Turbidity Estimation by Machine Learning Modeling and Remote Sensing Techniques Applied to a Treatment Plant Water Inlet

Víctor H. Gauto
Grupo de Investigación Sobre Temas Ambientales y Químicos
Universidad Tecnológica Nacional, Resistencia, Argentina
e–mail: victor.gauto@ca.frre.utn.edu.ar

Enid M. Utgés, Elsa I. Hervot, María D. Tenev, Alejandro R. Farías
Grupo de Investigación Sobre Temas Ambientales y Químicos
Universidad Tecnológica Nacional, Resistencia, Argentina
e–mail: gistaq@ca.frre.utn.edu.ar

ABSTRACT

Water availability and sanitation are among the UN Sustainable Development goals for 2030. Remote sensing techniques are used to monitor and retrieve quality estimators from water bodies. Clean water is a scarce resource fundamental for human development and well-being. Treatment plants depend on the current water quality state to properly provide clean water. Combining laboratory measurements, provided by a water plant in Resistencia city, Argentina, and remote sensing data, i.e., surface reflectance, from Sentinel-2 platform, several algorithms were developed, trained, and compared for turbidity estimation. The model with the highest performance metrics was a random forest model, with Pearson's coefficient of determination (R^2) 0.918 and root-mean squared error (RMSE) 138.8 nephelometric turbidity units (NTU). Global feature importance and partial dependencies profiles techniques were applied to the random forest model to understand the spectral bands effects. Turbidity maps and time series were made and analyzed.

KEYWORDS

Machine learning, random forest, remote sensing, turbidity, water treatment.

INTRODUCTION

Ensure water availability is one of the Objectives of 2030 Agenda for Sustainable Development by UN [1]. To achieve this, satellite remote sensing techniques can be applied to study and monitoring of water bodies, since it's possible to retrieve spectral data from large portions of the Earth surface. Applied remote sensing can be used to estimate biophysical water parameters, such as total suspended matter [2], chlorophyl-a [3], Secchi disc depth [4] and turbidity [5]. These algorithms can be relatively simple equations [6], or a more complex approach, like machine learning techniques [7]. Remote sensing techniques can be used for research in a wide range of environmental topics, such as land pollution [8] and glacier retreat [9].

Sentinel-2 (S2) is a spatial mission developed by the European Space Agency (ESA), consisting of two platforms. The MultiSpectral Instrument (MSI) is the optical sensor mounted in S2, with a maximum spatial resolution of 10 m, spectral range of 440 nm to 2200 nm, and a revisit time of 5 days for the constellation. S2 database is free and open access, available from the Copernicus Open Access Hub. S2-MSI has been used in water monitoring and parameter estimation, like color of water [10], chlorophyll-a concentration [11], and CDOM [12]. The generated products from S2-MSI are reliable [13] due to a low radiometric uncertainty [14].

The Paraná River is the second longest river in South America, running through 4000 km [15]. In Argentina, is the natural boundary of multiple provinces, until it reaches the Río de la Plata, to end in Atlantic Ocean. Paraguay River, with 2550 km [15], is a tributary of Paraná River, in its middle basin, North-East Argentina. Due to the high sediment concentration in Paraguay River, the discharge made into Paraná River changes the characteristics of its composition, creating two distinct regions of high (West) and low (East) sediment concentration [16].

The Metropolitan Area of Gran Resistencia (MAGR) is an urban region in Chaco Province, North-East Argentina. It's made of four cities, including the capital city of Chaco, Resistencia. MAGR has a population of 423000 inhabitants, according to the last census [17]. Paraná River has a large impact in its society: fishing industry, tourism, recreational activities of the local communities, and transportation route [18]. The water source for the MAGR potabilization plant is in an arm of Paraná River, Barranqueras River, which is connected to two main rivers in the metropolitan area, Negro and Tragadero Rivers. The water plant needs to adapt its potabilization process to ensure the removal of large amounts of sediments presents in the water. Monitoring and understanding the spatial distribution of water turbidity in Barranqueras River is a valuable input to the overall system since it can be used to give support to the making decision process. Remote sensing techniques can be applied to retrieve spectral data from the water, and thus create a model to estimate turbidity remotely.

In this study, daily water turbidity values were given by the Resistencia water plant. Using S2-MSI images, processing level L2A, surface reflectance (R_S) was extracted for the water inlet location, at surface level. A database of spectral values and turbidity measurements was built to train several mathematical models, including a machine learning approach, a random forest (RF) model. The model with the best performance metrics was selected, and turbidity maps and time series were made for further study. To understand the spectral bands effect in the whole model, two different techniques were applied for further explanation.

AREA OF STUDY

The region of interest is shown in Fig. 1. The inset image corresponds to Argentina map, with Chaco province (pink), MAGR location (white dot) and Paraná River extension (blue line). The main image is a satellite view of the area of study, with the potabilization treatment plant (yellow triangle) and main rivers. The sample point (red star) was selected over the Barranqueras River, where the pumps take in raw water, so the spectral properties from satellite products are representative of the water treated.

From the inlet point, the water is pumped to a chamber from which its distributed to the different plant sections. In this chamber, samples are collected and delivered to the in-site laboratory to estimate a series of parameters, mainly turbidity, pH, conductivity, and alkalinity. The next step is filtration, where the sediment concentration is critical since filters and flocculant reagents are needed.

