

Why Can't Neural Networks Forecast Pandemics Better

Joshua D. Zelek
John S. Zelek
Alex Wong
Email: {jzelek@princeton.edu, jzelek@uwaterloo.ca, a28wong@uwaterloo.ca }

Princeton University
University of Waterloo
University of Waterloo

Abstract

Why can't Neural Networks (NN) forecast better? In the major super-forecasting competitions, NN have typically under-performed when compared to traditional statistical methods. When they have performed well, the underlying methods have been ensembles of NN and statistical methods. Forecasting stock markets, medical, infrastructure dynamics, social activity or pandemics each have their own challenges. In this study, we evaluate the strengths of a collection of methods for forecasting pandemics such as Covid-19 using NN, statistical methods as well as parameterized mechanistic models. Forecasts of epidemics can inform public health response and decision making, so accurate forecasting is crucial for general public notification, timing and spatial targeting of intervention. We show that NN typically under-perform in forecasting Covid-19 active cases which can be attributed to the lack of training data which is common for forecasts. Our test data consists of the top ten countries for active Covid-19 cases early in the pandemic and is represented as a Time Series (TS). We found that Statistical methods outperform NN for most cases. Albeit, NN are still good pattern finders and we suggest that there are perhaps more productive ways other than purely data driven models of using NN to help produce better forecasts.

1 Introduction

"Prediction is very difficult, especially about the future" (Nils Bohr).

A TS is a collection of data where the independent variable is time. TS forecasting is an important application for financial analysis (e.g., stock markets, sales, economic), medical analysis (e.g., ECG, drug reactions, pandemics, elderly care), civil infrastructure (e.g., bridge vibrations, road conditions, electricity demand), environmental monitoring (e.g., air, water pollutant monitoring), sensor network monitoring, social activity mining to name a few. The forecasting of TS data helps to provide useful information for important decision making. Methods for forecasting include traditional methods such as parametric models informed by domain expertise (either compartmental or network), statistical methods and recently NN data driven models including deep learning methods. Forecasts can be either one-step-ahead or a multi-horizon TS variety.

The consensus in the forecasting community is that "neural networks (and other highly non-linear and non-parametric methods) are not well suited to TS forecasting due to the relatively short nature of most time series" (Rob Hyndman on M-challenges 2018). The M-competitions (<https://mofc.unic.ac.cy/m4>) are the most influential forecasting competitions and the testing data consists of 100K series of various frequencies. In general, NN are great for learning complex patterns when sufficient numbers of TS are available to learn from however these conditions do not hold for all applications including forecasting. Forecasting sometimes is geographically and temporally localized meaning that only one TS is available for learning which is not desirable for a NN to function well. NN have shown great promise in applications ranging from computer vision to natural language processing. Forecasting is special in that it has challenges of scaling, probability distribution outputs, sample efficiency and incorporation of prior knowledge. Statistical and NN models are both data driven yet statistical methods still on average outperform NN models. Some domains such as pandemic modelling have also been modelled using parameter based differential equations. The difficulty of these parameterized models is that it is difficult to estimate the parameters. Simulations have also been conducted which are governed by these parameterized models. Given that for pandemic modelling, there are modelling options such as parameterized models, statistical and NN forecasting methods, exclusivity does not make a lot of sense as each of these methods can benefit from the strengths of others. Ensembles of solutions from various techniques have been used with good success in the M4 competitions however explainable and innovative combinations of these various methods is an opportunity.

