Chapter 30
Sustainable and Added Value Small Pelagics Fisheries Pilots

Karl-Johan Reite, J. Haugen, F. A. Michelsen, and K. G. Aarsæther

Abstract This chapter describes four pilot cases covering the Norwegian pelagic fisheries for small fish species in the North Atlantic Ocean, such as mackerel, herring and blue whiting. The pilot cases aim to improve sustainability and value creation. Big data methods and tools have been used to demonstrate the potential impact on fuel consumption, fisheries planning and fish stock assessments. Specifically, the pilots have targeted immediate operational choices, short-term fisheries planning, fish stock assessments and longer-term market predictions.

30.1 Introduction

The main challenges for the small pelagic fisheries are related to both the fisheries management and the fisheries itself. Within the fisheries management, one seeks to maximize the production by optimizing the fishing quotas and regulations. At the same time, the resources available for this task are limited. For the fisheries itself, the shipowners want to maximize the value of their fish quotas while minimizing the costs associated with owning and operating their vessels.

The governing bodies (EC and national EU and EEC member states) require fishermen and landing sites by law to report catch data for monitoring purposes. The Norwegian small pelagic fisheries fleet follows the Norwegian law of wild caught fish ('Råfiskloven'), which monopolizes the sale of fish from vessels through sales associations with geographic and species-based areas of monopoly. These sales organizations collect detailed information about species, volume, time of capture, time of lading and price for the entire regional market. This data source is the foundation for the small pelagic fisheries planning and market prediction pilots.

The variations of demands for propulsion and electric energy onboard these ships [1] have led to the development of ships with very advanced energy and propulsion systems. A downside of this development is that the operation of these vessels has become more complex, making it difficult sometimes to take advantage of the

K.-J. Reite (✉) · J. Haugen · F. A. Michelsen · K. G. Aarsæther
SINTEF Ocean, Brattørkaia 17 C, 7010 Trondheim, Norway
e-mail: karlr@sintef.no

possibilities within the systems. The crew is also often engaged in fishing operations, where management of a power plant is not a priority, making decision support systems important [2]. Collecting extensive energy performance of ships and delivering advice based on big data technology is therefore a focus for one of the small pelagic fisheries pilots.

Short-term planning of the fisheries is mainly based on the fishermen's expectations about where they can most efficiently do their fishing. These decisions are mainly based on own experiences, meteorological forecasts and current fisheries activity as it is perceived through catch reports, available AIS data and communication with friendly fishermen on other vessels. Developments in the market situation are considered based on expectations for the amount of catches from other vessels and fish quality. These factors are subjectively considered by the individual fishermen.

Long-term planning involves such decisions as, for instance, catching more herring in the spring to have more time for mackerel fisheries in the autumn, due to expectations of being able to achieve higher mackerel prices in the autumn if one has time to make smaller catches. These decisions are very complex, based on a range of uncertain factors and currently with few tools available for decision support.

The small pelagic fisheries pilots focus on small pelagic species harvesting in the North Atlantic Ocean, with the Norwegian pelagic fishing fleet as the main stakeholder. The stakeholders are represented by the pelagic sales association (Norges Sildesalgslag) and companies which own fishing vessels with fishing rights in the North Atlantic. SINTEF Ocean has established the SINTEF Marine Data Centre in order to test, develop and deploy big data tools such as Apache Mesos, CouchDB and GlusterFS for storage and analysis of the available data.

The small pelagic fisheries pilots are highly dependent on big data, for both modelling the ocean environment and the fish stocks. The datasets, stakeholders and analytic needs are illustrated in Fig. 30.1. The data needed include satellite data
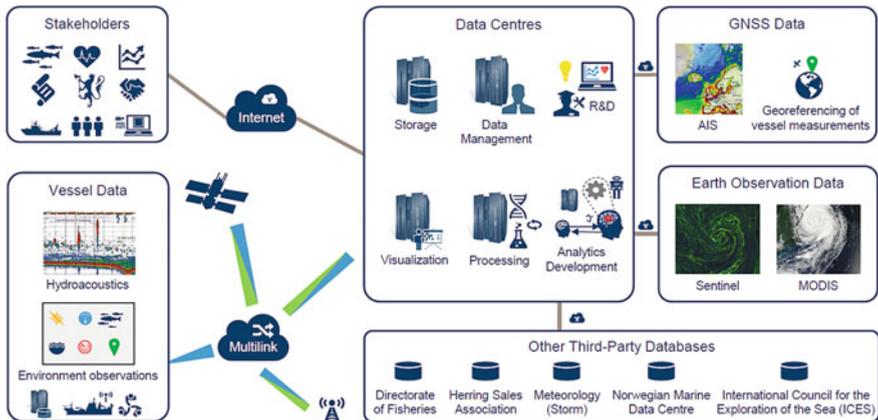

Fig. 30.1 Overview of datasets, stakeholders and components in pelagic fishery

(meteorological and oceanographic), model data (predictions and hindcasts), local measurements (shipborne instruments) and reports on fish catches, for example:

1. Information about all pelagic catches landed in Norway since 2012 is provided by the sales association. This includes information such as price, quantity, catch location, species and size distribution.
2. The ship-owning companies provide onboard measurements (e.g. echo sounders, navigation, machinery and propulsion).
3. Oceanographic hindcasts and daily forecasts are provided by the oceanographic model SINMOD.
4. Satellite-based oceanographic measurements are provided, for example, by CMEMS and NOAA.
5. Meteorological forecasts and hindcasts are provided by the Norwegian Meteorological Institute.

