Neural Relational Autoregression for High-Resolution COVID-19 Forecasting

Matthew Le
Facebook AI Research
New York, NY, United States
mattle@fb.com

Mark Ibrahim
Facebook AI Research
New York, NY, United States
marksibrahim@fb.com

Levent Sagun
Facebook AI Research
Paris, France
leventsagun@fb.com

Timothee Lacroix
Facebook AI Research
Paris, France
tlacroix@fb.com

Maximilian Nickel
Facebook AI Research
New York, NY, United States
maxn@fb.com

ABSTRACT

Forecasting COVID-19 poses unique challenges due to the novelty of the disease, its unknown characteristics, and substantial but varying interventions to reduce its spread. To improve the quality and robustness of forecasts, we propose a new method which aims to disentangle time-varying and region-specific factors – such as enacted policies and mobility – from disease-inherent factors that influence its spread. For this purpose, we combine deep learning with a vector autoregressive model and train the joint model with a novel regularization scheme that increases the coupling between regions. This approach is akin to using Granger causality as a relational inductive bias and allows us to train high-resolution models that borrow statistical strength across regions. Our method has been deployed since early in the pandemic to assist response teams and we observe strong performance in forecasting COVID-19 when compared to state-of-the-art forecasts.

ACM Reference Format:
Matthew Le, Mark Ibrahim, Levent Sagun, Timothee Lacroix, and Maximilian Nickel. 2021. Neural Relational Autoregression for High-Resolution COVID-19 Forecasting. In *epiDAMIK 2021: 4th epiDAMIK ACM SIGKDD International Workshop on Epidemiology meets Data Mining and Knowledge Discovery*. ACM, New York, NY, USA, 8 pages. https://doi.org/xx.xxxx/xxxxxxxxx.xxxxxxx

1 INTRODUCTION

Modeling the spread of COVID-19 at a high spatial and temporal resolution (i.e., confirmed cases at county or admin-3 level) is an important task in the public health response to the disease. High-quality forecasts at county-level are not only central to monitor the state of the pandemic but are also important to efficiently allocate scarce resources such as ventilators, personal protective equipment, and ICU beds; as well as to making progress towards early detection systems.

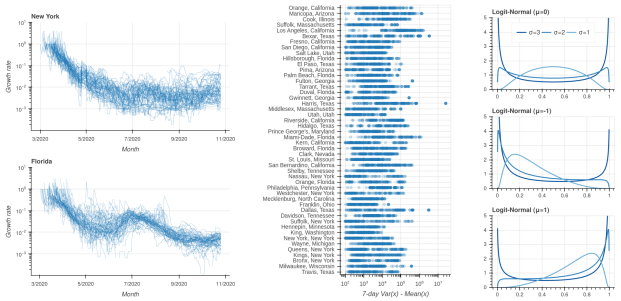
epiDAMIK 2021, Aug 15, 2021, Virtual

ACM ISBN 978-1-xxxx-XXXX-X…$15.00
https://doi.org/xx.xxxx/xxxxxxxxx.xxxxxxx

However, forecasting COVID-19 poses unique challenges – especially when considering confirmed cases at high spatial resolution. Although there has been considerable progress towards understanding the disease process, there still exists only limited data and knowledge about important factors that influence its spread. Due to the global nature of COVID-19, the disease occurs also among regions with substantially different properties, many of which may affect its spread and which change over time. This includes, for instance, demographics and population densities, enacted policies, adherence to those policies, mobility patterns, and geographic features such as temperature and UV radiation. In addition, testing and reporting can vary significantly across regions and time. All these factors lead to considerable variability and uncertainty in the data and make reliable forecasting at high spatial resolution difficult (see also fig. 1a-b). This is further exacerbated by the larger amount of noise in county-level data.

To alleviate these issues, we propose a new method for predicting time-varying disease processes that combines recurrent neural networks with a vector autoregressive (VAR) model and a novel regularization scheme. Our approach is motivated by two main aspects: First, we seek to develop a model which can use covariate data like mobility to estimate the time-varying force of infection of the disease. Basing this on recurrent networks allows us to develop an end-to-end differentiable model which is able to make efficient use of the limited available data while providing enough flexibility to capture the large variability of cases across locations and time. However, while such flexible models are needed to account for possible factors the influence the spread of the disease, there exists only little data to estimate them reliably and without overfitting. For this reason, we seek, second, to disentangle region- and time-specific factors from disease-inherent factors that influence its spread. This allows us to borrow statistical strength between regions by coupling their predictions. Moreover, it allows us to develop a relational inductive bias akin to *Granger causality* that improves the quality of the time-varying component of the model (based on the assumption that information about the spread in region j can help to improve predictions for a related region i once a model has correctly accounted for region-specific dynamics). For this purpose, we introduce also a novel regularization scheme to control the number of Granger-related time series in a VAR model.

