Forecasting Black Sigatoka Infection Risks with Latent Neural ODEs

Yuchen Wang, Matthieu Chan Chee, Ziyad Edher, Minh Duc Hoang, Shion Fujimori, Jesse Bettencourt

Computer Science
UNIVERSITY OF TORONTO


Black
Sigatoka

U.S. DEPARTMENT OF ENERGY

Applying fungicides regularly accounts for

40%

of the banana production cost

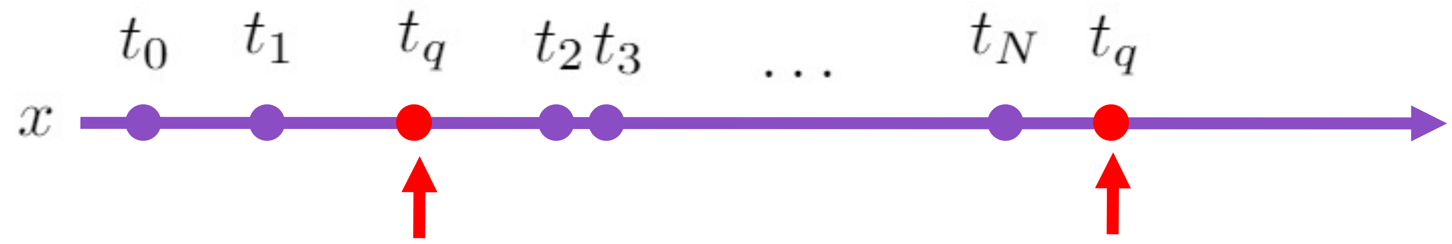
Forecast black Sigatoka infections risks

1 month

into the future

Bonus:
Outstanding interpolation capabilities, even when inferencing with **10%** of data window!

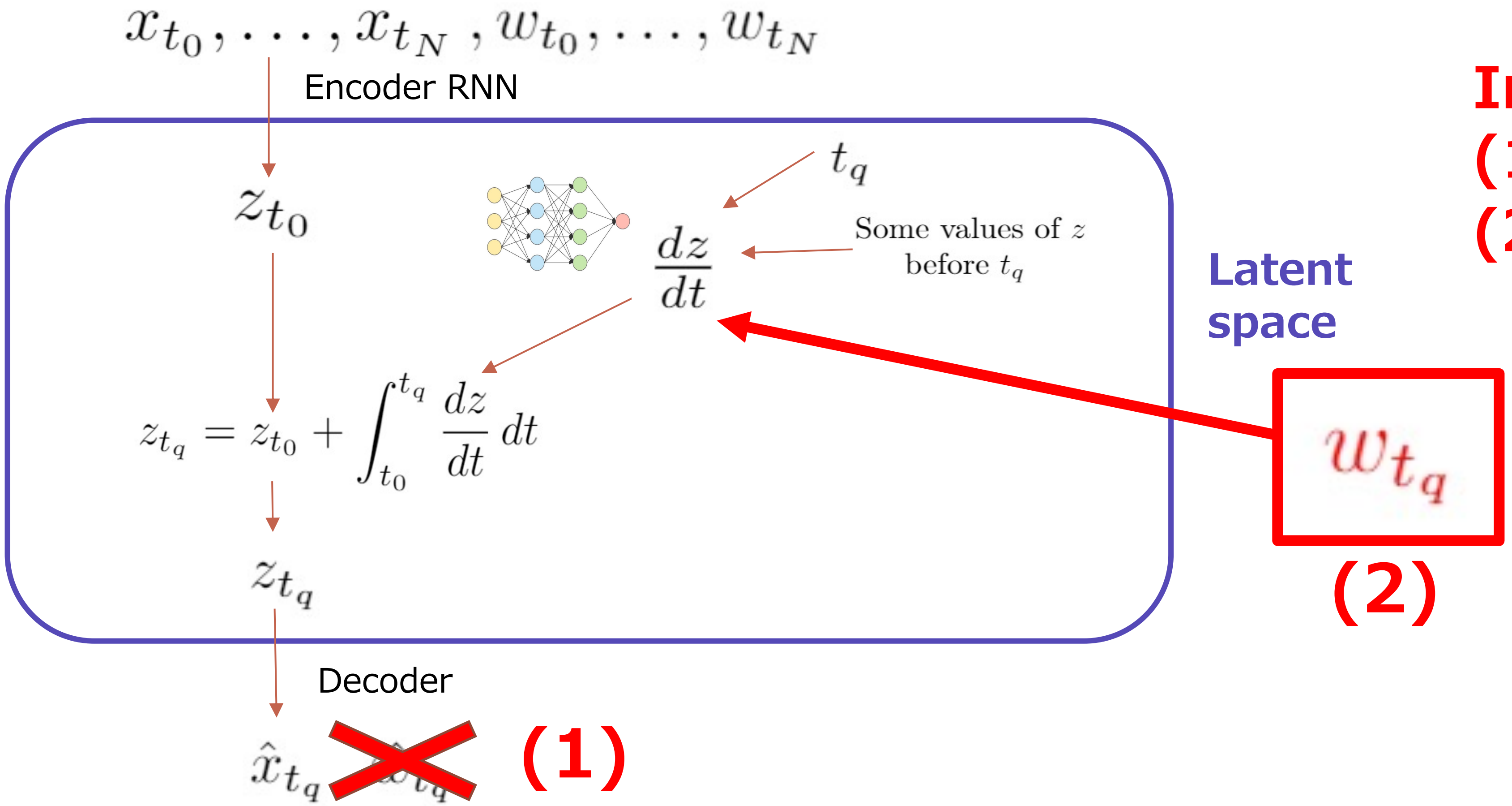
Multiple predictoR Neural ODE (MR. NODE)



