

# WORKING GROUP ON MACHINE LEARNING IN MARINE SCIENCE (WGMLEARN)

VOLUME 1 | ISSUE 45

ICES SCIENTIFIC REPORTS

RAPPORTS  
SCIENTIFIQUES DU CIEM



## International Council for the Exploration of the Sea Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46  
DK-1553 Copenhagen V  
Denmark  
Telephone (+45) 33 38 67 00  
Telefax (+45) 33 93 42 15  
[www.ices.dk](http://www.ices.dk)  
[info@ices.dk](mailto:info@ices.dk)

The material in this report may be reused for non-commercial purposes using the recommended citation. ICES may only grant usage rights of information, data, images, graphs, etc. of which it has ownership. For other third-party material cited in this report, you must contact the original copyright holder for permission. For citation of datasets or use of data to be included in other databases, please refer to the latest ICES data policy on ICES website. All extracts must be acknowledged. For other reproduction requests please contact the General Secretary.

This document is the product of an expert group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the view of the Council.

ISSN number: 2618-1371 | © 2019 International Council for the Exploration of the Sea

# ICES Scientific Reports

Volume 1 | Issue 45

## WORKING GROUP ON MACHINE LEARNING IN MARINE SCIENCE (WGM-LEARN)

### Recommended format for purpose of citation:

ICES. 2019. Working Group on Machine Learning in Marine Science (WGMLEARN).  
ICES Scientific Reports. 1:45. 13 pp. <http://doi.org/10.17895/ices.pub.5539>

### Editors

Ketil Malde • Jean-Olivier Irisson

### Authors

Alessandra Gomes • Bernhard Keuhn • Cédric Jamet • Edwin Van Helmond • Jean-Baptiste Romagnan • Jose Fernandes • Julien Simon • Laura Hoebeke • Laura Uusitalo • Linh Nguyen • Madiop Lo • Michiel Stock • Perikilis Panagiotidis • Rainer Kiko • Raphaëlle Sauzede • Rubens Lopes • Tristan Cordier • Vincent Rossi • William Michaels

# Contents

i	Executive summary .....	ii
ii	Expert group information .....	iii
1	Terms of Reference .....	1
2	Invited Presentations .....	2
2.1	T Cordier – Topic introduction: Genomics .....	2
2.2	C Jamet – Topic introduction: Remote sensing.....	2
2.3	P Panagiotidis – Demo of ICES data and services .....	3
3	Participants Presentations.....	4
3.1	J Simon <i>et al.</i> – Giving Artificial Monitoring intelligence to Fishing TRAWLS (Game of Trawls).....	4
3.2	E van Helmond and L. Nguyen <i>et al.</i> – Computer vision technologies in Dutch fisheries.....	4
3.3	L Hoebeke <i>et al.</i> – Automated hierarchical classification of animal species in camera trap images .....	4
3.4	K Malde <i>et al.</i> – Vector space embedding for plankton images .....	5
3.5	J Fernandes <i>et al.</i> – Machine learning is not only for big data and big data is not only a lot of data: successful examples in marine sciences with sparse and heterogeneous data.....	5
3.6	R Sauzede <i>et al.</i> – Machine learning-based methods to derive biogeochemical parameters from profiling floats and satellite data.....	6
3.7	M Lo <i>et al.</i> – Deep learning to improve remote-sensed ocean data for fisheries research .....	6
3.8	M Stock <i>et al.</i> – Pairwise learning to predict species interaction networks .....	7
3.9	L Uusitalo <i>et al.</i> – Hidden variables in a Dynamic Bayesian Network identify ecosystem level change .....	7
3.10	W Michaels <i>et al.</i> – Scientific capacity building with machine learning for the sustainability of living marine resources.....	7
4	Towards a shared literature references database.....	9
4.1	Marine genomics .....	9
4.2	Satellite imaging and remote sensing.....	9
4.3	Ecology and interactions.....	9
4.4	Imaging of benthic fauna .....	9
4.5	Imaging of plankton .....	9
4.6	Imaging and acoustics for nekton.....	10
4.7	Areas yet to be covered for the literature reference search .....	10
5	Developing a review article .....	11
6	Action points and plans for next meeting .....	12
Annex 1:	List of participants.....	13

