A Neural SIR Model for Global Forecasting

Philip Nadler P.NADLER@IMPERIAL.AC.UK
Imperial College London - Data Science Institute

Rossella Arcucci R.ARCUCCI@IMPERIAL.AC.UK
Imperial College London - Data Science Institute

Yike Guo Y.GUO@IMPERIAL.AC.UK
Imperial College London - Data Science Institute

Editors: Emily Alsentzer[⊗], Matthew B. A. McDermott[⊗], Fabian Falck, Suproteem K. Sarkar, Subhrajit Roy[‡], Stephanie L. Hyland[‡]

Abstract

Being able to understand and forecast epidemic developments is crucial for policymakers. We develop a predictive model combining epidemiological dynamics of compartmental models with highly non-linear interactions learned by a LSTM Network. A novel dynamic SIR model is fit to variables related to the population transmission of Covid-19. This is embedded in a Bayesian recursive updating framework which is then coupled with a LSTM network to forecast cases of Covid-19. The model significantly improves forecasts over simple univariate LSTM or SIR models. We apply the model to developed and developing countries and forecast confirmed infections and analyze future trajectories.

Keywords: Epidemiology, Neural Networks, Forecasting, LSTM, SIR, Compartmental Model

1. Introduction

The outbreak of the global 2019-nCov pandemic requires robust government intervention to mitigate future social and economic costs of epidemics. The global south is especially vulnerable, thus in order to inform policymakers we develop a robust forecasting and inference model which is deployable and generalisable across different populations worldwide.

Combining machine learning and epidemic modelling, we propose a new compartmental SIR model for forecasting and policy evaluation which incorporates new data in real-time through Bayesian updating. The model is applied to infer the amount of infected people and parameters such as the disease transmissibility rate and the rate of recovery.

These dynamics are coupled with a neural network which can model highly non-linear interactions of variables which are not explicitly modelled in the dynamic SIR framework. We label the model neural SIR by combining the LSTM Neural Network with the dynamic compartmental SIR model.

This novel architecture generates forecasts of confirmed Covid19 cases across multiple countries. The model combines the effectiveness of LSTM networks in forecasting with the epidemiological dynamics of SIR models, which the network can learn and use to enhance predictive accuracy. This dual approach combines the high predictive performance of blackbox machine learning models with well-defined parametric models, whose parameters are clearly derived and interpreted in our work, contributing to interpretability and transparency of epidemic

forecasting models. The parsimonious structure of our model allows for generalisation across multiple locations and scenarios by adding new compartments and thus eases access for practitioners and policy makers who need to generate forecasts for local epidemic outbreaks.

We use our model to compare accuracy of forecasts in developed countries such as the United States and the United Kingdom as well as a multitude of developing countries including India, Turkey and Brazil, where accurate forecasting is crucial to support policymakers.

The rest of the paper is structured as follows: Section 2 reviews related work, Section 3 introduces the dynamic SIR model. Section 4 introduces the LSTM and Neural SIR model. Section 5 discusses empirical results and section 6 concludes.

2. Related Work

Authors such as Imai et al. (2020), Li et al. (2020) and Wu et al. (2020) have done the first studies on the size of the outbreak in China using epidemiological compartmental models. The analysis is based on basic SIR or SEIR models where the models are fit to Chinese cities in order to infer population infection rates as well as parameters such as the reproduction number of Covid19. First studies of dynamic models using filtering approaches for epidemiological modelling have been conducted by other authors such as Rhodes and Hollingsworth (2009) and Bettencourt et al. (2007), which studied the frameworks in different cases such as influenza using standard SIR models.Further studies such as Bettencourt and Ribeiro (2008) and Cobb et al. (2014) investigate time varying parameters in more detail, although only for established SIR models with no relationship to the current coronavirus outbreak. A more recent line of research such as La Gatta et al. (2020) have investigated the use of deep learning for epidemiological analysis, learning spatio-temporal patterns of Covid19. The authors in Coskun et al. (2017) conducted initial studies coupling LSTMs and filtering algorithms, but limited to the context of pose detection and not compartmental modelling.

We are the first to conduct a study of the current spread of 2019-nCoV using Bayesian recursive updating combined with LSTM models. Preliminary studies using only the dynamic SIR have been conducted in Nadler et al. (2020). Because of the model's inability to capture non-linearities Arcucci et al. (2017) which are not explicitly specified in the model, model results may suffer inaccuracy. To overcome this limitation we couple the SIR model with a recursive Bayesian updating scheme and a LSTM network in order to improve the accuracy of the model, which we label Neural SIR.

