2024/05/16

Dynamic causal modeling of the COVID-19 pandemic in northern Italy predicts possible scenarios for the second wave

Daniela G., Giuseppe P., Tommaso F., Alessia G., Marco V., Egidio D.A., Jonathan M.

Frontiers in Public Health, 2020

Introduction: Background

- A model is necessary to identify the factors underlying the dynamics of the outbreak.
- A time series generative model and Bayesian model inversion are considered for parameter estimation.
- DCM was initially designed to infer the nature of connectivity in brain networks and is effective in various non-linear dynamical systems where different causal sources interact in complex ways.

- The case study is based on Northern Italian regions.

Northern Italian regions were more affected than others, with variable severity raising several unanswered questions.

- **How did the tight lockdown impact the local dynamics of Covid spread?**
- **Will there be a second wave? If yes, how and when should we should expect it?**
- **Will a further lockdown be necessary in case of a second wave, or will efficient testing and other confinement strategies be sufficient?**

Aims of DCM application

- Verifying the predictive ability of the model.

- Estimating the duration of immunity and its profound impact on the expected onsets of second waves.

(Bayesian model comparison is based on 32 models with different immunities as priors)

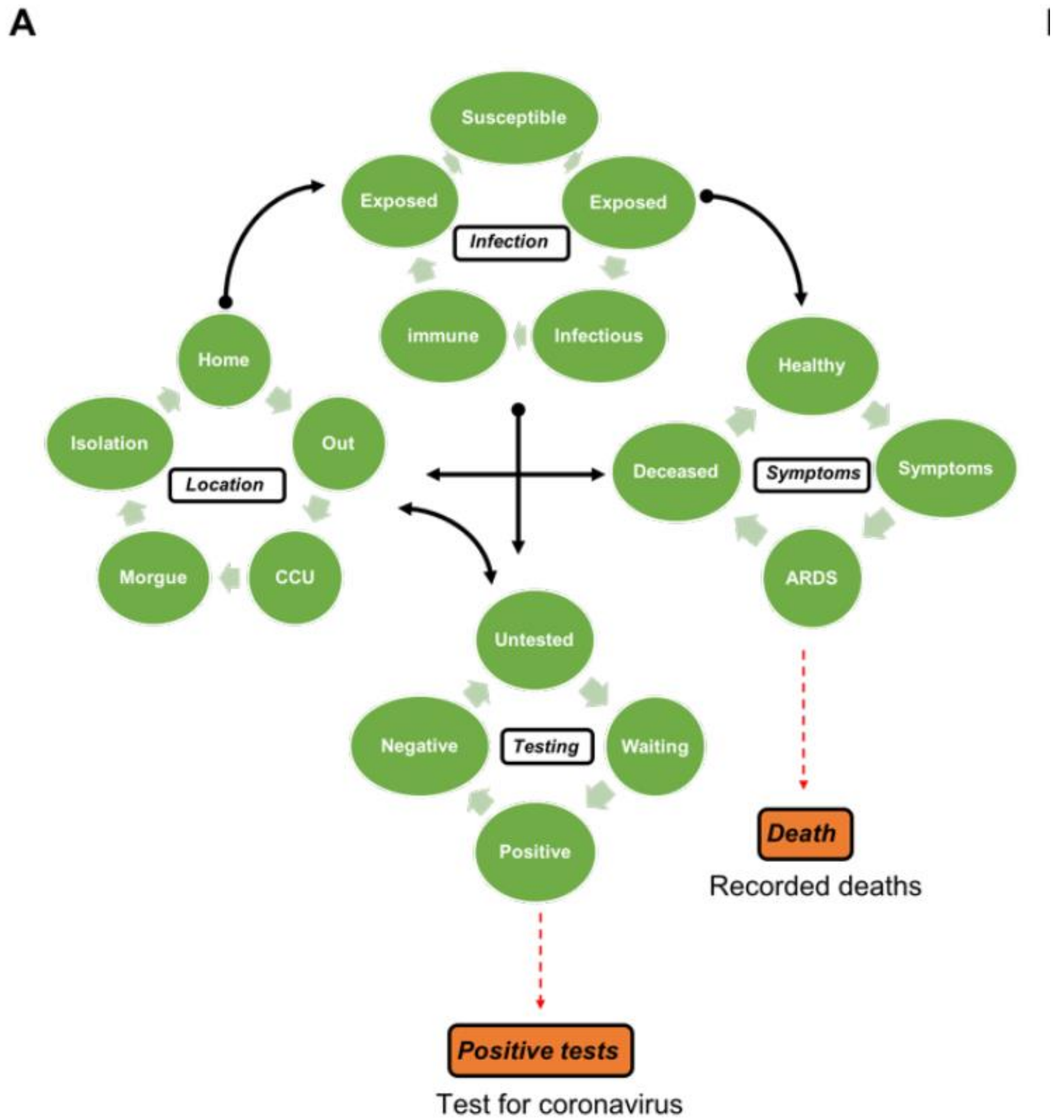
- Testing and tracking strategies

- Evaluating the effect of varying degrees of efficacy of the testing and tracking strategy on second-wave outbreaks

