



HAL
open science

A Hawkes process to make aware people of the severity of COVID-19 outbreak: application to cases in France

Laurent Lesage

► **To cite this version:**

Laurent Lesage. A Hawkes process to make aware people of the severity of COVID-19 outbreak: application to cases in France. [Research Report] Université de Lorraine; University of Luxembourg. 2020. hal-02510642

HAL Id: hal-02510642

<https://hal.science/hal-02510642>

Submitted on 18 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Hawkes process to make aware people of the severity of COVID-19 outbreak: application to cases in France

Laurent Lesage

March 18, 2020

Plan

1	Method	2
2	Results	3
2.1	Step 1: fitting on data until March 8 th	3
2.2	Step 2: back-testing on data from March 8 th to March 16 th	5
2.3	Step 3: simulation until March 26 th	6
2.3.1	Scenario 1: dynamics remain the same	6
2.3.2	Scenario 2: France shuts down the country like China partially did	6
3	Conclusion	10

Abstract We propose a simple model based on Hawkes processes to understand COVID-19 contagion, and to make people aware of the seriousness of the situation. Hawkes processes could show that this COVID-19 outbreak is a *self-exciting* phenomenon. We present results for cases observed in France: fitting on data of first weeks (until March 8th), then simulation over ten days to back-test our model (until March 16th) and over ten days to estimate the future number of confirmed cases (until March 26th) for two scenarios: either dynamics remain the same or French government takes measures as China (successfully) did.

1 Method

Our objective is to show that the COVID-19 is a *self-exciting* phenomenon, in order to estimate the future number of cases in France over the next month with a simple model, and to make French people aware of the seriousness of the situation. The study in [AG20] uses a Poisson autoregressive model to fit the evolution of the outbreak in China, Iran, Italy, and South Korea. For our model, we use a Hawkes process.

The Hawkes process (HP) is named after Alan G. Hawkes [Haw71], who introduced the concept of "self-exciting" point processes. The HP is a point process which models the occurrence of infections. We denote $N(t)$ the HP, which represents the number of cases at time t . We also denote $\mathcal{H}(t)$ the historical data of the process until time t .

A point process is characterized by its conditional intensity function, $\lambda^*(\cdot)$, which is defined by:

$$\lambda^*(t) = \lim_{dt \rightarrow 0^+} \frac{\mathbb{E}[N(t+dt) - N(t) | \mathcal{H}(t)]}{dt}. \quad (1.0.0.1)$$

From this definition, we show that:

$$\mathbb{P}[N(t+dt) - N(t) = 1 | \mathcal{H}(t)] \simeq \lambda^*(t)dt,$$

which illustrates that in our case, the probability of observing a new case between t and $t+dt$ is proportional to $\lambda^*(t)$. For a HP, the conditional intensity function is of form:

$$\begin{aligned} \lambda^*(t) &= \lambda + \int_0^t \mu(t-u) dN(u), \\ &= \lambda + \sum_{k=1}^{N(t)} \mu(t-t_k), \end{aligned} \quad (1.0.0.2)$$

where:

- $\lambda > 0$ is the *background intensity*, which models the infection from foreign countries,
- $\mu : (0, +\infty) \rightarrow [0, +\infty)$ is the *excitation function*, which models the way each past contamination influences the point process,
- $\{t_1, \dots, t_{N(t)}\}$ the sequence of the past occurrences until time t .

The difference between HP and a homogeneous Poisson process is the sum on past occurrences, which represents the *self-exciting* part of the model. This term means that each occurrence "excites" the process by increasing the likelihood that another event occurs in a short period of time. That is what is happening during an outbreak: each new case could transmit the virus to other people and then cause new infections.

2 Results

We have applied the model to the data from the GitHub repository of the Center of Systems Science and Engineering (CSSE) from Johns Hopkins University, in the folder dedicated to COVID-19 data [Joh20], where we could find the time series of confirmed cases in France from January 22nd to March 16th.

2.1 Step 1: fitting on data until March 8th

We learn our model on data observed until March 8th. We choose a Gamma density function as the *excitation function*:

$$\mu(t) = \alpha \frac{t^{k_1-1} \exp(-\frac{t}{k_2})}{\Gamma(k_1) k_2^{k_1}}, \quad (2.1.0.3)$$

where $\alpha \in \mathbb{R}$, $k_1, k_2 > 0$ and Γ is the Gamma function. The choice of this function is common in literature. It is indeed a generalization of the exponential example, which is the simplest choice for a HP. Parameters λ , α , k_1 and k_2 are estimated by maximizing the likelihood of the HP, which could be written only in function of $\lambda^*(\cdot)$, thanks to an Expectation-Maximization (EM) algorithm suited for HP [Hal12].