3. The SIR Model

The analysis is conducted using a standard SIR model Anderson (1991), which is a system of three interrelated, non-linear ordinary differential equations without an explicit analytical solution. The dynamics of the model are given by :

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{IS}{N} \\ \frac{dI}{dt} &= \beta \frac{IS}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{1}$$

Where S denotes the susceptible population size, I the infected people who are not isolated from the population and R the recovered population. The total population is given by N. The parameters β and γ denote the transmission and recover rate of the virus infection. Note that for the outbreak in

cities or countries such as Brazil, the susceptible number S is observable, which we define as the population of the location currently under analysis whereas the true number of total symptomatic and asymptomatic infections I in the population is unobservable. The recovered population R denotes patients not infectious anymore or being quarantined from the population.
We only observe confirmed cases R^{obs}, which for local Covid19 outbreaks is assumed to be the number of cases which are reported and being hospitalized or isolated, thus not infecting the general population anymore. Thus R is the absolute fraction of recovered population in the model and R^{obs} is the number of confirmed cases. We take this difference explicitly into account in the next section where the model is updated by confirmed cases. We thus aim to evaluate the prediction accuracy of the model to investigate model fit as well as to use the generated forecasts of the model to give an estimate for the trend of the epidemic. We outline the model in the following sections.

3.1. The General State Space Model

We first discuss the model updating framework, which integrates new observations into the model to enhance forecasts as well as computing model parameters of interest Asch et al. (2016). The methodology of our proposed Bayesian updating framework is derived from the literature on data assimilation (DA), a form of hidden markov or state space modelling Miller et al. (1999); Asch et al. (2016), which conducts model parameter updates when new observations become available.
It is often applied in physical sciences, but has seen increased use in many other fields such as machine learning or finance Abarbanel et al. (2018). Minimization of the appropriate cost function can lead to well known filtering algorithms such as the Kalman filter Kalman (1960). The general updating model is described by the following equation

$$\mathbf{x}_{t+1} = \boldsymbol{\mathcal{M}}_{t+1}\mathbf{x}_t \tag{2}$$

where $\mathbf{x}_t$ and $\boldsymbol{\mathcal{M}}_t$ are the state (background) vector and nonlinear model operator at timestep t respectively. In context of the SIR model, $\boldsymbol{\mathcal{M}}_t$ is the forecasting model given by Eq. 1, and the statevector is $\mathbf{x}_t = [S_t, I_t, R_t]$. Furthermore, the variable $\mathbf{y}_t^o$ represents observations at timestep t and $\boldsymbol{\mathcal{H}}_t$ is the nonlinear observation operator that maps observation to model space:

$$\mathbf{y}_t^o = \boldsymbol{\mathcal{H}}_t\mathbf{x}_t \tag{3}$$

Here, $\mathbf{y}_t^o$ correspond to observations R^{obs} and model predictions thereof, based on which model parameters will be updated. In order to formulate the cost function of the model, both the background error covariance matrix $\mathbf{Q}$ and the observation error covariance matrix $\mathbf{P}$ need to be computed depending on the data Bannister (2008) which will be outlined in the following sections.

The cost function is a form of Tikhonov regularisation Dong et al. (2017) and is defined as

$$J(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{Q}^{-1}} + \sum_{window} \|\mathbf{H}\mathbf{x} - \mathbf{y}^o\|_{\mathbf{P}^{-1}} \tag{4}$$

where $\mathbf{H}$ is the linearized version of the nonlinear observation operator $\boldsymbol{\mathcal{H}}$ with the background state vector give by $\mathbf{x}^b$. The window size determines the number of past observations for updating and depends on the updating approach Lorenc (2003). We next show how compartmental SIR models can be embedded in this approach.
Using data from India, Fig. 1 illustrates the difference between a static SIR model modelled by three ordinary differential equations and a dynamically updated model using Bayesian updates. Depending on the pa-

rameterization a static SIR model would predict a simplistic exponential growth pattern given the initial observations, whereas the dynamic model incorporating model updates adjusts infection estimates given the new observations.

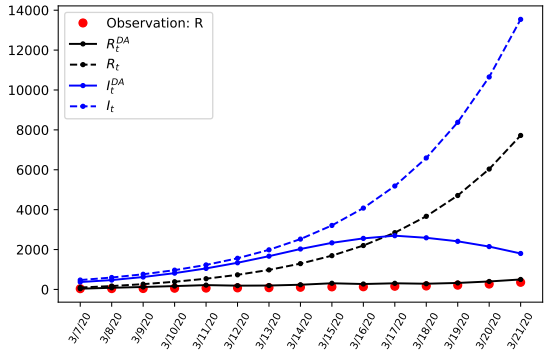

Figure 1: Comparing a standard SIR model as static ODE representation with a dynamic (DA) SIR model applying recursive Bayesian model updates.

3.2. The Dynamic SIR Model

The continuous SIR model in equation (1) can be discretized with respect to the time variable, giving the following equations:

$$\begin{aligned} S_{t+1} &= S_t - \beta \frac{I_t S_t}{N} \\ I_{t+1} &= I_t + \beta \frac{I_t S_t}{N} - \gamma I_t \\ R_{t+1} &= R_t + \gamma I_t \end{aligned} \tag{5}$$

