



Universiteit  
Leiden  
The Netherlands

## Environmental factors and mobility predict COVID-19 seasonality in the Netherlands

Hoogeveen, M.J.; Kroes, A.C.M.; Hoogeveen, E.K.

### Citation

Hoogeveen, M. J., Kroes, A. C. M., & Hoogeveen, E. K. (2022). Environmental factors and mobility predict COVID-19 seasonality in the Netherlands. *Environmental Research*, 211. doi:10.1016/j.envres.2022.113030

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/3479851>

**Note:** To cite this publication please use the final published version (if applicable).



Contents lists available at ScienceDirect

## Environmental Research

journal homepage: [www.elsevier.com/locate/envres](http://www.elsevier.com/locate/envres)

# Environmental factors and mobility predict COVID-19 seasonality in the Netherlands

Martijn J. Hoogeveen<sup>a,\*</sup>, Aloys C.M. Kroes<sup>b</sup>, Ellen K. Hoogeveen<sup>c</sup>

<sup>a</sup> Department Technical Sciences & Environment, Open University, the Netherlands

<sup>b</sup> Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands

<sup>c</sup> Department of Internal Medicine, Jeroen Bosch Hospital, Den Bosch, the Netherlands

## ARTICLE INFO

## Keywords:

COVID-19 reproduction number  
Solar radiation  
Allergens  
Allergies  
Mobility  
Temperature

## ABSTRACT

**Background:** We recently showed that seasonal patterns of COVID-19 incidence and Influenza-Like Illnesses incidence are highly similar, in a country in the temperate climate zone, such as the Netherlands. We hypothesize that in The Netherlands the same environmental factors and mobility trends that are associated with the seasonality of flu-like illnesses are predictors of COVID-19 seasonality as well.

**Methods:** We used meteorological, pollen/hay fever and mobility data from the Netherlands. For the reproduction number of COVID-19 ( $R_t$ ), we used daily estimates from the Dutch State Institute for Public Health. For all datasets, we selected the overlapping period of COVID-19 and the first allergy season: from February 17, 2020 till September 21, 2020 ( $n = 218$ ). Backward stepwise multiple linear regression was used to develop an environmental prediction model of the  $R_t$  of COVID-19. Next, we studied whether adding mobility trends to an environmental model improved the predictive power.

**Results:** Through stepwise backward multiple linear regression four highly significant ( $p < 0.01$ ) predictive factors are selected in our combined model: temperature, solar radiation, hay fever incidence, and mobility to indoor recreation locations. Our combined model explains 87.5% of the variance of  $R_t$  of COVID-19 and has a good and highly significant fit:  $F(4, 213) = 374.2$ ,  $p < 0.00001$ . This model had a better overall predictive performance than a solely environmental model, which explains 77.3% of the variance of  $R_t$  ( $F(4, 213) = 181.3$ ,  $p < 0.00001$ ).

**Conclusions:** We conclude that the combined mobility and environmental model can adequately predict the seasonality of COVID-19 in a country with a temperate climate like the Netherlands. In this model higher solar radiation, higher temperature and hay fever are related to lower COVID-19 reproduction, and higher mobility to indoor recreation locations is related to an increased COVID-19 spread.

## 1. Introduction

COVID-19 appears to be subject to multi-wave seasonality (Kissler et al., 2020; Grech et al., 2020; Liu et al., 2021; Coccia, 2022), comparable to other respiratory viral infections and pandemics since time immemorial (Moriyama et al., 2020; Fox et al., 2017). It is observed that the COVID-19 community outbreaks have a pattern that is similar to those of other seasonal respiratory viruses (Sajadi et al., 2020; Poole, 2020; Burra et al., 2021; Hoogeveen and Hoogeveen, 2021), whereby the seasonal dips coincide with allergy season in regions in the temperate climate zone (Hoogeveen et al., 2021; Hoogeveen, 2020;

Shah et al., 2021). The same factors that drive the seasonality of flu-like illnesses, appear to drive COVID-19 seasonality: solar radiation including ultraviolet (UV) light, temperature and humidity (Byun et al., 2021), seasonal allergens (i.e., pollens) and allergies, and behavior. Regarding behavior, mobility data show the beneficial effect of restrictive measures on the reproduction number ( $R_t$ ) of COVID-19 (Kajitani and Hatayama, 2021; Nouvellet et al., 2021; Linka et al., 2020), but the seasonal aspects of mobility are often overlooked. For example, during nice weather people spend more time outdoors. The advantage of including mobility data in the model is that it allows to discriminate between indoor and outdoor locations. This distinction is relevant since

\* Corresponding author. Attn. Marga Winnubst, Department of Technology & Environment, Open Universiteit Nederland, Valkenburgerweg 177, 6419, AT, Heerlen, the Netherlands.

E-mail address: [martijn.hoogeveen@icecat.com](mailto:martijn.hoogeveen@icecat.com) (M.J. Hoogeveen).

<https://doi.org/10.1016/j.envres.2022.113030>

Received 15 November 2021; Received in revised form 18 February 2022; Accepted 23 February 2022

Available online 4 March 2022

0013-9351/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

government policies typically restrict mobility regarding specific location types. A single indicator for human-to-human interactions, as for example “commercial trade” (Bontempi, 2020), would not allow us to discriminate mobility per location type. Moreover, selective restrictive policies may distort the association between “commercial trade” and viral spread during our research period. For flu-like illnesses, we previously showed that a compound predictor of solar radiation and seasonal allergens is highly significant though moderately strong  $r(222) = -0.48$  ( $p < 0.001$ ) (Hoogeveen et al., 2021). It is unclear why environmental factors, such as higher solar radiation, a higher level of seasonal allergens (pollens) and subsequently hay fever are consistently associated with a lower  $R_t$  of COVID-19, and, thus possibly associated to COVID-19 seasonality as well. Exposure to solar radiation might be associated with better COVID-19 outcomes (Abraham et al., 2021), and daylight is understood to regulate melatonin levels, and subsequently circadian (lung) immunity (Nosal et al., 2020). Further, increased UV light levels are associated with a more rapid degradation of SARS-CoV-2 particles (Kumar et al., 2021), although the clinical relevance of this effect is debatable.

