Oscillation Detection

Wei Cui

July 19, 2015

1 Oscillation Detection

In the topic of pollen tube growth, there are three kinds of signal which we can care about. The first is from F-actin which we still can not measure for now. The others are from ROP1 and Calclum which we can measure and they should be oscillation theoretically. Unfortunately, the signal is always not good enough. One of the best is shown in the top plot of 1. So the first task is to detect oscillation from the raw signal.

Convolution is a useful tool to this problem. There are some similar (but different) definations of convolution. We will use the following two of them.

Denote the original data list as $\{X_1, X_2, \ldots, X_n\}$.

Denote the center symmetric transform of X_n as $\{Y_1, Y_2, \ldots, Y_{2n-1}\}$ where $Y_i = X_i$ when $i = 1, 2, \ldots, n$, $Y_i = X_{2n-i}$ when $i = n + 1, n + 2, \ldots, 2n - 1$.

Denote the right expand list as $\{Z_1, Z_2, \ldots, Z_{2n-1}\}$, where $Z_i = X_i$ when $i = 1, 2, \ldots, n$, $Z_i = 0$ when $i = n + 1, n + 2, \ldots, 2n - 1$.

The first defination is about X and Y. The convolution $\{C_1, C_2, \ldots, C_n\}$ will be:

$$C_i = \sum_{j=0}^{n-1} X_{i+j} Y_{i+j} \tag{1}$$

The second defination is about X and Z. The convolution $\{C_1, C_2, \ldots, C_n\}$ will be:

$$C_i = \sum_{j=0}^{n-1} X_{i+j} Z_{i+j} \tag{2}$$

The result of these two definations (to the dataset mentioned above) is almost the same as the plots shown in 1. Then we can determine a threshold of peak values and do clustering to get period of the data.

Tt seems that the problem is solved but actually it is not. There are some other kind of data, for example the top plot of 2. The data has some trend so the pure convolution method is not efficient here. We have to eliminate the effect of trend. One method is linear regression but we can not always suppose the trend is linear.

Here we apply empirical mode decomposition (EMD) from Hilbert-Huang transform (HHT). HHT is used to get instant frequency of data and EMD will decompose the original data to a series of intrinsic mode function (IMF) and a trend function which we desire to get. We use this trend function to modify the original data and the result is in 1 and 2.

Figure 1: The result of the first dataset. Top: standarized data. Middle: result of defination one with modification. Bottom: result of defination two with modification.
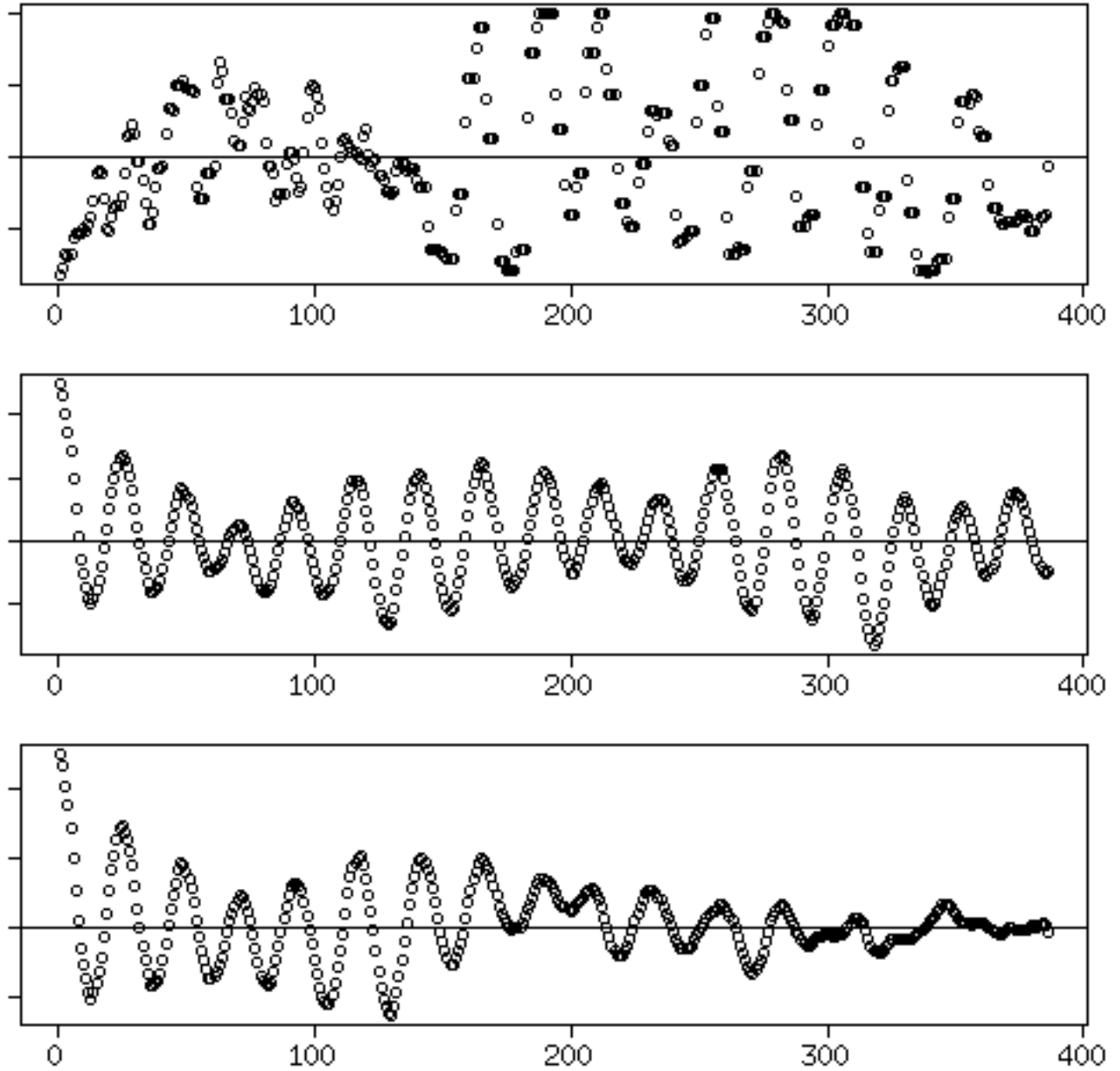
Positive:

1. The method is very simple.

2. It is actually very fast since the convolution manipulation can be done by fast Fourier transform (But EMD will take some time).

3. The result is acceptable.

Negative:

1. The result depends on the head part of original data which means if only some other part of the data is oscillation then the method will be not so efficient.

2. The tail part of the result is not very confidential since there are not enough data point in the tail.

3. There is no universal method to determine the threshold value.

2   Growth model include ROP1 and Calcium

In the previous model, only the signal of ROP1 is considered. The final target is to cover the signal of ROP1, F-actin and Calclum. Because the F-actin is not measurable for now, Nan Luo suggest a model include the signal of ROP1 and Calcium as follows.

$$
\begin{aligned}
\frac{\partial R(t)}{\partial t} &= k_{pf}(R_{total} - R(t)) - k_{nf}C(t)R(t) \\
\frac{\partial C(t)}{\partial t} &= k_a(R(t-\tau) + C_0) - k_d C(t)
\end{aligned}
\tag{3}
$$

Four assumptions:

1. The time interval between any two continuous measurements is very small;

2. The time interval between any two continuous measurements is constant.

3. The parameter τ is non-negative integer.

4. The upper bound of period length L can be estimated from data. (With the result of Oscillation Detection part, this assumption can be satisfied.)

The first assumption allows us to replace differential equation with difference equation. The parameter τ can be estimated with assumption 2 and 3. The parameter space of τ is limited with assumption 3 and 4.

The method:

1. let $t = 0, 1, 2, \ldots, N$ as the sampling time.

2. To $t = 1, 2, \ldots, N-1$, replace $\frac{\partial R(t)}{\partial t}$ with $\frac{R(t+1)-R(t-1)}{2}$, replace $\frac{\partial C(t)}{\partial t}$ with $\frac{C(t+1)-C(t-1)}{2}$. Replace $k_{pf}R_{total}$ with kR, replace $k_a C_0$ with kC. Then the formula becomes:

$$\frac{R(t+1)-R(t-1)}{2} = kR - k_{pf}R(t) - k_{nf}C(t)R(t)$$

$$\frac{C(t+1)-C(t-1)}{2} = k_a R(t-\tau) + kC - k_d C(t) \tag{4}$$

3. Estimate kR, k_{pf}, k_{nf} by ordinary least squares (OLS) or generalized least squares (GLS).

4. Let $\tau = 0$. Then we can get OLS estimation or GLS estimation of kC, k_a, k_d.

5. Let $\tau = \tau + 1$ and return to step 4 until $\tau = L$.

6. Choose the best τ from $L + 1$ estimations.

Positive:

1. The method is simple enouph.

2. The method can estimate all the parameters.

Negative:

1. The method will import new computational errors.

An alternative method is to eliminate $C(t)$ from the equation array to get an ODE of $R(t)$ and then solve this ODE. But it will be difficult to estimate all the parameters.

Figure 2: The result of the second dataset. Top: standarized data. Middle: result of defination one with modification. Bottom: result of defination two with modification.