ON THE LOCATION-INDEPENDENT RECONSTRUCTION OF PHOTOSYNTHETICALLY ACTIVE RADIATION IN THE WATER COLUMN USING NEURAL NETWORKS

Martin Maximilian Kumm[1], Christoph Tholen[2], Lars Nolle[1,2] and Frederic Stahl[2]

[1]Jade University of Applied Science, Department of Engineering Sciences
Wilhelmshaven, Germany, e-mail: {martin.kumm|lars.nolle}@jade-hs.de
[2]German Research Center for Artificial Intelligence (DFKI), Marine Perception
Oldenburg, Germany, e-mail: {christoph.tholen|frederic_theodor.stahl}@dfki.de

KEYWORDS

BGC-Argo float, photosynthetically active radiation prediction, artificial neural network, machine learning, generalisation.

ABSTRACT

Accurate reconstruction of photosynthetically active radiation (PAR) in aquatic environments is critical for understanding primary production and ecosystem dynamics. This study evaluates the generalisation abilities of artificial neural networks (ANNs) for location-independent PAR reconstruction using data from BGC-Argo floats. The proposed ANN model is trained on datasets from multiple geographic regions and validated against independent test data from diverse oceanic locations. Comparisons with multiple linear regression (MLR) and regression tree (RT) models demonstrate that the ANN consistently achieves superior predictive accuracy, with R^2 values exceeding 0.97 in most test cases. The results indicate that neural networks can effectively generalise across different marine environments, even in regions with distinct optical properties. Notably, the ANN outperforms alternative models except in one test case, highlighting the potential influence of regional environmental factors. This study underscores the potential of machine learning techniques to enhance bio-optical sensor configurations and reduce the necessity for dedicated PAR sensors.

INTRODUCTION

The dramatic increase in environmental challenges, ranging from climate change and sea level rise to pollution, deoxygenation, and ocean warming, has driven the development of innovative methods for monitoring critical environmental parameters (Lotze, et al., 2019; Wollschläger, et al., 2021). Modern operational oceanography now relies on an array of autonomous platforms, among which Argo floats have emerged as a cornerstone (Roemmich, et al., 2019). There are over 4,000 floats deployed globally and more than 1,600 specialised biogeochemical (BGC-Argo) floats in operation since 2012. To date, over 50,000 multispectral profiles have been collected using this configuration of bands, and these profiles have been evaluated in terms of data quality (Jutard, et al., 2021; Organelli, et al., 2016; Stoer, et al., 2023). These platforms continuously collect high-resolution vertical profiles from the ocean's surface to depths of approximately 2,000 meters. Data transmission via Iridium or Argos satellite systems ensure that information is publicly and freely available through two global data assembly centers (GDAC). Data is typically available within 24 hours (see Argo website https://argo.ucsd.edu).

Initially designed with a three-sensor configuration to capture fundamental physical oceanographic properties, Argo floats have evolved with the BGC-Argo initiative to include a suite of additional physical, chemical, and bio-optical sensors (Johnson, et al., 2017; Claustre, et al., 2011). A key instrument in this expanded sensor package is the Ocean Colour Radiometer (OCR), such as the OCR-504 from SATLANTIC Inc./Sea-Bird Scientific, which measures radiometric observations at four channels. Three of these channels,380 nm, 412 nm, and 490 nm, were selected in this study for their sensitivity to variations in the underwater light field, while the fourth channel is dedicated to recording Photosynthetically Active Radiation (PAR). PAR, which integrates downward irradiance between 400 nm and 700 nm, is essential for assessing the light available for primary production in natural waters (Behrenfeld & Falkowski, 1997; Morel & André, 1991).

