Experiments with REINFORCE algorithm

Rajesh Siraskar

14-Jun-2023

Agenda

1. Objective of this presentation
2. The RL environment
3. Evaluation strategy
4. Findings
5. *What could I possibly be doing wrong?*
6. Discussion

Objectives

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

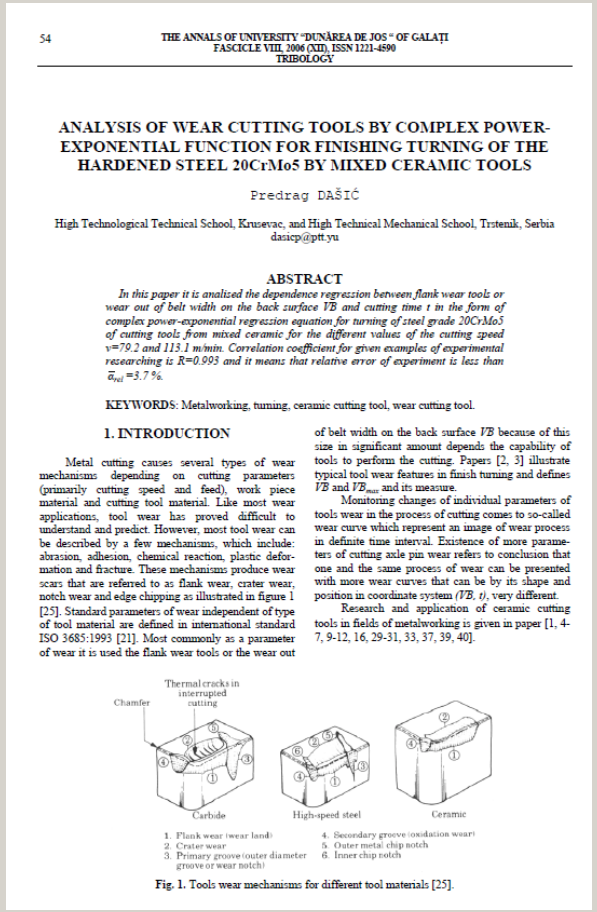
1. Experiment with the very *basic, naïve*, REINFORCE algorithm
2. Implemented from “scratch”
3. Compare against industry gold-standard Stable-Baselines-3 implementation of DQN, A2C and PPO
4. REINFORCE trained for 800 episodes, with an inner loop of 300. SB-3 trained with 10,000 episodes
5. This is a purely experimental (empirical) exercise.

* Note: My “proposal” goal is “Deep RL for predicting RUL”, this is a sub/related goal

RL environment

Two different sources – simulated and real milling data

$$VB = a \cdot t^{b_1} \cdot e^{b_2 \cdot t} \Big|_{t=0}^{t=t_k}$$



54 THE ANNALS OF UNIVERSITY "DUNĂREA DE JOS" OF GALAȚI
FASCICLE VIII, 2006 (XII), ISSN 1221-4590
TRIBOLOGY

ANALYSIS OF WEAR CUTTING TOOLS BY COMPLEX POWER-EXPONENTIAL FUNCTION FOR FINISHING TURNING OF THE HARDENED STEEL 20CrMo5 BY MIXED CERAMIC TOOLS

Predrag DAŠIĆ

High Technological Technical School, Kruševac, and High Technical Mechanical School, Trstenik, Serbia
dasicp@ptt.yu

ABSTRACT

In this paper it is analized the dependence regression between flank wear tools or wear out of belt width on the back surface VB and cutting time t in the form of complex power-exponential regression equation for turning of steel grade 20CrMo5 of cutting tools from mixed ceramic for the different values of the cutting speed v=79.2 and 113.1 m/min. Correlation coefficient for given examples of experimental researching is R=0.993 and it means that relative error of experiment is less than $\bar{a}_{rel}$ *=3.7 %.*

KEYWORDS: Metalworking, turning, ceramic cutting tool, wear cutting tool.

1. INTRODUCTION

Metal cutting causes several types of wear mechanisms depending on cutting parameters (primarily cutting speed and feed), work piece material and cutting tool material. Like most wear applications, tool wear has proved difficult to understand and predict. However, most tool wear can be described by a few mechanisms, which include: abrasion, adhesion, chemical reaction, plastic deformation and fracture. These mechanisms produce wear scars that are referred to as flank wear, crater wear, notch wear and edge chipping as illustrated in figure 1 [25]. Standard parameters of wear independent of type of tool material are defined in international standard ISO 3685:1993 [21]. Most commonly as a parameter of wear it is used the flank wear tools or the wear out of belt width on the back surface *VB* because of this size in significant amount depends the capability of tools to perform the cutting. Papers [2, 3] illustrate typical tool wear features in finish turning and defines *VB* and VB_{max} and its measure.

Monitoring changes of individual parameters of tools wear in the process of cutting comes to so-called wear curve which represent an image of wear process in definite time interval. Existence of more parameters of cutting axle pin wear refers to conclusion that one and the same process of wear can be presented with more wear curves that can be by its shape and position in coordinate system (*VB, t*), very different.

Research and application of ceramic cutting tools in fields of metalworking is given in paper [1, 4-7, 9-12, 16, 29-31, 33, 37, 39, 40].

Fig. 1. Tools wear mechanisms for different tool materials [25].

