Ann. For. Res. 65 (2): 103-118, 2022
https://doi.org/10.15287/afr.2022.2682

ANNALS OF FOREST RESEARCH
www.afrjournal.org

How accurate is the remote sensing based estimate of water physico-chemical parameters in the Danube Delta (Romania)?

Marian Necula[1], Iris Maria Tușa[2], Manuela Elisabeta Sidoroff [2], Corina Iţcuș[2], Daniela Florea[2], Alexandru Amărioarei[2,3], Andrei Păun[2,3], Octavian Pacioglu[2], Mihaela Păun [2,4]

Necula N., Tușa I.M., Sidoroff M.E., Iţcuș C., Florea D., Amărioarei A., Păun A., Pacioglu O., Păun M., 2022. How accurate is the remote sensing based estimate of water physico-chemical parameters in the Danube Delta (Romania)? Ann. For. Res. 65(2): 103-118.

Abstract The current paper estimated the physico-chemical properties of water in the Danube Delta (Romania), based on Sentinel 2 remote sensing data. Eleven sites from the Danube Delta were sampled in spring and autumn for three years (2018-2020) and 21 water physico-chemical parameters were measured in laboratory. Several families of machine learning algorithms, translated into hundreds of models with different parameterizations for each machine learning algorithm, based on remote sensing data input from Sentinel 2 spectral bands, were employed to find the best models that predicted the values measured in laboratory. This was a novel approach, reflected in the types of selected models that minimised the values of performance metrics for the tested parameters. For alkalinity, calcium, chloride, carbon dioxide, hardness, potassium, sodium, ammonium, dissolved oxygen, sulphates, and suspended matter the results were promising, with an overall percentage bias of the estimates of +/- 10% from the observed values. For copper, magnesium, nitrites, nitrates, turbidity and zinc the estimates were fairly accurate, with percentage biases in the interval +/- 10% - 20%, whereas for detergents, led, and phosphates the percentage bias was higher than 20%. Overall, the results of the current study showed fairly good estimates between remote sensing based estimates and laboratory measured values for most water physico-chemical parameters.

Keywords: Danube Delta; water quality; remote sensing; Sentinel 2.

Addresses: [1]The Bucharest University of Economic Studies, Romania| [2]National Institute of Research and Development for Biological Sciences, Bucharest, Romania| [3]University of Bucharest, Faculty of Mathematics and Computer Science, Bucharest, Romania| [4]University of Bucharest, Faculty of Business and Administration, Bucharest, Romania.

Corresponding Authors: Marian Necula (neculamarian18@stud.ase.ro); Mihaela Păun (mihaela.paun@incdsb.ro); Octavian Pacioglu (octavian.pacioglu@incdsb.ro).

Manuscript: received November 11, 2022; revised December 07, 2022; accepted December 29, 2022.

Introduction

The European Union created a legislative context, the Water Directive Framework (European Parliament and European Council 2000), in which all member states are required to have an up-to-date assessment of the quality of inland and coastal waters. The water quality is usually assessed through on a number of physico-chemical parameters, such as the concentrations of chlorophyll-a, dissolved oxygen, nitrogen (e.g. nitrites, nitrates, and ammonia) and phosphorus (e.g. phosphate) chemical species, to name just the most common measured parameters during monitoring programs (Ielenicz & Comanescu 2006). Nevertheless, some major issues that the traditional monitoring programs are faced with comprise the high costs associated with field work, time consuming and the difficulties of physical access to the sampling sites. Moreover, another crucial aspect that the monitoring sampling programs of water quality are faced with is the reliability of the results *per se* (Dumitru et al. 2019). Such parameters often characterize the water quality at the local level, hence faced with the issue of spatial representativeness, especially for extensive and diverse types of aquatic ecosystems (Niculescu et al. 2017).

Given these drawbacks, the implementation of real-time monitoring program through traditional methods in remote and extended areas, such as the deltas and estuaries, is not always feasible. Therefore, a different type of complementary tools emerged during the past decades, the remote sensing techniques, which became an important instrument in assessing the water quality of various lotic and lentic habitats (Oteman et al. 2021). The advantages of using such complementary techniques are obvious: reduced costs and increased spatial and temporal representativeness, by covering larger areas more frequently compared with traditional surveys (Rahul et al. 2022). The main groups of water quality proxy parameters estimated using remote sensing data (i.e. optically active water constituents) are total suspended solids, coloured dissolved organic matter and water clarity, but also the optically inactive physico-chemical parameters (Topp et al 2019). The main technical challenges that remote sensing techniques application in aquatic habitats are faced with nowadays are the atmospheric and adjacency corrections, pre-processing steps, inversion based retrieval methods and optical water type classification (Palmer et al 2015). However, the application of remote sensing techniques outside the field of aquatic ecology covers a wide range of domains, such as agriculture, forestry, urban areas, biogeochemical parameters and sea ice levels (EU Commission, 2022a), as well as atmosphere products on pollution and gases (EU Commission, 2022c).

The Danube River is the second largest river in Europe, with a total length of 2860 km and a total mean annual discharge of 6500 m^3s^{-1} (Jugaru-Tiron et al. 2009). The Danube is a major contributor with fresh water and sediments to the Black Sea, through the largest deltaic system in Europe, the Danube Delta (Jugaru-Tiron et al. 2009). The Danube Delta consists of three branches and a vast fluvial-marine contact zone, given the multiple splits and junctions of distributaries across the area (Zinevici & Parpală 2006). Across the Danube Delta there were counted approximately three hundred freshwater lakes, highly interconnected through a vast network of channels and canals, fuelled mostly with water from the Danube or just during flood events for remote water bodies (Gâștescu 2009).

The Danube Delta is a natural reserve, harbouring numerous wetlands and floodplain forests, which play a crucial role in lateral exchanges of nutrients with the adjacent water bodies. The Danube Delta is well-known for its biodiversity, comprising a large number of plant and animal species that unfortunately were negatively affected by the anthropogenic interventions over the past century (Zinevici & Parpală 2006, Gâștescu 2009). The human-induced disturbances, such as eutrophication,

changes in water and sediment fluxes as a result of cut-off channels or frequent dredging activities for improving navigation, the fast development rate of local tourism, but without proper sanitation conditions, have affected the complex of ecosystems and implicitly the water quality in the Danube Delta (Jugaru-Tiron et al. 2009, Pacioglu et al. 2022). The effects of forestry management actions undertaken in the Danube Delta after 1960, which included the planting of willow species and Canada poplar could also have influenced the water physico-chemistry, although this aspect was understudied (Călugăr et al., 2017). For such reasons, the proper assessment of water quality is considered nowadays essential for a sustainable implementation of biodiversity conservation and ecological restoration programs, as well as for the local eco-tourism (Stoica et al. 2016).

Another challenging aspect in the application of remote sensing techniques in assessing the quality of inland waters is the types of statistical analyses employed whenever comparing the compatibility between remote sensing estimates and laboratory measured values of water physico-chemical parameters. The linear regressions still comprise the standard statistical techniques used to test directly the reliability of estimated values for a given physico-chemical parameter from remote sensing data with those measured in the field, usually associated with other statistical metrics to estimate the goodness of fit of residual points between estimated and measured values as a result of linear regression techniques (Karaska et al. 2004, Gholizadeh et al. 2016, Soomets et al. 2020). Whilst we consider this to be a continuous learning curve for the ecologists that apply such models, we equally acknowledge the lack of studies that compared the efficiency of various models employed. With increasing availability of remote sensing data and cheap computational power, in the last three decades a data-driven continuous expansion of statistical analyses methods occurred (Ogashawara, 2021). The main statistical approaches are, besides the traditional linear regression models, the machine learning models, such as artificial neural networks, support vector machines, and genetic algorithms.

