



CoRoNNa: A Deep Sequential Framework to Predict Epidemic Spread from Mobility and Viral Diffusion Trajectories

Dario Aragona, Luca Podo, [Bardh Prenkaj](#), Paola Velardi



SAPIENZA
UNIVERSITÀ DI ROMA

Introduction on Epidemic Spread Prediction

- **Characteristics:**
 - Long-term sequence of measurements
 - Regular intervals between events
 - Hidden long-term dependencies within the sequence
- **Task:** Epidemic prediction on **viral spread** of COVID-19
- **This paper:** Analysis of the mobility containment policies effect on the viral spread and prediction of *daily positive cases* in the infection trajectories

Related work

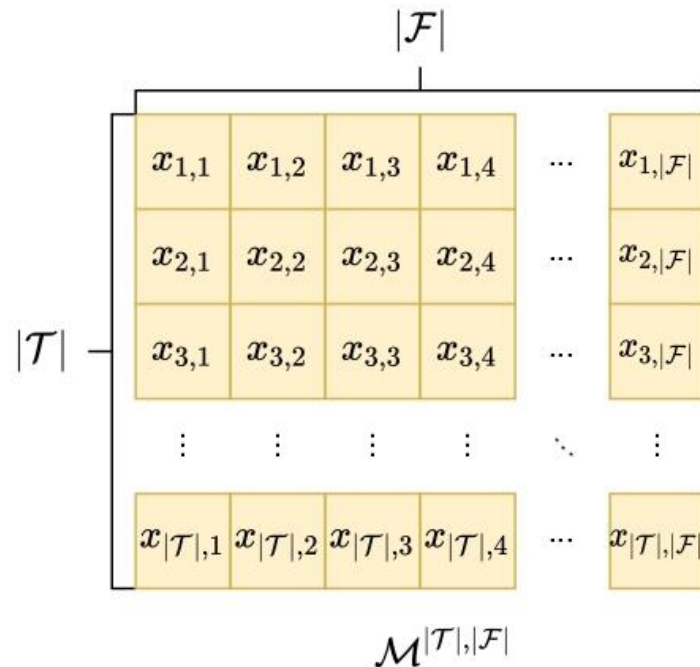
- **Diffusion models [1-6]:**
 - Based on specific infection parameters of the virus
 - Usually augmented with environmental and weather data
 - Example - SIR, SEIR, LT
- **Trajectory-based models:**
 - Stationary linear stochastic models [7-10] predict the value at instant t by regressing its past values
 - Deep sequential models learn long-term patterns in sequential data [11,12]

Related work

- Diffusion models are sensitive to the epidemiological parameters (i.e. **infectivity period**)
 - Unstable and less effective when the parameters are highly variable
- Stochastic models are **not effective with long historical data**
 - The COVID-19 spread depends on mobility containment policies to combat the disease
 - The effect of the containment policies is visible after several days/weeks
- Deep sequential methods are more **promising**:
 - No need to encode epidemiological parameters explicitly
 - Detection of underlying patterns and long-term dependencies

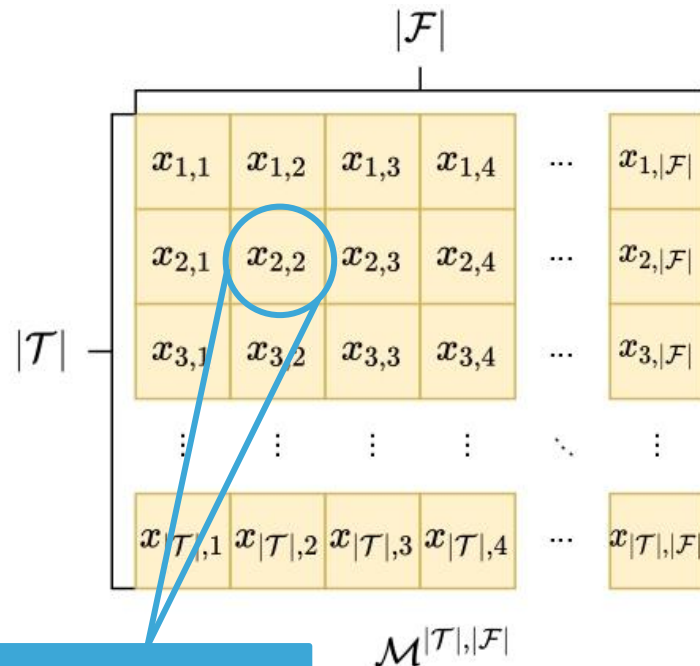
Input modelling formalisation

- The COVID-19 spread in a population is **time-dependent**
 - Number of infected, recovered and deceased patients change in time
- Time-series is a **discrete sequence** of random variables $\{x_t\}$ ordered in time $t \in T \subset \mathbb{R}$
- Let $M^{|T|,|F|}$ be a **time-matrix** of $|T|$ rows and $|F|$ columns:
 - T is the set of timestamps
 - F is the collection of time-variables (features)



Input modelling formalisation

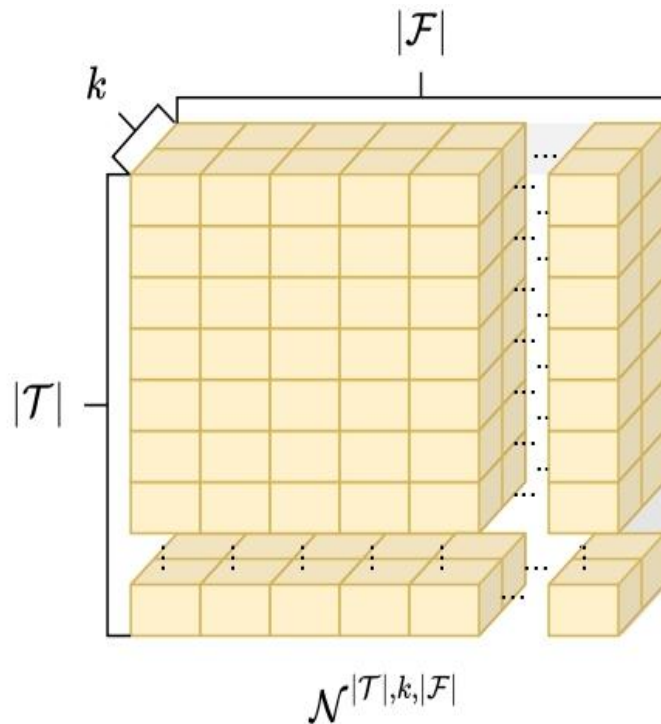
- The COVID-19 spread in a population is **time-dependent**
 - Number of infected, recovered and deceased patients change in time
- Time-series is a **discrete sequence** of random variables $\{x_t\}$ ordered in time $t \in T \subset \mathbb{R}$
- Let $M^{|T|,|F|}$ be a **time-matrix** of $|T|$ rows and $|F|$ columns:
 - T is the set of timestamps
 - F is the collection of time-variables (features)