Methodology

- Four factors have been considered: location, infection, symptoms, and testing state.
- These factors are coupled through probabilistic transitions specified by state probability transition matrices.
- Parameters are initialized with a priori expectations and variances.
- The DCM model inversion maps from the observed data to the estimated parameter values using gradient ascent until the marginal likelihood of the data is maximized.

Generative Model



Generative Model

$$P = \theta_{out}(1 - p_{infected}^{inf})^{\theta_{sde}}$$

$$Q = \sigma(p_{CCU}^{loc}, \theta_{cap})$$

rows

columns

- Home
- Work
- CCU
- Morgue

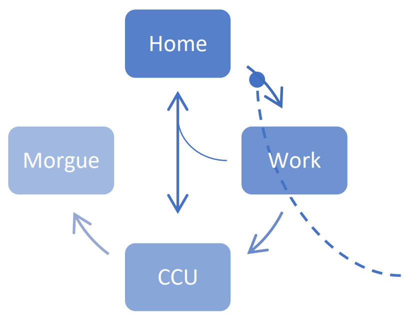
$$P\left(loc_{t+1} \mid loc_t, clin_t = asymptomatic\right) = \begin{bmatrix} 1-P & 1 & 1 & \\ P & & & \\ & & & \\ & & & 1 \end{bmatrix}$$

$$P\left(loc_{t+1} \mid loc_t, clin_t = symptomatic\right) = \begin{bmatrix} 1 & 1 & 1 & \\ & & & \\ & & & \\ & & & 1 \end{bmatrix}$$

$$P\left(loc_{t+1} \mid loc_t, clin_t = ARDS\right) = \begin{bmatrix} 1-Q & 1-Q & & \\ & & & \\ Q & Q & 1 & \\ & & & 1 \end{bmatrix}$$

$$P\left(loc_{t+1} \mid loc_t, clin_t = deceased\right) = \begin{bmatrix} & & & \\ & & & \\ & & & \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

(1) location



$$p_{t+1} = T(\theta, p_t)\, p_t$$

(Friston, K.J. et al.,2020)

Observation Model

Likelihood

$$P(O \mid \vartheta) = P(O_0, \ldots, O_T \mid \vartheta) = \prod_0^T N(\sqrt{n\pi_t}, I)$$

$$n\pi_t = (10^6 \cdot \theta_N)\boxed{(p_{deceased,t}^{clin} - p_{deceased,t-1}^{clin})}$$

$$p_{t+1} = T(\vartheta, p_t)p_t$$

Variational free energy

$$F = \overbrace{E_Q[\ln P(O \mid \vartheta)]}^{accuracy} - \boxed{\overbrace{D_{KL}[Q(\vartheta) \| P(\vartheta)]}^{complexity}}$$

Kulback - Leibler divergence

(Friston, K.J. et al.,2020)

Data collection

- **Daily pandemic figures**
- Daily number of confirmed positive cases, deaths, and recovered cases from January 22nd to August 09th.
- This dataset was split into two subsets.

(dataset 1: from January 22nd to July 20th; dataset 2: from January 22nd to August 09th)

To accommodate differences in the social and movement patterns before and after the end of July, a period when most of the Italian population leaves home for the summer vacations.

