AALBORG UNIVERSITY
DENMARK

AALBORG UNIVERSITY

Department of Electronic Systems

MASTER THESIS (30 ECTS)

Covid-19 Modelling, estimation and prediction

By Borja Barrios

June 2021

AALBORG UNIVERSITY
DENMARK

Department of Electronic Systems

Elektronik & IT

Fredrik Bajers Vej 7B

9220 Aalborg

https://www.es.aau.dk/

Title:

Covid-19 Modelling, estimation and prediction

Thesis period:

Spring Semester 2021

Thesis group:

1031a

Participants:

Borja Barrios

Supervisor:

Torben Knudsen

Page Number: 72

Date of completion: 3rd of June 2021

Abstract:

The main focus of this project is modelling the behaviour of the Covid-19 disease in order to carry out estimations and predictions. A deterministic model has been created for modelling purposes. An estimation algorithm as the Extended Kalman Filter has been used in order to cope with the non-linearities of the model, estimating its states based on measurement´s data extracted from the Danish Health Authorities. A long-term and a short-term estimation have been carried out in order to prove the adaptation of the model to different time frames. An estimation of the behaviour of the Covid-19 disease during the pandemic have been made for each of the Danish regions. A 40-days prediction for the hospitalized state in the region of Hovedstaden has been carried out in order to show the behaviour of the model when no measurement´s data is added after the EKF prediction step. The results shown in this report have proven that the model developed in this thesis shows a good estimation of how the Covid-19 disease performs in the Danish society, although certain aspects of the modelling rely on assumptions that can be subject to further investigations.

Contents

Index of figures

Index of tables

Index of equations

1. Introduction

In December 2019, an epidemiological outbreak was detected in a crowded food market of the Chinese city of Wuhan. A couple of days later, some cases of pneumonia with an unknown origin were detected in some workers of that place.

After some investigations, the Chinese Government arrived at the conclusion that the already mentioned symptoms were caused by the Sars-Cov-2, a type of Coronavirus that provokes the Covid-19 disease.

With a symptomatology that mostly affects the human respiratory system, going from dry cough to severe pneumonias that can lead into death, the Covid-19 disease was declared as a public health emergency of international concern by the World Health Organization the 30th of January 2020.

Due to its fast spread, the Sars-Cov-2 started to settle in countries from all over the world, being declared as a pandemic the 11th of March of the same year, causing the biggest world-wide pandemic since the 1918 Influenza pandemic. In total, it has infected more than 165 million of people around the globe causing more than 3,5 million of deaths. **(1)**

The Covid-19 pandemic is not just a simple disease, it has completely changed the way of living for the human beings. The majority of the Governments started to act in order to stop its spread, establishing measurements that limit human interactions, such as curfews or lockdowns for more than a year.

One of these affected countries is Denmark, which declared its lockdown the 11th of March 2020. **(2)** At the beginning of the pandemic, the Danish country was affected by the disease, but the prompt reaction of the Government allowed the country to almost extinct the virus in a couple of months.

Nevertheless, this Sars-Cov-2 is characterized by being able to mutate very fast, what implies the formation of new different strains of the virus with alternative characteristics. Some of these new strains, such as the British strain, were characterized with a faster spread and more dangerous symptoms than the common Sars-Cov-2.

These new strains of the virus mainly caused the second wave of Covid-19 cases in Denmark, having a more significant impact than the previous one. Concretely, it infected more than ten times the people that were infected in the first wave of the virus.

At present, Denmark has overcome this second wave of the virus, thanks principally to the arrival of the Covid-19 vaccines and the measurements that were stablished by the Government. Despite of this, as the virus is still spreading among its population and after suffering more than 282 thousand of cases, what implies almost the 5 per cent of the population, the Danish authorities have started to work in order to prevent a hypothetical third wave of the virus. **(3)**

This report aims to propose a new model for the Sars-Cov-2 which will involve more groups than the already known SIR and SEIR models. Concretely, the model will be divided into 7 different states that group the entire Danish population, considering its situation against the virus.

As the model would be time dependent, there is a need of using an estimate method that can overcome the noise that will appear after the discretization of the equations from the model.

The proposed estimator for this project would be an Extended Kalman Filter, which not only it will cope with the originated noise, but it will also overcome the non-linearities of the model. This mathematical algorithm will be used to prove the viability of the model in two different pandemic time frames, each of them having a different length, showing the behaviour of the states in terms of time.

Finally, the estimation results will be obtained particularly for each of the region in which Denmark is divided, so that it can clearly be seen how the cases are distributed through the 5 different Danish regions. Last but not least, a forecast will also be done in order to show the behaviour of the model when it predicts the behaviour of the states without using any measurement´s data.

2. Problem statement

The performance of Covid-19 is already an unknown for the population. Its spread and behaviour have overcome the expectations causing a world-wide pandemic with severe consequences. In fact, one of the main reasons why this pandemic has affected in a negative way the society, is the unexpected number of hospitalizations that the virus has caused during this period, what have caught the countries unaware, resulting in the lack of hospitalized resources to face the virus.

Regarding this, the following statement has been proposed to be the aim of this report:

"Develop a consistent model that suits the behaviour of the Covid-19 disease, estimating its states and parameters during the pandemic time frame, focusing on the hospitalized state so it can be considered as a method to forecast and prevent the hospitalizations originated by the disease."

Not all the mathematical models in epidemiology are valid for this specific disease. Its uncertain behaviour needs a complex model that can control how the virus is performing in the society.

Moreover, the new infections are mainly caused by the interaction between humans, so factors as the number of inhabitants or the geographic density will be important in order to determine the impact of the virus in a certain territory.

Consequently, the aim of this thesis is to create an innovative model that can be used to estimate the behaviour of the Covid-19 disease in the five Danish regions, so that it can be considered as a valid model in which forecasts can be performed with the sole aim of securing hospitalized resources for all the patients.

In order to achieve this goal, there will be some steps that needs to be completed:

- Create a model that kindly adapts to the behaviour of the Covid-19 disease
- Extract and filter the official data that would be use for the estimation of the states
- Study and give a value for every parameter of the model
- Estimate the model with an estimator that can overcome the design difficulties
- Examine the performance of the model regarding different pandemic´s time frames
- Extract the results for every Danish region and analyze them in detail
- Forecast a short-period of time and compare it with the real-life events

3. Modelling

In this section, the created model that will be used for this project will be described. Furthermore, before starting showing the entire model, an explanation on how this model was chosen will be given too.

3.1 Election of the SEIRHD model

When an infectious disease wants to be studied, SIR or SEIR models are commonly chosen.

SIR Model was developed at the beginning of the 20th century by Ronald Ross and William Hamer. This model involves three non-linear differential equations, each of them belonging to one of the labels in which the population is classified (Susceptible, Infectious, Recovered). **(4)**

The main advantage of this model is that a good approximation of how a disease is going to act can easily be obtained, with the use of a few states.

In the other hand, the main disadvantage of this model is that the main groups in which the population is divided are not very specific, and there are a lot of phases that are not considered in it, such as hospitalized people (which in SIR model are included in the infected group) or the dead state. This issue may be negligible for other diseases because they can be adapted to a SIR type model, but the behaviour of COVID-19 is far from adapting to such a simple model.

The same applies to the SEIR type model which includes the same as the SIR type, but it differentiates between exposed people (people who is already infected but is incubating the virus) and the infected people.

But mainly, the election has been made due to the aims that this project is trying to reach. A suitable model for the COVID-19 disease is mandatory in order to fulfil those goals, ending in the conclusion that the population needs to be divided accordingly to the phase of the virus that they are living. So that, the model could show the closest estimations to the reality that was lived in Denmark during the pandemic.

For that reason, the following type of model has been chosen as a reference for the final model that would be used in this report.

3.2 SEIHRD Model

The SEIHRD model was built as an expansion of the SIR and SEIR types, creating a model which suits in an acceptable way the COVID-19 characteristics. It considers six different states for the population, regarding its situation against COVID-19. **(5)**

3.2.1 Description of the population groups

As the acronym from SEIHRD shows (one letter for each group), the population will be divided in six different groups, which are the following:

Susceptible (S)	People who have not been infected with the virus yet, but that is susceptible to be infected, as they coexist with people from all of the other groups.
Exposed (E)	People that have been affected by the virus, but they are still in the incubation period.
Infectious (I)	People that have been infected and have passed the incubation period.
Hospitalized (H)	People that show serious problems coping with the virus and need to have urgent medical attention.
Recovered (R)	People that have successfully overcome the virus and are not infected anymore.
Dead (D)	People who have not survived to the virus.

TABLE 1

3.2.2 Modelling assumptions

Regarding the variety of parameters that must be taken into account in this project, considering that the ODE model is compound of 6 different equations, some assumptions have been made in order to simplify the modelling:

- As a result of the newborns immunity for the virus, the newborn people, that are born during the study time, will not be considered.

- People that dies daily in Denmark, for other reasons that are not associated to COVID-19, will not be considered.

- It is assumed that there cannot be people that goes from the infected group to the dead group. They must go through the hospitalized group first.

- The exposed group is formed by people who is already infected by the virus but is still in the incubation period. Then, it is assumed that susceptible people need to go through this group before going into the infected group.

- 14 days will be the approximate incubation period of the virus which is consider as a fixed parameter in the model.

-The total population (P) of Denmark would be the sum of the people that belongs to each of the groups, as it is expressed in (Eq. 1):

$$S + E + I + H + R + D = P \quad \text{(Eq. 1)}$$

3.2.3 Variables and equations of the model

For each of the population groups that have been shown before in 3.2.1, there is an ODE equation that expresses the inputs and outputs of people that belongs to that certain group.

Now, the equations that characterize each of the population groups will be shown:

$$\frac{dS}{dt}(t) = -\beta(t)\, I(t) \frac{S(t)}{P} \qquad \textbf{(Eq. 2)}$$

, being S(t) the amount of people that is in the susceptible group at time t;

$I(t)$ is the amount of people that belongs to the infected group at time t;

P is the total population of the country that is being studied, in this case Denmark;

β is the transmission rate from a susceptible population to an exposed population.

$$\frac{dE}{dt}(t) = \beta(t)\, I(t) \frac{S(t)}{P} - \alpha E(t) \qquad \textbf{(Eq. 3)}$$

, being $\alpha(t)$ the transmission rate of infected people from the exposed population;

$E(t)$ is the amount of people that belongs to the exposed group at time t.

$$\frac{dI}{dt}(t) = \alpha E(t) - \tau I(t) - \theta_I I(t) \qquad \textbf{(Eq. 4)}$$

, being θ_I the transmission rate of recovery from the infected population;

$$\frac{dH}{dt}(t) = \tau(t) I(t) - \theta_H H(t) - \gamma H(t) \qquad \textbf{(Eq. 5)}$$

, being $H(t)$ the amount of people that belongs to the hospitalized group at time t;

τ the transmission rate from the infected group to the hospitalized group;

θ_H the rate of recovery from the hospitalized population;

γ the mortality rate of the virus.

$$\frac{dR}{dt}(t) = \theta_H H(t) + \theta_I I(t) \qquad \textbf{(Eq. 6)}$$

, being R(t) is the amount of people that belongs to the recovery group at time t;

$$\frac{dD}{dt}(t) = \gamma H(t) \qquad \textbf{(Eq. 7)}$$

, being $D(t)$ the amount of people that belongs to the dead group at time t.

3.2.4 Improvements that can be applied to the SEIHRD type model

As the COVID-19 immunity behavior is already unknown, there are some factors that can be taken into account in the modelling in order to make it more realistic:

- **Recovered people´s immunity can be lost:** At the beginning of the pandemic, the thought was optimistic as the majority of the diseases perform in that way; every person that overcomes COVID-19 will be immune for a long time. However, time has proven that this is not true. **(6)**
  The mutations of the virus, combined with the loss of antibodies from the recovered people implies that the recovered population can have a possibility of being considered as susceptible people again after a certain time. This will lead in a positive term in (Eq. 2) of the modelling, which will restart the process for that people who have already overcame the virus.

- **Vaccinates:** After a year of pandemic, Denmark has started the vaccination period to fight the COVID-19 spread. This vaccine would transfer this susceptible people that receives the vaccine to a non-contemplated new state which would group the vaccinated people.
  This new group would suppose a new negative term in (Eq. 2) of the modelling, and it will increase with an approximate mean number of vaccines that are supplied daily in Denmark. As the performance of the vaccines is still an uncertainty, this model will also consider that, as it happens with the immunity that can be acquired after overcoming the virus, the vaccination´s immunity can also be lost, so that the people that enters the vaccinated state will have a probability of returning to the susceptible group.

3.2.5 Final modelling after introducing the improvements

As it was shown in Improvements that can be applied to the , a couple of improvements can be added to the modelling, in order to make it closer to the real behaviour of the Covid-19 disease.

To achieve that, (Eq. 2) would have this new form:

$$\frac{dS}{dt}(t) = \varepsilon_r R(t) + \varepsilon_v V(t) - M - \beta\, I(t)\frac{S(t)}{P} \qquad \textbf{(Eq. 8)}$$

, being $\varepsilon_r\ and\ \varepsilon_v$ the transmission rate from the recovery group and from the vaccinated group to the susceptible group, respectively;

V(t) is the amount of people that belongs to the vaccinated group at time t;

M is the approximated number of vaccines that are injected daily.

(Eq. 8) shows three new terms that, respectively, involve:

- people from the recovery group who lost the immunity
- people from the vaccinated group who lost the immunity
- people that are vaccinated and goes directly to the vaccinated group.

Also, as it is considered that there is a probability of the immunity being lost, the recovery group equation will also change, ending in:

$$\frac{dR}{dt}(t) = \theta_H H(t) + \theta_I I(t) - \varepsilon_r R(t) \quad \textbf{(Eq. 9)}$$

Finally, as it can be seen in (Eq. 8) there is a new group which is the vaccinated group, which will be defined by the following equation:

$$\frac{dV}{dt}(t) = M - \varepsilon_v V(t) \quad \textbf{(Eq. 10)}$$

The rest of the modelling remains the same, as the changes just applies to the susceptible, recovery and vaccinated groups.

3.2.6 Outputs of the model

There are some parameters that can be calculated after making the estimation of the mentioned model, which will be the following:

- $c_t(t)$: total amount of Covid-19 active cases at time t, which is given by:

$$c_t(t) = E(t) + I(t) + H(t) \quad \text{(Eq. 11)}$$

- $d_t(t)$: cumulative number of deaths at time t, which is given by:

$$d_t(t) = D(t) \quad \textbf{(Eq. 12)}$$

- $h_t(t)$: total amount of Covid-19 hospitalized people at time t, which is given by:

$$h_t(t) = H(t) \quad \textbf{(Eq. 13)}$$

- $maxc$: maximum number of Covid-19 cases that are reached in an interval of time{t_0,T}, which is given by:

$$maxc = max\ c_t(\mathrm{t0}, \mathrm{T}) \quad \textbf{(Eq. 14)}$$

- $maxh$: maximum number of hospitalized people that are reached in an interval of time{t_0,T}, which is given by:

$$maxh = max\ h_t(\mathrm{t0}, \mathrm{T}) \quad \textbf{(Eq. 15)}$$

- $R_t(t)$ or Effective reproduction number: it is the parameter that measures how many people can a single infected person infect during its infection period.
  Depending on its value, the disease cases will rapidly increase if R_t is higher than 1, and in the other hand, the disease will begin to stop its spread once the result of R_t is lower than 1. **(7)** The Effective reproduction number is given by the following equation:

$$R_t(t) = \frac{S_t}{P}\left(\frac{\beta(t) * T_i}{\theta + \gamma}\right) \quad \textbf{(Eq. 16)}$$

, where T_i is the average time that a person is infected by Covid-19.

3.3 Final model

After adding the improvements that were commented in Improvements that can be applied to the SEIHRD type model, the final model that is going to be used in this project is shown in the following diagram:

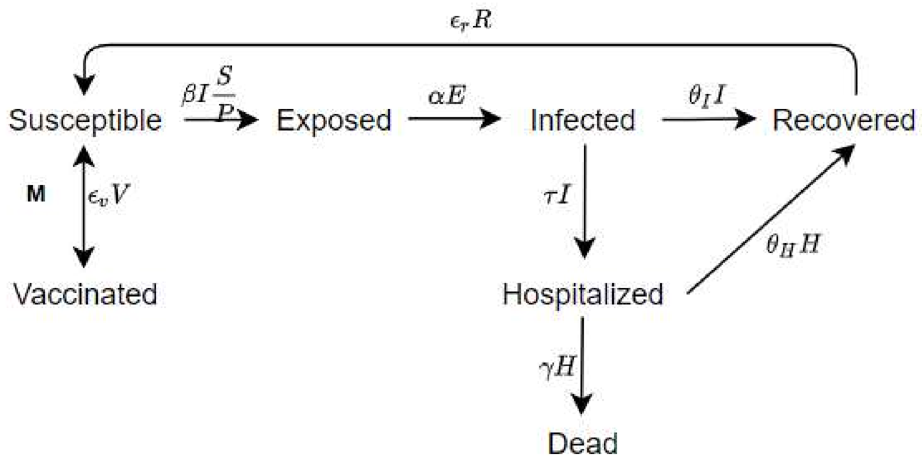

FIG. 1

4. State estimation

In this chapter, it is going to be introduced step by step the procedure that has been done to estimate the states of the model.

As it was seen in Modelling, all the states from the model depends on time. Due to this fact, and considering that the model is dynamic, a discretization of the model will be done before applying the model to the estimation algorithm.

4.1 Discretization of the model

As the groups that divides the population are continuously changing due to the spread of COVID-19, it has been taken into account that the states that represents those population groups varies with time.

The rest of the parameters that are involved in the model are considered as fixed parameters, except from the parameter that measures the rate of people that goes from the susceptible group to the exposed group ($\beta(t)$). This parameter measures the rate of infections that happens in a certain time, so depending on the behaviour of the infected and exposed states, it will increase or decrease in consonance. Consequently, it is necessary to consider that this parameter varies on time, so that a more realistic estimation can be achieved.

For that reason, this parameter ($\beta(t)$) is going to be consider as a state when the estimation is being done. Converting a parameter to a new state variable is a common method when estimating a parameter using EKF. **(8)**

To achieve the desired discretized model, one of the Runge-Kutta methods will be used. Concretely, the Euler method which is a common and simple way to solve ODE´s equation with a given initial value.

After applying the Euler´s Method **(9)** to the model that is shown in Final model, the resultant system of equations will be the following:

$$S(k+1) = S(k) + (\varepsilon_r R(k) + \varepsilon_v V(k) - M - \beta(k)\, I(k) \frac{S(k)}{P})\Delta t + \omega_1 \qquad \textbf{(Eq. 17)}$$

$$E(k+1) = E(k) + (\beta(k)\, I(k) \frac{S(k)}{P} - \alpha E(k))\Delta t + \omega_2 \qquad \textbf{(Eq. 18)}$$

$$I(k+1) = I(k) + (\alpha E(k) - \tau I(k) - \theta_I I(k))\, \Delta t + \omega_3 \qquad \textbf{(Eq. 19)}$$

$$H(k+1) = H(k) + (\tau I(k) - \theta_H H(k) - \gamma H(k))\Delta t + \omega_4 \qquad \textbf{(Eq. 20)}$$

$$R(k+1) = R(k) + (\theta_H H(k) + \theta_I I(k) - \varepsilon_r R(k))\Delta t + \omega_5 \qquad \textbf{(Eq. 21)}$$

$$D(k+1) = D(k) + \gamma H(k)\Delta t + \omega_6 \qquad \textbf{(Eq. 22)}$$

$$V(k+1) = V(k) + ((M - \varepsilon_v V(k))\Delta t + \omega_7 \qquad \textbf{(Eq. 23)}$$

As it has been mentioned, apart from the equations of the model, the parameter ($\beta(t)$) is going to be treated as a state of the system, implying that it will have its own equation which is the following:

$$\beta(k+1) = \beta(k) + \omega_8 \qquad \textbf{(Eq. 24)}$$

In all the discretized equations, it has been added a new parameter (ω) which represents the noise. This noise remains as an uncertainty in the model, and it is being assumed to be white, Gaussian and uncorrelated.

4.2 Extended Kalman Filter

As it has been shown in Discretization of the model, after discretizing the model, some noise appears on it. This noise needs to be removed or, at least, reduced in order to achieve a correct estimation for the states of the system. For this reason, a Kalman Filter is being used, as it can estimate the states of the system while it copes with the mentioned noise.

However, as the model is non-linear, a common Kalman Filter cannot be used to solve this problem. Consequently, an Extended Kalman Filter is being chosen in this project to overcome these non-linearities. This estimation algorithm will linearize the non-linear functions of the model around a working point, and finally will estimate the desired states of the system.

The discretized system shown in Discretization of the model can be expressed in the following way:

$$x(k+1) = f\big(x(k)\big) + \omega(k) \qquad \textbf{(Eq. 25)}$$

, being $x(k+1)$ the augmented state vector of the system, which equation is shown here:

$$x(k+1) = (\,S(k+1)\ E(k+1)\ \mathrm{I(k}\ +1)\ H(k+1)\ R(k+1)\ D(k+1)\ V(k+1)\ \beta(k+1))^T \qquad \textbf{(Eq. 26)}$$

, and being $f\big(x(k)\big)$ the non-linear term of the system which matches with the right-hand side part of the discretized equations without the term of the noise, which is represented by $\omega(k)$.

It is worth saying that the working point that Kalman Filter is using, matches with the estimation that is being done, ending in the conclusion that the system is being linearized around the state estimation vector ($\hat{x}(k)$). **(10)**

Then, if the first-order Taylor series expansion is applied to f at $\hat{x}(k)$, the following equation is obtained:

$$f(x(k)) = f\big(\hat{x}(k)\big) + J_f(\hat{x}(k)(x(k) - \hat{x}(k)) \quad \textbf{(Eq. 27)}$$

, being J_f the Jacobian matrix of f.

4.2.1 Jacobian Matrix

The Jacobian Matrix of the system is the matrix that collects each of the partial derivatives of the equations of the model derived around each of the eight states of the discretized system.