An architectural approach has been chosen, with a focus on the use of case pilots ranging from immediate energy optimization to trip planning and market predictions. The number of potential big data technologies usable for fisheries is vast. The available components and technologies were organized in the framework developed by the Big Data Value Association (BDVA). The potential components were identified during the pilot specification phase and also in the BDVA framework. This selection was refined as the pilot implementation was planned in more detail, ending up with a common architecture design for the pelagic pilots with focuses on the components needed for a minimal viable system, illustrated by the components in the red boxes in Fig. 30.2. The dataset representations are standardized to use JSON for thin data
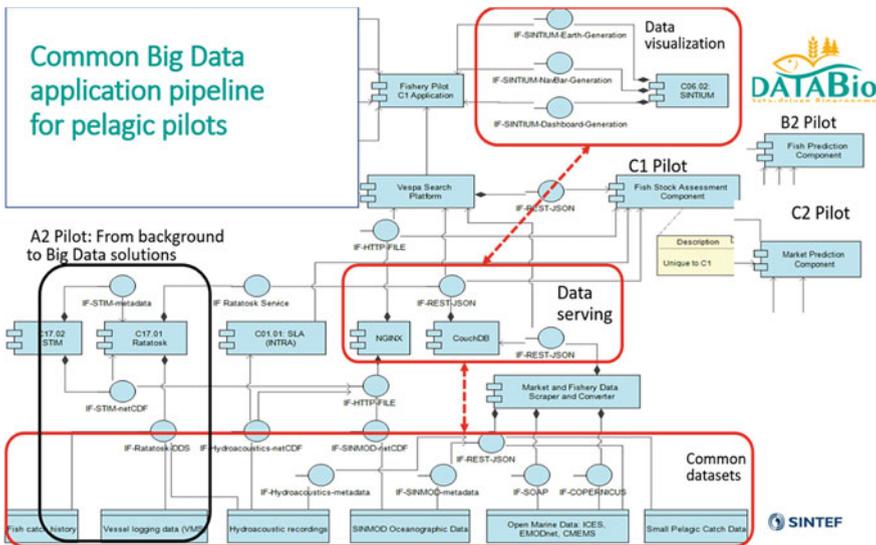

Fig. 30.2 Common architecture for small pelagic pilots

(i.e. catch reports, market and position data) and metadata, while NetCDF is used for large volume data like EO, hydroacoustic and oceanographic data. A combination of search (VESPA) and database technology (CouchDB) is suggested for use in data collation and discovery, both using JSON data representation.

As the small pelagic fisheries pilots had overlapping needs for data centre resources, the provisioning of such resources was a priority. The SINTEF Marine Data Centre was formed for such tasks and therefore chosen as a basis to develop the necessary shared resources for the pilots. The infrastructure includes storage servers, hosting of services and building nodes for software development. A central part of the SINTEF Marine Data Centre is the use of DC/OS for service provisioning and task distribution. This is based on a collection of masters, agents, load balancers and a single bootstrap node. This installation acts like a resource for deploying services in a scalable and repeatable way. It also has functionality for making services available from the Internet without exposing internal systems. The most important services provided by SINTEF Marine Data Centre for the small pelagic pilots are shown in Table 30.1.

The file storage uses the GlusterFS system for creating posix-compliant, replicated network storage. Periodic and dependent jobs are run using the Chronos service running on DC/OS. Vessel data are written by the vessels to an external server ("Incoming"). The data are then fetched to the file storage behind corporate firewalls for further curation, monitoring and analysis. Access is governed by public key cryptography. The high-performance computing cluster unity is used for simulating and predicting oceanographic processes and properties, such as salinity, temperature,

Table 30.1 Services and containers used in SINTEF Marine Data Centre for storage and analysis

| Id	Description
Datafetcher	Responsible for writing vessel data to Gluster
Serverconfig	Responsible for keeping the in-house server configurations updated
Vesselconfig	Responsible for keeping the external server configurations updated
Incoming	External server which the vessels can send their data to
GeoServer	Serves GIS data to map clients
Postgis	Serves GIS data primarily to GeoServer
Glusteraccess	Provides access to the Gluster file storage
Chronos	Responsible for running periodic and dependent jobs
Vesseldatamonitor	Functionality for performing health checks on incoming vessel data
Artifactory	Provides storage of and access to built software
Haproxymain	HAProxy instance providing controlled access to some external services
Aptly	Distribution of Debian installation packages to vessels
CouchDB	Database for local caching of incoming data from external sources
Stimanalysis	A basic Docker container containing various analysis capabilities for analysing vessel operational data

nutrients, plankton and fish stock migrations. This system uses earth observation data, as well as catch reports from the sales association [3, 4].

30.2   Small Pelagic Fisheries Immediate Operational Choices

This pilot aims to improve the operation of relatively complex machinery arrangements onboard small pelagic fishery vessels based on measurements of current state and historic performance. The energy needs of the vessel for propulsion power, deck machinery, fish processing and general consumption are met by the same power generation system, which on newer vessel can be configured to produce and distribute power in a variety of ways. The vessel machinery systems may meet crew requirements in a variety of ways but lack feedback on efficiency or suggested actions to reconfigure power production and distribution. Even if the increasing number of sensors can provide valuable information for crew, fishermen's main focus will always be fish harvesting and not the fine-tuning of complex machinery systems. This can lead to higher fuel consumption than necessary.

The four participating vessels have been equipped with instrumentation for continuous collection of navigation data, power production, fuel consumption and high-frequency motion data, as well as fuel and loading condition data where available. The collected data have been analysed and the vessels integrated into the SINTEF Marine Data Centre infrastructure. The signals recorded onboard the vessels are augmented with synthetic signals for decision support in order to cope with the inherent heterogeneous nature of data collected from different fishing vessels. Datasets are heterogenous due to different engine system layouts, different choices of suppliers for propellers, prime movers and auxiliary engines. The new synthetic signals enable the four vessels to slot into a data collection and processing pipeline in the SINTEF Marine Data Centre. This integration of heterogeneous vessel data, or sensor platforms, into a common system has highlighted the need for feedback of both analysis techniques and synthetic signal generation, but also of updated decision support databases to the vessels from SINTEF Marine Data Centre. The introduction of new signals, real or computed, may necessitate an update from the data centre to the vessels of both signal definitions, analyses and the database on which the decision support is based. The already collected data should not be forgone when making such updates, and a new decision support database should be populated from the data centre to the vessel with new signals and new analyses and decision support possibilities.

The first technological hurdle for the pilot is the implementation of harvesting and retrieval of data from the vessels. The retrieved data are of high value for the future and must be kept securely stored, as if it is lost there is no way of recovering it. The pilot has therefore integrated the measurement system onboard the vessels with the SINTEF Marine Data Centre to store all collected data securely for future use
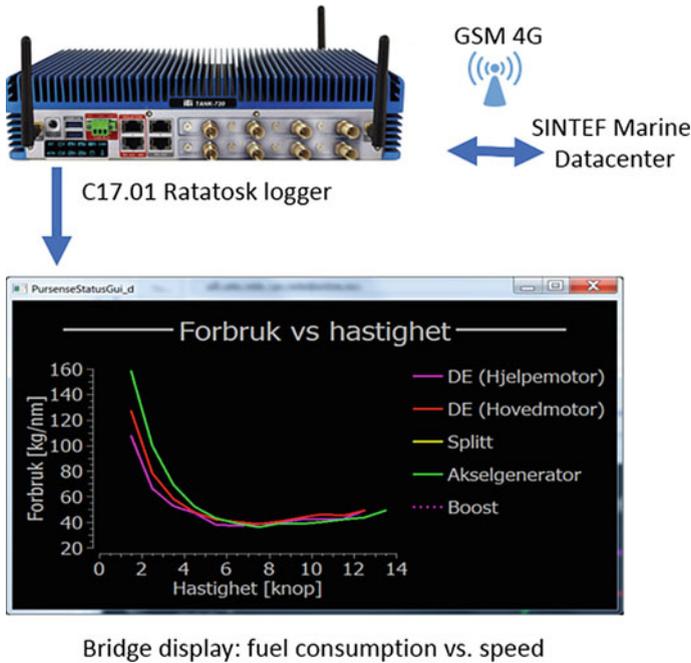
Fig. 30.3 Schematic view of the integration of the vessel's logging computer with SINTEF data centre and screenshot from bride—decision support system for DataBio vessels

