

Study sites and observations

- Three ocean time-series stations: BATS, HOT Station ALOHA, and CARIACO.
- In situ chlorophyll-*a* profiles were combined with daily ESA OC-CCI satellite chlorophyll-*a* and $K_d(490)$.
- Satellite data were spatially averaged around each station.

Station	Latitude	Longitude	Profile depths [m]
BATS	31.6667°N	64.1667°W	3–140
HOT	22.7500°N	158.0000°W	5–175
CARIACO	10.5000°N	64.6667°W	3–140

Observed vertical chlorophyll structure

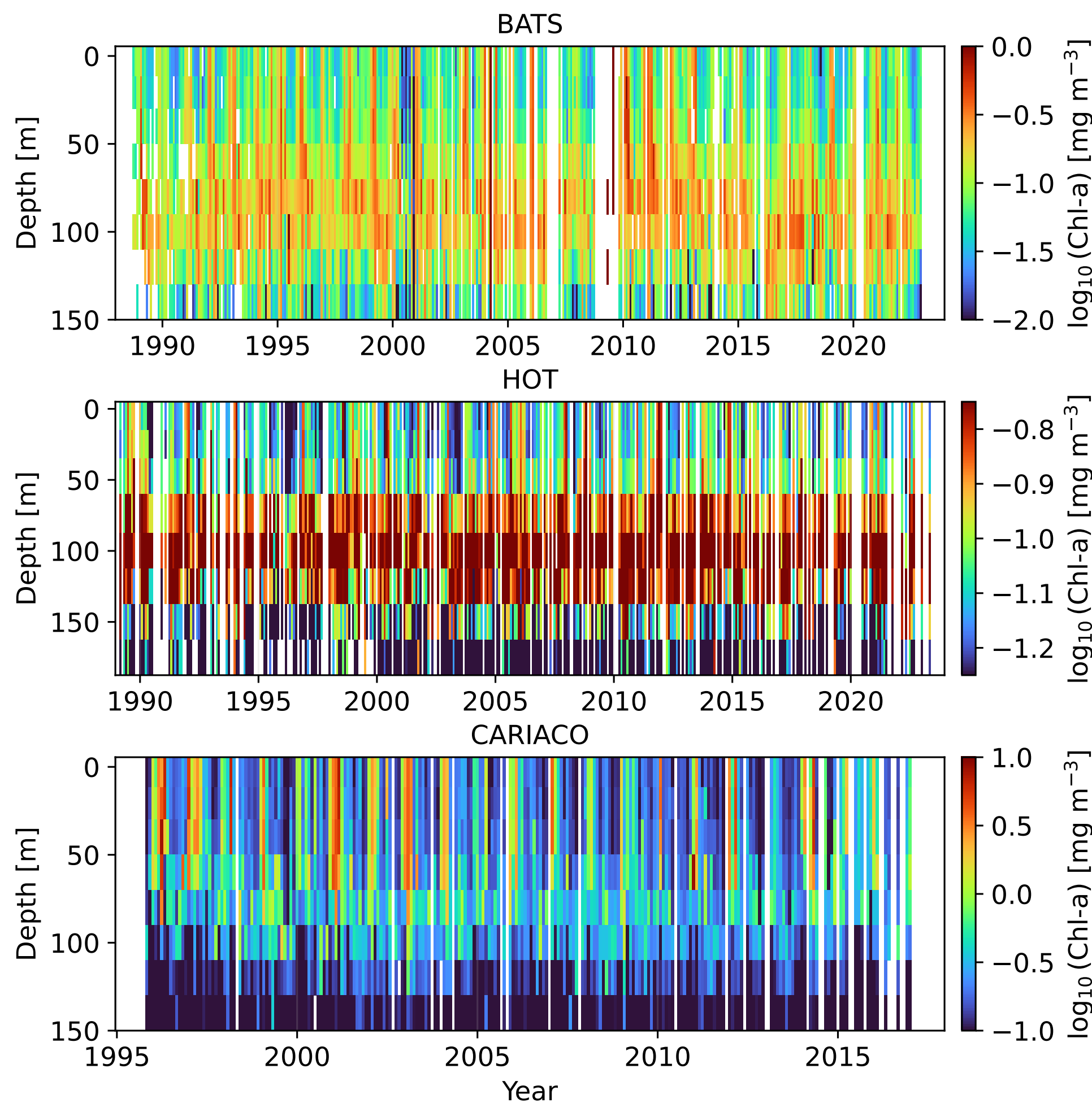


Figure. Hovmöller diagrams of observed in situ chlorophyll-*a* profiles at BATS, HOT, and CARIACO. Colors show $\log_{10}$ chlorophyll-*a*.

Satellite and vertically averaged in situ chlorophyll

The optical depth was estimated as

$$z_{\text{opt}} = \frac{1.7}{K_d(490)}.$$

The mean in situ chlorophyll concentration to this depth was

$$\overline{\text{Chl}}_{0-z_{\text{opt}}} = \frac{1}{z_{\text{opt}}} \int_0^{z_{\text{opt}}} \text{Chl}(z) dz.$$

