Received: 12 May 2022 | Revised: 1 September 2022 | Accepted: 13 December 2022

DOI: 10.1002/aqc.3912

RESEARCH ARTICLE

WILEY

A machine learning model and biometric transformations to facilitate European oyster monitoring

Santiago E. A. Pineda-Metz | Verena Merk | Bernadette Pogoda

Alfred Wegener Institute (AWI), Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

Correspondence
Santiago E. A. Pineda-Metz, Alfred Wegener Institute (AWI), Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany.
Email: sametz@awi.de

Funding information
German Federal Agency for Nature Conservation, Grant/Award Numbers: FKZ 3520892013, FKZ 3519892016, FKZ 3516892001

Abstract

1. Ecosystem monitoring, especially in the context of marine conservation and management requires abundance and biomass metrics, condition indices, and measures of ecosystem services of key species, all of which can be calculated using biometric transformation factors.
2. Following ecosystem restoration measures in the North Sea and north-east Atlantic waters, European oyster (*Ostrea edulis*) restoration and its monitoring have substantially increased over the past decade. Restoration activities are implemented by diverse approaches and practitioners ranging from governmental conservation agencies, research institutions and non-governmental institutions to regional groups, including citizen science projects. Thus, tools for facilitating data acquisition and estimation with non-destructive techniques can support monitoring quantitatively and qualitatively.
3. Weight-to-weight transformation factors for calculating dry weight of *O. edulis* from wet weight measurements are presented. Another important tool is the estimation of weight only from size measurements. The classical approach to achieve these transformation factors is the construction of allometric models, which, however, can greatly vary among regions and between years, making them extremely location/season specific.
4. Alternative and more flexible models constructed using random forests are proposed. This algorithm is a machine learning technique that is increasingly used in ecology, and has been proven to outperform other predictive models. From biometric variable measurements of 1,401 *O. edulis* individuals, allometric models were used to estimate total, shell and body wet weights, and compare them with 15 random forest models.
5. In general, the random forest models outperformed the allometric ones, with lower error when estimating weight. The developed random forest models can thus provide a tool for facilitating oyster restoration monitoring by increasing data acquisition without the need of sacrificing European oyster individuals. Their improvement can imply its implementation in other regions and support European oyster restoration and monitoring efforts throughout Europe.

KEYWORDS
length–weight relation, monitoring, North Sea, *Ostrea edulis*, oyster reefs, random forests, restoration, weight–weight transformation

1 | INTRODUCTION

The study of aquatic organisms' growth by means of allometric models including size and weight, has been defined as essential for conservation and management of living marine resources and for understanding how, e.g. shellfish populations, are affected by environmental variations (Sujitha, 2013; Coughlan et al., 2021). Similarly, research on population dynamics further requires the calculation of biomass, reproduction and condition indices (Riccardi & Bourget, 1998), energy production (Brey, Rumohr & Ankar, 1988), and, in the context of ecosystem services, e.g. blue carbon production and stocks (Macreadie et al., 2019). All these variables can be calculated by means of biometric transformation factors (e.g. Brey, Rumohr & Ankar, 1988; Riccardi & Bourget, 1998).

Even if data are only collected at the population level for size, weight, abundance and biomass, logistical and planning constraints greatly limit the time, and thus breadth, of the obtained data and parameters which could be calculated. In this sense, the publication and use of allometric regression models (e.g. Orton, 1935; Eklöf et al., 2017) as well as biometric transformation factors (e.g. Brey, Rumohr & Ankar, 1988; Riccardi & Bourget, 1998) are relevant tools to expand the knowledge gained from collected data. The development of allometric models and transformation factors aims to allow reuse of data, and estimation of further ecological parameters and indices when time constraints restrict sampling or laboratory analyses (Rumohr, Brey & Ankar, 1987; Brey, Rumohr & Ankar, 1988; Riccardi & Bourget, 1998; Eklöf et al., 2017).

While of great help, both allometric models and biometric transformation factors have shortcomings and knowledge gaps restricting the spatial and temporal extension to which these can be applied and used. Classical allometric weight–size models have been shown to be inaccurate for larger individuals of a given species (e.g. Galtsoff, 1931; Coughlan et al., 2021), to have substantial differences in model parameters even between closely located regions (Chatterji et al., 1985; Powell et al., 2016; Petteta et al., 2019) and at different seasons and years (Froese, 2006; Powell et al., 2016), or are only described for a restricted number of geographical regions and species (Froese, 2006; Coughlan et al., 2021). These shortcomings also apply for biometric transformation factors, which are restricted to specific regions (e.g. Rumohr, Brey & Ankar, 1987), are too general and/or only for selected taxa (e.g. Riccardi & Bourget, 1998), or were calculated from a very limited number of measurements (e.g. some transformation factors developed by Brey, 2001).

There is a lack of up-to-date and flexible allometric models, as well as biometric factors for marine regions and species, which are the focus of conservation management and restoration goals. This lack results in a considerable monitoring effort being allocated to activities, which could be facilitated and expedited. Furthermore, it limits the degree to which data collected by voluntary workers can be used by scientists. Thus, providing the means to transform size measurements to weight, and these to biomass, could provide a powerful and useful tool to facilitate and improve the monitoring effort and output. An area of application, which can greatly benefit from tools transforming size measurements to weight data, is the restoration and monitoring of the European oyster (*O. edulis*). Following its inclusion in the Oslo-Paris (OSPAR) Commission's list of endangered species and habitats (OSPAR, 2009) and its importance as a bioengineering species providing ecosystem functions and services (Pogoda, 2019), several European projects have been developed and implemented to restore and monitor local populations of *O. edulis* and its habitats in the North Sea and in north-east Atlantic waters (Pogoda et al., 2019; Pogoda et al., 2020a; zu Ermgassen et al., 2021).

While there are allometric models published for *O. edulis*, these were either developed decades ago (e.g. Orton, 1935; Andreu, 1968) under environmental regimes that have changed due to anthropogenic pressure, i.e. they are suboptimal for modern or restored populations, or were based on data from specific geographical regions (e.g. Orton, 1935; Andreu, 1968; Aydin & Biltekin, 2020). The classical approaches to construct allometric models for fish and shellfish species are based on the power law for weight–length transformations (Froese, 2006), which is expressed as $W = a\,L^b$. Where W is weight, L is length (or other size measurement), and a and b are constants that can be calculated from the linear regression of the logarithms of weight and length; $\text{Log (W)} = a + b\ \text{Log (L)}$ (Froese, 2006). While regressions are powerful and straightforward tools for quantifying the relationship between two variables, they tend to be too simplistic to represent meaningful ecological patterns for the dependent variable (De'ath & Fabricius, 2000; Cutler et al., 2007; Elith, Leathwick & Hastie, 2008). This is reflected in variations and differences of the values for constants a and b due to environmental and physiological variability (e.g. Chatterji et al., 1985; Powell et al., 2016). Due to this effect, several allometric models have to be developed, even for one specific region. A modern alternative to circumvent the shortcomings of classical statistical regressions is the use of machine learning algorithms.

One machine learning algorithm that is increasingly being used in ecology is random forest (RF; e.g. Oppel, Strobl & Huettmann, 2009; Wei et al., 2010; Wang et al., 2016). RF is a mathematical algorithm developed by Breiman (2001) that can be applied in ecology for regressions, survival analyses, detecting general multivariate structure and classification (Prasad, Iverson & Liaw, 2006; Cutler et al., 2007; Elith, Leathwick & Hastie, 2008). RFs have been suggested to outperform linear methods, especially in cases where strong interactions between variables are present (Cutler et al., 2007; Oppel, Strobl & Huettmann, 2009). Furthermore, RF has been proven to also outperform other machine learning algorithms (e.g. Wang et al., 2016). Due to its performance and accuracy, its ability to deal with and use large sets of independent variables (even when highly correlated), not being sensitive to overfitting, and its comparably easy use and application, RF has gained increasing popularity amongst terrestrial (e.g. Wang et al., 2016; Shah et al., 2019; Rather, Kumar & Khan, 2021) and marine ecologists (e.g. Oppel, Strobl & Huettmann, 2009; Miller et al., 2014; Kijewski et al., 2019).

