



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat d'Informàtica de Barcelona

MASTER IN INNOVATION AND RESEARCH IN INFORMATICS  
SPECIALIZATION IN DATA SCIENCE

---

# Optimizing vehicle profile speed settings based on historic data

---

*Author:*

Ester LORENTE GARCÍA

*Advisor:*

Esteve CODINA SANCHO  
(DEIO - UPC)

*Director:*

Jaume BARCELÓ I BUGEDA  
(PTV Group)

*Co-Director:*

Ralf MIKULLA  
(PTV Group)

5th July 2019



# Acknowledgments

En primer lugar me gustaría agradecer a PTV Group por la oportunidad que me dio al ofrecerme realizar este proyecto de fin de máster.

Agradecer a mis dos directores, Jaume Barceló y Ralf Mikulla, y a mi tutor, Esteve Codina, por su apoyo y supervisión durante todo el proyecto. Y sobre todo por su asesoramiento en las partes más técnicas relacionadas con los temas de transporte.

También agradecer a Christian Mähler, mi responsable dentro de PTV Group, porque a pesar de no ser mi director de proyecto me ha ayudado y dado algunos consejos para mejorar.

Agradecer especialmente a Lidia Montero, por recomendarme para este trabajo y por aconsejarme en temas más de estadística.

Y por supuesto, quiero agradecer a Marc Vila por todo el apoyo durante el desarrollo de este proyecto y en las correcciones de la memoria.

Por último, quiero agradecer a mi familia por toda el apoyo que me han dado, porque sin ellos no podría haberlo conseguido.

Gracias!



# Abstract

PTV Group offers a large variety of trip planning, tour optimisation and traffic management and simulation tools. In particular, they provide the PTV xRoute service responsible for calculating routes between different locations taking into account multiple road factors as well as different restrictions depending on the type of vehicle.

The normal use of this service comes from transportation companies that require the daily route planning of their trucks. Where a small deviation in traveltimes for long-distance routes is not critical. However, it is currently being used in real-time products, where the calculation of traveltimes has to be very accurate in order to reproduce the real traffic situation.

One way to influence these traveltimes is through their so-called speed profiles for different networkclasses (a hierarchy of importance of the links) of the traffic network. Thus, varying the speeds associated to the networkclasses, the shortest paths and the related traveltimes can change and be used as a calibration method.

It is commonly used Floating Car Data (FCD) as reference data for this type of calibration. However, for this project we will study the usage of historical traffic data from real-time navigation services (such as HERE, Google, etc). Since it is more easily accessible and has a higher coverage of the network as opposed to the FCD that has only information where the connected vehicles have crossed.

In this way, this project measures the differences between the historical traveltimes from the navigation service and the calculated traveltimes given some speed profiles with the PTV xRoute service.

As well as evaluating different mechanisms to optimise these speed profiles in an appropriate way to minimise the existing deviations in traveltimes between both services. Defining several optimization models based on reference data at two different levels of detail: using segment traveltimes or total traveltimes.

It also evaluates the quality of the different mechanisms found, determining the advantages and disadvantages of each method in general.

**Keywords: optimisation models, distance analysis, historical traffic data, speed profiles optimisation**

# Resumen

PTV Group ofrece una larga variedad de herramientas de planificación de viajes, optimización de rutas y gestión de tráfico y de simulación. En concreto, disponen del servicio de PTV xRoute, responsable de calcular rutas entre diferentes ubicaciones teniendo en cuenta múltiples factores de las carreteras así como diferentes restricciones según el tipo de vehículo.

El uso habitual de este servicio proviene de empresas logísticas que requieren de una planificación de las rutas diarias de sus camiones. Donde una pequeña desviación en los tiempos de viaje para rutas largas no es primordial. Pero actualmente se empieza a usar en productos de real-time, donde el cálculo de los tiempos de viaje tiene que ser muy preciso para poder reproducir la situación real del tráfico.

Una forma de influenciar estos tiempos de viaje es mediante sus llamados perfiles de velocidad para diferentes networkclasses (jerarquía de importancia de los links) de la red de tráfico. De este modo, variando las velocidades asociadas a las networkclasses los caminos mínimos y los tiempos de viaje asociados varían y pueden usarse como método de calibración.

Comúnmente se utilizan los Floating Car Data (FCD) como datos de referencia para este tipo de calibrado. Pero para este proyecto se estudia el uso de los datos históricos de tráfico provenientes de servicios de navegación en tiempo real (como HERE, Google, etc). Ya que son más fácilmente accesibles y tienen una mayor cobertura de la red, al contrario que los FCD que solo disponen de información por donde los vehículos conectados han cruzado.

De este modo, este proyecto mide las diferencias entre los tiempos de viaje históricos de un servicio de navegación y los tiempos de viaje calculados dados unos perfiles de velocidad con el servicio de PTV xRoute.

Así como también evalúa diferentes mecanismos para optimizar estos perfiles de velocidad de una forma apropiada para minimizar las desviaciones existentes en los tiempos de viaje entre ambos servicios. Definiendo diferentes modelos de optimización basados en datos de referencia a dos niveles de detalle diferentes: usando tiempos de viajes a nivel de segmento o con tiempos totales de viaje.

También evalúa la calidad de los diferentes mecanismos encontrados, determinando las ventajas y desventajas de cada método en general.

**Palabras clave: modelos de optimización, análisis de distancias, datos históricos de tráfico, optimización de perfiles de velocidad**

# Glossary

## **HERE Routing**

HERE Routing is a routing service provided by HERE. It allows to calculate routes between different locations taking into account several road factors (e.g. allowed zones...) and for the specific case of truck routing, indicate some restrictions (e.g. weight, width, height...).

## **Key Performance Indicator (KPI)**

A KPI is a key performance indicator that allows to quantify the results of a given strategy. In traffic, the KPIs collected are the different traffic variables. Such as the traveltime, distance, speed, flow, fuel consumption...

## **Maximum Absolute Error (MAXAE)**

The MAXAE is a measure of the difference between two continuous variables. It indicates the maximum deviation of the absolute differences between both variables. Being a deviation of 0 the ideal value.

## **Mean Absolute Error (MAE)**

The MAE is a measure of the difference between two continuous variables. It indicates the average deviation of the absolute differences between both variables. Being a deviation of 0 the ideal value.

## **Networkclass (NC)**

The networkclasses are a hierarchy of the network segments in PTV Group products that classifies them according to their importance.

There is a total of eight networkclasses available: 0-Motorway, 1-Highway, 2-Trunk road, 3-Country road, 4-City street, 5-Residential street, 6-Drivable but normally blocked roads and 7-Cycle lanes and walkways.

## **PTV xRoute**

PTV xRoute is a service included in the PTV xServer software. It allows to calculate routes between different locations taking into account multiple road factors (e.g. speeds, toll costs...) as well as different restrictions depending on the type of vehicle (e.g. weight, width, height...).

## **Segment**

A segment is an element of the PTV Group network that defines the link between two nodes of the network.

These segments are classified according to a hierarchy of networkclasses about their importance and a subcategory of speedclasses about the speed restrictions in the networkclass.

---

### **Speedclass (SC)**

Aside from the networkclasses, the speedclasses are a subhierarchy of the network segments in PTV Group products that classify them according to speed restrictions from the same networkclass.

There is a total of eight speedclasses: 0-Fastest, 1-Faster, 2-Fast, 3-Medium fast, 4-Medium slow, 5-Slow, 6-Slower and 7-Slowest.

### **Speed profiles / speed settings**

The speed profiles or speed settings is the speeds descriptor of PTV Group vehicles. Defining the speeds used in the network segments.

This speed parameter is specified as a min-max range for each networkclass. Being the speeds assigned to each speedclass of the networkclass computed by a linear interpolation from that range.

# Contents

<b>Glossary</b>	<b>i</b>
<b>Contents</b>	<b>iii</b>
<b>List of figures</b>	<b>vi</b>
<b>List of tables</b>	<b>xi</b>
<b>List of codes</b>	<b>xiii</b>
<b>Chapter 1: Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Contextualisation . . . . .	1
1.3 Problem formulation . . . . .	4
1.3.1 Motivation . . . . .	5
1.3.2 Main objectives . . . . .	6
1.4 Scope of the project . . . . .	7
1.4.1 Possible obstacles . . . . .	7
1.5 Master thesis outline . . . . .	8
<b>Chapter 2: State of art</b>	<b>9</b>
2.1 Link traveltime estimation problem . . . . .	9
2.1.1 Floating Car Data (FCD) . . . . .	9
2.1.2 Historical traffic data . . . . .	10
2.2 Conclusions . . . . .	11
<b>Chapter 3: System specification</b>	<b>13</b>
3.1 Data description . . . . .	13
3.1.1 PTV xRoute service . . . . .	13
3.1.2 HERE Routing service . . . . .	15
3.1.3 Problematic issue: integration of both services . . . . .	15
3.2 Data pipeline . . . . .	16
3.2.1 Data generation . . . . .	17
3.2.2 Data preparation . . . . .	17
3.2.3 Data refinement . . . . .	18
3.2.4 Speed profiles computation . . . . .	18
3.2.5 Data generation (calibrated) . . . . .	19
3.3 Filtering methods . . . . .	20
3.3.1 Maximum absolute distance error filter (MAXD) . . . . .	20
3.3.2 Maximum absolute distance percentage error filter (MAXDP) . . . . .	20
3.3.3 Euclidean distance filter (EUCL) . . . . .	21

3.3.4	Dynamic euclidean distance filter (DYN-EUCL)	21
3.4	Proposed optimisation models	22
3.4.1	Segment traveltime optimisation models (ST)	22
3.4.1.1	Segment traveltime optimisation by segment error (ST_SE)	22
3.4.2	Trip traveltime optimisation models (TT)	24
3.4.2.1	Trip traveltime optimisation by segment error (TT_SE)	24
3.5	Validation methods	25
<b>Chapter 4: System architecture</b>		<b>27</b>
4.1	Architecture	27
4.1.1	Extensibility	28
4.2	Technologies	30
<b>Chapter 5: Results</b>		<b>31</b>
5.1	Area coverage	31
5.1.1	Barcelona Metropolitan Area (BCN)	31
5.1.2	Reduced Barcelona Metropolitan Area (BCN-R)	33
5.1.3	Data source integration limitations	36
5.2	Distance analysis	38
5.2.1	Segment distance analysis	38
5.2.1.1	Initial analysis	38
5.2.1.2	Maximum absolute distance error filter (MAXD)	39
5.2.1.3	Maximum absolute distance percentage error filter (MAXDP)	43
5.2.2	Trip distance analysis	46
5.2.2.1	Initial analysis	46
5.2.2.2	Maximum absolute distance error filter (MAXD)	47
5.2.2.3	Maximum absolute distance percentage error filter (MAXDP)	50
5.2.2.4	Euclidean distance filter (EUCL)	54
5.2.2.5	Dynamic euclidean distance filter (DYN-EUCL)	55
5.2.3	Conclusions	58
5.3	Speed calibration results	59
5.3.1	Area 1: Barcelona Metropolitan Area (BCN)	61
5.3.2	Area 2: Reduced Barcelona Metropolitan Area (BCN-R)	68
5.3.3	Conclusions	75
5.4	Time performance	75
5.5	Validation	78
<b>Chapter 6: Conclusions</b>		<b>81</b>
6.1	Achievement of objectives	81
6.2	Contributions	82
6.3	Further research	82

