Project: Weather-API

Valentino Mascherini

26/01/2022

1 The variable: air temperature

Air temperature is a good benchmark variable, and for different reasons. First of all, it's quite easy to measure, meaning it doesn't need costly and complex, error prone sensors in order to be recorded. The sensors do not need a particularly time-spending set-up phase, and can be placed almost everywhere, as long as they are out of range of strongly-affecting heat sources. Furthermore, it is the closest measurement we can get in respect to the perceived temperature, making it a great forecast predictor of how the weather will actually affect you.

2 Unsuited station data

We can see from the table below (Table 1) that the results look at least decent, as every HOBO shows a value higher than 0.5. It has to be noted how last year's data have just a moderately strong correlation with the modeled data, while, on the other hand, 2022's data are very strongly correlated. I would say that the differences are significant, judging by how the most strongly correlated hobo from 2021 cannot break past the 0.8 mark, when for 2022 the weakest is over 0.9. A station from 2021, namely that year's weakest, was placed outside the city of Freiburg and is no wonder why is characterized by such values. This year's weakest is also quite far from this year's best, by around 0.5, but in this case it's not due to a topographical outlier. The plot 2 actually show the correlation of the single HOBOs to the model quite neatly, distinguishing for both terms and thus giving us the general situation at a glance. Here we can we To obtain the said values, Pearson's correlation test was used, instead of a covariance test. My choice is mostly dependent by the fact that a correlation is not affected by a change in scale, a procedure that is absolutely vital in a regionalization context. It's also a really nice tool to investigate a relationship before implementing any kind of model.

3 Regional differences

In order to visualize the regional differences between the model and the actual temperature measurements, I got the mean of the temperature of every HOBO present in a district and tested it through the Pearson correlation against the model data, and plotted it on the Freiburg map (Map 3). I first grouped by year and calculated the average district temperature, then I got the average of the correlation per district. The resolution of the regions are the different city districts, that are filled with different colour corresponding to a certain correlation level. Only the districts with HOBOs actually placed are coloured, thus the majority of the districts furthest to the city center is unfilled. The differences are quite harsh, from 0.9 to less than 0.7. We can see, looking at the areas containing isolated circles, how the 2022 HOBOs have the best correlation, and how the areas with an high ratio of circles over triangles usually have a better ratio, especially than triangles alone. We can conclude that the most correlated districts are the ones who have the most number of new data, making "year" a confounding variable. To have a better understanding of how the districts differ, the map 4 shows only the current year's HOBOs. We see that the model correlates best with the devices in

Table 1: Pearson correlation for 2021 and 2022

device_id	pear	device_id	pear
10350049	0.6586137	10347320	0.9503876
10760820	0.7032124	10088310	0.9704685
10350090	0.5612516	10234637	0.9531420
10350043	0.7181350	10347328	0.9632858
10350000	0.7050513	10801141	0.9465079
10350068	0.7367652	10347362	0.9692138
10347320	0.6814675	10347366	0.9661877
10350101	0.7091883	10760820	0.9599650
10347367	0.6587725	10347313	0.9634984
10347318	0.7232509	10350049	0.9156104
10801132	0.7146285	10347384	0.9638377
10347325	0.7088330	10760822	0.9552664
10347371	0.7253197	10350002	0.9423964
10347310	0.6469094	10347327	0.9403460
10350017	0.7195004	10350005	0.9675617
10760810	0.7244618	10801134	0.9593305
10760710	0.7125403	10760709	0.9619075
10347377	0.6735399	10347394	0.9558611
10347342	0.7306123	10760710	0.9709113
10350072	0.4636110	10347356	0.9650677
10350075	0.7171054	10610854	0.9632271
10347394	0.7008400		
10088310	0.7025215		
10350028	0.6618593		
10801134	0.6883759		
10350007	0.6923150		
10347335	0.6796495		

Table 2: Displaying records 1 - 10

t_avg	t_day	t_night	t_t_var	count	id
6.025167	6.536034	5.514300	3.025000	504	14
6.021342	6.405295	5.637390	4.375000	672	59
5.895161	6.080881	5.709441	1.555000	504	6
5.198464	5.839754	4.557175	5.351000	504	2
5.415662	6.085108	4.746216	6.200333	504	32
5.562612	6.112062	4.999761	4.892488	498	29
5.224146	5.764280	4.684012	7.571000	672	54
5.562612	6.112062	4.999761	6.598000	498	29
4.957056	5.378992	4.535119	4.612000	504	18
6.222058	6.711339	5.732777	4.455166	504	24

the Wiehre district, that has two sensors, a number alike three other areas. To be fair, the districts finding themselves worse off correaltion-wise are those with only one HOBO, so we could say that the more the sensors, the better the correlation. To further support this thesis, I also tried finding the Pearson correlation for each HOBO and then deriving its mean per district, and the distribution was a little different in the districts with more HOBOs, in fact they had a lower correlation than map 4. This means that having more sensors captions an overall temperature average closer to the actual model.

