ICES WKMLEARN 2018 Report

Ecosystem Observation Steering Group

ICES CM 2018/EOSG:20

Ref ACOM and SCICOM

Report of the Workshop on Machine Learning in Marine Science (WKMLEARN)

16-20 April 2018

ICES Headquarters, Copenhagen, Denmark

ICES International Council for the Exploration of the Sea

CIEM Conseil International pour l'Exploration de la Mer

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
info@ices.dk

Recommended format for purposes of citation:

ICES. 2018. Report of the Workshop on Machine Learning in Marine Science (WKMLEARN), 16-20 April 2018, ICES Headquarters, Copenhagen, Denmark. ICES CM 2018/EOSG. 28. pp. https://doi.org/10.17895/ices.pub.8178

Contents

Executive summary

The field of machine learning has seen tremendous advances the last few years and is increasingly being applied to data of all kinds. The ICES WKMLEARN workshop was formed to investigate the actual and potential use of such technologies in the context of marine sciences and advisory processes, and provide recommendations for how to apply these technologies. The workshop was attended by around 30 researchers from Europe, United States, and Canada.

Machine learning algorithms implement models that are adjusted according to data. In supervised learning, training data are given along with target values (e.g. classes to be identified), while unsupervised learning identifies structure (e.g. clusters) in the data itself. Machine learning models can be applied when there is insufficient knowledge or resources to develop mechaniztic models. This makes the technology attractive in many contexts. However, the many highly publicized successes may have resulted in an overly optimistic impression regarding the opportunities and limitations of the technology. In particular, models can often be opaque or difficult to understand, and it is important to be aware that the limitations and pitfalls of machine learning models may be poorly understood.

We identified the following areas that need addressing:

Expertise

- *Attracting and long-term engagement of expertise in machine learning:* expertise is attractive for industry and it can be difficult to recruit skilled individuals and especially to retain expertise over time;
- *Publication of work:* applications of machine learning combine methods from computer science with problems from other fields and is, by nature, cross-disciplinary. It can therefore be difficult to find appropriate venues for publication, especially since the applications rarely induce computer science breakthroughs or new ideas and concepts in fisheries science;
- *Preservation and sharing of acquired knowledge/competence:* projects are often small and isolated and resulting knowledge is often not retained when the project terminates.

Available data

- *Data quality, organization, and volume:* data are often made available in insufficient volumes and with data qualities adequate for manual curation, but not automatic analysis;
- Data are often collected with *insufficient labeling*, making it difficult to construct the training datasets needed for supervised learning;
- There is no centrally organized venue for sharing of architectures, trained models, and code for machine learning in marine science.

Processes

- *Knowing the requirements* for the subsequent analyses of data (such as the accuracies required) is needed to guide machine learning applications towards areas with high impact;
- *Communication between marine scientists and machine learning scientists* is needed to increase awareness of the methods and of their potential applications;
- *Implementation and deployment* of developed methods/analyses;

- *Machine learning applications could often guide data collection and equipment design*, in particular when machine learning is used for data processing. This is difficult because the scientific communities responsible for data collection and analysis through machine learning are often distant, both as subjects and as organizations.

Acceptance, quality

- Inferring mechanism from machine learning models is difficult, but crucial to improve our understanding of the patterns and processes modelled and discard the impression that machine learning models are always opaque;
- New methods need to be carefully verified for validity because, while the requirements of many machine learning methods are not very stringent (as opposed to classic inferential statistics for example), the methods can still lead to wrong results if not used properly;
- Machine learning models are different from conventional statistics used by scientists, and might be distrusted or avoided due to unfamiliarity.

1 Opening of the meeting

The meeting was attended by around 30 researchers from Europe, United States, and Canada. The age distribution of participants reflected how machine learning is appealing to early career fisheries scientists and ecologists. The participants were diverse in terms of age and education: computer science, mathematics, statistics, engineering, and GIS. Machine learning knowledge was patchy and the majority of participants had acquired skills through self-learning, such as online courses. Other participants were involved in machine learning research projects through university collaborations. The preferred programming languages were Python, Java and R; R being preferred by fisheries scientist and Python by computer scientists.

2 Specific responses to Terms of Reference

2.1 ToR a

Review ICES Fisheries Science processes to understand where machine learning and/or deep learning may be of greatest benefit, including: i) Survey and data collection, ii) Data handling, iii) Analysis and assessment, iv) Review and advice, v) Check degree to which expert groups are meeting their terms of reference, vi) Taking a forward look and consider emerging topics.

2.1.1 Large-scale data analysis

Many recent and highly publicized developments in machine learning include so-called deep learning algorithms, typically convolutional neural networks. These methods show great potential, in particular in cases with:

- Large data volumes (and potential for further increase)
- Regular data (matrices or time-series)
- High dimensional data
- Labor intensive analysis
- Labeled data availability (to be used as training sets)

However, convolutional neural networks are only a subset of all the machine learning methodologies that are being applied successfully in fisheries science. A large selection of supervised, unsupervised and semi-supervised learning methods and architectures can readily be applied to many important data types. A clear advantage of models like decision trees or Bayesian networks are that they deal with uncertainty explicitly, providing an intuitive interface to data in terms of transparency and comprehensibility of the final model. The intuitive properties of such methods enhance the confidence of domain experts on their forecasts (Fernandes *et al.*, 2010; 2013; 2015).

Camera equipment can be deployed in a multitude of situations, including underwater observatories, trawl cameras, UAVs, lab equipment, in the monitoring of commercial fisheries, etc.. While specialized equipment can be expensive, cheap "action cameras" are also used by researchers. In either case, the cost of manual curation by a human expert will easily dwarf equipment costs, making automation necessary to exploit the full potential of the technology.

Acoustics data are one of the most important data sources for fisheries advice, and while volume is already a challenge, new equipment types exacerbate this by increasing data resolution and dimensionality. Multibeam echosounders replace traditional two-dimensional data with three-dimensional data, while multifrequency and broadband equipment expand the measured signal from one to multiple values, or to a continuous frequency spectrum response. The resulting data are difficult to visualize effectively and hence difficult for a human expert to interpret. Automated solutions will more easily be able to exploit the information in the data.

2.1.2 Ecosystem and complex models

Many machine learning models like neural networks have the ability to capture complex, non-linear relationships in the input data. Such relationships often occur in ecosystem models, which are crucial building blocks for the implementation of ecosystem-based fisheries management. There is a need for models that can capture these complex dynamics and merge high-dimensional data from different sources.

Ecosystems predictions are rarely used in fisheries management or advice, despite the drastic population fluctuations potentially caused by environmental factors. Ecosystem-based fisheries management, enabled by machine learning, could alleviate the stresses from the traditionally boom-or-bust economy associated with many fishing industries.