## i Executive summary

The working group on machine learning in marine science (WGMLEARN) is tasked with charting the current status and exploring the potential for the use of machine learning methods in the various fields of marine science. The group members' presentations covered primarily computer vision for classification problems, derivation of new variables from remotely sensed data, and inference regarding species interactions. Those topics were complemented by three invited presentations: Tristan Cordier on genomics, Cedric Jamet on remote sensing, and Periklis Panagiotidis from the ICES Data Centre. In future meetings, we will strive to cover topics not covered this year (acoustics, fishing effort, etc.).

We started to assemble a comprehensive literature database to document all applications of machine learning in marine sciences, in particular in relation to the ecosystem approach to fisheries. It will serve as the basis for a review paper; both the database and the paper will help the numerous scientists interested in applying these relatively new techniques to their questions to get a broad and exhaustive overview of prior work. It will also highlight active topics and future research questions.

Approximately 500 papers were registered, covering various data types (acoustics, imaging, etc.), machine learning techniques (classic learning, deep learning, etc.), and topics (stock assessment, biogeochemistry, etc.). They are now in the process of being tagged according to these three categories, to facilitate searching the database. Topics for which the members present this year did not have sufficient expertise were identified and assigned to other group members, known to be interested and competent. An early outline for the review article was drafted, based on the distribution of topics for which papers were found.

A recurring theme was the need for training of marine scientists in the relatively new field of machine learning. For this purpose, possible new directions were discussed, including the creation and maintenance by members of the group of an online list of relevant conferences and training options (such as video lectures and MOOC courses) or the organisation of dedicated ICES training courses.

## ii Expert group information

<b>Expert group name</b>	Working Group on Machine Learning in Marine Science (WGMLEARN)
<b>Expert group cycle</b>	Multiannual Fixed Term
<b>Year cycle started</b>	2019
<b>Reporting year in cycle</b>	1/3
<b>Chair(s)</b>	Ketil Malde, Norway Jean-Olivier Irisson, France
<b>Meeting venue(s) and dates</b>	22-24 May 2019, Ostend, Belgium (19 participants)

# 1 Terms of Reference

ToR	Description	Background	<a href="#">Science plan codes</a>	Duration	Expected Deliverables
a	Review 1) new method developments in machine learning, 2) current applications of machine learning methods in marine science, and 3) their implementations and deployments in advisory and scientific processes.	Machine learning holds great potential, but it is necessary for practitioners to keep up with new developments and to gain an understanding of the opportunities and challenges with new methods.	4.1, 4.5, 3.2	1, 2, 3	On-line (live) report
b	Invite presentations (externally and internally) and review data or analysis challenges in order to discuss possible methods, approaches and technologies.	ML experts need to meet with stakeholders and data collection efforts for mutual understanding of data analysis challenges.	4.2, 4.3	1, 2, 3	On-line list of challenges
c	Communicate with DIG and the ICES Data Centre on data organization and requirements related to machine learning analysis.	For effective deployment, ML has to be integrated with data collection and data management efforts.	4.2	1, 2, 3	
d	Summarize current and future needs in marine science and identify how machine learning methods can provide solutions. Work actively to promote adoption of relevant technologies.	Future developments in the marine sciences, including project proposals, need to have an informed and up to date view of the state of the art, in order to make optimal use of the technology.	4.2, 4.3	3	

