

Satellite sea surface temperature shows stronger association with observed coral bleaching than in-situ loggers: Evidence from Japan’s national monitoring program

Hiroki Fukui

Independent researcher, Tokyo, Japan

Correspondence: [contact via marine-obs.org]

Draft v2.3 — 2026-02-02

Abstract

Accurate prediction of coral bleaching events is critical for reef management, yet the relative performance of satellite versus in-situ temperature measurements for this purpose remains poorly characterized. Using Japan’s national coral reef monitoring program (Monitoring Sites 1000), we compared the association between thermal indicators and observed bleaching for NOAA CoralTemp satellite SST versus in-situ temperature loggers across 44 sites over five years (2020–2024). Applying the weather forecast verification framework, we found that satellite SST showed substantially stronger association with bleaching observations than in-situ loggers (Equitable Threat Score: 0.99 vs 0.79; AUC: 0.88 vs 0.81). The optimal indicator was simply the number of days with SST $\geq 30^{\circ}\text{C}$, with a threshold of 5 days achieving maximum skill. This counterintuitive result—satellite outperforming in-situ—likely reflects unstandardized logger deployment depths (–2 to –8 m, undocumented) introducing heterogeneity that degrades statistical association. We emphasize that these results demonstrate retrospective association within a specific monitoring framework, not validated predictive skill or biological truth about thermal stress thresholds. Prospective validation on independent data is essential before operational deployment. Our findings suggest that (1) simple absolute temperature thresholds may be effective for regional bleaching monitoring, (2) the apparent superiority of satellite SST reflects its standardization rather than physiological relevance, and (3) future monitoring programs should document logger depths to enable proper validation of satellite products.

Keywords: coral bleaching, sea surface temperature, satellite remote sensing, in-situ monitoring, thermal stress, weather forecast verification, Japan

1. Introduction

Mass coral bleaching events have become increasingly frequent worldwide, with thermal stress recognized as the primary driver (Hughes et al. 2018). Satellite-derived sea surface temperature (SST) products, particularly NOAA’s Coral Reef Watch, have become the standard tool for monitoring thermal stress on coral

reefs globally (Liu et al. 2014). These products calculate Degree Heating Weeks (DHW), which accumulate thermal anomalies above a climatological baseline to predict bleaching risk.

This study treats coral bleaching monitoring not as an exercise in identifying physiological thresholds, but as a test of how different temperature products associate with documented bleaching events within a real-world monitoring framework. We do not claim to validate predictive skill for future events; rather, we ask a narrower question: given existing monitoring infrastructure and observational protocols, which data source shows stronger statistical association with recorded bleaching outcomes?

Satellite SST represents bulk temperature of the upper ocean layer, while corals experience temperatures at specific depths that may differ substantially from satellite estimates (Wyatt et al. 2023). In-situ temperature loggers deployed on reefs are often assumed to provide more accurate measures of the thermal environment experienced by corals. Yet direct comparisons of satellite versus in-situ temperature for bleaching association remain scarce, particularly in long-term monitoring contexts.

Japan’s Ministry of Environment operates the Monitoring Sites 1000 (Moni1000) program, which has conducted standardized coral reef surveys since 2008 across 26 sites spanning from Kushimoto (33°N, the northern limit of coral reefs in Japan) to Sekisei Lagoon (24°N). A subset of 44 survey locations have co-located temperature loggers, providing a unique opportunity to directly compare satellite and in-situ temperature as indicators of observed bleaching.

Here, we apply the weather forecast verification framework proposed by DeCarlo (2020) to systematically evaluate the association between thermal indicators and bleaching observations. This approach enables objective optimization of thresholds and direct comparison of different thermal stress indicators. We address two questions: (1) Does satellite SST or in-situ logger temperature show stronger association with observed bleaching within this monitoring framework? (2) What thermal threshold maximizes this association in Japanese waters?

2. Methods

2.1 Study sites and data sources

We analyzed data from 44 survey locations within the Moni1000 program that had co-located temperature loggers and bleaching observations during 2020–2024. Sites spanned four regions: Honshu/Kyushu (30–33°N), Amami Islands (27–28°N), Okinawa Main Island (26–27°N), and Yaeyama Islands (24–25°N). Bleaching data were obtained from annual visual surveys conducted by trained divers following standardized protocols established by Japan’s Biodiversity Center. For each survey location, bleaching presence was defined as any observed bleaching (>0% of colonies affected).

