Wavelet Analysis of Epidemiological Time-Series

COVID-19 Hackathon

Yanran Cui

Jun. 14, 2020

Introduction

- Emerging and re-emerging of epidemic diseases have been a great challenge to public health. To analyze the data and build a proper model is necessary.

- Traditional Time-Series analysis is only applicable to stationary process, but most epidemiological processes are non-stationary.

- Wavelet Analysis is an effective way to decompose non-stationary time-series.

Definition

The wavelet transform is defined as the convolution of the signal with a wavelet function. a is called the scale of the wavelet.

$$W_x(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-\tau}{a})dt$$

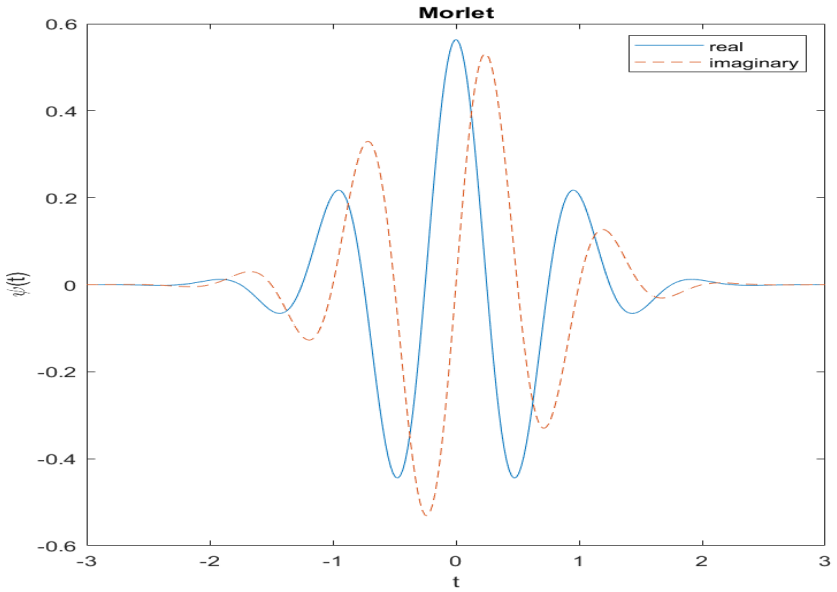
The discrete form is

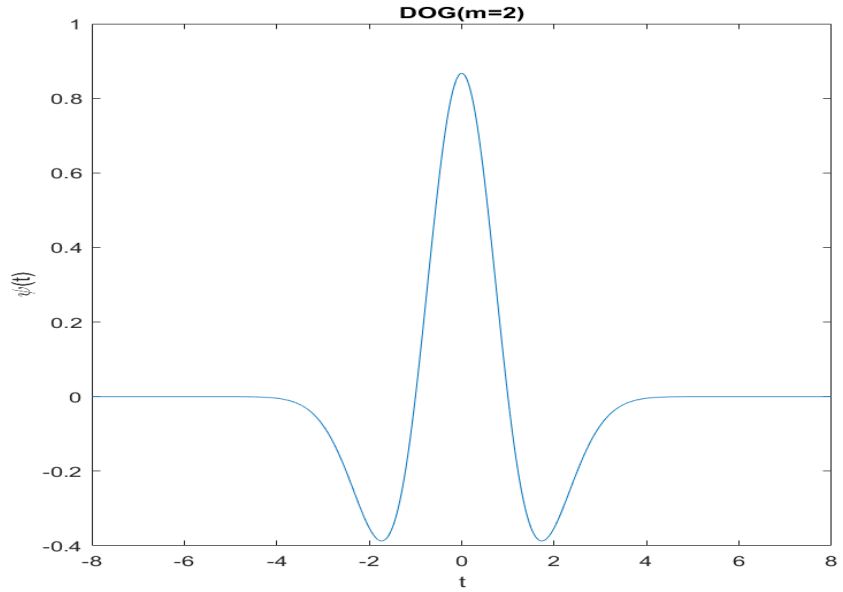
$$W_x(a,\tau) = \frac{\Delta t}{\sqrt{a}} \sum_{k=0}^{N-1} x_n \psi^*[\frac{(k-\tau)\Delta t}{a}]$$

The requirement for a wavelet function

$$\int \parallel \psi(t) \parallel^2 dt \; = \; 1 \quad \int \psi(t)dt = 0$$

Wavelet Function





Morlet is a complex function, *Derivative of Gaussian (DOG)* is a real-valued function. Below is the formula for *Morlet* and use it for the rest of the analysis.

$$\psi(t) = \pi^{-1/4} exp(-i2\omega_0 t) exp(-t^2/2)$$

'Damping' version of Fourier transform.

