

# FORCOAST



Earth Observation Services For Wild Fisheries, Oystergrounds  
Restoration And Bivalve Mariculture Along European Coasts

## PROJECT DELIVERABLE REPORT

**Deliverable Number:** D5.5

**Deliverable Title:** System performance including KPIs  
and 'fit for purpose' analysis

**Author(s):** Jun She

**Work Package Number:** WP5

**Work Package Title:** Service Operationalization,  
Demonstration & Validation



FORCOAST Project Information	
<b>Project full title</b>	Earth Observation Services For Wild Fisheries, Oystergrounds Restoration And Bivalve Mariculture Along European Coasts
<b>Project acronym</b>	FORCOAST
<b>Grant agreement number</b>	870465
<b>Project coordinator</b>	Ghada El Serafy, Deltares
<b>Project start date and duration</b>	1 <sup>st</sup> November 2019, 36 months
<b>Project website</b>	<a href="https://forcoast.eu/">https://forcoast.eu/</a>

Deliverable Information	
<b>Work package number</b>	WP5
<b>Work package title</b>	Service Operationalisation, Demonstration & Validation
<b>Deliverable number</b>	D5.5
<b>Deliverable title</b>	System performance including KPIs and 'fit for purpose' analysis
<b>Description</b>	This deliverable reports on the system performance including KPIs and a fit for purpose analysis. Examples of these KPIs include targeted percentages of data availability, timeliness of data delivery, acceptable downtimes (resulting from scheduled and unscheduled maintenance), time taken to resolve possible issues, etc.
<b>Lead beneficiary</b>	DMI
<b>Lead Author(s)</b>	Jun She – DMI
<b>Contributor(s)</b>	Daniel Twigt & Luis Rodriguez– Deltares Jens Murawski – DMI Diego Pereiro – Marine Institute Geneviève Lacroix - RBINS
<b>Revision number</b>	5
<b>Revision Date</b>	31/08/2022

<b>Status (Final (F), Draft (D), Revised Draft (RV))</b>	F
<b>Dissemination level (Public (PU), Restricted to other program participants (PP), Restricted to a group specified by the consortium (RE), Confidential for consortium members only (CO))</b>	PU

<b>Document History</b>			
<b>Revision</b>	<b>Date</b>	<b>Modification</b>	<b>Author</b>
0	23/03/2022	Report structure	Jun She
1	10/08/2022	Partners contributions completed	Diego Pereiro, Genevieve Lacroix, Jun She, Jens Murawski, Daniel Twigt, Luis Rodriguez
2	23/08/2022	Revision	Jun She
3	26/08/2022	Revision	Diego Pereiro
4	30/08/2022	Revision	Luis Rodriguez
5	31/08/2022	Revision	Jun She

<b>Approvals</b>				
	<b>Name</b>	<b>Organisation</b>	<b>Date</b>	<b>Signature (initials)</b>
<b>Coordinator</b>	Ghada El Serafy	Deltares	31/08/2022	GES
<b>WP Leaders</b>	Tomasz Dabrowski	Marine Institute	31/08/2022	TD



### **PROPRIETARY RIGHTS STATEMENT**

This document contains information, which is proprietary to the FORCOAST consortium. Neither this document, or the information contained within may be duplicated, used or communicated except with the prior written permission of the FORCOAST coordinator.



## Executive Summary

The purpose of this report is to validate the service system, including both products and platforms based on Key Performance Indexes (KPIs), which are quantifiable performance indicators used to define success factors and measure progress toward the achievement of the service goals. The KPIs in FORCOAST are used to help to understand whether the FORCOAST service is successful and serving the strategic and operative goals, i.e., fit for the purposes of the users. In the scope of the FORCOAST project, not all SMs can be evaluated with KPIs since the evaluation of KPIs needs observations which cannot be obtained during the project period. Therefore, only SM-A1 “Marine conditions”, SM-A4 “Assistant for spat capture” and SM-R1 “Retrieval of contaminant sources” will make the product KPI evaluation.

The KPIs evaluated for the SM-A1 Marine Conditions cover some of the most significant high impact issues on marine physical conditions for bivalve mariculture, e.g., storm warning, storm surge warning and SST prediction. In Limfjorden, DMI forecast KPIs reached the target of the sea storm early warning and SST prediction. Results showed that forecasted extreme sea level is lower than the observed ones at the validation location Lemvig. Such a negative bias can be corrected by using a local forecast protocol.

For SM-A4 Spat capture service, the KPI is defined as arrival timing of larvae in the nurseries. The results showed that the predicted timing of larvae allows improving parameterization of the service module and the product could be used to detect spat arrival.

For SM-R1 Retrieve Sources of Contaminants, the KPI is defined as a skill score based on the separation distance between the observed (real) trajectory and the modeled (simulated) trajectory. The results showed that, for 6 out of 8 drifter trajectories, the model has a KPI  $>0.5$ , which means that the model performance on predicting surface drifting of the objects is satisfactory.

To assess the platform fit-for-purpose, different KPIs categorized in service availability, input data availability, interruptions, timeliness and platform usage are identified and showcased based on recent available data. The presented results serve as the backbone to prove the suitability of the service platform as prototype to fulfill its function towards users when it comes to information system.

The KPIs obtained for the FORCOAST service platform prototype present in section 5.3 prove the suitability of the developed system solution for the end user requirements in terms of information retrieval and delivery system. Service availability parameters show that the services are fully operational except those times that there is a maintenance or testing of a new functionality/feature, which development tends to be short. Regarding the timeliness of the service information, when setting automated runs so the user receives bulletins on schedule via Telegram, the timeliness is instantaneous due to the ability to configure the run times and frequency on demand with no limitations. Lastly, platform usage has been steadily rising in term of users until the expected summer valley period. This is expected to increase with the last efforts in user engagement, marketing and dissemination. The average retention time of 18 minutes and 51 seconds per user, and 3 minutes and 13 seconds per session makes sense since it gives the user time to access the platform, running the service of their interest, explore the results, and getting engaged towards the full service with the scheduled runs which all of it is the objective of the platform prototype.

## Table of Contents

Executive Summary	5
Table of Contents	6
1. Introduction	8
2. SM-A1 Marine conditions	9
2.1 Product KPIs for Limfjorden	9
2.1.1 Selection and definition of SM-A1 KPIs	9
2.1.2 Methods and input data for calculating the KPIs	10
2.1.3 Results	11
2.1.3.1 Storm early warning	11
2.1.3.2 Storm surge early warning	12
2.1.3.3 SST forecast	13
3. SM-A4 Assistance for spat capture	15
3.1 Selection and definition of KPIs	15
3.2 Methods and input data	15
3.3 Results	15
4. SM-R1 Retrieve sources of contaminants	20
4.1 Selection and definition of KPIs	20
4.2 Methods and input data	20
4.3 Results	22
5. Platform performance	25
5.1 Selection and definition of KPIs	25
5.2 Methods and input data	25
5.3 Results	26
5.3.1 Service availability	26
5.3.2 Input data availability	27
5.3.3 Interruptions	29
5.3.4 Timeliness	29
5.3.5 Platform usage	30
6. Discussion: fit-for-purpose analysis	32
6.1 SM-A1: Marine condition service	32
6.2 SM-A4: Spat capture service	32
6.3 SM-R1	32



6.4 Platform KPIs	33
6.5 Long-term validation	33
References	34



## 1. Introduction

In the FORCOAST project, a service platform, including seven Service Modules (SM), has been implemented in work package 4. The platform was co-designed by the service module developers, platform developers and the users. Preliminary user evaluation on the service products and pre-operational platform are given in Deliverable D5.2. The purpose of this report is to validate the service system, including both products and platforms based on Key Performance Indexes (KPIs), which are quantifiable performance indicators used to define success factors and measure progress toward the achievement of the service goals. The KPIs in FORCOAST are used to help to understand whether the FORCOAST service is successful and serving the strategic and operative goals, i.e., fit for the purposes of the users.

FORCOAST provides value-added information services for 7 Pilot aqua-farms via web access. The recommended KPIs fall into two main categories: metrics related to quality of service (hereafter referred to as product KPIs) and metrics related to service use (hereafter referred to as platform KPIs). In the report, the product KPIs are defined and evaluated by different SM while the platform KPIs are defined and evaluated by the platform operator.

In the scope of the FORCOAST project, not all SMs can be evaluated with KPIs since the evaluation of KPIs needs observations which cannot be obtained during the project period. For SM-F1 “Suitable fishery areas”, the KPI can be defined as the successful rate of the fishery resource index (HSI). However, this cannot be validated as no observations are available. This is similar to SM-A3 “Prospection for site selection”. For SM-F2 “Front detection”, model data have been used for generating the frontal positions. Due to a lack of in-situ observations on the frontal positions for validating the service products, it is difficult to evaluate the KPIs for the service product. This validation would require a sea survey onboard a scientific vessel for measuring temperature changes in the sea surface, covering all the pilot site. This campaign would last several weeks, and the dedication (person/month) could not be covered in the frame of the Forcoast project. In addition, these scientific vessels are in high demand and as such have to be booked years in advance. The situation is similar for SM-A2 “Land pollution”.

Therefore, only SM-A1 “Marine conditions”, SM-A4 “Assistant for spat capture” and SM-R1 “Retrieval of contaminant sources” will make the product KPI evaluation. Especially for SM-R1, experiments with drifters have to be conducted in order to evaluate the KPI's.

The report is organized as follows: section 2 – section 4 are the product KPI evaluation for the 3 SMs and section 5 for the platform KPI evaluation. In each of the sections, KPI definition, methods, input data and evaluation will be performed. Conclusions are given in section 6.

## 2. SM-A1 Marine conditions

### 2.1 Product KPIs for Limfjorden

#### 2.1.1 Selection and definition of SM-A1 KPIs

For SM-A1, marine forecast and hindcast information (sea winds, water temperature, salinity, sea level and currents) are provided to support aquaculture for safe and efficient operations. Three KPIs are selected for the forecast information: storm warning quality, storm surge warning quality and Sea Surface Temperature (SST) forecast quality.

Sea storms affect aqua-farm offshore operations by increasing the risks for damages and loss of materials and lives. For example, mussel farms on New Zealand's South Island were hit by a severe storm in August 2021 that caused damages estimated in the millions of dollars. In the Baltic Sea, storms have been identified as one of the major reasons causing low production of mussel-farming (Hedberg et al., 2018). Another example is that Oyster Boat lost its water temperature monitoring instruments in a storm case in March 2021 in Limfjorden. Accurate and timely warning of the storms is thus very important for the safe and efficient aqua-farm operations.

