On big data models for infectious disease forecasting - Commentary on "Transparent modeling of influenza incidence: Big data or single data point from psychological theory?"

Souhaib Ben Taieb[1]

Big Data and Machine Learning Lab, Université de Mons (UMONS), Belgium

Kathryn S. Taylor

Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Infectious disease forecasting is challenging notably due to the complexity of diseases and population dynamics, data paucity, and the prophet dilemma (Lauer et al., 2020). In an influenza forecasting exercise, Katsikopoulos et al. found that the simple recency heuristic provides better forecasts than Google Flu Trends, a black box machine learning algorithm [1]. The authors give various other forecasting applications where simple heuristics have outperformed ML models, including U.S. presidential elections, consumer purchases, and terrorist attacks. They highlight the fact that complex ML methods tend to overfit the past, and are unable to deal with quickly-changing situations. As a result, they advocate using simple and transparent heuristics, based on psychological theory, as benchmarks when testing modern complex and black-box ML algorithms.

With the recent development and proliferation of new complex and black-box big data models, the authors discuss important issues related to model overparameterization and the lack of model transparency. We think Katsikopoulos et al essentially invite us to revisit the calls by Box (1976) for model parsimony, and avoid excessive elaboration and overparameterization. We join the authors in encouraging the systematic comparison of complex ML methods with simple heuristics in infectious disease forecasting, and other forecasting applications with high uncertainty. However, given the recent progress in interpretable and explainable AI, and the excellent performance of modern ML models in various predictive modelling applications, we argue that these models should not be undermined. In this note, we would like to draw attention to certain topics that could help better understand the differences between simple heuristics and complex ML methods. Although we focus on infectious disease forecasting, similar observations can be made in other similar forecasting applications.

Katsikopoulos et al only present point forecasts, but there will be inherent uncertainty associated with these forecasts. Gneiting & Katzfuss (2014) describe how optimal decision-making relies on probabilistic forecast, rather than just a single point forecast. Probabilistic forecasting is essential in many applications, including smart grid operations (Ben Taieb et al., 2020), economic and financial risk management (Groen et al., 2013), and demographic projection (Raftery et al., 2012). There have been also calls for probabilistic forecasting in epidemiological models, for example, from Ray et al. (2020) and also Ioannidis et al. (2020), who are cited by the authors. In the context of the COVID-19 pandemic, Taleb et al. (2020) also pointed out how inadequate point forecasts are as input to decision-making when the underlying processes are highly uncertain and complex. Furthermore, Katsikopoulos et al only present the results for a single time series. Without seeing data on the accuracy of their heuristic compared with that of a number of different time series and ML methods, it is hard

[1]We use big data, machine learning (ML) and artificial intelligence (AI) interchangeably.

to draw general conclusions.

ML methods often struggle to beat simple benchmarks especially on short and highly noisy time series. However, there have been significant recent progress in ML and neural forecasting methods (Benidis et al., 2020; Hewamalage et al., 2021). In particular, when forecasting a group of time series, a recent trend in ML is to build a single (global) model for all series (Mariet & Kuznetsov, 2019; Montero-Manso & Hyndman, 2020). This is different from the (classical) local approach where a different model is trained for each series. A major advantage of the global approach is that the global model can afford to be more complex with less chance of overfitting. For example, top entries in the M4 forecasting competition have used global models (Makridakis et al., 2020). These global models have been shown to perform well even on heterogeneous groups of time series (Montero-Manso & Hyndman, 2020). In infectious disease forecasting, a global "complex" ML model could be trained for example using data from different geographic regions. By exploiting the fact that a disease spreads at different speeds and on a different scale in different regions, a global model could extract similar progression patterns in these regions within different time intervals.

Katsikopoulos et al used the concept of bias and variance tradeoff to explain the better accuracy of simple heuristics compared to complex forecasting methods. Specifically, while simple heuristics have a high bias, their low variance often lead to smaller prediction errors. For complex ML methods, their low bias does not generally compensate their high variance. As a result, it often induces an overfitting of the training data, and leads to higher prediction errors. The classical bias and variance error decomposition is a useful tool to study and compare sources of forecast errors (Ben Taieb & Atiya, 2015). However, some of the observed learning behaviors of modern ML methods can not be explained using the classical bias and variance tradeoff (Belkin et al., 2019; Bartlett et al., 2020). Specifically, it is possible to train a complex ML model (e.g, a neural network) to exactly fit (i.e., interpolate) the data, and still obtain good or even better (out-of-sample) test prediction accuracy. In other words, overparametrization can play a beneficial role in the interpolation regime. This phenomenon is referred to as "double descent" and "benign overfitting". To the best of our knowledge, there are no studies on the impact of this phenomenon on modern ML forecasting methods, especially when global ML models are used.

