Forecasting Influenza-like Illness using Physics Informed Neural Ordinary Differential Equations

Anonymous Authors[1]

Abstract

We present a method of using Physics informed neural ODEs to forecast influenza at a state, regional, and national level in the US. It is better than previous works and is more stable/better etc etc.

1. Introduction

Estimating the current and future prevalence of an infectious disease enables public health organisations to prepare for and mitigate the effects of that disease. The impact of forecasting models was highlighted during the Covid-19 pandemic where forecasts informed policy decisions aimed at reducing the prevalence of Covid-19 through lock-downs and social distancing. While the pandemic is over there is still a need to forecast both established and novel infectious diseases. According to the World Health Organisation, Influenza is responsible for between 290,000 and 650,000 deaths each year with the potential to cause a pandemic. Forecasting the prevalence of influenza allows policy makers to decide when and whe to focus their efforts, such as vaccinations, hygiene programs and antivirals.

We focus on influenza forecasting due to the well established datasets going back to 2004 at a US national level and to 2010 at a state level. In the US the prevalence of influenza is monitored through a syndromic surveillance network of 2000 healthcare providers across the US. Instead of measuring the prevalence of influenza they substitute it for a proxy influenza-like-illness (ILI). ILI is defined as a fever (temperature of 37.8° or greater) and cough and/or sore throat without a known cause besides influenza. ILI is weighted by population to estimate the weighted ILI (wILI) at a regional and national level. ILI is a noisy signal which may account for other diseases such as RSV, and more recently Covid-19. There is also considerable noise in the ILI rate, which has been observed to be inversely related to the population of the region. For example, Alaska will have a more noisy ILI rate than California - which has a much larger population. The influence of other diseases and population dependent noise result in a signal which is difficult to forecast.

The difficulty is compounded because ILI data is collected with a delay of up to two weeks due to the time taken to collect and process the case counts over a large networks of healthcare providers. Conversely, exogenous data can be collected in real time and may provide a more timely indication of the ILI rate. Exogenous variables provide additional information which can be used to improve forecast accuracy. The use of exogenous Web search activity data is well established ILI modelling—(Ginsberg et al., 2009; Yang et al., 2015; Lampos et al., 2015; 2017; Morris et al., 2023). Web search data refers to frequencies of searches using a user defined keyword or keywords in a web browser i.e., Google, for a set region and time period. For example the frequency of searches for "flu remedies" in California on the first of January. These searches include symptoms, medicines, information about different strains etc. The methods proposed in (Morris et al., 2023) leverages this more recent Web-search data to improve their forecasts. We incorporate a similar method and incorporate Web search data into our models to improve forecasting performance.

Epidemic forecasting models fall into two camps, mechanistic and non-mechanistic models, often these are referred to as mechanistic and statistical, however mechanistic models are trained as statistical models so we avoid this distinction. Mechanistic estimate the spread of a disease by modelling the underlying process of individuals (or groups of individuals) interacting with the disease, becoming infected and later recovering or dying. These models are either compartmental - where the population is split up into groups where everyone in a group has the same interaction with the disease at a given time, described by an ordinary differential equation (ODE), or agent based models where individuals are modelled including. We focus on compartmental models which are better suited to forecasting as they are more trainable.

Non-mechanistic models learn patterns directly from train-

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

ing data with no prior knowledge of the underlying mechanism that they are modelling. These tend to be more accurate for epidemic modelling (Reich et al., 2019). Recent work using Bayesian neural networks(Morris et al., 2023) used Web search activity data to forecast ILI accurately with well calibrated uncertainty.

Neural networks have several advantages over mechanistic models, by removing assumptions and learning directly from data they are more flexible — meaning that they can learn to fit any data regardless of how complex it is. This flexibility extends to enabling them to incorporate exogenous variables with minimal change to the model. This benefit is made possible because neural networks are easy to scale, the limiting factory on the number of inputs is the size of the training set which is available (a larger training allows the model to learn more complex relationships between more diverse inputs).