Service outcome	Indicator	Target
Sea storm early warning	Ratio of early warning services for sea storms that were effectively early detected and reported by the system (>20.8 m/s)	100% of observed occurrences
	Number of false positives "Red Alerts" reported by the system (>32.6 m/s)	Less than 1% of the total reported event
	Number of false positives "Orange Alerts" reported by the system (24.5-32.6m/s)	Less than 2% of the total reported events
Storm surge early warning	Ratio of effectively detected early warning storm surge events	90% of the observed occurrences
Sea temperature forecasts	Level of forecast accuracy	Deviation>1C less than 20% of the time

Table 1. Definition of KPIs for Service Module –A1: Marine conditions

A sea storm early warning is normally issued when wind speed (10 minute mean) exceeds a certain criterion. In Denmark this criterion is 20.8 m/s. The warning is issued 24 or 48 hours beforehand. The warning is divided into three categories for wind speeds 20.8 – 24.5 m/s (yellow), 24.5-32.6m/s (orange) and >32.6 m/s (red). According to this practice, a sea storm early warning KPI can be defined (Table 1), including positively detecting rate (with a target of 100%), false positive detecting rates for "Red Alerts" and "Orange Alerts", respectively.

Storm surge is another frequent coastal hazard, causing flooding and damage to aqua-farms. Weather agencies have already had their routine KPIs for storm surge early warning. In Denmark, DMI defines the water level criterion for all coastal cities for issuing a storm surge warning. A false alarm is defined as if the forecasted sea level peak deviates from the observed one by more than 10%. This criterion is used in the storm surge KPI definition (Table 1). The target is to have 90% of the storm surge cases correctly warned.

SST is an important parameter for the growth and health of mussels and oysters. Extremely warm water can also be related to oxygen depletion and algae bloom, which is harmful to the shellfish health. It is not easy to forecast coastal sea temperature correctly. For example, even with satellite SST assimilation, the hindcast validation of BALMFC forecast model still gives a mean absolute bias of 0.54 °C and cRMSE of 1.1 °C for 20 coastal stations. At Aarhus station in the eastern Limfjorden, the BALMFC forecast model has a SST bias of -0.9 °C and cRMSE of 1.4 °C. Therefore our target of the SST KPI is set as less than 20% for forecast deviations >1 °C (Table 1).

### 2.1.2 Methods and input data for calculating the KPIs

For Pilot 6, west Limfjorden is the focused area. Oyster Boat is operating an oyster farm near Lemvig (Figure 1). This area is selected for KPI verification. Ocean forecast (including surface winds) is provided by DMI in real time. For sea storm early warning, the forecasted surface winds are used to make the sea storm early warning while the observed winds at 3 synoptic stations: 06052, 06019 and 06063 (Figure 1) are used to validate the positive and the false alarming rates of the storms. These stations are coastal stations. For simplicity and statistical representability, mean wind speeds at the 3 stations are used for issuing and validating the storm warning. The KPI evaluation period covers one year starting from 17 March 2021.

For storm surge KPI, forecast and observation data at Lemvig are used. The KPI assessments use the results of the forecast validation for the operational period of the model, which was commissioned on the 17th of March 2021. The validation period covers nearly 11 months, until the 8th of February 2022. The assessment focuses on sea level events exceeding pre-defined warning levels. Usually, these warning levels are user defined, but in the presented case, we have chosen to use 1 m as the critical value. Reason for this is that the period is too short for a comprehensive assessment of warning events. The number of warning events is too low. In the entire period, there is only 1 event, storm Malik in Denmark, at the 29<sup>th</sup> to 30<sup>th</sup> of January 2022, which exceeds the warning level of 1.35 m at Lemvig. In total there are 7 events in the assessment period with observed sea levels exceeding 1 m. These have been selected for the assessment.

The assessment is limited by the time resolution of the model data, which is hourly. Modelled sea level data in higher time resolution is not available for longer forecasting periods, after the first 12 hours. Only the archive data is saved. For that reason, some of the peak events might have been missed, if the peak falls in between two, hourly output time steps. To compensate this, the observations were averaged in time, over each hour, to compare data sets with identical time resolution.

The KPI assessment follows the procedure for operational model assessments. In the first step, modelled and observed sea level peak values are matched to each other. The method neglects phase errors by allowing for a time shift of up to  $\pm 6$  hours between modelled and observed peaks. A separation window of at least 12 hours between peak events ensures that storm surges are not counted twice. The assessment analyses peak errors and compares them with justified error ranges, to calculate a miss rate. If the peak error exceeds the error range, the forecast is called a “missed warning”. The number of missed warnings is studied as well. The peak error assessment works with a justified error range of 10% of the observed peak value, with a minimum of 10cm.

For calculating SST KPI, the water temperature from satellite and cruise observations in the western Limfjorden are compared with forecasts. The locations of Lemvig and cruise stations are shown in Figure 1. The cruise data (T/S profiles) are obtained from Danish Environmental Protection Agency data portal <https://miljoedata.miljoportal.dk/> . The satellite SST is obtained from CMEMS with product ID SST\_BAL\_SST\_L3S\_NRT\_OBSERVATIONS\_010\_032. For the KPI evaluation using satellite SST, only data from

the cruise station locations have been used, in order to enable an intercomparison between the satellite and cruise data.

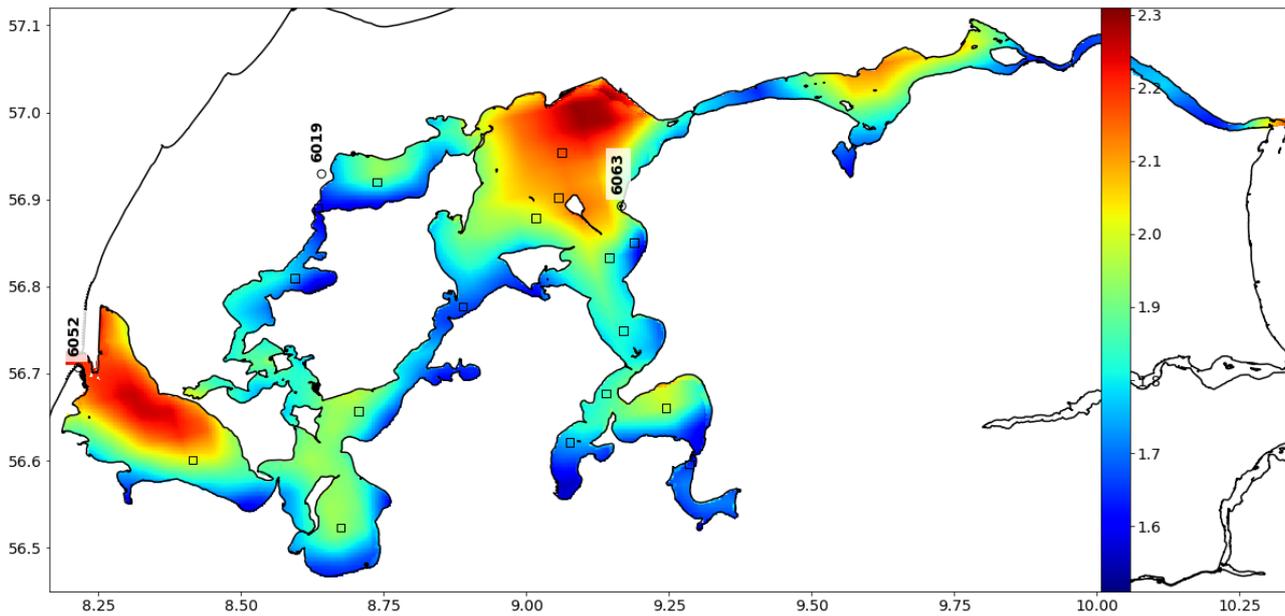


Figure 1. Observation stations used for calculating and validating the KPIs in the western Limfjorden, together with mean wind speeds from 2015-2019 (in m/s).

### 2.1.3 Results

#### 2.1.3.1 Storm early warning

For the forecast demonstration period, there is only one event reached the warning level, occurring in 29<sup>th</sup> to 30<sup>th</sup> of January 2022 when the observed mean wind speed exceeded 20.8 m/s but less than 24.6 m/s (Figure 2). The forecasted mean wind speed also fell in the same interval, thus a yellow category warning should be issued, according to the forecast. This also means that the position early warning rate is 100%, and the targeted KPI criterion (as defined in Table 1) has been reached

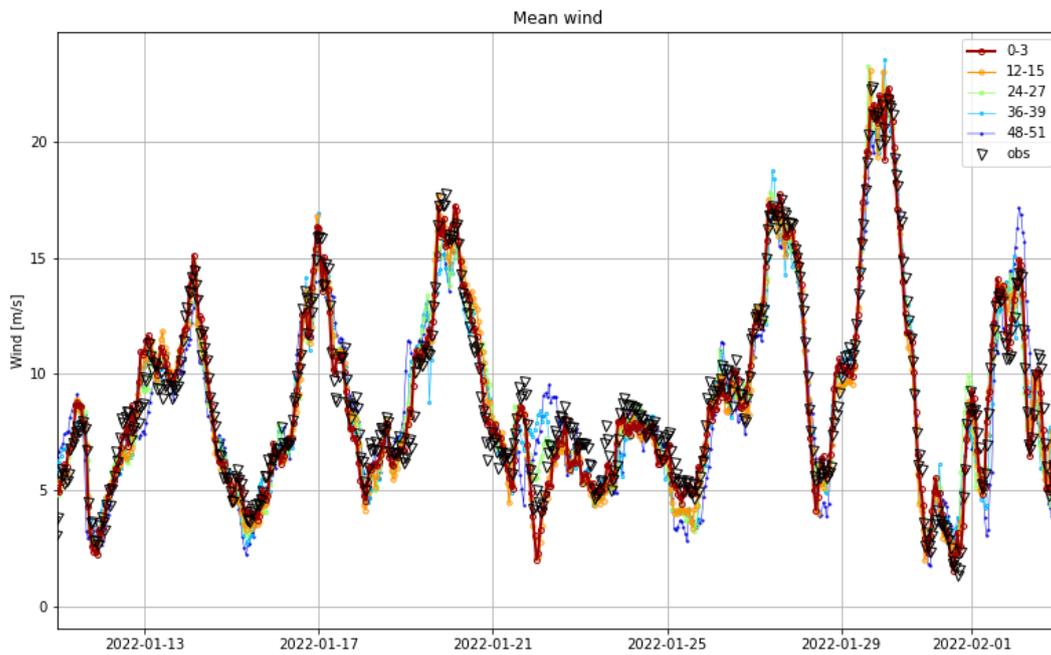


Figure 2. Observed and forecasted averaged wind speed time series (derived from 3 coastal stations in western Limfjorden) during the storm event on 29th – 30th January 2022. The forecasts are shown with different forecasting ranges from 1-51 hours.

