Project report:

The influence of weather conditions to trip mode choices

Group 3
Georgios Kritikos (s222562)
Elmo Lampén (s226404)
Lucian Vrbanc (s222660)
Kunimi Wataru (s221747)

42180 Quantitative Modelling of Behaviour
DTU Management
Technical University of Denmark (DTU)
12th May 2023

Table of Contents

1 Introduction 2
2 Data 2
3 Method 3
3.1 Scenario 1: Weather score 3
3.1.1 Base model 4
3.1.2 Common Betas 4
3.1.3 Alternative Specific Betas 5
3.2 Scenario 2: Continuous Weather Variables 6
3.2.1 Common Betas 6
3.2.2 Alternative Specific Betas 6
3.3 Scenario 3: Weather Condition Groups + Combination comparison 7
3.3.1 Mean Temperature Groups 7
3.3.2 Adding the rest of the weather conditions 9
3.3.3 Combining groups and continuous variables 11
4 Results 13
4.1 The weather score scenario 13
4.2 The continuous variables scenario 14
4.3 The weather combination groups with comparisons scenario 15
5 Discussion and conclusions 15
Appendix 1. The results of our scenarios 17
Appendix 1A. The base model 17
Appendix 1B. The individual weather scoring 18
Appendix 1C. The paper scoring 20
Appendix 1D. The parameter groups 21
Appendix 2. Descriptive statistics of the weather parameters 22

1 Introduction

This report presents a comprehensive study that investigates the impact of weather conditions on trip mode choices. The study uses two main datasets: the London Trip Mode Choice Data and Daily Weather Data. The former dataset is provided by the course, while the latter is obtained from the European Climate Assessment Dataset. By analysing these datasets and drawing on the reference paper by Thomas, Jaarsma, and Tutert (2013) titled "Exploring Temporal Fluctuations of Daily Cycling Demand on Dutch Cycle Paths: The Influence of Weather on Cycling," the study seeks to answer two primary research questions.

The first research question focuses on determining <u>which model best explains the relationship between weather conditions and mode choice</u>. To address this question, the study employs the multinominal logit model or nested logit model, both of which are statistical methods used to analyse the choice behaviour of individuals, particularly in the context of transportation.

The second research question explores <u>which weather component has the greatest influence on mode choices</u>. Through the application of various models and analyses of the datasets, the study aims to shed light on the complex relationship between weather conditions and mode choices.

The findings of this research are expected to be valuable for policymakers and transportation planners in designing sustainable transportation systems that can withstand changing weather patterns. By gaining a better understanding of how weather conditions affect mode choices, policymakers and planners can develop transportation systems that are resilient and responsive to changing weather patterns, ultimately leading to a more sustainable and efficient transportation network.

2 Data

Our research study incorporates data from the London Passenger Mode Choice (2012-2015) dataset, as well as weather conditions recorded at London St. James Park and Heathrow Airport, sourced from the European Climate Assessment and Dataset.

In the London Passenger Mode Choice dataset, we specifically utilized information such as the date, cost, and time associated with each subject's mode of transportation. By analysing this data, we were able to gain insights into individuals' mode choice patterns.

Regarding the weather data, we focused on key variables including daily wind mean speed (FG), daily mean temperature (TG), daily precipitation (RR), daily sunshine duration (SS), and daily cloud cover (CC). We employed a blending technique to address missing values by leveraging mathematical modelling, thereby ensuring a comprehensive and robust dataset.

To conduct our analysis, we combined the London Passenger Mode Choice and weather data sets based on the corresponding dates. This integration enabled us to explore the relationship between weather conditions and passenger mode choice in London more effectively.

By employing these combined datasets, we aim to shed light on the influence of weather conditions on transportation mode selection, providing valuable insights for urban planning, transportation management, and decision-making processes.

3 Method

To assess the effect of the different weather conditions on mode choices of the population the following three separate scenarios were examined:

1. Scenario 1, a scoring method was used for combining all the individual weather conditions into a single score by using a linear function. This method is used by Thomas et. al. (2013) in their paper "Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling".
2. In scenario 2, individual weather parameters (cloud cover, wind speed etc.) were treated separately as continuous variables and added one by one to the specific utility functions.
3. Finally in scenario 3, groups for each weather parameters were created, and dummy variables were used for each group in the utility functions. At this point, the use of continuous variables and groups was compared between each other and to their consequences.

For all of the above scenarios 2 models were tested. One was a simple MNL model and the second one was a Nested Logit model with 2 classes. The first class consisted of bicycle and walking as they are both affected by the weather conditions more and the second one was public transport and car as they were considered the ones that are not so much affected by the weather.

The test was carried out twice for each scenario. In the first run the whole data set was used to study the effect of the weather. In the second run the dataset was restricted only to the trips that are below 5 kilometres that were considered short distances and could be achieved by all available transport alternatives. The reason for choosing 5 km as upper limit is that it is an approximate equivalent of one hour walking time and was therefore considered to be a good restriction.

3.1 Scenario 1: Weather score

The weather conditions that are included in the dataset are TG (mean temperature), CC (cloud cover), SS (sunshine duration), RR (precipitation) and FG (wind speed). For the purposes of the paper of Thomas et. al. (2013), the cloud cover (CC) is not considered so it will be omitted from the calculations. As already mentioned, each one of the above parameters is converted into a score by using the following functions:

$$W_{TG} = \begin{cases} TG, & 3 \leq TG \leq 18 \\ TG - 0.2 * (TG - 3), & 3 > TG \\ 18, & T > 18 \end{cases}$$

$$W_{SS} = SS^{0.7}$$

$$W_{RR} = RR^{0.5}$$

$$W_{FG} = FG^{1.5}$$

After applying the above functions for all parameters, the scores are combined into a single weather score in a linear combination by applying:

$$Weather\ Score\ = a_{TG} * W_{TG} + \ a_{SS} * W_{SS} + a_{RR} * W_{RR} + \ a_{FG} * W_{FG}$$

The α values are the weather coefficients calculated in the paper, and the same values are used in this study:

$$a_{TG} = 0.78, \qquad a_{ss} = 0.39, \qquad a_{RR} = -0.39, \qquad a_{FG} = -0.38$$

Now that the weather score value has been calculated for each day and so for each trip, this score can be used to find if it affects the choices of the passengers.

