UNCERTAINTY QUANTIFICATION WITHIN AQUAVERSE FRAMEWORK: ESTIMATION, VALIDATION, AND CALIBRATION

Currently, the core algorithm of the Aquaverse (Aquatic inversion scheme for optical sensing of fresh and coastal waters) is the Mixture Density Networks (MDNs), which have been proven to be a powerful estimator for a variety of in-water variables[1,2]. Given their proven estimation ability, the practical application of such models requires the quantification of the uncertainty associated with the prediction of such models. This uncertainty document only refers to the properties of the specific MDN used for prediction (e.g., its architecture, hyperparameters, the relationship of training and test datasets), and a complete measure of the estimation uncertainty will need to include uncertainty from other sources, such as measurement uncertainties, atmospheric corrections.

Uncertainty Estimation for MDNs

The MDNs are a variant of the neural networks wherein the prediction for each test sample is a probability distribution rather than a point value. In particular, the MDN models the distribution as a mixture of Gaussians and predicts the parameters of the distribution as its output. For this distribution, one can estimate the predicted values as the mean of the largest Gaussian component in the MDN. Previous work[3,4] has that the probabilistic nature of the MDN output also encodes the information on the confidence of the model in the specific prediction. In particular, the variance/standard deviation associated with the model provides users with a good measure of the model's confidence. The variance is affected by two main factors (i) the variances of the individual Gaussians (which are interpreted as the uncertainty in the model estimate due to noise) and (ii) the dispersion between the component means (a measure of the model's uncertainty in the actual value). In combination, these two components provide a solid basis for understanding the model's confidence in its prediction. Experiments have shown that Rrs spectra for which the model predicts with high confidence have unimodal distribution with low component means, while low confidence models have high component variances or are highly multi-modal (or both).

Validating the MDN uncertainty metric

The variance/standard deviation-based uncertainty metric mentioned in the previous section has been tested for MDNs trained on a labeled dataset of about 10,000 samples made up of collocated measurements of *in situ* Rrs and water constituent concentrations (Chl*a*, TSS, a_{cdom}(440), and PC), namely the GLORIA dataset[5]. The model was trained to accept the remote sensing Rrs as input and predict the optical parameters. Further, the uncertainty metric was tested for its robustness under three different scenarios (i) enhanced noise in the test samples, (ii) noise due to atmospheric correction, (iii) out-of-distribution samples (i.e., sample, unlike any of the models, has seen in training—like most data models the predictions of the MDN for such samples are subject to some doubt). Further, it was verified that even at the spectral resolution of multi-spectral sensors like MSI and OLCI, there appears to be a clear correlation between the error and uncertainty. Experiments also indicate that the uncertainty for a specific sample seems to depend on the number of samples like it in the training set, and one can improve model performance (for both parameter and uncertainty estimation) by exposing the model to more samples in regions where it does not seem to contain much data. Qualitative testing on a few satellite images (e.g., HICO, OLCI, PRISMA) indicates that the uncertainty maps appear reasonable, and spot checks on pixel spectra indicate that the uncertainty values are affected by the similarity between test samples and the training dataset. For more details, please refer to Saranathan et al. (2023)[4], a thorough validation of the MDN uncertainty metric for models trained on aquatic remote sensing data.

Calibration of the MDN uncertainty metric

While the MDN uncertainty metric has many valuable properties mentioned above, unlike many other uncertainty metrics (like the ones in measurement), it cannot be interpreted as a bound on the error expected in the prediction. Further, the scale of the uncertainty metric (unitless) seems quite arbitrary and changes quite significantly for changes in the model hyperparameters like the number of layers, nodes, outputs, etc. The calibration is an attempt to standardize the uncertainty across models, sensors, etc. to make these more intuitive and easier to understand for end-users by reporting them in physical units. In an effort to make the MDN uncertainties analogous to an error bound, an interval-based statistical method has been designed wherein the estimated uncertainties are scaled by a chosen calibration factor. The specific calibration factor is the 80th percentile ratio between the predictive error and MDN uncertainty for the samples in that interval (i.e., there is an 80% probability that the calibrated uncertainty is an upper bound on the error). Such calibration has been seen to make the uncertainty maps reveal more intuitive spatial trends and are easier to compare to prediction and maps from other sensors.

Conclusion & future work

This document describes a simple metric to capture the uncertainty associated with predictions from the MDNs. While the metric has many valuable properties, some work needs to be done in terms of improving the interpretability of such models. Attempts at calibrating the uncertainty metric for interpretability are ongoing to ultimately identify an optimal method for reporting pixel-wise uncertainties for absorbing IOPs, Chla, TSS, $a_{cdom}(440)$, and PC generated from PACE/OCI Rrs fields. Future work will include attempts to capture the effect of uncertainty propagation across all levels of the prediction pipeline from data acquisition, atmospheric correction, pre-processing, and ML-based prediction.

References

1. N. Pahlevan, B. Smith, J. Schalles, C. Binding, Z. Cao, R. Ma, K. Alikas, K. Kangro, D. Gurlin, N. Ha` et al., 2020, "*Seamless retrievals of chlorophyll-a from sentinel-2 (MSI) and sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach*," Remote Sensing of Environment, vol. 240, p. 111604.
2. O'Shea, R.E., Pahlevan, N., Smith, B., Bresciani, M., Egerton, T., Giardino, C., Li, L., Moore, T., Ruiz-Verdu, A., Ruberg, S. and Simis, S.G., 2021. "*Advancing cyanobacteria biomass estimation from hyperspectral observations: Demonstrations with HICO and PRISMA imagery*". *Remote Sensing of Environment*, *266*, p.112693.
3. Choi, S., Lee, K., Lim, S. and Oh, S., 2018, May. Uncertainty-aware learning from demonstration using mixture density networks with sampling-free variance modeling. In *2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 6915-6922). IEEE
4. Saranathan, A.M., Smith, B. and Pahlevan, N., 2023. "*Per-pixel uncertainty quantification and reporting for satellite-derived chlorophyll-a estimates via mixture density networks". IEEE Transactions on Geoscience and Remote Sensing*
5. Lehmann, M.K., Gurlin, D., Pahlevan, N., Alikas, K., Anstee, J., Balasubramanian, S.V., Barbosa, C.C., Binding, C., Bracher, A., Bresciani, M. and Burtner, A., 2023. GLORIA-A globally representative hyperspectral in situ dataset for optical sensing of water quality. *Scientific Data*, *10*(1), p.100.