

MACHINE LEARNING APPLICATIONS IN OCEANOGRAPHY

Hafez Ahmad 

Cite this article as:

Ahmad, H. (2019). Machine learning applications in oceanography. *Aquatic Research*, 2(3), 161-169. <https://doi.org/10.3153/AR19014>

University of Chittagong, Faculty of
Marine Sciences and Fisheries,
Department of Oceanography,
Bangladesh

ORCID IDs of the author(s):

H.A. 0000-0001-9490-9335

Submitted: 16.06.2019

Revision requested: 17.06.2019

Last revision received: 24.06.2019

Accepted: 28.06.2019

Published online: 08.07.2019

Correspondence:

Hafez AHMAD

E-mail: hafazahmad100@gmail.com

ABSTRACT

Machine learning (ML) is a subset of artificial intelligence that enables to take decision based on data. Artificial intelligence makes possible to integrate ML capabilities into data driven modelling systems in order to bridge the gaps and lessen demands on human experts in oceanographic research. ML algorithms have proven to be a powerful tool for analysing oceanographic and climate data with high accuracy in efficient way. ML has a wide spectrum of real time applications in oceanography and Earth sciences. This study has explained in simple way the realistic uses and applications of major ML algorithms. The main application of machine learning in oceanography is prediction of ocean weather and climate, habitat modelling and distribution, species identification, coastal water monitoring, marine resources management, detection of oil spill and pollution and wave modelling.

Keywords: Machine learning, Application, Oceanography, Data driven



Introduction

Machine Learning (ML) is a discipline of the computer science that develops dynamic algorithms capable to produce data-driven decisions (Thessen, 2016). ML has proven itself to be an answer to many real world problems with its capabilities. ML has advantage over the traditional methods because it is able to build a model, which is highly dimensional and nonlinear data with complex relations and missing values. ML has proven useful for a very large number of applications in many parts of the Earth system (land, ocean, and atmosphere) and beyond, from retrieval algorithms, crop disease detection, new product creation, bias correction and code acceleration (Yi and Prybutok, 1996).

Large amount of data which is collected by scientific instruments then separated into train set and test set. Therefore ML algorithms are trained by this data. Then build model with high accuracy and its parameters are optimized based on sample data during the learning step. During prediction, the model parameters are used to infer results on the previously unseen data.

ML has multiple algorithms, techniques and methodologies that can be used to build models to solve real world problems using oceanographic data. A supervised Learning (SL) is a type of ML algorithm that uses labelled data. After that, the machine is provided with new set of data so that SL algorithms analyse the training data and produce a correct outcome from the labelled data. SL mainly tries to model the relationship between the inputs and their corresponding outputs from the training data so that we would be able to predict the output based on the knowledge it gained earlier with regard to relationships. SL are classified into two major categories. A. classification and B. regression.

Unsupervised learning (USL) is the training of the machine using data that is neither classified nor labelled. The task of the machine is to group unsorted data based on the similarities, patterns and differences without any guidance. USL can be classified into following the categories a. clustering, b. dimensionality reduction c. anomaly detection.

The reinforcement learning (RL) methods are slightly different from SL or USL. RL is a type of ML where an agent learns

how to behave in the environment by performing actions and thereby drawing intuitions and seeing the results.

Deep learning (DL) is the subset of ML concerned with algorithms inspired by the structure and function of the human brain called artificial neural network. Neural networks (NNs) come in several forms such as recurrent neural networks, convolutional neural networks, and artificial neural networks and feed forward neural networks. An ANN is an interconnected group of nodes. Here, each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another. Model comprises synaptic links which allow the inputs (x_1, x_2, \dots, x_n) to be measured by applying the weights (w_1, w_2, \dots, w_n).

Methodology

This study was based on the syntheses of secondary information. To collect data, an intensive literature review related to the machine learning applications and scope of machine learning in oceanography was done. The context were conducted through an online and offline mode. In addition, relevant documents and reports were also collected from the websites and published research articles personal contacts. Open source software python and R as well as commercial software adobe illustrator were used for data analysis and visualization (Figure 1).