3.1.1 Base model

The base model of weather score with walking as a base for testing the choice behaviour:

$$V_{cycle} = ASC_{cycle} + b_{time} * Time$$

$$V_{drive} = ASC_{drive} + b_{time} * Time + b_{cost} * Cost$$

$$V_{pt} = ASC_{pt} + b_{time} * Time + b_{cost} * Cost$$

$$V_{walk} = b_{time} * Time$$

In this base model, common betas across alternatives for time and cost are used to emphasise the effect of the remaining variables as much as possible to weather conditions. This is beneficial especially for later modifications of this model when more parameters will be added.

3.1.2 Common Betas

In the common betas model, weather conditions were studied as they would be constant across all the alternatives. The influence of each weather condition was assumed to be the same in each alternative in this model. One of the alternatives, walking this time, was chosen the base to which other alternatives were compared to. Therefore, the total number of betas was N-1.

The model of weather score with common betas and walking as a base:

$$V_{cycle} = ASC_{cycle} + b_{time} * Time + b_{weather} * weather\_score$$

$$V_{drive} = ASC_{drive} + b_{time} * Time + b_{cost} * Cost + b_{weather} * weather\_score$$

$$V_{pt} = ASC_{pt} + b_{time} * Time + b_{cost} * Cost + b_{weather} * weather\_score$$

$$V_{walk} = b_{time} * Time$$

The above model is applied to both the MNL and the Nested Logit model for the whole dataset and for the short distance dataset (<5 km trips). The main criterion for assessing the efficiency of each model was the Log Likelihood. The summarized results are presented in the following tables 1 and 2. A more detailed analysis of the results will be done in chapter 4.

Table 1. Log-likelihoods of the model for the whole dataset (weather score, common betas)

Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x^2 (0.05)
Base MNL	-74975.46	5	0	0	0	-
MNL	-74975.46	6	0	4	0	3.841
Base NL	-73082.67	6	0	0	0	-
NL	-73082.67	7	0	4	0	3.841

Table 2. Log-likelihoods of the model for short distances (weather score, common betas)						
Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x2 (0.05)
Base MNL	-55243.27	5	0	0	0	-
MNL	-55243.27	6	0	1	0	3.841
Base NL	-55229.15	6	0	0	0	-
NL	-54236.35	7	-992.8	1	1985.6	3.841

Tables 1 and 2 shows that the NL model gives better results than the MNL model and that the addition of the weather variable in most cases doesn't affect the Log Likelihood except in the NL model of short distances.

3.1.3 Alternative Specific Betas

The model of weather score with alternative specific betas and walking as a base:

$$V_{cycle} = ASC_{cycle} + b_{time} * Time + b_{weather,cycle} * weather\_score$$

$$V_{drive} = ASC_{drive} + b_{time} * Time + b_{cost} * Cost + b_{weather,drive} * weather\_score$$

$$V_{pt} = ASC_{pt} + b_{time} * Time + b_{cost} * Cost + b_{weather,pt} * weather\_score$$

$$V_{walk} = b_{time} * Time$$

The respective results are presented in tables 3 and 4 below.

Table 3. Log-likelihoods of the model for the whole dataset (weather score, alternative specific betas)						
Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x2 (0.05)
Base MNL	-74975.46	5	0	0	0	-
MNL	-74934.36	9	-41.1	4	82.2	9.488
Base NL	-73082.67	6	0	0	0	-
NL	-73031.64	10	-51.03	4	102.06	9.488

Table 4. Log-likelihoods of the model for short distances (weather score, alternative specific betas)						
Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x2 (0.05)
Base MNL	-55243.27	5	0	0	0	-
MNL	-55209.52	9	-33.75	4	67.5	9.488
Base NL	-55229.15	6	0	0	0	-
NL	-54197.1	10	-1032.05	4	2064.1	9.488

In this case tables 3 and 4 shows that all the Likelihood Ratio Tests are accepted, and the new models do in fact improve the first model based on the improvement in log-likelihoods when compared to the respective base model results.

3.2 Scenario 2: Continuous Weather Variables

In Scenario 2 the individual parameters of the weather were studied separately, and they were treated as continuous variables. In contrast to the paper, all the five parameters were included in the dataset, so the Cloud Cover (CC) was added among other variables. The test conditions stayed the same. Test runs were conducted for both the whole dataset and to the short distances: one run with common betas and the other run with alternative specific betas resulting four runs in total.

3.2.1 Common Betas

The model of continuous weather variables with common betas and walking as a base:

$$V_{cycle} = ASC_{cycle} + b_{time} * Time + b_{TG} * TG + b_{SS} * SS + b_{RR} * RR + b_{CC} * CC + b_{FG} * FG$$

$$V_{drive} = ASC_{drive} + b_{time} * Time + b_{cost} * Cost + b_{TG} * TG + b_{SS} * SS + b_{RR} * RR + b_{CC} * CC + b_{FG} * FG$$

$$V_{pt} = ASC_{pt} + b_{time} * Time + b_{cost} * Cost + b_{TG} * TG + b_{SS} * SS + b_{RR} * RR + b_{CC} * CC + b_{FG} * FG$$

$$V_{walk} = b_{time} * Time$$

The respective results are presented in tables 5 and 6 below.