4 Database VS Scripts

Both SQL queries and scripts can be exploited in order to achieve the same goal, being initially the display of an overview of the data, then various plots (like Plot 1) and maps. However, it has to be accounted for the possibility of automation that scripts entail, due to the possibility of easily storing objects and elements, opposed to the volatility of queries. Furthermore, resources like GitHub and Pastebin making it so that you don't need a specific database connection allow the scripts to be easily designable to function independently.

5 Tables, plots and maps

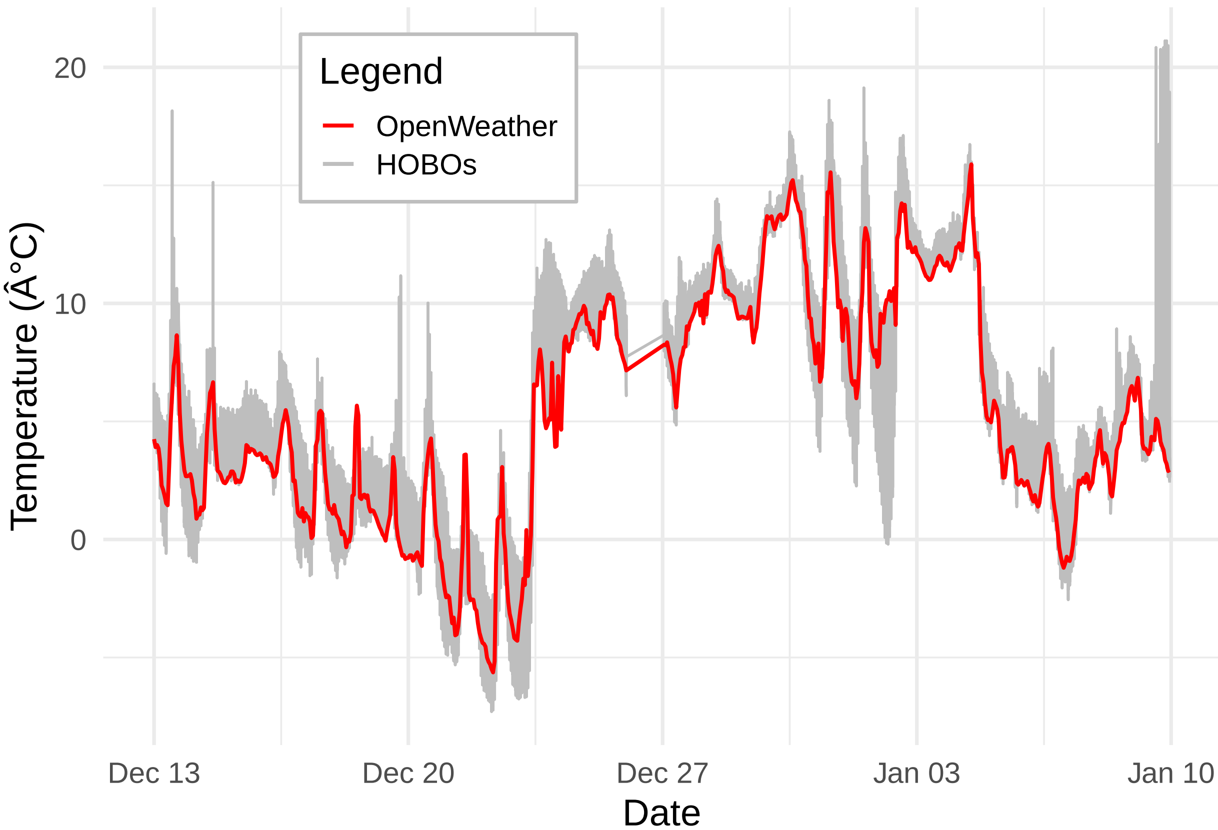

Figure 1: Differences in correlation between HOBOs

