

ALMA MATER STUDIORUM
UNIVERSITÁ DI BOLOGNA - CAMPUS DI CESENA
DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA
ELETTRICA E DELL'INFORMAZIONE
"GUGLIELMO MARCONI"

Corso di Laurea Magistrale in Ingegneria Biomedica

**MODELLO EPIDEMIOLOGICO PER
L'EPIDEMIA COVID-19
VALIDATO SULLA POPOLAZIONE
DELL' EMILIA ROMAGNA**

Tesi in

Algebra e Analisi Numerica LM

Presentata da:
CHIARA CAVOTTA

Relatore:
Chiar.ma Prof.ssa
SERENA MORIGI

Correlatore:
Chiar.mo Prof.
STEFANO SEVERI

Anno Accademico 2019/2020

**”La differenza tra un sogno e un
obiettivo è semplicemente la data”
Walt Disney**

Indice

Indice	3
Introduzione	4
1 Epidemia Covid-19 e modelli epidemiologici	9
1.1 Quadro storico Covid-19	9
1.2 Eventi nel trimestre Settembre-Novembre 2020	10
1.2.1 Situazione in Emilia Romagna	13
1.3 Modelli epidemiologici per il Covid-19	15
2 Modelli matematici in epidemiologia	29
2.1 Problema diretto	32
2.1.1 Problema di Cauchy o IVP	32
2.1.2 Risolubilità	34
2.1.3 Condizionamento	35
2.1.4 Metodi numerici per la soluzione di IVP	38
2.1.5 Metodi one step espliciti: metodi Runge-Kutta	40
2.2 Problema inverso	46

2.2.1	Problema di minimi quadrati non lineari	47
2.2.2	Risoluzione dei problemi di minimi quadrati non lineari	48
3	Casi di studio	55
3.1	Modello SIR	58
3.1.1	Implementazione modello SIR	60
3.2	Modello SIDARTHE	64
3.2.1	Implementazione modello SIDARTHE	68
3.2.2	Suddivisione del periodo Settembre - Novembre 2020	74
3.2.3	Previsione relativa al mese di Dicembre 2020	81
	Conclusioni	84
	Bibliografia	87

Introduzione

La malattia respiratoria acuta COVID-19, causata dal nuovo coronavirus SARS-CoV2, rappresenta un evento di portata storica che inizia a segnare e condizionare profondamente la vita di tutti gli individui dall'inizio dell'anno 2020, e provoca una emergenza sanitaria cui é data risposta immediata con una serie di misure urgenti fin dalla dichiarazione dello stato di emergenza del 31 gennaio 2020.

La necessità di determinare un contenimento degli effetti epidemiologici, di fornire strumenti utili per affrontare i vari problemi sanitari, sociali ed economici, chiama in causa, a vario titolo, l'intera comunità scientifica. Contestualmente, lo studio e l'applicazione di modelli epidemiologici danno vita ad una grande proliferazione di letteratura scientifica, che si va sempre aggiornando al fine di controllare l'evoluzione della pandemia, partendo dai dati raccolti, e di prevedere gli sviluppi a breve-medio termine.

Gli aggiornamenti sono motivati in ragione di nuove evidenze scientifiche, di modifiche nelle dinamiche epidemiche o di elementi sopravvenuti rilevanti per la strategia di contrasto all'epidemia da SARS-CoV-2.

Con lo scopo di ridurre la diffusione del virus, sono state introdotte misure di contenimento da parte del Governo italiano che influiscono sull'andamento dei dati epidemiologici. Proprio questo aspetto ha richiesto un adattamento dei modelli matematici epidemiologici e del processo di calibrazione e quindi di identificazione dei parametri.

L'obiettivo di questo elaborato é quello di studiare ed implementare i modelli matematici epidemiologici SIR (Suscettibili - Infetti - Rimossi) e SIDARTHE (Suscettibili - Infetti asintomatici non rilevati - Infetti asintomatici rilevati - Infetti

sintomatici non rilevati - Infetti sintomatici rilevati - Terapia Intensiva - Guariti - Deceduti) per controllare l'andamento dell'epidemia COVID-19 in Italia, utilizzando i dati messi a disposizione dal Dipartimento della Protezione Civile Italiano, relativi alla regione Emilia Romagna nell'arco temporale che va dal 1 settembre 2020 al 29 novembre 2020.

La tesi si sviluppa secondo la seguente struttura.

Il primo capitolo descrive a caratteri generali la situazione epidemiologica relativa all'epidemia COVID-19 e analizza quelle che sono state le misure di contenimento applicate dal Governo italiano alla penisola durante il trimestre sopra citato, soffermandosi principalmente sulle restrizioni imposte in Emilia Romagna.

Inoltre vengono descritti in dettaglio alcuni modelli epidemiologici implementati dall'inizio dell'epidemia, utilizzati per monitorare la curva dei contagi e per prevederne l'evoluzione.

Si prosegue con il secondo capitolo che descrive da un punto di vista matematico i modelli epidemiologici utilizzati, partendo quindi dall'analisi del problema diretto attraverso il quale risolvere il sistema di equazioni differenziali ordinarie ai valori iniziali (IVP), fino ad arrivare all'analisi del problema inverso, visto come problema di minimi quadrati non lineari, il cui scopo é l'identificazione dei parametri del modello epidemiologico in questione. In dettaglio, il problema inverso da affrontare, essendo la maggiore difficoltà di questo elaborato, consiste nel partire dai dati sperimentali prelevati dal repository della Protezione Civile italiana e successivamente nel calcolare l'insieme dei parametri che caratterizzano il modello differenziale generico dell'epidemia per una specifica epidemia, appunto il COVID-19.

Infine il terzo ed ultimo capitolo descrive i casi di studio considerati, quindi inizialmente il classico modello semplice a compartimenti di Kermack-McKendrick (SIR), e poi a seguire il modello complesso SIDARTHE, entrambi relativi al trimestre settembre - novembre 2020.

La validazione di entrambi i modelli mette in luce interessanti aspetti legati all'evoluzione della pandemia in corso. L'elaborato si conclude con la validazione del modello SIDARTHE relativo ad un periodo contiguo a quello studiato, ai fini di una previsione dell'andamento epidemico.

Capitolo 1

Epidemia Covid-19 e modelli epidemiologici

1.1 Quadro storico Covid-19

Tutto parte alla fine del mese di dicembre 2019, quando la Cina informa l'Organizzazione Mondiale della Sanità (OMS) dell'emergenza presente a Wuhan per una misteriosa polmonite.

Il resto del mondo osserva da lontano ma il rischio viene sottovalutato; inizia quindi la diffusione di un nuovo coronavirus, il 2019-nCoV, successivamente l'OMS lo rinomina SARS-CoV-2. A febbraio 2020, l'OMS definisce la malattia associata a SARS-CoV-2 come "COroNaVIrus Disease 19 (COVID-19)".

In pochi mesi, la malattia acuta respiratoria provocata dall'infezione di tale virus si diffonde in tutto il mondo. Infatti, quello che si ritiene un fenomeno isolato e lontano, già a fine febbraio arriva in Italia, e ci si accorge che il nuovo coronavirus fa molto più male di quanto si potesse immaginare. L'11 marzo 2020 l'OMS la dichiara pandemia.

L'Italia è uno dei primi paesi europei ad essere colpito, diventa quindi il nuovo epicentro del virus che, come uno sciame, si sposta da una regione all'altra, sconvolgendo la vita delle persone e trasformando in anormale ciò che fino a quel momento era ritenuto normale.

Il virus SARS-CoV-2, era per lo piú sconosciuto, tuttavia in pochissimo tempo aveva fatto due grandi conquiste: era passato dagli animali all'uomo ed era passato da uomo a uomo, inoltre si diffonde principalmente in Lombardia e quando vengono diagnosticati i primi casi la diffusione é già elevata.

Aumentano i pazienti bisognosi di cure ospedaliere e gli ospedali sono sotto pressione. La situazione costringe quindi il Governo Italiano ad imporre alcune norme per il contenimento della diffusione. Viene introdotto un lockdown inizialmente nei comuni e zone dove si riscontrano i primi focolai e poi progressivamente viene esteso in tutto il paese italiano a partire dal giorno 9 Marzo 2020. Questo provoca un sostanziale rallentamento dei casi d'infezione, dopo il raggiungimento del picco intorno al 19 aprile 2020 [11]. In seguito le misure restrittive vengono allentate ma la diffusione del virus, seppur sotto controllo e monitorata, non si spegne mai completamente durante i mesi successivi.

Il 29 luglio 2020 é il giorno migliore da quando é esplosa l'emergenza. I ricoverati in terapia intensiva sono "solo" 38 in tutta la penisola italiana.

Il 15 agosto Wuhan, primo focolaio di questa pandemia, é ormai fuori dal tunnel. A testimoniarlo arrivano le immagini di un ferragosto all'insegna della normalità: le immagini di centinaia di persone che ballano in piscina fanno velocemente il giro del mondo.

Parallelamente in Italia la curva dei contagi inizia a risalire e tornano dunque alcune misure di contenimento. Discoteche chiuse in tutta la penisola e mascherine obbligatorie dalle ore 18 alle 6 del mattino nei luoghi della movida e nei locali all'aperto. Si arriva al 29 agosto 2020 che, con "solo" 1 morto, é il miglior giorno per l'Italia dal punto di vista dei deceduti, dall'inizio della pandemia.

1.2 Eventi nel trimestre Settembre-Novembre 2020

Di seguito si presenta una cronistoria degli eventi successi in Italia durante il periodo che va da settembre 2020 a novembre 2020, descrivendo eventuali misure di contenimento messe in atto dal Governo Italiano.

- **Settembre 2020:**

Durante il mese di settembre, mentre il virus in Europa prende il sopravvento, in Italia il contagio sembra essere sotto controllo. I casi crescono, ma la curva italiana é molto piú stabile rispetto a quella di altri paesi.

- **Ottobre 2020:**

Ad ottobre, invece, l'Italia entra in piena "seconda ondata". Infatti il numero degli infetti inizia nuovamente a crescere in modo esponenziale, senza dare segni di rallentamento e la pressione sugli ospedali torna a farsi sentire. Il giorno peggiore é l'ultimo del mese, il 31 ottobre, con nuovi 31.756 positivi in tutta la penisola. Per questo tra la metà di ottobre ed i primi di novembre vengono introdotte altre restrizioni da parte del Governo per contrastare la nuova crescita dei contagi.

Il 19 ottobre arriva un nuovo DPCM il quale preannuncia nuove misure restrittive: nel provvedimento del Governo c'è la possibilità di poter chiudere le piazze e le vie delle città nelle quali si riscontra un accentuarsi di assembramenti incontrollati. I locali di ristorazione possono servire al tavolo dalle ore 5 alle 24 e fino alle ore 18 se non sono al tavolo. Non sono previsti limiti al servizio a domicilio né per le attività in ospedali, aeroporti e autogrill. Il decreto prevede altresí che da mercoledì 21 ottobre, ci sia l'ingresso alle 9 per le scuole superiori. Inoltre hanno una settimana di tempo piscine e palestre per adeguarsi alle norme, altrimenti dovranno chiudere. Invece é previsto il fermo dello sport amatoriale di squadra e per quanto riguarda le fiere si svolgeranno solo quelle a livello nazionale.

Il 24 ottobre, il presidente del Consiglio, Giuseppe Conte, firma il nuovo DPCM in cui vengono chiuse palestre e piscine e per le scuole superiori viene incrementato il ricorso alla didattica a distanza.

1.2.1 Situazione in Emilia Romagna

In Europa il primo caso di COVID-19 é stato accertato a gennaio 2020. Le regioni settentrionali dell'Italia, in particolare Lombardia, Veneto ed Emilia Romagna, sono state tra le aree piú colpite nella prima fase dell'epidemia.

Con l'obiettivo di registrare tempestivamente l'andamento dei decessi già durante le prime settimane dell'emergenza sanitaria, in Emilia Romagna é stato implementato un sistema di monitoraggio della mortalità basato sui dati delle anagrafi e del sistema di notifiche che ha rivelato, a partire dal mese di marzo 2020, un eccesso di mortalità rispetto all'atteso, concentrato maggiormente nelle aree nord occidentali dell'Emilia-Romagna.

La regione emiliana-romagnola si lascia alle spalle un primo semestre 2020 con un eccesso di mortalità che ha colpito in maniera differenziale i due sessi e le fasce di età. Si sono registrati 30.330 decessi a livello regionale. I decessi correlati all'epidemia COVID-19 corrispondono al 14% dei decessi totali e si sono registrati con maggiore frequenza tra le donne di età ≥ 85 anni e tra gli uomini di età compresa tra i 75 e gli 84 anni. La maggiore concentrazione di decessi correlati all'epidemia COVID-19 tra le donne piú anziane può essere dovuta sia alla struttura della popolazione, e quindi a un contingente piú cospicuo di soggetti fragili ultra ottantacinquenni tra le donne, sia alle dinamiche di diffusione dell'infezione.

Successivamente, superato il primo semestre 2020, gli eventi piú significativi in Emilia Romagna sono stati:

– 25 ottobre 2020 → seconda ondata:

La seconda ondata della pandemia si fa sentire forte anche in Emilia Romagna. Come in tutta Italia, dove si registrano 21.273 nuovi positivi e 128 decessi, nella regione emiliana-romagnola si sono registrati 1.192 contagi in piú rispetto al giorno precedente, su un totale di 9.644 tamponi eseguiti nelle ultime 24 ore. Dei nuovi positivi, sono 636 gli asintomatici individuati nell'ambito delle attività di contact tracing e screening re-

gionali. Prosegue l'attività di controllo e prevenzione: complessivamente, tra i nuovi positivi 170 persone erano già in isolamento al momento dell'esecuzione del tampone e 338 sono state individuate nell'ambito di focolai già noti.

– 6 novembre 2020 → zona gialla:

L'Emilia Romagna è inserita in zona gialla, quindi non in uno degli scenari di più elevata gravità: per questo, qui si applicano le misure restrittive basilari previste per tutto il resto del territorio nazionale.

– 15 novembre 2020 → zona arancione:

L'Emilia Romagna passa dunque da zona gialla, a rischio moderato, a zona arancione, a rischio elevato, nella quale si applicano misure più restrittive a tutela della salute pubblica e per frenare il contagio da Coronavirus.

Questo comporta spostamenti consentiti solo all'interno del proprio comune di residenza, mentre sono vietati quelli verso altri comuni e quelli in entrata e uscita dalla regione, salvo che per comprovate esigenze lavorative, di salute, studio o necessità. Si raccomanda inoltre di evitare gli spostamenti non necessari anche all'interno del proprio comune. Chiusi i ristoranti e le altre attività di ristorazione (bar, pub, gelaterie, pasticcerie): per loro resta consentita la sola vendita da asporto, dalle ore 5 alle ore 22, mentre le consegne a domicilio non hanno limiti d'orario.

Tutte queste sono dunque le nuove misure, in aggiunta a quelle già in vigore, scattate in Emilia Romagna da domenica 15 novembre 2020 e per almeno due settimane dopo l'ordinanza firmata dal ministro della Salute, Roberto Speranza, che prevede l'inserimento della regione in zona arancione, sulla base della valutazione dei parametri relativi al livello di rischio e allo scenario epidemico [6].

1.3 Modelli epidemiologici per il Covid-19

L'emergenza sanitaria da Covid-19 ha portato dunque molti ricercatori ad approfondire lo studio dell'epidemia anche dal punto di vista della modellazione matematica del fenomeno, con lo scopo di aiutare medici, infettivologi e virologi a conoscere meglio la portata di questa calamità, a presentarne possibili scenari futuri e a pianificare strategie di controllo efficaci.

La diffusione della malattia é generalmente formulata sotto forma di sistemi di equazioni differenziali ordinarie ai valori iniziali (ODE - IVP):

$$Y'(t) = f(Y, t) \quad (1.1)$$

con condizioni iniziali:

$$Y(0) = Y_0 \quad (1.2)$$

dove f rappresenta il modello differenziale.

Modelli semplici possono riflettere bene le dinamiche di diverse epidemie. Nell'ultimo periodo, questi modelli sono stati utilizzati per modellare l'evoluzione pandemica dovuta al COVID-19.

Il modello SIS (Suscettibili - Infetti - Suscettibili), utilizzabile per malattie di breve durata che non danno immunità, distingue solo due gruppi di individui: S (suscettibili) e I (infetti). Questo modello può essere utilizzato per le malattie che hanno le seguenti proprietà:

- dopo che si é guariti dalla malattia, ogni individuo ritorna immediatamente al gruppo sano e può essere nuovamente infettato;
- gli infetti sono immediatamente contagiosi;
- le persone sane si ammalano con un tasso di infezione lineare β ;

- le persone infette guariscono con un tasso di guarigione lineare α ;
- ogni gruppo interagisce con la stessa probabilità, ciò giustifica l'assunzione di relazioni lineari.

La diffusione della malattia é generalmente formulata sotto forma di equazioni differenziali ordinarie ai valori iniziali (IVP):

$$\begin{cases} S'(t) = -\beta IS + \alpha I, \\ I'(t) = \beta IS - \alpha I. \end{cases} \quad (1.3)$$

Come modelli semplici, oltre al modello SIS, vi sono anche il modello SIR (Suscettibili - Infetti - Rimossi), che verrà trattato nel capitolo 3, e il modello SEIR (Suscettibili - Esposti - Infetti - Rimossi). A differenza degli altri due, il modello SEIR considera un significativo periodo di incubazione durante il quale gli individui sono stati infettati ma non sono ancora infettivi. Durante questo periodo l'individuo si trova nel compartimento E che descrive la percentuale "esposta" di popolazione.

Supponendo che il periodo di incubazione sia una variabile casuale con distribuzione esponenziale con parametro a (cioé il periodo medio di incubazione é a^{-1}) con N costante, si ha il modello SEIR:

$$\begin{cases} S'(t) = -\beta \frac{SI}{N}, \\ E'(t) = \beta \frac{SI}{N} - aE, \\ I'(t) = aE - \gamma I, \\ R'(t) = \gamma I. \end{cases} \quad (1.4)$$

Casella et al. [1] ha ampliato il modello SEIR sviluppando un modello orientato al controllo che sottolinea gli effetti dei ritardi e confronta i risultati di diverse politiche di contenimento.

Il punto di partenza é dunque il modello SEIR con l'aggiunta di un ulteriore compartimento L :

$$\begin{cases} S'(t) = -\beta \frac{SI}{N}, \\ E'(t) = \beta \frac{SI}{N} - aE, \\ I'(t) = aE - \gamma I, \\ L'(t) = \gamma I - \delta L, \\ R'(t) = \delta L. \end{cases} \quad (1.5)$$

dove L é il numero di soggetti che sono ancora malati, ma non sono piú infettivi a causa del ricovero in ospedale, della quarantena o semplicemente perché i soggetti infetti sono per lo piú infettivi durante i primi giorni dopo la fine del periodo di latenza. [1]

Liu et al. [8] invece sviluppa un modello di equazioni differenziali dell' epidemia COVID-19 relativo alla Corea del Sud, Italia e Spagna. Il suo obiettivo é prevedere nel tempo il numero futuro di casi partendo dai primi dati di casi segnalati in regioni di tutto il mondo.

Il modello incorpora i seguenti elementi importanti:

- il numero di individui infettivi asintomatici (con sintomi molto lievi o assenti);
- il numero di individui infettivi segnalati sintomatici (con sintomi gravi);
- il numero di individui infettivi sintomatici non segnalati (con sintomi meno gravi).

