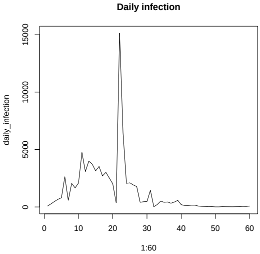
Forcasting the amplitude, the location and the width of the peak of time series using the weigthed median

A. Dermoune,Y. Esttafa, D. Ounaissi, Y. Slaoui

Le Mans 21/5/2024

Motivation

Let us consider the daily infections $t \in [1, 60] \to y(t)$ of COVID-19 in China during the first wave:



Location $L = 22$ and amplitude $A = 15136$.

Question

Is it possible to predict the true location $L = 22$ and the true amplitude $A = 15136$ before a date $T < L$?

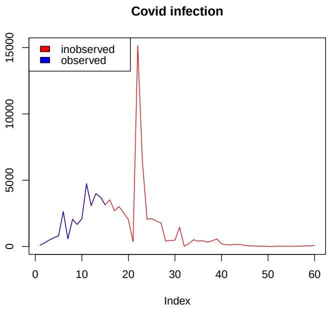


Figure: inobserved and observed infections

What is a peak

Popular forecasting models such as Autoregressive Integrated Moving Average (ARIMA) and Recurrent Neural Networks (RNN) are unable to define or forcast peak values in the data.
Before going further, we need to answer a more fundamental question: What is a peak? For finding peaks in a time series, the SciPy signal processing module offers the powerful scipy.signal.findpeaks function. In the SciPy implementation, there is a single mandatory requirement for a sample to be labeled as a peak. A sample $y[k]$ is considered a peak, if the value is a local maximum, e.g.

$$y[k] > y[k-1] \text{ and } y[k] > y[k+1].$$

Our peak definition

Let A, L, S three positive numbers and t a positive integer. Let us corrupt $A\exp(-\frac{(t-L)^2}{S^2})$, by a random number $e(t)$. We consider the observation $y(t) = A\exp(-\frac{(t-L)^2}{S^2}) + e(t)$. The peak of our time series y takes place at the location L with the amplitude A and the width S.

Time series with several peaks

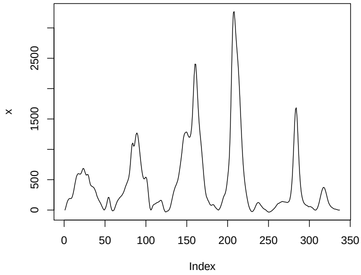


Figure: Time series with several peaks

How to forcast the amplitude, the location and the width

We propose to forcast A, L, S from the observations before the location L, i.e., from $y(t), t \in [1, T]$ with $T < L$.

Our method

To estimate the three parameters (A, L, S) based on the T observations, we consider LAD nonlinear regression

$$f(T,a,l,s) = \frac{\sum_{t=1}^{T} |y(t) - a\exp(-\frac{(t-l)^2}{s^2})|}{T}$$
$$= \frac{\sum_{t=1}^{T} \exp(-\frac{(t-l)^2}{s^2})|a - y(t)\exp(\frac{(t-l)^2}{s^2})|}{T}.$$

For each (l, s) fixed, the minimum of the function $a \to f(T,a,l,s)$ is attained at the weighted median $a(T,l,s)$ of the sequence $(x(t) = y(t)\exp(\frac{(t-l)^2}{s^2}) : t = 1, \ldots, T)$ endowed with the weights $(w(t) = \exp(-\frac{(t-l)^2}{s^2}) : t = 1, \ldots, T)$.

Algorithm 1 Our algorithm to predict the amplitude, the location and the width of the peak.

Require: T, and $h > T$

1: For each (l, s) fixed, the minimum of the function $a \to f(T, a, l, s)$ is attained at the weighted median $a(T, l, s)$ of the sequence $(x(t) = y(t)\exp(\frac{(t-l)^2}{s^2}) : t = 1, \ldots, T)$ endowed with the weights $(w(t) = \exp(-\frac{(t-l)^2}{s^2}) : t = 1, \ldots, T)$.

2: Calculate the minimizer $s(T, l)$ of the curve $s \to f(T, a(T, l, s), l, s)$ for each $l \in [1, h]$.

3: We propose the minimizer $\hat{l}(T)$ of the sequence $l \in [1, h] \to f(T, a(T, l, s(T, l)), l, s(T, l))$ as a predictor of the peak L.