In parallel with these technological advancements, the rapid expansion in sensor diversity has necessitated more sophisticated data management, quality control, and analysis methods, with machine learning emerging as a particularly promising tool (Claustre, et al., 2011; Jiang, et al., 2017). Recognizing the potential for streamlining sensor configurations, the BGC-Argo community has suggested reconfiguring the OCR by omitting the dedicated PAR channel. This proposal is based on the observation that PAR measurements can be reliably reconstructed from the three remaining spectral channels 380 nm, 412 nm, and 490 nm, and pressure. Previous studies have demonstrated that techniques such as Multiple Linear Regression (MLR) (Stahl, et al., 2021), Regression Trees (RT) (Stahl, et al., 2021) and Artificial Neural Networks (ANN) (Kumm, et al., 2022; Pitarch, et

al., 2025) can successfully predict PAR sensor readings. However, most recent research has shown an unsatisfactory accuracy of the models relaying on the three wavelengths 400 nm, 412 nm and 490 nm for the Freefall Profiler dataset, which covers more geolocations and is therefore more complex (Tholen, et al., 2024).
Building on these findings, the present study evaluates an ANN model with datasets from different geolocations to demonstrate the ability of generalisation of the ANN to reconstruct the PAR values and this work improves the architecture of the ANN. An MLR and an RT were also trained as a comparison to the ANN.

VARTICAL RADIOMETRIC MEASURMENT OF THE WATER COLUMN

Among the six essential variables measured by BGC-Argo floats, the underwater light field is a key parameter (Claustre, et al., 2019). To capture this, OCR-504 from SATLANTIC Inc./Sea-Bird Scientific is utilised, measuring downward irradiance at three specific wavelengths, 380 nm, 412 nm, and 490 nm, along with PAR, which is integrated across the 400–700 nm range (Satlinc, 2013). The arrangement of these four sensors is depicted on the right hand side of Figure 1. These three specific wavelengths were chosen due to their strong correlation with the primary variations in underwater optical properties (Xing, et al., 2012; Organelli, et al., 2016). The PAR sensor data is commonly used to estimate the amount of light available for primary production in aquatic environments (Mignot, et al., 2018).
To ensure a fair comparison of the methods, this study utilised a similar dataset for training as in (Kumm, et al., 2022; Stahl, et al., 2021). The float data was collected and made publicly available by the International Argo Program, along with contributions from national initiatives (http://www.argo.ucsd.edu and http://argo.jcommops.org).

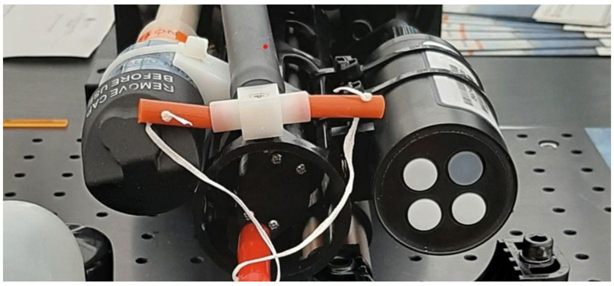

Figure 1: OCR-504 mounted on the BGC-Argo float

TRAINING AND TEST DATA

For training our model, similar BGC-Argo floats as used by Kumm, et al. (2022) were employed. However, the dataset was larger since new data has been acquired since the original publication. All samples collected above 100 dbar were removed, as no light reaches those depths. To prevent an unbalanced dataset, an equal number of samples from each dataset was used for training. Specifically, the number of samples from the smallest dataset, the North Atlantic dataset with 2,124 samples, served as the reference. For the other datasets, 2,124 samples were randomly selected, resulting in a balanced training dataset. For each set, an 80/20 train-test split was applied individually, with 10% of the training data further reserved for validation, before merging them into training, test, and validation sets.
The aim of the publication was to demonstrate the ability of generalisation of the developed ANN with respect to geolocations. Therefore, the ANN was subsequently tested with data from eight additional floats, located at different regions of the world. Table 1 shows the different datasets from the different sites used in this work.

Table 1: Training and test data

Identifier	Location	Samples
Training		
WMO 7900561	North Atlantic	2,124
WMO 7900562	Mediterranean Sea	2,124
WMO 7900579	Baltic Sea	2,124
WMO 7900580	Baltic Sea	2,124
Test		
WMO 7900585	North Atlantic	4,116
WMO 7900583	Tasmania Sea	1,843
WMO 4903711	Caspian Sea	544
WMO 5905505	Tasmanian Sea	700
WMO 5906661	Indian Ocean	1,471
WMO 6902906	South Pacific	10,042
WMO 6990638	North Atlantic	770
WMO 7902198	North Pacific	230

Figure 2 depicts the locations of the floats. It can be seen that we deliberately tested the ANN with data, which were not included in the training process.