Satellite SST was obtained from NOAA CoralTemp (5 km resolution, daily), extracted for the coordinates of each survey location. In-situ temperature was recorded by loggers deployed at each site, with sampling intervals ranging from 10 minutes to 1 hour, subsequently aggregated to daily means. Logger deployment

depths ranged from −2 to −8 m according to program documentation, but specific depths for individual loggers were not recorded—a limitation central to this study’s findings.

2.2 Thermal stress indicators

We calculated the number of days during the summer season (June–September) when temperature exceeded 30°C, separately for satellite SST and in-situ loggers. The 30°C threshold was chosen based on: (1) its proximity to the mean monthly maximum (MMM) temperatures for Okinawan reefs (Sakai et al. 2019), (2) preliminary analyses showing superior performance compared to anomaly-based DHW metrics, and (3) consistency with Lachs et al. (2021) who demonstrated the effectiveness of MMM-based (rather than MMM+1°C) thresholds. We acknowledge that this threshold lacks direct experimental validation for the coral species present in our study area.

2.3 Statistical analysis

We applied the weather forecast verification framework (DeCarlo 2020) to evaluate association strength. For each threshold of days $\geq 30^\circ\text{C}$ (1–20 days), we constructed a contingency table of indicator versus observed bleaching events:

- Hits (H): Threshold exceeded and bleaching observed
- False Alarms (FA): Threshold exceeded but no bleaching observed
- Misses (M): Threshold not exceeded but bleaching observed
- Correct Negatives (CN): Threshold not exceeded and no bleaching observed

From these, we calculated:

- **Equitable Threat Score (ETS):** $\text{ETS} = H / (H + \text{FA} + M - H_{\text{random}})$, where $H_{\text{random}} = (H + \text{FA})(H + M) / n$. ETS rewards correct classification while accounting for chance, ranging from −1/3 to 1 (perfect).
- **Bias:** $(H + \text{FA}) / (H + M)$. Values >1 indicate over-prediction, <1 under-prediction.
- **Probability of Detection (POD):** $H / (H + M)$. Equivalent to sensitivity.
- **False Alarm Ratio (FAR):** $\text{FA} / (H + \text{FA})$.
- **Area Under the ROC Curve (AUC):** Calculated across all thresholds.

The optimal threshold was defined as that maximizing ETS. **Critically, all thresholds were optimized on the same dataset used for evaluation, without cross-validation or hold-out testing.** Results therefore represent retrospective association strength, not validated predictive skill.

3. Results

3.1 Overall association strength

A total of 204 site-year observations were available for satellite SST analysis, of which 182 had paired in-situ logger data. Bleaching was observed in 76 of 204 observations (37%) for the full dataset and 65 of 182 (36%) for the paired dataset.

For satellite SST (CoralTemp), the optimal threshold was 5 days $\geq 30^{\circ}\text{C}$, achieving ETS = 0.84, POD = 0.74, FAR = 0.24, and Bias = 0.97 (Table 1). AUC was 0.81 for the full dataset.

Table 1. Association metrics for CoralTemp SST by threshold (n=204)

Threshold (days $\geq 30^{\circ}\text{C}$)	ETS	Bias	POD	FAR	H	FA	M	CN
1	0.803	1.18	0.78	0.34	59	31	17	97
3	0.806	1.13	0.76	0.33	58	28	18	100
5	0.843	0.97	0.74	0.24	56	18	20	110
7	0.742	0.87	0.66	0.24	50	16	26	112
10	0.548	0.66	0.50	0.24	38	12	38	116
14	0.438	0.54	0.41	0.24	31	10	45	118

Notes: ETS = Equitable Threat Score; POD = Probability of Detection; FAR = False Alarm Ratio. Bold indicates optimal threshold. All metrics represent retrospective association, not validated prediction.

3.2 Satellite versus in-situ comparison

When compared on the same 182 observations with both satellite and logger data, satellite SST showed substantially stronger association with bleaching observations across all metrics (Table 2, Figure 1). At the optimal threshold of 5 days, satellite ETS was 0.99 compared to 0.79 for loggers.

