Experiments with REINFORCE algorithm

Rajesh Siraskar

12-Jul-2023 v.3.0

Agenda

1. Objective ▶

2. The RL environment ▶

3. Evaluation strategy ▶

4. Findings ▶

5. Results, Plots, Tests ▶

6. Status of Planned Objectives and Schedule ▶

Objectives

Research goal*: An optimal predictive maintenance policy for replacement of milling tool

1. Experiment with the very *basic, naïve*, REINFORCE algorithm

2. Implemented from "scratch" (Bigger objective: Start with basic REINFORCE, then keep improving it)

3. Compare against industry grade implementation of DQN, A2C and PPO (Stable-Baselines-3)

4. This is a purely experimental (empirical) exercise.

RL environment

Two different sources – simulated and real milling data

$$VB = a \cdot t^{b_1} \cdot e^{b_2 \cdot t} \quad \bigg| \begin{array}{l} t=t_k \\ t=0 \end{array}$$

◄ Simulated. Dašić, Predrag (2006): Analysis of wear cutting tools by complex power exponential function for finishing turning of the hardened steel 20CrMo5 by mixed ceramic tools.

▼ Real data. IEEE – PHM Society

RL environment - Variants

Three environments and their variants:

1. Simulated. Based on Dašić (2006). Simple single-variate state (tool wear)

    - Variants: (1) No noise (2) Low noise and low break-down chance and (3) High noise and high break-down chance

2. PHM 2010 real data – Simple single-variate state (tool wear)

    - Variants: C-01, C-04 and C-06 data-sets

    - Variants: (1) No noise (2) Low noise and low break-down chance and (3) High noise and high break-down chance

3. PHM 2010 real data – Complex multivariate state (tool wear, 3-axis forces, 3-axis vibration and acoustic data)

    - Variants: C-01, C-04 and C-06 data-sets
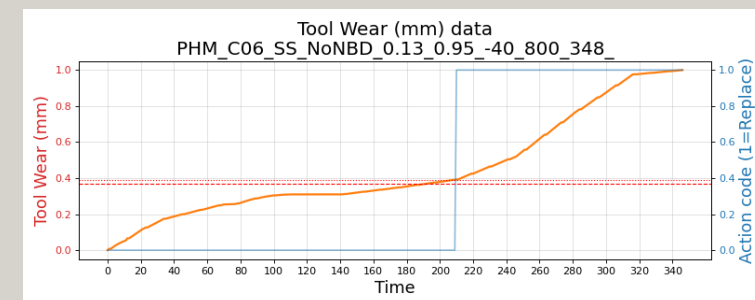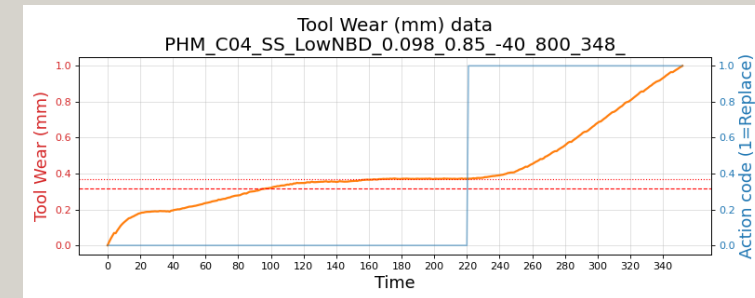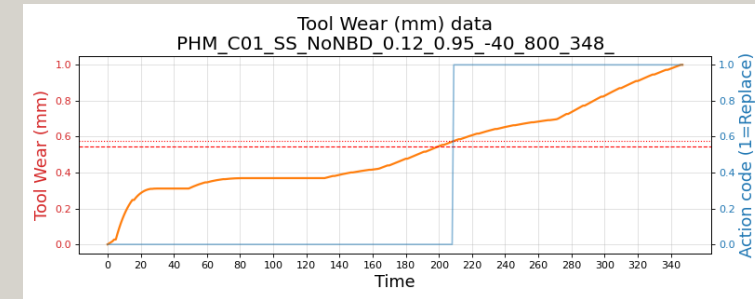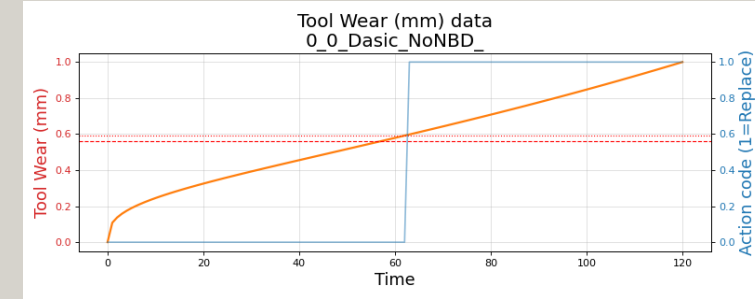
Wear plot - real data (PHM C06) and its variants
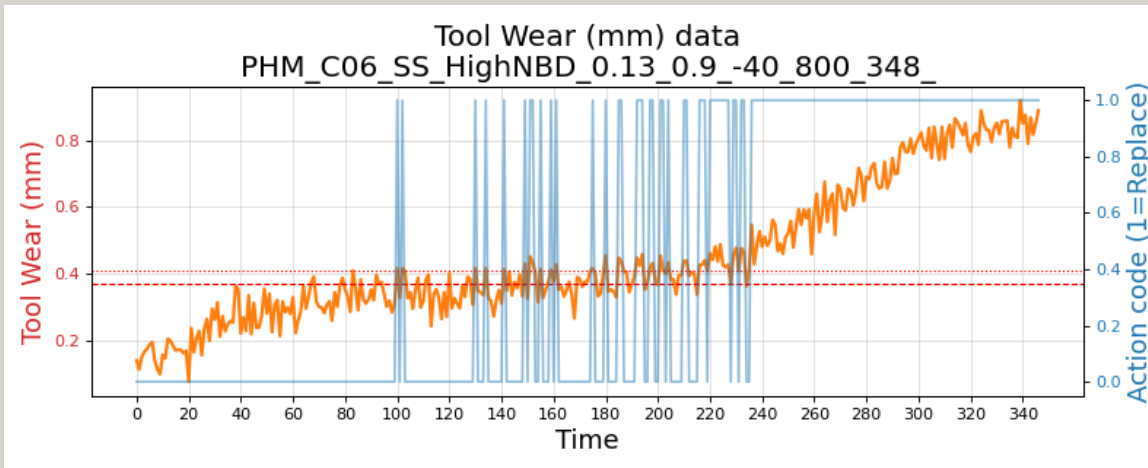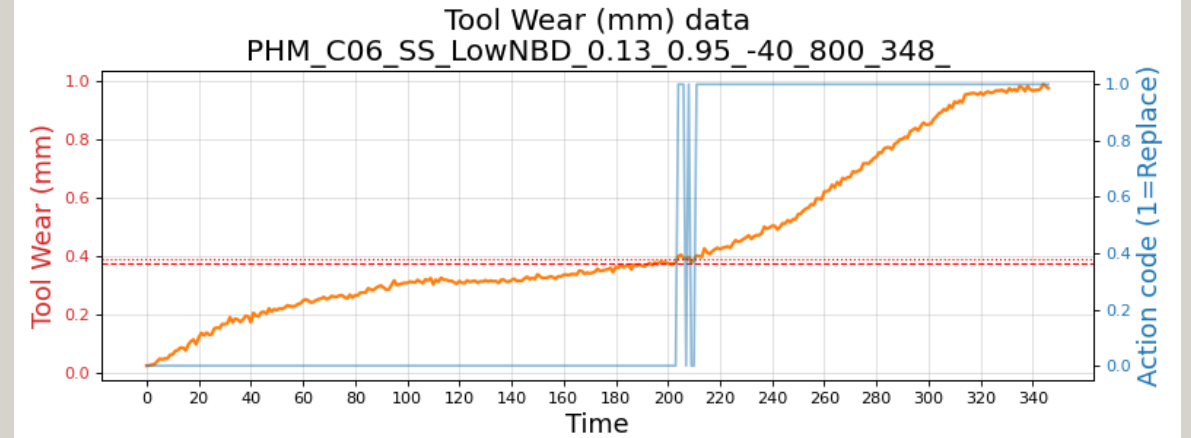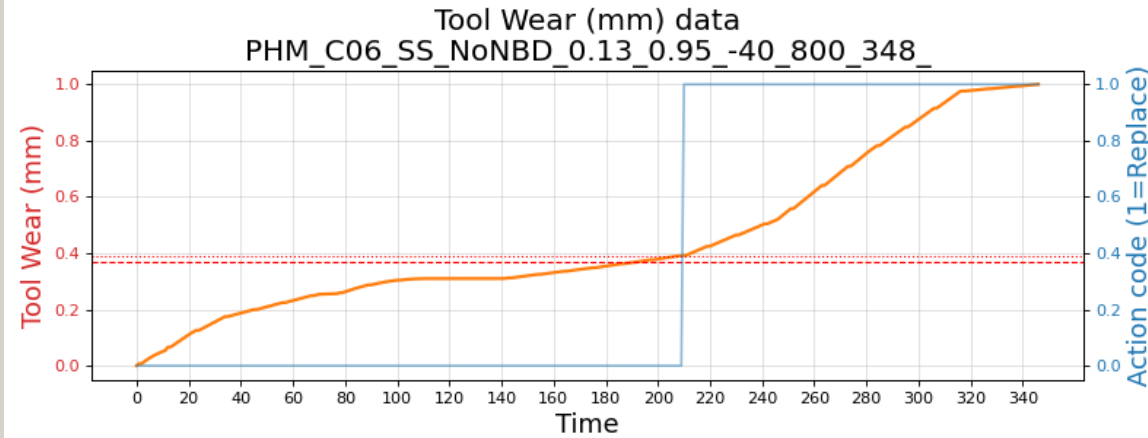
PHM 2010 C-06 data-set