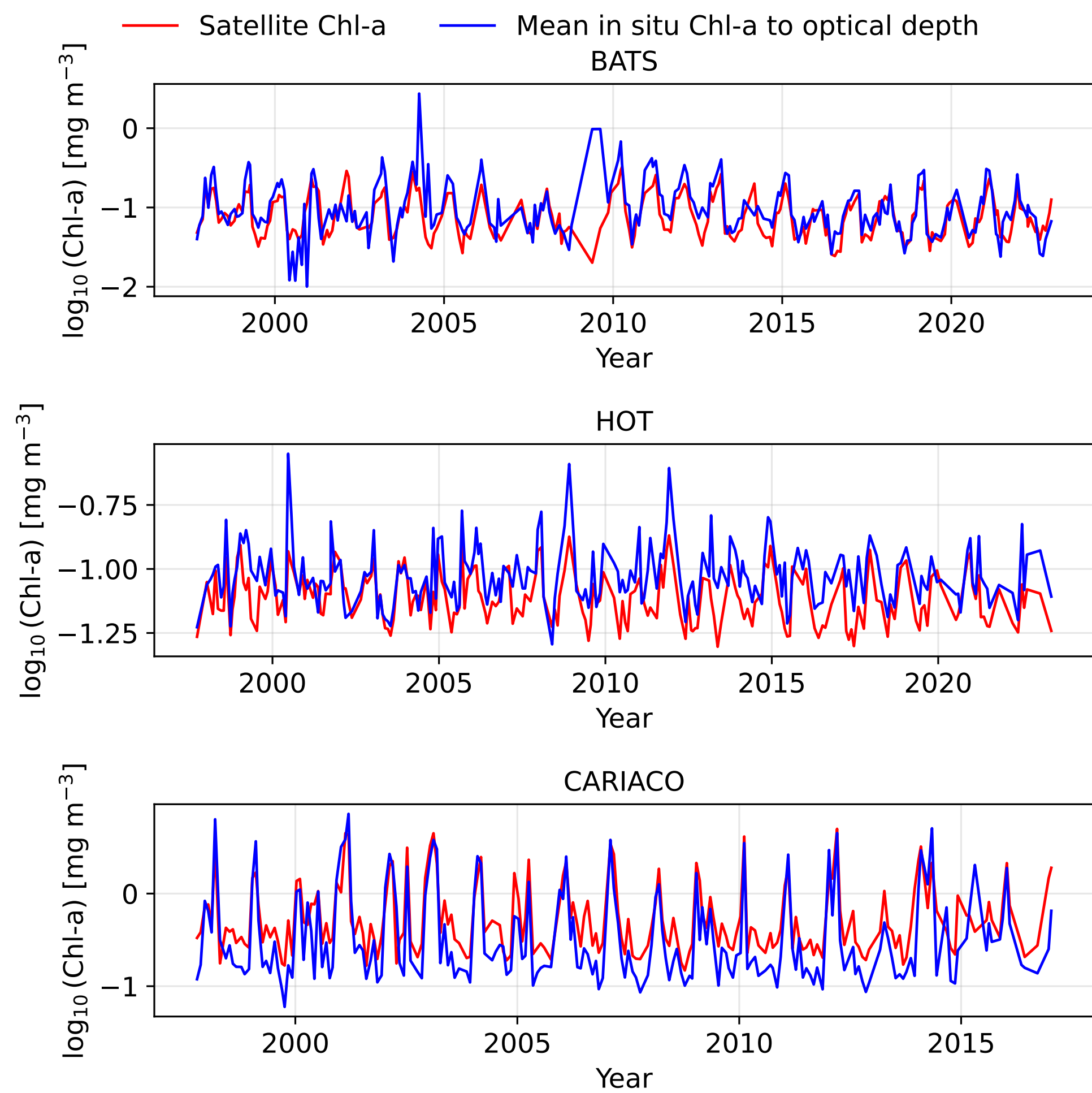


Figure. Time series of satellite and in situ chlorophyll-*a* at BATS, HOT, and CARIACO. Red lines show satellite chlorophyll-*a*; blue lines show mean in situ chlorophyll-*a* from the surface to z_{opt} .

Temporal satellite matching

Each in situ observation was paired with the nearest available satellite observation containing chlorophyll-*a*.

Station	<i>N</i>	0 d	1 d	2 d	> 2 d	Max.
BATS	280	176	81	15	8	6
HOT	229	148	57	11	13	15
CARIACO	207	146	43	12	6	6

- Within 2 days: BATS 97.1%, HOT 94.3%, CARIACO 97.1%.

