Per-Pixel Uncertainty Quantification and Reporting for Satellite-Derived Chlorophyll-a Estimates via Mixture Density Networks

Arun M. Saranathan^{ORCID}, *Member, IEEE*, Brandon Smith, *Member, IEEE*, and Nima Pahlevan^{ORCID}, *Senior Member, IEEE*

Abstract—**Mixture density networks (MDNs) have emerged as a powerful tool for estimating water-quality indicators, such as chlorophyll-a (Chl*a*) from multispectral imagery. This study validates the use of an uncertainty metric calculated directly from Chl*a* estimates of the MDNs. We consider multispectral remote sensing reflectance spectra (R_{rs}) for three satellite sensors commonly used in aquatic remote sensing, namely, the ocean and land colour instrument (OLCI), multispectral instrument (MSI), and operational land imager (OLI). First, a study on a labeled database of colocated in situ Chl*a* and R_{rs} measurements clearly illustrates that the suggested uncertainty metric accurately captures the reduced confidence associated with test data, which is drawn for a different distribution than the training data. This change in distribution maybe due to: 1) random noise; 2) uncertainties in the atmospheric correction; and 3) novel (unseen) data. The experiments on the labeled in situ dataset show that the estimated uncertainty has a correlation with the expected predictive error and can be used as a bound on the predictive error for most samples. To illustrate the ability of the MDNs in generating consistent products from multiple sensors, per-pixel uncertainty maps for three near-coincident images of OLCI, MSI, and OLI are produced. The study also examines temporal trends in OLCI-derived Chl*a* and the associated uncertainties at selected locations over a calendar year. Future work will include uncertainty estimation from MDNs with a multiparameter retrieval capability for hyperspectral and multispectral imagery.**

Index Terms—**Aquatic remote sensing, chlorophyll-a (Chl*a*), inland and coastal waters, Landsat-8, machine learning (ML), Sentinel-2, Sentinel-3, uncertainty.**

I. INTRODUCTION

FRESHWATER and coastal estuaries play a significant role in various aspects of human lives from furnishing drinking water to providing ecosystem services essential to human health and the economy. As such, it is critical to study

and monitor the health of these ecosystems. One important aspect in tracking the health of these ecosystems is to assess the amount of phytoplankton present in the water column; too little plant biomass may lead to their collapse, while excessive plant life may be detrimental to other wildlife [1], human health [2], and the ecosystem. In general, chlorophyll-a (Chl*a*) (the active agent for oxygenic photosynthesis presents in all phytoplankton types) serves as a proxy for biomass in the water column [3], [4], [5], [6].

Historically, estimates of near-surface concentrations of Chl*a* from optical remote sensing data have been used to monitor the health of a variety of aquatic ecosystems [5], [7], [8], [9]. Image datasets from Landsat-8's moderate-resolution operational land imager (OLI) [10], [11], [12], the multispectral instrument (MSI) aboard Sentinel-2 [13], [14], and the ocean and land colour instrument (OLCI) carried on Sentinel-3 [15] have been harnessed for monitoring Chl*a* in various water bodies. The optical remote sensing literature records various attempts and approaches to estimate Chl*a* from radiometric observations. Initial techniques designed for instruments like OLI, which do not support measurements in the red-edge (RE) spectral region, primarily focused on the use of specific ocean-color band ratios (generally a blue–green ratio) [16], [17], [18]. While such approaches have been successful for open-ocean waters, for more optically complex scenarios, such as lakes and rivers, they exhibit significant performance degradation [19], [20], [21]. For satellite instruments equipped with the RE band, models leveraging bands in the red-NIR regions have been shown to perform reasonably well in eutrophic waters with limited accuracy in ecosystems with Chl*a* < 8 mg m^{-3} [22], [23], [24]. Another popular avenue has been to leverage various machine learning (ML) algorithms, such as support vector machines [25], [26], XGB-tree algorithms [27], and different neural network-based approaches [28], [29], [30], [31], [32], [33].

While classical ML approaches are capable of modeling complex relationships between input data and target variables [34], [35], they are designed to predict a specific variable of interest, such as Chl*a*, i.e., they are designed to make a single prediction. In addition, such models also assume that the distribution of the target variable is unimodal. This assumption might be problematic when one is modeling inverse problems, such as Chl*a* retrievals, which are expected to be multimodal (due to the nonunique relationship between the input and the target variables) [36]. Recently, it has

been proposed to instead utilize a neural network variant, the mixture density network (MDN), which models the target variable as a distribution, in particular, a mixture of Gaussians conditioned on the input variable [37], [38]. This formulation accounts for the possibility that the distribution of the target variable is multimodal and has outperformed existing retrieval techniques across a variety of different water types [37], [38].

Most of the commonly used ML models in the aquatic remote sensing literature leverage labeled training datasets to uncover the complex relationships between radiometric data and biogeochemical variables. The ability of the models to accurately predict the target variable is highly data-dependent, and often, the predictions of these models are only appropriate if the test samples are drawn from the same distribution as the training samples.[1] Previous work has shown that basing real-world decisions solely on the predictions of such models without analyzing the uncertainty/confidence associated with the predictions may mislead product end-users [39], [40], [41]. Uncertainty analysis in ML attempts to capture the similarity of a given test sample to the training data distribution; this will provide the end-user important information on the validity of the ML model prediction for a specific sample. This uncertainty is distinct from the uncertainty already present in the optical measurements under consideration for such applications (see [42] and the references therein for details). Mélin et al. [43] have previously showed that the uncertainties associated with satellite remote sensing products are higher than the uncertainty associated with products from in situ measurements. Brewin et al. [44] provide an initial attempt to measure the uncertainty in the ranking/performance of a large suite of methods using statistics on the model performance. McKinna et al. [45] proposed a method to track the effects of radiometric uncertainties on the performance of ML models. Many of these papers note that the radiometric uncertainty affects the quality of the predicted products, but they do not track the inherent uncertainty associated with the ML model prediction, which is also relevant and should be determined.

The uncertainty associated with the prediction of an ML model can be broadly classified into two main categories, namely, aleatoric and epistemic uncertainty [46], [47]. Aleatoric (random) uncertainty is the uncertainty induced in the prediction due to the noise inherently present in the training data. Epistemic (knowledge-based) uncertainty refers to the ignorance of the model on a given data point. Various factors may affect the presence of epistemic uncertainty, such as: 1) a lack of sufficient training samples resembling the test sample; 2) inadequate model capacity (e.g., using a linear model to model a nonlinear relationship); and 3) the multimodality of the target variable distribution. Unlike aleatoric uncertainty, which is caused by the randomness in the data acquisition process, the epistemic uncertainty can be reduced by providing additional data while training the ML models or increasing the model's complexity and so on. For ML applications, the

---

[1]Section I in the Supplementary Material provides some insight into why uncertainty exists for predictions made by an ML model for unknown data using a toy problem.

uncertainty in the prediction is modeled in a probabilistic manner, i.e., using Bayesian approximations (since exact Bayesian inference is intractable) of the model [46]. The literature provides a variety of techniques for performing the Bayesian approximation ranging from Markov chain Monte Carlo simulations [48], variational inference [49], Monte Carlo dropout [50], and Laplacian approximations [51], among others (see [46] and references therein for a complete review of uncertainty estimation techniques for neural-network-based ML models). In the aquatic remote sensing/ocean-color literature, another uncertainty estimation approach cited in Mélin et al. [42] is to build a supplementary inverse network to reconstruct the original R_{rs} from the model output. The samples with high reconstruction error are considered "out-of-scope." While these approaches appear to provide notions of uncertainty and (in some cases) include significant theoretical analysis, they require both modification and retraining of the deep learning models to accommodate them. The training of these models is both much more challenging (due to increased complexity) and computationally expensive.

Here, we will leverage MDN's inherent ability to estimate the probability distributions associated with a target variable. These distributions can be directly used to approximate the uncertainty without retraining or modifying the model. Choi et al. [40] proposed an MDN-specific uncertainty technique, wherein the uncertainty associated with a prediction of the MDN is shown to be the variance of a mixture of Gaussians estimated by the MDN. The authors further show that the variance (and, in turn, the uncertainty) can be easily decomposed into aleatoric and epistemic components. While this approach [40] appears to provide a quick and efficient way to estimate the uncertainty associated with an MDN prediction, it is essential to verify that uncertainties estimated in this way are reasonable and consistent for the issues commonly affecting aquatic remote sensing data. The consistency and reasonability of the estimated metric are verified for some of the well-known distortion processes affecting satellite remote sensing data, namely: 1) highly noisy data; 2) novel (unseen) data samples different from the training data; and 3) structured distortions, due to uncertainties in the atmospheric correction, on the predictions of the MDN models. In addition, we will perform experiments that: 1) verify the link between estimated uncertainties and predictive error and 2) demonstrate that the estimated uncertainty metrics can be interpreted as a physical bound on prediction errors. For practical purposes, the proposed uncertainty metric will be used to: 1) create uncertainty maps for near-simultaneous acquisitions of OLCI, MSI, and OLI to showcase the generation of consistent multimission uncertainty products and 2) evaluate temporal variability in OLCI-derived Chl$a$ predictions and the associated uncertainties for selected locations.

With this content information, the rest of this article is organized as follows. Section II will describe the different radiometric datasets used for analysis in this study. Section III will describe the various ML tools and methods used for the estimation of Chl$a$, the associated uncertainty metric (σ_{UNC}), and the design of the various experiments on the in situ dataset used to validate the MDN uncertainty metric for R_{rs} data.

Section IV contains the results of these validation experiments. Section V illustrates some possible practical applications of the uncertainty metric both on labeled and satellite data. Section VI will describe a set of observations based on the experiments in Sections IV and V. Finally, Section VII will provide the conclusions and describe avenues for future research.

II. DATASETS

In this article, two datasets were primarily used: 1) an in situ radiometric dataset that was used for model development and testing of the proposed uncertainty metric and 1) satellite images used to qualitatively examine the realism and practicality of the MDN-generated uncertainty products.

A. Training/Testing Dataset

The primary dataset for testing the proposed uncertainty metric comprises colocated measurements of hyperspectral R_{rs} spectra and Chl*a*, which have been previously used for the training and testing of the MDN-based models. The primary radiometric quantity used in this study is the remote sensing reflectance (R_{rs} [sr^{-1}]), which is defined as the ratio of the water leaving radiance ($L_w(\lambda)$) to the downwelling irradiance just above the water ($E_d(\lambda)$) at each wavelength [52], i.e.,

$$R_{rs}(\lambda) = \frac{L_w(\lambda)}{E_d(\lambda)}. \tag{1}$$

The spectral dependency (λ) is discarded hereafter. This hyperspectral data is resampled according to the relative spectral response functions of the satellite instruments of interest i.e., OLCI, MSI, and OLI to gauge the performance of the MDN model for different spectral configurations. While the in situ data simulate near-ideal conditions in terms of measurement techniques and environments, these measurements are not void of uncertainties due to random/systematic noise in field instrument measurements, operation errors, nonideal environmental conditions, and inaccuracies in laboratory-based Chl*a* measurements. Nonetheless, it is noted that the uncertainty/randomness associated with the dataset is expected to be lower than the uncertainties associated with satellite-derived R_{rs} [37], [38]. The samples in this dataset have been collected from various inland and near-shore coastal waters, such as lakes, bays, estuaries, coast lines, and rivers from around the world, covering a wide range of trophic states (TSs) and geographic locations [37]. As such, the dataset contains samples with Chl*a* values ranging from 0.1 to 1000 mg m^{-3}, and the distribution of Chl*a* for the samples corresponding to the satellite instruments is shown in Fig. 1.

B. Multispectral Satellite Datasets

In addition to the in situ dataset, the proposed uncertainty metric is also tested on Chl*a* predictions made on selected satellite images. The most interesting aspect of these datasets is that, for each geographic location considered in this study (see Section V-B), we have near-simultaneous image acquisitions from all three sensors namely, OLCI, MSI, and OLI. These areas are two major estuaries (San Francisco

Bay and Chesapeake Bay) and one hypereutrophic lake (Upper Klamath Lake). The images are processed using the Atmospheric Correction for OLI "lite" (ACOLITE) [53]. The San Francisco Bay images were acquired on March 16, 2019, with OLCI and OLI measurements occurring within minutes of each other, while the MSI image was acquired 30 min later. The three Chesapeake Bay images were acquired by the three instruments in a 15-min window on November 7, 2016. The final location of interest is the Upper Klamath Lake on July 29, 2019, when Sentinel-3, Sentinel-2, and Landsat-8 imagery were acquired in a 20-min window. We also evaluated the temporal variability seen in the MDN predictions and uncertainties for a selected location in San Francisco Bay over the calendar year 2019.

III. METHODS

A. Mixture Density Networks

Classical ML algorithms attempt to solve a forward problem, wherein they predict/estimate the value of a specific variable (referred to in learning as the target/dependent variable) given the causal factors that produced them (referred to as independent variables). As such, these forward problems have a unique solution since a set of causal factors will always produce the same observable target value. With the addition of noise to the data and the model, we will have a target value (conditioned on the independent variable) drawn from a distribution with a single dominant value or mode (i.e., a unimodal distribution). The inverse problem of retrieving the inherent optical properties (IOPs) of the water column [37] or concentrations of water constituents (such as Chl*a*) from derived (or observed) quantities, such as R_{rs}, violates this assumption; thus, the mapping from the input space to output space (e.g., $R_{rs} \rightarrow$ Chl*a*) may be nonunique. In any scenario, wherein the mapping between input and the target variables is many-to-one (i.e., the problem has multiple solutions), the target distribution is multimodal.