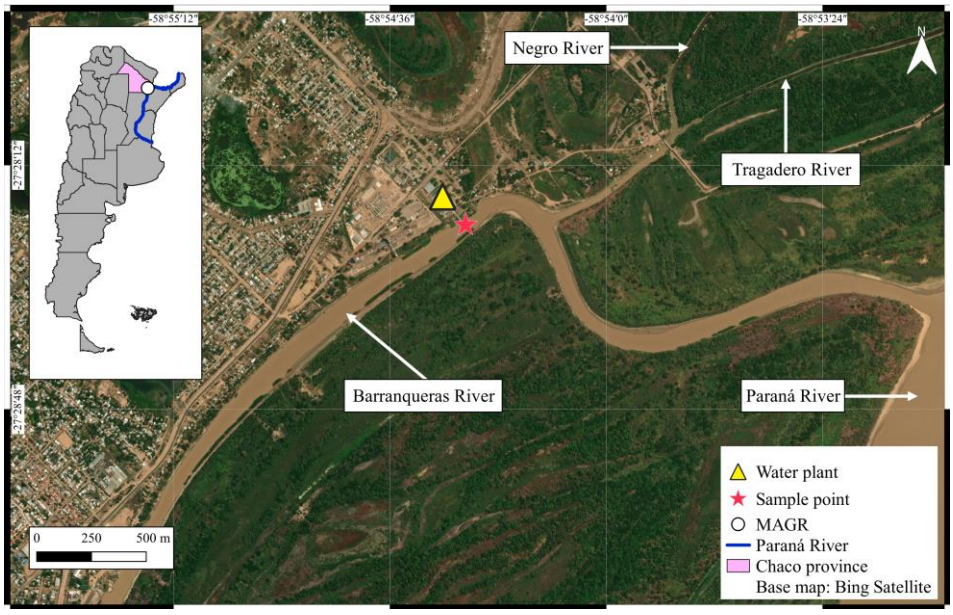

Figure 1. Region of study, indicating main rivers, water plant location and sample point. Inset: Chaco Province relative to Argentina, MAGR location and Paraná River extension.

MATERIAL AND METHODS

This section describes the source and characteristics of the laboratory and remote sensing data, and the mathematical model methodology.

Laboratory data

Daily measurements were given by the water treatment plant, from 2017-01-01 to 2021-09-03. In this time span 1732 dates were recorded. The parameters were pH, conductivity (µS/cm), alkalinity (ppm $CaCO_3$) and turbidity (NTU). Alongside these data, water samples of our own were taken to assess more sediments related parameters, such as total suspended matter (TSM), total dissolved matter (TDM), and total matter. These parameters are related since total matter is the sum of TSM and TDM. The number of samples obtained was 28, from 2021-08-24 to 2022-12-07. The measurements followed the Standard Methods techniques [19].

Remote sensing data

Satellite R_S data was obtained from S2-MSI, product L2A. Table 1 resumes the characteristics of both sensors since products of platform S2A and S2B were used. Maximum spatial resolution of 10 m (when available), 5 days revisit time, and 11 spectral bands were used. Bands B09 and B10 were discarded since no surface measurement is done at those wavelengths, 945 nm, and 1373 nm, respectively. S2-MSI has been used to estimate turbidity by band ratios [20].

S2-MSI L2A products are atmospherically corrected by Sen2Cor processor [21]. Copernicus Open Access Hub provides complete, open, and free access to S2 products. For the same period, 382 images were acquired. Using the quality assessment band, QA60, the images with clouds were discarded. This method was preferred over more complex approaches [22], since QA60 band is a coded bit mask detecting clear sky, dark clouds, and cirrus clouds.

Table 1. Spatial and spectral resolutions of platforms S2A and S2B, S2-MSI.

		S2A		S2B	
Band	Spatial resolution (m)	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)
1	60	442.7	21	442.3	21
2	10	492.4	66	492.1	66
3	10	559.8	36	559.0	36
4	10	664.6	31	665.0	31
5	20	704.1	15	703.8	16
6	20	740.5	15	739.1	15
7	20	782.8	20	779.7	20
8	10	832.8	106	833.0	106
8A	20	864.7	21	864.0	22
11	20	1613.7	91	1610.4	94
12	20	2202.4	175	2185.7	185

After removing the dates with clouds over the study area, 181 images remained to continue the analysis. The R_S values extraction was made using a 3x3 pixel window around the point near the plant water entrance, on Barranqueras River. The final pixel value was the mean of the individual values in the grid.

Modeling

The target parameter was turbidity with the spectral band as predictors, as a mathematical regression problem. As a first step, the relationship between turbidity and R_s was studied.

Two main modeling methods were used: linear, with algebraic relationships between the predictors, and a based-tree machine learning RF approach. To perform the linear modeling spectral bands and normalized difference turbidity index (NDTI) were used. NDTI is obtained by the red and green bands [23], B03 and B04, respectively. This index is used for water quality assessment and it's proportional to turbidity [24].

Machine learning techniques were used to improve traditional methods for parameter retrieval [25]. RF operates by an ensemble of decision trees [26], each one trained by a subset of the whole data, and the final regression value is the mean of all tree's individual estimates. RF can manage many predictor variables and maintain low levels of over-fitting [27], a negative aspect in modeling.

RF modeling used all spectral bands available (Table 1), since this method is appropriated to find non-linear relationships between multiple predictors. To improve the performance of RF, a tuning step was applied to obtain the best arguments required for a RF model, called hyperparameters. The tuned hyperparameters were 'min_n', the minimum number of samples taken from the dataset to form a node in a decision tree; and 'mtry', the number of predictors that will be sampled. The 'trees' hyperparameter was fixed at 1000 units. Both steps of selection of sample points and predictors are random, across all trees, so the final turbidity estimation is an average value of all tree's individual values.