4: We propose $\hat{s}(T) = s(T, \hat{l})$ as a predictor of the width S.

5: We propose $\hat{a}(T) = a(T, \hat{l}, s(T, \hat{l}))$ as a predictor of the amplitude A.

Ensure: $\hat{l}(T)$, $\hat{s}(T)$ and $\hat{a}(T)$.

Our algorithm forcasting with free noise

$A = 15000$, $L = 22$, $S = 10$ and the time horizon $h = 30$. In the following figure, we plot the curve $l \in [1, h] \to f(T, a(T, l, s(T, l)), l, s(T, l))$ with $T = 2, 3, 10$ and 20.

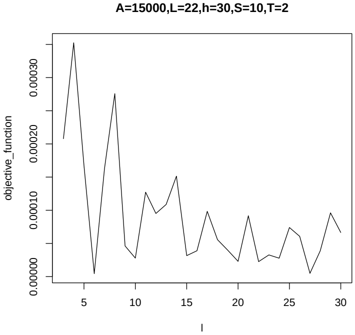


Figure: With T=2

Our algorithm forcasting with free noise

$A = 15000$, $L = 22$, $S = 10$ and the time horizon $h = 30$. In the following figure, we plot the curve
$l \in [1, h] \to f(T, a(T, l, s(T, l)), l, s(T, l))$ with $T = 3, 10$ and 20.

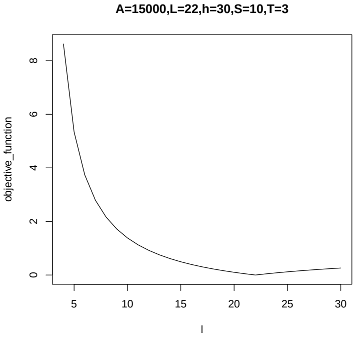


Figure: With T=3

Our algorithm forcasting with free noise

$A = 15000$, $L = 22$, $S = 10$ and the time horizon $h = 30$. In the following figure, we plot the curve $l \in [1, h] \rightarrow f(T, a(T, l, s(T, l)), l, s(T, l))$ with $T = 2, 3, 10$ and 20.

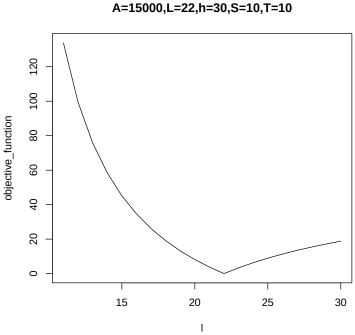


Figure: With T=10

Our algorithm forcasting with free noise

$A = 15000$, $L = 22$, $S = 10$ and the time horizon $h = 30$. In the following figure, we plot the curve $l \in [1, h] \rightarrow f(T, a(T, l, s(T, l)), l, s(T, l))$ with $T = 2, 3, 10$ and 20.

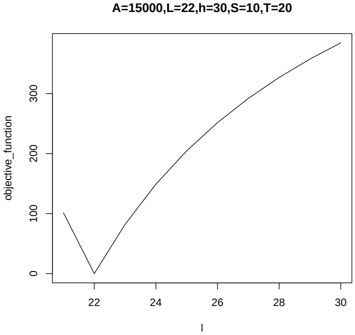


Figure: With T=20

Our algorithm forcasting with Laplace noise

Let us first recall the probabilistic interpretation of LAD regression. We have

$$y(t) = A\exp(-\frac{(t-L)^2}{S^2}) + e(t),$$

where the errors $(e(t))$ are i.i.d. with the common probability distribution

$$\frac{1}{2b}\exp(-\frac{|e|}{b}), \quad \text{with the scale} \quad b > 0.$$

Based on the data $(y(1), \ldots, y(T))$ the likelihood is equal to

$$\prod_{t=1}^{T} \frac{1}{2b} \exp\left(-\frac{|y(t) - a\exp(-\frac{(t-l)^2}{s^2})|}{b}\right).$$

It comes that the maximum likelihood estimator (MLE) $\hat{\theta}$ of the parameter $\theta = (a, l, s, b)$ are

$$\begin{cases} (\hat{a}, \hat{l}, \hat{s}) & = & \arg\min\{f(T, a, l, s) : a, l, s\} \\ \hat{b} & = & f(T, \hat{a}, \hat{l}, \hat{s}). \end{cases}$$

There are many local minimizer (Dermoune, Ounaissi, Slaoui 2023).

