A description of Validation Processes and Techniques for Ocean Forecasting

Marcos Garcia Sotillo[1], Marie Drevillon[2], Fabrice Hernandez[3]

[1]Nologin Oceanic Weather Systems, Santiago de Compostela, Spain
[2]Mercator Ocean International, Toulouse, France
[3]LEGOS, Institut de Recherche pour le Développement, Toulouse, France

Correspondence to: Marcos Garcia Sotillo (marcos.sotillo@nowsystems.eu)

Abstract. Operational forecasting systems architecture requires clear identification of best practices for assessing the quality of ocean products: it plays a key role not only for the qualification of predictions skill, but also for advancing in the scientific understanding of the ocean dynamics from global to coastal scales. The Authors discusses on the role of observing network for performing validation of ocean model outputs, identifying current gaps (i.e., different capacity in assessing physical essential ocean variables versus biogeochemical ones), but also emphasizing the need of new metrics (tailored for end-users comprehension and usages). An analysis on the level of maturity of validation processes from global to regional systems is provided. A rich variety of approaches exist, and the most we move towards the coast the higher is the complexity in calculating such metrics, due to increased resolution but also somehow limited by the lack of coastal observatories worldwide. It is provided as example how the Copernicus Marine Service currently organizes the product quality information from producers (with dedicated scientific documentation, properly planned and designed) to end-users (with publication of targeted estimated accuracy numbers for its whole product catalogue).

1 Introduction

Product quality assessment is a key issue for operational ocean forecasting systems (OOFS). There is a long tradition in scientific research related to model validation and, through coordinated community initiatives there have been in recent times important progress on this field related to operational oceanographic services (Hernandez, 2015; 2018).

Strong efforts to define operational oceanography's best practices have started; among others: the Ocean Best Practices (Pearlman, et al., 2019 and https://www.oceanbestpractices.org/) and the Guide on Implementing Operational Ocean Monitoring and Forecasting Systems delivered by ETOOFS (Expert Team on Operational Ocean Forecasting Systems; Alvarez-Fanjul et al., 2022). In the latest ETOOFS guide, several sections are dedicated to model validation: i.e.: Section 4.5 on Validation and Verification, and sub-sections on Validation Strategies for ocean physical models (Section 5.7), Sea ice models (Section 6.2.6), storm surge (Section 7.2.6), wave models (Section 8.7) and biogeochemistry models (Section 9.2.6) or the specific section (Section 12.9) on quality assessment for intermediate and end users.

The main goal of this paper is to describe the status of the validation of ocean forecasting products. In Section 2, it is discussed the crucial role that observational data sources plays in the validation of ocean models, and how identified gaps in the observations determine model validation processes, limiting them for some essential ocean variables in some temporal scales and on specific zones (i.e., on shelf and in the coastal zone). An analysis on the level of maturity of validation processes applied by OOFS is provided in Section 4; Operational processes implemented in the Copernicus Marine service for product quality across global and (European) regional model systems are analysed in Section 4.1; whereas Section 4.2 provides a view of model validation approaches applied by (non-european) regional and coastal operational services around the world. Finally, conclusions are delivered in Section 5.

2 Observations for model validation

The lack of observations is the primary, and obvious, difficulty to validate an OOFS on a specific site. In that sense, it is very difficult to overcome observational gaps and, if they exist, OOFS validation processes are seriously hindered by them. Validation is a global necessity and challenge. Capet et al. (2020) provides a complete overview and mapping of the current European capability in terms of OOFSs, including contributions from 49 organizations around Europe about 104 operational model systems, simulating mostly hydrodynamics, biogeochemistry and sea waves. This contribution shows how, and to what extent, different observational data sources are used for model skill assessment. As shown in Figure 1, most of the model validation systems mainly use fixed platforms, satellite remote sensing and coastal tide gauges.
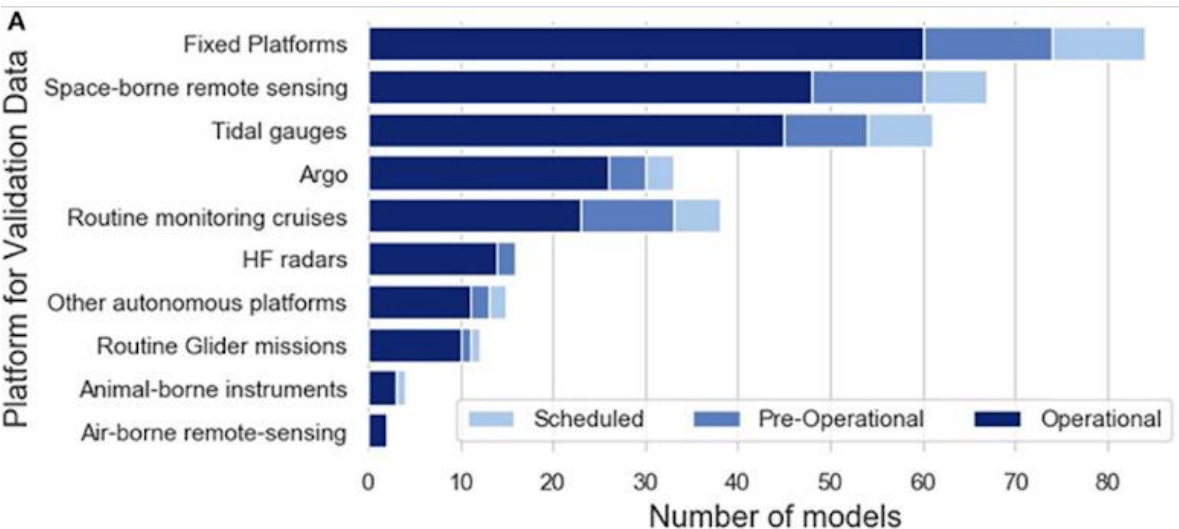
Figure 1: (from Capet et al. 2020). Observing platforms providing data used for model skill assessment and validation purposes; and number of models (from the EuroGOOS survey that use them).