Effect of satellite sampling time

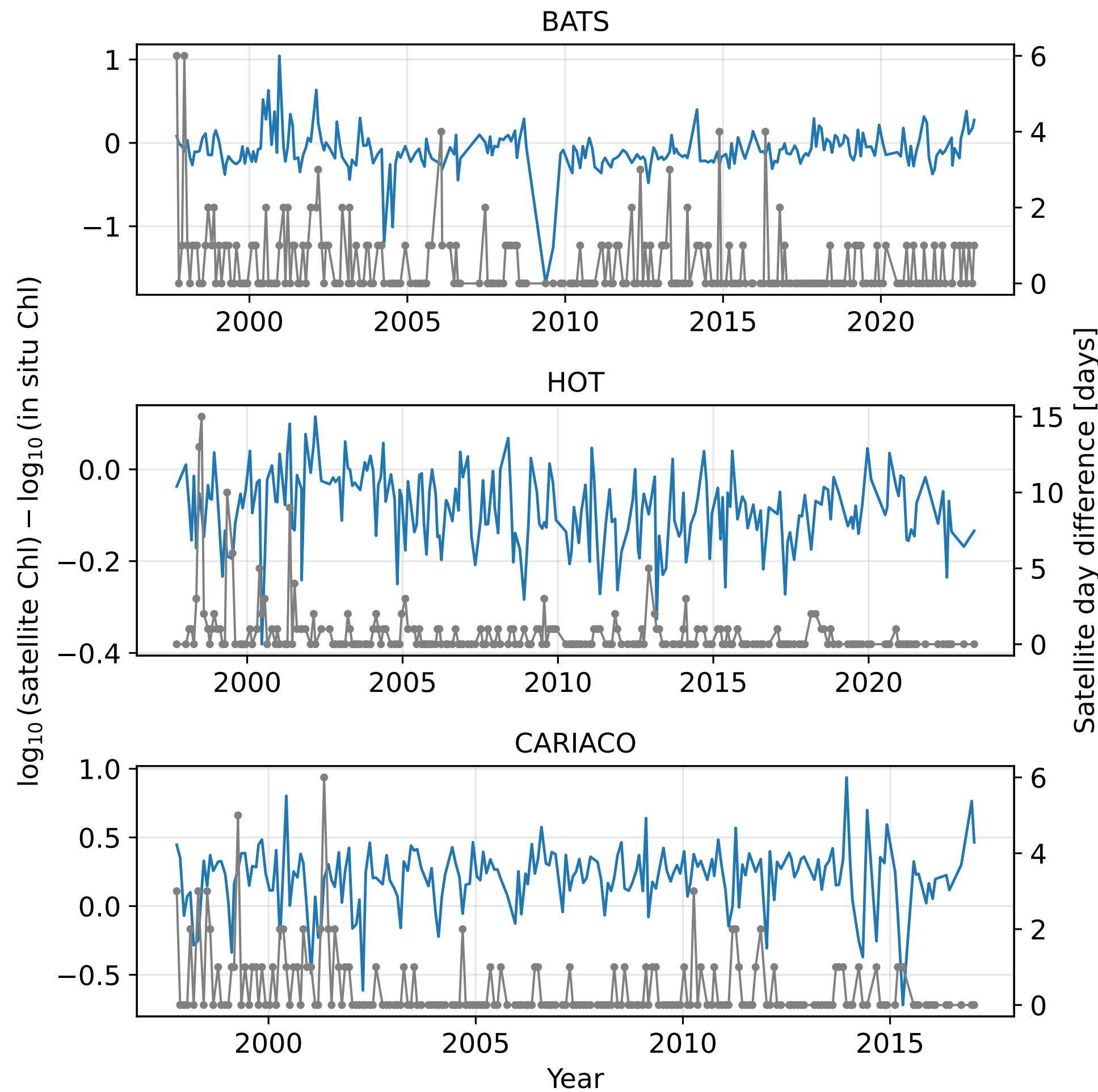


Figure. Temporal evolution of the satellite-in situ chlorophyll difference at BATS, HOT, and CARIACO. Black lines show $\Delta \log_{10}(\text{Chl-}a)$, while gray lines show the absolute number of days between the in situ sampling date and the nearest available satellite observation.

How deep does the satellite signal extend?

