How does weather affect the use of public transport in Berlin?

K M Nissen, N Becker, O Dähne, M Rabe, J Scheffler, M Solle and U Ulbrich

1 Department of Meteorology, Freie Universität Berlin, Germany
2 Berliner Verkehrsbetriebe, Berlin, Germany
3 Now at UK Centre for Ecology and Hydrology, Penicuik, United Kingdom

Keywords: weather, public transport, statistical model

Supplementary material for this article is available online

Abstract

The effect of weather on public transport usage in Berlin is analysed. The number of single and day tickets sold is used as a proxy for the number of occasional public transport users. Analysing more than three years of hourly ticket sale data, it is shown that the most important factor influencing ticket sales is temperature. Temperatures below $-5^\circ\text{C}$ lead to an increase in ticket sales by up to 30% on working days, while on hot days ($>28^\circ\text{C}$) passenger numbers drop by up to 5%. Precipitation increases the number of sales on working days by up to 5%. On weekends, the lowest ticket-sale numbers are associated with wet and either very cold or very hot conditions. Another factor influencing ticket sales is sunshine duration, while wind and snowfall do not seem to play a role for ticket sales in Berlin. It is demonstrated that it is possible to predict ticket sales depending on date, time and weather conditions using a statistical model.

On designated public transport routes the effect of weather on passenger numbers can be much stronger than the district average. This is shown for the example of a bus route to a public beach. With each degree of temperature increase, passenger numbers on this line go up by approximately 30%.

1. Introduction

Berlin has approximately 3.7 million inhabitants and was visited by 13.5 Million tourists in 2018. According to a survey, around 27% of all trips within Berlin are made by means of public transport, 30% by car, 30% on foot and 13% by bike (Senatsverwaltung für Umwelt, Verkehr und Klimaschutz 2019b). When using public transport, people have the choice of taking the tram, bus, the underground, regional trains or city trains (S-Bahn). Additionally, there are even 6 ferry lines, of which 3 run only during the summer season. Tram, buses, underground and ferry are operated by the BVG (Berliner Verkehrsbetriebe), taking 1,102 million passengers on trips each year.

In order to provide adequate public transport capacity and frequency but remain economical, it is important to have realistic estimates of passenger numbers at a given time. Studies for other cities around the world have shown that weather can be an important factor influencing passenger numbers. For Chicago, for example, Chagnon (1996) found a 3–5% decrease of public transportation between 0500-2100 local time on rainy days. In another study for Chicago based on daily averaged data from an automatic fare collection system, (Guo et al 2007) find that precipitation and snowfall have the highest negative influence on public transport usage followed by wind speed, while there is a positive relationship between temperatures and passenger numbers. The influence of weather was found to be stronger at weekends than on weekdays. For Brisbane, Australia, (Tao et al 2018) analysed hourly data for a 3-month period and also found that the effect of weather on bus usage is higher on weekends than on weekdays, with warmer weather and a light breeze promoting and wet conditions reducing usage. A stronger effect of weather on weekends than weekdays was also reported by Singhal et al (2014) for New York City subway usage which was found to significantly decrease under adverse weather conditions including rain, wind and snow. Kalkstein et al (2009) analysed daily data of rail usage in 3 US cities. The effect of weather was again strongest on weekends with dry, comfortable air masses increasing and moist, cool air masses decreasing passenger numbers. In the state of Washington, adverse weather conditions also lead to
lower public transport usage (Stover and McCormack 2012). Similar weather factors play a role in Spain and Ireland. For Spain, (Arana et al 2014) show that the number of bus trips on weekends decreases in windy and rainy conditions, while a temperature rise leads to an increase. For an unnamed city in Ireland, a slight decrease of bus usage on rainy days was found (Hofmann and O’Mahony 2005).

