

Article

Modelling the impact of weather and context data on transport mode choices: A case study of GPS trajectories from Beijing

Timothy Otim ^{1,†,‡} , Leandro Dörfer ^{1,†,‡}, Dina Bousdar Ahmed ^{1,†,‡}, and Estefania Munoz Diaz ^{*1,†,‡}

¹ Affiliation 1; {timothy.otim, leandro.doerfer, dina.bousdarahmed, estefania.munoz}@dlr.de

* Correspondence: estefania.munoz@dlr.de

† Current address: Institute of Communications and Navigation, German Aerospace Center, Oberpfaffenhofen 82234, Germany

‡ These authors contributed equally to this work.

Abstract: Over the years, researchers have been studying the effect of weather and context data on the transport mode choice. The majority of these works are based on survey data, however the accuracy of their findings relies on how respondents give accurate and honest answers. In this paper, the potential of using GPS trajectories as an alternative to travel surveys in studying the impact of weather and context data on transport mode choices is investigated in Beijing city. In the analysis, we apply both descriptive and statistical models such as the MNL and MNP models. Our findings indicate that temperature has the most prominent effect among weather conditions. For instance, for temperatures greater than 25 C, the walking share increases by 27% and the bike share reduces by 21%, which is line with the results from several survey studies. In addition, the evidence of government policy on transport regulation is revealed when the air quality becomes hazardous as people are encouraged to use environmentally friendly travel mode choices such as the bike instead of the bus and car, which are known CO₂ emitters. Moreover, due to a series of traffic restrictions introduced by the Beijing government during the 2008 summer Olympics, a decrease of 17.5% in the car share and an increase of 13% and 10% in the walking and bus shares, respectively are observed. These findings provide a scientific basis for effective transport regulation and planning purposes.

Keywords: Air quality; Geolife; Olympics; Traffic demand; Transport planning; Transport regulation

1. Introduction

At present, the transport sector accounts for 29% of all world greenhouse gas emissions, and 23% of global carbon dioxide (CO₂) emissions, consequently contributing towards the global climate change [1]. Climate change mitigation requires finding ways of achieving sustainable mobility options without compromising economic growth and social inclusion, which necessitates effective government planning and regulation of the transport sector. Therefore, it is crucial to investigate the scientific evidence which shows how changes in climate conditions, transport planning, and transport regulation are intrinsically interrelated.

With climate change, weather is now an important topic in transportation research. Over the years, researchers have been studying the effect of weather and context on the transport mode choice. We define context data as information that can provide perspective into a person or an event such as culture and habits. These studies relate context information such as individual characteristics [2], temporal factors [3], transportation supply [4], travel demands [5], and weather conditions [6] to existing or self-gathered travel behaviour data. Most of the works focus on weather effects on transport mode decisions across different cities [6–8].

The existing literature on weather and daily mobility focus particularly on precipitation [9–13], temperature [14–16], humidity [17,18], wind speed [18], air quality [16,18], seasons [19]. The findings of these studies give insights into the separate roles of not only weather parameters, but also the geographical context, cultures and habits in influencing transport mode choices. For instance, the authors in [15,20] indicate that in the Netherlands, higher but not too high temperature favours walking and biking over motorised transport,



Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Preprints* 2022, 1, 0. <https://doi.org/>

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

hence, temperature is more important than precipitation. In contrast, researchers in [9,10] conclude that precipitation creates more significant ridership fluctuations than temperature in Nanjing, China and Flanders, Belgium.

In search for a better understanding of how the geographical context and habits affect mobility decisions, Böcker et al. [12] investigates weather and daily mobility across Dutch, Norwegian and Swedish cities. The authors find that biking is favoured by dry and warm conditions and the use of the car is favored during wet and windy weather. The authors also highlight the presence of differences in the effects of weather on mobility across the different cities and countries.

With respect to seasonality, Hyland et al. [19] reports that it is highly likely for the car to be the chosen mode of transport during bad weather for commuters in Chicago. In fact, the authors show that millennials are inclined to choose the car in winter than the summer, while for non-millennials, seasonality had little impact on the choice of mode of transport. In [21], Ton et al. further highlights that apart from seasonality and weather characteristics, other categories of context information such as work conditions, trip and household characteristics, built environment do also influence the transport mode choices. In [22], Böcker et al. investigates how emotional travel experiences influence transport mode choices and finds that biking is influenced by sunny dry, calm, and warm but not too hot weather conditions, hence leading to more satisfying emotions.

While the aforementioned studies improve our understanding of weather and context data effects on daily mobility choices, one key issue stands out, many of the studies are based on survey or diary travel for characterising individual travel behaviour. The disadvantage of surveys is that the accuracy of their findings relies on how respondents give accurate and honest answers. However, information sources such as smartphones which can log satellite positioning data can now be used as an alternative to the surveys in transportation research. To the best of our knowledge, there is no study that analyses the relationship among weather conditions, context information, and different transport modes using satellite positioning data.

The goal of this paper is to show the possibility of using GPS trajectories in investigating the impact of weather and context data on transport mode choices in Beijing city. In this paper, not only do we study the impact of weather conditions such as temperature, precipitation, relative humidity, wind speed, and air quality on transport mode choices, but also the effect of context information such as rush hours, holidays, day/night, an event such as the Olympics, and trip distance on individual travel behaviour is presented. We apply both descriptive and statistical models such as the multinomial logit (MNL) and probit (MNP) models. The specific contribution of this paper is two fold:

1. The potential of using GPS trajectories to analyse and model the relationship between transport mode choices, weather and context information is investigated.
2. The relationship among weather and context information, transport planning, and transport regulation is analysed.

The structure of the remainder of this paper is as follows. Section 2 presents the databases applied in this work. The statistical analysis and models are described in Section 3 and Section 4, respectively. The results are discussed in Section 5 while the conclusions and future work are presented in Section 6.

2. Databases

The case study is Beijing [see – Figure 1], which is the capital of the People's Republic of China. With a population of over 21 million residents, it is regarded as the most populated capital city in the world. Due to its high population density, rapid urbanization and motorization, Beijing faces severe congestion and air quality problems.