Table 1 shows the values of estimated parameters:

Parameter	Estimated value
λ	0.1616
α	1.0518
k_1	1.0031
k_2	4.4141

Table 1: Estimated value of the HP parameters

The fact that k_1 is close to 1 means that the delay between the infection and the contagion is very short (there is no consensus on this question among scientists, the delay could vary from 0 to 10 days). Indeed it means that Gamma density is close to an exponential function. Consequently, each new case leads to a maximal jump of value α for the conditional intensity function. The fact that λ (*background intensity*) is lower than α (exciting part) means that the spread is mainly due to transmission from French people to French people, rather than infections from foreign countries, even if a first infection from outside is necessary to launch the outbreak.

To test whether $\{t_1, t_2, \dots\}$ follows a HP with intensity function $\lambda^*(\cdot)$, we could demonstrate that it is equivalent to test whether $\{\Lambda(t_1), \Lambda(t_2), \dots\}$ form a Process process of parameter 1, where Λ is the compensator of the HP, which is defined by $\Lambda(t) = \int_0^t \lambda^*(s) ds$ (see [Pat15] for a proof). We use the QQ-plot in Figure 2.1 below to check this assumption:

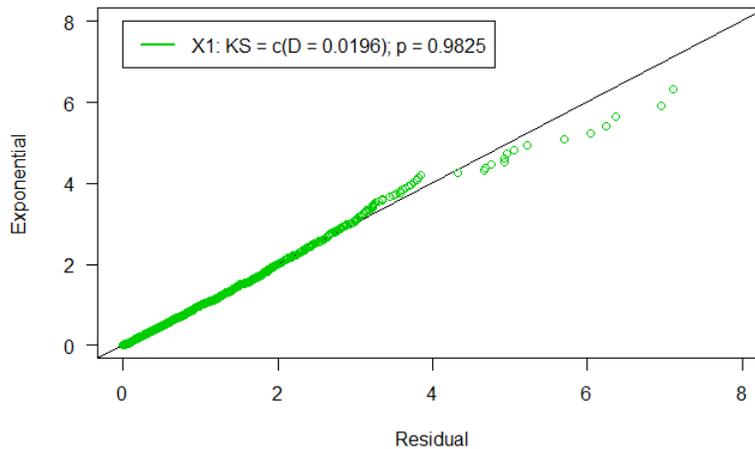


Figure 2.1: QQ-plot to check the goodness-of-fit of the model

The QQ-plot allows us to assume that the time series could be modelled by a HP. Moreover, we performed a Kolmogorov-Smirnov test on the two distributions. The maximal distance is $D = 0.0198$, with a p-value of $p = 0.9825$, which confirms our assumption.

Remark: Figure 2.2 shows the evolution of the number of infections in France from January 22nd to March 8th (right Y-axis), and the conditional intensity function (left Y-axis):

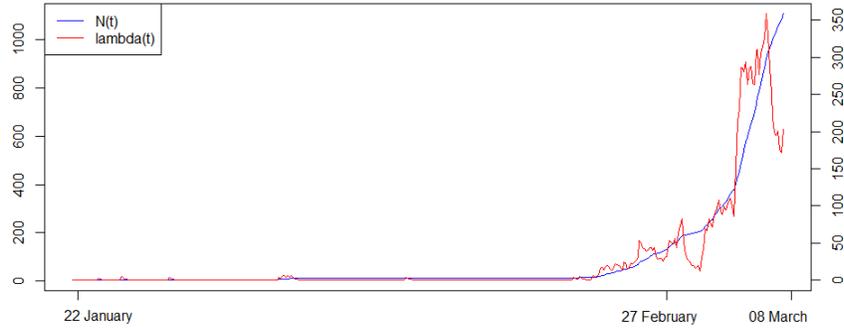


Figure 2.2: Number of infections in France from January 22nd to March 8th, and estimated conditional intensity function

The very first little peak of the conditional intensity function at January 24th corresponds to the first case in France. The high value of the conditional intensity function at March 8th shows that the outbreak is growing fast.

2.2 Step 2: back-testing on data from March 8th to March 16th

To back-test our model, we simulate the evolution of the outbreak from March 8th to March 16th, then we compare with real data. We simulate the HP thanks to Ogata's thinning algorithm [Oga81]. The idea of Ogata's algorithm is that randomly removing points from a "faster" simulated Poisson process (i.e. with a higher intensity function than targeted Hawkes process) allows to simulate a HP.

