HIBOP-COVID
User guide

Angel Borque, Ana C. Cebrian, Luis M. Esteban, Miguel Lafuente,
Javier López, Pedro Mateo, José A. Moler, Ana Pérez, Gerardo Sanz

January 7, 2023

Contents

1 Introduction

This document includes a brief set of instructions to use the software HIBOP-COVID:

Hospital and **I**CU **B**ed **O**ccupancy **P**rediction in **COVID**-19

Objective: The main purpose of the tool is to forecast the ICU and hospital bed occupancy by COVID-19 patients over a period of time in the future. Here "hospital bed occupancy" comprises both ICU and non-ICU beds. Also, the tool can be validated using past periods in order to assess its predictive power for the hospital or region under study.

Hardware/software request: The program has been coded in Java, for its execution you only need a *jre* (java runtime environment) 1.8 or higher. All the required libraries are included in the distributed executable version of the program. This document as well as the tool are under GPLv3 license https://www.gnu.org/licenses/gpl-3.0.html.

Additional info: All the information about the statistical methods used by the tool can be obtained in the document [1].

2 Data files

The tool requires a file with data of Covid-19 patients in the region in the previous months. This includes the relevant dates of patients and their classification in groups depending on their characteristics (sex, age...), both in Groupings I and II. See Section 3.3 in [1] for details on the definitions and usefulness of Groupings I and II. The file with patients' information must be in *.csv* format [2]. Each line contains the information of a patient:

Column 1:	*id_number*. An integer from 0 to 99999999 identifying the patient. This number is not used by the tool.
Columns 2, 3:	Group number of the individual in Groupings I and II. Group number must be an integer.
Columns 4-8:	Dates of positive testing, admission to hospital, admission to ICU, discharge from ICU and discharge from hospital. The date format is "yyyy-mm-dd"; missing dates must be written as NA.

In Figure 1 a screenshot of the sample file provided with the tool is shown[1]. The first patient shown, *id_number*=1388, belongs to groups 4 and 2, for grouping 1 and 2, respectively. His/her positive test date is 2021-03-08 and hospital admission date is 2021-03-09. He or she does not enter the ICU and leaves the hospital on 2021-03-18. It can be observed that patient number 1396 enters the ICU on 2021-03-28 and leaves the ICU on 2021-04-13.

```
1388,4,2,2021-03-08,2021-03-09,NA,NA,2021-03-18
1389,2,1,2021-03-08,NA,NA,NA,NA
1390,2,1,2021-03-08,NA,NA,NA,NA
1391,3,1,2021-03-08,NA,NA,NA,NA
1392,2,1,2021-03-08,NA,NA,NA,NA
1393,6,2,2021-03-08,NA,NA,NA,NA
1394,5,2,2021-03-08,2021-03-13,NA,NA,2021-03-30
1395,5,2,2021-03-08,2021-03-15,NA,NA,2021-03-18
1396,5,2,2021-03-08,2021-03-21,2021-03-28,2021-04-13,2021-04-26
1397,4,2,2021-03-08,NA,NA,NA,NA
1398,2,1,2021-03-08,NA,NA,NA,NA
```

Figure 1: Data file. csv structure.

3 Using the tool

In the main screen of the application window, Figure 2, the user can select the *Forecasting* or *Validation* option, so the first action will be to choose the desired option and check the corresponding button. At this first moment only *Go* and

[1]Folder *DataSamples* contains a sample data file and two files which contain new COVID infection patients.

About buttons are available because the remaining ones will only be activated after processing the user's data.

