Appendix

Appendix A. SNR assessment

Table A.1 contains the SD scene IDs used for SNR analysis. The 23-character IDs refer to the acquisition date, acquisition time, acquisition time seconds hundredths and satellite id.

Table A.1

SD scene IDs used for the SNR assessments presented in section 3.1.

SD Scene ID
20220830_184956_86_2426
20220820_183749_60_247b
20220815_180250_75_2453
20220726_180058_59_2451
20220711_183828_22_2489
20220626_183600_15_2495
20220621_182143_58_2251

Appendix B. Cross-calibration and validation

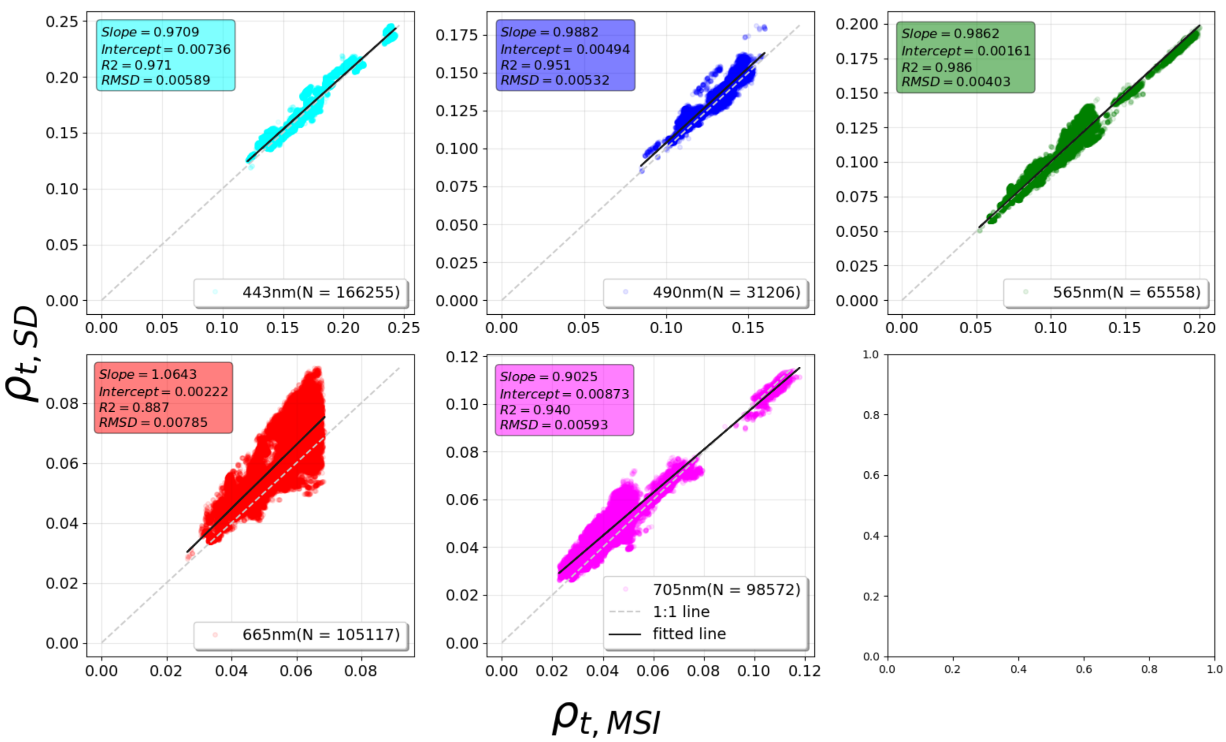

Figure B1 | Scatterplot of $\rho_{t,SD}$ versus $\rho_{t,MSI}$ for the five common SD-MSI bands from the calibration data, illustrating the radiometry difference between SD and MSI over water bodies.