Necessity of the Machine Learning Approach for Oceanographic Research

The ocean is vast, dynamic and complex. Data structure of the ocean becomes increasingly complex and large. Generally, coastal zone is vulnerable to natural disasters like sea level rise (SLR), coastal flooding, erosion etc. For the coastal zone management and flood erosion control, a reliable and accurate tool for prediction and forecasting of coastline evolution and inundation by water is needed in order to minimize coast protection and conservation. For this reason, traditional data analysing methods are time consuming and costly, even in some cases, analysis is not possible in conventional way. ML techniques are robust, fast and highly accurate.

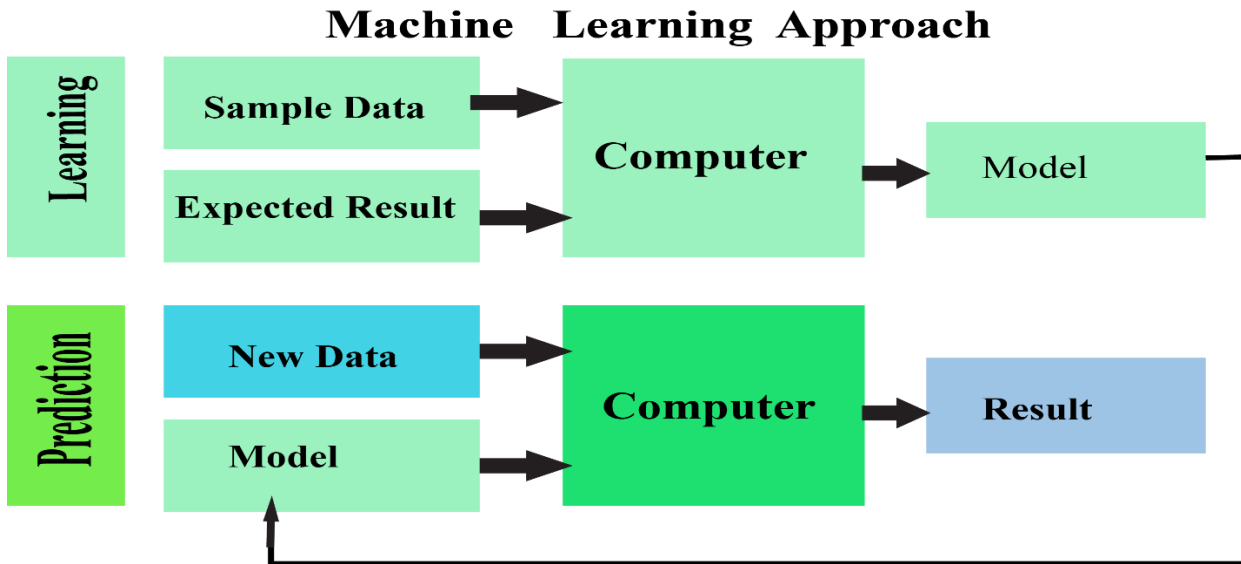


Figure 1. Simple Machine learning working approach (created by adobe illustrator CS6)