6.4	Personal assessment . . . . .	83
<b>References</b>		<b>85</b>
<b>Appendix</b>		<b>87</b>
A	Routing APIs examples . . . . .	87
A.1	PTV xRoute API - Calculate Route . . . . .	87
A.2	HERE Routing API - Calculate Route . . . . .	90
B	Speedclass computation . . . . .	93
C	Euclidean distance factor . . . . .	94
D	Area coverage . . . . .	96
D.1	Area 1: Barcelona Metropolitan Area (BCN) . . . . .	96
D.1.1	Sample BCN-1000: BCN area with 1,000 trips . . . . .	96
D.1.2	Sample BCN-3000: BCN area with 3,000 trips . . . . .	97
D.2	Area 2: Reduced Barcelona Metropolitan Area (BCN-R) . . . . .	98
D.2.1	Sample BCN-R-1000: BCN-R area with 1,000 trips . . . . .	98
D.2.2	Sample BCN-R-3000: BCN-R area with 3,000 trips . . . . .	99
E	Distance analysis . . . . .	101
E.1	Segment distance analysis . . . . .	102
E.1.1	Initial analysis . . . . .	102
E.1.2	Maximum absolute distance error filter (MAXD) . . . . .	104
E.1.3	Maximum absolute distance percentage error filter (MAXDP) . . . . .	106
E.2	Trip distance analysis . . . . .	108
E.2.1	Initial analysis . . . . .	108
E.2.2	Maximum absolute distance error (MAXD) . . . . .	110
E.2.3	Maximum absolute distance percentage error filter (MAXDP) . . . . .	112
E.2.4	Euclidean distance filter (EUCL) . . . . .	114
E.2.5	Dynamic euclidean distance filter (DYN-EUCL) . . . . .	116
F	Speed calibration results . . . . .	118
F.1	Area 1: Barcelona Metropolitan Area (BCN) . . . . .	118
F.1.1	Speed profiles variation . . . . .	118
F.1.2	Speeds distribution with ST models . . . . .	119
F.1.3	Traveltime improvements . . . . .	121
F.2	Area 2: Reduced Barcelona Metropolitan Area (BCN-R) . . . . .	124
F.2.1	Speed profiles variation . . . . .	124
F.2.2	Speeds distribution with ST models . . . . .	125
F.2.3	Traveltime improvements . . . . .	127

# List of Figures

## Chapter 1: Introduction

1.1 MaaS concept . . . . .	1
1.2 PTV Map&Guide . . . . .	3
1.3 Variation of mean traveltime depending on the time slot . . . . .	6

## Chapter 2: State of art

2.1 Map-matching example. . . . .	9
-----------------------------------	---

## Chapter 3: System specification

3.1 PTV xRoute response result lists . . . . .	13
3.2 Speedclass interpolation and max speed for one networkclass . . . . .	15
3.3 Matching routes from two networks . . . . .	16
3.4 Iterative data pipeline . . . . .	17

## Chapter 4: System architecture

4.1 System architecture . . . . .	29
-----------------------------------	----

## Chapter 5: Results

5.1 Barcelona area coverage with 1,000 trips . . . . .	32
5.2 Barcelona area coverage with 3,000 trips . . . . .	33
5.3 Barcelona reduced area coverage with 1,000 trips . . . . .	34
5.4 Barcelona reduced area coverage with 3,000 trips . . . . .	35
5.5 Route splitting by Batches . . . . .	37
5.6 Analysis of MAXD filter factor in segment distances of BCN-1000 . . . . .	41
5.7 Analysis of MAXD filter factor in segment distances of BCN-3000 . . . . .	41
5.8 Analysis of MAXD filter factor in segment distances of BCN-R-1000 . . . . .	42
5.9 Analysis of MAXD filter factor in segment distances of BCN-R-3000 . . . . .	42
5.10 Analysis of MAXDP filter factor in segment distances of BCN-1000 . . . . .	44
5.11 Analysis of MAXDP filter factor in segment distances of BCN-3000 . . . . .	44
5.12 Analysis of MAXDP filter factor in segment distances of BCN-R-1000 . . . . .	45
5.13 Analysis of MAXDP filter factor in segment distances of BCN-R-3000 . . . . .	45
5.14 Analysis of MAXD filter factor in trip distances from BCN-1000 . . . . .	48
5.15 Analysis of MAXD filter factor in trip distances from BCN-3000 . . . . .	48
5.16 Analysis of MAXD filter factor in trip distances from BCN-R-1000 . . . . .	49
5.17 Analysis of MAXD filter factor in trip distances from BCN-R-3000 . . . . .	49
5.18 Analysis of MAXDP filter factor in trip distances from BCN-1000 . . . . .	52
5.19 Analysis of MAXDP filter factor in trip distances from BCN-3000 . . . . .	52
5.20 Analysis of MAXDP filter factor in trip distances from BCN-R-1000 . . . . .	53

5.21 Analysis of MAXDP filter factor in trip distances from BCN-R-3000 . . . . .	53
5.22 Euclidean and shortest path distance ratio on the Barcelona area . . . . .	56
5.23 Euclidean and shortest path distance ratio on the Barcelona reduced area . . . . .	56
5.24 Comparison of mean traveltime variation throughout the day with PTV xRoute and HERE services in BCN area . . . . .	79
5.25 Comparison of mean traveltime variation throughout the day with PTV xRoute and HERE services in BCN-R area . . . . .	79

**Appendix**

D.1 Barcelona area coverage of Networkclass-0 with 1,000 trips . . . . .	96
D.2 Barcelona area coverage of Networkclass-1 with 1,000 trips . . . . .	96
D.3 Barcelona area coverage of Networkclass-2 with 1,000 trips . . . . .	96
D.4 Barcelona area coverage of Networkclass-3 with 1,000 trips . . . . .	96
D.5 Barcelona area coverage of Networkclass-4 with 1,000 trips . . . . .	96
D.6 Barcelona area coverage of Networkclass-5 with 1,000 trips . . . . .	96
D.7 Barcelona area coverage of Networkclass-6 with 1,000 trips . . . . .	97
D.8 Barcelona area coverage of Networkclass-7 with 1,000 trips . . . . .	97
D.9 Barcelona area coverage of Networkclass-0 with 3,000 trips . . . . .	97
D.10 Barcelona area coverage of Networkclass-1 with 3,000 trips . . . . .	97
D.11 Barcelona area coverage of Networkclass-2 with 3,000 trips . . . . .	97
D.12 Barcelona area coverage of Networkclass-3 with 3,000 trips . . . . .	97
D.13 Barcelona area coverage of Networkclass-4 with 3,000 trips . . . . .	98
D.14 Barcelona area coverage of Networkclass-5 with 3,000 trips . . . . .	98
D.15 Barcelona area coverage of Networkclass-6 with 3,000 trips . . . . .	98
D.16 Barcelona area coverage of Networkclass-7 with 3,000 trips . . . . .	98
D.17 Barcelona reduced area coverage of Networkclass-0 with 1,000 trips . . . . .	98
D.18 Barcelona reduced area coverage of Networkclass-1 with 1,000 trips . . . . .	98
D.19 Barcelona reduced area coverage of Networkclass-2 with 1,000 trips . . . . .	99
D.20 Barcelona reduced area coverage of Networkclass-3 with 1,000 trips . . . . .	99
D.21 Barcelona reduced area coverage of Networkclass-4 with 1,000 trips . . . . .	99
D.22 Barcelona reduced area coverage of Networkclass-5 with 1,000 trips . . . . .	99
D.23 Barcelona reduced area coverage of Networkclass-6 with 1,000 trips . . . . .	99
D.24 Barcelona reduced area coverage of Networkclass-7 with 1,000 trips . . . . .	99
D.25 Barcelona reduced area coverage of Networkclass-0 with 3,000 trips . . . . .	100
D.26 Barcelona reduced area coverage of Networkclass-1 with 3,000 trips . . . . .	100
D.27 Barcelona reduced area coverage of Networkclass-2 with 3,000 trips . . . . .	100
D.28 Barcelona reduced area coverage of Networkclass-3 with 3,000 trips . . . . .	100
D.29 Barcelona reduced area coverage of Networkclass-4 with 3,000 trips . . . . .	100
D.30 Barcelona reduced area coverage of Networkclass-5 with 3,000 trips . . . . .	100
D.31 Barcelona reduced area coverage of Networkclass-6 with 3,000 trips . . . . .	100
D.32 Barcelona reduced area coverage of Networkclass-7 with 3,000 trips . . . . .	100

E.1	Segment distance distr. and distance error distr. for BCN-1000 without filter	102
E.2	Segment distance distr. and distance error distr. for BCN-3000 without filter	102
E.3	Segment distance distr. and distance error distr. for BCN-1000 without filter	103
E.4	Segment distance distr. and distance error distr. for BCN-R-3000 without filter	103
E.5	Segment distance distr. and distance error distr. for BCN-1000 with MAXD-0.001 filter . . . . .	104
E.6	Segment distance distr. and distance error distr. for BCN-3000 with MAXD-0.001 filter . . . . .	104
E.7	Segment distance distr. and distance error distr. for BCN-R-1000 with MAXD-0.001 filter . . . . .	105
E.8	Segment distance distr. and distance error distr. for BCN-R-3000 with MAXD-0.001 filter . . . . .	105
E.9	Segment distance distr. and distance percentage error distr. for BCN-1000 with MAXDP-0.5 filter . . . . .	106
E.10	Segment distance distr. and distance percentage error distr. for BCN-3000 with MAXDP-0.5 filter . . . . .	106
E.11	Segment distance distr. and distance percentage error distr. for BCN-R-1000 with MAXDP-0.5 filter . . . . .	107
E.12	Segment distance distr. and distance percentage error distr. for BCN-R-3000 with MAXDP-0.5 filter . . . . .	107
E.13	KS test and trip distance distr. for BCN-1000 without filter . . . . .	108
E.14	KS test and trip distance distr. for BCN-3000 without filter . . . . .	108
E.15	KS test and trip distance distr. for BCN-R-1000 without filter . . . . .	109
E.16	KS test and trip distance distr. for BCN-R-3000 without filter . . . . .	109
E.17	KS test and trip distance distr. for BCN-1000 with MAXD-1 filter . . . . .	110
E.18	KS test and trip distance distr. for BCN-3000 with MAXD-1 filter . . . . .	110
E.19	KS test and trip distance distr. for BCN-R-1000 with MAXD-1 filter . . . . .	111
E.20	KS test and trip distance distr. for BCN-R-3000 with MAXD-1 filter . . . . .	111
E.21	KS test and trip distance distr. for BCN-1000 with MAXDP-0.1 filter . . . . .	112
E.22	KS test and trip distance distr. for BCN-3000 with MAXDP-0.1 filter . . . . .	112
E.23	KS test and trip distance distr. for BCN-R-1000 with MAXDP-0.1 filter . . . . .	113
E.24	KS test and trip distance distr. for BCN-R-3000 with MAXDP-0.1 filter . . . . .	113
E.25	KS test and trip distance distr. for BCN-1000 with EUCL-1.269 filter . . . . .	114
E.26	KS test and trip distance distr. for BCN-3000 with EUCL-1.269 filter . . . . .	114
E.27	KS test and trip distance distr. for BCN-R-1000 with EUCL-1.269 filter . . . . .	115
E.28	KS test and trip distance distr. for BCN-R-3000 with EUCL-1.269 filter . . . . .	115
E.29	KS test and trip distance distr. for BCN-1000 with DYN-EUCL filter . . . . .	116
E.30	KS test and trip distance distr. for BCN-3000 with DYN-EUCL filter . . . . .	116
E.31	KS test and trip distance distr. for BCN-R-1000 with DYN-EUCL filter . . . . .	117
E.32	KS test and trip distance distr. for BCN-R-3000 with DYN-EUCL filter . . . . .	117
F.1	Speed profiles variation with BCN-1000 sample . . . . .	118

