The Role of Inter-Regional Tourism in the Spread of COVID-19 in Italy during the 2020 Summer: A Confirmatory Study

Marco Roccetti
Dept. of Computer Science and Engineering
University of Bologna, Bologna, Italy
marco.roccetti@unibo.it

Luca Casini
Dept. of Computer Science and Engineering
University of Bologna, Bologna, Italy
luca.casini7@unibo.it

ABSTRACT

Going into 2021 many European countries have faced subsequent strong waves of COVID-19 infections that forced governments to reestablish many of the restrictions imposed in the 2020 spring, after a period of relative calm in the summer when the pandemic seemed to be under control. Here, we examine the period prior to the resurgence of the pandemic in autumn, following the inter-regional mobility due to Italian tourism. Further to a preliminary analysis, we investigate on the spread of the pandemic in Italy during the summer 2020, relating the number of infections with tourism data. In particular, we put the focus on the internal tourism (as international mobility was restricted at the time) between the various Italian regions for the summer holidays. The intuition is that tourist flows, due to the lift of travel restriction in the period when most people went on vacation, has been a determining factor in bringing back the virus, especially in areas where there were only a handful of cases, thus kick starting the so-called second wave that hit Italy in the autumn 2020.

CCS CONCEPTS

• HCI • Life and medical sciences • Probability and Statistics

KEYWORDS

COVID-19 · Tourism · GLM · Negative Binomial

ACM Reference format:

M. Roccetti et al. 2021. The Role of Inter-Regional Tourism in the Spread of COVID-19 in Italy During the 2020 Summer: A Confirmatory Study. In *Proceedings of GOODIT'21. ACM, Rome, Italy, 6 pages,* https://doi.org/10.1145/3462203.3475888

GoodIT '21, September 9–11, 2021, Rome, Italy

ACM ISBN 978-1-4503-8478-0/21/09… $15.00
https://doi.org/ 10.1145/3462203.3475888

1 Introduction

As the death toll from the ongoing COVID-19 pandemic rises worldwide, there are few motivations to be cheerful. However, this coronavirus has also brought with it some positive side effects. Among them one of the good things has been that many scientists of different disciplines have begun to study this complex phenomenon, providing their contribution. This paper should be interpreted as one of these attempts.

Italy indeed was the first European country that had to deal with the insurgence of the COVID-19 pandemic at the end of February 2020. After that the number of infected quickly escalated, the government decided for a total lockdown of the country at the beginning of March 2020 that was gradually lifted by the end of May and then terminated at the beginning of June 2020.

At that point, the situation had become manageable, with certain regions that were virtually free from the contagion. The numbers of daily new cases stayed low for a couple of months, giving the illusion that the worst was behind, but gradually started increasing at the end of July before rising up considerably in August, ending up an order of magnitude higher by the start of September. After that, in October 2020, Italy experienced an exponential rise that quickly reached, and surpassed, the numbers encountered at the beginning of the pandemic.

Our intuition is that the first bump of that second wave may have been the consequence of the summer holydays. Italians, traditionally, tend to go on vacation in that specific period, with a peak in the mid of August, and the subsequent first growth in the number of people positive to coronavirus was registered at the beginning of September, around two weeks after.

This pattern could be observed in almost every Italian region and, even though some of the smallest regions are an exception, they actually prove the rule, having very small numbers of positives to begin with, and not a lot of tourism. Thus, what we believe is that tourism may have played a decisive role in this first bump of contagion, perhaps setting in motion the so-called second wave registered in October-November 2020 [2, 3].

Within this framework, in a previous paper we studied tourist flows (inbound and outbound) for each Italian region and their relationship to the number of daily COVID-19 infection cases [1]. Leveraging on the databases of the Italian Institute of Statistics (ISTAT), for data on tourism and other demographic indicators, and of Civil Protection (Protezione Civile), for data on COVID-19 cases and testing, we conducted a statistical analysis, aiming at finding a proof of the aforementioned intuition.

