On an Artificial Neural Network Approach for Predicting Photosynthetically Active Radiation in the Water Column

Martin M. Kumm[1(✉)], Lars Nolle[1,2], Frederic Stahl[2], Ahlem Jemai[3], and Oliver Zielinski[2,3]

[1] Department of Engineering Sciences, Jade University of Applied Science, Wilhelmshaven, Germany
martin.kumm@jade-hs.de

[2] Marine Perception, German Research Center for Artificial Intelligence (DFKI), Oldenburg, Germany

[3] Institute for Chemistry and Biology of the Marine Environment, Carl Von Ossietzky University of Oldenburg, Oldenburg, Germany

Abstract. About 1,600 bio-geo-chemical Argo floats (BGC-Argo), equipped with a variety of physical sensors, are currently being deployed in the ocean around the world for profiling the water characteristics up to a depth of 2,000 m. One of the parameters measured by the Argo is the radiometric measurement of downward irradiance, which is important for primary production studies. The multispectral Ocean Color Radiometer measures the downwelling irradiance at three wavelengths 380 nm, 412 nm and 490 nm plus the photosynthetically available radiation (PAR) integrated from 400 nm to 700 nm. This study proposes a method to reconstruct the PAR sensor values from readings of the remaining onboard sensors, independent of the location the BGC-Argo is being deployed. This allows for the PAR channel being replaced by a fourth band in the visible range. Stahl et al. [1] have already shown, that a machine learning approach, based on a multiple linear regression (MLR) or on a regression tree (RT), is capable of predicting the PAR values based on other parameters measured by the physical sensors of the BGC-Argo float. In this study, a nonlinear Artificial Neural Network (ANN) was used for the prediction of PAR. The ANN achieved a better coefficient of determination R^2 of 0.9968, compared with the MLR approach, which achieved an R^2 of about 0.97 for a combined dataset consisting of measurements from three different geographical locations. Therefore, it was concluded that the ANN was better suited to generalise the underlying transfer function.

Keywords: BGC-Argo float · Photosynthetically active radiation prediction · Artificial neural network · Machine learning

1 Introduction

Due to a dramatic increase in environmental challenges, such as climate change and the associated rise of the sea level, new methods for monitoring environmental parameters

M. Bramer and F. Stahl (Eds.): SGAI-AI 2022, LNAI 13652, pp. 112–123, 2022.
https://doi.org/10.1007/978-3-031-21441-7_8

are needed to gain a better understanding of the complex interactions in the environment. Anthropogenic activities are rapidly changing the ocean, contributing to pollution, deoxygenation, ocean warming and the resulting rise in sea level [2, 3]. To monitor these changes, modern operational oceanography uses numerous types of autonomous platforms [4]. One of these autonomous platforms is the Argo float [5–7]. Over 1,600 BGC-Argo floats have been deployed by June 2022.

A typical 10-day cycle mission of BGC-Argo float is shown in the left part of Fig. 1. It starts with a descent to a depth of 1,000 m and subsequently maintains this level while drifting away due to the ocean's currents. After nine and a half days, the float dives up to 2,000 m. After reaching the target level, the sensors begin to acquire data. When the float eventually resurfaces, it begins to transmit the data acquired via Iridium satellite communication into the Argo network [6]. The transmitted Argo float data is publicly and freely available via two global data assembly centers (GDAC) typically within 24 h (see Argo website https://argo.ucsd.edu).

Modern versions of BGC-Argo floats are, unlike older Argo floats with a three-sensor setup, equipped with a variety of additional physical, chemical and bio-optical sensors [8, 9]. Due to this increase in sensors, accompanied by data management and quality control processes, demand for machine learning has been rising [9, 10].

