Remote Sensing of Environment 281 (2022) 113220

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements

Emy Alerskans[a,*,1], Ann-Sofie P. Zinck[b,1], Pia Nielsen-Englyst[c,a], Jacob L. Høyer[a]

[a] Danish Meteorological Institute, Copenhagen, Denmark
[b] IMAU, Utrecht University, Utrecht, The Netherlands
[c] DTU-Space, Technical University of Denmark, Lyngby, Denmark

ARTICLE INFO

Edited by Menghua Wang

Keywords:
Remote sensing
Passive microwave
Sea surface temperature
Machine learning
AMSR-E

ABSTRACT

Two machine learning (ML) models are investigated for retrieving sea surface temperature (SST) from passive microwave (PMW) satellite observations from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) and auxiliary data, such as ERA5 reanalysis data. The first model is the Extreme Gradient Boosting (XBG) model and the second is a multilayer perceptron neural network (NN). The performance of the two ML algorithms is compared to that of an existing state-of-the-art regression (RE) retrieval algorithm.

The performance of the three algorithms is assessed using independent in situ SSTs from drifting buoys. Overall, the three models have similar biases; 0.01, 0.01 and −0.02 K for the XGB, NN and RE, respectively. The XGB model performs best with respect to standard deviation; 0.36 K. While the NN model performs slightly better than the RE model with respect to standard deviation, 0.50 and 0.55 K, respectively, the RE model is found to be more sensitive to changes in the in situ SST. Moreover, the XGB model is the least sensitive with an overall sensitivity of 0.78, compared to 0.90 for the RE model and 0.88 for the NN model.

The good performance of the two ML algorithms compared to the state-of-the-art RE algorithm in this initial study demonstrates that there is a large potential in the use of ML algorithms for the retrieval of SST from PMW satellite observations.

1. Introduction

Sea surface temperature (SST) is an essential climate variable (Bojinski et al., 2014) used in various applications such as climate monitoring (e.g. Merchant et al., 2019), numerical weather prediction (NWP; Chelton and Wentz, 2005; Brasnett and Colan, 2016), ocean and coupled models (Le Traon et al., 2015; Yang et al., 2015; Liang et al., 2017) and in the understanding of air–sea interactions (Monzikova et al., 2017; Ning et al., 2018). SST has been measured in situ for more than 150 years, initially from ships and oceanographic profiles and later from moored and drifting buoys (Rayner et al., 2006). SST retrieved from Earth-orbiting satellites is a crucial supplement to the in situ network due to the more complete temporal and spatial coverage from satellites (Minnett et al., 2019). Thermal infrared (IR) satellite observations have been available since 1981, but these observations are biased from aerosols and limited by their inability to observe the surface through clouds (Merchant et al., 1999, 2006). Observations from passive microwave (PMW) sensors are widely recognised as an important supplement to IR observations since PMW observations of the surface are not prevented by non-precipitating clouds and the impact from aerosols is small (Wentz and Meissner, 2000; Chelton and Wentz, 2005). They are, however, impacted by precipitation (Gentemann et al., 2010) and sun glint contamination, which increases the swath gaps (Gentemann and Hilburn, 2015). The first global accurate PMW SST data using the 6 GHz channels became available in 2002 from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E; Kawanishi et al., 2003; Chelton and Wentz, 2005), carried onboard the National Aeronautics and Space Administration's (NASA's) Earth Observation System Aqua platform. AMSR-E ceased normal operations in October 2011 and was followed by the currently operational AMSR2 on the Global Change Observing Mission (GCOM-W1; Maeda et al., 2015), launched in May 2012. An AMSR2 follow-on mission (AMSR3) is planned by Japan Aerospace Exploration Agency (JAXA) (Maeda et al., 2020) and the Copernicus Imaging Microwave Radiometer (CIMR) is prepared by the European

[*] Corresponding author.
E-mail address: ea@dmi.dk (E. Alerskans).
[1] Previous address: Physics of Ice, Climate, and Earth, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark.

https://doi.org/10.1016/j.rse.2022.113220
Received 23 December 2021; Received in revised form 7 August 2022; Accepted 9 August 2022
Available online 5 September 2022

Space Agency (ESA) as a part of the Copernicus Expansion Program of the European Union (http://www.cimr.eu/; Donlon, 2020).

Different PMW SST retrievals have been developed and refined over the years using different frequency channels and approaches. Two types of retrieval algorithms have generally been used to retrieve SST from PMW observations; statistical algorithms and physically based algorithms. The most common approach to generate PMW SST products is by using a statistical retrieval algorithm (e.g. Shibata, 2006; Wentz and Meissner, 2007; Gentemann et al., 2009; Chang et al., 2015; Alerskans et al., 2020). Statistical retrieval algorithms are developed by comparisons of satellite measured brightness temperatures and collocated and temporally matched in situ observations and model data, such as atmospheric and oceanic reanalysis data of wind speed, atmospheric water vapour content and SST. The second type of retrieval algorithm uses a radiative transfer model (RTM) to simulate the top of atmosphere brightness temperatures. This approach requires instrument information (azimuth/earth incidence angles, frequency and polarisation) and environmental information (SST, sea surface salinity, wind speed/direction, water vapour density, liquid water density, pressure, and atmospheric profiles of temperature). Optimal estimation (OE) theory is an example of an approach that makes use of an RTM (forward model). In OE models, the RTM is inverted in order to retrieve SST from satellite measured brightness temperatures (Nielsen-Englyst et al., 2018). The inversion is performed using a priori information about the ocean and atmosphere (and corresponding uncertainties) to constrain the retrievals.

The OE retrieval allows for indication of measurement errors, such as imperfect calibration and channel contamination (Minnett et al., 2019). This also means that the performance of OE algorithms is constrained by the accuracy of the RTM as well as the representativeness of the observation and prior error covariances (Merchant et al., 2020). Moreover, measurement errors require ad-hoc corrections to the geophysical retrievals in the OE type of algorithms (Meissner and Wentz, 2012; Nielsen-Englyst et al., 2018). In contrast, statistically based algorithms may account for some of the measurement errors through the coefficient derivation process, but they are limited by the established statistical relationships between the variables. Hence, both physical and statistical models make a series of considerable assumptions about the nature of the radiative transfer process, which are provided directly by the RTM in physical models, whereas statistical models rely on established assumptions of how the geophysical quantities can be used as proxies for the actual physical processes that influence the surface emissivity and the radiative transfer through the atmosphere.

Machine learning (ML) models may improve or supplement existing retrieval algorithms through their higher flexibility and capability of recognising meaningful patterns and structures in complex problems (Lee et al., 2017; Azodi et al., 2020). Compared to both physical models and statistical models, there are much fewer assumptions about the functional form of how the geophysical quantities are related to the predicted quantity in ML models. This may allow development of complex functional forms that more closely approximate the actual physical processes and thereby provide a more accurate SST retrieval. ML models may also be a good alternative in situations where observation characteristics and the structure of the uncertainty components are not well known, e.g. during a commissioning phase of a new instrument such as CIMR. Therefore, there is a need for insight into the performance of different ML models for retrieving SST. Until recently, the use of ML techniques has been very limited within the field of SST retrievals, but investigations using ML to improve the accuracy of SST algorithms is listed as one of the priority recommendations provided by the SST community (O'Carroll et al., 2019). There has been an increasing amount of research applying ML techniques to specific parts of retrieval algorithms, such as for cloud detection (Paul and Huntemann, 2021), bias correction (Saux Picart et al., 2018), error estimation (Kumar et al., 2021), identification of eddies (Moschos et al., 2020) and ocean extremes (Prochaska et al., 2021). A recent study also used ML techniques to retrieve daily cloud-free IR SSTs from the MODIS Aqua sensor (Sunder et al., 2020). In addition, ML techniques have also been used for retrieval of other satellite-derived geophysical variables, such as soil moisture (Rodriguez-Fernandez et al., 2015) and precipitation (Sanò et al., 2016, 2018).

In this paper two types of ML SST retrieval techniques have been assessed and compared against an existing state-of-the-art statistical regression model retrieval algorithm. The first is the decision tree-based algorithm Extreme Gradient Boosting (XGBoost, here XGB; Chen and Guestrin, 2016), and the second is a multilayer perceptron (MLP) neural network (NN; Haykin, 1999; Nielsen, 2015). These methods differ in architecture and represent two of the main ML categories; decision-trees and neural networks. The XGB is a relatively new algorithm which has shown good performance for retrieval and bias correction of other geophysical variables (e.g. Just et al., 2018, 2020; Liu et al., 2021). The MLP is a fully-connected feed-forward NN and is one of the simplest and most used neural network architectures.

The paper is structured with a description of the dataset, as well as pre-processing and dataset splitting in Section 2. This is followed by a presentation of the three retrieval algorithms and model optimisation of the two ML algorithms in Section 3. The results are presented in Section 4 and discussed in Section 5 before the final concluding remarks are provided in Section 6.

2. Data

2.1. ESA CCI Multisensor Matchup Dataset (MMD)

The ESA climate change initiative (CCI) Multisensor Matchup Dataset (MMD), described in Nielsen-Englyst et al. (2018) and Alerskans et al. (2020), is the basis for this work. The MMD consists of quality controlled in situ measured SST observations from the International Comprehensive Ocean-Atmosphere DataSet (ICOADS) version 2.5.1 (Woodruff et al., 2011) and the Met Office Hadley Centre (MOHC) Ensembles dataset version 4.2.0 (EN4; Good et al., 2013). Brightness temperatures from the AMSR-E Level 2 A (L2 A) swath data product, AMSR-E V12 (Ashcroft and Wentz, 2013), spatially re-sampled to the 6.9 GHz resolution (75 × 43 km), are also included. The in situ and satellite observations are matched by imposing a maximal geodesic distance of 20 km and a maximal time difference of 4 h. The MMD includes matchups from the period June 2002–October 2011.

Additional data included in the MMD are information from both the ERA-Interim reanalysis (Dee et al., 2011) and the ERA5 reanalysis (Hersbach et al., 2020) on SST, total column water vapour (TCWV), total cloud liquid water (TCLW), wind speed (WS) and sea ice concentration (SIC). Sea surface salinity (SSS) from the GLOBAL-REANALYSIS-PHY-001-030 reanalysis product, provided by the Copernicus Marine Environment Monitoring Service (CMEMS; http://marine.copernicus.eu) is also included in the MMD. Additional wind data from the Cross-Calibrated Multi-Platform (CCMP) gridded surface wind vector product (Atlas et al., 2011) version 2.0 was included. The additional data were collocated in time and space with the MMD matchups using the nearest neighbour interpolation. For a list of the MMD variables extracted for this study and considered as input features to the two ML retrieval algorithms see Table 1.

