university of groningen / faculty of science and engineering / mathematics and applied mathematics

# An investigation into the impact of temperature on the rise and spread of COVID-19 in the Netherlands

Bachelor's Project Mathematics

April 2021

Student: S.J.Greengrass

First supervisor: Dr. M.A. Grzegorczyk

Second assessor: Dr. ir. B. Besselink

# An investigation into the impact of temperature on the rise and spread of COVID-19 in the Netherlands

Bachelor Project

Sarah Jane Greengrass s3413942,
Advisor: dr. M.A. Grzegorczyk

20 April 2021

## Abstract

This paper uses an ARMA model to investigate the rise and spread of COVID-19 in the Netherlands per province. The model allows the study of the effect of temperature on the rise of COVID-19. A total of four models are looked at in this paper, a base model which uses purely the number of infections at time $(t-1)$ to predict the number of infections at time $t$, a model which includes temperature and lockdown measures as independent variables, a model which lags these independent variable a suitable number of days and finally a model which includes temperature and lockdown measures as dummy independent variables. Model selection is then assessed by AIC and BIC values. The results suggested that temperature below a certain temperature can be seen to increase the number of infections but only for certain provinces.

# 1 Introduction

COVID-19 is primarily a respiratory and vascular disease associated with the SARS-CoV-2 virus, first discovered in Wuhan, China at the end of 2019 [1]. The disease quickly spread throughout the world and in March 2020 COVID-19 was declared a 'global pandemic'. Since then, countries have been managing outbreaks with various measures that have heavily impacted peoples lives.

At first glance, it looks as if the infection rates have risen and fallen with the seasons. With data of only about 1 year, research investigating into seasonal effects will be likely insignificant. However, an investigation looking into the relationship between temperature and infection rates can be undertaken. This paper aims to identify a relationship between infection rates and temperature in the Netherlands. This research could very well be repeated some time in the future to determine any seasonal effects of COVID-19, which will be important in determining any possible seasonal measures that governments enforce.

The main research question will therefore be "Is there a relationship between temperature and the number of cases of COVID-19 in the Netherlands?". The testable hypothesis is that temperature negatively influences the spread of COVID-19.

It is important to note that the content of this research was influenced heavily on the accessibility of data. Research papers that can be seen in chapter 2 can often access full hospital data and then model far more involved models taking into consideration recovery time and various other factors. This was not possible with the freely available data from the Netherlands.

The Dutch government provides data daily from February 27 2020 of the number of new infections (people tested positive), the number of hospital admissions and the number of deaths, they provide this for each municipality. This paper will focus of the number of new infections, as this measure is the most receptive with respect to changes in environment or measures in place. In other words, the effect in the number of new infections will likely be seen first before any change in hospital admissions or deaths.

In chapter 2, some relevant literature is discussed, which provides some important background information as well as defending some decisions taken in this research. A method section can be found in chapter 3, an outline of the data used can be found in chapter 4 and some simulation details follow in chapter 5. In chapter 6, the results can be seen. Various models are considered in the pre-study and the goodness of fit of these models is discussed in the empirical results section and the best model is determined. In chapter 7 the conclusion can be found and in chapter 8 this paper is discussed critically and any downfalls or improvements are explored. The references and appendix follow the discussion.

# 2 Literature Review

Due to the fact that COVID-19 has had such a large impact in the lives of billions of people, there has been a vast amount of research done on various topics involving COVID-19. Environmental factors have been a highly researched topic in order to identify any possible drivers of infection rates and possible reasons why some countries are worse hit than others.

In the paper by Shi et al. [18] a M-SEIR model is used to model the effect of temperature on the spread of COVID-19 in China. They modified the susceptible-exposed-infectious-recovered model to consider when travel restrictions came into force. A moving average was also considered due to some artificial distortion in the case numbers in Hubei province.
It is important to note that only around 1 month worth of data was investigated over 31 provinces. Temperature data of 344 cities was split among 31 provinces and the means was calculated. The highest temperature in any province between January 20 2020 and February 29 2020 was 26 degrees Celsius and the lowest was -22 degrees Celsius, due to the large and diverse nature of China, the temperature data has a large range. This is an important difference in the research of this paper since the Netherlands is a great deal smaller than China, the climate in the Netherlands has less variation. The paper found that the lowest incidence of infections was at -10 degrees Celsius and the highest incidence was at 10 degrees Celsius. The conclusion found was that transmission rate decreased as temperature increased.

In a second paper by Xie et al. [20] a generalized additive model was used (an extension to a GLM (generalized linear model)) to investigate the relationship between weather and COVID-19. Here, the paper also used a moving-average since the temperature effect could last for a few days and the incubation period for COVID-19 could be up to two weeks. In conclusion, this paper found that the temperature has a positive re-

lationship with the number of COVID-19 cases when the temperature is below 3 degrees Celsius. Interestingly, a 1 degree increase in temperature when the temperature was below 3 degrees was associated with around a 5% increase in COVID-19 cases. However, they did not find evidence to suggest that the cases declined if the weather became warmer.
The paper also experiments with various lagged temperature readings between 0 and 21 days. The paper finds that the relationship between temperature and COVID-19 cases is significantly non-linear.

An ARMA model was used by S.I. Alzahrani et al. [4] to predict the spread of COVID-19 in Saudi Arabia in early 2020. The optimal ARMA orders for their paper were p=2 and q=1, where p is the auto-regressive order and q is the moving-average order. This paper experimented with an auto-regressive (AR) model, a moving-average model (MA), an auto-regressive moving-average model (ARMA) and an auto-regressive integrated moving-average model (ARIMA). ARIMA is used in place of ARMA when there is non-stationarity present in the time series. Non-stationarity is dealt with by differencing any dependent and independent variables a number of times, the optimal number of times is normally when the standard deviation is lowest. There were 4 models in total and cross-validation was used to identify the best model, the data was split into two subsets. One subset was used as a training subset and the other was used as a testing subset. In conclusion, ARIMA(2,1,1) was found to be the best model for their data.
This paper only used the daily and cumulative number of infections of COVID-19. It is important to note that this research was performed in early 2020, therefore the amount of data is limited and it is likely that not many measures were put in place either.

# 3 Mathematical Preliminaries

This section provides an overview of the mathematical background of this paper. Firstly, the ARMA model is introduced then the estimation method of the Kalman filter is detailed then finally the two criterion used in this paper, namely AIC and BIC, are explained.

## 3.1 ARMA model

The main model that will be considered in this paper is the auto-regressive moving average model (ARMA) [8], the main driver of this choice of model was the data that was freely available. This model was first developed by Peter Whittle in 1951 [19]. As the name suggests the model combines an auto-regressive model and a moving-average model.
The ARMA model is used to model time series, it can be used to analyse or predict the trend in the time series. The ARMA(p,q) model can be represented by the following model equation:

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + \sum_{i=1}^{p} \theta_i(y_{t-i} - x_{t-i}\boldsymbol{\beta}) + \epsilon_t + \sum_{i=1}^{q} \alpha_i\epsilon_{t-i} \tag{1}$$

where $p$ is the order of the moving-average part and $q$ is the order of the auto-regressive part of the model. $\theta_i$, $\alpha_i$ and $\boldsymbol{\beta}$ are unknown parameters to be estimated, $\epsilon_t \sim$ i.i.d $N(0, \sigma^2)$ are error terms. $\mathbf{x}_t$ are the possible independent variables and $y_t$ are the dependent variables.
Clearly, if $p$ were to be 0 then the ARMA model would provide a moving-average model and if $q$ were to be 0 then the ARMA model would provide a purely auto-regressive model. The best model (i.e. the best orders $p$ and $q$) can be found via the AIC criterion (see section 3.4).
This model equation 1 can be rewritten in terms of a lag operator $L$ [2]. When dealing with a time series, the lag operator applied to an element of the time series produces the previous element, i.e. $Ly_t = y_{t-1}$. $L^i$ can also be defined as follows:

$$L^i y_t = y_{t-i} \tag{2}$$

The auto-regressive AR(p) part of the ARMA model is defined as $y_t = \mathbf{x}_t\boldsymbol{\beta} + \sum_{i=1}^{p} \theta_i(y_{t-i} - x_{t-i}\boldsymbol{\beta}) + \epsilon_t$ and can be re-expressed in terms of the lag operator as follows:

$$\epsilon_t = (1 - \sum_{i=1}^{p} \theta_i L^i)(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \theta(L)(y_t - \mathbf{x}_t\boldsymbol{\beta}) \tag{3}$$

where $\theta(L) := (1 - \sum_{i=1}^{p} \theta_i L^i)$ to create neat notation.
Similarly, the moving-average MA(q) part of the ARMA model is defined as $y_t = \epsilon_t + \sum_{i=1}^{q} \alpha_i\epsilon_{t-i}$ and can be re-expressed in terms of the lag operator as follows:

$$y_t = (1 + \sum_{i=1}^{q} \alpha_i L^i)\epsilon_t = \alpha(L)\epsilon_t \tag{4}$$

where $\alpha(L) := (1 - \sum_{i=1}^{q} \alpha_i L^i)$ to create neat notation.
Then, combining both the moving average and auto-regressive parts of the model, the ARMA(p,q) model can be expressed in terms of the lag operator as follows:

$$\theta(L)(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \alpha(L)\epsilon_t \tag{5}$$

## 3.2 Kalman Filter

Kalman filtering is a recursive approach to obtain estimates of some unknown variables, with careful definition of variables this method can also be used to obtain estimates of the unknown parameters. This section widely follows the method by Hamilton [12].
Firstly, the model equation needs to be expressed in a state space representation. A state space representation comprises of two parts: a state equation and an observation equation.
Using the notation in terms of the lag operator we can derive the state and observation equations.