2.1.3 Data quality control

Data going into the ICES systems and databases is often not collected with a view to manual curation and analysis. Collection and contribution tend to conform to varying and local standards or practices, and can contain errors or require interpretation. Machine learning techniques, such as anomaly detection algorithms or Bayesian networks, can help to (automatically) find potential errors in the data and establish the overall quality of the import data, and e.g. to flag data that exceeds a certain error threshold. This then can be used to have a closer (manual) look at the data, improving overall quality while limiting the required human effort.

2.1.4 In regards to the ICES fisheries process

The workshop presented an overview of the ICES advisory processes by a number of ICES staff members (e.g. ACOM and TAF processes), a Norwegian perspective on data preparation for assessments, and the US perspective on optical data acquisition opportunities that might lend themselves to machine learning applications. The focus was mainly on identifying the bottlenecks, both in the sample processing of current advisory processes and scientific advancement in the methodology of assessments.

The opportunities for machine learning impacts identified in this session could be grouped or summarized in several different ways by the ecosystem components of interest, by the current level of developments, or by the magnitude of the impact of machine learning models. To aid in the prioritization of these opportunities, however, it seems useful to prioritize on the basis of the respective role in the advisory process.

Presentations summarized the activities of the current routine advisory process in the form of information flow/pipeline system including the process of providing advice and the TAF procedure. Clearest / easiest opportunities for machine learning applications presented themselves in the sample analysis or data preparation part of the process as opposed to the data analytical part (assessment procedure). The benefits of the advisory process seemed to be the rapid nature and the reproducibility of choices, and if new methodologies could be spliced into the modular pipeline set up. The seamless replacement of human evaluators, the frequent availability of annotated images which could lend themselves to the development of training datasets as well as detailed documentation on the level of human performance indicators make these ideal opportunities for training supervised learning approaches. Examples include:

- Age data (surveys and catch)
- Fish egg abundance from surveys
- Nephrops burrow counts from surveys
- Scallop count and size distributions from surveys
- Species identification / quantification mainly in bycatch or catch data
- Acoustic survey interpretation (scrutineering)

Likely longer-term opportunities where additional work is required to evaluate the performance of machine learning methods are those where new types of samples/information are to be included in the advisory process. These methods usually require additional effort both in the development of the training data and the peer-review of the methods. Such opportunities aim to include environmental/ecosystem information into the advisory process, as well as new data methods for single species assessment in the benchmark process and external estimation of catchability. Less fixed/traditional advisory processes at least partially still in development such as ecosystem assessments and ecosystem overviews could be transformed by these methods. Examples Include:

- Biodiversity through automated identification of samples, zooplankton, phytoplankton
- Size spectral / functional type analysis of zooplankton
- Identification and abundance of species on untrawlable ground
- Higher spatial resolution of survey trawl samples to aid acoustic scrutineering
- Benthic habitat classification or status assessments from multibeam or video transect information
- Environmentally influenced forecast predictions
- Multispectral satellite analysis to evaluate changes in environmental or ecological conditions

2.2 ToR b

Identify areas of marine science, data and advice within the ICES remit where machine learning/deep learning has already been applied.

A large number of projects and ongoing efforts are now experimenting and developing prototypes of machine learning systems in marine sciences. Several ongoing projects were presented by the workshop participants, and the following topics were highlighted:

2.2.1 Electronic monitoring and vessel monitoring systems

Within the last decade, Electronic Monitoring (EM) of commercial fishing vessels has emerged as a cost-efficient supplement to the existing expensive observer programs documenting catches in commercial fisheries. Such monitoring can be performed using existing data sources like AIS, or through special equipment in the form of a Vessel Monitoring System (VMS). These systems collect data by video surveillance of the fisher's catch as well as logging equipment use.

Machine learning applications to electronic monitoring of fishery-dependent data are of increasing interest to management bodies in the United States and Europe. It has the potential to reduce the cost associated with observers and streamline the processing of video data. NOAA Fisheries’ Alaska Fisheries Science Center has just completed a pilot project using machine learning techniques to analyse video data from a chute system installed to monitor halibut bycatch during release from trawler catches, and they were able to identify fish species to a high degree of accuracy. NOAA Fisheries has also just created an internal national machine learning working group to further evaluate machine learning in the context of electronic monitoring. Concurrently, vendors from around the United States are developing techniques to better incorporate machine learning into the video review of electronic monitoring data. One such vendor is working with the electronic monitoring data of the groundfish fleet in New England. By targeting specific image analysis goals such as species identification, count, and length measurement, they were able to run a successful data science competition to demonstrate the feasibility of semi-automated systems. The outcome of this process was a project to establish an open source library of data, algorithms, and software modules available to regulators, EM providers, and other interested parties. Such a library will spur adoption of semi-automated video review, and help alleviate resource constraints. In Europe, automated species recognition algorithms have been developed. However, follow up, deployment, and implementation of automated video review should be improved. Challenges include establishing the process of using machine learning for the purpose of evaluating EM data and more generally facilitating the uptake of EM in commercial fisheries to improve data collection.

The recent introduction of Automatic Identification Systems (AIS) for identification and tracking vessels presents an opportunity to quantify pressure on the marine ecosystem deriving from human activities such as shipping and fishing. Identifying fishing activities from AIS data alone (such as position, speed, and bearing) may be possible, alternatively, it can be used in conjunction with more extensive monitoring systems (VMS) and Logbook data. Several comparisons are starting to be published (McCauley *et al.* 2016; Eigaard *et al.*, 2017).

2.2.2 Time-series forecasting and reconstruction

Time-series forecasting work was presented by several scientists. In particular, anchovy recruitment forecasting in the Bay of Biscay (Fernandes *et al.*, 2010) is an example of a successful application that influenced advice (Fernandes *et al.*, 2009a) to open the fishery, together with other scientific evidence (e.g. acoustics survey). This work was applied to seven species in the North Atlantic (Fernandes *et al.*, 2015). The use of Bayesian networks was expanded to use multidimensional Bayesian networks to double the chance of correctly forecasting the recruitment of three species simultaneously following an ecosystem-based approach (Fernandes *et al.*, 2013). These methods were also combined with a mechaniztic model (e.g. gadget) to make forecasts to the year 2020 (Andonegi *et al.*, 2011) which correctly forecasted the last peaks in anchovy recruitment 6-8 years before they occurred. Recent machine learning advances in combination with optimization methods are promising to balance the performance of forecast and the earliness of those forecasts (Mori *et al.*, 2017). Traditional Bayesian networks approaches and novel semi-supervised approaches are being applied to coastal litter forecasting in the LIFE-LEMA project (Hernández-González *et al.*, 2018).

Similar work in this regard included preliminary research showing that simple neural networks have the potential to simultaneously forecast changes in six fish functional groups of the Grand Bank, Northwest Atlantic, using small sets (i.e. 2 to 5) of fishing and environmental indicators as predictors.

2.2.3 Optical and acoustic data from surveys

Survey data looking at fish stocks and habitat are collected with optical and acoustic technologies. The processing of these data are costly from both a monetary and human effort perspective, but efforts are underway to use machine learning to streamline these processes and enhance capabilities in analysis and visualization.

