

Covid-19: l'evoluzione spaziale dell'epidemia in Italia

Filippo Massimiliano Gagliano¹

¹I.I.S. MARIO RIGONI STERN BERGAMO, via Borgo Palazzo 128, 24125 Bergamo
tel 035220213, e-mail: fmgagliano@gmail.com

Abstract in Italiano

Il 31 dicembre 2019, le autorità sanitarie cinesi hanno notificato un focolaio di casi di polmonite ad eziologia non nota nella città di Wuhan (Provincia dell'Hubei, Cina). Un'ipotesi plausibile, ma non ancora confermata, ma molto probabile, una possibile trasmissione da animali vivi al Wuhan's South China Seafood City market. Il 9 gennaio 2020, il Centro per il controllo e la prevenzione delle malattie della Cina ha identificato un nuovo coronavirus (denominato 2019-nCoV) come causa eziologica di queste patologie. Le autorità sanitarie cinesi hanno confermato la trasmissione inter-umana del virus. L'11 febbraio 2020, l'Organizzazione Mondiale della Sanità (OMS) ha annunciato che la malattia respiratoria causata dal 2019-nCoV è stata chiamata COVID-19 (*Corona Virus Disease*). In Italia, la regione più colpita è stata la Lombardia, il numero di infezioni e decessi è aumentato rapidamente. Il COVID-19 è la più grande minaccia mondiale che sta colpendo gravemente la salute umana, i settori produttivi, la vita, il funzionamento sociale e le relazioni internazionali. Nella lotta contro COVID-19, i *Geographic Information Systems (GIS)* e le tecnologie dei *big data* hanno svolto un ruolo importante in molti aspetti, tra cui la rapida aggregazione di *big data* multi-source, la visualizzazione rapida di informazioni epidemiologiche, la georeferenziazione spaziale delle tipologie di casi da COVID-19 sia a livello mondiale e sia a livello locale, modelli di previsione della contaminazione territoriale, creazione di ipotesi di correlazioni spaziali tra i fattori ambientali e i fenomeni di diffusione dell'epidemia. Il risultato, mappe intelligenti, potrebbero essere un solido supporto per i processi decisionali durante la gestione dello stato di emergenza e post-emergenza per fronteggiare un possibile ritorno dell'epidemia. Con lo strumento GIS si è sviluppata in modo rapido una metodologia per la costruzione di piattaforme per la realizzazione di mappe fruibili via web. Tuttavia, per la lotta contro l'epidemia diffusa, la principale sfida è trovare soluzioni per migliorare le metodologie alla precisione e al grado di affidabilità delle informazioni. Nell'era dei *big data*, le fonti dei dati sono diverse, dalle informazioni del governo a quelle delle imprese. Di conseguenza, l'uso del GIS potrebbe gestire le difficoltà nell'acquisizione dei dati e l'integrazione dei dati eterogenei. Le funzionalità del GIS dovrebbero essere migliorate per l'acquisizione rapida dei dati e informazioni, migliorare l'interoperabilità e la parametrizzazione di modelli e metodi.

Contesto nazionale

Nel 2020 lo scoppio del nuovo coronavirus (COVID-19) ha scatenato l'emergenza sanitaria pubblica di maggior interesse internazionale che ha causato oltre 1.000.000 infezioni e 67.841 decessi in oltre 208 paesi entro il 7 aprile 2020 (OMS, 2020), con gravissime conseguenze economiche. In Italia ci sono

132.547 casi confermati e 16.523 decessi entro il 7 aprile 2020 (OMS, 2020). L'Italia è stato il primo paese, dopo la Cina, a dover affrontare l'epidemia da COVID-19. All'inizio dell'epidemia, le strutture ospedaliere e i comitati scientifici hanno risposto immediatamente con le comunità. Hanno rapidamente isolato il nuovo coronavirus e condotto il sequenziamento completo del genoma, per la determinazione del virus, condividendo i dati con la comunità internazionale. Ogni provincia ha successivamente lanciato un servizio di sanità pubblica di primo livello per frenare la diffusione dell'epidemia. Nel mese di dicembre 2019, sono stati rilevati un numero di casi di polmonite virale inspiegabile nelle città del nord Italia, soprattutto nella provincia di Bergamo. Successivamente l'epidemia da COVID-19 si è diffusa nel resto del mondo, diventando una pandemia. Il numero di casi italiani per regione recentemente confermati è aumentato rapidamente dall'inizio del mese di marzo (figura 1).

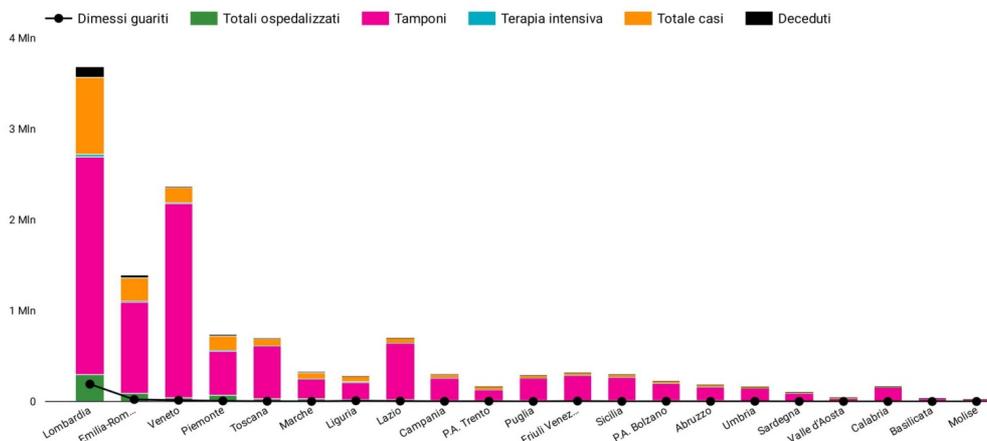


Figura 1: Grafico dei dati su scala regionale in Italia (dal 24/02/2020 al 06/04/2020).

