COVID-19 Data Analysis of the Netherlands 2020 - 2021

Group 15: Emma Bisgaard (ebis@itu.dk), Kirstine Torp Pedersen (kirp@itu.dk), Danielle Dequin (ddeq@itu.dk), Karlis Buiko(kabu@itu.dk), Nicola Clark (niccl@itu.dk)

Introduction

COVID-19 has been present in the Netherlands since February 2020. With the emergence of this new virus governments have had a difficult time choosing the most effective course of action, especially given that various factors may contribute to the spread of the virus. Previous studies, including one by Wiktor Mazin, indicate that weather is one of these contributing factors (Mazin, 2020). These studies informed the decision to explore the relationship of weather and COVID-19 with a specific focus on the Netherlands.

This report investigates and provides an analysis of the correlation between weather factors and COVID-19 cases, for the purpose of investigating the impact on the spread of the virus. Specifically this study focuses on answering the question of whether there is a significant statistical correlation between the number of cases and the UV index, as well as other confounding weather factors. The findings in this report may be used to inform further research to understand the spread of the virus. In particular, to inform whether the Netherlands should prepare for different rates and patterns of COVID-19 spread based on UV levels at different times of the year and in different regions. The findings may also be applicable to other countries, with further research.

Data

This study employs data of corona cases in the Netherlands provided by an open source data initiative, CoronaWatchNL, which collects data on COVID-19 case counts, and standardizes and publishes the data for analysis and academic use (Bruin, 2020). The weather data comes from IBM Pairs, which provides localized, geospatial, and temporal data (IBM, 2021). Furthermore additional data was found regarding lockdown measures in the Netherlands based on news posts from the government (Ministerie, 2021).

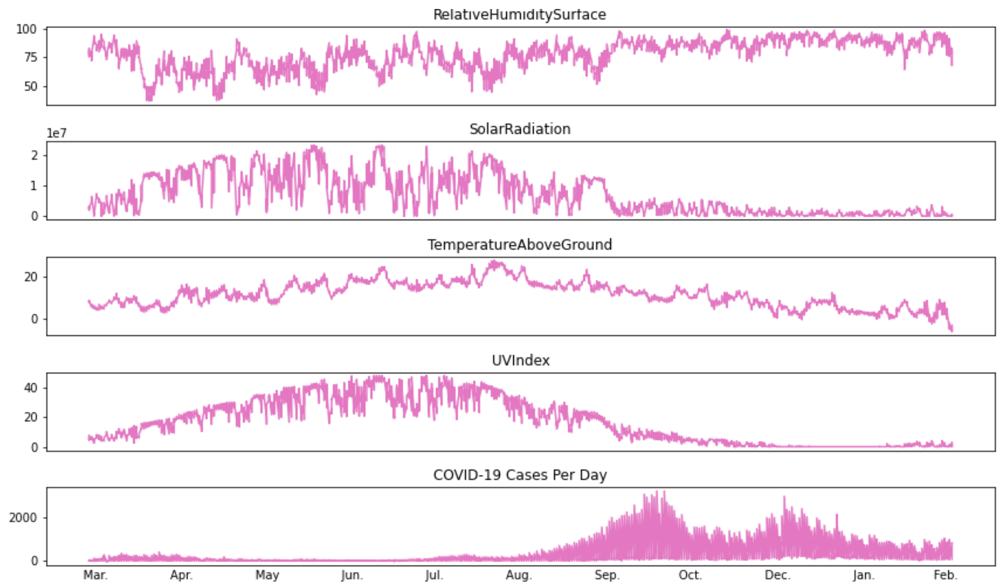
The data regarding COVID-19 cases spanned from February 27, 2020 to February 22, 2021. The first two weeks of this data had many empty fields, that could be attributed to the fact that the virus was still new in this period and that there were very few daily cases. This two week section of data was therefore removed, thus the cleaned data had no missing values. Additionally, country metadata, which included the ISO codes, population information, and geoJSON data of each region, was utilized in the study (Natural Earth, 2021). For later analysis and consistency, region names in the corona data were changed from Dutch to English spelling (CBS, 2020). The ISO codes were used to filter the weather data to include only regions from the Netherlands (Wikipedia contributors, 2021a, 2021b). This localized weather data spanned from February 13, 2020 to February 15, 2021, and contained no missing values, therefore no further cleaning was required.

