



# **Epidemiological Feature Detection: A Latent Mixture Approach to Modelling Surges in COVID-19 Cases and Deaths**

**Adam Melnyk**

Department of Biomedical and Biological Engineering  
McGill University, Montreal, Canada

May 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Engineering, Biomedical and Biological Engineering.

© Adam Melnyk 2022

# Table of Contents

<b>Abstract</b> .....	<b>iii</b>
English .....	iii
French.....	iv
<b>Acknowledgements</b> .....	<b>v</b>
<b>Contribution of Authors</b> .....	<b>vii</b>
<b>List of Figures</b> .....	<b>viii</b>
<b>List of Tables</b> .....	<b>viii</b>
<b>List of Abbreviations</b> .....	<b>ix</b>
<b>Graphical Abstract</b> .....	<b>x</b>
<b>Chapter 1   Introduction</b> .....	<b>1</b>
<b>Chapter 2   Literature Review</b> .....	<b>5</b>
Feature Detection .....	5
Epidemiological Models .....	7
SIRD Model.....	12
Case-Death Latency .....	14
<b>Chapter 3   Methods</b> .....	<b>15</b>
Data Selection and Processing .....	15
Peak Deconvolution.....	18
SIRD Modelling.....	19
<b>Chapter 4   Results</b> .....	<b>21</b>
Peak Analysis.....	21
Model Validation .....	24
Case-death Latency .....	27
<b>Chapter 5   Discussion</b> .....	<b>29</b>
<b>Chapter 6   Conclusions</b> .....	<b>34</b>
<b>References</b> .....	<b>36</b>

## **Abstract**

### **English**

The COVID-19 pandemic continues to emphasize the importance of epidemiological modelling in guiding timely and systematic responses to public health threats. Nonetheless, the predictive qualities of these models remain limited by their underlying assumptions of the factors and determinants shaping national and regional disease landscapes. Here, we introduce epidemiological feature detection, a novel latent variable mixture modelling approach to extracting and parameterizing distinct and localized features of real-world trends in daily COVID-19 cases and deaths. In this approach, we combine methods of peak deconvolution that are commonly used in spectroscopy with accepted epidemiological models of disease transmission. We analyze the second wave of the COVID-19 pandemic in Israel, Canada, and Germany and find that the lag time between reported cases and deaths, which we term case-death latency, is closely correlated with adjusted case fatality rates across these countries. Our findings illustrate the spatiotemporal variability of both these disease metrics within and between different disease landscapes and highlight the complex relationship between case-death latency, adjusted case fatality rate, and COVID-19 management across various degrees of decentralized governments and administrative structures.

## **French**

La pandémie de COVID-19 continue de souligner l'importance de la modélisation épidémiologique pour guider des réponses rapides et systématiques aux menaces à la santé publique. Néanmoins, les qualités prédictives de ces modèles demeurent limitées par leurs hypothèses sous-jacentes concernant les facteurs et les déterminants qui façonnent les paysages pathologiques nationaux et régionaux. Nous présentons ici la détection des caractéristiques épidémiologiques, une nouvelle approche de modélisation par mélange de variables latentes permettant d'extraire et de paramétriser des caractéristiques distinctes et localisées des tendances réelles des cas et des décès quotidiens de COVID-19. Dans cette approche, nous combinons les méthodes de déconvolution des pics couramment utilisées en spectroscopie avec des modèles épidémiologiques reconnus de transmission des maladies. Nous analysons la deuxième vague de la pandémie de COVID-19 en Israël, au Canada et en Allemagne et constatons que le temps de latence entre les cas déclarés et les décès, que nous appelons latence des cas-décès, est étroitement corrélé aux taux de létalité ajustés dans ces pays. Nos résultats illustrent la variabilité spatio-temporelle de ces deux mesures de la maladie au sein et entre les différents paysages pathologiques et mettent en évidence la relation complexe entre la latence des décès, le taux de létalité ajusté et la gestion du COVID-19 à travers différents degrés de décentralisation des gouvernements et des structures administratives.

## **Acknowledgements**

Perhaps one of the greatest joys of writing this thesis has been the opportunity to reflect on the community of friendly faces that made this work possible. First, I would like to thank my supervisor, Professor Sebastian Wachsmann-Hogiu, whose multidisciplinary expertise and genuine scientific curiosity were foundational to my pursuit of research questions beyond the scope of his lab's already esteemed work. Your willingness to explore unfamiliar territory, candidness in providing constructive feedback, and patience throughout this process have led me – by example – to become a better researcher.

Thank you as well to Lena Kozarov, whose hard work and keen observations exceeded her responsibilities as a summer student and ultimately brought my work to a higher level. I would also like to thank the Wachsmann-Hogiu Lab, including Juanjuan Liu, Reza Abbasi, Meruyert Imanbekova, and Sorina Suarasan, who provided me with an added sense of community and purpose while working towards this thesis and whose academic pursuits encouraged me to strive for excellence in my own work.

I would also like to acknowledge and extend my gratitude to my Graduate Program Director, Professor Brandon Xia, whose insights, support, and excitement during progress meetings made this experience all the more enjoyable and productive. Thank you as well to Professor Caroline Wagner, who generously shared her subject matter expertise early on in my foray into epidemiological modelling, and Professor Paul

François, who first introduced me to the idea of COVID-19 modelling in his biophysics course, PHYS 519, at the onset of the pandemic.

The privilege of studying at McGill University, and more broadly in Quebec, Canada, would not have been possible without the generous financial support I received from various local funding bodies. Thank you to McGill's Faculty of Engineering as well as the Department of Biological and Biomedical Engineering for selecting me as a recipient of a McGill Engineering Undergraduate Student Masters Award and Graduate Excellence Fellowship, and to the Fonds de Recherche du Québec – Nature et Technologies for awarding me one of their prestigious Master's Research Scholarships.

Last, but certainly not least, I would like to thank my mom, dad, and brother, Daniel, as well as my better half, Chloe, for their unconditional support of my goals and aspirations in life. My words here would fall short of conveying how much you've helped but know that you continue to bring out the best in me.

## **Contribution of Authors**

The research questions and ideas in the thesis herein were conceived by Sebastian Wachsmann-Hogiu (S.W.-H.) and Adam Melnyk (A.M.). S.W.-H. supervised the project, helped interpret the data, and provided written edits to the thesis. A.M. performed the analysis, interpreted the data, and wrote the thesis. Lena Kozarov (L.K.) helped analyze and interpret the data.

## List of Figures

<b>Figure 1</b> Epidemiological feature detection methodology.....	17
<b>Figure 2</b> Peak fits for national daily cases and deaths reported from the second wave of the COVID-19 pandemic. ....	23
<b>Figure 3</b> Reported peak-by-peak adjusted case fatality rates compared to adjusted case fatality rates calculated from SIRD model parameters in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic.....	26
<b>Figure 4</b> Case-death latency compared to modelled infection rates (A), death rates (B), and adjusted case fatality rates (C), in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic.....	28

## List of Tables

<b>Table 1</b> Correlation between the deconvoluted trends in daily cases and deaths in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic....	24
<b>Table 2</b> Weighted averages of CDLs, infection rates, and death rates calculated from peak fits and SIRD model parameters in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic.....	29

## List of Abbreviations

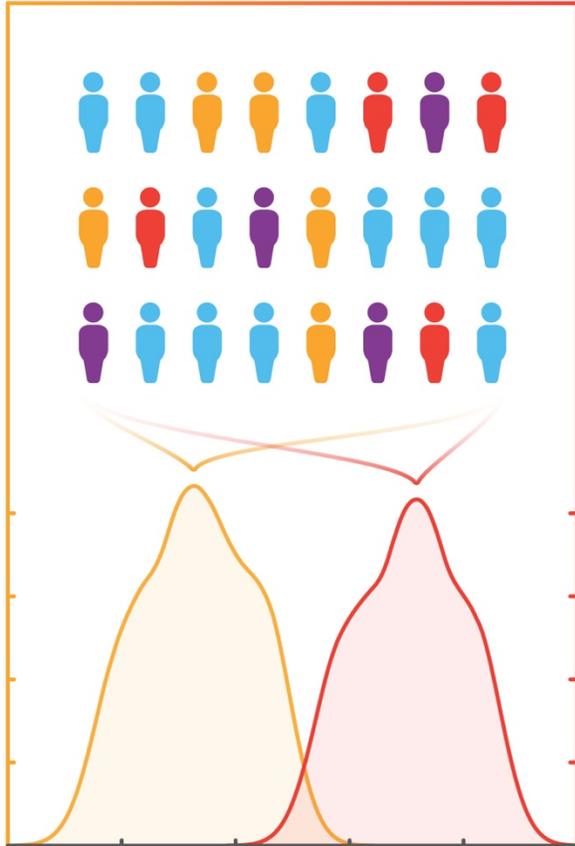
aCFR	Adjusted case fatality rate
CDL	Case-death latency
CFR	Case fatality rate
COVID-19	Coronavirus disease of 2019
DTW	Dynamic time warping
GDP	Gross domestic product
GLEAM	Global epidemic and mobility model
HOG	Histogram of oriented gradients
MATLAB	Matrix laboratory
PDE	Partial differential equation
SEIRD model	Susceptible-exposed-infected-recovered-deceased model
SERS	Surface-enhanced Raman scattering
SIFT	Scale invariant feature transform
SIR model	Susceptible-infected-recovered model
SIRD model	Susceptible-infected-recovered-deceased model
SURF	Speeded up robust features
SVM	Support vector machine

# Graphical Abstract

SIRD Model



Disease Landscape



\*Not to Scale

Time

Hidden Feature of Disease Landscape



Time

## Chapter 1 | Introduction

The COVID-19 pandemic has highlighted the importance of epidemiological modelling in responding to public health threats and continues to be a critical tool to study and anticipate the spread of disease, even with the introduction of vaccines and antiviral therapies. While there is an abundance of peer-reviewed research which sets out to make predictions about the short and long term spread of COVID-19 based on accepted epidemiological models, the study of past and emerging waves of the pandemic remains relatively unexplored despite the potential for understanding the impact of implemented public health responses. Some studies discuss the potential short term effects of nonpharmaceutical interventions such as mask mandates or lockdowns [1, 2] while others look deeper into possible disease futures, speculating about the longer term impacts of emerging variants, vaccination efficacies, and imperfect or waning immunities [3]. While these efforts broadly contribute to the decision making of public health authorities [4], they often place an emphasis on forecasting over retrospective investigations that evaluate the accuracy of predictions [5]. Models of prediction are an integral part of the epidemiologist's toolbox and serve as a basis for pandemic scenario planning; however, precise quantitative forecasting remains an imperfect assessment of future disease landscapes and public health risks [6, 7].