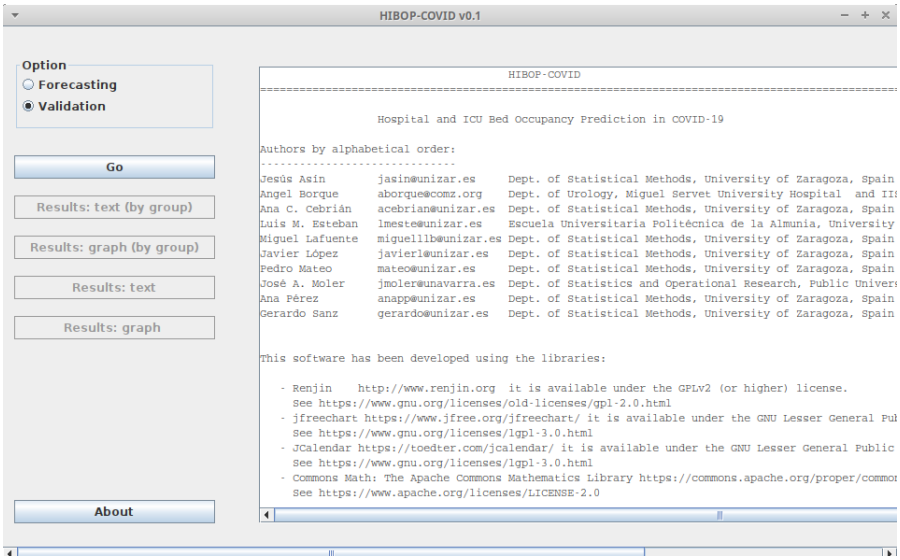


Figure 2: Main screen of HIPOB-COVID

3.1 Forecasting option

After selecting the *Forecasting* option and clicking the button *Go* a new window, Figure 3.

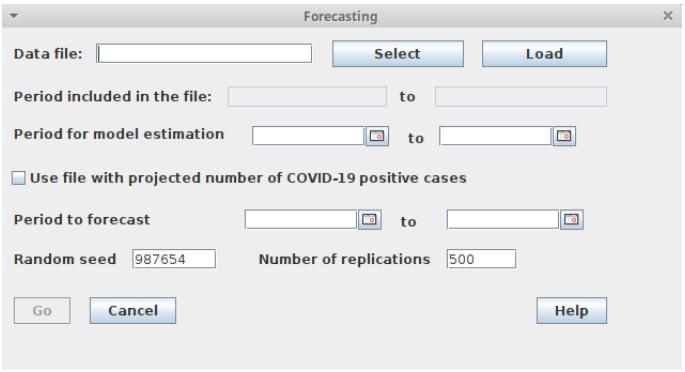


Figure 3: Forecasting sub-window.

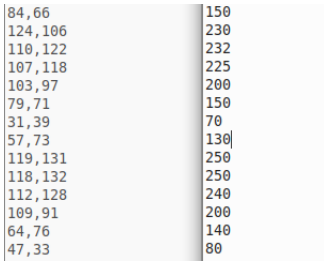
User has to follow the following steps:

1. Select the data file with the patient information, for that, *Select* button has to be clicked. Then, the Load button is pushed to pre-process the *csv* selected file. In the folder *DataSamples* provided with the application there is a sample file called *sample_data.csv* which can be used to follow this user guide.

2. After Step 1 the window will show the information about the range of dates contained in the data file. The result with the *sample_data.csv* is shown in Figure 5. This file contains information from 2021-03-02 to 2021-05-25.

3. Next step consists in selecting the cohort period for the model estimation, parameters t_I and t_F in Section 3.3 of [1]. The period for model estimation must be within the range of dates included in the file and the forecast period must begin after the last day included in the file. We can leave the dates the program shows by default (the whole available range).

4. The tool needs a projection of the new positive cases during the forecast period for each of the groups defined in Grouping II. If you have your own projections, then check the box "*Use file with projected number of COVID-19 positive cases*". Then select a file containing the information of the projected cases. The file must be in *csv* format. The number of lines in the file is the number of days in the forecast period and it can have either a single column or as many columns as the number of groups in Grouping II. In the latter case, the number in *Row* i and *Column* j indicates the number of projected new positive cases diagnosed on day i of the forecast period in group j.



84,66	150
124,106	230
110,122	232
107,118	225
103,97	200
79,71	150
31,39	70
57,73	130
119,131	250
118,132	250
112,128	240
109,91	200
64,76	140
47,33	80

Figure 4: Sample files of the projected new positive cases. Disaggregated by Group II data, left, and total number, right.

In the *DataSamples* folder there are two additional files, Figure 4, *projected_pos_1.csv* and *projected_pos_2.csv* with two examples of 2-week projected data. The first file contains information for each day and for each of the two groups defined in Grouping II (according to the *sample_data.csv* file) and the second one contains the information in only one column, showing the total number of projected new positive cases diagnosed every day (the tool internally splits this number for estimating the number in each group using the proportion of cases by group in the last weeks of the

estimation period). It can be observed that for the first day in the first file the number of new positive cases for the first group is 84 and for the second one is 66. For the second file the number of positive cases is not disaggregated and the total number in this first day is 150 new infected cases.

