



## AutoMated Vessels and Supply Chain Optimisation for Sustainable Short SEa Shipping

### D.4.3: Results of autonomous tugboat operation simulation

| Document Identification |                              |                     |   |
|-------------------------|------------------------------|---------------------|---|
| Status                  | Final                        | Due Date            | Thursday, June 30, 2022                           |
| Version                 | 1.0                          | Submission Date     | 08/11/2022  |
| Related WP              | WP4                          | Document Reference  | D4.3  |
| Related Deliverable(s)  | D2.2, D2.4, D4.1, D4.2       | Dissemination Level | CO  |
| Lead Participant        | CORE                         | Document Type:      | R   |
| Contributors            | NTUA<br>ESI<br>DANAOS<br>DNV | Lead Author         | Manthos Kampourakis (CORE)                        |
|                         |                              | Reviewers           | Gerco Hagesteijn (MARIN)<br>Iason Vlavianos (SAT) |



| <b>List of Contributors</b> |                  |                |
|-----------------------------|------------------|----------------|
| <b>First Name</b>           | <b>Last Name</b> | <b>Partner</b> |
| Manthos                     | Kampourakis      | CORE           |
| Elias                       | Mantouvalos      | CORE           |
| Nikos                       | Monios           | CORE           |
| Konstantinos                | Louzis           | NTUA           |
| Nikolaos                    | Ventikos         | NTUA           |
| Nikolaos                    | Themelis         | NTUA           |
| Haris                       | Oikonomidou      | NTUA           |
| Alexandros                  | Koimtzooglou     | NTUA           |
| Elias                       | Kotsidis         | ESI            |
| Artemis                     | Flori            | DANAOS         |
| Chara                       | Georgopoulou     | DNV            |
| Giannis                     | Kanellopoulos    | ICCS           |

| <b>Document History</b> |             |  |   |
|-------------------------|-------------|--|---|
| <b>Version</b>          | <b>Date</b> | <b>Change editors</b>  | <b>Changes</b>                              |
| 0.1                     | 03/03/2022  | M. Kampourakis   | Table of Contents (draft)                   |
| 0.2                     | 10/03/2022  | M. Kampourakis   | Sections 2 & 3 additions                    |
| 0.3                     | 29/04/2022  | M. Kampourakis   | Section 4, ToC updates                      |
| 0.4                     | 16/05/2022  | M. Kampourakis   | Summary, introduction & content additions   |
| 0.5                     | 17/05/2022  | M. Kampourakis,<br>H. Oikonomidou,<br>K. Louzis,<br>E. Kotsidis,<br>A. Flori                         | Baseline Scenarios                          |
| 0.6                     | 29/08/2022  | M. Kampourakis   | Training results update                     |
| 0.7                     | 21/09/2022  | K. Louzis,<br>N. Ventikos,<br>N. Themelis,<br>H. Oikonomidou,<br>A. Koimtzooglou,<br>C. Georgopoulou | Autonomous tugboats fail-safe functionality |
| 0.8                     | 17/10/2022  | G. Kanellopoulos   | Predictive Battery Optimisation             |

| Document History |            |   |                               |
|------------------|------------|---|-------------------------------|
| Version          | Date       | Change editors  | Changes                       |
| 0.9              | 27/10/2022 | M. Kampourakis  | MARIN, SAT review corrections |
| 1.0              | 08/11/2022 | M. Kampourakis,<br>K. Louzis,<br>N. Themelis,<br>H. Oikonomidou | Final version to be submitted |