x:	Infection rate
$t_0, \dots, t_N$:	Observed time steps (possibly irregular)
t_q:	Unobserved time step ($t_q \geq t_0$)
w:	Weather conditions

Goal: Find infection rate at time t_q using weather!!

Using MR. NODE:



Innovations:
(1) 'Partial' autoencoder
(2) Inject weather into dz/dt

Datasets


Costa Rica
Lake Nicaragua
Nicaragua
Caribbean Sea
Arenal
Quesada
Alajuela
Heredia
San Jose
Cartago
Pacuare
Chirripó
Puerto Limón
Cahuita National Park
Gulf of Nicoya
Manuel Antonios National Park
Dominical
Panama
Pacific Ocean
Inter-American Hwy
Corcovado National Park
Sweet Gulf
50 Miles

Data generation

Bebber (2019)

$$r(T) = \left(\frac{T_{max} - T}{T_{max} - T_{opt}}\right)\left(\frac{T - T_{min}}{T_{opt} - T_{min}}\right)^{\frac{T_{opt} - T_{min}}{T_{max} - T_{opt}}}$$

$$H(t,T) = r(T)\left(\frac{t}{\alpha}\right)^{\gamma}$$

$$F(t,T) = 1 - e^{-H(t,T)}$$

$$Y(t,T) = \beta F(t,T)$$

relative humidity

canopy temperature

moisture storage on canopy

(N = 91,556)

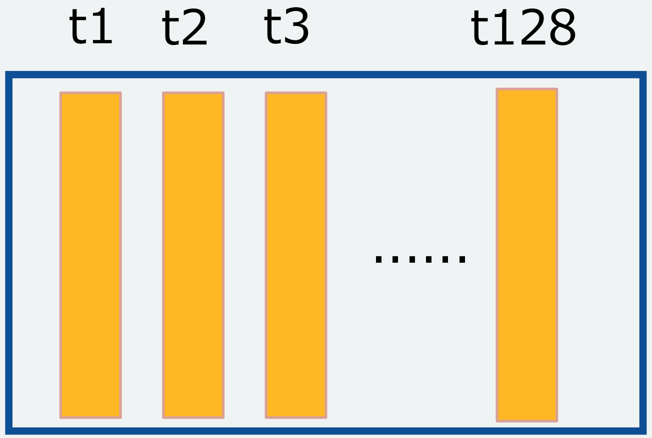
Relative number of infections

Experiments

- Encoder: LSTM
- Loss Functions:
  - Negative Log Likelihood (train & valid)
  - Mean Squared Error (test)