Upon the observation that allergic diseases are associated with lower rates of COVID-19 hospitalizations (Larsson and Gill, 2021; Keswani et al., 2020), several possible pathophysiological explanations are provided, such as a lower expression of membrane-bound angiotensin-converting enzyme 2 (ACE-2) (Jackson et al., 2020; Wan et al., 2020), higher eosinophil counts that are associated with a more favorable course of COVID-19 (Licari et al., 2020; Lindsley et al., 2020; Ferastroaru et al., 2021), a reduced risk of a cytokine storm or hyper-inflammation (Carli et al., 2020), and T cell-mediated immune responses to allergens which might be effective against COVID-19 as well (Balz et al., 2021). On the other hand, a recent international epidemiological study reported a positive correlation between pollen concentrations and COVID-19 incidence (Damialis et al., 2021). As another study, from Spain, could not confirm the latter finding (Moral de Gregorio et al., 2022), this finding is still a matter of considerable debate.

Further, we noticed that an estimate of  $R_t$  discriminates better between independent variables than incidence metrics (Hoogeveen et al., 2021), as typically used in flu-like and COVID-19 seasonality research (Byun et al., 2021). The  $R_t$  variable is not only a more sensitive metric, but also includes incubation time lags, and is corrected for test bias. Moreover, it is less dependent on seasonality than crude incidence. For such reasons, the  $R_t$  has become the standard in predictive modelling for COVID-19. As a consequence, it typically leads to quite different conclusions regarding the nature of, specifically, the relation between COVID-19 spread and humidity and temperature, as seen in research focusing on crude incidence.

Apart from the use of the  $R_t$ , the novelty of our research is based on the aim to almost fully explain the remarkable fact that during every spring, COVID-19 incidence appears to quickly melt away in the temperate climate zone. We aim to do this by including a very comprehensive set of environmental and mobility parameters in a single predictive model. Whereby we also include recently identified variables such as mobility, solar radiation and seasonal allergens and allergies, that appear to be reliable predictors as discussed above. Finally, among experts there is quite some disagreement on the relative importance of different SARS-CoV-2 transmission routes (Freeman et al., 2021). The statistical elimination of mobility locations and identifying the best predictor, can be helpful to reduce the confusion.

Our hypothesis is that a model, combining both environmental factors and mobility trends, improves the prediction of the seasonality of COVID-19 compared to each factor alone. Therefore, the main objective of this study is to explore a comprehensive model, including both environmental factors and mobility trends of people, to improve the prediction of the reproduction number for COVID-19 during spring season which coincides with the low-season of flu-like respiratory diseases in a country in the temperate climate zone such as the Netherlands

(latitude: 52°N).

## 2. Methods

### 2.1. Data

For the present analyses, we selected the overlapping period of available data sets. The baseline is defined as the first measurements of incidence of COVID-19 on February 17, 2020, and the end date coincides with the end of the first full allergy or pollen season on September 21, 2020, as will be further explained below.

#### 2.1.1. Reproduction number for COVID-19

For the observations of  $R_t$ , we used the respective dataset from the Dutch State Institute for Public Health (Rijksinstituut voor Volksgezondheid en Milieu; RIVM) (dataset]JD-19 r, 2021) from February 17, 2020 till September 21, 2020. RIVM uses a standard method to calculate the  $R_t$  metric on the basis of the input data described below (Wallinga and Lipsitch, 2007). RIVM's  $R_t$  metric is a daily estimate that is based on positive COVID-19 tests in the Netherlands in hospitals from national intensive care foundation (NICE) and from RIVM's own institutes in municipalities (GGD). When the first symptomatic day of a COVID-19 infected person is not known, RIVM estimates this date. Further, RIVM assumes an average 4 days delay period between infection and first symptoms, and estimates the mean incubation period to be 6.4 days (95% confidence interval (CI): 5.6–7.7) (Backer et al., 2020). As the surveillance system of COVID-19 incidence and hospitalizations in the Netherlands, on which  $R_t$  is based, is considered highly reliable and valid, we did not consider to look into alternative metrics such as viral particle counts in waste water (Hu et al., 2021).

#### 2.1.2. Meteorological data

Regarding meteorological data, we used datasets from the Royal Dutch Meteorological Institute (dataset]Daily Mete, 2021) from February 17, 2020 till September 21, 2020. The downloaded daily data included global solar radiation in  $J/cm^2$ , mean relative atmospheric humidity (% RH), and average temperature in degrees Celsius. For comparison, and given their effects on pollen distribution, we also added precipitation duration in 0.1 h, precipitation amount in 0.1 mm, mean wind speed, minimum and maximum temperatures in degrees Celsius, mean dew point temperature in degrees Celsius, and sunshine duration in 0.1 h. Additionally, we calculated the wind chill temperature per day. These datasets were obtained from the KNMI's centrally located De Bilt weather station. De Bilt is traditionally chosen as it provides an approximation of modal meteorological parameters in the Netherlands, which is a small country. Furthermore, all major population centers in the Netherlands, which account for around 70% of the total Dutch population, are within a radius of only 60 km from De Bilt. We therefore assumed in this study that the measurements from De Bilt are sufficiently representative for the meteorological conditions typically experienced by the Dutch population.

#### 2.1.3. Mobility data

We used Google mobility data for relative trends regarding visits to different types of locations in the Netherlands (dataset] Google. Mobility, 2021) for the same period from February 17, 2020 till September 21, 2020. These location types are: Residential, Workplaces, Indoor Recreation (called retail & recreation by Google, which includes restaurants, cafes, retail, shopping centers, theme parks, museums, libraries, and movie theaters), Outdoor Recreation (called Parks by Google, and including places such as national parks, public beaches, marinas, dog parks, plazas, and public gardens), and Transit Stations (places such as public transport hubs such as subway, bus, and train stations). For comparison, as these are less affected by lockdowns, we also included mobility trends for Grocery & Pharmacy (places such as grocery markets, food warehouses, farmers markets, specialty food

shops, drug stores, and pharmacies).

### 2.1.4. Seasonal allergens and allergies

For hay fever (allergic rhinitis) we used the data from Nivel (2021), for the same period, about weekly incidence reports at primary medical care level, per 100,000 citizens in the Netherlands. Primary medical care is the day-to-day, first-line healthcare given by local healthcare practitioners to their registered clients as typical for the Netherlands. The hay fever incidence metric is a weekly average based on a representative group of 40 primary care units, and calculated using the number of hay fever reports per primary care unit divided by the number of patients registered at that unit. This is then averaged for all primary care units and then extrapolated to the complete population. We used interpolation to generate a daily data set.