MDNs [54] are a class of neural network algorithms that were designed to accommodate the solution of these inverse problems [36], [55]. The main difference is that this class of models attempts to predict the full target distribution in lieu of a point target value. Furthermore, to accommodate the possibility of multimodal solutions, MDNs model the target distribution by using a Gaussian mixture model (GMM) and predict the parameters of the GMM. A GMM can be simply defined as

$$p(y|\theta) = \sum_{j=1}^{k} \pi_j \, \mathcal{N}\big(y|\mu_j, \ \Sigma_j\big)$$

$$\text{s.t. } \pi_j > 0 \ \forall \ j; \ \sum_{j=1}^{k} \pi_j = 1 \tag{2}$$

where $\theta = \{\pi_j, \ \mu_j, \ \sigma_j\}_{j=1}^{K}$ are the parameters corresponding to the GMM, namely, the mixture probability (π_j), the mixture mean (μ_j), and the mixture variance (σ_j) corresponding to each Gaussian component. Once the MDN predicts the components of the GMM as described above, the final model estimate is extracted from the probability distribution
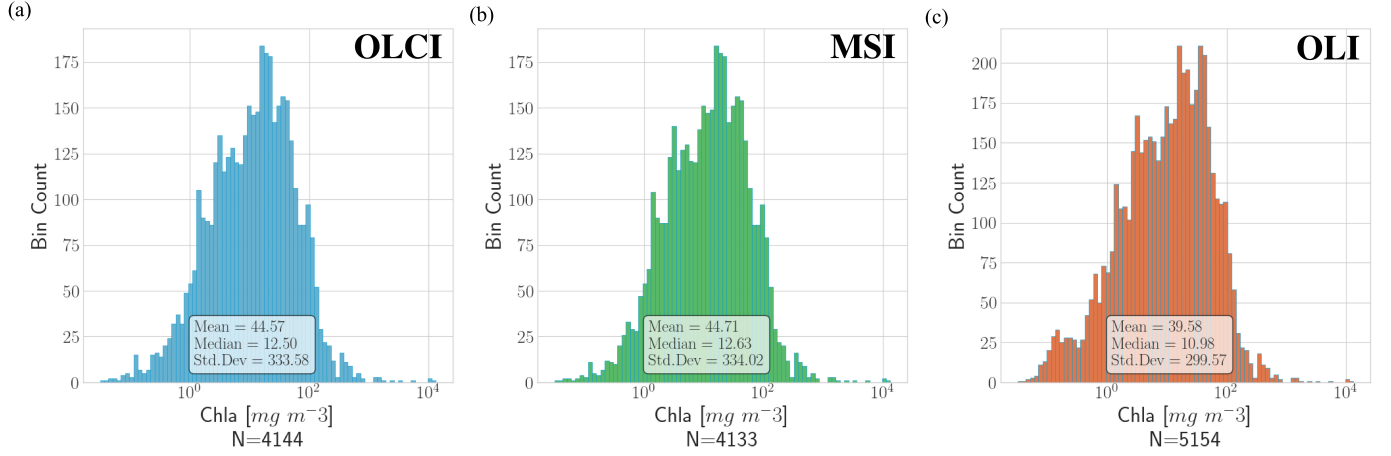
Fig. 1. Distribution and modes of Chl*a* for the in situ data for which colocated hyperspectral R_{rs} spectra are available for (a) OLCI, (b) MSI, and (c) OLI. The spectral range for each dataset varies, and so does the number of samples for each sensor.

described above. One method for extracting a point estimate from the MDN output is to take the mean of the dominant distribution in the MDN output, i.e., the mean of the Gaussian component, which represents the area of the highest probability mass. The overall scheme for extracting Chl*a* predictions from R_{rs} is shown in Fig. 2.

In this study, we use MDNs based on the architecture and the hyperparameters described in previous publications [37], [38]. Similar to these publications, we train a separate model for each of the sensor configurations under consideration. Each MDN model is designed to take R_{rs} as the input and generates a Chl*a* estimate. The model for each sensor has five hidden layers with 100 nodes in each layer (the number of nodes in the input layer is sensor specific), the number of components in the GMM is set to 5, and the learning rate and L2 normalization rate (a regularization term with model weights squared) of the network are set to 0.001. In addition, prior to the application of the data to the MDNs, both R_{rs} data and Chl*a* data are preprocessed using intraquartile range (IQR) scaling and log scaling, respectively. Finally, both variables are scaled using Minmax scaling (to be in the range $[-1, 1]$). The models were implemented using Python's Tensorflow [56] distribution and trained for 250 epochs with a batch size of 128. In the default configuration, all the available labeled data for the specific sensor configuration are used for training the MDN models (note that, unless explicitly stated otherwise, this default configuration is used in all experiments/applications throughout this article). The final estimate of Chl*a* is considered to be the mean of the component with the largest weight (postinversion of scaling mentioned above). The MDNs are trained to minimize the negative log-likelihood for the training samples. The prediction error, given the true value (y) and the predicted value ($\hat{y}$), is measured using the mean squared log error (MSLE) as

$$\text{MSLE} = \frac{1}{N} \sum_{i=1}^{N} (\log(y_i + 1) - \log(\hat{y}_i + 1))^2. \quad (3)$$

MSLE has been chosen as the measure for predictive error in this study; compared to classical metrics, MSLE is a relative measure, which is based on the ratio of the actual and predicted

values, and is also not overly affected by the magnitude of the error. This metric is also more robust to outliers compared to classical metrics.[2]

The one difference relative to prior publications is that, at this stage, we consider a single MDN model as opposed to ensembles—this is done to first verify that the uncertainty metric is reasonable for a single model before extending it to ensembles. For more information on the MDN, its training and performance relative to other Chl*a* estimators readers are referred to the prior publications, such as [37] and [38].

B. Uncertainty Estimation for MDNs

Since MDNs predict the probability distribution of the target variable using a GMM, the probability distribution can be leveraged to estimate the uncertainty [40]. If the output of the MDN for the ith sample y_i is a GMM, as shown in (2), the mean of the distribution can be estimated as

$$\mathbb{E}(y_i) = \sum_{j=1}^{k} \pi_j \mu_j \quad (4)$$

while the variance for this distribution can be estimated as

$$\begin{aligned} \mathbb{V} &= \mathbb{E}_{k \in \pi}(\mathbb{V}_k) + \mathbb{V}_{k \in \pi}(\mathbb{E}_k) \\ &= \sum_{j=1}^{k} \pi_j(x) \Sigma_j(x_j) \\ &+ \sum_{j=1}^{k} \pi_j(x) \left\| \mu_j(x) - \sum_{k=1}^{K} \pi_k(x)\mu_k(x) \right\|^2. \end{aligned} \quad (5)$$

In the above equation, $\mathbb{E}_{k \in \pi}(\mathbb{V}_k)$ corresponds to the average variance among the components of the mixture of Gaussian estimated by the MDN, while the second term $\mathbb{V}_{k \in \pi}(\mathbb{E}_k)$ is the contribution of the dispersion of the means in the predicted distribution. Furthermore, Choi et al. [40] point out that the first term can be interpreted as the random uncertainty present in the Gaussian components [or *aleatoric uncertainty* (σ_{ALT}^2)], while the second term corresponds to the multimodality

---

[2]While MSLE is the primary metric used in this article, Appendix B shows that the use of other error metrics reveals similar trends.
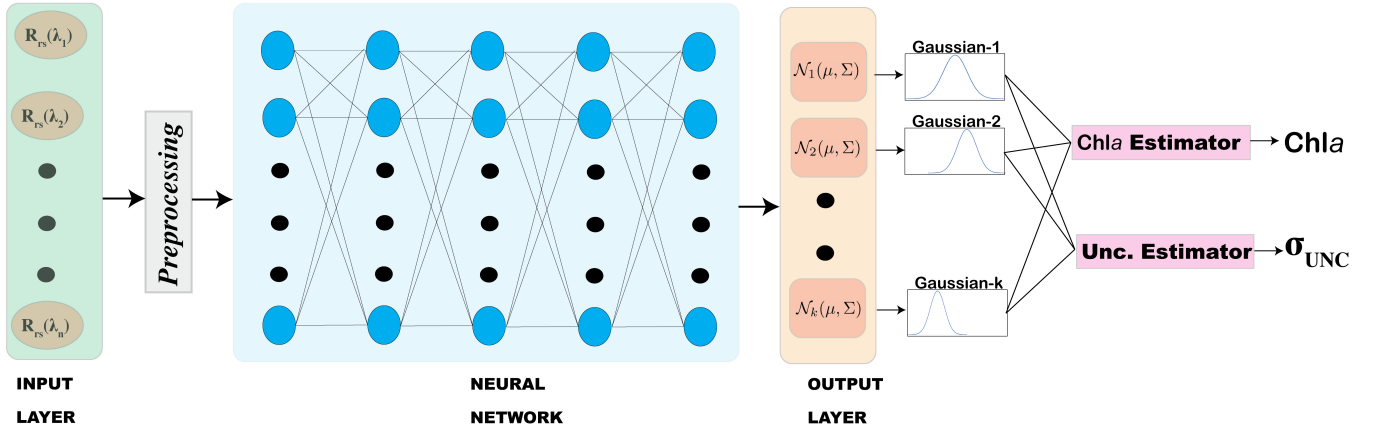
Fig. 2. Schematic representation of the main components involved in Chl*a* and σ_{UNC} estimation from R_{rs} data using MDNs.

present in the distribution, which would correspond to the fixed or knowledge-based uncertainty [or *epistemic uncertainty* (σ_{EPS}^2)].

Based on this, we can rewrite (5) as

$$
\mathbb{V} = \sigma_{ALT}^2 + \sigma_{EPS}^2
$$

$$
= \underbrace{\sum_{j=1}^{k} \pi_j(x) \Sigma_j(x_j)}_{\text{ALT}}
$$

$$
+ \underbrace{\sum_{j=1}^{k} \pi_j(x) \left\| \mu_j(x) - \sum_{k=1}^{K} \pi_k(x) \mu_k(x) \right\|^2}_{\text{EPS}}. \quad (6)
$$

The total uncertainty associated with the MDN prediction is the sum of the aleatoric and epistemic uncertainty

$$
\sigma_{\text{UNC}} = \sigma_{\text{ALT}} + \sigma_{\text{EPS}}. \quad (7)
$$

While the estimated uncertainty metric gives a user a measure of confidence in the models' prediction, this uncertainty metric is slightly different than the uncertainty metric associated with the physical measurements as the uncertainty does not depend on the properties of the measuring devices; as such, it is hard to assign a physical interpretation (for example, in terms of the units of the measured quantity) to the uncertainty. Rather, this notion of uncertainty depends more on the properties of the algorithm, such as hyperparameters of the algorithm, the algorithm complexity, and the dataset (e.g., the training data distribution). Therefore, this metric is only used as a measure of model-specific uncertainty rather than physical uncertainty.

To illustrate the ability of the MDNs in capturing the uncertainty associated with their predictions, we show the distributions estimated by the MDN for two in situ samples representing two extreme scenarios for the uncertainty estimation task. The first spectral sample [see Fig. 3(a)] shows an OLCI-like R_{rs} spectrum representing oligotrophic waters, for which the model has high confidence or low uncertainty in its prediction of Chl*a*. Note that all the estimated Gaussian components are quite sharp (i.e., have low variance). Furthermore, the means are very close, and the combined distribution will be close to unimodal. On the other
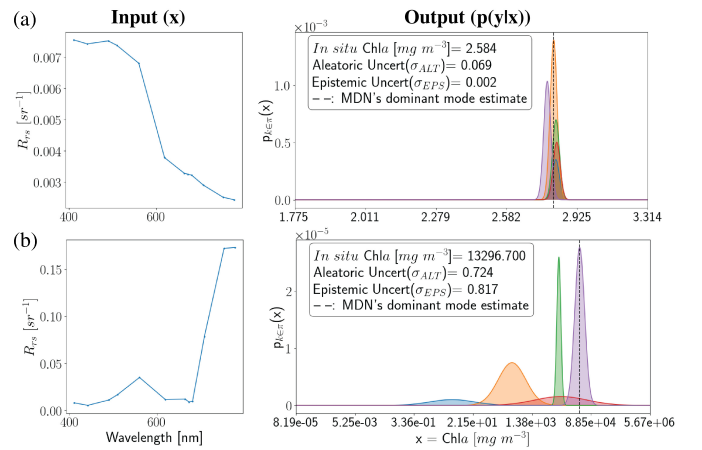

Fig. 3. Illustration of the differences in the distribution estimated by the MDN for (a) R_{rs} representing clear waters with low uncertainty and (b) R_{rs} sample with high uncertainty. The samples tested correspond to OLCI-like R_{rs}.

hand, for the high uncertainty/low confidence sample [see Fig. 3(b)], which, from the elevated NIR reflectance, appears to represent surface scums, the estimated MDN distribution is highly multimodal, i.e., the combined distribution will have many well-separated peaks. Also, the Gaussian distributions are significantly more spread [especially, the blue and red components in Fig. 3(b)]. The high uncertainty seen for this sample may be caused by various reasons, the presence of additional noise/distortion in this specific sample (for example, due to human error), or the model may be unable to confidently fit the specific data-sample due to the lack of sufficient similar data-samples in the training set. The main takeaway is that model lacks sufficient confidence to make a specific guess for this sample and rather predicts it to be something in a very large range.

C. Design of Experiments to Validate the Uncertainty Metric

This section describes the set of experiments designed to ensure/verify that the MDN-specific uncertainty metric is valid for Chl*a* predictions. It is essential to demonstrate that the σ_{UNC} metric accurately captures the uncertainty induced by the common distortion process seen in satellite data, such as

noise, atmospheric effects, and novel test data. The various experiments in this study are described in the following.