Machine learning algorithms and networks for automated image analysis are being trained using datasets that incorporate a wide array of fisheries species and habitats. NOAA Fisheries is in early stages of implementing an automated image analysis open source software toolkit to process the optical data used in stock assessments. Results so far vary based on the dataset used to train the algorithm; for example, optical data with homogenous habitat and fewer species of fish are easier to target than more complex systems. Datasets encompassing a broad range of underwater conditions and species and collection consistency are critical when applying machine learning to optical survey data.

Other advancements in machine learning for non-government fisheries data include pilot projects in preprocessing data (i.e. data filtering), automatic species identification within a single frame, species identification and tracking within a video frame and measuring of individual fish for assessments and enforcement purposes.

There is also an effort in using machine learning to assist in the laborious process of aging fish by reading otoliths or scales. Images of sectioned otoliths are analysed by algorithms that automatically count annuli and produce an age estimate for the individual. The error can be small, but the interpretation of the error in the algorithm-produced age vs. true age identified by the reader has important implications when considering the biology of the species. Error around the age at maturity for a species could misinform management decisions.

The UK has begun pilot work on using machine learning algorithms on habitat classifications using multibeam sonar. Several groups in Europe are trying machine learning applications with fisheries acoustics. Effort is typically focused on species differentiation and size estimation, often combined with imaging systems. Imaging systems are typically used in association with trawls and offer a spatial resolutionsuitable for combining with acoustics. Machine learning within acoustics is well suited for supervised manner (i.e. trained with ground-truthed examples) since the processing usually involves manual labeling of large amounts of data. This will help predict fish school echo types or species, and potentially decide when to deploy trawls that will result in improved classification. In addition, models could be used in an unsupervised manner to cluster three-dimensional fish data without any expert a priori. This could unveil new structure in fish schools and fish spatial distributions.

In AZTI (Spain) there have been different examples of applications of machine learning to acoustic databased on narrowband echosounders in the last years. In the Bay of Biscay, the temporal series of abundance of the juvenile fraction of anchovy population (2003-2015) was used to predict recruitment (Boyra *et al.*, 2013) and the aggregation patterns of anchovy were used to explain spatial differences in behavior for this species (Boyra *et al.*, 2016). Besides, image processing techniques were applied to sonar screenshots to automatically detect bluefin tuna from the commercial live-bait tuna fishing fleet (Uranga *et al.*, 2017). In tropical tuna fisheries, there have been developments to improve automatic abundance estimation from acoustic data from echosounder buoys in FADs (Lopez *et al.*, 2016) and to increase the size and species discrimination from sonar and echosounder data (Boyra *et al.*, 2018). In addition, there is ongoing research to develop machine learning techniques to deal with the challenging, recently released broadband echosounders. These allow new ways of discriminating species and size of fish based on the analysis of continuous frequency-response patterns, but at the cost of handling much larger datasets (Demer, 2017).

2.2.4 Plankton: imaging technologies

Plankton includes the *larval* stages of many harvested species, is a source of *food* for many others, and can act as an *indicator* of good overall ecosystem status (sensu Marine Strategy Framework Directive). The biomass of planktonic organisms is often assessed with *nets*, *bottles* and *microscopic counts*. This is time-consuming and the samples do not contain important but fragile taxa, which are damaged by nets. In addition, these catches are often not sufficient to resolve processes at the *small spatial and short temporal scales* that are relevant to planktonic organisms. This has led to the development of instruments that take images of plankton at high resolution and frequency. Some scanners or in-flow cameras can speed up the processing of plankton net or pump *samples* (e.g. ZooScan, FlowCam, ZooCam). Some underwater cameras take images of plankton directly *in situ* and can therefore describe fragile species that are destroyed by net sampling (e.g. UVP, LOKI, ISIIS).

These instruments produce an enormous amount of digital images (e.g. typically, one ZooScan: 1 billion pixels/year, UVP: 8.6 billion pixels/year, ZooCam: 7.2 trillion pixels/year, ISIIS:14 trillion pixels/year). An important bottleneck is *processing* this massive data. Most devices produce grey scale images with a quite uniform background and *semi-automated* processing systems have existed for more than ten years (e.g. ZooProcess, EcoTaxa, ZooImage). They process input images, extract objects of potential interest, measure morphological characteristics of these objects, and offer some classic machine learning tools (such as Random Forests) to propose an identification, based on a manually curated training set of images.

These systems are now used *routinely* (Ortner *et al.*, 1979; Benfield *et al.*, 2007; Irigoien *et al.*, 2009; Gorsky *et al.*, 2010; Uusitalo, 2016). They commonly reach 60 to 70% overall accuracy for classifying plankton images in about 40 groups (albeit with tremendous differences among groups). This accelerates the processing of samples compared to manual sorting but *human curation* is still *required* to increase data quality or detail, both of which are mandatory to inform ecological studies and advisory processes.

Recent developments, such as *deep learning* approaches, will further accelerate this process. Plankton images from quantitative imaging instruments lend themselves very well to these new techniques and a few studies are emerging, with promising results (<5 published papers, one open implementation in EcoTaxa). These techniques could improve classification and also perform segmentation and streamline the overall image analysis.

Large images datasets can also benefit from the help of *citizen scientists*: volunteers from the public who are presented images through a website and classify them. The PlanktonID initiative (https://planktonid.geomar.de/en) actually intertwines machine learning to suggest identifications and citizen science to validate them.

Finally, with such large datasets, it is important to consider that classifiers do not output a yes/no answer but a continuous *score* which can be used to guide human curation (focus on the lowest scores) or systematically discard unsure classifications and quickly provide data, which, albeit incomplete, can be sufficient for some purposes.

2.2.5 Modelling ecosystem processes

Ecological systems are typically influenced by multiple drivers that may combine cumulatively or interactively often resulting in threshold or non-linear responses. Investigation of relationships between ecosystem processes and external drivers are required to build an understanding of the mechanizm underlying those responses. Regression models, which can allow for non-linear relationships, are often employed but given the complexity of the ecological system may fail to detect deep interactive effects, non-linear or abrupt changes. Such methods often make assumptions on the distribution of data and rely on a single parsimonious model imposing limitations on the form of underlying relationships. These limitations can be overcome by introducing more flexible machine learning methods.

When machine learning analysis was applied to phytoplankton growth, the model was able to describe relationships with environmental variables that are known and verified experimentally (vouching for its relevance) (Thomas *et al.* 2018) and, at the same time, uncovering interactions between these variables that are difficult and costly to test experimentally.

