

COVID-LIBERTY, A Machine Learning Computational Framework for the Study of the Covid-19 Pandemic in Europe. Part 1: Building of an Artificial Neural Network and Analysis and Parametrization of Key Factors which Influence the Spread of the Virus.

Nicholas Christakis^{a,b,*}, Michael Politis^a, Panagiotis Tirchas^a, Minas Achladianakis^a, Eleftherios Avgenikou^a, Christina Kalafati Matthaïou^a, Maria Kalykaki^a, Argyri Kyriakaki^a, Panagiotis Paraschis^a, Evangelos Piliou^a, George Kossioris^{a,b}

^aDepartment of Mathematics and Applied Mathematics, University of Crete, Greece 700 13

^bLaboratory of Applied Mathematics, University of Crete, Greece 700 13

*: Corresponding Author's email: nchristakis@tem.uoc.gr

Received: July 10, 2021. Revised: August 27, 2021. Accepted: August 30, 2021. Published: September 1, 2021.

Abstract—Covid-19 is the most recent strain from the corona virus family that its rapid spread across the globe has caused a pandemic, resulting in over 200,000,000 infections and over 4,000,000 deaths so far. Many countries had to impose full lockdowns, with serious effects in all aspects of everyday life (economic, social etc.). In this paper, a computational framework is introduced, aptly named COVID-LIBERTY, in order to assist the study of the pandemic in Europe. Initially, the mathematics and details of the computational engine of the framework, a feed-forward, back-propagation Artificial Neural Network are presented. 5 European countries with similar population numbers were chosen and we examined the main factors that influence the spread of the virus, in order to be taken into consideration in the simulations. In this way lockdown, seasonal variability and virus effective reproduction were considered. The effectiveness of lockdown in the spread of the virus was examined and the Lockdown Index was introduced. Moreover, the relation of Covid-19 to seasonal variability was demonstrated and the parametrization of seasonality presented.

Keywords—COVID-LIBERTY, machine learning, Covid-19, Artificial Neural Network, Lockdown Index, Seasonal Variability, Effective Reproduction Number

I. INTRODUCTION

The Covid-19 respiratory disease, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was first discovered in Wuhan, China, near the end of December 2019. Following its global spread, it was declared as a public health emergency by the World Health Organization (WHO) on January 20th 2020 [1]. In the course of one and a half year, it has affected 219 countries and there have been over 200,000,000 confirmed cases and over 4,000,000 deaths globally [2]. Since its discovery, scientists and medical experts around the world have been working towards treatment methods, control of the virus spread, and development of a vaccine. Most countries have experienced so far three distinct periods, where there was a

rise in the number of cases and deaths (pandemic waves): the first in spring 2020, the second in autumn 2020 and the third in winter-spring 2021 [3].

In recent years, Machine Learning (ML) methods are being utilized in various fields of medicine (for a review see [4]). Because of their nature, they are considered most suitable for application in Covid-19 research. A brief overview of ML applications related to Covid-19 follows.

ML methods, such as Convolutional Neural Networks, Support Vector Machines and Random Forests, have been utilized to propose improvements in fast diagnosis of cases, in order to augment the screening process of a patient with Computed Tomography images, X-Ray scans and blood samples [5].

In an attempt to improve Covid-19 diagnosis based on clinical data, the use of statistical analysis and ML was proposed to uncover relevant features in common clinical test results [6]. The SOM (self-organizing map) method was employed, along with Random Forest and LASSO regression models in order to classify whether a case is Covid-19 or common influenza and thus, find distinguishing features in the datasets.

SIR-based (Susceptible, Infectious, Recovered) models may help with predicting the spread of a disease, assessing the threat posed to a social group or predicting the mortality rate. Due to their high degree of complexity and the possibility of uncertainties in regards to the pandemic, SIR-based models may prove difficult to apply. To address these issues, ML methods have been employed in order improve the usability of these models [7].

A study on mortality rate in Korea [8] investigated possible factors associated with mortality and high risk through multivariable analysis of demographic and clinical data (age, income, underlying diseases, cardiac symptoms and others), identifying associations between older age, chronic lung disease, asthma with high risk of mortality and between young age, infection from personal contact with low mortality risk, among others. Five different ML classifiers were employed using the above data as predictors in order to assess a case's mortality risk, each predictor showing different levels of importance between the classifiers. Despite the satisfying results, certain predictors may prove to be more important for other datasets (i.e., different countries).

In an article recently published in the Journal of Public Affairs [9], different ML algorithms were built with the aim to forecast cases in India. In particular, Nonlinear

Regression, Decision Trees and Random Forests models were trained. Due to the insufficient data from India, data from China were used, because of similarities in both countries' features (GDP, population, education level, etc.). Lockdown, severity of lockdown, number of deaths, recovered patients, confirmed cases and performed tests were the features which were utilized to train the model. The results analysis outlined the importance of a lockdown feature on training ML models and may be considered as evidence to the positive effects of lockdown.

Artificial Neural Network (ANN) models have also been utilized in the scientific community's attempt to analyze existing data and predict future trends of the disease. By utilizing epidemic data from the outbreak of the disease, researchers in [10] tested 14 ANN-based models to estimate the confirmed cases in China, Iran, South Africa, Italy, United States of America, Singapore and Japan. Each Network was trained solely on number of confirmed COVID-19 cases from previous days with the current day being the reference data, the first using only the previous day, and each subsequent Network going one day back, with the 14th being trained on 2 weeks of daily cases. The results showed that the Network trained on 2 weeks ranked first in terms of overall accuracy among all countries. The results themselves, however, were not satisfactory, possibly because using only reported cases as data is not enough to identify the difference in viral spread between different countries.

In this work, a modeling framework was designed around a feed-forward back-propagation ANN in order to study the evolution of Covid-19 in 5 European countries, namely Austria, Belgium, Greece, Portugal and Sweden. The framework was named **COVID-LIBERTY: Cases OVER tests Identifier-Lockdown Index Based Ensemble modeling with effective Reproduction and Temperature variability**. All countries have a population of the order of 10,000,000 with varying population densities (Sweden being the less and Belgium being the more densely populated of the 5, with Austria, Greece and Portugal being in between with similar orders of magnitude of population density) and they are located in different geographical parts of Europe, i.e. Greece and Portugal in the south, Austria in the centre and Belgium and Sweden in the north. All countries except for Sweden enforced periods of lockdown. These 5 countries were chosen in order to demonstrate the importance of lockdown in keeping down the numbers of infections and to put forward the proposition that seasonal variability is related to the ability of the virus to transmit,

as is the case with respiratory diseases (see e.g. [11]). In Part 1 of this work, it is established that lockdown and weather conditions are important factors that should be taken into consideration, along with the effective reproduction number. In the following sections, the architecture of the feed forward, back-propagation ANN, which constitutes the main computational engine of COVID-LIBERTY, is detailed. Then, the data acquisition procedure is given, followed by discussion on the virus effective reproduction. Then, separate analyses show the importance of lockdown and seasonal variability in the spread of the virus and the Lockdown Index and a parametrized effective country temperature are introduced. This paper acts as a precursor to Part 2, where the COVID-LIBERTY framework and the idea of ensemble modeling are introduced and tested.

