Applied Geography 154 (2023) 102941

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Seeing the forest and the trees: Holistic view of social distancing on the spread of COVID-19 in China

Danlin Yu [a,*,1], Yaojun Zhang [b,**,1], Jun Meng [c,1], Xiaoxi Wang [d], Linfeng He [d], Meng Jia [d], Jie Ouyang [d], Yu Han [d], Ge Zhang [e], Yao Lu [f]

[a] Department of Earth and Environmental Studies, Montclair State University, Montclair, NJ, 07043, USA
[b] School of Applied Economics, Renmin University of China, Beijing, 100086, China
[c] Department of Obs.&Gyn., Beijing Youan Hospital, Capital Medical University, Beijing, 100069, China
[d] School of Sociology and Population Studies, Renmin University of China, Beijing, 100086, China
[e] School of Management, Minzu University of China, Beijing, 100081, China
[f] School of Ethnology and Sociology, Minzu University of China, Beijing, 100081, China

A R T I C L E I N F O

Handling Editor: J Peng

Keywords:
COVID-19
Global and local spatial analysis
Geographically weighted panel regression
Social distancing
China

A B S T R A C T

The human social and behavioral activities play significant roles in the spread of COVID-19. Social-distancing centered non-pharmaceutical interventions (NPIs) are the best strategies to curb the spread of COVID-19 prior to an effective pharmaceutical or vaccine solution. This study investigates various social-distancing measures' impact on the spread of COVID-19 using advanced global and novel local geospatial techniques. Social distancing measures are acquired through website analysis, document text analysis, and other big data extraction strategies. A spatial panel regression model and a newly proposed geographically weighted panel regression model are applied to investigate the global and local relationships between the spread of COVID-19 and the various social distancing measures. Results from the combined global and local analyses confirm the effectiveness of NPI strategies to curb the spread of COVID-19. While global level strategies allow a nation to implement social distancing measures immediately at the beginning to minimize the impact of the disease, local level strategies fine tune such measures based on different times and places to provide targeted implementation to balance conflicting demands during the pandemic. The local level analysis further suggests that implementing different NPI strategies in different locations might allow us to battle unknown global pandemic more efficiently.

1. Introduction

The outbreak and rapid spread of COVID-19 in the early days of 2020 in China stricken the world with unprecedented damage. During the initial periods of the outbreak, China faced with the dilemma of containing the spread of the disease with little knowledge of how. Epidemiological practices suggest that effective non-pharmaceutical intervention (NPI) strategies that center on social distancing and early quarantine are critical to curb the spread of an unknown, infectious disease (Davies et al., 2020; Dayaratna, Gonshorowski, & Kolesar, 2022; Fair, Karatayev, Anand, & Bauch, 2022; Flaxman et al., 2020; Nazia, Law, & Butt, 2022; Redlin, 2022). A recent study on how social distancing impacts the spread of influenza pandemics (Fong et al., 2020) found that social distancing, including the isolation of the infected, active contact tracing and quarantine, prevention of public gathering including school and workplace closures helped reduce the rate of transmission and flatten the curve of the spread. Social distancing allows the disease to spread at a lower rate over a longer period, which enables the health system to respond in a more efficient way and gives time for pharmaceutical solutions to be devised to combat the disease. An investigation of the social distancing measures during the first 50 days of the pandemic in China suggests that intra-city public transportation ban,

* Corresponding author.
** Corresponding author.
E-mail addresses: yud@mail.montclair.edu (D. Yu), zhyaojun@ruc.edu.cn (Y. Zhang), mengjun_98@yahoo.com (J. Meng), wangxiaoxi2016@ruc.edu.cn (X. Wang), helingfeng@ruc.edu.cn (L. He), jiameng0209@ruc.edu.cn (M. Jia), jie.ouyang@ruc.edu.cn (J. Ouyang), hyzxshyx@ruc.edu.cn (Y. Han), gezhang@muc.edu.cn (G. Zhang), aria_ly0118@muc.edu.cn (Y. Lu).
[1] The authors contribute equally to this work.

https://doi.org/10.1016/j.apgeog.2023.102941
Received 31 March 2022; Received in revised form 21 January 2023; Accepted 18 March 2023
Available online 27 March 2023

large gathering ban successfully delayed the spread of the disease (Tian et al., 2020).

Still, the real situation, as the pandemic stretches, was more complex because the spread rate of COVID-19 and what might work better to curb its spread might not be the same everywhere as witnessed in many studies in China (Kraemer et al., 2020; Xiong, Wang, Chen, & Zhu, 2020), UK (Davies et al., 2020), Italy (Ciufolini & Paolozzi, 2020; Giordano et al.), Germany (Chae & Park, 2020), the US (Auger et al., 2020). It is hence critical to consider the effectiveness of different social distance based NPI policies at different times and places from the location specific perspective. Different locations have different speed of responsiveness, different policies regarding a never-before-seen disease, different natural and infrastructural conditions, but more importantly, different human social and behavioral responses in dealing with a contagious disease that are often not only related to governmental policies but also individual knowledge and the general cultural environments (Cevasco et al., 2021; S. Guo et al., 2020; X. Guo, Zhang, & Wu, 2021). Based on studies in the early stages of the COVID-19 pandemic (Bertozzi, Franco, Mohler, Short, & Sledge, 2020; Brauner et al., 2021; Drew et al., 2020; J. Huang et al., 2020; Poletto, Scarpino, & Volz, 2020), accurate estimation and prediction of the spread of the disease on different locations is critical to estimate necessary medical requirement and capacities, enable decision makers to allocate limited medical resources during the initial outbreak, reduce the loss of human lives, and flatten the curve of the spread. Many predictive epidemiological models have been established prior to or during the outbreak of COVID-19 (Ciufolini & Paolozzi, 2020; Dehning et al., 2020; Dowd et al., 2020; Enserink & Kupferschmidt, 2020; S. Guo et al., 2020). The experiences and conclusions from the models are in consensus: reliable short-term prediction is the key to the success of combatting the spread, flattening the curve especially in the initial stages of the spread, regardless of where the prediction was made and how it was made (Baker et al., 2020; Chen et al., 2022; Ciufolini & Paolozzi, 2020; Davies et al., 2020; Dayaratna et al., 2022).

Most such models and predictions, however, investigate the impacts of NPI on the spread of COVID-19 from a global perspective (we define this as the *forest* view). That is to say, the models are built on the premise that NPI strategies will work universally, regardless of locations that have varied degrees of the diseases occurring and spread, and social, economic, demographic, and natural conditions. Results from those global models are valuable because they provide timely and important predictions of what might work and what might not work from a macro perspective, which will enable the country to take immediate actions at the initial stages of the outbreak as China did to save lives and flatten the curve (Chinazzi et al., 2020; Wei & Wang, 2020; Xiong et al., 2020). In addition, under the premise of providing legitimate short-term prediction, oftentimes models need to rely on available information. Many varying conditions regarding the places and times are either unavailable or less reliable to be used for better modeling practices under the global modeling framework. As to whether similar strategies will work or work to the same degree in different locations, the global level models are not equipped to answer these questions. Yet answers to these questions could mean not only actual difference between life and death for individuals in different places during the spread of COVID-19, but also a balance between strict lockdown and meeting the local demands for adequate and flexible socioeconomic activities. Without the detailed knowledge, all decisions might come down to a choice between strict national level lock-down, which carries prohibitively high socioeconomic cost as witnessed in China (Chinazzi et al., 2020), or a half-hearted, hesitant lock-down that might not work at all as the US situation pans out in the past year (S. Guo et al., 2020; Manchein, Brugnago, da Silva, Mendes, & Beims, 2020), and we are now registering over a million lost lives. For this matter, to combat the spread of the disease more efficiently and save human lives at various locations, models need to also look at the individual "*trees*" (individual locations) that might have different relationships between the mitigation strategies and the spread of the COVID-19 disease. This "*trees*" view provides a micro perspective of the modeled relationship and might provide a critical supplement for the holistic view for policy implementation and decision-making.

The current investigation explores the *forest* view from a spatial analytical perspective as some have attempted (X. Guo et al., 2021; Xiong et al., 2020), but focuses on the *trees* view to establish a new modeling strategy and new modeling paradigm that can strike a balance between better saving lives from the pandemic and meeting the demands for other socioeconomic needs. We envision the study contributes to the applied geographic research community by proposing a research paradigm that starts with a global vision, delves deeper with local variations, couples advanced global with novel local spatiotemporal analytical models to provide a ***holistic*** view of relationships between stimulus and outcomes by considering the spatiotemporal effects that are embedded in the collected data in the applied geographical communities. This study fills a gap in empirical spatial epidemiological studies that emphasizes on pragmatic and actionable local strategies. In addition, this study promotes the spatial holistic research paradigm to explore the effectiveness of various NPI strategies and how such strategies prevent the spread of COVID-19 at both the macro and micro levels. The spatial holistic research paradigm suggests that "seeing both the forest and the trees" contributes to the growing spatial data analysis knowledge base of spatial epidemiology studies. The research is expected to supply the governments more efficient strategies against future global pandemic, allow the governments to act more responsively and responsibly, flatten the curve of spread, save lives, and enable the society to return to normal earlier.

After this introduction section, we will detail our data and methodology. This is followed by the results from the models in the third section, and discussion of these results in the fourth section. We conclude our study with a summary of the findings of the study.