- No noise, no break-down chance

- Low noise: 1e-3 and break-down chance 5%

- High noise: 1e-2 and break-down chance 10%

Evaluation strategy

- RL policy decides when to replace tool

- Compare against "human" preventive replacement policy

- REINFORCE trained for 800 episodes. SB-3 trained with 10,000 episodes

- Train over 10 rounds to understand training stability

- Evaluate: During each round, test from another test set, randomly sampled 40 points, and tested over 10 rounds

- Compute Precision, Recall and F1, F1-Beta

- Compare mean and std. deviations

Findings and Results

Findings

1. The naïve implementation of REINFORCE algorithm was implemented in PyTorch with an extremely simple architecture: One hidden layer and ReLU activation and an Adam optimizer.

2. Despite its simplicity, REINFORCE performs significantly better than the three advanced algorithms.

Average across 10 trained models, across all variants:

1. Precision: 0.687 against A2C: 0.449, DQN: 0.418, PPO: 0.472

2. F1-score: 0.609 against A2C: 0.442, DQN: 0.374, PPO: 0.345

3. Variability: Precision and F1 lower by 0.08 and 0.016, when compared to the average of A2C, DQN, PPO

"Best" model from the 10, across all variants:

1. Precision: 0.884 against A2C: 0.520, DQN: 0.651, PPO: 0.558

2. F1-score : 0.873 against A2C: 0.639, DQN: 0.740, PPO: 0.580

Results: Overall – all environments and their variants

Overall:

(1) Simulated x 3 with noise levels

(2) PHM Real-data: Uni-variate state x 3 data sets x 3 noise levels

(3) PHM Real-data: Multi-variate state x 3 data sets

	Precision		Recall		F1 score		F Beta (0.5)	
	μ	σ	μ	σ	μ	σ	μ	σ
A2C	0.449	0.088	0.480	0.084	0.442	0.070	0.436	0.071
DQN	0.418	0.185	0.504	0.032	0.374	0.035	0.348	0.058
PPO	0.472	0.144	0.316	0.087	0.345	0.091	0.393	0.105
REINFORCE	**0.687**	**0.059**	**0.629**	**0.051**	**0.609**	**0.050**	**0.631**	**0.052**

Results: Simple single variate state. Including noise variants

Simulated – Dašić, Predrag (2006).

	Precision		Recall		F1 score		F Beta (0.5)	
	μ	σ	μ	σ	μ	σ	μ	σ
A2C	0.416	0.120	0.385	0.073	0.363	0.072	0.373	0.082
DQN	0.432	0.184	0.510	0.031	0.374	0.034	0.351	0.056
PPO	0.500	0.178	0.215	0.081	0.285	0.099	0.370	0.122
REINFORCE	**0.806**	**0.040**	**0.915**	**0.038**	**0.841**	**0.035**	**0.816**	**0.037**

PHM 2010: Single-variate environment, across three data sets C-01, C-04 and C-06

	Precision		Recall		F1 score		F Beta (0.5)	
	μ	σ	μ	σ	μ	σ	μ	σ
A2C	0.447	0.077	0.477	0.091	0.452	0.072	0.446	0.070
DQN	0.419	0.179	0.507	0.032	0.379	0.036	0.352	0.057
PPO	0.450	0.146	0.314	0.082	0.333	0.087	0.374	0.102
REINFORCE	**0.605**	**0.046**	**0.603**	**0.046**	**0.570**	**0.041**	**0.576**	**0.040**

Results: Complex, multi-variate environment

PHM 2010: Complex, multi-variate environment, across three data sets C-01, C-04 and C-06
No noise or break-down

	Precision		Recall		F1 score		F Beta (0.5)	
	μ	σ	μ	σ	μ	σ	μ	σ
A2C	0.487	0.086	0.582	0.075	0.488	0.063	0.467	0.065
DQN	0.399	0.204	0.491	0.032	0.361	0.035	0.332	0.060
PPO	0.512	0.107	0.422	0.107	0.441	0.096	0.472	0.096
REINFORCE	**0.813**	**0.119**	**0.421**	**0.079**	**0.495**	**0.090**	**0.609**	**0.101**