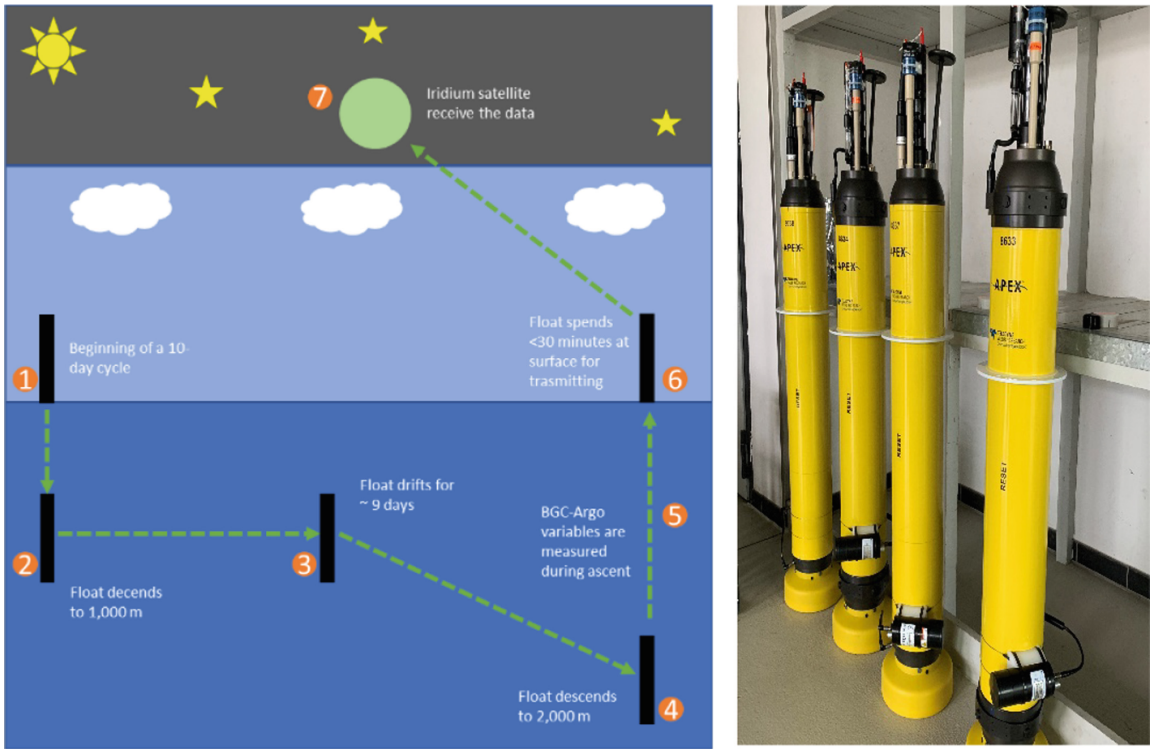

Fig. 1. Left: A typical mission of a Biogeochemical-Argo (BGC-Argo) float (adapted from [6]), Right: BGC-Argo floats ready for deployment

The BGC-Argo community suggested to re-configurate the Ocean Color Radiometer to dismiss the fourth channel, which is designed to record PAR, since this parameter could potentially be predicted from the three remaining channels, which measure the intensity of radiant energy at wavelengths at 380 nm, 412 nm and 490 nm, and the pressure. It

was shown previously that both, MLR [11] and RT [12], are capable of reconstructing the PAR sensor readings [1].

This study compares the machine learning algorithms recently utilized in [1] with a new approach, based on ANNs.

2 Related Work

Due to its influence on the botanical photosynthesis process, PAR plays a fundamental role in modeling vegetation growth [13, 14]. Several studies have already shown that PAR values can be predicted by using different meteorological and radiometric parameters. López et al. [15] and Jacovides et al. [16] developed ANN models, which use global irradiance and the solar zenith angle as inputs to estimate the global PAR. Jacovides et al. achieved for their best model an accuracy of R^2 0.979, whereas the best model trained by López et al. achieved an accuracy R^2 of about 0.999. Yu et al. [17] already showed that both, ANN models and conventional regression models, can predict PAR on the surface from incoming solar radiation and that the ANN models have a higher accuracy R^2 of about 0.999 compared with the regression models with $R^2 = 0.994$. These results show that the PAR radiant flux is strongly correlated with the broadband global radiant flux. In these studies, PAR was predicted on the surface, using the global irradiance and the solar zenith angle, whereas in this study only wavelengths in a specific narrowband together with the pressure are used to predict PAR. A review of radiometric measurements on Argo floats was recently published by Jemai et al. [18].

The BGC-Argo measured the PAR in the water column and has no information about the solar zenith. Nevertheless, Stahl et al. [1] showed the correlation between the irradiance under water at several spectral wavelengths and the PAR value.

The model by Stahl et al. [1] uses MLR and RT to estimate the PAR value on the BGC-Argo with the irradiance of three other spectral wavelengths at 380 nm, 412 nm and 490 nm.

3 Vertical Radiometric Measurement of the Water Column

One of the six essential variables measured by the BGC-Argo float is the underwater light field [6]. The Ocean Color Radiometer (OCR-504) from SATLINC Inc./Sea-Bird Scientific is used for measuring the downward irradiance at three bands 380 nm, 412 nm and 490 nm plus PAR integrated from 400 nm to 700 nm [19]. It can be seen in the right-hand side of Fig. 2 how the four sensors are arranged. These three wavelengths were selected, because they are related to the main variations in underwater optical properties [20, 21]. The information from the PAR sensor is commonly used to make predictions about the light available for primary production in natural waters [22].