1) Effect of Noise in the Test Dataset: The simplest disruption scenario is when the data are affected by random (Gaussian) noise. This experiment will provide key insights into how the model performs on test data that are like the training data except for the presence of noise. This is to verify that the uncertainty associated with predictions on noisy data is higher than the uncertainty associated with predictions on data with little to no noise. For this test, we used the trained default models described in Section III-A for each sensor. Furthermore, to reduce uncertainty caused by the presence of new/previously unseen samples, we simply test σ_{UNC} associated with predictions of the training samples themselves, with additional noise injected into the samples incrementally. In each trial, increasing amounts of (Gaussian) white noise (typified by noise drawn from a normal distribution with higher variance) was added to all the R_{rs} samples in the training data. Then, predictions are made using the pretrained sensor-specific MDNs on the "noisy" data. Finally, the aleatoric and epistemic uncertainties associated with each prediction are estimated. Since tracking the uncertainty or retrieval errors on individual samples would only provide anecdotal evidence, which has no statistical basis, we choose the mean of the uncertainty metric $\mathbb{E}(\sigma_{\text{UNC}})$ and error [see (3)] over all the "noisy" samples.

2) Effect of Unseen/Novel Data on Uncertainty Estimation: Since ML algorithms attempt to leverage labeled samples to learn the relationship between the input data and the target variables, the performance of such models is highly dependent on the training data to which these models are exposed. Therefore, the model's performance should be "uncertain" when exposed to novel data that are significantly different from the training data. In our interpretation, a test sample can be considered significantly different if it lies outside the convex hull of the training data and, therefore, appears to not share support with the training data. To design a test for this scenario, we calculated the first few principal components (PCs) of the in situ dataset (see Section II-A) [57], [58] to be able to visualize the data in 2-D space. To eliminate extreme effects, outliers that are very far away from the bulk of the dataset are eliminated. These labeled datasets are then split into two portions based on their position in the PC plot. One dataset referred to as the training dataset ($N_{\text{tr}} = 2456$) is shown in blue on the left column plots of Fig. 6. The other dataset is the test dataset ($N_{\text{te}} = 992$ samples) shown in orange on the left column plots of Fig. 6. We trained the MDN models for each sensor according to the procedure described in Section III-A, except only that the samples in the training dataset (as shown on the left-hand side of Fig. 6) were used for training. Finally, prediction errors (MSLE) and uncertainty (σ_{UNC}) based on the MDN prediction were estimated for samples in both the training and test datasets for each sensor.

3) Effect of Structured Distortions/Artifacts: Another well-known disruption factor in aquatic remote sensing is the presence of systematic noise from uncertainties in the atmospheric correction or instrument miscalibration, the former of which is more pronounced for modern satellite sensors [59]. These distortions are caused by inaccurate
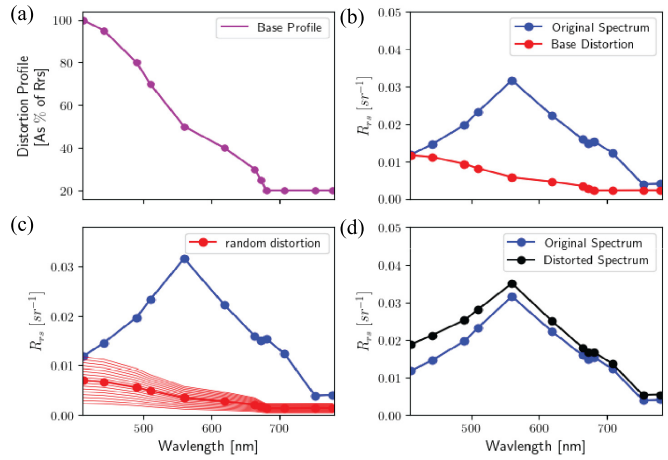

Fig. 4. (a) Base distortion profile for an OLCI spectrum. (b) Estimated-based distortion for a sample OLCI-like R_{rs}. (c) Effect of the randomizing with a random multiplier. (d) Comparison of the original and distorted spectra.

compensation for aerosol and sky-glint contributions [60]. In general, satellite-derived R_{rs} products over freshwater and coastal estuaries are known to carry larger distortions in the blue bands that decay toward the red and NIR bands [53], [60], [61], [62], [63], [64], [65]. To simulate these effects, we assumed a base distortion shape representing positively correlated systematic noise shown for OLCI in Fig. 4(a) where the distortion in the shorter wavelengths is higher and diminishes toward the longer wavelengths (the same shape was subsampled to the spectral resolution of the other sensors). The value at each spectral band indicates the percentage of the true R_{rs} that is added back to the specific spectral band distortion. Fig. 4(b) shows the base distortion for a specific OLCI-like R_{rs}. To accommodate the randomness inherent to these processes, the base distortion is further multiplied with a random multiplier drawn from a uniform distribution over the interval $[\alpha, 1]$ of width $1 - \alpha$ where $0 \leq \alpha < 1$. The multiplier affects the magnitude of distortion. Choosing a small value of α would mean that the multiplier for the distortion is drawn from a large range, where only a few samples suffer from a high distortion (i.e., have a multiplier close to 1). On the other hand, choosing a large α would mean that the multiplier is drawn from a narrow range of large values; therefore, all samples will have a distortion multiplier close to 1. Fig. 4(c) shows some of the possible distortions for $\alpha = 0.2$. From the possible distortions, one is chosen at random, such a random selection is shown in bold Fig. 4(c), and added to the original R_{rs}. This specific distortion adds about 0.0066 [sr^{-1}] to R_{rs} (412 nm) and decays down to 0.0019 [sr^{-1}] at 778 nm. Fig. 4(d) shows the comparison between the original (blue) and the distorted R_{rs} (black) spectrum. For the other sensors, the base distortion shape shown in Fig. 4(a) is sampled at the spectral resolution of the specific sensor and added using the same process as the one described above for OLCI. The performance of the default (described in Section III-A) MDN models in terms of uncertainty estimation is tested across a variety of values of α for each sensor configuration.

4) Verifying Correlation Between Predictive Error and Uncertainty: While the experiments described above

investigate the effect of noise/distortions in the data on the uncertainty involved in the MDN predictions, it is also important to investigate the relationship (if any) between the error and uncertainty of the MDN Chl*a* predictions. To investigate this, we first split the labeled samples (using their measured Chl*a*) into various bins (of 1 mg m^{-3} width). The number of samples in each bin is shown in Fig. 8(a). Since both σ_{UNC} and MSLE appear to be drawn from random processes, an averaging scheme is applied to clarify trends: therefore, in this experiment, we focus on the averages for the samples in a bin with a narrow range of Chl*a* (i.e., 1 mg m^{-3}). In particular, we focus on the 50 consecutive bins with Chl*a* ranging from 0 to 50 mg m^{-3}, as each of these bins has an adequate number of labeled samples, which can be used for analysis. The samples/bins used in this experiment are highlighted in red in Fig. 8(a). In this experiment too, the default MDNs (described in Section III-A) are used for analysis.

5) Estimated Uncertainty as Bounds on Predictive Error: In general, the uncertainty metric encapsulating the ML models' confidence in its prediction does not appear to have a physical basis but rather appears to be based on factors such as training data distribution and specific model properties. This experiment attempts to investigate if there is any reason to believe that the estimated uncertainty metric also has some physical relevance and if it can be used as an error bound on the predictions (i.e., we will check if the true value is consistently a member of the set of the interval centered on the predicted value and bounded by the estimated uncertainty). For an uncertainty analysis, a prediction $\hat{y}$ is said to have an uncertainty σ_{UNC}, with a coverage factor of k if, for the true value of for that sample (referred to here as y), we can say that the following condition is applicable to a specific percentage of the samples [42]: $\hat{y} - k\sigma_{\text{UNC}} \leq y \leq \hat{y} + k\sigma_{\text{UNC}}$ [N.B.: in our experiments, these tests were performed on the scaled Chl*a* predictions]. The confidence level associated with a model at a specific coverage factor can be measured as the percentage of samples for which the above condition holds true. In this experiment, we consider the confidence level associated with the default MDNs (as described in Section III-A) predictions for different values of $k \in \{0.5, 1, 2\}$ for the three OLCI, MSI, and OLI band settings.

6) Investigating the Effects of "Noisy" Training Data on the Uncertainties Estimated by MDNs: This experiment attempts to verify that the effect of the noise in the training data on the uncertainty estimated by the MDN models. It is reasonable to expect the uncertainty for models trained on noisy data to be higher. We created two suites of MDN models trained on slightly different datasets to test this hypothesis. First, the samples from the labeled dataset described in Section III-A were divided using a 70:30 split into a training set and a test set. Note that, for brevity, we only run this experiment for OLCI-simulated R_{rs}. Second, a suite of nine models was trained on the training dataset mentioned above; this suite is referred to as the "orig_train" models. Finally, another suite of nine models was trained on a noisy version of the training data (adding white noise drawn from a Gaussian distribution ($\sigma = 2.5\ e^{-3}$). This suite is referred to as the "noisy_train."
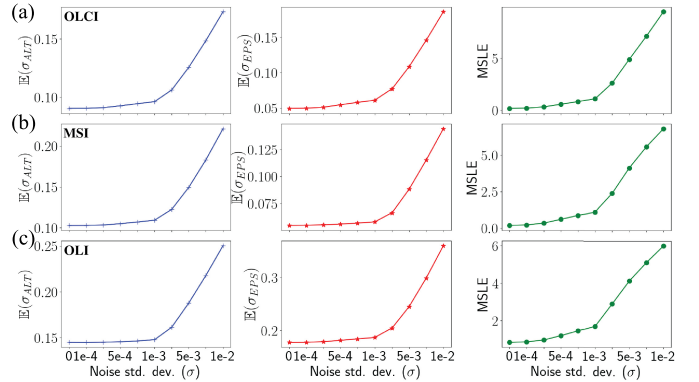

Fig. 5. Comparison of $\mathbb{E}(\sigma_{\text{ALT}})$ and $\mathbb{E}(\sigma_{\text{EPS}})$ to the MSLE for MDNs trained on (a) OLCI-like, (b) MSI-like, and (c) OLI-like spectra at different noise levels.
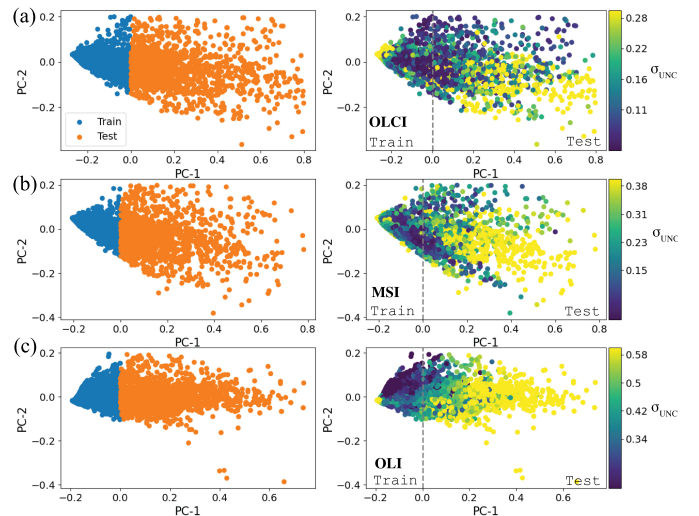

Fig. 6. Effect of novel data on the estimated total uncertainty. The left-hand side plot shows the PC visualization of the data for the specific sensor (e.g., OLCI in the top row); training samples are in blue, and test samples are shown in orange. The right-hand side scatter plot is colored according to the estimated σ_{UNC} as shown in the color bar for the three spectral band configurations (a) OLCI, (b) MSI, and (c) OLI. (To improve the visualization in the PC-space, we eliminate data points that are very far away from the bulk of the data cloud. This outlier elimination is only done to improve visualization, which reduced the number of labeled samples to 3448. Note that the visualizations from the three sensors are very similar because they are essentially approximations of the same optical dataset.)

Both sets of models were tested in terms of prediction error and uncertainty on the common test dataset.

IV. RESULTS

This section contains the results for the various validation experiments described in Section III-C.

A. Experiment-I: Effect of Noise in the Test Dataset

The comparison of $\mathbb{E}(\sigma_{\text{ALT}})$ and $\mathbb{E}(\sigma_{\text{EPS}})$ to MSLE for the various spectral band settings (corresponding to the three satellite sensors) is shown in Fig. 5. Note that there appears to be a clear correlation between $\mathbb{E}(\sigma_{\text{ALT}})$, $\mathbb{E}(\sigma_{\text{EPS}})$, and MSLE for the MDN models of all three sensors. They increase linearly up to a point (in this case, up to a Gaussian noise with $\sigma = 0.001$), beyond which both exponentially increase. It should be noted that, while the trends for error and uncertainty are similar, the scales and rates of change are

TABLE I
COMMON ABBREVIATIONS AND NOTATIONS USED

Abbreviation/Notation	Units [if any]	Explanation
$Chla$	$mg\ m^{-3}$	Chlorophyll-a
E_d	$W\ m^{-2}\ \mu m^{-1}$	Downwelling irradiance
$\mathbb{E}(x)$		Expected value of a random variable x
GMM		Gaussian Mixture Model
IOP		Inherent Optical Properties
IQR		Intra Quartile Range
L_w	$W\ m^{-2}\ \mu m^{-1}\ sr^{-1}$	Water Leaving Radiance
MDN		Mixture Density Network
ML		Machine Learning
MSI		MultiSpectral Instrument
$MSLE$		Mean Squared Log Error (predictive error measure)
$\mathcal{N}(\mu, \Sigma)$		Gaussian dist. with mean μ and variance Σ
$OLCI$		Ocean and Land Colour Instrument
OLI		Operational Land Imager
$R_{rs}(\lambda)$	sr^{-1}	Spectral remote sensing reflectance
σ_{ALT}		MDN-estimated aleatoric uncertainty
σ_{EPS}		MDN-estimated epistemic uncertainty
σ_{UNC}		Total MDN-estimated uncertainty
$\mathbb{V}(x)$		Variance of a Random variable x

TABLE II
MEAN ESTIMATED UNCERTAINTIES FOR KNOWN AND NOVEL
R_{RS} SPECTRA RESAMPLED TO VARIOUS SENSORS

Sensor	MSLE		σ_{UNC}	
	Train	Test	Train	Test
OLCI	0.11	0.70	0.14	0.21
MSI	0.135	0.48	0.172	0.323
OLI	0.61	1.81	0.35	0.54

not the same, indicating that $\mathbb{E}(\sigma_{UNC})$ is not a substitute for the error. Based on this experiment, we infer that, when there is noise present in the test dataset, the uncertainty metric is able to accurately respond to the presence of such a distortion, and furthermore, the mean uncertainty appears to be correlated with the mean predictive error.