### 2.1.3.2 Storm surge early warning

The assessment focuses on sea level anomalies, which are generated by subtracting the long-term mean from the model forecasts and observations. For the current assessment we have chosen to subtract the long-term mean of the assessment period 17th of March 2021 to the 8th of February 2022.

The model is able to predict 6 of the 7 events but miss one of them for the 24 hour forecast. The miss rate is 14%. The forecast range does not matter that much during the first 1.5 days of the forecast. Blue, yellow and red circles are overlapping closer in Figure 3. But after 1.5 days, the forecast quality seems to reduce. For the 48h forecast, one more storm is missing compared to the 24h and 36h forecast. The missing rate increases to 29%. Therefore, it is beneficial to provide users with regularly updated information.

The here presented assessment should be extended to cover larger periods and numbers of extreme events. This would provide stronger statistical significance to the assessment.

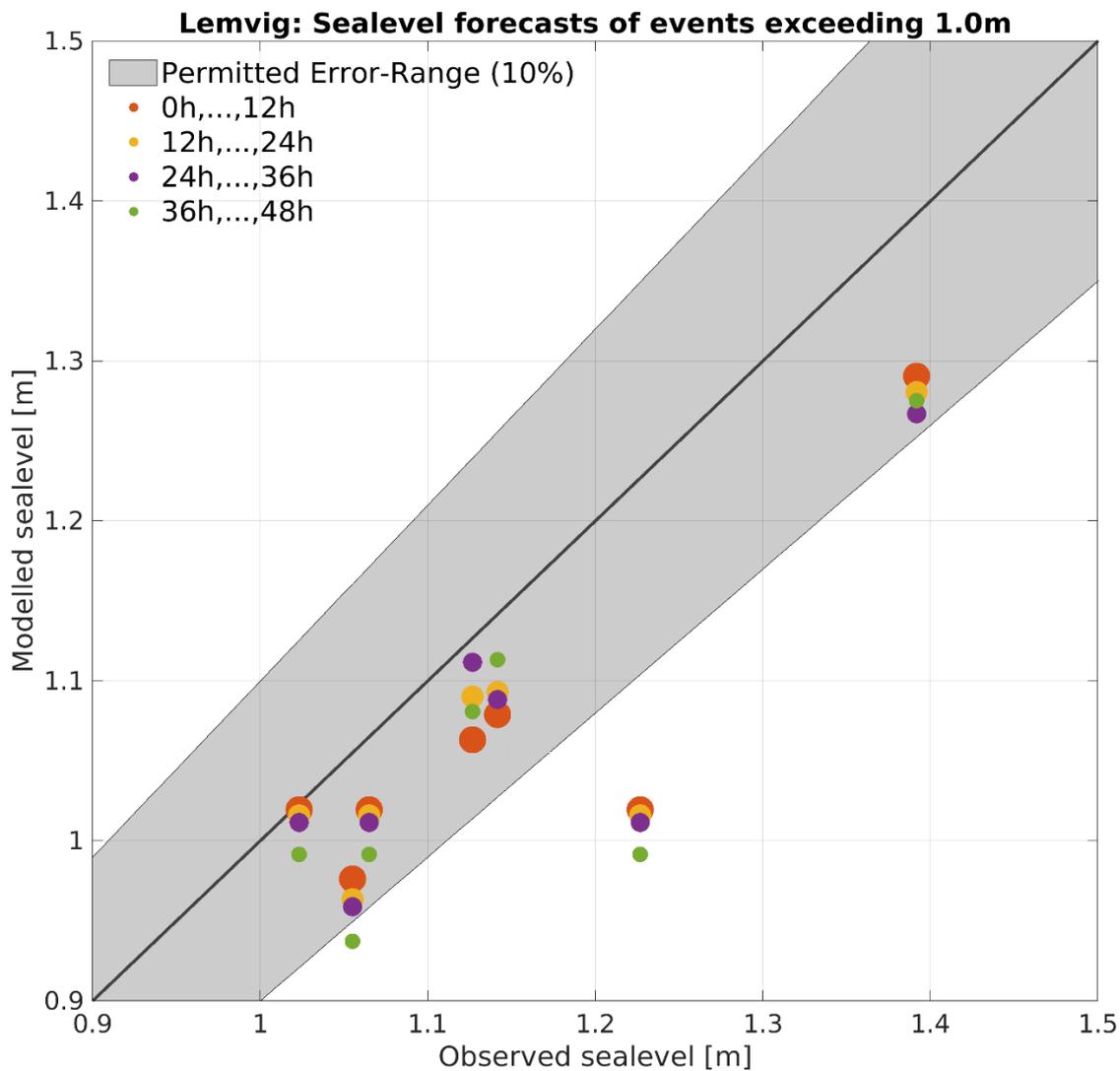


Figure 3. Model performance with regards to predicting sea level extreme events (storm surges) at Lemvig tide gauge station in the period 17th of March 2021 to 8th of February 2022. The model runs twice daily, so each circle represents a forecast that is 12 hour.

### 2.1.3.3 SST forecast

The SST KPI, i.e., the fraction of SST forecast error  $> 1\text{ }^{\circ}\text{C}$ , derived from forecasts, satellite data and cruise data for forecast ranges up to 132h, is shown in Figure 4. The intercomparison with satellite SST showed that the fraction of SST forecast error  $> 1\text{ }^{\circ}\text{C}$  is less than 20% for most of the forecast ranges except for 108-120h. However, results from the cruise data are worse than the satellite data, showing a KPI fraction of 24-29%. Further investigation indicated that the large SST error is mainly detected in winter time. This can be illustrated in Figure 3 using the KPI evaluated only from  $\text{SST} > 5\text{ }^{\circ}\text{C}$ . The results showed that the KPIs derived from satellite and cruise SST are similar, which are between 12-16%, much smaller than the full datasets. This means that the SST KPI in winter is much worse than the non-winter seasons. The KPI differences between using the satellite SST and using the cruise observations are mainly caused by low water temperature, i.e., winter situation. This is understandable as the satellite SST is heavily biased in the number of samples in

winter and non-winter seasons due to the impacts of clouds while the cruise observations are more evenly distributed in the seasons.

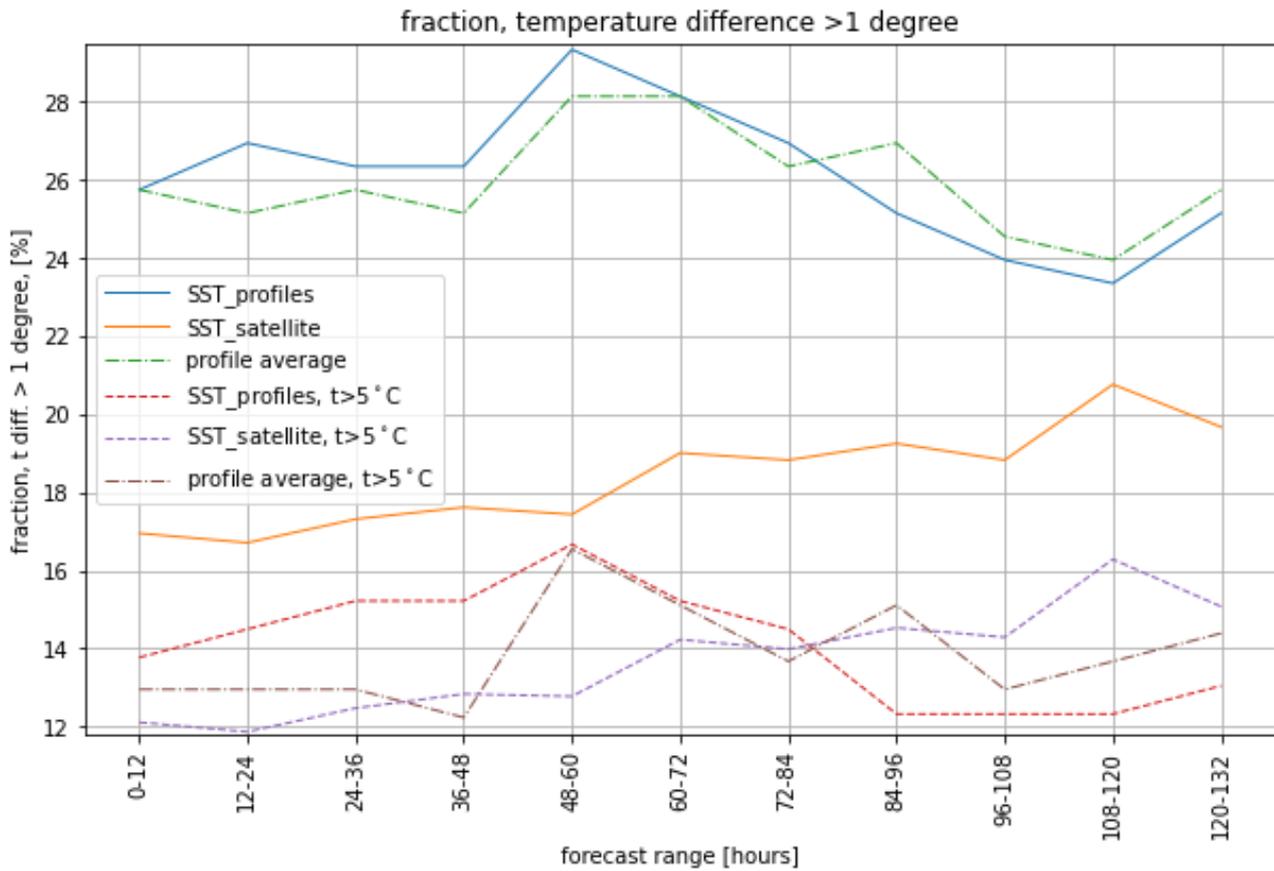


Figure 4. SST KPIs, i.e., the fraction of forecast error >1 oC, derived from forecast s with different forecast ranges up to 132h and observations from satellite SST, cruise SST and cruise T/S profile average.

### 3. SM-A4 Assistance for spat capture

#### 3.1 Selection and definition of KPIs

For SM-A4, assistance for spat capture (period of arrival at the collection site) are provided to support aquaculture. To assess service module performance, arrival of larvae in the nurseries are compared to model prediction based on Receiver Operating Characteristic (ROC) curve. Quantification of spat arrivals is not included in SM-A4 due to the actual state of scientific knowledge which does not allow to evaluate the quantity of spats which arrive but only the timing, hence the true/false positive rate and true/false negative rate are computed weekly.

The particularity of the SM-A4 module is that several parameterizations can be used. In this context data will help to choose the best parameterization to assess SM-A4 performance.