- **Cellphone-based estimates of daily movements**

Daily individual movements have been collected anonymously via mobile cellphone networks (to assess the ability of the DCM to infer the probability of leaving home)

- **Serological tests data**

antibody-positive individuals in Italy

Results: Posterior distribution

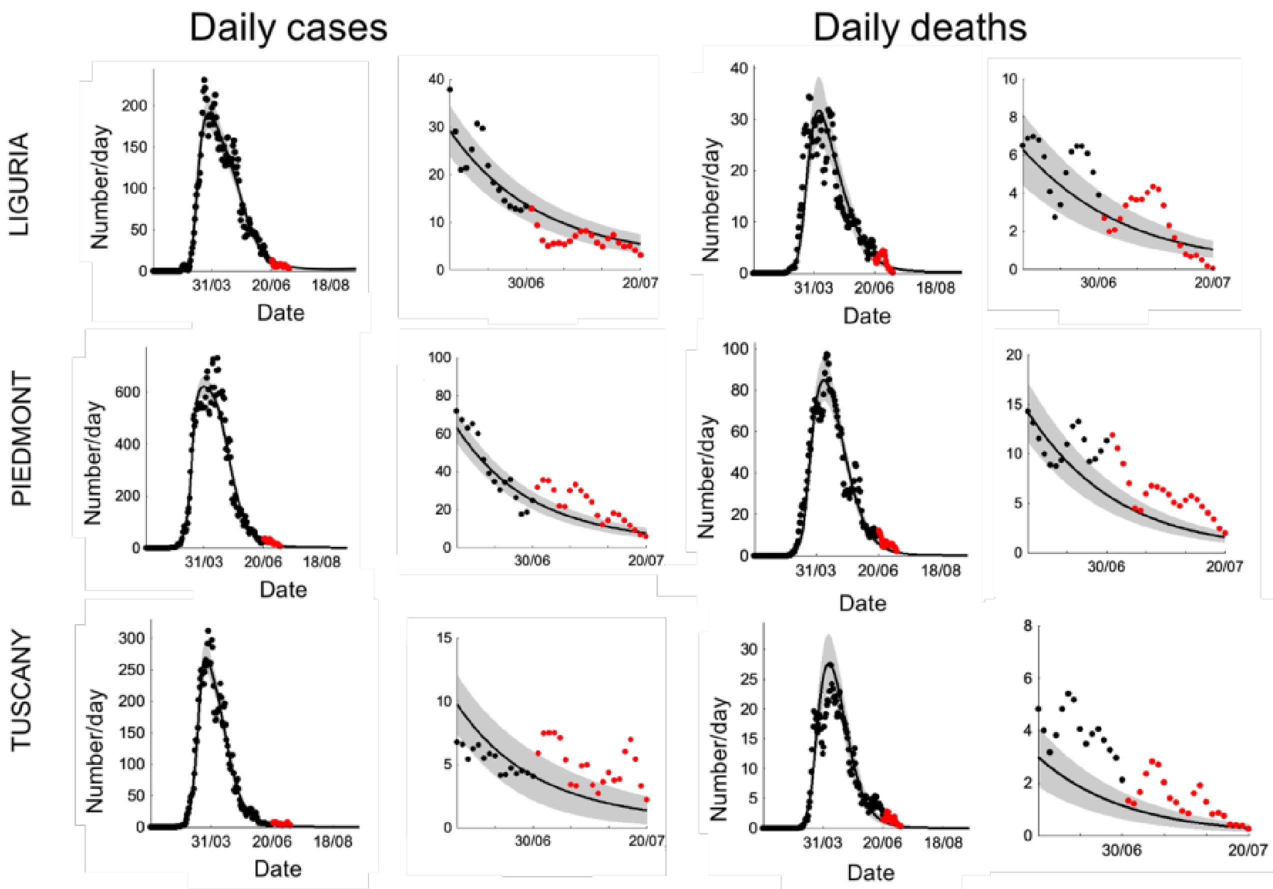
Data usage

Parameter estimation – from January 22nd to June 30th
Model validation – from June 30th to July 20th

FACTORS	PARAMETERS	LOMBARDY	VENETO	EMILIA - ROMAGNA	LIGURIA	PIEDMONT	TUSCANY
Location parameters	Prob going out from home (%)	32.81	35.35	33.46	33.06	33.37	33.07
	Social distancing threshold	0.0244	0.0297	0.0340	0.0302	0.0319	0.0265
	Bed availability threshold (per capita)	0.0001671	0.0001667	0.000165	0.000167	0.000168	0.000168
Infection parameters	Effective number of contacts being at home	5.92	3.18	5.55	6.03	6.13	5.99
	Effective number of contacts being out	51.58	99.78	55.39	42.05	45.44	54.34
	Probability of getting contagion for each contact (%)	0.52	0.56	0.53	0.44	0.48	0.57
	Infected pre contagious period (days)	3.94	3.35	3.80	3.71	3.66	3.21
	Infected contagious period (days)	3.43	3.65	4.16	3.82	3.89	3.48
Clinical parameters	Time till symptoms (days)	16.71	19.61	18.35	17.89	18.81	21.41
	Prob severe symptoms from symptomatic conditions (%)	2.87	2.56	2.79	2.76	2.66	2.26
	Symptomatic period (days)	8.68	9.57	8.95	8.85	9.29	10.38
	CCU period (days)	14.82	15.68	14.62	12.61	16.85	13.72
	Prob of death from CCU with severe symptoms (%)	50.91	44.06	48.32	49.93	50.25	39.09
	Prob of survival from home with severe symptoms (%)	12.61	12.56	12.59	12.65	12.71	12.56
Testing parameters	Test, track and trace	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	Test delays (days)	1.15	1.10	1.13	1.44	1.32	1.24
	Test selectivity (for infection)	11.38	3.36	9.05	8.72	10.56	4.03
	Baseline testing	0.029	0.074	0.064	0.057	0.023	0.034
Immunity	Proportion of resistant cases (%)	41.86	73.24	64.76	52.46	57.56	65.35
	Proportion of people with innate immunity (%)	57.21	44.80	46.14	48.71	48.85	45.37