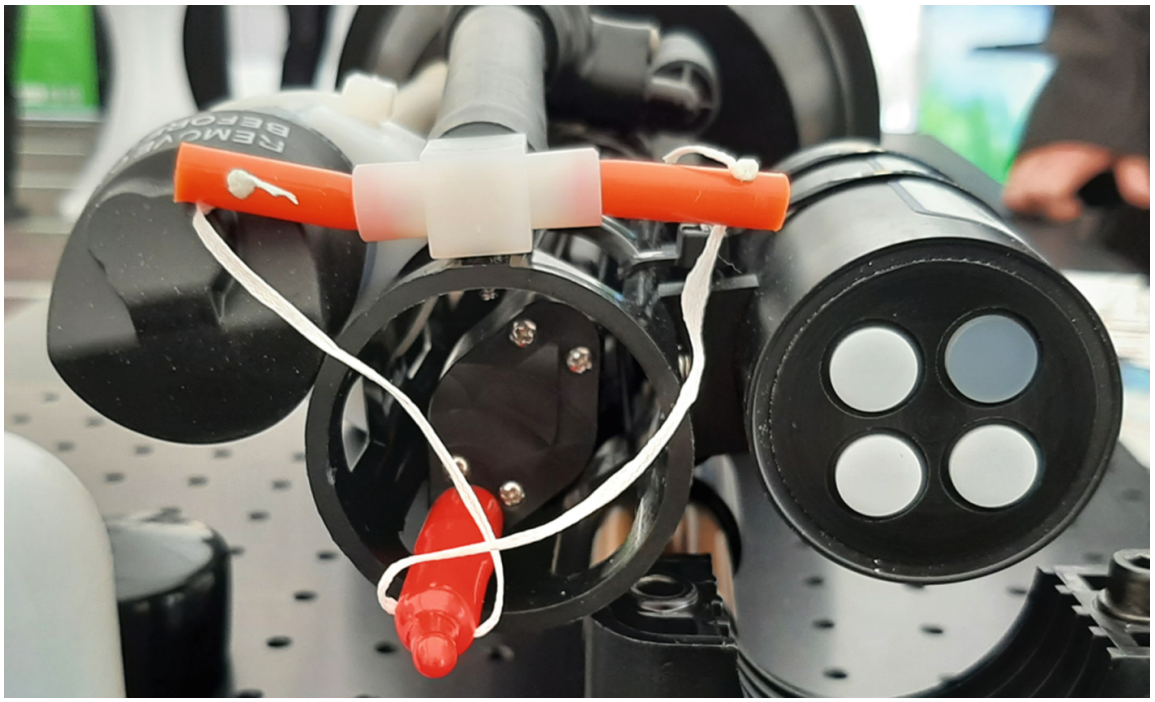

Fig. 2. OCR-504 mounted on the BGC-Argo float

In this study, the same dataset as in [1] was used to allow for a fair comparison of the methods. Float data was collected and made freely available by the International Argo Program and the national programs that contribute to it (http://www.argo.ucsd.edu and http://argo.jcommops.org). Table 1 shows the datasets from the three different sites used in this work. The sites can be identified by the World Meteorological Organization (WMO) number. The WMO number also identifies the platform type.

Table 1. Datasets

Identifier	Location	No. of instances
WMO 7900585	North Atlantic	4,403
WMO 7900562	Mediterranean Sea	13,068
WMO 7900579	Baltic Sea	1,373
WMO 7900580	Baltic Sea	1,274

4 Data Interpretation

Each instance of each dataset consists of 7 attributes. The first one indicates the cycle, the second the current number of the cycle, the third the value of the pressure in dbar, the fourth, fifth and sixth contain the different wavelengths at 380 nm, 412 nm, 490 nm and the seventh attribute represents the value of the PAR sensor.

Due to errors in the datasets, for example missing values, the associated instances were removed, leaving 1,331 instances in the WMO 7900579 dataset and 1,268 instances in the WMO 7900580 dataset. No errors were found in the other two datasets. Figure 3, 4, 5 and 6 show the correlations of each sensor with the PAR sensor. It can be seen that all sensor readings from all geographical locations, except for the pressure reading, show a good correlation with PAR. It can also be seen for readings above 100 dbar that PAR had low values. The reason for this is that light at this depth is fully absorbed by the water [1]. The fact that the pressure does not correlate with the other sensors was subsequently confirmed by the Institute for Chemistry and Biology of the Marine Environment. Therefore, Stahl et al. [1] decided to exclude the pressure from their modelling. In contrast to MLR and RT, ANNs are able to learn and interpret non-linear relations [23]. Since a non-linear relationship between PAR and the remaining parameters was expected, it was decided to use an ANN and to include the pressure values in the training set.