Deterministic compartmental models of disease transmission are among the most common modelling techniques in epidemiology. In these models, individuals are labeled and compartmentalized based on their disease status (e.g., susceptible, infected,

recovered) and set to move between compartments over time according to model parameters and dynamics representative of a specified epidemiological landscape. Mean-field compartmental models assume that labeled populations are sufficiently large and homogeneously mixed, such that variations in individual behaviours are approximated by a single averaged effect across an entire population [8]. Averaged analytical solutions of compartmental dynamics are generated from point estimates of model parameters, which provide simple approximations of disease progression but limit the ability to quantify model uncertainty, especially for long term forecasting [8, 9]. While the simplifying assumptions of these models pose limitations on their predictive abilities, they provide a parsimonious framework for measuring and monitoring past and real-time trends in disease landscapes based on spatiotemporal data [2].

To complement ongoing scenario planning initiatives for pandemic preparedness, science advisors and policymakers need to take an empirical approach to modelling disease landscapes. Empirical models are not intended to derive projections of disease progression but instead use analytical methods to interpolate and better estimate drivers of disease over time. These approaches to disease spread use epidemiological frameworks, such as compartmental models, to quantify and simulate real-world data as opposed to forecasting based on assumptions of epidemiological parameters such as reproduction number, contact rate, or critical vaccination threshold. They are often used to estimate such parameters [10] and may also be used to assess the validity of projections set forth by predictive models [5, 11].

While compartmental models serve as a basis for studying disease landscapes, the ability of these models to provide spatiotemporal information on disease progression is limited by the granularity of the data being studied and by their capacity to extract and relate latent epidemiological features across different data types (e.g., cases and deaths). Motivated by the concept of feature detection used in the field of computer vision [12], a feature represents a distinct and localized grouping of new cases, deaths, and other epidemiological data (e.g., hospitalizations) from a larger disease landscape. Latent (or hidden) features could include city-wide outbreaks that become indiscernible at a national scale or local superspreader events smoothed over amidst provincial data, and they often manifest asynchronously across cases and deaths to shape disease landscapes. On a broader scale, these features are often the result of many confounding real-world events which share the same spatiotemporal localizations.

Epidemiologists should have access to other tools and techniques, such as peak deconvolution, to extract hidden features from disease landscapes and improve existing epidemiological frameworks where richer data inputs may be unavailable. Peak deconvolution describes the process of deconstructing overlapping data features into individual components to extract hidden information about underlying phenomena. For example, in surface-enhanced Raman scattering (SERS), deconvolution of complex SERS spectra is used for the detection and characterization of molecular species based on the position and intensity of extracted peaks [13]. In an epidemiological context, similar peak deconvolution methodologies are useful for the extraction of hidden features

from disease landscapes and provide a previously unexplored perspective into the progression of the COVID-19 pandemic.

Here, we introduce a peak deconvolution method to deconstruct previous waves of daily case-death trends into smaller sub-waves and isolate latent features of disease landscapes to track distinct and localized changes in disease progression over time. Our work demonstrates how peak deconvolution can be used to quantify epidemiological feature parameters, such as CDL, from the second wave of the COVID-19 pandemic in Israel, Canada, and Germany. We also present a variation of the susceptible-infected-recovered (SIR) model to empirically simulate peak fits taken from CDL analyses. We show how this parsimonious model can be used to quantify feature parameters of deconvolved sub-waves to better understand drivers of disease progression such as rates of infection, death, and recovery. Together, these tools serve as a novel method of epidemiological feature detection, with which each component pair of case-death peaks can be attributed to specific biological, behavioural, and social – among other – factors and determinants of disease. Finally, we discuss how these findings can inform researchers and policymakers with actionable insights regarding the outcomes of public health policies being implemented across various governments and administrative structures, the effectiveness of COVID-19 testing programs and intensive care infrastructure, as well as the spread of COVID-19 variants, among other epidemiological considerations.

## Chapter 2 | Literature Review

### Feature Detection

The concept of feature detection as it is applied here in an epidemiological context was inspired by techniques developed in the field of computer vision under the same name. In the field of computer vision, an image feature or keypoint is broadly defined as a pixel neighbourhood with a distinct and localized pattern, such as an edge or a corner, which conveys specific information about its unique position within the image. In this sense, image features are the parts of an image that are generally more likely to be identifiable within and across different images with similar depictions. Identifying the same features across multiple images is the basis for a range of computer vision techniques as these features allow images and the objects within them to be – among other applications – matched based on the relationships between their featural similarities [12]; transformed using methods of image registration and stitching [14]; as well as tracked from one frame to another in the case of motion and video (also known as optical flow) [15].

Features typically possess various properties, sometimes referred to as feature descriptors, which parametrize their distinctiveness. For instance, Lowe (1999 & 2004) famously introduced a method of feature detection called the scale invariant feature transform (SIFT), where each pixel neighbourhood identified as a feature is described by 4 parameters: position, scale, orientation, as well as an image gradient descriptor that summarizes the spatial structure of the feature [12, 16]. Many other similar approaches have since been developed under different names (e.g., SURF [17] and MSER [18]),

which use different methods of selecting and describing features to accomplish the same goal.

Features are not only useful for extracting and characterizing the components parts of an image, but also for inferring context from those parts. Features can be grouped into feature sets, which may be used as multi-dimensional inputs into more complex detection and classification algorithms based on machine learning and deep learning frameworks [19, 20]. For example, Viola and Jones (2001) used collections of image filters (i.e., features) to train a classifier for face detection [20], and Dalal and Triggs (2005) later trained a linear support vector machine (SVM) to detect pedestrians based on complex feature sets of test images, which included SIFT descriptors [19]. These examples show that feature detection, as it is used in computer vision and the broader field of computer science, is not only a power tool for deconstructing complex images into meaningful parts, but in many cases, it also forms the basis for reconstructing meaning from those parts.

While the use of featural representations is popular in computer science, the notion of features is broadly applicable to a variety of scientific disciplines, such as spectroscopy [13, 21] and chromatography [22], in which complex data is decoded into component parts. In SERS spectroscopy, spectra can be deconstructed into series of overlapping peaks, used for the detection and characterization of molecular species and biological materials based on feature descriptors such as peak positions, widths, and magnitudes. In many cases, spectra are taken from multiple parts of a sample to provide additional

information about the spatial localization of molecular species and properties of the tissues being studied [21]. For instance, this approach has been used in oncology to identify localized tissue features (i.e., biochemical compositions) associated with different stages of cancer progression such as cell proliferation, lipid reduction, and neovascularization [23]. This example demonstrates how spectral fingerprints can be used to form a hierarchy of features at various biological scales, which all serve to shed light on a larger system.

Many of the ideas regarding feature detection as they are applied to computer science and the other scientific disciplines discussed are also applicable to epidemiology. In the case of epidemiological data, the various data types which arise from a disease landscape – be it cases, deaths, hospitalizations, vaccinations, among many others – may each be interpreted as distinct snapshots of the same scene. Each of these data inputs can be treated as feature sets, which may be stratified spatially as well as temporally into features that reveal locally distinctive patterns within larger disease landscapes.

### **Epidemiological Models**

There are two main types of epidemiological models for studying disease outbreaks such as the COVID-19 pandemic [24], which include compartmental models [8, 25] as well as more computationally complex and data-intensive agent-based models (sometimes referred to as individual-based models) [26, 27]. While compartmental models assume homogeneously-mixed populations comprised of labeled, but otherwise

indistinguishable, hosts [28], agent-based models avoid aggregation assumptions and instead approximate heterogeneous populations as graph networks of interconnected nodes, where each node represents a distinct autonomous individual [29]. In agent-based models, the disease status of each agent is updated and stored independently as opposed to being processed in aggregate based on their compartmental status (e.g., susceptible, infected, recovered), and they also allow for the explicit representation of complex transmission networks [30]. For example, the COVID-19 Agent-based Simulator (Covasim) is an agent-based model, which takes into account country-specific demographics, disease transmission across various social levels including households, schools, and communities, as well as age-specific disease outcomes (among other considerations), to inform the unique qualities and interactions of each individual simulated by the model [31].

The notion of epidemiological features can be applied in the context of both compartmental and agent-based modelling frameworks. Both of these models can be adapted to study the spatiotemporal patterns (i.e., features) of disease landscapes with various degrees of resolution and granularity mainly by incorporating metapopulation approaches [32, 33]. In epidemiology, metapopulation models divide bulk populations into structured sub-populations, which are broadly grouped based on their ability to influence the dynamics of a disease landscape [28]. Subpopulations are often structured based on their geographic location at various scales (e.g., state, county, community); however, they may also be stratified using demographic descriptors such as age and

sex [34] as well as more elaborate population breakdowns such as travel networks [35] and contact patterns [36], which help define their influence on a disease landscape.

