The Role of Temperature and Humidity in the Early Evolution of SARS CoV-2

Andrea Bravo, Lucas Fernández & Sebastià Nicolau
Supervised by:
Dr. Sergio Alonso, Dr. Enric Álvarez & Dr. David Pino.
Universitat Politècnica de Catalunya.
Projects of Engineering Physics II.

The new coronavirus SARS CoV-2 which causes the COVID-19 was declared as a global pandemic at 11 of March 2020. Since then, many studies have been conducted in order to determine the importance of environmental factors,mainly the temperature and humidity, in the dynamics of the pandemic. Although there is some evidence that the environment can modulate the transmission of the virus, it is yet to be determined if seasonal and meteorological variations could play an important role in the spread of the virus. In this study, we analyse observational data from four different European countries in order to explore whether temperature and humidity influence the reported incidence. It is observed that in Spain and Italy there is a negative correlation between temperature and humidity with the incidence of the disease. However, this is not the case for the four analyzed countries. Consequently, from the data analyzed here we cannot conclude that temperature or humidity are main drivers in the transmission dynamics of SARS CoV-2.

I. INTRODUCTION

The pandemic of the COVID-19, acronym where CO' stands for corona, 'VI' for virus, and 'D' for disease, caused by the SARS CoV-2 can have devastating consequences for the world's health and economic status, that is why it is presented as an unprecedented challenge. Almost all countries in the world currently present a continuous local transmission, with numbers that do not stop growing, and the implications in the statistics of mortality are expected to be severe. This is why there is a growing interest in knowing what factors can influence the dynamics of the pandemic to better understand and control the transmission of the virus.

There has been a growing interest in the role played by the environment in the modulation of the transmission dynamics, both to predict the future of the pandemic and to know which decisions should be made with the arrival of summer. It could be considered that the influence of environment on the expansion of pathogens is reflected in different aspects. In addition to the environment, human behavior should be taken into account as it changes significantly with seasons: in winter more time is spent indoors with less ventilation and personal space, while in summer outdoor activities are more frequent and population mobility increases.

In a recent article, *Gutiérrez et al., 2020* review different contributions that relate environmental conditions to the spread of the virus. They found that there may be a consensus on the influence of meteorologic conditions in the distribution of SARS CoV-2, although a good part of the investigations based on observational data have not been able to reliably isolate their effect. The authors highlight the unreliability of some research studies because of their claim of clearly existing correlations. The majority of the articles revised concur that the ideal conditions for the spread of the virus is a cool and dry environment, although they emphasize that it is necessary to insist that SARS CoV-2 is in an active propagation phase and it is to early to conclude that its distribution is in equilibrium with the weather, although patterns can be observed.

There are two types of studies that analyze the correlation of the environment conditions with the incidence of the virus. One type consists of running simulations with different hypotheses of correlations, while in the second type correlations based on observational data are searched. *R. E. Baker et al., 2020* performed different simulations, using a Susceptible-Infected-Recovered-Susceptible (SIRS) model to case data, and they found that while variations in weather may be important for endemic infections, during the pandemic stage of an emerging pathogen, the environmental conditions drive only modest changes to the pandemic size. Simulations showed that tropical scenarios experience lower intensity outbreaks that the higher latitudes which experience higher transmission rates, although this fact has been demonstrated not to be correct.

Alvarez et al., 2020 analyzed observational data of accumulated cases and temperature from each province of Spain to find possible correlations and make a predictive model. In this study, no consistent trend or evidence of a relationship between the accumulated cases and the temperature values at the provincial level are found.

Here we perform, a spatio-temporal analysis of the early evolution of SARS CoV-2 across the regions of Spain, Italy, Germany and Sweden. The main goal is to explore the possible existence of a relationship between temperature and humidity and the evolution of the daily number of COVID-19 cases for approximately the month of the epidemic before the confinement was ordered in each country.

II. METHODOLOGY

A. Data

COVID-19 data corresponding to the provinces of Spain was downloaded from a repository publicly available and accessible through this link: `https://code.montera34.com:4443numeroteca/covid19`. This repository is being maintained by multiple volunteers that are extracting and homogenising COVID-19 data from multiple web sources belonging to both official and private media.

COVID-19 data corresponding to the Länder of Germany was downloaded from a repository publicly available and accessible through this link: `https://corona-sh.de/covid-19-statistiken`. This repository belongs and is being maintained by Lübeck University.

COVID-19 data corresponding to the regions of Italy and Sweden are obtained from World Health Organization (WHO) surveillance reports and from European Centre for Disease Prevention and Control (ECDC).