B. Experiment-II: Effect of Unseen/Novel Data

The left-hand side column of Fig. 6 shows a scatterplot visualization of the training and test datasets. The right-hand side plot in Fig. 6 shows the same PC visualization of the data but with each sample colored according to the value of σ_{UNC} associated with the MDN predictions. Based on these scatterplots, it is clear that σ_{UNC} for samples in the training set is significantly lower than that for the samples in the test set. The metrics reported in Table II point to the MSLE and mean uncertainty being significantly higher on the novel samples in the test set for all three spectral configurations. Another interesting observation is that the uncertainty is lower for novel samples closer to the training set, while, toward the edges (which are farther away), there are higher uncertainties, indicating that as the difference of a sample from the training dataset increases, so does the uncertainty.

C. Experiment-III: Effect of Structured Distortions/Artifacts

Fig. 7 shows the effect of changing the value of α on the mean predictive error (MSLE) and uncertainty ($\mathbb{E}(\sigma_{UNC})$)
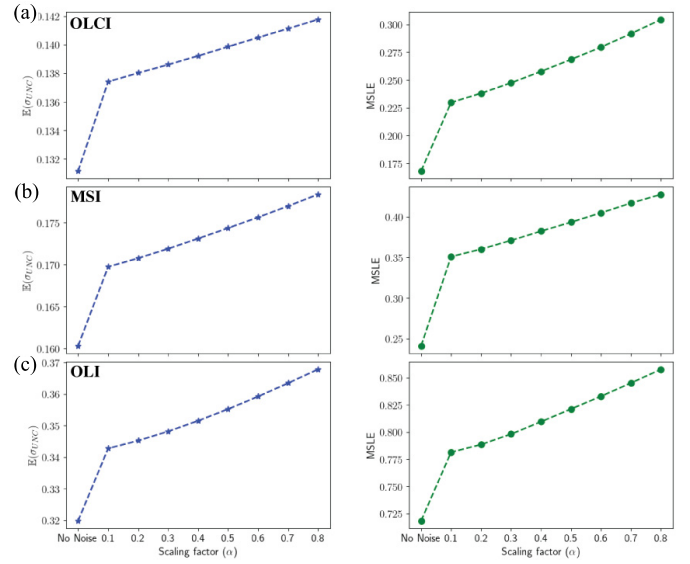

Fig. 7. Effect of the systematic noise due to spectrally dependent uncertainties from atmospheric correction (of the type shown in Fig. 4) in terms $\mathbb{E}(\sigma_{UNC})$ and MSLE for the different band settings (a) OLCI, (b) MSI, and (c) OLI.
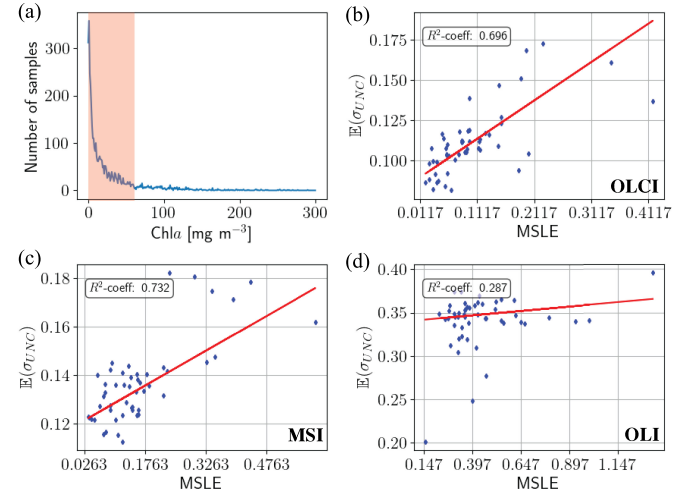

Fig. 8. Correlations between uncertainty estimate ($\mathbb{E}(\sigma_{UNC})$) and mean predictive errors denoted by MSLE. (a) Samples in the in situ database are partitioned into bins of width 1 [mg m^{-3}]. Plots correspond to the relationship between MSLE and $\mathbb{E}(\sigma_{UNC})$ in the selected bins for models trained for (b) OLCI, (c) MSI, and (d) OLI spectral resolutions.
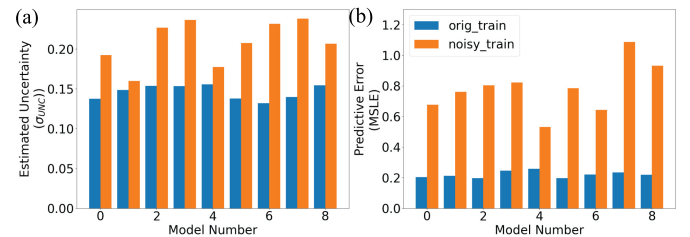

Fig. 9. (a) Mean estimated uncertainty and (b) mean predictive error for the models trained on training sets with different noise levels.

for all the three sensors. Note that, for all three sensors, the presence of some noise or distortion leads to an increase in both the error and uncertainty. In this case too, the mean error and uncertainty appear to be correlated for all three sensors. Furthermore, the magnitude of error and uncertainty
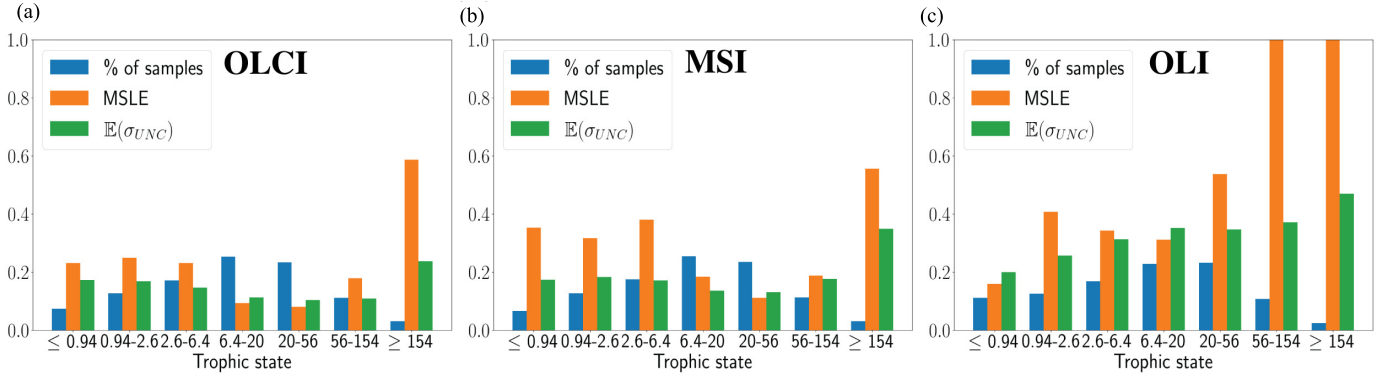
Fig. 10. Comparing the performance of the MDN models in terms of MSLE and $\mathbb{E}(\sigma_{\text{UNC}})$ based on TSs for (a) OLCI, (b) MSI, and (c) OLI.

TABLE III
CONFIDENCE ASSOCIATED WITH THE ESTIMATED UNCERTAINTIES
FOR DIFFERENT COVERAGE FACTORS

Sensor Type	$k = 0.5$	$k = 1.0$	$k = 2.0$
OLCI	92.16	95.89	97.97
MSI	89.31	94.94	97.74
OLI	81.29	90.01	96.89

also seems dependent on the multiplier α (as mentioned earlier, a larger α corresponds to generally increased distortion). Similarly, the uncertainty metric accurately captures the presence of confounding factors, such as uncertainties in the atmospheric correction with an increased uncertainty.

D. Experiment-IV: Verifying Correlation Between Uncertainty and Predictive Errors

In this experiment, we investigate the relationship between MSLE and $\mathbb{E}(\sigma_{\text{UNC}})$ within narrow bins, as shown in Fig. 8(b)–(d). Each point in these subplots represents the value of MSLE and $\mathbb{E}(\sigma_{\text{UNC}})$ for a single bin. Note that there is almost a linear relationship between the two quantities, with the line of best fit between the two terms shown in red in each subfigure. While this observation appears to be generally true, it should be noted that there are specific samples that show high $\mathbb{E}(\sigma_{\text{UNC}})$ for low MSLE, while a few samples also indicate low MSLE for high $\mathbb{E}(\sigma_{\text{UNC}})$. This indicates that, while there is a general correlation between the two terms, they are not direct analogs of each other. The figures also show the correlation between σ_{UNC} and MSLE in terms of the R^2-correlation coefficient.

E. Experiment-V: Estimated Uncertainty as Bounds on Predictive Error

The confidence levels associated with different values of k for all three sensors are shown in Table III. For both MSI and OLCI, for about ~90% of the samples, the uncertainty provides a reasonable bound on the error even for a coverage factor of $k = 0.5$. Based on the sensor and the application, an appropriate coverage factor can be chosen for each application (unless explicitly mentioned, the coverage factor chosen in this article is always $k = 1$). While these experiments indicate that the uncertainty metric has some
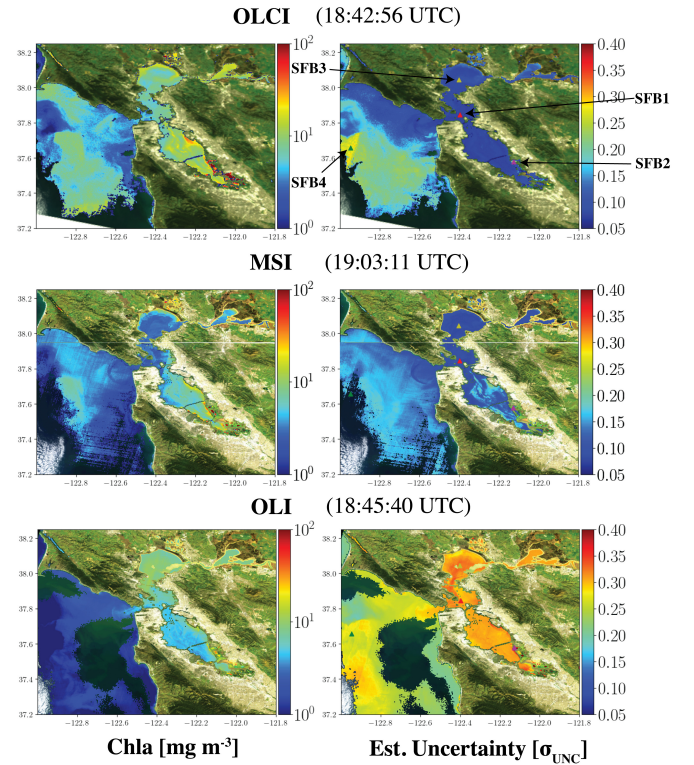

Fig. 11. Estimated Chl*a* [mg m^{-3}] and σ_{UNC} maps produced via MDNs for near-concurrent images of OLCI, MSI, and OLI onboard Sentinel-3B, Sentinel-2B, and Landsat-8, respectively, acquired on March 16, 2019. The images were processed to R_{rs} using ACOLITE. Also marked in the OLCI σ_{UNC} map are the stations "SFB1" [122.396° W, 37.846° N], "SFB2" [122.399° W, 37.847° N], "SFB3" [122.403° W, 38.043° N], and "SFB4" [122.944° W, 37.658° N] that are used for further analyses.

relations to the physical uncertainty metrics estimated for other measurement instruments, it is still not completely clear if this will be true for all test samples. To enable the use of these metrics as a physical uncertainty, we need to identify a method to estimate the confidence level with satellite data.

F. Experiment-VI: Investigating the Effects of "Noisy" Training Data on the Uncertainties Estimated by the MDNs

Both suites of models are tested on the (noiseless) test set in terms of the total estimated uncertainty and the prediction error, as shown in Fig. 9. From this plot, the uncertainty
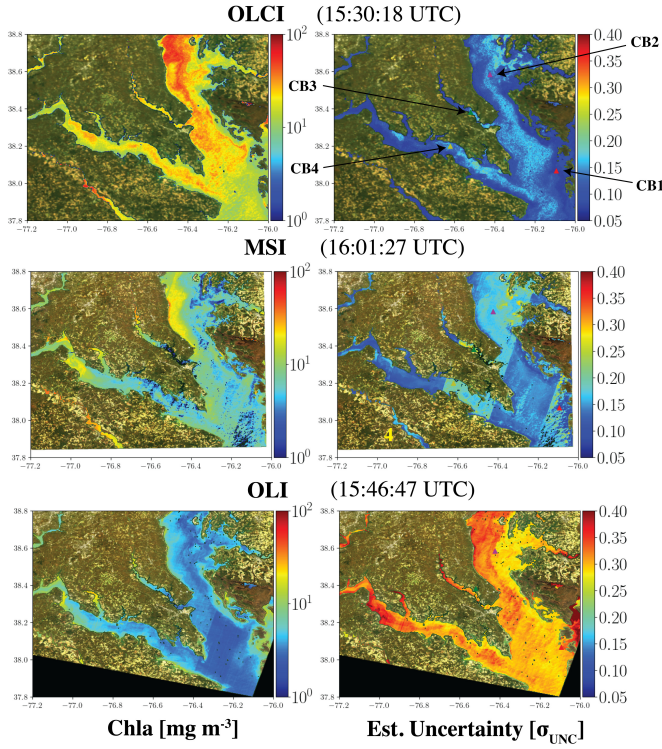
Fig. 12. Similar to Fig. 11 but over the Chesapeake Bay acquired on November 7, 2016. Also marked in the OLCI σ_{UNC} map are the stations "CB1" [76.095° W, 38.067° N], "CB2" [76.423° W, 38.582° N], "CB3" [76.620° W, 38.201° N], and "CB4" [76.522° W, 38.382° N] that are used for further analyses.