Machine learning methods can also be used to assist mechaniztic prediction models for marine ecosystems that are based upon an understanding of the underlying dynamics. Due to influences of climate changes, marine environments are now more than ever being subjected to conditions that are outside their previously experienced time-series. This forces traditional regression, as well as machine learning models, to extrapolate rather than to interpolate. This can be a risk for both approaches, but integrating the machine learning approach with insights from ecological theory and lab experiments can mitigate this risk. This can be done by developing the underlying understanding necessary for mechaniztic models using machine learning analysis techniques, which are well-suited to identifying complex relationships in unwieldy datasets.

With a shift towards ecosystem-based management, incorporation of complex ecosystem models is paramount. Machine learning is an invaluable tool for both exploratory analysis and modelling ecosystem processes necessary for incorporation of these complex ecosystem variables into fisheries management.

2.2.6 Ecological indicators: WFD and MSFD

Machine learning has been used on work related to estimate or model ecological indicators. For example, Bayesian networks have been used in the Gulf of Finland in relation to the Water Framework Directive (WFW) (Fernandes *et al.*, 2012). Recently, a proof-of-concept and a review of the potential for automatic classification of plankton for Marine Strategy Framework Directive (MSFD) indicators have been published (Uusitalo *et al.*, 2016). However, not all the attempts to use machine learning to model ecological indicators have been successful, probably due to sparse data with not enough spatial and temporal resolution (Rodríguez *et al.*, 2012).

Machine learning also has the potential to guide the selection of the most influential pressures indicators for a given set of response indicators. For example, several different methods exist for quantifying predictor importance from neural network analyses (e.g. see Olden *et al.*, 2004 and references therein; deOna and Garrido, 2014). A workshop presentation highlighted one of these methods (product of standardized weights; e.g. Olden and Jackson, 2006) for selecting the most influential pressures on the Grand Bank fish community over three time periods. Such results can feed into other scientific analyses and modelling studies, such as those quantifying thresholds or "tipping points" for ecosystems under multiple stresses (e.g. Large et. al., 2015).

2.2.7 Spatial planning and decision-making

Machine learning, and more specifically Bayesian networks are being used for marine spatial planning. Other decision-making tools based on Bayesian networks have not considered the spatial component. Bayesian networks in combination with GIS tools are being used to: (1) analyse conflicting uses (e.g. when we have an area declared as interest for aquaculture and we wish to know which fleets would be affected and how to reallocate them with minimal impact to them); (2) develop new activities such as wind energy; and, (3) consider other social and economic aspects (Galparsoro *et al.*, 2009; Galparsor *et al.*, 2010; Galparsoro *et al.*, 2015; Pascual *et al.*, 2011; Cocoli *et al.*, 2018).

2.2.8 Text/literature analysis

The workshop has shown that machine learning has useful applications for textual data too, such that is found in the form of reports and scientific publications. Machine learning can provide automatic ways of analysing large collections of documents, hundreds or thousands, and infer topical content from these documents (Syed & Weber 2018; Syed *et al.*, 2018), or provide a sense of the document content by word frequency. The topical content provides an overview of topics or themes that are present in these documents, whereas a word frequency count can help in quickly finding relevant documents. Such analysis can help in the manual task of reading and understanding the documents and can save time and effort. The automatic analysis can then be used for document classification, searching for documents, comparing documents, quality assurance of document content.

2.2.9 Bid data and machine learning

The difficulties for direct observation of biological interactions and mechanizms have pushed marine science towards the development of large bodies of measurements from where the underlying mechanizms can be deduced. In fact, major international programs (IGBP, JGOFS, GLOBEC, ICES, and others), engaging thousands of marine scientists throughout the world over the past decade, have delivered a massive amount of information on the biogeochemical foundations, functioning and structure of marine foodwebs. Parallel technological developments, ranging from satellite imagery to autonomous underwater vehicles, have increased by orders of magnitude the resolution and amount of data available on relevant properties of the ocean ecosystem. However, there are still challenges in the use of this huge amount of diverse data providing solutions to final users. Other science domains had similar recent technological and data monitoring advancements which have led to a deluge of data over the past two decades. The term big data were coined to capture the meaning of this emerging trend (Hu *et al.*, 2014). In addition to its sheer volume, big data also exhibit other unique characteristics as compared with traditional data. For instance, big data are commonly unstructured and requires more real-time analysis. This development calls for new system architectures for data acquisition, transmission, storage, and large-scale data processing mechanizms (Hu *et al.*, 2014). Big data techniques enhanced by machine learning methods can increase the value of such data and its applicability to society, industry, and management challenges. Big data methodologies allow real-time updating and application of models required to satisfy the fast-paced real-world needs of industry and managers. H2020 project Databio (https://www.databio.eu) aims to use the innovative ICTs and information flows in order to provide a streamlined Big Data Infrastructure for data discovery, retrieval, processing, and visualizing in fisheries (e.g. tuna pilot; Fernandes *et al.*, 2017) and other bioeconomy sectors. In this workshop several times data integration approaches and data processing pipelines where presented that can be considered Big data approaches.

2.2.10 Missing data

Missing data are a common problem in fishery surveys. It has been a common practice for countries to “borrow” data from other countries to fill in their gaps when submitting data for stock assessment. While this problem can better be handed by using traditional statistical models, machine learning approaches can be explored especially where complete variables are missing. An example of a machine learning based method for data missing imputation is the supervised method CMean (Kononenko *et al.*, 1984; Delavallade and Dang, 2007; Fernandes *et al.*, 2013) based in information theory.

Time-series data are crucial in stock assessment yet this is lacking in some countries on some variables. Machine learning was used to reconstruct a time-series for recreational fishers. Other areas where machine learning can be explored include the reconstruction of time-series for discards.

Estimating missing biological information in the commercial fishery is one of the main work done by stock coordinators before data are fed through the stock assessments models. The RDBES will in future replace InterCatch and the RDB and will incorporate statistically sound raising and estimations by using design-based sampling data. It was found that machine learning will best serve in anomaly detection on the submitted data and no direct recommendation was made on any known machine learning algorithm that could improve this process.

2.3 ToR c

Identify options to better include social scientists into ICES processes, through the use of machine learning/deep learning.

The composition of participants was such that the workshop was unable to adequately address this term of reference. Topics that were adressed and of relevance to this ToR, include the Bayesian networks applied to spatial planning and ecosystem services work , and the work presented in text mining presented in previous section. Similarly, so-called sentiment analysis that aims to analyse opinions, sentiments, and emotions expressed in text (Ortigosa-Hernández *et al.*, 2012), may be relevant.

2.4 ToR d

Recommend ways forward, particularly to include experts from outside ICES, and consider further areas of work within ICES where machine learning/deep learning would be particularly applicable. Future data storage options to facilitate machine learning/deep learning could also be considered.