Le transizioni tra le fasi sono generalmente difficili da determinare, ma possono essere stimate dai dati dei casi segnalati, col passare del tempo. Il modello di Liu et al. é costituito dalle seguenti equazioni differenziali e condizioni iniziali:

$$\left\{ \begin{array}{ll} S'(t) = -\tau(t)S(t)[I(t) + U(t)], & S(t_0) = S_0, \\ E'(t) = \tau(t)S(t)[I(t) + U(t)] - \alpha E(t), & E(t_0) = E_0, \\ I'(t) = \alpha E(t) - \nu I(t), & I(t_0) = I_0, \\ R'(t) = \nu_1 I(t) - \eta R(t), & R(t_0) = R_0, \\ U'(t) = \nu_2 I(t) - \eta U(t), & U(t_0) = U_0, \end{array} \right. \quad (1.6)$$

Si ha che $t \geq t_0$ é il tempo in giorni, t_0 é la data di inizio dell'epidemia, $S(t)$ é il numero di individui suscettibili all'infezione al tempo t , $E(t)$ é il numero di non infettivi asintomatici (esposti o infetti latenti) individui al tempo t , $I(t)$ é il numero di individui asintomatici ma infettivi al tempo t , $R(t)$ é il numero di individui infettivi sintomatici segnalati al tempo t e $U(t)$ é il numero di individui infettivi sintomatici non segnalati al tempo t . [8]

Il parametro della velocità di trasmissione dipendente dal tempo é $\tau(t)$. Individui asintomatici non infettivi recentemente infettati $E(t)$ sono incubati in un periodo medio di $1/\alpha$ giorni. Gli individui infettivi asintomatici $I(t)$ sono infettivi per un periodo medio di $1/\nu$ giorni. Gli individui infettivi sintomatici segnalati $R(t)$ sono infettivi per un periodo medio di $1/\eta$ giorni, cosí come gli individui infettivi sintomatici non segnalati $U(t)$.

Si presuppone che gli individui infettivi sintomatici segnalati $R(t)$ siano segnalati e isolati immediatamente e non causino ulteriori infezioni. Gli individui asintomatici $I(t)$ possono anche essere visti come aventi uno stato sintomatico di basso livello. Tutte le infezioni vengono acquisite da individui infettivi $I(t)$ o $U(t)$. La frazione f di infezione asintomatica diventa infezione sintomatica segnalata e la frazione $(1-f)$ diventa infezione sintomatica non segnalata.

La velocità con cui le infezioni asintomatiche diventano sintomatiche segnalate é $\nu_1 = f\nu$, la velocità con cui le infezioni asintomatiche diventano sintomatiche non segnalate é $\nu_2 = (1 - f)\nu$, dove $\nu_1 + \nu_2 = \nu$. [8]

Anche *Magri et al.* [9] definisce un modello epidemiologico per l'epidemia Covid-19; la popolazione di un paese viene divisa in gruppi che si escludono a vicenda: suscettibile (S), infetto (I), guarito (R) e deceduto (D). Si presume che ogni gruppo abbia le stesse caratteristiche, cioè che i gruppi siano omogenei. Ogni persona suscettibile può contrarre il virus (il gruppo immunitario viene trascurato).

Matematicamente si ha,

$$N = S + I + R + D.$$

L'equazione precedente è un'equazione di continuità. La popolazione N , che non varia nel tempo, è la somma dei gruppi S , I , R , D , che variano nel tempo. Questo approccio compartimentale è noto come modello SIRD epidemico con dinamiche vitali e popolazione costante. Le ipotesi di lavoro e i primi principi sono espressi matematicamente da quattro equazioni differenziali ordinarie (IVP) con parametri variabili nel tempo (sistema dinamico non autonomo). [9]

$$\begin{cases} S'(t) = -\beta \frac{I}{N} S, \\ I'(t) = \beta \frac{I}{N} S - (\mu + \gamma) I, \\ R'(t) = \gamma I, \\ D'(t) = \mu I. \end{cases} \quad (1.7)$$

Il sistema (1.7) è soggetto alle condizioni iniziali

$$\begin{cases} S(0) = S_0, \\ I(0) = I_0, \\ R(0) = R_0, \\ D(0) = D_0. \end{cases} \quad (1.8)$$

Il rapporto $\frac{I}{N}$ è la probabilità di entrare in contatto con un individuo infetto.

- β è il numero medio di contatti per persona per unità di tempo pesato dalla trasmissione del virus (tasso di contatto);

- γ é il numero medio di persone recuperate per unità di tempo (tasso di recupero);
- μ é il numero medio di decessi dovuti a COVID-19 per unità di tempo (tasso di mortalità).

Questi parametri dipendono dal tempo e dipendono da diverse variabili, come le politiche governative (blocco, chiusure di scuole / università, allontanamento sociale, ecc.), l'eterogeneità della popolazione (età, stile di vita, immunità della mandria, standard igienici, ecc.), e proprietà dell'epidemia (genoma del virus, meccanismi di diffusione, ecc.).

I parametri SIRD stimano le scale temporali dell'epidemia:

- $\frac{1}{\gamma}$ é il tempo medio di recupero;
- $\frac{1}{\beta}$ é il tempo medio tra un contatto (con un infetto) e l'altro;
- $\frac{1}{\mu}$ é il tempo medio alla morte (per coloro che non guariscono).

Il numero totale di nuovi casi per unità di tempo dovuti al contatto di persone suscettibili S con persone infette é $\beta I/NS$. Questo é l'unico termine non lineare delle equazioni. (Altre non linearità sono nascoste nella dipendenza dal tempo dei parametri β , γ e μ .)

Le equazioni del sistema (1.7) vengono interpretate come segue.

La prima equazione é il tasso di cambiamento del gruppo suscettibile. Il numero di suscettibili, S , cambia piú velocemente nel tempo se ci sono piú persone infette, I , e di conseguenza ci saranno piú suscettibili S che possono essere infettati. Chiaramente, il gruppo suscettibile é costante nel tempo se il tasso di contatto del virus é zero, e / o se il numero di infetti é zero e / o se il numero di suscettibili é zero.

La seconda equazione é la derivata temporale dell'equazione di continuità. Esprime il fatto che, in questo modello epidemico, si assume che la popolazione N sia costante.

La terza equazione é il tasso di cambiamento delle persone guarite. Il numero di guariti é proporzionale al numero di infetti, I , perché una persona guarita deve essere stata infettata.

La quarta equazione é il tasso di cambiamento del gruppo deceduto. Il numero di morti é proporzionale al numero di infetti, I , perché un individuo deceduto deve essere stato infettato (in questo modello). [9]

Esistono tuttavia modelli piú complessi che possono rappresentare con precisione la diffusione dinamica dell'epidemia.

Lin e colleghi [7] hanno esteso il modello SEIR, facendolo diventare SEIRNDC, considerando quindi la dimensione totale della popolazione N con due extra classi, quali la percezione del rischio per quanto riguarda il numero di casi gravi, critici e decessi (D) e il numero cumulativo dei casi sia segnalati che non segnalati (C).

Adottano la funzione di velocità di trasmissione, rinominano l'effetto del termine scolastico come effetto dell'azione governativa, poiché il primo appartiene al secondo; assumono anche un periodo di trasmissione zoonotica nel mese di dicembre 2019; modellano la trasmissione zoonotica (indicata come F) come una funzione graduale, che tende a zero dopo la chiusura del mercato ittico di Huanan (presumibilmente).

Quindi modificano solo la trasmissione uomo-uomo di COVID-19 dopo questa data, insieme all'emigrazione di 5 milioni di abitanti prima che Wuhan fosse ufficialmente bloccata. Pertanto, un modello compartimentale é formulato come segue:

$$\left\{ \begin{array}{l} S'(t) = -\frac{\beta_0 SF}{N} - \frac{\beta(t)SI}{N} - \mu S, \\ E'(t) = \frac{\beta_0 SF}{N} + \frac{\beta(t)SI}{N} - (\sigma + \mu)E, \\ I'(t) = \sigma E - (\gamma + \mu)I, \\ R'(t) = \gamma I - \mu R, \\ N'(t) = -\mu N, \\ D'(t) = d\gamma I - \lambda D, \\ C'(t) = \sigma E \end{array} \right. \quad (1.9)$$

dove

$$\beta(t) = \beta_0(1 - \alpha) \left(1 - \frac{D}{N}\right)^k. \quad (1.10)$$

La velocità di trasmissione, $\beta(t)$ nell'Eq. (1.10), incorpora l'impatto dell'azione governativa (tutte le azioni che possono essere modellate come una funzione graduale) e la diminuzione dei contatti tra gli individui che rispondono alla proporzione di decessi (cioè la gravità dell'epidemia).

In questo lavoro, si sono concentrati sulla trasmissione di COVID-19 a Wuhan, in Cina. Il quadro concettuale può essere applicato ad altre città / paesi o essere integrato in un modello a più patch per modellare il contesto di più città / paesi. Il modello inoltre può essere adattato ai dati giornalieri quando sono disponibili maggiori informazioni (ad esempio, il numero giornaliero di test).

Il merito di questo modello è che vengono considerati alcuni elementi essenziali, tra cui la risposta comportamentale individuale, le azioni governative, la trasmissione zoonotica e l'emigrazione di un'ampia percentuale della popolazione in un breve periodo di tempo. Nel frattempo, il modello è relativamente semplice, pertanto dovrebbe essere considerato come un modello di base per un ulteriore miglioramento.[7]

Tang et al. [15] ha studiato un modello epidemiologico generale di tipo SEIR in cui sono considerati quarantena, isolamento e trattamento.

Più in dettaglio, ha stratificato le popolazioni come compartimenti:

- suscettibili (S),
- esposti (E),
- infettivi ma non ancora sintomatici (pre - sintomatici) (A),
- infettivi con sintomi (I),

- ospedalizzati (H),
- guariti (R)

Inoltre la popolazione viene ulteriormente stratificata per includere i compartimenti:

- suscettibili in quarantena (S_q),
- esposti isolati (E_q),
- infetti isolati (I_q).

Con il tracciamento dei contatti, una percentuale q di individui esposti al virus viene messa in quarantena. Gli individui in quarantena possono spostarsi nel compartimento E_q o S_q , a seconda che siano effettivamente infetti o meno, mentre l'altra proporzione $(1 - q)$, che é costituita da individui esposti al virus che non sono stati rilevati dal tracciamento del contatto, può tuttavia spostarsi nel compartimento esposto E una volta effettivamente infettati, o rimanere nel compartimento S altrimenti.

Le dinamiche di trasmissione sono governate dal seguente sistema di equazioni:

$$\left\{ \begin{array}{l} S'(t) = -(\beta c + cq(1 - \beta))S(I + \theta A) + \lambda S_q, \\ E'(t) = \beta c(1 - q)S(I + \theta A) - \sigma E, \\ I'(t) = \sigma \rho E - (\delta_I + \alpha + \gamma_I)I, \\ A'(t) = \sigma(1 - \rho)E - \gamma_A A, \\ S'_q(t) = (1 - \beta)cqS(I + \theta A) - \lambda S_q, \\ E'_q(t) = \beta cqS(I + \theta A) - \delta_q E_q, \\ H'(t) = \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H, \\ R'(t) = \gamma_I I + \gamma_A A + \gamma_H H. \end{array} \right. \quad (1.11)$$

Sia β la probabilità di trasmissione e c la velocità di contatto costante. Quindi, gli individui in quarantena, se infetti (o non infetti), si spostano nel compartimento E_q (o S_q) a una velocità di βcq o $((1 - \beta)cq)$.

Coloro che non sono in quarantena, se infetti, si sposteranno nel compartimento E ad una velocità di $\beta c(1 - q)$. Gli individui infetti possono essere rilevati e quindi isolati a una velocità d_I e possono anche spostarsi nel compartimento R grazie al recupero. [15]

Infine si analizza il modello di *Gatto et al.* [2], il quale presenta un modello risolto spazialmente con nove variabili di stato per ciascuna delle 107 province del paese.

Qui si propone e si utilizza un modello che viene elaborato partendo dallo schema locale di base del modello di *Tang et al.*

L'evidenza empirica mostra che l'intervallo seriale di COVID-19 tende ad essere piú breve del periodo di incubazione, suggerendo cosí che una proporzione sostanziale di trasmissione secondaria puó avvenire prima dell'insorgenza della malattia. La trasmissione pre sintomatica sembra svolgere un ruolo importante nell'accelerare la diffusione della malattia all'interno di una comunitá, rappresentando circa il 12,6% dei casi clinici in Cina, il 48% a Singapore e il 62% a Tianjin, in Cina.

Il nucleo del modello é quindi chiamato SEPIA e comprende i seguenti compartimenti:

- Suscettibile (S),
- Esposto (E),
- Pre sintomatico (P),
- Infetto con sintomi pesanti (I),
- Asintomatico / lievemente sintomatico (A),
- Ricoverato in ospedale (H),
- In quarantena a casa (Q),

- Guariti (R),
- Morti (D).

La dinamica locale di trasmissione é data da

$$\left\{ \begin{array}{l} S'(t) = -\lambda S, \\ E'(t) = \lambda S - \delta_E E, \\ P'(t) = \delta_E E - \delta_P P, \\ I'(t) = \sigma \delta_P P - (\eta + \gamma_I + \alpha_I) I, \\ A'(t) = (1 - \sigma) \delta_P P - \gamma_A A, \\ H'(t) = (1 - \zeta) \eta I - (\gamma_H + \alpha_H) H, \\ Q'(t) = \zeta \eta I - \gamma_Q Q, \\ R'(t) = \gamma_I I + \gamma_A A + \gamma_H H, \\ D'(t) = \alpha_I I + \alpha_H H. \end{array} \right. \quad (1.12)$$

Nel modello, gli individui suscettibili (S) vengono esposti all'agente virale al contatto con individui infettivi, che si presume siano quelli nelle classi pre sintomatiche, fortemente sintomatiche o asintomatiche / lievemente sintomatiche.

Sebbene l'ipotesi possa non essere valida per alcune comunitá, si assumono tassi di contatto dipendenti dalla frequenza (come fa la maggior parte degli autori), in modo che l'esposizione si verifichi a una velocitá descritta dalla forza dell'infezione λ ,

$$\lambda = \frac{\beta_P P + \beta_I I + \beta_A A}{S + E + P + I + A + R}$$

dove β_P , β_I e β_A sono le velocitá di trasmissione specifiche delle tre classi infettive.

Gli individui esposti (E) sono infetti latentemente, cioé ancora non contagiosi, fino a quando non entrano nello stadio pre sintomatico (al tasso δ_E) e solo allora diventano infettivi.

Gli individui pre sintomatici (P) progrediscono (al tasso δ_P) per diventare individui infettivi sintomatici che sviluppano sintomi gravi (con probabilità σ). In alternativa, diventano individui asintomatici / lievemente sintomatici (con probabilità $1 - \sigma$) [2].

Gli individui infettivi sintomatici (I) escono dal loro compartimento se / quando:

- sono isolati dalla comunità (al tasso η) perché una frazione $1 - \zeta$ di loro é ricoverata in ospedale, mentre una frazione ζ é messa in quarantena a casa;
- guariscono dall'infezione (al tasso γ_I);
- muoiono (al tasso α_I).

Gli individui asintomatici / lievemente sintomatici (A), invece, lasciano il proprio compartimento dopo essersi ripresi dall'infezione (al tasso γ_A).

Gli individui ospedalizzati (H) possono guarire dall'infezione (al tasso γ_H) o morire a causa di esso (al tasso α_H), mentre gli individui isolati in casa (Q) lasciano il loro compartimento dopo il recupero (al tasso γ_Q). Le persone che guariscono dall'infezione o muoiono a causa del COVID-19 popolano la classe di individui guariti (R) e morti (D), rispettivamente, indipendentemente dal loro compartimento epidemiologico di origine [2].

Capitolo 2

Modelli matematici in epidemiologia

In epidemiologia, un modello matematico é un modello simbolico costituito da una o piú equazioni che prendono in considerazione i diversi parametri che sono coinvolti nella genesi e nell'evoluzione del fenomeno di interesse sanitario studiato (in genere: una malattia).

La formulazione dei modelli matematici é oggetto di studio della biomatematica, dove i modelli traggono origine dalla descrizione deterministica dell'evoluzione temporale dell'evento epidemico studiato, ossia della cinetica delle trasformazioni che possono comporlo.

I modelli matematici utilizzati in epidemiologia sono costruiti per scopi diversi, ad esempio: prevedere l'andamento di una malattia in determinate condizioni, oppure prevedere l'effetto sulla prevalenza o sull'incidenza qualora vengano adottate determinate misure di controllo, oppure calcolare il rischio di morte o l'aspettativa di vita nel corso di una epidemia o in specifiche condizioni ambientali.

Un buon modello permette di simulare ciò che avverrà in natura e quindi può rappresentare un utilissimo strumento nello studio delle malattie.

I principali si pongono tra modelli deterministici o stocastici, chiusi o aperti, omogenei o eterogenei.

- *Modello deterministico*: i modelli deterministici sono i piú semplici; in essi, le

variabili di input assumono valori fissi, determinati. In un modello deterministico, gli individui nella popolazione sono assegnati a diversi sottogruppi o compartimenti, ognuno dei quali rappresenta uno stadio specifico dell'epidemia. I tassi di transizione da una classe all'altra sono espressi matematicamente come derivate, quindi il modello é formulato usando equazioni differenziali. Durante la costruzione di tali modelli, si deve presumere che la dimensione della popolazione in un compartimento sia differenziabile rispetto al tempo e che il processo epidemico sia deterministico. In altre parole, i cambiamenti nella popolazione di un compartimento possono essere calcolati utilizzando solo la cronologia utilizzata per sviluppare il modello.

- *Modello stocastico*: i modelli stocastici tengono in considerazione le variazioni delle variabili di input, e forniscono risultati in termini di "probabilitá". Un modello stocastico é uno strumento che permette di stimare le distribuzioni di probabilitá di potenziali esiti consentendo variazioni casuali in uno o piú input nel tempo. I modelli stocastici dipendono dalle variazioni del rischio di esposizione, malattia e altre dinamiche della malattia. Potendo inserire la variabilitá dei dati di input, i modelli stocastici hanno una struttura piú complessa di quelli deterministici ma riescono a essere piú aderenti alla realtá.

Nei modelli compartimentali si assumono delle ipotesi per semplificare la simulazione matematica della dinamica delle malattie infettive, sinteticamente tali ipotesi presumono che la popolazione sia divisa in compartimenti e che ogni individuo nello stesso compartimento abbia le stesse caratteristiche.

I modelli compartimentali sono di solito costruiti con equazioni differenziali ordinarie (che sono deterministiche), ma, utilizzando come input delle probabilitá, possono anche essere visualizzati in un quadro stocastico che é piú realistico ma anche piú complicato da analizzare.

I modelli compartimentali possono essere utilizzati per prevedere le proprietá della diffusione di una malattia, ad esempio la prevalenza (numero totale di infetti) o la durata di un'epidemia. Inoltre, il modello consente di comprendere in che modo diverse situazioni possono influenzare l'esito dell'epidemia, ad esempio quale sia la

percentuale di vaccinazioni che in una determinata popolazione procura l'immunità di gregge oppure quale variazione nel numero di riproduzione effettivo produce un contenimento dell'epidemia.

Sono quasi cento anni che si usano modelli matematici per descrivere la dinamica delle epidemie.

Per lo sviluppo dei modelli sono in genere necessari:

- il tasso di infezione o la velocità di trasmissione;
- il tasso di recupero;
- il tasso di latenza o di incubazione;
- il tasso di ospedalizzazione.

Nelle epidemie che si sviluppano in tempi molto lunghi o in equilibrio endemico normalmente si considerano anche il tasso di natalità e il tasso di mortalità, e nelle condizioni di stato stazionario ed endemico, anche l'età media in cui i soggetti si ammalano e l'aspettativa di vita media della popolazione. Quindi più il modello è complesso, più aumentano i parametri che lo descrivono.

Tuttavia, sia con un modello epidemiologico semplice sia con uno complesso, il procedimento per analizzarlo è analogo, ossia bisogna risolvere due tipi di problemi: diretto ed inverso.

2.1 Problema diretto

Si intende problema diretto quello in cui si forniscono sufficienti informazioni per poter avviare un procedimento ben definito e stabile che porti ad un'unica soluzione del problema.

$$\begin{array}{ccccc} \text{informazioni} & \rightarrow & \text{procedimento} & \rightarrow & \text{soluzione} \\ (\text{input}) & & & & (\text{output}) \end{array}$$

Se il processo descrive un fenomeno fisico o comunque del mondo reale, si può descrivere il problema diretto come:

$$\begin{array}{ccccc} \text{causa} & \rightarrow & \text{modello} & \rightarrow & \text{effetto} \\ x & \rightarrow & K & \rightarrow & y \end{array}$$

cioé

$$K(x) = y$$

Il problema diretto consiste nel calcolare l'effetto y , assegnata la causa x e il modello K .

