Abstract: This paper introduces a discussion about the need for atmospheric corrections by comparing data-driven sea surface salinity (SSS) derived from Top- and Bottom-of-Atmosphere imagery. Atmospheric corrections are used to remove the effect of the atmosphere in reflectances acquired by satellite sensors. The Sentinel-2 Level-2A product provides atmospherically corrected Bottom-of-Atmosphere (BOA) imagery, derived from Level-1C Top-of-Atmosphere (TOA) tiles using the Sen2Cor processor. SSS at high resolution in coastal areas (100 m) is derived from multispectral signatures using artificial neural networks. Four scenarios with different input variables are tested for both TOA and BOA imagery, for interpolation (previous information on all platforms is available in the training dataset) and extrapolation (certain platforms are isolated and the network does not have any previous information on these) problems. Results show that TOA always outperforms BOA in terms of higher coefficient of determination ($R^2$), lower mean absolute error (MAE) and lower most common error ($\mu_e$). The best TOA results are $R^2 = 0.99$, $MAE = 0.4$ PSU and $\mu_e = 0.2$ PSU. Moreover, the evaluation of the neural network in all the pixels of Sentinel-2 tiles shows that BOA results are accurate only far away from the coast, while TOA data provides useful information on nearshore mixing patterns, estuarine processes and is able to estimate freshwater salinity values. This suggests that land adjacency corrections could be a relevant source of error. Sun glint corrections appear to be another source of error. TOA imagery is more accurate than BOA imagery when using machine learning algorithms and big data, as there is a clear loss of information in the atmospheric correction process that affects the multispectral–in situ relationships. Finally, the time and computational resources gained by avoiding atmospheric corrections can make the use of TOA imagery interesting in future studies, such as the estimation of chlorophyll or coloured dissolved organic matter.

Keywords: Sentinel-2; atmospheric corrections; Top of Atmosphere (TOA); Bottom of Atmosphere (BOA); Sea Surface Salinity (SSS); artificial neural network (ANN); coastal oceanography

1. Introduction

The Sentinel-2 mission is a constellation of two multispectral polar-orbiting satellites placed in the same sun-synchronous orbit, phased 180° to each other at a mean altitude of 786 km [1]. Sentinel-2 presents a swath width of 290 km and a revisit time of 5 days in the equator. The main mission objectives are the systematic global acquisition of high-resolution, multispectral images linked to a high revisit frequency; the continuity of multispectral imagery provided by the SPOT series of satellites and the LANDSAT Thematic Mapper instrument; and the provision of observation data for
the next generation of operational products, such as land cover maps, land change detection maps and geophysical variables [2].

The Sentinel-2 multispectral instrument (MSI) collects rows of data across the orbital swath and uses the forward motion of the spacecraft along the path of the orbit to acquire new rows. The MSI covers visible and near-infrared (VNIR) and short-wave infrared (SWIR) with 12 detectors staggered in two rows. The 13-spectral band configuration of the mission arose as a result of consultation with the user community during the design phase. The existing Copernicus Service Elements (CSEs) services were developed around the use of LANDSAT and SPOT wavelengths, and the service requirements for Sentinel-2 have been based on this [2]. The bands cover optical (red, green and blue), NIR, SWIR, red-edge, aerosols, water vapour and cirrus clouds. Narrowing the width of the Sentinel-2 spectral bands limits the influence of atmospheric constituents, as it was observed that the original LANDSAT NIR (near-infrared) band was highly contaminated by water vapour [2]. The narrowness of the 8a band at 865 nm in the NIR is designed to avoid contamination from water vapour. Precise aerosol correction of acquired data is enabled by the inclusion of a spectral band in the blue domain at 443 nm (Band 1) [2].

There are two data products available for users from Sentinel-2, Level-1C (L1C) Top-of-Atmosphere (TOA), and Level-2A (L2A) Bottom-of-Atmosphere (BOA), which provides the equivalent surface-leaving reflectances to those of L1C. In both L1C and L2A products, the granules (or tiles) are orthorectified images of 100 km × 100 km in UTM/WGS84 projection. Both are resampled with a constant Ground Sampling Distance (GSD) of 10, 20 and 60 m, depending on the original resolution of the different spectral bands. The products also contain a cloud mask. The L2A BOA imagery is obtained following a set of atmospheric corrections from L1C imagery.

The atmosphere is a composition of different gases and suspended solids and liquids. It is primarily composed of nitrogen, oxygen and some inert gases, which account for ~99% of the atmospheric composition [3]. The rest is water vapour and carbon dioxide in different concentrations depending on the area. The solid and liquid particles suspended in the atmosphere are called aerosols. Electromagnetic waves interact with the atmosphere, changing the way different satellites “see” the Earth surface depending on aerosols, angle of view, Sun incidence angle and albedo, among others. The radiative transfer equation expresses the radiance as a function of these parameters. Atmospheric correction consists of two parts: the estimation of atmospheric parameters and the retrieval of surface reflectance. Reflectance retrieval is based mainly on the correction of aerosol and water vapour impacts, as well as reflection effects. The most commonly used radiative transfer model is MODTRAN [4]. MODTRAN has been adapted for wide use as part of interface–based software packages, such as FLAASSH (Fast Line-of-sight Atmospheric Analysis of Hypercubes), ATCOR (Atmospheric and Topographic Correction) or ACORN (Atmospheric Correction Now). These three software packages have similar features, and include a set of look-up tables for a fast atmospheric correction [3]. As atmospheric corrections are mostly empirical, these software packages differ in the use made of certain variables, like thermal regions; terrain corrections; and other correction features, such as aerosols, land adjacency or cloud removal, among others [5].