## 2 Invited Presentations

### 2.1 T Cordier – Topic introduction: Genomics

A key challenge of the upcoming decades is to maintain marine ecosystems services and the biodiversity that contribute to their sustainability. National and international regulations have been adopted to implement the monitoring of biodiversity as a key component of the assessment. To comply with those regulations, biotic indices, that combine biodiversity measures into a single integrative scale, have been developed. The majority of these indices rely on the direct observation of morphologically distinguishable bio-indicators species. These inventories proved hampered by multiple shortcomings, including both biological (life stage, or different species being undistinguishable) and technical (labour-intensive, various level of taxonomic expertise). The recent pace of development of high-throughput sequencing technologies made the collection of large biodiversity datasets from environmental samples faster than ever before. Pioneering studies assessing the potential of environmental genomics tools for the screening of bio-indicators taxa showed that reliable biotic indices and ecological quality status can be obtained. However, the amount of data used to compute the indices remains limited, because a large proportion of the sequences remains unassigned or assigned to taxa of unknown ecology. More recently, the use of machine learning tools for the supervised training of predictive models from genomics and ground-truth, morphology-based quality assessment data have proven useful. Using such an approach provide few advantages. First, the hurdles of the lack of taxonomic framework is bypassed, because the ecological signal of the sequenced (linear or not) is being inferred during the training process. Second, as opposed to the morphological identification, sequencing data is unambiguous and can be easily stored and compared across space and time. Finally, thanks to the continuous improvement in cost-effectiveness and the miniaturizing of the hardware, in situ sampler for the environmental DNA collection, processing and sequencing is at reach, unlocking the disruptive potential of environmental genomics for the marine observation and monitoring.

### 2.2 C Jamet – Topic introduction: Remote sensing

Remote sensing covers the wide array of observations techniques that provide information about an object/surface without being in contact with it. Here, the focus is put on satellite observations, and in particular on ocean colour information (although the techniques presented would apply more widely). Such data provide synoptic information on the global ocean continuously for the past 20+ decades and at a spatial resolution of 300-1000 meters. Its analysis is often done through very standard processing routines that, over the years, have been challenged by machine learning approaches.

The machine learning tools used for satellite imaging are mostly Artificial Neural Networks such as the Multi Layer Perceptron and the Self Organising Maps. After an overview of their use through existing review papers, I presented some examples of applications for (1) atmospheric correction, through direct inversion of the satellite measurements or the optimisation of inversion models parameters, (2) derivation of ocean variables such as total chlorophyll a concentration, phytoplankton functional groups, or pCO<sub>2</sub> in surface waters.



## **2.3 P Panagiotidis – Demo of ICES data and services**

The ICES secretariat has been collecting data for over 100 years. Currently, the databases are maintained for stock assessment groups and well as academics. ICES maintains data portals for each of these databases, by topics (oceanographic observations, biological observations, catch records, etc.) as well as a central portal for all. The data is accessible through APIs and a catalogue of such services is also maintained. Links to these portals and catalogues were provided and a few were demonstrated, to raise awareness about the existence of these large databases.

## 3 Participants Presentations

### 3.1 J Simon *et al.* – Giving Artificial Monitoring intelligence to Fishing TRAWLS (Game of Trawls)

The trawl fisheries are still recording high levels of bycatch despite the numerous selectivity projects conducted in the recent years. The main issue is that trawlers are towing their fishing gears for hours without knowing if what is actually entering their trawls is what they are targeting. Moreover, inside trawls the fish often adopt a behaviour, which consists of not coming into contact with the mesh, which makes the selective devices inefficient.

In the GAME OF TRAWLS project (Giving Artificial Monitoring intelligence to Fishing TRAWLS) we propose to adapt the technological advances made in recent years in the fields of artificial intelligence to fishing trawls. This project will propose several approaches, including computer vision to be able to detect and identify in real time the species that enter the fishing gear. Such systems could allow fishermen to detect in real time high abundance of bycatch in their trawls so they could operate an escape device (diversion hatch, bright flash, acoustic signals ...) or they could change of fishing area.

### 3.2 E van Helmond and L. Nguyen *et al.* – Computer vision technologies in Dutch fisheries

There is great potential in using machine learning for automated species identification of ray species in commercial fish catches. North Sea ray stocks are currently defined as data limited. Stock assessment will benefit from improved data collection and will lead, eventually, to more sustainable fisheries management of these species. Wageningen University and Wageningen Marine Research conducted a feasibility study on automated registration of ray species from the Dutch demersal fisheries. We explore the possibility of using a convolutional neural network to locate the fishes of interest in images. To demonstrate this technique, we collected image data that represent the different levels of fish composition on the sorting belt, which include randomly positioned single fish and multiple fish of different species positioned in different degree of occlusion. This automated approach shows promising results. Compared to current practice of random manually sampling, automated video monitoring could provide estimations of the complete catch, while minimising labour costs.

