Data driven digital twins for the maritime domain

Alexandros TROUPIOTIS-KAPELIARIS[a], Nicolas Zygouras[b], Manolis KALIORAKIS[b], Spiros MOUZAKITIS[c], Giannis TSAPELAS[c] , Alexander Artikis[d], Eva Chondrodima[d], Yannis Theodoridis[d], Dimitris ZISSIS[a,b]
[a] *University of the Aegean*
[b] *MarineTraffic*
[c] *National Technical University of Athens*
[d] *University of Piraeus*

Abstract. Digital twins are computational models that replicate the structure, behaviour and overall characteristics of a physical asset in the digital world. In the maritime domain, conventional approaches have relied on mathematical modeling (e.g., linearised equations of motion) and heavy computations for estimating ship resistance and propulsion, seakeeping and maneuverability and overall hull form optimization, treating the vessel as a point body. For instance, the ability to predict a vessel's future track in confined or congested waters presents a significant challenge due to the fact that as time passes, these models often fall out of sync with their digital counterparts due to changes that happen to the ship (e.g., foulding affecting maneuverability). In addition to this, mostly due to computational resources required, in real world deployments models are simplified, thus reducing their overall prediction accuracy. In our work, we implement AI-enabled coupled abstractions of the asset-twin system, which rely on machine learning methods for constant learning of the evolving over time behavior of a vessel based on historical trip data and information related to vessel's structure and loading capacity. The evaluation results indicate that the inclusion of vessel and journey specific information is beneficial for the predictions.

Keywords. vessel, digital twins, machine learning, big data, trajectory prediction

1. Introduction

The advancement of computational resources establishes the Digital Twins (DTs) as an important digitalization trend across industries [1][2]. In the maritime industry, a digital twin is a digital representation of any physical object, asset or system, which can be translated into a ship engine, embedded vessel sensors, a hull, a propeller, or even an entire ship. It can contain various digital models such as 3D, simulation or Artificial Intelligence (AI) models [3][4]. Several digital twin approaches are already applied covering the entire lifecycle of a vessel (design, manufacturing, operation, and maintenance phase) [5][6][7][8]. These approaches bring several benefits to the stakeholders [9][10], who try to cope with major challenges (e.g. reduction of costs, increase of overall efficiency, reliability, and sustainability) coming from new market demands [11], or stricter environmental and safety regulations [12].

This paper presents a data-driven digital twin approach for the entire trajectory forecasting of the vessel till the destination port. Subsections 1.1 and 1.2 illustrate the related work and the problem definition, while sections 2 and 3 present the methodology and the evaluation results. Finally, Section 4 showcases the conclusions of this work.

1.1. Related work

DTs were initially and widely used in marine industry for design purposes, where Computer Aided Design (CAD) systems have to a large extent replaced traditional manual design, solving complex engineering tasks faster and more precisely using computer software such as 3D models and simulation tools [13]. Nowadays, digital twins were evolved from the static and descriptive models of the early CAD approaches to dynamic models for the behavior of the virtual entity during its entire lifecycle and operation [14][15].

The implemented approaches to reduce the gap between physical and virtual entities can be classified based on the software design patterns followed for the DT creation [16]. ML and physics-based DTs are two distinct implementation approaches of DTs in the maritime industry. The main advantages and disadvantages between the ML and physics-based DTs are summarized in Table 1 [16], indicating that both approaches can be complementary and are always oriented to the specific scope and/or restrictions of the study. The tradeoff among the levels of performance, scalability and level of accuracy for each DT are the main factors driving the decisions on the type of the model that will be finally implemented. Some recent approaches to achieve high performance follow the co-simulation pattern for distributed resources, data, and models [9][17], however these approaches do not reach the high scalability levels for a large fleet of vessels, which can be achieved with a more abstract approach based on data science and ML principles.

Table 1. Advantages and disadvantages of machine learning and physics based digital twins.