After adding the β parameter to the states of the system, the resultant model has a total of 8 equations for 8 different states. So that, the Jacobian Matrix of the system will be a 8x8 size matrix defined by:

$$J = \begin{pmatrix} J_{11}(\hat{x}(k)) & J_{12}(\hat{x}(k)) & J_{13}(\hat{x}(k)) & J_{14}(\hat{x}(k)) & J_{15}(\hat{x}(k)) & J_{16}(\hat{x}(k)) & J_{17}(\hat{x}(k)) & J_{18}(\hat{x}(k)) \\ J_{21}(\hat{x}(k)) & J_{22}(\hat{x}(k)) & J_{23}(\hat{x}(k)) & J_{24}(\hat{x}(k)) & J_{25}(\hat{x}(k)) & J_{26}(\hat{x}(k)) & J_{27}(\hat{x}(k)) & J_{28}(\hat{x}(k)) \\ J_{31}(\hat{x}(k)) & J_{32}(\hat{x}(k)) & J_{33}(\hat{x}(k)) & J_{34}(\hat{x}(k)) & J_{35}(\hat{x}(k)) & J_{36}(\hat{x}(k)) & J_{37}(\hat{x}(k)) & J_{38}(\hat{x}(k)) \\ J_{41}(\hat{x}(k)) & J_{42}(\hat{x}(k)) & J_{43}(\hat{x}(k)) & J_{44}(\hat{x}(k)) & J_{45}(\hat{x}(k)) & J_{46}(\hat{x}(k)) & J_{47}(\hat{x}(k)) & J_{48}(\hat{x}(k)) \\ J_{51}(\hat{x}(k)) & J_{52}(\hat{x}(k)) & J_{53}(\hat{x}(k)) & J_{54}(\hat{x}(k)) & J_{55}(\hat{x}(k)) & J_{56}(\hat{x}(k)) & J_{57}(\hat{x}(k)) & J_{58}(\hat{x}(k)) \\ J_{61}(\hat{x}(k)) & J_{62}(\hat{x}(k)) & J_{63}(\hat{x}(k)) & J_{64}(\hat{x}(k)) & J_{65}(\hat{x}(k)) & J_{66}(\hat{x}(k)) & J_{67}(\hat{x}(k)) & J_{68}(\hat{x}(k)) \\ J_{71}(\hat{x}(k)) & J_{72}(\hat{x}(k)) & J_{73}(\hat{x}(k)) & J_{74}(\hat{x}(k)) & J_{75}(\hat{x}(k)) & J_{76}(\hat{x}(k)) & J_{77}(\hat{x}(k)) & J_{78}(\hat{x}(k)) \\ J_{81}(\hat{x}(k)) & J_{82}(\hat{x}(k)) & J_{83}(\hat{x}(k)) & J_{84}(\hat{x}(k)) & J_{85}(\hat{x}(k)) & J_{86}(\hat{x}(k)) & J_{87}(\hat{x}(k)) & J_{88}(\hat{x}(k)) \end{pmatrix} \quad \textbf{(Eq. 28)}$$

As the equations are compound by the linear combination of a couple of the states, not all of them, most of the terms that are shown in the matrix are going to be 0.

Therefore, the Jacobian Matrix will have the following form:

$$J = \begin{pmatrix} J_{11}(\hat{x}(k)) & 0 & J_{13}(\hat{x}(k)) & 0 & J_{15}(\hat{x}(k)) & 0 & J_{17}(\hat{x}(k)) & J_{18}(\hat{x}(k)) \\ J_{21}(\hat{x}(k)) & J_{22}(\hat{x}(k)) & J_{23}(\hat{x}(k)) & 0 & 0 & 0 & 0 & J_{28}(\hat{x}(k)) \\ 0 & J_{32}(\hat{x}(k)) & J_{33}(\hat{x}(k)) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & J_{43}(\hat{x}(k)) & J_{44}(\hat{x}(k)) & 0 & 0 & 0 & 0 \\ 0 & 0 & J_{53}(\hat{x}(k)) & J_{54}(\hat{x}(k)) & J_{55}(\hat{x}(k)) & 0 & 0 & 0 \\ 0 & 0 & 0 & J_{64}(\hat{x}(k)) & 0 & J_{66}(\hat{x}(k)) & 0 & 0 \\ J_{71}(\hat{x}(k)) & 0 & 0 & 0 & 0 & 0 & J_{77}(\hat{x}(k)) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & J_{88}(\hat{x}(k)) \end{pmatrix} \quad \textbf{(Eq. 29)}$$

, being the non-zero terms of the Jacobian Matrix described as:

$$J_{11}\big(\hat{x}(k)\big) = 1 - \frac{\beta(k)\, I(k)\Delta t}{P}$$

$$J_{13}\big(\hat{x}(k)\big) = -\frac{\beta(k)\, S(k)\Delta t}{P}$$

$$J_{15}\big(\hat{x}(k)\big) = \varepsilon_r \Delta t$$

$$J_{17}\big(\hat{x}(k)\big) = \varepsilon_v \Delta t$$

$$J_{18}\big(\hat{x}(k)\big) = -\frac{I(k)\ S(k)\Delta t}{P}$$

$$J_{21}\big(\hat{x}(k)\big) = \frac{I(k)\ \beta(k)\Delta t}{P}$$

$$J_{22}\big(\hat{x}(k)\big) = 1 - \alpha\Delta t$$

$$J_{23}\big(\hat{x}(k)\big) = \frac{\beta(k)\ S(k)\Delta t}{P}$$

$$J_{28}\big(\hat{x}(k)\big) = \frac{I(k)\ S(k)\Delta t}{P}$$

$$J_{32}\big(\hat{x}(k)\big) = \alpha\Delta t$$

$$J_{33}\big(\hat{x}(k)\big) = 1 - (\tau + \theta_I)\ \Delta t$$

$$J_{43}\big(\hat{x}(k)\big) = \tau\Delta t$$

$$J_{44}\big(\hat{x}(k)\big) = 1 - (\theta_H + \gamma)\ \Delta t$$

$$J_{53}\big(\hat{x}(k)\big) = \theta_I\Delta t$$

$$J_{54}\big(\hat{x}(k)\big) = \theta_H\Delta t$$

$$J_{55}\big(\hat{x}(k)\big) = 1 - \varepsilon_r\Delta t$$

$$J_{64}\big(\hat{x}(k)\big) = \gamma\Delta t$$

$$J_{66}\big(\hat{x}(k)\big) = 1$$

$$J_{71}\big(\hat{x}(k)\big) = 0$$

$$J_{77}\big(\hat{x}(k)\big) = 1 - \varepsilon_v\Delta t$$

$$J_{88}\big(\hat{x}(k)\big) = 1$$

4.2.2 Covariance matrices

After the discretization made in Discretization of the model, it can be seen that in every equation of the model appears a term for the noise (ω). This state noise, as well as the noise that would be added for the measurements, is considered to be white noise with zero mean and a variance defined by the covariance matrices R and Q. This relation is shown in the following equations:

$$\omega_S(k) \sim N(0, Q) \qquad \textbf{(Eq. 30)}$$

$$\omega_M(k) \sim N(0, R) \qquad \textbf{(Eq. 31)}$$

, where ω_S would be the state noise, while ω_M would be the measurement noise.

The covariance matrix Q, that would be use for the state noise, is considered as a tuning matrix, what implies that its values has been adjusted until the output of the EKF was sufficiently correct to be extracted.

This covariance matrix is shown in (Eq. 32):

$$Q = \begin{pmatrix} 100 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 500 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1e^{-5} \end{pmatrix} \quad \textbf{(Eq. 32)}$$

In the other hand, the covariance matrix that has been used to model the noise from the measurements, has been modeled according to how reliable are the states of the model. This modelling will be later discussed in Election of the white noise´s variance for the measurement´s noise. The resulting covariance matrix can be seen in (Eq. 33):

$$R = \begin{pmatrix} 1000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \quad \textbf{(Eq. 33)}$$

4.2.3 Prediction step

The prediction step of the Extended Kalman Filter is mainly based on the given model, in the state calculated in the previous step, and in the previous covariance. The predicted state is expressed in (Eq. 34):

$$\hat{x}(k+1|k) = f(\hat{x}(k|k)) \quad \textbf{(Eq. 34)}$$

Similarly, the predicted covariance is based on the given covariance, the Jacobian matrix, and the state noise. The predicted covariance is expressed in (Eq. 35):

$$\hat{\Sigma}(k+1|k) = J\big(\hat{x}(k|k)\big)\hat{\Sigma}(k|k)J\big(\hat{x}(k|k)\big)^T + Q(k) \quad \textbf{(Eq. 35)}$$

After predicting the state and the covariance of the model, it is time to introduce the measurements into the estimate. This process is being done in the updating step. **(11)**

4.2.4 Updating step

The updating step of the Extended Kalman Filter is characterized by the Kalman gain. This gain will determine how much are involved the measurements in the state estimates. This Kalman gain is calculated as shown in (Eq. 36) with the predicted covariance, the measurement noise, and the measurement matrix:

$$K_g = \hat{\Sigma}(k+1|k)L(k)^T\left(L(k)\hat{\Sigma}(k+1|k)L(k)^T + R(k)\right)^{-1} \quad \textbf{(Eq. 36)}$$

, being L(k), the measurement matrix shown in (Eq. 37):

$$L(k) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \textbf{(Eq. 37)}$$

After calculating the Kalman gain, the updating process of the state and the covariance can be determined. Firstly, the EKF updating state is calculated by implementing the measurement data to the predicted state that was calculated in the previous step. (Eq. 38) shows how the updating of the state is done:

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K_g(\tilde{y}(k+1)) \quad \textbf{(Eq. 38)}$$

, where $\tilde{y}(k+1)$ is an updated output vector.

This output vector would allow to incorporate into the model the measurement data and its update step is shown in (Eq. 39):

$$\tilde{y}(k+1) = y(k+1) - L(k)\hat{x}(k+1|k) \quad \textbf{(Eq. 39)}$$

Finally, after the output vector and the state are updated, it´s time for the covariance to follow the same path. This updating process for the covariance would include the Kalman gain, the measurement matrix, and the predicted covariance. This covariance would indicate the variances of each of the states in which the model is split, concluding that it will show the precision of the estimate that is being done. **(11)** The covariance´s update is shown in (Eq. 40):

$$\hat{\Sigma}(k+1|k+1) = (I - K_gL(k))\hat{\Sigma}(k+1|k) \quad \textbf{(Eq. 40)}$$

5. Model´s estimation

This chapter of the report aims to introduce the estimation that has been done in Matlab in order to test the behaviour of the model that has been shown in Modelling.

The section will be divided in two main subsections: Parameters will show the parameters for each region and the value of the parameters, while Performance of the estimations will show the performance of the algorithm compared to the data that has been extracted, in a short-term and a long-term estimation.

5.1 Parameters

For the appropriate behaviour of the system, some parameters have been approximated according to the real data that have been obtained in the official webpage from the Danish Health Authorities. **(12)**

As the estimation depends on the initial data of the different regions, there are some parameters that are fixed and equal for all the regions while there are some others than varies depending on the data.

5.1.1 Fixed parameters

The following parameters are the same for the 5 regions in which Denmark is divided:

Parameter	Description	Value
ε_r	Transmission rate from the recovery group to the susceptible group (loss of immunity)	0.0017
ε_v	Transmission rate from the vaccinated group to the susceptible group (loss of immunity)	0.0005
τ	Transmission rate from the infected group to the hospitalized group	0.0042
T_i	Infection time and incubation period (with standard deviation of 1 day)	14
α	Transmission rate from the exposed group to the infected group	$\frac{1}{T_i}$

TABLE 2

ε_r parameter

According to some investigations that have been done during the pandemic, around to a 30% of the people that have been infected by Covid-19 have lost its immunity after 6 months. **(13)**

As this statement can be considered as a progressive fact (some people will lose their immunity in 2 months while other people will lose it in 5 months); the value of ε_r has been set to a constant rate of 0.17% of the people.

ε_v parameter

At the date that this report is being done, there is no evidence that the immunity obtained by the vaccine can be lost, because there has only been 3 months since the people started to get vaccinated.

The immunity that is obtained by a vaccine is prepared to attack the main protein of the virus (protein S in the case of Sars-Cov-2) while if you obtain antibodies after being infected by the Covid-19, these antibodies will protect the individual from different proteins of the virus not necessarily from the main one. **(14)**

As the Covid-19 is a virus that has infected the population with some different strains, attacking its main protein (the one that shares every strain) seems to be the way to defeat the pandemic.

So, according to this statement and assuming that the vaccines will show good immunity results in the future, in this report it has been set that the immunity obtained from the vaccines is much more reliable that the one obtained after the infection. So that, the value of ε_v has been set to a rate of 0.05%.

τ parameter

According to a research made in Copenhagen a couple of months ago, the British variants of the Covid-19 has a higher rate of hospitalizations than the rest of the strains. **(15)**

As Denmark, has suffered a second-wave of the virus with much more infected than the previous one, due to this strain and others with similar characteristics, the hospitalized rate has been set to a 0.42% which is the value that matches with the rate of hospitalizations caused by the British strain of the virus.

T_i and α parameters

The incubation period for Sars-Cov-2 can vary from 2 days to a total of 14 days. **(16)**

As there is no data for the exposed group, in this report it has been assumed that the waiting incubation time will be around 14 days. So that, a person needs to be around 14 days on the exposed group before entering on the infected one. In order to have a more realistic view of this parameter, a standard deviation has been added to it, so it can vary a bit from that value.

This value will also be considered as the number of days that the person is infected by the virus (infection time).

The α parameter depends directly on the infection time, so its value will be around $\frac{1}{14} \approx 0.071$.

5.1.2 Variable parameters

The coming parameters are specific for each of the regions, varying depending on the initial values that are obtained:

Parameter	Description	Value
CFR	The ratio between the deaths that a disease has provoked divided by the total amount of finished cases affected by that disease.	$\frac{D(t)}{D(t)+R(t)}$
θ_I	The rate of recovery from the infected group	$(1-CFR)*\frac{1}{T_i}$
θ_H	The rate of recovery from the hospitalized group	$(1-CFR)*\frac{1}{T_i}$
γ	The mortality rate of the virus	$CFR*\frac{1}{T_i}$

TABLE 3

CFR (Case Fatality Rate)

This parameter is not one of the parameters previously described on Modelling, but it directly affects some of them, as it can be seen in Table 3.

It is obtained dividing the number of people that has dead during a certain time, by the number of people that has recovered and dead in that certain time. As the population varies in the different regions of Denmark, the recovered cases and the death cases will also vary accordingly, ending in different values for the CFR. **(17)**

θ_I and θ_H parameters

These parameters correspond to the rates of people that goes from the infected or the hospitalized group to the recovered group, respectively. As there is no real data available for the value of these parameters, they have been modelled as the total cases that are not going to the dead group. That is why the expression above can be simplified as $\theta = 1 - \gamma$.

γ parameter

Gamma shows the rate of people that goes from the hospitalized group to the dead group.

In order to determine it, it is necessary to obtain first the case fatality rate which directly identifies the ratio of people that dies affected by the disease.

As γ directly depends on the measurement´s data for two of the groups of the model, while both θ depends on γ, the three of them need to be considered as variable parameters, as they are going to be different in every Danish region.

5.1.3 Population

The population parameter or P (as it is known in the Final modelling after introducing the improvements) is the starting point of the whole modelling process, as it will differ depending on the region that is being studied.

At the date in which the Covid-19 started to have cases at the Danish country (this day matches with the 12th of February according to the data extracted from the Danish Authorities) Denmark has a total population of 5817134.

This population is distributed into five different regions in the following way:

Region	Description	Population
Hovedstaden	Main region in Denmark. It includes the capital city, Copenhagen. It is emplaced in the northern part of Zealand.	1846023
Midtjylland	Central region of Denmark. It is emplaced in the middle of Jutland.	1321453
Nordjylland	Northern region of Denmark. It is emplaced in the north of Jutland.	589755
Sjaelland	Zealand´s Region. It is emplaced in the southern part of Zealand.	836738
Syddanmark	Southern region of Denmark. It is emplaced in the south of Jutland.	1223165

TABLE 4

As it can be seen in Table 4, the value for the population varies a lot depending on the region that is studied. Consequently, the estimation results could vary depending on the region which is being analyzed.

5.2 Performance of the estimations

In this section, it will be shown the performance of the different states in an estimation of just a couple of months and in an estimation that includes the entire pandemic time frame.

The chosen months for the short estimation have been the ones corresponding to the year 2021, what means that the estimation will include the period of time from the 1st of January, until the 6th of May which is the last day in which the data is known.

In the other hand, the period that covers the long estimation will start with the date in which the first Danish Covid-19 case approximately became infected (12th of February 2020) **(18)**, and it will last until the day in which data is no longer available which is the 6th of May 2021, meaning that the short estimation will be a subset of the long one.

There are two main reasons why a short-term and a long-term estimation are required, and here they are citated:

- **The performance of the algorithm regarding the amount of data received:** While the short estimation will take samples of 126 days, the long estimation will have data from 449 days, which implies more than three times the data from the other.
- **The vaccination state:** The vaccines didn´t start until the first dates of 2021. Regarding this fact, it will be interesting to see how the vaccination state performs in the long estimation, as there is no vaccine administered in great part of the time frame that it covers.

To show the comparison between the short and the long estimations, the region of Hovedstaden has been selected, as it is the region with more people in Denmark.

The images that will be shown in the following sections will be analyzed further in detail in Short-term vs Long-term estimations.

5.2.1 Short-term estimation

To simulate the model that appears on Modelling.3, some initial conditions are required.

In this case, as the short-term estimation starts on the 1st of January 2021, the initial conditions has been chosen according to the Danish Authorities´ data that has been used in this report.

So that, on the 1st of January of 2021 in Hovedstaden, the different states of the system have the following amount of people **(12)**:

State	Number of people
Susceptible	1764034
Exposed	339
Infected	13516
Hospitalized	483
Deaths	904
Recovered	68828
Vaccinated	0

TABLE 5

The fixed parameters will remain with the same value that was shown in Fixed parameters while the variable ones will vary in accordance with the data that is obtained from the Danish Authorities.

The values of the mentioned "Variable parameters" for this estimation are shown in the following table:

Parameter	Value
CFR	0.0113
θ_I	0.062
θ_H	0.062
γ	0.000706

TABLE 6

After running the short estimation, which lasted an elapsed time of $59.56\ seconds \approx 1\ min$, the following plots were obtained:

Susceptible state

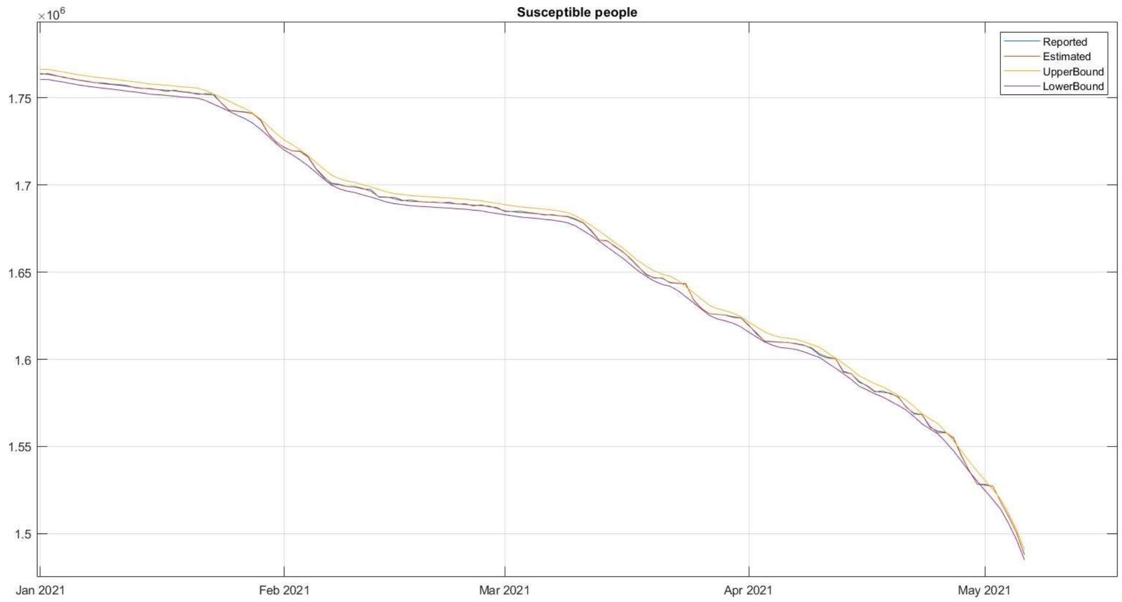

FIG. 2

In Fig. 2, it can be seen that the susceptible state (red) behaves in a good way according to the reference curve (blue). It starts in a value that can be considered that matches with the initial condition mentioned before, while the rest of the curve tend to have almost the same shape as the reported one except for some fluctuations that are sharper than the blue plot. These fluctuations rarely cross the confidence intervals, so it can be said that it mostly stays inside the boundaries.

It is worth saying that this graph is plotted with a scale of 10^6 so any minimal fluctuation or difference that can be shown between both plots, implies a big difference between the estimated and the real data.

Infected state

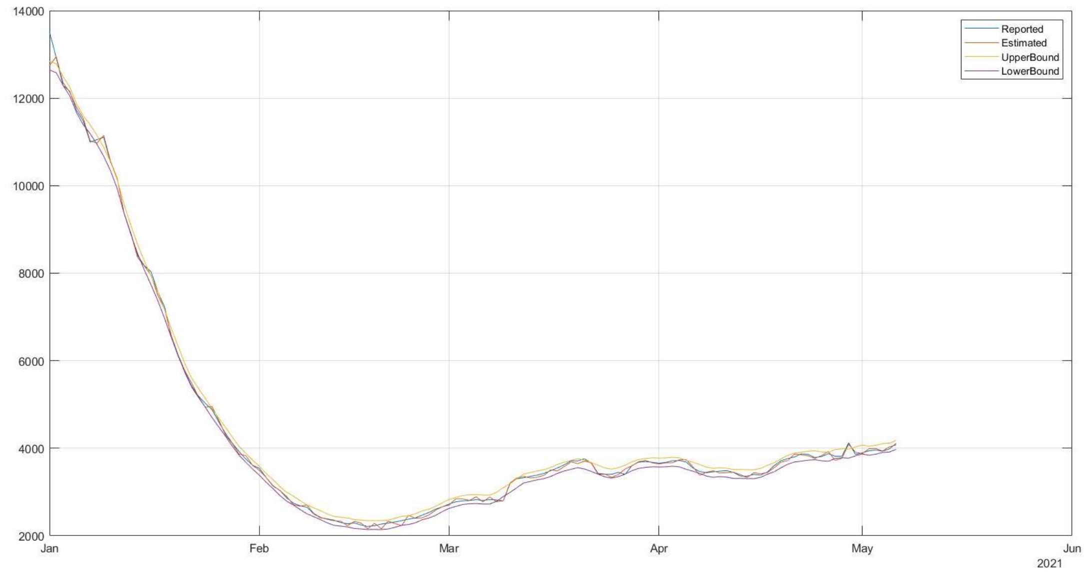

FIG. 3

Regarding Fig. 3, the initial value seems to be a bit smaller than the one that is shown in the reported plot. Despite of this, the estimated plot catches very fast the shape of the reported plot, and it also behaves well during the entire estimation, only crossing a few times the boundaries.