Katsikopoulos et al criticize SIR-type disease models [2], but the performance of their heuristic is not compared with any SIR-type model. In rejecting SIR-type models in favour of their own heuristic, the authors cite Ioannidis et al. (2020), who have highlighted the challenges of modelling the effects of COVID-19, as so much is unknown about this novel disease. It would have been nice to have some consideration of the studies that have used SIR-type models of influenza. In contrast to the situation with COVID-19, there is a considerable volume of historical data on the influenza virus. Simple SIR-type models are also transparent. In fact, Katsikopoulos et al question SIR-type models in such uncompromising terms, without accepting their established role in epidemiological modelling. These models are not simply used to make forecasts under the current circumstances, but by changing their assumptions, they can also provide an experimental platform to obtain insight into the effects of changes to the current scenario. Recently, hybrid models which integrate ML into SIR type models have been shown to improve forecast accuracy in predicting the progression of COVID-19, while providing explainable models (Arik et al., 2020).

[2] Includes SIR and SEIR models which have compartments representing those of the population who are Susceptible, Exposed, Infected and Removed or Recovered from the disease.

The authors have focused on identifying the best model and discussing the best type of model, but of course another alternative would be to combine forecasts from multiple different models. Aggregation can incorporate information underlying different prediction methods in a pragmatic way, diversify the risk inherent in relying on a single model, and offset the statistical bias associated with individual models, with overestimation and underestimation potentially cancelling out (Bates & Granger, 1969). Given the uncertainties in epidemiological modelling, aggregation makes sense, and there is an increasing number of epidemiological applications of forecast aggregation, aiming to produce more accurate forecasts (Lutz et al., 2019), for example with influenza models (Yamana et al., 2017; Reich et al., 2019) and, more recently, COVID-19 models (Ray et al., 2020; Taylor & Taylor, 2020). The COVID-19 Hub ensemble forecast is a simple average aggregation of probabilistic forecasts from time-series, SIR-type, ML, and other models (Ray et al., 2020). In the influenza ensemble model, Yamana et al. (2017) produce a weighted average, with weights based on records of historical accuracy of the individual component models. Weighting forecasts and comparing the accuracy of aggregated forecasts against the accuracy of the individual forecasts are not straightforward for COVID-19 models, as COVID-19 is a novel disease and subsequently, there are limited data on historical accuracy. Both the ensemble influenza models described by Yamana et al. (2017) and Reich et al. (2019) were reported as being more accurate than the individual models. McGowan et al. (2019) have also found that ensemble forecasting techniques consistently outperformed simple benchmarks.

One would expect the COVID-19 pandemic to have an impact on future influenza infections, and on influenza forecasting. Our world has changed, as social distancing, shielding, and increased attention to hand hygiene have become the norm. These measures, along with government lockdown restrictions, have led to fewer reported cases of influenza in England (Iacobucci, 2020). It is possible that the winter peak of influenza infections this winter will be markedly reduced and that the peak of deaths by influenza will be masked by COVID-19. Therefore, the current and future influenza years could be very different from the years prior to 2020. It is not clear how COVID-19 will impact the relative performance of sophisticated methods and simplistic benchmarks. Uncertainties about how the pandemic might evolve in 2021, and beyond, present various challenges (Scudellari, 2020).

We believe more efforts is needed to explore the strength and weaknesses of simple heuristics and ML methods for infectious disease forecasting. Forecasting competitions such as FluSight for Influenza (McGowan et al., 2019) or RAPIDD for Ebola (Viboud et al., 2018) have been very useful in that regard. These competitions highlight the importance of probabilistic forecasting and seem to indicate that forecast combination leads to consistently better forecast accuracy. Finally, as pointed out by Saltelli et al. (2020), being aware of model assumptions and model ignorance while aiming for an appropriate model complexity will ensure that our forecasting models will better serve society.

Acknowledgment

James W. Taylor provided helpful comments. The opinions expressed of course remain our own.

References

Arik, S., Li, C.-L., Yoon, J., Sinha, R., Epshteyn, A., Le, L., Menon, V., Singh, S., Zhang, L., Nikoltchev, M., & Others (2020). Interpretable sequence learning for COVID-19 forecasting. *Advances in neural information* *processing systems*, *33*.

Bartlett, P. L., Long, P. M., Lugosi, G., & Tsigler, A. (2020). Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 30063–30070.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *The Journal of the Operational Research Society*, *20*, 451–468.

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 15849–15854.

Ben Taieb, & Atiya, A. F. (2015). A bias and variance analysis for Multistep-Ahead time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, *PP*, 1–1.

Ben Taieb, S., Taylor, J. W., & Hyndman, R. J. (2020). Hierarchical probabilistic forecasting of electricity demand with smart meter data. *Journal of the American Statistical Association*, (pp. 1–17).

