

Un apport de l'IA aux crises épidémiologiques Kiss the Covid

P. Mathieu^{a,1,*}, N. Mauhé^a

^aUniv. Lille, CNRS, Centrale Lille, UMR 9189 – CRISTAL Lab, F-59000 Lille, France

Abstract

Lors d'une crise sanitaire liée à l'apparition d'une maladie infectieuse, il est impératif d'avoir des systèmes d'aide à la décision des politiques publiques (évolution de la maladie, immunité collective, nombres dates et durées de périodes de confinement).

L'évolution d'une maladie infectieuse peut-être modélisée de différentes façons (approche centrée-individus, réseaux sociaux, equations mathématiques, modèles de flux).

Toutes ces modélisations necessite l'usage de paramètres pour calibrer le modèle. Plus il y a de paramètres, plus il est facile de coller aux données (tout comme il existe toujours un polynome de degrés n qui passe par n points).

Certains modélisateurs pensent que plus les paramètres sont précis (et nombreux), plus le modèle sera fiable. C'est un course sans fin.

L'important est de trouver l'essence du problème, le modèle le plus économe en paramètres (parcimonieux) rendant compte des phénomènes observés.

L'IA est d'une grande aide pour éviter les biais et régler au mieux ces modèles.

Keywords: Infectious disease, SIR modelling, Decision support system, artificial intelligence

2010 MSC: 00-01, 99-00

*Corresponding author

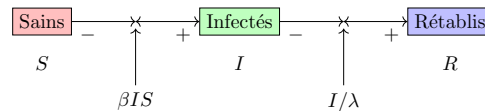
Email addresses: philippe.mathieu@univ-lille.fr (P. Mathieu),
nicolas.mauhe@univ-lille.fr (N. Mauhé)

1. Introduction (ce qui se fait en épidémiologie)

Depuis exactement un siècle et le papier séminal de Kermack et McKendrick, la modélisation épidémiologique a pris une part prépondérante dans l'étude de ces phénomènes complexes. La crise du H1N1 en 2008 a nettement amplifié ce phénomène, faisant apparaître un nombre important de modèles plus complexes les uns que les autres. Dans ce papier nous montrons que les techniques d'intelligence artificielle Utilisées à bon escient, Permettent d'une part de concevoir des modèles plus faciles à comprendre et à utiliser par les thématiques et d'autre part de faciliter grandement la conception d'outils d'aide à la décision pour les politiques sanitaires.

Le modèle présenté s'appuie sur des taux de contagion adaptable chaque jour, dont la valeur est automatiquement calculée par apprentissage génétique.

Le modèle SIR de Kermack et McKendrick [1] a été proposé en 1927 pour expliquer la hausse et la baisse rapides du nombre de patients infectés observées lors d'épidémies telles que la peste (à Londres en 1665, à Bombay en 1906) ou le choléra (à Londres en 1865). Il a été remis à l'avant-plan à la fin des années 1970 par Anderson et May [2].



2. L'apport de l'informatique

L'usage de l'IA est une méthode nouvelle pour l'épidémiologie, mais ancienne pour l'informatique.

Certains pensent que si on complexifie le modèle, on s'approche de la réalité : or plus il y a de paramètres, plus c'est facile de "fitter" : Problématique de l'overfitting (mettre une belle courbe qui passe par n points alors qu'en fait c'est une droite qui généralise le mieux).

Article	Type	Statut	Zone	Paramètres	Estimations proposées
Benvenuto et al. [1]	...	Published
Verity et al. [2]	...	Published
Zlojutro et al. [3]	...	Published
Chinazzi et al. [4]	...	Published
Scire et al. [5]	...	Published
Li et al. [6]	...	Published
Flaxman et al. [7]	Bayésien	Preprint	Europe	/	R0, Part de contaminés, Impact du confinement, Prévisions des décès
Ferguson et al. [8]	...	Preprint	
Di Domenico et al. [9]	...	Preprint	
Wilder et al. [10]	...	Preprint	
Simha et al. [11]	...	Preprint	
Roux et al. [12]	...	Preprint	
Pullano et al. [13]	...	Preprint	
Salje et al. [14]	...	Preprint	
Team and Murray [15]	...	Preprint	
Luo [16]	...	Preprint	
Branas et al. [17]	...	Preprint	
Keskinocak et al. [18]	...	Preprint	
Wang et al. [19]	...	Preprint	
Li et al. [20]	...	Preprint	
Woody et al. [21]	...	Preprint	
Pei and Shaman [22]	...	Preprint	
Platen [23]	...	Preprint	
Arenas et al. [24]	...	Preprint	
Sanche et al. [25]	...	Preprint	
Aleta et al. [26]	...	Preprint	