The observation equation, let $r = \max\{p, q+1\}$:

$$\begin{aligned}
&\theta(L)(y_t - \mathbf{x}_t\boldsymbol{\beta}) = \alpha(L)\epsilon_t \\
\iff\ & y_t - \mathbf{x}_t\boldsymbol{\beta} = \theta^{-1}(L)\alpha(L)\epsilon_t \\
\iff\ & y_t = \mathbf{x}_t\boldsymbol{\beta} + \alpha(L)\theta^{-1}(L)\epsilon_t \\
\iff\ & y_t = \mathbf{x}_t\boldsymbol{\beta} + \mathbf{x}_t\boldsymbol{\beta} + \alpha(L)\xi_t, \quad \text{let } \xi_t := \theta^{-1}(L)\epsilon_t \\
\iff\ & y_t = \mathbf{x}_t\boldsymbol{\beta} + \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & ... & \alpha_{r-1} \end{bmatrix} \begin{bmatrix} \xi_t \\ \xi_{t-1} \\ \vdots \\ \xi_{t-r+1} \end{bmatrix}
\end{aligned}$$

With some manipulation, the state equation can be obtained:

$$\begin{aligned}
&\xi_t = \theta^{-1}(L)\epsilon_t \\
&\iff \theta(L)\xi_t = \epsilon_t \\
&\iff (1 - \theta_1 L - ... - \theta_r L^r)\xi_t = \epsilon_t \\
&\iff \xi_t = \theta_1\xi_{t-1} + ... + \theta_r\xi_{t-r} + \epsilon_t \\
&\iff \begin{bmatrix} \xi_t \\ \xi_{t-1} \\ \vdots \\ \xi_{t-r+1} \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_r \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \xi_{t-1} \\ \xi_{t-1} \\ \vdots \\ \xi_{t-r} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\end{aligned}$$

Now, let's define the Kalman filter matrices as follows:

$$\begin{aligned}
F &= \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_r \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \\
v_t &= \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
H' &= \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & ... & \alpha_{r-1} \end{bmatrix} \\
w_t &= 0 \\
A' &= \boldsymbol{\beta}
\end{aligned}$$

The state space equations can now be written as:

$$\xi_t = F \cdot \xi_{t-1} + v_t \tag{6}$$

$$y_t = A' \cdot x_t + H' \cdot \xi_t + w_t \tag{7}$$

We will assume that:

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim N\left\{0, \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}\right\}$$

where $Q$ is an $r$x$r$ matrix, the covariance matrix of $v_t$ and $R$ is an $n$x$n$ matrix, the covariance matrix of $w_t$.
The Kalman filter recursions can be seen as an algorithm to calculate the linear least squares forecast of the state vector $\begin{bmatrix} \xi_t \\ \xi_{t-1} \\ \vdots \\ \xi_{t-r+1} \end{bmatrix}$ with respect to the data observed at $t$.

$$\hat{\xi}_{t|t-1} = \hat{E}(\xi_t|\mathcal{Y}_{t-1}) \tag{8}$$

where $\mathcal{Y}_{t-1} = (y'_{t-1}, ..., y'_1, x'_{t-1}, ..., x'_1)'$, and $\hat{E}(\xi_t|\mathcal{Y}_{t-1})$ represents the linear projection of $\xi_t$ on $\mathcal{Y}_{t-1}$ and a constant.
The Kalman filter then obtains $\hat{\xi}_{0|1}, ...., \hat{\xi}_{T|T-1}$ and associated mean squared error matrix ($r$ x $r$):

$$P_{t|t-1} = E[(\xi_t - \hat{\xi}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})'] \tag{9}$$

The above equations 8 and 9 can be rewritten in terms of the Kalman filter matrices:

$$\hat{\xi}_{t|t-1} = F\xi_{t-1} + v_{t-1} \tag{10}$$
$$P_{t|t-1} = FP_{t-1}F' + Q \tag{11}$$

The estimator of $y_t$ can be given as:

$$\hat{y}_{t|t-1} = A'x_t + H'\hat{\xi}_{t|t-1} \tag{12}$$

The error of this estimator is as follows:

$$y_t - \hat{y}_{t|t-1} = A'x_t + H'\xi_t + w_t - A'x_t - H'\hat{\xi}_{t|t-1} = H'(\xi_t - \hat{\xi}_{t|t-1}) + w_t \tag{13}$$

with mean squared error:

$$E[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'] = H'P_{t|t-1}H + R \tag{14}$$

The expected value of $\xi_t$ conditioned on $y_t$ can be defined as follows:

$$\hat{\xi}_{t|t} = \hat{E}(\xi_t|y_t, x_t, \mathcal{Y}_{t-1}) = \hat{E}(\xi_t|\mathcal{Y}_t) \tag{15}$$

This can be calculated by the following:

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + E[(\xi_t - \hat{\xi}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})'] \times (E[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'])^{-1} \times (y_t - \hat{y}_{t|t-1}) \tag{16}$$

This can be simplified by substituting in the equations 9, 14 and 12 to obtain:

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + P_{t|t-1}(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t + H'\hat{\xi}_{t|t-1}) \tag{17}$$

The mean squared error can be found by the following:

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}HM_t^{-1}H'P_{t|t-1} \tag{18}$$

There needs to be initial conditions defined in order to start any recursive formula.
The initial conditions are stated as follows:

$$\xi_{1|0} = 0 \tag{19}$$
$$vec(P_{1|0}) = (I_{r^2} - F \otimes F)^{-1}vec(Q) \tag{20}$$

where the $vec()$ operator produces a column matrix by stacking each successive column of the target matrix.
The Kalman filter can then be used to calculate the log-likelihood function.
The log-likelihood for observation t is as follows:

$$\ln L_t = -\frac{1}{2}\left\{\ln(2\pi) + \ln(|H'P_{t|t-1}H + R|) - (H'(\xi_t - \hat{\xi}_{t|t-1}) + w_t)'(H'P_{t|t-1}H + R)^{-1}(H'(\xi_t - \hat{\xi}_{t|t-1}) + w_t)\right\} \tag{21}$$

This function can then be maximised numerically with respect to the Kalman filter matrices in order to estimate these unknown parameters.

### 3.3 Illustrative example: ARMA(1,1) model

Using equation 1, ARMA(1,1) model can be expressed as:

$$y_t - \mu = \theta(y_{t-1} - \mu) + \epsilon_t + \alpha\epsilon_{t-1}$$

where $\epsilon_t \sim$ i.i.d $N(0, \sigma^2)$.
The Kalman filter matrices can then be defined as:

$$F = \begin{bmatrix} \theta & 0 \\ 1 & 0 \end{bmatrix}$$
$$v_t = \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$$
$$Q = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$
$$w_t = 0$$
$$A' = \mu$$
$$x_t = 1$$
$$H' = \begin{bmatrix} 1 & \alpha \end{bmatrix}$$
$$R = 0$$

The initial conditions are as follows:

$$\hat{\xi}_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$P_{1|0} = \begin{bmatrix} \frac{\sigma^2}{1-\theta^2} & \frac{\theta\sigma^2}{1-\theta^2} \\ \frac{\theta\sigma^2}{1-\theta^2} & \frac{\sigma^2}{1-\theta^2} \end{bmatrix}$$

The recursion then follows from the algorithm above.
The log-likelihood can then be calculated by equation 21 and then maximised numerically to obtain $\theta$ and $\alpha$.

### 3.4 Akaike Information Criterion (AIC)

When comparing the quality of the models, the AIC can provide a objective opinion on which is best. The AIC is an estimator of the prediction error. The AIC is calculated as follows:

$$\text{AIC} = 2n - 2ln(L_{\max}) \tag{22}$$

where $n$ is the number of parameters being estimated in the model and $L_{\max}$ is the maximum value of the likelihood function of the model.
Suppose there are a selection of models, the preferable model would be the one with the minimum AIC, since AIC is an estimator of prediction error and a better model would minimise prediction errors.[3]
Interestingly when comparing AIC values to decide whether one model is preferable or not, the difference is analysed not the relative values compared to the size of the AIC value. [7]

- $\Delta\text{AIC} \leq 2$ : There is little evidence to choose either model
- $4 \leq \Delta\text{AIC} \leq 7$ : There is strong evidence to choose the model with the lower AIC
- $\Delta\text{AIC} \geq 10$ : There is very strong evidence to choose the model with lower AIC

### 3.5 Bayesian Information Criterion (BIC)

Another model selection criterion that can be used to compare models is called the BIC.
The BIC is calculated as follows:

$$\text{BIC} = nln(T) - 2ln(L_{\max}) \tag{23}$$

where $n$ is the number of parameters being estimated in the model, $T$ is the number of observations and $L_{\max}$ is the maximum value of the likelihood function of the model.
Similarly to AIC, the preferable model would be the one with the minimum BIC and when comparing BIC values to decide whether one model is preferable or not, the difference is analysed not the relative values compared to the size of the BIC value. [17]

- $\Delta\text{BIC} \leq 2$ : There is little evidence to prefer one model over another
- $2 \leq \Delta\text{BIC} \leq 6$ : There is evidence to choose the model with the lower BIC
- $6 \leq \Delta\text{BIC} \leq 10$ : There is strong evidence to choose the model with the lower BIC
- $\Delta\text{BIC} \geq 10$ : There is very strong evidence to choose the model with the lower BIC

# 4 Data

This section provides details of the data used in this paper. A description of the data is given, along with various descriptive statistics. Furthermore, the data pre-processing is described and then finally the method of creating the dummy lockdown variable.

## 4.1 Data and Variables

This paper will make use of the database produced by RIVM [1]. The RIVM (Rijksinstituut voor Volksgezondheid en Milieu) has collected the number of people infected, the number of hospital admissions and also the number of deaths on a daily basis since 27/02/2020. This data is recorded per municipality, this data will be summed over the 12 provinces.
Population data (see below in table 1) will also be used to calculate cases per 100,000 people per province. This population data will be taken from CBS (Centraal Bureau voor de Statistiek) [2].
This paper will also make use of the data produced by KNMI (Koninklijk Nederlands Meteorologisch Instituut) [3]. The KNMI has many local stations all over the Netherlands that measure various climatological measurements every day. There were 34 stations that were found to have full temperature readings from 27/02/2020 till 17/02/2021. The stations will be split into 12 provinces where the means will be calculated.

| Province | Population |
|---|---|
| Drenthe | 493,682 |
| Flevoland | 423,021 |
| Fryslân | 649,957 |
| Gelderland | 2,085,952 |
| Groningen | 585,866 |
| Limburg | 1,117,201 |
| Noord-Brabant | 2,562,955 |
| Noord-Holland | 2,879,527 |
| Overijssel | 1,162,406 |
| Utrecht | 1,354,834 |
| Zeeland | 383,488 |
| Zuid-Holland | 3,708,696 |

Table 1: Population data from CBS

## 4.2 Descriptive statistics

In Table 2 the descriptive statistics for the temperature readings of the 12 provinces have been given.

| Province | Mean | Standard deviation | Min | Max |
|---|---|---|---|---|
| Drenthe | 10.4 | 6.3 | -7.0 | 26.4 |
| Flevoland | 10.9 | 6.3 | -6.7 | 25.9 |
| Fryslân | 10.7 | 6.0 | -4.6 | 25.6 |
| Gelderland | 11.0 | 6.5 | -6.7 | 26.8 |
| Groningen | 10.6 | 6.2 | -5.7 | 26.5 |
| Limburg | 11.5 | 6.7 | -7.0 | 27.5 |
| Noord-Brabant | 11.5 | 6.5 | -6.3 | 27.8 |
| Noord-Holland | 11.0 | 6.0 | -4.4 | 25.9 |
| Overijssel | 10.7 | 6.5 | -7.4 | 26.7 |
| Utrecht | 11.2 | 6.3 | -5.6 | 26.8 |
| Zeeland | 11.9 | 6.0 | -4.5 | 27.2 |
| Zuid-Holland | 11.5 | 6.0 | -4.8 | 26.1 |

Table 2: Descriptive statistics for the daily mean temperature readings in degrees Celsius.

