# Statistical models for the minimization of energy consumption in cyber-physical systems

Marwan El Khazen

StatInf, Inria

October 19, 2021

# Outline

1. Introduction
   - Cyber Physical Systems
   - Energy reduction techniques
   - Scheduling and Timing Analysis
2. Statistical Models
   - Extreme Value Theory
   - Clustering and Representativity
   - Combining pWCETs
3. Perspectives for PhD

# Introduction - What are Cyber Physical systems?

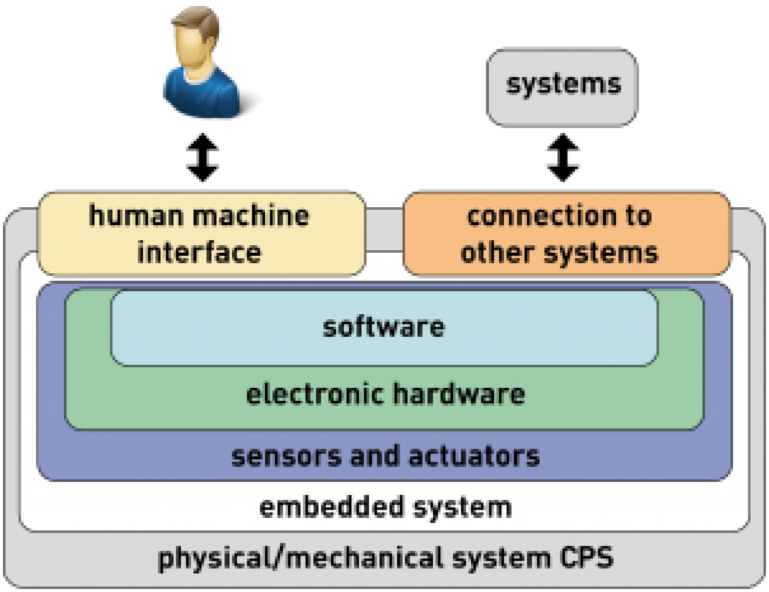


Figure 1: *https://addi-data.com/cps-cyber-physical-systems/*

CPS is about the **intersection**, not the union, of the **physical** and the **cyber**. It is not sufficient to separately understand the physical components and the computational components. We must instead understand their **interaction**.

*Introduction to Embedded Systems: A Cyber-Physical Systems Approach, Edward A. Lee & Sanjit Arunkumar Seshia*

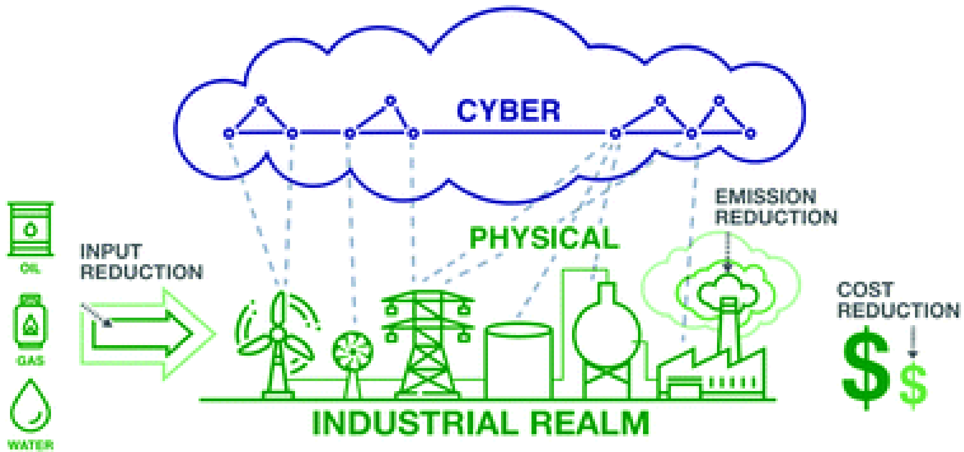


Figure 2: *Oliver Inderwildi et al.*

# Introduction - Energy reduction techniques

Dynamic Voltage and Frequency Scaling (DVFS):

- **Static** and/or **Dynamic**
- Speed set
- Computation time
- Time/energy overhead