and to establish the ability to curate data and update the database of the vessels with synthetic signals derived from the original data as seen in Fig. 30.3. The installed system onboard the vessels accumulates the data and makes a statistical database of the vessel's experienced operations. This database is continuously monitored with the current operation mode in order to give crew a quick feedback when it is practical to operate the vessel in a more efficient manner [5]. This relies on the assumption that the optimum, practical, attainable, operational configuration of the power plant onboard the vessel can often be deduced from its historical data.

30.3 Small Pelagic Fisheries Planning

The main objective of this pilot is to evaluate the effect of utilizing big data technologies in pelagic fisheries planning. The pilot's work focuses on developing services that can help improve vessel operation planning with better fishing ground targeting and improved timing of the fishing execution. The working hypothesis of the pilot is the causality between oceanographic parameters, such as temperature and low-trophic organisms (e.g. *Calanus spp.* copepods), with the location and migration

patterns of pelagic species. Therefore, a useful service would be to visualize oceano-graphic and biology parameters together with historical catch data of various species. The primary pilot's goal was to create a Web portal enabling end users to browse through this information on a map. This includes the ability to select a time period of reported catch data for specific pelagic species, which then are displayed on a map that includes oceanographic attributes. A playback feature lets the user see the time evolution of the selected attributes.

The fishing operation region for which the pilot provides decision support includes large portions of the Norwegian Sea and the North Sea, totalling approximately 1.5 million square kilometres. Pelagic fisheries usually only operate in small subregions of this area, depending on targeted species.

The consortium involved in this pilot consists of:

- SINTEF Ocean is a contract research organization committed to technical research within marine applications. SINTEF Ocean leads the pilot and is also the main contributing research organization.
- Norges Sildesalgslag (Norwegian Fishermen's Sales Organization for Pelagic Fish) is a sales organization, owned and operated by fishermen (a coopera-tive), selling fish on a first-hand basis from fishermen to buyers—for further sales/export. They contribute with knowledge and accumulated data on fish catches.
- The fishing vessel owners Liegruppen Fiskeri, Eros, Ervik & Sævik and Kings Bay operate in fisheries targeting pelagic fish species in the North Atlantic. Their role in this pilot is to contribute with their knowledge about fisheries planning and to serve as an end user for the pilot's Web portal.

Important activities in the pilot have been to identify *Data Sources*, select appro-priate components/assets and configure necessary *Data Management* and *Data Processing Architecture*. This work facilitated the primary goal of the project, namely provisioning of the Web portal and its *Data Visualization*. Definitions of key perfor-mance indicators that directly quantify the fishery operation performance were quickly dismissed, because any evaluation of such indicators depends on unmeasur-able and non-deterministic factors. Any potentially improved measurement of fishery efficiency can only be speculatively attributed to the introduction of the pilot service. As a consequence, "key performance indicators" were instead defined as measurable progress/completeness of the technological components used in the pilot.

The following technologies have been found relevant for this pilot:

- **SaltStack** provides configuration management of data centre servers, facilitating version control and remote access.
- **Docker** provides containerization and facilitates version control of onshore systems.
- **SINMOD** provides biomarine simulations and simulation of fish migrations.
- **DC/OS** provides container orchestration and communication.
- **CouchDB** provides storage of and access to catch data.

- **GlusterFS** provides replicated and distributed storage of and access to collected data and the results of biomarine simulations.
- **KRAKK** provides data scraping functionality, especially for data from Sildelaget.
- **GeoServer** provides an open source server for sharing geospatial data.
- **Python** scripts that make use of RESTful API and GDAL for ingesting SINMOD oceanographic and biology data rasters into GeoServer.
- **Python Flask** is used as a Web Server Gateway Interface (WSGI) Web application framework to develop the Web portal.
- **uWSGI** is used for serving the Web portal.
- **Crossfilter, D3.js, dc.js and Leaflet** are important JavaScript libraries for presenting data in the Web portal.

The implementation of this pilot is based on a number of data sources:

- **Catch data** are made available by Sildelaget through an API developed by Sildelaget for DataBio. This API makes available all pelagic catches landed in Norway since 2012, and it is continuously updated as new catches are landed. This provides locations, amounts and price for each catch. The catch data from Sildelaget is proprietary datasets that will not be available after the project. On the other hand, the Norwegian Directorate of Fisheries recently open sourced catch data historic records.
- **SINMOD** oceanographic and biological hindcast and forecast data for the Norwegian Basin, including temperature, salinity, ice thickness and concentration, NO3, *Calanus finmarchicus*, *C. glacialis and* chlorophyll. These parameters were provided both historically, since 2012, and regularly with short-term forecasts two days into the future with a spatial resolution of 4 km in polar stereographic projection. The SINMOD data source relies on several satellite and buoy-based inputs, and see the next pilot for details (Fig. 30.4).