2.2. Pre-processing

To ensure an accurate derivation of the retrieval algorithms erroneous in situ, satellite and auxiliary data are excluded. The quality of the brightness temperatures were assessed using the L1 AMSR-E instrument quality flags and low-quality data were excluded. Moreover, brightness temperatures outside the accepted range (0–320 K) were flagged. In addition, data were excluded if the difference between vertical (V) and horizontal (H) polarisations for the 18–36 GHz brightness temperatures were negative, as this indicates invalid oceanographic

Table 1
MMD variables considered as input features to the ML retrieval algorithms. The asterisk marks the features used in the XGB model (see Section 3.2).

Feature	Acronym	Source
AMSR-E orbit (asc./desc.)	orbit	AMSR-E
Latitude*	lat	AMSR-E
Longitude*	lon	AMSR-E
Solar zenith angle	solza	AMSR-E
Satellite zenith angle*	satza	AMSR-E
Satellite azimuth angle	sataz	AMSR-E
Sun glint angle	sga	AMSR-E
Brightness temperature, channel 6V*	tb6V	AMSR-E
Brightness temperature, channel 6H*	tb6H	AMSR-E
Brightness temperature, channel 10V*	tb10V	AMSR-E
Brightness temperature, channel 10H*	tb10H	AMSR-E
Brightness temperature, channel 18V	tb18V	AMSR-E
Brightness temperature, channel 18H	tb18H	AMSR-E
Brightness temperature, channel 23V	tb23V	AMSR-E
Brightness temperature, channel 23H	tb23H	AMSR-E
Brightness temperature, channel 36V*	tb36V	AMSR-E
Brightness temperature, channel 36H*	tb36H	AMSR-E
Brightness temperature, channel 89V*	tb89V	AMSR-E
Brightness temperature, channel 89H*	tb89H	AMSR-E
Wind speed*	WS	ERA5
Wind direction	ϕ_W	ERA5
Relative angle between sataz and ϕ_W^*	ϕ_{REL}	ERA5/AMSR-E
Total column water vapour*	TCWV	ERA5
Cloud liquid water*	CLWT	ERA5

retrievals. To exclude low-quality brightness temperatures possibly contaminated by e.g. rain and sea ice, an additional quality control check for the AMSR-E 23 and 36 GHz brightness temperatures were performed. The spatial standard deviation was calculated over a 21 × 21 pixel subregion around each matchup in order to exclude matchups with an anomalously high standard deviation. Data were flagged if the standard deviations of the 23 V and H and 36 V and H channels were larger than 55, 35, 25 and 25 K, respectively. These thresholds were chosen based on the distribution of the spatial standard deviations in order to exclude matchups in the end tails, as they have a very high and anomalous spatial standard deviation and therefore likely are contaminated. The chosen thresholds resulted in exclusion of less than 1% of the matchups. Low quality in situ data and matchups with an in situ or ERA5 SST outside the range −2–34 °C were also excluded, where the lower limit of −2 °C is used in order to exclude matchups potentially contaminated by sea ice. Furthermore, matchups with an ERA5 WS greater than 20 ms^{-1} were also flagged. The upper wind speed limit is based on the fact that extreme surface roughness and the existence of foam on the surface caused by high wind speeds impact the brightness temperatures and make the SST retrievals uncertain (Kilic et al., 2018). Together, these checks constitute the gross error checks in Table 2, which are performed to remove obviously erroneous satellite, in situ and auxiliary data. To exclude matchups that might have been contaminated due to atmospheric or surface effects, additional checks were performed. Matchups contaminated by sea ice or land were excluded using the AMSR-E land/ocean flag and the ERA5 sea ice fraction. To account for contamination due to rain, matchups were removed if the 18 V GHz brightness temperature was greater than 240 K. Sun glitter contamination was avoided by excluding matchups with a sun glint angle less than 25°. Diurnal warming effects were accounted for by excluding daytime matchups with ERA5 WS less than 4 ms^{-1}. Matchups potentially contaminated by ground-based and space-based RFI were excluded using observation location and reflection longitude and latitude according to Table 2 in Gentemann and Hilburn (2015). Lastly, obviously erroneous in situ SSTs were removed using a 3-sigma filter, based on the mean difference between ERA5 and in situ SSTs. To ensure a balanced and latitudinally representative dataset, such that the models are trained and validated on data not only from a few specific regions in which in situ observations are dense, the number of matchups per latitude degree was restricted. As the number of matchups increase with time, the restriction is temporally dependent with different number of matchups allowed for different years. For 2002, which is the year with fewest matchups, a maximum of 2,000 matchups per latitude degree were allowed, whereas for 2011, which is the year with the most matchups, a maximum limit of 20,000 matchups per latitude degree was used. The percentage of matchups removed and the total number of matchups left after each filtering check is shown in Table 2. Furthermore, the geographical distribution of satellite versus drifting buoy matchups after filtering for the validation dataset is shown in Fig. 1.

Table 2
Number of matchups remaining after each check and the percentage of matchups each check removes. The percentages removed for checks 1–8 plus the summary checks ("Checks 1–7", "Checks 1-8" and "All checks 1–9") are with respect to all matchups. The percentage of matchups removed by the SST 3σ-filter, on the other hand, is with respect to "Checks 1–7" and the even-out-by-latitude check is with respect to "Checks 1–8".

Filter	No. of matchups	Percentage of matchups removed (%)
(0) All matchups (no filter)	40,480,306	–
(1) Gross error checks	31,070,944	23.24
(2) Rain	401,42,612	0.83
(3) Sun glint	38,145,778	5.77
(4) RFI	37,387,456	7.64
(5) Land	34,839,510	13.93
(6) Sea ice	31,390,993	22.45
(7) Diurnal warming	37,311,599	7.83
Checks 1–7	19,397,886	52.08
(8) SST 3σ-filter	18,999,399	2.05
Checks 1–8	18,999,399	53.07
(9) Even-out-by-latitude	15,316,989	19.38
All checks 1–9	15,316,989	61.16

The MMD is divided into six subsets in order to perform all steps on independent data. A random splitting of the data is performed such that all datasets retain the same distribution for each variable. The number of matchups in each subset, as well as the percentage of data with respect to the filtered dataset, is indicated in parentheses.

1. Training dataset: used for training the NN and XGB models (6,126,795/40.0%).
2. Test dataset: used for evaluating the performance of the ML models during training (1,021,133/6.7%).
3. Hyperparameter optimisation dataset: used for optimising model hyperparameters (see Section 3.4; 3,063,398/20.0%).
4. Feature selection dataset: used for selecting input variables for the NN and XGB models (see Section 3.1; 1,021,133/16.7% of the training dataset).
5. Validation dataset: used for validating the performance of the NN, XGB and RE models (5,105,663/33.3%).
6. Sensitivity dataset: used for the estimating the SST sensitivity of the NN, XGB and RE models (1,021,133/20.0% of the validation dataset).

Another important part of the pre-processing step in order to ensure a good performance for the ML retrieval algorithms is data normalisation (Kotsiantis et al., 2006; Huang et al., 2020). Normalisation of the data is a transformation of the data in order to transform the data to the same scale. Different methods can be used for normalisation of the data. Here, we have used quantile transform normalisation. This method uses quantile information in order to transform the data to follow a uniform distribution and as it reduces the impact from outliers it is a robust normalisation method. It has previously been used for feature normalisation within satellite-based applications and for classification (Ferreira et al., 2019; Sismanidis et al., 2021). Other popular methods include min–max normalisation and standardisation. It should be mentioned that no universal normalisation method exists

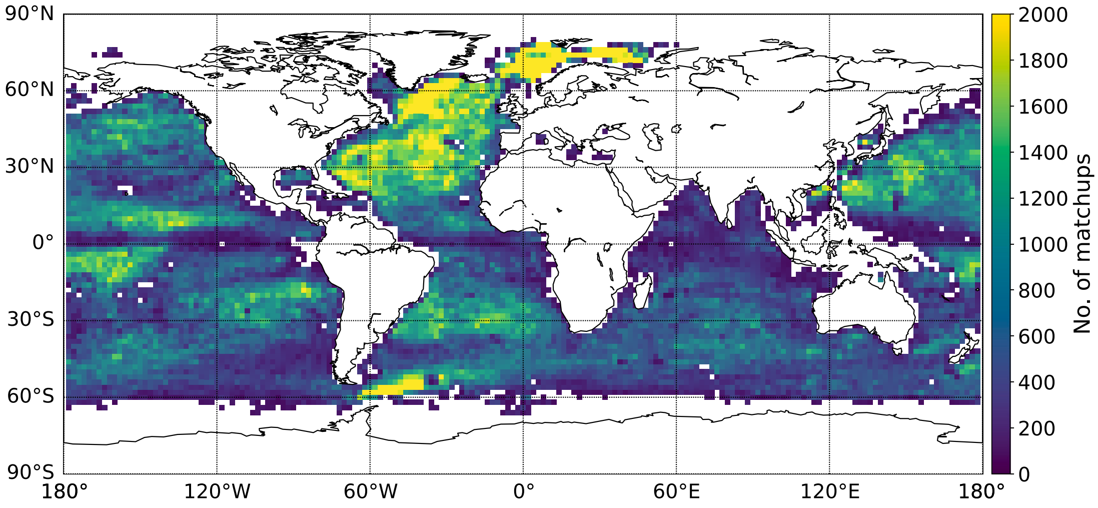

Fig. 1. Spatial distribution of satellite matchups with drifting buoys for the validation dataset. The statistics have been calculated on a 2 × 2 degree grid with a minimum of 50 matchups per grid cell.

and that the performance of models might vary depending on the normalisation method and the problem.

3. Retrievals

In this section the three models used for retrieving SST will be introduced. First the selection of input variables (also called input features) is presented, followed by a description of the two ML models, XGB and NN, and the optimisation thereof. Lastly, the state-of-the-art regression retrieval model used as a benchmark is presented.

