

Predictive modelling of grain-size distributions from marine electromagnetic profiling data using end-member analysis and a radial basis function network

B. Baasch,^{1,*} H. Müller^{1,2} and T. von Dobeneck¹

¹MARUM - Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, D-28359 Bremen, Germany. E-mail: baasch@uni-bremen.de

²Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

Accepted 2018 April 12. Received 2018 April 8; in original form 2017 November 15

SUMMARY

In this work, we present a new methodology to predict grain-size distributions from geophysical data. Specifically, electric conductivity and magnetic susceptibility of seafloor sediments recovered from electromagnetic profiling data are used to predict grain-size distributions along shelf-wide survey lines. Field data from the NW Iberian shelf are investigated and reveal a strong relation between the electromagnetic properties and grain-size distribution. The here presented workflow combines unsupervised and supervised machine-learning techniques. Non-negative matrix factorization is used to determine grain-size end-members from sediment surface samples. Four end-members were found, which well represent the variety of sediments in the study area. A radial basis function network modified for prediction of compositional data is then used to estimate the abundances of these end-members from the electromagnetic properties. The end-members together with their predicted abundances are finally back transformed to grain-size distributions. A minimum spatial variation constraint is implemented in the training of the network to avoid overfitting and to respect the spatial distribution of sediment patterns. The predicted models are tested via leave-one-out cross-validation revealing high prediction accuracy with coefficients of determination (R^2) between 0.76 and 0.89. The predicted grain-size distributions represent the well-known sediment facies and patterns on the NW Iberian shelf and provide new insights into their distribution, transition and dynamics. This study suggests that electromagnetic benthic profiling in combination with machine learning techniques is a powerful tool to estimate grain-size distribution of marine sediments.

Key words: Neural networks, fuzzy logic; Statistical methods; Electrical properties; Magnetic properties; Marine electromagnetics; Controlled source electromagnetics (CSEM).

1 INTRODUCTION

Coastal and shelf areas as transition zones between land and ocean play an important economic role in modern societies. This is reflected in the increasing development of ports, wind farms, oil rigs, pipelines and cable routes. For the planning, construction and maintenance of these marine infrastructures accurate information on shallow marine sediments is needed.

The grain-size distribution (GSD) plays a decisive role in terms of sediment characterization, influencing most of the sediment physical properties and behaviour. Its spatial distribution is of high interest for researchers and engineers. Traditional sampling methods to determine grain size (GS) are time consuming and labour intensive. This led to an increased interest in fast and less expensive remote seafloor characterization techniques.

Acoustic methods such as multibeam echo-sounders are the most cost-effective tools to map large seabed areas (Brown *et al.* 2011). However, acoustic sediment characterization is still challenging because of the complex relation between the acoustic backscatter and the seabed characteristics. A major remaining issue is the separation of seafloor roughness from sediment properties (e.g. Jackson & Briggs 1992).

* Now at: Institute of Transportation Systems, German Aerospace Center (DLR e.V.), Lilienthalplatz 7, 38108 Braunschweig, Germany.

In recent years, electromagnetic methods have been employed as a complementary tool for the estimation of marine sediment properties. Initially developed to investigate deep tectonic features and hydrocarbon reservoirs, marine electromagnetic methods have been increasingly applied to survey marine near surface sediments (Butler 2009). Most of these systems are designed to recover the electric conductivity (EC) of the seafloor (e.g. Cheesman *et al.* 1990; Evans 2001; Schwalenberg *et al.* 2010; Swidinsky *et al.* 2015), which can be linked to sediment porosity (Archie 1942) and GS (e.g. Jackson *et al.* 1978). Müller *et al.* (2012) developed a frequency domain central loop electromagnetic method for marine applications capable to provide information about magnetic susceptibility (MS) in addition to EC. Magnetic properties have been used as particle size proxy in many different environments (e.g. Oldfield *et al.* 1985; Booth *et al.* 2005). Recently, Baasch *et al.* (2017) showed that MS and EC correlate with mean GS, mud content and sorting of marine sediments and gave evidence that combining both electromagnetic parameters improves the accuracy of GS predictions.

A common problem of all seafloor sensing techniques is to convert the measured geophysical parameters to qualitative or quantitative sediment characteristics. In seabed as well as terrestrial land cover mapping unsupervised techniques have been previously used to create attribute classes and maps from geophysical and remote sensing data which could then be linked to geology by means of ground-truth samples (e.g. Paasche *et al.* 2006; De & Chakraborty 2009; Eberle *et al.* 2015). More recently supervised machine learning techniques became popular for predictive modelling of marine sediment (e.g. Huang *et al.* 2012, 2013, 2014; Dising & Stephens 2015). With these methods the models are trained by ground-truth data and then used to predict seafloor properties for the whole survey area.

Previous efforts to predict GS from geophysical data aimed to estimate only fractions of the GSD, such as gravel (per cent), sand (per cent), mud (per cent) or statistical measures such as the mean GS (e.g. Heil & Schmidhalter 2012; Huang *et al.* 2012; Baasch *et al.* 2017). Independent prediction of single fractions of the GSD may produce misleading results because of the compositional character of the GSD. Because the frequencies of the GSD components have a constant sum, no single GS fraction is free to vary separately from the rest of the total composition. Hence, the prediction of just one or a few GS fractions will always be biased by the remaining fractions. The best solution to overcome this problem is to predict the entire GSD. Following this idea we present a new methodology to predict compositional data from geophysical survey data. Specifically, we predict GSD from electromagnetic profiling data. The training GSDs used for our studies is derived from core samples and contains 92 logarithmically spaced GS classes from 0.38 to 2000 μm . Many of these GS classes are insufficiently populated or their abundances do not vary significantly between sediment samples and hence do not carry useful information. It is therefore desirable to reduce the number of predicted variables to the minimum amount necessary to describe the variability of the GSD. In order to achieve this, we employ a low-rank approximation namely a non-negative matrix factorization (NMF) approach that determines GSD end-members. The aim is then to predict the abundances of these end-members from the electromagnetic data along the survey profiles. The prediction is done by a radial basis function network (RBFN) optimized for prediction of compositional data. The predicted GSD end-member abundances can finally be back transformed to GSD. The here presented workflow to predict GSD from electromagnetic profiling data can be summarized as follows:

- (i) EC and MS are recovered through inversion of corrected and calibrated electromagnetic profiling data.
- (ii) End-member modelling is applied to the GSDs of sediment samples to reconstruct independent end-members with characteristic GSDs.
- (iii) The relationship between EC, MS and the abundance of each GSD end-member is modelled using an RBFN trained by the EC, MS (input variables) and end-member abundances (target variables) at the sample locations.
- (iv) The established model is then used to predict end-member abundances for each data point along the survey profiles.
- (v) Finally, a complete GSD for each electromagnetic data point can be reconstructed using the end-members' GSDs and their predicted abundances.

Applying this approach, we reconstructed GSDs along three profiles on the NW Iberian shelf validating the potential of electromagnetic profiling for quantitative sediment characterization.

2 METHODOLOGY

2.1 End-member modelling

In the following, we assume that the seafloor sediments are a simple mixture of different sediment types (end-members) and that the GSDs of the sediment samples can be expressed as a linear combination of the GSDs of the end-members. The aim of the end-member unmixing is then to unravel the number of end-members and their compositions (GSDs) from the GSDs of the sediment samples. The mixing model can be written in algebraic form as

$$\mathbf{X} = \mathbf{AS}, \quad (1)$$

where \mathbf{X} is the $N \times M$ data matrix that contains the GSDs, namely the relative amount of each of the N GS classes in all M samples. The elements of \mathbf{X} are normally quantities of dimension one or expressed in percentage. The $L \times M$ matrix \mathbf{S} denotes the frequencies of the N GS classes in the L end-members and has the same units as \mathbf{X} . \mathbf{A} is a $N \times L$ matrix representing the dimensionless abundances of the L end-members in each of the M samples. To factor \mathbf{X} in \mathbf{A} and \mathbf{S} we use an unsupervised unmixing approach based on an NMF approach using an alternating least squares (ALS) technique. The basic idea of the ALS algorithms is to solve each of the factors \mathbf{A} and \mathbf{S} alternately while keeping the other factor constant (Paatero & Tapper 1994). In the simplest ALS algorithm (e.g. Berry *et al.* 2006) the matrix \mathbf{A} is initialized first with a random matrix before \mathbf{S} is computed from \mathbf{X} and \mathbf{A} via a simple least-square operation. After the least squares computation of \mathbf{S} all its negative elements are set to zero. Subsequently, \mathbf{A} is computed in the same way from \mathbf{X} and \mathbf{S} and finally all negative elements in \mathbf{A} are set to zero, too. The computation of \mathbf{S} and \mathbf{A} is repeated until the maximum number of iteration is reached. This simple ALS algorithm is very fast and does not lock elements when they become zero (locking phenomenon), as especially NMF techniques of the multiplicative update class do. Due to the limited number of sediment samples and GS classes, the matrices \mathbf{A} and \mathbf{S} are relatively small compared to other disciplines, such as text mining or image processing. Therefore, we can afford a computationally more expensive technique than the simple ALS algorithm. Instead of using an ad-hoc enforcement of non-negativity, we use the non-negative least squares (NNLS) algorithm of Lawson & Hanson (1995) that properly enforces non-negativity and is known to converge to a local minimum. We also normalize \mathbf{S} and \mathbf{A} in each iteration such that their row-elements sum to one, which

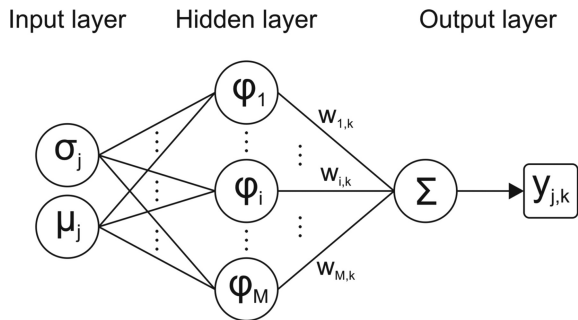


Figure 1. RBFN network structure. The j th input layer contains normalized EC and MS values from the j th electromagnetic profiling data point, in the hidden layer EC and MS from the j th data point are connected to EC and MS from the M centres. The output is the abundance of the k th end-member at j th electromagnetic data point.

is an inherent constraint for compositional data. The method described above offers reliable results without any preconditioning of the data. The number of end-members included in the final mixing model can be found via principle component analysis, goodness-of-fit tests or *a priori* knowledge. In general, the goal is to keep the number of components low while maintaining a reasonably good approximation of the input data (Just *et al.* 2012).