However, neural networks are limited by their dependence on large datasets to learn patterns in the data - this is seldom available for disease forecasting, even the US only has weekly state level data going back to 2010 ($\approx$ 800 data points). It is also unclear how well these models will generalise to a state level where the ILI signal is far more noisy, and the Web search activity data is more sparse.

To make predictions they must learn parameters from training data which does not always generalise across different seasons or diseases. This means that for every new flu season there is no guarantee of the model's performance. The problem is compounded in unusual or epidemic seasons where accurate forecasts are potentially the most useful.

Compartmental models are less dependent on large datasets. Due to their grounding in physical assumptions they behave reasonably in out-of-sample scenarios. Physical assumptions do not guarantee accuracy, but they do reduce the model's dependence on training data to learn patterns.

The advantages and disadvantages of mechanistic models and neural networks are contrasting, mechanistic models are less flexible but well understood with statistical guarantees and less reliance on data. Neural networks are flexible and accurate, but poorly understood without statistical guarantees and require large datasets and careful training to provide good results. We present a physics informed neural network which benefits from the advantages of both models while providing more accurate ILI forecasts than the top performing statistical models at a state, regional, and national level in the United States.

The benefits of mechanistic models mean that they are popular in the epidemiology community, this is despite the fact that non-mechanistic models are known to be more accurate for forecasting. The disadvantages of mechanistic and non-mechanistic models are unavoidable, but it is possible to design a model to selectively use mechanistic and non-mechanistic components to benefit from the advantages of both.

1.1. Compartmental Models

Next I describe how a model can be designed to benefit from both approaches. I focus on using an SIR model, but the method is generally applicable to any compartmental model. The equations of an SIR model are:

$$\frac{ds}{dt} = -\beta \cdot s \cdot i$$
$$\frac{di}{dt} = \beta \cdot s \cdot i - \omega \cdot i$$
$$\frac{dr}{dt} = \omega \cdot i$$

where s, i and r are the susceptible, infected and recovered proportion of the population. The transmission rate β represents the rate at which susceptible individuals become infected upon contact with infectious individuals. The recovery rate ω represents the rate at which infectious individuals recover and move to the recovered state. I use ω instead of γ because we often use γ as the forecast horizon.

In ILI modelling there are four areas that we cannot easily model with a mechanistic model:

1. We do not know, and cannot measure the exact number of susceptible, infected and recovered people in a population, models with more compartments have more unknowns.
2. We do not know the β and ω parameters which govern the disease spread. But, in some situations, β or γ may be known and we want to be able to easily modify the model to use this additional information. In others we may have a rough idea but do not know exactly, again we would like to be able to leverage this information.
3. We do not know, and cannot measure the relationship between the s, i and r parameters and the measured ILI rate. This is similar to the first problem, except in the first we want to estimate s, i and r from the ILI rate, and here we want to estimate ILI rate from s, i, and r.
4. No compartmental model exists which can exactly fit to the ILI trajectory, instead we want to be able to model the discrepancy between the compartmental model and the ground truth.

1.2. Contributions

Our main contribution is a generalised framework that solves each of these problems using separate neural networks, as well as a method for training the combined model

directly from data in a reliable way. If for any reason one of these components is not required, for example if we know the initial conditions, or know the relationship between the ground truth and the model components then we can add or remove the neural network and replace it with a mechanistic component.

The combined model uses a known mechanistic model up to the limit of its forecasting ability and uses neural networks that learn patterns from the data to improve accuracy. The compartmental model still behaves in the same way as existing compartmental models, we can decompose the augmentation and SIR and see what is happening, including observing the estimated parameters for any timestep. This guarantees reasonable behaviour including in unseen situations. Introducing physical constraints should also reduce the dependence on large datasets and thus improve the forecasting accuracy.

2. Methods

First we describe the datasets, then the model setup and architectures. Different latencies from different data sources in ILI forecasting can introduce ambiguity to forecasting setup. We obtain ILI rates up to t_0, however there is a reporting latency of $\delta = 14$ days in ILI data. We can download web search data up to the current time i.e., for $t > t_0$. Thus in our experiments we use ILI data up to t_0 and web search data up to $t_0 + \delta$. When we refer to the forecast horizon γ it is from time t_0. This is consistent with other forecasting literature, but it should be noted that forecasts up to $\gamma = 14$ are actually estimating previous (but still unknown) ILI rates.