The influence of weather on public transport usage in the above mentioned studies is moderate and often not even quantified. However, there are also cities in which this influence can become quite large. Such a city is Münster, a university city with 310,000 inhabitants which is located in the northwest of Germany. Münster has a high number of cyclists. Bicycles are used in 38% of all trips. The total number of bus passengers in Münster increases by 30% when it rains in the morning. When considering only the number of occasional bus users, the numbers even go up by 110% (Adorf et al 2019).

Thus, the effect of weather is not only stronger in Münster but also different in sign compared to the aforementioned studies with precipitation, leading to an increase in passenger numbers rather than a decrease.

In Berlin such an analysis is difficult as there is no automatic ticket control and passengers using public transport are normally not counted. However, each passenger over the age of 6 needs to buy a ticket prior to travelling and there is a comprehensive data set on ticket sales. In this study, we analyse this data set with respect to weather related variation. It can be assumed that the effect of weather on ticket sales is mostly reflected in spontaneous purchases of single-trip or day tickets. We will thus concentrate the analysis on these types of tickets. With this approach we will mainly capture variations in the number of occasional travellers as most regular travellers choose to buy season tickets. The analysis thus concentrates on occasional public transport users, as regular users normally buy season tickets. The ticket-sale data have an hourly temporal resolution. The period analysed in this study starts in January 2016 and ends in March 2019.

The second data set evaluated contains the number of passengers counted by APC systems fitted to buses. For each count, the bus stop location and the exact time are given. To exclude measurement errors, we only consider bus trips in which the total number of passengers entering and leaving the bus is balanced at the end of the trip. The APC data are analysed in order to compare the variability of occasional public transport users with total passenger numbers. This is done for two districts in Berlin (shaded regions in figure 1): Neukölln as an inner-city example and Spandau as a suburb which is mostly serviced by buses. For Neukölln, 203 bus stops served by 28 bus routes are included in the analysis. In Spandau, the data comprises 245 bus stops served by a total of 29 bus routes. In both districts the APC data are first aggregated to hourly values and then interpolated in time and space with the help of time tables to account for the passengers in buses without APC. For the interpolation we assume that the number of passengers in the buses without APC at a given hour is equal to the mean passenger volume inside the sampled buses. The percentage of bus trips for which APC data have successfully passed the quality check gradually increases from 0 before January 2018 to approximately 25% in May 2019 (not shown). Due to the short time series, the incomplete data set and the assumptions made for the interpolation, any statistical analysis based on this data has to be interpreted with care.

Absolute transport numbers are a relevant for business competition and will not be published here. Instead, normalized variations in the number of passengers and sales are taken as the basis of our analysis, i.e. actual passenger or ticket counts are expressed in percent of the respective long term median, which is quantified as 100%. We base comparisons on medians rather than means in order to limit the effect of outliers, e.g. due to major public events.

With respect to weather, data from several weather stations are available for the region (figure 1). Of these stations, 7 are operated by the German Weather Service (DWD) and 9 are operated by the Freie Universität Berlin. The water supplier for Berlin (Berliner Wasserbetriebe; BWB) additionally measures precipitation at 44 stations. The station-based data are further complemented by

2. Data

In this study we analyse two different public-transport data sets. The first one is the number of tickets sold at stationary ticket machines, in buses and from the BVG mobile phone app. We restricted the analysis to single-trip and day tickets, neglecting seasonal tickets. The analysis thus concentrates on occasional public transport users, as regular users normally buy season tickets. The ticket-sale data have an hourly temporal resolution. The period analysed in this study starts in January 2016 and ends in March 2019.