Results: Overall – Training across 10 rounds
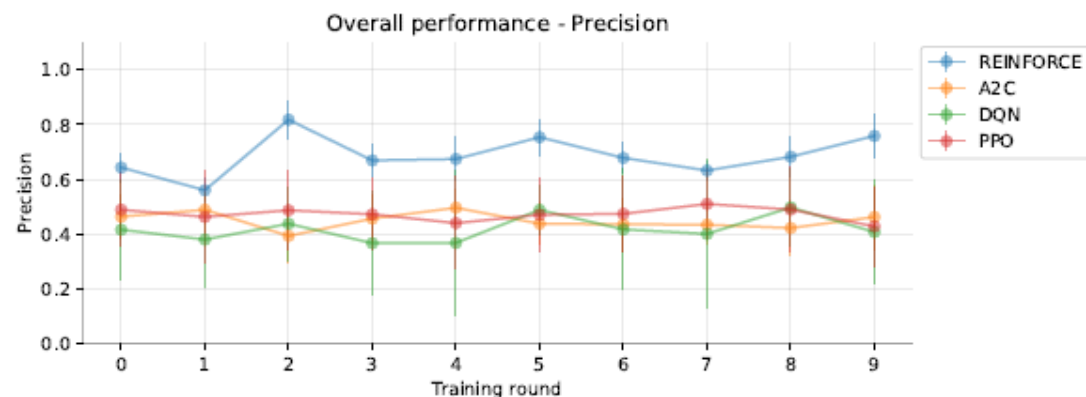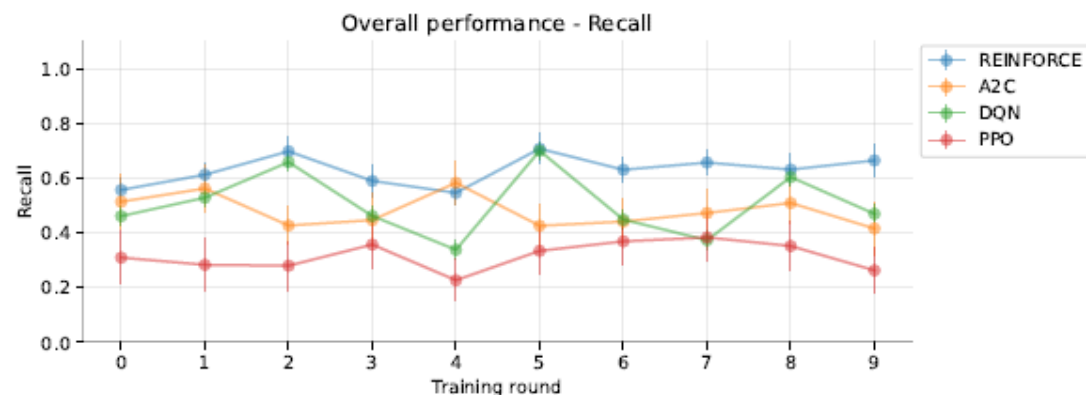
(a) Precision

(b) Recall

(c) F1-score

(d) F1-beta (0.5)

Results: Simulated – Training across 10 rounds

(a) Precision

(b) Recall

(c) F1-score

(d) F1-beta (0.5)

Results: PHM Real Data – **3** data-sets - Simple univariate

(a) Precision

(b) Recall

(b) Recall

(c) F1-score

(d) F1-beta (0.5)