For comparison, we included daily mean pollen concentrations based on the data from two Dutch pollen stations: Elkerliek Ziekenhuis in Helmond (latitude 51.487059, longitude 5.662036) (Elkerliek, 2020), and Leiden University Medical Center in Leiden (latitude 52.166309, longitude 4.477315) (dataset]Pollen con, 2021). The mean pollen concentration is measured in grains/m<sup>3</sup>, whereby we used the daily totals for the 42 types of pollen particles for which by both stations the numbers are counted and averaged per day per 1 m<sup>3</sup> of air. The common Burkard spore trap is used by these stations. It was noticed before that a metric including all available allergenic particle types, lower allergenic or higher allergenic, correlates stronger with the incidence or R<sub>t</sub> of COVID-19, than a metric only based on higher allergenic particle types (Hoogeveen et al., 2021; Shah et al., 2021). In the Netherlands, these are the only two pollen stations, whereby it is understood that the station in Leiden represents the maritime coastal zone in the Netherlands, below sea level, and the one in Helmond the more continental zone, above sea level. In previous studies, we saw that using the data of a single station already provided a good parameter for trend analysis (Hoogeveen et al., 2021; Hoogeveen, 2020). But, we believe that including data of both stations leads to a better estimate for the Netherlands as a whole.

### 2.2. Data sets consolidation

As said before, we selected for all datasets the same overlapping period of COVID-19 and the first full allergy season (Hoogeveen and Hoogeveen, 2021), during 2020. The overlapping period runs therefore from February 17, 2020 till September 21, 2020 (n = 218 days). The end of allergy or pollen season, we defined as total pollen concentrations structurally dropping below 10 grains/m<sup>3</sup>. Further, all variables were consolidated into a single parameter for the whole country not accounting for spatial variability given that not for all datasets data is available on provincial or municipality level.

For sensitivity analyses, we also extended the datasets to periods till June 10, 2021 (n = 480 days).

For mobility datasets the clearly intra-week patterns required a 7 days moving average to reduce noise. Therefore, for reasons of consistency, we calculated 7 days moving averages for all other variables as well.

### 2.3. Statistical analysis

Variables are presented with their sample sizes (n), means (M), and standard deviations (SD). We calculated correlation coefficients to assess the strength and direction of relations of each independent variable with R<sub>t</sub>, and with each other.

Stepwise backward multiple linear regression for all independent variables on R<sub>t</sub> was used to keep only candidate predictors that are significant (p < 0.05) in the model and remove insignificant predictors. Next, we removed predictors that were multicollinear as defined below. With the remaining independent variables the F-value, standard deviations and errors, degrees of freedom (DF), and significance level, are calculated to test the goodness of fit hypothesis for our predictive model

for R<sub>t</sub>. Further, the multiple R, Multiple R squared (R<sup>2</sup>) and adjusted R<sup>2</sup> correlation coefficients are calculated to estimate the predictive power of our model. Additionally, the algebraic equation to predict R<sub>t</sub> is determined, which is just to be understood as an empirical formula. Next, per independent variable the (standard) coefficient, t-stat and its 95% CI, probability, and the variance inflation factor value (VIF) are calculated as well.

Further, as linear regression assumes normality of the residuals, we applied the Shapiro-Wilk test and to test the homoscedasticity requirement – homogeneity of variance of residuals– the White test is applied. To analyze multicollinearity we used a VIF value of 2.5 as a threshold. Additionally, the priori power is calculated of each predictor alone and compared with the full model. Although the independent variable R<sub>t</sub> assumes time lags, we also studied the autocorrelations of residuals, whereby we interpret an autocorrelation beyond a time lag of 7 days as an indication that our model probably might miss a key predictor. Finally, we created calibration plots to visually review the fit of the model.

For selected independent variables with a p < 0.05 and VIF score <2.5, standard log<sub>10</sub>, square root and quadratic (ˆ2) data transformations are applied to reduce non-linearity in relations between variables which helps to reduce skewness, and, especially, meet the normality and homoscedasticity requirement. Such data transformations do not change the nature and direction of relations between independent variables and R<sub>t</sub>. In case of the relative mobility trend data we added a constant before such data transformations to avoid loss of data because of negative numbers. For other variables that was not necessary as they only included positive numbers.

We reported the results in APA style, adapted to journal requirements, and applied the TRIPOD guidelines in so far applicable.

All statistical analyses were done with Stats Kingdom 2021, which we benchmarked on R version 3.5.

## 3. Results

### 3.1. Variables and their correlations

The sample sizes (N), means, and SDs of the independent variables as used in our multiple linear regression models are summarized in Table 1. The values are given for the data sets after applied data transformations.

During the allergy season, the factors that negatively correlate with R<sub>t</sub>, are in order of strength: hay fever (r(218) = -0.65, p < 0.00001), solar radiation (r(218) = -0.63, p < 0.00001), pollen (r(218) = -0.62, p < 0.00001), and temperature (r(218) = -0.12, p = 0.085). Positively correlated to R<sub>t</sub> are relative humidity (r(218) = 0.55, p < 0.00001) and the related dew point temperature (r(218) = 0.12, p = 0.082). Further, higher relative humidity is associated with rain or fog, and thus reduced solar radiation and lower temperature. Temperature and solar radiation are associated as well, although only moderately strong: r(218) = 0.39, p < 0.00001).

**Table 1**

Overview means (M), standard deviations (SDs) and skewness values.

Variable	N	Mean	SD
Hay Fever	218	131	73.8
Log <sub>10</sub> (Hay Fever)	218	2.06	0.215
Log <sub>10</sub> (Pollen)	218	1.84	0.464
Log <sub>10</sub> (Solar Radiation)	218	3.15	0.273
Log <sub>10</sub> (Solar Radiation <sub>7dma</sub> )	218	3.18	0.198
Temperature <sup>2</sup>	218	221	142
Dew point temperature	218	8.56	5.70
Sqrt(Mobility: Indoor recreation)	218	214	35.8
Sqrt(R <sub>t</sub> )	218	1.03	0.163

**Table 1:** Overview of mean (M), and standard deviation (SD) per independent variable as used in the multiple linear regression models. The function Sqrt returns the square root of the variable.

Pollen and hay fever are, as to be expected, associated:  $r(218) = 0.50$ ,  $p < 0.00001$ , although moderately strong. We did not add allergenicity weights to different pollen particles, and the pollen stations do not cover all types of allergenic particles such as, for example, mold spores. Therefore, having both data sets next to each other has added value, at least for our environmental model.