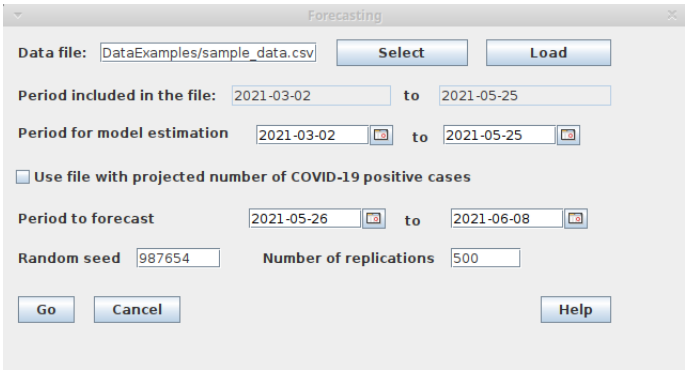


Figure 5: Forecasting window after loading *sample_data.csv* file.

If you do not have a projection of the new positive cases for the forecast period, then leave the box "*Use file with projected number of COVID-19 positive cases*" unchecked and the application will project the number of positive cases using time-series methodology (See Section 3.5 in [1]). In this case the "Period to forecast" dates must be included.

5. Whether the user has checked the box "*Use file with projected number of COVID-19 positive cases*" or not, a random seed to initialize the random elements of the tool and the number of replications (parameter *nsim* in Section 3.5 of [1]) must be introduced. In Figure 6 the required information when no projected new cases file is available, left, and when projected information is available, right, is shown.

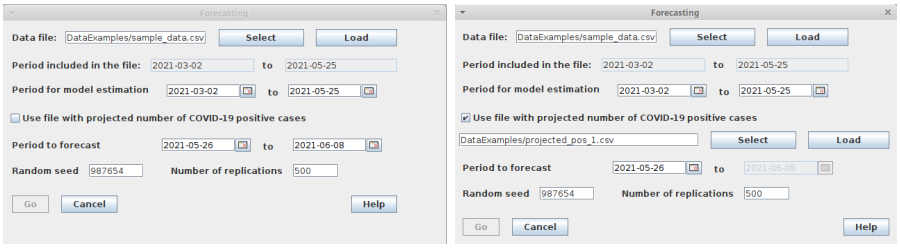


Figure 6: Forecast window after introducing required information.

6. After clicking the *Go* button, if a file with the projected number of new positive cases is not provided, an additional windows appears requesting

the time frequency; this is an integer specifying the seasonal parameter of the series of positive cases, which will be used for projections during the forecast period (see [3],[4]). The default is 7, corresponding to week seasonality; if your series has no seasonality, write 1. (Figure 7).

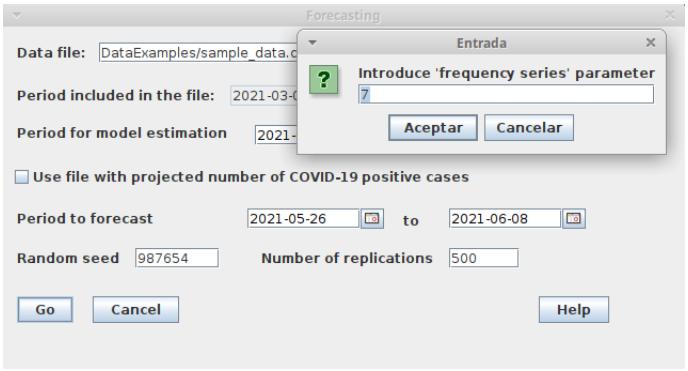


Figure 7: Introducing the time frequency needed parameter for Holt-Winters estimation method.

3.2 Forecasting results

Once simulations have been completed, you can access the results, both in text and graph formats.

The program provides four outputs with the results of the calculations:

Result: text For each day of the forecast period, the following information is given (see Figure 8, left):

- Projected number of positive cases (from the user file or the time-series projections)
- Mean, 5th, 95th percentiles of forecasted ICU bed occupancy
- Mean, 5th, 95th percentiles of forecasted hospital bed occupancy
- **Warning:** The 5th-95th percentile bands in forecasting take into account only the variability of the evolution of the system considering the number of positive cases in the period as fixed; in particular, they do not include the error in the time-series projections of the number of positive cases.

