Information on platform and submission

RespiNow SP4

Where to find what

Resources for the project are spread out across a few repositories and websites:

- The official RespiNow website: (https://respinow.de/) contains some basic information, but is mainly for communicating to the outside.
- The beta version of the visualization dashboard currently lives at https://jobrac.shinyapps.io/respinow_viz/. It now does most of what it needs to do, but still has a few bugs.
- The repository https://github.com/KITmetricslab/RESPINOW-Hub is the backbone of the project. It contains the data we have aggregated and will collect the nowcasts / forecasts. See details in the following sections.
- The repository https://github.com/KITmetricslab/RESPINOW-Data contains some data stuff which we pulled outside of the main repository. You likely don't need to look into this repository.

Truth data format

Truth data is available in the folder `data` of the `RESPINOW-Hub` repository. Files are sorted by the data source and disease. For each indicator we provide three csv files:

- `latest_data-<source>-<disease>.csv`: This file contains simple time series data in the format of weekly counts. See example here.

```
          date year week location age_group value
100 2021-09-26 2021   38       DE       00+   508
101 2021-10-03 2021   39       DE       00+   695
102 2021-10-10 2021   40       DE       00+   981
103 2021-10-17 2021   41       DE       00+   878
104 2021-10-24 2021   42       DE       00+   658
105 2021-10-31 2021   43       DE       00+   390
```

The `date` variable corresponds to the Sunday concluding the respective week. These data may be stratified by `location` and `age_group`. Note that the last 1–2 values of this time series (in some cases more) are typically subject to revisions and may be overwritten with new, typically higher values.

- `reporting_triangle-<source>-<disease>.csv`: For nowcasting we need to keep track of these revisions, which is why we provide so-called reporting triangles. These contain the weekly counts stratified by how many weeks it took until they appeared in the data. See example here.

	location	age_group	year	week	date	value_0w	value_1w	value_2w	value_3w
56	DE	00+	2022	47	2022-11-27	391	80	0	0
57	DE	00+	2022	48	2022-12-04	538	145	8	0
58	DE	00+	2022	49	2022-12-11	904	171	2	1
59	DE	00+	2022	50	2022-12-18	937	138	1	46
60	DE	00+	2022	51	2022-12-25	889	88	21	58
61	DE	00+	2022	52	2023-01-01	457	83	17	0
62	DE	00+	2023	1	2023-01-08	690	123	2	0
63	DE	00+	2023	2	2023-01-15	514	73	-1	0

	value_4w	value_5w	value_6w	value_7w	value_8w	value_9w	value_10w	value_.10w
56	-1	1	0	2	0	0	-1	0
57	4	2	2	0	7	0	0	0
58	86	8	2	2	1	0	0	0
59	66	3	2	0	-1	1	0	1
60	3	2	1	0	0	0	0	2
61	1	2	0	0	0	0	0	0
62	1	0	0	0	0	0	4	5
63	0	0	0	0	0	12	0	3

Values reported with a delay of more than 19 weeks are aggregated. Again data are potentially stratified by `location` and `age_group`. Note that this table occasionally contains negative numbers, meaning that data have been corrected downwards rather than upwards as we would expect.

-`reporting_triangle-<source>-<disease>-preprocessed.csv`: As most nowcasting methods cannot deal with such negative increments we also provide a "processed'' version of the reporting triangle, where these negative values have been removed via a simple heuristic shifting counts between neighbouring cells. This file is typically what will be used as the input for nowcasting. See example here.

	location	age_group	year	week	date	value_0w	value_1w	value_2w	value_3w
56	DE	00+	2022	47	2022-11-27	391	79	0	0

```
57       DE       00+ 2022   48 2022-12-04      538      145        8        0
58       DE       00+ 2022   49 2022-12-11      904      171        2        1
59       DE       00+ 2022   50 2022-12-18      937      138        1       46
60       DE       00+ 2022   51 2022-12-25      889       88       21       58
61       DE       00+ 2022   52 2023-01-01      457       83       17        0
62       DE       00+ 2023    1 2023-01-08      690      123        2        0
63       DE       00+ 2023    2 2023-01-15      514       72        0        0
   value_4w value_5w value_6w value_7w value_8w value_9w value_10w value_.10w
56        0        1        0        1        0        0         0          0
57        4        2        2        0        7        0         0          0
58       86        8        2        2        1        0         0          0
59       66        3        1        0        0        1         0          1
60        3        2        1        0        0        0         0          2
61        1        2        0        0        0        0         0          0
62        1        0        0        0        0        0         4          5
63        0        0        0        0        0       12         0          3
```

Note that the `latest_data` files typically reach further back in time. For the `reporting_trianle` files we are constrained to the time period during which we collected weekly data snapshots, which depending on the indicator in question is something like the last 12–24 months.