Machine learning based		**Physics based**	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
• Model derived from data only – no need for domain knowledge • Generic and flexible handling heterogenous data types to extract knowledge • Model improves over time (reinforcement learning) • Good at discovering complex relationships and patterns • Scaling for different vessels and parameters can be easily achieved leveraging data science principles.	• Availability of large size training data to develop the model • Correlations and not causalities can be extracted. Back-box, no explanations for the behavior for some cases • Approximation methods, no exact mathematics • Predictive capabilities deteriorate quickly outside training set scope • Difficult to predict extreme/critical conditions due to few observations	• Models capture deep existing knowledge on physics • Causal relationships provide insight and understanding • Uncertainty controlled by input and modeling accuracy • Model has universal validity	• Require extensive domain (physics) knowledge • Computationally intensive (increased accuracy leads to extra needs for resources) • Scaling restrictions (extra resources and time are required to run multiple experiments for different vessels and parameters) • Assumptions about input-output should be made upfront

DTs are also applied on the route planning process [18], when traditionally the route is plotted beforehand to be the most efficient in terms of fuel consumption, speed or distance travelled. With the rise of the first digital twins, the first geographic information systems [19] were created to assist the nautical navigation, while complying with high safety standards and strict regulations. Traditional applied methods for modelling ship movement are based on kinematic models and motion measurements such as acceleration [20][21], without accounting dynamic effects such as loading condition and environmental disturbances due to wind, waves and current; information that is handled by advanced dynamic models [22]. However, limitations due to the lack of direct measurements of these dynamic effects (e.g., current drift) leads to oversimplified models with substantial discrepancies between the behavior of the real ship and the virtual model. ML based and data driven approaches can bridge such discrepancies, providing accurate results for different vessels at large scale.

After the establishment of the Automatic identification system (AIS), and as moving data volumes and availability increased, transportation intelligence also evolved with increased need for data-driven methods combined with DT technology to optimize moving objects operation. Various ML and data-driven methods based on AIS have been proposed in the past years to address the route (trajectory) forecasting problem in the maritime [24][25][26][27][28][29][30][31]. In addition, some previous works discuss digital twins' implementations for planning and security monitoring of the transport corridors [7][32][33]. Finally, [34] recently proposed a data driven methodology to forecast the vessel trajectories in the open sea free-space. However, the aforementioned methods are focused on extracting knowledge only from the spatio-temporal information of the AIS messages without taking into consideration additional static information related to the structural characteristics of the vessel (such as the size), and dynamic information such as the draught, deadweight tonnage (dwt), and schedule details, which can potentially affect the final trajectory of the vessel.

1.2. Problem description

A trajectory is a path that a moving object follows through space as a function of time. To capture though the accurate and complete trajectory of a moving object, is almost impossible in real conditions, due to the inherent limitations of data acquisition and storage mechanisms. As a result, the continuous movement of an object is usually obtained as an approximate form of discrete samples of spatiotemporal locations (Figure 1). Thus, it can be captured as a time stamped series of location points as in Eq. (1), where x_i, y_i represent geographic coordinates of the moving object at time t_i and N is the total number of elements in the series. Notice that the approximated trajectory can also be represented as a series of line segments between the stamped positions as in Eq. (2), given that there is a unique identifier grouping these positions into the same path.

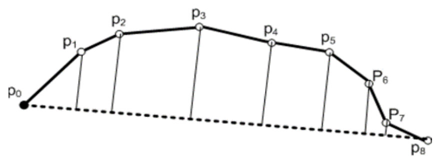

Figure 1. Approximated trajectory captured by a sensor

$$p_0(x_0, y_0, t_0), p_1(x_1, y_1, t_1), \ldots, p_N(x_N, y_N, t_N) \tag{1}$$

$$traj_1 = p_0p_1, p_1p_2, p_2p_3, p_3p_4, p_4p_5, p_5p_6, p_6p_7, p_7p_8 \tag{2}$$

In this work, an efficient technique is developed to forecast the path of the vessel till its arrival at the destination port by receiving as input the current location of a particular vessel along with its destination port. The proposed model is trained based on a huge data of historical trajectories for the same journey, while incorporating additional static and dynamic features to improve accuracy of the prediction, leading to a Digital Twin model aligned with its physical counterpart.