Beijing is an interesting case study due to the sheer number of solutions that have been adopted to overcome its transport related problems such as the development of bus rapid transit corridors, new extensive bus lanes, policies like congestion charging, among

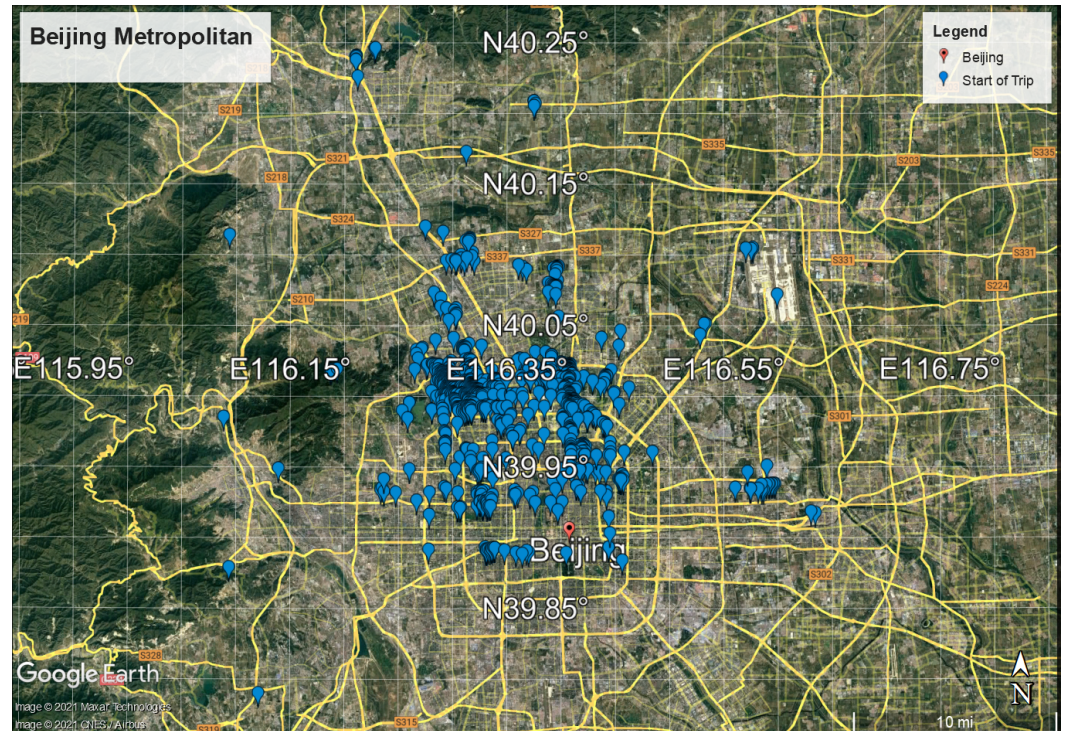


Figure 1. Geographical location and demarcation of Beijing. Also shown is the distribution of points corresponding to the start of each trajectory from the Geolife dataset.

others. In this section, we describe the three databases with the GPS trajectories, weather and context information in Beijing.

2.1. GPS trajectories

The dataset adopted in this study has GPS trajectories of 182 users collected during the GeoLife project conducted between April 2007 and August 2012 [23]. The GPS trajectories were recorded by GPS loggers and phones and are described by a sequence of time-stamped points, with each containing longitude, latitude, altitude, and transportation mode label. Although the dataset contains trajectories distributed in over 30 cities worldwide, in this work, we use a total of 2,671 labeled trajectories from Beijing. The total traveled trip distance and duration is 21350 km and 1296.6 h, respectively. Each traveler has an average of 74 trips, with an average distance and duration of 5.75 km and 0.5 h, respectively. In Figure 1, we show the geographical location and demarcation of Beijing and the starting point of the trajectories.

The transport modes considered in this work include: walk, bike, car, bus, and train. The percentage of the transport mode in the dataset is shown in Figure 2. During the period in which the data was collected, walking accounts for 46% of the transport mode labels in Beijing. This result is expected given that travel guides recommend getting around Beijing on foot as the best and most efficient commuting mode [24].

It is worth noting that the Geolife trajectory dataset is natural since it was recorded while the users performed their life routines [25]. For instance, trajectories were recorded as users made trips from home to work and vice-versa, to entertainment and sports venues, shopping, sightseeing.

The Geolife trajectory dataset has been used in different research fields such as in privacy preserving location data [26], measuring trajectory stops and moves [27], user identification [28], trajectory completion [29], and transport mode detection [30].

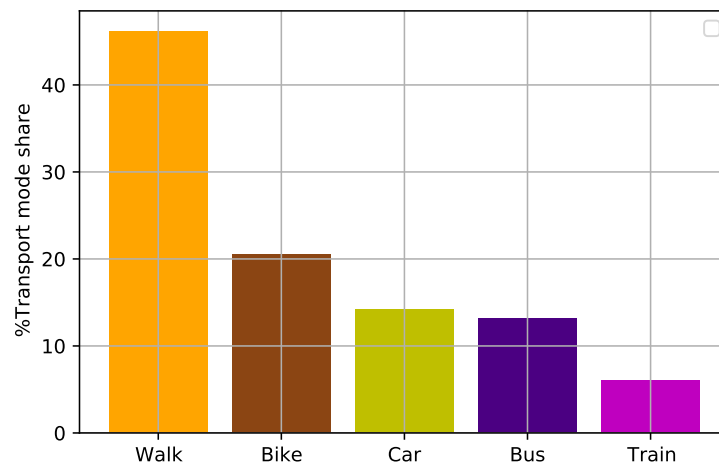


Figure 2. Modal share per transport mode in the dataset. A total of 2,671 trajectories were recorded from which the modal share of each transport mode was computed.

2.2. Weather information

The weather conditions considered include, temperature, precipitation, wind speed, air quality, and relative humidity, collected from a meteorological station located within the city of Beijing. This database can be accessed from the National Aeronautics and Space Administration (NASA) website [31].

Thanks to its continental climate, Beijing has hot, sultry, and rainy summers, cold and sunny winters, and a precipitation of about 545 mm annually [32]. However, a summary of the conditions in Table 1 shows that during the period of study: i) average temperature was about 19C, ii) the rains were not abundant with a mean of 0.083 mm/h, iii) the air was less humid with a mean relative humidity of 53%, iv) the wind speed was very light according to Beaufort scale given the mean wind speed is 3 m/s, and v) the air quality was unhealthy given that the mean air quality was $153 \mu\text{g}/\text{m}^3$.