The SM-A4 is evaluated on the Belgian pilot site.

#### 3.2 Methods and input data

The main difficulty to assess SM-A4 performance is the data collection. The capture of spats requires the installation of collectors. Those collectors have been installed in the pilot site. The pilot site is an experimental site, and we evaluate efficiency of collectors at the same time as data collection. Collectors are assessed each week during the spat arrival season to count the number of spats that settled on the substrates. Performance indicators compare weekly the arrival of spats on the collector with model prediction.

The data cover partially one year for each species (2020 for oysters and 2021 for mussels), we make the assumption that we cover the whole arrival period for oysters, however data do not cover the whole arrival period for mussels.

Two parameterizations for spawning events are evaluated for each species. For mussels, a spawning event occurs when a temperature of 9°C or 10°C is reached. In the case of oysters, spawning occurs either when a temperature of 15°C is reached or when a cumulative temperature of 576°C (cumulative temperature upper 7°C from 1<sup>st</sup> of January, the value of 7°C corresponds to a threshold for gonad development) is reached.

For more details on the parameterization used and references consult the deliverable D5.6 parameterizations “Platform User Manual.”

The temperature date used to determine spawning event condition and their validation are specifically described in the deliverable D5.4 “Final Coordinated Pilot Model Evaluation Report.”

#### 3.3 Results

Spat capture on the collector can be compared to SM-A4 prediction.

##### **Oysters**

Figure 5 shows the capture of spats in the pilot site and the SM-A4 prediction. Table 2 and Table 3 show the performance of the service module. Spawning at 15°C showed better performance than a parameterization based on cumulative temperature (TP and NP higher; lower value for FP and FN).

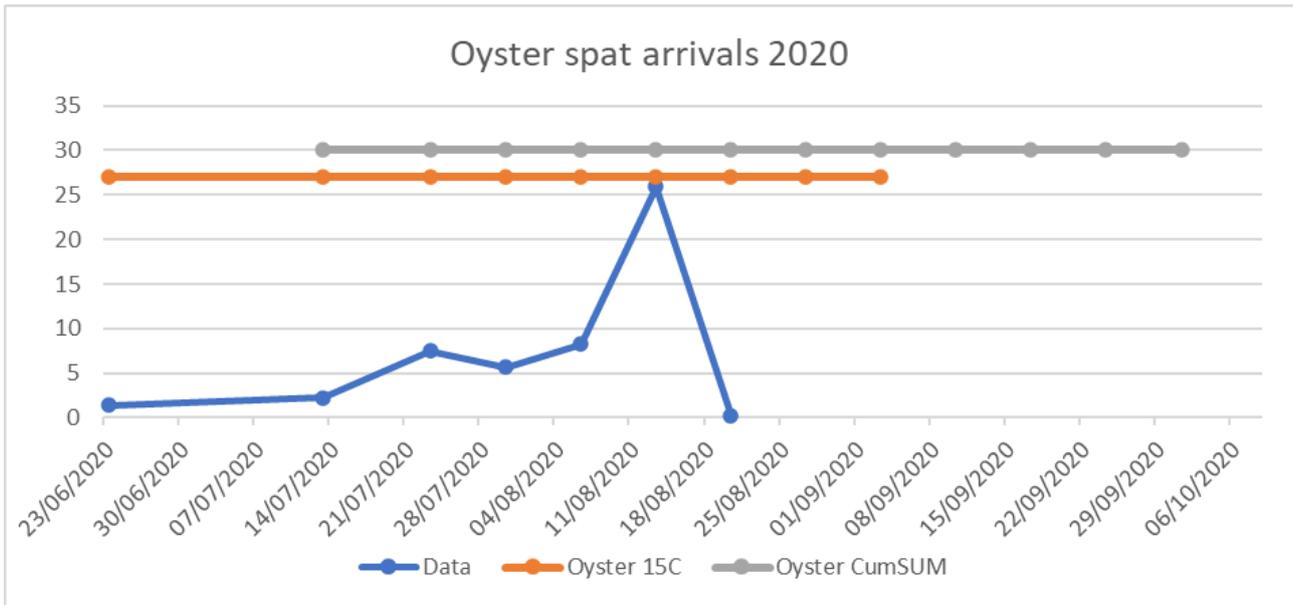


Figure 5. weekly arrival of oysters spats on the collector for the year 2020. The blue line represents the spat capture on the collector, the y axis is the number of spat captures on the collector. Orange line represents the prediction of SM-A4 under the hypothesis of spawning events at 15°C and the grey line in the case of a cumulative temperature hypothesis. SM-A4 only predicts the period of arrival.

<p><b>False positive (FP) (arrivals predicted but not observed)</b></p> <p style="text-align: center;"><b>2</b></p>	<p><b>True Positive (TP) (arrivals predicted and observed)</b></p> <p style="text-align: center;"><b>7</b></p>
<p><b>True Negative (NP) (arrivals not predicted and not observed)</b></p> <p style="text-align: center;"><b>41</b></p>	<p><b>False negative (FN) (arrivals observed but not predicted)</b></p> <p style="text-align: center;"><b>0</b></p>

Table 2. The table summarizes the performance of the SM-A4 module when the spawning period is based on a temperature of 15°C.

<p><b>False positive (FP) (arrival predicted but not observed)</b></p> <p style="text-align: center;"><b>4</b></p>	<p><b>True Positive (TP) (arrival predicted and observed)</b></p> <p style="text-align: center;"><b>6</b></p>
--	---

<p><b>True Negative (NP) (arrival not predicted and not observed)</b></p> <p><b>39</b></p>	<p><b>False negative (FN) (arrival observed but not predicted)</b></p> <p><b>1</b></p>
--	--

Table 3. The table summarizes the performance of the SM-A4 module when the spawning period is based on a cumulative temperature.

### Mussels

Figure 6 shows the capture of spats in the pilot site and the SM-A4 prediction. Table 4 and Table 5 show the performance of the service module. Determination of spawning events based on a temperature of 9°C shows a better performance than at 10°C, which predicts a too late arrival in comparison with observation. However, data for this year are incomplete, the first data was collected on the week of the 9<sup>th</sup> of April 2021 where the highest value is observed, which seems to show that the collector was installed too late and we can make the hypothesis with regards to the data distribution that we missed the beginning of the spat arrivals. SM-A4 at 9°C predicting arrivals before the week of 9<sup>th</sup> of April 2021 seems realistic and, in this case, we only considered arrivals after the week of 15<sup>th</sup> June as FP. In any case FP, NP and FN are not possible to determine with this incomplete data set.

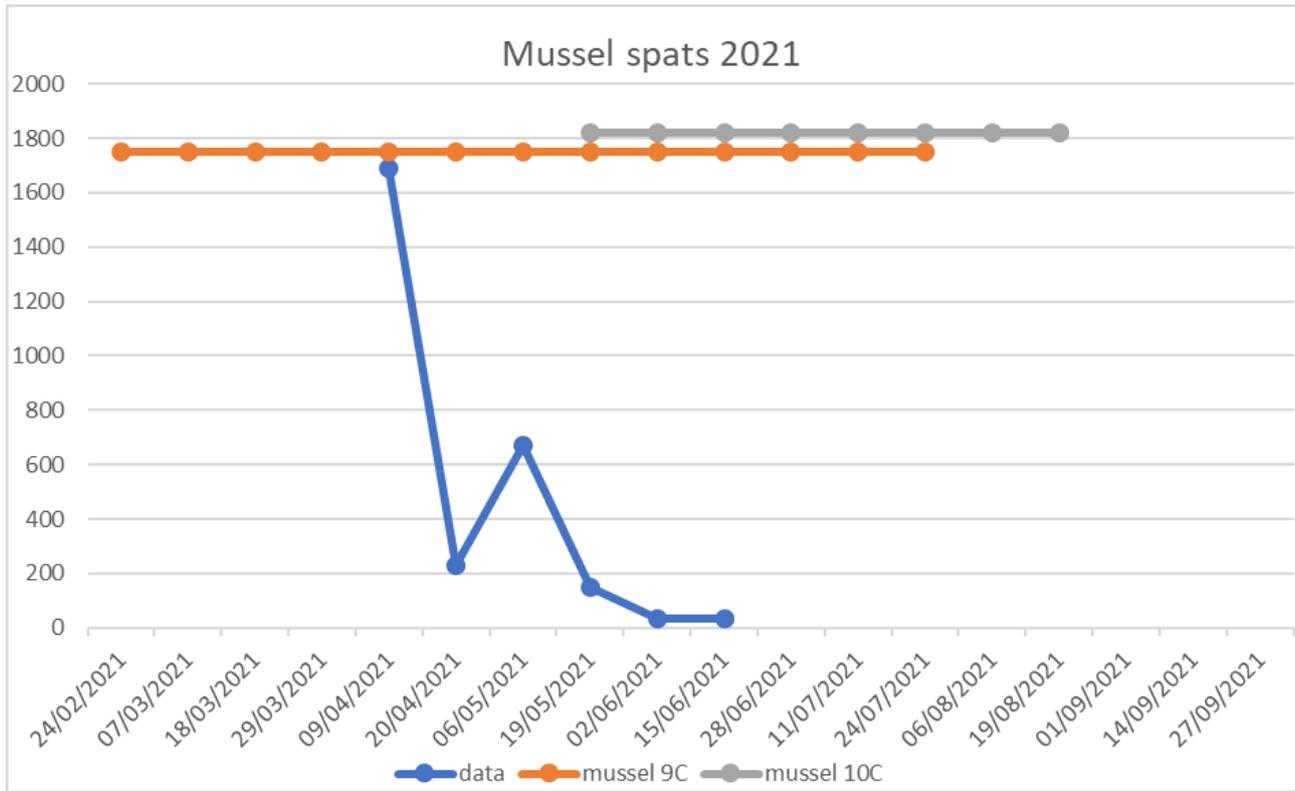


Figure 6. Weekly arrival of mussel spats on the collector for the year 2021. The blue line represents the spat capture on the collector, the y axis is the number of spat captures on the collector. Orange line represents the prediction of SM-A4 under the hypothesis of spawning events at 9°C and the grey line in the case of spawning events at 10°C. SM-A4 only predicts the period of arrival.

<p><b>False positive (FP) (arrivals predicted but not observed)</b></p> <p style="text-align: center;"><b>3</b></p>	<p><b>True Positive (TP) (arrivals predicted and observed)</b></p> <p style="text-align: center;"><b>6</b></p>
<p><b>True Negative (NP) (arrivals not predicted and not observed)</b></p> <p style="text-align: center;"><b>Not possible to determine</b></p>	<p><b>False negative (FN) (arrivals observed but not predicted)</b></p> <p style="text-align: center;"><b>Not possible to determine</b></p>

Table 4. The table summarizes the performance of the SM-A4 module when the spawning period is based on a temperature of 9°C.

