

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



Caratterizzazione dell'evoluzione temporale e spaziale dell'epidemia da SARS-CoV-2 in Italia

Relatore: Prof.ssa Alessandra R. Brazzale
Dipartimento di Scienze Statistiche

Correlatore: Dott. Andrea Sottosanti
Dipartimento di Medicina

Laureando Marco Zambito
Matricola N. 2023835

Anno Accademico 2023/2024

Elenco delle figure

1	Mappa geografica dell'Italia, divisa in regioni tramite i confini politici	14
2	Cronoprogramma che illustra la finestra temporale di collezione dei dati	27
3	Nuovi positivi cumulati e incidenza durante la prima ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda	29
4	Nuovi positivi cumulati e incidenza durante la seconda ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda	30
5	Nuovi positivi cumulati e incidenza durante la terza ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda	31
6	Nuovi positivi cumulati e incidenza durante la quarta ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda	32
7	Nuovi positivi cumulati e incidenza durante la quinta ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda	33
8	Grafo non orientato dei centroidi delle regioni italiane	35
9	Indici di Moran locali per ogni regione; quarta ondata	37
10	Serie temporale giornaliera dei nuovi positivi a livello nazionale	38
11	Serie temporale giornaliera della prevalenza a livello nazionale	38
12	Serie temporale giornaliera dei nuovi positivi a livello regionale	40
13	Serie temporale giornaliera della prevalenza a livello regionale	41
14	Nuovi positivi giornalieri cumulati	42
15	Stima dei nuovi positivi giornalieri in Lombardia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	43
16	Stima della mediana del trend lineare α_t in Lombardia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	44

17	Stima della mediana della pendenza β_t del trend lineare in Lombardia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	44
18	Stima DGLM dell'incidenza giornaliera in Molise, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata.	45
19	Stima DGLM della mediana del trend lineare α_t in Molise, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata.	46
20	Stima DGLM della mediana della pendenza β_t del trend lineare in Molise, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata.	46
21	Stima di R_t in Lombardia, con intervallo di confidenza al 95%; quarta ondata	47
22	Media a posteriori del trend temporale stimato; l'area intorno al trend rappresenta il suo intervallo di credibilità al 95%	49
23	Media a posteriori degli effetti spaziali stimati	50
24	Indici di Moran locali per ogni regione; prima ondata	54
25	Indici di Moran locali per ogni regione; seconda ondata	55
26	Indici di Moran locali per ogni regione; terza ondata	55
27	Indici di Moran locali per ogni regione; quinta ondata	56
28	Stima dei nuovi positivi giornalieri in Abruzzo tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	56
29	Stima dei nuovi positivi giornalieri in Basilicata tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	57
30	Stima dei nuovi positivi giornalieri in Calabria tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	57
31	Stima dei nuovi positivi giornalieri in Campania tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	58
32	Stima dei nuovi positivi giornalieri in Emilia-Romagna tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	58
33	Stima dei nuovi positivi giornalieri in Friuli-Venezia Giulia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	59
34	Stima dei nuovi positivi giornalieri nel Lazio tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	59
35	Stima dei nuovi positivi giornalieri in Liguria tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	60

36	Stima dei nuovi positivi giornalieri nelle Marche tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	60
37	Stima dei nuovi positivi giornalieri in Molise tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	61
38	Stima dei nuovi positivi giornalieri a Bolzano tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	61
39	Stima dei nuovi positivi giornalieri a Trento tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	62
40	Stima dei nuovi positivi giornalieri in Piemonte tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	62
41	Stima dei nuovi positivi giornalieri in Puglia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	63
42	Stima dei nuovi positivi giornalieri in Sardegna tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	63
43	Stima dei nuovi positivi giornalieri in Sicilia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	64
44	Stima dei nuovi positivi giornalieri in Toscana tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	64
45	Stima dei nuovi positivi giornalieri in Umbria tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	65
46	Stima dei nuovi positivi giornalieri in Valle d'Aosta tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	65
47	Stima dei nuovi positivi giornalieri in Veneto tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata	66
48	Stima della mediana del trend lineare α_t in Abruzzo tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	66
49	Stima della mediana del trend lineare α_t in Basilicata tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	67
50	Stima della mediana del trend lineare α_t in Calabria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	67

51	Stima della mediana del trend lineare α_t in Campania tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	68
52	Stima della mediana del trend lineare α_t in Emilia-Romagna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	68
53	Stima della mediana del trend lineare α_t in Friuli-Venezia Giulia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	69
54	Stima della mediana del trend lineare α_t nel Lazio tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	69
55	Stima della mediana del trend lineare α_t in Liguria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	70
56	Stima della mediana del trend lineare α_t nelle Marche tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	70
57	Stima della mediana del trend lineare α_t in Molise tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	71
58	Stima della mediana del trend lineare α_t a Bolzano tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	71
59	Stima della mediana del trend lineare α_t a Trento tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	72
60	Stima della mediana del trend lineare α_t in Piemonte tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	72
61	Stima della mediana del trend lineare α_t in Puglia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	73
62	Stima della mediana del trend lineare α_t in Sardegna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	73
63	Stima della mediana del trend lineare α_t in Sicilia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	74
64	Stima della mediana del trend lineare α_t in Toscana tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	74
65	Stima della mediana del trend lineare α_t in Umbria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	75

66	Stima della mediana del trend lineare α_t in Valle d'Aosta tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	75
67	Stima della mediana del trend lineare α_t in Veneto tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	76
68	Stima della mediana della pendenza β_t del trend lineare in Abruzzo tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	76
69	Stima della mediana della pendenza β_t del trend lineare in Basilicata tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	77
70	Stima della mediana della pendenza β_t del trend lineare in Calabria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	77
71	Stima della mediana della pendenza β_t del trend lineare in Campania tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	78
72	Stima della mediana della pendenza β_t del trend lineare in Emilia-Romagna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	78
73	Stima della mediana della pendenza β_t del trend lineare in Friuli Venezia Giulia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	79
74	Stima della mediana della pendenza β_t del trend lineare nel Lazio tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	79
75	Stima della mediana della pendenza β_t del trend lineare in Liguria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	80
76	Stima della mediana della pendenza β_t del trend lineare nelle Marche tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	80
77	Stima della mediana della pendenza β_t del trend lineare in Molise tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	81
78	Stima della mediana della pendenza β_t del trend lineare a Bolzano tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	81
79	Stima della mediana della pendenza β_t del trend lineare a Trento tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	82
80	Stima della mediana della pendenza β_t del trend lineare in Piemonte tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	82

81	Stima della mediana della pendenza β_t del trend lineare in Puglia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	83
82	Stima della mediana della pendenza β_t del trend lineare in Sardegna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	83
83	Stima della mediana della pendenza β_t del trend lineare in Sicilia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	84
84	Stima della mediana della pendenza β_t del trend lineare in Toscana tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	84
85	Stima della mediana della pendenza β_t del trend lineare in Umbria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	85
86	Stima della mediana della pendenza β_t del trend lineare in Valle d'Aosta tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	85
87	Stima della mediana della pendenza β_t del trend lineare in Veneto tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata	86
88	Stima di R_t in Abruzzo, con intervallo di confidenza al 95%; quarta ondata	86
89	Stima di R_t in Basilicata, con intervallo di confidenza al 95%; quarta ondata	87
90	Stima di R_t in Calabria, con intervallo di confidenza al 95%; quarta ondata	87
91	Stima di R_t in Campania, con intervallo di confidenza al 95%; quarta ondata	88
92	Stima di R_t in Emilia-Romagna, con intervallo di confidenza al 95%; quarta ondata	88
93	Stima di R_t in Friuli-Venezia Giulia, con intervallo di confidenza al 95%; quarta ondata	89
94	Stima di R_t nel Lazio, con intervallo di confidenza al 95%; quarta ondata	89
95	Stima di R_t in Liguria, con intervallo di confidenza al 95%; quarta ondata	90
96	Stima di R_t nelle Marche, con intervallo di confidenza al 95%; quarta ondata	90
97	Stima di R_t in Molise, con intervallo di confidenza al 95%; quarta ondata	91
98	Stima di R_t a Bolzano, con intervallo di confidenza al 95%; quarta ondata	91
99	Stima di R_t a Trento, con intervallo di confidenza al 95%; quarta ondata	92

100	Stima di R_t in Piemonte, con intervallo di confidenza al 95%; quarta ondata	92
101	Stima di R_t in Puglia, con intervallo di confidenza al 95%; quarta ondata	93
102	Stima di R_t in Sardegna, con intervallo di confidenza al 95%; quarta ondata	93
103	Stima di R_t in Sicilia, con intervallo di confidenza al 95%; quarta ondata	94
104	Stima di R_t in Toscana, con intervallo di confidenza al 95%; quarta ondata	94
105	Stima di R_t in Umbria, con intervallo di confidenza al 95%; quarta ondata	95
106	Stima di R_t in Valle d'Aosta, con intervallo di confidenza al 95%; quarta ondata	95
107	Stima di R_t in Veneto, con intervallo di confidenza al 95%; quarta ondata	96

Elenco delle tabelle

1	Percentuale dei dati mancanti per ogni variabile	27
2	Indici di Moran globali per le diverse ondate di Covid-19	35
3	Indici di Moran locali per ogni regione; quarta ondata	36
4	Medie a posteriori degli effetti fissi	48
5	Medie a posteriori stimate degli effetti spaziali per regione, con intervallo di credibilità al 95%	51

1 Introduzione

1.1 Caratterizzazione di un'epidemia

Negli ultimi decenni, la crescente interconnessione globale e i cambiamenti demografici hanno reso le epidemie un fenomeno sempre più frequente e di grande impatto sulla salute pubblica e sull'economia mondiale. La recente pandemia di COVID-19 ha evidenziato quanto sia cruciale comprendere e monitorare la diffusione di malattie infettive per poter adottare tempestive e appropriate misure di contenimento.

Questa tesi si propone di analizzare e caratterizzare l'epidemia da SARS-CoV-2 utilizzando diversi strumenti statistici, con l'obiettivo di fornire una comprensione dettagliata dei meccanismi di diffusione e degli effetti delle varie misure di controllo. La statistica applicata alle epidemie, nota come epidemiologia, permette di modellare la dinamica della trasmissione, stimare alcune quantità importanti come il tasso di riproduzione R_t e valutare l'efficacia delle strategie di intervento.

Il lavoro si sviluppa attraverso diverse fasi, partendo dalla raccolta e dalla pulizia dei dati, passando per l'analisi esplorativa, fino alla costruzione di modelli di regressione. Verranno utilizzati metodi statistici per analizzare l'aspetto temporale e l'aspetto spaziale dei dati osservati in Italia per garantire una caratterizzazione completa dell'epidemia in esame.

L'epidemia oggetto di studio sarà analizzata non solo in termini di diffusione temporale e spaziale, ma anche considerando variabili che possono influenzare il rischio di infezione e la gravità dell'impatto sulla popolazione. Questo approccio multidimensionale è essenziale per individuare quali sono le zone maggiormente a rischio durante il periodo considerato e per mostrare tecniche statistiche volte ad analizzare questa tipologia di dati.

La tesi si conclude con la discussione e l'interpretazione dei risultati ottenuti, proponendo, inoltre, raccomandazioni per migliorare le analisi in essere.

1.2 L'epidemia da Covid-19 nel mondo

La pandemia di Covid-19, causata dal nuovo coronavirus SARS-CoV-2, è emersa per la prima volta a Wuhan, in Cina, alla fine del 2019. La malattia si è rapidamente diffusa a livello globale, portando l'Organizzazione Mondiale della Sanità (OMS) a dichiarare lo stato di pandemia l'11 marzo 2020. Il COVID-19 ha colpito quasi tutti i paesi del mondo, con gravi conseguenze per la salute pubblica e l'economia globale.

Le misure di contenimento adottate dai vari paesi hanno incluso lockdown, distanziamento sociale, chiusura delle scuole e delle attività non essenziali, e la promozione dell'igiene personale e dell'uso di mascherine. La velocità e l'efficacia di queste misure hanno variato notevolmente tra le diverse nazioni, influenzando il tasso di infezione e la mortalità.

Lo sviluppo e la distribuzione dei vaccini contro il SARS-CoV-2 hanno rappresentato una svolta cruciale nella lotta contro la pandemia. A partire dalla fine del 2020, vari vaccini hanno ricevuto l'autorizzazione per l'uso di emergenza e sono stati somministrati su larga scala, contribuendo a ridurre significativamente i tassi di infezione e di mortalità nei paesi con alta copertura vaccinale.

1.3 L'epidemia da Covid-19 in Italia

In Italia, il primo caso di Covid-19 è stato confermato il 30 gennaio 2020. Tuttavia, è stato alla fine di febbraio che il paese ha iniziato a sperimentare una rapida crescita dei casi, concentrata inizialmente nelle regioni settentrionali, in particolare Lombardia, Veneto ed Emilia-Romagna.

L'Italia è stato uno dei primi paesi occidentali a essere gravemente colpito dalla pandemia, con il sistema sanitario rapidamente sovraccaricato. In risposta, il governo italiano ha imposto rigide misure di lockdown a partire dal marzo 2020, chiudendo scuole, negozi e attività non essenziali, e limitando la mobilità delle persone.

Le misure di contenimento hanno portato a una riduzione dei casi durante l'estate del 2020, ma una seconda ondata ha colpito duramente il paese nell'autunno dello stesso anno. La risposta del governo ha incluso ulteriori lockdown, la promozione del telelavoro e il potenziamento del sistema sanitario.

La campagna vaccinale in Italia è iniziata a dicembre 2020, con priorità agli operatori sanitari e alle categorie più vulnerabili. Nel corso del 2021, la copertura vaccinale è aumentata significativamente, contribuendo a un controllo più efficace della diffusione del virus e a una riduzione dei decessi.

L'Italia ha affrontato numerose sfide durante la pandemia, tra cui il coordinamento tra le diverse regioni e la gestione delle risorse sanitarie. Tuttavia, la risposta alla crisi ha anche messo in luce la resilienza del paese e la capacità di adattamento di fronte a un'emergenza senza precedenti.

1.4 Analisi longitudinale

Le analisi longitudinali, anche note come studi di coorte, sono fondamentali in epidemiologia per comprendere l'evoluzione temporale delle malattie e i loro fattori di rischio, seguendo lo stesso gruppo di individui nel tempo e raccogliendo dati ripetuti a intervalli regolari. Questo approccio è particolarmente efficace per indagare le dinamiche di infezioni come il Covid-19, poiché consente di tracciare la progressione della malattia, valutare l'impatto delle politiche di intervento e identificare i fattori che influenzano la suscettibilità e la diffusione del virus.

Tramite metodi di analisi esplorative e di modellazione, è possibile rappresentare e interpretare l'incidenza e la prevalenza della malattia, identificare le regioni maggiormente colpite dal virus e valutare l'efficacia delle misure di contenimento dell'epidemia, come i lockdown e le campagne di vaccinazione.

Le analisi spazio-temporali permettono di integrare le dimensioni geografiche e temporali dei dati, fornendo una visione più dettagliata delle dinamiche dell'epidemia. Attraverso metodi di analisi esplorative e di modellazione appartenenti all'ambito della geostatistica, è possibile identificare aree geografiche con tassi di contagio particolarmente elevati e interpretare la variazione del rischio relativo in relazione a variabili relative al contenimento dell'epidemia, come l'isolamento domiciliare o la somministrazione di vaccini.

1.5 Studi statistici sull'epidemia da SARS-CoV-2

Gli studi statistici sul COVID-19 hanno adottato una vasta gamma di approcci per comprendere e modellare la diffusione del virus e l'efficacia delle misure di controllo. Tra i modelli più utilizzati vi sono i modelli SIR (*Susceptible-Infectious-Recovered*), che suddividono la popolazione in compartimenti e descrivono le transizioni tra questi stati nel tempo. Questi modelli si basano sul calcolo di equazioni differenziali per ciascun compartimento per stimare il numero di soggetti in ogni compartimento. Questi modelli hanno permesso di stimare parametri chiave come il tasso di riproduzione R_0 [5, 10, 16, 20] e di prevedere l'andamento delle infezioni, fornendo una base solida per le decisioni di sanità pubblica. Alcuni esempi sono forniti dagli studi di Grimm et al.(2021)[14], di Grace Y. et al.[12] e di Coffeng et al.(2022)[9]. Il primo studio [14] propone l'utilizzo del modello SEIR(*Susceptible-Exposed-Infectious-Recovered*) come estensione del modello SIR, aggiungendo la componente degli "esposti" alle equazioni differenziali del modello. Il secondo studio [12] estende a sua volta il modello SEIR, proponendo il modello SEASAR (*Susceptible-Exposed-Asymptomatic-Symptomatic-Active-Recovered*). Il punto principale di questa proposta è espandere ancora di più il modello matematico sottostante, dividendo gli infetti in asintomatici, sintomatici e "attivi" (soggetti ancora positivi ma senza capacità infettiva). Il terzo studio [9] muove una critica a tali modelli, ritenendoli troppo "pessimistici": infatti, Coffeng afferma che il modello SEIR sovrastima il tasso di contagio, ignorando l'eterogeneità relativa alla trasmissione del virus.

Le reti neurali, come ad esempio le Reti Neurali Artificiali[19], sono state impiegate per catturare le complessità temporali e le non linearità nella diffusione del virus. Questi modelli hanno dimostrato una notevole capacità di previsione, adattandosi dinamicamente ai dati in arrivo e migliorando continuamente le proprie stime grazie alla loro natura di apprendimento automatico. Il lavoro di Mollalo et al.(2020)[19] si muove in questa direzione, con lo scopo di stimare e prevedere i tassi di incidenza del Covid-19, tenendo conto delle correlazioni fra fattori socio-economici, ambientali e fattori di comorbidità (presenza contemporanea di più malattie).

I modelli Bayesiani hanno fornito un quadro probabilistico per la stima dei parametri epidemici e l'analisi dell'incertezza. Attraverso l'uso delle tecniche di inferenza Bayesiana, è stato possibile integrare dati provenienti da diverse fonti e incorporare priori informative per migliorare la robustezza delle stime. Questi modelli hanno anche facilitato la valutazione dell'impatto delle misure di intervento, offrendo una visione completa dell'evoluzione dell'epidemia sotto vari scenari. Grande et al.(2023) [13] propongono un approccio Bayesiano per combinare la flessibilità dei modelli lineari generalizzati (GLM) con i modelli autoregressivi a media mobile (ARMA). L'approccio di modellazione Bayesiana è stato utile anche per studiare malattie precedenti al Covid-19: è il caso dello studio di Martinez-Bello et al.(2017) [18] sulla "febbre dengue", che si incentra sull'adattamento di modelli Poisson dinamici per tenere conto dei coefficienti variabili nel tempo relativi a variabili meteorologiche. Chiogna e Gaetan, negli studi del 2004 e del 2010[7, 6] utilizzano un modello gerarchico spazio-temporale per l'analisi di dati relativi al morbillo, tenendo conto sia della componente spaziale che di quella temporale. Inoltre, nel 2020 [8] propongono un modello lineare generalizzato dinamico, come approccio allo studio della prima fase dell'epidemia da Covid-19.

2 Metodi per l'analisi spazio-temporale

Il presente progetto di tesi ha lo scopo di analizzare i dati relativi all'epidemia da SARS-CoV-2 in Italia sia attraverso una prospettiva longitudinale, focalizzandosi sulla dimensione temporale dei dati, sia tramite l'analisi dell'evoluzione dell'epidemia dal punto di vista spaziale.

Si mostrano, dunque, due approcci di analisi esplorativa. Il primo approccio è esclusivamente spaziale, in cui si tiene conto della dimensione temporale attraverso il condizionamento per diversi periodi temporali dell'epidemia.

Il secondo metodo di analisi esplorativa riguarda la caratterizzazione dell'epidemia dal punto di vista temporale, in cui si tiene conto della componente spaziale condizionando le analisi per le regioni secondo le quali sono suddivisi i dati osservati.

Successivamente, viene mostrata l'analisi di modellazione in cui si introducono sia la componente spaziale che la componente temporale tramite diversi approcci che vengono spiegati nel paragrafo 2.3.

2.1 Analisi esplorativa spaziale

Nel seguente paragrafo si mostrano i metodi utilizzati nel progetto di tesi relativi all'analisi esplorativa spaziale dei dati.

Viene fornita una caratterizzazione spaziale dell'epidemia tramite la visualizzazione del conteggio dei nuovi positivi e della prevalenza di ogni regione; inoltre, vengono mostrati degli indici che misurano la correlazione spaziale globale e locale dei dati, due strumenti che possono rivelarsi utili per valutare l'influenza della variazione dei positivi fra le regioni vicine.

2.1.1 Visualizzazione dei dati spaziali

Per la visualizzazione dei dati da un punto di vista spaziale vengono rappresentati i nuovi positivi giornalieri e la prevalenza, utilizzando come base grafica la mappa dell'Italia divisa in regioni tramite i confini regionali delle stesse, come visibile

in Figura 1. Si precisa che Trento e Bolzano sono considerate separatamente in quanto province autonome.

I dati che vengono analizzati nel progetto di tesi sono chiamati anche *dati areali*, cioè, come definito da Bivand et al.[2], sono dati raccolti su aree specifiche con confini ben definiti che vengono aggregati ed assegnati ad un certa entità areale, come le 21 regioni italiane prese in considerazione nel progetto di tesi.



Figura 1: Mappa geografica dell'Italia, divisa in regioni tramite i confini politici

2.1.2 Misure di correlazione spaziale

Per analizzare la correlazione spaziale, è importante capire quali regioni possano essere considerate vicine alle altre e, per formalizzare questo concetto, come descritto da Bivand et al.[2] viene creata una matrice di adiacenza (o anche matrice di prossimità o matrice di vicinanza), che verrà chiamata W , rappresentata da una matrice quadrata di dimensione 21×21 in cui l'elemento $w_{ii} = 0$ e l'elemento $w_{ij} \neq 0$, con $i \neq j$, se la regione j -esima è un vicino della regione i -esima. Ad ogni regione è associato un centroide, definito come punto centrale dell'area regionale e individuato tramite una coppia di coordinate spaziali di latitudine e longitudine.

Partendo dalla matrice di adiacenza viene creato un grafo non orientato con lo scopo di rappresentare graficamente la vicinanza fra le regioni, tramite il collegamento dei rispettivi centroidi. È stato imposto un vincolo che rappresenta la distanza massima perché una regione possa essere considerata vicina: tale vincolo è pari a $\frac{3}{4}$ della distanza massima rilevata fra due centroidi. In questo modo si tiene conto della morfologia delle regioni italiane, in cui le regioni più estreme (Sardegna, Sicilia, Valle d'Aosta) risulteranno meno connesse nel grafo creato.

Si stima l'indice di Moran globale per analizzare se la distribuzione dei nuovi

positivi mostra una correlazione spaziale a livello globale.

L'indice I di Moran è calcolato come il rapporto fra due quantità: il prodotto (pesato) fra la variabile di interesse (centrata attorno alla media) nelle regioni i e j , e la devianza della variabile di interesse nella i -esima regione. Tale rapporto viene moltiplicato per il rapporto fra il totale delle regioni n e la somma dei pesi w_{ij} [2].

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (y_i - \bar{y})^2},$$

dove, nell'analisi effettuata, $y_i(y_j)$ è il numero di nuovi casi di positività cumulati all'interno di un'ondata nella regione i -esima (j -esima), \bar{y} è la media dei nuovi casi cumulati fra tutte le regioni, w_{ij} , con $i \neq j = 1, \dots, n$, è l'elemento ij della matrice di adiacenza W , indica il peso spaziale del collegamento tra le regioni i e j e può essere calcolato in diversi modi: si può assegnare un valore pari a 1 se due regioni sono considerate vicine. Oppure $w_{ij} = \frac{1}{K}$, dove K è il numero totale di vicini dell' i -esima regione. La differenza sostanziale fra i due metodi è che utilizzando il secondo metodo la somma dei pesi per ogni regione sarà sempre pari a 1, mentre con il primo metodo tale somma sarà pari al numero di vicini. Nel presente progetto di tesi verrà utilizzato il primo metodo di assegnazione dei pesi per seguire la morfologia dell'Italia che presenta una disposizione asimmetrica delle regioni.

Poiché il modello è centrato sulla media, l'indice di Moran individua la presenza di informazione residua dopo il centramento sulla media: tale informazione non spiegata è causata dalle relazioni spaziali rappresentate dalla matrice di adiacenza W . La statistica I viene testata, verificando la sua divergenza dal valore atteso $E(I) = \frac{1}{(n-1)}$, con n pari al numero di regioni con almeno un vicino [2].

Viene, inoltre, calcolato l'indice di Moran locale, allo scopo di indagare più nello specifico sulla presenza di effetti spaziali a livello regionale, che possono impattare maggiormente in particolari aree dell'Italia. L'indice di Moran locale possiede una logica simile all'indice di Moran globale: è da intendere come gli scostamenti della regione i -esima dalle aree circostanti, secondo i pesi imposti dalla matrice di adiacenza. Inoltre, i valori sono costruiti per sommare al valore dell'indice di Moran globale.

La statistica ha la seguente formulazione, come presente in Bivand et al.(2013)[2]

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}.$$

Come l'indice di Moran globale, l'ipotesi alternativa H_1 è che il valore osservato del test sia maggiore del suo valore atteso $E(I_i)$, calcolato nel seguente modo da Sokal et al.(1998) [23]: $E(I_i) = -\frac{\sum_j w_{ij}}{(n-1)}$

2.2 Analisi esplorativa temporale

Nel seguente paragrafo si mostrano i metodi utilizzati per esplorare la dimensione temporale dei dati tramite grafici per visualizzare l'andamento nel tempo del conteggio dei nuovi positivi e della prevalenza, sia a livello nazionale che a livello regionale.

Per la visualizzazione dei dati a livello temporale, si rappresentano i valori puntuali giornalieri dei nuovi positivi e della prevalenza formando le cosiddette curve epidemiologiche. I nuovi positivi rappresentano il conteggio giornaliero dei nuovi soggetti identificati positivi al virus SARS-CoV-2 in senso assoluto, mentre la prevalenza indica la percentuale di positivi giornaliera sul totale della popolazione rilevata a inizio mese (assunta costante per tutti i giorni del mese considerato). Entrambe le quantità sono rappresentate a livello nazionale e regionale.

Nei seguenti paragrafi si specifica un modello lineare generalizzato dinamico (DGLM), con lo scopo di trovare un migliore adattamento per i dati di conteggio e si illustra il metodo di stima del numero di riproduzione R_t per vedere l'andamento dell'evoluzione dei contagi.

2.2.1 Il modello lineare generalizzato dinamico

Per rappresentare una curva stimata che si adatta meglio alla natura di conteggio dei dati e aggiungere la componente di autocorrelazione, è stato considerato un modello lineare generalizzato dinamico (DGLM) [11, 17] basato sull'assunzione che i dati si distribuiscono secondo una distribuzione di Poisson.