Table 5. Log-likelihoods for the whole dataset (continuous weather variables, common betas)						
Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x2 (0.05)
Base MNL	-74975.46	5	0	0	0	-
MNL	-74968.66	10	-6.8	5	13.6	11.071
Base NL	-73082.67	6	0	0	0	-
NL	-73072.02	11	-10.65	5	21.3	11.071

Table 6. Log-likelihoods for short distances (continuous weather variables, common betas)						
Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x2 (0.05)
Base MNL	-55243.27	5	0	0	0	-
MNL	-55234.83	10	-8.44	5	16.88	11.071
Base NL	-55229.15	6	0	0	0	-
NL	-54224.62	11	-1004.53	5	2009.06	11.071

Tables 5 and 6 shows again that the results aren't very promising for the models with common betas as they barely pass the Likelihood Ratio Test in most cases.

3.2.2 Alternative Specific Betas

The model of continuous weather variables with alternative specific betas and walking as a base:

$$V_{cycle} = ASC_{cycle} + b_{time} * Time + b_{TG,cycle} * TG + b_{SS,cycle} * SS + b_{RR,cycle} * RR + b_{CC,cycle} * CC + b_{FG,cycle} * FG$$

$$V_{drive} = ASC_{drive} + b_{time} * Time + b_{cost} * Cost + b_{TG} * TG + b_{SS,drive} * SS + b_{RR,drive} * RR + b_{CC,drive} * CC + b_{FG,drive} * FG$$

$$V_{pt} = ASC_{pt} + b_{time} * Time + b_{cost} * Cost + b_{TG,pt} * TG + b_{SS,pt} * SS + b_{RR,pt} * RR + b_{CC,pt} * CC + b_{FG,pt} * FG$$

$$V_{walk} = b_{time} * Time$$

The respective results are in tables 7 and 8 below.

Table 7. Log-likelihoods for the whole dataset (continuous weather variables, alternative specific betas)						
Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x2 (0.05)
Base MNL	-74975.46	5	0	0	0	-
MNL	-74911.75	20	-63.71	15	127.42	24.996
Base NL	-73082.67	6	0	0	0	-
NL	-73008.66	21	-74.01	15	148.02	24.996

Table 8. Log-likelihoods for short distances (continuous weather variables, alternative specific betas)						
Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x^2 (0.05)
Base MNL	-55243.27	5	0	0	0	-
MNL	-55181.8	20	-61.47	15	122.94	24.996
Base NL	-55229.15	6	0	0	0	-
NL	-54166.47	21	-1062.68	15	2125.36	24.996

Tables 7 and 8 points out that the results are much better compared to the common betas model and in most cases the results are better than even the scoring method used in scenario 1.

3.3 Scenario 3: Weather Condition Groups + Combination comparison

In the scenario 3, model was developed first by classifying the weather parameters into groups and then second by using dummy variables for these groups to check if the model could be improved.

3.3.1 Mean Temperature Groups

For the team, the most important condition and the most difficult one to assess was the mean temperature. That is because in comparison with the rest of the of the weather conditions the temperature's utility is not linear. This means for example in the case of precipitation that the higher it is, the less utility it gives to the choice of bicycle. The rest of the conditions (cloud cover, wind speed, sunshine duration) work similarly. The temperature, however, does not really work like that. As it can also be seen in the analysis of Thomas et. al. (2013), its utility is not linear, which means that there is an ideal temperature and when the temperature changes away from it to either side (hotter or colder), the utility decreases. To capture this attribute, classes will be created for the mean temperature of the dataset.

In the dataset the mean temperature (TG) varies between -1.9 and 25.4 degrees of Celsius. Some different variations of classes will be tried out:

- 4 classes: [-1.9, 5.6], (5.6, 13.1], (13.1, 20.6], (20.6, 281]
- 5 classes: [-1.9, 4.1], (4.1, 10.1], (10.1, 16.1], (16.1, 22.1], (22.1, 28.1]
- 6 classes: [-1.9, 3.1], (3.1, 8.1], (8.1, 13.1], (13.1, 18.1], (18.1, 23.1], (23.1, 28.1]
- 7 classes: [-1.9, 2.1], (2.1, 6.1], (6.1, 10.1], (10.1, 14.1], (14.1, 18.1], (18.1, 22.1], (22.1, 26.1]
- 5 classes changed (due to the small number of observations in the 1st, the 2nd, the 6th, and the 7th class above, we combine them to create 5 classes as follows): [-1.9, 6.1], (6.1, 10.1], (10.1, 14.1], (14.1, 18.1], (18.1, 26.1]

Then all the above groups were tested with the following model where public transport and temperature group 3 were the base:

$$V_{cycle} = ASC_{cycle} + b_{time} * Time + \sum_{\substack{i=1 \\ i\neq 3}}^{n} b_{TG_i,cycle} * TG_i \qquad , n = 4,5,6,7$$

$$V_{drive} = ASC_{drive} + b_{time} * Time + b_{cost} * Cost + \sum_{\substack{i=1 \\ i\neq 3}}^{n} b_{TG_i,drive} * TG_i \qquad , n = 4,5,6,7$$

$$\boldsymbol{V_{pt} = ASC_{pt} + b_{time} * Time + b_{cost} * Cost}$$

$$\boldsymbol{V_{walk} = b_{time} * Time + \sum_{\substack{i=1 \\ i\neq 3}}^{n} b_{TG_i,walk} * TG_i \qquad , n = 4,5,6,7}$$

The daily temperatures (TG_i) are dummy binary variables that show if a temperature belongs to the corresponding class (1 if it does, 0 otherwise) in this model.

The results of the comparison between the groups are presented in table 9 below.

Table 9. Results for the comparisons between number of different groups						
Number of groups	Log Likelihood	Parameters	LL difference	Par difference	LR test	x^2 (0.05)
4	-74924.78	14	0	0	0	0
5	-74917.64	17	-7.14	3	14.28	7.815
6	-74925.12	20	7.48	3	-14.96	7.815
7	-74913.43	23	-4.21	6	8.42	12.592
5 modified	-74926.39	17	8.75	0	-17.5	3.841

The above test was carried out only for the whole dataset with an MNL model as it has been seen that it is a good indicator for all the models.