<p><b>False positive (FP) (arrivals predicted but not observed)</b></p> <p style="text-align: center;"><b>5</b></p>	<p><b>True Positive (TP) (arrivals predicted and observed)</b></p> <p style="text-align: center;"><b>3</b></p>
<p><b>True Negative (NP) (arrivals not predicted and not observed)</b></p> <p style="text-align: center;"><b>Not possible to determine</b></p>	<p><b>False negative (FN) (arrivals observed but not predicted)</b></p> <p style="text-align: center;"><b>Not possible to determine</b></p>

Table 5. The table summarizes the performance of the SM-A4 module when the spawning period is based on a temperature of 10°C.

## 4. SM-R1 Retrieve sources of contaminants

### 4.1 Selection and definition of KPIs

The objective of Service Module R1 - Retrieve Sources of Contaminants, is to provide end users with a probabilistic estimation of the origin of contamination affecting aquaculture farms. Of course, this is only interesting in areas where the source of contamination is unknown, such as water bodies surrounded by rural areas where agriculture and farming are the main economic activities. Agricultural and farming practices are often inextricably linked to the release of large amounts of nutrients, organic matter and agro-chemicals to the neighboring waters. Fecal pollution occurs when disease-causing organisms (e.g. *Escherichia coli*) in the gastrointestinal tract of mammals enter the rivers and other water courses, to finally reach the marine environment. An important aspect of this process is the *diffuse* nature of this contamination in a mostly rural area with many potential sources spread along the coastline, and this is particularly true for an estuary like Galway Bay, where farming lands are widespread across all the territory, and there are countless small freshwater courses that would bring the fecal contamination into the estuary. Once in the marine environment, filter-feeders such as oysters and mussels may be affected by the contamination. Bacterial contamination of shellfish (e.g. where *Escherichia coli* has been found present) affects the route to the market and the revenue.

Again, the main issue here is the inability to determine the land source of contamination, and this is where Service Module R1 can be helpful by providing probability maps of the distribution of contaminants using backward tracking from the affected farms. However, the fact that it is impossible to determine the real-world source of contamination means that, in fact, there is no way to check that the service is providing accurate estimations. This service is not a forecasting service like SM-A1 Marine Conditions, where the accuracy of a forecasted storm surge can then be checked in the real world. Rather, in this service, for a past contamination event, the service searches backward-in-time the potential sources of contamination based on the oceanic currents field. In other words, the event that is being investigated has already occurred and most likely has not left any footprints at the origin, only at the affected aquaculture farm (the contaminated shellfish).

Therefore, defining service KPIs based on the accuracy of the estimations is simply not possible, as all the service is providing is an educated guess about an event in the past. Instead, in order to define KPIs for this service, it is possible to measure how accurately the background model predicts a similar process that takes place under controlled conditions. Such procedure and the methods involved are explained in the section below.

### 4.2 Methods and input data

The rationale behind Service Module R1 is particle-tracking modeling. In particle-tracking modeling, a flow velocity field (e.g. atmospheric winds or oceanic currents defined on a 3-dimensional space over time) is used to track the dispersion of an object forced by that flow velocity field. Such objects can be marine litter, missing persons at the sea, oil spills or a water contamination plume. Forward-in-time particle-tracking modeling is the most usual approach, when the source is known and the goal is to determine how the object will move in the future. This is the approach followed by SM-A2 Land Pollution. In backward-in-time particle-tracking modeling, the objective is to find the source, and for this the velocity field is reversed and the object is tracked starting from the aquaculture farm.

Under controlled conditions, it is possible to release drifters in the marine environment and then track their positions using satellite positioning systems. Drifters are buoyant objects that remain on or near the ocean surface and move freely under the effect of oceanic currents and, to some extent, atmospheric

winds. The drifters are recovered later and then the same model used to track the contamination is used to simulate the path followed by these drifters. It is possible to evaluate the performance of the model by defining a skill score that measures the goodness of fit between the observed trajectory and the modeled trajectory. This skill score can then be used to define a KPI for this service.

This is the approach that has been followed here. During a survey conducted on 23 March 2022, a few WAVY drifters (see MELOA project at <https://www.ec-meloa.eu/pages/wavy-drifters>, Fig. 7) were released in Galway Bay at different sites and tracked for the next few hours. These drifters are designed to stay beneath the surface and reduce the windage to a minimum. However, a small fraction of the drifter remains above the surface, thus the wind exerts some pressure on the drifter and this should be taken into consideration when simulating its advection.



Figure 7. A WAVY drifter floating at the ocean surface. Source: bluewisemarine.ie

Then, the OpenDrift particle-tracking model, coupled with the hydrodynamic fields derived from the Galway Bay model, is used to simulate the trajectories followed by the drifters. Finally, a skill score can be developed based on the separation distance between the observed (real) trajectory and the modeled (simulated) trajectory. The idea is described in Liu & Weisberg (2011) and is explained using the figure below (Figure 8).

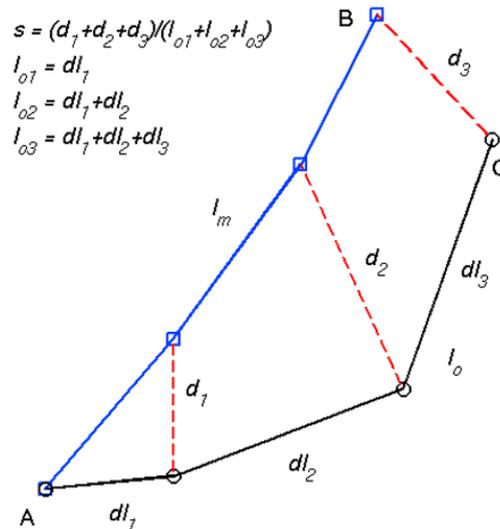


Figure 8. The Liu & Weisberg (2011) skill score. Here, the observed (black) and modeled (blue) trajectories are represented. The separation distances  $d_1, d_2, d_3, \dots$  between the observed and modeled trajectories are determined. The skill score is defined as the sum of the separation distances normalized by the corresponding total lengths of the observed trajectory  $l_{o1}, l_{o2}, l_{o3}, \dots$

The Liu & Weisberg (2011) skill score is based on the normalized separation distance between the observed and modeled trajectories. Following the description in Figure 8, the index  $s$  is defined as:

$$s = \frac{\sum_{i=1}^N d_i}{\sum_{i=1}^N l_{oi}}$$

And the skill score  $ss$  is defined as  $ss = \max\{0, 1 - s\}$ . In an ideal world, with a perfect match between the observed and the modeled trajectories, the successive separation distances  $d_1, d_2, d_3, \dots, d_{N-1}, d_N$  are all 0 and  $s = 0$ , and the skill score would be  $ss = 1$ . On the contrary, a skill score  $ss = 0$  (no skill) occurs when the sum of the separation distances are equal or greater than the corresponding total lengths of the observed trajectory. Therefore, under these definitions, a skill score of 0 means no skill of the service, and a skill score of 1 means that the service is perfect, although in practice this is unachievable.

It is important to notice that this approach to evaluate SM-R1 Retrieve Sources of Contaminants has some drawbacks. As explained above, there are some differences between the experiment being evaluated and occurring under controlled conditions (i.e. the motion of the drifter) and the process of interest (i.e. the dispersion of agricultural and farming contaminants in the bay). First of all, the drifter is not a passive tracer, it is buoyant, remaining at the surface of the ocean and being partially affected by winds, whereas the contaminants move along with the water. In addition, the motion of the drifter is studied in a forward-in-time mode, whilst the rationale of the service is the backward-in-time modeling of the contaminants, from the aquaculture farm to the unknown source.

### 4.3 Results

In addition to the WAVY drifters released on 23 March 2022, the results here are supplemented with trajectory data from additional drifter experiments conducted in Galway Bay in January 2015. The model trajectory is set to start at the same position as the real one, and then the trajectory is simulated for the whole duration of the observed trajectory. The resulting Liu & Weisberg (2011) Skill Scores have been calculated separately for each drifter and presented in Table 6.

Drifter ID	Liu & Weisberg (2011) Skill Score
WAVY #129 - 23 Mar 2022	0.351
WAVY #130 - 23 Mar 2022	0.803
WAVY #131 - 23 Mar 2022	0.505
WAVY #133 - 23 Mar 2022	0.793
WAVY #134 - 23 Mar 2022	0.134
TRACKER #281226 - 21 Jan 2015	0.803
TRACKER #284469 - 21 Jan 2015	0.554

TRACKER #284519 - 21 Jan 2015	0.697
-------------------------------	-------

*Table 6. Liu & Weisberg skill scores for the drifters deployed in Galway Bay*

The results show that the model is quite heterogeneous in relation to its ability to predict the path of different objects, with a poor performance ( $ss < 0.5$ ) for drifters WAVY #129 and WAVY #134, and an excellent performance ( $ss > 0.8$ ) for drifters WAVY #130 and TRACKER #281226. In order to illustrate how good this performance is, the observed (green) and modeled (red) paths of these top-performing drifters are shown below (Figure 9).