Wavelet Power Spectrum

The wavelet power is defined as

$$S_x(a,\tau) = \| W_x(a,\tau) \|^2$$

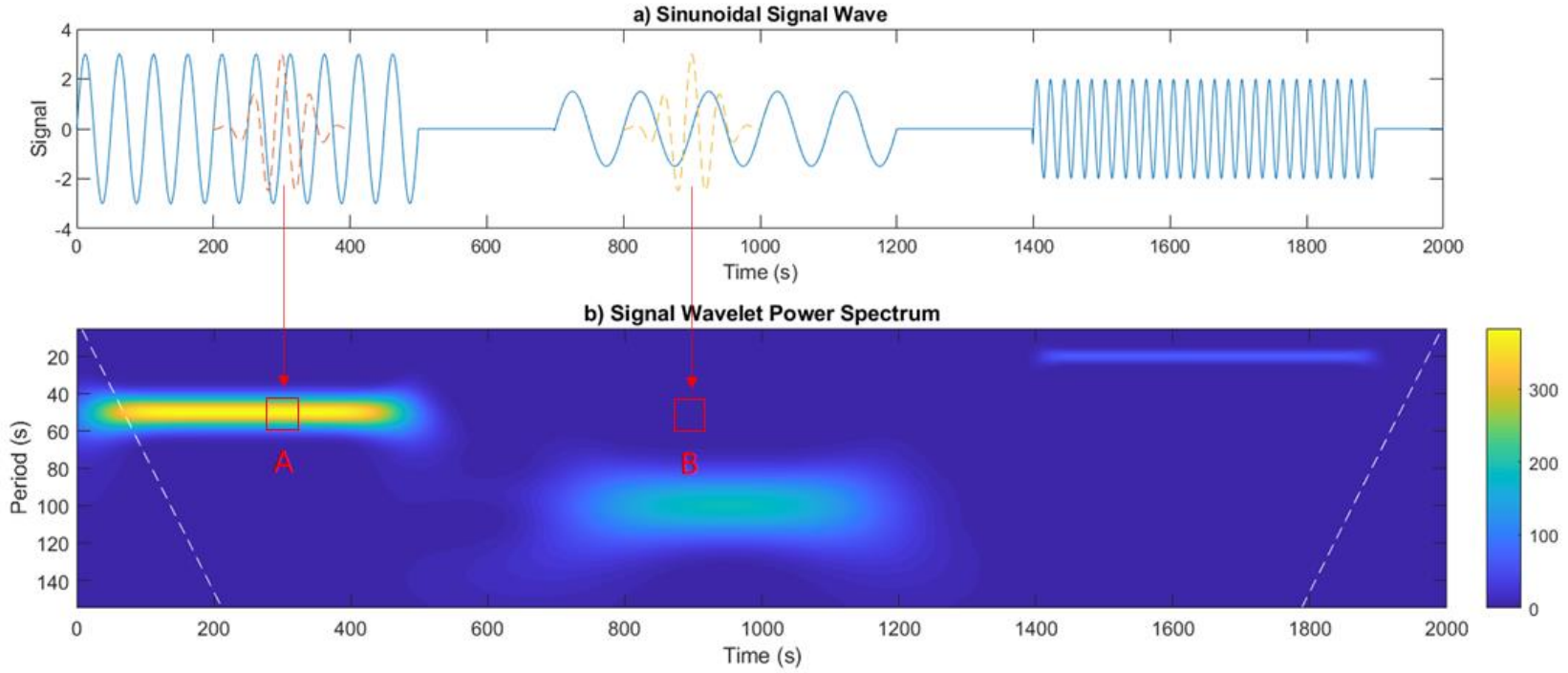
Two averages:

Global Wavelet Spectrum $$\bar{S}_x(a) = \frac{1}{N}\sum_{\tau=0}^{N-1} \| W_x(a,\tau) \|^2$$

Scale Average Spectrum $$\bar{S}_x(\tau) = \frac{\Delta a \Delta t}{C_\delta}\sum_{j=j_1}^{j_2} \frac{\| W_x(a_j,\tau) \|^2}{a_j}$$

It is a two-dimensional power spectrum.

Scalogram



- When the wavelet matches well with the signal, it will produce high power. When the wavelet doesn't match well with the signal, it will produce low power.
- The dashed white line is called cone of influence, which denotes the influence of the edge effects.
- The scale can be converted to a corresponding Fourier period for interpretation purpose.

Statistical Significance

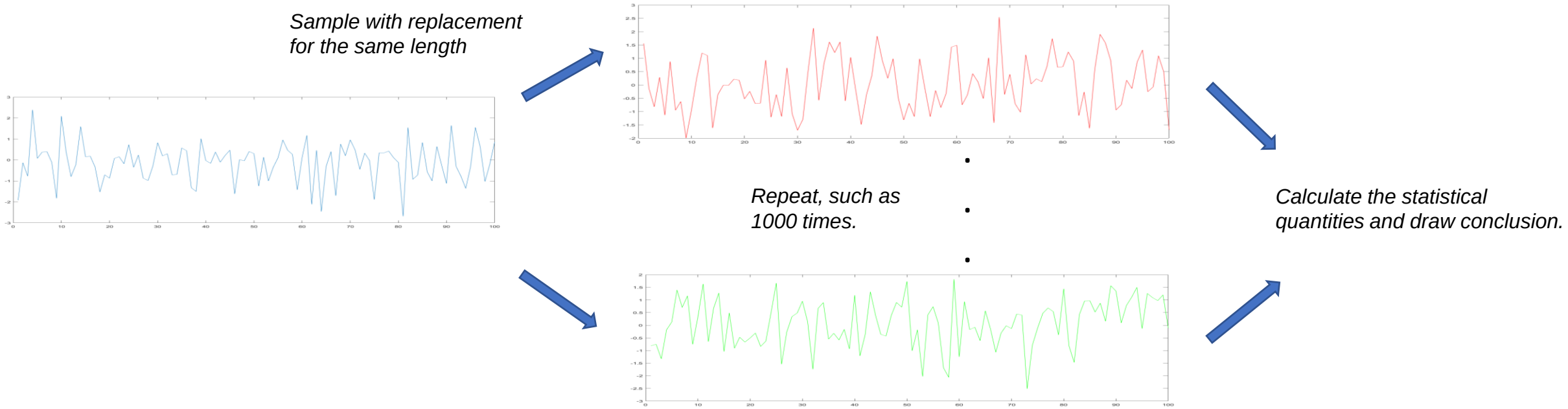
Generally there are two types of statistical significance test method for the wavelet analysis:

1. Model based

First pre-assume a model such as AR process, then based on the model to derive the test level on the power. Need some previous knowledge of the process.