This study is based in the development of allometric and RF models, as well as weight-to-weight transformation factors that can be applied in European oyster monitoring and restoration. Against the background of biogenic reef restoration, European oyster habitats are currently a key focus of marine landscape restoration as they provide a wide range of relevant ecosystem functions and services and are defined as hotspots of biodiversity. Over the past decade, the activities in the field have increased substantially, including large-scale restoration attempts, supported by national legislation, European obligations, as well as by the UN decade on ecosystem restoration. We address the lack of information on up-to-date and North Sea-specific transformation factors, by developing a tool, that facilitates data acquisition in a non-destructive and easy to apply way. The main objective of the study is to develop and to share tools that facilitate monitoring tasks in the context of ecological restoration, namely of European oyster populations.

2 | MATERIAL & METHODS

For the construction of allometric and RF models, and weight-to-weight transformation factors, a set of size and weight variables, the location where the oysters were measured/collected, and whether these were single or clustered individuals (also including the number of individuals per cluster) were considered. All measurements were taken between April 2020 and September 2021, within the frame of several field excursions and expeditions (see Acknowledgements) of the RESTORE project (Merk, Colsoul & Pogoda, 2020; Pogoda et al., 2020b; Pineda-Metz et al., 2023). The size variables considered were shell height (H, umbo hinge to longest edge), length (L, longest distance across the valve) and width (Wi, maximum distance between external surfaces of the umbo), these were all measured to the closest 0.01 mm using a digital caliper (Figure 1). The definitions of L, Wi and H match those of the European native oyster habitat restoration monitoring handbook (zu Ermgassen et al., 2021), but differ from previous studies published within the frame of the RESTORE Project (e.g. Merk, Colsoul & Pogoda, 2020) and other studies on allometric models (e.g. Galtsoff, 1931; Mann et al., 2009; Perez & Santelli, 2018) (Figure 1). The weight variables considered were total wet and dry weights (TWW and TDW), shell wet and dry weights (SWW and SDW), and soft tissue wet and dry weights (BWW and BDW); these were measured to the nearest 0.01 g. The dry weight data set comprises a combination of: (i) data extracted

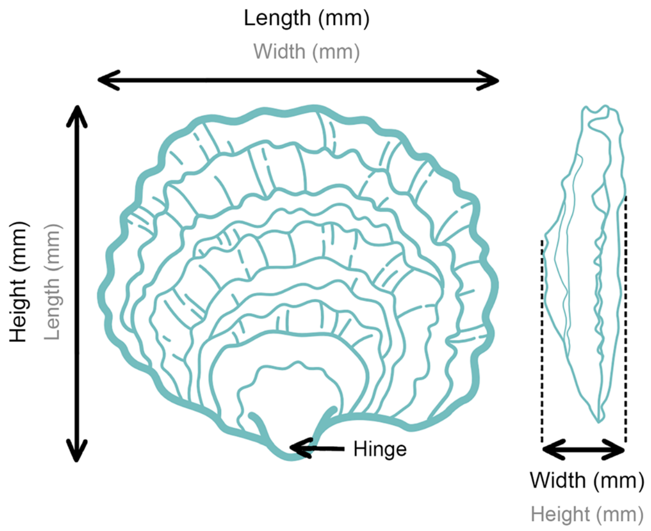

FIGURE 1 Graphical representation of the size measures, including the terminology used in this study (in bold text). The schematic also provides a comparison with the definitions provided for the same size measures in other oyster and bivalve studies (in grey; e.g. Galtsoff, 1931; Mann et al., 2009; Merk, Colsoul & Pogoda, 2020). Modified from zu Ermgassen et al. (2021).

from a previous study of *O. edulis* (Merk, Colsoul & Pogoda, 2020); and (ii) measurements taken for other studies (see Acknowledgements).

European oyster individuals included in this study were reintroduced to two regions of the German North Sea: at three sites in the proximity of the island of Helgoland (N = 738; Merk, Colsoul & Pogoda, 2020) and at the pilot oyster reef in the Natura 2000 site Borkum Reef Ground (BRG, N = 663; Pogoda et al., 2020c; Pineda-Metz et al., 2023) (Figure 2). For models relating size and weight, the region in which individuals grew is more important than the provenance of the individuals. *O. edulis* individuals considered here did not originate from but grew in the study regions from 2 mm size. As such, provenance can be considered a secondary trait and was thus excluded from this iteration of the models. Further information such as, origin of the oysters, oyster growth and condition index, sampling period, experimental set-up, measurements taken, and environmental setting in the area is provided in Merk, Colsoul & Pogoda (2020) and Pineda-Metz et al., 2023. *O. edulis* individuals were deployed in BRG in 2020 within oyster cages attached to landers located at ~30 m water depth. Size and weight of these oysters was measured in September 2021 during the annual monitoring of the pilot reef located in the area (Pineda-Metz et al., 2023). Measurements of TDW, SDW and BDW of individuals sampled in BRG were provided by colleagues from the Alfred Wegener Institute (see Acknowledgements). In total 1,401 individuals were measured, from which 1,241 measurements included TWW. For SWW, 260 measurements were used to calculate constants and construct the RF models, whereas 130 were used for BWW. The calculation of weight-to-weight transformation factors was done considering 130 measurements in which TWW and TDW, SWW and

FIGURE 2 Study area in the North Sea. Red points represent restoration locations from which oysters were collected and measured: Nature Conservation Area Helgoländer Felssockel, Natura 2000 site Borkum Reef Ground (black polygons), and offshore wind farm Meerwind Süd I Ost.

SDW, and BWW and BDW were recorded from the same animals. These factors represent the dry weight to wet weight ratio.

Based on the allometric function $W = a\ H^b$ (where W = wet weight in g; and H = height in mm), three allometric models for estimating TWW, three for estimating SWW, and another three for estimating BWW were developed. In all cases, one model was a general model considering data from both BRG and Helgoland, and two were location-specific models (i.e. one per location). Location-specific models were developed based on the fact that constants of the allometric model can vary based on location. For estimating TWW, n = 1,241 measurements were considered, whereas for estimating SWW and BWW, n = 260 and 130 measurements, respectively, were considered. To test for differences between general and location-specific models, the terms of the linear form of the allometric function were compared by means of an analysis of variance (ANOVA).

Fifteen RF models each were developed for estimating TWW (n = 1,241), SWW (n = 260), and BWW (n = 130). These models include a combination of the variables H, L, Wi, location and 'type' (representing single and clustered oysters; Supporting Information 1). Size measurements were given as absolute values in mm, whereas location (two levels: BRG and Helgoland) and type (three levels: single, clusters of two oyster and clusters three oysters) were categorical variables. Since location was included as an explanatory variable, we did not differentiate into data sets as it was done for the allometric models. For simplicity reasons, these models were named RF-A to RF-O, based on the combination of variables (Table 1). A detailed list of allometric and RF models including variable combinations is provided in Table 1.

For the allometric models, variance explained was represented with the R^2 coefficient of the linearized form of the function, $\text{Log}(W) = a + b\ \text{Log}(H)$. The significance of the regression and values for the constants *a* and *b* were also calculated from the linearized form of the allometric model (Packard, 2014). For the RF models, out-of-bag data (OOB; i.e. 1/3 of the data provided) were used to calculate the error and variance explained by the models. Variance explained is also given as R^2 and was calculated as $R^2 = 1 - MSE_{OOB}/Var(y)$, where MSE_{OOB} is the mean square error between observations and OOB predictions, and Var (y) is the variance of the observed values (Liaw & Wiener, 2002). While R^2 is an inadequate measure to fit a non-linear regression, it is simpler and more intuitive (Eklöf et al., 2017; Coughlan et al., 2021). Since this study aims to compare allometric and RF models, the variable contribution plots for all RF models are provided in Supporting Information 2.

To compare the precision of the allometric and RF models, plots of observed vs. predicted values were examined, and the mean absolute error (MAE) of each model was calculated based on the estimated values calculated from all organisms used for constructing each model. The MAE is a criterion which shows how close the estimated values are to the observed ones (Willmontt & Matsuura, 2005). MAE was chosen rather than the root mean square error, since it provides a more robust measure of error (Willmontt & Matsuura, 2005).

All calculations, figures and models were developed using the packages *vegan* (Oksanen et al., 2020), *randomForest* (Liaw & Wiener, 2002) and *ggplot2* (Wickham, 2016) for RStudio (R Core Team, 2020).