The second data set evaluated contains the number of passengers counted by APC systems fitted to buses. For each count, the bus stop location and the exact time are given. To exclude measurement errors, we only consider bus trips in which the total number of passengers entering and leaving the bus is balanced at the end of the trip. The APC data are analysed in order to compare the variability of occasional public transport users with total passenger numbers. This is done for two districts in Berlin (shaded regions in figure 1): Neukölln as an inner-city example and Spandau as a suburb which is mostly serviced by buses. For Neukölln, 203 bus stops served by 28 bus routes are included in the analysis. In Spandau, the data comprises 245 bus stops served by a total of 29 bus routes. In both districts the APC data are first aggregated to hourly values and then interpolated in time and space with the help of time tables to account for the passengers in buses without APC. For the interpolation we assume that the number of passengers in the buses without APC at a given hour is equal to the mean passenger volume inside the sampled buses. The percentage of bus trips for which APC data have successfully passed the quality check gradually increases from 0 before January 2018 to approximately 25% in May 2019 (not shown). Due to the short time series, the incomplete data set and the assumptions made for the interpolation, any statistical analysis based on this data has to be interpreted with care.

Absolute transport numbers are a relevant for business competition and will not be published here. Instead, normalized variations in the number of passengers and sales are taken as the basis of our analysis, i.e. actual passenger or ticket counts are expressed in percent of the respective long term median, which is quantified as 100%. We base comparisons on medians rather than means in order to limit the effect of outliers, e.g. due to major public events.

With respect to weather, data from several weather stations are available for the region (figure 1). Of these stations, 7 are operated by the German Weather Service (DWD) and 9 are operated by the Freie Universität Berlin. The water supplier for Berlin (Berliner Wasserbetriebe; BWB) additionally measures precipitation at 44 stations. The station-based data are further complemented by
rare measurements for precipitation. Here we use the German Weather Service’s RADOLAN product which is available on a 1x1 km grid and has an hourly temporal resolution. For this product radar reflectivity is converted to precipitation amounts adjusted to station measurements (Weigl and Winterrath 2009). RADOLAN is used to calculate regional averages of precipitation amounts and to estimate the percentage of a region affected by precipitation. A time step is categorized as wet if precipitation was recorded by one or more stations and/or RADOLAN within the region of interest. For the present study, we have analysed the effect of weather in terms of temperature, various precipitation measures, wind speed, sunshine duration and snowfall.

As one of the probable reasons for variations in public transport usage is a passenger shift to travel by bicycle, we also analysed data from the 17 automatic bike traffic counting locations which exist in Berlin (Senatsverwaltung für Umwelt, Verkehr und Klimaschutz 2019a).

3. Variability unrelated to weather

The dominant part of the variability both for ticket sales and for passenger counts is due to factors unrelated to weather such as the time of the day and the day of the week (figure 2). Most weekdays (Monday to Thursday) show a similar daily cycle. On Fridays and Saturdays the nighttime hours are more busy than usual (not shown). Sundays exhibit the lowest number of ticket sales as well as the lowest passenger volume. An additional factor are school holidays. Here the effect is different for ticket sales and passenger counts. While ticket sales increase slightly during school holidays, passenger counts go down significantly, especially during the peak hour between 7 and 8 a.m., caused by the absence of pupils on their way to school.

4. The effect of weather

In order to extract the effects of weather from the variability, it is necessary to eliminate the variability due to non-weather-related factors. This can be done, for example, by stratifying the data according to the type of day and the time before comparing two different weather situations. Examples for such a comparison are shown in figure 3. The difference between the medians of ticket sales for two groups of time steps is depicted, those with and without a certain weather condition (e.g. wet/dry). The statistical significance of the differences is tested with a bootstrapping technique for which we randomly split the sample into two groups. The groups include the same number of time steps as the data sets that will be tested (e.g. size of group 1 = number of time steps with precipitation, size of group 2 = all other time steps). The difference in the medians of the random groups is computed and the process is repeated 1000 times. The range of differences covered by 95% of the randomly obtained results is regarded as the range in which differences are not statistically significant. It is indicated by shading in figure 3.