3.1. Feature selection

Table 1 shows the 24 features that were extracted from the MMD and considered as input to the two ML models. To exclude redundant features and only select the important ones, such that the dimensionality of the input data is reduced and the risk of overfitting likewise is reduced (Goodfellow et al., 2016), a feature importance analysis was performed in order to obtain the explanatory power of each input feature. The analysis was based on the SHapley Additive exPlanations (SHAP) values (Lundberg and Lee, 2017). SHAP uses shapley values (Shapley Ll, 1953), which are based on cooperative game theory and are used in many state-of-the-art feature attribution methods (Ribeiro et al., 2016; Shrikumar et al., 2016, 2017). SHAP is based on the idea that the performance of all possible combinations of input features should be considered when determining the importance of a single feature on a single prediction. To determine the importance of each feature, the ML model to be used (here XGB and NN) is trained for each combination of input features and the marginal contribution of each feature is evaluated. The marginal contribution is defined as the difference between the performance of the model which includes the feature to be assessed and the model in which the feature is excluded. The marginal contribution of a feature is therefore obtained by considering the difference between all models in which this feature is present and all models in which it is excluded. From this, the average contribution of a single feature can be obtained. Based on this, an importance is assigned to each feature for each prediction and from this the average explanatory power of each feature can be estimated. For a further explanation of SHAP see Lundberg and Lee (2017).

The SHAP analysis is performed on the XGB and NN base models, i.e. the models with default settings. For XGB the default settings are given by its python implementations using scikit-Learn (Pedregosa et al., 2011), whereas for the NN model the corresponding default parameters were used, with the exception of number of hidden layers and neurons, which were chosen as 2 layers and 20 and 15 neurons, respectively (see Section 3.3). Therefore, the SHAP feature importance analysis might be slightly different after the ML model settings have been optimised. However, performing the optimisation using all features might also yield different results than performing it on the subset of selected features. As the purpose of the feature analysis is to estimate the explanatory power of the input features and perform a feature selection we therefore perform the SHAP analysis before optimising the models. Fig. 2 shows the SHAP values for each input feature for the two models. In both models the most dominant feature is the vertical polarisation of the 6 GHz brightness temperature (tb6V), on average changing the predicted values by 23 percentage points (pp) and 21 pp for XGB and NN, respectively. Other than tb6V, the SHAP values of the different input features differ greatly between the two models. Furthermore, the general magnitudes of importances are very different for the two models, with lower average importances in the XGB model. Therefore, the choice is made to keep all features in the NN model and only reduce the number of features in the XGB model by using a threshold of 0.1 pp. The features with an average impact on the XGB model predictions higher than this threshold are therefore included in the XGB model. These are marked with an asterisk in Table 1.

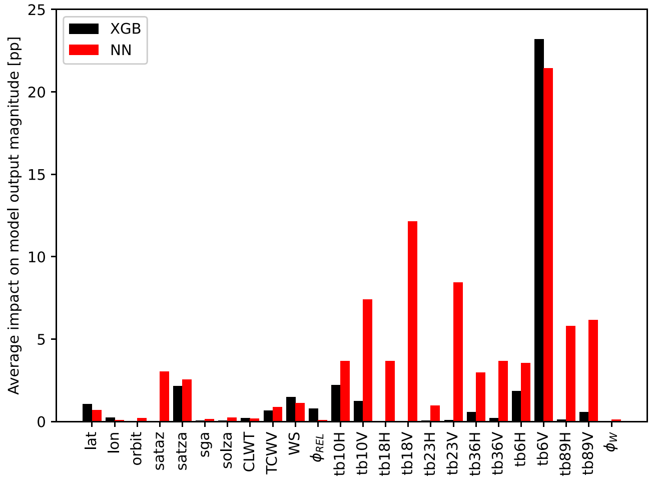

Fig. 2. SHAP feature importance analysis for XGB (black) and NN (red) with the average magnitude of impact on the model prediction (in percentage points) on the ordinate. Each bar represents the importance of a single input feature.

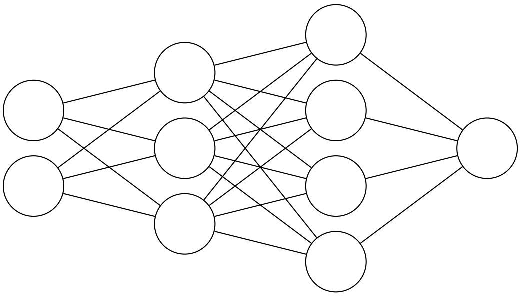

Fig. 3. Example of a neural network with an input layer consisting of two input neurons, two hidden layers with three and four neurons, respectively, and an output layer consisting of one neuron.

3.2. XGBoost

Extreme Gradient Boosting (XGBoost, here referred to as XGB) is a supervised machine learning model for when working with tabulated data. It has shown to provide state-of-the-art results on both classification and regression problems (Zhang et al., 2017; Liu et al., 2020). Here, a brief introduction of XGB is given and for a more detailed description the reader is referred to e.g. Chen and Guestrin (2016).

XGB is a so-called tree-based ML model, which means that it consists of decision trees (Breiman et al., 1984). A decision tree divides the input data into different regions with separate parameters for each region, such that the structure of the model resembles that of a tree. In XGB, trees are built sequentially, which allows a tree to learn from previous trees through a method called gradient boosting (Friedman, 2001). XGB is based on extreme gradient boosting, which is a highly scalable gradient boosting method with a sophisticated sparsity-aware algorithm for parallel tree learning (Chen and Guestrin, 2016). In gradient boosting algorithms, prior knowledge of trees and splitting is used to build better trees, since every tree is validated as it is built. Thus, each new tree will be better than the previous. All trees then contribute to the final prediction through a weighted average. The extreme gradient boosting algorithm uses a learning rate (also called shrinkage) to update the trainable model parameters in the same way as a neural network (Chen and Guestrin, 2016).

A well-known issue with machine learning is the risk of overfitting, which means that the model becomes good only at predicting data from the training dataset and performs poorly when presented to unseen data, i.e. the generalisation ability of the model becomes poor (Goodfellow et al., 2016). The problem of overfitting in tree-based models is a well-studied topic, with several different methods suggested for preventing overfitting, such as pre-pruning, post-pruning and early stopping (Esposito et al., 1997; Ying, 2019). To avoid overfitting in the XGB model early stopping, based on the mean absolute difference (MAD) metric of a test dataset, is used in the training of the model.

3.3. Neural network (NN)

Neural networks (NNs) are inspired by the functionality of the neural system and are one of the most well-known ML techniques for supervised learning. Here, a short description of the neural network used in this study is given. For a more comprehensive and detailed description of neural networks see e.g. LeCun et al. (2015), Nielsen (2015) and Goodfellow et al. (2016).

Fig. 3 shows an illustration of an NN, which consists of an input layer, two hidden layers and an output layer. Each layer consists of one or more neurons (also called nodes or units). It is through the input layer that the NN receives its input and the output layer produces the output of the model. The number of neurons corresponds to the number of inputs and outputs, respectively. In Fig. 3, the NN receives two inputs and as the output layer only consists of one neuron, only one output is produced. The layers in between the input and output layers are called hidden layers as they are neither input nor output layers, but are hidden in between. The connections between neurons in the different layers each has a weight associated with it, which indicates the weight given to the respective input information. Furthermore, the connections between the neurons in the hidden layer(s) are associated with an activation function. The purpose of the activation function is to introduce non-linearity to the system, as well as to allow for variable importance and to introduce an on–off behaviour in the response of the model to the input data. In this study, we have used the multilayer perceptron model, which is a fully-connected feed-forward NN, applying the backpropagation method (Hecht-Nielsen, 1992) during training. The NN retrieval algorithm used in this study is implemented using the TensorFlow interface (Abadi et al., 2015). As for the XGB model, early stopping is applied to ensure that the NN is not overfitting.

3.4. Optimisation

ML models have two types of parameters; (i) parameters which the ML model estimates during the training process; and (ii) hyperparameters, which need to be assigned prior to the model training. These hyperparameters can be tuned in order to improve the performance of the model. This is done through a process called hyperparameter optimisation (HPO). There exist several methods for performing HPO, two of them being through gridded and randomised searches (Liashchynskyi and Liashchynskyi, 2019; Yang and Shami, 2020). The gridded search offers a thorough scan of the entire desired parameter space, whereas the randomised search only scans a fraction of the desired parameter space, based on the chosen hyperparameter distribution function, thereby decreasing the computational cost.

HPO for the XGB model is performed using the scikit-learn RandomizedSearchCV. For the NN model, on the other hand, the scikit-learn GridSearchCV was used. The reason for this was that it was difficult to define the parameter space and obtain a satisfactory performance with the randomised algorithm for the NN model. Hence, a gridded search was performed instead. The parameter space was easier to define for the XGB model, which is why a randomised algorithm was used in order to reduce the computational cost. To ensure that the models are not overfitting to the training data in the HPO, k-fold cross-validation (e.g. Grimm et al., 2017; Berrar, 2018) in five folds is used. The hyperparameters considered, their prior distributions and the final value for each hyperparameter obtained from the HPO is shown in Table 3. It should be noted, that the entire possible hyperparameter space has not been investigated. Searching for more combinations in an extended space, might alleviate the strong link currently seen between the prior distributions and their final chosen values.

3.5. Regression model

The statistical regression model retrieval algorithm described in Alerskans et al. (2020) is used as benchmark in order to compare the performance of the ML retrievals. The regression (RE) model consists of a two-stage WS regression model followed by a two-step SST retrieval regression model. The first step uses a global algorithm to retrieve an initial estimate of wind speed. In the second step, these initial estimates are used to derive localised retrieval algorithms. Both steps in the WS retrieval algorithm use AMSR-E brightness temperatures and are regressed against CCMP wind speeds. The SST retrieval algorithm applies localised retrievals for both steps. In the first step, regression coefficients are derived locally for fixed latitude intervals and ascending and descending passes, respectively, whereas the second step uses localised SST and WS algorithms. Both steps in the SST retrieval algorithm use AMSR-E brightness temperatures, Earth incidence angle, retrieved wind speeds from the WS retrieval algorithm, and the relative angle between satellite azimuth angle and wind direction. For more information on the RE model, see Alerskans et al. (2020).

Table 3
The hyperparameters optimised for the XGB and NN models, their prior distributions and final values obtained through the HPO. The prior distributions for the XGB randomised HPO include a uniform distribution, with the minimum and maximum values specified, a Poisson distribution, with the expected separation indicated, and a normal (Gaussian) distribution, with the mean and standard deviation indicated. For the NN, the gridded HPO search intervals are shown.