The tuning process applied was made by the racing technique [28]. Tuning racing techniques evaluate the model in a subset of resamples and obtained the performance metrics, continuing

only with the hyperparameters candidates that showed good results. Usually racing techniques are faster to compute than traditional methods, such as grid search [29].

To measure the model's performance the following metrics were used: Pearson's coefficient of determination (R^2) and root mean squared error (RMSE). Following equations show the mathematical expressions for these metrics:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \quad (2)$$

Where i represents each measurement, from a total of n samples. y_i are the real turbidity values, and x_i the estimated values from the correspondent model. For R^2, Eq. 1, $\bar{y}_i$ is the mean value of y_i.

To produce a model, linear or RF, the dataset was split into two parts: 75% of the samples was used for training and tuning of the corresponding model; the remaining 25% was used for testing and to finalizing the model, that is, to get the last version of the model specification. Since the original dataset corresponded to a time series of turbidity values, the testing dataset corresponded to the most recent dates. This methodology followed the best practices for data modeling [30]. The testing dataset was preserved and was not used in the training, so the performance metrics were calculated with new and later data points. The final model was selected according to the performance metrics.

Several maps showcasing the turbidity spatial distribution were made applying the selected model to S2-MSI products, in the Barranqueras River, for four different dates. The spectral index MNDWI (modified normalized difference water index) was used to mask the water from the scene [31]. A time series of turbidity was retrieved by the selected model and compared with the training values.

RESULTS AND DISCUSSION

This section contains the main results of water characterization: the process of model selection, tuning, performance metrics and comparisons. Two techniques were applied to the final model to explain the main features. Maps were made to visualize the turbidity spatial distribution.

Water characterization

The parameters measured by the water treatment plant are shown in Fig. 2 as time series. The number of samples (n) is shown in the top right corner of each panel.

In 2019, Paraná River scarce rains and drought caused an historic low level of water [32]. This can be seen in the steady increase in turbidity and conductivity from 2019. Conductivity from the same year started to be more disperse than previous years. Turbidity presented yearly cycles, with high values at the beginning of each year, between January and April-May, then followed by a low-turbidity period.

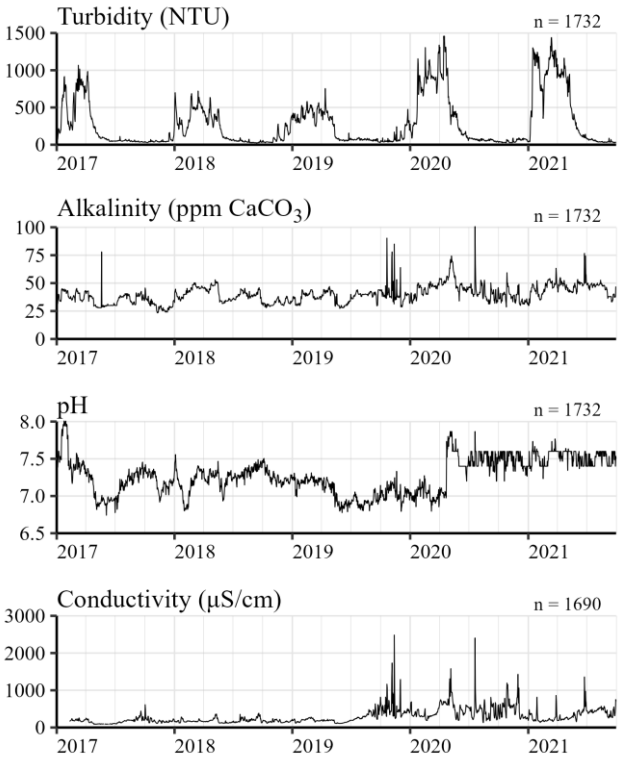

Figure 2. Time series of measured water parameters: turbidity, alkalinity, pH, and conductivity.

Main statistical values of the time series, by parameter, are resumed in Table 2. Mean, median, standard deviation (SD), initial and final sampling date and number of samples (n).

Table 2. Statistical summaries of measured water properties.

Parameter (unit)	Mean	Median	SD	Initial date	Last date	n
Turbidity (NTU)	280.7	89.8	328.2	2017-01-01	2021-09-30	1732
Alkalinity (ppm $CaCO_3$)	40.0	40.0	7.9	2017-01-01	2021-09-30	1732
pH	7.3	7.3	0.2	2017-01-01	2021-09-30	1732
Conductivity (µS/cm)	293.6	220.2	205.6	2017-02-10	2021-09-30	1690
TSM (ppm)	84.2	30.0	171.6	2021-08-24	2022-12-07	25
TDM (ppm)	195.9	166.8	100.8	2021-08-24	2022-12-07	26
Total matter (ppm)	281.9	209.5	212.1	2021-08-24	2022-12-07	25

To estimate turbidity from R_S is necessary to verify a relationship between them. Fig. 3 shows the spectral signatures of all the observations in light grey. Grouping the data points by turbidity ranges, mean spectral signatures were obtained. Lower turbidities (<150 NTU) presented the lowest R_S. As the turbidity range increased, the spectral signature response raised, until values higher than 1050 NTU.