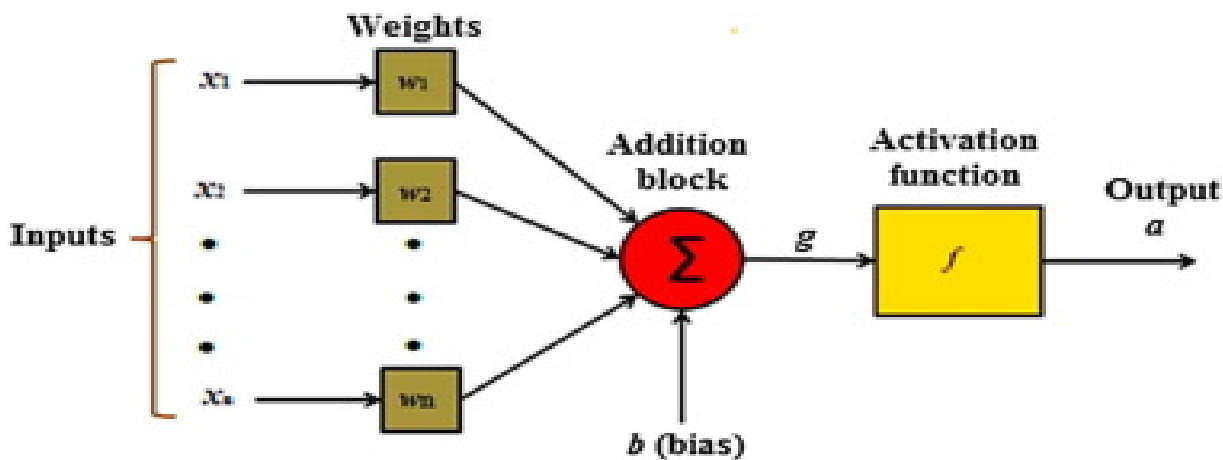


Figure 2. simple artificial neural network (Burkitt, 2006; Oja, 1982; Turkson et al., 2016)

Common Machine Learning Applications in Oceanography

Oceanic climate prediction and forecasting

Advancements in ML, in combination with optimization methods are promising to balance the performance of forecast and the earliness of those forecasts (Mori et al., 2017). The most common ML methods used in meteorological forecasting are genetic algorithms, which have been used to model rainy vs. non-rainy days (Haupt, 2009). Machine learning methods have been applied to forecast coastal sea level fluctuations

(Hsieh, 2009). ML is used to study important processes such as El Niño, sea surface temperature anomalies, and monsoon models (Cavasos et al., 2002; Hsieh, 2009; Krasnopolsky, 2009; Thessen, 2016). The oceanography community makes extensive use of neural networks for forecasting sea level, waves, and sea surface temperature (Hsieh, 2008; Forget et al., 2015). Wu et al. (2006) developed an MLP NN model to forecast the sea surface temperature (SST) of the entire tropical pacific ocean where sea level pressure and SST were used as predictor to predict.

Species identification

Identification of small and large size marine taxa require specialized knowledge, which is one of the bottlenecks in oceanographic studies. This limitation can be solved by ML approach with high accuracy (automatic identification techniques). Recent advances in the ML are promising with regard to improving accuracy of automated detection and classification of marine organisms from high volume data such as images and video (Olson and Sosik 2007). Generally, ML algorithms are trained on images, videos, sounds and other types of data labelled with taxon names. Trained algorithms can then automatically annotate new data and this methods are used to identify plankton, shellfish larvae from images, bacteria from gene sequences, cetacean from audio, fish and algae from acoustic and optical characteristics (Simmonds and Armstrong, 1996; Boddy, 1999; Jennings et al., 2008; Goodwin and North 2014).

Detection of ocean pollution

ML can be used in detection of ocean pollution with the help of satellite and radar images such as oil spills, plastics pollution, algal bloom etc. Oil spill detection currently requires a highly trained human operator to assess each region in each image (Kubat et al., 1998). Del Frate et al. (2000) used MLP NN models to detect oil spill on the ocean surface from synthetic aperture radar (SAR) images.

Marine and coastal water monitoring

A multilayer preceptor neural networks model was developed to derive the concentrations of phytoplankton pigment, suspended sediments and gelbstoff, and aerosol over turbid coastal waters from satellite data (Tanaka et al., 2004). ML methods are also used in coastal water monitoring (Kim et al., 2014). 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 (Lewis et al., 2001).