Table 1: Description des modèles épidémiologiques proposés durant l'épidémie de COVID-19

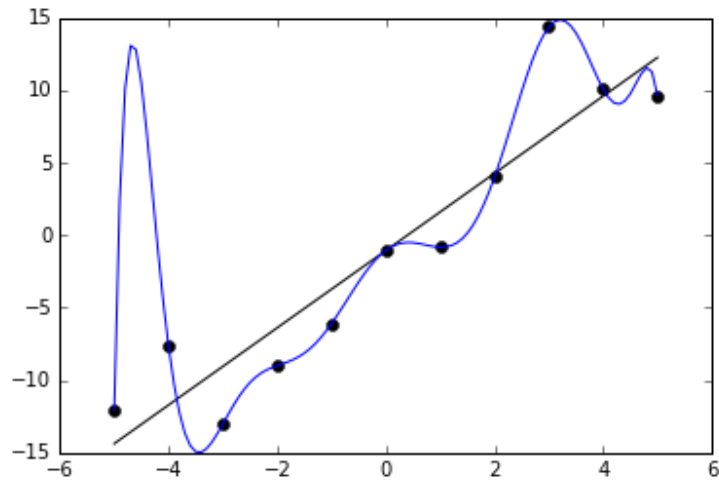


Figure 1: ©<https://en.wikipedia.org/wiki/Overfitting> Des données bruitées (approximativement linéaires) sont approximées par une fonction linéaire et une fonction polynomiale. Bien que la fonction polynomiale soit parfaitement ajustée, on constate que la fonction linéaire généralise surement bien mieux

Un modèle qui ne fonctionne pas (parce que par exemple des R0 sont délirants) peut très bien néanmoins “fitter” les données et donc continuer à prédire correctement les personnes en Réa

L’apport de l’informatique :

- 30 • des techniques et des outils : algos d’apprentissage
- s’attacher à l’essence du problème : principe de parcimonie

3. Cinq principes issus de l’informatique

*(Ici on peut décomposer en 1) définition du problème en épidémiologie puis
2) solutions apportées par l’informatique)*

35 3.1. Nécessité

Les modèles d’épidémie vont du simple modèle SIR que l’on peut résoudre sur un coin de feuille à des modèles réunissant des centaines d’agents indépendants,

placés sur des cartes reproduisant des pays entiers (REF nécessaire). Les progrès informatiques rendent possible toutes les constructions auxquelles les épidémiologistes
40 peuvent penser. Quel doit être le principe qui guide les choix des modélisateurs ?

Un principe central dans la littérature informatique est celui de *nécessité*. Chaque élément d'un modèle doit être nécessaire, et cette nécessité doit être justifiée.

45 3.2. Calibration

Tout modèle qui vise à estimer un élément de la réalité non accessible (comme le taux de propagation) ou à prédire un élément empirique doit être calibré sur des éléments réels. Si ce principe n'est pas respecté, il est possible de produire un modèle qui semble tout à fait répondre aux objectifs fixés mais qui décrit en
50 fait un tout autre phénomène. (Illustration modèle python).

Il s'agit d'un aspect central dans modélisation informatique et l'apprentissage informatique. Les principes de calibration issus de cette littérature peuvent être adaptés aux modèles épidémiologiques.

3.3. Approximation

55 Les possibilités induites par la modélisation donnent souvent l'illusion de la précision. On lit par exemple dans la presse des estimations très précises après la virgule produites par des modèles épiémiologiques portant sur des éléments très difficiles à estimer : c'est le cas par exemple des estimations du nombre d'infectés ou dut aux de propagation, souvent rapportées avec trois chiffres
60 après la virgule (REF 4,4% presse).

