

數據科學看新型冠狀病毒

宏觀視野 • 秦泗釗 張清鵬

新型冠狀病毒疫情數據與日俱增，數據科學有助我們了解疫情發展及進行防控工作。香港城市大學(城大)數據科學學院在此次疫情中，領導了多項研究及分析工作，並為疫情發展作初步研判。此次疫情數據中的各種現象，也為數據科學的應用提供範例。

數據中的異常值

【圖1】是湖北省內、外的新增確診情況，異常值出現於2月12日湖北省內及2月20日湖北省外確診數據。對應實際情況，前者因統計數據的指標改變所致，後者則因為山東和浙江等地補充通報監獄中發病人數。可見數據中的異常值不一定是剔除處理，要看實際情況。

數據統計的規範性和一致性

2月12日湖北統計數據的指標改變，也引致死亡人數上升【圖2】，2月12日之前數據分析所呈現高度線性關係亦隨之消失，但這不代表疫情本身發生劇變。因此要小心分析統計數據突變的情況。

數據的時效性

2月5日前後湖北省外累計死亡數出現兩個截然不同的線性關係【圖3】，2月5日之後死亡個案急升，這是由於病人從確診到死亡有一個周期，導致初期數據時效性偏低。

數據科學能助我們分析、了解疫情的發展。例如湖北省外新增確診情況的趨勢變化【圖4】，2月2日起，湖北省外新增確診人數呈顯著下降趨勢，可見湖北省外的疫情已開始受控。

拐點

從湖北省外數據【圖5上】看到，2月3日後新增確診人數顯著下降；從累計確診數據【圖5下】看到，2月3日數據點的切線穿越了累計病例走勢曲線。但數據上的「拐點」並非意味着疫情的拐點，放鬆警惕或讓之前的防疫努力功虧一簣。

SIR模型

模型在數據科學領域至關重要，在理解、挖掘數據方面發揮巨大作用。SIR模型是傳染病模型中的經典模型，由Kermack和McKendrick在1927年研究倫敦黑死病時提出。該模型將固定人群分成三類：

S-Susceptible (易感者)，即未得病但與感染者接觸後有機會受到感染；I-Infectious (感染者)，可以將傳染病傳播給S人群；R-Removal，指被隔離或因病痊癒而有免疫力的人群。模型基於設置的傳播率、痊癒率等參數來模擬疾病在人群中的擴散速度。傳統的模型需要大量數據才能應用。由SIR模型衍生的諸多模型便為是次防疫發揮重要作用。

是次疫情由武漢流向全國，給防疫帶來更大挑戰。張清鵬博士團隊整合人口流動數據和傳統的SIR模型，對新型冠狀病毒疫情作準確評估。【圖6-B】是截至1月22日從武漢市向全國各地人口流出人數中潛伏期和感染期人數預測，武漢流入湖北省內城市較多，估計疫情防控壓力較大，這也符合實際情況。張博士的模型同時預測全國非湖北省的感染人數【圖6-A】，而鑑於前文所說數據統計的規範性、一致性和統計指標造成數據偏差，湖北省內包括武漢市的報道感染人數遠低於預測值。

新型冠狀病毒疫情下，各省市的醫療能力備受挑戰。張博士團隊收集了從武漢流出口數排名前100城市的定點收治醫院的病床數，並利用SIR模型拓展出的SEIR-Metapopulation模型，對這些城市的感染人數作預測，從而推算出湖北省外的其他城市，如河南省信陽市、安徽省六安市等承受的醫療壓力較大，需要重點關注。(來源：<https://www.medrxiv.org/content/10.1101/2020.02.11.20022111v1>)

結合新型冠狀病毒及2003年SARS疫情的數據，模型預估在大規模封城措施下，內地地區性新型冠狀病毒疫情已在減弱，武漢市和湖北省疫情有望4月上中旬結束，非湖北地區疫情相對更早一些結束，不少非湖北省市疫情更有望3月末結束。值得一提的是，全國多區已開展復工潮，伴隨人口流動及海外輸入感染案例，疫情有反彈風險，為下一階段疫情