Precipitation and temperature influence the number of ticket sales. Wet conditions increase the number of tickets sold on weekdays (figure 3(a)). On weekends, the effect of precipitation is reversed, with less tickets being sold during wet hours (not shown). On hot summer days, less passengers buy a ticket during the afternoon hours. Figure 3(b) shows the effect of daily maximum temperature $> 28$°C for the summer school holiday season, which includes most of these hot days. During very cold periods (e.g. temperatures $< -5$°C), ticket sales increase considerably (figure 3(c)). No statistically significant changes could be found for snowfall (not shown). The depicted weather related signals on ticket sales are robust with respect to the district that was analysed and to the means of transport (i.e. bus vs. underground): changes in ticket sales due to temperature variations and precipitation remain qualitatively the same when analysing only sales in buses or when analysing the districts of Spandau and Neukölln separately (not shown).

In order to determine if the response of occasional public transport users to weather is also representative for absolute passenger numbers, the analysis was repeated using APC data for Spandau and Neukölln. The results can be found in the online supplement to this article (stacks.iop.org/ERL/15/085001/mmedia). Due to the short APC time series the analysis turned out to be inconclusive as the results were mostly not statistically significant.

A comparison of medians is descriptive but only possible for categorical variables. In order to depict
Figure 2. Diurnal cycle of the median of ticket sales across Berlin (top panels (a), (b)) and passenger counts in Spandau (bottom panels (d), (c)) on Monday to Thursday working days except public holidays (left panels (a), (c)) and on Sundays and public holidays (right panels (b), (d)). Individual hourly numbers (%) are given in relation to the theoretical hourly median value (100%). The latter is estimated by dividing the long term median daily value by 24 hours.

the effect of temperatures, arbitrary thresholds (28°C and −5°C) had to be chosen. To analyse the effect of continuous weather variables such as sunshine duration, wind speed, precipitation amount, area affected by precipitation, and temperature, a statistical model as the one described in the next section is better suited.

5. Statistical model

We first reduce the non-weather related variability within the data by removing the variability that can be determined by ‘calendar’ variables (table 1). This is motivated by the fact that such variations can be accounted for by suitable timetables. To isolate the non-weather related variability, a generalized linear model is fitted to the data using the statistical software package R (R Core Team 2018). We chose a Poisson regression which is an accepted choice to model count data. Day of the week, time of the day and holidays are treated as categorical variables. All dependencies between the above mentioned calendar variables are considered (e.g. the effect of holidays is different on work days and weekends). This results in 192 regression coefficients that need to be determined. Two additional regression coefficients (βsin and βcos) describe the seasonal cycle (e.g. caused by seasonal differences in the number of people buying a seasonal travel pass) which is modelled as βsin(sin(ωt)) + βcos(cos(ωt)) with t being the day of the year and ω = (2π)/365.25. In the following, we will refer to this model as the calendar model.

In the next step, we analyse to what extent the residuals (deviations of the observations from the calendar model results) are due to the effects of weather by fitting a multiple linear regression model to the residuals of the calendar model. This time we choose weather parameters as independent variables. The dependent variable is expressed in percent of the anticipated numbers ((countdata − countdatafitted) * 100/countdatafitted) instead of using the absolute deviations as we want to avoid putting a higher weight on certain hours and days (i.e. absolute residuals tend to be large during working hours on weekdays when the absolute number of sales is high). Calendar variables on their own should thus have no effect on the result. Dependencies between the different weather variables and between weather variables and calendar variables are, however, taken into account. With this approach, we consider the fact that people’s reaction to different weather may depend, for example, on the day of the week. When travelling to school on a weekday, pupils may switch from using the bike to using the bus when it rains but may decide to stay at home entirely on a Sunday. We tested different weather variables as predictors (table 1) using cross-validation. For cross-validation we removed seven days (four Monday–Thursdays, one Friday, one Saturday, one Sunday) from each month of the data set. The statistical model was fitted to the remaining data and then used to
forecast the days we previously removed. This test was repeated 10 times and the average RMSE (root mean square error) of the predictions was used to judge the performance of the model. We always compared both the RMSE of the percentual and the absolute deviations. We also tested if there are superfluous interaction terms between variables which can be removed from the statistical model. This was done by comparing three different levels of complexity:

1) The full model with all interactions.
2) Removal of predictors based on the Akaike Information Criterium using the function stepAIC and the penalty factor $k = 2$ (NCAR-Research Applications Laboratory 2015),
3) as 2) but using the stronger penalty factor of $k = \log(n)$, with $n$ being the number of time steps.