It is important to note that the aggregate results of the study do not provide differences between basin/regional systems and the more coastal ones. Indeed, in this contribution most of the Near-Real-Time systems of the Copernicus Marine regional
55   Monitoring and Forecasting Centres are included, making that some observational data sources not so coastal oriented such as the Argo will be used by a high number of European OOFSs. The same may happen with the use of space-borne remote sensing products, which is more limited in its use for validation OOFS as we move to more limited small coastal model domains.
Use of satellite products for OOFS validation is common in the case of global, basin and regional systems, but limited in the case of coastal ones. If used, this is done mainly by those coastal systems that present a bigger spatial geographical coverage
60   (going beyond the shelf break). Furthermore, new incoming observational technologies (i.e., the new Sentinels missions, swath altimetry, HF Radars, BGC-Argo, etc.) and opportunities to use of new coastal observing systems (links with Member State networks and/or specific R&D projects) will enhance model validation capacities. New validation tools may also be developed for coordinated Observing System Experiments (OSEs) and Observing System Simulation Experiments (OSSEs) related to the optimization of these observation networks. Taking advantage of these OSSEs framework, AI emulated variables will be
65   developed which will increase validation capacities. Increased awareness on the need for enhancing observing networks, bringing new initiatives and efforts to better integrate existing ocean observing systems with the OOFS validation processes is needed.

3 OOFS Validation: A matter of EOVs

In terms of ocean model validation, there is a different level of application depending on which Essential Ocean Variable
70   (EOV) is targeted. The Copernicus Marine Service, a comprehensive multi-product service dealing with more than 150 operational products and involves more than 60 EOVs for the blue, green and white ocean, can illustrate such differences across EOVs. The document that provides the terms of reference for all the Product Quality (PQ) assessment done within the Service and the long-term strategy for the PQ enhancement (Copernicus Marine PQ Strategic Plan; Sotillo et al.; 2021), includes in its analysis of strengths and weaknesses the following points about the different level of maturity in terms of model
75   validation across EOVs: s:
The physic blue world versus the green biogeochemistry component. The assessment of physical parameters is more developed than the one for biogeochemistry parameters; The Copernicus Marine Service identifies the need of special efforts for biogeochemical model product validations. The lack of biogeochemistry observations conditions not only the biogeochemical model validation, but even the modelling itself. Due to the lack of in situ data, some phenomena such as primary production
80   and bloom of phytoplankton are most of the time assessed using chlorophyll Ocean Colour satellite data, which have some limitations related to coverage and resolution, especially for the coastal zones. Furthermore, it is necessary to also assess in biogeochemical models the factors that cause these blooms (i.e., transport of nutrients). Carbon, oxygen, and ocean acidification are parameters of interest both at regional and global scales that need a better validation. BGC-argo floats can enhance the monitoring, but mostly off shelves and far from coastal areas. Finally mention that in the bio model validation, it

85   is important to evaluate together the errors in the physical system, particularly vertical transport and mixing which strongly impact the coupled biogeochemical models. Thus, a monitoring of errors on key parameters of the physical forcing should help to characterize causes of errors of biogeochemical products.

- Sea ice concentration, due mainly to observation by satellite, is assessed and brings to sea ice extent, sea ice drift, sea ice thickness and sea ice edge validation. New validation metrics (some related to end-user needs) should be developed for sea ice temperature and icebergs concentration maps, and specific assessments on interannual time scales, multi-year sea ice parameters need to be specifically addressed.

- Water temperature. Sea surface temperature is the most used EOV, being the most monitored parameter, and usually assessed with (in-situ and remotely sensed) multi-product approaches, considering regional specificities (for high frequency products, particular attention should be paid to diurnal cycle and tidal mixing effects). Generally, validation on surface layers is privileged with respect to the rest of layers across the water column, existing a clear decreasing gradient towards deeper levels. Availability of in-situ observations has greatly improved since the 2000s with the Argo program. At depth, T/S data are the most used observations in Product Quality assessment. However, at synoptic scales, water mass distribution stays partially sampled in the upper ocean. There are significant regional differences, not always being the coastal areas the privileged ones (indeed, the autonomous Argo measure network changed the usual fact of coastal and on-shelf areas being traditionally the more sampled).

- In the case of salinity, in-situ measurements from fixed moorings, ARGO drifters or offshore, coastal profiles with CTD or XBT instruments, and surface transects with thermo-salinometer are the most common data sources used for OOFS model validation. Averaged maps of sea surface salinity derived from remotely sensed satellite data (such as the SMOS ones) can be used to validate models, especially far from coastal areas.

- The approach to regionally validate sea level model solutions is based on comparison to satellite altimetry, at the scales of interest, from open ocean to coastal dynamical responses. Enhancement of sea level validation in coastal and on-shelf areas is needed and preparation for the use of the new wide swath altimetry products should be done in following years. On the other hand, comparisons of coastal OOFS model products with in-situ sea-level measurements from tide-gauge are quite common. External metrics linked to Storm Surge services (including total sea levels, tidal solution, and residuals) are considered. For many coastal forecast systems, especially for those with more limited spatial coverages, the comparison of the simulated sea level with local observation from a tide-gauge usually installed in ports and the unique NRT ocean measurement available, is the only feasible direct model-observation comparison.