Quindi se si vuole predire il comportamento futuro di un sistema fisico conoscendo il suo stato presente e le leggi fisiche che lo governano, allora si dice che si vuole risolvere un problema diretto.

2.1.1 Problema di Cauchy o IVP

Molti fenomeni fisici sono descrivibili con modelli matematici costituiti da una o più equazioni differenziali ordinarie del tipo [10]

$$\left\{ \begin{array}{l} y_1'(x) = f_1(x, y_1(x), \dots, y_m(x)) \\ \vdots \\ y_m'(x) = f_m(x, y_1(x), \dots, y_m(x)) \end{array} \right. \quad (2.1)$$

con condizioni iniziali in un unico punto $x=a$

$$\left\{ \begin{array}{l} y_1(a) = y_{1,0} \\ \vdots \\ y_m(a) = y_{m,0} \end{array} \right. \quad (2.2)$$

In questi casi l'obiettivo matematico diventa la determinazione di funzioni (derivabili) $y_1(x), \dots, y_m(x)$ soluzione del sistema (2.1) e passanti per i punti definiti dalla (2.2).

Il problema di Cauchy o Initial Value Problem IVP determina dunque la soluzione di un'equazione differenziale ordinaria, scalare o vettoriale, completata da condizioni iniziali appropriate; é vettoriale quando si parla di sistemi di equazioni differenziali.

Si determina una funzione $y(x)$, continua e derivabile sull'intervallo I in \mathbb{R} tale che

$$\left\{ \begin{array}{ll} y'(x) = f(x, y(x)) & \forall x \in I \equiv [a, b] \\ y(x_0) = y_0 & x_0 \in [a, b] \end{array} \right. \quad (2.3)$$

In generale si può avere f come una qualsiasi funzione della variabile dipendente y e della variabile indipendente x , di conseguenza f può anche essere non lineare.

Avendo in questo elaborato un sistema di equazioni differenziali, si avranno dunque m funzioni soluzioni dipendenti da x che soddisfano contemporaneamente le m equazioni differenziali; si avranno inoltre tante condizioni iniziali quante sono le funzioni.

Prima di procedere allo studio di metodi numerici, si ricordano alcuni importanti risultati di carattere teorico riguardanti l'esistenza e il comportamento delle soluzioni di problemi del tipo (2.3).

2.1.2 Risolubilità

Per ogni problema bisogna sempre verificare l'esistenza e l'unicità della soluzione. Si dice che esiste un'unica soluzione quando la f (funzione del problema) è continua e anche la derivata prima è continua.

Si introduce il teorema di esistenza ed unicità della soluzione in termini di *Lipschitzianità*.

Una funzione f è Lipschitziana nell'intervallo I di esistenza, se esiste una costante L tale per cui, per una qualunque coppia di valori (x_1, x_2) , il rapporto incrementale della funzione valutata in questa coppia di valori sia sempre limitato, quindi minore o uguale di una certa costante L .

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq L \quad (2.4)$$

$$x \in I = [a, b], \quad \forall x_1, x_2 \in \mathbb{R} \quad (2.5)$$

Con L costante di Lipschitz, essa è un valore finito.

Quindi il rapporto incrementale della funzione, in un certo intervallo, presi i due punti, è sempre superiormente limitato; chiaramente se f è derivabile, allora f è anche Lipschitziana.

Il teorema dell'esistenza ed unicità della soluzione analitica del problema lo si rivede dunque sotto questa condizione e non più sotto le classiche ipotesi in cui f sia continua e con derivata prima continua, ma che f sia Lipschitziana nella variabile. La variabile è la funzione soluzione del problema, ossia y_1 e y_2 .

Il teorema dice che per ogni punto iniziale preso nell'intervallo di integrazione $[a, b]$, e per ogni valore della condizione iniziale y_0 , esiste esattamente una ed una sola soluzione che è la funzione $y(x)$ differenziabile, e che risponde al problema di Cauchy:

$$\begin{cases} y'(x) = f(x, y(x)) \\ y(x_0) = y_0 \end{cases} \quad (2.6)$$

Quindi $\forall x_0 \in [a,b]$ e $\forall y_0 \in \mathbb{R}^m$, esiste esattamente una funzione $y(x)$ che soddisfa le seguenti condizioni:

- $y(x) \in C^1 [a,b]$,
- $y' = f(x,y(x))$,
- $y(x_0) = y_0$.

Affinché esista una soluzione si richiede che la f del problema, ossia la funzione che mette in relazione la variabile indipendente x con quella dipendente y , sia Lipschitziana.

Dal punto di vista teorico ci si chiede se il problema diretto gode di alcune proprietà che corrispondono alla definizione di problema *ben posto*.

Si introducono brevemente le nozioni di problemi ben posti; nel 1923, J. Hadamard diede la seguente definizione di problema ben posto:

1. Esiste una soluzione del problema (esistenza) ;
2. La soluzione è unica (unicità);
3. La soluzione dipende con continuità dai dati (stabilità).

Queste imposizioni permettono di risolvere il problema con metodi numerici e in maniera opportuna. Il problema IVP considerato risulta essere ben posto, si procede quindi allo studio del condizionamento.

2.1.3 Condizionamento

Ci si chiede anche se il problema diretto è *ben condizionato*, ossia si vuol capire, introducendo delle piccole perturbazioni sui dati iniziali, come si ripercuote questo sui risultati e quindi cosa succede alla soluzione indipendentemente dal metodo che si utilizza. Ci si concentra sul condizionamento di equazioni differenziali non lineari.

Nelle applicazioni pratiche, le misure sperimentali sono sempre più o meno perturbate da una certa quantità di rumore, dovuta all'azione di agenti esterni quali ad

esempio lo strumento di misurazione stesso. È importante dunque andare a studiare l'influenza di tali perturbazioni sulla soluzione del problema ai minimi quadrati perturbato e vedere quanto distante essa sia dalla soluzione reale ossia dal problema non perturbato.

Il problema non perturbato è il seguente:

$$\begin{cases} y'(x) = f(x, y(x)) \\ y(x_0) = y_0 \end{cases} \quad (2.7)$$

Mentre il problema perturbato è il seguente:

$$\begin{cases} z'(x) = f(x, z(x)) \\ z(x_0) = y_0 + \epsilon \end{cases} \quad (2.8)$$

nel quale viene aggiunta una piccola perturbazione ϵ solo nel termine noto ossia nella condizione iniziale.

Si pone $\delta(x) = z(x) - y(x)$, dove $\delta(x)$ è la funzione scostamento ossia è la distanza tra $z(x)$ e $y(x)$. Se ci si aspetta che lo scostamento tra $z(x) - x_0$ e $y(x) - x_0$, ossia l' ϵ introdotto precedentemente, sia piccolo, allora se il problema è ben condizionato si vorrebbe che anche $\delta(x)$ rimanesse piccolo $\forall x$.

Detto ciò, bisogna fare un'approssimazione della funzione f , la quale rappresenta il sistema stesso, la si sviluppa nell'intorno del punto $(x, y(x))$.

Essendo non lineare, l'obiettivo è linearizzarla con Taylor.

$$f(x, z(x)) = f(x, y(x)) + f_y(x, y(x)) \cdot \delta(x) + O(\|\delta\|^2) \quad (2.9)$$

dove f_y é la matrice Jacobiana, mentre $O(\|\delta\|^2)$ sono i termini del secondo ordine che non verranno considerati.

Sostituendo la parte lineare nel sistema non lineare si ottiene:

$$\begin{cases} z' = y' + \delta'(x) \cong f(x, y(x)) + f_y(x, y(x)) \cdot \delta(x) \\ z(x_0) = y(x_0) + \delta(x_0) = y_0 + \epsilon \end{cases} \quad (2.10)$$

La matrice Jacobiana viene definita come:

$$\{f_y(x, y(x))\}_{ij} = \{J_f\}_{ij} = \frac{\partial(f(x, y))}{\partial(y_j)} \quad (2.11)$$

Essa é una matrice di derivate, ancora non lineare poiché é una matrice di funzioni.

Si semplifica dunque il sistema precedente, elidendo i termini tra loro uguali, e si ottiene il seguente sistema semplificato:

$$\begin{cases} \delta'(x) \cong f_y(x, y(x)) \cdot \delta(x) \\ \delta(x_0) = \epsilon \end{cases} \quad (2.12)$$

Adesso si ha un problema da risolvere che indica quanto vale $\delta(x)$ ossia lo scostamento tra la soluzione perturbata e quella non perturbata; si ricordi però che $f_y(x, y(x))$, matrice di derivate non costante, é ancora presente all'interno del sistema.

Per cui viene fatta un'assunzione forzata, poiché si sta studiando cosa succede in un intorno, si suppone che quella matrice sia pressocché costante, ossia viene valutata nel punto iniziale, quindi non é piú una matrice di derivate bensí di costanti.

$$f_y(x, y(x)) \cong f_y(x_0, y_0) \quad (2.13)$$

Quindi il sistema precedentemente analizzato, non lineare, diventa un sistema lineare:

$$\begin{cases} \delta'(x) \cong f_y(x_0, y_0) \cdot \delta(x) \\ \delta(x_0) = \epsilon \end{cases} \quad (2.14)$$

La soluzione del sistema é quindi:

$$\delta(x) = \epsilon \cdot \exp^{f_y(x-x_0)} \quad (2.15)$$

Il problema si dice dunque *ben condizionato* se:

$\|\delta(x)\| \rightarrow \inf$ per $x \rightarrow \infty$ quando $\Re(\lambda_l) < 0$ per $l=1,2,\dots,m$, quindi è asintoticamente stabile.

Mentre il problema si dice *mal condizionato* se:

$\|\delta(x)\| \rightarrow \infty$ per $x \rightarrow \infty$ quando $\Re(\lambda_l) > 0$ per almeno un indice l , quindi potrebbe divergere.

Si conclude questo esame preliminare ribadendo che lo studio della propagazione di una perturbazione iniziale ϵ é stato possibile solo supponendo trascurabile il termine $O(\|\delta\|^2)$ in (2.9) e costante la matrice Jacobiana $f_y(x, y(x))$. In realtà queste ipotesi spesso non sono verificate e il comportamento di $\delta(x)$ può non essere validamente rappresentato dagli autovalori di $f_y(x_0, y_0)$.

2.1.4 Metodi numerici per la soluzione di IVP

Una volta compreso che il problema é sia ben condizionato che ben posto, é possibile risolverlo con dei metodi numerici; questo vuol dire plottare per punti la funzione a partire dal dato iniziale (x_0, y_0) .

Si ha un intervallo di integrazione, di cui solitamente x_0 é l'estremo sinistro, che viene discretizzato poiché bisogna ricostruire una funzione su quell'intervallo di integrazione. Si prende un certo passo di discretizzazione h , si conosce inoltre quanto vale la soluzione nel punto iniziale (x_0, y_0) .

A partire dal primo punto, si vede quanto vale la soluzione nel primo nodo, e la si disegna, poi si passa al secondo nodo, al terzo nodo e così via. Quindi viene disegnata tutta la funzione incognita $y(x)$ per punti. Questo é il criterio sul quale i metodi numerici basano la loro risoluzione; dunque la soluzione approssimata viene chiamata $u(x)$.

La coppia (x_j, u_j) approssima la soluzione in $(x_j, y(x_j))$. Quindi il calcolo numerico dará luogo a $u_j = u_1, u_2, u_3, \dots$ che permettono di ricostruire la funzione incognita e naturalmente sará presente un piccolo scostamento tra la soluzione approssimata e la soluzione esatta in quel punto.

Con il metodi numerici che verranno presentati di seguito ci si accontenterá di determinare, in un numero finito di punti $\{x_n\}$ dell'intervallo di interesse $[\alpha, \beta]$, delle approssimazioni $\{y_n\}$ dei valori che la soluzione $y(x)$ assume nei nodi $\{x_n\}$. Ovviamente, note le $\{y_n\}$, mediante l'uso di tecniche di interpolazione sará sempre possibile dedurre approssimazioni della $y(x)$ in punti x diversi dagli $\{x_n\}$. [10]

I predetti metodi vengono raggruppati in due classi principali:

1. *Metodi one step*, o ad un solo passo. Il valore y_{n+1} viene calcolato utilizzando unicamente l'approssimazione precedente y_n ; ossia il metodo si presenta in una delle tre forme seguenti:

$$\begin{aligned} y_{n+1} &= y_n + h\Phi(x_n, y_n; h) && (\textit{forward Euler esplicito}) \\ y_{n+1} &= y_n + h\Phi(x_{n+1}, y_{n+1}; h) && (\textit{backward Euler implicito}) \quad (2.16) \\ y_{n+1} &= y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1})] && (\textit{Crank - Nicolson}) \end{aligned}$$

generalmente $h=h(n)$.

2. *Metodi multistep*, o a piú passi. Il valore y_{n+1} viene definito utilizzando piú approssimazioni precedenti $y_n, y_{n-1}, \dots, y_{n-k+1}$ relative ai punti $x_n, x_{n-1}, \dots, x_{n-k+1}$, $x_i = a + ih$, con formule del tipo

$$y_{n+1} = \Psi(x_n, y_n, y_{n-1}, \dots, y_{n-k+1}; h) \quad (\text{metodo multistep esplicito}) \quad (2.17)$$

oppure

$$y_{n+1} = \Psi(x_n, y_{n+1}, y_n, \dots, y_{n-k+1}; h) \quad (\text{metodo multistep implicito}) \quad (2.18)$$

l'intero k va supposto fisso per tutti gli n .

Il metodo viene denominato *multistep* a k passi. I metodi espliciti calcolano lo stato del sistema in un secondo momento dallo stato del sistema al momento corrente senza la necessità di risolvere equazioni algebriche. I metodi impliciti, contrariamente ai metodi espliciti, trovano la soluzione risolvendo un'equazione che coinvolge lo stato corrente del sistema e quello successivo.

Tuttavia i metodi espliciti, a differenza di quelli impliciti, sono significativamente piú semplici da implementare, inoltre l'esecuzione di un solo passo di integrazione é in genere molto piú veloce.

2.1.5 Metodi one step espliciti: metodi Runge-Kutta

Nel processo che conduce dal metodo di Eulero in avanti verso metodi di ordine piú elevato del primo, i metodi multistep lineari ed i metodi Runge-Kutta (RK) si ispirano a criteri opposti [14].

Come il metodo di Eulero, i metodi multistep sono lineari rispetto ad u_n e $f_n = f(t_n, u_n)$, inoltre richiedono una sola valutazione funzionale per ogni passo temporale e guadagnano accuratezza incrementando il numero dei passi.

I metodi RK, al contrario, guadagnano accuratezza conservando la struttura ad un passo, ma sacrificando la linearità al prezzo di un aumento del numero di

valutazioni funzionali per ogni passo. In questo modo é facile modificare il passo di integrazione per i metodi RK, ma si perde la possibilitá di valutare in modo semplice l'errore locale con le tecniche proprie dei metodi multistep.

Nella forma piú generale un metodo RK puó essere scritto come:

$$u_{n+1} = u_n + hF(t_n, u_n, h; f), \quad n \geq 0 \quad (2.19)$$

dove F é la funzione incremento definita nel modo seguente:

$$F(t_n, u_n, h; f) = \sum_{i=1}^s b_i K_i, \quad (2.20)$$

con

$$K_i = f(t_n + c_i h, u_n + h \sum_{j=1}^s a_{ij} K_j), \quad i = 1, 2, \dots, s$$

ed s indica il numero di stadi del metodo. I coefficienti a_{ij} , c_i e b_i caratterizzano completamente un metodo RK e vengono generalmente raccolti nella cosiddetta *matrice di Butcher* [14]

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & & a_{2s} \\ \vdots & \vdots & & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array} \quad o \quad \begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

essendo $A = (a_{ij}) \in \mathbb{R}^{s \times s}$, $\mathbf{b} = (b_1, b_2, \dots, b_s)^T \in \mathbb{R}^s$ e $\mathbf{c} = (c_1, \dots, c_s)^T \in \mathbb{R}^s$.

Si suppone inoltre che valga la seguente condizione

$$c_i = \sum_{j=1}^s a_{ij} \quad i = 1, \dots, s \quad (2.21)$$

Se i coefficienti a_{ij} in A sono nulli per $j \geq i$, con $i=1, 2, \dots, s$, allora ogni K_i può essere calcolato esplicitamente in funzione dei soli $i - 1$ coefficienti K_1, \dots, K_{i-1} già precedentemente calcolati. Per questo motivo, in tal caso, lo schema viene detto *esplicito*.

In caso contrario, lo schema RK é *implicito* ed il calcolo dei K_i richiede la risoluzione di un sistema non lineare di dimensione s . L'aggravio computazionale da esso comportato rende eccessivamente onerosi questi schemi.

Per l'analisi di *consistenza*, si definisce l'errore di troncamento locale $\tau_{n+1}(h)$ nel nodo τ_{n+1} utilizzando l'equazione del residuo

$$h\tau_{n+1}(h) = y_{n+1} - y_n - hF(t_n, y_n, h; f),$$

con $y(t)$ soluzione del problema di Cauchy di partenza. Il metodo (2.19) é detto *consistente* se $\tau(h) = \max_n |\tau_n(h)| \rightarrow 0$ per $h \rightarrow 0$. Si verifica facilmente che ciò accade se e solo se

$$\sum_{i=1}^s b_i = 1 \tag{2.22}$$

Al solito, si dirá poi che (2.19) é un metodo di ordine p (≥ 1) rispetto ad h se $\tau(h) = O(h^p)$ per $h \rightarrow 0$.

Per quanto riguarda la *convergenza* di un metodo RK, é sufficiente osservare che, trattandosi di metodi ad un passo, la consistenza implica la stabilitá e quindi la convergenza. É possibile ottenere, come per i metodi multistep, stime dell'errore di troncamento $\tau(h)$: tuttavia, queste stime sono troppo complicate per poter essere utilizzate con profitto nella pratica. Si noti soltanto che, come per i metodi multistep, se un metodo RK ha un errore di troncamento locale $\tau_n(h) = O(h^p)$, per ogni n , allora anche l'ordine di convergenza sará p . [14]

Per quanto riguarda l'ordine degli schemi espliciti, vale la seguente proprietá:

Teorema 1.1 *Un metodo Runge-Kutta esplicito a s stadi non può avere ordine maggiore di s . Non solo, non esistono metodi Runge-Kutta espliciti a s stadi con ordine s se $s \geq 5$.*

In particolare, per gli ordini da 1 a 8, il minimo numero di stadi s_{min} necessario per ottenere un metodo di ordine corrispondente é riportato nella seguente tabella

<i>ordine</i>	1	2	3	4	5	6	7	8
s_{min}	1	2	3	4	6	7	9	11

Si noti che 4 é il massimo numero di stadi in corrispondenza del quale l'ordine non é inferiore al numero di stadi stesso.

Derivazione di un metodo Runge-Kutta esplicito

I metodi di Runge-Kutta possono essere ricavati o per integrazione numerica o tramite l'utilizzo di opportuni sviluppi in serie. Si analizza un esempio della prima procedura. [14]

Si consideri il seguente metodo approssimato per calcolare la soluzione di un problema di Cauchy: per ogni $n \geq 0$, trovare u_{n+1} tale che

$$u_{n+1} = u_n + h \int_{t_n}^{t_{n+1}} f(t, y(t)) dt$$

e si suppone di approssimare l'integrale con il metodo del punto medio, ossia

$$\int_{t_n}^{t_{n+1}} f(t, y(t)) dt \simeq \frac{h}{2} f(t_{n+1/2}, y_{n+1/2}) \simeq \frac{h}{2} f(t_{n+1/2}, u_{n+1/2})$$

Il problema é che non si conosce neppure $u_{n+1/2}$. D'altra parte si potrebbe pensare di approssimare tale valore con un semi-passo di Eulero esplicito, ad esempio si troverebbe:

$$u_{n+1/2} = u_n + \frac{h}{2} f(t_n, u_n) = \frac{h}{2} f_n.$$

Si é costruito in tal modo un metodo Runge-Kutta esplicito a due stadi. Si tratta in effetti di un metodo di questa forma

$$\begin{aligned} u_{n+1} &= u_n + \frac{h}{2} K_2 \\ K_1 &= f_n, \\ K_2 &= f\left(t_n + \frac{h}{2}, u_n + \frac{h}{2} K_1\right) \end{aligned}$$

Adattività del passo per i metodi Runge-Kutta

Essendo ad un passo, i metodi RK ben si prestano al cambio del passo di integrazione h , purché si disponga di uno stimatore efficiente dell'errore locale commesso al singolo passo. Questi stimatori sono in generale *a posteriori* in quanto le stime a priori dell'errore locale sono troppo complicate da utilizzare nella pratica e possono essere ottenute in due modi [14]:

- utilizzando lo stesso metodo RK con due passi diversi (tipicamente $2h$ e h);
- impiegando contemporaneamente due metodi RK di ordine diverso, ma con lo stesso numero s di stadi.