Nowcasting / forecasting targets and submission

Definition of the target

For most indicators data can potentially be revised over very long time periods. Late revisions are rare and hardly predictable, but can be substantial. Our suggestion is to define the prediction target as the **number of events reported up to a given maximum delay** D, which could be $D = 4$ or 6 weeks. This would be the target for both the nowcasting part and the short-term forecasting part.

Prediction horizons

For each of the indicators we suggest to issue nowcasts and forecasts for the following time points:

- **nowcasts:** from $D - 1$ weeks back until the current week, i.e., all weeks for which revisions are still technically possible under the definition evoked in the previous section.

- **forecasts:** we will not set a formal limit on how far into the future you want to predict. Seasonal diseases may be predictable a bit futher into the future than emerging ones, but we do not really know how this will play out post-COVID. The US FluSight project considers forecasts up to 4 weeks into the future, but we may limit at least public display to 2 weeks.

Submission

Folder structure

Predictions will be collected in the folder `submissions`, which will be further structured by target data source and disease. Within each of the resulting subfolders, each team will have a folder where to deposit predictions for that specific target.

File naming

Submission files should in csv format and placed in the correct subfolder. The naming convention is

`<submission date>-<data source>-<disease>-<team>-<model name>.csv`.

An example of a correct name is

`2022-11-13-survstat-influenza-KIT-sipmple_nowcast.csv`.

Please note that for technical reasons the team and model names should not contain the character `"-"`,

File structure

Teams are asked to provide their forecasts in a quantile-based format (even though we also accept submissions containing only point forecasts). The tabular version of the data model is a simple, long-form data format, with the following required columns: `['location', 'age_group', 'forecast_date', 'target_end_date', 'target', 'type', 'quantile', 'value']`. The contents of these columns are defined as follows.

`forecast_date`

The data version on which the nowcasts/forecasts are based in `YYYY-MM-DD` format. This `forecast_date` should be a Sunday and correspond to the date in the filename. Our submission system will only accept nowcasts marked with the current date as the `forecast_date`.

`target`

Values in the `target` column describe which quantity is being predicted as there may be several (e.g., total number of tests and positive tests). **We have not quite figured out how to handle this**.

`target_end_date:`

The date corresponding to the end time of the target, in `YYYY-MM-DD` format. This is the Sunday at the end of the week the prediction refers to.

`location`

A unique id for the location. We use `"DE"` for the national level and the following codes for the federal states:

	location_long	location
1	Alle (Deutschland)	DE
2	Baden-Württemberg	DE-BW
3	Bayern	DE-BY
4	Brandenburg / Berlin	DE-BB-BE
5	Hessen	DE-HE
6	Mecklenburg-Vorpommern	DE-MV
7	Niedersachsen / Bremen	DE-NI-HB
8	Nordrhein-Westfalen	DE-NW
9	Rheinland-Pfalz	DE-RP-SL
10	Sachsen	DE-SN
11	Sachsen-Anhalt	DE-ST
12	Schleswig-Holstein / Hamburg	DE-SH-HH
13	Thüringen	DE-TH

Note that not all indicators may be available with age stratification. In this case just use `"DE"`.

`age_group`

One of the following to indicate the age group: `"00+"` (all age groups), `"00-04"`, `"05-14"`, `"15-34"`, `"35-59"`, `"60-79"`, `"80+"`. Note that not all indicators may be available with age stratification. In this case just use `"00+"`.

type

Either "mean" or "quantile".

quantile

A value in [0.025, 0.1, 0.25, 0.5, 0.75, 0.9, 0.975], stating which quantile is displayed in this row. If type=="mean" then NA. We encourage all groups to make available all 7 quantiles.

value

A numeric value representing the value of the quantile or mean prediction.