It has to be noted that it is not the aim of the authors to compare their results with results from other similar models but rather to demonstrate the ability of the current framework to operate as a useful tool for predicting future trends of the pandemic. Its assessment and comparison with other similar tools will be performed in a future work. Moreover, different parameters to the ones used in other models have been utilized and new parametrizations have been derived exclusively for the scope of this work, thus making comparisons with other models difficult. It has to be noted that to the best of our knowledge, the idea of ensemble modeling within the same framework for the study of Covid-19, parametrizing a number of important factors and conditions, has not been utilized in any other current study.

II. THE COMPUTATIONAL ENGINE OF COVID-LIBERTY: BUILDING THE ANN MODEL

ANNs have been utilized for a number of years in a number of diverse applications, ranging from image processing [12] to tuberculosis studies [13]. For a recent comprehensive review, see [14]. In recent years, Convolutional Neural Networks (CNNs) have been used in areas where ANNs were dominant, especially when image classification and large datasets are involved [15]. However, ANNs are still widely utilized and perform better in cases where image recognition is not necessary or when datasets are limited [16].

In COVID-LIBERTY, a feed-forward, back-propagation ANN was utilized. The feed-forward, back-propagation algorithm is still the most popular algorithm employed in ANNs, due to its performance and simplicity

to implement (see e.g., [17]). In a generic ANN (with input-hidden-output layers of neurons), each inter-neural connection is defined by a particular real number, a synaptic weight w_{ij} . The total number of these weights is the degrees of freedom of the ANN and in these connections resides the ‘knowledge’ of a trained ANN. Each neuron internally performs two basic operations: (a) a combination which computes a linear summation over its input signals and (b) a scaling through an activation function which bounds the result of the summation to a required interval. The most commonly used activation function (employed in the present work, as well) is the non-linear sigmoid:

$$y(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

whose range of values is the interval $[0, 1]$.

To illustrate how for a pattern the values propagate through successive layers (input to hidden to output), imagine that I is the number of nodes of layer- i and J the number of nodes of layer- j (we are moving from layer- i to layer- j), where x_i are the values of layer- i that are fed in each node of layer- j to yield a value y_j . First, each of the i -neurons will pass its value through the sigmoid function of Eq.(1) to yield the following output:

$$x_i^{in} = \frac{1}{1 + e^{-x_i}}, i = 1, 2, \dots, I \quad (2)$$

Therefore, each of the j -neurons receives a signal equal to:

$$y_j = \sum_{i=1}^I w_{ij} x_i^{in} + w_{0j}, j = 0, 2, \dots, J \quad (3)$$

In the previous equations, the subscript 0 corresponds to a bias, where x_0^{in} is taken to be equal to 1. The process described above is repeated until the values reach the output layer. The training process starts with a random distribution of the synaptic weights w_{ij} between layers. Via the process described above, a unique and wrong solution vector is produced at the output of the ANN, for a particular pattern. However, since the correct vector corresponding to the input vector for this pattern is known, a measure of the error can be calculated:

$$E = \frac{1}{2} \sum_{k=1}^K (y_k - t_k)^2 \quad (4)$$

where K is the number of output neurons and y_k and t_k are the components of their corresponding output and target values, respectively. The back-propagation algorithm with Gradient Descent is utilized in order to find the particular distribution of weights that minimize the value of E . The

objective is to update w_{ij} after the n^{th} iteration by $\Delta^n w_{ij}$, such that the total error becomes smaller. This is achieved by determining $\Delta^n w_{ij}$ through the change in the signals y_j corresponding to the neurons of layer- j ($j=1,2,\dots,J$) (i.e., Δy_j) and Δw_{ij} of the $(n-1)^{\text{th}}$ iteration (i.e., $\Delta^{n-1} w_{ij}$):

$$\Delta^n w_{ij} = \eta \cdot \Delta y_j + a \cdot \Delta^{n-1} w_{ij} \quad (5)$$

where η and a are the ‘learning rate’ and ‘learning momentum’, respectively. The values of these parameters, following the discussions in [18] and [19], are set to 0.2.

Next, the procedure is repeated for all patterns and thus an ‘epoch’ is completed. The procedure is then iterated over the necessary number of epochs to satisfy a convergence criterion. It is critical that during every epoch, the patterns are introduced at a random order. This ensures faster learning rates, avoids ‘memorizing’ and increases the capability of the ANN to handle situations it has not been trained for. After training has been completed, the obtained set of synaptic weights may be utilized in the ANN, in order to obtain predictions for patterns that the network has not been trained for.

For the current study, the ANN was also utilized during the initial stages of data acquisition, to fill in missing values for Sweden. Various configurations were tested, based on heuristics [20] and previous experience [19], in order to find the optimum combination of hidden layers (HL) and number of nodes in each HL, which gave the most accurate results in the shortest time. The configurations chosen for the ANN simulations were always 1 HL and either 20 nodes in the HL for 8 inputs and 1 output or 30 nodes in the HL for 4 inputs and 3 outputs. The convergence criterion at training was either for the cumulative error from all patterns to become 0.10% of its initial value or for 5,000,000 epochs in total to be completed. For all trainings performed, the error fell sharply to below 1% of its initial value within less than 1,000,000 epochs and then decreased slowly but it never reached the required 0.10% and all 5,000,000 epochs were completed. However, the maximum final error for all trainings was of the order of 0.69% of its initial value. Hence, the process of training was considered as successfully completed and we proceeded to the validation and prediction phase of the ANN.

III. DATA ACQUISITION AND THE IMPORTANCE OF EFFECTIVE REPRODUCTION

All data in this study were collected for the period between March 1, 2020 and April 30, 2021. For four

countries, namely Austria, Belgium, Sweden and Portugal data went up to January 10, 2021, whereas for Greece data were collected up until April 30, 2021, in order to study the third wave of the pandemic. Data up until December 20, 2021, were utilized both in Parts 1 and 2 of this work. In Part1 these data were used to establish the important factors which should be considered in the computational framework, i.e., effective reproduction, lockdown and seasonal variability. Then, in Part 2 the COVID-LIBERTY framework was trained on data up until December 20, 2020, to obtain predictions for the Christmas period (December 21, 2020 until January 10, 2021) and introduce the algorithm of ensemble modeling. Moreover, a further study was conducted for Greece (utilizing a training dataset with data until March 1, 2021), in order to demonstrate the ability of the framework to accurately predict the disease evolution during the prolonged third pandemic wave (March-April 2021). Data in all graphs are presented from March 15, 2020 onwards, when consecutive non-zero values of all variables were present in the dataset.

Covid-19 data concerning the 5 countries of this study were retrieved from [21] for daily numbers of performed Covid-19 tests, daily numbers of reported positive cases and daily numbers of reported deaths. In general, full datasets were available for all countries, with occasional, not consecutive, values missing. For such missing values, the average of their border values (i.e., of the previous and the following day) was chosen in order to fill the gaps. For Austria, the numbers of daily performed Covid-19 tests were missing before April 2, 2020 hence, it was decided for Austria to use data after that date. For Sweden, data on daily numbers of performed tests were absent (only weekly numbers were given) hence, in order not to introduce any further error, we did not perform any analysis for Sweden concerning the number of daily positive cases with respect to the number of daily performed tests. There were also missing data for the daily cases in Sweden for 73 days between August 28 and December 20, 2020. These days were not all consecutive but were split into one period of 3 days, sixteen periods of 4 days and one period of 6 days, scattered between periods where data were available. The final day for all periods contained the sum of positive cases of all consecutive days. For the case of the single periods (i.e. those of 3 and 6 days), the cumulative values were equally split between the 3 and 6 days. For the sixteen periods, where only the cumulative number of positive cases was available on the 4th day, it was decided to train the COVID-

LIBERTY ANN with data from March 16, 2020 (when Sweden started reporting consecutive non-zero positive cases) to August 28, 2020 in order to be able to predict the missing values. For this reason, that interval was split into periods of 7 days and patterns for the training of the network were formulated, where as inputs the values of the first three days were taken, plus the sum of the values of the following 4 days. The outputs requested from the ANN were the values of the 4th, 5th and 6th days. Obviously, the value of the 7th day may be calculated through the subtraction of the 4th, 5th and 6th predicted values from the cumulative value given for the final day. Hence, 173 such patterns were formulated and the ANN was trained with 4 inputs, 3 requested outputs, 1 HL and 30 nodes in the HL with final training error of 0.65% after 5,000,000 epochs. Then, with the obtained weights, the 73 missing values were predicted and the dataset of daily reported positive cases for Sweden was completed.