Ces analyses ignorent une question déterminante : le phénomène est-il capturable entièrement par des moyens computationnels ? Parfois le mécanisme à l'oeuvre est une dynamique changeante dont la prévision est hors de notre portée : c'est le cas des mutations possibles du COVID-19 au cours de l'épidémie,
65 ou de sa réaction face à l'environnement. C'est particulièrement vrai lorsque l'on s'attaque à des dynamiques sociales, déterminantes dans le calcul du taux

de propagation. La difficulté de modéliser les comportements humains et leur aggrégation au niveau de la société est un challenge considérable, qui nourrit une littérature particulièrement active parmi les disciplines informatiques. Il est possible de tirer de nombreux enseignements de cette littérature à destination de l'épidémiologie : le principal est sans doute le recul nécessaire relatif à la précision des estimations et des prévisions des modèles impliquant des décisions humaines (REF article JASSS).

Un autre aspect relativement méconnu du grand public relatif aux modèles épidémiologiques est la sensibilité aux paramètres initiaux. Un changement relativement faible des conditions initiales peut conduire à des modélisations très différentes. Il s'agit d'un aspect bien connu des disciplines informatiques, étudié notamment par les sciences de la complexité.

3.4. *Validation*

L'épidémie de COVID-19 peut être modélisée de manière très différente, même si l'on s'accordait sur les caractéristiques de base du virus. La question déterminante est donc de savoir si une modélisation est fidèle à la réalité qu'elle prétend simuler, voire prévoir. Il s'agit d'une problématique centrale en modélisation informatique, celle de la *validation*.

La littérature informatique peut nous aider à nous prémunir contre un certain nombre d'écueils, comme celui de l'*overfitting*. (...)

Pour aider à valider un modèle épidémiologique, plusieurs outils issus de la littérature informatique sont disponibles. Un exemple est celui de la *cross-validation*. (...)

3.5. *Répliquabilité*

4. Présence de ces principes dans les modèles actuels

[ANALYSE UNIQUEMENT POSITIVE]

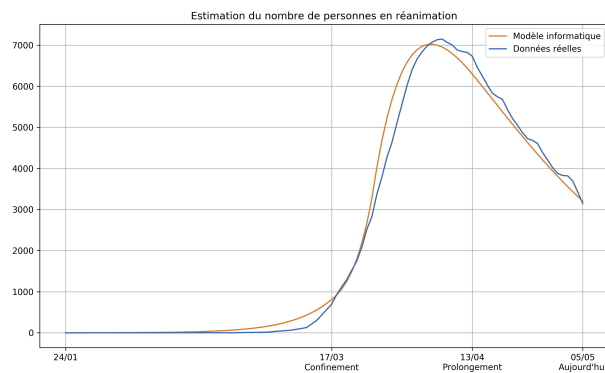
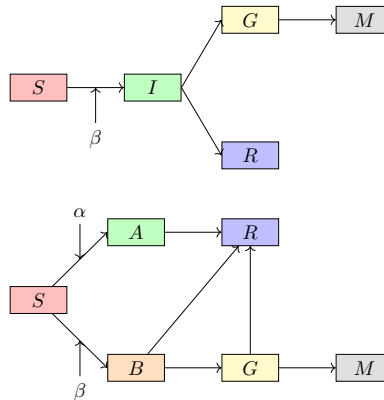
Article	Nécessité	Calibration	Approximation	Validation	Réplicabilité
Benvenuto et al. [1]
Verity et al. [2]
Zlojutro et al. [3]
Chinazzi et al. [4]
Scire et al. [5]
Li et al. [6]
Flaxman et al. [7]		✓	✓	✓	
Ferguson et al. [8]
Di Domenico et al. [9]
Wilder et al. [10]
Simha et al. [11]
Roux et al. [12]
Pullano et al. [13]
Salje et al. [14]
Team and Murray [15]
Luo [16]
Branas et al. [17]
Keskinocak et al. [18]
Wang et al. [19]
Li et al. [20]
Woody et al. [21]
Pei and Shaman [22]
Platen [23]
Arenas et al. [24]
Sanche et al. [25]
Aleta et al. [26]

