



SESAR Innovation Days

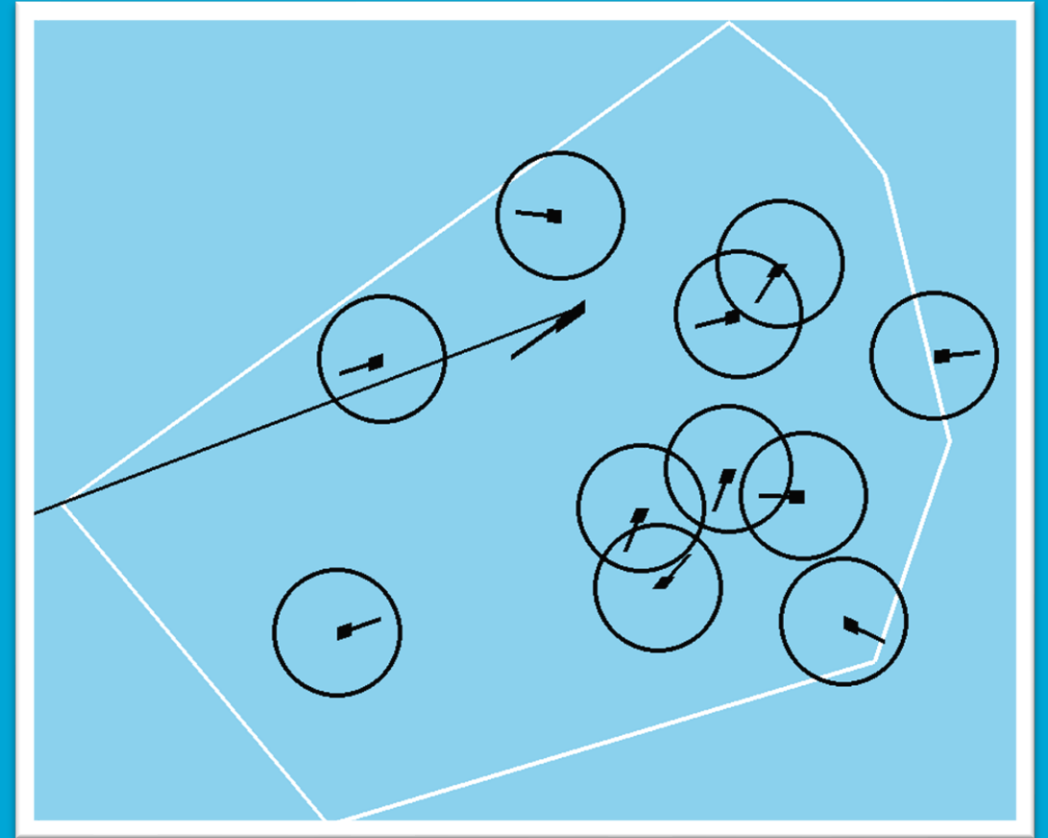
1-4 December 2025
Bled

Mixed-Fidelity Reinforcement Learning for Aircraft Conflict-Resolution

Adam Moëc – a.a.g.moec@student.tudelft.nl

Jan Groot

Joost Ellerbroek



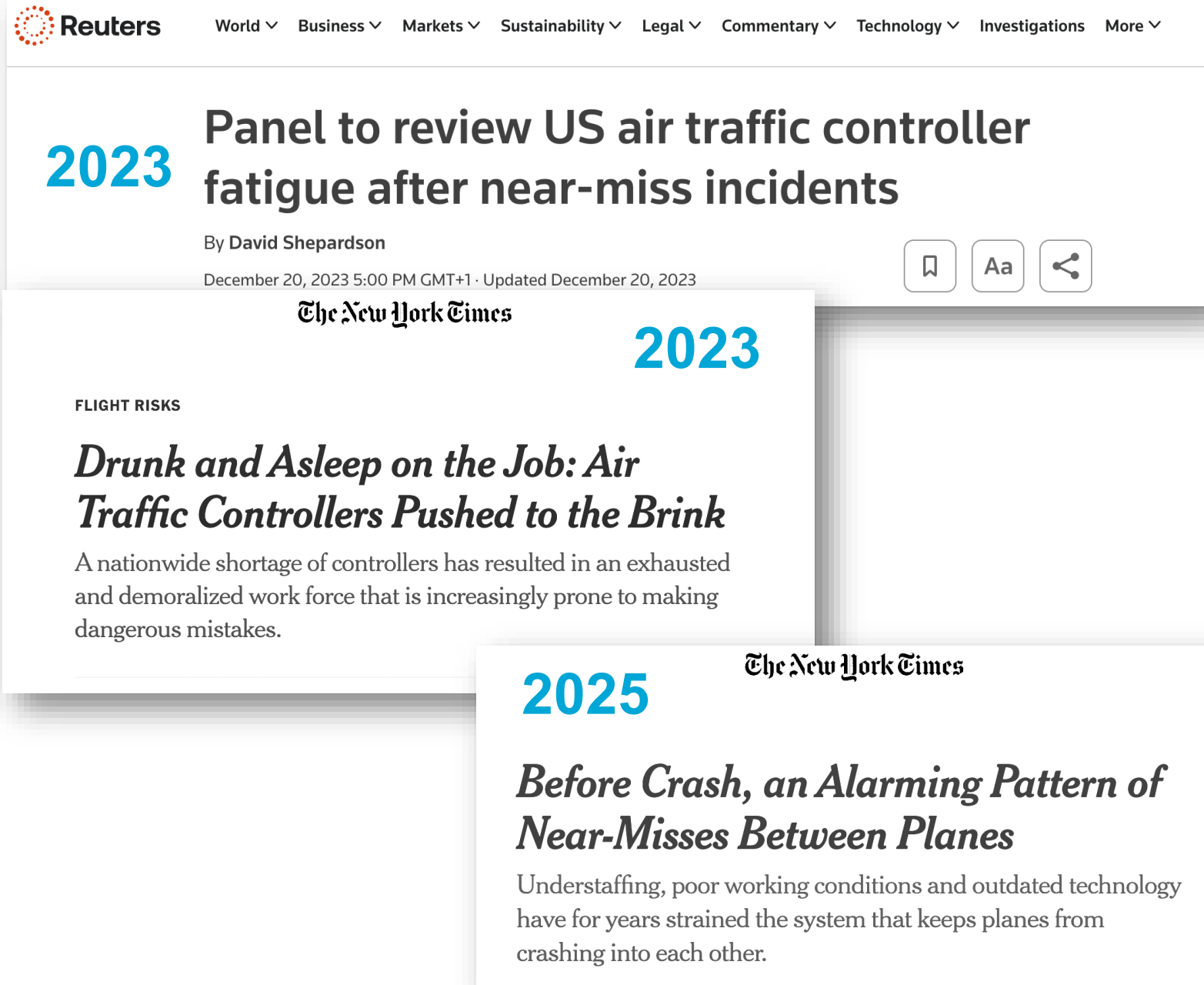
Overview

- Competing pressures in ATM research
 - Introducing mixed-fidelity RL for conflict-resolution
- Research Method & Objectives
 - And the difference between low- and high-fidelity
- Experimental Results
- Conclusion and Limitations

