

PRÉDICTION D'ÉPIDÉMIES DE GRIPPE EXTRÊMES

Maud Thomas ¹ & Holger Rootzén ²

¹ *Laboratoire de Probabilité, Modélisation et Statistiques, Sorbonne Université*

4 place Jussieu 75005 Paris

maud.thomas@sorbonne-universite.fr

² *Université technologique de Chalmers, SE-412 96 Göteborg, Suède*

hrootzen@chalmers.se

Résumé. Les épidémies de la grippe provoquent chaque année plus de 500,000 décès à l'échelle mondiale et présentent une forte morbidité, ce qui fait peser un fardeau supplémentaire sur des systèmes de santé déjà fragiles. Un des enjeux de la planification en santé publique est donc de prédire le risque de la survenue d'épidémies sévères ou extrêmes. Une épidémie est dite extrême lorsque le taux d'incidence dépasse un certain seuil (élevé).

Notre objectif est de prédire l'occurrence d'une épidémie extrême à court terme, par exemple dans les prochaines semaines. Pour cela, nous travaillons sur les taux d'incidence hebdomadaires de syndromes grippaux en France publiés par le réseau Sentinelles depuis 1985.

La théorie des valeurs extrêmes a été développée pour prédire, à partir d'une série d'observations, la probabilité d'événements plus extrêmes que ceux précédemment enregistrés. En nous basant sur l'approche proposée par Rootzén et al. (2018) et Kiriliouk et al. (2018), nous ajustons d'abord une loi de Pareto généralisée sur les premières semaines de l'épidémie. Puis, des estimations de la probabilité d'excès d'un seuil élevé sont obtenues à partir de la loi conditionnelle du modèle précédemment ajusté. Ces prédictions sont évaluées à partir de graphiques représentant les probabilités d'excès et de scores de Brier et comparées avec une régression logistique standard sur données simulées et réelles.

Mots-clés. Épidémiologie, Théorie des valeurs extrêmes, Prédiction

Abstract. Influenza viruses are responsible for annual epidemics, causing more than 500,000 deaths per year worldwide and also lead to high morbidity which puts substantial strain on public health systems. A crucial question for resource planning in public health is then to predict the morbidity burden of extreme epidemics. We say that an epidemic is extreme whenever the influenza incidence rate exceeds a high threshold for at least one week.

Our objective is to predict whether an extreme epidemic will occur in the near future, say the next couple of weeks. To that purpose, we collected the weekly numbers of influenza-like illness (ILI) incidence rates in France published by the Sentinelles network since 1985.

A main goal of Extreme Value Theory is to assess, from a series of observations, the probability of events that are more extreme than those previously recorded. Building on

the approach presented in Rootzén et al. (2018) and Kiriliouk et al. (2018), we first fit a 3-dimensional generalized Pareto distribution. Estimates of the probability of exceeding a high value are then obtained as the conditional distribution of the previously fitted model. The predictions are assessed by plots of the predicted probabilities and by Brier scores REF et and compared with a standard logistic regression on simulated and real data.

Keywords. Epidemiology, Extreme value analysis, Prediction

1 Structure du texte long

Les épidémies de la grippe provoquent chaque année plus de 500,000 décès à l'échelle mondiale et présentent une forte morbidité, ce qui fait peser un fardeau supplémentaire sur des systèmes de santé déjà fragiles. Un des enjeux de la planification en santé publique est donc de prédire le risque de la survenue d'épidémies sévères ou extrêmes. Une épidémie est dite extrême lorsque le taux d'incidence dépasse un certain seuil (élevé). Des études statistiques basées sur la théorie des valeurs extrêmes ont fourni des résultats prometteurs pour prédire des événements inhabituels en sciences environnementales (Katz et al., 2002) et en finance (Embrechts et al., 1997). En revanche, les applications au domaine de la santé publique sont peu nombreuses : la théorie des valeurs extrêmes a été utilisée par Guillou et Kratz (2013) pour détecter des valeurs aberrantes (vues ici comme des valeurs extrêmes), Chen et al. (2016) pour prédire l'incidence mensuelle des cas de grippe aviaire, et Thomas et al. (2016) pour prédire des épidémies de grippe extrêmes dans les prochaines années.