|  |           |
|--|-----------|
| <b>Executive Summary.....</b>                                      | <b>10</b> |
| <b>1. Introduction .....</b>                                       | <b>11</b> |
| 1.1 Purpose of the document .....                                  | 11        |
| 1.2 Intended readership.....                                       | 11        |
| 1.3 Document Structure.....  | 11        |
| <b>2. Reinforcement Learning Main Concepts .....</b>               | <b>12</b> |
| 2.1 Differences from classical Machine Learning applications ..... | 13        |
| 2.2 Examples .....   | 14        |
| 2.2.1 Pick-and-Place Robot.....                                    | 14        |
| 2.2.2 Pole-Balancing .....   | 14        |
| 2.2.3 Remote control.....  | 15        |
| 2.3 Applications.....  | 15        |
| 2.3.1 Traffic Light Control .....                                  | 15        |
| 2.3.2 Industry Automation .....                                    | 16        |
| 2.3.3 Chemistry.....   | 17        |
| 2.3.4 Personalised Recommendations .....                           | 17        |
| 2.3.5 Games.....   | 17        |
| 2.3.6 Shipping Industry .....                                      | 18        |
| 2.3.6.1 Empty Container Repositioning.....                         | 18        |
| 2.3.6.2 Cargo Management.....                                      | 18        |
| 2.3.6.3 Ship Navigation.....                                       | 19        |
| 2.4 Model-free vs Model-based RL.....                              | 19        |
| 2.5 Policy .....   | 21        |
| 2.5.1 Policy approximation .....                                   | 21        |
| 2.5.2 Value approximation .....                                    | 22        |
| 2.5.3 States & Actions.....  | 22        |
| 2.5.4 Reward Signals.....  | 23        |
| 2.6 Popular algorithms.....  | 25        |

|           |   |           |
|-----------|---|-----------|
| 2.6.1     | Deep Q-Network.....   | 25        |
| 2.6.2     | PPO .....   | 27        |
| 2.6.3     | SAC.....  | 28        |
| 2.7       | General RL limitations .....  | 30        |
| 2.8       | Traditional VS DL/RL based control methods .....                                      | 32        |
| <b>3.</b> | <b>ML-Agents Toolkit .....</b>  | <b>37</b> |
| 3.1       | Agent Design Functionalities.....   | 38        |
| 3.1.1     | Curiosity-Driven Learning .....   | 38        |
| 3.1.2     | Imitation Learning .....  | 40        |
| 3.1.2.1   | Generative Adversarial Imitation Learning (GAIL).....                                 | 41        |
| 3.1.2.2   | Behavioral Cloning (BC).....  | 42        |
| 3.1.2.3   | Recording Demonstrations.....   | 42        |
| 3.1.3     | Curriculum Learning .....   | 42        |
| 3.1.4     | Parameter Randomisation.....  | 43        |
| 3.2       | Multi-Agent Scenarios.....  | 44        |
| 3.2.1     | Cooperative .....   | 44        |
| 3.2.2     | Competitive .....   | 44        |
| <b>4.</b> | <b>MOSES autonomous tugboat training .....</b>  | <b>46</b> |
| 4.1       | Model/Behavior Parameterization .....   | 46        |
| 4.1.1     | Request decisions from the model.....   | 46        |
| 4.1.2     | Behavior script.....  | 47        |
| 4.2       | LiDAR implementation .....  | 48        |
| 4.3       | Tugboat Brain.....  | 48        |
| 4.3.1     | Observations.....   | 50        |
| 4.3.2     | Actions .....   | 51        |
| 4.3.3     | Reward Formulation.....   | 52        |
| 4.3.3.1   | Push tug rewards.....   | 53        |
| 4.3.3.1   | Pull tug rewards.....   | 57        |
| 4.3.3.2   | Reward summary .....  | 59        |
| 4.4       | Virtual baseline scenarios .....  | 60        |
| 4.4.1     | Baseline Scenario A .....   | 60        |
| 4.4.1.1   | MOSES Phase 1 (Position B: Bring the vessel to a position parallel to the dock)<br>61 |           |
| 4.4.1.2   | MOSES Phase 2 (Parallel transfer to position C) .....                                 | 62        |
| 4.4.2     | Baseline Scenario B.....  | 64        |
| 4.4.2.1   | MOSES Phase 1 (Position B: Bring the vessel to a position parallel to the dock)<br>64 |           |
| 4.4.2.2   | MOSES Phase 2 (Parallel transfer to position C) .....                                 | 66        |