Figure 2.3 shows the comparison between the simulation and the real data:

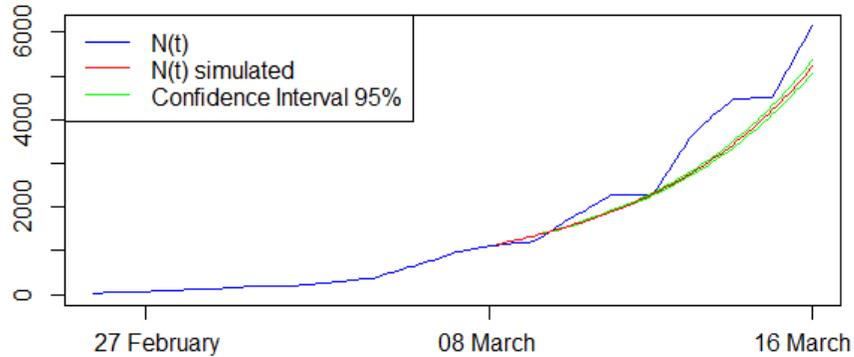


Figure 2.3: Simulation of the number of infections between March 8th and March 16th

Figure 2.3 shows that the model catches the trend of the outbreak evolution, and validates our approach. At March 16th, the estimation of the number of cases is 5,208.96. The confidence interval of 95% (CI 95%) is $[5,208.96 - 140; 5,208.96 + 140] = [5,068.96; 5,348.96]$. The actual number of infected people is 6,130.

2.3 Step 3: simulation until March 26th

The last step is the most interesting (and frightening) part. It allows us to have an estimation of the number of infections, in ten days (Thursday March 26th). We focus on two scenarios: either the dynamics of the outbreak remains the same, or French government takes measures similar to what Chinese authorities decided to.

2.3.1 Scenario 1: dynamics remain the same

In a first scenario, we estimate of the number of infections, if the dynamics of the outbreak remain the same, in ten days (Thursday March 26th), that's to say if French people have not changed their social interactions in the first weeks of March.

2.3.2 Scenario 2: France shuts down the country like China partially did

In a second scenario, we assume the French government takes similar measures than China at March 16th. To take into account these decisions, we adapt our HP model into a non-linear form:

$$\lambda^*(t) = \left[\lambda + \sum_{k=1}^{N(t)} \mu(t - t_k) \right] \times \min \left(1, \left(\frac{t}{t_m} \right)^a \right), \quad (2.3.2.4)$$

where t_m is the time when the measures are taken (i.e. March 16th) and $a > 1$ is a parameter to estimate. This model allows the conditional intensity function to be inhibited from t_m , thanks to the inhibiting part $\min \left(1, \left(\frac{t}{t_m} \right)^a \right)$.

To estimate a , we focus on data from the Chinese province of Jiangxi, which is close to the province of Hubei, the beginning of the outbreak. We fit a non-linear HP for these data. We denote:

$$\lambda_c^*(t) = \left[\lambda_c + \sum_{k=1}^{N_c(t)} \mu_c(t_c - t_{c,k}) \right] \times \min \left(1, \left(\frac{t}{t_{c,m}} \right)^{a_c} \right), \quad (2.3.2.5)$$

where index c refers to China, and then we will assume that $a_c = a$. To fit this model, we first fix $t_{c,m} = \text{February } 5^{\text{th}}$, which is the approximate date from when China decided to shut down the country. Then, we fit the parameters with data until February 5th, so that the inhibiting part is equal to 1. After getting the parameters λ_c and μ_c , we simulate the HP from February 5th to February 20th with a range of values for a_c . We choose the value of a_c which gives the estimated number of cases which is the closest to the actual number of cases at March 16th. Figure 2.4 represents the simulation of the fitted non-linear HP and the actual number of cases in Jiangxi, with the final choice for a_c parameter: $a_c = -1.403$.

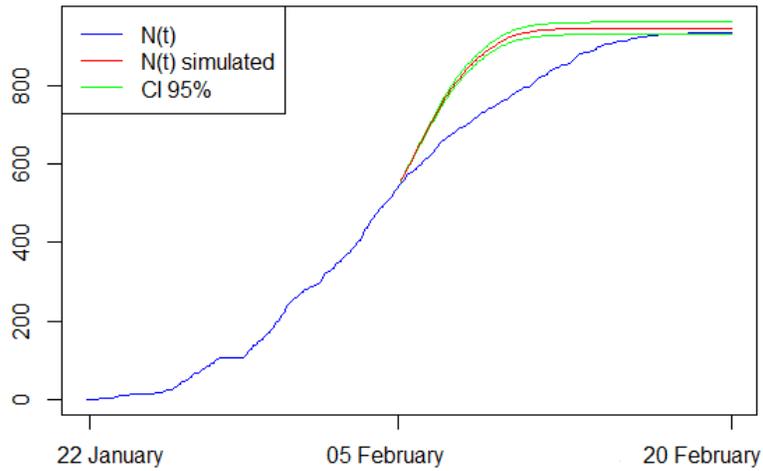


Figure 2.4: Calibration of a by simulating the number of infections between February 5th and February 20th for Chinese data