- Ocean currents and associated transports, especially near surface, are parameters with strong impact in many applications. Their assessment is usually done using independent observations (as most of today's systems do not assimilate this kind of observations). To this aim, in-situ observation from current meters and ADCPs, installed at mooring buoy stations as well as remoted sensed data from coastal HF radar systems are used to validate simulated currents at specific locations. The proliferation at global and regional scales of surface velocity products derived from drifters' observations as pseudo-eulerian estimated maps are also contributing to the validation of ocean models.

Finally, Satellite altimetry can also be used to assess geostrophic/non geostrophic properties of the ocean, and some
120   derived estimation of currents from satellite SAR, or from SST/Ocean colour images, can also be used on specific
areas.

4 Operational Validation: status across different OOFSs

A discussion on the status of operational validation across different existing OOFS systems is here provided. There are
significant differences in the status of the operational validation procedures applied by global/basin/regional systems and the
125   ones applied by the coastal services.

To illustrate how operational validation is being performed by basin and regional OOFSs, Section 4.1 provides as example the
approach followed by the Copernicus Marine service. In this service outcomes from the validation of several global/regional
models contributes to generate a variety of product quality information across products that is delivered to users

On the other hand, in the case of more localized national/coastal OOFSs, there is a variety of model validation approaches.
130   Section 4,2 reviews them, providing for the European space information from the EuroGoos coastal model capability mapping,
and different examples from systems located all over the world (including North and South America, Africa and Asia).

4.1 Validation of global/basin/regional model systems: The Copernicus Marine example

The Copernicus Marine Service (, Le Traon et al., 2019), deliver a consistent, reliable and state-of-the-art information derived
both from space/in-situ observations and from models – including forecasts, analyses and reanalyses – on the physical and
135   biogeochemical state over the global ocean and the European regional seas. As stated in previous section, the extensive multi-
product portfolio offered, comprising more than 150 operational products and involving more than 60 EOVs for the blue, green
and white ocean, established this Copernicus Marine service as a benchmark in operational oceanography. The Service relies
on a network of producers, interconnecting several European OOFS at global and regional scales: 7 Monitoring and Forecasting
Centres (MFC) run ocean numerical models, assimilating data, to generate reanalyse, analyse and 10-day forecasts of the
140   ocean.

Model validation in Copernicus Marine is closely linked to the operational production performed at the OOFS level. This
connection spans all service phases, from design to the operational delivery of products, including associated communication
and training activities. Furthermore, a scientifically sound and effectively communicated Product Quality assessment stands
as one of the key cross-cutting functions of the Marine Service, Further, details on its achievements during the first phase of
145   the service can be found in Sotillo et al., 2022

Individual OOFS, producers of the regional components of the Marine Service, daily verify the scientific quality of their model
products (i.e., NRT forecast/analysis and MY reanalysis), using quantitative validation metrics, described in standard protocols
and plans, and using extensively any available observational data sources, as referred in previous sections. Regular updates of
a subset of the validation metrics assessed by the own producers, including Class2 validation of model products at mooring

150   sites and Class4 regional validation metrics, is made available to end-users through a dedicated website (the Copernicus Marine Product Quality Dashboard; pqd.mercator-ocean.fr).

Furthermore, the Marine Service is responsible for informing end-users about relevant PQ information in a transparent way. To this aim, reference scientific PQ documentation is issued for each delivered product. These documents, stating the expected quality of a product by means of validation metrics computed along the qualification phase of the new model system, are

155   updated for every quality change associated with any new operational release.

The Copernicus Marine service model production needs to be carefully monitored at each step, and then, quality of any upstream data used in the model runs shall be properly assessed (even if such upstream data is quality controlled by the data providers). Indeed, regular exchanges are organized between observations and model producers within the service to discuss data assimilation and validation issues. Scientific quality is one of the key performance indicators for the OOFS, and producers

160   report quarterly to the Service on quality monitoring activities. Any change affecting model solutions required to be justified from a Product Quality perspective.

The consistency in the choice of model validation metrics, and in the way they are presented, can be important because it makes easier for users to understand the Product Quality information provided across products and to browse in the service portfolio which are fit-for-purpose products. However, given the wide range of Copernicus Marine products and production

165   methods, it is not always scientifically meaningful to provide same type of information across products and for all involved systems. The Product Quality cross-cutting strategy (Sotillo et al., 2021) thus aims to strike a balance between the level of homogeneity of the information delivered and its relevance. Indeed, the Copernicus Marine Service is a first achievement towards the interconnection of operational oceanography services at basin scale, and digital ocean platforms based on cloud technology will enable new validation capacities, facilitating the set-up of dynamic uncertainty for most of their products. The

170   frequency of the updates will also increase to better serve coastal OOFS, where short-term forecast and quality information should be delivered on a daily (preferred) or weekly basis.

4.2 Validation of coastal OOFS: a world of variety

There is not a common operational validation approach in coastal OOFSs, and the degree of operationality for the model validation is highly dependent on the type of forecasting system set-up (i.e. system with data assimilation scheme activated,

175   generating analysis or on the other hand, OOFS based on a free forecast model system), the extension of the area of interest (being different for very limited coastal systems or going into a larger regional extent), the service purpose (system targeted on a primary end-user with specific interests or needs, or if the OOFS delivers a general multi-parameter/purpose service) and finally in the availability, and degree of operational access, to local in-situ observational data sources.