2. Sample based

Bootstrap method. Sample randomly from the original data with replacement. Then perform analysis on the sample data. Repeat this process to construct test level.



Wavelet Coherency

The wavelet cross-spectrum

$$W_{x,y}(a,\tau) = W_x(a,\tau)W_y^*(a,\tau)$$

Coherency

$$R_{x,y}(a,\tau) = \frac{\| \langle W_{x,y}(a,\tau) \rangle \|}{\| \langle W_{x,x}(a,\tau) \rangle \|^{1/2} \| \langle W_{y,y}(a,\tau) \rangle \|^{1/2}}$$

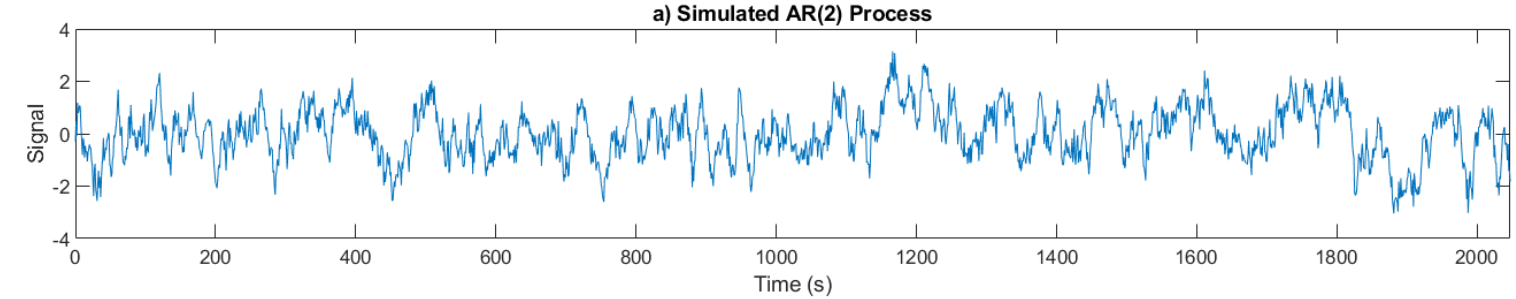
Phase Difference

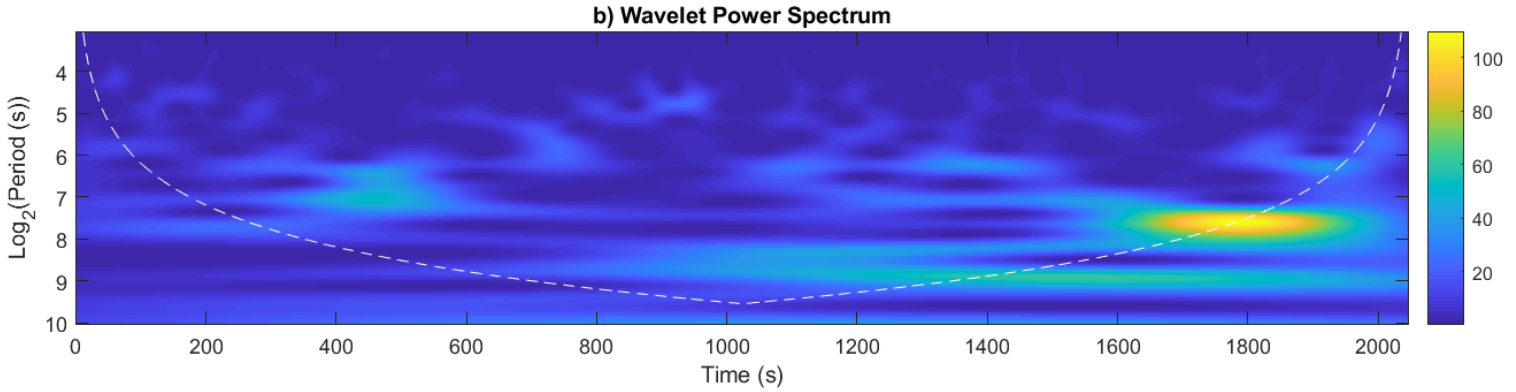
$$\phi_{x,y}(a,\tau) = tan^{-1} \frac{\Im(\langle W_{x,y}(a,\tau) \rangle)}{\Re(\langle W_{x,y}(a,\tau) \rangle)}$$

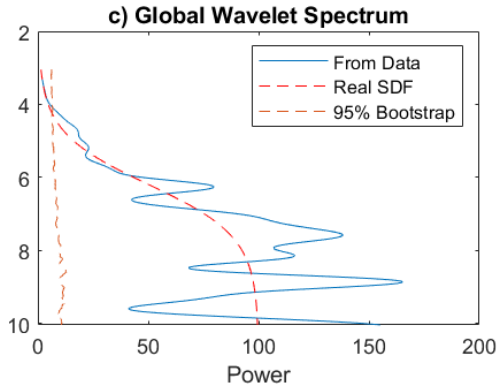
High Coherency implies high correlation, unimodal phase distribution implies phase lock between two time-series.