2.1. Datasets

Influenza-like illness rates

Influenza-Like Illness (ILI) is characterized by a fever (temperature of 37.8° or higher)) and cough and/or sore throat without a known cause besides influenza. To monitor the prevalence of ILI, various surveillance efforts, including the Outpatient Influenza-Like-Illness Surveillance Network (ILINet), have been established. ILINet collects weekly ILI rates at the state level from over 2,000 healthcare providers across all states.

For forecasting purposes, ILI is projected at the state, regional, and national levels. The Office of Intergovernmental and External Affairs oversees 10 Regional Offices, referred to as Health and Human Services (HHS) regions. Each HHS region encompasses multiple states within a larger geographic area. To provide a ILI rates at larger geographic regions, state-level ILI rates are weighted by population size, resulting in the calculation of Weighted ILI (wILI) at HHS regions and National level.

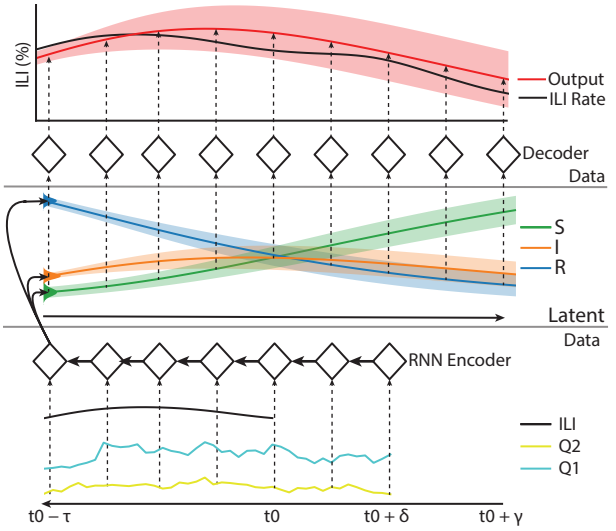

Figure 1. **VAE Diagram**
The RNN Encoder observes ILI rates and queries backwards in time from $t_0 + \delta$ to $t_0 - \tau$

Web Search query frequency

Search query frequencies for the US are obtained from the Google Health Trends API similarly to other studies (Morris et al., 2023; Lampos et al., 2021; 2017). A frequency represents the fraction of searches for a certain term or set of terms divided by the total amount of searches (for any term) for a day and a certain location. We employ the same query selection method as in our prior work using Bayesian neural networks (Morris et al., 2023). We initially downloaded a predetermined pool of the 500 best queries for national level forecasting. State level query frequencies are aggregated based on population size to HHS level. For each test season we choose the top m queries based on the semantic similarity to flu (Lampos et al., 2017) correlation between each query and the preceding five years of ILI rates (Morris et al., 2023). Query frequencies are smoothed using a 7-day moving average, and min-max normalisation is applied to each query's time series during training (i.e. without using any future data).

2.2. Architectures

We use a variational autoencoder (VAE) architecture where the encoder observes sequences of ILI rates and web search query frequencies from each region to estimate latent initial conditions. A latent ODE model integrates the initial conditions forwards in time and a decoder converts the latent trajectories back into ILI rates. We maintain the same architecture for the encoder and decoder, but vary the ODE model to compare different methods.

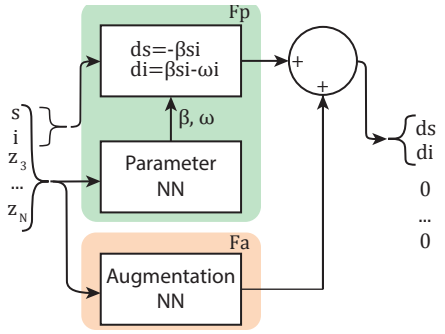

Figure 2. **UONN Diagram**
The UONN, here shown for a single region, is a universal differential equation made up of a physical model (Fp) and an augmentation model (Fa). The physical model is derived an SIR model where the parameters β, ω are estimated by the parameter neural network. The augmentation NN minimises the error between Fp and the ILI rate. Both NNs use the SIR inputs as well as additional variables $Z_{1:N}$ which are determined by the encoder. This enables gives the model additional information about the disease which would not be available from the ODE compartments. CONN uses only Fp, whereas SONN uses only Fa, the architectures are otherwise identical.