Fig. 13. Similar to Fig. 11 but over the Upper Klamath Lake on July 29, 2019. Also marked in the OLCI σ_{UNC} map are the stations "UKL1" [121.972° W, 42.419° N], "UKL2" [121.848° W, 42.372° N], "UKL3" [122.007° W, 42.452° N], and "UKL4" [121.973° W, 42.527° N] that are used for further analyses.

for models trained on noisy data is significantly higher than those trained on low-noise data. This clearly highlights the relationship between the uncertainty estimated by the MDN and the training data. This behavior is also aligns with our expectations that the predictive error and uncertainty increase when the models are trained on noisy data. For all these applications in this section, we use the default MDN models (as described in Section III-A).

V. PRACTICAL APPLICATIONS

Here, we describe practical applications of MDN-based uncertainties associated with Chl*a* prediction. The first practical application is the identification of the range of Chl*a* for which the MDN technique can be applied with high confidence. Other applications include evaluating MDN Chl*a* products and the spatial distribution of uncertainties from near-concurrent images of Sentinel-3/OLCI, Sentinel-2/MSI, and Landsat-8/OLI and analyzing time-series of Chl*a* and their uncertainties for select locations.

A. Application 1: Range-Specific Uncertainty Estimates

We aim at identifying the factors that are responsible for the behavior of the MDN models and specifying appropriate steps/strategies for improving the model performance. To perform this analysis, the entire in situ dataset of available labeled samples was split into seven nonoverlapping TSs [66] based on their Chl*a* value. $\mathbb{E}(\sigma_{\text{UNC}})$ and MSLE were then estimated for samples corresponding to each TS. The details of
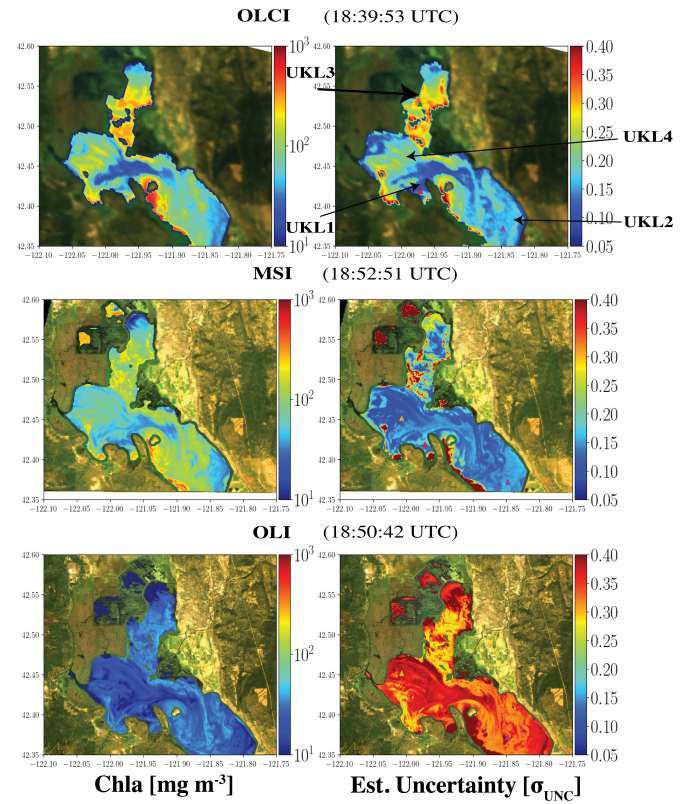
the various trophic ranges and the performance of the various MDN models for different sensors across all these TSs are shown in Fig. 10. This figure indicates that, for OLCI and MSI band configurations, the trends for $\mathbb{E}(\sigma_{\text{UNC}})$ and MSLE are inversely proportional to the number of samples available in a specific range. Based on this analysis it is clear that, for the OLCI- and MSI-like spectra, the models perform best in terms of both $\mathbb{E}(\sigma_{\text{UNC}})$ and MSLE for a Chl*a* range of 6.4–56 [mg m^{-3}]. This is also the range in which most training samples are available. For the OLI-like spectra, on the other hand, the model performs best in the range ≤ 1 [mg m^{-3}].

B. Application 2: Uncertainty Maps and the Spatial Context

The MDN-derived Chl*a* and the associated uncertainty maps from satellite observations (see Section II-B) are shown in Figs. 11–13. The left-hand side columns show the predicted Chl*a* values overlaid on RGB composites, while the right-hand side columns show the same RGB composites overlaid with the σ_{UNC}. The rows correspond to the three sensors under consideration. In addition, on the σ_{UNC} maps, we have marked the locations of four select stations considered for further analysis in the following sections. These locations, namely, "SFB1-4," "CB1-4," and "UKL1-4," are chosen to represent a broad spectrum of TSs in each region. The estimated Chl*a* and σ_{UNC} maps of the of San Francisco Bay are shown in Fig. 11. Similarly, the Chl*a* and σ_{UNC} maps for Chesapeake
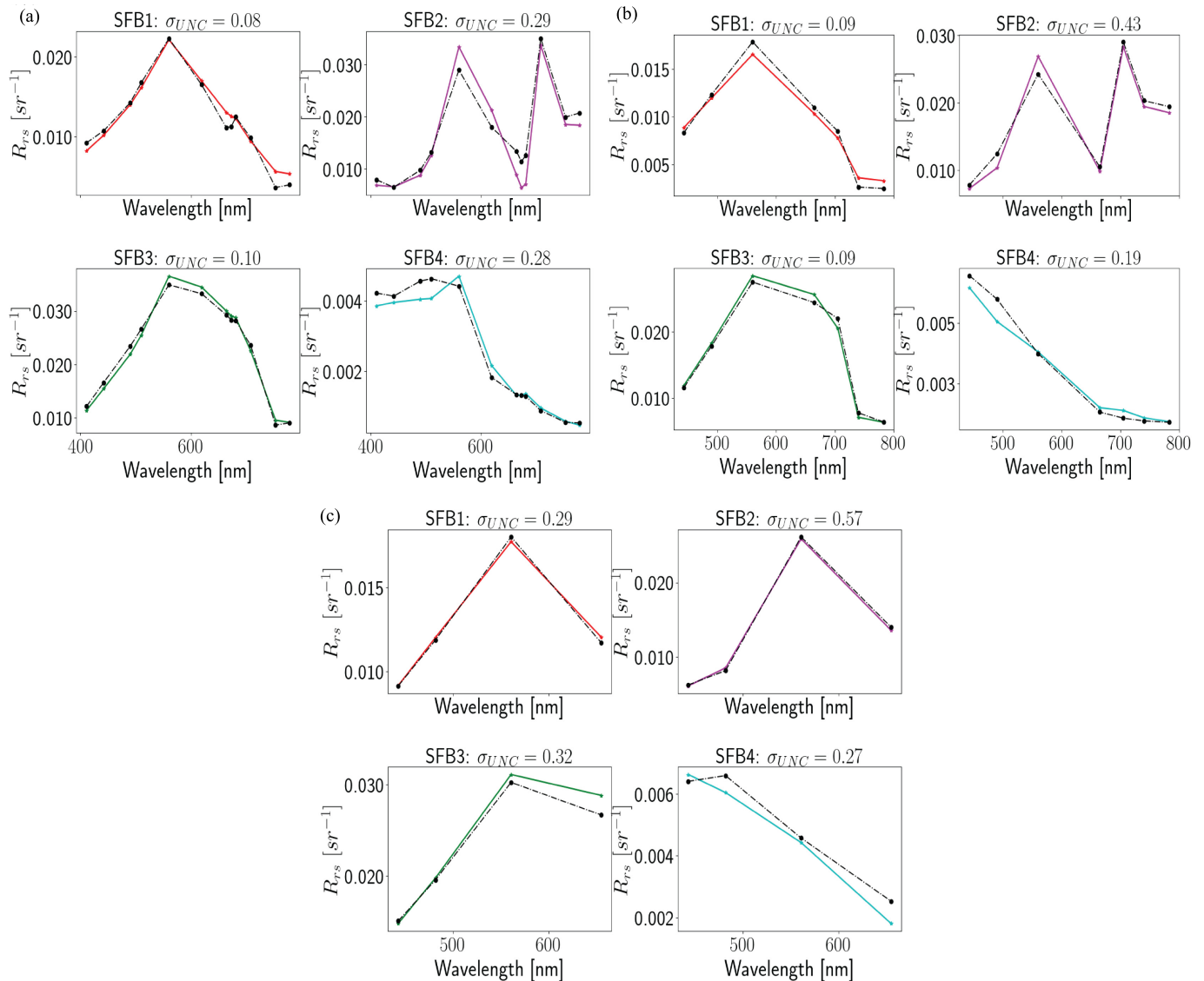
Fig. 14. Comparison of satellite-derived R_{rs} (solid lines) from selected locations marked on maps across San Francisco Bay and the nearby ocean together with their closest counterparts in our in situ database (dashed lines) for (a) OLCI, (b) MSI, and (c) OLI.

Bay and Upper Klamath Lake are shown in Figs. 12 and 13, respectively.

There are some interesting observations based on these satellite maps. The Chl*a* maps in Fig. 11 show that estimates from all the three sensors range within 5–10 mg m^{-3}, with both the OLCI and MSI maps, mapping higher Chl*a* values in the southern part of the bay than in the northern part. While the OLI map shows similar predictions in terms of the magnitude of Chl*a*, it seems to show the lower Chl*a* in the southern bay. The OLCI/MSI maps also show certain localized regions with very high Chl*a* (around 100 [mg m^{-3}]) as shown for regions near the "SFB-2" station. It is harder to appreciate such localized high Chl*a* in the OLI map. In terms of the uncertainty maps, there is significant agreement between the OLCI and MSI maps, with most of the pixels inside the San Francisco Bay and part of the coastal ocean exhibiting low uncertainty. The uncertainty increases farther away from the coast, which is reasonable as the training data contains smaller samples size over the open ocean (see Fig. 1). The OLI uncertainty

maps show significantly higher uncertainty; in addition, the OLI uncertainty maps do not show the same pattern as the OLCI and MSI maps. Furthermore, regions/pixels with comparatively higher estimations of Chl*a* are also in general more uncertain; it is expected that this behavior is because the training dataset has far more samples with Chl*a* in the 0–10-mg m^{-3} range as opposed to the 10–1000-mg m^{-3} range.

The MDN-estimated Chl*a* maps for the Chesapeake Bay (as shown in Fig. 12) are in general agreement in terms of the global distribution, such as comparatively higher Chl*a* in the northern part of the bay (e.g., near/around station "CB-2"). There are, however, some disagreements in the exact value of Chl*a* as predicted for the three sensors. In terms of uncertainty, both OLCI and MSI, in general, expect low to intermediate uncertainties, while the OLI indicates a higher uncertainty. Similarly, for the Upper Klamath Lake (see Fig. 13), both the OLCI and MSI maps show Chl*a* around 100 mg m^{-3} for most pixels, while the OLI maps show a far lower value. Overall,
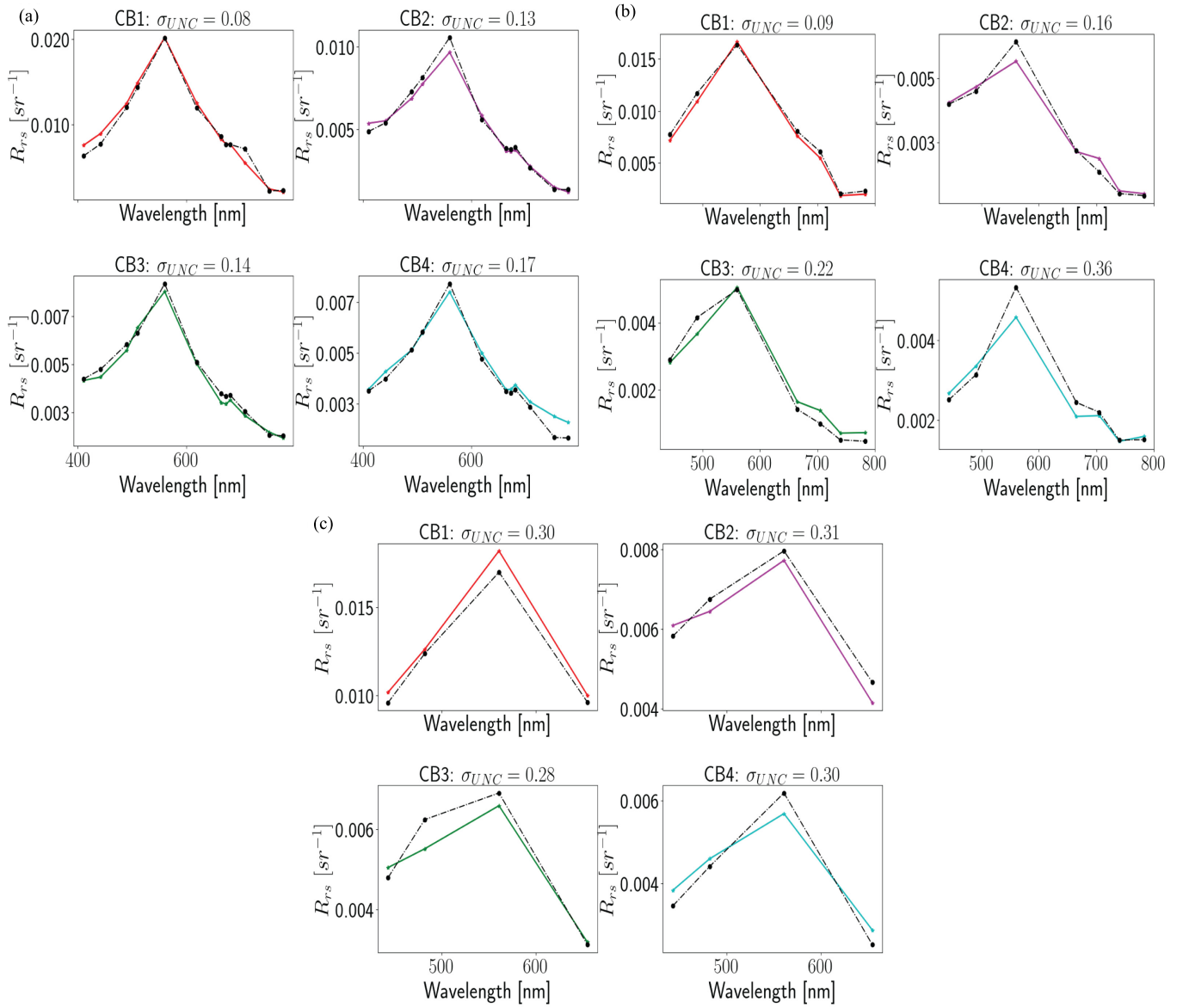
Fig. 15. Same as Fig. 14 but for the Chesapeake Bay.

the OLCI and MSI uncertainty maps also show similar trends, while the OLI shows much larger uncertainties.