For a given time step t and assuming to have observations of R_t^{obs}, the model updating problem consists in computing the minimum of the cost function

$$J(I) = \underset{I}{\operatorname{argmin}} \sum_{i=t+1}^{t+\tau} ||R_t^{obs} - H(i|I, \beta, \gamma)||_{\mathbf{P}_t^{-1}} + ||I - I_t||_{\mathbf{Q}_t^{-1}} \tag{6}$$

and

$$I_t^{DA} = \underset{I}{\operatorname{argmin}}\, J(I) \tag{7}$$

within window length τ, where $H : I \to R$ is a linear transformation function usually called observation function Asch et al. (2016) which is here represented by the SIR model, and where $\mathbf{Q}$ and $\mathbf{P}$ denote the the background and the observation covariance matrices, being an estimate of noise in the data. Since results are sensitive to the parameterization of the covariance matrices, we outline their calibration in appendix B. To compute parameters, we minimize

$$\beta, \gamma = \underset{\beta_i, \gamma_i}{\operatorname{argmin}} \sum_{i=t+1}^{t+\tau} ||R_t^{obs} - H(i|I_i^{DA}, \beta_i, \gamma_i)||_{\mathbf{P}_t^{-1}} \tag{8}$$

Solving the cost function leads to a modified Kalman filtering algorithm, where updating the model state generates a latent series of latent asymptomatic infection numbers I, given parameters β and γ. The result is visible in Fig. 1 or in more detail in appendix Fig. 7, where the model estimates a much higher number of asymptomatic infections not captured only by reported cases. The examples show results for India and the United Kingdom. As is visible in the later case, if the epidemic is in a more advanced stage and the model infers a high amount of infections, the accuracy of the model suffers due to the imposed structure of the differential equations, which in turn affects model estimates. The next section introduces the coupled LSTM network and discuss how the models accuracy can be improved by combining an LSTM and the generated latent variables of the dynamic SIR model.

4. The LSTM Model

A generalized form of supervised learning for neural networks is given by:

$$(\mathbf{x_t^{tr}}(\mathbf{l_0}), \mathbf{x_{t+1}^{tr}}(\mathbf{l_F})) \tag{9}$$

where training of the network on data $\mathbf{x_t^{tr}}$ is done on a set of input and output pairs which are available at input layer l_0 and output layer l_F which are the observations with $\mathbf{x}_t = [S_t, I_t, R_t]$ and $\mathbf{x}_{t+1} = [S_{t+1}, I_{t+1}, R_{t+1}]$. The structure and activation of each hidden unit in the hidden layers l_i are determined by the neurons in the previous layers, denoted as $\mathbf{a}_k(l_i)$.The activity of each layer is given by the nonlinear activation function σ as in:

$$\mathbf{a}_k(l_i) = \sigma(\sum_j w_{j,k}(l_i)\mathbf{a}_j(l_{i-1}) + \mathbf{b}_k(l_i)) \quad (10)$$

where $w_{j,k}(l_i)$ is the weight from the j^{th} node of layer l_{i-1} to the k^{th} neuron of the layer l_i, with $\mathbf{b}_k(l_i)$ and $\mathbf{a}_k(l_i)$ representing the bias and the activation terms of the k^{th} neuron in layer l_i. The summation over weights $w_{j,k}(l_i)$ determines how the activities in layer l_i are combined before allowing σ to act, yielding the activities at layer l_i.

A popular and well tested model for sequential data is the Long Short Term Memory Network Hochreiter and Schmidhuber (1997); Gers et al. (1999). The memory cells of the LSTM can be described as Recurrent Neural Networks (RNN) using memory cells which allow the processing of data for long temporal sequences, circumventing the issue of vanishing gradients due to its gate structure Xingjian et al. (2015). The LSTM can be described by the following system of equations:

$$\begin{aligned} i_{t+1} &= \sigma(w_{xi}\mathbf{x}_t + w_{bi}b_t + w_{ci}c_t + b_i) \\ f_{t+1} &= \sigma(w_{xi}\mathbf{x}_t + w_{hf}h_t + w_{cf}c_{t-1} + b_f) \\ c_{t+1} &= f_{t+1}c_t + i_{t+1}tanh(w_{xc}\mathbf{x}_t + w_{hc}h_t + b_c) \\ o_{t+1} &= \sigma(w_{xo}\mathbf{x}_t + w_{ho}h_t + w_{co}c_{t+1} + b_0) \\ h_{t+1} &= o_{t+1}tanh(x_{t+1}) \\ x_{t+1} &= d(h_{t+1}) \end{aligned} \quad (11)$$

Where c_t is a memory cell that accumulates the information of the state. The three elements i_t, f_t and o_t denote the input, forget and output gates respectively. If the input gate i_t is activated, the information from a new input will be accumulated, if the forgot gate f_t is activated then the status of the cell c_{t-1} is forgotten. The propagation of the last cell c_t into the final state h_t is controlled by the output gate o_t. As this network is based on a recurrent neural network architecture, the output block is recurrently connected back to the input and the three gates.

4.1. The Neural SIR Model

Coupling the LSTM and the dynamic SIR model allows to capture possible higher order non-linearities and interactions between the variables which are not captured by the SIR model dynamics.

