Neural Relational Autoregression for High-Resolution COVID-19 Forecasting

Anonymous Authors[1]

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 region-specific factors – such as demographics, enacted policies, and mobility – from disease-inherent factors that influence its spread. For this purpose, we combine recurrent neural networks with a vector autoregressive model and train the joint model with a specific 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 by borrowing statistical strength across regions. In our experiments, we observe that our method achieves strong performance in predicting the spread of COVID-19 when compared to state-of-the-art forecasts.

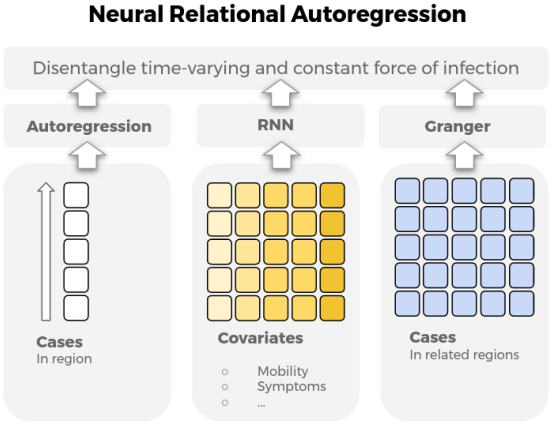

Figure 1. Ranking of county-level forecasts by average MAE over various forecast horizons. The proposed neural relational autoregressive model (β-AR) shows strong performance over all horizons when compared to state-of-the-art forecasts. Mean rank over all horizons in parentheses.

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) has become an important task in the public health response to the disease. For instance, accurate county-level forecasts 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; and to make progress towards efficient early detection systems.

However, forecasting COVID-19 poses unique challenges – in particular when considering confirmed cases at high spatial resolution. Although there has been considerable progress towards understanding the spread of the disease, there still exists only limited data and knowledge about important factors that influence its spread. This is only exacerbated by the naturally larger noise-levels in county-level data as compared to more highly aggregated state-level data. Due to the global nature of COVID-19, the available data is also distributed among regions with very different properties, many of which may affect its spread. This includes, for instance, demographics and population densities, enacted policies, adherence to those policies, mobility patterns, and geographic features such as temperature. In addition, testing and reporting can vary considerably across regions and time. All these factors lead to considerable variablity (see also **??**) and uncertainty in the data and makes reliable forecasting at high spatial resolution difficult.

To alleviate these issues, we propose a new method for predicting the spread of COVID-19 by combining recurrent neural networks with a vector autoregressive model and a specific regularization relational scheme. Our approach is motivated by two main aspects: First, we seek to develop an end-to-end differentiable model, as this allows us to make efficient use of the limited available data while also

[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.

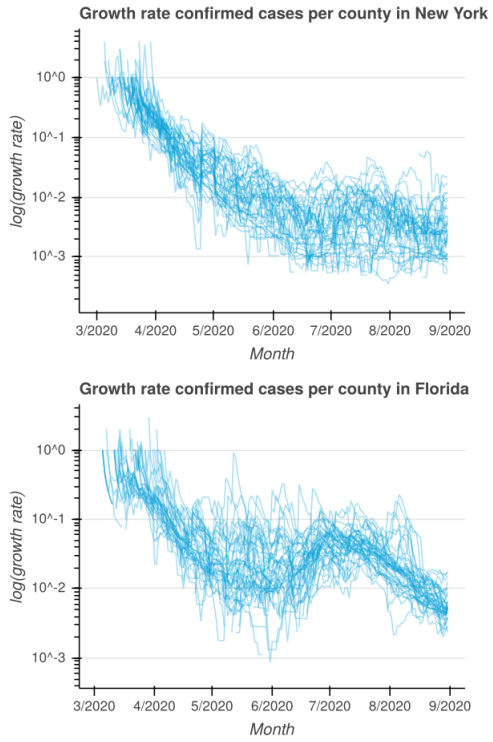

Figure 2. Variability of growth in confirmed cases per region over time. Each line represents one county in the state. For y_t denoting the number of cases at time t, growth rate is computed as $(y_{t+1} - y_t)/y_t$.

enabling us to estimate parameters of powerful models that can capture the large variability of cases across locations and time. However, while such flexible models are needed to account for possible influencing factors there is 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 – based on the assumption that once a model has correctly accounted for region-specific dynamics, information about the spread of the disease in region j can also help to improve predictions for a related region i. This approach is akin to using *Granger causality* as an inductive bias to improve forecast quality and robustness.

Compared to existing state-of-the-art forecasting models, our method takes a highly data-driven approach with fewer modeling assumptions as, for instance, in very detailed compartmental models. As such, we see our approach as complementary to existing models which provides strong forecasting performance at the cost of reduced interpretability.

2. Neural Relational Autoregression

We consider the forecasting of m time series that are different realizations of the same underlying disease process. Let $\mathcal{Y} = \{(y_i^1, \dots, y_i^T)\}_{i=1}^m$ denote the 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 the set of all observed case counts up to time $\tau \leq T$. We then model the case counts 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*[1] at time t in location i and where $f(x)$ denotes a probability distribution with parameter x (e.g., a Poisson or Negative Binomial distribution).