The scenario 3 was started with 4 groups as a base. It can be observed that going from 4 to 5 groups, 3 parameters were added, but the result of log-likelihood get better and that it passes the Likelihood Ratio Test. So, by using now 5 groups as the base, it can be seen that using 6 groups gives worse LL. Then 5 groups were compared with 7 groups. The LL is better with 7 groups but by adding 6 more parameters the Likelihood Ratio Test shows that it is not worth to do. The modified 5 groups give an even worse LL so in the end, 5 groups were chosen to stand for the mean temperature.

3.3.2 Adding the rest of the weather conditions

The rest of the weather conditions was now added by using the same classification method. As already explained, the rest of the weather conditions can be considered linear in utility so there were only 4 groups for each parameter. Then it only needed to be tested which one of the conditions to add to the model by doing Likelihood Ratio Tests.

To do that, already created TG groups was used as a base. Then iteratively one condition was added at a time and compared each LL to the base. The one that had the best LL was added and the combination of the two became the base for the next iteration.

The first iteration results are presented in table 10.

Table 10. Log-likelihood comparison after the first iteration

Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base (TG)	-74917.64	17	0.00	0	0	0
TG + CC	-74901.11	26	-16.53	9	33.06	16.919
TG + RR	-74880.55	26	-37.09	9	74.18	16.919
TG + SS	-74893.28	26	-24.36	9	48.72	16.919
TG + FG	-74896.39	26	-21.25	9	42.5	16.919

The best LL is given by adding 4 groups of precipitation (RR) ([0, 1], (1, 2], (2, 3], [3, +∞)) and by doing the Likelihood Ratio Test it can be observed that the new model is better than the base.

Then by using the TG+RR as the base, the results of the second iteration are the ones in table 11.

Table 11. Log-likelihood comparisons after the second iteration

Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base (TG+RR)	-74880.55	26	0.00	0	0	0
TG + RR + SS	-74859.51	35	-21.04	9	42.08	16.919
TG + RR + FG	-74860.91	35	-19.64	9	39.28	16.919
TG + SS + CC	-74870.56	35	-9.99	9	19.98	16.919

The best LL is given by adding the 4 groups of sunshine duration (SS) ([0, 36.25], (36.25, 72.5], (72.5, 108.75], (108.75, 145]) and by doing the Likelihood Ratio Test the new model is better than the base.

For the third iteration the combination of TG + RR + SS was used as the base and compared to FG (wind speed) and CC (cloud cover). Results are shown in table 12.

Table 12. Log-likelihood comparisons after the third iteration						
Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base (TG + RR + SS)	-74859.51	35	0.00	0	0	0
TG + RR + SS + FG	-74840.37	44	-19.14	9	38.28	16.919
TG + RR + SS + CC	-74848.8	44	-10.71	9	21.42	16.919

The best LL is given by adding the 4 groups of wind speed (FG) ([7, 24.6667], (24.6667, 42.3333], (42.3333, 60], (60, +∞]) and by doing the Likelihood Ratio Test the new model is better than the base.

Finally, it was tested whether it is worth to add the last weather parameter CC (cloud cover) to the model and the results are shown in table 13.

Table 13. Log-likelihood comparisons after the fourth iteration						
Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base (TG + RR + SS + FG)	-74840.37	44	0.00	0	0	0
All parameter groups	-74829.21	53	-11.16	9	22.32	16.919

It can be seen from the Likelihood Ratio Test that even by adding the 4 cloud cover groups ([0, 2], (2, 4], (4, 6], (6, 8]) the model is improved and accepted.

So, the final model becomes:

$$\boldsymbol{V_{cycle} = ASC_{cycle} + b_{time} * Time + \sum_{\substack{i=1\\ i\neq 3}}^{5} b_{TG_i,cycle} * TG_i + \sum_{\substack{j=1\\ i\neq 3}}^{4} b_{RR_j,cycle} * RR_j}$$
$$\boldsymbol{+ \sum_{\substack{k=1\\ i\neq 3}}^{4} b_{SS_k,cycle} * SS_k + \sum_{\substack{l=1\\ i\neq 3}}^{4} b_{FG_l,cycle} * FG_l + \sum_{\substack{m=1\\ i\neq 3}}^{4} b_{CC_m,cycle} * CC_m}$$

$$\boldsymbol{V_{drive} = ASC_{drive} + b_{time} * Time + b_{cost} * Cost + \sum_{\substack{i=1\\ i\neq 3}}^{5} b_{TG_i,drive} * TG_i + \sum_{\substack{j=1\\ i\neq 3}}^{4} b_{RR_j,drive}} * RR_j$$
$$+ \sum_{\substack{k=1\\ i\neq 3}}^{4} b_{SS_k,drive} * SS_k + \sum_{\substack{l=1\\ i\neq 3}}^{4} b_{FG_l,drive} * FG_l + \sum_{\substack{m=1\\ i\neq 3}}^{4} b_{CC_m,drive} * CC_m$$

$$V_{pt} = ASC_{pt} + b_{time} * Time + b_{cost} * Cost$$

$$V_{walk} = b_{time} * Time + \sum_{\substack{i=1 \\ i\neq 3}}^{5} b_{TG_i,walk} * TG_i + \sum_{\substack{j=1 \\ i\neq 3}}^{4} b_{RR_j,walk} * RR_j + \sum_{\substack{k=1 \\ i\neq 3}}^{4} b_{SS_k,walk} * SS_k + \sum_{\substack{l=1 \\ i\neq 3}}^{4} b_{FG_l,walk} * FG_l + \sum_{\substack{m=1 \\ i\neq 3}}^{4} b_{CC_m,walk} * CC_m$$

3.3.3 Combining groups and continuous variables

One idea after applying the groups method was that it might be better to combine the groups for some parameters and add the rest as continuous variables. To test this idea, some previous iterations were taken and for each stage continuous variable was tested by comparing the groups for the same variable. For the first iteration, it was started with the base model and the mean temperature was tried to add either in groups or as a continuous variable.

The results for the first iteration (TG comparison) are presented in table 14.