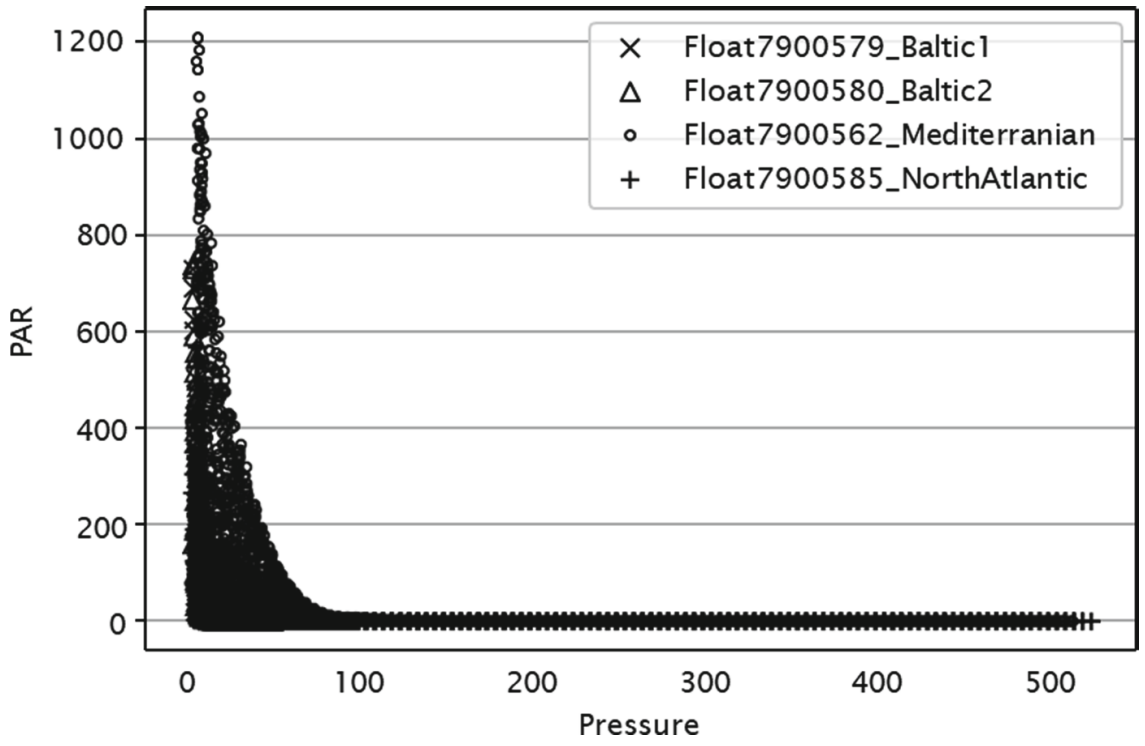

Fig. 3. Correlation between PAR and Pressure

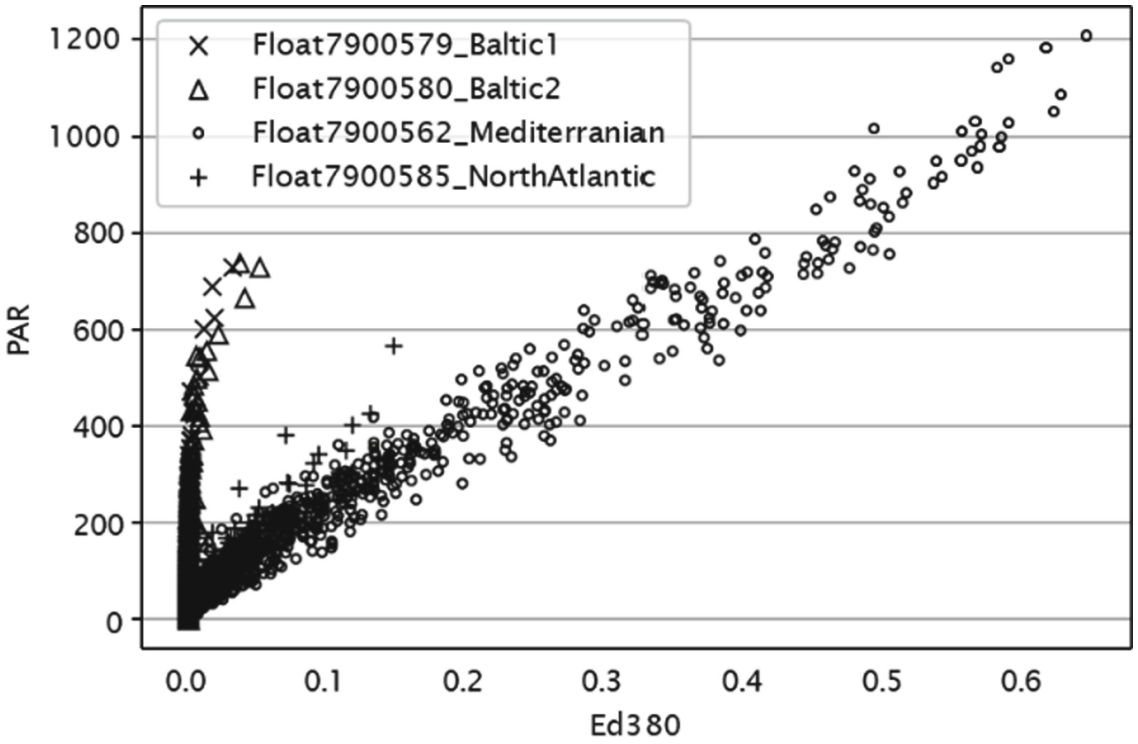

Fig. 4. Correlation between PAR and 380 nm

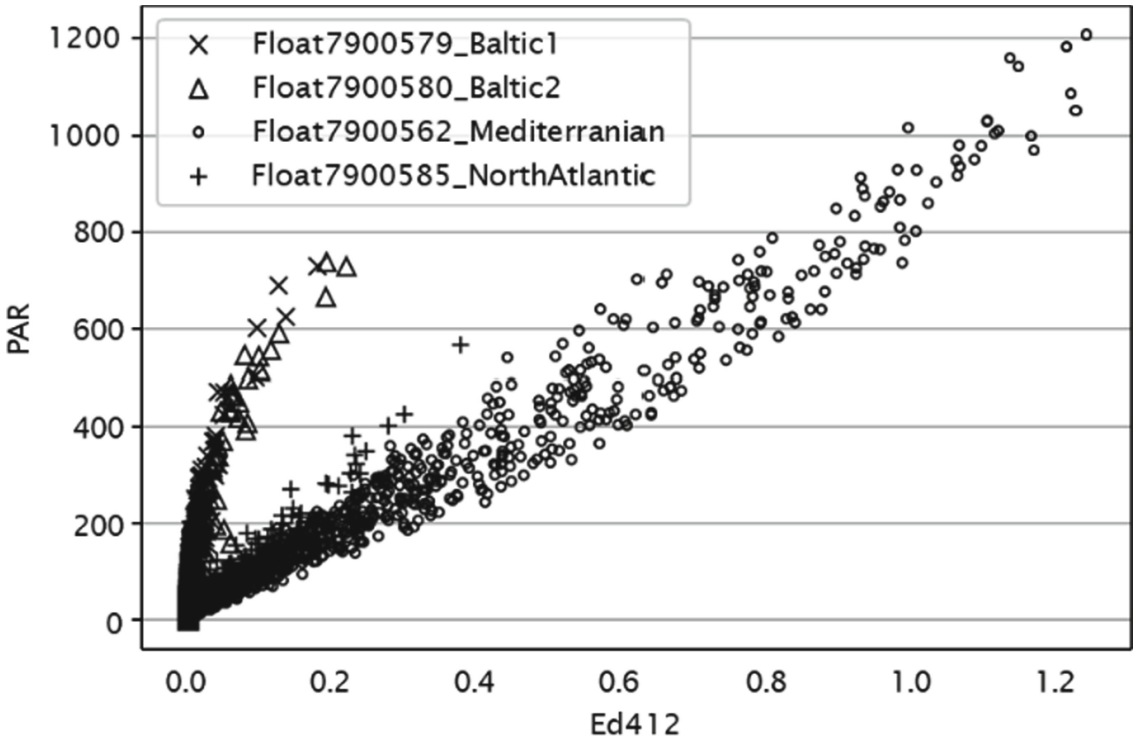

Fig. 5. Correlation between PAR and 412 nm

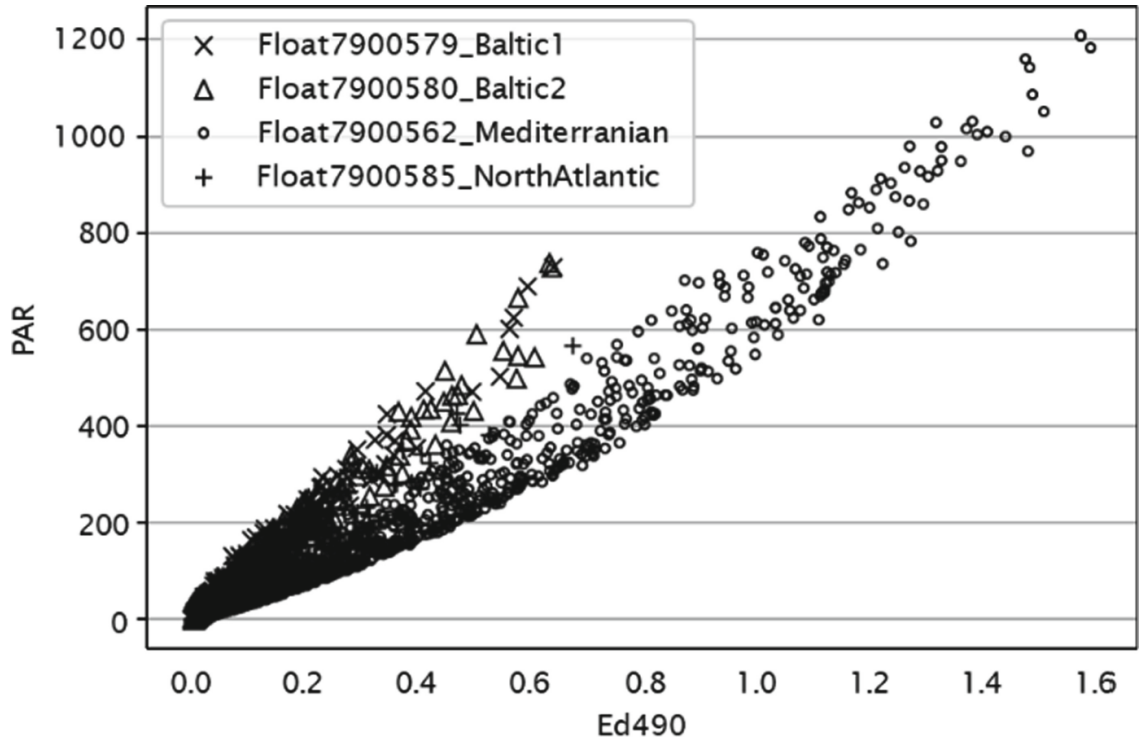

Fig. 6. Correlation between PAR and 490 nm

5 Architecture of the Developed ANN

The aim of this study was to establish whether an ANN is better suited for the generalization of the underlying relationships between PAR and the other parameters compared with MLR or RT. All error-free instances were used for modelling. The wavelength parameters at 380 nm, 412 nm and 490 nm and the pressure were used as the features for training and the PAR as the target value.

For building the model, Python was used with the Matplotlib library (https://matplotlib.org) for visualization, Pandas (https://pandas.pydata.org) and Scikit-learn (https://scikit-learn.org) for data pre-processing to detect and remove erroneous values from the dataset and the Keras implementation of TensorFlow (https://www.tensorflow.org/) for implementing the ANN.