Il modello DGLM di Chiogna e Gaetan (2020) [8], adattato allo studio dell'andamento dei casi di positività da Covid-19, è utilizzato a scopo esclusivamente esplorativo in quanto nel loro lavoro non viene inserita alcuna covariata, ma l'obiettivo è quello di individuare sostanziali variazioni nel processo casuale sottostante al fenomeno dell'aumento dei casi positivi. Il modello è così definito: Y_t è il numero di nuovi casi giornalieri al tempo t e segue una distribuzione di Poisson con numero atteso di conteggi al tempo t pari $\mu_t > 0$:

$$Y_t | \mu_t \sim \text{Poisson}(\mu_t), \quad t = 1, 2, \dots$$

con funzione legame log-lineare per i conteggi attesi μ_t

$$\begin{aligned} \log(\mu_t) &= \alpha_t \\ \alpha_t &= \alpha_{t-1} + \beta_{t-1} + e_{1,t} \\ \beta_t &= \beta_{t-1} + e_{2,t} \end{aligned}$$

dove $e_t = (e_{1,t}, e_{2,t})^\top \sim N_2(0, Q)$, $\theta_0 = (\alpha_0, \beta_0)^\top \sim N_2(m_0, P_0)$ sono quantità casuali indipendenti fra loro, m_0 è il vettore delle medie di α_0 e β_0 , e P_0 è la

matrice di covarianza di α_0 e β_0 . Sia m_0 che P_0 sono fissati a priori. La matrice di covarianza

$$Q = \begin{pmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{pmatrix}$$

contiene i parametri ignoti da stimare $\psi_i > 0$, $i = 1, 2$.

Il modello specifica un trend lineare locale $\alpha_t = \log(\mu_t)$ e β_t è la pendenza di α_t che varia nel tempo.

Viene utilizzato l'algoritmo di ottimizzazione "Broyden-Fletcher-Goldfarb-Shanno" (BFGS) [25], un algoritmo di ottimizzazione non lineare ampiamente utilizzato per risolvere problemi di ottimizzazione non vincolata. In questo caso, viene utilizzato per massimizzare la funzione di verosimiglianza nella stima di un DGLM con una distribuzione di Poisson. La funzione di verosimiglianza per il modello di Poisson, dati i parametri α_t e β_t , è data da:

$$L(y | \theta) = \prod_{t=1}^T P(Y_t = y_t | \alpha_t, \beta_t),$$

dove θ è il vettore dei parametri del modello. La probabilità $P(Y_t = y_t | \alpha_t)$ per la distribuzione di Poisson è:

$$P(Y_t = y_t | \alpha_t) = \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!}.$$

Inoltre, si utilizza l'algoritmo *importance sampling* con lo scopo di calcolare la verosimiglianza e le stime lisciate del trend lineare α_t [15]. Come mostrato da Yuan e Druzdzel (2006) [26], il problema si basa sulla stima del seguente integrale

$$V = \int_{\Omega} \frac{f(X)}{D(X)} D(X) dX,$$

dove $f(x)$ è una funzione di X nel dominio $\Omega \subset \mathbb{R}^n$ che rappresenta la densità di probabilità della distribuzione dei dati osservati e $D(X)$, chiamata anche funzione di importanza, è una distribuzione di probabilità tale per cui $D(X) > 0$ per ogni $X \in \Omega$, da cui è più facile campionare (in questo caso una Poisson) e che dovrebbe essere sufficientemente simile alla distribuzione di probabilità dei dati osservati.

Dunque, si generano i campioni X_1, X_2, \dots, X_N dalla distribuzione di probabilità di campionamento $D(X)$ e si calcolano i pesi che rappresentano l'importanza di ogni campione: si calcola il rapporto tra la densità di probabilità della distribuzione dei dati osservati e la densità di probabilità della distribuzione di campionamento nel punto in cui è stato campionato il valore. In formule:

$$\hat{V} = \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{D(X_i)}.$$

L'algoritmo assegna maggiore importanza ai campioni in cui $f(x) > D(X)$ e minor importanza ai campioni in cui $f(x) < D(X)$.

I campioni pesati secondo questo metodo vengono ordinati secondo il peso assegnato e viene stimato il numero di nuovi positivi giornalieri tramite la mediana dei valori campionati.

Come riportato da Chiogna e Gaetan (2020) [8] il vantaggio di questo modello è la capacità di adattarsi ai cambiamenti nel processo casuale sottostante, ma uno svantaggio è quello di generare intervalli di confidenza molto ampi, soprattutto nel caso delle previsioni di lunghi periodi temporali.

2.2.2 Il numero di riproduzione R_t

Il numero di riproduzione R_t è un indicatore dell'andamento dei contagi all'interno di un'epidemia, infatti, con R_t si intende il numero medio di individui contagiato da un unico individuo positivo al tempo t , assumendo che tutti gli individui siano suscettibili. È possibile interpretare il numero di riproduzione verificando la sua distribuzione osservata: quando $R_t > 1$ la curva dei contagi epidemica è in crescita e quando $R_t < 1$ la curva è in decrescita. L'indice R_t può essere stimato in diversi modi, per esempio tramite modelli SIR e loro estensioni [14]

Nel presente progetto di tesi, R_t viene stimato utilizzando una tecnica di stima Bayesiana che viene suggerita da Cori et al.(2013) [10], fornendo la distribuzione dei nuovi positivi giornalieri osservati e specificando a priori la media e la deviazione standard dell'intervallo seriale, seguendo dunque una stima bayesiana parametrica di R_t . L'intervallo seriale di una malattia è interpretato come il tempo (in questo caso i giorni) che intercorre fra l'insorgenza dei sintomi di due soggetti consecutivi in un'ipotetica catena di contagi [4, 24]. L'assunzione alla base di questo criterio di stima è che gli individui infetti abbiano un "profilo infettivo" dato da una distribuzione di probabilità g_s di infezione al tempo s ma indipendente dal giorno $t = 1, 2, 3, \dots$ considerato, cioè un soggetto positivo avrà la stessa distribuzione di infettività indipendentemente da quando risulta positivo. La distribuzione assunta è una Poisson, dove il tasso al quale una persona, infetta al tempo s , genera nuove infezioni al tempo t è dato da $R_s g_{t-s}$, dove R_s è il numero di riproduzione istantaneo al tempo s e g_{t-s} è la distribuzione di probabilità che descrive il profilo medio di infettività dopo l'infezione. Pertanto, i nuovi positivi al tempo t sono distribuiti secondo una Poisson con media:

$$Y_t = R_t \sum_{s=1}^{t-1} Y_s g_{t-s}.$$

La verosimiglianza dei nuovi positivi Y_t dato il numero di riproduzione R_t condizionato al numero di nuovi positivi precedenti Y_1, \dots, Y_{t-1} è:

$$L(Y_t | R_t, Y_1, \dots, Y_{t-1}) = \frac{(R_t \sum_{s=1}^{t-1} Y_s g_{t-s})^{Y_t} e^{-(R_t \sum_{s=1}^{t-1} Y_s g_{t-s})}}{Y_t!}$$

Se si assume che l'infettività sia costante in un periodo di tempo τ , la verosimiglianza dei nuovi positivi durante questo periodo dato il numero di riproduzione R_t condizionato ai nuovi positivi precedenti è

$$L(Y_{[t,t+\tau]}|R_t, Y_1, \dots, Y_{t-1}) = \prod_{u=t}^{t+\tau-1} \frac{(R_t \sum_{s=1}^{u-1} Y_s g_{u-s})^{Y_u} e^{-(R_t \sum_{s=1}^{u-1} Y_s g_{u-s})}}{Y_u!}.$$

Utilizzando un approccio Bayesiano con una distribuzione a priori Gamma con parametri α e β per R_t , la distribuzione a posteriori congiunta di R_t è

$$\pi(R : t|Y_{[t,t+\tau]}, Y_1, \dots, Y_{t-1}) \propto R_t^{\alpha-1} e^{-R_t/\beta} \cdot \prod_{u=t}^{t+\tau-1} \frac{(R_t \sum_{s=1}^{u-1} Y_s g_{u-s})^{Y_u} e^{-(R_t \sum_{s=1}^{u-1} Y_s g_{u-s})}}{Y_u!}.$$

Quindi, la distribuzione a posteriori di R_t è una distribuzione Gamma con parametri:

$$\alpha = \alpha_0 + \sum_{u=t}^{t+\tau-1} Y_u$$

$$\beta = \left(\frac{1}{\beta_0} + \sum_{u=t}^{t+\tau-1} \sum_{s=1}^{u-1} Y_s g_{u-s} \right)^{-1}$$

Pertanto, la media a posteriori di R_t è:

$$\mathbb{E}[R_t|Y_{[t,t+\tau]}, Y_1, \dots, Y_{t-1}] = \frac{\alpha}{\beta},$$

con $Y_{[t,t+\tau]}$ che rappresentano i valori futuri dell'intervallo temporale di lunghezza τ . Infine, dalla distribuzione a posteriori di R_t è possibile ricavare le sue caratteristiche di interesse, come mediana, varianza o intervallo di credibilità al 95%. I valori di α_0 e β_0 specificati a priori sono stati ricavati dallo studio di metanalisi di Challen et al.(2020) [4] che raggruppa i parametri degli intervalli seriali per diverse epidemie e, in questo caso, sono i valori presenti nello studio di Cereda et al.(2020) [3] relativo alle prime fasi dell'epidemia da SARS-CoV-2 in Lombardia.

2.3 Modellazione spazio-temporale

2.3.1 Il modello CAR

Si introduce, per prima cosa, il modello Autoregressivo Condizionato (CAR), la cui specificazione viene seguita dalla distribuzione condizionata del termine

di errore spaziale. Un processo Autoregressivo Condizionato (CAR) definisce una struttura di dipendenza spaziale tra le osservazioni attraverso distribuzioni condizionate autoregressive. In altre parole, la distribuzione di una variabile in un'area dipende dai valori delle variabili nelle aree vicine e questo è dato dalla proprietà di condizionalità, definita come segue da Besag (1974)[1]: Ogni valore Y_i per l'area i è condizionato sui valori Y_j delle aree vicine j :

$$Y_i | \{Y_j, j \neq i\} \sim N \left(\sum_j \frac{w_{ij} Y_j}{\sum_j w_{ij}}, \frac{\tau_i^2}{\sum_j w_{ij}} \right),$$

dove w_{ij} è il generico elemento della matrice di adiacenza W introdotta nel paragrafo 2.1.2 e τ_i^2 è la varianza condizionata.

Inoltre, le distribuzioni condizionate devono essere compatibili, permettendo di costruire una distribuzione congiunta per tutte le aree. Attraverso il lemma di Brook, la distribuzione congiunta può essere espressa come:

$$p(Y) \propto \exp \left[-\frac{1}{2} Y' \{D^{-1}(I - W)\} Y \right],$$

dove D è una matrice diagonale delle varianze condizionate tale che $D_{ii} = \sum_j w_{ij}$, e I è la matrice identità.

La matrice di varianza-covarianza del modello è data da:

$$\Sigma_Y = (I - \rho W)^{-1} D,$$

dove $\rho \in (\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}})$, con $\lambda_{(1)} < \dots < \lambda_{(n)}$ che sono gli autovalori ordinati di $D^{-1/2} W D^{-1/2}$, e scelto in modo tale da rendere Σ_Y^{-1} non singolare.

Σ_Y è strettamente collegata al concetto di indipendenza condizionata di Y_i e Y_j : infatti, se $(\Sigma_Y^{-1})_{ij} = 0$, allora Y_i e Y_j sono condizionatamente indipendenti dati gli altri valori. Questo implica che la scelta della struttura della matrice di adiacenza influisce direttamente sulla specificazione dell'indipendenza condizionata.

2.3.2 Separabilità delle componenti spazio-temporali

Come proposto sia da Gneiting (2002) che da Bivand et al. (2013) [2], data la complessità di modellazione dei tassi di infettività nello spazio e nel tempo, si può assumere la separabilità delle componenti spaziali e temporali. In altre parole si assume che la struttura di correlazione spazio-temporale sia il prodotto fra la correlazione spaziale e la correlazione temporale.

Siano

- x_i e x_j le coordinate spaziali dell' i -esima e j -esima regione, con $i \neq j$ e $i, j = 1, \dots, 21$,
- $u = \|x_i - x_j\|$ la distanza euclidea fra i centroidi di qualsiasi due delle regioni considerate,

- $v = |t - (t - k)|$, con $t = 1, \dots, T$ e $k = 1, 2, 3, \dots, t - 1$, la distanza temporale considerata,

la struttura di correlazione spazio-temporale $\rho(u, v; \tau)$ è definita come il prodotto della funzione di correlazione spaziale $\rho_1(u; \tau_1)$ e della funzione di correlazione temporale $\rho_2(v; \tau_2)$, dove τ_1 e τ_2 parametrizzano rispettivamente ρ_1 e ρ_2 .

Cioè:

$$\rho(u, v; \tau) = \rho_1(u; \tau_1)\rho_2(v; \tau_2) = \frac{1}{(1 + v/\psi)^{\eta+1}} \exp\left\{-\frac{u}{\phi}\right\},$$

dove ϕ e (η, ψ) sono dei parametri positivi che determinano la velocità di tendenza verso lo 0 rispettivamente della correlazione spaziale e temporale.

Seguendo il modello proposto da Bivand et al.[2], si sceglie di modellare la dipendenza spaziale tramite un processo CAR intrinseco e la dipendenza temporale tramite un processo *Random Walk* di ordine 1.

2.3.3 Test per l'omogeneità dei rischi relativi

Prima di formalizzare il modello e di assumere una distribuzione per il rischio relativo θ si effettua un test sull'omogeneità dei rischi relativi tramite un test proposto da Potthoff e Whittinghill(1966) [21]. Viene effettuato il test per studiare la presenza di eterogeneità dei rischi relativi, che può essere dovuta a diversi fattori, come ad esempio una diversa esposizione ai fattori di rischio, una diversa conformazione territoriale delle zone considerate o anche una diversa distribuzione fra le aree geografiche delle manovre di contenimento dell'epidemia. L'ipotesi alternativa è che i rischi relativi θ_i seguano una distribuzione Gamma con media μ_θ e varianza σ_θ^2 :

$$H_0 : \theta_1 = \dots = \theta_n = \mu_\theta$$

$$H_1 : \theta_i \sim \text{Gamma}\left(\frac{\mu_\theta^2}{\sigma_\theta^2}, \frac{\mu_\theta}{\sigma_\theta^2}\right).$$

La statistica test è definita come segue:

$$PW = \sum_i E_i \sum_i \frac{Y_i(Y_i - 1)}{E_i}$$

2.3.4 Formalizzazione del modello

In questo paragrafo viene spiegato il processo di modellazione che tiene conto sia della componente latente spaziale che di quella temporale.

L'obiettivo della modellazione è la valutazione degli eventuali effetti della correlazione spaziale, dell'autocorrelazione seriale e dell'effetto che alcune variabili esplicative di interesse possono avere sulla variabile risposta, che in questo caso è la serie giornaliera dei nuovi positivi per ciascuna regione. La scelta del modello di regressione ricade su un modello Poisson gerarchico con criterio di

stima Bayesiana dei parametri di interesse. Nel contesto del *Disease Mapping* spazio-temporale, che consiste in una mappatura geografica della malattia, si può utilizzare questo approccio di modellazione in cui, come presente in Bivand et al. (2013) [2], il numero di nuovi casi, nella regione i -esima al tempo t , viene indicato come $Y_{i,t}$. La media $\mu_{i,t}$ nella regione i e al tempo t , è definita come $\mu_{i,t} = E_{i,t}\theta_{i,t}$, con $E_{i,t}$ che indica il numero atteso di nuovi casi della regione i -esima al tempo t e $\theta_{i,t}$ come il rischio relativo. $E_{i,t}$ è calcolato come segue:

$$E_{i,t} = \frac{\sum_{t=1}^T Y_{i,t}}{\sum_{t=1}^T N_{i,t}},$$

dove T è il numero totale dei giorni di ciascun mese e $N_{i,t}$ è la popolazione al tempo t nella i -esima regione. Poiché, come discusso all'inizio del paragrafo 2.2, la popolazione è osservata mensilmente, e dunque costante fra i giorni di ciascun mese, il numero atteso di nuovi casi sarà costante fra i giorni di ciascun mese considerato. Il modello intende stimare l'effetto dell'isolamento domiciliare della campagna di vaccinazione durante la quarta ondata, e stima l'effetto latente della correlazione spaziale e dell'autocorrelazione temporale fra le osservazioni. Il modello si può formalizzare come segue:

$$\begin{aligned} Y_{i,t} &\sim \text{Poisson}(\mu_{i,t}) \\ \log(\theta_{i,t}) &= \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i + v_t \\ u_i &\sim \text{CAR}(\sigma_u^2) \\ v_t &\sim \text{RW}(1, \sigma_v^2) \end{aligned}$$

Dove:

- $Y_{i,t}$ rappresenta il numero di nuovi casi osservati nella regione i al tempo t .
- $\mu_{i,t} = E_{i,t}\theta_{i,t}$, con $E_{i,t}$ che rappresenta il numero atteso dei nuovi casi nella regione i al tempo t e con $\theta_{i,t}$ che rappresenta il rischio relativo nella regione i al tempo t .
- α è l'intercetta.
- β_1 è il coefficiente di regressione della covariata x_{i1} , che rappresenta il numero giornaliero di soggetti in quarantena e segue una distribuzione a priori $N(\mu_{\beta_1}, \sigma_{\beta_1}^2)$.
- β_2 è il coefficiente di regressione della covariata x_{i2} , che rappresenta il numero giornaliero di vaccini somministrati e segue una distribuzione a priori $N(\mu_{\beta_2}, \sigma_{\beta_2}^2)$.
- u_i rappresenta gli effetti casuali (o latenti) spaziali e segue un processo CAR, definito nel paragrafo 2.3.1, e σ_u^2 è la sua varianza, mentre v_t rappresenta gli effetti temporali e segue un processo *Random Walk* di ordine 1 per fornire una stima flessibile del trend temporale, e σ_v^2 è la sua varianza.

- $\sigma_u^2 \sim \text{InvGamma}(\gamma_u, \delta_u)$, $\sigma_v^2 \sim \text{InvGamma}(\gamma_v, \delta_v)$, dove $\gamma_u = \delta_u = \gamma_v = \delta_v = 0.001$ per avere una distribuzione a priori meno informativa possibile, come mostrato da Bivand (2013) [2].

La funzione di verosimiglianza del modello è data da:

$$L(Y | \theta, E) = \prod_{i,t} \frac{(\mu_{i,t})^{y_{i,t}} e^{-\mu_{i,t}}}{y_{i,t}!}.$$

Il rischio relativo $\theta_{i,t}$ ha una distribuzione a priori:

$$p(\theta) \propto p(\beta_1)p(\beta_2)p(u)p(v)p(\rho),$$

dove $\rho(u, v; \tau)$ è la correlazione spazio-temporale introdotta del paragrafo 2.3.2. Dunque, la distribuzione a posteriori è:

$$p(\theta, \Psi | Y, E) \propto p(Y | \theta, E) \cdot p(\theta | \sigma_u^2, \sigma_v^2) \cdot p(\sigma_u^2) \cdot p(\sigma_v^2),$$

dove $\Psi = (\mu_{\beta_1}, \mu_{\beta_2}, \sigma_{\beta_1}^2, \sigma_{\beta_2}^2, \sigma_u^2, \sigma_v^2)$ è il vettore dei parametri del modello. Esplicitamente, può essere scritta come:

$$p(\theta, \Psi | Y, E) \propto \left[\prod_{i,t} \frac{(\mu_{i,t})^{y_{i,t}} \exp\{-\mu_{i,t}\}}{y_{i,t}!} \right] \cdot \frac{\exp\left\{-\frac{(\beta_1 - \mu_{\beta_1})^2}{2\sigma_{\beta_1}^2}\right\}}{\sqrt{2\pi\sigma_{\beta_1}^2}} \cdot \frac{\exp\left\{-\frac{(\beta_2 - \mu_{\beta_2})^2}{2\sigma_{\beta_2}^2}\right\}}{\sqrt{2\pi\sigma_{\beta_2}^2}} \cdot \exp\left\{-\frac{1}{2\sigma_u^2} \mathbf{u}^\top W \mathbf{u}\right\} \cdot \exp\left\{-\frac{1}{2\sigma_v^2} \sum_{t=2}^T (v_t - v_{t-1})^2\right\} \cdot \frac{\delta_u^{\gamma_u}}{\Gamma(\gamma_u)} \frac{\exp\left\{-\frac{\delta_u}{\sigma_u^2}\right\}}{\sigma_u^{2\gamma_u+1}} \cdot \frac{\delta_v^{\gamma_v}}{\Gamma(\gamma_v)} \frac{\exp\left\{-\frac{\delta_v}{\sigma_v^2}\right\}}{\sigma_v^{2\gamma_v+1}},$$

dove \mathbf{u} è il vettore degli effetti spaziali regionali u_i , $i = 1, \dots, 21$ e W è la matrice di adiacenza.

2.3.5 Stima dei parametri

La stima dei parametri del modello appena mostrato viene eseguita utilizzando un metodo per l'inferenza Bayesiana basato sui metodi di stima presenti in Rue et al.(2017)[22], chiamato *Approssimazione di Laplace integrata annidata* (INLA). Questo metodo è un'alternativa ai metodi di stima che si basano su catene Markoviane, in particolare il campionamento di Gibbs, offrendo maggiore velocità computazionale, soprattutto in modelli con numerosi parametri da stimare, e spesso maggiore accuratezza sulle approssimazioni della distribuzione a posteriori.

L'idea principale del metodo INLA è approssimare l'integrale della distribuzione a posteriori utilizzando una serie di approssimazioni successive.

Si inizia approssimando la distribuzione marginale degli iperparametri σ_u^2 e σ_v^2 utilizzando l'approssimazione di Laplace:

$$p(\sigma_u^2, \sigma_v^2 | Y) \approx \frac{p(Y, \theta^*, \sigma_u^2, \sigma_v^2)}{p(\theta^* | \sigma_u^2, \sigma_v^2) |\mathbf{H}(\sigma_u^2, \sigma_v^2)|^{-1/2}}$$

dove θ^* è il massimo a posteriori di θ dato σ_u^2 e σ_v^2 , e $\mathbf{H}(\sigma_u^2, \sigma_v^2)$ è la matrice hessiana della log-verosimiglianza.

Le distribuzioni condizionate dei parametri latenti θ sono approssimate usando:

$$p(\theta | Y, \sigma_u^2, \sigma_v^2) \approx N(\mu_i, \sigma_i^2)$$

dove i parametri μ_i e σ_i^2 sono ottenuti tramite approssimazioni locali di Laplace.

Un limite del metodo INLA è che esso non fornisce la distribuzione a posteriori, ma questa non è necessaria nel caso della stima del rischio relativo nel contesto della mappatura geografica di una malattia, come affermato da Bivand(2013).

3 Caso studio

3.1 Dataset

Gli insiemi di dati utilizzati per l'analisi contengono le informazioni sul conteggio dei positivi e provengono dal pacchetto R *covid19ita*, consultabile dal sito: <https://github.com/UBESP-DCTV/covid19ita>. Il pacchetto R contiene i dati italiani per il monitoraggio dell'epidemia da Covid-19.

I dati relativi al conteggio della somministrazione dei vaccini provengono dal sito ufficiale github del Ministero della Salute, all'indirizzo <https://github.com/italia/covid19-opendata-vaccini/tree/master/dati>.

I dati della popolazione mensile regionale provengono dalla piattaforma ufficiale per la diffusione dei dati aggregati dell'Istat, all'indirizzo <https://esploradati.istat.it/databrowser>.

3.1.1 Dati a livello nazionale

I dati a livello nazionale presentano 1040 osservazioni, ognuna relativa a un giorno, a partire dal 24/02/2020 fino al 29/12/2022.

Le variabili presenti sono 23: data, numero di ricoverati con sintomi, numero di pazienti in terapia intensiva, totale dei pazienti ospedalizzati, soggetti in isolamento domiciliare, totale dei positivi correnti (pazienti ospedalizzati + soggetti in isolamento domiciliare), variazione totale positivi (nuovo ammontare dei casi positivi correnti), nuovi positivi (casi attivi oggi - casi attivi ieri), dimessi guariti, morti, casi da "sospetto diagnostico" (casi positivi emersi dall'attività clinica), casi da screening (casi positivi emersi da indagini e test, pianificati a livello regionale o nazionale), ammontare totale dei casi positivi, tamponi effettuati, note.

Inoltre, a partire dal 09/04/2020 è registrato il numero di casi testati (nessuna differenziazione del test utilizzato) e a partire dal 15/01/2021 i positivi sono anche differenziati in base alla positività al test molecolare o al test rapido, oltre ai tamponi molecolari e rapidi effettuati.

3.1.2 Dati a livello regionale

I dati raccolti a livello regionale presentano 21.840 osservazioni, suddivise in 21 osservazioni per giorno, corrispondenti alle 21 diverse denominazioni delle regioni (Trento e Bolzano considerate come due regioni diverse), per 1.040 giorni, dunque la finestra temporale è la stessa dei dati nazionali.

Le variabili presenti sono 28: le stesse presenti a livello nazionale con l'aggiunta del codice e denominazione della regione e i valori di latitudine e longitudine del centroide associato alla regione.

3.1.3 Dati sui vaccini

I dati relativi alla somministrazione dei vaccini presentano 14874 osservazioni e 15 variabili, raccolti in una finestra temporale che va dal 27/12/2020 al 29/12/2022. Le variabili presenti sono: il codice identificativo del soggetto vaccinato, la sigla della regione nella quale è avvenuta la somministrazione, il codice ISTAT di tale regione, la sigla di classificazione europea della regione e la denominazione della regione in cui è avvenuta la somministrazione, la data di somministrazione, il numero totale di dosi di vaccino somministrate per giorno e per regione, il totale di soggetti di sesso maschile e femminile a cui è stato somministrato il vaccino e infine alcune informazioni riguardanti il numero di prime, seconde dosi e di dosi "booster". Per l'analisi verranno considerate esclusivamente le variabili relative a data, nome regione e totale dei vaccini giornalieri. Si rappresenta un cronoprogramma per illustrare il grado di sovrapposizione temporale delle varie fonti di dati.