### 3.3 L Hoebeke *et al.* – Automated hierarchical classification of animal species in camera trap images

Automatic imaging techniques, such as camera traps, are increasingly being used in biological monitoring. The great advantage of camera traps is that accurate data can be collected without animals being disturbed or researchers being present. However, such imaging frameworks produce high volumes of images, which often need to be reviewed and annotated manually. Convolutional neural networks, nowadays the go-to technique for computer vision problems, can be used to automate this labour-intensive process.

The limited number of labelled camera trap images, combined with the difficulty of the classification task, does not allow for detailed classification of all species by the neural network. To overcome this problem, we incorporated hierarchical classification into the network. This way,

the network can still reduce the manual workload, while misclassification is being avoided. Depending on the reliability of the classification, the level of detail can be adapted. The classification can be restricted to a higher level, for example family level, if there is insufficient information to classify that image to species level, while other images can still be classified in more detail using the same network. Finally, the network can automatically label images or provide suggestions to users when incorporated into annotation applications to speed up the annotation process.

This method of hierarchically classifying camera trap images can easily be extended to other imaging data by incorporating the corresponding classification tree into the network.

### **3.4 K Malde *et al.* – Vector space embedding for plankton images**

Traditional classifiers map inputs to discrete classes, which must be specified in advance and are intrinsic to the classifier. Plankton images, like many other real world data types, have properties that make the design of effective classifiers difficult. For instance, the number of classes is potentially very large, and classes often overlap. In addition, the choice of taxonomy can differ between researchers and between institutions. Inspired by recent work in face recognition, we instead use a deep convolutional network to learn a vector embedding of the data. The vector embedding preserves semantics, so that objects in the same class are placed close to each other, and objects in different classes are placed far apart. This reveals inherent structure in the underlying data, which allows the classifier to work with different taxonomies or to classify data in previously unseen classes.

### **3.5 J Fernandes *et al.* – Machine learning is not only for big data and big data is not only a lot of data: successful examples in marine sciences with sparse and heterogeneous data**

Marine research is challenging due to the difficulties and high costs to get data and perform experiments. Despite technological developments that have notably increased the resolution and amount of data available, it is difficult and costly to access labelled data for automatic learning and validation. Recent reborn of machine learning under the umbrella of the 'Big data' term focuses on large datasets or heterogeneous datasets where sparse data problems seem to be a challenge. However, these arguments do not consider recent advances in machine learning and early histories of successful applications on marine research to deal with sparse and heterogeneous data. These applications have proven their capacity to successfully solve problems with sparse data and high uncertainty. Here, we present several successful examples based on machine learning methodologies for forecasting of fish recruitment and marine litter beaching, fishing activity tracking, marine spatial planning as well as current work forecasting biotoxins in offshore aquaculture or fuel consumption reduction (e.g. H2020 DataBio project). We also discuss novel and robust machine learning approaches with high potential to be applied to marine science domain such as weakly-supervised classification (including semi-supervised methods), multi-dimensional Bayesian networks, multi-objective time-series forecasting, aggregated outputs by linear models and others. We also present the risks related with the inappropriate application and validation of machine learning (e.g., overfitting, poor validation schemes, bias in training sets, and others). We also discuss some of the bottlenecks: 1) Find skilled/motivated people; 2) Data availability, knowledge and quality; 3) High inversion at beginning / slow to start having results; 4) Right combination of domain experts and ML experts; 5) Lack of stable ML

group to continue and keep the momentum; 6) Traditional ways hard to change. The success strongly relies on good understanding of the problem domain and the data characteristics for correct selection of the methods to use, its validation and interpretation, as well as a good integration with current work processes (protocols, software and other).