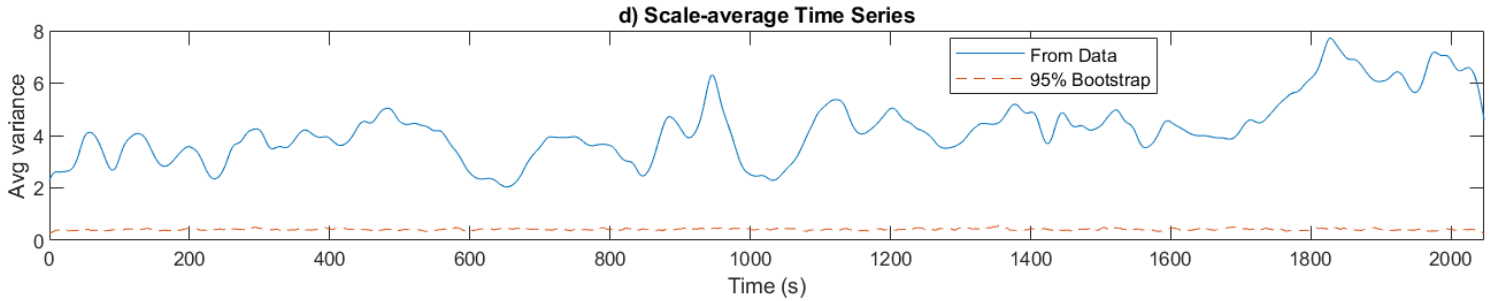
Example: Wavelet Analysis of AR(2) process

$$x_n = 0.7x_{n-1} + 0.2x_{n-1} + \epsilon_n \qquad \epsilon_n \sim N(0,1)$$





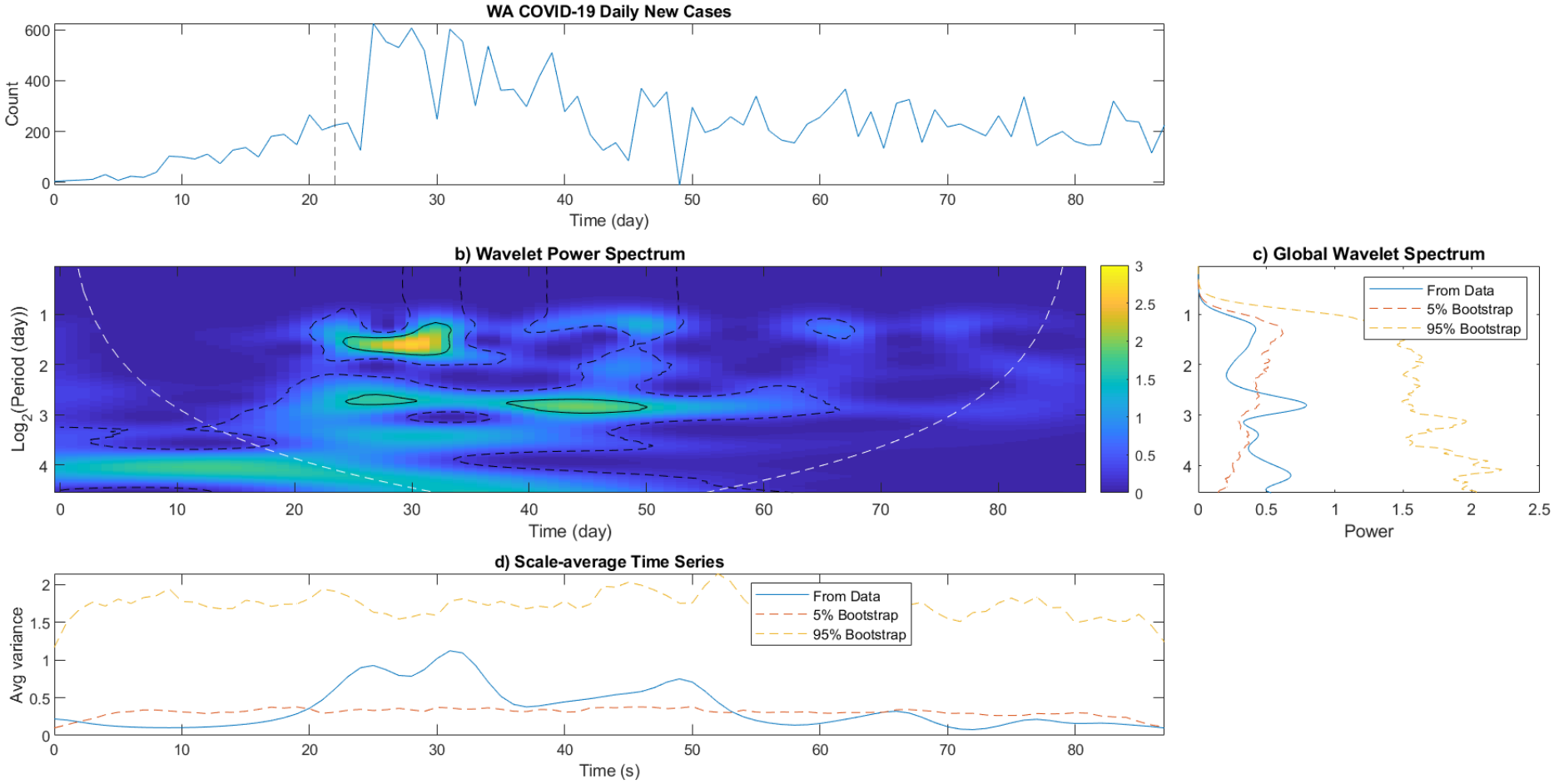




Time averaged power spectrum converges to the true SDF.