F.2	Speed profiles variation with BCN-3000 sample . . . . .	118
F.3	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-1000 without filter . . . . .	119
F.4	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-1000 with MAXDP-0.005 filter . . . . .	119
F.5	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-3000 without filter . . . . .	120
F.6	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-3000 with MAXDP-0.005 filter . . . . .	120
F.7	Total traveltime error distr. for BCN-1000 and BCN-3000 with ST-SE model calibration without filter . . . . .	121
F.8	Total traveltime error distr. for BCN-1000 and BCN-3000 with ST-SE model calibration with MAXDP-0.005 filter . . . . .	121
F.9	Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration without filter . . . . .	122
F.10	Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration with MAXD-1 filter . . . . .	122
F.11	Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration with MAXDP-0.1 filter . . . . .	123
F.12	Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration with DYN-EUCL filter . . . . .	123
F.13	Speed profiles variation with BCN-R-1000 sample . . . . .	124
F.14	Speed profiles variation with BCN-R-3000 sample . . . . .	124
F.15	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-1000 without filter . . . . .	125
F.16	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-1000 with MAXDP-0.005 filter . . . . .	125
F.17	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-3000 without filter . . . . .	126
F.18	Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-3000 with MAXDP-0.005 filter . . . . .	126
F.19	Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with ST-SE model calibration without filter . . . . .	127
F.20	Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with ST-SE model calibration with MAXDP-0.005 filter . . . . .	127
F.21	Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE model calibration without filter . . . . .	128
F.22	Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE model calibration with MAXD-1 filter . . . . .	128
F.23	Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE model calibration with MAXDP-0.1 filter . . . . .	129

*LIST OF FIGURES*

---

F.24 Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE  
model calibration with DYN-EUCL filter . . . . . 129

# List of Tables

## Chapter 1: Introduction

1.1 Networkclass meanings . . . . .	5
1.2 Speedclass meanings . . . . .	5

## Chapter 5: Results

5.1 Networkclass coverage with 1,000 trips on Barcelona area . . . . .	32
5.2 Networkclass-Speedclass coverage with 1,000 trips on Barcelona area . . . . .	32
5.3 Networkclass coverage with 3,000 trips on Barcelona area . . . . .	33
5.4 Networkclass-Speedclass coverage with 3,000 trips on Barcelona area . . . . .	33
5.5 Networkclass coverage with 1,000 trips on Barcelona reduced area . . . . .	34
5.6 Networkclass-Speedclass coverage with 1,000 trips on Barcelona reduced area . . . . .	35
5.7 Networkclass coverage with 3,000 trips on Barcelona reduced area . . . . .	35
5.8 Networkclass-Speedclass coverage with 3,000 trips on Barcelona reduced area . . . . .	35
5.9 Summary of invalid requests for each sample . . . . .	36
5.10 Summary of Batches for each sample . . . . .	37
5.11 KS test statistics of segment distances without filter . . . . .	39
5.12 KS test statistics of segment distances with MAXD-0.001 filter . . . . .	40
5.13 Networkclass coverage after using with MAXD-0.001 filter over segment distances in BCN area . . . . .	40
5.14 Networkclass coverage after using with MAXD-0.001 filter over segment distances in BCN-R area . . . . .	40
5.15 KS test statistics of segment distances with MAXDP-0.005 filter . . . . .	43
5.16 Networkclass coverage after using with MAXDP-0.005 filter over segment distances in BCN area . . . . .	46
5.17 Networkclass coverage after using with MAXDP-0.005 filter over segment distances in BCN-R area . . . . .	46
5.18 KS test statistics of trip distances without filter . . . . .	47
5.19 KS test statistics of trip distances with MAXD-1 filter . . . . .	47
5.20 Networkclass coverage after using with MAXD-1 filter over trip distances in BCN area . . . . .	50
5.21 Networkclass coverage after using with MAXD-1 filter over trip distances in BCN-R area . . . . .	50
5.22 KS test statistics of trip distances with MAXDP-0.1 filter . . . . .	51
5.23 Networkclass coverage after using with MAXDP-0.1 filter over trip distances in BCN area . . . . .	51
5.24 Networkclass coverage after using with MAXDP-0.1 filter over trip distances in BCN-R area . . . . .	54

5.25	KS test statistics of trip distances with EUCL-1.269 filter . . . . .	54
5.26	Networkclass coverage after using with EUCL-1.269 filter over trip distances in BCN area . . . . .	55
5.27	Networkclass coverage after using with EUCL-1.269 filter over trip distances in BCN-R area . . . . .	55
5.28	KS test statistics of trip distances with DYN-EUCL filter . . . . .	57
5.29	Networkclass coverage after using with DYN-EUCL filter over trip distances in BCN area . . . . .	57
5.30	Networkclass coverage after using with DYN-EUCL filter over trip distances in BCN-R area . . . . .	58
5.31	Design of experiments considered . . . . .	60
5.32	Initial speed profiles per networkclass in Barcelona . . . . .	61
5.33	Initial MAE and MAXAE with default speed profiles in BCN area . . . . .	61
5.34	MAE iterative results in BCN area . . . . .	64
5.35	MAXAE iterative results in BCN area . . . . .	65
5.36	Speed profiles per networkclass computed from BCN-1000 sample . . . . .	66
5.37	Speed profiles per networkclass computed from BCN-3000 sample . . . . .	67
5.38	Initial MAE and MAXAE with default speed profiles in BCN-R area . . . . .	68
5.39	MAE iterative results in BCN-R area . . . . .	71
5.40	MAXAE iterative results in BCN-R area . . . . .	72
5.41	Speed profiles per networkclass computed from BCN-R-1000 sample . . . . .	73
5.42	Speed profiles per networkclass computed from BCN-R-3000 sample . . . . .	74
5.43	Time performance with TT models depending on the sample size . . . . .	76
5.44	Time performance with ST models depending on the sample size . . . . .	77
5.45	Speedup per procedure depending on the sample size and model . . . . .	78
5.46	MAE comparison before and after calibration at different departure hours in BCN area . . . . .	80
5.47	MAE comparison before and after calibration at different departure hours in BCN-R area . . . . .	80

**Appendix**

B.1	Default speeds distribution per networkclass example in Barcelona . . . . .	93
B.2	Default speeds table example in Barcelona . . . . .	93
C.1	Estimated $\beta$ values for the euclidean distance metric model . . . . .	95

# List of Codes

## Chapter 3: System specification

3.1	Estimator: Mean Absolute Error (MAE) . . . . .	19
3.2	Maximum absolute distance error filter (MAXD) . . . . .	20
3.3	Maximum absolute distance percentage error filter (MAXDP) . . . . .	20
3.4	Euclidean distance filter (EUCL) . . . . .	21
3.5	Dynamic euclidean distance filter factor (DYN-EUCL) . . . . .	21
3.6	Model: Segment traveltime optimisation by segment error (ST_SE) . . . . .	23
3.7	Model: Trip traveltime optimisation by segment error (TT_SE) . . . . .	24

## Chapter 5: Results

5.1	KS statistic . . . . .	38
5.2	Speedup formula . . . . .	78

## Appendix

A.1	PTV xRoute API - Calculate Route request example . . . . .	87
A.2	PTV xRoute API - Calculate Route response example . . . . .	88
A.3	HERE Routing API - Calculate Route request example . . . . .	90
A.4	HERE Routing API - Calculate Route response example . . . . .	90
C.1	Distance metric generalisation . . . . .	94
C.2	Inflated distance metric generalisation . . . . .	94
C.3	Inflated euclidean distance metric . . . . .	95



# Chapter 1

## Introduction

### 1.1 Introduction

This project is done as a Final Master Thesis (TFM) of the Master in Innovation and Research in Informatics (MIRI) specialised in Data Science, at the Barcelona School of Informatics (FIB) of the Polytechnic University of Catalonia (UPC).

This project is being carried out in collaboration with PTV Group, a German company specialised in software solutions and consulting services for logistics, mobility and traffic and transportation. Its headquarters are located in Karlsruhe (Germany), but the project is being developed from its office in Barcelona (Spain).

### 1.2 Contextualisation

This project is developed as part of the tasks in the PTV MaaS Operator department at PTV Group. In this department we deal with projects that improve urban mobility through the MaaS (Mobility as a Service).

MaaS is a new concept of mobility services that connects users with different forms of transport services, public and private modes, into a single service for their use. It is an on-demand real-time platform that combines different multimodal transport providers (e.g. bus, train, taxi or car/bike sharing) into a single account that manages their trip planning along with the payments.



Figure 1.1: MaaS concept<sup>1</sup>.

<sup>1</sup>Figure property of PTV Group.

This concept originates as a solution to reduce traffic congestion and its environmental impact and also to increase the network efficiency in the smart cities. It proposes an alternative to the use of private car's ownership, where the conventional use of the private car is 1 passenger (1.2 in average in the case of many cities) travelling from home to work and vice-versa, increasing in that way the number of trips with the consequent impacts on the state of the network (contribution to congestion) and the environment.

The solution proposed by MaaS is then to make users leave aside their personal vehicles and move into an alternative and innovative mobility while still satisfying their main objectives of freedom (travelling from A to B whenever I need) in the most convenient and efficient way. Where within a unique digital platform, users will have access to a large variety of transport providers for choosing the most suitable end-to-end trip depending on their personal preferences (e.g. traveltime or cost)[1].

Consequently, this decrease of the number of private vehicles, making users use shared means of transportation, would cause a series of benefits such as reducing congestion in cities and therefore the emissions to the environment[2] improving life in the cities.

This concept is developed in the PTV MaaS Accelerator Program[3] consisting of four modules that provide solutions at different levels of detail to bring shared mobility services into smart cities:

1. **PTV MaaS Modeller**

Calculates relevant KPIs to create an effective business model for MaaS fleets within a multimodal transport city system.

2. **PTV MaaS Simulator**

Provides a detailed simulation of different mobility scenarios considering connected vehicles as well as public transport, pedestrians and bikes.

3. **PTV MaaS Operator**

Is the core component of the real-time multimodal MaaS fleet that coordinates and optimises the trips as well as the vehicles occupancy and the system efficiency.

4. **PTV MaaS Controller**

Integrates all of the components into a cloud-base service to build an overall operating system with real-time control of connected vehicles and traffic prediction and route optimisation.