Before performing the analysis of the effect of temperature and humidity on the incidence of the pandemic, we wanted to study how the incidence evolves over time in order to conduct a proper exploration. We observed that the incidence follows the usual exponential behaviour with an underlying oscillatory character of an approximate period of 7 days. This lead us to do a mean of the daily cases over one week to analyse the incidence evolution and performing the same mean for the atmospheric variables over the same time span.

Temperature and humidity data has been obtained from around 1000 automatic weather stations installed across the four countries of study. The data was downloaded from the OpenData platform of the National Oceanic and Atmospheric Administration (NOAA) at National Centers for Environmental Information. We obtained the daily mean temperature and dew point temperature data (in $^{\circ}C$) measured by each of these weather stations from the beginning of February to the beginning of April. The dew point is the temperature to which air must be cooled to become saturated with water vapor. When cooled further, the airborne water vapor will condense to form liquid water, since this variable does not depend on temperature, it is away to compute humidity independently from temperature.

SARS CoV-2 shows a mean incubation period of approximately 5 days (ranging from 2 to 14 days) (*Nishiura et al., 2020*).Therefore, the effect of temperature and humidity variables has been considered with a time delay in respect to the incidence.

B. Software

The MatLab programming language has been used to perform the statistical analysis involved in the present investigation. MatLab offers enough tools to process a considerable amount of data, adequately plot the obtained results and adjust specific trends.

Regarding the atmospheric variables, the mean variables for each region per day are obtained calculating the mean of all the stations present in that region, excluding those located at a height superior to a determined value (different for each country), to keep out non-populated mountain sectors from the analysis. The final data used are the mean values between 14 February and 14 March. It has been used a month-period to minimize the effect of climatic disturbances, so obtaining the most averaged representation of each region's climate possible.

The data of the accumulated cases was processed to obtain two different variables. The first one is the incidence of the epidemic, that is, the number of cases per 100.000 inhabitants. The second one is the incidence multiplied by the surface area of each region, so it depends on the average population density of the region.

To represent these data, it is first designated a starting day of the epidemic. It is selected as the day when the incidence is equal or higher than 5 in a specific region. Starting from this day, the daily values for Incidence and Incidence·Surface after a specific number of days are plotted as a function of the mean temperature and dew point temperature. The study considers lapses of 7, 10, 14 and 20 days after the beginning of the epidemic. This particular method is the one that has been perceived to show the highest correlation between the studied variables from all the different methods that have been tested. As mobility policies adopted in each country might diminish or neutralize the effect of climate on the evolution of the epidemic, there is a limit in the day which data can be represented. This limit is set to 1 April in Spain, Sweden and Germany, and 24 March in Italy.

Three different types of adjustments are used in the analysis: lineal, quadratic and exponential. The adjustment plotted in the graphics is the one showing a lower adjustment residual of these three. There have been set qualitative boundaries by adding and subtracting a finite value to the tendency curve, which are plotted in dashed lines. The names of the regions that are located outside these boundaries are shown in order to properly identify the regions that "are further" from the trend.

The reason why more complex functions or higher order polynomials have not been used is that the main intention of the project is to identify general trends of the data. Higher order polynomials might have shown a lower residual in the adjustment, but would have failed to depict the global tendency the data might show.

III. RESULTS

In order to perform a global analysis, Fig.1 shows the incidence and incidence.surface as a function of the mean T and Td 10 days after the epidemic starts for all the countries considered in the study.

As it can be seen, the results change drastically when the incidence is multiplied by the surface because, by doing this, the population density is taken into consideration (at some scale) and the fitting curve bends. However, it is important to emphasise that the population distribution within a region isn't usually uniform. The majority of the population is concentrated around the most important cities of the region, while the rest of the territory is usually less populated.

Moreover, although the the polynomial adjust is quadratic for all the countries, the shape of the fitting curve varies greatly for each country. In Fig.1, the plots involving the variable incidence.surface show that the adjust functions for Sweden and Spain are very similar. For Germany, Spain and Sweden the fitted parabolas are concave, while for Italy it is convex. These differences could be caused by other effects (apart from the temperature and humidity) that are not being taken into account and that are more important in some countries than others. Fig. 2a,2b,2c and 2d include respectively the expressions of the fitted polynomials for Germany, Spain, Sweden and Italy, together with their corresponding residues.

By looking at the range of values for T and Td, it is observed that Spain and Italy have a wider temperature range than Sweden and Germany. Consequently, if a correlation between the incidence and T or Td exists, for the case of Germany and Sweden there should not be significant changes between the incidence in different regions of the same country. This fact can be observed in Fig.1, as all the points corresponding to the regions of Sweden and Germany are located in a small areas of the 2D plane.