Wavelet Analysis of COVID-19 Data

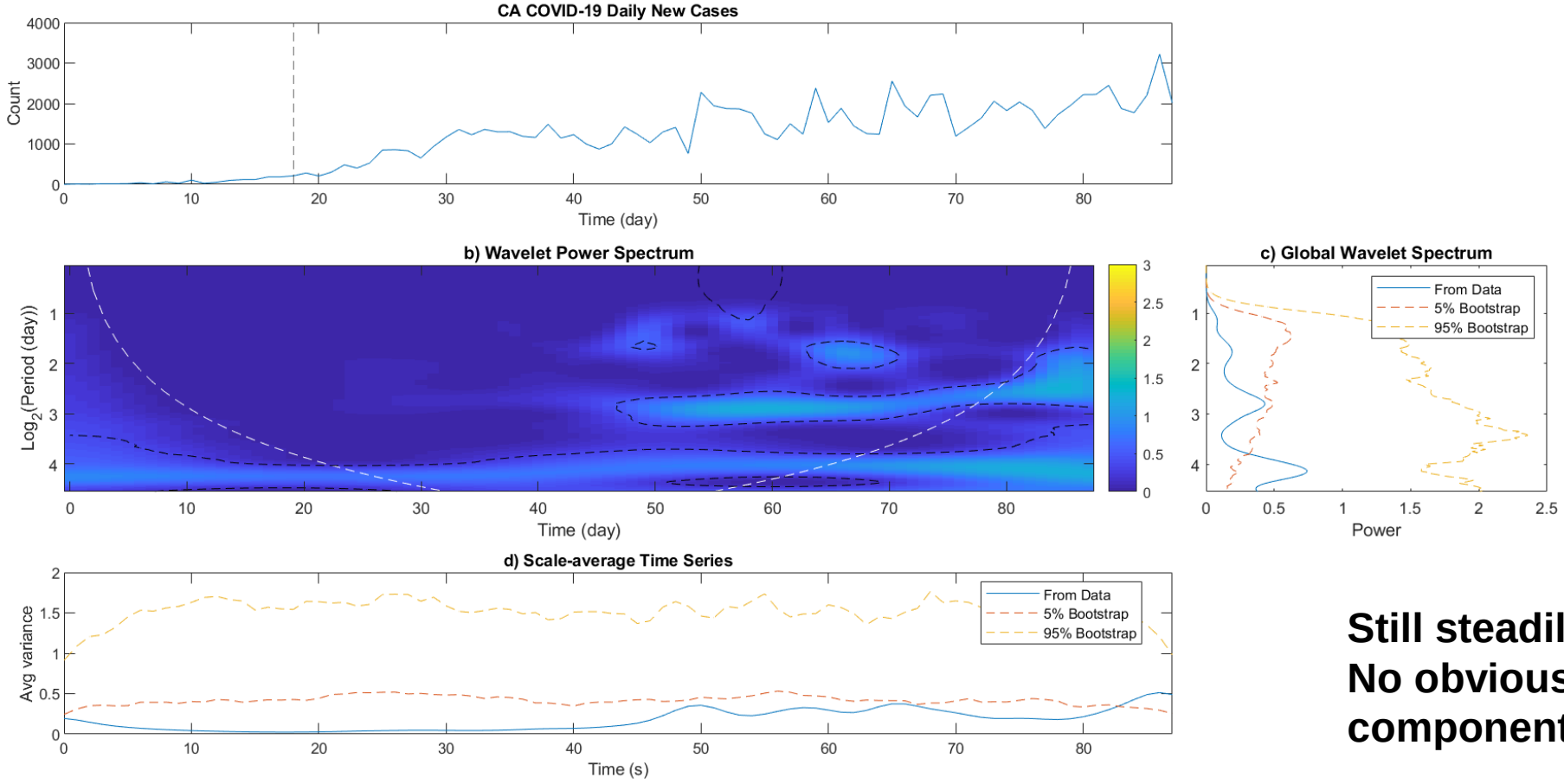
Three states were chosen: Washington (WA), California (CA) and New York (NY). Analysis start from Mar. 1st, 2020.



- **High power at 4-day period appeared during the explosion stage then disappeared.**
- **After 60 days, the daily new cases became random and imply stabilization of the process.**