2.2 Radial basis function network

An RBFN is a machine learning method belonging to the class of artificial neural networks which uses a radial basis function (RBF) as activation function. RBFNs are used for function approximation, data interpolation or smooth fitting of data. An RBFN generally consist of three layers (Broomhead & Lowe 1988). The first layer is the input layer which only transfers the data to the next layer. The second layer is a hidden layer with a non-linear RBF, the so-called activation function. The third layer is a linear output layer. Its output is given by

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^M w_i \varphi(\|\mathbf{x} - \mathbf{c}_i\|) + p(\mathbf{x}), \quad (2)$$

where \mathbf{x} is the input vector; $\varphi(\|\mathbf{x}\|)$ is an RBF, where $\|\cdot\|$ denotes the Euclidean norm; w_i is the weight corresponding to the i th centre \mathbf{c}_i , a vector of the same size as \mathbf{x} ; and $p(\mathbf{x})$ is an optional polynomial of low degree, which can be included to model global trends of the data (Billings *et al.* 2002) and thus enhancing the extrapolation performance of the RBFN. Because of its simple form, smoothness and other advantages, the RBF in eq. (2) is commonly a Gaussian function. The disadvantage of the Gaussian function is that the spread parameter which controls the width of the RBF needs to be carefully chosen. If the data density of the input variables is non-uniform, a proper selection of this parameter might be difficult or even impossible. In contrast, here a thin-plate spline is used, which is a scale independent RBF and thus handles better varying data density (Billings *et al.* 2002). A thin-plate spline is a special polyharmonic spline of the form:

$$\varphi(\|\mathbf{x}\|) = \|\mathbf{x}\|^2 \ln(\|\mathbf{x}\|). \quad (3)$$

Let us now consider the case in which the input to the network is the MS and EC data obtained from the inversion of the electromagnetic profiling data and the output is the abundance of the sediment GSD end-members (Fig. 1).

The input vector elements are individually normalized to zero mean and unit variance in order to calculate the Euclidean norm in the RBFN's hidden layer. The set of input vectors then has the form $\mathbf{x}_j = \begin{bmatrix} \sigma_j \\ \mu_j \end{bmatrix}$ ($j = 1, \dots, N$) with σ_j and μ_j the normalized EC and MS values of the j th data point, respectively. A linear polynomial is added to the RBF in the hidden layer so that eq. (2) can be written as

$$\hat{y}_{j,k}(\mathbf{x}_j) = \sum_{i=1}^M w_{i,k} \varphi(\|\mathbf{x}_j - \mathbf{c}_i\|) + \mathbf{v}_k^T \begin{bmatrix} 1 \\ \mathbf{x}_j \end{bmatrix}, \quad i=1, 2, \dots, M, \quad (4)$$

$$j=1, 2, \dots, N, \quad k=1, 2, \dots, L,$$

where $\hat{y}_{j,k}$ is an approximation of the abundance of the k th GS end-member at the j th data point, $w_{i,k}$ links the i th hidden neuron with the k th output neuron and $\mathbf{v}_k^T = [v_{1,k}, v_{2,k}, v_{3,k}]$ contains the three weights of the linear polynomial at the k th output neuron. The number of output neurons L represents the number of end-members. The centres \mathbf{c}_i were chosen to be the normalized EC and MS values at the i th sediment sample location. Theoretically, the training of the RBFN then only involves the estimation of the weights in the output layer. This can be obtained by solving the symmetric, linear system of equations

$$\begin{bmatrix} \boldsymbol{\varphi}_t & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_k \\ \mathbf{v}_k \end{bmatrix} = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{0} \end{bmatrix}, \quad (5)$$

for each end-member separately, where $\varphi_{j,i} = (\|\mathbf{c}_j - \mathbf{c}_i\|)$ ($i = 1, \dots, M, j = 1, \dots, M$) are the components of $\boldsymbol{\varphi}_t$, $\mathbf{C} = \begin{bmatrix} 1 & \dots & 1 \\ \mathbf{c}_1 & \dots & \mathbf{c}_M \end{bmatrix}$ and $\mathbf{w}_k, \mathbf{v}_k$ are the vectors of weights and \mathbf{y}_k contains the abundances of the k th GSD end-member obtained from the M sediment samples. Eq. (5) also fulfills the orthogonality conditions $\sum_{i=1}^M w_{i,k} = 0$ and $\sum_{i=1}^M w_{i,k} \mathbf{c}_i = \begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}$. Since the number of centres equals the number of input vectors, the misfit between the observed and the predicted data will be zero. Considering that the data are contaminated by noise, there is little reason to exactly fit the data. In contrast, it may be desired that the output meet some constraints based on *a priori* information. For our analysis, it can be assumed that the noise in the electromagnetic data of neighbouring measurements has a higher variation than the signal. Therefore, we included a minimum variation term in the training process, which penalizes variations of the predicted end-member abundances of adjacent electromagnetic measurement points. Another reasonable constraint is that the predicted end-member abundances at each data point are positive and sum to one. The sum-to-one constraint requires that the RBFNs of the L end-member abundances are trained together. The training of the weights may then be expressed by a linear least-square problem of the form

$$\min_{\mathbf{u}} \frac{1}{2} \|\boldsymbol{\Phi}_t \mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2} \|\mathbf{R} \boldsymbol{\Phi}_d \mathbf{u}\|_2^2, \quad (6)$$

such that

$$\hat{y}_{j,k} \geq 0, \quad (7)$$

$$\sum_{k=1}^L \hat{y}_{j,k} = 1, \quad j = 1, \dots, N \quad (8)$$

$$\sum_{i=1}^M w_{i,k} = 0, \quad (9)$$

$$\sum_{i=1}^M w_{i,k} \mathbf{c}_i = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (10)$$

The first term in eq. (6) is a measure of the data misfit, where

$$\Phi_t = \begin{bmatrix} \varphi_t \mathbf{C}^T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varphi_t \mathbf{C}^T \end{bmatrix},$$

a $(M \cdot L) \times (M \cdot L + 2)$ matrix containing the RBFs of the training data and the centres. The second term contains the minimum variation penalty function, where

$$\Phi_d = \begin{bmatrix} \varphi_d \mathbf{C}^T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varphi_d \mathbf{C}^T \end{bmatrix},$$

a $(N \cdot L) \times (N \cdot L + 2)$ matrix and $\varphi_{j,i} = \varphi(\|\mathbf{x}_j - \mathbf{c}_i\|)$ ($i = 1, \dots, M$, $j = 1, \dots, N$) are the components of φ_d , namely the RBFs connecting each input value with each centre. The vector $\mathbf{u} = [\mathbf{w}_1^T, \mathbf{v}_1^T, \dots, \mathbf{w}_L^T, \mathbf{v}_L^T]^T$ contains the weights and hence is subject of the training process and $\mathbf{d} = [\mathbf{y}_1^T, \dots, \mathbf{y}_L^T]^T$. The regularization parameter λ controls the trade-off between minimizing the prediction error and the spatial variation of the output. The matrix \mathbf{R} is a first-order finite difference operator. The smoothness constraint is applied to the output data in the spatial domain, namely along the acquisition sail line, and not in the feature space, which is the MS–EC plane. The constrained linear least-squares problem described in eqs (6)–(10) was solved using the MATLAB optimization toolbox (MATLAB R2015b).

3 NW IBERIAN SHELF STUDY

3.1 Study area and sedimentary settings

The NW Iberian shelf is a classic example of a low accumulation non-glaciated clastic shelf system (Lantzsch *et al.* 2009a,b). Its extension reaches from Cape Finisterre in the north (43°N) to the Douro River mouth in the south (41°N). The width of the shelf is relatively narrow increasing from 30 km in the north to 50 km in the south. The shelf break occurs at a water depth of 160–200 m. Sediment transport, deposition and erosion is controlled by riverine discharge, winter storms, longshore currents and the geomorphology of the shelf (e.g. Oliveira *et al.* 2002; Oberle *et al.* 2014a,b), as well as anthropogenic activities, in particular bottom trawling and dredging (Oberle *et al.* 2016a,b). One of the most prominent sedimentary features is the Galicia Mud Belt (GMB; Lantzsch *et al.* 2009b), a well-defined 50-km-long and 2–3-km-wide coast-parallel mud depocentre (Dias *et al.* 2002). The GMB is located mid-shelf, north of the Minho River estuary at 110–120 m water depth. It is made up of 90 per cent fluvio-genic silty and clayey material mainly originating from the Douro River (Dias *et al.* 2002). Seaward the GMB is boarded by glaucony sands consisting of up to 50 per cent paramagnetic glaucony (Odin & Lamboy 1988). On the outer shelf, very fine to fine sands predominate, which consist mainly of carbonate microfossil fragments, quartz and mica (Lantzsch *et al.* 2010, fig. 2). Previous electromagnetic surveys in this area showed that the main sediment facies could be identified by means of bivariate data analysis of EC and MS (Müller *et al.* 2012; Baasch *et al.* 2015, 2017). Specifically, the GMB is identified by high EC and MS, the glaucony-sand facies by high MS and low EC, while the sands on the outer shelf show moderate EC and low MS values.