Sedimentation modelling

Sedimentation is an important phenomenon in the coastal oceanography among ML methods, ANN has widely used in various water related research such as rain runoff modelling, modelling stage discharge relationship (Bhattacharya and Solomatine 2005). ML models that predict sedimentation in the harbor basin of the port of Rotterdam (Bhattacharya and

Solomatine, 2006). Random forest ML approach has been applied to the mapping marine substrates (Hasan et al., 2012; Diesing et al., 2014).

Coastal morphological and morphodynamic modeling

A variety of coastal morphology and morphodynamic models have been built by using the ML (Goldstein et al., 2018). ML models are widely used in the applications of sediment transport, morphology and detection of coastal changes through videos, images. Nonlinear ML forecasting techniques were used to predict suspended sediment concentration based on instantaneous water velocity (Goldstein et al. 2018). ANN was also used to predict the depth integrated alongshore sediment transport using water depth, wave height, wave period and alongshore current velocity (van Maanen et al., 2010). ANN was used to determine the correlation between sandbar morphology and a given wave climate, culminating in examining the nonlinear dependencies of bar position on past wave conditions (Múnera et al., 2014).

Habitat modelling and species distribution

Understanding the habitat and distribution of marine species are important tasks for management and conservation of oceanography. An algorithm can be trained using a large data set matching environmental variables to taxon abundance or presence/absence data. If the algorithm tests well, it can be given a suite of environmental variables from a different location to make predictions on what taxa are present. This technique has been used to identify current suitable habitat for specific taxa, model future species distributions including predicting invasive and rare species presence, and predict biodiversity of an area (Thessen, 2016).

Wind and wave modelling

Ocean wave modelling and prediction are important for a maritime country because there are numerous reasons behind this. For example shipping routes can be optimized by avoiding rough sea thereby reducing time spent during transportation (James et al., 2018). Accurate forecasts of ocean wave heights and directions are a valuable resource for many marine-based industries (O'Donncha, 2017). We may apply machine learning techniques is to predict wave conditions in order to replace a computationally intensive physics-based model by straightforward multiplication of an input vector by mapping matrices resulting from the trained machine learning models (James et al., 2018). Horstmann et al. (2003) used multilayer perceptron (MLP) NN models to retrieve wind

speed s globally at about 30 m resolution from SAR data (Horstmann et al., 2003).

Ocean current prediction

Generally, ROMS is widely used for ocean dynamic process analysis. It is possible to improve the prediction of ocean currents using (historical data) data-driven machine learning methods (Hollinger et al., 2012). For example, neural networks have been used to build Reynolds average turbulent models (Bolton and Zanna, 2019).

Marine and coastal resources management

ML models have ability to capture complex, nonlinear relationships in the input data which are the crucial building blocks for the implementation of ecosystem based fisheries management (Lewis et al., 2001). Taking right inferences about marine conservation and management can be very difficult as there is not sufficient data for certainty and the consequences of their existence can be disastrous. ML methods can provide a tool for increasing certainty and improving results especially techniques that incorporate Bayesian probabilities (Thessen, 2016). ML and more specifically Bayesian networks are being used for marine spatial planning in cooperation with GIS (Lewis et al., 2001).

The goal of this review paper is to give a clear idea about ML applications in oceanographically different areas. Traditional Data driven research is time consuming, even not integrated and dynamic nature. Furthermore, the extent of our training, testing, and field evaluation data ensures that the approach is robust and reliable across a range of conditions (i.e., changes in taxonomic composition and variations in image quality related to lighting and focus (Olson and Sosik, 2007). ML methods has great potentials for applications in oceanography but effective adoption is limited by several factors that need to be eliminated. 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 (Headquarters, 2018).

Common ML methods for resources management are genetic algorithms (Haupt, 2009), neural networks (Brey and Jarre-Teichmann, 1996), support vector machines (Guo and Kelly, 2005), fuzzy inferences systems (Tscherko and Kandeler, 2007), decision tree (Jones and Fielding, 2006) and random forest (Quintero et al., 2014).

