HOLISTIC DIGITAL TWIN OF THE OCEAN

Emilio Fortunato Campana *and* Gianpaolo Coro

DIITET-CNR, Rome, Italy

e-mail: {emiliofortunato.campana, gianpaolo.coro}@cnr.it

We introduce the concept of holistic Digital Twin of the Ocean (*h*-DTO), a novel physics-based and data-driven integrated model, which uses AI and Decision Support Systems to model the complex relations between all systems composing a marine area. The *h*-DTO models a marine area as a system of systems. The AI components process data flows to simulate ecosystem functions and answer complex questions about the ecosystem's state, response to perturbations, and the future change of biodiversity and ecosystem services. The *h*-DTO overall aims to enhance the ecosystem understanding and improve our knowledge of the oceans. In this paper, we give indications of the main *h*-DTO features and propose design and implementation pathways.

Keywords: *holistic digital twin of the ocean*, *artificial intelligence*, *decision support systems*, *marine ecosystems*, *ecosystem functions*

1. Introduction

The unsustainable exploitation of ocean resources - with overfishing, chemical and physical pollution, and heavy maritime traffic - is threatening European oceans, seas, and coasts [1]. Climate change is exacerbating this issue. To mitigate this pressure, there is an urgent need for new intelligent digital technologies [2]. Integrated Environmental Assessment (IEA) systems and Ecosystem Models (EMs) can model causal links between anthropogenic driving forces, environmental pressures, and the impact on and response of ecosystems [3]. However, these models heavily rely on data interoperability and scalability capacity and use heuristic and non-automatic approaches with limited result transparency. Artificial Intelligence (AI) can help overcome these limitations and develop scalable and robust virtual representations as Digital Twins of the Ocean (DTOs), combining data from different, heterogeneous, and real-time data flows.

The present paper introduces a new physics-based and data-driven integrated model corresponding to a holistic concept of DTO (*h*-DTO). It empowers AI models with physical/biological laws to model the complex relations between the systems composing a marine area. The *h*-DTO considers the ocean as a system of systems (e.g., morphological, geological, chemical, physical, biological, and economical). Its AI components process data flows to simulate a functioning marine ecosystem and answer complex questions about its current state, its resilience and expected responses to perturbations, and the future change and impact on resident species and ecosystem services. The *h*-DTO adapts to the available data of a marine area - rather than being constrained to a minimum type and quantity of data - performs predictions and informs decision makers.

2. Research questions

An *h*-DTO addresses the following research questions:
1. Can we operate sustainably in a local marine area, given the interconnections between ecosystems and activities and existing constraints from complex natural and anthropogenic driving forces acting on the area?
2. Can we predict ecosystem responses to perturbations and disruptions?
3. Can we predict long-term changes to the ecosystem and blue economy due to climate change effects?

These macro-questions require answering more detailed research questions, such as:

1. Can we automate the discovery of natural relations between the ecosystem, environmental conditions, and anthropogenic stressors?
2. Which data are required within a Decision Support System (DSS) to generate reliable predictions and suggestions?
3. Can AI fill the possible gaps in these data?
4. How can the DSS scale?
5. Can the data and results produced for one area inform the DSS of another area?
6. How can consequences from AI-driven biases, assumptions and omissions be identified and mitigated?

3. Theoretical framework

The *h*-DTO adopts a Bayesian-like approach for decision support. It follows a marine environment over time through environmental monitoring and data accumulation. Meanwhile, it infers information about the exploitation and ecological sustainability of the marine area, the status of the ecosystem, and its capability to adapt to climate change (e.g., sea level rise, extreme events, and long-term change). The *h*-DTO constantly collects new data, integrates these with existing data - available from public and local providers, IoT, and possibly citizen-science networks -, assimilates past and new data with data-driven and physics-based computational models, and intelligently reuses historical heterogeneous data and models via cloud-based analyses. Overall, it can predict the evolution of a marine environment under the effects of heterogeneous driving forces acting on the areas and helps optimise and conserve resources while considering the policies in place.

An active *h*-DTO should live within a virtual environment of an Open Science-compliant e-Infrastructure to ensure repeatability, reproducibility, and reusability features for all models and publish, communicate, and disseminate maps and data through long-term sustainable catalogues. These features should guarantee access and sharing facilities for the marine research activities and campaigns, the re-processed data flows, and the forecasts. The catalogues and the virtual environments should promote the *h*-DTO predictions, assessments, knowledge gap filled, and the emerging possibilities to build productive interactions with the research and industrial sectors. Moreover, they should increase coastal communities' involvement, consensus, and sense of responsibility.

The recipe for building such a system requires working on four main dimensions: (i) Data census, which requires identifying available oceanographic, biodiversity, maritime, social, and fisheries data flow; (ii) strategies to standardise and gap-fill the identified data flows and improve their usability in AI models; (iii) design of AI models to automatically discover ecosystem functional connections; (iv) design of DSSs to simulate complex relations within the ecosystem and answer the research questions.