Competing pressures in ATM research

1. Reducing ATCo fatigue

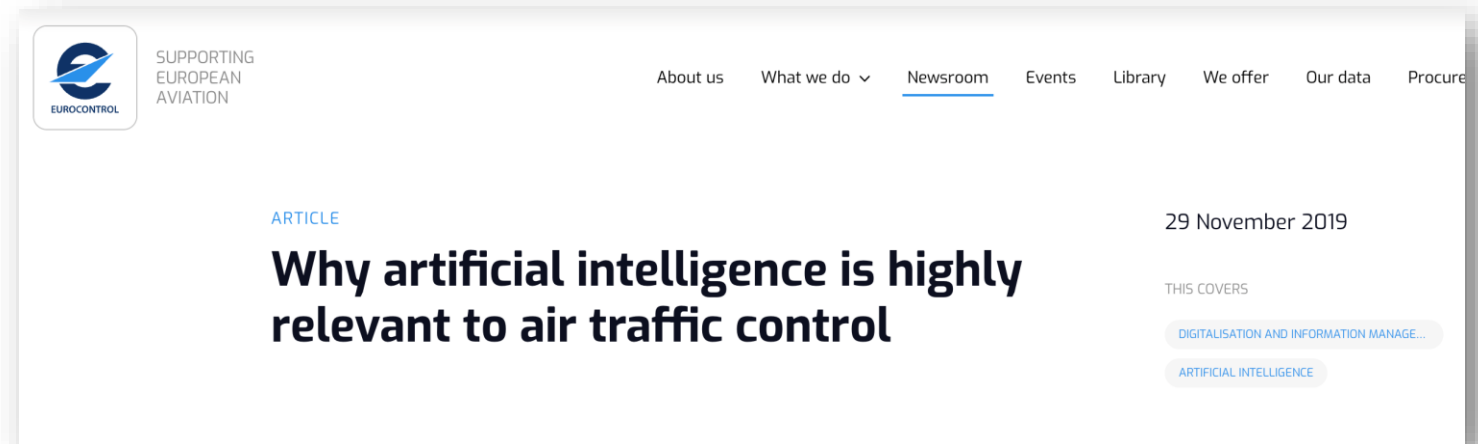
- Human-centric solutions
- ML/RL aides



Competing pressures in ATM research

1. Reducing ATCo fatigue

- Human-centric solutions
- ML/RL aides



Competing pressures in ATM research

1. Reducing ATCo fatigue

- Human-centric solutions
- ML/RL aides

2. ML/RL in ATM research

- Infrastructure uncertainties
- Financial aspect
- Environmental aspect

3. Deployability

- Aviation is safety-critical; certification mandatory



Competing pressures in ATM research

1. Reducing ATCo fatigue

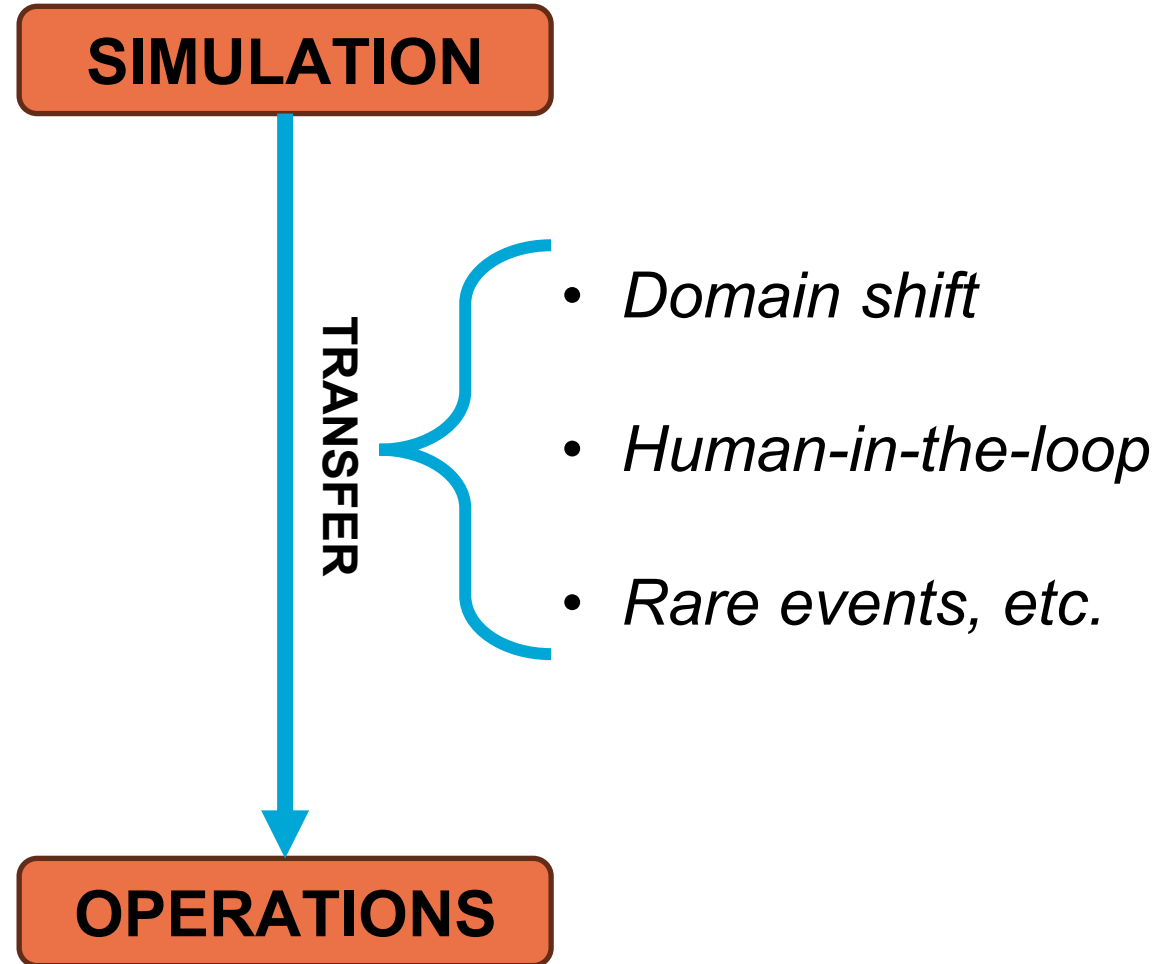
- Human-centric solutions
- ML/RL aides

2. ML/RL in ATM research

- Infrastructure uncertainties
- Financial aspect
- Environmental aspect

3. Deployability

- Aviation is safety-critical; certification mandatory
- Can we be sure of real-world performance?