In the current study we have compared several families of machine learning algorithms, based on remote sensing data input from Sentinel 2 spectral bands, to assess the goodness of fit between satellite-based estimates and laboratory measurements for 21 physico-chemical water parameters, both optically active and inactive. We have further discussed the benefits and compared the output of several families of machine learning algorithms used to estimate the water quality parameters. The novelty from the current study is that for each physico-chemical parameter several families of machine learning models were created via automatic search for model parameters. From the array of created models, there were selected the optimal models that minimised the values of performance metrics for the tested water physico-chemical parameters.

Materials and Methods

Dataset description

The current study employed the Copernicus Sentinel 2 multispectral sensor observations source for estimating the physico-chemical parameters of water collected from 11 sampling sites within the Danube Delta Biosphere Natural Reserve (Fig. 1 and Table 1).

The remote sensing measurements used in the current study were based on the natural optical properties of water (Mobley 2022), which are disturbed by organic and inorganic dissolved and suspended matter in the aqueous medium and that influence the absorption and scattering of the electromagnetic radiation through water column. Often, in remote sensing techniques, the spectral reflectance is used as a proxy for estimating the concentrations of dissolved and suspended organic and inorganic matter in the aqueous medium. The spectral reflectance (ISO 9288: 1989), a unit measure construct with

subunit values and expressed as sr^{-1} (steradian), is the ratio between the reflected and incident radiance over a surface, described by equation 1:

$$R = \frac{L^r_{e,\Omega,\nu}}{L^i_{e,\Omega,\nu}} \quad (1)$$

where:
R – reflectance,
$L^r_{e,\Omega,\nu}$ – reflected radiance;
$L^i_{e,\Omega,\nu}$ – incident radiance.

Table 1 Sampling sites coordinates, *in-situ* (IS) and remote sensing (RS) collection dates.

Sampling Site	Longitude	Latitude	2018		2019		2020	
			Spring	Autumn	Spring	Autumn	Spring	Autumn
Ceatal Sf. Gheorghe	28.89285	45.18573	IS:03/06/2018 RS:02/06/2022	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Murighiol	29.28568	45.00522	IS:03/06/2018 RS:02/06/2022	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Pardina	28.96891	45.33078	IS:03/06/2018 RS:02/06/2022	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Chilia Veche	29.1504	45.38864	IS:03/06/2018 RS:02/06/2022	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Periprava	29.55623	45.40492	IS:03/06/2018 RS:02/06/2022	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Crisan	29.44736	45.17285	IS:03/06/2018 RS:02/06/2022	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Lacul Taranova	29.32033	45.03659	IS:12/06/2018 RS:12/06/2018	IS:06/11/2018 RS:15/10/2018 RS11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Caraorman	29.34418	45.17248	IS:12/06/2018 RS:12/06/2018	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Sf. Gheorghe - Canal Paladin	29.53588	44.93135	IS:12/06/2018 RS:12/06/2018	IS:06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Lacul Razim - Gura Portitei	28.98356	44.68834	IS:12/06/2018 RS:12/06/2018	IS:06/11/2018 RS:15/10/2018 RS11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020
Ceatal Ismail	28.73951	45.22984	IS:12/06/2018 RS:12/06/2018	IS: 06/11/2018 RS:15/10/2018 RS:11/11/2018	IS:29/06/2019 RS:27/06/2019 RS:02/07/2019	IS:12/11/2019 RS:09/11/2019 RS:04/12/2019	IS:07/06/2020 RS:06/06/2020 RS:08/06/2020	IS:08/11/2020 RS:08/11/2020

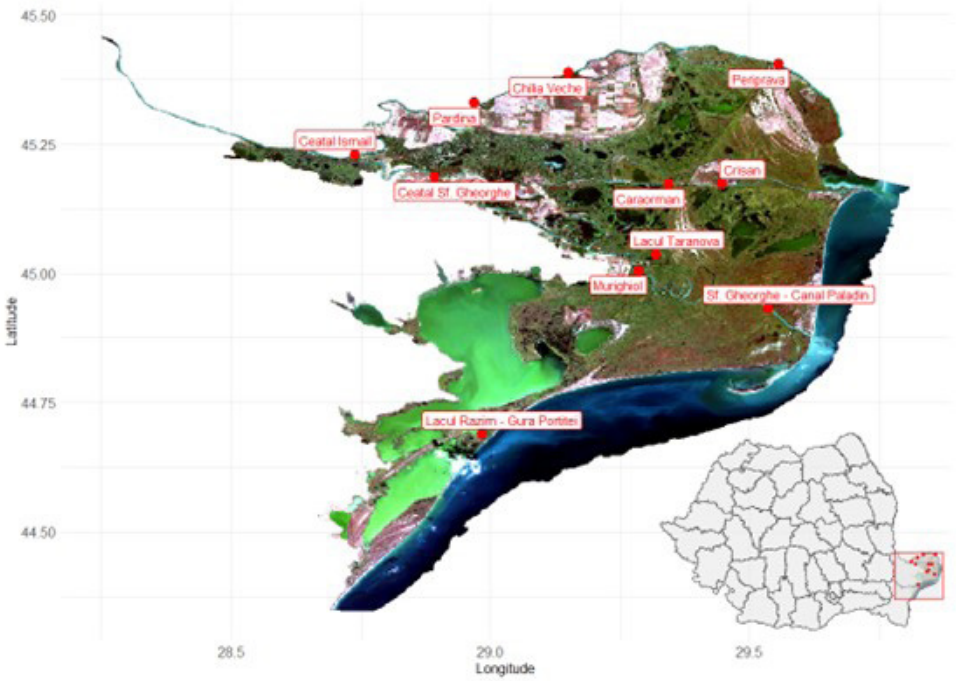

Figure 1 The locations of sampling sites on the map and the position of the Danube Delta in Romania (lower figure).

In Figure 2, the spectral reflectance was plotted for each sampling site. The spectral response for band 10 was related to the presence of cirrus clouds, hence always null for the bottom of atmospheric measurements, given two types of spectral reflectance: one measured at instrument level, and called top of the atmosphere, and the other measured at surface, and called bottom of the atmosphere. The latter type of spectral reflectance is obtained by removing distortions and noises of the electromagnetic interactions between light and atmospheric constituents. Approximating the surface spectral reflectance obtained from remote sensing satellite data to *in-situ* measured spectral reflectance comprises a continuously improving process, given that accurate estimations of various types of water quality physico-chemical parameters are based in turn on the reliability of spectral reflectance estimates at surface level (Brockmann et al. 2016, Pahlevan et al. 2020, Pereira-Sandoval et al. 2020). In the current study we have used the atmosphere correction algorithm provided with Sentinel Application Platform Sentinel-2 toolbox (Niculescu et al. 2017), which was in turn developed and maintained by the European Space Agency (Djamai et al. 2018, Steinmentz and Ramon 2018).

Eleven sampling sites, situated in the Danube Delta and Razelm - Sinoe lagoon complex were sampled during spring and autumn between 2018 and 2020 (Fig. 1 and Table 1). Water samples were collected in autoclaved HDPE bottles from each sampling site and transported to the laboratory, where 21 physico-chemical parameters were measured (Annex 1). The remote sensing data were collected from the European Spatial Agency Open Access Hub, an online data repository and long-term archive system, which enables the access to Copernicus Programme space monitoring missions: Sentinel 1, Sentinel 2, Sentinel 3 and Sentinel 5p. Sentinel 2 comprises a

polar orbiting twin constellation of satellites (Sentinel 2-A and Sentinel 2-B) deployed in a sun-synchronous orbit with a phase of 180 degrees in-between. The average revisit time at mid-latitude is two-three days, with covering swaths of 290 km wide. Sentinel 2 payload consists of a Multi Spectral Instrument covering visible, near-infrared red (NIR) and short-wave infrared (SWIR) electromagnetic radiation, split into 13 spectral bands (van der Laan et al. 2016) (Table 2). The main objective of the Sentinel 2 mission is to measure terrestrial land cover and coastal zones, by using a set of multiple spectral bands which capture the spectral response of light – vegetation interaction in the visible, near and short-wave infrared domain.