Nel primo caso, si suppone di impiegare un metodo RK di ordine p e si pretende che, partendo dal dato esatto $u_n = y_n$ (ovviamente indisponibile, se $n \geq 1$), l'errore locale che si commette sia inferiore ad una tolleranza prestabilita. Osserviamo che vale la relazione

$$y_{n+1} - u_{n+1} = \Phi(y_n)h^{p+1} + O(h^{p+2}), \quad (2.23)$$

essendo Φ una funzione incognita valutata in y_n . (Si osservi che, in questa speciale situazione, risulta $y_{n+1} - u_{n+1} = h\tau_{n+1}(h)$).

Eseguendo lo stesso calcolo con passo $2h$ a partire da t_{n-1} ed indicando con \hat{u}_{n+1} la soluzione corrispondente, si trova

$$y_{n+1} - \hat{u}_{n+1} = \Phi(y_{n-1})(2h)^{p+1} + O(h^{p+2}) = \Phi(y_n)(2h)^{p+1} + O(h^{p+2}) \quad (2.24)$$

avendo supposto di sviluppare in serie anche y_{n-1} rispetto a t_n . Sottraendo la (2.23) alla (2.24), si perviene alla relazione

$$(2^{p+1} - 1)h^{p+1}\Phi(y_n) = u_{n+1} - \hat{u}_{n+1} + O(h^{p+2}), \quad (2.25)$$

da cui

$$y_{n+1} - u_{n+1} \simeq \frac{u_{n+1} - \hat{u}_{n+1}}{(2^{p+1} - 1)} = \mathcal{E} \quad (2.26)$$

Se $|\mathcal{E}|$ é minore della tolleranza ϵ fissata, lo schema prosegue al passo successivo, mentre in caso contrario esegue nuovamente la stima con un passo dimezzato. In generale si raddoppia il passo di integrazione quando $|\mathcal{E}|$ é minore di $\epsilon/2^{p+1}$.

Questo approccio comporta un considerevole aggravio computazionale, per via delle $s-1$ valutazioni funzionali addizionali necessarie per generare il valore \hat{u}_{n+1} . Inoltre, nel caso in cui fosse necessario dimezzare il passo, dovrà essere ricalcolato anche il valore di u_n . Un'alternativa che non comporti valutazioni funzionali addizionali é quella di utilizzare simultaneamente due metodi RK a s stadi, di ordine p e $p+1$ rispettivamente, che presentino lo stesso insieme di valori K_i .

Questi metodi vengono rappresentati in modo sintetico tramite la matrice di Butcher modificata

$$\begin{array}{c|c} c & A \\ \hline & b^T \\ & \hat{b}^T \\ \hline & E^T \end{array}$$

identificando il metodo di ordine p con i coefficienti \mathbf{c} , A e \mathbf{b} e quello di ordine $p+1$ con \mathbf{c} , A e $\hat{\mathbf{b}}$, ed essendo $\mathbf{E} = \mathbf{b} - \hat{\mathbf{b}}$ [14].

É semplice verificare che la differenza fra le soluzioni approssimate in t_{n+1} ottenute usando i due schemi, fornisce una stima dell'errore di troncamento locale per lo schema di ordine inferiore. D'altra parte, essendo i K_i uguali, questa differenza é data proprio da $h \sum_{i=1}^s E_i K_i$ e non richiede dunque ulteriori valutazioni funzionali.

Si noti che se viene usata la u_{n+1} generata dallo schema di ordine p per innescare lo schema al passo $n + 2$, il metodo avrá complessivamente ordine p . Se al contrario venisse usata la soluzione generata dallo schema di ordine $p + 1$, si troverebbe uno schema ancora di ordine $p + 1$. [14]

2.2 Problema inverso

Il problema inverso consiste nel determinare la causa x dato il modello K e le osservazioni y , ovvero

$$x = K^{-1}(y)$$

Quindi consiste nel risalire allo stato passato di un sistema dalla conoscenza della condizione presente, oppure nel determinare il valore di certi parametri, conoscendo l'evoluzione del sistema.

Nel caso specifico analizzato in questo elaborato il problema inverso da risolvere consiste nel voler calcolare l'insieme dei parametri che caratterizzano il modello generico epidemico SIR o SIDARTHE per l'epidemia specifica studiata Covid-19, partendo da dati sperimentali noti prelevati dal database della Protezione Civile.

Parlare di problema inverso dal punto di vista matematico significa dunque partire dai dati sperimentali \bar{y} presenti nel database, dove \bar{y} rappresenta il vettore contenente le variabili del modello epidemico generico che si vuole rappresentare, e voler costruire un modello differenziale non lineare $M(\Lambda, t)$ con $\Lambda = (\Lambda_1, \dots, \Lambda_m)$ vettore dei parametri del modello, che descriva tali dati. La scelta di $M(\Lambda, t)$ dipenderá ovviamente dal tipo di problema considerato.

I parametri Λ da determinare sono ottenuti risolvendo il problema ai minimi quadrati non lineare

$$\min_{\Lambda} \|y_j - M(\Lambda, t)\|_2^2 \quad (2.27)$$

Inoltre il problema inverso é solitamente *mal posto*.

Un problema si dice mal posto quando non soddisfa una delle proprietà elencate precedentemente descritte da Hadamard, caratterizzanti invece un problema ben posto. Ad esempio, é mal posto un problema che non ha soluzioni, oppure che ne ha piú di una. Ma la condizione piú esigente é quella della stabilità.

Ci sono numerosi problemi che hanno una e una sola soluzione, ma tale soluzione é instabile. Instabilità significa che cause molto diverse possono provocare effetti molto simili rendendo cosí vano il tentativo di risalire ad esse.

Il problema inverso che viene trattato in questo elaborato, relativo a due differenti modelli epidemiologici, consiste nella risoluzione di un problema di minimi quadrati non lineari, analizzato di seguito piú in dettaglio.

2.2.1 Problema di minimi quadrati non lineari

Il problema di minimi quadrati non lineari che si vuole risolvere é un problema di minimizzazione della forma:

$$\min_{x \in \mathbb{R}^n} f(x), \quad f(x) = \frac{1}{2} \|r(x)\|_2^2 = \frac{1}{2} \sum_{j=1}^m r_j^2(x) \quad (2.28)$$

dove le funzioni $r_j : \mathbb{R}^n \rightarrow \mathbb{R}$ dette *residui*, sono funzioni non lineari.

La particolaritá di questo problema é che la funzione obiettivo $f(x)$ é una funzione non lineare all'interno di una norma due.

Sia J la matrice Jacobiana di r , ossia la matrice $m \times n$ definita come segue

$$J(x) = \begin{bmatrix} \nabla r_1(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad (J(x))_{ij} = \frac{\partial r_i(x)}{\partial x_j} \quad (2.29)$$

Il gradiente e la matrice Hessiana della funzione obiettivo, utilizzando queste notazioni, si scrivono come segue:

$$\nabla f(x) = \frac{1}{2} \sum_{j=1}^m \nabla r_j^2(x) = \sum_{j=1}^m r_j(x) \nabla r_j(x) = J(x)^T r(x) \quad (2.30)$$

$$\nabla^2 f(x) = \sum_{j=1}^m \nabla r_j(x) \nabla r_j(x)^T + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x) = J(x)^T J(x) + S(x) \quad (2.31)$$

con $S(x) = \sum_{j=1}^m r_j(x) \nabla^2 r_j(x)$.

I metodi numerici di risoluzione di problemi ai minimi quadrati non lineari sono basati in generale sull'assunzione che il termine $S(x)$ sia trascurabile rispetto al termine $J(x)^T J(x)$.

Problemi di questa natura trovano numerose applicazioni, in modo particolare nel *data fitting*, dove spesso i modelli matematici utilizzati per il fit dei dati sono di natura non lineare.

2.2.2 Risoluzione dei problemi di minimi quadrati non lineari

I problemi di minimi quadrati non lineari sono, in generale, problemi non convessi a cui possono essere applicati metodi di ottimizzazione non vincolata; in particolare i piú utilizzati sono i metodi di ottimizzazione con proprietá regolarizzanti [3].

Di seguito verranno analizzati il metodo di Gauss-Newton e di Levenberg-Marquardt per la risoluzione del problema (2.28).

Metodo di Gauss-Newton

Per capire quale sia l'idea alla base del metodo si consideri lo sviluppo di Taylor di secondo grado della funzione obiettivo f nel punto x_k :

$$f(x_k + d) = f(x_k) + d^T \nabla f(x_k) + \frac{1}{2} d^T \nabla^2 f(x_k) d + O(d^3) \quad (2.32)$$

Il metodo di Newton é un metodo iterativo che, a partire da un iterato iniziale x_0 , calcola l'iterato successivo $x^{(k+1)} = x^{(k)} + \alpha^{(k)} d^{(k)}$ con $\alpha \in \mathbb{R}$, calcolando il passo di discesa d che minimizzi la funzione

$$f^*(d) = f(x_k) + d^T \nabla f(x_k) + \frac{1}{2} d^T \nabla^2 f(x_k) d \quad (2.33)$$

Condizione necessaria affinché d sia un punto di minimo per f^* é che il gradiente di f^* si annulli in d o equivalentemente che valga

$$\begin{aligned} \nabla^2 f(x_k) d &= -\nabla f(x_k) \\ (J(x_k)^T J(x_k) + S(x_k)) d &= -J(x_k)^T r(x_k) \end{aligned} \quad (2.34)$$

Poiché il calcolo del termine $S(x_k)$ potrebbe essere troppo costoso o le derivate seconde dei residui potrebbero non essere accessibili, il metodo di Gauss-Newton si fonda sull'assunzione che il termine $S(x_k)$ sia trascurabile. Questo ovviamente presuppone che l'iterato x_k sia sufficientemente vicino al punto di minimo (equivalentemente che i residui siano sufficientemente piccoli).

Il metodo iterativo di Gauss-Newton é dunque il seguente.

Viene calcolato, ad ogni iterazione k , il passo d_k come soluzione del sistema lineare.

$$J(x_k)^T J(x_k) d = -J(x_k)^T r(x_k) \quad (2.35)$$

Quindi la soluzione viene aggiornata

$$x_{k+1} = x_k + \alpha d_k. \quad (2.36)$$

Osservazioni:

- Il metodo di Gauss-Newton approssima localmente il problema ai minimi quadrati non lineare con un problema ai minimi quadrati lineare.
- Se il rango della matrice $J(x_k)$ é massimo, d_k é sicuramente una direzione di discesa per la f , infatti:

$$\nabla f(x_k)^T d_k = r(x_k)^T J(x_k) d_k = -d_k^T J^T(x_k) J(x_k) d_k = -\|J(x_k) d_k\|^2 \quad (2.37)$$

- A causa del carattere locale dell'approssimazione di Taylor, la lunghezza del passo potrebbe non essere corretta, cioè $f(x_k + d_k) > f(x_k)$.

Per ovviare a questo tipo di problema esistono algoritmi specifici per la scelta della lunghezza del passo che consentono di assicurarsi una sufficiente riduzione della funzione ad ogni passo k : un metodo molto utilizzato, ad esempio, é quello di scegliere un passo di discesa α definito precedentemente, tale che :

$$f(x_k + \alpha_k d_k) < f(x_k) + \alpha_k \nabla f(x_k)^T d \quad (2.38)$$

- Risolvere ad ogni passo iterativo il sistema (2.35) equivale a risolvere con il metodo delle equazioni normali il problema di minimi quadrati

$$\min_{d \in \mathbb{R}^n} \|J_k d + r_k\|_2^2. \quad (2.39)$$

Regolarizzazione di Tikhonov

In caso di problemi ai minimi quadrati mal condizionati, il metodo di Gauss-Newton può essere utilizzato ma risulta poco efficiente, soprattutto se il più piccolo valore singolare della matrice $J(x)^T J(x)$ tende a zero. In questi casi si preferisce ricorrere a metodi di regolarizzazione, il più famoso è il *metodo di Tikhonov* [3].

L'idea alla base del metodo è quella di aggiungere un vincolo alla norma della soluzione del problema, trasformandolo così in un problema di ottimizzazione vincolata. Utilizzando il metodo dei moltiplicatori di Lagrange si dimostra che risolvere tale problema di minimo vincolato è equivalente a risolvere

$$\min_x \frac{1}{2} (\|r(x)\|_2^2 + \lambda \|x\|_2^2) \quad \lambda \in (0, 1) \quad (2.40)$$

Il problema (2.40) è regolarizzato mediante il termine aggiunto che è detto di *regolarizzazione alla Tikhonov* in quanto, scegliendo un opportuno valore di λ , è possibile compensare il mal condizionamento della matrice Jacobiana J .

Il comportamento tipico della soluzione al variare della scelta di λ è il seguente:

- $\lambda \rightarrow 1$ il problema (2.40) risulta meglio condizionato ma la soluzione si allontana dalla soluzione del problema originale.
- $\lambda \rightarrow 0$ la soluzione di (2.40) tende alla soluzione di (2.28) ma il problema tende ad essere mal condizionato come il problema originale.

Metodo di Levenberg-Marquardt

Il metodo di Levenberg-Marquardt può essere visto come una modifica del metodo di Gauss-Newton in cui la matrice Hessiana $\nabla^2 F(x)$ è approssimata con $J(x)^T J(x)$.

Essendo la matrice $J_k^T J_k$ molto mal condizionata, la si regolarizza tramite il parametro λ_k sfruttando la regolarizzazione di Tikhonov precedentemente descritta.

Pertanto il metodo di Levenberg-Marquardt implementa iterativamente la direzione d risolvendo

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} (\|J_k d + r_k\|_2^2 + \lambda \|d\|_2^2) \quad (2.41)$$

ovvero

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} \left\| \begin{bmatrix} J_k d + r_k \\ \lambda_k I_d + 0 \end{bmatrix} \right\|_2^2 = \min_{d \in \mathbb{R}^n} \frac{1}{2} \left\| \begin{bmatrix} J_k \\ \sqrt{\lambda_k} I_d \end{bmatrix} d + \begin{bmatrix} r_k \\ 0 \end{bmatrix} \right\|_2^2 \quad (2.42)$$

e poi

$$x_{k+1} = x_k + \alpha d_k.$$

Chiamando

$$J_k(d) = \|J_k d + r_k\|_2^2 + \lambda_k \|d\|_2^2, \quad (2.43)$$

si procede dunque con la risoluzione di

$$\min_d J_k(d),$$

la quale è data dal metodo delle equazioni normali.

Si ha quindi

$$J_k(d) = (J_k d + r_k)^T (J_k d + r_k) + \lambda_k d^T d$$

$$\nabla J_k(d) = J_k^T (J_k d + r_k) + \lambda_k d$$

Ponendo il gradiente uguale a zero

$$\nabla J_k(d) = 0$$

si ottiene infine

$$(J_k^T J_k + \lambda_k I)d = -J_k^T r_k. \quad (2.44)$$

L'algoritmo *Levenberg-Marquardt* é dunque un metodo di discesa, la cui direzione di discesa d é soluzione del sistema lineare (2.44).

Si noti che la (2.44) é la forma regolarizzata della (2.35) con in aggiunta il termine regolarizzatore di Tikhonov.

Si ottiene quindi la trasformazione di un problema non lineare in una minimizzazione di un funzionale lineare.

Proprio come nel caso Gauss-Newton, l'equivalenza tra (2.42) e (2.44) fornisce un modo per risolvere il sotto problema senza calcolare il prodotto matrice-matrice $J^T J$ e la sua fattorizzazione di Cholesky [12].

Un'altra interpretazione del metodo potrebbe essere quella di vedere il termine $\lambda_k I$ come un'approssimazione del termine $S(x_k)$ che era stato trascurato nel metodo di Gauss-Newton. A tal proposito, il metodo di Levenberg-Marquardt viene inquadrato nella classe dei metodi di ottimizzazione di secondo grado, o di tipo Newton, che prevedono l'utilizzo della matrice Hessiana, o come in questo caso, di una sua approssimazione.

Al variare di λ_k si otterrà una soluzione d_k vicina alla soluzione ottenuta con Gauss-Newton se λ_k é vicino a zero; al crescere di λ_k , invece, il sistema lineare da risolvere risulta meglio condizionato e la soluzione risulta sempre piú lontana da quella ottenuta con Gauss-Newton.

Il problema (2.44) puó essere descritto e analizzato tramite il metodo dei minimi quadrati noto come *trust region reflective*.

In effetti, il metodo Levenberg-Marquardt é talvolta considerato il progenitore dell'approccio *trust region* per l'ottimizzazione generale non vincolata [12].

Nel metodo *trust region* ciò che si fa é aggiustare la regione di ricerca, quindi ad ogni passo k , per evitare che il vettore d_k sia troppo grande o abbia una direzione che non assicuri una sufficiente riduzione della funzione obiettivo, si aggiunge un vincolo sulla norma di d_k e invece di risolvere il problema (2.35), per il calcolo di d_k il problema secondario da risolvere ad ogni iterazione é

$$\min_{d \in \mathbb{R}^n} \|J_k d + r_k\|_2^2, \quad \text{tale che} \quad \|d\| \leq \Delta_k, \quad (2.45)$$

con $\Delta_k > 0$ sfera di trust region e J_k matrice Jacobiana.

Il problema (2.44) é verificato nel lemma seguente [12]:

Lemma

Il vettore d^{LM} é una soluzione del sotto problema trust region (2.45), se e solo se d^{LM} é ammissibile e c'è uno scalare $\lambda \geq 0$ tale che,

$$\begin{aligned} (J^T J + \lambda I) d^{LM} &= -J^T r, \\ \lambda(\Delta - \|d^{LM}\|) &= 0. \end{aligned} \quad (2.46)$$

Capitolo 3

Casi di studio

In questo capitolo verrà presentata l'analisi sperimentale relativa a due modelli epidemiologici:

- il modello epidemiologico SIR;
- il modello epidemiologico SIDARTHE.

Il primo é un modello generico di base, che successivamente viene adattato alla epidemia in corso Covid-19, mentre nel secondo vengono direttamente considerati parametri specifici per l'epidemia di Covid-19.

I modelli epidemiologici sono modelli matematici compartimentali che descrivono la diffusione di un' infezione; sono detti compartimentali poiché dividono la popolazione in gruppi (compartimenti) e descrivono i cambiamenti di tali gruppi nel tempo. Ad ogni compartimento generalmente é associata un' equazione che ne descrive la dinamica.

I modelli epidemiologici inoltre dipendono da parametri che tengono conto delle caratteristiche del virus e della popolazione all' interno della quale si verifica l'epidemia.

In alcuni casi questi modelli sono molto complessi ed é necessario suddividere la popolazione in molti compartimenti per poter spiegare la diffusione del virus.

La costruzione dei modelli puó avere diversi scopi, ad esempio, prevedere l'evoluzione della pandemia, oppure prevedere i cambiamenti dopo l'introduzione di una

vaccinazione, o, piú in generale, studiare come cambia la diffusione del virus quando viene introdotto un fattore esterno.

Per entrambi i modelli viene scelto di monitorare l'epidemia nella regione Emilia Romagna, i dati sperimentali epidemiologici sono reperiti dal repository open source (Github) gestito dal Dipartimento della Protezione Civile italiano che utilizza i dati ufficiali forniti dal Ministero della Salute [5].