Sentinel-2 L2A uses algorithms of scene classification and atmospheric correction from the L1C product. L2A products are generated automatically by its operational processor, but can be generated by users using the Sen2Cor processor [6]. The Sen2Cor algorithm performs the atmospheric, terrain and cirrus correction of TOA data. Sen2Cor delivers BOA terrain and cirrus corrected reflectance tiles, as well as information on aerosol optical thickness (AOT), water vapour (WV), scene classification and quality indicators for cloud and snow probabilities [7]. The processing starts with scene classification, followed by AOT and WV retrieval, finishing with TOA to BOA conversion [8]. Moreover, the implemented algorithm requires dense dark vegetation pixels in the image for AOT retrieval, which presents a problem in ocean tiles.

Sen2Cor is based on the ATCOR algorithm, including its terrain, adjacency and empirical bidirectional reflectance distribution function (BRDF) correction, see Figure 1. ATCOR is a method
used to reduce atmospheric and illumination effects on satellite imagery to retrieve variables such as atmospheric conditions, thermal and atmospheric radiance and transmittance functions in order to simulate the simplified properties of a 3D atmosphere, [9].

\[ G = \left( \frac{\cos \beta_i}{\cos \beta_T} \right)^b \]  

where \( \beta_i \) is the local zenith angle (obtained from the tile metadata), \( \beta_T \) is the threshold for surface reflectance and \( b \) is an exponent that can take different values (0, 1, 1/2, 1/3 or 3/4). The most appropriate values of the exponent \( b \) are recommended as 3/4 for channels with \( \lambda < 720 \) nm, and \( b = 1/3 \) for channels with \( \lambda > 720 \) nm. The threshold illumination value \( \beta_T \) is related to the solar zenith angle \( \theta_s \) as

- if \( \theta_s < 45^\circ \) → \( \beta_T = \theta_s + 20^\circ \),
- if \( 45^\circ \leq \theta_s \leq 55^\circ \) → \( \beta_T = \theta_s + 15^\circ \) and
- if \( \theta_s > 55^\circ \) → \( \beta_T = \theta_s + 10^\circ \). 

The geometric function \( G \) also needs a lower boundary, \( g \), to prevent strong reductions in reflectance. \( g \) is advised to take values in between 0.2 and 0.25. \( G \) also has an upper boundary of 1, so \( 0.2 \leq G \leq 1 \). Any values of \( G \) above or below those parameters will be automatically set to the nearest boundary. \( G \) is then used to correct the reflectance following the expression

\[ \rho_S = \rho_L G \]  

where \( \rho_L \) is the isotropic (Lambert) reflectance.

The biggest issue presented by the empirical BRDF correction is the fact that it has been tested for land and vegetation, but there are no indicative values for water. While Sen2Cor was developed for tiles over land, it can be applied over water surface using the values estimated over land pixels.

**Figure 1.** Sen2Cor processing scheme, adapted from the work in [8].
in the image. However, the processor does not contain any considerations of water surface effects like sun glint [10]. Consequently, the BOA BRDF correction values over open ocean and coastal waters are not completely reliable. In contrast, the TOA values may present BRDF effects, but they are unaltered and thus may be more reliable in terms of training machine learning algorithms using big amounts of satellite data, i.e., the algorithm might be able to learn sun glint pattern behaviours. Another relevant issue for coastal waters is the land adjacency effect. Pixels affected by adjacency effects have a water-leaving reflectance spectrum with a different shape to the reference spectrum [12]. This deviation is used as a measure of the adjacency effect. For Sentinel-2, reflectance is only corrected for adjacency influence at the end of the process, and the correction factor is directly proportional to the ratio of diffuse to direct ground-to-sensor transmittance [13]. This is applied in all neighbouring pixels, obtaining once again an approximation that implies the loss of data to an empirical process.

In terms of sea surface salinity estimation from remote sources, various satellite missions have focused on sea surface salinity. The European Space Agency’s SMOS mission (Soil Moisture Ocean Salinity) uses its Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) to provide salinity in the ocean with a spatial resolution of 35 km at the centre of field of view [14]. NASA’s Aquarius mission also produced salinity with a spatial resolution of 150 km [15]. The mission lasted 3 years, producing a global scale salinity product by using radiometers to detect changes in the oceans microwave thermal emissions frequencies due to salinity. The low resolution of available satellite missions on sea surface salinity motivates the need to explore other alternatives to derive high-resolution products, particularly relevant in coastal areas. Previous research showed an existing empirical relationship between salinity and reflectance ratio (blue/green) in the Zambezi estuary in Mozambique, in the Indian Ocean [16]. Other papers, such as those in [17–20], demonstrated that there is a relationship between yellow substance (the optically active component of Dissolved Organic Carbon) and salinity (which has no direct colour signal, and thus tracers are needed [21]). These research papers focused in different seas worldwide, including waters off the west coast of Ireland, the Baltic Sea, the North Sea and the Firth of Clyde in the Atlantic Ocean, supporting the general applicability of the method independently of the type of water and location.

This paper explores the differences in data-driven methods using L1C and L2A Sentinel-2 imagery to estimate sea surface salinity (SSS) in coastal areas. SSS estimation from Sentinel-2 L1C TOA data in [22] followed research from in [23] to estimate SSS from multispectral sources, including high-resolution results, worldwide coverage for coastal areas and independence from sea temperature in the SSS estimation. Results in [22] showed a good agreement between multispectral properties and salinity content, with a coefficient of determination above 80% and most common errors around 0.4 PSU. The present paper takes the next step by providing a comparison between SSS derived by atmospherically corrected and TOA imagery, aiming at starting a discussion about the need for atmospheric corrections when machine learning is used to link satellite and in situ data. The paper starts with a description of the methodology, use of in situ salinity data as well as TOA and BOA imagery, and matching between these. The proposed neural network is described, and results, including a table with detailed parameters used and performance metrics, are provided. A discussion section is included at the end, where the optimal models for both L1C and L2A imagery are applied on tiles in three locations: Kuwait Bay, the mouth of the Amazon river and Canterbury Bight. A discussion on the differences between TOA and atmospherically corrected results is provided, and the paper finishes with a summary of findings and conclusions.