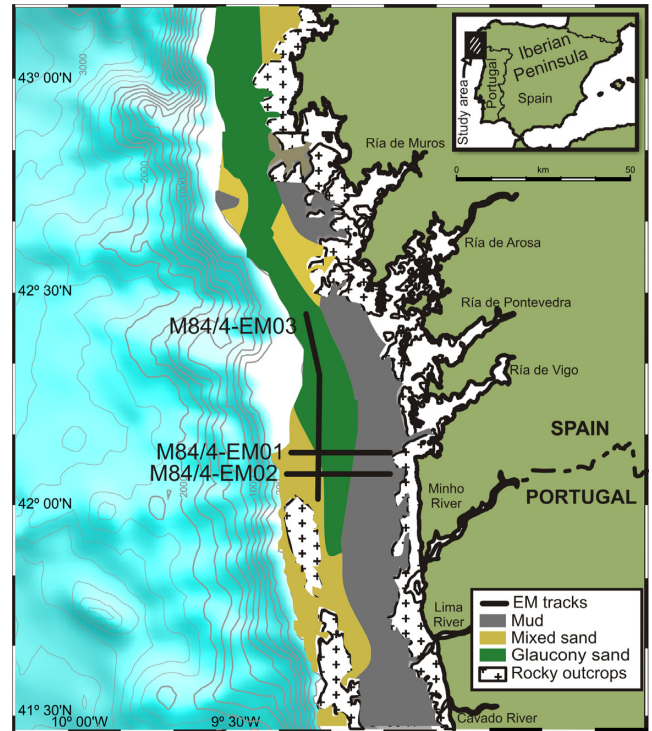


Figure 2. NW Iberian shelf. Colours indicate sediment facies, black lines indicate electromagnetic profiles from RV Meteor cruise M84/4b. Profile names are shown next to the lines. Modified from Lantzsch *et al.* (2010).

3.2 Electromagnetic measurements

The electromagnetic data used for this study have been recently described in Baasch *et al.* (2017). The data were acquired during the RV Meteor cruise M84/4b along three profiles, two of them run perpendicular and one parallel to the shore line (Fig. 2). The acquisition was carried out with the benthic profiler NERIDIS III developed at MARUM - Center for Marine Environmental Science at the University of Bremen, Germany (Baasch *et al.* 2015). The profiler is a bottom towed non-conductive, non-magnetic sled carrying a commercial broad-band frequency-domain electromagnetic induction sensor (Won *et al.* 1997) customized for marine operations. The sled is towed on the seafloor ensuring a constant distance between the sensor and the ground of typically 0.2 m. The electromagnetic data were acquired continuously along the ship track with a sampling rate of 25 Hz providing a very high spatial resolution.

The electromagnetic induction sensor consists of three co-axial and coplanar coils, namely source, receiver and bucking coil. The bucking coil cancels out the source transmitted primary field at the receiver coil location such that the receiver only measures the weaker secondary field. The secondary field can be expressed as a sum of an in-phase and a quadrature or out-of-phase component. The quadrature component is dominated by the induction of eddy currents in the conductive subsurface over the entire frequency range. The low-frequency in-phase response is controlled by the magnetization of the subsurface. With increasing frequencies the influence of induction on the in-phase response increases. While the amplitude of the induction-related in-phase and quadrature response increases with increasing frequency (within the frequency ranges considered here), the depth to which sediments influence this response decreases. Since the magnetization of the subsurface is frequency independent, the depth of investigation of the related in-phase response is frequency independent, too. The multifrequency

transmitter signal was created by superimposing frequencies of 75, 175, 1025, 5025 and 10025 Hz with a pulse-width modulation technique (Won *et al.* 1997).

The EC and MS of the subsurface sediments were reconstructed following a calibration and inversion approach described by Baasch *et al.* (2015) and Baasch *et al.* (2017), respectively: The electromagnetic data were corrected for instrument-related bias by comparing and linear fitting the electromagnetic data measured in the water column to the desired sea water response calculated from sea water EC measured with a conductivity-temperature-depth (CTD) probe attached to the profiler and a constant MS of -9×10^{-6} . Additionally, the secondary electromagnetic field was similarly calibrated using EC and MS ground-truth data measured on the collected sediment samples. EC was determined with a hand-held direct current probe with miniaturized four-electrode-in-line Wenner configuration and 4 mm electrode spacing. MS was measured with a laboratory susceptometer. To recover EC and MS of similar depth ranges, the half-space inversion was performed in three steps: (1) in-phase and quadrature values of all five frequencies were inverted simultaneously for EC and MS. These values were used as *a priori* information in the following inversion steps. (2) The high-frequency quadrature response was used to recover EC. (3) The low-frequency in-phase response was inverted for MS. The recovered EC and MS are then a function of the sedimentary characteristics of the subsurface to a depth of approximately 0.9 and 0.5 m, respectively (Müller *et al.* 2012).

3.3 Grains-size analysis

During the Meteor cruise M84/4, sediment samples at 105 different locations across the NW Iberian shelf were taken. 33 of these sample locations lie on the three here presented electromagnetic profiles. Sediment sampling was performed using a Rumohr corer (100-cm-long gravity corer), a grab sampler and a box corer. In contrast to a conventional grab sampler and box corer, the Rumohr corer allows probing the seafloor without disturbance of the surface sediments. The length of the recovered cores varied between 10 and 70 cm depending on sediment texture and coring device. For the GS analysis, the cores were subsampled every 10–20 cm based on visual inspection. GS analyses were carried out with a Coulter LS 200 laser particle sizer. The volume distribution was divided in 92 logarithmically spaced size classes ranging from 0.39 to 2000 μm . Particles larger than 2000 μm (gravel) were not present in the sediment samples. The arithmetic mean of the vertical subsamples at each sample location was taken for every individual GS class to get a single GSD for each sample location. These averaged GSDs served as input for the end-member modelling. Three of the 33 samples have not been considered in the RFBN training due to uncertainties in the sample localization.

4 RESULTS

4.1 Grain-size and end-member analysis

GSs of the 33 surface samples presented here have been recently discussed in Baasch *et al.* (2017). The sediments consist of predominantly homogeneous medium to coarse silts and fine to medium sands with no or low content of coarse sands (Fig. 3).

To determine the optimal number of end-members for the factorization of the GSDs, the Frobenius norm of residuals

$$\|\mathbf{E}(L)\|_F = \|\mathbf{X} - \mathbf{A}(L)\mathbf{S}(L)\|_F \quad (11)$$

and the coefficient of determination were calculated and plotted versus the number of end-members L (Fig. 4). Naturally, it can be assumed that the points with the highest curvature of both graphs indicate the best compromise between number of end-members and goodness-of-fit. In both graphs, two distinct corner points are visible. The first occurs at $L = 3$ (point with highest curvature) and the second at $L = 5$. Considering that three end-members explain less than 90 per cent ($R^2 = 0.85$) of the GSD variance, we only further investigated the four ($R^2 = 0.94$) and five end-member ($R^2 = 0.98$) solutions. It was found out that the RFBN-GSD prediction performed better using the four end-member model. This is mainly because adding a fifth end-member leads to more very small values in the end-member frequency distributions which make the regression less stable. The choice of four end-members is also supported by cluster analysis carried out by Lantzsch *et al.* (2010), who discovered four main sedimentary facies.

The four end-members were sorted according to their modes such that end-member 1 has the smallest and end-member 4 the highest mode (Table 1). End-member 1 with a mean GS of 15.84 μm (silt fraction) represents the sediments from the GMB and is very similar to the GSD from core GeoB 15641-1 (Fig. 3), the core with the highest percentage of silt and clay. End-member 2 is characterized by a small standard deviation of 1.89 μm representing well-sorted very fine sands from the outer shelf. End-members 3 and 4 are bimodal end-members with dominant peaks in the fine and medium sand fraction (Fig. 3), respectively. Both have a distinct side peak in the silt fraction. Most similar sediments can be found in the glaucony-sediment facies of profile M84/4-EM03. In particular, end-member 3 represents the sediments in the troughs and end-member 4 those on the crests of the seafloor undulations. The four end-members have distinctive electromagnetic properties. Sediment samples with high abundances of a particular end-member build well-defined clusters when represented in EC–MS cross-plots (Figs 7–10). This suggests that there is high correspondence between GS end-members and electromagnetic clusters.

4.2 Prediction of GSDs

4.2.1 Training and validation

Leave-one-out cross validation was performed to find the optimal regularization parameter and to evaluate the performance of end-member abundance prediction. Choosing successively one sample and predicting its end-member abundances from the RFBN which is trained using all the other samples provide a set of predicted data with the same size as the sample data set. The regularization parameter was tested using values between 0.1 and 1. The Frobenius norm of residuals between the measured and predicted data and the coefficient of determination were calculated from the leave-one-out experiment for each regularization parameter (Fig. 5). The lowest Frobenius norm of residuals and hence the best fit are achieved for a regularization parameter value of 0.2 (Fig. 5, left). Although a value of 0.3 results in a slightly higher Frobenius norm of residuals, this value was chosen because it maximizes the lowest coefficient of determination (R^2) of the four end-members (Fig. 5, right) and thus offers a more consistent prediction accuracy between the different end-members.

The leave-one-out experiment using the optimal regularization parameter of 0.3 reveals a high goodness-of-fit with R^2 values of 0.89, 0.78, 0.76 and 0.87 for end-members 1, 2, 3 and 4, respectively (Fig. 6). The fact that the highest R^2 values are achieved for end-

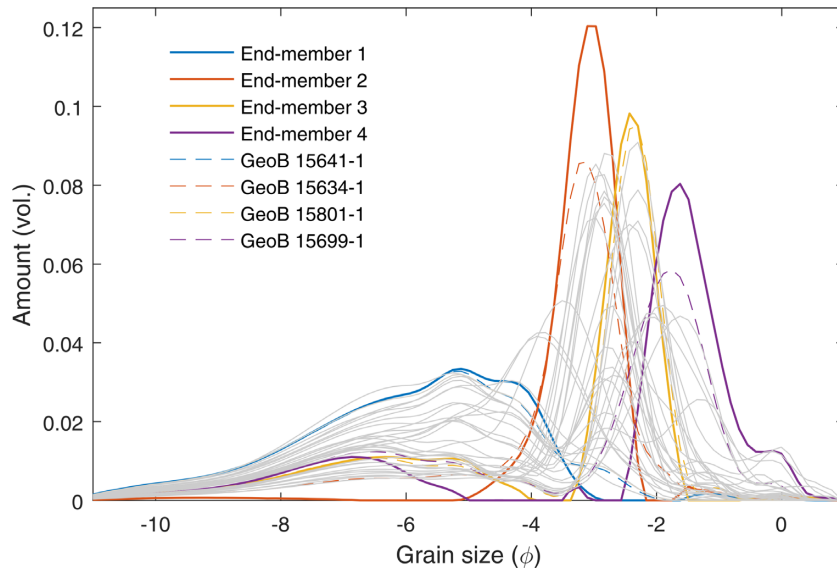


Figure 3. GSDs of the 33 surface samples (grey) and the four end-members (coloured solid line). The coloured dashed lines represent the GSDs of sediment samples most similar to the four end-members.