2. Data and methods

2.1. Measuring social distancing via big data: the social, behavior and policy responses

Social distancing measures vary from study to study. The essence of social distancing, however, is to prevent individual physical closeness as would normally be required under usual social interaction scenarios. The current study extends previous investigation on social distancing measures' effect (Tian et al., 2020) to May and attempts the "*forest and trees*" research paradigm with data mined from non-traditional sources (other than official reporting and records) for China, as the first country that reported the outbreak and the country that implemented one of the strictest social distancing measures to curb the spread of the disease. We chose the dates from the start of the pandemic to early May (May 6th) because on April 8th, Wuhan, the epicenter of the pandemic, finally decided to relax its strict local quarantine lock-down. We did not stop our data collecting points at exactly April 8th, however. Instead, considering many studies suggest that COVID-19 has a 5–8 days of incubation period (Alene et al., 2021; Y. Y. Cai et al., 2020; Cheng et al., 2021; Cimolai, 2021; Guan et al., 2020; Quesada et al., 2021; Wan, Liu, & Liu, 2021; Wu et al., 2022; L. Zhang et al., 2021), and in clinical practices, a 14-day extended quarantine period is often adopted in the policy sphere (Charvadeh, Yi, Bian, & He, 2022), we extended our data collection period to the early May (May 6th) until one of the data entry, the Baidu migration index, is no longer available.

With the help of search engine, website text and policy analyses, the study team identified eight proxies that could roughly represent the many different social distancing measures in China during the outbreak of the pandemic. These proxies include the intensity of within city travel (*travelcity*), inter-city travel ban (*intercity*, 1 for ban, 0 not), the intensity of daily migration (*addmig*), the daily net change of the number of individuals for each prefecture (*netmin*), school closures (*school*, 1 for closure, 0 not), workplace closures (*work*, 1 for closure, 0 not), the number of tourist sites comments (*tourists*), and quarantine status (*quarantine*, 1 for full prefecture lock down, 0 otherwise). All data are acquired daily from January 11th, 2020, to May 6th, 2020, the period when it is closest to the mandatory quarantine of Wuhan City on January 23rd, 2020, and the gradual reopening since April 8th, until one of our data entry, the Baidu migration index was no longer available after May 6th, 2020.

The data used in this study is acquired at the prefecture administrative level. In China, a prefecture is the sub-provincial administrative unit that is composed of a central district and many counties. Counting the four provincial level municipalities (Beijing, Tianjin, Shanghai, and Chongqing) as such prefectures and excluding Hong Kong, Macau, and Taiwan where not all data is available, the study counts 359 analytical units in China. The study uses the number of daily confirmed newly added cases of COVID-19 for each prefecture as the outcome variable. This daily data is readily available from Tencent's COVID-19 real-time reporting website (https://news.qq.com/zt2020/page/feiyan.htm?from=timeline&isappinstalled=0#/?nojump=1). The eight proxies for social distancing measures are acquired through website analysis, policy related text analysis and document analysis. The intensity of within city travel (*travelcity*), the intensity of daily migration for each prefecture (*addmig*), and the daily net change of the number of migrants for each prefecture (*netmin*) are acquired through the Baidu migration website (https://qianxi.baidu.com/#/2020chunyun). A Python code is devised by one of our team members to acquire and organize the three types of data. A team of researchers also combed through news reports and published policy documents from January 11th to May 6th, 2020 via Baidu web searches using the keywords "school closure," (*Fengxiao*) "workplace closure" (*Tinggong*), "quarantine" (*Fengcheng*), or "intercity transportation closure" (*Quji Tingyun*) and the relevant prefecture's name to acquire school closures (*school*, 1 for closure, 0 not), workplace closures (*work*, 1 for closure, 0 not), quarantine status (*quarantine*, 1 for full prefecture lock down, 0 otherwise), and closure of inter-city transportation (*intercity*, 1 for closure, 0 not) for each prefecture, and the time points those measures were enacted and lifted through these searches. This process was done through a Python code for intelligent sematic analysis that picks up the key words, the prefectures' names, and sort the queries according to the time stamp it was posted during our study periods' Baidu Search results. Furthermore, a Python code searching the Dazhong Dianping (www.dianping.com) website for each prefecture's popular tourism sites' daily comments was also devised to collect the number of comments. This data (*tourists*) is used as a proxy for potential travels initiated by individuals without utilizing public transportation. The purpose of social distancing is to prevent close contact among individuals from both global (city) and local (individual) levels. Increased social distancing policies, societal, and individual behaviors that prevent close contact will be the efficient ways to curb the spread of COVID-19. The detailed description of each social distancing measures we have collected data for this study and how it is related with the spread of COVID-19 are as follows.

The *travelcity* variable measures the percentage of the total number of people who attempted any types of travel activities in the total population of a prefecture. More people who are moving about during the pandemic will increase the risk of spreading the disease. This social distancing measure is hypothesized to be positively related with the outcome variable.

The *addmig* variable represents the total number of people that move in or out of a prefecture for any given day. This variable measures the intensity of voluntary movement of individuals even under the strict lock-down policy has been instituted. Higher intensity will exacerbate the spread of COVID-19. This social distancing measure is hypothesized to be positively related with the spread of COVID-19.

The *netmin* variable measures the difference between the numbers of people who moves in and out of a prefecture. Longer term net change of the number of individuals for each prefecture is usually a reflection of differences in place attractivity due to socioeconomic or natural reasons (Yu, Zhang, & Wu, 2020), daily change of the number of individuals, however, might reflect a dynamic of spontaneous movement primarily because of personal reasons (such as visiting friends, doing business, personal shopping trips, among others, especially with the convenience of high-speed rail system) instead of regional socioeconomic differences. The hypothesis is that when more people move in a prefecture, it will be more likely for COVID-19 to spread. This social distancing measure is hypothesized to be positively related with the outcome variable.

The *school*, *work*, *intercity*, and *quarantine* are all binary variables in that 1 represents a positive action that is designated to curb the spread of COVID-19 (so 1 represents closure of schools and workplaces, shutdown of intercity transportation, and enabling prefecture-wide quarantine), and 0 otherwise. These four variables are all hypothesized to be negatively related with the outcome variable.

The *tourists* data attempts to capture individual behaviors that might not follow regional or national policies entirely. A heighted number of *tourists* might suggest the higher probability for the spread of COVID-19. The relationship is hypothesized to be positive with the outcome variable.

Because the relationship between these social distancing measures and the spread of COVID-19 is well defined (Tian et al., 2020), for ensuing regression models, the tests of coefficients' significance are all based on one-way test instead of the commonly reported two-way test. This test applies to both the global and local models (the *forest* and *trees* views).

2.2. The spread of COVID-19 and the indirect inversed normal transformation supported regression model

Daily data of the COVID-19 in China at prefecture administrative level is acquired through Tencent's COVID-19 website. The current study focuses on the daily newly confirmed cases since this item is representative of the spread of the diseases. Studies of the initial cases in Wuhan suggest that COVID-19 has an average incubation period of 5–6 days (Anderson, Heesterbeek, Klinkenberg, & Hollingsworth, 2020; Charvadeh et al., 2022; Guan et al., 2020; Wu et al., 2022). The quarantine practice adopted by many countries requires either mandatory or self-quarantine for a period of 14 days (Guan et al., 2020; Wan et al., 2021). To see how the NPI strategies prevent the spread of COVID-19, we use the 14 days lagged NPI strategies as the intervention input (meaning for any one day's daily newly confirmed cases, the NPI strategies used in the model are the ones implemented 14 days ago). For this matter, our initial analysis date moves from January 11th, 2020, to January 25th, 2020, which is also one day after the traditional Chinese Lunar New Year, when normally the largest inter and inner-city movements are recorded. All the data in their raw format of the 103 days over the 359 prefectures are summarized in Table 1.

To investigate the potential impacts of social distancing on the spread of COVID-19, this study adopts regression models with the daily newly confirmed cases as the outcome, and the social distancing

Table 1
Summary of the raw data.

	Min	1st Quantile	Median	Mean	3rd Quantile	Max
newly added cases	−107	0	0	2.149	0	13,436
travelcity	0.300	2.960	4.212	3.939	4.889	8.878
addmig	0.003	0.345	0.795	1.351	1.588	31.228
netmin	−23.115	−0.087	−0.013	0.000	0.041	6.533
tourist	0	5	12	18.550	25	125
school	0	1	1	0.839	1	1
work	0	0	0	0.267	1	1
quarantine	0	0	1	0.531	1	1
intercity	0	0	0	0.306	1	1

measures as explanatory variables. In addition to these time variant daily COVID-19 cases and social distancing measures, it is important to note that these daily variant data (both cases and responses) are influenced by and associated with the geographically variant but temporally invariant background information. The WHO developed a "social determinants of health" (SDOH) framework (S. Guo et al., 2020; WHO Commission on Social Determinants of Health W. H. O, 2008) that emphasizes the strong relationship between societal background and health outcomes, and relates social and behavioral responses towards potential health risks. The societal backgrounds often include socioeconomic status, health system capability, and infrastructure status. This study adds a demographic status to the SDOH framework since the amount of population in China also plays a crucial role in the spreading of the disease. Based on these four categories of social determinants of health and demographic background information, we collected data of the total population (demographic category), local economic status (prefecture GDP per capita, portion of the secondary industry in GDP, portion of the tertiary industry in GDP, local financial income per capita, local financial expenditure per capita, socioeconomic category), health care system status (the number of hospitals per 10,000 individuals, the number of doctors per 10,000 individuals, the number of hospital beds per 10,000 individuals, health system category), and infrastructure status (the density of roads, and the average road network travel time to the nearest high-speed rail station, infrastructure category). All data are collected from the 2019 China's statistical yearbook or calculated based on the 2019 open street map road network (openstreetmap.org) and high-speed rail information, the most recent such data (Yu, Murakami, et al., 2020; Yu, Zhang, Wu, Li, & Li, 2021).

To evaluate the impact of social distancing on the daily spread of COVID-19, a three-steps indirect inverse-normal transformation (I-INT) strategy (McCaw, Lane, Saxena, Redline, & Lin, 2019) is adopted to preprocess the data. The preprocessed data is then used to model how social distancing measures curb the spread of COVID-19 with both global and local models. Detailed methodology discussions follow.