2 Time Series Forecasting

TS forecasting is about predicting future values of a target $y_{i,t}$ for a given entity i at time t . Each entity represents an organized grouping of temporal information that can be observed at the same time such as vital signs from different medical patients or the vibrations from an automobile. A one-step-ahead forecasting model, which is the simplest case, takes the form in Equation 1.

$$\hat{y}_{i,t+1} = f(y_{i,t-k:t}, x_{i,t-k:t}, s_i) \quad (1)$$

,where $\hat{y}_{i,t+1}$ is the model forecast, $y_{i,t-k:t}$ are the observations of the target and $x_{i,t-k:t}$ are the exogenous inputs; all of this data is over a look-back window k . s_i is static metadata associated with the entity, for example the sensor location. $f()$ is the prediction function learnt by the model. The equations are given for uni-variate forecasting (i.e., 1D targets) however all of the notation can be easily extended to multi-variate models. In addition, multi-horizon forecasting slightly revises the one-step-ahead model to the form given by Equation 2.

$$\hat{y}_{i,t+\tau} = f(y_{i,t-k:t}, x_{i,t-k:t}, s_i, \tau) \quad (2)$$

,where $\tau \in \{1, \dots, \tau_{max}\}$.

The approaches for TS forecasting range from fitting the parameters of mechanistic models to statistical methods using either compartmental or agent methods and recently machine learning including NN approaches.

3 Mechanistic models: Pandemic Modelling

A mechanistic or compartmental model of describing the propagation of ILI (Influenza Like Illness) can follow many different processes. Examples include models such as Susceptible-Infectious (SI), Susceptible-Infectious-Susceptible (SIS), Susceptible-Infectious-Recovered/Immune (SIR) and an extension to SIR called Susceptible-Exposed-Infected-Resistant (SEIR). SI is a two-state model with a one-way transition from Susceptible to Infected and no recovery after infection and is used to model HIV or herpes simplex virus type 2 (HSV-2). SIS is a two-state model where recovery does not induce immunity so individuals can transition both ways between susceptible and infected states and is used to model the common cold and curable STIs (Sexually Transmitted Illnesses) like gonorrhea. SIR is a three-state model with an additional one-way transition from infected to recovered with immunity and is used to model diseases like measles. SEIR adds an initial stage of infection, the E (Exposed) compartment. During the Exposed phase, individuals experience a long incubation duration in which the individual is infected but not yet infectious. Examples of this type of disease include Ebola. Both SIR and SEIR have been used to model Covid-19 [1–3]. The SIR model is governed by Equations 3-5 while the SEIR model is governed by Equations 6 & 7.

$$\frac{dS}{dt} = -\beta \frac{IS}{N} \quad (3)$$

$$\frac{dI}{dt} = \beta \frac{IS}{N} - \gamma I. \quad (4)$$

$$\frac{dR}{dt} = \gamma I \quad (5)$$

$$\frac{dS}{dt} = \beta \frac{IS}{N}, \quad \frac{dE}{dt} = \beta \frac{IS}{N} - aE \quad (6)$$

$$\frac{dI}{dt} = aE - \gamma I, \quad \frac{dR}{dt} = \gamma I \quad (7)$$

The SIR model assumes a population of size N where S is the total number of susceptible, I is the number of infected and R is the

number of resistant. Deaths are viewed as a subset of resistant, which is reasonable for a small death rate. β is the transmission rate constant, γ is the recovery rate constant, $R_0 = \frac{\beta}{\gamma}$ is the reproduction number and a is the inverse of the average incubation time. SEIR has been shown to predict the death rate well in the 1918 pandemic while the SIR model has shown to predict influenza cases well. Unfortunately, these models have several sources of uncertainty, including parameter uncertainty, variation based on data or model type used, even uncertainty in the data being collected to validate the model, uncertainty in the severity and length of social distancing measures and the parameters can vary drastically by location [1]. In general, the behaviors and interactions between viruses, hosts and the environment all play a part in determining the disease transmission. Surveillance data is typically imperfect, reported with delay or missing [4]. There is other uncertainty including biological and medical facts (relevance of viral load for chronic-degenerative disease, obesity) and vagueness in medical care structures [5].

One way to solve for the process model is to estimate the parameters by fitting the differential equations to data. Numerical methods such as Euler's method or Runge-Kutta method are used to solve the ordinary differential equations to get a solution. The fitting of a dynamical epidemiological model, e.g., SIR process, is done by using least squares, optimization algorithms or likelihood [6, 7].