In that preliminary analysis, we had to work with tourism data from 2019, as the numbers for 2020 were not available for most regions in Italy. While assuming a tourist flow distribution similar to that of 2019 was not outlandish, we are now able to confirm some of those hypotheses with the tourism data of summer 2020 which are currently, and slowly, released by the various Italian regions. More precisely, in this paper we analyze the relation between the tourist flows and the resurgence of the infections using the 2020 tourism data from eight different Italian regions (i.e., Emilia-Romagna, Lazio, Liguria, Lombardia, Piemonte, Umbria, Valle d'Aosta and Veneto). The results we have achieved with this second analysis and the actual data definitely confirm the plausibility of our initial intuition.

The remainder of this paper is structured as follows. In Section 2 we provide a survey of the most recent scientific literature in the field. In Section 3 we illustrate the scientific background behind our study and the statistical methodologies we have employed, while in Section 4 we provide the new evidence we have got by using current data from the 2020 summer tourism for the eight aforementioned Italian regions. Finally, Section 5 concludes the paper with a discussion on the strengths and limitations of our work.

2 Related Work

Indeed, there have been a number of relevant researches that have addressed the topic of interest, yet in the context of international travels or domestic travels, but in different nations than Italy. Among those researches, relevant is the work developed in [4], where a model was proposed that tried to estimate how the number of infections evolved as a function of people interaction and the COVID-19 infection prevalence, registered in different geographical areas. This specific model heavily depends on the geographical distance between the areas taken into consideration and, as such, it draws upon a well-known migration modeling approach, termed gravity modeling, that is, essentially, a type of regression used to model international migration flows. In those models, the probability of bringing the infection to a given area is subjected to an exponential decay, depending on the relative distance, like with the physical gravitational concept of mass.

In another work, a similar kind of gravity model fitted on the number of infections was used that performs a regression analysis on the various means of transport in order to investigate on how much is relevant their relative contribution to the spread of the virus [5].

Even more interestingly, in [6, 7], scientific investigations were conducted to reason about the role played by tourism in the spread of the virus between countries across the world, using a regression analysis that has verified the relevance of tourist flows on the total number of infected people, yet in conjunction with other factors like, for example, population density and aging, healthcare expenditures, and others.

The same approach was followed also in [8], where alpine tourism during the winter holidays was examined as a primary factor in favor of a surge of COVID-19 cases in the Scandinavian countries, as many citizens brought back the virus, at their return from their holidays in Italy and Austria. In the same vein of the already examined studies, also in [9] it has been investigated on the influence of tourists moving from China to other countries, while comparing the resulting COVID-19 spread with previous pandemic.

Obviously, we are very far from expressing comments on all these researches, as the only reasonable comment is that the majority of the mathematical models employed in the studies, conducted in all these months, have revealed various limitations that are due to a variety of external factors: including the stochastic nature of the phenomenon under observation, the scarcity or unavailability of data, their unreliability with some of them changing meaning and/or value over time limitation of drawing on mathematical models which are linear by nature, while the subtle implications of this phenomenon are probably stochastic and depend on non-linear dynamics. This is probably and, in some sense, unfortunately a limitation which is shared by many researches in this field.

On our side, in the preliminary analysis, described in [1], we fitted a generalized linear model, where the number of infections were assumed to be distributed on the basis of a negative binomial distribution and were predicted utilizing a traditional maximum likelihood estimation method [10].

Beside the accuracy of the predictions we achieved, the interesting part of the work we have done with this kind of analysis was that of providing a further evidence in favor of the Italian domestic tourism as a factor that may have determined the resurgence of the virus in many Italian regions during the 2020 summer period.

In this sense, instead, the limitation of our work, rather than an issue of capturing some kind of non-linearity of the phenomenon, was more in the fact that, while the number of new infections per each Italian region was precisely known, the distribution of the 2020 Italian summer tourist flows, unknown during the period of our investigation, was replaced with the distribution of the Italian flow of tourists for the same observed period, but occurring in the year 2019, as recorded by the Italian Institute of Statistics (ISTAT) [11].

While that preliminary work has gained an international reputation, being considered representative of the Italian situation during the 2020 summer, as witnessed by a specific citation in the British Medical Journal [12], we feel that our duty is that of confirming/denying the findings of that preliminary analysis with the real data of the 2020 Italian summer tourist flows.