Results: PHM Real Data – **3** data-sets – Multi-variate

(a) Precision

(b) Recall

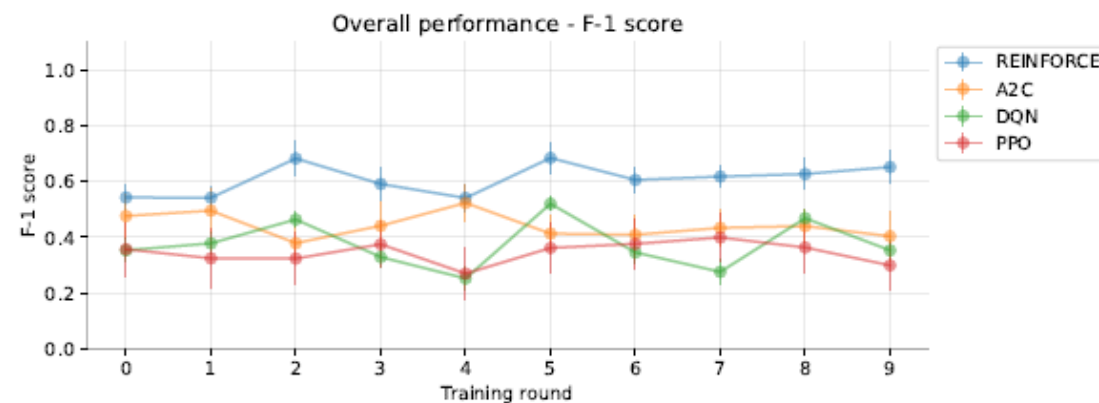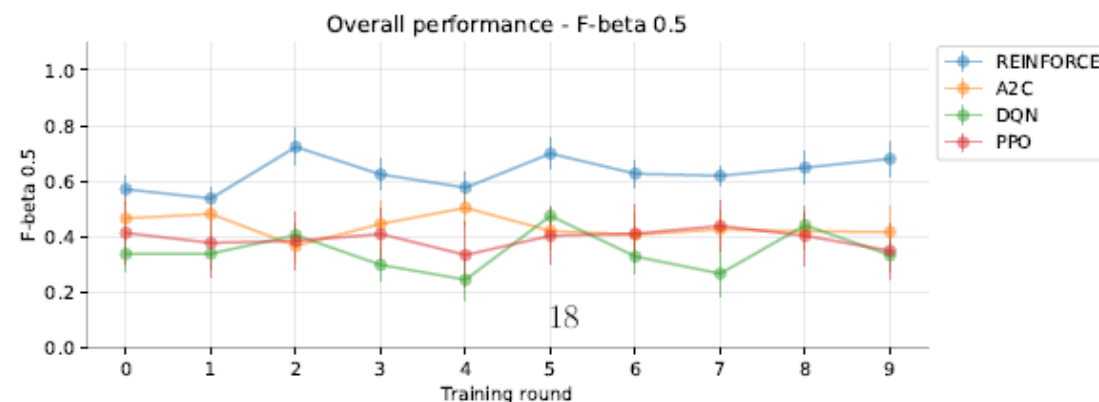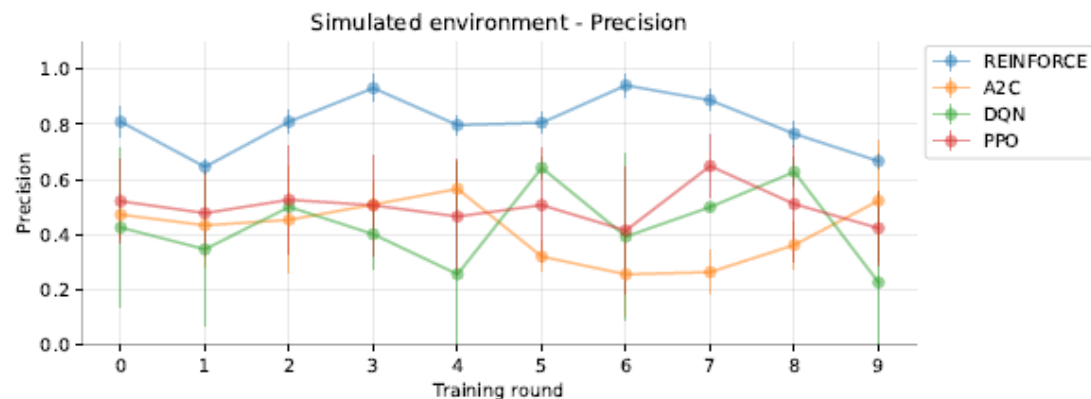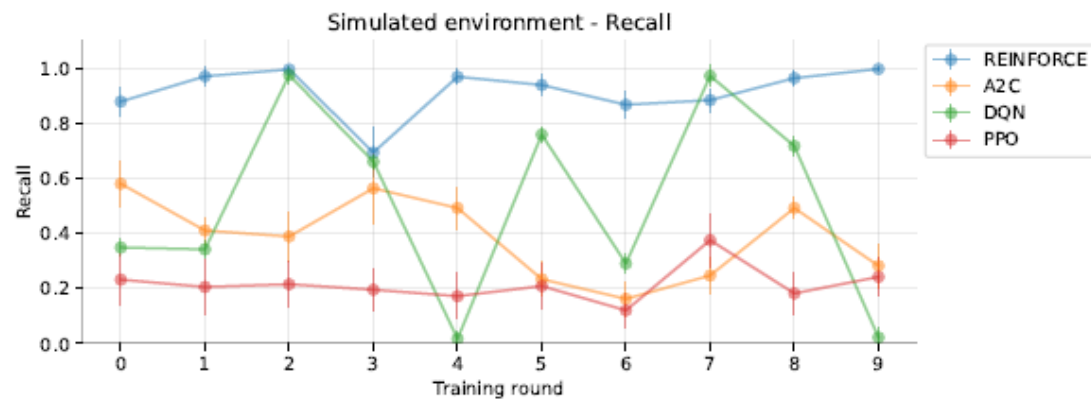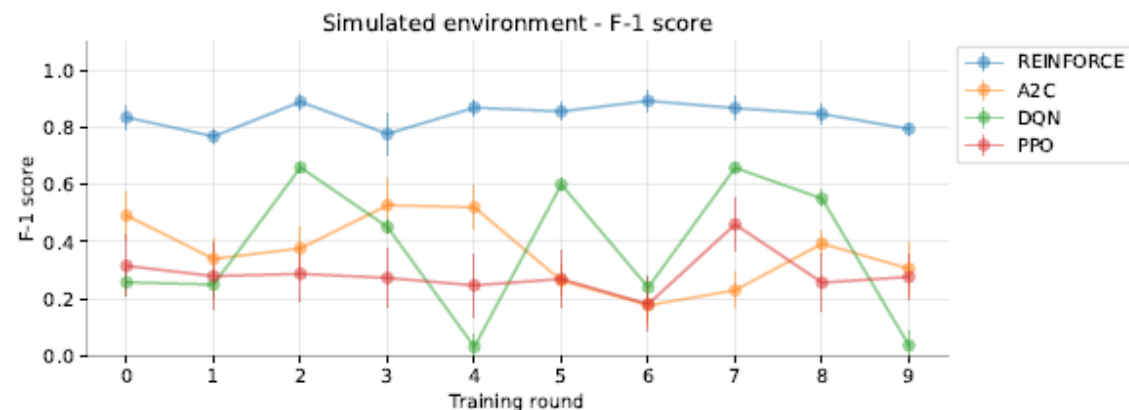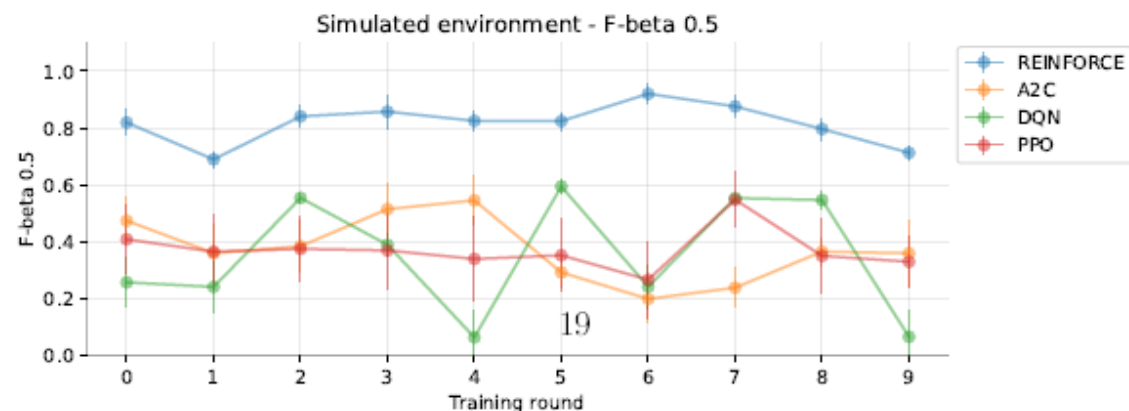