FORECASTING THE COVID 19 PANDEMIC

Using Ensemble Data Assimilation to Enhance and Guide Models to More Reliable Predictions

Shaniah Reece

Mentors: Moha Gharamti, Ben Johnson

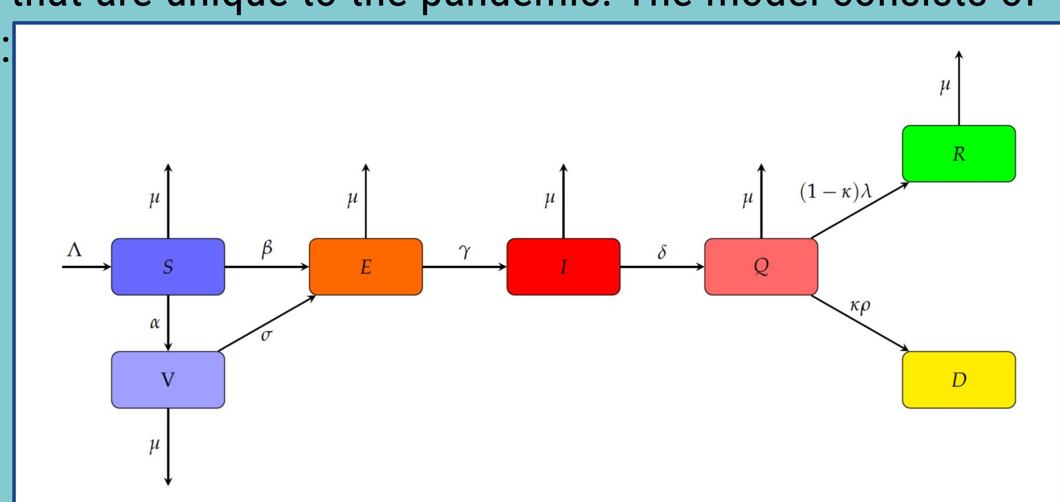
Summary

Numerical modeling is a common approach employed by researchers and scientists to analyze and describe the behavior of a physical system. However, model biases, observational error and data sparsity are major limitations to this approach making it difficult to accurately represent the state of a system and generate reliable forecasts. Data assimilation is a widely implemented computational technique in the Earth Sciences that can enhance model predictions by combining observations with prior forecasts. Most notably, it is used to initialize atmospheric models for weather forecasting. This research demonstrates how the coupling of NCAR's Data Assimilation Research Testbed (DART) with an extended SEIR model generates better forecasts of the COVID-19 pandemic than the SEIR model alone. We compare the outcome of data assimilation performed on four different countries around the world with different levels of COVID preparedness and response to demonstrate its effectiveness. The findings of this research denote that even in poorly tuned systems and for datasets with sparse or missing values, ensemble data assimilation can still effectively compensate for these issues and create a fairly informative picture of the pandemic. It is hoped that this study will motivate the adoption of data assimilation to aid epidemiological models.

Epidemiological Modeling

Numerical models have long been used by scientists to simulate real world systems. In this study, a recently modified version of the epidemiological model –SIR– is used to simulate the COVID-19 pandemic. The extended SEIR model –SEIQRDV– accounts for additional compartments that are unique to the pandemic. The model consists of seven state variables namely:

- S-Susceptible
- E-Exposed
- I-Infected
- Q-Quarantined
- R-recovered
- D-Death
- V-Vaccinated



Ghostine, R., Gharamti, M., Hassrouny, S., & Hoteit, I. (2021). An extended SEIR model with vaccination for forecasting the COVID-19 pandemic in Saudi Arabia using an ensemble Kalman filter. Mathematics, 9(6), 636.

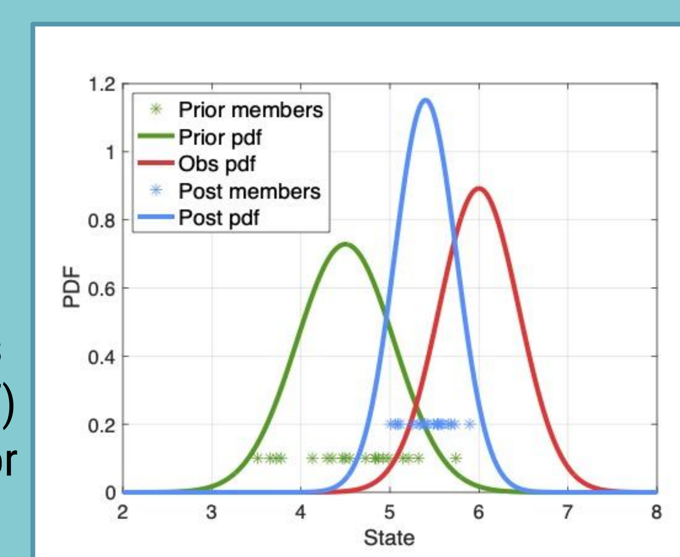
In addition to the prognostic state, the model utilizes a comprehensive set of parameters that describes several aspects of the COVID-19 virus for instance, transmission rate. These parameters however can be highly uncertain which can degrade the performance of the SEIQRDV model.