	Hyperparameter	Prior distribution	Final value
XGB	Number of gradient boosted trees	Poisson(100)	103
	Maximum tree depth	Poisson(25)	22
	Minimum number of incidences in a final leaf	uniform(1,5)	3
	Subsampling[a]	norm(0.6,0.1)	0.58
	Subsampling by tree[b]	norm(0.6,0.1)	0.7
	Subsampling by level[c]	norm(0.6,0.1)	0.63
	Learning rate[d]	norm(0.1,0.03)	$8.5 \cdot 10^{-2}$
	Hyperparameter	**Search range**	**Final value**
NN	Number of hidden layers	1–3	2
	Number of neurons in each hidden layer	[15,20,25,30]	20 (1st layer) 15 (2nd layer)
	Activation function in the hidden layers	[ReLU, tanh]	tanh
	Optimiser[e]	[Adam, SGD]	Adam
	Initial learning rate[f]	0.0001–0.01	$8 \cdot 10^{-4}$

[a] Fraction used to randomly select a subset of training data.
[b] Fraction used to randomly select features to train each tree.
[c] Fraction used to randomly select a subsample of the features for every new depth level reached in a tree.
[d] The rate at which the trainable model parameters are updated during the training process.
[e] Algorithm by which the weights are optimised in order to minimise a loss function.
[f] Adam uses an adaptive learning rate, hence the initial value of the learning rate is optimised here.

The RE model was developed using a previous MMD version, in which ERA-Interim data was used instead of ERA5 data, as it had not been produced yet. The subsetting for the RE model is therefore different from the two ML retrieval algorithms, also due to different needs for number of subsets. Therefore, the same data are not used for training of the RE and ML models. The two training dataset are, however, representative of each other and the RE and ML models are therefore trained on similar data. The RE model is, on the other hand, validated on the same subset as the ML models. However, this means that the validation of the RE model is likely not performed on completely independent data as some matchups in the RE training dataset likely are included in the validation dataset. However, as validation on the same data makes the results more comparable the RE model was validated on the same subset as the two ML models.

4. Results

4.1. Overall

The two ML models and the RE retrieval algorithm have been run for the validation dataset introduced in Section 2.2. The overall performance of the three retrieval algorithms, as validated against drifter in situ SSTs (SST_{insitu}), is shown in Table 4. The overall bias of the retrieved AMSR-E PMW SSTs is 0.01, 0.01 and −0.02 K for the XGB, NN and RE models, respectively. The standard deviation of the retrieved PMW SSTs versus drifter in situ SSTs is 0.36, 0.50 and 0.55 K for the XGB, NN and RE models, respectively. The XGB retrieval algorithm performs the best, with a small bias and lowest standard deviation, whereas the NN and RE retrievals perform more similarly, where the NN model has a slightly smaller standard deviation. Overall, the XGB model shows better performance with respect to the other verification metrics as well and the NN and RE models show more similar overall results, with the NN model performing slightly better.

Table 4
Overall performance of the three retrieval algorithms. The table shows the mean difference (MD), standard deviation of the difference (STD), mean absolute difference (MAD), mean squared difference (MSD) and the R^2 score of retrieved minus in situ SST. The overall sensitivity of the three models to changes in situ SST (see Section 4.3) is shown as well.

	NN	XGB	RE
MD $[K]$	0.01	0.01	−0.02
MSD $[K^2]$	0.25	0.13	0.30
MAD $[K]$	0.37	0.24	0.42
STD $[K]$	0.50	0.36	0.55
R^2	0.997	0.998	0.996
Sensitivity	0.88	0.78	0.90

Fig. 4 shows the geographical distribution of mean and standard deviation of retrieved minus in situ SSTs for the XGB, NN and RE models. For the XGB model, only few areas have biases and these are generally small, as can be seen in Fig. 4(a). At higher latitudes, especially in the Southern Ocean, areas with a slight warm bias can be seen. Small cold biases, on the other hand, can be seen for e.g. the Arabian Sea and the Pacific warm pool area. The corresponding results for the NN model (Fig. 4(c)) show more and larger areas with both warm and cold biases. Most notable are the areas of large warm biases in the Southern Ocean. The XGB model also show a warm bias for some of these areas, although not a as wide-spread nor as large. Cold biases can be seen for the NN model for e.g. the higher northern latitudes and close to the tip of South America, as well as for the Arabian Sea. The RE model (Fig. 4(e)) also has a larger warm bias in the high latitudes, especially for the southern hemisphere, which was confirmed to be linked to undetected sea ice. This area is the same area where both the XGB and NN models also exhibit warm biases. In addition, areas of large warm biases are seen for the west coast of North and Central America. Furthermore, the RE model also exhibits a similar cold bias for the Arabian Sea and the Pacific warm pool area, much like the two ML models, however more pronounced and wide-spread. Otherwise, areas of both warm and cold biases can be seen. In general, no clear latitudinal pattern for the spatial distribution of bias can be seen for any of the models.

The geographical distribution of standard deviation, on the other hand, shows a clear latitudinal pattern for all three models. Higher standard deviations are seen for the higher latitudes, and lower standard deviations are found at lower latitudes. Furthermore, all three models exhibit higher standard deviations for the dynamical ocean regions, such as the Gulf Stream extension, the Aghulas Current and the Kuroshio Current, as well as off the east coast of Argentina. These dynamical ocean regions are areas with large SST gradients over smaller scales. Comparing the retrievals from AMSR-E, which has a resolution of several kilometres, with a point observation in one of these regions will therefore add to the discrepancies, as discussed in Alerskans et al. (2020) and Nielsen-Englyst et al. (2018). Higher standard deviations are therefore expected for the dynamical ocean regions and is not

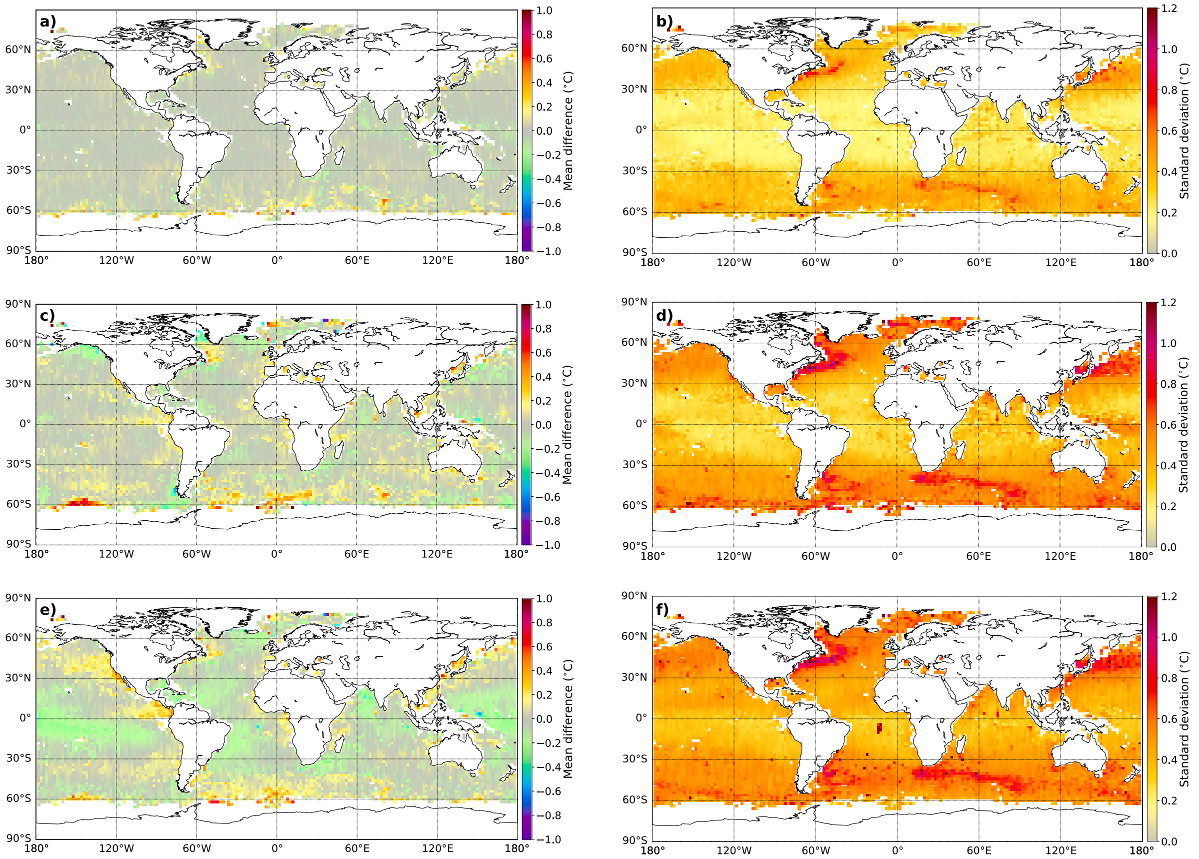

Fig. 4. The geographical distribution of mean and standard deviation of SST_r minus in situ SST for XGB (a) and b)), NN (c) and d)) and RE (e) and f)). The statistics have been calculated on a 2 × 2 degree spatial grid with a minimum of 50 matchups per grid cell.

necessarily an indication of the quality of the retrievals. Overall, the XGB model shows smaller standard deviations, whereas the magnitude of the standard deviation for the NN and RE models are more similar. However, local areas with high standard deviations are seen for the RE model. Most notably is the relatively larger area of higher standard deviation in the South Atlantic. Neither the NN nor the XGB model shows such high standard deviations for this area, although they seem to have locally slightly larger standard deviations for the same area.

4.2. Dependencies

To further investigate the performance of the three retrievals, the dependency of the retrieved minus in situ SST as a function of in situ SST and WS is shown in Fig. 5. The dependence of the retrieved SSTs on wind speed reflects the change in the sea surface roughness and hence the emissivity of the ocean. It should be noted here that the wind speed used for the XGB and NN models is the ERA5 wind speed, whereas the wind speed used for the RE model is the CCMP wind speed, which is also what was used in the RE retrieval algorithm (Alerskans et al., 2020).

The binned statistics for retrieved SST minus in situ SST as a function of in situ SST (Fig. 5a) show a warm bias for cold SSTs ($SST_{insitu} <$ 1 °C) and a cold bias for warm SSTs ($SST_{insitu} >$ 30 °C) for all three models. The standard deviation can be seen to decrease with increasing SST, except for very warm SSTs where a sharp increase can be seen, at least for the two ML models. This is the same interval for which the cold bias is seen. Overall, all three models show similar biases. In general, the XGB model has slightly lower standard deviation, whereas the NN and RE models both have similar standard deviations. However, for the edges of the SST interval, all three models exhibit a sharp increase in standard deviation for very cold temperatures, whereas for very warm temperatures a large increase in standard deviation is seen only for the two ML models.