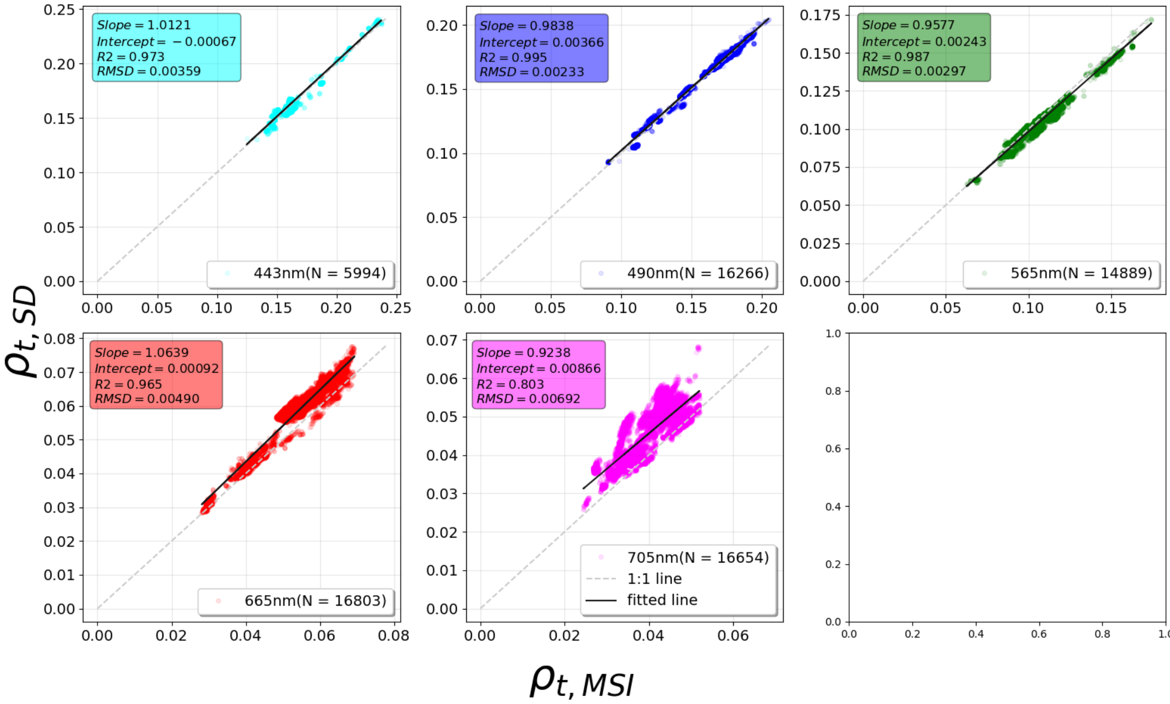

Figure B2 | Scatterplot of $\rho_{t,SD}$ versus $\rho_{t,MSI}$ for the five common SD-MSI bands from the validation data before applying calibration. This is exhibiting the ρ_t differences between SD and MSI over water bodies before applying calibration.

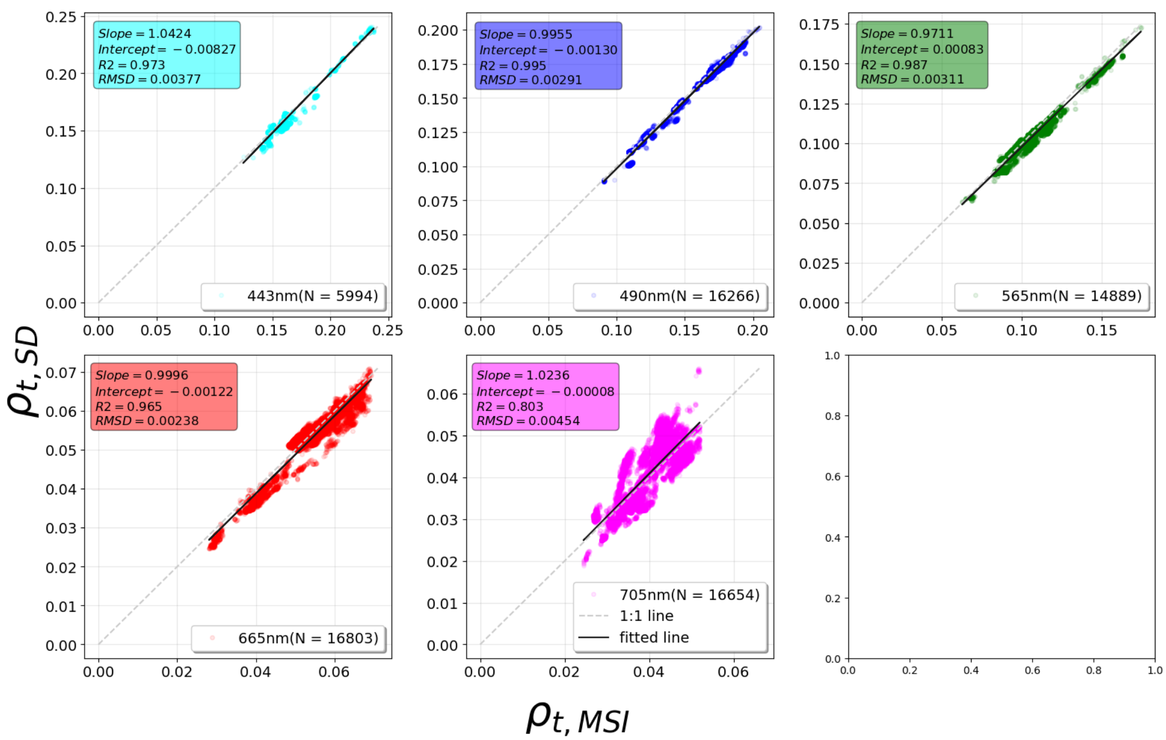

Figure B3 | Scatterplot of $\rho_{t,SD}$ versus $\rho_{t,MSI}$ for the five common SD-MSI bands from the validation data after applying calibration. This is exhibiting the ρ_t differences between SD and MSI over water bodies after applying calibration.

