

# Uncertainty-aware SST Anomaly Monitoring

NOAA OISST → Daily anomalies → Kalman RTS smoothing → diagnostics

Boris Baudel

February 10, 2026

**Goal.** Build an end-to-end pipeline for regional SST anomaly monitoring using NOAA OISST daily fields: (i) regional mean aggregation, (ii) daily climatology baseline (1991–2020), (iii) anomaly computation, (iv) uncertainty-aware signal extraction with a local-level Kalman filter + RTS smoother, (v) trend/extreme metrics and event detection, (vi) reproducible plots and machine-readable outputs (JSON/CSV).

## 1 Data and Region of Interest

### 1.1 NOAA OISST

We use NOAA Optimum Interpolation Sea Surface Temperature (OISST v2.x) daily SST fields (gridded satellite + in situ). Inputs in this project:

- **Daily observations:** `sst.day.mean.YYYY.nc`
- **Daily climatology (baseline 1991–2020):** `sst.day.ltm.1991-2020.nc`

### 1.2 Spatial domain

The region is a latitude/longitude box:

$$\lambda \in [\lambda_{\min}, \lambda_{\max}], \quad \phi \in [\phi_{\min}, \phi_{\max}],$$

e.g. Mediterranean example:  $\lambda \in [-6^\circ, 6^\circ]$ ,  $\phi \in [34^\circ, 44^\circ]$ .

**Longitude convention.** Some OISST products use longitudes in  $[0, 360)$ ; for regional extraction it is often convenient to convert to  $[-180, 180)$ :

$$\lambda_{180} = ((\lambda + 180) \bmod 360) - 180.$$

This ensures standard slicing over negative longitudes.

## 2 Regional Mean SST

Let  $T(t, \phi_i, \lambda_j)$  denote SST at day  $t$  on grid point  $(i, j)$ . We compute a latitude-weighted regional mean (area weight proportional to  $\cos \phi$ ):

$$\bar{T}(t) = \frac{\sum_{i,j} T(t, \phi_i, \lambda_j) w(\phi_i)}{\sum_{i,j} w(\phi_i)}, \quad w(\phi) = \cos(\phi).$$

This accounts for the decrease of grid-cell area with latitude on a regular lat/lon grid.

### 3 Daily Climatology and Anomalies

#### 3.1 Daily climatology

Define  $d(t) \in \{1, \dots, 365\}$  (or 366) as the day-of-year index. From the baseline period (1991–2020), we obtain a daily climatology:

$$T_{\text{clim}}(d) = \mathbb{E}[\bar{T}(t) \mid d(t) = d, t \in \text{baseline}].$$

In practice, the provided LTM file encodes  $T_{\text{clim}}(d)$  directly; we subset to the region and compute the same weighted mean as above, then re-index by  $d$ .

**Leap day handling.** If the climatology provides 365 values but the observation window contains Feb 29, define

$$T_{\text{clim}}(366) := T_{\text{clim}}(365)$$

(or any consistent convention), enabling anomaly computation for leap years.

#### 3.2 Anomalies

The anomaly time series is

$$A(t) = \bar{T}(t) - T_{\text{clim}}(d(t)).$$

This removes the seasonal cycle and highlights interannual/intraseasonal variability.

## 4 State-Space Model (Local Level) and Kalman Filtering

#### 4.1 Model

We model anomalies  $y_t = A(t)$  as a local-level (random walk) state-space model:

$$x_t = x_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, q), \quad (1)$$

$$y_t = x_t + v_t, \quad v_t \sim \mathcal{N}(0, r), \quad (2)$$

where:

- $x_t$  is the latent (smooth) climate signal,
- $q$  is the process noise variance (signal variability),
- $r$  is the measurement noise variance (observation + unresolved variability).

This is a special case of linear Gaussian SSM:

$$x_t = Fx_{t-1} + w_t, \quad y_t = Hx_t + v_t, \quad F = 1, H = 1, Q = q, R = r.$$

#### 4.2 Kalman filter (prediction + update)

Define:

- $\hat{x}_{t|t-1}$  and  $P_{t|t-1}$ : predicted mean/variance of  $x_t$  given observations up to  $t-1$ ,
- $\hat{x}_{t|t}$  and  $P_{t|t}$ : filtered mean/variance given observations up to  $t$ .