Parameter	Description
Λ	New births and new residents
β_1	Transmission rate before intervention
β_2	Transmission rate during and after intervention
α	Vaccination rate
μ	Natural death rate
γ^{-1}	Incubation period
σ	Vaccine inefficacy
δ^{-1}	Infection time
κ	Case fatality rate
λ^{-1}	Recovery time
ρ^{-1}	Time until death

Data Assimilation

Data Assimilation (DA) involves combining model predictions with available observations to produce accurate state estimates. In its simplest form, DA follows Bayes theorem:

$$p(x|y) \approx p(x) \cdot p(y|x)$$



Gharamti, Moha. "Welcome to the Data Assimilation Research Testbed." Welcome to the Data Assimilation Research Testbed - DART 10.0.4 Documentation, https://docs.dart.ucar.edu/en/latest/README.html.

In this work, the ensemble Kalman filter (EnKF) is utilized. The EnKF approximates the prior Probability Density Function (pdf) of the system using ensemble members or model realizations. This simplifies the problem in the presence of nonlinearities.

Data Assimilation was implemented using a Matlab-based version of the DART software (DART LAB). The SEIQRDV was added to an established list of models available in DART LAB. A few new algorithmic features we augment to DART LAB include: Gaussian Anamorphosis, additive inflation, various initial state distributions. We conducted sensitivity experiments by altering DA configurations such as multiplicative inflation, filtering types, ensemble size and observation types.

Conclusion

Epidemiological models are important, however they can be highly uncertain and prone to biases. Data assimilation can be a valuable aid to mitigate these uncertainties and to further improve their predictions.

The case studies explored demonstrated the following:

- A properly tuned model is indispensable to obtain reliable predictions
- DA allows for correction in unobserved variables
- DA uncovered unreliable pandemic data (e.g. Recovery data)
- Model nonlinearities such as the vaccination compartment can be resolved better using non-linear filters (e.g. Rank Histogram Filter)

Acknowledgements

Thank you to my mentors Moha Gharamti and Ben Johnson for their guidance throughout this research project. The DAREs team for their support and encouragement, Virginia Do, director of the SIParCS program at NCAR|UCAR and the National Science Foundation.

Results

USA

A. Poorly Tuned

Model parameterized using two transmission rates: before and after lockdowns

B. Properly Tuned

More sophisticated parameterization of beta accounting for waves, variants and lockdowns

