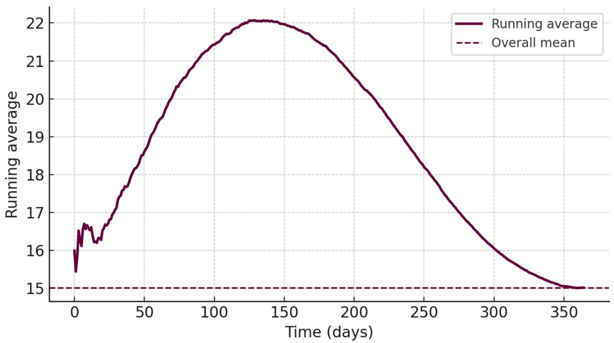
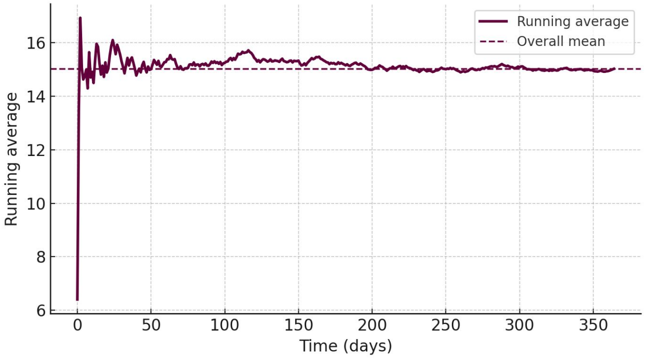
Online Shuffling for Online Machine Learning and Progressive Visualization

Pavlo Poliuha, Francesca Bugiotti, Benoit Groz, Jean-Daniel Fekete
Université Paris-Saclay, CNRS, Inria, LISN

Problem → Solution → Applications & Impact



Original (sorted) order → bias



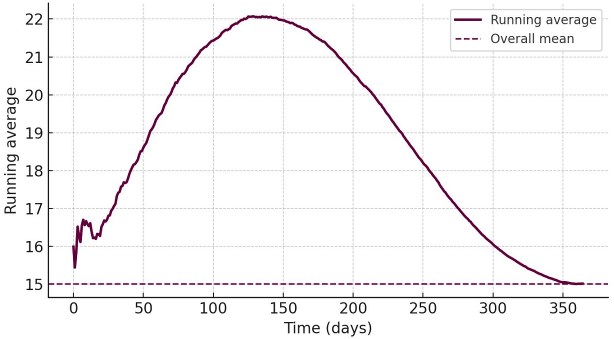
Shuffled order → fast convergence

- Ordered online data:
  - biased early results
  - slow convergence.

- Middleware that **shuffles remote data ‘on the fly’**
- Streams progressively unbiased samples to the downstream system.

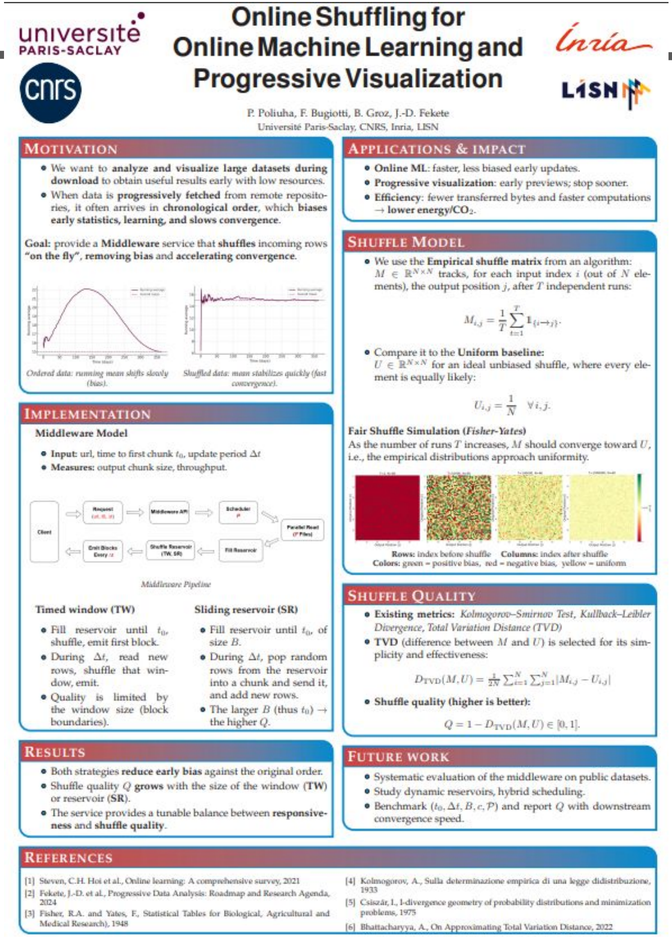
- **Online ML**: faster, less biased updates.
- **Visualization**: early previews, quicker insights.
- **Efficiency**: fewer transferred bytes, faster computations → **reduce energy and CO2.**

Problem



Original (sorted) order → bias

- Ordered online data:
  - biased early results
  - slow convergence.



Applications & Impact

- **Online ML**: faster, less biased updates.
- **Visualization**: early previews, quicker insights.
- **Efficiency**: fewer transferred bytes, faster computations → **reduce energy and CO2.**