Fig. 5b shows no significant dependence of the retrieved SST for the XGB and NN models as a function of WS with respect to bias. Only a small bias can be seen for high wind speeds for both models. For the RE model, on the other hand, a small bias can be seen for wind speeds of around 4–8 ms^{-1}, as well as for high wind speeds. The standard deviation increases with increasing wind speed for all three models, most notably for the NN and RE models which show standard deviations of up to almost 1 ms^{-1} for very high wind speeds. Overall, the XGB model has smaller standard deviations, with the NN and RE models exhibiting larger standard deviations.

4.3. Sensitivity

The SST sensitivity is a measure of the change in retrieved SST per unit change in the true SST (Merchant et al., 2009). Ideally, the SST sensitivity is 1 K K^{-1}, however, several geophysical factors can have an impact on the sensitivity, such as water vapour, cloud water and sea surface roughness. Here, a modified version of the forward model developed by Wentz and Meissner (2000) (Nielsen-Englyst et al., 2018) is used to estimate the SST sensitivity of each of the retrieval algorithms. The forward model relates the relevant geophysical factors to brightness temperatures, and the sensitivities to the geophysical factors show good agreement with those found by Prigent et al. (2013) using the fast radiative transfer model, RTTOV (Nielsen-Englyst et al., 2021). Two sets of brightness temperature simulations were performed for the sensitivity subset. The first set used ERA5 TCLW, TCWV and WS input together with modified drifting buoy SSTs, where 1 °C was added (SST_{+1}). The second set, on the other hand, used the same ERA5 data, but now together with modified drifting buoy SSTs where

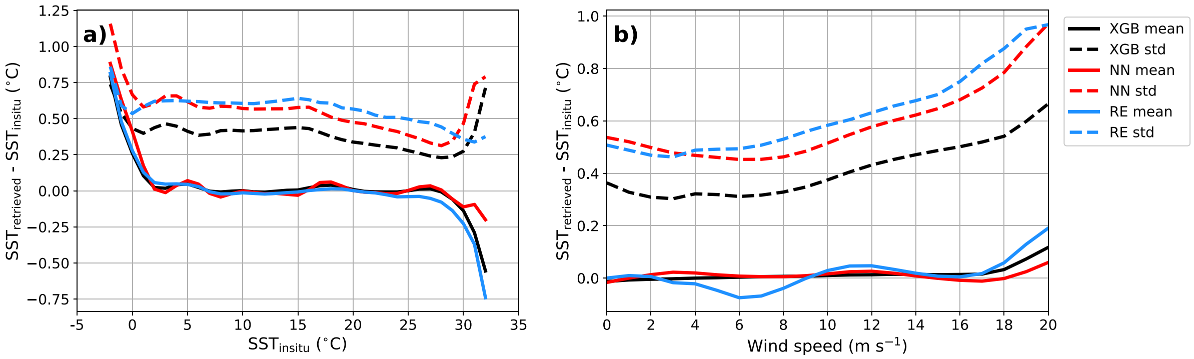

Fig. 5. Retrieved SST minus in situ SST as a function of (a) in situ SST and (b) wind speed. Solid lines show the mean and dashed lines show the standard deviation for the XGB (black), NN (red), and RE (blue) retrieval algorithms. A minimum of 50 matchups were used for the statistics calculations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

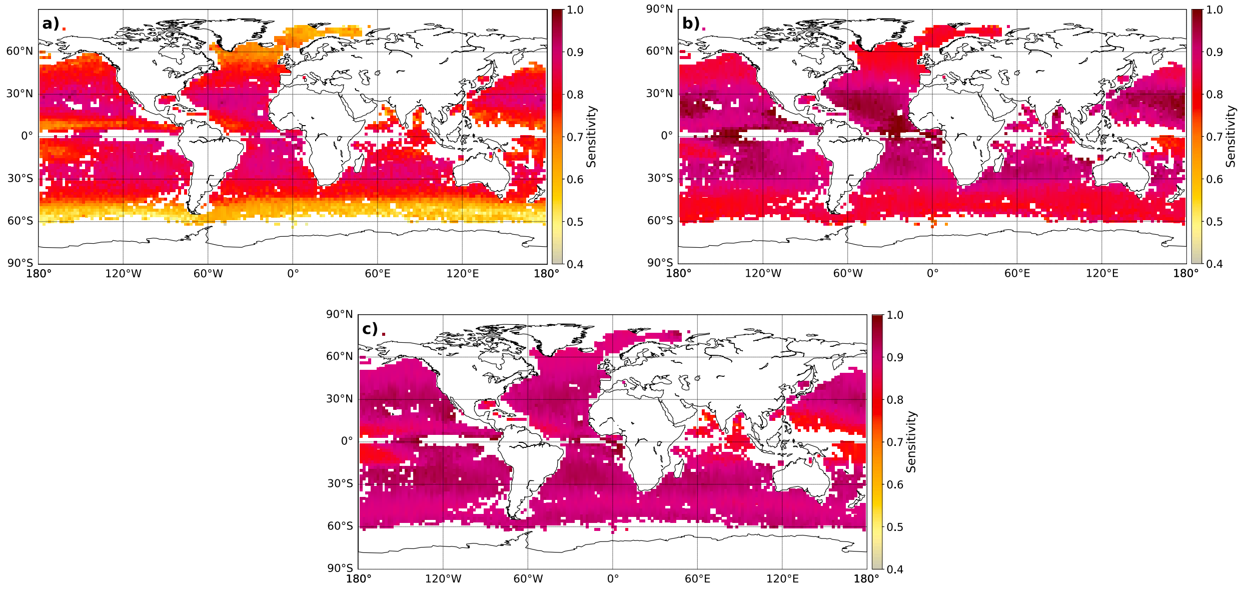

Fig. 6. The geographical distribution of sensitivity with respect to changes in SST for (a) the XGB, (b) the NN, and (c) the RE retrieval algorithms. The statistics have been calculated on a 2 × 2 degree spatial grid with a minimum of 50 matchups per grid cell.

1 °C was subtracted (SST$_{-1}$). These two subsets of simulated brightness temperatures were propagated through the retrievals to obtain new SSTs - SST$_{r,+1}$ and SST$_{r,-1}$. The sensitivity was then calculated based on these new SST retrievals, such that the sensitivity is given by (SST$_{r,+1}$ − SST$_{r,+1}$)/2, which ideally should be 1 as the two retrieved SSTs ideally should differ by 2 °C. The average sensitivity for the XGB, NN and RE models were found to be 0.78, 0.88 and 0.90, respectively.

Fig. 6 shows the geographical distribution of sensitivity for the three models. Both the XGB and NN models have higher sensitivities for lower latitudes and smaller sensitivities for higher latitudes. Areas with relatively lower sensitivities can be seen in the Pacific warm pool area as well as in the Arabian Sea. The two ML models show the same geographical patterns in the sensitivity results, however, overall the sensitivity for the NN model is higher than for the XGB model. The RE model shows some similar geographical dependencies as the other two models, such as lower sensitivities for the Pacific warm pool area and the Arabian Sea, where minimum sensitivities of 0.50 can be found. Overall, higher sensitivities are mainly found for lower latitudes, however, areas with lower sensitivities are also present at lower latitudes. Furthermore, high sensitivities are also found for higher latitudes. Hence, the same clear latitudinal pattern as for the other two models is not present for the RE model. Overall, the sensitivity of the RE model is slightly higher compared to the NN model, especially for the higher latitudes.

The dependency of the sensitivity on in situ SST is shown in Fig. 7. Here, a clear dependence can be seen, with lower sensitivities for colder SSTs and higher sensitivities for warm SSTs. However, a sharp decrease in sensitivity can be seen for very warm SSTs for all three models. The XGB model shows lowest sensitivity for all SSTs, with minimum sensitivities of less than 0.5 seen for the very cold SSTs. The NN model can be seen to have consistently higher sensitivities than the XGB model; more than 0.2 for colder SSTs where the largest difference can be seen. The RE model, however, has consistently higher sensitivities than the NN model, except for very warm SSTs. Furthermore, the RE model does not exhibit the same dependence on SST as for the other two models. Instead, a more consistent sensitivity with SST can be seen, except for very warm SSTs for which the sensitivity drops significantly.

5. Discussion

The XGB model provides the lowest bias and standard deviation with a mean bias of 0.01 K and standard deviation of 0.36 K. The NN and RE models have biases of 0.01 and −0.02 K and standard deviations of 0.50 and 0.55 K, respectively. The results obtained here for all three models are comparable to, and in the case of the XGB model even better than, previous validation results of AMSR-E PMW SST retrievals. O'Carroll et al. (2008) report a bias of 0.02 K and a standard deviation of 0.46 K, whereas Gentemann (2014) obtained

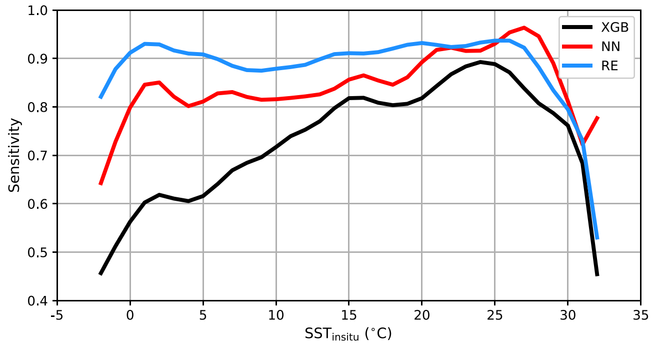

Fig. 7. Sensitivity as a function of in situ SST for the XGB (black), NN (red), and RE (blue) retrieval algorithms. A minimum of 50 matchups were used for the statistics calculations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

retrieved SSTs with a bias of −0.05 K and a standard deviation of 0.48 K. Both Nielsen-Englyst et al. (2018) and Alerskans et al. (2020) report similar validation results with biases ±0.02 K and standard deviations of around 0.46 K.

All three retrieval algorithms exhibit warm biases for higher latitudes, especially in the Southern Ocean. The areas close to the poles have previously been confirmed to be partly related to sea ice contamination (Alerskans et al., 2020). Furthermore, it is also believed that wind induced effects could play a role in the larger biases. Fig. 5 showed an increase in bias for strong wind speeds and since the Southern Ocean is characterised by very strong wind speeds (Young, 1999) it is likely that the retrievals are effected. Furthermore, the RE model, and the NN model to some extent, have a cold bias in the Arabian Sea, as well as in the Pacific warm pool area. These regions are characterised by warm temperatures and the RE model especially has a cold bias for very warm SSTs (Fig. 5). However, neither the NN nor the XGB models show the same cold bias for the Pacific warm pool area, which indicates that the RE model is impacted by other factors and at the moment it is still unclear why this large cold bias is seen.