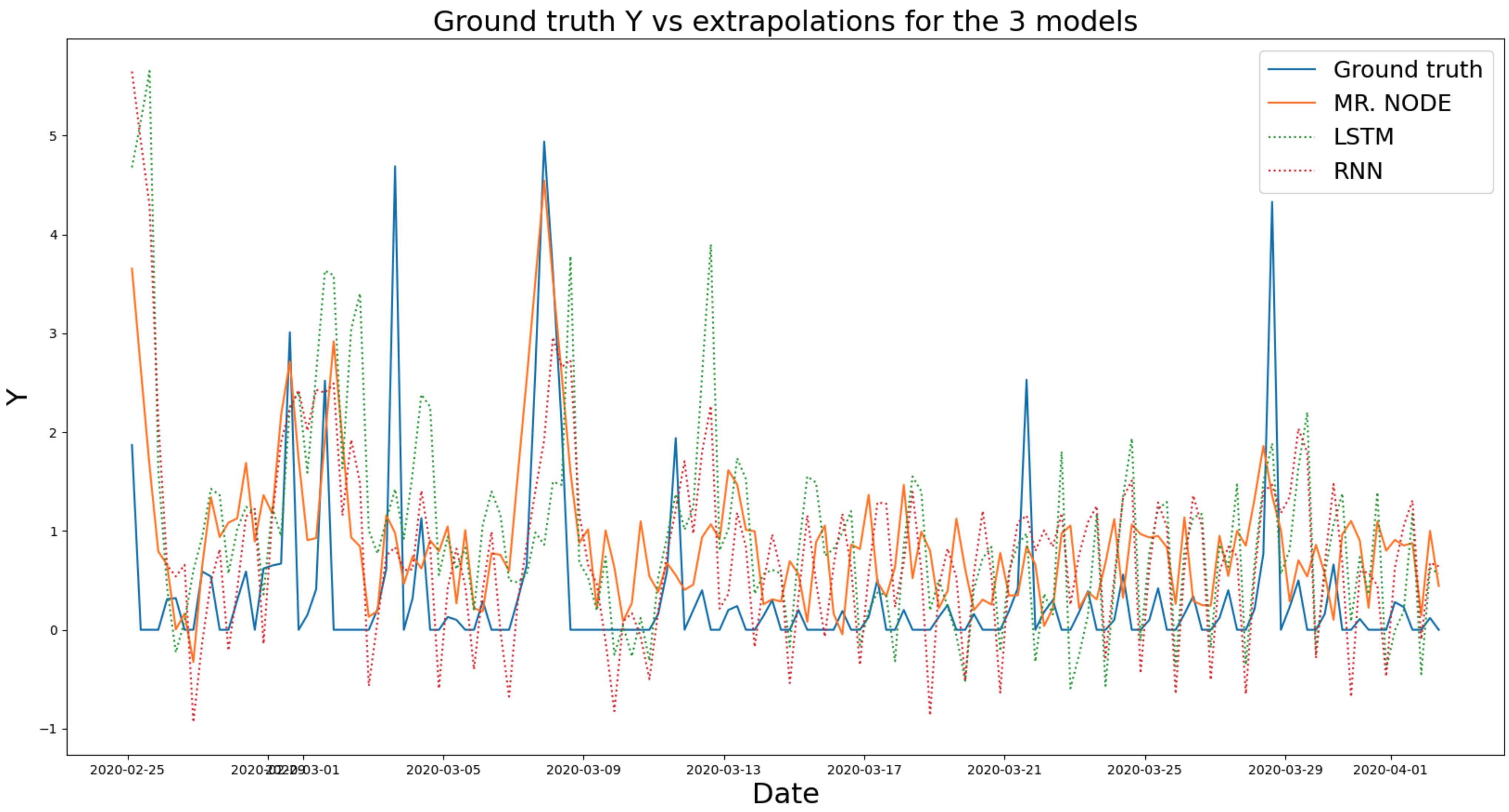
Long Extrapolation

	# encoded	#reconstructed	#extrapolated
Training	128	128	0
Validation	100	100	150
Extrapolation Test	100/70/50/30 (drop rate 0/0.3/0.5/0.7)	-	150 (37.5 days)
Interpolation Test	100/70/30/10 (drop rate 0/0.3/0.7/0.9)	100	0