◄ 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

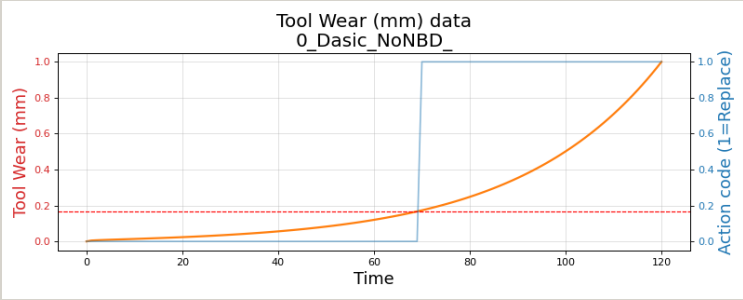
RL environment - Variants

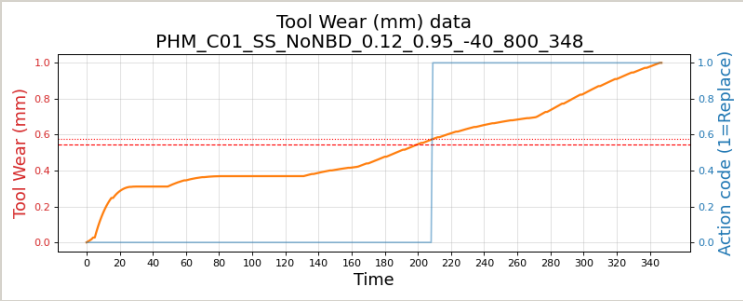
Three environments and their variants:

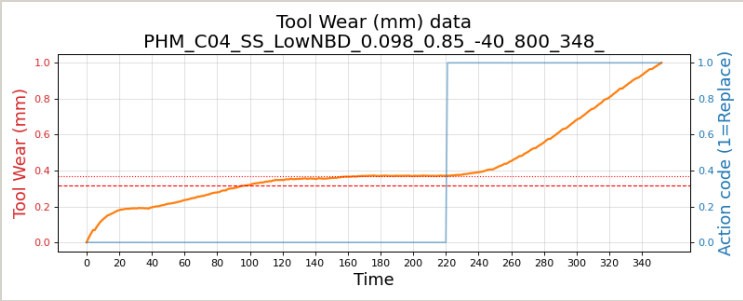
1. Simulated. Based on Dašić (2006). Simple single-variate state (tool wear)
    - Variants: (1) No noise (2) Low noise and low break-down chance and (3) High noise and high break-down chance

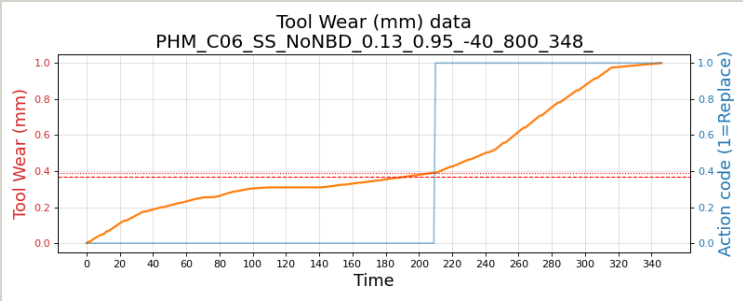
2. PHM 2010 real data – Simple single-variate state (tool wear)
    - Variants: C-01, C-04 and C-06 data-sets
    - Variants: (1) No noise (2) Low noise and low break-down chance and (3) High noise and high break-down chance