ARCHITECTURE OF THE ANN

In this study, the architecture proposed by Kumm et al. (2022) was largely followed. The ANN is implemented as a feed-forward neural network consisting of three main components: an input layer, a hidden layer, and an output layer. The input layer receives the four sensor data inputs Pressure, Ed380, Ed412, and Ed490, while the output layer provides the value for PAR.
The ReLU function was used as the activation function to capture the non-linearities in the model. Based on Kumm et al. (2022), the Root Mean Squared Propagation algorithm was employed for training with a learning rate of 0.01, and the Mean Absolute Error (MAE) was chosen as the loss function.
To determine the optimal number of nodes in the hidden layer, ANNs with varying node counts from 1 to 50 were trained. For each node count, the ANN was trained 20 times to calculate the mean and standard deviation of performance. A batch size of 32 was used during training. Figure 3 depicts the development of the mean value and

the standard deviation from 13 to 50 nodes. Each ANN was trained for 1000 epochs. The R^2 from the ANNs with less than 13 nodes in the hidden layer were below 0.9935. This systematic variation of the node count enabled a detailed investigation of the impact of ANN topology on model performance, thereby facilitating a well-founded decision regarding the optimal ANN structure. The best R^2 of 0.9974 was reached with 48 nodes in the hidden layer. The standard deviation with this number of nodes was 0.0003.