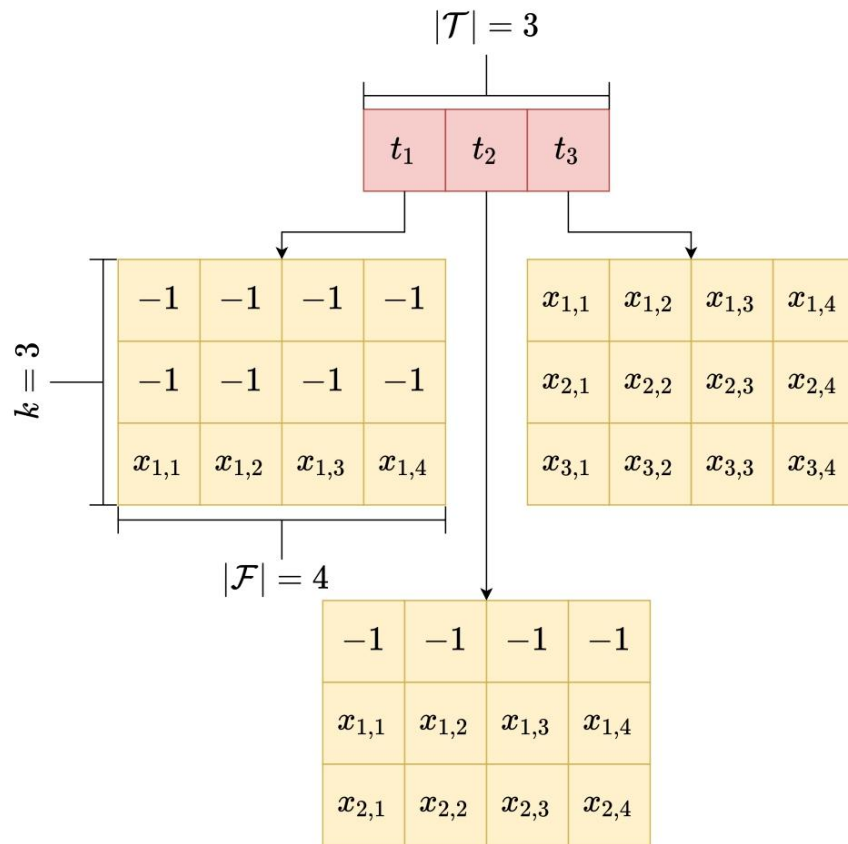
Input modelling formalisation

- Temporal **lag** and **lead**
 - Lag (k) = the number of days to look back
 - Lead (ℓ) = the number of days to predict in the future
- Extend $M^{|\mathcal{T}|,|\mathcal{F}|}$ into a time-tensor $N^{|\mathcal{T}|,k,|\mathcal{F}|}$ to represent the lag
- For each $t \in \mathcal{T}, i \in [1,k], j \in [1,|\mathcal{F}|]$ we have:

$$N_{t,i,j} = \begin{cases} \mathcal{M}_{t-k+i,j} & \text{if } t - k + i > 0 \\ -1 & \text{otherwise} \end{cases}$$