1) Interpretation of Estimated Uncertainties: To understand the reasons behind the estimated uncertainties, we performed a basic spectral comparison between the satellite-derived spectra and the spectra in the labeled in situ database used for training for the arbitrarily selected locations, which have been marked in Figs. 11–13. The stations were chosen to consider spectra from regions with varying levels of uncertainty in these maps. For each station, we estimated the average spectra corresponding to a 300×300 m spatial patch and compared them to the spectra in the labeled in situ database (see Section II-A). For R_{rs} corresponding to each station, we identified the spectrum in the in situ database that is closest based on the Euclidean distance, computed using all the available bands (see Table IV) for each sensor. The spectrum for each station is then compared to its closest in situ analog (see Figs. 14–16). In these figures, each plot window shows

two spectra, namely, the spectrum corresponding to a specific station from the specific sensor (the colored solid line) and its closest analog from the appropriate in situ database (the black dashed lines).

Consider, for example, the spectral comparison for the stations in San Francisco Bay from the OLCI sensor shown in Fig. 14(a). "SFB1" and "SFB3" correspond to examples from inside the Bay and, in general, are from regions with low σ_{UNC}, as shown in Fig. 11. Consequently, the spectra from these locations also are very similar to their closest counterparts in the in situ database. R_{rs} associated with "SFB2," which is from highly eutrophic near-shore waters of the southern section of the bay, has a very high Chl$a$ and σ_{UNC}, and is quite different from its closest in situ counterpart in many spectral bands. "SFB4" situated over the open ocean close to a cloud edge also shows more pronounced differences between the two spectra. Similarly, Figs. 15 and 16 show the spectral comparison for stations in Chesapeake Bay and Upper Klamath Lake.
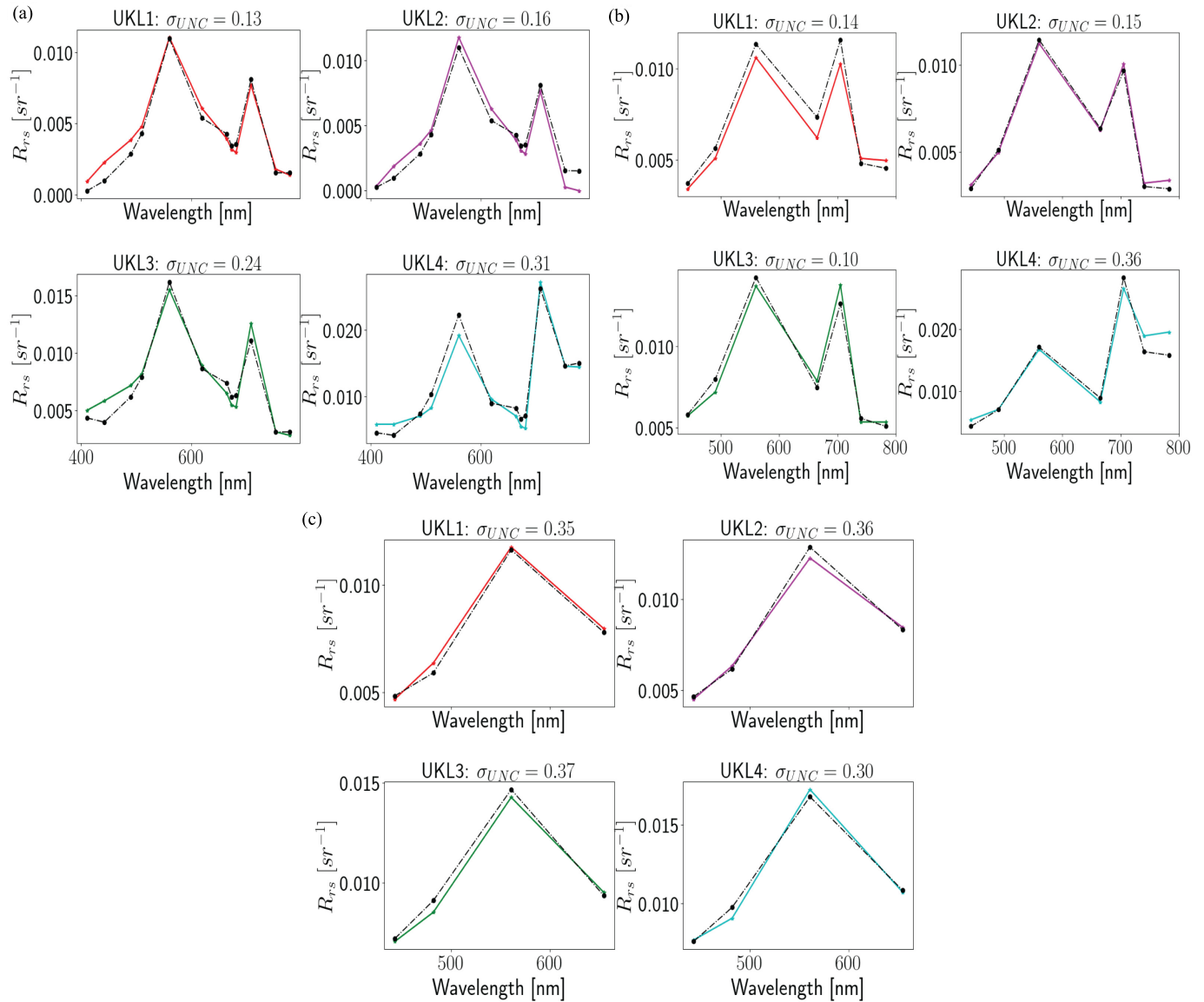
Fig. 16. Same as Fig. 14 but for the Upper Klamath Lake.

C. Application 3: Uncertainty Maps and Temporal Variability

For the final application, as an example, we tracked the variations in the estimated Chl*a* and uncertainty from OLCI data for "SFB1" and "SFB3" in San Francisco Bay (see Fig. 11) over a period of one year (from January 7, 2019). According to the revisit rate and cloud cover, usable data for the chosen locations are only available on specific days of the year. The left-hand side column in Fig. 17 shows the daily variations, and the right-hand side column illustrates monthly variations of estimated Chl*a* (shown in red) and σ_{UNC} (shown in blue). Note that, for "SFB1" and "SFB3," σ_{UNC} is generally low with some high-frequency variations in the daily reporting. Another interesting observation is that the left-hand side column shows a clear correlation between the estimated Chl*a* and σ_{UNC}, i.e., spikes in estimated Chl*a* are also accompanied by spikes in σ_{UNC}. In addition, the fact that

average σ_{UNC} remains quite low $(0.07 - 0.11)$ for the entire year indicates that the predictions made for these locations are reliable.

VI. DISCUSSION

A. Validity of the MDN-Specific Uncertainty Metric

Based on the results of the experiments shown in Section IV, it is clear that the performance of the MDN models is highly dependent on the amount of noise in the spectral data. Experiment-I (in Section IV-A) particularly indicates that increased noise in the spectral data leads to performance deterioration consistent with expectations on how the noise affects the performance of the neural network models [40]. Here, the uncertainty metric shows a corresponding increase, which would indicate poorer performance to the user. Furthermore, the mean uncertainty over the entire labeled dataset seems to be proportional to the mean predictive error.
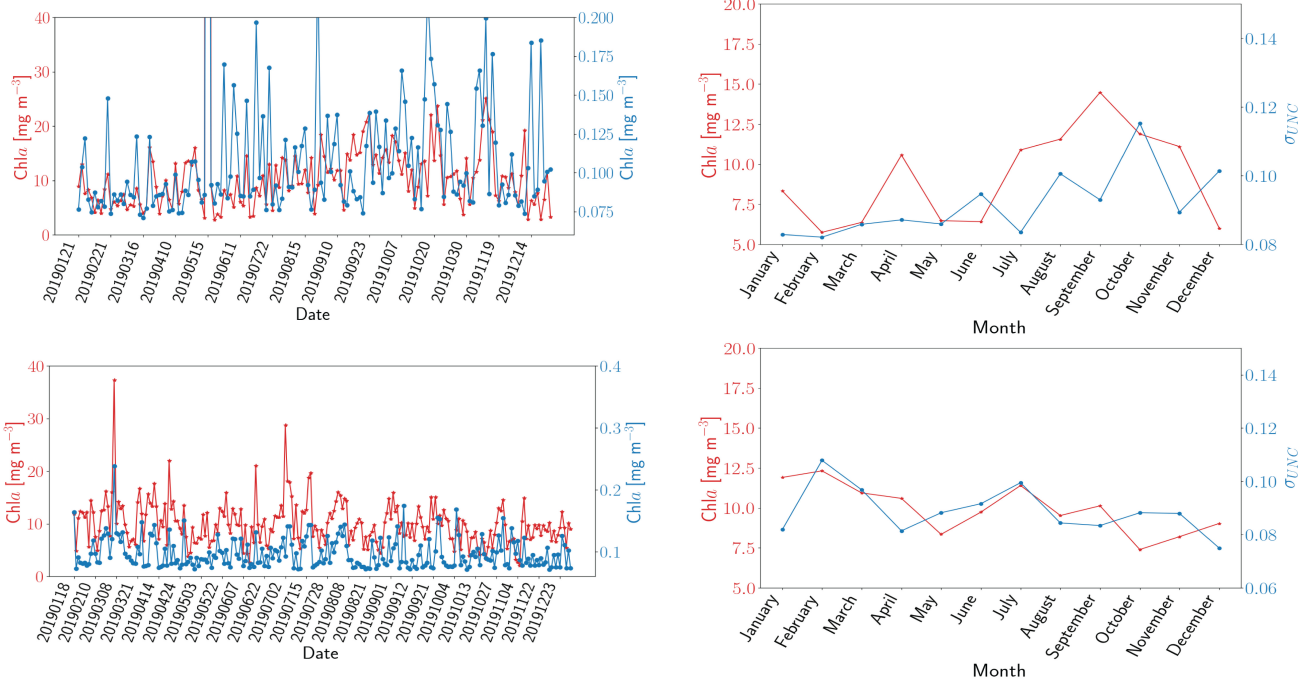
Fig. 17. Tracking the temporal changes in the estimated Chl$a$ (red) and uncertainty (blue) from OLCI for stations "SFB1" and "SFB3" in San Francisco Bay (see Fig. 11) over 2019. The left-hand side column shows the changes over all the available days in the year. The right-hand side column shows the monthly variations.

A similar observation can also be made in the presence of structural distortions (Experiment-III in Section IV-C). It is also interesting to note that the amount of uncertainty is proportional to the average distortion in the data. As the value of α increases, we note that both $\mathbb{E}(\sigma_{\text{UNC}})$ and MSLE increase slightly (see Fig. 7). This is consistent with the fact that a smaller α (which corresponds to the distortion multiplier being drawn from larger intervals) permits much smaller random multiplicative values and, thus, injects lesser distortion into R_{rs}. Taken in concert, these experiments make a strong case that the estimated uncertainty metric is proportional to the amount of noise/distortion present in the data. On the other hand, Experiment-II (see Section IV-B) indicates that the uncertainty metric is also able to identify novel test samples that are completely unlike the training samples. In addition, σ_{UNC} also increases as the test samples get farther away from the training samples, i.e., σ_{UNC} for samples at the edge of the test set (i.e., furthest away from the training set), which is significantly higher than that for the test samples closer to the training set.

One important observation across all the experiments has been that the predictions for the OLI spectral resolution show significantly higher uncertainty relative to those from OLCI and MSI spectral band settings. This suggests that our MDN model is not as robust when Chl$a$ predictions are made from R_{rs} that lacks critical RE information. The spectral resolution also affects the correlation between the predictive error (MSLE) and the estimated uncertainty (Experiment-IV in Section IV-D). For both OLCI and MSI, the R^2-correlation coefficient between σ_{UNC} and MSLE indicates a high correlation, while the R^2-correlation coefficient for OLI is quite low. Also, the estimated uncertainty

appears to be an upper bound on the predictive error for most samples (Experiment-V in Section IV-D). Taken in concert, these observations indicate that the proposed uncertainty metric captures the uncertainty associated with the challenging situations encountered when handling satellite-derived products, which contains higher noise levels relative to in situ data. We also demonstrated that the proposed uncertainty metric can be reasonably used as a proxy for the error for unlabeled test samples.

B. Applications MDN-Specific Uncertainty Metric

Based on the results of Application 1 (see Section V-A), for both OLCI and MSI, the uncertainty or predictive errors in a specific range are found to be inversely proportional to the number of samples in that range (i.e., the performance (in terms of both $\mathbb{E}(\sigma_{\text{UNC}})$ and MSLE) is worst in the ranges where the training data are sparsest). For OLI, on the other hand, the model performs poorly across nearly all the Chl$a$ ranges. These observations also indicate that any advancements in the model performance in a specific range can be achieved for OLCI and MSI by adding more labeled data samples in that range.