Images were acquired in June 2018, 2019, 2020 and November 2018, 2019, 2020 (Table 3), from Sentinel Open Data Hub. Data comprised cloud free Sentinel 2 MSI L1C images overlapped closely with in-situ measurements dates and region of interest (Figure 3 and Table 2). There were cases when remote sensing data were not available for the exact day when water samples were taken,

Table 2 Sentinel 2 spectral bands wavelength and spatial resolution.

Band name	Type	Central wavelength (nm)	Resolution (m)
B1	Coastal aerosol	442	60
B2	Visible blue	492	10
B3	Visible green	558	10
B4	Visible red	664	10
B5	NIR Vegetation red edge	703	20
B6	NIR Vegetation red edge	739	20
B7	NIR Vegetation red edge	779	20
B8	NIR	833	10
B8A	NIR Vegetation red edge	864	20
B9	NIR Water vapour	943	60
B10	SWIR Cloud cirrus	1377	60
B11	SWIR	1610	20
B12	SWIR	2186	20

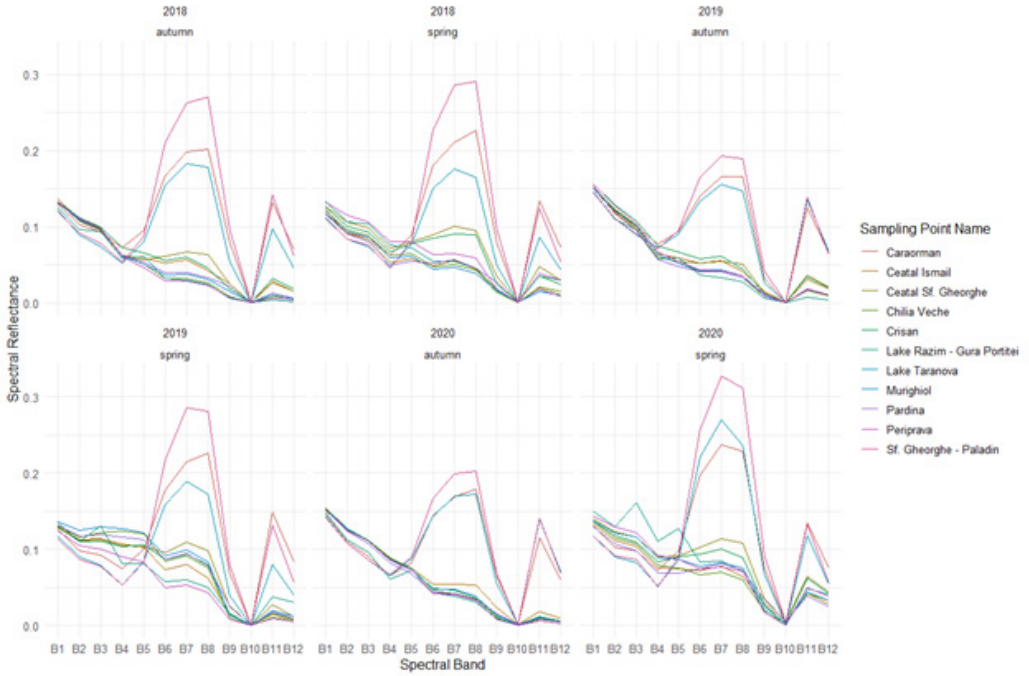

Figure 2 The spectral reflectance response of each Sentinel 2 spectral bands for *in-situ* sampling sites and seasons.

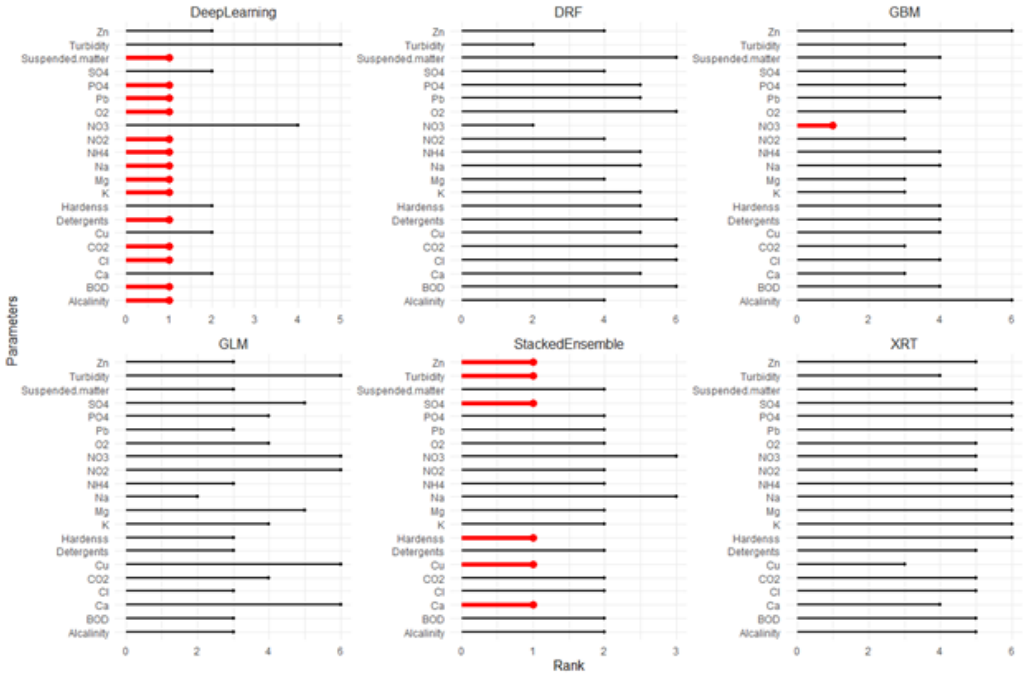

Figure 3 Model performance ranking for the measured water physico-chemical parameters.

because of strong cloud coverage (> 10%). To overcome this issue, we have calculated the mean of remote sensing data parameters (i.e. pixel values of the Sentinel 2 spectral bands, for the two nearest dates when cloud free data were available). The images were re-sampled to 10 m as part of the pre-processing step and the mean of 3 × 3 pixels window was centred on the sampling site geographical coordinates and the mean values were used as model dependent variables. The pre-processing stage comprised the computing of the bottom-of-the-atmosphere (water surface) reflectance values for the following spectral bands: B2, B3, B4, B5, B6, B7, B8, B11, B12, and using ESA SNAP Sentinel 2 toolbox (Table 2). These spectral bands were used as predictors in our models. The dependent variables were the organic and inorganic suspended and dissolved matter (Table 3).

Table 3 Model parameterization of optimal model.