We used the RMSE from the cross-validation process to decide which model-complexity performs best. As keeping all predictors from the full model results in overfitting, it was always identified as the worst option. The strongest improvements in terms of the RMSE were found for option 3).

When modelling ticket sales for the entire region of Berlin, the calendar model has a Pseudo-$R^2$ value of more than 0.98 and thus already explains most of the data variability. When it comes to the effect of weather on ticket sales, the most important variable is temperature. The best of all tested statistical models includes both a linear and a quadratic temperature term. The best result is obtained by using temperature anomalies $(temperature - mean(temperature))$.

Removal of predictors based on the Akaike Information Criterium using the function stepAIC and the penalty factor $k = 2$ (NCAR-Research Applications Laboratory 2015),

$mean(temperature) \approx 11{°}C$ for the linear term. The quadratic term gives the best results when subtracting 18° C. The value of 18° C was determined by systematically testing different offsets. Including this term weakens the response of the ticket sales on temperature at moderate temperatures (the effect can be seen in figure 5). With respect to precipitation, we found that the binary information of whether it rains or not performs better than precipitation amount or area affected by precipitation. The information of whether it rains in the morning between 6 a.m. and 10 a.m., which was reported to be an important factor in the study conducted for Münster (Adorf et al 2019), turned out to have a weaker effect than instantaneous rain. Sunshine duration also has an effect on ticket sales. Here systematic testing of different offsets revealed that we obtain the best results when we subtract 10 to 20 minutes from the observed values. For the statistical model we use 15 minutes (table 1). One can speculate that this improves the results because people that base the decision of whether to travel or not on sunshine duration, distinguish between substantially sunny and mostly cloudy conditions rather than absolutely clear and completely overcast. By subtracting 15 minutes from the sunshine duration the response changes sign for sunshine durations above/below 15 minutes. Under normal conditions wind speed has no effect on ticket sales. The only exception within the analysed period was windstorm Xavier (October 5th, 2017) which led to a shutdown of the entire bus-service in the afternoon (not shown). Overall, the statistical model taking weather variables into account improves the prediction of the percentual residuals by 8% and the error of the predicted absolute sales by 5%.

An example for the performance of the two model steps to capture sales variability is depicted in figure 4, showing two weeks of the time series. The first Monday within this period, the 21st of May 2018, was a public holiday and is thus treated as a Sunday by the statistical model. Other than that, Berlin had no further school holidays within this period. During the two weeks, temperatures did not fall below 10° C and rose up to 32° C. The first week of the time series was mostly dry. More hours with precipitation were observed during the last 5 days. It can be seen that the prediction of the calendar model (green curve) already captures most of the observed variability (black curve). By design, it predicts the same number of sales for Monday–Thursday. During weekdays the afternoon peak in sales within the period is often overestimated by the calendar model as the high temperatures during the period lead to a

![Figure 3. Influence of (a) precipitation, (b) high temperatures and (c) low temperatures on the diurnal cycle of ticket sales in Berlin during weekdays (Monday to Thursday). The vertical axis shows normalized differences between the median values during hours fulfilling the respective criterion from those not fulfilling it. (a) Difference between rainy and dry hours, (b) anomaly for days on which the temperature exceeds 28° C (analysis includes only summer school holidays), (c) anomaly at hours on which the temperature falls below −5° C during the winter month (DJF). Shaded area marks the range in which anomalies/differences are not statistically significant on the 95% level.](image-url)
Table 1. Variables tested as predictors in the statistical models. Variables which are included in the best performing model are printed in bold face and in the form they have been included.