In Figures 1, 2 and 3 the data until December 20, 2020 for all countries are presented (except for daily tests for Sweden), along with their 7-days moving averages, which follow closely the trend of the data curves. It was decided, based on the discussion in [18], to utilize moving averages of the data for future predictions, rather than their daily values. The 7-days moving averages were preferred to the 5- and 10-days moving averages, because it was a good compromise between the two, following closely the data curve, without smoothing it out in a way that the trend of the data was lost.

It was deemed necessary to include in our study a measure of the spread rate of the disease, which would give an indication of its future evolution. One such number, widely used, is the effective reproduction number R_t . R_t is defined as the average number of secondary infectious cases that may potentially result from a primary infectious case. For instance, a value of 2 for R_t means that an infected person may infect 2 more persons, who in turn may infect 2 persons more each and so on. For the method followed in order to derive the values of R_t see [22]. It should be noted that for the derivation of R_t , the population density of a country (i.e., the number of people per unit area) is not taken into consideration. There is also a 7-day lag in the R_t values, because there is an incubation period of up to 14 days [23]. This means that every case is considered to have contracted the virus on average 7 days prior to the date that the case is reported. Values of R_t were retrieved from [24] from March 19, 2020 onwards.

Figure 4 presents the data set of R_t for all countries and its 7-days moving averages, until December 20, 2021. As observed, data were missing for Greece for the period between May 22 and June 06, 2020. In order to fill the missing data, the two border values were considered (i.e., the values of May 21 and June 07), their average was calculated and all missing values were considered to take that value. This approach was preferred to regression, because the behaviour of R_t is observed to be quite erratic hence, regression will not offer better error minimization and better approximation of the missing values. The importance of the variations of R_t may be understood, when they are compared to the variations of the number of reported cases, where it can be seen that for all countries there was a several days' lag between the peak of R_t and the peak of the number of cases. The quantification of this variation will be made more explicit in Part 2 of this work, where R_t is explicitly compared to the number of reported cases.

Population statistics data (i.e. population of each country and population density) were retrieved from [2]. Population density represents the number of people per unit area.

Meteorological data were retrieved from [25]. Only temperature data were retrieved, since as a first approximation, only temperature was considered in this study as a measure of seasonality. All countries chosen for this study fall well within the synoptic meteorological scale (i.e., similar weather conditions to all parts of the country for time scales of a few days at a time, see e.g. [26]). Hence, a daily effective temperature was calculated for each country by averaging the mean daily temperatures of 4 different geographical locations. These locations were chosen to be either in the North-East-South-West or only in the North-South directions, depending on the morphology of each country. In this way, the locations considered were: for Austria Graz, Innsburg, Salzburg and Vienna; for Belgium Anwerp, Kotrijk, Malmedy and Vieux Virton; for Greece Athens, Heraklion, Patras and Salonica; for Portugal Elvas, Faro, Lisbon and Porto; for Sweden Gothenburg, Malmö, Stockholm and Umeå. Taking into consideration the Covid-19 incubation period as for R_t , it was decided to attribute to each moving average of daily reported positive cases a 3-day moving average country temperature, 7 days prior to the date the cases were reported. The 3-day moving average for temperature was chosen to represent the average duration of synoptic scale meteorological phenomena. Through this parametrization,

we believe that the meteorological/seasonal conditions at the time of infection were correctly represented.

IV. THE IMPORTANCE OF LOCKDOWN

As already mentioned, of the countries considered in this study, only Sweden did not officially enforce any lockdown. By lockdown, a restrictive policy is meant that forces people to stay where they are and various aspects of everyday life are restricted (people movement restrictions in place, commercial shops / restaurants closed, schools/universities operating tele-education, etc.). Lockdown was proposed early on, as a preventive measure for the spread of the disease [27]. Lockdown may be classified as soft (some commercial shops, apart from supermarkets, are in operation, limited restrictions in people movement, etc.) or hard (no shops apart from supermarkets are in operation, strict restrictions in people movement and travel, restrictions to some industrial operations). The periods of lockdown for each country were obtained through [28] and [3]. All lockdown information was valid until January 10, 2021, for Austria, Belgium, Sweden and Portugal and until April 30, 2021 for Greece. Austria imposed a hard lockdown from March 16 until April 20, 2020. Then, a second hard lockdown was introduced from November 17, 2020, which turned into soft between December 6 and December 26, 2020 and since that date a hard lockdown was again in place. Belgium introduced a hard lockdown on March 18 until April 17, 2020, which then turned into soft and eventually ended on June 07, 2020. Then a soft lockdown was reintroduced on October 22, 2020, it turned into hard on November 1, 2020 and continued as such. Greece introduced a hard lockdown on March 19 until May 4, 2020. Then, it reintroduced a hard lockdown on November 6, 2020, this turned into soft from December 14, 2020 until January 1, 2021 and then it turned into hard and continued as such. In Portugal, only one period of hard lockdown was considered between April 2 and May 18, 2020. Then, in October, some minor restrictions were introduced (earlier closing times for shops and restaurants) that could not be classified as lockdown. The periods of lockdown (from March 15 to December 20, 2020) for these countries may be graphically seen in Figure 5, where the 7-day moving average of reported cases is presented, without differentiating on each panel between hard and soft lockdowns. This however, was taken into consideration in the COVID-LIBERTY framework with the formulation of a Lockdown Index. This will be discussed later in this section.

Figure 6 presents the 7-day moving averages of reported cases and deaths for all 5 countries from March 07 to December 20, 2020 and the 7-days moving averages of R_t from March 23 to December 20, 2020. As observed, Belgium exceeded significantly in absolute numbers the number of reported cases and deaths of all other countries, whereas R_t displayed an erratic behavior, similar for all countries, after the initial stages of the pandemic, oscillating between 0.35 and 1.75. However, in these results, the population density of each country was not taken into consideration. Since all countries have similar populations (ranging from just over nine million for Austria to about eleven and a half million for Belgium), it was decided to convert these results with respect to population density. If all numbers (daily cases, deaths and R_t rates) are divided by the population density of the respective countries, we believe that these relative numbers give a good indication of the effect of Covid-19, depending on how sparsely or densely populated a country is. The population densities of all 5 countries are given in Table 1.

Table 1. Population densities of countries considered in this study (retrieved from [2]).