Il database é suddiviso in cartelle, una di queste é denominata "dati-regioni", che al suo interno contiene i dati epidemiologici regionali. In particolare ci sono dei file .csv giornalieri che raccolgono i dati di un solo giorno, e un file .csv cumulativo denominato "dpc-covid19-ita-regioni.csv" [11].

All'interno di ogni file i dati sono strutturati in 16 campi, riportati nella seguente tabella.

Tabella 3.1: Struttura dei dati contenuti all'interno del file 'dpc-covid19-ita-regioni.csv', all'interno della cartella 'dati-regioni' del database ufficiale di Github gestito dalla Protezione Civile.

Nome campo	Descrizione	Formato
data	Data dell'informazione	YYYY-MM-DDTHH:MM:SS (ISO 8601) Ora italiana
stato	Codice a 3 lettere dello stato	XYZ (ISO 3166-1 alpha-3)
codice_regione	Codice a 2 cifre della regione (ISTAT 2019)	Numero intero
denominazione_regione	Denominazione della Regione	Testo
lat	Latitudine	Numero decimale (WGS84)
long	Longitudine	Numero decimale (WGS84)

Continua nella pagina successiva

Tabella 3.1 - *Continuo della pagina precedente*

ricoverati_con_sintomi	Pazienti ricoverati con sintomi	Numero intero
terapia_intensiva	Pazienti ricoverati in terapia intensiva	Numero intero
totale_ospedalizzati	Totale pazienti ospedalizzati (ricoverati con sintomi + ricoverati in terapia intensiva)	Numero intero
isolamento_domiciliare	Persone in isolamento domiciliare	Numero intero
totale_positivi	Totale attualmente positivi (totale ospedalizzati + isolamento domiciliare)	Numero intero
variazione_totale_positivi	Variazione del totale positivi (totale positivi giorno corrente - totale positivi giorno precedente)	Numero intero
nuovi_positivi	Nuovi attualmente positivi (totale casi giorno corrente - totale casi giorno precedente)	Numero intero
dimessi_guariti	Persone dimesse guarite	Numero intero
deceduti	Persone decedute	Numero intero
casi_da_sospetto_diagnostico	Casi positivi al tampone emersi da attività clinica	Numero intero
casi_da_screening	Casi positivi emersi da indagini e test, pianificati a livello nazionale o regionale	Numero intero
totale_casi	Totale casi positivi	Numero intero
tamponi	Numero di tamponi effettuati	Numero intero
casi_testati	Totale dei soggetti sottoposti al test diagnostico	Numero intero
note	Note	Testo

Nella Tabella 3.1 le voci in grassetto sono quelle utilizzate nei modelli trattati in seguito. Oltre alle voci che determinano i vari compartimenti caratterizzanti i modelli epidemiologici SIR e SIDARTHE, nel database della Protezione Civile viene prelevata anche la voce 'tamponi' che rappresenta il numero di tamponi effettuati nella regione Emilia Romagna durante il periodo che va dal 1 Settembre 2020 al 29 Novembre 2020 (90 giorni).

Di seguito viene riportato l'andamento nel tempo:

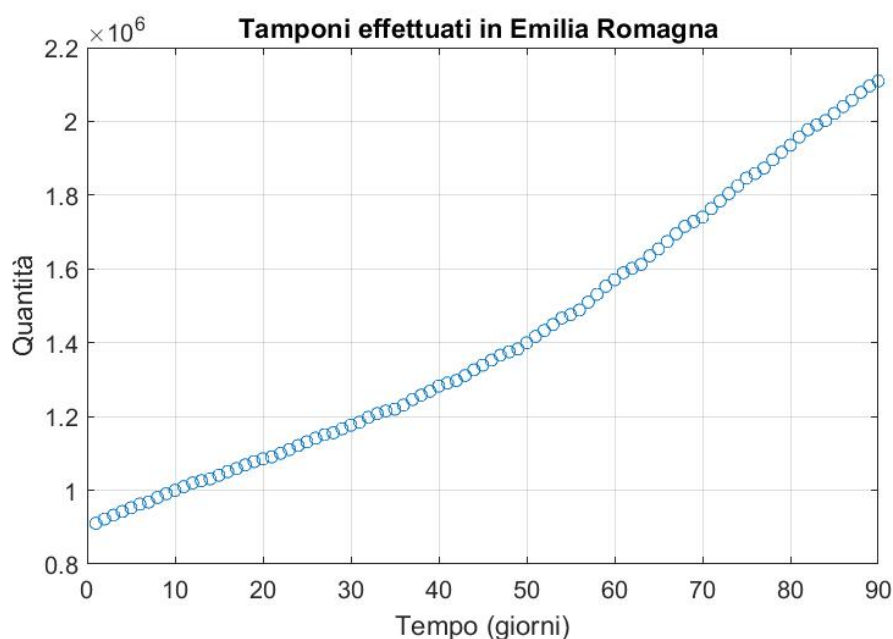


Figura 3.1: Andamento tamponi effettuati in Emilia Romagna dal 1 Settembre al 29 Novembre 2020.

3.1 Modello SIR

Il modello epidemiologico piú semplice é il modello SIR, nel quale gli individui sono categorizzati in tre compartimenti:

- S : l'insieme degli individui *Suscettibili*, ovvero coloro che non sono entrati in contatto con il virus.

- I : l'insieme degli individui *Infetti*, ovvero coloro che hanno contratto il virus.
- R : l'insieme degli individui *Rimossi*, ovvero coloro che non sono piú infetti o perché guariti dall'infezione o perché deceduti.

Il modello SIR descrive la dinamica con cui un individuo si sposta da un compartimento ad un altro ossia come varia il numero degli individui nei compartimenti S, I ed R al variare del tempo. La transizione tra i compartimenti puó avvenire in una sola direzione, di seguito é riportato un diagramma che mostra visivamente il modello [11]:



Figura 3.2: Schema a compartimenti del modello SIR.

Il modello é descritto dal seguente sistema di equazioni differenziali ordinarie (IVP):

$$\begin{cases} S'(t) = -\beta SI \\ I'(t) = -\beta SI - \gamma I \\ R'(t) = \gamma I \end{cases} \quad (3.1)$$

con condizioni iniziali:

$$\begin{cases} S(0) = S_0 \\ I(0) = I_0 \\ R(0) = R_0 \end{cases} \quad (3.2)$$

dove β e γ sono i parametri del modello che rappresentano il tasso di trasmissione e il tasso di rimozione sotto l'ipotesi di popolazione omogenea e chiusa, ossia non si considera la variazione della demografia [13].

Questa semplificazione ha senso se viene considerato un periodo relativamente breve in cui i cambiamenti demografici non sono sostanziali.

In generale vale sempre la relazione $S + I + R = 1$ (oppure $S + I + R = N$ se non si considera la porzione di individui ma il numero esatto), quindi spesso una delle tre equazioni precedenti viene omessa e il compartimento relativo viene dedotto utilizzando quest'ultima relazione [11].

Il modello SIR diventa:

$$\begin{cases} S'(t) = -\beta SI \\ I'(t) = -\beta SI - \gamma I \end{cases} \quad (3.3)$$

con condizioni iniziali:

$$\begin{cases} S(0) = S_0 \\ I(0) = I_0 \end{cases} \quad (3.4)$$

mentre il compartimento R viene determinato di conseguenza dalla relazione:

$$R = N - S - I.$$

3.1.1 Implementazione modello SIR

Dopo aver descritto teoricamente il problema diretto e quello inverso nel capitolo 2, si procede ora alla loro implementazione e validazione con dati reali.

Inizialmente si procede con la risoluzione del *problema diretto* relativo ad un'epidemia generica, ovvero si determinano i compartimenti S,I,R in un intervallo temporale di 30 giorni, risolvendo il sistema di ODE del modello SIR (3.3).

Si considera la popolazione unitaria ($N=1$), il numero iniziale di suscettibili é $S_0= 1$ mentre il numero iniziale di infetti é $I_0 = 10^{-6}$. I parametri scelti sono $\beta = 1.4247$, $\gamma = 0.14286$.

Il sistema di equazioni differenziali IVP viene risolto con il metodo di Runge Kutta esplicito a passo adattivo di ordine 4 e 5 (mediante la funzione `ode45` di `Matlab`). La figura di seguito mostra i risultati:

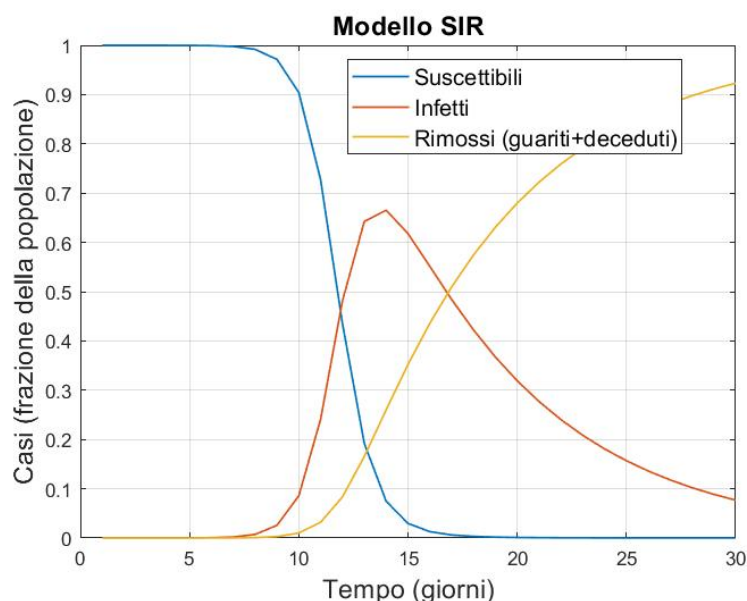


Figura 3.3: Andamento dei compartimenti S, I, R nell’arco di 30 giorni relativi ad una epidemia generica.

Le curve dei tre compartimenti in figura 3.3 sono state determinate in un intervallo temporale di 30 giorni.

Si nota che dopo circa 5 giorni il numero degli infetti inizia a crescere fino a raggiungere un valore massimo per poi decrescere, mentre quello degli individui suscettibili diminuisce considerevolmente e quindi l’epidemia é in atto [11].

Successivamente si passa all’analisi dell’evoluzione temporale dell’epidemia COVID-19 relativa alla regione Emilia Romagna per il periodo che va dal 1 Settembre 2020 al 29 Novembre 2020 (90 giorni).

I dati sperimentali vengono reperiti dal database della Protezione Civile descritto precedentemente: il compartimento dei *Rimossi* é l’insieme degli individui che non sono piú infetti o perché guariti dall’infezione o perché deceduti. Nel caso del COVID-19 questi dati corrispondono alla somma dei casi riportati in Tabella 3.1 sotto la voce ‘dimessi_guariti’ e ‘deceduti’.

Analogamente, poiché il compartimento degli *Infetti* é l’insieme degli individui che hanno contratto il virus, detti impropriamente positivi (poiché positivi al te-

st diagnostico), i dati sono quelli riportati nel campo 'totale_positivi'. Infine il compartimento dei *Suscettibili* viene ricavato come $S = N - R - I$.

Si procede preliminarmente con l'implementazione del *problema inverso*, il quale consiste nell'identificazione dei parametri $\Lambda^* = [\beta^*, \gamma^*] \in \mathbb{R}^2$ sulla base dei dati reali disponibili. Solitamente le condizioni iniziali ovvero I_0 e R_0 si possono conoscere, ma i parametri β^* e γ^* del modello non sono generalmente noti.

Il problema interessante é dunque inverso: a partire da alcuni dati sperimentali dei compartimenti, si vogliono stimare i parametri Λ^* in modo che le curve individuate dal modello SIR approssimino nel miglior modo i dati già noti.

Questo tipo di problema é detto *identificazione dei parametri*. Una volta individuati i parametri β^* e γ^* si procede alla risoluzione del problema diretto e la si confronta con i dati sperimentali.

Definiamo il vettore $\Lambda = [\beta, \gamma] \in \mathbb{R}^2$ come il vettore dei parametri del modello.

Sia $\bar{y} = [\bar{S}, \bar{I}, \bar{R}]$ il vettore dei dati sperimentali dei compartimenti dedotto dai dati in Tabella 3.1 e sia $Y_m(t) = [S_m(t), I_m(t), R_m(t)]$ il vettore dei compartimenti teorici determinati risolvendo il problema diretto (3.3) in funzione dei parametri $\Lambda = [\beta, \gamma]$.

Si deve perciò risolvere il seguente problema di minimizzazione:

$$\Lambda^* \in \underset{\Lambda \in \mathbb{R}^2}{\operatorname{argmin}} \|\bar{y} - f(Y; \Lambda)\|_2^2. \quad (3.5)$$

dove $f(Y; \Lambda)$ rappresenta il modello differenziale considerato (3.3).

Come già descritto nel paragrafo 2.2.1, é un problema di minimi quadrati non lineare e non vincolato che pertanto viene risolto tramite il metodo di Levenberg - Marquardt.

I dati teorici $Y_m(t) = [S_m(t), I_m(t), R_m(t)]$ sono ricavati risolvendo il sistema di equazioni differenziali ODE tramite la funzione 'ode45' di `Matlab` che fa uso del metodo Runge-Kutta.

Si ottengono dunque i parametri $\Lambda^* = [\beta^*, \gamma^*]$, soluzioni del problema inverso. In particolare:

- Tasso di trasmissione $\beta^* = 0.043297$;

- Tasso di rimozione $\gamma^* = 0.006599$.

Dopo aver raccolto i dati sperimentali dal repository e risolto il problema inverso (3.5), si passa al *problema diretto* in cui si risolve il sistema di equazioni differenziali (3.3) utilizzando i parametri ottimali β^* e γ^* appena determinati, e si ricavano i compartimenti $Y_m(t) = [S_m(t), I_m(t), R_m(t)]$.

Di seguito si riportano i risultati in cui è possibile vedere che i dati teorici seguono fedelmente quelli reali.

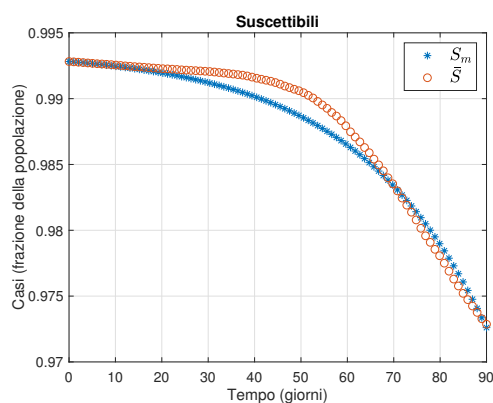


Figura 3.4: Andamento compartimento 'Suscettibili' nell'arco di 90 giorni.

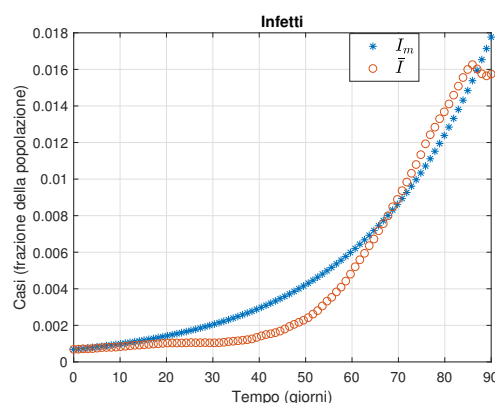


Figura 3.5: Andamento compartimento 'Infetti' nell'arco di 90 giorni.

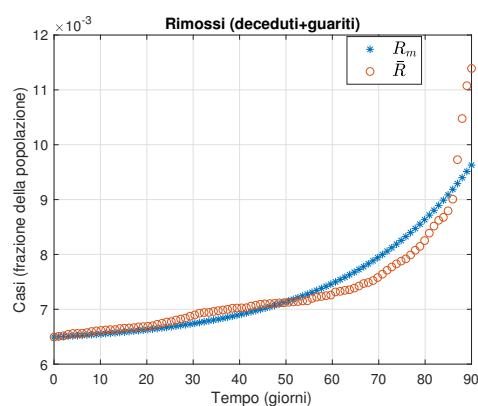


Figura 3.6: Andamento compartimento 'Rimossi' nell'arco di 90 giorni.

3.2 Modello SIDARTHE

Modelli più complessi di quello SIR possono rappresentare con maggior precisione la dinamica della diffusione di epidemie specifiche.

Il modello presentato di seguito, denominato SIDARTHE, discrimina tra casi di infezione rilevati e non rilevati, tra diverse gravità della malattia (SOI: severity of illness) quindi tra casi non pericolosi per la vita (asintomatici e paucisintomatici; infezione lieve e moderata) e casi potenzialmente pericolosi per la vita (maggiore ed estrema) che richiedono l'ammissione in terapia intensiva.[4]

Il modello considera otto stadi di infezione:

- *S*: suscettibile;

- *I*: infetto (infetto asintomatico non rilevato);

- *D*: diagnosticato (infetto asintomatico rilevato);

- *A*: malato (infetto sintomatico non rilevato);

- *R*: riconosciuto (infetto sintomatico rilevato);

- *T*: minacciato (infettato da sintomi potenzialmente letali - rilevato);

- *H*: guarito;

- *E*: estinto.

Le interazioni tra queste fasi sono mostrate nella figura seguente.

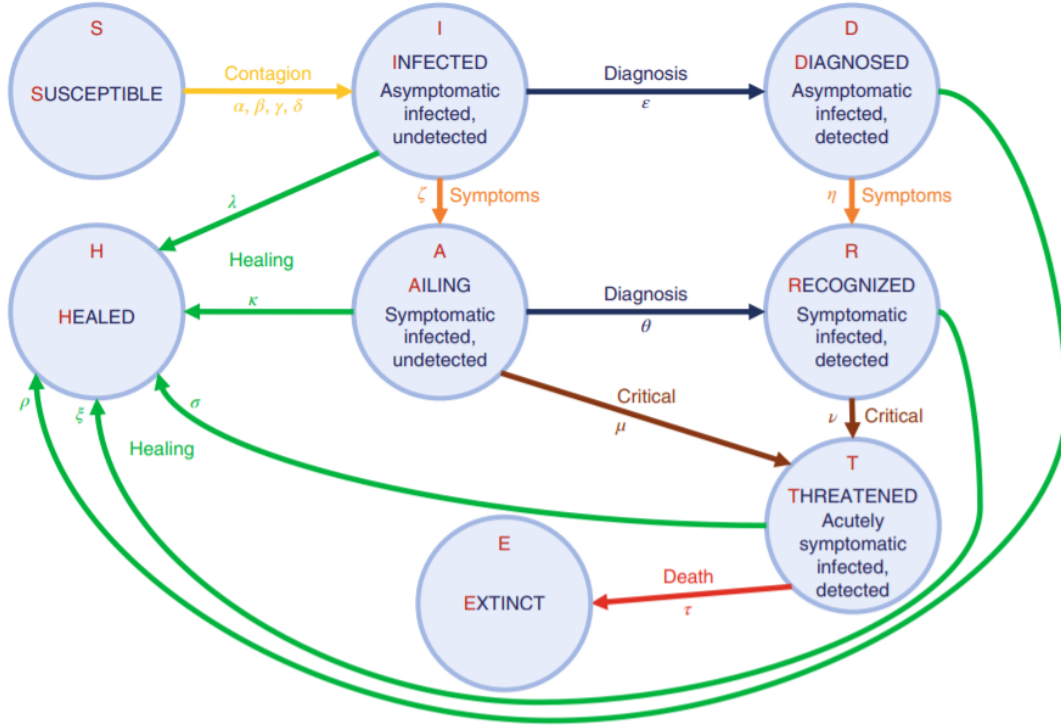


Figura 3.7: Schema a compartimenti del modello SIDARTHE.

