

A APPENDIX

A.1 META-DATA

To incorporate epidemiological and geographical information into early-stage pandemic forecasting, we collected 7 epidemiological and 13 geographical metadata for each location and pandemic. Data unavailable during the early stages were marked as missing in the dataset. A detailed list of the metadata collected is provided in Table A.1.

Table A.1: **Meta data table for training HG-DCM** The meta-data table including the epidemiological metadata and geological metadata used in training HG-DCM

Epidemiological Meta Data	Geological Meta Data
Basic Reproduction Rate (R0)	Population
Transmission Pathways	Net lending/borrowing
Mortality Rate	Current Health Expenditure per capita
Average Hospitalization Length	Population Density
Hospitalization Rate	GNI per capita
Latent Period	GDP per capita
Incubation Period	Physician per 1,000 people
	Urban Population Living in Slums
	GDP
	External Health Expenditure per capita
	Air Transport
	Government Effectiveness
	Domestic General Government Health Expenditure per Capita

A.2 PARAMETER RANGES

To fit an accurate and interpretable model, it is crucial that the predicted parameters in both HG-DCM and DELPHI fall within epidemiologically reasonable ranges. For this study, we adopted the parameter ranges defined in the original DELPHI paper (Li et al.), with adjustments tailored specifically for early-stage forecasting. A comprehensive list of the parameter ranges used during model training is presented in Table A.2.

Table A.2: **DELPHI Parameters Ranges** The ranges of DELPHI parameters used in training HG-DCM and DELPHI

DELPHI Parameters	Range
α	[0.1, 1.5]
days	[0, 100]
r_s	[1, 15]
r_dth	[0.02, 0.5]
p_dth	[0.01, 0.25]
r_dthdecay	[-0.2, 5.0]
k1	[0.001, 5]
k2	[0.001, 5]
jump	[0, 5]
t_jump	[0, 300]
std_normal	[0.1, 100]
k3	[0.2, 2.0]

A.3 COVID-19 EARLY-STAGE FORECASTING PARAMETER ANALYSIS

To better interpret the predictions, we analyzed the parameters inferred from HG-DCM compared to DELPHI in an early-stage COVID-19 forecasting task using four weeks of data. All 12 parameters

are significantly different between DELPHI and HG-DCM based on Wilcoxon signed-rank test ($p < 0.05$). HG-DCM model tends to predict a lower α , $days$, r_dth , t_jump , p_ath , $k1$, and std_normal , whereas predicting a higher r_s , $r_athdecay$, $k2$, $jump$, t_jump , and $k3$. The divergent prediction set of parameters from two different forecasting methods provides a more comprehensive understanding of the pandemic. (Figure A.1).

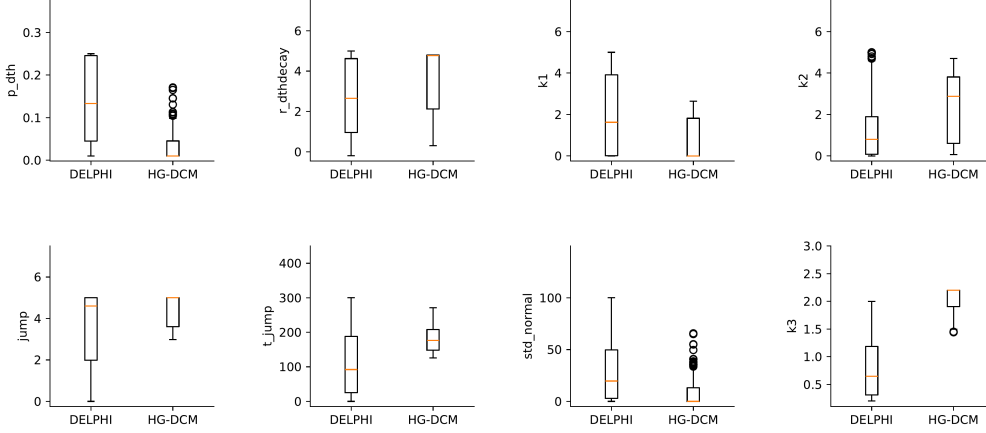


Figure A.1: **Comparison of fitted parameters in traditional DELPHI model and HG-DCM** The 8 bar graphs show the mean and standard deviation of the remaining 8 predicted parameters from two different approaches that are not shown in the main text. Mann-Whitney U test is used to calculate the p-value of the two groups. Pairs with p-values < 0.05 are considered significantly different.