2.2.1. The indirect inversed normal transformation (INT) method

The I-INT method is recently proposed in a Genome-Wide Association Studies (McCaw et al., 2019) that shows efficient and unbiased properties when handling many tie-values and skewed variables because of zero-inflation. Our COVID-19 caseload data (lagged daily newly added cases) has many tie-values, and is zero-inflated. The raw data by no means follows a normal distribution and using the raw data for regression estimation will not be tenable. In addition, while we have collected daily changing caseload information and NPI strategies data through web searches and mining, we have also collected some socioeconomic background data for the prefectures as outlined in the data section. To investigate the relationship between the spread of COVID-19 and the social distancing measures, we intend to not only consider the association between the NPI social distancing measures and the spread of COVID-19, but also the socioeconomic background information. While the spread of COVID-19 and the social distancing measures changed daily (temporally variant), the socioeconomic background information stayed constant on a daily interval (temporally invariant). To incorporate both the temporally variant and invariant information to the modeling scheme, the indirect-INT method takes three steps.

In the first step, separately for each time point $t \in \{1, \dots, T\}$, regress the outcome variable y_{it} (lagged daily newly added cases) on the time-invariant covariates w_i (the socioeconomic background information) to obtain the residuals ε_{it}.

In the second step, conduct INT on the residuals $z_{it} \equiv INT(\varepsilon_{it})$ to obtain the Z-scores, again separately for each time point t. Detailed for the INT procedure follow. Suppose u is a skewedly distributed variable with many tie-values (0s in our study). Let rank(u_i) denote the sample rank of u_i when the measurements are placed in ascending order. The inverse normal transformation (INT) for each u_i is then defined as:

$$INT(u_i) = \phi^{-1}\left[\frac{rank(u_i) - k}{n - 2k + 1}\right]$$

Here ϕ^{-1} is the normal density function, $k \in (0, ½)$ is an adjustable offset, and n is the sample size. By default, the Blom offset of $k = 3/8$ is adopted and the transformed variables tend to be very close to normally distributed (S. Guo et al., 2020). Replacing u_i with ε_{it}, we obtain the INT transformed outcome variable, which is now continous and normally distributed.

In the third step, because the INT transformation of the residuals ε_{it} is not a linear transformation, the transformed outcome variable will be again related with the daily invariant background information (S. Guo et al., 2020; McCaw et al., 2019). To mitigate this issue, again separately for each time point, regress the time-variant NPI social distancing measures x_{ikt} on the time-invariant covariates w_i and obtain the residuals for each social distancing measures and treat those residuals as the new proxies for the social distancing measures, $\widetilde{x}_{ikt}$, (k is the number of social distancing measures). In so doing, both time variant and invariant information is incorporated, and the zero-inflated, highly skewed outcome variable (14 day lagged newly confirmed case data) is also transformed to a normally distributed variable. The I-INT transformed variables of the 103 days over the 359 prefectures is summarized in Table 2. As can be seen from Table 2, the outcome variable is now perfectly normally distributed.

After the data is preprocessed through these three stages of indirect-INT approach, we further generated smoothed scatterplots between the outcome (daily newly confirmed cases) and the five NPI strategies. Approximate linear relationships can be reasonably assumed (straight lines for all scatter plots generated through smoothing across all data points) from the smoothed scatterplots. We can then proceed to apply the spatial panel and geographically weighted panel regression models to investigate both the *forest* and the *trees* views of how social-distancing measures curb the spread of COVID-19.

2.2.2. The spatial panel regression method

Regional studies suggest that data collected over geographic space often exhibits strong spatial effects (Anselin, 1988b; Elhorst, 2014) that prevent regular regression estimators from producing consistent and unbiased estimation because the residuals of the regular regression estimators are spatially dependent on one another (spatial autocorrelation) (Elhorst, 2014). Spatial panel regression model is developed in the early 2000s and fully explained in LeSage and Pace (2009) and Elhorst (2014). The fundamental premise is that as data is collected over geographic units, the First Law of Geography "that everything is related with everything else, but closer things are more related" (Tobler, 1970) dominates the data generating process, which causes the nonspatial regression analysis to produce spatially autocorrelated residuals, leading to biased, inefficient and/or misleading results when using the ordinary least squares estimator (Yu & Wei, 2008). The maximum likelihood based alternative estimator is the valid choice for estimation.

Table 2
Summary of I-INT transformed data.

	Min	1st Quantile	Median	Mean	3rd Quantile	Max
newly added cases	−2.922	−0.608	0.000	0.000	0.608	2.922
travelcity	−4.042	−0.390	0.028	0.000	0.447	4.047
addmig	−9.108	−0.330	−0.032	0.000	0.245	15.257
netmin	−13.676	−0.118	0.002	0.000	0.111	5.982
tourists	−83.229	−7.364	−1.167	0.000	6.077	112.207
school	−0.954	0.000	0.000	0.000	0.036	1.013
work	−1.278	−0.033	0.000	0.000	0.006	0.993
quarantine	−1.150	−0.050	0.000	0.000	0.023	1.074
intercity	−1.001	−0.082	0.000	0.000	0.061	1.041

Anselin (2002) and Elhorst (2014) detailed the estimation and testing procedures through Lagrange Multiplier tests. According to Elhorst (2014), three types of interaction effects exist and two of them might contribute to spatial autocorrelation of the residuals. The first is the spatial autocorrelation among the dependent/outcome variables (endogenous interaction). The second is the spatial autocorrelation among the errors (error interaction) because of omitted spatially autocorrelated predictors (independent variables). The third type of interaction is the exogenous interaction in which the dependent variable of a spatial unit depends on the independent variables of neighboring spatial units. This interaction, however, will not cause the regression residuals to be spatially autocorrelated. Depending on which interaction causes the residuals' spatial autocorrelation, two types of spatial panel regression model are often adopted: the spatial lag model that assumes the residuals' spatial autocorrelation is the result of the outcome variable's spatial autocorrelation (endogenous interaction), and the spatial error model that assumes the residuals' spatial autocorrelation is the result of missing spatially autocorrelated explanatory variables (error interaction).

Per Elhorst (2014), a full panel model considering all three types of interactions can be written as:

$$\boldsymbol{Y}_t = \rho \boldsymbol{W}\boldsymbol{Y}_t + \alpha \boldsymbol{\iota}_N + \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\theta} + \boldsymbol{\mu} + \xi_t\boldsymbol{\iota}_N + \boldsymbol{\mu}_t$$

$$\boldsymbol{\mu}_t = \lambda \boldsymbol{W}\boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$$

Where $\boldsymbol{Y}_t$ is the outcome variable at time t. $\boldsymbol{WY}_t$ is the endogenous interaction, also called the spatial lag. $\boldsymbol{W}$ is a spatial weight matrix that defines the neighboring relationship among states. For contiguous areas, either the rook or queen adjacency rule works fine (Anselin, 1988b, 1992; Anselin, Bera, Florax, & Yoon, 1996). For non-contiguous areas, a graphic based sphere of influence (SOI) approach is often employed (Bivand, Pebesma, Gomez-Rubio, & Pebesma, 2008). The SOI approach defines neighborhood relationships in a set of geographic objects (go) as such: for any geographical object in the set, go_i, let r_i be the distance from go_i to its nearest neighbor in the set, and C_i is the circle centered on go_i with the radius of r_i, then i and j are SOI neighbors if and only if C_i and C_j intersect in at least 2 places. ρ is the coefficient for the spatial lag. $\boldsymbol{\iota}_N$ is a vector of 1s, $\boldsymbol{X}_t$ is the matrix of predictor variables, $\boldsymbol{\beta}$ is the vector of coefficients of the predictors. $\boldsymbol{WX}_t$ is the exogenous interaction, and $\boldsymbol{\theta}$ the vector of its coefficients. $\boldsymbol{\mu}$ is the unobservable, individual specific effects, and ξ_t is the time-specific effects. $\boldsymbol{\mu}_t$ is the error term. $\boldsymbol{W\mu}_t$ is the error interaction, and λ its coefficient. $\boldsymbol{\varepsilon}_t$ is the independent and identically distributed (i.i.d.) random noise. Because of nesting all three interaction effects, this model is called General Nesting Spatial (GNS) model (Elhorst, 2014).

Although it is tempting to estimate the GNS model with defined spatial weight matrix because the sources of spatial autocorrelation in the regression residuals are likely from both endogenous and error interactions, the sources usually can only be weakly identified (Elhorst, 2014). For this argument, in practice, either the endogenous (called the spatial autoregressive model, SAR) or error interaction (called the spatial error model, SEM) will be considered to control for spatial autocorrelation in the regression residuals. The choice of either SAR or SEM needs to consider carefully depending on which source might be more likely to introduce spatial autocorrelation to the residuals from a theoretical perspective. A robust Lagrange Multiplier (LM) test (Anselin, 1988a) that is based on the loglikelihood of the alternatives and null models can provide guidance from an empirical perspective. Technique details can be found in Anselin (1988a) and Croissant and Millo (2019) and will not be repeated here. In addition, depending on how the individual effects are assumed to be fixed or generated from a random distribution, panel model can be estimated with either fixed or random effects. A Hausman's test is often used to determine a better alternative (Baltagi, 2005).

2.2.3. The geographically weighted panel regression analysis

To investigate the "*trees*" view of how social distancing measures will curb the spread of COVID-19 at individual prefectures, this study attempts a newly developed exploratory spatial data analytical strategy, the geographically weighted panel regression (GWPR) analysis to investigate potentially varying relationships between the mitigation strategies and the spread of COVID-19. The GWPR analytical procedure (R. H. Cai, Yu, & Oppenheimer, 2014; Yu, 2010, 2014) is a newly developed exploratory spatial data analysis approach extended from cross-sectional geographically weighted regression (GWR) analysis (Fotheringham, Brunsdon, & Charlton, 2002). The fundamental premise for GWR is that regressed relationships are not likely the same from place to place as suggested by conventional regression analysis because of different geographic backgrounds (including socioeconomic, cultural, demographic, and natural conditions). Through introducing a small bias, cross-sectional GWR analysis often reduces the variance of estimated coefficient quite significantly hence the analysis provides better confidence of the estimation (Fotheringham et al., 2002). The approach has seen wide application in many disciplines. Extending the analysis from cross-section to including temporal dimension has been pioneered by Yu (2014), Yu (2010), R. H. Cai et al. (2014), B. Huang, Wu, and Barry (2010), Fotheringham, Crespo, and Yao (2015), among others. Situating GW approach with the panel setting, however, is only explored in (R. H. Cai et al., 2014; Yu, 2010, 2014).