Deterministic Compartment Models (DCMs) solve differential equations representing analytic epidemic systems in continuous time. The fixed mathematical functions of input parameters and initial conditions define the models as Deterministic. The models are compartmental as they divide the population into groups representing discrete disease states (e.g., susceptible, infected). ICM (Individual Contact Models) differ from DCMs in that the parameters are stochastic random draws (rates and risks governing state transitions are random draws from distributions), time is discrete and units are individuals (versus DCMs that treat populations as an aggregate).

Network models (i.e., agent based models) for example, modelled with EpiModel [3], represent contact phenomena within and across dyads (pairs of individuals who remain in contact) over time. EpiModel uses "separable-temporal exponential-family random graph models STERGMs (Spatial Temporal Exponential Random Graph Models) to estimate and simulate complete networks based on individual-level, dyad-level, and network-level patterns of density, degree, associative, and other features influencing edge formation and dissolution". A network model consists of elements (actors in the model), states (attributes of system elements) and transitions (movement rate between states). The domain of networks can either be human social networks (links can be contact, exchange, affect, role-based, genetic), animal networks (links can be contact, movement, genetic), institutional networks (links can be hospital patient transfer, goods, money) or multi-level designs. The states or attributes of the elements can either be infection related (susceptible, infected, recovered), demographic (sex, age), behavioral (level of sociability, occupation), clinical (tested or not, treated), geospatial (community, coordinates) and others. The transitions can either be deterministic (a fixed rate of transition between states) or stochastic (probability rate of transition). Using a model such as SIR makes assumption among individuals which does not necessarily hold true in the real world. Incorporating some commonly observed networks patterns such as heterogeneity in the number of contacts per individual and the clustering of contacts affects behaviors of the model such as the speed of the disease spread, long run health outcomes and the effect of disease on economic activity [8]. This also holds for interventions such as social distancing and vaccinations. As an example of network modelling, epidemic dynamics can be modelled as a diffusion process on a specific unidirectional contact network $G(V,E)$ on a population V of nodes, each edge $e = (u,v) \in E$ implies that individuals $u,v \in V$ come in contact. An infection can be spread from u to v along edge $e(u,v)$ with a probability $\beta(e,t)$ at time t after u becomes infected conditional on node v remaining uninfected until time t . The dynamical system initiates with at least one or more nodes in state I (infected) and reaches a point where all nodes are in states S (susceptible) or R (recovered) for a SIR model [9]. One such model for Covid-19 was put forward by Ferguson at Imperial College that brought attention to the severity of the upcoming Covid-19 pandemic [10].

Mechanistic models are also referred to as causal models. On the positive side, they employ mathematical disease models and make multi-fidelity predictions. The models capture an aspect of

human decision making and provide paths for counterfactual forecasts. Unfortunately they can be computationally expensive due to the requirement of estimating parameters from a high dimensional space. For the case of network models, obtaining the required data for a realistic social network can be difficult [11].

4 Statistical Models

Statistical methods, also referred to as classical methods [12], include both linear and non-linear models such as Koopmans model [13]. Most of the winners of the influential forecasting competitions including the M4, M3 and Tourism datasets have seen statistical methods, winning more often than average, however, recently there have been some machine learning winners [14]. However, most of the winners in these competitions have actually been ensemble methods, usually statistical and machine learning approaches combined. The strength of ensembles is their diversity.