Table 1. Descriptive characteristics of continuous observable variables

	Mean	Std.dev	Min	Max
Temperature (C)	19.153	8.477	-9.22	38.45
Precipitation (mm/h)	0.083	0.307	0	5
Relative Humidity (%)	53.837	22.076	6.25	100
Wind speed (m/s)	3.064	1.742	0.05	12.37
Air quality ($\mu\text{g}/\text{m}^3$)	153.016	64.545	13	547
Distance (km)	7.993	12.341	0	117.216

2.3. Context information

This study also includes the impact of context information such as the Beijing 2008 summer Olympics event held from 8th to 24th August on transport mode choices. This event was awarded to Beijing in the year 2001, thereafter its leaders embarked on massive projects to transform the city's transport system. For instance, the rail network was expanded from 50 km to 200 km and there was an introduction of 286 km of dedicated on-road lanes [33]. In fact, a new set of restrictions were passed months leading to the games. For instance, in some parts of the city, people were not allowed to use their private vehicles, trucks from outside Beijing were to avoid the city, flexible retail and shopping hours were also introduced to spread traffic loads.

Other context information attributes we look at include, rush hour or off-rush hour, holiday or non-holidays, day time or night time, and trip distance. In Beijing, morning and evening rush hour traffic demand is from about 7AM to 9AM and 4PM to 8PM, respectively according to the studies made by [33], which has been applied during our analysis. As for the holidays, there are 10 public holidays spread throughout the year, which include: New Year's, Chinese New Year, Lunar New Year, Qingming Festival, Labour Day, Dragon Boat Festival, Mid-Autumn Festival, Golden Week, and Christmas [34]. In total, our analysis considers 28 days of the year as public holidays. Day time in Beijing is considered as the period between sunrise, i.e. 6AM and sunset, i.e. 6PM compared to night time, which is between sunset and sunrise.

2.4. Matching weather and context information with GPS trajectories

As mentioned earlier, the GPS trajectories from the Geolife dataset contain the location, date, time, and the mode of transport chosen by the traveler. The starting point of the recorded trajectory provides the departure information of the traveler such as their departure location, date, and time.

In matching the weather data to the GPS trajectories, the starting point of each trajectory is linked to the hourly historical weather data from the nearest meteorological station. The result generates weather-related variables corresponding to the departure time of each trip. It is worth noting that hourly weather data are preferred over daily weather data because the former generates higher temporal accuracy, given the constantly changing weather.

Similarly, the timing corresponding to each of context information attributes is linked to the departure time of each trip. From the combined datasets, we analyse the impact of weather and context data on transport mode choices by considering weather and context information as input and modal choices as output conditions.

3. Statistical analysis

In this section, the weather conditions, context information and GPS trajectories are analysed statistically. We apply two main statistical methods which are commonly used in data analysis: i) descriptive statistics, where we summarize the findings using bar graphs, and ii) statistical interference, where we apply the MNL and MNP models.

In order to track the weather changes, we classify the weather conditions into different levels according to their rating generated from published weather knowledge. For instance, the temperature scales with seven levels observed in Figure 3a are adopted from [35,36], precipitation with five levels in Figure 3b from [37,38], wind speed with six levels of classification in Figure 3c – are adopted according to Beaufort number [39], air quality yardstick with six levels that run from 0 to 500 in Figure 3d adopted from monitoring the fine particulate matter concentrations ($PM_{2.5}$) [40], and relative humidity in Figure 3e falls into four levels adopted from [41]. These classifications as well as the timing corresponding to the context information are used to compute the share for each transport mode in Figure 3 and Figure 4 due to the effects of weather and context data, respectively.

3.1. Influence of the weather condition on the transport mode choices

In Figure 3a, it is observed that during snowfall, cars have a higher share compared to other transport modes. However, walking becomes the most preferred transport as the temperature increases. At temperatures greater than 25C, the bike and walking shares reduce and the bus share increases. In Figure 3b, we see that the share of the transport modes is similar across the precipitation levels except when precipitation is greater than 2 mm/h, i.e. little rain hardly impacts passengers' travel choices, which is in line with the observations of Junlong Li et al. [9].

In Figure 3c, it is observed that as the intensity of the wind increases from calm, greater than 0.27 m/s to strong wind, greater than 7.8 m/s, the bike and walk share reduce, while the car share increases. In Figure 3d, it is noticeable that as the air quality turns from good