**Initialization.** Choose  $(\hat{x}_{0|0}, P_{0|0})$ . A standard weakly-informative choice is  $\hat{x}_{0|0} = y_0$  and  $P_{0|0}$  large.

**Prediction step.** From eq. (1),

$$\hat{x}_{t|t-1} = F\hat{x}_{t-1|t-1} = \hat{x}_{t-1|t-1}, \quad (3)$$

$$P_{t|t-1} = FP_{t-1|t-1}F^\top + Q = P_{t-1|t-1} + q. \quad (4)$$

**Innovation and Kalman gain.**

$$\nu_t = y_t - H\hat{x}_{t|t-1} = y_t - \hat{x}_{t|t-1}, \quad (5)$$

$$S_t = HP_{t|t-1}H^\top + R = P_{t|t-1} + r, \quad (6)$$

$$K_t = P_{t|t-1}H^\top S_t^{-1} = \frac{P_{t|t-1}}{P_{t|t-1} + r}. \quad (7)$$

**Update step.**

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t\nu_t, \quad (8)$$

$$P_{t|t} = (I - K_tH)P_{t|t-1} = (1 - K_t)P_{t|t-1}. \quad (9)$$

### 4.3 Riccati recursion (covariance dynamics)

The covariance update above is the scalar form of the discrete-time Riccati equation for the Kalman filter. In scalar local-level form:

$$P_{t|t} = \left(1 - \frac{P_{t|t-1}}{P_{t|t-1} + r}\right) P_{t|t-1} = \frac{r P_{t|t-1}}{P_{t|t-1} + r}.$$

Combined with prediction  $P_{t|t-1} = P_{t-1|t-1} + q$ , we get the closed recursion

$$P_{t|t} = \frac{r(P_{t-1|t-1} + q)}{(P_{t-1|t-1} + q) + r}$$

which is a Riccati-type nonlinear recursion.

**Steady-state solution (optional).** If  $(q, r)$  are constant and the system reaches steady state  $P_{t|t} = P$ , then with  $P^- = P + q$ :

$$P = \frac{r(P + q)}{P + q + r}.$$

Rearranging yields a quadratic equation in  $P$ :

$$\begin{aligned} P(P + q + r) &= r(P + q) \implies P^2 + Pq + Pr = rP + rq \\ \implies P^2 + Pq + P(r - r) - rq &= 0 \implies \boxed{P^2 + qP - rq = 0}. \end{aligned}$$

Thus

$$P = \frac{-q + \sqrt{q^2 + 4rq}}{2} \quad (\text{positive root})$$

and the steady-state gain becomes  $K = \frac{P+q}{P+q+r}$ .

## 5 RTS Smoother (Backward pass)

Filtering gives  $\hat{x}_{t|t}$  using data up to  $t$ . Smoothing uses all data 1: $T$  and improves estimates:

$$\hat{x}_{t|T} = \mathbb{E}[x_t | y_{1:T}], \quad P_{t|T} = \text{Var}(x_t | y_{1:T}).$$

For linear Gaussian models, the Rauch–Tung–Striebel smoother recursion is:

## 5.1 Smoother gain

Define predicted covariance  $P_{t+1|t} = P_{t|t} + q$ . The smoother gain is

$$G_t = P_{t|t} F^\top (P_{t+1|t})^{-1} = \frac{P_{t|t}}{P_{t|t} + q}$$

(using  $F = 1$ ).

## 5.2 Backward recursions

For  $t = T - 1, \dots, 0$ :

$$\hat{x}_{t|T} = \hat{x}_{t|t} + G_t (\hat{x}_{t+1|T} - \hat{x}_{t+1|t}) \quad (10)$$

$$P_{t|T} = P_{t|t} + G_t^2 (P_{t+1|T} - P_{t+1|t}) \quad (11)$$

In scalar form,  $G_t^2$  appears because  $G_t P G_t^\top = G_t^2 P$ .