[1] https://data.rivm.nl/covid-19/COVID-19_aantallen_gemeente_per_dag.csv
[2] https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072ned/table?dl=3B9A1
[3] http://projects.knmi.nl/klimatologie/daggegevens/selectie.cgi

In Table 3 the mean and maximum values for the number of new infections of the 12 provinces have been given, this provides some context to the provinces that have been particularly effected by the pandemic.

| Province | Mean | Max | Mean per 100,000 people | Max per 100,000 people |
|---|---|---|---|---|
| Drenthe | 57 | 328 | 11.5 | 66.4 |
| Flevoland | 70 | 392 | 16.5 | 92.7 |
| Fryslân | 68 | 325 | 10.5 | 50.0 |
| Gelderland | 330 | 1897 | 15.8 | 90.9 |
| Groningen | 64 | 370 | 10.9 | 63.2 |
| Limburg | 192 | 1233 | 17.2 | 110 |
| Noord-Brabant | 456 | 2224 | 17.8 | 86.8 |
| Noord-Holland | 495 | 1953 | 17.2 | 67.8 |
| Overijssel | 206 | 1410 | 17.7 | 121 |
| Utrecht | 233 | 1164 | 17.2 | 85.9 |
| Zeeland | 44 | 234 | 11.5 | 61.0 |
| Zuid-Holland | 695 | 3188 | 18.7 | 86.0 |

Table 3: Descriptive statistics for the number of new COVID-19 infections and also the number of new COVID-19 infections per 100,000 people per province - N.B. the numbers per 100,000 people will be the values used throughout this paper.

| | Dren | Flev | Frys | Geld | Gron | Limb | N-B | N-H | Ovij | Utr | Zeel | Z-H |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Drenthe | 1 | | | | | | | | | | | |
| Flevoland | 0.8718 | 1 | | | | | | | | | | |
| Fryslân | 0.9426 | 0.8223 | 1 | | | | | | | | | |
| Gelderland | 0.9204 | 0.9412 | 0.8773 | 1 | | | | | | | | |
| Groningen | 0.9240 | 0.8550 | 0.9327 | 0.9014 | 1 | | | | | | | |
| Limburg | 0.9152 | 0.9203 | 0.8978 | 0.9372 | 0.8957 | 1 | | | | | | |
| N-Brabant | 0.8813 | 0.9084 | 0.8510 | 0.9436 | 0.8570 | 0.8874 | 1 | | | | | |
| N-Holland | 0.8889 | 0.8950 | 0.8503 | 0.9317 | 0.8719 | 0.8597 | 0.9546 | 1 | | | | |
| Overijssel | 0.9151 | 0.9148 | 0.8643 | 0.9515 | 0.8616 | 0.9180 | 0.9216 | 0.9125 | 1 | | | |
| Utrecht | 0.8620 | 0.8955 | 0.8082 | 0.9312 | 0.8380 | 0.8416 | 0.9507 | 0.9645 | 0.9136 | 1 | | |
| Zeeland | 0.8874 | 0.8587 | 0.8831 | 0.9211 | 0.8661 | 0.9284 | 0.8849 | 0.8515 | 0.8963 | 0.8292 | 1 | |
| Z-Holland | 0.8192 | 0.8468 | 0.7629 | 0.8854 | 0.7917 | 0.7774 | 0.9448 | 0.9661 | 0.8755 | 0.9567 | 0.7881 | 1 |

Table 4: Correlation matrix of number of infections of provinces

## 4.3 Data processing

As with the majority of data available, it is not always in an ideal format in order to model easily, the data sets obtained here were no exception.
The data from RIVM contained some duplicate recordings that needed to be deleted and since the data was also recorded per municipality and this research will be considering new infections per province, the data was then collapsed into provinces.
The climate data was downloaded per station then merged into one file, the files needed to be pre-processed before being merged. Once the stations were identified to be in a province, the mean was taken over the stations in the same province.

### 4.3.1 Lockdown measures dummy variable

The overview of measures taken by the government website can be found in the appendix. A binary dummy variable was created by generalising the measures taken by the government, this can be seen in table 5 below. This is indeed a drastic generalisation and with further research this generalisation could be more specific to encapsulate each measure, however the aim of this paper is to identify a relationship between the infection rate and temperature therefore an in-depth study on the government measures is not required. It is still important to include the effect of such measures in order to get a more accurate model and the fuller picture.

| Dates | Government action | Binary Value |
|---|---|---|
| 27/02/2020 - 12/03/2020 | No measures | 0 |
| 12/03/2020 - 19/05/2020 | First measures introduced and tightening of measures | 1 |
| 19/05/2020 - 28/09/2020 | Relaxation of measures | 0 |
| 28/09/2020 - 17/02/2021 | Lockdown imposed again and tightening of measures | 1 |

Table 5: Generalisation of measures imposed by the government - please find the full list of all measures taken by the government in the appendix.

# 5 Simulation details

STATA/SE 16.1 is the software that will be used in this paper and the command that will be primarily used is *arima*. For a stationary series, this command obtains maximum likelihood estimates of $\theta_i$, $\alpha_i$ and $\boldsymbol{\beta}$ via a Kalman filter.

## 5.1 Stationarity of data

The data can be described as panel data, as the data set has observations (number of infections) per time unit (per day) for a geographical unit (per province). The concept of stationarity is important in order to model the spread of COVID-19 by an ARMA model.
Stationarity means the statistical properties of the time series do not change over time, the series does indeed change over time but the way in which it changes over time does not change. Stationarity is important when modelling using an auto-regressive model as the model includes a auto-covariance function, if the statistical properties of the data changes of time then the auto-covariance function would be inaccurate, causing bad results.

Various unit root tests were performed in order to test whether the data is stationary: Levin-Lin-Chu unit root test [15], Harris-Tzavalis unit root test [13], Breitung unit root test [5] [6], Im-Pesaran-Shin unit root test [14], Hadri LM test [11] and fisher-type tests [9]. The Levin-Lin-Chu, Harris-Tzavalis, Breitung, Im-Pearson-Shin and fisher-type unit root tests all have the null hypothesis that all the panels contain a unit root. Whereas Hadri LM test has the null hypothesis that all the panels are stationary.
To observe this let's consider a simple equation describing the trend of the panel [16]:

$$y_{it} = (1 - \phi_i)\mu_i + \phi_i y_{i,t-1} + \epsilon_{it} \tag{24}$$

where $i = 1, ..., N$ ($N$ = number of panels) and $t = 1, ..., T$ ($T$ = number of observations). $\epsilon_{it}$ being the error terms, $\mu_i$ is the mean and $\phi_i$ being the parameters to be estimated.
This equation is not ideal to find out which should be the null hypothesis and alternative hypothesis in order to determine whether or not the time series are stationary.
Equation 24 will be rewritten in terms of the change in $y_{it}$:

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \epsilon_{it} \tag{25}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\alpha_i = (1 - \phi_i)\mu_i$ and $\beta_i = -(1 - \phi_i)$.
The null hypothesis in the Levin-Lin-Chu, Harris-Tzavalis, Breitung, Im-Pearson-Shin and fisher-type unit root tests is that all panels contain a unit root, in this case that $\beta_i = 0$ for all $i = 1, ..., N$. The alternative hypothesis is that all panels are stationary, $\beta_i = \beta < 0$. The drawback of tests like these is that the null hypothesis may be rejected even when not all panels are stationary.

Various unit root tests assume a different asymptotic of N/T, this is important to consider in order to find the unit root test which is the most appropriate. In this case, we have that N = 12 and T = 357 and if the long run is considered $N/T \to 0$, as $N$ remains fixed. Levin-Lin-Chu requires that $N/T \to 0$, therefore this test would be the most appropriate to use.

Table 17 in the appendix shows the results of the unit-root tests. The p-values of the Levin-Lin-Chu tests are 0.008, therefore we can reject the null hypothesis and conclude with some certainty that every panel is stationary.

# 6 Results

This results section contains two parts, a pre-study and an empirical results subsection. The pre-study subsection aims to provide details of each model which was considered chronologically and also provide details of some modelling decisions.
The empirical results section aims to focus on the model selection and the performance of each of these models.

## 6.1 Pre-study

In the following section, four models will be discussed as follows:

| Model | Details |
|---|---|
| 1 : **'base model'** | ARMA(1,1) - no independent variables |
| 2 : **'non-lagged model'** | ARMA(1,1) - with temperature and lockdown dummy independent variables |
| 3 : **'lagged model'** | ARMA(1,1) - with lagged temperature and lockdown dummy independent variables |
| 4 : **'dummy variable model'** | ARMA(1,1) - with temperature dummy variable and lagged lockdown dummy variable |

Table 6: Model overview of section 6.1

### 6.1.1 ARMA(1,1) model without independent variables - Base model

This model provides a baseline in order to determine if the addition of independent variables into the model will improve the AIC value or not.
From section 3, the model equation is as follows:

$$y_t = c + \theta_1 y_{t-1} + \epsilon_t + \alpha_1 \epsilon_{t-1} \tag{26}$$

with $p$ and $q$ set to 1 initially.
The best and the worst fitting provinces will be included below in table 7 and the graphs in figures 1 and 2. In table 7 the estimated parameters can be found for provinces Fryslân and Overijssel, the AIC and BIC are also given. The SSR (sum of squared residuals) was then computed by:

$$SSR = \sum_{i=1}^{T} (y_t - \hat{y}_t)^2$$

where $\hat{y}_t$ is the predicted value of $y_t$ based on the estimated parameters and $y_t$ is the actual value from the data. The rest of the provinces will be included in the appendix (see table 23), where the estimated parameters, AIC and BIC can be found.
The figures 1 and 2 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' and the actual values in the red line called 'normed_infections'. In the entirety of this paper normed infections are the $y_t$ values which are the number of cases per 100,000 people per province.
Visually, both the graphs of provinces of Fryslân and Overijssel look quite similar fit-wise. However looking at the SSR, the predicted values of infection rates of Fryslân fit much better that the province of Overijssel. This could be down to the fact that Fryslân is about half the size of Overijssel or perhaps the actual data of the provinces could be quite different. The data of Fryslân appears to have more peaks than Overijssel and the quantity of infections is a lot less per 100,000 in Fryslân than in Overijssel.

| Variables | Fryslân | Overijssel |
|---|---|---|
| $\theta_1$ (AR part) | 0.9927 | 0.9841 |
| $\alpha_1$ (MA part) | -0.6220 | -0.5285 |
| $c$ | 10.48 | 14.81 |
| AIC | 1998 | 2533 |
| BIC | 2014 | 2549 |
| SSR | 5569 | 24719 |

Table 7: Base model - ARMA(1,1) without indep. vars., this table provides the estimated parameters (as in the model equation) and also provides AIC, BIC and SSR values for the best and worst fitting provinces

The figures 1 and 2 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' (this results from using the estimated parameters with the value of $y_{t-1}$ to predict the value $y_t$) and the actual values in the red line called 'normed_infections'. In the entirety of this paper normed infections are the $y_t$ values which are the number of cases per 100,000 people per province.