This is exactly what we have done, in the present paper, with the 2020 summer inbound domestic tourist flows data, released by the eighth Italian regions: Emilia-Romagna, Lazio, Liguria, Lombardia, Piemonte, Umbria, Valle d'Aosta and Veneto.

In Section 4, in fact, we present the results we achieved with this new 2020 summer data, thus providing new insights into this phenomenon, as well as a confirmation of the plausibility of our initial research hypothesis.

3 Background and Methodologies

Two are, essentially, the statistical methods we have used in our analysis about the possible correlation between the 2020 Italian summer tourism and the resurgence of the virus: a change point detection method and a generalized linear model.

From a technical viewpoint, a change point detection method, is a statistical analysis of a series of temporal data (the daily infection cases, in our analysis) which can be employed to look for a mean shift in series under observation.

Using this statistical method and observing the period August-September 2020, we have found such a change point in the number of new daily infection cases for all the Italian regions, set approximately around the beginning of September 2020, thus providing a first evidence of the fact that the diffusion of the virus may have been favored by the traditional peak in tourism, typically experienced in Italy, around August 15th.

We present, here, a summarization of this kind of results under the form of a COVID-19 infections plot that gives a visual impression of the general situation, showing the mean value of when a change points occurred on a national basis (red line, Figure 1). This value has been computed as the average of all the regional change points, found with the method mentioned before. As already anticipated, on a national basis, the average change point falls on September 1st (Figure 1; computed as the average value of all the regional change points).

In other terms, with this first change point detection method, we have shown that for most Italian regions, a separation exists between a *before* and an *after*, thus making self-evident that the COVID-19 infections curve in Italy begins to increase robustly around the end of August 2020, a couple of weeks after the peak in tourism (also considering the lag due to the incubation time of the virus and the time needed for the registering process).

At that point, to have a confirmation of the role that summer tourism seemed to play, we developed a statistical learning model for regression (also called, a generalized linear model or GLM) with two different goals:

- estimating the expected number of daily COVID-19 cases based on the 2019 tourist flows and comparing that number with the real infection cases experienced in the summer 2020 in all the Italian regions,

- providing insights on the importance of certain factors in this prediction, like tourisms and others.

Of particular interest, here, is the consideration that, while there are much more sophisticated models than a GLM that can probably provide better predictions, our interest has been more in making guesses on what could be the causes of what lies at the basis of the exponential growth of the COVID-19 infections suffered during the so-called second wave in autumn 2020 in Italy. Based on the above considerations, our analysis should hence be interpreted as an attempt to assess with more precision if tourism had effectively had an impact on the number of new daily cases in the summer 2020.

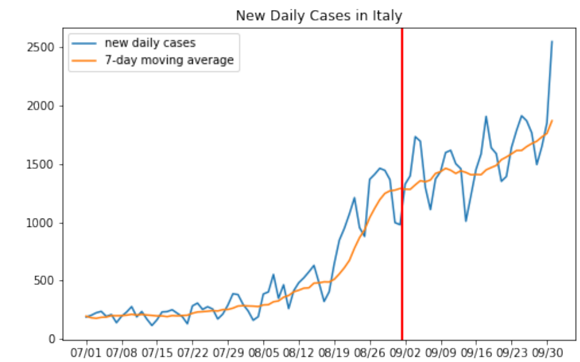

Figure 1. New daily infection cases in Italy (blue), with 7-day moving average (orange) and the average national change point (red), on September 1st.

In other terms, our intent behind the use of the GLM methodology was to demonstrate that some given patterns for the resurgences of the virus could be predicted, if not with the exact amount of the number of infections, at least as an indication of the potential risk. All this with the final objective to inform our policy makers about what should be done next time in similar situations [1].

Technically speaking, we predicted, per each region, the cumulative number of new infections as the result of the influence of the quantity of tourists who have previously visited that region. Further, given the characteristics of that model, into account were taken other variables that could help to get more precise predictions.