The confirmed observed cases are used as inputs for the LSTM as well as the SIR model, whose outputs are used as inputs for the LSTM model. Both models interact by generating weighted model outputs of confirmed cases which is used as forecast. The network input $\mathbf{x^{tr}}(\mathbf{l_0})$ in Eq. 9 is given by state variables of the dynamic SIR model in Eq. 4, which are the variables S, I, R respectively as specified in Eq. 6. The state vector $\mathbf{x} = [S, I, R]$ is propagated through the LSTM network and yields output $\mathbf{x^{tr}}(\mathbf{l_F})$. The loss of the resulting model is represented by

$$\begin{aligned} L_{NS} = min||R^{obs} - \mathbf{x^{tr}}(\mathbf{l_F}), \\ R^{obs} - H(i|I_i^{DA}, \beta_i, \gamma_i)||_2 \end{aligned} \quad (12)$$

The coupled models given in Fig. 2 illustrate the Neural SIR architecture. Both models receive new updates from observed cases, the network then ingests the state variables of the SIR model as additional inputs and both model forecasts are combined. The final output of the coupled model is obtained by combining the network output of confirmed cases $\mathbf{x^{tr}}(\mathbf{l_F})$ with the SIR model forecast R_{t+1}^{obs}:

$$R_{t+1}^{*} = w_1\mathbf{x^{tr}}(\mathbf{l_F}) + w_2R_{t+1}^{obs} \quad (13)$$

where both weights w_1 and w_2 are relative weights defined by the previous periods forecasting error of each model.
The network architecture we use for our model is a LSTM using ReLU activation functions which is trained using Adam optimizer with a mean squared error loss function. The LSTM network was implemented with 175 neurons followed by dense layers with finally one neuron. The model is not constrained to this particular setup and we searched over multiple hyperparameters to vary the number of neurons, with similar results. We provide model details and reproducible code with various hyperparameter settings at: https://github.com/pnadler-imperial/ode_nsir.

This enables the LSTM network to learn the dynamics of the SIR model, without explicitly imposing the exponential growth curves of the SIR model, allowing for short and long term deviation of trends.
Appendix Figures 8 and 9 visualise how the dynamics of the model change when coupling a LSTM network with a SIR model. The linear dynamics of the forecasted confirmed infection cases of the simple LSTM are in stark contrast to the dynamics of the Neural SIR model which has coupled the LSTM network with the adaptive SIR model.
In the next section we describe and discuss the neural SIR results compared to other benchmarks such as the dynamic SIR and univariate LSTMs, applied to multiple countries.

5. Results

To illustrate the forecasting capability of the Neural SIR model, we compare the number of predicted confirmed Covid19 cases under various measures for within- and out out of sample scenarios. The in-sample fit of the model is an important indicator for the validity of the model's estimation of latent parameters, whereas the out-of-sample forecasts are an important guide for policy makers. The data for all analysis was obtained from the John Hopkins University (JHU) Coronavirus Resource Center[1] which collects daily updates on infected, recovered and deceased patients affected by Covid19. Appendix Table 3 gives a brief overview of the data. A more detailed summary is publicly available from JHU (2020).

5.1. In-Sample Fit

To make models comparable, we apply a mean squared forecasting error (MSFE) metric as well as mean absolute percentage errors (MAPE).

We report MAPE percentage errors to allow cross-comparisons between countries but pay special attention to absolute numbers in MSFE and MAFE. This is due to the absolute level of confirmed cases containing important information for policymakers and prediction errors mattering less when infection numbers are very low, whereas higher accuracy for a high number of infections is a desirable property of an epidemiological forecasting model . The main results are given in Table 1 and Table 2. We report metrics for Neural SIR, as well as forecasts produced by the dynamic SIR. We furthermore present results for a univariate LSTM which is trained only on confirmed cases R^{obs}. If not explicitly specified, the network architecture is the same as in the Neural SIR setup. We also include a multivariate LSTM metric, which is the forecast of the coupled multivariate LSTM without combining its model forecast with the SIR model.

The first two rows of Table 1 show that the Neural SIR outperforms all other model setups in terms of MSFE and MAFE for all countries. In some cases the dynamic SIR model has good prediction accuracy, but is

1. https://coronavirus.jhu.edu/map.html

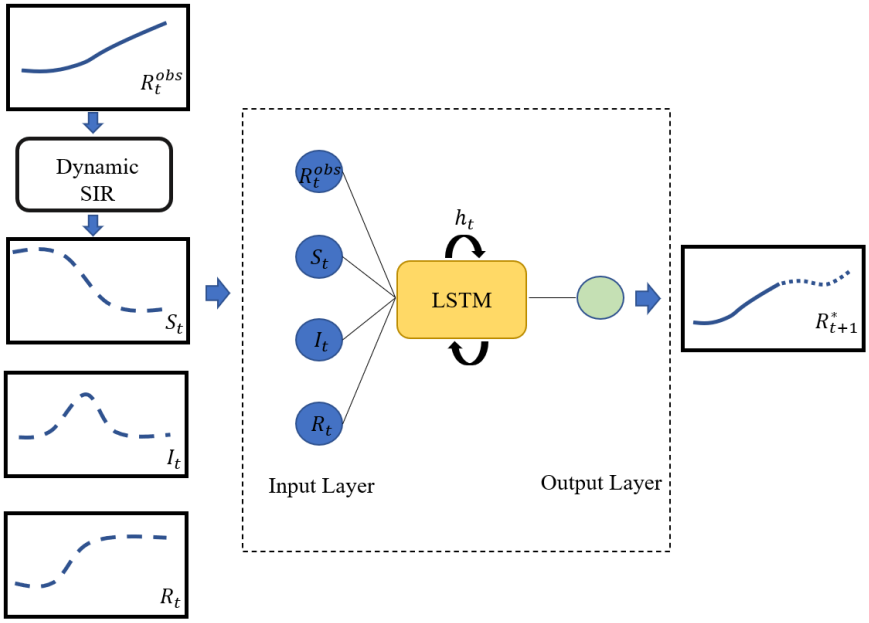

Figure 2: An illustration of the Neural SIR model structure, coupling the dynamic SIR model with the LSTM network to generate forecasts of confirmed cases. Given the observed cases R_t^{obs} the dynamic SIR model generates latent model series at each timestep which serve as input layer $\in \mathbb{R}^4$ for the LSTM which is propagated through the model until the output layer $\in \mathbb{R}^1$