Table 2: Présence des cinq principes dans les modèles épidémiologiques proposés durant l'épidémie de COVID-19

5. Hypothèse de simplicité

Avec quelque chose de très simple on “fitte” les données

95



References

- [1] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, Application of the ARIMA model on the COVID-2019 epidemic dataset, Data in Brief 29 (2020) 105340, ISSN 23523409, URL <https://linkinghub.elsevier.com/retrieve/pii/S2352340920302341>.
- [2] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá,

- R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: a model-based analysis, *The Lancet Infectious Diseases* (2020) S1473309920302437ISSN 14733099, URL <https://linkinghub.elsevier.com/retrieve/pii/S1473309920302437>.
- [3] A. Zlojutro, D. Rey, L. Gardner, A decision-support framework to optimize border control for global outbreak mitigation, *Scientific Reports* 9 (1) (2019) 2216, ISSN 2045-2322, URL <http://www.nature.com/articles/s41598-019-38665-w>.
- [4] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y. Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, *Science* 368 (6489) (2020) 395–400, ISSN 0036-8075, 1095-9203, URL <https://science.sciencemag.org/content/368/6489/395>.
- [5] J. Scire, S. Nadeau, T. Vaughan, G. Brupbacher, S. Fuchs, J. Sommer, K. N. Koch, R. Misteli, L. Mundorff, T. Götz, T. Eichenberger, C. Quinto, M. Savic, A. Meienberg, T. Burkard, M. Mayr, C. A. Meier, A. Widmer, R. Kuehl, A. Egli, H. H. Hirsch, S. Bassetti, C. H. Nickel, K. S. Rentsch, W. Kübler, R. Bingisser, M. Battegay, S. Tschudin-Sutter, T. Stadler, Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft, *Swiss Medical Weekly* 150 (1920), URL <https://smw.ch/article/doi/smw.2020.20271>.
- [6] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), *Science* 368 (6490) (2020) 489–493, ISSN 0036-8075, 1095-9203, URL <https://science.sciencemag.org/content/368/6490/489>.

- 135 [7] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan,
H. Zhu, T. Berah, J. Eaton, P. Perez Guzman, N. Schmit, L. Cilloni,
K. Ainslie, M. Baguelin, I. Blake, A. Boonyasiri, O. Boyd, L. Cattarino,
C. Ciavarella, L. Cooper, Z. Cucunuba Perez, G. Cuomo-Dannenburg,
A. Dighe, A. Djaafara, I. Dorigatti, S. Van Elsland, R. Fitzjohn, H. Fu,
140 K. Gaythorpe, L. Geidelberg, N. Grassly, W. Green, T. Hallett, A. Hamlet,
W. Hinsley, B. Jeffrey, D. Jorgensen, E. Knock, D. Laydon, G. Nedjati Gilani,
P. Nouvellet, K. Parag, I. Siveroni, H. Thompson, R. Verity, E. Volz,
C. Walters, H. Wang, Y. Wang, O. Watson, P. Winskill, X. Xi, C. Whittaker,
P. Walker, A. Ghani, C. Donnelly, S. Riley, L. Okell, M. Vollmer,
145 N. Ferguson, S. Bhatt, Report 13: Estimating the number of infections and
the impact of non-pharmaceutical interventions on COVID-19 in 11 European
countries, Report, URL <http://spiral.imperial.ac.uk/handle/10044/1/77731>, 2020.
- [8] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie,
150 M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg,
A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley,
L. Okell, S. Van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang,
P. Walker, P. Winskill, C. Whittaker, C. Donnelly, S. Riley, A. Ghani, Report 9: Impact of non-pharmaceutical
155 interventions (NPIs) to reduce COVID19 mortality and healthcare demand,
Tech. Rep., Imperial College London, URL <http://spiral.imperial.ac.uk/handle/10044/1/77482>, 2020.
- [9] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, V. Colizza,
Expected impact of lockdown in Île-de-France and possible exit strategies,
preprint, Infectious Diseases (except HIV/AIDS), URL <http://medrxiv.org/lookup/doi/10.1101/2020.04.13.20063933>, 2020.
160
- [10] B. Wilder, M. Charpignon, J. Killian, H.-C. Ou, A. Mate, S. Jabbari,
A. Perrault, A. Desai, M. Tambe, M. Majumder, The Role of Age Distribution
and Family Structure on COVID-19 Dynamics: A Preliminary