Mixed-Fidelity (MiFi) RL for Aircraft Conflict-Resolution

1. Reducing ATCo fatigue

- Human-centric solutions
- ML/RL aides

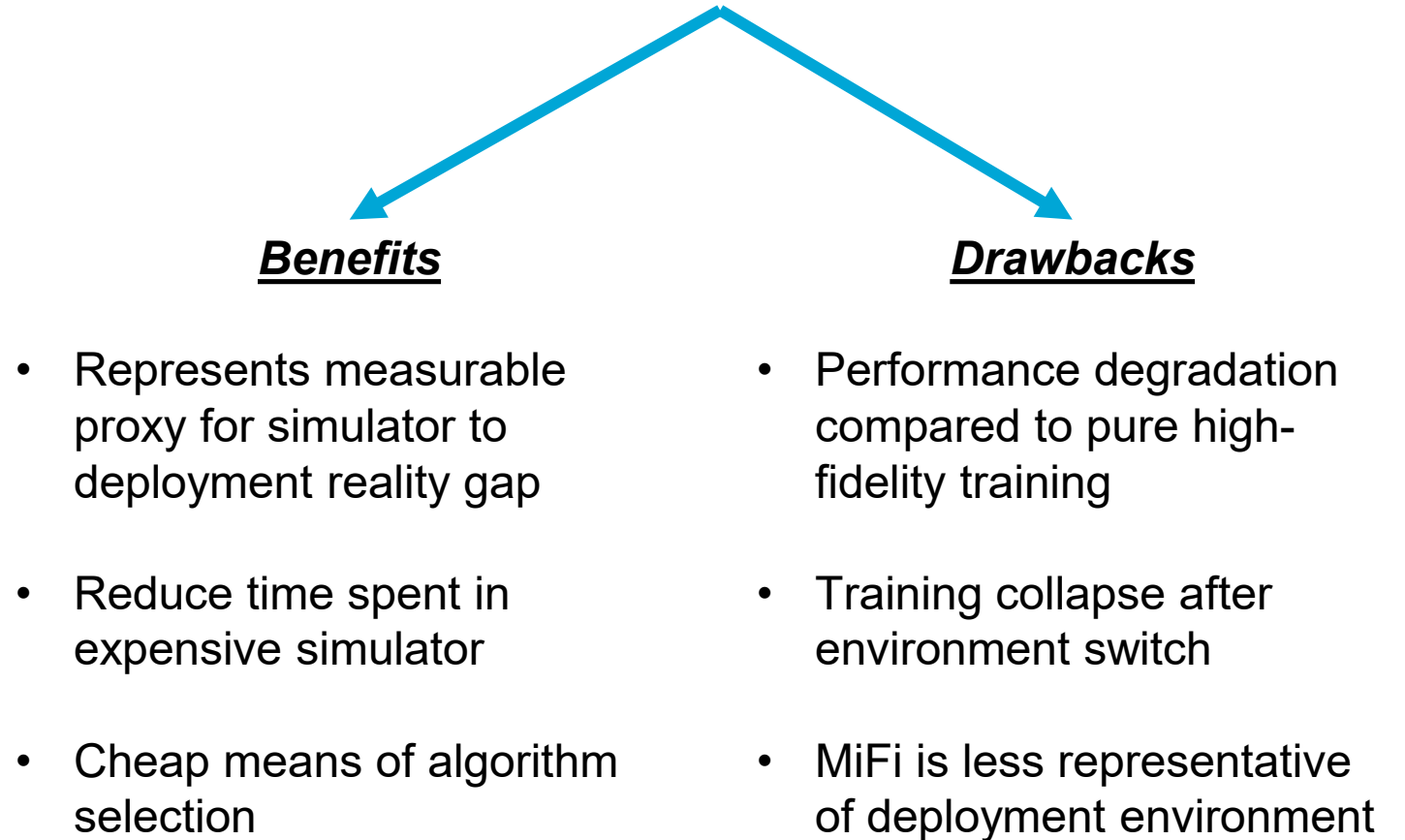
2. ML/RL in ATM research

- Infrastructure uncertainties
- Financial aspect
- Environmental aspect

3. Deployability

- Aviation is safety-critical; certification mandatory
- Can we be sure of real-world performance?

Training occurs in two distinct environments of different fidelities...



Research Method & Objectives

Research Method & Objectives

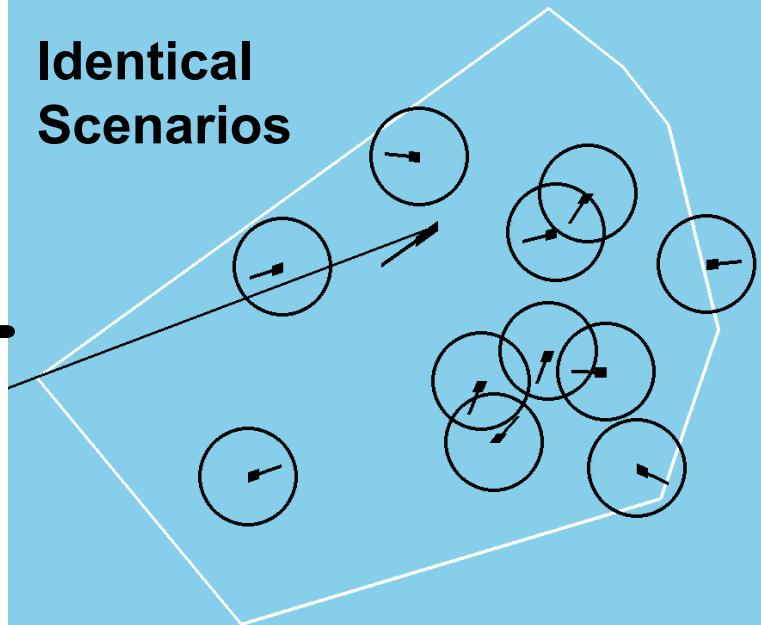
Train a RL agent for a total of 3×10^6 timesteps to perform aircraft conflict-resolution in:

1. **A low-fidelity (LoFi) simulator** – Instantaneous actions, no simulated aircraft dynamics... but cheap to run!