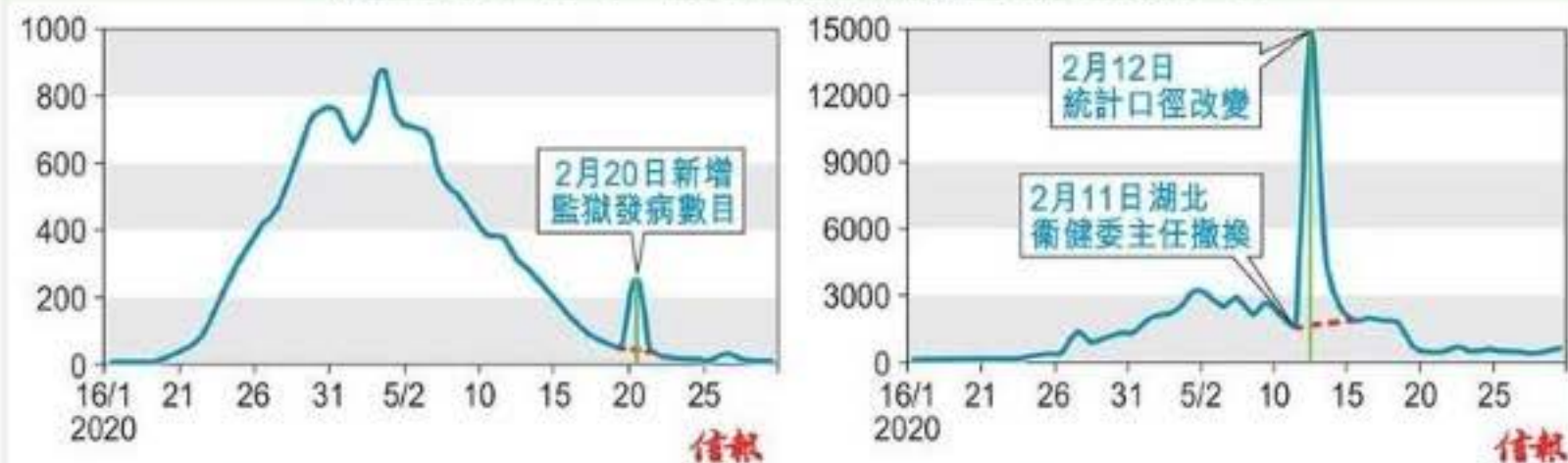
防控帶來嚴峻挑戰。

香港疫情則相對可控，尤其因本港感染案例基本可追溯來源(如近期的鑽石公主號群組、火鍋家族群組和北角佛堂群組)，即使出現社區傳染，大規模社區爆發的風險暫時不高。但切勿掉以輕心，由於全球疫情已爆發，香港作為金融中心幾乎不可避免將面臨下一輪國外輸入性傳播。數據科學被譽為「21世紀最性感的工作」，可在新型冠狀病毒疫情中發揮重

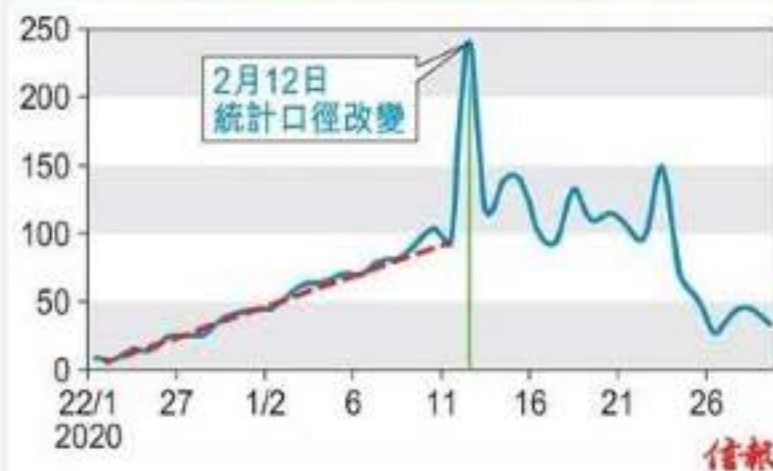
要作用。這門新學科基於各類數據和領域知識，整合諸多學科的優勢，挖掘數據背後所包含的訊息與知識，既能從數據角度分析數據差異，因地制宜，去偽存真，又能整合傳統的動力學模型，對疫情防控起指導作用。在後續的回顧性研究，數據科學可以從機器學習角度對新型冠狀病毒發展作更深入的詮釋。