**Uncertainty band.** The report plots  $\pm 2\sigma_t$  with  $\sigma_t = \sqrt{P_{t|T}}$  (smoothed).

## 6 Parameter Estimation for $(q, r)$

In practice  $q$  and  $r$  can be:

- user-specified (hyperparameters),
- estimated via robust variance heuristics (e.g., based on differences and residuals).

A common heuristic for a random-walk + noise model uses the variance of first differences  $\Delta y_t = y_t - y_{t-1}$ :

$$\text{Var}(\Delta y_t) \approx q + 2r,$$

and additional robust measures can split the contribution between  $q$  and  $r$  (implementation-specific).

## 7 Trend and Diagnostics

### 7.1 Linear trend per decade

Given timestamps  $t$  (in days) we fit:

$$y(t) = \alpha + \beta t + \varepsilon(t),$$

then convert to per-decade:

$$\text{trend}_{10y} = \beta \cdot (365.25 \times 10).$$

Trends are computed on both raw anomalies and smoothed signal.

### 7.2 Extreme events: P90 + consecutive days

Compute threshold  $\theta$  as the 90th percentile of  $y_t$  over the analysis window:

$$\theta = \text{quantile}_{0.90}(\{y_t\}_{t=1}^T).$$

Define an event as any contiguous segment where  $y_t > \theta$  for at least  $k$  days (e.g.  $k = 3$ ). Each event is summarized by: start/end dates, duration, peak anomaly, mean anomaly.

## 8 Algorithm (End-to-end Pipeline)

---

**Algorithm 1** SST anomaly monitoring pipeline (regional, uncertainty-aware)

---

**Require:** Region bounds  $(\lambda_{\min}, \lambda_{\max}, \phi_{\min}, \phi_{\max})$ , window  $[t_s, t_e]$ , climatology file, OISST yearly files (offline cache)

**Ensure:** Anomaly series  $y_t$ , smoothed signal  $\hat{x}_{t|T}$ , uncertainty  $\sigma_t$ , diagnostics, plots, JSON/CSV outputs

- 1: Load yearly OISST daily SST, subset region/time, compute weighted mean  $\bar{T}(t)$
  - 2: Load climatology file, subset region, compute weighted mean  $T_{\text{clim}}(d)$  (index by day-of-year)
  - 3: Compute anomalies  $y_t = \bar{T}(t) - T_{\text{clim}}(d(t))$
  - 4: Estimate  $(q, r)$  (robust heuristics) unless provided
  - 5: Kalman filter: compute  $\hat{x}_{t|t}, P_{t|t}$  for  $t = 1:T$
  - 6: RTS smoother: compute  $\hat{x}_{t|T}, P_{t|T}$  backward for  $t = T - 1:1$
  - 7: Trend diagnostics on  $y_t$  and  $\hat{x}_{t|T}$
  - 8: Extreme detection using P90 threshold and consecutive-day rule
  - 9: Save `summary.json`, `events.csv`, plots
- 