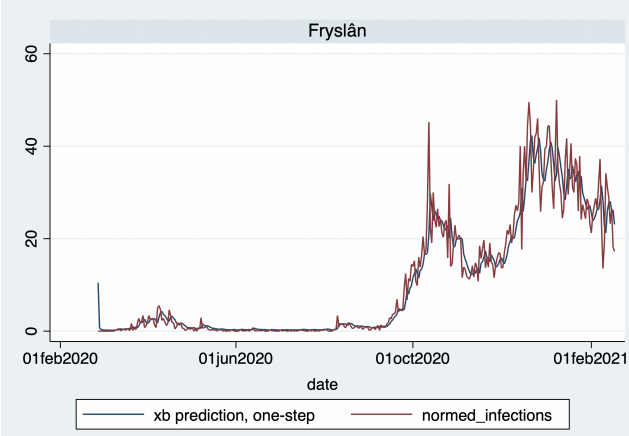


Figure 1: Predicted vs actual normed infections of Fryslân (base model)

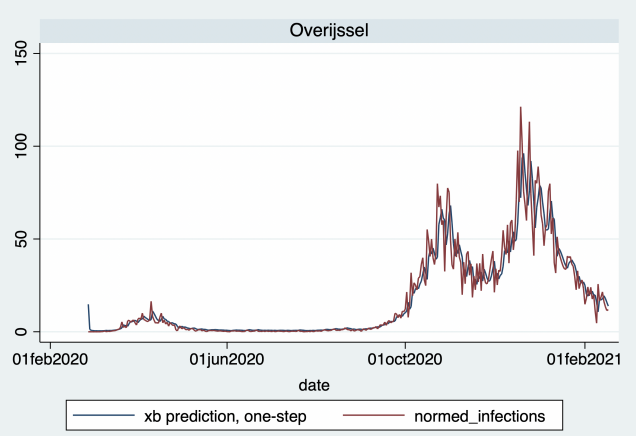


Figure 2: Predicted vs actual normed infections of Overijssel (base model)

### 6.1.2 Confirming the auto-regressive and moving average order

The auto-regressive order of this model has been 1 and the moving average order has been also 1, however the model could perhaps be better with a different order. Find the results tables 18 and 19 in the appendix. Below are the figures 3 and 4 to visually display the best orders, the tables of the exact values can be found in the appendix in table 18 and 19.
The AIC values were not checked for all provinces but the only the best and worst fitting provinces which is sufficient in order to check the orders.
It is clear from the tables that for both provinces the best AIC value occurs when both the moving average order and auto-regressive order is 1.

The figures 3 and 4 visually show the AIC values of the moving average and auto-regressive orders of the best and worst fitting provinces.

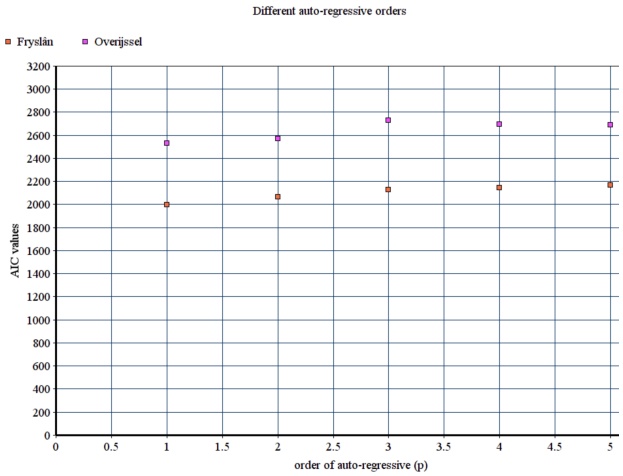


Figure 3: Figure of AIC values for different auto-regressive orders of the provinces Fryslân and Overijssel

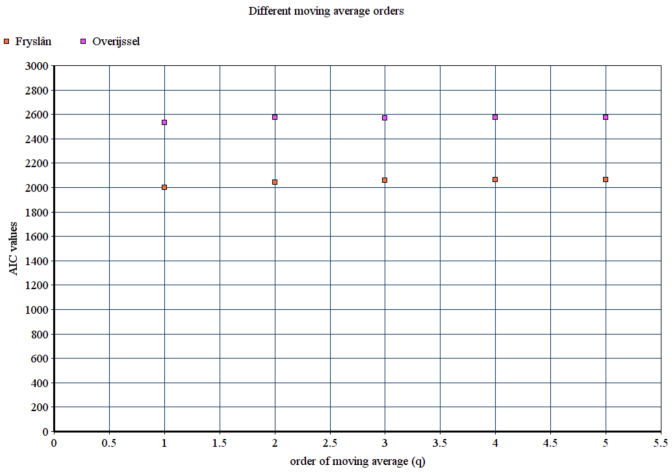


Figure 4: Figure of AIC values for different moving average orders of the provinces Fryslân and Overijssel

### 6.1.3 ARMA(1,1) model with temperature and lockdown dummy variables - non-lagged model

The model equation will be as follows:

$$y_t = c + \theta_1 y_{t-1} + \epsilon_t + \alpha_1 \epsilon_{t-1} + \beta x_t + \gamma lockdown_t \tag{27}$$

Similar to above but now with $x_t$ the daily temperature readings and $lockdown_t$ the lockdown dummy variable (see data section above for explanation of this variable).
The best and the worst fitting provinces will be included below in table 8 and the graphs 5 and 6. In table 8 the estimated parameters can be found for provinces Fryslân and Overijssel, the AIC, BIC and SSR are also given.
The rest of the provinces will be included in the appendix (see table 24), where the estimated parameters, AIC and BIC can be found.
The figures 5 and 6 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' and the actual values in the red line called 'normed_infections'.
The SSR values of both Fryslân and Overijssel are better than in model 1 by a small margin. Yet the model for Fryslân fits a great deal better than Overijssel which still could be down to the reasons given for model 1.

| Variables | Fryslân | Overijssel |
|---|---|---|
| $\beta$ | -0.04625 | -0.2749 |
| $\gamma$ | 1.320 | 1.9205 |
| $\theta_1$ (AR part) | 0.9921 | 0.9813 |
| $\alpha_1$ (MA part) | -0.6249 | -0.5093 |
| $c$ | 10.12 | 16.66 |
| AIC | 2001 | 2535 |
| BIC | 2025 | 2558 |
| SSR | 5548 | 24577 |

Table 8: non-lagged model - ARMA(1,1) with indep. vars., this table provides the estimated parameters (as in the model equation) and also provides AIC, BIC and SSR values for the best and worst fitting provinces

The figures 5 and 6 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' (this results from using the estimated parameters with the value of $y_{t-1}$ to predict the value $y_t$) and the actual values in the red line called 'normed_infections'.

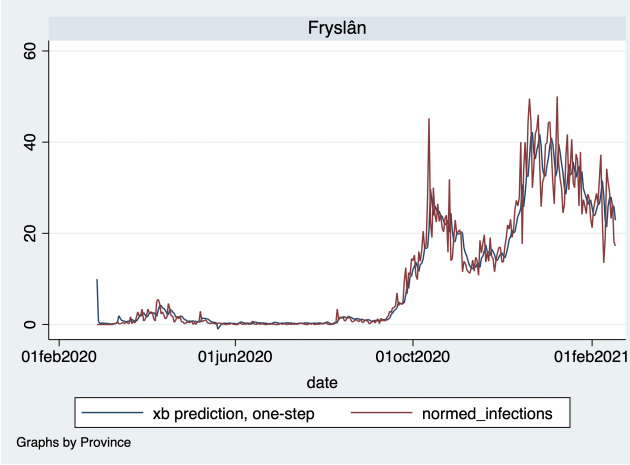


Figure 5: Predicted vs actual normed infections of Fryslân (non-lagged model)

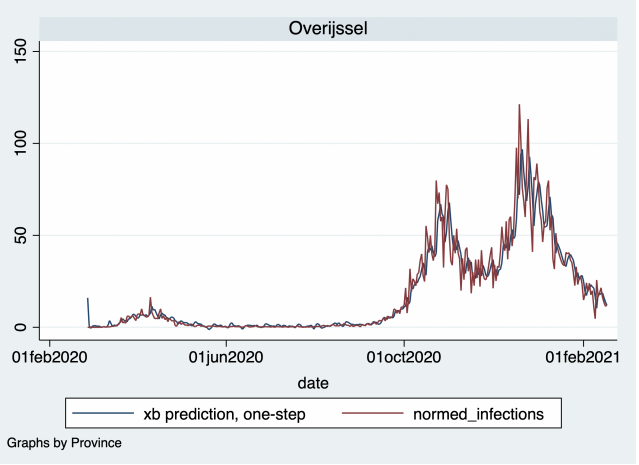


Figure 6: Predicted vs actual normed infections of Overijssel (non-lagged model)

#### 6.1.4 ARMA(1,1) model with lagged temperature readings and lagged lockdown measures - lagged model

When looking at the effect of temperature of the number of infections, it is logical to expect that there should be a lag on the temperature readings since the effect would likely take some time to see in the number of infections. Similarly with lockdown measures, it takes some time for the measure to make a difference.
In the Netherlands after someone has been abroad they should self-quarantine for 10 days, it is usually thought that it takes around 10 days for someone to be symptomatic. Using this as a starting point, 7-21 days will be experimented with and the AIC values recorded, this can be found in figure 7 below, the table of the exact values can be found in the appendix in table 20.
The AIC values were not checked for all provinces but the only the best and worst fitting provinces which is sufficient in order to check the lags.
Interestingly, for the majority the number of lags didn't change the AIC value greatly, which implies that it doesn't have a great impact on this model.
For the province of Fryslân, the difference in AIC values don't seem to level out, however for the province of Overijssel, the difference in AIC values does level out at 10 days, as the difference in AIC values remains 6 for 3 days.
Therefore the lag value of temperature will be chose at 10 days.

With the lockdown measures, it becomes a little more difficult to discern how long it takes for a certain measure to work, especially when the measures have been generalised as much as this.
It is fairly certain that a certain measure takes a number of weeks, therefore as a starting point 1-5 weeks will be experimented with.
As above, the AIC values will be checked for the best and worst fitting provinces, find figure 8 below, the table of exact values can be found in the appendix (see table 21).
For both provinces after 3 weeks the differences between AIC values become similar, therefore the lag value that will be taken is 3 weeks.

The figure 7 shows the AIC values of the best and worst fitting provinces for different number of days in which the temperature variable was lagged. The figure 8 shows the AIC values of the best and worst fitting provinces for different number of weeks in which the dummy lockdown variable was lagged.

