8 % | 3.8 % |
| $x264_2$ | 14.9 % | 14.3 % | 4.8 % | **5.1 %** | 4.7 % | 4.7 % | 15.9 % | 4.7 % | **4.6 %** | 14.3 % | 14.0 % | 10.2 % | 13.8 % | 12.0 % | 4.7 % | 7.6 % | **4.7 %** | 4.6 % |
| $x264_3$ | 8.6 % | 8.3 % | 7.8 % | 8.1 % | **7.3 %** | 7.4 % | 11.2 % | 7.6 % | 7.4 % | 9.9 % | 9.3 % | 8.0 % | 9.6 % | 8.3 % | 7.5 % | **7.7 %** | 7.4 % | **7.3 %** |
| $x264_4$ | 18.4 % | 16.7 % | **6.6 %** | **4.5 %** | 6.8 % | 6.8 % | 14.1 % | **6.7 %** | 6.7 % | 17.5 % | 16.7 % | 7.0 % | 16.9 % | 6.9 % | 6.9 % | 7.8 % | 6.9 % | 6.9 % |
| $x264_5$ | 11.3 % | 11.0 % | 10.8 % | **4.9 %** | 6.6 % | 5.7 % | 12.3 % | 9.4 % | **4.8 %** | 11.8 % | 11.5 % | 10.9 % | 11.6 % | 10.6 % | 10.0 % | 9.4 % | **6.4 %** | 5.2 % |
| $x264_6$ | 24.6 % | **5.3 %** | **5.2 %** | **5.4 %** | 5.4 % | 5.3 % | 25.6 % | **5.3 %** | 5.3 % | 17.6 % | 16.8 % | 5.5 % | 16.1 % | 5.4 % | 5.4 % | 6.3 % | **5.3 %** | 5.3 % |
| $x264_7$ | 9.4 % | 9.0 % | 8.7 % | **8.1 %** | 8.4 % | 8.3 % | 8.4 % | **8.2 %** | **8.2 %** | 9.4 % | 9.4 % | 8.9 % | 9.3 % | 8.6 % | 8.5 % | 9.1 % | 8.4 % | 8.3 % |
| $x264_8$ | 10.4 % | 9.7 % | 8.9 % | 8.7 % | 8.0 % | 8.1 % | 11.2 % | **7.6 %** | **8.0 %** | 12.4 % | 12.0 % | 9.5 % | 12.0 % | 9.9 % | 8.5 % | **8.5 %** | 8.3 % | 8.2 % |
| $x264_9$ | 11.6 % | 10.5 % | 9.5 % | 7.6 % | 8.6 % | 8.5 % | **6.9 %** | **8.4 %** | **8.4 %** | 11.3 % | 11.6 % | 9.6 % | 10.8 % | 9.7 % | 8.7 % | 8.8 % | 8.5 % | 8.4 % |
| $x264_{10}$ | 5.2 % | 5.2 % | 4.9 % | 5.2 % | 5.0 % | 4.8 % | 5.0 % | 4.6 % | 4.6 % | 6.0 % | 5.8 % | 5.0 % | 5.7 % | 5.1 % | 4.7 % | **4.9 %** | 4.6 % | 4.6 % |
| $x264_{11}$ | 12.4 % | 11.8 % | 11.1 % | 11.1 % | 10.8 % | 11.0 % | **8.8 %** | 9.9 % | 11.4 % | 12.8 % | 11.8 % | 9.0 % | 12.0 % | 10.2 % | **8.6 %** | 10.9 % | **9.4 %** | 8.8 % |
| $x264_{12}$ | 25.7 % | 3.6 % | 3.6 % | 5.3 % | 3.5 % | 3.6 % | 28.9 % | 3.6 % | 3.5 % | 16.5 % | 14.6 % | 3.5 % | 15.4 % | 3.5 % | 3.4 % | **4.8 %** | **3.5 %** | **3.4 %** |
| $x264_{13}$ | 4.7 % | 4.7 % | 4.6 % | **4.5 %** | 4.7 % | 4.7 % | 5.4 % | 4.8 % | 4.7 % | 5.1 % | 5.0 % | 4.8 % | 5.0 % | 4.7 % | 4.7 % | 5.0 % | **4.7 %** | **4.6 %** |
| $x264_{14}$ | 10.2 % | 9.6 % | 9.4 % | 5.1 % | **7.4 %** | **8.8 %** | **3.6 %** | 9.6 % | 9.5 % | 10.6 % | 10.6 % | 10.0 % | 9.8 % | 9.6 % | 9.6 % | 9.3 % | 9.0 % | 9.5 % |
| $x264_{15}$ | **4.1 %** | **4.0 %** | **4.0 %** | 7.5 % | 4.5 % | 4.3 % | 40.9 % | 4.3 % | 4.2 % | 21.7 % | 8.3 % | 4.1 % | 19.1 % | 4.1 % | 4.1 % | 5.4 % | 4.2 % | 4.1 % |
| $x264_{16}$ | 8.3 % | 8.1 % | 7.9 % | **7.7 %** | 7.8 % | 7.6 % | 9.2 % | 7.7 % | 7.6 % | 8.8 % | 8.7 % | 8.2 % | 8.7 % | 7.9 % | 7.7 % | 8.3 % | **7.7 %** | **7.6 %** |
| Mean | 11.5 % | 8.7 % | 7.2 % | **6.7 %** | 6.8 % | 6.7 % | 13.8 % | 7.0 % | 6.8 % | 12.6 % | 10.7 % | 7.6 % | 12.0 % | 7.7 % | 6.8 % | 7.7 % | 6.6 % | **6.5 %** |

Results table for encoding size

# Results




(a) flower_sif.y4m $x264_2$




(b) 720p50_parkrun_ter.y4m $x264_{15}$

# Replicability


- Fully replicable experiment
- Dataset for video encoding time and size available
- Docker image with all data and scripts for performance prediction and results aggregation : https://github.com/jualvespereira/ICPE2020

# What’s next?

- How do version and hardware affect the sampling effectiveness?
- How does machine learning technique affect the sampling effectiveness?
- How to leverage the fact that some sampling strategies overperform by focusing on important options?

# Conclusion

- Random sampling is a strong baseline, hard to challenge
- Diversified distance-based sampling is a strong alternative
- Researchers should be aware that effectiveness of sampling strategies can be biased by inputs and performance property used