In general, the static panel model is formulated as follows:

$$\mathbf{y}_{1:T} = \mathbf{X}_{1:T}\boldsymbol{\beta} + \mathbf{G}_{1:T}\boldsymbol{\gamma}_G + \mathbf{H}_{1:T}\boldsymbol{\gamma}_H + \boldsymbol{\varepsilon}_{1:T}, \boldsymbol{\varepsilon}_{1:T} \sim N\left(0_{1:T}, \sigma^2\mathbf{I}_{1:T}\right)$$

where T represents time and $\mathbf{y}_{1:T}$ is an $NT \times 1$ vector that stacks the outcome variables vectors at time 1 to T. N is the number of individuals. Other matrices and vectors with the subscript $1:T$ are defined similarly. $\mathbf{X}$ is the matrix of explanatory variables. $\mathbf{G}_{1:T}\boldsymbol{\gamma}_G$ captures the individual or group-specific effects where $\mathbf{G}_{1:T}$ is an $NT \times K_G$ matrix of K_G individual dummy variables indicating district, race, sex, and so on (Greene, 2003) and $\boldsymbol{\gamma}_G$ is a $K_G \times 1$ coefficient vector. $\mathbf{H}_{1:T}\boldsymbol{\gamma}_H$ captures temporal effects where $\mathbf{H}_{1:T}$ is an $NT \times K_H$ matrix indicating time, such as day, month, and year. $\boldsymbol{\gamma}_H$ is a $K_H \times 1$ coefficient vector. The static panel model is often estimated as an individual one-way model (namely, the term $\mathbf{H}_{1:T}\boldsymbol{\gamma}_H$ is often dropped).

Noted that the coefficient vector $\boldsymbol{\beta}$ in equation (1) does not change from location to location, hence the static model. The geographically weighted extension of the static panel model allows the coefficient vector $\boldsymbol{\beta}$ to change over locations (but remain constant over temporal periods for panel analysis). The geographically weighted panel regression model can then be written as:

$$\mathbf{y}_{1:T} = \mathbf{X}_{1:T}\boldsymbol{\beta}_{(u_i,v_i)} + \mathbf{G}_{1:T}\boldsymbol{\gamma}_G + \mathbf{H}_{1:T}\boldsymbol{\gamma}_H + \boldsymbol{\varepsilon}_{1:T}, \boldsymbol{\varepsilon}_{1:T} \sim N\left(0_{1:T}, \sigma^2\mathbf{I}_{1:T}\right)$$

(u_i, v_i) is the coordinate pairs of location i. Everything else remains the same.

The introduction of the spatially varying coefficient immediately introduces either a parsimonious problem that we have more unknowns than data if the number of explanatory variables is more than the number of temporal periods, or the collapse of the panel data to a collection of individual time series estimation if we have longer time periods as in the current study. In the parsimonious scenario, restrictions of the spatially varying mechanisms must be introduced for the coefficients to be estimable (the geographical weighting). In the collapse scenario, though the coefficients are now varying from place to place, the variation is not really "spatial" in the way a geographically weighted approach intends hence estimation that expands from individual locations will be required. However, if the geographical weighting is not assumed to be temporal invariant, we can then avoid the collapse scenario with geographically weighted panel regression as well. Details follow.

2.2.4. Algorithm for implementing geographically weighted panel regression

In a nutshell, geographically weighted approaches assume the observed data is generated by many overlapping and smooth spatial processes that follow a distance-decaying mechanism (Fotheringham et al., 2002). This mechanism is generally referred to as the First Law of Geography, which states that everything is related, but closer things are more related than distant ones (Tobler, 1970) and is often mathematically represented by a kernel function and graphically a symmetric bell-shaped curve (Fotheringham et al., 2002). The observed data on location *i* is hence the result of overlapping a smooth distance-decaying process that centered on location *i*, and many other smooth distance-decaying processes that centered on other locations but neighbor location *i* in various degrees depending on the distance between those locations and location *i*.

Based on this argument, once a functional form for the distance-decaying process is identified, it is possible to find all the spatial processes that project influences on location *i*. So that we can create a unique sample for location *i* that only pertains to location *i* hence will not involve spatial effects. Since the distance-decaying process is symmetric (distance-decaying from location *i* to location *j* is equivalent to distance-decaying from location *j* to location *i*), finding all the spatial processes that project influences on location *i* is equivalent to singling out the one spatial process that centered on *i* and finding out all the other locations that this spatial process reaches. Once that spatial process is singled out (represented as a symmetric bell-shaped curve), one can then extract the part of information that belongs only to this spatial process at location *i*, and the parts of observed information in any other locations that this spatial process reaches. Once this is done, for each location *i*, we could then construct a subsample of observations that only pertains to location *i*. Because this subsample does not involve spatial effects, estimation based on this subsample could utilize the regular ordinary least squares model. In addition, also because the information for this subsample will be extracted from existing data, the distance-decaying mechanism will be used to weigh the original data at every location the *i*th spatial process reaches, hence the name "geographically weighted." The First Law of Geography and its approximation, the distance-decaying kernel function are the keys to this "geographical weighting."

The "geographical weighting" can solve the problem of the parsimonious scenario well because now for each location, there will be a weighted subsample that will have enough data to estimate the coefficients. "Geographical weighting" might also provide a more tenable way for estimation than simply rely on estimating non-related individual time series data on each location in the collapse scenario.

How to weigh existing data on all locations, however, is often considered slightly differently in different studies. Time is added as an additional dimension as in (Yu, 2014), (B. Huang et al., 2010), and (Fotheringham et al., 2015), or the geographical weighting is applied to every temporal period as time invariant mechanism to generate a subsample of panel data for each location as in (R. H. Cai et al., 2014). In this study, however, we do not add time as an additional dimension, nor do we intend to treat the geographical weighting as time invariant mechanism. Instead, we argue that the geographical weighting should be time variant and should be based on each cross-sectional dataset. We argue that although the geographical arrangement of observations does not change abruptly over the temporal period, assuming the data generating spatial processes remain the same for different temporal periods sounds less tenable. As an exploratory approach, a time varying spatial process might be closer to the true data generating mechanism (DGM). It is very likely, however, that the time variant geographical weighting will generate different sizes of samples for different temporal period. The subsample for each location is likely an unbalanced panel dataset. Still, such treatment could effectively avoid the collapse scenario when estimating geographically weighted panel regression.

The entire estimation procedure for the proposed geographically weighted panel regression follows these steps.

1. For each temporal period, the cross-sectional data is extracted. Conventional geographically weighted regression approach is applied to this cross-section dataset to determine the geographically weighted subsample for each location at this temporal period. Specifically:
   - 1.1 A distance-decaying spatial kernel function of the form $k(d_{ij}) = f((d_{ij}/b)^{-\alpha})$ or $k(d_{ij}) = f(-(d_{ij}/b)^{\alpha})$ is chosen to decide the local region around a location and weigh the observations that fall within this local region (where d_{ij} is the distance between locations *i* and *j*; *b* is called the bandwidth that determines the flatness of the bell-shaped symmetric kernel (lager the *b*, flatter the kernel); α is a parameter that controls how quickly the values on the kernel curve drops from 1, often α takes the value of 2 (bi-square) or 3 (tri-cube). For a location *i*, any other location *j* will be assigned the weight of $k(d_{ij})$ and will be used to weigh the observations on location *j*; the weighted observations are then assembled with the observation on location *i* to form the subsample for location *i*).
   - 1.2 With the chosen spatial kernel function, a starting bandwidth *b* is arbitrarily selected to determine the local region for each location. The weights are assigned for all the observations fall within the local region using the kernel function with the arbitrarily chosen bandwidth *b*. A local subsample for each location is created.
   - 1.3 When all the locations have their own subsamples, ordinary least squares regression analysis at each location is conducted to produce the spatially varying coefficients of each explanatory variable.
   - 1.4 An "optimal" bandwidth *b* will be determined through optimization strategies that either maximize the model fit (such as the leave-one-out cross-validation approach) or minimize the information loss (such as the Akaike Information Criterion approach) by repeating steps 1.2 and 1.3 with different *b*.
   - 1.5 Once an "optimal" *b* is determined, the local region and weights for observations fall within this local region for all locations can be determined (the geographically weighted subsamples) for the specific temporal period.
2. For each location, the geographically weighted subsamples from all temporal periods are combined to be a (likely unbalanced) panel dataset.
3. For each location, regular panel regression analysis will be applied to estimate the coefficients of the explanatory variables.
4. The estimation will be repeated for all locations and the coefficients estimated in such way will be spatially varying.
5. For each local panel regression, the regular statistical tests for the significance of the coefficients will also be produced. The significance test results will be used for mapping purposes. Specifically, only the statistically significant local coefficients' spatial variation will be mapped. In such way, the maps answer directly the question *where the social distancing measures work* and *how well these measures work* to curb the spread of COVID-19.