- 165 Modeling Assessment for Hubei and Lombardy, SSRN Electronic Journal
ISSN 1556-5068, URL <https://www.ssrn.com/abstract=3564800>.
- [11] A. Simha, R. V. Prasad, S. Narayana, A simple Stochastic SIR model
for COVID 19 Infection Dynamics for Karnataka: Learning from Europe,
arXiv:2003.11920 [math, q-bio] URL <http://arxiv.org/abs/2003.11920>,
170 arXiv: 2003.11920.
- [12] J. Roux, C. Massonnaud, P. Crépey, COVID-19: One-month im-
pact of the French lockdown on the epidemic burden, preprint, Epi-
demiology, URL [http://medrxiv.org/lookup/doi/10.1101/2020.04.](http://medrxiv.org/lookup/doi/10.1101/2020.04.22.20075705)
22.20075705, 2020.
- 175 [13] G. Pullano, E. Valdano, N. Scarpa, S. Rubrichi, V. Colizza, Population
mobility reductions during COVID-19 epidemic in France under lockdown
(2020) 22.
- [14] H. Salje, C. T. Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau,
A. Andronico, N. Hoze, J. Richet, C.-L. Dubost, Y. L. Strat, J. Lessler,
180 D. Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boëlle, S. Cauchemez, Esti-
mating the burden of SARS-CoV-2 in France (2020) 44.
- [15] I. C.-. h. s. u. f. Team, C. J. Murray, Forecasting the impact of the
first wave of the COVID-19 pandemic on hospital demand and deaths
for the USA and European Economic Area countries, medRxiv (2020)
2020.04.21.20074732URL [https://www.medrxiv.org/content/10.1101/](https://www.medrxiv.org/content/10.1101/2020.04.21.20074732v1)
185 2020.04.21.20074732v1.
- [16] J. Luo, Predictive Monitoring of COVID-19 (2020) 12.
- [17] C. C. Branas, A. Rundle, S. Pei, W. Yang, B. G. Carr, S. Sims, A. Ze-
browski, R. Doorley, N. Schluger, J. W. Quinn, J. Shaman, Flatten-
190 ing the curve before it flattens us: hospital critical care capacity limits
and mortality from novel coronavirus (SARS-CoV2) cases in US coun-

- ties, preprint, Epidemiology, URL <http://medrxiv.org/lookup/doi/10.1101/2020.04.01.20049759>, 2020.
- [18] P. Keskinocak, B. E. O. Aglar, A. Baxter, J. Asplund, N. Serban, The
195 Impact of Social Distancing on COVID19 Spread: State of Georgia Case
Study, medRxiv (2020) 2020.04.29.20084764 URL <https://www.medrxiv.org/content/10.1101/2020.04.29.20084764v1>.
- [19] L. Wang, G. Wang, L. Gao, X. Li, S. Yu, M. Kim, Y. Wang, Z. Gu,
Spatiotemporal Dynamics, Nowcasting and Forecasting of COVID-19 in
200 the United States, arXiv:2004.14103 [stat] URL <http://arxiv.org/abs/2004.14103>, arXiv: 2004.14103.
- [20] M. L. Li, H. T. Bouardi, O. S. Lami, N. Trichakis, M. F. Zarandi, D. Bert-
simas, Overview of DELPHI Model V2.8 - COVIDAnalytics (2020) 6.
- [21] S. Woody, M. G. Tec, M. Dahan, K. Gaither, M. Lachmann, S. Fox,
205 L. A. Meyers, J. G. Scott, Projections for first-wave COVID-19 deaths
across the US using social-distancing measures derived from mobile
phones, medRxiv (2020) 2020.04.16.20068163 URL <https://www.medrxiv.org/content/10.1101/2020.04.16.20068163v2>.
- [22] S. Pei, J. Shaman, Initial Simulation of SARS-CoV2 Spread and Interven-
210 tion Effects in the Continental US, preprint, Epidemiology, URL <http://medrxiv.org/lookup/doi/10.1101/2020.03.21.20040303>, 2020.
- [23] E. Platen, Stochastic Modelling of the COVID-19 Epidemic, SSRN Schol-
arly Paper ID 3586208, Social Science Research Network, Rochester, NY,
URL <https://papers.ssrn.com/abstract=3586208>, 2020.
- [24] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gómez, C. Granell, J. T.
215 Matamalas, D. Soriano-Panos, B. Steinegger, A mathematical model for
the spatiotemporal epidemic spreading of COVID19, medRxiv (2020)
2020.03.21.20040022 URL <https://www.medrxiv.org/content/10.1101/2020.03.21.20040022v1>.