Application 2 (see Section V-B) also corroborates that OLI shows significantly higher uncertainty than that from OLCI and MSI. Furthermore, while the exact values of the Chl$a$ predicted for the data at the three different spectral resolutions appear numerically different, in terms of uncertainty, the spatial distributions and magnitudes of maps from OLCI and MSI are far more similar compared to the OLI, which is the clear outlier. It is also interesting to note that, for both San Francisco and the Chesapeake Bay, the OLCI predictions

TABLE IV
RELEVANT TECHNICAL CHARACTERISTICS OF THE MULTISPECTRAL SENSORS EVALUATED THROUGHOUT THIS STUDY

	OLI	MSI	OLCI
Nominal spatial resolution (m)	30	10, 20, 60	300
Swath (km)	185	290	1270
# of relevant bands ¡ 800 nm	4	7	13
Spectral bands (nm)	443, 482, 560, 655	443, 492, 560, 665, 705, 740, 783	410, 443, 490, 510, 560,620, 665, 673, 681, 708, 753, 778
Full-width-half-maximum (nm)	20 – 60	20 – 65	15
Lifetime	2013 –	2015 –	2016 –
Data source	https://earthexplorer.usgs.gov	https://scihub.copernicus.eu	

show generally low uncertainties (≤ 0.15) for inland and near-shore coastal waters. High and intermediate uncertainties are only seen over the open ocean. This seems reasonable as the in situ training dataset comprises primarily of samples from inland and near coastal waters. Comparing R_{rs} from various regions in the satellite datasets shows that, for low estimated uncertainties (i.e., $\sigma_{UNC} \leq 0.15$), there is a close match between the spectral shapes of the satellite-derived R_{rs} and its closest counterpart in the in situ database. Note that a close match in spectral shape by itself does not guarantee a low σ_{UNC} value. A known spectrum may have high or intermediate uncertainty due to the multimodality of the true target distribution or because there are not enough samples like the test spectrum in the training set. This indicates that the model accurately flags predictions for samples unlike the ones in training as "uncertain." On the other hand, when there are clear differences between a satellite-derived R_{rs} and its nearest in situ sample in terms of shape, the uncertainty is significantly higher across the various locations and sensors.

VII. CONCLUSION AND FUTURE WORK

Built upon previous efforts, we analyzed the feasibility of using techniques previously proposed by Choi et al. [40] to extend the capability of the MDNs for performing uncertainty estimation concurrently to the prediction of Chl*a* from existing multispectral satellite sensors, such as Sentinel-3/OLCI, Sentinel-2/MSI, and Landsat-8/OLI [37], [38]. The proposed uncertainty estimation method has the advantage of not requiring additional processing, modification of the MDN network architectures, or any retraining of the MDNs. Based on several experimental validations, the uncertainty metric captures the uncertainty caused by common disruptions in satellite-derived radiometric products, such as increased noise, structural/atmospheric distortions, and unseen/novel data. Furthermore, our experiments showed that the estimated uncertainties are not only strongly correlated with the error associated with predictions on the labeled dataset but also agree across the three sensors in terms of general trends. Nevertheless, the less robust uncertainty estimation was associated with Chl*a* approximations from OLI-like R_{rs} spectra due to OLI's lack of spectral measurements at the RE band critical for quantifying Chl*a* in inland and coastal waters. Another simple test on the labeled dataset suggested that, for most of the labeled data samples, the estimated uncertainty provides realistic bounds for the MDN model predictions. The practicality of the proposed uncertainty

metric was further demonstrated for recorded satellite imagery. We gauged model performance categorically for different TSs and determined Chl*a* ranges where the model is most accurate in quantifying uncertainties. Chl*a* uncertainty maps were further produced from near-simultaneous OLCI, MSI, and OLI imagery for three different sites with various aquatic and atmospheric conditions. In general, we found OLCI and MSI uncertainty maps consistent in terms of magnitude and relative spatial distribution, whereas corresponding OLI maps demonstrated much larger uncertainties. Temporal analyses of produced uncertainties from OLCI at select locations corroborated the viability of our uncertainty estimation for practical applications.

Future work will focus on two primary aspects of uncertainty estimation. The first stream will focus on extending the analysis on uncertainty estimation to MDN-based prediction of other biogeochemical variables, such as the concentration of total suspended solids (TSS) and the absorption by the colored dissolved organic materials (a_{cdom}) and other IOPs from multispectral or hyperspectral spectra. The second stream will involve combining the uncertainty associated with predictions of the MDNs with the uncertainty associated with other aspects of the data acquisition (e.g., uncertainties in R_{rs}). Such attempts may either focus on using the uncertainty for each stage of the data acquisition, processing, and processing pipelines as inputs to the next stages or modifying the MDN loss functions to include uncertainty from previous stages.

APPENDIX A
TECHNICAL CHARACTERISTICS OF THE SENSORS USED FOR THE EVALUATION IN THIS STUDY

Table IV provides the reader with the relevant technical characteristics of the three sensors considered in this study.

APPENDIX B
COMPARING UNCERTAINTY TO OTHER ERROR METRICS

While the primary metric for predictive error in this article has been the MLSE, it is important to note that similar trends/observations can be seen with any predictive error metric, such as median symmetric accuracy (MdSA), root mean square error (RMSE), and mean absolute percentage error (MAPE). Given the standard formulation, wherein the true Chl*a* values for N samples are represented by y and the predicted values by $\hat{y}$, the various metrics can be defined as
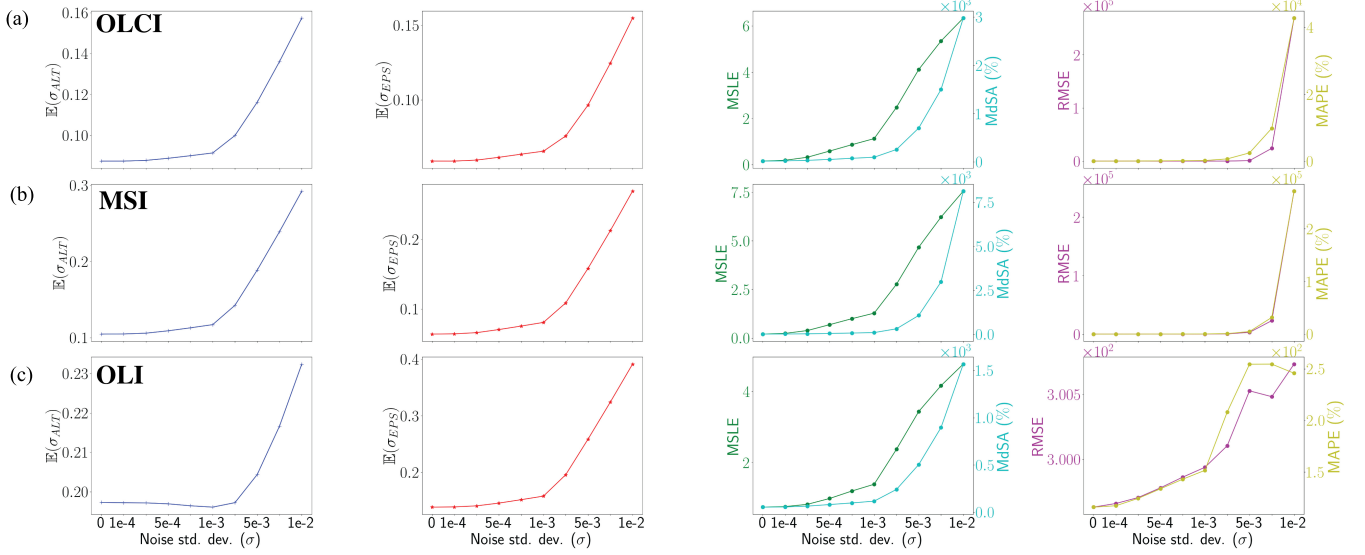
Fig. 18. Comparison of $\mathbb{E}(\sigma_{\text{ALT}})$ and $\mathbb{E}(\sigma_{\text{EPS}})$ to the different error metrics for MDNs trained on (a) OLCI-like, (b) MSI-like, and (c) OLI-like spectra at different noise levels.
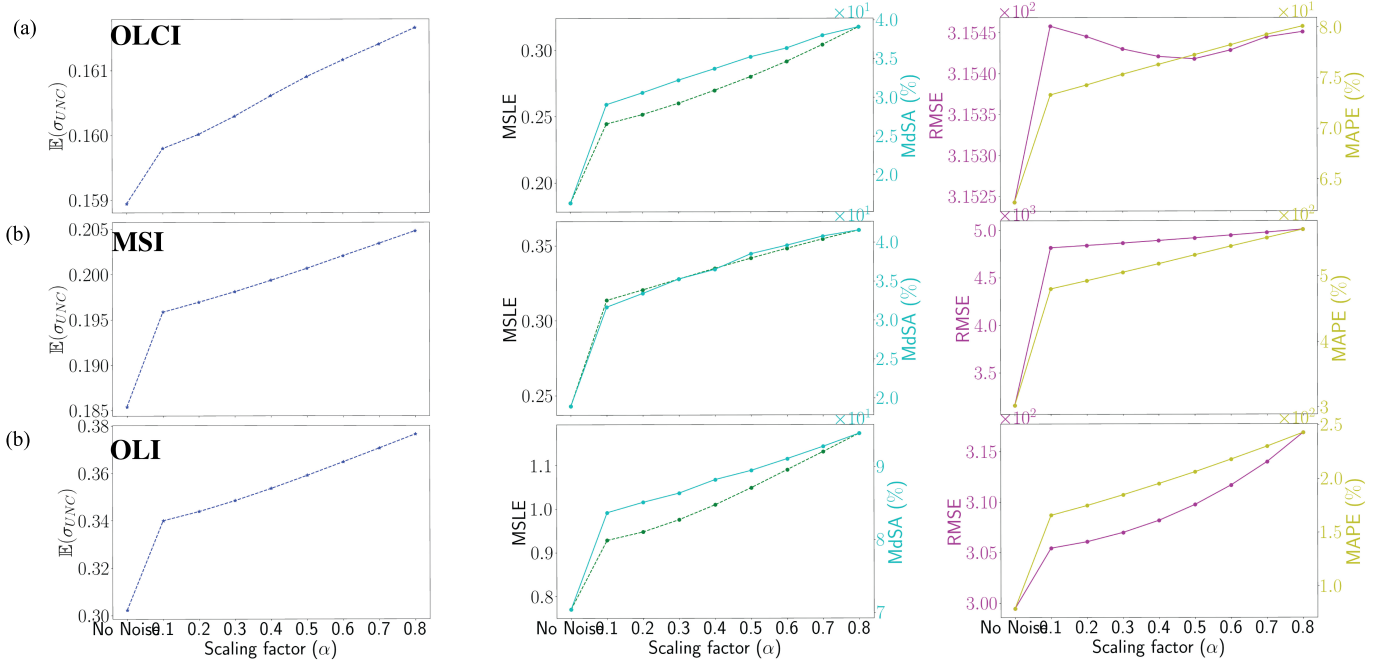

Fig. 19. Effect of the systematic noise due to spectrally dependent uncertainties from atmospheric correction (of the type shown on Fig. 4) in terms $\mathbb{E}(\sigma_{\text{UNC}})$ and various predictive error metrics for the different band settings (a) OLCI, (b) MSI, and (c) OLI.

follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\|y - \hat{y}\|^2}$$

$$\text{MdSA} = 100\left[\exp\left(Md\left(\log\left(\frac{\hat{y}}{y}\right)\right)\right) - 1\right]$$

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i}. \tag{8}$$

To show the similarity of the trends, we will repeat Experiment-I (see Section III-C1) and Experiment-III (see Section III-C2) with all the above metrics. The results for Experiment-I with the additional metrics are shown in Fig. 18.

Similarly, Fig. 19 shows similar trends for Experiment-III as well. Note that, despite having very different scales, the different metrics show the same exact trends and do not in any way change the conclusions drawn from these experiments. With that said, note that the RMSE is somewhat noisier than the log-scaled metrics, perhaps due to its susceptibility to outliers.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization: Brandon Smith; supervision: Nima Pahlevan; methodology: Arun M. Saranathan; formal analysis: Arun M. Saranathan; visualization: Arun M. Saranathan and Brandon Smith; funding acquisition: Nima Pahlevan;

writing—original draft: Arun M. Saranathan; and writing—review and editing: Arun M. Saranathan, Brandon Smith, and Nima Pahlevan.

REFERENCES

[1] W. W. Carmichael, "Health effects of toxin-producing cyanobacteria: 'The CyanoHABs,'" *Hum. Ecol. Risk Assessment, Int. J.*, vol. 7, no. 5, pp. 1393–1407, 2001.

[2] B. W. Brooks et al., "Are harmful algal Blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems?" *Environ. Toxicol. Chem.*, vol. 35, no. 1, pp. 6–13, Jan. 2016.

[3] G. L. Clarke, G. C. Ewing, and C. J. Lorenzen, "Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration," *Science*, vol. 167, no. 3921, pp. 1119–1121, Feb. 1970.

[4] H. R. Gordon, D. K. Clark, J. W. Brown, O. B. Brown, R. H. Evans, and W. W. Broenkow, "Phytoplankton pigment concentrations in the middle Atlantic bight: Comparison of ship determinations and CZCS estimates," *Appl. Opt.*, vol. 22, no. 1, pp. 20–36, Jan. 1983.

[5] R. C. Smith and K. S. Baker, "Oceanic chlorophyll concentrations as determined by satellite (Nimbus-7 coastal zone color scanner)," *Mar. Biol.*, vol. 66, no. 3, pp. 269–279, 1982.