3 | RESULTS

Based on the 130 measurements containing TWW and TDW, SWW and SDW, and BWW and BDW, the following weight-to-weight transformation factors were calculated: The mean transformation ratio for TWW to TDW was 0.662 (standard error (SE) = 0.009), for SWW to SDW it was 0.718 (SE = 0.006), and for BWW to BDW it was 0.176 (SE = 0.011).

Apart from the allometric model for BWW of *O. edulis* individuals from BRG, all models were significant, with height explaining at least 77.5% of the variance (Table 1). The best fit was found for models used to estimate TWW, with little difference between general and location-specific models, same for models to estimate SWW (Figure 3; Table 2). This was supported by the comparison between constants of

TABLE 1 Description of allometric and random forest models developed for this study and their corresponding variance explained (R^2). While height (H), width (Wi), and length (L) are numerical variables, type (3 levels: Single, Cluster of two individuals, Cluster of three individuals) and location (2 levels: Helgoland, Borkum) are categorical variables. The n was different for models to estimate total wet weight (n = 1,241), shell wet weight (n = 260), and soft tissue wet weight (n = 130).

		Independent variable(s) used					Variance explained (R^2)		
Model Type	Name	H (mm)	Wi (mm)	L (mm)	Type	Location	Total wet weight (TWW; in g)	Shell wet weight (SWW; in g)	Soft tissue wet weight (BWW; in g)
Allometric	General	X					0.909	0.910	0.800
	Helgoland	X					0.902	0.840	0.806
	BRG	X					0.915	0.803	0.021
Random Forest (RF)	RF-A	X					0.715	0.866	0.662
	RF-B	X	X				0.818	0.924	0.783
	RF-C	X	X	X			0.840	0.932	0.810
	RF-D			X			0.668	0.744	0.613
	RF-E		X	X			0.753	0.863	0.756
	RF-F		X				0.550	0.780	0.591
	RF-G	X			X		0.728	0.607	0.540
	RF-H	X	X		X		0.801	0.755	0.664
	RF-I	X	X	X	X		0.830	0.824	0.716
	RF-J	X				X	0.705	0.848	0.696
	RF-K	X	X			X	0.800	0.907	0.765
	RF-L	X	X	X		X	0.830	0.919	0.796
	RF-M	X			X	X	0.638	0.671	0.524
	RF-N	X	X		X	X	0.749	0.772	0.633
	RF-O	X	X	X	X	X	0.806	0.813	0.705

Abbreviation: BRG, Borkum Reef Ground.

the linear form of the allometric function (i.e. *a* and *b*), which showed no significant differences between the general and the location-specific models (all *P*-values >0.05; Table 2). Only the models estimating BWW showed significant differences between the general and the location-specific models (all *P*-values <0.001; Table 2); i.e. soft tissue growth differs significantly between oysters growing off Helgoland and those growing in BRG.

The allometric models showed that the European oysters from BRG and Helgoland have a tendency to show isometric growth (Table 2), i.e. individuals grow at the same rate in terms of both TWW and height. This is represented by values of the constant *b* being close or equal to 3. This also applies to shell growth based on the general allometric model, but not for soft tissue growth. In contrast, the location-specific models for shell and soft tissue wet weights showed growth to be faster in terms of height than weight (i.e. they get bigger faster than they get heavier; Table 2). When observing the plots of estimated vs. observed values, it is evident that the allometric models lose prediction precision for larger organisms, regardless of the weight estimated from L (Figure 3).

The 15 RF models developed explained between 55.0 and 83.0% of the variance (Table 1) when predicting TWW, SWW and BWW. Due to the nature of the algorithms, no *P*-values can be provided. For all models, in which H was included, this was the most important variable of the models, followed by other size measurements when available (Supporting Information 2). For the model consisting of L and Wi, the former was the most important variable (Supporting Information 2). Both categorical variables (location and type) contributed the least to the models (Supporting Information 2) and, when included, resulted in lower prediction precision (Figures 4, 5 & 6) and larger MAE (Table 3). This could point to both categorical variables, location and type, to act as (statistical) noise variables, increasing predictive error in the RF models. Although the variance explained (R^2) by the RF models was lower than that of the allometric models (Tables 1 & 2), the MAE of the RF models including one or more size measurements were lower than that of the allometric models (Table 3).

Based on the plots of estimated vs. observed TWW, SWW and BWW values, differences in the precision of prediction can be observed. This includes differences between allometric and RF models (Figures 3, 4, 5 & 6), as well as between different RF models (Figures 4, 5 & 6). The allometric models for estimating TWW showed a rather large spread compared to RF models A–F, G/H, I–L and O (Figures 3 & 4). This is also reflected by the larger MAE that the general allometric model has (MAE = 5.78 g; Table 3) compared to that of most RF models including only size measures (MAE ranges from 2.13 to 4.28 g; Table 3). The same can be observed when

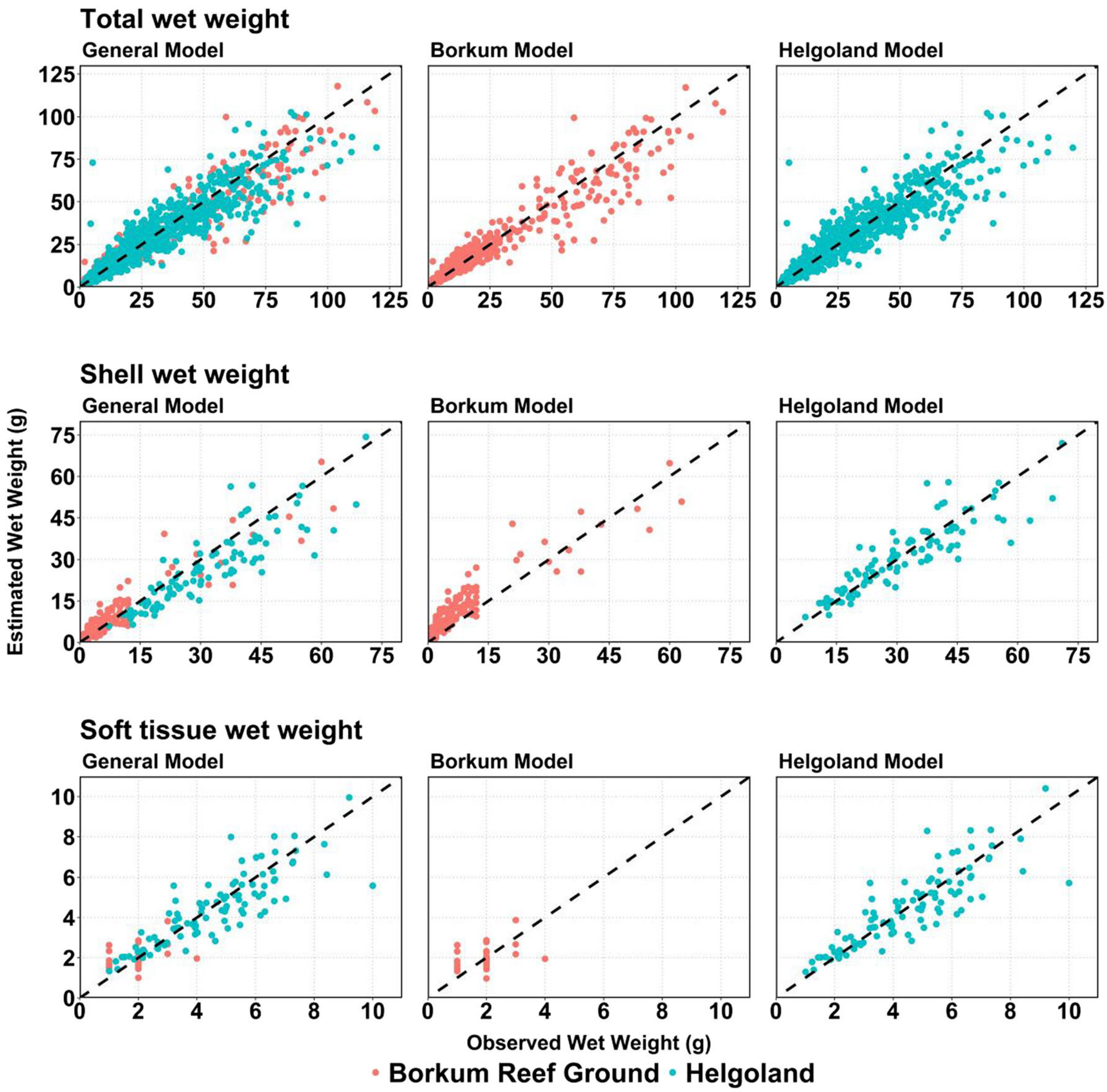

FIGURE 3 Estimated vs. observed total, shell, and soft tissue wet weights (n = 1,241; 260; 130, respectively) for the general and location-specific allometric models. Estimations were calculated based on the general allometric model W = a H^b.