(Fonte dati: Dipartimento della Protezione Civile, <https://github.com/pcm-dpc/COVID-19>)

Tuttavia, valutare la diffusione spaziale dell'epidemia COVID-19 è fondamentale per creare modelli previsionali locale di focolai e sviluppo di politiche di salute pubblica per le fasi di gestione dell'epidemia o di un possibile ritorno da COVID-19.

Precedenti studi hanno descritto la diffusione spaziale della sindrome respiratoria acuta grave (SARS) a Pechino, in Cina continentale (Meng et al., 2005; Fang et al., 2009, Emmanuel et al., 2019). Uno studio prende in considerazione i sistemi di trasporto tra le città per calcolare l'associazione spaziale (Meng et al., 2005; Zhao et al., 2020). Altri studi hanno analizzato i dati epidemiologici del Medio Oriente per la sindrome respiratoria coronavirus (MERS-CoV) in Arabia Saudita usando diversi approcci spaziali (Adegboye et al., 2017; Lin et al., 2018; Al-Ahmadi et al., 2019). Questo studio analizza la dinamica epidemiologica spaziale del COVID-19 in Lombardia (Italia), confrontando dati giornalieri a livello nazionale, regionale, provinciale e comunale, incrociando diverse fonti di informazioni, utilizzando diversi strumenti di rappresentazione delle informazioni territoriali. Sono state prese in considerazione diverse ipotesi di correlazione spaziale tra le province per quanto riguarda le possibili vie di diffusione di COVID-

19 (Meng et al., 2005, Roger et al., 2019). L'obiettivo di questo studio è fornire modelli per la rappresentazione delle informazioni spaziali registrate nel tempo da COVID-19 attraverso le tecnologie GIS (Dayun et al., 2020), a livello nazionale e un focus nel territorio della regione Lombardia.

Dati e metodologia

La caratteristica della forte infettività del coronavirus, un lungo e incerto periodo di incubazione e una rilevazione incerta nelle persone dal COVID-19, combinata con un territorio densamente popolato, la presenza di importanti sistemi di trasporto e altri fattori spaziali, hanno portato alla necessità urgente di supporto scientifico e tecnologico per controllare e prevenire la diffusione di l'epidemia. Durante la diffusione dell'epidemia, il GIS occupa un ruolo strategico per la gestione dei dati spaziali per la realizzazione di una banca dati, acquisizione dei big data e la gestione dell'interoperabilità dei dati, creazione di mappe nelle diverse scala territoriali, modelli di previsioni spazio-temporale, stime del tasso di contagio e dei decessi, pianificazione per la fornitura di materiali sanitario e assistenziale, monitoraggio in tempo reale dei flussi demografici.

Il primo set di dati COVID-19 è stato acquisito dal Dipartimento di Protezione Civile Nazionale (DPCN), forniti in un formato ordinato e standard con estensione "json¹". I dati sono disponibili e aggiornati ogni 24 ore nel repository ufficiale del DPCN presente nella piattaforma **GitHub**² (<https://github.com/pcm-dpc/COVID-19>). Le informazioni interessano 20 regioni e 107 province nel territorio italiano, 1.516 comuni nella regione Lombardia. I dati di studio partono dal 24 febbraio al 6 aprile 2020, interessando soprattutto le fasi di evoluzione dell'epidemia. In questo studio il GIS gioca un ruolo fondamentale per tradurre i dati e le informazioni in mappe e grafici dinamiche per l'analisi e il carattere dell'evoluzione dell'epidemia. Un sistema informativo territoriale può essere costruito rapidamente, soprattutto in termini di gestione del geo-database, in termini di elaborazione dei dati con gli strumenti di analisi e correlazione spaziale ai fini della rappresentazione e condivisione via web. In risposta all'epidemia, molte istituzioni governative e gruppi di ricerca di tutto il mondo hanno creato e condiviso una serie di informazioni e sistemi informativi per la visualizzazione di mappe epidemiche, valutando il numero dei posti delle terapie intensive, il numero di contagi, il numero dei decessi, studiando modelli di diffusione del virus, costruendo una banca dati con le caratteristiche dei contagiati, cercando di dare un contributo importante alla prevenzione delle epidemie.

Attraverso l'uso di QGIS, software open - source, utilizzato in questo studio per la creazione e l'analisi della banca dati, si è proceduto alla creazione di stringhe di connessione alla fonte dati (dpc-covid19-ita-regioni.csv) contenuta nel repository tramite protocollo HTTPS per importare e aggiornare i dati (figura 2).