Most of the OOFS have some system validation. Even those models used for research purposes or in the process of maturing

180   their operationality (pre-operational state) have some kind of model validation, typically the early stages of the model set-up configuration, often running in hindcast mode on specific time periods to take benefit of existing observational data campaigns. In pre-operational systems, or in early stages of OOFS services, an operational validation system is not so common, while

model providers are working on the configuration of operational processes for an automatic PQ model assessment, and meanwhile, maintaining some offline model validation (using available observations or focused on specific targeted periods when outstanding events occurred, or when observational campaigns are available).

It is worthy to note that Capet et al. (2020), concludes that only 20% of models provide a dynamic uncertainty together with the forecasted EOVs, which would be required for a real-time provision of confidence levels associated with the forecasts (e.g., as is usual for instance in weather forecasts). Usually, model providers perform operational and off-line validations, focusing mostly on the best estimate solution, and not so on the forecast skill assessment; scientific statistical metrics are computed using available in-situ observational data sources from their own networks or external observational data providers (using observational products from core services such as Copernicus, other national/regional/local public providers or from the industry, if available); in coastal high resolution systems, with quite limited geographical domains, the use of satellite data is not so common for model validation due to both the lack of remotely sensed product coverage and the higher uncertainty of remotely sensed coastal data. This is the case of many OOFS all over the world. For instance, the South African ocean forecast system (SOMISANA-Sustainable Ocean Modelling Initiative: A South African Approach; https://somisana.ac.za/) delivers downscaling of global model products for specific coastal applications in key coastal areas. In these cases, scientific model validation is mostly done offline by model producers, comparing their best estimated hindcast solutions with the existing historic observations. Given the coastal nature of the models, the validation process is focussed mainly on coastal moorings with the model domains. These include coastal temperature recorders, bottom mounted thermistor strings, wire-walker moorings and ADCP moorings from previously published datasets (e.g. Lucas et al., 2014, Pitcher et al., 2014, Goschen et al., 2015) as well as unpublished datasets from local institutes. Currently, there is not a direct transfer of information about the product quality from the service to the OOFS users, neither computation of forecast skill assessment nor end-user-oriented metrics. However, some interesting initiatives, mostly linked with the engaging of stakeholder and product disseminations through end-user services platforms, are on-going and in the SOMISANA OOFS roadmap is included the implementation of an operational validation protocol, including forecast assessment.

The most common situation is that model validation is performed by the OOFS providers themselves. However, in some cases (usually targeted services) there may be options for some external validation, performed not by the provider itself, but directly by the targeted end-user(s). This is the case of the DREAMS service (Hirose et al., 2013; 2021) in the west Japan coast, where model solutions are validated directly by the end-users of the service: in this case, fisheries, through a program with fishing boats as ship of opportunity (Ito et al., 2021). As DREAMS model provider states: "The fishermen watch carefully the coastal ocean prediction to meet better catches. They are inevitably the serious users to claim the quality of prediction".

In the case of the Brazilian REMO service (Lima et al., 2013; Franz et al., 2021), the validation is done in house only for targeted end-users; either by the Navy or by the PETROBRAS oil company teams. On the PETROBRAS side, they have several current meters sites where they compare in situ measurements with not only the REMO forecast, but also from all the other available Ocean Forecasting Systems that deliver forecasts on that given day. At the Navy side, they do several validations that include the thermohaline structure, the Taylor Diagrams for few properties as well as the transport for the

Brazil Current and the tidal analysis of both level and currents where they have data available. Furthermore, there can be very high-resolution coastal OOFSs that can be implemented for specific purposes, running only along designated periods, to provide model data as input, for instance, during the design and construction phases of large infrastructures. In such cases, the implementation of the specific model solution can go together with some monitoring activity in the targeted area , allowing some model validation throughout the construction phase and after operations commence. In this type of services, modelling and validation are typically done in-house, with products and results rarely being publicly disseminated, not contributing to the literature.

There are coastal systems, that have big domains (going into regional) and that may include data assimilation schemes, or pure forecast local coastal systems (run by providers of regional/basin systems in which the local systems are nested) that tend to have operational validation systems (taking benefit of the extensive use of the observational data sources done for assimilation purposes). There are examples of OOFSs supported by state agencies, such as the Canadian Government CONCEPTS (Canadian Operational Network for Coupled Environmental Prediction Systems) that develops and operates a hierarchy of OOFS. These include the whole downscaling approach: going in this Canadian case from the Global Ice Ocean Prediction System (GIOPS; Smith et al., 2016) used to initialize coupled deterministic medium range predictions (Smith et al., 2018), as well as ensemble predictions (Peterson et al., 2022). The global system provides boundary conditions to the Regional Ice-Ocean Prediction System (RIOPS; Smith et al., 2021), which in turn provides boundary conditions and nudging fields for a Coastal Ice Ocean Prediction System (CIOPS; Paquin et al., 2024). Recently, six port-scale prediction systems have also been put in place. Paquin et al. 2019 presented the prototype of the mentioned port models, whereas Nudds et al. (2020) presented the initial intercomparison projects that took place to compare the NEMO model implementation described in Paquin et al. (2019) with an unstructured model implementation using FVCOM. CONCEPTS also develops and operates deterministic and ensemble wave and storm surge prediction systems. Proposed changes to these systems must follow a set of formalized verification standards. Evaluation of forecast skill as a function of lead time is also made. Monitoring systems are also in place to ensure the quality of real-time analyses. Forecasts are evaluated in near real-time as part of the OceanPredict Class4 intercomparison activity (Ryan et al., 2015), and evaluations are predominantly made against available observations, but also include comparison to analyses for the longer-range coupled forecasts. These include assimilated satellite (sea level anomaly, sea surface temperature, sea ice concentration) and in situ observations (Argo, buoys, moorings, gliders, field campaigns, etc.). Additional independent evaluations are made against tide gauges, ADCP, HF Radar, drifters, ice beacons (Chikhar et al., 2019) and estimates of sea ice and snow thickness. Evaluations are also made of transports across reference sections and of surface fluxes (both against observations as well as in terms of budgets, e.g., Roy et al. (2015) and Dupont et al. (2015)). Finally, user-relevant verification is made in terms of sea ice (e.g. probability of ice, ice formation and melt dates) and ocean (e.g. eddy identification and properties) features (Smith and Fortin, 2022). An ongoing effort is underway to quantify unconstrained variability in the systems and to provide uncertainty estimates to users.

