We would like to thank all the Reviewers for their valuable comments that have helped improve our submission.

Reproducibility [R1]: We have put special attention into including all details (initialization, optimization algorithms, parameter values) for reproducibility, and we are working on open-sourcing our code so others can use it conveniently.

Related work [R2]: We agree that our Related Work would benefit from descriptions of comparison models. We briefly discuss them here and will expand the explanations in the paper: **(i) YYG** is an SEIR model with learnable parameters and accounts for reopenings. The parameters are fit using hyperparameter optimization. Unlike ours, YYG uses fixed (time-invariant) rates as SEIR parameters and is limited to modeling standard SEIR compartments. It does not have a systematic mechanism to integrate additional covariates and it cannot benefit from cross-location information sharing. **(ii) IHME** is based on fitting a curve to model the non-linear mixing effects. Unlike ours or others based on compartmental modeling, it does not explicitly model the transitions between the compartments and thus cannot reflect well important inductive biases from epidemiology. **(iii) LANL** is based on statistical-dynamical growth modeling for the underlying numbers of susceptible and infected cases. Unlike ours, it does not model all the available compartments and it does not have any mechanism to learn from covariates. **(iv) UMass** is a Bayesian compartmental model separately fit to each location considering observation noise and a detection rate. Unlike ours, it does not have any mechanism to share information across locations and it does not utilize information extracted from covariates.

Ablation study [R2]: We've added extra ablation studies to clarify the contributions of different model parts. Due to space limit, we only show results for 3 dates in Table 1. Our model with additional compartments outperforms standard SEIR models (only with S, E, $I^{(d)}$ and $R^{(d)}$ compartments), with or without encoder (1st vs. 2nd and 3rd vs. 4th rows). Encoders bring significant performance gains, for both cases. Regularization also helps especially during phases of rapid trend change (e.g. 05/25 and 06/01).

Table 1: 14-day forecasting RMSE for state-level deaths.

Models / Prediction date	05/25	06/01	06/08
Standard SEIR compartments (*w/o* encoder)	219.8	190.8	127.1
Standard SEIR compartments (*with* encoder)	116.4	75.4	81.2
Our model (*w/o* encoder)	128.5	77.0	50.0
Our model	**57.6**	**41.1**	42.0
More ablation cases			
Our model *w/o* fine-tuning	179.9	143.0	122.2
Our model *w/o* partial teacher forcing	1311.6	3141.2	2163.6
Our model *w/o* regularization	130.8	132.5	**40.3**

Model comparison [R2]: We will correct the bold fonts in Appendix. We have been running our model on new dates since our submission and we include those results here to help increase the confidence in our results. Overall, Table 2 shows that *our model consistently outperforms YYG (as well as others that we cannot include due to space limits) by a large margin* in most cases, and indeed, the gap is even larger margin for these recent dates. We attribute this to learning from more training data – by using covariate encoders our model takes advantage of the increasing training data in ways other methods cannot.

Table 2: RMSE for forecasting the number of deaths at state level with different prediction dates and forecasting horizons.

Start date	Predict 5 days		Predict 7 days		Predict 14 days	
	Ours	YYG	Ours	YYG	Ours	YYG
05/24/2020	**51.4**	142.4	**41.5**	143.9	**76.7**	158.3
05/25/2020	**36.6**	138.2	**31.3**	140.6	**61.5**	150.8
05/31/2020	**37.7**	152.9	**37.6**	154.4	**53.2**	163.1
06/01/2020	**28.5**	153.4	**34.7**	155.2	**44.9**	161.7
06/07/2020	**33.8**	157.2	**32.8**	159.1	**50.0**	165.3
06/08/2020	**27.5**	150.8	**30.3**	152.1	**42.5**	157.6
06/15/2020	**39.6**	151.5	**31.4**	153.3	**139.6**	211.0
06/22/2020	**184.3**	225.3	**200.8**	270.1	**224.9**	296.3
06/29/2020	**90.8**	158.4	**110.1**	160.1	**96.6**	168.3

Using covariates [R2, R4]: We agree that using more covariates is advantageous as they contain additional information. Systematic integration of covariates is under-explored in existing epidemiology literature and is one of our major contributions. We convey their value by demonstrating large accuracy gains. We also show that our model outperforms other methods that use covariates such as IHME. We will clarify our contribution as efficient information extraction from extra covariates for epidemiological modeling. We will also add the limitations mentioned by R4.

Policy making use cases [R4]: We will include more explanations on how forecasts can guide better policy decisions. In summary, the benefits for public health are more optimal (i) resource allocation and (ii) non-pharmaceutical interventions. For (i), healthcare organizations rely on accurate forecasts to reallocate health personnel, protective equipment, ventilators and treatment drugs based on the expected severity of COVID-19. E.g. if one county has higher hospitalization forecasts compared to its neighbor, the resources can be shifted to that county to help better address the outbreak. For (ii), local governments can make better-informed decisions on gathering bans and school/business/restaurant closures, as well as improve their messaging to the public based on the expected severity of COVID-19. *Our forecasts already being actively used by several large healthcare organizations and governments, and are publicly available – we will add some notes on these later (we cannot disclose here due to NeurIPS anonymity principles).*

Toning down our claims [R4]: As explained to R2, we have been running benchmarks continuously since our submission and we do seem to be consistently outperforming other models by a large margin. That being said, we appreciate this comment and we agree that we should not make strong claims based on the limited data given the non-stationary environment. We will revise all our claims to tone them down.

Typos [R1,R4]: Line 127 refers to the Table 2 in Appendix. Thanks for all found typos, we will correct them.