3. Results: seeing the forest and the trees with advanced geospatial analyses

3.1. Seeing the forest

Although our data covers the entire period from January 25th, 2020, to May 6th, 2020, preliminary data exploration suggests that the newly added cases for the entire nation flattens on around March 5th, 2020. Following March 5th, 2020, most of the prefectures in China experienced rather sporadic occurring of newly added COVID-19 cases. The large number of cases appear only in or around Wuhan, the original epicenter of the outbreak. For this matter, the study presents results from two models with different temporal span: model 1 reports the period before the flattening on March 5th, 2020 (Table 3). Model 2

Table 3
Influence of the social distancing measures on the daily newly confirmed cases in China at prefecture level, January 25th – March 5th, 2020, fixed effect non-spatial and spatial lag panel regression models.

Fixed effect panel regression model results				
Residuals				
Min.	1st Qu.	Median	3rd Qu.	Max.
−4.092	−0.169	−0.009	0.149	4.447
Social distancing measures	Estimate	Std. Error	t-value	Pr(>|t|)
travelcity	0.114	0.009	12.393	0.000 ***
addmig	0.028	0.006	4.288	0.000 ***
netmin	0.005	0.006	0.823	0.411
work	−0.285	0.020	−14.566	0.000 ***
tourists	−0.001	0.000	−2.023	0.043
quarantine	0.120	0.051	2.342	0.019
intercity	0.067	0.017	3.941	0.000
R-Squared: 0.03268				
Adj. R-Squared: 0.0080809				
F-statistic: 69.2719 on 7 and 14,353 DF, p-value: <2.22e-16				
Robust Lagrange multiplier test: RLMlag = 32.68, RLMerr = 5.87				
Fixed effect spatial lag panel regression model results				
Residuals				
Min.	1st Qu.	Median	3rd Qu.	Max.
−3.930	−0.160	−0.010	0.142	4.434
Social distancing measures	Estimate	Std. Error	t-value	Pr(>|t|)
travelcity	0.098	0.009	11.009	0.000 ***
addmig	0.030	0.006	4.731	0.000 ***
netmin	0.008	0.006	1.398	0.162
work	−0.223	0.019	−11.742	0.000 ***
tourists	−0.001	0.000	−1.548	0.122
quarantine	0.100	0.049	2.024	0.043
intercity	0.065	0.016	3.941	0.000
Spatial autoregressive coefficient Lambda	0.234	0.011	21.363	0.000 ***

Significance codes: ***: 0.001; **: 0.01; *: 0.05; .: 0.1 (one-tailed test).

Table 4
Influence of the social distancing measures on the daily newly confirmed cases in China at prefecture level, January 25th – May 6th, 2020, fixed effect non-spatial and spatial lag panel regression models.

Fixed effect panel regression model results				
Residuals				
Min.	1st Qu.	Median	3rd Qu.	Max.
−3.520	−0.380	−0.057	0.319	4.455
Social distancing measures	Estimate	Std. Error	t-value	Pr(>|t|)
travelcity	0.106	0.009	11.425	0.000 ***
addmig	0.024	0.008	3.010	0.003 **
netmin	−0.004	0.008	−0.539	0.590
school	−0.058	0.018	−3.229	0.001 **
work	−0.087	0.022	−3.929	0.000 ***
tourists	−0.001	0.000	−2.236	0.025
quarantine	−0.069	0.017	−4.147	0.000 ***
intercity	−0.003	0.016	−0.194	0.847
R-Squared: 0.0074346				
Adj. R-Squared: −0.0026868				
F-statistic: 33.6053 on 8 and 35,892 DF, p-value: <2.22e-16				
Robust Lagrange Multiplier test: RLMlag = 176.03, RLMerr = 67.91				
Fixed effect spatial lag panel regression model results				
Residuals				
Min.	1st Qu.	Median	3rd Qu.	Max.
−3.360	−0.330	−0.049	0.287	4.787
Social distancing measures	Estimate	Std. Error	t-value	Pr(>|t|)
travelcity	0.077	0.009	8.945	0.000 ***
addmig	0.019	0.007	2.521	0.012 **
netmin	−0.004	0.007	−0.569	0.569
school	−0.024	0.016	−1.491	0.136
work	−0.022	0.020	−1.089	0.276
tourists	0.000	0.000	0.190	0.849
quarantine	−0.033	0.015	−2.172	0.030 *
intercity	0.005	0.014	0.328	0.743
Spatial autoregressive coefficient Lambda	0.418	0.006	68.197	0.000 ***

Significance codes: ***: 0.001; **: 0.01; *: 0.05; .: 0.1 (one-tailed test).

reports the entire period from January 25th to May 6th, 2020 (Table 4). The significant test is based on $p < 0.05$, one-tailed test, and marked with the "*" symbol in the tables. We do not report the period after the flattening because the daily newly confirmed cases show only small amount of variation throughout China. The non-spatial individual fixed effect panel regression results are also reported for comparison purposes.

3.2. Seeing the trees

The geographically weighted panel regression (GWPR) was first proposed in (Yu, 2010) as one of the alternative local analytical approaches that deals with geo-panel dataset. In addition to allowing the distance-decaying processes to be singled out for each location, the GWPR also allows the processes to be different for each temporal period (each day in the current study). This is critical for long panels as in the current study comparing to short panels. The results of the GWPR approach for the period before the flattening is reported in Fig. 1, and for the entire period is reported in Fig. 2. Only prefectures whose coefficients are pseudo-significant at 95% confidence level (one-tailed, the absolute cut value for one-tailed test is 1.68) are grey scaled. The variable *netmin* does not show statistical significance for any prefectures for both periods, hence its maps are not reported.

4. Discussion

4.1. The forest view

4.1.1. The overall impression

When reading Tables 3 and 4, we found that the non-spatial panel models' (for both periods) adjusted R^2s are very low. This is to be expected, however, because this study focuses on how social distancing based NPI strategies could curb the spread of COVID-19 on a daily basis. Undoubtedly, other than the NPI strategies, there are many other important contributors that also work to curb the spread of the disease. The most important ones are medical infrastructure related factors such as numbers of health providers, capacity of the health care systems. Unfortunately, those factors are not readily collected at daily interval so that they cannot be included in the model. While the I-INT pre-process attempted to include the background information, the procedure includes the information in a rather coarse way. Because we do not have sufficient number of contributing factors other than the mined NPI strategies at the required daily interval, those missing important variables are all captured by the residuals. This is also a problem that the nonspatial global level model cannot address adequately. Still, when spatial effects are introduced, the missing variables' information is indirectly modeled through addressing the spatially autocorrelated residuals in the global spatial panel model, and directly modeled in the

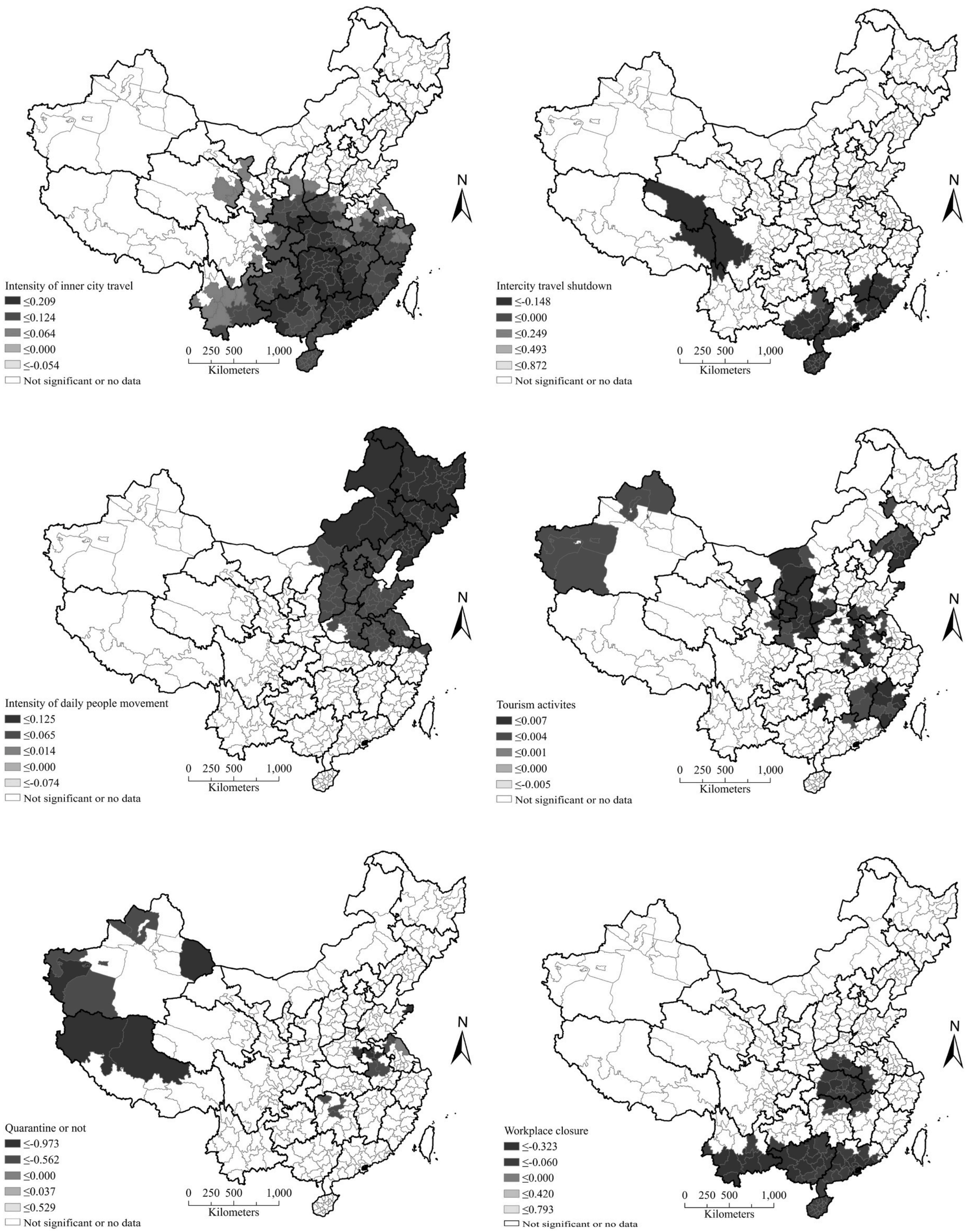

Fig. 1. Spatially varying influence of the social distancing measures on the daily newly confirmed cases in China at prefecture level, January 25th – March 5th, 2020.