Dynamic Power Management (DPM):

- **Online** or **Offline**
- State transition overhead
- Task early terminations

# Introduction - Energy reduction techniques

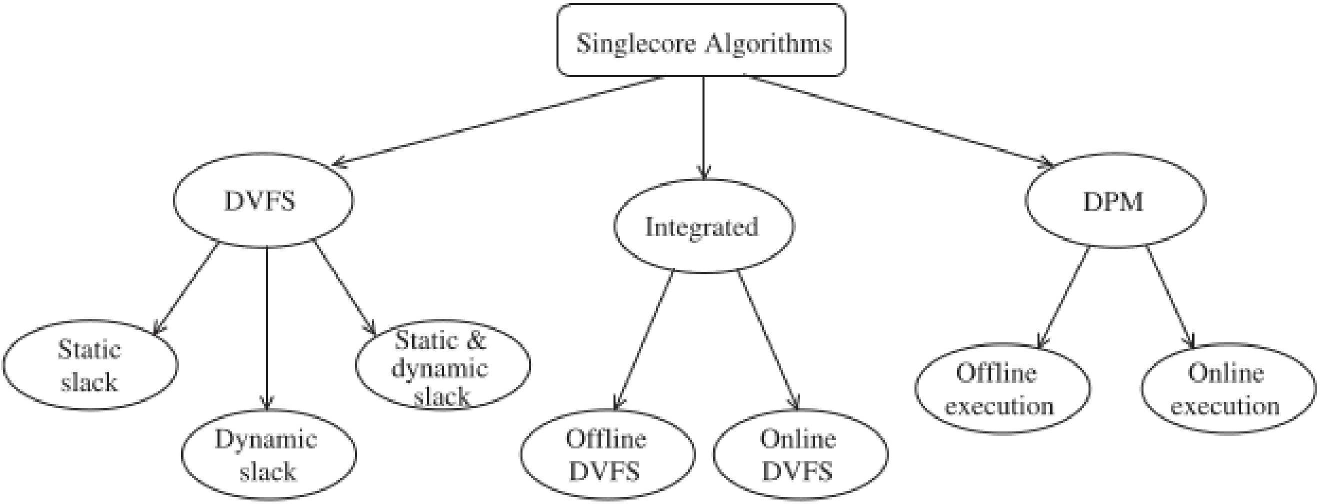


Figure 3: *Energy- aware scheduling for real-time systems: A survey, M. Bambagini et al.*

# Introduction - Energy reduction techniques

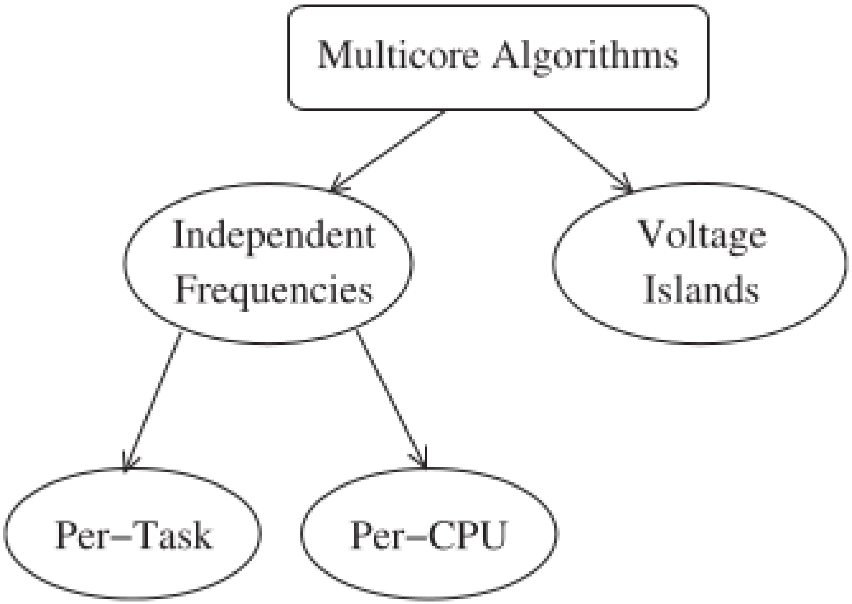


Figure 4: *M. Bambagini et al.*

# What about Time?

A task $\tau$ is defined by $(O, C, T, D)$:

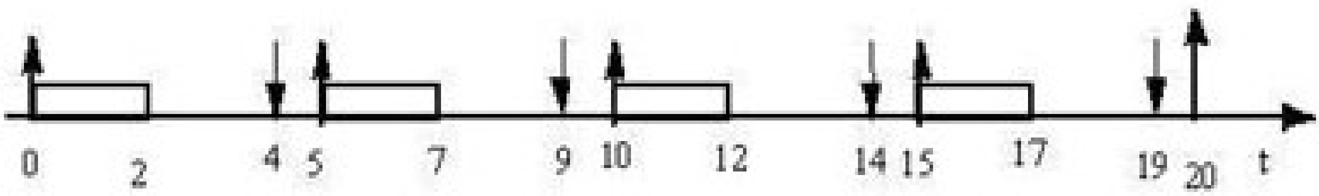


Figure 5:
*http://y.legouzouguec.free.fr/cours/ns/chap3/ordonnancement$_t$emps$_r$eel.html*

- $O$ is the origin, $O = 0$
- $C$ for computation time, $C = 2$
- $T$ is the period, $T = 5$
- $D$ is the deadline, $D = 4$

# Introduction - Scheduling and Timing Analysis

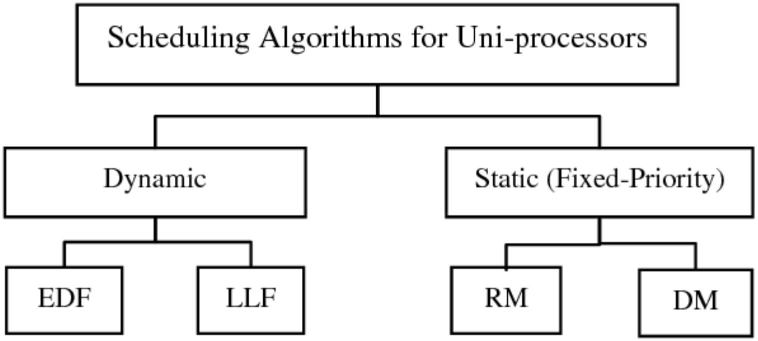


Figure 6: *A Survey on Scheduling Approaches for Hard Real-Time Systems, Mehrin Rouhifar et al.*