The World Health Organization states that COVID-19 has an incubation period of 5 - 14 days (“Transmission”, 2020). To account for this the corona data was associated and merged with the weather data from two weeks prior. This resulted in the merged data spanning from February 29, 2020 until February 22, 2021. To investigate confounding factors, the additional data regarding lockdown restriction dates was added to the merged data set (Ministerie, 2021). This data included the dates in which different lockdown measures were active. In this analysis the measures were defined as follows: *School closure*; all schools closed, from preschool to adult education. *Lockdown*; all businesses closed and people advised to stay home. *Travel ban;* travellers not allowed into the country if arriving from countries listed as “red” unless meeting strict requirements.

Results and discussion

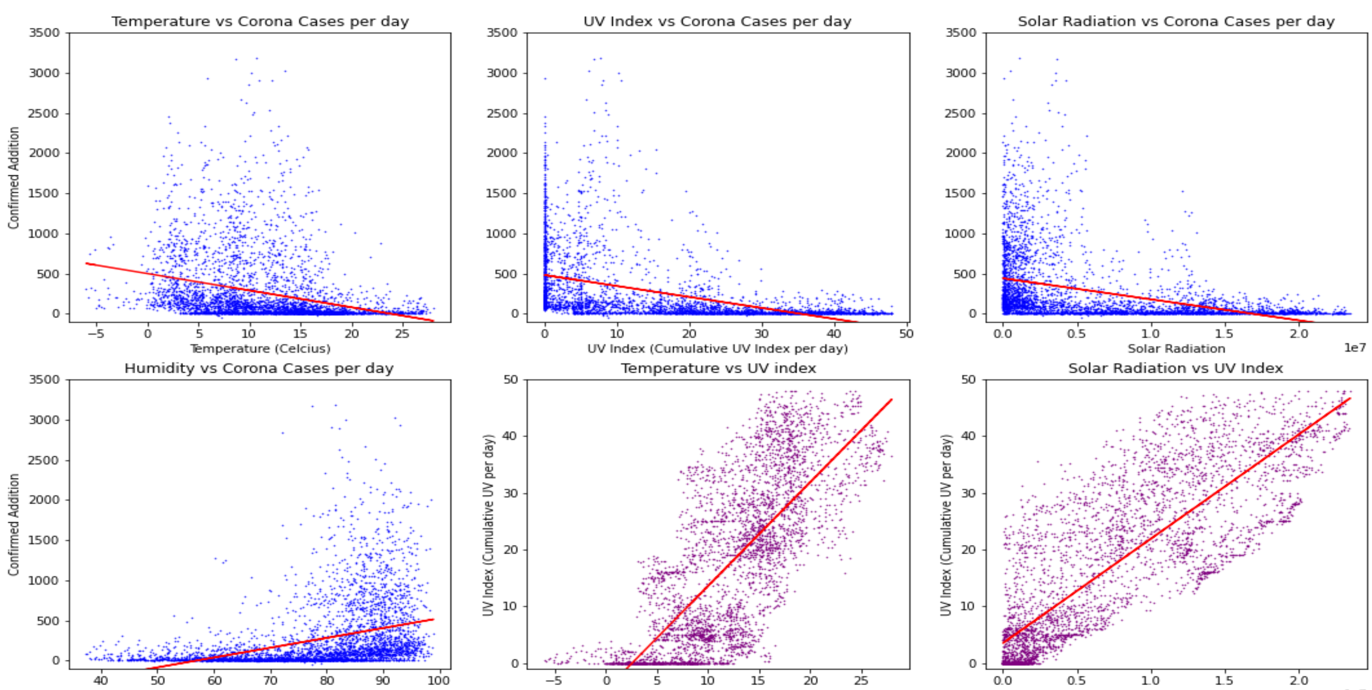
After the data was merged and cleaned, select variables relating to the research question were visualised. Figure 1 below shows the levels per day of the following variables over the span of the data: relative surface humidity, solar radiation, temperature, UV index, and number of COVID-19 cases. The figure shows a potential negative correlation between the number of COVID-19 cases and the variables UV index and solar radiation.