The most naive method of forecasting, the random walk forecast, predicts the future values of the time series based on the last known observation:

$$\hat{y}_{n+h} = y_n \quad (8)$$

There is some empirical evidence that the random walk forecast provides a reasonable fit for financial data [15]. NN typically do well with financial data and the knock against them is that they appear to be only modelling a random walk. The seasonal naive model bears similarity to the naive model except that it is based on the last observed value in the same season.

$$\hat{y}_{n+h} = y_{n+h-m} \quad (9)$$

The ARMA (Auto-Regressive Moving Average) combines two techniques AR(p) and MA(q). The AR(p) model estimates the value of a given time series y_n by using a linear combination of the past p observations with an error term ϵ_n and a constant term.

$$y_n = c + \sum_{i=1}^p \phi_i y_{n-i} + \epsilon_n \quad (10)$$

where $\phi_i \forall i \in \{1, \dots, p\}$ are the model parameters and p is the order of the model. The MA(q) model is based on past errors.

$$y_n = \mu + \sum_{i=1}^q \theta_i \epsilon_{n-i} + \epsilon_n \quad (11)$$

where μ is the mean of the observations, $\theta_i \forall i \in \{1, \dots, q\}$ are the parameters of the models and q denotes the model order. The ARMA(p,q) model is a combination of the AR(p) and MA(q) models.

$$y_n = c + \sum_{i=1}^p \phi_i y_{n-i} + \sum_{i=1}^q \theta_i \epsilon_{n-i} + \epsilon_n \quad (12)$$

The ARMA(p,q) model is defined for stationary data however real world data exhibits non-stationary behavior. ARIMA(p,d,q) model overcomes the non-stationary typically exhibited by trend and seasonality by including an integration parameter of order d . ARIMA applies d difference transformations to the time series before applying ARMA(p,q). A simplification of the ARIMA model is based on the observation that it can be rewritten as [16]:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_t \quad (13)$$

where y'_t is the differenced series. This is called the ARIMA(p,d,q) model where p is the order of the auto-regressive component, d is the degree of the first differencing involved and q is the order of the moving average component. Special cases of the ARIMA model include white noise (ARIM(0,0,0)), random walk (ARIMA(0,1,0) with no constant), random walk with drift (ARIMA(0,1,0) with a constant), auto-regression (ARIMA(p,0,0)) and moving average (ARIMA(0,0,q)). Rearranging the terms and expressing ARIMA in backshift notation [16] results in:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \epsilon_t \quad (14)$$

,where AR(p) is the first term on the left hand side and d differences are the second term while the right hand side is just MA(q).

Other statistical methods include TBAT (which is a combination of Trigonometric seasonal formulation, Box-Cox transformation,

ARMA errors and Trend component) and HWASS (Holt-Winters additive model) [17]. TBAT is effective for maximum likelihood estimation and handles seasonal data, has a larger parameter space than ARIMA and better handles non-linearities. HWASS handles trend and seasonality well however it is sensitive to the choice of initial values and is sensitive to anomalies. ARIMA is usually still a strong statistical method however variants such as TBAT and HWASS and others have strengths. Methods other than ARIMA are usually ensembles or variants of ARIMA. There are other packages such as Facebook's Prophet. Prophet is an additive regressive model that is able to automatically detect trends and seasonal changes. Parameter estimation is handled probabilistically and parameter uncertainty is handled using the Hamiltonian Monte Carlo algorithm. Statistical methods learn patterns from the TS, are easy to implement and are fast to train and forecast [11]. Unfortunately they also assume a simple input and output relationship and are unable to make heterogeneous high resolution forecasts.

5 Neural Network Models

Recently, a deep learning method called N-BEATS [14], a neural architecture based on backward and forward residual links and a deep stack of fully-connected layers has shown good results on the standard M4 datasets. Another method called DeepAR [18] is based on an auto-regressive neural network for probabilistic forecasting in the form of Monte Carlo samples to calculate consistent estimates in the forecasting scope, not assuming Gaussian noise. Gluonts is a probabilistic time series modeling toolkit, focusing on deep learning-based models based on MXNET (ts.gluon.ai); our implementation of DeepAR is taken from this toolkit. Recently, a method called TDESFI [11] uses an agent based simulation model to train a neural network. At a coarse resolution (state, county levels), the technique outperforms standard statistical methods. TDESFI uses a network model to provide synthetic data for training. If a network model can provide synthetic data for training then it can also provide forecasts as well, negating the advantage of TDESFI.