A latitudinal distribution of the standard deviation of the retrieved minus in situ SSTs is seen for all three models, with lower values for low latitudes and higher values for high latitudes. Gentemann (2014) reported similar results, with lowest standard deviation between 40°S and 40°N and higher values for increasing latitude. The PMW SST validation results of Alerskans et al. (2020) and Nielsen-Englyst et al. (2018) both show the same latitudinal dependence as shown in Fig. 4. The greater variability near the poles can be attributed to the presence of sea ice, which can contaminate the microwave observations (see e.g. Alerskans et al., 2020) and the fact that the sensitivity of the brightness temperature to SST is lower for colder temperatures compared to warmer temperatures (Prigent et al., 2013). Higher standard deviations are also seen for the dynamical ocean regions, with a very similar pattern reported in Nielsen-Englyst et al. (2018) and Alerskans et al. (2020). This increase is not believed to be due to the performance of the models but rather due to sampling errors in these large variability regions. The large AMSR-E footprint (43 × 75 km for the 6.9 GHz resolution) is compared against point measurements from drifting buoys. The maximum allowed geodesic distance of 20 km and maximum allowed temporal difference of 4 h between the matchups of AMSR-E and buoy data will contribute to the discrepancies seen. A high-resolution IR based SST reference has been used to calculate the variability within an AMSR-E microwave footprint and a similar pattern in the variability was found in the dynamical regions. This indicates that the increased standard deviation seen in Fig. 4 is related to the larger variability found in these regions and sampling errors, and not to poor model performances.

A higher standard deviation is also seen for the RE model for a small area in the South Atlantic Ocean. In addition, a slightly larger cold bias is seen as well. Neither the NN nor the XGB models show the same clear pattern, although for the XGB a slight increase in standard deviation can be seen. This area is a known region with strong RFI (Gentemann and Hilburn, 2015). The pre-processing of the data included an RFI mask, which has previously been applied and successfully removed RFI (Gentemann and Hilburn, 2015; Nielsen-Englyst et al., 2018). Alerskans et al. (2020) showed that it is also possible to exclude RFI contaminated matchups by comparing the baseline retrieved SSTs to additionally retrieved SSTs, for which the 10 GHz and 18 GHz channels were excluded. A similar filter could be applied here. Neither the NN nor the XGB models show the same pronounced increase in standard deviation, which indicates that they might not be as sensitive to RFI contamination. Further work is needed to investigate this.

The dependence of the retrieved minus in situ SST on in situ SST (Fig. 5) shows elevated bias and standard deviations for very warm and very cold SSTs for all three models. The behaviour for cold SSTs is believed to be partly due to sea ice contamination, as previously discussed. Furthermore, it is well-known that ML models have a hard time predicting extreme values (Ribeiro and Moniz, 2020). If not enough training data for a certain range, e.g. cold SSTs, are included, the ML models will have a hard time learning how to predict these cold SSTs. In both the NN and XGB, the optimisation is performed based on the minimisation of a loss function. As this loss function measures the average performance of the model across the domain of the target variable, the most abundant cases will have the largest impact on the model performance. Rare cases will have an almost negligible effect and the performance of the model for these cases will therefore suffer (Ribeiro and Moniz, 2020). As such, the ML models will have a hard time retrieving SSTs for the very cold and very warm SSTs, as there are not many matchups for these cases. In addition, extrapolation of predictions for data outside the training data ranges poses a problem for ML models (Xu et al., 2020). To improve the results of the ML models for the extreme ends, more training data is needed for these cases. Another option would be to modify the loss function by scaling it during the training process such that a wrong prediction of the rarer cases would have a larger impact on the model. The uneven distribution of the training data could therefore perhaps be partly offset. Yet another possibility is to train the ML models in a similar way as the RE model. By training separate instances of the ML models for e.g. the very cold and very warm SSTs a better performance might be obtained. The RE model is also seen to perform worse for the very cold and very warm SSTs. In this case, it is most likely related to the training of the model. As the last step of the RE model uses local SST and wind speed retrievals, the training data was binned into SST and wind speed bins. For the extreme ends, there are not many training examples and a minimum number of matchups for each bin was required in order to obtain robust statistics. Therefore, if there were not enough matchups in a bin, the regression coefficient from the closest SST and wind speed bin is used instead. Hence, for the very cold and very warm SSTs, as well as for the very high wind speeds, the regression coefficients are obtained from nearby bins, which might not accurately represent the relationship for the current bin.

All three models show an increase in standard deviation with wind speed, as well as a slightly higher bias for very high wind speeds. The increased uncertainty in retrieved SST for larger wind speeds is well known (see e.g. Alerskans et al. (2020)). It is likely to be related to the surface roughness and the physical characteristics that the sensitivity of the brightness temperature to wind speed increases as the wind speed (i.e. surface roughness) increases and when white foam appears on the surface (Kilic et al., 2018).

The geographical pattern of sensitivities for the NN and XGB models, with higher sensitivities for lower latitudes and lower sensitivities for higher latitudes, is similar to what was reported by Nielsen-Englyst et al. (2018) using an OE algorithm. As the sensitivity of the brightness temperatures to SST decreases with colder SSTs (Prigent et al., 2013; Nielsen-Englyst et al., 2021) higher latitudes are expected to be associated with lower sensitivities and lower latitudes with higher sensitivities. Gentemann et al. (2009) reported sensitivities of 0.39 for SSTs of 0 °C and 0.65 for SSTs of 30 °C. The geographical distribution of sensitivity for the two ML models are therefore consistent with the expected distribution. The RE model, on the other hand, does not show such a clear latitudinal pattern. There are regions with both relatively higher and lower sensitivities found at lower latitudes. As discussed in Alerskans et al. (2020), the absence of a clear latitudinal dependence is thought to be related to the retrieval algorithm itself, more specifically to the binning performed. The RE model is trained on binned data such that separate regression coefficients are obtained for each bin. If the SST variability within a bin is small, the RE algorithm will fit very well to the SST but may experience a lower sensitivity (Alerskans et al., 2020).

It was seen in Fig. 2 that the input feature which impacts the output of the XGB model the most is, by far, tb6V. The other input features are at least an order of magnitude less important. For the NN model, tb6V is also the most important feature, however, several other features also contribute significantly to the model performance, such as tb10V, tb18V and tb23V. All of the channels have a relatively large impact on the performance of the NN model and are therefore included, whereas all channels except the 23 and 18 GHz channels are included in the XGB model. Nielsen-Englyst et al. (2021) has previously investigated the importance of different frequency channels using both a physically based and a statistically based retrieval algorithm by including different subsets of the AMSR-E frequency channels (considering the 6–36 GHz frequency range) and validating the resulting SST retrievals against independent drifting buoy observations. Nielsen-Englyst et al. (2021) found that the most important channels for SST retrievals are the 6 GHz channels, which is in agreement with the feature importance analysis for both the NN and XGB models. Following the 6 GHz channels, the 10 and 18 GHz channels were found to be the most important using both the physically and statistically based models (Nielsen-Englyst et al., 2021). The statistical algorithm showed a clear improvement in performance when more channels are included, while the physical algorithm showed less variation among the channel subsets, and it actually performed quite well by only including 6 and 10 GHz. This is similar to the XGB model, which also relies mostly on the 6 and 10 GHz channels (see Fig. 2). The NN model, on the other hand, is more similar to the regression based algorithm in the sense that it performs better when more information is included, as is evident on the more even distribution of importances.

The XGB model was found to perform best with respect to standard deviation but worst with respect to sensitivity. This might be related to a poor generalisation ability of the XGB model, which implies a problem with overfitting. However, as mentioned in Section 3.2, the XGB model is run with early stopping in order to prevent overfitting and no overfitting was observed when analysing the training and generalisation errors. To investigate this issue further, simpler XGB models could be trained and a comparison between performances with respect to standard deviation, bias and sensitivity could be made in order to see if the problem is related to overfitting. The low sensitivity might also be related to the input features used. Even though a feature importance analysis was performed in order to only select the most important input features, the exclusion of some input features might negatively influence the performance of the XGB model. Nielsen-Englyst et al. (2021) found that the 18 GHz channels were important for the both a physically based and a statistically based model. The XGB model includes neither the 18 GHz channels nor the 23 GHz channels, the latter which have been found to be sensitive to atmospheric water content (Nielsen-Englyst et al., 2021). To investigate if the inclusion of some of the excluded features could affect the performance of the model, several new instances of the XGB model could be trained where some of the now-excluded input features are included. The performance of these models could then be compared with the performance of the current XGB model, especially with respect to sensitivity.

The RE and the two ML models are all statistically based retrieval algorithms. However, the RE model used here is constrained to predefined linear relationships (although it can be expanded to include non-linearities), whereas the ML models allow non-linear relationships between input features and retrieved SST. The main advantage of ML models is that they allow approximation of complex functions as they are considered universal approximators (Hornik et al., 1990; Cybenko, 1989; Hornik, 1991). They therefore allow for the learning of new relationships without prior assumptions. This is one of the main advantages of ML-based models in comparison to more traditional regression-based algorithms. On the other hand, one of the disadvantages of ML models is related to the computational cost. The HPO of the models for example, is very computationally heavy, especially if opting for the gridded search. However, not performing an HPO can have an impact on the performance of the model, as there is no optimal model structure that suits all problems (Yang and Shami, 2020). Moreover, the training of the ML models is also more computationally expensive than the training of a linear regression based model such as the RE. Furthermore, the more complex the model, e.g. the larger the architecture and the more input features used, the slower the optimisation and training is. This not only applies to the ML models but also to the RE model. For retrieving SSTs, on the other hand, the ML models are equally as fast as the RE model.

In this study, we focused on the retrieval of PMW SSTs from AMSR-E, however, it is also possible to apply ML models to retrievals of SSTs from other satellite sensors. Initial validation results using AMSR2 show good performances with biases of 0.01 and −0.08 K and standard deviations of 0.34 and 0.44 k for an XGB and an NN model, respectively. The better validation results of AMSR2 compared to AMSR-E is in agreement with those reported in Alerskans et al. (2020). The retrieval of satellite SSTs from PMW observations using ML can also be extended to future satellite missions, such as CIMR, which is currently prepared by the ESA as a part of the Copernicus Expansion Program of the European Union (http://www.cimr.eu/; Donlon, 2020) and to the retrieval of IR satellite SSTs (Sunder et al., 2020).