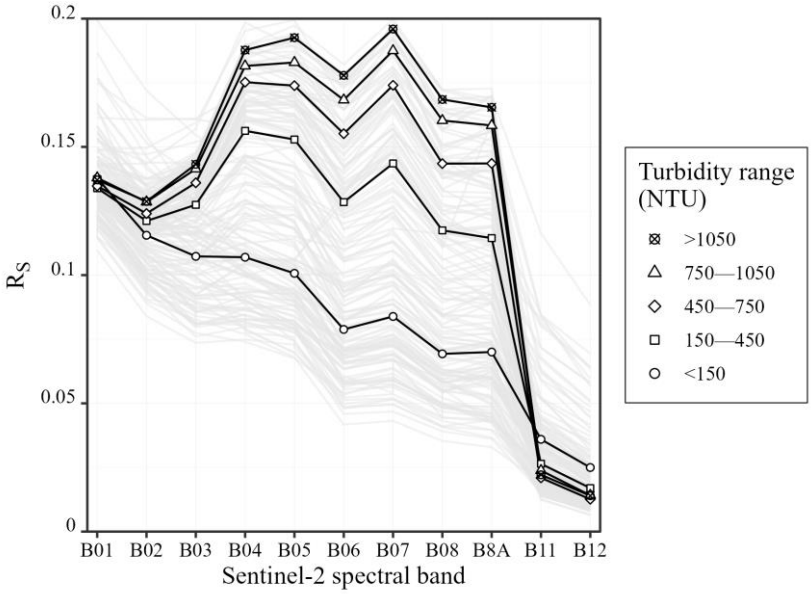

Figure 3. Mean spectral signatures per turbidity range, for S2-MSI bands.

The spectral signatures in Fig. 3, bands B01, B11 and B12 are not sensitive to the changes in turbidity, since the points for different turbidity ranges remained in the same position. Bands B05, B06 and B07 presented the highest changes. These bands are related with algorithms to estimate turbidity [5].

Model selection

Several models were tested to estimate the inlet water turbidity for the water treatment plant. The predictors variables were selected according to the model used. For RF, all S2-MSI bands (Table 1) were used as predictors. Traditional linear models used the following variables: an interaction between B06 and B07; individual bands B05, B06 and B08; and the spectral index NDTI. The spectral bands used in the linear models were selected according to the results from Fig. 3.

The characteristics and performance metrics for all the tested models are resumed in Table 3. The performance metrics were used to select the specific model, and in a later step the model is finalized using the preserved dataset for testing, with observations not used in the training.

Table 3. Regression models tested and performance metrics.

Model characteristics		Performance metrics[1]	
Specification	Expression	RMSE	R^2
RF	Turbidity ~ all bands	**111.4**	**0.840**
Linear model	Turbidity ~ B06 + B07 + B06×B07	121.9	0.802
Linear model	Turbidity ~ B08	142.9	0.736
Linear model	Turbidity ~ B06	145.7	0.732
Linear model	Turbidity ~ B05	155.8	0.693
Linear model	Turbidity ~ NDTI	218.7	0.296

[1]Unfinalized models. Final performance values are obtained only after the validation step.

Individual bands B05, B06 y B08 presented 0.693, 0.732 and 0.736 as R^2, respectively. In a similar work [5], also in turbid lakes, for the same bands presented the following values: 0.83, 0.66 and 0.63 for S2-MSI bands.

RF model is the selected model for the following results, since it presented the best performance metrics (in bold), with the lowest deviations (RMSE) and highest correlation (R^2). The interaction model between B06 and B07 was the second-best model, combining bands in the red edge, related to sediments in water [33]. Using single band linear models performed poorly in comparison. The NDTI index presented the lowest R^2 and the highest RMSE, unlike other studies [34].

The selected RF model was tuned using a racing technique [28]. The hyperparameters tuned were the number of predictors randomly sampled for each division in the trees (mtry) and the minimum number of observations required for a node to be further divided (min_n). The number of trees was set to 1000. Table 4 shows the values of the hyperparameters and the main characteristics of the final RF model.

Table 4. Tuned hyperparameters of RF model, by racing technique.

RF type	Regression
Training observations	116
Variables	11
Trees	1000
min_n	9
mtry	3

After the model selection the last fitting was performed. For the RF model, the final performance metrics were obtained by the testing dataset. These observations were kept apart so they have no influence on the modeling. The performance metrics are RMSE 136 NTU and R^2 0.918, with 39 data points. Noted that these values are different from Table 3, since those are obtained from the training data set, and are used only for model selection. Figure 4 shows the values measured by the water plant company and the estimation by the present RF model.

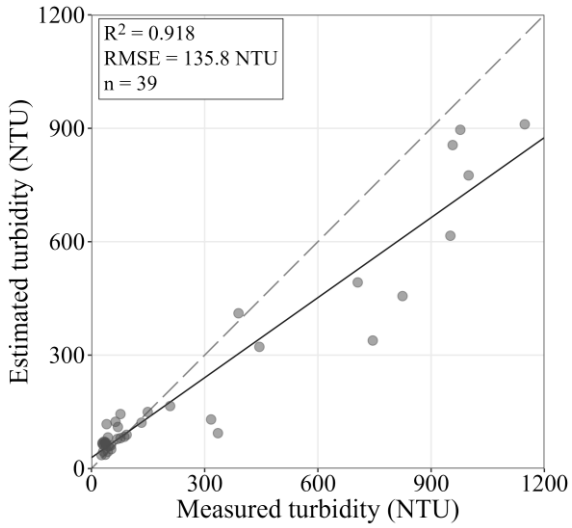

Figure 4. Estimated and measured turbidity values using the validation dataset.