|           |  |            |
|-----------|--|------------|
| 4.5       | Autonomous Agent Performance Evaluation .....                | 67         |
| 4.5.1     | Training performance of baseline scenario A re-creation..... | 67         |
| 4.5.1.1   | Using push tugboats only .....                               | 68         |
| 4.5.1.2   | Using push & pull tugboats .....                             | 73         |
| 4.5.2     | Training performance of baseline scenario B re-creation..... | 75         |
| 4.5.2.1   | Using push tugboats only & curriculum .....                  | 78         |
| 4.5.2.2   | Using push tugboats only .....                               | 81         |
| 4.5.2.3   | Using push & pull tugboats .....                             | 84         |
| 4.6       | Limitations/constraints of agent training .....              | 88         |
| 4.6.1     | MOSES training limitations.....                              | 88         |
| <b>5.</b> | <b>Predictive Battery Optimisation .....</b>                 | <b>92</b>  |
| <b>6.</b> | <b>Fail-safe Operation .....</b>                             | <b>94</b>  |
| 6.1       | Methodology.....   | 94         |
| 6.2       | Requirements.....  | 98         |
| 6.3       | STPA Results.....  | 99         |
| 6.3.1     | Purpose of the Analysis .....                                | 99         |
| 6.3.2     | Control structure modelling .....                            | 101        |
| 6.3.3     | Unsafe Control Actions identification .....                  | 105        |
| 6.3.4     | Loss scenarios identification.....                           | 106        |
| 6.4       | Minimum Risk Condition (MRC) States.....                     | 109        |
| 6.4.1     | MRC States Identification .....                              | 109        |
| 6.4.2     | Correspondence of MRC States to loss scenarios .....         | 113        |
| 6.5       | Development outline for fail-safe software.....              | 116        |
| <b>7.</b> | <b>Conclusion.....</b>                                       | <b>119</b> |
|           | <b>References.....</b>                                       | <b>121</b> |

## List of Tables

|  |     |
|--|-----|
| Table 1: Reward Assignment Table .....   | 59  |
| Table 2 MOSES Desktop Setup .....  | 91  |
| Table 3 Description of the terms used in the STPA process [36]. .....  | 96  |
| Table 4 Indicative MRC states described in the DNV and BV guidelines [33], [34].....                         | 97  |
| Table 5 Functional requirements and specifications related to the fail-safe operation<br>functionality ..... | 98  |
| Table 6 STPA - Identified Losses .....   | 100 |
| Table 7 STPA - Identified Hazards.....   | 101 |
| Table 8 STPA Identified System-Level Constraints .....   | 101 |
| Table 9 STPA - Identified Controller Responsibilities .....  | 102 |
| Table 10 STPA Identified Unsafe Control Actions (UCAs).....  | 105 |
| Table 11 STPA Identified Loss Scenarios .....  | 107 |
| Table 12 Identified MRC States .....   | 110 |
| Table 13 MRC States corresponding to the identified Loss Scenarios .....                                     | 113 |

## List of Figures

|  |    |
|--|----|
| Figure 1: Stock market forecasting is typical supervised learning task<br>( <a href="https://pythondata.com/stock-market-forecasting-with-prophet/">https://pythondata.com/stock-market-forecasting-with-prophet/</a> ).....   | 12 |
| Figure 2: Data clustering is an unsupervised learning technique that can point out useful<br>insights on raw data ( <a href="https://deeptai.org/machine-learning-glossary-and-terms/unsupervised-learning">https://deeptai.org/machine-learning-glossary-and-terms/unsupervised-learning</a> )..... | 13 |
| Figure 3: The agent-environment interaction in a Markov Decision Process .....   | 14 |
| Figure 4: Cart-pole balance example.....   | 15 |
| Figure 5: Traffic network .....  | 16 |
| Figure 6: Model-free and model-based strategies to solve a hypothetical action-selection<br>problem using a rat in a maze [1].....   | 20 |
| Figure 7: Virtual Environment MDP .....  | 24 |
| Figure 8: Schematic illustration of the convolutional neural network used in DQN.....  | 26 |
| Figure 9: Score of different DQN implementations on Atari games .....  | 30 |
| Figure 10: Failure dimensions of ML (“Why is Machine Learning ‘Hard’?”) .....  | 31 |
| Figure 11: An open-loop control system (a), and a closed-loop control system (b). Source:<br>Dorf and Bishop (2010).....   | 32 |
| Figure 12: A PID controller in a closed loop control system.<br>Source: <a href="https://plcynergy.com/pid-controller/">https://plcynergy.com/pid-controller/</a> .....  | 34 |
| Figure 13: Pyramids Unity Environment.....   | 40 |
| Figure 14: Training Reward .....   | 41 |
| Figure 15: Example of variations of the 3D Ball example environment. The environment<br>parameters are gravity, ball mass and ball scale.....  | 43 |
| Figure 16: MOSES Unity test scene displayed during training of 3 push agents next to the<br>“Advanced Ship Controller” and “Behavior parameters” component.....  | 46 |
| Figure 17: Decision Requester Component.....   | 47 |
| Figure 18: Behavior script modifiable fields .....   | 47 |
| Figure 19: LiDAR virtual placement on tugboat .....  | 48 |
| Figure 20: Unity LiDAR component specifications .....  | 48 |
| Figure 21: Behavior parameters component designed for the MOSES training environment   | 49 |