Due to the different interventions during the course of the epidemic, we regard $\mathcal{Y}$ as a time-varying process that is influenced by external factors such as policies, mobility, etc. 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

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

Hence, β_i^t can be understood as a dampening factor of the underlying force of infection which models the effect of interventions 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,[2] i.e.,

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

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; (**???**)) such that

$$\text{RNN} : \beta_i^t = f_\theta(\{x_i^k\}_{k=0}^t) \tag{2}$$

[1] Given y_i^t infected individuals, the force of infection (or hazard) models the probability that a susceptible individual at time t will become infected by time $t + 1$

[2] AR models where

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

can be interpreted as approximations of Reed-Frost chain binomial SIR models (**?**). For a detailed discussion see (**?**).

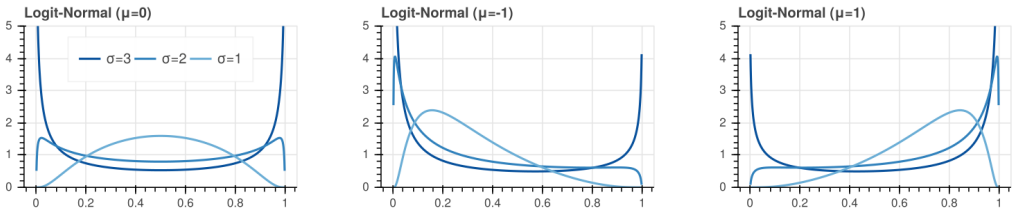

Figure 3. The Logit-Normal distribution is a probability distribution of a random variable whose logit has a normal distribution, i.e., $\phi(\mathcal{N}(\mu, \sigma))$.

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). Although an RNN as in **??** has enough capacity to model the evolution of β_i^t, the limited data about the spread of COVID-19 makes it challenging to estimate its parameters without overfitting. We seek therefore an inductive bias which allows us to estimate β_i^t from few observations.

2.1. 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 predicting 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 other regions y_j^t. We interpret this as an inductive bias akin to Granger causality (**?**)[3] and extend **??** to a *vector autoregressive* model where it is known that Granger causality is directly linked to its coefficients. In particular, let

$$\text{VAR}(p) : \lambda_i = \sum_{\ell=0}^{p-1} \sum_{j=1}^{m} w_{ij}^{\ell} y_j^{t-\ell} \tag{3}$$

be a vector autoregressive model of order p. *A time series y_j is then Granger-causing y_i if and only if $w_{ij} \neq 0$* (**?**). For causal discovery, coefficients w_{ij} are therefore often ℓ_1-regularized. Here, we take the opposite approach and seek solutions in which many time-series can be considered Granger-causal related. However, we do not force all time series to be related since this is likely an unrealistic constraint. Instead, we assume $\forall i \neq j : w_{ij}$ are drawn from a logit-normal distribution (**?**), what allows us to specify a prior on the proportion of related and unrelated time series.

In particular, 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. Putting everything together, we then model the full *time-varying* force of infection as

$$\beta\text{-AR}(p) : \quad \lambda_i^{t+1} = \beta_i^t \sum_{\ell=0}^{p-1} \sum_{j=1}^{m} w_{ij}^{\ell} y_j^{t-\ell} \tag{4}$$
$$\alpha_{ij} \sim \mathcal{N}(\mu, \sigma^2) \quad \forall i \neq j$$

Hence, the β-AR model consists of a standard AR component ($w_{ii} > 0$) and a relational component ($w_{ij} \in [0, 1]$) which aims to couple the different regions. The number of non-zero entries in the "adjacency matrix" w_{ij} can then be controlled through the logit-normal prior.

2.2. Accounting for Overdispersion

Count data such as confirmed cases is naturally modeled using Poisson distributions. However, COVID-19 case counts exhibit substantial overdispersion, i.e., the variance of the observed counts can significantly exceed their mean (e.g., see **??**). For this reason, we will model case counts with Negative Binomial distributions what allows us to account for varying degrees of overdispersion (**?**). 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 the β-AR model of **??**. The likelihood function in **??** is then of the form

$$p_\theta(y) = \frac{\Gamma(y+\nu)}{y!\Gamma(\nu)} \left(\frac{\mu}{\mu+\nu}\right)^y \left(1+\frac{\mu}{\nu}\right)^{-\nu} \qquad \mu > 0, \nu > 0$$

2.3. Parameter Estimation and Implementation Details

To estimate the parameters of the model, we regularize the model log-likelihood such that w_{ij} is drawn from a logit-normal distribution with location μ and scale σ. Let θ denote

[3] Granger causality is defined as follows: Let $X^t = \{X_t\}_{i=1}^t$, $Y^t = \{Y_t\}_{i=1}^t$, $Z^t = \{Z_t\}_{i=1}^t$ denote stochastic processes and let L denote a loss function. Furthermore, let

$$\mathcal{R}(Y^{t+1}|Y^t, Z^t) = \mathbb{E}(L(Y_{t+1}, f(Y^t, Z^t)))$$

denote the expected loss (risk) of a predictor f. We then say X *Granger-causes* Y if its inclusion in the predictor significantly improves the forecast, i.e., if

$$\mathcal{R}(Y^{t+1}|Y^t, X^t, Z^t) \ll \mathcal{R}(Y^{t+1}|Y^t, Z^t)$$

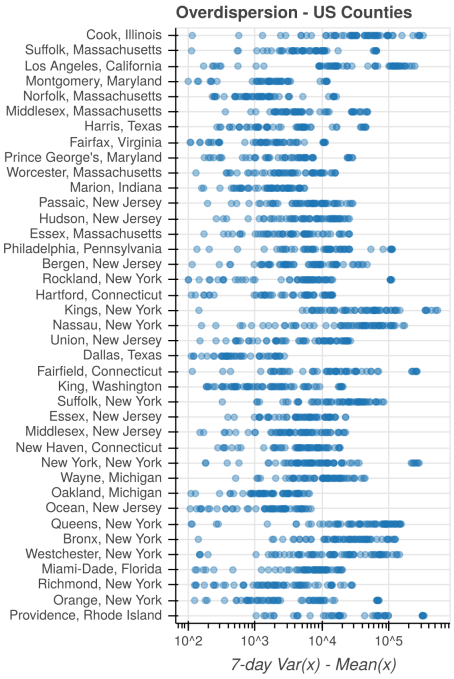

Figure 4. Overdispersion of daily case counts in US states and counties with most number of cases.