For testing the network, 30% of the data instances were randomly selected without replacement and separated from the dataset as test data to guarantee that the ANN training is not biased towards the test data. For the training phase, 20% of the training instances were randomly selected without replacement for the validation set.

Figure 7 shows the topology of the ANN used. Since it was proven that feed-forward networks with a single hidden layer are capable of approximating any given function with any desired degree of accuracy [24], a three-layer feed forward ANN with one input layer, which can take the four input features, one hidden layer with 100 nodes and one output layer returns the predicted PAR value. The number of nodes in the hidden layer were determined empirically. The rectified linear activation function was used in the hidden-layer. For the training, Root Mean Squared propagation (RMSProp) with a learning rate of 0.01was used. The Mean Absolute Error (MAE) was used as loss

function. Each ANN was trained for 1,000 epochs, since the network tended to overfit when more epochs were used.

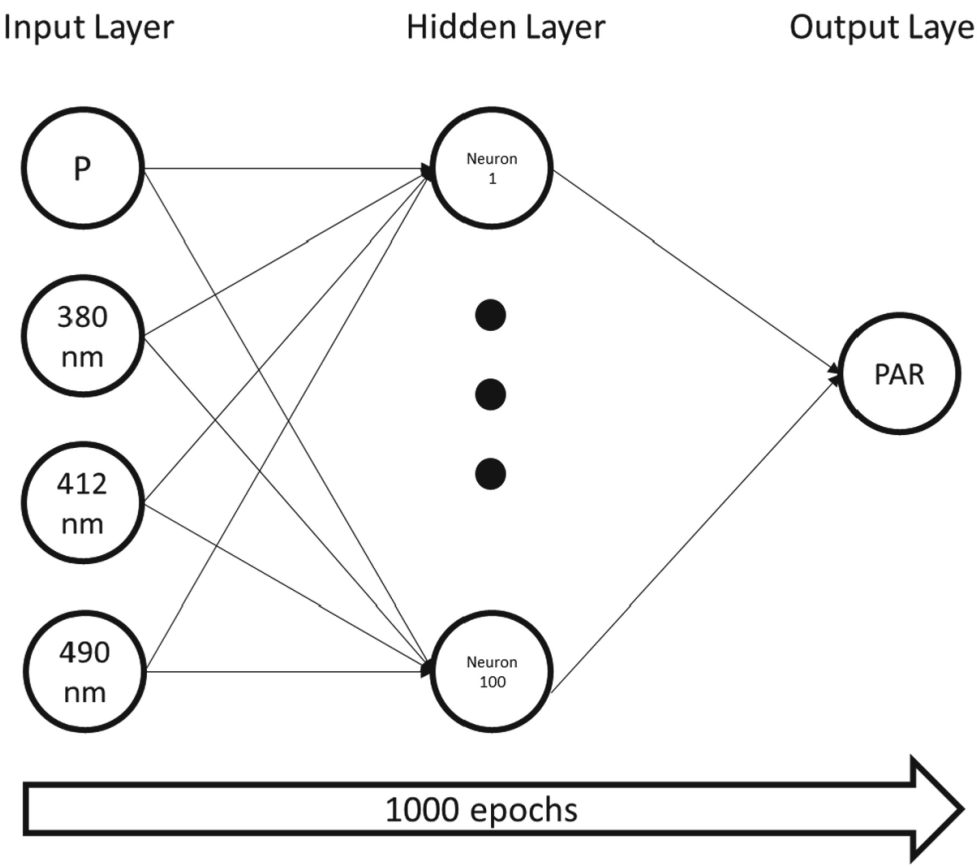

Fig. 7. Topology of the ANN

$$PAR = \sum_{j=1}^{\mathrm{m}} \left(max\left(0, \sum_{i=1}^{\mathrm{n}} x_i * w_{ij} * learning\ rate\right)\right) * w_{j\,PAR} * learning\ rate$$

$$with\, x_1 = P, x_2 = 380\,\mathrm{nm}, x_3 = 412\,nm, x_4 = 490\,nm, w = weights, learning\ rate = 0.01, n = 4, m = 100 \quad (1)$$

Eq. (1) above shows how the neural network estimates the PAR value. The two weight-matrices w_{ij} and w_{jPAR} were determined by the ANN during the training phase. n and m are equivalent to the input nodes and the hidden nodes.

6 Evaluation of the Artificial Neural Network

Figure 8 depicts the correlation between the predictions and the ground truth for the test sets for each trained ANN. The diagram in the top row on the left side shows the results of the dataset from the North Atlantic (WMO 7900585) with an R^2 value of 0.998 and on the right side the Mediterranean Sea (WMO 7900562) with an R^2 value of 0.999. The row below the diagrams shows the results from the Baltic Sea. The dataset from the

