POLITECNICO DI TORINO

Master Degree course in Mathematical Engineering

Master's Degree Thesis

# Modeling the new flu wave using data science and complex networks theory.

**Supervisors**
Prof. Lorenzo Zino
Prof. Alessandro Rizzo

**Candidate**
Francesco Celino

Academic Year 2023-2024

# Acknowledgements

Ancora da fare...

**Abstract**

This Master's thesis explores the use of the SEINR (Susceptible, Exposed, Infectious, Non-infectious, Recovered) compartmental model to forecast the evolution of flu-like illness cases in Italy during the 2023-24 winter season. The model's innovative aspects lie in its region-based meta-population framework, which simulates *intra-* and *inter*-regional mobility, capturing network dynamics critical to understanding how a disease spreads in Italy's diverse demographic and geographic landscape. This approach, previously successful in evaluating the efficacy of NPIs during the *COVID-19* pandemic, is further refined by incorporating features such as class divisions based on age, activity levels, and vulnerability of different age groups.

Key epidemiological parameters, such as infection duration and transmission rates, are then estimated with Bayesian inference methods to improve forecast accuracy. Empirical results demonstrate the modelss effectiveness in capturing flu trends across Italian regions, especially in critical timeframes during which increased number of contacts are observable.

This work emphasizes the role of adaptive modeling in epidemiology, and how public health strategies driven by past data can help in managing seasonal epidemics.

# Contents

# Chapter 1

# Introduction

Each year, Influenza affect millions of people across the world, posing great challenges for healthcare systems in various developed countries, with great risks for elderly and more vulnerable people. In response, researchers and public health officials work to predict and control the spread of these diseases, using various tools and methods. One of the most powerful tools we have available is mathematical modeling, which allows us to simulate how a disease might spread through a population through the use of "differential equations". This thesis focuses on adopting one of such models [5] for flu-like illnesses in Italy, using a framework called the SEINR model.

The SEINR model is a type of "compartment" model. This means that it divides the population into different groups (or compartments) of people, based on the stages of the disease and whether or not they are infected. In this case, the compartments are *Susceptible* (people who can catch the flu), *Exposed* (people who have been infected but are not yet contagious), *Infectious* (people who can spread the flu), *Non-infectious* (people who have recovered or are isolated), and *Removed* (people who are either immune to the flu after recovering, or dead due to complications).

The history of compartmental models dates back to 1927, when Kermack and McKendrick [7] defined the SIR model for the first time, in its simplest form.

However, what makes the model we used for this thesis particularly innovative is its "meta-population" structure, already used with great success in 2021 when trying to model the first Covid-19 wave in Italy [5]. Instead of treating Italy as one large group of people, the meta-population approach breaks it down into regions and even considers how people move between these regions. This is crucial for a country like Italy, where people frequently travel between cities and regions for work, school, and other reasons.

Most importantly, when trying to predict the outcomes of a Influenza season, we need to consider other crucial factors, such as vaccines, the age and activity levels of individuals, as well as their vulnerability to illness. These factors are important because some groups, like the elderly or those with pre-existing health conditions, are more at risk during a flu outbreak, but could very well be less exposed to the disease due to having less social contacts. If we want our predictions to be accurate, we need to take into account all of those factors.

Another key feature of this thesis is the use of "Bayesian inference", a statistical method that allows the model to improve its accuracy (and precision) over time. As new weekly data about flu cases becomes available, the model updates its parameters in a probabilistic fashion: that means, if in one particular week we observe more infected people than we would expect, we probably need to rethink our predictions. That's why, when using Bayesian inference, we do not provide exact estimates for the number of infected people we will have next week, but rather a "probability distribution" that includes reasonable confidence intervals: as more weeks pass and we gather more real data, those probability distributions get narrower, reflecting our increased confidence on the disease's characteristics.

In short, the goal of my thesis is to provide a more accurate and flexible tool for predicting flu-like illnesses in Italy, one that can adapt to different variants of Influenza each year (as you probably know, each year the flu outbreak is slightly different due to changing vaccines, mobility patterns, timing and public awareness). This tool can help public health officials better prepare for and respond to outbreaks, reducing the strain on healthcare systems and, who knows, maybe also save some lives.

I am also proud to say that, thanks to the invaluable guidance of my supervisors, we were able to use this model to contribute to two major forecasting projects at the European level. Through these projects, I had the opportunity to collaborate with the *ISI Foundation* and the *Istituto Superiore di Sanita'*, as we provided weekly estimates for the evolution of last year's flu season.

[5]

# Chapter 2

# Methods

Inizia parlando del SIR, della struttura a equazioni diff. alle derivate ordinarie, introduci il modello SEINR generale coi 5 diversi compartimenti

## 2.1 Metapopulation Framework

Spiega cosa c'e' di nuovo nell'approccio a meta-popolazione, come funzionano i contatti fra individui, come ci si sposta

### 2.1.1 Matrice di pendolarismo

metti figura e spiega

### 2.1.2 Vaccines

Parla di come abbiamo implementato i vaccini e dove abbiamo preso i dati su efficacia [3], copertura in Italia e timeline [1]

## 2.2 SEINR Metapopulation model

Introduci le equazioni differenziali del modello SEINR aggiornate al contesto a MetaPopolazione

## 2.3 Our data

Spiega da dove abbiamo preso i dati, il modo in cui sono formattati, metti una tabella sulla struttura dei file .csv

### 2.3.1 Data manipulation

Spiega che i dati reali sono regionali e il nostro modello ragiona in base provinciale, quindi abbiamo dovuto attuare una suddivisione degli infetti in proporzione al numero di abitanti (approssimazione molto forte, spiega i limiti di questo approccio)

### 2.3.2 Initial conditions and Model Calibration

Spiega come abbiamo usato i dati reali delle stagioni di influenza passate per impostare le condizioni iniziali del modello, e anche per una calibrazione iniziale dei parametri del modello. Spiega anche l'approccio alternativo di calibrare il modello in base a studi di natura virologica [4] [2] facendo un confronto col covid.