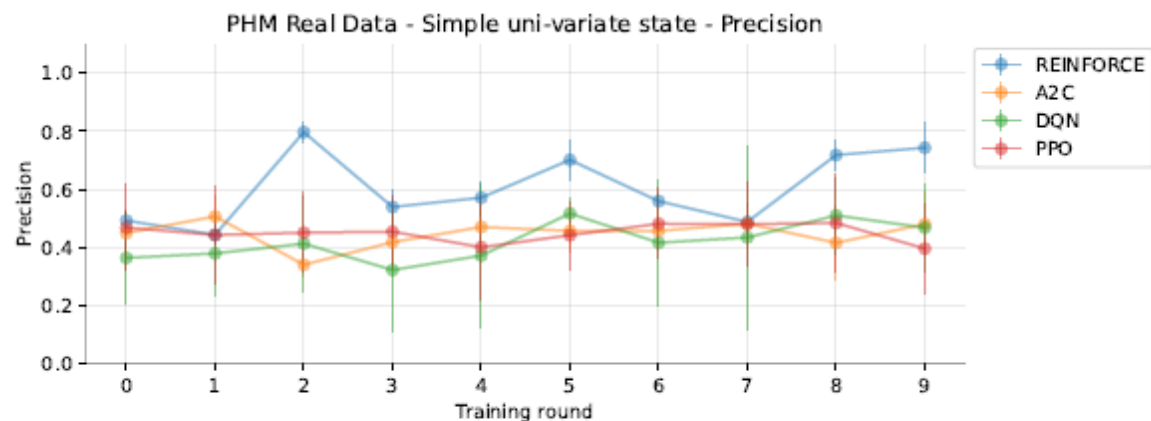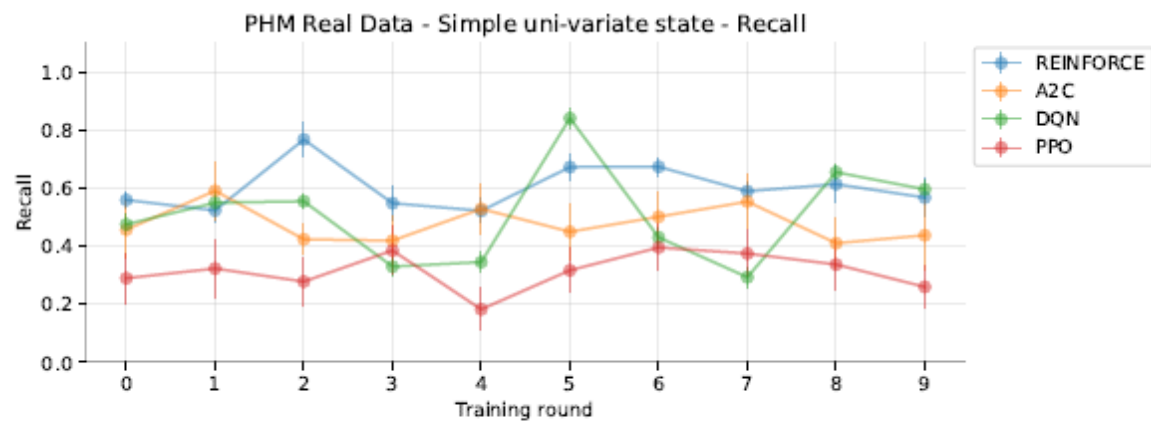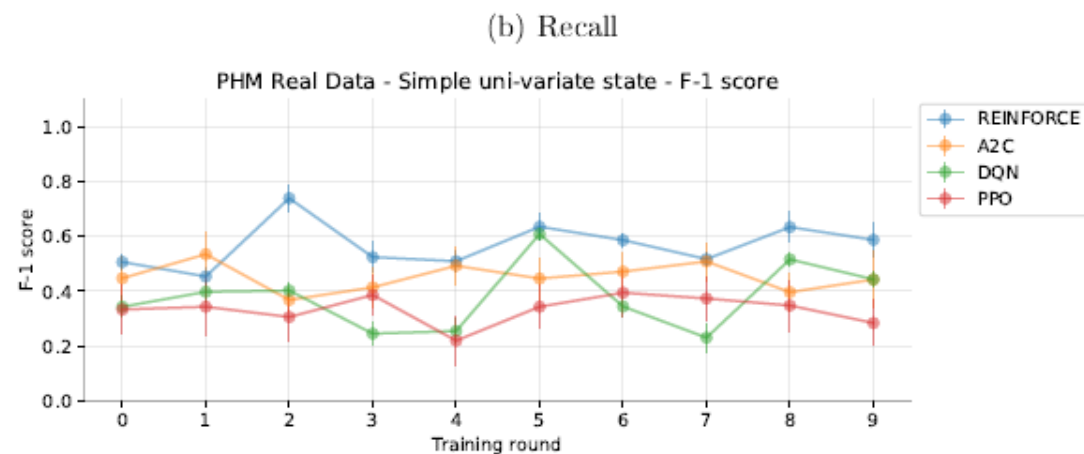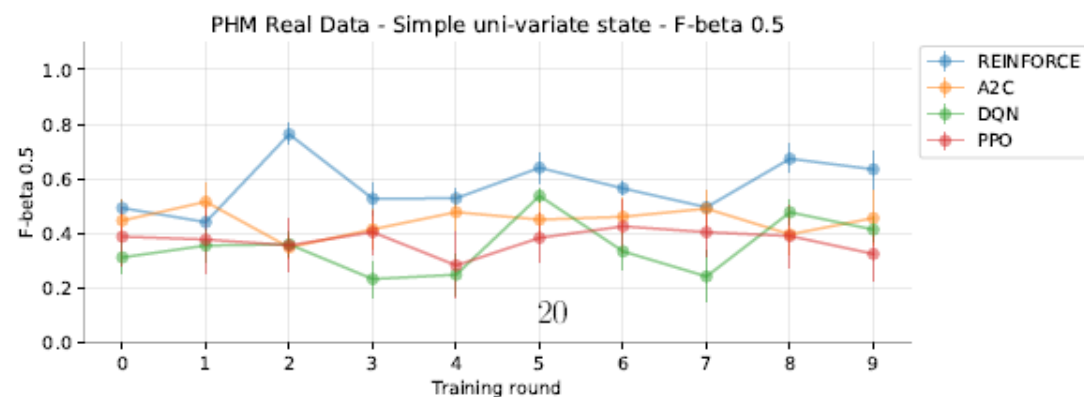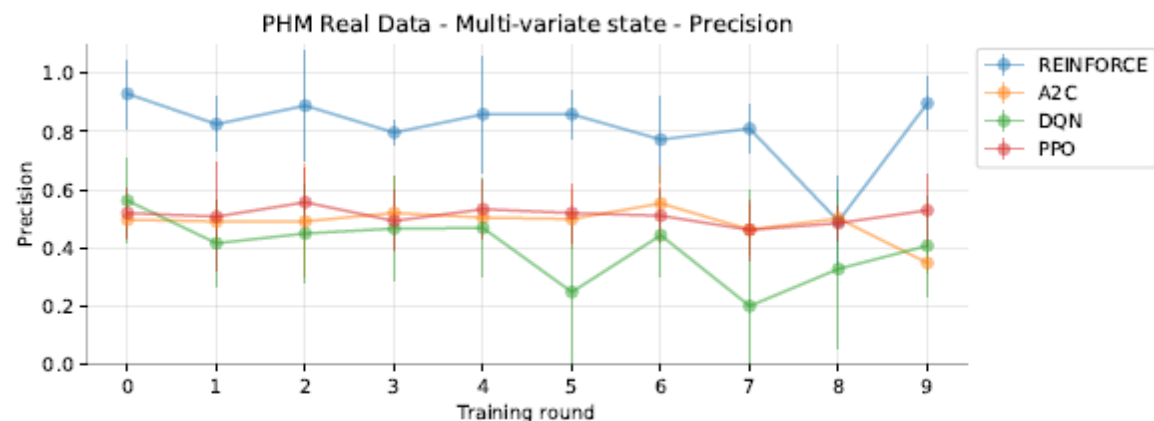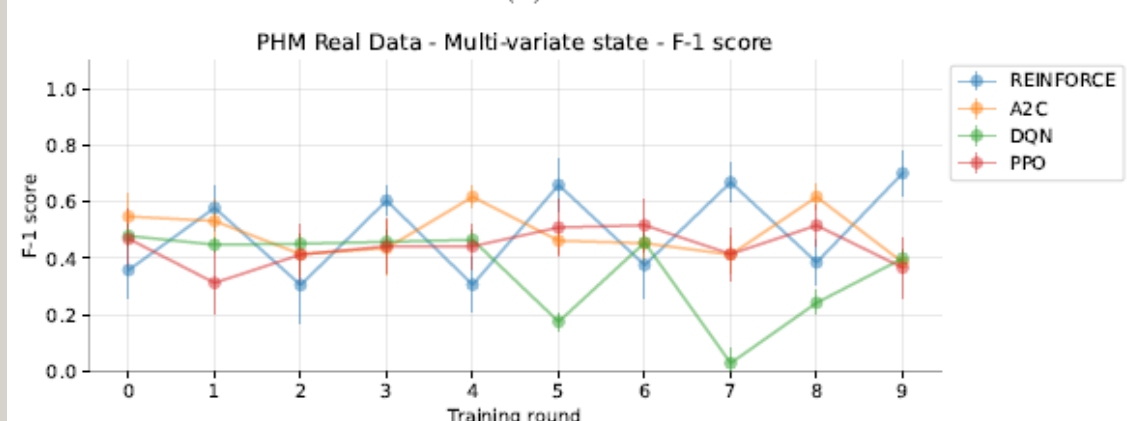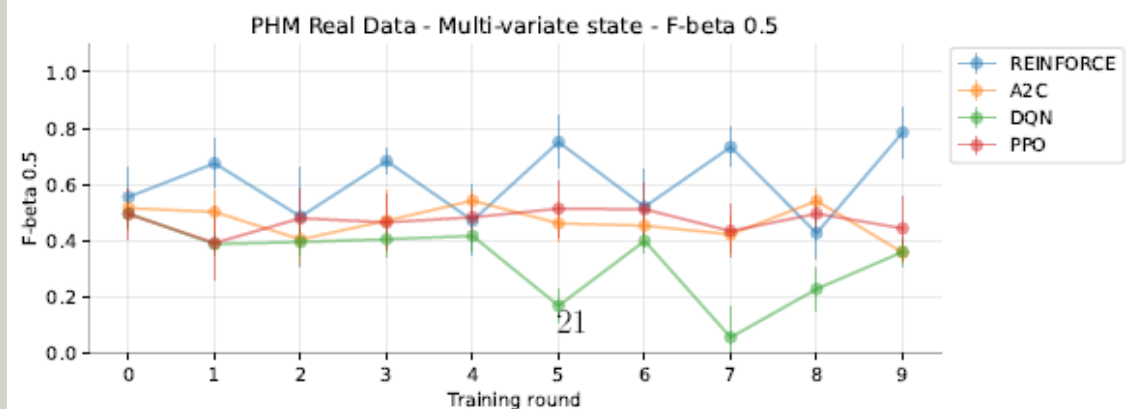(c) F1-score