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

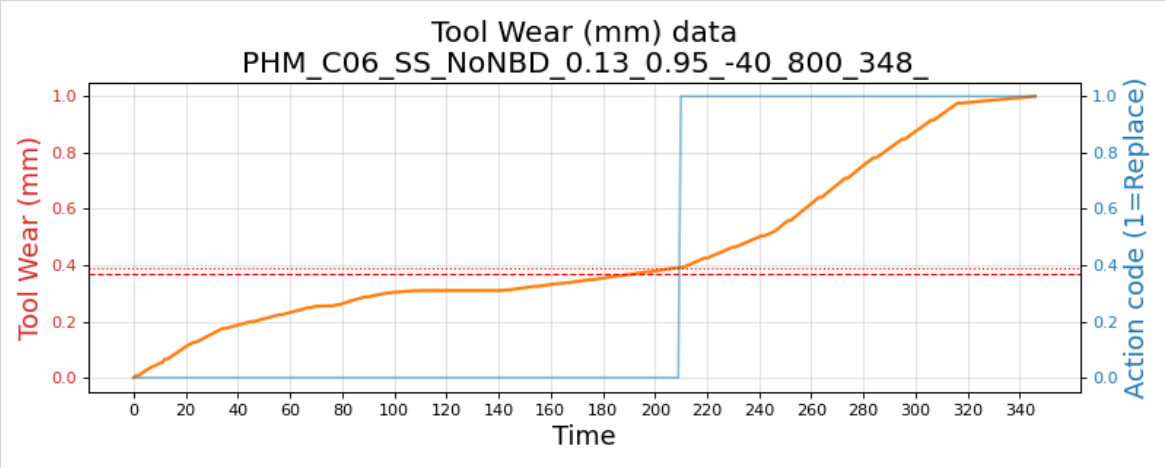


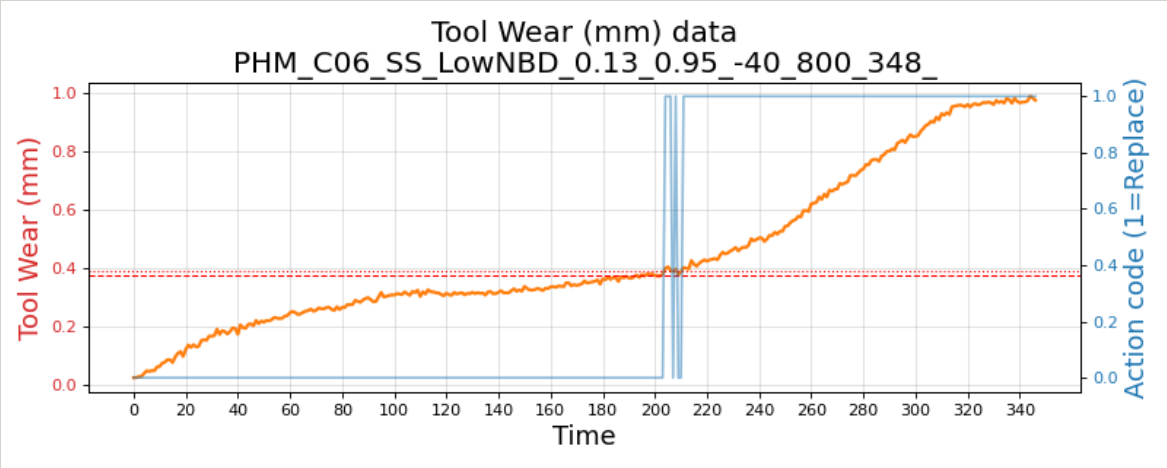


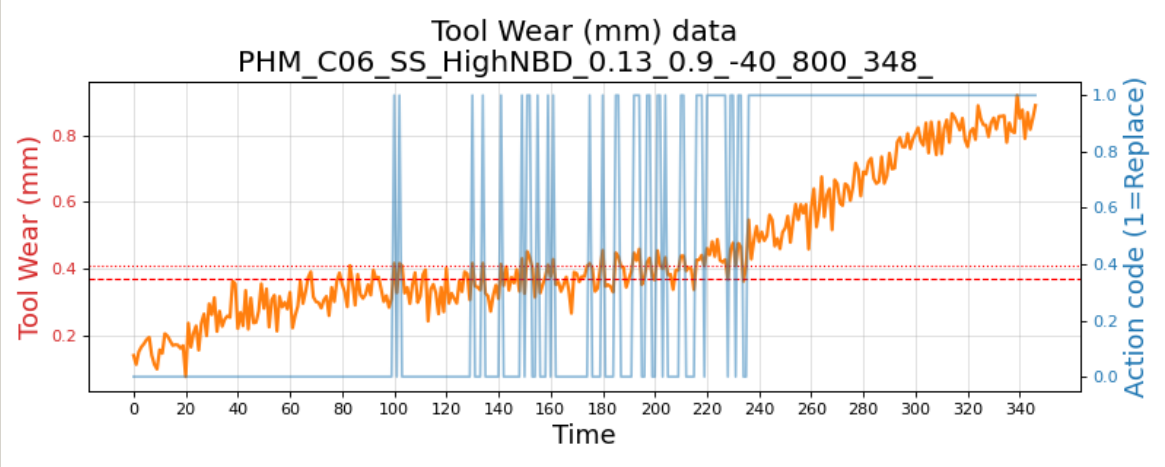




Wear plot - real data and its variants







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” replacement action
- Test from randomly sampled 40 points, 5 rounds
- Compute Precision, Recall and F1
- Compare mean and std. deviations

Findings

1. The naïve implementation of REINFORCE algorithm was implemented in PyTorch with an extremely simple architecture: Three layers and ReLU activation and an Adam optimizer.
2. In training, the SB-3 algorithms were extremely stable, but the F1 performance was always between 0.38-0.67. The naïve REINFORCE, on the other-hand, is unstable in training.
3. We saved models that performed better than 0.7 and these cases, the REINFORCE performs surprisingly better than DQN, A2C and PPO in all cases
4. The SB-3 algorithms almost always perform at about F1 0.47-0.50
5. REINFORCE performed between 0.72 (for complex real data environment) to 0.89 (for simulated environment)
6. Across precision, recall and F1; REINFORCE was better than the *best* performing SB-3 algorithm by 0.252 basis-points. The difference[1] in variance (σ^2) was also *lower*, though near negligible, for REINFORCE at -0.0003
7. Across precision on tool replacement, REINFORCE was better by 0.354, a lower variance of -0.004

1. w.r.t. lowest of SB-3 algorithms
2. Reference data-sheet: "1. Analysis_model_performance_summary_V3 (PPT ref.).xls"

Results: Overall – all environments and their variants

	Precision		Recall		F1-score	
Algorithm	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
A2C	0.431	0.090	0.397	0.065	0.383	0.059
DQN	0.475	0.175	0.657	0.030	0.471	0.029
PPO	0.504	0.121	0.391	0.066	0.402	0.070
REINFORCE	**0.861**	0.039	**0.843**	0.051	**0.836**	0.041

Results: Simple single variate state. Including noise variants

Simulated – Dašić, Predrag (2006).

	Precision		Recall		F1-score	
Algorithm	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
A2C	0.292	0.041	0.083	0.024	0.128	0.030
DQN	0.517	0.012	0.963	0.032	0.673	0.015
PPO	0.478	0.244	0.093	0.048	0.150	0.074
REINFORCE	**0.928**	0.032	**0.883**	0.041	**0.895**	0.039

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

	Precision		Recall		F1-score	
Algorithm	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
A2C	0.459	0.115	0.422	0.077	0.417	0.067
DQN	0.486	0.236	0.559	0.029	0.413	0.035
PPO	0.502	0.094	0.479	0.071	0.467	0.073
REINFORCE	**0.845**	0.040	**0.892**	0.051	**0.855**	0.038

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	
Algorithm	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
A2C	0.485	0.067	0.637	0.072	0.534	0.065
DQN	0.398	0.156	0.647	0.031	0.444	0.025
PPO	0.537	0.080	0.423	0.068	0.458	0.055
REINFORCE	**0.842**	0.040	**0.657**	0.062	**0.721**	0.052

Results Summary (13-Jun-2023 run)