The SINMOD operationalization produces NetCDF4 files that largely follow the Climate and Forecast Convention 1.5. Nonetheless, there have been several issues related to standardized naming conventions of the variables, consistent spatial resolution, as well as correct projection parameters between the historic and predictive datasets. The process of making SINMOD data available to the map service involves extraction of selected depths and timepoints so that only relevant data are being served by GeoServer. Instead of using the NetCDF plugin of GeoServer, we rather used GDAL to manually reproject NetCDF files into the destination projection as GeoTIFF files. File handling logic was developed to facilitate ingesting large datasets. GeoServer's built-in colorbar legend currently lacks the necessary flexibility to show customized styling in a satisfactory manner, which again warranted manual customization. GeoWebCache, the built-in tile caching integration, does not play well with periodic regeneration of new rasters. This is at least true when using GeoWebCache REST API and CQL filters to selectively "reseed" updated datasets. We experienced intermittent issues with newly ingested rasters, where it cached transparent tiles, probably because tiles were cached before their ingestion into the
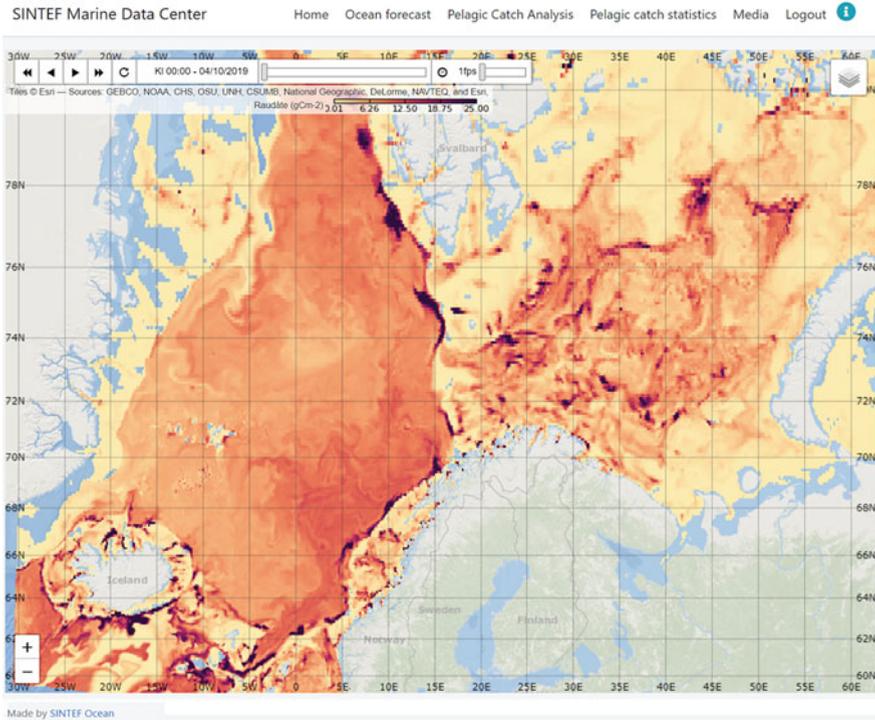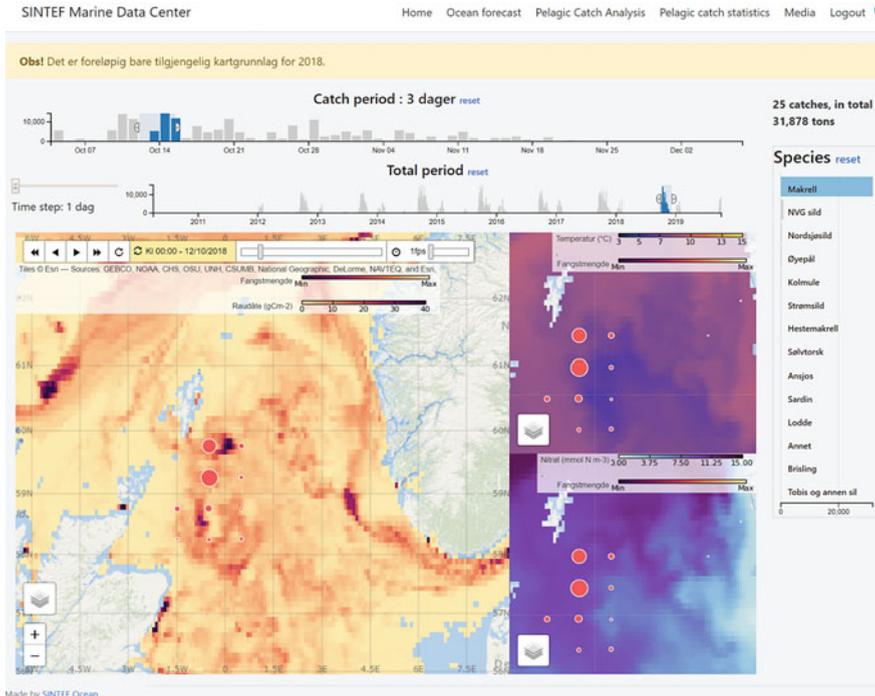
Fig. 30.4 Web portal: Calanus finmarchicus concentration distribution

PostGIS database was done. This issue was not easily reproducible, nor did it produce any error messages, causing undetected issues with the Web map service (Fig. 30.5).

We chose tiled WMS to serve the raster data. The styling of the layers was done on the server side, so no styling configurations were needed in the Web application Leaflet. Designing styles that work globally for a single attribute all year round is challenging, because of the span of interesting values changing throughout the year. WMS playback was achieved using a Leaflet plugin, but the flexibility in zoom levels with different tiles made it challenging for the plugin to buffer many timepoints in a manner that enabled good user experience. Some browser caching occurred, as well as server-side caching, but a different choice of technology or data format may have improved the UX smoothness.

We estimate the impact of the new service provided by this pilot to be minimal; that is, the pilot's end users do not yet actively use of the Web portal for their fishery planning. The reason for this is multifaceted. First, the time period for which the service has been available, with fair service reliability, is very brief still. The user experience in these initial versions of the Web application can be frustrating, due to sluggishness and lack of responsiveness. There is a lack of fundamental features that could be of interest for the user to check for specifically interesting phenomena. For

Fig. 30.5 Web portal: Catch data together with temperature, nitrate and Calanus finmarchicus

example, a simple extension would be the ability to select a region and provide key information/analysis on demand. The portal was specifically designed for desktop application use, but in hindsight it should have been readily available on all platforms, including smartphones and tablets. The UX design could also have been more targeted to specific use cases. For instance, by providing several subpages, each designed to provide a very limited set of information. One such tailored design could lower the threshold for use.

The pilot was designed on top of systems and infrastructure designed for use in production. DC/OS are made ready for production use cases, which includes scalability, load-balancing, resource management, etc. What the pilot technology design does not cover is situations in which users employ low bandwidth networks, which is often the case for ocean-going fishing vessels. Therefore, the Web portal is more practical and applicable in an onshore, by-the-computer setting, with high-quality bandwidth. We believe that despite these initial challenges, the concept of collating information and providing insight into multi-origin data in a clear manner still has great potential for improving fishery planning. Establishment of a minimally viable product that the end user is interested in could spawn the foundation for future applications that have a large impact on how fishermen make use of big data and technology in planning their operations.