# Introduction - Scheduling and Timing Analysis

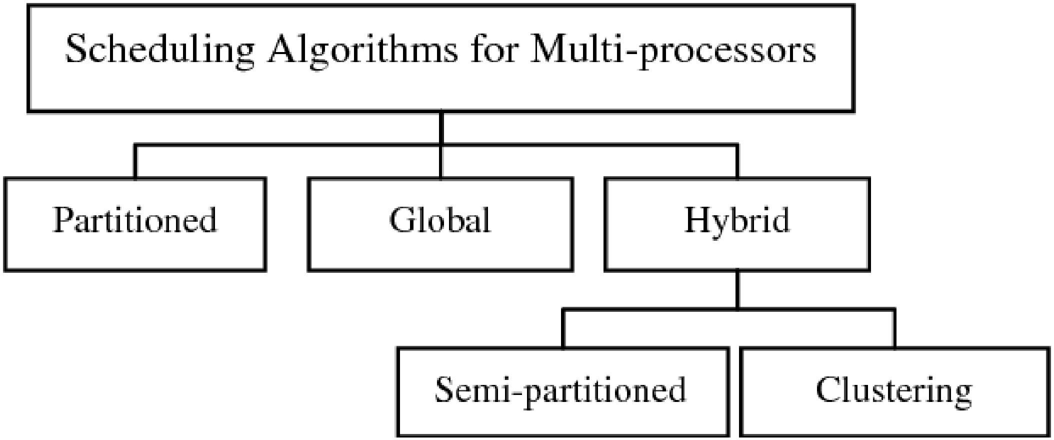


Figure 7: *Mehrin Rouhifar et al.*

# Introduction - Scheduling and Timing Analysis

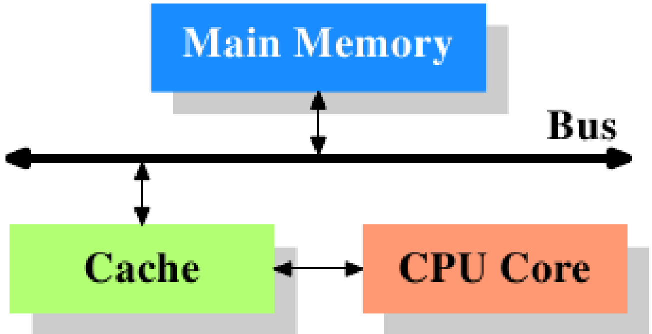


Figure 8: *https://lwn.net/Articles/252125/*

- Static or Measurement Based Timing Analysis
- Measurement Based Probabilistic Timing Analysis **MBPTA**
- Worst-Case Execution Time **WCET**
- Worst-Case Response Time **WCRT**
- Worst-Case Execution Cycles **WCEC?**
- Worst-Case Energy Consumption **WCEC?**

# Why statistical models?

# Statistical Models

- Complexifying architectures
- Extreme Value Theory and statistical WCET/WCRT/WCEC
- MBPTA, clustering and representativity
- Concentration inequalities
- Distributions for Execution Times, Response Times, Energy Consumption
- Statistical methods for scheduling
- And more...

# Extreme Value Theory: Two methods

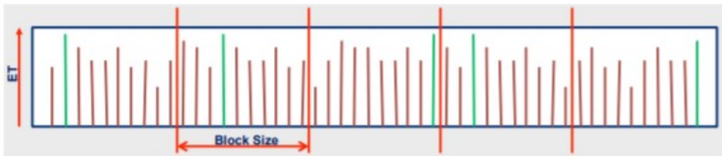


Figure 9: The *block maxima* method

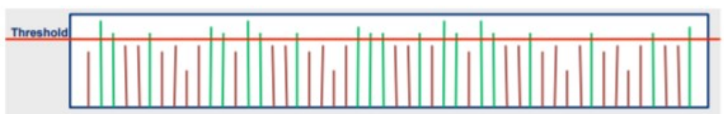


Figure 10: The *peaks-over-threshold* method

# Extreme Value Theory: Theorems

Theorem 1: If there exist sequences of constants $a_n > 0$ and $b_n$ such that $\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z)$ as $n \to \infty$, where G is a non-degenerate distribution function, then G belongs to the Generalized Extreme Value (GEV) family of distributions:

$$G\left(z\right) = \exp\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\},$$

defined on $\{z : 1 + \xi\left(\frac{z-\mu}{\sigma}\right) > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$

# Extreme Value Theory: Theorems

Theorem 2: The approximate distribution function of $X - u | X > u$ is, for $u$ large enough, the Generalized Pareto Distribution (GPD):

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma + \xi (u - \mu)}\right)^{-\frac{1}{\xi}},$$

defined on $\{y : y > 0 \text{ and } 1 + \frac{\xi y}{\sigma + \xi (u - \mu)} > 0\}$