Encoder

We employ a Gated Recurrent Unit (GRU) (Cho et al., 2014)-based Recurrent Neural Network (RNN) to process queries and Influenza-Like Illness (ILI) rates backwards in time from $t_0 + \delta$ and t_0 to $t_0 - \tau$, respectively. Missing values in the ILI rate time series are padded with -1, resulting in a $[b \times (\tau + \delta) \times N_r(1 + m)]$ matrix. Here, b represents the batch size, τ is the window size, N_r is the number of regions (1 for national, 10 for regional, and 49 for state[1]), and m is the number of search queries used by each region. The encoder comprises L_{GRU} GRU layers. The output of the last GRU layer at time $t_0 - \tau$ is then fed into L_{FF} feed-forward layers. The selection of the number and size of these GRU and FF layers is determined through hyperparameter tuning. The output of the encoder $Z_{t_0-\tau}$ is a $[2 \times N_r \times z_{\text{dim}} - 1]$ matrix, where each region is associated with z_{dim} values, containing means and standard deviations of the latent distribution for each region.

We use the reparameterisation trick to sample from the latent distribution. The first c values of the latent distribution for each region correspond to the initial conditions for each of the c compartments in the compartmental model. The constraint $\sum_{i=1}^{c} z_{i,r} = 1$ holds, where r denotes the region. Therefore, we compute the initial value of the c^{th} (recovered) compartment as $1 - \sum_{i=1}^{c-1} z_{i,r}$. The remaining values in the latent distribution supply additional information to the ODE model so that it can make more accurate predictions. If we only supply the model with the initial conditions for the c compartments the ODE cannot make informed estimates of the disease transmission characteristics. The sampled initial conditions are reshaped into a $[b * N_s, N_r * z_{\text{dim}}]$ matrix, where N_s is the number of samples taken from the latent distribution.

[1] no ILI data available for Florida

Latent ODE Models

We employ three ODE models, A traditional neural-ODE without a mechanistic component, denoted Simple-ODE-NN (SONN). A neural-ODE that estimates parameters of a compartmental model denoted Compartmental-ODE-NN (CONN), and a universal-differential-equation (UDE) model derived from CONN, incorporating an additional augmentation neural network denoted Universal-ODE-NN (UONN).

The SONN model is an L_{ODE} layered FF neural network, it observes the latent states Z_t at time t, and estimates gradients $^{dZ_t}/_{dt}$.

The CONN model is an SIR model which uses the same SIR equations but the parameters vary with time and are estimated by a neural network in the ODE function. The ODE function uses a neural network to estimate the SIR parameters $[\beta_t, \omega_t] = F_{\theta_{ode}}(x_t)$, where $F_{\theta_{ode}}(x_t)$ represents a neural network with parameters θ_{ode}. The network consists of three layers. The first two use the eLu activation function. The final layer uses an `abs` activation function to prevent β and ω from being negative. Parameters for each regions are computed simultaneously, however there is no interaction between the regions. For example infections from California cannot affect individuals in Nevada.

The UONN model uses the same physical model as CONN with an additional non-mechanistic component defined by a neural network. At a national level this augments the ODE to become a universal differential equation. For the state and regional models the augmentation model can also estimate transmission between regions. This enables the model to behave as a metapopulation model in addition to being a UDE. The compartmental component acts on each region individually,

Decoder

The decoder is an single layered FF neural network, it observes the latent trajectories and converts estimates N_r ILI rates. For the CONNand UONNthe decoder only observes the $c * N_r$ latent trajectories associated with the compartmental models for each region. For the SONN it observes all $Z_{\text{dim}} * N_r$ trajectories.