Compared to existing state-of-the-art forecasting models, our method takes a highly data-driven approach with fewer modeling



(a) (b) (c)

Figure 1: a) Growth of confirmed cases on the example of New York and Florida. Each line represents one county. For y_t denoting the number of cases at time t, growth rate is computed as $(y_{t+1}-y_t)/y_t$. b) Overdispersion of daily case counts in US states and counties with most number of cases. c) The Logit-Normal distribution is a probability distribution of a random variable whose logit has a normal distribution, i.e., $\phi(\mathcal{N}(\mu, \sigma))$. As $\sigma \to \infty$ the Logit-Normal exhibits samples close to 0 and 1.

assumptions as, for instance, in very detailed compartmental models. Our method has been deployed since early in the pandemic to support response teams and has been very successful in forecasting the spread of the disease with high spatial and temporal resolution under real-world conditions. The full implementation of our model is open-source and available at anonymized.

2 NEURAL RELATIONAL AUTOREGRESSION

We consider the forecasting of m time series where interventions and regional factors lead to varying realizations of the same underlying disease process. In particular, let $\mathcal{Y} = \{(y_i^1, \ldots, y_i^T)\}_{i=1}^m$ denote observed case counts where i indexes locations and where T denotes the maximum observation time. Furthermore, let $\mathcal{Y}(\tau) = \{(y_i^t : t \leq \tau)\}_{i=1}^m$ denote tobserved case counts up to time $\tau \leq T$. We then model the number of confirmed cases as random variables

$$Y_i^{t+1} \mid \mathcal{Y}(t) \sim f(\lambda_i^t)$$

where λ_i^t denotes the *force of infection* at time t in location i; and where $f(x)$ denotes a probability distribution for count data with parameter x (e.g., a Poisson or Negative Binomial distribution).

Due to different interventions during the pandemic, we regard $\mathcal{Y}$ as a time-varying process that is influenced by external factors such as policies (e.g., lockdowns) and behavior (e.g., mobility, mask-wearing). For this reason, we decompose λ_i^t into a time-specific component β_i^t and a time-indepedent component λ_i such that

$$\lambda_i^t = \beta_i^t \lambda_i \quad \text{where} \quad \beta_i^t \in [0, 1],\ \lambda_i > 0 \tag{1}$$

Hence, β_i^t can be understood as a dampening factor that models the effect of interventions on the underlying force of infection λ_i and depends on time and location. While some influencing factors for the evolution of β_i^t might be known (e.g., mobility, population density, etc.), we assume that the full set of influencing factors is unknown and will regard β_i^t as a latent variable. Using this decomposition, we then model the time-independent force of infection as a autoregressive model of order p, i.e.,

$$\text{AR}(p) : \lambda_i = \sum_{\ell=0}^{p-1} w^\ell y_i^{t-\ell} \tag{2}$$

where $\{w^\ell > 0\}_{l=0}^{p-1}$ are the parameters of the model which are shared across locations i. For the time-depdendent dampening β_i^t we employ recurrent neural networks (RNNs; [5, 6, 14]) such that

$$\text{RNN} : \beta_i^t = f_\theta(\{x_i^k\}_{k=0}^t) \tag{3}$$

where θ are the parameters of the network which are again shared across locations and where $\{x_i^k\}_{k=0}^t$ denote observed input features to the RNN (e.g., mobility in location i at time k).

By combining a rigid AR model with a flexible RNN that can model the evolution of β_i^t, this decomposition is a first step towards separating time-independent and time-varying aspects of the disease process. However, limited data about the spread of COVID-19 makes it challenging to estimate the parameters of the model without overfitting — especially for high-capacity RNNs. We seek therefore an inductive bias which allows us to estimate β_i^t from few observations.

Relational Inductive Bias. Since all regions are affected by the same underlying process, we assume that we can borrow statistical strength between regions and use information about the spread in region i to help predict the spread in region j – once we have accounted for time- and location-dependent dynamics. A good model of β_i^t should therefore help to improve the predictions of y_i^{t+1}/β_i^t from cases in regions y_j^t where $i \neq j$. We interpret this as an inductive bias akin to Granger causality [13] and extend eq. (2) to a *vector autoregressive* (VAR) model. For VAR, it is known that Granger causality is directly linked to the coefficients of the model. In particular, let

$$\text{VAR}(p) : \lambda_i = \sum_{\ell=0}^{p-1} \sum_{j=1}^{m} w_{ij}^\ell y_j^{t-\ell} \tag{4}$$

be a vector autoregressive model of order p. *A time series y_j is then Granger-causing y_i if and only if $w_{ij} \neq 0$* [29]. For causal discovery, coefficients w_{ij} are therefore often regularized using ℓ_1-penalty terms to remove spurious relations. Here, we take the opposite approach and seek solutions in which many time-series can be considered to be Granger-causally related. This serves as an inductive bias to learn good models for the time-varying component β_i^t. However, we do not force all time series to be related since this is likely an unrealistic constraint. Instead, our goal is to define a model which allows us to specify a budget (or ratio) of time series that we assume to be related.