## 9 Results (2019–2020) and Figures

### 9.1 Time series: raw anomaly vs RTS smooth (+ uncertainty)

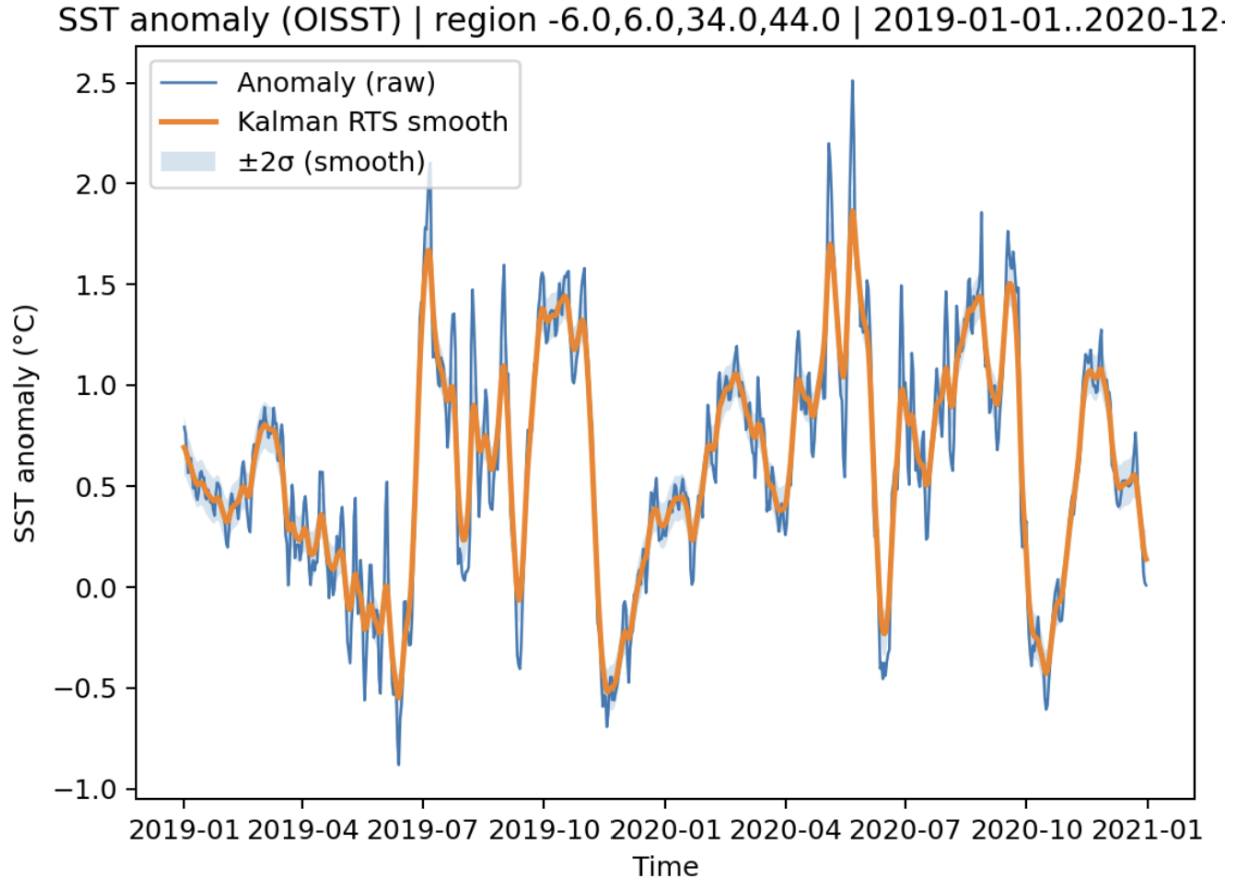


Figure 1: Regional SST anomaly (raw) and Kalman RTS smoothed estimate with  $\pm 2\sigma$  uncertainty band.