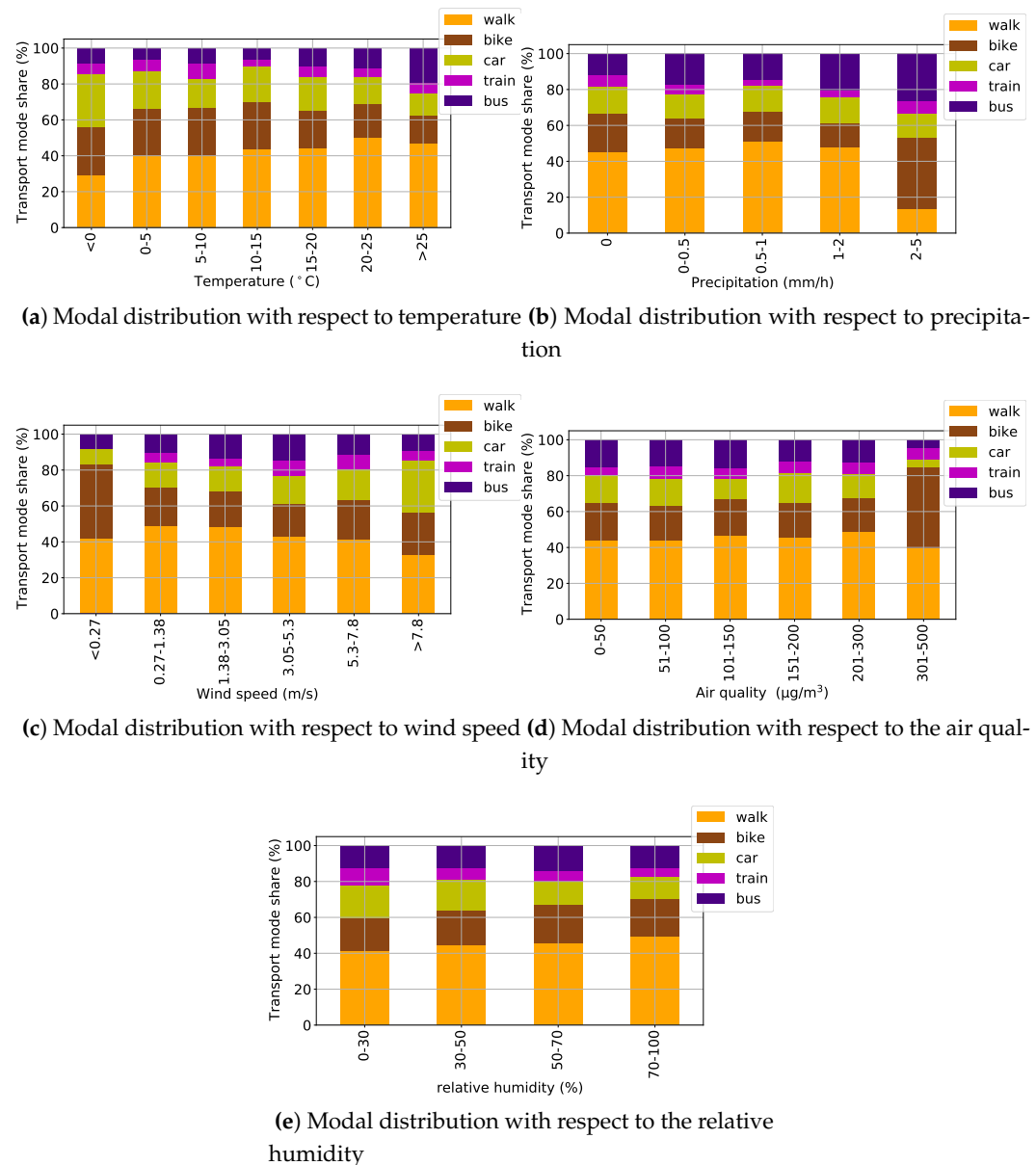


Figure 3. Impact of weather conditions on transport mode choices.

(0-50) $\mu\text{g}/\text{m}^3$ to unhealthy (201-300) $\mu\text{g}/\text{m}^3$, the transport mode share remains constant. The potential explanation is that Beijing averages an air quality of 150 $\mu\text{g}/\text{m}^3$ [see – Table 1], so it is highly likely that most of the time the air quality is unhealthy as people carry on with their daily lives. However, when the air quality becomes hazardous (300+) $\mu\text{g}/\text{m}^3$, bike and walk have a larger share and there is a reduction in the car and bus shares in order to reduce the (CO_2) emissions. In Figure 3e, the share of transport mode is the same regardless of the relative humidity, which is in line with the findings in Aultman-Hall et al. [36], who found that humidity has a limited effect on transport choices.

3.2. Influence of context information on the transport mode choices

During the Olympics, an increase in the bus share and a reduction in the car share is shown in Figure 4a. Its intuitive that during international events of this magnitude some roads are closed off or restricted to the public. Our findings show that the alternative travel means is by public transport, the bus in particular. In Figure 4b, an increase in the bus share and a reduction in the train and car shares is noticeable at rush hours in comparison

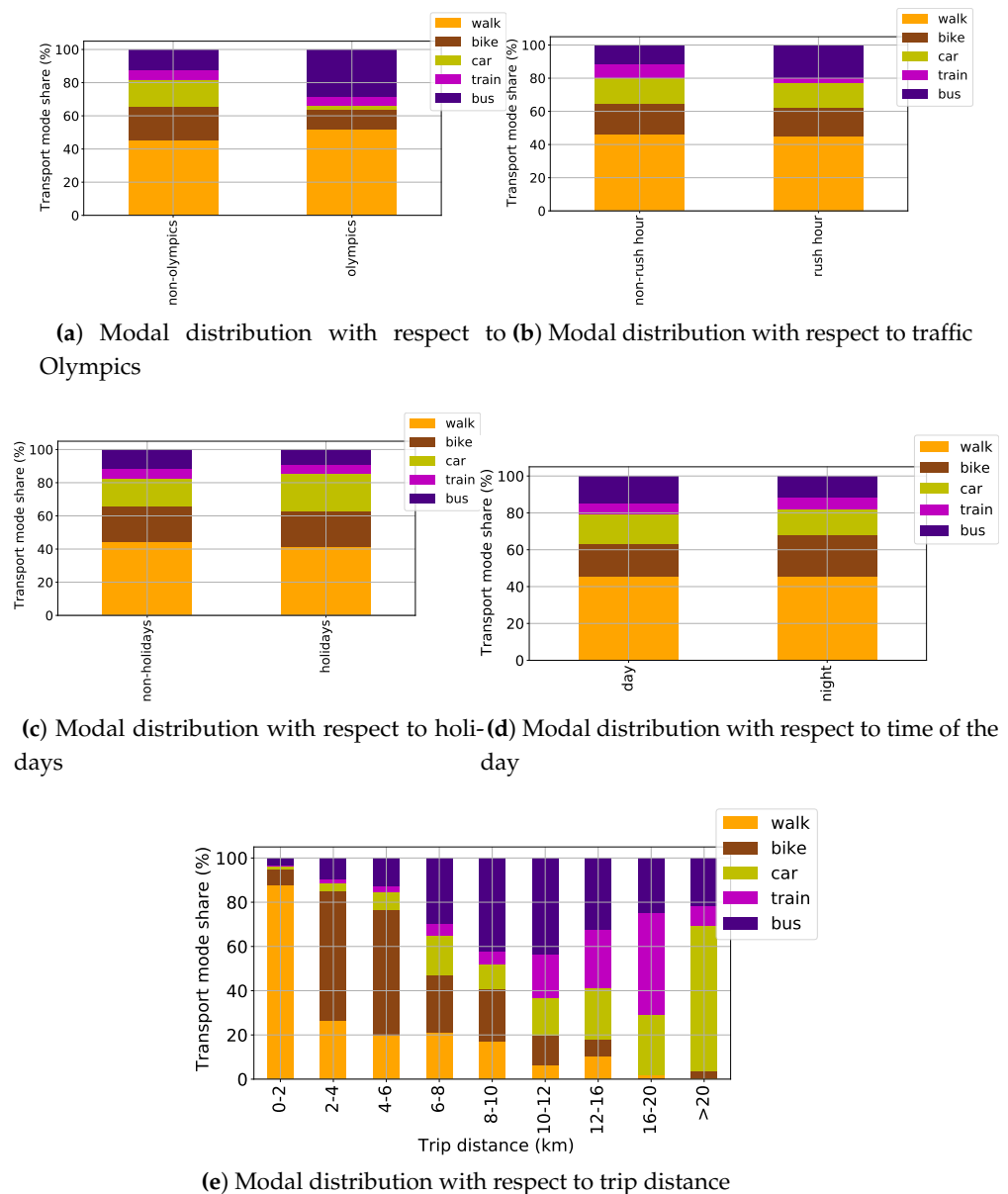


Figure 4. Impact of context information on transport mode choices.

with non-rush hours. This finding is in line with the study made by the Asian development bank in [33] in which it is shown that during rush hours passengers prefer taking the bus than the train.