Then for the French process, we do the same simulation as the first scenario, but with the non-linear HP and with the parameters $t_m = \text{March } 16^{\text{th}}$ and $a = a_c = -1.403$ for the inhibiting part. Figure 2.5 shows the comparison between the first and the second scenario.

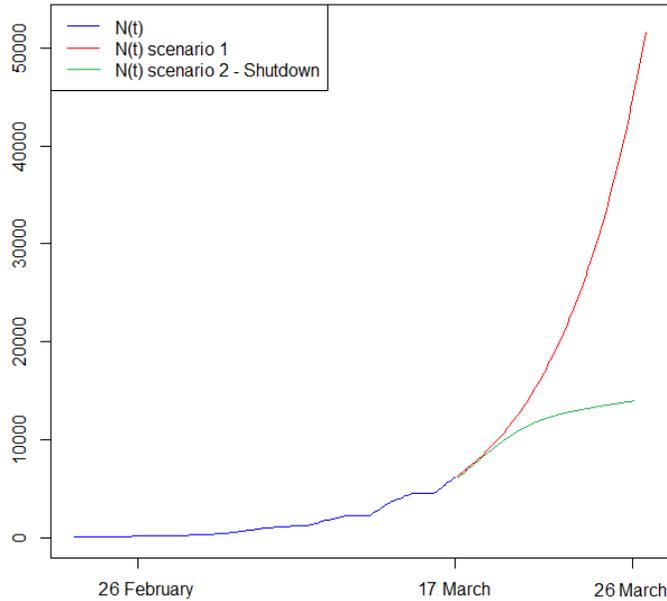


Figure 2.5: Simulation of the number of infections between March 16th and March 26th for the two scenarios

It shows that taking measures as in China would reduce significantly the number of estimated cases. We obtain an estimated number of cases of 51,642 cases in the first scenario, and 14,621 cases in the second scenario, approximately three times less!

But data we have worked on are just the number of official cases. We could reasonably think that the actual number of cases is much higher, because hospitals don't have enough doctors and material to test every people who would want to. Having COVID-19 symptoms is not enough to be tested (see [20a] for instance). According to [20b], 11,895 tests were performed on French population at March 9th, for 1,209 official cases, that's to say a rate of 10.1% positive tests. Let's assume that this rate is the same for our simulation. It means that $\frac{51,642}{10.1\%} = 511,306$ tests would have been performed, that's to say only 0.7% of the population! Consequently, the actual number of cases would be much higher than 18,529. It is hard to quantify, but we could imagine that doctors could only focus on one case of 3, or one case of 5.

3 Conclusion

As a conclusion, our model with a HP shows that COVID-19 seems to have a *self-exciting* dynamics. The simulations confirm that the evolution of the outbreak is exponential if people don't change their habits, and that it could be decreased significantly by taking similar measures than China. Everyone should become aware of the gravity of the situation and adopt the measures taken by the government to decrease the outbreak.

References

- [Haw71] Alan G. Hawkes. “Spectra of some self-exciting and mutually exciting point processes”. In: *Biometrika* 58 (1971), pp. 83–90.
- [Oga81] Ogata. “On Lewis’ simulation method for point processes”. In: *IEEE Transactions on Information Theory* 27 (1981), pp. 23–31.
- [Hal12] Peter F. Halpin. “An EM algorithm for Hawkes Process”. In: *Psychometrika* (2012).
- [Pat15] Philip K. Pollett Patrick J. Laub Thomas Taimre. “Hawkes Processes”. In: *arXiv preprint arXiv:1507.02822* (2015).
- [AG20] Agosto and Giudici. “A Poisson autoregressive model to understand COVID-19 contagion dynamics”. In: (2020).
- [20a] “<https://www.lemonde.fr/planete/article/2020/03/11/coronavirus-les-strategies-de-depistage-laissent-perplexes-patients-et-professionnels-6032554-3244.html>”. In: (2020).
- [20b] “<https://www.worldometers.info/coronavirus/covid-19-testing/>”. In: (2020).
- [Joh20] CSSE from John Hopkins University. <https://github.com/CSSEGISandData/COVID-19>. 2020.