2. Methodology

2.1. Sentinel-2 Level-1C and Level-2A Imagery

As described in the introduction, the Sentinel-2 L2A data is processed from L1C and provided online in the Copernicus Open Access Hub [24], but can also be obtained by users using the Sentinel-2 Toolbox [1]. On 26 March 2018, an evolution of the L2A products was released over the
Euro-Mediterranean region. The pilot Level-2A products had been distributed since 2 May 2017, and are published on the Copernicus Open Access Hub 48–60 h after the publication of their corresponding L1C product. Table 1 presents the bands for L1C and L2A. Note that L2A does not contain band B10 for cirrus clouds. Both B10 and the cloud mask band QA60 have not been used as inputs for the neural network to ensure the same information is available for both L1C and L2A products. Table 2 presents the Sentinel-2 metadata used in this study.

Table 1. Sentinel-2 band information for level 1C and 2A. (*) Bands QA60 and B10 have not been used as input for the neural network.

| Level-1C | Level-2A | Resolution (m) | Wavelength (nm) | Description |
|----------|----------|----------------|-----------------|-------------|
| B1       | B1       | 60             | 443.9           | Aerosols    |
| B2       | B2       | 10             | 496.6           | Blue        |
| B3       | B3       | 10             | 560             | Green       |
| B4       | B4       | 10             | 664.5           | Red         |
| B5       | B5       | 20             | 703.9           | Red Edge 1  |
| B6       | B6       | 20             | 740.2           | Red Edge 2  |
| B7       | B7       | 20             | 782.5           | Red Edge 3  |
| B8       | B8       | 10             | 835.1           | NIR         |
| B8a      | B8a      | 20             | 864.8           | Red Edge 4  |
| B9       | B9       | 60             | 945             | Water vapour |
| B10 (*)  | -        | 60             | 1373.5          | Cirrus      |
| B11      | B11      | 20             | 1613.7          | SWIR1       |
| B12      | B12      | 20             | 2202.4          | SWIR2       |
| QA60 (*) | QA60     | 60             | -               | Cloud mask  |

Table 2. Sentinel-2 metadata used in as inputs for the neural network.

| Data                                | Description                                                                 |
|-------------------------------------|-----------------------------------------------------------------------------|
| Cloud pixel percentage             | Granule-specific cloudy pixel percentage.                                   |
| Cloud coverage assessment          | Cloudy pixel percentage for the whole archive.                              |
| Mean Incident Azimuth angle for every band (×12 bands) | Mean value containing viewing incidence azimuth angle average for each band. |
| Mean Incident Zenith angle for every band (×12 bands) | Mean value containing viewing incidence zenith angle average for each band. |
| Mean Solar Azimuth angle            | Mean value containing sun zenith angle average for all bands.               |
| Reflectance conversion correction  | Earth–Sun distance correction factor.                                       |

2.2. Copernicus Marine Environmental Monitoring Service In Situ Data

The methodology follows that introduced in [22]. In situ data have been downloaded from the Copernicus Marine Environmental Monitoring Service (CMEMS) [25]. Data from Global Ocean, Arctic Ocean, Baltic Sea, European North-West Shelf Seas, Iberia—Biscay—Ireland regional Seas, Mediterranean Sea and Black Sea have been used. Data for the Near-Real-Time (NRT) component of datasets with SSS were used, including a preselection of the data with the best Quality Checks (QC). NRT products are updated with new observations at a maximum daily frequency, depending on connection capabilities in the platform. The data is collected from main global networks (Argo, GOSUD, OceanSITES, and World Ocean Database) completed by European data provided by EUROGOOS regional systems and national system by the regional in situ components [25], and the products are delivered by authenticated FTP. Figure 2 shows the distribution of platforms measuring SSS from the period May 2017 (start of L2A availability in the Copernicus Open Access Hub) to May 2020. The information in those platforms was downloaded and filtered to look for matches with satellite passing times. There is a higher concentration of platforms around European waters, as well as North America and Japan.
Figure 2. Map of the available platforms since May 2017 containing sea surface salinity data. Blue rectangles shows validation areas in this study, see Section 3. Green rectangles show evaluation areas (see Section 4).

Please note that the dataset used in this paper contains salinity in both coastal and open ocean waters. Research by the authors of [16–20] proved that the relationship between reflectance and salinity is applicable to different types of waters and locations worldwide. Moreover, machine learning algorithms need big amounts of information, which brings the need to include open ocean data in order to obtain accurate coastal values; otherwise, it would be impossible to train a neural network only with the available coastal data.

2.3. Satellite–In Situ Matching Process and Neural Network Approach

The information provided for L2A in the Copernicus Open Access Hub is less extensive than that for L1C. Although L1C has been available since 2015, the L2A processing did not commence until 2017, and in many areas data is not available until late 2018. To ensure correlation between datasets for both levels, and that the same data is used to train the neural networks for L1C and L2A, only images after May 2017 were matched with in situ data for both L1C and L2A, which gives a 3-year data coverage. The matched L1C and L2A datasets were reviewed in a second iteration to ensure the same information was present in both. A total of approximately 2700 points (of an initial batch of about 25,000, pre-filtering) with global coverage were finally used for the May 2017–2020 period. This amount of information is less than that used in previous work (see, e.g., in [22]), however the optimal tuning of the neural network presented in this paper provides better results, even with the reduced a dataset.

The matching process followed here is the same as that presented in [22], with the addition of the extra data cleansing to ensure the same information is used for L1C and L2A. The process was implemented in Python, via the Google Colab platform and Google Earth Engine [26,27]. A summary of the steps taken is given below.