In Figure 4c, we see that there is an increase in the car share and a reduction in the bus share during holidays perhaps because people are choosing to travel by car on holidays. In Figure 4e, it can be observed that trip distance has a tremendous influence on the transport mode choice. For instance for shorter distances (0-6 km), walk and bike have a larger share but between (6-12 km) the bus share is larger. However between 12-20 km, the bus share reduces as the car and train shares increase. Beyond 20 km, train share reduces and car share increases.

4. Statistical Modelling

Given that the transport mode choices are a typical example of discrete outcomes, in order to link the probabilities of choosing a given mode of transport to the weather and context information, in this study, we adopt MNL and MNP models. These are among the

most popular models in analysing and predicting travel decisions. These models analyse simultaneous effects of meteorological and context information on transport decisions in one integrated model, which is important in understanding mobility decisions.

4.1. Methods

According to Ben-Akiva et al. [42], the framework of these models is based on four general assumptions: i) decision maker who is the individual or group of people making the choice, ii) alternatives, which are the available choices to choose from, iii) attributes, which are parameters that characterise each alternative, iv) decision rule, which is the process used by the decision maker evaluates the alternatives.

The decision rule in models used in travel behavior analysis is based on the Utility model, U_{in} , and takes the form

$$U_{in} = V_{in} + \epsilon_{in} \quad (1)$$

The alternative with the highest utility is chosen, therefore the probability that alternative i is chosen by an individual n from a choice set $C_n = \{1, 2, \dots, i, \dots, j_n\}$ takes on the general form

$$\begin{aligned} P_{in} &= P[V_{in} + \epsilon_{in} \geq V_{jn} + \epsilon_{jn}, \forall j \in C_n] \\ &= P[(V_{in} + \epsilon_{in}) = \max_{j \in C_n} (V_{jn} + \epsilon_{jn})] \\ &= P[\epsilon_{jn} - \epsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n] \end{aligned} \quad (2)$$

V_{in} is the deterministic or systematic part of the utility function. It is defined in (3) by a vector of observable variables z_n and their corresponding coefficients γ_i :

$$V_{in} = \gamma_i^T z_n \quad (3)$$

To identify the model, one set of coefficients of the systematic term needs to be normalized to zero e.g. ($\gamma_1 = 0$), which makes the corresponding transport mode choice ($i = 1$) to become the base mode. The coefficients of other alternatives are interpreted in reference to the base outcome.

ϵ_{in} is the random term which expresses the errors of the utility function. Its distribution is often known and makes the problem more reasonable to empirically characterise. If the error terms assumes an independent and identically distributed "Extreme Value", i.e. Gumbel, the model is MNL, hence P_{in} takes on the form

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \quad (4)$$

However, when the error terms assume a multivariate normal distribution, the model is MNP and P_{in} takes on the form

$$\int_{\xi_{1n}}^{\infty} \dots \int_{\xi_{(i-1)n}}^{\infty} \int_{\xi_{(i+1)n}}^{\infty} \dots \int_{\xi_{jn}}^{\infty} \phi(\tilde{\epsilon}_{1n}, \dots, \tilde{\epsilon}_{(i-1)n}, \tilde{\epsilon}_{(i+1)n}, \dots, \tilde{\epsilon}_{jn}) d\tilde{\epsilon}_{1n} \dots d\tilde{\epsilon}_{(i-1)n} d\tilde{\epsilon}_{(i+1)n} \dots d\tilde{\epsilon}_{jn} \quad (5)$$

where $\phi()$ is the density function which follows a multivariate normal distribution with means 0 and a covariance matrix with a size of $J_n \times J_n$, $\xi_{1n} = V_{in} - V_{jn}$ and $\tilde{\epsilon}_{in} = \epsilon_{jn} - \epsilon_{in}$ as also seen in (2) are the difference between the systematic and error terms, respectively.

The difficulty of estimation of these models grows with the number of discrete choices, so dedicated commercial software packages are recommended for their estimation. In our

multivariate analysis, MNL and MPL models have been implemented via the software package Stata. We refer the reader to Stata base manual [43] for a detailed discussion of estimation procedure.

Note that during the analysis, five modes of transports are used as the dependent variable [see Figure 2], and two groups of individual level variables: weather and context information attributes are used as the independent (explanatory) variables [see Table 1]. However, in the model, we excluded the use of relative humidity, precipitation, and holidays information because as seen previously have a limited impact on the transport mode decisions. In fact, the categories of the remaining weather conditions such as temperature, wind speed, and air quality were adjusted according to relationships and patterns observed during the statistical analysis.

Table 2. MNL and MNP model results of the impact of weather and context data on transport mode choices

Explanatory variables	Multinomial logit (MNL)				Multinomial probit (MNP)			
	Bike	Car	Bus	Train	Bike	Car	Bus	Train
Intercept	0.4154	-2.560**	-1.265**	-2.928**	0.320	-1.607**	-0.788*	-1.701**
Temperature (C)								
< 0	ref	ref	ref	ref	ref	ref	ref	ref
0-15	-1.122**	-1.068	-2.292**	-1.349*	-0.974**	-0.823*	-1.685**	-0.981*
15-25	-1.628**	-1.371*	-2.152**	-2.068**	-1.335**	-0.998*	-1.642**	-1.464**
> 25	-1.969**	-1.558**	-1.331*	-1.338*	-1.585**	-1.144*	-1.103**	-1.013*
Wind speed (m/s)								
0-3	ref	ref	ref	ref	ref	ref	ref	ref
3-7.8	0.013	0.311	0.256	0.554*	0.020	0.177	0.160	0.320**
> 7.8	0.280	0.621	-0.018	0.228	0.200	0.400	-0.004	0.069
Air quality ($\mu\text{g}/\text{m}^3$)								
< 300	ref	ref	ref	ref	ref	ref	ref	ref
> 300	0.921**	-1.333	-0.564	0.405	0.735*	-0.524	-0.270	0.315
Trip distance (km)								
0-4	ref	ref	ref	ref	ref	ref	ref	ref
4-8	2.143**	2.875**	2.588**	2.829**	1.741**	1.964**	1.890**	1.890**
8-12	1.912**	4.222**	4.305**	4.926**	1.480**	2.829**	3.082**	3.150**
12-20	0.872*	5.211**	4.319**	6.381**	0.838**	3.444**	2.992**	4.091**
> 20	3.609**	9.380**	7.094**	7.912**	2.150**	5.862**	4.293**	4.553**
Olympics (<i>ref. non Olympics</i>)	-0.471	-2.347**	0.2617	-0.432	-0.380	-1.635**	0.210	-0.353
Rush hour (<i>ref. non rush hour</i>)	0.022	-0.999**	-0.259	-1.351**	-0.005	-0.624**	-0.146	-0.827**
Day (<i>ref. night</i>)	-0.140	0.355	0.515**	0.059	0.102	0.211	0.330*	0.0382
Modelling performance								
Log likelihood	-2498.486				-2505.669			
AIC	5108.971				5123.339			
BIC	5438.823				5453.19			
McFadden R-squared	0.3289				0.012			