TS data augmentation to provide sufficient training data to avoid over-fitting can be categorized into basic and advanced approaches [19]. Basic approaches include cropping, flipping, warping, jittering and perturbing in either the time, frequency or combined domains. Advanced approaches include decomposition, model or learning approaches. Learning approaches either are embedding space, GAN (Generative Adversarial Networks) approaches or the use of RL (Reinforcement Learning) or meta-learning. TDESFI is a model based approach. Concerns about neural network methods include the dependence on a large amount of training data, the inability to make heterogeneous high resolution forecasts, the inability to include domain expertise, the lack of explainable results and over-fitting is a concern due to the usual availability of only a small size of training data [11]. Neural networks require lots of data to estimate the parameters of the network. Data in this context refer to the number of series but the number of observations [20]. Machine learning models including NNs can benefit when larger TS are available with large and high frequency data [15]. As with other ML interests, topics of interpretability, explainability and causality are relevant for TS forecasting. TS also have hierarchical structure with logical groupings between trajectories, for example trends that are not handled well by NN [21]. Computational requirements are considerably greater for NNs when compared to statistical methods [22]. It has been shown that machine learning including NN approaches improve predictability when sample training size grows [15]. Super-forecasting [23],[24] shows that humans can do quite well providing forecasts by relying on domain expertise which both statistical and NN do not as they are purely data driven. NNs also require TS to be discretised at regular intervals making it difficult to forecast when observations are missing or arrive at random intervals [21]. It should be noted that NNs are not always the most appropriate choice for long-range macro-economic forecasts or other problems requiring external domain knowledge not learn-able from the data [20]. It has been found that typically ML methods do not stack well in terms of both accuracy measures and forecasting horizons [22].

NNs are good at learning patterns from historical data and are able to capture non-linear input output relationships [11]. As mentioned earlier, they need a large amount of training data, are unable to make heterogeneous high resolution forecasts, lack explainability and overfitting is a possibility if insufficient training data is presented. Hybrid methods that combine data driven (Statistical,

