Within our project we have many datasets providing the number of new cases of Covid19 for the different countries within the world. All datasets do have one row per day, the beginning of counting the cases differs from country to country depending on when the first occurence was detected.

Predicting the development of cases in each of the countries is of high interest. As all countries do have different developments within their cases each prediction has to be made for each country. We could choose different models for each country or we choose a model which allows to adjust the model parameter according to the different behaviour.

The model should be able to handle increasing and decreasing of new cases in two different ways.

We chose the SARIMA model (Seasonal Auto Regressive Integrated Moving Average), which can handle seasonal effects. The SARIMA model will analyse by statistical methods the trends and the seasonal effects a pandemie like Covid19 consists of. SARIMA consists of two parts, the first, ARIMA, tries to figure out how a stationary time series can be described by functions and parameters, the later integrates the seasonal effect.

Usage of SARIMA to predict development of a time series:

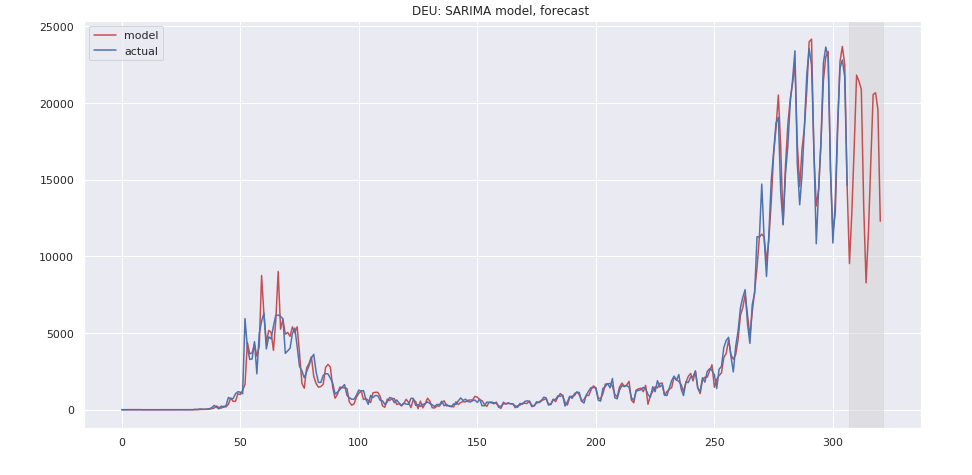
1. Define Model
   1. in total 7 parameters to be chosen
   2. choose error function to evaluate the model accuracy
   3. per grid search calculate error for different combinations of parameters
   4. select parameter set which gives the lowest error
2. Fit Model
   1. run the model with the chosen parameters with the training data, the actual new cases
   2. returns a set of figures which can be compared with the actual data
3. Predict future development
   1. run the model with the number of time steps desired.
   2. returns the predictions

The outcome of the model is a set of parameters for each country investigated as well as a prediction of the development of the new cases within the next days.

Parameter set and prediction

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| country | p | q | d | P | Q | D | s | actual:  new cases 29.11.2020 | predicted: new cases 7.12.2020 |
| Germany | 6 | 2 | 1 | 1 | 1 | 2 | 7 | 14611 | 8275 |
| USA | 0 | 2 | 1 | 1 | 1 | 4 | 24 | 154893 | 228418 |
| Sweden | 5 | 3 | 1 | 1 | 0 | 2 | 25 | 5464 | 6114 |

The model will predict how the actual time series will proceed. The graph shows the development of new cases within Germany:



The prediction (red line) follows the 7-days motion (blue line) within the numbers as well as the declining after reaching a high point some weeks before.