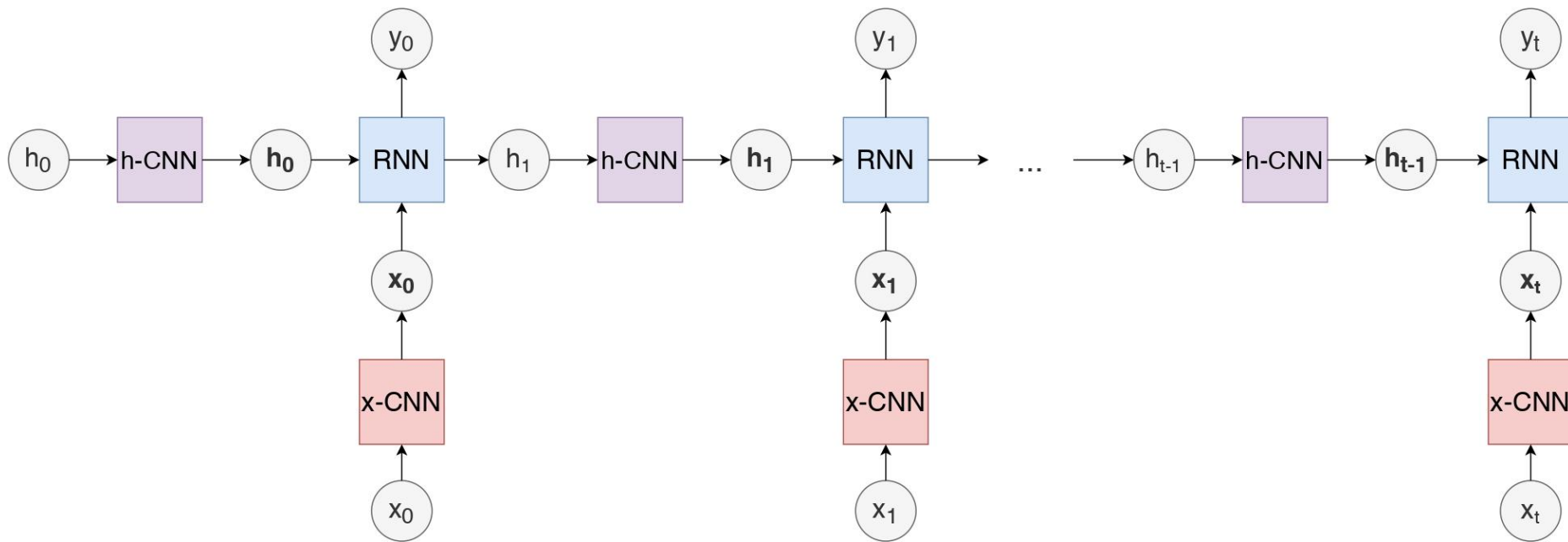
Input modelling formalisation



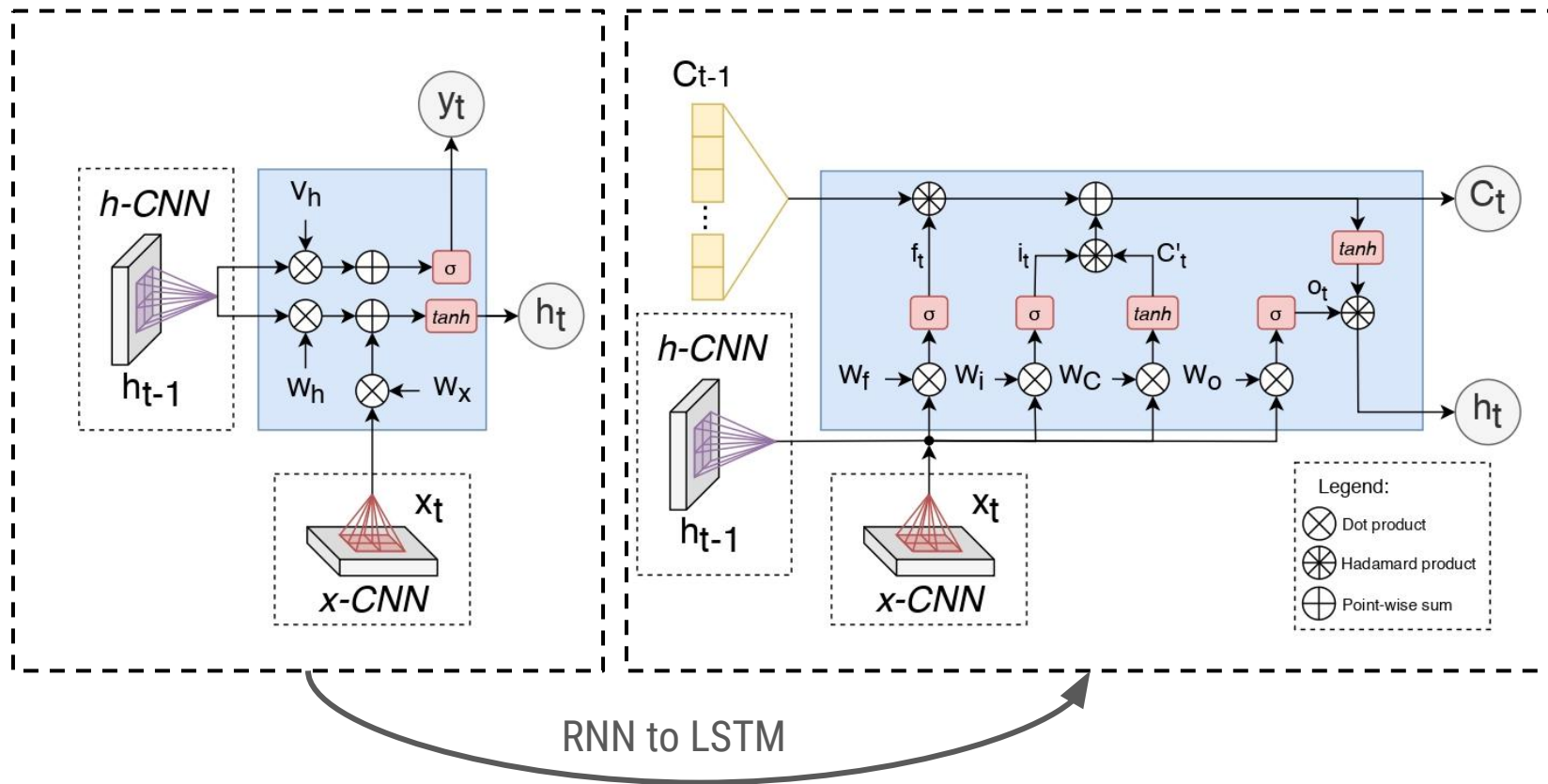
Predictive strategy

- CNN + RNN = CoRoNNa
- **3 modules:**
 - x-CNN = convolution on the input trajectory at time t
 - h-CNN = convolution on the hidden vector at time $t-1$
 - Main RNN = finds features on the curves of daily infected patients
- Polymorphic and agile structure
 - Plug-in-plug-out architecture
 - Fully-configurable components

Predictive strategy



Predictive strategy



Datasets

	Covid spread infection data	Mobility patterns	COVID-19 features
Italy (A)	Protezione Civile - Emergenza Coronavirus	Google and Apple mobility patterns	<ul style="list-style-type: none">• N° of daily positive cases• N° of deceased persons• N° of daily swab tests• N° of daily ICU admissions
Switzerland (A)	Our World in Data	//	//
South Korea (B)	//	//	//
Japan (B)	//	//	//
Brazil (C)	//	//	*missing N° of daily swab tests
Texas, US (C)	//	//	//

A = countries with severe lockdown policies; B = light containment; C = no constraints

Mobility patterns

- **Google mobility data:** mobility trends related to specific places
 - Retail & Recreation (restaurants, cafes, shopping centers...)
 - Grocery & Pharmacy
 - Parks, Transit Stations, Workplaces
 - Residential
- **Apple mobility data:** trends related to directions request
 - Walking
 - Driving (private car)
 - Transit (public transport)

Experimental setup

- **Objectives:**

- Assess the predictive power of infection (F1) and mobility (F2) trajectories in countries with containment strategies, compare with countries with no or little containment
- Compare the quality of predictions of the models