2.4.1 General machine learning challenges

Machine learning methods have great potential for applications in fisheries science but effective adoption is limited by several factors that need to be overcome. This concerns not only the methods themselves, which can often seem opaque or are not well understood, but also the necessary data sources, as well as deployment and how methods are integrated into the existing advisory and scientific process. To ease collaboration and avoid duplication of effort, common data frameworks are necessary, including data standards, APIs, and common databases. It is important to tie close connections between informatics and marine sciences, both to leverage new methods as they are developed, and to stimulate machine learning research in directions relevant to the marine sciences. Automating the interpretation of data will make it easier to combine and compare data from different sources. Interdisciplinary approaches, e.g. combining imaging and genomics, should therefore also be encouraged.

We have identified the following areas that need addressing:

Expertise

- *Aattracting and long-term engagement of expertise in machine learning:* expertise is attractive for industry and it can be difficult to recruit skilled individuals and especially to retain expertise over time;
- *Publication of work:* applications of machine learning combine methods from computer science with problems from other fields and is, by nature, cross-disciplinary. It can therefore be difficult to find appropriate venues for publication, especially since the applications rarely induce compute science breakthroughs or new ideas and concepts in fisheries science.

- *Preservation and sharing of acquired knowledge/competence:* projects are often small and isolated and resulting knowledge is often not retained when the project terminates.

Available data

- *Data quality, organization, and volume:* data are often made available in insufficient volumes and with data qualities adequate for manual curation, but not automatic analysis.
- Data are often collected with *insufficient labeling,* making it difficult to construct the training datasets needed for supervised learning
- There is *no centrally organized venue* for sharing of architectures, trained models, and code for machine learning in marine science

Processes

- *Knowing the requirements* for the subsequent analyses of data (such as the accuracies required) is needed to guide machine learning applications towards areas with high impact
- *Communication between marine scientists and machine learning scientists* is needed to increase awareness of the methods and of their potential applications
- *Implementation and deployment* of developed methods/analyses
- *Machine learning applications could often guide data collection and equipment design*, in particular when machine learning is used for data processing. This is difficult because the scientific communities responsible for data collection and analysis through machine learning are often distant, both as subjects and as organizations.

Acceptance, quality

- Inferring mechanizm from machine learning models is difficult, but crucial to improve our understanding of the patterns and processes modelled and discard the impression that machine learning models are always opaque;
- New methods need to be carefully verified for validity because, while the requirements of many machine learning methods are not very stringent (as opposed to classic inferential statistics for example), the methods can still lead to wrong results if not used properly;
- Machine learning models are different from conventional statistics used by scientists, and might be distrusted or avoided due to unfamiliarity.

2.4.2 Risks and opportunities

From the advisory perspective there are significant benefits and risk of machine learning. These are not homogenous across all types models and given the diversity of machine learning approaches, this summary is a general one and may not apply to all specific cases. The clear benefits are the improved consistency of supervised learning techniques compared to human experts, and the ability to deal with larger data volumes. Unsupervised learning mechanizms lack this consistency as by definition they change their characteristics with new data which may have unexpected consequences on the remainder of the advisory process through time, e.g. revision of historic data. When analysing samples, for example otolith images to ages, humans can easily quality control the process making it suitable for the advisory process.

Predictive models including those that investigate unknown relationships between environmental variables are currently more difficult to diagnose than traditional models due to an unfamiliarity of the majority of environmental practitioners. Outcomes are more difficult to verify by humans. There is a perception that machine learning is insensitive to this risk and often seen as a cure-all. However, there is a developing understanding that these models suffer also from the problems of over parameterization and co-linearity of variables. These models are currently less suitable for the advisory process but should be considered in the scientific process. The focus here should be to develop greater familiarity with the model characteristics and the development of better diagnostic on the risks of over parameterization and col-linearity of variables.

Machine learning applications and in particular deep learning methods, have the possibility to find highly complex patterns in data. However, in general, the methods are data greedy and it is more difficult to explain how machine learning methods achieve their objective, compared to more traditional statistical models.

We must be wary of trusting new models blindly, since their performance or the assumptions that they make can change over time. A clear example of the dangers is the rise and failure of flu prediction by Google (Ginsberg *et al.*, 2009; Lazer *et al.*, 2014).

3 Conclusions and way forward

Machine learning has been successfully applied to data analysis in almost every field and has a large potential also for fisheries sciences, which is heavily based on data. End-users should identify areas and data types of highest relevance to or impact on the advisory processes, and the required or desired accuracy of analysis to guide machine learning research. Improving current processes (rather than establish new ones) is quick wins. However, machine learning techniques cannot just be applied and accepted after they give good predictions on a given dataset (which is often what is emphasized in their applications). To provide solid, long lasting advice one must investigate how they work and whether the mechanizms that they infer make sense. Results and processes to get to these results should be emphasized equally in further research.

Some ways to move forward:

- Provide a venue for sharing knowledge and experience, publishing (organize a conference or special issue on machine learning).
- Stimulate or organize training on using machine learning for fisheries scientists.
- Summer school.
- Encourage long-term research in machine learning specifically targeting marine sciences (collaboration computer science and marine biology).
- Provide a central data storage for images and their annotations. Plankton, but also fish, benthos, otoliths, and scales. Care to make creation of training datasets easy, and publish such training sets.
- Kaggle competition. Data science competitions are good for raising awareness and engaging citizen scientists, but the costs span more than just the prizes, and usually result in outputs that require significant effort to turn into usable products.
- Provide a model zoo for trained models aimed at marine science problems.
- Shared Github or bitbucket repository for code.

4 References

Andonegi, E., Fernandes, J. A., Quincoces, I., Irigoien, X., Uriarte, A., Pérez, A., ... & Stefánsson, G. (2011). The potential use of a Gadget model to predict stock responses to climate change in combination with Bayesian networks: the case of Bay of Biscay anchovy. ICES Journal of Marine Science, 68(6), 1257-1269.

Benfield, M. C., Grosjean, P., Culverhouse, P. F., Irigoien, X., Sieracki, M. E., Lopez-Urrutia, A., ... & Pilskaln, C. H. (2007). RAPID: research on automated plankton identification. Oceanography, 20(2), 172-187.

Boyra, G., Martínez, U., Cotano, U., Santos, M., Irigoien, X., and Uriarte, A. 2013. Acoustic surveys for juvenile anchovy in the Bay of Biscay: Abundance estimate as an indicator of the next year's recruitment and spatial distribution patterns. ICES Journal of Marine Science, 70.

Boyra, G., Peña, M., Cotano, U., Irigoien, X., Rubio, A., and Nogueira, E. 2016. Spatial dynamics of juvenile anchovy in the Bay of Biscay. Fisheries Oceanography, 25.

Boyra, G., Moreno, G., Sobradillo, B., Pe, I., Sancristobal, I., and Demer, D. A. 2018. Target strength of skipjack tuna (Katsuwanus pelamis) associated with fish aggregating devices (FADs).

Coccoli, C., Galparsoro, I., Murillas, A., Pinarbasi, K., Fernandes, J. A. 2018. Conflict analysis and reallocation opportunities in the framework of marine spatial planning: a novel, spatially explicit Bayesian belief network approach for artisanal fishing and aquaculture. Marine Policy. In press.