2. Methodology

Although vessels in theory may be considered to be moving in free-space, historical data indicate that is not the case. In general, vessels tend to follow specific pathways across the seas in order to reach their destination. These pathways may originate from international passages, hydrographic studies or simply serve the closest path towards the destination port. The proposed approach translates historical data of a specific route (i.e., trajectories from a single pair of origin-destination ports) into a directed network. Figure 2 presents all stages of the proposed approach for the actual training of the algorithm and the route prediction of a vessel based on its location and destination port. After discovering the frequent pathways of movement (corridors), the relations between neighboring pathways are modelled through the creation of a directed graph, which represents the movement of vessels. In the next step, and in order to improve the prediction accuracy, the algorithm is enhanced with classification models. Each model is responsible for deciding which pathway the vessel is to follow in a specific junction of the created graph. Features considered by these models include static and dynamic information regarding the moving vessel and the journey. The following sections describe in more detail all stages of the approach.

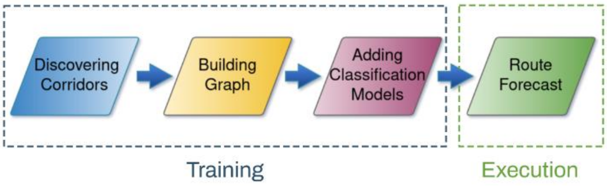

Figure 2. Training and execution stages of the proposed methodology.

2.1. Extracting corridors

In the first stage of the methodology, the major pathways of the vessels are extracted through a preprocessing on a huge volume of historical data. More precisely, the AIS messages that were collected are cleaned and partitioned according to the route they belong. This is achieved by splitting the historical data of each vessel on trajectories beginning from and ending at a port, using the port geometries as an indicator based on [35]. After grouping the trajectories from all vessels based on the origin and destination port, a separate model for each group is constructed.

Since the goal of this step is to determine the common pathways between vessels of the same route, clustering techniques are used to create spatial groups between the positions found in the historical trajectories. First, one of the historical trajectories is selected as the baseline trajectory. Moving along consecutive points of this trajectory, corresponding rectangles are generated, and then a clustering for all points falling in the same rectangle is performed (using the DBSCAN algorithm [36]). By keeping the clusters that include the baseline trajectory in each case, series of corresponding areas are extracted, forming a corridor. After removing all the trajectories that were covered by the previous steps, a new baseline trajectory is selected and the same process is performed to create all corridors of the journey (highlighted with color in Figure 3). In the end, multiple corridors are extracted from the dataset. The final number of the extracted corridors varies depending on both the clustering method, the parameters defined for the geometry extraction, as well as the characteristics of the route itself. Finally, for each corridor a representative path is extracted based on all the trajectories encapsulated in this corridor.

2.2. Creating movement graph

After extracting the corridors, the vessel movement is translated into a directed graph. To achieve this, the sub-areas where vessels have changed corridors during their travel are firstly detected. The number of vessels following this transition is captured as an indicative weight for each case. Next, a directed weighted graph is created for the vessel movement, so that each node represents these previously extracted points of transition between different corridors.

2.3. Adding classification models

To enhance the prediction with additional static and dynamic information, a mechanism that updates the graph weights is incorporated. Specifically, after the creation of the graph, these additional features are passed at the transition points to train a separate classification model at each point. This way, during the prediction step, the graph is updated based on the query's characteristics and achieve more accurate results close to the physical replica of the vessel. The features taken into account in this study are the length, draught, deadweight tonnage (dwt) of the vessel as well as the day of the week the journey took place. The proposed system trains two classification models (based on Decision Trees [37] and Support Vector Machines [38]) for each junction point of the graph and keeps the most accurate one for the prediction.

2.4. Route forecasting

Using the created graph and the classification models from the previous stage and based on a vessel's current position, a prediction for a vessel's path until the destination port is produced by firstly identifying the closest corridor to the vessel's position. Next, the algorithm moves along the corridor's baseline trajectory until the next transition point, when the graph's weights are used to decide whether to continue on the current corridor, or to change and continue the journey accordingly. These steps are repeated until reaching the destination port.