Solar radiation is an important factor as it has, during allergy season, stimulating effects on pollen ( $r(218) = 0.40$ ,  $p < 0.00001$ ) and subsequently hay fever ( $r(218) = 0.40$ ,  $p < 0.00001$ ), in addition to its associations with temperature and  $R_t$ .

The mobility places that are correlated with  $R_t$  are Indoor Recreation ( $n(218) = 0.761$ ,  $p < 0.00001$ ), Residential ( $n(218) = -0.684$ ,  $p < 0.00001$ ), Transit Stations ( $r(218) = 0.563$ ,  $p < 0.00001$ ), Workplaces ( $r(218) = 0.532$ ,  $p < 0.00001$ ), Grocery & Pharmacy ( $r(218) = 0.472$ ,  $p < 0.00001$ ), and, not significantly, Outdoor Recreation ( $r(218) = -0.048$ ,  $p = 0.5$ ). Indoor Recreation and Residential are most strongly inversely correlated:  $r(218) = -0.817$ ,  $p < 0.00001$ , and thus highly collinear ( $p > 0.8$ ). Indoor Recreation has moderately strong positive correlations with all other mobility variables, and should therefore be seen as a representant of the mobility cluster.

Temperature and dew point temperature had a high correlation of  $r(218) = 0.84$  ( $p < 0.00001$ ), and appear thus to be collinear. These variables although they have, standalone, no significant correlation with  $R_t$ , still play a role in our combined and environmental model, probably because of their indirect effects on mobility and pollen maturation and dispersion, with their opposite associations with  $R_t$ .

### 3.2. Outcomes combined model

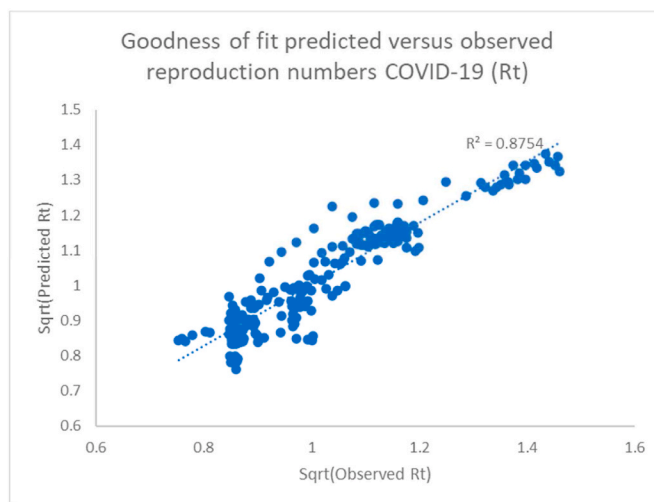
After several iterations with stepwise backward multiple linear regression, four independent variables were selected from the combined pool of environmental and mobility variables that are both significant ( $p < 0.05$ ) and have a VIF value below 2.5. These selected predictors are: temperature, solar radiation, hay fever, and Indoor Recreation (see Table 2). From the mobility datasets, residential was significant as well but was deselected based on its very high multicollinearity with all other mobility variables, homoscedasticity concerns and lowered explanatory power. In other words, staying at home has a beneficial effect, but, does not explain at which out-of-home location most COVID-19 infections occur. Without the hay fever data, the pollen data would have been significant, but using only the pollen data led to homoscedasticity concerns, which were fully mitigated when using the hay fever data instead.

On the basis of the multiple linear regression test, we can reject the null-hypothesis ( $H_0$ ) that our combined predictive model with the four selected factors does not provide a good fit:  $F_{(4, 213)} = 374.2$ ,  $p < 0.00001$ .  $R^2$  equals 0.875, which means that our predictors explain 87.5% of the variance of  $R_t$ . The adjusted R square equals 0.873, and the coefficient of multiple correlation (R) equals 0.936. A simple Pearson correlation between our model's predicted and the observed values for  $R_t$  is equally strong and highly significant:  $r(218) = 0.996$ ,  $p < 0.00001$ . It means that there is a strong, and highly significant, relationship between our combined model's predicted and the observed  $R_t$  of COVID-19

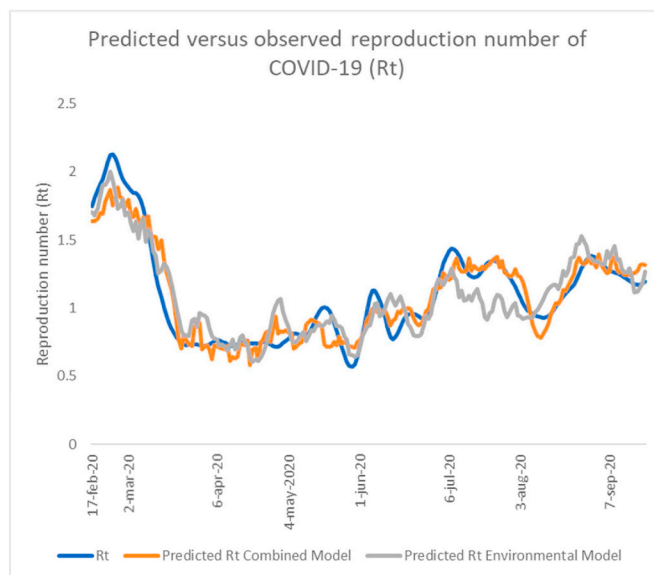
**Table 2**  
Multiple linear regression for mobility and environmental predictors.

	Coeff.	SE	t-stat	lower $t_{0.025(213)}$	upper $t_{0.975(213)}$	Stand. Coeff.	P	VIF
B	0.804	0.0961	8.37	0.615	0.994	0	<0.00001	
Sqrt(Mobility: Indoor recreation)	0.00385	0.000174	22.1	0.0035	0.00419	0.842	<0.00001	2.48
Log <sub>10</sub> (Hay Fever)	-0.132	0.0241	-5.46	-0.179	-0.084	-0.173	<0.00001	1.72
Log <sub>10</sub> (Solar Radiation)	-0.0637	0.0201	-3.17	-0.103	-0.024	-0.106	0.00177	1.93
Temperature <sup>2</sup>	-0.000561	0.0000401	-14.0	-0.00063	-0.000482	-0.489	<0.00001	2.09

**Table 2:** Overview of outcomes per predictor after multiple linear regression for both mobility and environmental variables. Selection of predictors is based on being (highly) significant and having multicollinearity (VIF) score below 2.5. The function Sqrt returns the square root of the variable.