1. In situ data containing salinity since May 2017 to 2020 (i.e., 3 years of data, linked to the Sentinel-2 L2A availability) is downloaded from the Copernicus Marine In Situ data portal [24]. Data are extracted from the Global component, but also from the different seas: Arctic, Baltic, Black Sea, Iberian–Biscay–Ireland, Mediterranean and Northwest Shelf seas.
2. For each in situ point coordinate, Sentinel-2 L1C and L2A image collections are filtered to the tiles that contain the point on the day and time when the measurement was taken. The image is only considered if the in situ measurement was taken within 1 hour of the Sentinel-2 pass time.
3. If there are any valid tiles for that point, these are clipped in sections of area 100 m × 100 m, centred in the point location to obtain high-resolution estimators of SSS.
4. For each tile section, properties and band data summarised in Tables 1 and 2 are extracted by reducing the properties in the area to their average value.
5. The time difference between the in situ measurement and the satellite image is recorded. In case of multiple tiles covering the point of interest, the matched data is sorted by time difference, and the match with the smallest time difference is selected.
6. A table containing satellite data (band information and metadata) and equivalent SSS in situ information for each valid point is composed for both L1C and L2A collections.
7. Band QA60 containing a cloud mask has been used as a filter to select points only with a clear sky (i.e., points were clouds are persistent have not been considered: no opaque clouds or cirrus clouds are present).
8. Duplicates are dropped.
9. Matching datasets for L1C and L2A are compared and filtered to ensure the same information is available for both.
10. Outlier removal: any values outside a range of ±3 standard deviations are not considered. Assuming data follows a normal distribution, any data points in the tail of the distribution over 3 standard deviations from the mean represent ~0.1% of the information.
11. Data normalisation is conducted using $X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$, where $X_{\text{norm}}$ is the normalised value, $X$ is the original value, $X_{\text{min}}$ is the minimum value of the normalised vector and $X_{\text{max}}$ is the maximum value of the normalised vector. Normalised data is fed to the neural network.

A neural network has been used to establish relationships between SSS and multispectral properties of the water. As described in the introduction, previous research showed an existing empirical relationship between salinity and reflectance ratio (blue/green) [16]. The neural network introduced in [22] strengthened that theory. The architecture of the neural network is derived from the one presented in [22], see Figure 3: a deep neural network with shortcuts is used to avoid the vanishing gradient problem, as demonstrated in [22]. Residual networks avoid this issue by skipping connections, or jumping layers, see Figure 3. By doing this, previous activations are reused, adapting the weights of adjacent layers [28]. The input layer is composed of 40 variables equivalent to the items summarised in Tables 1 and 2, combining band information and image metadata. The 40 input variables are band data (B1, B2, B3, B4, B5, B6, B7, B8, B8a, B9, B11, B12), Cloud pixel percentage, Cloud coverage assessment, Mean incident azimuth angle for each band (12 input values), Mean incident zenith angle for every band (12 input values), Mean solar azimuth angle and Reflectance conversion correction. The output layer is SSS. A more more restraining data cleansing process has been implemented in this paper, as described in previous sections. Moreover, the size and hyperparameters of the neural network presented here have been optimised (details given in Tables 3 and 4), providing better results to those in [22].

Two different networks have been trained for L1C and L2A. While both networks have the same architecture in terms of hidden layers and neurons per layer, the intrinsic training parameters have been optimised in each case to obtain the best possible results, and compare L1C and L2A at their best. For each case, two types of problems have been studied: interpolation and extrapolation. In the interpolation problem, data is randomised, and the test split is selected as an aleatory portion of the general dataset. In this case, training and test datasets follow the common 90/10 split, providing 90% of the data for training, and 10% as test. In the extrapolation scenario, a set number of in situ platforms are selected as test to check if the network is able to infer values without any prior information about the behaviour of the given platforms. The training/test split in the extrapolation case contains less information in the test, accounting for merely ~2% of the data. This makes the extrapolation problem much more complex than the interpolation one, as the amount of information to fit during test is
smaller, and thus the metrics used to understand how well the network performs are less likely to achieve good results.

\[ R^2 = \left( \frac{E[(\hat{y} - \mu_y)(y - \mu_y)]}{\sigma_y \sigma_y} \right)^2 \]  
(3)

- Mean Absolute Error (MAE):

\[ MAE = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j| \]  
(4)

- Most common error (\( \mu_e \)), defined as the expectation (or mean) of the error distribution:

\[ \mu_e = E[f(e)] \]  
(5)

where \( E \) is the expectation, \( y \) is the ground truth variable, \( \hat{y} \) is the estimated variable, \( \sigma_y \) is the variance of the ground truth variable, \( \sigma_{\hat{y}} \) is the variance of the estimated variable, \( \mu \) is the mean, \( \sigma \) is the standard deviation, \( n \) is the number of observations, \( y_j \) is the ground truth, \( \hat{y}_j \) is the predicted variable and \( f(e) \) is the error distribution function (or histogram, as data is discrete).

\( R^2 \) can take values from 0 to 1. The closer \( R^2 \) is to 1, the stronger the correlation between predicted and ground truth variables is and the better the model. The MAE is a valid metric, but can be easily moved towards higher values if outliers are present, and thus the reason to introduce the most common error definition. Errors are presented in an histogram, and the value corresponding to the average of the distribution is selected as the most common one.

3. Results

Four different cases have been studied to test the capabilities of the neural network and how different parameters might improve its performance. These four cases have been trained for both interpolation and extrapolation problems, providing a total of eight scenarios to test the neural network performance. The scenarios are as follows.

