Evolving Weather Trends in Germany: Insights from Historical Data

Student Yannick Rudolf
Matriculation number 22722156

Introduction

Nowadays, as it affects us all climate change is a hot topic. According to the News last June was 1.33 degrees warmer in Germany than in the long-term comparison period. Nevertheless, it feels like there are more and more storms and natural disasters. Most recently there were floods in the south of Germany. Therefore this project answers the question of how the number and types of weather phenomena in Germany changed over time.

Used data

The data used for this project is provided by the German weather service which is called Deutscher Wetterdienst (DWD). The data is under the Creative Commons BY 4.0 "CC BY 4.0" License and is therefore free to use under certain conditions. From the open data portal hosted by the DWD all data sets that have information on weather phenomena are extracted. The data sets contain annual data on the occurrences of weather phenomena in Germany.

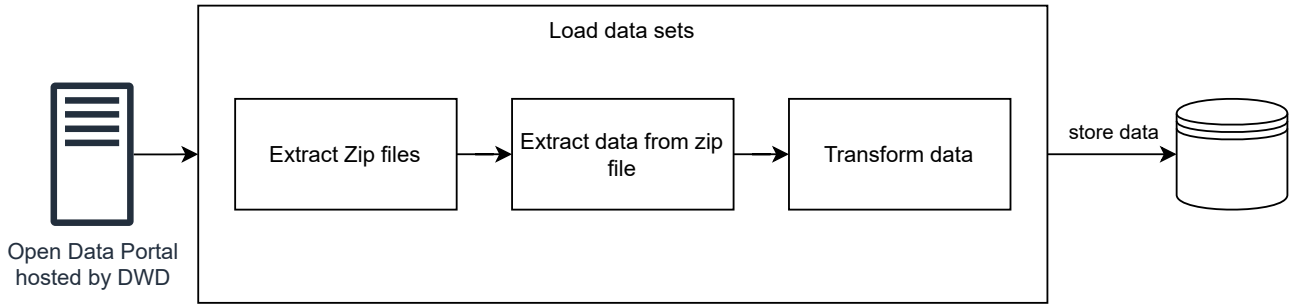

Figure 1: Steps within the data pipeline

Figure 1 shows the process of the data extraction. Each weather station has its own data set in the form of a zip file containing meta and weather data. In total, the pipeline extracts 999 data sets for 998 weather stations and 1 data set which contains metadata in the form of a text file. The metadata, which is shown as an example in Table 1, contains information about the position of the respective station.

STATIONS_ID	geoBreite	geoLaenge	Stationshoehe
1	47.8413	8.8493	478
3	50.7827	6.0941	202
44	52.9336	8.2370	44
52	53.6623	10.1990	46
70	48.2052	9.0371	712

Table 1: Sample metadata

Each weather station's data can be categorized into two main categories. Data on the observation. Which station, which period is measured, and what is the data quality?

As shown in Table 2 most of the columns are relatively self-explaining. The column QN_4 gives information on the data quality in the observed year. The DWD categorizes the quality of the data into 10 different levels. An exhaustive overview can be found on the webpage of the DWD. In the context of this project and regarding the analysis I categorized the data into three categories as shown in Table 3.

Each data set contains information on the weather phenomena on an annual basis. Overall 8 weather phenomena are observed. Thunderstorms, black ice, sleet, hail, fog, storms > 6 on the Beaufort scale, storms > 8 on the

STATIONS_ID	MESS_DATUM_BEGINN	MESS_DATUM_ENDE	QN_4
1	1981-01-01	1981-12-31	10
1	1982-01-01	1982-12-31	10
1	1983-01-01	1983-12-31	10
1	1984-01-01	1984-12-31	10
1	1985-01-01	1985-12-31	10

Table 2: Sample data for weather station 1

< 3	At most, the data was systematically checked if the values were measured automatically. In addition, however, there is no systematic placing of quality bytes
5	Historical subjective quality assurance procedures. No IT-supported testing.
> 7	At least, data that is systematically checked. All values are provided with quality bytes. Contradictory values are marked but are generally not yet corrected.