Appendix C. Water quality products

This study leveraged GLObal Reflectance community dataset for Imaging and optical sensing of Aquatic environments (GLORIA) database (Lehmann et al., 2023) for developing Z_{sd} retrieval model from the co-located radiometry (R_{rs}) and biogeochemical (Z_{sd}) samples. These *in situ* $R_{rs} - Z_{sd}$ (n = 3818) data originated from wide range of optical regimes, including lakes, reservoirs, rivers, estuaries, coastal waters, etc. These data were spilt randomly (50/50) to train and test the models.

The measured *in situ* hyperspectral R_{rs} were resampled with seven SD and MSI Relative Spectral Responses (RSRs) within the 400-800 nm spectral range. The resampled *in situ* R_{rs} data went through median-centering with interquartile range scaling and the *in situ* Z_{sd} values were log-scaled and transformed to be within (-1, 1). The MDN models take those preprocessed *in situ* data as input. The model architecture and hyperparameters were searched and minor performance gains were observed with a complex model compared to the configuration used in previous publications (Pahlevan et al., 2022, 2020; Smith et al., 2021b). Hence, the following model architecture and configuration is used. The neural network section of the model consists of five hidden layers with 100 neurons in each layer, and products are estimated with maximum likelihood estimation of five Gaussian mixtures. Ten models were trained with a learning and L2 normalization rate of 0.001, and median of them were used to predict the products.

Figure C1 shows the performance of MDNs for SD and MSI-like spectra simulated from hyperspectral *in situ* R_{rs} spectra. The median absolute percentage error (MAPE), bias, root mean squared log-difference (RMSLD) and slope are annotated. They are defined as follows.

$MAPE = 100 \times \tilde{r}$ where $\tilde{r}$ is the median of $[|\mathrm{Est_i} - \mathrm{Meas_i}|/\mathrm{Meas_i}]$ for i = 1…n,

bias = $100 * (10^{Z})$ $where\ Z = \sum_{i=1}^{n}(\log_{10}(\mathrm{Est_i}) - \log_{10}(\mathrm{Meas_i}))/n$

$\mathrm{RMSLD} = \left[\sum_{\mathrm{i=1}}^{\mathrm{N}}(\log_{10}(\mathrm{Est_i}) - \log_{10}(\mathrm{Meas_i}))^2/n\right]^{1/2}$

The MAPE, slope and RMSLD for the MSI and SD Z_{sd} products are very similar. Bias centered around 610 nm. Overall, these results suggest that the SD and MSI Z_{sd} products are expected to be similar when the R_{rs} products are similar.

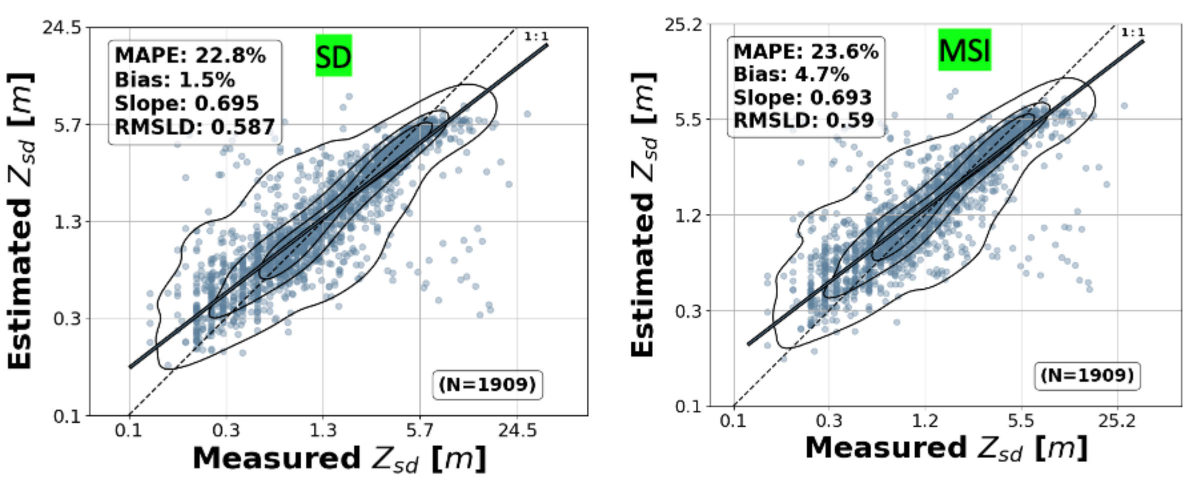

Figure C1 | Performance of MDNs for SD and MSI-like spectra simulated from hyperspectral *in situ* R_{rs} spectra. The number of test samples (N), MAPE, bias, slope and RMSLD are annotated.