TABLE 2 Constants, variance explained (R^2) and significance values for general and location-specific allometric models to estimate total, shell, and soft tissue wet weights. All R^2 were significant at P < 0.001.

	Estimated variable	*Constant* a	*Constant* b	R^2
Helgoland	Total wet weight	9.69×10^{-5}	3.00	0.902
	Shell wet weight	8.31×10^{-4}	2.46	0.840
	Soft tissue wet weight*	1.05×10^{-4}	2.49	0.806
BRG	Total wet weight	6.53×10^{-5}	3.08	0.915
	Shell wet weight	2.60×10^{-4}	2.61	0.803
	Soft tissue wet weight*	1.05×10^{-1}	0.73	0.021
Both	Total wet weight	8.09×10^{-5}	3.04	0.908
	Shell wet weight	6.08×10^{-5}	3.03	0.884
	Soft tissue wet weight*	1.46×10^{-4}	2.41	0.775

Abbreviations: BRG, Borkum Reef Ground; a, intercept; b, allometric constant.
*Significantly different between each other at P < 0.001.

comparing the location-specific allometric models to most size-measures-RF models to predict location specific values (Table 3). The RF model including only size measures that performed worse than the allometric models was RF-F, which included only Wi in mm. Similar observations of spread and differences in MAE can be observed when comparing the allometric models to estimate SWW or BWW to RF

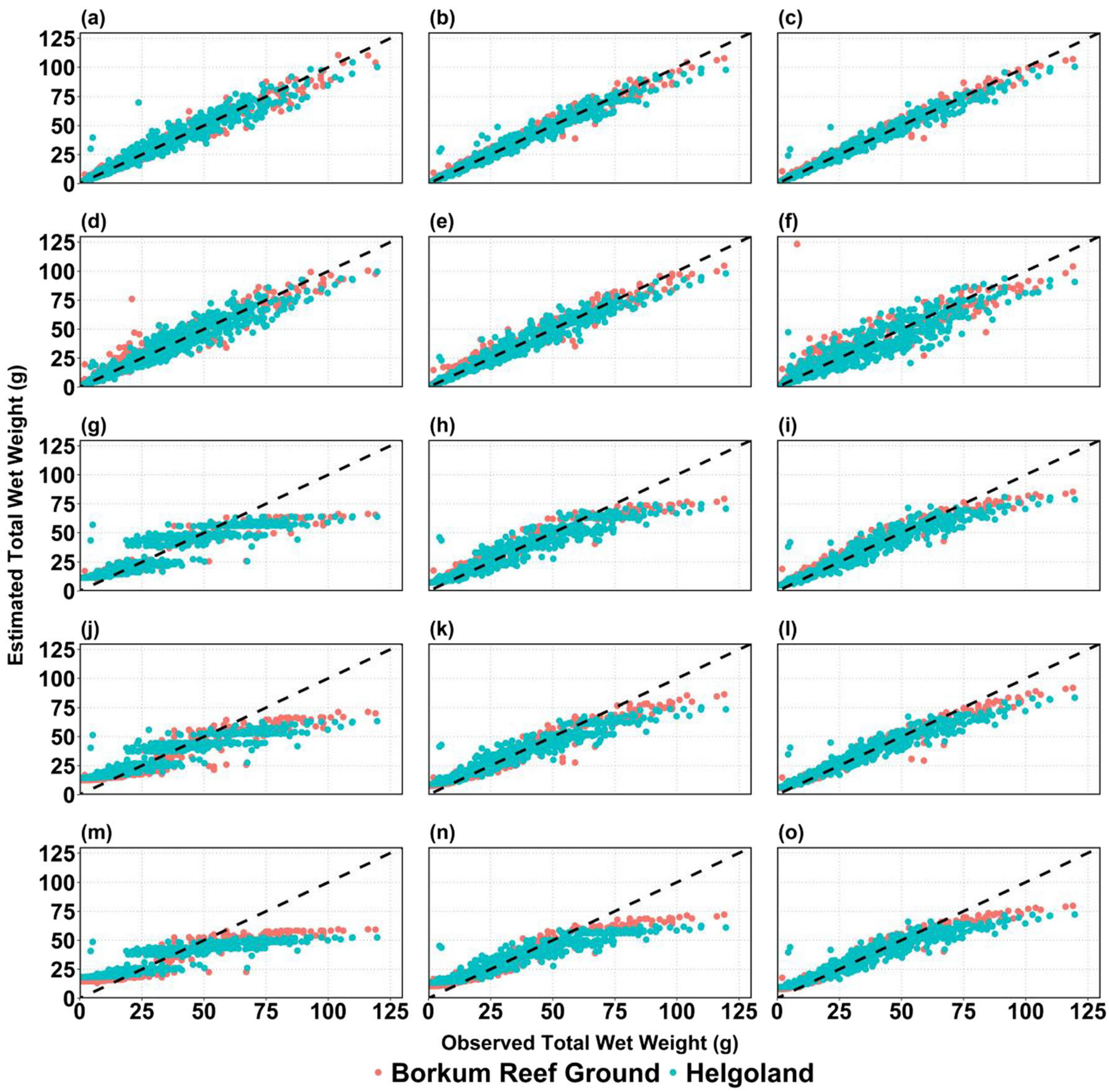

FIGURE 4 Estimated vs. observed total wet weight (TWW) based on 15 random forest models (n = 1,241). The models predict TWW based on: (a) oyster height (in mm); (b) oyster height (in mm) and width (in mm); (c) oyster height (in mm), width (in mm) and length (in mm); (d) oyster length (in mm); (e) oyster length (in mm) and width (in mm); (f) oyster width (in mm); (g) oyster height (in mm) and type; (h) oyster height (in mm), width (in mm) and type; (i) oyster height (in mm), width (in mm), length (in mm) and type; (j) oyster height (in mm) and location; (k) oyster height (in mm), width (in mm) and location; (l) oyster height (in mm), width (in mm), length (in mm) and location; (m) oyster height (in mm), type and location; (n) oyster height (in mm), width (in mm), type and location, and; (o) oyster height (in mm), width (in mm), length (in mm), type and location.

models A-F, K-L (Table 3). While some RF models including either type or location (e.g. RF-H, RF-I, RF-K, RF-L and RF-O) outperformed the allometric models (Table 3) for estimating TWW, but only RF-K and RF-L outperformed the allometric models for the estimation of SWW and BWW (Table 3). Other RF models including one or two categorical variables were outperformed by the allometric models, regardless of the estimated wet weight (Table 3).

In terms of precision (measured by the MAE), the allometric models predicting TWW performed poorly in comparison with its equivalent RF model, which included only H (RF-A in Table 3) as a predictive variable, with a MAE of at least 1.49 g larger than that of the RF model. Random forest models including only size measures (i.e. H, L and Wi; RF-A to RF-F in Table 3) were the most precise (i.e. had the lowest MAE), outperforming the simple one-variable RF model and all allometric models (Table 3). The loss of precision RF models showed that when at least one categorical variable (location and type) was present it was improved by the inclusion of further size measures (Table 3). For example, the MAE of the RF model including location and H (RF-J) was larger compared to the MAE of the RF model including location and at least one extra size measure (RF-K and RF-L; Table 3). This precision increment by adding further size measures was also observed for the models that included both categorical variables (RF-M to RF-O in Table 3). Despite the higher precision of RF