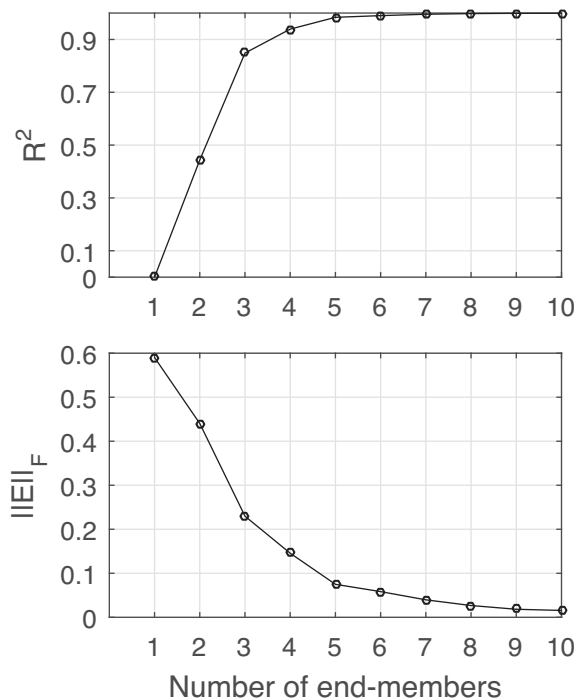


Figure 4. Coefficient of determination (top) and Frobenius norm of residuals (bottom) between measured GS and GS reconstructed from end-member modelling for different numbers of end-members.

member 1 and 4 confirms the results of former studies that EC and MS are highly sensitive to the mud content (represented by end-member 1) and MS is sensitive to the paramagnetic glaucony represented by end-member 4.

4.2.2 RBFN prediction of end-member abundances

Representing the RBFN-derived end-member abundances for each electromagnetic measurement in a EC–MS cross-plot (Figs 7–10)

shows that clusters with high loadings are situated at distinct segments. High abundance of end-member 1 corresponds to high EC and high MS (Fig. 7), associated sediment samples are from cores retrieved from the GMB. For MS lower than 300×10^{-6} , the end-member 1 abundances do not exceed a value of 0.3. The abundance of end-member 2 (Fig. 8) has the highest values for susceptibilities below 250×10^{-6} and conductivities between 0.6 and 1 S m^{-1} . All measurements within this EC–MS range originate from the outer shelf. High end-member 3 abundance corresponds with EC below 0.3 S m^{-1} and MS below 300×10^{-6} . End-member 4 shows similar electromagnetic characteristics with high abundances for low EC but higher MS (330×10^{-6} to 400×10^{-6}) than end-member 3. End-member 4 abundance is very low for EC higher than 0.6 S m^{-1} . The abundances of end-member 2 and 3 in sample 54 and 04 are very different (Figs 8 and 9), even though the EC and MS values are very similar. It is not clear why the electromagnetic properties do not reflect this GS variability, but inaccurate localization of one or both samples might be the reason. It can be seen that the RBFN interpolation predicts smoother values than the underlying samples suggest, this helps to produce spatially consistent models and avoids overfitting in areas where the electromagnetic data do not fully explain the GS variability.

4.2.3 Spatial distribution of predicted end-members and grain sizes

The east–west profiles M84/4-EM01 (Fig. 11) and M84/4-EM02 (Fig. 12) both cross the Galician Mud Belt between 495 000 m to 500 000 m East. In this area, EC and MS reach the highest values. The high amount of mud in this area is reflected in high end-member 1 contributions of up to over 90 per cent. Both profiles show a GS increase from the western edge of the mud belt towards the inner shelf which is related to a stronger wave impact towards the shore (Lantzsck *et al.* 2009a). Correspondingly, the abundance of end-member 1 decreases while the end-member 2 abundance increases. This trend is stronger for profile M84/4-EM02 which could be linked to the steeper shoreward rise of the water bottom.

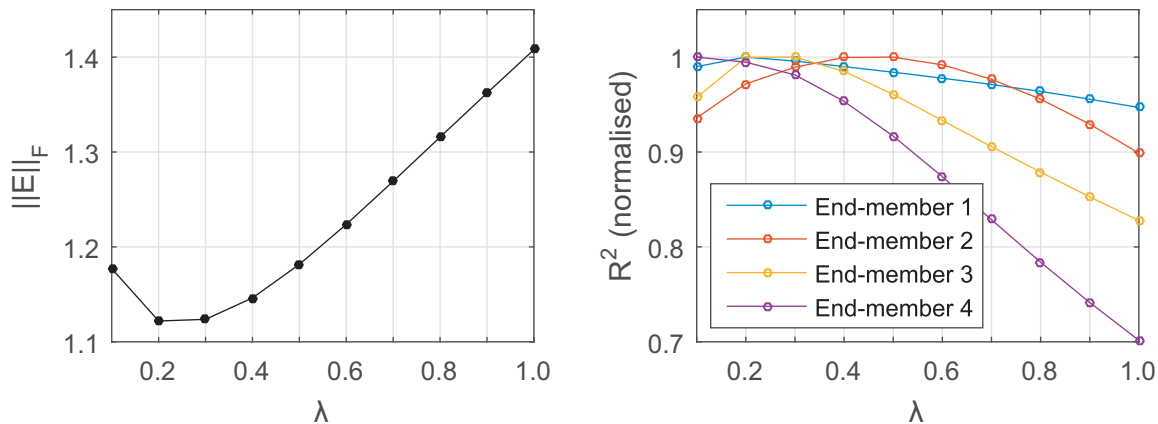


Figure 5. Frobenius norm of residuals between measured and predicted end-member abundances (left) and coefficient of determination for each end-member (R^2 normalized to self-peak, right) versus regularization parameter λ .

Table 1. Grain-size statistics of the four end-members.

	End-member 1	End-member 2	End-member 3	End-member 4
Mode (μm)	28.70	116.27	185.35	324.40
Mean (μm)	15.84	108.08	74.71	141.45
Median (μm)	18.00	111.1	153.83	295.53
Standard deviation (μm)	3.43	1.89	4.64	6.23

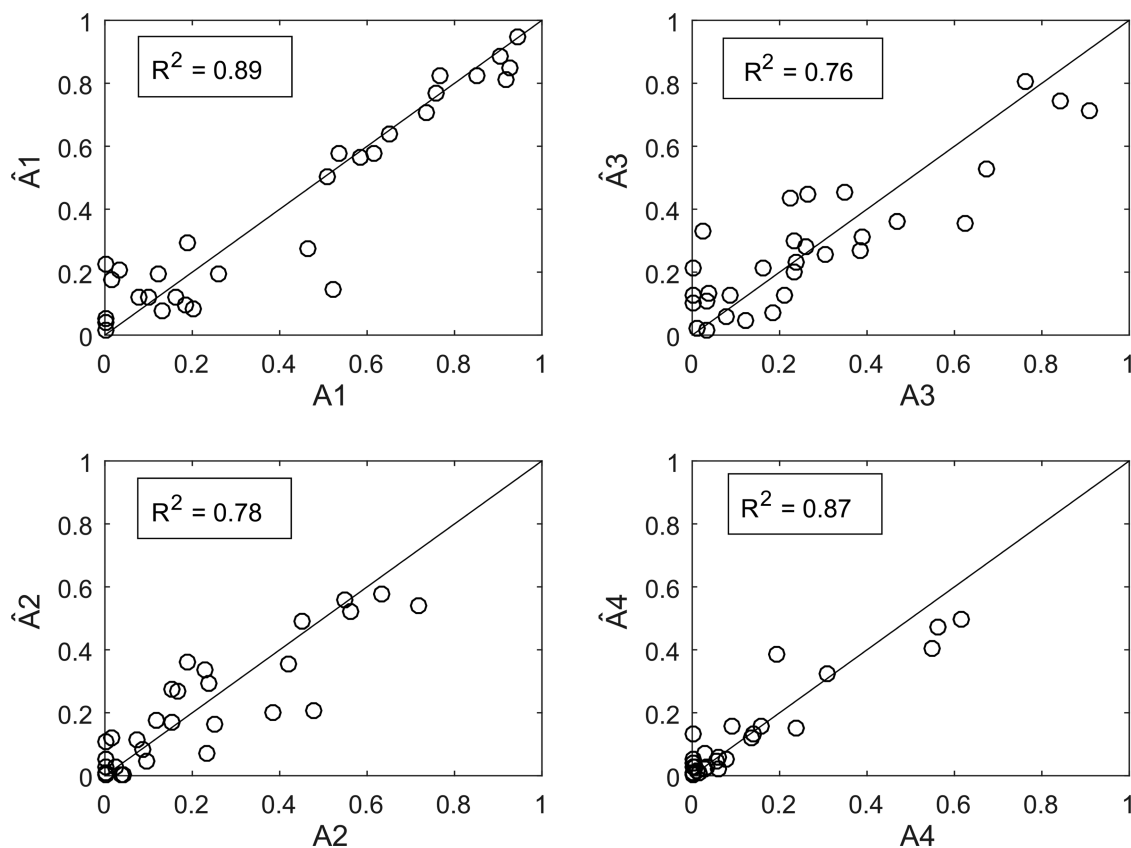


Figure 6. Comparison of end-member abundances (\hat{A}) derived from the leave-one-out experiment and measured end-member abundances (A) of the four different end-members.

The highest accumulation of mud occurs at the break of slope between the steeper inner shelf and flatter mid shelf on both profiles. In western direction the mud belt is followed by an area which is characterized by sediment dunes and undulating EC and MS values.