Table 14. The results for average temperature (TG) comparison

Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base	-74975.46	5	0	0	0	0
TG groups	-74917.64	17	-57.82	12	115.64	21.026
TG continuous	-74933.2	8	-42.26	3	84.52	7.815

Groups seemed to perform better. The next iteration was the comparison for precipitation (RR) and it is presented in table 15.

Table 15. The results for precipitation (RR) comparison

Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base (TG)	-74917.64	17	0	0	0	0
TG + RR groups	-74880.55	26	-37.09	9	74.18	16.919
TG + RR continuous	-74909.00	20	-8.64	3	17.28	7.815

Again, the groups performed better. Next is the comparison for sunshine duration (SS) presented in table 16.

Table 16. The results for sunshine duration (SS) comparison						
Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base (TG + RR)	-74880.55	26	0	0	0	0
TG + RR + SS groups	-74859.51	35	-21.04	9	42.08	16.919
TG + RR + SS continuous	-74876.39	29	-4.16	3	8.32	7.815

Once again, the groups gave a better LL

The comparison for wind speed (FG) is presented in table 17.

Table 17. The results for wind speed (FG) comparison						
Model	LL	Parameters	LL difference	Par difference	LR	x^2 (0.05)
Base (TG + RR + SS)	-74859.51	35	0	0	0	0
TG + RR + SS + FG groups	-74840.37	44	-19.14	9	38.28	16.919
TG + RR + SS + FG continuous	-74859.28	38	-0.23	3	0.46	7.815

In this case the FG continuous is also rejected by the Likelihood Ratio Test

Finally, the comparison for cloud cover (CC) is presented in table 18.

Table 18. The results for cloud cover (CC) comparison						
Model	LL	Parameters	LL difference	Par difference	LR	x2 (0.05)
Base (TG + RR + SS)	-74840.37	44	0	0	0	0
All parameter groups	-74829.21	53	-11.16	9	22.32	16.919
TG + RR + SS + FG + CC continuous	-74835.56	47	-4.81	3	9.62	7.815

As can be seen from tables 14 to 18, in every case the groups perform better than continuous variables. Since many parameters were added each time, the calculation time also greatly increased. All the tests were managed to run by the all-groups model. If some variables had to be continuous instead of the groups, they would be the ones that have the lowest LL difference between the base which in this case would be the mean temperature (TG) (-57.82 – (42.26) = -15.56) and the cloud cover (CC) (-11.16 – (-4.81) = -6.35).

The results for the all-groups model are presented in tables 19 and 20.

Table 19. The results of all-groups model comparison for the whole dataset

Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x^2 (0.05)
Base MNL	-74975.46	5	0	0	0	-
MNL	-74829.21	53	-146.25	48	292.5	65.171
Base NL	-73082.67	6	0	0	0	-
NL	-72927.98	54	-154.69	48	309.38	65.171

Table 20. The results of all-groups model comparisons for short distances

Model	Log Likelihood	Parameters	LL difference	Par difference	LR test	x^2 (0.05)
Base MNL	-55243.27	5	0	0	0	-
MNL	-55121.9	53	-121.37	48	242.74	65.171
Base NL	-55229.15	6	0	0	0	-
NL	-54104.87	54	-1124.28	48	2248.56	65.171

4 Results

The results of this study were obtained from a dataset containing trip mode choice behaviour and daily weather information. The latter dataset was used first as a whole dataset and second as a sample containing only trips that were shorter than five kilometres. The specific sample was set to study the mode choice behaviour in short distances and then to compare these results to the whole dataset. For each scenario both MNL model and NL model were used. The results of the first and the third scenarios are in Appendix 1. The weather score distributions can be found in Appendix 2.

Three scenarios were studied: the weather score, continuous variables, and weather combination groups with comparisons. In the weather score scenario, the base model was set up, and then models with common and alternative specific betas. The base model only describes the influence of travel time and cost to mode choice. The other weather score models studied the influence of weather based on the non-linear scoring presented in the paper by Thomas et. al. (2013) both with common and alternative specific betas. The continuous variables describe instead the effect of different weather components as linear and continuous. This was also studied both with common and alternative specific betas. The weather combination groups studied iteratively the influence of different weather combinations to mode choice starting from the combination of base model and mean temperature and adding new weather parameters one by one based on the best comparison result of the previous iteration round. The main results are expressed below scenario by scenario.

4.1 The weather score scenario

The base scenario was the first part of the weather scenario and the simplest one that was tested covering only cost and time as variables. Result tables of the final log-likelihoods, alternative specific constants, and beta coefficients for both models and mu-value (μ) for nested logit model are presented in Appendix 1A. Walking was chosen as base mode for comparison and therefore no alternative specific constant was defined for that mode. As the results show in Appendix 1A, both time and cost are considered significant factors behind mode choice in all journeys but also in short journeys. Time is a more significant factor than cost in both datasets. The rest of the models in our study within this and other scenarios was compared to these two base models, one with MNL and another one with NL model.

For the whole dataset, the MNL model with common betas shows no change in Log Likelihood (LL) when the weather variable is added. The NL model with common betas also shows no change in LL when the weather variable is added. Overall, the NL model performs better than the MNL model. For short distances (shorter than 5 km), the MNL model with common betas shows no change in LL when the weather variable is added. The NL model with common betas instead shows a significant decrease in LL when the weather variable is added. The likelihood ratio test (LR test) indicates that the difference in LL is statistically significant. The base MNL and base NL models show no change in LL compared to the models with the weather variable.

To conclude, in both the whole dataset and short distances, the NL model generally performs better than the MNL model. The addition of the weather variable has little to no effect on the LL in most cases. When alternative-specific betas are introduced, the models show improvement in LL compared to the base models. The LR test confirms the significance of the new models. These findings suggest that weather conditions, as represented by the weather score, have a limited impact on mode choice behaviour in the given dataset.