The formula on which such regression model was based is given below:

$$ln\,(E(Y_{\mathrm{iw}}|T_{iw}, D_i, H_i, O_i, A_i)) = b_0 + b_1 ln(T_{iw}) + b_2 ln(D_i) + b_3 ln(H_i) + b_4 O_i + b_5 A_i$$

Where: b_0, b_1, b_2, b_3, b_4, and b_5 are the parameters to be estimated with the regression, and:

Y_{iw} is the cumulative number of the new infection cases, occurring in a region i during the time comprised within a period w.

T_{iw} is the sum of inbound and outbound tourists, for a given region i during the aforementioned period w.

D_i is the population density for each region i, measured as the number of inhabitants per km^2.

H_i is the healthcare expenditure for region i expressed as a percentage of the region's Gross Domestic Product.

O_i is the ratio between the total population and the population over the age of 65.

A_i is a variable that takes into account the geographical area to which a given region belongs, and finally ln stands for the natural logarithm.

At the end, as expected, the use of our GLM model was particularly useful for its ability to properly estimate the involved parameters, thus revealing how significant each of the factor mentioned above is.

In particular, tourism and density of population (T and D) appeared as important predictors with the capability to have an influence on the solution of the regression problem of interest, while the predictors associated with healthcare expenditures and the percentage of persons with an age over 65 (H and O) do not seem to play a relevant role.

In summary, the use of our GLM model was a confirmation of the influence of the tourism on the registered infections, being instead of secondary importance, for the count of those infections, how much money a given region spends on healthcare and the percentage of people with an age over 65.

In the next Section, we will see if these preliminary results, obtained using data from the summer 2019, can be finally confirmed with real tourist data registered in the summer 2020 for the regions of Emilia-Romagna, Lazio, Liguria, Lombardia, Piemonte, Umbria, Valle d'Aosta and Veneto.

4 Results with Provisional Data from the 2020 Summer

While the model in our previous work was fitted on 2019 tourist data released by ISTAT, we are now going to test our research hypothesis with provisional 2020 data, made available by the statistical offices of the following Italian regions: Emilia-Romagna, Lazio, Liguria, Lombardia, Piemonte, Umbria, Valle d'Aosta, Veneto [13-20].

To begin, it is important to notice that these 2020 data comprise the numbers of monthly arrivals of Italian tourists to all the aforementioned regions, while in our previous work we could count also on the 2029 outgoing tourist flows (which are, unfortunately, currently not made available by ISTAT).

To give a general idea of these 2020 tourist flows, we have shown in Figure 2 the number of incoming tourist, for each region and per each month, expressed in millions, in the period May-September 2020.

The dataset at the basis of our analysis was then assembled, based on one month-long temporal windows that can slide forward, week by week, comprising the relative number of tourists/COVID-19 cases. This temporal structure was used per each region to put in relationship the number of tourists with the number of COVID-19 infections, as shown in Figure 3.

To span all the period of interest (June-September 2020), these one-month long windows were shifted one week at time, thus engaging two different months at the same time. Utilizing a linear combination of the quantity of the tourists from two different contiguous months, we yielded the total number of tourists to be considered for that given period. As an example, to get the number of tourists between June 15th and July 15th, a quantity equal to the 50% of the total number of tourists for each of the two months was used to yield the final number of tourists.

As to the number of new COVID-19 cases, this was again considered based on one month-wide windows that slide week by week, but shifted forward by 14 days, as this lag is a good approximation of the so-called COVID-19 incubation time (i.e., the time needed for COVID-19 to manifest its symptoms after the relative infection).

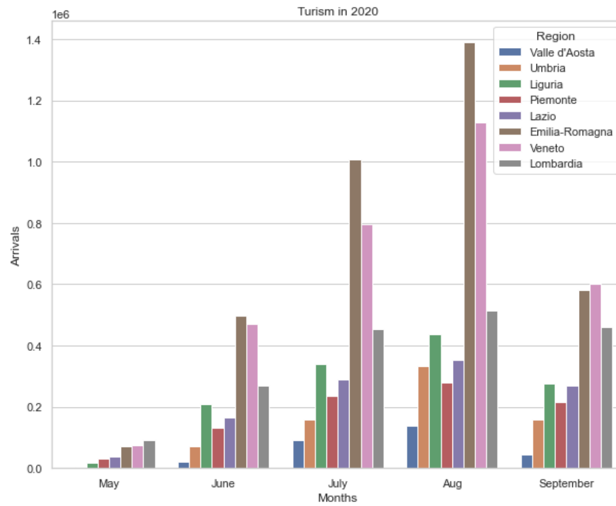

Figure 2: Tourist arrivals on a monthly basis, in the regions of interest. Period: May-September 2020.