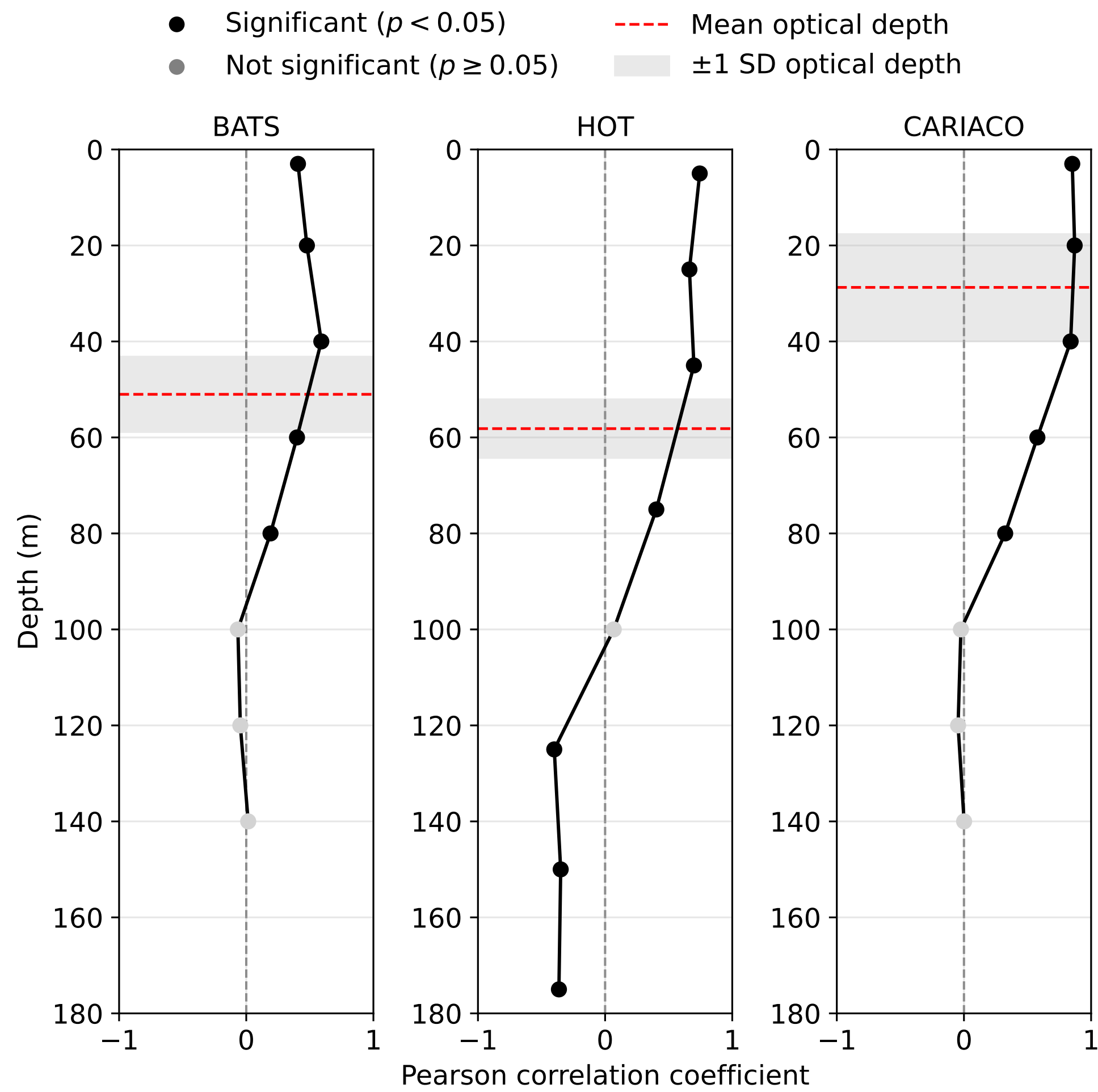


Figure. Depth-dependent Pearson correlation between satellite and in situ chlorophyll-*a*. Black symbols indicate $p < 0.05$ and gray symbols non-significant correlations. Red dashed lines show mean optical depth and gray bands show ± 1 standard deviation.

Station	z_{opt} [m]	Maximum <i>r</i>	Depth [m]
BATS	51.0 ± 7.8	0.590	40
HOT	58.2 ± 6.1	0.744	5
CARIACO	28.7 ± 11.1	0.871	20

Conclusions

- Satellite chlorophyll-*a* contains substantial information on the upper part of the vertical chlorophyll profile.
- The satellite-in situ relationship decreases strongly with depth.
- Reconstruction skill decreases toward deeper profile levels.

- Random Forest does not provide a systematic improvement over linear regression across all stations.
- Adding $K_d(490)$ does not consistently improve vertical-profile reconstruction.
- Pooled model skill can conceal large depth-specific reconstruction errors.

Profile reconstruction

Each sampling date was treated as one data object.

The complete eight-level vertical profile was the multi-output response:

$$\mathbf{y}_i = [\text{Chl}_{z_1}, \dots, \text{Chl}_{z_8}]^T.$$

Two predictor configurations were tested:

- satellite chlorophyll-*a*;
- satellite chlorophyll-*a* + $K_d(490)$.

Chronological partition:

60% training | 20% validation | 20% test.