Country	Population density (people/km ²)
Austria	109
Belgium	383
Greece	81
Portugal	111
Sweden	25

Figure 7 shows the converted with respect to population density 7-days moving averages of daily cases, deaths and R_t rates. It is very clear from the results, that Sweden was well above all other countries in relative numbers. The relative number of cases for Sweden always remained above that of the other countries, had a small peak towards the end of June and from September onwards until December 20, 2020 (when the last data for Sweden were collected) followed a sharply increasing trend, in direct contrast with the behaviour of the other countries. In the relative number of deaths, two large peaks were observed, towards the beginning and the end of the examined period (first and second pandemic waves), which were 2 to 4 times higher than the highest peak of the rest of the countries. The relative R_t rate of Sweden was always much higher than those of the other countries, which exhibited a similar behaviour with most of their relative R_t rates intersecting

throughout the examined period. It should be observed that in relative numbers Belgium exhibited a remarkably low R_t rate (lowest of all). This may be attributed to the fact that it has the second highest population density in Europe (second only to the Netherlands) and that it enforced the longest lockdown of all countries in this study at the beginning of the pandemic (March 18-June 07, 2020) and reintroduced its second lockdown earlier than all the rest (October 22, 2020).

All presented data pointed to the fact that lockdowns did indeed have a significant effect in keeping down the number of infections and deaths related to Covid-19 in European countries. It also becomes clear from the presented results that the numbers of cases and deaths kept increasing after October (and up at least December 20, 2020) in Sweden, where no lockdown or any other restrictions were enforced (in contrast to the rest of the countries). The relation between the length of lockdowns and the reduction in infection numbers and deaths, in conjunction with other factors, such as how fast lockdowns should be imposed, vaccination rates, etc., will be the subject of a following study. The authors also believe that population density serves as a crucial factor to demonstrate the importance of lockdowns.

As mentioned earlier, in order to parametrize the effect of lockdown in COVID-LIBERTY, a Lockdown Index (LI) was formulated, which took values between 0.0 (no lockdown) and 1.0 (hard lockdown) or 0.5 (soft lockdown). LI has the advantage of not needing further normalization for its inclusion in the COVID-LIBERTY framework, since its values are already between 0 and 1 (see [18]). Based on data from the countries involved in this study (i.e. variation of infections, deaths and R_t rates), for hard lockdowns it was decided for LI to increase gradually from 0.0 to 1.0 within four weeks (i.e., 0.033 increase per day). From then on, it remained fixed to 1.0 until the date when lockdown ended. Then, it gradually decreased to 0.0, again within a period of four weeks (i.e., 0.033 increase per day). For soft lockdowns, the exact same procedure was followed, but with LI reaching a maximum value of 0.5 (i.e., 0.0165 increase/decrease per day). During transitions from soft to hard lockdowns, an increase of 0.033 per day was assumed, until the value of 1.0 was reached. Similarly, for transitions from hard to soft lockdowns a decrease of 0.0165 was assumed until the value of 0.5 was reached. These assumptions were supported by the framework performance, as will be demonstrated in Part 2 of this work.

V. THE IMPORTANCE OF SEASONALITY

By observing the evolution of the Covid-19 pandemic in Europe from its start until recently, it is believed that weather conditions/seasonal variations play a crucial role in the ability of the virus to infect more people. The idea that climate and weather conditions are important in the evolution of diseases is not new (see e.g., [29]). It has been recently argued that seasonality might induce two outbreaks of respiratory-linked diseases per year [30]. Of the atmospheric variables that characterize weather (i.e., surface temperature, atmospheric pressure, precipitation, humidity, wind speed and direction, see [31]), temperature is the most readily available and is the one exhibiting the required variability in order to describe seasonal changes (see e.g., [26]). Hence, in this study seasonal changes were correlated to the increase/decrease in the number of reported positive cases and deaths through temperature. An effective daily temperature for each country was calculated and attributed to reported cases and deaths, as described in Section 3.

Figure 8 presents the 7-day moving average of daily cases and the corresponding temperatures during periods when no lockdown was in place for all 5 countries, during summer-autumn 2020. The results are given outside lockdown periods, in order to isolate the effect of seasonal variability. It is clear from the results that there existed a trend of inverse proportionality between number of cases and temperature (i.e., when temperatures decreased, the number of cases seemed to be increasing). This trend became clearer when cases were presented with respect to temperatures in Figure 9. Both cases and temperatures were normalized between 0 and 1, utilizing their respective maxima and minima for each country. The inverse proportionality between cases and temperatures was evident, with the lower temperatures corresponding to higher number of cases, whereas for higher temperatures, the number of cases was significantly reduced. Regression analysis with least squares was applied and a $y=a/x+b$ curve was fitted to the data. The coefficients a and b for all five countries are given in Table 2.

Table 2. Coefficients of equation $y = a/x+b$, where x is normalized temperature and y is the normalized number of reported cases for all countries in this study.

Countries	a	b
Austria	0.086	-0.089

Belgium	0.074	-0.086
Greece	0.049	0.0042
Portugal	0.213	-0.231
Sweden	0.065	0.0075

We cannot perform the exact same analysis for the summer of 2021, since we are halfway through it and there exists the new Delta variant, which, with its very high degree of transmissivity, has made the number of infections rise significantly (see e.g. [32], [33], [34]). However, the number of deaths is particularly low, compared to the rise in infections, and this is believed to be due to weather conditions, since the herd immunity threshold of 80% due to vaccination has yet to be reached in Europe (see [3] and [35]). A rigorous analysis will be performed at a later stage (beginning of autumn), when all data from summer 2021 will be available.

The performed analyses indicate that weather conditions do play a significant role and in the COVID-LIBERTY framework seasonality will be considered through the parametrization discussed in Section 3.

VI. CONCLUSIONS

In Part 1 of this work, full analyses were performed in order to establish the link between lockdown / seasonal variability and the spread of Covid-19. Their effects were appropriately parametrized through LI (Section 4) and an effective daily temperature (Section 3), in order to be utilized in the COVID-LIBERTY framework. The importance of the effective reproduction number was also discussed. Moreover, the foundations for COVID-LIBERTY were laid by fully describing the feed-forward, back propagation ANN, which constitutes the engine of the framework. To the best of our knowledge, the proposed approach is the only one found in literature thus far, which considers a number of (readily available from public datasets) important parameters for the spread of the disease and incorporates them as parametrizations in a computational framework to predict the evolution of the pandemic. In Part 2, the setup of the COVID-LIBERTY framework will be presented, along with the algorithm of ensemble modeling. Two different periods of the pandemic will be studied, in order to assess the ability of the framework to predict the evolution of the disease.