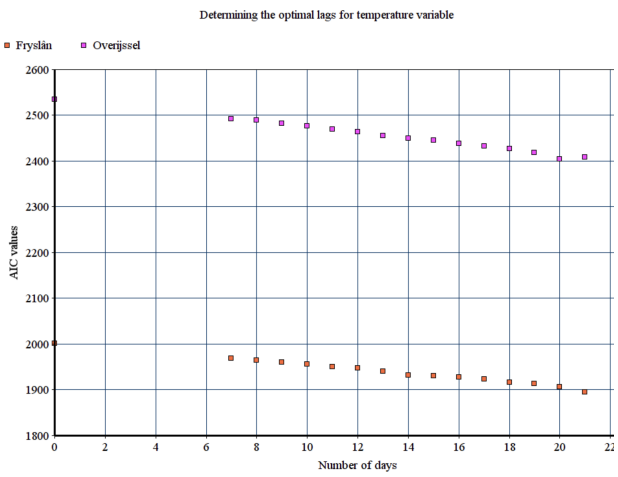


Figure 7: Figure of AIC values for different lags for temperature variable

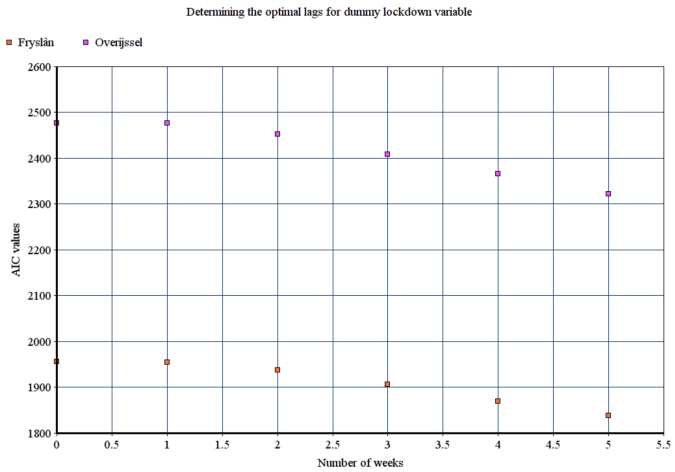


Figure 8: Figure of AIC values for different lags for lockdown dummy variable

The model equation will be as follows:

$$y_t = c + \theta_1 y_{t-1} + \epsilon_t + \alpha_1 \epsilon_{t-1} + \beta x_{t-10} + \gamma lockdown_{t-21} \tag{28}$$

The same model equation as above but now with the lagged temperature variable and the lagged lockdown dummy variable.
The best and the worst fitting provinces will be included below in table 9 and the graphs 9 and 10. In table 9 the estimated parameters can be found for provinces Fryslân and Overijssel, the AIC, BIC and SSR are also given.
The rest of the provinces will be included in the appendix (see table 25), where the estimated parameters, AIC and BIC can be found.
The figures 9 and 10 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' and the actual values in the red line called 'normed_infections'.
The SSR value of Fryslân is worse than for model 1 and the SSR value of Overijssel is slightly better than model 1. Still the model for Fryslân fits a great deal better than Overijssel which could be down to the reasons given for model 1.

| **Variables** | **Fryslân** | **Overijssel** |
|---|---|---|
| $\beta$ | 0.00805 | 0.004231 |
| $\gamma$ | -0.3303 | 0.3710 |
| $\theta_1$ (AR part) | 0.9924 | 0.9833 |
| $\alpha_1$ (MA part) | -0.6219 | -0.5292 |
| $c$ | 11.08 | 15.68 |
| AIC | 1906 | 2409 |
| BIC | 1928 | 2432 |
| SSR | 5574 | 24695 |

Table 9: lagged model - ARMA(1,1) with lagged indep. vars., this table provides the estimated parameters (as in the model equation) and also provides AIC, BIC and SSR values for the best and worst fitting provinces

The figures 9 and 10 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' (this results from using the estimated parameters with the value of $y_{t-1}$ to predict the value $y_t$) and the actual values in the red line called 'normed_infections'.

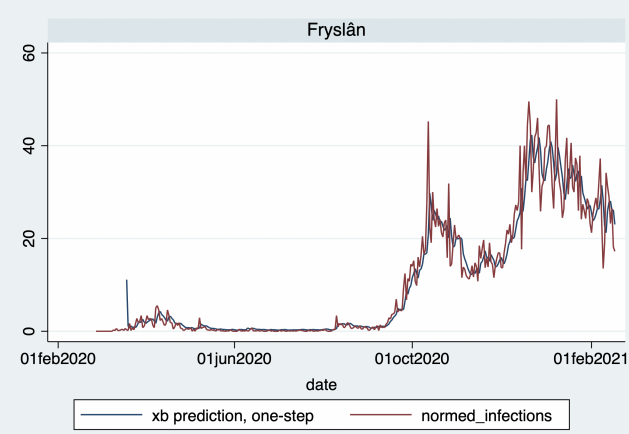


Figure 9: Predicted vs actual normed infections of Fryslân (lagged model)

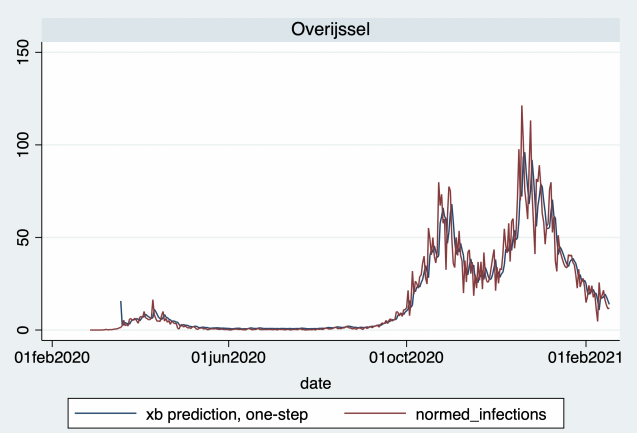


Figure 10: Predicted vs actual normed infections of Overijssel (lagged model)

### 6.1.5 ARMA(1,1) model with temperature dummy variable - dummy variable model

From above, the lockdown measures seem to create a bigger difference in the AIC values than the temperature readings. Following the approach of Xie et al.[20], the paper found that when the temperature is below 3 degrees Celsius that then the temperature has a positive relationship with the number of COVID-19 cases.
In this section, a dummy variable for temperature will be created to test the theory of Xie et al. In which the variable will be 1 when below the cutoff temperature and 0 when it is above the cutoff temperature.
As above, the AIC values of the best and worst fitting provinces will be analysed. See figure 11, the table of exact values can be found in the appendix (see table 22). For the best fitting province (Fryslân) the best cutoff temperature is clearly -3 degrees, however when looking at the worst fitting province, the best cutoff temperature is at 3 degrees. This perhaps implies that each province has a different cutoff temperature. Thus each province will be investigated, see table 10.
The model equation will be as follows:

$$y_t = c + \theta_1 y_{t-1} + \epsilon_t + \alpha_1 \epsilon_{t-1} + \beta temp_t + \gamma lockdown_{t-21} \tag{29}$$

Similar to the other model equations, however now with dummy temperature variable $temp_t$.

| Province | Optimal cutoff temp | AIC |
|---|---|---|
| Drenthe | -4 | 2066 |
| Flevoland | 3 | 2259 |
| Fryslân | -3 | 1890 |
| Gelderland | -1 | 2133 |
| Groningen | -3 | 2082 |
| Limburg | 4 | 2135 |
| Noord-Brabant | 1 | 2074 |
| Noord-Holland | -4 | 1958 |
| Overijssel | 3 | 2407 |
| Utrecht | -4 | 2171 |
| Zeeland | 3 | 2146 |
| Zuid-Holland | -4 | 2079 |

Table 10: Optimal cutoff temperature in degrees Celsius per province - each optimal cutoff temperature was found by AIC (as was demonstrated for Fryslân and Overijssel)

The best and the worst fitting provinces will be included below in table 11 and the graphs 12 and 13. In table 11 the estimated parameters can be found for provinces Fryslân and Overijssel, the AIC, BIC and SSR are also given.
The rest of the provinces will be included in the appendix (see table 26), where the estimated parameters, AIC

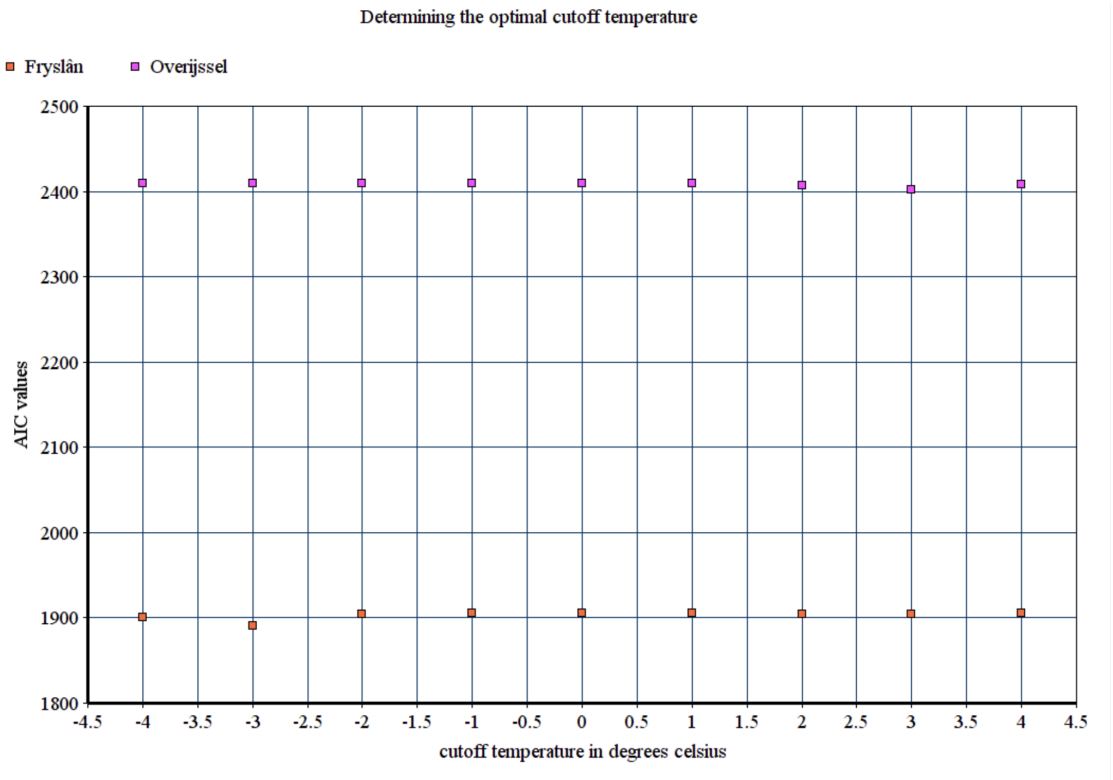


Figure 11: Figure of AIC values for different cutoff temperatures (in order to create a temperature dummy variable - which is 1 when below that certain temperature and 0 when above that certain temperature) of provinces Fyslân and Overijssel

and BIC can be found.
The figures 12 and 13 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' and the actual values in the red line called 'normed_infections'.
The SSR values of Fryslân and Overijssel are the best out of all the models, this could suggest that this model is the best choice of model but this will be left for chapter 6.

| **Variables** | **Fryslân** | **Overijssel** |
|---|---|---|
| $\beta$ | 8.250 | 3.048 |
| $\gamma$ | -0.3283 | 0.3510 |
| $\theta_1$ (AR part) | 0.9921 | 0.9832 |
| $\alpha_1$ (MA part) | -0.6129 | -0.5287 |
| $c$ | 10.83 | 15.51 |
| AIC | 1890 | 2407 |
| BIC | 1913 | 2425 |
| SSR | 5328 | 24523 |

Table 11: dummy variable model - ARMA(1,1) with lagged lockdown dummy and temperature dummy, this table provides the estimated parameters (as in the model equation) and also provides AIC, BIC and SSR values for the best and worst fitting provinces

The figures 12 and 13 show the predicted values of $y_t$ in the blue line called 'xb prediction, one-step' (this results from using the estimated parameters with the value of $y_{t-1}$ to predict the value $y_t$) and the actual values in the red line called 'normed_infections'.