Linear regression

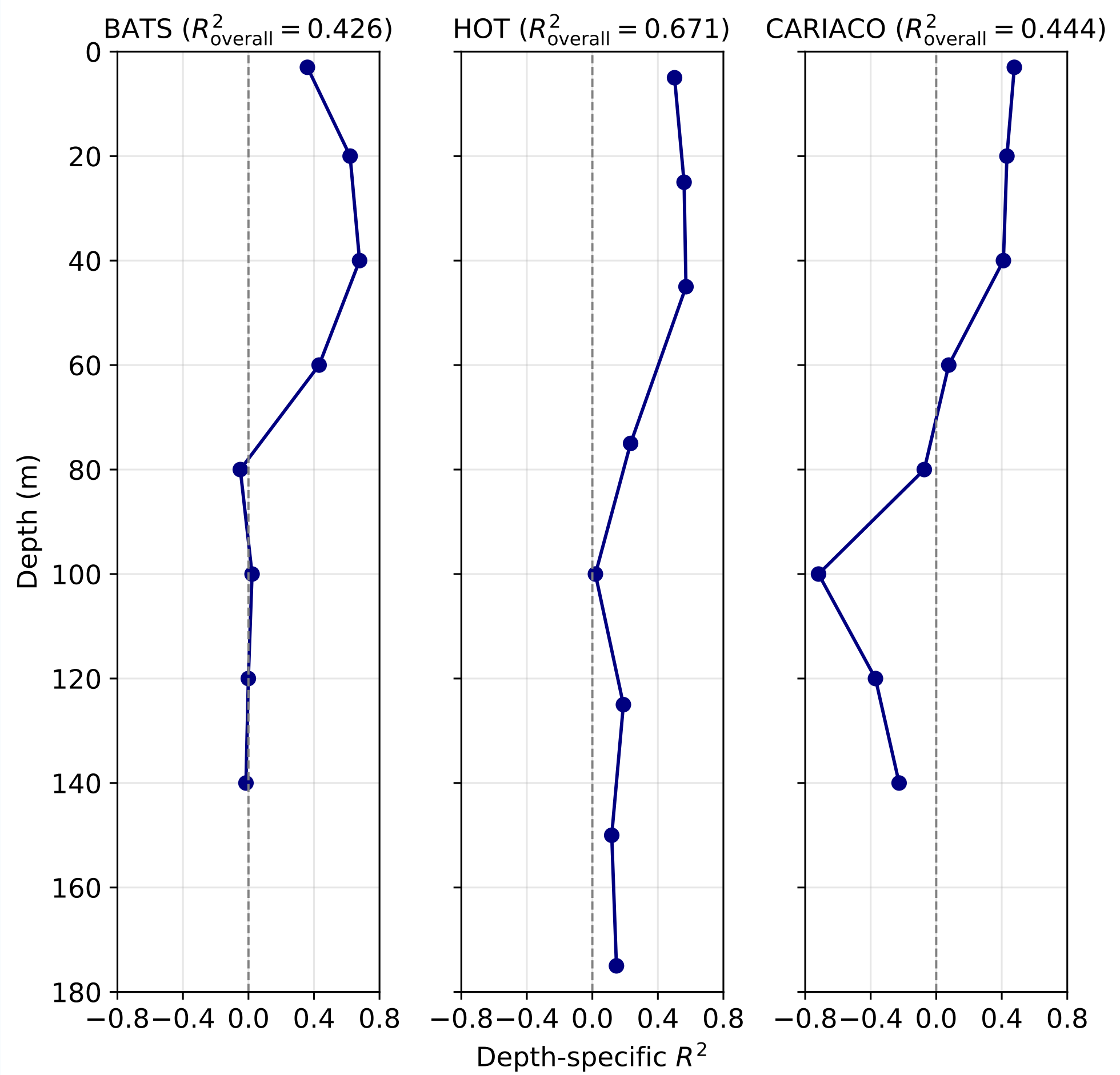


Figure. Depth-specific test-set coefficient of determination (R^2) for multi-output Ridge regression using satellite chlorophyll-*a*. Titles show pooled R^2 calculated across the complete vertical profile.