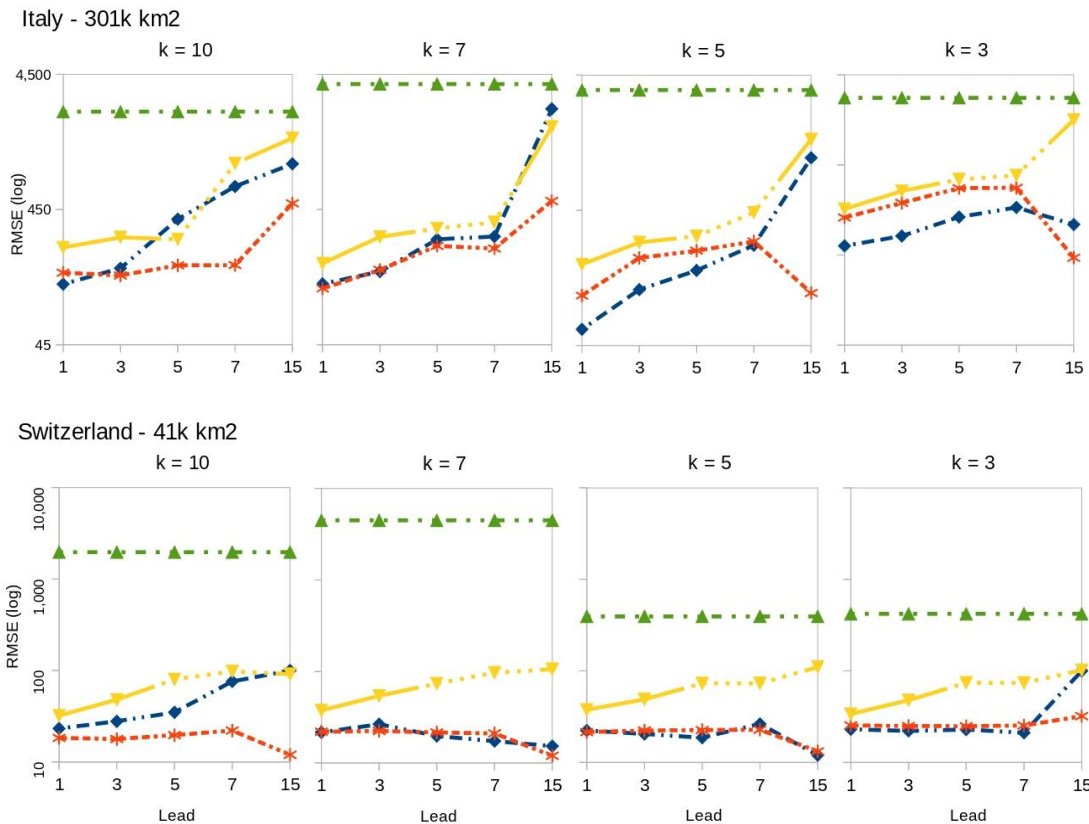
Experimental setup

- **Metrics:**
 - RMSE
- **Time-window:**
 - Combination of all $k \in \{3,5,7,10\}$ and $l \in \{1,3,5,7,15\}$
- **Training:**
 - 60:10:30 split into training, validation, and test sets
 - Splitting is done sequentially in the time axis

Experimental setup

- **Competitors:**
 - LSTM-Yang [12] with 25 cells and a sigmoid activation function
 - VAR
 - Dependency evaluation between DPC and other series
 - Quadratic detrending for stationarity to be met

Results (F1)



Legend:

* LSTM-Yang

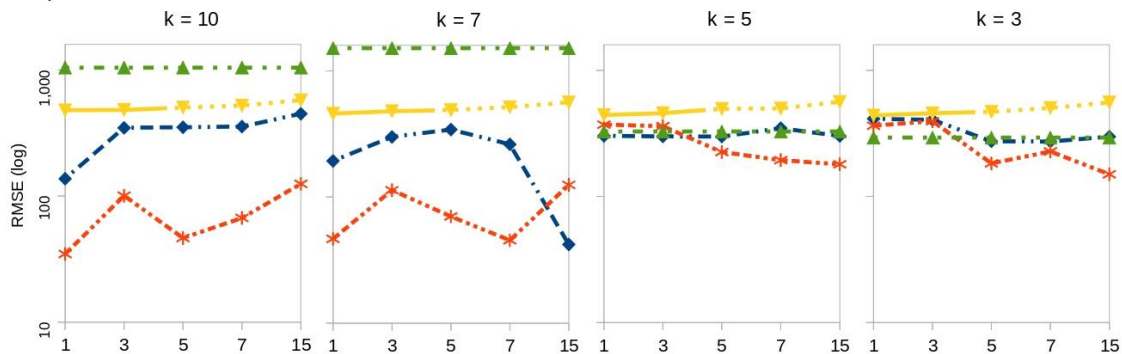
* CoRoNNA

* Random Forest

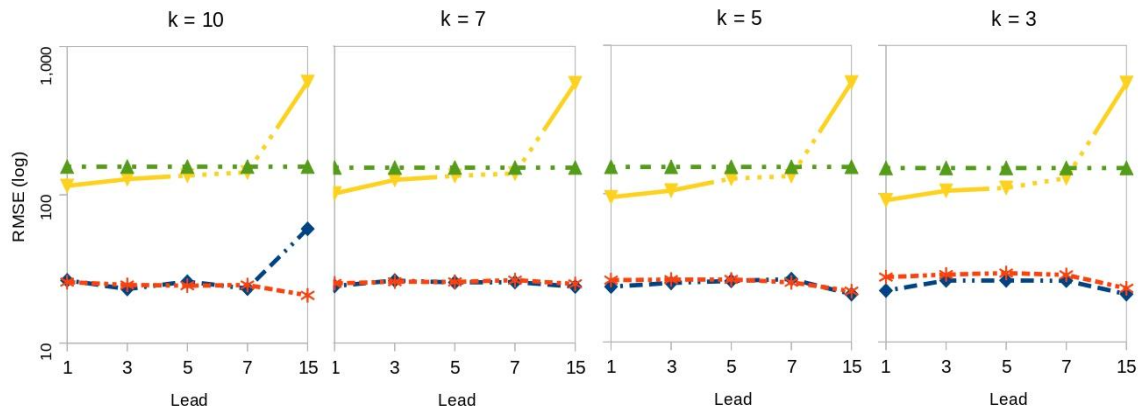
* VAR

Results (F1)