For this purpose, we propose a novel regularization scheme for VAR models in which the coefficients w_{ij} are drawn from a Logit-Normal distribution [2] for all $i \neq j$. This allows us to specify a prior on the proportion of related and unrelated time series as follows: Let $\phi(\cdot)$ denote the logistic function, let $\forall i \neq j : w_{ij} = \phi(\alpha_{ij})$, and let $\mathcal{N}(\mu, \sigma^2)$ denote the Normal distribution with mean μ and variance σ^2. For appropriate hyperparameter choices (i.e., high variance distributions where $\sigma > 3.2$) the Logit-Normal distribution exhibits samples close to 0 and 1. Consequently, this allows us to regularize models such that they learn adjacency-like matrices

$w_{ij} = \phi(\alpha_{ij})$ which either fully include or exclude Granger-causal relations between any region in the data. Furthermore, the number of non-zero entries in w_{ij}, i.e., the ratio of related and unrelated time series, can be controlled through the mean of the Logit-Normal prior (see also fig. 1c). Putting everything together, we model then the *time-varying* force of infection as

$$\beta\text{-AR}(p): \quad \lambda_i^{t+1} = \beta_i^t \sum_{\ell=0}^{p-1} \sum_{j=1}^{m} w_{ij}^{\ell} y_j^{t-\ell} \tag{5}$$
$$\alpha_{ij} \sim \mathcal{N}(\mu, \sigma^2) \quad \forall i \neq j$$

Likelihood and Parameter Estimation. Count data is naturally modeled using Poisson distributions. However, COVID-19 case counts exhibit substantial overdispersion, i.e., the variance of observed counts can significantly exceed their mean (see also fig. 1b). For this reason, we model case counts using the Negative Binomial distribution what allows us to account for varying degrees of overdispersion [22]. Specifically, we set

$$y_i^{t+1} \sim \text{NB}(\lambda_i^t, \nu_i)$$

where λ_i^t and ν_i are mean and dispersion parameter of the distribution and λ_i^t is modeled using eq. (5).

To estimate the parameters of the model, we regularize the log-likelihood of the model such that w_{ij} is drawn from a Logit-Normal distribution with location μ and scale σ. Let θ denote the model parameters (i.e., VAR parameters α_{ij}, dispersion parameters ν_i, as well as the parameters of the RNN) and let $p_\theta(y)$ denote the likelihood function of the β-AR model. Furthermore, let q denote the prior normal distribution for α_{ij}. To estimate the model parameters from observed time series $\mathcal{Y}$, a natural approach would be to maximize the regularized log-likelihood

$$\max_\theta \sum_y \log p_\theta(y) + \gamma \Omega(\theta)$$
$$\text{where } \Omega(\theta) = \sum_{ij} \log q(\alpha_{ij} \mid \mu, \sigma). \tag{6}$$

and regard $\mu, \sigma > 0, \gamma$ as hyperparameters which allow us to control the ratio of related time series and regulariztion strength. However, for high-variance Logit-Normal distributions (which are required to sample approximate $\{0, 1\}$ values), this approach is ineffective. This is because the regularization term is equivalent to

$$\log q(a_{ij} | \mu, \sigma) \approx \frac{1}{\sigma^2}(a_{ij} - \mu)^2 \tag{7}$$

such that it has little regularizing effect as $\sigma \to \infty$. Moreover, the moments of the Logit-Normal have no closed-form expression in terms of μ such that it is difficult for a practitioner to specify hyperparameters that lead to the intended ratio of related time series. For this reason, we use a different approach and directly regularize the sample mean of $\phi(a_{ij})$ by replacing Ω in eq. (6) with

$$\Omega(\theta) = \left(\frac{1}{m^2} \sum_{i,j} \phi(\alpha_{ij}) - \psi \right)^2 \tag{8}$$

where $\psi \in [0, 1]$ is a hyperparameter that specifies the desired ratio of related time series. It is straightforward to regularize the variance of the Logit-Normal distribution in a similar way (see suppl. material). However, we have found that it is not necessary to include such a term in the objective as even without explicit regularization the values end up concentrating near $\{0, 1\}$. This is because non-regularized α_{ij}s can be interpreted as samples from a uniform distribution which approximates well a Normal distribution with high variance.

Since eq. (6) is end-to-end differentiable we can jointly estimate the parameters of the entire model using gradient-based optimization. We compute gradients via automatic differentiation using the PyTorch framework [27]. To maximize eq. (6) we use the stochastic optimization method Adam [16]. However, we have found that the adaptive updates of Adam do not work well for the regularization term. For this reason we take an approach similar to AdamW [23] and compute non-adaptive gradient updates of the regularization term Ω. The gradient for these updates can be computed in closed form via