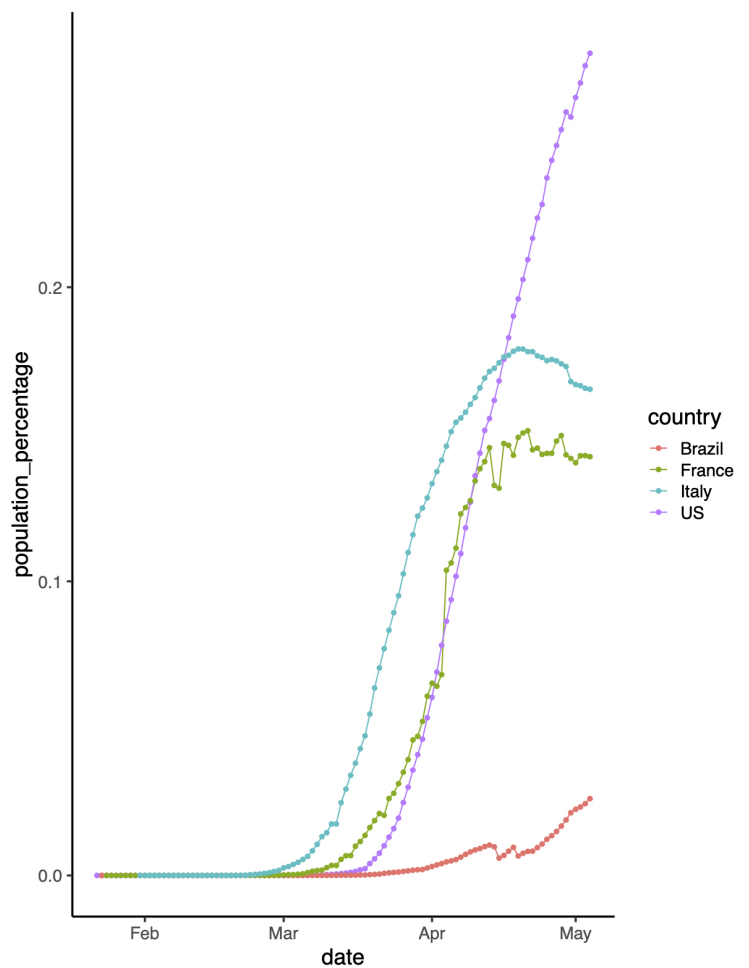


Fig. 1: Covid-19 cases: 4 of the 10 top countries: as a % of population

NN) and mechanistic methods have the possibility to integrate all their strengths and epidemic forecasting is a good domain to test this. The TDESFI method is one such hybrid method however it appears that the synthetic training data obtained from a network method is only capturing the probability distributions and unknowns which could also be captured with a network method by itself. There must be a better way to integrate different methods.

Interesting methods that incorporate state models into neural network forecasting include ones with a particle filter [25] or Kalman filter [26]. Forecasting may also be better served by using features as opposed to raw time series data [27–29]

6 Experiments

Based on the implementation of [17], six approaches are applied to a Kaggle dataset of Covid-19 data [30]. Of the six approaches, 4 are statistical based (ARIMA, HWAAS, TBAT, Prophet) while 2 are neural networks (N-Beats, DeepAR). Ten countries with the greatest number of total confirmed cases in the period of April to May 2020 were chosen as the dataset. Each TS model was trained, evaluated and tested using a TS representing a percentage of active cases with respect to the total population of the relevant country. For training and validation purposes, 72 instances were used for training, 25 for validation and a window of seven days was used for prediction. Figure 1 illustrates the active cases for the top ten countries as a percentage of active cases with respect to the total population.

The forecasting results for each country represented in terms of RMSE (Root Mean Squared Error) are shown in Table 1. The table shows that there is no one size fits all for the forecasting results. From the table, it is evident that the statistical methods (i.e., ARIMA, TBAT, HWAAS) in general out perform the deep learning NN approaches (N-beats, DeepAR) in 8 out of 10 instances. As mentioned earlier, this could be a result of a lack of high volume training data that NN approaches require to perform well. Unfortunately for the task of forecasting where the data is localized temporally and geographically does not necessarily allow this. Figure 2 shows

country	ARIMA	HWAAS	TBAT	N-Beats	DeepAR
US	0.0074	0.1729	0.0098	0.0369	0.0448
Spain	0.0800	0.0314	0.0292	0.0369	0.1088
Italy	0.0056	0.0066	0.0058	0.0086	0.0435
UK	0.0054	0.0043	0.0043	0.0376	0.0376
France	0.0608	0.0110	0.0070	0.0042	0.0105
Germany	0.0064	0.0371	0.0033	0.0131	0.0575
Russia	0.0015	0.0022	0.0021	0.0270	0.0344
Turkey	0.0044	0.0008	0.0019	0.0182	0.0938
Brazil	0.0041	0.0057	0.0056	0.0057	0.0028
IRan	0.0026	0.0010	0.0004	0.0037	0.0022