Japan - 377k km²

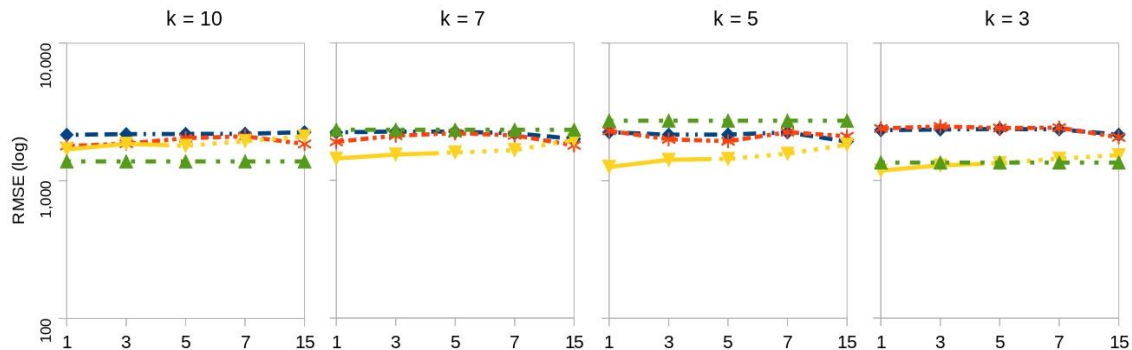


South Korea - 100k km²



Results (F1)

Texas - 695k km2



Legend:

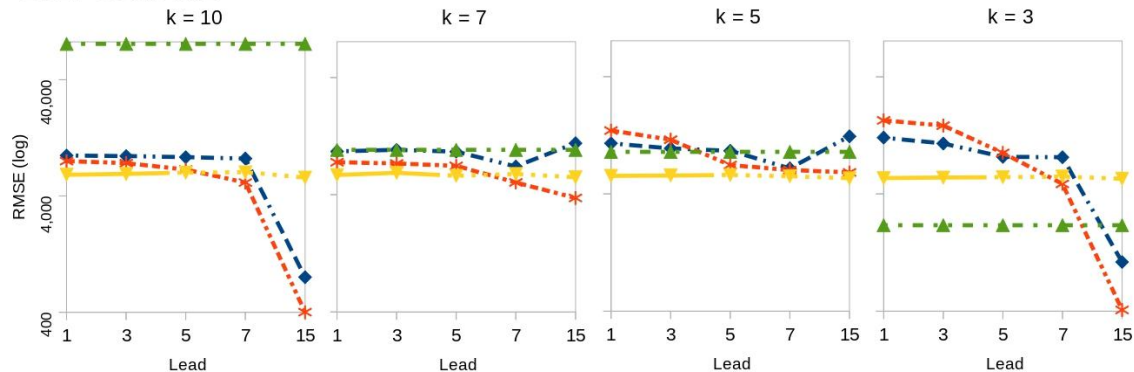
* LSTM-Yang

* CoRoNNa

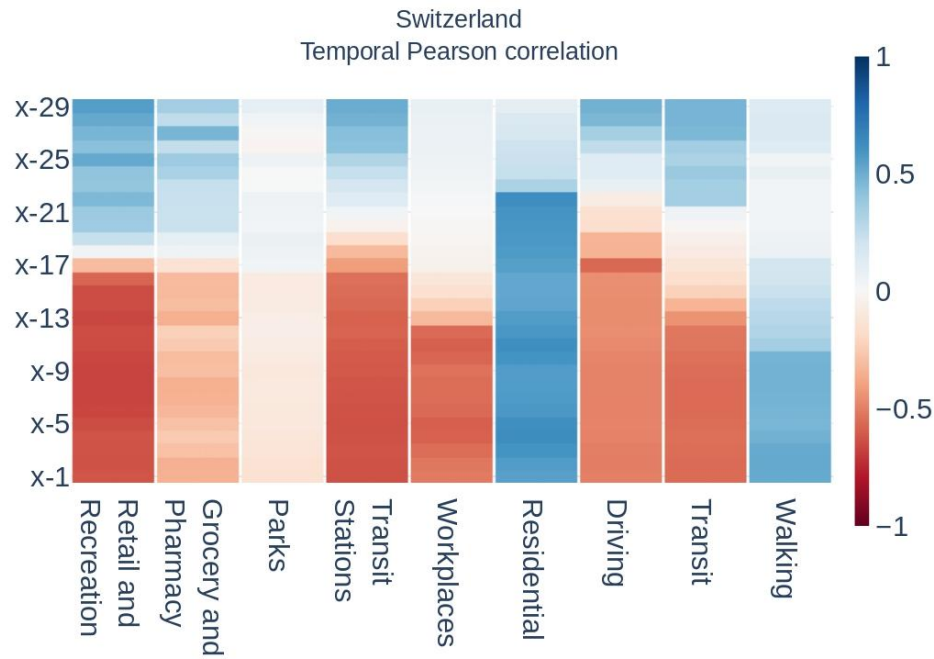
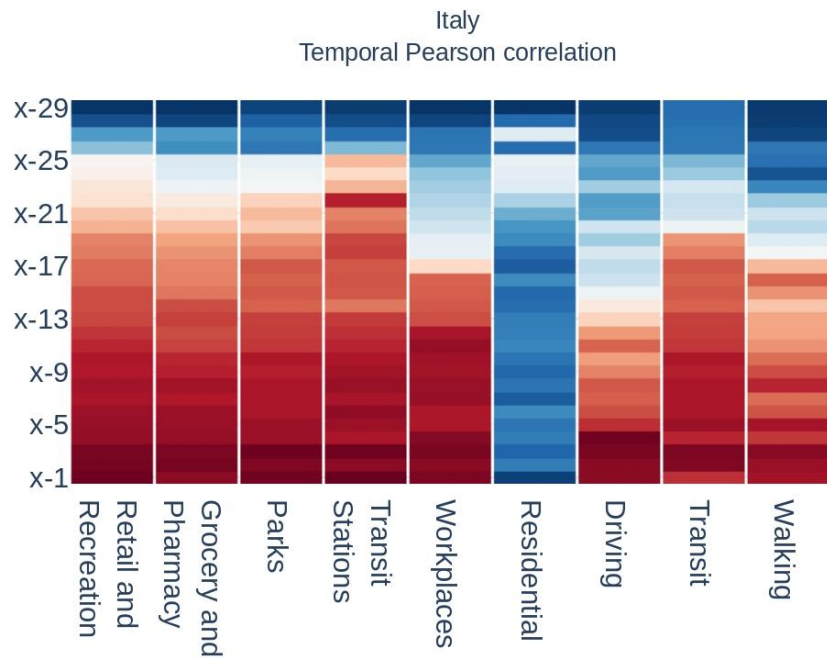
* Random Forest