Figure 22: Example during training of tugboat agent moments before crashing the dock .... 53

Figure 23: Example during training of tugboat agent when escaping port during space exploration ..... 53

Figure 24: Approach point of mother vessel (highlighted in red) ..... 54

Figure 25: Virtual LiDAR ray highlighted in white..... 55

Figure 26: Target of the front pushing tugboat..... 56

Figure 27: Case where the front tugboat needs to brake (highlighted in red) ..... 58

Figure 28: Case where the back tugboat needs to brake (highlighted in red)..... 58

Figure 29: Baseline Scenarios Positions ..... 60

Figure 30: Baseline Scenario A - Phase 1: Bring the vessel to a position parallel to the dock. FP and AP correspond to the fore and aft peak of the vessel..... 61

Figure 31: Baseline Scenario A - Phase 2: Parallel transfer of the vessel to position C ..... 62

Figure 32: Baseline Scenario A - Phase 2: Corrective action mode (negative yaw angle)..... 63

Figure 33: Baseline Scenario B - Phase 1: Bring the vessel to a position parallel to the dock 65

Figure 34: Baseline Scenario B - Phase 2: Parallel transfer of the vessel to position C..... 66

Figure 35: Custom thruster control script as seen in the Unity editor..... 68

Figure 36: Example where vessel’s rotation angle is beyond allowed bounds..... 68

Figure 37: Target point highlighted on the dock ..... 69

Figure 38: Environment training state at t=0 ..... 70

Figure 39: Zoomed-in plot of push agent rewards..... 71

Figure 40: Zoomed-out plot of push agent rewards ..... 71

Figure 41: Smoothed out policy loss ..... 72

Figure 42: Agent model entropy ..... 72

Figure 43: Environment training state at t=0 ..... 73

Figure 44: Plot of push agent rewards ..... 74

Figure 45: Smoothed out value loss ..... 75

Figure 46: Placement of 3 push agents relative to the mother vessel..... 76

Figure 47: Smoothed out rewards for 3 push agents attached on the bow (orange), stern (blue) and middle (red) side of the ship ..... 76

Figure 48: Smoothed out policy loss for 3 push agents attached on the bow (orange), stern (blue) and middle (red) side of the ship ..... 77

Figure 49: Curriculum learning in comparison to training steps..... 77

Figure 50: Environment training state at t=0 ..... 78

Figure 51: Agent attachment achieved at training step 115.000 (highlighted in red)..... 79

Figure 52: First lesson rewards of front (orange) and back (blue) agent..... 80

Figure 53: Second lesson smoothed-out rewards of front (orange) and back (blue) agent... 80

Figure 54: Second lesson smoothed-out policy loss of front (orange) and back (blue) agent 81

Figure 55: Second lesson smoothed-out mean value estimate of front (orange) and back (blue) agent ..... 81

Figure 56: Environment training state at t=0 with no curriculum..... 82

Figure 57: Full run rewards of front (blue) and back (orange) agent..... 82

Figure 58: Zoomed-in smoothed out policy loss of front (blue) and back (orange) agent ..... 83