Table 1: RSME with Different Prediction Methods

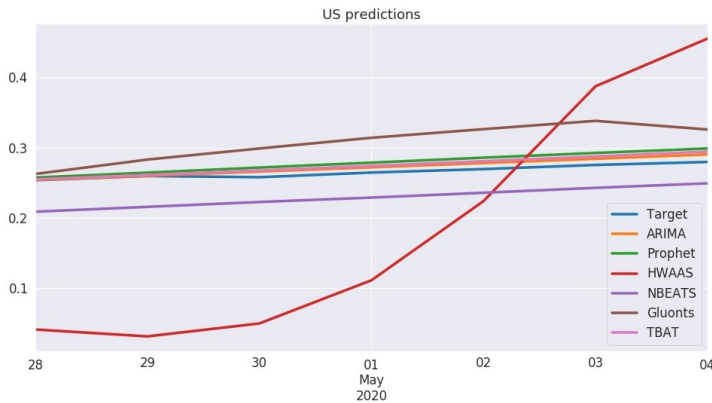


Fig. 2: Predicted Active Cases for US using Different Methods

the forecast results for US data for the various methods. TBAT and ARIMA appear to be the best performing techniques overall in terms of RMSE. ARIMA was superior with respect to RMSE in the US, Italy and Russia. The NN methods appeared to show good results for the data from France and Brazil. There are so many factors that can help explain why some methods performed better than others including climate, geographical culture, varying population densities, testing and measuring differences as well as variety in social distancing measures, quarantine in terms of timing, during and severity of measures acted upon. It is quite possible that the NN methods work best only on the Brazil and France data because of the random walk nature of the data (see Figure 1), which has been found to be a trait of NN methods forecasting financial data.

7 Discussions

Superforecasters [24] rely on domain knowledge and perhaps data driven techniques should also utilize this expertise if available. There are opportunities to improve data driven approaches with state space models using Kalman [26] and Particle Filtering [25]. The NN approach called TDEFSI [11] utilizes a network simulation to provide synthetic input to train a network to forecast. An interesting approach however, if the data is already simulated why not just simulate further into the horizon. New approaches that integrate physics modelling with NN [31] are inspiring for directions on how to integrate mechanistic models with NN techniques for forecasting, in particular pandemics where mechanistic models do exist. When the mechanistic models are based on ODE (Ordinary Differential Equations), somehow embedding this into the NN architecture appears necessary and recent techniques suggest that this may be possible [32]. Pandemic modelling is difficult by itself, even though that in most cases, the data is not reliable. Being able to incorporate domain expertise in some manner makes sense as a purely data driven approach has its limits. Forecasting usually does not afford a lot of training data and this handcuffs NN approaches. Perhaps recasting the way NN are used, lets say as anomaly detectors with respect to a mechanistic model may provide more training data? Accepting the unknowns and unobserved and dealing with the forecasting problem using a NN probabilistically is also worth of exploration [33].

"Those who have knowledge, don't predict. Those who predict, don't have knowledge" (Lao Tzu).

Acknowledgments

The authors would like to thank NSERC for their generous funding via the USRA award program.

References

- [1] A. L. Bertozzi, E. Franco, G. Mohler, M. B. Short, and D. Sledge, "The challenges of modeling and forecasting the spread of covid-19," *arXiv preprint arXiv:2004.04741*, 2020.
- [2] S. Kandula, T. Yamana, S. Pei, W. Yang, H. Morita, and J. Shaman, "Evaluation of mechanistic and statistical methods in forecasting influenza-like illness," *Journal of The Royal Society Interface*, vol. 15, no. 144, p. 20180174, 2018.
- [3] S. M. Jenness, S. M. Goodreau, and M. Morris, "EpiModel: an r package for mathematical modeling of infectious disease over networks," *Journal of statistical software*, vol. 84, 2018.
- [4] S. A. Lauer, A. C. Brown, and N. G. Reich, "Infectious disease forecasting for public health," *arXiv preprint arXiv:2006.00073*, 2020.
- [5] M. Wang and S. Flessa, "Modelling covid-19 under uncertainty: what can we expect?" *The European Journal of Health Economics*, p. 1, 2020.
- [6] R. F. Grais, M. J. Ferrari, C. Dubray, O. Bjørnstad, B. Grenfell, A. Djibo, F. Fermon, and P. J. Guerin, "Estimating transmission intensity for a measles epidemic in niamey, niger: lessons for intervention," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 100, no. 9, pp. 867–873, 2006.
- [7] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927.
- [8] B. R. Craig, T. Phelan, J.-P. Siedlarek, and J. Steinberg, "Improving epidemic modeling with networks," *Economic Commentary*, no. 2020-23, 2020.
- [9] A. Adiga, D. Dubhashi, B. Lewis, M. Marathe, S. Venkatraman, and A. Vullikanti, "Mathematical models for covid-19 pandemic: a comparative analysis," *Journal of the Indian Institute of Science*, pp. 1–15, 2020.
- [10] N. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg *et al.*, "Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand," *Imperial College London*, vol. 10, p. 77482, 2020.
- [11] L. Wang, J. Chen, and M. Marathe, "Tdefsi: Theory-guided deep learning-based epidemic forecasting with synthetic information," *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, vol. 6, no. 3, pp. 1–39, 2020.
- [12] C. Faloutsos, V. Flunkert, J. Gasthaus, T. Januschowski, and Y. Wang, "Forecasting big time series: Theory and practice," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 3209–3210.
- [13] S. Spanbauer and I. Hunter, "Extended koopman models," *arXiv preprint arXiv:2010.06845*, 2020.
- [14] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," *arXiv preprint arXiv:1905.10437*, 2019.
- [15] V. Cerqueira, L. Torgo, and C. Soares, "Machine learning vs statistical methods for time series forecasting: Size matters," *arXiv preprint arXiv:1909.13316*, 2019.
- [16] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