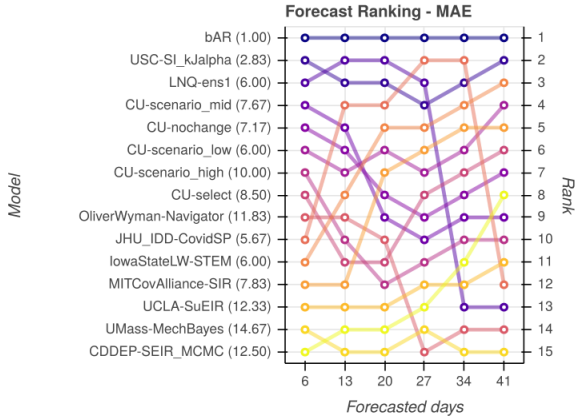

Figure 5. Ranking of county-level forecasts by average MAE over various forecast horizons. The proposed neural relational autoregressive model (β-AR) shows strong performance over all horizons when compared to state-of-the-art forecasts. Mean rank over all horizons in parentheses.

the model parameters (i.e., α_{ij} as well as 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}. We then maximize the regularized log-likelihood

$$\max_\theta \sum_y \log p_\theta(y) + \sum_{ij} \log q(\alpha_{ij} \mid \mu, \sigma). \qquad (5)$$

We regard $\mu, \sigma > 0$ as hyperparameters which allow us to control the ratio of related and unrelated time series.

Since **??** 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 (**?**). To maximize **??** we then use the stochastic optimization method AdamW (**?**) where we decouple the updates of the normally distributed parameters α_{ij} from the adaptive updates of the remaining parameters.

3. Results

In the following, we evaluate the forecast quality of our method compared to multiple state-of-the-art forecasts for confirmed cases on county-level. All comparison forecasts are collected from the COVID-19 Forecast Hub[4] as submitted by the respective teams. The COVID-19 Forecast Hub features county-level forecasts from July 5th onwards and we selected those models for which at least 10 forecasts where available since then. The full list of comparison forecasts is shown in **??**.

Forecast setup and model selection To compute forecasts for the different dates in the test set, we use the following fully automated model selection scheme: For each forecast date d, we perform cross-validation by holding out additional 21 days of validation data and train the model on the remaining data. We then select the best hyperparameters as measured by RMSE on the 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.[5]. For all training details of the model, please see the supplementary material.

Input data As input features for β-AR, we use multiple data sources as listed in **??**. Confirmed cases enter the model only in the autoregressive part. All other covariates enter the model only as input features for the time-varying β-part. For cases and weather data, we use the preprocessed data from the Google COVID-19 Open Data repository (**?**). All datasets are publicly available, de-identified, and aggregated at county- or state-level.

[4] https://github.com/reichlab/covid19-forecast-hub

[5] This setting places natural limits on the duration of the forecasting horizon. We reserve the joint forecasting of cases and covariates – what could extend the horizon – for future work.

Table 1. Forecasting models for confirmed cases on county-level.

Group	Model	
Center for Disease Dynamics, Economics & Policy	*CDDP-SEIR_MCMC*	(**?**)
Columbia University	*CU-**	(**?**)
COVID Alliance at MIT	*MITCovAlliance-SIR*	(**?**)
Iowa State University Lily Wang Research Group	*IowaStateLW-STEM*	(**?**)
Johns Hopkins ID Dynamics COVID-19 Working Group	*JHU-IDD_CovidSP*	(**?**)
LockNQuay	*LNQ-ens1*	(**?**)
Oliver Wyman	*Pandemic Navigator*	(**?**)
UCLA Statistical Machine Learning Lab	*UCLA-SuEIR*	(**?**)
University of Southern California Data Science Lab	*USC-SI_kJalpha*	(**?**)
University of Massachussets Amherst	*UMass-MechBayes*	(**?**)

Table 2. Data sources for β-AR.

Dataset	Source	Resolution
Confirmed Cases	**?** *Confirmed cases based on reports from state & local health agencies*	County
Symptom Survey	CMU COVIDcast (**?**) **?** *Prevalence of COVID-like symptoms from self-reported surveys*	County, State
Movement Range Maps	**?** *Mobility metrics related to physical distancing measures (change in movement and staying put)*	County, State
Community Mobility	**?** *Movement trends across different categories of places (retail and recreation, groceries and pharmacies, etc.)*	County, State
Doctor visits	CMU COVIDcast (**?**) Percentage of COVID-related doctor's visits in a given location	County, State
Testing	**?** *Total number of COVID PCR tests per state*	State
Weather	NOAA GHCN (**?**) *Average, minimum, maximum temperature & rainfall per county*	County

Forecast evaluation **??** shows the forecast quality as measured by MAE for multiple forecast horizons.[6] It can be seen that the proposed β-AR models shows a consistently strong performance and is for all forecasting dates and horizons either the best model or among the best. **??**, which shows the ranking of all models by the average MAE for each forecast horizon, further illustrates this property. It can be seen that β-AR model is consistently ranked first over all horizons. Furthermore, other models show much larger variability in their performance.