- 220 [25] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, R. Ke,
High Contagiousness and Rapid Spread of Severe Acute Respiratory Syn-
drome Coronavirus 2, *Emerging Infectious Diseases journal* 26 (7), URL
https://wwwnc.cdc.gov/eid/article/26/7/20-0282_article.
- [26] A. Aleta, D. Martin-Corral, A. P. y. Piontti, M. Ajelli, M. Litvinova,
225 M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini, S. Merler,
A. Pentland, A. Vespignani, E. Moro, Y. Moreno, Modeling the impact
of social distancing, testing, contact tracing and household quarantine
on second-wave scenarios of the COVID-19 epidemic, *medRxiv* (2020)
2020.05.06.20092841URL [https://www.medrxiv.org/content/10.1101/](https://www.medrxiv.org/content/10.1101/2020.05.06.20092841v1)
230 2020.05.06.20092841v1.
- [27] R. M. May, R. M. Anderson, Population biology of infectious diseases: Part
II, *Nature* 280 (5722) (1979) 455–461.
- [28] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis,
J.-F. Pinton, N. Khanafer, W. Van den Broeck, et al., Simulation of an
235 SEIR infectious disease model on the dynamic contact network of confer-
ence attendees, *BMC medicine* 9 (1) (2011) 87.
- [29] W. Duan, Z. Fan, P. Zhang, G. Guo, X. Qiu, Mathematical and computa-
tional approaches to epidemic modeling: a comprehensive review, *Frontiers*
of Computer Science 9 (5) (2015) 806–826.
- 240 [30] W. O. Kermack, A. G. McKendrick, A contribution to the mathemati-
cal theory of epidemics, *Proceedings of the royal society of london. Series*
A, Containing papers of a mathematical and physical character 115 (772)
(1927) 700–721.
- [31] M. J. Keeling, P. Rohani, Modeling infectious diseases in humans and ani-
245 mals, Princeton University Press, 2011.
- [32] C. Kamp, M. Moslonka-Lefebvre, S. Alizon, Epidemic spread on weighted
networks, *PLoS computational biology* 9 (12).

- [33] B. Wilder, M. Charpignon, J. A. Killian, H.-C. Ou, A. Mate, S. Jabbari, A. Perrault, A. Desai, M. Tambe, M. S. Majumder, The role of age distribution and family structure on COVID-19 dynamics: A preliminary modeling assessment for hubei and lombardy, Available at SSRN 3564800 .
- [34] B. M. Ndiaye, L. Tendeng, D. Seck, Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting, arXiv preprint arXiv:2004.01574 .
- [35] J.-Q. Liao, X.-B. Hu, M. Wang, M. S. Leeson, Epidemic modelling by ripple-spreading network and genetic algorithm, Mathematical Problems in Engineering 2013.
- [36] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, Application of the ARIMA model on the COVID-2019 epidemic dataset, Data in Brief 29 (2020) 105340, ISSN 2352-3409.
- [37] N. Pearce, F. Merletti, Complexity, simplicity, and epidemiology, 2006.
- [38] S. A. Levin, T. G. Hallam, L. J. Gross, Applied mathematical ecology, vol. 18, Springer Science & Business Media, 2012.
- [39] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu, et al., Estimates of the severity of coronavirus disease 2019: a model-based analysis, The Lancet Infectious Diseases .
- [40] D. Chicco, Ten quick tips for machine learning in computational biology, BioData mining 10 (1) (2017) 35.