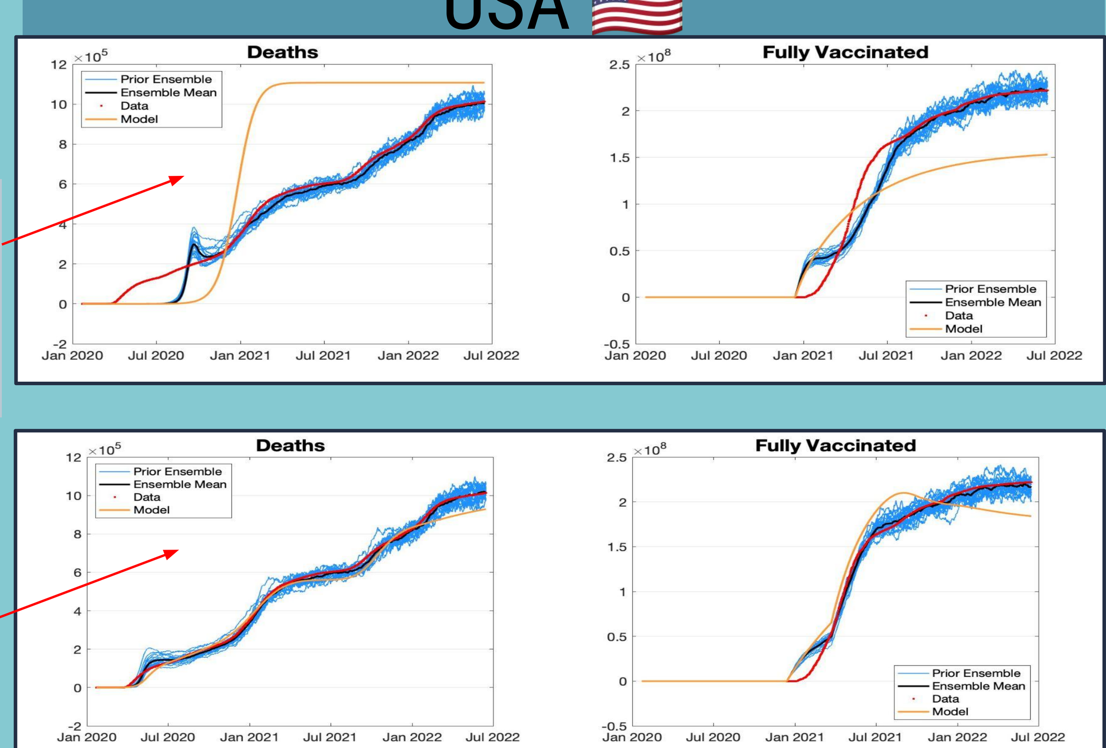


Fig1. The effects of DA on a properly tuned vs poorly tuned model

Guyana

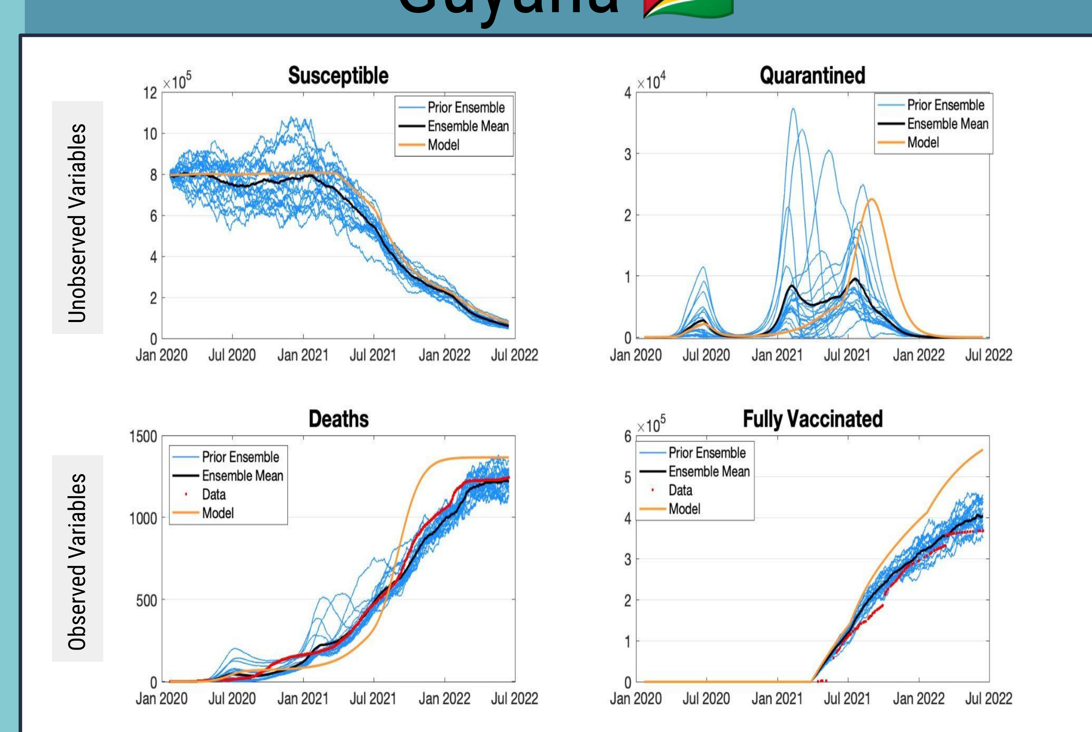


Fig 2. DA performance on datasets with gaps and unobserved variables

Ethiopa

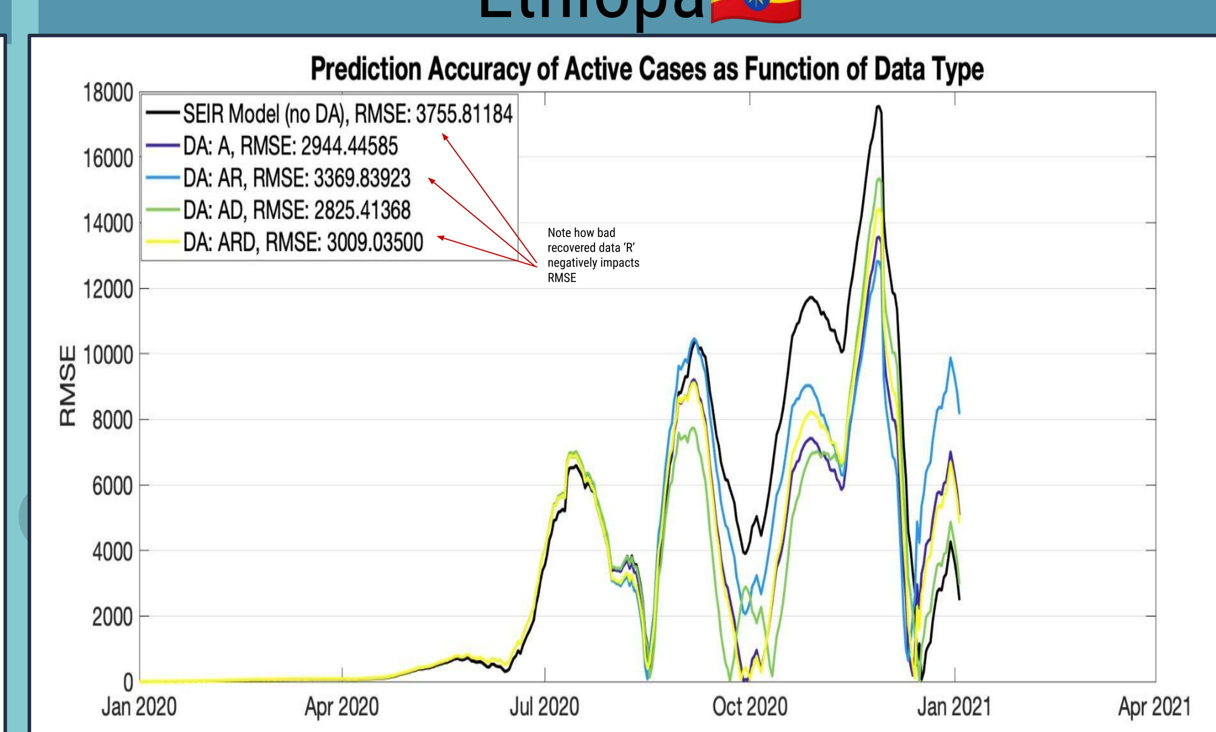