30.4 Small Pelagic Fish Stock Assessment

Pelagic fish stock assessments are traditionally based on a combination of research cruises with dedicated research vessels, catch statistics and non-spatial stock models. These methods are criticized for low cost efficiency, being based on too few measurements and unable to adapt to rapid climate change effects. The objective of this pilot has been to demonstrate that the combination of information from a great variety of assets can be used to produce better population dynamics estimates for pelagic species. Specifically, crowd-sourced data collection effort from fishing vessels combined with public/private data assets, biomarine modelling and data analytics are assumed to be able to increase both the accuracy and precision of fish migration and stock assessments.

The pilot has concentrated on three research questions:

1. How can hydroacoustic data be cost-efficiently collected from a fleet of fishing vessels?
2. How can a fleet of fishing vessels be part of a crowd-sourced data collection system?
3. How can biomarine modelling and spatio-temporal modelling of pelagic species be used for stock assessments?

To cost-efficiently collect hydroacoustic data from fishing vessels, the integration against existing hydroacoustic sensors was important. Due to the large variations in equipment and interfaces, as well as lack of interface possibilities for much of this equipment, this proved to be a serious challenge. The pilot created a preliminary interface against one type of equipment, but cost-efficient integration against the hydroacoustic equipment of a substantial part of the fishing fleet is not solved.

To make a fleet of fishing vessels part of a crowd-sourced data collection system, cost-efficient installation and maintenance in the vessel are needed. The most important challenges are the variation in vessel systems, sensors and their set-up, as well as how these change over time. This pilot addressed these challenges by using configuration management systems using version-controlled configuration descriptions. This gave a way to perform remote maintenance, updating and reconfiguration, as well as simplify initial installations.

To model the fish stocks and their behaviour, both adequate biomarine models and correction of these, based on measurements, are needed. This pilot developed a preliminary migration model of one pelagic species. Also, a preliminary method for correcting this model using data assimilation was developed, and this correction was performed based on historical data. The results showed that more data for correction are needed, and this has become the focus of new research initiatives.

The consortium involved in this pilot consists of:

- SINTEF Ocean is a contract research organization committed to technical research within marine applications. SINTEF Ocean leads the pilot and is also the main contributing research organization.

- INTRASOFT International offers IT solutions to a wide range of international and national public and private organizations. INTRASOFT has performed comparisons of different methods for classification of hydroacoustic measurements.
- Norges Sildesalgslag (Norwegian Fishermen's Sales Organization for Pelagic Fish) is a sales organization, owned and operated by fishermen (a cooperative), selling fish at a first-hand basis from fishermen to buyers—for further sales/export. They contribute with knowledge and accumulated data on fish catches.
- The fishing vessel owners Liegruppen Fiskeri, Eros, Ervik & Sævik and Kings Bay operate in fisheries targeting pelagic fish species in the North Atlantic. Their role in this pilot is to contribute with their knowledge about fish migration patterns and how this is observed from the fishing vessels, as well as the technical installations available onboard the fishing vessels.

This DataBio pilot has been aimed at assessing if and how stock assessments of pelagic fish species could benefit from low-cost data collection during fishing vessels' day-to-day normal operations, combined with biomarine simulations and migration pattern simulations of pelagic fish species. To this end, this pilot aimed at developing a demonstration version of an infrastructure consisting of both vessels and shore systems.

Relating to the above specified research questions, the following technologies have been found to be relevant for this pilot and its implementation:

- **SaltStack** provides configuration management of both shore servers and vessel equipment, facilitating version control and remote access.
- **Ratatosk** provides onboard data acquisition, data exchange and monitoring of these functions.
- **STIM** provides efficient analysis of collected data (except for hydroacoustic data).
- **Docker** provides containerization and facilitates version control of onshore systems.
- **SINMOD** provides biomarine simulations and simulation of fish migrations.
- **Ratacoustics** provides integration between hydroacoustic equipment and Ratatosk.
- **DC/OS** provides container orchestration and communication.
- **CouchDB** provides storage of and access to catch data.
- **GlusterFS** provides replicated and distributed storage of and access to collected data and the results of biomarine simulations.
- **KRAKK** provides data scraping functionality, especially for data from Sildelaget.

The implementation of this pilot is based on a number of data sources:

- **Catch data** are made available by Sildelaget through an API developed by Sildelaget for DataBio. This API makes available all pelagic catches landed in Norway since 2012, and it is continuously updated as new catches are landed.
- **Hydroacoustic data** are found to be important for correcting the biomarine models and the fish migration model. Some data have been collected using ad hoc methods, but creating general tools for large-scale deployment has proved to be challenging.

- **Vessel operational data** are important for determining what the hydroacoustic data represent in both time and space. Also, for example, ship motions can be important for interpreting the data. The vessels Eros, Kings Bay, Ligrunn and Christina E are contributing with such data.
- **Global ocean tidal components** M2, S2, N2, K2, K1, O1, P1, Q1, Mf, Mm and SSa at the open boundaries of the SINMOD model are imported from [6], which is based on [7].
- **Boundary conditions** for the large-scale 20 km model are acquired from the Mercator Global Ocean model system.
- **Atmospheric input for the large-scale models** is acquired from NOAA Global Forecast System.
- **Atmospheric input for the local scale models** is provided by the Norwegian Meteorological Institute from the 2.5 km MetCoOp EPS system.
- **Sea surface temperatures** are downloaded from the product METOFFICE-GLO-SST-L4-NRT-OBS-SKIN-DIU-FV01.1 [8].

The selected technologies seem to be adequate for the tasks, and there are no obvious benefits associated with making technology changes. But as there are possible alternatives for most of them, the final choice is as much dependent on preferences and existing tools as on the task itself. Without loss of benefits, one may, for example, replace SaltStack with Ansible, Puppet or Chef; Docker could be replaced by Mesos Containerizer; DC/OS could be replaced by Mesos or Kubernetes; CouchDB could be replaced by another database or file storage; GlusterFS could be replaced by Ceph. But for now, no clear benefits are seen from making such changes in the choice of technologies.