The uncertainties of the retrieved SSTs have not been considered in this study but it is an important aspect that needs to be addressed in the future as uncertainties are important for a wide variety of applications, such as the use of SSTs within oceanic and atmospheric models (Merchant et al., 2017). Therefore, future work should aim at estimating the uncertainties of the retrieved SSTs for each of the two ML models. Statistical models have previously been used to estimate the uncertainty in SST retrievals, such as a regression based algorithm for the estimation of the uncertainty of the RE SSTs (Alerskans et al., 2020). More recently, Kumar et al. (2021) investigated the use of two ML models for estimating the uncertainty in satellite derived IR SSTs. Good results were obtained, showing the usefulness of ML based algorithms in uncertainty estimates. Another approach could be to train multiple algorithms to obtain an ensemble from which the uncertainties can be estimated. Future work on estimating the uncertainties of the XGB and NN SST retrievals could therefore include an investigation of these approaches.

6. Conclusions

In this study, two types of machine learning (ML) models have been assessed for the retrieval of SSTs using passive microwave (PMW) satellite observations from AMSR-E. The results have been compared with an existing state-of-the-art regression (RE) retrieval algorithm. The ML models considered were the decision tree-based algorithm Extreme Gradient Boosting (XGB) and a multilayer perceptron neural network (NN). The performance of the models was evaluated using independent in situ observations of SST from drifting buoys. The performance of the RE and NN retrieval algorithms with respect to bias and standard deviation is similar, with the NN generally performing slightly better. The XGB model performs significantly better than both the RE and

NN models with respect to standard deviation and has a similar bias. However, the SST sensitivity of both the RE and NN models is significantly higher than that of the XGB model, with the RE model having the highest sensitivity. This demonstrates the importance of including the sensitivity in the validation analysis. It is not yet understood why the XGB model performs well with respect to standard deviation but significantly worse than the other two models with respect to sensitivity. This should be further investigated, especially with respect to overfitting and selected input features.

This is an initial study meant to investigate the possibilities of using ML based algorithms for retrieval of SST from PMW observations. It shows that there is a large potential for the use of ML models but also that further work is needed in order to explore the full potential of ML based retrievals. The NN used here is a very simple form of a neural network and does not represent the full spectrum of neural networks. In order to investigate the use of neural networks for PMW SST retrievals, a study comparing different types of neural networks is needed. Similarly, more work is needed for evaluating the potential of the XGB model and other decision tree based ML models, especially with respect to sensitivity.

The main strength of ML models is that they allow for the approximation of complex functions without prior assumptions. For statistical based algorithms, such as the RE, the relationship between the input and output variables needs to be explicitly specified in the model formulation. For ML models, on the other hand, the model itself will find the best relationship between input and output variables without prior assumptions.

The ML methodology, where the algorithms select the important features based on the information in the observations and the training dataset may also be of great value in complex problems where not all physical or instrumental effects are well determined e.g. in the commissioning phase of new satellites and instruments. This initial study demonstrates that there is a large potential in the use of ML algorithms for the retrieval of SST from PMW observations.

CRediT authorship contribution statement

Emy Alerskans: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Ann-Sofie P. Zinck:** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing - review & editing. **Pia Nielsen-Englyst:** Software, Investigation, Writing – original draft, Writing – review & editing. **Jacob L. Høyer:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

Emy Alerskans reports financial support was provided by European Space Agency. Pia Nielsen-Englyst reports financial support was provided by European Space Agency. Jacob L. Hoeyer reports financial support was provided by European Space Agency.

Data availability

ICODAS version 2.5.1 is available via https://icoads.noaa.gov/products.htmland EN4 version 4.2.0 is available at https://www.metoffice.gov.uk/hadobs/en4/. The resampled L2A data product AMSR-E V12 is produced by Remote Sensing Systems (RSS) and distributed by NASA's National Snow and Ice Data Center (NSIDC). Data are available at https://nsidc.org/data/ae_l2a. The ERA-Interim reanalysis data is available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim and the ERA5 reanalysis data is available via https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

Acknowledgements

The authors are grateful for computing resources and technical assistance provided by the Danish Center for Climate Computing, a facility built with support of the Danish e-Infrastructure Corporation, Danish Hydrocarbon Research and Technology Centre, VILLUM Foundation, and the Niels Bohr Institute. The PMW algorithm development was partially funded by The European Space Agency Climate Change Initiative for Sea Surface Temperature, grant 4000109848/13/I-NB.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org, URL: https://www.tensorflow.org/.

Alerskans, E., Høyer, J.L., Gentemann, C.L., Pedersen, L.T., Nielsen-Englyst, P., Donlon, C., 2020. Construction of a climate data record of sea surface temperature from passive microwave measurements. Remote Sens. Environ. 236, 111485.

Ashcroft, P., Wentz, F.J., 2013. AMSR-E/aqua L2A global swath spatially-resampled brightness temperatures, version 3. http://dx.doi.org/10.5067/AMSR-E/AE_L2A.003, Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center.

Atlas, R., Hoffman, R.N., Ardizzone, J., Leidner, S.M., Jusem, J.C., Smith, D.K., Gombos, D., 2011. A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. Bull. Am. Meteorol. Soc. 92 (2), 157–174.

Azodi, C.B., Tang, J., Shiu, S.-H., 2020. Opening the black box: Interpretable machine learning for geneticists. Trends Genet. 36 (6), 442–455.

Berrar, D., 2018. Cross-validation. ISBN: 9780128096338, http://dx.doi.org/10.1016/B978-0-12-809633-8.20349-X.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. Bull. Am. Meteorol. Soc. 95 (9), 1431–1443. http://dx.doi.org/10.1175/BAMS-D-13-00047.1, URL: https://journals.ametsoc.org/view/journals/bams/95/9/bams-d-13-00047.1.xml.

Brasnett, B., Colan, D.S., 2016. Assimilating retrievals of sea surface temperature from VIIRS and AMSR2. J. Atmos. Ocean. Technol. 33 (2), 361–375.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, CA.

Chang, P., Jelenak, Z., Alsweiss, S., 2015. Algorithm Theoretical Basis Document: GCOM-W1/AMSR2 Day-1 EDR version 1.0.. Technical Report, URL: https://www.star.nesdis.noaa.gov/jpss/documents/ATBD/ATBD_AMSR2_Ocean_EDR_v2.0.pdf.

Chelton, D.B., Wentz, F.J., 2005. Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. Bull. Am. Meteorol. Soc. 86 (8), 1097–1116. http://dx.doi.org/10.1175/BAMS-86-8-1097, URL: https://journals.ametsoc.org/view/journals/bams/86/8/bams-86-8-1097.xml.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. pp. 785–794. http://dx.doi.org/10.1145/2939672.2939785.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control Signals Systems 2 (4), 303–314.

Dee, D.P., Uppala, S.M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d.P., et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597.

Donlon, C., 2020. Copernicus imaging microwave radiometer (CIMR) mission requirements document, version 4. European Space Agency, Noordwijk, The Netherlands.

Esposito, F., Malerba, D., Semeraro, G., Kay, J., 1997. A comparative analysis of methods for pruning decision trees. IEEE Trans. Pattern Anal. Mach. Intell. 19 (5), 476–491.

Ferreira, P., Le, D.C., Zincir-Heywood, N., 2019. Exploring feature normalization and temporal information for machine learning based insider threat detection. In: 2019 15th International Conference on Network and Service Management. CNSM, IEEE, pp. 1–7.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Statist. 1189–1232.

Gentemann, C.L., 2014. Three way validation of MODIS and AMSR-E sea surface temperatures. J. Geophys. Res. Oceans 119 (4), 2583–2598.

Gentemann, C.L., Hilburn, K.A., 2015. In situ validation of sea surface temperatures from the GCOM-w 1 AMSR 2 RSS calibrated brightness temperatures. J. Geophys. Res. Oceans 120 (5), 3567–3585.

Gentemann, C.L., Meissner, T., Wentz, F.J., 2009. Accuracy of satellite sea surface temperatures at 7 and 11 GHz. IEEE Trans. Geosci. Remote Sens. 48 (3), 1009–1018.

Gentemann, C.L., Wentz, F.J., Brewer, M., Hilburn, K., Smith, D., 2010. Passive microwave remote sensing of the ocean: An overview. Oceanogr. Space 13–33.

Good, S.A., Martin, M.J., Rayner, N.A., 2013. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J. Geophys. Res. Oceans 118 (12), 6704–6716.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, http://www.deeplearningbook.org.

Grimm, K.J., Mazza, G.L., Davoudzadeh, P., 2017. Model selection in finite mixture models: A k-fold cross-validation approach. Struct. Equ. Model. A Multidisciplinary Journal 24 (2), 246–256. http://dx.doi.org/10.1080/10705511.2016.1250638.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice hall.

Hecht-Nielsen, R., 1992. Theory of the backpropagation neural network. In: Neural Networks for Perception. Elsevier, pp. 65–93.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Mu noz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4 (2), 251–257.

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Netw. 3 (5), 551–560.

Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., Shao, L., 2020. Normalization techniques in training dnns: Methodology, analysis and application. arXiv preprint arXiv:2009.12836.

Just, A.C., De Carli, M.M., Shtein, A., Dorman, M., Lyapustin, A., Kloog, I., 2018. Correcting measurement error in satellite aerosol optical depth with machine learning for modeling PM2. 5 in the Northeastern USA. Remote Sens. 10 (5), 803.

Just, A.C., Liu, Y., Sorek-Hamer, M., Rush, J., Dorman, M., Chatfield, R., Wang, Y., Lyapustin, A., Kloog, I., 2020. Gradient boosting machine learning to improve satellite-derived column water vapor measurement error. Atmos. Meas. Tech. 13 (9), 4669–4681.

Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., Spencer, R., 2003. The advanced microwave scanning radiometer for the earth observing system (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. IEEE Trans. Geosci. Remote Sens. 41 (2), 184–194. http://dx.doi.org/10.1109/TGRS.2002.808331.

Kilic, L., Prigent, C., Aires, F., Boutin, J., Heygster, G., Tonboe, R.T., Roquet, H., Jimenez, C., Donlon, C., 2018. Expected performances of the copernicus imaging microwave radiometer (CIMR) for an all-weather and high spatial resolution estimation of ocean and sea ice parameters. J. Geophys. Res. Oceans 123 (10), 7564–7580.

Kotsiantis, S.B., Kanellopoulos, D., Pintelas, P.E., 2006. Data preprocessing for supervised leaning. Int. J. Comput. Sci. 1 (2), 111–117.

Kumar, C., Podestá, G., Kilpatrick, K., Minnett, P., 2021. A machine learning approach to estimating the error in satellite sea surface temperature retrievals. Remote Sens. Environ. 255, 112227.