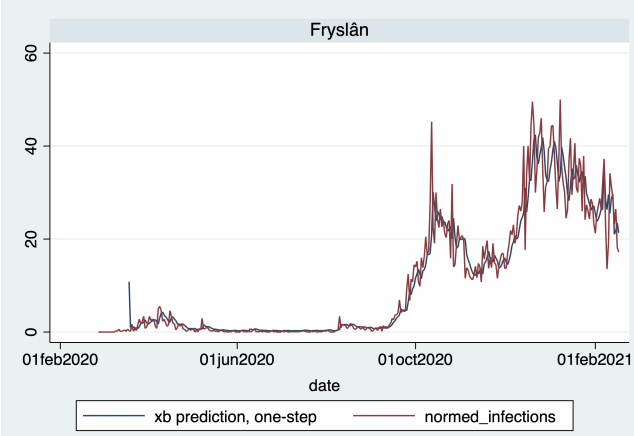


Figure 12: Predicted vs actual normed infections of Fryslân (dummy variable model)

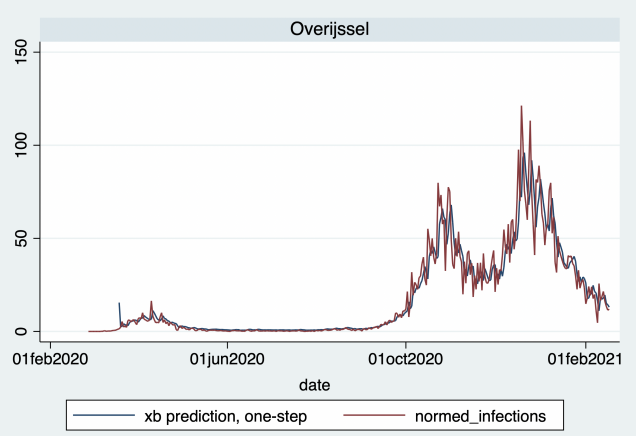


Figure 13: Predicted vs actual normed infections of Overijssel (dummy variable model)

## 6.2 Empirical results

This section aims to discuss the results obtained and to establish whether or not there is a relationship between temperature and COVID-19 infections. From the earlier section on AIC and BIC values, the difference between AIC and BIC values need to be assessed.

The tables below provide the difference in AIC and BIC values between the base model and the non-lagged model. The difference in AIC and BIC values are required in order to assess which model is optimal.

| Province | ΔAIC - non-lagged model |
|---|---|
| Drenthe | +2 |
| Flevoland | +1 |
| Fryslân | +3 |
| Gelderland | +4 |
| Groningen | +3 |
| Limburg | +2 |
| Noord-Brabant | +3 |
| Noord-Holland | +2 |
| Overijssel | +2 |
| Utrecht | +2 |
| Zeeland | +3 |
| Zuid-Holland | +4 |

Table 12: Difference in AIC from base model

| Province | ΔBIC - non-lagged model |
|---|---|
| Drenthe | +9 |
| Flevoland | +8 |
| Fryslân | +11 |
| Gelderland | +12 |
| Groningen | +10 |
| Limburg | +11 |
| Noord-Brabant | +10 |
| Noord-Holland | +11 |
| Overijssel | +9 |
| Utrecht | +10 |
| Zeeland | +11 |
| Zuid-Holland | +11 |

Table 13: Difference in BIC from base model

The AIC difference of the base model and the non-lagged model is 2-4 depending on the province implies that there is no significant evidence to choose the non-lagged model over the base model or vice versa. However, the BIC value has a difference of 8-12 depending on the province, which provides stronger evidence to choose the base model over the non-lagged model.
This is expected due to the fact that a difference in cases regardless of the factors will not be seen on the day. It is expected that lockdown measures or temperature take time to create an effect in the number of cases.

Considering that the measures and temperature take time to produce an effect, the next model that will be studied is the lagged model. There is now a problem of comparing AIC and BIC values since there are 357 observations for the base model for each province, from 27 February 2020 till 21 February 2021. Each province of the lagged model has only 336 observations, from 19 March 2020 till 21 February 2021. To combat this problem, the base model is re-evaulated with only 336 values so the AIC and BIC values can be compared with the lagged model and the dummy variable model. Find tables 14 and 15 below with the new values for AIC and BIC for each province.
These AIC and BIC values seem to differ between each province more than previously. The new AIC difference of the base model and the lagged model is between 1 and 4 which implies there is weak evidence to choose the base model over the lagged model. There is one outlier of South Holland which has a difference of 8 between the base model and the lagged model which suggests there is strong evidence to choose the base model over the lagged model. The new BIC difference of the base model and the lagged model is between 8 and 12, this suggests there is strong evidence to choose the base model over the lagged model. There are two outliers, one being the province of Utrecht and the other being the province of Zeeland. For the province of Utrecht there is a difference of -12 between the base model and the lagged model whihc suggests that there is strong evidence to choose the lagged model over the base model. However, for the province of Zeeland there is a difference of 19 between the base model and the lagged model which suggests that there is very strong evidence to choose the base model over the lagged model.

Now considering the final model, the dummy variable model. The new AIC difference of the base model and the dummy variable model varies greatly between provinces. For provinces of Gelderland, Limburg and North-Holland there is no significant evidence to choose the dummy variable model over the base model and vice versa. For provinces Utrecht and Zeeland there is weak evidence to choose the base model over the dummy variable model. For South-Holland there is significant evidence to choose the base model over the dummy variable model. On the other hand, for provinces Drenthe, Flevoland, Friesland, Groningen, North-Holland and Overijssel there is evidence to choose the dummy variable model over the base model. In the province Friesland there is the strongest evidence to choose the dummy variable model over the base model.
The BIC difference varies greatly between provinces like the AIC difference. Provinces of Gelderland, South-

| Province | AIC of base model (n=336) | ΔAIC - lagged model | ΔAIC - dummy variable model |
|---|---|---|---|
| Drenthe | 2072 | +1 | -6 |
| Flevoland | 2265 | +1 | -6 |
| Fryslân | 1902 | +4 | -12 |
| Gelderland | 2131 | +3 | -2 |
| Groningen | 2088 | +3 | -6 |
| Limburg | 2136 | +4 | -1 |
| Noord-Brabant | 2075 | +1 | -1 |
| Noord-Holland | 1963 | +2 | -5 |
| Overijssel | 2405 | +4 | -3 |
| Utrecht | 2168 | +4 | +3 |
| Zeeland | 2143 | +3 | +3 |
| Zuid-Holland | 2073 | +8 | +6 |

Table 14: Comparison of AIC values for lagged model and dummy variable model - ΔAIC refers to the difference between lagged model and base model, the AIC of the base model being recalculated without the missing readings from the lagged model. Similarly for the dummy variable model.

| Province | BIC of base model (n=336) | ΔBIC - lagged model | ΔBIC - dummy variable model |
|---|---|---|---|
| Drenthe | 2088 | +8 | +1 |
| Flevoland | 2280 | +9 | +2 |
| Fryslân | 1917 | +11 | -4 |
| Gelderland | 2146 | +11 | +10 |
| Groningen | 2104 | +10 | +1 |
| Limburg | 2151 | +12 | -16 |
| Noord-Brabant | 2090 | +9 | +7 |
| Noord-Holland | 1978 | +10 | +3 |
| Overijssel | 2421 | +11 | +4 |
| Utrecht | 2184 | -12 | +10 |
| Zeeland | 2158 | +19 | +11 |
| Zuid-Holland | 2093 | + 11 | +9 |

Table 15: Comparison of BIC values for lagged model and dummy variable model - ΔBIC refers to the difference between lagged model and base model, the BIC of the base model being recalculated without the missing readings from the lagged model. Similarly for the dummy variable model.

Holland, North-Brabant, Utrecht and Zeeland have strong evidence to choose the base model over the dummy variable model. Whereas Friesland has weak evidence to choose the dummy variable model over the base model and Limburg has strong evidence to choose dummy variable model over the base model. The other provinces have little evidence to choose one model over the other.

The conclusion depends on whether AIC of BIC differences will be considered better for this research. BIC has a greater penalty for increased number of parameters in the model making it less ideal for the comparison in this case. Considering in the base model there are no independent variables and in subsequent models 2 parameters are added or included. In this case, AIC criterion will be preferred over BIC.

In conclusion, provinces of Drenthe, Flevoland, Friesland, Groningen and North-Holland are best fitted by the dummy variable model. Whereas the other provinces: Gelderland, Limburg, North-Brabant, Overijssel, Utrecht, Zeeland and South-Holland are best fitted by the base model.

# 7 Conclusion

Returning to the research question "*Is there a relationship between temperature and the number of cases of COVID-19 in the Netherlands?*" and furthermore the correctness of the testable hypothesis.

The results do not suggest that there is a relationship that fits all provinces. For the provinces that are best fitted by model 4, the results suggest relationship between temperature that is below a certain cut-off temperature and also lockdown measures. Whereas for the provinces that are best fitted by model 1, the results suggest there is no relationship that can be found between temperature and number of infections.
For the provinces that are best fitted by model 4, they are in line with the hypothesis yet does not completely agree. Temperature below a certain temperature positively influences the number of cases however otherwise temperature has no relationship. For the provinces that are best fitted by model 1, they are not in line with the hypothesis.

In conclusion, this study has not been able to prove that there is a relationship between temperature and number of COVID-19 cases in the Netherlands overall, however there are some provinces that appear to produce a stronger relationship than others.

# 8 Discussion

Naturally, there are points of discussion to this analysis. The generalisation of the lockdown measures and the choice of certain lags will be discussed along with further points of research.

Firstly, the generalisation of lockdown measures, which was modelled as a binary dummy variable. As can be seen in section 4 the lockdown measures dummy variable was created by analysing the measures taken and whether they were being tightened or relaxed. The reason for this wide generalisation of these measures was that temperature was the main interest in this paper but measures could not be overlooked. However, generalising the measures so severely might have caused the difference to be minor in some provinces. An opportunity for further research might be to repeat this method with levels of lockdown measures in order to see if some measures create more difference than others.
Furthermore, in the provinces that best fitted model 4, there cannot be a concrete answer that temperature alone plays a role since the conclusion must be that lockdown measures and temperature influence the number of cases. Here the comparison between a model with only temperature and then a model with only lockdown measures could not be accurately made. Since both the AIC and BIC values penalise against number of parameters, this was a downside to choosing AIC and BIC model selection. A further research opportunity would be to use cross-validation for model selection in a similar way of the paper by S.I. Alzahrani et al. [4]

Another point of discussion is the number of lags taken in model 3. It can be seen in tables 20 and 21 that the AIC values keep decreasing, since the value of n for each province keeps decreasing. This makes it harder to choose the "optimal" number of lags for a certain variable. The solution of this problem was to choose the lag value where the difference between AIC values became the same, so that the gain (reduced AIC value) from choosing one lag over the other became small. Similarly, looking at the graph of the AIC values (figure 7 or 8) and identify points of inflection or turning points, this was difficult for these graphs and the lag chosen comes down to opinion. This is clearly not an ideal solution for the problem, this is where there are shortfalls in choosing AIC or BIC for model selection. There are many different methods for model selection, however with the time constraints of this project, AIC and BIC values were necessary.