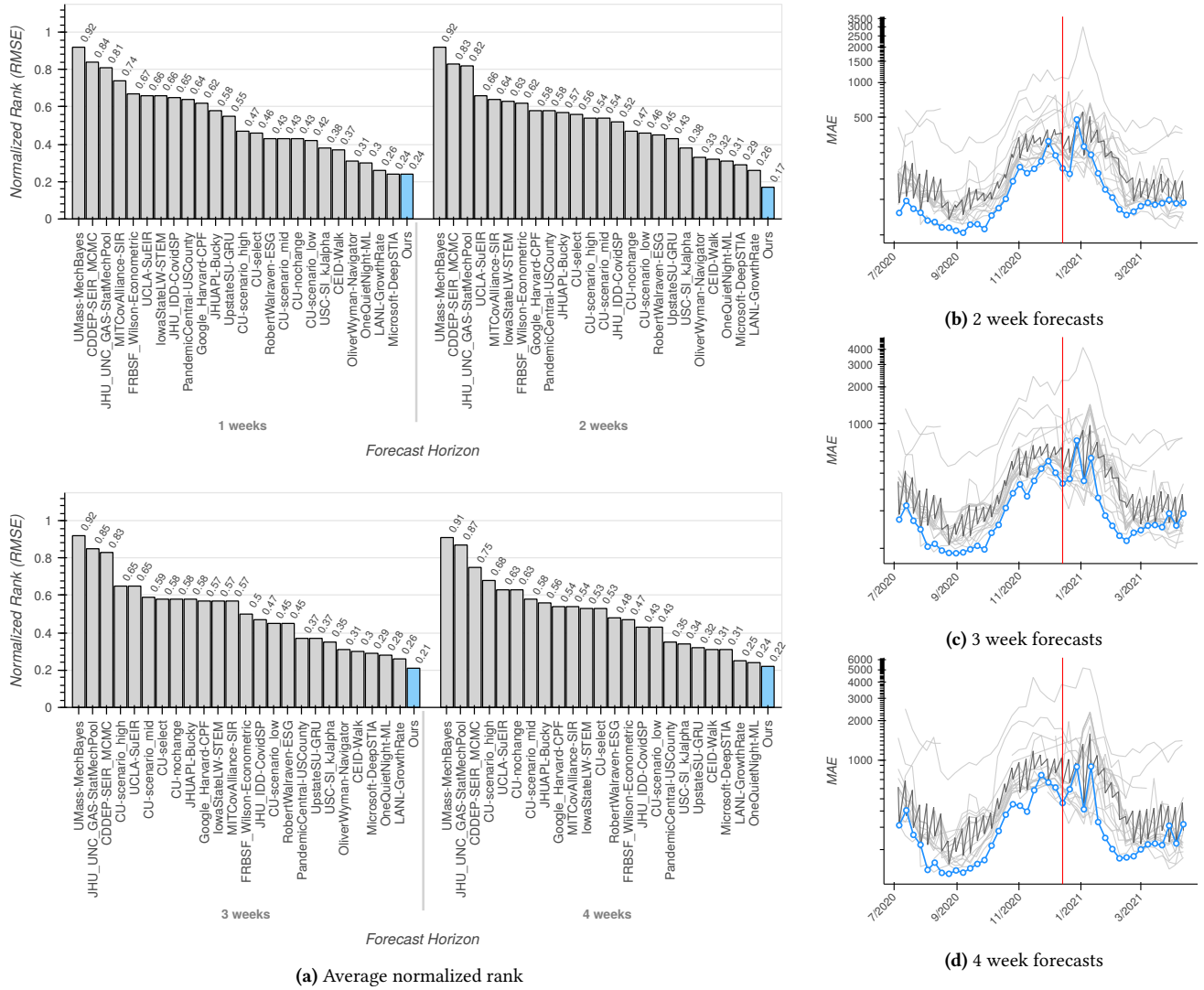
$$\frac{\partial}{\partial \alpha_{ij}} \Omega = \frac{2(\phi(\alpha_{ij}) - \psi) e^{\alpha_{ij}}}{m^2 (e^{\alpha_{ij}} + 1)^2} \tag{9}$$

3 RESULTS

We evaluate the quality of our method compared to multiple state-of-the-art forecasts of confirmed cases on county-level in the United States. For comparison, we collected all forecasts from non-ensemble models that have been submitted to the COVID-19 Forecast Hub (CFH). See tables 2 and 3 in the appendix for a full list of comparison models and datasets used to train our model. All datasets are publicly available, de-identified, and aggregated at county- or state-level. To compute forecasts using our model, we use the following fully automated model selection scheme: For each forecast date d, we perform cross-validation by holding out 41 days of validation data and train the model on the remaining data. We then select the best hyperparameters as measured by MAE on the held out validation set and retrain the whole model with those hyperparameters on the combined training and validation set to compute the final forecast. When computing the forecasts, we hold all additional input data (e.g., symptom survey, mobility, weather, etc.) constant after the last observed day d. For data that only exists at the state-level, we use the state-level value for each county.