¹ *JSON (JavaScript Object Notation) è un semplice formato per lo scambio di dati. Per le persone è facile da leggere e scrivere, mentre per le macchine risulta facile da generare e analizzarne la sintassi. Si basa su un sottoinsieme del Linguaggio di Programmazione JavaScript, Standard ECMA-262 Terza Edizione - Dicembre 1999.*

² *Il sito principalmente utilizzato dagli sviluppatori, che caricano il codice sorgente dei loro programmi e lo rendono scaricabile dagli utenti.*

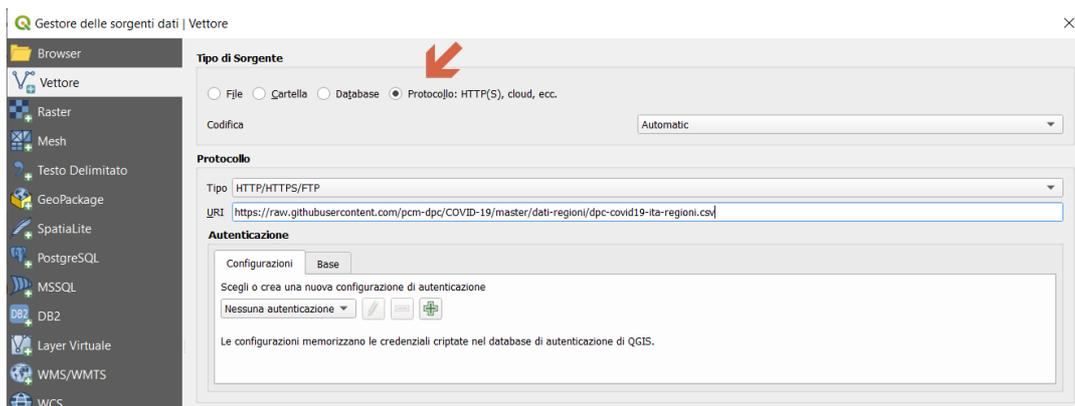


Figura 2: Ambiente QGIS desktop per la connessione dati

Attraverso l'elaborazione e la creazione della banca dati e utilizzando le funzioni di geostatistica si procede al calcolo dei valori incrementali giornalieri per essere tematizzati e rappresentati nei grafici. La figura 1 mostra un grafico dei casi per regione (guariti, ospedalizzati, con tamponi, in terapia intensiva, e deceduti), mentre la figura 3 mostra una mappa con variazione temporale giornaliera a livello regionale del numero di casi divisi per tipologia e con il calcolo incrementale del numero di casi rispetto al giorno precedenti della regione Lombardia. Il report completo delle altre regioni è consultabile al link <http://igisweb.it/covid-19/COVID19-ITA.pdf>.

Un altro parametro importante in un'epidemia di una malattia infettiva è il cosiddetto R_0 ovvero il "numero di riproduzione di base" che rappresenta il numero medio di infezioni secondarie prodotte da ciascun individuo infetto in una popolazione completamente suscettibile cioè mai venuta a contatto con il nuovo patogeno emergente. Questo parametro misura la potenziale trasmissibilità di una malattia infettiva. In altre parole se l' R_0 di una malattia infettiva è circa 2, significa che in media un singolo malato infetterà due persone. Quanto maggiore è il valore di R_0 e tanto più elevato è il rischio di diffusione dell'epidemia. Se invece il valore di R_0 fosse inferiore ad 1 ciò significa che l'epidemia può essere contenuta. Da quando l'epidemia del nuovo coronavirus (2019-nCoV) emerso in Italia ha cominciato a diffondersi e sono iniziati a circolare i dati sui primi casi confermati, l'Organizzazione Mondiale della Sanità (OMS) e numerosi istituti di ricerca di tutto il mondo hanno diffuso stime di R_0 dell'infezione. In QGIS sono state calcolate le stime di R_0 di tutte le province italiane in figura 4, per la consultazione completa delle altre province è disponibile al link <http://www.igisweb.it/covid-19/R0/provinces.png>.

Si riporta il p codice di programma scritto in python³ per il calcolo e la creazione del dataset dell'indicatore R_0 :

```
import requests
from datetime import date, timedelta
import os.path
from os import path
import csv
import datetime
```

³ Linguaggio di programmazione ad alto livello object-oriented, <https://www.python.it>.

```

sdate = date(2020, 2, 24)
edate = date(2020, 12, 31)
delta = edate - sdate
for i in range(delta.days + 1):
    date = (sdate + timedelta(days=i))
    url = date.strftime('https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-
province/dpc-covid19-ita-province-%Y%m%d.csv')
    path = date.strftime('/tmp/dpc-covid19-ita-province-%Y%m%d.csv')
    if not os.path.exists(path):
        r = requests.get(url)
        if r.status_code == 404:
            break
        open(path, 'wb').write(r.content)
    lines = []
    for i in range(delta.days + 1):
        date = (sdate + timedelta(days=i))
        path = date.strftime('/tmp/dpc-covid19-ita-province-%Y%m%d.csv')
        if not os.path.exists(path):
            break
        with open(path) as csv_file:
            csv_reader = csv.reader(csv_file, delimiter=',')
            next(csv_reader)
            for row in csv_reader:
                date = datetime.datetime.strptime(row[0], '%Y-%m-%dT%H:%M:%S')
                date_str = date.strftime('%Y-%m-%d')
                location = row[5]
                fips = 1
                cases = row[9]
                deaths = 0
                if cases.isnumeric() and "aggiornamento" not in location:
                    lines.append("{}{}{}{}{}".format(date_str, location, fips, cases, deaths))
                    lines.insert(0, "date,state,fips,cases,deaths")
            with open('/tmp/joined.csv', 'w') as f:
                for line in lines:
                    f.write("%s\n" % line)