## 9.2 Time series with rolling mean and P90 threshold

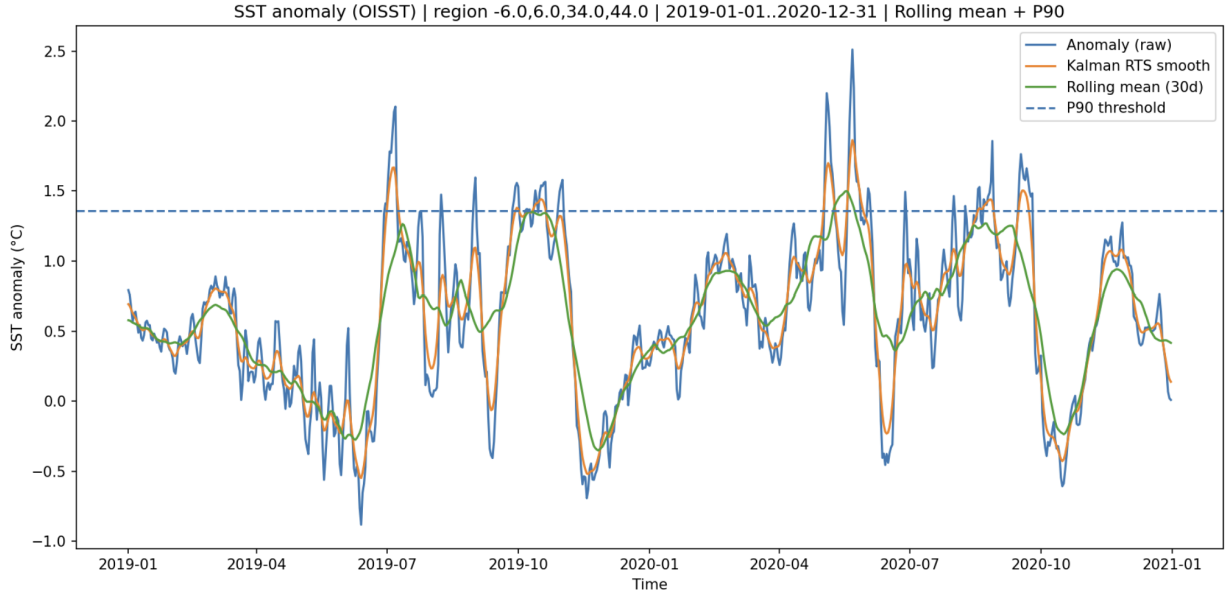


Figure 2: Regional SST anomalies with Kalman smoothing, 30-day rolling mean, and P90 warm-event threshold.

## 9.3 Distributions

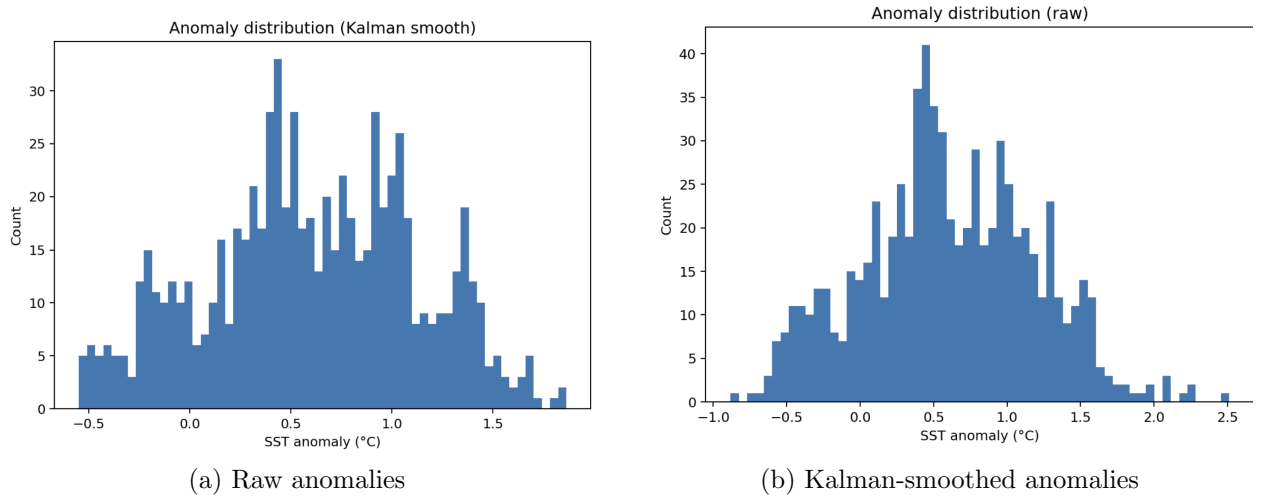


Figure 3: Anomaly distributions over the analysis window.

## 10 Reproducibility and Outputs

The pipeline writes:

- **summary.json**: region, time window, baseline label,  $(q, r)$ , trends, metrics.
- **events.csv**: event catalogue (start/end, duration, peak, mean, threshold).
- **Figures**: time series and histograms (PNG).