One possible exception is with the hydroacoustic data collection, where a Simrad echo sounder was used in the project. This echo sounder facilitates two main approaches for collecting hydroacoustic data in a systematic manner. One is to use the record functionality in the graphical user interface, and the other one is through a subscription-based application programming interface. The first approach is simplistic in that a vessel crew member basically pushes a record button and the system will record data. The downside is that it requires human intervention from the crew, and real-time processing is cumbersome. At the beginning of the project, it was deemed as a risky approach. Therefore, it was decided that API-based data acquisition was a more robust and long-term investment and better suited as an extension of the existing data acquisition system (Ratatosk), as visualized in Fig. 30.6. The subscription API is a comprehensive implementation that enables access to processed and unprocessed data streams and parameters using Ethernet User Datagram Protocol (UDP). Our approach is to implement this subscription API and make the data streams available to the Ratatosk logging component, enabling both real-time processing and storage to file. Most of the functionality towards the subscription API is in place, but the adaptations to connect to the Ratatosk component are currently lacking.

The currently available hydroacoustic echo sounder dataset, see snapshot in Fig. 30.7, has been used as a preliminary comparison of classification methods. The dataset consists of five hydroacoustic frequencies (18, 38, 70, 120 and 200 kHz),

Fig. 30.6 Extension of the
vessel logging system to
facilitate logging of
hydroacoustic data
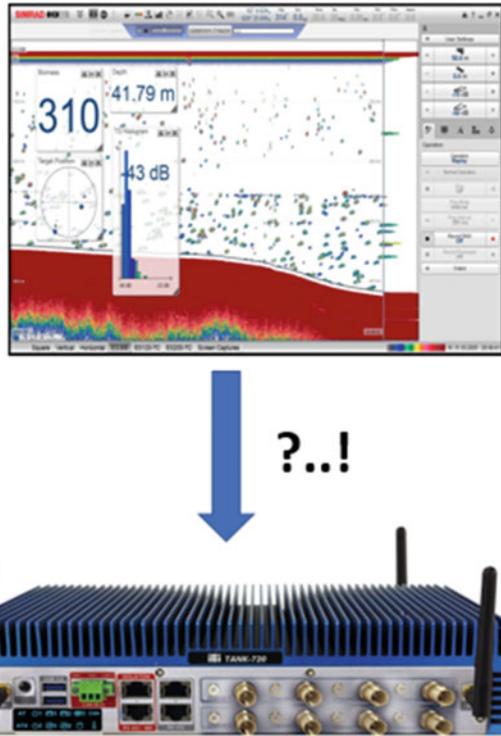
Fig. 30.7 Snapshot excerpt
of echo sounder dataset
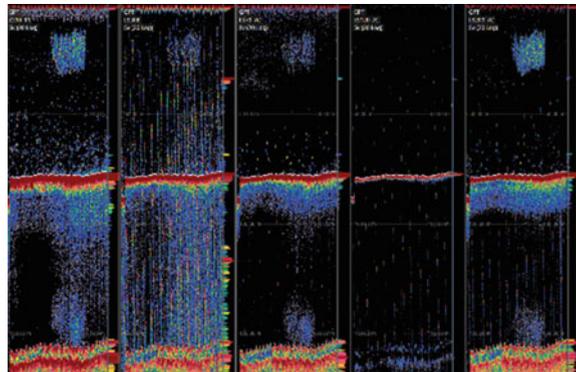
which are computed into mean volume backscatter strengths. Four different algorithms have been tested on the dataset: Naïve Bayes, k-nearest neighbours, support vector machine and principal component analysis. The goal is divided into two tasks:

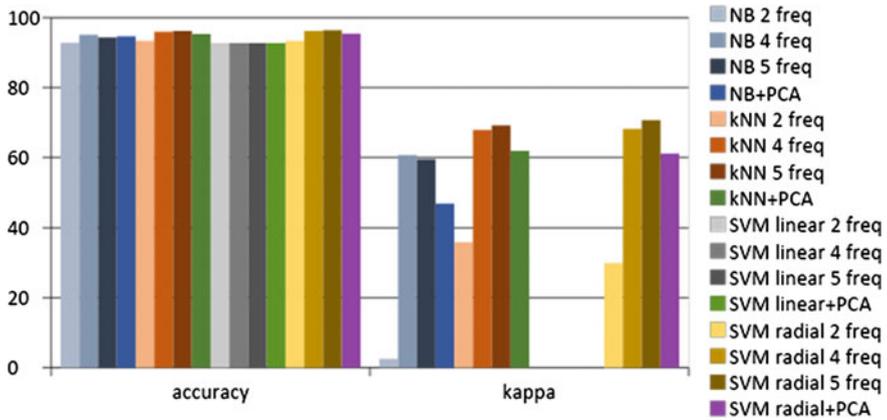i.   Identify and remove seabed echoes and determine fish shoal presence.

Fig. 30.8 Comparison of classification methods

ii. Discriminate plankton from fish, identify fish species, and perform a biomass evaluation.

Figure 30.8 shows that accuracy is high for all tested methods, but this is due to the few positives of the dataset. Kappa is a more sophisticated metric that shows how much the algorithm improves the average expected accuracy. Kappa shows more varying results when comparing the different methods.

For simulation of the marine ecosystem and the migratory behaviour of selected species, the tool SINMOD was used. This tool perfectly suits the task, as it is able to integrate the simulation of oceanography, low-trophic biology and how this affects higher-order processes. For demonstration purposes, a preliminary fish migration model for herring (*Clupea harengus*) was developed, based on simple behavioural rules and corrected by reported catches. Even if very simplified, the model was able to recreate migration patterns. The model will need to be developed further before it can provide actual value for fish stock assessments, but the results are promising.

The aim of this pilot was to demonstrate that the combination of data collection, existing datasets and biomarine simulations can benefit pelagic fish stock assessments. The business value of this pilot will only materialize once the developed methodologies and technologies become integrated into the fish stock assessment process. At that time, the business impact of reducing the inherent uncertainty associated with stock assessments and thereby improving management and production of the oceans can be very large. If, for example, the production (and thus the catch) of pelagic fish species could be increased by say 10% as a result of this work, this would amount to approximately a € 60 million increase in first-hand value of pelagic fish species in Norway alone.

As stated above, alternatives exist for many of the technologies used in this pilot. Still, the combination of provided functionalities is a good fit for the pilot's objectives. Most notably, the abilities of such a system are to:

- Adapt to the great variations of sensors and configurations onboard fishing vessels, as well as introduced changes over time. This includes both hydroacoustic equipment and operational sensors, such as motion reference systems and GPS.
- Handle a large fleet of vessels in a structured way, with respect to installation, configuration, maintenance and data collection.
- Simulate oceanography, marine biology and fish migrations, while assimilating available data for model and output corrections.
- Extract useful information from hydroacoustic equipment with respect to, for example, fish species and amount of fish.
- Provide systems for data flow, analysis and storage which are suitable for large-scale deployment.