| calender-variables                  | weather-variables                  |
|-------------------------------------|-------------------------------------|
| day of week (Monday–Thursday, Friday, Saturday, Sunday) | temperature anomaly |
| hour of day                          | precipitation yes/no               |
| school holidays yes/no               | (temperature-18°C)^2               |
| seasonal cycle                      | sunshine duration - 15 minutes     |
|                                     | precipitation amount averaged over the region |
|                                     | area affected by precipitation in percent |
|                                     | precipitation between 6 a.m. and 10 a.m. |
|                                     | wind speed                         |
|                                     | snowfall yes/no                    |

Figure 4. Top: Two-week time series of tickets sold (black), prediction of sales by calender model (green), prediction of sales by adding the result of the statistical model with weather variables to that of the calender model (red). Bottom: Temperature (brown) and presence of precipitation (blue) within this two-week period.

Figure 5. Influence of temperature and precipitation (no precipitation = orange; precipitation = blue) on ticket sale prediction for 2 p.m. on a workday (solid line) and on a Sunday (dashed line). The spread marks variations that are possible due to the seasonal cycle and differences in sunshine duration. Median of observations within 5°C bins are marked with symbols (dot for weekdays, triangle for Sunday) and the maximum and minimum values are connected by vertical lines.

reduction in ticket sales. This is improved by taking weather variables into account (red curve). On the three Sundays/public holidays, the peak in ticket sales is slightly underestimated by the calendar model. This is improved by, again, taking weather into account.

The unusually high number of ticket sales between the 31st of May and 2nd of June under hot/warm and rainy conditions cannot be explained by our statistical model and is possibly not weather related. A plausible reason may be that the 31st of May was
a public holiday in some federal states in Germany (not in Berlin) leading to an increased number of visitors, who extended their weekend by taking an extra day off.

The effect of temperature and precipitation on sales in the afternoon is visualized as a showcase in figure 5. The response of the statistical model to different temperatures is shown under dry (orange) and wet (blue) conditions for weekdays (Monday–Thursday; solid lines) and Sundays (dashed lines) at 2 p.m. The possible range of values due to variations in the sunshine duration and the seasonal cycle is indicated by shading. The median of the observations within 5 °C temperature bins (symbols) and the range between the highest and the lowest observed value (vertical lines) are also shown on the plot as a reference. Higher temperatures lead to a decrease in the number of tickets sold on weekday afternoons, explaining the improvements seen for many afternoons in figure 4. On the 30th of May the rapid temperature drop in the afternoon caused the model to predict close to average ticket sales, however the number of ticket sales remained low. This suggests that one might consider adding a persistence term to the statistical model. Precipitation under moderate-temperature conditions leads to a small increase in ticket sales. Corroborating graphs can be drawn for other times of the day. They indicate, for example, that during the night, sales increase with increasing temperatures both during the week and during the weekend (not shown).

The full set of regression coefficients for the weather related predictors is included in the supplement to this article.

6. Special routes

In the previous sections we determined an area average response of ticket sales and passenger volume to weather variations. There is also a small number of specific routes for which one can expect a much stronger dependency between traveller numbers and weather. An example for such a designated route is the bus service between the city train station Nikolassee and the public bathing beach in Wannsee. The beach is located in the South-West of Berlin and the bus line, which has only two stops and operates only during summer holidays, is marked in orange in figure 1. Data were available for the summer holiday seasons of 2018 and 2019. During that period an average of 18% of the buses were equipped with APC technology. The daily median of the number of passengers per bus on route to the beach (expressed in percent of the average median) and its dependence on the daily mean temperature at the closest weather station is shown as dots in figure 6. It can be seen that more people use the bus line on weekends (blue dots) than during the week (red dots). The relationship between temperature and passenger numbers can be quantified by fitting a Poisson regression model to the observations (red and blue line). The number of passengers strongly increases with temperature at a rate of $e^\beta$.