2.3. Training

We train the models using the Evidence-Lower-Bound (ELBO) where the likelihood term is calculated between the distribution over trajectories (calculated by sampling from the latent initial conditions) and the ground truth. We

sample K times from the latent initial conditions, integrate each of the K samples forward to $t_0 + \gamma$, then decode the latent trajectories. We then compute the mean and standard deviation over the K samples which are used to compute the NLL. We find that a larger value of K results in more stable training at the cost of computational complexity. For this reason, we begin the training with $K = 32$ samples, and double K after 100 epochs, and again after 400 epochs (we train for a total of 500 epochs). We find this to be increase the training speed while allowing the models to converge.

The prior distribution over the latent initial conditions is $\mathcal{N}(0, 1)$. However, for the mechanistic components of the ODE models — s and i for an SIR model — we use prior distributions specific to that component which are chosen via tuning. For the susceptible and infectious fractions the prior distribution is $p(s_{t_0-\tau}, i_{t_0-\tau}) = \mathcal{N}\left(z_{\mu_{t_0-\tau}}, [0.1, 0.01]\right)$. Thus, the means are the estimated means taken from the output of the encoder and only the standard deviations are regularised. We make this design choice because the standard deviations of the latent initial conditions are small but the mean is situational, for instance the susceptible fraction may change from ≈ 1 to ≈ 0 over a flu season.

In addition to regularising the latent initial conditions, we regularise the model to choose reasonable values for β and ω. Due to the samples from the initial conditions we have a time-varying distribution over β and ω. We regularise the parameters to a prior $p_{\theta_P}(\beta, \omega) = \mathcal{N}([0.8, 0.55], [0.1, 0.1])$ using the KL divergence:

$$\mathcal{L}_{KL_p} = \mathrm{D}_{\mathrm{KL}}(q_{\theta_{ode}}(\beta, \omega)||p(\beta, \omega)), \tag{1}$$

where $q_{\theta_{ode}}(\beta, \omega)$ is the posterior distribution over the parameters. We chose the prior based on related work (Osthus et al., 2019), we experimented with variations on this prior but found no noticeable improvement.

Stability
Variational Autoencoders (VAEs) often exhibit instability during training, characterized by high reconstruction loss (NLL in our case) or KL losses. Following prior work on VAEs (Child, 2020), we skip updates with a gradient norm surpassing a threshold set by a hyperparameter. The threshold is chosen empirically, and a plot of gradient norms during training is provided in Supplementary Figure ABC.

In addition to skipping updates with exceptionally large gradient norms, we also employ cyclical annealing (Fu et al., 2019) to improve stability and mitigate KL vanishing. Through hyperparameter tuning, we observe an average 25% improvement in NLL using a cosine KL annealing scheduler.

For the physics-informed models (CONNand UONN), we introduce an additional loss function to penalize instances where the latent trajectories generated by the compartmental model exceed 1 or fall below 0:

$$\mathcal{L}_{\text{reg}}(z) = \sum_t \begin{cases} |z_t| - 1 & \text{if } z_t > 1 \\ |z_t| & \text{if } z_t < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

This is only necessary when the initial conditions are poorly specified and $s_0 + i_0 > 1$. Despite accelerating convergence and improving stability, this addition to the loss function does not impact model performance after training; for trained models, $\mathcal{L}_{\text{reg}}(z)$ should always equal 0. Additionally, we set the outputs of the ODE model to zero when the corresponding input is > 2 or -1 to enhance stability during early training stages. However, this adjustment has no effect once the model can produce reasonable forecasts.

Lastly, we perform pre-training on the encoder to minimise the KL divergence from the prior. Integrating and backpropagating with the Ordinary Differential Equation (ODE) solver are the slowest phases of training. Pre-training using only the encoder significantly accelerates convergence.