Another factor that could be of influence on the results is the increasing test capacity for COVID-19. It was covered heavily in the media that at the beginning of the pandemic there was a very limited test capacity, tests were exclusively for very sick people and staff members of the hospital. Today, everyone who wants a test can get a test. This is of course of great influence on the spread of the virus, since people now know earlier on if they are contagious and can then quarantine, which will limit the spread. To solve this problem one would need to include test capacity as a variable and perhaps number of tests carried out per day, however this data is not freely available for the whole observation period of this paper.

Concluding the discussion is that there are various factors that have not be taken into account yet in this study, but that could all be a valuable addition when further investigating this topic.

# References


[1] 2021. URL: https://www.ecdc.europa.eu/en/covid-19/timeline-ecdc-response.

[2] Ratnadip Adhikari and Ramesh K Agrawal. "An introductory study on time series modeling and forecasting". In: *arXiv preprint arXiv:1302.6613* (2013).

[3] Hirotugu Akaike. "A Bayesian analysis of the minimum AIC procedure". In: *Selected Papers of Hirotugu Akaike*. Springer, 1998, pp. 275–280.

[4] Saleh I Alzahrani, Ibrahim A Aljamaan, and Ebrahim A Al-Fakih. "Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions". In: *Journal of infection and public health* 13.7 (2020), pp. 914–919.

[5] J. Breitung. "The local power of some unit root tests for panel data." In: *In Advances in Econometrics* 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels, ed. B. H. Baltagi, Amsterdam: JAI Press. (2000), pp. 161–178.

[6] J. Breitung and S. Das. "Panel unit root tests under cross-sectional dependence". In: *Statistica Neerlandica* 59 (2005), pp. 414–433.

[7] Kenneth P Burnham and David R Anderson. "Multimodel inference: understanding AIC and BIC in model selection". In: *Sociological methods & research* 33.2 (2004), pp. 261–304.

[8] ByoungSeon Choi. *ARMA model identification*. Springer Science & Business Media, 2012.

[9] I. Choi. "Unit root tests for panel data". In: *Journal of International Money and Finance* 20 (2001), pp. 249–272.

[10] *Coronavirus covid-19 — government.nl*. 2020. URL: https://www.government.nl/topics/coronavirus-covid-19/news.

[11] K. Hadri. "Testing for stationarity in heterogeneous panel data". In: *Econometrics Journal* 3 (2000), pp. 148–161.

[12] J.D. Hamilton and Princeton University Press. *Time Series Analysis*. Time Series Analysis v. 10. Princeton University Press, 1994. ISBN: 9780691042893.

[13] R. D. F. Harris and E. Tzavalis. "Inference for unit roots in dynamic panels where the time dimension is fixed." In: *Journal of Econometrics* 91 (1999), pp. 201–226.

[14] K.S. Im, M. Pesaran, and Y. Shin. "Testing for unit roots in heterogeneous panels". In: *Journal of Econometrics* 115.1 (2003), pp. 53–74. DOI: 10.1016/s0304-4076(03)00092-7.

[15] A. Levin, C. Lin, and C.J. Chu. "Unit root tests in panel data: asymptotic and finite-sample properties". In: *Journal of econometrics* 108.1 (2002), pp. 1–24.

[16] M Hashem Pesaran. "On the interpretation of panel unit root tests". In: *Economics Letters* 116.3 (2012), pp. 545–546.

[17] Adrian E Raftery. "Bayes factors and BIC: Comment on "A critique of the Bayesian information criterion for model selection"". In: *Sociological Methods & Research* 27.3 (1999), pp. 411–427.

[18] Peng Shi et al. "Impact of temperature on the dynamics of the COVID-19 outbreak in China". In: *Science of The Total Environment* 728 (2020), p. 138890. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2020.138890. URL: http://www.sciencedirect.com/science/article/pii/S0048969720324074.

[19] P Whittle. "Hypothesis Testing in Time Series Analysis, 1951". In: *Almquist and Wiksell, Upssala* (1951).

[20] Jingui Xie and Yongjian Zhu. "Association between ambient temperature and COVID-19 infection in 122 cities from China". In: *Science of the Total Environment* 724 (2020), p. 138201.

# Appendix

The following section provides the additional tables that have been referenced in the paper. Following the tables, an overview of all the governmental lockdown measures can be found. The tables can be found as follows:



Table 16: Appendix table of contents

| **Test** | **P-value** | **Asymptotics** |
|---|---|---|
| Levin-Lin-Chu unit-root test | 0.008 | $N/T \to 0$ |
| Levin-Lin-Chu unit-root test* | 0.008 | $N/T \to 0$ |
| Levin-Lin-Chu unit-root test** | 0.008 | $N/T \to 0$ |
| Harris-Tzavalis unit-root test | 0.000 | $N \to$ Infinity, T fixed |
| Breitung unit-root test with 4 lags | 0.0019 | T, $N \to$ Infinitely Sequentially |
| Im-Pesaran-Shin unit-root test | 0.000 | T, $N \to$ Infinitely Sequentially |
| Philips-Perron unit-root Test with 1 lag | 0.000 | $T \to$ Infinity |
| Hadri LM Test*** | 0.000 | T, $N \to$ Infinitely Sequentially |

Table 17: Unit root tests on COVID-19 new infections (per 100,000 people) cases - discussion of stationarity can be found in section 5.1 (*with lags chosen by AIC (1.00), **with lags chosen by BIC (1.00), ***using Bartlett's kernel with 1 lag)

| **p** | **AIC (Fryslân)** | **AIC (Overijssel)** |
|---|---|---|
| 1 | 1998 | 2533 |
| 2 | 2063 | 2570 |
| 3 | 2125 | 2727 |
| 4 | 2146 | 2696 |
| 5 | 2166 | 2689 |

Table 18: Different auto-regressive orders - graph and explanation can be found in section 6.1.2

| q | AIC (Fryslân) | AIC (Overijssel) |
|---|---|---|
| 1 | 1998 | 2533 |
| 2 | 2040 | 2576 |
| 3 | 2060 | 2569 |
| 4 | 2063 | 2576 |
| 5 | 2062 | 2576 |

Table 19: Different moving average orders - graph and explanation can be found in section 6.1.2

| Number of days | AIC (Fryslân) | Difference from previous | AIC (Overijssel) | Difference from previous |
|---|---|---|---|---|
| 0 | 2001 | | 2535 | |
| 7 | 1969 | -32 | 2492 | -43 |
| 8 | 1964 | -5 | 2489 | -3 |
| 9 | 1960 | -4 | 2482 | -7 |
| 10 | 1956 | -4 | 2476 | -6 |
| 11 | 1950 | -6 | 2470 | -6 |
| 12 | 1947 | -3 | 2464 | -6 |
| 13 | 1941 | -6 | 2456 | -8 |
| 14 | 1932 | -9 | 2449 | -7 |
| 15 | 1931 | -1 | 2446 | -3 |
| 16 | 1928 | -3 | 2438 | -8 |
| 17 | 1923 | -5 | 2433 | -5 |
| 18 | 1917 | -6 | 2427 | -6 |
| 19 | 1913 | -4 | 2418 | -9 |
| 20 | 1907 | -6 | 2404 | -14 |
| 21 | 1895 | -12 | 2408 | +4 |

Table 20: Number of days lags for temperature variable - Model 3 (more information can be found in section 6.1.4)

| Number of weeks | AIC (Fryslân) | Difference from previous | AIC (Overijssel) | Difference from previous |
|---|---|---|---|---|
| 0 | 1956 | | 2476 | |
| 1 | 1955 | -1 | 2476 | 0 |
| 2 | 1937 | -18 | 2452 | -24 |
| 3 | 1906 | -31 | 2409 | -43 |
| 4 | 1870 | -36 | 2366 | -43 |
| 5 | 1838 | -32 | 2322 | -44 |

Table 21: Number of weeks lags for dummy lockdown variable - Model 3 (more information can be found in section 6.1.4)

| Degrees | AIC (Fryslân) | Difference from previous | AIC (Overijssel) | Difference from previous |
|---|---|---|---|---|
| 4 | 1905 | 0 | 2408 | -1 |
| 3 | 1904 | -1 | 2402 | -6 |
| 2 | 1904 | 0 | 2407 | +1 |
| 1 | 1905 | +1 | 2409 | +2 |
| 0 | 1905 | 0 | 2409 | 0 |
| -1 | 1905 | 0 | 2409 | 0 |
| -2 | 1904 | -1 | 2409 | 0 |
| -3 | 1890 | -14 | 2409 | 0 |
| -4 | 1901 | +11 | 2409 | 0 |

Table 22: Cutoff in degrees Celsius for dummy temperature variable - Model 4 (more information can be found in section 6.1.5)

| Variables | Drenthe | Flevoland | Fryslân | Gelderland | Groningen | Limburg | N-Brabant | N-Holland | Overijssel | Utrecht | Zeeland | Z-Holland |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | 0.9850 | 0.9912 | 0.9927 | 0.9873 | 0.9908 | 0.9905 | 0.9877 | 0.9897 | 0.9841 | 0.9881 | 0.9853 | 0.9871 |
| $\alpha_1$ | -0.5096 | -0.6222 | -0.6220 | -0.4744 | -0.6635 | -0.5058 | -0.4053 | -0.4089 | -0.5285 | -0.4973 | -0.6083 | -0.3733 |
| $c$ | 10.93 | 12.86 | 10.48 | 13.02 | 10.64 | 14.65 | 15.26 | 15.71 | 14.81 | 14.32 | 10.51 | 15.88 |
| AIC | 2180 | 2384 | 1998 | 2242 | 2197 | 2248 | 2183 | 2064 | 2533 | 2282 | 2255 | 2185 |
| BIC | 2196 | 2400 | 2014 | 2257 | 2213 | 2263 | 2199 | 2079 | 2549 | 2298 | 2270 | 2201 |

Table 23: ARMA(1,1) without independent variables - base model see section 6.1.1

| Variables | Drenthe | Flevoland | Fryslân | Gelderland | Groningen | Limburg | N-Brabant | N-Holland | Overijssel | Utrecht | Zeeland | Z-Holland |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta$ | -0.1466 | 0.2540 | -0.04625 | -0.05651 | -0.09689 | 0.1230 | -0.1131 | -0.09416 | -0.2749 | -0.1694 | -0.08653 | -0.02868 |
| $\gamma$ | -1.693 | 2.1863 | 1.320 | 0.9776 | 1.877 | 0.5070 | 0.9337 | 1.764 | 1.9205 | 0.7276 | 1.202 | -0.1525 |
| $\theta_1$ | 0.9846 | 0.9914 | 0.9921 | 0.9867 | 0.9897 | 0.9908 | 0.9868 | 0.9888 | 0.9813 | 0.9873 | 0.9840 | 0.9871 |
| $\alpha_1$ | -0.5032 | -0.6286 | -0.6249 | -0.4725 | -0.6644 | -0.5074 | -0.4003 | -0.4033 | -0.5093 | -0.4913 | -0.6009 | -0.3723 |
| $c$ | 13.16 | 9.500 | 10.12 | 13.09 | 10.40 | 13.20 | 15.96 | 15.67 | 16.66 | 15.64 | 10.78 | 16.24 |
| AIC | 2182 | 2385 | 2001 | 2246 | 2200 | 2250 | 2186 | 2066 | 2535 | 2284 | 2258 | 2189 |
| BIC | 2205 | 2408 | 2025 | 2269 | 2223 | 2274 | 2209 | 2090 | 2558 | 2308 | 2281 | 2212 |