To also evaluate the performance of our model on days prior to July 5th, we compare to forecasts of Google Cloud AI (**?**) and Columbia University (**?**) which provide county-level forecasts of confirmed cases from May 11th to June 27th. **??** shows the average MAE over all counties for 7 and 14 day forecasts for these models.[7] It can be seen that the β-AR model shows again consistently strong performance on these earlier days and is typically ranked first for both 7 and 14 day forecasts.

Ablations In addition to comparisons to state-of-the-art county-level forecasts, we also evaluate the contributions of different aspects of our model. First, we test the effect of the relational autoregressive part. For this purpose, we trained additional models were we disabled the relational part (by setting $\forall i \neq j : w_{ij} = 0$) and compared their forecasts to the full model of **??**. To measure the relative improvement of the full model over the non-relational model, we compute then the relative error of both models, e.g.,

$$\text{Relative Mean Absolute Error} = \frac{\text{MAE}_{\text{full}}}{\text{MAE}_{\text{non-relational}}}$$

[6] MAE numbers are computed in accordance with https://github.com/youyanggu/covid19-forecast-hub-evaluation

[7] For this comparison, average MAE is computed as described in (**?**)

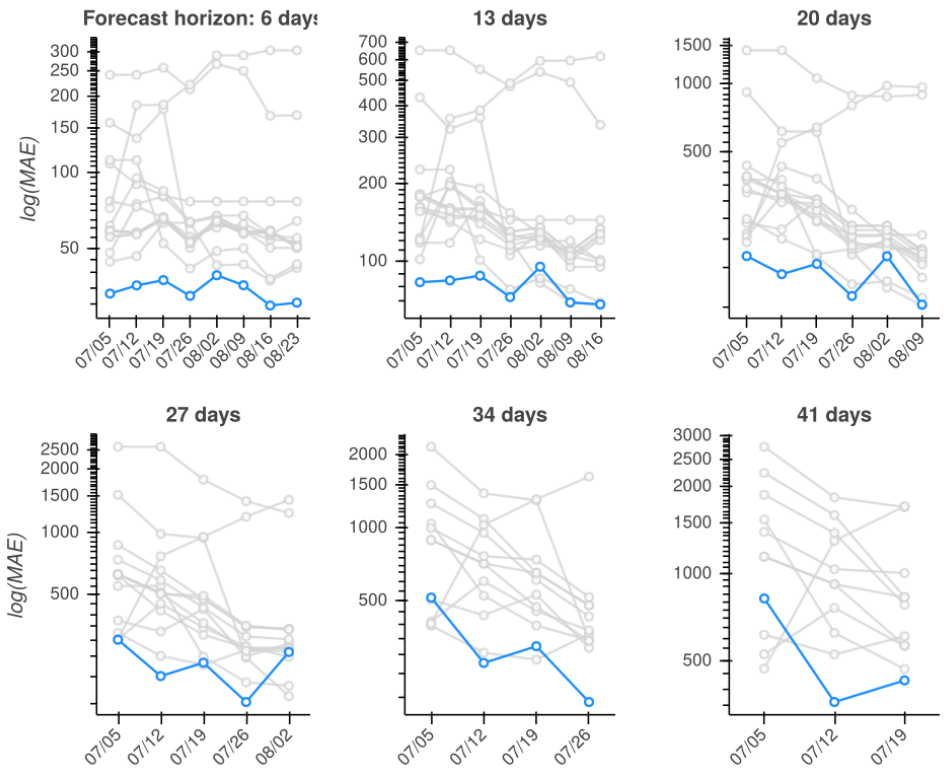

Figure 6. Comparison of β-AR model (blue) to 15 county-level models from COVID-19 forecast hub (gray). Forecast quality is measured in MAE (log-scale) where the absolute errors are averaged over all counties. For similar analysis using RMSE please see the supplementary material.

It can be seen from **??** that full model offers substantial improvements over the non-relational model as the relative forecast quality grows exponentially with the forecasting horizon. While the non-relational model can offer acceptable forecast for horizons of 1-2 days, it quickly deteriorates with larger horizons. This show the importance of the relational component for disentangling the different growth factors and learning high quality models.

In addition to the non-relational component, we also evaluated the contributions of the logit-normal regularization method. For this purpose, we trained a model where we explicitly set the requlariation parameter $\sigma = 0$. We then compare the forecast quality to the standard model where the regularization parameter has been selected via cross-validation. **??** shows the results of the comparison. It can be seen that the logit-normal regularization can be very beneficial to improve forecast quality. While the differences to the standard model are much smaller than for the non-relational model, the addition of the regularization term can lead to substantial improvements, especially for horizons of 13 days and longer.

Finally, we also evaluated the contributions of the Negative Binomial distribution compared to a standard Poisson distribution for modeling confirmed cases. Similar to the logit-normalization method, we trained an additional model with Poisson likelihood and compared the forecast quality to the standard model. It can be seen from **??**, that Negative Binomial likelihood significantly improves the quality of the model over all forecast horizons. This is likely due to the fact that the Negative Binomial can better model the noise in the observed data, while the stricter Poisson likelihood causes the (recurrent) model to overfit to these variations.