# Extreme Value Theory: Distributions

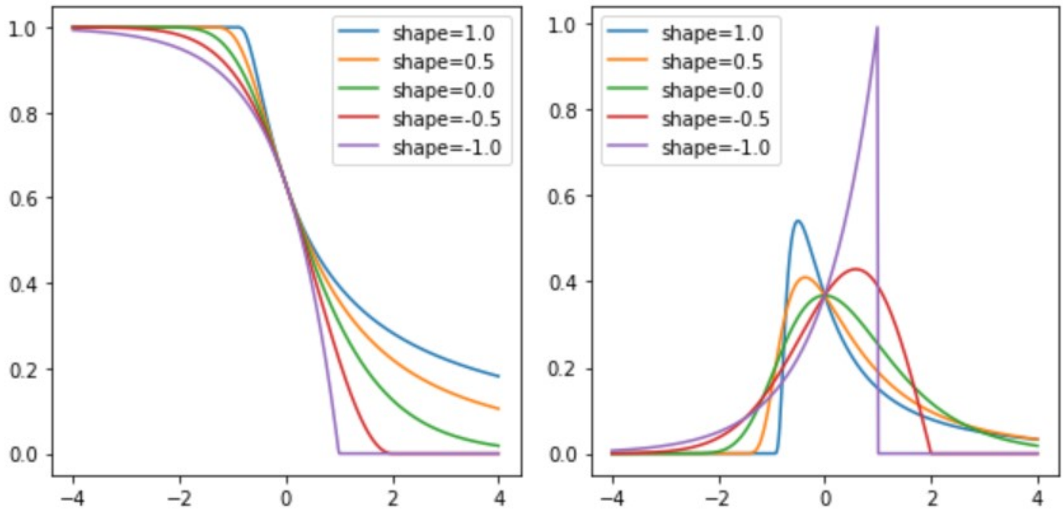


Figure 11: The *GEV* distribution

# Extreme Value Theory: Distributions

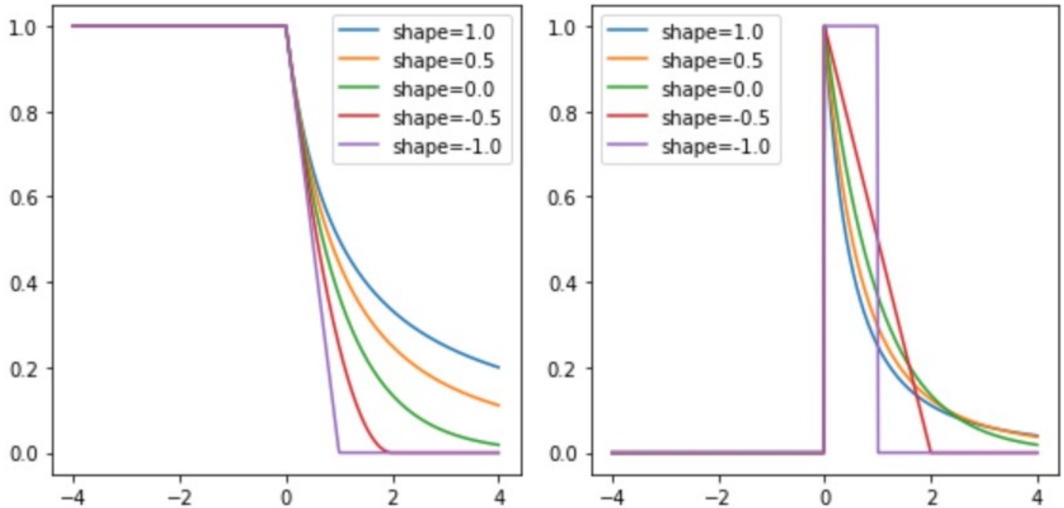


Figure 12: The *Generalized Pareto* distribution

# Statistical WCET of a program

The statistical or probabilistic worst-case execution time (pWCET) $\mathcal{C}$ of a program is an upper bound on all possible probabilistic execution times $\mathcal{C}_i$ for all possible execution scenarios $S_i, \forall i \geq 1$. The relation $\succeq$ describes the relation between the probabilistic execution times (pETs) of a program and its probabilistic worst case execution time (pWCET), $\mathcal{C}_i \succeq \mathcal{C}_i^j$, $\forall j$, defined as follows.

One writes $\mathcal{C} \succeq \mathcal{C}_i$ or $\mathcal{C}$ is said to be worse than $\mathcal{C}_i$ if its complementary cumulative distribution function (survival function 1-CDF) has a higher or equal probability associated to each possible value, i.e., $P(C \geq c) \geq P(C_i \geq c), \forall c$.