There are also coastal OOFSs delivered by national agencies or organisms that run their own observational networks. This is the case in Spain of ocean model systems from different state and regional government agencies: i.e., Puertos del Estado

(SAMOA; Sotillo et al., 2019 and García-León et al., 2022), SOCIB (WMOP; Mourre et al., 2018), MeteoGalicia (MG; Costa et al., 2012) or the case of the Marine Institute ocean forecasting systems for Ireland (Nagy et al., 2020) with coastal systems focused on very limited -highly monitored bay areas. In these cases, usually they take advantage of synergies of the combination of high-resolution model solutions and operational observational data sources (being the in-situ operational observational capacity developed by running operational networks or through the sustained periodic measurement at fixed stations) progressing towards more operational validation procedures. Even in these optimal cases, operational validation is mainly limited to model best estimate solutions, and generation of end-user metrics or uncertainty estimation is still missing, but in the long-term evolution roadmaps.

5 Summary, conclusions and Outlook.

This paper reviews the status in the validation of operational ocean forecasting products. Recent advancements in the field of operational oceanographic services have significantly contributed to scientific research on model validation. This is achieved by the OOFS individually, but also through coordinated efforts, such as those developed within Copernicus Marine and other international initiatives, like the evolution of GODAE to OceanPredict.

The crucial role of observations in ocean model validation is discussed, highlighting how gaps in observational capabilities significantly impact the validation processes in OOFS. These limitations particularly affect the validation of essential ocean variables at specific temporal scales and in certain regions, such as shelf areas and coastal zones. Most model validation systems primarily rely on observations from fleets of floats, drifters, fixed in-situ mooring platforms, coastal tide gauges, and satellite remote sensed data products. It is pointed how there are notable differences between the validation of global/basin systems and more coastal-focused ones. Some observational data sources, such as ARGO, are crucial for validating global, basin, and regional systems, however, are less relevant for coastal systems due to coverage limitations. Similarly, while satellite products are commonly used for validating global, basin, and regional OOFS, their use is more constrained in coastal OOFS with much smaller coastal model domains.

Across EOVs there are also significant differences. The assessment of physical parameters is more developed than the one for biogeochemistry parameters; Being the lack of biogeochemistry observations certainly a shortcoming for the validation of biogeochemical models. Generally, it is seen that model validation tends to prioritize surface layers over the rest of the water column. Likewise, there are significant regional differences, and coastal areas are not always the most prioritized. Indeed, the ARGO network shifted the traditional favored focused on coastal and on-shelf areas to open ocean. Among the physical EOVs, temperature is the most widely sampled and therefore validated. Ocean current, particularly the near surface one, is a critical parameter for many applications, however, it is less well-monitored and then less validated. Recently, the scatter in-situ monitoring has been reinforced locally in coastal zones with HF radar systems. For sea level, the regional modeling solutions are typically validated with satellite altimetry, while comparisons of coastal OOFS model products and in-situ sea-level

measurements from tide gauges are also quite common. Lastly, simulated sea ice parameters are primarily validated with satellite remote sensing.

285 An analysis of the maturity of validation processes from global to regional forecasting systems is presented, using the approach followed by the Copernicus Marine Service as an example. This service connects more than seven regional production centers, that run models for ocean physics, including sea ice and wave modelling systems, as well as biogeochemistry. It delivers forecast and reanalysis products for several EOVs, ensuring homogenized product quality information across the entire range. The Copernicus Marine Service organizes product quality information from producers, providing dedicated scientific product

290 quality documentation that is well-planned and designed. This PQ information is then communicated to end-users: every product in the Copernicus Marine product portfolio is accompanied by the relevant product quality documents, and on-line publication of updated estimated accuracy values for the entire product catalogue is also ensured through the Copernicus Product Quality Dashboard.

A wide variety of approaches exist as OOFS work closer to the coast. For high-resolution coastal models with very limited

295 geographical domains, the complexity of calculating validation metrics increases. This is due to the need of higher resolution to validate local processes, but also operational validation is often constrained by the scarcity of near-real-time coastal observations. The review presents examples of model validation approaches used by regional and coastal operational services worldwide, particularly from outside Europe, to complement the European Copernicus approach described earlier. Detailed examples of OOFS from Canada, Brazil, South Africa, and Japan are also included. The case of coastal OOFSs delivered by

300 national agencies or organizations that operate their own observational networks is also highlighted as successful examples of operational model validation. These OOFS benefit from synergies between high-resolution model solutions and operational observational data sources, advancing towards more robust operational validation procedures. However, even in these optimal cases, operational validation in most coastal OOFS is primarily limited to the validation of best-estimate model solutions, typically on a daily basis at best.