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, B., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Callot, L., & Januschowski, T. (2020). Neural forecasting: Introduction and literature overview, . `arXiv:2004.10240`.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*, 791–799.

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*, 125–151.

Groen, J. J. J., Paap, R., & Ravazzolo, F. (2013). Real-Time inflation forecasting in a changing world. *Journal of business & economic statistics: a publication of the American Statistical Association*, *31*, 29–44.

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International journal of forecasting*, *37*, 388–427.

Iacobucci, G. (2020). Covid lockdown: England sees fewer cases of colds, flu, and bronchitis. *BMJ*, *370*, m3182.

Ioannidis, J. P. A., Cripps, S., & Tanner, M. A. (2020). Forecasting for COVID-19 has failed. *International journal of forecasting*, .

Lauer, S. A., Brown, A. C., & Reich, N. G. (2020). Infectious disease forecasting for public health, . `arXiv:2006.00073`.

Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., Fernandez, D., Greene, S. K., Kipshidze, N., Liu, L., Mgbere, O., McHugh, L. A., Myers, J. F., Siniscalchi, A., Sullivan, A. D., West, N., Johansson, M. A., & Biggerstaff, M. (2019). Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. *BMC public health*, *19*, 1659.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International journal of forecasting*, *36*, 54–74.

Mariet, Z., & Kuznetsov, V. (2019). Foundations of Sequence-to-Sequence modeling for time series. In K. Chaudhuri, & M. Sugiyama (Eds.), *Proceedings of Machine Learning Research* (pp. 408–417). PMLR volume 89 of *Proceedings of Machine Learning Research*.

McGowan, C. J., Biggerstaff, M., Johansson, M., Apfeldorf, K. M., Ben-Nun, M., Brooks, L., Convertino, M., Erraguntla, M., Farrow, D. C., Freeze, J., Ghosh, S., Hyun, S., Kandula, S., Lega, J., Liu, Y., Michaud, N., Morita, H., Niemi, J., Ramakrishnan, N., Ray, E. L., Reich, N. G., Riley, P., Shaman, J., Tibshirani, R., Vespignani, A., Zhang, Q., Reed, C., & Influenza Forecasting Working Group (2019). Collaborative efforts to forecast seasonal influenza in the united states, 2015-2016. *Scientific reports*, *9*, 683.

Montero-Manso, P., & Hyndman, R. J. (2020). Principles and algorithms for forecasting groups of time series: Locality and globality, . `arXiv:2008.00444`.

Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 13915–13921.

Ray, E. L., Wattanachit, N., Niemi, J., Kanji, A. H., House, K., Cramer, E. Y., Bracher, J., Zheng, A., Yamana, T. K., Xiong, X., Woody, S., Wang, Y., Wang, L., Walraven, R. L., Tomar, V., Sherratt, K., Sheldon, D., Reiner, R. C., Aditya Prakash, B., Osthus, D., Li, M. L., Lee, E. C., Koyluoglu, U., Keskinocak, P., Gu, Y., Gu, Q., George, G. E., España, G., Corsetti, S., Chhatwal, J., Cavany, S., Biegel, H., Ben-Nun, M., Walker, J., Slayton, R., Lopez, V., Biggerstaff, M., Johansson, M. A., Reich, N. G., & COVID-19 Forecast Hub Consortium (2020). Ensemble forecasts of coronavirus disease 2019 (COVID-19) in the U.S. *medRxiv*, (p. 2020.08.19.20177493).

Reich, N. G., McGowan, C. J., Yamana, T. K., Tushar, A., Ray, E. L., Osthus, D., Kandula, S., Brooks, L. C., Crawford-Crudell, W., Gibson, G. C., Moore, E., Silva, R., Biggerstaff, M., Johansson, M. A., Rosenfeld, R., & Shaman, J. (2019). Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. *PLoS computational biology*, *15*, e1007486.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke, R., Jr, Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., Stirling, A., van der Sluijs, J., & Vineis, P. (2020). Five ways to ensure that models serve society: a manifesto. *Nature*, *582*, 482–484.

Scudellari, M. (2020). How the pandemic might play out in 2021 and beyond. *Nature*, *584*, 22–25.

Taleb, N. N., Bar-Yam, Y., & Cirillo, P. (2020). On single point forecasts for fat-tailed variables. *International journal of forecasting*, .

Taylor, K. S., & Taylor, J. W. (2020). A comparison of aggregation methods for probabilistic forecasts of COVID-19 mortality in the united states, . `arXiv:2007.11103`.

Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., Vespignani, A., & RAPIDD Ebola Forecasting Challenge group (2018). The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, *22*, 13–21.

Yamana, T. K., Kandula, S., & Shaman, J. (2017). Individual versus superensemble forecasts of seasonal influenza outbreaks in the united states. *PLoS computational biology*, *13*, e1005801.