(d) F1-beta (0.5)

Statistical validation: Two sample, one-tail test

Two sample, one-tail test

$\alpha = 0.05$, 95% confidence

$$H_0 : \mu_{RF} - \mu_{AA} = 0,$$
$$H_a : \mu_{RF} - \mu_{AA} > 0,$$

$$\forall \ AA \in [A2C, DQN, PPO]$$

Average metric

	REINFORCE	A2C	DQN	PPO
Overall				
Precision	0.687	0.449	0.418	0.472
Recall	0.629	0.480	0.504	0.316
F_1_Score	0.609	0.442	0.374	0.345
Simulated				
Precision	0.806	0.415	0.431	0.500
Recall	0.915	0.385	0.510	0.215
F_1_Score	0.841	0.363	0.374	0.284
PHM-SS				
Precision	0.605	0.447	0.419	0.450
Recall	0.603	0.477	0.507	0.314
F_1_Score	0.570	0.452	0.379	0.333
PHM-MS				
Precision	0.813	0.487	0.399	0.511
Recall	0.421	0.582	0.491	0.422
F_1_Score	0.495	0.488	0.360	0.441

p values

	Samples	A2C	DQN	PPO
Overall	1500			
Precision		4.31E-126	2.17E-109	2.81E-106
Recall		4.20E-35	3.37E-16	4.36E-150
F_1_Score		1.99E-64	1.46E-88	5.29E-155
Simulated	300			
Precision		3.20E-98	1.69E-63	2.65E-81
Recall		8.12E-104	2.56E-41	1.57E-264
F_1_Score		9.60E-134	8.56E-99	2.96E-242
PHM-SS	900			
Precision		2.27E-32	7.29E-31	9.95E-31
Recall		1.27E-16	1.55E-06	8.19E-71
F_1_Score		1.94E-19	4.67E-34	2.19E-67
PHM-MS	300			
Precision		1.64E-60	3.34E-54	7.88E-59
Recall		2.69E-10	2.69E-02	9.68E-01
F_1_Score		7.27E-01	1.44E-08	1.35E-03

t statistic

	A2C	DQN	PPO
Overall			
Precision	25.071	23.170	22.804
Recall	12.522	8.206	27.650
F_1_Score	17.364	20.634	28.160
Simulated			
Precision	25.611	19.032	22.427
Recall	26.665	14.558	62.541
F_1_Score	32.402	25.719	56.575
PHM-SS			
Precision	12.082	11.770	11.742
Recall	8.357	4.821	18.607
F_1_Score	9.121	12.423	18.098
PHM-MS			
Precision	18.451	17.207	18.122
Recall	-6.425	-2.219	-0.041
F_1_Score	0.349	5.748	3.220