Table 1. Machine learning algorithms and scope of applications in oceanography

No.	Types	Major Machine learning algorithms	Scope and potentials of application
1	Supervised	Linear regression	
2		Support vector machine Support vector regression	1.Oil spill mapping and detection 2.Satellite image processing for land use 3.Retrieve ocean surface chlorophyll concentration 4.Habitat modeling
3		Decision tree	1.Resources management 2.Sediment properties
		Random forest	Mapping of marine substrates
		Naïve Bayes	
4	Unsupervised	k-means	Clustering ocean biomes
5		PCA	
6	Reinforcement	Markov decision process	1.Quickly detect hazardous weathers 2.Detection of whale acoustics
7	Deep learning		1.Wave modelling 2.Coastal water monitoring 3. prediction of coastal morphologic properties

Recommendations: Some steps can be taken to improve ML models in oceanography.

1. Constant Engagement of oceanographic expertise in ML.
2. Preservation and sharing acquired knowledge of ML among community.
3. Collected data of Ocean should be available for ML model experiments such as “www.kaggle.com”.
4. Communication between oceanographers and machine learning scientist is needed for awareness and potentials of applications.
5. Machine learning scientists could cooperate ocean scientists for data collection and equipment designing.
6. Motivation and encourage for long term ML research in oceanographic applications.
7. Some events in schools, college and university, competition of ML in oceanography can be effective.

Conclusion

This work investigates various machine learning techniques for the oceanographic data analysis and future opportunities. ML offers a diverse number of methods that are accessible to researchers and fitted in oceanographic applications which is heavily based on data. This approach offers significant advantages in real life operational applications. They have great potential to improve the quality of oceanographic research approaches by creating more accurate models. ML might be used in large oceanographic datasets to discover hidden patterns and trends. The success of the ML approach strongly depends on the adequacy of the data set used for the training. The data availability, precision, quality, representativeness, and amount are the crucial elements for success in this type of ML application. ML also requires interdisciplinary collaboration, communication, technical knowledge on programming and financial support.

Compliance with Ethical Standard

Conflict of interests: The author declare that for this article they have no actual, potential or perceived conflict of interests.

Acknowledgement: I would like to express my sincere thanks to all those who provided me documents and published papers to complete this review. And I am highly motivated by the popularity of “<https://www.kaggle.com>”. This website provided me with the taste

of machine learning. And there has been no financial support for this work.

References

- Bhattacharya, B., Solomatine, D.P. (2005). Neural networks and M5 model trees in modelling water level-discharge relationship. *NeuroComputing*, 63, 381-396.
<https://doi.org/10.1016/j.neucom.2004.04.016>
- Bhattacharya, B., Solomatine, D.P. (2006). Machine learning in sedimentation modelling. *Neural Networks*, 19(2), 208-214.
<https://doi.org/10.1016/j.neunet.2006.01.007>
- Boddy, L.M.C. (1999). *Machine Learning Methods for Ecological Applications* (p. 37-88 pp). Springer US, New York.
https://doi.org/10.1007/978-1-4615-5289-5_2
- Bolton, T., Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, 11(1), 376-399.
<https://doi.org/10.1029/2018MS001472>
- Brey, T., Jarre-Teichmann A.B.O. (1996). Artificial neural network versus multiple linear regression: Predicting P/B ratios from empirical data. *Marine Ecology Progress Series*, 140, 251-256.
<https://doi.org/10.3354/meps140251>
- Burkitt, A.N. (2006). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological Cybernetics*, 95(1), 1-19.
<https://doi.org/10.1007/s00422-006-0068-6>
- Hasan, R.C., Ierodiaconou, D., Monk, J. (2012). Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. *Remote Sens*, 4, 3427-3443.
<https://doi.org/10.3390/rs4113427>
- Cavazos, T., Comrie, A.C., Liverman, D.M. (2002). Intraseasonal Variability Associated with Wet Monsoons in Southeast Arizona, *Journal of Climate*, 15, 2477-490.