Most of the systems and infrastructure developed in the pilot are ready for use in production, and many of these are easily available. But for such a system to really have an impact on fish stock assessment, improvements are needed in the interpretation of hydroacoustic data and the fish migration modelling.

30.5 Small Pelagic Market Predictions and Traceability

Norwegian fishermen in the pelagic sector work in fisheries for different pelagic species. The timing for these fisheries is to some extent determined by the availability of fish species and their migrations. In addition, to some extent, the shipowners make strategic decisions about when and where to do their fishing based on expectations of both market development and fishing possibilities. These are important choices, but there is a lack of tools helping the fishermen select the right one.

Preliminary exploratory analyses for mackerel showed expected seasonal variations, as well as other variations so far unexplained. Figure 30.9 shows daily average mackerel price variations and daily catch from 2012 to 2019 for Norwegian mackerel landings. Only the second half of each year is plotted, as this is the main season for this fishery. The size of each point marker reflects the amount of daily/weekly catch. The seasonal variations are obvious, while the variations with other variables in this dataset other than time are not.

The goal of this pilot is to enable fishermen to make the right strategic decisions, which can make a substantial difference in both profitability and landed quality.

The consortium involved in this pilot consists of:

- SINTEF Ocean is a contract research organization committed to technical research within marine applications. SINTEF Ocean leads the pilot and is also the main contributing research organization.
- Norges Sildesalgslag (Norwegian Fishermen's Sales Organization for Pelagic Fish) is a sales organization, owned and operated by fishermen (a cooperative), selling fish at a first-hand basis from fishermen to buyers—for further sales/export.
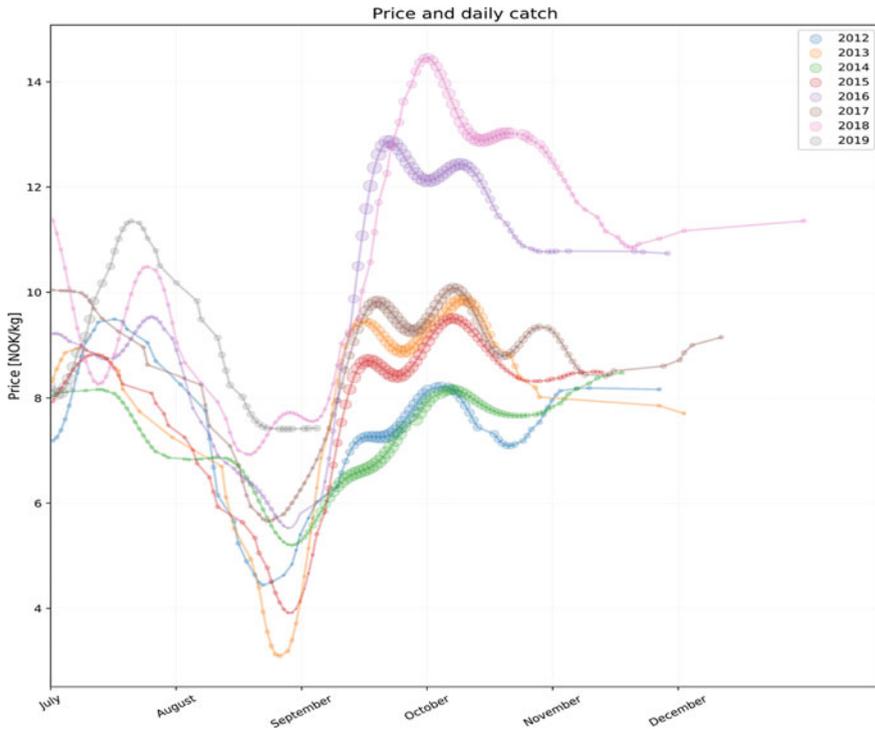
Fig. 30.9 Seasonal variations of Norwegian mackerel prices from 2012 to 2019

They contribute with knowledge and historic and present data on mackerel catches and price.

- The fishing vessel owners Liegruppen Fiskeri, Eros, Ervik & Sævik and Kings Bay operate in fisheries targeting pelagic fish species in the North Atlantic. Their role in this pilot is to contribute with their knowledge on mackerel fisheries and the pelagic market.

This pilot has developed a Web portal to provide fishermen with the tools to analyse historical data. In addition, machine learning has been employed to predict the development of pelagic market segments, so that the fisheries may be targeted based on the species that will allow the highest yield given a predicted economic outlook. The Norwegian mackerel market has been used as a case benchmark, as this is an important pelagic species with large price fluctuations. The basis for the market predictions has been to combine different data sources relevant for price development, such as time, season, predicted catch volume and financial data. Machine learning and predictive analytics have been used to model the relationship between market development and other factors. These models can then be used to provide predictions for how the market will develop in future.

Relating to the above specified research questions, the following technologies have been found to be relevant for this pilot and its implementation:

- **SaltStack** provides configuration management of shore servers, facilitating version control and remote access.
- **Docker** provides containerization and facilitates version control of onshore systems.
- **DC/OS** provides container orchestration and communication.
- **CouchDB** provides storage of and access to catch data.
- **GlusterFS** provides replicated and distributed storage of and access to collected data.
- **KRAKK** provides data scraping functionality, especially for data from Sildelaget.
- **Python Flask** is used as a Web Server Gateway Interface (WSGI) Web application framework to develop the Web portal.
- **scikit-learn** and **Keras** are important Python libraries used for training prediction models.
- **uWSGI** is used for serving the Web portal.
- **Crossfilter, D3.js, dc.js and Leaflet** are important JavaScript libraries for analysing and presenting results in the Web portal.

The implementation of this pilot is based on a number of data sources:

- **Catch data** are made available by Sildelaget through an API developed by Sildelaget for DataBio. This API makes available all pelagic catches landed in Norway since 2012, and it is continuously updated as new catches are landed. This provides locations, amounts and price for each catch. Each catch is typically defined in terms of approximately 70 variables, such as catch size, where it is caught, sale price, storage method and sales method.
- **Catch areas and other definitions** are provided by the Norwegian Fisheries Directorate, such as definitions of various codes representing fish species, catch areas, conservation methods, storage methods, seller, vessel and so on. These data are necessary to interpret the data from Sildelaget.
- **Historical value exchange rates** are made available by the Norwegian bank DNB. These data are potentially valuable for interpreting and forecasting market variations [9].
- **World Bank, EMODnet, Comtrade, Eumofa, Eurostat, ICES and Statistics Norway** offer various data which can be of interest when developing price forecasts for pelagic species. Data scrapers have been developed for these data sources to use in price prediction pipelines.