Metapopulation models help to deconstruct the spread of disease within and between subpopulations and often provide insights into distinct and localized trends that are overlooked when studying aggregated population data. For instance, Schüler and colleagues (2021) combined a metapopulation approach with a modified SEIRD compartmental model to study the localized spatial patterns of disease progression across Germany's 412 districts during the first wave of the COVID-19 pandemic [37]. They used this model to highlight specific disease hotspots in the South of Germany early in the outbreak and showed that simulation results at the district level could be aggregated to accurately reflect the cases reported at the state and national levels. In this example, the SEIRD models fitted at a district level could be interpreted as distinct features of the national disease landscape that provide otherwise hidden insights into the spatiotemporal patterns of disease progression.

A current example of an agent-based metapopulation network approach is the global epidemic and mobility model (GLEAM) [35, 38]. GLEAM structures subpopulations around more than 3800 major transportation hubs across roughly 230 countries (as of May 2022) based on real-world data. These subpopulations are interconnected based on the data-driven patterns of travellers moving between them. Chinazzi and colleagues (2021) demonstrated the application of this mobility network to assess the impacts of early travel restrictions in Wuhan, China on the spread of COVID-19 to the rest of

mainland China and the world [39]. They showed that strict travel quarantine measures reduced international case importations from Wuhan by 80% for almost a month after they were first implemented in late January 2020. This study provides insights into how spatial representations generated from mobility data can serve to define and interrelate features of the disease landscape at a provincial, national, and international level.

There are also spatial representations other than metapopulations, which generally involve compartmental models defined by systems of diffusion-mediated partial differential equations (PDEs) [40-43]. Unlike metapopulation approaches which utilize discrete subpopulations, PDE models provide continuous spatiotemporal descriptions of disease landscapes which can be further informed by real-world considerations such as geographical boundaries and population distributions [43]. For example, Viguerie et al. (2021) used a SEIRD-based model of PDEs to study the Italian region of Lombardy and showed that lockdown restrictions should be tailored to local population densities to optimize the success of reopening strategies [42]. In a more recent paper, the same research group also showed how the PDE models could be discretized and aggregated at various scales to consolidate spatially-localized patterns such as hotspots, similar to the analyses performed by Schüller et al. (2021) for Germany [37].

Like any model, each type and variation of epidemiological model comes with simplifying assumptions and sources of uncertainty and needs to be carefully considered within the context in which it is applied [44]. In this sense, complexity does not guarantee model accuracy, nor is it a necessary feature of accurate models. As more advanced network

models are used, the need for greater computational resources and detailed input data increases [29]. Although simpler models are often limited by their inability to account for the effects of complex human interactions and spatial mixing patterns on disease dynamics, these limitations also make them less sensitive to misestimations of parametric assumptions which dominate more complex models [44]. In many cases, the underlying biology of disease dynamics, which includes incubation period, contagious period, and transmissibility, likely plays a larger role in determining the accuracy of epidemiological models compared to explicit considerations of spatial or social heterogeneities between individuals or structured subpopulations [45]. Such considerations form the basis of even the simplest compartmental models.

Metapopulations and other spatially-heterogenous approaches to studying infectious disease dynamics provide comprehensive frameworks for explicitly representing spatiotemporal features of disease landscapes. However, these approaches require detailed input data, which has the potential to bias model results and compound model complexity. For instance, Cooper (2006) discussed the tendency of mobility networks based on commuting data to overestimate smallpox transmission as the spread of disease requires close contact that survey data showed was less likely to occur at work compared to at home [45]. Similarly, for GLEAM, the use of airport networks to study the spread of COVID-19 ideally requires additional data regarding local travel restrictions, testing and vaccination requirements, as well as quarantine guidelines for return travelers. While close attention to the cascade of considerations surrounding network-based disease dynamics may mitigate misestimations, it is also important to consider

other means of deconstructing disease landscapes for situations in which limited data or complex assumptions make these models more difficult to use.

Lastly, these models mainly provide broad representations of spatially constrained time-series trends, but it is less clear how they can be used to study temporally localized variations within these trends without defaulting to complex modelling scenarios. Ascribing time-based features to models based on real-world data, analogously to structuring subpopulations based on geography or demography, could provide deeper insights into retrospective studies of disease dynamics and inform new methods of forecasting disease transmission. Time-based features are especially relevant to mapping relationships between compartmental statuses such as cases, hospitalizations, recoveries, and deaths, which manifest asynchronously across disease landscapes. Time-based segmentation of epidemiological data prior to modelling is considered more in depth throughout this thesis.

### **SIRD Model**

The susceptible-infected-recovered-deceased (SIRD) model is a variation of the classic SIR model which explicitly accounts for disease-induced deaths among infected populations. In the SIRD model, infected individuals either recover from disease with natural immunity or die due to infection. The progression of the disease landscape in this model is represented by a set of four ordinary differential equations which each describe the dynamics of one of the SIRD populations as follows:

$$\frac{dS}{dt} = -\beta IS, \quad \frac{dI}{dt} = \beta IS - \gamma I - \mu I, \quad \frac{dR}{dt} = \gamma I, \quad \frac{dD}{dt} = \mu I \quad [1]$$

Here,  $\beta$  is the transmission rate constant,  $\gamma$  is the recovery rate constant, and  $\mu$  is the mortality rate constant. Unlike the SIR model, the SIRD model includes an explicit analytical solution for the deceased population,  $D$ , which enables simultaneous modelling of death data in addition to case data. Cases and deaths are among the most commonly reported epidemiological data worldwide and serve as the basis for our CDL analyses.

The SIRD model does not explicitly consider many of the complexities surrounding disease progression, such as incubation periods, asymptomatic transmission rates, and restrictive public health measures. Additional model compartments, such as an exposed (E) compartment used in SEIRD models of COVID-19 dynamics, are omitted from this analysis for simplicity. In many cases, simpler models have been shown to be sufficient for short term modelling of COVID-19 trends on the scale of several days to a couple months [46] as well as observationally equivalent to their more complex counterparts at these timescales [2, 47].

Vital dynamics, which include births and natural deaths, are not considered in the SIRD model. These processes are generally used for modelling endemic diseases, which persist in populations over longer periods of time upwards of a decade [48]. With the SIRD model, we assume that natural deaths are negligible compared to those caused

by disease and similarly assume that birth processes are insignificant to overall population dynamics at the timescales presented.

We also assume that natural immunity is conferred to recovered individuals at least temporarily such that they are not reintroduced into the susceptible population following recovery. Immune memory has been shown to persist for upwards of 3 months in most individuals following COVID-19 infection [49, 50] with little evidence of reinfection within similar timeframes [51]. In our analysis, each pair of case-death peak fits are simulated with a unique SIRD model and typically span between 10-15 days of the 105-120 days over which most second waves are observed. Given these considerations, the simplifying assumptions of the SIRD model without vital dynamics are reasonable for the timescales being studied.

### **Case-Death Latency**

One epidemiological parameter which has remained central to calibrating and fitting pandemic models of disease progression is the apparent temporal lag between publicly reported daily COVID-19 cases and deaths, a phenomenon which we term “case-death latency” (CDL). CDL is in part determined by the inherent dynamics of the SARS-CoV-2 virus [52] as well as by the unique physiological responses it imposes on each infected individual [53]. However, it is also influenced by other factors such as the quality of public health infrastructure, and centralized disease reporting and management [54, 55]. While this latency is reported in the literature [56] and estimated to range on average between 13-16 days from the onset of COVID-19 symptoms [57, 58], little has been done to

quantify variations in CDL over time or consider these variations within a broader epidemiological framework.

Oversimplification of CDL can misinform calculations of important epidemiological metrics such as case fatality rate (CFR), which is currently calculated based on crude estimates of CDL, if any [59]. Given that it is influenced by many confounding – and often unknown – variables and used to estimate important epidemiological metrics, CDL serves as a rich and complex metaparameter which implicitly codes for a variety of factors and determinants of disease. Decoding CDL could provide new insights into COVID-19 disease landscapes.