4. Related Work

We build on prior work that has proposed to use autoregressive models for spatially and temporally aggregated disease surveillance data of endemic-epidemic processes (**???**). Such autoregressive models are, for instance, used to monitor infectious diseases by public health agencies like the Robert Koch Institute (**?**).

Moreover, the negative binomial distribution has become a popular way to model infectious diseases, largely to its ability to model count data with varying degrees of overdispersion (**?**). Autoregressive models in combination with negative binomial distributions have, for instance, been used

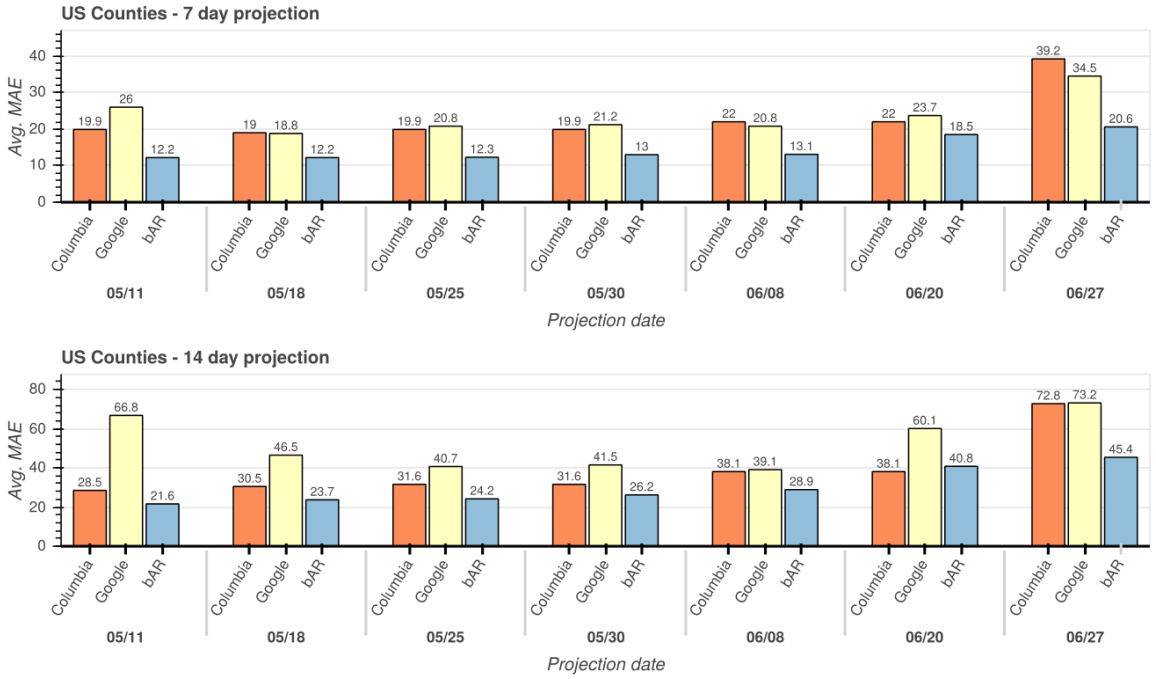

Figure 7. Comparions of β-AR model to forecasts from Google Cloud AI and Columbia for 7 and 14 day horizons and earlier forecast dates. Forecast quality is measured in MAE where the absolute errors are averaged over all counties.

by **???** to model infectious disease count data.

? proposed a combination of VAR(1) models and ℓ_1 regularization to for the discovery of Granger-causal relations to understand brain connectivity. **?** proposed an improved estimator which can be applied for VAR models of order $p > 1$.

5. Conclusion

To improve the quality and robustness of forecasts, we propose a new method which aims to disentangle region-specific factors – such as demographics, enacted policies, and mobility – from disease-inherent factors that influence its spread. For this purpose, we combine recurrent neural networks with a vector autoregressive model and train the joint model with a specific regularization scheme that increases the coupling between regions. In our experiments, we observe that our method achieves strong performance in predicting the spread of COVID-19 when compared to state-of-the-art forecasts. Through ablations of the model, we show that the relational approach in general, the added logit-normal regularization, and the negative binomial likelihood are all important factors that contribute to the forecast quality. Compared to existing state-of-the-art forecasting models, our method takes a highly data-driven approach with fewer modeling assumptions as, for instance, in very detailed and 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

Abbey, H. An examination of the reed-frost theory of epidemics. *Human biology*, 24 3:201–33, 1952.

Arik, S. O., Li, C.-L., Yoon, J., Sinha, R., Epshteyn, A., Le, L. T., Menon, V., Singh, S., Zhang, L., Yoder, N., Nikoltchev, M., Sonthalia, Y., Nakhost, H., Kanal, E., and Pfister, T. Interpretable sequence learning for covid-19 forecasting, 2020.

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

Baek, J., Farias, V. F., Georgescu, A., Levi, R., Peng, T., Sinha, D., Wilde, J., and Zheng, A. The limits to learning an sir process: Granular forecasting for covid-19, 2020.

Bauer, C. and Wakefield, J. Stratified space–time infectious disease modelling, with an application to hand, foot and mouth disease in china. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(5):1379–1398, 2018.

Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y.