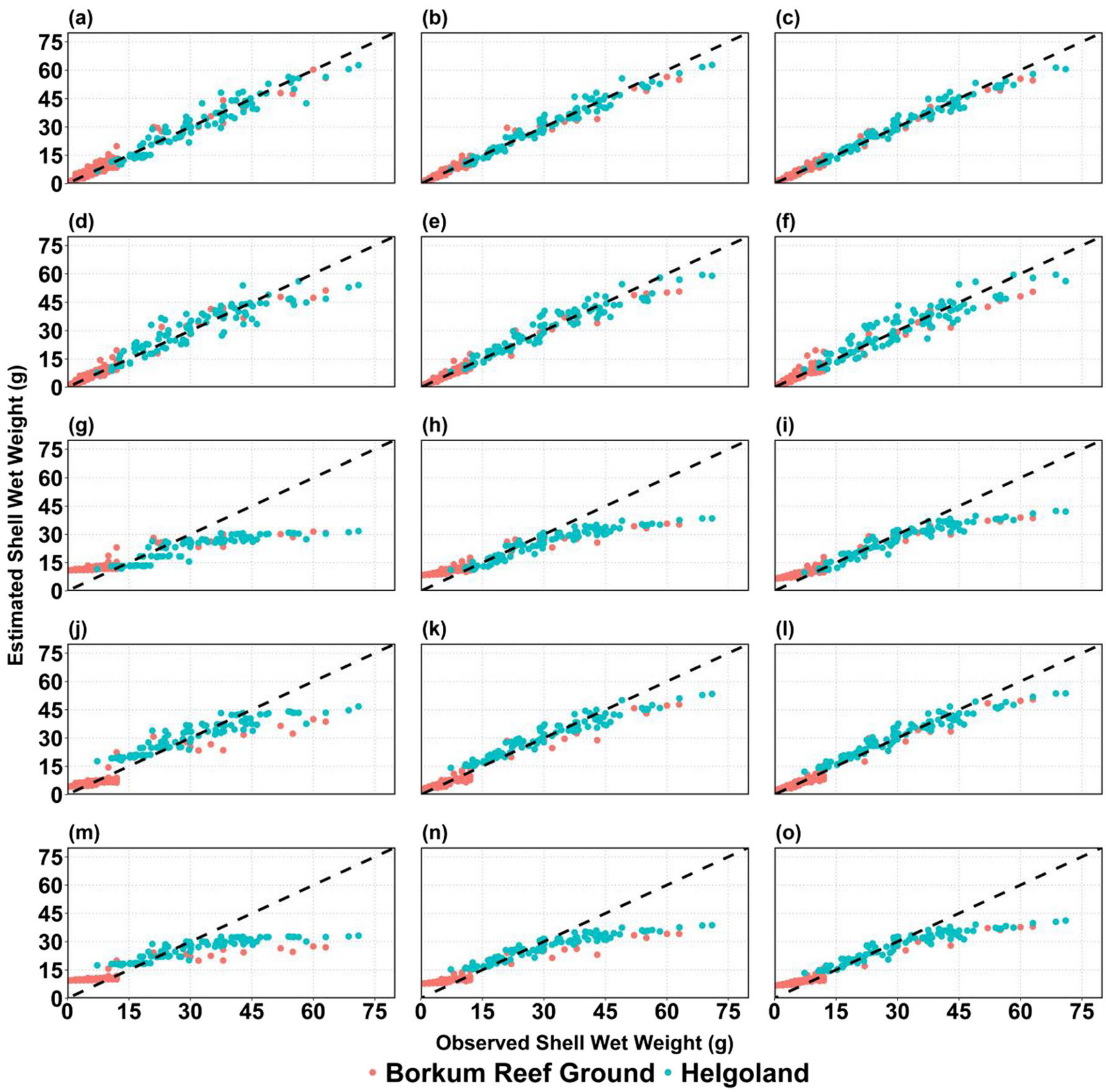

FIGURE 5 Estimated vs. observed shell wet weight (SWW) based on 15 random forest models (n = 260). The models predict SWW based on: (a) oyster height (in mm); (b) oyster height (in mm) and width (in mm); (c) oyster height (in mm), width (in mm) and length (in mm); (d) oyster length (in mm); (e) oyster length (in mm) and width (in mm); (f) oyster width (in mm); (g) oyster height (in mm) and type; (h) oyster height (in mm), width (in mm) and type; (i) oyster height (in mm), width (in mm), length (in mm) and type; (j) oyster height (in mm) and location; (k) oyster height (in mm), width (in mm) and location; (l) oyster height (in mm), width (in mm), length (in mm) and location; (m) oyster height (in mm), type and location; (n) oyster height (in mm), width (in mm), type and location, and; (o) oyster height (in mm), width (in mm), length (in mm), type and location.

models including only size measures, these also tend to underestimate the weight of larger organisms (Figures 4, 5 & 6).

4 | DISCUSSION

The results of this study represent, to the best of our knowledge, the first machine learning models to estimate total, shell and soft tissue wet weights for the European oyster *O. edulis*. In addition to the RF models developed, this study provides updated weight-to-weight transformation factors to transform total, shell and soft tissue wet weights to dry weights. This represents an improvement of the database provided by Brey (2001), which included transformation factors for *O. edulis* based on only three measurements. In general, this study presents statistically validated tools, which can facilitate the monitoring of *O. edulis* populations, especially against the background of conservation management and European restoration projects (Pogoda et al., 2020a; zu Ermgassen et al., 2021).

Over half of the RF models developed outperformed the classic allometric model used for estimating weights of fish and shellfish based on H, L or Wi. The RF models including H, L, and Wi outperformed all other models. These results show the advantage of RF over classical regressions (Cutler et al., 2007; Oppel, Strobl & Huettmann, 2009). However, RF models including the categorical