```

La figura 5 mostra la variazione temporale del numero di casi cumulativi per provincia a livello nazionale. Per calcolare la correlazione spaziale, viene calcolato un indice spaziale di autocorrelazione I di Moran, dove i e j rappresentano gli indici spaziali e W_{ij} indicano le zone di confine tra la zona i e la zona j , mentre \bar{Y} rappresenta il valore medio dei casi confermati (figura 6).

$$I = \frac{n \sum_{ij} W_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i \neq j} W_{ij} (Y_i - \bar{Y})^2}$$

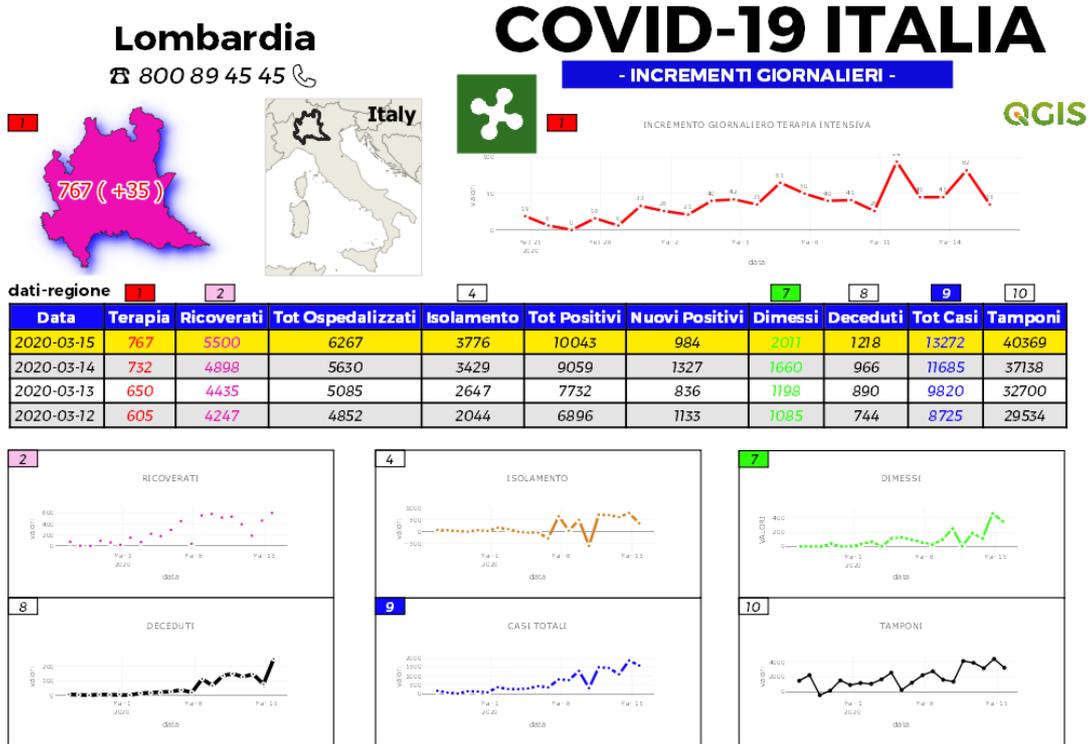


Figura 3: Valori giornalieri della regione Lombardia