- Scenario 1: Baseline scenario.
- Scenario 2: Temperature included as input.
- Scenario 3: Latitude included as input.
- Scenario 4: Longitude included as input.
Scenario 1 presents a basic network, including only satellite bands and metadata as inputs. Scenario 2 includes the input variables from Scenario 1, plus sea surface temperature. Scenario 3 includes the input variables from Scenario 1, plus the platform latitude. Scenario 4 includes the input variables from Scenario 1, plus the platform longitude. Finally, Scenario 5 includes the input variables from Scenario 1, plus both latitude and longitude. In every case, the only output of the network is SSS.

Tables 3 and 4 include a summary of the most relevant variable combinations and performance metrics. Please note that only the optimal combination for each scenario has been included in the tables. The results are depicted graphically and discussed in the following subsections. In every case, the neural network architecture that provides the best results is composed of 20 hidden layers. The activation function is \( \text{tanh} \) for every layer except for the output layer, where a \( \text{sigmoid} \) is used. Different losses, optimisers and learning rates (the tuning parameter that determines the step size at each iteration while moving toward a minimum of the loss function) were tested to obtain the best behaviour for L1C and L2A cases. In every case, there was no dropout selected (as this produced worse results), and the optimal batch size was found to be 100. The batch size is a hyperparameter that controls the number of training samples to work through before the model internal parameters are updated. The optimal batch size was found by iterating with different sizes and choosing the one that made the network stay away from overfitting while still learning. Results represent a considerable improvement from those presented in [22] thanks to parameter optimisation.

**Table 3.** Neural network results for different Learning Rate (LR) combinations for the L1C dataset.

| Scenario | LR    | \( R^2_{\text{test}} \) | \( R^2_{\text{train}} \) | MAE Test | MAE Train |
|----------|-------|--------------------------|--------------------------|----------|-----------|
| Interpolation |       |                          |                          |          |           |
| 1        | 0.02  | 0.8549                   | 0.9445                   | 1.872    | 1.123     |
| 2        | 0.013 | 0.8463                   | 0.9778                   | 1.8119   | 0.9153    |
| 3        | 0.015 | 0.8446                   | 0.9823                   | 1.4231   | 0.7171    |
| 4        | 0.02  | 0.9928                   | 0.9952                   | 0.4627   | 0.3599    |
| Extrapolation |       |                          |                          |          |           |
| 1        | 0.02  | 0.7228                   | 0.8195                   | 2.85     | 2.48      |
| 2        | 0.02  | 0.7746                   | 0.9556                   | 1.9259   | 1.3182    |
| 3        | 0.02  | 0.6664                   | 3.9719                   | 2.585    | 1.0567    |
| 4        | 0.02  | 0.9717                   | 0.9918                   | 0.8927   | 0.6672    |

**Table 4.** Neural network results for different Learning Rate (LR) combinations for the L2A dataset.

| Scenario | LR    | \( R^2_{\text{test}} \) | \( R^2_{\text{train}} \) | MAE Test | MAE Train |
|----------|-------|--------------------------|--------------------------|----------|-----------|
| Interpolation |       |                          |                          |          |           |
| 1        | 0.013 | 0.811                    | 0.9435                   | 2.211    | 1.263     |
| 2        | 0.013 | 0.821                    | 0.8718                   | 2.2566   | 1.8302    |
| 3        | 0.01  | 0.804                    | 0.9514                   | 1.9456   | 1.1764    |
| 4        | 0.013 | 0.9731                   | 0.9924                   | 0.6367   | 0.492     |
| Extrapolation |       |                          |                          |          |           |
| 1        | 0.013 | 0.6506                   | 0.8948                   | 2.789    | 1.878     |
| 2        | 0.013 | 0.6999                   | 0.9018                   | 2.6      | 1.7265    |
| 3        | 0.01  | 0.6238                   | 0.9751                   | 2.9952   | 0.9423    |
| 4        | 0.013 | 0.9569                   | 0.9882                   | 1.024    | 0.6838    |

3.1. Interpolation

3.1.1. Experiment 1

In experiment 1, the baseline scenario, both L1C and L2A perform similarly in training, while the performance for test is slightly poorer in both cases, with \( R^2 \) dropping to 85% for L1C and 81% for L2A.
The MAE also increases considerably for L2A compared to L1C (from 1.87 PSU to 2.21 PSU). In terms of the most common error, $\mu_e$, the value duplicates for L2A (0.4 PSU) compared to L1C (0.2 PSU), see Figure 4. Note that values below 25 PSU are linked to in situ salinity around the Arctic and polar regions, as well as estuaries with strong freshwater inputs.

![Figure 4](image)

**Figure 4.** Interpolation, experiment 1: Sentinel-2 level 1C and level 2A neural network results. Training comparison between in situ and predicted data (left), test comparison between in situ and predicted data (centre) and distribution of errors in test data (right).

### 3.1.2. Experiment 2

Experiment 2 includes sea surface temperature as an input for the estimation of salinity. As in experiment 1, experiment 2 shows higher coefficient of determination for L1C, with L2A presenting higher MAE, above 2 PSU, see Figure 5. L2A also presents more scattered values in medium to low salinity ranges, where less information is available. This reinforces the fact that L1C is better than L2A when predicting values with little available information. In experiment 2, however, the most common error is similar in both cases, approximately 0.2 PSU, showing how L2A is more prone to errors in values far away from the standard salinity values in ocean waters. Generally, L1C provides values very close to the ground truth. Compared to experiment 1, results are slightly better in experiment 2, also showing a smaller overestimation of salinity in the mid to low range (below 25 PSU).
3.1.3. Experiment 3

Experiment 3 includes latitude as input. Results are slightly worse, in terms of the values of the coefficient of determination, than those for experiments 1 and 2. However, results are better in terms of MAE by almost 0.5 PSU. Most common errors are in the same range as previous experiments, which may imply that the marginal improvement is due to the randomisation of the test data rather than to an actual increase in the network reliability. Same as in previous cases, predicted salinity values are higher than their ground truth in medium values for L2A, while for L1C that problem is not present, see Figure 6.