[6] C. T. Wezernak, F. J. Tanis, and C. A. Bajza, "Trophic state analysis of inland lakes," *Remote Sens. Environ.*, vol. 5, pp. 147–164, Jan. 1976.

[7] A. G. Dekker, T. J. Malthus, M. M. Wijnen, and E. Seyhan, "Remote sensing as a tool for assessing water quality in loosdrecht lakes," *Hydrobiologia*, vol. 233, nos. 1–3, pp. 137–159, May 1992.

[8] S. C. Jain and J. R. Miller, "Subsurface water parameters: Optimization approach to their determination from remotely sensed water color data," *Appl. Opt.*, vol. 15, no. 4, pp. 886–890, 1976.

[9] J. A. Yoder, C. R. McClain, G. C. Feldman, and W. E. Esaias, "Annual cycles of phytoplankton chlorophyll concentrations in the global ocean: A satellite view," *Global Biogeochem. Cycles*, vol. 7, no. 1, pp. 181–193, Mar. 1993.

[10] E. Knight and G. Kvaran, "Landsat-8 operational land imager design, characterization and performance," *Remote Sens.*, vol. 6, no. 11, pp. 10286–10305, 2014.

[11] B. Markham et al., "Landsat-8 operational land imager radiometric calibration and stability," *Remote Sens.*, vol. 6, no. 12, pp. 12275–12308, 2014.

[12] D. P. Roy et al., "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sens. Environ.*, vol. 145, pp. 154–172, Apr. 2014.

[13] K. Toming, T. Kutser, A. Laas, M. Sepp, B. Paavel, and T. Nõges, "First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery," *Remote Sens.*, vol. 8, no. 8, p. 640, 2016.

[14] N. Pahlevan, S. Sarkara, B. A. Franza, S. V. Balasubramanian, and J. Hec, "Sentinel-2 multispectral instrument (MSI) data processing for aquatic science applications: Demonstrations and validations," *Remote Sens. Environ.*, vol. 201, pp. 47–56, Nov. 2017.

[15] J. Nieke et al., "Ocean and land color imager on Sentinel-3," in *Optical Payloads for Space Missions*. Chichester, U.K.: Wiley, 2015, pp. 223–245.

[16] C. Neil, E. Spyrakos, P. D. Hunter, and A. N. Tyler, "A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types," *Remote Sens. Environ.*, vol. 229, pp. 159–178, Aug. 2019.

[17] J. E. O'Reilly et al., "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res., Oceans*, vol. 103, pp. 24937–24953, Oct. 1998.

[18] F. Watanabe, E. Alcântara, T. Rodrigues, N. Imai, C. Barbosa, and L. Rotta, "Estimation of Chlorophyll-a concentration and the trophic state of the Barra Bonita hydroelectric reservoir using OLI/Landsat-8 images," *Int. J. Environ. Res. Public Health*, vol. 12, no. 9, pp. 10391–10417, Aug. 2015.

[19] R. P. Bukata, J. H. Jerome, J. E. Bruton, S. C. Jain, and H. H. Zwick, "Optical water quality model of lake Ontario. 1: Determination of the optical cross sections of organic and inorganic particulates in lake Ontario," *Appl. Opt.*, vol. 20, no. 9, pp. 1696–1703, 1981.

[20] F. H. Freitas and H. M. Dierssen, "Evaluating the seasonal and decadal performance of red band difference algorithms for chlorophyll in an optically complex estuary with winter and summer Blooms," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111228.

[21] C. Le et al., "Towards a long-term chlorophyll-a data record in a turbid estuary using MODIS observations," *Prog. Oceanogr.*, vol. 109, pp. 90–103, Feb. 2013.

[22] A. A. Gitelson, J. F. Schalles, and C. M. Hladik, "Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake bay case study," *Remote Sens. Environ.*, vol. 109, no. 4, pp. 464–472, Aug. 2007.

[23] D. Gurlin, A. A. Gitelson, and W. J. Moses, "Remote estimation of chl-a concentration in turbid productive waters—Return to a simple two-band NIR-red model?" *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3479–3490, Dec. 2011.

[24] W. J. Moses, A. A. Gitelson, S. Berdnikov, V. Saprygin, and V. Povazhnyi, "Operational MERIS-based NIR-red algorithms for estimating chlorophyll-a concentrations in coastal waters—The Azov sea case study," *Remote Sens. Environ.*, vol. 121, pp. 118–124, Jun. 2012.

[25] E. J. Kwiatkowska and G. S. Fargion, "Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, Dec. 2003.

[26] H. Zhan, P. Shi, and C. Chen, "Retrieval of oceanic chlorophyll concentration using support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, Dec. 2003.

[27] Z. Cao et al., "A machine learning approach to estimate Chlorophyll-a from Landsat-8 measurements in inland lakes," *Remote Sens. Environ.*, vol. 248, Oct. 2020, Art. no. 111974.

[28] L. G. Vilas, E. Spyrakos, and J. M. T. Palenzuela, "Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician rias (NW Spain)," *Remote Sens. Environ.*, vol. 115, no. 2, pp. 524–535, Feb. 2011.

[29] L. Gross, S. Thiria, R. Frouin, and B. G. Mitchell, "Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration," *J. Geophys. Res., Oceans*, vol. 105, no. C2, pp. 3483–3495, Feb. 2000.

[30] I. Ioannou, A. Gilerson, B. Gross, F. Moshary, and S. Ahmed, "Neural network approach to retrieve the inherent optical properties of the ocean from observations of MODIS," *Appl. Opt.*, vol. 50, no. 19, pp. 3168–3186, 2011.

[31] C. Jamet, H. Loisel, and D. Dessailly, "Retrieval of the spectral diffuse attenuation coefficient $K_d(\lambda)$ in open and coastal ocean waters using a neural network inversion," *J. Geophys. Research: Oceans*, vol. 117, no. C10, Oct. 2012, Art. no. C10023.

[32] T. Kajiyama, D. D'Alimonte, and G. Zibordi, "Algorithms merging for the determination of chlorophyll-a concentration in the black sea," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 677–681, May 2019.

[33] H. Schiller and R. Doerffer, "Neural network for emulation of an inverse model operational derivation of case II water properties from MERIS data," *Int. J. Remote Sens.*, vol. 20, no. 9, pp. 1735–1746, Jan. 1999.

[34] M. Hieronymi, D. Müller, and R. Doerffer, "The OLCI neural network swarm (ONNS): A bio-geo-optical algorithm for open ocean and coastal waters," *Front. Mar. Sci.*, vol. 4, p. 140, May 2017.

[35] Y. Fan et al., "OC-SMART: A machine learning based data analysis platform for satellite ocean color sensors," *Remote Sens. Environ.*, vol. 253, Feb. 2021, Art. no. 112236.

[36] M. Sydor, R. W. Gould, R. A. Arnone, V. I. Haltrin, and W. Goode, "Uniqueness in remote sensing of the inherent optical properties of ocean water," *Appl. Opt.*, vol. 43, no. 10, pp. 2156–2162, 2004.

[37] N. Pahlevan et al., "Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach," *Remote Sens. Environ.*, vol. 240, Apr. 2020, Art. no. 111604.

[38] B. Smith et al., "A Chlorophyll-a algorithm for Landsat-8 based on mixture density networks," *Frontiers Remote Sens.*, vol. 1, Feb. 2021, Art. no. 623678.

[39] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in AI safety," 2016, *arXiv:1606.06565*.

[40] S. Choi, K. Lee, S. Lim, and S. Oh, "Uncertainty-aware learning from demonstration using mixture density networks with sampling-free variance modeling," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 6915–6922.

[41] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–12.

[42] F. Mélin, Ed., "Uncertainties in ocean colour remote sensing," Int. Ocean-Colour Coordinating Group (IOCCG), IOCCG Rep. 18, 2019.

[43] F. Mélin, J.-F. Berthon, and G. Zibordi, "Assessment of apparent and inherent optical properties derived from SeaWiFS with field data," *Remote Sens. Environ.*, vol. 97, no. 4, pp. 540–553, Sep. 2005.

[44] R. J. W. Brewin et al., "The ocean colour climate change initiative: III. A round-robin comparison on in-water bio-optical algorithms," *Remote Sens. Environ.*, vol. 162, pp. 271–294, Jun. 2015.

[45] L. I. W. McKinna, I. Cetinić, A. P. Chase, and P. J. Werdell, "Approach for propagating radiometric data uncertainties through NASA ocean color algorithms," *Frontiers Earth Sci.*, vol. 7, p. 176, Jul. 2019.

[46] M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021.

[47] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, 2021.

[48] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in *Proc. 25th Int. Conf. Mach. Learn. - ICML*, 2008, pp. 880–887.

[49] K. Posch, J. Steinbrener, and J. Pilz, "Variational inference to measure model uncertainty in deep neural networks," 2019, *arXiv:1902.10189*.

[50] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[51] K. Shinde, J. Lee, M. Humt, A. Sezgin, and R. Triebel, "Learning multiplicative interactions with Bayesian neural networks for visual-inertial odometry," 2020, *arXiv:2007.07630*.

[52] C. D. Mobley, "Estimation of the remote-sensing reflectance from above-surface measurements," *Appl. Opt.*, vol. 38, no. 36, pp. 7442–7455, 1999.

[53] Q. Vanhellemont and K. Ruddick, "Atmospheric correction of Sentinel-3/OLCI data for mapping of suspended particulate matter and Chlorophyll-*a* concentration in Belgian turbid coastal waters," *Remote Sens. Environ.*, vol. 256, Apr. 2021, Art. no. 112284.

[54] C. M. Bishop, "Mixture density networks," Aston Univ., Birmingham, U.K., Tech. Rep., 1994.

[55] M. Defoin-Platel and M. Chami, "How ambiguous is the inverse problem of ocean color in coastal waters?" *J. Geophys. Res., Oceans*, vol. 112, no. C3, 2007, Art. no. C03004.

[56] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.

[57] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, p. 417, 1933.

[58] F. R. S. K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Philosoph. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.

[59] H. R. Gordon, "Atmospheric correction of ocean color imagery in the Earth observing system era," *J. Geophys. Res.*, vol. 102, no. D14, pp. 17081–17106, 1997.

[60] N. Pahlevan et al., "ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters," *Remote Sens. Environ.*, vol. 258, Jun. 2021, Art. no. 112366.

[61] L. De Keukelaere et al., "Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: Validation for coastal and inland waters," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 525–542, Jan. 2018.

[62] C. Goyens, C. Jamet, and T. Schroeder, "Evaluation of four atmospheric correction algorithms for MODIS-Aqua images over contrasted coastal waters," *Remote Sens. Environ.*, vol. 131, pp. 63–75, Apr. 2013.

[63] P. R. Renosh, D. Doxaran, L. D. Keukelaere, and J. I. Gossn, "Evaluation of atmospheric correction algorithms for Sentinel-2-MSI and Sentinel-3-OLCI in highly turbid estuarine waters," *Remote Sens.*, vol. 12, no. 8, p. 1285, Apr. 2020.

[64] M. A. Warren et al., "Assessment of atmospheric correction algorithms for the Sentinel-2A multispectral imager over coastal and inland waters," *Remote Sens. Environ.*, vol. 225, pp. 267–289, May 2019.

[65] J. Wei, Z. Lee, and S. Shang, "A system to measure the data quality of spectral remote-sensing reflectance of aquatic environments," *J. Geophys. Res. Oceans*, vol. 121, no. 11, pp. 8189–8207, 2016.

[66] R. E. Carlson, "A trophic state index for lakes 1," *Limnol. Oceanogr.*, vol. 22, no. 2, pp. 361–369, 1977.

Arun M. Saranathan (Member, IEEE) is currently pursuing the Ph.D. degree in electrical engineering with the University of Massachusetts Amherst, Amherst, MA, USA.

He is currently a Senior Research Scientist with Science Systems and Applications Inc. (SSAI), NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA. Since joining SSAI in July 2021, he has been primarily focused on analyzing the uncertainties associated with predictions of machine learning models used for predicting water quality parameters from remote sensing data. His primary research focus has been on adapting/customizing techniques from various fields, such as machine learning, deep learning, signal, and image processing for automated analysis of large planetary and terrestrial datasets.

Brandon Smith (Member, IEEE) received the master's degree with a focus on machine learning and algorithm development from the Georgia Institute of Technology, Atlanta, GA, USA, in 2017.

He is currently the Lead Computer Scientist with Science Systems and Applications Inc. (SSAI), NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA. Since then, he has been working in various roles with SSAI with a primary focus on creating machine learning models for estimating water quality parameters from multispectral and hyperspectral remote sensing data. He has also been involved with the Atmospheric Correction Intercomparison Exercise (ACIX), an international initiative to compare various atmospheric correction (AC) algorithms for moderate-resolution optical sensors.

Nima Pahlevan (Senior Member, IEEE) received the Ph.D. degree in imaging science from the Rochester Institute of Technology (RIT), Rochester, NY, USA, in 2012.

He is currently a Remote Sensing Scientist with Science Systems and Applications Inc. (SSAI), Terrestrial Information Systems Laboratory, NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA. He was formerly trained in geospatial engineering and remote sensing. Prior to joining GSFC in 2014, he finished a two-year postdoctoral term working on in situ and satellite-based ocean color measurements at the University of Massachusetts Boston, Boston, MA, USA. He is a member of Landsat, PACE, and Terra/Aqua/SNPP Science Teams, participates in GEO AquaWatch activities, and contributes to the surface biology geology (SBG) preformulation studies. He also leads the Global Freshwater Sensing Group that focuses on developing novel machine-learning models to generate globally robust aquatic science products (e.g., water quality indicators) from multispectral and hyperspectral instruments. His main area of research lies within the aquatic remote sensing domain with a focus on algorithm developments, atmospheric correction, calibration/validation, impacts of climate variability on water resources, harmful algal blooms (HABs), and relevant applied science practices.