The PTV MaaS Modeller, the lowest-level component, to model the entire multimodal traffic system, integrates two other products from the PTV Group company: PTV Visum[4], a macroscopic travel-demand modelling simulating software, and PTV xServer[5], a set of geographical and logistical software components to optimise logistic processes.

This PTV xServer software offers a large variety of services depending on the problem to tackle. Two versions are available according to the client's needs [6]: the desktop version

where the modules can be integrated to any software through interfaces, or the internet version where all the modules are hosted on a Windows Azure server that can be accessed through a Web API with no need for the client to have its own server.

In particular, PTV xRoute is responsible for calculating routes between two locations. For this reason, this service uses PTV Maps&Data [7] software that provides digital maps with a high level of detail that are regularly updated using the leading map providers HERE and TomTom.

This service is also used by PTV Map&Guide[8], a route planning software that computes routes suitable for trucks and HGVs (Heavy Goods Vehicles) taking into account the truck-specific restrictions (e.g. weight and height limits) and computing the exact transportation and toll costs. An example can be seen in Figure 1.2 showing the results for a route from Ljubljana (Slovenia) to Amsterdam (Netherlands) and its associated costs.

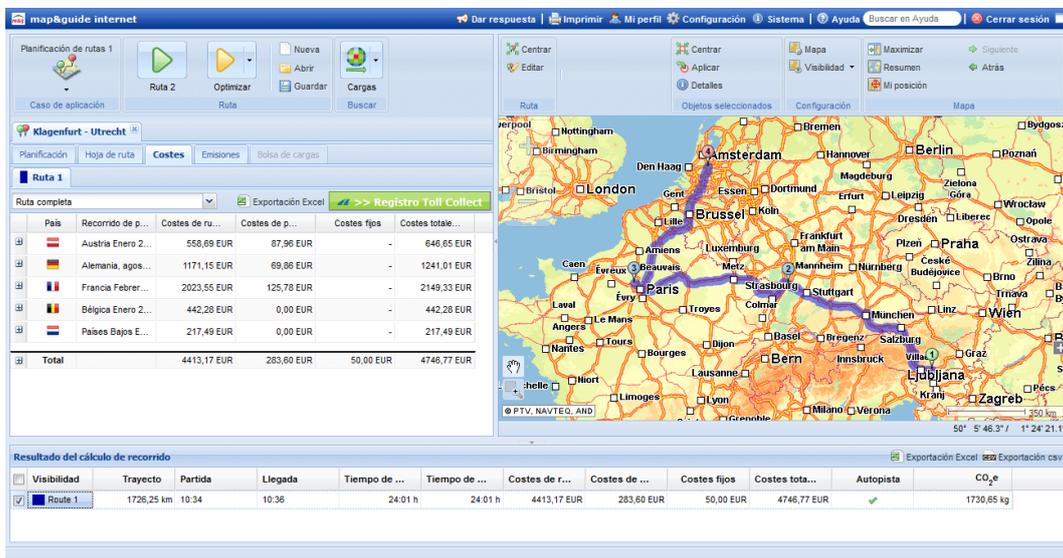


Figure 1.2: PTV Map&Guide<sup>2</sup>.

Until now, the normal use came from transportation companies that compute long-distance routes that could cross through several countries. In this scenario, a small deviation on traveltime does not affect the final goal. However, the usage of the PTV MaaS Operator is meant for a real-time service for urban routes within a city that need to be more accurate. And a deviation on traveltime can influence a lot the routes of the MaaS fleets and give an inefficient solution wasting the available resources.

Therefore, the PTV MaaS Operator system requires a calibrated PTV xRoute service as close to reality as possible. Thus, this idea will constitute the main objective of this project.

<sup>2</sup>Figure property of PTV Group.

## 1.3 Problem formulation

PTV Group offers a large variety of trip planning, tour optimisation and traffic management and simulation tools. All using PTV Maps&Data service that provides digital maps with a high level of detail and regularly updated with external leading map providers (e.g. TomTom, HERE).

Different information is also collected from these suppliers in order to determine the information related to the segments of the network. In particular there is the speed parameter.

This speed parameter is really important since it provides some of the information required to calculate the shortest paths within the network and the associated traveltime. It can be either static or dynamic:

- The static speed is derived from time independent data from these providers. Logically, it does not cover all possible fluctuations throughout the day and only covers part of the real situation in urban areas.
- The dynamic speed on the contrary takes into account the instant of time. This is also derived from standard suppliers, although it can be enriched with specific logistics data such as the base speeds obtained from FCD (Floating Car Data) for the estimation of speed patterns in the network segments. As a result, this kind of speeds approximate better the real situation.

For the general usage only static speeds are available. And only in special cases the dynamic speeds are used because they require a more complex and specific processing of the data.

This static speeds parameter is enough for the normal usage of PTV Group clients where a small deviation in the traveltime calculation does not affect the final goal. But not for real-time products where this calculation has to be very accurate.

Alternatively, these traveltimes can be further influenced with the so called vehicle profiles that define the characteristics of the vehicle used and some states of the network, such as the speed settings for different networkclasses of the network.

These networkclasses are a hierarchy of the network segments of PTV Group products that classifies them depending on their importance. And at the same time, each networkclass is subclassified into speedclasses, a classification depending on the speed restrictions (see Tables 1.1 and 1.2 for clarification).

By changing the speed settings associated with this hierarchy of segments in the network, the shortest path search and its associated traveltime can be influenced and can be used as a calibration capability.

But even though it is possible to modify these speeds by oneself, a calibration process of this parameter is required for calculating speeds to offer trips more accurate accordingly to the

real situation of the network. Not only because it is required in the PTV project MaaS, but also to offer a better service to PTV clients as a general usage.

NC	MEANING
0	Motorway
1	Highway
2	Trunk road
3	Country road
4	City street
5	Residential street
6	Drivable but normally blocked roads
7	Cycle lanes and walkways

Table 1.1: Networkclass meanings.

SC	MEANING
0	Fastest
1	Faster
2	Fast
3	Medium fast
4	Medium slow
5	Slow
6	Slower
7	Slowest

Table 1.2: Speedclass meanings.

It should also be emphasised that these default speed profiles are only a good estimation for long distance routes. And that such procedure would not work for large areas such as countries. Because local situations (driving habits, road quality, etc.) are completely different everywhere. And therefore, it will only be effective for more targeted areas.

### 1.3.1 Motivation

The main motivation for the realisation of this project is to extend the PTV xRoute service, implementing an automatic calibration method for the speed profiles. Before that, the normal usage of this service by PTV customers did not require a high accuracy in the calculation of traveltimes. The vast majority of its customers are transportation companies that only require to plan their daily long-routes where differences of a few minutes on long routes is a negligible error.

But with the PTV MaaS project these small deviations are important if working in urban areas. Where the service should offer routes as realistic as possible with the minimum error in the calculation of these traveltimes. Not only to improve the usage and efficiency of the fleet but also for the satisfaction of the end user who makes use of it.

Another motivation is because right now the assignment of some default values of the speed settings for each area is done manually. Now an expert analyses the situation, at that precised moment, to calculate some values that reflects the reality.

That is why it is required to automate this manual process, not only to facilitate this work, but also to be able to update these profiles more frequently. And thus better reflect the variation in the state of any urban network as it changes completely at every instant of the day (as seen in Figure 1.3): at peak hours the mean traveltime increases, thus decreasing

the speed of vehicles. In contrast, in the rest of the hours this time decreases, increasing the average speed.

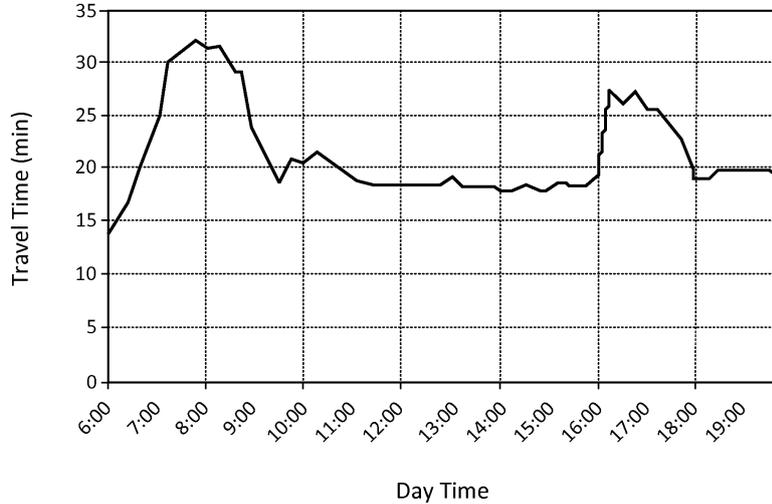


Figure 1.3: Variation of mean traveltime depending on the time slot<sup>3</sup>.

In this manner, if it is desired to replicate this temporal variability throughout the day, this procedure should be executed every certain interval of hours for a correct calibration for such moments. Therefore this system should not only be automatisable, but also efficient so that it can be recomputed at different time slots.

### 1.3.2 Main objectives

This project aims to improve the behaviour of the PTV xRoute service calibrating one of its vehicle profiles parameter, the speed settings for different networkclasses of the network (a hierarchy of the segments importance).

From a set of input data, this work will measure the differences between such traveltimes (from external traffic data providers) and traveltimes with given vehicle profile speed settings (from PTV xRoute service). Using these traveltimes as a reference, different optimisation methods of the speed profiles will be proposed in order to minimise the deviation in the traveltimes.

As it may be necessary in the future, the calibration process should be agnostic towards the external reference data service. In this way, it would be possible to change this provider without causing major changes in the calibration process.

Furthermore, the calibration of some default suitable settings that characterises the state of the network is, as of now, performed manually by an expert analyst. Moreover, these speed settings are mean values used for all day. And logically, using the same speeds at all instants in a day is useless (as observed before in Figure 1.3).

<sup>3</sup>Figure extracted from <https://patents.google.com/patent/US20140046593>.

Thus, this work will evaluate mechanisms to automatically calibrate the speed settings appropriately that could be re-computed at different time slots.

## 1.4 Scope of the project

For fulfilling the previously mentioned objectives, the system will need to satisfy the following requirements:

- **Study different areas**

It will allow to study the speed profiles in different areas. For now it will be tested with the Barcelona area, but should be extendable to other areas in the future.

- **Optimise the speed settings**

It will optimise the speed settings of PTV vehicles profiles used for routing to minimise the traveltime differences compared to historical input data.

- **Evaluate the quality of the new solution**

It will be estimated the performance of the system in terms of time execution and quality of the solution using different KPIs.

- **Automate the process**

It will allow to automate the process, currently manual, in order to compute speed profiles within different areas and time slots.

### 1.4.1 Possible obstacles

During the process of developing a project can appear some obstacles that delay the planning and may prevent the fulfilment of the requirements.

The first obstacle that could appear could be that the network of the external traffic data provider does not match in its majority with the PTV xRoute network. And in this case, could not be used as reference data for calibration.

The other obstacle that could be encountered would be the overall performance of the whole process. The calibration system requires to gather data from two web services with the consequent performance penalty for each request.

Since this is a data analysis and optimisation project, a large amount of data is required which can cause the entire calibration process computationally costly. Nevertheless, in transport planning tools a few hours are acceptable on a weekly or monthly basis.