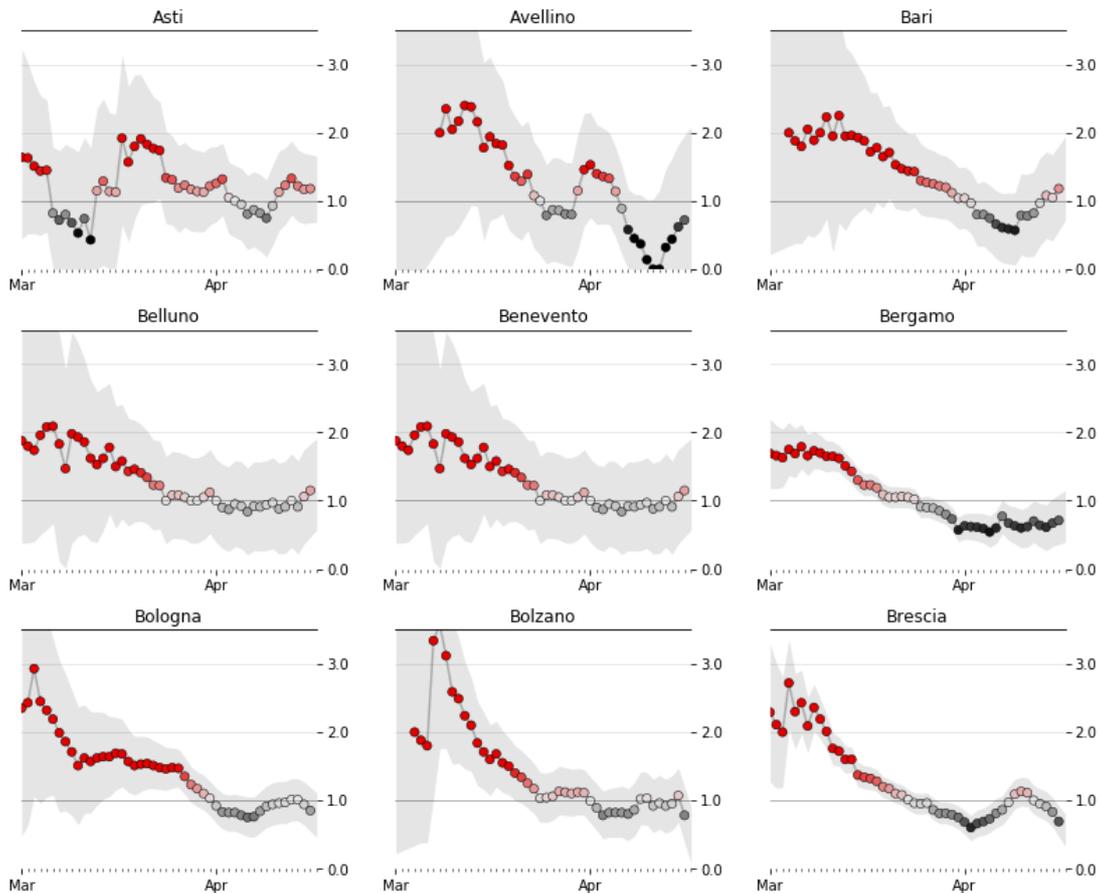


Figura 4: Valori dell'indice R0

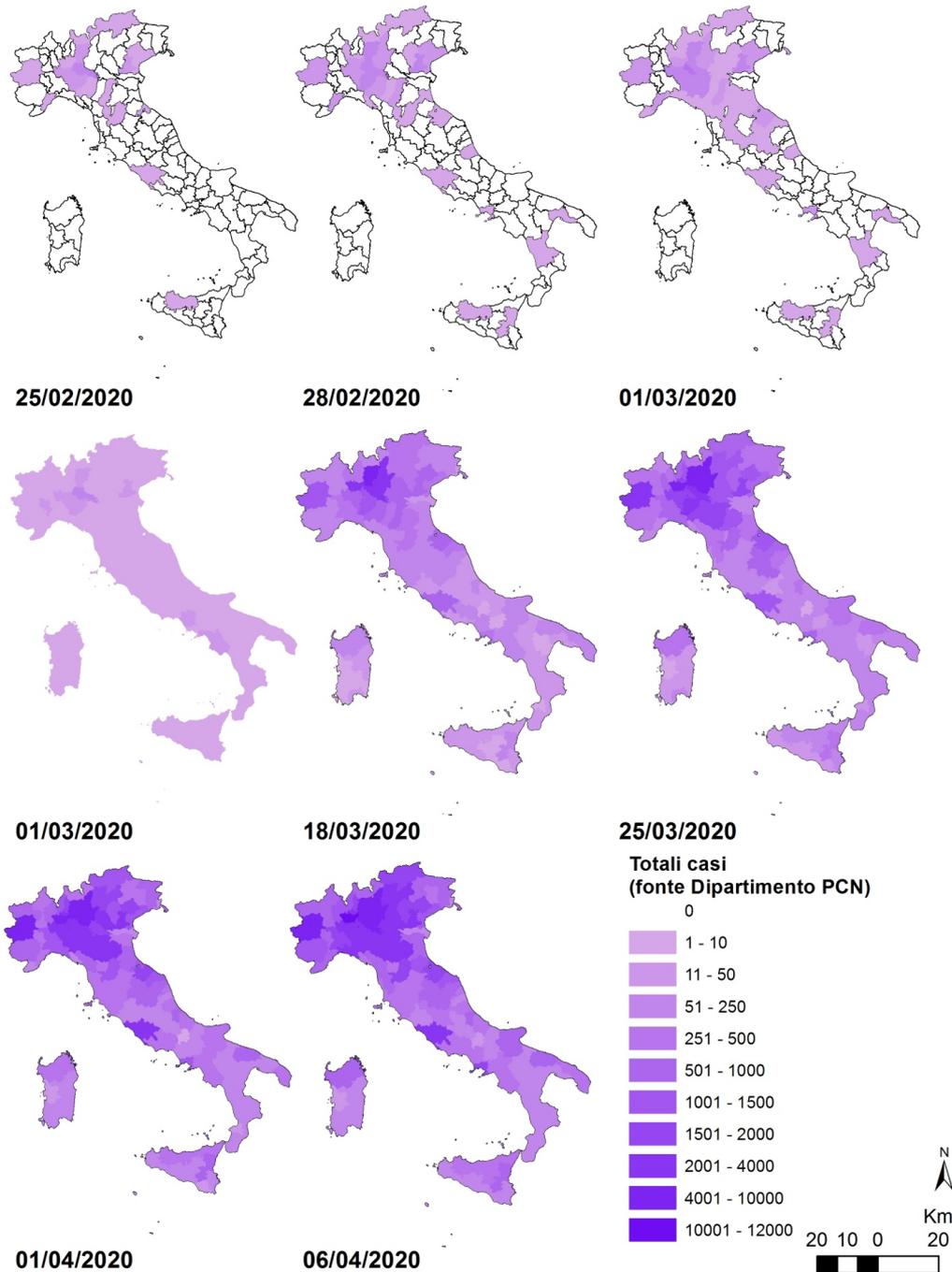


Figura 5: Andamento temporale dei casi totali affetti da COVID-19 per regione su scala nazionale

Per valori pari a 0 indica che non esiste spazio autocorrelazione nei dati, per valori di I positivi di Moran indicano il raggruppamento di valori simili, mentre un valore I negativo indica il raggruppamento di valori diversi. Più grande è il valore assoluto di Moran, più forte è l'autocorrelazione spaziale (figura 5). Il report completo del calcolo dell'indice I è disponibile al link <http://www.igisweb.it/covid-19/morans.html>.