Some plots

Training plots for REINFORCE algorithm

PHM 2010 C-04 data-set: Variant: Single-state tool-wear with low noise (1e-3) and low break-down chance (5%)
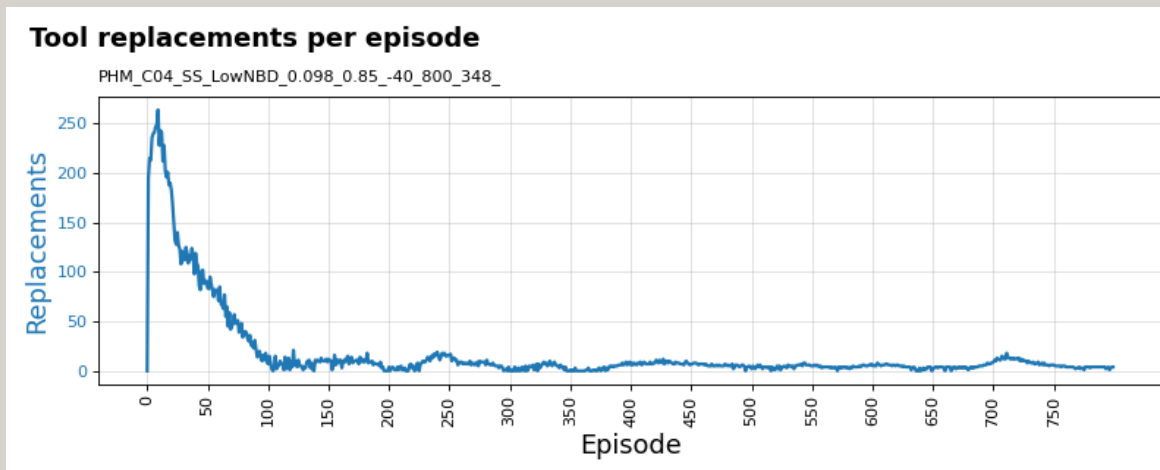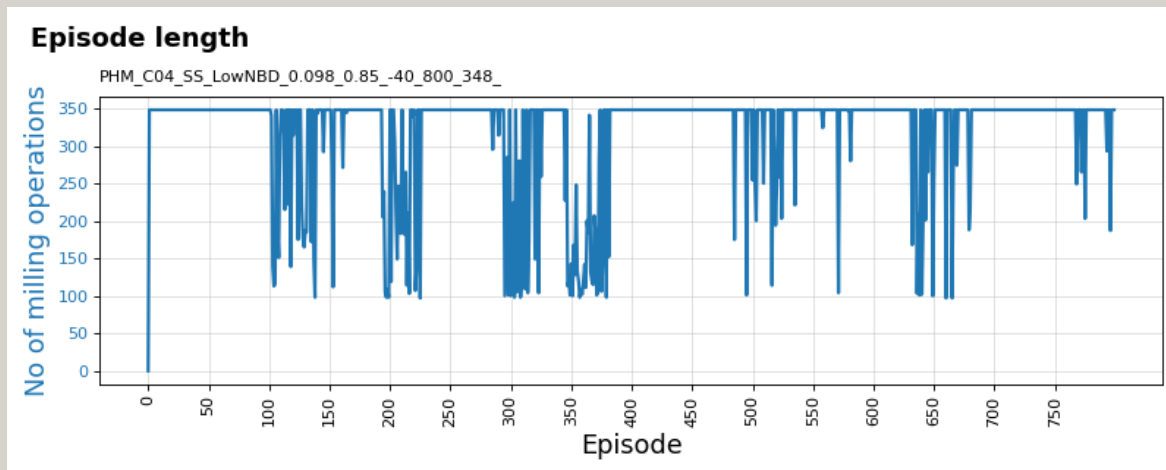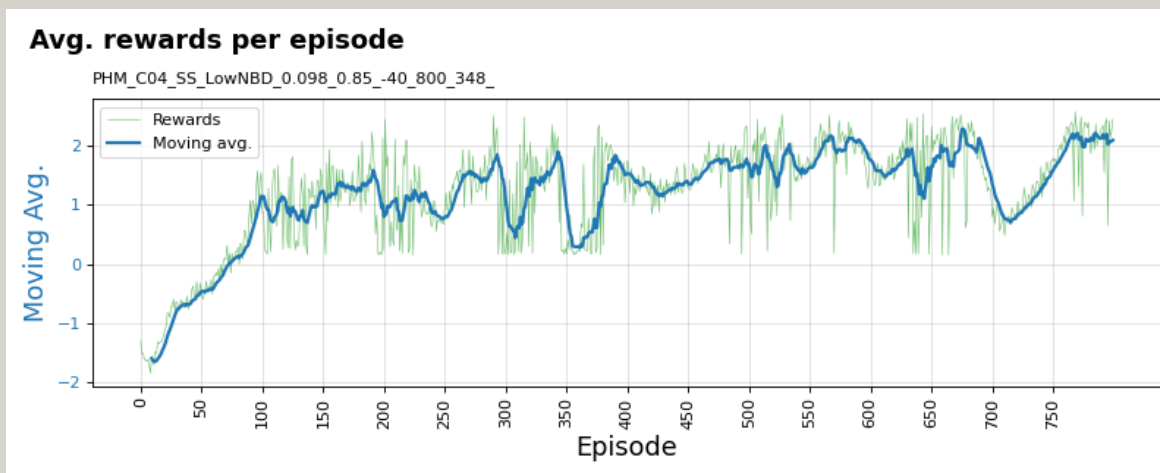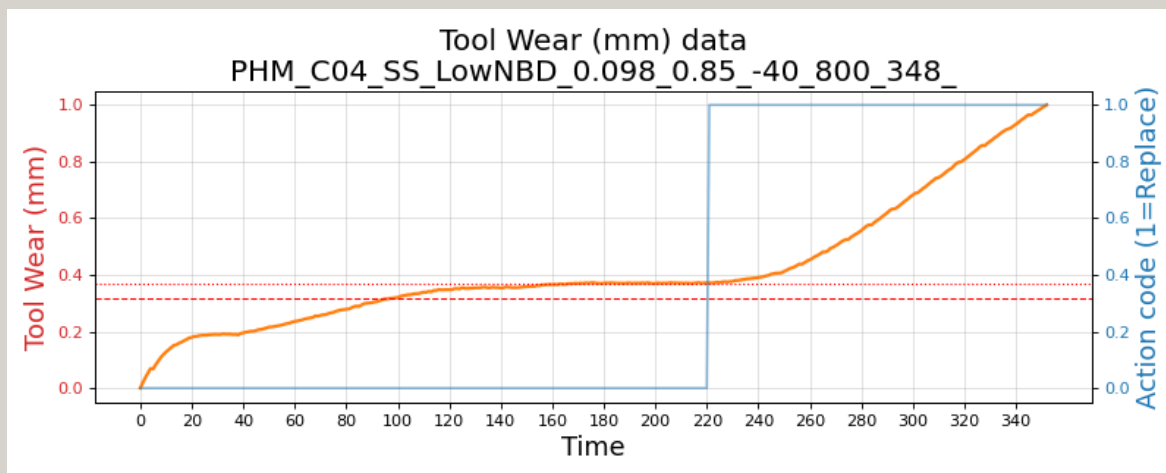
Training plots for REINFORCE algorithm