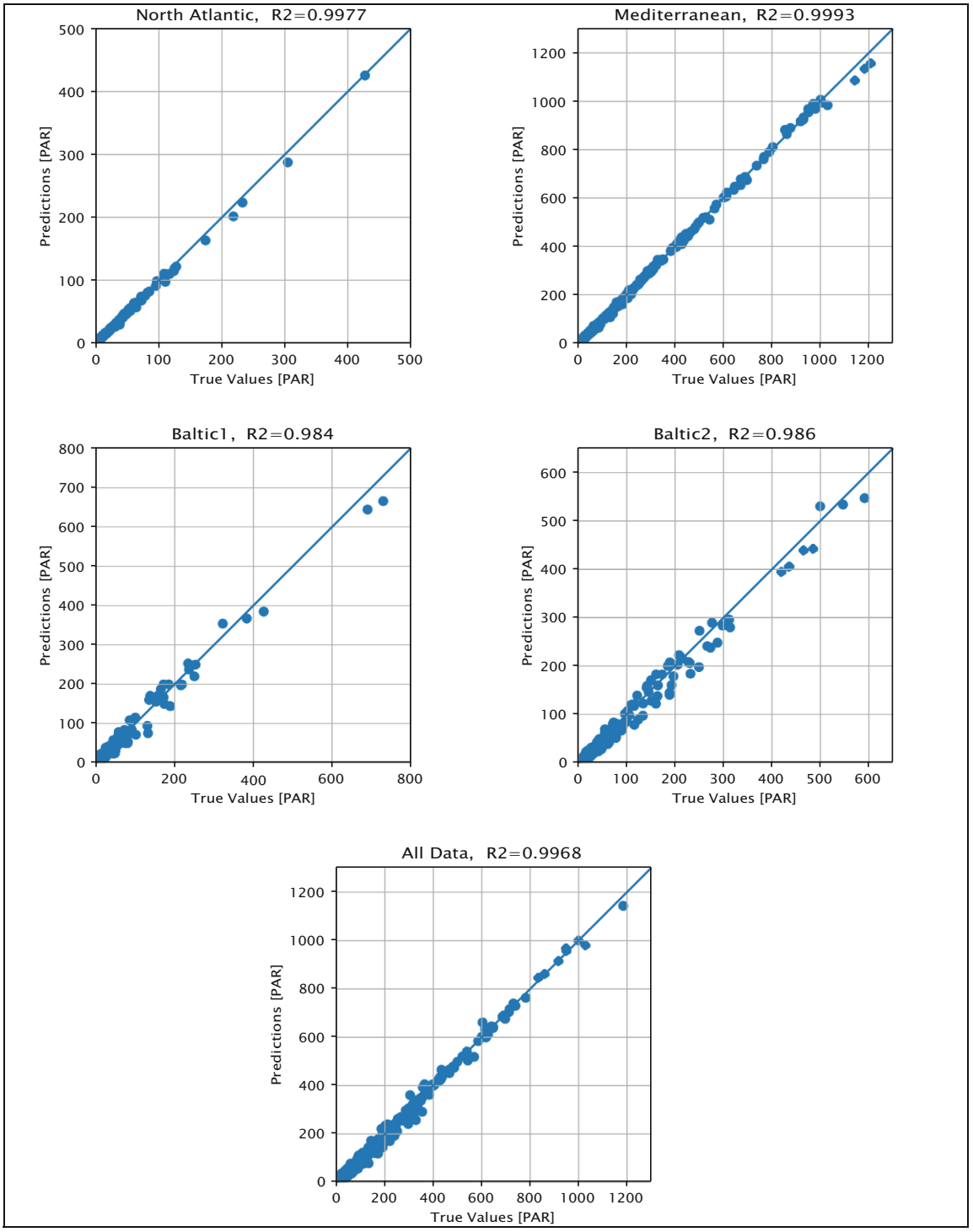

Fig. 8. Predicted vs. True PAR from each site

WMO 7900579 on the left side with a R^2 value of 0.984 and on the right side the WMO 7900580 with a R^2 value of 0.986. The diagram at the bottom shows the results for the data combined from all 4 Argo floats. Here, the R^2 value was 0.997.

When compared to MLR, the achieved R^2 value for the combined dataset was approximately 0.027 higher. This indicates that the ANN has generalised the underlying transfer function better. The reason for that might be that in this work pressure was used as an

additional input parameter for the ANN. In order to allow for a fair comparison with the results from [1], another ANN was trained without using pressure as an input. It can be seen from Table 2 that the performance degraded slightly when compared to the ANN using pressure, but it performed still better or equal than MLR and RT, except for the Baltic Sea Float 1 dataset.

Table 2. R^2 values for the different models (MLR and RT from [1])

R^2 values				
Dataset	ANN with pressure	ANN without pressure	MLR	RT
Combined	0.997	0.987	0.970	0.960
Mediterranean Sea	0.999	0.998	0.997	0.989
Baltic Sea Float 1	0.984	0.977	0.981	0.973
Baltic Sea Float 2	0.986	0.983	0.983	0.963
Atlantic Ocean	0.998	0.997	0.996	0.988

7 Conclusion

Due to the purpose of replacing the PAR sensor with a fourth band in the visible range, a method to reconstruct the PAR sensor values from readings of the remaining onboard sensors, based on ANNs was proposed. These generalization properties of the combined model developed make it possible to predict the PAR value independent of the geographical location where the data was acquired, with high accuracy. This indicates, that the environmental differences at those locations, for example salinity or turbidity, have no effects on the model. Therefore, the PAR sensor can be replaced by fourth band to measure the downward irradiance without losing the information about the photosynthetically active radiation in the water column.