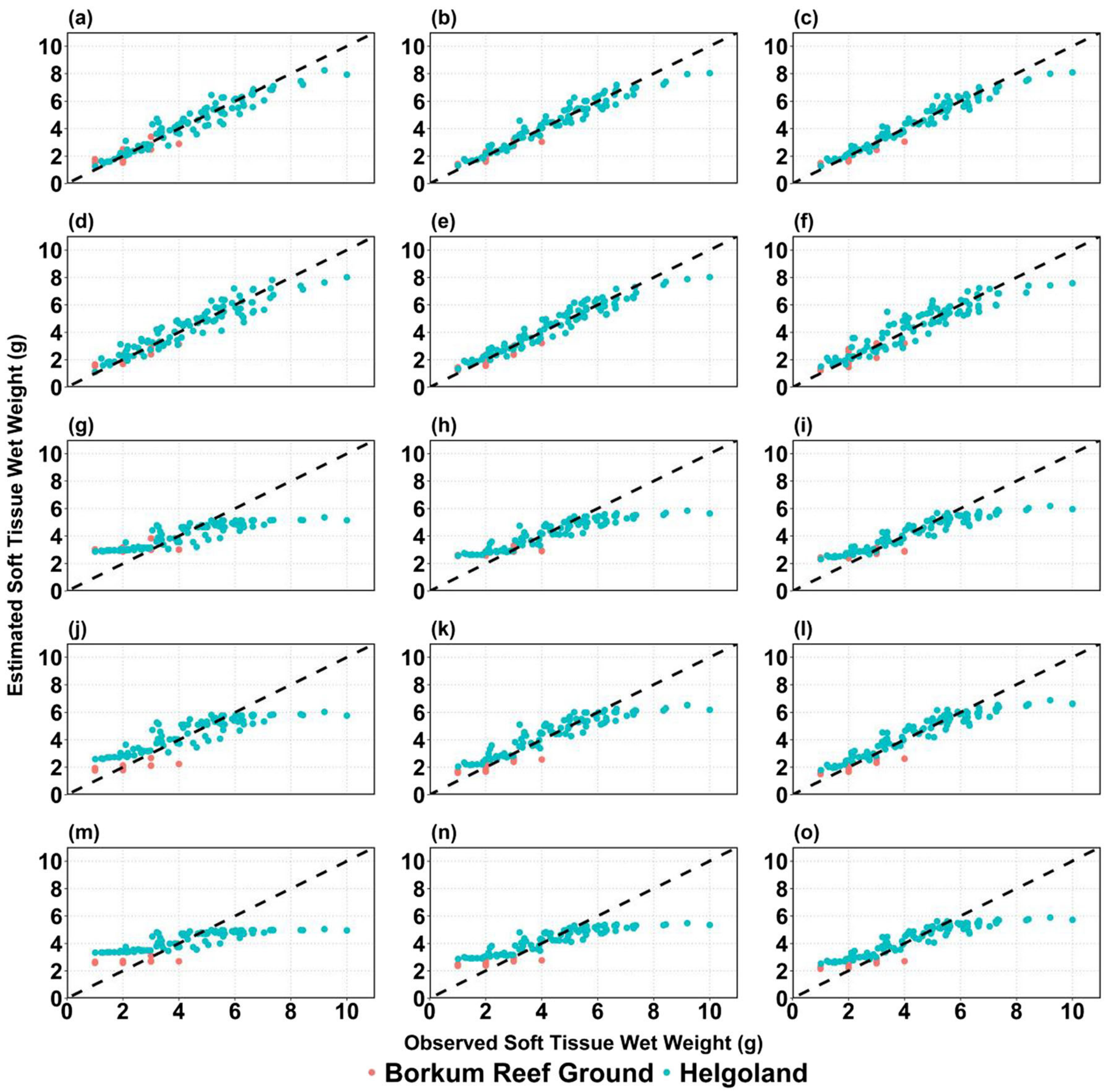

FIGURE 6 Estimated vs. observed soft tissue wet weight (BWW) based on 15 random forest models (n = 130). The models predict BWW based on: (a) oyster height (in mm); (b) oyster height (in mm) and width (in mm); (c) oyster height (in mm), width (in mm) and length (in mm); (d) oyster length (in mm); (e) oyster length (in mm) and width (in mm); (f) oyster width (in mm); (g) oyster height (in mm) and type; (h) oyster height (in mm), width (in mm) and type; (i) oyster height (in mm), width (in mm), length (in mm) and type; (j) oyster height (in mm) and location; (k) oyster height (in mm), width (in mm) and location; (l) oyster height (in mm), width (in mm), length (in mm) and location; (m) oyster height (in mm), type and location; (n) oyster height (in mm), width (in mm), type and location, and; (o) oyster height (in mm), width (in mm), length (in mm), type and location.

variables type and location suffered a loss of estimation precision. Based on the variable importance for each model, both categorical variables can be assumed to act as noisy variables. This agrees with finding no difference in the allometric models between regions, the allometric models for estimating BWW being the only exception. We assume that this exception was due to the non-significant location-specific allometric model of BRG used for estimating BWW, suggesting that this requires further sampling and testing. Other studies in allometry (e.g. Chatterji et al., 1985; Powell et al., 2016; Petteta et al., 2019) have shown different locations to show statistically significant different allometric models. It would be realistic to hypothesize that: (i) there are no large environmental differences between the Helgoland sites and BRG; and (ii) that the inclusion of data of *O. edulis* individuals from further locations and enlarging the number of levels for the variable might indeed increase predictive accuracy of the RF models that include these categorical variables.

While relatively accurate, the developed RF models can be further improved and complemented by, for example, the weight-to-weight transformation factors also presented in this study. Including a larger data set with clustered *O. edulis* organisms might prove the categorical variable type (here oyster clusters; Supporting Information 1) to be important to estimate weight, rather than be regarded as a statistically noisy variable. From this data set, only 60 data points corresponded to clusters of 2 (n = 12) and 3 (n = 48) individuals, whereas the

TABLE 3 Mean absolute error (MAE) in g of the general and location-specific allometric models, and for the 15 (A–O) random forest (RF) models for estimating total wet weight (TWW), shell wet weight (SWW) and soft tissue wet weight (BWW). Additional MAE calculated for the random forest (RF) models based on location specific data is also provided. Random forest models A-O include the following factor combinations: A, oyster height (in mm); B, oyster height (in mm) and width (in mm); C, oyster height (in mm), width (in mm) and length (in mm); D, oyster length (in mm); E, oyster length (in mm) and width (in mm); F, oyster width (in mm); G, oyster height (in mm) and type; H, oyster height (in mm), width (in mm) and type; I, oyster height (in mm), width (in mm), length (in mm) and type; J, oyster height (in mm) and location; K, oyster height (in mm), width (in mm) and location; L, oyster height (in mm), width (in mm), length (in mm) and location; M, oyster height (in mm), type and location; N, oyster height (in mm), width (in mm), type and location, and; O, oyster height (in mm), width (in mm), length (in mm), type and location. Values in bold represent RF models with lower MAE than that of the allometric models.

	MAE for total wet weight estimations (in g)			MAE for shell wet weight estimations (in g)			MAE for soft tissue wet weight estimations (in g)		
Model	Helgoland	BRG	Both	Helgoland	BRG	Both	Helgoland	BRG	Both
Allometric	6.54	4.62	5.78	4.63	2.68	3.94	0.72	0.45	0.68
RF-A	**4.05**	**2.73**	**3.52**	**3.00**	**1.47**	**2.06**	**0.42**	**0.34**	**0.40**
RF-B	**2.71**	**2.00**	**2.42**	**2.20**	**0.91**	**1.41**	**0.36**	**0.23**	**0.33**
RF-C	**2.41**	**1.72**	**2.13**	**2.00**	**0.82**	**1.27**	**0.34**	**0.22**	**0.31**
RF-D	**4.51**	**3.94**	**4.28**	**4.41**	**1.35**	**2.53**	**0.53**	**0.23**	**0.46**
RF-E	**3.55**	**3.12**	**3.38**	**3.16**	**0.97**	**1.81**	**0.43**	**0.24**	**0.38**
RF-F	**6.07**	5.47	5.83	**4.28**	**1.34**	**2.47**	**0.56**	**0.26**	**0.49**
RF-G	8.23	6.91	7.69	9.00	7.40	8.02	1.00	1.10	1.02
RF-H	**5.82**	5.04	**5.50**	6.47	5.54	5.90	0.81	0.83	0.81
RF-I	**4.67**	**4.03**	**4.41**	5.22	4.17	4.57	**0.71**	0.69	0.71
RF-J	9.25	6.99	8.33	5.34	2.93	**3.86**	0.83	**0.37**	0.72
RF-K	**6.18**	**4.22**	**5.39**	**3.90**	**1.86**	**2.65**	**0.63**	**0.32**	**0.56**
RF-L	**4.83**	**3.25**	**4.19**	**3.44**	**1.45**	**2.21**	**0.56**	**0.30**	**0.50**
RF-M	10.99	8.19	9.85	7.61	6.10	6.68	1.12	0.81	1.05
RF-N	8.17	5.72	7.18	6.36	4.82	5.41	0.92	0.69	0.87
RF-O	**6.19**	**4.33**	**5.43**	5.55	4.10	4.65	0.77	0.57	0.73

Abbreviation: BRG, Borkum Reef Ground.

remaining data points (n = 1,341) corresponded to single individuals. Including more measurements, especially for individuals with TWW > 75 g, SWW > 45 g, and BWW > 6 g, could also improve the precision of the RF models, which tend to underestimate the weight of large individuals. Another improvement would be the addition of measures of *O. edulis* individuals from other areas of the North Sea, the North-east Atlantic and other biogeographic regions where *O. edulis* populations exist (e.g. the Mediterranean or the Black Sea; Aydin & Biltekin, 2020), or data from monitoring conducted across Europe within the frame of several restoration initiatives (see Pogoda et al. (2019) for a detailed list of programmes). Further parameters that could be included are tidal location (intertidal or subtidal) or season of the year. The latter might be of great importance, since it has been shown that the condition of *O. edulis* (expressed as the ratio of soft tissue to shell weight) varies with season (Pogoda, Buck & Hagen, 2011; Merk, Colsoul & Pogoda, 2020), and relates to the reproductive state of *O. edulis* individuals. Allometric and RF models tended to underestimate weight for larger individuals, which could be related to the reproductive activity, as energetic investment of larger organisms is focused on reproduction rather than growth. Once RF models are improved, they could also be applied to estimate dry weights by including weight-to-weight transformation factors. The application of this combined approach can potentially be used for calculation of condition indices (Walne & Mann, 1975; Davenport & Chen, 1987) for oyster populations without the need of destructive sampling and measurements, i.e. without sacrificing any animals.