The selected technologies seem to be adequate for the tasks, and there are no obvious benefits from making additional technology changes. But as there are possible alternatives for most of them, the final choice is as much dependent on preferences and existing tools at the time, as on the task itself.

In a case study, the possibilities for direct predictions of the mackerel prices were investigated. The focus was on long-term predictions, aiming to enable fishermen to adopt long-term successful strategic decisions. As the market is greatly influenced

by unpredictable psychological factors, the results were not expected to be good. This can be compared to predicting the stock market, which understandably is a close-to-impossible task.

A Web portal was developed to allow fishermen to investigate how prices have developed with factors such as species, landed quanta, year, time of year, moon phase and catch location. This Web portal is based on providing the possibility to filter historical catch data along the relevant factors. For example, by selecting only last year's catches of mackerel using a short time window, and then slide this window to see how the prices varied with time. Also, similar procedures can be employed to consider variation with moon phase. Or one can use the opposite approach and select only the catches giving the highest prices to investigate under which circumstances high prices were achieved (Fig. 30.10).

The service developed in this pilot is, as far as we know, the first of its kind. It is notably difficult to estimate the potential business impact. Even if one can investigate how fisheries have historically performed, any changes in fishery timing would influence the market, and we do not know how efficient the fishery could be predicted for alternative timings. As an example, in 2015, the price distribution for herring in the spring (66,000 tons) and in the autumn (119,000 tons) is shown in Fig. 30.11. If one assumes that the market would not be affected by shifting the fisheries to autumn, and that the fisheries could be performed in autumn without affecting other fisheries, a 10% shift of this fishery to autumn would approximately generate an extra 700,000 €.
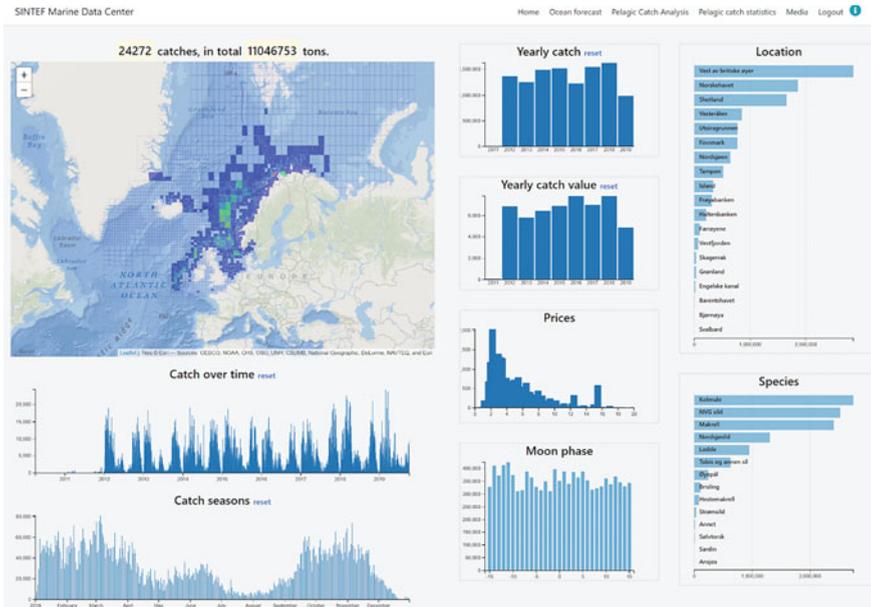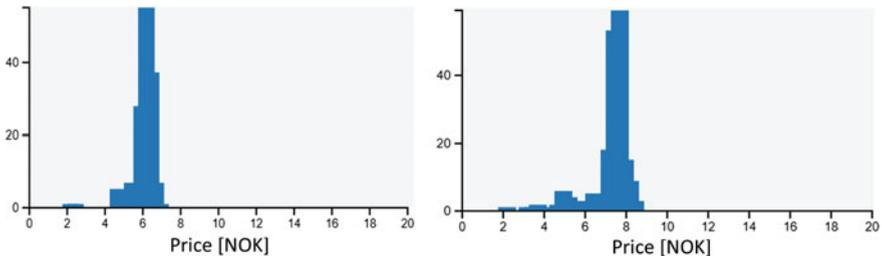

Fig. 30.10 Filtering of historical catch and price data facilitated in the Web portal

Fig. 30.11 Changes in Norwegian mackerel prices between spring and autumn 2015

References

1. Aursand, I. G., Digre, H., Ladstein, J., Kyllingstad, L. T., Erikson, U. G., Tveit, G. M., Backi, C. J., & Reite, K. J. (2015). Development and assessment of novel technologies improving the fishing operation and on board processing with respect to environmental impact and fish quality (DANTEQ).

2. Skjong, S., Kyllingstad, L. T., Reite, K. J., Haugen, J., Ladstein, J., & Aarsæther, K. G. (2019). Generic on-board decision support system framework for marine operations. In *Proceedings of the international conference on offshore mechanics and arctic engineering—OMAE*. https://doi.org/10.1115/OMAE2019-95146

3. Slagstad, D., & McClimans, T. A. (2005). Modeling the ecosystem dynamics of the Barents sea including the marginal ice zone: I. Physical and chemical oceanography. *Journal of Marine Systems, 58,* 1–18.

4. Wassmann, P., Slagstad, D., Riser, C. W., & Reigstad, M. (2006). Modelling the ecosystem dynamics of the Barents Sea including the marginal ice zone: II. Carbon flux and interannual variability. *Journal of Marine Systems, 59,* 1–24.

5. Reite, K.-J., Ladstein, J., & Haugen, J. (2017). Data-driven real-time decision support and its application to hybrid propulsion systems. In *Proceedings of the international conference on offshore mechanics and arctic engineering—OMAE* (Vol. 7B-2017, p. V07BT06A024). https://doi.org/10.1115/OMAE201761031

6. Egbert & Erofeeva. (2019). The TPXO 7.2-model. https://volkov.oce.orst.edu/tides/TPXO7.2.html

7. Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology, 19*(2), 183–204.

8. Copernicus. (2019). Copernicus marine environment monitoring service. https://nrt.cmems-du.eu/motu-web/Motu
9. DnB Markets. (2019). Historical valuta exchange rates. https://www.dnb.no/bedrift/markets/valuta-renter/valutakurser-og-renter/HistoriskeValutakurser/Hovedvalutaerdaglig/Historikk