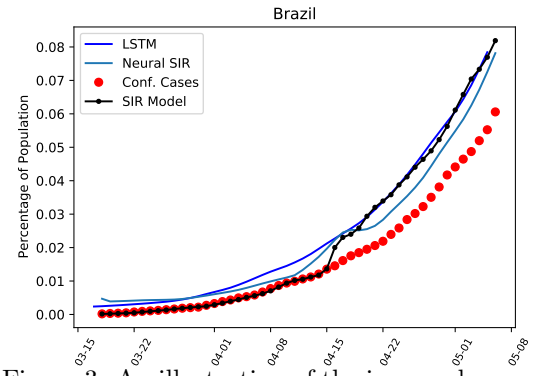

Figure 3: An illustration of the in-sample sample forecasting performance of different models for Brazil

still below the Neural SIR. The other LSTM field entries show similar results. The univariate LSTM by itself performs weak in terms of accuracy, showing that the incorporation of the SIR dynamics can significantly increase the accuracy of the network.

The MAPE results in Table 2 show that the Neural SIR still performs well but is outperformed by the dynamic SIR in the case of India where reported cases are very low compared to the overall population. This highlights that in phases with low infection numbers the dynamic SIR model performs better than LSTM architectures which exhibit decreased predictive performance when infection numbers are low. The fitted forecasted

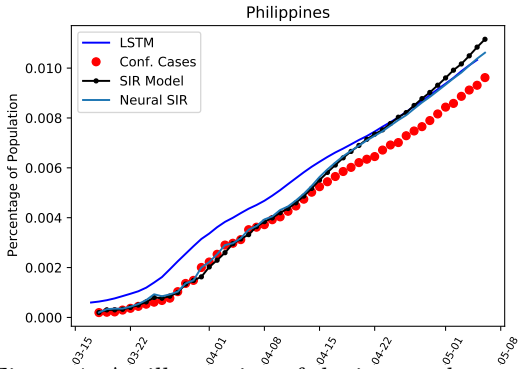

Figure 4: An illustration of the in-sample sample forecasting performance of different models for the Philippines

values of models are also depicted in Fig. 3 for Brazil. All prediction models overestimate the number of infected patients. We see that the SIR model follows exponential dy-

Table 1: Summary of all in-sample forecasting errors when forecasting one step ahead. Accuracy is given in mean squared (MSFE) and mean absolute (MAFE) forecasting errors for developed and developing countries. The Neural SIR model consistently outperforms other model setups in terms of accuracy.

		India	Brazil	Turkey	Philipp.	US	UK
Neural	MSFE	1.30E+08	4.47E+08	**1.06E+08**	**7.19E+05**	**5.64E+09**	**2.50E+08**
SIR	MAFE	**7.40E+03**	**1.34E+04**	**7.44E+03**	**6.70E+02**	**5.93E+04**	**1.38E+04**
Dynamic	MSFE	1.52E+08	2.34E+09	3.23E+09	2.36E+06	3.20E+10	4.27E+08
SIR	MAFE	7.57E+03	1.42E+04	3.67E+04	1.07E+03	1.27E+05	1.75E+04
Univ.	MSFE	2.43E+08	1.99E+09	1.56E+08	1.08E+06	1.56E+10	7.02E+08
LSTM	MAFE	1.13E+04	2.98E+04	1.08E+04	1.03E+03	1.24E+05	2.46E+04
Multiv.	MSFE	**1.25E+08**	**3.06E+08**	1.73E+08	9.51E+05	8.12E+09	3.18E+08
LSTM	MAFE	8.53E+03	1.42E+04	1.17E+04	9.40E+02	8.44E+04	1.65E+04

Table 2: Mean absolute percentage errors (MAPE) for one step ahead forecasts for all experiment setups

	India	Brazil	Turkey	Philippines	US	UK
Neural SIR	4.42E-01	**1.53E-01**	**6.78E-02**	**1.19E-01**	**7.94E-02**	**1.64E-01**
Dynamic SIR	**3.20E-01**	2.69E-01	3.39E-01	1.39E-01	1.43E-01	2.54E-01
Univ. LSTM	8.89E-01	2.88E-01	9.74E-01	5.83E-01	4.97E-01	2.89E-01
Multiv. LSTM	4.89E-01	8.54E-01	9.82E-01	4.86E-01	7.62E-01	2.56E-01

namics, similar to the univariate LSTM network. The Neural SIR model is more flexible and approximates the observed predictions the closest. The results for the Philippines in Fig. 4 reveal similar patterns although the dynamic SIR framework is a better predictor in the early sample with its performance deteriorating when absolute infection numbers are higher.