	REINFORCE			SB-3 A2C			SB-3 DQN			SB-3 PPO		
Model	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Simulated Dasic 2006 - No noise	1.00	0.74	0.85	0.51	0.62	0.56	0.30	0.03	0.05	0.57	0.43	0.49
Simulated Dasic 2006 - Low NBD	0.93	0.99	0.96	0.00	0.00	0.00	0.51	0.99	0.67	0.72	0.46	0.56
Simulated Dasic 2006 - High NBD	0.90	0.99	0.94	0.70	0.82	0.76	0.30	0.03	0.05	0.59	0.31	0.40
PHM C01 simple - No noise	0.87	0.98	0.92	0.47	0.26	0.33	0.40	0.03	0.06	0.27	0.13	0.17
PHM C01 simple - Low NBD	0.91	0.85	0.88	0.00	0.00	0.00	0.40	0.03	0.06	0.50	0.17	0.25
PHM C01 simple - High NBD	0.80	0.99	0.88	0.50	1.00	0.67	0.52	0.99	0.68	0.52	0.53	0.52
PHM C04 simple - No noise	0.90	0.93	0.91	0.53	0.62	0.57	0.05	0.01	0.02	0.37	0.16	0.22
PHM C04 simple - Low NBD	0.67	0.99	0.80	0.59	0.68	0.63	0.10	0.01	0.02	0.27	0.02	0.04
PHM C04 simple - High NBD	0.69	0.82	0.75	0.00	0.00	0.00	0.51	1.00	0.68	0.46	0.26	0.33
PHM C06 simple - No noise	1.00	0.61	0.76	0.50	1.00	0.67	0.51	0.99	0.67	0.30	0.05	0.08
PHM C06 simple - Low NBD	0.99	0.85	0.91	0.42	0.38	0.40	0.50	0.99	0.67	0.38	0.23	0.28
PHM C06 simple - High NBD	0.78	0.90	0.83	0.40	0.29	0.33	0.50	0.96	0.66	0.27	0.05	0.08
PHM C01 multi-variate state	0.83	0.94	0.88	0.49	0.64	0.55	0.50	0.98	0.66	0.58	0.37	0.44
PHM C04 multi-variate state	0.78	0.68	0.72	0.48	0.64	0.55	0.60	0.03	0.06	0.61	0.29	0.39
PHM C06 multi-variate state	1.00	0.58	0.73	0.50	1.00	0.67	0.50	1.00	0.67	0.42	0.38	0.40

Training time: Avg. over different variants

Average training time in secs.

Variant	A2C	DQN	PPO	REINFORCE
Dasic 2006 Simulated - Single variate state	33.53	3.04	34.23	182.76
PHM-2010 Real data - Single variate state	21.53	2.37	24.27	313.99
PHM-2010 Real data - Multi-variate state	37.23	6.06	42.46	632.05
Overall average	**27.07**	**3.24**	**29.90**	**351.36**

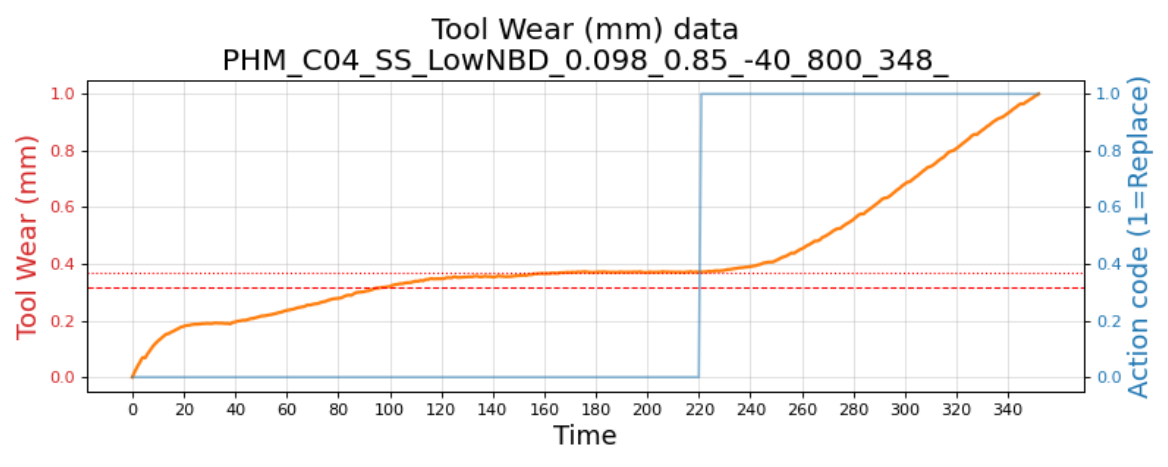
Average training time per 1,000 time-steps, in secs. Stable-baseline algorithms are trained for 10,000 time-steps, while REINFORCE uses 96,800 for simulated and 278,400 for PHM based environments

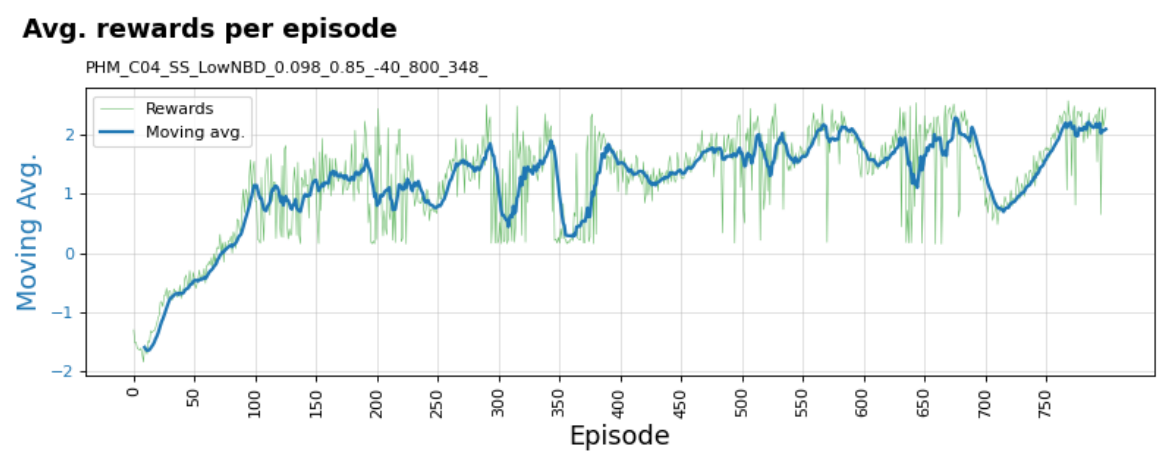
Variant	A2C	DQN	PPO	REINFORCE
Dasic 2006 Simulated - Single variate state	3.35	0.30	3.42	1.89
PHM-2010 Real data - Single variate state	2.15	0.24	2.43	1.13
PHM-2010 Real data - Multi-variate state	3.72	0.61	4.25	2.27
Overall average	**9.23**	**1.15**	**10.10**	**1.73**