## 1.5 Master thesis outline

Considering the previously formulated problem and its main objective, the document of this project is organised as follows:

- **Chapter 2: State of art** analyses and reviews the current literature related to our problem formulation.
- **Chapter 3: System specification** describes the data used in the system and the algorithm applied. Along with the definition of the proposed models for computing the speed profiles.
- **Chapter 4: System architecture** describes the architecture decided for the implementation of the system.
- **Chapter 5: Results** presents the results obtained for each studied area. As well as an analysis of the performance of the system in terms of quality and efficiency.
- **Chapter 6: Conclusions** reviews the achievement of the objectives. Along with the specification of the contributions of the realisation of this project and further researchs.

# Chapter 2

## State of art

### 2.1 Link traveltime estimation problem

The main objective of this system is to know how to calculate segment speeds of a road network depending on a certain hierarchy of their importance. This problem can be therefore expressed as a link traveltime estimation method (where the link is expressed as a segment in PTV xRoute). By estimating these link traveltimes, then the link speeds can be also estimated as this system requires.

This problem can be solved in different ways, but in either case it requires reference data of the state of the network for a proper calibration of the links. The most common data source is Floating Car Data (FCD). But there is also another possibility such as historical data from routing providers. Both kinds of data are described below with their advantages and disadvantages and why it is decided to use historical data as the reference data.

#### 2.1.1 Floating Car Data (FCD)

The Floating Car Data (FCD) also known as Probe Vehicle Data (PVD) is data generated by a fleet of vehicles obtained from sensors located inside the vehicles such as mobile phones or traditional GPS. This data contains the vehicle position and the corresponding timestamp with any specific vehicle state related data like the instant speed [9].

With this raw data, a map-matching process is carried out to associate these raw time indexed positions with those of the road network to be estimated [9]. This allows to associate the data collected to each link in the network and thus estimate the traveltime in the links (see Figure 2.1).

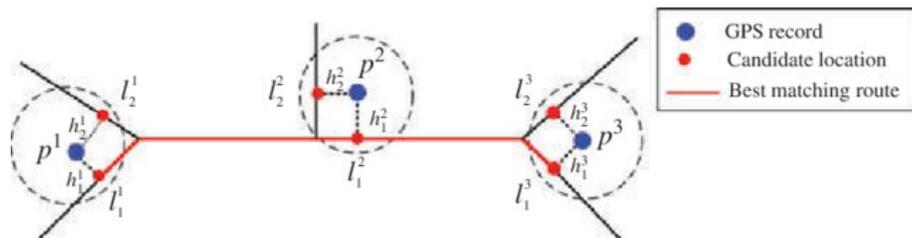


Figure 2.1: Map-matching example [10].

This kind of data allows to obtain the information at a detailed level of the network and with a large fleet of vehicles, to obtain a sufficiently large sample to cover an urban region over long periods of time [11]. And in particular, it allows to obtain a detailed information about the route choice of the vehicles over the same origin-destination routes.

The most common data sources for FCD in urban areas come from taxi or delivery companies as these have a large fleet of vehicles moving around the urban network throughout most of the day [9]. And therefore, generating a huge amount of data from which a rich knowledge about the state of the network can be extracted at different time slots.

But it is necessary to emphasise that these are only a sample of the state of the network since data is only obtained for those links through which the vehicle fleet traverses [11].

And the most significant challenge of this data source, which is not in the focus of this work, is the map-matching algorithm since it is the one that matches the network status data with the road network. Therefore, it must be as accurate as possible to avoid using as reference data those corresponding to another link.

In addition, the transmission of vehicle information is not constant taking place every few minutes (normally each 2-3 minutes) as it is not the prime requirement for taxi companies to have the location of their fleet at such a detailed level [9].

Involving the application of some algorithms in order to predict the possible routes that vehicles may have followed through the known links.

### 2.1.2 Historical traffic data

Historical traffic data is another data source to know the state of the network for any instant of time. This data is obtained from routing providers such as GoogleMaps, HERE and TomTom who offer route calculation services and estimated trip traveltimes through their APIs (Application Program Interface).

In this case, this data is already processed by the routing service providers and is not raw data such as FCD. This way these providers are able to estimate the travel time in the links of their internal networks and thus being able to calculate routes as reliable as possible.

Within large-scale service providers, these estimations are generated by integrating different data sources [12] such as FCD or real-time traffic informations from traffic sensors (e.g. inductive loops, traffic cameras...).

But also using real-time information obtained through the users who use their navigation applications (e.g. GoogleMaps, HERE WeGo...) as anonymous data [12]. In the case of Google, obtaining data through Android mobile phones or through its GoogleMaps and Waze applications.

In this case, in order to obtain the reference data of these routing services, it is only neces-

sary to make multiple requests to the API over multiple origin-destination pairs of routes. Simplifying in this way all data processing since the data obtained from these services are already the estimations of the state of the network required for our system.

Also avoiding the map-matching procedure since these services already have a very sophisticated one implemented for matching the coordinates that users use in their services.

## 2.2 Conclusions

After describing both possible data sources, FCD or historical traffic data, as our reference data, different conclusions can be drawn.

The main reason is the simplicity of the data collection. If this procedure is simplified, more emphasis can be placed on system calibration methods.

In the case of using FCD, several processes would be required in order to obtain a reference sample. First a map-matching process to match the raw coordinates to those of the road network. And then applying algorithms to predict the routes that the vehicles would have used since the vehicles only send information every few minutes and the samples that are collected may not be from contiguous links.

On the other hand with the historical traffic data, the data would already be processed and we would have the samples with the estimations of the traveltimes on each link required for our system. In addition, the map-matching of the locations is carried out by these suppliers.

And secondly, because these routing service providers have been offering these services for years and are very reliable when predicting traveltimes for multiple routes. Moreover, it has a very wide coverage for any area that one wants to analyse and not just for the links through which the fleet of vehicles has crossed.



# Chapter 3

## System specification

### 3.1 Data description

This section describes the data that will be used to calibrate the speeds profile parameter of the PTV xRoute service. First, it will be specified the PTV xRoute and HERE services and at the end the problematic issues of the integration of both networks.

#### 3.1.1 PTV xRoute service

PTV xRoute is a service that creates routes between two locations considering toll costs, emissions and traffic situation allowing customers to customise the properties of the vehicle used: car or truck, sizes, weight, ...

It has several resources (e.g. calculation of routes, calculation of reachable areas...). But the system for calibrating this service, object of this thesis, will be based only on the use of the resource of calculating routes since it is the only one that fits with our objective of calibrating routes for a series of origins and destinations.

PTV xRoute response result is based on a composition of the hierarchy of geographical elements depicted in Figure 3.1.

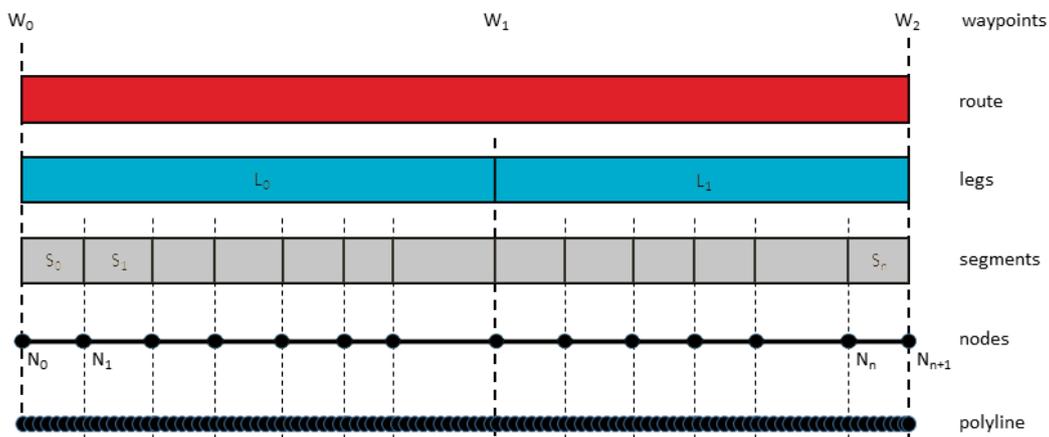


Figure 3.1: PTV xRoute response result lists

Thus, a *Route* between a sequence of *Waypoints* (origin, destination and a series of intermediate points) is a set of *Legs* that splits that route between two consecutive waypoints in order to have its associated aggregate information. Each *Leg* is also a sequence of *Segments* that connects different *Nodes*. To compare it with a real urban network, each *Segment* would be like a section of a street and the *Nodes* the connection points between sections with the turns information.

This system will focus mainly on the terms *Segments* and *Nodes* since both contains the information that is required for the calibration of the speeds for the different networkclasses.

In the Appendix A.1 it is shown an example of a request for a given origin-destination, enabling the corresponding parameters so that the PTV xRoute service answers with the detailed information that is required for our system (nodes and segments with descriptors). And a subset of the response is also shown.

Within the terminology of nodes there are three types: the input node, the link node and the network node. They all contain the same information (coordinates), but have a different meaning:

- *Input node* (also called *Waypoints*): the node at the exact coordinates of the origin and destinations of the requests.
- *Link node*: the node that connects the input nodes and the network nodes.
- *Network node*: the actual nodes in the network.

And within the terminology of the Segments there are two types: the link segments and the network segments:

- *Link segment*: the segment that connects the input nodes with the link nodes. These does not have any information relative to a networkclass as it is an auxiliary segment for connecting the input origin-destination to the network.
- *Network segment*: the segment that actually connects the nodes inside the network (link and network nodes). These do contain the required detailed information relative to the networkclass.

But with this information only a part of the required data for calibration process is obtained (coordinates and distances of the segments and their associated networkclass). In order to correctly calibrate the service depending the networkclass, it is also necessary to know its associated speedclass. In this case, this information is not provided by any PTV xServer service as customers do not usually use it. Therefore, it must be obtained with an alternative process specified below.

The PTV traffic network categorises the segments into different networkclasses and speedclasses, and then given some speed profiles (speed ranges within each networkclass) calculates

the speed associated with each segment. This mapping is a linear interpolation between the min-max speed range of that networkclass.

In Figure 3.2 it can be seen this mapping between the speeds and the speedclasses for one networkclass, where  $N_{min}$  and  $N_{max}$  are the minimum and maximum speeds and *maximumSpeed* is the maximum speed allowed.

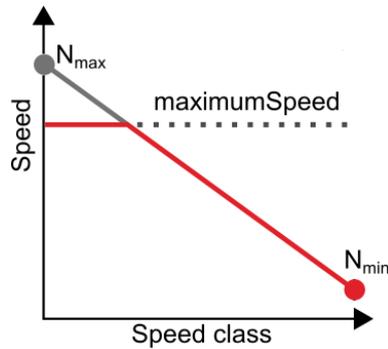


Figure 3.2: Speedclass interpolation and max speed for one networkclass

In our case, as the calculated speed is known, then it will be applied an inverse process to obtain the speedclass to which the segment is associated. For this purpose, we will use a table with the distribution of speeds on each networkclass for each area to analyse (for a description of this process see the Appendix B: Speedclass computation).