Table 2. Comparison of CoralTemp SST and in-situ logger at 5-day threshold (n=182)

Data source	ETS	AUC	POD	FAR	Bias	H	FA	M	CN
CoralTemp SST	0.990	0.878	0.86	0.24	0.97	56	18	9	99
In-situ logger	0.791	0.805	0.77	0.33	0.87	50	25	15	92
Difference	+0.199	+0.073	+0.09	-0.09	+0.10	+6	-7	-6	+7

Notes: Comparison performed on 182 site-year observations with both satellite and logger data. Bleaching observed in 65/182 (36%). All metrics represent retrospective association.

The stronger satellite association was driven by fewer false alarms (18 vs 25) and more hits (56 vs 50) compared to loggers.

3.3 Comparison with prior studies

Our maximum ETS of 0.84–0.99 substantially exceeds the global optimum of 0.218 reported by DeCarlo (2020) using DHW as a predictor across 100 sites worldwide (Table 3). This difference likely reflects the regional specificity of our analysis, the use of absolute temperature thresholds rather than anomaly-based metrics, and potential overfitting to this particular dataset.

Table 3. Comparison with DeCarlo (2020) global analysis

Study	Scope	Indicator	Optimal threshold	Maximum ETS
DeCarlo (2020)	Global, 100 sites	DHW	3.5 °C-weeks	0.218
This study	Japan, 44 sites	Days $\geq 30^{\circ}\text{C}$	5 days	0.84–0.99*

*Note: Our ETS values were obtained without cross-validation and likely overestimate true predictive skill.

4. Discussion

4.1 Why does satellite show stronger association than in-situ?

The counterintuitive finding that satellite SST shows stronger association with bleaching than in-situ loggers demands explanation. We propose three non-mutually exclusive hypotheses:

Hypothesis 1: Unstandardized logger depths. The Moni1000 program documents logger depths as ranging from -2 to -8 m, but specific depths for individual loggers are not recorded. Temperature decreases with depth due to solar heating of surface waters, particularly during calm summer conditions when bleaching risk is highest. A logger at -8 m may record substantially lower temperatures than one at -2 m during the same heat event. This depth variability introduces heterogeneity into bleaching associations based on logger data, while satellite SST consistently measures near-surface temperature. In such a setting, in-situ loggers do not constitute ground truth, but an unmodeled mixture of depth-dependent thermal regimes.

Hypothesis 2: Surface heat input as the driver. Coral bleaching may be driven more by cumulative heat input from the surface than by the absolute temperature at coral depth. Satellite SST directly measures this surface thermal forcing, while subsurface loggers capture a buffered signal. This interpretation aligns with Skirving et al. (2019), who emphasized the importance of surface heat flux for bleaching. However, this hypothesis remains speculative and requires targeted experimental validation.

Hypothesis 3: Spatial averaging. Satellite SST represents a 5 km spatial average, potentially providing a more representative measure of regional thermal stress than point measurements from individual loggers. Micro-scale variability in logger temperatures may not correlate with reef-wide bleaching patterns.

4.2 Critical limitations and epistemological caveats

This study has fundamental limitations that preclude interpretation as validated prediction.

First, all thresholds were optimized on the same dataset used for evaluation, without cross-validation, hold-out testing, or temporal splitting. The exceptionally high ETS values (up to 0.99) almost certainly reflect overfitting to this particular dataset. We therefore present these results as evidence of retrospective association, not predictive skill. Leave-one-year-out cross-validation or prospective testing on 2025 data would be essential to estimate true predictive performance.

Second, bleaching observations in Moni1000 are institutionally mediated outcomes—annual, visual, binary assessments by trained but potentially variable observers—that may not directly reflect the cumulative physiological damage corals experience. The apparent superiority of satellite SST thus reflects its stability and standardization as an indicator within this specific observational framework, not necessarily its physiological relevance.

Third, the 30°C threshold, while operationally effective in this dataset, lacks direct experimental validation for the coral species and thermal histories present in our study area. Regional thermal adaptation, species composition, and historical exposure all influence bleaching susceptibility in ways our simple threshold cannot capture.