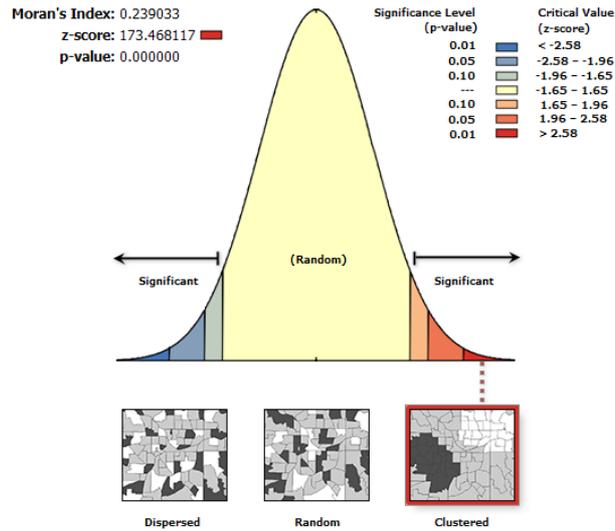


Figura 6: Indice di Moran I.

Dal punteggio z di 173.468117465, esiste una probabilità inferiore all'1% che questo modello cluster possa essere il risultato di una casualità. Successivamente si è analizzato tutto il territorio comunale della regione Lombardia, raccogliendo i dati dei decessi ufficiali, dei decessi dall'ufficio anagrafe e inserendo il dato ISTAT. Costruendo la banca dati regionale (figura 7), si è proceduto al calcolo del tasso di contagio della popolazione in percentuale (figura 8).