Result: Text (Group) Detailed information for each group (in Grouping II), including *mean, sd 5th, 10th, 50th, 90th, 95th percentiles* of forecasted ICU and hospital bed occupancy (see Figure 8, right). The information can be exported as a **txt** file using the button **Export results**.

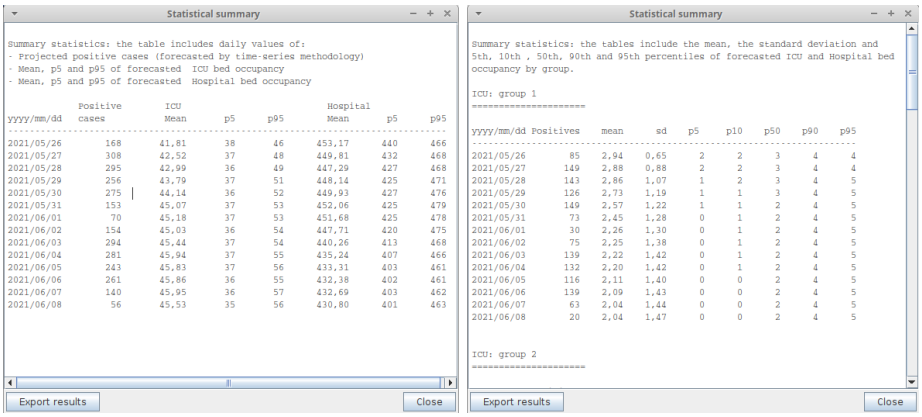


Figure 8: Output obtained by **Result: text**, left, and **Result: Text (Group)**, right, buttons.

Result: graph An interactive graph is produced, where the x-axis represents time and the y-axis represents the number of cases or occupied beds (See Figure 9). The lines on the graph are drawn by checking the corresponding names at the left panel:

- Positive. Projected number of positive cases either from the user file or time-series projection.
- Forecast Hospital. Forecast of hospital bed occupancy together with 5th-95th percentile bands.
- Forecast ICU. Forecast of ICU bed occupancy together with 5th-95th percentile bands.

Drag the Previous period slider to modify the initial date in the x-axis and draw the observed data in the extended period. The appearance of the graph can be changed by clicking the right mouse button over the graph and selecting the desired option in the resulting menu. This menu also provides the option of storing the graph in **png** (bitmap) and **svg** (vectorial) formats.

Result: Graph (Group) Same graph as above, separated by groups (in Grouping II). Check the desired information and select the groups in the left panel of the graph (see Figure 10).

3.3 Validation option

The introduction of the information is similar to the one of the Forecast option.

First the data file with the patients' information is selected and loaded by clicking the *Select* and *Load* buttons. The tool will upload the file and will show the range of dates included. Then select the period for model estimation (values

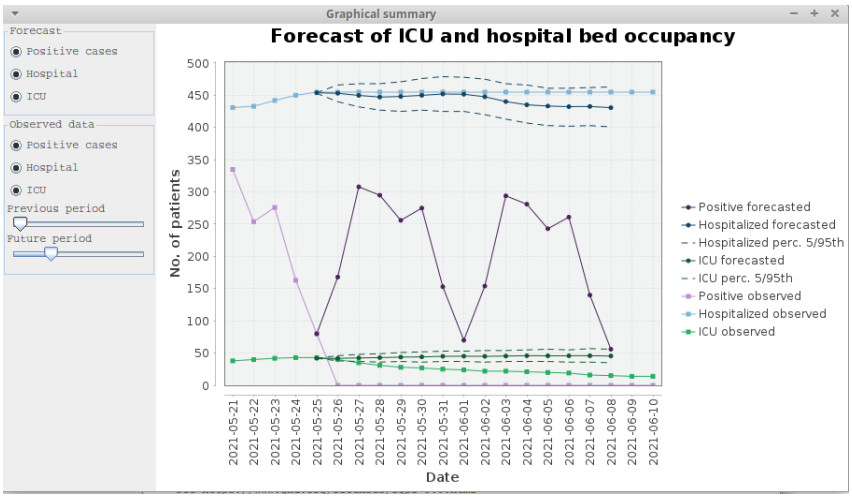


Figure 9: Interactive graph obtained by **Result: graph** button.

of t_I and t_F in Section 3.3. of [1]) and the period of validation. Note that both the periods for model estimation and for validation must be within the range of dates included in the file, and that the period for validation must begin after the end of the period for model estimation.