No.	Water quality parameter	Model	Parameterisation	Residual deviance	Predictor variable importance – spectral bands
1	Alkalinity	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.01546	B5, B12, B2
2	BOD	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	26.055	B5, B12, B2
3	Ca	Stacked ensembles	**Super learner:** Generalised linear model **Weak learners:** deep learning (1 model), deep random forest (1 model), gradient boosting (2 models), generalised linear model (1 model)	3.30073	Deep learning, Generalised linear model, Gradient boosting
4	Cl	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	1.7181	B2, B3, B12

No.	Water quality parameter	Model	Parameterisation	Residual deviance	Predictor variable importance – spectral bands
5	CO_2	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	2.8269	B3, B2, B6
6	Cu	Stacked ensembles	**Super learner:** Generalised linear model **Weak learners:** deep learning (5 models)	0.515999	Deep learning
7	Detergents	Deep learning	3 hidden layers (3x100 neurons) with backpropagation; activation function - rectifier without dropout; 10000 learning epochs	0.74094	B2, B5, B3
8	Hardness	Stacked ensembles	**Super learner:** Generalised linear model **Weak learners:** deep learning (8 models)	0.49331	Deep learning
9	K	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.44209	B3, B2, B6
10	Mg	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.14187	B11, B2, B12
11	Na	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.44419	B5, B2, B12
12	NH_4	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.00353	B5, B4, B3
13	NO_2	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.00031	B4, B2, B3
14	NO_3	Gradient boosting	Regression forest, 25 trees	0.92622	B5, B2, B4
15	O_2	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.29176	B2, B4, B5
16	Pb	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.00986	B3, B5, B7
17	PO_4	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	0.00048	B2, B4, B3
18	SO_4	Stacked ensemble	**Super learner:** Generalised linear model **Weak learners:** deep learning (2 models)	1.44372	Deep learning
19	Suspended matter	Deep learning	3 hidden layers (3x100 neurons) with back propagation; activation function - rectifier without dropout; 10000 learning epochs	1.32201	B2, B5, B3
20	Turbidity	Stacked ensemble	**Super learner:** Generalised linear model **Weak learners:** deep learning (5 models)	3.00846	Deep learning
21	Zn	Stacked ensemble	**Super learner:** Generalised linear model **Weak learners:** deep learning (3 models)	3.85811	Deep learning

Statistical models

To test for multiple models simultaneously, we have used the machine learning framework H_2O, through an R Application Programming Interface, and quickly set up the methodology for searching the optimal fitting parameters for a mainstream set of machine learning algorithms, as described below. The single most important feature of H_2O consists in performing parameter cartesian or random grid search with respect to a pre-specified cost function, which alleviates the burden of manually fine-tuning the models.

Stacked ensemble models use a simple set of weak learners (i.e. models) and combine, usually within a model, called a super learner, which uses the predicted values from the weaker learners as input, in order to improve the predictions through cross-validation. The super learner model is independently specified

or selected from the given set of weak learners (Wolpert 1992, van der Laan et al. 2016).

Generalised linear models are used to relax the constraints assumed in common linear models, where the response and the residuals must satisfy certain conditions (i.e., linearity, independence, homoscedasticity). The generalised linear models use values generated by link functions, usually from an a priori known exponential distribution (i.e., differentiable and monotonic), parameterized with the original predictor matrix, as alternate predictors for the expected value of dependent variables (Nedler & Wedderburn 1972, Hastie & Tibshirani 1990, LeDell 2015).

Deep learning (deep neural networks) comprise a class of artificial neural networks in which the inner layer architecture consists of an arbitrary number of sub-layers. Depending on the type of connection between neurons, data flow and aggregation functions of the output, a plethora of deep neural networks are currently in existence from an application-wise standpoint (Goodfellow et al. 2016).

Gradient boosting models entail an optimisation technique for approximating the minimum of the cost functions over a set of potential candidate ensemble models, which can be either regression or classification algorithms, usually in the form of linear or tree-based models. Iteratively, a better model is fitted by combining results obtained from multiple weaker similar models. The search stops when a certain criterion is reached or if, following a certain number of iterations, the loss functions are relatively numerically stable.

Distributed regression forest is a parallel extension of regression forest algorithm for parallel environments. It is similar to the traditional regression forest, but the decisions at each split node are softer and randomized for each parallel worker, creating regression trees using only a sample of the data The result is returned as an average of all regression trees (Geurts et al. 2006).

Extreme randomised trees comprise a type of random forest algorithms, where cut-points and attributes of a regression or classification tree are randomly selected and the results are averaged according to the minimum of an arbitrary loss function (Geurts et al. 2006).

In terms of goodness-of-fit metrics we used residual deviance (Jørgensen 1997) to assess individual models and for the comparisons between observed and predicted values. Given the different measurement scales of physico-chemical parameters, the normalised root mean squared error, equation 2, was also used (here after NRMSE) and scaled by max-min (Hyndman & Koehler 2006), along with percent bias (PBIAS).

$$NRMSE = \frac{RMSE}{y_{max,i,p} - y_{min,i,p}} \quad (2)$$

where

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_{i,p} - \hat{y}_{i,p}\right)^2}{N}} \quad (3)$$

$$PBIAS = \frac{100 * \sum_{i=1}^{N}\left(y_{i,p} - \hat{y}_{i,p}\right)}{\sum_{i=1}^{N} y_{i,p}} \quad (4)$$

$y_{i,p} - i^{th}$ observed value for physico-chemical parameter p;
$\hat{y}_{i,p} - i^{th}$ predicted value for physico-chemical parameter p.

The procedure for validating the results included splitting the datasets into train and test sample, about 1000 replicates for each type of model, afterwards selecting the model which scored better on NRMSE. We also used mean, standard deviation and confidence intervals obtained through bootstrapping (Efron and Tibshirani 1993) to compare the estimates with the measurements in the field for each season-year combination (Table 4).

Results

The algorithms for selecting the most reliable models minimised the residual deviances on a

batch of several families of potential models. A cartesian search for parameter selection for each model iteration was defined: learning rate, number of sub-samples, number of iterations, stop criterion for minimization, number of maximum hidden layers, and the number of maximum artificial neurons per layer. The best models were ranked (Fig. 3), based on the minimum residual deviance. The deep learning method, along with the stacked ensembles, dominated the pack of selected models. Table 3 presents for each parameter the corresponding most reliable models and it can be observed that the stacked ensembles also comprise a generalised linear model composite of deep learning results. The results were expectable, given that deep learning models are proved to be universal approximates, when parameterized with an optimal set of weights and can be used to fit any type of non-linear function. However, the neural networks require a representative number of input data, and in the current case it can be observed that the residual deviance varied greatly among model estimates, mainly for those water parameters that were zero-inflated and with outliers (i.e. detergents and CO_2). Also, as other types of machine learning algorithms, deep learning models are sensitive to the number of predictors versus observation ratios. The spectral bands B2 (visible blue light), B3 (visible green light), B4 (visible red light), B5 (near infrared) and B12 (shortwave infrared) were among the most important top variables (Table 3).

To compare how well the models fitted the data, a plot of observed versus predicted was graphically constructed, using a theoretical line of best fit, NRMSE as the numerical goodness of fit metric (Fig. 4). For alkalinity, calcium, chloride, carbon dioxide, hardness, potassium, sodium, ammonium, dissolved oxygen, sulphates, and suspended matter the overall percentage bias of the estimates was +/- 10% from the observed values. For copper, magnesium, nitrites, nitrates, turbidity and zinc the percentage bias varied in the interval +/- 10% - 20%, whereas for BOD, detergents, led, and phosphates the percentage bias was higher, with maximum above 90% for detergents (Fig. 4 and Table 4). The highest deviations from the predicted values for the category of parameters with the lowest percentage bias (i.e. +/- 10%), was recorded for carbon dioxide, with values either being underestimated or overestimated (Table 4). The explanation for this high deviation of carbon dioxide could be the exclusion of seasons or years as predictors during the model parameterisation. For the parameters where the percentage bias varied in the range +/- 20%, the deviations from observed values are explicable either by values close to 0 or by outliers with one order of magnitude greater than 0 (e.g. detergents, Table 4).