The $SNR = \frac{A}{b\sqrt{2}}$.

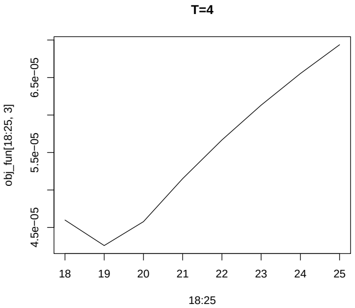
$$\hat{l}(T, A, L, h, S, SNR, seed) = \arg\min_{l \in [1,h]}$$

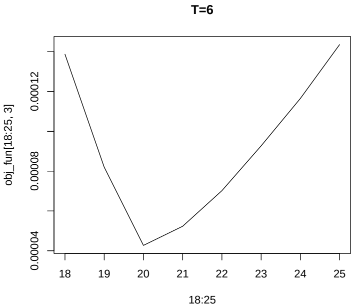
$$\frac{1}{T}\sum_{t=1}^{T} |a(T, l, s(T, l)) \exp(-\frac{(t-l)^2}{s^2(T, l)}) - y(t)|,$$

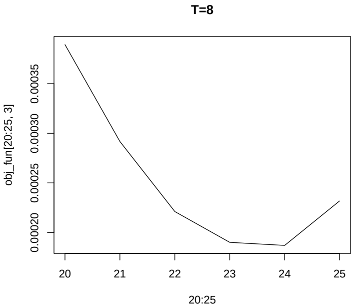
$$\hat{s}(T, A, L, h, S, SNR, seed) = s(T, \hat{l}(T, A, L, h, S, SNR, seed)),$$

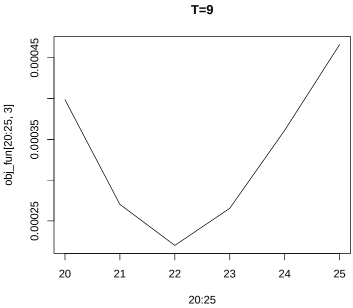
$$\hat{a}(T, A, L, h, S, SNR, seed) = a(T, \hat{l}(T, A, L, h, S, SNR, seed).$$

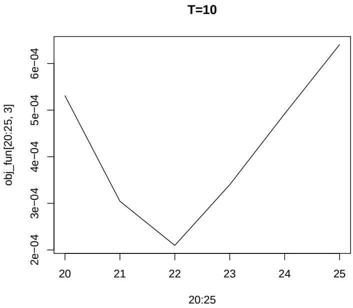
Here we plot the objective function
$l \in [18, 25] \rightarrow f(T, a(T, l, s(T, l)), l, s(T, l))$ with
$A = 1, h = 30, S = 10, L = 22$, $SNR = 1000$, seed=280, and
$T = 4, 6, 8, 9, 10$:




T=6
obj_fun[18:25, 3]
0.00004
0.00008
0.00012
18
19
20
21
22
23
24
25
18:25


T=8
obj_fun[20:25, 3]
0.00020
0.00025
0.00030
0.00035
20
21
22
23
24
25
20:25


T=9
0.00045
0.00035
0.00025
obj_fun[20:25, 3]
20
21
22
23
24
25
20:25


T=10
obj_fun[20:25, 3]
2e−04
3e−04
4e−04
5e−04
6e−04
20
21
22
23
24
25
20:25

Minimum number of observations needed for the exact prediction

For each fixed $A, L, h, S, SNR, seed$, let us define:

1) The minimum number of observations needed for a good prediction of the true amplitude A equals

$$T_1(A, L, h, S, SNR, seed) = \min\{T : \hat{a}(T, A, L, h, S, SNR, seed) \approx \hat{a}(T+1, A, L, h, S, SNR, seed) \ldots \approx \hat{a}(h, A, L, h, S, SNR, seed)\}.$$

2) The minimum number of observations needed for the exact prediction of the true location L equals

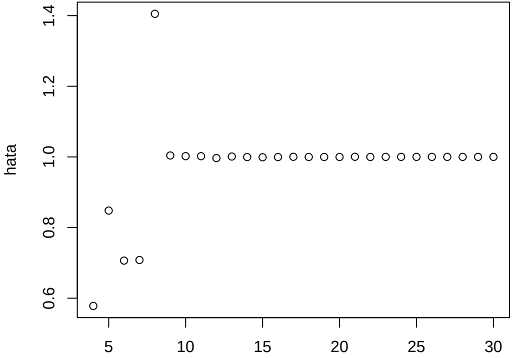
$$T_2(A, L, h, S, SNR, seed) = \min\{T : \hat{l}(T, A, L, h, S, SNR, seed) = \ldots \hat{l}(h, A, L, h, S, SNR, seed)\}.$$