2. **A high(er)-fidelity (HiFi) simulator** – Using the BlueSky-Gym aircraft kinematic model... but more expensive!

LoFi

- No turning radius
- Heading/Velocity changes are instant
- Basic 2D airspace

Identical Scenarios



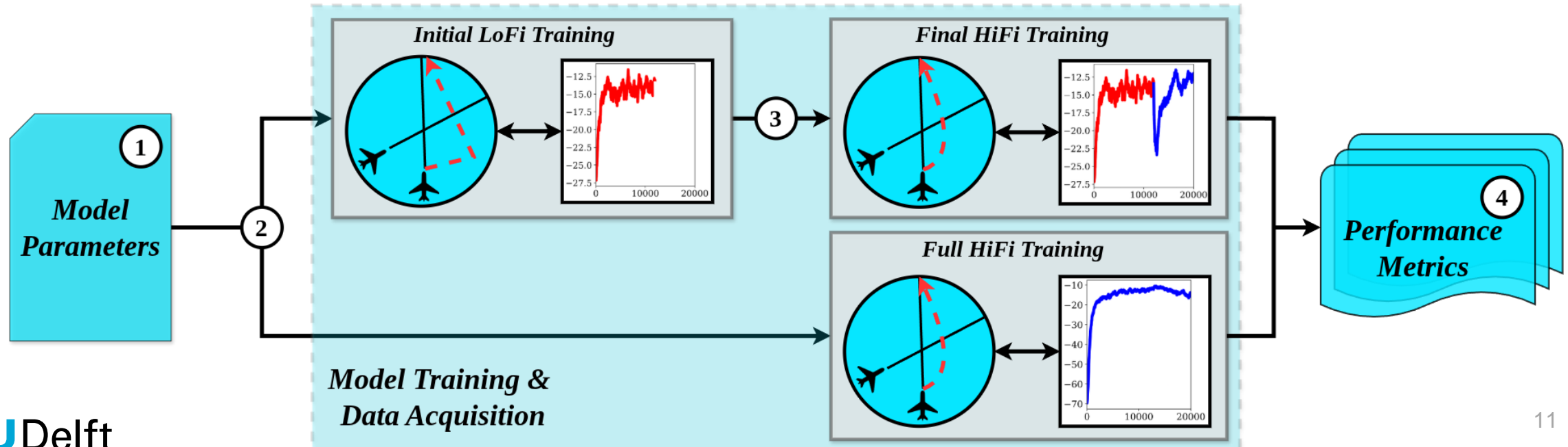
HiFi

- Aircraft are A320s
- Kinematic turn & acceleration dynamics
- Earth's curvature accounted for

Research Method & Objectives

Train a RL agent for a total of 3×10^6 timesteps to perform aircraft conflict-resolution in:

1. **A low-fidelity (LoFi) simulator** – Instantaneous actions, no simulated aircraft dynamics... but cheap to run!
2. **A high(er)-fidelity (HiFi) simulator** – Using the BlueSky-Gym aircraft kinematic model... but more expensive!