As it happened with the susceptible case, there are some dates where the values fluctuate in a higher or lower way, but it rarely loses the shape. In this case, the difference between the red and the blue plot can be appreciated better, as the prediction will only differ around some tens of cases.

Recovered state

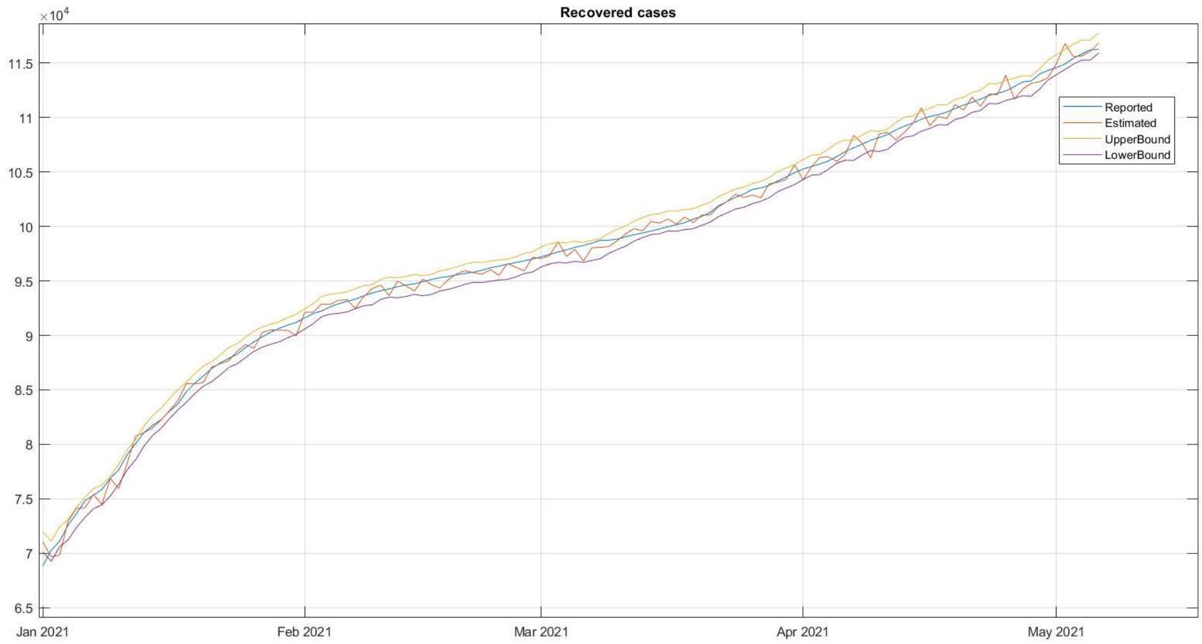

FIG. 4

In Fig. 4, the recovered state can be observed. At first sight, it is the picture that is more different than the one that is reported by the Danish Authorities. Nevertheless, it has a similar beginning and ending values, and the red line is always following the blue line although it oscillates with higher and lower values around it.

As a consequence of the sharp oscillations that can be seen in Fig. 4, there has been added in the plot some 95% confidence intervals that bound the plot, in order to see if the results can be considered as valid or not.

Vaccinated state

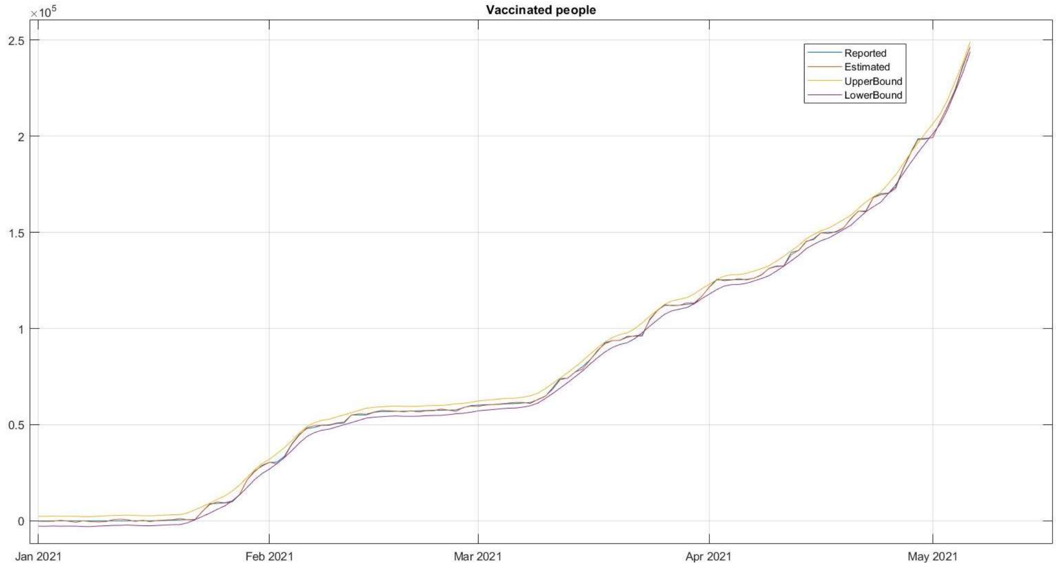

FIG. 5

Lastly, the plot for the vaccinated state can be seen in Fig. 5. At first, it can be seen that the days when the vaccination did not start (the first vaccinated person was registered on the 13th of January 2021) the plot remains at the value of 0. As it has been mentioned before, this was one of the main reasons to plot two different time frames for the estimation.

Furthermore, as it happened with the susceptible plot, in Fig. 5 it can be appreciated that the estimated plot fits well the reported one. Also, the little variations that can suffer this plot would lead in a significant change due to the scale (10^5) so although the fluctuations are minimum, it can lead in a variation of a couple of hundred cases. Despite of this, the red plot mostly remains inside the boundaries, so these oscillations could be expected.

The rest of the states´ plots (Hospitalized, Dead and Beta (β)) will be shown in Results as they are part of the outputs of the model.

5.2.2 Long-term estimation

This estimation will also be done with the Hovedstaden data but starting from the day when the Covid-19 arrived at Denmark. The first confirmed case in this region appeared the 25th of February 2020. As it has been mentioned before in Fixed parameters, the algorithm is working with an incubation period of $T_i = 14\ days$, so it can be considered that the first Covid-19 case detected in Hovedstaden on the 25th was already in the exposed group around the 12th of that month, which is the first day where data is available.

Consequently, the 12th of February 2020, the region of Hovedstaden had the following initial data:

State	**Number of people**
Susceptible	1846022
Exposed	1
Infected	0
Hospitalized	0
Deaths	0
Recovered	0
Vaccinated	0

TABLE 7

As it happened with the short-term estimation, the variable parameters will vary depending on the CFR.

This parameter needs two numbers to be calculated: the whole number of deaths registered in Hovedstaden during the test (which corresponds to the value that is registered on the final day of the short estimation: 6th of May of 2021), and the cumulative number of people that has already entered the recovered group at the last day of the simulation.

As the long estimation ends in the same day, the 6th of May 2021, the variable parameters will have the same values for both estimations. These values were shown in Table 3

After running the simulation, which lasted an elapsed time of $797.11\ seconds \approx 13{,}3\ mins$, the following plots were obtained:

Susceptible state

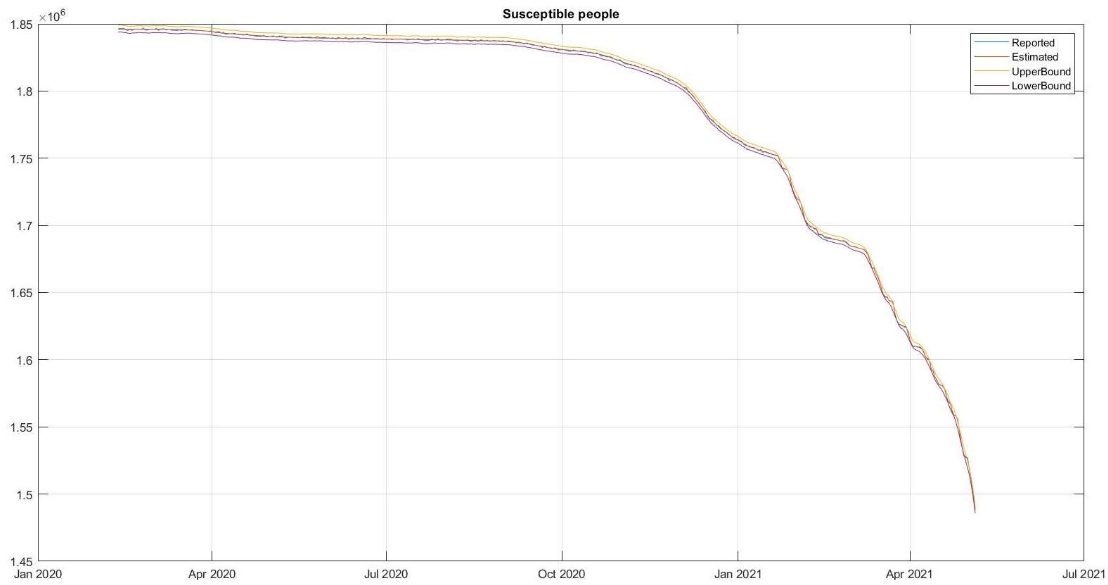

FIG. 6

In Fig. 6, firstly, it can be appreciated that the modelled plot starts around a value that seems to be near to the initial value that has been set before. Also, it matches the final value of the reported data, so as the entire plot doesn´t seem to exceed the confidence intervals and there are not high differences in the shape between the blue and the red lines, it can be assumed that the model again matches well the reported plot.

Furthermore, this longer estimation shows better the time frames where the susceptible group significantly decreased, due to the serious effects of the pandemic in the Danish region.

Infected state

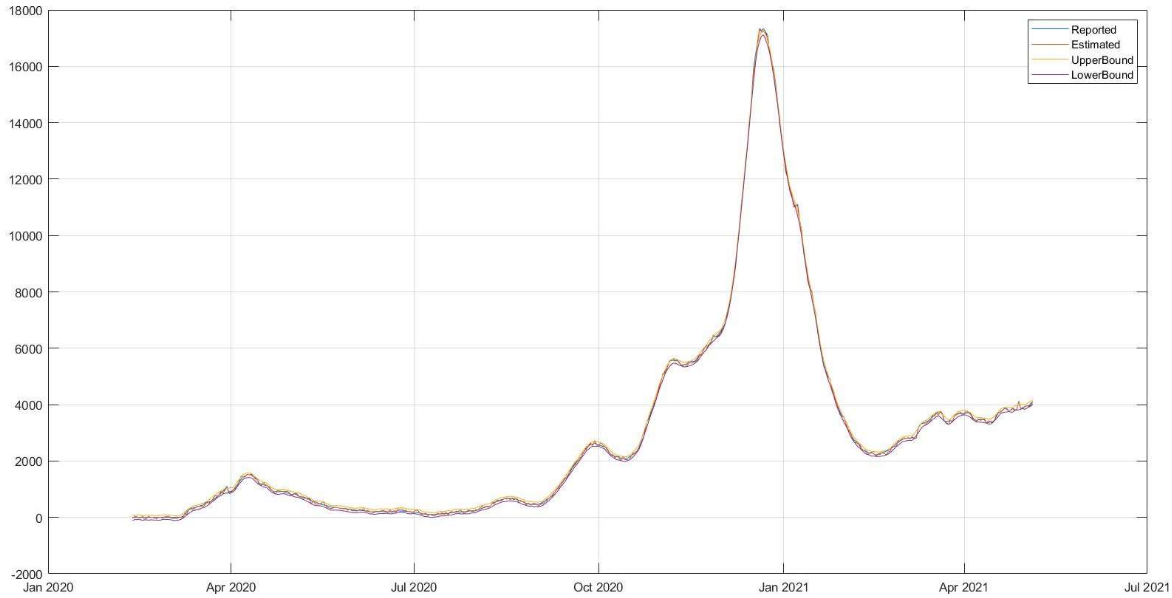

FIG. 7

Regarding the infected plot shown in Fig. 7, the first values seem to behave well when the value of infected people is equal to 0. There are some fluctuations that maybe cross the 0 line and become negative, but as the negativity of these values is too small that can even be neglected, there is no need to add negative constraints to the model.

In relation to the plot, it can be seen that the red line has almost the same shape as the reference line, so despite of some fluctuations that can differ the result in some tens of cases, it can be considered that the infected state adapts well to the Danish Authorities´ data.

Also, it can be appreciated that the red plot crosses a couple of times the 95% confidence intervals, especially when the cases increase, but generally it stays inside the desired boundaries

Recovered state

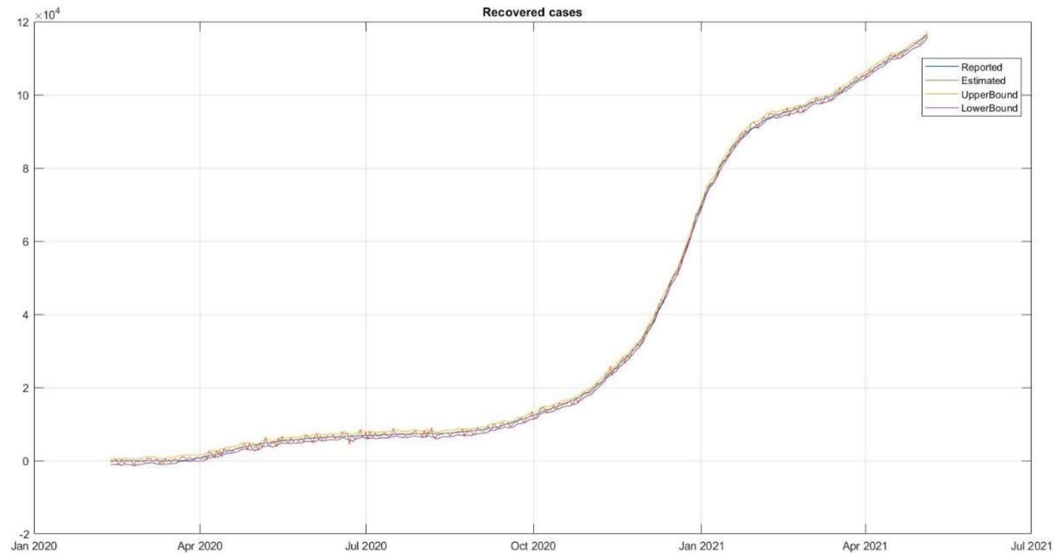

FIG. 8

As it happened with the short estimation, some bounds needed to be added in the plot shown in Fig. 8, because of the fluctuations that the recovered state makes around the blue line.

Despite of these variabilities, the estimation plot rarely crosses the 95% confidence interval margins. Moreover, the red line seems to be oscillating around the reported one, ending in no significant changes between both plots.

Vaccinated state

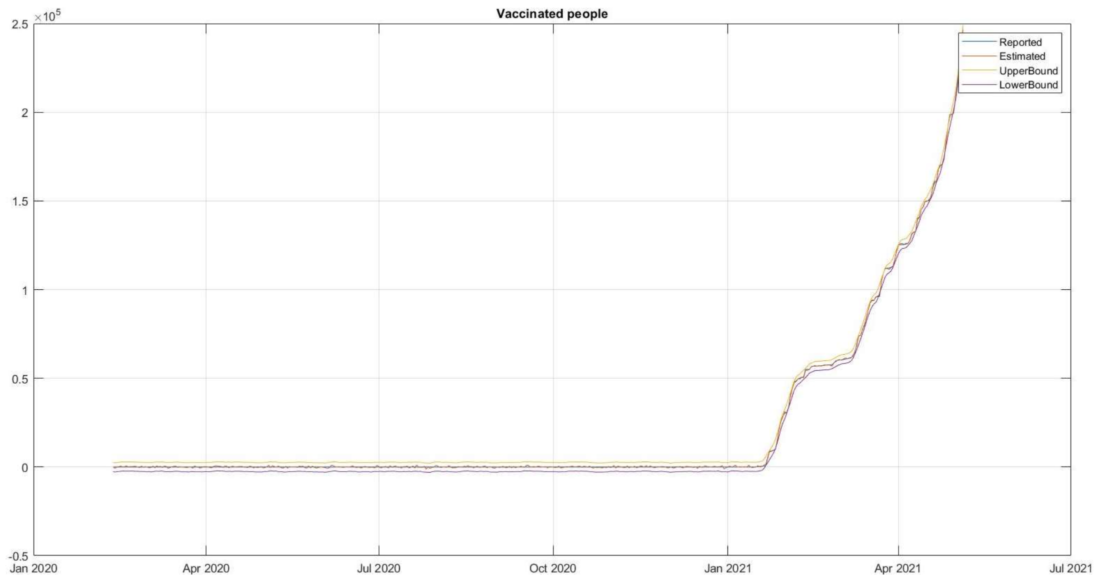

FIG. 9

Fig. 9 shows the vaccinated plot of the model for the long estimation. At first, it can be seen that the system remains at 0 in the time when there is no vaccination administered. It just varies with small oscillations that can be considered as null.

Furthermore, the red line adapts almost perfectly to the shape of the reported line, so it can be stated that the model follows well the reported data. It can also be appreciated that the model approaches the boundaries a couple of times, even it crosses a few of them, but in general terms, it fulfils the 95% boundary conditions.

In this chapter, some figures appeared in order to show the behaviour of the model and how it worked after estimating it with an Extended Kalman Filter. Now, that this performance is already shown, the outputs of these states are going to be presented in the Results.

6. Results

This chapter will present the plots obtained after simulating the model presented in Final model. Concretely, the results that are going to appear, are the ones obtained from the outputs of the model that were previously stated on Outputs of the model.

Apart from these outputs, it has been decided to show the results from the parameter beta. As it was explained in Modelling, the parameter Beta can be considered as unique in this model. It is the parameter that measures the transmission rate between the susceptible group and the exposed group; in other words, it measures the rate of infections caused by Covid-19.

For that reason, at it is one of the most important parameters in the model, it has been estimated as another state of the system and its behaviour is going to be seen in this chapter.

As this report aim to expose the results for the five different regions in which Denmark is divided, this section is going to be divided into five sections, each of them for a single region.

At the end of the section, it is going to be shown how much impact has the Kalman Gain in the model´s prediction; also, a forecast of 40 days will be carried out for the region of Hovedstaden.

The images given in this chapter are going to be analyzed more in detail in the Discussion.

6.1 Results per region

6.1.1 Hovedstaden

As it was already mentioned in Population, Hovedstaden is the capital region of Denmark where its capital city, Copenhagen, is located.

This region is also the most populated in Denmark with an initial population of 1846023 inhabitants at the date where the pandemic started at Denmark (12th February 2020).

Active cases

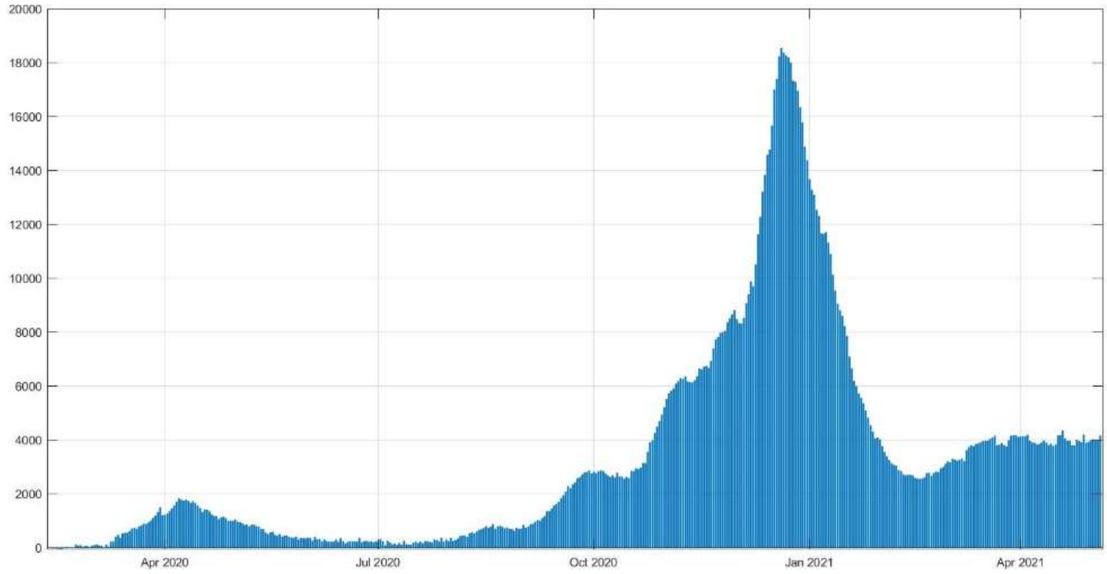

FIG. 10

In the first image of this chapter, Fig. 10, it can be appreciated how the pandemic evolved in the region of Hovedstaden. Concretely, it can be seen the number of active cases that the region had during each day of the pandemic.

At the first days of the pandemic, where the Lockdown was active, there was a peak of cases of less than 2000 active cases in a day, but after the second-wave of the virus arrived at Denmark the graph changes a lot. In fact, at the end of the year 2020 it can be seen an increase in the number of active cases until it reaches its top on the 20th of December 2020, when there was a total of 18530 active cases at the same time.

Finally, it decreases its value until an average of 4000 active cases, which is the value that is maintained until the 6th of May 2021.