Figure 5. Interpolation, experiment 2: Sentinel-2 level 1C and level 2A neural network results. Training comparison between in situ and predicted data (left), test comparison between in situ and predicted data (centre) and distribution of errors in test data (right).

Figure 6. Interpolation, experiment 3: Sentinel-2 level 1C and level 2A neural network results. Training comparison between in situ and predicted data (left), test comparison between in situ and predicted data (centre) and distribution of errors in test data (right).
3.1.4. Experiment 4

Experiment 4 introduces longitude as input to the neural network. The inclusion of longitude produces the best values of all the possible combinations tried, with test coefficients of determination near 1, and MAEs of 0.4 PSU for L1C. The values are clearly closer to the ground truth than previous experiments. L2A still performs worse than L1C, with higher MAE (0.6 PSU), and some rogue values still showing in the intermediate salinity range, see Figure 7.

The reason for the inclusion of longitude giving such good results, and not latitude, is of special interest, as it gives a clue to what is the key parameter affecting the performance of the neural network linked to atmospheric corrections. The improvement might be related to the slight misalignment between bands and the sun position, due to the relative staggered alignment of the bands. As latitude is not as relevant, it can be concluded that relative temperature is not a driving factor for the network performance.

Longitude effects are related to the satellite orbit: the sun-synchronous orbit is achieved by having the oscillating orbital plane precess (or rotation) of approximately 1° eastwards each day with respect to the celestial sphere to keep pace with the Earth’s movement around the Sun [29]. The precession is achieved by tuning the inclination to the altitude of the orbit such that Earth’s equatorial bulge, which perturbs the inclined orbit, causes the orbital plane of the satellite to rotate with the desired rate. It seems reasonable then to assume that longitude adds some degree of correction to the relative position of Sentinel-2 with the Sun in different images: the zenith and azimuth angles are provided as inputs to the neural network, but longitude corrects the potential differences between images, as they are taken in different parts of the world and times of the year.

![Figure 7](image)

Figure 7. Interpolation, experiment 4: Sentinel-2 level 1C and level 2A neural network results. Training comparison between in situ and predicted data (left), test comparison between in situ and predicted data (centre) and distribution of errors in test data (right).

Error maps are provided for experiment 4, as it is the best performing test, to see how it behaves in different platforms around the world. Figure 8 shows a comparison between errors for L1C (left) and L2A (right). Errors for L1C are in the range of 0.2 PSU, with some outliers present in the Mediterranean.
region. In the L2A cases, errors are higher (lighter green) and many more outliers are present in all areas studied.

![Map of differences between measured and predicted salinity in test locations for the United States, Europe, and Japan for S2-L1C and S2-L2A.](image)

**Figure 8.** Difference between measured and predicted salinity in test locations for the United States (top), Europe (centre) and Japan (bottom) for S2-L1C (left) and S2-L2A (right).
3.2. Extrapolation

As presented in Tables 3 and 4, all four experiments have been performed for the extrapolation problem, selecting six random platforms as test dataset (approximately 200 data points). All the data from those platforms is isolated, and thus the network does not have any previous information about the platforms’ behaviour. Extrapolation experiments show the true generalisation capabilities of a machine learning algorithm. Figures 9 and 10 only show results for experiment 4, given that one was the experiment that provided the best results.

Figure 9. Extrapolation, experiment 4: Sentinel-2 level 1C and level 2A neural network results. Training comparison between in situ and predicted data (left), test comparison between in situ and predicted data (centre), and distribution of errors in test data (right).

Figure 10. Difference between measured and predicted salinity in test locations for Europe for S2-L1C (left) and S2-L2A (right).
The extrapolation experiment including longitude performs very well, with MAEs of approximately 0.9 PSU and most common errors slightly below 0.2 PSU for L1C. Results for L2A are worse as in previous cases, and particularly noticeable in the most common error, showing that the generalisation capability of L2A is more limited than that L1C, see Figure 9.

The random test buoys are located in the Mediterranean, Atlantic, English Channel and Baltic Sea, see Figure 10. As in previous cases, the L1C results are better, except for a point in the Mediterranean, where L2A behaves slightly better—mean estimation error in the Mediterranean: 0.7 PSU for L1C compared to 0.6 PSU for L2A; Atlantic: 0.1 PSU for L1C compared to 0.5 PSU for L2A; English Channel (same results for Baltic Sea): 0.3 PSU for L1C compared to 0.8 PSU for L2A.

4. Discussion: Evaluation and Comparison of Outputs from L1C and L2A in Complete Tiles

The best results from the neural network training have been used to evaluate the model in a mosaic of tiles in different areas. The selected areas are Kuwait Bay, the Amazon river mouth and Canterbury Bight. These three locations provide different latitudes and longitudes, as well as climates, times of the year and singularities that make them ideal to demonstrate the differences between L1C and L2A results. Kuwait Bay offers a secluded environment at the northern end of the Persian gulf, with inputs from small rivers and a very complex geography. The Amazon mouth shows a coastal area with input from a very large river with very high sediment concentration. Canterbury Bight, at the south of the city of Christchurch (New Zealand), faces the South Pacific, with a very constant shore exposed to long-shore transport. The best performing neural network models for L1C and L2A have been applied to each pixel in the tiles, creating the salinity maps depicted in the figures below.

Please keep in mind that the aim of this discussion and the following figures is not to check if the values of salinity are correct or not (for that purpose we have Figures 4–10), but to qualitatively compare the results using L1C and L2A tiles. We cannot know if the derivation of salinity in a complete Sentinel-2 tile is accurate, as we do not have any other high-resolution models to compare with, but we will clearly see that L2A is not providing the details that L1C has. Therefore, even if the actual values might not be valuable (as they cannot be validated further from the neural network test sites), they are very useful in terms of observing the phenomenological effects around coastal areas. The goal of the following figures is to test the hypothesis of this paper: that atmospherically corrected data is not good enough for water quality analyses in general, and salinity estimation in particular.