REFERENCES

- [1] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, R. Agha, World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19), *International Journal of Surgery*, 76 (2020) 71-76
- [2] Worldometers, Information on Coronavirus and European countries statistics <https://www.worldometers.info> (accessed 19 August 2021)
- [3] Reuters Covid-19 Tracker, general Covid-19 information for countries worldwide <https://graphics.reuters.com/world-coronavirus-tracker-and-maps> (accessed 19 August 2021)
- [4] J.A.M. Sidey-Gibbons, C.J. Sidey-Gibbons, Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*, 19, 64, 18 pages (2019). <https://doi.org/10.1186/s12874-019-0681-4>
- [5] S. Lalmuanwma, J. Hussain, L. Chhakchhuak, Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, 139, 110059 (2020). <https://doi.org/10.1016/j.chaos.2020.110059>
- [6] W.T. Li, J. Ma, N. Shende, G. Castaneda, J. Chakladar, J.C. Tsai, L. Apostol, C.O. Honda, J. Xu, L.M. Wong, T. Zhang, A. Lee, A. Gnanasekar, T.K. Honda, S. Kuo, M.A. Yu, E.Y. Chang, M.R. Rajasekaran, W.M. Ongeko, Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. *BMC Medical Informatics and Decision Making*, 20, 13 pages (2020). <https://doi.org/10.1186/s12911-020-01266-z>
- [7] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Lio, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, J. He, Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 12, 3 (2020) 165-174
- [8] C. An, H. Lim, D.-W. Kim, J.H. Chang, Y.J. Choi, S.W. Kim, Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Scientific Reports*, 10, article no.18716, 11 pages (2020)
- [9] R. Mahji, R. Thangeda, R.P. Sugasi, N. Kumar, Analysis and prediction of COVID-19 trajectory: A machine learning approach. *Journal of Public Affairs*, e2537, (2020). <https://doi.org/10.1002/pa.2537>

- [10] H. R. Niazkar, M. Niazkar, Application of artificial neural networks to predict the COVID-19 outbreak. *Global Health Research and Policy*, 5, 50 (2020).
<https://doi.org/10.1186/s41256-020-00175-y>
- [11] E.B. Postnikov, D.V. Tatarenkov, Prediction of flu epidemic activity with dynamical model based on weather forecast. *Ecological Complexity*, 15 (2013) 109-113
- [12] I. Ebert-Uphoff, K. Hilburn, Evaluation, Tuning and Interpretation of Neural Networks for Working with Images in Meteorological Applications. *Bulletin of the American Meteor. Soc.*, 101, 12 (2020) 2149-2165
- [13] P. Dande, P. Samant, Acquaintance to Artificial Neural Networks and use of artificial intelligence as a diagnostic tool for tuberculosis: A review. *Tuberculosis*, 108 (2018) 1-9
- [14] A.I. Olundare, J. Aman, E.O. Abiodun, V.D. Kemi, A.M. Nachaat, A. Humaira, State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4, e00938, 41 pages, (2018)
- [15] J. Teuwen, N. Moriakov, Chapter 20 - Convolutional Neural Networks. In: *Handbook of Medical Image Computing and Computer Assisted Intervention*. The Elsevier and MICCAI Society Book Series, Academic Press (2020) 481-501
- [16] V. Meel, ANN and CNN: Analyzing differences and similarities, February 1, 2021, <https://viso.ai/deep-learning/ann-and-cnn-analyzing-differences-and-similarities/#:~:text=Similarities%20and%20Differences%20of%20ANN%20vs%20CNN,-ANN%20uses%20weights&text=The%20%E2%80%9Cclayes%20in%20ANN%20are,ANN%20uses%20weights%20to%20learn.&text=Comparatively%2C%20other%20is%20no%20neuron,filtration%20to%20analyze%20image%20inputs.>
- [17] N.B. Shaik, S.R. Pedapati, S.A.A. Taqvi, A.R. Othman, F.A.A. Dzibir, A Feed-Forward Back Propagation Neural Network Approach to Predict the Life Condition of Crude Oil Pipeline. *Processes*, 8, 661, 13 pages (2020)
- [18] N. Christakis, V. Barbaris, A. Spentzos, A New Approach in Financial Modelling with the Aid of Artificial Neural Networks. *Journal of Algorithms and Computational Technology*, 5, 3 (2011) 513-529
- [19] N. Antoniou, A Big Data Analytics Application in Transportation and Astrophysics, MSc Dissertation, Department of Physics, University of Crete, Greece, 124 pages (2017)
- [20] T. Kavzoglu, P.M. Mather, The use of backpropagating artificial neural networks in land cover classification. *Int. J. Remote Sensing*, 24, 23 (2003) 4907-4938
- [21] Our World in Data, Statistics and Data concerning Covid-19
<https://ourworldindata.org/coronavirus>
(accessed 27 July 2021)
- [22] T.V. Inglesby, Public Health Measures and the Reproduction Number of SARS-CoV-2. *JAMA*, 323, 21 (2020) 2186-2187
- [23] C. McAloon, A. Collins, K. Hunt, A. Barber, A.W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O'Grady, S.J. More, Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*, 10, e0396532, 9 pages (2020)
- [24] Covid-19 Spread Mapper, *R*, data retrieval
<http://metrics.covid19-analysis.org>
(accessed 27 July 2021)
- [25] Time and Date, Temperature data retrieval
<https://www.timeanddate.com>
(accessed 27 July 2021)
- [26] L.M. Grenzi, J.M. Nese, A World of Weather: Fundamentals of Meteorology. Kendall Hunt Publishing, 548 pages (2018)
- [27] K.H.R. Dineros, J.P.P. Dipasupil, COVID-19 crisis and prevention plan, March 15, 2020,
https://www.slideshare.net/KevinHilRDineros/covid19-crisis-management-and-prevention-plan?fbclid=IwAR2a74JhmEJwZCHLqA11RMtm_MX_qM1LZvAL7CJB_tj4pITiMozguGjgIJE
- [28] Wikipedia, lockdown information retrieval
<https://www.wikipedia.org>
(accessed 15 May 2021)
- [29] M.J. Bouma, H.J. van der Kaay, El Niño Southern Oscillation and the historic malaria epidemics on the Indian subcontinent and Sri Lanka: an early warning system for future epidemics? *Tropical Medicine and International Health*, 1 (1996) 86-96
- [30] T. Dbouk, D. Drikakis, Fluid dynamics and epidemiology: Seasonality and transmission dynamics. *Physics of Fluids*, 33, 021901, 9 pages (2021)
- [31] K.L. Ebi, L.O. Mearns, B. Nyenzi, Weather and climate: changing human exposures. Chapter 2 in: *Climate change and human health*, World Health Organization, 18-42 (2003)

- [32] L. Di Domenico, V. Colizza, Epidemic scenarios of Delta variant in France in the summer 2021. Report #31, www.epicx-lab.com (2021).
- [33] J. Wise, Covid-19: France and Greece make vaccination mandatory for healthcare workers. *BMJ* 2021;374:n1797 (2021).
- [34] E. Mahase, Delta variant: What is happening with transmission, hospital admissions and restrictions? *BMJ* 2021;373:n1513 (2021).
- [35] S. Bolotin, S. Wilson, M. Murti, Achieving and sustaining herd immunity to SARS-Cov-2. *CMAJ*, 193, 28, July 19 2021, (2021)

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Nicholas Christakis conceived the idea, wrote the code in Fortran77 for the ANN, organized all simulations and wrote the paper.

Michael Politis, Evangelos Pilios and Panagiotis Paraschis retrieved all weather and Covid-19 data for Sweden, performed the ANN simulations for Sweden and produced all graphs.

Panagiotis Tirchas, Minas Achladianakis and Eleftherios Avgenikou retrieved and analysed all Covid-19 data for Belgium and Portugal.

Argyri Kyriakaki, Christina Kalafati Matthaiou and Maria Kalykaki retrieved and analysed all Covid-19 data for Greece and Austria.

George Kossioris helped in organizing the layout of the paper and simulations and wrote the conclusions.

Sources of funding for research presented in a scientific article or scientific article itself

This work was not performed under any specific project and did not receive any grant.

Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US

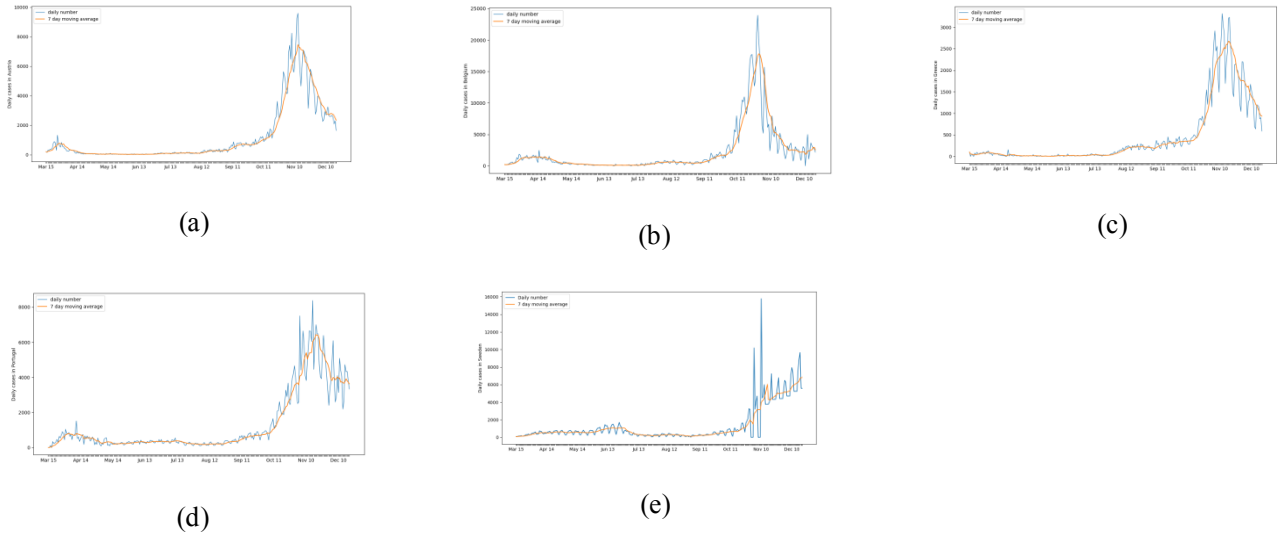


Figure 1. Daily reported Covid-19 positive cases and 7-days moving averages for (a) Austria, (b) Belgium, (c) Greece, (d) Portugal and (e) Sweden.

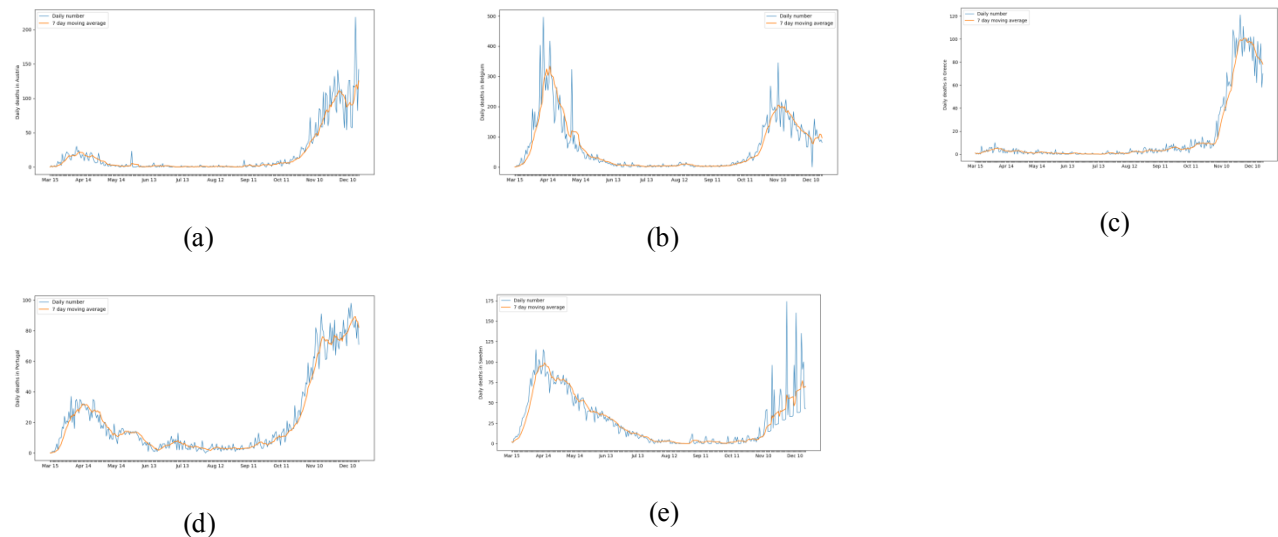


Figure 2. Daily reported deaths due to Covid-19 and 7-days moving averages for (a) Austria, (b) Belgium, (c) Greece, (d) Portugal and (e) Sweden.

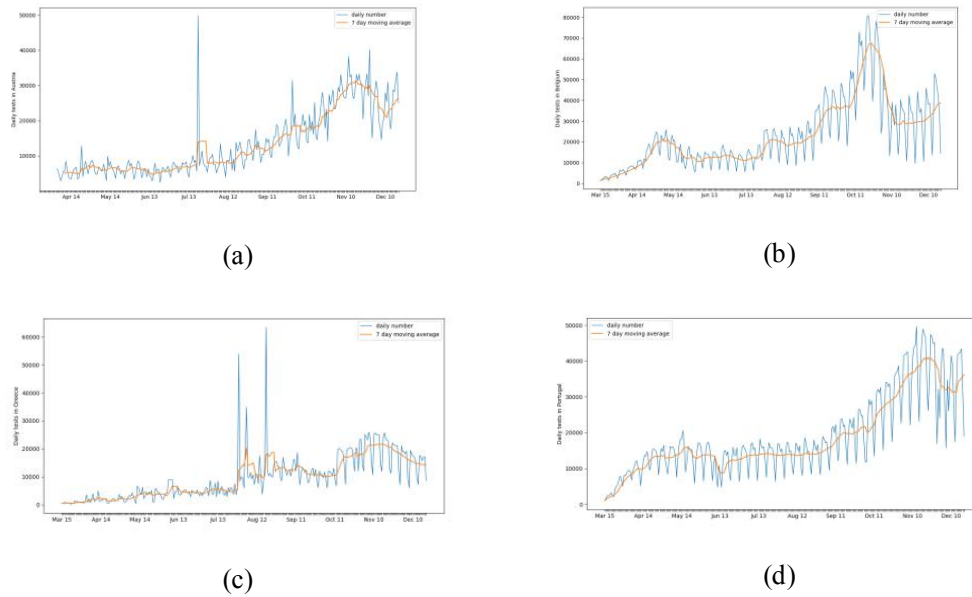


Figure 3. Daily performed Covid-19 tests and 7-days moving averages for (a) Austria, (b) Belgium, (c) Greece and (d) Portugal.