The solid line represents the linear relationship between estimated and measured turbidity, with the dashed line at 45° per comparison reasons. Lower estimated turbidity values are closer to the real values. For higher turbidities, the differences increase, with estimates being lower than measured values, with the solid line below the dashed line. The outcome variable presented a wide range, with many observations under 100 NTU, and measurements as high as 1100 NTU.

The measured and estimated turbidities, from Fig. 4, are shown as a time series plot in Fig. 5. The crosses represented the estimations made by the RF model; the turbidity measures were plotted as a solid line. The number of samples in the testing dataset is shown in the top right corner.

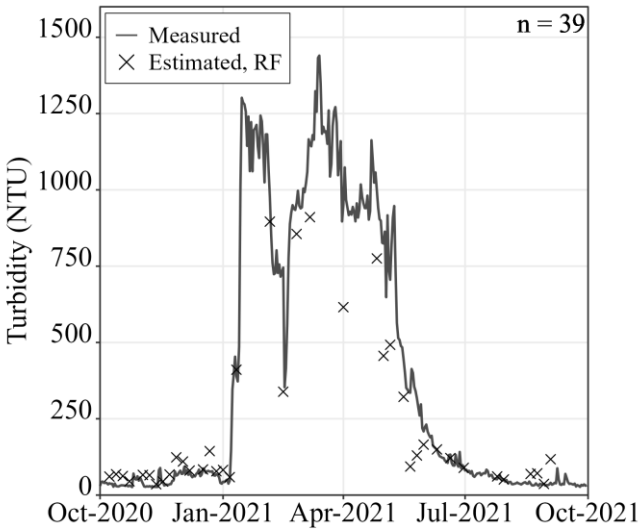

Figure 5. Time series of estimated and measured turbidities in validation dataset.

The biggest differences between estimated and measured turbidities in Fig. 5 are within the larger values, equivalent to the Fig. 4. The estimations followed the same trend seen in the turbidity time series in Fig. 2, with high turbidities in the beginning of the year, followed by lower values in the middle and end year.

Understanding random forest model

The complexity of a RF model is difficult to explain since the explicit form is not as clean as a simpler model, i.e., linear model. The **global feature importance** is an explanatory technique that assists in understanding the driving predictors variables of a RF, aggregated in the entire training observations.

The results of the global feature importance for the obtained RF model, analyzing each spectral band is shown in Fig. 6. This technique is based on the notion of the overall change in the model due to the perturbation of a specific variable [35]. A permutation-based approach is a valuable tool for model explanation, since it expected that after the permutation of said variable the model performance will decrease [26]. Spectral band B06 presents the most effect in the model, according to Fig. 6, since the boxplot has the highest RMSE (105.5 NTU). Close to B06 is B07, B08 and B05. Spectral bands B05 [36] and B08 [20] have been reported to be related to turbidity. The least effects are given by B02, B03 and B01, since the perturbation of these bands have a much lesser impact on the overall model. The vertical dashed line represents the base RMSE.

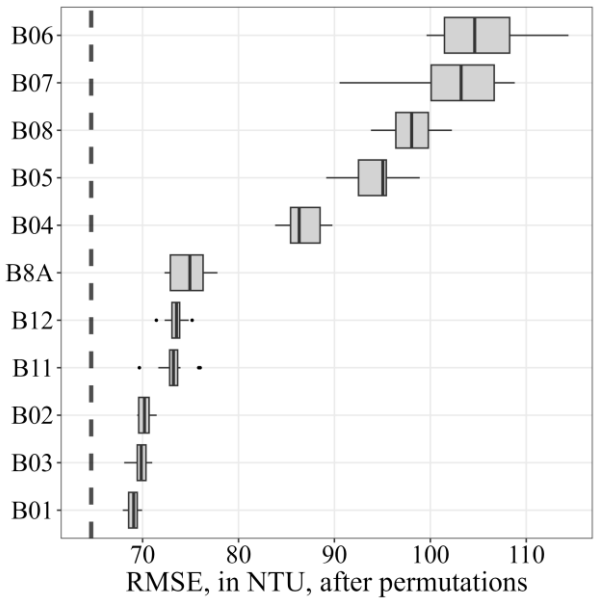

Figure 6. Global feature importance of S2-MSI spectral bands in RF model.

Partial dependencies profiles allowed to show the change in the expected value of a model estimate alongside a single explanatory variable [35]. According to the global feature importance technique, B06 is the spectral band that has the highest effect on the RF model. For this band the partial dependencies profile was obtained and is shown in Fig. 7.A.

The grey lines in Fig. 7 correspond to 100 randomly selected observations from the training dataset. The thick line indicates the mean. The effect of B06 (Fig. 7.A) on turbidity estimates is constant until 0.12 R_S, then starts to increase until its highest effect at around 0.18 R_S. In this range of reflectance, the change in turbidity presented was from 214.2 NTU to 377.9 NTU. For comparison, Fig. 7.B corresponds to the same analysis for B01, the band with the lowest feature importance as mentioned above (Fig. 6). The partial dependency profile of B01 is constant, that is, the turbidity presented no change in the entire range of B01 R_S values. This result is consistent with Fig. 3 and Fig. 6.