Deaths

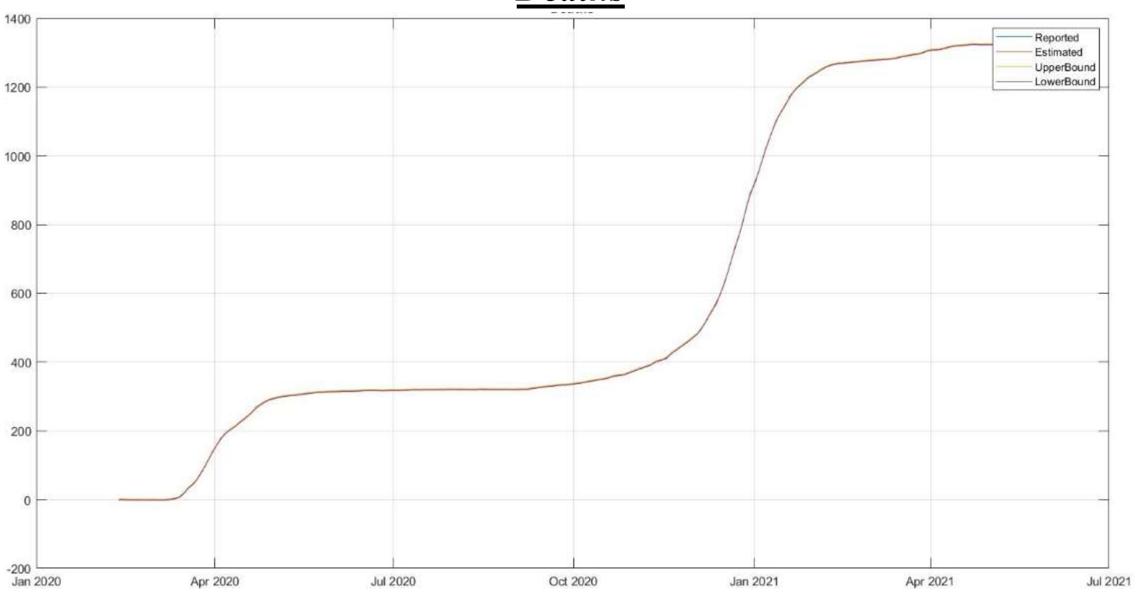

FIG. 11

Fig. 11 shows the dead state of the model. This state is characteristic inside the model as it is the state with the lowest variance in the entire model.

As a consequence of that, it can be seen that both plots, the blue and the red one, have almost the same shape in the entire picture. It can also be seen that there is a period of time in which the deaths stopped making almost a straight line in a value near of 350 cases, but when September 2020 arrived it started to increase again until an ending value of 1324 deaths at the date of 6th of May 2021.

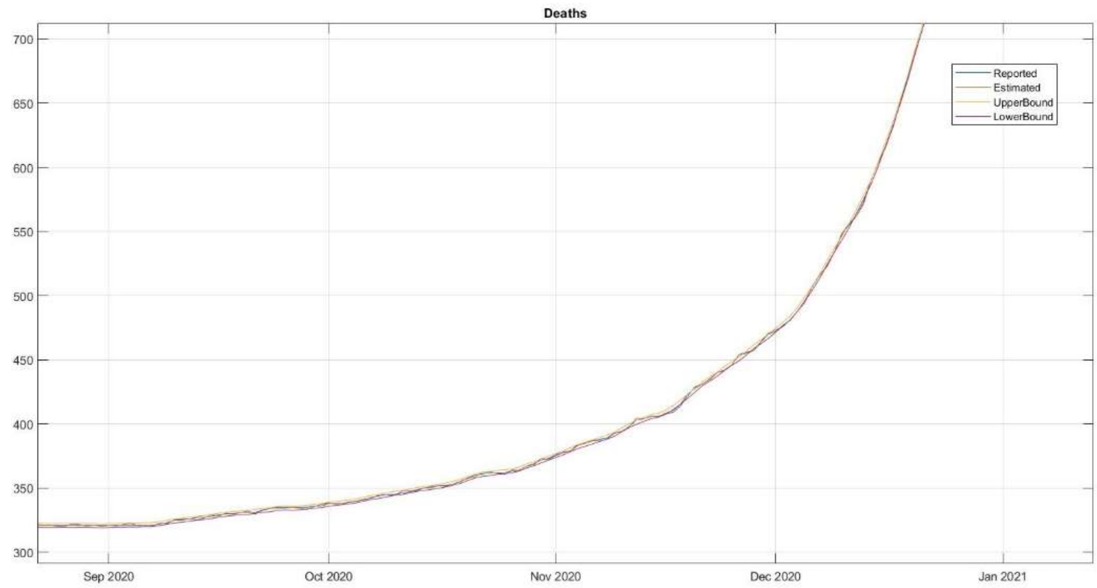

FIG. 12

In this picture also there are an upper and a lower confidence interval plotted, but as it happened with the blue and the red line, they are too joint to be appreciated. For this reason, Fig. 12 has been also included as an approach of the dead state plot, concretely taking the months from September 2020 to January 2021, in order to proof that the estimated plot mostly fulfils the confidence intervals.

Furthermore, in Fig. 12, it can also be appreciated that there are some fluctuations that tend to go from a higher number to a lower one. In a plot that is showing the number of dead people, it should not be like that, as there cannot be fewer deaths than the previous day. Nevertheless, as these oscillations are very small, they can be considered as negligible.

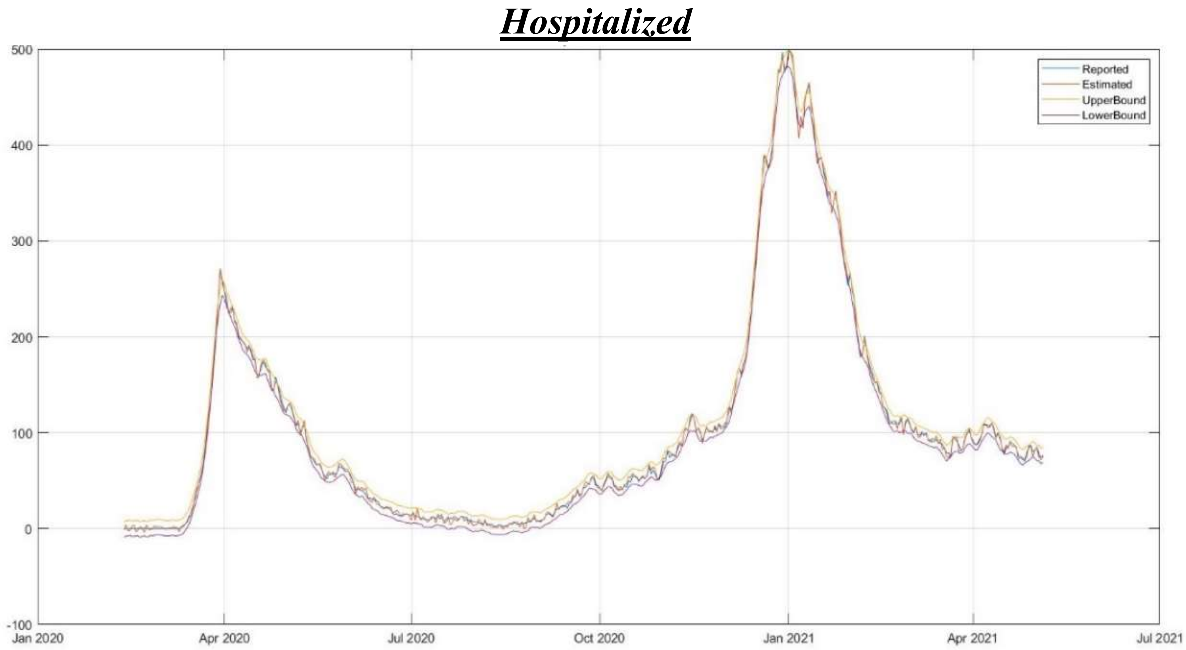

FIG. 13

In Fig. 13 it can be seen the plot for the hospitalized people. At first sight, it can be seen that there are two periods of time where the hospitalized people grew a lot. The first one matches with the beginning of the pandemic (April 2020) while the second one appears at the end of the year 2020.

Looking at the estimated and reported plots, it can be appreciated that both of them are a bit bumpy. Despite this, the estimated plot seems to follow the blue line nicely, although there are a few times where they have different values. Nevertheless, this picture also includes an upper and a lower 95% confidence interval, which will determine how confident are the values. The red line crosses these intervals just a few times during the draw, especially when the graph increases exponentially, but it remains inside the intervals most of the time.

Last but not least, thanks to this picture it can also be known the data for the maximum number of hospitalized people being in Hovedstaden at the same time. Concretely, this data is obtained from the first day of the year 2021, when 498 people were estimated to be hospitalized in this region.

<u>Beta and Effective reproduction number</u>

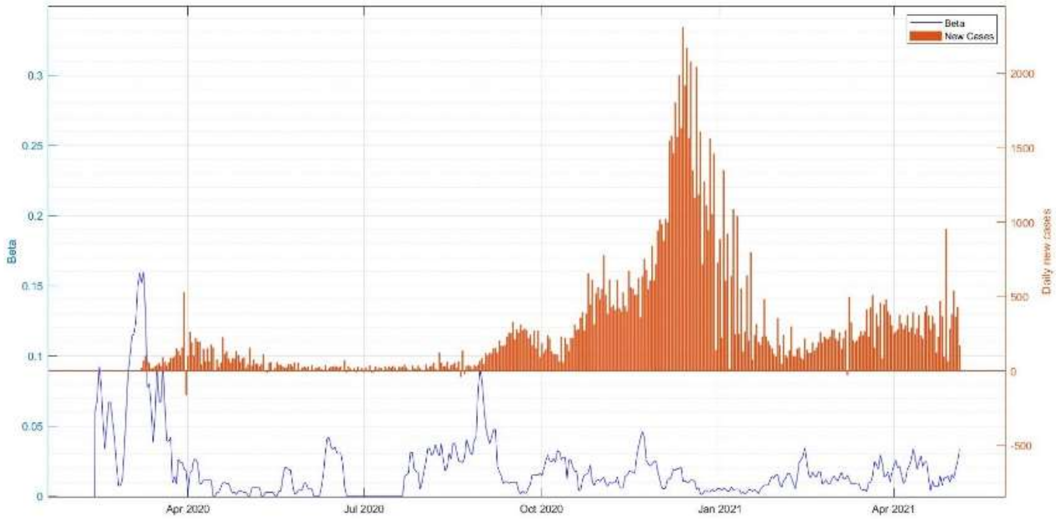

FIG. 14

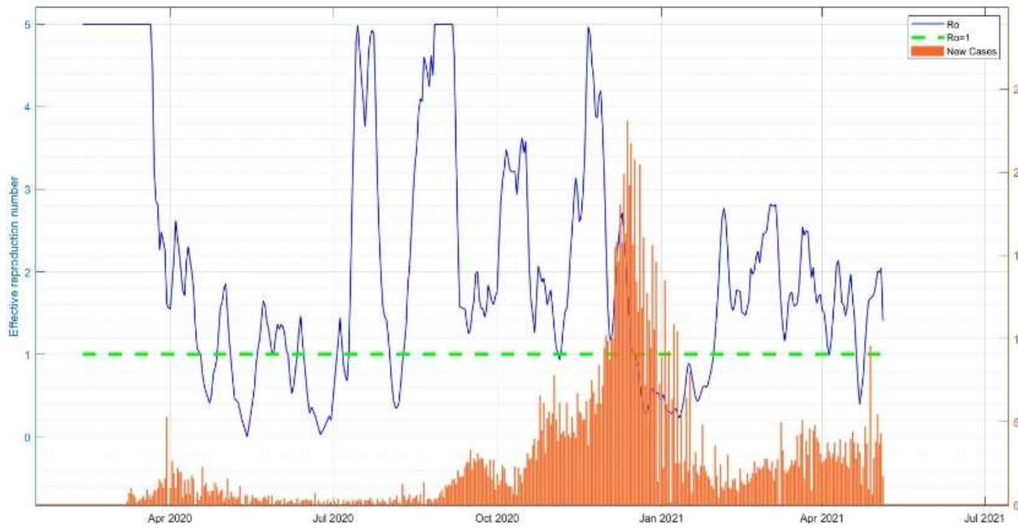

FIG. 15

In Fig. 14, it appears a comparison between the daily new cases of Covid-19 and the behaviour of the Beta parameter.

At the beginning of the plot when there are no cases yet, Beta seems to have an unusual behaviour. It increases its value to almost 0.1, then it rapidly reduces it to again start increasing until an unrealistic value around 0.16. But when the middle of March arrives, the plot changes drastically.

This blue plot shows that Beta is oscillating between the values of 0.005 and 0.01 when there is not an increase in the daily new cases. These oscillations sometimes get to higher values, but when the daily new cases remain in small numbers, the peaks are low.

In the other hand, when the daily new cases increase, Beta start to jump a bit more, oscillating around values above 0.15 and suffering some peaks that once reach a value of 0.09. This particular case can be considered as an exemption, as the rest of the plot, when the daily new cases are in an uptrend, doesn´t overcome the value of 0.05.

Similarly, Fig. 15 compares the performance of the Effective reproduction number compared to the number of new cases that appeared daily in Hovedstaden.

At first sight, it can be seen that the Effective reproduction plot is really unstable, it can vary from a value near to 0.4 to a value of 4 in a couple of days. Despite of this, as it happened with Beta, the lower values of the Effective reproduction number appears when there are less daily new cases, and the other way round, when the daily new cases tend to increase, there is an increase also in the Effective reproduction number´s plot.

In this picture, it can also be appreciated that there is a green line at the value of 1. This value is fundamental for this parameter, as this is the frontier value that determines if the infection is increasing or decreasing its spread. After this assumption, back in the picture, it can be seen that most of the time the blue plot has a value higher than this green line, what would mean that the majority of the time the virus is spreading fast.

Nevertheless, when the plot reaches values near to 0 or at least, below 1, it means that Covid-19 was decreasing its spread, and these numbers matches with the time frames where there were low values for the daily new cases.

6.1.2 Midtjylland

Midtjylland is the second region in Denmark in terms of population. On the 12th of February 2020, 1321453 people were living in this region.

This region is located in the middle of Jutland (its name can be translated into English as the Central Region of Denmark) and it has the largest extension in the country (13000 km^2).

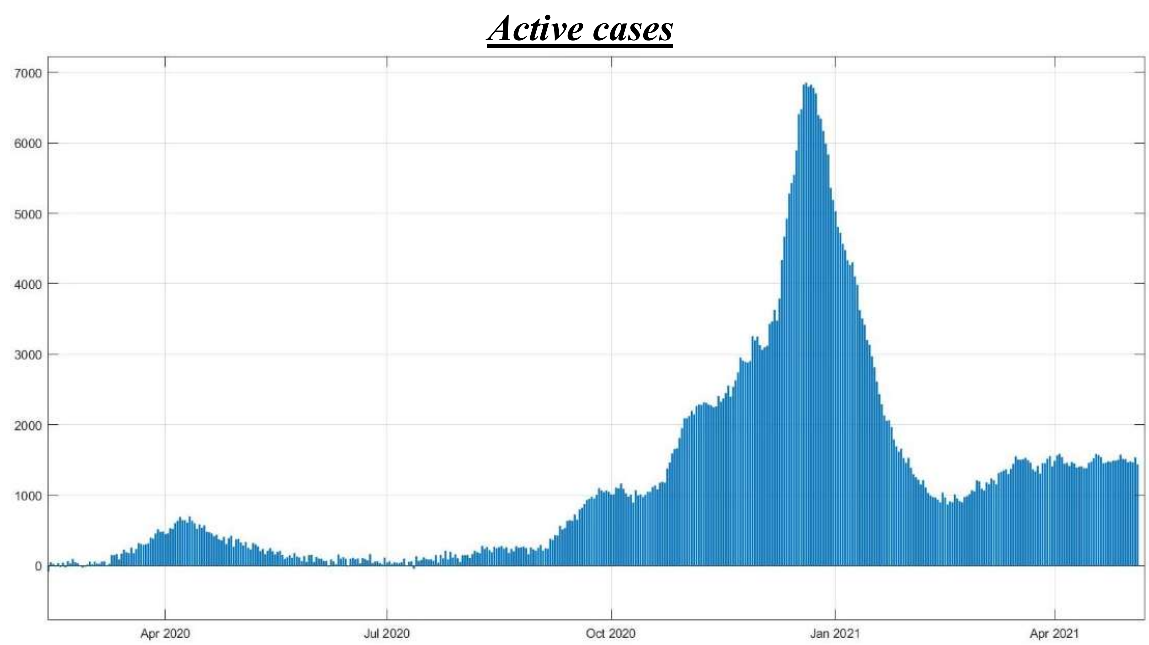

FIG. 16

In Fig. 16, there is a bar plot showing the number of active cases that the region of Midtjylland had during the days of the pandemic.

The plot shows, at first, a lower move to a local maximum number of around 700 active cases in the first part of the pandemic time frame, which is reduced to a hundred of cases after June.

The bars seem to oscillate around this value, until they start to draw a curve that grows almost exponentially until it gets to an absolute maximum value of 6852 active cases. This value is obtained the 22nd of December 2020 and is the peak value of the graph.

Although it seems to stabilize for a couple of days around this value, it suddenly starts to decrease really fast to an average of 1500 cases which it maintains until the end of the plot.

Deaths

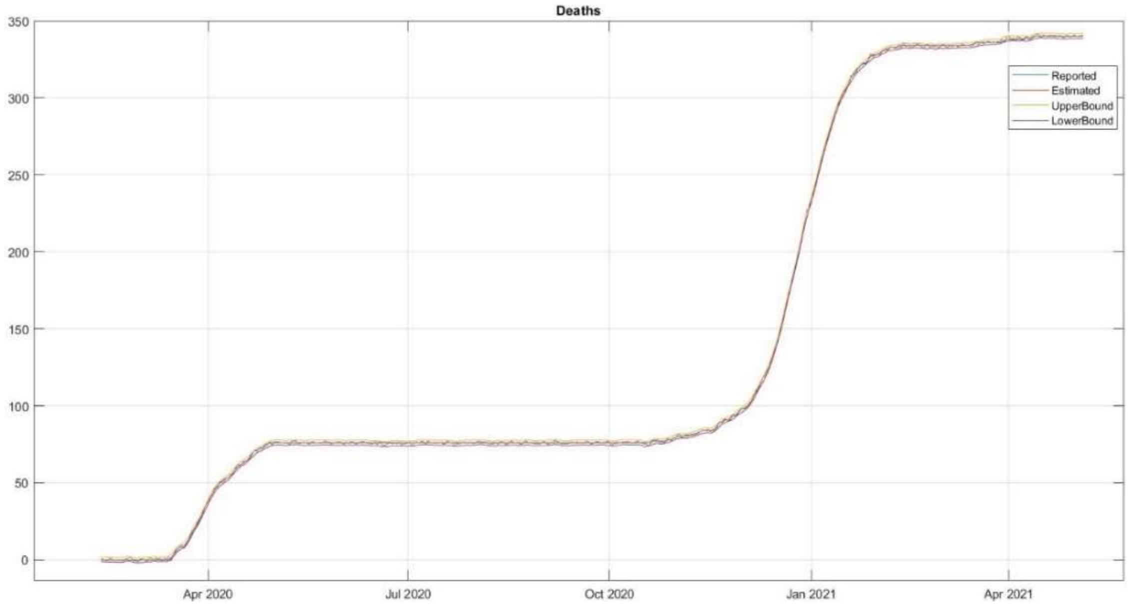

FIG. 17

Fig. 17 shows how the dead state performs during the estimation.

Firstly, it can be seen that there is an increase in the number of deaths of around 75 cases, in the time frame of the pandemic between March and May 2020. Then it seems to fully stabilize showing no increasing attitude until the month of November, when it begins to increase again.

The plot increases until a value near to 330 deaths, and it has a slight increase at the end of the graph that reaches the peak and final value of deaths which is 340.

Finally, the estimated plot has been bounded with an upper and lower confidence interval. During the entire time frame, the estimated line seems to be inside these two bounds, although there are a few cases that goes out of them. Nevertheless, the red line has most of the time the same shape as the blue line which are the reported deaths extracted from the Government data.

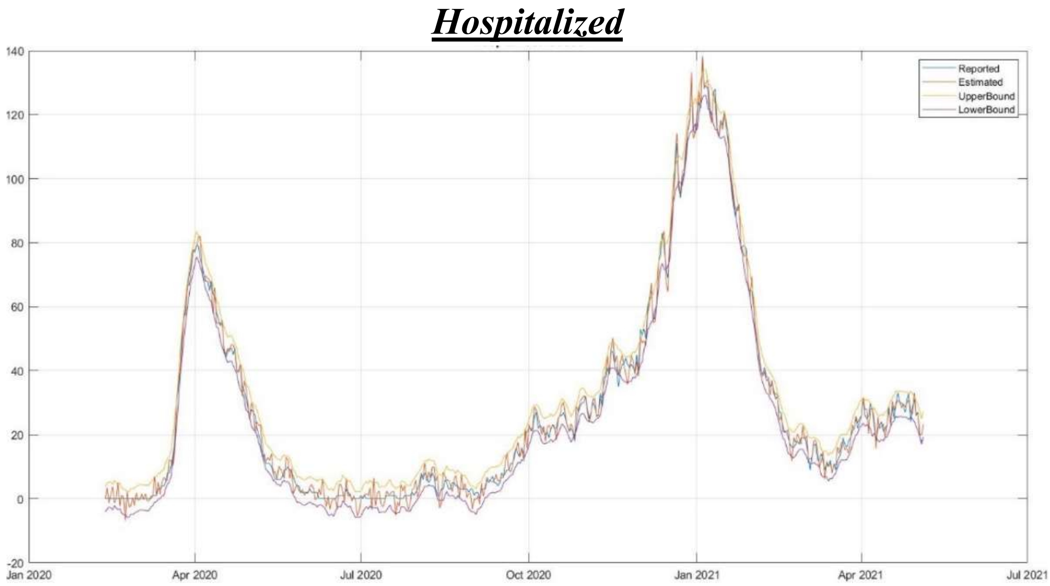

FIG. 18

In Fig. 18, it can be seen the amount of people that stayed in a hospital due to Covid-19 in Midtjylland.

At first sight, it can be appreciated that the red line which belongs to the estimated cases, oscillates around the blue line of reported cases. It seems to catch sometimes the shape of this reported line, but it doesn´t remain with that shape for a long time. Despite of this, the red plot seems to fulfil the confidence intervals requirements, and it stays inside of them most of the time. There are a few cases where it goes above or below these intervals, but these cases can be included in that 5% of cases that could not be inside of these boundaries.