香港城市大學數據科學學院院長秦泗釗教授、助理教授張清鵬博士

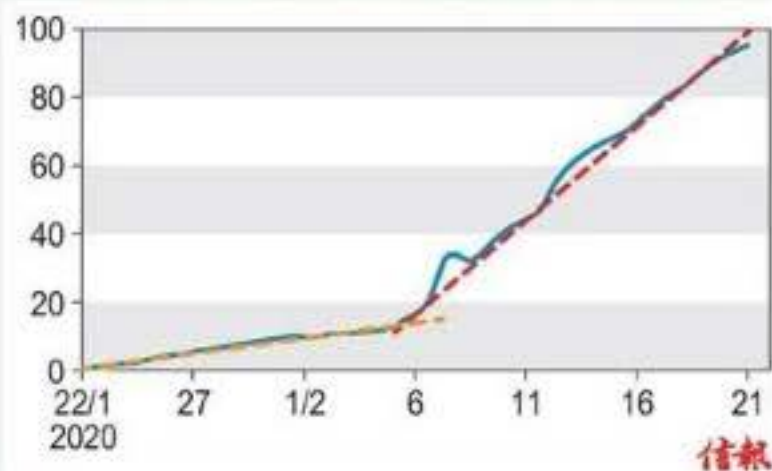
湖北省外(左)、省內(右)新增確診人數 圖1



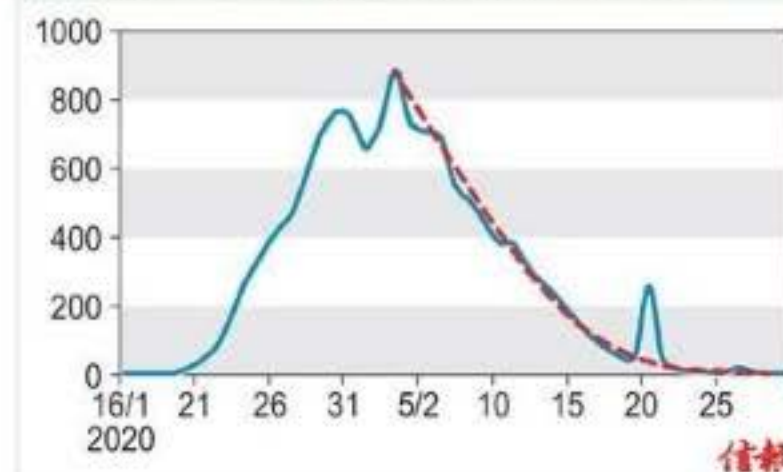
湖北新增死亡病例數變化 圖2



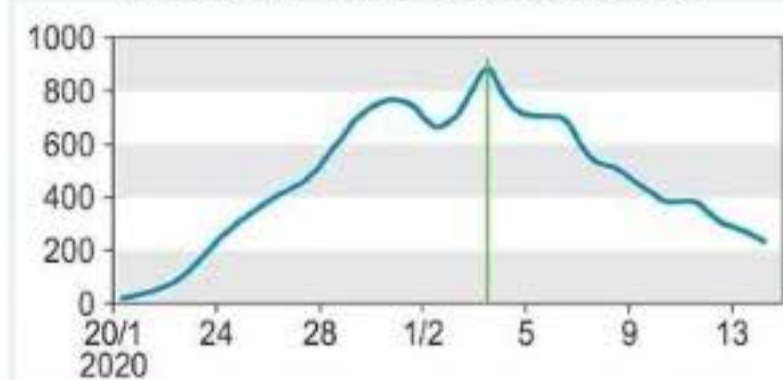
湖北省外累計死亡病例數變化 圖3



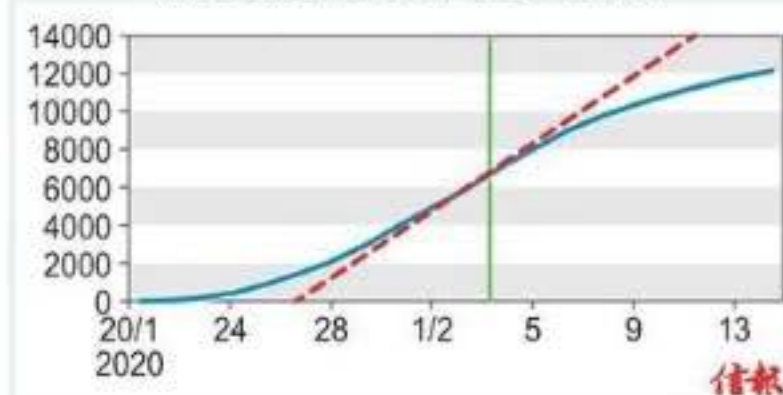