- [17] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, "Covid-19: A comparison of time series methods to forecast percentage of active cases per population," *Applied Sciences*, vol. 10, no. 11, p. 3880, 2020.
- [18] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "Deepar: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.
- [19] Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," *arXiv preprint arXiv:2002.12478*, 2020.
- [20] K. Benidis, S. S. Rangapuram, V. Flunkert, B. Wang, D. Maddix, C. Turkmen, J. Gasthaus, M. Bohlke-Schneider, D. Salinas, L. Stella *et al.*, "Neural forecasting: Introduction and literature overview," *arXiv preprint arXiv:2004.10240*, 2020.
- [21] B. Lim and S. Zohren, "Time series forecasting with deep learning: A survey," *arXiv preprint arXiv:2004.13408*, 2020.
- [22] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLoS one*, vol. 13, no. 3, p. e0194889, 2018.
- [23] I. Katsagounos, D. D. Thomakos, K. Litsiou, and K. Nikolopoulos, "Superforecasting reality check: Evidence from a small pool of experts and expedited identification," *European Journal of Operational Research*, 2020.
- [24] P. E. Tetlock and D. Gardner, *Superforecasting: The art and science of prediction*. Random House, 2016.
- [25] X. Ma, P. Karkus, D. Hsu, and W. S. Lee, "Particle filter recurrent neural networks," *arXiv preprint arXiv:1905.12885*, 2019.
- [26] J. T. Jalles, "Structural time series models and the kalman filter: a concise review," 2009.
- [27] V. Cerqueira, N. Moniz, and C. Soares, "Vest: Automatic feature engineering for forecasting," *arXiv preprint arXiv:2010.07137*, 2020.
- [28] B. D. Fulcher and N. S. Jones, "hctsa: A computational framework for automated time-series phenotyping using massive feature extraction," *Cell systems*, vol. 5, no. 5, pp. 527–531, 2017.
- [29] B. D. Fulcher, C. H. Lubba, S. S. Sethi, and N. S. Jones, "Compengine: a self-organizing, living library of time-series data," *arXiv*, pp. arXiv–1905, 2019.
- [30] SRK. (2019) Novel corona virus 2019 dataset. [Online]. Available: <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>
- [31] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating physics-based modeling with machine learning: A survey," *arXiv preprint arXiv:2003.04919*, 2020.
- [32] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in neural information processing systems*, 2018, pp. 6571–6583.
- [33] M. Haas and S. Richter, "Statistical analysis of wasserstein gans with applications to time series forecasting," 2020.