305 Looking ahead, uncertainty estimation of OOFS products is identified as a key focus and is included in the long-term evolution roadmap of services like Copernicus Marine. The operational delivery of end-user tailored metrics is still largely lacking, with this being more feasible in coastal OOFSs targeted and co-designed with specific end-user purposes in mind (e.g., services for ports or support for specific activities, such as aquaculture) than in regional/basin/global systems. New observational technologies (e.g., the upcoming Sentinel missions, swath altimetry, HF radars, BGC-Argo, etc.) and the opportunities

310 presented by new coastal observing systems (through links with Member State networks and/or specific R&D projects) will enhance model validation capabilities. Improvements in sea level validation in coastal and on-shelf areas are expected using new wide-swath and higher-frequency altimetry products in coming years. Finally, the integration of operational validation tools with future Observing System Experiments (OSEs) and Observing System Simulation Experiments (OSSEs) aimed at optimizing observation networks could provide significant benefits. Leveraging these OSSE frameworks, AI-derived emulated

315 variables may be developed, enhancing validation capacities. Overall, increasing awareness and fostering new initiatives to better integrate existing ocean observing systems with OOFS validation processes will be a key focus for the future.

References

Alvarez Fanjul, E., Ciliberti, S., Bahurel, P.: Implementing Operational Ocean Monitoring and Forecasting Systems. IOC-UNESCO, GOOS-275. https://doi.org/10.48670/ETOOFS, 2022.

Capet, A., Fernández, V., She, J., Dabrowski, T., Umgiesser, G., Staneva, J., Mészáros, L., Campuzano, F., Ursella, L., Nolan, G., and El Serafy, G.: Operational Modeling Capacity in European Seas - An EuroGOOS Perspective and Recommendations for Improvement. Frontiers in Marine Science, 7:129. https://doi.org/10.3389/fmars.2020.00129, 2020.

Chikhar, K., Lemieux, J.F., Dupont, F., Roy, F., Smith, G.C., Brady, M., Howell, S.E. and Beaini, R.: Sensitivity of Ice Drift to Form Drag and Ice Strength Parameterization in a Coupled Ice–Ocean Model. Atmosphere-Ocean, 57(5), 1-21. https://doi.org/10.1080/07055900.2019.1694859, 2019.

Costa, P., Ǵomez, B., Venancio, A., Ṕerez, E., Munuzuri, V.: Using the Regional Ocean Modeling System (ROMS) to improve the sea surface temperature from MERCATOR Ocean System. Advanced in Spanish Physical Oceanography. Sci. Mar. 76 S1 (2012). https://doi.org/10.3989/scimar.03614.19E, 2012.

Dupont, F., Higginson, S., Bourdallé-Badie, R., Lu, Y., Roy, F., Smith, G.C., Lemieux, J.-F., Garric, G., Davidson, F.: A high-resolution ocean and sea-ice modelling system for the Arctic and North Atlantic Oceans. Geosci. Model Dev., 8, 1577-1594. https://doi.org/10.5194/gmd-8-1577-2015, 2015.

Franz, G., Garcia, C. A. E., Pereira, J., Assad, L. P. F., Rollnic, M., Garbossa, L. H. P., Cunha, L. C., Lentini, C. A. D., Nobre, P., Turra, A., Trotte-Duhá, J. R., Cirano, M., Estefen, S. F., Lima, J. A. M., Paiva, A. M., Noernberg, M. A., Tanajura, C. A. S., Moutinho, J. M., Campuzano, F., Pereira, E. S., Lima, A. C., Mendonça, L. F. F., Nocko, H., Machado, L., Alvarenga, J. B. R., Martins, R. P., Böck, C. S., Toste, R., Landau, L., Miranda, T., Santos, F., Pellegrini, J., Juliano, M., Neves, R., Polejack, A.: Coastal Ocean Observing and Modeling Systems in Brazil: Initiatives and Future Perspectives. Frontiers in Marine Science, 8, 681619. https://doi.org/10.3389/fmars.2021.681619, 2021.

García-León, M., Sotillo, M. G., Mestres, M., Espino, M., & Fanjul, E. Á. (2022). Improving Operational Ocean Models for the Spanish Port Authorities: Assessment of the SAMOA Coastal Forecasting Service Upgrades. *Journal of Marine Science and Engineering*, *10*(2), 149. https://doi.org/10.3390/jmse10020149

Goschen, W., Bornman, T., Deyzel, S., & Schumann, E. (2015). Coastal upwelling on the far eastern Agulhas Bank associated with large meanders in the Agulhas Current. Continental Shelf Research, 101 , 34–46. doi:10.1016/j.csr.2015.04.004

Hernandez F, Blockley, E., Brassington, G.B., Davidson, F., Divakaran, P., Drévillon, M., Ishizaki, S., Garcia-Sotillo, M., Hogan, P.J., Lagemaa, P., Levier, B., Martin, M., Mehra, A., Mooers, C., Ferry, N., Ryan, A., Regnier, C., Sellar, A., Smith, G.C., Sofianos, S., Spindler, T., Volpe, G., Wilkin, J., Zaron, E.D., and Zhang, A.: Recent progress in performance evaluations and near real-time assessment of operational ocean products. Journal of Operational Oceanography, 8:sup2, s221-s238. https://doi.org/10.1080/1755876X.2015.1050282, 2015.