GWPR model, albeit not at the same statistic power as when actual variables are included in the model.

4.1.2. The detailed interpretations

By reading Tables 3 and 4, we can draw some interesting conclusions of the investigated relationships between the spread of COVID-19 and the NPI strategies as well. First, the modeling tests suggest that a non-spatial panel model could potentially produce misleading results due to significant spatial autocorrelation in the regression residuals. Both robust Lagrange Multiplier tests support the spatial lag panel model. For the period prior to the flattening, however, the spatial and non-spatial models agree well. The primary contributing factors for the daily spread of COVID-19 are the intensity of inner-city travel, the intensity of daily migration, and workplace closure. These results agree well with previous studies found in China (X. Guo et al., 2021; Xiong et al., 2020; Y. H. Zhang, Zhang, & Wang, 2020), in UK (Davies et al., 2020), Italy (Giordano et al., 2020), and Germany (Chae & Park, 2020). The results suggest that the rigorous lock-down policy implemented in Wuhan after

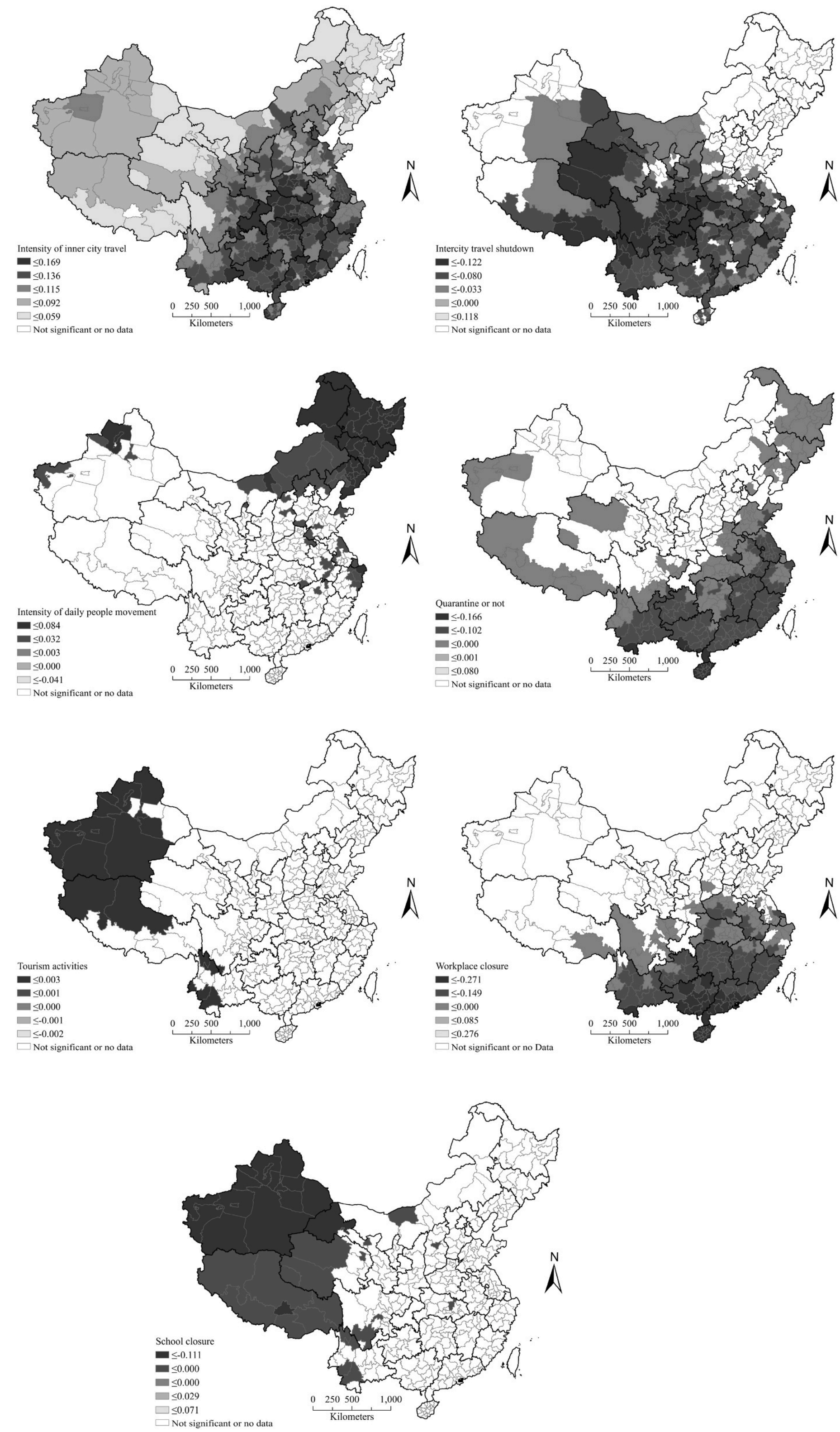

Fig. 2. Spatially varying influence of the social distancing measures on the daily newly confirmed cases in China at prefecture level, January 25th – May 4th, 2020.

January 23rd, 2020, and later to the entire nation contributes significantly to curb the overall spread of the COVID-19 in China. This result also suggests that social-distancing policies/strategies are working during the initial outbreak of a pandemic, though it might not be immediate (Vogel, 2020), but ignoring it could prove to be detrimental (Mahase, 2020; Manchein et al., 2020).

Second, when looking at the results from the entire periods, the non-spatial model suggests that other than the above three social distancing polices/strategies, school closure (which does not appear in the first period because schools remained closed during that period), and whether the prefecture is under mandatory quarantine, also work to curb the spread. After controlling for the spatial effects, however, the workplace closure and school closure strategies do not seem to be significant factors. Whether a prefecture is under strict quarantine still serves as a significant contributor to curb the spread. Quarantine and lockdowns are still effective means when curbing a contagious disease as commonly practiced in epidemiological studies (Guner, Hasanoglu, & Aktas, 2020; Manchein et al., 2020).

The discrepancy regarding the closure of the working places and quarantine policy suggests that during the initial period of the outbreak, the severity of the pandemic was only gradually being realized by the society. Some working places remained open in the hope that the pandemic will end soon. While in the meantime, the quarantine policy was strict and almost universal except for some relatively remote prefectures in the western parts of China. The model of the first period captures the significant workplace closure effect and insignificant quarantine policy effect. However, when the entire period is considered, as more workplaces were closed because of clear correlation between opening of workplaces and the spread of the disease as witnessed in ensuing studies (Mendez-Brito, El Bcheraoui, & Pozo-Martin, 2021), the effect of closing workplaces seems to become less important in curbing the spread. Quarantine policy, on the contrary, was relaxed in many relatively remote prefectures after the flattening of the curve but remained a strict policy measure in densely populated eastern and southern parts of China. Relaxation of quarantine policy likely facilitates the spread of the disease, like discussed in Cimolai (2021), our model is able to capture this effect.

Third, it is also worth noting that the absolute values of the significant coefficients are larger for the period prior to the flattening of the curve than for the entire period. This agrees with the common knowledge that early implementation of social distancing strategies tends to be more effective because the NPI strategies could block the transmission pathways of the virus (Baker et al., 2020; Brauner et al., 2021; Mendez-Brito et al., 2021; Redlin, 2022). In addition, the "forest" view suggests that except for the above three measures, all other social-distancing measures do not seem to play significant roles in curbing the spread of COVID-19 in China at the prefecture level. A decision-maker or modeler who works with the global model might suggest those social distancing strategies are less effective hence might not give them adequate consideration. This is more evident when we model the entire study period. Understandably, when the disease was gradually under control, with both significantly decreased cases and gradually relaxed policies in China, the relationship between the spread of the disease and the social distancing measures will be diluted and eventually become nonexistent. The results suggest that China's quick and decisive social distancing campaign and quarantine policies worked well and quickly as discussed in detail in Tian et al. (2020). The problem is, however, that such campaign and policies are hardly replicable elsewhere no matter it is due to the cultural and political system differences or economic resistance. The recent strict tracking and local lock-down approaches adopted by the Chinese government due to the resurgence of sporadic newly added cases suggests the potential cost could be prohibitively high elsewhere.

In a sense, the *forest view* provides a general, global perspective as to what might work and what might not. It is a standard and effective way in epidemiology to curb the spread of a contagious disease at early stages of the disease when data was scarce, and many remained unknown (Fong et al., 2020). It is very likely that some social distancing policies/strategies work in certain places and certain times, but not necessarily in all places and all times (Davalgi, Undi, Annadani, & Nawaz, 2020). The models established under the *forest view* tend to average out and mask these effects. Considering the startling cost of implementing social distancing measures in nations as vast as China, a global, *forest view* might not always produce the most ideal results for efficient strategies to curb the spread of COVID-19 elsewhere, albeit effective.

4.2. *The trees view*

4.2.1. *The overall impression*

When we turn our attention to the results produced by the GWPR models in Figs. 1 and 2, a more detailed picture emerges. It is evident that other than the net daily movement of people (*netmin*, which is also not significant at the global level), all the included social-distancing measures show significant contribution to curb the spread of COVID-19 at the 95% confidence level in some prefectures in China, but not all prefectures. Comparing with the *forest view*, this result suggests that the global model might mask out important details that merit close attention, especially in the time of a global pandemic from an unknown infectious disease. The variation of the coefficients' values suggests that estimates of the *forest view* are averaged results from the *trees view*. While the *forest view* is valuable to tell at a national/regional level what NPI strategies shall be implemented immediately, the *trees view* takes to the next level to suggest at when and where, what work most effectively. Seeing both the *forest* and the *trees* provides strategical as well as detailed information for both effective and efficient decision-making.