Research Method & Objectives

Determine the following **across RL algorithms and LoFi/HiFi training splits**:

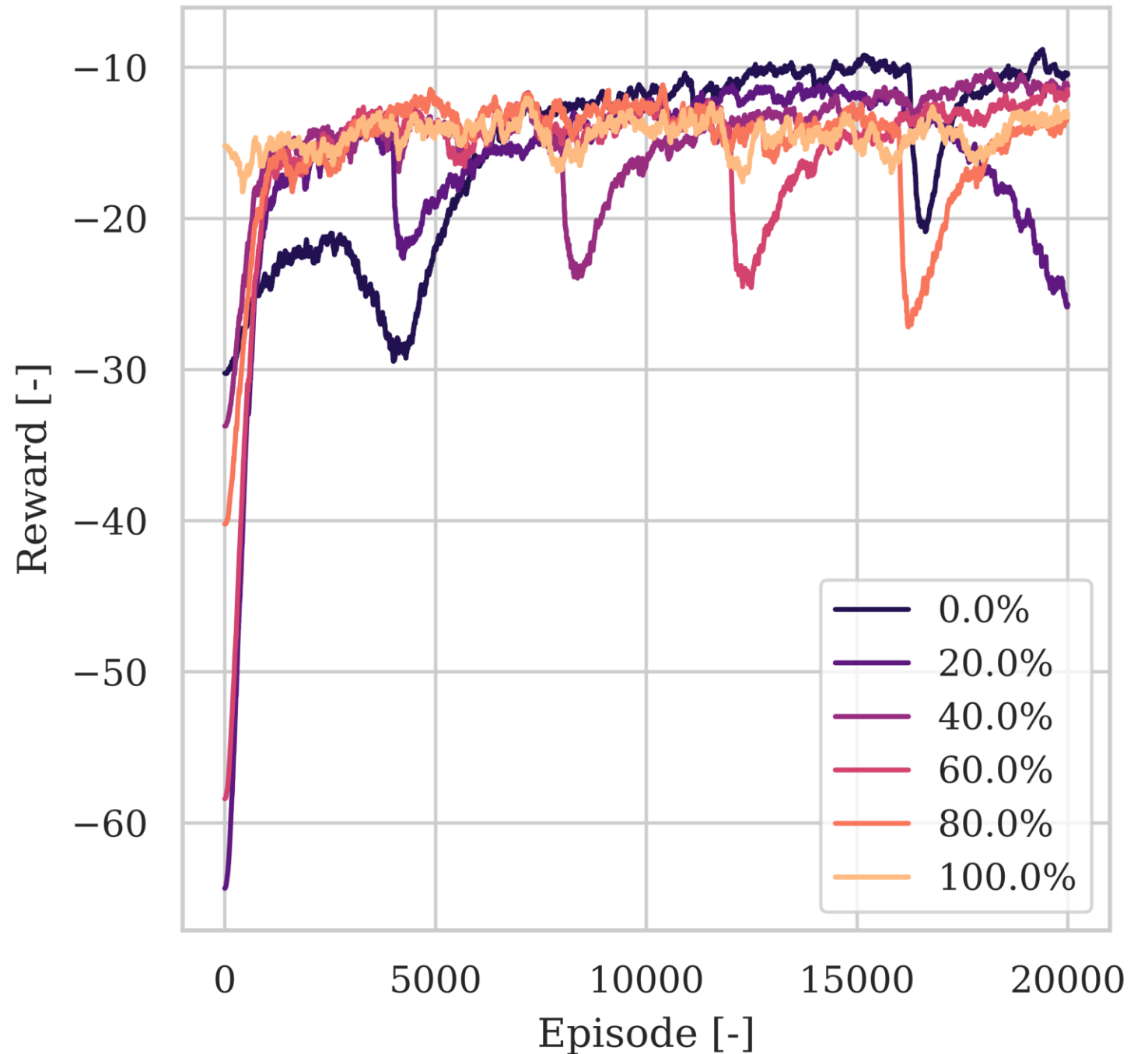
1. Performance penalty upon fidelity transfer:
 - *Proxy for HiFi/real-world fidelity gap*
2. Capacity to generalize from LoFi to HiFi environments:
 - *Impact of MiFi on evaluation performance*
3. Wall-clock training time improvements:
 - *Capacity to reduce expensive compute*
4. Predictive power of LoFi performance on HiFi performance
 - *Shortcut to algorithm selection*