Table 24: ARMA(1,1) with independent variables - non-lagged model see section 6.1.3

| Variables | Drenthe | Flevoland | Fryslân | Gelderland | Groningen | Limburg | N-Brabant | N-Holland | Overijssel | Utrecht | Zeeland | Z-Holland |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta$ | 0.1645 | 0.09254 | 0.00805 | 0.1134 | -0.08191 | 0.004725 | -0.1155 | 0.03775 | 0.004231 | 0.03247 | 0.1266 | 0.1044 |
| $\gamma$ | -2.893 | 4.632 | -0.3303 | 1.901 | -2.276 | 0.1245 | 4.062 | 2.902 | 0.3710 | -2.445 | -0.5720 | 1.242 |
| $\theta_1$ | 0.9867 | 0.9901 | 0.9924 | 0.9864 | 0.9909 | 0.9899 | 0.9853 | 0.9884 | 0.9833 | 0.9879 | 0.9856 | 0.9863 |
| $\alpha_1$ | -0.5079 | -0.6232 | -0.6219 | -0.4730 | -0.6691 | -0.5062 | -0.4071 | -0.4124 | -0.5292 | -0.4920 | -0.5991 | -0.3729 |
| $c$ | 11.62 | 10.63 | 11.08 | 11.97 | 13.31 | 13.99 | 15.83 | 15.00 | 15.68 | 16.41 | 10.20 | 15.31 |
| AIC | 2073 | 2266 | 1906 | 2134 | 2091 | 2140 | 2076 | 1965 | 2409 | 2172 | 2146 | 2081 |
| BIC | 2096 | 2289 | 1928 | 2157 | 2114 | 2163 | 2099 | 1988 | 2432 | 2172 | 2169 | 2104 |

Table 25: ARMA(1,1) with lagged independent variables - lagged model see section 6.1.4



| Variables | Drenthe | Flevoland | Fryslân | Gelderland | Groningen | Limburg | N-Brabant | N-Holland | Overijssel | Utrecht | Zeeland | Z-Holland |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Temp cut-off | -4 | 3 | -3 | -1 | -3 | 4 | 1 | -4 | 3 | -4 | 3 | -4 |
| $\beta$ | 7.909 | -4.478 | 8.250 | -3.153 | 8.675 | -2.845 | 2.637 | 8.201 | 4.946 | -2.410 | 0.7552 | 4.148 |
| $\gamma$ | -3.120 | 4.501 | -0.3283 | 1.754 | -2.191 | -0.09189 | 4.204 | 2.866 | 0.4053 | -2.503 | -0.6154 | 1.190 |
| $\theta_1$ | 0.9858 | 0.9908 | 0.9921 | 0.9862 | 0.9912 | 0.9906 | 0.9854 | 0.9888 | 0.9823 | 0.9878 | 0.9852 | 0.9865 |
| $\alpha_1$ | -0.5040 | -0.6329 | -0.6129 | -0.4747 | -0.6667 | -0.5108 | -0.4003 | -0.4208 | -0.5176 | -0.4901 | -0.6077 | -0.3819 |
| $c$ | 13.05 | 12.00 | 10.83 | 13.26 | 12.27 | 16.47 | 14.50 | 15.24 | 15.17 | 16.82 | 11.51 | 16.27 |
| AIC | 2066 | 2259 | 1890 | 2133 | 2082 | 2135 | 2074 | 1958 | 2402 | 2171 | 2146 | 2079 |
| BIC | 2089 | 2282 | 1913 | 2156 | 2105 | 2135 | 2097 | 1981 | 2425 | 2194 | 2169 | 2102 |

Table 26: ARMA(1,1) with temperature dummy and lagged lockdown dummy - dummy variable model see section 6.1.5

### Overview of measures in the Netherlands

A list follows that summarizes the timeline of when the government implemented certain measures. [10]

- 27/02/2020 - First patient of COVID-19 reported in the Netherlands
- 12/03/2020 - Premier Rutte announces the first measures for limiting the spread of COVID-19:
  - Stay at home if you have a cold or symptoms
  - Social distancing
  - Gatherings of over 100 people are forbidden
  - Working from home encouraged
  - Higher education institutions to offer online education where possible
- 15/03/2020 - Additional measures introduced
  - Primary and secondary schools to close from 15 March - 6 April
  - All bars, cafes and restaurants to close from 15 March 18:00 - 6 April inclusive
  - Sports clubs, gyms, saunas, sex clubs and coffee shops to close from 15 March 18:00 - 6 April inclusive
  - 1.5m rule specified
- 17/03/2020 - Travel abroad only permitted if essential
- 18/03/2020 - The Netherlands closes its borders to people outside Europe
- 19/03/2020 - No visits to nursing homes
- 23/03/2020 - Additional measures to control the spread:
  - Stay at home as much as possible
  - All gatherings prohibited until June 1
  - Public transport and shops to take measures such as limit the maximum number of people allowed in
  - All other contact-based professions must close until 6 April
  - Casinos to close from 24 March
  - Local authorities given power to close areas such as parks, lakes, etc so that people may not contact each other
- 31/03/2020 - All existing measures to be extended to 28 April inclusive
- 21/04/2020 - Measures extended
  - Primary schools will reopen May 11
  - Secondary schools to reopen June 2
  - From 29 April under 18s may play sports outdoors
  - From 29 April over 70s living alone may be visited by the same one or two people occasionally
  - Gatherings and events to be banned till 1 September
- 19/05/2020 - The Netherlands relaxes some measures
  - People may meet outside provided 1.5m rule obeyed
  - Public buildings may have up to 30 people inside provided 1.5m apart
  - Restaurants, bars and cafes may open provided people make a reservation, maximum 30 guests and 1.5m rule obeyed
  - From 15 June, higher education institutions may reopen for a limited number of exams and support to vulnerable students
  - From 1 June 12.00, cinemas, theatres and concert halls may reopen provided a maximum of 30 guests, reservation required and 1.5m rule respected
  - Must wear a face mask on public transport, fine for no mask worn, only for essential travel
  - From 15 June, a visiting policy will be implemented for nursing homes

- 24/06/2020 - More relaxed measures
  - Maximum number of people in one space indoors is 100, 1.5m rule obeyed
  - Maximum number of people outdoors in 250
  - No maximum number applies if they have their own seat
  - Face masks must be worn on shared transport
  - Children under 12 do not have to obey the 1.5m rule
  - Older children (up to 18) do not have to obey the 1.5m rule with other under 18s
  - From the start of the academic year more activities can take place on site
  - Choirs and singing groups can start to rehearse together again
  - Nightclubs and similar venues to remain closed until September 1
- 06/08/2020 - The Netherlands reinstates local measures to contain second rise in cases
  - Educational institutions should hold activities online as much as possible. Any introduction activities should end at 22.00 at the latest and no alcohol is allowed.
  - Restaurants, bars and cafes should use reservations and get contact information of the customers
  - If GGD tracks an outbreak to a particular establishment, that establishment may have to close for 14 days
  - Quarantine or testing for arrivals from high risk countries
  - Local authorities still have power to implement more measures if necessary
- 18/08/2020 - Maximum of 6 guests at home
- 18/09/2020 - Extra measures for 6 security regions: Amsterdam-Amstelland, Rotterdam-Rijnmond, Haaglanden,Utrecht, Kennemerland, Hollands Midden
  - Catering venues should not welcome new guests after midnight and they must close at 01:00. No music after midnight
  - Groups may not exceed 50, such as parties and outings
  - Gatherings over 50 are subject to a reporting requirement, with some exceptions including funerals and religious activities
  - Extra measures for Amsterdam-Amstelland and Rotterdam-Rijnmond:
    * Public establishments will have special opening times for vulnerable and elderly
    * Parks will close at midnight
- 25/09/2020 - 8 more security regions added:Gelderland South, Zaanstreek-Waterland, Flevoland, Gooi- en Vechtstreek, South Holland South, Brabant South-East, Brabant North, Groningen.
- 28/09/2020 - Additional measure to combat the spread of the virus
  - No more than 3 guests at your home does not include children under 13
  - Catering establishments must accept no new customers after 21.00 and they are to close at 22.00
  - Sports clubhouses are to close
  - Every indoor space must make use of reservations except shops and markets
  - Maximum group size of 40 outside
  - Sports events must be held without spectators
  - The maximum number of people in 1 room is 30 excludes staff
  - Keep travel to a minimum
- 30/09/2020 - Urgent advice on use of face masks
- 13/10/2020 - Partial lockdown introduced nationwide
  - No more that 3 guests at your home all day
  - Maximum number of people for an indoor venue is 30
  - Indoors and outdoors a group must not exceed 4 people from different households

  - All catering establishments must close
  - Retail stores must close at 20.00 at the latest, no koopavond
  - No alcohol to be sold from 20.00 and 07.00
  - Between 20.00 and 07.00 you may not drink alcohol or have it on your person
  - Coffee shops may only provide a takeaway service and must close by 20.00
  - Team sports is allowed but no more than 4 people and keep distance of 1.5m
  - Showers and changing rooms at gyms, sporting venues to be closed
  - Secondary and higher education face masks are required to be worn outside of class
- 03/11/2020 - Tightening of partial lockdown
  - No more than 2 guests at your home all day
  - Indoors and outdoors a group must not exceed 2 people from different households
  - Stay at home as much as possible
  - All public venues to be closed
  - Everyone over 12 must wear a face mask in public indoor spaces and on public transport
  - Indoor and outdoor events are banned
  - Sports only for maximum 2 people
  - Under 18s may take part in group sports
  - Group lessons and spectators are not allowed
- 19/11/2020 - Partial lockdown to continue with some easing of some measures
  - No more that 3 guests at your home all day
  - Indoors and outdoors a group must not exceed 4 people from different households
  - Maximum number of people for an indoor venue is 30
  - Catering establishments to remain closed, takeaway allowed
  - Retail stores must close at 20.00 at the latest, no shopping evenings
  - No alcohol to be sold from 20.00 and 07.00
  - Between 20.00 and 07.00 you may not drink alcohol or have it on your person
  - Coffee shops may only provide a takeaway service and must close by 20.00
  - Team sports is allowed but no more than 4 people and keep distance of 1.5m
  - Showers and changing rooms at gyms, sporting venues to be closed
  - Secondary and higher education face masks are required to be worn outside of class
  - Public venues may reopen but with reservation of time slots to avoid constant flow of people
- 15/12/2020 - Lockdown to minimise contact between people
  - No more than 2 guests at your home all day
  - Allowed to go outdoors with a maximum of 2 other people
  - Non-essential shops closed
  - Contact professions also closed (e.g. hairdressers, nail salons)
  - Advised to work from home
- 20/01/2021 Lockdown tightened due to new variants
  - No more than 1 guest at your home all day
  - Allowed to go outdoors with a maximum of 1 person
  - Non-essential shops remain closed
  - Contact professions remain closed
  - Curfew in place from 20.30 - 04.30