3. Evaluation

3.1. Experimental setup

To evaluate the accuracy of our approach and quantify the inclusion of several features on the final prediction, two AIS datasets were used, each concerning a specific pair of origin-destination ports. The first dataset consists of all trips of passenger vessels from the port of Mykonos to Piraeus for a single year (2019), while the other includes all trips of container vessels from Marsaxlokk to Thessaloniki for a four-year period (2016-2019). For the former dataset 82 trajectories were considered during training and 36 for evaluation, while for the latter the trajectories were split to 129 and 7, respectively. We perform full trip predictions from the origin port using different features as input and the results are compared to the real trajectory of the vessel collected through AIS using the Dynamic Time Warping (DTW) method [39].

3.2. Results

Table 2 summarizes the results of the study for the aforementioned dataset, while Figure 3 illustrates the extracted corridors of the two journeys.

Table 2. Experimental results on real data for the route Mykonos-Piraeus targeting passenger vessels and Marsaxlokk-Thessaloniki for container vessels, respectively. The results shown are calculated through DTW and measured in km. The best combination of features in terms of prediction accuracy is highlighted in bold.

Classification features	**Mykonos-Piraeus**	**Marsaxlokk-Thessaloniki**
(None)	16.174813	10.930125
dayOfWeek	14.000156	10.925935
dayOfWeek, draught	**13.954733**	10.328922
dayOfWeek. dwt	14.000127	10.914007
dayOfWeek, length	13.986446	**10.307028**
dayOfWeek, draught, length	14.412126	10.925935
dayOfWeek, dwt, length	14.397106	10.914007
dayOfWeek, draught, dwt	13.987193	10.930887
dayOfWeek, draught, dwt,length	13.99941	10.944829

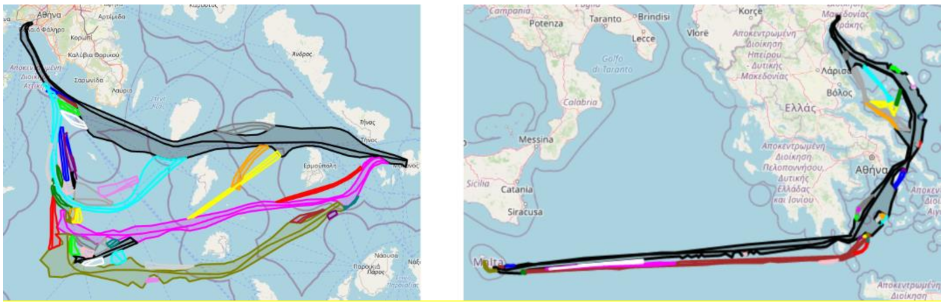

Figure 3. Resulting corridors for the Mykonos-Piraeus (left) and Marsaxlokk-Thessaloniki (right). Each corridor is represented by a different color.

In general, the use of most features benefits our approach, decrease the prediction error up to 13.73% compared to the implementation without including any static or dynamic information using the real trajectory as reference point. As seen in Figure 3, the selected routes are of high complexity, in terms of the number of different extracted corridors, mainly due their passage through the Aegean archipelago, which consists of numerous islands, giving many alternative paths for the journey. Furthermore, some combinations of more than one feature result in a more accurate prediction. For instance, it can be seen that the inclusion of both the journey's day of the week along with the vessel's draught leads to significant improvement to the resulting predictions in both journeys. On the contrary, in some cases allowing multiple features to be included may damage the model's accuracy, as is illustrated in the last row of the table.

4. Conclusions

This work explores the use of DTs in the maritime domain, focusing mainly on the vessel trajectory prediction problem. Though several works provide accurate results for short-term route forecasting, the presented work is dedicated to predicting the vessel's future path until the destination port, regardless of the distance. Furthermore, based on clustering and classification techniques, the proposed method uses past movement patterns from historical data and vessel-specific characteristics to return better-suited predictions. Experimental results on real data show that our approach performs prediction of high accuracy even for complex trips (i.e., Mykonos-Piraeus), while further analysis indicates that the inclusion of additional static and dynamic information features can refine the prediction accuracy. Finally, the addition of dynamic information related to weather conditions may be considered as an extension of the current work.

Acknowledgements

This research is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 957237, project ENABLING MARITIME DIGITALIZATION BY EXTREME-SCALE ANALYTICS, AI AND DIGITAL TWINS (VesselAI).