4.1. Kuwait Bay, Persian Gulf

Figure 11 shows the true colour composite from Sentinel-2 L1C at Kuwait Bay. Land is depicted in black. Figure 12 shows the SSS product derived from L1C (top) and L2A (bottom). The L1C product shows riverine inputs to the bay at lower salinity content, as well as a salinity front where the transition between estuarine and coastal waters happen. Values range from 20 PSU to 38 PSU. In contrast, in the L2A product, values are much higher, being the same as in the L1C case closer to the bottom of the figure (further away from the coast). This and similar results in the next figures suggest that land adjacency correction in L2A could be one source of error, given that results for both L1C and L2A get closer farther away from the coast.
Figure 11. Original S2-L1C true colour composite of Kuwait Bay (9th December 2019). Land depicted in black.

Figure 12. Sea surface salinity at Kuwait Bay, 100 m resolution. Results using S2-L1C (top) and S2-L2A (bottom).
4.2. Mouth of the Amazon River, West Atlantic

Figure 13 shows a true colour composite from L1C at the Amazon River mouth. The brown waters have a very high sediment load. The BRDF effect is clearly visible in this image, represented as subtle stripes in the water. Figure 14 shows the L1C and L2A products for the Amazon. As in the previous case, L1C shows the differences between riverine waters, and their interaction with sea water. The BRDF effect is still present in the L1C product, although the salinity is almost the same in the different bands, and the transition is very subtle. This fact is relevant, as it shows that the L1C network is learning about the sun glint effect, and is correcting for it. This effect is not present in the L2A product thanks to the atmospheric corrections, but the L2A information has averaged the behaviour of the coastal waters, and the SSS values are only similar to the L1C product far away from the river mouth.

Figure 13. Original S2-L1C true colour composite of the mouth of river Amazon (15th November 2019). Land depicted in black.

Figure 14. Cont.
4.3. Canterbury Bight, South Pacific

Figure 15 shows a true colour composite at Canterbury Bight (New Zealand). Figure 16 presents the L1C and L2A SSS products. This is possibly the most interesting result of all: the L1C product clearly shows a Kármán vortex street moving from the bottom-left of the image, parallel to the coast. This can be caused by temperature or density anomalies, as well as by the presence of obstacles in the flow. This phenomenon is not present at all in the L2A product, which shows results similar to previous cases: high salinity values that fit well open ocean salinity values, but are not accurate enough for coastal areas. This may be caused by the atmospheric corrections, which are creating an output BOA product that is losing information compared to that provided by the L1C product. Same as in previous cases, subtle sun glint lines are observed in the L1C product, but the SSS values around them seem to transition accordingly.
5. Conclusions

High-resolution sea surface salinity in coastal areas obtained from Top- and Bottom-of-Atmosphere multispectral Sentinel-2 data is presented in this paper. The aim is to show the effect of atmospheric corrections when using data-driven innovation techniques and machine learning approaches. A neural network is trained with Level-1C and Level-2A (atmospherically corrected) imagery and in situ information from different platforms around the world, in order to build the input–output pipeline for the neural network. The network is fed with band information and metadata as inputs, and the output is sea surface salinity. The resolution of the output product is 100 m. Four scenarios testing the addition of different input variable to the neural network are presented. Both L1C and L2A show good agreements between predicted and in situ data, with L1C always outperforming L2A. The best scenario is the one where longitude is included as an extra input to the network. Results show coefficients of determination close to 1, mean errors below 0.4 PSU and most common errors below 0.2 PSU for L1C. This is thought to be caused because of improvements on the slight misalignment between bands and sun position, due to the relative staggered alignment of the instrumentation for each band physically on the satellite. Longitude seems to add some degree of correction to the relative position of Sentinel-2 with the Sun in different images: the zenith and azimuth angles are provided as inputs to the neural network, but longitude corrects the potential differences between images, as they are taken in different parts of the world and times of the year, making the dataset more uniform.

The results of the network are tested in three mosaics from different coastal waters in Kuwait, Brazil and New Zealand. The aim of this is to compare qualitatively the differences between L1C- and L2A-derived salinity. The L1C SSS product shows a higher degree of detail, clearly depicting river outputs and estuarine circulation. However, the L2A product shows similar SSS values only closer to open ocean, and coastal values are overestimated. Representative patterns clearly visible
in L1C are not present in the L2A product. The atmospheric correction seems to be “averaging” the reflectance values of the L2A product, which leads to the loss of all the details that the L1C still has, as the information has not be adulterated by any external processes. Moreover, results from L1C and L2A become closer farther away from the coast, which suggests that land adjacency corrections in the L2A correction process could be one source of error. On the other hand, the BRDF correction could be another source of error, as the values for water pixels are calculated using its closest land values. Despite the BRDF effect present in L1C tiles, the salinity derived using the neural network algorithm presents a soft transition, making the BRDF negligible. This last point is relevant because it supports the fact that the network is able to “learn” when BRDF is present, and correcting it automatically by the information learnt from unaltered pixels.