Fig 3. The impact of adding different data types and unreliable data on RMSE

Trinidad

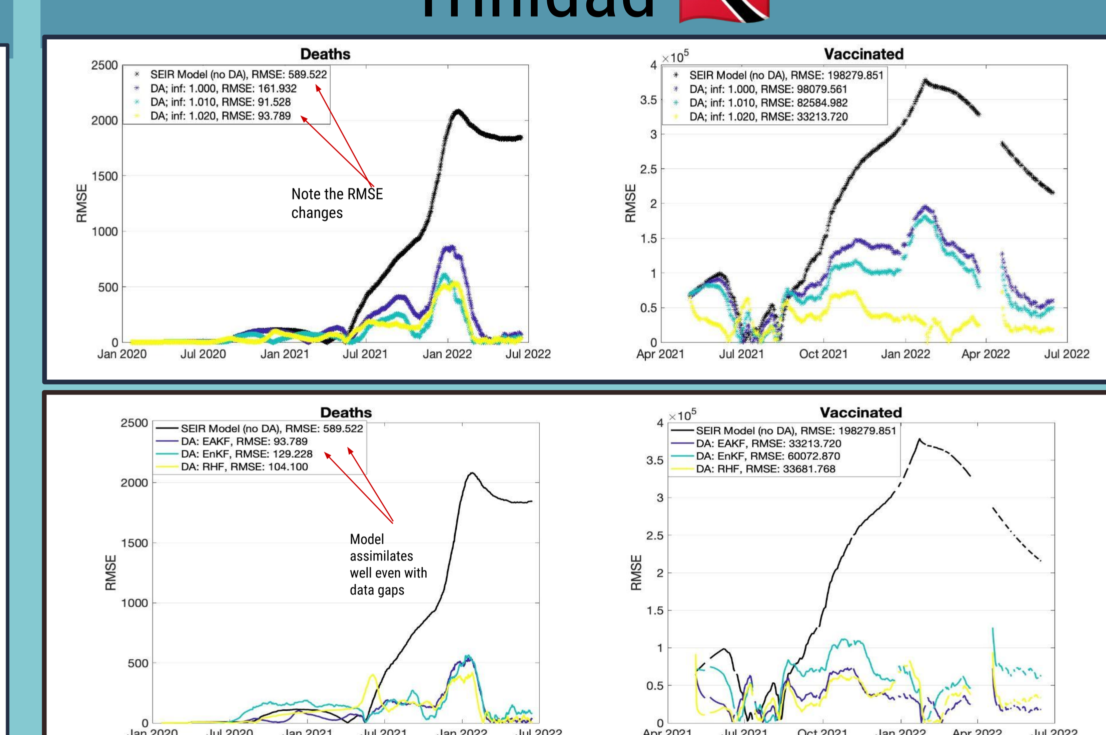


Fig. 4 The impact of Inflation and Filtering algorithms on DA

NCAR | NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

MCNAIR SCHOLARS PROGRAM

UMBC AN HONORS UNIVERSITY