References

[1] Barricelli BR, Casiraghi E, Fogli D. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. In: IEEE Access; 2019, vol. 7, pp. 167653-167671, doi: 10.1109/ACCESS.2019.2953499

[2] Semeraro C, Lezoche M, Panetto H, Dassisti M. Digital twin paradigm: A systematic literature review. In: Computers in Industry, Elsevier, 2021, 130, pp.103-469.

[3] Aragão Í, Gaspar FH. Challenges when creating a cohesive digital twin ship: a data modelling perspective, Ship Technology Research, 2021, 68:2, 70-83, doi: 10.1080/09377255.2020.1815140

[4] Smogeli O. Digital twins at work in maritime and energy, DNV-GL Feature, Feb. 2017 pp. 1-7

[5] Hatledal L, Skulstad R, Li G, Styve A, Zhang H. Co-simulation as a Fundamental Technology for Twin Ships. In: Journal of Modeling, Identification and Control (MIC), vol. 41, No. 4, 2020, pp. 297- 311.

[6] Giering JE, Dyck A. Maritime Digital Twin architecture: A concept for holistic Digital Twin application for shipbuilding and shipping. In: Automatisierungstechnik, vol. 69, no. 12, 2021, pp. 1081-1095. https://doi.org/10.1515/auto-2021-0082.

[7] Guo ZZ, Wang K, Qin JY, Li XB, Ran DC, Shen Y. An Intelligent Maritime Scene Frame Prediction Based on Digital Twins Technology. In: IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI), 2021, pp. 25-28. https://doi.org/10.1109/DTPI52967.2021.9540188

[8] Veenstra AW. New Business Models for Shipping: Innovation in the Netherlands. In: New Maritime Business. WMU Studies in Maritime Affairs, 2021, vol 10. Springer, Cham. https://doi.org/10.1007/978-3-030-78957-2_11

[9] Perabo F, Park D, Zadeh MK, Smogeli O, Jamt L. Digital twin modelling of ship power and propulsion systems: Application of the open simulation platform (osp). In: IEEE 29th International Symposium on Industrial Electronics (ISIE), 2020, pages 1265-1270. doi:10.1109/ISIE45063.2020.9152218.

[10] Bekker A. Exploring the blue skies potential of digital twin technology for a polar supply and research vessel. In: Proceedings of the 13th International Marine Design Conference Marine Design XIII (IMDC 2018), vol 1, pages 135-146, 2018. doi:10.1201/9780429440533-11.

[11] Van Os J. The Digital Twin throughout the Lifecycle. SNAME Maritime Convention, Nov. 2018.

[12] International Maritime Organization, Marpol. ANNEX VI and NTC 2008, 2013: With Guidelines for Implementation, Intl Maritime Organization, 2013.

[13] Marine Insight. Top 20 Software Used In Naval Architecture. Available from: https://www.marineinsight.com/naval-architecture/top-20-software-used-in-naval-architecture/