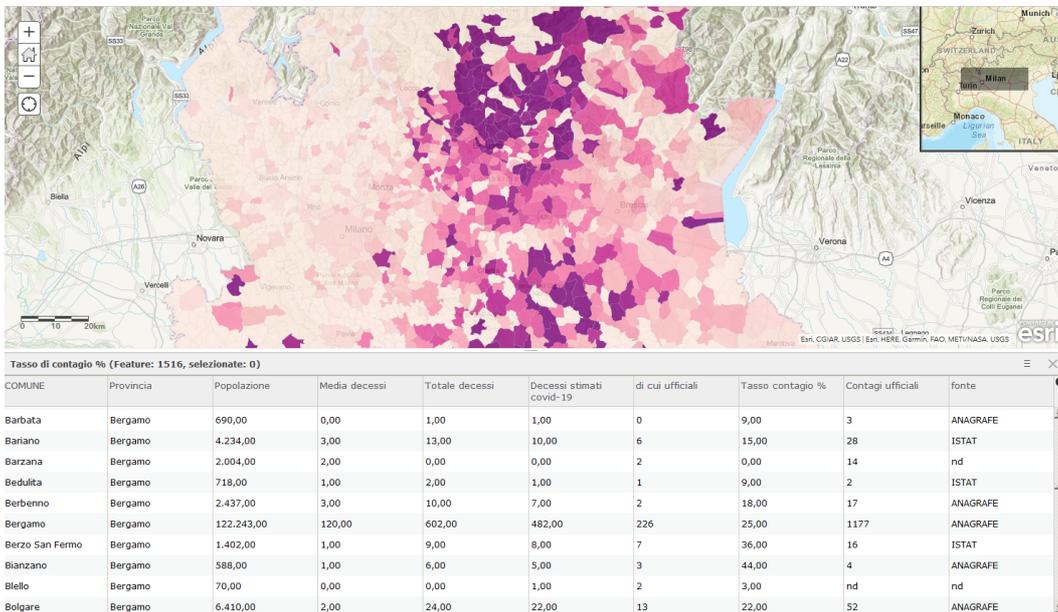


Figura 7: Banca dati COVID-19 regione Lombardia

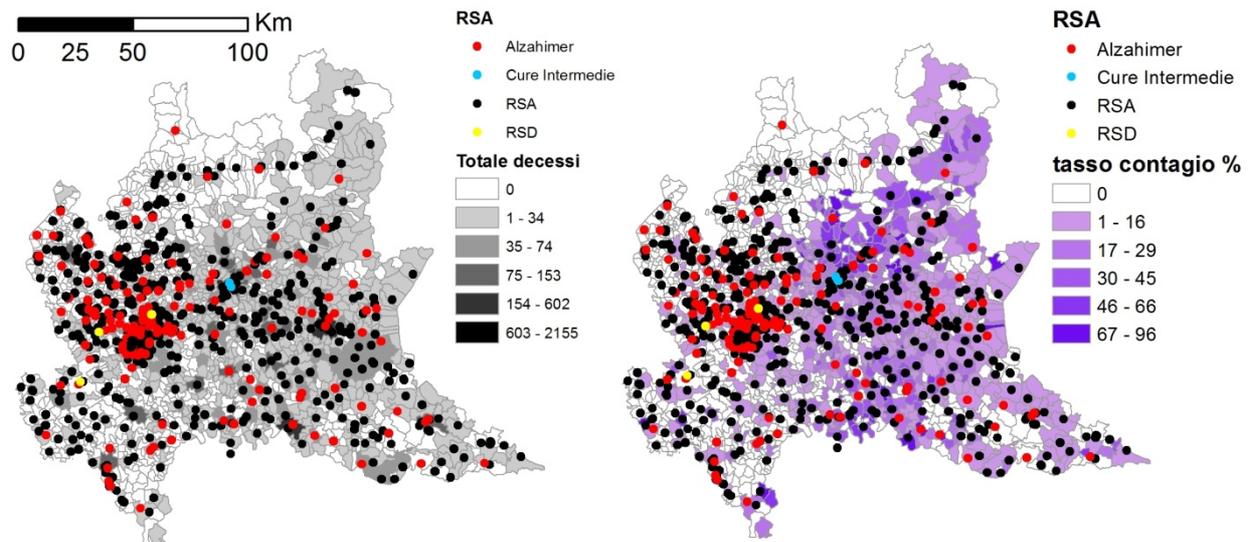


Figura 8: Tasso di contagio (%) e numero di decessi dal 24 febbraio al 6 aprile 2020

La banca dati dei decessi a livello regionale è stata costruita considerando i riferimenti dell'anagrafe comunale e incrociandoli con i dati ISTAT di aprile 2020. Dall'analisi delle mappe, il tasso di contagio registrato è maggior nei comuni della provincia di Bergamo e Cremona, mentre il numero di decessi prevale nella provincia di Milano. Per valutare possibili correlazioni spaziali si georeferenziano i dati delle strutture RSA (Residenze Sanitarie Assistenziali) e dei presidi ospedalieri della regione lombarda. L'indice I calcolato a livello regionale sul tasso di mortalità è pari a 0.122719, questo dimostra che la prossimità spaziale del tasso di contagio ha contribuito solo per una parte nelle provincie della regione lombarda, il restante contagio con una alta percentuale di mortalità è correlato con le strutture RSA e gli ospedali.

Risultati

Dalla banca dati si procede alla creazione delle piattaforme per rappresentare e condividere i dati sul web con la tecnologia GIS. Per rappresentare e gestire la serie temporali dei casi nuovi confermati dall'inizio dell'epidemia su tutto il territorio nazionale, si è costruita una time map (figura 9). Analizziamo la diffusione esponenziale spaziale nelle prime fasi in l'Italia: l'epicentro del scoppio rimane la provincia di Bergamo, il numero di casi recentemente confermati nelle altre provincie limitrofi è cresciuto in maniera esponenziale. Le provincie di Bergamo, Milano e Brescia sono state confermate come aree rappresentative della prima fase. Il numero giornaliero di casi confermati è mostrato nel pannello a sinistra (figura 10). L'analisi a livello regionale è disponibile al link http://www.igisweb.it/covid-19/covid_lombardia.html.

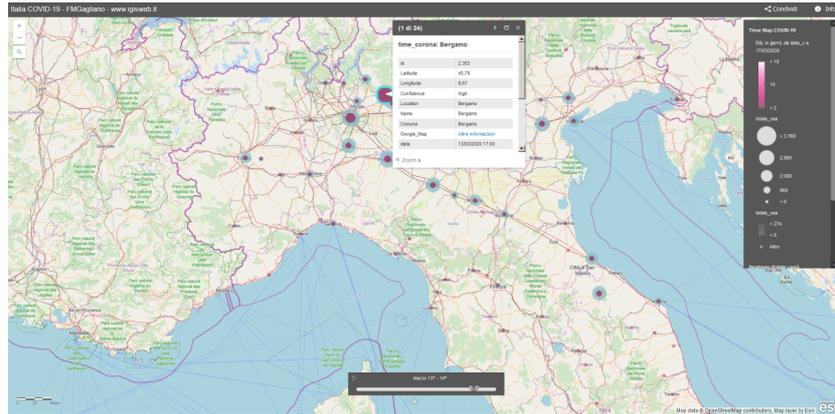


Figura 9: Time Map dei casi confermati da COVID-19

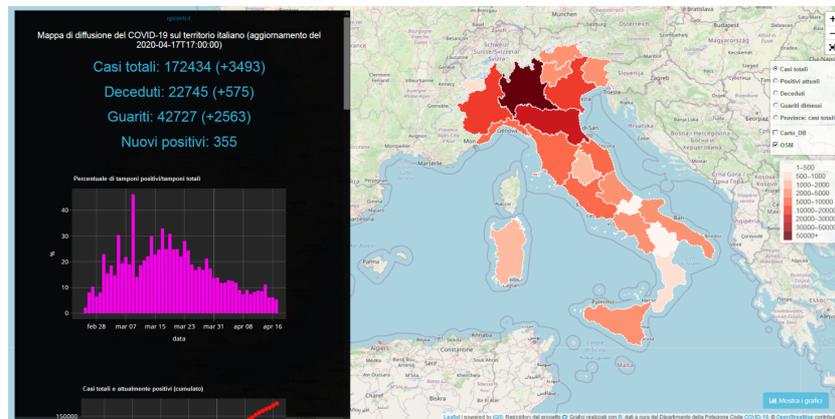


Figura 10: WebGIS per gestione in tempo reale della serie temporali dei dati COVID-19 (<http://www.igisweb.it/covid-19/italia/webmap/>)