In addition, there are a couple of times that the number of hospitalized people is set to a negative value. These oscillations are present when the reported cases approximate to a 0 value, so they can be considered as negligible, as the number that they reaches is around a couple of units.

To conclude, this figure draws two maximums in the graph, one for each phase of the pandemic. The first of them appears around the first days of April and it reaches a value above 80 hospitalized people. The biggest one is reached on the 4th of January 2021, when the value of hospitalized people is estimated to reach 138 cases.

Beta and Effective Reproduction Number

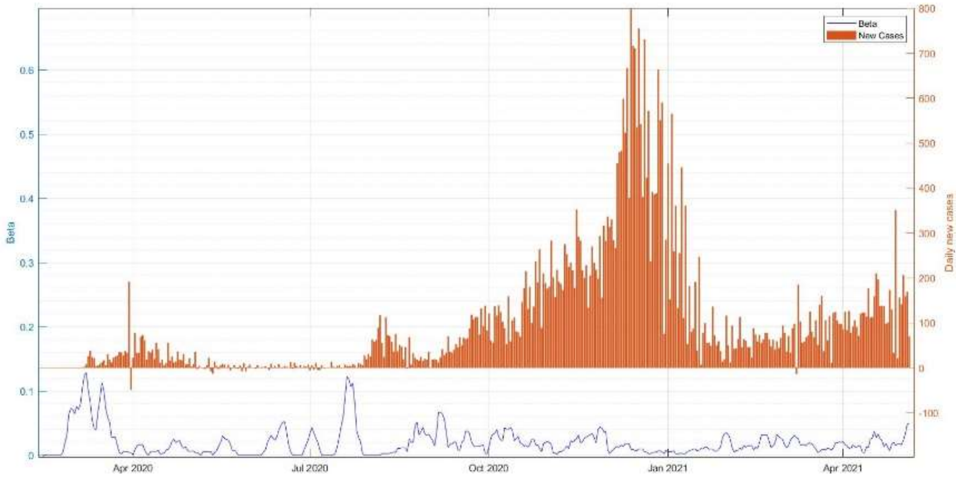

FIG. 19

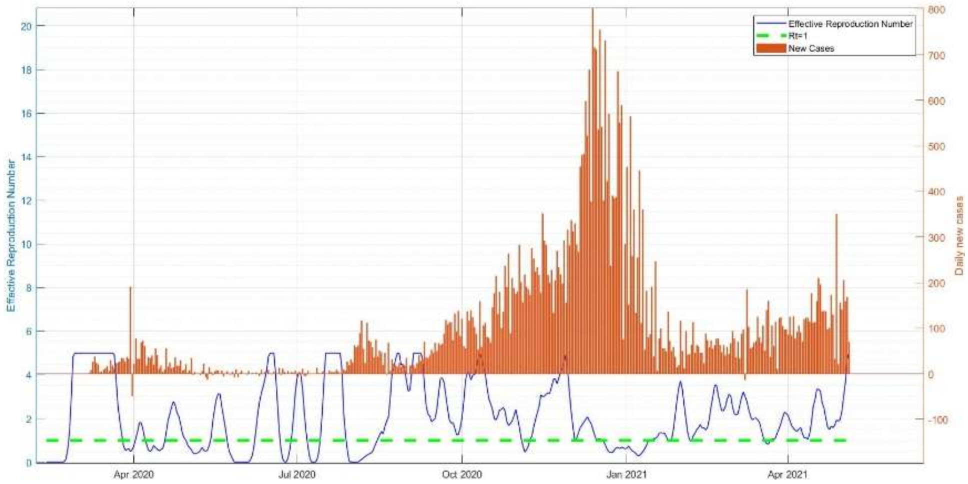

FIG. 20

The Beta plot for the population of Midtjylland can be seen on Fig. 19. Moreover, it is compared with the Daily new Covid-19 cases that this region had during the pandemic time frame.

At contrary as it happened with Hovedstaden region, Beta seems to have a reasonable value when the pandemic started, but it suddenly grew until the plot reaches an unrealistic 0.13 value. This peak is the biggest peak that the plot shows, and it has 2 similar peaks in the time frame between March and August of 2020.

Excluding these cases, the rest of the plot seems to oscillate between the values of 0 and 0.01 when the daily new cases are in a downtrend.

In the other hand, when the daily new cases increase to greater values, Beta seems to fluctuate around a value of 0.02, with the presence of some peaks that can reach the number of 0.04.

In the bar plot, it can also be appreciated that there are some exceptional cases that have a negative value. Fortunately, this fact just happens a few times and Beta doesn´t care much about it, as when this happens, it remains in positive values.

In the other hand, Fig. 20 compares the Effective reproduction number with the bar plot which shows the daily new cases of Covid-19 in Midtjylland.

Firstly, it can be concluded that the Effective reproduction number has an oscillating attitude until August 2020. It shows oscillations that goes from 0 to a value of 5. This behaviour matches with the first wave of the virus, where the daily new cases never overcome the number of 80. In addition, later in this time frame, before August, it can also be seen some rare cases that has negative values for the daily new cases.

Fortunately, after this month is passed, the plot seems to bounce with a bit more criteria. Although there is still a presence of sharp fluctuations, in general terms, the Effective reproduction number seems to increase when the daily new cases increase, while it decreases or oscillates when the daily new cases decrease or stays near a value of 0, respectively.

6.1.3 Nordjylland

The North Denmark Region or Nordjylland is the Danish area located in the North of Jutland. Among other things, it is the region where Aalborg University is placed.

Nordjylland is also the region with the lowest population number. At 12th of February 2020, this region only had 589755 people living at it.

Active cases

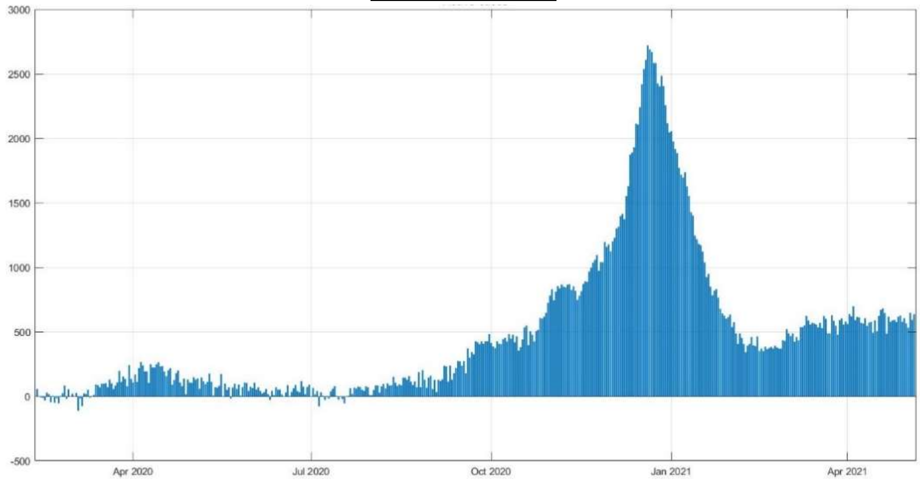

FIG. 21

Fig. 21 shows the progress of the active cases due to Covid-19 in the region of Nordjylland.

First of all, there are a few cases that shows a negative value, which can be neglected as the value is very low and also because it happens just a few times in the graph.

For the rest of the sample, the active cases seem to draw an uptrend in the first wave of the Covid-19 pandemic, which matches with the time frame between March and May of 2020, while after that, it will go near to a value of 0 when the summer of 2020 arrived.

Then, the graph shows again an uptrend, starting from the month of August until the beginning of 2021. It´s worth saying that this growing movement reaches values that are much bigger than the values that were reached in the first wave. Indeed, this is the part of the graph where the peak of active cases is reached.

This peak of active cases belongs to the data of the 20th of December 2020, where 2519 people were estimated to be infected or hospitalized at the same time due to Covid-19 in Nordjylland.

Finally, the plot goes down from these peaks' values until the third month of 2021, where it make a slight increase to oscillate around a value of 600 active cases, which is the number that remains until the end of the graph.

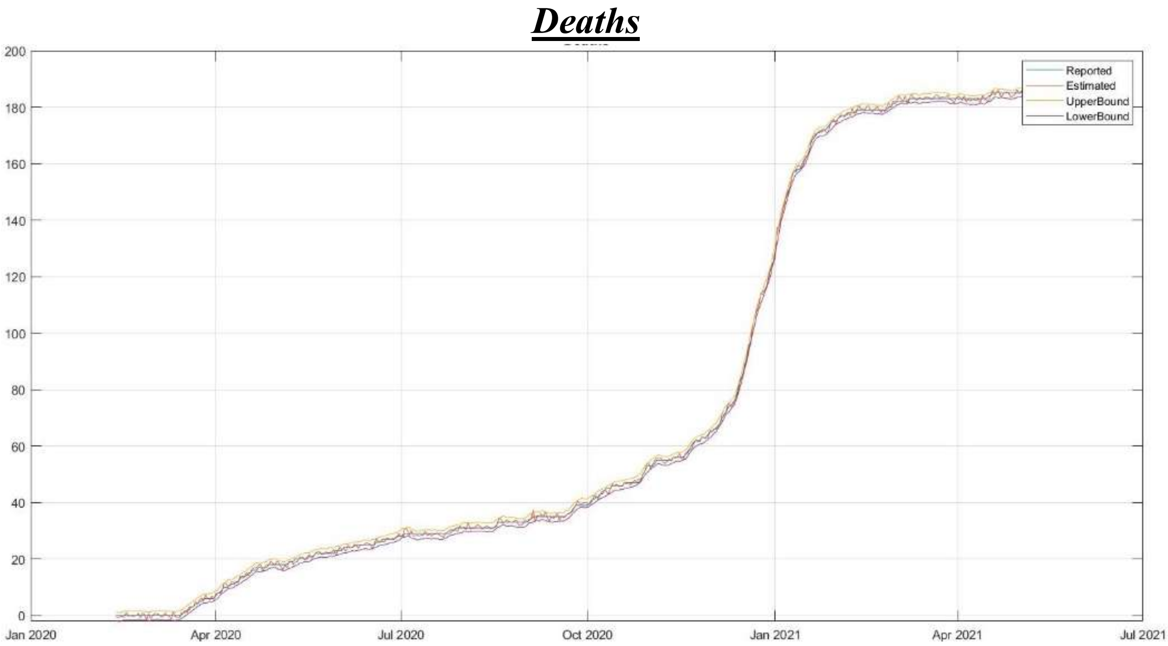

FIG. 22

In Fig. 22, the plot of deaths caused by the infection of Sars-Cov-2 in Nordjylland can be seen.

At first, the graph is stabilized around a value of 0 with slight fluctuations that can reach negative values. As these fluctuations are really small, they can be considered negligible, as they just mean a couple of negative decimals.

Before April of 2020, the graph starts to increase its value, following an uptrend that would characterize the entire plot until it gets to April 2021, where it adds just a couple of deaths more.

The plot increases almost linearly between the end of March of 2020 until the end of November 2020. Then, when the second wave affected Denmark, the graph starts to grow exponentially for a couple of months, until it reaches a value near 160 deaths.

Finally, after February 2021 it increases much slower, until the end of the graph where it shows a value of 185 deaths caused by Covid-19 in Nordjylland at the date of 6th May 2021.

Regarding the comparison between the reported and the estimated data, it can be seen that the red plot follows the shape of the reported one most of the time. Sometimes, the estimated line oscillates around the blue one, but these fluctuations have a value of just a couple of decimals. In fact, the red plot stays inside the 95% confidence intervals most of the time, except for when the graph grows very fast, where it can be found a couple of cases where the estimated plot escape from these boundaries.

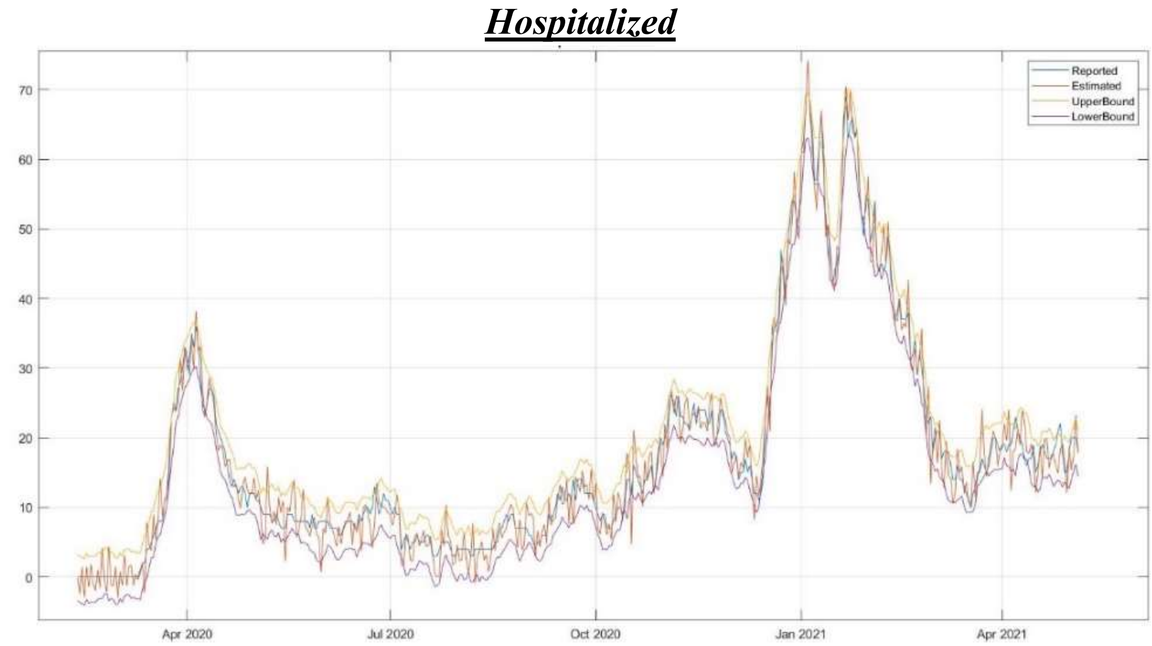

FIG. 23

Fig. 23 shows the development of the hospitalized cases in Nordjylland through the entire Covid-19 time frame.

At first sight, it can be seen that the estimated plot, which is the one obtained from the model showed on Final model, is constantly fluctuating around the Authorities´ data plot. Nevertheless, there are a few cases that even crosses the boundaries, but most of the time, the red line remains inside the confidence intervals.

Regarding the shape of the plot, firstly, in March 2020 it starts to grow until the next month where it found a local peak of 35 hospitalized cases. Then, the graph goes back to even less than 10 cases, until October 2020 arrives. After that, the plot starts again its uptrend until the absolute maximum is reached. This number is obtained the 4th of January 2021, when 74 people are estimated to be hospitalized at the same time due to the Sars-Cov-2 infection.

Then, the plot draws a smaller peak the 21st of January, to finally decrease until a value near 20, in which it will keep oscillating until the end of the figure.

<u>Beta and Effective reproduction number</u>

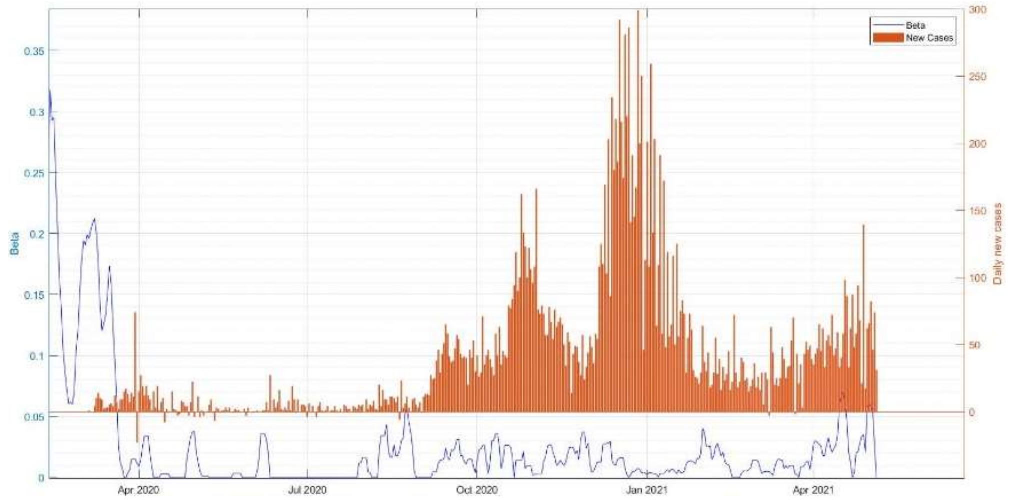

FIG. 24

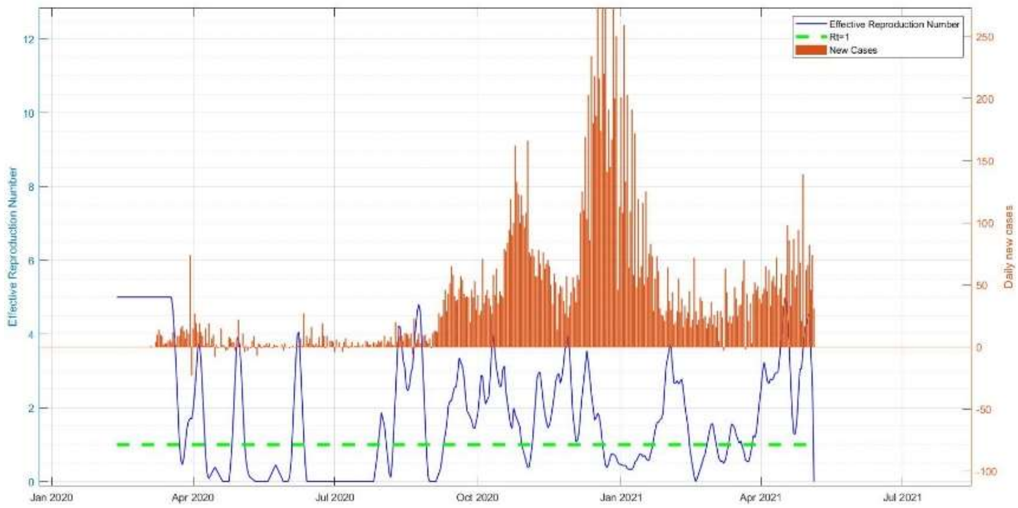

FIG. 25

Fig. 24 shows the comparison between the performance of Beta and the daily new cases for the region of Nordjylland.

Firstly, it can be appreciated that Beta behaves badly when there are not new cases, reaching unreal values higher than 0.2. Then, when the new cases appear, Beta starts to change its behaviour and remains near the value of 0.005, having some peaks that get to 0.04.

After that, when August of 2020 arrives, Beta increases its value and starts to oscillate around the value of 0.015 which is three times the value that was approaching in the first wave of the pandemic. This increase in Beta matches with the increase in the daily new cases suffered by Covid-19.

Finally, Beta have two more peaks that overcome the value of 0.05.

In the other hand, Fig. 25 shows the development of the Effective reproduction number compared to the appearance of new Covid-19 cases in Nordjylland.

At the beginning of the graph, when the new cases are low, the Effective reproduction number seems to bounce from low values near to 0, to values near to 4 in just a few days. This behaviour is maintained until September 2020 is reached, when the Effective reproduction number starts to stay above the green line, which means that it stays above 1.

In this period, the Effective reproduction number draw some peaks that are reached in a progressive way, not as before. Also, these peaks match with the maximums that can be observed in the bar plot.

Finally, in the 2021 time frame, the Effective reproduction number value can be seen below the green line, when the cases decreased. This behaviour is maintained until April, when the new cases grow again, and the Effective reproduction number draw two new peaks that go above 4.

6.1.4 Sjaelland

Sjaelland is the nearest region to the Capital Region (Hovedstaden). Its name comes from Zealand, which is the name that receives the right-side part of Denmark.

Sjaelland is the second lowest Region in Denmark in terms of population, with just 836738 people living there at the beginning of the pandemic; and also, it is the second smallest Region in Denmark, surpassed only by Hovedstaden.

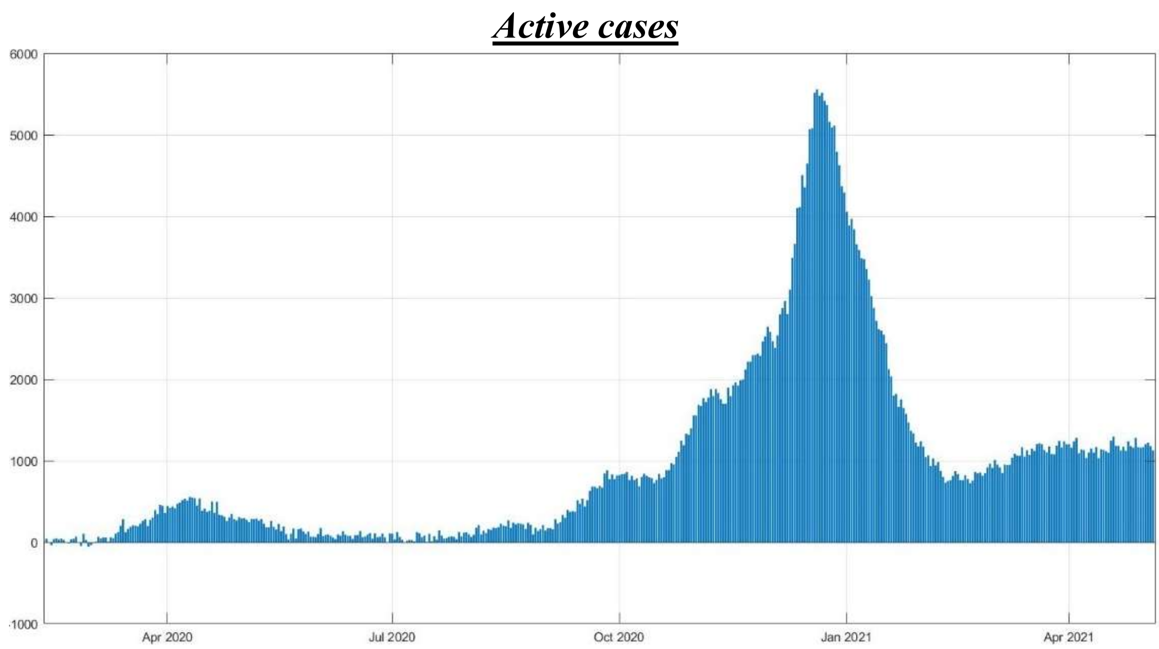

FIG. 26

In Fig. 26, it can be seen the daily number of Covid-19 active cases in Sjaelland.