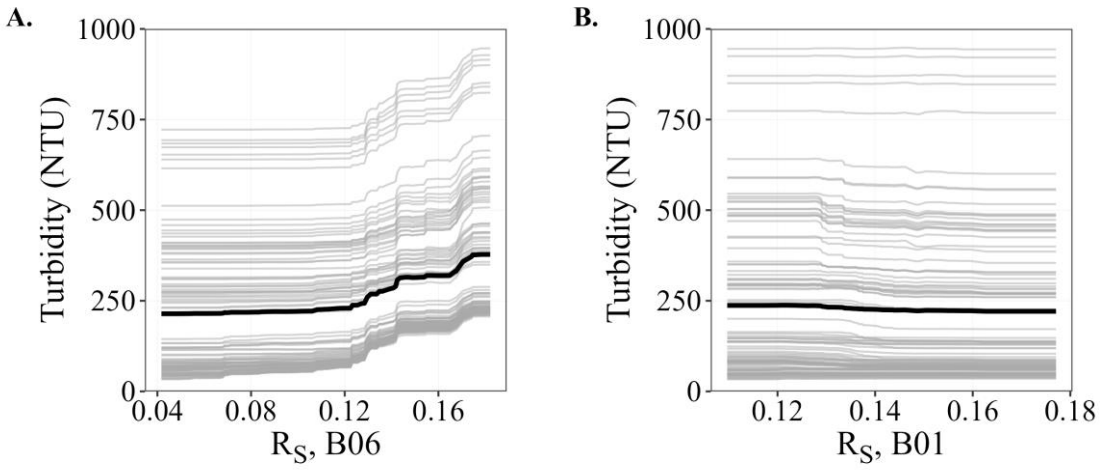

Figure 7. Partial dependencies profiles for spectral bands B06 and B01.

Turbidity maps

The obtained RF model was applied to the spectral values from Barranqueras River to evaluate the spatial turbidity distribution. Maps for four different dates from 2020 are shown in Fig. 8.

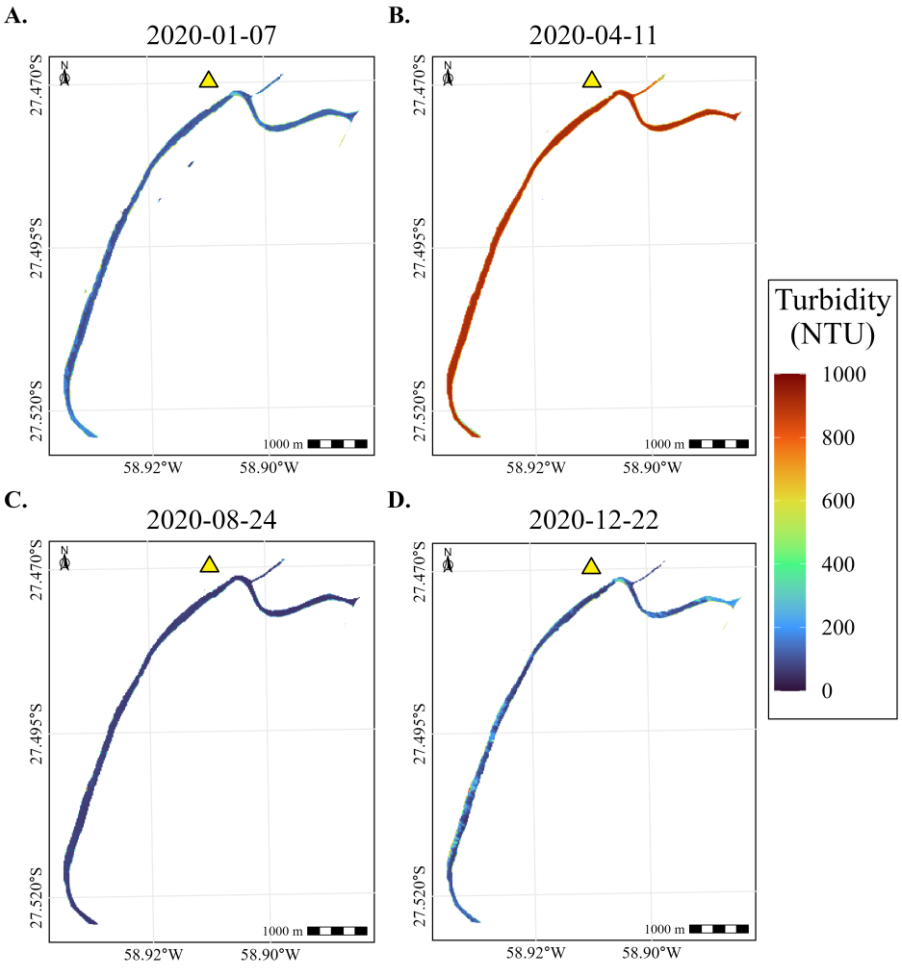

Figure 8. Barranqueras River turbidity maps for four different dates.

The yellow triangle on the top-center of each panel represents the water plant building. A water mask was applied to the region of interest to only extract pixel values from the Barranqueras River. The mask was made using the MNDWI spectral index [37].

The Barranqueras River presented a homogeneous spatial turbidity, with low dispersion in the values. For any given date in Fig. 8, the range of turbidities was narrow, so no large change of color was seen.

Map A (2020-01-07) in Fig. 8 presented relatively lower values, below 200 NTU. Map B (2020-04-11) had the highest turbidities, above 900 NTU. Then, map C (2020-08-24) with the lowest values, under 100 NTU. Finally, map D (2020-12-22) had an increased turbidity, around 100 NTU. These estimated turbidity maps followed the same measured turbidity trend seen in Fig. 2 for the year 2020.

CONCLUSION

Inlet water properties are an important input in a water potabilization plant, to set the filtration process and the reagents needed for the flocculation step. Specifically, water turbidity is a valuable parameter for the decision-making process.