One aspect to consider, is the potential use of RF models to estimate the condition of any given population of *O. edulis* (see above), and thus the potential to facilitate monitoring efforts across the Native Oyster Restoration Alliance (NORA), supporting the implementation of a coherent habitat restoration monitoring programme (zu Ermgassen et al., 2021). Oyster shell size frequency, density, and condition index are included in the suggested metrics and required for monitoring metrics. The measured shell size frequency could provide the basic data input for the RF models developed in this study, thus resulting in an additional weight frequency measurement. As per the NORA monitoring handbook (zu Ermgassen et al., 2021), density should be provided as abundance and should be primarily measured by means of seabed image (SBI) techniques. Estimation of abundance by means of SBI implies the presence of a spatial reference, which allows for measuring seabed area (Solan et al., 2003). Using a spatial reference can enable height, width and/or length of the oysters identified via SBI to be measured. As a consequence, such measurements can also be used as input data for the RF models to provide biomass estimations in a non-destructive way. Applying suitable transformations, biomass stock estimations can be transformed to blue carbon stocks (biologically fixed carbon), and can be used to quantify ecosystem service potential (in terms of climate change mitigation) provided by a given *O. edulis* population. It is worth pointing out that, while the models of this study were developed for *O. edulis*, they can be further developed and adapted for associated fauna of *O. edulis* populations, or also for other marine species of interest for conservation and restoration programmes. This could potentially provide further metrics for estimating community development and status, facilitating biodiversity projections and conservation as such.

Additional to the potential facilitation of the monitoring activities conducted by researchers across Europe, the models can also make use of data collected by voluntary workers. During the last NORA4 conference (Native Oyster Restoration Alliance, 2021), volunteer (citizen science) work done by private organizations and research projects was presented. Part of this work entails either *in situ* measurement of *O. edulis* organisms, or doing short SBI surveys. The non-destructive nature of these surveys naturally does not include measurements of either SWW or BWW (nor any dry weight). Thus, by using these data as input for RF models, combined with the use of transformation factors, available data collected by citizen scientists can be used for broader regional studies (e.g. calculation of biomass or condition indices). This would open the door to engaging the community to contribute to the restoration efforts of natural populations by supporting conservation and restoration with valuable data and increase the output of single metrics.

5 | CONCLUSION

The RF models developed and presented in this study have the potential to outperform classical allometric models, providing a flexible tool to support and to facilitate monitoring of European oyster restoration programmes throughout the North Sea, the north-east Atlantic and potentially other regions where *O. edulis* populations occur. In combination with the weight-to-weight transformation factors presented here, they increase the data output and minimize destructive sampling. The presented RF models can be applied to existing data and to data provided by volunteers or citizen scientists to enhance the output by transforming to additional metric data.

In general, RF and machine learning tools can be easily adapted and updated to increase their precision e.g. for different ecological regions or seasons. This is key to improve models including categorical variables which, in their current status, performed poorly (e.g. those including location). The improvement of RF models for estimating *O. edulis* total, shell and body weights will provide tools to obtain the respective monitoring metrics (zu Ermgassen et al., 2021). Due to their flexibility and accuracy, RF models can be further developed to provide helpful metrics for conservation and restoration programmes, not only at the population level, but also at community level.

ACKNOWLEDGEMENTS

We thank the German Federal Agency for Nature Conservation (BfN) for funding and supporting this study in close cooperation. We thank the captain and crew of RV Heincke for their help during the cruises HE426, 482, 485, 490, 496/1 and 506, as well as the captains and crews of the vessels Mya II, Uthörn, MV Multrasalvor 4 and of SSV Arctic Ocean. We also thank the team of the Centre for Scientific Diving of the Alfred Wegener Institute, namely Prof. Dr Philipp Fischer, Dr Markus Brand, Christoph Walcher, Lisa Spotowitz, Christopher Groß and Marco Warmuth for their constant and valuable support to implement and investigate the fieldwork and especially the following colleagues for their help throughout this study: Corina Peter, Bérenger Colsoul, Tanja Hausen, Anna-Lena Böger, Lena Blum, and Anna Feuring. We thank Mathias Wegener and Timon Ameis for providing the dry weight measurements. We thank the Federal German Water and Shipping Agency (WSV) and WindMW GmbH for supporting field experiments within the Natura 2000 site GRB and within the OWF Meerwind Süd I Ost. This research is part of the testing and development project RESTORE (2016-2025), funded by the German Federal Agency for Nature Conservation, grant numbers FKZ 3516892001, FKZ 3519892016, FKZ 3520892013. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflicting interests.

AUTHOR CONTRIBUTIONS

Santiago E. A. Pineda-Metz: Conception (lead); data curation (lead); investigation (lead); methodology (lead); validation (lead); visualization (lead); writing - original draft preparation (lead); writing - review & editing (lead); resources (equal); and supervision (equal). **Verena Merk:** Data curation; investigation; resources; writing - original draft; writing - review and editing. **Bernadette Pogoda:** Funding acquisition; investigation; project administration; resources; supervision; writing - original draft; writing - review and editing.

DATA AVAILABILITY STATEMENT

Data on oyster size and weight measurements have been submitted to the PANGAEA data repository (Pineda-Metz, Merk & Pogoda, 2022). Data can also be provided upon request to the corresponding author.

ORCID

Santiago E. A. Pineda-Metz https://orcid.org/0000-0001-7780-6449
Verena Merk https://orcid.org/0000-0002-9034-3431
Bernadette Pogoda https://orcid.org/0000-0003-3997-426X

REFERENCES

Andreu, B. (1968). Fishery and culture of mussels and oyster in Spain. *Proceedings Mollusca; Marine Biological Association India*, 3, 835–846.

Aydin, A. & Biltekin, D. (2020). First morphometric aspects and growth parameters of the European flat oyster (*Ostrea edulis* Linnaeus, 1758) for the Black Sea, Turkey. *Natural and Engineering Sciences*, 5(2), 101–109. https://doi.org/10.28978/nesciences.756736

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brey, T. (2001). *Population dynamics in benthic invertebrates. A virtual handbook*. *Version 01.2*. https://www.thomas-brey.de/science/virtualhandbook [Accessed 04 January 2022].

Brey, T., Rumohr, H. & Ankar, S. (1988). Energy content of microbenthic invertebrates: general conversion factors from weight to energy. *Journal of Experimental Marine Biology and Ecology*, 117(3), 271–278. https://doi.org/10.1016/0022-0981(88)90062-7

Chatterji, A., Ansari, A., Ingole, B.S. & Parulekar, A.H. (1985). Length-weight relationship of giant oyster, *Crassostrea gryphoides* (Schlotheim). *Mahasasgar-Bulletin of the National Institute of Oceanography*, 18(4), 521–524.

Coughlan, N.E., Cunningham, E.N., Cuthbert, R.N., Joyce, P.W.S., Anastácio, P., Banha, F. et al. (2021). Biometric conversion factors as a unifying platform for comparative assessment of invasive freshwater bivalves. *Journal of Applied Ecology*, 58(9), 1945–1956. https://doi.org/10.1111/1365-2664.13941

Cutler, D.R., Edwards, T.C., Jr., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. et al. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Davenport, J. & Chen, X. (1987). A comparison of methods for the assessment of condition of condition in the mussel (*Mytilus edulis* L.). *Journal of Molluscan Studies*, 53(3), 293–297. https://doi.org/10.1093/mollus/53.3.293

De'ath, G. & Fabricius, K.E. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81(11), 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Eklöf, A., Austin, Å., Bergström, U., Donadi, S., Eriksson, B.D.H.K., Hansen, J. et al. (2017). Size matters: relationships between body size and body mass of common coastal, aquatic invertebrates in the Baltic Sea. *PeerJ*, 5, e2906. https://doi.org/10.7717/peerj.2906

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

zu Ermgassen, P.S.E., Bos, O., Debney, A., Gamble, C., Glover, A., Pogoda, B. et al. (2021). *European native oyster habitat restoration monitoring handbook*. London, UK: The Zoological Society of London.

Froese, R. (2006). Cube law, condition factor and weight-length relationships: history, meta-analysis and recommendations. *Journal of Applied Ichthyology*, 22(4), 241–253. https://doi.org/10.1111/j.1439-0426.2006.00805.x

Galtsoff, P.S. (1931). The weight-length relationship of the shells of the Hawaiian pearl oyster, *Pinctada* sp. *The American Naturalist*, 65(700), 423–433. https://doi.org/10.1086/280387

Kijewski, T., Zbawicka, M., Strand, J., Kautsky, H., Kotta, J., Rätsep, M. et al. (2019). Random forest assessment of correlation between environmental factors and genetic differentiation of populations: case of marine mussels *Mytilus*. *Oceanologia*, 61(1), 131–142. https://doi.org/10.1016/j.oceano.2018.08.002

Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.