Discussion

Overall, the employed models were a bag of mixed results; for few parameters (i.e., alkalinity, calcium, chloride, carbon dioxide, hardness, potassium, sodium, ammonium, dissolved oxygen, sulphates, and suspended matter) the goodness of fit measure indicated reliable approximations for *in-situ* observations, whereas for other models it failed (Fig. 4 and Table 4). The example of biological oxygen demand is relevant for the current study, given the estimate of the employed models were only fairly accurate, whereas previous studies (Rivera et al. 2014), which used a combination of Sentinel 2 data, multispectral data collected *in-situ* from a spectral instrument mounted on unmanned aerial vehicle, and water samples, it offered better fits. Data collected from unmanned aerial vehicle was used to calibrate data collected from Sentinel 2, before using any estimation techniques on *in-situ* versus remote sensing data. Other researchers reported excellent fits (Kabolizadeh et al. 2022), by using a combination of artificial neural networks and spectral indices (i.e. spectral bands aggregation) for water alkalinity, calcium, chloride, potassium, sodium, sulphates, suspended

matter and turbidity, with strong correlations between observed and predicted values. These authors also discarded the spectral bands with lower resolution (> 10 m per pixel), a venue that we consider an interesting area for further research. High outliers in observed data (e.g. possibly detected with an arbitrary threshold of ±3 standard deviations from the mean), may induce strong biases in predictions. This issue can be solved either be eliminating the outliers, but possibly losing valuable information or by generating models on homogenous spatial-temporal sub-samples, such as models for each homogenous water type and time frame of data collection, where the likelihood of observing outliers is greater compared to the entire sample. In the latter case, supplementary spatial data could be entered, such as the proximity to point or diffuse pollution sources (e.g., towns, agricultural terrains) to improve the model efficiency.

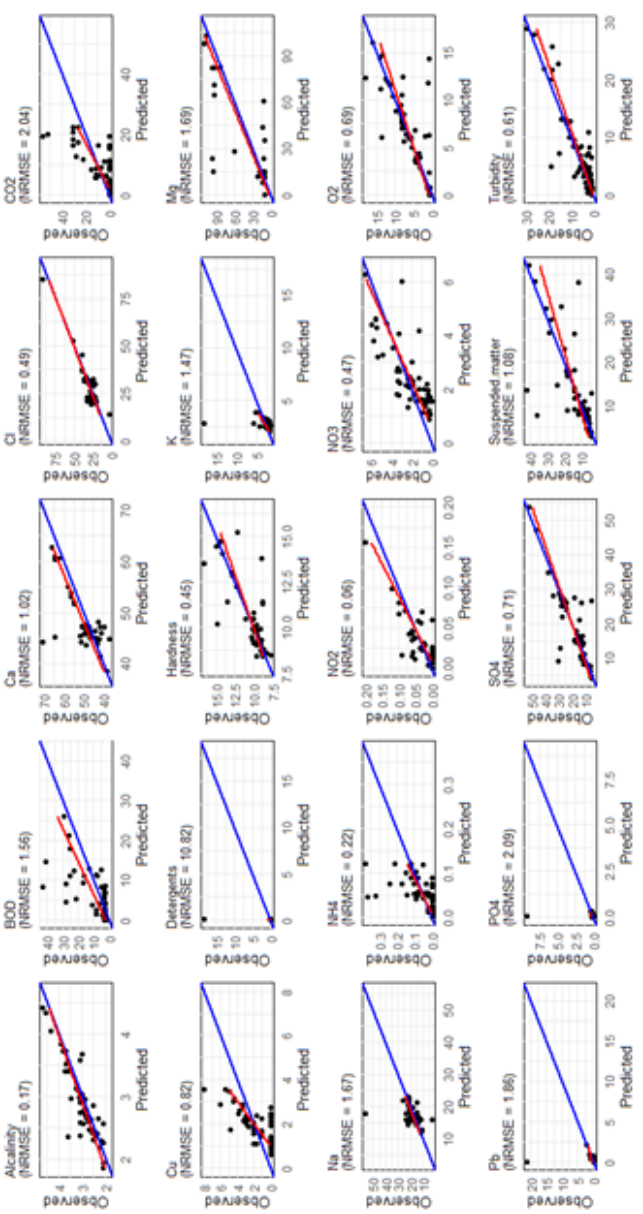

Figure 4 Observed versus predicted values. Red lines represent actual fit, blue lines theoretical best fit..

The current research based on Sentinel 2 data for estimating inland water quality parameters is divided among scientists that either derived empirical models to estimate these parameters and those that rather prefer to enhance the Sentinel 2 data attributes, such as by deriving L3 data products to be used as predictors. Previous comparisons between *in-situ* measurements and remotely

detected concentrations of chlorophyll-a, turbidity and suspended matter highlighted that despite good reliability at local level, the extrapolation to other biogeographic areas was cumbersome, due to a complex set of local specificities (Soltana et al. 2017, Alvado et al. 2021, Rahul et al. 2022). This was the rationale for defining the optical water types as an intermediary L3 product before applying the models, suggesting that the spectral reflectance from Sentinel 2 data is a reliable method that improves the robustness of water physico-chemical parameters estimates. Other studies (Alvado et al. 2021, Rahul et al. 2022) suggested an algorithm for differentiating between the organic and inorganic components of total suspended matter with the aid of a specialised software tool that created suitable sets of spectral indices as predictors of water quality. It was also argued that Sentinel-2 data, although lacking the high resolution needed for the proper assessment of physico-chemical parameters in small water bodies, is still suitable for measuring the water quality as required by Water Framework Directive and to compensate for the high cost and human labour required during routine monitoring programs (Soltana et al. 2017). In the Danube Delta several previous studies showed promising results. One study (Niculescu et al. 2017) mapped the wetland habitats from this region and assessed the anthropogenic effects on local ecosystems, by combining Sentinel 1 and Sentinel 2 data. The potential of Sentinel 2 data for discriminating between coastal and river water types at the Danube Delta-Black Sea confluence was successfully employed (Dumitru et al. 2019). Based on Sentinel 1 and Sentinel 2 time data series data fusion, an automatic algorithm was implemented to overcome the issue of cloud presence (Oteman et al. 2021). Also, based on Sentinel 1 and Sentinel 2 data (Oteman et al. 2021), there were measured the areas covered by reed belts within the Danube Delta and the zones covered by other types of local vegetation were efficiently identified, overcoming the inherent issues induced by atmospheric phenomena and/or adjacency effects.

A cumbersome aspect whenever using remote sensing data is the necessity of overlapping *in-situ* collection dates with the dates when the satellites actually passed over the area of interest. The Sentinel-2 satellites have a mean revisit time between two passes over the same area of two-three days and the cloud presence could negatively impact the quality of data. To overcome this issue, a mean aggregation of two adjacent by date passes was used, creating a suitable time frame for the *in-situ* collection of samples. Whereas various water quality parameters can vary greatly not only between a few days' window, but also over a daily basis, we considered this method to be suitable when no auxiliary data are available.

Another type of issue encountered in the current study was the limited dataset, a situation that occurs frequently during sampling programs undertaken in remote locations and characterised by arduous working conditions. In order to overcome this issue, several techniques of artificially created datasets can be constructed, by exploring promising methods in the field of synthetic data generation (Wolpert 1992, Smith et al. 2009, Soltana et al. 2017, Zhang et al. 2018).