In the first wave of the pandemic, between March and May 2020, the bar plot shows that a peak number near to 550 active cases are reached. After that, the graph decreases to values near 0 until autumn of 2020 arrives.

Then, it can be appreciated an uptrend in the plot that gets to its maximum value. This peak value is almost 10 times higher than the peak value that was reached in the first period of the pandemic.

This absolute maximum is obtained on the 19th of December 2020, when 5557 active Covid-19 cases are estimated.

At the beginning of the year 2021, the graph starts to decrease rapidly until the middle of February, when it starts to fluctuate around a value above 1000 active cases until the end of the plot.

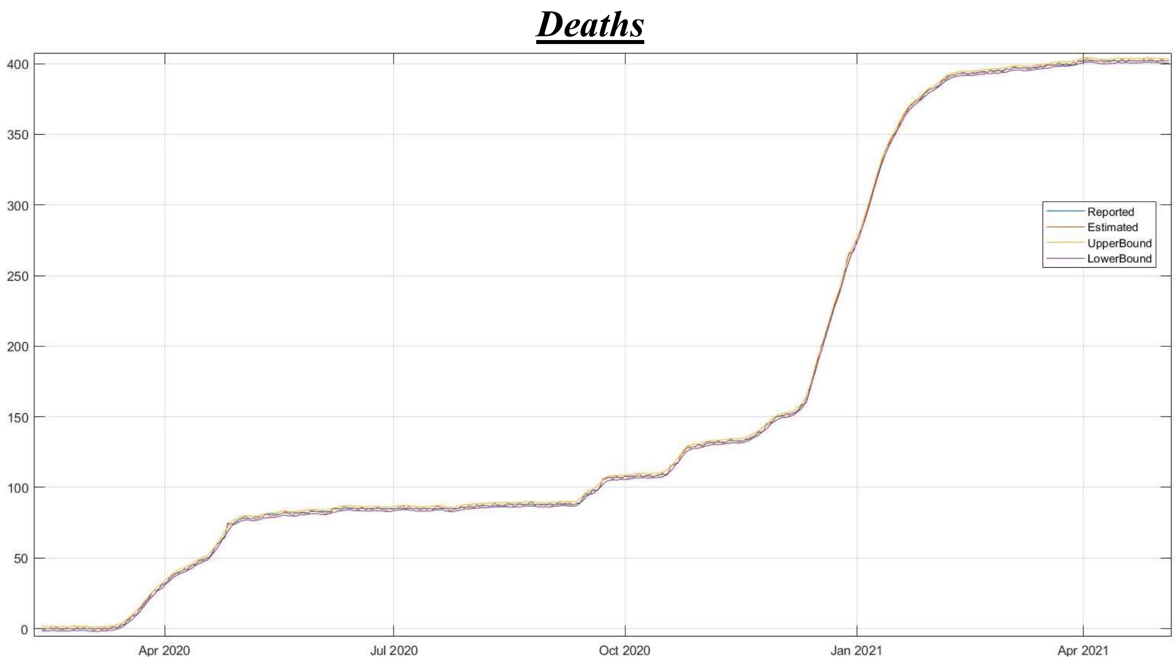

FIG. 27

Fig. 27 shows the development in the number of deaths caused by Covid-19 in Sjaelland.

Firstly, all the lines stay at a value of 0 until the month of March 2020 when people are estimated to start to die due to Covid-19. This graph increases until a value of 80 deaths and then stabilize for a couple of months around that value. When September arrives, the graph starts to increase again, and it is progressively gaining slope until the end of January 2021 where it starts to make an stabilization move again.

From this month until the end of the graph, there are only a few more deaths caused by Covid-19, reaching the final value of 402.

In general, in the figure it can be appreciated that the estimated results plot matches the shape from the Authorities´ data plot. It also remains inside the confidence intervals most of the time, although there are a few cases where it overcomes the boundaries when the value is drastically increasing.

Hospitalized

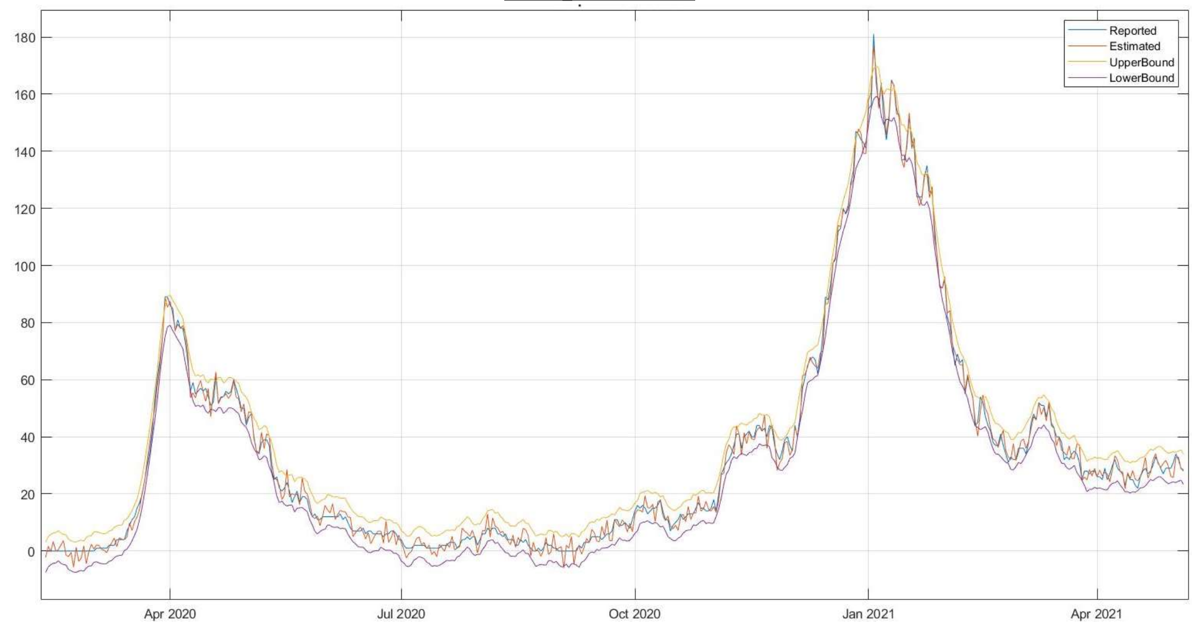

FIG. 28

In Fig. 28, it can be seen the development of the hospitalized cases due to Covid-19 in Sjaelland.

Firstly, it can be seen that the estimated plot catches in a good way the shape of the reported plot. Most of the time, this red line is sticked to the blue line or at least, fluctuating around it. Nevertheless, as the confidence intervals are present in the graph, it can be appreciated that these fluctuations rarely cross the 95% boundaries, meaning that they are under control.

Regarding the behaviour of the plot, it can be seen that before April 2020 it starts to increase fast until a value over 80 hospitalized people is reached. Then, after this month, it can clearly be seen that the hospitalizations are reduced and almost disappeared. The graph seems to oscillate between the numbers of 0 and 10 during the summer of 2020.

After October, the plot starts to increase again. Moreover, in this period of the graph, it reaches the maximum number of hospitalizations registered in Sjaelland. This maximum number is obtained at the date of the 3rd of January 2021, when a peak number of 181 hospitalizations are estimated.

Finally, the plot starts to decrease after this first month of 2021 until it reaches a value of 30 hospitalizations, which is the value that it will oscillate around up to the 6th of May.

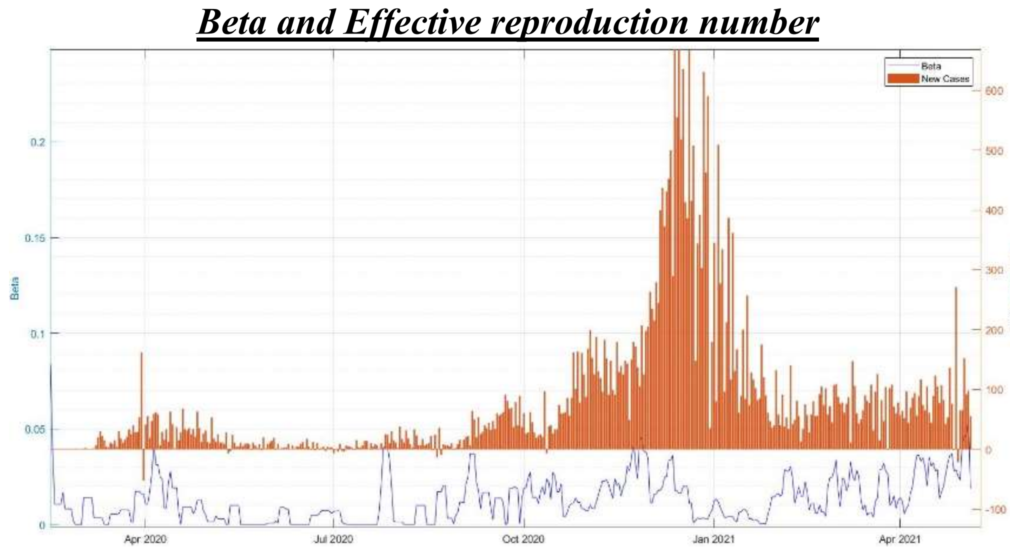

FIG. 29

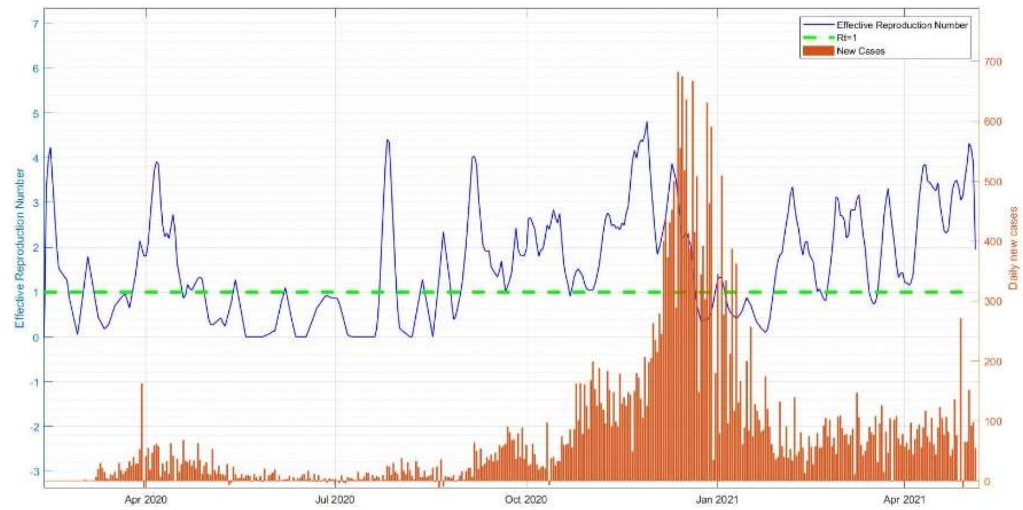

FIG. 30

In the first image, Fig. 29, it can be seen the comparison between Beta and the daily new Covid-19 cases in Sjaelland.

Beta starts to show an initial value of 0.08 when there are no new cases, but when these cases arrive Beta starts to oscillate around a value of 0.005 with the presence of two peaks that even get to 0.04.

When the cases substantially decrease, Beta goes down to the zone between 0 and 0.01 until August 2020. In this month, the cases start to grow again and so that Beta increase its value in order to fluctuate around the values of 0.015 and 0.03. This behaviour will remain until the beginning of the year 2021, including some peaks of almost 0.04. When the new year arrives, Beta again decreases to the same zone as it was in summer of 2020.

Finally, as the cases again increase at February, Beta increase accordingly, oscillating around the zone between 0.01 and 0.03 which is the zone in which it will stay until the end of the graph.

In the other hand, Fig. 30 shows the comparison between the Effective reproduction number and the Covid-19 cases that appeared daily at Sjaelland during the pandemic.

Most of the time, the Effective reproduction number seems to behave as the daily new cases plot. In other words, when the new cases increase the Effective reproduction number seems to increase too, and the other way around.

Despite of this, at the beginning of the plot Beta behaves in a rare way a couple of times, as it goes from a peak value to a very low value in a few days. Luckily, this behaviour is not constant, and when the important Covid-19 waves appear, it seems to move accordingly to how the cases vary. In fact, this behaviour can be clearly seen after the month of October 2020, where the Effective reproduction number remains above the green line until the time frame between January and February of 2021 appears, when there is a considerably decrease in the new cases.

This is the unique period where the Effective Reproduction Number gets to a value below 1 in the year 2021. At the arrival of March, the cases increase again, what makes the Effective reproduction number overcome the green line ubicated in 1 and oscillate around the values of 1 and 3.5 until the 6th of May 2021.

6.1.5 Syddanmark

Syddanmark, also known as the Region of Southern Denmark, is placed in the south of Jutland, bordering Germany.

It is the second biggest region in Denmark and at the beginning of the pandemic, it has a population of 1223165 people.

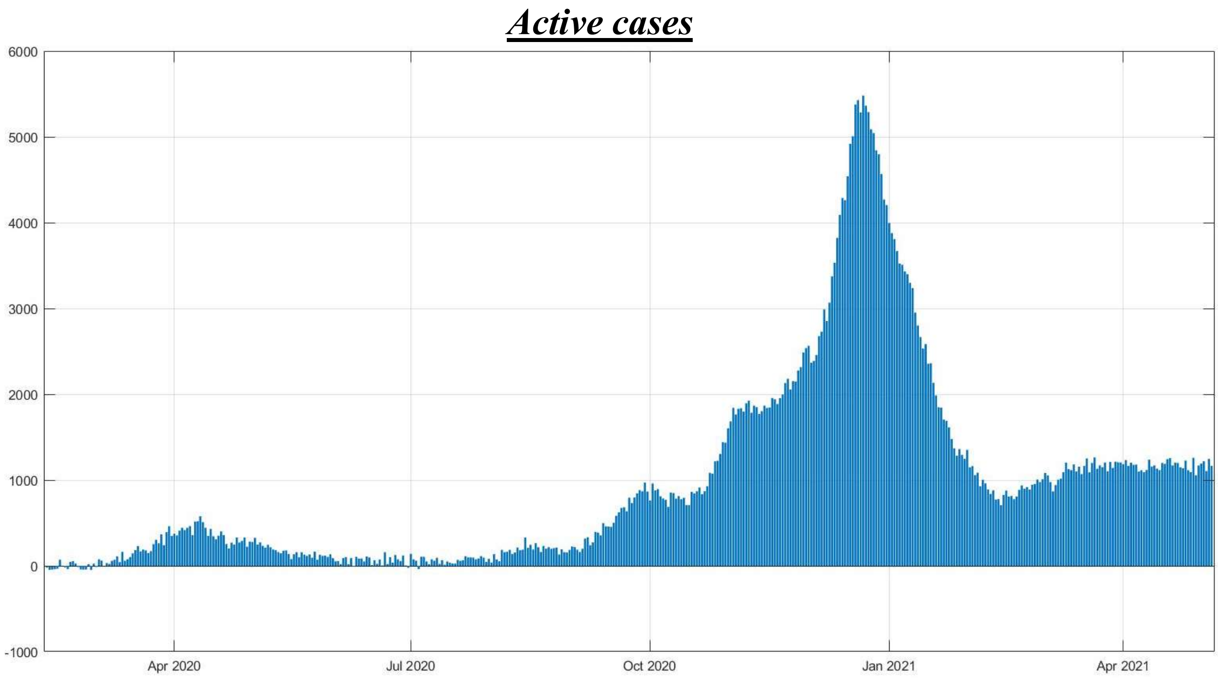

FIG. 31

Fig. 31 shows how the active Covid-19 cases evolved during the pandemic time frame.

Firstly, it can be seen that after April 2020, the bar plot reaches the first maximum of the graph. This maximum can be found on the first days of this month, and it reaches a value above 500 active cases.

After this, the graph decreases and maintains some tens of cases from June 2020 until August. Then, after this month, it starts to grow again until it reaches its top before the year 2020 ends.

This top is concretely reached on the 22nd of December when it is estimated that there were 5481 active cases at the same time in the Southern Region.

Finally, after reaching this top, the plot starts to decrease until February 2021, when there is a slight rise to finally oscillate around a value above of 1000 active cases, which will remain until the end of the figure.

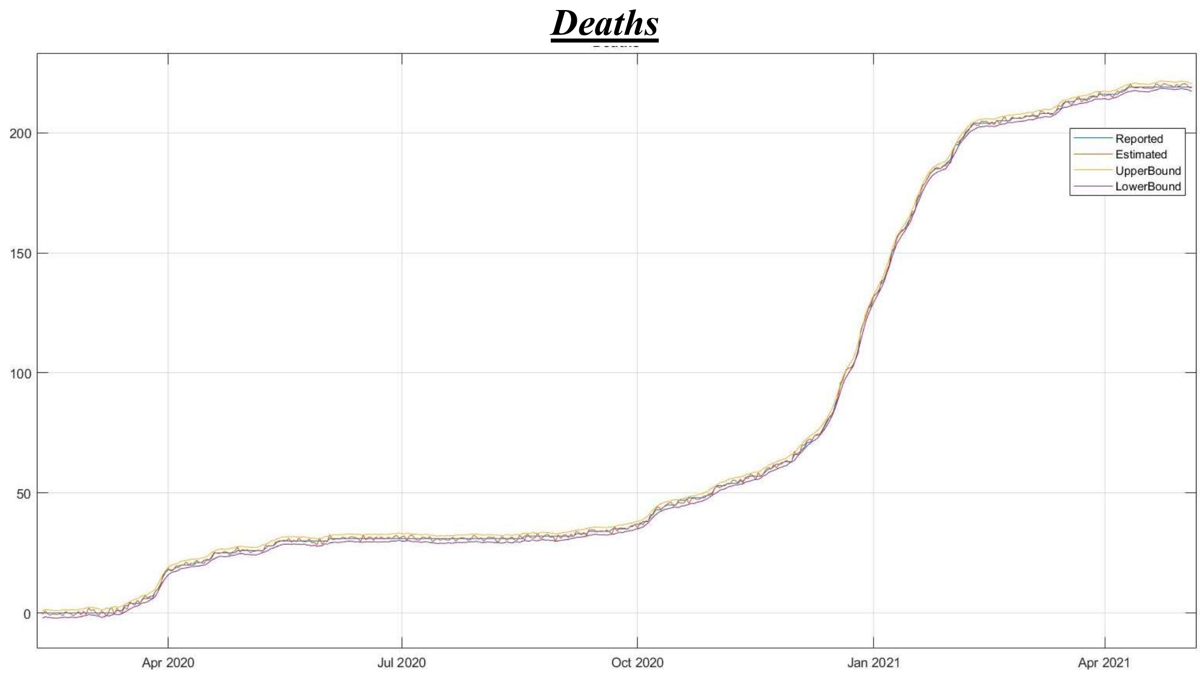

FIG. 32

In Fig. 32, it can be seen how the number of deaths increase during the pandemic in Syddanmark.

The graph needs like a month in order to have the first death cases in it, but then, after the first wave of Covid-19 which lasts until May 2020, it stabilizes around a value near 30 deaths. This value will be held until the autumn of this year arrives.

Then, the graph starts to grow exponentially until the beginning of 2021, when the number of 200 deaths is overcome. For the rest of the plot, it can be appreciated a slight uptrend to get into the final number of 219 deaths.

Regarding the shape of the plot, the red and the blue lines seems to be sticked most of the time. There are a couple of times where the estimated data fluctuates around this reported line, but as it rarely surpasses the confidence intervals, they can be considered as regular fluctuations.

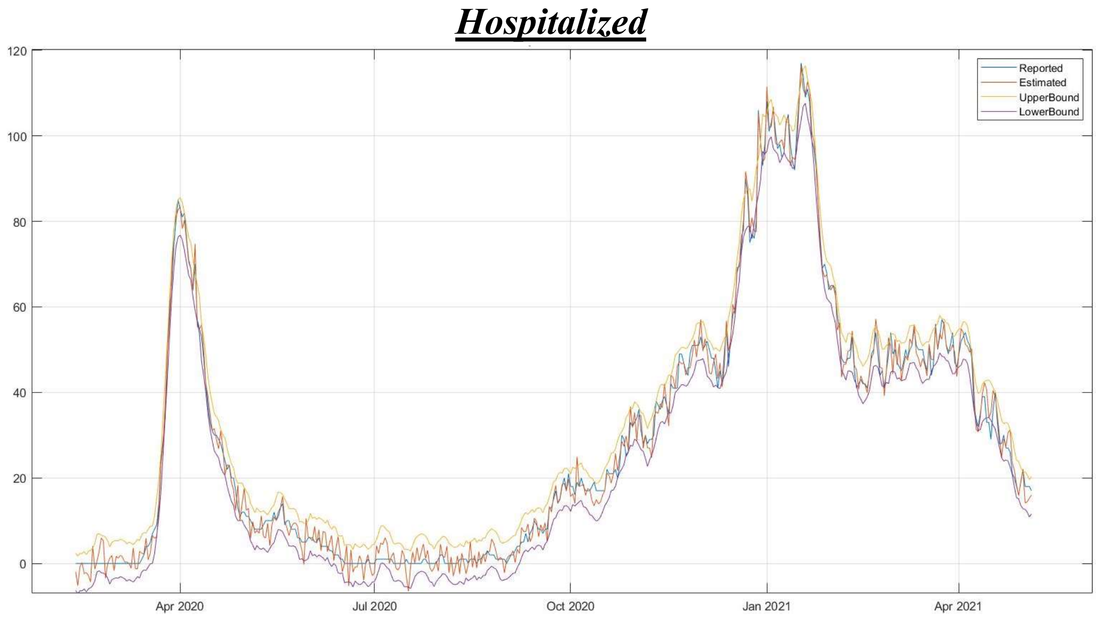

FIG. 33

Fig. 33 shows the number of hospitalizations that were held daily on Syddanmark during the pandemic.

At first sight, it can be seen that the red estimated line rarely follows the shape of the blue line. Nevertheless, it is always oscillating around it, and although there are some cases in which it surpasses the boundaries of the plot, the majority of the cases remains inside these 95% confidence intervals.