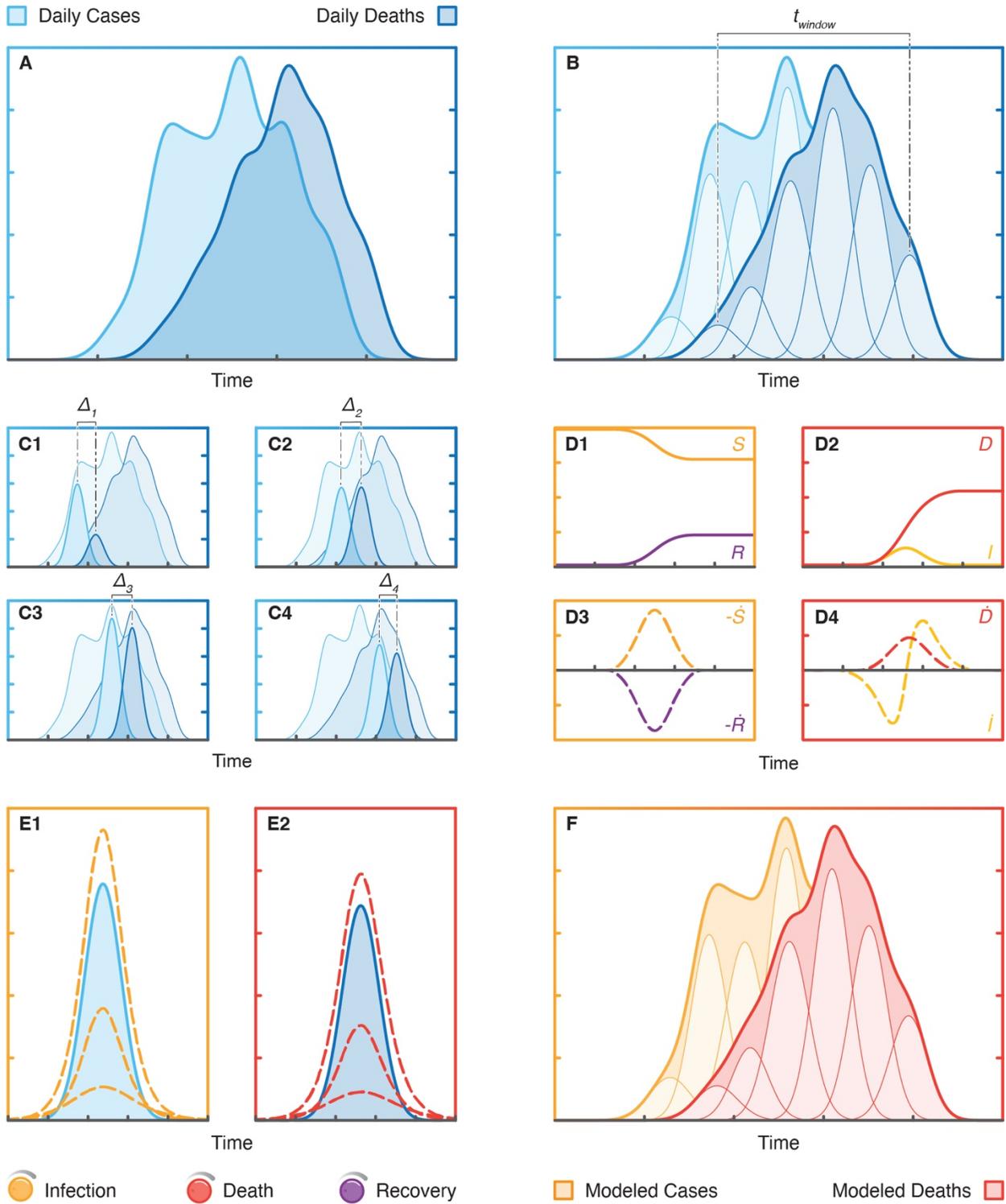
## **Chapter 3 | Methods**

### **Data Selection and Processing**

The COVID-19 case and death data used in this manuscript was downloaded from the Johns Hopkins Coronavirus Resource Center's open access database [60] and subsequently analyzed using our novel approach to epidemiological feature detection, which is outlined in Figure 1. Prior to analysis, all data was smoothed in MATLAB using a gaussian-weighted moving average filter with a window length of 25 days to obtain quasi-continuous trends, as shown in Figure 1A, which were more suitable for peak deconvolution than the noisy raw data. Pairs of corresponding case-death trends for each region of interest were min-max normalized between 0 and 1 to maintain consistency in our analyses across these regions.

The regions and waves of interest we analyzed were mainly selected for the purposes of demonstrating this novel approach and follow a few general selection criteria. Firstly, regions of interest where the daily reporting of cases and deaths was relatively consistent and showed distinct surges were preferred as they facilitated data smoothing and peak deconvolution. Higher income countries reporting peak surges upwards of 5000 cases per day generally fit these criteria often due to the greater capacities of their healthcare networks to document disease progression.

The second wave of the pandemic in each region of interest was selected for analysis based on the assumption that COVID-19 cases and deaths were more accurately reported following the first wave due to constant improvements in testing and reporting infrastructure through the early stages of the pandemic. Additionally, the second wave was chosen instead of later waves to investigate COVID-19 disease landscapes prior to widespread vaccines rollouts. Although these selection criteria were used to guide the present analysis, this approach to epidemiological feature detection is broadly applicable to the study of disease landscapes beyond nationally reported surges in COVID-19 cases and deaths, including regions with sparser disease reporting.



**Figure 1** Epidemiological feature detection methodology: **A** reported daily cases and deaths data for a region of interest are cleaned and smoothed, **B** case and death trends are deconstructed into component peaks, **C** peak pairs are isolated for SIRD modelling, **D** time derivatives for analytical solutions of SIRD model are used to model peaks, **E** SIRD model parameters are optimized to fit each peak pair, **F** case and death trends are reconstructed using modelled peak pairs.

## Peak Deconvolution

Peak deconvolution was used to deconstruct national time series trends from the second wave of the COVID-19 pandemic into their component peaks, or sub-waves, as shown in Figure 1B. Following data selection and processing in MATLAB, peak deconvolution was performed using Fityk, an open source curve fitting and data analysis software [61]. The quality of the fit for each of the deconvoluted time series trends was assessed based on their coefficient of determination,  $R^2$ , where an  $R^2$  value greater than 0.99 was considered acceptable for this analysis.

To fit corresponding case-death surges within a region of interest, time windows of the same length for each data type were offset to account for the average CDL between trends. The offset between case and death time windows was defined by maximizing the cross-correlation function between case-death surges. Although time windows were constrained to be the same length for the case and death trends within each region of interest, they were allowed to vary between regions as the duration of the second wave was unique to each region studied.

For the purposes of peak deconvolution, the time window ultimately defined the distance between the center of the first and last deconvoluted peaks flanking case-death surges, which, for example, is indicated in Figure 1B as  $t_{window}$  for the trend in daily deaths. Each trend can be fit with at least 3 peaks (i.e., two flanking peaks and 1 unconstrained peak), and the number of peaks can be increased to enhance the resolution of feature extraction. Corresponding case-death trends within a region of interest must be fit with

the same number of peaks to ensure correspondence between individual features extracted from the case and death landscapes. For instance, Figure 1B shows a pair of case-death trends fit with 6 peaks each, of which the 4 central peaks from each trend are shown as isolated peak pairs in Figure 1C. Each of the trends in this thesis were deconvoluted into 6 peaks, which proved to be the minimum number of peaks that accurately fit all these trends (i.e., the number of peaks was limited by the most feature-rich landscapes). Consistent peak fitting across each region of interest provided uniformity of model outputs, which facilitated comparisons of regional disease progression across these regions.

Each isolated peak pair comprises a case and death peak of identical widths and represents a unique epidemiological feature. Peak widths may vary across isolated peak pairs but are kept constant between the case and death peaks of each individual pair to consolidate the temporal parameterization of each feature. CDL is defined as the center-to-center distance between the case peak and death peak of each isolated peak pair, which is indicated as  $\Delta$  in each of the subplots in Figure 1C. The peak deconvolution process yields a sequence of distinct and localized features from real-world data that can be further parameterized with compartmental models of disease transmission.

## **SIRD Modelling**

While the case and death data for peak deconvolution are shown as the change in daily counts (i.e., new cases and deaths), the explicit analytical solutions of the SIRD model are expressed as total daily counts as shown in Figures 1D1 and 2. To convert the

population dynamics of the model to daily changes in the SIRD populations, we take the numeric derivative of the model's analytical solutions using the forward difference formula, which are shown in Figures 1D3 and 4. New daily cases are approximated using the negative time derivative of the susceptible population,  $S$ , given that changes in  $S$  are governed by a single term as shown in **Eq. 1**, which describes the rate of irreversible transition between susceptible and infected populations. Analogously, the time derivative of the deceased population,  $D$ , is used to approximate new daily deaths.

Modelled cases and deaths are fit to pairs of reported case-death peaks using non-linear least squares regression, which is illustrated in Figure 1E. The optimization process works to minimize the residual sum of squares between the model and reported data over the time interval spanning the case-death peak pair of interest. Minimization of the model error is defined by the objective function,  $f(x)$ , as follows:

$$\min \|f(x)\|^2 = \min \left[ (C_d - C_m(x))^2 + \lambda (D_d - D_m(x))^2 \right] \quad [2]$$

where  $C$  and  $D$  represent case and death peaks, respectively, and are denoted with a  $d$  for deconvoluted peaks or  $m$  for SIRD-modelled peaks. Here,  $x$  is a vector variable of the parameters and initial conditions used in the SIRD model and  $\lambda$  is a scaling factor intended to compensate for the general disparity in the magnitude of cases compared to deaths. Model parameters include the transmission, recovery, and mortality rate constants while initial conditions include the initial susceptible, infected, and recovered

populations. For simplicity, we constrained initial recovered populations to zero and kept initial infected populations arbitrarily low compared to their corresponding initial susceptible populations (i.e., a ratio of approximately 1:100,000 infected to susceptible individuals). The initial susceptible population of each peak pair model was scaled proportionally to the relative area occupied by its death peak as mortality was assumed to be a more reliable indicator of disease prevalence than cases due to variability in testing rates within and across regions of interest throughout the pandemic [62].

Each pair of modelled case-death peaks represents a unique SIRD simulation generated from a single set of optimized parameter values (i.e.,  $\beta$ ,  $\gamma$ , and  $\mu$ ), which ensures that corresponding case-death data for each isolated peak pair is satisfied under the same epidemiological conditions. Model parameter values of different peak pairs are independent of one another. During the simulation process, modelled peaks are also temporally shifted to be aligned with the time series data and ultimately result in a purely mathematical recreation of the original time series trends as illustrated in Figure 1F.