3) The minimum number of observations needed for a good prediction of the true width S equals

$$T_3(A, L, h, S, SNR, seed) =$$
$$\min\{T : \hat{s}(T, A, L, h, S, SNR, seed) \approx \ldots$$
$$\hat{s}(h, A, L, h, S, SNR, seed)\}.$$

Numerical illustration

Here we plot the curves $T \in [4, h] \to \hat{l}(T, A, L, S, SNR, seed)$, $T \in [4, h] \to \hat{s}(T, A, L, S, SNR, seed)$, $T \in [4, h] \to \hat{a}(T, A, L, S, SNR, seed)$ with $A = 1$, $L = 22$, $S = 10$, $SNR = 1000$, and $seed = 280$. The curves $\hat{l}$,$\hat{a}$ and $\hat{s}$ converge at $T_1 = T_2 = T_3 = 9$.



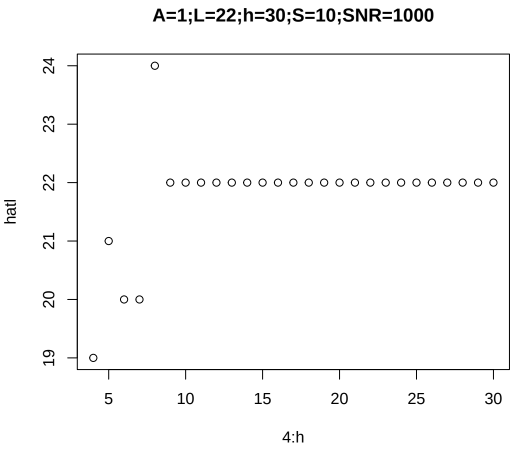


Figure: The curve $T \in [4, h] \to \hat{l}(T)$.

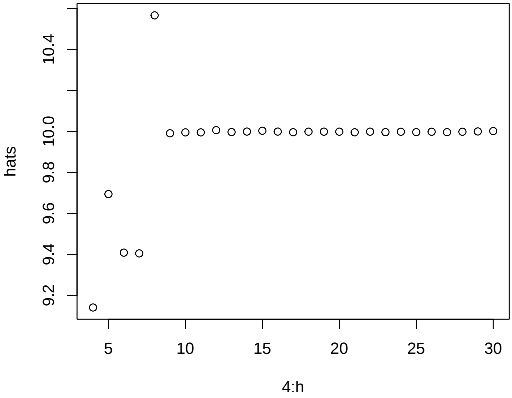


Figure: The curve $T \in [4, h] \to \hat{s}(T)$.

Amplitude, width, signal to noise ratio effects

We recall that $y(t) = A\exp(-\frac{(t-L)^2}{S^2}) + e_b(t)$, with $e_b(t)$ is the Laplace noise having th scale b. The equality

$$Tf(T,a,l,s) = \sum_{t=1}^{T} |a\exp(-\frac{(t-l)^2}{s^2}) - A\exp(-\frac{(t-L)^2}{S^2}) - e_b(t)|$$

shows that our objective function depends also on A, L, S, e_b. Hence the weighted median $a(T,l,s) = a(T,l,s,A,L,S,e_b)$, $s(T,l) = s(T,l,A,L,S,e_b)$ and $\hat{l}(T)) = \hat{l}(T,A,L,S,e_b)$ depend also on A, L, S and e_b.

Now, we can announce the following results.

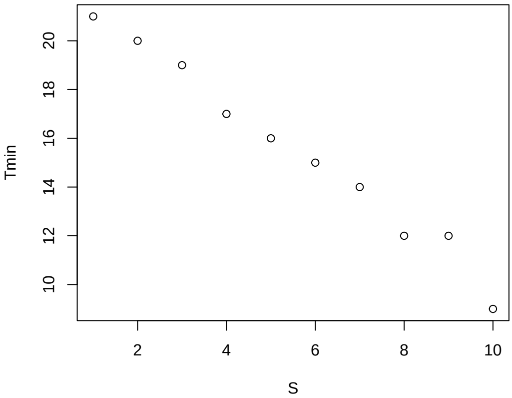
We have

$$a(T,l,s,A,L,S,e_b) = Aa(T,l,s,1,L,S,SNR,seed),$$
$$s(T,l,A,L,S,e_b) = s(T,l,1,L,S,SNR,seed),$$
$$\hat{l}(T,A,L,S,e_b) = \hat{l}(T,1,L,S,SNR,seed).$$

Width effect

Here we plot the curve $S \in [1, 10] \rightarrow Tmin(S)$ with $A = 1$, $L = 22$, $h = 30, SNR = 1000$, seed=280.