Forecast quality. Comparing the quality of forecasts from the CFH in a robust way is non-trivial. Since teams submit their forecasts in varying intervals and on different dates, averaging error metrics such as MAE across dates is not meaningful. The difficulty of forecasting varies considerably depending on the specific day that a forecast has been submitted such that a date-averaged metric is easily dominated by the dates of submission. Instead, we compare the relative performance of each forecast to all other forecasts submitted on that *same day*. For this purpose, we compute the average normalized rank for each model, which is computed as follows. Let $r_d(m)$ denote the rank of model m among all submissions on day d when ranked by an error metric such as MAE. Furthermore, let R_d denote the total number of submissions on day d and let $\mathcal{D}_m$ denote the set of submissions for model m. We compute then the average normalized rank (ANR) for each model via

$$\text{ANR}(m) = \frac{1}{|\mathcal{D}_m|} \sum_{d \in \mathcal{D}_m} \frac{r_d(m) - 1}{R_d - 1} \quad \in [0, 1] \tag{10}$$



(a) Average normalized rank

(b) 2 week forecasts

(c) 3 week forecasts

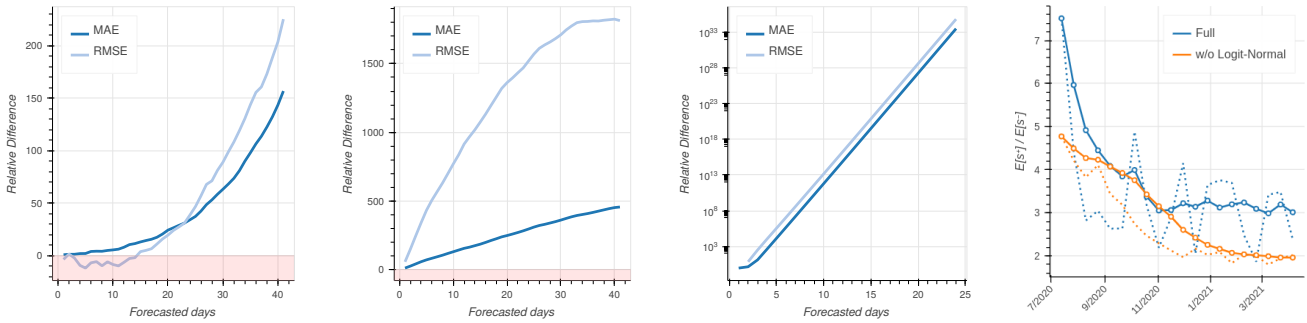
(d) 4 week forecasts

Figure 2: a) Averaged normalized rank based on RMSE for each model for different forecast horizons (lower is better). b-d) Average MAE for two to four weeks forecast horizons (epiweeks) for all models submitted to the COVID-19 Forecast Hub. β-AR model is indicated in blue, average MAE for all models is indicated in black. Red line indicates start of vaccinations in the US.

Table 1: Correlation of β-AR forecast errors with demographic properties of counties. Worst case correlations among all other forecasts from the COVID-19 Forecast Hub are reported in parantheses for comparison.

	Population Race/Ethnicity %					**Education %**	**Income %**
	Non white	Asian	Black	Latinx	Native American	Non college	Below Poverty
MAE	0.119 (0.23)	0.278 (0.363)	0.043 (0.15)	0.17 (0.355)	-0.027 (0.042)	-0.164 (-0.217)	-0.029 (0.08)
Mean Error	0.026 (-0.113)	0.002 (0.318)	0.018 (-0.061)	0.026 (0.211)	0.007 (-0.037)	0.038 (-0.166)	0.028 (0.065)

Figure 2a shows the forecast quality of our model (blue) when measured by ANR compared to all other COVID-19 Forecast hub models over the entire time period from July 2020 to April 2021. It can be seen that the proposed β-AR model shows strong performance and is the best performing model across all forecast horizons according to ANR. Figure 2b-d, which shows the MAE of all models



(a) No Logit-Normal **(b) Poisson Regression** **(c) Non-Relational** **(d) Social Connectedness**

Figure 3: a-c) Relative Improvement (MAE and RMSE) of the β-AR model compared to a models with no Logit-Normal regularization, Poisson regression instead of Negative Binomial, and a non-relational variant. d) Ratio of expected social connectedness of related and unrelated counties ($\mathbb{E}[s^+]/\mathbb{E}[s^-]$) in the learned β-AR models. Solid lines indicate smoothed ratio (7 day window), dotted lines indicate non-smoothed ratios.

per day, further illustrates this property. It can be seen that our model shows again strong performance and is generally the best or among the top forecasts across the entire evaluation period and for all forecast horizons. A time period where our forecasts are not among the best models is briefly at the end of January 2021 where its performance is equal to the average forecast submitted to the CFH. This coincides with the beginning of vaccinations in the US whose effect on the spread of the pandemic would start to materialize around that time. A possible conjecture of this observation is that our model required time to adjust to the change in disease dynamics that was introduced through vaccinations. However, as can be seen by the results in February and March, the model performed this adjustment quicky and was very soon among the top performing forecasts again.