Regarding the development of the hospitalizations, there is a big increase at the end of March 2020 that makes the graph reach a local maximum of more than 80 hospitalizations. Then, the graph decreases until it gets near the value of 0. In this time frame, it can be seen that there are a couple of times that the estimated line goes below the 0 line. As there are only a few of these cases, and their size is small, they can be considered as negligible.

This behaviour is maintained until October of 2020 arrives. Then, the graph starts to grow quickly until it reaches its peak value. This value is obtained the 17th of January 2021, when it is estimated that 116 people were hospitalized in Syddanmark at the same time due to Covid-19.

Finally, the graph starts to decrease after February, reaching a value around 50 hospitalizations that remains for a month. After this, the graph starts again to show a downtrend that will stay until the 6th of May, when a value below 20 hospitalizations is estimated.

Beta and Effective Reproduction Number

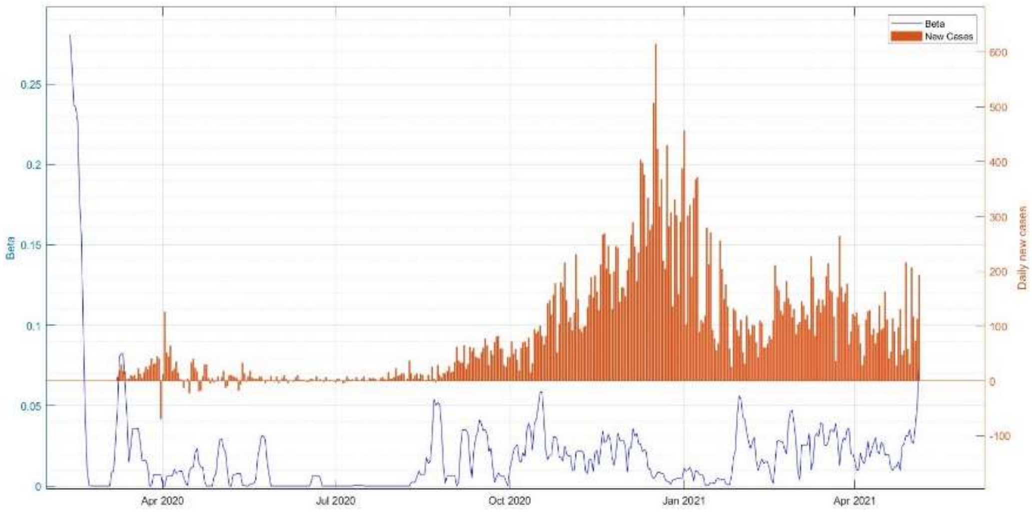

FIG. 34

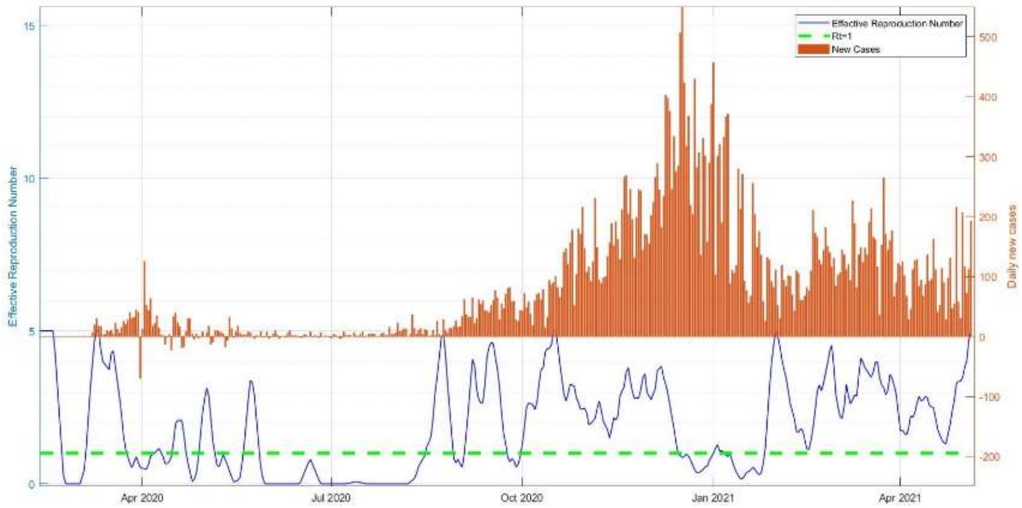

FIG. 35

In Fig. 34, it can be seen the comparison between Beta and the daily new Covid-19 cases that appear in Syddanmark.

This Beta has an enormous value at the beginning of the graph, when there are not cases yet. After having the first positive cases, Beta tends to go down and it starts oscillating between 0.005 and 0.03.

This behaviour is stablished until summer of 2020 where Beta gets close to a value of 0. But when this summer ends, Beta returns its way up, showing some peaks of 0.05 and fluctuating around the 0.01 and 0.03. This increase in the value of Beta comes also with an increase in the daily new cases.

The bar plot of daily new cases, as well as Beta, go down after the 2021 arrives and they get to minimum values, until February of the same year. After that, Beta will oscillate in the zone between 0.01 and 0.04 which is the zone that will remain until the end of the plot.

Summarizing, except for some peak values that are achieved during the entire plot, Beta seems to increase when the daily new cases increase, and the other way around.

In the other hand, Fig. 35 shows the plot of daily new cases compared to the behaviour of the Effective reproduction number.

This Effective reproduction number shows at the beginning of the graphs some peaks that goes from 0 to 5 in a couple of days. Then, after April, it starts to oscillate in a more reasonable way, between the values of 0 and 3. This routine is maintained until the summer arrives, when the Effective reproduction number goes almost to 0.

When the second wave of Covid-19 arrives, in September, the Effective reproduction number goes up again and crosses the green line which indicates the value of 1. Sometimes it gets to peak values of almost 5, but most of the time is oscillating around 1 and 4. This behaviour is maintained from October 2020 until the end of the graph, except from February 2021, when the decrease in the daily new cases makes the Effective reproduction number go again below the green line.

In general terms, the Effective reproduction number maintains a value above 1 when the bar plot is showing an increase in the daily new cases, and the other way around. Nevertheless, there is a presence of some peaks that makes the plot jumpy.

6.2 Prediction, errors and forecast for hospitalized results

As it has been shown in Extended Kalman Filter, the Extended Kalman Filter operates in 2 different steps. While the first step is mainly affected by the model and the previous step calculations, the updating step is based on the measurements and the calculated covariance. Then, the Kalman Gain would determine how much impact has these measurements on the predicted step.

In order to show how much impact has the Kalman Gain on the estimated results that it has been obtained previously, it has been plotted in Fig. 36 a comparison between the predicted step, the estimated results and the measurements.

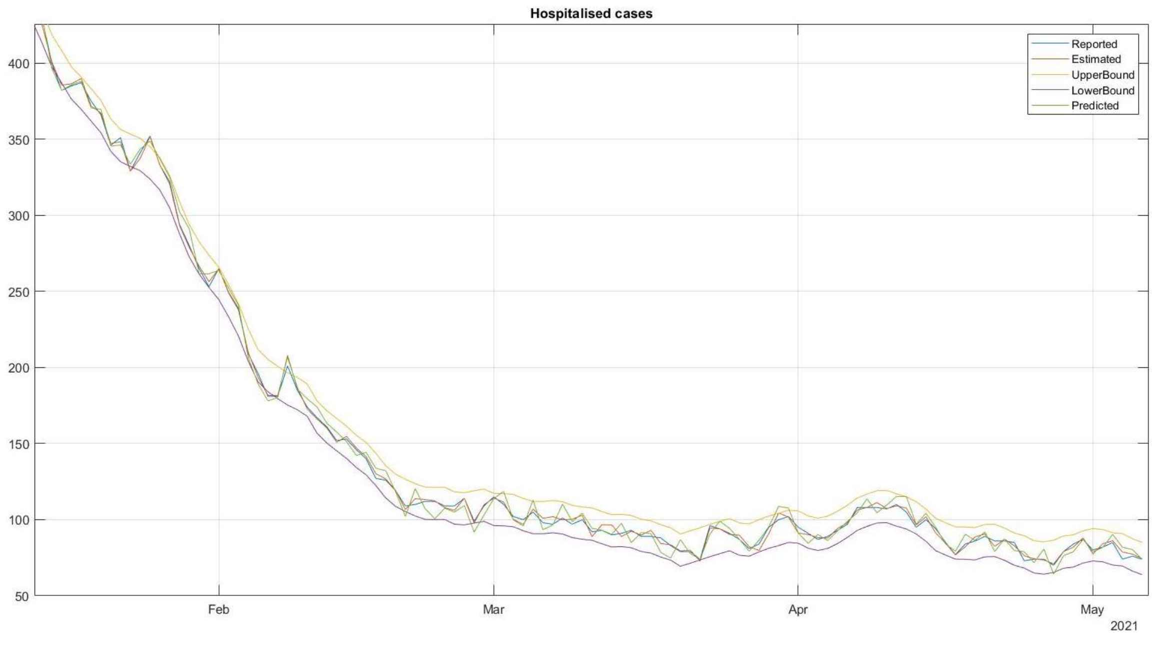

FIG. 36

In the figure, it can be seen that the green line that corresponds to the predicted plot is more abrupt regarding the fluctuations than the other two lines (estimated and reported). Nevertheless, it can be seen that this predicted line is not that far from having the shape from the estimated plot, as it differs just a bit from it. Furthermore, the majority of the predicted points are inside the confidence interval. This last fact can be better seen in Fig. 37 , where a plot of just the predicted results and the confidence intervals is shown.

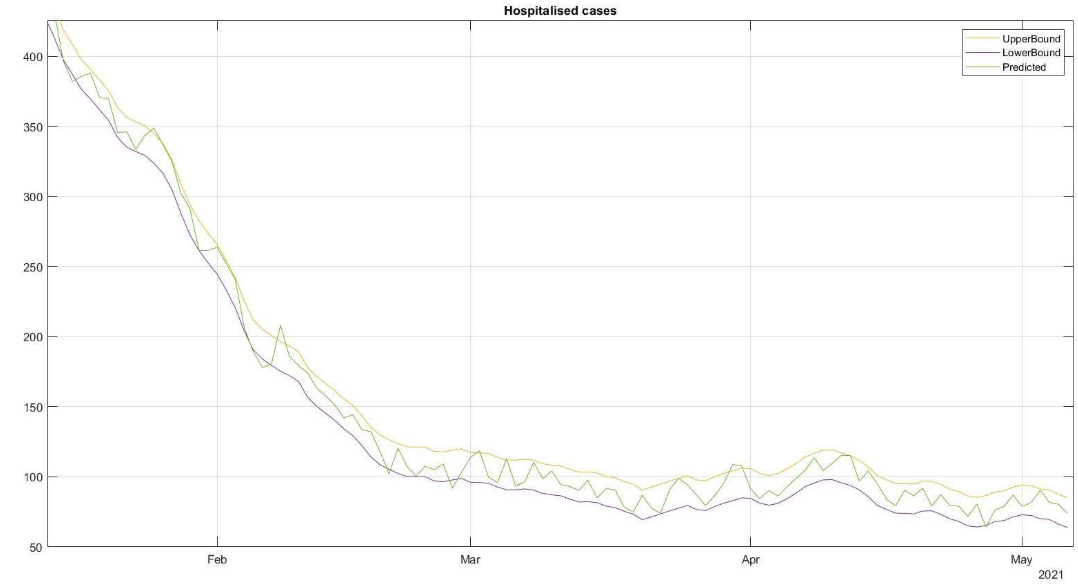

FIG. 37

For these two figures, it has been used the time frame from the 1st of January 2021 to the 6th of May of the same year, which also matches with the short time frame used in Short-term estimation.

The region that has been selected to take this additional test is Hovedstaden as it is the region that has been selected to show the results from the model states in Performance of the estimations.

In order to prove also that the difference between the estimated step and the predicted one is sensible, in Fig. 38 and Fig. 39, it is shown the absolute error from each of the curves respect the measurement data.

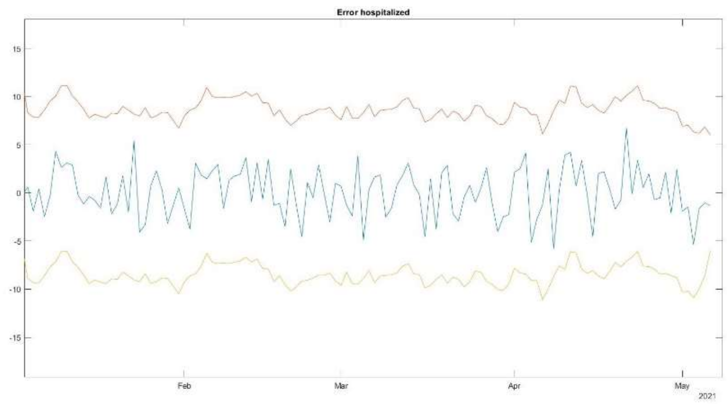

FIG. 38

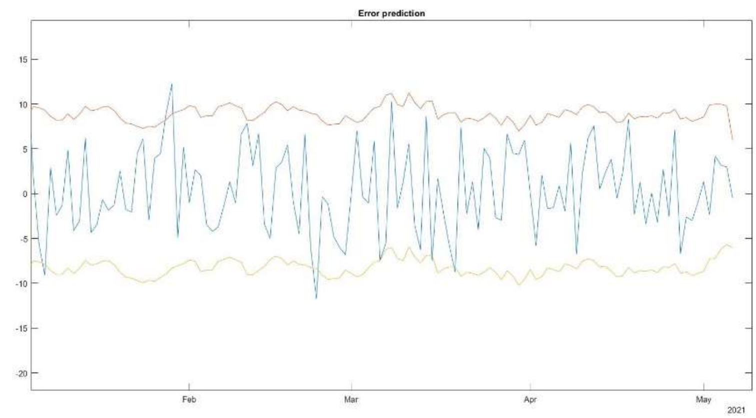

FIG. 39

In both figures it can be seen that the error is mostly bounded between the confidence intervals, but the error from the estimated hospitalized state is smaller than the one from the predicted plot. Nevertheless, this second graph show that just a couple of samples exceed the boundaries.

As it has been shown that the prediction part of the algorithm works in a similar way than the estimated state results, it has been decided to plot a single test where the prediction step of the hospitalized results is plotted, with the addition of a 40-days forecast at the end of the graph. In this forecast, it has been excluded the measurement data from the Danish Authorities.

This plot is shown in Fig. 40:

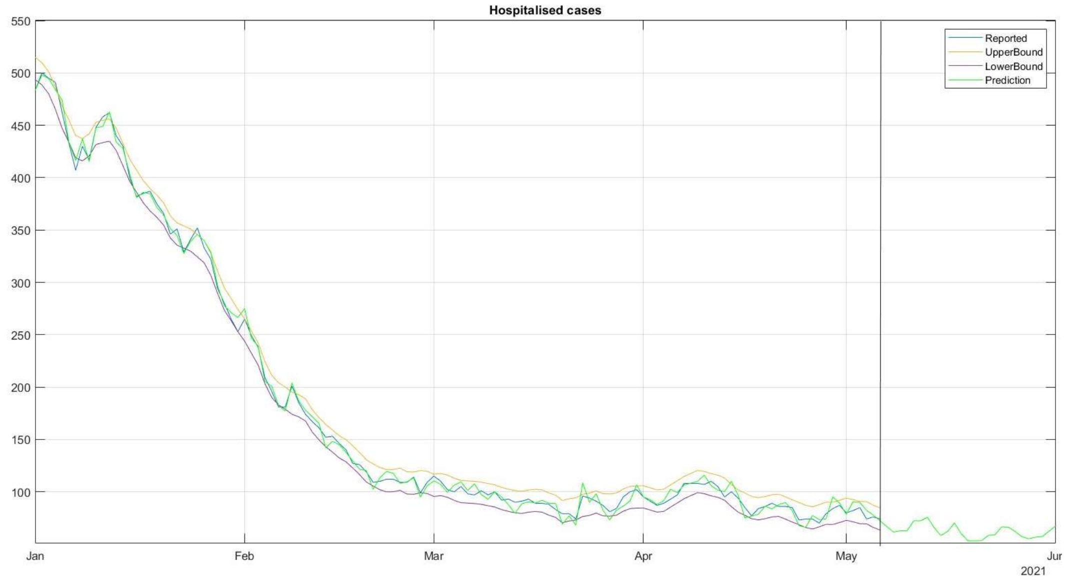

FIG. 40

In the mentioned figure, it can be seen the performance of the prediction results against the measured data from the Danish authorities. Once it gets to the black line, it means that there is no data available for the following dates, showing a pure forecast from the model. The time frame for this forecast goes from the 7th of May 2021 until the 15th of June 2021.

In this forecast it can be seen that the downtrend, that the prediction plot had, is continued, while it fluctuates between values around 50 and 80 cases. So, according to the model presented in Final model, the region of Hovedstaden would have this quantity of hospitalized people during these mentioned dates.

It is worth saying that if Fig. 10 is observed, there is a stabilization in the number of active cases at the end of the graph, which matches with the time frame that is immediately before to the one that has been forecasted. This implies that the number of hospitalizations should also stabilized at this time frame, ending in the conclusion that the forecast can be acting properly.

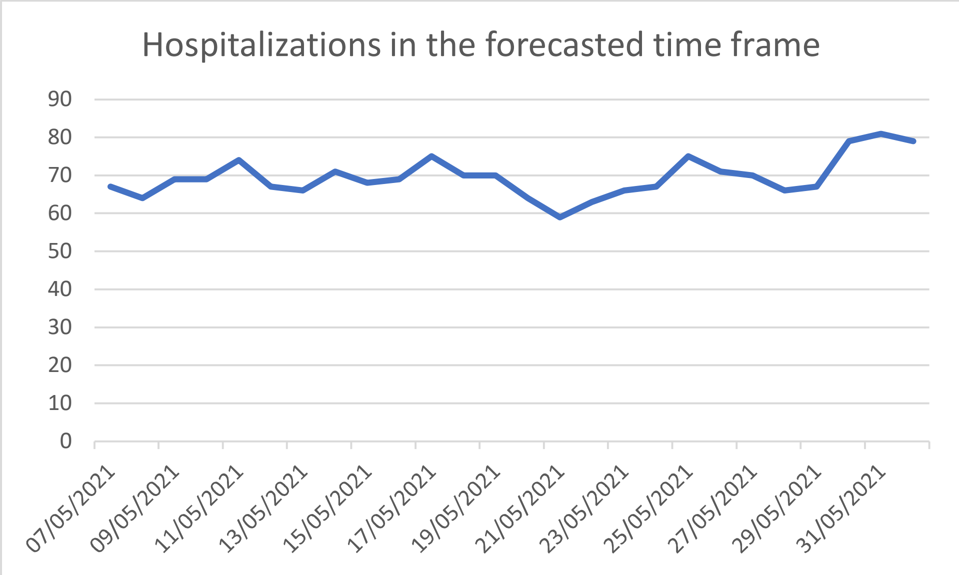

FIG. 41

Moreover, Fig. 41 shows the hospitalizations that has occurred during the first part of this forecasted time frame. Concretely, it starts on the 7th of May and it ends in the 2nd of June.

In the figure it can be seen that the value for the hospitalized people oscillates around the value of 70. As it has been said before, the forecast expects for these days to fluctuates between the values of 50 and 80, so although it is not showing the same shape as this last plot, it can be considered as a good approach of how the hospitalized cases are developing in Hovedstaden at present.

7. Discussion

In this section of the report, it will be discussed the results shown in Performance of the estimations and in Results. Furthermore, there are some other topics that will be included in this section, like the choice of the variance of each state, or the validity of the data that has been used to estimate the model.

7.1 Validity of the data

As it was said on Model´s estimation, the data that has been used as a reference in this report comes directly from the Danish Authorities. For that reason, it has been assumed to be sufficiently reliable to be used in this Master´s Thesis.

However, there is some information about a couple of states that was impossible to find. In consequence, the information that has been used for these two states is a linear combination of the rest of the data, or a sensible manipulation of it. Concretely, these two states are the susceptible and the exposed one.

The first case, the susceptible state was obtained by a linear combination of the rest of the groups. For example, if Hovedstaden has 5 reported new exposed people, this people will be removed from the susceptible group and will be included in the exposed one.

However, for the exposed group, it was not that easy to obtain something sensible. In fact, there is no way to measure how many exposed cases are there, as the majority of them do not know that they have been infected by the virus. Thereby, an assumption was made. As the incubation period for Covid-19 has been set to a value of 14 days in this report, then the new Covid-19 cases has been displaced fourteen days to the past, forming the exposed group.

As the incubation period can vary, and it is known that every person incubates the virus from 2 to 14 days, this exposed state will not act as the rest, accumulating the people that is inside the group as it will disturb the data even more. It will just give the number of new Covid-19 infections, 14 days before they are known, assuming that all of these patients suffer the largest incubation period.

This is the reason why the active cases plots that were shown in Results, have a lower value than the expected one. Although the infected and the hospitalized cases are summed in these plots, it should also include the exposed cases. However, as there is no data on how this state evolves, only the new exposed cases were summed to the active cases plot, ending in a lower number than the one that it should have.

Furthermore, this is also the reason why the Beta parameter and the Effective reproduction number plots shown in Results are displaced fourteen days from the bar plot. As Beta appears in the equation of the exposed group, this Beta will increase its value when new exposed cases appear. This increase will be produced 14 days before the case is detected, ending in a disagreement between the graphs, as Beta acts 2 weeks before the bar plot consonantly moves. Moreover, as the Effective reproduction number depends directly on Beta, this case can also be applied to it. Anyways, the results obtained from these two numbers are going to be further analyzed during this section.

7.2 Election of the white noise´s variance for the measurement´s noise

The measurements that have been extracted from the Danish Authorities has been considered as reliable. Nevertheless, in this kind of data, there can be errors or failures that can vary the sample of data. These failures can be a consequence of human errors as there can be a typo when the data is being inputted in the system or a case that is not registered at the time it has been validated; or they can also be caused by testing errors, such as a false positive or negative in a Covid-19 test.

For these reasons, a measurement noise appears in the system. This noise is considered white and Gaussian with zero mean and a variance that is calculated in terms of the factors that can affect the sample of data that has been extracted. These variance´s results can be seen in the covariance matrix R, in (Eq. 33).