Il modello é descritto dal seguente sistema di equazioni differenziali ordinarie (IVP):

$$\begin{cases} S'(t) = -S(t)(\alpha I(t) + \beta D(t) + \gamma A(t) + \delta R(t)) \\ I'(t) = S(t)(\alpha I(t) + \beta D(t) + \gamma A(t) + \delta R(t)) - (\epsilon + \zeta + \lambda)I(t) \\ D'(t) = \epsilon I(t) - (\eta + \rho)D(t) \\ A'(t) = \zeta I(t) - (\theta + \mu + \kappa)A(t) \\ R'(t) = \eta D(t) + \theta A(t) - (\nu + \xi)R(t) \\ T'(t) = \mu A(t) + \nu R(t) - (\sigma + \tau)T(t) \\ H'(t) = \lambda I(t) + \rho D(t) + \kappa A(t) + \xi R(t) + \sigma T(t) \\ E'(t) = \tau T(t) \end{cases} \quad (3.6)$$

con condizioni iniziali:

$$\left\{ \begin{array}{l} S(0) = S_0 \\ I(0) = I_0 \\ D(0) = D_0 \\ A(0) = A_0 \\ R(0) = R_0 \\ T(0) = T_0 \\ H(0) = H_0 \\ E(0) = E_0 \end{array} \right. \quad (3.7)$$

Nel sistema di equazioni differenziali (3.6) le lettere latine maiuscole (variabili di stato) rappresentano la frazione di popolazione in ogni fase, mentre tutti i parametri considerati, indicati con lettere greche, sono numeri non negativi. [4]

I parametri sono definiti come segue:

- α, β, γ e δ indicano rispettivamente la velocità di trasmissione (la probabilità di trasmissione della malattia in un singolo contatto moltiplicata per il numero medio di contatti per persona) dovuta ai contatti tra un soggetto suscettibile (S) e un infetto asintomatico non rilevato (I), un diagnosticato asintomatico rilevato (D), un malato sintomatico non rilevato (A) o un soggetto riconosciuto sintomatico rilevato (R). Tipicamente, α è maggiore di γ (assumendo che le persone tendano ad evitare contatti con soggetti che mostrano sintomi, anche se la diagnosi non è stata ancora fatta), a sua volta γ è maggiore di β e δ (assumendo che i soggetti a cui è stata diagnosticata siano adeguatamente isolati). Questi parametri possono essere modificati da politiche di allontanamento sociale (ad esempio, chiusura di scuole, lavoro a distanza, blocco). Il rischio di contagio dovuto a soggetti minacciati, trattati in apposite unità di terapia intensiva, è considerato trascurabile.
- ϵ e θ indicano il tasso di probabilità di rilevamento della malattia, relativo rispettivamente ai casi asintomatici e sintomatici. Questi parametri, anche

modificabili, riflettono il livello di attenzione sulla malattia e il numero di test effettuati sulla popolazione: possono essere aumentati imponendo una massiccia campagna di tracciamento dei contatti e test. Si noti che θ é tipicamente maggiore di ϵ , poiché é piú probabile che un individuo sintomatico venga testato.

- ζ e η indicano il tasso di peggioramento: l'infetto asintomatico non rilevato e rilevato diventa sintomatico. É dunque il tasso di probabilità con cui un soggetto infetto, rispettivamente non consapevole e consapevole di essere infetto, sviluppa sintomi clinicamente rilevanti e sono comparabili in assenza di un trattamento specifico. Questi parametri dipendono dalla malattia, ma possono essere parzialmente ridotti da terapie migliori e dall'acquisizione dell'immunità contro il virus.
- μ e ν indicano rispettivamente la velocità con cui soggetti infetti non rilevati e rilevati sviluppano sintomi potenzialmente letali; sono comparabili se non esiste un trattamento specifico noto che sia efficace contro la malattia, altrimenti μ potrebbe essere maggiore. Al contrario, ν può essere maggiore perché é piú probabile che siano stati diagnosticati individui infetti con sintomi piú acuti, che hanno un rischio maggiore di peggioramento. Questi parametri possono essere ridotti mediante migliori terapie e acquisizione dell'immunità contro il virus.
- τ indica il tasso di mortalità (per soggetti infetti con sintomi potenzialmente letali) e può essere ridotto mediante migliori terapie.
- λ , κ , ξ , ρ e σ indicano la velocità di recupero per le cinque classi di soggetti infetti; possono differire in modo significativo se un trattamento appropriato per la malattia é noto e adottato per i pazienti diagnosticati, altrimenti sono probabilmente paragonabili. Questi parametri possono essere aumentati grazie a migliori trattamenti e all'acquisizione dell'immunità contro il virus.

In generale vale sempre la relazione $S + I + D + A + R + T + H + E = 1$ (oppure $S + I + D + A + R + T + H + E = N$ se non si considera la porzione di individui ma

il numero esatto), quindi i dati sperimentali e di conseguenza il dato iniziale relativi al compartimento dei *Suscettibili* sono dedotti utilizzando quest'ultima relazione.

Bisogna precisare inoltre che le variabili I (infetto asintomatico non rilevato) e A (infetto sintomatico non rilevato) non sono presenti nel database della Protezione Civile perché appunto non registrati; quindi verrà assegnato loro solo il valore iniziale relativo alle condizioni iniziali per implementare il modello SIDARTHE, non si avrà l'andamento nel tempo dei dati sperimentali, ma solo dei dati teorici.

3.2.1 Implementazione modello SIDARTHE

Analogamente a quanto fatto per il modello SIR, anche nel modello SIDARTHE bisogna risolvere i problemi diretto ed inverso.

I vari passaggi che seguiranno quindi sono:

- Risoluzione del *problema inverso* (FASE 1), ossia partendo dalla conoscenza delle condizioni iniziali delle variabili e partendo dalla conoscenza dei parametri iniziali (FASE 0), si risolve il problema di minimi quadrati non lineari per identificare i parametri $\Lambda^* = [\alpha^*, \beta^*, \gamma^*, \delta^*, \epsilon^*, \zeta^*, \eta^*, \theta^*, \kappa^*, \lambda^*, \mu^*, \nu^*, \xi^*, \rho^*, \sigma^*, \tau^*]$ ottimali;
- Risoluzione del *problema diretto* (FASE 2), ossia risoluzione del sistema (3.6), il quale utilizza i parametri ottimali trovati nel problema inverso, per raffigurare l'andamento nel tempo dei compartimenti $S_m, I_m, D_m, A_m, R_m, T_m, H_m, E_m$;
- Validazione andamento del modello teorico con dati sperimentali, quindi ci si aspetta che il modello descritto dai compartimenti $S_m, I_m, D_m, A_m, R_m, T_m, H_m, E_m$ approssimi nel miglior modo i dati sperimentali prelevati dal database della Protezione Civile relativi alla regione Emilia Romagna per il periodo che va dal 1 Settembre 2020 al 29 Novembre 2020 (90 giorni).

Sia $\bar{y} = [\bar{S}, \bar{I}, \bar{D}, \bar{A}, \bar{R}, \bar{T}, \bar{H}, \bar{E}]$ il vettore dei dati sperimentali dei compartimenti e sia $Y_m(t) = [S_m(t), I_m(t), D_m(t), A_m(t), R_m(t), T_m(t), H_m(t), E_m(t)]$ il vettore dei compartimenti teorici in funzione dei 16 parametri sopra elencati.

I dati sperimentali prelevati dal database della Protezione Civile vengono normalizzati rispetto alla popolazione dell'Emilia Romagna per far sí che i risultati vengano espressi in percentuale.

Definiamo il vettore $\Lambda = [\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \theta, \kappa, \lambda, \mu, \nu, \xi, \rho, \sigma, \tau] \in \mathbb{R}^{16}$ come il vettore dei parametri del modello.

Si ha il seguente problema di minimizzazione non lineare:

$$\Lambda^* \in \operatorname{argmin}_{\Lambda \in \mathbb{R}^{16}} \|\bar{y} - f(Y; \Lambda)\|_2^2. \quad (3.8)$$

dove $f(Y; \Lambda)$ rappresenta il modello differenziale considerato (3.6). Come già descritto, é un problema di minimi quadrati non lineare e non vincolato.

Prima di risolvere il problema di identificazione dei parametri é necessario determinare i parametri iniziali poiché non si conoscono a priori.

Quindi la FASE 0 é caratterizzata dai seguenti step ripetuti per 100 volte:

- Definire un vettore random Λ di 16 parametri iniziali;
- Risolvere il problema diretto (3.6) con il vettore Λ e ottenere quindi i compartimenti teorici $Y_m = [S_m, I_m, D_m, A_m, R_m, T_m, H_m, E_m]$;
- Valutare l'errore E definito come:

$$E = \|S_m - \bar{S}\|_2 + \|D_m - \bar{D}\|_2 + \|R_m - \bar{R}\|_2 + \|T_m - \bar{T}\|_2 + \|H_m - \bar{H}\|_2 + \|E_m - \bar{E}\|_2 \quad (3.9)$$

relativo ai 90 giorni e alle variabili del modello SIDARTHE su cui si ha un controllo, escludendo quindi la variabile I e A poiché fanno riferimento a valori non rilevati. Per le sei variabili in gioco, l'errore é pari alla norma euclidea della differenza tra il dato teorico e quello sperimentale, si ricava quindi un vettore errore $E \in \mathbb{R}^{540}$;

- Salvare il minimo errore E calcolato.

Dopo aver ricavato i parametri iniziali Λ si passa alla risoluzione del *problema inverso* (3.8) e quindi all'*identificazione dei parametri*. Si procede come già fatto per il modello SIR, quindi il problema (3.8) é un problema di minimi quadrati non lineare risolto con il metodo di Levenberg - Marquardt *trust region reflective*.

I dati sperimentali vengono presi dal database della Protezione Civile descritto precedentemente (si veda Tabella 3.1): il compartimento dei *Diagnosticati* (D) ossia infetti asintomatici rilevati é l'insieme degli individui che si trovano in isolamento domiciliare, quindi i dati sono quelli riportati nel campo "isolamento_domiciliare";

il compartimento dei *Riconosciuti* (R) ossia infetti sintomatici rilevati é l'insieme degli individui che sono ricoverati con sintomi, quindi i dati sono quelli riportati nel campo "ricoverati_con_sintomi";

il compartimento dei *Minacciati* (T) ossia infettati da sintomi potenzialmente letali é l'insieme degli individui che si trovano in terapia intensiva, quindi i dati sono quelli riportati nel campo "terapia_intensiva";

il compartimento dei *Guariti* (H) é l'insieme degli individui che sono dimessi perché guariti, quindi i dati sono quelli riportati nel campo "dimessi_guariti";

il compartimento dei *Deceduti* (E) contiene i dati riportati nel campo "decaduti".

Infine il compartimento dei *Suscettibili* (S) viene ricavato come

$$S = N - I - D - A - R - T - H - E.$$

Le condizioni iniziali relative al sistema di equazioni differenziali IVP (3.6) sono le seguenti:

$$\left\{ \begin{array}{l} I(0) = I_0 = \bar{I}, \\ D(0) = D_0 = \bar{D}(1), \\ A(0) = A_0 = \bar{A}, \\ R(0) = R_0 = \bar{R}(1), \\ T(0) = T_0 = \bar{T}(1), \\ H(0) = H_0 = \bar{H}(1), \\ E(0) = E_0 = \bar{E}(1), \\ S(0) = S_0 = N - I_0 - D_0 - A_0 - R_0 - T_0 - H_0 - E_0. \end{array} \right. \quad (3.10)$$

Le condizioni iniziali riguardanti le variabili D,R,T,H,E si riferiscono al primo valore relativo al giorno 1 Settembre 2020 normalizzato rispetto alla popolazione dell'Emilia Romagna, mentre le variabili I e A, non essendo registrati nel repository della Protezione Civile sono assunti pari a: $I_0 = 3.2511e-06$, $A_0 = 1.1211e-07$, coerentemente col fatto che le condizioni iniziali delle variabili I e A normalizzate rispetto all'intera popolazione italiana fossero : $I_0 = 8.2625e-03$, $A_0 = 4.1312e-05$. [4]

Al problema di minimi quadrati non lineare si aggiungono dei vincoli di bound per i parametri: $lb \leq \Lambda \leq ub$ dove i limiti inferiore e superiore sono almeno $lb = 0_{16}$ e $ub = 1_{16}$.

Questo porta il metodo di Levenberg-Marquardt ad essere descritto ed analizzato tramite il metodo dei minimi quadrati noto come *Trust region reflective*, descritto precedentemente.

Si ottiene Λ^* ovvero i 16 parametri, soluzione del problema inverso. In particolare

- tasso di trasmissione suscettibile - infetto asint. non rilevato (I) $\alpha^*=0.789416$
- tasso di trasmissione suscettibile - infetto asint. rilevato (D) $\beta^*=0.002207$
- tasso di trasmissione suscettibile - infetto sint. non rilevato (A) $\gamma^*=0.599709$
- tasso di trasmissione suscettibile - infetto sint. rilevato (R) $\delta^*=0.002296$
- tasso probabilità rilevamento asintomatici $\epsilon^*=0.668547$
- tasso peggioramento infetto non rilevato $\zeta^*=0.120813$
- tasso peggioramento infetto rilevato $\eta^*=0.000125$
- tasso probabilità rilevamento sintomatici $\theta^*=0.751369$
- tasso di guarigione per infetto sintomatico non rilevato (A) $\kappa^*=0.525305$
- tasso di guarigione per infetto asintomatico non rilevato (I) $\lambda^*=0.014496$
- tasso sviluppo sintomi letali soggetto infetto non rilevato $\mu^*=0.605672$
- tasso sviluppo sintomi letali soggetto infetto rilevato $\nu^*=0.000004$

- tasso di guarigione per infetto sintomatico rilevato (R) $\xi^*=0.060269$
- tasso di guarigione per infetto asintomatico rilevato (D) $\rho^*=0.000313$
- tasso di guarigione per infettato da sintomi gravi (T) $\sigma^*=0.591456$
- tasso di mortalità $\tau^*=0.098084$

Dopo aver risolto il problema inverso si passa all'implementazione del *problema diretto*, nel quale si risolve il sistema di equazioni differenziali IVP (3.6) utilizzando i parametri ottimali appena determinati, e si ricavano quindi i compartimenti $Y_m(t) = [S_m(t), I_m(t), D_m(t), A_m(t), R_m(t), T_m(t), H_m(t), E_m(t)]$.

Di seguito si riportano i risultati del modello implementato in cui é possibile visualizzare, per le diverse variabili, i dati teorici con quelli sperimentali, e come i dati teorici seguano abbastanza fedelmente i dati reali. Per quanto riguarda invece le variabili I e A si visualizzano solo i dati teorici perché come già detto non si hanno dati registrati nel repository della Protezione Civile.

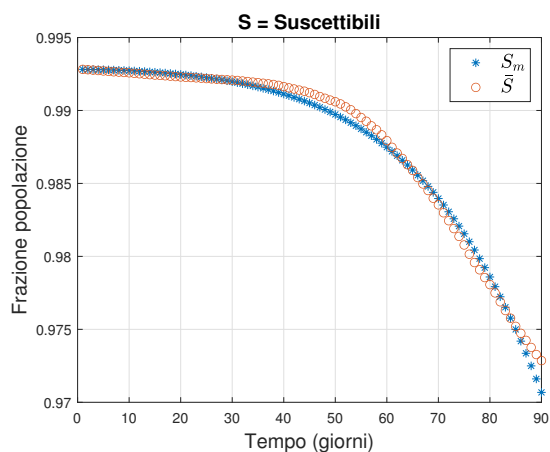


Figura 3.8: Andamento compartimento "S" nell'arco di 90 giorni

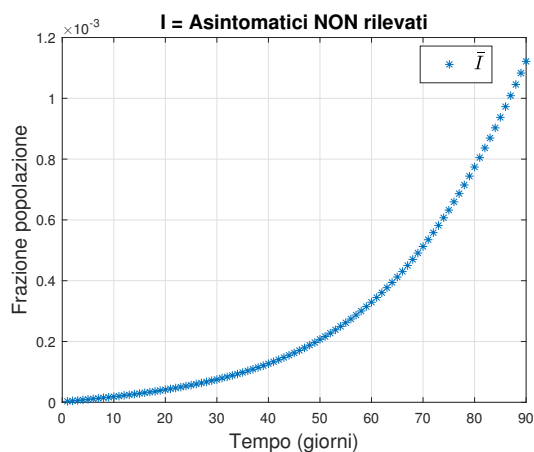


Figura 3.9: Andamento compartimento "I" nell'arco di 90 giorni

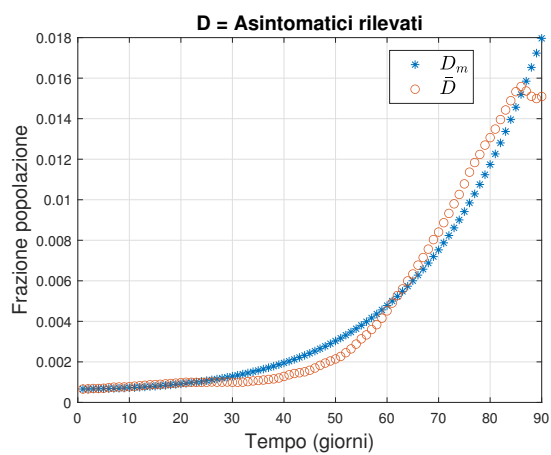


Figura 3.10: Andamento compartimento "D" nell'arco di 90 giorni

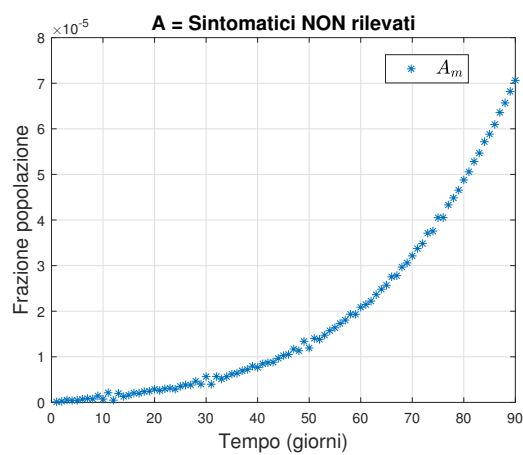


Figura 3.11: Andamento compartimento "A" nell'arco di 90 giorni

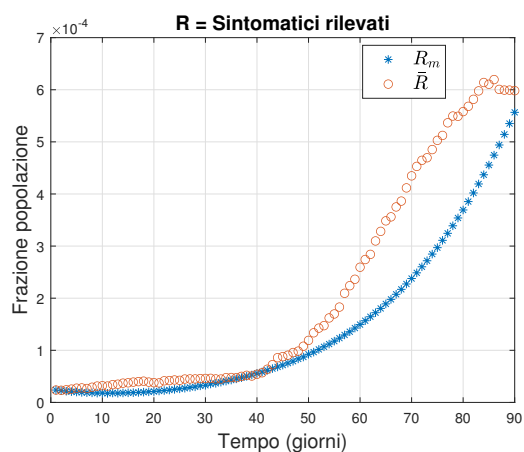


Figura 3.12: Andamento compartimento "R" nell'arco di 90 giorni

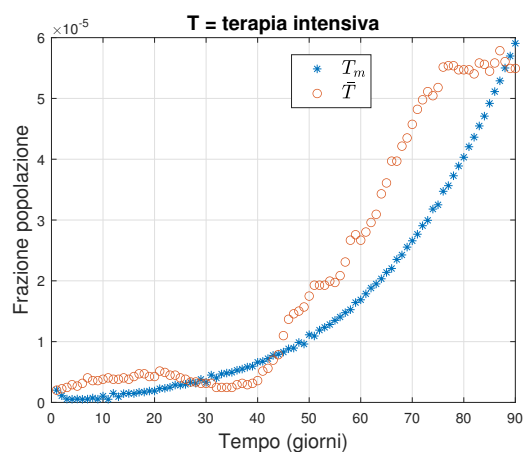


Figura 3.13: Andamento compartimento "T" nell'arco di 90 giorni

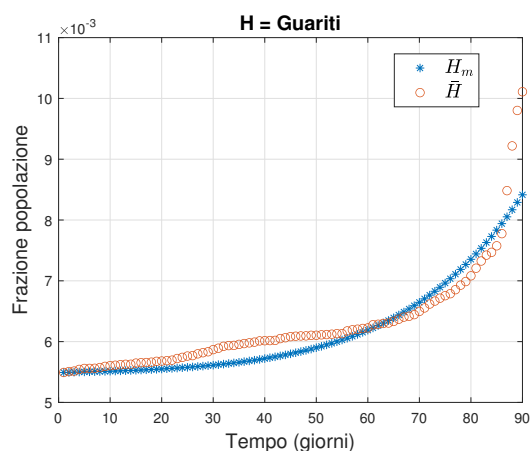


Figura 3.14: Andamento compartimento "H" nell'arco di 90 giorni

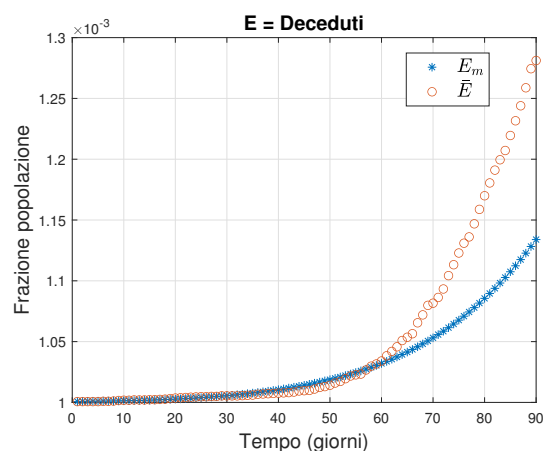


Figura 3.15: Andamento compartimento "E" nell'arco di 90 giorni

3.2.2 Suddivisione del periodo Settembre - Novembre 2020

Gli andamenti delle variabili rappresentati dalla Fig 3.8 alla Fig 3.15 si limitano a descrivere la dinamica della pandemia Covid-19 nei mesi Settembre, Ottobre e Novembre 2020 senza considerare, tuttavia, le varie restrizioni verificatesi durante quei tre mesi.