### 3.1.2 HERE Routing service

To complement the data required to calibrate the PTV xRoute service with historical data, a suitable service is the HERE Routing one that also includes a route calculation resource. Since PTV Group works with their services, its data is easily available for our calibration system.

The HERE response is very similar to PTV xRoute's providing the information related to the *Waypoints*, *Legs* and *Nodes*, but not about the *Segments*.

In the Appendix A.2 it is shown an example of a request using the HERE's service to calculate a route for a series of waypoints obtained from the route computed by PTV xRoute.

This service allows us to know both the total traveltime and the traveltime for each segment between the waypoints that have been indicated. That is the information required for our calibration system.

### 3.1.3 Problematic issue: integration of both services

The main problematic found in this project is that both services use two different traffic networks and therefore the coordinates are not exactly the same.

This can lead to the following cases when using the coordinates provided by the PTV xRoute service in the HERE network to calculate the reference KPIs:

- If the HERE network does not have any segment in those coordinates, it will try to match the nearest segments (within a maximum distance value). Thus, resulting in routes very different from those required.
- If existing, but due to small differences in their coordinate system, the nearest HERE segment may be on the opposite direction or on the side street (see Figure 3.3 as an example).

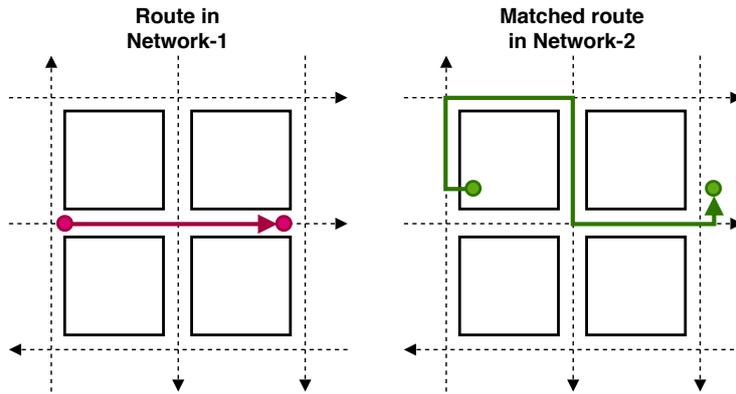


Figure 3.3: Matching routes from two networks

In this case, if there are a lot of intermediate points to match in the HERE network, the resulting route can make many detours in circles to pass through each of those points. Thus, resulting also in routes that are very different from those required.

In both cases, the routes obtained are not as the required for using as reference data along with the PTV xRoute ones. And therefore, both cases will be taken into account when analysing the reference data that will be used for the calibration system, invalidating them if needed.

As a method of detection of these erroneous cases, one possible approach could be through the analysis of distances. Excluding those segments or routes with a distance deviation greater than a certain threshold value. As described in Section 3.3 with the description of the considered procedures to solve this problem.

## 3.2 Data pipeline

In this section it will be explained the data pipeline that has been considered for the system.

The aim of this system is to calibrate the parameter of the speed profiles of the PTV xRoute service. This means that if this parameter varies, so can the routes computed by the service,

as well as the associated KPIs (total traveltime and distance). Therefore, it is decided to apply an iterative algorithm to determine these speed profiles..

In this way, the calculation of the speed profiles will be iterated until these speed profiles and associated KPIs are maintained minimum and stable. In the figure below it can be seen an overview of the iterative data pipeline that has been implemented.



Figure 3.4: Iterative data pipeline

### 3.2.1 Data generation

An important process of this system is the generation of data from both data sources and their integration.

The first step is to obtain an initial sample of the trips to be used for the speed calibration process. Generating a set of origins and destinations.

Using the available origin-destination routes in the area, the sample is then generated from the two data sources:

- **PTV xRoute**: computes the trips information for the set of routes. Computing the set of segments and the total information of the trips.
  - Segment-level information: distance, traveltime, speed, networkclass and speed-class
  - Trip-level information: distance and traveltime
- **HERE**: computes the reference data for the trips.
 

In the case of a segment-based model (as will be described in Section 3.4), it will also be required to calculate the reference traveltime for each of the trip segments.

  - Segment-level information: distance and traveltime (if required)
  - Trip-level information: distance and traveltime

### 3.2.2 Data preparation

The next step is preparing the raw data sample from the integration of both providers to obtain an appropriate sample for the speed calibration process:

1. Remove invalid trips: It may be the case that any of the two services was unable to find a route for some of the trips and therefore must be treated as missing data.

2. Compute new attributes: from the existing information, new attributes can be computed at a different level of detail.
  - Segment-level: speed, error in traveltime and distance (comparing both services) and euclidean distance between the initial and destination nodes of the segment.
  - Trip-level: error in traveltime and distance (comparing both services) and euclidean distance between the origin and destination nodes of the trip.

Since this system is working with two routing services with a large coverage in multiple areas, there should not be a high proportion of the sample as invalid. However, as a result of this procedure an analysis of the missing data will be carried out to find a plausible explanation to try to fix it if possible.

In addition, the quality of the sample will be analysed in terms of total coverage within the area and the distribution of segments within the networkclasses and segmentclasses to be optimised.

#### 3.2.3 Data refinement

This component is responsible for removing reference information from erroneously collected data. For using only a reliable sample for the speed calibration procedure.

With the already prepared data a cleaning process is carried out in order to discard erroneous data in the sample. For the detection of these integration errors between both sources, the distances between segments or trips are available.

The first step is to analyse the distance distributions in order to understand the differences between both services. To know if the routes computed by the HERE external service resemble those of the PTV xRoute in any way. And therefore, if the associated traveltimes can be used as a reference.

Afterwards, the portion of the sample with a large discrepancy in distances will be detected and removed from the dataset for the calibration procedure.

For this purpose, several methods for detecting these differences will be considered. So that their influence on future calibration can be studied. Such methods are described in Section 3.3: Filtering methods.

#### 3.2.4 Speed profiles computation

This component is responsible for calculating the speed profiles according to the data sample on the analysed area.

The objective of this system is to calibrate the speed profiles of PTV xRoute service in order to reproduce the real traffic state. Thus, for this calibration it is considered to use HERE's

traveltimes as the reference information for the traffic conditions, reducing the existing differences between both services.

As a method for minimising these deviations, it is decided to use optimisation models. Since this problem has an objective function to minimise and several constraints that must satisfy the speed profiles for giving the desired shape to the speeds table (the speeds are not individual, but related to those of the same networkclass. As described in Appendix B).

Therefore, several optimisation models have been considered depending on the level of detail of the reference data (total traveltime or segment traveltime). Such models are described in Section 3.4: Proposed optimisation models.

### 3.2.5 Data generation (calibrated)

As a final step, the data is regenerated using the optimised speed settings computed before in order to measure the calibration improvement.

Firstly the sample is generated again with the PTV xRoute service:

- **PTV xRoute**: computes the calibrated information as in Step 1.

Then the calibration results are analysed in order to compare if the calibrated speeds has improved or not the situation.

The main measure in which we will focus will be the Mean Absolute Error (MAE) in traveltime as it indicates the mean deviation in traveltime between both services (PTV xRoute and HERE). Being the closest to 0 the ideal.

$$MAE = \frac{\sum_{i=1}^N |xroute\_traveltime_i - here\_traveltime_i|}{N}$$

Code 3.1: Estimator: Mean Absolute Error (MAE).

But it is also convenient to take into account the Maximum Absolute Error (MAXAE) as it indicates the maximum deviation in the traveltime between both services.

These measures allow to know if the calibration has reached a stable point, or still varies slightly between iterations. Irrespective of the result obtained (minimised or not), if it is not stable in the MAE estimator, a further iteration will be executed. Otherwise, if the procedure converges its execution finishes.

If at least one more iteration is followed, it is required to generate the sample again with the external data provider for the routes that may have changed with the calibration process:

- **HERE**: computes the reference information as in Step 1.

### 3.3 Filtering methods

As filtering methods for discarding the part of the data with large deviations between both data sources it has been decided to consider the following four filters.

As mentioned before, depending on the level of detail of the sample this filtering will be applied to the distances of the segments for the samples at segment-level or to the total trip distances for the samples at trip-level.

#### 3.3.1 Maximum absolute distance error filter (MAXD)

As a first filtering method it is tried a filter over the deviations between both distances. Eliminating the segments or the trips with a distance deviation greater than a threshold.

The filtering rule that is applied is as follows. Where the segments / trips with a distance difference greater than a factor *MAX\_DIST\_ERROR* are eliminated.

$$\text{abs}(xRoute\_distance - HERE\_distance) \leq MAX\_DIST\_ERROR$$

Code 3.2: Maximum absolute distance error filter (MAXD).

This filter is a bit simple but using a low value as a threshold allows to preserve data with a small deviation.

However, the disadvantage of this method is that the sample contains both long and short routes. And in this case, the same filter is being applied which does not take this variation into account.

#### 3.3.2 Maximum absolute distance percentage error filter (MAXDP)

As an attempt of correcting the previous filter, it is decided to consider a filtering on the percentages of error in the deviations of the distances. Eliminating segments or trips with an error percentage higher than a threshold.

The filtering rule that is applied is as follows. Where the segments / trips with a distance difference percentage greater than a factor *MAX\_DIST\_PERC\_ERROR* are eliminated.

$$\text{abs}(xRoute\_distance - HERE\_distance) / xRoute\_distance \leq MAX\_DIST\_PERC\_ERROR$$

Code 3.3: Maximum absolute distance percentage error filter (MAXDP).

This filter is still simple, but it tries to fix the problem with samples with high variability in the length of the routes.

In this case, using the same percentage threshold allows a small deviation for short routes and a slightly larger deviation for long routes.

### 3.3.3 Euclidean distance filter (EUCL)

The problem with the previous filters is that the factors are specific to each area and an analysis of these will be required in order to select feasible values for each area to be treated. For this reason, it is decided to consider a new type of filtration commonly used in the logistics environment based on Euclidean distances.

The filtering rule that is applied is as follows. Where the segments / trips with a distance greater than a factor *EUCLIDEAN\_FACT* multiplied by the euclidean distance are eliminated.

$$\begin{aligned} xRoute\_distance &\leq EUCLIDEAN\_FACT * euclidean\_distance \\ HERE\_distance &\leq EUCLIDEAN\_FACT * euclidean\_distance \end{aligned}$$

Code 3.4: Euclidean distance filter (EUCL).

In this case, the Euclidean distance has been studied for a long time and studies have been made to analyse the existing relationship between the shortest path and the euclidean distance. Showing that this factor is associated with the topology of the area with a unique value. See a description in Appendix C and to see some of the known values.

However, this filter only makes sense at trip level and not at segment level. Since the Euclidean distance is the value of the straight line between two points and the euclidean factors determine how large any possible distance between those two points can be compared to the diagonal. In the case of a segment, there are no different possible routes.

### 3.3.4 Dynamic euclidean distance filter (DYN-EUCL)

To improve the previous filter, being that it is not always possible to access to previous studies on the value of the Euclidean distance factors, it is decided to consider a new type of filter over the Euclidean distances but with a dynamic factor calculated by our system.