Figure 1: Weather variables and number of COVID-19 cases per day from 29.2.2020 to 22.2.2021



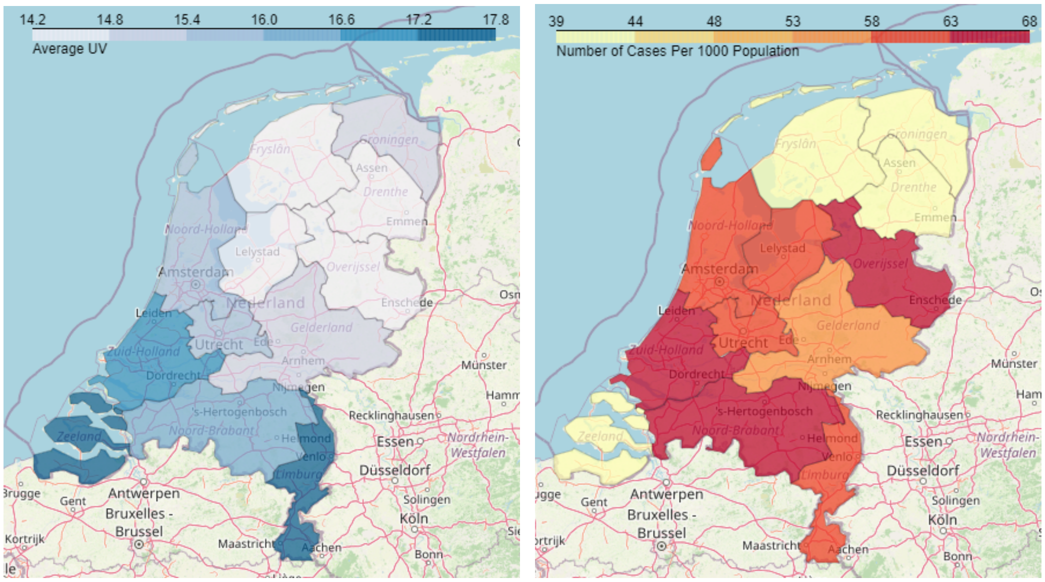
To further visualize these possible correlations, certain variables were plotted using scatterplots along with the lines of best fit, as shown below in figure 2. The plots indicate a slight negative correlation between the number of cases and the variables; temperature, UV and solar radiation. In addition, there seems to be a strong positive correlation between UV index and both temperature and solar radiation.

Figure 2: Weather variables compared to COVID-19 case counts and UV Index



To visualize the average cumulative UV index per day and the number of cases per thousand people, maps were created showing the variance of UV and number of cases with a colour scale by region. From the maps below, a correlation between the two variables does not seem apparent. However, to investigate whether there were significant correlations that could not be seen, appropriate statistical tests were run to find the p-values and associated correlation coefficients.

Figure 3: Maps of the Average Cumulative UV per day and Number of Cases per Thousand People by Region