湖北省外新增確診人數 圖4



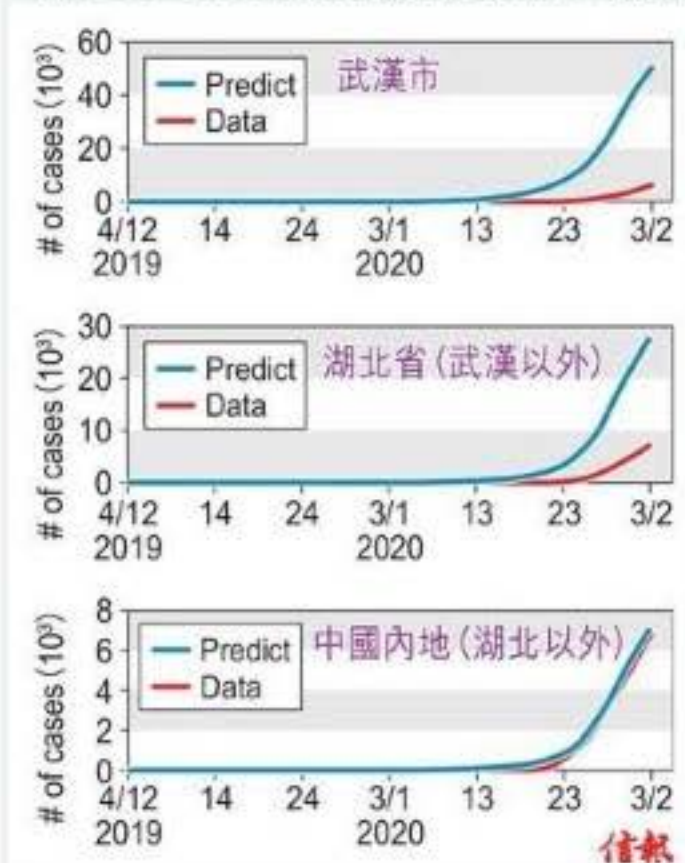
湖北省外新增確診人數 圖5



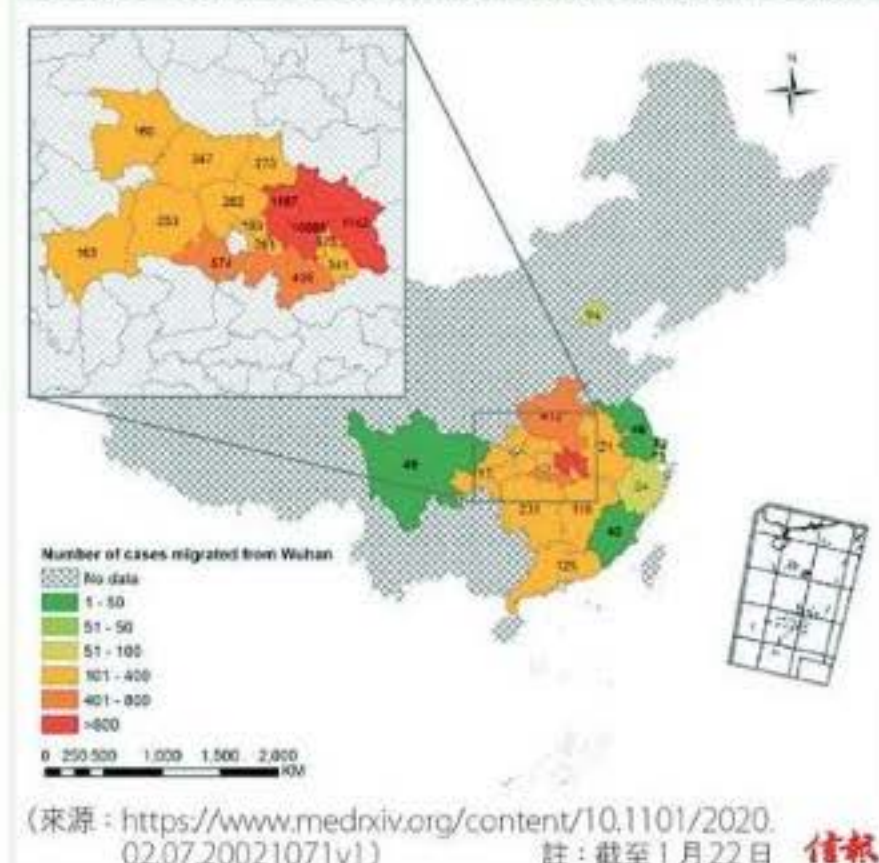
湖北省外累計確診人數



不同地區的感染人數規模預測 圖6-A



從武漢主要流出地區潛伏期和感染期人數預測 圖6-B



(來源：<https://www.medrxiv.org/content/10.1101/2020.02.07.20021071v1>) 註：截至1月22日 信報