- [https://doi.org/10.1175/1520-0442\(2002\)015<2477:IVAWWM>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2477:IVAWWM>2.0.CO;2)
- Diesing, M., Green, S.L., Stephens, D., Lark, R.M., Stewart, H.A., Dove, D. (2014). Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, 84, 107-119.
<https://doi.org/10.1016/j.csr.2014.05.004>
- Goldstein, E.B., Coco, G., Plant, N.G. (2018). A review of machine learning applications to coastal sediment transport and morphodynamics.
<https://doi.org/10.31223/osf.io/cgzvs>
- Del Frate, F., Petrocchi, A., Lichtenegger, J., Calabresi, G. (2000). Neural networks for oil spill detection Using ERS-SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5), 2282-2287.
<https://doi.org/10.1109/36.868885>
- Forget, G., Campin, J., Heimbach, P., Hill, C.N., Ponte, R.M. (2015). ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, 8, 3071-3104.
<https://doi.org/10.5194/gmdd-8-3653-2015>
- Goodwin, J., North, E., Thompson, C.M. (2014). Evaluating and improving a semi-automated image analysis technique for identifying bivalve larvae. *Limnology and Oceanography: Methods*, 12, 548-562.
<https://doi.org/10.4319/lom.2014.12.548>
- Guo, Q., Kelly, M., Graham, C.H. (2005). Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling*, 182(1), 75-90.
<https://doi.org/10.1016/j.ecolmodel.2004.07.012>
- Haupt, S.E. (2009). Environmental Optimization: Applications of Genetic Algorithms. In: Haupt S.E., Pasini A., Marzban C. (eds) *Artificial Intelligence Methods in the Environmental Sciences*. Springer, Dordrecht, p. 379-380.
https://doi.org/10.1007/978-1-4020-9119-3_18
- Headquarters, I. (2018). ICES WKMLEARN 2018 REPORT Report of the Workshop on Machine Learning in Marine Science (WKMLEARN) International Council for the Exploration of the Sea, (April), 16-20.
- Hollinger, G. A., Pereira, A., Ortenzi, V., Sukhatme, G. S. (2012). Towards Improved Prediction of Ocean Processes Using Statistical Machine Learning. In *Robotics: Science and Systems Workshop on Robotics for Environmental Monitoring*, Sydney, Australia, Jul 2012.
<http://robotics.usc.edu/publications/downloads/pub/775/> (accessed 23.12.2018)
- Hsieh, W.W. (2009). *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511627217>
- James, S.C., Zhang, Y., O'Donncha, F. (2018). Forecast Wave Conditions. *Coastal Engineering*, 137, 1-10.
<https://doi.org/10.1016/j.coastaleng.2018.03.004>
- Jennings, N., Parsons, S., Pocock, M.J.O. (2008). Human vs. machine: identification of bat species from their echolocation calls by humans and by artificial neural networks. *Canadian Journal of Zoology*, 86(5), 371-377.
<https://doi.org/10.1139/Z08-009>
- Horstmann, J., Schiller, H., Schulz-Stellenfleth, J., Lehner, S. (2003). Global wind speed retrieval from SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 41(10), <https://doi.org/10.1109/TGRS.2003.814658>
- Jones, M., Fielding, A., Sullivan, M. (2006). Analysing extinction risk in parrots using decision trees. *Biodivers Conserv*, 15(6), 1993-2007.
<https://doi.org/10.1007/s10531-005-4316-1>
- Kim, Y.H., Im, J., Ha, H.K., Choi, J.K., Ha, S. (2014). Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GIScience and Remote Sensing*, 51(2), 158-174.
<https://doi.org/10.1080/15481603.2014.900983>
- Krasnopolsky V.M. (2009) Neural Network Applications to Solve Forward and Inverse Problems in Atmospheric and Oceanic Satellite Remote Sensing. In: Haupt S.E., Pasini