To be noticed is also the fact that, differently from [1], the present analysis has also included the month of June. This decision was taken to compensate, with four more time temporal windows, the reduced number of observations given by the fact that only 8 regions had available data.

In Figure 3, this fact is emphasized with the use of the red color, indicating the four windows added for this present analysis.

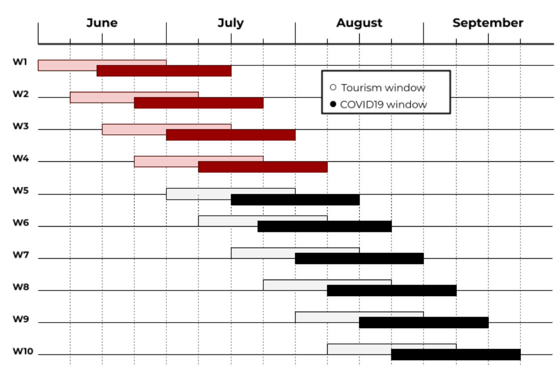

Figure 3: Sliding windows structure used to put in relationship the number of 2020 tourists with the number of COVID-19 infections. Contiguous months are interpolated linearly. In red, windows added for this study.

Upon construction of this structured dataset, we passed to the choice of an adequate statistical model. Inspired by our previous work [1], we developed a GLM model, yet with a reduced number of predictors to take into account the fact that we could count on a limited number of regions to be studied. Essentially, we removed those predictors that in our previous study have appeared to contribute no meaningful effect. Precisely, the annual expenditure on healthcare and the geographic categorical variables were excluded, also considering that the latter factor would not make much sense, as all the regions under consideration are not in the south or in the two islands of Italy. All this considered, the equation at the basis of our GLM was as follows (with the meaning of the predictors Y_{iw}, T_{iw}, D_i, O_i and of the b_i coefficients unaltered with respect to that expressed in Section 3).

$$ln\ (E(Y_{\mathrm{iw}}|T_{iw}, D_i, O_i)) = b_0 + b_1 ln(T_{iw}) + b_2 ln(D_i) + b_3 O_i$$

Fitting our GLM, we achieved the results shown in Table 1.

Table 1: Parameter Estimates for the GLM. The z-values show tourism (T) and density (D) as highly significant.

Coefficients	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-11.1929	1.6939	-6.608	3.9e-11
log(*T*)	1.0166	0.1223	8.315	< 2e-16
log(*D*)	1.2596	0.1409	8.939	< 2e-16
O	-6.7474	3.9403	-1.712	0.0868

Analyzing the two rightmost columns in Table 1 in combination (i.e., high z-values and low p-values), it is evident how tourism is a significant predictor in the fitting of the model, as confirmed by the extremely large values that T_{iw} gets in terms of z-values. Furthermore, the estimates of all the parameters are in line with those we had got in our previous study, suggesting that there is coherence between the tourist data of the 2019 and the 2020 summer. This provides a further support to our initial research hypothesis and justifies the intuition that domestic tourism has played a role in the spread of the pandemic in summer 2020. At the end, in fact, the 2020 tourist data were somewhat proportionate to that of the previous year, given the absence of restrictions to movement put in place by the government in that period. Nonetheless, to be honest, we cannot avoid to point out how this present GLM model, may be quite imprecise in its predictions, especially if we put the focus on final part of the 2020 summer. In fact, technically speaking, while it is true that with a GLM the residuals are not required to be normally distributed, nonetheless they seem to suggest that our model tends to be more inaccurate with those values concerning the rightmost temporal windows of Figure 3, for those regions with were affected with many COVID-19 cases, like Lombardia. A plausible explanation for this limitation is that in September 2020 life was starting to go back to normal and, in turn, this may have caused other factors to overtake the influence of tourism, or perhaps the dynamics of the virus spread are such that the non-linearity of the growth appears after a certain number of active cases. Giving precise answers is a hard task and it goes well beyond the scope of this paper. Nonetheless, it is obvious to observe that, independently on how complex the model is, the scarcity of data allows us to think that we are missing something. Not only, but the 2020 data on tourism have revealed another interesting fact, shown in Figure 4. If we plot the number of tourist arrivals (x) against the number of new COVID-19 cases (y), we observe an evident differentiation between the trajectories of Emilia-Romagna and Veneto, with respect to those of Lombardia and Lazio. In the case of both Emilia-Romagna and Veneto, the number of tourists grows robustly, yet without a corresponding spike in new COVID-19 infections. On the contrary, with Lombardia and Lazio, a large increase in the number of new COVID-19 infections is registered, yet with a moderate increase in the number of tourists. (In the plot, each dot corresponds to one of the 10 windows for each of the 8 considered regions).