Results: Model validation and predictive validity

- Inaccurate predictions are most likely due to isolated local outbreaks typically occurring during vacation periods, people returning from foreign countries after vacations, or migrants' fluxes

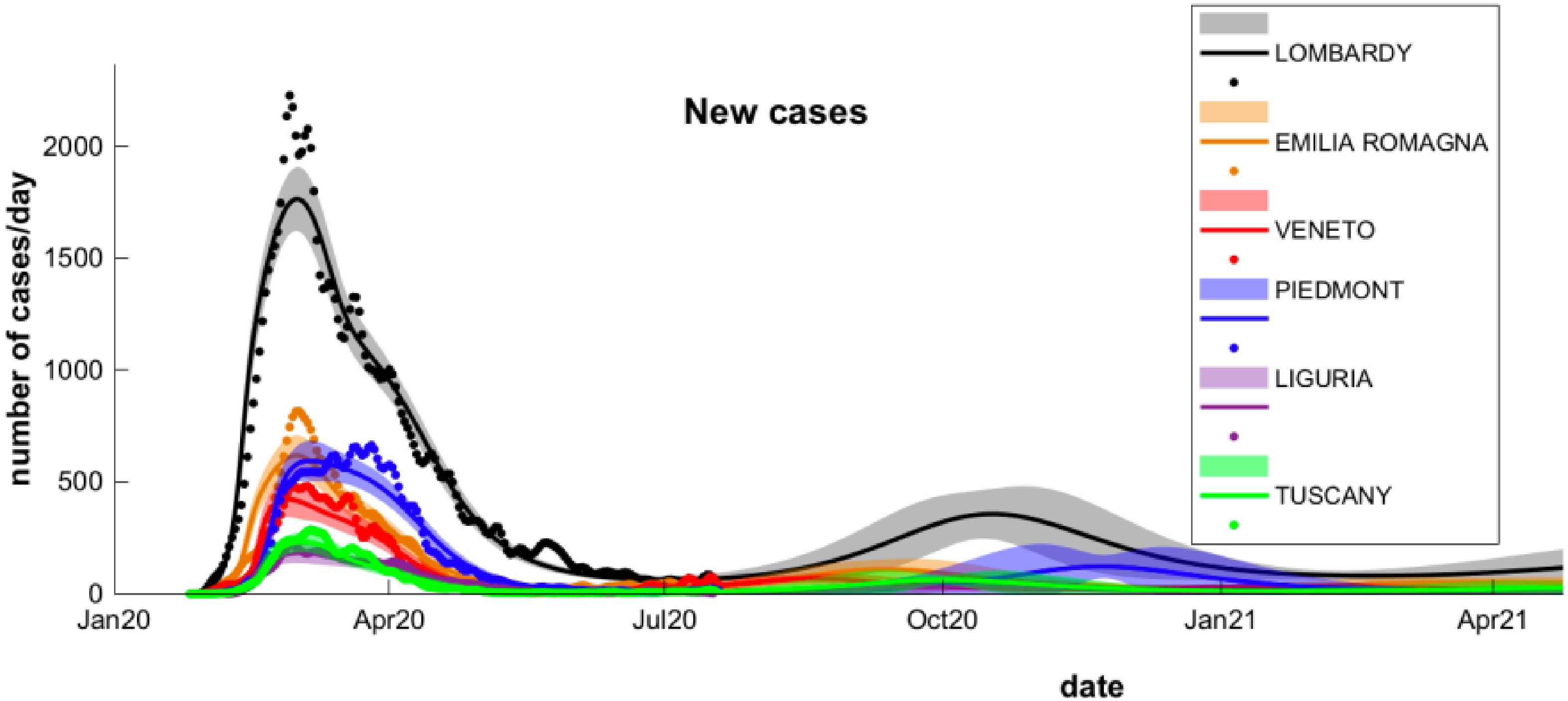


Estimating the duration of immunity

- Bayesian model comparison has been done by considering 32 models.
- Each model differed in its prior assumption about the immunity duration, which varied in monthly increments from 1 to 32 months.
- The model evidence for each of the 32 models was pooled over the six Italian regions under consideration
- This estimated output has been considered for forecasting the second wave in various Italian regions.
- The best estimate has been obtained as 7 months.