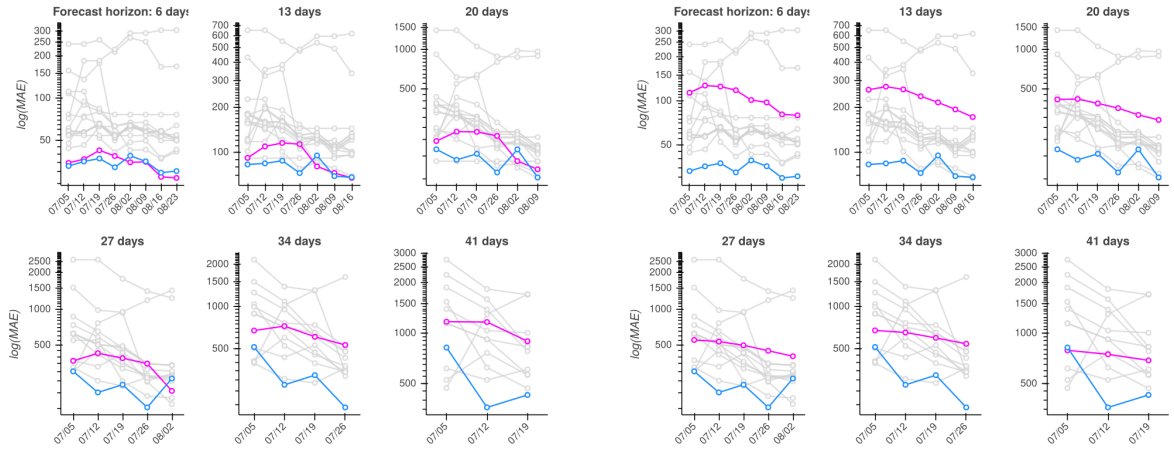

Figure 8. Comparison of β-AR model with (blue) and without (magenta) Granger regularization. Forecast quality is measured in MAE.

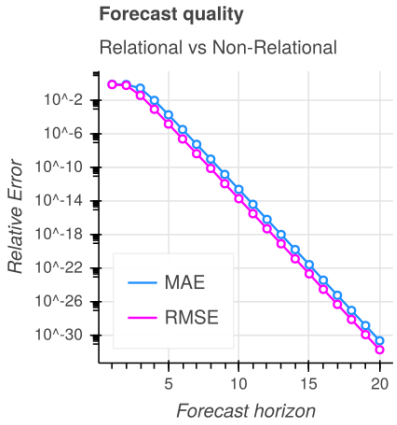

Figure 9. Relative Error (MAE and RMSE) of the fully relational β-AR model compared to a non-relational variant.

Figure 10. Comparison of β-AR model with Negative Binomial (blue) and Poisson (magenta) likelihood. Forecast quality is measured in MAE.

On the properties of neural machine translation: Encoder-decoder approaches. In *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8), 2014*, 2014.

Elman, J. L. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

Facebook Data for Good. Movement range maps, 2020a. URL https://dataforgood.fb.com/tools/movement-range-maps/.

Facebook Data for Good. Symptom survey, 2020b. URL https://dataforgood.fb.com/tools/symptommap/.

Farrow, D. C., Brooks, L. C., Rumack, A., Tibshirani, R. J., and Rosenfeld, R. Delphi epidata api, 2015. URL https://github.com/cmu-delphi/delphi-epidata.

Google. Community mobility reports, 2020. URL https://www.google.com/covid19/mobility/.

Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

Haufe, S., Müller, K.-R., Nolte, G., and Krämer, N. Sparse causal discovery in multivariate time series. In *Causality: Objectives and Assessment*, pp. 97–106, 2010.

Held, L., Höhle, M., and Hofmann, M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical modelling*, 5(3):187–199, 2005.

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

Klein, E., Lin, G., and Yang, Y. CDDEP SEIR Markov Chain Monte Carlo, 2020. URL https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/CDDEP-SEIR_MCMC.

Koyluoglu, U. and Milliken, J. Oliver wyman pandemic navigator, 2020. URL https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/OliverWyman-Navigator/.

Lemaitre, J. C., Grantz, K. H., Kaminsky, J., Meredith, H. R., Truelove, S. A., Lauer, S. A., Keegan, L. T., Shah, S., Wills, J., Kaminsky, K., Perez-Saez,

Table 3. COVID-19 Forecast Hub MAE (1 week horizon).

basedate	CU-select	β-AR	Google_Harvard-CPF	LANL-GrowthRate	Microsoft-DeepSTIA
2020-08-06	70.681	58.610	-	90.897	-
2020-08-13	61.309	53.553	-	74.575	-
2020-09-03	51.492	43.194	-	67.363	-
2020-09-27	43.166	36.189	42.139	-	-
2020-10-01	-	59.339	-	101.979	-
2020-10-11	37.986	45.747	76.775	-	-
2020-10-15	91.446	80.918	-	141.719	-
2020-10-25	52.614	64.412	52.371	-	-
2020-11-08	90.748	125.892	121.305	-	-
2020-11-15	98.254	113.979	118.928	-	-
2020-11-19	181.705	133.064	-	180.716	-
2020-12-06	123.796	152.413	117.942	-	-
2020-12-07	-	134.868	-	105.553	-
2020-12-13	105.273	130.788	136.166	-	-
2020-12-14	-	126.242	-	109.234	116.788
2020-12-17	184.772	179.296	-	168.666	-
2020-12-21	-	76.497	-	138.659	123.644
2021-01-03	136.050	174.788	259.101	-	-
2021-01-04	-	156.656	-	108.324	74.025
2021-01-11	-	85.915	190.840	110.568	159.404
2021-01-18	-	74.570	207.132	111.241	134.228
2021-01-25	-	66.138	92.175	66.162	79.503

J., Lessler, J., and Lee, E. C. A scenario modeling pipeline for covid-19 emergency planning. *medRxiv*, 2020. doi: 10.1101/2020.06.11.20127894. URL https://www.medrxiv.org/content/early/2020/06/12/2020.06.11.20127894.