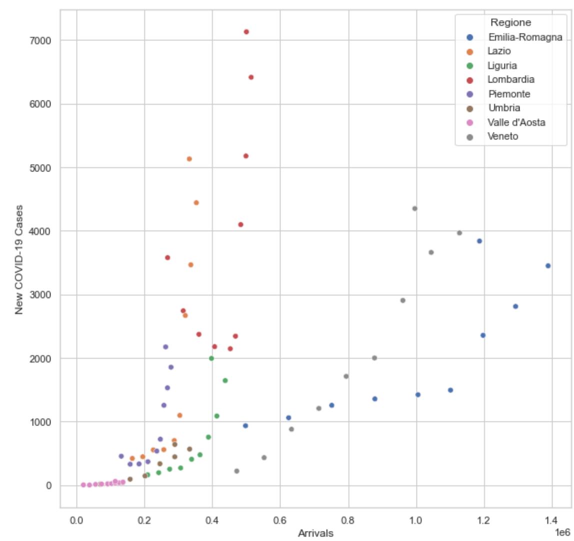

Figure 4: Tourist arrivals vs. new COVID-19 cases. Each dot corresponds to one of the windows for each of the eight regions under consideration.

The reasons for this behavior could be various and surely the fact that the data we have considered are limited to inbound Italian tourists may have played a role. Not only, but it is reasonable to think that Lombardia and Lazio, with their capital cities, could have attracted a large number of tourists from foreign countries, and this could be enough to bridge the gap. It is also well known that Emilia-Romagna and Veneto are the two regions that attract the most part of tourists in Italy (especially domestic tourists), but the duration of their vacations is short, and connected to maritime attractions, art cities and nightlife. This raises the reasonable doubt that the contagion could have happened between tourists, but the relative infections could have been detected only when those tourists have come back home, being thus registered as new COVID-19 cases in their region of residence. A final answer to all these questions will be provided only when the complete dataset will be made available by ISTAT, with a complete knowledge of both outbound and inbound tourist flows, for Italian and foreign tourists alike.

5 Conclusions

We have investigated the relationship between the COVID-19 spread and the 2020 summer tourism across Italian regions. Our hypothesis was that the uptick in the number of new daily cases experienced at the beginning of September 2020 by all regions was linked to the summer holydays traditionally held in Italy in August. Previously, we had found that, assuming a peak of tourist movements around August 15th, the usual highpoint of the season, an upshift of new infections had occurred about two weeks later, at the beginning of September 2020 [1]. A confirmation came from a regression analysis that revealed how tourism was a relevant predictor for the number of new infections, while of secondary importance were other predictors, like population density and healthcare expenditures. An important limitation was that we worked with data from the 2019 summer, assuming that 2020 summer tourist flows (unavailable during our investigation) were proportional. Here, we have validated that supposition, with some new information made available for the 2020 summer: data on the inbound domestic tourist flows of Emilia-Romagna, Lazio, Liguria, Lombardia, Piemonte, Umbria, Valle d'Aosta and Veneto. Drawing on past experiences with this kind of techniques [21-36], we fit a GLM confirming that tourism was a relevant predictor for the number of new infections. We also registered a difference between certain regions when the number of tourist arrivals is compared with the number of new infections. This deserves further studies once the complete ISTAT dataset will be available, allowing us to take into account the 2020 inbound and outbound tourist flows.