Finally a random seed to initialize simulations and the number of replications have to be set. After clicking the **Go** button a window will pop up, requesting the time frequency (the integer specifying the seasonal parameter of the series of positive cases, which will be used for projections during the validation period). The default is 7, corresponding to week seasonality; if your series has no seasonality, write 1. Figure 11 shows the Validation windows after introducing the information.

Once simulations have been completed, you can access the results, both in text and graph formats.

3.4 Validation results

The obtained results after the execution of the Validation option are the following:

Result: Text For each day of the validation period, the following information is given (see Figure 12):

- Observed and time series projection of the number of positive cases
- Observed ICU bed occupancy
- Mean, 5th, 95th percentiles of forecasted ICU bed occupancy both when the observed number of positive cases is used (column OPC)

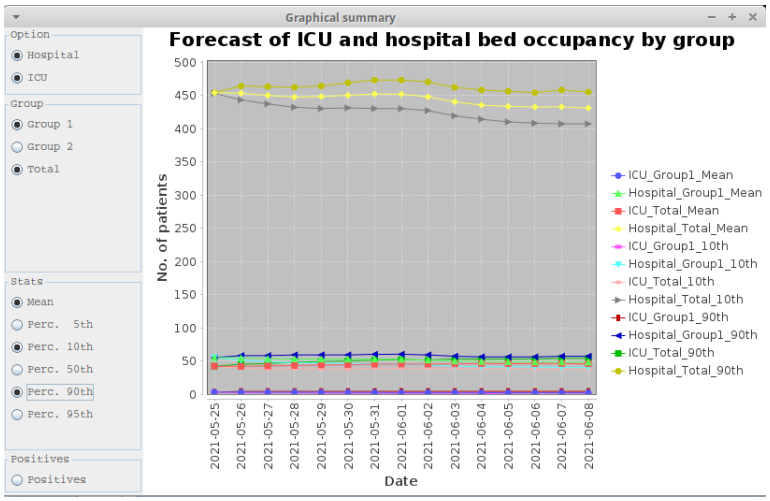


Figure 10: Interactive graph obtained by **Result: graph (Group)** button.

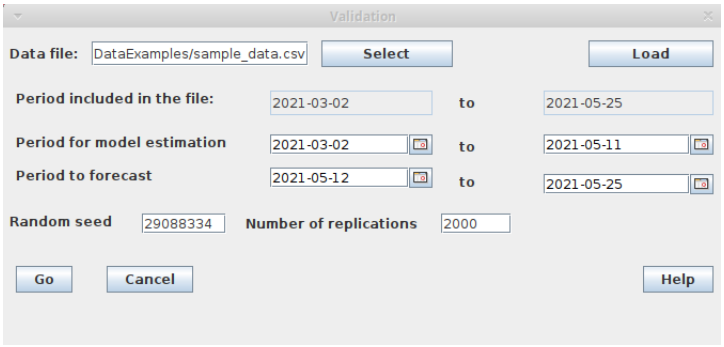


Figure 11: Validation window after selecting parameters.

and when the time-series projection of the number of positive cases is used (column PPC).

- Observed hospital bed occupancy
- Mean, 5th, 95th percentiles of forecasted hospital bed occupancy both when the observed number of positive cases is used (column OPC) and when the time-series projection of the number of positive cases is used (column PPC).
- **Warning**: The MAPE can be high when the number of positive cases is projected by the time series and the validation period is long, since the time series algorithm cannot predict a change of trend in the number of positive cases. If such a change occurs in the validation period, the projected positive cases will be very different from the

observed cases and this will result in bad forecasts of hospital and ICU bed occupancy. Also, note that the 5th-95th percentile bands in forecasting take into account only the variability of the evolution of the system considering the number of positive cases in the period as fixed; in particular, they do not include the error in the time-series projections of the number of positives.