In Italy, a 1°C increase in T (in a range of T between 8-10°C) reduces the incidence value by about 7.748509 (± 0.003422), and a 1°C increase in Td (in a range of T between 4-6°C) reduces the incidence value by about 7.90841 (± 0.14462).

In Spain, a 1°C increase in T (in a range of T between 8-10°C) reduces the incidence value by about 14.6324 (± 1.5208), and a 1°C increase in Td (in a range of T between 4-6°C) reduces the incidence value by about 10.4223 (± 1.1214).

In Germany, a 1°C increase in T (in a range of T between 5-7°C) increases the incidence value by about 5.215 (± 14.03), and a 1°C increase in Td (in a range of T between 1-3°C) reduces the incidence value by about 1.14845 (± 1.5498).

In Sweden, a 1°C increase in T (in a range of T between 1-3°C) reduces the incidence value by about 1.6117 (± 0.2736), and a 1°C increase in Td (in a range of T between 1-3°C) reduces the incidence value by about 2.0899 (± 0.3278).

The results obtained for each country can be compared

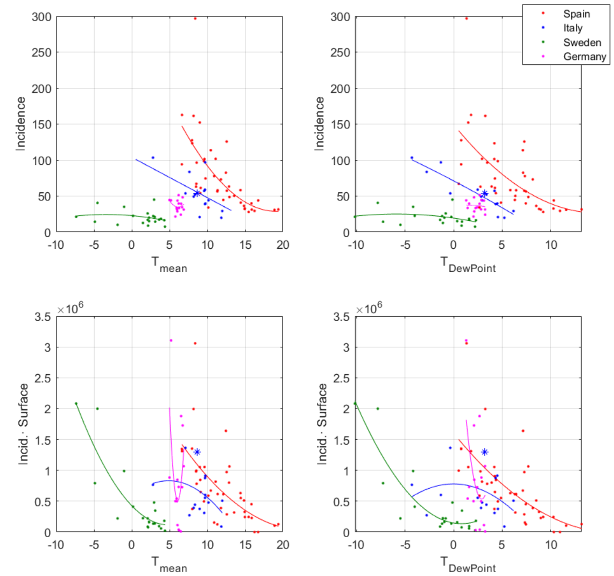


FIG. 1: Incidence and Incidence·Surface vs. Mean Temperature and Dew Point Temperature in Spain, Italy, Sweden and Germany regions 10 days after epidemic starts