Fairness. As forecasts can inform policy and resource allocation decisions during the pandemic, a key question is whether forecasts are similarly accurate across counties with different demographic characteristics in order to minimize the potential for unfair distribution of those resources. Such evaluation of machine learning models is especially important when they are used to inform real-world decisions. To analyze this aspect of our forecasts, we correlate the error of forecasts with demographic properties of counties related to income (e.g., percent of residents below poverty level), education (e.g., percent of non-college educated residents), and race and ethnicity (e.g., percent of non-white population) as published by the US Census Bureau's American Community Survey[1]. Table 1 shows the correlation of MAE with these quantities. We also report Mean Error ($y_i^t - \hat{y}_i^t$) as under-predicting cases is relevant in the context of resource allocations and would be masked by MAE. It can be seen that the error of our forecasts have low correlations with all demographic aspects listed here and that the low correlations for Mean Error do not indicate systematic under- or over-prediction. To compare with other available forecasts, we evaluated the same correlation on predictions that are available through the CFH. We found that the low correlation of error and demographic properties holds for most but not for all of them (see also table 1). Note that there may be other (unmeasured) demographic properties of counties for which the correlations may look different. It should also be noted that there may be confounding factors such as under-reporting issues in marginalized communities that would imply a bias in the reported case counts. As we are open-sourcing our model, we encourage every user to perform similar analyses for their specific use-cases and forecasting regions.

Ablations. In addition to comparisons against state-of-the-art forecasts, we also evaluate the contributions of different aspects of our model to gain insights into their relative importance. Specifically, we are interested in the contributions of the novel Logit-Normal regularization method, the Negative Binomial regression, and the relational approach in general. In all cases, we report the improvement in terms of the relative difference for MAE and RMSE, i.e., $\text{MAE}_{\text{Ablation}} - \text{MAE}_{\text{Full model}}$.

To evaluate the contributions of the Logit-Normal regularization, we trained a model where we fix the corresponding regularization parameter to 0 and compared its forecast quality to the standard model where the hyperparameter has been selected via cross-validation. fig. 3a shows the results of this comparison. It can be seen that the Logit-Normal regularization can be very beneficial to improve forecast quality. It provides substantial improvements for both MAE and RMSE and especially for longer forecast horizons. This supports our motivation that regularizing the β-AR model through a Granger-like approach improves the generalization properties of the RNN and stabilizes the forecasts.

To evaluate the contributions of the Negative Binomial distribution, we compared it to a model where cases are modeled using a standard Poisson distribution. It can be seen from fig. 3b, that the NB likelihood also improves the quality of the model substantially for all forecast horizons. This supports our motivation that the NB distribution can better account for the random variability in the observed data, while the rigid Poisson likelihood causes the (recurrent) model to overfit to these variations.

To evaluate the contributions of the relational approach in general, we trained additional models where we disabled the relational part by setting $\forall i \neq j : w_{ij} = 0$. It can be seen from fig. 3c that the full model offers again substantial improvements over the non-relational model as the forecast quality grows exponentially with

[1] https://www.census.gov/programs-surveys/acs

the forecasting horizon. While the non-relational model can offer acceptable forecasts for 1-2 days, it quickly deteriorates with larger horizons. This shows the importance of the relational component for disentangling the different growth factors and learning high quality models.

Adjacency Structure. In addition to these ablations, we analyze the structure of the learned adjacency matrix $\phi(\alpha_{ij})$. While we found some correlation with spatial distance, we found a stronger connection to the social connectedness of counties. To illustrate this property, we employ the Social Connectedness Index (SCI)[2] which provides for each pair of counties a score s_{ij} capturing the social connectedness of both counties according to their relative frequency of friendships on Facebook [4]. To evaluate whether our learned adjancency matrix captures this structure, we group counties as related if $\phi(\alpha_{ij}) > 0.5$ and not related otherwise (we also remark that $\phi(\alpha_{ij})$ are approximately bimodal at 0 and 1). Then we calculcate the expected social connectedness using values from SCI of related counties ($\mathbb{E}[s^+]$) and unrelated counties ($\mathbb{E}[s^-]$) according to our model. It can be seen from fig. 3d that the expected social connectedness of related counties is between 2-8 times higher than that of unrelated counties. Moreover, the Logit-Normal regularization leads to higher social connectedness ratios what can also explain its contributions to model quality.

4 CONCLUSION

To improve the quality and robustness of COVID-19 forecasts, we propose a new method which aims to disentangle time-varying and region-specific factors – such as demographics, policies, and mobility – from disease-inherent factors that influence its spread. For this purpose, we combine deep learning with VAR and train the joint model with a new regularization scheme that increases the coupling between regions. In our experiments, we observe that our method achieves strong performance in forecasting COVID-19 when compared to state-of-the-art models. Our method takes a highly data-driven approach with fewer modeling assumptions as, for instance, in mechanistic compartmental models. As such, we see our approach as complementary to existing models with focus on strong forecasting performance at the cost of reduced interpretability.