Experimental Results

Experimental Results

1. Transfer Penalty

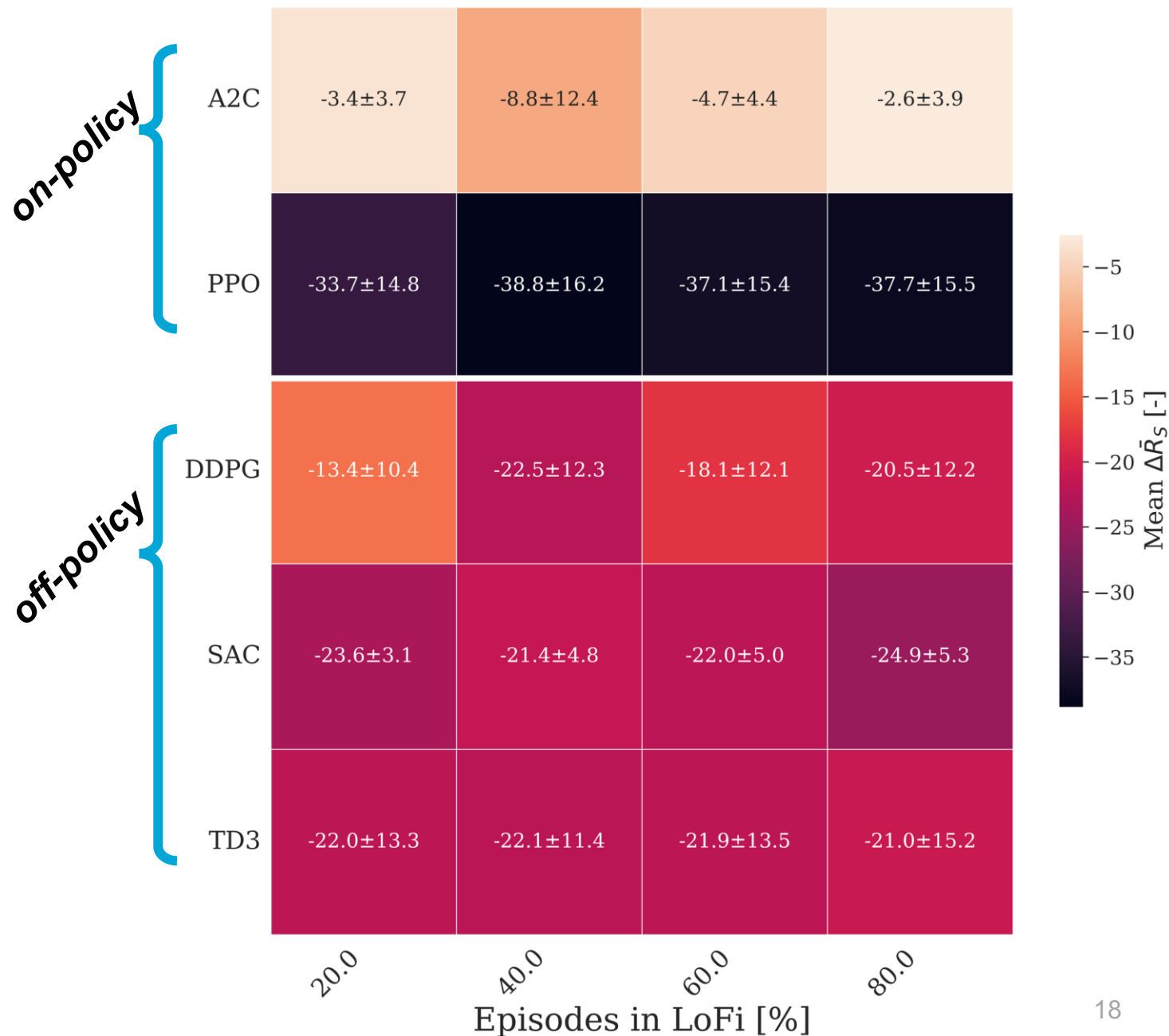
- Drop in training reward visible upon environment transfer