The dune crests are represented by EC lows and MS highs, while the troughs show high EC and lower MS. This pattern has been earlier explained by local accumulations of mud in troughs (Baasch *et al.* 2015) and relatively higher paramagnetic glaucony content

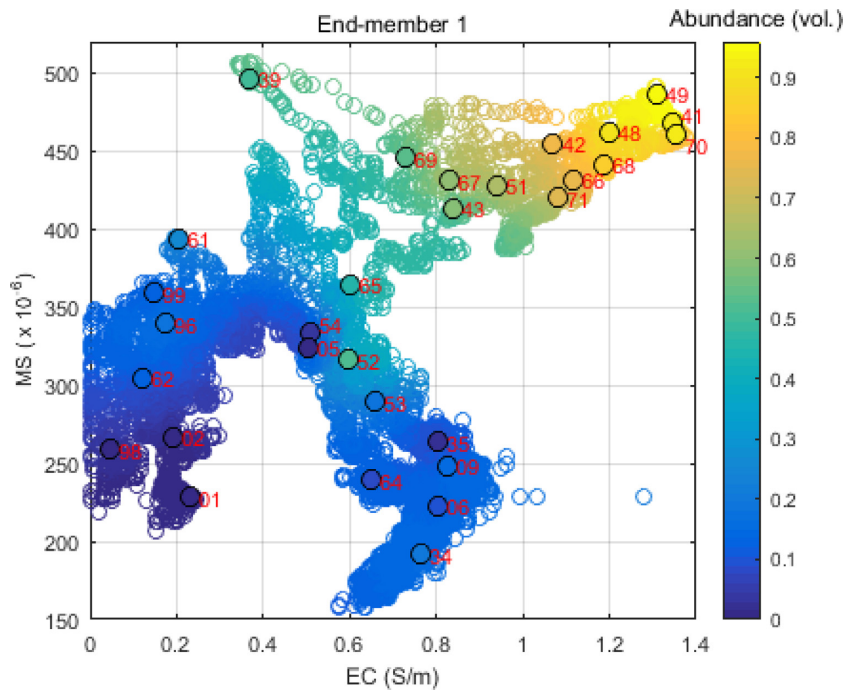


Figure 7. Cross plot, apparent MS versus apparent EC. Colour of circles indicates the loading of end-member 1. Filled circles indicate the end-member loadings from the training data (sediment-core samples). Red numbers show the last two digits of the GeoB-core number.

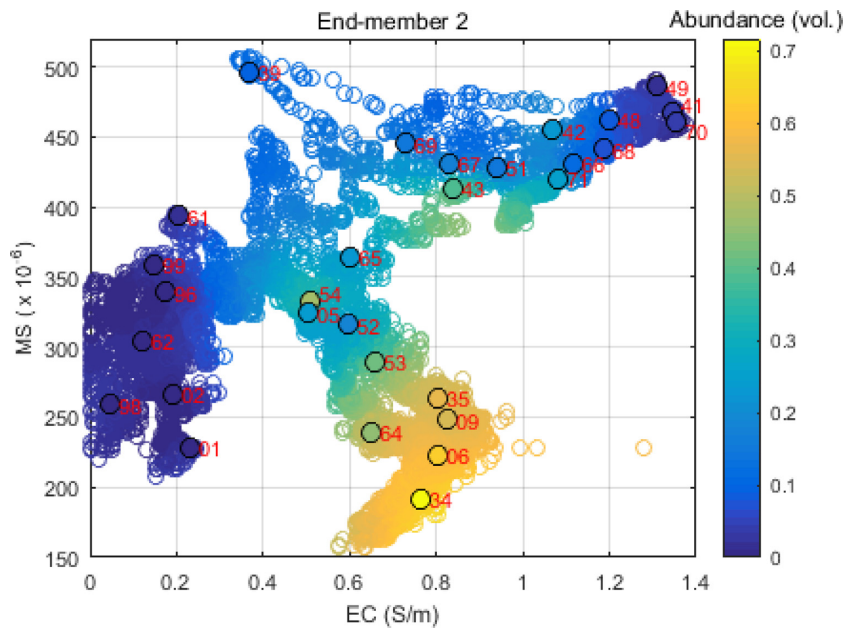


Figure 8. Cross plot, apparent MS versus apparent EC. Colour of circles indicates the loading of end-member 2. Filled circles indicate the end-member loadings from the training data (sediment-core samples). Red numbers show the last two digits of the GeoB-core number.

and higher compaction on crests (Müller *et al.* 2012). The predicted end-member abundances confirm this relationship between seabed topography, glaucony sand and mud abundance. End-members 1 and 2 peak out in troughs and on crests, respectively. The GSDs are generally wide in this area, reflecting the sediment variety of this transition zone between mid-shelf mud belt and outer shelf. Towards the outer shelf GS and sorting of the trough’s infill sediments increase. On the outer shelf end-member 2 dominates with contributions of over 60 per cent and hence GSDs are narrow representing highly sorted very fine to fine sands.

Profile M84/4-EM03 (Fig. 13) runs in a north–south direction on the outer shelf parallel to the coast line. In the south, it is crosscut by the two other profiles. The predictive GSDs are very similar between the three profiles in the area where they overlap. Towards the north profile M84/4-EM03 passes a canyon. From the deepest point of the canyon towards the north, the sedimentology changes distinctly. Abundance of end-member 4 increases and end-member 2 vanishes. In general, the main mode of the GS increases from south to north. Additionally, sediments change in north direction from a narrow unimodal distribution to a bimodal distribution with a side

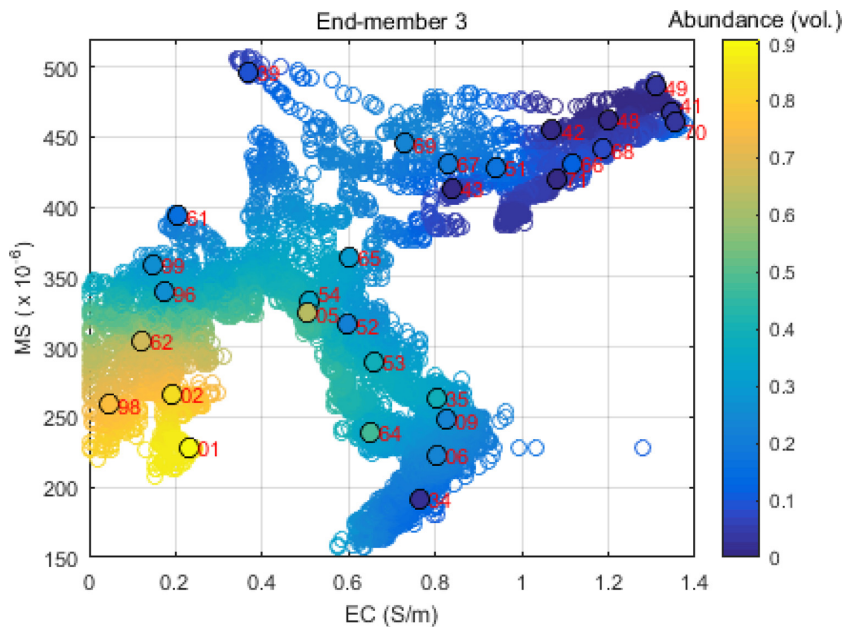


Figure 9. Cross plot, apparent MS versus apparent EC. Colour of circles indicates the loading of end-member 3. Filled circles indicate the end-member loadings from the training data (sediment-core samples). Red numbers show the last two digits of the GeoB-core number.

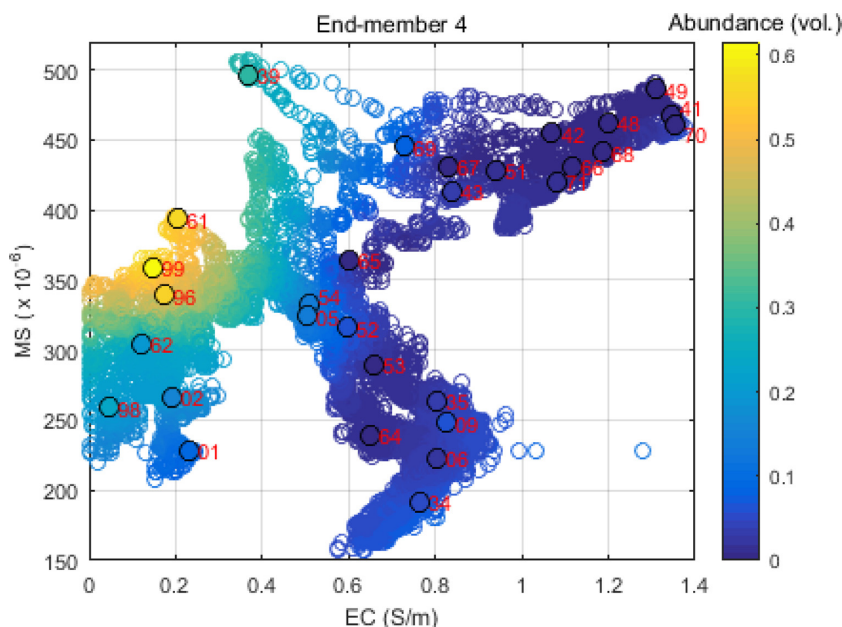


Figure 10. Cross plot, apparent MS versus apparent EC. Colour of circles indicates the loading of end-member 4. Filled circles indicate the end-member loadings from the training data (sediment-core samples). Red numbers show the last two digits of the GeoB-core number.

peak in the silt fraction. The northern end of the profile is characterized by rough seabed topography. Sediments at morphologic highs have a higher proportion of end-member 4 (up to 60 per cent) and morphologic lows have a higher proportion of end-member 3 (up to 90 per cent). The accumulation of end-member 4 at more exposed locations corresponds to the general assumption that glaucony sands (mainly represented by end-member 4) are indicative for non-accumulating or erosional systems related to strong hydrodynamic forces.