For example, if quantile is 0.3 and value is 10, then this row is saying that the 30% quantile of the predictive distribution is 10. If type is "mean" and value is 15, then this row is saying that the predictive mean from this model is 15.

See this example file:

	location	age_group	forecast_date	target_end_date	target
1	DE	00+	2023-02-05	2023-01-15	-3 week ahead inc case
2	DE	00+	2023-02-05	2023-01-22	-2 week ahead inc case
3	DE	00+	2023-02-05	2023-01-29	-1 week ahead inc case
4	DE	00+	2023-02-05	2023-02-05	0 week ahead inc case
5	DE	00+	2023-02-05	2023-01-15	-3 week ahead inc case
6	DE	00+	2023-02-05	2023-01-22	-2 week ahead inc case
7	DE	00+	2023-02-05	2023-01-29	-1 week ahead inc case
8	DE	00+	2023-02-05	2023-02-05	0 week ahead inc case
9	DE	00+	2023-02-05	2023-01-15	-3 week ahead inc case
10	DE	00+	2023-02-05	2023-01-22	-2 week ahead inc case
11	DE	00+	2023-02-05	2023-01-29	-1 week ahead inc case
12	DE	00+	2023-02-05	2023-02-05	0 week ahead inc case
13	DE	00+	2023-02-05	2023-01-15	-3 week ahead inc case
14	DE	00+	2023-02-05	2023-01-22	-2 week ahead inc case
15	DE	00+	2023-02-05	2023-01-29	-1 week ahead inc case
16	DE	00+	2023-02-05	2023-02-05	0 week ahead inc case
17	DE	00+	2023-02-05	2023-01-15	-3 week ahead inc case
18	DE	00+	2023-02-05	2023-01-22	-2 week ahead inc case
19	DE	00+	2023-02-05	2023-01-29	-1 week ahead inc case
20	DE	00+	2023-02-05	2023-02-05	0 week ahead inc case
21	DE	00+	2023-02-05	2023-01-15	-3 week ahead inc case
22	DE	00+	2023-02-05	2023-01-22	-2 week ahead inc case
23	DE	00+	2023-02-05	2023-01-29	-1 week ahead inc case

```
24       DE       00+    2023-02-05      2023-02-05  0 week ahead inc case
25       DE       00+    2023-02-05      2023-01-15 -3 week ahead inc case
26       DE       00+    2023-02-05      2023-01-22 -2 week ahead inc case
27       DE       00+    2023-02-05      2023-01-29 -1 week ahead inc case
28       DE       00+    2023-02-05      2023-02-05  0 week ahead inc case
29       DE       00+    2023-02-05      2023-01-15 -3 week ahead inc case
30       DE       00+    2023-02-05      2023-01-22 -2 week ahead inc case
31       DE       00+    2023-02-05      2023-01-29 -1 week ahead inc case
32       DE       00+    2023-02-05      2023-02-05  0 week ahead inc case
       type quantile    value
1      mean       NA 5971.000
2      mean       NA 2735.423
3      mean       NA 2089.162
4      mean       NA 2083.992
5  quantile    0.500 5971.000
6  quantile    0.500 2735.000
7  quantile    0.500 2087.000
8  quantile    0.500 2068.000
9  quantile    0.025 5971.000
10 quantile    0.025 2730.000
11 quantile    0.025 2073.000
12 quantile    0.025 1987.975
13 quantile    0.100 5971.000
14 quantile    0.100 2731.000
15 quantile    0.100 2077.000
16 quantile    0.100 2004.000
17 quantile    0.250 5971.000
18 quantile    0.250 2733.000
19 quantile    0.250 2081.000
20 quantile    0.250 2027.000
21 quantile    0.750 5971.000
22 quantile    0.750 2737.000
23 quantile    0.750 2095.000
24 quantile    0.750 2117.000
25 quantile    0.900 5971.000
26 quantile    0.900 2740.000
27 quantile    0.900 2104.000
28 quantile    0.900 2179.100
29 quantile    0.975 5971.000
30 quantile    0.975 2745.000
31 quantile    0.975 2118.025
32 quantile    0.975 2294.025
```

Submission rhythm

We plan to have weekly updates. Maybe we can have a system where nowcasts are updated on Monday and forecasts from models which require nowcasts as input are updated on Tuesdays.