- A., Marzban C. (eds) Artificial Intelligence Methods in the Environmental Sciences. Springer, Dordrecht. p. 191-205.
https://doi.org/10.1007/978-1-4020-9119-3_9
- Kubat, M., Holte, R.C., Matwin, S. (1998). Machine Learning for the detection of oil spills in satellite radar Images. *Machine Learning*, 30(2-3), 195-215.
<https://doi.org/10.1023/A:1007452223027>
- Lewis, J.M., Weinberger, K.Q., Saul, L.K. (2001). Mapping Uncharted Waters : Exploratory Analysis, Visualization, and Clustering of Oceanographic Data 2821 Mission College Blvd.
- 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.
<https://doi.org/10.1007/s10618-016-0462-1>
- Múnera, S., Osorio, A.F., Velásquez, J.D. (2014). Data-based methods and algorithms for the analysis of sandbar behavior with exogenous variables. *Computers and Geosciences*, 72, 134-146.
<https://doi.org/10.1016/j.cageo.2014.07.009>
- Oja, E. (1982). Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3), 267-273.
<https://doi.org/10.1007/BF00275687>
- Olson, R.J., Sosik, H.M. (2007). Automated taxonomic classification of phytoplankton sampled with imaging-inflow cytometry. *Limnology and Oceanography: Methods*, 5, 204-216.
<https://doi.org/10.4319/lom.2007.5.204>
- O'Donncha, F. (2017). Using deep learning to forecast ocean waves. <https://phys.org/news/2017-09-deep-ocean.html> (accessed 23.12.2018)
- Quintero, E., Thessen, A.E., Arias-Caballero, P., Ayala-Orozco, B. (2014). A statistical assessment of population trends for data deficient Mexican amphibians. *PeerJ*, 2, E703.
<https://doi.org/10.7717/peerj.703>
- Simmonds, J.E., Armstrong, F., Copland, P.J. (1996). Species identification using wideband backscatter with neural network and discriminant analysis. *ICES Journal of Marine Science*, 53(2), 189-195.
<https://doi.org/10.1006/jmsc.1996.0021>
- Tanaka, A., Kishino, M., Doerffer, R., Schiller, H., Oishi, T., Kubota, T. (2004). Development of a neural network algorithm for retrieving concentrations of chlorophyll, suspended matter and yellow substance from radiance data of the ocean color and temperature scanner, *Journal of Oceanography*, 60(3), 519-530.
<https://doi.org/10.1023/B:JOCE.0000038345.99050.c0>
- Thessen, A. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, 1, e8621.
<https://doi.org/10.3897/oneeco.1.e8621>
- Tscherko D, Kandeler E, Bárdossy, A. (2007). Fuzzy classification of microbial biomass and enzyme activities in grassland soils. *Soil Biology and Biochemistry*, 39(7), 1799-1808.
<https://doi.org/10.1016/j.soilbio.2007.02.010>
- Turkson, R F., Yan, F., Ali, M.K.A., Hu, J. (2016). Artificial neural network applications in the calibration of spark-ignition engines: An overview. *Engineering Science and Technology, an International Journal*, 19(3), 1346-1359.
<https://doi.org/10.1016/j.jestch.2016.03.003>
- van Maanen, B., Coco, G., Bryan, K.R., Ruessink, B. G. (2010). The use of artificial neural networks to analyze and predict alongshore sediment transport. *Nonlinear Processes in Geophysics*, 17, 395-404.
<https://doi.org/10.5194/npg-17-395-2010>
- Wu, A., Hsieh, W.W., Tang, B. (2006). Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Networks*, 19(2), 145-154.
<https://doi.org/10.1016/j.neunet.2006.01.004>

Yi, J., Prybutok, V.R. (1996). A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution*, 92(3), 349-357.

[https://doi.org/10.1016/0269-7491\(95\)00078-X](https://doi.org/10.1016/0269-7491(95)00078-X)