In realtà, come già anticipato nei capitoli precedenti, durante l'arco temporale analizzato in questo elaborato, diverse sono state le restrizioni messe in atto dal Governo italiano nella regione Emilia Romagna. Ripercorriamole:

- 20 ottobre 2020 → in seguito alla seconda ondata avvenuta in tutta la penisola italiana, la curva dei contagi quindi riprende a salire vertiginosamente senza dare segni di rallentamento. Il presidente del Consiglio Giuseppe Conte emana, a distanza di pochi giorni l'uno dall'altro, due DPCM con restrizioni via via crescenti;
- 15 novembre 2020 → la regione Emilia Romagna passa da zona gialla ad arancione.

A seguito delle restrizioni applicate si é suddiviso il periodo di 90 giorni in tre tratti, affinché i risultati ottenuti e quindi l'identificazione dei 16 parametri del modello SIDARTHE si adattassero ai limiti imposti.

Naturalmente si é tenuto conto del fatto che, in conseguenza all'applicazione di una specifica restrizione, il monitoraggio degli andamenti delle curve relative alle variabili del modello non é immediato e non avviene al giorno 0 della restrizione applicata, ma é necessario un range che va da 7 giorni a 14 giorni dall'imposizione della restrizione stessa.

In questo elaborato si é applicata una traslazione di 7 giorni dall'applicazione delle restrizioni avvenute il 20 ottobre 2020 e il 15 novembre 2020.

Dopo aver fatto queste premesse i tre tratti presi in considerazioni relativi ai 90 giorni che vanno dal 1 settembre al 29 novembre 2020 sono:

- primo tratto: 1 settembre - 27 ottobre 2020 (57 giorni);
- secondo tratto: 28 ottobre - 22 novembre 2020 (25 giorni);
- terzo tratto: 23 novembre - 29 novembre 2020 (8 giorni).

I vari passaggi da eseguire sono gli stessi descritti precedentemente, ossia:

- Risoluzione del *problema inverso* (FASE 1), ossia partendo dalla conoscenza delle condizioni iniziali delle variabili e partendo dalla conoscenza dei parametri iniziali (FASE 0), si risolve il problema di minimi quadrati non lineari per identificare i parametri $\Lambda^* = [\alpha^*, \beta^*, \gamma^*, \delta^*, \epsilon^*, \zeta^*, \eta^*, \theta^*, \kappa^*, \lambda^*, \mu^*, \nu^*, \xi^*, \rho^*, \sigma^*, \tau^*]$ ottimali;
- Risoluzione del *problema diretto* (FASE 2), ossia risoluzione del sistema (3.6), il quale utilizza i parametri ottimali trovati nel problema inverso, per raffigurare l'andamento nel tempo dei compartimenti $S_m, I_m, D_m, A_m, R_m, T_m, H_m, E_m$;
- Validazione andamento del modello teorico con dati sperimentali, quindi ci si aspetta che il modello descritto dai compartimenti $S_m, I_m, D_m, A_m, R_m, T_m, H_m, E_m$ approssimi nel miglior modo i dati sperimentali prelevati dal database della

Protezione Civile relativi alla regione Emilia Romagna per il periodo che va dal 1 Settembre 2020 al 29 Novembre 2020 (90 giorni).

Il vettore $\Lambda^* = [\alpha^*, \beta^*, \gamma^*, \delta^*, \epsilon^*, \zeta^*, \eta^*, \theta^*, \kappa^*, \lambda^*, \mu^*, \nu^*, \xi^*, \rho^*, \sigma^*, \tau^*]$ ottenuto, relativo al primo tratto, al secondo tratto e al terzo tratto viene rappresentato nella figura seguente:

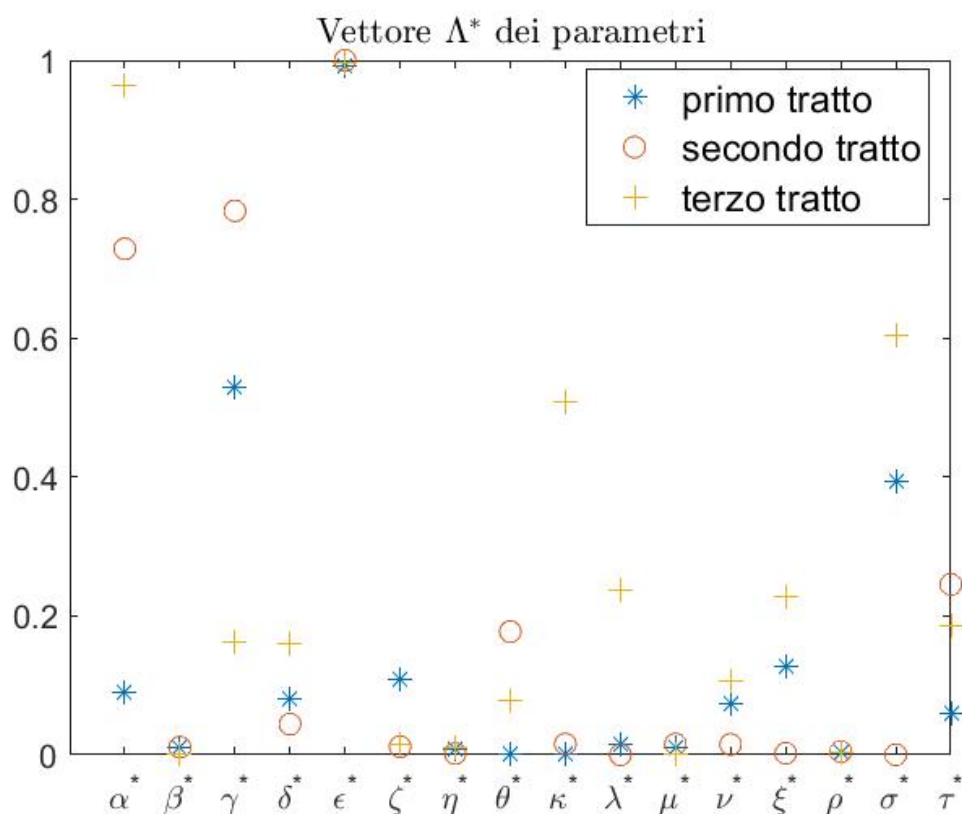


Figura 3.16: Valori dei 16 parametri relativi ai 3 tratti definiti.

La suddivisione dell'arco temporale di 90 giorni ha portato alla modifica di alcuni dei 16 parametri caratterizzanti il vettore Λ^* che descrivono il modello SIDARTHE, modifica lecita perché ha sottolineato il fatto che le restrizioni imposte stavano dando i risultati sperati.

Si nota infatti che tra i 16 parametri Λ^* identificati, quelli che si discostano maggiormente durante i tre tratti selezionati sono:

- α^* (velocità di trasmissione della malattia dovuta ai contatti tra un soggetto suscettibile S e uno infetto asintomatico non rilevato I), che aumenta considerevolmente per il secondo tratto, questo é dovuto molto probabilmente all'elevata probabilità che un soggetto suscettibile entri in contatto con un soggetto infetto ma asintomatico non rilevato essendo il periodo analizzato caratterizzato dalla zona gialla, e quindi é possibile passare piú tempo in modo inconsapevole con un soggetto asintomatico; mentre aumenta lievemente per il terzo tratto, in relazione al fatto che la trasmissione continua ad esserci ma in maniera piú contenuta essendo l'Emilia Romagna passata in zona arancione;
- γ^* (velocità di trasmissione della malattia dovuta ai contatti tra un soggetto suscettibile S e uno infetto sintomatico non rilevato A), che aumenta durante il passaggio dal primo al secondo tratto, e questo é sempre giustificato dal fatto che la regione parte inizialmente dall'essere in zona gialla, e successivamente cala durante il passaggio dal secondo al terzo tratto. Questo é giustificato dal fatto che la regione si trovi in zona arancione con restrizioni dunque maggiori e che i soggetti suscettibili tendano ad evitare contatti con soggetti che mostrano sintomi, anche se la diagnosi non é ancora stata fatta.

I rimanenti parametri restano pressocché costanti, come ad esempio β^* (velocità di trasmissione della malattia dovuta ai contatti tra un soggetto suscettibile S e uno infetto asintomatico rilevato D) e δ^* (velocità di trasmissione della malattia dovuta ai contatti tra un soggetto suscettibile S e uno infetto sintomatico rilevato R), questo é dovuto al fatto che ci si aspetta che i soggetti a cui é stata diagnosticata la malattia siano adeguatamente isolati, a seguito anche delle restrizioni apportate.

Dopo aver risolto il problema inverso e quindi aver identificato i parametri appartenenti al vettore Λ^* , si risolve il problema diretto.

Si mostrano dunque di seguito gli andamenti delle curve relativi alle variabili del modello SIDARTHE, i quali rappresentano sia i dati sperimentali $\bar{y} = [\bar{S}, \bar{I}, \bar{D}, \bar{A}, \bar{R}, \bar{T}, \bar{H}, \bar{E}]$ sia i dati teorici $Y_m = [S_m(t), I_m(t), D_m(t), A_m(t), R_m(t), T_m(t), H_m(t), E_m(t)]$.

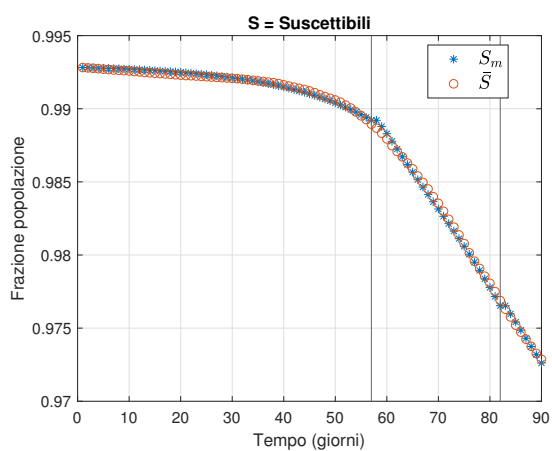


Figura 3.17: Andamento compartimento "S" nell'arco di 90 giorni

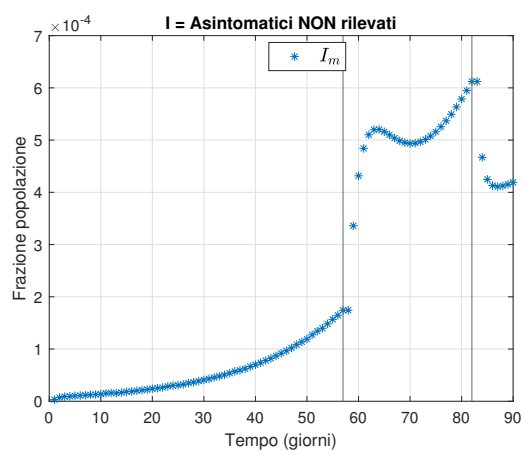


Figura 3.18: Andamento compartimento "I" nell'arco di 90 giorni

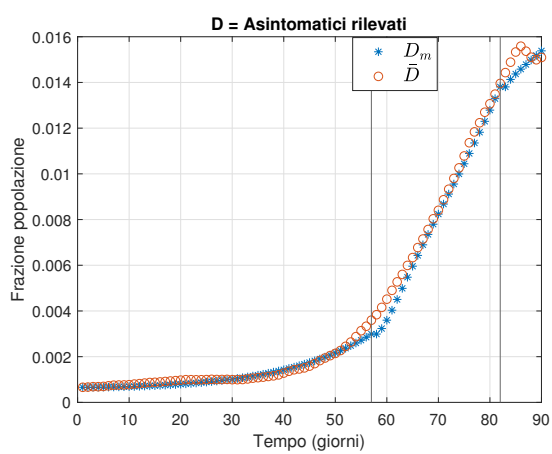


Figura 3.19: Andamento compartimento "D" nell'arco di 90 giorni

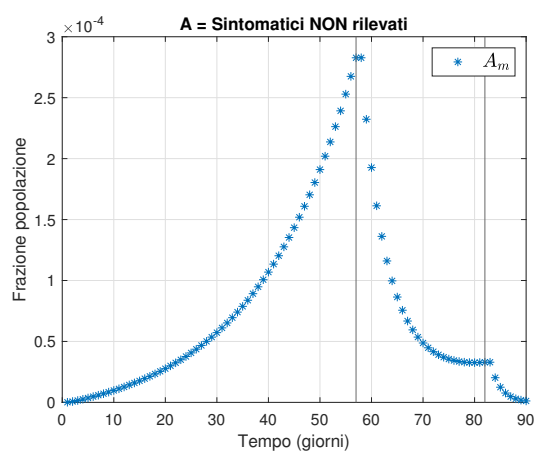


Figura 3.20: Andamento compartimento "A" nell'arco di 90 giorni

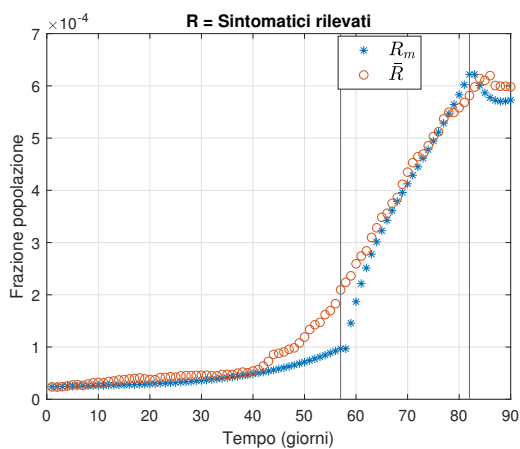


Figura 3.21: Andamento compartimento "R" nell'arco di 90 giorni

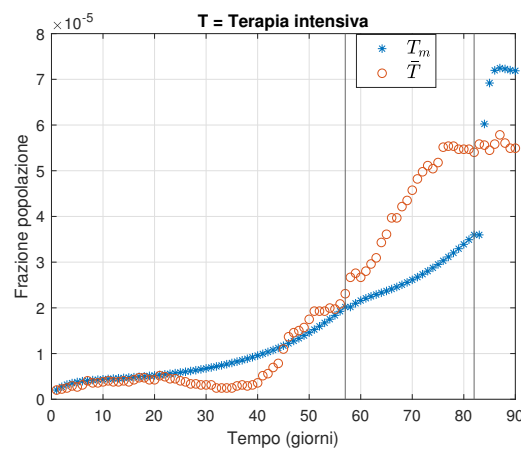


Figura 3.22: Andamento compartimento "T" nell'arco di 90 giorni

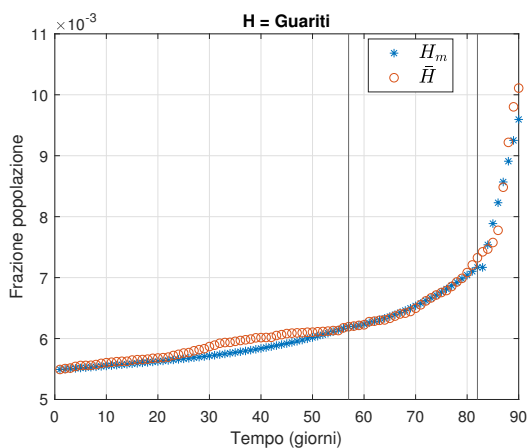


Figura 3.23: Andamento compartimento "H" nell'arco di 90 giorni

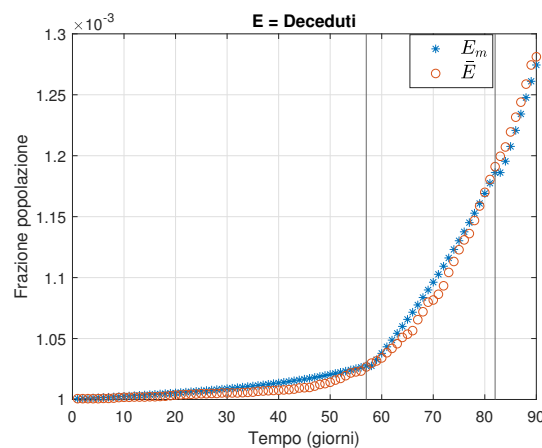


Figura 3.24: Andamento compartimento "E" nell'arco di 90 giorni

Aver suddiviso il periodo dei 90 giorni in tre tratti distinti caratterizzati da parametri diversi ha portato a dei risultati migliori. Si osserva infatti che i dati teorici seguono abbastanza fedelmente i dati sperimentali, inoltre si osserva che,

sia nel compartimento degli 'asintomatici non rilevati' (I) che nel compartimento dei 'sintomatici non rilevati' (A), gli andamenti delle curve decrescono a livello approssimativamente dei punti in cui ci si aspetta che le restrizioni abbiano avuto effetto, rispettivamente circa il giorno 58 (a seguito dell'applicazione dei DPCM), e circa il giorno 82 (a seguito del passaggio della regione Emilia Romagna in zona arancione).

Tuttavia, si osserva che dalla Fig 3.17 alla Fig 3.24, solo la curva dei dati sperimentali relativi al compartimento della 'terapia intensiva' (T) si discosta significativamente da quella dei dati sperimentali.

Questo può essere giustificato dal fatto che in Emilia Romagna le terapie intensive hanno ospitato non solo pazienti emiliani-romagnoli, ma anche pazienti provenienti da altre regioni, infatti è stato realizzato, a metà aprile 2020, il progetto *'Hub regionale e nazionale per la terapia intensiva'*.

Articolato su 6 strutture ospedaliere e ospedaliero-universitarie del territorio, a Bologna, Modena, Parma e Rimini, il progetto rafforza il sistema sanitario regionale, aumentando la dotazione complessiva di 144 nuovi posti letto di terapia intensiva e sub-intensiva, dislocati all'Ospedale Infermi di Rimini (34 nuovi posti letto di terapia intensiva), al Policlinico Sant'Orsola (14) e all'Ospedale Maggiore di Bologna (34), al Policlinico di Modena (30), all'Ospedale Civile di Baggiovara (18), sempre a Modena, e all'Ospedale Maggiore di Parma (14).

Vanno poi considerati i 45 posti letto di terapia intensiva creati nella provincia di Piacenza, una delle aree più colpite in Italia, durante la crisi da Covid-19, che diventano strutturali all'interno della sanità regionale pubblica.

Sono dunque spazi utilizzabili da subito per curare pazienti in situazioni critiche che richiedono questo tipo di assistenza specialistica nell'ambito dell'attività ordinaria, e ai quali ricorrere per gestire un'eventuale nuova ondata epidemica di pazienti Covid, provenienti anche da altre regioni, o gestire qualsiasi altra emergenza [6].