REFERENCES

[1] Sercan O. Arik, Chun-Liang Li, Jinsung Yoon, Rajarishi Sinha, Arkady Epshteyn, Long T. Le, Vikas Menon, Shashank Singh, Leyou Zhang, Nate Yoder, Martin Nikoltchev, Yash Sonthalia, Hootan Nakhost, Elli Kanal, and Tomas Pfister. 2021. Interpretable Sequence Learning for COVID-19 Forecasting. (2021). arXiv:2008.00646 [cs.LG]

[2] J Atchison and Sheng M Shen. 1980. Logistic-normal distributions: Some properties and uses. *Biometrika* 67, 2 (1980), 261–272.

[3] Jackie Baek, Vivek F. Farias, Andreea Georgescu, Retsef Levi, Tianyi Peng, Deeksha Sinha, Joshua Wilde, and Andrew Zheng. 2020. The Limits to Learning an SIR Process: Granular Forecasting for Covid-19. arXiv:2006.06373 [stat.ME]

[4] Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives* 32, 3 (2018), 259–80.

[5] Kyunghyun Cho, B van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. In *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8), 2014.*

[6] Jeffrey L Elman. 1990. Finding structure in time. *Cognitive science* 14, 2 (1990), 179–211.

[7] Facebook Data for Good. 2020. *Movement Range Maps.* https://dataforgood.fb.com/tools/movement-range-maps/

[8] Facebook Data for Good. 2020. *Symptom Survey.* https://dataforgood.fb.com/tools/symptommap/

[9] David C. Farrow, Logan C. Brooks, Aaron Rumack, Ryan J. Tibshirani, and Roni Rosenfeld. 2015. *Delphi Epidata API.* https://github.com/cmu-delphi/delphi-epidata

[10] Joseph Galasso and Duy Cao. [n.d.]. *PandemicCentral-USCounty.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/PandemicCentral-USCounty

[11] Zhifeng Gao, Chaozhuo Li, Wei Cao, Shun Zheng, Jiang Bian, Xing Xie, Tie-Yan Liu, and Juan Lavista Ferres. 2020. *Microsoft-DeepSTIA.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/Microsoft-DeepSTIA

[12] Google. 2020. *Community Mobility Reports.* https://www.google.com/covid19/mobility/

[13] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.

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

[15] Areum Jo and Jae Cho. [n.d.]. *OneQuietNight-ML.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/OneQuietNight-ML

[16] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1412.6980

[17] Matt Kinsey, Kate Tallaksen, R.F. Obrecht, Laura Asher, Cash Costello, Michael Kelbaugh, and Shelby Wilson. 2020. *JHUAPL-Bucky.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/JHUAPL-Bucky

[18] Eili Klein, Gary Lin, and Yupeng Yang. 2020. *CDDEP SEIR Markov Chain Monte Carlo.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/CDDEP-SEIR_MCMC

[19] Ugur Koyluoglu and John Milliken. 2020. *Oliver Wyman Pandemic Navigator.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/OliverWyman-Navigator/

[20] Joseph Chadi Lemaitre, Kyra H Grantz, Joshua Kaminsky, Hannah R Meredith, Shaun A Truelove, Stephen A Lauer, Lindsay T Keegan, Sam Shah, Josh Wills, Kathryn Kaminsky, Javier Perez-Saez, Justin Lessler, and Elizabeth C Lee. 2020. A scenario modeling pipeline for COVID-19 emergency planning. *medRxiv* (2020). https://doi.org/10.1101/2020.06.11.20127894

[21] Justin Lessler, Jess Edwards, Keya Joshi, Claire Smith, Paul Zivich, Marc Coram, Ellen Klein, Michael Brenner, Lusann Yang, Anton Geraschenko, Stephan Hoyer, Dmitri Kochkov, and Jamie Smith. 2020. *JHU_UNC_GAS.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/JHU_UNC_GAS-StatMechPool

[22] James O. Lloyd-Smith. 2007. Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly Overdispersed Data, with Applications to Infectious Diseases. *PLOS ONE* 2, 2 (02 2007), 1–8. https://doi.org/10.1371/journal.pone.0000180

[23] Ilya Loshchilov and Frank Hutter. 2018. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations.*

[24] Matthew J Menne, Imke Durre, Russell S Vose, Byron E Gleason, and Tamara G Houston. 2012. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology* 29, 7 (2012), 897–910.

[25] Eamon O'Dea. 2020. *CEID-Walk.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/CEID-Walk

[26] Dave Osthus, Sara Del Valle, Carrie Manore, Brian Weaver, Lauren Castro, Courtney Shelley, Manhong (Mandy) Smith, Julie Spencer, Geoffrey Fairchild, Travis Pitts, Dax Gerts, Lori Dauelsberg, Ashlynn Daughton, Morgan Gorris, Beth Hornbein, Daniel Israel, Nidhi Parikh, Deborah Shutt, and Amanda Ziemann. 2020. *LANL-GrowthRate.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/LANL-GrowthRate

[27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems.* 8026–8037.