## 2.4 A different approach: Bayesian estimates

parla della seconda parte del lavoro, cioe' riscrivere il modello per ottenere un aggiornamento in tempo reale dei parametri scelti

### 2.4.1 Theory of Bayesian Estimates

Parla molto brevemente del teorema di Bayes, che cos'e' una prior, una posterior e la likelyhood, e di come si possono usare questi concetti nel nostro modello. Spiega le differenze con l'approccio frequentista in probabilita'.

### 2.4.2 Implementation

Spiega come implementare questo approccio nel modello SEINR tramite Metodi Monte Carlo per via della forma complicatissima della likelyhood

# Chapter 3

# Coding and Implementation

Parla delle librerie di python utilizzate, dell'approccio "discreto" usato per risolvere le eq. diff. ordinarie

## 3.1 Coding Bayesian Inference

Spiega la parte di codice relativa a Bayes

## 3.2 Uploading our results

Spiega finalmente come abbiamo formattato i dati per inviarli a Influcast e Respicast, metti qualche tabella

## 3.3 Related work

Menzionare brevemente i lavori precedenti a cui ci siamo ispirati per il modello

# Chapter 4

# Experiments and Contributions

Spiega che esperimenti abbiamo fatto, le stime che abbiamo fornito e a quali progetti abbiamo lavorato

## 4.1 Collaborative Projects

### 4.1.1 Influcast

### 4.1.2 Respicast

### 4.1.3 Collaborative paper published

# Chapter 5

# Model Validation

Confronta i risultati ottenuti con i due metodi con le varie metriche di errore (log score, weighted interval score, mean relative error, mean absolute error, coverage) spiegando pregi e difetti di ognuna

## 5.1 Christmas Holiday issue

Parla del picco di casi a Natale e mostra che l'utilizzo di Bayes risolve parzialmente il problema, mostra che tutti i modelli di Influcast hanno fatto fatica in quel periodo

## 5.2 Methodology

Spiega come verificare quale dei due approcci (bayesiano e puramente deterministico) si comporta meglio nel periodo critico natalizio

## 5.3 Numerical results

Metti tutti i grafici e tutte le tabelle per le varie regioni con le metriche dell'errore [6]

# Chapter 6

# Conclusion

Riporta molto brevemente che risultati abbiamo ottenuto, ricalca l'importanza dell'approccio bayesiano.

Spiega i punti deboli del nostro modello: la struttura provinciale della matrice di pendolarismo si sposa male col fatto che i dati reali sono su base regionale. Il fatto che i dati reali vengano forniti su base settimanale influisce negativamente, dato che il modello ragiona in maniera giornaliera. Servono anche piu' dati sull'efficacia del vaccino, e l'applicabilita' del nostro modello e' limitata a singole stagioni influenzali in cui non si prevede la possibilita' di reinfezione.

L'approccio bayesiano probabilmente migliora la precisione del modello (cio' va confermato nei prossimi anni se vogliamo esserne sicuri) ma aumenta in maniera drammatica i costi computazionali, e se si vuole ovviare a questo problema e' necessario applicare l'inferenza bayesiana a pochissimi parametri. Menziona brevemente la questione underfitting e overfitting, e chiedersi se includere il compartimento N comporti dei guadagni in accuratezza.

Proponi spunti per espandere il modello e implementare stime meno onerose per i parametri con il machine learning o l'intelligenza artificiale

# Bibliography


[1] Dati coperture vaccinali influenza, https://www.salute.gov.it/portale/influenza.

[2] Yu Bai and Xiaonan Tao. Comparison of covid-19 and influenza characteristics. *Journal of Zhejiang University-SCIENCE B Biomedicine and Biotechnology*, 2021.

[3] Kissling Esther, Maurel Marine, Emborg Hanne-Dorthe, Whitaker Heather, McMenamin Jim, Howard Jennifer, Trebbien Ramona, Watson Conall, Findlay Beth, Pozo Francisco, Bolt Botnen Amanda, Harvey Ciaran, and Rose Angela. Interim 2022/23 influenza vaccine effectiveness: six european studies, october 2022 to january 2023. *Euro Surveill. 2023;28(21):pii=2300116*, 2023.

[4] Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, and Valleron AJ. Time lines of infection and disease in human influenza: A review of volunteer challenge studies. *American Journal of Epidemiology*, 2007.

[5] Parino F., Zino L., Porfiri M., and Rizzo A. Modelling and predicting the effect of social distancing and travel restrictions on covid-19 spreading. *J. R. Soc. Interface 18: 20200875.*, 2021.

[6] Stefania Fiandrino, Andrea Bizzotto, Giorgio Guzzetta, Stefano Merler, Federico Baldo, Eugenio Valdano, Alberto Mateo-Urdiales, Antonino Bella, Francesco Celino, Lorenzo Zino, Alessandro Rizzo, Yuhan Li, Nicola Perra, Corrado Gioannini, Paolo Milano, Daniela Paolotti, Marco Quaggiotto, Luca Rossi, Ivan Vismara, Alessandro Vespignani, and Nicolo Gozzi. Collaborative forecasting of influenza-like illness in italy: the influcast experience. *medRxiv*, 2024.

[7] William Ogilvy Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A115700–721*, 1927.