In summary, results suggest that atmospheric corrections add a degree of uncertainty to the final products, and lead to the loss of information key for the development of SSS from multispectral imagery. This is an important fact to take into account in the development of any other products from multispectral data (such as Chlorophyll and Coloured Dissolved Organic Matter), as the same issues with atmospheric corrections could be observed.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. European Space Agency. Sentinel—2 Data Products. 2020. Available online: https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products (accessed on 10 June 2020).
2. ESA. Sentinel—2 User Handbook; ESA Standard Documents; ESA: Paris, France, 2015. Available online: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook (accessed on 10 June 2020).
3. Liang, S.; Li, X.; Wang, J. (Eds.) Chapter 5—Atmospheric Correction of Optical Imagery. In Advanced Remote Sensing; Academic Press: Boston, FL, USA, 2012; pp. 111–126; Available online: http://www.sciencedirect.com/science/article/pii/B9780123859549000058 (accessed on 10 June 2020).
4. MODTRAN 5.2.0.0 User’s Manual. In Advanced Remote Sensing; Spectral Sciences, Inc.: Burlington, MA, USA, 2008.
5. Applications, R. ATCOR—Code Comparison. 2020. Available online: https://www.rese.ch/software/atcor/compare.html (accessed on 10 June 2020).
6. European Space Agency. Sentinel—2 Level—2A Processing Overview. 2020. Available online: https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a-processing (accessed on 10 June 2020).
7. European Space Agency. Sen2Cor, 2020, Available online: http://step.esa.int/main/third-party-plugins-2/sen2cor/ (accessed on 10 June 2020).
8. Pflug, B.; Main-Knorn, M.; Bieniarz, J.; Debaecker, V.; Louis, J. Early Validation of Sentinel-2 L2A Processor and Products. In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016.
9. Satellite Imaging Corporation. ATCOR. 2020. Available online: https://www.satimagingcorp.com/services/atcor/ (accessed on 10 June 2020).
10. Kremezi, M.; Karathanassi, V. Correcting the BRDF Effects on Sentinel-2 Ocean Images. In Proceedings of the Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019), 11741C, Paphos, Cyprus, 18–21 March 2019. [CrossRef]
11. European Space Agency. Level 2A Input Output Data Definition. 2020. Available online: https://step.esa.int/thirdparties/sen2cor/2.5.5/docs/S2-PDGS-MPC-L2A-IODD-V2.5.5.pdf (accessed on 10 June 2020).
12. Sterckx, S.; Knaeps, E.; Ruddick, K. Detection and correction of adjacency effects in hyperspectral airborne data of coastal and inland waters: The use of the near infrared similarity spectrum. Int. J. Remote Sens. 2011, 32, 6479–6505. [CrossRef]
13. Richter, R.; Louis, J.; Berthelot, B. Sentinel-2 MSI—Level 2A Products Algorithm Theoretical Basis Document. 2011. Available online: https://earth.esa.int/c/document_library/get_file?folderId=349490&name=DLFE-4518.pdf (accessed on 10 June 2020).
14. Barre, H.M.J.P.; Duesmann, B.; Kerr, Y.H. SMOS: The Mission and the System. *IEEE Trans. Geosci. Remote Sens.* 2008, 46, 587–593. [CrossRef]

15. Le Vine, D.M.; Lagerloef, G.S.E.; Ral Colomb, F.; Yueh, S.H.; Pellerano, F.A. Aquarius: An Instrument to Monitor Sea Surface Salinity From Space. *IEEE Trans. Geosci. Remote Sens.* 2007, 45, 587–593. [CrossRef]

16. Siddorn, J.; Bowers, D.; Hoguane, A. Detecting the Zambezi River Plume using Observed Optical Properties. *Mar. Pollut. Bull.* 2001, 42, 942–950. [CrossRef]

17. Aarup, T.; Holt, N.; Højerslev, N. Optical measurements in the North Sea-Baltic Sea transition zone. II. Water mass classification along the Jutland west coast from salinity and spectral irradiance measurements. *Cont. Shelf Res.* 1996, 16, 1343–1353. [CrossRef]

18. Warnock, R.E.; Gieskes, W.W.; van Laar, S. Regional and seasonal differences in light absorption by yellow substance in the Southern Bight of the North Sea. *J. Sea Res.* 1999, 42, 169–178. [CrossRef]

19. McKee, D.; Cunningham, A.; Jones, K. Simultaneous Measurements of Fluorescence and Beam Attenuation: Instrument Characterization and Interpretation of Signals from Stratified Coastal Waters. *Estuar. Coast. Shelf Sci.* 1999, 48, 51–58. [CrossRef]

20. Bowers, D.; Harker, G.; Smith, P.; Tett, P. Optical Properties of a Region of Freshwater Influence (The Clyde Sea). *Estuar. Coast. Shelf Sci.* 2000, 50, 717–726. [CrossRef]

21. Sullivan, S.A. Experimental Study of the Absorption in Distilled Water, Artificial Sea Water, and Heavy Water in the Visible Region of the Spectrum. *J. Opt. Soc. Am.* 1963, 53, 962–968. [CrossRef]

22. Medina-Lopez, E.; Urena-Fuentes, L. High-Resolution Sea Surface Temperature and Salinity in Coastal Areas Worldwide from Raw Satellite Data. *Remote Sens.* 2019, 11, 2191. [CrossRef]

23. Chen, S.; Hu, C. Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean colour measurements. *Remote Sens. Environ.* 2017, 201, 115–132. [CrossRef]

24. Copernicus. Copernicus, Europe’s Eyes of Earth. 2020. Available online: https://www.copernicus.eu/en (accessed on 10 June 2020).

25. Copernicus Marine Environment Monitoring Service. 2020. Available online: https://marine.copernicus.eu/ (accessed on 10 June 2020).

26. Google. Google Earth Engine. 2020. Available online: https://code.earthengine.google.com (accessed on 10 June 2020).

27. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 2017, 202, 18–27. [CrossRef]

28. Srivastava, R.; Greiff, K.; Schmidhuber, J. Highway networks. arXiv 2015, arXiv:1505.00387.

29. Toomer, G.J. The Solar Theory of az-Zarqal A History of Errors. *Centaurus* 1969, 14, 306–336. [CrossRef]