Random Forest

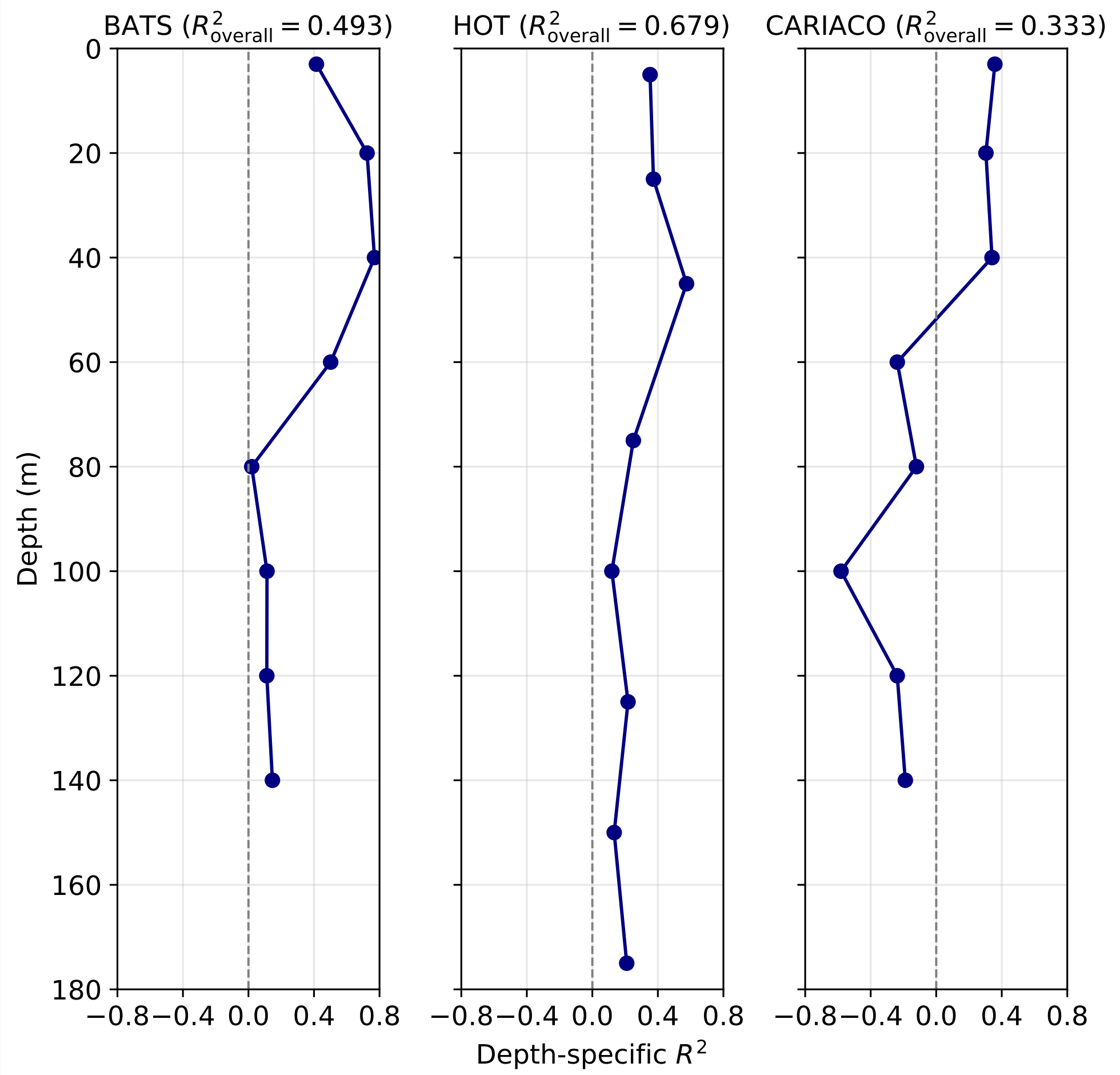


Figure. Depth-specific test-set coefficient of determination (R^2) for the Random Forest configuration selected using validation performance. BATS used satellite chlorophyll-*a*; HOT and CARIACO used satellite chlorophyll-*a* plus $K_d(490)$. Titles show pooled test-set R^2 .