TORSTEN HOEFLER

Performance Reproducibility in HPC - Challenges and State-of-the-Art

PASC16
Platform for Advanced Scientific Computing Conference
Lausanne Switzerland
08-10 June 2016

- CLIMATE & WEATHER
- SOLID EARTH
- LIFE SCIENCE
- CHEMISTRY & MATERIALS
- PHYSICS
- COMPUTER SCIENCE & MATHEMATICS
- ENGINEERING
- EMERGING DOMAINS

sighpc
acm

Performance reproducibility is a pipe dream!

- **Cannot really be attained in the real world**
  - Systems change (especially software versions)
  - Supercomputers are not generally available (think Gordon Bell runs)
  - In general nearly impossible, exceptions may exist

- **So what now?**
  - Performance interpretability as a weaker goal

    "We call an experiment interpretable if it provides enough information to allow scientists to understand the experiment, draw own conclusions, assess their certainty, and possibly generalize results." [1]

- **Are not all scientific papers interpretable in this definition?**
  - Unfortunately not
  - Most are not interpretable and can easily be questioned ☹

    See survey in [1]

[1]: TH, R. Belli: *"Scientific Benchmarking of Parallel Computer Systems", SC15*

Scientific Benchmarking Guidelines

- **The state of the practice is nearly comical**
  - Inspired many funny talks/reports
  - Bailey's "12 ways to fool the masses"
  - Nelson Amaral's "How did this get published?"
  - Wellein/Hager "Fooling the masses with performance results"

- **We define 12 rules in a State of the Practice paper**

- **Key points:**
  - Careful factorial design
  - Use correct data summarizations
  - Report data variance and distribution
  - Do not assume normality, use nonparametric statistics
  - Measure parallel time correctly

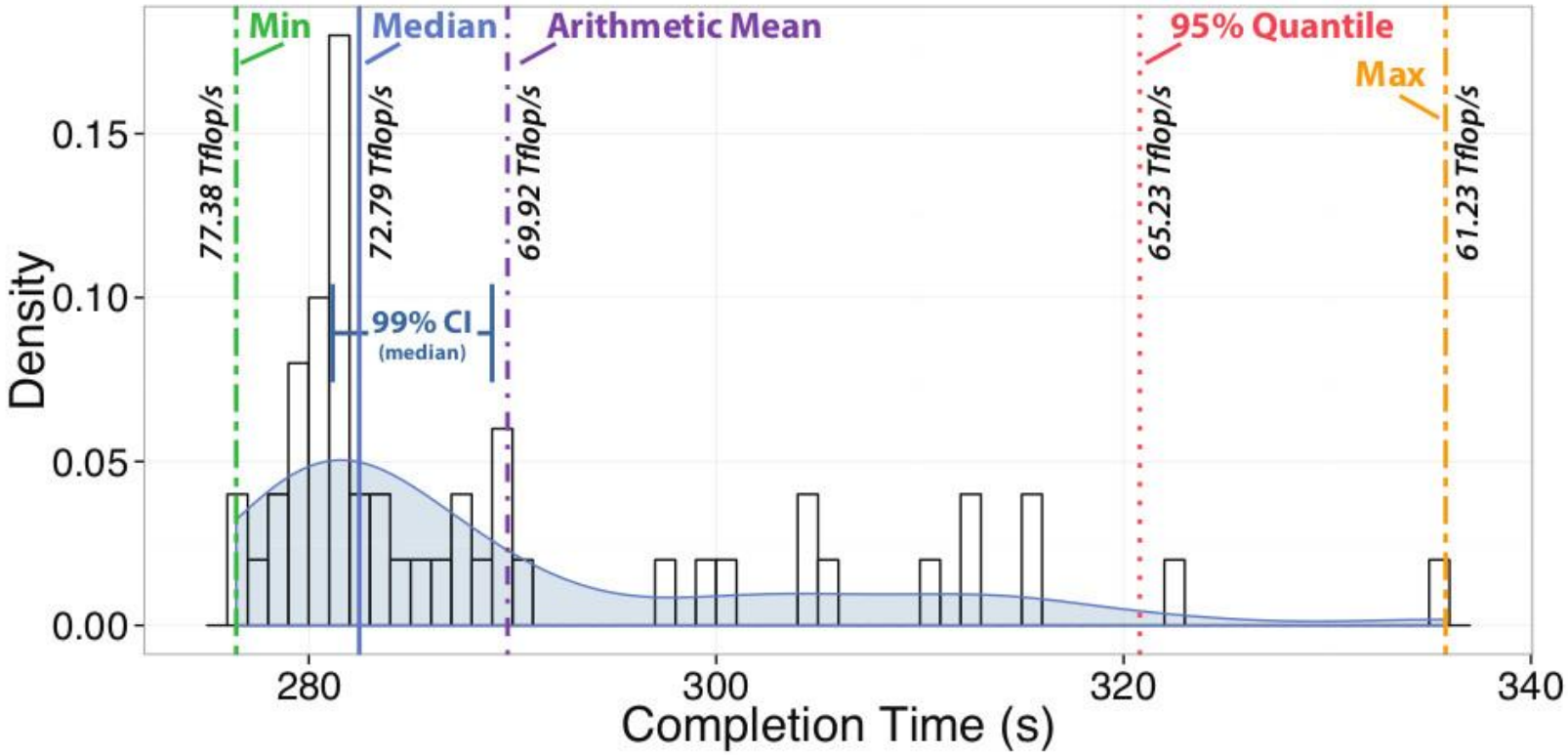
Want to know details?

"Scientific Benchmarking of Parallel Computer Systems"

Thursday 1:30-2pm, Room 18AB

Dealing with non-normal data – nonparametric statistics

- **Rank-based measures (no assumption about distribution)**
  - Almost always better than assuming normality
- **Example: median (50th percentile) vs. mean for HPL**
  - Rather stable statistic for expectation
  - Other percentiles (usually 25th and 75th) are also useful



Call for action!

- **Improve quality of reporting performance results**
  - Community effort needed
  - Teach students
  - Enforce at conferences
- **Continue the discussion**
  - Look at our experimental methodology very carefully
  - Establish minimal quality guidelines
- **Discuss in this BoF- now!**