Figure 2: Location of the datasets

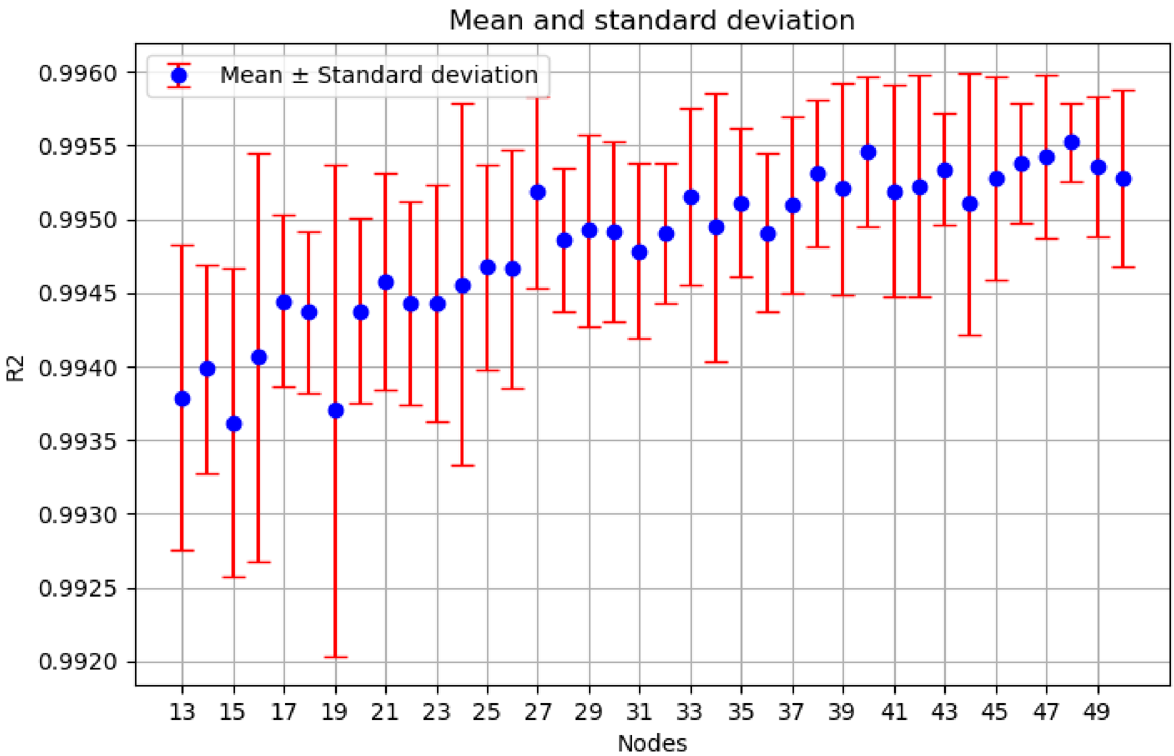

Figure 3: Mean and standard deviation of the ANN

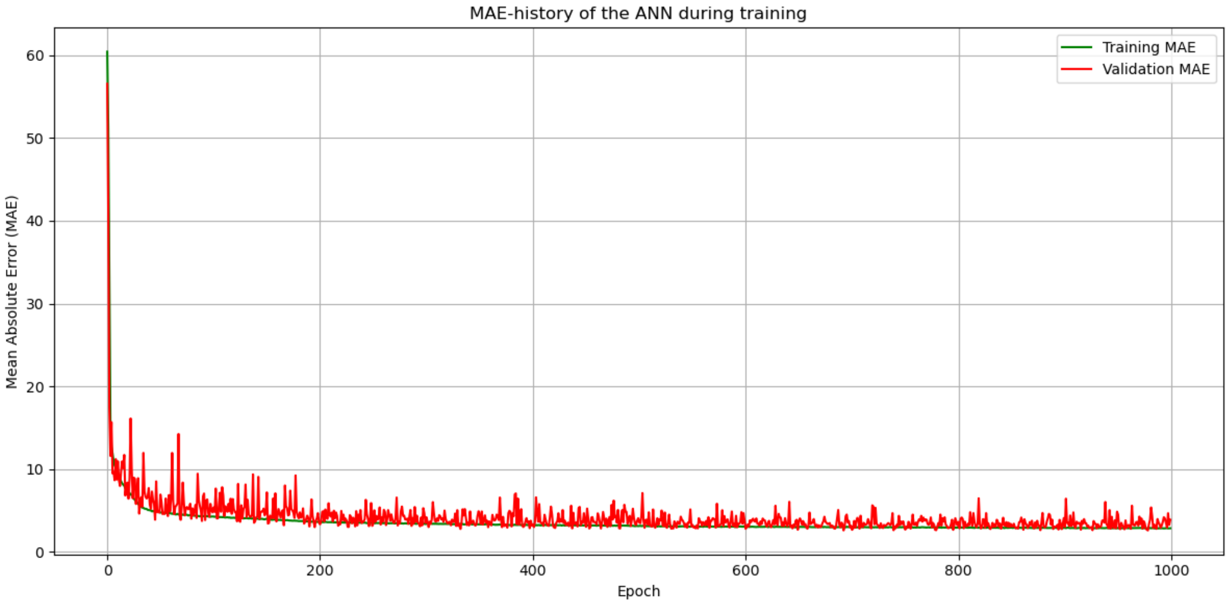

Figure 4: MAE-history of the ANN during training

EVALUATION OF THE ANN

Figure 4 illustrates the error progression of the ANN for one training run for both the training and validation datasets over time. The graph indicates that after 1000 epochs, no significant improvement in performance is observed, suggesting that the model has reached convergence.

Furthermore, the results show no signs of overfitting. The validation error remains stable even after 1000 epochs, indicating that the model maintains its ability to generalise well to unseen data. This stability suggests that the ANN effectively learns the underlying patterns in the data.

An MLR and an RT were also trained. The exact same training data as for the ANN was used for this. A comparison of the predictions for the test set can be seen in Figure 5. It can be observed, that the three methods are able to predict the PAR.

To assess the generalisation capability of the trained ANN, the MLR and the RT, eight datasets from previously unseen locations were used for testing. The results, presented in Table 2, demonstrate that the ANN exhibits strong generalisation properties compared with MLR and RT due to his consistently higher R^2 value.