Wavelet Analysis of COVID-19 Data

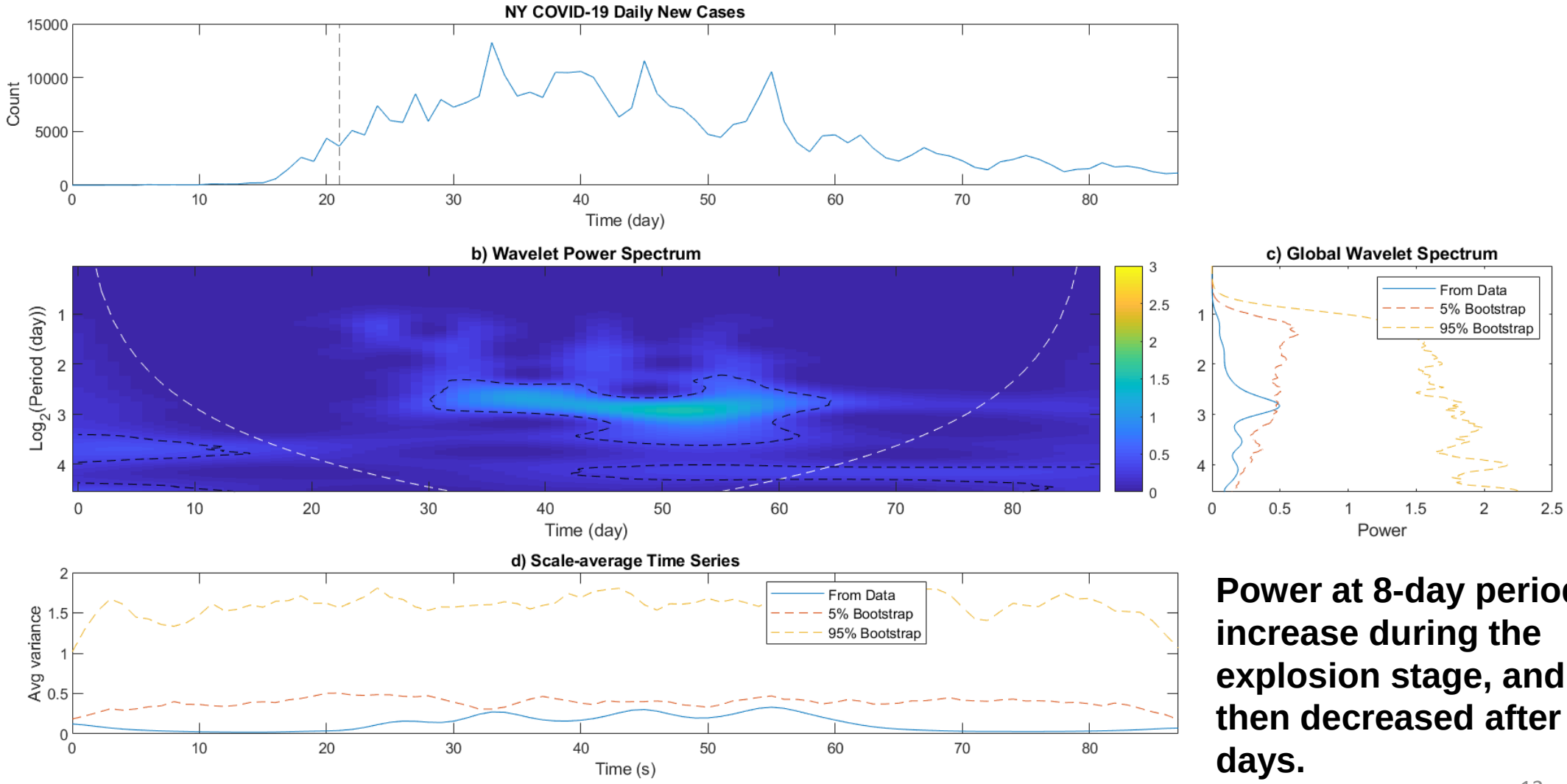
California



Still steadily increasing. No obvious fluctuation component observed.