Both cases highlight the drawback of the SIR model being modelled by differential equations. When the amount of observable infections relative to the population increases the model will infer a very high amount of unobservable infections I, negatively affecting prediction accuracy. Since the Neural SIR incorporates information of the differential equations as in Eq.1 but is not explicitly constrained by the differential equations, its prediction performance is superior.

The results show that for both the Philippines and Brazil only a small number of infected patients are reported, with 0.01 and 0.08 percent respectively. The confirmed numbers show a strong upwards trend, especially for Brazil, therefore it is likely that the number of infections is going to increase at an accelerating rate in the close future.

5.2. Out of Sample Forecasts

We conduct an additional analysis performing a multiple day out of sample forecast of the model. Fig. 5 and Fig. 6 depict the trajectory of infections for both Brazil and the

Philippines.
For Brazil, the depicted curve of the reported infections is increasing. The coupled LSTM predicts the lowest number of new confirmed cases, whereas the dynamic SIR predicts a much higher dynamic, with an increasing slope at the end of the forecasting period. The univariate network has a similar behaviour with predicting an increase in infection numbers at the end of the forecasting period, although at a lower level. The dynamics of the Philippines as given in Fig. 6 are less pronounced, with a lower rate of new confirmed infections and the overall level being lower at 0.013 percent of the population. A strong uptick in confirmed cases makes the long term out of sample forecasts inaccurate for the Philippines, where the total number of infections relative to the population is very low, whereas in Brazil, the coupled LSTM predicts the number of confirmed cases very well. For Brazil as well as the Philippines, the results suggests that Brazil is facing an increase in infection numbers whereas a strong uptick in confirmed Covid cases in the Philippines with an overall lower level of reported cases indicates another strong increase in the future confirmed cases.

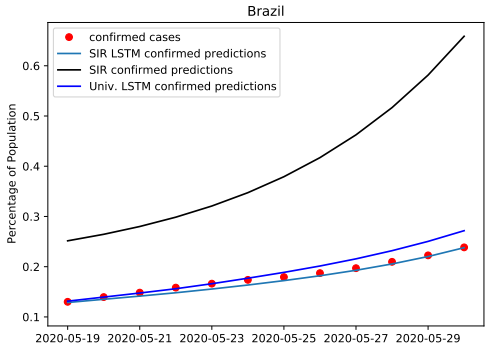

Figure 5: An illustration of a multiple days ahead out of sample forecast comparing different models, applied to Brazil.

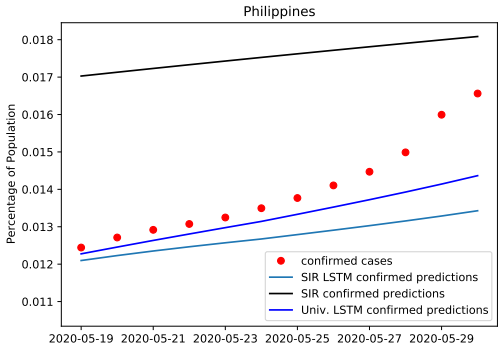

Figure 6: An illustration of a multiple days ahead out of sample forecast comparing different models, applied to the Philippines.

6. Conclusion

We introduced a novel epidemiological updating scheme allowing for forecasts and evaluation of the current covid19 outbreak worldwide. We combined compartmental models with recursive Bayesian updating, showing the advantage of real-time forecasting and parameter estimation which is combined with a LSTM network, creating a novel model type which we named Neural SIR model. It combines the forecasting accuracy of LSTM networks with the epidemiological model dynamics of the SIR model. When both models are coupled the LSTM learns the dynamics of the SIR model and can significantly increase accuracy, as our experiments have confirmed.
The model is transparent and parsimonious, allowing for additional compartments in the SIR model or additional input data in the network which makes it generalisable and accessible to policy makers worldwide.

Further research can analyze the coupling of specific hidden layers in different network architectures with the prediction and output cycle of the dynamic SIR model. We also suggest to investigate not just the increase in forecasting performance, but how the highly non-linear capabilities of the LSTM network can be used to conduct inference on latent parameters of the SIR model. We hope that

researchers and local policymakers can apply the model for initial estimates of infection rates and confirmed cases, which help to implement policies that help to mitigate further social and economic costs of the pandemic.

Acknowledgments

We are grateful for helpful discussions and feedback from Joseph Wu, Neil Ferguson and other participants at the Royal Society based conference "Scientists against Covid-19 and beyond". We furthermore like to thank Shuo Wang and Xian Yang for insightful discussions, as well as valuable feedback from Ovidiu Serban.

References

John hopkins university coronavirus resource center. https://coronavirus.jhu.edu/map.html, 2020. Accessed: 2020-05-24.

Henry DI Abarbanel, Paul J Rozdeba, and Sasha Shirman. Machine learning: deepest learning as statistical data assimilation problems. *Neural Computation*, 30(8): 2025–2055, 2018.