The coefficient of determination (R^2) values for all datasets range between 0.97 and 0.99, indicating a high level of agreement between predicted and observed values. However, one dataset yielded a slightly lower R^2 value of 0.8955.

Table 2: R^2 from the different models

Dataset	ANN	MLR	RT
Training	0.9974	0.9872	0.9923
WMO 7900585	0.9739	0.9378	0.9130
WMO 7900583	0.9944	0.9823	0.9864
WMO 4903711	0.9910	0.8838	0.9318
WMO 5905505	0.9975	0.9921	0.9944
WMO 5906661	0.9983	0.9914	0.9823
WMO 6902906	0.9773	0.9424	0.9701
WMO 6990638	0.8955	0.9636	0.9513
WMO 7901106	0.9823	0.9548	0.8867

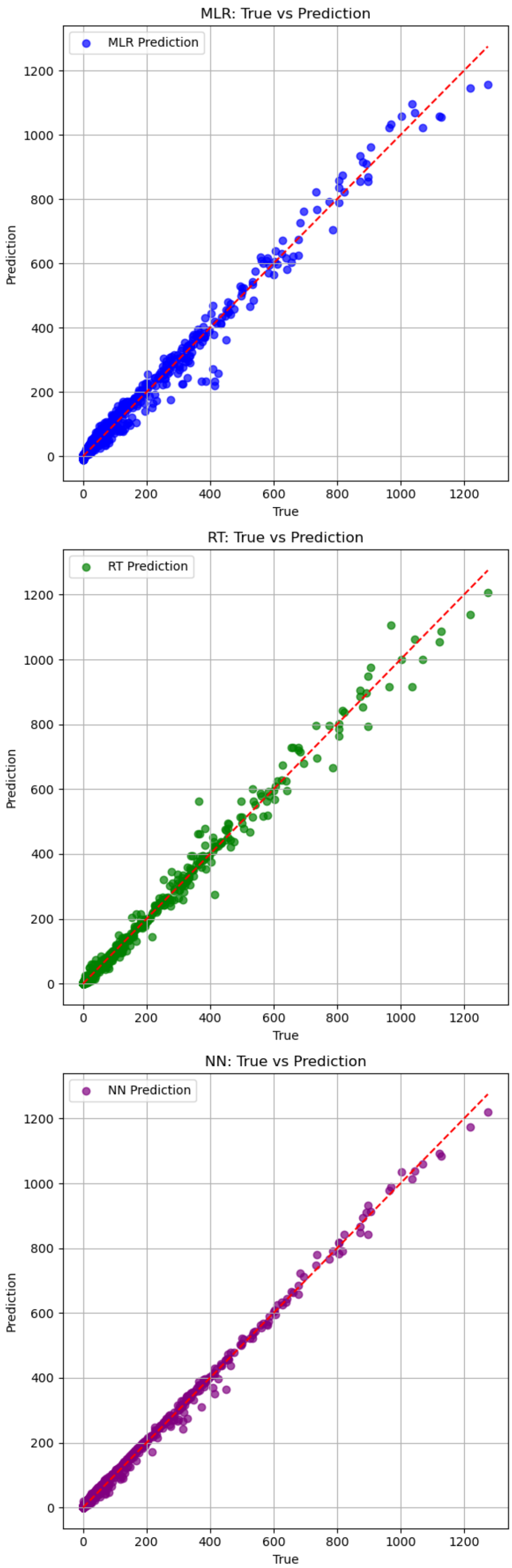

Figure 5: Prediction vs. observation for MLR, RT and ANN

DISCUSSION AND FUTURE WORK

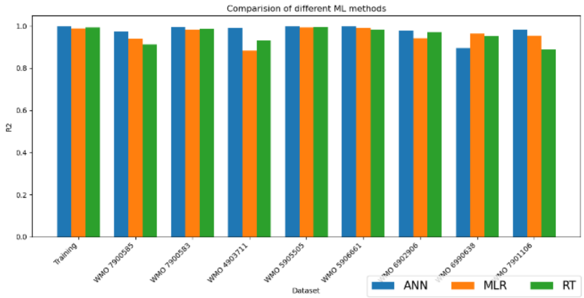

Figure 6: Comparison of the models regarding their location

As shown in Table 2 and Figure 6, the ANN outperforms the MLR and the RT on every tested site except for WMO 6990638 with 0.8955. This Argo float is located to the east of Greenland. This deviation could be attributed to the unique environmental conditions at the data collection site. In this region, highly variable light conditions, influenced by factors such as sea ice, may have introduced additional complexities that were not fully captured by the model during training. This needs to be investigated in more detail in future studies.
On the other hand, the results of the WMO 4903711 should also be emphasised. This float is located in the Caspian Sea. The Caspian Sea has a different water composition due to its geographical location. Nevertheless, the PAR values are predicted with an accuracy of 0.9910. This is a further indication of the generalisation properties of the developed ANN.

Future research should focus on further improving the predictive performance and robustness of the ANN. One potential avenue is the integration of additional spectral bands, which could enhance the model's ability to capture complex underwater optical properties. Another important aspect is the extension of the dataset to include more diverse oceanic regions, particularly areas with extreme environmental conditions, such as polar waters. This would help assess the model's adaptability to highly variable light conditions and improve its overall reliability.

ACKNOWLEDGMENT

This work was partially funded by the Ministry of Science and Culture, Lower Saxony, Germany, through funds from zukunft.niedersachsen (ZN3480).