La théorie des valeurs extrêmes a été développée pour prédire, à partir d'une série d'observations, la probabilité d'événements plus extrêmes que ceux précédemment enregistrés. Par exemple, 40% de la superficie des Pays-Bas se trouvent sous le niveau de la mer et doivent donc être protégés par des digues. Il a été décidé que la hauteur de la digue devait assurer qu'il n'y ait pas plus d'une inondation tous les 10 000 ans. Les données disponibles ne couvrant qu'une période de 100 ans, la théorie des valeurs extrêmes permet de déterminer une hauteur de digue qui ne soit dépassée qu'une fois tous les 10 000 ans.

Notre objectif est de prédire l'occurrence d'une épidémie extrême à court terme, par exemple dans les prochaines semaines.

1.1 Données du réseau Sentinelles

Le réseau Sentinelles représente environ 1,500 médecins généralistes volontaires en France et Les taux d'incidence de syndromes grippaux hebdomadaires sont disponibles sur le site de réseau depuis 198. La Figure 1) présente les données de 1985 à 2017.

Le réseau Sentinelles définit les périodes selon le modèle de Serfling (Serfling, 1963). Nous proposons de définir la première semaine de l'épidémie comme la première des deux premières semaines consécutives ayant dépassé un certain seuil. Ce seuil a été choisi égal

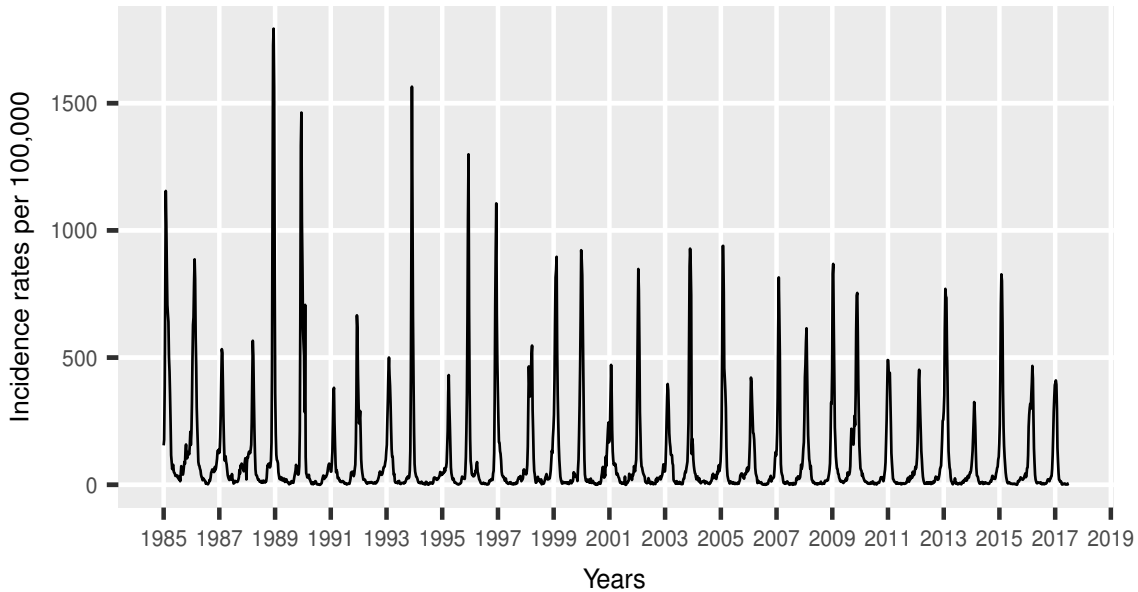


FIGURE 1 – Taux d’incidence de syndromes grippaux pour 100,000 en France de 1985 à 2017

au quantile d’ordre 0.3 des taux d’incidence, soit 255 pour 100,000. Cette définition permet de mieux capturer la partie la plus importante de l’épidémie (Figure 2).

1.2 Prédiction d’épidémies extrêmes à court terme

En nous basant sur l’approche proposée par Rootzén et al. (2018) et Kiriliouk et al. (2018), nous ajustons d’abord une loi de Pareto généralisée sur les taux d’incidence des semaines 1, 2 et 3 de l’épidémie et les semaines 1, 2 et la somme totale de l’épidémie. Notre but est donc d’estimer la probabilité que le taux d’incidence de la semaine 3 (ou que la somme totale des taux d’incidence) dépasse un certain seuil élevé sachant que l’on a observé les deux premières semaines de l’épidémie.