The next steps in this work are to further evaluate the ANN model with additional datasets collected from different ocean basins before finally using the model operationally on the fleet of Argo floats.

Acknowledgments. This work was partly funded by the Ministry for Science and Culture, Lower Saxony, Germany, through funds from the Niedersächsische Vorab (ZN3480), and the German Federal Ministry of Education and Research, project SpektralArgo-N (Grant No. 03F0825A and 03V01478), and project DArgo2025 (Grant-No. 03F0857B).

References

1. Stahl, F., Nolle, L., Jemai, A., Zielinski, O.: A model for predicting the amount of photosynthetically available radiation from BGC-ARGO float observations in the water column. Commun. ECMS 36 (2021)

2. Lotze, H.K., et al.: Global ensemble projections reveal trophic amplification of ocean biomass declines with climate change. Proc. Natl. Acad. Sci. USA **116**, 12907–12912 (2019)
3. Wollschläger, J., Neale, P., North, R., Striebel, M., Zielinski, O.: Climate change and light in aquatic ecosystems: variability & ecological consequences. Front. Mar. Sci. **8**, 688712 (2021)
4. Roemmich, D., et al.: On the future of argo: a global, full-depth, multi-disciplinary array. Front. Mar. Sci. **6**, 439 (2019)
5. Sloyan, B.M., Roughan, M., Hill, K.: Global ocean observing system. New Front. Oper. Oceanogr. 75–89 (2018)
6. Claustre, H., Johnson, K., Takeshita, Y.: Observing the global ocean with biogeochemical-argo. Annu. Rev. Mar. Sci. **12**, 23–48 (2019)
7. Organelli, E., Leymarie, E., Zielinski, O., Uitz, J., D'Ortenzio, F., Claustre, H.: Hyperspectral radiometry on biogeochemical-argo floats: a bright perspective for phytoplankton diversity. Observing 90 (2021)
8. Johnson, K.S., et al.: Biogeochemical sensor performance in the SOCCOM profiling float array. J. Geophys. Res.: Oceans **122**(8), 6416–6436 (2017)
9. Claustre, H., et al.: Bio-optical sensors on argo floats. In: Claustre, H. (ed.) Reports and Monographs of the International Ocean-Colour Coordinating Group, pp. 1–89 (2011)
10. Jiang, Y., Gou, Y., Zhang, T., Wang, K., Chengquan, H.: A machine learning approach to argo data analysis in a thermocline. Sensors **17**(10), 2225 (2017)
11. Freedman, D.: Statistical Models: Theory and Practice. Cambridge University Press (2009)
12. Breiman, L., Friedman, J., Olshen, R.A., Stone, C.: Classification and Regression Trees.Routledge (2017)
13. Wang, L., Gong, W., Li, C., Lin, A., Hu, B., Ma, Y.: Measurement and estimation of photosynthetically active radiation from 1961 to 2011 in Central China. Appl. Energy **111**, 1010–1017 (2013)
14. Holinde, L., Zielinski, O.: Bio-optical characterization and light availability parameterization in Uummannaq Fjord and Vaigat-Disko Bay (West Greenland). Ocean Sci. **12**, 117–128 (2016)
15. López, G., Rubio, M., Martínez, M., Batlles, F.: Estimation of hourly global photosynthetically active radiation using artificial neural network models. Agric. For. Meteorol. **107**(4), 279–291 (2001)
16. Jacovides, C., Tymvios, F., Boland, J., Tsitouri, M.: Artificial Neural Network models for estimating daily solar global UV, PAR and broadband radiant fluxes in an eastern Mediterranean site. Atmos. Res. **152**, 138–145 (2015)
17. Yu, X., Guo, X.: Hourly photosynthetically active radiation estimation in Midwestern United States from artificial neural networks and conventional regressions models. Int. J. Biometeorol. **60**(8), 1247–1259 (2015)
18. Jemai, A., Wollschläger, J., Voß, D., Zielinski, O.: Radiometry on argo floats: from the multispectral state-of-the-art on the step to hyperspectral technology. Front. Marine Sci. **8**, 676537 (2021)
19. SATLANTIC: Operation manual for the OCR-504. In: SATLANTIC Operation Manual SAT-DN-00034, p. 66 (2013)
20. Xing, X., Morel, A., Claustre, H., D'Ortenzio, F., Poteau, A.: Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats: 2. Colored dissolved organic matter absorption. J. Geophys. Res. **117**, 1–16 (2012)
21. Organelli, E., et al.: A novel near-real-time quality-control procedure for radiometric profiles measured by Bio-Argo floats: protocols and performances. J. Atmos. Ocean. Technol. **33**, 937–951 (2016)
22. Mignot, A., Ferrari, R., Claustre, H.: Floats with bio-optical sensors reveal what processes trigger the North Atlantic bloom. Nature Commun. **9**, 190 (2018)

23. Hong-ze, L., Sen, G., Chun-jie, L., Jing-qi, S.: A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. Knowl.-Based Syst. **37**, 378–387 (2013)
24. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. **2**(5), 359–366 (1989)