# Existing estimators in the literature

- Maximum Likelihood Estimator (MLE)
- Method of Moments (MoM)
- L-Moments estimator (LM)
- Hill, Pickands and Zipf estimators
- Empirical Bayesian Computation (EBC)

# Novel estimators proposed

- **A hybrid method ML/LM**
- Approximate Bayesian Computation (ABC)
- Random Forest Regression
- Gradient Tree Boosting Regression
- Systematic Testing (ST)
- Neural networks

# Outputs of the tool

- Identically distributed test
- Independence test
- Dependencies graph
- Histogram and density
- pWCET based on BM and POT methods

*Work in Progress Lessons learnt from creating Extreme Value Libraries in Python - Marwan W. El Khazen, Adriana Gogonel and Liliana Cucu-Grosjean*

- **New:** Stationarity tests and estimation

# Python-tool

```
Length: 2782
Identically distributed hypothesis cannot be rejected, with average p-value 0.6
Independence hypothesis cannot be rejected, with p-value 0.23
GEV parameters: shape=-0.416, location=29183.88, scale=14.72
GPD parameters: shape=-0.517, location=29159.94, scale=29.68
```

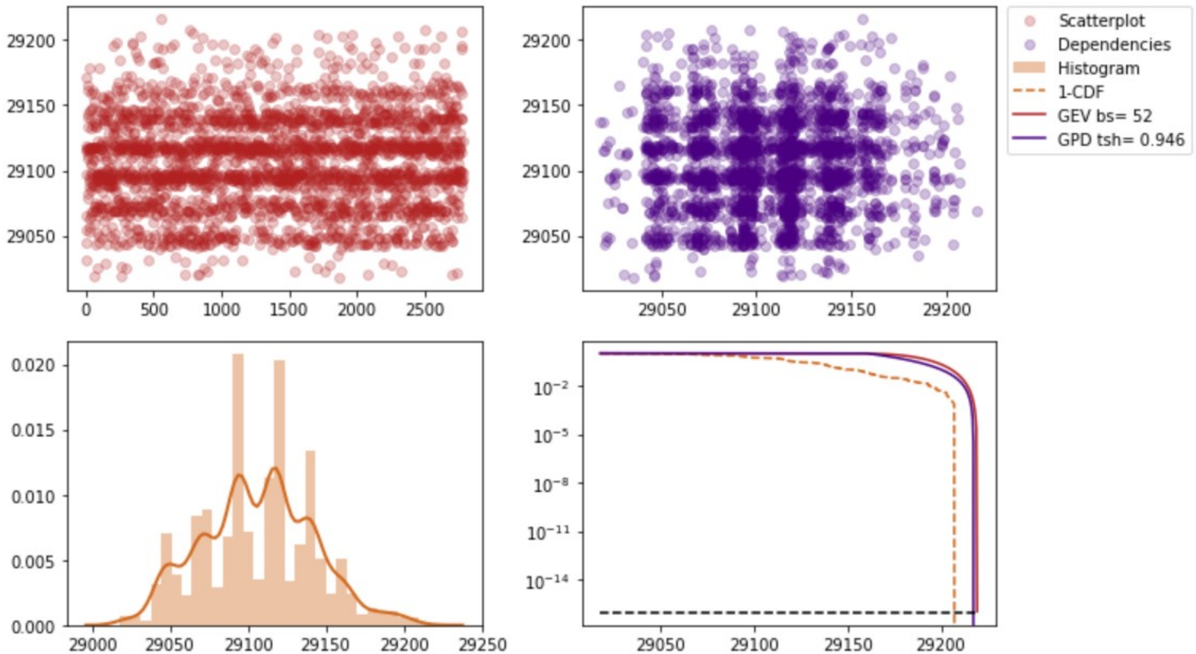


Figure 13: A complete output

# Clustering and Representativity

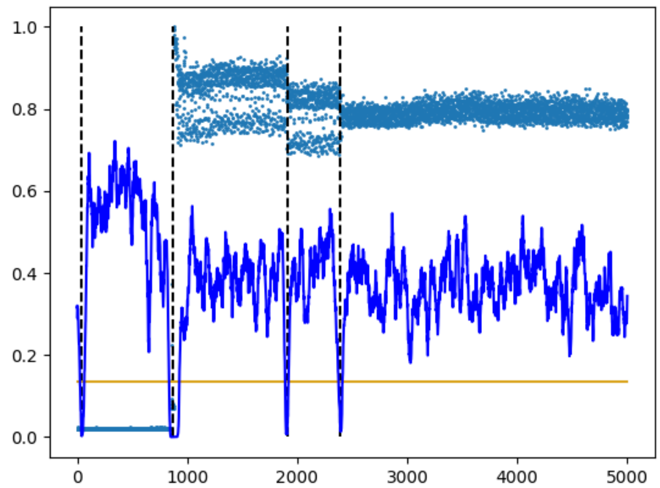


Figure 14: Walking Kolmogorov Smirnov Test (WKS)