Colas, F., Tardivel, M., Perchoc, J., Lunven, M., Forest, B., Guyader, G., ... & Sourisseau, M. (2017). The ZooCAM, a new in-flow imaging system for fast onboard counting, sizing and classification of fish eggs and metazooplankton. Progress in Oceanography.

Delavallade, T., Dang, T.H., 2007. Using entropy to impute missing data in a classification task. In: Proceedings of the IEEE International Conference on Fuzzy Systems, vol. 7.

Demer, D. a. 2017. 2016 USA–Norway EK80 Workshop Report: Evaluation of a wideband echosounder for fisheries and marine ecosystem science. 69 pp.

de Ona, J., and Garrido, C. (2014). Extracting the contribution of independent variables in neural network models : a new approach to handle instability. Neural Comput. Appl. 25, 859–869. doi:10.1007/s00521-014-1573-5.

Eigaard, O.R., Bastardie, F., Hintzen, N.T., Buhl-Mortensen, L., Buhl-Mortensen, P., Catarino, R., Dinesen, G.E., Egekvist, J., Fock, H.O., Geitner, K., Gerritsen, H.D., González, M.M., Jonsson, P., Kavadas, S., Laffargue, P., Lundy, M., Gonzalez-Mirelis, G., Nielsen, J.R., Papadopoulou, N., Posen, P.E., Pulcinella, J., Russo, T., Sala, A., Silva, C., Smith, C.J., Vanelslander, B. & Rijnsdorp, A.D. (2017) The footprint of bottom trawling in European waters: distribution, intensity, and seabed integrity. ICES Journal of Marine Science, 74, 847-865.

Fernandes, J., Irigoien, X., Uriarte, A., Ibaibarriaga, L., Lozano, J., & Inza, I. (2009a). Anchovy Recruitment Mixed Long Series prediction using supervised classification. Tech. rep., Working document to the ICES benchmark workshop on short lived species (WKSHORT), Bergen, Norway.

Fernandes, J. A., Irigoien, X., Boyra, G., Lozano, J. A., & Inza, I. (2009b). Optimizing the number of classes in automated zooplankton classification. Journal of Plankton Research, 31(1), 19-29.

Fernandes, J. A., Irigoien, X., Goikoetxea, N., Lozano, J. A., Inza, I., Pérez, A., & Bode, A. (2010). Fish recruitment prediction, using robust supervised classification methods. Ecological Modelling, 221(2), 338-352.

Fernandes, J. A., Kauppila, P., Uusitalo, L., Fleming-Lehtinen, V., Kuikka, S., & Pitkänen, H. (2012). Evaluation of reaching the targets of the Water Framework Directive in the Gulf of Finland. Environmental science & technology, 46(15), 8220-8228.

Fernandes, J. A., Irigoien, X., Lozano, J. A., Inza, I., Goikoetxea, N., & Pérez, A. (2015). Evaluating machine-learning techniques for recruitment forecasting of seven North East Atlantic fish species. Ecological Informatics, 25, 35-42.

Fernandes. J.A., Quincoces, I., Fradua, G, Ruiz, J., Lopez, J., Murua, H., Inza, I., Lozano, J.A., Irigoien, X., Santiago, J. (2017) Fishery pilot B1: Planning of oceanic tuna fisheries - Arrantza B1 kasua: Atun tropikalaren arrantza plangintza. DataBio general assembly 02 (Helsinki), 27-29 June, DOI: 10.13140/RG.2.2.22519.32165.

Galparsoro, I., Á. Borja, J. Bald, P. Liria, G. Chust, 2009. Predicting suitable habitat for the European lobster (Homarus gammarus), on the Basque continental shelf (Bay of Biscay), using Ecological-Niche Factor Analysis. Ecological Modelling, 220: 556-567.

Galparsoro, I., Á. Borja, I. Legorburu, C. Hernández, G. Chust, P. Liria, A. Uriarte, 2010. Morphological characteristics of the Basque continental shelf (Bay of Biscay, northern Spain); their implications for Integrated Coastal Zone Management. Geomorphology, 118: 314-329.

Galparsoro, I., J. G. Rodríguez, I. Menchaca, I. Quincoces, J. M. Garmendia, Á. Borja, 2015. Benthic habitat mapping on the Basque continental shelf (SE Bay of Biscay) and its application to the European Marine Strategy Framework Directive. Journal of Sea Research, 100: 70-76.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012.

Gorsky, G., Ohman, M. D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J. B., ... & Prejger, F. (2010). Digital zooplankton image analysis using the ZooScan integrated system. Journal of plankton research, 32(3), 285-303.

Hernández-González, J., Inza, I., & Lozano, J. A. (2016). Weak supervision and other non-standard classification problems: a taxonomy. Pattern Recognition Letters, 69, 49-55.

Hernández-González, J., Inza, I., Granado, I., Oihane C. Basurko, Fernandes, J. A., Lozano, J. A. 2018 Aggregated outputs by linear models: An application on waste accumulation prediction. Knowledge-Based Systems, in review.

Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. IEEE access, 2, 652-687.

Irigoien, X., Fernandes, J. A., Grosjean, P., Denis, K., Albaina, A., & Santos, M. (2008). Spring zooplankton distribution in the Bay of Biscay from 1998 to 2006 in relation with anchovy recruitment. Journal of plankton research, 31(1), 1-17.

Jensen, F., Nielsen, T., 2001. Bayesian Networks and Decision Graphs. Springer-Verlag, New York, NY, USA.

Kononenko, I., Bratko, I., Roskar, E., 1984. Experiments in automatic learning of medical diagnostic rules. In: International School for the Synthesis of Expertise Knowledge Workshop, Bled, Slovenia.

Large, S. I., Fay, G., Friedland, K. D., and Link, J. S. (2015). Critical points in ecosystem responses to fishing and environmental pressures. Mar. Ecol. Prog. Ser. 521, 1–17. doi:10.3354/meps11165.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. Science, 343(6176), 1203-1205.

Lopez, J., Moreno, G., Boyra, G., and Dagorn, L. 2016. A model based on data from echosounder buoys to estimate biomass of fish species associated with fish aggregating devices. Fishery Bulletin, 114.

McCauley, D. J., Woods, P., Sullivan, B., Bergman, B., Jablonicky, C., Roan, A., ... & Worm, B. (2016). Ending hide and seek at sea. Science, 351(6278), 1148-1150.

Mori, U., Mendiburu, A., Keogh, E., & Lozano, J. A. (2017). Reliable early classification of time series based on discriminating the classes over time. Data Mining and Knowledge Discovery, 31(1), 233-263.

Olden, J. D., Joy, M. K., and Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Modell. 178, 389–397. doi:10.1016/j.ecolmodel.2004.03.013.

Olden, J. D., Joy, M. K., and Death, R. G. (2006). Rediscovering the Species in Community-Wide Predictive Modeling. Ecol. Appl. 16, 1449–1460.