3. Experiments

3.1. Metrics

We use NLL and

3.2. Baselines

What I've done:

1. Tuning for HHS CONN, was slow (2 weeks CPU 24x) because I didn't know which encoder architecture would work best and I added a lot of stability/speed improvements - see methods. Future tuning is faster (approximately 3x faster but dependent on server space).
2. Identified best encoder - GRU reading backward.
3. Set up experiments for other tasks, haven't done national yet but this *should be* the easiest one.
4. written methods - not complete as experiments are not complete.
5. HHS Results done
6. Changed UONN training to be CONN + post training with Fa - this works better, needs to be quantified.

Problems
Dante not working at state or regional level. ODEs not working at national level. Regional level ODEs are better than national level. Tried different tuning, tried pre training. Looks like ODEs are too confident - NLL is terrible for national level ODEs. could be overfitting? Could be a lack of uncertainty in the actual ODE.

Horizon	7		14		21		28	
	NLL	Skill	NLL	Skill	NLL	Skill	NLL	Skill
State								
IRNN_S		0.03		0.04		0.05		0.05
SONN	2.14	0.20	2.17	0.18	2.17	0.18	2.18	0.17
CONN	0.88	0.39	1.04	0.36	1.23	0.32	1.51	0.28
UONN	1.51	0.26	2.37	0.25	2.44	0.23	2.51	0.21
Regional								
IRNN_S		0.49		0.44		0.40		0.34
SONN	0.90	0.46	0.97	0.42	1.06	0.39	1.13	0.36
CONN	1.31	0.30	1.47	0.27	1.66	0.24	1.89	0.21
UONN	1.56	0.24	1.62	0.22	1.71	0.21	1.82	0.19
National								
IRNN_S	**0.24**	**0.71**	**0.24**	**0.71**	0.67	**0.62**	0.87	**0.58**
SONN	1.73	0.64	1.78	0.58	1.86	0.50	1.93	0.42
CONN	0.28	0.67	0.43	0.65	**0.59**	0.60	**0.79**	0.53
UONN	0.50	0.57	0.58	0.55	0.73	0.50	0.92	0.46

Table 1. **Results Table**
Results averaged over all flu seasons.

Currently running

1.

To do

1. Dante - had problems with this and it only worked for 2017/18, the other seasons' results are terrible, this will work at all geographic levels in 1 go though. Not sure what the problem is but potential to just remove it and use the IRNN as a baseline + could include something basic like Bayesian linear regression.

Improvements

1. Changed training for UONN, works now.

Deadlines

1. ICML - 1st February
2. CHIL - 6th February (healthcare orientated)
3. KDD - 9th February (Data Science: Methods for analyzing social networks, time series, sequences, streams, text, web, graphs, rules, patterns, logs, IoT data, spatio-temporal data, biological data, scientific and business data; recommender systems, computational advertising, multimedia, finance, bioinformatics.)
4. If I miss these the next ones are in April, and NeurIPS in May. Funding till end of March, I think without a deadline this would take me until March.
5. **PhD Viva is on 5th February, need a few days to prepare at some point**

3.3. Results

Averages (for the paper)

4. Related Work

5. Bibliography

References

Child, R. Very deep vaes generalize autoregressive models and can outperform them on images. *arXiv preprint arXiv:2011.10650*, 2020.

Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

Fu, H., Li, C., Liu, X., Gao, J., Celikyilmaz, A., and Carin, L. Cyclical annealing schedule: A simple approach to mitigating kl vanishing. *arXiv preprint arXiv:1903.10145*, 2019.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. Detecting influenza epidemics using search engine query data. *Nature*, 457 (7232):1012–1014, 2009.

Lampos, V., Miller, A. C., Crossan, S., and Stefansen, C. Advances in nowcasting influenza-like illness rates using search query logs. *Scientific reports*, 5(1):12760, 2015.

Lampos, V., Zou, B., and Cox, I. J. Enhancing feature selection using word embeddings: The case of flu surveillance. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 695–704, 2017.

Lampos, V., Majumder, M. S., Yom-Tov, E., Edelstein, M., Moura, S., Hamada, Y., Rangaka, M. X., McKendry, R. A., and Cox, I. J. Tracking covid-19 using online search. *NPJ digital medicine*, 4(1):17, 2021.