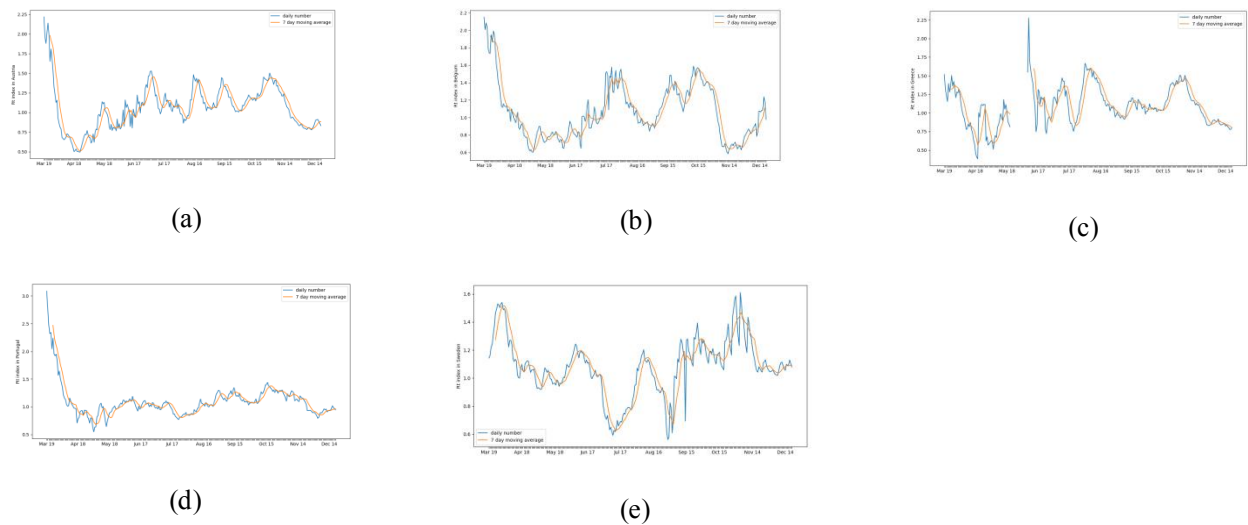


Figure 4. Daily R_t and 7-days moving averages for (a) Austria, (b) Belgium, (c) Greece, (d) Portugal and (e) Sweden.

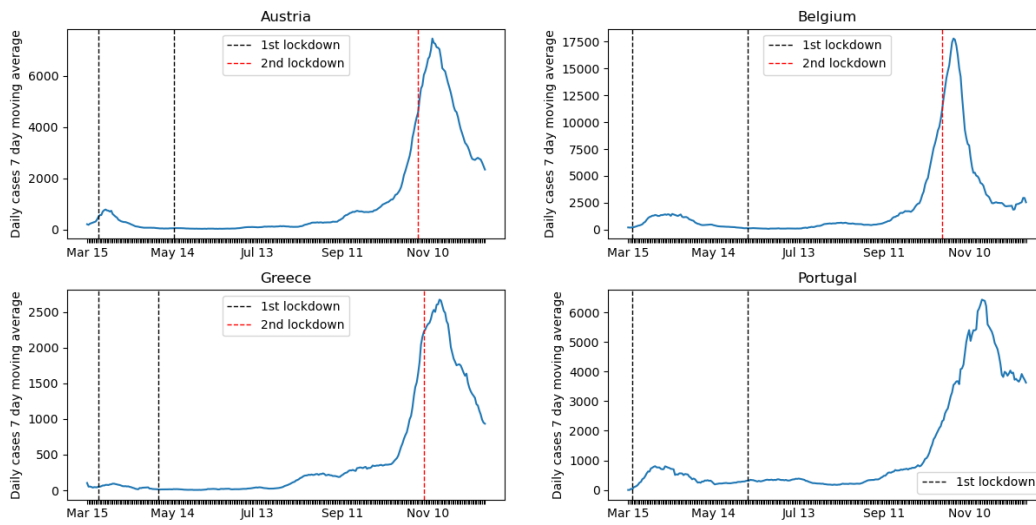


Figure 5. 7-days moving averages of reported positive cases between March 7 and December 20, 2020 with lockdown periods being marked.

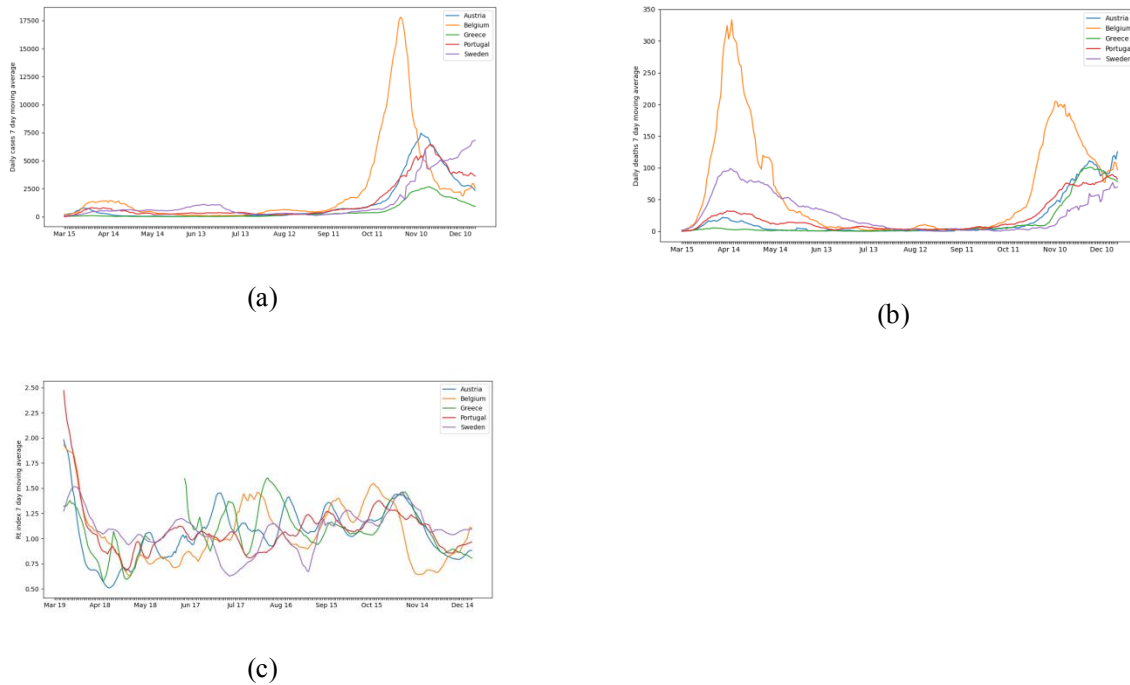


Figure 6. Comparisons between countries of 7-days moving averages of (a) reported cases, (b) reported deaths and (c) R_t