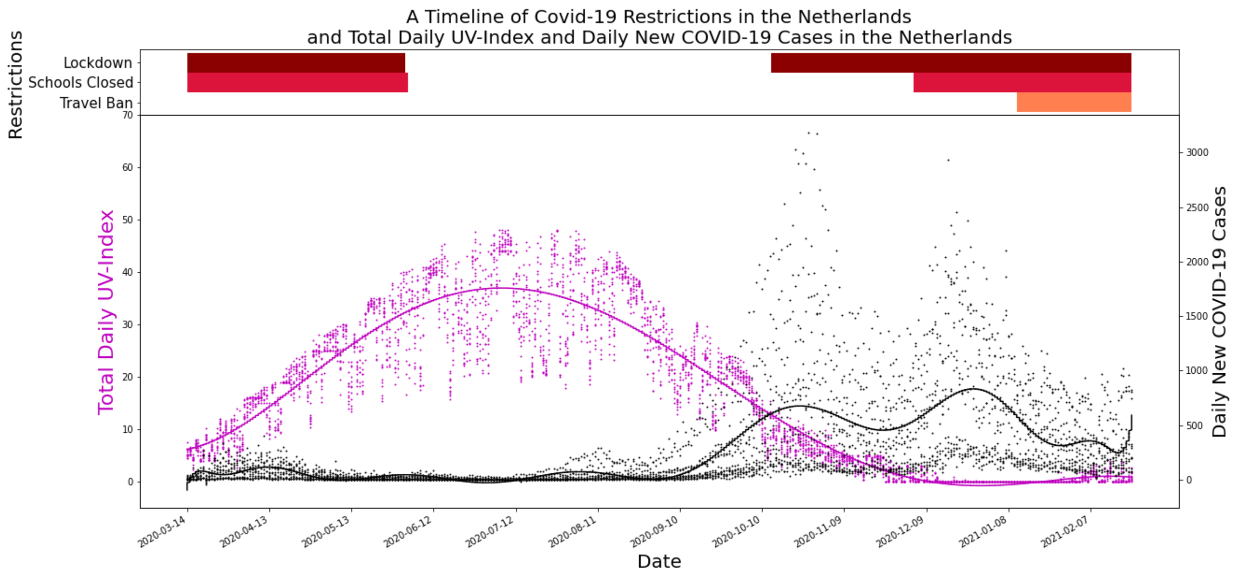


For statistical analysis three tests were utilized, since it was unknown whether the variables followed a linear pattern, had a monotonic relationship, or were skewed. The tests included the Pearson and Spearman correlation tests, as well as another Pearson test with log transformed data. To conduct the various tests, a significance level of 0.001 was chosen, meaning there is only a 1% chance or less that the values are correlated by coincidence. As three tests were conducted, a Bonferonni correction was applied.

The Pearson test evaluates a linear relationship between two variables. It showed a negative correlation between the number of cases and the following variables: solar radiation, temperature and UV index. The UV index's correlation was the most significant with a negative coefficient of -0.467. Meanwhile, the Spearman correlation evaluates the monotonic relationship between two variables based on rank. This test provided further proof of the correlations found by the Pearson test. For instance, UV index had a -0.703 coefficient when run with the Spearman test. In addition, variables that did not seem correlated in the Pearson test remained uncorrelated in the Spearman. Next, to further reduce uncertainty in the results, a Pearson test with log-transformed data was run, as it is able to handle skewed data. This test showed a similar pattern to the previous test, with a negative correlation between the number of cases and solar radiation, temperature, and UV index. UV index stood out with a coefficient of -0.655. This test also further supported results of variables that were previously uncorrelated.

With the results of all three statistical tests aligning with the visualized relationships provided by Figure 2, the study can confidently report that there is a statistically significant relationship between the cumulative UV index per day and the number of corona cases per day. In the figure below, the relationship between UV index and corona cases per day is further visualized, in reference to the dates of restrictions. It shows a possible negative correlation between these two variables, which is further investigated using statistical techniques.

Figure 4: UV Index Compared to Number of Corona Cases per day, including a Timeline of Restrictions



Lastly, a multivariate linear regression test was performed to isolate each variable. This test was run twice; first with the two week delay in corona cases using only the weather variables, and then again after adding in the variables that represent government lockdown measures to see if this would impact the coefficients. For the first test, using a log transformation of the data, the R squared-value was 0.463. This number indicates that 46.3% of the variability of the observed data can be explained by the results of this test. Once again, UV index, solar radiation, and temperature had a p-value of 0, however, the coefficients were much lower. For example the UV index had a coefficient of -0.166, which would indicate that there was less of a correlation. However, it had the largest absolute value of all the t-values in the results. The larger the t-value, the less likely it is that the results occurred by chance, indicating that the correlation with the number of cases and UV was indeed strong.

For the second test, the results did not show a strong correlation, either positive or negative, between government measures and the number of corona cases, which could be due to confounding factors. One possible explanation for a lack of impact of the lockdowns in the results could be that the lockdowns occurred as a response to a strong increase in cases. Therefore, the number of cases could have continued to increase after lockdown measures were in place, since the rate of spread was already high. However, UV index stands out again as being negatively correlated, with a p-value of 0 and a coefficient of -0.169. While the coefficient was low, the absolute value of the t-value is quite high, once again indicating that this correlation is strong.