Table 3: Categories of quality levels that are defined by the DWD

Beaufort scale and dew. Each weather station tracks for each weather phenomenon the number of occurrences within every observed year. If a value is missing it is indicated with the value -999.

Finally, the output of the data pipeline is an SQL database with two tables. One table named "description" contains the metadata for the weather stations. And one table called "weather_phenomena" which contains the observed values from all weather stations. In total, the weather_phenomena table has 34372 entries. The data was intentionally stored in two tables to normalize the database and avoid storing values twice unnecessarily. The two tables can then be linked relatively simply with a join command using the station ID.

Analysis

For the analysis the data gets initially filtered by the data quality. As entries with a data quality of less than 7 are not systematically checked and contain some kind of subjective measurement, these entries are not taken into account. Only 42% (14574 entires) of the data meet the quality level requirement. Subsequently, the entries with missing values are discarded as well. 0.02% of the entries have missing values. In total 14223 values are left for the analysis. The data analyzed is between 1981-01-01 - 2021-12-31. Each observed year has at least 15 and at most 539 observations. The newer the data, the fewer valid entries there are. For example, the years 2020 and 2021 only have 16 and 15 valid observations for the weather phenomena.

Figure 2 illustrates the min, max, and mean values for all weather phenomena over the years. It is clear to see that there is a clear upward trend around 2010.

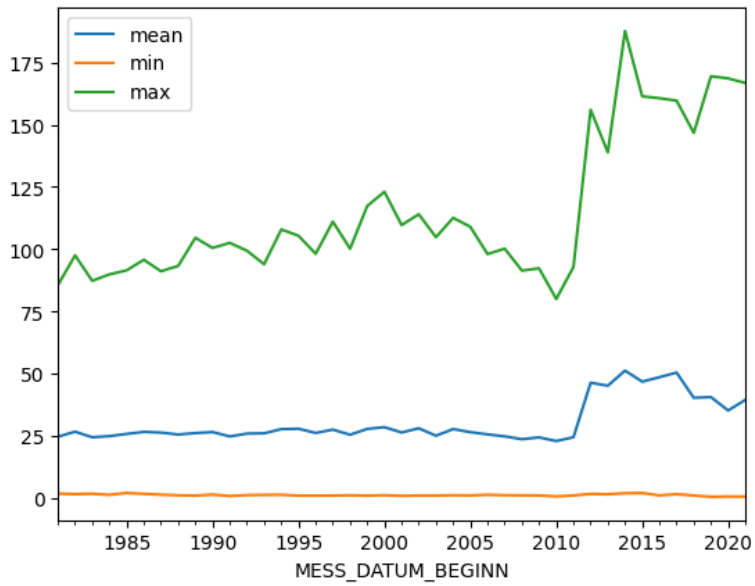

Figure 2: Overall number of occurrences of weather phenomena

More detailed is Figure 3. Here the plots are separated to have a detailed look at the different weather phenomena. It is also confirmed here that the number of weather events will increase from 2010 for thunderstorms, storms, dew, sleet, and fog.