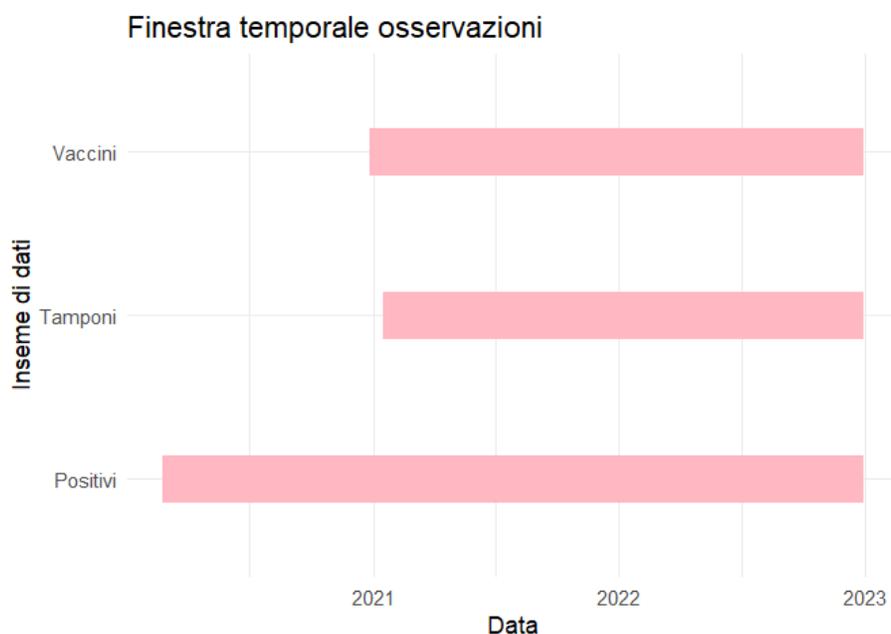


Figura 2: Cronoprogramma che illustra la finestra temporale di collezione dei dati

In Tabella 1 si mostrano le percentuali di dati mancanti nel dataset regionale.

Variabile	Percentuale di dati mancanti
Data	0
Denominazione regione	0
Totale positivi	0
Variazione totale positivi	0
Nuovi positivi	0
Dimessi guariti	0
Deceduti	0
Tamponi	0
Popolazione mensile	0
Casi testati	0.0529
Ingressi in terapia intensiva	0.2721
Totale positivi al test molecolare	0.3135
Totale positivi al test antigenico rapido	0.3135
Tamponi molecolari effettuati	0.3135
Tamponi rapidi effettuati	0.3135
Totale vaccini somministrati	0.3190

Tabella 1: Percentuale dei dati mancanti per ogni variabile

Si nota un ammontare di dati mancanti pari al 31.35% del totale delle osservazioni relative ai tamponi effettuati (antigenico e molecolare) e ai tamponi positivi (antigenico e molecolare), in quanto la raccolta di tali dati è iniziata a partire dal 15/01/2021. Inoltre, l'ammontare di dati mancanti per il conteggio dei vaccini somministrati è pari al 31.9%, poiché la finestra temporale di raccolta di tali dati inizia il 27/12/2020.

Non si rilevano ulteriori problemi, in quanto le variabili con una percentuale eccessiva di dati mancanti non sono di interesse per il progetto di tesi.

3.2 Risultati dell'analisi esplorativa

3.2.1 Analisi esplorativa spaziale

Seguono i grafici per la visualizzazione spaziale dei nuovi positivi cumulati e dell'incidenza, condizionati alle diverse finestre temporali corrispondenti alle cinque ondate (Figure 3-7).

Nuovi positivi e incidenza prima ondata
24/02/2020-24/04/2020

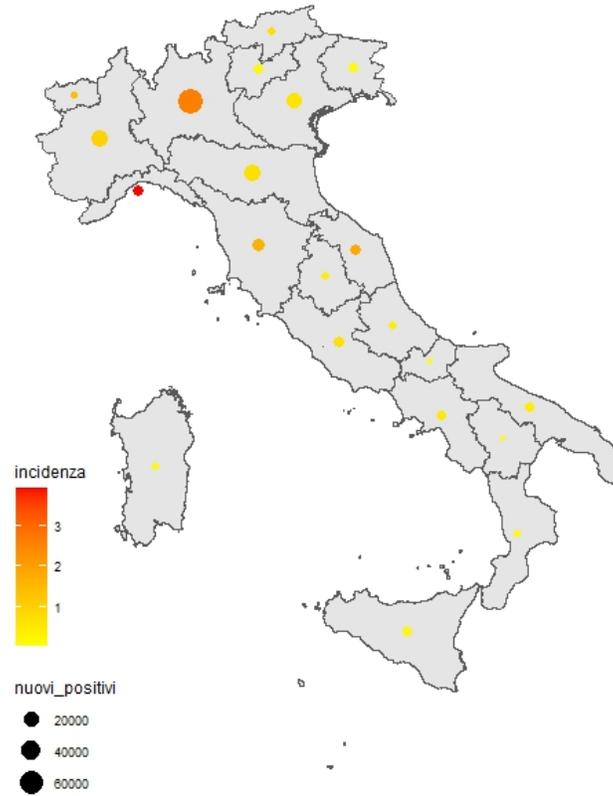


Figura 3: Nuovi positivi cumulati e incidenza durante la prima ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda

Nuovi positivi e incidenza seconda ondata
24/09/2020-23/12/2020

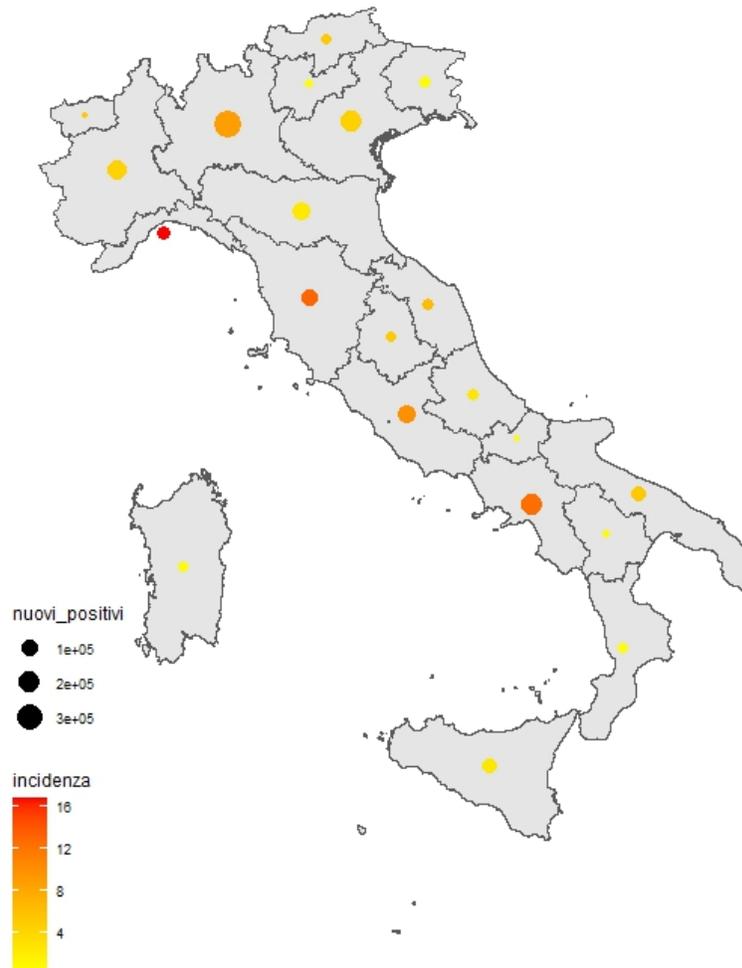


Figura 4: Nuovi positivi cumulati e incidenza durante la seconda ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda

Nuovi positivi e incidenza terza ondata
24/12/2020-24/05/2021

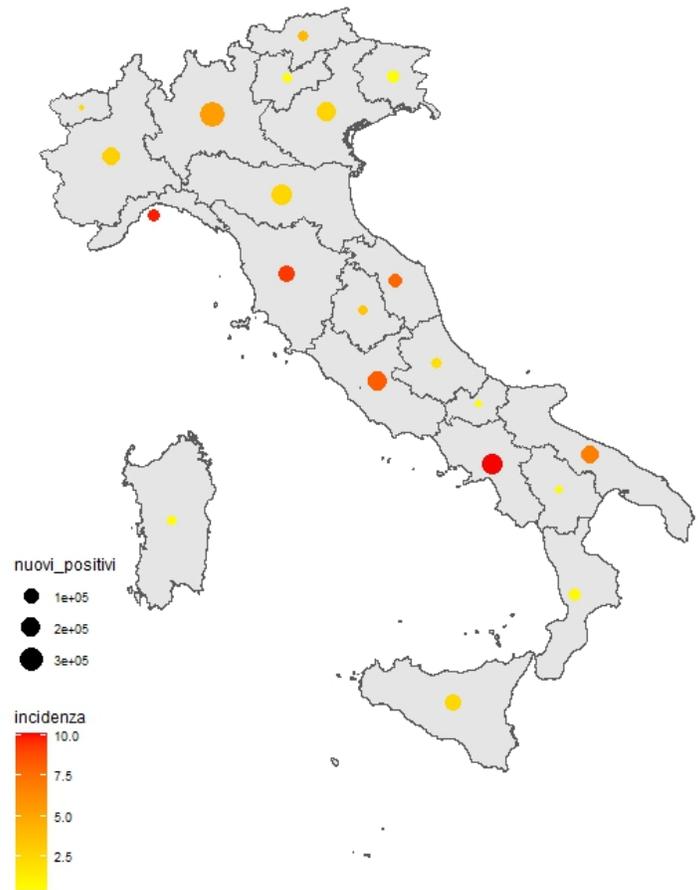


Figura 5: Nuovi positivi cumulati e incidenza durante la terza ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-ondata

Nuovi positivi e incidenza quarta ondata
24/10/2021-23/05/2022

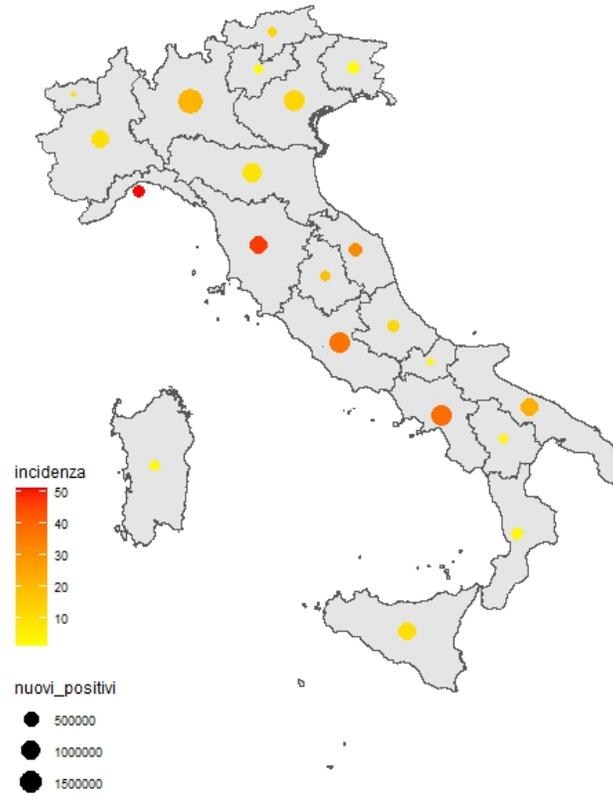


Figura 6: Nuovi positivi cumulati e incidenza durante la quarta ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-onda

Nuovi positivi e incidenza quinta ondata
24/05/2022-29/12/2022

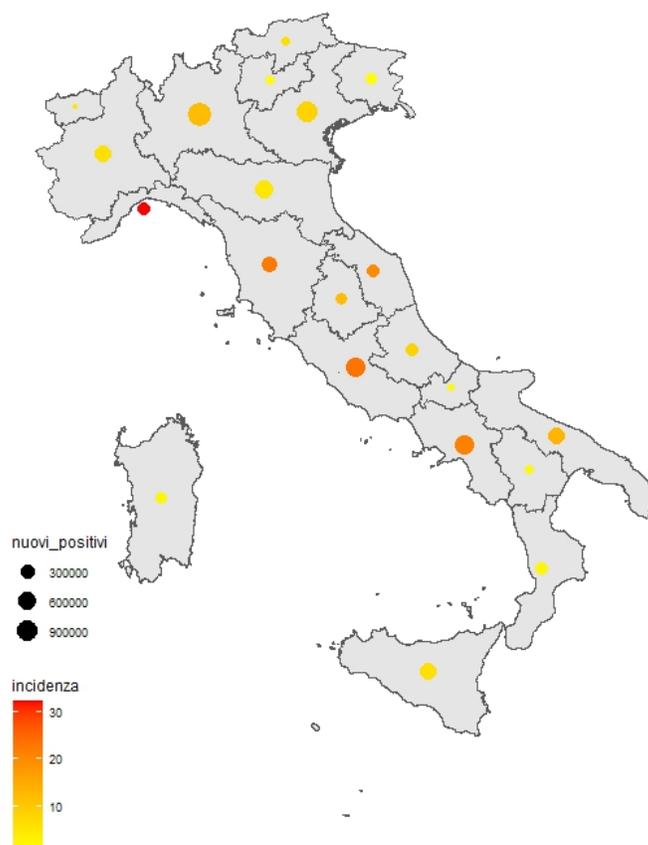


Figura 7: Nuovi positivi cumulati e incidenza durante la quinta ondata; la grandezza del centroide indica il totale di nuovi positivi durante la prima ondata e il colore indica l'incidenza per 10.000 persone-ondata

Le mappe mostrano i nuovi positivi cumulati (indicata dalla grandezza del cerchio) e l'incidenza (indicata dall'intensità di colore) durante i periodi delle ondate di Covid-19 che sono stati definiti precedentemente. L'incidenza è stata calcolata tramite il rapporto fra i nuovi casi cumulati e la massa tempo-persona a rischio durante l'ondata, ed è stata moltiplicata per 10.000 per facilitarne l'interpretazione. Si nota che le regioni maggiormente colpite, in scala assoluta, in tutte le ondate, sono la Lombardia, il Veneto, la Campania, l'Emilia-Romagna e il Lazio. Inoltre, si nota che la Liguria è una delle regioni con l'incidenza più alta, in tutte le ondate dell'epidemia.

Focalizzandosi sulla quarta ondata (Figura 7), nel periodo compreso fra il

24/10/2021 e il 23/05/2022, si nota un'incidenza di circa 50 nuovi casi ogni 10.000 abitanti in Liguria, con un numero simile in Toscana, Lazio e Campania. Tale informazione conferma quanto visto nei grafici sull'andamento temporale dei positivi, poiché, in tutte le regioni, la quarta ondata racchiude il punto di massimo assoluto raggiunto dalle curve dei positivi.

Anche la Lombardia e la Puglia mostrano dei valori di incidenza elevata (circa 30 nuovi casi ogni 10.000 abitanti), suggerendo una possibile forma di correlazione spaziale fra le regioni appena citate, soprattutto nella zona del Nord Italia e fra le regioni del Centro-Sud Italia.

3.2.2 Risultati dell'indice I di Moran

Viene svolta un'analisi esplorativa sulla correlazione spaziale del numero dei nuovi positivi cumulati nelle cinque diverse ondate, con lo scopo di indagare una possibile correlazione spaziale globale o locale fra sottogruppi di regioni. In altre parole, ci si chiede se un aumento o una diminuzione dei nuovi casi di positività in una certa regione possa essere correlato con l'aumento o la diminuzione dei nuovi casi nelle regioni considerate come vicine.

Viene creato un grafo non orientato (Figura 8) da cui si ricava la matrice di adiacenza, come discusso nel paragrafo 2.1.2, con lo scopo di creare una matrice che indica quali regioni possono essere considerate vicine, e dunque connesse, nel grafo. Viene impostata una distanza massima fra i centroidi delle regioni pari a $\frac{3}{4}$ della distanza massima rilevata, per tenere conto della distribuzione spaziale delle regioni italiane in cui le regioni più estreme, come Sardegna, Sicilia o Valle d'Aosta, risultano meno connesse.

In questo modo si nota che la Sardegna non presenta alcun collegamento al grafo, la Sicilia presenta soltanto il collegamento con la Calabria e la Valle d'Aosta è collegata esclusivamente a Liguria, Piemonte e Lombardia. Una volta ricavata la matrice di adiacenza, è possibile vedere l'indice di Moran globale per le diverse ondate e il corrispettivo p-valore nella Tabella 2. L'ipotesi alternativa H_1 è che il valore osservato della statistica I sia superiore al suo valore atteso $E(I)$ come definito nel paragrafo 2.1.2.



Figura 8: Grafo non orientato dei centroidi delle regioni italiane

Tabella 2: Indici di Moran globali per le diverse ondate di Covid-19

Ondata	Valore osservato	Valore atteso	p-valore ($\alpha = 0.05$)
1	0.088	-0.053	0.09033
2	-0.054	-0.053	0.506
3	-0.129	-0.053	0.7666
4	-0.117	-0.053	0.7309
5	-0.141	-0.053	0.7988

Si nota che il test globale di correlazione spaziale per tutte le ondate risulta non significativo, indicando che non è presente una correlazione spaziale globale nei dati sui positivi cumulati fra le varie ondate.

La Tabella 3 mostra gli indici di Moran locali (uno per ciascuna regione) relativi alla quarta ondata e la Figura 9 mostra l'intensità del coefficiente di correlazione spaziale indicato dalla colorazione delle regione in mappa (le Tabelle e i grafici relativi alle altre ondate sono presenti in Appendice).

Tabella 3: Indici di Moran locali per ogni regione; quarta ondata

Regione	Valore osservato	Valore atteso	p-value ($\alpha = 0.05$)
Abruzzo	0.4116	-0.0830	0.6602
Basilicata	0.5098	-0.2242	0.6969
Calabria	-0.6073	-0.0647	0.6091
Campania	-2.0305	-0.4782	0.5519
Emilia-Romagna	0.3740	-0.3696	0.7202
Friuli Venezia Giulia	0.0347	-0.1021	0.9154
Lazio	-1.1105	-0.3895	0.7614
Liguria	-1.8695	-0.0712	0.0956
Lombardia	-0.7604	-3.0111	0.6520
Marche	-0.4190	-0.0840	0.7498
Molise	-0.1006	-0.3897	0.9000
P.A. Bolzano	-2.9183	-0.1871	0.1139
P.A. Trento	-3.6546	-0.2557	0.0807
Piemonte	0.4771	-0.0280	0.4844
Puglia	-0.4564	-0.0243	0.5089
Sicilia	-0.2294	-0.0081	0.5909
Toscana	1.8094	-0.1109	0.1109
Umbria	-1.3177	-0.1836	0.4793
Valle d'Aosta	-2.9236	-0.1836	0.1279
Veneto	0.7288	-0.6615	0.6241

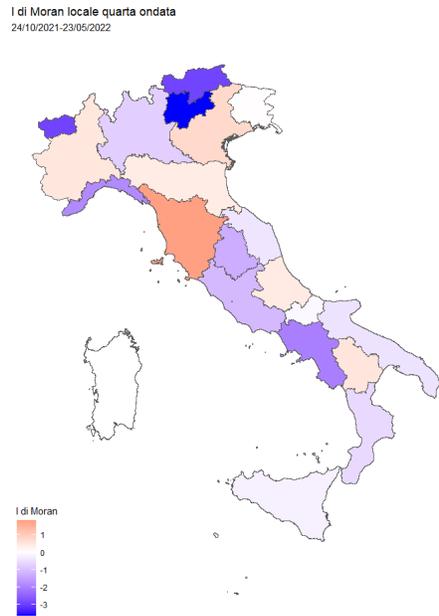


Figura 9: Indici di Moran locali per ogni regione; quarta ondata

Nonostante in Figura 9 siano visibili diverse intensità dell'indice locale di Moran, come visibile in Tabella 3, non ci sono correlazioni spaziali locali statisticamente significative.

Questi risultati possono essere influenzati dalla distribuzione della popolazione nelle diverse regioni, dunque si calcola un indice di Moran globale sul rischio relativo, che dà un valore osservato della statistica pari a 0.046, un valore atteso pari a -0.045 ottenuto tramite un *bootstrap* parametrico con 1000 ripetizioni e il p-valore pari a 0.38, che dunque conferma l'assenza di correlazione spaziale globale.

3.2.3 Analisi esplorativa temporale

Vengono mostrati i grafici rappresentanti l'andamento delle serie temporali dei nuovi positivi e della prevalenza dell'infezione da SARS CoV-2 nella popolazione, sia a livello nazionale che a livello regionale.

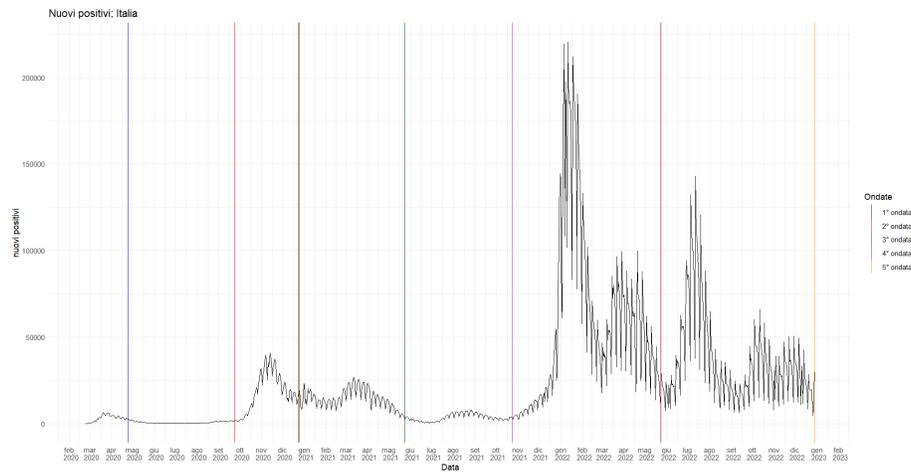


Figura 10: Serie temporale giornaliera dei nuovi positivi a livello nazionale

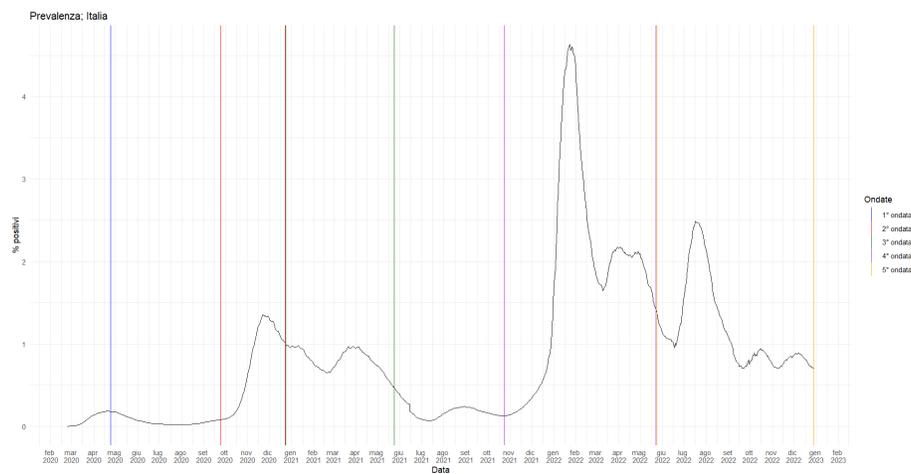


Figura 11: Serie temporale giornaliera della prevalenza a livello nazionale

Le Figure 10 e 11 mostrano l'evoluzione dell'epidemia da SARS-CoV-2 nell'intera finestra temporale presa in considerazione, visualizzata tramite il numero di nuovi casi positivi giornalieri e la prevalenza della malattia. In entrambi gli

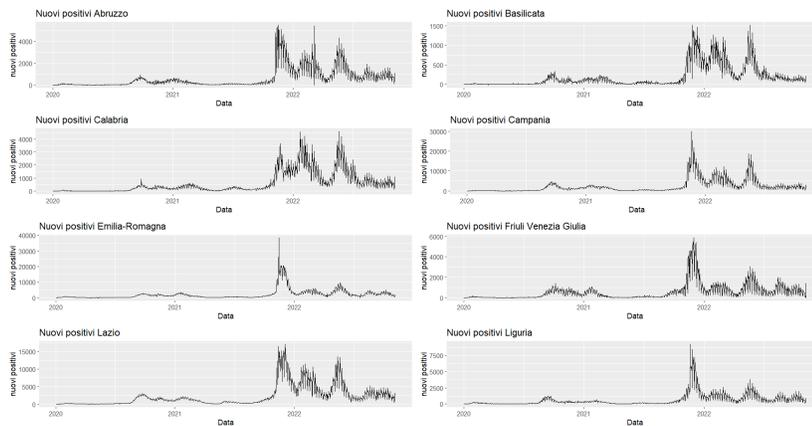
andamenti si notano diversi picchi nelle curve epidemiche che possono essere ricondotti ai diversi periodi di “ondate” del virus.

Le ondate di Covid-19 sono state definite a partire da un sostanziale aumento dei casi sino ad una diminuzione e al raggiungimento di un “plateau”, divise in periodi temporali come segue:

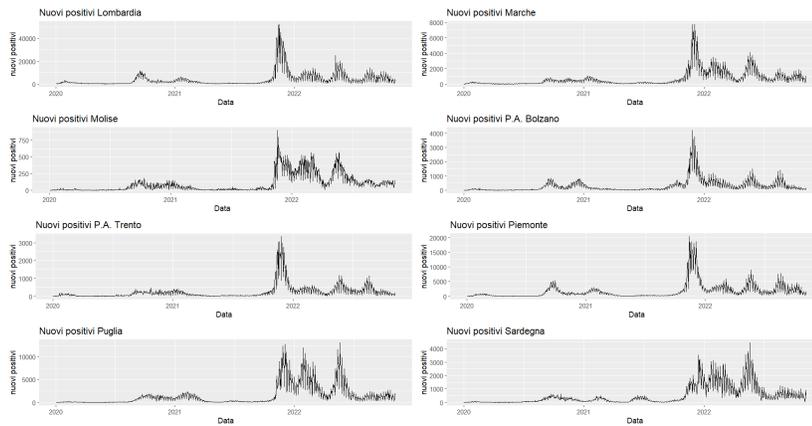
- La prima ondata parte dal primo giorno di rilevazione disponibile, ovvero il 24/02/2020 e termina il 24/04/2020. Mostra un picco pari a 6.557 nuovi casi di positività giornalieri e una prevalenza massima pari allo 0,18% della popolazione.
- La seconda ondata va dal 24/09/2020 al 23/12/2020 e mostra un massimo di nuovi positivi pari a 40.902 nuovi casi giornalieri e un massimo di prevalenza che è pari all'1,4%
- La terza ondata parte il 24/12/2020 e termina il 24/05/2021. Questa ondata sembra essere meno intensa della precedente, mostrando un massimo di 26.824 nuovi positivi giornalieri e un picco della prevalenza che raggiunge l'1% della popolazione nazionale.
- La quarta ondata va dal 24/10/2021 al 23/05/2022 e mostra di gran lunga numeri più elevati. Si notano, lungo il periodo, due picchi dei nuovi casi di positività giornalieri e due aumenti della prevalenza: si raggiunge il punto più alto della curva epidemica con 220.532 nuovi casi di positività giornalieri e una prevalenza pari al 4,6% della popolazione italiana; successivamente a una drastica diminuzione della curva, c'è un ulteriore aumento di incidenza che raggiunge circa i 100.000 nuovi casi giornalieri e un aumento della prevalenza intorno al 2,2%.
- La quinta ondata è definita nella finestra temporale compresa fra il 24/05/2022 e l'ultimo giorno di rilevazione disponibile, ovvero il 29/12/2022, e mostra cifre superiori alla terza ondata, ma inferiori alla quarta, con un massimo nei nuovi positivi giornalieri pari a 142.967 casi e un massimo nella prevalenza giornaliera pari a circa il 2,5% della popolazione.