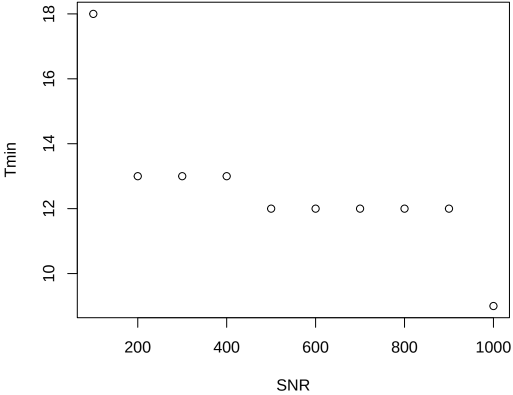
$$21, 20, 19, 17, 16, 15, 14, 12, 12, 9.$$



SNR effect: $SNR \geq 100$

$$A = 1; L = 22; h = 30; S = 10$$

$$SNR = 100, \ldots, 1000\text{: } 18,\ 13,\ 13,\ 13,\ 12,\ 12,\ 12,\ 12,\ 12{,}9.$$



SNR effect: $SNR = 10$

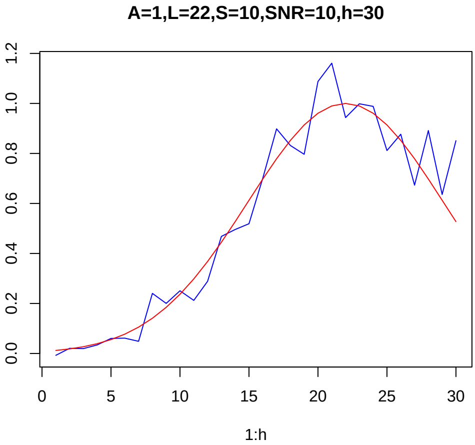


Figure: data plus noise.

SNR effect: $SNR = 5$

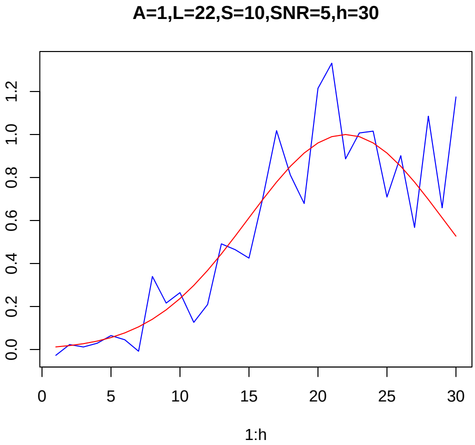


Figure: data plus noise.

SNR effect: $SNR = 1$

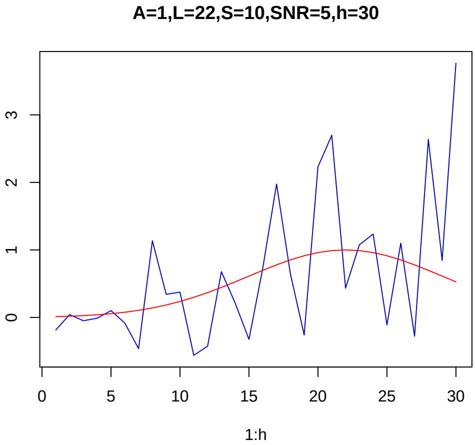


Figure: data plus noise.

SNR effect: $SNR \leq 10$

$SNR = 1, \ldots, 10$: 30, 30, 28, 28, 30, 30, 30, 30, 22, 22.