Ces probabilités d’excès d’un seuil élevé sont estimées à partir de la loi conditionnelle du modèle précédemment ajusté. Afin de pouvoir comparer ces estimations à un modèle classique, nous effectuons également une régression logistique du taux d’incidence de la semaine 3 (et de la somme totale) sur ceux des semaines 1 et 2. Ces prédictions sont évaluées à partir de graphiques représentant les probabilités d’excès estimées et de scores de Brier définis comme

$$\frac{1}{N} \sum_{i=1}^N (\hat{p}_i - o_i)^2$$

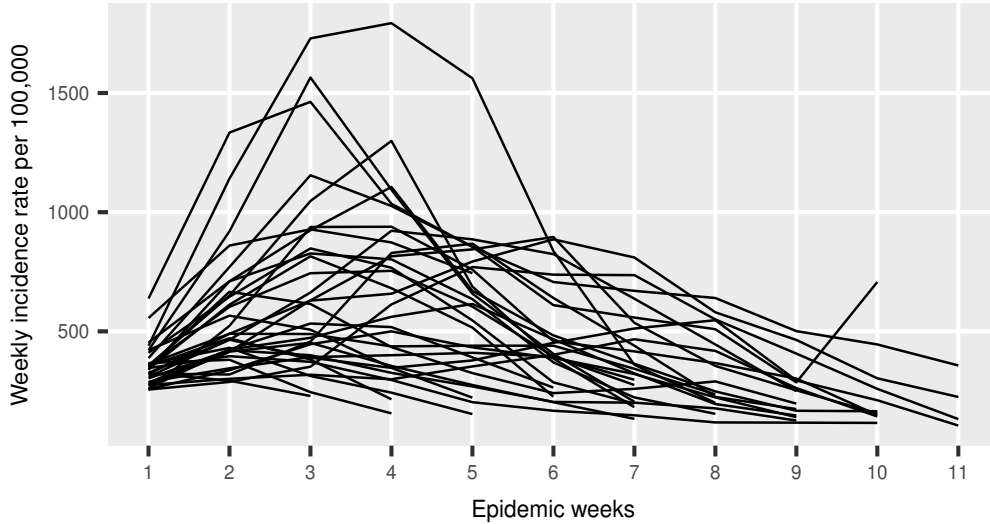


FIGURE 2 – Epidémies de 1985 à 2017 synchronisées sur la première semaine au-dessus du seuil 258.

où N est le nombre d'épidémies étudiées (ici 27), \hat{p}_i la probabilité d'excès estimée pour l'année i et $o_i = 1$ si l'excès a eu lieu l'année i et 0 sinon (Steyerberg et al., 2010). Ces comparaisons ont été réalisées sur données simulées et sur données réelles. La figure 3 compare la performance du modèle extrême ajusté avec la régression logistique pour prédire le dépassement du seuil $0.95 \times$ maximum observé sur 5000 données simulées. Les scores de Brier sont égaux à 0.015 pour le modèle extrême et à 0.014 pour le modèle logistique.

Ces résultats montrent que la validation de la prédiction d'événements extrêmes est une question délicate due au faible nombre de réalisations d'événements extrêmes. En effet, sur 5000 données simulées, seules 5% ont dépassé le seuil $0.95 \times$ le maximum observé. Ainsi, selon le score de Brier le modèle logistique semble être légèrement plus performant que le modèle extrême alors que la figure 3 montre que le modèle logistique peine à prédire le dépassement du seuil $0.95 \times$ le maximum observé.

1.3 Conclusion

La théorie des valeurs extrêmes permet à partir d'une série d'observations de prédire la probabilité d'événements plus extrêmes que ceux précédemment enregistrés. Un des enjeux de la planification en santé publique est de prédire le risque de la survenue d'épidémies sévères ou extrêmes. Grâce aux modèles proposés par Rootzén et al. (2018) et Kiriliouk et al. (2018), nous proposons une méthode permettant de prédire l'occurrence d'une épidémie