4.2 The continuous variables scenario

In the second scenario, all weather variables (TG, SS, RR, CC, FG) are treated as continuous variables. This scenario studied models with both common betas and alternative-specific betas. In the model with common betas across alternatives for each weather variable, the MNL model shows a slight improvement in Log Likelihood (LL) compared to base model, but the difference is not significant based on the Likelihood Ratio Test. The NL model with common betas also shows a small improvement in LL for both the whole dataset and short distances, but again the difference is not statistically significant. Overall, the common betas models do not show significant improvements compared to the base models.

In the case of alternative-specific betas, the weather variables (TG, SS, RR, CC, FG) are also treated as continuous variables. The models include alternative-specific betas for each weather variable. For both the whole dataset and short distances, the MNL model with alternative-specific betas shows a significant improvement in LL compared to the base model, as confirmed by the Likelihood Ratio Test. The NL model with alternative-specific betas also shows a significant improvement in LL for both the whole dataset and short distances. Overall, the alternative-specific betas models perform better than the common betas models and show significant improvements in LL.

To conclude, weather variables were handled as continuous variables and using alternative-specific betas to yield better results compared to the common betas approach. The addition of individual weather variables improves the model's ability to explain mode choice behaviour, as evidenced by the significant improvements in LL. The NL model with alternative-specific betas outperforms the MNL model. These findings suggest that considering individual weather variables separately as continuous variables is a more effective approach in modelling the relationship between weather and mode choice behaviour.

4.3 The weather combination groups with comparisons scenario

The third scenario of the project is an improved version of a scoring system that was presented on Tue 25th of April 2023. In this scenario, weather conditions were grouped together one parameter at time and then compared to each other. Mean temperature was chosen to be the first iteration together with base model. Different variations of mean temperature classes were tested, ranging from 4 to 7 groups. The Likelihood Ratio Test was used to compare the classes. Among the tested variations, the model with 5 temperature groups provided the best log likelihood. This was chosen to represent the mean temperature.

The remaining weather conditions (precipitation, cloud cover, wind speed, sunshine duration) were added to the model using 4 groups for each condition. The Likelihood Ratio Test was used to determine which weather condition to add next. Iterative comparisons were performed by adding one condition at a time to the base model. The condition that improved the log likelihood the most was chosen for each iteration. The final model included the 5 temperature groups, 4 precipitation groups, 4 sunshine duration groups, 4 wind speed groups, and 4 cloud cover groups.

Comparison between Groups and Continuous Variables was conducted to compare the performance of using groups versus continuous variables for each weather condition. In each iteration, the base model was compared to a model where the corresponding weather condition was included as a continuous variable. The Likelihood Ratio Test was used to compare the models. For all the weather conditions (mean temperature, precipitation, sunshine duration), the models using groups outperformed the models using continuous variables.

The idea of combining groups for some parameters and using continuous variables for others was tested. The iterations from the previous steps were revisited, and for each stage, the continuous variable was compared to the corresponding groups. In all cases, the models using groups provided better log likelihoods compared to the models using continuous variables.

5 Discussion and conclusions

Based on the findings, it can be concluded that using classes (groups) for the weather conditions, as well as combining groups and continuous variables, improves the performance of the model in predicting the choice of transportation mode.

We can see in every case the groups perform better than continuous variables. However, since we are adding a lot more parameters each time the calculation time also greatly increases. We managed to use the all-groups model to run all our tests but if we had to choose some variables to be continuous instead of in groups, it would be the ones that have the lowest LL difference between the groups the continuous and the base which in this case would be the mean temperature (TG) and the cloud cover (CC). With these being said we concluded that this report best represents how with more variables and complexity of the model we get better results but overall, the findings stay the same in a sense that the results do not show a notable change in behaviour which leads us to believe that with further investigation and testing with more complex models and variables we could further improve the results and perhaps find a change in the behaviour but do to the limitations we weren’t able to confirm this.

The analysis of the provided data indicated a positive correlation between higher temperatures and the usage of active modes of transportation, such as walking or cycling. This finding aligns with previous research, which suggests that favourable weather conditions, including warmer temperatures, can encourage individuals to engage in physical activities and choose active modes of transportation. Warmer weather often enhances the comfort and enjoyment of walking and cycling, making these modes more appealing options for short-distance travel. Moreover, higher temperatures may also decrease the reliance on motorized transportation due to concerns related to congestion, parking availability, and environmental sustainability.

On the other hand, wind speed demonstrated a negative relationship with active modes of transportation in the analysis. This finding implies that as wind speed increases, individuals may be less likely to choose walking or cycling as their preferred mode of transportation. Intense winds can create discomfort and inconvenience for pedestrians and cyclists, making these modes less attractive, particularly for longer journeys. The influence of wind speed on mode choices may be linked to concerns such as safety, the physical effort required to navigate against the wind, and the potential for unpleasant experiences, such as windblown debris or unstable cycling conditions.

Comparing the influence of the different weather components, precipitation, particularly rain, appears to have a more significant impact on mode choices compared to temperature and wind speed. While higher temperatures and favourable weather conditions may promote the use of active modes of transportation, the effect size is smaller compared to the influence of rain. Rainfall presents a more tangible barrier to personal vehicle usage due to the inconveniences, discomfort, and safety concerns associated with wet road conditions and reduced visibility. Consequently, individuals are more likely to shift towards alternative modes of transportation, such as public transit or cycling, when faced with rainy weather.

However, it is important to note that weather components do not operate in isolation but rather interact with one another to shape mode choices. For example, the influence of temperature and wind speed on mode choices may be moderated by the presence of precipitation. A combination of rain, high winds, and low temperatures can significantly deter individuals from choosing active modes of transportation and may even lead to increased reliance on private vehicles or public transit.

In conclusion, the analysis of the provided data suggests that precipitation, particularly rain, has a significant influence on mode choices. During rainy weather conditions, individuals are more likely to opt for alternative modes of transportation, such as public transit or cycling. This observation aligns with existing research, which highlights the role of adverse weather conditions in discouraging personal vehicle usage. Nonetheless, it is crucial to acknowledge that other weather components, such as temperature and wind speed, may also contribute to variations in transportation behaviour. Future research should delve further into these relationships, considering additional factors and employing advanced modelling techniques, to provide a more comprehensive understanding of the complex interplay between weather conditions and mode choices.