Hernandez, F., Smith, G., Baetens, K., Cossarini, G., Garcia-Hermosa, I., Drevillon, M., etc. and von Schuckman, K.:
350   Measuring performances, skill and accuracy in operational oceanography: New challenges and approaches. New Frontiers in
Operational Oceanography, 759-795. DOI: 10.17125/gov2018.ch29, 2018.

Hirose, N,, Takayama, K., Moon, J.H.,Watanabe, T., Nishida, Y.: Regional data assimilation system extended to the East Asian
marginal Seas. Sea and Sky, 89(2), 43-51, 2013.

Hirose, N., Liu, T., Takayama, K., Uehara, K., Taneda, T., Kim, Y.H.: Vertical viscosity coefficient increased for high-
355   resolution modeling of the Tsushima/Korea Strait. Journal of Atmospheric and Oceanic Technology, 38 (6), 1205-1215.
https://doi.org/10.1175/JTECH-D-20-0156.1, 2021.

Ito, T., Nagamoto, A., Takagi, N., Kajiwara, N., Kokubo, T., Takikawa, T., Hirose, N.: Construction of CTD hydrographic
observation system by fishermen in northwest Kyushu Japan. Bull. Jpn. Soc. Fish. Oceanogr., 85(4), 197-203 (in Japanese;
with English abstract). https://doi.org/10.34423/jsfo.85.4_197, 2021.

360   Le Traon, P.-Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando,
V.E., Kreiner, M.B., Benkiran, M., Carval, T., Ciliberti, S.A., Claustre, H., Clementi, E., Coppini, G., Cossarini, G., De Alfonso
Alonso-Muñoyerro, M., Delamarche, A., Dibarboure, G., Dinessen, F., Drevillon, M., Drillet, Y., Faugere, Y., Fernández, V.,
Fleming, A., Garcia-Hermosa, M.I., Sotillo, M.G., Garric, G., Gasparin, F., Giordan,C., Gehlen, M., Gregoire, M.L., Guinehut,
S., Hamon, M., Harris, C., Hernandez, F., Hinkler, J.B., Hoyer, J., Karvonen, J., Kay, S., King, R., Lavergne, T., Lemieux-
365   Dudon, B., Lima, L., Mao, C., Martin, M.J., Masina, S., Melet, A., Buongiorno Nardelli, B., Nolan, G., Pascual, A., Pistoia,
J., Palazov, A., Piolle, J.F., Pujol, M.I., Pequignet, A.C., Peneva, E., Pérez Gómez, B., Petit de la Villeon, L., Pinardi, N.,
Pisano, A., Pouliquen, S., Reid, R., Remy, E., Santoleri, R., Siddorn, J., She, J., Staneva, J., Stoffelen, A., Tonani, M.,
Vandenbulcke, L., von Schuckmann, K., Volpe, G., Wettre, C., and Zacharioudaki, A.: From Observation to Information and
Users: The Copernicus Marine Service Perspective. Front. Mar. Sci., 6. https://doi.org/10.3389/fmars.2019.00234, 2019.

370   Lima, J. A. M., Martins, R. P., Tanajura, C. A. S., Paiva, A. M., Cirano, M., Campos, E. J. D., Soares, I. D., França, G. B.,
Obino, R. S., Alvarenga, J. B. R.: Design and implementation of the Oceanographic Modeling and Observation Network
(REMO) for operational oceanography and ocean forecasting. Revista Brasileira de Geofísica, 31, 209-228.
http://dx.doi.org/10.22564/rbgf.v31i2.290, 2013.

Lucas A. J., Pitcher G. C., Probyn T. A., Kudela R. M. (2014). The influence of diurnal winds on phytoplankton dynamics in
375   a coastal upwelling system off southwestern Africa. Deep Sea Res. Part II: Topical Stud. Oceanogr. 101, 50–62. doi:
10.1016/j.dsr2.2013.01.016Mourre B., Aguiar E., Juza M., Hernandez-Lasheras J., Reyes E., Heslop E., et al.: Assessment of
high-resolution regional ocean prediction systems using muli-platform observations: illustrations in the Western
Mediterranean Sea. In "New Frontiers in Operational Oceanography," Eds E. Chassignet, A. Pascual, J. Tintoré, and J. Verron,
GODAE Ocean View, 663-694. https://doi.org/10.17125/gov2018.ch24, 2018.

380   Nagy H, Lyons K, Nolan G, Cure M, Dabrowski T.: A Regional Operational Model for the North East Atlantic: Model
Configuration and Validation. Journal of Marine Science and Engineering, 8(9), 673. https://doi.org/10.3390/jmse8090673,
2020.

Nudds S, Lu Y, Higginson S, Haigh SP, Paquin J-P, O'Flaherty-Sproul M, Taylor S, Blanken H, Marcotte G, Smith GC, Bernier NB, MacAulay P, Wu Y, Zhai L, Hu X, Chanut J, Dunphy M, Dupont F, Greenberg D, Davidson FJM, Page F (2020)
385   Evaluation of structured and unstructured models for application in operational ocean forecasting in Nearshore Waters. J Mar Sci Eng 8(7):484. https://doi.org/10.3390/jmse8070484

Paquin, JP., Roy, F., Smith, G.C., MacDermid, S., Lei, J., Dupont, F., Lu, Y., Taylor, S., St-Onge-Drouin, S., Blanken, H., Dunphy, M., Soontiens, N. A new high-resolution Coastal Ice-Ocean Prediction System for the East Coast of Canada. Ocean Dynamics 74, 799–826 (2024). https://doi.org/10.1007/s10236-024-01634-7

390   Paquin J-P, Lu Y, Taylor S, Blanken H, Marcotte G, Hu X, Zhai L, Higginson S, Nudds S, Chanut J, Smith GC, Bernier N, Dupont F (2019) High-resolution modelling of a coastal harbour in the presence of strong tides and significant river runoff. Ocean Dyn 70:365–385. https://doi.org/10.1007/s10236-019-01334-7Pearlman, J., Bushnell, M., Coppola, L., Karstensen, J., Buttigieg, P.L., Pearlman. F., et al.: Evolving and Sustaining Ocean Best Practices and Standards for the Next Decade. Frontiers in Marine Science, 6:277. https://doi.org/10.3389/fmars.2019.00277, 2019.