As it has been said, these values are not chosen at random, they have been calculated among different factors that can vary the sample of data for each of the states. These factors are presented here:

- The susceptible value will directly depend on how the Covid-19 tests performs. Moreover, there is a number of cases considered as exposed, that mostly don´t know that they have left the susceptible group.
  For that reason, the susceptible state will have an uncertainty of 1000 cases.
- The infected state has data extracted directly from the Danish Authorities. However, as there is a small probability of being a false positive in a Sars-Cov-2 test, and also there can be people that is not infected anymore but they do not know about it yet, the value for this variance has been set to 100. As the infected state is directly linked in the model with the exposed state, this value of a 100 is also chosen for the exposed sample of data.
- The case of the hospitalized and the dead states are the same. As they are built with reliable data extracted from the Danish Authorities, the cases should not vary a lot for them. However, there is a possibility that the data is collected some days later or that a number is not registered at the moment that it happens. For example, there could be a person that leaves the hospital in the Wednesday´s afternoon, but this absence is not registered until the next day. Consequently, the values for the variance of these two states are low but higher than 0.
- The recovered and the vaccinated states share the same explanation. A person who is recovered from the virus knows it because its proven negative in the Covid-19 tests. There can be errors in these tests, but this is not the main reason why this high value has been chosen. Both states are modeled with a variable that allows the people that are inside this group, to lose their immunity and go back to the susceptible group. As these changes are not registered in the official data, this variance parameter has been set to a value of 100, just in case some people lose its immunity, and it has not been aware of it.

7.3 Short-term vs Long-term estimations

In Performance of the estimations, it can be seen how the model performs when an estimation of its parameters is made. In this case, it has been chosen to make two different estimations, one for a short time frame (from the 1st of January 2021 until the 6th of May 2021) and the other one for a longer time frame that, in fact, includes the short timeframe too (from the 12th of February 2020 until the 6th of May 2021).

At first sight, seeing the results for the different states, it can be determined that the model acts correctly in both estimations. In the plotted states, the estimated line seems to catch the reported line or at least oscillate around it. Furthermore, to prove the validity of the results, it has also been included a plot of the 95% confidence intervals, that are surpassed only a few times. As there is a 5% of the data that is expected to be outside these boundaries, the fact that every plot has some cases that goes out of these lines, also shows a good performance of the model.

However, both of the estimations have its downsides.

The long estimation is considered to be the best choice for this project, as it can be seen how the model reacts to the entire pandemic time frame. Despite of this, it lasts around thirteen more times to generate the plots, than the shorter one. So, in the case that a result is needed fast, this long estimation is not a good choice.

In the other hand, the short estimation is finished in almost a minute. The model performs really well as in the other one and it shows clearly, without the needing of any zoom or an approach in the images, the development of the states. Even so, if the behaviour of the Covid-19 pandemic wants to be studied, this short estimation is not a good choice. Or at least, it is not a good choice if this time frame is chosen.

The Covid-19 has two significant waves that impacted Denmark in a severe way. The first of them happened around the months of March and May of 2020 while the second, and the most important one caused mainly by the British strain, goes from October of 2020 until the beginning of 2021. Both of these waves are not contemplated in the short estimation, so if there is a need of taking the peak value of any of the states, as aims some of the outputs of this report, it is not a good choice to pick this estimation, as for example the peak number of hospitalizations happened on December of 2020.

Furthermore, one of the main reasons why the short estimation was done, was to prove how the vaccinated state performs. As in the long estimation it has been proven that the vaccine value remains at 0 until the first vaccines are administered, there is no need to have a short estimation for this particular reason.

To sum up, as both estimations seem to have good results, in this report, the long estimation has been preferred for building the Results which shows the estimated results in each of the different regions.

Nevertheless, in the case that a prediction wants to be done, the short estimation would be a better choice, as the elapsed time to complete an estimation for the entire pandemic is much bigger.

7.4 Performance of the states

In Performance of the estimations, some plots are given in order to show the behaviour of some of the model states. Concretely, there are four states that are shown in this section: susceptible, infected, recovered, and vaccinated.

Although some of the results are focused on the hospitalized cases, the behaviour of these states is also important in order to prove the validity of the model. In fact, the model that is presented in this thesis has been a challenge, as it is trying to improve the already known models like the SIR or the SEIR model.

Moreover, the majority of these states have been created or modelled in a different way that in the common mathematical models of infectious diseases. For example, in the susceptible model it has been added a term that can add people from the recovered and the vaccinated states. So that, it was important also to show how the states behaves to these changes.

7.4.1 Susceptible

The susceptible state is the initial state where the people are located at the beginning of the pandemic. In this particular model, shown in Final model, it has been built as a state with an output of people which goes to the exposed group; but also, it allows people to go again inside the state, as it would mean that they have already lost its immunity.

As it can be seen in Fig. 2 and Fig. 6, the susceptible state is constantly showing a downtrend in the plots. However, there is an important fact to be taken into account, the scale of the plot. The plot has a scale of 10^6 meaning that the oscillations of tens or hundreds of people are not significantly varying the plot.

As this state could suffer from negative testing, or people that is going back to it due to the loss of immunity, the variance of this state was chosen to be high, in order to handle these events. Concluding that, the low-size oscillations that can be seen in the figures are provoked by the variance that has been set for this state, ending in the conclusion that the state is behaving as expected.

7.4.2 Infected

The infected state has been included in order to show how the development of the pandemic affected Denmark. Concretely, it has been shown for the region of Hovedstaden which is the Capital Region of the Danish country.

Generally, the infected state seems to behave nicely as the estimated infected people is almost always joined to the reported cases plot. Furthermore, the boundaries that are also plotted, make the state more confident, as the estimated line rarely cross these confidence intervals.

So, concluding, the infected state allows the reader of this thesis to see the estimated main time frames where Denmark was heavily affected by the Covid-19 pandemic, with a good and reliable result that matches the real events.

7.4.3 Recovered

The recovered state of the model is characterized by both θ parameters that indicates the rate of people that have recovered from the infected state and from the hospitalized state, respectively. In addition, there is a parameter ε_r that shows the rate of people that loses its immunity and goes back to the susceptible group.

At first sight, the recovered plots that can be found on Fig. 4 and Fig. 8 have also a big scale, what could lead into the conclusion that it would be hard to see how the state is behaving, as it happened with the susceptible group.

Despite of this, in both figures, it can be clearly appreciated that the estimated state is varying its value around the blue line, but it never stays sticked to the reported line shape. It can be considered that these fluctuations are caused by the variance of the state, as this variance was built in order to cope with the constant inputs and outputs that the state would suffer according to the model, so it can be concluded that this state is also acting well, although is the one showing the highest difference regarding the reported measurements.

7.4.4 Vaccinated

The vaccinated state was considered as the most unpredictable state of the model.

In real terms, these vaccinations appeared almost a year after the pandemic started, so the performance of this state in the first year of estimation was unknown.

Fortunately, as it can be appreciated in Fig. 5 and Fig. 9, the state remains near the value of 0 until the first vaccine is administered on the 13th of January 2021. Until this day arrived, the estimated plot is fluctuating with a couple of hundreds of cases around the blue line, what can be considered a result of the variance.

As it happened with the recovered state, this state is also built with a tiny probability of losing people in benefit of the susceptible group. Moreover, the input of this state is built with a constant number, what would mean that every day is inputting to the state a fixed number of vaccinated people. These two facts were taken into account when the variance of this state was considered, so that it can be concluded that according to the figures, the vaccinated state is performing well in the entire pandemic timeframe, despite of these small oscillations

7.4.5 Dead

The development of the dead state can be appreciated in five figures inside the Results of this report, each of them for a different region.

At the beginning of the graph, the dead state plot seems to be working nicely, as it doesn´t cross below the zero value. Furthermore, in all the plots, the estimated plot for death people is showing constantly an uptrend, what is also a good sign that is behaving well.

Regarding Fig. 12, which is an approach of the previous figure that shows the development of Covid-19 deaths in Hovedstaden, it can be seen that the red curve sometimes tends to make tiny decreasing movements. These moves can be considered as negligible, as they have a difference of decimals between the estimated value and the reported value.

In fact, these differences are a consequence of introducing the variance into this state, as it was previously mentioned on Covariance matrices.

Summing up, it can be affirmed that the dead state is working well through the entire pandemic time frame, as the plot is really similar to the one that is reported, and there are just a few cases where the estimated plot crosses the confidence intervals.

7.5 Beta and the Effective reproduction number

One of the most important parameters inside the model exposed in Final model, is Beta. This parameter shows the rate of people that goes from the susceptible group to the exposed one, in other words, it tells how fast the virus is spreading among the population.

For that reason, in this report it has been chosen to estimate the Beta parameter, instead of assuming an initial and unique value for it. Although, it is a parameter and not a state, in order to estimate its value, Beta has been considered as a separate state instead of being a normal parameter. However, this fact implies that Beta can behave in a more unexpected way than if it has a fixed value.

After seeing the plots that have been given in Results showing the estimation of Beta for each of the different Danish region, it can be determined that Beta is not behaving as well as it was desired. In fact, in the majority of the pictures it can be seen that at least for a period of time, it shows values that can be considered as impossible in the real life.

Despite this, there is a main reason that can explain the behaviour of Beta, and it is directly connected with what is mentioned in Validity of the data.

Beta is the parameter that measures the input of data to the exposed group. As it has been explained in that previous section, the exposed group has been modelled in an atypical way due to the lack of data for this state. Although the Beta plots are not showing enough good results, there is a positive fact that can be remarked in them, which is that Beta follows the attitude of the daily new cases, what means that if the daily new cases are increasing, Beta will do the same movement and the other way around. Despite of this, this follow-up movement is not shown in the same time frame as in the daily new cases, in fact, it can be seen that Beta anticipates this move 14 daus.

This anticipation can be considered as a consequence of the incubation period´s modelling for the exposed group. In other words, Beta increases when the people enter the exposed group, while the daily new cases increase when the people are entering the infected group. Consequently, there is a gap between both movements of 14 days that directly affects the final results of Beta.

Nevertheless, taking as an example Fig. 19 which is the Beta plot for Midtjylland, there is another assumption that can also be made after analyzing its behaviour.

Beta doesn´t seem to behave well when there is an abrupt change in the daily new cases. In fact, Beta performs better when the cases have high values that are varying progressively than when the values are close to 0 and all of a sudden it appears a bar with a much higher value than the previous. Consequently, the plot of Beta has a better performance when the time frame of the second Covid-19 wave arrives than in the beginning of the pandemic.

Although in this second wave period, it can also be seen some peak values that exceeds the real-life expectations, Beta shows a smoother adaptation to the changes than in the beginning of the plot where it can go directly from a value near 0 to 0.05 in just a day because of a single unexpected grow in the new cases.

The Effective reproduction number is heavily linked to the behaviour of Beta. In fact, the model of this number depends directly on Beta, as it is the only parameter in its equation that is constantly changing its value. So, it can be assumed that the changes that can experience Beta during the plot, would approximately be the changes that the Effective reproduction number will suffer.

In fact, comparing the plots of Beta and the Effective reproduction number that were shown in Results, it can be concluded that there are a lot of similarities between them.

Furthermore, for these Effective reproduction number´s plot it has been added a green line that identifies the value of 1. This is a critical value for this number, as it is the frontier that indicates whether the spread of the virus is being reduced or not.

In general, the plots for this number are showing a sensible response to the daily new cases, as the plot seems to be in values above 1 when the infections are increasing, while it gets to values below the green line when the new cases are decreasing or close to 0. Despite this, the Effective reproduction number is closely affected by the behaviour of Beta as it was mentioned before, what ends in the conclusion that the unexpected changes that Beta is showing during the majority of the plot, can also be appreciated in the Effective reproduction number´s plots.

In addition to this, there is also another fact that negatively affects the results for this number. The Effective reproduction number equation that can be seen in (Eq. 16) shows that apart from Beta, it depends on two other parameters (Theta and Gamma).

These two parameters are considered as fixed parameters in the report, what implies that its value is given at the beginning of the estimation, and they are not progressively changing it, as it happens with Beta. This leads to the assumption that has been made in the previous paragraphs, which is that the Effective reproduction number would depend in a much major way on Beta than in the other two parameters, because as they are not changing its value, they are not going to change the behaviour of the resulting number either.

Consequently, the Effective reproduction number seems to have a similar performance as Beta, what implies that it behaves better when the daily new cases are high than when the new cases are at low values. In fact, if the plot is analyzed in this period of time (the second wave), the Effective reproduction number is oscillating between the values of 1 and 4, which are sensible values for that period of time, as the Covid-19 was spreading fast.

To sum up, Beta and the Effective reproduction number are not showing results that can be considered as perfect, due to these modelling problems. Nevertheless, they seem to behave accordingly to the development of the virus, so although their results can be improved solving these problems that have already been mentioned, they can be considered as a good approach of how they should behave.

7.6 Hospitalized state: results, error tests and forecast

As it was remarked in Problem statement where the problem proposed for this report was stated, the hospitalized state has a greater importance than the rest of the states.

One of the objectives of this report is to be able to estimate and predict the number of hospitalizations that has been achieved during the pandemic, in order to prevent and secure hospitalized resources for all the patients. This could also be important in the case that a hypothetical new wave of the virus appears, as the Covid-19 is already present in the society.

For that reason, after showing that the rest of the groups are behaving properly in the model, it has been made a couple of more tests for this specific state, in order to fully secure the validity of its results.

Before that, it can be seen in Results per region the different plots for the hospitalized state in the five Danish regions. These plots are characterized by the presence of the reported plot, the estimated plot and also the boundaries for the two of them. In all of them, it can be agreed that the reported plot and the estimated plot fulfils the 95% confidence intervals, while both plots (reported and estimated) have similar shapes and follow almost the same path.

This can be concluded as a good estimation for the hospitalization cases that happens in Denmark towards the pandemic. Moreover, as these results are considered as a good approximation to the reality, it has been chosen to make some other tests in order to see how the state behaves when a forecast is done.

These tests are located in Prediction, errors and forecast for hospitalized results, and firstly, they show the impact of the Kalman Gain in the prediction step in contrast with the estimated plot. This can be seen in Fig. 36, where it can also be seen the reported plot and the boundaries. Both, the estimated and the predicted plots seems to behave in a similar way, but there can be seen some difference between them as sharper edges or higher fluctuations.

Consequently, there has been made a test showing the absolute error against the reported data from the two plots. These results are shown in Fig. 38 and Fig. 39, where it can be appreciated that the difference between the absolute errors from the predicted and the estimated plots are just of a few units, which indicates that a priori the predicted step of the model is not that far from the estimated results. As it can also be seen that the error´s plots are mostly inside the boundaries, this predicted step has been assumed as sufficiently reliable for making a forecast without measurement data.

It´s worth mentioning that the measurement data goes into the Extended Kalman Filter in the updating step, as it was shown before in Updating step, so in order to forecast, it was needed to check firstly the behaviour of the predicting step and the influence that the Kalman Gain has on it.

When this viability was proven, the forecast was made, showing the plot that can be seen in Fig. 40. This plot shows the evolution of the hospitalized cases in Hovedstaden during the 2021 dates of the pandemic, while it also shows a forecast of the next 40 days, starting from the last day when data was available (6th of May). This forecast seems to fluctuate between reasonable values of 50 and 80 cases, which can be assumed as possible because of the downtrend that is continuously following.

Other reason that helps to make this forecast more reliable is that the graphs that saw the daily new cases and the active Covid-19 cases were stabilizing its values at the end of the plots, where the forecast time frame was approaching. Seeing that the forecast is also stabilizing or at least fluctuating around the same or smaller values than the previous dates is also a good notice for it.

In fact, Fig. 41 shows that the forecast is not that far from the reality. This figure plots the hospitalizations during the first part of the forecasted time frame (7th of May until 2nd of June) in Hovedstaden. Comparing both results, the forecasted and the reported one, it can be seen that both plots are fluctuating around the same values, although the forecasted one reaches lower values than the other. Nevertheless, it can be considered as a good forecast as it remains in the same range of values for the first 27 days of forecast.

8. Conclusion

In this chapter, a conclusion is made, based upon the problem statement written in Problem statement . For the fulfilling of that statement, an innovative model for the Covid-19 has been created and its results have been presented during all this report.

Regarding the model, it can be concluded thanks to the results shown in Performance of the estimations and Results, that each of the states is working nicely, giving sensible results that are mostly similar than the ones that has been extracted from Danish Authorities. Nevertheless, it can also be concluded that the process noise modelling can be improved in further works, as it can be seen that it makes the results fluctuate more, when the population of each region decreases. What means that the process noise is well designed for Hovedstaden region, but it can be a little higher for the regions with less inhabitants.

With regard to the Extended Kalman Filter, it has been proved that is a good estimation algorithm for the problem that was stated at the beginning. It has shown a good response against the process noise, ending in estimations that are very near the reality despite of the non-linearities and difficulties of the modelling. The worst part of this Extended Kalman Filter results has to do with Beta, but as it was discussed in Beta and the Effective reproduction number, Beta has a better performance when the values are high than when the values are close to 0. Nevertheless, these problems can also be a consequence of the modelling of Beta, since estimating it as a state when the rest of the parameters are fixed, can cause disturbances in the results.

The hospitalized state has been also successfully estimated for each of the five Danish regions. It seems to perform well against the reported results, and it also seems to be showing good results when the state is predicted instead of being estimated.

Furthermore, for this state it has also been given the maximum estimated number of hospitalizations that has been produced in every region during the pandemic. These numbers are very close to the number of hospitalizations that was reported by the Authorities on those exact dates, so it seems like the estimation is also acting well regarding abrupt changes in the plot or peak values.

Last but not least, in order to prove that the model, although it needs some changes that has been commented, is going in the right direction for being used in a hypothetical future in real-life tasks, it has been added a 40-days forecast that shows a sensible result for what it is happening in the Hovedstaden region at present.

Summing up, in general words, the model and its written report catch up with the requirements that were stated in the problem statement, showing that is performing well in each of the key points of the initial project´s problem.

Nevertheless, as it has been mentioned in this conclusion, this project can be improved by the resolution of a couple of problems that have been appearing during its development. These problems are going to be stated in the next section when a couple of possible improvements for this report are going to be suggested.

9. Further development

During the development of this project, it has been shown that there are some decisions that have been made or problems that have appeared, that in the case of being solved, they should improve the results obtained in this project.

These are some proposals to be taken into account in further works related to this report:

- **The estimation of Beta:** as it has been said a couple of times through this report, Beta is not behaving as well as it can be expected. A proposal for its improvement is the estimation of the rest of the parameters of the model. It can be a complex task, since there are a few different parameters in the project, but leaving Beta as the only one estimated, leaves to inaccuracies that can be spread to other results as it happened with the Effective reproduction number. Another possible solution could be estimate Beta maximizing its likelihood function, applying the maximum likelihood estimation method.

- **Estimating a state rather than searching measurement data for each of them**: this can be applied especially to the exposed state which have been very hard to find any data that could be applied for it. In that case, the improvement proposal for this problem is to directly estimate this state, as it has been done with the Beta parameter. This can be applied thanks to the deterministic behaviour of the given model, what allows the estimation of the different states just knowing the data from one of them.

- **The adaptative adjustment of the process white noise´s variances depending on the development of the states:** the results obtained in this report could have been more accurate if the process noise were modelled according to the changes that the plots are experimenting. That means, for example that regarding the population, is not the same talking about an uncertainty of 5 in a group of people of a 100 than in a group of 500. This adjustment could guarantee that the results are smoother and more reasonable for samples of population that are lower, while it can also solve some unusual behaviours that have been seen on the plots, like the negative values for the hospitalized cases or the dead plot showing less cases than the day before. By the way, these are performances that were expected to happen, but that should also be improved with this proposal.

References

1. **Wikipedia. https://en.wikipedia.org/wiki/COVID-19.**

2. The Mayor.eu. https://www.themayor.eu/en/a/view/denmark-goes-on-lockdown-4549.

3. Worldometer. https://www.worldometers.info/coronavirus/country/denmark/.

4. Weiss, Howard. *The SIR model and the Foundations of Public Health.* s.l. : Departament de Matemàtiques de la Universitat Autònoma de Barcelona, 2013.

5. B. Ivorra, M.R. Ferrández , M. Vela-Pérez , A.M. Ramos. *Mathematical modeling of the spread of the coronavirus disease 2019. (COVID-19) taking into account the undetected infections. The case of China.* Madrid : Research Gate, 2020.

6. Anadolu Agency. https://www.aa.com.tr/en/latest-on-coronavirus-outbreak/recovered-covid-19-patients-lose-immunity-fastuk-study/2021112.

7. Jones, James Holland. https://web.stanford.edu/. https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf.

8. Simon, D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* s.l. : John Wiley & Sons, 2006.

9. https://ocw.mit.edu/. https://ocw.mit.edu/courses/aeronautics-and-astronautics/16-90-computational-methods-in-aerospace-engineering-spring-2014/numerical-integration-of-ordinary-differential-equations/discretizing-odes/1690r-the-forward-euler-method/.

10. Morrel, Darryl. https://www.cs.cmu.edu. https://www.cs.cmu.edu/~motionplanning/papers/sbp_papers/kalman/ekf_lecture_notes.pdf.

11. Dieter Fox, Sebastian Thrun and Wolfram Burgard. *Probabilistic Robotics.* s.l. : MIT Press Ltd, 2005.

12. Authorities, Danish Health. https://www.sst.dk/en/English.

13. https://www.businesstoday.in/latest/trends/20-30-recovered-people-lose-natural-immunity-against-covid-19-in-6-months/story/436312.html.

14. https://www.elconfidencial.com/tecnologia/ciencia/2021-02-17/inmunidad-natural-adquirida-coronavirus-vacunas_2953835/ .

15. https://www.reuters.com/article/us-health-coronavirus-denmark-idUSKBN2AO2GW.

16. https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus#:~:text=The%20time%20from%20exposure%20to,before%20starting%20to%20experience%20symptoms. .

17. A. Hasan, H. Susanto, V.R. Tjahjono, R. Kusdiantara, E.R.M. Putri, P. Hadisoemarto, N. Nuraini. *A new estimation method for COVID-19 time-varying reproduction number using active cases.* s.l. : Research Gate, 2020.

18. https://www.thelocal.dk. https://www.thelocal.dk/20200227/breaking-denmark-announces-first-coronavirus-case/.