REFERENCES

Behrenfeld, M. J. & Falkowski, P. G., 1997. A Consumer's Guide to Phytoplankton Primary Productivity Models. *Limnology and Oceanography 42.*

Claustre, H. et al., 2011. Bio-Optical Sensors on Argo Floats, In: Claustre, H. (ed.). *Reports and Monographs of the International Ocean-Colour Coordinating Group*, pp. 1-89.

Claustre, H., Johnson, K. & Takeshita, Y., 2019. Observing the Global Ocean with Biogeochemical-Argo.

Jiang, Y. et al., 2017. A machine learning approach to argo data analysis in a thermocline. *Sensors, Vol. 17, No. 10, Article 2225*.

Johnson, K. S. et al., 2017. Biogeochemical sensor performance in the SOCCOM profiling float array. *Journal of Geophysical Research: Oceans Vol. 122 Issue 8*, p. 6416–6436.

Jutard, Q., Organelli, N., Briggs, N. & al., e., 2021. Correction of Biogeochemical-Argo Radiometry for Sensor Temperature-Dependence and Drift: Protocols for a Delayed-Mode Quality Control. *Sensors 21.*

Kumm, M. M. et al., 2022. On an Artificial Neural Network Approach for Predicting Photosynthetically Active Radiation in the Water. *Artificial Intelligence XXXIX, Lecture Notes in Computer Science.*

Lotze, H. K., Tittensor, D. P., Bryndum-Buchholz, A. & al., e., 2019. Global ensemble projections reveal trophic amplification of ocean biomass declines with climate change. *Proc. Natl Acad. Sci. USA 116*, pp. 12907 - 12912.

Mignot, A., Ferrari, R. & Claustre, H., 2018. Floats with bio-optical sensors reveal what processes trigger the North Atlantic bloom. *Nature Communications*.

Morel, A. & André, J.-M., 1991. Pigment Distribution and Primary Production in the Western Mediterranean as Derived and Modeled From Coastal Zone Color Scanner Observations. *Journal of Geophysical Research: Oceans 96.*

Organelli, E. et al., 2016. A Novel Near-Real-Time Quality-Control Procedure for Radiometric Profiles Measured by Bio-Argo Floats: Protocols and Performances. *J. Atmos. Ocean. Technol. 33* , p. 937–951.

Pitarch, J., Leymarie, E., Vellucci, V. & Massi, L., 2025. Accurate estimation of photosynthetic available radiation from multispectral downwelling irradiance profiles. *Limnology and Oceanography: Methods*.

Pitarch, J. et al., 2025. Accurate estimation of photosynthetic available radiation from multispectral downwelling irradiance profiles. *Limnology and Oceanography: Methods*.