395   Peterson, K.A., Smith, G.C., Lemieux, J.F., Roy, F., Buehner, M., Caya, A., Houtekamer, P.L., Lin, H., Muncaster, R., Deng, X. and Dupont, F.: Understanding sources of Northern Hemisphere uncertainty and forecast error in a medium-range coupled ensemble sea-ice prediction system. Quarterly Journal of the Royal Meteorological Society, 148(747), 2877-2902. https://doi.org/10.1002/qj.4340, 2022.

Pitcher G. C., Probyn T. A., du Randt A., Lucas A. J., Bernard S., Evers-King H., et al. (2014). Dynamics of oxygen depletion
400   in the nearshore of a coastal embayment of the southern benguela upwelling system. J. Geophys. Res.: Oceans 119, 2183–2200. doi: 10.1002/7332013JC009443Roy, F., Chevallier, M., Smith, G. C., Dupont, F., Garric, G., Lemieux, J.-F., Lu, Y. and Davidson, F.: Arctic sea ice and freshwater sensitivity to the treatment of the atmosphere-ice-ocean surface layer. J. Geophys. Res.: Oceans, 120(6), 4392-4417. https://doi.org/10.1002/2014JC010677, 2015.

Ryan, A. G., Regnier, C., Divakaran, P., Spindler, T., Mehra, A., Hernandez, F., Smith, G. C., Liu, Y. and Davidson, F.:
405   GODAE OceanView Class 4 forecast verification framework: Global ocean inter-comparison. Journal of Operational Oceanography, 8.sup1, s98-s111. https://doi.org/10.1080/1755876X.2015.1022330, 2015.

Smith, G. C., Bélanger, J. M., Roy, F., Pellerin, P., Ritchie, H., Onu, K., Roch, M., Zadra, A., Surcel Colan, D., Winter, B. and Fontecilla, J. S.: Impact of Coupling with an Ice-Ocean Model on Global Medium-Range NWP Forecast Skill. Monthly Weather Review, 146, 1157-1180. https://doi.org/10.1175/MWR-D-17-0157.1, 2018.

410   Smith, G.C. and Fortin, A.S.: Verification of eddy properties in operational oceanographic analysis systems. Ocean Modelling, 172, 101982. https://doi.org/10.1016/j.ocemod.2022.101982, 2022.

Smith, G.C., Liu, Y., Benkiran, M., Chikhar, K., Surcel Colan, D., Gauthier, A.A., Testut, C.E., Dupont, F., Lei, J., Roy, F. and Lemieux, J.F.: The Regional Ice Ocean Prediction System v2: a pan-Canadian ocean analysis system using an online tidal harmonic analysis. Geoscientific Model Development, 14(3), 1445-1467. https://doi.org/10.5194/gmd-14-1445-2021, 2021.

415 Smith, G. C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., Belanger, J. M., Skachko, S., Liu, Y., Dupont, F. and Lemieux, J.-F.: Sea ice forecast verification in the Canadian global ice ocean prediction system. Quarterly Journal of the Royal Meteorological Society, 142(695), 659-671. https://doi.org/10.1002/qj.2555, 2016.

Sotillo, M. G., Cerralbo, P., Lorente, P., Grifoll, M., Espino, M., and Sanchez-Arcilla, A. et al.: Coastal ocean forecasting in Spanish ports: the SAMOA operational service. J. Operat. Oceanography 13, 37–54.
420 https://doi.org/10.1080/1755876X.2019.1606765, 2019.

Sotillo, M. G., Garcia-Hermosa, I., Drévillon, M., Régnier, C., Szczypta, C., Hernandez, F., Melet, A., Le Traon, P.-Y.: Communicating CMEMS Product Quality: evolution & achievements along Copernicus-1 (2015- 2021). In "57th Mercator Ocean Journal Issue: Copernicus Marine Service Achievements 2015 - 2021". Available at https://marine.copernicus.eu/news/copernicus-1-marine-service-achievements-2015-2021, 2021 (last access: 28/07/2024).

425 Sotillo, M.G, García-Hermosa, I., Drevillon, M.: Copernicus Marine Product Quality Strategy Document. Available at https://marine.copernicus.eu/sites/default/files/media/pdf/2022-02/%5BAD14%5D%20CMEMS-PQ-SD-V2.0.pdf, 2022 (last access: 28/07/2024).

Competing interests

The contact author has declared that none of the authors has any competing interests.

430 **Acknowledgements**

The authors acknowledge to the COSS-TT group, and specially to Jennifer Veitch, Gilles Fearon, Naoki Hirose, Greg Smith, Mauro Cirano, and Tomasz Dabrowsky for their valuable comments and inputs for the description of validation processes across different coastal systems in the world (Section 4.2).