3.2.3 Previsione relativa al mese di Dicembre 2020

Dopo aver analizzato il modello SIDARTHE per il periodo che va dal 1 settembre al 29 novembre 2020, si é voluto validare il modello nell'arco temporale che va dal 1 dicembre 2020 al 15 dicembre 2020, con l'intento di testare il modello implementato precedentemente anche su un altro periodo contiguo dell'anno.

Per quanto riguarda le restrizioni imposte dal Governo Italiano avvenute nel mese di dicembre si puó dire che il:

- 6 dicembre:

la regione Emilia Romagna entra in zona gialla e ne rimarrá fino al giorno 21 dicembre 2020; in zona gialla é consentito spostarsi tra le regioni appartenenti alla zona gialla; é consentito spostarsi dalle ore 5 alle 22 senza dover giustificare gli spostamenti. Confermato il divieto di spostamenti dalle ore 22 alle 5 del giorno successivo, se non per comprovate esigenze lavorative, situazioni di necessitá o motivi di salute (per le quali occorre l'autodichiarazione prevista). Si ha didattica a distanza al 100% per gli studenti delle scuole secondarie di secondo grado – il ritorno delle lezioni in presenza, al 75% - e Università; servizi di ristorazione, quindi bar, pub, ristoranti, gelaterie e pasticcerie sono aperti dalle ore 5 alle 18, con l'asporto consentito fino alle 22. Inoltre nelle giornate festive e prefestive sono chiusi gli esercizi commerciali all'interno dei centri commerciali e dei mercati, sempre in base al Decreto governativo, ad eccezione di farmacie, parafarmacie, presidi sanitari e punti vendita di generi alimentari, tabacchi e edicole. Restano sospesi gli spettacoli e le mostre aperti al pubblico in teatri, sale da concerto, musei; stop alle attivitá di palestre e centri benessere; raccomandazione al piú ampio uso dello smart working per le attivitá lavorative e professionali, sia nel privato sia nel pubblico impiego. Infine é obbligatorio l'uso della mascherina per coprire naso e bocca sempre dal momento in cui si esce dalla propria abitazione, fatto salvo che per i bambini al di sotto dei sei anni, le persone con patologie o disabilitá incompatibili con l'uso della mascherina e durante l'esercizio dell'attivitá sportiva, nelle modalitá

consentite, così come restano raccomandati il distanziamento fisico e l'igiene costante e accurata delle mani. [6]

Ciò che si è fatto è stato prendere i parametri caratterizzanti il vettore Λ^* relativi al terzo tratto del periodo settembre-novembre 2020, ed utilizzarli per risolvere il problema diretto e quindi il sistema di equazioni differenziali ordinarie, non considerando quindi il problema inverso con conseguente identificazione dei parametri.

Mentre le condizioni iniziali prese in considerazione si riferiscono al primo giorno del periodo analizzato, ossia 1 dicembre 2020.

Per avere un quadro generale e quindi un riscontro da parte dei dati sperimentali, viene fatto il confronto dei dati teorici appena determinati con i dati sperimentali prelevati dal repository della Protezione Civile relativi al periodo 1 dicembre - 15 dicembre 2020.

Nel database della Protezione Civile viene prelevata anche la voce 'tamponi' che rappresenta il numero di tamponi effettuati nella regione Emilia Romagna durante il periodo che va dal 1 dicembre 2020 al 15 dicembre 2020.

Di seguito viene riportato l'andamento nel tempo:

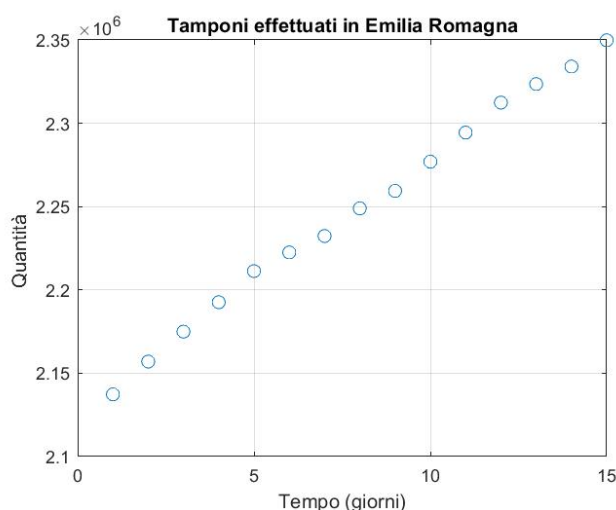


Figura 3.25: Andamento tamponi effettuati in Emilia Romagna dal 1 Dicembre al 15 Dicembre 2020.

Si mostrano dunque di seguito gli andamenti delle curve relativi alle variabili del modello SIDARTHE, i quali rappresentano sia i dati sperimentali $\bar{y} = [\bar{S}, \bar{I}, \bar{D}, \bar{A}, \bar{R}, \bar{T}, \bar{H}, \bar{E}]$ sia i dati teorici $Y_m = [S_m(t), I_m(t), D_m(t), A_m(t), R_m(t), T_m(t), H_m(t), E_m(t)]$.

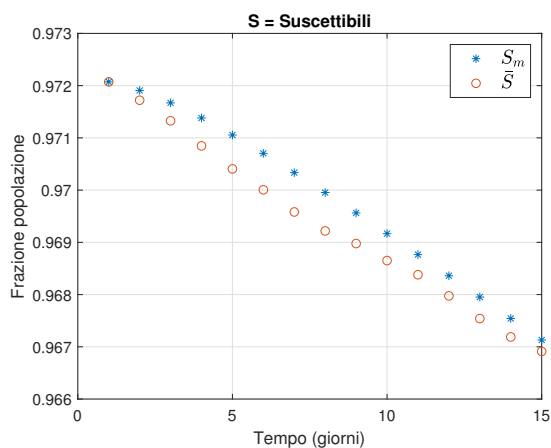


Figura 3.26: Andamento compartimento "S" nell'arco di 15 giorni

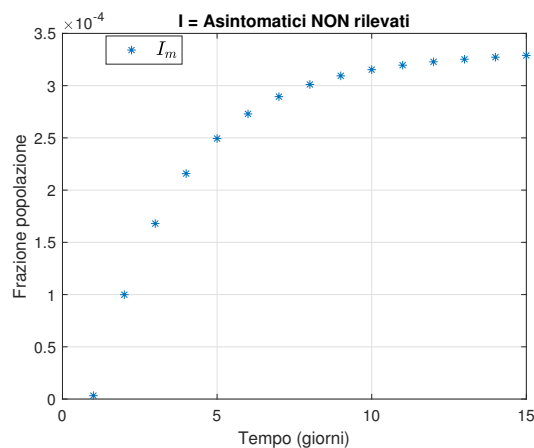


Figura 3.27: Andamento compartimento "I" nell'arco di 15 giorni

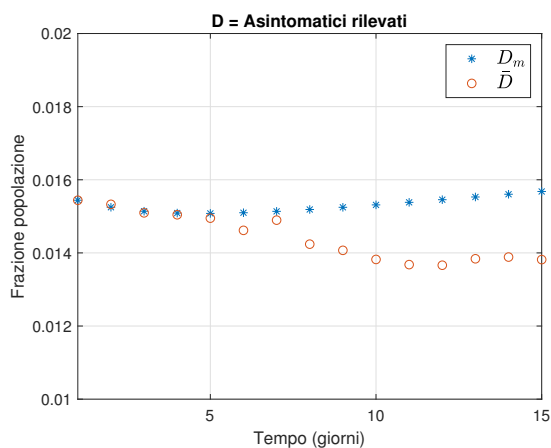


Figura 3.28: Andamento compartimento "D" nell'arco di 15 giorni

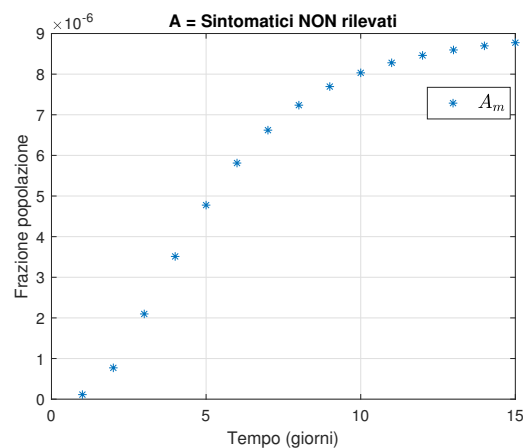


Figura 3.29: Andamento compartimento "A" nell'arco di 15 giorni

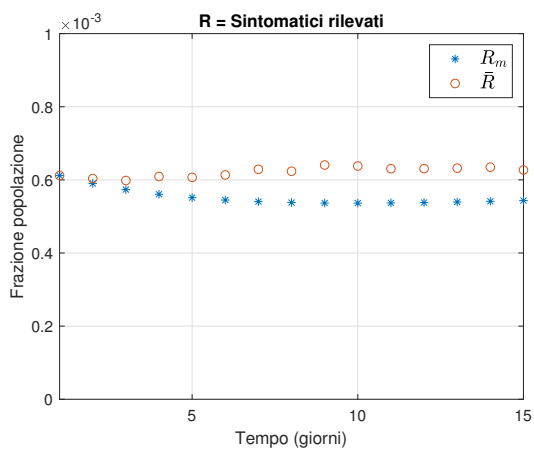


Figura 3.30: Andamento compartimento "R" nell'arco di 15 giorni

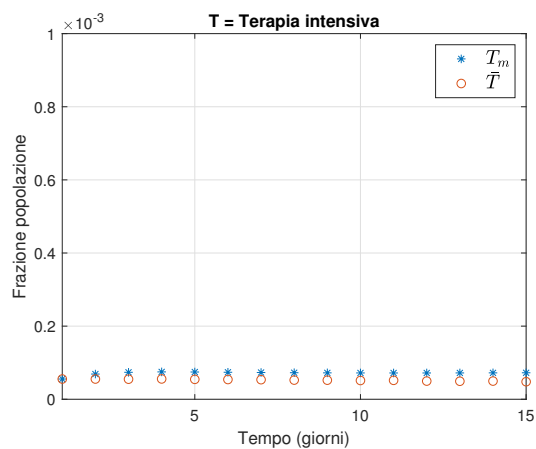


Figura 3.31: Andamento compartimento "T" nell'arco di 15 giorni

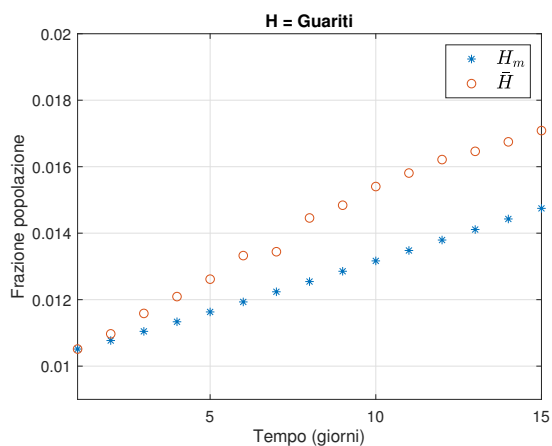


Figura 3.32: Andamento compartimento "H" nell'arco di 15 giorni

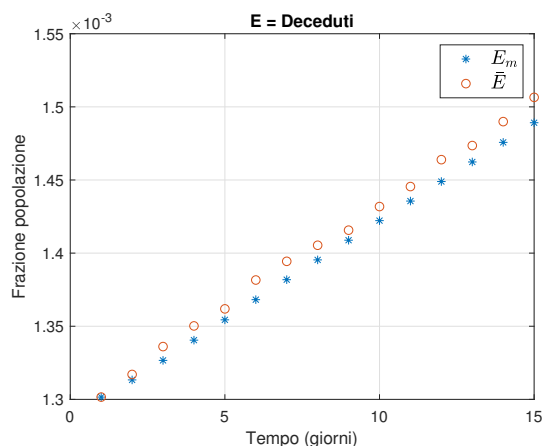


Figura 3.33: Andamento compartimento "E" nell'arco di 15 giorni

Dalle figure rappresentate si osserva che i dati teorici seguono abbastanza fedelmente i dati sperimentali, dunque é possibile concludere che il modello matematico validato permette di prevedere efficacemente periodi di tempo contigui e futuri a quello analizzato in questo elaborato.

Conclusioni

Il lavoro proposto mostra, con l'ausilio di modelli matematici, come sia possibile analizzare alcuni dati di una situazione reale e ottenerne anche delle predizioni per periodi futuri.

I modelli epidemiologici utilizzati in questo elaborato sono stati il modello semplice classico SIR e quello complesso SIDARTHE, i quali possono essere applicati ad epidemie generiche, ma in questo elaborato ci si è concentrati sulla specifica pandemia COVID-19. Entrambi i modelli sono stati implementati nel periodo settembre - novembre 2020 in relazione alla regione Emilia Romagna. La loro analisi è stata caratterizzata prima dal problema inverso attraverso l'identificazione dei parametri che caratterizzano il modello differenziale per questa particolare epidemia, e poi dal problema diretto attraverso la risoluzione del sistema di equazioni differenziali ordinarie.

Tuttavia la difficoltà affrontata non è stata tanto la costruzione del modello matematico attraverso la risoluzione del problema diretto, quanto la ricerca di parametri che rendessero il modello più fedele possibile alla situazione reale documentata dai dati del repository della Protezione Civile. È importante sottolineare che per problema inverso si intende porre i parametri specifici appena identificati per quella pandemia specifica, per quella nazione specifica e per quella regione specifica.

Si è voluto anche validare l'efficienza e l'attendibilità del modello matematico su periodi di tempo contigui da quelli presi in esame in questo elaborato, concentrandosi sui primi 15 giorni di dicembre 2020. Possiamo concludere dicendo che il modello matematico si adatta abbastanza bene ai dati sperimentali nonostante non siano state considerate le restrizioni relative al periodo precedente.

Ovviamente si potrebbe estendere l'analisi del modello matematico ad altri dati della pandemia, facendo considerazioni ad esempio su altri stati europei o extraeuropei.

In ogni caso si conclude dicendo che i modelli matematici possono agevolare la nostra comprensione del fenomeno che stiamo vivendo e possono anche aiutarci ad affrontarlo nel migliore dei modi.

Bibliografia

- [1] Francesco Casella. Can the covid-19 epidemic be controlled on the basis of daily test reports? *IEEE Control Systems Letters*, 5(3):1079–1084, 2021.
- [2] Marino Gatto, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, 117(19):10484–10491, 2020.
- [3] Chiara Gentile. *Metodo del gradiente coniugato per problemi ai minimi quadrati non lineari*. PhD thesis.
- [4] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, January 2020.
- [5] Dati dal repository. <https://github.com/pcm-dpc/COVID-19>.
- [6] Situazione covid-19 in emilia romagna. <https://www.regione.emilia-romagna.it>.
- [7] Qianying Lin, Shi Zhao, Daozhou Gao, Yijun Lou, Shu Yang, Salihu S. Musa, Maggie H. Wang, Yongli Cai, Weiming Wang, Lin Yang, and Daihai He. A conceptual model for the coronavirus disease 2019 (covid-19) outbreak in wuhan, china with individual reaction and governmental action. *International Journal of Infectious Diseases*, 93:211–216, 2020.

- [8] Z. Liu, P. Magal, Ousmane Seydi, and Glenn Webb. A model to predict covid-19 epidemics with applications to south korea, italy, and spain. *medRxiv*, 2020.
- [9] Luca Magri and Nguyen Anh Khoa Doan. First-principles machine learning modelling of covid-19, 2020.
- [10] Giovanni Monegato. *Fondamenti di calcolo numerico*. Clut, 1998.
- [11] Irene Moretto. *Aspetti numerici nell'applicazione del modello SIR*. PhD thesis.
- [12] Jorge Nocedal. *Numerical optimization / Jorge Nocedal, Stephen J. Wright*. Springer series in operations research. Springer, New York, 2nd ed.. edition, 2006.
- [13] Elena Loli Piccolomini and Fabiana Zama. Preliminary analysis of covid-19 spread in italy with an adaptive seird model. *arXiv preprint arXiv:2003.09909*, 2020.
- [14] Alfio Quarteroni, Riccardo Sacco, and Fausto Saleri. *Matematica numerica*. Springer Science & Business Media, 2010.
- [15] Biao Tang, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *Journal of Clinical Medicine*, 9(2), 2020.

Ringraziamenti

Giunta al termine di questo meraviglioso ed intenso percorso, è indispensabile per me ringraziare:

La mia relatrice, la professoressa Morigi, la quale mi ha guidata con attenzione e premura in questo lungo percorso di tesi, condividendo con me soddisfazioni e preoccupazioni, consigliandomi sempre la scelta più giusta da prendere. Grazie per avermi fatto entrare in punta di piedi nel grande mondo della Matematica e avermi fatto toccare con mano argomenti impegnativi ma al tempo stesso attuali e coinvolgenti.

La mia famiglia, mia mamma, mio papà e mio fratello per avermi permesso di intraprendere questa bellissima avventura, per avermi insegnato ad essere tenace e resiliente, per aver creduto sempre in me e avermi fatto da ancora nei momenti in cui mi credevo perduta. A voi devo tutto. A voi devo la persona che sono diventata oggi. Naturalmente ringrazio anche il mio cagnolino Oscar che, nei momenti felici mi ha trasmesso la sua scodinzolante felicità, ma ancor di più in quelli tristi, facendo sempre trasparire la sua innata fedeltà.

Armando, il quale nonostante questo virus ci abbia reso ancora più distanti, mi è sempre stato vicino con il cuore e con il suo amore, affrontando insieme a me questo lungo percorso, condividendo con me gioie e preoccupazioni, vittorie e sconfitte, supportandomi e sopportandomi, subendo silenziosamente e pazientemente i miei momenti 'no', trasmettendomi anzi la serenità di cui avevo bisogno in quel momento. Con te oggi condivido questa gioia che spero sia la prima di una lunga lista.

Nonna Ebe, la quale, seppur non sia più qui con me fisicamente, mi ha permesso di percepire i suoi abbracci e la sua protezione ogni qual volta mi si presentasse

davanti una difficoltà.

Nonna Anna, i miei zii e i miei cugini, sempre pronti a supportarmi con un sorriso, un caldo abbraccio o semplicemente una parola di conforto in momenti in cui un sorriso, un caldo abbraccio o una parola di conforto era proprio ciò di cui avevo bisogno.

I miei migliori amici Grasholo, Anna Rita e Adele, punti fermi della mia vita, con loro sono cresciuta e con loro ancora una volta condivido un momento importante per me, perchè so che la mia felicità è anche la loro. Grazie per esserci sempre stati, il nostro legame è ormai indissolubile.

Sara, la quale seppur sia entrata nella mia vita da poco più di un anno, sembra che mi conosca da sempre, con lei ho condiviso l'ultimo periodo universitario, direi il più duro, e so per certo che il suo aiuto e il suo sostegno lo hanno reso sicuramente più leggero da affrontare. So che potrò sempre contare su di lei perchè il suo sguardo mi trasmette sincerità e lealtà. Io ti ringrazio per tutto Saretta bella e ti dico che la nostra amicizia la proteggerò con cura.

I miei compagni Marta, Giacomo, Claun, Edwin, Fedez, Giacomino, Gambs, e tutti coloro che hanno vissuto con me questo intenso percorso, accompagnandomi in tutti questi anni, condividendo con me momenti bellissimi, vittorie e anche sconfitte. Vi porterò nel cuore.

Le mie coinquiline Chiara e Camilla, che mi hanno letteralmente sopportata e supportata durante l'ultimo periodo universitario, sono state la mia valvola di sfogo e tramite tisane e dolcetti hanno sicuramente contribuito nel loro piccolo ad alleggerire la mia tensione. Vi prometto che le CCCoinqui non si perderanno.

La mia Cesena, piccolo scrigno che mi ha accolta, protetta e fatta crescere. Sarà difficile per me lasciarla perchè ormai è Casa. Porterò con me ogni singolo momento passato qui.

Sono pronta dunque adesso ad affrontare nuove avventure, a mettermi in gioco, e sono curiosa di sapere cosa mi riserverà il futuro perchè come dice Cremonini "Per quanta strada ancora c'è da fare. Amerai il finale!".