Therefore this filter will apply the same rule as in Code 3.4, but using a dynamically calculated factor for each analysed area as shown in the following formula:

$$DYNAMIC\_EUCLIDEAN\_FACT = \frac{\sum_{i=1}^N \frac{euclidean\_distance_i}{shortestpath\_distance_i}}{N}$$

Code 3.5: Dynamic euclidean distance filter factor (DYN-EUCL).

From a sufficiently large sample, the euclidean distance factor can be approximated as the average value of the ratio between both distances. For a more detailed description of this process see the Appendix C.

For the above-mentioned reason, this filter only makes sense at trip level and not at segment level.

## 3.4 Proposed optimisation models

For the calculation of the speed profiles it has been decided to apply optimisation models to minimise the deviation of the speeds with the historical reference traveltimes. For this purpose, different models have been implemented based mainly on the following two principles:

- **Segment traveltime optimisation models (ST)**

Learning from the segment traveltime computed from external providers, these models compute speed profiles that fit to the segment traveltimes.

In this case, it is required to have the reference traveltime for each segment.

- **Trip traveltime optimisation models (TT)**

Learning from the trip total traveltimes computed from external providers, these models compute speed profiles that fit to the trip traveltimes.

In this case, it is only required to have the total reference traveltime and not for each segment.

For both cases, the same conditions have been imposed to give the speed profiles format used internally in the PTV xRoute service. As explained previously in the Tables 1.1 and 1.2 those are (see Figure 3.2 or also an example of the speed profiles in the Appendix B for clarification):

- Within the same networkclass, speedclass- $i$  is faster than speedclass- $(i+1)$ .
- The relation within the same networkclass is linear. That is, the difference between two consecutive speedclasses is equal.

Between different networkclasses, there is no relationship among the associated speeds. The networkclass- $i$  does not necessarily have to be faster than the networkclass- $(i+1)$ , instead these speeds generally overlap (see a sample of the default speeds in Appendix B).

Therefore, no restriction will be introduced between speeds of different networkclasses. Allowing a later analysis to study the relationship between the networkclass speeds.

### 3.4.1 Segment traveltime optimisation models (ST)

Only a unique model of the segment based models has been considered using the segment traveltimes as a reference. As it is described below.

#### 3.4.1.1 Segment traveltime optimisation by segment error (ST\_SE)

The basis of this first model is to minimise the existing deviation in traveltimes on each segment. Minimising the deviation between traveltimes calculated with the calibrated speed profiles and the reference traveltimes.

The problem can be modelled with the following parameters definitions:

$minNC, maxNC$	Minimum and maximum networkclass.
$minSC, maxSC$	Minimum and maximum speedclass.
$NS$	Number of segments.
$NC_j$	Networkclass of segment- $j$ .
$SC_j$	Speedclass of segment- $j$ .
$T_j$	Traveltime of segment- $j$ .
$L_j$	Length of segment- $j$ .
$vMin, vMax$	Minimum and maximum speed.

The following decision variables are also defined:

$v_{k,l}$  Speed of Networkclass- $k$  and Speedclass- $l$

Finally the model is as follows:

$$\text{minimize } \sum_{j=1}^{NS} \left( \frac{L_j}{v_{NC_j, SC_j}} - T_j \right)^2 \quad (3.1)$$

subject to:

$$v_{k,l} \geq v_{k,l+1} \quad \forall k = minNC \dots maxNC, \quad (3.2)$$

$$l = minSC \dots maxSC - 1$$

$$v_{k,l} - v_{k,l+1} == v_{k,l+1} - v_{k,l+2} \quad \forall k = minNC \dots maxNC, \quad (3.3)$$

$$l = minSC \dots maxSC - 2$$

$$v_{k,minSC} \leq vMax \quad \forall k = minNC \dots maxNC \quad (3.4)$$

$$v_{k,maxSC} \geq vMin \quad \forall k = minNC \dots maxNC \quad (3.5)$$

Code 3.6: Model: Segment traveltime optimisation by segment error (ST\_SE).

(3.1): Minimisation of the sum of squares of the traveltime error on each segment.

(3.2): Speed of Speedclass- $l$  is greater than speed of Speedclass- $(l+1)$  for each Networkclass- $k$ .

(3.3): Difference between two consecutive speedclasses for each Networkclass- $k$  is the same.

(3.4): Maximum speed in the speeds table.

(3.5): Minimum speed in the speeds table.

### 3.4.2 Trip traveltime optimisation models (TT)

Only a unique model of the trip based models has been considered using the total traveltimes as a reference. As it is described below.

#### 3.4.2.1 Trip traveltime optimisation by segment error (TT\_SE)

The basis of this model is to minimise the existing deviation in traveltimes at a trip level. Minimizing the deviation between the sum of the calculated segment traveltimes with the calibrated speed profiles and the reference total traveltimes.

The problem can be modelled with the following parameters definitions:

$minNC, maxNC$	Minimum and maximum networkclass.
$minSC, maxSC$	Minimum and maximum speedclass.
$NT$	Number of trips.
$NS_i$	Number of segments in trip- $i$ .
$T_i$	Total traveltime of trip- $i$ .
$L_{i,j}$	Length of segment- $j$ in trip- $i$ .
$NC_{i,j}$	Networkclass of segment- $j$ in trip- $i$ .
$SC_{i,j}$	Speedclass of segment- $j$ in trip- $i$ .
$vMin, vMax$	Minimum and maximum speed.

The following decision variables are also defined:

$v_{k,l}$  Speed of Networkclass- $k$  and Speedclass- $l$

Finally the model is as follows:

$$\text{minimize } \sum_{i=1}^{NT} \left( T_i - \sum_{j=1}^{NS_i} \frac{L_{i,j}}{v_{NC_{i,j}, SC_{i,j}}} \right)^2 \quad (3.6)$$

subject to:

$$v_{k,l} \geq v_{k,l+1} \quad \forall k = minNC \dots maxNC, \quad (3.7)$$

$$l = minSC \dots maxSC - 1$$

$$v_{k,l} - v_{k,l+1} = v_{k,l+1} - v_{k,l+2} \quad \forall k = minNC \dots maxNC, \quad (3.8)$$

$$l = minSC \dots maxSC - 2$$

$$v_{k,minSC} \leq vMax \quad \forall k = minNC \dots maxNC \quad (3.9)$$

$$v_{k,maxSC} \geq vMin \quad \forall k = minNC \dots maxNC \quad (3.10)$$

Code 3.7: Model: Trip traveltime optimisation by segment error (TT\_SE).

(3.11): Minimisation of the sum of squares of the traveltime error on each trip.

- (3.12): Speed of Speedclass- $l$  is greater than speed of Speedclass- $(l + 1)$  for each Networkclass- $k$ .
- (3.13): Difference between two consecutive speedclasses for each Networkclass- $k$  is the same.
- (3.14): Maximum speed in the speeds table.
- (3.15): Minimum speed in the speeds table.

### 3.5 Validation methods

In order to validate the project, a system validation will be performed first. It will be checked that each component performs its assigned task giving logical results for each area analysed in its specific time slot.

As this is a project involving the processing of large amounts of data, this verification will be applied only to a random subset of the data.

A validation of the results of the system will also be carried out in two phases. First, the quality of the calculated speed profiles will be analysed, measuring the final deviations in traveltimes on the same data set, thus having a first measurement of system quality.

And then analysing the variability in traveltimes at different instants of the day. Estimating the capacity of the calibration to replicate the different traveltime variations throughout the day.

Afterwards a final validation process will be applied to a set of independent data from PTV Group clients in order to measure its effectiveness with real data.



# Chapter 4

## System architecture

### 4.1 Architecture

In this section it will be explained the architecture that has been decided to implement for the system. The main idea for the design of the system was that it should be applicable to different areas as well as agnostic concerning the external provider of reference data (the reference traveltimes).

Thus, it is decided to separate distinguishable responsibilities into different modules (see Figure 4.1). And so the following three modules were implemented:

- **Module 1: Area coordinates and initial speed settings**

This first module is in charge of all the tasks related to the area to be analysed. Its main controller (**AreaController**) is in charge of managing two more controllers:

- **TripRequestController**: in charge of loading the coordinates of the reference trips in the area to be analysed. As each area is treated in a different way, it delegates its task to the specific trip request controller of that area.
- **SpeedsTableController**: in charge of loading the speed profiles of the area to be analysed. This controller is used to calculate the speedclass of the segments provided by PTV xRoute.

- **Module 2: Data generator**

This module is in charge of generating the data from each of the suppliers for the trips provided by the previous module and integrating them into a single set of data for a later analysis. Its main controller (**DataController**) is in charge of managing two more controllers:

- **XRouteDataController**: in charge of generating the data from the PTV xRoute service.
- **ExternalReferenceDataController**: in charge of generating the reference data with the external services available in the system. In this case, it is decided to use the HERE service and therefore a controller is implemented for its generation.

- **Module 3: Speed profiles computation**

This last module is in charge of calculating the speed profiles for the reference data generated. In this case a single controller (**SpeedComputationController**) has been implemented, which can be of the following two types according to the principle on which the model used is based:

- **SpeedComputationBySegmentTraveltimeController**: if it is based on traveltimes per segment for the computation.
- **SpeedComputationByTripTraveltimeController**: if it is based on the total traveltimes per trip for the computation.

### 4.1.1 Extensibility

The system specified above has been designed separating responsibilities according to the nature of the tasks and in such a way that it would be adaptable to future changes in the area to be analysed or the external data provider. In this way, each of the possible extensions that may occur will be explained below:

#### **EXTENSION 1: NEW AREA TO ANALYSE**

The main objective of the system is to analyse an area and calculate its speed profiles. Therefore, it is essential that new areas can be added easily.

By separating all the work related to the treatment of the area to analyse into a single module, to add a new area it is only necessary to add a new specific trip request controller. And also loading the speeds table with the default format (see Table B.1) in the corresponding controller.

#### **EXTENSION 2: NEW EXTERNAL DATA PROVIDER**

Another of the main requirements in the design of the system is that it should be agnostic to the external provider reference data so that it can be easily replaced by a new one if needed.

In this way, if the external data provider is changed, only the specific controller needs to be added.

#### **EXTENSION 3: NEW SPEED SETTINGS COMPUTATION MODEL**

Another possible extension is to apply a new computation model of the speed settings. Therefore, it will only be necessary to add a new controller for its use.