In Resistencia, Chaco province in Argentina, the inlet water properties of the local water treatment plant were studied as a time series. Annual turbidity cycles were observed, with high values between January to April-May, and lower values for the rest of the year. Combining with satellite optical data, a RF model was developed and tuned to retrieve water turbidity from R_S, measured by MSI sensor in S2 platform.

Several linear and machine learning models were tested for turbidity estimation, with the spectral bands as predictor variables. A tuned RF model presented the highest performance metrics, with 0.918 R^2 and 135.8 NTU RMSE. Applying global feature importance technique to the RF model, band B06 was stablished as the most important variable. The partial dependence profile for B06 indicated the highest change in the outcome variable. The highest turbidity measures presented the biggest differences between measured and estimated turbidities.

The maps generated from the RF applied to S2-MSI products follow the same trend as the measured turbidities for the year 2020.

Remote sensing techniques, and machine learning modeling, allowed the development of a RF model, with high performance metrics, to estimate water turbidity for the contribution of the vital process of water treatment.

REFERENCES

[1] United Nations, "2030 Agenda for Sustainable Development," vol. 11371, no. July, pp. 1–13, 2017, doi: 10.1109/TNSRE.2015.2480755.

[2] Y. Du *et al.*, "Total suspended solids characterization and management implications for lakes in East China," *Science of the Total Environment*, vol. 806, Feb. 2022, doi: 10.1016/j.scitotenv.2021.151374.

[3] W. G. Buma and S. Il Lee, "Evaluation of Sentinel-2 and Landsat 8 images for estimating Chlorophyll-a concentrations in Lake Chad, Africa," *Remote Sens (Basel)*, vol. 12, no. 15, Aug. 2020, doi: 10.3390/RS12152437.

[4] G. Rodrigues *et al.*, "Temporal and spatial variations of secchi depth and diffuse attenuation coefficient from sentinel-2 MSI over a large reservoir," *Remote Sens (Basel)*, vol. 12, no. 5, Mar. 2020, doi: 10.3390/rs12050768.

[5] Y. Ma *et al.*, "Remote sensing of turbidity for lakes in Northeast China using sentinel-2 images with machine learning algorithms," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 14, pp. 9132–9146, 2021, doi: 10.1109/JSTARS.2021.3109292.

[6] M. Pereira-Sandoval *et al.*, "Evaluation of atmospheric correction algorithms over spanish inland waters for sentinel-2 multi spectral imagery data," *Remote Sens (Basel)*, vol. 11, no. 12, pp. 1–23, 2019, doi: 10.3390/rs11121469.

[7] Z. Cao *et al.*, "A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes," *Remote Sens Environ*, vol. 248, no. July, p. 111974, Oct. 2020, doi: 10.1016/j.rse.2020.111974.

[8] I. Barut, H. Keskin-Citiroglu, M. Oruc, and A. M. Marangoz, "Determination by Landsat Satellite Imagery to Local Scales in Land and Pollution Monitoring: a Case of Buyuk

Melen Watershed (Turkey)," *Journal of Sustainable Development of Energy, Water and Environment Systems*, vol. 3, no. 4, pp. 389–404, Dec. 2015, doi: 10.13044/j.sdewes.2015.03.0029.

[9] C. S. Carrión *et al.*, "Multi-Temporal Analysis of the Glacier Retreat Using Landsat Satellite Images in the Nevado of the Ampay National Sanctuary, Peru," *Journal of Sustainable Development of Energy, Water and Environment Systems*, vol. 10, no. 1, Mar. 2022, doi: 10.13044/j.sdewes.d8.0380.

[10] C. Giardino *et al.*, "The Color of Water from Space: A Case Study for Italian Lakes from Sentinel-2," in *Earth Observation and Geospatial Analyses [Working Title]*, IntechOpen, 2019. doi: 10.5772/intechopen.86596.

[11] M. H. Tavares, R. C. Lins, T. Harmel, C. R. Fragoso, J. M. Martínez, and D. Motta-Marques, "Atmospheric and sunglint correction for retrieving chlorophyll-a in a productive tropical estuarine-lagoon system using Sentinel-2 MSI imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 174, no. April 2020, pp. 215–236, 2021, doi: 10.1016/j.isprsjprs.2021.01.021.

[12] K. Toming, T. Kutser, A. Laas, M. Sepp, B. Paavel, and T. Nõges, "First experiences in mapping lakewater quality parameters with sentinel-2 MSI imagery," *Remote Sens (Basel)*, vol. 8, no. 8, pp. 1–14, 2016, doi: 10.3390/rs8080640.

[13] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sensing*, vol. 12, no. 14. MDPI AG, Jul. 01, 2020. doi: 10.3390/rs12142291.

[14] J. Gorroño, A. C. Banks, N. P. Fox, and C. Underwood, "Radiometric inter-sensor cross-calibration uncertainty using a traceable high accuracy reference hyperspectral imager," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 393–417, Aug. 2017, doi: 10.1016/j.isprsjprs.2017.07.002.

[15] A. A. Bonetto, "1. The Parana River system," 1986.

[16] S. N. Lane, D. R. Parsons, J. L. Best, O. Orfeo, R. A. Kostaschuk, and R. J. Hardy, "Causes of rapid mixing at a junction of two large rivers: Río Paraná and Río Paraguay, Argentina," *J Geophys Res Earth Surf*, vol. 113, no. 2, Jun. 2008, doi: 10.1029/2006JF000745.

