Towards a Prediction Engine for Flight Delays Based on Weather Analysis

Cristina Cabanillas, Enver Campara, Claudio Di Ciccio, Bartholomäus Koziel, Jan Mendling, Johannes Paulitschke, and Johannes Prescher*

Institute for Information Business
Vienna University of Economics and Business, Austria
johannes.prescher@wu.ac.at

Extended abstract

Arrival performance for the United States shows that over 83 percent of flights are actually on time. However, 17 percent delayed flights are still an indisputable high number, having almost eight million commercial travel flights per year, only in the US. Knowledge of the conditions leading to flight delays may be used in a monitoring and prediction tool to diminish its impact on commercial flight operations. From a broader perspective, we also consider multi-modal logistics chains, which involve different modes of transportation. These modes are adopted in consecutive *legs*, which have to be thus synchronized. Determining whether a delay is going to be verified for the aircraft can thus be of advantage, in order to rearrange the overall transportation process involving such leg. This work reports on investigations carried out in the context of the GET (Green European Transport) Service[1] project. GET Service is an ongoing European research project, whose objective is to improve the ecological impact and efficiency of logistics processes.

Among the possible causes of flight delays, we focus on weather. Weather is observed throughout the world and the need to make future predictions is noteworthy.By now, research has analysed the influence of weather on airports, on flight delays in general, and on how a flight may be influenced by certain weather conditions. Furthermore, models for flight delays with respect to the weather and traffic index have been devised. However, there is little insight on the quantification of the prediction of flight delays.

In our work, we investigate in how far weather conditions have an actual impact on the punctuality of a flight. Following up on the insights gained in this step, we determine categories of impacts to allow for more generalisation. Subsequently, we use the categories and apply them in a prediction model. We fill the model with historical data. Accordingly, the model and corresponding data are the foundation for live predictions on actual flights.

*Corresponding author
[1]http://www.getservice-project.eu/

Our work builds upon the combination of two data sources being weather information and flight data. The relevant weather information is retrieved accessing the Meteorological Aerodome Report (METAR), which is an internationally established reporting instrument for weather information. METAR data is gathered at every airport and airfield and is usually generated every 30 minutes. A dataset of METAR contains station meta data (which we use to map the information to flights) as well as the information related to the weather itself. The flight data we use is of two different types, historical data and current flight data. We use historical data sets to analyse the cause for a delay and to validate our prediction model. The timeframe for our dataset ranges from 2005 to 2008. We analyse flights choosing a single route, which contains both *(i)* enough observable weather stations and *(ii)* a high amount of flights to be analysed. In order to observe and analyse the weather at a specific point in time and the position of each flight, we merge the collected information from METAR with the flight information by time, date and location for flights and weather stations. We evaluate whether there is a significant dependency between the delays of flights and certain weather conditions occurring in the meantime. We also examine at which stage of a flight specific weather conditions show the strongest impact. In order to conduct the analysis of our dataset containing 869 recorded flights we use SPSS.[2] In addition to the integration of information we suggest a conceptual description of a monitoring tool for current flights. The tool is able to predict flight delays considering the categorised impacts in conjunction with previous flight delays and may present the predicted delay. Basing on our data set we analyse specific weather conditions which potentially impact flights through multiple linear regression [DS98]. The conditions are light rain, rain, heavy rain, haze/fog, thunderstorms, light snow and snow. Our findings indicate that there is no significant influence on delays for light rain, rain and heavy rain. However, haze/fog, thunderstorm, light snow and snow seem to have a noticeable impact on flight delays. We investigate these conditions to figure out in how far they explain the delay of a flight. As a result of this analysis we derive a factor which allows for a calculation of delay time. However, while light snow will lead to delays when it appears close to the airport, it does not influence the flight at all while the airplane is flying at 30,000 feet. We therefore consider weather conditions within different stages of a flight (close to an airport or *en route*).

Our investigation indicates that the conditions' impact increases significantly once they appear closer to the airports. It identifies four weather conditions which have a significant impact on flights. These conditions lead to different lenghts of delay, which are considered within the linear equitation to predict the delays for prospective flights.

A major limitation of our work is that the influence of wind has not been considered within the analysis. Wind can be an important factor during the landing procedure in which airplanes may not be able to reduce the speed to an optimal level due to tailwind. Additionally, wind at the cruise altitude may also be an important factor for delays. The results of our analysis are strongly dependent on the quality of the data. The historical data set of the flights includes the minutes of delays which are based on bad weather conditions. Few cases in the data set are flights which are stated to be delayed due to weather but do not show bad weather conditions according to the weather stations on their

[2]http:/www.ibm.com/software/analytics/spss/

way. Furthermore, in order to predict delays, the described model obtains weather data from another prediction model, namely the weather forecast. This forecast is afflicted with uncertainty and may lead to deviations in our prediction. Although the predictions seem to be precise for the flight in our scenario, the model needs to be tested with different departure and destination airports to enhance the universal usage of the model.

Keywords: Flight Delay Prediction, Aircraft, Prediction Model, Weather, Data Acquisition

Acknowledgements

The research leading to these results has received funding from the European Unionâs Seventh Framework Programme (FP7/2007-2013) under grant agreement 318275 (GET Service).

References

[DS98] Norman R. Draper and Harry Smith. *Applied Regression Analysis (Wiley Series in Probability and Statistics).* Wiley-Interscience, third edition, 1998.

References

Cabanillas, Cristina, Enver Campara, Claudio Di Ciccio, Bartholomäus Koziel, Jan Mendling, Johannes Paulitschke, and Johannes Prescher (2014). "Towards a Prediction Engine for Flight Delays based on Weather Delay Analysis". In: *EMoV*. Ed. by Anne Baumgraß, Nico Herzberg, Gerti Kappel, Jan Mendling, Andreas Meyer, and Stefanie Rinderle-Ma. Vol. 1185. CEUR Workshop Proceedings. CEUR-WS.org, pp. 49–51. URL: http://ceur-ws.org/Vol-1185/paper6.pdf.

BibTeX

```
@InProceedings{   Cabanillas.etal/EMoV2014:TowardsPredictionEngine,
  author        = {Cabanillas, Cristina and Campara, Enver and Di Ciccio,
                  Claudio and Koziel, Bartholom{\"{a}}us and Mendling, Jan
                  and Paulitschke, Johannes and Prescher, Johannes},
  title         = {Towards a Prediction Engine for Flight Delays based on
                  Weather Delay Analysis},
  booktitle     = {EMoV},
  year          = {2014},
  pages         = {49--51},
  publisher     = {CEUR-WS.org},
  crossref      = {EMoV2014},
  keywords      = {Flight delay prediction; Aircraft; Prediction model;
                  Weather; Data acquisition},
  url           = {http://ceur-ws.org/Vol-1185/paper6.pdf}
}
@Proceedings{     EMoV2014,
  title         = {Joint Proceedings of the 1st International Workshop on
                  Modeling Inter-Organizational Processes and 1st
                  International Workshop on Event Modeling and Processing in
                  Business Process Management co-located with Modellierung
                  2014, Vienna, Austria, March 19, 2014.},
  year          = {2014},
  editor        = {Anne Baumgra{\ss} and Nico Herzberg and Gerti Kappel and
                  Jan Mendling and Andreas Meyer and Stefanie Rinderle{-}Ma},
  volume        = {1185},
  series        = {{CEUR} Workshop Proceedings},
  publisher     = {CEUR-WS.org},
  url           = {http://ceur-ws.org/Vol-1185/paper6.pdf}
}
```