With each degree of temperature increase the number of passengers increases by approximately 25% and 33% on weekends and weekdays, respectively. The observation period includes only three days with rain (definition used here: longer than 15 minutes during the opening hours of the beach) for which APC data are available. They are marked with a star in figure 6. The daily mean temperature for these days is around 20 °C and the median of passengers is low. For a quantitative evaluation of the effect of precipitation, a longer time series with more rainy days would be needed.

7. Discussion and conclusions

In this study we have analysed the effect of weather on ticket sales and passenger numbers in Berlin. Prior to analysing the weather effects, time-related variability was quantified with a calendar and time model that disregards weather. Time of the day, day of the week, season and differences between school days and holidays are the main factors influencing the variability.

The number of single-trip or day tickets sold can be seen as a proxy for the number of occasional passengers. When integrating over the entire city domain, the main weather effects for these type of customers are:

- Increase in the number of tickets sold by up to 5% when it rains on weekdays.
- Increase in the number of tickets sold with decreasing temperatures on weekdays (up to 30% on extremely cold days with temperatures $< -5 ^\circ C$).
- Increase in the number of tickets sold with increasing temperatures on Sundays under dry conditions (12% increase between 0 °C and 20 °C).
- Wet and cold conditions lead to the highest/lowest ticket sales on weekdays/Sundays, respectively.
• Increase in ticket sales with increasing temperatures during the night on all days.

The sign and size of the changes depend on the combination of temperature, precipitation, sunshine duration, day of the week and time of the day, as well as on holidays. In this study we built and tested a generalized linear model that is able to determine expected sales for any combination of these factors.

The response of occasional public transport users in Berlin to weather differs from that reported in many of the published studies mentioned in the introduction. These mostly report a decrease in passenger numbers under rainy and/or cold conditions. In Berlin, the effect of precipitation on the number of occasional travellers depends on the type of day. On weekdays, i.e. the majority of days, the number of occasional travellers rises under wet conditions. A decrease can only be seen during the warm season on weekends.

Uncomfortable weather can change the travel behaviour of people in two different ways. They can chose to stay at home or they can chose to switch modes of transport. In Berlin, the effect on weekdays seems to be mainly a switch in the mode of transport. Most likely the switch is from foot and bike to public transport. A decrease in bike traffic with decreasing temperatures and in wet conditions is indeed measured at the automatic bike counting points in Berlin (not shown). This is also in line with results from Münster (Adorf et al. 2019). A transition from non-motorized traffic (bike, foot) to motorized traffic (mostly private cars) under wet conditions was also found for the Netherlands (Sabir 2011). In the cities examined in the US and Australia, in contrast, people seem to prefer staying at home or switching to private cars, thus reducing the number of passengers in local public transport under inclement weather conditions.

On weekends, our results suggest that warm and dry weather encourages people to make a trip, using public transport. Thus, for weekends our results agree qualitatively with those of the other studies cited in the introduction. The strength of the weather effect between weekdays and weekends, is similar in our study, which is probably due to the fact that the ticket sales data set represents occasional public transport users rather than absolute passenger numbers.

The quantitative effect of weather on ticket sales is much weaker than for the smaller city of Münster, which is known as a cycling hot spot. While 38% of all distances in Münster are travelled by bike, this share is only 13% in Berlin. The effect of a changeover from bicycle to public transport must therefore be lower in Berlin.

Whereas the ticket sales are a complete inventory for a period of more than three years, the APC data available at the time of this study is a random sample of data with non-constant coverage for 17 months. Due to the short time series and the unusually dry weather conditions during the sampling period, it was not possible to determine if total passenger volume shows a similar response to weather as the number of occasional public transport users. Currently the APC data time series is also too short and sparse to use for the development of a statistical model analogous to the one we build for the calculation of anticipated ticket sales. However, as the number of vehicles fitted with APC systems is continually increasing, such a model can be developed in the future. Such a statistical model can then be used to predict passenger volumes using weather forecasts in order to adjust capacity based on expected demand. To some extent, adjusting the transport capacity to weather conditions is possible without extra costs by a directed distribution of buses with different passenger capacities according to the predicted passenger volume on specific lines. Adjusting the frequency of vehicles is currently not economically viable as the duty rosters for the personnel need to be planned weeks ahead and weather can only be forecasted at sufficient accuracy days in advance. Short term adjustments to the time tables might be possible in the future though, if/when autonomous driving vehicles replace conventional buses and trains. Research on such systems is already under way (BVG 2019).