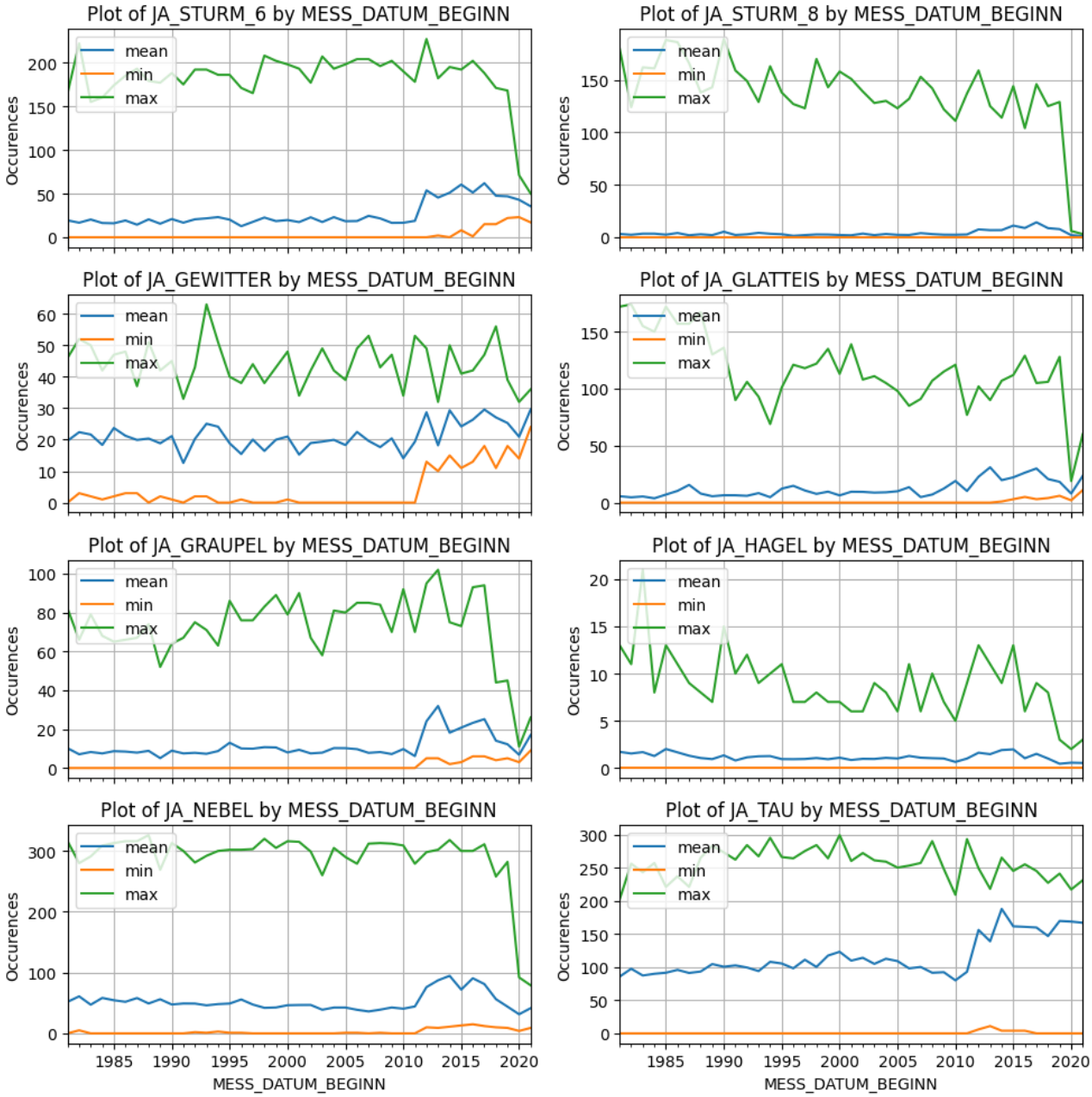

Figure 3: Min, Mean, Max values for each weather phenomenon

Conclusion

The data shows that there is a shift towards more occurrences of weather phenomena nowadays. However, since there was such an enormous jump in the graphs around 2010, it is difficult to say whether the number of weather events has increased drastically or whether something has changed in the type of measurement. In the example of dew as a weather event, for example, you can see very clearly that the mean value has changed significantly. This may have been due to newer technology or changed threshold values for detecting the event.

To summarize, the answer to the question of whether there are more weather events today than in the past is positive. However, the data also shows that it may not be due to climate change, but also to new, better technology.