A data window

Long Extrapolation

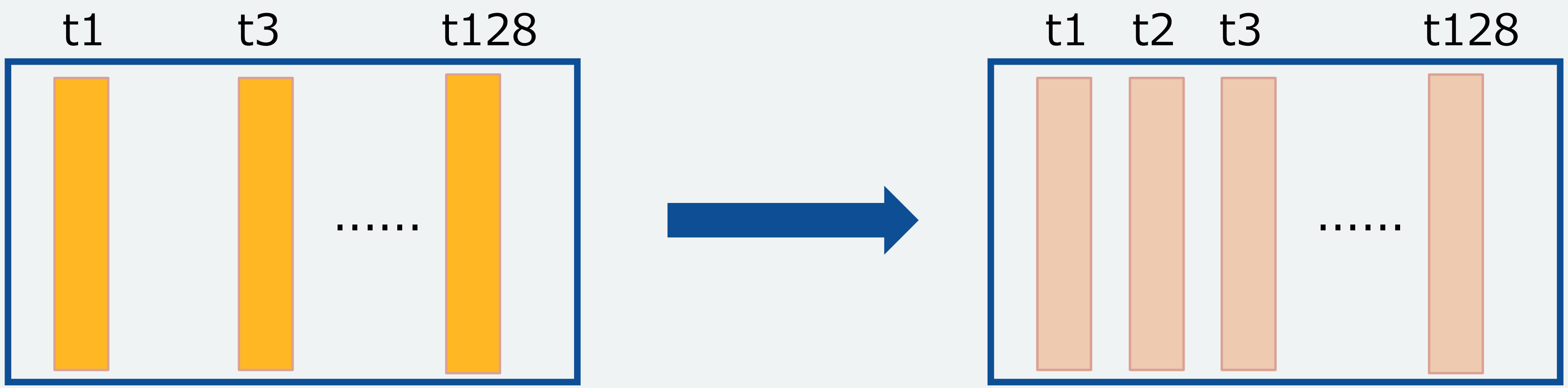


Long Extrapolation

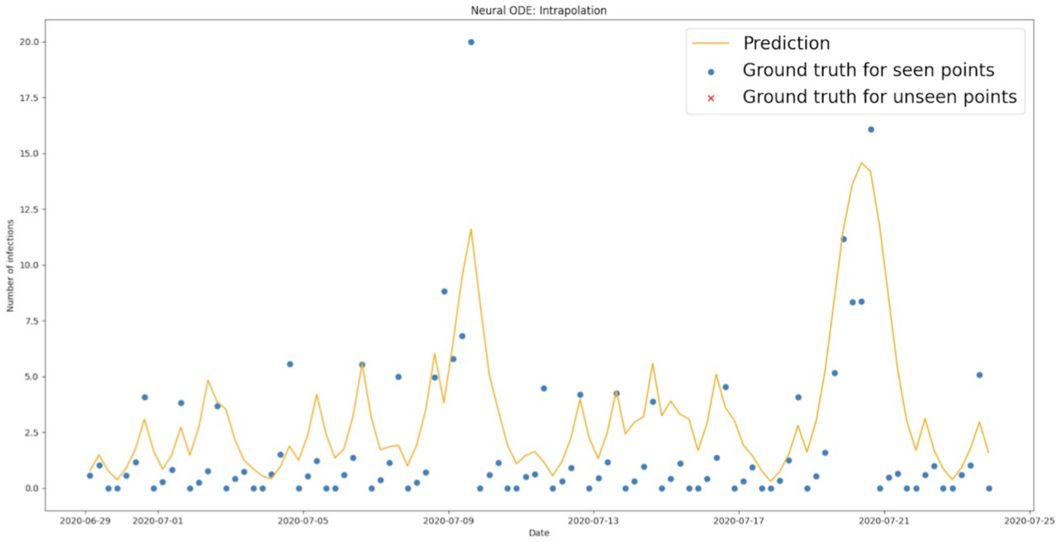
Method	Drop rates	MSE on test set
RNN	0 0.3 0.5 0.7	13.55 13.51 12.62 13.70
LSTM	0 0.3 0.5 0.7	12.76 12.71 17.94 14.12
MR. NODE	0 0.3 0.5 0.7	**12.16** **12.29** **12.40** **12.68**