**Fig. 1.** Scatter diagram predicted versus observed reproduction number. Fig. 1: The combined mobility and environmental model is superior as its predictions ( $\hat{R}_t$ ) explain 87.5% of the variance of the observed reproduction number of COVID-19 ( $R_t$ ) during allergy season.



**Fig. 2.** Time series predicted versus observed reproduction number COVID-19. Fig. 2. The time series of the predicted ( $\hat{R}_t$ ) versus the observed reproduction number of COVID-19 ( $R_t$ ) in the Netherlands show the very good fit of both the combined and environmental model during allergy season in the Netherlands. However, the Combined Model predicts  $R_t$  even better. The seasonality effect in March is visible in both model.

(see Fig. 1 and Fig. 2).

The combined predictive model's regression formula looks as follows:

$$\hat{R}_t = \left( 0.804 + 0.00385 \sqrt{M_R} - 0.132 \log_{10} HF - 0.0637 \log_{10} SI - 0.000561 T^2 \right)^2$$

Where  $\hat{R}_t$  is the predicted effective reproduction number for COVID-19,  $M_R$  is the indexed mobility trend data for Indoor Recreation locations to which the mobility constant of 60,000 is added,  $HF$  is the hay fever incidence per 100K citizens,  $SI$  is the mean global solar radiation in  $J/cm^2$ , and  $T$  is the mean temperature in degrees Celsius. In our dataset, the transformed variables only contain positive numbers.

### 3.3. Statistical outcomes environmental model

For the environmental model we excluded mobility data. Again solar radiation and hay fever were selected as predictor of  $R_t$ . The pollen metric added explanatory power, and dew point temperature was selected at the expense of its collinear, temperature (see Table 3). Relative humidity was again deselected as an insignificant predictor.

On the basis of the multiple linear regression test, we can reject the  $H_0$  that our environmental predictive model with the four selected factors does not provide a good fit:  $F_{(4, 213)} = 181.3$ ,  $p < 0.00001$ , and  $R^2$  equals 0.773, which means that our environmental predictors explain 77.3% of the variance of  $R_t$ . The adjusted  $R^2$  equals 0.769, and the coefficient of multiple correlation ( $R$ ) equals 0.879. It means that there is a very strong direct and highly significant relation between our environmental model's predicted and the observed reproduction numbers of COVID-19.

The environmental model's regression formula looks as follows:

$$\hat{R}_t = (3.00 - 0.0587 \log_{10} SA - 0.592 \log_{10} SI + 0.00674 T_d - 0.000262 HF)^2$$

Where  $\hat{R}_t$  is the predicted reproduction number for COVID-19,  $SA$  is average seasonal allergens or pollen concentrations in particles/ $m^3$ ,  $SI$  is the 7 days moving average of global solar (ir)radiation in  $J/cm^2$ ,  $T_d$  is the average dew temperature in degrees Celsius, and  $HF$  is the hay fever incidence per 100K citizens. In our dataset, the transformed variables only contain positive numbers.

## 4. Discussion

The predictive power of the combined environmental-mobility model including solar radiation, hay fever, temperature and visits to Indoor Recreation locations (87.5%) surpasses the environmental model (77.3%) with more than 10%. Furthermore, the improved accuracy of the combined model shows that adding mobility trends not only helps to control the environmental model for lockdown effects, but also clearly improves it by helping to show the importance of seasonal behavior better. For example, nice weather (sun shine, warmth, low humidity) in The Netherlands is related to higher pollen concentrations, and more visits to crowded non-residential locations where social distancing is hard to apply. The latter is in turn associated with increased COVID-19 infections. Interestingly, increased visits to Outdoor Recreation

locations are not associated with an increase in COVID-19 infections ( $R_t$ ). This finding suggests that outdoor transmission of SARS-CoV-2 is far less likely than indoor transmission, and that restrictive policies that limit visiting Outdoor Recreation locations have less added value.

Although, overall, the environmental model is weaker than the combined model, it is still somewhat better at the onset of COVID-19 during February and March 2020. This is probably explained by the exclusion from the mobility data the visits to ski holiday locations abroad, in Italy and Austria, where many of the first patients contracted COVID-19, which leads to an underestimation of both the Indoor Recreation and Outdoor Recreation trend. On the other hand, the combined model is somewhat better in July when lockdown restrictions were relaxed and people were less strict, which is caught well by the mobility trends variable. Both models are almost equally strong in predicting the seasonal decline in March/April, which indicates that the relative importance of restrictive measures was probably not the main driver of that particular decline, but the seasonality effect was.

Of the non-residential locations, especially Indoor Recreation is by far the best predictor of increasing COVID-19 infections ( $R_t$ ), which makes sense as social distancing in busy shopping locations, bars, discos and other such locations, is hard to maintain. Especially, when the seasonality effects are offset by relaxed lockdown measures and social distancing discipline. Even more, if people are under the influence of alcohol and party drugs in crowded party locations, social distancing becomes a distant reality. Additionally, the strong inverse correlation of Residential with  $R_t$ , shows that staying at home, because of lockdown measures, is effective. That all other indoor locations have a positive correlation with  $R_t$ , shows basically the same: when lockdown measures are relaxed, infection rates increase as people will meet more other people.

The single effect of high temperature on  $R_t$  appears to be not significant. The role of temperature can be understood only when its associations with other variables such as mobility trends and pollen maturation and dispersion are taken into account. Humidity in general, relative or specific ( $T_d$ ), appears to be positively associated to COVID-19 reproduction, as it is associated with reduced solar radiation and seasonal allergens, and more traffic to indoor locations which are associated with an increase in infections. Even despite observations that, indoors, very dry air, with a low absolute humidity, might favor SARS-CoV-2 transmission, which is likely caused by increased aerosolisation of infectious aqueous particles. It appears that using  $R_t$  as dependent variable instead of crude incidence, is a good approach to resolve the inconsistencies that are found in literature regarding especially humidity and COVID-19 (Byun et al., 2021). For example, the observation that rainy season in tropical countries coincides with an uptick in flu-like incidence fits our environmental model: the higher outdoor humidity during rainy season is associated with less solar radiation (and UV) and less pollens, which in its turn explains the increase in incidence of viral respiratory diseases. Finally, although we assume that day length is already covered by the solar radiation variable, it might still be interesting to look if this solar-related variable could add something to the predictive power of our models as well.