## **Chapter 4 | Results**

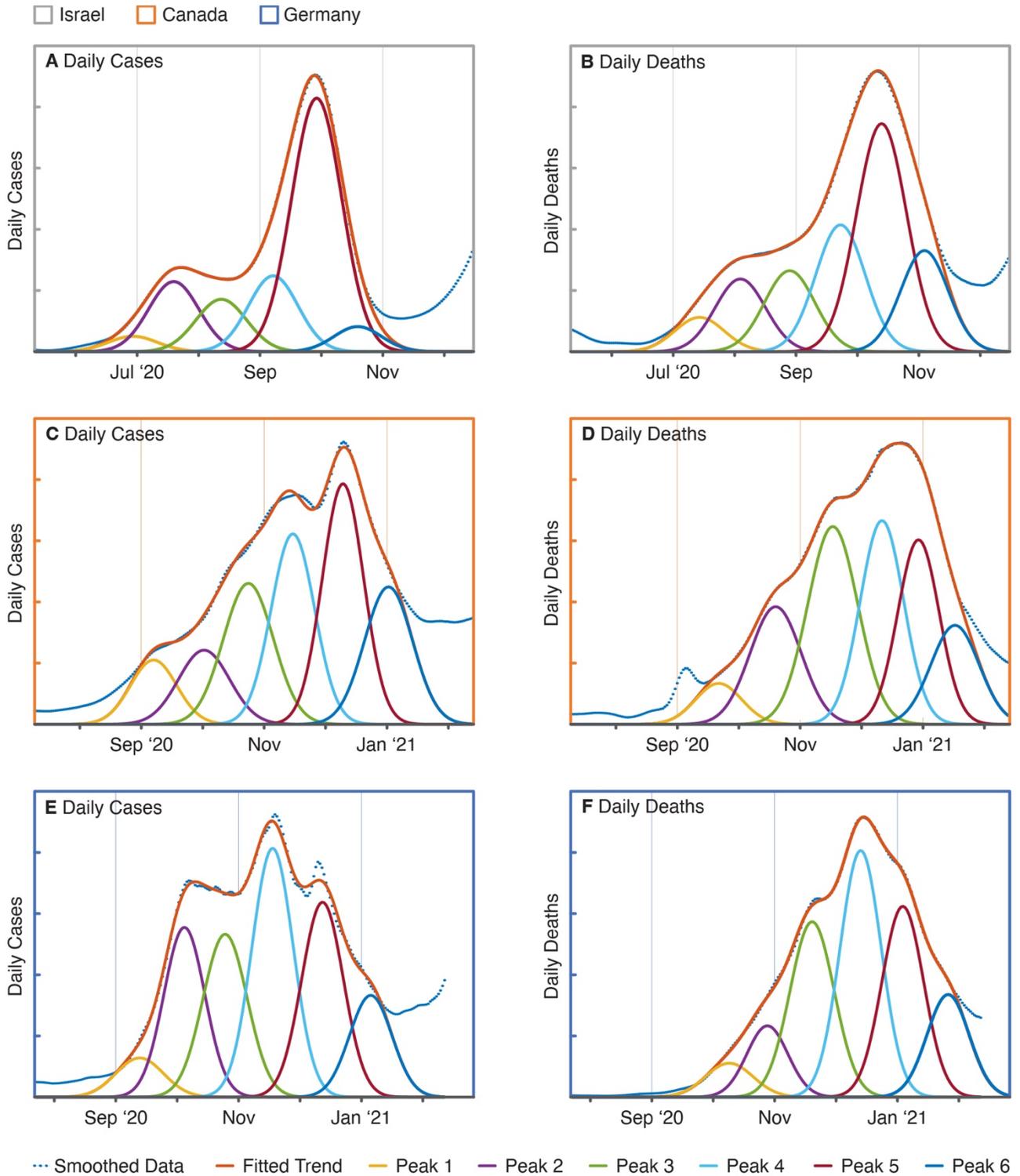
### **Peak Analysis**

Nationally reported data from the second waves of the COVID-19 pandemic in Israel, Canada, and Germany were analyzed using the presented peak deconvolution and SIRD modelling methodology to investigate the relationship between CDL and other epidemiological parameters shaping these disease landscapes. Daily case and death

trends for each country were fit with 6 peaks as shown in Figure 2. While the scale of the x-axes is the same for each of the subplots, Israel's second wave started approximately 100 days before those of Canada and Germany.

The durations of the second waves of cases and deaths, which were calculated as the number of days between the centers of the two peaks flanking each fit (i.e., peaks 1 and 6), were between 110 and 115 days for all three countries. The consistency of the second wave durations across these countries is suggestive of an epidemiologically invariant phenomenon underlying these transient surges in the spread of COVID-19. Despite this similarity, the deconvoluted case and death trends shown in Figure 2 depict a unique disease landscape for each country within each of their second waves. The surge in Israel's daily cases shown in Figure 2A has a backloaded bimodal distribution, whereas Canada's cases (Figure 2C) increase steadily; both of which are distinct from Germany's sudden and relatively sustained spike in daily cases (Figure 2E). As expected, surges in daily deaths shown in Figures 2B, D, and F lag behind their respective surges in daily cases and also bare shaped-based resemblances to them.

The similarity between the respective trends in second wave cases and deaths was confirmed and compared by calculating the Pearson correlation coefficients and dynamic time warping (DTW) minimum distances for each of the pairs of case-death time series shown in Table 1. These similarity measures were calculated using independently normalized case and death trends, which were temporally shifted to maximize their



**Figure 2** Peak fits for national daily cases and deaths reported from the second wave of the COVID-19 pandemic: **A** Israel daily cases, **B** Israel daily deaths, **C** Canada daily cases, **D** Canada daily deaths, **E** Germany daily cases, and **F** Germany daily deaths.

cross-correlation functions. The Pearson correlation coefficients for each country are greater than 0.95, which indicates a strongly positive correlation between the national case and death trends of each country. The similarity between respective trends confirms a level of consistency between the reporting of case and death data within each of these countries required to fit and model the same number of analogous case-death peaks in the analyses presented. Additionally, the Pearson correlation coefficients and DTW minimum distances both show that Canada’s case-death trends are most similar, followed by those of Israel and then Germany.

**Table 1** Correlation between the deconvoluted trends in daily cases and deaths in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic.

<b>Country</b>	<b>Pearson Correlation Coefficient</b>	<b>Dynamic Time Warping Minimum Distance</b>
Israel	0.977	1.332
Canada	0.991	1.173
Germany	0.961	1.946

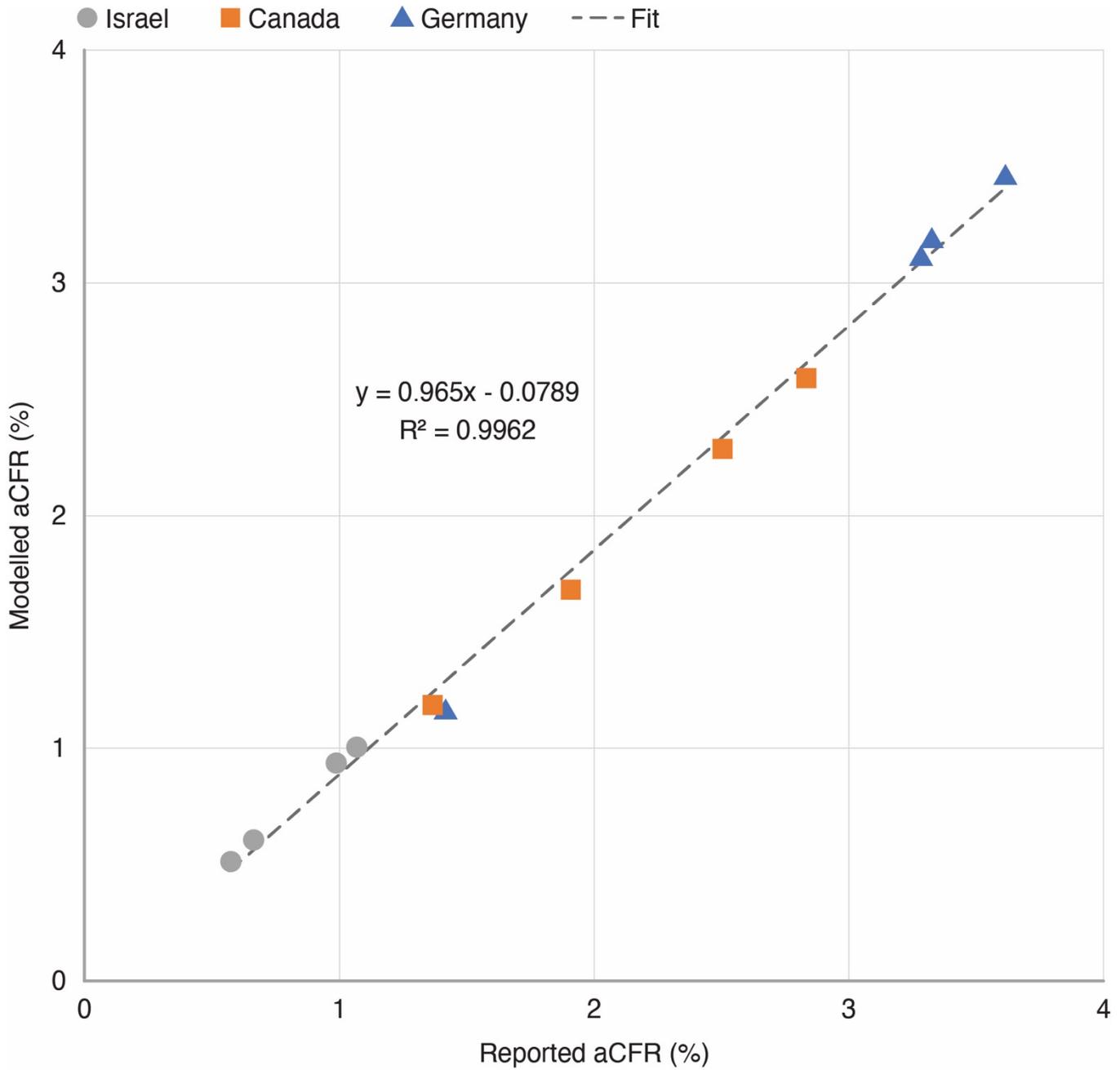
### **Model Validation**

To validate the SIRD model, we first calculated and compared the adjusted case fatality rate (aCFR) for each of the deconvoluted and modelled peak pairs. CFR is defined as the ratio of deaths to cases over a specified period of time while aCFR accounts for CDL and is defined as the ratio of deaths to cases where the deaths are temporally offset by their respective latency. This epidemiological measure is generally reported as a percentage, where larger percentages are indicative of more severe disease outcomes. The majority of nationally reported CFRs for COVID-19 based on aggregate data range

between 0.5-5.0%: Israel reporting 0.6%, Canada 1.7%, and Germany 2.2%, as of October 2021. For the deconvoluted peak pairs, aCFRs were calculated by dividing the area under each death peak by that of its corresponding case peak. The corresponding rates for each modelled peak pair were similarly calculated using their respective SIRD model parameters as follows:

$$\text{Modelled CFR} = \frac{\mu I}{\beta I S} \approx \frac{\mu}{\beta S_0} \quad \text{where } I \ll S \quad [3]$$

Here, in addition to the rate constants,  $\beta$ ,  $I$ , and  $\mu$ , introduced in **Eq. 1**,  $S_0$  is the initial susceptible population for each modelled peak pair. Figure 3 provides a comparison of the reported and modelled aCFRs within and between each country. As expected, the plot shows a strong linear correlation ( $R^2$  greater than 0.99) with an equal proportionality (slope of 1.0), which demonstrates that the aCFRs calculated for each of the isolated peak pairs are closely and consistently approximated by their respective SIRD model parameters. Figure 3 also shows a clear separation between the reported second wave disease landscapes of each of these countries, where Israel has the lowest aCFRs on average, followed by Canada, and then Germany. Moreover, the variation in aCFRs between peak pairs from the same country highlights the potential for peak deconvolution methodologies to identify increasingly granular features of these disease landscapes.