Roy M Anderson. Discussion: the kermack-mckendrick epidemic threshold theorem. *Bulletin of mathematical biology*, 53(1-2): 1, 1991.

Rossella Arcucci, Luisa D'Amore, Jenny Pistoia, Ralf Toumi, and Almerico Murli. On the variational data assimilation problem solving and sensitivity analysis. *Journal of Computational Physics*, 335:311–326, 2017.

Mark Asch, Marc Bocquet, and Maelle Nodet. *Data assimilation: methods, algorithms, and applications*. 12 2016. ISBN 978-1-611974-53-9.

R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. ii: Modelling the forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1971–1996, 2008. doi: 10.1002/qj.340.

Luis MA Bettencourt and Ruy M Ribeiro. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One*, 3(5), 2008.

Luis MA Bettencourt, Ruy M Ribeiro, Gerardo Chowell, Timothy Lant, and Carlos Castillo-Chavez. Towards real time epidemiology: data assimilation, modeling and anomaly detection of health surveillance data streams. *NSF Workshop on Intelligence and Security Informatics*, pages 79–90, 2007.

Loren Cobb, Ashok Krishnamurthy, Jan Mandel, and Jonathan D Beezley. Bayesian tracking of emerging epidemics using ensemble optimal statistical interpolation. *Spatial and spatio-temporal epidemiology*, 10:39–48, 2014.

Huseyin Coskun, Felix Achilles, Robert DiPietro, Nassir Navab, and Federico Tombari. Long short-term memory kalman filters: Recurrent neural estimators for pose regularization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5524–5532, 2017.

Guozhi Dong, Bert Jüttler, Otmar Scherzer, and Thomas Takacs. Convergence of tikhonov regularization for solving ill-posed operator equations with solutions defined on surfaces. *Inverse Problems & Imaging*, 11(2):221–246, 2017. ISSN 1930-8337. doi: 10.3934/ipi.2017011.

Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. 1999.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Natsuko Imai, Ilaria Dorigatti, Anne Cori, Steven Riley, and Neil M Ferguson. Estimating the potential total number of novel coronavirus cases in wuhan city, china, 2020.

Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

Valerio La Gatta, Vincenzo Moscato, Marco Postiglione, and Giancarlo Sperli. An epidemiological neural network exploiting dynamic graph structured data applied to the covid-19 outbreak. *IEEE Transactions on Big Data*, 2020.

Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 2020.

Andrew C Lorenc. The potential of the ensemble kalman filter for nwp—a comparison with 4d-var. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 129 (595):3183–3203, 2003.

Robert N Miller, Everett F Carter, and Sally T Blue. Data assimilation into nonlinear stochastic models. *Tellus A: Dynamic Meteorology and Oceanography*, 51 (2):167–194, 1999.

Philip Nadler, Shuo Wang, Rossella Arcucci, Xian Yang, and Yike Guo. An epidemiological modelling approach for covid19 via data assimilation. *European Journal of Epidemiology*, 35:749–761, 2020.

CJ Rhodes and T Déirdre Hollingsworth. Variational data assimilation with epidemic models. *Journal of theoretical biology*, 258(4):591–602, 2009.

Xuguang Wang, David Parrish, Daryl Kleist, and Jeffrey Whitaker. Gsi 3dvar-based ensemble–variational hybrid data assimilation for ncep global forecast system: Single-resolution experiments. *Monthly Weather Review*, 141(11):4098–4117, 2013.

Joseph T Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 2020.

SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.

Appendix A. Additional Figures

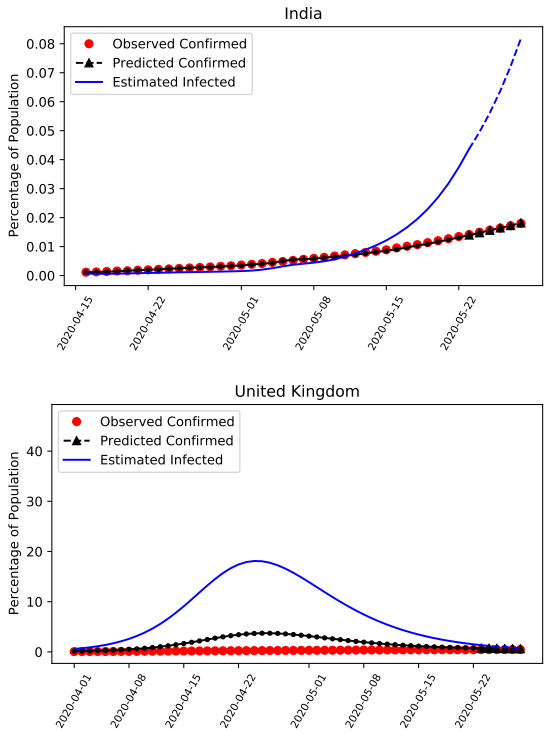

Figure 7: An Illustration of the latent infection numbers I generated by a SIR model using recursive Bayesian model updates

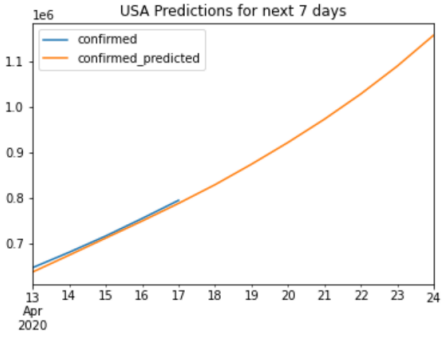

Figure 8: Illustration of the out of sample forecasting dynamics of a simple univariate LSTM network.