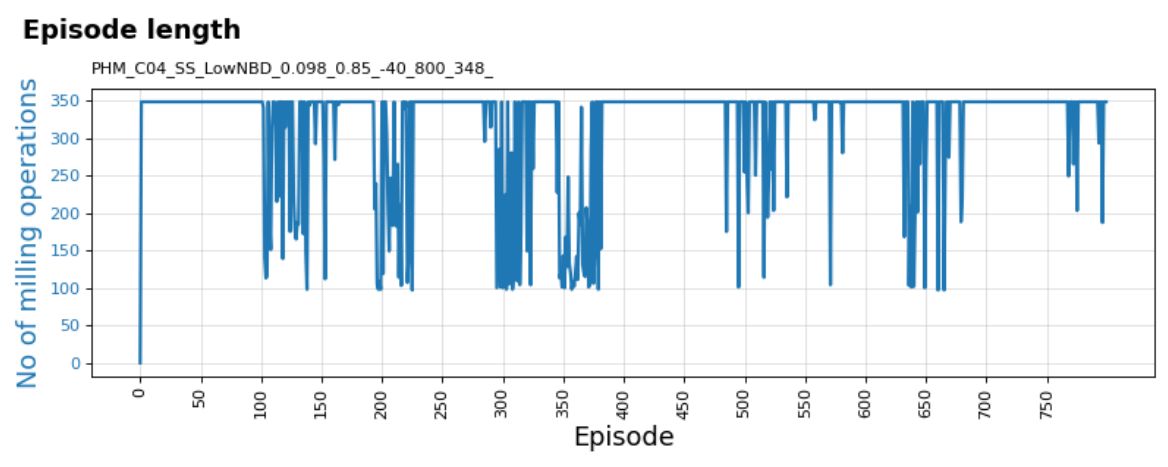
Some plots

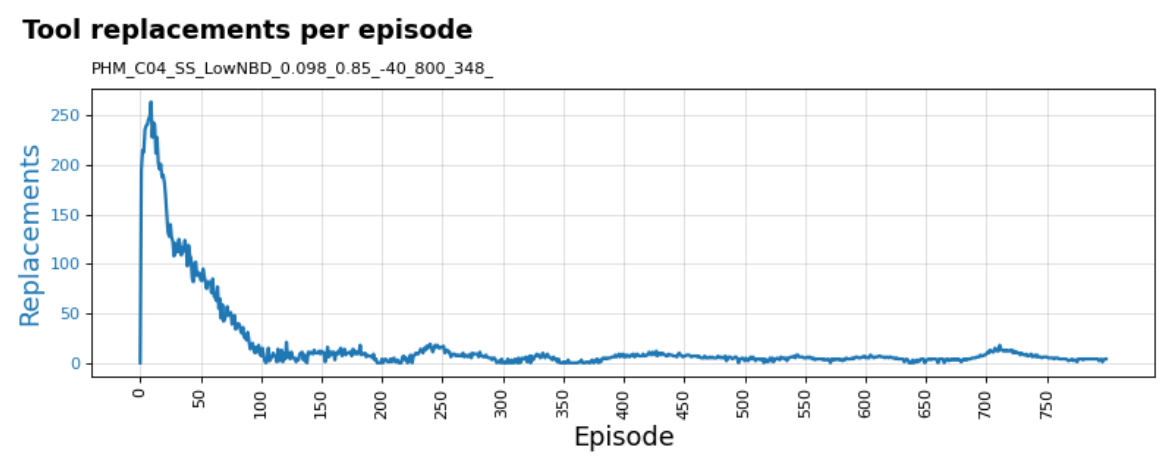
Training plots for REINFORCE algorithm

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



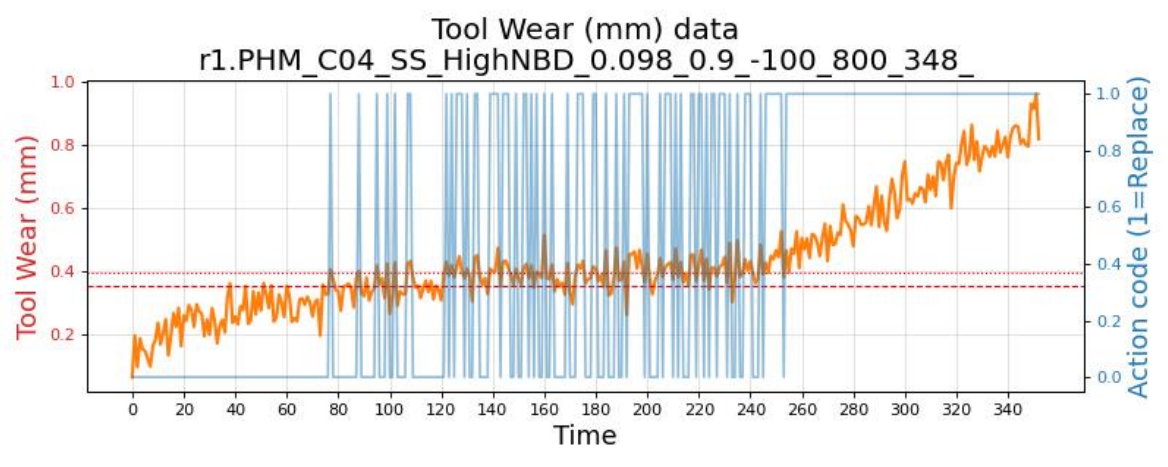


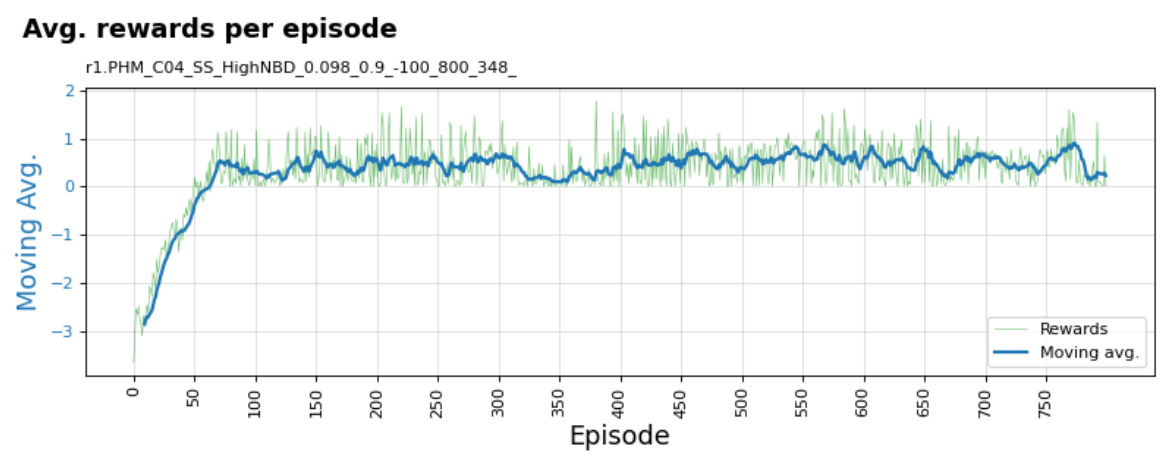


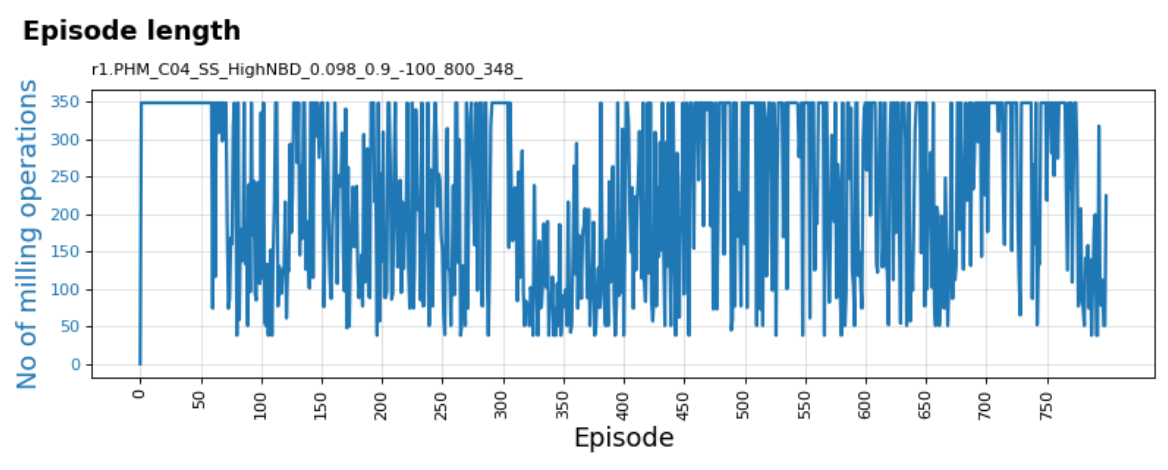


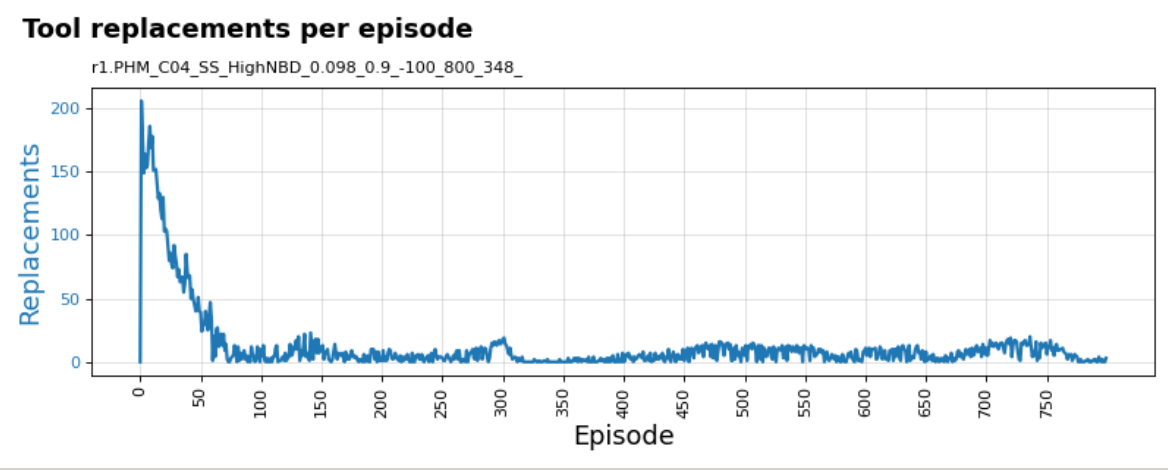
Training plots for REINFORCE algorithm

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

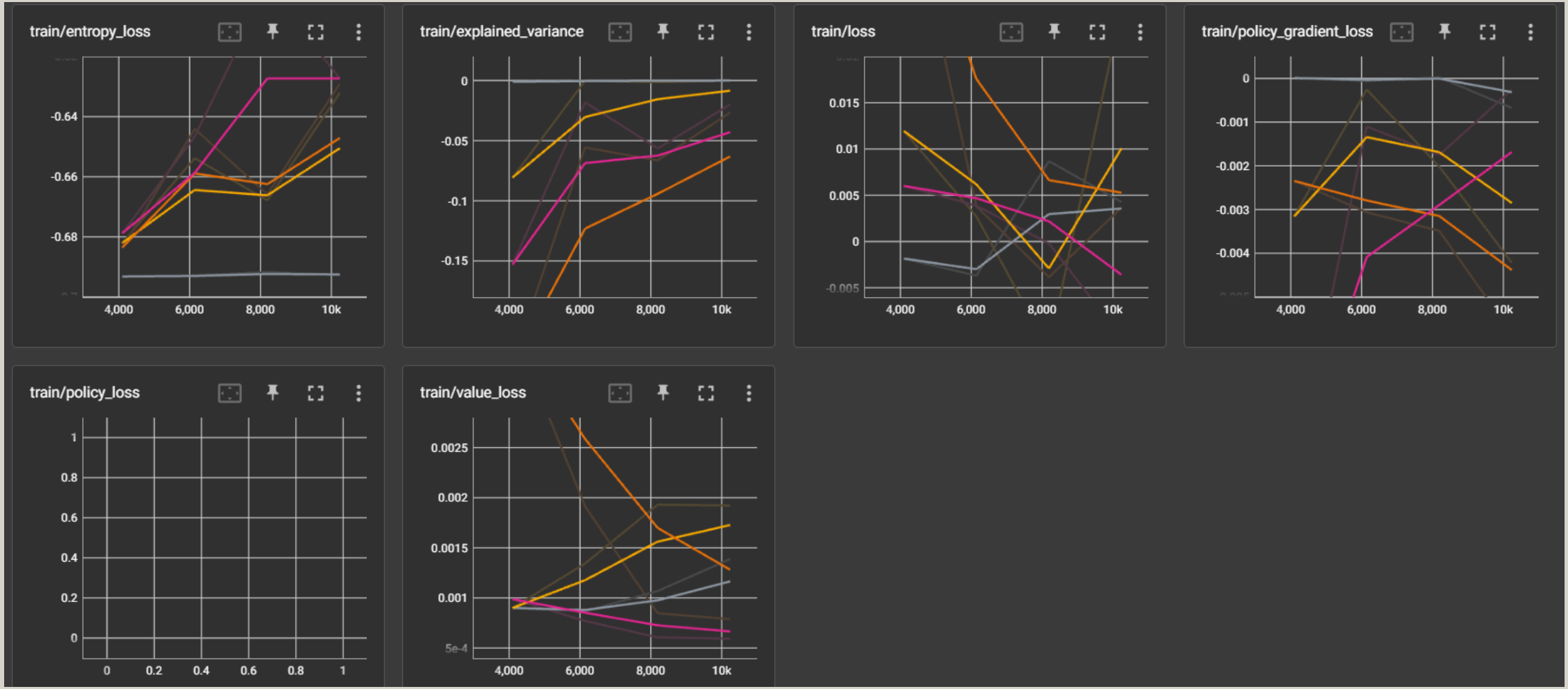








Tensorboard plots for SB-3 algorithms - PPO



SB-3 stability doubts

SB-3 stability issue?

1. Earlier runs with same episodes = 800
2. But REINFORCE has an internal loop of 348
3. Increased SB-3 algo runs to 10,000, then 800 x 348 then 20,000
4. Re-installed SB-3 and upgraded the version
5. Tested on cart-pole and mountain-car environment