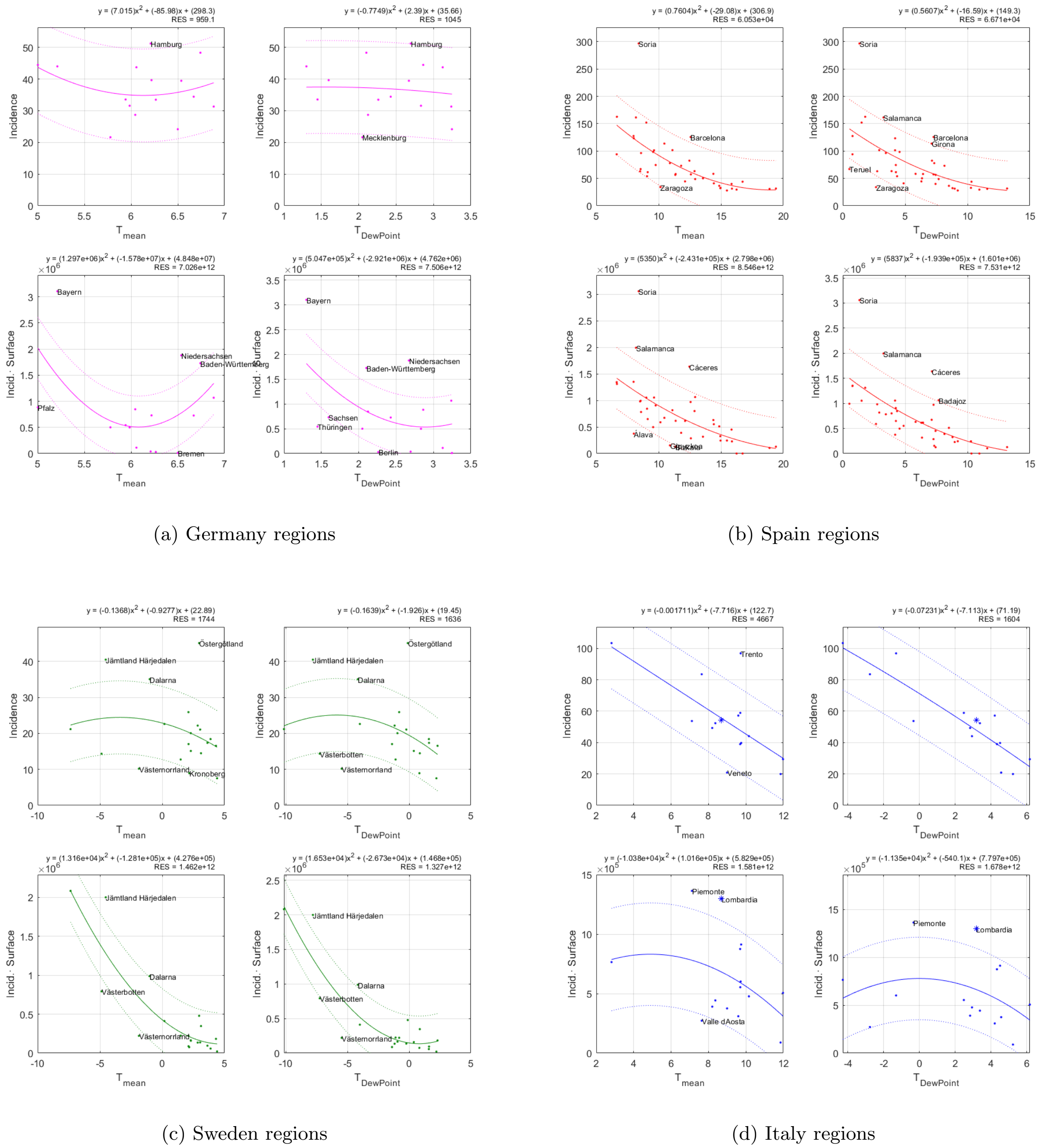
with the results presented by the article *Jingyuan Wang et al., 2020*. In this study, for China and the U.S, the effect of T and the relative humidity (instead of Td) on the variation of the effective reproductive number R (instead of the incidence) is analysed.

In order to perform a local analysis, the data (and its fitting) 10 days after the epidemic starts was plotted for all the regions of each country separately. Fig. 2a,2b,2c and 2d show respectively the maps of clusters corresponding to Germany, Spain, Sweden and Italy.

In the cases of Germany (Fig.2a) and Sweden (Fig.2c) the comments stated before can be clearly observed. Therefore, the incidence and the incidence.surface change much less between the regions of these two countries than for Spain or Italy, and so do the T and Td. As a result, a correlation between these variables is not discarded (although it is not proved). The more dispersed regions in the clustering map could be associated to fluctuations of the variables.

In the cases of Spain (Fig.2b) and Italy (Fig.2d), T and Td present a wider range of values than for Germany and Sweden. As can be observed in this figures, incidence and incidence·surface present large variations between the regions of these two countries. From these data, a relation between these variables can be inferred.

It should be noted that for Spain, the analysis was performed by provinces rather than by autonomic communities. This way, the unit of territory represented in the figure had a less irregular population distribution and when multiplying the incidence by the surface, the effect of the population density is corrected more effectively.



(a) Germany regions

(b) Spain regions

(c) Sweden regions

(d) Italy regions

FIG. 2: Incidence and Incidence·Surface vs. Mean Temperature and Dew Point Temperature 10 days after epidemic starts.

In the cases of Italy and Spain, a quadratic negative tendency of the incidence and the incidence.surface with the T and Td can be observed. We will focus on the clustering maps involving the variable incidence.surface, as for this plots the effect of population density is taken into account.

As can be observed, there are some regions and provinces where the value of incidence·surface is located far away from the adjust function. In the case of Italy, the regions that do not follow the pattern are: Piemonte, Lombardia and Valle d'Aosta. All these regions are located at the North. Lombardia, and Piamonte present an incidence·surface higher than the adjust function. In these regions (specially in Lombardia) the epidemic started to wreak havoc earlier than in any other European region. On the other hand, the Valle d'Aosta region shows an incidence.surface lower than the norm. This phenomena could also be expected as, even though the region has a border with the Piamonte, it consists in a small valley poorly communicated with other regions.

In the case of Spain, Soria and Salamanca present an incidence·surface higher than the adjust function. This phenomena could be explained by considering the mobility of population from Madrid to these locations. Cáceres and Badajoz (both belonging to the autonomic community of Extremadura) also have an incidence.surface higher than expected, while Álava and Guipúzcoa show an incidence.surface lower than the norm. An explanation to this phenomena is yet to be found. Also, by looking at Fig. 2b, it can be seen that the provinces of Teruel, Zaragoza, Barcelona and Girona move closer to the fitting curve when multiplying the incidence by the surface.