*Work-in-Progress Abstract: WKS, a local unsupervised statistical algorithm for the detection of transitions in timing analysis - Marwan W. El Khazen et al.*

# Clustering and Representativity

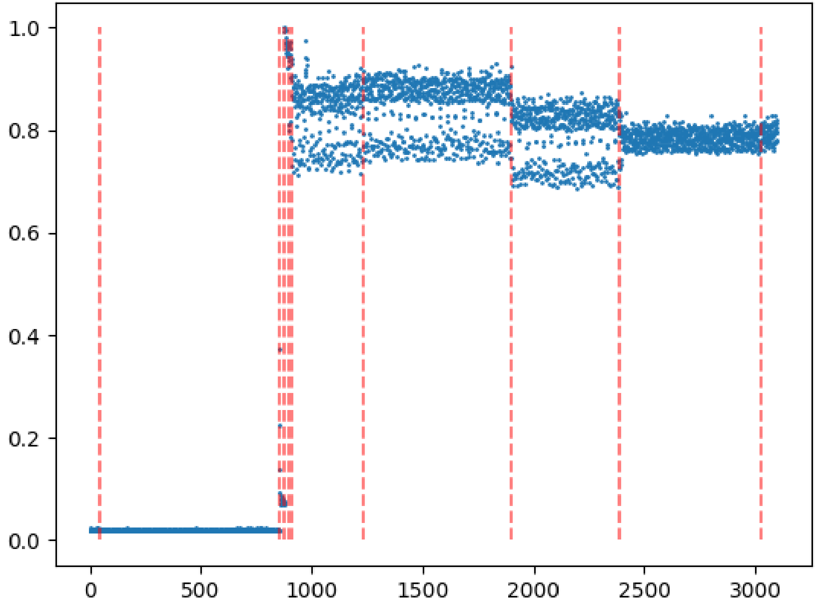


Figure 15: Distribution change point detection for time series

# Combining pWCETs

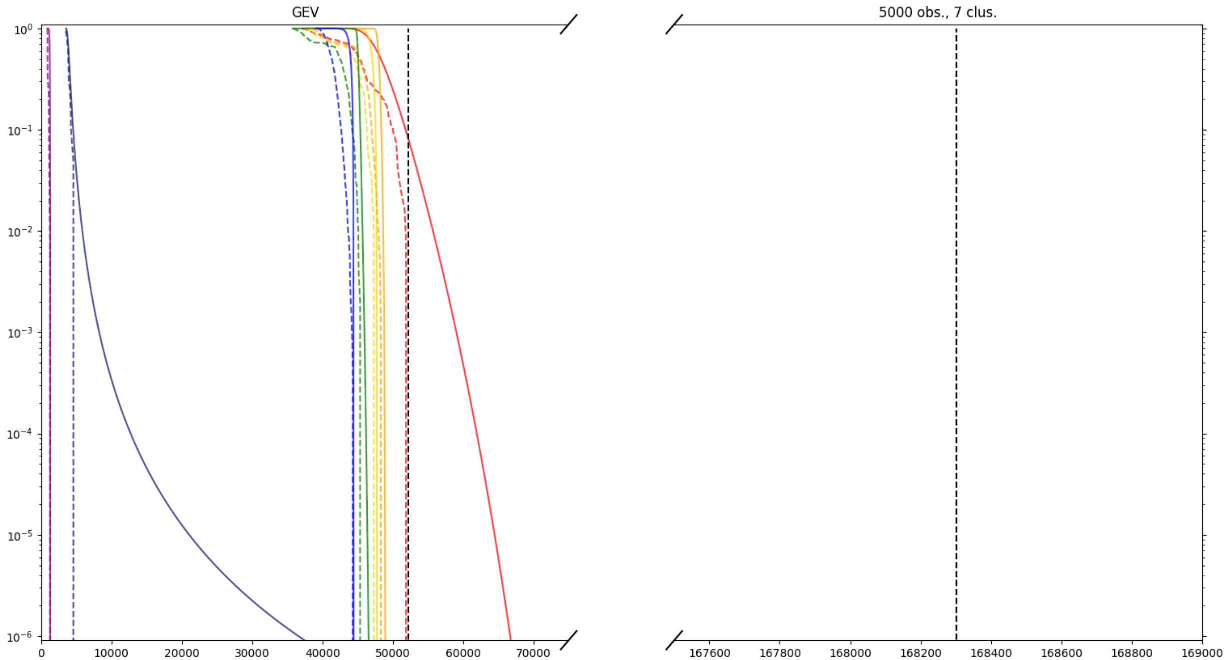


Figure 16: Multiple GEV pWCETs after clustering

# Combining pWCETs

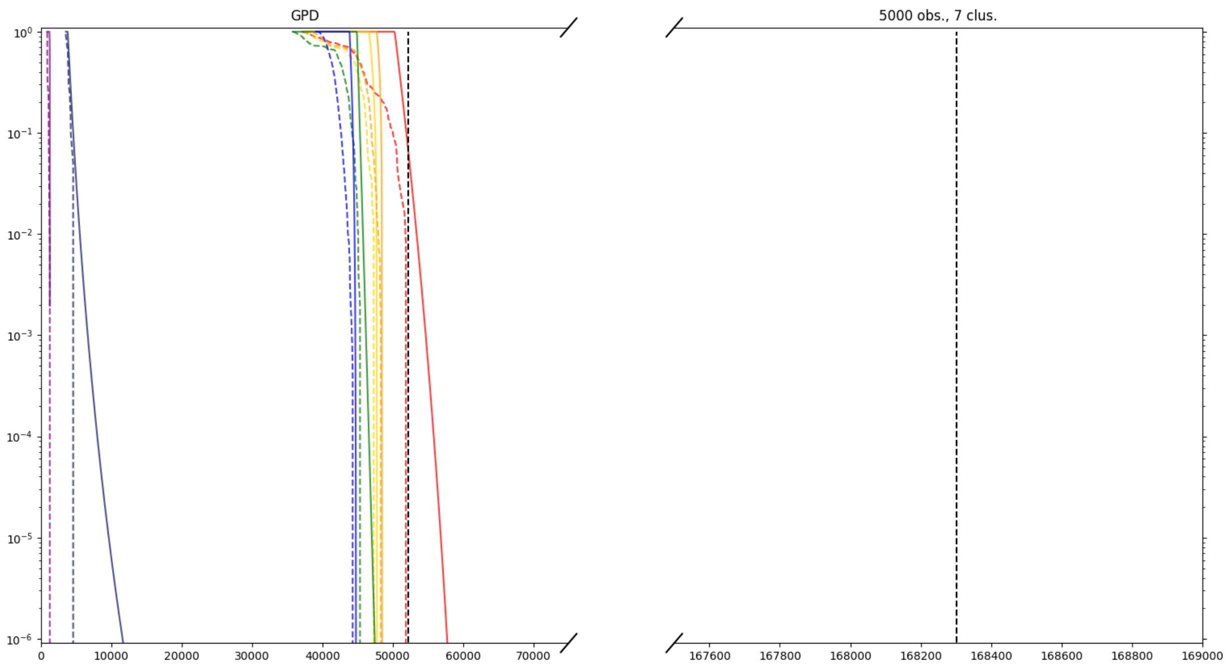


Figure 17: Multiple GPD pWCETs after clustering

# Combining pWCETs

- Stopping criterion
- Multiple testing problem
- P-hacking
- Probabilistic guarantees
- Asymptotic approximation

# Perspectives for PhD

- EVT for dependent WCRT/WCEC, working example
- Non-asymptotic estimation with concentration inequalities
- Distribution change point detection for better scheduling
- Machine Learning for time and energy predictions
- Prediction-based Integrated DVFS-DPM + Memory/Bus
- Comparing performance with static analyses

Thank you for your time