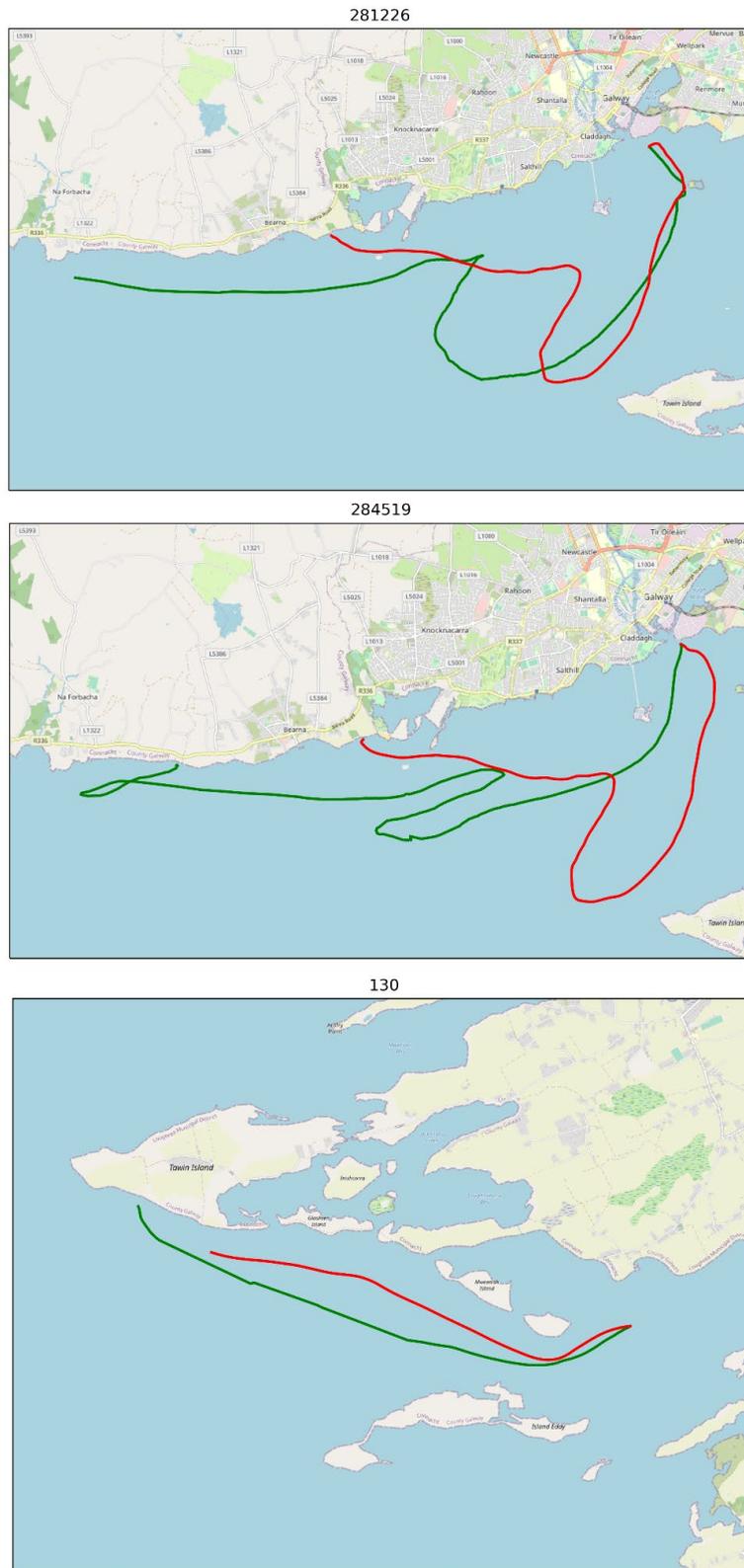


Figure 9. Observed (green) and modeled (red) trajectories of drifters TRACKER #282226 (top,  $ss = 0.803$ ), TRACKER #284519 (middle,  $ss = 0.697$ ) and WAVY #130 (bottom,  $ss = 0.803$ ).

## 5. Platform performance

### 5.1 Selection and definition of KPIs

Table 7 displays the proposed KPIs obtained to assess the service platform prototype fit-for-purpose, which results are showcased in section 5.3.

Category	KPI	Definition
<b>Service availability</b>	Available services	% of scheduled available services
	Service uptime	% of availability, per service
<b>Input data availability</b>	Available data	% of available data inputs from Geoserver
	Data uptime	% of availability, per data input from external sources
<b>Interruptions</b>	Downtime per service	How long the longest straight unavailability period for each service was.
<b>Timeliness</b>	Platform service timeliness	Time between when the data is available and ready to be used (in the platform))
	Telegram service timeliness	Time between when the data is available and ready to be used (scheduled bulletins)
<b>Platform Usage</b>	Number of users	How many people access the platform
	Engagement time	Time spent in the platform per user and per session

Table 7. Platform prototype KPIs

### 5.2 Methods and input data

Depending on the category and specific KPI to obtain, input data from different sources are required and the way they are processed changes as well.

For service availability, for scheduled services, bulletins produced by the service available via Telegram have been checked. In the case of services operational only in the platform they are included in the overall statistics of successful runs, back-end K8 logs (Kubernetes cluster) are used. For service interruptions, the same input data is checked.

For input data availability, a distinction is made between layers hosted in the FORCOAST back-end (Geoserver) and external data layers (partner's infrastructure). The layers hosted in the Geoserver are monitored four times per day, sending a monitoring bulletin to the partners in charge of the layers (Figure 10). The layers can have three different status: available, out-of-date (more than 96 hours behind real-time) or unavailable.

Pilot	Source	Layers	TinyURL	Response code	Last available date	Date status	Status
1. Sado Estuary	MARETEC THREDDS	(YYYYmddd)00_Surface.nc:ssh	tinyurl.com/2tft8eje	Success (200)	22-06-21:02:00:00	Up-to-date < 96h	✔
1. Sado Estuary	MARETEC THREDDS	(YYYYmddd)00.nc:wind_speed	tinyurl.com/29pvpb5x	Success (200)	22-06-21:02:00:00	Up-to-date < 96h	✔
2. Bay of Biscay	EusKOOS THREDDS	(YYYYmddd)_fronts_exp.nc:sst	tinyurl.com/5enj2szh	Success (200)	22-06-23:02:00:00	Up-to-date < 96h	✔
2. Bay of Biscay	Geoserver	forcoast:biscay_sst	tinyurl.com/FORCOAST	Success (200)	22-06-21:23:00:00	Up-to-date < 96h	✔
2. Bay of Biscay	Geoserver	forcoast:biscay_fronts_sst	tinyurl.com/FORCOAST	Success (200)	22-06-21:23:00:00	Up-to-date < 96h	✔
3. Black Sea BG	Geoserver	forcoast:pilot_3_CMEMS_sst	tinyurl.com/FORCOAST	Success (200)	22-06-30:00:00:00	Up-to-date < 96h	✔
3. Black Sea BG	Geoserver	forcoast:HSI_whiting	tinyurl.com/FORCOAST	Success (200)	22-06-11:00:00:00	Outdated > 96h	⏸
4. North Sea	RBINS ERDDAP	NOS_HydroState_V1.nc:sst	tinyurl.com/3hfeuj2n	Success (200)	22-06-24:02:00:00	Up-to-date < 96h	✔
5. Galway Bay	Marine Institute THREDDS	galway_bay_(YYYYmdddHH)_AN.nc:?	tinyurl.com/22e8936j	Success (200)	22-06-13:02:00:00	Outdated > 96h	⏸
5. Galway Bay	Marine Institute THREDDS	IMI_ROMS_HYDRO/GALWAY_BAY_NATIVE_70M_8L_1H forecast	null	Success (200)	22-06-22:02:00:00	Up-to-date < 96h	✔
6. Limfjord	DMI FTP via Geoserver	forcoast:dk_elev	tinyurl.com/FORCOAST	Success (200)	22-06-26:00:00:00	Up-to-date < 24h	✔
6. Limfjord	DMI FTP via Geoserver	forcoast:dk_salt	tinyurl.com/FORCOAST	Success (200)	22-06-26:00:00:00	Up-to-date < 24h	✔
6. Limfjord	DMI FTP via Geoserver	forcoast:dk_temp	tinyurl.com/FORCOAST	Success (200)	22-06-26:00:00:00	Up-to-date < 24h	✔
6. Limfjord	DMI FTP via Geoserver	forcoast:dk_currentspeed	tinyurl.com/FORCOAST	Success (200)	22-06-26:00:00:00	Up-to-date < 24h	✔
6. Limfjord	DMI FTP via Geoserver	forcoast:dk_windspeed	tinyurl.com/FORCOAST	Success (200)	22-06-26:00:00:00	Up-to-date < 24h	✔
7. Black Sea RO	SEAMOD THREDDS	_2_EFORIE_grid_T	null	Success (200)	22-06-24:02:00:00	Up-to-date < 96h	✔
7. Black Sea RO	SEAMOD THREDDS	_1_NWS_grid_T	null	Success (200)	22-06-24:02:00:00	Up-to-date < 96h	✔
7. Black Sea RO	SEAMOD THREDDS	BS_1h_grid_T	null	Success (200)	22-06-24:02:00:00	Up-to-date < 96h	✔
8. Northern Adriatic	Geoserver via OGS	forcoast:northernadriatic_thetao	tinyurl.com/FORCOAST	Success (200)	13-01-01:00:00:00	Up-to-date < 96h	✔
8. Northern Adriatic	Geoserver	forcoast:northernadriatic_fronts_sst	tinyurl.com/FORCOAST	Success (200)	22-06-21:23:00:00	Up-to-date < 96h	✔
8. Northern Adriatic	OGS THREDDS	chl	null	Success (200)	22-06-23:23:00:00	Up-to-date < 96h	✔
8. Northern Adriatic	OGS THREDDS	o2	null	Success (200)	22-06-23:23:00:00	Up-to-date < 96h	✔
Multiple pilots	CMEMS THREDDS	CHL	null	Success (200)	22-06-20:00:00:00	Up-to-date < 96h	✔
Multiple pilots	CMEMS THREDDS	analysed_sst	null	Success (200)	22-01-01:00:00:00	Outdated > 96h	⏸

Figure 10. Input data layer monitoring bulletin

For timeliness, the computation time is taken into account when running the services from the front-end, while for the scheduled services timeliness is instantaneous as the services can be scheduled for any time of the day and at any frequency.

Lastly, for platform usage, Google Analytics is used to retrieve different statistics from user interaction with the platform in its web application form.

### 5.3 Results

#### 5.3.1 Service availability

Figure 11 shows, for the month of August 2022, how often the different scheduled services have been available. Please note that due to scheduling system maintenance, there were between two and three days that there have not been automated runs. However, the services were operational in the central platform.

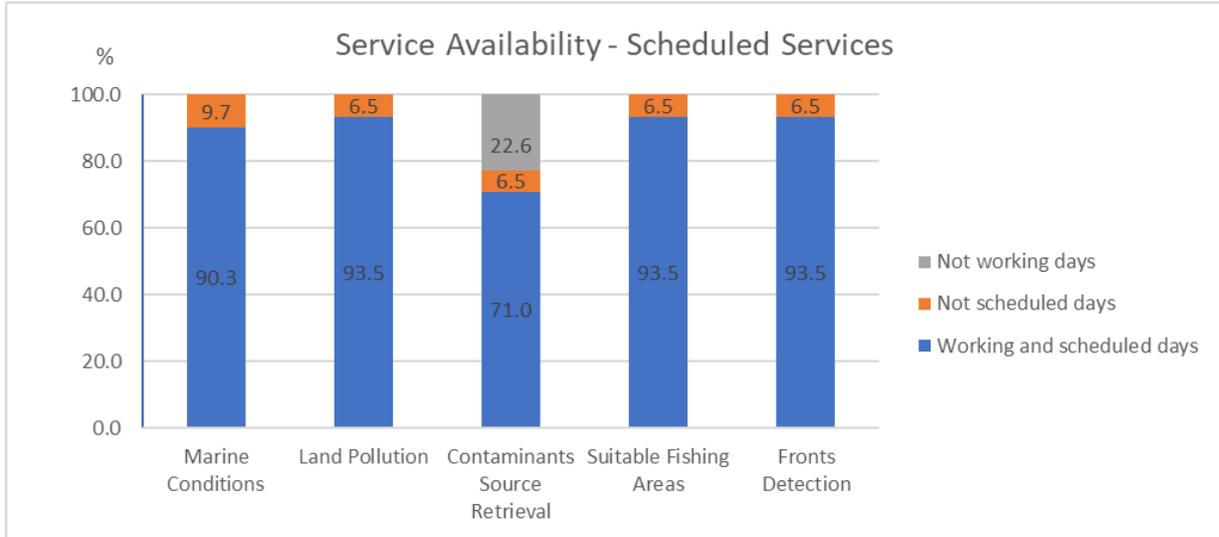


Figure 11. Service availability for scheduled services

Looking at total service run requests per service, the results can be seen in Figure 12 in the form of uptime (%) in the month of August. These statistics are retrieved from comparing the total number of job requests and the successful ones. Note that most of the percentage is marked as downtime due to the monitoring system being under maintenance for two days, thus making it not possible to retrieve successful logs from those days even if the services were working operationally in the platform prototype. This accounts for approximately 70% of the total downtime seen in Figure 12. This makes the effective service uptime for most of the cases very close, or reaching 100%.