Conclusion

In the current study, Sentinel-2 remote sensing data were used to estimate 21 water quality parameters from eleven sampling sites situated in the Danube Delta (Romania) using remote sensing spectral reflectance as predictors in order to assess the suitability of Sentinel 2 MSI data and to compare the correspondence to those measured in laboratory. Due to small sample size of *in-situ* measurements, the observations were pooled within a single dataset and several machine learning algorithms were further applied with mixed results, given the high number of null values

and some outliers. For certain parameters (i.e., alkalinity, calcium, chloride, carbon dioxide, hardness, potassium, sodium, ammonium, dissolved oxygen, sulphates, and suspended matter) the results were promising, with an overall percentage bias of the estimates of

Table 4 Predicted values summary statistics: mean, standard deviation, and confidence intervals of observed vs. estimates, along with percentage bias (PBIAS)

Parameter	2018				2019				2020				PBIAS (%)
	Spring		Autumn		Spring		Autumn		Spring		Autumn		
	Obs	Est	Obs	Est	Obs	Est	Obs	Est	Obs	Est	Obs	Est	
Alkalinity	2.54 (0.73) [2.22- 2.97]	2.59 (0.75) [2.23- 3.05]	3.03 (0.2) [2.95- 3.16]	2.9 (0.27) [2.74- 3.05]	3.63 (0.34) [3.45- 3.84]	3.44 (0.52) [3.13- 3.69]	3.06 (0.13) [2.98- 3.12]	2.98 (0.14) [2.9- 3.05]	3.19 (0.61) [2.89- 3.58]	3.03 (0.53) [2.78- 3.34]	2.56 (0.17) [2.45- 2.65]	2.51 (0.11) [2.44- 2.57]	3.06
BOD	3.22 (1.82) [2.49- 4.4]	2.96 (3.08) [1.4- 4.78]	2.69 (0.66) [2.35- 3.09]	0.98 (3.12) [-0.72 -2.9]	14.47 (13.2) [7.91- 22.06]	9.64 (4.05) [7.3- 11.93]	3.08 (1.79) [2.39- 4.23]	1.02 (2.82) [-0.42 -2.63]	18.28 (11.82) [11.92- 25.46]	10.77 (7.56) [6.84- 15.45]	5.25 (8) [2.54- 10.34]	0.71 (3.27) [-0.84 -3.01]	44.08
Ca	45.01 (7.13) [41.78 - 49.76]	44.17 (2.49) [42.74- 45.46]	48 (2.55) [46.48- 49.14]	46.5 (1.27) [45.78 - 47.1]	48.85 (3.37) [46.73- 50.56]	46.85 (1.09) [46.19 -47.36]	48.35 (2.31) [46.94- 49.54]	46.72 (1.07) [46.09 - 47.28]	-	-	46.01 (3.54) [43.85- 47.69]	46.01 (0.77) [45.56 - 46.44]	2.57
Cl	24.81 (10.87) [18.7 - 31.52]	26.77 (10.04) [21.83 - 33.11]	31.39 (1.53) [30.74 - 32.4]	30.3 (3.08) [28.34 - 31.45]	38.45 (18.3) [31.18 - 49.89]	36.26 (17.77) [28.66 - 46.53]	29.79 (4.05) [27.22 - 31.82]	28.99 (4.4) [26.12 - 31.02]	26.73 (4.47) [24.26 - 29.14]	26.51 (4.38) [24.09 - 28.91]	21.43 (1.35) [20.73 - 22.33]	20.64 (0.81) [20.24 - 21.18]	1.79
CO2	33.05 (10.97) [27.7 - 39.58]	17.81 (3.57) [15.7 - 19.76]	0 (0) [0 - 0]	6.51 (6.59) [3.24 - 10.18]	0 (0) [0 - 0]	6.81 (6.03) [3.64 - 10.06]	0 (0) [0 - 0]	1.5 (1.44) [0.77 - 2.34]	7.5 (3.96) [5.2 - 9.6]	8.7 (3.8) [6.71 - 10.9]	1.5 (4.74) [0 - 4.5]	3.14 (2.9) [1.66 - 4.98]	-5.42
Cu	3.97 (1.21) [3.31 - 4.64]	2.67 (0.69) [2.29 - 3.04]	0 (0) [0 - 0]	1.48 (0.53) [1.23 - 1.8]	1.43 (2.63) [0.15 - 3.17]	1.52 (0.94) [1 - 2.05]	1.45 (0.93) [0.91 - 2]	1.62 (0.36) [1.43 - 1.83]	0.18 (0.6) [0 - 0.55]	1.4 (0.45) [1.17 - 1.7]	2.27 (1.15) [1.51 - 2.86]	2.13 (0.32) [1.95 - 2.32]	-16.86
Detergents	0.04 (0.08) [0 - 0.09]	0.03 (0.01) [0.02 - 0.04]	0.03 (0.07) [0 - 0.08]	0.03 (0.01) [0.02 - 0.04]	1.67 (5.55) [0 - 5.02]	0.01 (0.02) [0 - 0.02]	0.01 (0.04) [0 - 0.03]	0.01 (0.01) [0.01 - 0.02]	0 (0) [0 - 0]	0.01 (0.01) [0 - 0.02]	0.01 (0.03) [0 - 0.03]	0.02 (0.02) [0.01 - 0.03]	93.5
Hardness	9.66 (1.37) [9.08 - 10.5]	9.65 (0.97) [9.17 - 10.21]	9.61 (0.31) [9.41 - 9.76]	9.83 (0.59) [9.54 - 10.19]	10.03 (0.52) [9.74 - 10.32]	10.56 (1.13) [10.07 - 11.28]	9.49 (0.3) [9.31 - 9.65]	9.39 (0.46) [9.09 - 9.63]	13.63 (1.74) [12.63 - 14.57]	13.11 (1.86) [12.03 - 14.03]	8.62 (0.38) [8.37 - 8.82]	8.68 (0.23) [8.56 - 8.84]	-0.27
K	2.65 (0.87) [2.29 - 3.19]	2.75 (0.45) [2.54 - 3.04]	3.17 (0.41) [2.99 - 3.44]	2.93 (0.24) [2.79 - 3.05]	2.55 (0.79) [2.26 - 3.05]	2.72 (0.43) [2.52 - 2.96]	3.33 (0.88) [2.92 - 3.85]	3.04 (0.29) [2.89 - 3.23]	1.93 (0.38) [1.75 - 2.16]	2.43 (0.29) [2.27 - 2.6]	4.09 (4.85) [2.49 - 7.17]	2.71 (0.05) [2.68 - 2.73]	5.84
Mg	11.88 (2.21) [10.91 - 13.38]	11.92 (5.44) [8.81 - 15.13]	14.1 (0.81) [13.74 - 14.61]	15.94 (9.3) [12.74 - 21.6]	11.65 (3.05) [10.62 - 13.56]	18.9 (15.52) [12.1 - 27.83]	13.09 (0.4) [12.85 - 13.3]	11.82 (2.14) [10.61 - 12.86]	-	-	12.19 (0.32) [12.02 - 12.38]	11.91 (0.26) [11.76 - 12.05]	-12.3

+/- 10% from the observed values. For other parameters (i.e., copper, magnesium, nitrites, nitrates, turbidity and zinc) the estimates were fairly accurate, with percentage bias in the interval +/- 10% - 20%, whereas for detergents, biological oxygen demand, led, and phosphates