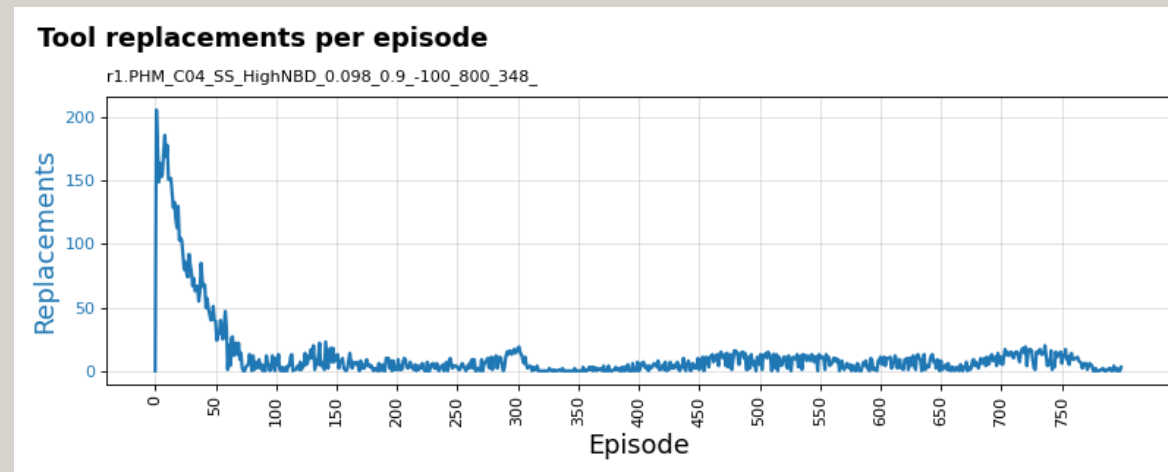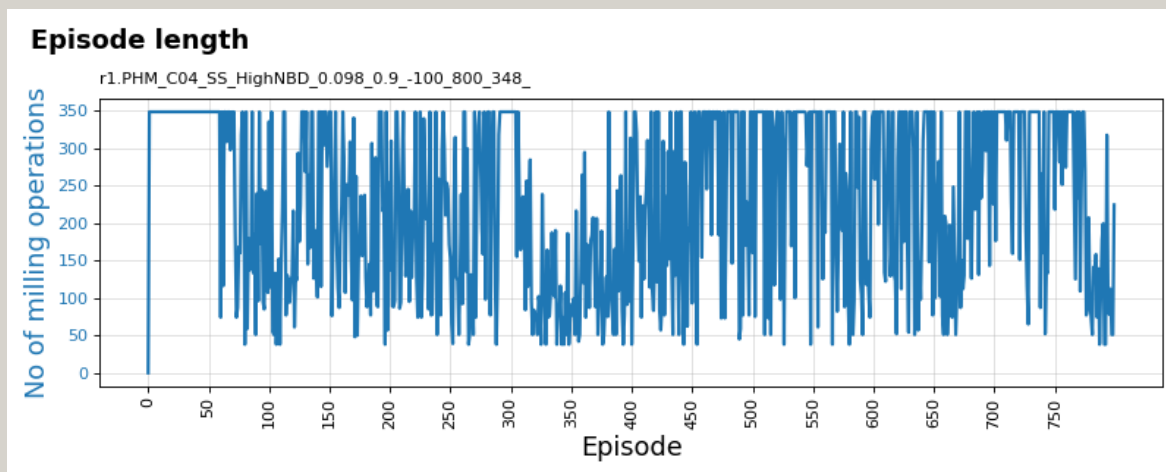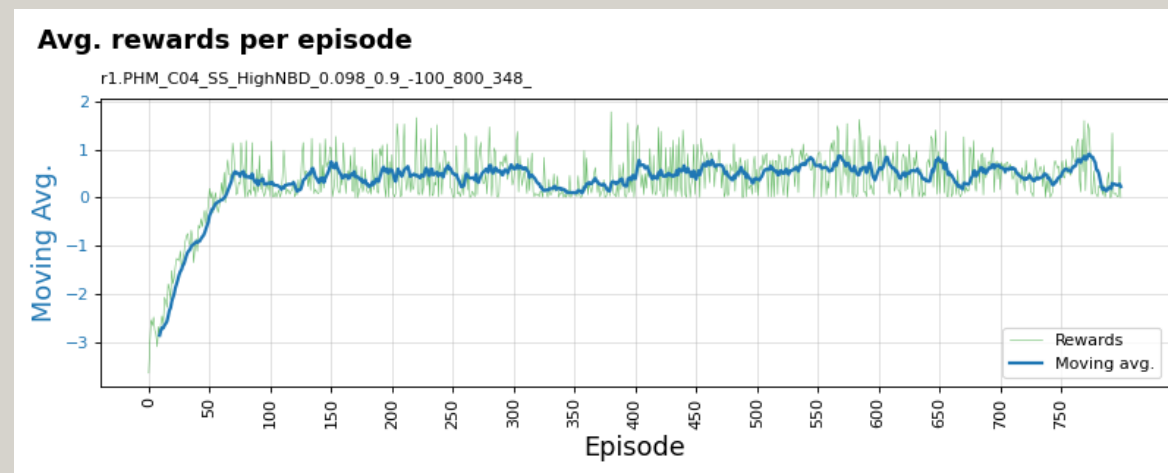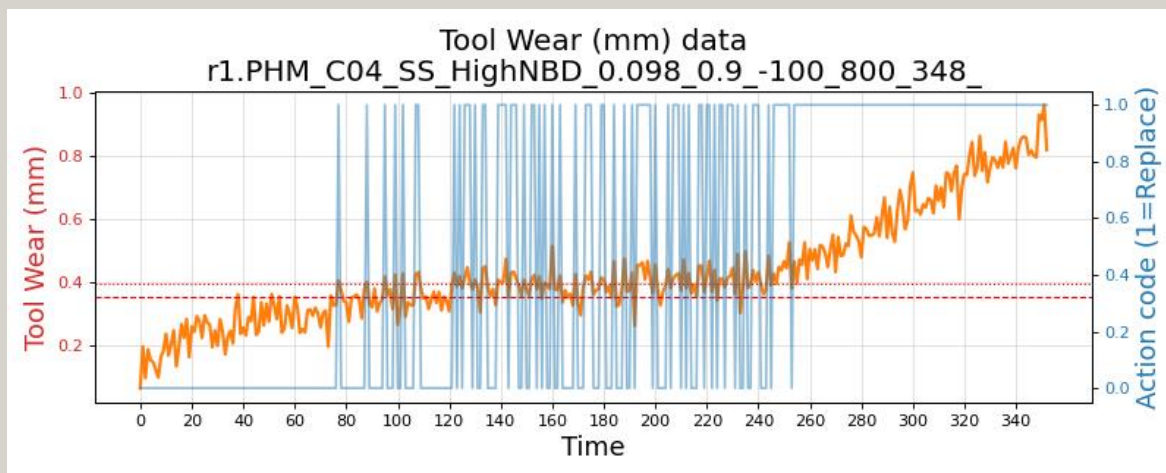PHM 2010 C-04 data-set: Variant: Single-state tool-wear with high noise (1e-2) and low break-down chance (10%)

Status of Planned Objectives and Schedule

Planned schedule and status: 12-Jul-2023

	Jan-23 Apr-23	May-23 Aug-23	Sep-23 Dec-23	Jan-24 Apr-24	May-24 Aug-24	Sep-24 Dec-24	Jan-25 Apr-25	May-25 Aug-25	Sep-25 Dec-25	Jan-26 Apr-26	May-26 Aug-26	Sep-26 Dec-26
Literature Research	✓											
Research on suitable public data sets	✓											
Research on suitable RL algorithms to use for PdM		✓										
Build PdM environment		✓										
Design RL elements (state, action and reward signal)		WIP										
Develop an experimental set up												
Evaluate on public data set												
Iterative phase: Experiment / fine-tune hyper-parameters												
Development of final model												
Stretch objective: Evaluate on primary data												
Modify model for adapting to primary data												
Thesis writing and submission												

OBJECTIVES:

1. Evaluate various RL algorithms suitable for solving predictive maintenance (PdM)

2. Design and develop a PdM application with above recommendation.

3. Extract features and explore techniques for standardizing the features across various applications.

4. To develop a data-driven model on real world data set

Current status:

1. SLR paper published in Q1 journal

2. Empirical study on 4 RL algorithms completed

Thank you