** Significant at $\alpha < 0.01$

* Significant at $\alpha < 0.05$

4.2. Results

Table 2 shows the results of the both models with a summary of the relationships between the transport mode choices, weather, and the context information. The modelling performance shows that MNL model performs better than MNP because it has a larger McFadden R-squared and smaller AIC and BIC values, which shows that results will be in favor of the MNL model. The values in the tables represent the standardised regression coefficients, with the star indicating their respective statistical significance. Note that the coefficients are a relative measure of transport choice compared to walking, which was reference mode. For instance our findings indicate that a negative temperature effect on the bike, car, bus, and train make them less likely to be used as transport modes than walking.

However, this means that other than the sign, the coefficients don't have a lot of useful interpretation since the magnitude of the coefficients cannot be interpreted or quantified. Thus the marginal effects of the observatory variables z_n on the dependent variable i is required. Marginal effects are defined in (6) as the amount of change in dependent variable due to one unit change of the observatory variable in the model system.

$$\frac{\partial P_{in}}{\partial z_n} = P_{in}(\gamma_i - \bar{\gamma}_i) \quad (6)$$

Table 3 reports the marginal effects for each variable according to the MNP and MNL models. The positive values represent an increase in the probability of selecting an alternative by the marginal effect expressed as a percentage while the negatives indicate the opposite. It should be noted that qualitative results are consistent between the MNP and MNL.

Table 3. Marginal Effects of MNL and MNP models of the impact of weather and context data on transport mode choices

Explanatory variables	Multinomial logit (MNL)					Multinomial probit (MNP)				
	Walk	Bike	Car	Bus	Train	Walk	Bike	Car	Bus	Train
Temperature (C)										
< 0	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
0-15	0.234**	-0.006	0.003	-0.219**	-0.013	0.273**	-0.050	0.077	-0.218**	-0.011
15-25	0.320**	-0.107	-0.004	-0.177**	-0.031	0.351**	-0.135	0.004	-0.183**	-0.037
> 25	0.279**	-0.216**	-0.026	-0.030	-0.006	0.319**	-0.234**	-0.025	-0.053	-0.005
Wind speed (m/s)										
0-3	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
3-7.8	-0.039	-0.024	0.018	0.027	0.018*	-0.037	-0.018	0.0158	0.020	0.020*
> 7.8	-0.062	0.039	0.046	-0.025	0.002	-0.057	0.0342	0.05	-0.024	-0.002
Air quality ($\mu\text{g}/\text{m}^3$)										
< 300	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
> 300	-0.099	0.312**	-0.125**	-0.096*	0.005	-0.100	0.295**	-0.121**	-0.084*	0.010
Trip distance (km)										
0-4	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
4-8	-0.517**	0.303**	0.061**	0.125**	0.027**	-0.518**	0.298**	0.062**	0.128**	0.028**
8-12	-0.629**	-0.015	0.123**	0.398**	0.123**	-0.629**	-0.017	0.125**	0.394**	0.126**
12-20	-0.664**	-0.162**	0.225**	0.246**	0.353**	-0.662**	-0.163**	0.229**	0.245**	0.350**
> 20	-0.725**	-0.176**	0.672**	0.162**	0.066**	-0.724**	-0.177**	0.671**	0.163**	0.067**
Olympics (<i>ref. non Olympics</i>)	0.130*	-0.046	-0.175**	0.100**	-0.004	0.140*	-0.040	-0.21**	0.115**	-0.005
Rush hour (<i>ref. non rush hour</i>)	0.072	0.058**	-0.071*	-0.014	-0.044**	0.063	0.0503**	-0.073**	0.004	-0.052*
Day (<i>ref. night</i>)	-0.031	-0.064**	0.024	0.073*	-0.006	-0.029	-0.057*	0.023	0.064*	0.010

** Significant at $\alpha < 0.01$

* Significant at $\alpha < 0.05$

5. Discussion

In the previous section, we have presented statistical models that help in understanding the relationship between the weather and context information with transport mode choices. In this section, the results in Section 4.2 are discussed.

5.1. Air Quality

When the air quality changes from the very unhealthy category to hazardous, Table 3 shows that this change would increase the bike share by 31% and reduce the car and bus shares by 12% and 10%, respectively. The likely explanation is that hazardous air quality causes serious health concerns such as loss of lung capacity and decreased lung function according to epidemiological studies. Therefore, the government in Beijing introduces vehicle controls to reduce CO₂ emissions to prevent the population from long-term exposure to polluted air. Consequently, the population is encouraged to use environmentally friendly transport mode choices, such as the bike, to get to their preferred destinations.

5.2. Temperature

Among all weather conditions, temperature has the most prominent effect on mobility choices with strong significant effects on almost all travel behaviour modes. According to Table 3, when the temperature increases, people walk more, which decreases the relative likelihood of choosing other transport modes. For instance, we see that with a unit increase in temperature for 0-15 C and 15-25 C, the probabilities of walking is expected to increase by 23% and 32%, respectively while the probabilities of choosing the bus is expected to decrease by 21% and 17%, respectively. However, a unit increase in temperature greater than 25 C, would reduce the bike share by 21% while the walking share would increase by 27% according to MNL model, which is line with the survey results in [44].

5.3. Wind speed

Table 3 shows that a unit increase in the wind speed between 3-7.8 m/s would result into an increase of 1.8% in the train modal share according to the MNL model. This result shows that the wind speed just like relative humidity and precipitation has a very limited effect on transport mode choices in Beijing. One possible reason is that during the period in which this study was conducted, the average wind speed was 3 m/s, which is classified as very light according to the Beaufort scale.

5.4. Trip distance

Trip distances significantly influences all transport choices as shown in Table 3. For instance, a unit increase in distances of 4-8 km is expected to decrease walking and biking by 51% and 30%, respectively, while increase the bus, car, and train shares by 6%, 12.5%, and 2.7%, respectively. A unit increase in distances of 8-12 km is expected to lead to a further decrease in walking by 62%, while increasing, car, bus, and train shares by 12.3%, 39.8%, and 12.3%, respectively. For a unit increase in distances of 12-20 km, walking and biking are expected to decrease by 66% and 16.2%, respectively, while car, bus, and train shares are expected to increase by 22.5%, 24.6%, 35.3%, respectively.