Given that the same environmental factors and mobility are identified as predictors of COVID-19 spread in other countries, we hypothesize

**Table 3**  
Multiple linear regression for environmental predictors only.

	Coeff.	SE	t-stat	lower $t_{0.025(213)}$	upper $t_{0.975(213)}$	Stand. Coeff.	P	VIF
B	3.00	0.100	30.0	2.80	3.19	0	<0.00001	
Log <sub>10</sub> (Pollen)	-0.0587	0.0144	-4.08	-0.0870	-0.0303	-0.167	0.0000633	1.56
Log <sub>10</sub> (Solar radiation <sub>7dma</sub> )	-0.592	0.0370	-16.0	-0.664	-0.519	-0.717	<0.00001	1.89
Dew point temperature	0.00674	0.00109	6.19	0.00459	0.00888	0.235	<0.00001	1.35
Hay fever	-0.000262	0.0000903	-2.91	-0.000440	-0.0000844	-0.118	0.00405	1.56

**Table 3:** Overview of outcomes per selected environmental predictor after multiple linear regression. Selection of predictors is based on being (highly) significant and having multicollinearity (VIF) score below 2.5.

that our models would have a similar predictive power in other countries in the temperate climate zone with highly similar seasonality patterns. But, as this generalization was outside the scope of our study, further study is needed to test this hypothesis (Hooegeven and Hooegeven, 2021). Further, it can be argued that during flu-like season, outside allergy season, a model with solar radiation, mobility to indoor entertainment locations, and humidity (dew point temperature or specific humidity and relative humidity), but without the allergens or hay fever factor, can be determined with high predictive power. This is both demonstrated in a recent Lombardian study focusing on a combination of UV radiation and mobility trends, which jointly account for 82.6%–85.5% of the variance of  $R_t$  (Falzone et al., 2022). Which is only slightly lower than the outcomes of our combined environmental and mobility model. It is good to note that UV radiation is a collinear of our factor solar radiation.

#### 4.1. Methodological concerns

Test bias, especially for new viruses such as COVID-19, is a major methodological challenge. The approach to use more reliable metrics such as the number of hospitalizations to generate the  $R_t$  metric appears to be a good method to reduce test bias. But, the change of methodology in June 2020, when more test stations were included with their fluctuating test capacities, most likely led to the introduction of test bias in the  $R_t$  metric. Such reliability concerns may have reduced the predictive power of our combined and environmental model.

The usefulness of the pollen concentration metric might be improved by taking into account the allergenicity per particle type. The allergenicity classification is available, but it is not on a ratio scale and there are discussions about the accuracy of this classification. Furthermore, other allergenic particles like mold spores, are hardly ever covered by European pollen stations because of budget constraints.

We observed that the Indoor Recreation and Outdoor Recreation metric might need to be expanded to holiday locations in foreign countries. Unfortunately, that is something that is currently not possible via the Google Mobility datasets. A solution could be to make use of data about airport travelers. Further, it would be of interest to understand whether the predictive power of our models could be improved by using a single, consolidated parameter for human-to-human contacts. In future research, a comparison could be made between the mentioned “commercial trade” substitute (Bontempi, 2020), and a consolidated factor that combines all traffic to out-of-home indoor locations, all correlating positively with  $R_0$ .

In our research we precluded the period of intensive vaccination from January 2021 onward. Given that it is widely observed that the protective immunity of vaccinations or infections is short-lasting, we are still confronted with resurgences of COVID-19 (Edridge et al., 2020). A comparative Italian study concludes that “the COVID-19 pandemic is driven by seasonality and environmental factors that reduce the negative effects in the summer period, regardless control measures and/or the vaccination campaign” (Coccia, 2022). Nevertheless, it would be of interest to test our predictive models for vaccination events. In any case, it is likely that new waves, will be less intense given longer lasting B-cell and T-cell memory of people that have been infected or are vaccinated already. Therefore, it might be good to control also for herd immunity levels when testing the predictive models for subsequent allergy seasons. Additionally, it might be of interest to differentiate the  $R_t$  per virus variant, given that the genetic drift typically leads to more contagious but less deadly variants, that change the dynamics of COVID-19.

Finally, testing the predictive models for a wider geographical scope would be of interest, but would require metrics that are not widely available such as a standardized metrics for  $R_t$ , hay fever incidence, and pollen datasets.

## 5. Conclusion

The combined, mobility and environmental, model explains 87.5% of the variance of  $R_t$  of COVID-19 during spring season in a country in the temperate climate zone like the Netherlands, and provides a very good fit ( $F_{(4, 213)} = 374.2$ ,  $p < 0.00001$ ), as the predicted and observed  $R_t$  correlate strongly and highly significantly. The significant predictors in the combined model are temperature, solar radiation, hay fever incidence, and the Indoor Recreation trend. The environmental factors are inversely associated with  $R_t$ . On the other hand, more visits to Indoor Recreation locations is associated with more infections ( $R_t$ ). This seems to be the best mobility predictor for the effects of lockdown measures on the spread ( $R_t$ ) of COVID-19. On the other side of the spectrum, moving to Outdoor Recreation locations is not significantly associated with changes in  $R_t$ , and including such locations in lockdown regimes appears to be ineffective.

The solely environmental model, is around 10% less powerful than the combined model. Nevertheless, the environmental model shows that pollen concentrations and dew point temperature as a collinear of temperature, have an added explanatory value. Further, there are short periods in which the environmental model beats the combined model.

#### Credit author statement

Martijn Hooegeven: Conceptualization, Methodology, Data curation, Formal analysis, Writing, Investigation, Visualization, Resources. Ellen Hooegeven: Methodology, Writing, Resources. Aloys Kroes: Conceptualization, Reviewing, Data curation.

#### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

#### Data statement

The links to downloadable datasets are provided in the reference list or references to the sources are included. Upon request, the data used for this manuscript is available for inspection, but for other purposes we kindly refer to the respective copyright-holder(s).

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

Thanks to Geert H. Keetels, Bas J.H. van de Wiel of Delft University of Technology for help with dew point temperature calculation, and discussing insights related to specific humidity and COVID-19. Further, thanks to Letty de Weger of Leiden University Medical Center for providing pollen statistics and providing useful insights.