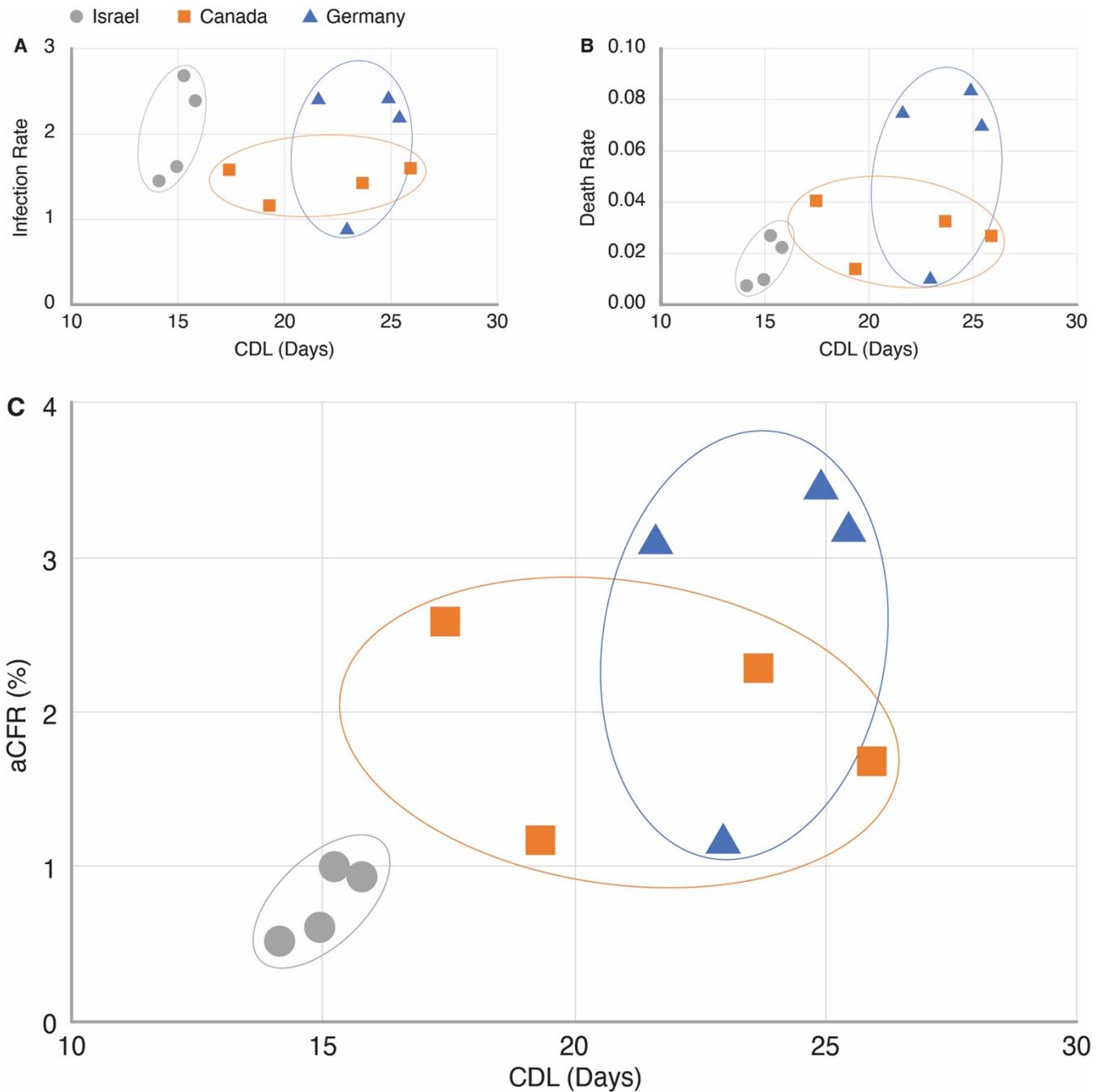


**Figure 3** Reported peak-by-peak aCFRs compared to aCFRs calculated from SIRD model parameters in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic.

## Case-death Latency

While aCFRs provide an overview of these disease landscapes, studying their underlying infection and death rates enables a deeper understanding of the epidemiological parameters most heavily influencing each country's pandemic response. Figure 4 shows the modeled infection, death, and aCFRs versus CDLs for each of the isolated peak pairs for Israel, Canada, and Germany. The plot of the recovery rates versus latencies is not shown as it closely resembled that of the infection rates shown in Figure 4A, given that more than 95% of reported cases led to non-fatal disease outcomes across all three countries. In Figure 4, peaks pairs – each represented by an individual data point – are uniquely defined within each plot and clustered by country, which illustrates the dynamic variability of these disease landscapes even within a single wave of the COVID-19 pandemic. In Germany, such time dependent variations in aCFRs during the second wave outbreak have been linked to changes in the age distribution of confirmed cases [63].

Between countries, death rates (Figure 4B) varied significantly more than infection rates (Figure 4A). Based on the weighted average rates for each country reported in Table 2, national death rates showed a 76% deviation from the average, whereas infection rates, which showed an 18% deviation, were relatively constant. These observations indicate that higher death rates were the main drivers of larger aCFRs across these disease landscapes during the second wave of the COVID-19 pandemic. They also suggest that the risk of COVID-19 infection was relatively independent of country and time.



**Figure 4** CDL compared to modelled infection rates (A), death rates (B), and aCFRs (C), in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic.

Studying case death latency provided further segmentation of these epidemiological parameters by country. For the three countries presented, we observed that the average death rates and aCFRs were directly proportional to CDL. In particular, the average

aCFRs showed a strong linear correlation ( $R^2 = 0.91$ ) to the latencies reported in Table 2, which is also illustrated in Figure 4C. Overall, this trend is suggestive of a more complex relationship between disease severity and the temporal dynamics of disease progression and reporting both at an individual and societal level, which needs to be considered when comparing aCFRs across multiple countries.

**Table 2** Weighted averages of CDLs, infection rates, and death rates calculated from peak fits and SIRD model parameters in Israel, Canada, and Germany during the second wave of the COVID-19 pandemic.

Country	CDL (Days)	Infection Rate	Death Rate	aCFR (%)
Israel	14.631	1.791	0.013	0.661
Canada	21.997	1.404	0.026	1.776
Germany	23.827	2.024	0.062	2.801

## Chapter 5 | Discussion

To further investigate the strongly positive correlation between CDL and adjusted case fatality rate across Israel, Canada, and Germany during the second wave of the COVID-19 pandemic, multiple underlying factors must be considered. One of the main factors governing the trends in the disease landscapes across these three countries are the governments themselves [64]. The governments of Israel, Canada, and Germany each embody unique administrative structures, which influence the unity of their pandemic responses at local, state, and federal levels. Based on our analyses, we propose a working theory that governments with increasingly decentralized and codependent

administrative structures are more likely to experience worse disease outcomes due to a lack of decisive policy making and timely access to systematic healthcare data.

While the degree of government centralization in each of these countries is difficult to quantify, broad comparisons of their different administrative structures and subsequent pandemic responses – informed by factors including population size, distribution, and segmentation – provide insights into the correlation across countries between CDLs and aCFRs. For example, Israel has a population of 9 million people living across 6 districts within a unitary state of centralized federal governance. Under this government, laws and public health policies are exclusively implemented at the federal level (e.g., countrywide lockdowns). During the second wave of the pandemic, Israel opted to decentralize the management of COVID-19 to the country's four universal health plans, which oversee the administration of primary care services such as testing and patient education [65], and municipalities had the option to implement additional health measures based on a classification system of local disease severity set forth by Israel's Ministry of Health [66]. However, these health plans ultimately belong to a system of direct oversight by the State, which resulted in a nationally homogeneous second wave pandemic response and lower CDLs as well as aCFRs in Israel compared to those of Canada and Germany.

Canada and Germany are both governed as federations, which operate on a spectrum of shared power distributed between state and federal levels. In Canada, which has a population of 38 million people living across 10 provinces and 3 territories, neither provincial governance nor federal jurisdiction are subordinate to the other. Instead,

Canada's provincial and federal governments act autonomously to exercise their respective constitutional responsibilities and only coordinate policies through voluntary negotiations where there is mutual interest in intergovernmental collaboration [67]. Based on this system of governance, the early stages of the pandemic in Canada saw mask mandates and regional lockdowns implemented at the provincial level while the federal government maintained national border closures as well as international travel restrictions including mandatory quarantining for return travellers [68, 69]. Canada's heterogeneous second wave pandemic response, which was highlighted by provincial autonomy and cooperative national decision making, led to moderate aCFRs with the largest variation in CDLs compared to Israel and Germany.