Results-Estimation of the duration of immunity and second-wave forecast

The Bayesian model comparison procedure, within a set of models with varying periods of immunity (PIL) from 1 to 32 months (see Methods), yielded a PIL best estimate of 7 month



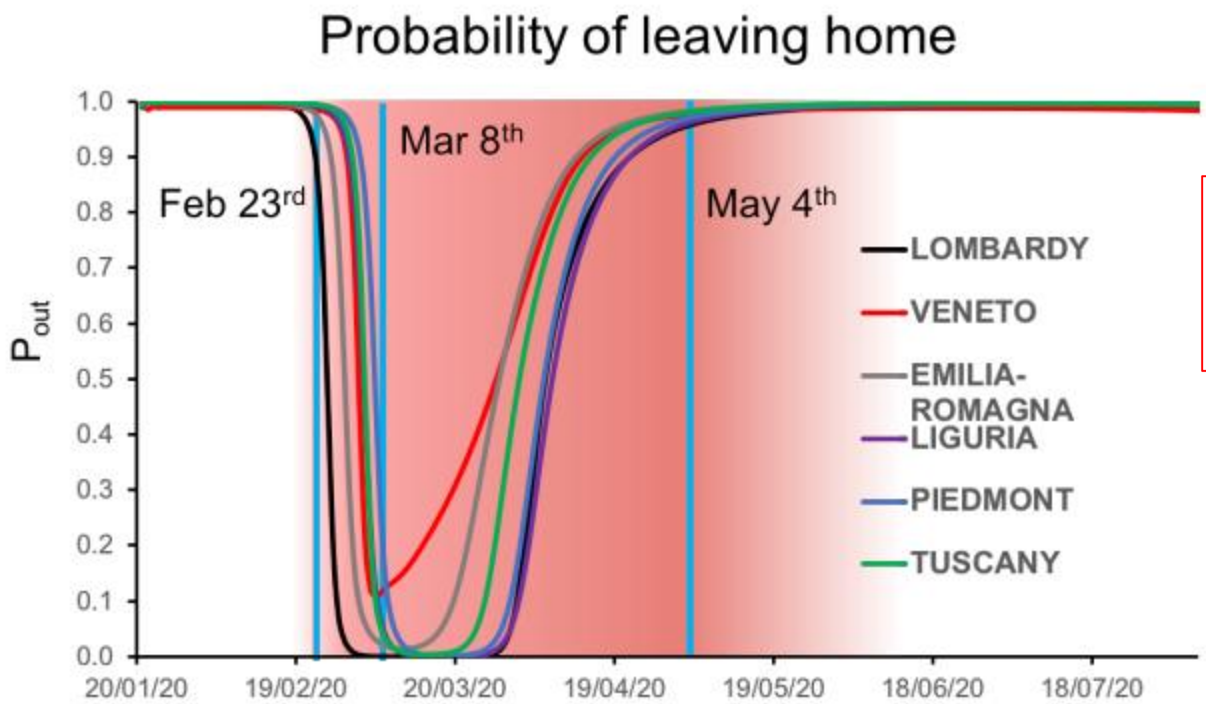
Latent causes estimated by the model

- The probability of leaving home (Location factor) - prior baseline rate of probability of going out, proportion of infected people
- The proportion of infected people (Infection factor) - number of social contacts, proportion of time spent at home
- The proportion of immune people (Infection factor)
- The proportion of resistant people (Infection factor)
- The proportion of people showing symptoms (Symptoms factor)