Experimental Results

1. Transfer Penalty

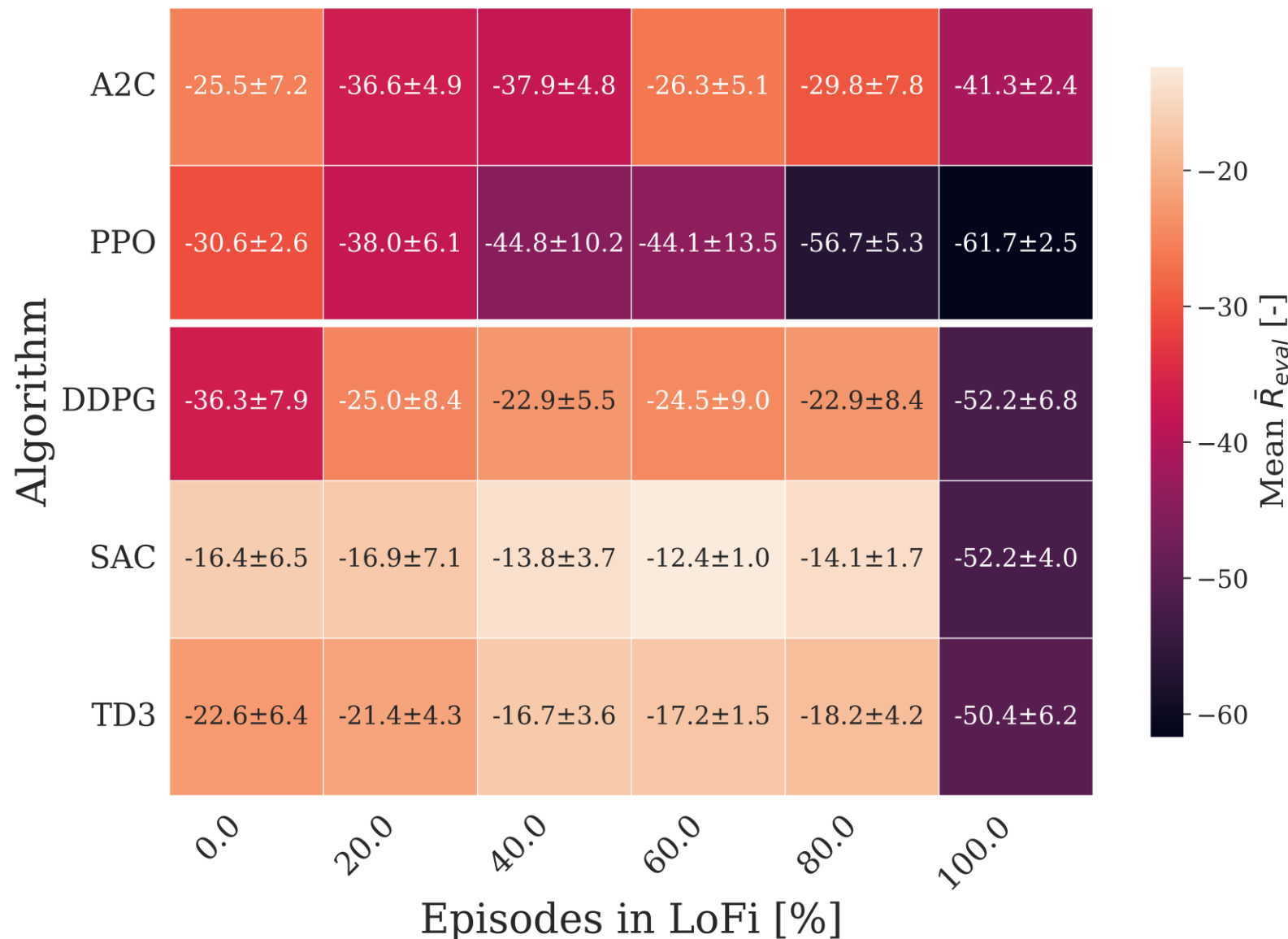
- Drop in training reward visible upon environment transfer
- No clear difference between on- and off-policy
- No clear link between time spent in LoFi & gap
- SAC experience larger penalties; but more reliable



Experimental Results

2. Evaluation Performance

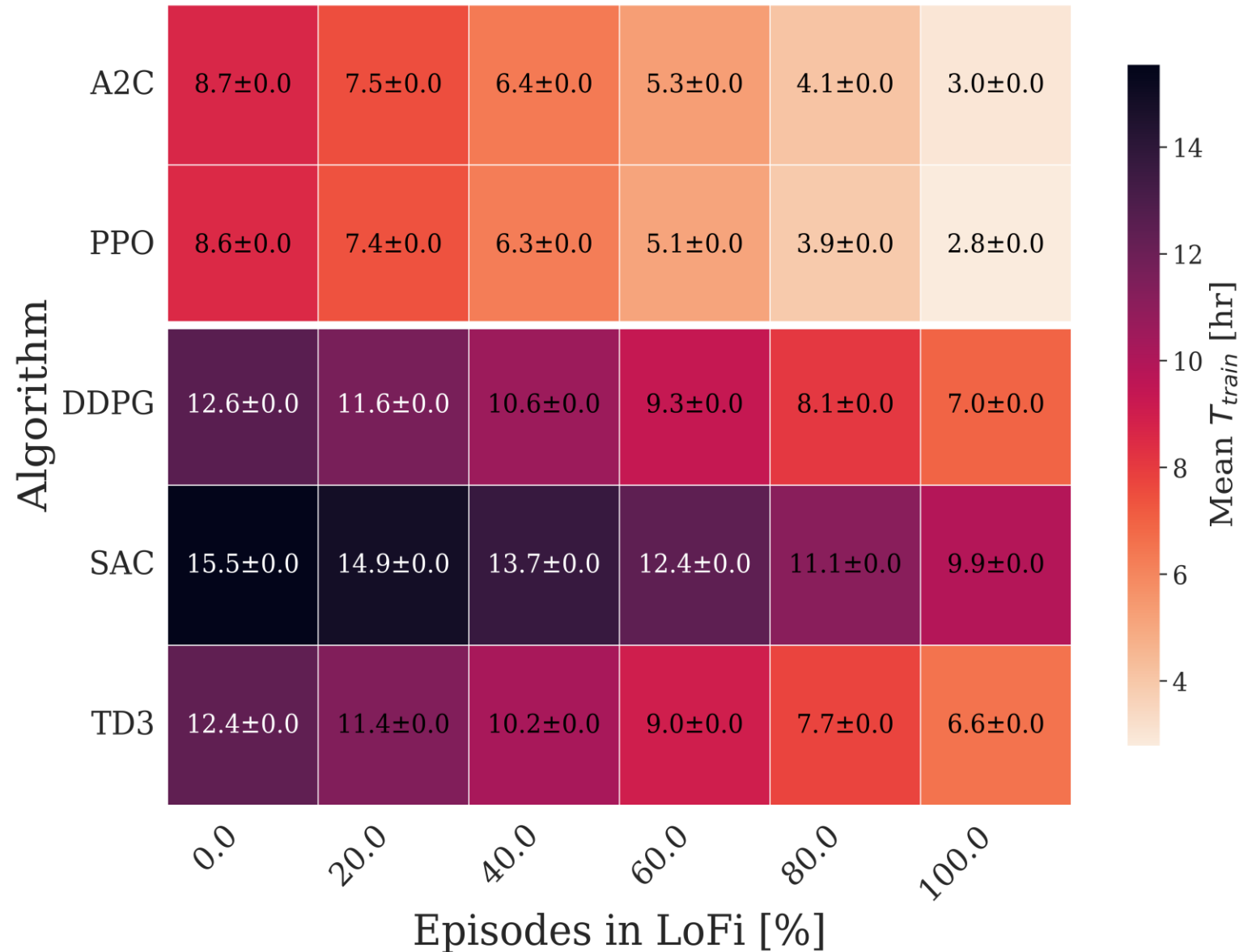
- Dichotomy between on- vs off-policy algorithms
- Off-policy algorithms *do better than generalize... they outperform*