[28] Sen Pei and Jeffrey Shaman. 2020. Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US. *medRxiv* (2020). https://doi.org/10.1101/2020.03.21.20040303

[29] Anil Seth. 2007. Granger causality. *Scholarpedia* 2, 7 (2007), 1667. https://doi.org/10.4249/scholarpedia.1667 revision #127333.

[30] Dan Sheldon, Graham Gibson, and Nick Reich. 2020. *UMass Mech-Bayes.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/UMass-MechBayes/

[2] https://dataforgood.fb.com/tools/social-connectedness-index/

[31] Ajitesh Srivastava, Tianjian Xu, and Viktor K. Prasanna. 2020. Fast and Accurate Forecasting of COVID-19 Deaths Using the SIkJa Model. arXiv:2007.05180 [q-bio.PE]
[32] The COVID Tracking Project. 2020. *State testing data (CC-BY 4.0 license).* https://covidtracking.com/
[33] The New York Times. 2020. *Coronavirus (Covid-19) Data in the United States.* https://github.com/nytimes/covid-19-data
[34] Robert Walraven. [n.d.]. *RobertWalraven-ESG.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/RobertWalraven-ESG
[35] Li Wang, Guannan Wang, Lei Gao, Xinyi Li, Shan Yu, Myungjin Kim, Yueying Wang, and Zhiling Gu. 2020. Spatiotemporal Dynamics, Nowcasting and Forecasting of COVID-19 in the United States. arXiv:2004.14103 [stat.AP]
[36] Daniel J. Wilson. 2021. Weather, Mobility, and COVID-19: A Panel Local Projections Estimator for Understanding and Forecasting Infectious Disease Spread. (2021). https://doi.org/10.24148/wp2020-23
[37] Yanli Zhang-James, Asif Salekin, Jonathan Hess, Samuel Chen, Dongliang Wang, Christopher P. Morley, and Stephen V. Faraone. [n.d.]. *UpstateSU-GRU.* https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/UpstateSU-GRU
[38] Difan Zou, Lingxiao Wang, Pan Xu, Jinghui Chen, Weitong Zhang, and Quanquan Gu. 2020. Epidemic Model Guided Machine Learning for COVID-19 Forecasts in the United States. *medRxiv* (2020). https://doi.org/10.1101/2020.05.24.20111989

A ADDITIONAL EXPERIMENTAL DETAILS

Table 2: Data sources for β-AR.

Dataset	Source	Resolution
Confirmed Cases	The New York Times [33] *Confirmed cases based on reports from state & local health agencies*	County
Symptom Survey	CMU COVIDcast [9] Facebook Data for Good [8] *Prevalence of COVID-like symptoms, mask-wearing, and vaccinations from self-reported surveys*	County, State
Movement Range Maps	Facebook Data for Good [7] *Mobility metrics related to physical distancing measures (change in movement and staying put)*	County, State
Community Mobility	Google [12] *Movement trends across different categories of places (retail and recreation, groceries and pharmacies, etc.)*	County, State
Doctor visits	CMU COVIDcast [9] Percentage of COVID-related doctor's visits in a given location	County, State
Testing	The COVID Tracking Project [32] *Total number of COVID PCR tests per state*	State
Weather	NOAA GHCN [24] *Average, minimum, maximum temperature & rainfall per county*	County

Table 3: Forecasting models for confirmed cases on county-level from the COVID-19 Forecast Hub. (https://github.com/reichlab/covid19-forecast-hub)

Group	Model	
Center for Disease Dynamics, Economics & Policy	*CDDEP-SEIR_MCMC*	[18]
Columbia University	*CU-**	[28]
COVID Alliance at MIT	*MITCovAlliance-SIR*	[3]
Federal Reserve Bank of San Francisco/Wilson	*FRBSF_Wilson-Econometric*	[36]
Google and Harvard University	*Google_Harvard-CPF*	[1]
Iowa State University Lily Wang Research Group	*IowaStateLW-STEM*	[35]
Johns Hopkins University and University of North Carolina	*JHU_UNC_GAS-StatMechPool*	[21]
Johns Hopkins University Applied Physics Lab	*JHUAPL-Bucky*	[17]
Johns Hopkins ID Dynamics COVID-19 Working Group	*JHU-IDD_CovidSP*	[20]
Los Alamos National Labs	*LANL-GrowthRate*	[26]
Microsoft	*Microsoft-DeepSTIA*	[11]
Oliver Wyman	*OliverWyman-Navigator*	[19]
One Quiet Night	*OneQuietNight-ML*	[15]
Pandemic Central	PandemicCentral-USCounty	[10]
Robert Walraven	*RobertWalraven-ESG*	[34]
SUNY Upstate and SU Covid-19 Prediction Team	*UpstateSU-GRU*	[37]
UCLA Statistical Machine Learning Lab	*UCLA-SuEIR*	[38]
University of Georgia	*CEID-Walk*	[25]
University of Massachusetts Amherst	*UMass-MechBayes*	[30]
University of Southern California Data Science Lab	*USC-SI_kJalpha*	[31]