Morris, M., Hayes, P., Cox, I. J., and Lampos, V. Neural network models for influenza forecasting with associated uncertainty using web search activity trends. *PLoS Computational Biology*, 19(8):e1011392, 2023.

Osthus, D., Gattiker, J., Priedhorsky, R., and Del Valle, S. Y. Dynamic Bayesian influenza forecasting in the United States with hierarchical discrepancy (with discussion). *Bayesian Analysis*, 14(1):261–312, 2019.

Reich, N. G., Brooks, L. C., Fox, S. J., Kandula, S., McGowan, C. J., Moore, E., Osthus, D., Ray, E. L., Tushar, A., Yamana, T. K., et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the united states. *Proceedings of the National Academy of Sciences*, 116(8):3146–3154, 2019.

Yang, S., Santillana, M., and Kou, S. C. Accurate estimation of influenza epidemics using google search data via argo. *Proceedings of the National Academy of Sciences*, 112 (47):14473–14478, 2015.

$\gamma = \mathbf{7}$

Model	2015	2016	2017	2018	2015-18
SONN	0.27	0.19	0.10	0.15	0.17
CONN	0.36	0.23	0.23	0.22	0.25
UONN	0.33	0.30	0.21	0.26	0.27
IRNN	0.06	0.07	0.04	0.01	0.03
Dante					

$\gamma = \mathbf{14}$

Model	2015	2016	2017	2018	2015-18
SONN	0.27	0.17	0.09	0.14	0.16
CONN	0.33	0.21	0.21	0.21	0.24
UONN	0.32	0.28	0.19	0.24	0.25
IRNN	0.03	0.08	0.06	0.02	0.04
Dante					

$\gamma = \mathbf{21}$

Model	2015	2016	2017	2018	2015-18
SONN	0.27	0.16	0.09	0.14	0.15
CONN	0.30	0.21	0.19	0.19	0.22
UONN	0.30	0.26	0.16	0.23	0.23
IRNN	0.04	0.07	0.07	0.03	0.05
Dante					

$\gamma = \mathbf{28}$

Model	2015	2016	2017	2018	2015-18
SONN	0.26	0.15	0.08	0.13	0.14
CONN	0.26	0.21	0.18	0.17	0.20
UONN	0.28	0.25	0.14	0.22	0.22
IRNN	0.04	0.08	0.07	0.03	0.05
Dante					

Table 2. State Level Skill for all Seasons and Horizons, missing values to be populated as they become available. Something wrong with state IRNN, i'm not sure what yet.

6. Appendix

$\gamma = \mathbf{7}$

Model	2015	2016	2017	2018	2015-18
SONN	0.46	0.42	0.50	0.40	0.44
CONN	0.49	0.49	0.40	0.48	0.46
UONN	0.63	0.44	0.45	0.48	0.50
IRNN	0.63	0.48	0.45	0.44	0.49
Dante					

$\gamma = \mathbf{14}$

Model	2015	2016	2017	2018	2015-18
SONN	0.41	0.39	0.46	0.33	0.39
CONN	0.45	0.45	0.36	0.41	0.42
UONN	0.57	0.40	0.42	0.44	0.46
IRNN	0.58	0.42	0.40	0.38	0.44
Dante					

$\gamma = \mathbf{21}$

Model	2015	2016	2017	2018	2015-18
SONN	0.38	0.37	0.42	0.26	0.35
CONN	0.42	0.42	0.34	0.36	0.38
UONN	0.52	0.37	0.38	0.40	0.41
IRNN	0.52	0.38	0.38	0.34	0.40
Dante					

$\gamma = \mathbf{28}$

Model	2015	2016	2017	2018	2015-18
SONN	0.35	0.35	0.39	0.24	0.33
CONN	0.38	0.41	0.32	0.33	0.36
UONN	0.48	0.33	0.34	0.37	0.38
IRNN	0.42	0.31	0.33	0.30	0.34
Dante					

Table 3. HHS Level Skill for all Seasons and Horizons, missing values to be populated as they become available.