[17] Instituto Nacional de Estadística y Censos (Argentina), *Censo Nacional de Población, Hogares y Viviendas 2022: Resultados provisionales / c Instituto Nacional de Estadística y Censos (Argentina)*. 2022.

[18] J. J. Neiff, A. S. G. Poi De Neiff, and S. L. Casco, "Importancia ecológica del Corredor Fluvial Paraguay-Paraná como contexto del manejo sostenible," *Humedales fluviales de América del Sur*, pp. 193–210, 2005.

[19] R. B. Baird, C. E. W. Rice, and A. D. Eaton, *Standard Methods for the Examination of Water and Wastewater, 23rd*, no. 1. Water Environment Federation, American Public Health Association, American Water Works Association, 2017. doi: https://doi.org/10.2105/SMWW.2882.003.

[20] N. M. Hussein, M. N. Assaf, and S. S. Abohussein, "Sentinel 2 analysis of turbidity retrieval models in inland water bodies: The case study of Jordanian dams," *Can J Chem Eng*, vol. 101, no. 3, pp. 1171–1184, Mar. 2023, doi: 10.1002/cjce.24526.

[21] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, "Sen2Cor for Sentinel-2," in *Image and Signal Processing for Remote Sensing XXIII*, L. Bruzzone, F. Bovolo, and J. A. Benediktsson, Eds., SPIE, Oct. 2017, p. 3. doi: 10.1117/12.2278218.

[22] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "AGGREGATING CLOUD-FREE SENTINEL-2 IMAGES WITH GOOGLE EARTH ENGINE," *ISPRS Annals of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2/W7, no. 2/W7, pp. 145–152, Sep. 2019, doi: 10.5194/isprs-annals-IV-2-W7-145-2019.

[23] J. P. Lacaux, Y. M. Tourre, C. Vignolles, J. A. Ndione, and M. Lafaye, "Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal," *Remote Sens Environ*, vol. 106, no. 1, pp. 66–74, Jan. 2007, doi: 10.1016/j.rse.2006.07.012.

[24] X. Chen, W. Chen, Y. Bai, and X. Wen, "Changes in turbidity and human activities along Haihe River Basin during lockdown of COVID-19 using satellite data," *Environmental Science and Pollution Research*, vol. 29, no. 3, pp. 3702–3717, Jan. 2022, doi: 10.1007/s11356-021-15928-6.

[25] S. Magrì, E. Ottaviani, E. Prampolini, B. Federici, G. Besio, and B. Fabiano, "Application of machine learning techniques to derive sea water turbidity from Sentinel-2 imagery," *Remote Sens Appl*, p. 100951, Apr. 2023, doi: 10.1016/j.rsase.2023.100951.

[26] L. Breiman, "Random Forests," 2001. doi: https://doi.org/10.1023/A:1010933404324.

[27] A. B. Ruescas, M. Hieronymi, G. Mateo-Garcia, S. Koponen, K. Kallio, and G. Camps-Valls, "Machine learning regression approaches for colored dissolved organic matter (CDOM) retrieval with S2-MSI and S3-OLCI simulated data," *Remote Sens (Basel)*, vol. 10, no. 5, May 2018, doi: 10.3390/rs10050786.

[28] O. Maron and A. W. Moore, "Hoeffding Races: Accelerating Model Selection Search for Classification and Function Approximation."

[29] M. Kuhn, "Futility Analysis in the Cross-Validation of Machine Learning Models," May 2014, [Online]. Available: http://arxiv.org/abs/1405.6974

[30] M. Kuhn and J. Silge, *Tidy modeling with R*, 1st ed. O'Reilly Media, Inc., 2022. Accessed: Jul. 25, 2023. [Online]. Available: https://www.tmwr.org/

[31] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int J Remote Sens*, vol. 27, no. 14, pp. 3025–3033, Jul. 2006, doi: 10.1080/01431160600589179.

[32] S. Montico, N. C. Di Leo, and J. A. Berardi, "Sequía, bajante y efectos de los incendios en suelos del delta del Paraná, Argentina," *Cuadernos del CURIHAM*, May 2023, doi: 10.35305/curiham.vi.199.

[33] T. S. Rahul, J. Brema, and G. J. J. Wessley, "Evaluation of surface water quality of Ukkadam lake in Coimbatore using UAV and Sentinel-2 multispectral data," *International Journal of Environmental Science and Technology*, vol. 20, no. 3, pp. 3205–3220, Mar. 2023, doi: 10.1007/s13762-022-04029-7.

[34] M. Elhag, I. Gitas, A. Othman, J. Bahrawi, and P. Gikas, "Assessment of water quality parameters using temporal remote sensing spectral reflectance in arid environments, Saudi Arabia," *Water (Switzerland)*, vol. 11, no. 3, Mar. 2019, doi: 10.3390/w11030556.

[35] P. Biecek and T. Burzykowski, *Explanatory Model Analysis*. New York: Chapman and Hall/CRC, 2021. Accessed: Jul. 20, 2023. [Online]. Available: https://pbiecek.github.io/ema/

[36] M. Chowdhury, C. Vilas, S. van Bergeijk, G. Navarro, I. Laiz, and I. Caballero, "Monitoring turbidity in a highly variable estuary using Sentinel 2-A/B for ecosystem management applications," *Front Mar Sci*, vol. 10, Jul. 2023, doi: 10.3389/fmars.2023.1186441.

[37] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int J Remote Sens*, vol. 27, no. 14, pp. 3025–3033, 2006, doi: 10.1080/01431160600589179.