Dalla Figura 10, inoltre, si può notare un'oscillazione settimanale delle rilevazioni dei nuovi positivi dovuta a un ridotto numero di casi riportati la domenica e ad un più elevato numero di casi riportati il lunedì. Tale caratteristica dei dati verrà presa in considerazione nelle successive analisi esplorative (paragrafo 3.2.4).

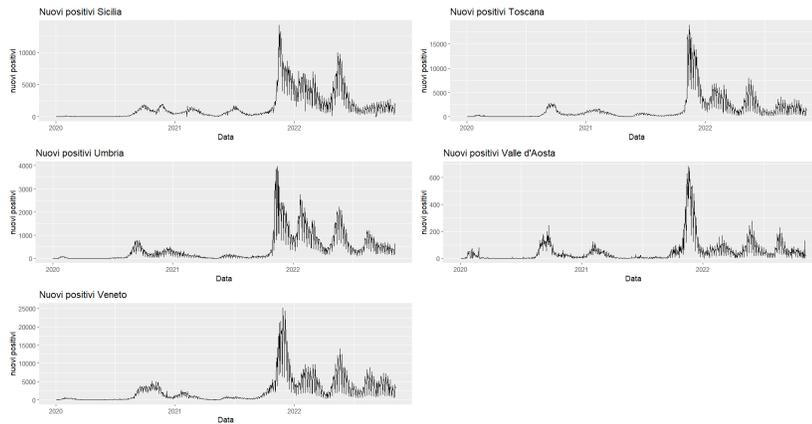
Successivamente si mostrano i grafici degli andamenti dei nuovi positivi e della prevalenza, divisi per regione di appartenenza al momento della rilevazione, per poter effettuare un confronto a livello regionale.



(a) Nuovi positivi; Abruzzo, Basilicata, Calabria, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria

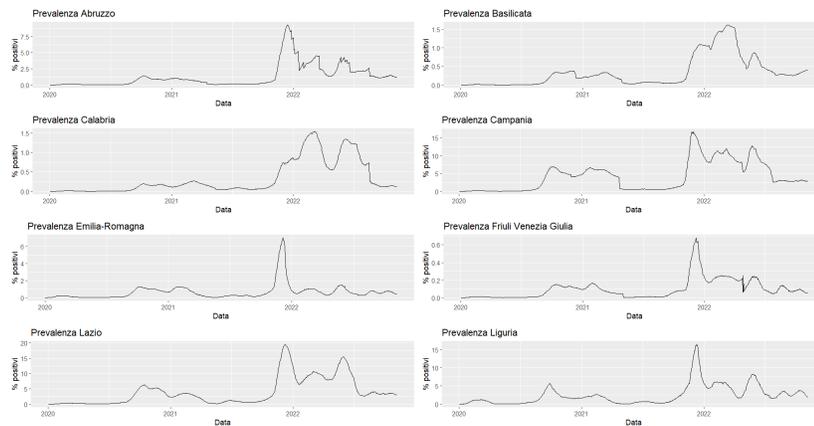


(b) Nuovi positivi; Lombardia, Marche, Molise, P.A. Bolzano, P.A. Trento, Piemonte, Puglia, Sardegna

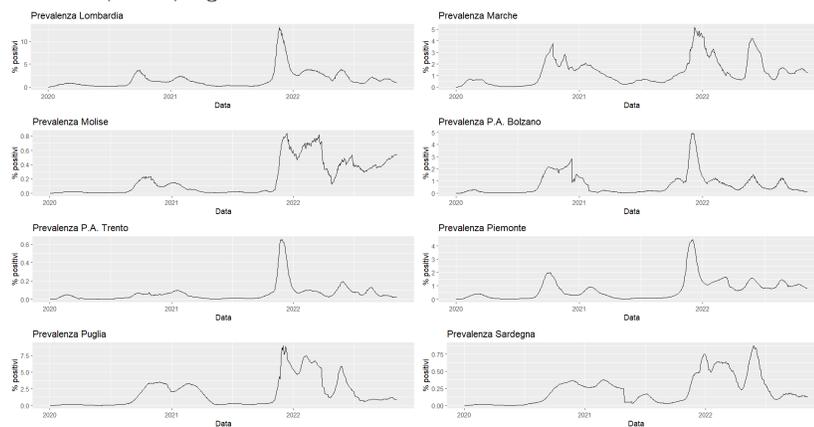


(c) Nuovi positivi; Sicilia, Toscana, Umbria, Valle d'Aosta, Veneto

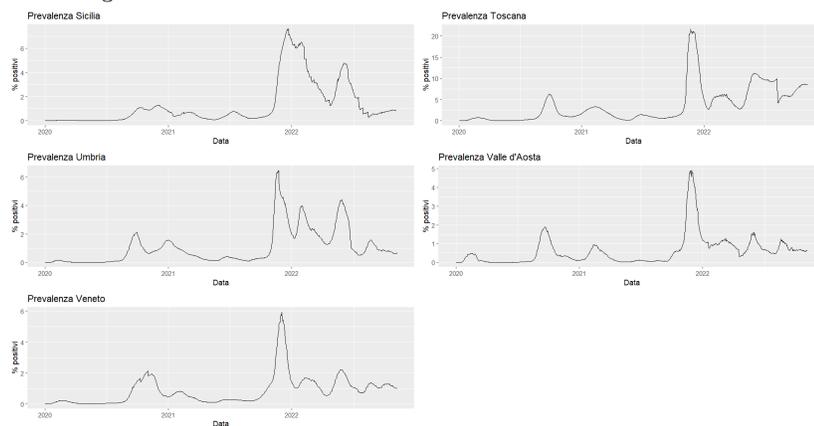
Figura 12: Serie temporale giornaliera dei nuovi positivi a livello regionale



(a) Prevalenza; Abruzzo, Basilicata, Calabria, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria



(b) Prevalenza; Lombardia, Marche, Molise, P.A. Bolzano, P.A. Trento, Piemonte, Puglia, Sardegna



(c) Prevalenza; Sicilia, Toscana, Umbria, Valle d'Aosta, Veneto

Figura 13: Serie temporale giornaliera della prevalenza a livello regionale

Sull'asse delle ordinate delle Figure 12 e 13, sono presenti rispettivamente i nuovi positivi giornalieri e la prevalenza giornaliera, cioè la percentuale del numero totale di positivi giornalieri (rispetto alla popolazione mensile regionale, assunta costante per tutti i giorni del mese).

Si nota che l'andamento dei nuovi positivi sembra rispecchiare quello nazionale per la maggior parte delle regioni, con qualche differenza fra le regioni soprattutto nel periodo della quarta ondata (dal 24/10/2021 al 23/05/2022): alcune regioni, come Lazio, Basilicata, Puglia e Umbria, mostrano due evidenti aumenti del numero di nuovi casi lungo il periodo appena considerato, raggiungendo dei punti di massimo simili all'interno delle regioni. Altre regioni, invece, evidenziano un unico, brusco, aumento nel periodo fra gennaio e febbraio 2022, come si può notare dall'andamento dei nuovi positivi delle regioni Piemonte e Lombardia e delle province di Trento e Bolzano.

La Figura 13 evidenzia delle differenze fra le regioni sia nell'andamento delle curve che nei valori di prevalenza. Infatti, si notano alcune regioni con dei picchi di prevalenza giornaliera che vanno dal 15% circa (Campania, Lombardia, Liguria) a quasi il 20% della popolazione come Lazio e Toscana, mentre altre regioni mostrano valori di prevalenza molto più bassi, come la Sardegna, il Molise o la provincia di Trento, le quali presentano un picco massimo di prevalenza inferiore all'1%.

In Figura 14, viene riportato il grafico dell'andamento cumulato dei nuovi positivi lungo tutto il periodo considerato, in cui emerge che la regione maggiormente colpita, in numeri assoluti, è la Lombardia, seguita da Veneto, Campania, Lazio ed Emilia Romagna. Il risultato è ovviamente influenzato dall'ammontare della popolazione di queste regioni, fra le più popolose d'Italia.

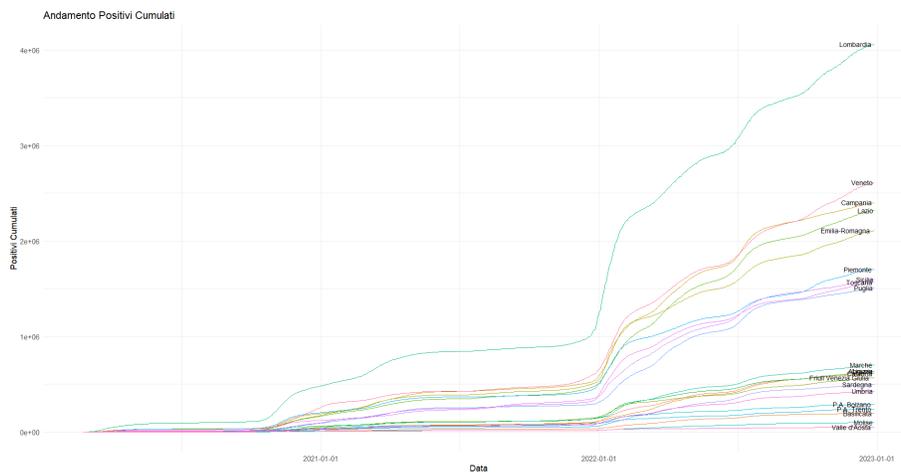


Figura 14: Nuovi positivi giornalieri cumulati

3.2.4 Risultati del modello lineare generalizzato dinamico

Viene adattato un modello lineare generalizzato dinamico (DGLM) con distribuzione di Poisson sui nuovi casi di positività giornalieri, con lo scopo di studiare la dinamica dell'epidemia. Secondo quanto riportato nel paragrafo 2.2.1 si utilizza l'algoritmo di *importance sampling* per simulare 100 campioni contenenti il vettore dei nuovi casi giornalieri per regione. Ai campioni simulati viene assegnato un peso e viene calcolata la mediana dei 100 valori campionati, e ordinati secondo il loro peso, per ogni giorno del periodo temporale considerato. Inoltre, viene stimato il trend lineare locale $\alpha_t = \log(\mu_t)$, con μ_t che rappresenta il conteggio dei nuovi positivi attesi, e β_t che rappresenta la pendenza di α_t che varia nel tempo, come definito nel paragrafo 2.2.1.

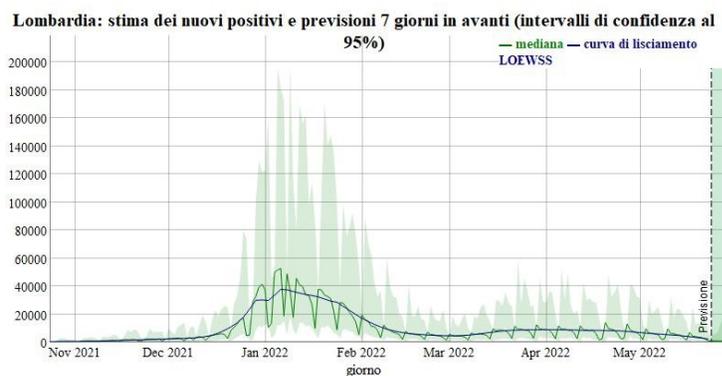


Figura 15: Stima dei nuovi positivi giornalieri in Lombardia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

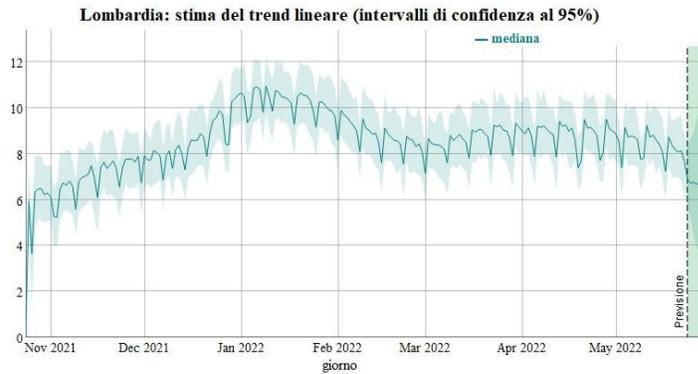


Figura 16: Stima della mediana del trend lineare α_t in Lombardia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

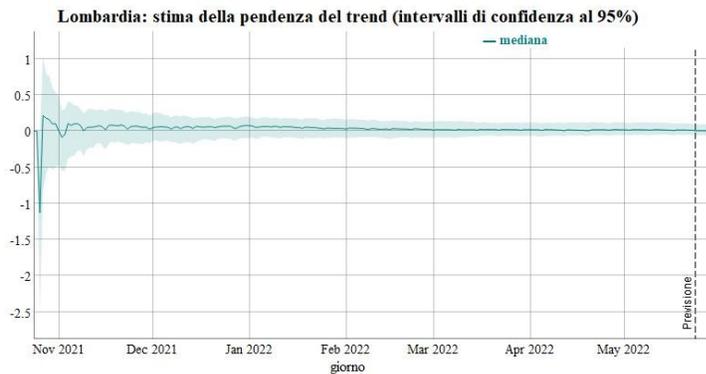


Figura 17: Stima della mediana della pendenza β_t del trend lineare in Lombardia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

La Figura 15 mostra i nuovi casi positivi giornalieri stimati dal DGLM in Lombardia (le altre regioni sono in Appendice) con le previsioni fino a 7 giorni in avanti, mostrando un intervallo di confidenza per i conteggi dei nuovi positivi abbastanza ampio, sia per le stime che per le previsioni (come discusso nel Paragrafo 2.2.1, il DGLM non è adatto a previsioni di lungo periodo).

È stata rappresentata una curva lisciata tramite metodo LOWESS, con parametro di liscio $h = 1/10$, per lisciare la curva dalle oscillazioni settimanali del lunedì e della domenica.

Il grafico dei conteggi stimati riesce ad individuare il forte aumento dei nuovi

positivi del periodo di gennaio 2022, come evidenziato in Figura 4, e inoltre rileva il secondo leggero aumento fra marzo e aprile 2022.

Questo risultato è visibile anche dalla Figura 16, che mostra la mediana del trend lineare stimato dal DGLM, mostrando un trend in crescita già a partire da novembre 2021 fino ad arrivare ad un massimo a gennaio 2022, mostrando poi il secondo aumento, meno intenso, ad aprile 2022.

La Figura 17 mostra che la pendenza del trend non si discosta significativamente da 0, indicando una crescita costante.

Un caso particolare può essere rappresentato dal Molise, che, come già visibile in Figura 13.b, mostra due aumenti del numero dei nuovi positivi di simile intensità, a gennaio 2022 e ad aprile 2022, e i conteggi dei nuovi positivi stimati vedono un iniziale aumento che resta quasi costante fino a maggio 2022.

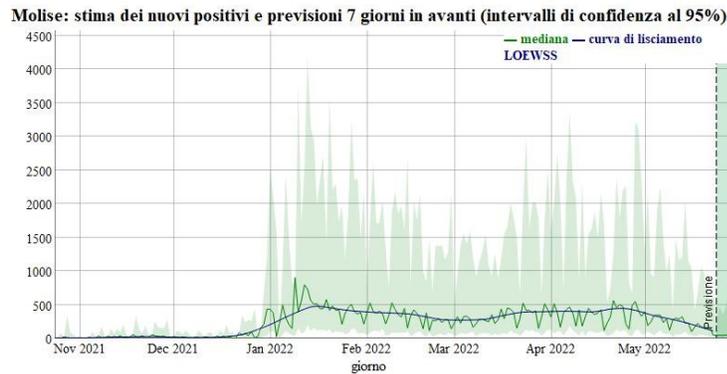


Figura 18: Stima DGLM dell'incidenza giornaliera in Molise, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata.

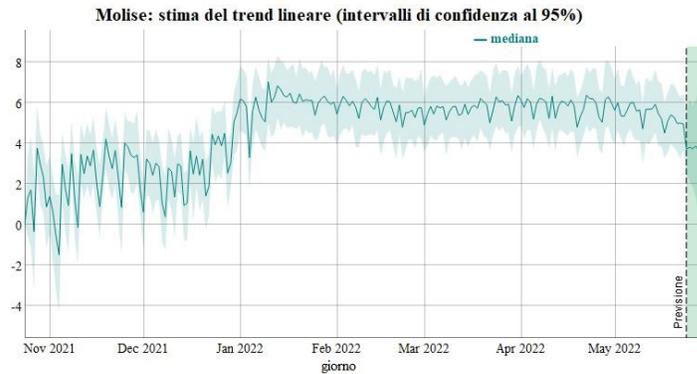


Figura 19: Stima DGLM della mediana del trend lineare α_t in Molise, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata.

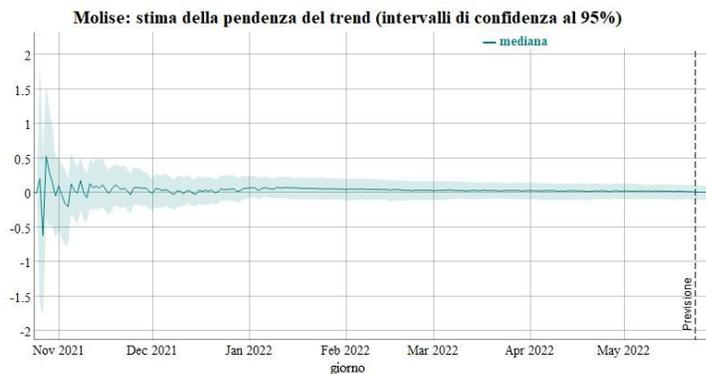


Figura 20: Stima DGLM della mediana della pendenza β_t del trend lineare in Molise, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata.

3.2.5 Stima del numero di riproduzione R_t

Successivamente, con i metodi discussi nel paragrafo 2.2.2, viene stimato il numero di riproduzione R_t utilizzando un intervallo seriale parametrico, assumendo che la media della distribuzione empirica della durata (in giorni) fra l'infezione del caso primario e del caso secondario infettato da esso sia pari a 6.68 e che la deviazione standard della stessa distribuzione sia pari a 4.88. Tali valori sono stati ricavati dallo studio di Cereda et al.(2020), come spiegato nel paragrafo 2.2.2.

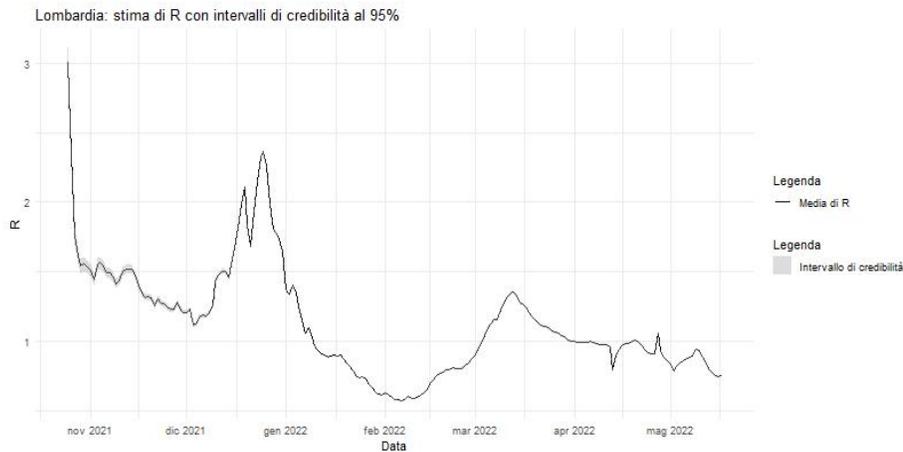


Figura 21: Stima di R_t in Lombardia, con intervallo di confidenza al 95%; quarta ondata

La Figura 21 mostra la stima del numero di riproduzione, R_t , in Lombardia (la stima di R_t per le altre regioni è presente in Appendice). Si nota un valore iniziale di R_t che decresce rapidamente, dovuto ai valori di media e deviazione standard assunti inizialmente.

L'indice R_t inizia ad aumentare nella seconda settimana di dicembre 2021 fino ad arrivare a un valore di poco inferiore a 2.5 ad inizio gennaio 2022, indicando che ogni caso di positività infetta, in media, fra i 2 e i 3 individui. Successivamente, l'indice diminuisce verso metà gennaio 2022 e scende sotto la soglia dell'1, che indica una diminuzione della curva epidemica, per poi aumentare nuovamente raggiungendo un picco di poco inferiore all'1.5 a fine marzo 2022.

Come ci si poteva aspettare, si nota come l'andamento dell'indice R_t rispecchi la curva epidemica dei nuovi positivi, anticipando il suo andamento di circa 2 settimane.

3.3 Risultati del modello Poisson gerarchico

Viene adattato un modello di regressione Poisson gerarchico con criterio di stima Bayesiana a componenti separabili, come descritto nel paragrafo 2.3.2, in cui si vuole stimare l'effetto a posteriori della componente spaziale, della componente temporale e degli effetti fissi delle variabili esplicative sul numero di nuovi casi di positività giornalieri.

Il modello tiene conto della distribuzione della popolazione tramite un offset sul logaritmo del conteggio atteso dei nuovi positivi, in modo che il $\log(\theta)$ possa essere interpretabile come rischio relativo.

Si osserva un valore della statistica test di Potthoff-Whittinghill (PW) pari a 3.4×10^{14} e un p-valore del test pari a 0.001, potendo dunque affermare che i rischi relativi non sono omogenei e seguono una distribuzione Gamma.

Vengono introdotte nel modello due variabili esplicative, entrambe con natura di conteggio: il totale dei vaccini somministrati giornalmente e il numero di soggetti in isolamento domiciliare ufficialmente rilevati, con lo scopo di osservare se la manovra di prevenzione dei contagi basata sulla somministrazione di vaccini e la manovra di contenimento dell'espansione del virus basata sull'isolamento domiciliare abbiano un effetto sul rischio relativo.

Si specifica che, a fronte di 125 osservazioni mancanti nei vaccini (2.8% del totale), è stata effettuata un'interpolazione lineare degli stessi per stimarne il numero mancante.

Nella Tabella 4 vengono mostrate le medie a posteriori, con relativo intervallo di credibilità, degli effetti fissi relativi all'isolamento domiciliare e alla somministrazione dei vaccini.

Variabile	Media a post.	Int. Cr. 2.5%	Int. Cr. 97.5%
Intercetta	-0.453	-0.454	-0.451
Vaccini	1.44×10^{-6}	1.39×10^{-6}	1.49×10^{-6}
Isolamento domiciliare	1.97×10^{-6}	1.96×10^{-6}	1.98×10^{-6}

Tabella 4: Medie a posteriori degli effetti fissi

Si può notare che l'intervallo di credibilità dei coefficienti stimati $\hat{\beta}_1$ e $\hat{\beta}_2$ non comprende lo 0 e dunque i coefficienti si possono considerare significativi. Si nota che entrambi i coefficienti stimati sono positivi, il che significa che un aumento della somministrazione dei vaccini e un aumento dei soggetti in isolamento domiciliare fa aumentare il rischio relativo. Il risultato osservato è contro-intuitivo, visti i numerosi studi a supporto dell'efficacia dei vaccini come strumento di prevenzione e dell'isolamento domiciliare come manovra di contenimento dei contagi. Si ipotizza, dunque, che un fenomeno così complesso come la diffusione dei contagi da Covid-19 richieda molte più variabili da tenere in considerazione, non presenti nei dati a disposizione. Ci si aspetterebbe che, in una situazione ideale in cui fossero presenti tutte e sole le variabili che spiegano il fenomeno studiato, i coefficienti stimati relativi alla somministrazione dei vaccini e all'isolamento domiciliare diventino negativi e rimangano significativi.

In Figura 22 è rappresentata la media a posteriori del trend temporale, in cui si nota un andamento molto simile alla mediana del trend lineare stimato con il modello lineare generalizzato dinamico in fase di analisi esplorativa (Figura 14): è presente un aumento a partire da novembre 2021 fino al raggiungimento di un picco intorno a gennaio 2022, per poi diminuire fino a marzo 2022 e vedere un nuovo aumento fra aprile e maggio 2022. Gli intervalli di credibilità, infine, sono molto stretti, confermando l'effetto significativo della componente temporale sul rischio relativo, eccetto in corrispondenza dei giorni nei quali la media a posteriori interseca lo 0.

La Figura 23 mostra la media a posteriori degli effetti spaziali, con i valori

riportati nella Tabella 5. L'effetto spaziale è da interpretare come l'effetto che ha sulla regione i -esima l'aumento dei nuovi casi di positività nelle regioni vicine. Nel Nord-Italia si nota un forte effetto spaziale negativo nella Provincia Autonoma di Trento e nel Friuli-Venezia Giulia: questo risultato suggerisce che l'aumento dei casi nelle regioni vicine alla provincia di Trento e al Friuli-Venezia Giulia ha influenzato negativamente l'aumento dei casi in queste due aree. Si ipotizza che in queste aree siano state adottate misure di contenimento più stringenti e che gli spostamenti verso di esse siano diminuiti. Nelle restanti regioni del Nord-Italia (Valle d'Aosta, Piemonte, Lombardia, Emilia-Romagna, Veneto e la provincia di Bolzano) sembra esserci un effetto spaziale vicino allo zero, fatta eccezione per la Liguria che, invece, è la regione italiana con il valore di media a posteriori più alta. Gli effetti spaziali positivi, dunque, si possono interpretare come l'influenza positiva dell'aumento dei nuovi casi di Covid-19 nelle aree vicine sull'aumento dei casi nell'area considerata, al netto degli effetti delle covariate.

Anche le regioni del Centro-Italia (in particolare Toscana, Umbria, Marche e Lazio) mostrano una media a posteriori positiva. L'idea ipotizzata è speculare rispetto a quanto detto per le aree di Trento e del Friuli-Venezia Giulia: in queste regioni si suppone che ci siano stati maggiori flussi in entrata della popolazione che dunque ne hanno influenzato positivamente il numero di contagi.

Il Sud-Italia e le isole, infine, si presentano eterogenee, con la maggioranza di effetti negativi (Sardegna, Basilicata, Calabria, Molise e Sicilia), mentre si notano medie a posteriori positive in Campania e Puglia. Per spiegare la maggioranza di effetti negativi al Sud si ipotizza una maggiore difficoltà a raggiungere il meridione, a causa anche dei numerosi voli cancellati o a ridotta capienza a causa delle restrizioni in essere.