4.2.2. *The detailed patterns*

Detailed patterns appear when we look at the figures individually. The grey scaled patterns for all the social distancing measures produced by GWPR reveal rich amount of information. First, for the period prior to the flattening of the curve (March 5th, 2020), Fig. 1 suggests that in prefectures around the epicenter, Wuhan, extending to the East and Southeastern coastal regions, the Southwestern regions and some prefectures in Shaanxi and Gansu of the Northwestern China, the inner-city travel intensity is the primary factor for the spread of the disease. On the other hand, the number of daily inter-prefecture migration is significantly contributing to the spread of COVID-19 in Northeastern and Northern China prefectures. Inter-city travel bans and workplace closure, however, worked primarily in prefectures immediately around the epicenter and the Peral River Delta Urban Agglomeration, where the largest temporary migrant destinations in China locate. The quarantine policy does not seem to have a distinctive pattern, though where it worked, it seems to work the best to curb the spread since quarantined prefectures will have a maximum of 0.973 cases less than non-quarantined ones daily, this observation, even at a varying local level, agrees with mainstream epidemiological evidence regarding the general effectiveness of quarantine (Charvadeh et al., 2022; Chinazzi et al., 2020; Guner et al., 2020). The tourism activities' relationship with the spread of the disease shows an inconsistent pattern. This could be the result of the way the data is collected (the number of comments on popular tourism sites) and it might not be representative of happened trips.

Second, the patterns in Fig. 2 do not change much since newly confirmed cases after the flattening changed very little, though prefectures with significant relationships between the various social distancing measures and the spread of COVID-19 expanded from the original clusters. The strengths of such significant relationships as measured by the varying coefficients are also smaller than the previous period (the absolute values of all maximum significant coefficients are smaller than in the previous period). This result suggests that the quick and strict social distancing measures taken by the Chinese government in the early stages of COVID-19 development was working and continued to work even after the flattening of the curve until the end of our study period.

Third, the patterns in both Figs. 1 and 2 suggest that for the entire country, mass movement of population, either within the city (inner city travel), or across cities (intercity travel and migration), are the most important factors that facilitate the spread of COVID-19. The global model failed to pick out the intercity travel ban as an important factor, primarily because the factor only works in areas around the epicenter and large temporary migrations centers as we now see in Figs. 1 and 2. Consequently, the global model averages out the local effectiveness of this factor. Reducing the amount of mass movement of population is the priority for the entire nation to curb the spread of COVID-19 (Chinazzi et al., 2020; Kraemer et al., 2020). For workplace closure and city-wide strict quarantine, these measures worked the best in the epicenters and cities that are either physically or socioeconomically close to the epicenter (Wuhan).

4.2.3. Significant policy implications of these patterns revealed by the GWPR

These specific spatial patterns are of critical importance for devising targeted social distancing policies for different prefectures in China. While it is well known that reduced mass movement of population is the key to slow down the spread of an infectious disease (Kraemer et al., 2020), reducing mass movement of population can be implemented differently in different regions. In China, for the prefectures that are either physically or socioeconomically close to the epicenter, the strictest social distancing policy, namely, restricting daily outings, closure of working places, and shutting down inter-city travels are necessary. In the Northeastern and Northern China centered on the capital city, Beijing, a strict temporary migration ban might prove to be effective to slow down the spread of COVID-19 because Beijing is one of the busiest hubs for temporary migrants from neighboring regions.

When the gravity of the COVID-19 disease weighed heavily on China's society in late January to February, the Chinese government took decisive and strict social distancing and quarantine measures *ad hoc*, which was effective as supported by our models and other studies (Kraemer et al., 2020; Tian et al., 2020). Such strict measures, however, can hardly be repeated elsewhere. Our models, especially the local models that provide the *trees view* can be applied *a priori* as data accumulates and might provide a more efficient strategy when implementing social distancing measures, alongside with the global model.

The impact of school closure was not measurable prior to the flattening of the curve because that was during the traditional Chinese New Year, no school opened before the flattening. For the entire study period, however, the impact of school closure seems to only have significant impact in the relatively remote cities in Qinghai, West Gansu, Xinjiang and Tibet, and not significant at the global level (Table 4), despite findings suggesting that school closure has significant impact on curbing the spread (Auger et al., 2020). Considering that these remote Western cities often have relatively weaker health care system than the Central and Eastern China, including tests and trace intervention, reopening of schools there was more likely associated with the spread of the disease as modeled in a recent UK study (Panovska-Griffiths et al., 2020).

The practice of implementing the GWPR model coupled with the global spatial panel model suggests a holistic spatial analysis of geographical phenomena (spatial epidemiology in our study) is of pragmatical significance. Data collected over geographic space shall always to scrutinized more carefully to make sure spatial effects, if present, are taken care of (Anselin, 2007). More importantly, however, the current study further suggests that it is of pragmatical importance to always implement a holistic spatial analytical approach that includes both global and local methods to provide a better understanding of the geographic process and data generating mechanisms.

5. Conclusion: implications for future research paradigm

Through integrating a global spatial panel regression model and a local geographically weighted panel regression model, this study attempts to provide a holistic view of investigating how NPI strategies influence the spread of COVID-19 diseases in China during the first three months of the outbreak. The results suggest that NPI strategies are in general effective to prevent the spread of the disease, but the effectiveness varies from place to place.

The take-home message from this study is that while social distancing based NPI measures work to curb the spread of COVID-19, these measures are usually effective at certain places and certain times as the pandemic progresses. Knowing where those measures are effective is critical for efficient decision making especially when facing unknown epidemiological events like the COVID-19. The outbreak of COVID-19 caught the entire human race off-guard. Medical equipment and personnel were limited at the beginning of the outbreak. All countries had to devise and implement various social distancing based NPI measures *ad hoc*. China has implemented one of the strictest social distancing strategies that are effective but can hardly be repeated elsewhere. The introduction of local models with advanced geospatial analysis has the potential to provide *a priori* strategies guiding more efficient implementation for social distancing measures during unknown pandemic. Both the *forest* and *trees views* suggest that restriction of mass movement of population and strict quarantine work the best in restricting the spread of COVID-19. The *trees view* further suggests that measures such as closure of workplace and strict quarantine policy work the best around the epicenter and large migration origins and destinations. Knowing what strategies work at when and where could potentially put the limited medical equipment and personnel into the most efficient use, which could save lives and reduce social disturbance and economic downturn resulted from the strict national lock-down policy as implemented in China during the February–May 2020 period. The practice proposed in this work is of great potential to facilitate both the understanding of how social distancing based NPI measures work (when and where), and the implementation of targeted social distancing campaign in certain areas to achieve maximum benefits of curbing the spread of the COVID-19. Considering during this early period of the pandemic, neither effective COVID-19 vaccines nor other pharmaceutical solutions were available at the decision-makers' disposal, such information is essential to flatten the curve of the COVID-19 diseases, to prevent the collapse of the health system, and to allocate the limited health care resources to places where they are most needed, with minimal cost of both human lives and economic performance. For this very practical reason, the proposal practice in the current study merits further attention of the scientific community.

Author statement

Danlin Yu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration Yaojun Zhao: Conceptualization, Validation, Investigation, Resources, Writing – Original Draft, Supervision, Project administration Jun Meng: conceptualization, Validation, Investigation, Writing – Original Draft, Supervision Xiaoxi Wang: Software, Investigation, Data Curation Linfeng He: Software, Formal analysis, Data Curation Meng Jia: Data Curation Jie Ouyang: Data Curation Yu Han: Data Curation Ge Zhang: Data Curation Yao Lu: Data Curation.

Acknowledgement

This research is funded by the IDC fund (CC10502 Earth IDC) of the Department of Earth and Environmental Studies, College of Science and Mathematics, Montclair State University.

References

Alene, M., Yismaw, L., Assemie, M. A., Ketema, D. B., Gietaneh, W., & Birhan, T. Y. (2021). Serial interval and incubation period of COVID-19: A systematic review and meta-analysis. *BMC Infectious Diseases, 21*(1).
Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet, 395*(10228), 931–934.
Anselin, L. (1988a). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis, 20*(1), 1–17.
Anselin, L. (1988b). *Spatial econometrics: Methods and models*. Dordrecht, The Netherland: Kluwer Academic Publisher.
Anselin, L. (1992). Space and applied econometrics - introduction. *Regional Science and Urban Economics, 22*(3), 307–316.