6. Re-trained SB-3 models and re-ran experiments for 10,000 and 20,000 episodes– see next slide

Results: SB-3 (10 K episodes)

	REINFORCE			SB-3 A2C			SB-3 DQN			SB-3 PPO		
Model	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Simulated Dasic 2006 - No noise	1.00	0.70	0.82	0.49	0.44	0.46	0.07	0.01	0.02	0.46	0.28	0.34
Simulated Dasic 2006 - Low NBD	0.94	0.90	0.92	0.00	0.00	0.00	0.45	0.03	0.05	0.37	0.06	0.10
Simulated Dasic 2006 - High NBD	0.90	1.00	0.94	0.50	0.53	0.51	0.50	0.96	0.66	0.58	0.15	0.24
PHM C01 simple - No noise	0.91	0.96	0.93	0.52	0.56	0.54	0.37	0.03	0.05	0.50	0.25	0.33
PHM C01 simple - Low NBD	0.89	0.80	0.84	0.38	0.13	0.19	0.40	0.02	0.04	0.46	0.14	0.21
PHM C01 simple - High NBD	0.78	0.93	0.85	0.00	0.00	0.00	0.51	0.96	0.66	0.52	0.21	0.29
PHM C04 simple - No noise	0.82	0.96	0.88	0.51	0.09	0.15	0.50	0.97	0.66	0.49	0.47	0.48