* VAR

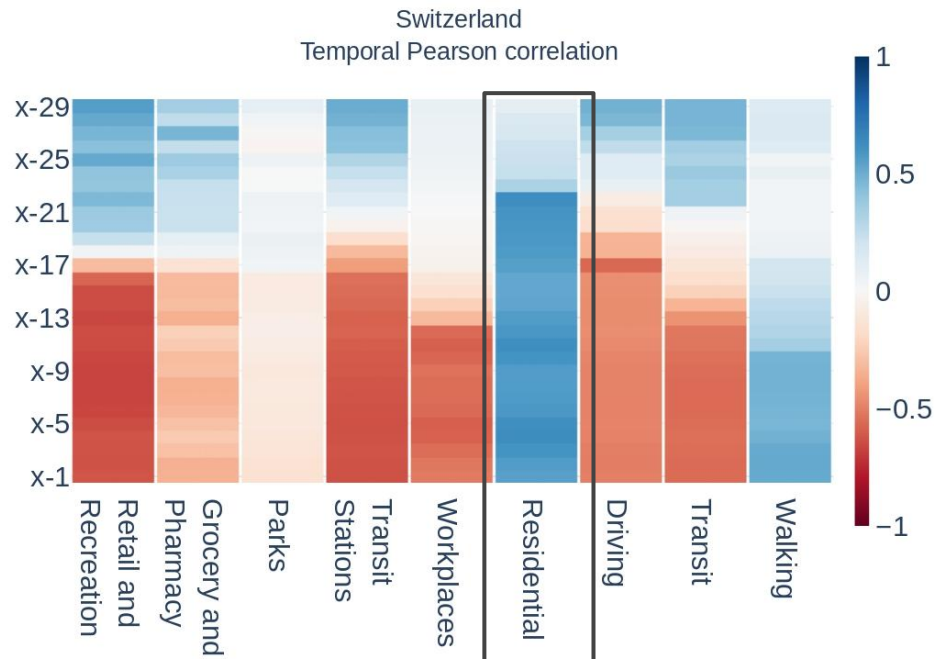
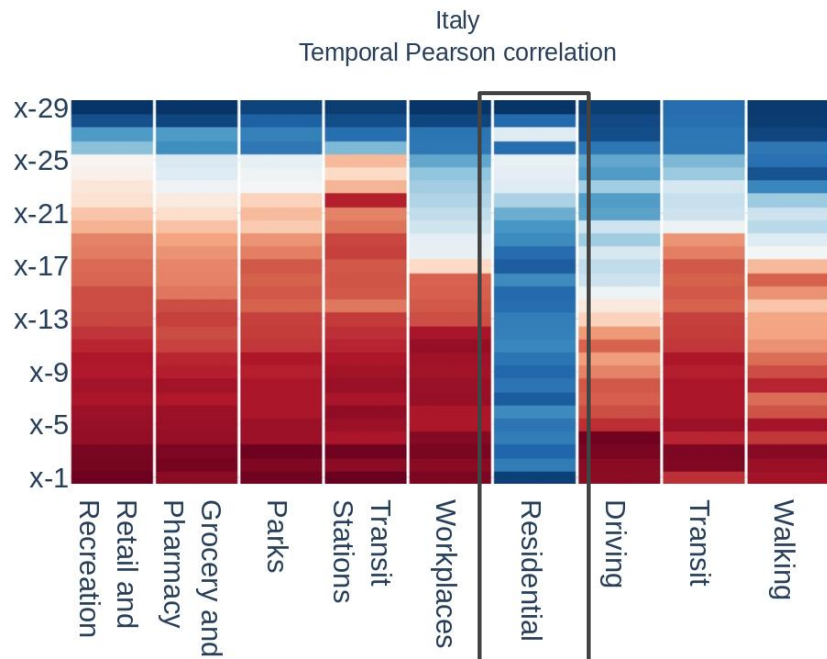
Brazil- 8516m km2



Results (F2)