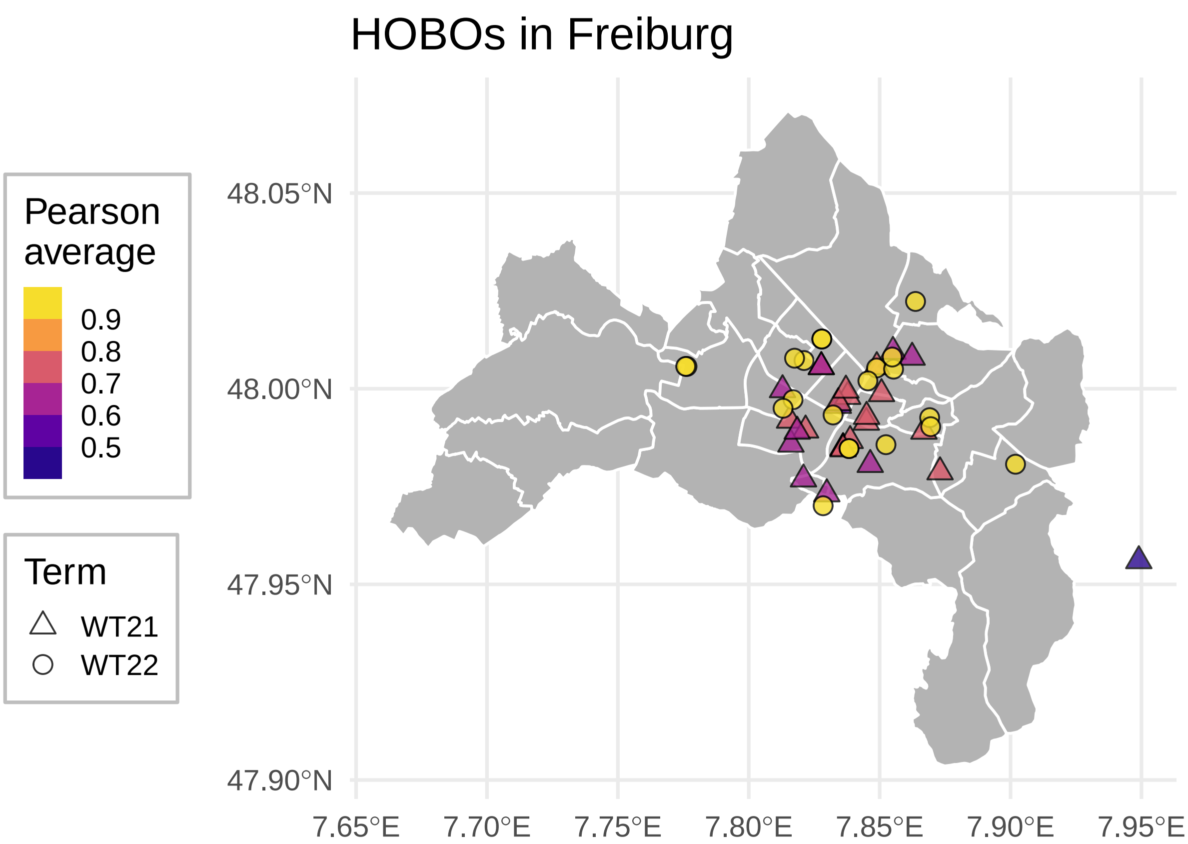

Figure 2: Differences in correlation between HOBOs

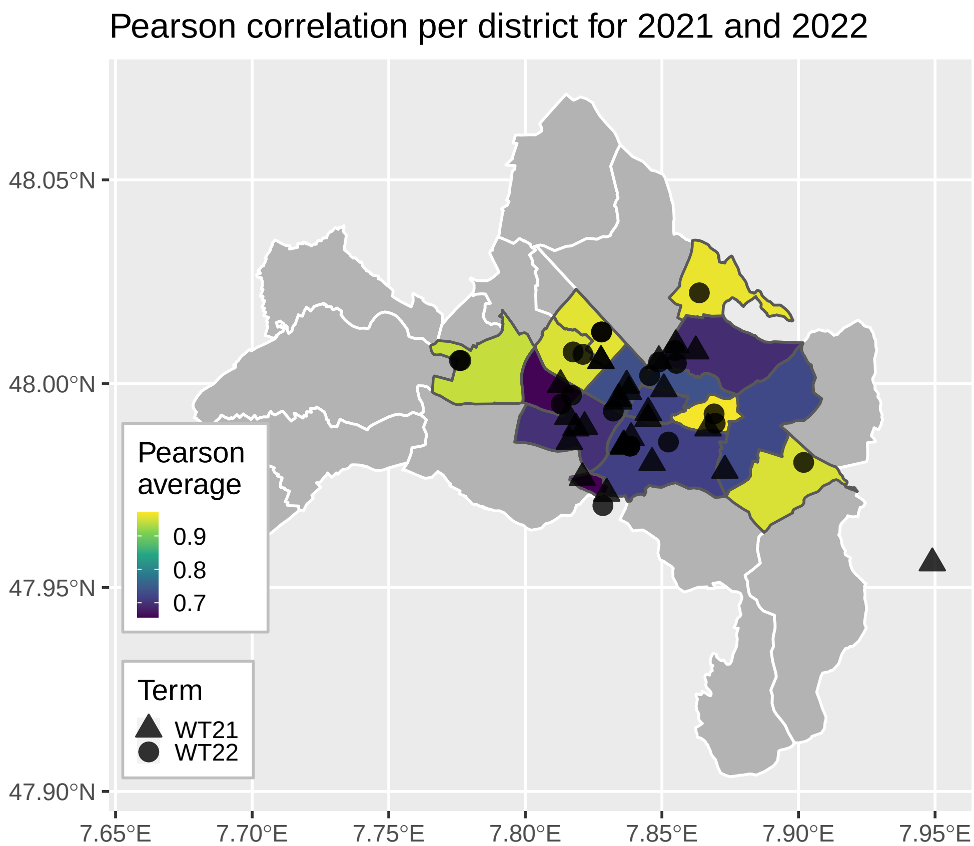

Figure 3: Differences in correlation between city districts 2021 and 2022

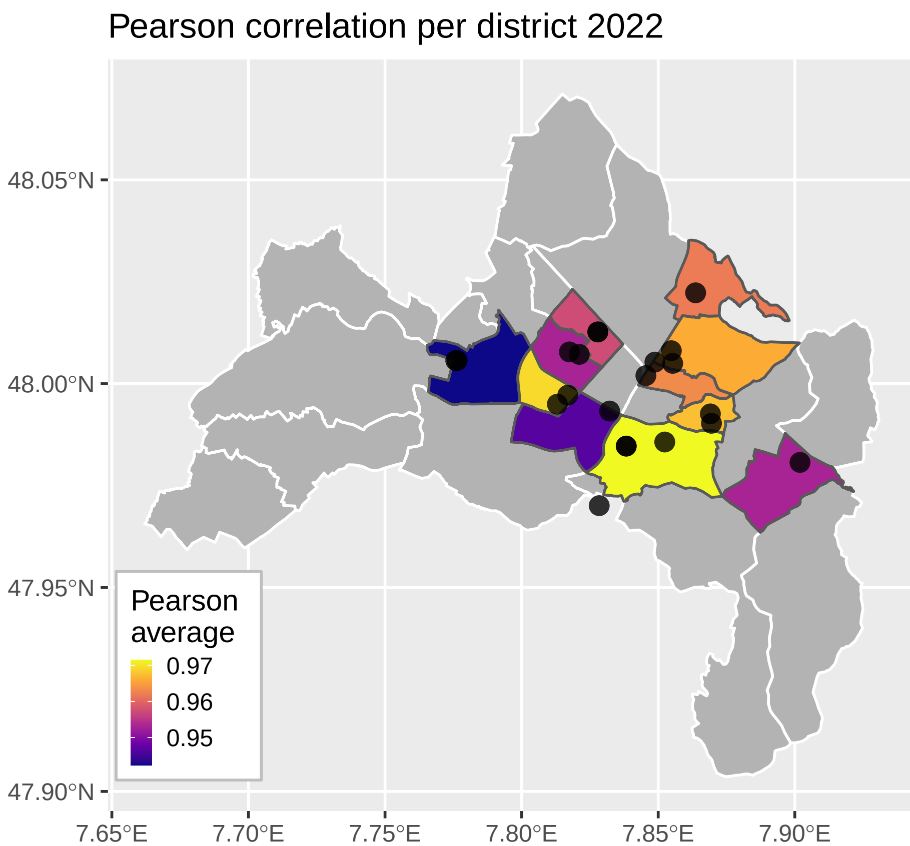

Figure 4: Differences in correlation between city districts 2022