Fourth, our comparison is fundamentally asymmetric: satellite SST is a standardized, validated product, while logger data suffer from undocumented deployment heterogeneity. **The satellite “advantage” may be entirely attributable to this design flaw in the monitoring program, not to any inherent superiority of satellite measurements.**

4.3 Implications for monitoring program design

Despite these limitations, our results have practical implications. The lack of documented logger depths in Moni1000—a well-resourced national program—suggests this issue may be widespread. We recommend that monitoring programs either (1) standardize logger deployment at a fixed depth (e.g., −3 m), or (2) explicitly record deployment depth for each logger to enable depth-stratified analyses.

For operational early warning purposes, if prospective validation confirms the utility of simple temperature thresholds, a rule such as “ ≥ 5 days above 30°C” could provide reef managers with an actionable alert requiring minimal computational resources. However, we emphasize that such operational use should await independent validation.

4.4 The value of field monitoring

Our finding should not be interpreted as an argument against field monitoring. In-situ measurements remain essential for understanding depth-dependent thermal regimes, light environments, water flow, and local stressors that satellites cannot capture. The value of in-situ data for mechanistic understanding is not diminished

by its reduced statistical association in this retrospective analysis. Rather, we suggest that the current monitoring design—with undocumented logger depths—inadvertently undermines the validation role that in-situ data are intended to serve.

5. Conclusions

Using Japan’s national coral reef monitoring data, we demonstrate that satellite SST shows stronger retrospective association with observed bleaching events than in-situ temperature loggers within this monitoring framework. This counterintuitive result likely reflects unstandardized logger deployment depths, highlighting an often-overlooked issue in coral reef monitoring design. We recommend that monitoring programs standardize or document logger depths, and suggest that simple absolute temperature thresholds warrant prospective validation for operational bleaching monitoring—while recognizing that retrospective association should not be conflated with predictive skill or biological truth.

Data availability

Moni1000 data were provided by the Biodiversity Center of Japan, Ministry of Environment (approval no. 2601271). Satellite SST data are publicly available from NOAA Coral Reef Watch.

Acknowledgments

We thank the Ministry of Environment and the Biodiversity Center of Japan for providing Moni1000 data and for their sustained commitment to coral reef monitoring in Japan. This work was conducted independently without external funding.

Author contributions

HF conceived the study, analyzed data, and wrote the manuscript.

Competing interests

The author declares no competing interests.

References

- DeCarlo TM (2020) Treating coral bleaching as weather: a framework to validate and optimize prediction skill. *PeerJ* 8:e9449.
- Hughes TP, Kerry JT, Baird AH, et al. (2018) Global warming transforms coral reef assemblages. *Nature* 556:492–496.
- Lachs L, Skirving WJ, et al. (2021) Emergent increase in coral thermal tolerance reduces mass bleaching under climate change. *Nature Communications* 12:1–12.
- Liu G, Heron SF, Eakin CM, et al. (2014) Reef-scale thermal stress monitoring of coral ecosystems: new 5-km global products from NOAA Coral Reef Watch. *Remote Sensing* 6:11579–11606.
- Sakai K, Singh T, Iguchi A, et al. (2019) Bleaching and mortality of a photosymbiotic bioeroding sponge under future carbon dioxide emission scenarios. *Coral Reefs* 38:1–12.
- Skirving WJ, Heron SF, Marsh BL, et al. (2019) The relentless march of mass coral bleaching: a global perspective of changing heat stress. *Coral Reefs* 38:547–557.
- Wyatt ASJ, Leichter JJ, Toth LT, et al. (2023) Hidden heatwaves and severe coral bleaching linked to mesoscale eddies and thermocline dynamics. *Nature Communications* 14:25.
-

Figure Legends

Figure 1. Association strength comparison. (A) Equitable Threat Score (ETS) by threshold for CoralTemp SST (blue circles, n=204) and in-situ loggers (orange squares, n=182). Dashed green line indicates the global optimum ETS of 0.218 from DeCarlo (2020). Vertical dotted line marks the optimal threshold of 5 days. (B) Comparison of CoralTemp SST and in-situ logger metrics at the 5-day threshold (n=182), showing ETS, Probability of Detection (POD), and 1–False Alarm Ratio. Note: All metrics represent retrospective association without cross-validation.

Word count: ~3,600 (excluding tables, figures, and references)