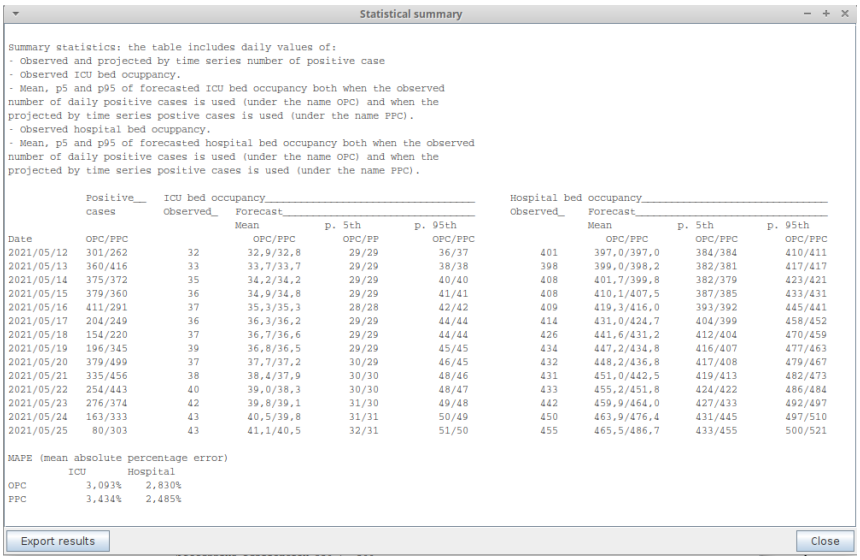


Figure 12: Information obtained with **Result: Text** button.

Result: Text (Group) Detailed information for each group (in Grouping II), including *mean, sd 5th, 10th, 50th, 90th, 95th percentiles* of forecasted ICU and hospital bed occupancy (see Figure 13).

Result: Graph An interactive graph is produced, where the x-axis represents time and the y-axis represents the number of cases or occupied beds. The lines on the graph are drawn by checking the corresponding names at the left panel (see Figure 14):

- Forecast (P). Positive. Time-series projection of the number of positive cases in the validation period.
- Forecast (P). Hospital. Forecast of hospital bed occupancy using the time- series projections of the number of positive cases in the validation period, together with 5th-95th percentile bands.
- Forecast (P). ICU. Forecast of ICU bed occupancy using the time-series projections of the number of positive cases in the validation period, together with 5th-95th percentile bands.

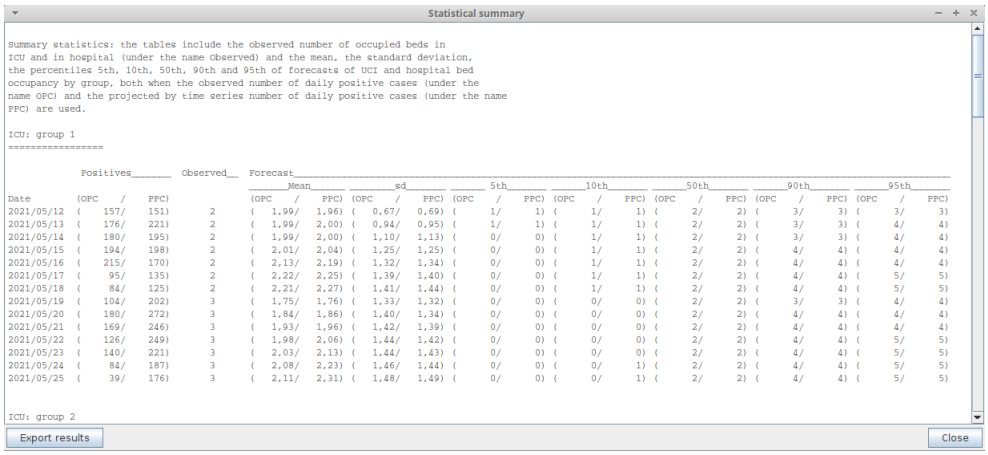


Figure 13: Information obtained with **Result: Text (Group)** button.

- Forecast (O). Hospital. Forecast of hospital bed occupancy using the observed number of positive cases in the validation period, together with 5th-95th percentile bands.
- Forecast (O). ICU. Forecast of ICU bed occupancy using the observed number of positive cases in the validation period, together with 5th-95th percentile bands.
- Observed data. Positive. Number of observed positive cases.
- Observed data. Hospital. Observed hospital bed occupancy.
- Observed data. ICU. Observed ICU bed occupancy.

Drag the Previous period and Future period sliders to modify the initial and final dates in the x-axis and draw the observed data in the extended periods.

Result: Graph (Group) Similar graph to the previous one, separated by groups (in Grouping II). Check the desired information and select the groups in the left panel of the graph (see Figure 15).