#### References

- Abraham, J., Dowling, K., Florentine, S., 2021. Can optimum solar radiation exposure or supplemented vitamin D intake reduce the severity of COVID-19 symptoms? *Int. J. Environ. Res. Publ. Health* 18 (2), 740. <https://doi.org/10.3390/ijerph18020740>.
- Backer, J.A., Klinkenberg, D., Wallinga, J., 2020 Feb. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travelers from Wuhan, China, 20-28 January 2020. *Euro Surveill.* 25 (5), 2000062 <https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062>.
- Balz, K., Kaushik, A., Chen, M., et al., 2021. Homologies between SARS-CoV-2 and allergen proteins may direct T cell-mediated heterologous immune responses. *Sci. Rep.* 11, 4792. <https://doi.org/10.1038/s41598-021-84320-8>.

- Bontempi, E., 2020. Commercial exchanges instead of air pollution as possible origin of COVID-19 initial diffusion phase in Italy: more efforts are necessary to address interdisciplinary research. *Environ. Res.* 188 <https://doi.org/10.1016/j.envres.2020.109775>.
- Burra, P., Soto-Díaz, K., Chalen, I., Gonzalez-Ricon, R.J., Istanto, D., Caetano-Anollés, G., 2021 January. Temperature and latitude correlate with SARS-CoV-2 epidemiological variables but not with genomic change worldwide. *Evol. Bioinf. Online.* <https://doi.org/10.1177/1176934321989695>.
- Byun, W.S., Heo, S.W., Jo, G., Kim, J.W., Kim, S., Lee, S., Park, H.E., Baek, J.H., 2021 May. Is coronavirus disease (COVID-19) seasonal? A critical analysis of empirical and epidemiological studies at global and local scales. *Environ. Res.* 196, 110972 <https://doi.org/10.1016/j.envres.2021.110972>. Epub 2021 Mar 9.
- Carli, G., Cecchi, L., Stebbing, J., Parronchi, P., Farsi, A., 2020. Is asthma protective against COVID-19? *Allergy* 76 (3), 866–868. <https://doi.org/10.1111/all.14426>.
- Coccia, M., 2022 Jan 13. COVID-19 pandemic over 2020 (with non-pharmaceutical measures) and 2021 (with vaccinations): seasonality and environmental factors. *Environ. Res.*, 112711 <https://doi.org/10.1016/j.envres.2022.112711> (Epub ahead of print).
- Damialis, A., Gilles, S., Sofiev, M., Sofieva, V., Kolek, F., Bayr, D., Plaza, M.P., Leier-Wirtz, V., Kaschuba, S., Ziska, L.H., Bielory, L., Makra, L., Del Mar Trigo, M., 2021 Mar 23. COVID-19/POLLEN study group, Traidl-Hoffmann C. Higher airborne pollen concentrations correlated with increased SARS-CoV-2 infection rates, as evidenced from 31 countries across the globe. *Proc. Natl. Acad. Sci. U. S. A.* 118 (12), e2019034118 <https://doi.org/10.1073/pnas.2019034118>.
- [dataset] Google. Mobility trends. [https://www.gstatic.com/covid19/mobility/Global\\_Mobility\\_Report.csv](https://www.gstatic.com/covid19/mobility/Global_Mobility_Report.csv). Accessed August 2, 2021.
- [dataset] RIVM. COVID-19 reproduction number on the basis of a 95% confidence interval. [https://data.rivm.nl/covid-19/COVID-19\\_reproductiegetal.json](https://data.rivm.nl/covid-19/COVID-19_reproductiegetal.json). Accessed August 2, 2021.
- [dataset] KNMI. Daily Meteorological Data download center. <https://www.daggegevens.knmi.nl/klimatologie/daggegevens>. Accessed August 2, 2021.
- [dataset] LUMC. Pollen concentrations in the Netherlands. <https://www.lumc.nl/org/longziekten/patientenzorg/pollen-en-hoikoorts/pollentelling/>. Accessed August 2, 2021.
- Edridge, A.W.D., Kaczorowska, J., Hoste, A.C.R., et al., 2020. Seasonal coronavirus protective immunity is short-lasting. *Nat. Med.* 26 (11), 1691–1693. <https://doi.org/10.1038/s41591-020-1083-1>.
- [dataset] Elkerliek. Pollen concentrations in The Netherlands. <https://www.elkerliek.nl/elkerliek.nl/specialismen/AKL/Overzicht%202020%20Week%2053.pdf>. Accessed August 2, 2021.
- Falzone YM, Bosco L, Sferruzza G, Russo T, Vabanesi M, and Filippi M. Evaluation of the combined effect of mobility and seasonality on the COVID-19 pandemic: a lombardy-based study. [doi.org:10.2139/ssrn.3778753](https://doi.org/10.2139/ssrn.3778753) (pre-print accessed January 16, 2022).
- Ferastraoaru, D., Hudes, G., Jerschow, E., Jariwala, S., Karagic, M., de Vos, G., Rosenstreich, D., Ramesh, M., 2021 Mar. Eosinophilia in asthma patients is protective against severe COVID-19 illness. *J. Allergy Clin. Immunol. Pract.* 9 (3) <https://doi.org/10.1016/j.jaip.2020.12.045>, 1152–1162.e3, Epub 2021 Jan 23.
- Fox, S.J., Miller, J.C., Meyers, L.A., 2017. Seasonality in risk of pandemic influenza emergence. *PLoS Comput. Biol.* 13 (10), e1005749 <https://doi.org/10.1371/journal.pcbi.1005749>.
- Freeman, A.L., Parker, S., Noakes, C., Fitzgerald, S., Smyth, A., Macbeth, R., Spiegelhalter, D., Rutter, H., 2021 Dec 1. Expert elicitation on the relative importance of possible SARS-CoV-2 transmission routes and the effectiveness of mitigations. *BMJ Open* 11 (12), e050869. <https://doi.org/10.1136/bmjopen-2021-050869>.
- Grech, V., Cuschieri, S., Gauci, C., 2020 Nov 12. COVID-19: the possible seasonal shape of things to come. *Early Hum. Dev.*, 105262 <https://doi.org/10.1016/j.earlhumdev.2020.105262>.
- Hoogeveen, M.J., 2020 July 20. Pollen likely seasonal factor in inhibiting flu-like epidemics. A Dutch study into the inverse relation between pollen counts, hay fever and flu-like incidence 2016–2019. *Sci. Total Environ.* 727, 138543 <https://doi.org/10.1016/j.scitotenv.2020.138543>.