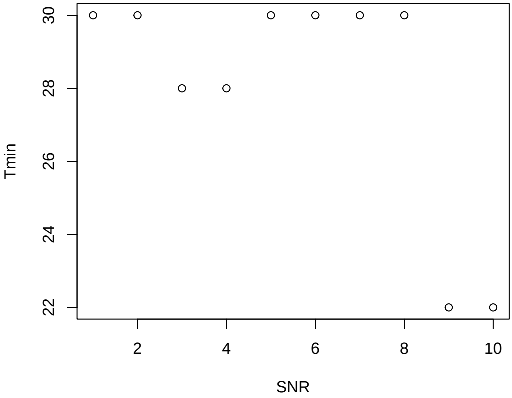


Figure: The curve $SNR \in [1, 10] \rightarrow Tmin(SNR)$.

Minimum number *Tmin* of observations needed for the exact prediction of the true location L: the case $T^* \geq L$

In the case where the number of observations *Tmin* needed for the exact prediction is such that $T^* \geq L$, our prediction is not helpful. We recall that we want to predict the location L from the date $T < L$.

Time of entry in the region of the peak

Instead of predicting the peak, we propose to define a time of the entry in the region of the peak and the exit time from the region of the peak as follows. The time of the entry in the region of the peak is based on two threshold values τ_1,τ_2. We look for the first time $T(\tau_1, \tau_2)$ such that the sequence $T \to \hat{I}(T)$ satisfies

$$|\hat{I}(T) - \hat{I}(T+i)| \leq \tau_2, \quad \forall\, i = 1, \dots, \tau_1.$$

The exit time from the region of the peak

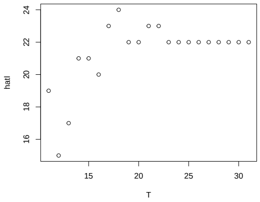
Having the observations $T > T(\tau_1, \tau_2)$, if T is not the peak, we expect that the peak will happen soon. The exit time from the region of the peak is defined by the first time T after $T(\tau_1, \tau_2)$ such that

$$|\hat{I}(T) - \hat{I}(T+i)| \leq \tau_3, \quad \forall\, i = 1, \ldots, \tau_1.$$

Illustration

In the case $A = 1$, $L = 22$, $h = 35$, $S = 10$, $SNR = 40$, $\tau_1 = 3$; $\tau_2 = 1$; $\tau_3 = 0$, we have

$$\hat{l}(11:31): 19, 15, 17, 21, 21, 20, 23, 24, 22, 22,$$
$$23, 23, 22, 22, 22, 22, 22, 22, 22, 22, 22.$$



The entry time in the region of the peak equals 16. The exit time from the region of the peak equals 22.

What happens if we permute s and l in our algorithm?

Algorithm 2 This new algorithm is not suitable and converges to another local minimizer of our objective function.

Require: T, and $h > T$

1: Calculate the minimizer $l(T, s)$ of the curve $l \rightarrow f(T, a(T, l, s), l, s)$ for each $s \in [1, 20]$.

2: We propose the minimizer $\hat{s}(T)$ of the sequence $s \in [1, 20] \rightarrow f(T, a(T, l(T, s), s), l(T, s), s)$ as a predictor of the width S.

3: We propose $\hat{l}(T) = l(T, \hat{s}(T))$ as a predictor of the width L.

4: We propose $\hat{a}(T) = a(T, l(T, \hat{s}(T)), \hat{s}(T))$ as a predictor of the amplitude A.

Nelder-Mead simplexe algorithm iterated using `optim()` function

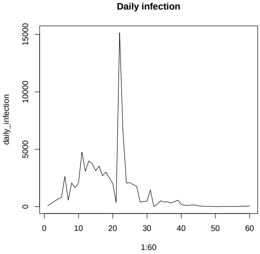
Algorithm 3 This new algorithm is also not suitable and converges to another local minimizer of our objective function.

Require: $T \geq 4$, init

1: Calculate the minimizer $\texttt{optim}(\text{init}(s), f(s+1, \cdot))$ for each $s = 3, \ldots, T-1$ with $\text{init}(3) = \text{init}$.

Ensure: $(\hat{a}(T), \hat{l}(T), \hat{s}(T)) = \texttt{optim}(\text{init}(T-1), f(T, \cdot))$.