References

[1] Borque A., Cebrian A.C. et al., *A Multistate model and its standalone tool to predict hospital and ICU occupancy by patients with COVID-19.* Heliyon, 2022. **Under review**.

[2] `https://docs.fileformat.com/spreadsheet/csv/`. Accessed December, 2022.

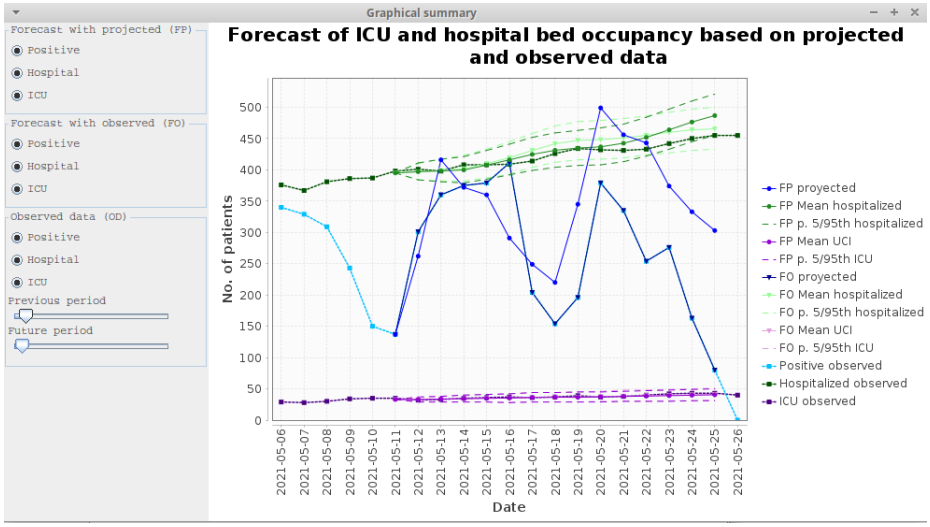


Figure 14: Information obtained with **Result: graph** button.

[3] Holt, C.C., *Forecasting Trends and Seasonal by Exponentially Weighted Averages*. Int J Forecast 20(1) 5–10, 1957.

[4] Winters, P.R., *Forecasting Sales by Exponentially Weighted Moving Averages*. Manag Sci 6(3) 324–342, 1960.

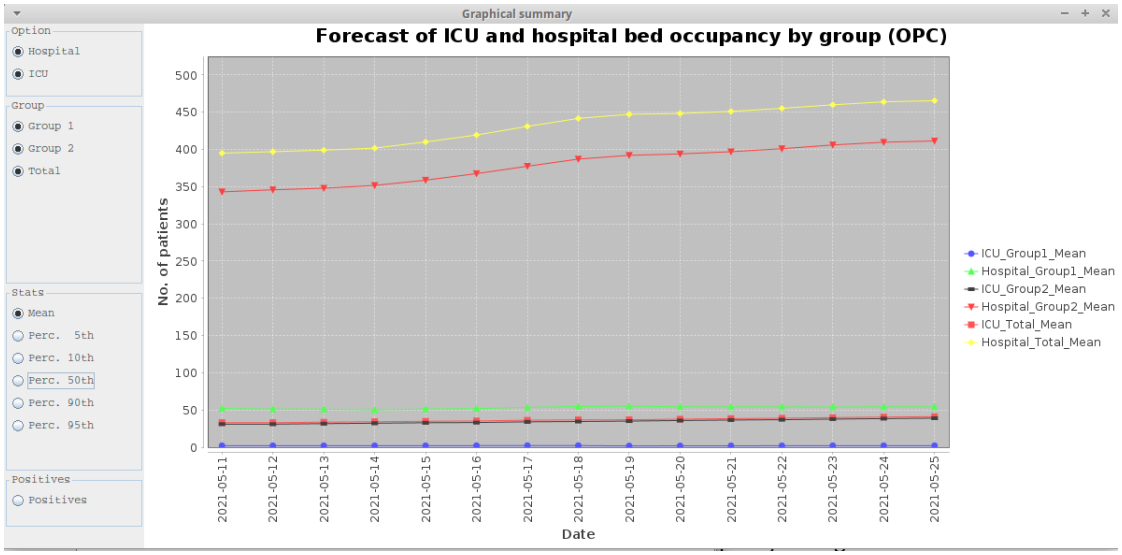


Figure 15: Interactive graph obtained by **Result: graph (Group)** button.