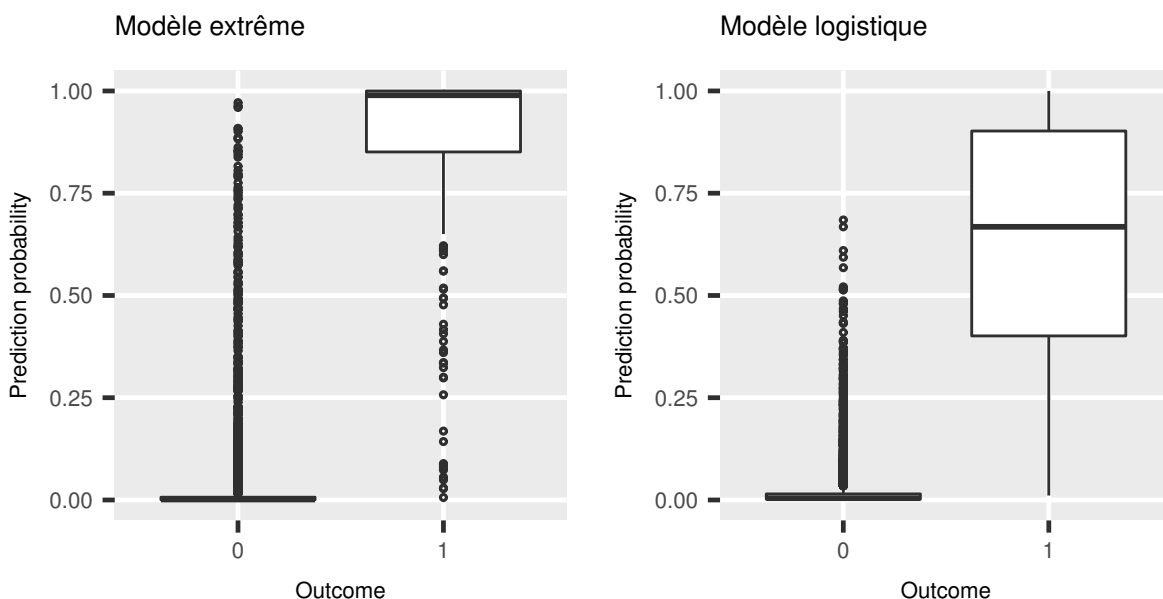


FIGURE 3 – Comparaison de la performance du modèle extrême et du modèle logistique pour la prédiction du dépassement du seuil $0.95 \times$ maximum observé sur 5000 données simulées.

extrême à partir de l’observation des premières semaines de l’épidémie, soit que le taux d’incidence dépasse un certain seuil (très) élevé au cours de l’épidémie. Nous évaluons la performance de notre méthode sur données simulées et sur données réelles à travers des graphiques représentant les probabilités d’excès estimées et des scores de Brier.

Ce travail met également en évidence que la validation d’événements extrêmes (donc peu ou non observés) reste une question délicate due au faible nombre de réalisations de ces événements.

Bibliographie

- Chen, J., Lei, X., Zhang, L. et Peng, B. (2015), Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China *PLoS ONE*, 10(2).
- Embrechts, P., Klüpperberg, C et Mikosch, T. (1997), *Modelling extremal events for insurance and finance*. Springer Verlag, Berlin.
- Guillou A., Kratz, M. et Le Strat, Y. (2014), An extreme value theory approach for the early detection of time clusters with applications to the surveillance of salmonella *Statistics in Medicine*, 33, pp 50015-5027.

- Katz, R. , Parlange, M. et Naveau, P. Statistics of extremes in hydrology (2002), *Advances in Water Resources*, 25, pp 1287–1304.
- Kiriliouk, A., Rootzén, H., Segers, J. et Wadsworth, J.L. (2018), Peaks over threshold modelling with multivariate generalized Pareto distributions, *To appear in Technometrics Réseau Sentinelles*, INSERM/Sorbonne Université, <http://www.sentiweb.fr>
- Rootzén, H., Segers, J. et Wadsworth, J.L. (2018), Multivariate peaks over threshold models, *Extremes*, 2(1), pp. 115-145.
- Serfling, E.R. (1963). Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public health reports*, 78(6), pp. 494.
- Steyerberg, E.W., Vickers, A.J., Cook, T.G., Gerds, T., Gonen, Obuchowski, N, Pencina, M.J. and Kattan, M.W. (2010), Assessing the performance of prediction models : a framework for traditional and novel measures , *Epidemiology*, 21(1), pp. 128-138.
- Thomas, M., Lemaitre, M., Wilson, M., Viboud, C., Yordanov, Y., Wackernagel, H. et Carrat, F. (2016). Applications of extreme value theory in public health. *PLoS ONE*, 11(7).