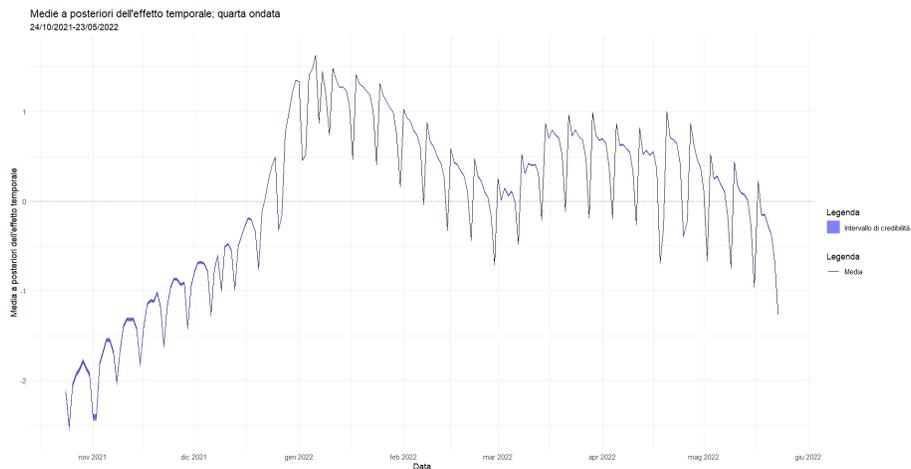


Figura 22: Media a posteriori del trend temporale stimato; l'area intorno al trend rappresenta il suo intervallo di credibilità al 95%

Medie a posteriori dell'effetto spaziale; quarta ondata
24/10/2021-23/05/2022

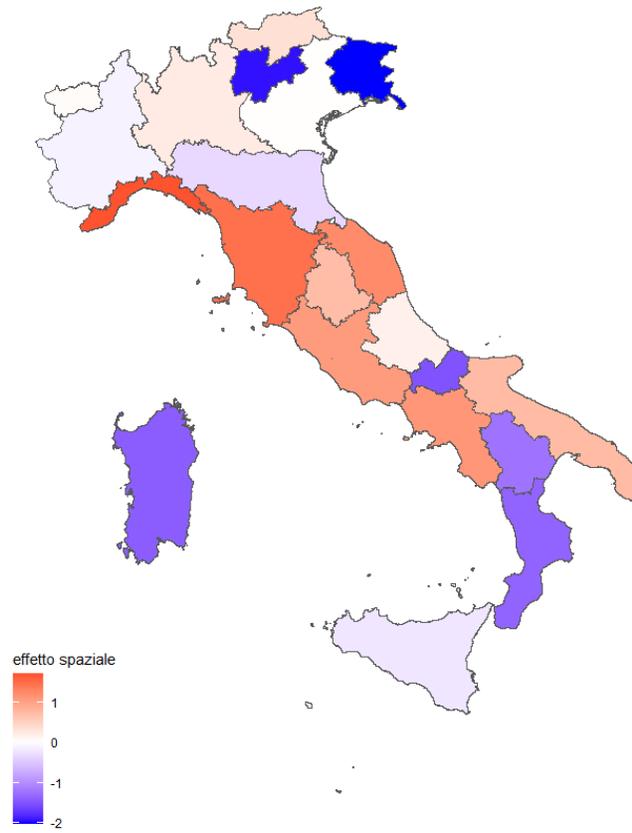


Figura 23: Media a posteriori degli effetti spaziali stimati

Regione	Media a posteriori	Int. Cr. 2.5%	Int. Cr. 97.5%
Abruzzo	0.16409076	0.16061525	0.16756627
Basilicata	-1.23701092	-1.24286297	-1.23115887
Calabria	-1.37707202	-1.38064945	-1.37349458
Campania	1.10655384	1.10443220	1.10867548
Emilia-Romagna	-0.34798073	-0.35017011	-0.34579135
Friuli Venezia Giulia	-2.03766111	-2.04147352	-2.03384870
Lazio	1.06356789	1.06139561	1.06574017
Liguria	1.69418030	1.69066891	1.69769169
Lombardia	0.22932972	0.22655821	0.23210122
Marche	1.21949482	1.21614910	1.22284055
Molise	-1.51082099	-1.51918881	-1.50245317
P.A. Bolzano	0.34068850	0.33568333	0.34569367
P.A. Trento	-1.94636201	-1.95192554	-1.94079848
Piemonte	-0.11569373	-0.11804390	-0.11334357
Puglia	0.73924049	0.73685245	0.74162853
Sardegna	-1.43195986	-1.43621012	-1.42770960
Sicilia	-0.22145851	-0.22392661	-0.21899041
Toscana	1.45895079	1.45667607	1.46122551
Umbria	0.71894684	0.71479318	0.72310049
Valle d'Aosta	0.04334716	0.03139040	0.05530391
Veneto	0.01566897	0.01364233	0.01769561

Tabella 5: Medie a posteriori stimate degli effetti spaziali per regione, con intervallo di credibilità al 95%

4 Conclusioni

Sulla base dei risultati ottenuti nel progetto di tesi, si possono trarre alcune conclusioni riguardanti la caratterizzazione dell'epidemia di Covid-19 in Italia.

In primo luogo, l'analisi esplorativa spaziale ha evidenziato una notevole variabilità nella diffusione del virus tra le diverse regioni italiane. Questa disomogeneità può essere dovuta alle diverse misure di contenimento adottate a livello locale. Ad esempio, regioni come Lombardia, Veneto ed Emilia-Romagna hanno mostrato un impatto iniziale più severo, il che ha portato a misure di lockdown più stringenti rispetto ad altre aree. Inoltre, l'utilizzo dei centroidi associati alle regioni italiane, per localizzare il conteggio dei nuovi positivi, potrebbe aver limitato l'effetto relativo alla componente spaziale nel contesto dell'analisi esplorativa e di modellazione.

Inoltre, l'analisi esplorativa temporale ha rivelato che i picchi del numero di nuovi positivi sono stati fortemente correlati con periodi di maggiore mobilità e interazione sociale, come le festività. Questo sottolinea l'importanza delle misure di distanziamento sociale e delle restrizioni alla mobilità per controllare l'epidemia.

La stima del numero di riproduzione R_t si è rivelata un utile strumento per monitorare l'andamento dei contagi: questo approccio può essere esteso all'analisi di altre epidemie, offrendo una metodologia robusta per la gestione delle crisi sanitarie future.

Il modello di regressione Poisson gerarchico ha portato alla luce alcuni risultati contro-intuitivi, attribuibili alla mancanza di ulteriori variabili utili a spiegare il complesso fenomeno dell'epidemia da Covid-19. Nonostante questo risultato inatteso, la stima dei coefficienti del modello si è rivelata precisa e robusta, con intervalli di confidenza stretti che confermano la solidità delle stime. Anche le stime degli effetti temporali e spaziali si sono rivelate precise e hanno restituito risultati plausibili.

Il modello ha catturato accuratamente l'evoluzione temporale dell'epidemia nella quarta ondata, evidenziando i picchi nell'effetto casuale temporale in corrispondenza dei picchi di nuovi casi di positività mostrati durante l'analisi esplorativa e ha migliorato la spiegazione dell'effetto casuale spaziale, inizialmente non rilevato dagli indici di correlazione spaziale globale e locale.

In conclusione, lo studio conferma l'importanza di un approccio multidimensionale nell'analisi delle epidemie, tenendo conto sia della componente spaziale che della componente temporale dei dati. Le raccomandazioni future includono la possibilità di espandere il lavoro presentato tramite un'analisi sui cluster della malattia e l'ampliamento del modello di regressione. Nello specifico, si propone di inserire un maggior numero di variabili nel modello in modo da spiegare più dettagliatamente la dinamica del rischio relativo, con l'obiettivo di ricavare dei risultati in linea con la nota efficacia dei vaccini e delle misure di contenimento sociale in risposta all'emergenza sanitaria affrontata.

Appendice

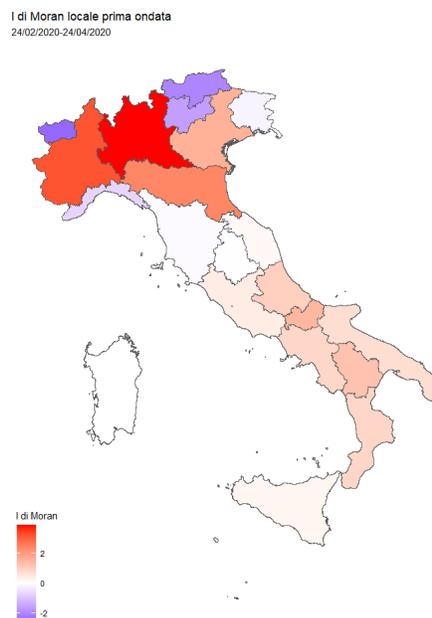


Figura 24: Indici di Moran locali per ogni regione; prima ondata

I di Moran locale seconda ondata
24/09/2020-23/12/2020

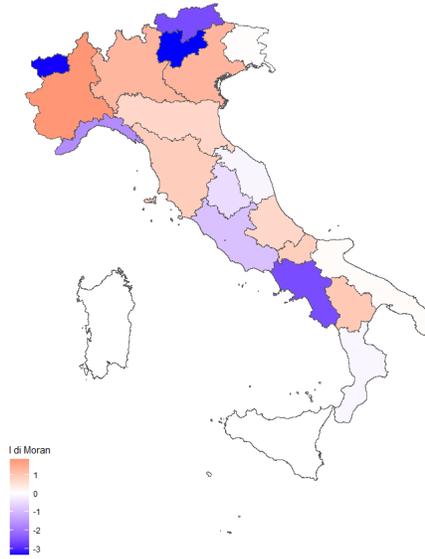


Figura 25: Indici di Moran locali per ogni regione; seconda ondata

I di Moran locale terza ondata
24/12/2020-24/05/2021

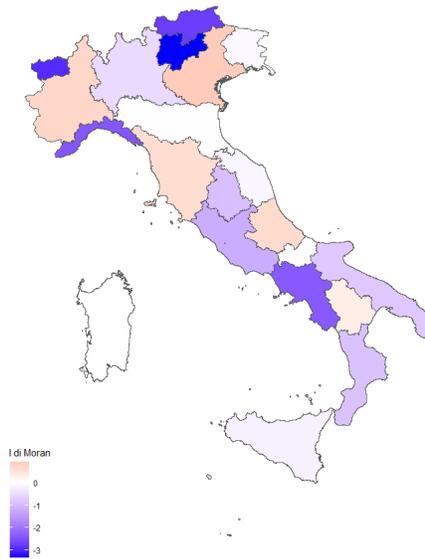


Figura 26: Indici di Moran locali per ogni regione; terza ondata

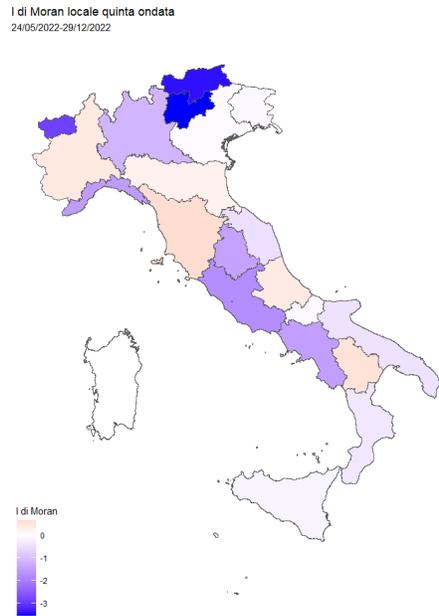


Figura 27: Indici di Moran locali per ogni regione; quinta ondata

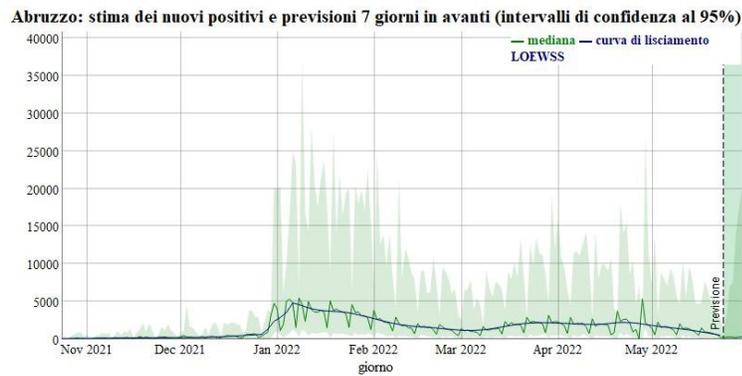


Figura 28: Stima dei nuovi positivi giornalieri in Abruzzo tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

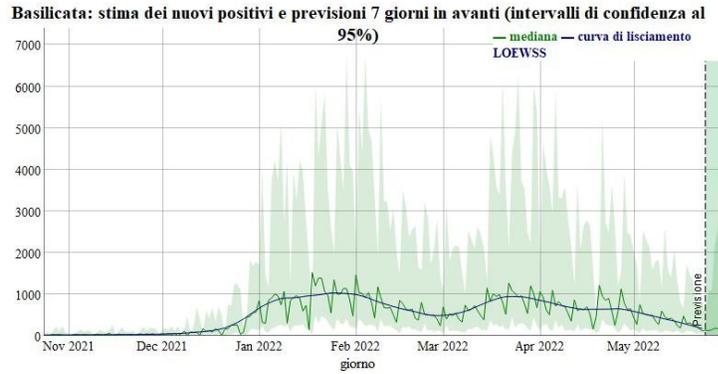


Figura 29: Stima dei nuovi positivi giornalieri in Basilicata tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

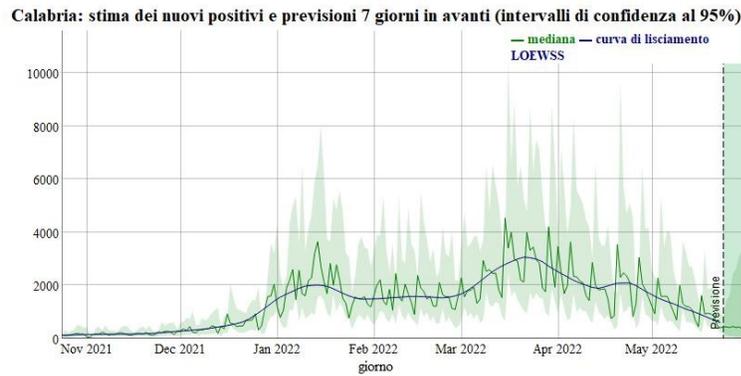


Figura 30: Stima dei nuovi positivi giornalieri in Calabria tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

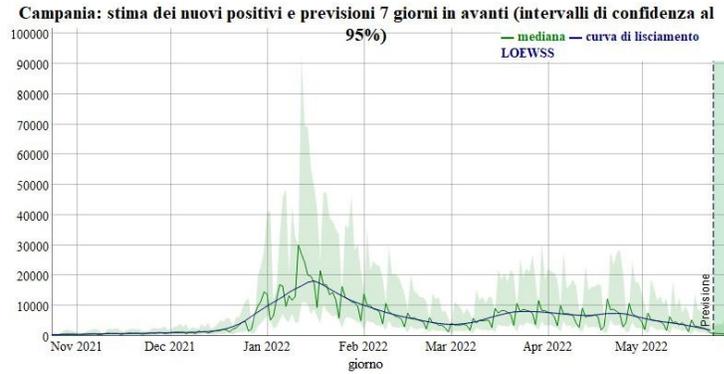


Figura 31: Stima dei nuovi positivi giornalieri in Campania tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata



Figura 32: Stima dei nuovi positivi giornalieri in Emilia-Romagna tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

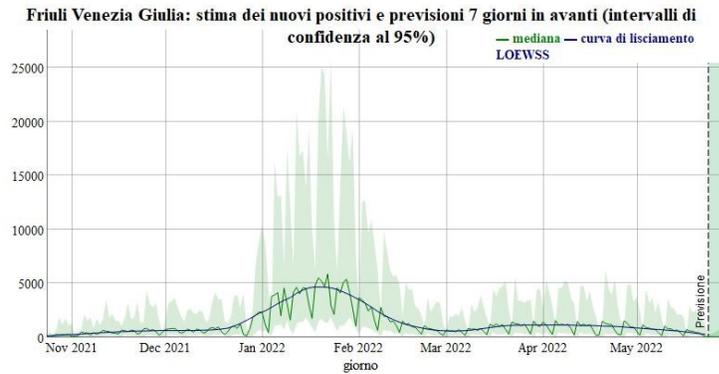


Figura 33: Stima dei nuovi positivi giornalieri in Friuli-Venezia Giulia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

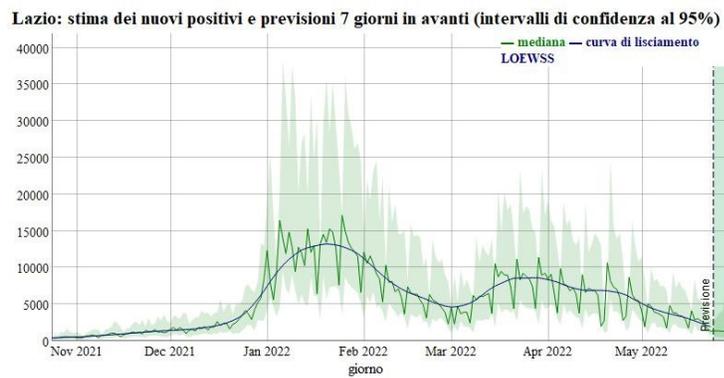


Figura 34: Stima dei nuovi positivi giornalieri nel Lazio tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

Liguria: stima dei nuovi positivi e previsioni 7 giorni in avanti (intervalli di confidenza al 95%)

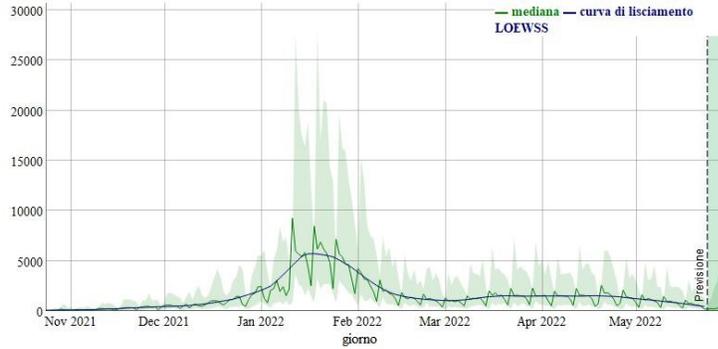


Figura 35: Stima dei nuovi positivi giornalieri in Liguria tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

Marche: stima dei nuovi positivi e previsioni 7 giorni in avanti (intervalli di confidenza al 95%)

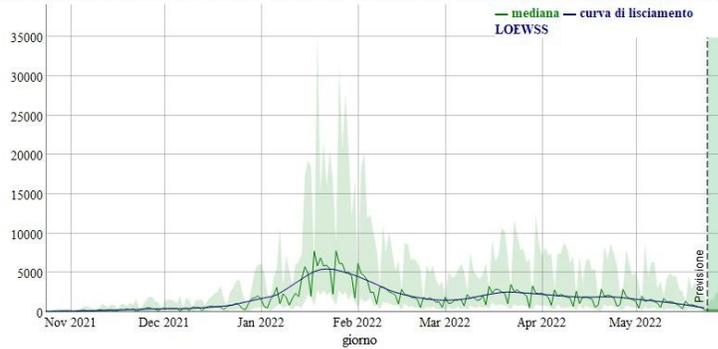


Figura 36: Stima dei nuovi positivi giornalieri nelle Marche tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

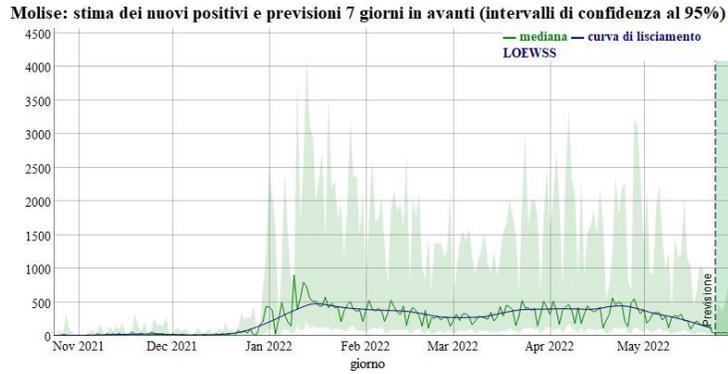


Figura 37: Stima dei nuovi positivi giornalieri in Molise tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

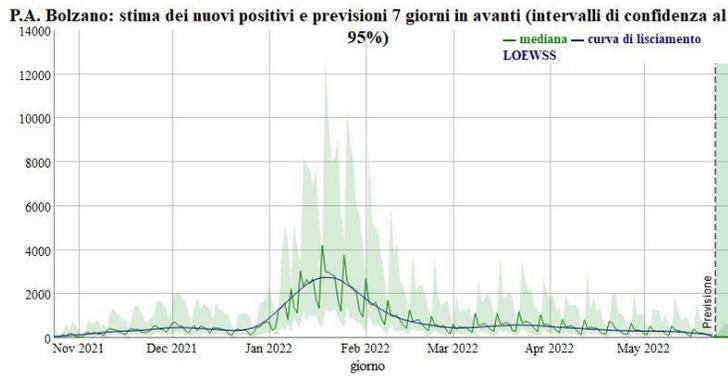


Figura 38: Stima dei nuovi positivi giornalieri a Bolzano tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

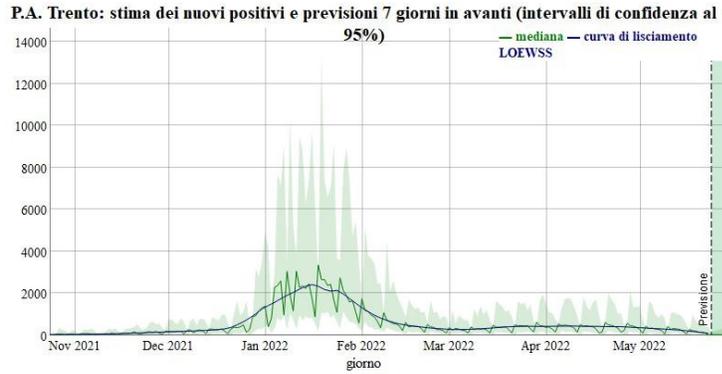


Figura 39: Stima dei nuovi positivi giornalieri a Trento tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

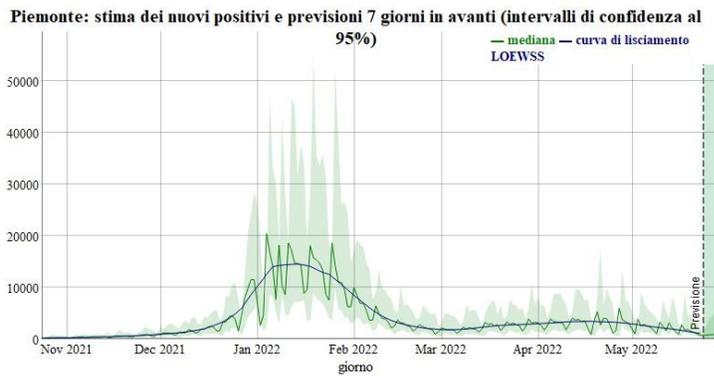


Figura 40: Stima dei nuovi positivi giornalieri in Piemonte tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

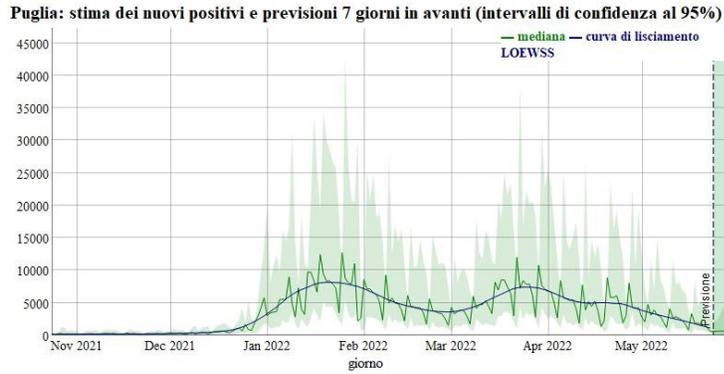


Figura 41: Stima dei nuovi positivi giornalieri in Puglia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

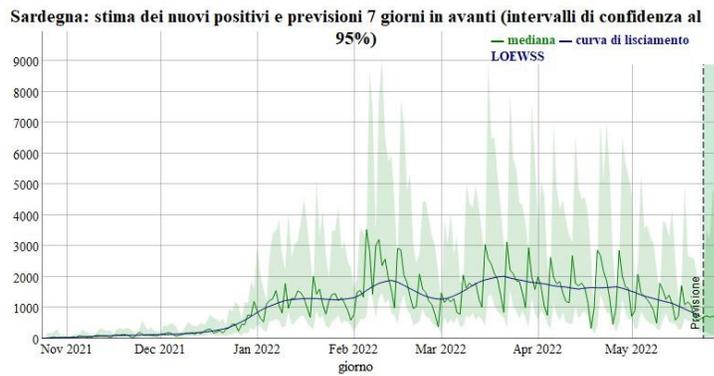


Figura 42: Stima dei nuovi positivi giornalieri in Sardegna tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

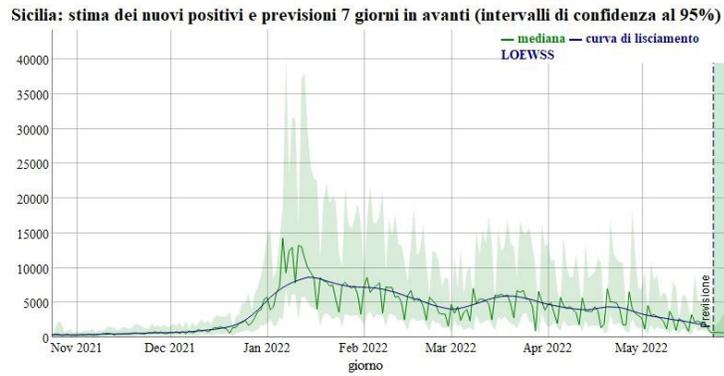


Figura 43: Stima dei nuovi positivi giornalieri in Sicilia tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

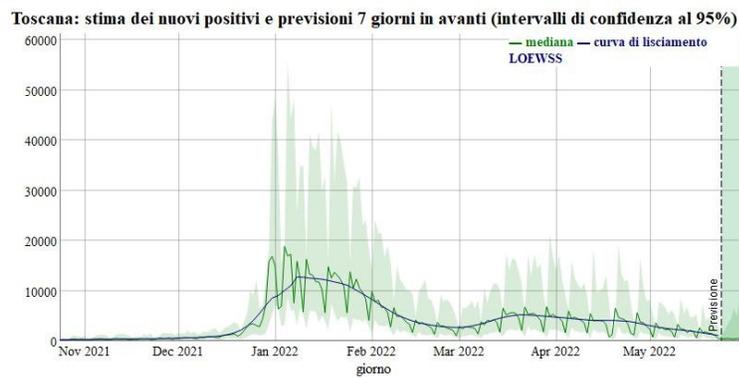


Figura 44: Stima dei nuovi positivi giornalieri in Toscana tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