Figure 59: Zoomed-in smoothed out policy value estimate of front (blue) and back (orange) agent..... 83

Figure 60: Zoomed-in entropy of front (blue) and back (orange) agent..... 84

Figure 61: Environment training state at t=0 ..... 85

Figure 62: Rewards of front push (red) and back push (orange) agent ..... 85

Figure 63: Group push rewards..... 86

Figure 64: Zoomed-in smoothed out policy loss of front (blue) and back (cyan) agent ..... 86

Figure 65: Series of screenshots showing different stages of the simulation, starting from the upper left image ..... 87

Figure 66: Virtual camera point of view showcase ..... 89

Figure 67: Vessel longitude evolution between two close attachment points (Unity coordinates)..... 89

Figure 68: Attachment position A (above) and B (below) ..... 90

Figure 69: Front (orange) and back (blue) push agent rewards..... 91

Figure 70: Typical Load profile of a harbor tug ..... 92

Figure 71 Generic control structure [36]..... 96

Figure 72 The terms used in the STPA process and their relationships [36]..... 96

Figure 73 STPA Control Structure for the autonomous tugboat swarm ..... 104

Figure 74 Control Structure depicting the connections required for the identified MRC-States ..... 112

Figure 75 Basic functionalities for the implementation of the fail-safe functionality ..... 117

Figure 76 Autonomous tugboat architecture; modified for fail-safe functionality ..... 118



## List of Acronyms

| Abbreviation / acronym | Description                               |
|------------------------|---|
| EC                     | European Commission                       |
| D1.1                   | Deliverable number 1 belonging to WP 1    |
| WP                     | Work Package                              |
| AI                     | Artificial Intelligence                   |
| ML                     | Machine learning                          |
| RL                     | Reinforcement learning                    |
| MDP                    | Markov Decision Process                   |
| DQN                    | Deep Q-Network                            |
| MCTS                   | Monte Carlo tree search                   |
| PPO                    | Proximal Policy Optimisation              |
| SAC                    | Soft Actor Critic                         |
| PFA                    | Policy function approximation             |
| CFA                    | Cost function approximation               |
| VFA                    | Value function approximation              |
| DLA                    | Direct lookahead approximation            |
| BC                     | Behavioral Cloning                        |
| GAIL                   | Generative Adversarial Imitation Learning |
| MA-POCA                | MultiAgent Posthumous Credit Assignment   |
| NN                     | Neural Network                            |
| Eq.                    | Equation                                  |
| ECR                    | Empty Container Repositioning             |
| LCG                    | Longitudinal center of gravity            |
| TCG                    | Transverse center of gravity              |
| DL                     | Deep Learning                             |

## Executive Summary

This document reports the activities and the outcome of the MOSES Task 4.3: Swarm intelligence algorithm development & simulation, which is part of WP4. In this task, we present the design, development and optimisation of the algorithms that enable intelligence of virtual tugboat agents in order to automate the docking process of a large ship. The main objective of this task is to teach a swarm of autonomous tugboats to establish path-planning, execute sophisticated manoeuvres and ultimately control a large vessel while coordinating actions among the actors of the swarm.

Leveraging both the virtual environment developed in the Unity3D framework under Task 4.2 and the ML-Agents toolkit, we showcase that it is possible to train the behavior of the tugboat agents using the Proximal Policy Optimisation Reinforcement Learning algorithm. During algorithm training, the agents are provided with real-time feedback from the environment and thanks to their own reward functions, they are able to dynamically adapt to its policies and navigation strategy.

The performance evaluation shows that by formulating appropriate reward functions, selecting the correct training strategy, tuning model hyperparameters, and shaping penalties, the agents can successfully learn the desired task and improve the performance and quality of an important maritime operation such as the docking of large ships.

Finally, within this document a Fail-safe operation analysis is presented that provides the foundations for the fail-safe functionality. The latter is intended to be used by the autonomous tugboat swarm to maintain the safety of the manoeuvring operation in case a failure occurs (e.g. connectivity loss, other alarm etc.).