[14] Hellenic Marine Environment Protection Association. HELMERA simulator. Available from: https://helmepa.gr/en/maritime-section/simulator/
[15] Smart Maritime Network. Lind M, Becha H, Watchon RT, Kouwenhoven N, Zuesongdham P, Baldauf U. Digital twins for the maritime sector. Available from: https://smartmaritimenetwork.com/wp-content/uploads/2020/07/Digital-twins-for-the-maritime-sector.pdf
[16] Erikstad SO. Design Patterns for Digital Twin Solutions in Marine Systems Design and Operations. In: 17th International Conference on Computer and IT Applications in the Maritime Industries: Pavone, 14-16 May 2018
[17] Kongsberg. C-SCOPE. Available from: https://www.kongsberg.com/kda/what-we-do/maritime-surveillance/c-scope/
[18] Busse A, Gerlach B, Lengeling JC, Poschmann P, Werner J, Zarnitz S. Towards Digital Twins of Multimodal Supply Chains. Logistics. 2021; 5(2):25. https://doi.org/10.3390/logistics5020025 Chicago/Turabian Style
[19] IMO. Electronic Nautical Charts (ENC) and Electronic Chart Display and Information Systems (ECDIS). Available from: https://www.imo.org/en/OurWork/Safety/Pages/ElectronicCharts.aspx
[20] Nomoto K, Taguchi T, Honda K, Hirano S. On the steering qualities of ships. In: International Shipbuilding Progress, 4(35):354-370, 1957.
[21] Amerongen JV. Adaptive steering of ships|a model reference approach. In: Automatica, 20(1), 1984.
[22] Kapteyn MG, Knezevic DJ, Willcox K. Toward predictive digital twins via component-based reduced-order models and interpretable machine learning. In AIAA Scitech 2020 Forum, page 04-18, 2020.
[23] Pallotta G, Vespe M, Bryan K. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. In Entropy, 2013, 15(6), 2218-2245
[24] Tu E, Zhang G, Rachmawati L, Rajabally E, Huang GB. Exploiting ais data for intelligent maritime navigation: A comprehensive survey from data to methodology. In: IEEE Trans. Intell. Transp. Syst., vol. 19, no. 5, pp. 1559–1582, 2018
[25] Petrou P, Tampakis P, Georgiou H, Pelekis N, Theodoridis Y. Online long-term trajectory prediction based on mined route patterns. In: International Workshop on Multiple-Aspect Analysis of Semantic Trajectories, (pp. 34-49). Springer, Cham, 2019.
[26] Valsamis A, Tserpes K, Zissis D, Anagnostopoulos D, Varvarigou T. Employing traditional machine learning algorithms for big data streams analysis: The case of object trajectory prediction. In: Journal of Systems and Software, 2017, 127, 249-257
[27] Tang H, Yin Y, Shen H. A model for vessel trajectory prediction based on long short-term memory neural network. In: Journal of Marine Engineering & Technology, 2019, pp. 1-10
[28] Stogiannos M, Papadimitrakis M, Sarimveis H, Alexandridis A. Vessel Trajectory Prediction Using Radial Basis Function Neural Networks. In: IEEE EUROCON 2021 - 19th International Conference on Smart Technologies, 2021, pp. 113-118, doi: 10.1109/EUROCON52738.2021.9535562
[29] Papadimitrakis M, Stogiannos M, Sarimveis H, Alexandridis A. Multi-Ship Control and Collision Avoidance Using MPC and RBF-Based Trajectory Predictions. In: Sensors 2021, 21, 6959. https://doi.org/10.3390/s21216959
[30] Xiao Z, Fu X, Zhang L, Goh RS. Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey. IEEE Transactions on Intelligent Transportation Systems. 2019 Apr 18;21(5):1796-825.
[31] Riveiro M, Pallotta G, Vespe M. Maritime anomaly detection: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2018 Sep;8(5):e1266.
[32] Kupriyanovsky V, Klimov A, Gots I, Volodin A, Jabitsky M, Kupriyanovsky J. Technologies of digital twins in transport corridors for sea and waterways in Russia. In: International Journal of Open Information Technologies, 2019, 8(12), 113-132
[33] Liu J, Li C, Bai J, Luo Y, Lv H, Lv Z. Security in IoT-Enabled Digital Twins of Maritime Transportation Systems. In: IEEE Transactions on Intelligent Transportation Systems, 2021
[34] Zygouras N, Spiliopoulos G, Zissis D. Detecting representative trajectories from global AIS datasets. In: IEEE International Intelligent Transportation Systems Conference (ITSC), 2021 pp. 2278-2285, doi: 10.1109/ITSC48978.2021.9564657.
[35] Zissis D, Chatzikokolakis K, Spiliopoulos G, Vodas M. A distributed spatial method for modeling maritime routes. IEEE Access. 2020 Mar 9;8:47556-68.
[36] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Inkdd 1996 Aug 2 (Vol. 96, No. 34, pp. 226-231).
[37] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Routledge; Oct 2017.
[38] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. InProceedings of the fifth annual workshop on Computational learning theory 1992 Jul 1 (pp. 144-152).
[39] Berndt DJ, Clifford J. Using dynamic time warping to find patterns in time series. InKDD workshop 1994 Jul 31 (Vol. 10, No. 16, pp. 359-370).