Beyond 20 km, this trend is similar except that there is a further increase in the car, bus, and train shares by 67.2%, 16.2%, 6.6%, respectively while a further reduction in walking and bike shares by 72.5% and 17.6%, respectively is expected. These results are in line with the findings in survey studies by [16] and [45], whose authors find that shorter distances are likely to be performed by walking and bike, while the longer distances by car, bus, and trains.

5.5. Olympics

Generally, it is highly likely that different transport modes will be disrupted due to the magnitude of people that an event like the summer Olympics attracts. This study helps to analyse how the Olympics influences the transport mode share. During the Olympics, we see that there is a decrease of 17.5% in the car share and an increase of 13% and 10% in the walk and bus shares, respectively.

The decrease in the car share can be attributed to the private car restrictions imposed by the government, and the increase in the bus share is explained by the additional bus lines put in place as an alternative to accommodate the number of passengers. Our findings are consistent with the findings of the authors in [33] who mention that during the preparation for the 2008 summer Olympics, a series of traffic restrictions on private vehicles were introduced by the Beijing government.

5.6. Rush hour

During rush hours, we see a reduction in the likelihood of traveling by car and train by 7.1% and 4.4%, respectively and an increase in biking by 5.8%. A more general explanation for this travel behaviour may have to do with the need to reduce or avoid congestion i.e. traffic jams and congestion of people in trains.

5.7. Day/Night

During the day, Table 3 shows that the probability of biking would decrease by 6.4% and the use of the bus would increase by 7.3%. A possible reason is that during the day there are more buses operating than in the night. During the night there is an increase in the people biking because transport operators reduce the number of buses.

However, our result is contrary to the findings obtained using survey data in Scandinavian cities of Stockholm and Oslo [45], in which during the night there are fewer trips by active transport modes such as biking than during the day.

6. Conclusions

Our main objective is to provide a better understanding on the possibility of using GPS data for studying the impact of weather and context data on transport mode choices. In the methodology, we linked GPS trajectories with weather and context information and then analysed their relationship by descriptive means and statistical models.

The first conclusion is that trip distances and the transport mode choice were very much interrelated. In fact, this study highlights that the trip distance is the most significant factor in choosing a transport mode. Of the five weather conditions, temperature had the most prominent effect on the transport mode choice.

The second conclusion is that we can observe the effect of governmental regulations on the choice of transport mode once the air quality becomes hazardous. We observe that traffic restrictions imposed by the government encourage the population to choose environmentally friendly transport modes such as the bike instead of the car and bus, which are greenhouse gas emitters. Moreover, our study concludes that hosting an event such as the summer Olympics would require transport regulators and operators to provide additional bus lines to accommodate the number of passengers.

Future studies could include adding more explanatory variables such as income, gender, occupation, age, trip purpose, to improve the performance of the models. In addition, it would be interesting to use advanced models such as Structural equation modelling to reveal potential unobservable heterogeneity of the effects of weather and context data on transport mode choices.

Acknowledgments: The authors wish to thank Dr. Anton Galich at DLR German Aerospace Center, Institute of Transport Research for his helpful comments on the research presented in this paper.

References

1. ADB Sustainable Development Working Paper Series: Rethinking Transport and Climate Change. <https://www.adb.org/sites/default/files/publication/28489/adb-wp10-rethinking-transport-climate-change.pdf>. Accessed: 2022-01-12.
2. Dixit, M.; Sivakumar, A. Capturing the impact of individual characteristics on transport accessibility and equity analysis. *Transportation Research. Part D, Transport and Environment* **2020**, *87*.
3. Fatima, K.; Moridpour, S.; Saghapour, T. Spatial and Temporal Distribution of Elderly Public Transport Mode Preference. *Sustainability* **2021**, *13*, 4752.
4. Şimşekoğlu, Ö.; Nordfjærn, T.; Rundmo, T. The role of attitudes, transport priorities, and car use habit for travel mode use and intentions to use public transportation in an urban Norwegian public. *Transport Policy* **2015**, *42*, 113–120.
5. Guzman, L.A.; Oviedo, D.; Rivera, C.M.R. Assessing equity in transport accessibility to work and study: The Bogotá region. *Journal of Transport Geography* **2017**, *58*, 236–246.
6. Koetse, M.J.; Rietveld, P. The impact of climate change and weather on transport: An overview of empirical findings. *Transportation Research Part D-transport and Environment* **2009**, *14*, 205–221.
7. Liu, C.; Susilo, Y.O.; Karlström, A. Weather variability and travel behaviour – what we know and what we do not know. *Transport Reviews* **2017**, *37*, 715 – 741.
8. Böcker, L.; Dijst, M.; Prillwitz, J. Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A Literature Review. *Transport Reviews* **2013**, *33*, 71 – 91.
9. Li, J.; Li, X.; wei Chen, D.; Godding, L. Assessment of metro ridership fluctuation caused by weather conditions in Asian context: Using archived weather and ridership data in Nanjing. *Journal of Transport Geography* **2017**, *66*, 356–368.
10. Cools, M.; Moons, E.; Creemers, L.; Wets, G. Changes in Travel Behavior in Response to Weather Conditions: Do Type of Weather and Trip Purpose Matter? *Transportation Research Record* **2010**, *2157*, 22–28, [<https://doi.org/10.3141/2157-03>]. doi:10.3141/2157-03.