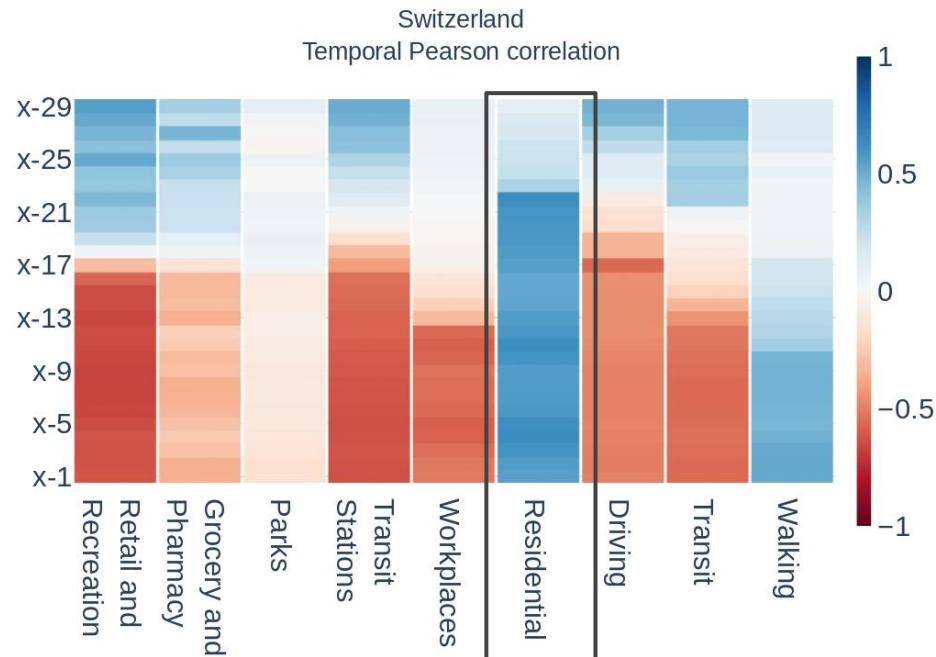
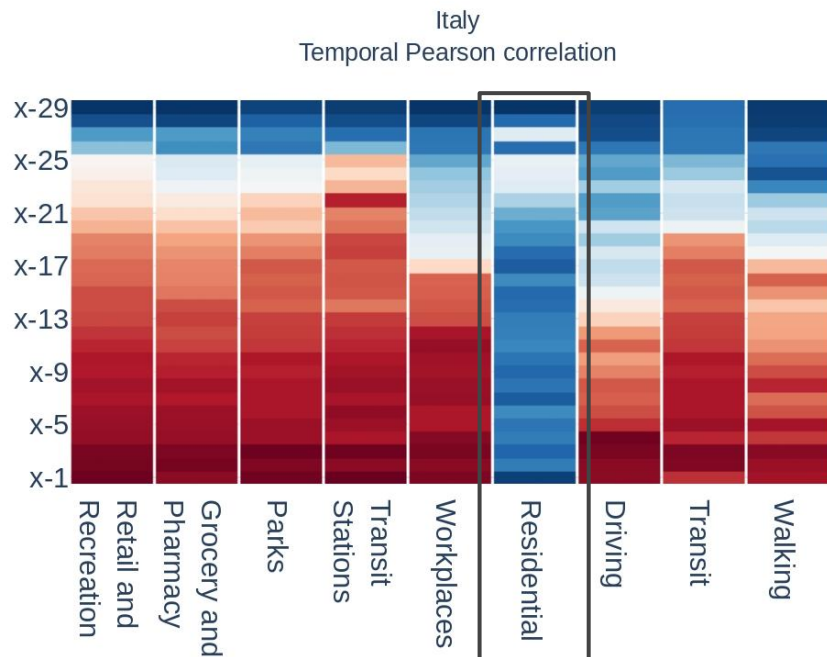


Results (F2)



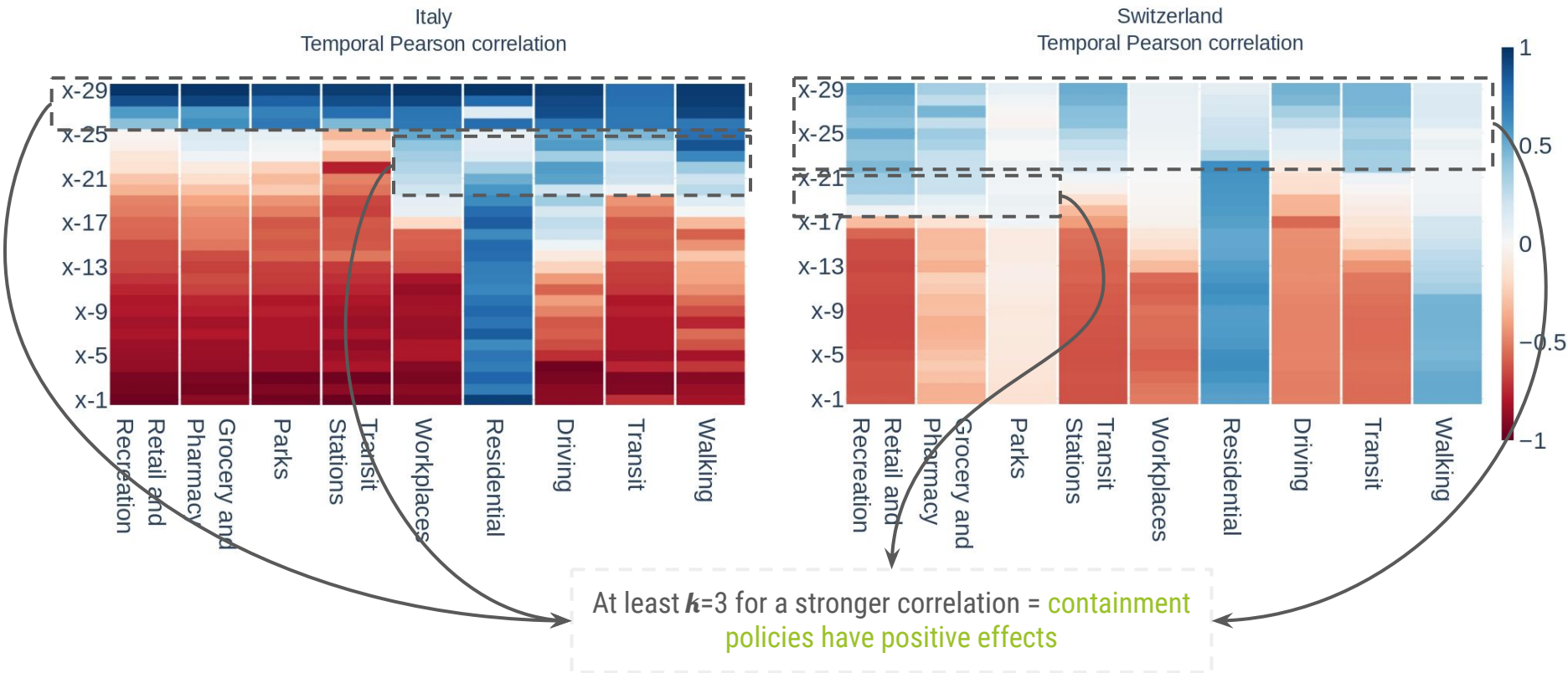
Highest correlation ≥ 0.5

Results (F2)