### **3.6 R Sauzede *et al.* – Machine learning-based methods to derive biogeochemical parameters from profiling floats and satellite data**

The ~4000 profiling floats from the Argo program have considerably increased our observation capabilities at the scale of the global ocean, but these floats record down to 1000m depth temperature and salinity only. A new generation of floats, the 350 BGC-Argo floats, are additionally equipped with sensors that measure biogeochemical properties such as oxygen, nitrate, fluorescence of chlorophyll-a, particulate backscattering, etc. These BGC-Argo floats represent a growing effort to build a global constant array for biogeochemical observations in the global ocean. In this context, machine learning-based methods are developed to take benefit from this growing amount of data in the aim to better understand the ocean biogeochemistry. In this talk, I presented how neural networks can be used to: (1) derive a global 4-dimensional distribution of phytoplankton biomass (i.e. chlorophyll-a concentration and particulate organic carbon) from merged hydrological properties of the water column (from Argo floats) and surface satellite data and (2) infer profiles of nutrient concentrations and carbonate system variables from profiling floats temperature, salinity and oxygen profiles. Products derived from these methods should soon become routinely available through Copernicus Marine Environment Monitoring Service.

### **3.7 M Lo *et al.* – Deep learning to improve remote-sensed ocean data for fisheries research**

During the last decades, the study of vessel's movements has been facilitated by the implementation of Vessel Monitoring Systems (VMS) and satellite Automatic Information Systems (AIS). These tracking systems have allowed scientists to investigate fishing activities with unprecedented details, including the study of vessel movements, of their searching strategies as well as the estimation of fishing effort. What has been less studied however is how the variations of fishing activities over space and time are controlled by the environmental variability. Remote-sensed oceanic data (e.g. SST, SSS, SSW, surface currents and ocean colour) constitute probably the best dataset to make such investigations since they are observations (approaching the "ground truth" better than ocean models), publicly available in real time and with wide and continuous spatio-temporal coverages. Nevertheless, those datasets are known to be limited to the near surface ocean, having multiple resolutions and including missing values. In addition, other key variables (e.g. Primary Production) are not directly measured by satellite. These limitations hampered our abilities to map fishing activities against aggregated environmental data. Our PhD project consists in using Deep Learning methods, along with in-situ data (from moving autonomous platforms like Argo, gliders and drifters, as well as from fixed moorings), to improve satellite dataset by leveraging the above-mentioned limitations. This includes increasing the resolution and filling the gaps for inter-operability and comparability. Other improvements consist in extrapolating surface information along the vertical and estimating new variables relevant for fisheries thanks to advanced deep learning technique. By comparing these upgraded remote-sensed dataset against fisheries data (e.g. catch, VMS, AIS, FAD...), this PhD project will improve our understanding of the environmental control of fishing. We will simultaneously learn about

how fishers uses the oceanic environment and how exploited marine species are distributed by linking the characteristics of the pelagic seascapes with fishing activities.

### **3.8 M Stock *et al.* – Pairwise learning to predict species interaction networks**

Plankton communities can be described as food webs, specific instances of species interaction networks. These networks, a collection of species as nodes and their interactions as edges, encode the structure of an ecosystem. We study how to use supervised machine learning tools to be able to predict new species interactions. Based on an observed network, we learn a function that takes as inputs the description of two species (e.g. traits, phylogenetic similarity or a morphological description) and predicts whether these two species are likely to interact or not. This framework for pairwise learning is based on kernels and similar methods have been highly successful for predicting molecular networks and for recommender systems, as used by companies such as Netflix and Amazon. During this workshop, we would like to explore how traits can automatically be extracted from (microscopy) imaging data, in order to predict trophic relations between species.

### **3.9 L Uusitalo *et al.* – Hidden variables in a Dynamic Bayesian Network identify ecosystem level change**

Ecosystems are known to change in terms of their structure and functioning over time. Modelling this change is a challenge, however, as data are scarce, and models often assume that the relationships between ecosystem components are invariable over time. Dynamic Bayesian Networks (DBN) methodology was applied to the Central Baltic Sea ecosystem, which has gone through a major regime shift. The hidden variables (HV) implemented in the DBN can capture the unobserved processes behind the ecosystem change. Different model structures and HV setups were used and their effect on the results was evaluated. The exact setup of the hidden variables did not considerably affect the result, and the hidden variables picked up a pattern that agrees with previous research on the system dynamics. The data being scarce, a Naive Bayes model performed slightly better than the expert-judgement-based model structures, probably due to its simplicity.