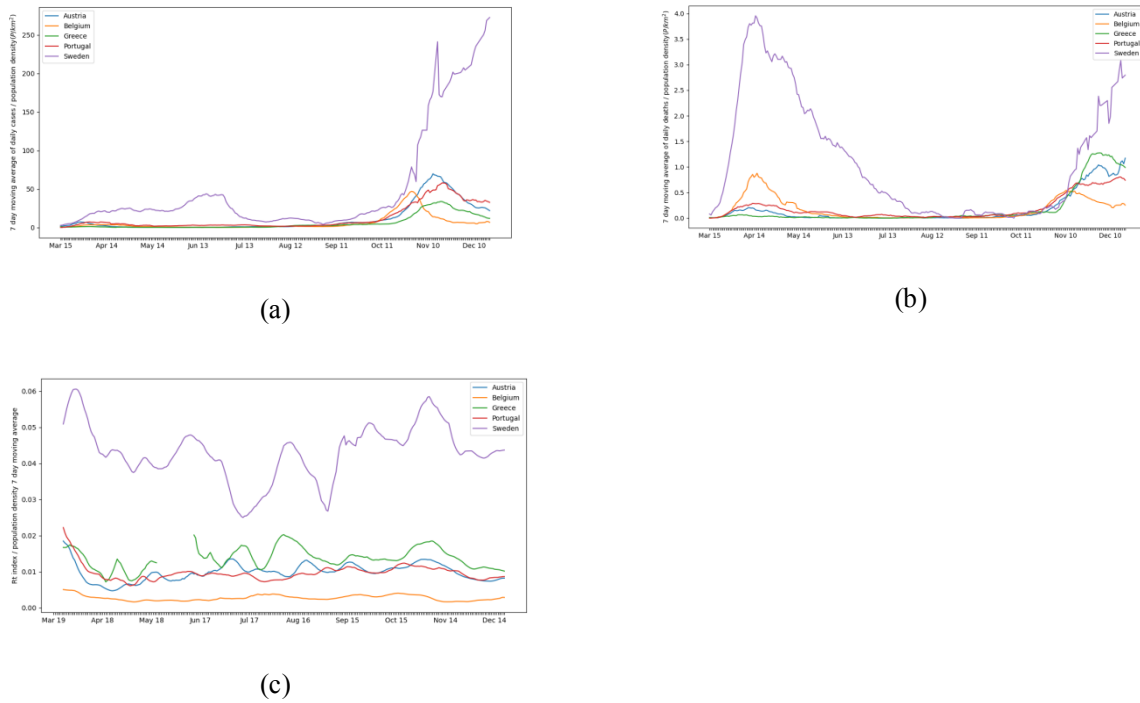


Figure 7. Comparisons between countries of 7-days moving averages of converted with respect to population density: (a) reported cases, (b) reported deaths and (c) R_t

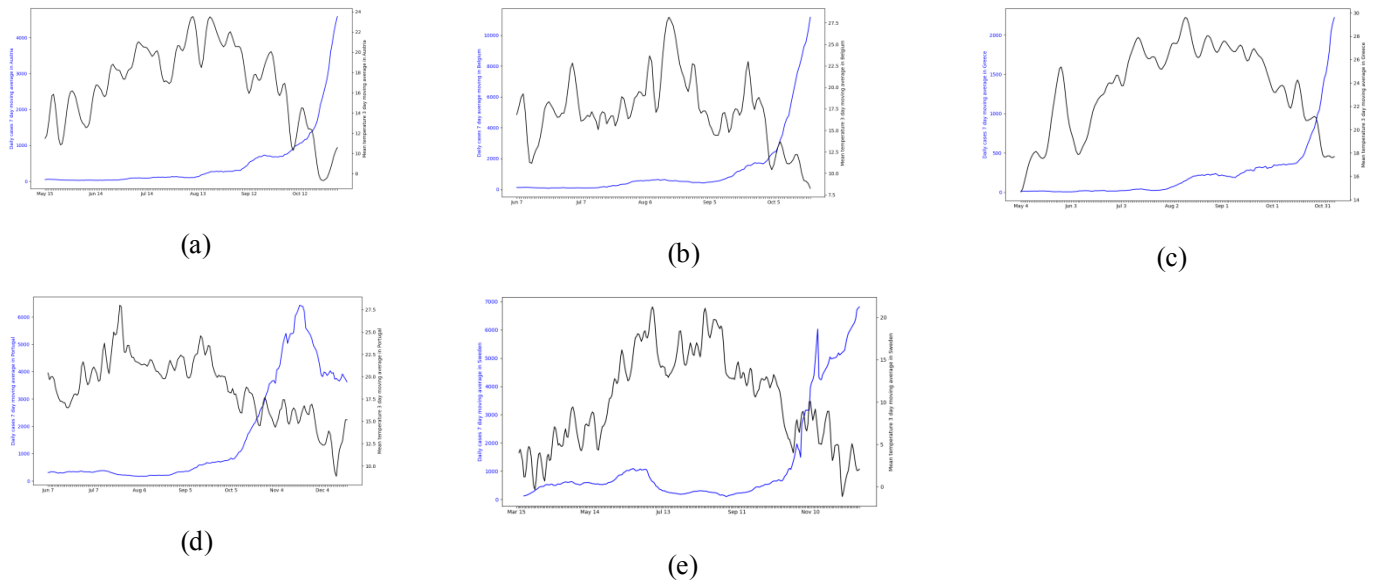


Figure 8. 7-days moving averages of reported cases and corresponding temperatures for periods when no lockdown was in place for (a) Austria, (b) Belgium, (c) Greece, (d) Portugal and (e) Sweden.

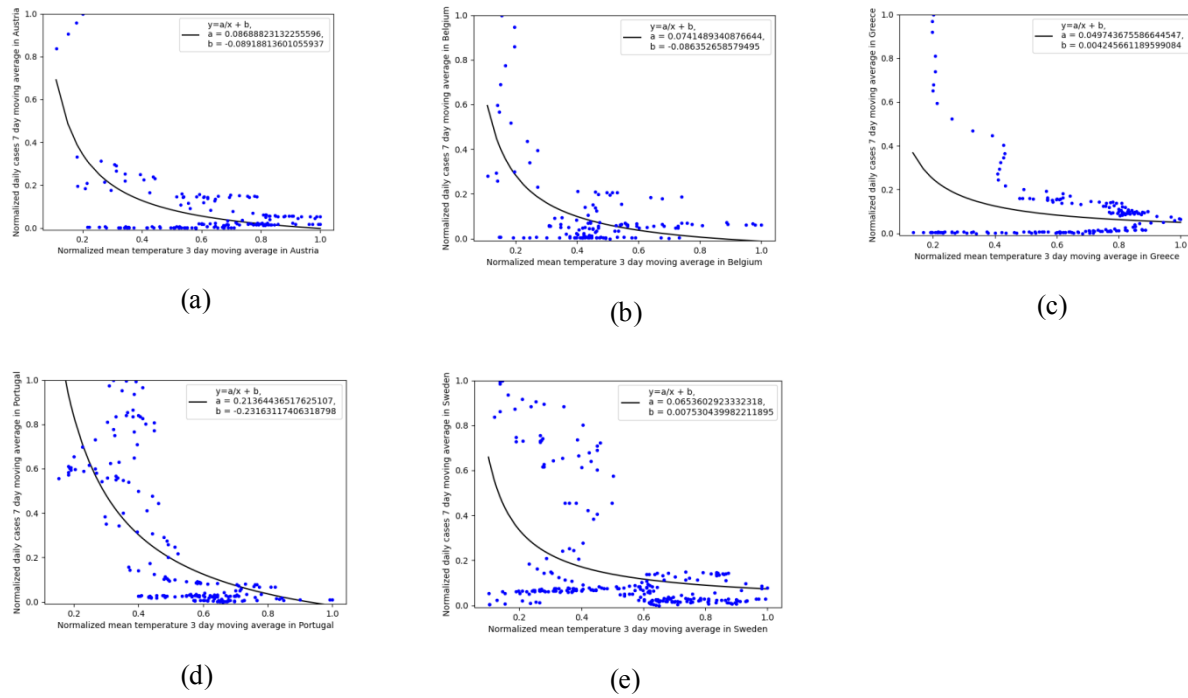


Figure 9. Normalized number of reported cases vs. normalized temperatures of Figure 8 for (a) Austria, (b) Belgium, (c) Greece, (d) Portugal and (e) Sweden.