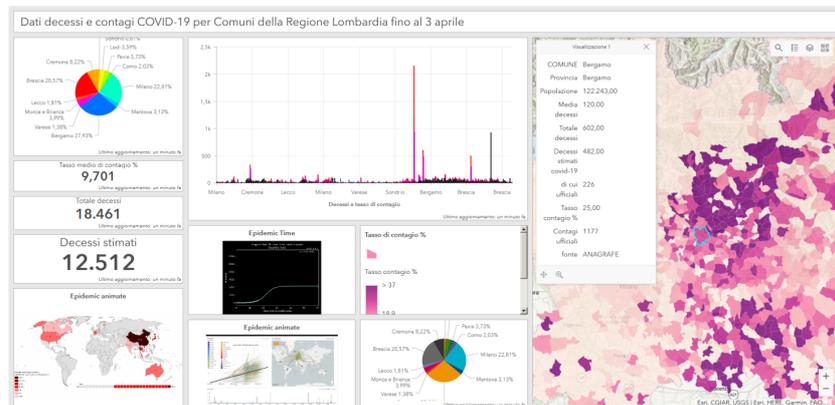


Figura 11: WebGIS del territorio di regione Lombardia (http://www.igisweb.it/covid-19/covid_lombardia.html)

Conclusioni

Esaminare la diffusione spaziale nelle prime fasi è molto importante per impedire un'ulteriore trasmissione. Questo studio cerca di fornire modelli spazio-temporali della pandemia COVID-19 in Italia e in Lombardia, la regione più colpita. Questo studio dimostra la correlazione spaziale tra le province attraverso l'indice I di Moran, soprattutto nelle zone con un alto indice di densità di popolazione. Altra considerazione, il tasso alto di contagio nelle zone in presenza di strutture sanitarie, che viene fuori da un'analisi dettagliata descritta prima, risultato

coerente studi precedenti (Meng et al., 2005). Tuttavia, è importante investire nelle analisi geostatistiche per lo studio dell'evoluzione dei fenomeni da epidemia, per meglio gestire il controllo e la prevenzione di questa malattia e stabilirne i fattori di contagio.

Bibliografia

Adegboye O.A., Gayawan E., Hanna F. (2020), "Spatial modelling of contribution of individual level risk factors for mortality from Middle East respiratory syndrome coronavirus in the Arabian Peninsula", *PLoS One*, 12(7): e0181215, doi: <http://dx.doi.org/10.1371/journal.pone.0181215>.

Al-Ahmadi K., Alhmadi S., Al-Zahrani A. (2012–2019), "Spatiotemporal clustering of Middle East respiratory syndrome coronavirus (MERS-CoV) incidence in Saudi Arabia", *Int J Environ Res Public Health* 2019, 16(14):2520.

Dayun K., Hyunho C., Jong-Hun K., Jungsoon C.D. (2020), "Spatial epidemic dynamics of the COVID-19 outbreak in China: International Journal of Infectious Diseases", 96–102

Paradis E., Blomberg S., Bolker B., Brown J., Claude J., Cuong H.S., et al. (2019), "Analyses of Phylogenetics and Evolution", Available from: <https://CRAN.R-project.org/package=ape>. [Accessed 13 April 2020].

Fang L.Q., Vlas S.J., Feng D., Liang S., Xu Y.F., Zhou J.P., et al. (2009), "Geographical spread of SARS in mainland China", *Trop Med Int Health*, 14:14–20.

Li H., Calder C.A., Cressie N. (2018) "Testing for spatial dependence based on the spatial autoregressive model", *Geographic Analysis* 2007;39(4):357–75.

Lin Q., Chiu A.P., Zhao S., He D. (2018), "Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia", *Stat Methods Med Res* 2018, 27(7): 1968–78.

Meng B., Wang J., Liu J., Wu J., Zhong E. (2005), "Understanding the spatial diffusion process of severe acute respiratory syndrome in Beijing", *Public Health*, 119 (12):1080–7.

Bivand R., Rundel C., Pebesma E., Stuetz R., Hufthammer K.O., Giraudoux P., et al. (2019), "Interface to Geometry Engine". Available from: <https://CRAN.R-project.org/package=rgeos>. [Accessed 13 April 2020].

Zhao S., Zhuang Z., Ran J., Lin J., Yang G., Yang L., et al. (2020), "The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020", *Travel Med Infect Dis*, 101568, doi:<http://dx.doi.org/10.1016/j.tmaid.2020.101568>. <http://statista.comStatista.2020> (accessed 21 February 2020).

Sitografia

- ∞ Centro Europeo per la Prevenzione e il Controllo delle Malattie
<https://gap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html>
- ∞ Confini amministrativi dell' ISTAT al 1 gennaio 2019:
<https://www.istat.it/it/archivio/222527>
- ∞ Fonti dati PCM-DPC: <https://github.com/pcm-dpc/COVID-19>
- ∞ ISTAT: <https://www.istat.it/>
- ∞ Istituto Superiore della Sanità:
<https://www.epicentro.iss.it/coronavirus/sars-cov-2-sorveglianza-dati>
- ∞ Plugin DataPlotly: <https://github.com/ghtmtt/DataPlotly>
- ∞ Protezione civile
<http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1>
- ∞ QGIS: <https://qgis.org/it/site/>
- ∞ Regione Lombardia: <https://www.regione.lombardia.it/>
- ∞ Visual Studio Code : <https://code.visualstudio.com/>
- ∞ Visual Style Guide :
<https://www.qgis.org/en/site/getinvolved/styleguide.html#trueno-fonts>
- ∞ WHO: <https://www.who.int/emergencies/diseases/novel-coronavirus2019/situation-reports>
- ∞ Word Map: <https://www.naturalearthdata.com/downloads/10m-cultural-vectors/>