Table of content

Chapter 1: Introduction 1
1.1. On our effect on the planet 1
1.2. Sustainability 2
1.3. Solar water heating 3
1.4. Policy optimization 4
1.5. Reinforcement Learning and Deep Reinforcement Learning 4
1.6. Objectives 5
1.6.1. General objective 6
1.6.2. Specific objectives 6
1.7. Structure of the Thesis 6
Chapter 2: Literature Review 8
2.1. Literature on solar water heating 8
2.2. Literature on policy optimization 8
2.3. Summary 10
Chapter 3: Theoretical framework 11
3.1. Dense Deep Neural Networks 11
3.1.1. Gradient Descent 12
3.1.2. Momentum 13
3.1.3. Backpropagation 14
3.1.4. Softmax activation 17
3.1.5. DNN initialization 19
3.2. Deep Reinforcement Learning Algorithms 19
3.2.1. Basic Concepts 20
3.2.2. Policy gradient methods and value methods 21
3.2.3. REINFORCE algorithm 22
3.2.4. Actor-Critic Algorithm 22
3.2.5. Q-Learning and Deep Q-Learning 24
3.2.5.1. Q-Values 24
3.2.5.2. Q-Learning 24
3.2.5.3. Deep Q-Learning algorithm 25
3.2.5.4. Double DQN 27
3.2.5.5. Prioritized Experience Replay 27

3.3. Reliability theory 28
3.3.1. Hazard rate 30
3.3.2. Exponential distribution 31
3.3.3. Series system 31
3.3.4. Parallel system 32
3.4. Discrete-time Markov chains 32
3.4.1. Geometric distribution 33
3.4.2. Steady-state probabilities 34
3.5. Heat pipe evacuated tube solar collectors 37
3.6. Heat pumps and refrigeration systems 38
3.7. TRNSYS 40
Chapter 4: Development of the training platform 42
4.1. System under study 42
4.2. Actions 44
4.3. Rewards 46
4.4. Environment state 48
4.5. TRNSYS-Python connection 49
4.6. Introducing stochastic failures 53
4.6.1. Failure rates of individual devices 54
4.6.2. Construction of the Markov chains 56
4.7. Pseudo-code versions of the Python Scripts 60
4.7.1. Initializer 60
4.7.2. Code to train the network 61
4.7.2.1. Main code 61
4.7.2.2. Interaction at 8.00 AM 62
4.7.2.3. Interaction from 10.00 AM to 8.00 PM 62
4.7.2.4. Interaction at 10.00 PM 63
Chapter 5: Methodology 64
5.1. Stages of the study 64
5.2. Result Analysis 64
Chapter 6: Results and Discussion 68
6.1. Comparison of DRL algorithms 68
6.2. Comparison to a non-smart-controlled baseline 74
6.2.1. Testing method 75

6.2.2. Results 76
6.3. Comparison of different network architectures and training hyperparameters 77
6.3.1. Comparing Different Architectures 79
6.3.2. Comparing Different Discount Factors 81
6.3.3. Comparing traditional DQN to Double DQN 84
6.3.4. Effect of momentum 85
6.4. Behavior comparison under different reward parameters 86
6.4.1. Changing the value of α1 87
6.4.2. Changing the value of α2 94
6.4.3. Changing the value of α3 97
6.4.4. Effect of the α4 parameter 98
6.5. System subject to failures 99
6.5.1. Training is carried out with the Markov chains of the real system 101
6.5.2. Training is carried out with planned failure cycles 108
6.5.3. Alternative Markov chains 114
6.5.4. Effect of momentum for failure-subject agents 117
6.5.5. Agent selection and final testing 119
Chapter 7: Conclusions 126
7.1. Accomplishment of objectives 126
7.2. Future work 127
8. Glossary 129
9. Bibliography 130
Annexes 134
Annexed A. Further details about the water heating system simulation 134
Annexed A.1. Detailed flow diagrams 134
Annexed A.2. Parameters of the elements in the system simulation (Types) 135
Annexed A.3. Elements (Types) with external files 138
Annexed A.4. Imposed Temperatures 142
Annexed B. All results of Section 6.5. 143