Experimental Results

3. Training Time

- Unsurprisingly, more time spent in LoFi results in a lower overall training time
- Algorithmic vs simulation overhead is very relevant



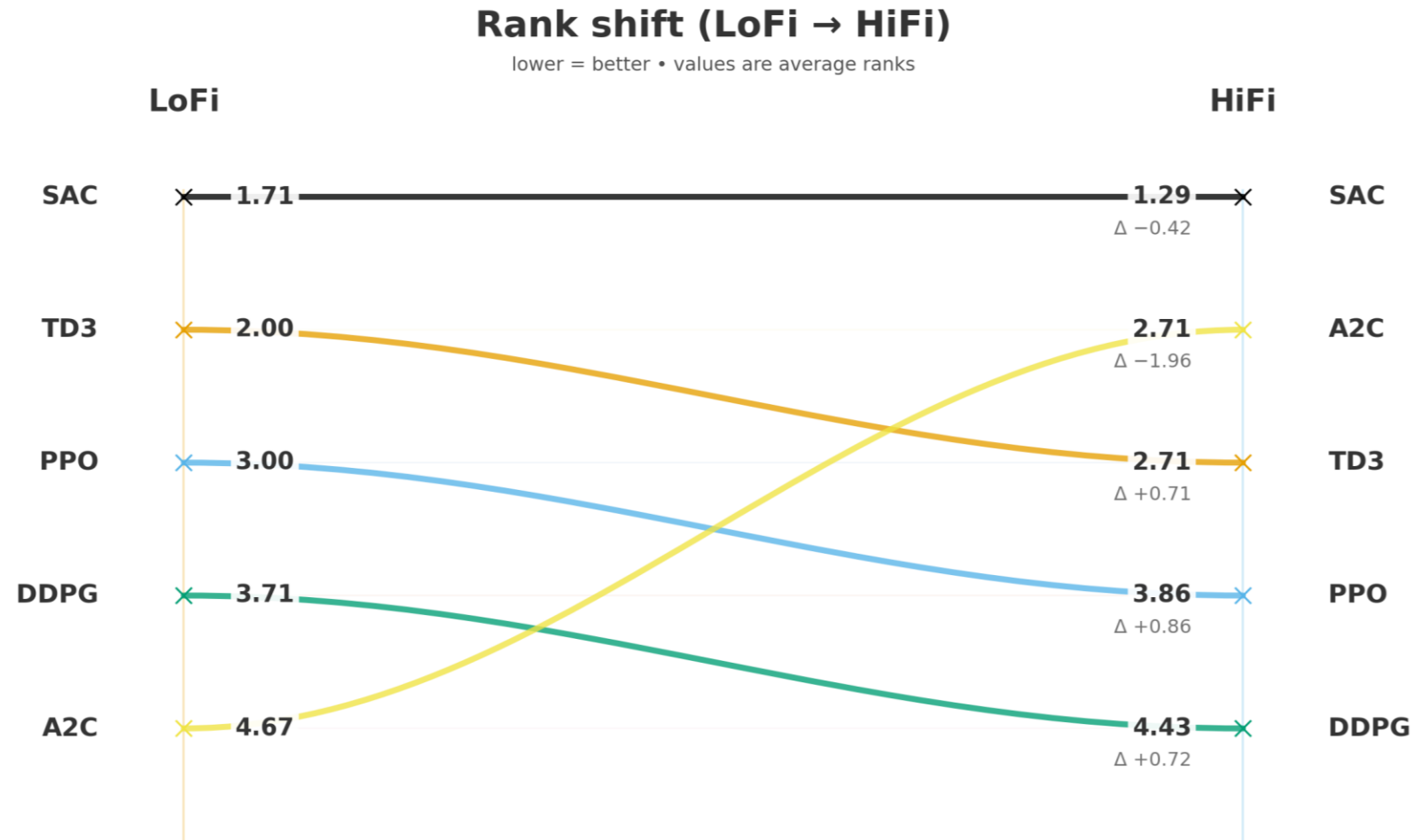
Experimental Results

3. Training Time

- Unsurprisingly, more time spent in LoFi results in a lower overall training time
- Algorithmic vs simulation overhead is very relevant

4. Predictive Capacity

- Although SAC ranks the best across both LoFi and HiFi, not statistically significant



Conclusions and Limitations

Mixed-Fidelity



- MiFi training not only **saves time** but also **outperforms pure HiFi-training for off-policy algorithms** (SAC, TD3, DDPG)
 - **No** performance degradation
 - **Reliable** trained behaviour
- On-policy (A2C, PPO) algorithms see significant performance degradation
 - **However**, A2C experiences the smallest transfer penalties: **significant for HiFi to real!**

- More samples (seeds) required to determine LoFi predictive capacity for HiFi
- **No hyperparameter tuning**
- ‘Fidelity’ is still nebulously defined
- Only a **proxy** for HiFi to real deployment!

Mixed-Fidelity Reinforcement Learning for Aircraft Conflict- Resolution

Thank you!

Adam Moëc – a.a.g.moec@student.tudelft.nl

Jan Groot

Joost Ellerbroek

