Appendix:

In this appendix, we summarize the technique details about cumulative sum, structure change model, Bayesian CPA and EARS algorithm. In-depth derivations can be found in their original methodology papers [10, 20-22].

1. Cumulative Sum

For time-series data Y_i with *i=1, …, N*, the mean-shift model is written as

$$Y_i = \mu + \varepsilon_i ,$$

where μ is the sample average as $\mu = \sum_i Y_i/N$ and ε_i is the residual term as $\varepsilon_i = Y_i - \mu$ for the i^{th} observation. The cumulative sums of residuals are calculated as $S_i = S_{i-1} + \varepsilon_i$ for *i=1, …, N* where $S_0 = 0$. The change point at location *m* is detected through searching for the maximum absolute CUSUM of residuals where $|S_m| = \max_{i=0,\ldots,N} |S_i|$. The time-series data is split into two segments on each side of the change point, and the analysis is repeated for each segment.

2. Structure Change Model

A multiple linear regression model with *m* change points (*m+1* regimes) is considered as

$$y_i = x_i'\beta + z_i'\delta_j + \varepsilon_i$$

for $j = 1, \ldots, m+1$ and $i = T_{j-1} + 1, \ldots, T_j$. In this model, y_i is the observed dependent variable and x_i, z_i are covariate vectors with their corresponding

coefficient β and $\delta_j (j = 1, \dots, m+1)$, respectively. The error term ε_i is the disturbance at time *i*. The change points $(T_1, \dots, T_m)$ are unknown with the exception for $T_0 = 0$ for the beginning of time series and $T_{m+1} = N$ for the last observation. The length of time series data is *N*. The objective is to estimate the model coefficient and location of *m* break points.

The estimated change points $(\hat{T}_1, \hat{T}_2, \dots, \hat{T}_m)$ are obtained by minimizing the sum of squared residuals $S_T(T_1, T_2, \dots, T_m)$ as

$$(\hat{T}_1, \hat{T}_2, \dots, \hat{T}_m) = argmin_{T_1, T_2, \dots, T_m} S_T(T_1, T_2, \dots, T_m),$$

where

$$S_T(T_1, T_2, \dots, T_m) = \sum_{j=1}^{m+1} \sum_{i=T_{i-1}+1}^{i=T_i} (y_i - x_i'\beta - z_i'\delta_i)^2$$

In this paper, we didn't consider covariates in the structure change model since the information about covariates were not available in the data. Therefore, the sum of squared residuals is simplified as

$$S_T(T_1, T_2, \dots, T_m) = \sum_{j=1}^{m+1} \sum_{i=T_{i-1}+1}^{i=T_i} (y_i - \bar{y}_i)^2$$

where $\bar{y}_i$ is the sample mean in the i^{th} regime as

$$\bar{y}_i = \frac{1}{T_i - T_{i-1}} \sum_{i=T_{i-1}+1}^{i=T_i} y_i$$

3. Bayesian CPA

In the Bayesian CPA, the observations are assumed independently normal distributed with $Y_i \sim N(\mu_i, \sigma^2),\ \ i = 1, \ldots, N$. A partition $\rho = (U_1, U_2, \ldots, U_N)$ is used in the algorithm where $U_i = 1$ indicates a change point at position *i+1* while $U_i = 0$ for no change. Initially all U_i are set to 0 except $U_N = 1$. Based on Bayes rule, the posterior estimate for change point partition ρ can be obtained as

$$f(\rho|Y) \propto f(Y|\rho) f(\rho).$$

The likelihood can be derived from normal distribution as

$$f(Y|\rho) \propto \int_0^\lambda \frac{\varpi^{(b-1)/2}}{[W + B\varpi]^{(N-1)/2}} d\varpi,$$

where *b* is the number of blocks in partition ρ and *W* and *B* are the within and between blocks sums of squares residuals.

The prior for the partition ρ is the integration of Bernoulli distribution at each location as

$$f(\rho) = \frac{1}{p_0} \int_0^{p_0} f[\rho|p] dp = \frac{1}{p_0} \int_0^{p_0} p^{b-1} (1-p)^{N-b} dp$$

Markov chain Monte Carlo (MCMC) samples are drawn to estimate the posterior distribution of change points and their means.

4. Early Aberration Reporting System (EARS)

CDC selected two widely used methods to be included in EARS: the historical limits method and the seasonally adjusted CUSUM method. In the historical limits method, the current reported cases in a 4-week interval (x_0) is compared with the historical mean and an alert is triggered if

$$\frac{x_0}{\mu} > 1 + 2 * \sigma_x/\mu$$

where μ is the mean of 15 totals of 4-week intervals, and σ_x is the standard deviation of these 15 historical incidence data values.

The seasonally adjusted CUSUM method is based on the formula

$$S_t = \max(0, S_{t-1} + (X_t - (\mu_0 + k\sigma_{xt})/\sigma_{xt}))$$

where an anomaly is detected if $S_t \geq 0.5$ for counts of 5 or more. In the formula, S_t is the current CUSUM while S_{t-1} is the previous CUSUM. The current count is denoted as X_t and the historical mean and standard deviation are μ_0 and σ_{xt}. k is the detectable shift from the mean which has to be pre-specified.

In this paper, we use the modified EARS C2 algorithm which is a historical limits method and the past 28 days after 2 days lag are used as historical data to compare with the current daily count.