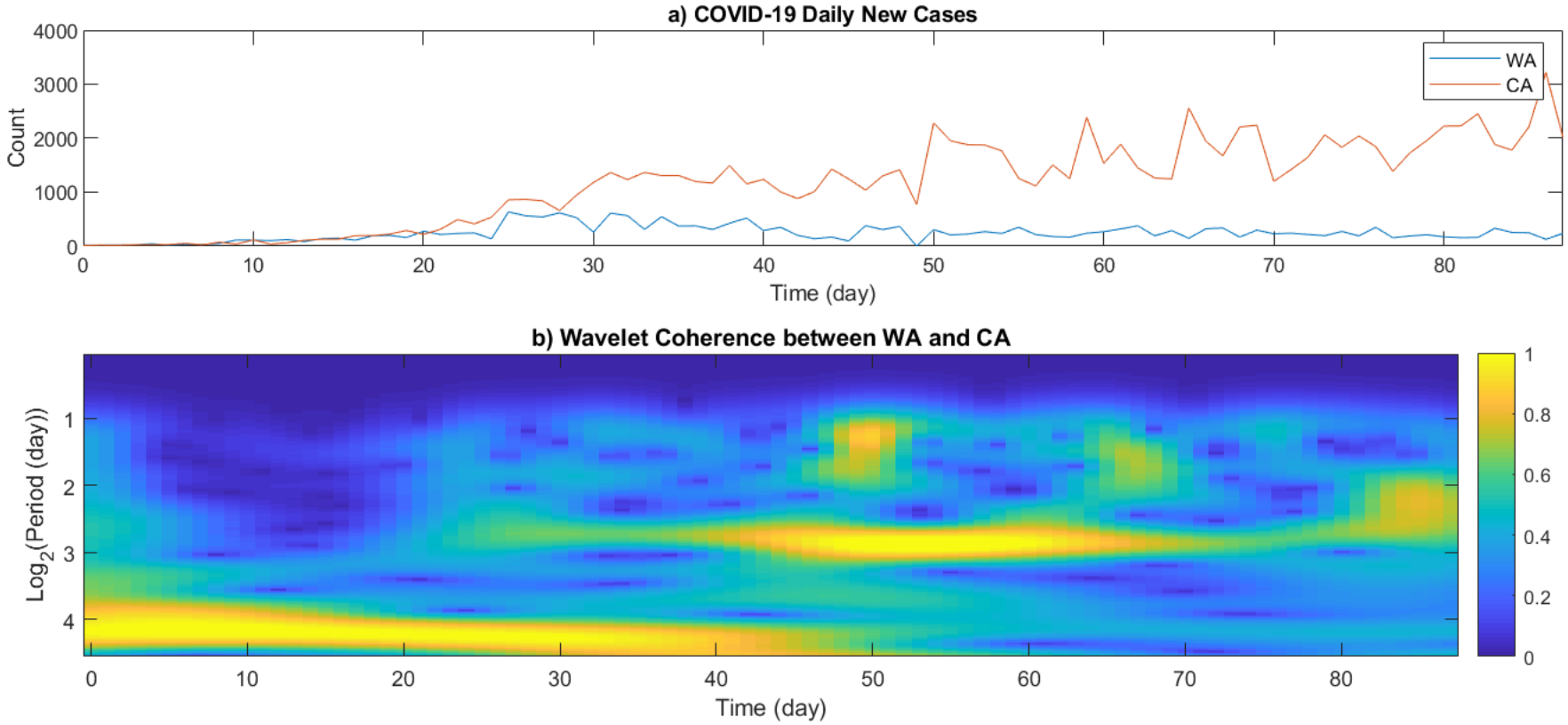
Wavelet Analysis of COVID-19 Data

New York State



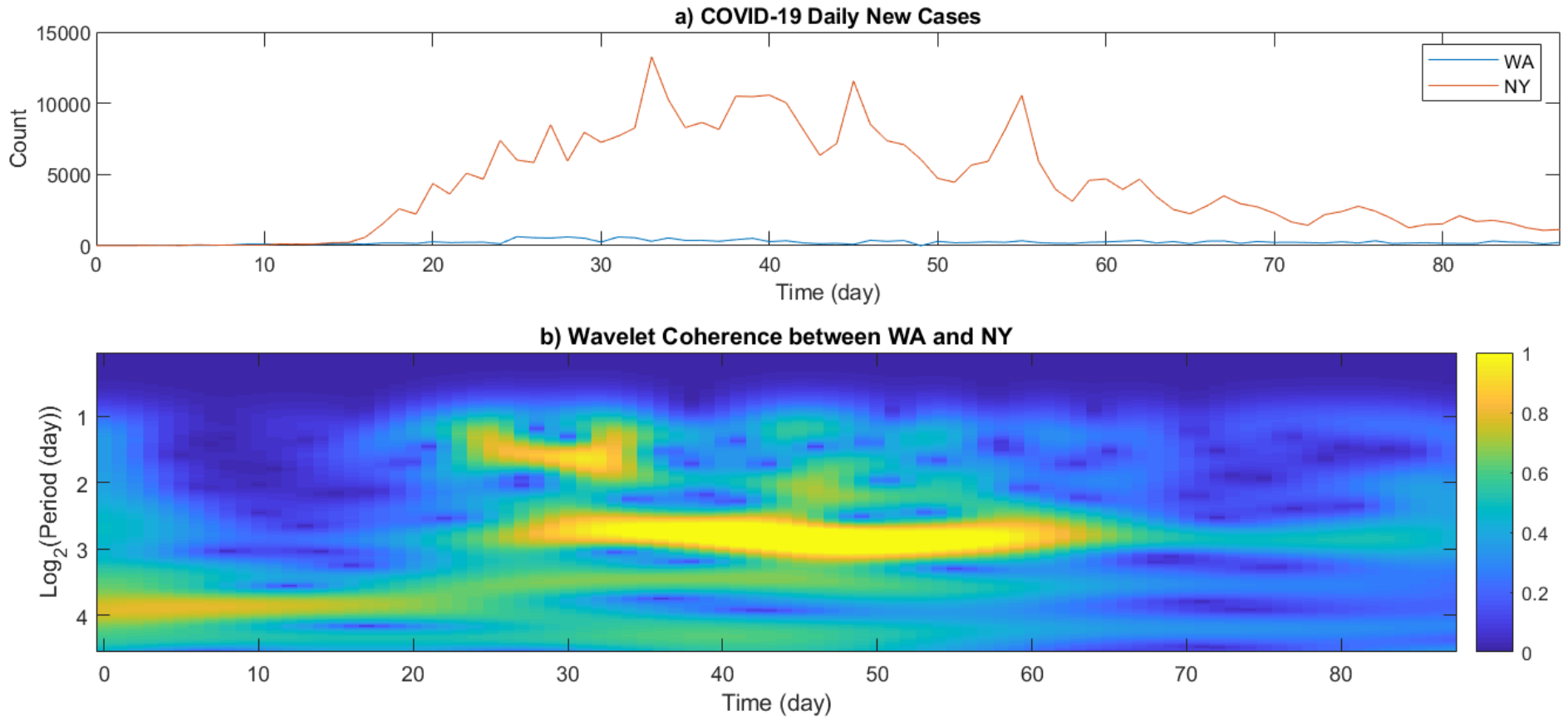
Power at 8-day period increase during the explosion stage, and then decreased after 60 days.

Wavelet Coherency between WA and CA



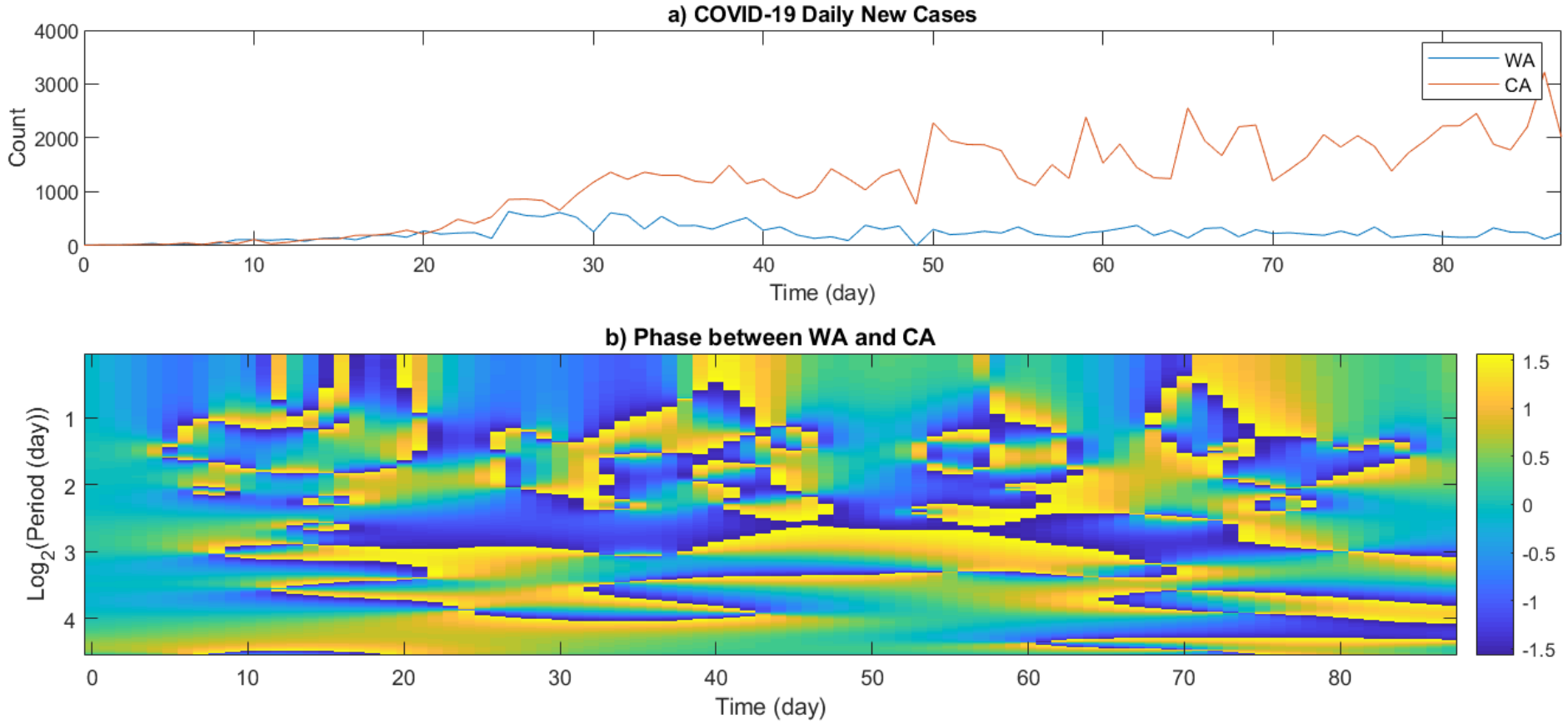
Correlation at high period at the beginning and then shift to low period. The correlation is weak after 70 days.