Irregular Interpolation

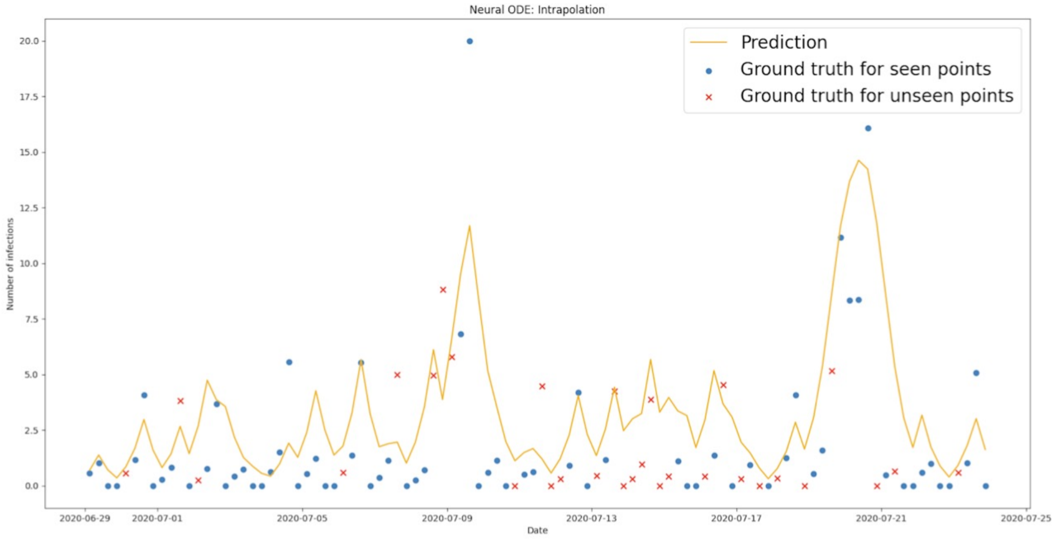
	# encoded	#reconstructed	#extrapolated
Training	128	128	0
Validation	100	100	150
Extrapolation Test	100/70/50/30 (drop rate 0/0.3/0.5/0.7)	-	150 (37.5 days)
Interpolation Test	100/70/30/10 (drop rate 0/0.3/0.7/0.9)	100	0



