Intelligent Job Scheduling for HPC Systems:

Methodology & Statistical Framework (Submission 2)

Justin M. Cheney

2026-05-14

University of the Western Cape

Honours 2026 - Submission 2

Recap: Research Focus & Proposed Approach

Research Focus

Goal: Identify the best-performing DRL family for HPC scheduling across heterogeneous workloads, with rigorous statistical evidence.

- Six DRL algorithms (PPO, DQN, A2C + maskable variants)
- Two real Slurm traces (CPU + GPU workloads)
- Fixed allocator (best_fit) and time-aware splits

Proposed Approach

Pipeline: smoke -> train -> evaluate -> aggregate -> stats -> plots

- Deterministic evaluation per seed
- Aggregated summaries across seeds
- Statistical comparisons per metric

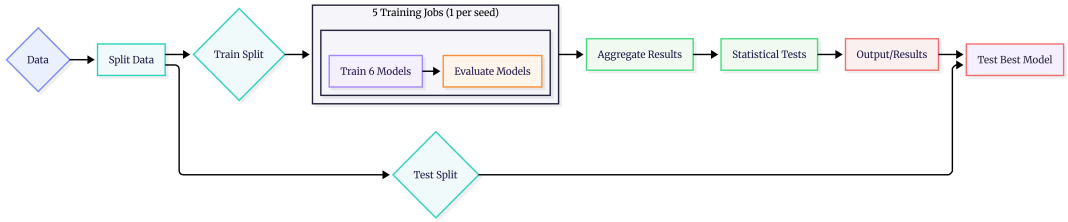
Snakemake + Nix + Apptainer for reproducibility [4, 6]

Methodology: From Concept to Executable Pipeline

What's New vs Intro/Lit Review

- Formalized train/eval/aggregate contracts (files + metadata)
- Time-aware split policy with holdout isolation
- Reproducibility stack: Nix + Snakemake + manifest hashes
- Explicit treatment_id schema for masked/unmasked variants

Experimental Scope



Full Snakemake pipeline (dev split -> stats; holdout isolated).

- 6 DRL algorithms × 5 seeds × 2 traces
- Deterministic eval per seed
- Baselines reported separately (descriptive)

Statistical Framework (Carrasco-Aligned) [1]

Core Tests (Per Metric)

Shapiro-Wilk (diagnostics)

Friedman (omnibus)

Nemenyi (pairwise) [2]

Kendall's W (effect size) [5]

New Additions vs Intro/Lit Review

- Wilcoxon non-parametric CI (paired differences) [1]
- Confidence curves (delta vs p-value)
- Page trend test (ordered treatments) [3]
- CD diagram input for rank visualizations

All tests are run per metric with seeds as blocks.

Future Steps (Submission 2 → Results)

Cluster Execution

- Apptainer export of Nix environment
- Slurm-backed Snakemake runs
- Full seed sweep on dev splits

Best Algorithm Evaluation

- Pareto selection (primaries + secondaries)
- Tie-breakers: avg_waiting → avg_slowdown → cpu_utilization
- Holdout evaluation across all seeds

Interpretation & Reporting

- PNG plots + CSV tables for Typst
- Combined DRL vs baseline visuals (descriptive)
- Final discussion of algorithm choice + masking impact

Bibliography

[1] Jacínto Carrasco, Salvador García, M Mar Rueda, Swagatam Das, and Francisco Herrera. 2020. Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review. *Swarm and Evolutionary Computation* 54, (2020), 100665.

[2] Janez Demšar. 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* 7, (2006), 1–30.

[3] Joaquín Derrac, Salvador García, Daniel Molina, and Francisco Herrera. 2011. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation* 1, 1 (2011), 3–18. https://doi.org/10.1016/j.swevo.2011.02.002

[4] Eelco Dolstra, Merijn de Jonge, and Eelco Visser. 2004. Nix: A Safe and Policy-Free System for Software Deployment. In *Proceedings of the 18th USENIX Large Installation System Administration Conference (LISA)*, 2004. 79–92.

[5] J. R. Landis and G. G. Koch. 1977. "An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers". *Biometrics* 33, 2 (June 1977), 363. https://doi.org/10.2307/2529786

[6] F Mölder, KP Jablonski, B Letcher, MB Hall, PC van Dyken, CH Tomkins-Tinch, V Sochat, J Forster, FG Vieira, C Meesters, S Lee, SO Twardziok, A Kanitz, J VanCampen, V Malladi, A Wilm, M Holtgrewe, S Rahmann, S Nahnsen, and J Köster. 2025. Sustainable data analysis with Snakemake [version 3; peer review: 2 approved]. *F1000Research* 10, 33 (2025). https://doi.org/10.12688/f1000research.29032.3

Thank You