Wavelet Coherency between WA and NY



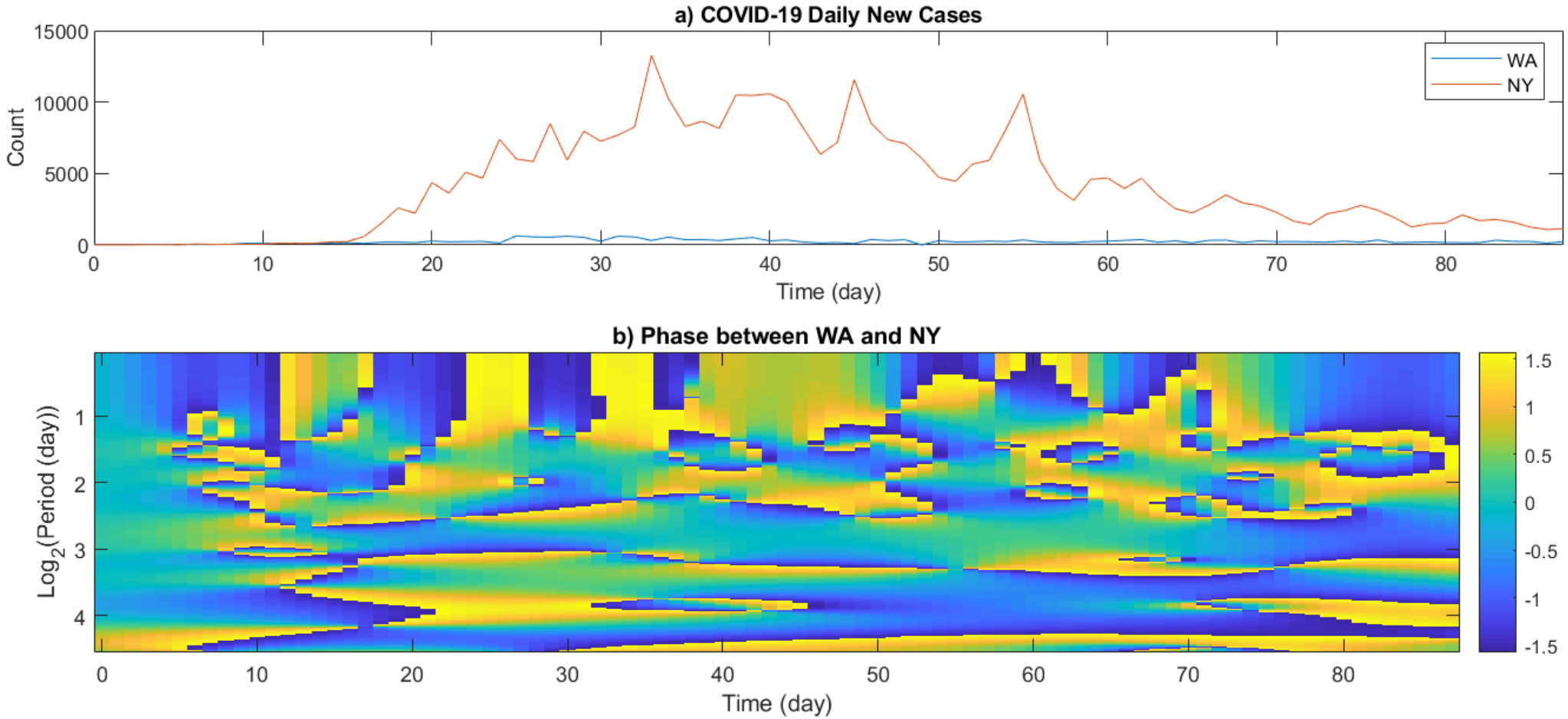
Correlation at low period (periodicity of 4 days) between 30-60 days. The correlation is weak after 70 days.

Wavelet Phase Different between WA and CA



Initially the same phase, then no obvious patterns.

Wavelet Phase Difference between WA and NY



Initially the same phase, then no obvious patterns.

Summary

- Wavelet can decompose non-stationary time-series into time and scale domains.
- Wavelet analysis of the COVID-19 data showed different dynamics in different states. WA tends to be stable first and NY also implies stabilization. However, CA is still not stabilized and increasing.
- Correlation exists between the spread of the virus in different states.

Thank you!