### **3.10 W Michaels *et al.* – Scientific capacity building with machine learning for the sustainability of living marine resources**

NOAA Fisheries has experienced a dramatic increase in imagery and acoustic data in recent years from the implementation of sampling technologies to enhance surveys and ocean observation operations. This presentation provide examples for the ML applications and the strategic roadmap for advancing the use of ML. To address the increasing processing costs of the big data, NOAA Fisheries initiated the Automated Image Analysis Strategic Initiative in 2014 to develop an open source toolkit to streamline data processing with automated detection and classification using machine learning (ML) algorithms. The AIASI working group partnered with Kitware Computer Vision Inc. to deliver the Video and Image Analytics for a Marine Environment

(VIAME) software in 2018 (<http://www.viametoolkit.org/>). VIAME is an end-to-end open-source software package for automated image analysis of marine and fisheries science data that utilizes advanced computer vision and ML analytics for automated object detection, tracking, and classification. The use of VIAME was initially applied to underwater fisheries surveys to improve the quality and timeliness of abundance estimates for stock assessments, and has recently expanded to aerial surveys for marine mammals. VIAME is also presently being appraised for electronic monitoring of fishing vessel catch and bycatch. Another ML example, is NOAA Fisheries partnership with the University of California with the CoralNet software for automated classification and classification of benthic coral habitat. Commonalities in the challenges of utilizing ML are pertinent to a number of new NOAA Fisheries' strategic initiatives, and these include:

- NOAA Strategic Initiative on Artificial Intelligence (AI) – NOAA's cross-functional mission priorities include the application of ML analytics to provide higher quality and more timely scientific products. This includes improving data accessibility for ML analytics, discovery by the broader scientific community, building partnerships for scientific exchange and building competence in ML.
- NOAA Fisheries Data Modernization – Optimizing the NOAA Fisheries Data Enterprise with improvements in storage and accessibility of its data and workflow for ML analytics.
- NOAA Unmanned Systems (UxS) Strategic Initiative - NOAA is expanding its UXS program to cost-effectively augment its existing surveys and ocean observation infrastructure, and the data enterprise modernization and ML analytics are closely linked with this initiative.

## 4 Towards a shared literature references database

Participants were instructed to prepare for the workshop by installing Zotero. Jean-Olivier Irisson gave a quick review of the functionality, and after some discussion, a set of tags were decided on (covering the data types, data collection platforms, machine learning techniques, and application domains). The participants were divided in topical groups, which were tasked to collect and tag relevant literature.

### 4.1 Marine genomics

The field of genomics have traditionally used some machine learning methods for the analysis of raw sequencing data (e.g. sequence clustering, annotation using HMMs and classification of gene expression data with SOMs), there appears to be a limited number of publications focusing on the application of ML specifically to genomics data collected from marine environment. While there are potential applications, currently this field appears to be too limited for a comprehensive review.

### 4.2 Satellite imaging and remote sensing

The use of machine learning methods in the field of remote sensing dates back to the late 80ties/early 90ties. The need naturally arises since satellite data of the ocean surface require much pre-processing to derive standardised, operational data products. Spatio-temporal interpolation/extrapolation, atmospheric scattering and absorption as well as cloud removal are a few of the challenges machine learning algorithms have to face. As like in other fields oceanographers are interested in pattern derived from satellite products of the ocean. Machine learning can help to identify structures like frontal zones or mesoscale eddies as well as biogeochemical parameters to e.g. predict ocean productivity or harmful algae blooms (HABs).

### 4.3 Ecology and interactions

Machine learning has been used to find spatial and temporal patterns in ecological data, as well as non-linear or otherwise more obscure connections between ecological components. References of Dynamic Bayesian Network applications, as well as multidimensional and tree-augmented Naive Bayes models were found.

### 4.4 Imaging of benthic fauna

Automatic, image-analysis based classification has been developed and evaluated for benthic invertebrates. They include the development of the imaging system and assessment of different classification algorithms. A first conclusion is that, while the literature is abundant, most papers are not comparable to each other because of the lack of common benchmark datasets.