The results obtained for Spain and Italy are in concordance with the results shown by many other studies, for example the ones collected in the article (*Gutiérrez et al., 2020*). As it can be seen in the plots, higher values of the T and Td correspond to lower values of the incidence and the incidence·surface. Hence, the transmission of the virus is favoured in cold and dry climates.

Apart from the environmental conditions there could be many other determining factors for the evolution of the pandemic. Some of the variables that could be taken into consideration when studying the dynamics of the SARS CoV-2 in a certain region are: the mobility of the population, the population density, the number of doctors per capita, the number of UCI beds per capita, the percentage of population over 65 years old in the region and the GPD per capita, among others. If any of these variables has a greater effect on the transmission of the virus than the variables considered, the possible correlation between the transmission dynamics and the T or Td would be hidden by the effect of the other variables. Therefore, as many variables as possible should be taken into consideration in the analysis.

In order include the effect of all of these factors, one should look for an index that quantifies each one of the factors and use the Fama-Macbeth regression methodology to perform the statistical analysis. Exactly this procedure is carried out in the study *Jingyuan Wang et al., 2020*.

IV. CONCLUSIONS

Studies of the influence of meteorological conditions may have affected the initial evolution of COVID-19 in different countries has received increasing attention in recent months. As many studies focus only in one country or region and some others only include the effect of temperature, in our analysis we wanted to compare different European countries, exploring both temperature and humidity, expressed in the variable of dew point temperature.

In this study, no consistent evidence has been found of a significative correlation of temperature and humidity in the transmission dynamics in the early stages of the SARS CoV-2 pandemic. In this regard, this agrees with the studies mentioned in the introduction. Nevertheless, patterns of a negative correlation between the meteorological variables and the incidence can be observed, especially for the Mediterranean countries, specially Spain. Furthermore, it is also observed, especially in the case of Sweden thanks to the fact that the temperature range reaches up to -10°C, that cold temperatures represent a lower limit for the transmission of the virus. this trend is observed more significantly if we multiply the incidence by the surface, turning it into a variable that is related to population density. What does stand out is the difference between countries, although they have differences in the temperature range, a great disagreement is observed in the behavior of the correlation and therefore it can be deduced that it is necessary to consider other variables that have a equal or greater influence in the transmission dynamics of SARS CoV-2.

This study only analyzes the evolution of the epidemic during winter. Therefore, it's advisable to repeat the analysis in areas or months having higher temperature of dew point temperature mean values. However, this is difficult at this moment because the epidemic is evolving in a world with different type of non-pharmaceutical interventions everywhere. Furthermore, it is not recommended to extrapolate the results outside the temperature range considered here.

Currently available data on the epidemic is subject to a large degree of uncertainty. The number of confirmed cases is globally underestimated, and comparisons across countries, cities or regions are difficult to be determined due to differences in data collection procedures or health policies, among others. Under these circumstances, every statistical analysis of COVID-19 data needs to be interpreted with caution (*Royal Statistical Society, 2020*). It should also be noted that the covariance between the environmental variables and the transmission of the virus is highly complex, since in addition to influencing the biological conditions for its development, they also influence human social behavior.

[1] Baker, Rachel E. and Yang, Wenchang and Vecchi, Gabriel A. and Metcalf, C. Jessica E. and Grenfell, Bryan T., Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science*, 2020

[2] Briz-Redón, Álvaro, Serrano-Aroca, Ángel. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Elsevier. Science of the Total Environment*, 2020, vol. 728, p. 138811.

[3] Gutiérrez-Hernández, O., & García, L.V. ¿Influyen tiempo y clima en la distribución del nuevo coronavirus (SARS CoV-2)? Una revisión desde una perspectiva bio-geográfica. *Investigaciones Geográficas*, 2020.

[4] Nishiura, H., Linton, N.M., Akhmetzhanov, A.R., 2020. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.*, 2020, vol. 93, p. 284–286.

[5] Oto-Peralías, D., Regional correlations of COVID-19 in Spain. *OSF Preprints (Not revewed).*, 2020.

[6] Jingyuan Wang1, Ke Tang2*, Kai Feng1, Xin Lin1, Weifeng Lv1, Kun Chen3,4 and Fei Wang5, D.,High Temperature and High Humidity Reduce the Transmission of COVID-19. *SSRN*, March 2020.