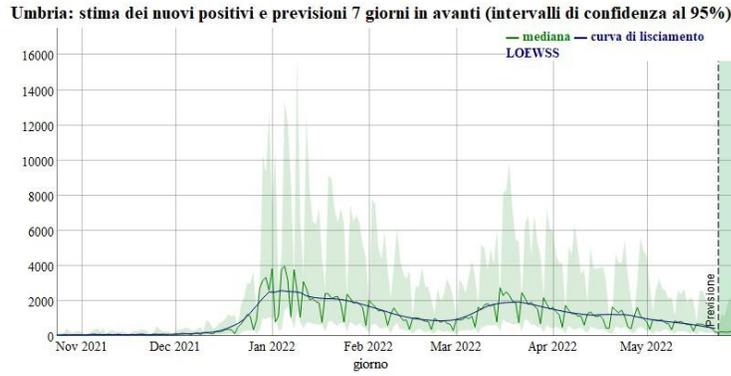


Figura 45: Stima dei nuovi positivi giornalieri in Umbria tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

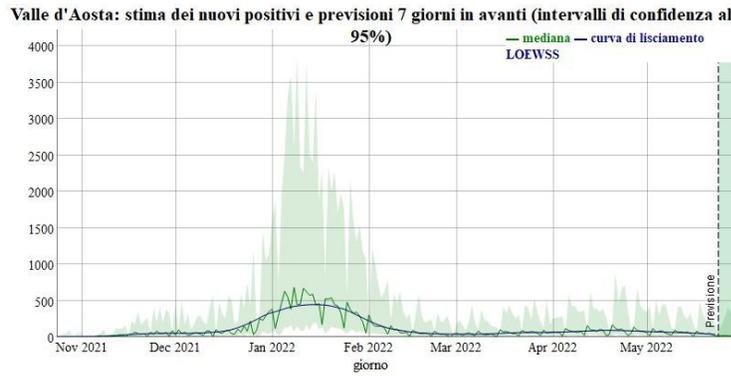


Figura 46: Stima dei nuovi positivi giornalieri in Valle d'Aosta tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

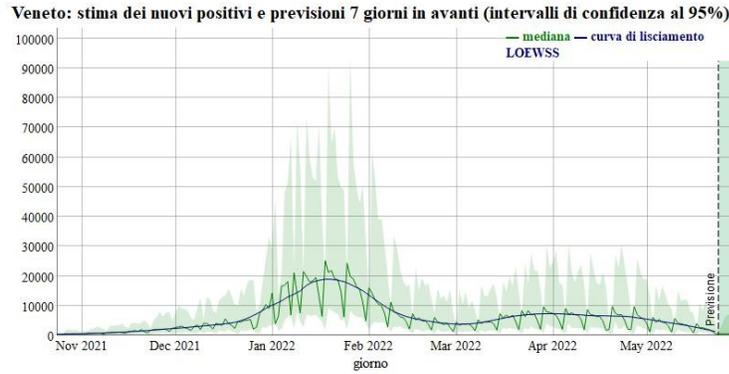


Figura 47: Stima dei nuovi positivi giornalieri in Veneto tramite il modello DGLM, con intervalli di confidenza al 95%, curva lisciata tramite LOWESS e previsioni 7 giorni in avanti; quarta ondata

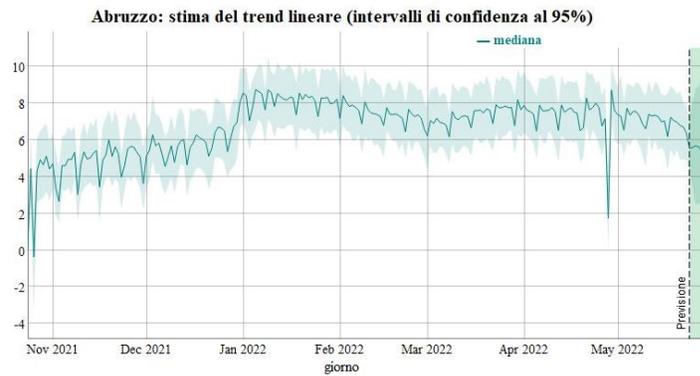


Figura 48: Stima della mediana del trend lineare α_t in Abruzzo tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata



Figura 49: Stima della mediana del trend lineare α_t in Basilicata tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

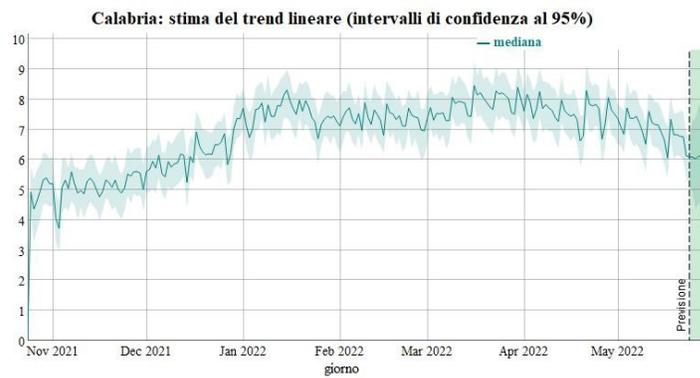


Figura 50: Stima della mediana del trend lineare α_t in Calabria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

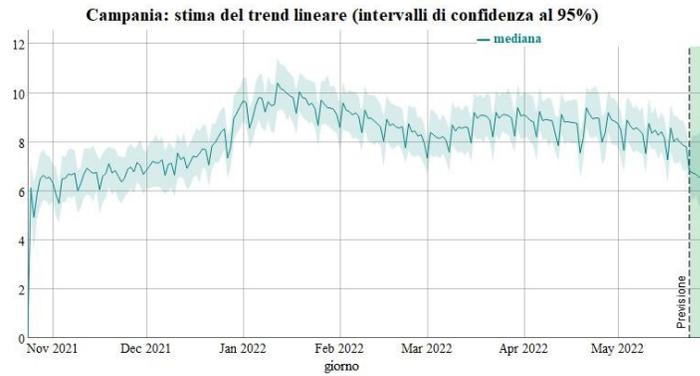


Figura 51: Stima della mediana del trend lineare α_t in Campania tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

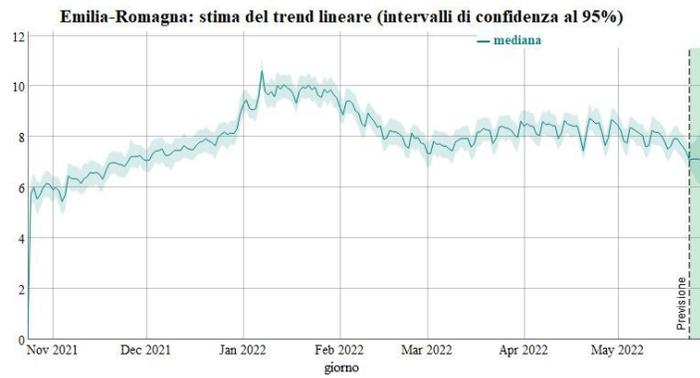


Figura 52: Stima della mediana del trend lineare α_t in Emilia-Romagna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

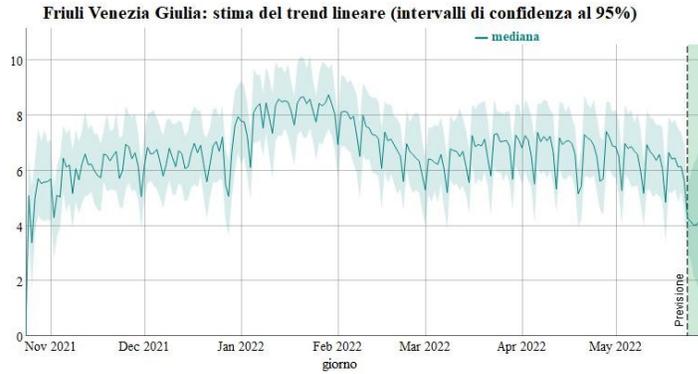


Figura 53: Stima della mediana del trend lineare α_t in Friuli-Venezia Giulia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

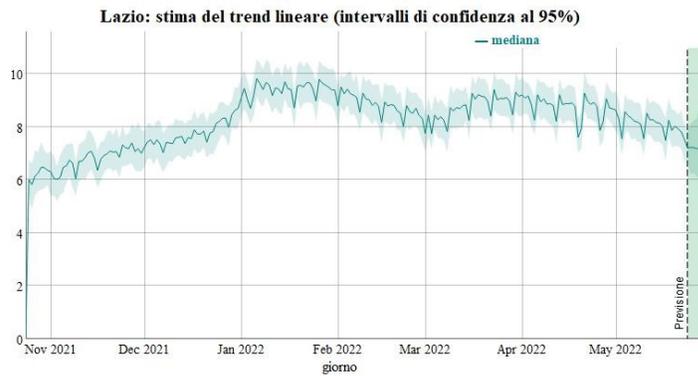


Figura 54: Stima della mediana del trend lineare α_t nel Lazio tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

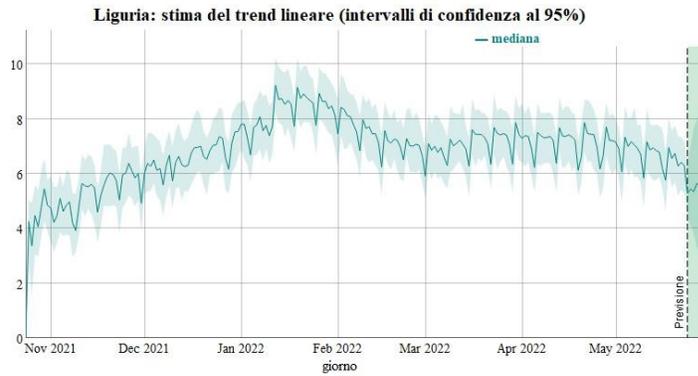


Figura 55: Stima della mediana del trend lineare α_t in Liguria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

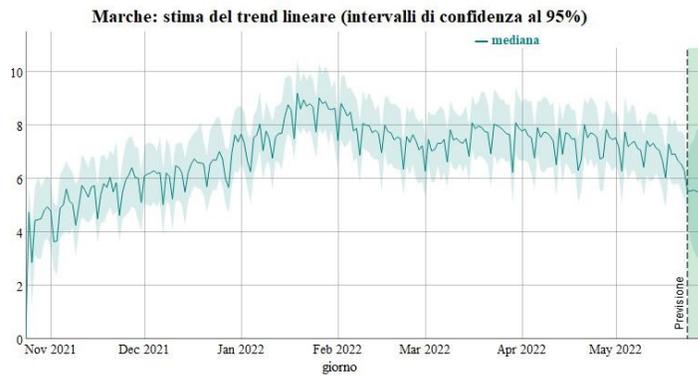


Figura 56: Stima della mediana del trend lineare α_t nelle Marche tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

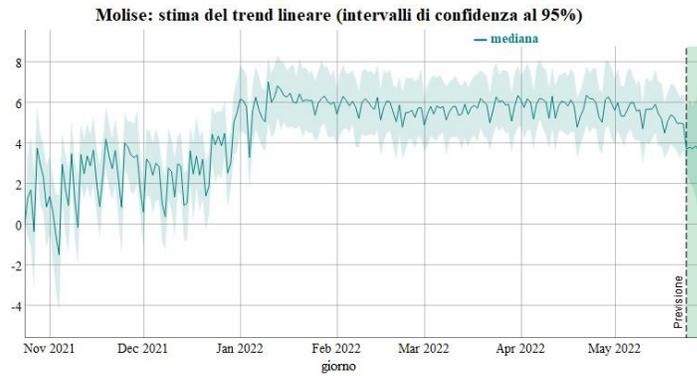


Figura 57: Stima della mediana del trend lineare α_t in Molise tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

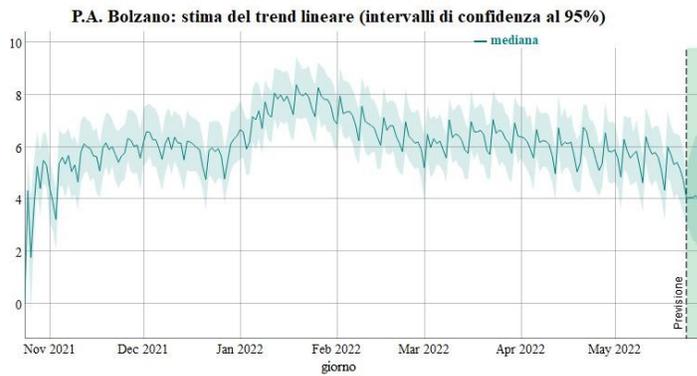


Figura 58: Stima della mediana del trend lineare α_t a Bolzano tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

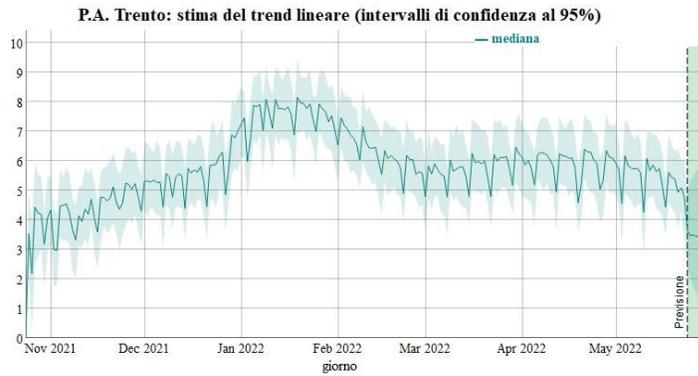


Figura 59: Stima della mediana del trend lineare α_t a Trento tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

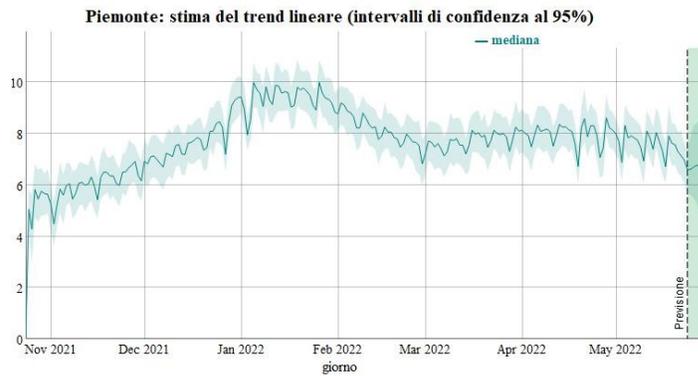


Figura 60: Stima della mediana del trend lineare α_t in Piemonte tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

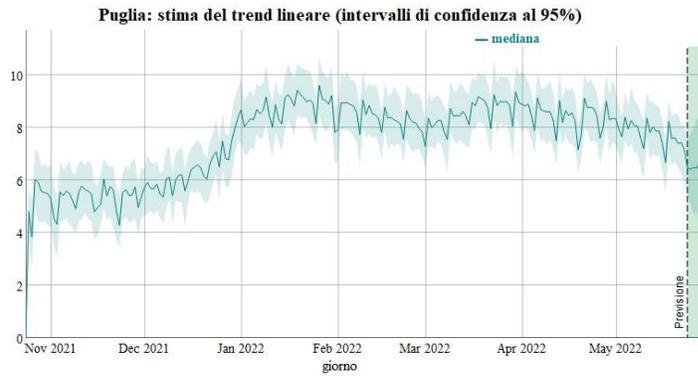


Figura 61: Stima della mediana del trend lineare α_t in Puglia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

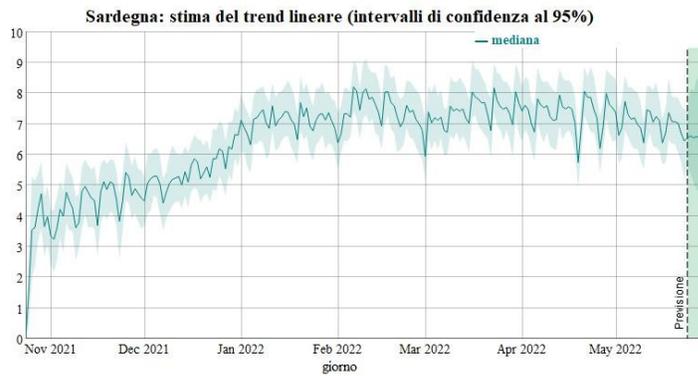


Figura 62: Stima della mediana del trend lineare α_t in Sardegna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

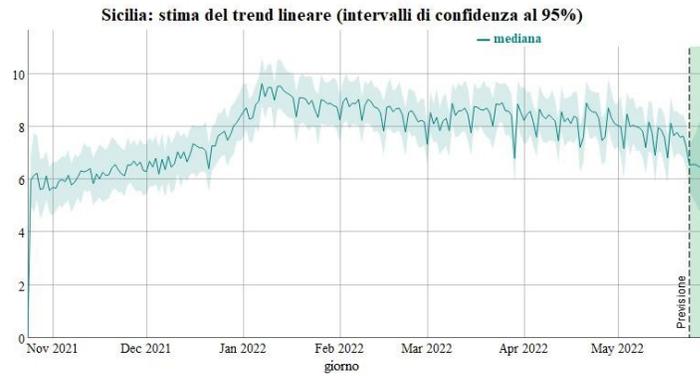


Figura 63: Stima della mediana del trend lineare α_t in Sicilia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

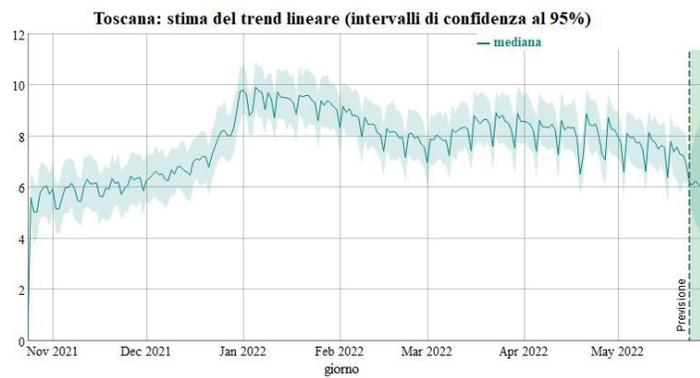


Figura 64: Stima della mediana del trend lineare α_t in Toscana tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

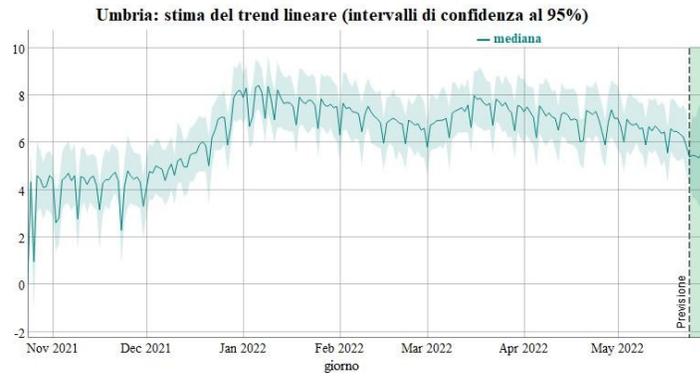


Figura 65: Stima della mediana del trend lineare α_t in Umbria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

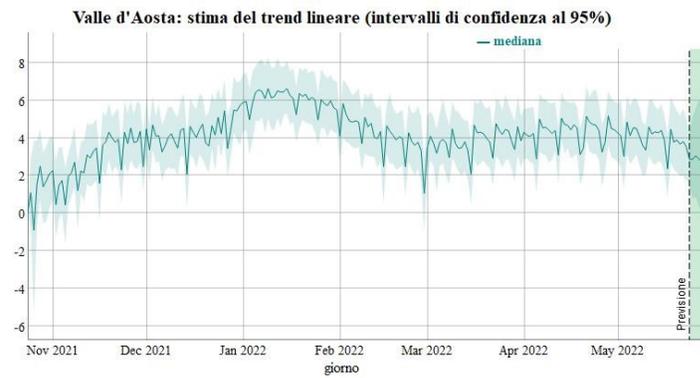


Figura 66: Stima della mediana del trend lineare α_t in Valle d'Aosta tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

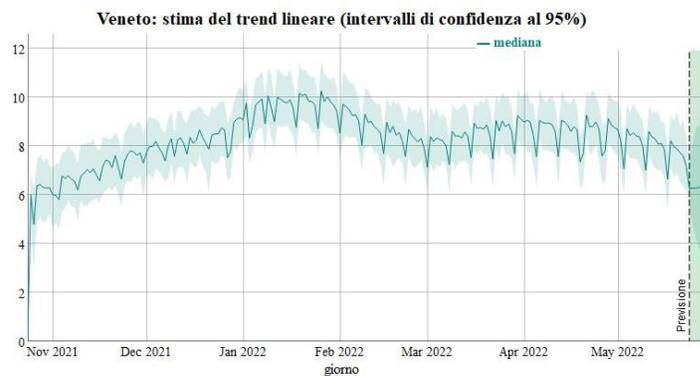


Figura 67: Stima della mediana del trend lineare α_t in Veneto tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

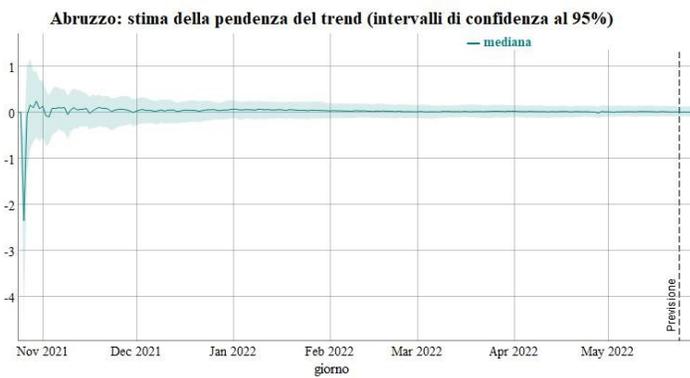


Figura 68: Stima della mediana della pendenza β_t del trend lineare in Abruzzo tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

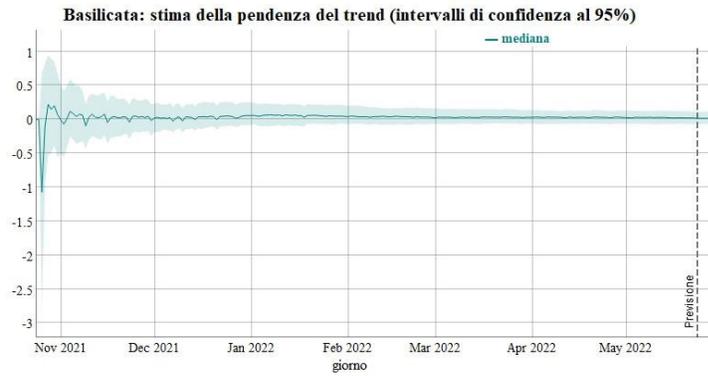


Figura 69: Stima della mediana della pendenza β_t del trend lineare in Basilicata tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

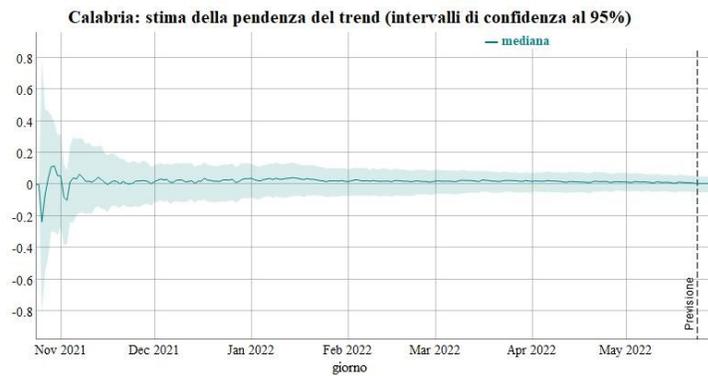


Figura 70: Stima della mediana della pendenza β_t del trend lineare in Calabria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

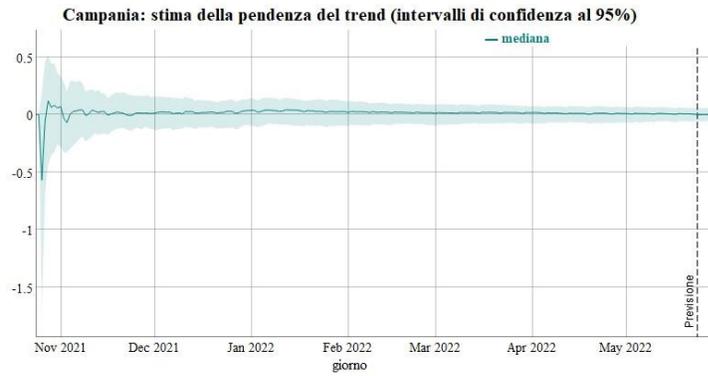


Figura 71: Stima della mediana della pendenza β_t del trend lineare in Campania tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

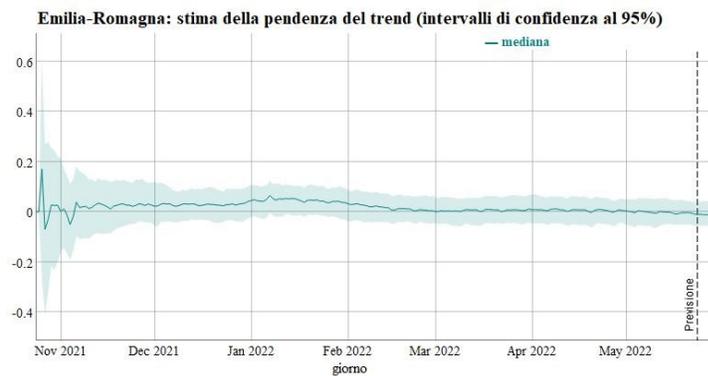


Figura 72: Stima della mediana della pendenza β_t del trend lineare in Emilia-Romagna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

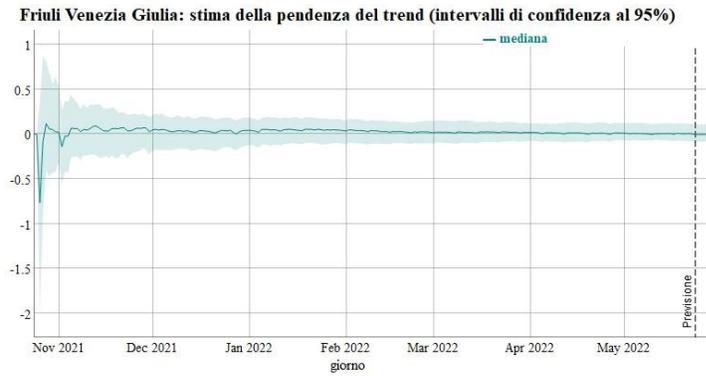


Figura 73: Stima della mediana della pendenza β_t del trend lineare in Friuli Venezia Giulia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