5 DISCUSSION

5.1 Electromagnetic profiling for sediment characterization

Electromagnetic profiling as sediment characterization tool has distinct advantages compared to other methods. It is a fast non- or minimal invasive technique which offers high spatial resolution and can be operated from small to large vessels. Although this is also true for acoustic methods, we showed that electromagnetic methods, unlike acoustic methods, can recover earth models with physical (electromagnetic) sediment properties through inversion.

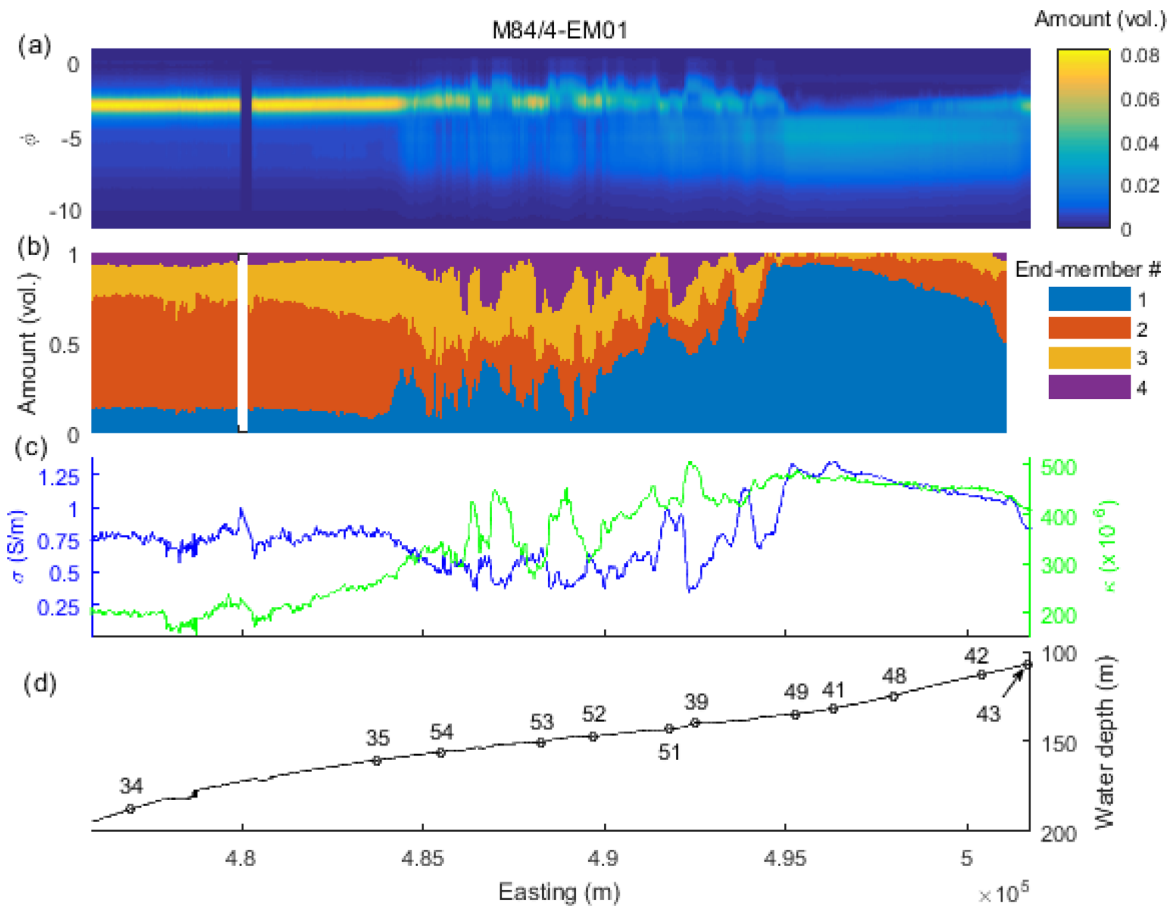


Figure 11. RBFN prediction results together with electromagnetic and bathymetric data along electromagnetic profile M84/4-EM01: (a) Predicted GSD, (b) predicted GSD end-member abundances, (c) EC (blue line) and MS (green line), (d) water depth, dots indicate surface sample locations and numbers indicate the two last digits of the GeoB-core number.

In contrast, acoustic methods measure backscatter features of the seabed which are not only influenced by sediment properties but also by the (micro) morphology of the seabed.

EC–MS cross-plots revealed that different sediment types are represented by distinct clusters and hence enable an electromagnetic-based sediment classification. These clusters well reflect GS-based sediment types of this and former studies indicating a strong link between electromagnetic and textural sediment properties. The electromagnetic profiling-based sediment classification can be used to define locations for sediment sampling and further investigations. In this study, the strong relationship between electromagnetic and textural properties can be mainly attributed to the sensitivity of EC and MS to mud content, the sensitivity of EC to sediment sorting and the sensitivity of MS to (coarse) paramagnetic glaucony minerals. Although EC and MS have been used in many different environments as particle size proxy, the electromagnetic signal cannot always be related to GS. In areas where no magnetic minerals are present, for example, coastal areas with only quartz-bearing and carbonate sediments, there might be only a weak or no relationship between MS and GS. Additionally, in areas with fresh-water or gas seepages the relationship between EC and GS might be masked by the effect of variable EC of the pore-fluid.

One constraint of the electromagnetic data presented here is that a homogenous half-space is assumed in the inversion algorithm to recover EC and MS. Thus, lateral and vertical sediment changes within

the electromagnetic sensor's footprint affect the reconstructed apparent EC and apparent MS. Consequently, the electromagnetic data are influenced by sediments of a certain volume. The size of this volume depends on the electromagnetic properties of the sediment and frequencies employed. Müller *et al.* (2012) specified the footprint of the EC-related 5025 Hz quadrature signal and MS-related 75 Hz in-phase signal to 3.5–5.5 m and 1 m, respectively, and the depth from which 90 per cent of the EC and MS information comes from to 92 cm and 50 cm, respectively. Comparing the electromagnetic data with sediment samples therefore means comparing bulk volume data with point data. If the sediments have a high vertical or lateral variability, the samples might represent a fraction untypical for the volume illuminated by the electromagnetic sensor.

Another constraint is the localization accuracy of the profiler and sampling devices to enable ground truthing at correct locations. Deviations between the profiler track and ground-truth locations naturally increase with current strength and water depth. A meaningful comparison of electromagnetic and ground-truth data is only feasible if the spatial sediment variability is small compared to the electromagnetic systems' footprint and positioning errors of the profiler and sampling device. As with other remote sensing techniques for seafloor characterization, the success of the electromagnetic method depends on a sufficient number of ground-truth data, which reflect the variety of sediment types and minimize statistical errors. Since the profiler is towed on the seafloor, it is exposed to natural

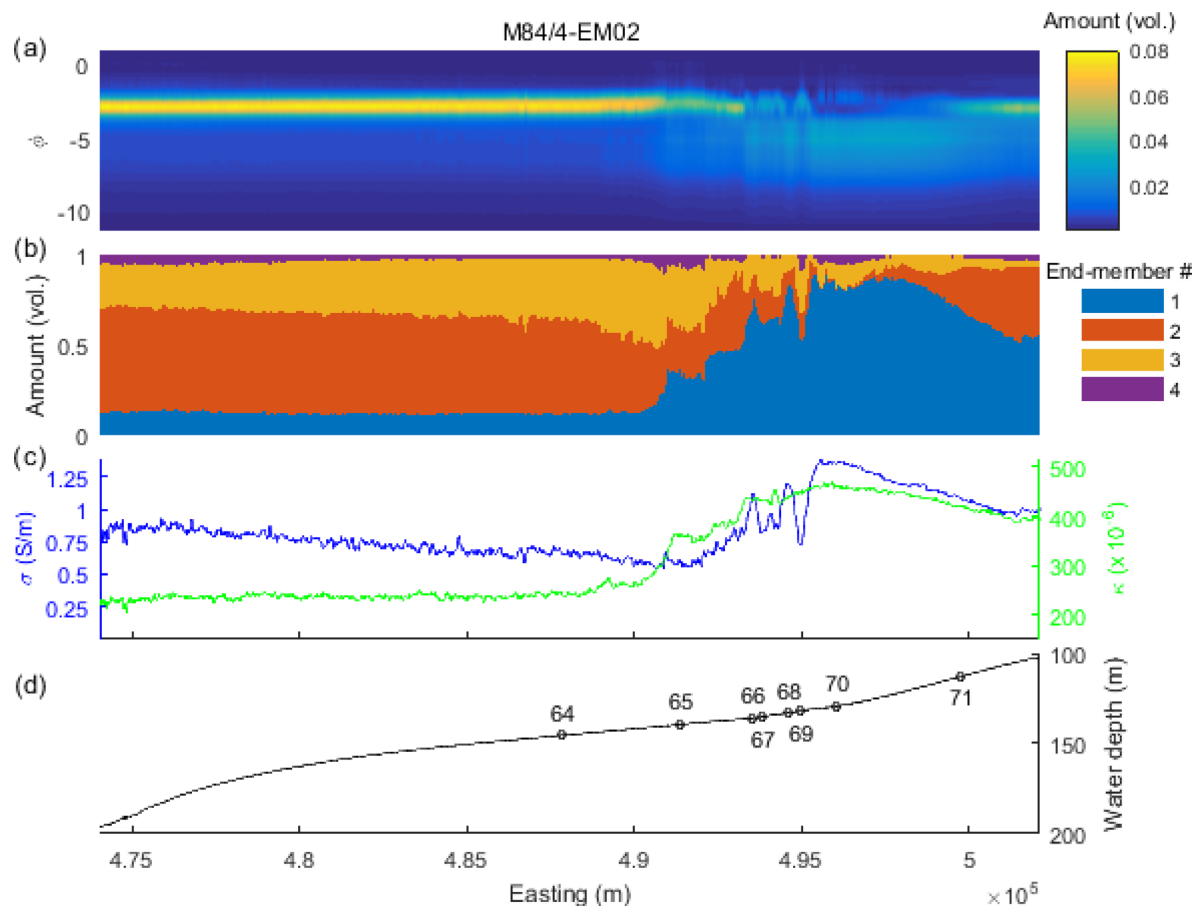


Figure 12. RBFN prediction results together with electromagnetic and bathymetric data along electromagnetic profile M84/4-EM02: (a) Predicted GSD, (b) predicted GSD end-member abundances, (c) EC (blue line) and MS (green line), (d) water depth, dots indicate surface sample locations and numbers indicate the two last digits of the GeOB-core number.

and artificial obstacles, which limits the application of the profiler in areas with rocky outcrops or man-made seafloor installations.