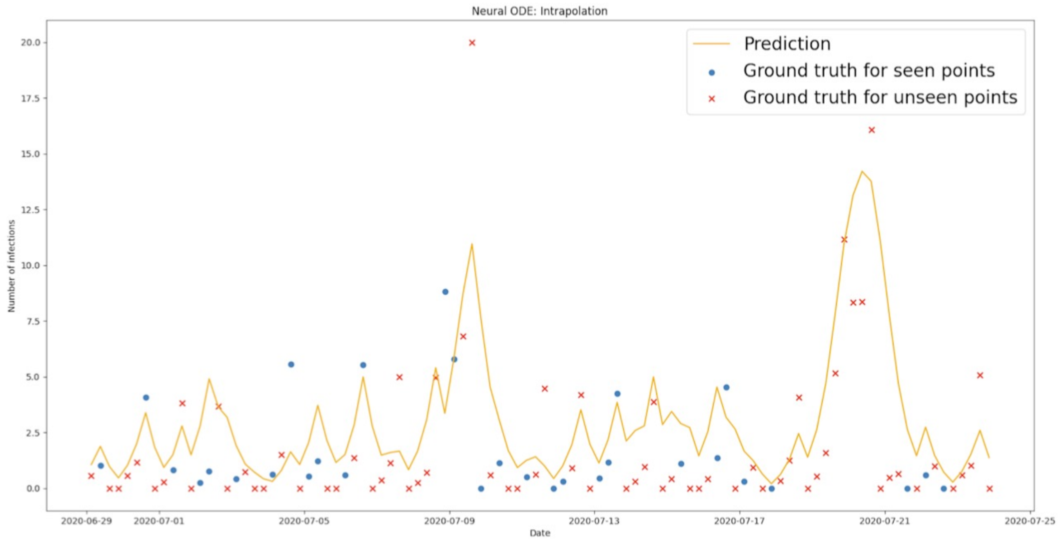
Irregular Interpolation



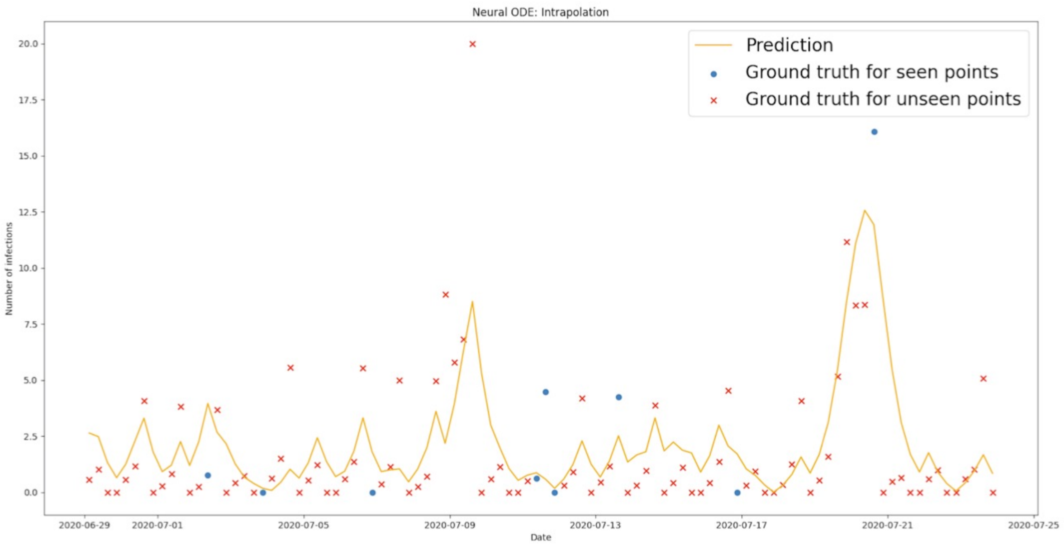
(a) dropping rate = 0



(b) dropping rate = 0.3



(c) dropping rate = 0.7



(d) dropping rate = 0.9

Takeaway

- New ML method for time-series with multiple predictors
- Can predict **1 month** into the future
- Outstanding interpolation capabilities, even when using **only 10%** of the data window

Thank you! ☺