Ortigosa-Hernández, J., Rodríguez, J. D., Alzate, L., Lucania, M., Inza, I., & Lozano, J. A. (2012). Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers. Neurocomputing, 92, 98-115.

Pascual, M., A. Borja, S. V. Eede, K. Deneudt, M. Vincx, I. Galparsoro, I. Legorburu, 2011. Marine biological valuation mapping of the Basque continental shelf (Bay of Biscay), within the context of marine spatial planning. Estuarine, Coastal and Shelf Science, 95: 186-198.

Rodríguez J.G., Fernandes J.A., Garmendia J.M., Muxica I., Borja A. (2012) Development of a Bayesian Networks based method for assessing the status of hard bottom substrata biota. XVII Simposio Ibérico de Estudios de Biología Marina. 11-14 September. Donostia-San Sebastián (Spain).

Syed, S., Borit, M., Spruit, M. (2018) Narrow lenses for capturing the complexity of fisheries: A topic analysis of fisheries science from 1990 to 2016. Fish and Fisheries.

Syed, S., Weber, C. T. (2018) Using Machine Learning to Uncover Latent Research Topics in Fishery Models. Reviews in Fisheries Science & Aquaculture, 26(3), 319-336.

Uusitalo, L., Fernandes, J. A., Bachiller, E., Tasala, S., & Lehtiniemi, M. (2016). Semi-automated classification method addressing marine strategy framework directive (MSFD) zooplankton indicators. Ecological indicators, 71, 398-405.

Thomas, M.K., Fontana, S., Reyes, M., Kehoe, M., & Pomati, F. (2018). The predictability of a lake phytoplankton community, over time-scales of hours to years. Ecology Letters, 21, 619-628.

Uranga, J., Arrizabalaga, H., Boyra, G., Hernandez, M. C., Goñi, N., Arregui, I., Fernandes, J. A., *et al.* 2017. Detecting the presence-Absence of bluefin tuna by automated analysis of medium-range sonars on fishing vessels. PLoS ONE, 12.

Annex 1: List of participants

Member	Institute	E-mail
Benjamin Woodward	CVisionconsulting, United States	benjamin.woodward@cvisionconsulting.com
Callum Scougal	Centre for Environment, Fisheries and Aquaculture Science	callum.scougal@cefas.co.uk
Carlos Pinto	International Council for the Exploration of the Sea	carlos@ices.dk
Cooper Hoffman Van Vranken	Danmarks Teknologisk Universitet	s160957@student.dtu.dk
Daniel Benden	Wageningen Marine Research	daniel.benden@wur.nl
Danielle Dempsey	Dalhousie University	Danielle.Dempsey@dal.ca
Dionysios Krekoukiotis	DTU Aqua -National Institute of Aquatic Resources	dikr@aqua.dtu.dk
Edwin van Helmond	Wageningen University & Research	edwin.vanhelmond@wur.nl
Esha Mohamed	Swedish University of Agricultural Sciences. Department of Aquatic Resources-SLU Aqua	esha.mohamed@slu.se
Frankwin van Winsen	Institute for Agricultural and Fisheries Research (ILVO)	frankwin.vanwinsen@ilvo.vlaanderen.be
Jean-Baptiste Romagnan	Ifremer Nantes Centre	jean.baptiste.romagnan@ifremer.fr
Jean-Olivier Irisson	Laboratoire d'Océanographie	irisson@normalesup.org
Jenni Fincham	Centre for Environment, Fisheries and Aquaculture Science	jenni.fincham@cefas.co.uk
Jens Rasmussen	Marine Science Scotland, Marine Laboratory	J.Rasmussen@marlab.ac.uk
Jose Fernandes	Plymouth Marine Laboratory	jfernandes@azti.es

Kadji Okou	International Council for the Exploration of the Sea	kadji.okou@ices.dk
Ketil Malde	Institute of Marine Research	ketil@malde.org
Kristian Plet-Hansen	DTU Aqua -National Institute of Aquatic Resources	kspl@aqua.dtu.dk
Lisa Peterson	NOAA Fisheries	Lisa.Peterson@noaa.gov
Mark Payne	DTU Aqua -National Institute of Aquatic Resources	mpa@aqua.dtu.dk
Maurizio Gibin	Institute for the Protection and Security of the Citizen	Maurizio.GIBIN@ec.europa.eu
Mridul Thomas	DTU Aqua -National Institute of Aquatic Resources	mrit@aqua.dtu.dk
Neil Holdsworth	International Council for the Exploration of the Sea	neilh@ices.dk
Nils Olav Handegard	Institute of Marine Research	nils.olav.handegard@hi.no
Olga Lyashevska	Marine and Freshwater Reseach Center	olga.lyashevska@gmit.ie
Rainer Kiko	GEOMAR Helmholtz Centre for Ocean Research	rkiko@geomar.de
Sarah Margolis	NOAA Fisheries	sarah.margolis@noaa.gov
Shaheen Syed	Utrecht University - Information and computing sciences (www.saf21.eu)	s.a.s.syed@uu.nl
Sven Kupschus	Lowestoft Laboratory	Sven.Kupschus@cefas.co.uk

Annex 2: Meeting agenda

MONDAY 16 APRIL 2018

Time	Name(s)	**Introduction**
1300	Shaheen Syed, Ketil Malde, Julie Krogh Hallin	Welcome and practical information
1315		Round table introductions
1415		Coffee break
1430	Ketil Malde, Shaheen Syed	Recent developments in machine learning and current state of the art.

We aim to end at about 1700.

TUESDAY 17 APRIL 2018

Time	Name(s)	**ICES overview, groups and activities**
0900	Neil Holdsworth	ICES data strategy / ICES strategic plan
	Mark Payne	IWGS2D - Seasonal To Decadal Predictions of Marine Ecosystems
	Morgane Travers-Trolet	WGIPEM - Integrative Physical-Biological and Ecosystem Modeling
	Maurizio Gibin	WGSFD - Spatial Fisheries Data
		Coffee break
	Nils Olav Handegard	WGFAST - Fisheries, Acoustics, Science and Technology
	Sven Kupschus	EOSG - Ecosystem Observation Steering group

	Neil Holdsworth	*An overview of ICES data collections and EU data infrastructures*
		Discussion and summary

1230-1330 Lunch break

	Name(s)	**The fisheries science processes**
1330	Sven Kupschus, Lotte Worsøe Clausen	Overview of the fisheries science process
	Nils Olav Handegard	The proessing pipeline from observations to stock indices - the IMR model.
	Colin Millar or Arni Magnusson?	*TAF - the ICES Transparent Asessment Framework*
	Sarah Margolis	*The NOAA fisheries management and the role of optical underwater survey data*
	Josean Fernandes	*Bayesian networks for Anchovy stock assessments in the Bay of Biscay*
		Discussion and summary

Approx. 1600-1800 Socializing and refreshments in the ICES lunchroom.