One critical aspect is model scalability. The *h*-DTO must have modularity, scalability, and adaptability features, such as:

- *spatial* scalability, i.e., generalisation from local to extensive areas;
- *complexity* adaptability, i.e., managing from uniform to heterogeneous data flows and drivers;
- *integration* scalability, i.e., managing from one area to different areas;
- *sensitivity* adaptability, i.e., adapting from data-poor to data-rich scenarios.

The number of satisfied scalability features is proportional to the predictions' robustness.

As for the models and processes, these should also be oriented to satisfying the research questions. Models should include (i) optimisation algorithms (e.g., operation research, simulated annealing) for resource conservation and optimisation; (ii) AI models (e.g., Machine- and Deep-learning, Maximum Entropy, and hybrid mechanistic correlative models) to simulate ecosystem functions and produce information for enhancing and speeding up traditional EMs (e.g., Ecopath with Ecosim); (iii) AI models combined with oceanographic and environmental forecast models (e.g., DIVA, ECHAM5) and signal processing models (e.g., Singular Spectrum Analysis, Fourier Analysis) for long-term projections; (iv) state-space models (e.g., Markov Chain Monte Carlo methods) and unsupervised models to unfold natural complex relations between the data; (v) feature selection and model sensitivity analysis to identify the data carrying the highest information amount; (vi) multidisciplinary AI approaches to fill data gaps;
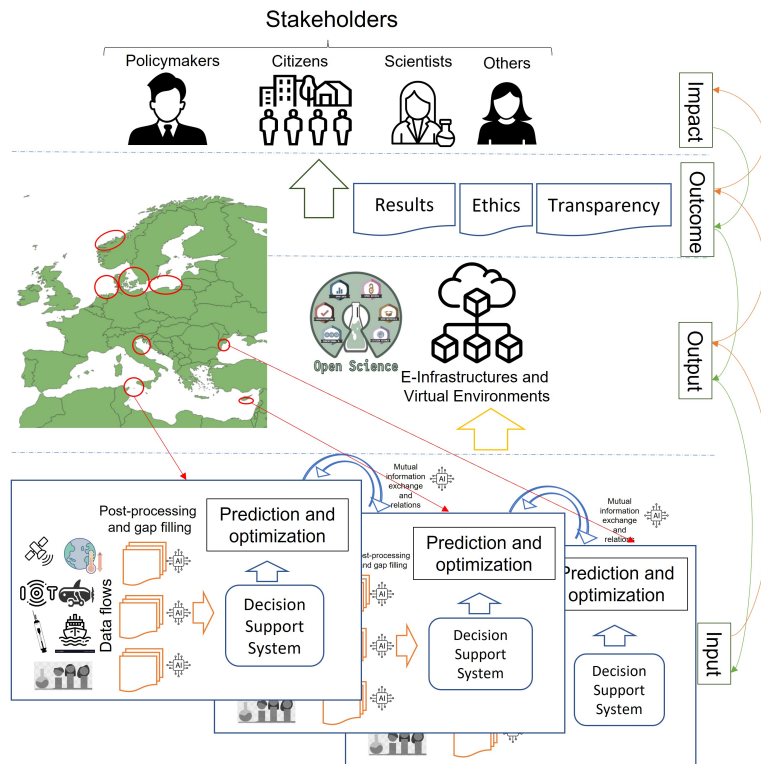
Figure 1: Conceptual schema of a multi-area *h*-DTO.

(vii) Open Science model and data reusability platforms to create interconnections between different DSSs.

4. Conclusions

The *h*-DTO is a novel, dynamic paradigm to unfold the understanding of essential ecosystem functions, the interactions between the driving forces acting on a marine area, and overall make oceans more understandable and predictable. The research questions addressed are indeed classic, unsolved questions in ecosystem modelling. A prototype *h*-DTO could be built on small PMAs guaranteeing easier control of the mathematical, numerical, and data interoperability challenges. The information available for each area should include data on marine biodiversity, cultural heritage, ship traffic, fishing and overfishing, plastic, artificial and organic waste, and climate change effects. The uncertainty of the DSSs should decrease with the increase in the completeness of the data flows. Multidisciplinarity is integral to developing an *h*-DTO because designing all components requires combining different skills from different disciplines. The realisation of an *h*-DTO project alone demonstrates the added value of multidisciplinary research actions for meeting environmental and societal challenges and improving the current ecosystem understanding at the levels of data, ecosystem functions, spatial planning, and predictions.

REFERENCES

1. Piante, C. and Ody, D. Blue growth in the mediterranean sea: the challenge of good environmental status, *MedTrends Project. WWF-France*, p. 192, (2015).

2. Campana, E. F., Ciappi, E. and Coro, G. The role of technology and digital innovation in sustainability and decarbonization of the blue economy, *Bulletin of Geophysics and Oceanography*, **123**, (2021).

3. Coro, G., Ellenbroek, A. and Pagano, P. An open science approach to infer fishing activity pressure on stocks and biodiversity from vessel tracking data, *Ecological Informatics*, **64**, 101384, (2021).