Appendix 1. The results of our scenarios

The significance of the results is studied with a t-test and those values are presented in parenthesis. All t-test values higher than 1.96 or smaller than -1.96 are considered here as meaningful results.

Appendix 1A. The base model

Clarifications for reading the tables below:

- *ASCs* refer to alternative specific constants for each mode

- *b_cost* and *b_time* refer to the cost and time coefficients and they are same for all modes except walking and cycling was assumed to have no cost at all

- *mu* is the value describing the relation between nests

Base model results for the whole dataset			Base model results for short journeys (shorter than 5 km)		
	MNL	NL		MNL	NL
Final log-likelihood	-74975.46	-73082.67	Final log-likelihood	-55243.27	-55229.15
ASC_PT	-0.501 (-37.5)	-1.98 (-50.4)	ASC_PT	-0.941 (-58.7)	-0.944 (-58.4)
ASC_cycle	-3.8 (-144)	-4.93 (-119)	ASC_cycle	-4.23 (-133)	-4.5 (-76.9)
ASC_drive	-1.22 (-62.1)	-3.07 (-57.9)	ASC_drive	-1.6 (-69.4)	-1.62 (-68.9)
b_cost	-0.171 (-51)	-0.342 (-60.5)	b_cost	-0.224 (-31.2)	-0.226 (-30.9)
b_time	-5.31 (-107)	-9.14 (-73.9)	b_time	-7.38 (-101)	-7.5 (-98.4)
mu		0.389 (64.9)	mu		-0.88 (47.3)

Appendix 1B. The individual weather scoring

Clarifications for reading the tables below:

- *ASCs* refer to alternative specific constants for each mode

- *b_cost* and *b_time* refer to the cost and time coefficients and they are the same for all modes. Walking and cycling were considered to have no cost at all.

- *mu* is the value describing the relation between nests

Individual parameters, common betas for the whole dataset			Individual parameters, common betas for short journeys (shorter than 5 km)		
	MNL	NL		MNL	NL
Final log-likelihood	-74968.66	-73072.02	Final log-likelihood	-55234.83	-54224.62
ASC_PT	-0.446 (-8.12)	-1.9 (-26.4)	ASC_PT	-0.87 (-14.8)	-2.23 (--26.3)
ASC_cycle	-3.74 (-63.1)	-4.86 (-67)	ASC_cycle	-4.16 (-64.1)	-4.88 (-65.5)
ASC_drive	-1-17 (-20.6)	-2.99 (-37.1)	ASC_drive	-1.52 (-24.9)	-2.72 (-34.3)
b_cost	-0.1717 (-51)	-0.341 (-60)	b_cost	-0.224 (-31.2)	-0.342 (-33.1)
b_time	-5.31 (-107)	-9.13 (-72.4)	b_time	-7.39 (-101)	-10.2 (-81.4)
b_cc	-0.00884 (-1.29)	-0.00789 (-1.03)	b_cc	-0.00995 (-1.37)	-0.00993 (-1.24)
b_fg	-0.000367 (0.592)	0.000708 (1.02)	b_fg	0.000478 (0.727)	0.000747 (1.03)
b_rr	-3.54e-05 (-2.22)	-4.19e-05 (-2.4)	b_rr	-3.9e-05 (-2.33)	-4.36e-05 (-2.41)
b_ss	-0.00123 (--2.88)	-0.00134 (-2.81)	b_ss	-0.00141 (-3.11)	-0.00155 (-3.11)
b_tg	-0.00017 (0.887)	-6.62e-05 (-0.307)	b_tg	0.000112 (0.547)	-8.72e-05 (0.000225)
mu		0.391 (60.3)	mu		0.436 (46.3)

Individual parameters, alternative specific betas for the whole dataset			Individual parameters, alternative specific betas for short journeys (shorter than 5 km)		
	MNL	NL		MNL	NL
Final log-likelihood	-74911.75	-73008.66	Final log-likelihood	-55181.8	-54166.47
ASC_PT	-0.393 (-6.74)	-1.8 (-20.4)	ASC_PT	-0.825 (-12.3)	-2.09 (-19.3)
ASC_cycle	-3.93 (-30.4)	-5.08 (-35.7)	ASC_cycle	-4.35 (-29.4)	-5.14 (-32.6)
ASC_drive	-1.21 (-18.7)	-3.09 (-29.9)	ASC_drive	-1.55 (-22.3)	-2.81 (-27.9)
b_cost	-0.172 (-51.1)	-0.341 (-60.1)	b_cost	-0.224 (-31.3)	-0.345 (-33.1)
b_time	-5.31 (-107)	-9.14 (-72.4)	b_time	-7.39 (-101)	-10.2 (-81.4)
b_cc_cycle	-0.0499 (-3.06)	-0.0531 (-3.03)	b_cc_cycle	-0.0646 (-3.44)	-0.0636 (-3.23)
b_cc_drive	-0.000774 (-0.0977)	0.00474 (0.421)	b_cc_drive	-0.00489 (-0.582)	-0.0029 (-0.259)
b_cc_pt	-0.0135 (-1.86)	-0.0174 (-1.72)	b_cc_pt	-0.0103 (-1.24)	-0.0138 (-1.13)
b_fg_cycle	0.000723 (0.473)	0.00134 (0.819)	b_fg_cycle	0.00216 (1.22)	0.00288 (1.56)
b_fg_drive	0.000876 (1.23)	0.00217 (2.16)	b_fg_drive	0.00123 (1.62)	0.00255 (2.55)
b_fg_pt	-0.000158 (-0.238)	-0.000831 (-0.89)	b_fg_pt	-0.000618 (-0.808)	-0.00195 (-1.72)
b_rr_cycle	3.36e-05 (0.742)	2.24e-05 (0.478)	b_rr_cycle	6.4e-05 (1.11)	6.13e-05 (1.04)
b_rr_drive	-4.14e-05 (-2.3)	-5.34e-05 (-2.16)	b_rr_drive	-3.39e-05 (-1.76)	-2.78e-05 (-1.08)
b_rr_pt	-3.35e-05 (-1.95)	-3.3e-05 (-1.4)	b_rr_pt	-5.21e-05 (-2.87)	-7.08e-05 (-2.9)
b_ss_cycle	-0.00214 (-2.17)	-0.00239 (-2.26)	b_ss_cycle	-0.00286 (-2.54)	-0.00296 (-2.5)
b_ss_drive	-0.0012 (-2.42)	-0.00138 (-1.95)	b_ss_drive	-0.163 (-3.1)	-0.00205 (-2.92)
b_ss_pt	-0.0.00119 (-2.63)	-0.00122 (-1.93)	b_ss_pt	-0.000939 (-1.82)	-0.000684 (-0.903)
b_tg_cycle	0.00327 (7.2)	0.00337 (6.94)	b_tg_cycle	0.0034 (6.59)	0.0035 (6.58)
b_tg_drive	1.37e-05 (0.062)	-0.000243 (-0.771)	b_tg_drive	-3.09e-05 (-0.131)	-6.9e-0.5 (-0.219)
b_tg_pt	6.73e-05 (0.329)	-0.000148 (-0.519)	b_tg_pt	-4.93e-05 (-0.21)	-0.000504 (-1.47)
mu		0.391 (60.3)	mu		0.435 (46.3)

