Satellite sea surface temperature outperforms in-situ loggers for coral bleaching prediction: Evidence from Japan's national monitoring program

Hiroki Fukui

Independent researcher, Tokyo, Japan

Correspondence: [contact via marine-obs.org]

Draft v2.1 — 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 bleaching prediction skill of NOAA CoralTemp satellite SST against in-situ temperature loggers across 44 sites over five years (2020–2024). Applying the weather forecast verification framework, we found that satellite SST substantially outperformed in-situ loggers (Equitable Threat Score: 0.99 vs 0.79; AUC: 0.88 vs 0.81). The optimal predictor was simply the number of days with SST ≥30°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 into predictions. We emphasize that these results demonstrate operational predictability within a specific monitoring framework, not biological truth about thermal stress. Our findings suggest that (1) absolute temperature thresholds may be operationally effective for regional bleaching prediction, (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.

However, 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 prediction 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 predictors of observed bleaching.

Here, we apply the weather forecast verification framework proposed by DeCarlo (2020) to systematically evaluate bleaching prediction skill. This approach enables objective optimization of prediction thresholds and direct comparison of different thermal stress indicators. We address two questions: (1) Does satellite SST or in-situ logger temperature better predict observed bleaching within this monitoring framework? (2) What is the optimal thermal threshold for operational bleaching prediction 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. 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.

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. This absolute threshold approach was chosen based on preliminary analyses showing superior performance compared to anomaly-based DHW metrics, consistent with Lachs et al. (2021) who demonstrated the effectiveness of MMM-based (rather than MMM+1°C) thresholds.

2.3 Statistical analysis

We applied the weather forecast verification framework (DeCarlo 2020) to evaluate prediction skill. For each threshold of days ≥30°C (1–20 days), we constructed a contingency table of predicted versus observed bleaching events:

- Hits (H): Bleaching predicted and observed
- False Alarms (FA): Bleaching predicted but not observed

- Misses (M): Bleaching not predicted but observed
- Correct Negatives (CN): Bleaching not predicted and not observed

From these, we calculated:

- **Equitable Threat Score (ETS)**: ETS = H / (H + FA + M − H_random), where H_random = (H + FA)(H + M) / n. ETS rewards correct predictions while accounting for chance, ranging from −1/3 to 1 (perfect).
- **Bias**: (H + 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)**: FA / (H + FA).
- **Area Under the ROC Curve (AUC)**: Calculated across all thresholds.

The optimal threshold was defined as that maximizing ETS, following standard practice in weather forecast verification.

---

3. Results

3.1 Overall prediction skill

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 ≥30°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. Prediction skill metrics for CoralTemp SST by threshold (n=204)

Threshold (days ≥30°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.

3.2 Satellite versus in-situ comparison

When compared on the same 182 observations with both satellite and logger data, satellite SST substantially outperformed in-situ loggers 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%).

The superior satellite performance 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 model and the use of absolute temperature thresholds rather than anomaly-based metrics.

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 ≥30°C	5 days	0.84–0.99

4. Discussion

4.1 Why does satellite outperform in-situ?

The counterintuitive finding that satellite SST outperforms in-situ loggers for bleaching prediction 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 predictions 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 Epistemological caveats

We emphasize that this study evaluates operational predictability under real-world monitoring constraints, not biological truth about thermal stress thresholds. Bleaching observations in Moni1000 are institutionally mediated outcomes—annual, visual, binary—that may not directly reflect the cumulative physiological damage corals experience. The apparent superiority of satellite SST thus reflects its stability and standardization as a predictor within this specific observational framework, not necessarily its physiological relevance.

The exceptionally high skill scores (ETS up to 0.99) warrant particular caution. These values were obtained by threshold optimization on the same dataset used for evaluation, without cross-validation or out-of-sample testing. We therefore present these results as evidence of strong association within this dataset rather than validated predictive skill for future events. Prospective testing of the 5-day threshold on independent data remains essential before operational deployment.

4.3 Implications for monitoring program design

Our results have practical implications for coral reef monitoring programs. First, 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.

Second, for operational early warning purposes, a simple threshold of "≥5 days above 30°C" could provide reef managers with an actionable alert requiring minimal computational resources. When CoralTemp SST exceeds 30°C for 5 consecutive days during summer, managers could initiate surveillance dives or prepare intervention measures.

4.4 Limitations and the value of field monitoring

Several limitations should be noted. First, our analysis covers only five years (2020–2024), a period that included both non-bleaching years (2021) and severe bleaching years (2022, 2024). Longer time series would enable more robust threshold validation. Second, we analyzed presence/absence of bleaching without considering severity. Third, our results are specific to Japanese reefs spanning 24–33°N and may not generalize to tropical reef systems with different thermal regimes.

Critically, 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 predictive skill in this operational context. 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 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 (days ≥30°C) may provide operationally effective bleaching prediction for management applications—while recognizing that operational utility should not be conflated with biological truth.

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.

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. Prediction skill 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 performance at the 5-day threshold (n=182), showing ETS, Probability of Detection (POD), and 1−False Alarm Ratio.

Word count: ~3,200 (excluding tables, figures, and references)