PHM C04 simple - Low NBD	0.74	0.99	0.85	0.47	0.51	0.49	0.71	0.72	0.71	0.53	0.28	0.37
PHM C04 simple - High NBD	0.67	0.78	0.72	0.52	0.55	0.53	0.50	0.98	0.66	0.47	0.25	0.32
PHM C06 simple - No noise	1.00	0.65	0.78	0.47	0.46	0.46	0.51	0.98	0.67	0.38	0.14	0.20
PHM C06 simple - Low NBD	0.98	0.84	0.90	0.49	0.53	0.51	0.95	0.58	0.72	0.50	0.38	0.43
PHM C06 simple - High NBD	0.72	0.88	0.79	0.50	0.47	0.48	0.51	0.98	0.67	0.37	0.11	0.17
PHM C01 multi-variate state	0.80	0.92	0.85	0.51	0.59	0.55	0.58	0.97	0.73	0.49	0.24	0.31
PHM C04 multi-variate state	0.77	0.69	0.73	0.48	0.53	0.50	0.51	0.97	0.66	0.49	0.34	0.40
PHM C06 multi-variate state	1.00	0.57	0.73	0.49	0.60	0.54	0.49	0.96	0.65	0.45	0.48	0.46

SB-3 algorithms do perform well sometimes *(green)*. On an average their performance is poor *(red-orange-yellow)*.