Appendix 1C. The paper scoring

Paper scoring, alternative specific betas for the whole dataset			Paper scoring, alternative specific betas for short journeys (shorter than 5 km)		
	MNL	NL		MNL	NL
Final log-likelihood	-74934.36	-73031.64	Final log-likelihood	-55209.52	-54197.1
b01	-4.04 (-92)	-5.17 (-91.6)	b01	-4.46 (-87)	-5.2 (-51.8)
b02	-1.18 (-48)	-2.99 (-51.1)	b02	-1.54 (-55.5)	-2.72 (-51.8)
b03	-0.486 (-25.3)	-1.95 (-42.7)	b03	-0.926 (-40.8)	-2.29 (-36.8)
b2	-0.172 (-51.1)	-0.342 (-60.1)	b2	-0.224 (-31.3)	-0.344 (-33.2)
b3	-5.31 (-107)	-9.13 (-72.4)	b3	-7.39 (-101)	-10.2 (-81.4)
b_wPT	-0.00869 (-5.59)	-0.00433 (-1.73)	b_wPT	-0.00756 (-3.91)	0.00544 (1.78)
b_wcycle	0.0271 (8.07)	0.033 (9.32)	b_wcycle	0.0265 (6.8)	0.0412 (10.3)
b_wdrive	-0.0121 (-7.76)	-0.0123 (-4.93)	b_wdrive	-0.0136 (-7.54)	-0.00309 (-1.16)
b_wwalk	-0.00632 (-3.54)	-0.000699 (-0.351)	b_wwalk	-0.00537 (-2.75)	0.00926 (4.39)
mu		0.391 (60.3)	mu		0.436 (46.4)

Paper scoring, common betas for the whole dataset			Paper scoring, common betas for short journeys (shorter than 5 km)		
	MNL	NL		MNL	NL
Final log-likelihood	-74975.46	-73082.1	Final log-likelihood	-55243.27	-54236.35
b01	-3.8 (-144)	-4.93 (-117)	b01	-4.23 (-133)	-4.97 (-120)
b02	-1.22 (-62.1)	-3.06 (-55.6)	b02	-1.6 (-69.4)	-2.81 (-57.7)
b03	-0.501 (-37.5)	-1.97 (-47.7)	b03	-0.941 (-58.7)	-2.32 (-40.7)
b2	-0.171 (-51)	-0.341 (-60)	b2	-0.224 (-31.2)	-.0.342 (-33)
b3	-5.31 (-107)	-9.12 (-72.5)	b3	-7.38 (-101)	-10.2 (-81.6)
b_W	-2.26e-15 (-1.48e-323)	-1.29e-15 (-927)	b_W	9.91e-16 (51.1)	-1.46e-12 (-1.01e+05)
mu		0.392 (60.3)	mu		0.436 (46.7)

Appendix 1D. The parameter groups

Model	LL	Par	LL diff	Par diff	LR	Chi squared (0.05)	Accepted
The first iteration							
Base (TG)	-74917.64	17	0.00	0	0	0	-
TG + CC	-74901.11	26	-6.53	9	33.06	16.919	yes
TG + RR	-74880.55	26	-37.09	9	74.18	16.919	yes
TG + SS	-74893.28	26	-24.36	9	48.72	16.919	yes
TG + FG	-74896.39	26	-21.25	9	42.5	16.919	yes

The second iteration							
Base (TG + RR)	-74880.55	26	0.00	0	0	0	-
TG + RR + SS	-74859.51	35	-21.04	9	42.08	16.919	yes
TG + RR + FG	-74860.91	35	-19.64	9	39.28	16.919	yes
TG + SS + CC	-74870.56	35	-9.99	9	19.98	16.919	yes

The third iteration							
Base (TG + RR + SS)	-74859.51	35	0.00	0	0	0	-
TG + RR + SS + FG	-74840.37	44	-19.14	9	38.28	16.919	yes
TG + RR + SS + CC	-74848.8	44	-10.71	9	21.42	16.919	yes

The fourth iteration							
Base (TG + RR + SS + FG)	-74859.51	44	0.00	0	0	0	-
All parameters	-74829.21	53	-30.30	9	60.6	16.919	yes

Appendix 2. Descriptive statistics of the weather parameters

Reading instructions for the graphs below:

- Classification limits presented for all graphs as such that [and] includes the values to the limits, and (and) excludes the values from the limits.