Lloyd-Smith, J. O. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLOS ONE*, 2(2):1–8, 02 2007. doi: 10.1371/journal.pone.0000180. URL https://doi.org/10.1371/journal.pone.0000180.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2018.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910, 2012.

Meyer, S. and Held, L. Power-law models for infectious disease spread. *Ann. Appl. Stat.*, 8(3):1612–1639, 09 2014. doi: 10.1214/14-AOAS743. URL https://doi.org/10.1214/14-AOAS743.

Meyer, S. and Held, L. Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18(2):338–351, 12 2016. ISSN 1465-4644. doi: 10.1093/biostatistics/kxw051. URL https://doi.org/10.1093/biostatistics/kxw051.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pp. 8026–8037, 2019.

Pei, S. and Shaman, J. Initial simulation of sars-cov2 spread and intervention effects in the continental us. *medRxiv*, 2020. doi: 10.1101/2020.03.21.20040303. URL https://www.medrxiv.org/content/early/2020/03/27/2020.03.21.20040303.

Salmon, M., Schumacher, D., and Höhle, M. Monitoring count time series in r: Aberration detection in public health surveillance. *Journal of Statistical Software, Articles*, 70(10):1–35, 2016. ISSN 1548-7660. doi: 10.18637/jss.v070.i10. URL https://www.jstatsoft.org/v070/i10.

Seth, A. Granger causality. *Scholarpedia*, 2(7):1667, 2007. doi: 10.4249/scholarpedia.1667. revision #127333.

Sheldon, D., Gibson, G., and Reich, N. Umass mechbayes, 2020. URL https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/UMass-MechBayes/.

Srivastava, A., Xu, T., and Prasanna, V. K. Fast and accurate forecasting of covid-19 deaths using the sikja model, 2020.

Table 4. COVID-19 Forecast Hub MAE (2 week horizon).

basedate	CU-select	β-AR	Google_Harvard-CPF	LANL-GrowthRate	Microsoft-DeepSTIA
2020-08-06	163.989	125.481	-	196.062	-
2020-08-13	144.941	114.155	-	172.747	-
2020-09-03	127.597	123.920	-	165.917	-
2020-09-27	127.281	96.989	135.297	-	-
2020-10-01	-	174.502	-	236.432	-
2020-10-11	157.076	141.673	200.563	-	-
2020-10-15	248.751	221.959	-	336.129	-
2020-10-25	267.457	228.184	309.802	-	-
2020-11-08	401.312	374.674	511.953	-	-
2020-11-15	355.734	302.691	444.156	-	-
2020-11-19	558.966	391.979	-	490.497	-
2020-12-06	488.648	441.784	549.195	-	-
2020-12-07	-	379.099	-	447.404	-
2020-12-13	414.71	321.997	386.938	-	-
2020-12-14	-	313.182	-	341.375	311.987
2020-12-17	619.721	500.558	-	557.816	-
2020-12-21	-	251.703	-	344.751	301.977
2021-01-03	473.132	537.502	530.039	-	-
2021-01-04	-	449.326	-	559.572	576.666
2021-01-11	-	227.165	201.820	325.607	276.098
2021-01-18	-	147.844	149.761	259.540	134.744

The COVID Tracking Project. State testing data (cc-by 4.0 license), 2020. URL https://covidtracking.com/.

The New York Times. Coronavirus (Covid-19) Data in the United States, 2020. URL https://github.com/nytimes/covid-19-data.

Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., and Canales-Rodríguez, E. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):969–981, 2005.

Wahltinez, O., Lee, M., Erlinger, A., Daswani, M., Yawalkar, P., Murphy, K., and Brenner, M. Covid-19 open-data: curating a fine-grained, global-scale data repository for sars-cov-2. 2020. URL https://github.com/GoogleCloudPlatform/covid-19-open-data. Work in progress.

Wakefield, J., Dong, T. Q., and Minin, V. N. Spatio-temporal analysis of surveillance data. *Handbook of Infectious Disease Data Analysis*, pp. 455–476, 2019.

Wang, L., Wang, G., Gao, L., Li, X., Yu, S., Kim, M., Wang, Y., and Gu, Z. Spatiotemporal dynamics, nowcasting and forecasting of covid-19 in the united states, 2020.

Wolfinger, R. and Lander, D. Lnq-ens1, 2020. URL https://github.com/reichlab/covid19-forecast-hub/blob/master/data-processed/LNQ-ens1/.

Zou, D., Wang, L., Xu, P., Chen, J., Zhang, W., and Gu, Q. Epidemic model guided machine learning for covid-19 forecasts in the united states. *medRxiv*, 2020. doi: 10.1101/2020.05.24.20111989. URL https://www.medrxiv.org/content/early/2020/05/25/2020.05.24.20111989.