11. Arana, P.; Cabezudo, S.; Peñalba, M. Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transportation Research Part A-policy and Practice* **2014**, *59*, 1–12.
12. Böcker, L.; Uteng, T.P.; Liu, C.; Dijst, M. Weather and daily mobility in international perspective: A cross-comparison of Dutch, Norwegian and Swedish city regions. *Transportation Research Part D-transport and Environment* **2019**, *77*, 491–505.
13. Singhal, A.; Kamga, C.; Yazici, M.A. Impact of weather on urban transit ridership. *Transportation Research Part A-policy and Practice* **2014**, *69*, 379–391.
14. Böcker, L.; Prillwitz, J.; Dijst, M. Climate change impacts on mode choices and travelled distances: a comparison of present with 2050 weather conditions for the Randstad Holland. *Journal of Transport Geography* **2013**, *28*, 176–185. doi:10.1016/j.jtrangeo.2012.1.
15. Hagenauer, J.; Helbich, M. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Syst. Appl.* **2017**, *78*, 273–282.
16. Ma, L.; Xiong, H.; Wang, Z.; Xie, K. Impact of weather conditions on middle school students' commute mode choices: Empirical findings from Beijing, China. *Transportation Research Part D: Transport and Environment* **2019**, *68*, 39–51. doi:10.1016/j.trd.2018.05.008.
17. Li, L.; Wang, J.; Song, Z.; Dong, Z.; Wu, B.J. Analysing the impact of weather on bus ridership using smart card data. *Iet Intelligent Transport Systems* **2015**, *9*, 221–229.
18. Li, X.; Ma, Q.; Wang, W.; Wang, B. Influence of Weather Conditions on the Intercity Travel Mode Choice: A Case of Xi'an. *Computational Intelligence and Neuroscience* **2021**, 2021.
19. Hyland, M.F.; Frei, C.; Frei, A.; Mahmassani, H.S. Riders on the storm: Exploring weather and seasonality effects on commute mode choice in Chicago. *Travel Behaviour and Society* **2018**.
20. Creemers, L.; Wets, G.; Cools, M. Meteorological variation in daily travel behaviour: evidence from revealed preference data from the Netherlands. *Theoretical and Applied Climatology* **2014**, *120*, 183–194.
21. Ton, D.; Duives, D.C.; Cats, O.; Hoogendoorn-Lanser, S.; Hoogendoorn, S.P. Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A: Policy and Practice* **2019**.
22. Böcker, L.; Dijst, M.; Faber, J. Weather, transport mode choices and emotional travel experiences. *Transportation Research Part A: Policy and Practice* **2016**, *94*, 360–373. doi:https://doi.org/10.1016/j.tra.2016.09.021.
23. GeoLife: Building Social Networks Using Human Location History. <https://www.microsoft.com/en-us/research/project/geolife>. Accessed: 2021-09-10.
24. Beijing Travel Guide. https://travel.usnews.com/Beijing_China/Getting_Around/. Accessed: 2022-01-12.
25. Zheng, Y.; Zhang, L.; Xie, X.; Ma, W.Y. Mining interesting locations and travel sequences from GPS trajectories. WWW '09, 2009.
26. Shaham, S.; Ding, M.; Liu, B.; Dang, S.; Lin, Z.; Li, J. Privacy Preserving Location Data Publishing: A Machine Learning Approach. *IEEE Transactions on Knowledge and Data Engineering* **2021**, *33*, 3270–3283. doi:10.1109/TKDE.2020.2964658.
27. Lehmann, A.L.; Alvares, L.O.; Bogorny, V. SSM: a similarity measure for trajectory stops and moves. *International Journal of Geographical Information Science* **2019**, *33*, 1847–1872, [<https://doi.org/10.1080/13658816.2019.1605074>]. doi:10.1080/13658816.2019.1605074.
28. Qi, M.; Wang, Z.; He, Z.; Shao, Z. User Identification across Asynchronous Mobility Trajectories. *Sensors* **2019**, *19*. doi:10.3390/s19092102.
29. Nawaz, A.; Huang, Z.; Wang, S.; Akbar, A.; AlSalman, H.; Gumaei, A. GPS Trajectory Completion Using End-to-End Bidirectional Convolutional Recurrent Encoder-Decoder Architecture with Attention Mechanism. *Sensors* **2020**, *20*. doi:10.3390/s20185143.
30. Namdarpour, F.; Mesbah, M.; Gandomi, A.H.; Assemi, B. Using genetic programming on GPS trajectories for travel mode detection. *IET Intelligent Transport Systems* **2022**, *16*, 99–113, [<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/itr2.12132>]. doi:https://doi.org/10.1049/itr2.12132.
31. POWER Data Access View. <https://power.larc.nasa.gov/data-access-viewer/>. Accessed: 2021-11-12.
32. Climate - Beijing (China). <http://www.climatestotravel.com/climate/china/beijing>. Accessed: 2022-01-12.
33. Travel Demand Management Options in Beijing. <https://www.adb.org/sites/default/files/publication/325266/travel-demand-management-beijing.pdf>. Accessed: 2022-01-12.
34. China Public Holidays Calendar. <https://www.beijingesc.com/china-holiday-calendar.html>. Accessed: 2022-01-12.
35. Cavallaro, C.; Vitrià, J. Corridor Detection from Large GPS Trajectories Datasets. *Applied Sciences* **2020**, *10*. doi:10.3390/app10145003.
36. Aultman-Hall, L.; Lane, D.; Lambert, R.R. Assessing Impact of Weather and Season on Pedestrian Traffic Volumes. *Transportation Research Record* **2009**, *2140*, 35–43, [<https://doi.org/10.3141/2140-04>]. doi:10.3141/2140-04.
37. Ma, Y.; Ni, G.; Chandra, C.V.; Tian, F.; Chen, H. Statistical characteristics of raindrop size distribution during rainy seasons in the Beijing urban area and implications for radar rainfall estimation. *Hydrology and Earth System Sciences* **2019**, *23*, 4153–4170. doi:10.5194/hess-23-4153-2019.
38. Harikumar, R.; Sampath, S.; Sasi Kumar, V. Variation of rain drop size distribution with rain rate at a few coastal and high altitude stations in southern peninsular India. *Advances in Space Research* **2010**, *45*, 576–586. doi:https://doi.org/10.1016/j.asr.2009.09.018.
39. Royal Meteorological Society. <https://www.rmets.org/resource/beaufort-scale/>. Accessed: 2022-01-12.
40. Air Quality Index (AQI). <https://www.epa.gov/sites/default/files/2014-05/documents/zell-aqi.pdf>. Accessed: 2022-01-12.
41. Lu, J.; Zhang, X.; Li, P.; Zhu, Y. Case-Based FCTF Reasoning System. *Applied Sciences* **2015**, *5*, 825–839. doi:10.3390/app5040825.
42. Hall, R.W. Handbook of transportation science. 1999.
43. StataCorp, Stata base reference manual. <https://www.stata.com/manuals/r.pdf>. Accessed: 2021-12-12.

-
44. Helbich, M.; Böcker, L.; Dijst, M. Geographic heterogeneity in cycling under various weather conditions: evidence from Greater Rotterdam. *Journal of Transport Geography* **2014**, *38*, 38–47.
 45. Böcker, L.; Uteng, T.P.; Liu, C.; Dijst, M. Weather and daily mobility in international perspective: A cross-comparison of Dutch, Norwegian and Swedish city regions. *Transportation Research Part D: Transport and Environment* **2019**, *77*, 491–505. doi:10.1016/j.trd.2019.07.012.