REFERENCES

[1] Luca Casini, M Marco Roccetti, "A Cross-Regional analysis of the COVID-19 spread during the 2020 Italian vacations: results from three Computational Models Compared", *Sensors* 20(24), 1–22, 2020.

[2] The Economist. "Tourism flows and death rates suggest COVID-19 is being under-reported". 2020. https://www.economist.com/graphic-detail/2020/03/07/tourism- ows-and-death-rates-suggest-covid-19- is-being-under-reported (accessed 28/4/2021).

[3] Luca Casini, Marco Roccetti, "Reopening Italy's schools in September 2020: a Bayesian estimation of the change in the growth rate of new SARS-CoV-2 cases", *BMJ Open* 11(7), 1–7, 2021.

[4] Susmita Dasgupta, David Wheeler, "Modeling and predicting the spread of COVID-19: comparative results for the United States, the Philippines, and South Africa", *World Bank Group: Washington, DC, USA*, 2020.

[5] Yahua Zhang, Anming Zhang, Jiaoe Wang, "Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China", *Transp. Policy*, 94, 34–42, 2020.

[6] Tamas Krisztin, Philipp Piribauer, Michael Wögerer, "The spatial econometrics of the coronavirus pandemic", *Lett. Spat. Resour. Sci.*, 2020.

[7] Mohammad R. Farzanegan, Hassan F. Gholipour, Mehdi Feizi, Robin Nunkoo, Amir E. Andargoli, "International tourism and outbreak of coronavirus (COVID-19): A Cross-Country analysis", *J. Travel Res.*. 2020.

[8] Martin T. Falk, Eva Hagsten, "The unwanted free rider: COVID-19", *Curr. Issues Tour.* 1–6, 2020

[9] Stefan Gössling, Daniel Scott, C. Michael Hall, "Pandemics, tourism and global change: A rapid assessment of COVID-19", *J. Sustain. Tour.*, 29, 1–20, 2020.

[10] Walter W. Piegorsch, "Maximum likelihood estimation for the negative binomial dispersion parameter". *Biometrics*, 46, 863–867, 1990.

[11] Istituto Italiano di Statistica, Una breve guida alle statistiche sul turismo 2020. Available online: https://www.istat.it/it/archivio/243826

[12] Fran Baum, Toby Freeman, Connie Musolino, Mimi Abramovitz, Wim De Ceukelaire, Joanne Flavel, Sharon Friel, Camila Giugliani, Philippa Howden-Chapman, Nguyen Thanh Huong, Leslie London, Martin McKee, Jennie Popay, Hani Serag, Eugenio Villar, "Explaining COVID-19 performance: what factors might predict national responses?", The BMJ, 372(91), 2021.

[13] Tourism data for Emilia-Romagna. webpage accessed on 28/04/2021. https://statistica.regione.emilia-romagna.it/turismo/dati-preliminari/dati-provvisori-2020

[14] Tourism data for Lazio. webpage accessed on 28/04/2021. http://www.regione.lazio.it/statistica/it/lazio-in-numeri/turismo/arrivi-turisti-fonte-dati-regione-lazio

[15] Tourism data for Liguria. webpage accessed on 28/04/2021. https://www.regione.liguria.it/homepage/turismo/osservatorio-turistico-regionale/analisi-del-movimento-turistico/report-2020

[16] Tourism data for Lombardia, webpage accessed on 28/04/2021. https://explora.in-lombardia.it/flussi-turistici-in-lombardia-anno-2020/

[17] Tourism data for Piemonte, webpage accessed on 28/04/2021. https://servizi.regione.piemonte.it/catalogo/osservatorio-turismo

[18] Tourism data for Umbria, webpage accessed on 28/04/2021. https://www.regione.umbria.it/turismo-attivita-sportive/statistiche-turismo-2020

[19] Tourism data for Valle d'Aosta, webpage accessed on 28/04/2021. https://gestionewww.regione.vda.it/asstur/statistiche/a2020/default_i.aspx

[20] Tourism data for Veneto, webpage accessed on 28/04/2021. http://statistica.regione.veneto.it/banche_dati_economia_turismo.jsp

[21] Marco Roccetti, Giovanni Delnevo, Luca Casini, Giuseppe Cappiello, "Is bigger always better? A controversial journey to the center of machine learning design, with uses and misuses of big data for predicting water meter failures", *J Big Data* 6(1), 20219.