- Hoogeveen, M.J., Hoogeveen, E.K., 2021 Dec. Comparable seasonal pattern for COVID-19 and flu-like illnesses. *One Health* 13, 100277. <https://doi.org/10.1016/j.onehlt.2021.100277>. Epub 2021 Jun 8.
- Hoogeveen, M.J., Van Gorp, E.C.M., Hoogeveen, E.K., 2021 Feb 10. Can pollen explain the seasonality of flu-like illnesses in The Netherlands? *Sci. Total Environ.* 755 (Pt 2), 143182 <https://doi.org/10.1016/j.scitotenv.2020.143182>.
- Hu, L., Deng, W.J., Ying, G.G., Hong, H., 2021 Aug. Environmental perspective of COVID-19: atmospheric and wastewater environment in relation to pandemic. *Ecotoxicol. Environ. Saf.* 219, 112297 <https://doi.org/10.1016/j.ecoenv.2021.112297>. Epub 2021 Apr 30.
- Jackson, D.J., Busse, W.W., Bacharier, L.B., et al., 2020. Association of respiratory allergy, asthma, and expression of the SARS-CoV-2 receptor. *ACE2. J Allergy Clin Immunol.* <https://doi.org/10.1016/j.jaci.2020.04.009>.
- Kajitani, Y., Hatayama, M., 2021 Mar 18. Explaining the effective reproduction number of COVID-19 through mobility and enterprise statistics: evidence from the first wave in Japan. *PLoS One* 16 (3), e0247186. <https://doi.org/10.1371/journal.pone.0247186>.
- Keswani, A., Dhana, K., Rosenthal, J.A., Moore, D., Mahdavinia, M., 2020. Atopy is predictive of a decreased need for hospitalization for coronavirus disease 2019. *Ann. Allergy Asthma Immunol.* 125 (4), 479–481. <https://doi.org/10.1016/j.anaai.2020.07.012>.
- Kissler, S.M., Tedijanto, C., Goldstein, E., Grad, Y.H., Lipsitch, M., 2020. Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period. *Science* 368 (6493), 860–868. <https://doi.org/10.1126/science.abb5793>.
- Kumar, A., Sagdeo, A., Sagdeo, P.R., 2021 Jun. Possibility of using ultraviolet radiation for disinfecting the novel COVID-19. *Photodiagnosis Photodyn. Ther.* 34, 102234 <https://doi.org/10.1016/j.pdpdt.2021.102234>. Epub 2021 Feb 24.
- Larsson, S.C., Gill, D., 2021 Dec 31. Genetic predisposition to allergic diseases is inversely associated with risk of COVID-19. *Allergy.* <https://doi.org/10.1111/all.14728>.
- Licari, A., Votto, M., Brambilla, L., Castagnoli, R., Piccotti, E., Olcese, R., Tosca, M.A., Ciprandi, G., Marseglia, G.L., 2020 Sep. Allergy and asthma in children and adolescents during the COVID outbreak: what we know and how we could prevent allergy and asthma flares. *Allergy* 75 (9), 2402–2405. <https://doi.org/10.1111/all.14369>.
- Lindsley, A.W., Schwartz, J.T., Rothenberg, M.E., 2020 Jul. Eosinophil responses during COVID-19 infections and coronavirus vaccination. *J. Allergy Clin. Immunol.* 146 (1), 1–7. <https://doi.org/10.1016/j.jaci.2020.04.021>.
- Linka, K., Peirlinck, M., Kuhl, E., 2020 Jul 28. The reproduction number of COVID-19 and its correlation with public health interventions. *Comput. Mech.* 1–16. <https://doi.org/10.1007/s00466-020-01880-8> (Epub ahead of print).
- Liu, X., Huang, J., Li, C., Zhao, Y., Wang, D., Huang, Z., Yang, K., 2021 Apr. The role of seasonality in the spread of COVID-19 pandemic. *Environ. Res.* 195, 110874 <https://doi.org/10.1016/j.envres.2021.110874>. Epub 2021 Feb 19.
- Moral de Gregorio, A.J., Guzmán Rodríguez, R., Senent Sánchez, C.J., Lemus Calderón, J. A., Valero Santiago, A.J., 2022. Is environmental pollen concentration a risk factor for SARS-CoV-2 infection? *Investig Allergol Clin Immunol* 32 (1). <https://doi.org/10.18176/jiaci.0743>.
- Moriyama, M., Hugentobler, W.J., Iwasaki, A., 2020. Seasonality of respiratory viral infections. *Annual Review of Virology* 7:1. <https://doi.org/10.1146/annurev-virology-012420-022445>.
- [dataset] Nivel. Hay fever incidence in the Netherlands. [https://www.nivel.nl/sites/default/files/algemene-content/NivelSurveillanceBulletin\\_2021\\_wk29b.pdf](https://www.nivel.nl/sites/default/files/algemene-content/NivelSurveillanceBulletin_2021_wk29b.pdf). Accessed August 2, 2021.
- Nosal, C., Ehlers, A., Haspel, J.A., 2020 Feb 10. Why lungs keep time: circadian rhythms and lung immunity. *Annu. Rev. Physiol.* 82, 391–412. <https://doi.org/10.1146/annurev-physiol-021119-034602>. Epub 2019 Sep. 27.
- Nouvellet, P., Bhatia, S., Cori, A., et al., 2021 Feb 17. Reduction in mobility and COVID-19 transmission. *Nat. Commun.* 12 (1), 1090. <https://doi.org/10.1038/s41467-021-21358-2>.
- Poole, L., March 15, 2020. Seasonal Influences on the Spread of SARS-CoV-2 (COVID19), Causality, and Forecastability (3-15-2020). <https://doi.org/10.2139/ssrn.3554746>.
- Sajadi, M.M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., Amoroso, A., 2020 March 5. Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19. <https://doi.org/10.2139/ssrn.3550308>.
- Shah, R.B., Shah, R.D., Retzinger, D.G., Retzinger, A.C., Retzinger, D.A., Retzinger, G.S., 2021. Competing bioaerosols may influence the seasonality of influenza-like illnesses, including COVID-19. The Chicago experience. *Pathogens* 10 (9), 1204. <https://doi.org/10.3390/pathogens10091204>.
- Wallinga, J., Lipsitch, M., 2007 Feb 22. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* 274 (1609), 599–604. <https://doi.org/10.1098/rspb.2006.3754>.
- Wan, Y., et al., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* [Epub ahead of print] <https://jvi.asm.org/content/94/7/e00127-20.long>.