Germany, which has a population of 83 million people living across 16 states (known as Länder), responded in much the same way as Canada with the main differences arising from Germany's consensus-based federal system. With lawmaking abilities predominantly resting in the states' hands and requiring unanimity from state leaders [67], Germany's second wave pandemic response was guided by fragmented governance and fractious federal-state relations. Despite the implementation of delayed and compromising nationwide lockdowns, rules around local social restrictions, face masks, and other public health policies varied across each Länder [70]. The decentralized healthcare systems of Germany's states have presented a heterogeneous response throughout the pandemic, which has ultimately resulted in substantial differences in disease landscapes across the country, especially during the second wave of the pandemic [71]. Of the three countries presented, Germany showed the largest

CDLs and aCFRs during their second wave. However, the range of CDLs in Germany's second wave (21.6 to 25.4 days) was smaller than that of Canada's (17.4 to 25.9 days), which may reflect Germany's more unified response across Länder.

Decentralized governments, such as the federal systems in Canada and Germany, tend to exhibit patchwork responses to large scale public health outbreaks [72]. Non-standardized disease reporting across these heterogeneous policy and data collection landscapes often leads to a lack of timely and systematic healthcare data, which can hamper pandemic responsiveness [70]. In Germany, for example, delays in registered COVID-19 deaths (i.e., from the date of death to date of publication) of one to three weeks were common throughout the second wave of the pandemic [55]. Overall, the decentralization of COVID-19 management may be a contributing factor to the higher CDLs and aCFRs observed during the second waves of the pandemic in Canada and Germany compared to Israel.

For Germany, the codependency of state governments to enact policies at a national level may have also contributed to the country having the highest CDLs and aCFRs of the three countries analyzed. However, timely pandemic responses are important to smaller unified governments. For example, Israel's more severe second wave has been attributed to delayed government action during the early weeks of the country's second wave outbreak [73]. Ultimately, unified governments and healthcare networks are more likely to effectively respond to emerging outbreaks and report on them in a way that reflects their true disease landscapes, whereas decentralized networks experience

greater latencies, which are more prone to underestimates in disease severity and costly delays in implementing public health interventions and countermeasures.

While administration and governance are overarching factors which have undoubtedly contributed to the progression of the pandemic, CDLs and aCFRs are also influenced by many other confounding factors, especially at subnational levels. Such factors may include the emergence of different variants of concern throughout the second wave of the pandemic, which are associated with a higher risk of mortality compared to earlier variants for cases running longer clinical courses (i.e., more than 2 weeks since diagnosis) [74]. Population level factors such as age demographics are also important to consider, as higher case fatality rates are disproportionately observed among older populations [75] while the relationship between age and CDL is relatively unexplored [76]. Other socioeconomic factors such as access to healthcare resources and GDP per capita also play a significant role in shaping these disease landscapes [77]. These additional factors and determinants of disease are areas of interest for future work, especially as relevant stratified data for each of these considerations continues to become more available.

As the pandemic progresses, new types of data also become available, allowing for further analyses of emerging waves. For instance, the methodologies introduced in this work could be used to analyze subsequent waves of COVID-19 for various regions. Analyses of third wave data could provide unique insights on the early impacts of

vaccinations, which could be studied alongside cases and deaths by applying compartmental models that account for vaccinal and waning immunities [3].

Furthermore, analyzing additional regions at various scales would allow for more conclusive findings to be drawn regarding the methodology presented here. Analyzing and comparing smaller regions, such as cities, provinces, or states, could provide more detailed insights into the spatiotemporal characteristics of disease progression, such as when individuals travel from one homogenous region to another. These analyses could additionally be used to explore temporal features of disease landscapes previously studied using mobility networks and metapopulation approaches [38, 39]. Ultimately, the versatility of the epidemiological feature detection method presented here makes it applicable to a wide variety of applications within epidemiology.

## **Chapter 6 | Conclusions**

In this thesis, we introduced and applied a novel latent mixture approach, which we coin epidemiological feature detection, to analyze the second wave of the COVID-19 pandemic in Israel, Canada, and Germany. Applying this approach, we used peak deconvolution methods to extract and relate distinct and localized features from trends in daily cases and deaths, and we further characterized these features using a SIRD model of disease transmission to quantify spatiotemporal variations in epidemiological parameters including infection, death, and recovery rates. We found that the average death rate across all three countries varied more than 4 times as much compared to the

average infection rate, which suggests that higher death rates, as opposed to lower infection rates, are the main drivers of increases in adjusted case fatality rates. Additionally, we found a strongly positive correlation ( $R^2 = 0.91$ ) between average adjusted case fatality rate and the lag time between reported cases and deaths of isolated features, which we term case-death latency. Of the three countries presented, Israel showed the lowest average case-death latency and adjusted case fatality rate (14.6 days and 0.7%), followed by Canada (22.0 days and 1.8%), and Germany (23.8 days and 2.8%). We further discuss this trend in the context of increasingly decentralized governments and administrative structures in these respective countries. We highlight the importance of cooperative decision making and timely access to systematic healthcare data for effective responses to emerging public health outbreaks. Overall, this work emphasizes the need for new empirical approaches to complement ongoing pandemic scenario planning initiatives and illustrates the potential for epidemiological feature detection to improve health security and pandemics preparedness for COVID-19 and beyond.

## References

- [1] N. Ferguson *et al.*, "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," 2020.
- [2] A. L. Bertozzi, E. Franco, G. Mohler, M. B. Short, and D. Sledge, "The challenges of modeling and forecasting the spread of COVID-19," *Proceedings of the National Academy of Sciences*, vol. 117, no. 29, pp. 16732-16738, 2020. [Online]. Available: <https://escholarship.org/content/qt2fk7664g/qt2fk7664g.pdf?t=qeu4jk>.
- [3] C. M. Saad-Roy *et al.*, "Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years," *Science*, vol. 370, no. 6518, pp. 811-818, 2020. [Online]. Available: [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7857410/pdf/370\\_811.pdf](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7857410/pdf/370_811.pdf).
- [4] C. J. Whitty, "What makes an academic paper useful for health policy?," ed: Springer, 2015.
- [5] S. Funk, A. Camacho, A. J. Kucharski, R. Lowe, R. M. Eggo, and W. J. Edmunds, "Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15," *PLoS computational biology*, vol. 15, no. 2, p. e1006785, 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6386417/pdf/pcbi.1006785.pdf>.
- [6] M. Nicola *et al.*, "Health policy and leadership models during the COVID-19 pandemic-review article," *International Journal of Surgery*, 2020.
- [7] D. Adam, "Special report: The simulations driving the world's response to COVID-19," *Nature*, vol. 580, no. 7802, pp. 316-319, 2020.

- [8] J. Tolles and T. Luong, "Modeling epidemics with compartmental models," *Jama*, vol. 323, no. 24, pp. 2515-2516, 2020.
- [9] G. Chowell, "Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts," *Infectious Disease Modelling*, vol. 2, no. 3, pp. 379-398, 2017. [Online]. Available:  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5726591/pdf/main.pdf>.
- [10] M. D'Arienzo and A. Coniglio, "Assessment of the SARS-CoV-2 basic reproduction number,  $R_0$ , based on the early phase of COVID-19 outbreak in Italy," *Biosafety and Health*, vol. 2, no. 2, pp. 57-59, 2020.
- [11] J. Friedman *et al.*, "Predictive performance of international COVID-19 mortality forecasting models," *Nature communications*, vol. 12, no. 1, pp. 1-13, 2021.
- [12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [13] Z. J. Smith, C.-W. Chang, L. S. Lawson, S. M. Lane, and S. Wachsmann-Hogiu, "Precise monitoring of chemical changes through localization analysis of dynamic spectra (LADS)," *Applied spectroscopy*, vol. 67, no. 2, pp. 187-195, 2013.
- [14] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International journal of computer vision*, vol. 74, no. 1, pp. 59-73, 2007.
- [15] C. Tomasi and T. Kanade, "Detection and tracking of point," *Int J Comput Vis*, vol. 9, pp. 137-154, 1991.

- [16] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, 1999, vol. 2: IEEE, pp. 1150-1157.
- [17] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, 2006: Springer, pp. 404-417.
- [18] P.-E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *2007 IEEE 11th International Conference on Computer Vision*, 2007: IEEE, pp. 1-8.
- [19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 2005, vol. 1: IEEE, pp. 886-893.
- [20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, 2001, vol. 1: IEEE, pp. I-I.
- [21] H. J. Butler *et al.*, "Using Raman spectroscopy to characterize biological materials," *Nature Protocols*, vol. 11, no. 4, pp. 664-687, 2016/04/01 2016, doi: 10.1038/nprot.2016.036.
- [22] W. Zhang, J. Chang, Z. Lei, D. Huhman, L. W. Sumner, and P. X. Zhao, "MET-COFEA: a liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation," *Analytical chemistry*, vol. 86, no. 13, pp. 6245-6253, 2014.