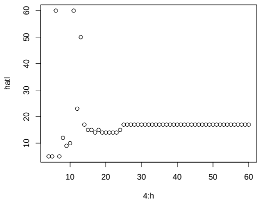
Application with real data: China

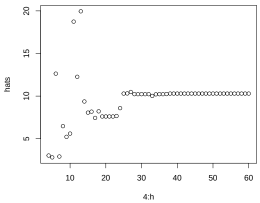


Let us consider the daily infections $t \in [1, 60] \to y(t)$ of COVID-19 in China. The sequence $T \in [4, h] \to \hat{I}(T)$, with $\tau_1 = 4$; $\tau_2 = 1$; $\tau_3 = 1$; $h = 35$, equals

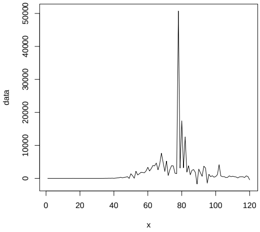
$$\hat{I}(4:35): 5, 5, 35, 5, 12, 9, 10, 35, 23, 35, 17, 15, 15, 14, 15, 14$$
$$14, 14, 14, 14, 15, 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 17.$$

The entry time in the zone of the peak:15. The exit time from the zone of the peak:25


hatl
4:h


hats
4:h

Application with real data: France

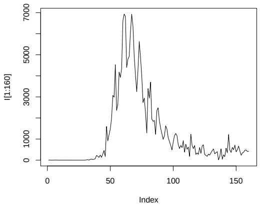


$$\tau_1 = 4;\ \tau_2 = 1;\ \tau_3 = 1;\ h = 85$$

$\hat{l}(50:85)$: 90, 66, 52, 90, 56, 63, 75, 67, 63, 69, 73, 64, 72, 73, 78, 82, 78, 74, 78, 74, 74, 75, 73, 67, 69, 69, 67, 69, 70, 70, 71, 71, 71, 71, 71, 71, 72, 72, 72, 72.

The entry time in the zone of the peak:69
The exit time from the zone of the peak: 81
We recall that the peak happens at the location 78.

Application with real data: Germany

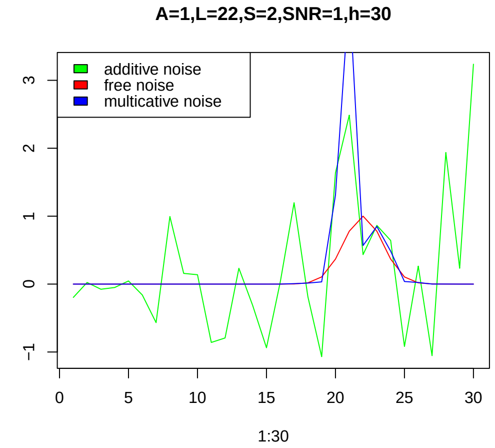


$\hat{I}(40:75): 75, 40, 42, 41, 44, 75, 44, 72, 75, 75, 75,$
$75, 75, 75, 75, 65, 54, 59, 58, 59, 60, 75, 70, 68, 62, 62,$
$63, 65, 68, 65, 64, 64, 64, 65, 65, 66$

the entry time in the region of the peak: 48
The exit time from the region of the peak: 69
We recall that the peak happens at the location 61.

Multiplicative noise model

Let A, L, S three positive numbers and t a positive integer. Let us corrupt $A\exp(-\frac{(t-L)^2}{S^2})$, by the random number $\exp(e(t))$. We consider the observation $y(t) = A\exp(-\frac{(t-L)^2}{S^2} + e(t))$.



COVID-19 in China:entry time and exit time

The entry time in the peak zone is equal to 14, while the exit time from the zone of the peak is 24.

Conclusion

Our algorithm 1 is able to predict in finite time the exact location, width and amplitude when the noise is moderate. If the noise is strong, then our algorithm is able to detect the entry time in the region of the peak and the exit time from the region of the peak. What is important in this work is its simplicity and its applicability to real data.

References

A. Dermoune, D. Ounaissi and Y. Slaoui.
Confidence intervals from local minimums of objective function.
Statistics, Optimization and Information Computing,
11 : 798-810. (2023).

A. Dermoune, D. Ounaissi and Y. Slaoui.
Gaussian and Lerch Models for Unimodal Time Series Forcasting.
Entropy,
25 : doi: 10.3390/e25101474.

Tucker Reed Stewart.
Peak Volume Prediction via Time Series Decomposition.
Bachelor of Science Computer Science and Systems,
June 2020 : University of Washington Tacoma.

B.Yu, G. Graciani, A. Nascimento,J. Hu.
Cost-adaptive neural networks for peak volume prediction.
Proceedings of Beijing 19: ACM International Conference
on Information and Knowledge Management.