Roemmich, D., Alford, M. H., Claustre, H. & al., e., 2019. On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array. *Frontiers in Marine Science, Vol. 6, Article 439.*.

Satlinc, 2013. Operation manual for the OCR-504. In: *SATLANTIC Operation Manual SAT-DN-00034.* s.l.:s.n., p. 66.

Stahl, F., Nolle, L., Jemai, A. & Zielinski, O., 2021. A Model for Predicting the Amount of Photosynthetically Available Radiation from BGC-ARGO Float Observations in the Water Column.

Stoer, A. C., Takeshita, Y., Maurer, T. L. & al., e., 2023. A Census of Quality-Controlled Biogeochemical-Argo Float Measurements. *Frontiers in Marine Science 10.*

Tholen, C., Nolle, L., Wollschläger, J. & Stahl, F., 2024. Model Generalisation for Predicting the Amount of Photosynthetically Available Radiation in the Water Column from Freefall Profiler Observations. *ECMS 2024 Proceedings*.

Wollschläger, J. et al., 2021. Climate Change and Light in Aquatic Ecosystems: Variability & Ecological Consequences. *Frontiers in Marine Science vol. 8*.

Xing, X. et al., 2012. Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats: 2. Colored dissolved organic matter absorption. *JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 117*.

AUTHOR BIOGRAPHY

MARTIN MAXIMILIAN KUMM graduated from Jade University of Applied Sciences in Wilhelmshaven, Germany, with a Master degree in Mechanical Engineering in 2022. Since 2020 he is a research fellow at the Jade University of Applied Sciences responsible for the research aircraft. He also works in a joint project between Jade University of Applied Sciences, the German Research Center for Artificial Intelligence (DFKI) and marinom GmbH for the development of explainable artificial intelligence decision support systems for nautical officers.

CHRISTOPH THOLEN is a Senior Researcher at the German Research Center for Artificial Intelligence (DFKI), in the Marine Perception research department. His current research interests including the application of Artificial Intelligence applied to the maritime context, with a special focus on the identification and quantification of plastic litter using remote sensing. He received his doctoral degree in 2022 from the Carl von Ossietzky University of Oldenburg. From 2016 to 2022, he worked on a joint project between the Jade University of Applied Science and the Institute for Chemistry and Biology of the Marine Environment (ICBM), at the Carl von Ossietzky University of Oldenburg for the development of a low cost and intelligent environmental observatory.

LARS NOLLE graduated from the University of Applied Science and Arts in Hanover, Germany, with a degree in Computer Science and Electronics. He obtained a PgD in Software and Systems Security and an MSc in Software Engineering from the University of Oxford as well as an MSc in Computing and a PhD in Applied Computational Intelligence from The Open University. He worked in the software industry before joining The Open University as a Research Fellow. He later became a Senior Lecturer in Computing at Nottingham Trent University and is now a Professor of Applied Computer Science at Jade University of Applied Sciences. He also is affiliated with the Marine Perception research department at the German Research Center for Artificial Intelligence (DFKI). His main research interests are computational optimisation methods for real-world scientific and engineering applications.

FREDERIC STAHL received his Dipl.-Ing. (FH) degree in bioinformatics from the University of Applied Sciences, Weihenstephan, Germany, in 2006, and the Ph.D. degree in computer science from the University of Portsmouth, U.K., in 2010. From 2010 to 2012, he was a Senior Research Associate in the Department of Computer Science at the University of Portsmouth. In 2012, he worked as a Lecturer in the Department of Design Engineering and Computing at Bournemouth University, U.K. From 2012 to 2019, he served as a Lecturer and Associate Professor at the University of Reading, U.K. In 2019, he joined the German Research Centre for Artificial Intelligence (DFKI GmbH) as a Senior Researcher and became the Scientific Director of the DFKI Research Department Marine Perception in 2023. Since 2023, he has also been the Head of the AI Competence Center for Environment and Sustainability (DFKI4Planet). He has published over 100 articles in peer-reviewed conferences, journals, and book chapters and has been working in the fields of data science and artificial intelligence for nearly twenty years.