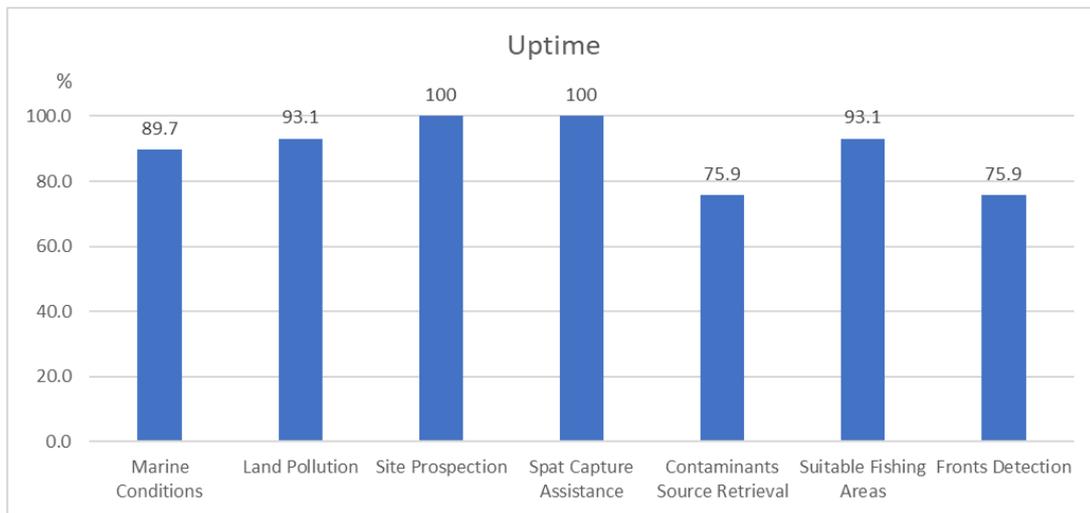


Figure 12. Service uptime

### 5.3.2 Input data availability

Looking at the data layers stored locally in the FORCOAST Geoserver, the statistics retrieved from the monitoring bulletins for the month of August 2022 are displayed in Figure 13. All of the data layers are available in the Geoserver, which in this case are either CMEMS data layers or datasets from those partners

that do not have THREDDS infrastructure. In most cases, every or almost every layer is up-to-date. The exception is Romania that, even if available, the specified CMEMS product is not being updated. In any case this data does not feed into the services implemented in the Romanian pilot, thus not preventing the services from being operational. Also, there are no input data layers of this type for the Galway Pilot.

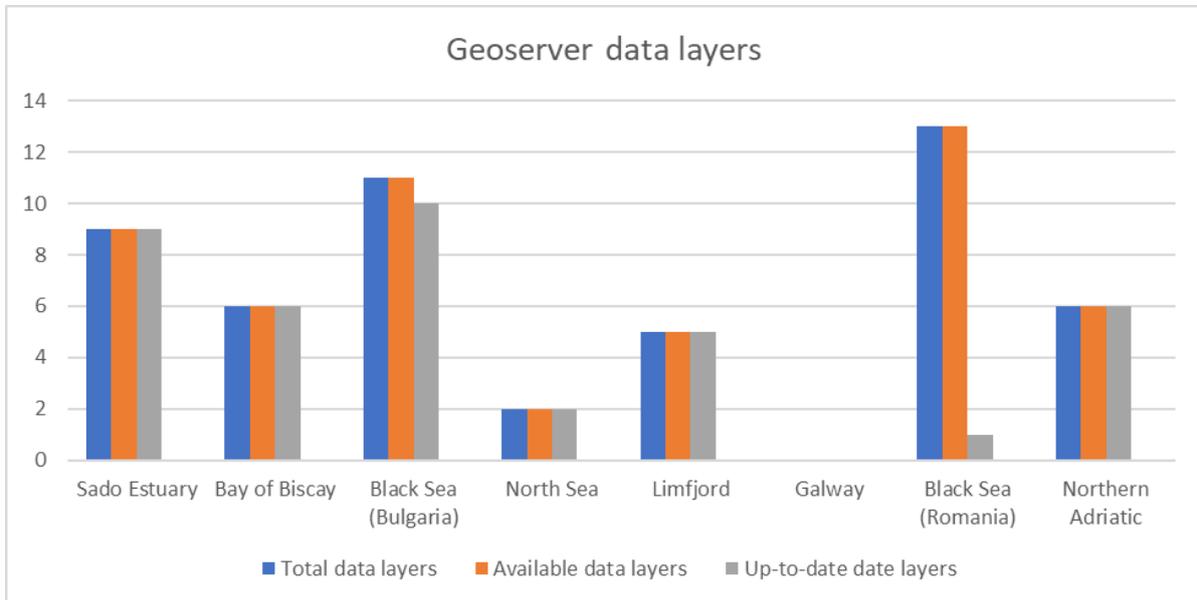


Figure 13. Data layer status at the FORCOAST Geoserver

Most of the input datasets used by the different service modules (and displayed as independent layers in the platform) are retrieved from the FORCOAST partners' infrastructures (i.e. THREDDS, ERRDAP, etc). An overview of the status situation for those layers in August 2022 is displayed in Figure 14. Every layer necessary is available, and the few ones that are not up-to-date are due to an sporadic interruption in the respective partners' model or delivery system which they are notified from the data monitoring system (Figure 10) so they can take action as soon as possible.

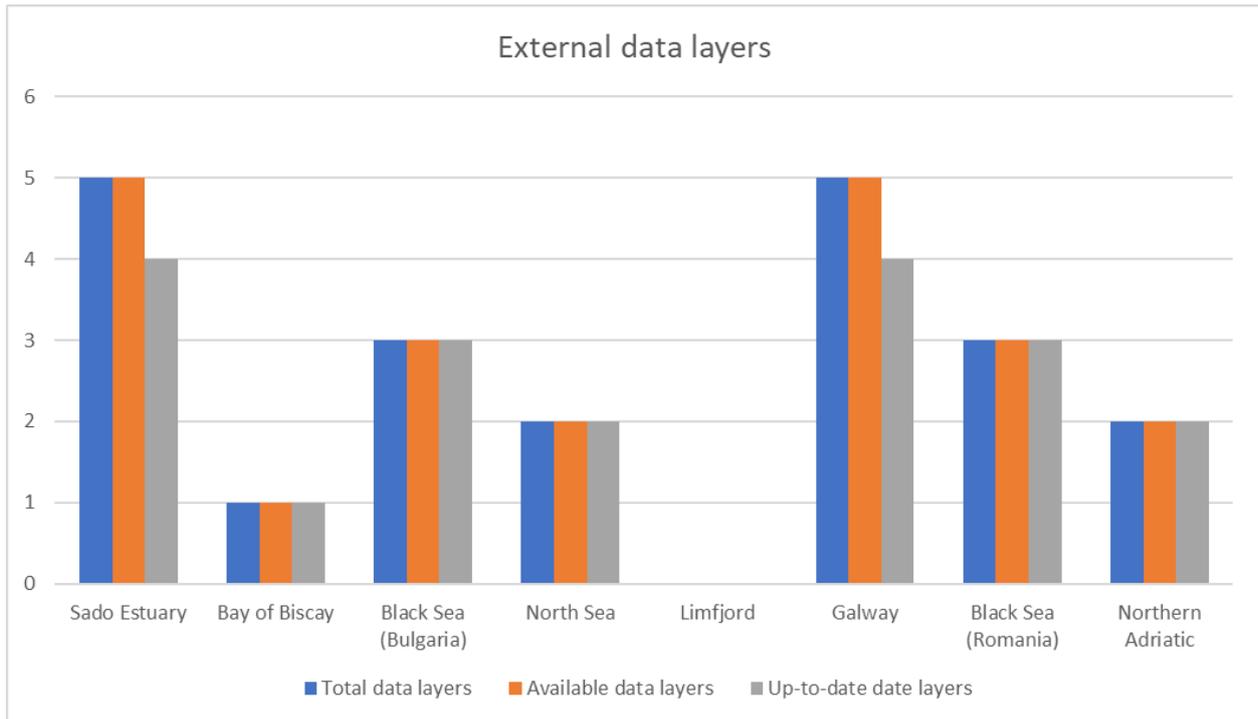


Figure 14. Data layer status from the FORCOAST partners

### 5.3.3 Interruptions

As it can be derived from the information in Section 5.3.1 and 5.3.2, in August 2022 most of the services have been running without any interruptions besides the two days of maintenance of the scheduling system, which did not prevent their correct functioning via the central platform.

The only service that is not working at the moment, since August 25<sup>th</sup>, is Contaminants Source Retrieval due to the implementation of features demanded by the users and subsequent testing before going live. This is expected to take no more than ten days.

### 5.3.4 Timeliness

By definition, timeliness means the time between when the data is available and ready to be used. In the case of the FORCOAST services use, this definition can be approached in two ways.

The first approach is receiving the information (bulletins) via Telegram at the right time via the automated runs. Since the scheduler allows for setting up the time and frequency of those runs, and thus the information retrieval, the timeliness in this regard is instantaneous.

The second approach comes from the service use via the central platform and the definition in this case is the time that takes for each service to run, i.e. time between the user requests the information and becomes available. To determine the time that takes to run each of the services, data on the computation time of the most recent runs was retrieved from the back-end. Those values have been averaged out and the results of that analysis are presented in Table 8.