Na	13.67 (4.55) [11.68 - 16.57]	14.79 (3.06) [13.18 - 16.56]	21.57 (0.73) [21.21- 21.99]	19.04 (2.56) [17.46 - 20.32]	19.43 (12.15) [15.45 - 26.96]	17.13 (1.79) [16.26 - 18.13]	20.89 (2.12) [19.54 - 21.75]	19.85 (2.89) [18.14 - 21.42]	18.92 (2.59) [17.51 - 20.51]	17.65 (2.23) [16.52 - 19.07]	14.14 (3.99) [11.52 - 15.82]	15.71 (0.86) [15.31 - 16.25]	4.27
NH4	0.09 (0.06) [0.06 - 0.12]	0.08 (0.03) [0.06 - 0.09]	0.14 (0.07) [0.11 - 0.18]	0.09 (0.02) [0.08 - 0.1]	0.01 (0.02) [0 - 0.03]	0.04 (0.02) [0.02 - 0.05]	0.1 (0.15) [0.02 - 0.19]	0.07 (0.03) [0.05 - 0.09]	0.01 (0.04) [0 - 0.04]	0.04 (0.02) [0.03 - 0.05]	0.01 (0.01) [0 - 0.02]	0.02 (0.02) [0.01 - 0.04]	6.94
NO2	0.06 (0.04) [0.04 - 0.08]	0.05 (0.03) [0.03 - 0.06]	0.05 (0.01) [0.05 - 0.06]	0.04 (0.01) [0.03 - 0.04]	0 (0) [0 - 0]	0.02 (0) [0.01 - 0.02]	0.03 (0.06) [0 - 0.07]	0.04 (0.05) [0.01 - 0.06]	0 (0) [0 - 0]	0.01 (0.01) [0.01 - 0.02]	0 (0) [0 - 0]	0.02 (0) [0.01 - 0.02]	-15.17
NO3	1.27 (0.76) [0.8 - 1.7]	1.8 (0.35) [1.59 - 1.97]	2.67 (1.17) [2.07 - 3.32]	3.06 (1.36) [2.4 - 3.88]	0.66 (0.68) [0.36 - 1.1]	1.62 (1) [1.16 - 2.23]	3.2 (2.65) [1.78 - 4.65]	3.08 (1.29) [2.37 - 3.77]	2.95 (1.96) [1.92 - 4.04]	2.48 (1.36) [1.88 - 3.28]	1.16 (0.65) [0.82 - 1.56]	1.51 (0.24) [1.38 - 1.65]	-13.6
O2	8.03 (0.22) [7.89 - 8.13]	8.14 (0.97) [7.64 - 8.68]	8.02 (0.19) [7.9 - 8.13]	8.41 (1.57) [7.57 - 9.41]	2.27 (3.4) [0.73 - 4.45]	5.06 (5.27) [2.1 - 8.11]	2.97 (1.86) [1.86 - 3.97]	3.08 (2.32) [1.82 - 4.3]	12.74 (3.54) [10.45 - 14.58]	11.34 (3.73) [9.17 - 13.34]	5.11 (1.75) [3.91 - 6.06]	5.3 (1.81) [4.14 - 6.18]	-5.65
Pb	0.66 (0.62) [0.35 - 1.06]	0.59 (0.57) [0.3 - 0.93]	0 (0) [0 - 0]	0.06 (0.17) [-0.01 - 0.17]	1.93 (6.39) [0 - 5.78]	0.07 (0.2) [0 - 0.19]	0 (0) [0 - 0]	0.01 (0.01) [0.01 - 0.02]	0 (0) [0 - 0]	0.05 (0.09) [0.01 - 0.11]	0.07 (0.24) [0 - 0.22]	0.06 (0.11) [0.01 - 0.14]	68.58
PO4	0.94 (2.71) [0.09 - 2.59]	0.09 (0.08) [0.05 - 0.14]	0.15 (0.04) [0.12 - 0.17]	0.15 (0.06) [0.12 - 0.18]	0.09 (0.06) [0.06 - 0.13]	0.08 (0.07) [0.04 - 0.13]	0.23 (0.07) [0.19 - 0.27]	0.22 (0.06) [0.19 - 0.25]	0.06 (0.11) [0.01 - 0.12]	0.06 (0.09) [0.01 - 0.12]	0.1 (0.04) [0.08 - 0.13]	0.11 (0.04) [0.09 - 0.14]	54.92
SO4	21.14 (14.54) [14.2 - 29.55]	21.78 (14.97) [14.43 - 30.44]	15.71 (4.85) [13.45 - 18.62]	14.98 (4.08) [13.01 - 17.59]	18.45 (9.52) [13.35 - 24.5]	16.83 (9.42) [11.72 - 22.37]	8.45 (6.37) [5.94 - 12.37]	10.29 (8.71) [5.71 - 15.73]	25.74 (6.8) [21.4 - 29]	22.91 (7.46) [18.03 - 26.72]	26.52 (3.42) [24.33 - 28.21]	27.6 (1.12) [26.94 - 28.22]	1.51
Suspended matter	19 (6.31) [15.82 - 22.73]	19.55 (7.51) [15.64 - 24]	10.36 (1.75) [9.45 - 11.36]	10.28 (5.36) [8.44 - 13.62]	24.91 (12.52) [18.27 - 32.55]	21.61 (13.28) [14.23 - 29.51]	8.73 (4.41) [6.36 - 11.09]	9.66 (3.69) [7.62 - 11.92]	6.55 (1.97) [5.45 - 7.64]	6.24 (2.3) [4.95 - 7.48]	7.6 (1.26) [6.8 - 8.4]	7.86 (0.85) [7.37 - 8.37]	2.58
Turbidity	3.24 (3.4) [1.68 - 5.54]	4.39 (3.16) [2.89 - 6.12]	2.34 (1.14) [1.72 - 2.99]	3.77 (1.91) [2.75 - 4.9]	10.26 (8.7) [5.55 - 15.64]	7.91 (7.79) [4.18 - 13.29]	12.69 (8.35) [8.11 - 17.58]	13.89 (10.04) [8.42 - 19.55]	3.18 (5.13) [1.34 - 6.35]	5.13 (6.16) [2.34 - 9.2]	6.27 (2.93) [4.75 - 7.91]	7.29 (3.33) [5.43 - 9.08]	-11.49
Zn	5.5 (2.03) [4.44 - 6.62]	5.5 (1.6) [4.63 - 6.41]	0.02 (0.03) [0.01 - 0.04]	2.76 (1.81) [1.84 - 3.78]	6.37 (3.03) [4.91 - 8.46]	4.28 (1.72) [3.31 - 5.22]	2.45 (2.07) [1.36 - 3.64]	2.94 (1.5) [2.18 - 3.77]	0.27 (0.9) [0 - 0.82]	1.94 (1.04) [1.42 - 2.59]	3.64 (2.35) [2.27 - 5.02]	2.95 (0.88) [2.59 - 3.54]	-12.13

the percentage bias was higher, with maximum recorded above 90% for detergents. The current study also employed for each physico-chemical parameter several families of machine learning models that were created via automatic search for model parameters. This was a novel approach, reflected in the array of created models that were selected based on the optimal models that minimised the values of performance metrics for the tested parameters.

Acknowledgments

The research was funded by the projects "Analysis of the potential for sustainable use of vegetation specific to the Danube-Danube Delta-Black Sea system", Code 108630, co-financed by the European Regional Development Fund and by the National Core Program—Romanian Ministry of Research and Innovation, Romania, Project 25N/2019 BIODIVERS 19270103 project. Octavian Pacioglu was also funded by a grant from the Romanian National Authority for Scientific Research and Innovation (UEFISCDI), Romania, Project Number PN-III-P1-1.1-TE-2021-0221.

Supplementary materials

Annex 1S. Mean, standard deviation and confidence intervals for water physico-chemical parameters in each sampling site for 2018-2020, as well as the ISO standards and equipment used for their measurement in the laboratory.

References

Alvado B., Sòria-Perpinyà X., Vicente E., Delegido J., Urrego P., Ruíz-Verdú A., Moreno J., 2021. Estimating organic and inorganic part of suspended solids from Sentinel 2 in different inland waters. Water 13(18): 2453; https://doi.org/10.3390/w13182453

Brockmann C., Doerffer R., Peters M., Kerstin S., Embacher S., Ruescas A., 2016. Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. Living Planet Symposium 740: 54.

Călugăr A., Ivan O., Acatrinei L., Huțanu, M., 2017. From a green perspective: management pressures on forest ecosystems from Danube Delta Biosphere Reserve linked with soil mesofauna dynamics and foliar gas-exchange parameters. In 15th International Conference on Environmental Science and Technology, 31 August - 2 September 2017, Rhodes, Greece. CEST, pp. 29-35.