$$p_{t+1} = T(\theta, p_t) p_t$$

Results – latent causes

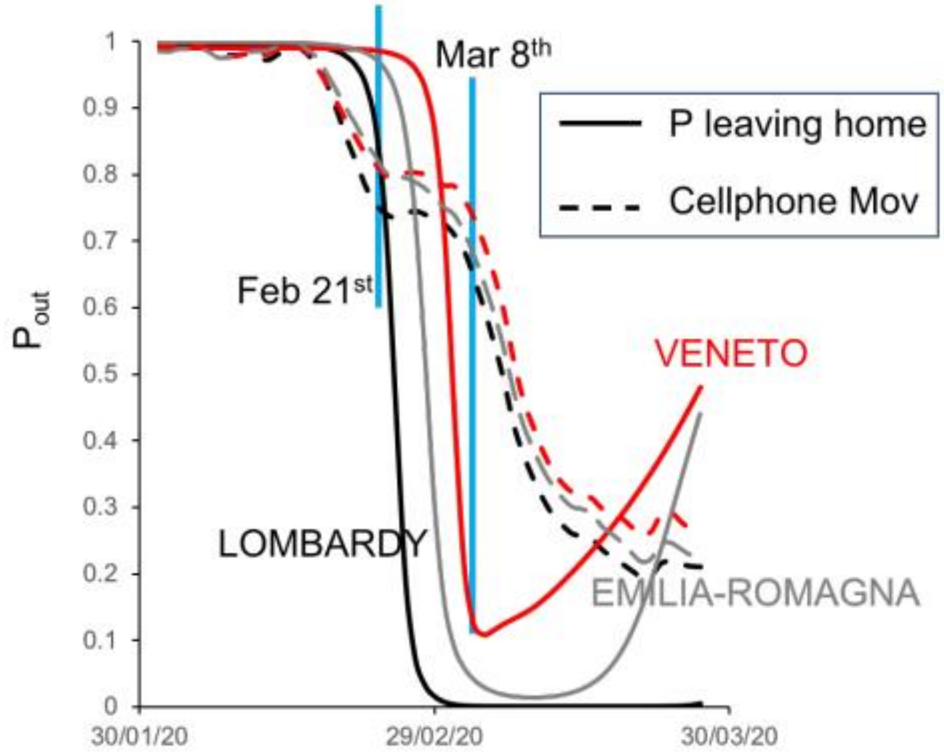
❖ The model was able to estimate many hidden parameters that underwrite epidemic dispersion



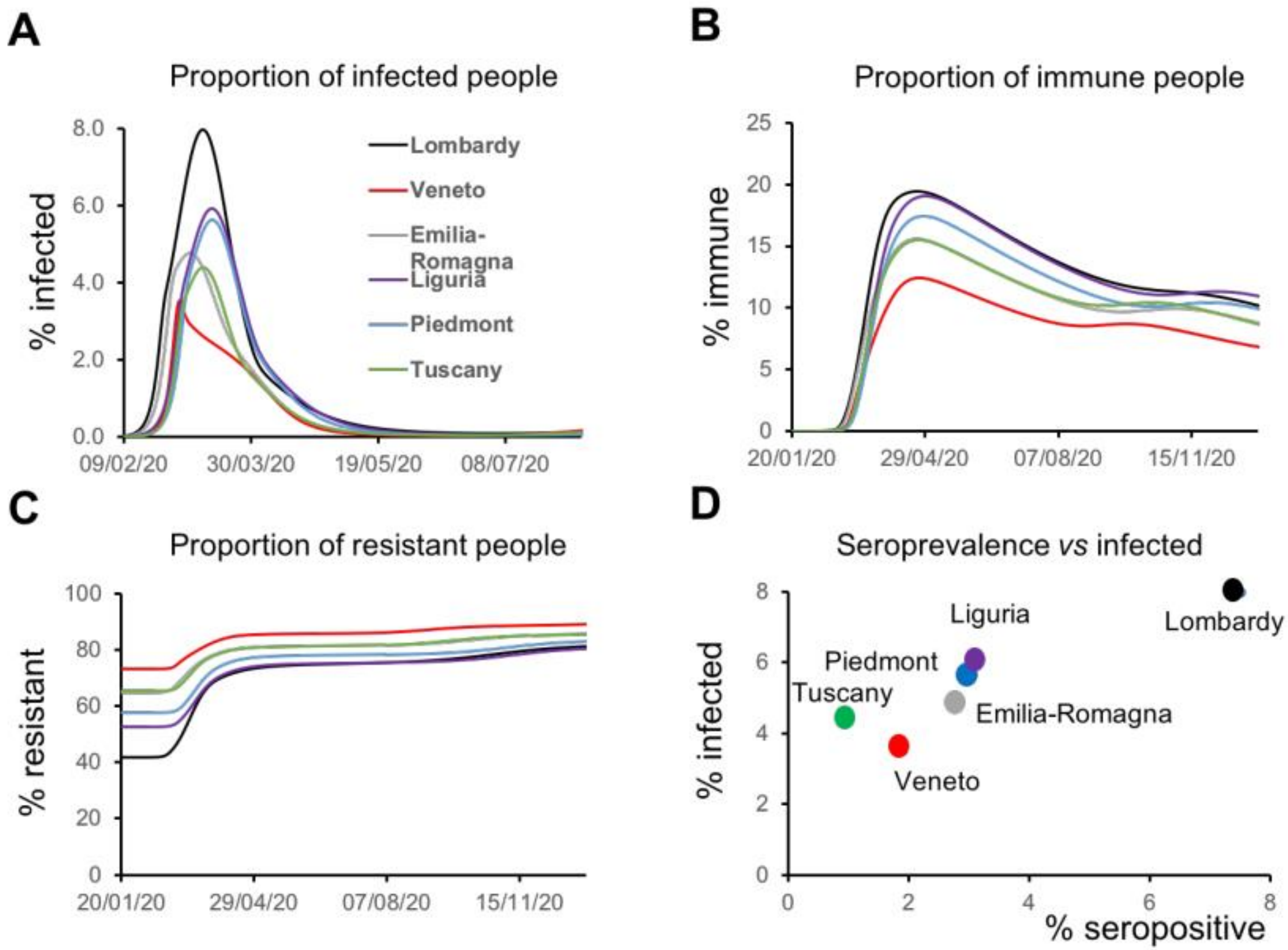
$$P = \theta_{out}(1 - p_{infected}^{inf})^{\theta_{sde}}$$