Anselin, L. (2002). Under the hood - issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27*(3), 247–267.
Anselin, L. (2007). Spatial econometrics in RSUE: Retrospect and prospect. *Regional Science and Urban Economics, 37*(4), 450–456.
Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics, 26*(1), 77–104.
Auger, K. A., Shah, S. S., Richardson, T., Hartley, D., Hall, M., Warniment, A., et al. (2020). Association between statewide school closure and COVID-19 incidence and mortality in the US. *JAMA, 324*(9), 859–870.
Baker, R. E., Park, S. W., Yang, W. C., Vecchi, G. A., Metcalf, C. J. E., & Grenfell, B. T. (2020). The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections. *Proceedings of the National Academy of Sciences of the United States of America, 117*(48), 30547–30553.
Baltagi, B. H. (2005). *Econometric analysis of panel data* (3rd ed.). Chichester, West Sussex, UK: John Wiley & Sons.
Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B., & Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences, 117*(29), 16732–16738.
Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R, 747248717*. Springer.
Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenciak, T., et al. (2021). Inferring the effectiveness of government interventions against COVID-19. *Science, 371*(6531), 802–+.
Cai, Y. Y., Liu, J. Y., Yang, H. T., Wang, M., Guo, Q. P., Huang, D., et al. (2020). Association between incubation period and clinical characteristics of patients with COVID-19. *Journal of International Medical Research, 48*(9).
Cai, R. H., Yu, D. L., & Oppenheimer, M. (2014). Estimating the spatially varying responses of corn yields to weather variations using geographically weighted panel regression. *Journal of Agricultural and Resource Economics, 39*(2), 230–252.
Cevasco, K. E., Roess, A. A., North, H. M., Zeitoun, S. A., Wofford, R. N., Matulis, G. A., et al. (2021). Survival analysis of factors affecting the timing of COVID-19 non-pharmaceutical interventions by US universities. *BMC Public Health, 21*(1).
Chae, S. H., & Park, H. J. (2020). Effectiveness of penalties for lockdown violations during the COVID-19 pandemic in Germany. *American Journal of Public Health, 110*(12), 1844–1849.
Charvadeh, Y. K., Yi, G. Y., Bian, Y., & He, W. Q. (2022). Is 14-days a sensible quarantine length for COVID-19? Examinations of some associated issues with a case study of COVID-19 incubation times. *Statistics in Biosciences, 14*(1), 175–190.
Chen, P. Y., Guo, X. D., Jiao, Z. T., Liang, S. H., Li, L. F., Yan, J., et al. (2022). A multilayer network model for studying the impact of non-pharmaceutical interventions implemented in response to COVID-19. *Frontiers in Physics, 10*.
Cheng, C., Zhang, D. D., Dang, D. J., Geng, J., Zhu, P. Y., Yuan, M. Z., et al. (2021). The incubation period of COVID-19: A global meta-analysis of 53 studies and a Chinese observation study of 11 545 patients. *Infectious Diseases of Poverty, 10*(1).
Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science, 368*(6489), 395–+.
Cimolai, N. (2021). In pursuit of the right tail for the COVID-19 incubation period. *Public Health, 194*, 149–155.
Ciufolini, I., & Paolozzi, A. (2020). Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a Gauss error function and Monte Carlo simulations. *European Physical Journal Plus, 135*(4), 8.
Croissant, Y., & Millo, G. (2019). *Panel data econometrics with R*. Hoboken, NJ: John Wiley & Sons.
Davalgi, S., Undi, M., Annadani, R., & Nawaz, A. S. (2020). Comparison of Measures adopted to combat COVID 19 Pandemic by different countries in WHO regions. *Indian Journal of Community Health, 32*(2), 288–299.
Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., Jombart, T., et al. (2020). *Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: A modelling study*. The Lancet Public Health.
Dayaratna, K. D., Gonshorowski, D., & Kolesar, M. (2022). Hierarchical Bayesian spatio-temporal modeling of COVID-19 in the United States. *Journal of Applied Statistics*, 1–18.
Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., et al. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science, 369*(6500), Article eabb9789.
Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X. J., et al. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America, 117*(18), 9696–9698.
Drew, D. A., Nguyen, L. H., Steves, C. J., Menni, C., Freydin, M., Varsavsky, T., et al. (2020). Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science, 368*(6497), 1362–+.
Elhorst, J. P. (2014). *Spatial econometrics: From cross-sectional data to spatial panels*. New York, NY: Springer.
Enserink, M., & Kupferschmidt, K. (2020). With COVID-19, modeling takes on life and death importance. *Science, 367*(6485), 1414–1415.
Fair, K. R., Karatayev, V. A., Anand, M., & Bauch, C. T. (2022). Estimating COVID-19 cases and deaths prevented by non-pharmaceutical interventions, and the impact of individual actions: A retrospective model-based analysis. *Epidemics, 39*.
Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 1–5.
Fong, M. W., Gao, H. Z., Wong, J. Y., Xiao, J. Y., Shiu, E. Y. C., Ryu, S., et al. (2020). Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings-social distancing measures. *Emerging Infectious Diseases, 26*(5), 976–984.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationship*. West Essex, England: John Wiley & Sons.
Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis, 47*(4), 431–452.
Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine, 32.
Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., et al. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine, 32*.
Greene, W. H. (2003). *Econometric analysis*. India: Pearson Education.
Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine, 382*(18), 1708–1720.
Guner, R., Hasanoglu, I., & Aktas, F. (2020). COVID-19: Prevention and control measures in community. *Turkish Journal of Medical Sciences, 50*(3), 571–577.
Guo, S., An, R., McBride, T. D., Yu, D., Fu, L., & Yang, Y. (2020). *Mitigation interventions in the United States: An exploratory investigation of determinants and impacts, 1049731520957415*. Research on Social Work Practice.
Guo, X., Zhang, J., & Wu, X. (2021). Spatio-temporal characteristics of the novel coronavirus attention network and its influencing factors in China. *PLoS One, 16*(9), Article e0257291.
Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science, 24*(3), 383–401.
Huang, J., Zhang, L., Liu, X., Wei, Y., Liu, C., Lian, X., et al. (2020). Global prediction system for COVID-19 pandemic. *Science Bulletin, 65*(22), 1884–1887.
Kraemer, M. U. G., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science, 368*(6490), 493–+.
LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: Taylor & Francis/CRC Press.
Mahase, E. (2020). Covid-19: UK starts social distancing after new model points to 260 000 potential deaths. *BMJ British Medical Journal, 368*, 2.
Manchein, C., Brugnago, E. L., da Silva, R. M., Mendes, C. F. O., & Beims, M. W. (2020). Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies. *Chaos, 30*(4), 9.
McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S., & Lin, X. H. (2019). Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics, 11*.
Mendez-Brito, A., El Bcheraoui, C., & Pozo-Martin, F. (2021). Systematic review of empirical studies comparing the effectiveness of non-pharmaceutical interventions against COVID-19. *Journal of Infection, 83*(3), 281–293.
Nazia, N., Law, J., & Butt, Z. A. (2022). Identifying spatiotemporal patterns of COVID-19 transmissions and the drivers of the patterns in toronto: A bayesian hierarchical spatiotemporal modelling. *Scientific Reports, 12*(1), 1–13.
Panovska-Griffiths, J., Kerr, C. C., Stuart, R. M., Mistry, D., Klein, D. J., Viner, R. M., et al. (2020). *Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: A modelling study*. The Lancet Child & Adolescent Health.
Poletto, C., Scarpino, S. V., & Volz, E. M. (2020). Applications of predictive modelling early in the COVID-19 epidemic. *The Lancet Digital Health, 2*(10), e498–e499.
Quesada, J. A., Lopez-Pineda, A., Gil-Guillen, V. F., Arriero-Marin, J. M., Gutierrez, F., & Carratala-Munuera, C. (2021). Incubation period of COVID-19: A systematic review and meta-analysis. *Revista Clínica Española, 221*(2), 109–117.
Redlin, M. (2022). Differences in NPI strategies against COVID-19. *Journal of Regulatory Economics, 62*(1–3), 1–23.
Tian, H. Y., Liu, Y. H., Li, Y. D., Wu, C. H., Chen, B., Kraemer, M. U. G., et al. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science, 368*(6491), 638–+.
Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography, 46*, 234–240.
Vogel, L. (2020). How long will social distancing take to work? Experts weigh in on Canada's COVID-19 response. *Canadian Medical Association Journal, 192*(14), E382–E383.
Wan, S. Y., Liu, J., & Liu, M. (2021). Progress on the incubation period of COVID-19. *Chinese Science Bulletin-Chinese, 66*(15), 1802–1811.
Wei, S., & Wang, L. (2020). Examining the population flow network in China and its implications for epidemic control based on Baidu migration data. *Humanities & Social Sciences Communications, 7*(1).
WHO Commission on Social Determinants of Health, W. H. O. (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health: Commission on social determinants of health final report*. World Health Organization.
Wu, Y., Kang, L. Y., Guo, Z. R., Liu, J., Liu, M., & Liang, W. N. (2022). Incubation Period of COVID-19 caused by unique SARS-CoV-2 strains A systematic review and meta-analysis. *JAMA Network Open, 5*(8).
Xiong, Y., Wang, Y., Chen, F., & Zhu, M. (2020). Spatial statistics and influencing factors of the COVID-19 epidemic at both prefecture and county levels in Hubei Province, China. *International Journal of Environmental Research and Public Health, 17*(11), 3903.

Yu, D. L. (2010). Exploring spatiotemporally varying regressed relationships: The geographically weighted panel regression analysis. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38*, 134–139.
Yu, D. L. (2014). Understanding regional development mechanisms in Greater Beijing Area, China, 1995–2001, from a spatial–temporal perspective. *Geojournal, 79*(2), 195–207.
Yu, D. L., Murakami, D., Zhang, Y. J., Wu, X. W., Li, D., Wang, X. X., et al. (2020). Investigating high-speed rail construction's support to county level regional development in China: An eigenvector based spatial filtering panel data analysis. *Transportation Research Part B: Methodological, 133*, 21–37.
Yu, D. L., & Wei, Y. D. (2008). Spatial data analysis of regional development in Greater Beijing, China, in a GIS environment. *Papers in Regional Science, 87*(1), 97–117.
Yu, D. L., Zhang, Y. J., & Wu, X. W. (2020). How socioeconomic and environmental factors impact the migration destination choices of different population groups in China: An eigenfunction-based spatial filtering analysis (vol 15, pg 129, 2020). *Population and Environment, 41*(3), 396, 396.
Yu, D. L., Zhang, Y. J., Wu, X. W., Li, D., & Li, G. D. (2021). The varying effects of accessing high-speed rail system on China's county development: A geographically weighted panel regression analysis. *Land Use Policy, 100*, 11.
Zhang, Y. H., Zhang, A. M., & Wang, J. E. (2020). Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. *Transport Policy, 94*, 34–42.
Zhang, L., Zhu, J. H., Wang, X. Y., Yang, J., Liu, X. F., & Xu, X. K. (2021). Characterizing COVID-19 transmission: Incubation period, reproduction rate, and multiple-generation spreading. *Frontiers in Physics, 8*.