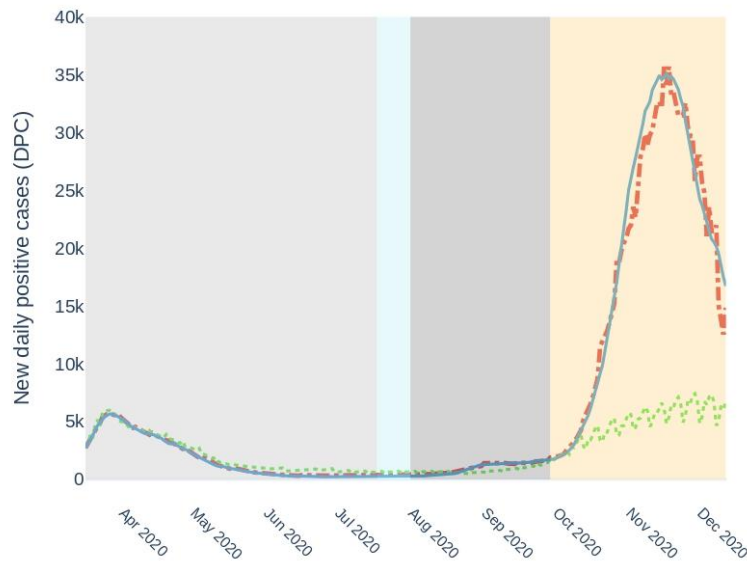
Low traffic in residential areas =
people staying home

Results (F2)

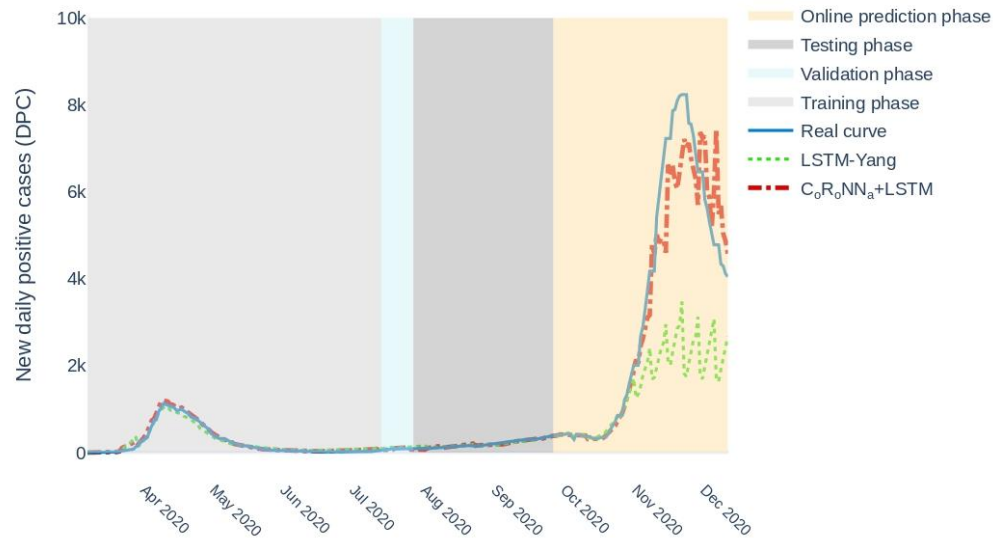


Results (Comparison of methods)

Italy

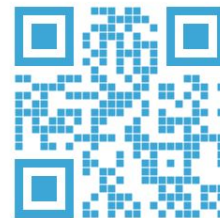


Switzerland



Conclusions

- Diffusion models are **less effective** at predicting the viral spread of infections
 - Epidemiological parameters have different manifestations from individual to individual
- CoRoNNa is **flexible** and makes far fewer assumptions on input variables
- Changes in **mobility data** are good predictors of future infections
 - Stronger correlation with daily positive cases at longer lags
 - CoRoNNa shows superior ability at capturing **long-term dependencies**



THANK YOU

MIUR under the grant of “Dipartimenti di eccellenza 2018-2022” of the Department of Computer Science of Sapienza University

Bibliography

- [1] Yan Bai, Lingsheng Yao, Tao Wei, Fei Tian, Dong-Yan Jin, Lijuan Chen, and Meiyun Wang. 2020. Presumed asymptomatic carrier transmission of COVID-19. *Jama* 323, 14 (2020), 1406–1407.
- [2] A Kucharski, T Russell, C Diamond, Y Liu, J Edmunds, S Funk, and R Eggo. 2020. Analysis and projections of transmission dynamics of nCoV in Wuhan. *CMMID repository* 2 (2020).
- [3] Ashleigh R Tuite and David N Fisman. 2020. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. *Annals of Internal Medicine* 172, 8 (2020), 567–568.
- [4] Joseph T Wu, Kathy Leung, and Gabriel M Leung. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 395, 10225 (2020), 689–697.
- [5] S. Zhao, S. S Musa, Q. Lin, J. Ran, G. Yang, W. Wang, Y. Lou, L. Yang, D. Gao, D. He, et al. 2020. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak. *Journal of clinical medicine* 9, 2 (2020), 388.
- [6] Ying Zhou, Renzhe Xu, Dongsheng Hu, Yang Yue, Qingquan Li, and Jizhe Xia. 2020. Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. *The Lancet Digital Health* 2, 8 (2020), e417–e424.

Bibliography

- [7] Zeynep Ceylan. 2020. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment* (2020), 138817.
- [8] Tanujit Chakraborty and Indrajit Ghosh. 2020. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons & Fractals* (2020), 109850.
- [9] Haytham H Elmousalami and Aboul Ella Hassanien. 2020. Day level forecasting for Coronavirus Disease (COVID-19) spread: analysis, modeling and recommendations. *arXiv preprint arXiv:2003.07778* (2020).
- [10] Fotios Petropoulos and Spyros Makridakis. 2020. Forecasting the novel coronavirus COVID-19. *PloS one* 15, 3 (2020), e0231236.
- [11] Vinay Kumar Reddy Chimmula and Lei Zhang. 2020. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals* (2020), 109864.
- [12] Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, et al. 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease* 12, 3 (2020), 165.