- [23] M. Bergholt *et al.*, "In vivo diagnosis of esophageal cancer using image-guided Raman endoscopy and biomolecular modeling," *Technology in cancer research & treatment*, vol. 10, no. 2, pp. 103-112, 2011.
- [24] H. Rahmandad and J. Sterman, "Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models," *Management science*, vol. 54, no. 5, pp. 998-1014, 2008.
- [25] F. Brauer, "Compartmental models in epidemiology," in *Mathematical epidemiology*: Springer, 2008, pp. 19-79.
- [26] S. Eubank *et al.*, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180-184, 2004.
- [27] P. Maheshwari and R. Albert, "Network model and analysis of the spread of Covid-19 with social distancing," *Applied network science*, vol. 5, no. 1, pp. 1-13, 2020.
- [28] F. Ball *et al.*, "Seven challenges for metapopulation models of epidemics, including households models," *Epidemics*, vol. 10, pp. 63-67, 2015.
- [29] M. Ajelli *et al.*, "Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models," *BMC infectious diseases*, vol. 10, no. 1, pp. 1-13, 2010.
- [30] J. Gomez, J. Prieto, E. Leon, and A. Rodríguez, "INFEKTA—An agent-based model for transmission of infectious diseases: The COVID-19 case in Bogotá, Colombia," *PloS one*, vol. 16, no. 2, p. e0245787, 2021.
- [31] C. C. Kerr *et al.*, "Covasim: an agent-based model of COVID-19 dynamics and interventions," *PLOS Computational Biology*, vol. 17, no. 7, p. e1009149, 2021.

- [32] I. Hanski, *Metapopulation ecology*. Oxford University Press, 1999.
- [33] M. J. Keeling, O. N. Bjørnstad, and B. T. Grenfell, "Metapopulation dynamics of infectious diseases," in *Ecology, genetics and evolution of metapopulations*: Elsevier, 2004, pp. 415-445.
- [34] G. C. Calafiore and G. Fracastoro, "Age structure in SIRD models for the COVID-19 pandemic—A case study on Italy data and effects on mortality," *Plos one*, vol. 17, no. 2, p. e0264324, 2022.
- [35] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21484-21489, 2009.
- [36] J. Zhang *et al.*, "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China," *Science*, vol. 368, no. 6498, pp. 1481-1486, 2020.
- [37] L. Schöler, J. M. Calabrese, and S. Attinger, "Data driven high resolution modeling and spatial analyses of the COVID-19 pandemic in Germany," *Plos one*, vol. 16, no. 8, p. e0254660, 2021.
- [38] D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, and A. Vespignani, "Modeling the spatial spread of infectious diseases: The GLocal Epidemic and Mobility computational model," *Journal of computational science*, vol. 1, no. 3, pp. 132-145, 2010.
- [39] M. Chinazzi *et al.*, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395-400, 2020.

- [40] Y. Mammeri, "A reaction-diffusion system to better comprehend the unlockdown: Application of SEIR-type model with diffusion to the spatial spread of COVID-19 in France," *Computational and Mathematical Biophysics*, vol. 8, no. 1, pp. 102-113, 2020.
- [41] P. G. Kevrekidis, J. Cuevas-Maraver, Y. Drossinos, Z. Rapti, and G. A. Kevrekidis, "Reaction-diffusion spatial modeling of COVID-19: Greece and Andalusia as case examples," *Physical Review E*, vol. 104, no. 2, p. 024412, 2021.
- [42] A. Viguerie *et al.*, "Simulating the spread of COVID-19 via a spatially-resolved susceptible–exposed–infected–recovered–deceased (SEIRD) model with heterogeneous diffusion," *Applied Mathematics Letters*, vol. 111, p. 106617, 2021.
- [43] M. Grave, A. Viguerie, G. F. Barros, A. Reali, and A. L. Coutinho, "Assessing the spatio-temporal spread of COVID-19 via compartmental models with diffusion in Italy, USA, and Brazil," *Archives of Computational Methods in Engineering*, vol. 28, no. 6, pp. 4205-4223, 2021.
- [44] N. P. Jewell, J. A. Lewnard, and B. L. Jewell, "Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections," *Jama*, vol. 323, no. 19, pp. 1893-1894, 2020.
- [45] B. Cooper, "Poxy models and rash decisions," *Proceedings of the National Academy of Sciences*, vol. 103, no. 33, pp. 12221-12222, 2006.
- [46] S. Moein *et al.*, "Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan," *Scientific Reports*, vol. 11, no. 1, p. 4725, 2021/02/25 2021, doi: 10.1038/s41598-021-84055-6.

- [47] I. Korolev, "Identification and estimation of the SEIRD epidemic model for COVID-19," *Journal of econometrics*, vol. 220, no. 1, pp. 63-85, 2021.
- [48] H. W. Hethcote, "Three basic epidemiological models," in *Applied mathematical ecology*: Springer, 1989, pp. 119-144.
- [49] L. B. Rodda *et al.*, "Functional SARS-CoV-2-specific immune memory persists after mild COVID-19," *Cell*, vol. 184, no. 1, pp. 169-183. e17, 2021.
- [50] A. Wajnberg *et al.*, "Robust neutralizing antibodies to SARS-CoV-2 infection persist for months," *Science*, vol. 370, no. 6521, pp. 1227-1230, 2020.
- [51] S. F. Lumley *et al.*, "Antibody status and incidence of SARS-CoV-2 infection in health care workers," *New England Journal of Medicine*, vol. 384, no. 6, pp. 533-540, 2021.
- [52] R. F. Bruinsma, G. J. Wuite, and W. H. Roos, "Physics of viral dynamics," *Nature Reviews Physics*, vol. 3, no. 2, pp. 76-91, 2021.
- [53] A. Sette and S. Crotty, "Adaptive immunity to SARS-CoV-2 and COVID-19," *Cell*, 2021.
- [54] K. W. Schechtman, "How Lagging Death Counts Muddled Our View of the COVID-19 Pandemic," ed: The COVID Tracking Project, 2021.
- [55] D. Lippold *et al.*, "Spatiotemporal modeling of first and second wave outbreak dynamics of COVID-19 in Germany," *Biomechanics and modeling in mechanobiology*, pp. 1-15, 2021.
- [56] C. Testa, N. Krieger, J. Chen, and W. Hanage, "Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An

- animation using per capita state-level data (January 22, 2020–July 8, 2020)," 2020.
- [57] M. Khalili, M. Karamouzian, N. Nasiri, S. Javadi, A. Mirzazadeh, and H. Sharifi, "Epidemiological characteristics of COVID-19: a systematic review and meta-analysis," *Epidemiology & Infection*, vol. 148, 2020.
- [58] N. M. Linton *et al.*, "Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data," *Journal of clinical medicine*, vol. 9, no. 2, p. 538, 2020.
- [59] N. Wilson, A. Kvalsvig, L. T. Barnard, and M. G. Baker, "Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality," *Emerging infectious diseases*, vol. 26, no. 6, p. 1339, 2020.
- [60] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *The Lancet infectious diseases*, vol. 20, no. 5, pp. 533-534, 2020.
- [61] M. Wojdyr, "Fityk: a general-purpose peak fitting program," *Journal of Applied Crystallography*, vol. 43, no. 5-1, pp. 1126-1128, 2010.
- [62] H. Lau, T. Khosrawipour, P. Kocbach, H. Ichii, J. Bania, and V. Khosrawipour, "Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters," *Pulmonology*, vol. 27, no. 2, pp. 110-115, 2021.
- [63] C. Staerk, T. Wistuba, and A. Mayr, "Estimating effective infection fatality rates during the course of the COVID-19 pandemic in Germany," *BMC Public Health*, vol. 21, no. 1, pp. 1-9, 2021.

- [64] K. Dodds *et al.*, "The COVID-19 pandemic: territorial, political and governance dimensions of the crisis," ed: Taylor & Francis, 2020.
- [65] P. Cohen, "Comparing the Responses to COVID-19 in Israel, Canada and Japan," in *Managed Healthcare Executive*, ed, 2021.
- [66] R. Waitzberg *et al.*, "COVID-19 pandemic health system responses in the Mediterranean countries: a tale of successes and challenges," *Health Policy*, 2021.
- [67] A. Benz, "Federalism and varieties of parliamentary democracy – Canada and Germany compared," presented at the European Consortium for Political Research General Conference, Prague, 2016.
- [68] A. S. Detsky and I. I. Bogoch, "COVID-19 in Canada: Experience and Response to Waves 2 and 3," *JAMA*, vol. 326, no. 12, pp. 1145-1146, 2021.
- [69] A. S. Detsky and I. I. Bogoch, "COVID-19 in Canada: experience and response," *Jama*, 2020.
- [70] A. Heitmueller and L. Roemheld, "Covid-19 and the false dichotomy between centralised and decentralised healthcare systems," in *BMJ Opinion*, ed, 2020.
- [71] A. Schuppert, K. Polotzek, J. Schmitt, R. Busse, J. Karschau, and C. Karagiannidis, "Different spreading dynamics throughout Germany during the second wave of the COVID-19 pandemic: A time series study based on national surveillance data," *The Lancet Regional Health-Europe*, vol. 6, p. 100151, 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8454815/pdf/main.pdf>.

- [72] R. L. Haffajee and M. M. Mello, "Thinking globally, acting locally—The US response to COVID-19," *New England Journal of Medicine*, vol. 382, no. 22, p. e75, 2020. [Online]. Available: <https://www.nejm.org/doi/pdf/10.1056/NEJMp2006740?articleTools=true>.
- [73] D. Birenbaum-Carmeli and J. Chassida, "Health and socio-demographic implications of the Covid-19 second pandemic wave in Israel, compared with the first wave," *International Journal for Equity in Health*, vol. 20, no. 1, pp. 1-12, 2021.
- [74] R. Challen, E. Brooks-Pollock, J. M. Read, L. Dyson, K. Tsaneva-Atanasova, and L. Danon, "Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study," *bmj*, vol. 372, 2021.
- [75] M. O'Driscoll *et al.*, "Age-specific mortality and immunity patterns of SARS-CoV-2," *Nature*, vol. 590, no. 7844, pp. 140-145, 2021.
- [76] I. C. Marschner, "Estimating age-specific COVID-19 fatality risk and time to death by comparing population diagnosis and death patterns: Australian data," *BMC medical research methodology*, vol. 21, no. 1, pp. 1-10, 2021.
- [77] G. Sorci, B. Faivre, and S. Morand, "Explaining among-country variation in COVID-19 case fatality rate," *Scientific reports*, vol. 10, no. 1, pp. 1-11, 2020.