Service	Number of runs	Average computation time [min]
Marine Conditions	104	2
Land Pollution	51	34.8
Site Prospection	48	0.9
Spat Capture Assistance	37	0.7
Contaminants Source Retrieval	74	14.8
Suitable Fishing Areas	51	5.2
Fronts Detection	90	4.2

Table 8. Average computation time of each service

### 5.3.5 Platform usage

In order to get the necessary data about the use of the platform prototype in terms of users, visits and interaction with the web application, Google Analytics statistics have been gathered covering the period since April 2022.

Figure 15. Daily users of the FORCOAST platform prototype shows the daily evolution in the number of platform prototype users. It shows a positive evolution of increasing users until summer period, where a drop is present. Peak daily value is 33 users, on 27<sup>th</sup> June.

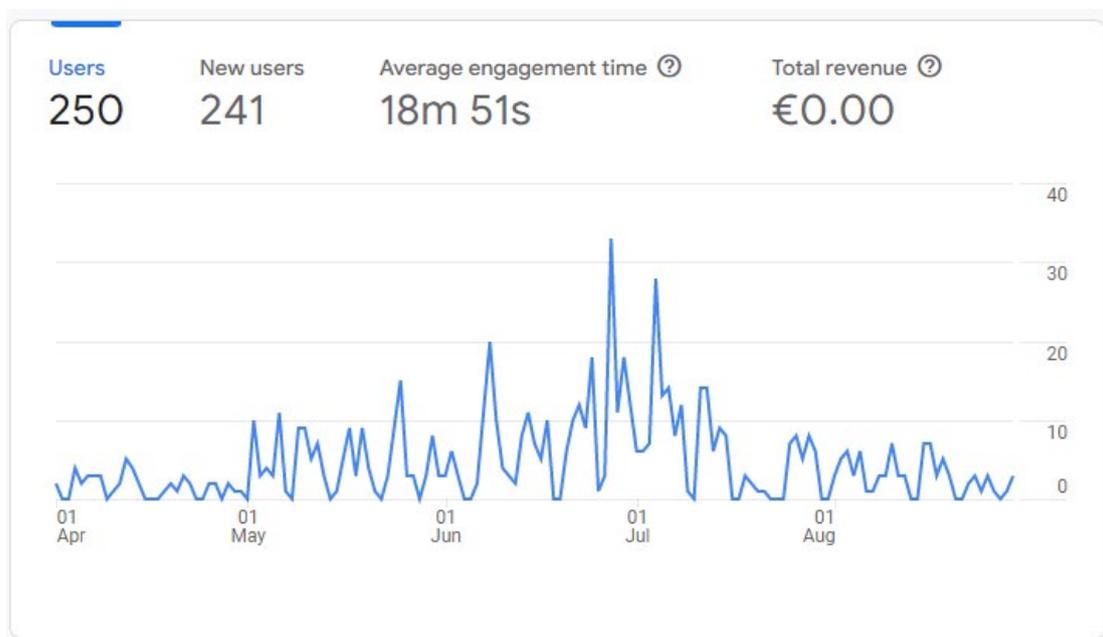


Figure 15. Daily users of the FORCOAST platform prototype

Figure 16 displays the daily evolution of the average engagement time, which is the mean time a user spends in the platform prototype. It is also derived that on average, each user accesses the platform prototype four times and for 3 minutes and 13 seconds in each session.

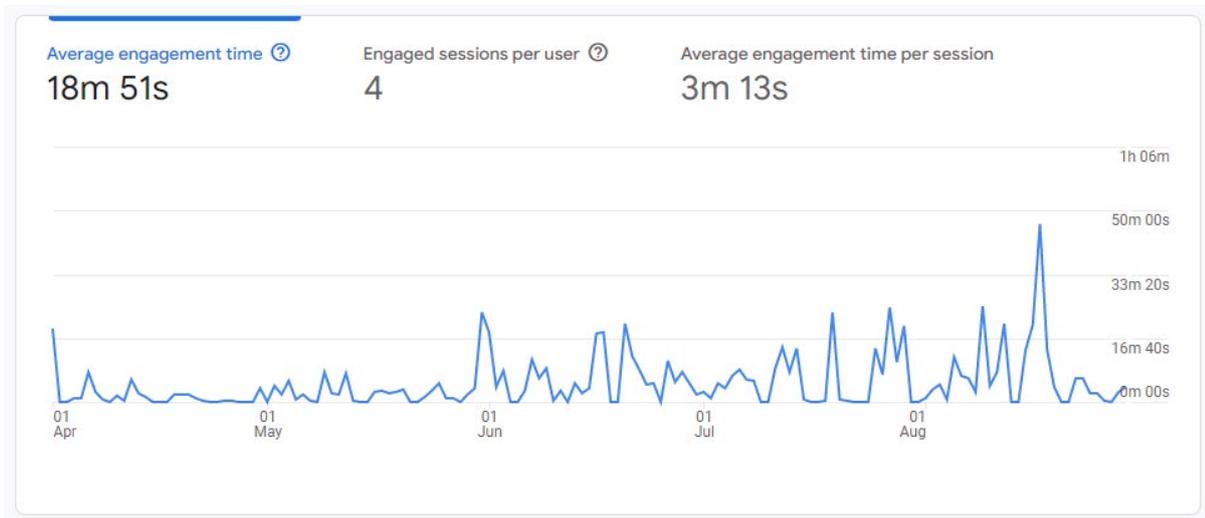


Figure 16. Average user engagement time

## 6. Discussion: fit-for-purpose analysis

### 6.1 SM-A1: Marine condition service

The KPIs evaluated for the SM-A1 Marine Conditions covers some of the most significant high impact issues on marine physical conditions for bivalve mariculture, e.g., storm warning, storm surge warning and SST prediction.

In Limfjorden, DMI forecast KPIs reached the target of the sea storm early warning and SST prediction. For the storm surge warning, the standard alone Limfjord forecast model performs not as good as the two-way nested setup, as shown in D5.4 forecast and hindcast validations. However, it still reached the 10% error criterion for extreme sea level in 6 out of 7 events. Results showed that forecasted extreme sea level is lower than the observed ones at the validation location Lemvig. Such a negative bias can be corrected by using a local forecast protocol.

It should be noted that current KPIs do not cover all the user requirements. For example, sea ice is an important parameter in Limfjord identified by Oyster Boat. Although it is part of the Limfjord ocean forecast model, due to the lack of resources and observations, this product has not been fully validated and optimized. More details on the user requirements on the sea ice can be found in D5.1.

### 6.2 SM-A4: Spat capture service

The validation of KPIs allows to improve parameterization of the service module and seems to show that the product could be used to detect spat arrival. SM-A4, based on the parameterization tested, predicts arrivals at the same time of observation, or before which is useful for bivalve's aquaculture.

Due to the novelty of installation of spat collectors in the Belgian pilot site, uncertainties remain around data, especially on the period where collectors are installed that could lead to a misinterpretation of performance indicators.

In addition, regarding the best parameterization found, it is difficult to test other parameters as PLD. With only one year of data, it is not possible to have a parameterization of PLD and spawning peak for which effect could be confused.

Currently, only 1 year of data is available to assess quality of prediction and to determine the best parameterization. An additional year (2022) will be used to assess model performance and improve parameterization if needed and to further assess the quality of service module with regards to interannual variability. To improve service module performance, additional years are required, if possible, in the site where the service module will be used to improve a site-specific parameterization.

### 6.3 SM-R1

In this service, the KPI is defined as a skill score based on the separation distance between the observed (real) trajectory and the modeled (simulated) trajectory. The skill score presented here constitutes a numerical measurement of the performance of the model in particle-tracking modeling, which is the objective in SM-R1 Retrieve Sources of Contaminants. Therefore, it can be regarded as a KPI for this service, with the first assessment of the KPI's presented in this document. Future drifter deployments can be carried out in the bay to continuously update the value of this KPI to assess any alteration in the model performance. A possible definition for a KPI target could be to ensure that the Liu & Weisberg skill score is above 0.5 for at least  $\frac{2}{3}$  of the drifters deployed in the successive experiments. A field experiment with surface drifters has been conducted in order to qualify the KPI. The results showed that, for 6 out of 8 drifter trajectories, the model

has a KPI >0.5, which means that the model performance on predicting surface drifting of the objects is satisfactory.

#### 6.4 Platform KPIs

The KPIs obtained for the FORCOAST service platform prototype present in section 5.3 prove the suitability of the developed system solution for the end user requirements in terms of information retrieval and delivery system.

Service availability parameters show that the services are fully operational except those times that there is a maintenance or testing of a new functionality/feature, which development tends to be short. These are the only service interruptions experienced in recent times, which due to the finished status of the platform prototype, they will be even more sparse. Input data availability supports the aforementioned point having a continuous supply of timely input data from the required sources: CMEMS and partners' models.

Regarding the timeliness of the service information, when setting automated runs so the user receives bulletins on schedule via Telegram, the timeliness is instantaneous due to the ability to configure the run times and frequency on demand with no limitations. When it comes to the use of the service via the platform prototype, Table 8 displays the average computation time for each service. Note that for the Suitable Fishing Areas and Fronts Detection Services, the result layers are available daily in the front-end, so the computation time does not affect the timeliness since the service is fully automated in the service platform. The rest of the services show a very low computation time in the order of seconds, except for Land Pollution and Contaminants Source Retrieval. An indicative computation time is stated in the front-end for each of the services, which makes it particularly informative in these cases where results take longer to be generated.

Lastly, platform usage has been steadily rising in term of users until the expected summer valley period. This is expected to increase with the last efforts in user engagement, marketing and dissemination. The average retention time of 18 minutes and 51 seconds per user, and 3 minutes and 13 seconds per session makes sense since it gives the user time to access the platform, running the service of their interest, explore the results, and getting engaged towards the full service with the scheduled runs which all of it is the objective of the platform prototype.

#### 6.5 Long-term validation

Product and service quality is one of the major user concerns. Relevant quality information has to be timely updated in the post-FORCOAST service. Currently the quality information is provided by D5.4 for data products and D5.5 for KPIs. Regular quality information should be updated. For model data products, this should aim for monthly or quarterly interim validation, as a practice performed in CMEMS. When the models are updated, extra validation should be provided. For product KPIs, a yearly update is appropriate. For platform KPIs, the update can be monthly, as performed in CMEMS service. However, it should be noted that the validations can be quite resource demanding. In CMEMS, a large amount of resources have been invested in the validation.



## References

Nils Hedberg, Nils Kautsky, Linda Kumblad and Sofia A. Wikström (2018). Limitations of using blue mussel farms as a nutrient reduction measure in the Baltic Sea. Published by Baltic Sea Centre, Stockholm University.