Acknowledgments

We would like to thank the anonymous reviewers for their constructive comments, that helped to improve the manuscript. This work was conducted within the research project ÖPNV-WETTER which was funded by the Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) under the mFUND program contract numbers 19F1053A and 19F1053B.

Data availability

Weather data from the DWD are available from www.dwd.de/EN/ourservices/opendata/opendata.html. The other data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID iD

K M Nissen https://orcid.org/0000-0003-1132-9701

References

Adorf L, Klemm O, Rose P and Jöne S 2019 Einfluss des Wetters auf die Fahrgastzahlen, Einflussfaktoren auf die Nutzung des ÖPNV in Münster Der Nahverkehr 4 (https://eurailpress-archiv.de/SingleView.aspx?show=876126)

Arana P, Cabezudo S and Penalba M 2014 Influence of weather conditions on transit ridership: A statistical study using data from Smartcards Transport. Res. A-Policy Practice 59 1–12
BVG 2019 Kleiner Gelber fährt sich selber (www.bvg.de/de/Aktuell/Newsmeldung?newsid=3379) (Accessed: 15 November 2019)

Chagnon S A 1996 Effects of summer precipitation on urban transportation Clim. Change 32 481–94

Guo Z, Wilson N H M and Rahbee A 2007 Impact of weather on transit ridership in Chicago, Illinois Transport. Res. Rec. J. Transport. Res. Board 2034 3–10

Hofmann M and O’Mahony M 2005 The impact of adverse weather conditions on urban bus performance measures – an analysis using ITS technology Conf. on Intelligent Transport Systems (Vienna, Austria, 16 September 2005) (https://doi.org/10.1109/ITSC.2005.1520087)

Kalkstein A J, Kuby M, Gerrity D, Clancy J J and James J 2009 An analysis of air mass effects on rail ridership in three US cities J. Transport. Geogr. 17 199–207

NCAR - Research Applications Laboratory 2015 Verification: Weather Forecast Verification Utilities R package version 1.42 (https://cran.r-project.org/web/packages/verification/verification.pdf)

R Core Team 2018 R: A Language and Environment for Statistical Computing (Vienna: R Foundation for Statistical Computing) (https://r-project.org/)

Sabir M 2011 Weather and travel behaviour PhD thesis Vrije Universiteit (https://research.vu.nl/en/publications/weather-and-travel-behaviour)

Senatsverwaltung für Umwelt, Verkehr und Klimaschutz 2019a Automatische Dauerzählstellen für den Radverkehr in Berlin (www.berlin.de/senuvk/verkehr/lenkung/vlb/de/radzahlungen.shtml) (Accessed: 15 November 2019)

Senatsverwaltung für Umwelt, Verkehr und Klimaschutz 2019b Zahlen und Fakten zum Verkehr, (https://www.berlin.de/senuvk/verkehr/politik_planung/zahlen_fakten/mobilitaet_2013/) (Accessed: 15 November 2019)

Singhal A, Kamga C and Yazici A 2014 Impact of weather on urban transit ridership Transport. Res. A 69 379–91

Stover V W and McCormack E D 2012 The impact of weather on bus ridership in Pierce County, Washington J. Public Transport. 15 95–110

Tao S, Corcoran J, Rowe F and Hickman M 2018 ‘To travel or not to travel: ‘Weather’ is the question. modelling the effect of local weather conditions on bus ridership Transport. Res. C 86 47–167

Weigl E and Winterrath T 2009 Radargestützte Niederschlagsanalyse und –vorhersage (RADOLAN, RADVOR-OP) promet 35 78–86