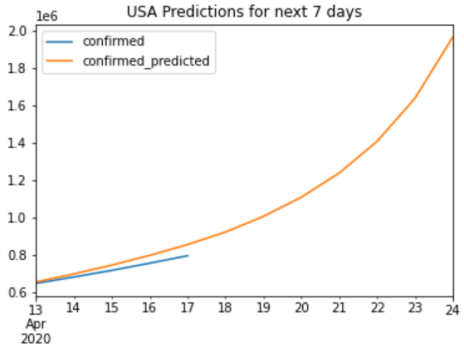

Figure 9: Illustration of the out of sample forecasting dynamics of a coupled SIR-LSTM network.

Appendix B. Covariance Matrix Estimation

To compute robust covariances for each model run, we estimate values for state and observation covariance matrices $\mathbf{Q}$ and $\mathbf{P}$ using an ensemble approach Wang et al. (2013). The values for $\mathbf{P}$ are based on an estimate of the residual covariance matrix of the observed time series.

Following the cost function give by Eq. 6, with $\mathbf{x}^b$ representing an individual background state vector, the full ensemble of state vectors is given by

$$\mathbf{x}^b_{(1)}, \mathbf{x}^b_{(2)}, .., \mathbf{x}^b_{(N)} \tag{14}$$

If the ensemble mean is defined as $\overline{\mathbf{x}}^b$, then $\mathbf{V}_{ens}$, the background state perturbations are computed via

$$\mathbf{V}_{ens} = \mathbf{X}^b = \frac{1}{\sqrt{N-1}}(\mathbf{x}^b_{(1)} - \overline{\mathbf{x}}^b, \mathbf{x}^b_{(2)} - \overline{\mathbf{x}}^b, ..., \mathbf{x}^b_{(N)} - \overline{\mathbf{x}}^b) \tag{15}$$

In this case, $\mathbf{V}_{ens}$ and $\mathbf{X}^b$ are a n x N matrix called the ensemble background perturbation matrix. The rank-deficient version of the background error covariance matrix is defined as $\mathbf{Q}^*$ with

$$\mathbf{Q}^* = \mathbf{X}^{bT}\mathbf{X}^b \tag{16}$$

The ensemble is static, meaning that it does not evolve dynamically with time, but it still incorporates flow-dependent information at the start time which is still beneficial for an extended Kalman filter analysis. The way the ensembles are chosen and computed determines the accuracy of ensemble updating schemes.

The ensemble needs to be computed in such a way that the time dependent variability of the background error covariance matrix, as well as the correlation of variables is captured by the sampling procedure. The method we devise is to divide the collection of background states, $\underline{\mathbf{x}}^b$ based on the size of the ensemble into N equally sized groups with each group being denoted by $\underline{\mathbf{x}}^b_{(i)}$ meaning that ensemble members belong to the ith group. The mean and standard deviation of each group is then estimated and used to sample the ensemble members from.

Algorithm 1 Build Ensemble

Inputs: $\underline{\mathbf{x}}^b$
$i = 0$, N = ensemble size, $n = length(\mathbf{x}^b)$
for $\underline{\mathbf{x}}^b_{(i)}$ in $array\_split(\underline{\mathbf{x}}^b, N)$ **do**
  $\mu_{(i)} = mean(\underline{\mathbf{x}}^b_{(i)})$
  $\sigma_{(i)} = standard\_deviation(\underline{\mathbf{x}}^b_{(i)})$
  $ensemble[:, i] = normal\_distribution(\mu_{(i)}), \sigma^2_{(i)})$
  $i = i + 1$
end for
$ensemble\_mean = mean(ensemble)$
for $i = 0, 1, .., N$ **do**
  $\mathbf{V}_{ens}[:, i] = ensemble[:, i] - ensemble\_mean$
end for
return $\mathbf{V}_{ens}$

Algorithm 1 describes in detail how $\mathbf{V}_{ens}$ is computed and ensembles are formed. The full background state matrix, $\underline{\mathbf{x}}^b$ is split into N groups each of size $n \times \frac{n}{N}$. Both, the means as well as the standard deviations of the n rows are estimated and used to generate draws from a multivariate Gaussian distribution to form the ensemble. In order to form $\mathbf{V}_{ens}$, for each ensemble member the corresponding mean is estimated and then subtracted, computing the standard deviation.

Appendix C. Data Overview

Table 3: Overview of country level data for the last available date 2020/05/18, showing the ratio of confirmed reported cases

	Population	Conf.Cases	Ratio
India	1.34E+09	1.65E+05	1.24E-04
Brazil	2.09E+08	4.38E+05	2.10E-03
Turkey	8.00E+07	1.61E+05	2.01E-03
Philipp.	1.04E+08	1.56E+04	1.50E-04
US	3.28E+08	1.72E+06	5.25E-03
UK	6.60E+07	2.69E+05	4.08E-03