5.2 Prediction of compositional data

A GSD is a classic example of compositional data, meaning that no single GS fraction is free to vary separately from the rest of the total composition. This closure effect can cause misleading results if fractions of the GSD are predicted independently. The best solution is to predict the entire GSD. However, prediction of all GS classes (92 in this study) is not only computationally expensive, it is also numerically difficult because the abundances of the GS classes at both tails of the distribution are very low or zero for the majority of the sediment samples. Additionally, not all GS classes are representative of certain sediment facies but rather have quasi constant abundances across all different sediment types. To overcome this problem we used an NMF algorithm to reduce the predicted parameters from 92 GS classes to four end-members with the goal to predict the abundances of these end-members rather than frequencies of the GS classes directly. Other low-rank factorizations, such as singular value decomposition or principal component analysis have not been tested here but could also be used for parameter reduction and might offer better results under certain conditions. It is worth mentioning that these techniques might require preconditioning of the compositional data in form of, for example, log-ratio transforms.

The main advantage of the end-member modelling is that it produces end-members which are directly interpretable. We showed that the NMF-derived end-members represent natural GSD end-members of the NW Iberian shelf sediments. The degree to which an end-member contributes to each sample offers valuable information for the sediment characterization and can be used to define sediment patterns and corresponding transition/mixing zones. The spatial distribution of the end-member contributions can be used to create GS-based facies maps. Furthermore, the end-members and its abundances carry all GS information and can be back transformed to GS classes to reconstruct complete GSDs or to recover other statistical measures such as mean GS, sorting, skewness, number of modes, etc. Just as the GSD, the GS end-members are compositional data and thus their abundances are always positive and sum to one. These two inherent characteristics of compositional data need to be taken care of during the RBFN prediction. Since a conventional RBFN would violate both constraints, we imposed positivity and the sum-to-one constraint of the end-member abundances in the RBFN learning algorithm. This means that RBFN predictions of all end-members need to be trained together.

5.3 Spatial constraining of the RBFN

In regression analysis, various techniques exist to encounter overfitting attributed to noise in the data. Smoothing through regularization is one of the most popular techniques to avoid overfitting (Girosi

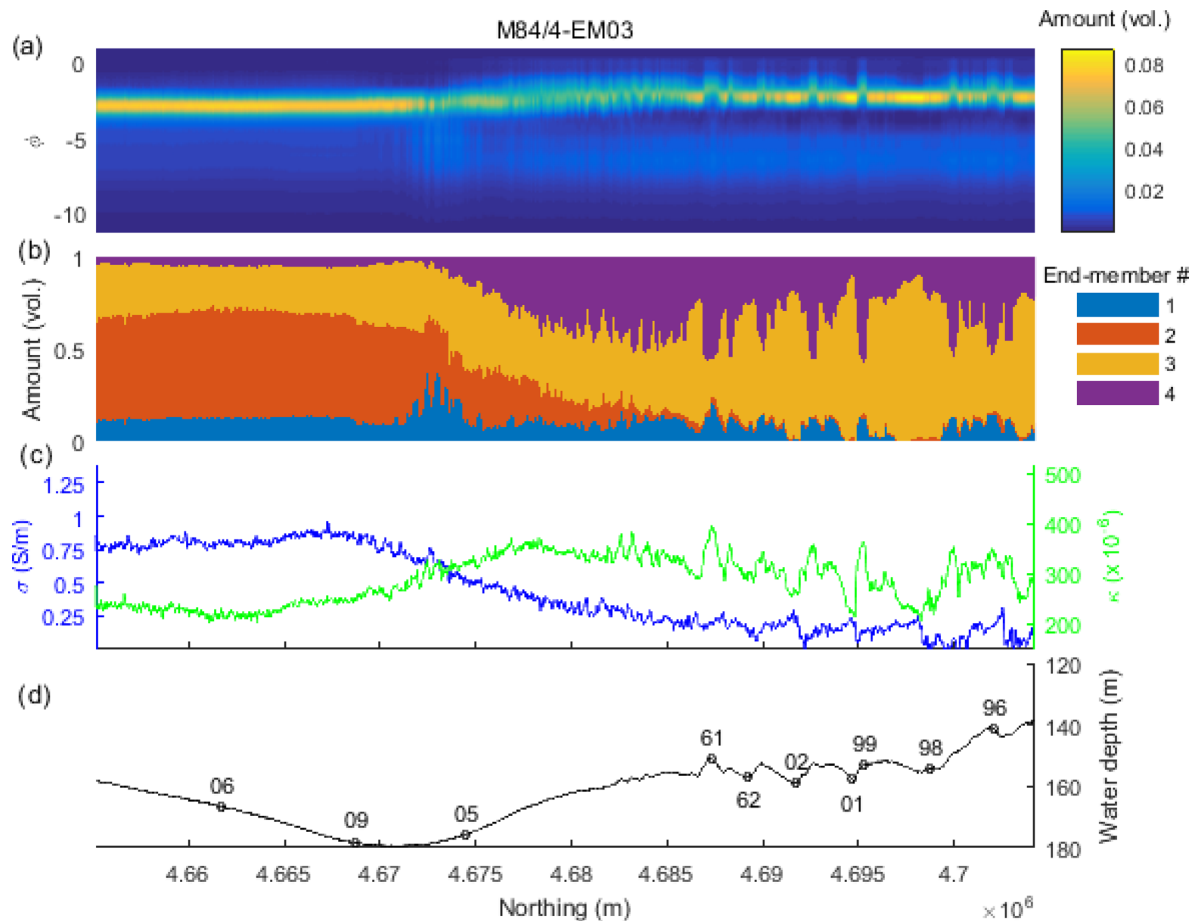


Figure 13. RBFN prediction results together with electromagnetic and bathymetric data along electromagnetic profile M84/4-EM03: (a) Predicted GSD, (b) predicted GSD end-member abundances, (c) EC (blue line) and MS (green line), (d) water depth, dots indicate surface sample locations and numbers indicate the two last digits of the GeoB-core number.

et al. 1995). Regularization can be performed on the model parameters namely the weights in the RBFN. This approach is used in, for example, support vector regression where the norm of the weights is minimized together with the estimation error. Regularization can also be performed on the output variables of the predictive model. In both cases, the regularization is conventionally implemented in the domain of the input variables (the feature space). Thus, output variables with similar input variables are constrained to be similar as well. This is a reasonable assumption in many applications.

In this paper, we used a different regularization approach, namely a minimum lateral variation constraint. It constrains the variability of the output variables (here the predicted GS end-member abundances) of neighbouring measurement points along the electromagnetic profile. This approach has the advantage that it respects spatial correlations between the output data. Hence, it is not prone to instrument-related acquisition footprints.

Another more straightforward way of incorporating spatial data in the RBFN training would be to include the spatial coordinates as predictor variables. In this case, no distinction is made between the spatial domain and the geophysical attributes pertaining to it. This would constrain the RBFN to learn from a limited range of coordinates simply because the RBFs of the hidden nodes produce a larger output when the input signal is close to the centre of the RBF. The problem is the 2-D nature of the electromagnetic profiling data acquisition resulting in very high spatial density along the profiles and a rather sparse density in cross-profile direction. This would

affect the RBFN to produce models biased towards values along the profiles and hence would be similar to produce separate models for each profile.

In fact, incorporating spatial coordinates as input data requires that the training samples are randomly distributed within the spatial domain over the entire survey area (Gahegan 2000; Cracknell & Reading 2014), which is rarely the case. In contrast, our approach considers the centres and hence the sediment samples from all profiles equally because the Euclidean distance in the RBF is solely calculated from the electromagnetic properties rather than spatial coordinates. This enables the RBFN to predict the output variables in regions spatially disjoint from the training samples and thus minimizes the total amount of sediment samples to be taken.

The drawback of the lateral variation constraint is that the matrix Φ_d in eq. (10), which connects all electromagnetic measurement points to each centre, can become very large and hence its computation and memory requirements can get expensive. In such a case, we recommend splitting the profile into segments of reasonable size.

6 CONCLUSION

We have presented a new methodology to predict seafloor GSDs from geophysical data. An unsupervised end-member modelling approach based on non-negative matrix factorization is first used to reduce the number of target GS parameters from typically 92 GS

classes to a minimum number of end-members. The end-member abundances are then predicted employing an RBFN, which is trained using the geophysical data and end-member abundances at the sample location as input and target variables, respectively. The RBFN predicts the abundances for all end-member simultaneously enabling to incorporate a sum-to-constant constraint on the abundances. This approach respects the closure of the GSD, which is a major advantage compared to previous studies that try to predict fractions of the GSD separately. In addition, the objective function to be minimized in the RBFN training includes a constraint penalizing lateral variation between neighbouring predicted end-member abundances. This constraint allows a higher variability in the feature domain while avoiding overfitting by incorporating spatial information. Not using this spatial information as explicit feature makes this approach preferable for spatially highly clustered sample data. The outputs of the RBFN, namely the predicted end-member abundances, can finally be back transformed to GSD.

The methodology was tested and validated using electromagnetic profiling data acquired on the NW Iberian shelf. Electromagnetic profiling is a new but established seafloor mapping technique. The electromagnetic data can be inverted for EC and MS of the seabed substrata, which both have proven to correlate with GS properties.

We have demonstrated that electromagnetic profiling in combination with machine-learning approaches is capable to predict GSDs with high accuracy and resolution along shelf-wide survey lines. This study suggests that electromagnetic benthic profiling in addition to acoustic methods should play a larger role for seabed characterization. However, our presented machine-learning approach can be readily adapted to other geospatial problems that involve predictive modelling of compositional data from geophysical or remote sensing surveys. Therefore, this study should also have an impact on, for example, acoustic seabed characterization and benthic habitat mapping, mineral exploration and characterization of spatial soil variability in precision agriculture.