### 4.5 Imaging of plankton

The literature on plankton imaging is large and has grown in recent years through publications in computer vision conferences. Covering it completely raised some questions about the relative merits of the various papers (and the peer reviewing quality). We collected all references in an

effort to be exhaustive. A first conclusion is that, while the literature is abundant, most papers are not comparable to each other because of the lack of common benchmark datasets.

## 4.6 Imaging and acoustics for nekton

Since 1993, studies on automated recognition on marine fauna, mainly fish, were published. An important driver behind implementation of machine learning applications is to improve the efficiency of data collection: reduce the extensive amount of time needed and to minimise labour costs to process image information. At the same, time automated image recognition allows increasing spatial and temporal coverage of data collection, and potentially real-time coverage, to improve marine resource management. Study aims could be divided in several defined categories or domains: acoustics, video monitoring on board, underwater environment video monitoring, stock assessments, biological data collection, and fish quality.

## 4.7 Areas yet to be covered for the literature reference search

So far, 500 papers were retrieved and tagged; the database is available [online](#). Not all topics could be tackled by the people present and in the time imparted. We identified the following missing themes and potential people to address them:

- acoustics, active and passive (N O Handegard and W Michaels)
- stock assessments (J Simon, J-B Romagnan, W Michaels)
- vessel monitoring and activity reporting (B Woodward, B Alger)
- autonomous platforms (underwater observatories, gliders, floats, drones, stationary continuous sampling stations, ships of opportunity...) (R Sauzede, W Michaels)
- Aquaculture (K Malde)
- Data access and policies (W Michaels)

We also identified that we could build a catalogue of reference training sets using this same online tool. Now they just need to be found and added.



## 5 Developing a review article

A review article based on the literature database is being drafted. The process is structured around topical domains: 1) environmental assessment, 2) ecology and interactions, 3) stock assessment and resource management, 4) fisheries/activities monitoring, 5) policy implications. This reflects competence in the group and the general process of the ecosystem-based management approach to fisheries. Editorial responsibilities were distributed for each domain.

## 6 Action points and plans for next meeting

Tentatively, the next meeting will be in Copenhagen, on April 14-17, 2020.

The following milestones were set for the deliverables:

- July 1: literature review checkpoint, paper structure and writing plan finalized
- October 1: manuscript sections draft ready, decision on where to publish
- November 1: manuscript finalized and set for homogenization by a reduced writing team
- December 1: polished manuscript sent to all authors for last comments
- December 15: submission

An additional deliverable within the WGMLEARN remit was suggested: a live/online catalogue of conferences, workshops, committees, etc. in relation to machine learning in marine sciences. This is not currently covered by the activities of the Working Group.

As highlighted above, some topics and data infrastructures could not be presented this year. The plan is to fill those gaps for the next meeting. This will involve:

- get key resource people contribute to the literature database and review paper; this is highlighted above already.
- contacting other working groups, infrastructures, projects and consortia to have them participate in the next meeting; this will be done by the chairs.

Soliciting new participants is difficult without the possibility to pay some of their expenses. A COST action proposal on machine learning in marine science, involving most WGMLEARN participants, was already submitted four times, unfortunately without success. It will be submitted again this year, taking advantage of the enlarged network of participants.

A recurring theme was the need for training of marine scientists in the relatively new field of machine learning. This year, this necessity was partly filled by the co-location with the [POGO workshop on machine learning](#). In the coming years, we will consider the possibility of organising training courses (e.g. ICES training courses) or at least collect links towards online training courses (MOOCs, etc.).

## Annex 1: List of participants

Alessandra GOMES

Bernhard KUEHN

Cédric JAMET

Edwin VAN HELMOND

Jean-Baptiste ROMAGNAN

Jean-Olivier IRISSON (chair)

Jose FERNANDES

Julien SIMON

Ketil MALDE (chair)

Laura HOEBEKE

Laura UUSITALO

Linh NGUYEN

Madiop LO (remote)

Michiel STOCK

Perikilis PANAGIOTIDIS

Rainer KIKO

Raphaelle SAUZEDE

Rubens LOPES

Tristan CORDIER

Vincent ROSSI (remote)

William MICHAELS