Macreadie, P.I., Anton, A., Raven, J.A., Beaumot, N., Connolly, R.M., Fries, D.A. et al. (2019). The future of blue carbon science. *Nature Communications*, 10, 3998. https://doi.org/10.1038/s41467-019-11693-w

Mann, R., Southworth, M., Harding, J.M. & Wesson, J.A. (2009). Population studies of the native eastern oyster, *Crassostrea virginica*, (Gmelin, 1791) in the James River, Virginia, USA. *Journal of Shellfish Research*, 28(2), 193–220. https://doi.org/10.2983/035.028.0203

Merk, V., Colsoul, B. & Pogoda, B. (2020). Return of the native: survival, growth and condition of European oysters reintroduced to German offshore waters. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 30(11), 2180–2190. https://doi.org/10.1002/aqc.3426

Miller, K., Huettmann, F., Norcross, B. & Lorenz, M. (2014). Multivariate random forest models of estuarine associated fish and invertebrate communities. *Marine Ecology Progress Series*, 500, 159–174. https://doi.org/10.3354/meps10659

Native Oyster Restoration Alliance. (2021). *NORA 4: reconnecting across Europe (Abstracts)*. https://noraeurope.eu/wp-content/uploads/2021/11/NORA-4-Abstracts.pdf [Accessed 05 January 2022].

Oksanen J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. et al. (2020). vegan: Community Ecology Package. *R package version 2.5–7*. https://CRAN.R-project.org/package=vegan

Oppel, S., Strobl, C. & Huettmann, F. (2009). *Alternative methods to quantify variable importance in ecology*. Munich, Germany: University of Munich. Technical Report 65.

Orton, J. (1935). Laws of shell-growth in English native oyster (*Ostrea edulis*). *Nature*, 135(3409), 340–341. https://doi.org/10.1038/135340a0

OSPAR. (2009). *Background document for Ostrea edulis and Ostrea edulis beds*. OSPAR Commission.

Packard, G.C. (2014). On the use of log-transformation versus nonlinear regression for analyzing biological power laws. *Biological Journal of the Linnean Society*, 113(4), 1167–1178. https://doi.org/10.1111/bij.12396

Perez, D.E. & Santelli, M.B. (2018). Allometric shell growth in infaunal burrowing bivalves: examples of the archiheterodonts *Claibornicardia paleopatagonica* (Ihering, 1903) and *Crassatela kokeni* Ihering, 1899. *PeerJ*, 6, e5051. https://doi.org/10.7717/peerj.5051

Petteta, A., Bargione, G., Vasapollo, C., Virgili, M. & Lucchetti, A. (2019). Length-weight relationships of bivalve species in Italian razor clam *Ensis minor* (Chenu, 1843) (Mollusca: Bivalvia) fishery. *The European Zoological Journal*, 86(1), 363–369. https://doi.org/10.1080/24750263.2019.1668066

Pineda-Metz, S., Colsoul, B., Niewöhner, M., Hausen, T., Peter, C. & Pogoda, B. (2023). Setting the stones to restore and monitor European flat oyster reefs in the German North Sea. *Aquatic Conservation: Marine and Freshwater Ecosystems*. This issue – details to follow.

Pineda-Metz, S.E.A., Merk, V. & Pogoda, B. (2022). Oyster size and weight measurements used for the development of weight-to-weight transformation factors and weight estimation models. *PANGAEA*. https://doi.org/10.1594/PANGAEA.949238

Pogoda, B. (2019). Current status of the European oyster decline and restoration in Germany. *Humanities*, 8(1), 9. https://doi.org/10.3390/h8010009

Pogoda, B., Boudry, P., Bromley, C., Cameron, T.C., Colsoul, B., Donnan, D. et al. (2020a). NORA moving forward: developing an oyster restoration network in Europe to support the Berlin oyster recommendation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 30(11), 2031–2037. https://doi.org/10.1002/aqc.3447

Pogoda, B., Brown, J., Hancock, B., Preston, J., Pouvreau, S., Kamermans, P. et al. (2019). The native oyster restoration Alliance (NORA) and the Berlin oyster recommendation: bringing back a key ecosystem engineer by developing and supporting best practice in Europe. *Aquatic Living Resources*, 32, 13. https://doi.org/10.1051/alr/2019012

Pogoda, B., Buck, B.H. & Hagen, W. (2011). Growth performance and condition of oysters (*Crassostrea gigas* and *Ostrea edulis*) farmed in an offshore environment (North Sea, Germany). *Aquaculture*, 319(3-4), 484–492. https://doi.org/10.1016/j.aquaculture.2011.07.017

Pogoda, B., Colsoul, B., Hausen, T., Merk, V. & Peter, C. (2020b). *Wiederstellung der Bestände der Europäischen Auster (Ostrea edulis) in* *der deutschen Nordsee (RESTORE Voruntersuchung)*, Bonn: Bundesamt für Naturschutz. https://doi.org/10.19217/skr582

Pogoda, B., Merk, V., Colsoul, B., Hausen, T., Peter, C., Pesch, R. et al. (2020c). Site selection for biogenic reef restoration in offshore environments: the Natura 2000 area Borkum reef ground as a case study for native oyster restoration. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 30(11), 2163–2179. https://doi.org/10.1002/aqc.3405

Powell, E.N., Mann, R., Ashton-Alcox, K.A., Kim, Y. & Bushek, D. (2016). The allometry of oysters: spatial and temporal variation in the length-biomass relationships for *Crassostrea virginica*. *Journal of the Marine Biological Association of the United Kingdom*, 95(5), 1127–1144. https://doi.org/10.1017/S0025315415000703

Prasad, A.M., Iverson, L.R. & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

R Core Team. (2020). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rather, T.A., Kumar, S. & Khan, J.A. (2021). Using machine learning to predict habitat suitability of sloth bears at multiple spatial scales. *Ecological Processes*, 10(1), 40. https://doi.org/10.1186/s13717-021-00323-3

Riccardi, A. & Bourget, E. (1998). Weight-to-weight conversion factors for marine benthic macroinvertebrates. *Marine Ecology Progress Series*, 163, 245–251. https://doi.org/10.3354/meps163245

Rumohr, H., Brey, T. & Ankar, S. (1987). A compilation of biometric conversion factors for benthic invertebrates of the Baltic Sea. *The Baltic Sea Biological Publications*, 9, 1–56.

Shah, S.H., Angel, Y., Houborg, R., Ali, S. & McCabe, M.E. (2019). A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sensing*, 11(8), 920. https://doi.org/10.3390/es11080920

Solan, M., Germano, J.D., Smith, C., Michaud, E., Parry, D., Wenzhöfer, F. et al. (2003). Towards a greater understanding of patterns, scale and process in marine benthic systems: a picture is worth a thousand worms. *Journal of Experimental Marine Biology and Ecology*, 285-286, 313–338. https://doi.org/10.1016/S0022-0981(02)00535-x

Sujitha, T. (2013). Allometric relationships of short neck clam *Paphia malabarica* from Dharmadom estuary, Kerala. *Journal of the Marine Biological Association of India*, 55(1), 50–54. https://doi.org/10.6024/jmbai.2013.55.1.01755-08

Walne, P.R. & Mann, R. (1975). Growth and biocheical composition in *Ostrea edulis* and *Crassostrea gigas*. In: Barnes, H. (Ed.) *Proceedings of the Ninth European Marine Biology Symposium*. Aberdeen, Scotland: Aberdeen University Press, pp. 587–607.

Wang, L., Zhou, X., Zhu, X., Dong, Z. & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, 4(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

Wei, C.-L., Rowe, G.T., Escobar-Briones, E., Boetius, A., Soltwedel, T., Caley, M.J. et al. (2010). Global patterns and predictions of seafloor biomass using random forests. *PLoS ONE*, 5(12), e15323. https://doi.org/10.1371/journal.pone.0015323

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. New York, United States: Springer-Verlag.

Willmontt, C.J. & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82. https://doi.org/10.3354/cr030079

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pineda-Metz, S.E.A., Merk, V. & Pogoda, B. (2023). A machine learning model and biometric transformations to facilitate European oyster monitoring. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 1–13. https://doi.org/10.1002/aqc.3912