Le Traon, P.-Y., Antoine, D., Bentamy, A., Bonekamp, H., Breivik, L., Chapron, B., Corlett, G., Dibarboure, G., DiGiacomo, P., Donlon, C., et al., 2015. Use of satellite observations for operational oceanography: recent achievements and future prospects. J. Oper. Oceanogr. 8 (sup1), s12–s27.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, A., Taylor, P., Kalpathy-Cramer, J., Tufail, A., 2017. Machine learning has arrived!. Ophthalmology 124, 1726–1728. http://dx.doi.org/10.1016/j.ophtha.2017.08.046.

Liang, X., Yang, Q., Nerger, L., Losa, S.N., Zhao, B., Zheng, F., Zhang, L., Wu, L., 2017. Assimilating copernicus SST data into a pan-arctic ice–ocean coupled model with a local SEIK filter. J. Atmos. Ocean. Technol. 34 (9), 1985–1999.

Liashchynskyi, P., Liashchynskyi, P., 2019. Grid search, random search, genetic algorithm: A big comparison for NAS. CoRR abs/1912.06059, arXiv:1912.06059.

Liu, H., Li, Q., Bai, Y., Yang, C., Wang, J., Zhou, Q., Hu, S., Shi, T., Liao, X., Wu, G., 2021. Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. Remote Sens. Environ. 256, 112316.

Liu, Y., Xia, X., Yao, L., Jing, W., Zhou, C., Huang, W., Li, Y., Yang, J., 2020. Downscaling satellite retrieved soil moisture using regression tree-based machine learning algorithms over Southwest France. Earth Space Sci. 7 (10), e2020EA001267.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc., pp. 4765–4774, URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Maeda, T., Taniguchi, Y., Imaoka, K., 2015. GCOM-W1 AMSR2 level 1R product: Dataset of brightness temperature modified using the antenna pattern matching technique. IEEE Trans. Geosci. Remote Sens. 54 (2), 770–782.

Maeda, T., Tomii, N., Seki, M., Sekiya, K., Shibata, A., 2020. Sea-surface-temperature retrieval at higher spatial resolution in the satellite-Borne microwave radiometer AMSR2 follow-on mission. IEEE Geosci. Remote Sens. Lett. 18 (2), 336–340.

Meissner, T., Wentz, F.J., 2012. The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and earth incidence angles. IEEE Trans. Geosci. Remote Sens. 50 (8), 3004–3026.

Merchant, C.J., Embury, O., Bulgin, C.E., Block, T., Corlett, G.K., Fiedler, E., Good, S.A., Mittaz, J., Rayner, N.A., Berry, D., et al., 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Sci. Data 6 (1), 1–18.

Merchant, C., Embury, O., Le Borgne, P., Bellec, B., 2006. Saharan dust in nighttime thermal imagery: Detection and reduction of related biases in retrieved sea surface temperature. Remote Sens. Environ. 104 (1), 15–30.

Merchant, C., Harris, A., Murray, M., Zavody, A., 1999. Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison. J. Geophys. Res. Oceans 104 (C10), 23565–23578.

Merchant, C., Harris, A., Roquet, H., Le Borgne, P., 2009. Retrieval characteristics of non-linear sea surface temperature from the advanced very high resolution radiometer. Geophys. Res. Lett. 36 (17).

Merchant, C.J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., De Leeuw, G., et al., 2017. Uncertainty information in climate data records from earth observation. Earth Syst. Sci. Data 9 (2), 511–527.

Merchant, C.J., Saux-Picart, S., Waller, J., 2020. Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. Remote Sens. Environ. 237, 111590.

Minnett, P., Alvera-Azcárate, A., Chin, T., Corlett, G., Gentemann, C., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., et al., 2019. Half a century of satellite remote sensing of sea-surface temperature. Remote Sens. Environ. 233, 111366.

Monzikova, A., Kudryavtsev, V., Reul, N., Chapron, B., 2017. On the upper ocean response to tropical cyclones: Satellite microwave observation. In: 2017 Progress in Electromagnetics Research Symposium-Fall. PIERS-FALL, IEEE, pp. 2437–2444.

Moschos, E., Schwander, O., Stegner, A., Gallinari, P., 2020. Deep-SST-Eddies: A deep learning framework to detect oceanic eddies in sea surface temperature images. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 4307–4311.

Nielsen, M.A., 2015. Neural Networks and Deep Learning, Vol. 25. Determination press San Francisco, CA.

Nielsen-Englyst, P., Høyer, J.L., Alerskans, E., Pedersen, L.T., Donlon, C., 2021. Impact of channel selection on SST retrievals from passive microwave observations. Remote Sens. Environ. 254, 112252.

Nielsen-Englyst, P., L Høyer, J., Toudal Pedersen, L., L Gentemann, C., Alerskans, E., Block, T., Donlon, C., 2018. Optimal estimation of sea surface temperature from AMSR-E. Remote Sens. 10 (2), 229.

Ning, J., Xu, Q., Wang, T., Zhang, S., 2018. Upper ocean response to super typhoon soudelor revealed by different SST products. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 6063–6066.

O'Carroll, A.G., Armstrong, E.M., Beggs, H.M., Bouali, M., Casey, K.S., Corlett, G.K., Dash, P., Donlon, C.J., Gentemann, C.L., Hø yer, J.L., et al., 2019. Observational needs of sea surface temperature. Front. Mar. Sci. 6, 420.

O'Carroll, A.G., Eyre, J.R., Saunders, R.W., 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. J. Atmos. Ocean. Technol. 25 (7), 1197–1207.

Paul, S., Huntemann, M., 2021. Improved machine-learning-based open-water–sea-ice–cloud discrimination over wintertime antarctic sea ice using MODIS thermal-infrared imagery. Cryosphere 15 (3), 1551–1565.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Prigent, C., Aires, F., Bernardo, F., Orlhac, J.-C., Goutoule, J.-M., Roquet, H., Donlon, C., 2013. Analysis of the potential and limitations of microwave radiometry for the retrieval of sea surface temperature: Definition of MICROWAT, A new mission concept. J. Geophys. Res. Oceans 118 (6), 3074–3086.

Prochaska, J.X., Cornillon, P.C., Reiman, D.M., 2021. Deep learning of sea surface temperature patterns to identify ocean extremes. Remote Sens. 13 (4), 744.

Rayner, N.A., Brohan, P., Parker, D., Folland, C.K., Kennedy, J.J., Vanicek, M., Ansell, T., Tett, S., 2006. Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. J. Clim. 19 (3), 446–469.

Ribeiro, R.P., Moniz, N., 2020. Imbalanced regression and extreme value prediction. Mach. Learn. 109 (9), 1803–1835.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. " Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144.

Rodriguez-Fernandez, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jimenez, C., Mahmoodi, A., Drusch, M., 2015. Soil moisture retrieval using neural networks: Application to SMOS. IEEE Trans. Geosci. Remote Sens. 53 (11), 5991–6007.

Sanò, P., Panegrossi, G., Casella, D., Marra, A.C., D'Adderio, L.P., Rysman, J.F., Dietrich, S., 2018. The passive microwave neural network precipitation retrieval (PNPR) algorithm for the CONICAL scanning global microwave imager (GMI) radiometer. Remote Sens. 10 (7), 1122.

Sanò, P., Panegrossi, G., Casella, D., Marra, A.C., Paola, F.D., Dietrich, S., 2016. The new passive microwave neural network precipitation retrieval (PNPR) algorithm for the cross-track scanning ATMS radiometer: Description and verification study over europe and africa using GPM and TRMM spaceborne radars. Atmos. Meas. Tech. 9 (11), 5441–5460.

Saux Picart, S., Tandeo, P., Autret, E., Gausset, B., 2018. Exploring machine learning to correct satellite-derived sea surface temperatures. Remote Sens. 10 (2), 224.

Shapley Ll, S., 1953. A value for n-person games. In: Contributions to the Theory of Games II, Annals of Mathematical Studies, Vol. 28.

Shibata, A., 2006. Features of ocean microwave emission changed by wind at 6 GHz. J. Oceanogr. 62 (3), 321–330.

Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. In: International Conference on Machine Learning. PMLR, pp. 3145–3153.

Shrikumar, A., Greenside, P., Shcherbina, A., Kundaje, A., 2016. Not just a black box: Learning important features through propagating activation differences. arXiv preprint arXiv:1605.01713.

Sismanidis, P., Bechtel, B., Keramitsoglou, I., Göttsche, F., Kiranoudis, C.T., 2021. Satellite-derived quantification of the diurnal and annual dynamics of land surface temperature. Remote Sens. Environ. 265, 112642.

Sunder, S., Ramsankaran, R., Ramakrishnan, B., 2020. Machine learning techniques for regional scale estimation of high-resolution cloud-free daily sea surface temperatures from MODIS data. ISPRS J. Photogramm. Remote Sens. 166, 228–240.

Wentz, F.J., Meissner, T., 2000. Algorithm Theoretical Basis Document(ATBD): AMSR Ocean Algorithm (Version 2). RSS Tech. Proposal 121599A-1, Remote Sensing Systems, Santa Rosa, CA.

Wentz, F.J., Meissner, T., 2007. Supplement 1: Algorithm theoretical basis document for AMSR-E ocean algorithms. 30, NASA: Santa Rosa, CA, USA.

Woodruff, S.D., Worley, S.J., Lubker, S.J., Ji, Z., Eric Freeman, J., Berry, D.I., Brohan, P., Kent, E.C., Reynolds, R.W., Smith, S.R., et al., 2011. ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol. 31 (7), 951–967.

Xu, K., Zhang, M., Li, J., Du, S.S., Kawarabayashi, K.-i., Jegelka, S., 2020. How neural networks extrapolate: From feedforward to graph neural networks. arXiv preprint arXiv:2009.11848.

Yang, C.-S., Kim, S.-H., Ouchi, K., Back, J.-H., 2015. Generation of high resolution sea surface temperature using multi-satellite data for operational oceanography. Acta Oceanol. Sinica 34 (7), 74–88.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316. http://dx.doi.org/10.1016/j.neucom.2020.07.061.

Ying, X., 2019. An overview of overfitting and its solutions. J. Phys. Conf. Ser. 1168, 022022. http://dx.doi.org/10.1088/1742-6596/1168/2/022022.

Young, I., 1999. Seasonal variability of the global ocean wind and wave climate. Int. J. Climatol.: J. Royal Meteorol. Soc. 19 (9), 931–950.

Zhang, C., Liu, C., Zhang, X., Almpanidis, G., 2017. An up-to-date comparison of state-of-the-art classification algorithms. Expert Syst. Appl. 82, 128–150. http://dx.doi.org/10.1016/j.eswa.2017.04.003, URL: https://www.sciencedirect.com/science/article/pii/S0957417417302397.