[22] Marco Roccetti, Giovanni Delnevo, Luca Casini, Silvia Mirri, "An alternative approach to dimension reduction for pareto distributed data: a case study", *J Big Data* 8(1), 2021.

[23] Giovanni Delnevo, Marco Roccetti, Silvia Mirri, "Intelligent and good machines? The role of domain and context codification". *Mobile Networks and Applications*. 25(7), 977-985, 2020.

[24] Marco Roccetti, Giovanni Delnevo, Luca Casini, Paola Salomoni, "A cautionary tale for machine learning design: why we still need human-assisted big data analysis", *Mobile Networks and Applications*. 25(3), 1075-1083, 2020.

[25] Silvia Mirri, Marco Roccetti, Giovanni Delnevo, "The new york city COVID-19 spread in the 2020 spring: A study on the potential role of particulate using time series analysis and machine learning". *Applied Sciences* 11(3), 1-19, 2021.

[26] Giovanni Delnevo, Silvia Mirri, Marco Roccetti, "Particulate matter and COVID-19 disease diffusion in Emilia-Romagna (Italy). Already a cold case?" *Computation* 8(2), 59, 2020.

[27] Silvia Mirri, Giovanni Delnevo, Marco Roccetti, "Is a COVID-19 second wave possible in Emilia-Romagna (Italy)? Forecasting a future outbreak with particulate pollution and machine learning", *Computation* 8(3), 74, 2020.

[28] Marco Roccetti, Gustavo Marfia, Marco Zanichelli, "The art and craft of making the Tortellino: playing with a digital gesture recognizer for preparing pasta culinary recipes". *Comput. Entertainment* 8(4), 28, 2010.

[29] Claudio E. Palazzi, Marco Roccetti, Gustavo Marfia, "Realizing the unexploited potential of games on serious challenges". *Computers in Entertainment*, 56(4), 2233–2240, 2010.

[30] Claudio E. Palazzi, Stefano Ferretti, S. Cacciaguerra, Marco Roccetti, "On maintaining interactivity in event delivery synchronization for mirrored game architectures", *GLOBECOM - IEEE Global Telecommunication Conferences, IEEE*, 157-165, 2004.

[31] Flavio Corradini, Roberto Gorrieri, Marco Roccetti, "Performance preorder and competitive equivalence", *Acta Informatica*, 34(11), 805-831, 1997.

[32] Marco Roccetti, Vittorio Ghini, Giovanni Pau, Maria E. Bonfigli, "Design and experimental evaluation of an adaptive playout delay control mechanism for packetized audio for use over the Internet", *Multimedia Tools and Applications,* 14(1), 23-53, 2001.

[33] Silvia Mirri, Marco Roccetti, Paola Salomoni, "Collaborative design of software applications: the role of users", *Human-Centric Computing and Information Sciences*, 8(1), 6, 2018.

[34] Gustavo Marfia, Marco Roccetti, Alessandro Amoroso, Mario Gerla, Giovanni Pau, J-H. Lim, "Cognitive cars: constructing a cognitive playground for VANET research testbeds", *ACM International Conference Proceedings Series, ACM*, 29, 2011.

[35] Marco Roccetti, Mario Gerla, Claudio E. Palazzi, Stefano Ferretti, "First responders' crystal ball: how to scry the emergency from a remote vehicle", *Conference Proceedings IEEE International Performance, Computing and Communications Conference*, 556-561, 2007.

[36] Marco Roccetti, Gustavo Marfia, Paola Salomoni, Catia Prandi, Rocco m. Zagari, Faustine L.G. Kengni, Franco Bazzoli, Marco Montagnani, "Attitudes of Crohn disease patients: infodemiology case study and sentiment analysis of facebook and twitter posts", *JMIR Public Health and Surveillance*, 5(3), e51, 2017.