ACKNOWLEDGEMENTS

This work was funded through DFG Research Center/Cluster of Excellence 'The Ocean in the Earth System' and was part of MARUM Research Area SD. A preliminary version of this paper formed a chapter of the PhD thesis of BB, which although unpublished, is available at <http://nbn-resolving.de/urn:nbn:de:gbv:46-00105671-12>. We would like to thank captain and crew of RV Meteor cruise M84/4. Finally, we would also like to thank the two anonymous reviewers and the editor for their helpful comments and suggestions that improved this manuscript.

REFERENCES

- Archie, G., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics, *Pet. Trans. AIME*, **146**, 54–62.
- Baasch, B., Müller, H., Oberle, F.K.J. & von Dobeneck, T., 2015. Inversion of marine multifrequency electromagnetic profiling data: a new approach to resolve surficial sediment stratification, *Geophys. J. Int.*, **200**(1), 438–451.
- Baasch, B., Müller, H., von Dobeneck, T. & Oberle, F.K., 2017. Determination of grain-size characteristics from electromagnetic seabed mapping data: a NW Iberian shelf study, *Cont. Shelf Res.*, **140**(Supplement C), 75–83.
- Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P. & Plemmons, R.J., 2006. Algorithms and applications for approximate nonnegative matrix factorization, *Comput. Stat. Data Anal.*, **52**, 155–173.
- Billings, S.D., Newsam, G.N. & Beatson, R.K., 2002. Smooth fitting of geophysical data using continuous global surfaces, *Geophysics*, **67**(6), 1823–1834.
- Booth, C., Walden, J., Neal, A. & Smith, J., 2005. Use of mineral magnetic concentration data as a particle size proxy: a case study using marine, estuarine and fluvial sediments in the Carmarthen Bay area, South Wales, U.K., *Sci. Total Environ.*, **347**(1–3), 241–253.
- Broomhead, D.S. & Lowe, D., 1988. Radial basis functions, multi-variable functional interpolation and adaptive networks, Royal Signals & Radar Establishment, Memorandum no. 4148, Malvern, Worcester, Great Britain.
- Brown, C.J., Smith, S.J., Lawton, P. & Anderson, J.T., 2011. Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques, *Estuarine Coast. Shelf Sci.*, **92**(3), 502–520.
- Butler, K.E., 2009. Trends in waterborne electrical and em induction methods for high resolution sub-bottom imaging, *Near Surf. Geophys.*, **7** (4), 241–246.
- Cheesman, S.J., Edwards, R.N. & Law, L.K., 1990. A test of a short-baseline sea-floor transient electromagnetic system, *Geophys. J. Int.*, **103**(2), 431–437.
- Cracknell, M.J. & Reading, A.M., 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information, *Comput. Geosci.*, **63**, 22–33.
- De, C. & Chakraborty, B., 2009. Acoustic characterization of seafloor sediment employing a hybrid method of neural network architecture and fuzzy algorithm, *IEEE Geosci. Remote Sens. Lett.*, **6**(4), 743–747.
- Dias, J.M.A. *et al.*, 2002. Present day sedimentary processes on the northern Iberian shelf, *Prog. Oceanogr.*, **52**(2–4), 249–259.
- Diesing, M. & Stephens, D., 2015. A multi-model ensemble approach to seabed mapping, *J. Sea Res.*, **100**, 62–69.
- Eberle, D., Hutchins, D., Das, S., Majumdar, A. & Paasche, H., 2015. Automated pattern recognition to support geological mapping and exploration target generation - a case study from southern Namibia, *J. Afr. Earth Sci.*, **106**, 60–74.
- Evans, R.L., 2001. Measuring the shallow porosity structure of sediments on the continental shelf: a comparison of an electromagnetic approach with cores and acoustic backscatter, *J. geophys. Res.*, **106**(C11), 27047–27060.
- Gahegan, M., 2000. On the application of inductive machine learning tools to geographical analysis, *Geogr. Analysis*, **32**(2), 113–139.
- Girosi, F., Jones, M. & Poggio, T., 1995. Regularization theory and neural networks architectures, *Neural Comput.*, **7**, 219–269.
- Heil, K. & Schmidhalter, U., 2012. Characterisation of soil texture variability using the apparent soil electrical conductivity at a highly variable site, *Comput. Geosci.*, **39**, 98–110.
- Huang, Z., Nichol, S.L., Siwabessy, J.P., Daniell, J. & Brooke, B.P., 2012. Predictive modelling of seabed sediment parameters using multibeam acoustic data: a case study on the carnarvon shelf, Western Australia, *Int. J. Geogr. Inf. Sci.*, **26**(2), 283–307.
- Huang, Z., Siwabessy, J., Nichol, S., Anderson, T. & Brooke, B., 2013. Predictive mapping of seabed cover types using angular response curves of multibeam backscatter data: testing different feature analysis approaches, *Conti. Shelf Res.*, **61–62**, 12–22.
- Huang, Z., Siwabessy, J., Nichol, S.L. & Brooke, B.P., 2014. Predictive mapping of seabed substrata using high-resolution multibeam sonar data: a case study from a shelf with complex geomorphology, *Mar. Geol.*, **357**, 37–52.
- Jackson, D.R. & Briggs, K.B., 1992. High-frequency bottom backscattering: roughness versus sediment volume scattering, *J. acoust. Soc. Am.*, **92**(2), 962–977.
- Jackson, P.D., Smith, D.T. & Stanford, P.N., 1978. Resistivity-porosity-particle shape relationships for marine sands, *Geophysics*, **43**(6), 1250–1268.
- Just, J., Heslop, D., von Dobeneck, T., Bickert, T., Dekkers, M.J., Frederichs, T., Meyer, I. & Zabel, M., 2012. Multiproxy characterization and budgeting of terrigenous end-members at the NW African

- continental margin, *Geochem., Geophys., Geosyst.*, **13**(9), Q0A001, doi:10.1029/2012GC004146.
- Lantzsch, H., Hanebuth, T.J. & Bender, V.B., 2009a. Holocene evolution of mud depocentres on a high-energy, low-accumulation shelf (NW Iberia), *Quat. Res.*, **72**(3), 325–336.
- Lantzsch, H., Hanebuth, T.J., Bender, V.B. & Krastel, S., 2009b. Sedimentary architecture of a low-accumulation shelf since the Late Pleistocene (NW Iberia), *Mar. Geol.*, **259**(1–4), 47–58.
- Lantzsch, H., Hanebuth, T.J. & Henrich, R., 2010. Sediment recycling and adjustment of deposition during deglacial drowning of a low-accumulation shelf (NW Iberia), *Conti. Shelf Res.*, **30**(15), 1665–1679.
- Lawson, C.L. & Hanson, R.J., 1995. Chap. 23: Linear least squares with linear inequality constraints, in *Solving Least Squares Problems, Classics in Applied Mathematics*, pp. 158–173, SIAM.
- Müller, H., von Dobeneck, T., Hilgenfeldt, C., SanFilipo, B., Rey, D. & Rubio, B., 2012. Mapping the magnetic susceptibility and electric conductivity of marine surficial sediments by benthic EM profiling, *Geophysics*, **77**(1), E43–E56.
- Oberle, F.K., Hanebuth, T.J., Baasch, B. & Schwenk, T., 2014a. Volumetric budget calculation of sediment and carbon storage and export for a late holocene mid-shelf mudbelt system (NW Iberia), *Cont. Shelf Res.*, **76**, 12–24.
- Oberle, F.K., Storlazzi, C.D. & Hanebuth, T.J., 2014b. Wave-driven sediment mobilization on a storm-controlled continental shelf (Northwest Iberia), *J. Mar. Syst.*, **139**, 362–372.
- Oberle, F.K., Storlazzi, C.D. & Hanebuth, T.J., 2016a. What a drag: quantifying the global impact of chronic bottom trawling on continental shelf sediment, *J. Mar. Syst.*, **159**, 109–119.
- Oberle, F.K., Swarzenski, P.W., Reddy, C.M., Nelson, R.K., Baasch, B. & Hanebuth, T.J., 2016b. Deciphering the lithological consequences of bottom trawling to sedimentary habitats on the shelf, *J. Mar. Syst.*, **159**, 120–131.
- Odin, G. & Lambooy, M., 1988. Glaucony from the margin off northwestern Spain, in *Green Marine Clays*, pp. 249–275, ed. Odin, G., Elsevier.
- Oldfield, F., Maher, B.A., Donoghue, J. & Pierce, J., 1985. Particle-size related, mineral magnetic source sediment linkages in the Rhode River catchment, Maryland, USA, *J. Geol. Soc.*, **142**(6), 1035–1046.
- Oliveira, A., Rocha, F., Rodrigues, A., Jouanneau, J., Dias, A., Weber, O. & Gomes, C., 2002. Clay minerals from the sedimentary cover from the Northwest Iberian shelf, *Prog. Oceanogr.*, **52**(2–4), 233–247.
- Paasche, H., Tronicke, J., Holliger, K., Green, A.G. & Maurer, H., 2006. Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses, *Geophysics*, **71**(3), H33–H44.
- Paatero, P. & Tapper, U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, **5**(2), 111–126.
- Schwalenberg, K., Haeckel, M., Poort, J. & Jegen, M., 2010. Evaluation of gas hydrate deposits in an active seep area using marine controlled source electromagnetics: results from Opouawe Bank, Hikurangi Margin, New Zealand, *Mar. Geol.*, **272**(1–4), 79–88.
- Swidinsky, A., Hölz, S. & Jegen, M., 2015. Rapid resistivity imaging for marine controlled-source electromagnetic surveys with two transmitter polarizations: an application to the North Alex mud volcano, West Nile Delta, *Geophysics*, **80**(2), E97–E110.
- Won, I., Keiswetter, D., Hanson, D., Novikova, E. & Hall, T., 1997. Gem-3: a monostatic broadband electromagnetic induction sensor, *J. Environ. Eng. Geophys.*, **2**(1), 53–64.