Djamai N., Fernandes R., 2018. Comparison of SNAP-derived Sentinel-2A L2A product to ESA product over Europe. Remote Sensing 10: 926. https://doi.org/10.3390/rs10060926

Dumitru C.O., Dax G., Schwarz G., Cazacu C., Adamescu M.C.,Datcu M., 2019. Accurate monitoring of the Danube Delta dynamics using Copernicus data. In Bostater C.R., Neyt X., Viallefont-Robinet F., (ed.), Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions, 9-10 September 2019, Strasbourg. SPIE, pp. 111-122.

Efron B., Tibshirani R. J., 1993. An Introduction to the Bootstrap. Chapman & Hall/CRC, United Kingdom.

European Parliament and European Council. 2000. Water Framework Directive.

Gâștescu P., 2009. The Danube Delta biosphere reserve. Geography, biodiversity, protection, management. Revue Roumaine de Géographie 53: 139-152.

Geurts P., Ernst D., Wehenkel L., 2006. Extremely randomized trees. Machine Learning 63: 33-42.

Gholizadeh M.H., Melesse A.M., Reddi L., 2016. A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors 16: 1298. https://doi.org/10.3390/s16081298

Goodfellow I., Bengio Y.,Courville A., 2016. Deep Learning. MIT Press, Massachusetts, United States of America, 356 p.

Hastie T.J., Tibshirani R.J., 1990. Generalized Additive Models. Chapman & Hall/CRC, London, United Kingdom, 307 p.

Hyndman R.J., Koehler A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22: 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Ielenicz M., Comanescu L., 2006. Romania touristic potential. Ed. Universitara, Bucharest, Romania, 174 p.

Ogashawara I., 2021. Bibliometric Analysis of Remote Sensing of Inland Waters Publications from 1985 to 2020. Geographies 1: 346-361. https://doi.org/10.3390/geographies1030019

International Standards Organisation (ISO), 1989. Thermal insulation – Heat transfer by radiation – Physical quantities and definitions.

Jørgensen B., 1997. The Theory of Dispersion Models. Chapman &Hall, London, United Kingdom, 473 p.

Jugaru-Tiron L., Le Coz J., Provansal M., Panin N., Raccasi G., Dramais G., Dussouillez, P., 2009. Flow and sediment processes in a cutoff meander of the Danube Delta during episodic flooding. Geomorphology 06: 186-

197. https://doi.org/10.1016/j.geomorph.2008.10.016
Kabolizadeh M., Rangzan K., Zareie S., Rashidian M., Delfan H., 2022. Evaluating quality of surface water resources by ANN and ANFIS networks using Sentinel-2 satellite data. Earth Science Informatics 15: 523-540. https://doi.org/10.1007/s12145-021-00741-z
Karaska M.A., Huguenin R.L., Beacham J.L., Wang M.H., Jensen J.R., Kaufmann R.S., 2004. AVIRIS measurements of chlorophyll, suspended minerals, dissolved organic carbon, and turbidity in the Neuse River, North Carolina. Photogrammetric Engineering & Remote Sensing 70: 125-133. https://doi.org/10.14358/PERS.70.1.125
LeDell E., 2015. Scalable Ensemble Learning and Computationally Efficient Variance Estimation. PhD, University of California, Berkley, United States of America, 251 p.
Mobley C.D., 2022. The Oceanic Optics Book. International Ocean Colour Coordinating Group (IOCCG), Dartmouth, NS, Canada, 924 p.
Niculescu S., Lardeux C., Hanganu, J., 2017. Synergy between Sentinel-1 radar time series and Sentinel-2 optical for the mapping of restored areas in Danube delta. In Brewer C.A. (ed.), Proceedings of the International Cartographic Association, Strasbourg, 9-11 September, pp. 1-10.
Nedler J., Wedderburn R., 1972. Generalized Linear Models. Journal of the Royal Statistical Society: Series A 153: 370-384.
Oteman B., Scrieciu A., Bouma T.J., Stanica A., van der Wal D., 2021. Indicators of expansion and retreat of Phragmites based on optical and radar satellite remote sensing: a case study on the Danube delta. Wetlands 41: 72. https://doi.org/10.1007/s13157-021-01466-x
Pacioglu O., Duțu L., Duțu F., Pavel A.B., 2022. Habitat preferences and trophic interactions of the benthic invertebrate communities inhabiting depositional and erosional banks of a meander from Danube Delta (Romania). Global Ecology and Conservation 38: e02213. https://doi.org/10.1016/j.gecco.2022.e02213
Pahlevan N., Smith B., Schalles J., Binding C., Cao Z., Ma R., Stumpf R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sensing of Environment 240: 111604. https://doi.org/10.1016/j.rse.2019.111604
Pereira-Sandoval M., Ruescas A., Urrego P., Ruiz-Verdú A., Delegido J., Tenjo C., Moreno J., 2020. Evaluation of atmospheric correction algorithms over Spanish inland waters for sentinel-2 multi spectral imagery data. Remote Sensing 11: 1469. https://doi.org/10.3390/rs11121469
Rahul T.S., Brema J., Wessley G., 2022. Evaluation of surface water quality of Ukkadam lake in Coimbatore using UAV and Sentinel-2 multispectral data. International Journal of Environmental Science and Technology, 1-16. https://doi.org/10.1007/s13762-022-04029-7
Rivera J.P., Verrelst J., Delegido J., Veroustraete F., Moreno J., 2014. On the Semi-Automatic Retrieval of Biophysical Parameters Based on Spectral Index Optimization. Remote Sensing 6: 4927–4951. https://doi.org/10.3390/rs6064927
Smith D.M., Clarke G.P., Harland K., 2009. Improving the synthetic data generation process in spatial microsimulation models. Environment and Planning A: Economy and Space 41: 1251-1268. https://doi.org/10.1068/a4147
Soltana G., Sabetzadeh M., Briand L.C., 2017. Synthetic data generation for statistical testing. In IEEE/ACM International Conference on Automated Software Engineering, Urbana-Champaign, IL, USA, 12-19 October 2017. IEEE, pp. 872-882.
Soomets T., Uudeberg K., Jakovels D., Brauns A., Zagars M., Kutser T., 2020. Validation and comparison of water quality products in Baltic lakes using Sentinel-2 msi and Sentinel-3 OLCI data. Sensors 20: 742. ttps://doi.org/10.3390/s20030742
Stoica C., Camejo J., Banciu A., Nita-Lazar M., Paun I., Cristofor S., ... Guevara M., 2016. Water quality of Danube Delta systems: ecological status and prediction using machine-learning algorithms. Water Science and Technology 73: 2413-2421. https://doi.org/10.2166/wst.2016.097
Topp, Simon N. Pavelsky, Tamlim M. Jensen, Daniel Simard, Marc, Ross, Matthew R.V., 2019. Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. Water 12(1):169. https://doi.org/10.3390/w12010169
van der Laan M.J., Polley E.C., Hubbard L., 2016. Super learner. Statistical Applications in Genetics and Molecular Biology 9: 1-6. https://doi.org/10.2202/1544-6115.1309
Wolpert D.H., 1992. Stacked Generalization. Neural Networks 5: 241 -259.
Zhang L., Gonzalez-Garcia A., Van De Weijer J., Danelljan M., Khan F.S., 2018. Synthetic data generation for end-to-end thermal infrared tracking. IEEE Transactions on Image Processing 28: 1837-1850. https://doi.org/10.48550/arXiv.1806.01013
Zinevici V., Parpală L., 2006. The zooplankton structure and productivity in Danube Delta lacustrine ecosystems. In Tudorancea C., Tudorancea M. (eds.), Danube Delta. Genesis and Biodiversity. Backhuys Publishers, Leiden, pp. 177-210.