Table 5. COVID-19 Forecast Hub MAE (3 week horizon).

basedate	CU-select	β-AR	Google_Harvard-CPF	LANL-GrowthRate	Microsoft-DeepSTIA
2020-08-06	254.378	202.747	-	296.812	-
2020-08-13	226.903	194.915	-	274.213	-
2020-09-03	211.574	223.471	-	274.117	-
2020-09-27	245.834	200.585	258.717	-	-
2020-10-01	-	320.958	-	400.305	-
2020-10-11	320.967	294.185	353.256	-	-
2020-10-15	498.298	435.184	-	611.468	-
2020-10-25	576.229	492.587	637.124	-	-
2020-11-08	701.287	660.135	868.373	-	-
2020-11-15	757.398	615.995	838.653	-	-
2020-11-19	1008.57	723.106	-	962.261	-
2020-12-06	875.903	783.316	970.478	-	-
2020-12-07	-	712.124	-	844.418	-
2020-12-13	862.825	661.099	823.145	-	-
2020-12-14	-	706.945	-	757.576	872.500
2020-12-17	1165.76	1026.700	-	1096.356	-
2020-12-21	-	765.264	-	839.919	776.512
2021-01-03	790.559	924.455	911.204	-	-
2021-01-04	-	836.486	-	934.932	933.769
2021-01-11	-	538.522	436.358	630.162	723.034

Table 6. COVID-19 Forecast Hub MAE (4 week horizon).

basedate	CU-select	β-AR	Google_Harvard-CPF	LANL-GrowthRate	Microsoft-DeepSTIA
2020-08-06	343.761	279.968	-	397.426	-
2020-08-13	295.25	261.564	-	359.325	-
2020-09-03	292.415	320.063	-	377.327	-
2020-09-27	385	336.712	410.264	-	-
2020-10-01	-	493.807	-	589.957	-
2020-10-11	572.784	523.691	586.931	-	-
2020-10-15	823.338	712.588	-	956.171	-
2020-10-25	955.633	805.278	1010.447	-	-
2020-11-08	1088.82	996.985	1285.209	-	-
2020-11-15	1255.35	970.215	1279.068	-	-
2020-11-19	1486.72	1036.903	-	1487.456	-
2020-12-06	1338.24	1152.662	1429.709	-	-
2020-12-07	-	1073.909	-	1318.017	-
2020-12-13	1426	1088.174	1369.633	-	-
2020-12-14	-	1222.210	-	1309.565	1205.770
2020-12-17	1701.34	1538.276	-	1617.285	-
2020-12-21	-	1271.774	-	1351.437	1422.710
2021-01-03	1052.719	1255.068	1243.796	-	-
2021-01-04	-	1168.079	-	1257.376	1209.617

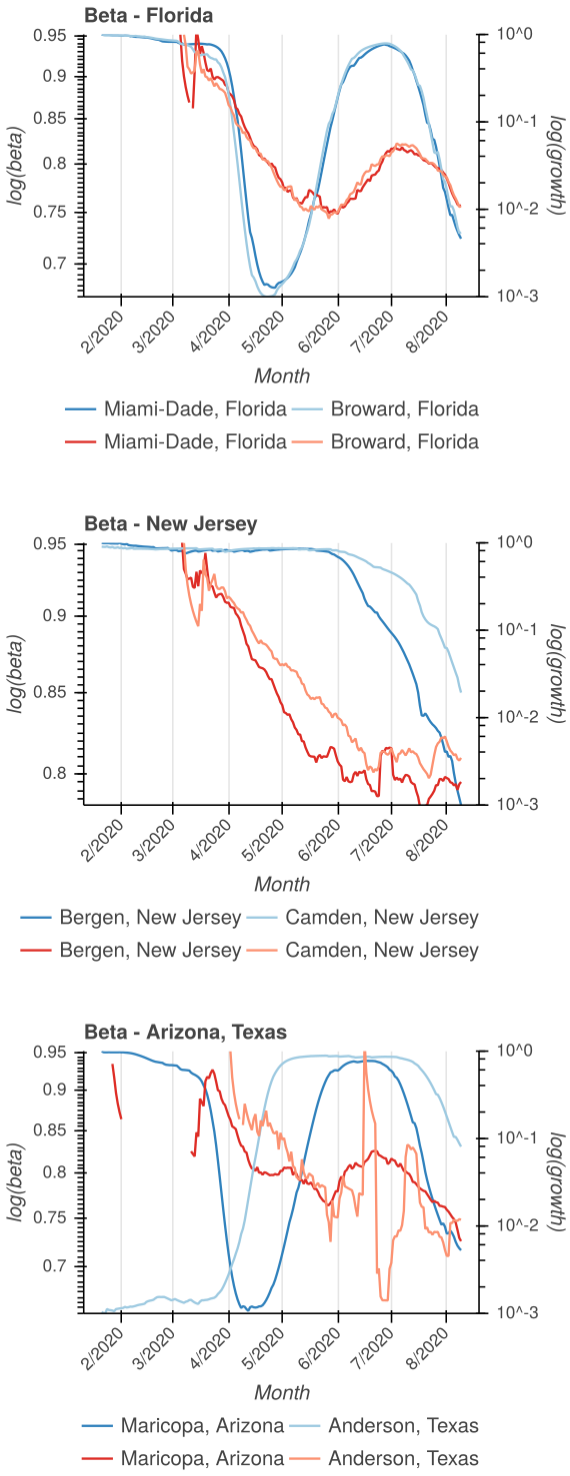

Figure 11. Evolution of β over time