Results - Latent causes

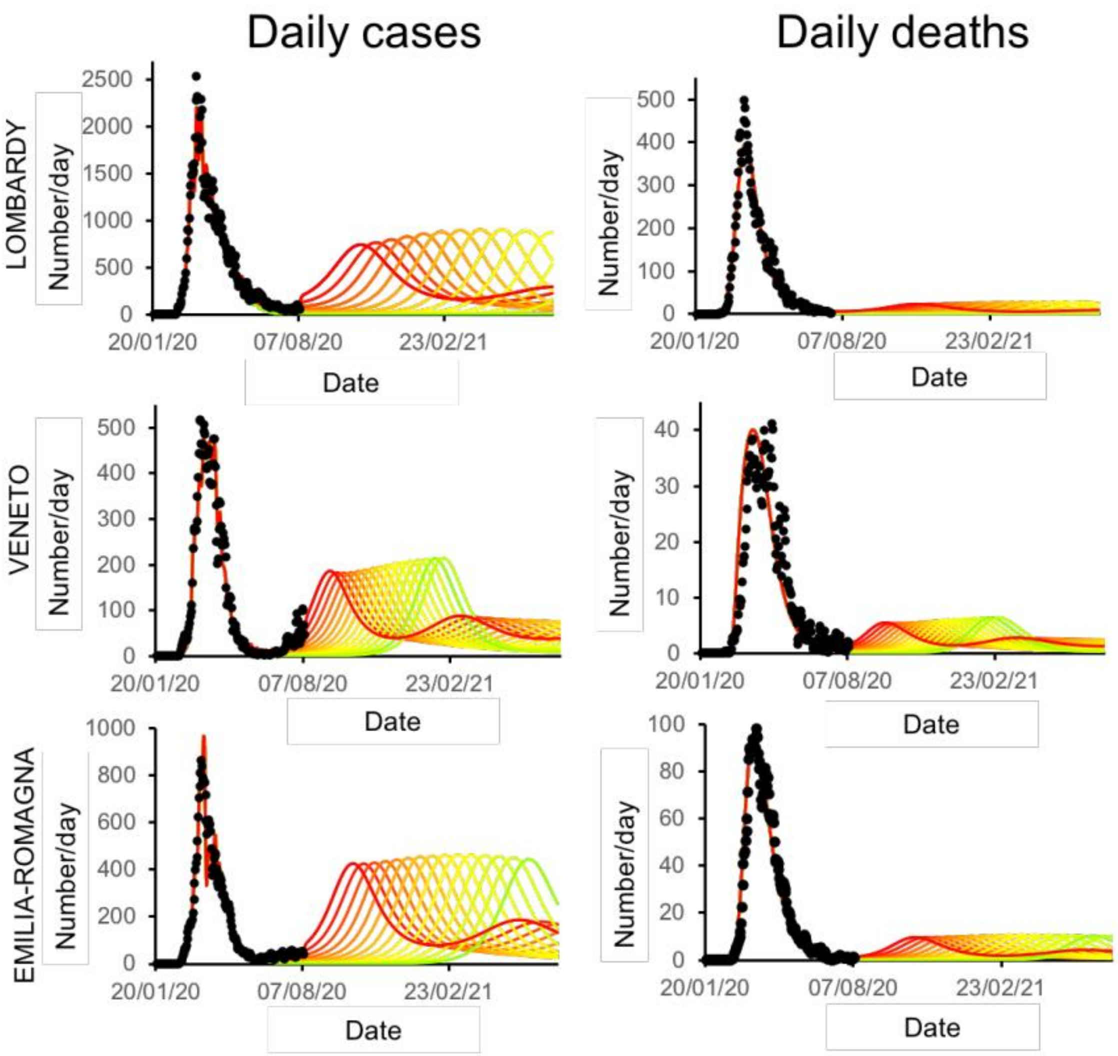
Comparison with collected cell phone movements