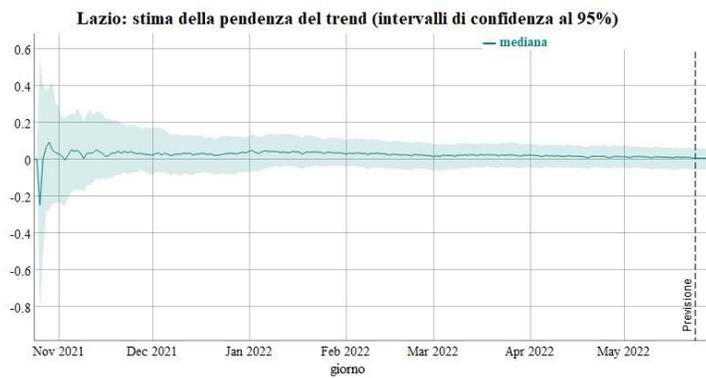


Figura 74: Stima della mediana della pendenza β_t del trend lineare nel Lazio tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

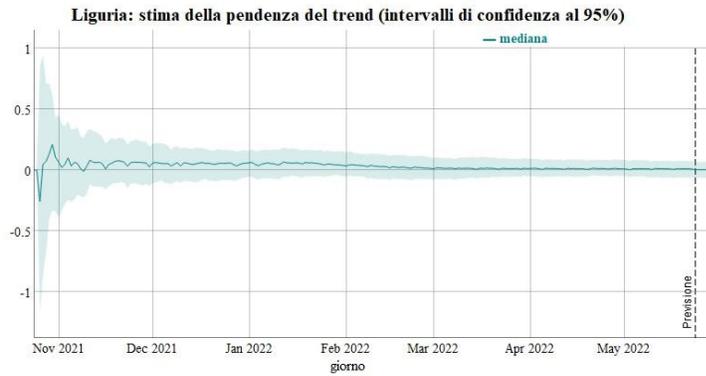


Figura 75: Stima della mediana della pendenza β_t del trend lineare in Liguria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

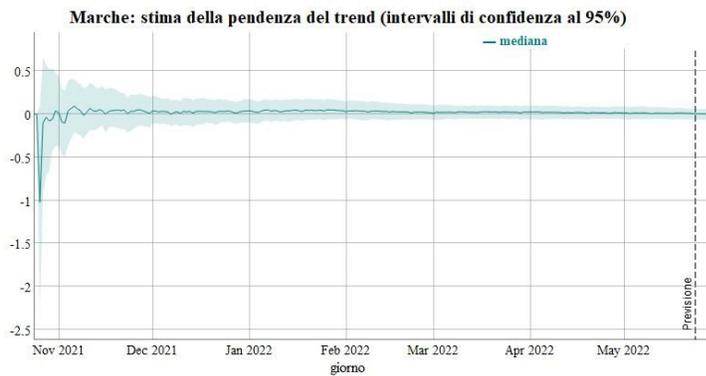


Figura 76: Stima della mediana della pendenza β_t del trend lineare nelle Marche tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

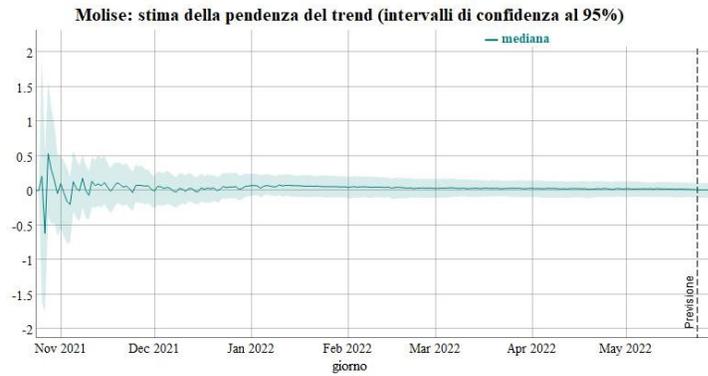


Figura 77: Stima della mediana della pendenza β_t del trend lineare in Molise tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

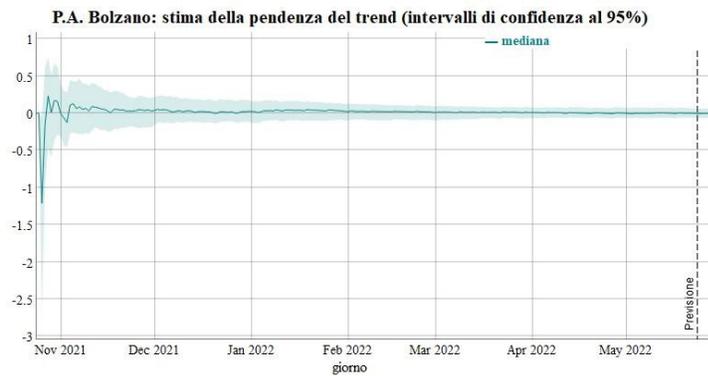


Figura 78: Stima della mediana della pendenza β_t del trend lineare a Bolzano tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

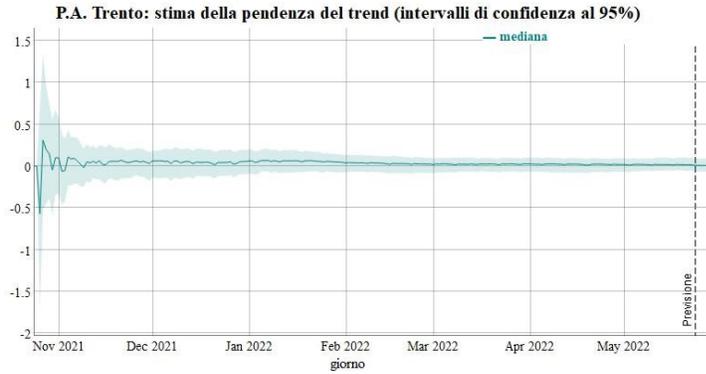


Figura 79: Stima della mediana della pendenza β_t del trend lineare a Trento tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

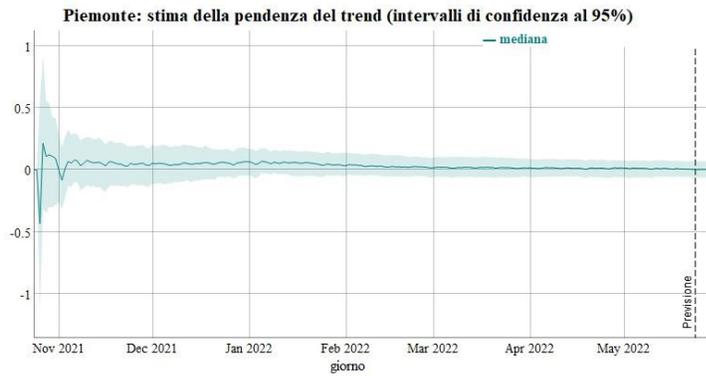


Figura 80: Stima della mediana della pendenza β_t del trend lineare in Piemonte tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

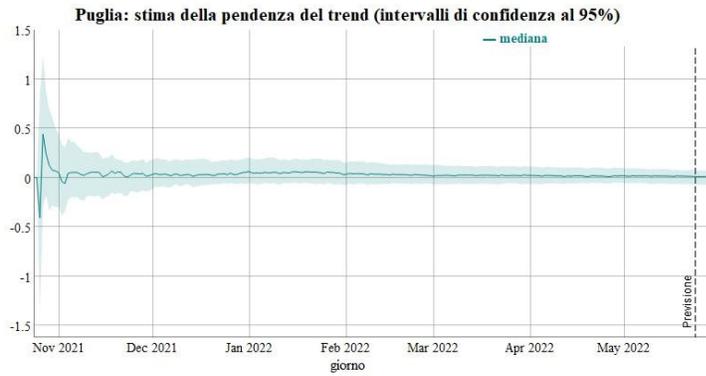


Figura 81: Stima della mediana della pendenza β_t del trend lineare in Puglia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

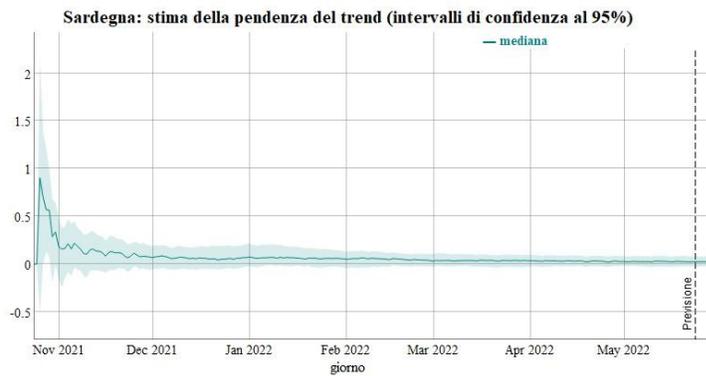


Figura 82: Stima della mediana della pendenza β_t del trend lineare in Sardegna tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

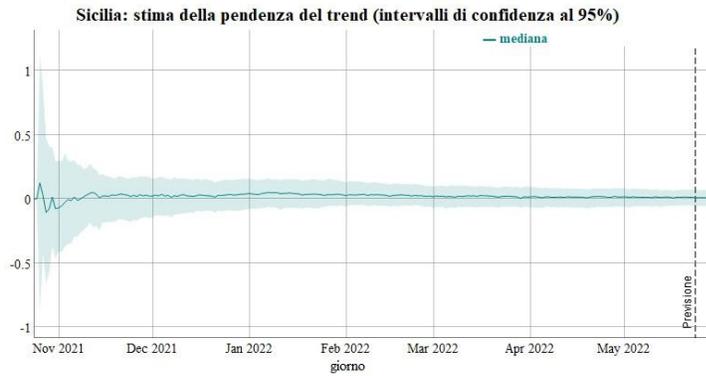


Figura 83: Stima della mediana della pendenza β_t del trend lineare in Sicilia tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

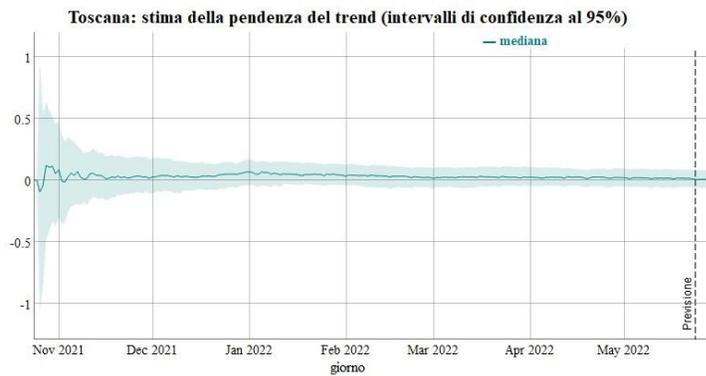


Figura 84: Stima della mediana della pendenza β_t del trend lineare in Toscana tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

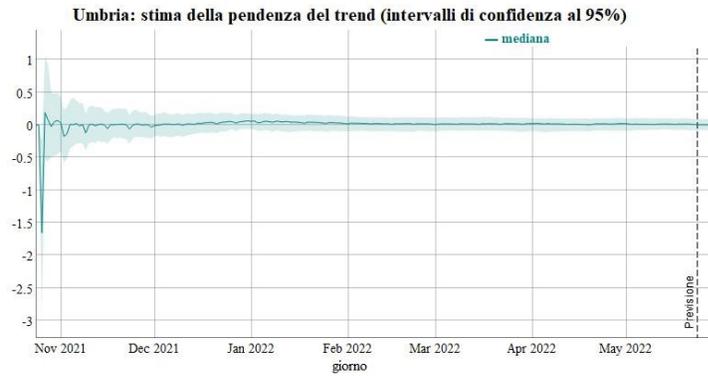


Figura 85: Stima della mediana della pendenza β_t del trend lineare in Umbria tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

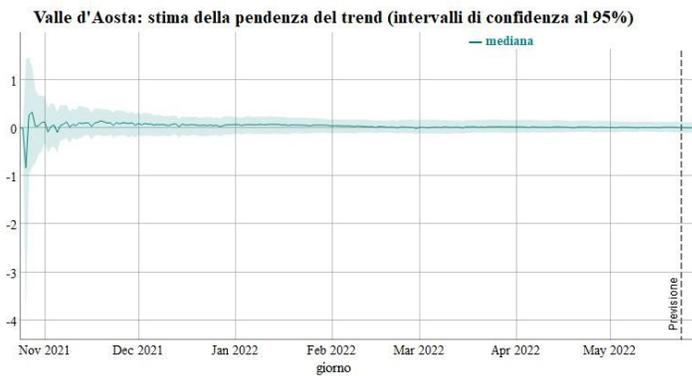


Figura 86: Stima della mediana della pendenza β_t del trend lineare in Valle d'Aosta tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

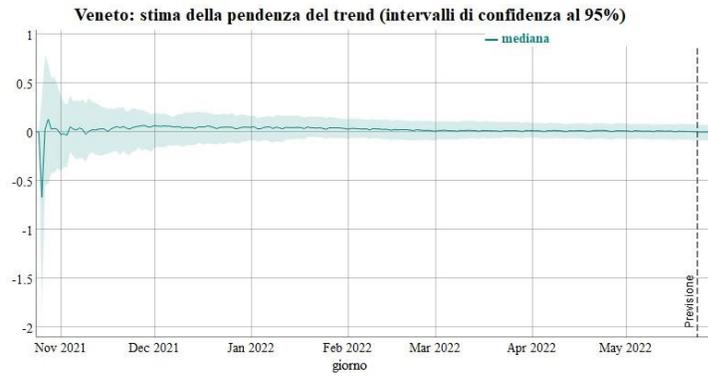


Figura 87: Stima della mediana della pendenza β_t del trend lineare in Veneto tramite il modello DGLM, con intervalli di confidenza al 95% e previsioni 7 giorni in avanti; quarta ondata

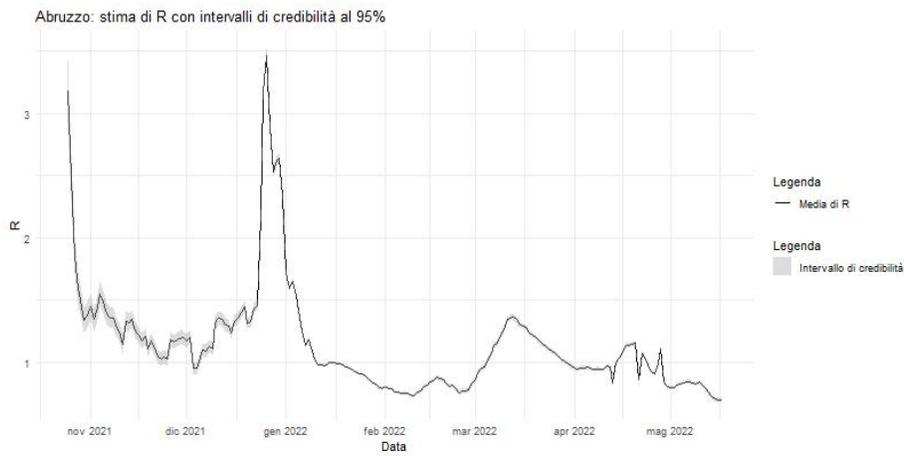


Figura 88: Stima di R_t in Abruzzo, con intervallo di confidenza al 95%; quarta ondata

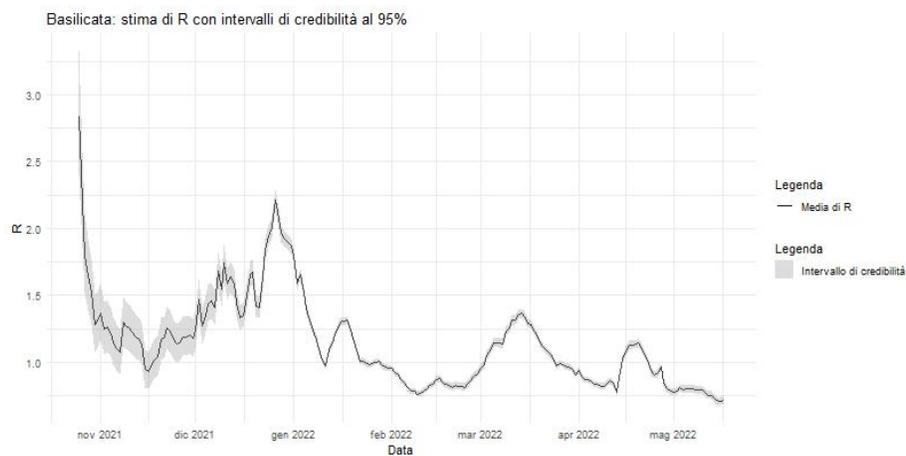


Figura 89: Stima di R_t in Basilicata, con intervallo di confidenza al 95%; quarta ondata

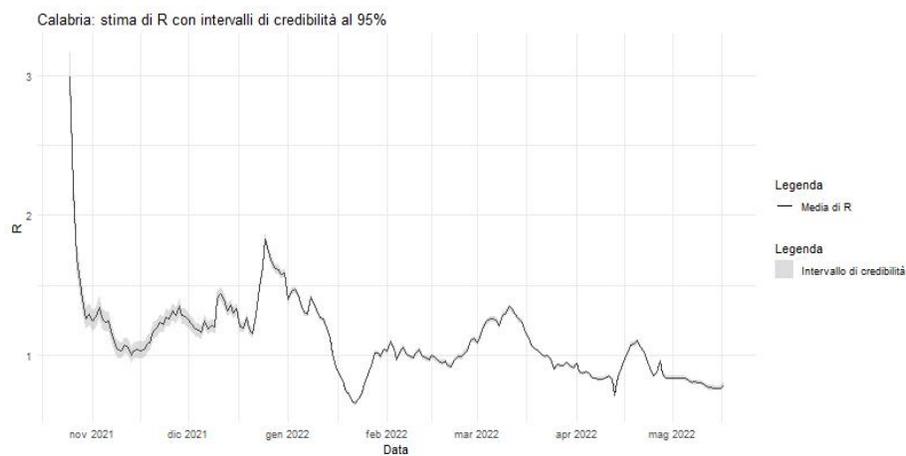


Figura 90: Stima di R_t in Calabria, con intervallo di confidenza al 95%; quarta ondata

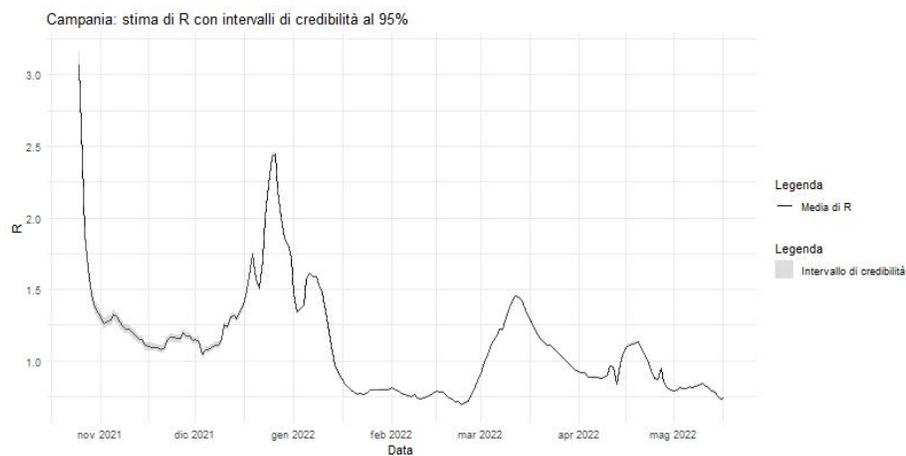


Figura 91: Stima di R_t in Campania, con intervallo di confidenza al 95%; quarta ondata

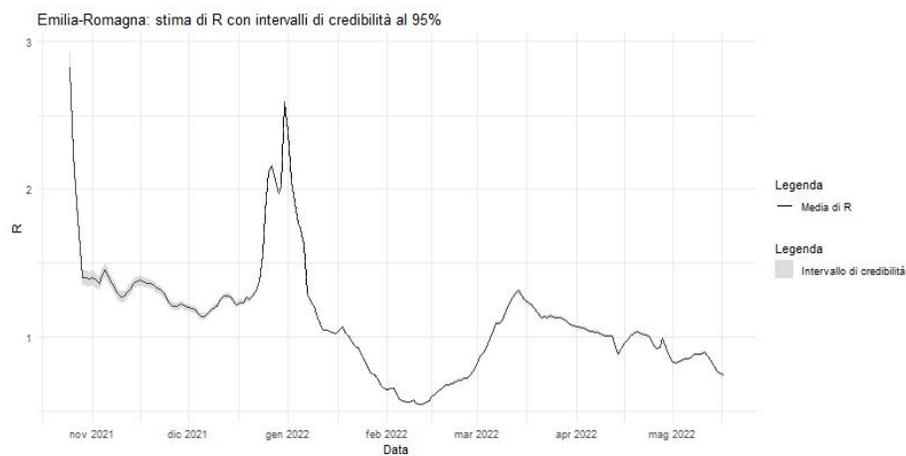


Figura 92: Stima di R_t in Emilia-Romagna, con intervallo di confidenza al 95%; quarta ondata

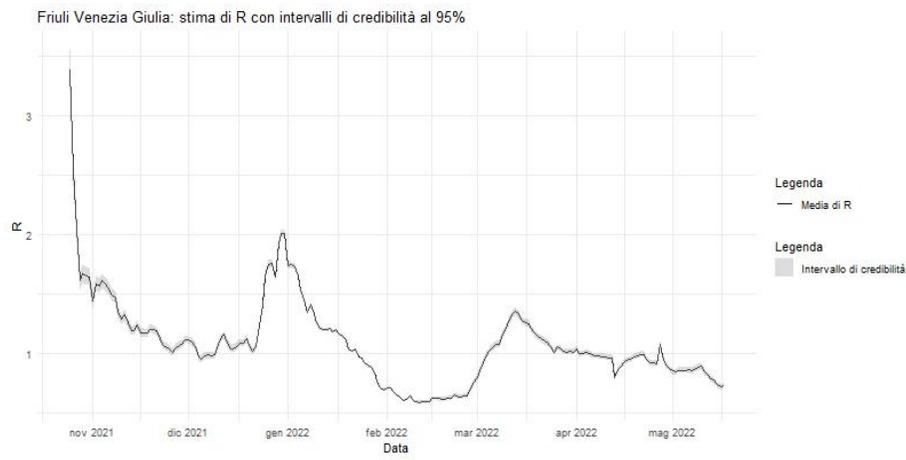


Figura 93: Stima di R_t in Friuli-Venezia Giulia, con intervallo di confidenza al 95%; quarta ondata

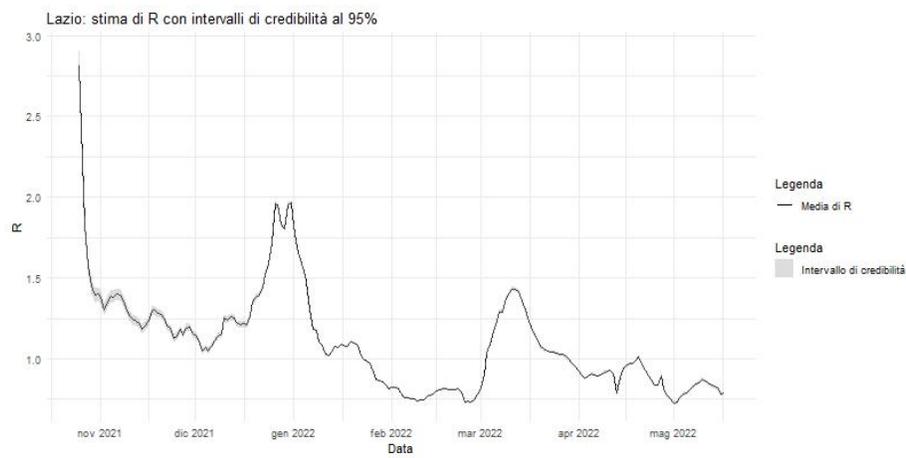


Figura 94: Stima di R_t nel Lazio, con intervallo di confidenza al 95%; quarta ondata

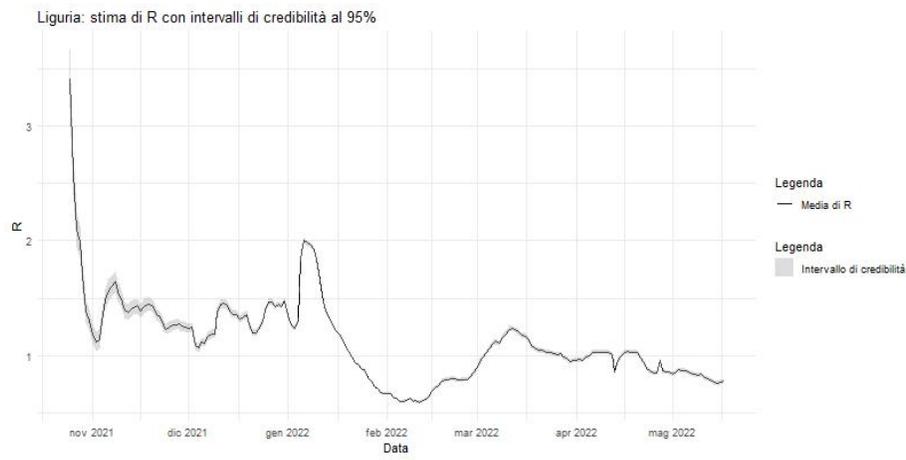


Figura 95: Stima di R_t in Liguria, con intervallo di confidenza al 95%; quarta ondata

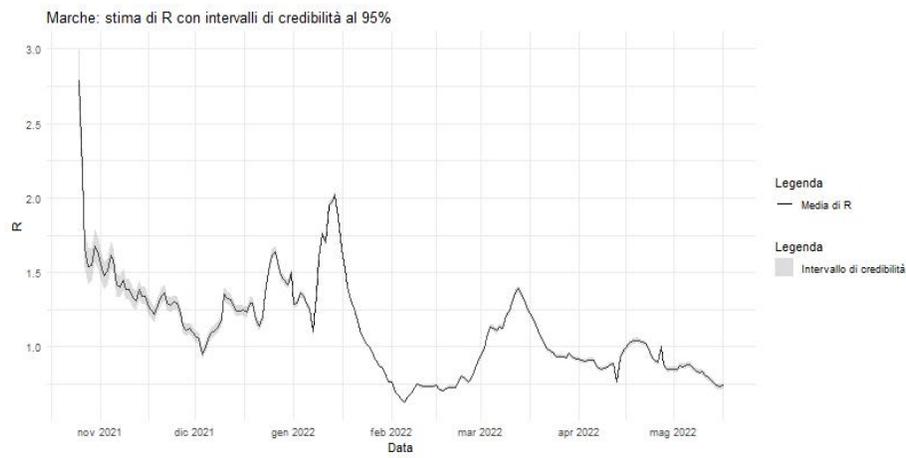


Figura 96: Stima di R_t nelle Marche, con intervallo di confidenza al 95%; quarta ondata

