Exponentially Weighted Moving Average Change Detection (EWMACD)

Notation and definitions: For an $m \times n$ matrix A, an n-vector x, $I \subset \{1, \ldots, m\}$, $J \subset \{1, \ldots, n\}$, let A_{IJ} denote the submatrix of A formed from the rows indexed by I and the columns indexed by J, and x_J denote the subvector of x indexed by J. $A_{I\cdot}$ ($A_{\cdot J}$) are the rows (columns) of A indexed by I (J), respectively. An image is an $R \times C$ matrix D, where each D_{rc} (pixel) is an $S \times B$ matrix, whose (s, b) element $(D_{rc})_{sb}$ is the signal value at time index s and frequency band index b.

Algorithm EWMACD.

for band $b = 1 : B$
  for row $r = 1 : R$
    for col $c = 1 : C$ **do**

Step 1: Write the time series data in the column $(D_{rc})_{\cdot b}$ as

$$(D_{rc})_{\cdot b} = \begin{pmatrix} u \\ v \end{pmatrix},$$

where the M-dimensional vector u is deemed training data and the $(S - M)$-dimensional vector v as the test data. Let

$$X = \begin{pmatrix} 1 & \sin t_1 & \cos t_1 & \cdots & \sin Kt_1 & \cos Kt_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \sin t_M & \cos t_M & \cdots & \sin Kt_M & \cos Kt_M \end{pmatrix}$$

be the Gram matrix for the time points $t_1, \ldots, t_M$, using K harmonics, where $M > 2K + 1$. The least squares fit to the training data u is then written as

$$u(t) = \alpha_0 + \sum_{i=1}^{K} (\alpha_{2i-1} \sin it + \alpha_{2i} \cos it)$$

with coefficients

$$\alpha = \left(X^t X\right)^{-1} X^t u$$

and residual

$$E(\alpha) = u - X\alpha.$$

Remark 1. In practice α is computed via a QR factorization of X, not by computing $(X^t X)^{-1}$ explicitly.

Next let

$$I = \{i \mid |E(\alpha)_i| < \tau_1\},$$

where τ_1 is a user defined threshold and $|I| > 2K + 1$. Calculate the coefficients for an improved fit to the underlying signal as

$$\alpha^* = \left((X_{I\cdot})^t X_{I\cdot}\right)^{-1} (X_{I\cdot})^t u_I\,.$$

With the refined coefficients α^*, calculate the residuals for

(i) the complete time series $(D_{rc})_{\cdot b}$ as

$$E^*(\alpha^*) = (D_{rc})_{\cdot b} - \bar{X}\alpha^*,$$

where $\bar{X}_{s\cdot} = (1, \sin t_s, \cos t_s, \ldots, \sin Kt_s, \cos Kt_s)$, for $s = 1, \ldots, S$.

(ii) the outlier-free time series as $\big(E^*(\alpha^*)\big)_{\bar{I}}$, where $\bar{I} = \{s \mid |E^*(\alpha^*)_s| < \gamma_2\}$, γ_2 is a user defined threshold, and

(iii) the outlier-free training set $\hat{I} = \bar{I} \cap \{1, \ldots, M\}$ as

$$\big(E^*(\alpha^*)\big)_{\hat{I}} = u_{\hat{I}} - X_{\hat{I}\cdot}\alpha^*,$$

where $|\hat{I}| > 2K + 1$.

Remark 2. In our implementation, we have defined γ_2 as

$$\gamma_2 = \begin{cases} 1.5\eta, & i \in [1, M], \\ 20\eta, & i \in (M, S], \end{cases}$$

where, η denote the standard deviation of first M elements of the residual vector $E^*(\alpha^*)$.

Step 2: Define the control limit vector τ by

$$\tau_i = \mu + \sigma L \sqrt{\frac{\lambda}{2-\lambda}\big(1 - (1-\lambda)^{2i}\big)}, \qquad i = 1, 2, \ldots, |\bar{I}|,$$

where $\mu = 0$ is used here, σ is the standard deviation of the outlier-free training data errors $\big(E^*(\alpha^*)\big)_{\hat{I}}$, L is the multiple of this standard deviation σ, and $\lambda \in (0, 1]$ is the weight given to the most recent residual in the exponentially weighted moving average (EWMA) defined next. L is typically set to 3 or slightly smaller depending on the value of λ.

Step 3: Let $\bar{I} = \{j_1, j_2, \ldots, j_{|\bar{I}|}\}$, $j_1 < j_2 < \cdots < j_{|\bar{I}|}$. Define the vector z by

$$\begin{aligned} z_1 &= \big(E^*(\alpha^*)\big)_{j_1}, \\ z_i &= (1-\lambda)z_{i-1} + \lambda\big(E^*(\alpha^*)\big)_{j_i}, \quad i = 2, \ldots, |\bar{I}|. \end{aligned}$$

This is the exponentially weighted moving average (EWMA) of the residual $\big(E^*(\alpha^*)\big)_{\bar{I}}$.

Step 4: Define the flag history S-vector f by

$$f_s = \begin{cases} \operatorname{sgn}(z_i)\big\lfloor |z_i/\tau_i| \big\rfloor, & s = j_i \in \bar{I}, \\ 0, & \text{otherwise.} \end{cases}$$

If there is a run of $+1$ or -1 in the values $\operatorname{sgn}(\Delta f_s) = \operatorname{sgn}(f_{s+1} - f_s)$ of length ϖ, called the 'persistence', signal a change at the index s beginning the (nonzero) run.

Remark 3. Missing data is automatically handled by not assuming that the time points t_i are equally spaced. Alternatively, missing data for time point t_k can be handled by including t_k in the sequence $(t_1, t_2, \ldots, t_S)$, but excluding t_k from the training sequence $(t_1, t_2, \ldots, t_M)$ and k from the sets I, $\bar{I}$, and $\hat{I}$, which is equivalent to treating $(D_{rc})_{kb}$ as an outlier and to setting the flag $f_k = 0$.

end

end

end