Results: SB-3 (20 K episodes)

	REINFORCE			SB-3 A2C			SB-3 DQN			SB-3 PPO		
Model	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Simulated Dasic 2006 - No noise	1.00	0.71	0.83	0.00	0.00	0.00	0.50	0.96	0.66	0.07	0.01	0.02
Simulated Dasic 2006 - Low NBD	0.94	0.94	0.94	0.54	0.57	0.55	0.88	0.11	0.19	0.00	0.00	0.00
Simulated Dasic 2006 - High NBD	0.88	0.98	0.93	0.44	0.22	0.29	0.07	0.04	0.05	0.33	0.04	0.07
PHM C01 simple - No noise	0.90	0.97	0.93	0.00	0.00	0.00	0.09	0.03	0.04	0.53	0.73	0.61
PHM C01 simple - Low NBD	0.95	0.92	0.93	0.50	0.45	0.47	0.98	0.47	0.63	0.40	0.02	0.04
PHM C01 simple - High NBD	0.83	0.92	0.87	0.51	0.44	0.47	0.88	0.24	0.37	0.20	0.01	0.02
PHM C04 simple - No noise	0.85	0.98	0.91	0.50	1.00	0.67	0.80	0.05	0.09	0.51	0.25	0.33
PHM C04 simple - Low NBD	0.71	1.00	0.83	0.52	0.60	0.55	0.50	0.97	0.66	0.10	0.01	0.02
PHM C04 simple - High NBD	0.73	0.90	0.81	0.50	1.00	0.67	0.40	0.02	0.04	0.41	0.11	0.17
PHM C06 simple - No noise	1.00	0.58	0.73	0.59	0.53	0.56	0.20	0.01	0.02	0.56	0.66	0.60
PHM C06 simple - Low NBD	0.94	0.92	0.93	0.00	0.00	0.00	0.33	0.03	0.05	0.20	0.01	0.02
PHM C06 simple - High NBD	0.69	0.88	0.78	0.00	0.00	0.00	0.40	0.03	0.06	0.37	0.20	0.26
PHM C01 multi-variate state	0.86	0.90	0.88	0.31	0.08	0.13	0.36	0.54	0.43	0.58	0.12	0.19
PHM C04 multi-variate state	0.72	0.61	0.66	0.49	0.94	0.64	0.50	0.98	0.66	0.44	0.15	0.22
PHM C06 multi-variate state	1.00	0.58	0.73	0.50	1.00	0.67	0.51	0.99	0.67	0.53	0.36	0.43

SB-3 algorithms do perform well sometimes *(green)*. On an average their performance is poor *(red-orange-yellow)*.

Thank you

Results: SB-3 (10 K episodes)

	REINFORCE			Advanced Algorithms (SB-3 implementations)									Key	
				A2C			DQN			PPO				
Environment	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score		
Simulated Dasic 2006 - No noise	1.000	0.690	0.815	0.800	0.160	0.259	0.505	0.980	0.667	0.551	0.170	0.258		> 0.70
Simulated Dasic 2006 - Low noise/break-down	0.927	0.970	0.947	0.554	0.560	0.555	0.200	0.010	0.019	0.067	0.010	0.017		< 0.05
Simulated Dasic 2006 - High noise/break-down	0.926	0.990	0.957	0.503	0.990	0.667	0.433	0.040	0.072	0.606	0.320	0.418		
PHM C01 simple - No noise	0.823	0.970	0.890	0.378	0.350	0.363	0.200	0.010	0.019	0.400	0.110	0.170		
PHM C01 simple - Low noise/break-down	0.945	0.860	0.899	0.466	0.510	0.484	0.821	0.960	0.885	0.320	0.110	0.159		
PHM C01 simple - High noise/break-down	0.792	0.950	0.863	0.529	0.500	0.512	0.600	0.030	0.057	0.484	0.190	0.270		
PHM C04 simple - No noise	0.874	0.970	0.919	0.528	0.800	0.635	0.100	0.010	0.018	0.494	0.720	0.586		
PHM C04 simple - Low noise/break-down	0.853	0.980	0.909	0.000	0.000	0.000	0.513	1.000	0.678	0.537	0.320	0.393		
PHM C04 simple - High noise/break-down	0.735	0.770	0.751	0.347	0.040	0.068	0.292	0.360	0.322	0.521	0.880	0.655		
PHM C06 simple - No noise	1.000	0.690	0.816	0.527	0.540	0.530	0.505	0.990	0.669	0.537	0.470	0.499		
PHM C06 simple - Low noise/break-down	0.928	0.790	0.853	0.000	0.000	0.000	0.508	1.000	0.674	0.534	0.420	0.466		
PHM C06 simple - High noise/break-down	0.763	0.900	0.824	0.513	0.430	0.467	0.374	0.580	0.454	0.200	0.030	0.052		
PHM C01 multi-variate state - No noise	0.893	0.910	0.898	0.505	0.460	0.478	0.100	0.010	0.018	0.481	0.160	0.237		
PHM C04 multi-variate state - No noise	0.824	0.830	0.825	0.503	0.860	0.635	0.503	0.970	0.662	0.494	0.600	0.541		
PHM C06 multi-variate state - No noise	1.000	0.490	0.657	0.463	0.670	0.547	0.508	0.990	0.671	0.408	0.130	0.197		
Maximum values	**1.000**	**0.990**	**0.957**	0.800	0.990	0.667	0.821	1.000	0.885	0.606	0.880	0.655		
Average values	**0.886**	**0.851**	**0.855**	0.441	0.458	0.413	0.411	0.529	0.392	0.442	0.309	0.328		