WEDNESDAY 18 APRIL 2018

Time	Name(s)	**Plankton**
0900	N.N.	ICES advisory processes and the role of plankton data

	Josean Fernandes	Automatic classification of plankton and MSFD indicators
	Jean-Baptiste Romagnan	Plankton imaging devices: combining instruments to describe a complete plankton ecosystem.
	Jean-Olivier Irisson	EcoTaxa: a human-computer interface to classify images along a taxonomy with the help of machine learning
		Coffee break
	Rainer Kiko	Combining deep learning, in site imaging and citizen science to resolve the distribution of zooplankton in major upwelling regions.
	Callum Scougal	Application of machine learning in regards to Zooplankton classification.
	Jean-Olivier Irisson	Imperfect automatic image classification successfully describes plankton distribution patterns.
	Mridul Thomas	Assessing the predictability of and mechanisms driving phytoplankton dynamics
		Discussion and summary

1230-1330 Lunch break

Time	Name(s)	**Acoustics and related data acquisition**
1330	Jenni Fincham	Developing a habitat classification based on multibeam data transformed to images
	Jean-Baptiste Romagnan	Acoustics in the framework of integrated ecosystemic surveys - data types, strength and weaknesses.

	Benjamin Woodward	Underwater optical trawl surveys for fisheries populations
	Josean Fernandes	Acoustics, sonar and underwater imagery for stock assessments
	Ketil Malde	The Deep Vision trawl camera and automatic fish species recognition

1500 Coffee break

	Name(s)	**Fisheries monitoring**
1530	Lisa Peterson	Electronic monitoring of commercial fishing vessels in the U.S.
	Maurizio Gibin	Estimating fishing effort using AIS and data fusion as well the consequent fishing effort validation through VMS.
	Edwin van Helmon	Current state of REM in the EU, summary of project and research, technical limitations, strengths, weaknesses of REM over the last 10 years.
	Benjamin Woodward	Bycatch/scallops
	Kristian Schreiber Plet-Hansen	ML/DL for CCTV
	Danielle Dempsey	Using neural networks to model fish community biomass on the Grand Bank, Newfoundland (PhD research).
		Discussion and summary

THURSDAY 19 APRIL 2018

Time	Name(s)	**Ecosystems**
0900	Olga Lyash-evska	Explaining trends in length-at-age of herring using gradient boosting regression trees.
	Ketil Malde	Predicting the age of Greenland halibut from otolith images
	Josean Fer-nandes	Decision Support Tools, ecosystem indicators and spa-tial planning with Bayesian networks.
	Benjamin Wood-ward	Underwater visual imagery.
	Dionysios Krekoukio-tis	Assessing the Role of Environmental Factors on Baltic Cod Recruitment using Artificial Neural Networks

1030-1100 Coffee break

	Name(s)	**Text mining - human data**
1100	Shaheen Syed	Topic modelling with Latent Dirichlet Allocation.
	Carlos Pinto, Sarah Lousie Millar	Text Mining of the ICES knowledge base.
		Discussion

1230-1330 Lunch

	Name(s)	**Working with missing data**
1330	Esha Mohamed	Using machine learning techniques to reconstruct time series for recreational fishers: our experiences.
	Kadji Okou, Henrik Kjems-Nielsen	Regionally coordinated database for Fishery assessment in the North Atlantic Ocean, the North Sea and the Baltic Sea.

1430 Discussion and workshop summary

FRIDAY 20 APRIL 2018

0900- Show your application / code / architecture

Annex 3: Terms of Reference

WKMLEARN - Workshop on Uses of Machine Learning in Marine Science

2017/2/EOSG11

A **Workshop on Uses of Machine Learning in Marine Science** (WKMLEARN), co-chaired by Ketil Malde*, Norway and Shaheen Syed*, Netherlands/UK, will be established and will meet in ICES HQ, Copenhagen, 16-20 April 2018 to:

a) Review ICES Fisheries Science processes to understand where machine learning and/or deep learning may be of greatest benefit, including:
   i) Survey and data collection,
   ii) Data handling,
   iii) Analysis and assessment,
   iv) Review and advice
   v) Check degree to which expert groups are meeting their terms of reference
   vi) Taking a forward look and consider emerging topics;

b) Identify areas of marine science, data and advice within the ICES remit where machine learning/deep learning has already been applied;

c) Identify options to better include social scientists into ICES processes, through the use of machine learning/deep learning, ;

d) Recommend ways forward, particularly to include experts from outside ICES, and consider further areas of work within ICES where machine learning/deep learning would be particularly applicable. Future data storage options to facilitate machine learning/deep learning could also be considered.

WKMLEARN will report by 31 May 2018 for the attention of the Advisory and Science Committees.

Supporting information

Priority	The Workshop will explore an area of science and technology that is rising rapidly in its ability to support science and which has the potential to replace a number of traditional activties within the fishery science process. ICES needs to understand how best to respond to theese developments.
Scientific justification	Term of Reference a) Machine Learning (and/or Deep Learning) can be used in many ways – from text analysis to finding hidden patterns in large datasets, to analysing images and video, and to deriving analytical algorithms. All forms of machine learning will be considered in examing each stage of fish stock assessment and advice. Term of Reference b) Machine Learning has been applied to determining numbers of salmon lice on farmed fish, identifying fish species from trawl cameras, interpreting fish scales, classifying fish behavior, and interpreting acoustics data through use of image analysis. It has also been applied to analysis of marine science literature to determine trends

	in research and publication. Participants who can provide further examples will be specifically sought. Term of Reference c) Among the challenges in bringing more social science into the traditional fisheries science and advice process has been the lack of a common language – with specialist terms being used in both areas that may not have any meaning elsewhere. Machine learning can help overcome such barriers. One option might be to match trends in social science data and publications with trends in fisheries science literature. The overall aim would be to facilitate the further inclusion of the social sciences in ICES processes. It may also be useful to identify research areas that can be addressed through multi-/inter-disciplinary computational social science approaches to study social processes relevant to fisheries. Term of Reference d) The Terms of Reference for this workshop have been kept deliberately constrained so as not to overload its work. Lessons learned from the workshop should be considered and a path forward recommended.
Resource requirements	It is hoped that participants will have sufficient access to computing resources so as to not require any further input.
Participants	Participants will be sought from as wide a community as is possible. We would hope to attach scientists with skills in surveying, stock assessment, social aspects, experience in ICES processes including advice and inter-disciplinary scientists. Scientists with access to complex datasets would be welcomed also. Early career scientists with skills in machine learning would be particularly welcome
Secretariat facilities	The Atlantic Room for 3 days, and the usual welocome Secretariat support.
Financial	No financial implications.
Linkages to advisory committees	Directly linked
Linkages to other committees or groups	Directly linked, and potentially to all SCICOM steering groups. Science Impact and Publications Group would be interested in bibliometric and citation analysis.
Linkages to other organizations	None at present