Limitations

The data in this report is only based on data from the Netherlands, therefore concrete conclusions cannot be generalized for other countries without further investigation. Additionally for future analysis, it would be beneficial to use a more precise model of incubation time in order to accurately analyze correlations. The unclear definition of incubation period, as well as the variance in when COVID-19 tests are done and reported, makes accurate statistical testing difficult. Further data on behaviors of citizens could also be beneficial when exploring the correlation between weather and the spread of the virus, to rule out confounding factors. For example, the relationship between time spent outside and weather conditions could be beneficial to explore.

In addition, precise locations of the cases would have been useful to enable more accurate analysis of the extent that the weather had an impact on the spread of COVID-19. In this way, specific and smaller regions could be isolated as being high risk, which could lead to further investigations into potential causations. To account for further irregularities in testing and reporting, a seven day rolling sum would be beneficial.

Concluding Remarks and Future Work

Using statistical techniques, including the Pearson and Spearman correlation and multivariate linear regression, this study has shown that UV Index is negatively correlated with the number of corona cases per day in the Netherlands. Although, this does not prove causation, as confounding variables need to be taken into account. Further specific analysis with additional data would be needed to conclude that weather variables actually cause a change in the spread of the COVID-19 virus.

With further testing, the government could make use of the results from this study to inform decisions regarding COVID-19. For example, if future tests show that certain weather variables alter the spread of the virus, then the government could preemptively enforce lockdown measures instead of reactively. As a result of this, the government could better prepare for, and more accurately predict the spread of the virus. Infection numbers could decrease as a result of this, and those more vulnerable could be better protected. If the virus is kept at manageable levels, life could continue to run without restrictions, allowing citizens a sense of normalcy. All of this would benefit both physical and mental health. Specifically in regards to the research question, with the knowledge that the winter periods may have a lower UV index, the government could take studies such as this one into account when deciding on restrictions in these periods. For example by increasing surface disinfection measures at this time.

In the future, with a larger data set, including data on citizen behaviour, stronger conclusions on the spread of COVID-19 could be drawn. This would be applicable in the fight against not only COVID-19 but future pandemics as well.

Disclosure statement

One of our group members, Kirstine Torp Pedersen, has been unreachable during this last week and has not contributed equally. Hopefully all is well, but we are mentioning this here to be fair and explicitly state an unequal workload.

References

Bruin, D. J. (2020, March 15). *Number of diagnoses with coronavirus disease (COVID-19) in The Netherlands*. Zenodo. https://zenodo.org/record/3711575

CBS. (2020, December 18). *Regional key figures for the Netherlands*. StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072NED/table?fromstatweb

IBM. (2021). *IBM Geospatial Analytics* [Dataset]. https://ibmpairs.mybluemix.net

Mazin, W. (2020, June 19). *Spearman correlations between the UV Index and COVID-19 per region using IBM PAIRS & Jupyter Notebooks.* Linkedin. https://www.linkedin.com/pulse/spearman-correlations-between-uv-index-covid-19-per-mazin-phd-mmt/

Ministerie. (2021, March 18). *Latest - Government.nl*. Government of the Netherlands. https://www.government.nl/latest/news?keyword=&start-date=&end-date=&element=All+ministries

Natural Earth. (2021). *Natural Earth » 1:10m Cultural Vectors - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales*. Natural Earth: Free Vector and Raster Map Data at 1:10m, 1:50m, and 1:110m Scales. https://www.naturalearthdata.com/downloads/10m-cultural-vectors/

Transmission of SARS-CoV-2: implications for infection prevention precautions. (2020, July 9). World Health Organization. https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions#:%7E:text=The%20incubation%20period%20of%20COVID,to%20a%20confirmed%20case

Wikipedia contributors. (2021a, March 15). *ISO 3166-2*. Wikipedia. https://en.wikipedia.org/wiki/ISO_3166-2

Wikipedia contributors. (2021b, March 16). *ISO 3166-1 alpha-2*. Wikipedia. https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2