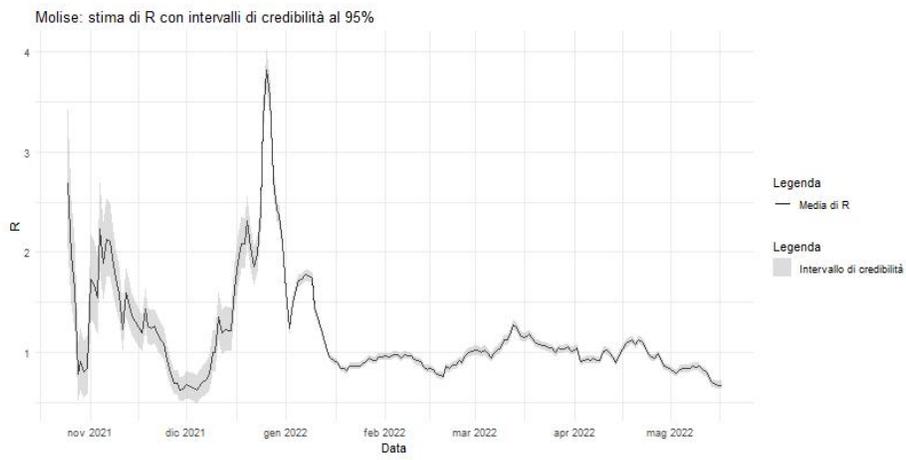


Figura 97: Stima di R_t in Molise, con intervallo di confidenza al 95%; quarta ondata

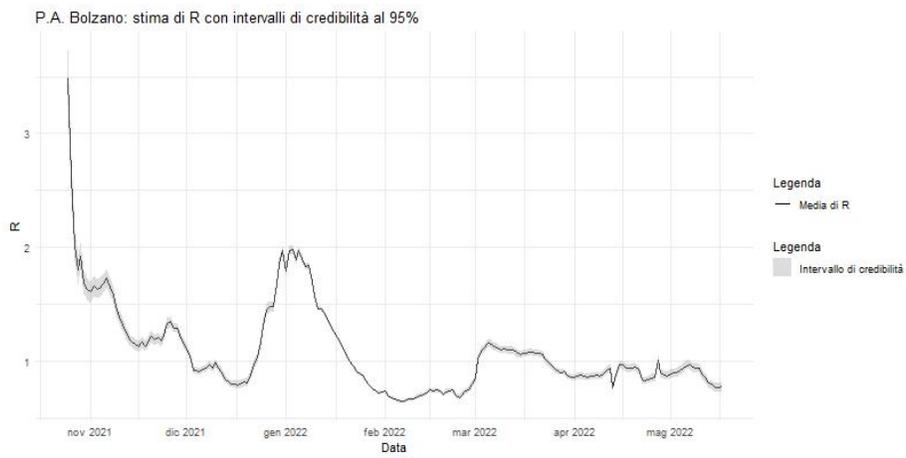


Figura 98: Stima di R_t a Bolzano, con intervallo di confidenza al 95%; quarta ondata

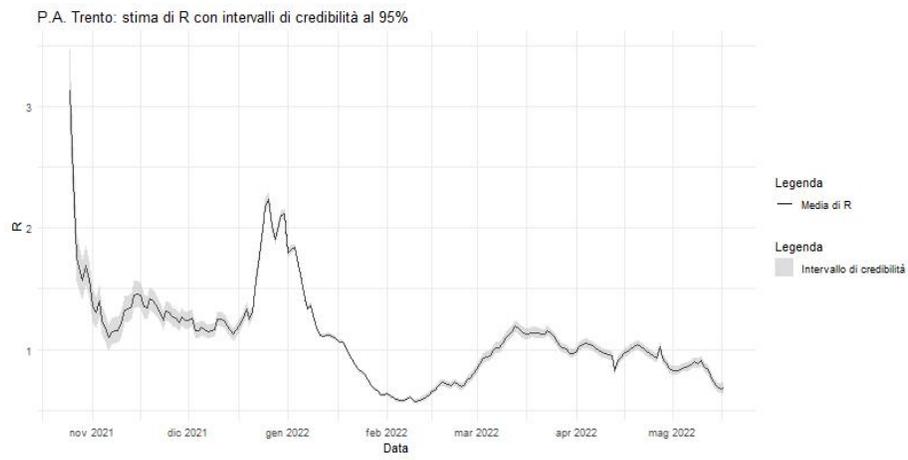


Figura 99: Stima di R_t a Trento, con intervallo di confidenza al 95%; quarta ondata

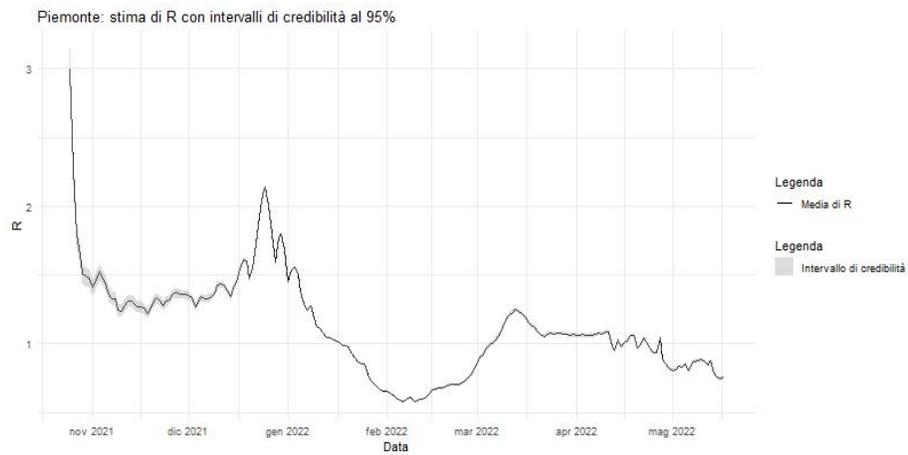


Figura 100: Stima di R_t in Piemonte, con intervallo di confidenza al 95%; quarta ondata

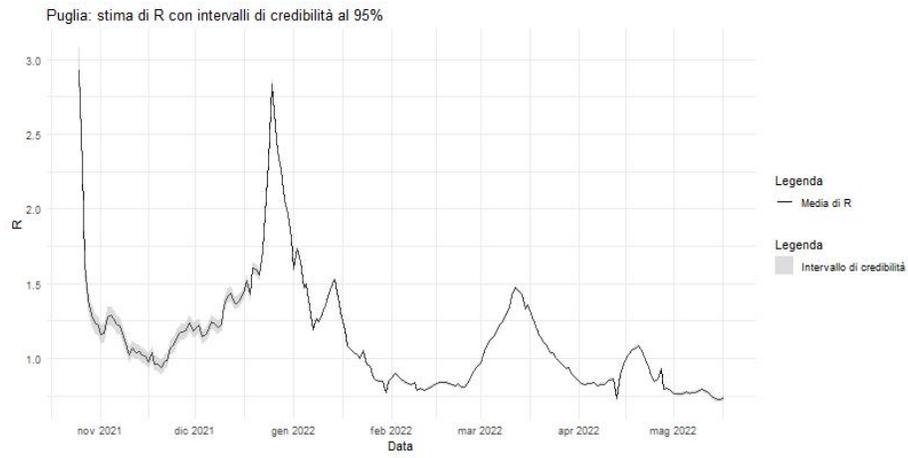


Figura 101: Stima di R_t in Puglia, con intervallo di confidenza al 95%; quarta ondata

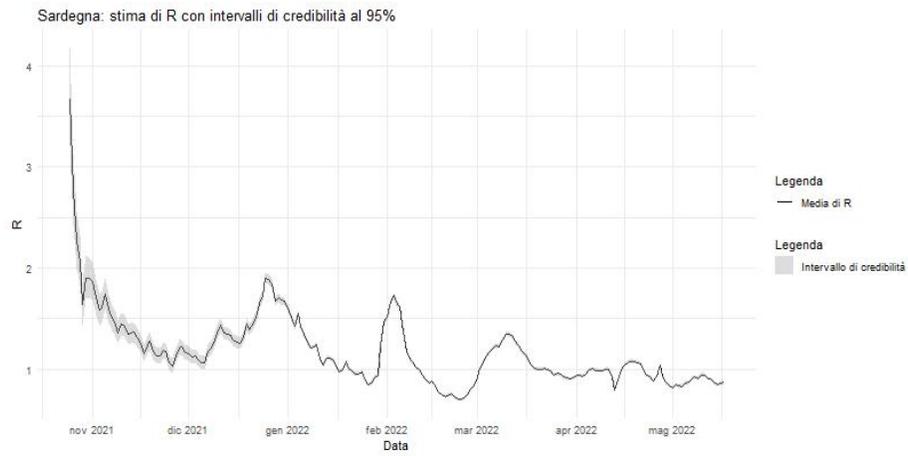


Figura 102: Stima di R_t in Sardegna, con intervallo di confidenza al 95%; quarta ondata

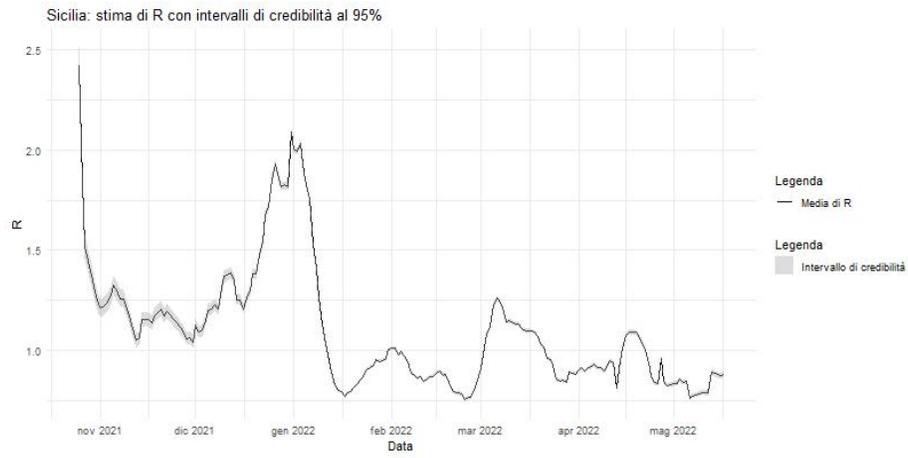


Figura 103: Stima di R_t in Sicilia, con intervallo di confidenza al 95%; quarta ondata

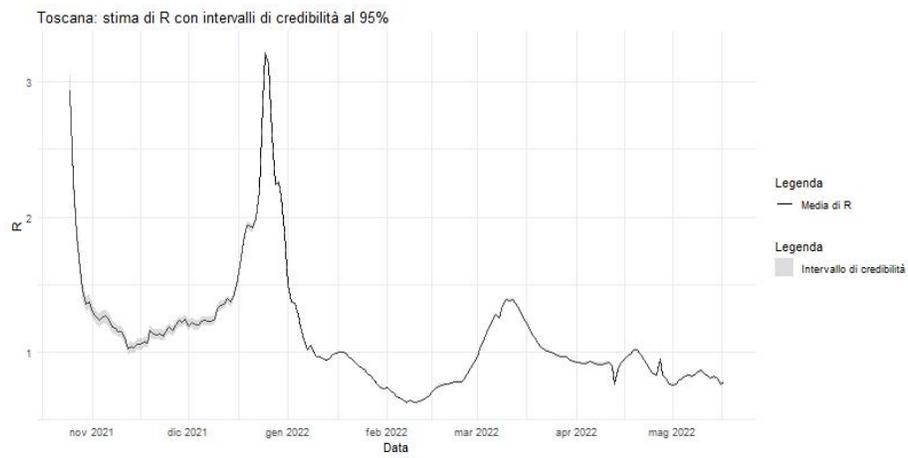


Figura 104: Stima di R_t in Toscana, con intervallo di confidenza al 95%; quarta ondata

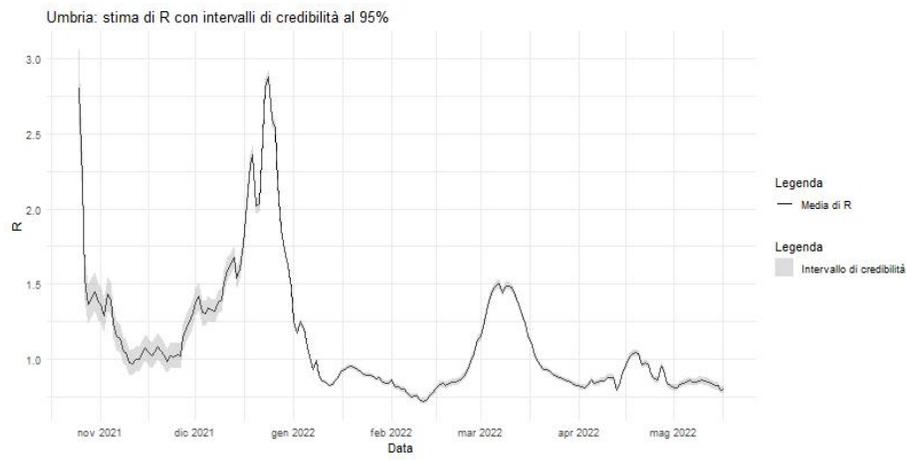


Figura 105: Stima di R_t in Umbria, con intervallo di confidenza al 95%; quarta ondata

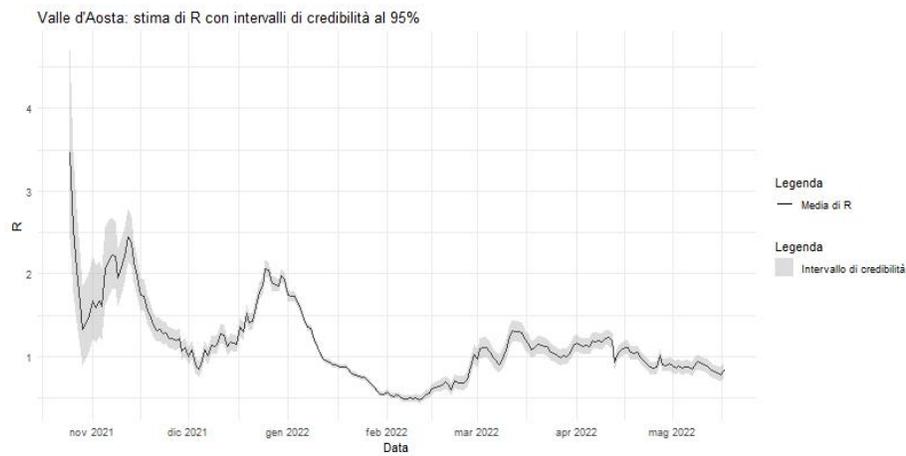


Figura 106: Stima di R_t in Valle d'Aosta, con intervallo di confidenza al 95%; quarta ondata

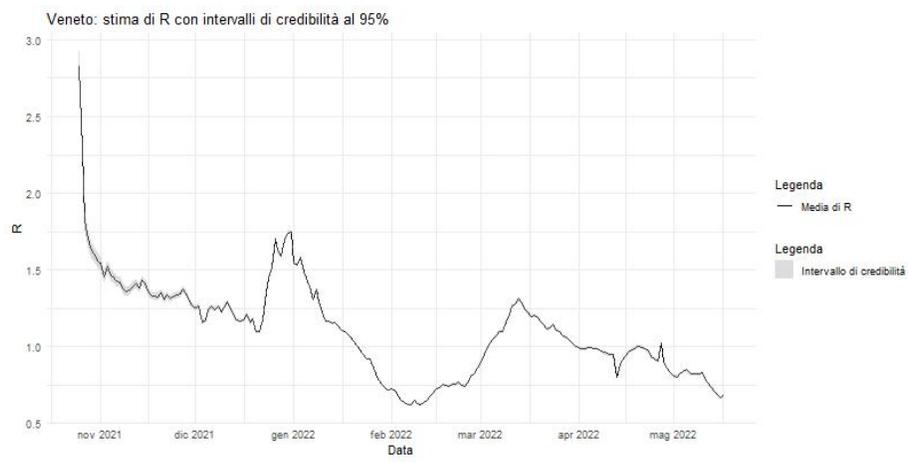


Figura 107: Stima di R_t in Veneto, con intervallo di confidenza al 95%; quarta ondata

Bibliografia

- [1] Julian Besag. “Spatial Interaction and the Statistical Analysis of Lattice Systems”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2 (Jan. 1974), pp. 192–225. ISSN: 0035-9246. DOI: [10.1111/j.2517-6161.1974.tb00999.x](https://doi.org/10.1111/j.2517-6161.1974.tb00999.x). URL: <https://doi.org/10.1111/j.2517-6161.1974.tb00999.x> (visited on 06/26/2024).
- [2] Roger S. Bivand et al. *Applied spatial data analysis with R / Roger S. Bivand, Edzer Pebesma, Virgilio Gómez-Rubio*. 2. ed. Use R! New York: Springer, 2013. ISBN: 978-1-4614-7617-7.
- [3] D. Cereda et al. *The early phase of the COVID-19 outbreak in Lombardy, Italy*. arXiv:2003.09320 [q-bio]. Mar. 2020. DOI: [10.48550/arXiv.2003.09320](https://doi.org/10.48550/arXiv.2003.09320). URL: <http://arxiv.org/abs/2003.09320> (visited on 02/11/2024).
- [4] Robert Challen et al. *Meta-analysis of the SARS-CoV-2 serial interval and the impact of parameter uncertainty on the COVID-19 reproduction number*. en. Pages: 2020.11.17.20231548. Nov. 2020. DOI: [10.1101/2020.11.17.20231548](https://doi.org/10.1101/2020.11.17.20231548). URL: <https://www.medrxiv.org/content/10.1101/2020.11.17.20231548v2> (visited on 02/11/2024).
- [5] Nalini Chintalapudi et al. “COVID-19 outbreak reproduction number estimations and forecasting in Marche, Italy”. In: *International Journal of Infectious Diseases* 96 (July 2020), pp. 327–333. ISSN: 1201-9712. DOI: [10.1016/j.ijid.2020.05.029](https://doi.org/10.1016/j.ijid.2020.05.029). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7211603/> (visited on 02/08/2024).
- [6] Monica Chiogna and Carlo Gaetan. “An interchangeable approach for modelling spatio-temporal count data”. en. In: *Environmetrics* 21.7-8 (2010). _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.1078>, pp. 849–867. ISSN: 1099-095X. DOI: [10.1002/env.1078](https://doi.org/10.1002/env.1078). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.1078> (visited on 06/27/2024).
- [7] Monica Chiogna and Carlo Gaetan. “Hierarchical space-time modelling of epidemic dynamics: an application to measles outbreaks”. en. In: *Statistical Methods and Applications* 13.1 (Apr. 2004), pp. 55–71. ISSN: 1613-981X. DOI: [10.1007/s10260-004-0085-3](https://doi.org/10.1007/s10260-004-0085-3). URL: <https://doi.org/10.1007/s10260-004-0085-3> (visited on 02/15/2024).
- [8] Monica Chiogna e Gaetan. *COVID-19/COVID-19-ITALY.pdf at master · cgaetan/COVID-19*. en. 2020. URL: <https://github.com/cgaetan/COVID-19/blob/master/COVID-19-ITALY.pdf> (visited on 06/26/2024).

- [9] Luc E. Coffeng and Sake J. Vlas. “Predicting Epidemics and the Impact of Interventions in Heterogeneous Settings: Standard SEIR Models are too Pessimistic”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 185.Supplement_1 (Nov. 2022), S28–S35. ISSN: 0964-1998. DOI: [10.1111/rssa.12854](https://doi.org/10.1111/rssa.12854). URL: <https://doi.org/10.1111/rssa.12854> (visited on 03/08/2023).
- [10] Anne Cori et al. “A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics”. In: *American Journal of Epidemiology* 178.9 (Nov. 2013), pp. 1505–1512. ISSN: 0002-9262. DOI: [10.1093/aje/kwt133](https://doi.org/10.1093/aje/kwt133). URL: <https://doi.org/10.1093/aje/kwt133> (visited on 02/19/2024).
- [11] Sourish Das and Dipak K. Dey. “On Dynamic Generalized Linear Models with Applications”. en. In: *Methodology and Computing in Applied Probability* 15.2 (June 2013), pp. 407–421. ISSN: 1573-7713. DOI: [10.1007/s11009-011-9255-6](https://doi.org/10.1007/s11009-011-9255-6). URL: <https://doi.org/10.1007/s11009-011-9255-6> (visited on 06/03/2024).
- [12] Yi Grace Y. *Characterizing the COVID-19 dynamics with a new epidemic model: Susceptible-exposed-asymptomatic-symptomatic-active-removed - Yi - 2022 - Canadian Journal of Statistics - Wiley Online Library*. Apr. 2022. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/cjs.11698> (visited on 03/08/2023).
- [13] Aline Foerster Grande, Guilherme Pumi, and Gabriela Bettella Cybis. *Bayesian Analysis of Beta Autoregressive Moving Average Models*. arXiv:2307.07042 [stat]. July 2023. URL: <http://arxiv.org/abs/2307.07042> (visited on 02/08/2024).
- [14] Veronika Grimm, Friederike Mengel, and Martin Schmidt. “Extensions of the SEIR model for the analysis of tailored social distancing and tracing approaches to cope with COVID-19”. en. In: *Scientific Reports* 11.1 (Feb. 2021), p. 4214. ISSN: 2045-2322. DOI: [10.1038/s41598-021-83540-2](https://doi.org/10.1038/s41598-021-83540-2). URL: <https://www.nature.com/articles/s41598-021-83540-2> (visited on 03/08/2023).
- [15] Jouni Helske. “KFAS: Exponential Family State Space Models in R”. en. In: *Journal of Statistical Software* 78 (June 2017), pp. 1–39. ISSN: 1548-7660. DOI: [10.18637/jss.v078.i10](https://doi.org/10.18637/jss.v078.i10). URL: <https://doi.org/10.18637/jss.v078.i10> (visited on 06/25/2024).
- [16] Nicholas P. Jewell and Joseph A. Lewnard. “On the use of the reproduction number for SARS-CoV-2: Estimation, misinterpretations and relationships with other ecological measures”. In: *Journal of the Royal Statistical Society. Series A, (Statistics in Society)* (June 2022), 10.1111/rssa.12860. ISSN: 0964-1998. DOI: [10.1111/rssa.12860](https://doi.org/10.1111/rssa.12860). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9350332/> (visited on 03/08/2023).
- [17] Brian King and Daniel R. Kowal. “Warped Dynamic Linear Models for Time Series of Counts”. In: *Bayesian Analysis* -1.-1 (Jan. 2023). Publisher: International Society for Bayesian Analysis, pp. 1–26. ISSN: 1936-0975, 1931-6690. DOI: [10.1214/23-BA1394](https://doi.org/10.1214/23-BA1394). URL: <https://projecteuclid.org/journals/bayesian-analysis/advance-publication/Warped-Dynamic->

- [Linear-Models-for-Time-Series-of-Counts/10.1214/23-BA1394.full](#) (visited on 02/12/2024).
- [18] Daniel Adyro Martínez-Bello, Antonio López-Quílez, and Alexander Torres-Prieto. “Bayesian dynamic modeling of time series of dengue disease case counts”. en. In: *PLOS Neglected Tropical Diseases* 11.7 (2017). Publisher: Public Library of Science, e0005696. ISSN: 1935-2735. DOI: [10.1371/journal.pntd.0005696](https://doi.org/10.1371/journal.pntd.0005696). URL: <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0005696> (visited on 02/12/2024).
- [19] Abolfazl Mollalo, Kiara M. Rivera, and Behzad Vahedi. “Artificial Neural Network Modeling of Novel Coronavirus (COVID-19) Incidence Rates across the Continental United States”. en. In: *International Journal of Environmental Research and Public Health* 17.12 (Jan. 2020). Number: 12. Publisher: Multidisciplinary Digital Publishing Institute, p. 4204. ISSN: 1660-4601. DOI: [10.3390/ijerph17124204](https://doi.org/10.3390/ijerph17124204). URL: <https://www.mdpi.com/1660-4601/17/12/4204> (visited on 03/08/2023).
- [20] Lorenzo Pellis et al. “Estimation of Reproduction Numbers in Real Time: Conceptual and Statistical Challenges”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 185.Supplement_1 (Nov. 2022), S112–S130. ISSN: 0964-1998. DOI: [10.1111/rssa.12955](https://doi.org/10.1111/rssa.12955). URL: <https://doi.org/10.1111/rssa.12955> (visited on 03/08/2023).
- [21] Richard F. Potthoff and Maurice Whittinghill. “Testing for Homogeneity: II. The Poisson Distribution”. In: *Biometrika* 53.1/2 (1966). Publisher: [Oxford University Press, Biometrika Trust], pp. 183–190. ISSN: 0006-3444. DOI: [10.2307/2334063](https://doi.org/10.2307/2334063). URL: <https://www.jstor.org/stable/2334063> (visited on 06/26/2024).
- [22] Håvard Rue et al. “Bayesian Computing with INLA: A Review”. en. In: *Annual Review of Statistics and Its Application* 4.Volume 4, 2017 (Mar. 2017). Publisher: Annual Reviews, pp. 395–421. ISSN: 2326-8298, 2326-831X. DOI: [10.1146/annurev-statistics-060116-054045](https://doi.org/10.1146/annurev-statistics-060116-054045). URL: <https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-060116-054045> (visited on 06/27/2024).
- [23] Robert R. Sokal, Neal L. Oden, and Barbara A. Thomson. “Local spatial autocorrelation in biological variables”. In: *Biological Journal of the Linnean Society* 65.1 (Sept. 1998), pp. 41–62. ISSN: 0024-4066. DOI: [10.1111/j.1095-8312.1998.tb00350.x](https://doi.org/10.1111/j.1095-8312.1998.tb00350.x). URL: <https://doi.org/10.1111/j.1095-8312.1998.tb00350.x> (visited on 06/26/2024).
- [24] J Wallinga and M Lipsitch. “How generation intervals shape the relationship between growth rates and reproductive numbers”. In: *Proceedings of the Royal Society B: Biological Sciences* 274.1609 (Nov. 2006). Publisher: Royal Society, pp. 599–604. DOI: [10.1098/rspb.2006.3754](https://doi.org/10.1098/rspb.2006.3754). URL: <https://royalsocietypublishing.org/doi/full/10.1098/rspb.2006.3754> (visited on 02/18/2024).
- [25] Yuchen Xie. *Analysis of the BFGS Method with Errors*. en. Jan. 2019. DOI: [10.1137/19M1240794](https://doi.org/10.1137/19M1240794). URL: <https://epubs.siam.org/doi/epdf/10.1137/19M1240794> (visited on 06/19/2024).

- [26] Changhe Yuan and Marek J. Druzdzel. “Importance sampling algorithms for Bayesian networks: Principles and performance”. In: *Mathematical and Computer Modelling. Optimization and Control for Military Applications* 43.9 (May 2006), pp. 1189–1207. ISSN: 0895-7177. DOI: [10.1016/j.mcm.2005.05.020](https://doi.org/10.1016/j.mcm.2005.05.020). URL: <https://www.sciencedirect.com/science/article/pii/S0895717705005443> (visited on 06/03/2024).