Results - Latent causes



Results - Second wave forecasts and future scenarios



Model limitations

- The DCM does not consider interactions with seasonal flu or other annual fluctuations.
- As with other modeling approaches, the outcomes of the Bayesian model comparison and posterior inferences are strictly model-dependent.
- The model does not include geospatial aspects, but rather, each outbreak is treated as a point process

Conclusion

- The model predicts that a second wave could arise due to a loss of effective immunity after about seven months.
- The DCM proposed by Friston et al. was initially developed to investigate brain functions and their innovative consideration in the predictions of the COVID-19 pandemic.
- Retrospective analysis of the pandemic’s hidden factors could furnish an innovative epidemiological tool.
- The DCM approach enables the evaluation of the role of factors that can be manipulated by institutional and healthcare policies

Research application

<u>Analysis of Spatio-Temporal Dynamics of Urban Growth by integrating Satellite Image Processing in Residential Choice Modeling using Dynamic Causal Structure</u>

- Preparation of Land use variables using satellite image processing.
- Development of Dynamic causal model and estimation using Bayesian inference. (Variational inference / MCMC sampling)
- Development of Discounted recursive logit model.
- Consideration of Model comparison between some disaster-prone countries / different periods / multiple disasters.
- Model Comparison between modified DCM location choice model with traditional RL model.

Thank you

Research plan

Analysis of Spatio-Temporal Dynamics of Urban Growth by integrating Satellite Image Processing in Residential Choice Modeling using Dynamic Causal Structure

Background: Due to the lack of behavioral basis models, significant obstacles remain in realistic demand forecasting and the necessity of an adequate body of research with advanced methodologies on dynamic prediction has been raised with time.

Purpose: To provide an effective investment plan for cities based on dynamic mobility after disasters

$$\dot{z} = (A + \sum_j u_j B^j) z + Cu$$

(Friston, J.K et al, 2003)

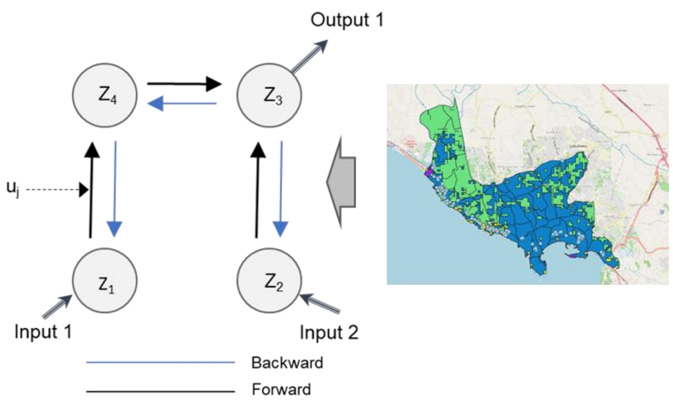


Figure 1: Model outline

Method: This model framework consists of two models including the observation model which gives the land use variations using satellite image processing. The next model is the location choice model, which is developed based on the **Dynamic Causal Model** (DCM) structure.

1. Land use variables are obtained using **satellite image** processing.
2. Parameters are obtained using **Maximum likelihood** estimation and **Bayesian estimation** for different DCM structures.
3. Analysis continues for various periods/ multiple disasters/ different countries
4. Model comparison is done between the traditional RL and DCM to measure the effect.