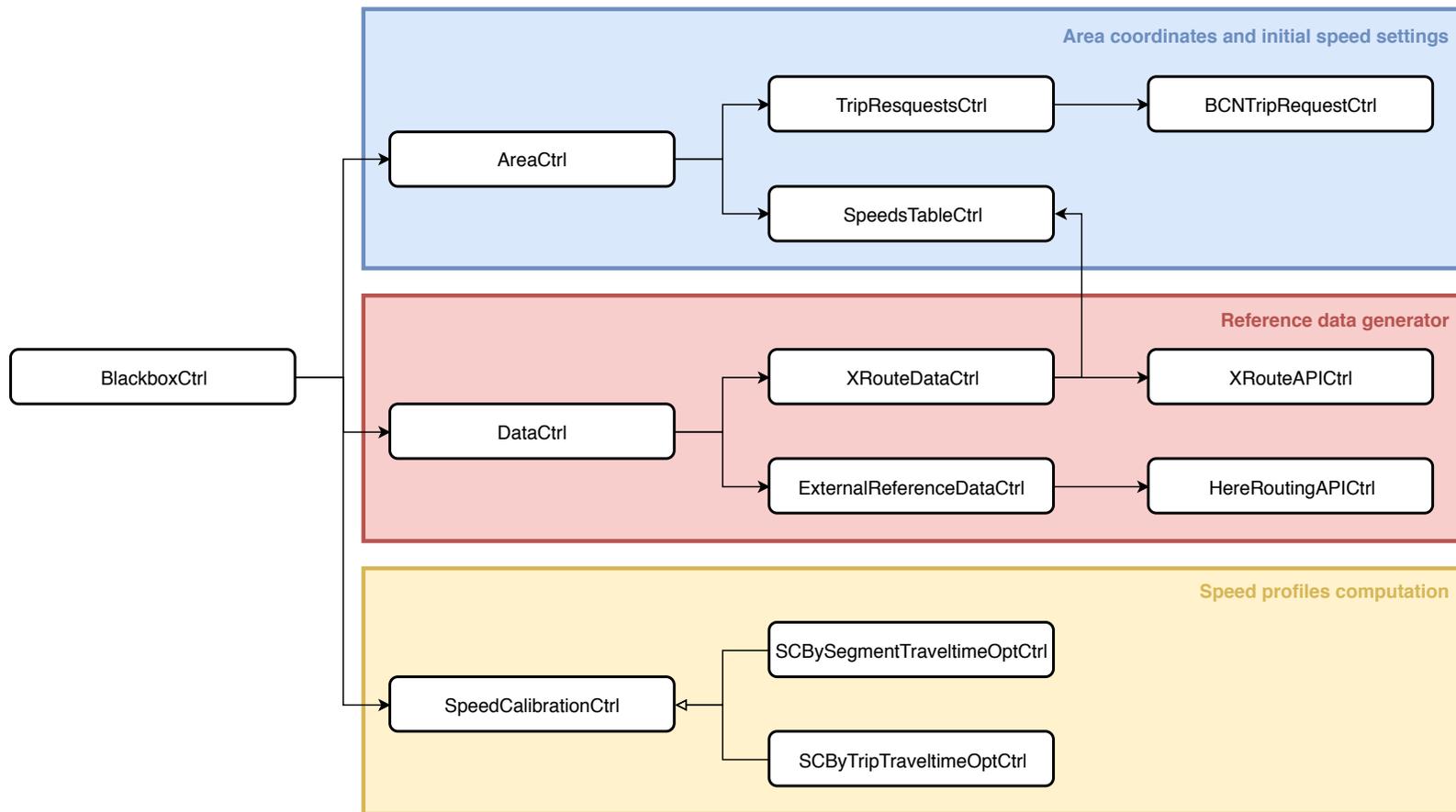


Figure 4.1: System architecture

## 4.2 Technologies

During the initial planning, several programming languages were proposed for the implementation of the system with which PTV Group usually works and with which the system would be easily integrated into its own. Thus, the technologies that have been used mainly for the development of the system will be specified below.

### **PYTHON**

For the implementation of the base system it is decided to use Python, an interpreted programming language, for the following reasons:

- It has lots of libraries for requesting external data providers as required for gathering the data from PTV xRoute and HERE services.
- It can deal easily with large data structures using the library *pandas*.
- It has different libraries for visual analysing data structures. In this case it is decided to use the *plotnine* library that is based on *ggplot2* (from *R*) for its familiarity of use of the developer.

### **AMPL**

For the optimisation models for the speed profiles computation, it is decided to use the language AMPL since it allows to execute non-linear models with the *MINOS* solver and it can be easily integrated in Python scripts with the library *amplpy*.

In addition, if due to the performance of the computer used for the prototype it does not have the necessary resources, the TRANTOR server of the UPC with a research license can be used.

# Chapter 5

## Results

### 5.1 Area coverage

This system has been developed to be extensible to multiple areas for calculating their speed profiles. However, for this prototype, its effectiveness has been evaluated in two areas only: the Metropolitan Area of Barcelona and a reduced area of the city of Barcelona.

#### 5.1.1 Barcelona Metropolitan Area (BCN)

The main analysed area covers a large portion of the Metropolitan Area of Barcelona. This way, the sample should contain segments of all the networkclasses to be analysed, since we would collect samples of all kinds of roads.

For the selection of the origin-destination pairs of the trips, we had a sample of 100,000 locations distributed throughout the Barcelona area. Resulting in a sample of almost 1,000,000 possible trips.

But due to performance issues with the computer used (described in Section 5.4), for the initial test it was decided to use two smaller samples randomly selected from all possibilities: a small sample with 1,000 trips and a larger sample with 3,000 trips.

A visualisation of the coverage of the area comprised in both samples can be seen in Figures 5.1 and 5.2. In both samples this coverage is quite wide, although logically in the sample of 3,000 trips a higher coverage of the road network is found.

Altogether with this area a total of 217,000 segments is obtained from the 1,000 trips sample and a total of 658,000 segments from the 3,000 trips sample.

Irrespectively of the size of the sample, as it was randomly collected among a very dispersed sample of possible routes, the proportion of segments obtained for each networkclass remains stable (see Tables 5.1 and 5.3).

However, this ratio is not uniform between networkclasses:

- Of networkclass 1, no sample has been obtained in any case.

## 5.1. AREA COVERAGE

- Of networkclasses 6 and 7 there is not even 1% of the total sample.
- Of networkclasses 2 and 5, there is also a very low proportion, less than 10%.
- Of networkclasses 0, 3 and 4 a large proportion is found between 25 and 40%.

It can also be appreciated that for the most important networkclasses, the collected segments are better distributed among the different speedclasses. Unlike the less important ones where there is small variation of speedclasses (see Tables 5.2 and 5.4). As preliminary predictions, analysing the distribution of segments on each classification of networkclass-speedclass, the following behaviours can be expected:

- In networkclass 1, as there are no segments of this networkclass in the sample to calibrate its speeds, its approximation will inevitably be erroneous.
- In networkclasses 6 and 7, having only information for one speedclass, it will not be possible to measure a real speed range.

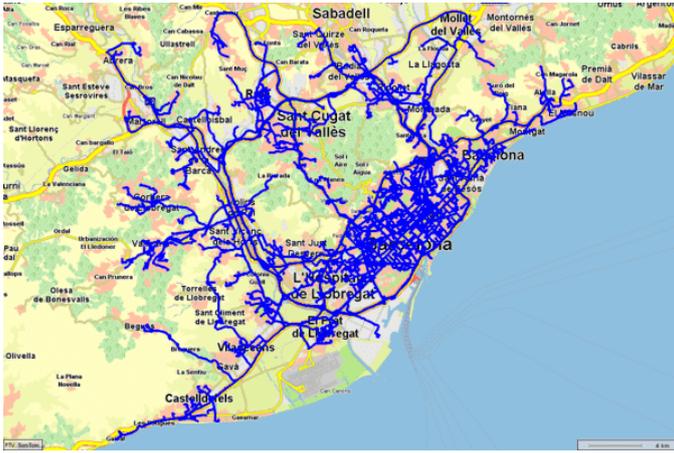


Figure 5.1: Barcelona area coverage with 1,000 trips

NC	COUNT	%
0	81,469	37.51
1	0	0
2	10,002	4.61
3	54,685	25.18
4	55,266	25.45
5	15,422	7.10
6	23	0.01
7	299	0.14

Table 5.1: Networkclass coverage with 1,000 trips on Barcelona area

NC/SC	0	1	2	3	4	5	6	7
0	0	0	14,260	4,093	9,243	2,360	41,919	9,594
1	0	0	0	0	0	0	0	0
2	0	3,394	1,784	1,582	2,257	272	367	346
3	3	812	3,560	1,258	3,763	2,383	4,116	38,790
4	0	113	1,047	148	967	2,124	10,060	40,807
5	0	1	0	181	3	11,658	3,345	234
6	0	0	0	0	0	0	23	0
7	0	0	0	0	0	299	0	0

Table 5.2: Networkclass-Speedclass coverage with 1,000 trips on Barcelona area

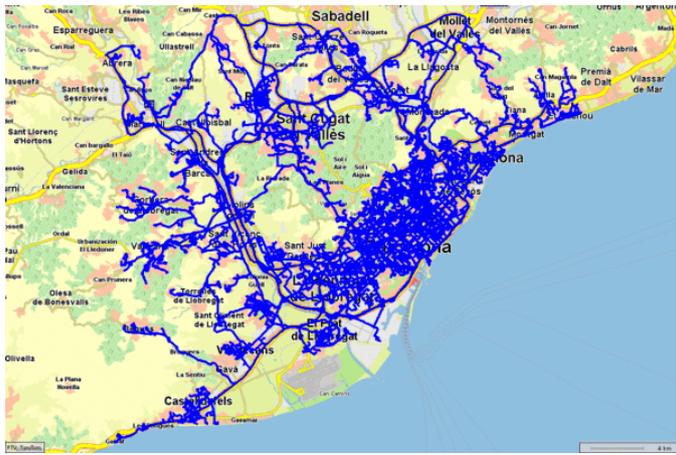


Figure 5.2: Barcelona area coverage with 3,000 trips

NC	COUNT	%
0	250,753	38.06
1	0	0
2	31,978	4.85
3	162,795	24.71
4	166,480	25.27
5	45,681	6.93
6	72	0.01
7	1,055	0.16

Table 5.3: Networkclass coverage with 3,000 trips on Barcelona area

NC/SC	0	1	2	3	4	5	6	7
0	0	0	44,101	12,121	28,919	7,571	127,541	30,500
1	0	0	0	0	0	0	0	0
2	8	10,998	5,770	4,982	6,818	884	1,367	1,151
3	14	2,470	10,592	3,931	11,942	7,215	12,865	113,766
4	0	305	3,310	451	2,865	6,801	26,801	125,947
5	0	25	1	574	50	34,103	10,210	718
6	0	0	0	0	0	0	72	0
7	0	0	0	0	0	1,055	0	0

Table 5.4: Networkclass-Speedclass coverage with 3,000 trips on Barcelona area

To see the coverage of the segments with the Barcelona area within each networkclass, in Appendix D.1 a visualization for each class can be seen.

### 5.1.2 Reduced Barcelona Metropolitan Area (BCN-R)

Using the entire Metropolitan Area of Barcelona can be quite huge, since the traffic behaviour is not the same on the city periphery as in the city centre. That is why it is decided to reduce this area to only the city of Barcelona between the *Rondas* (Barcelona's ring roads).

The selection of the origin-destination pairs of the trips has been applied in the same way as with the Barcelona area. However, in this case the trips with origin or destination outside the *Rondas* have been filtered.

For this area, two samples have also been collected randomly from all possibilities: a small sample with 1,000 trips and a larger sample with 3,000 trips.

A visualisation of the coverage of the area comprised in both samples can be seen in Figures

## 5.1. AREA COVERAGE

5.3 and 5.4. In both samples this coverage is quite wide, although logically in the sample of 3,000 trips a higher coverage of the road network is found.

Altogether with this area a total of 142,000 segments is obtained from the 1,000 trips sample and a total of 428,000 segments from the 3,000 trips sample.

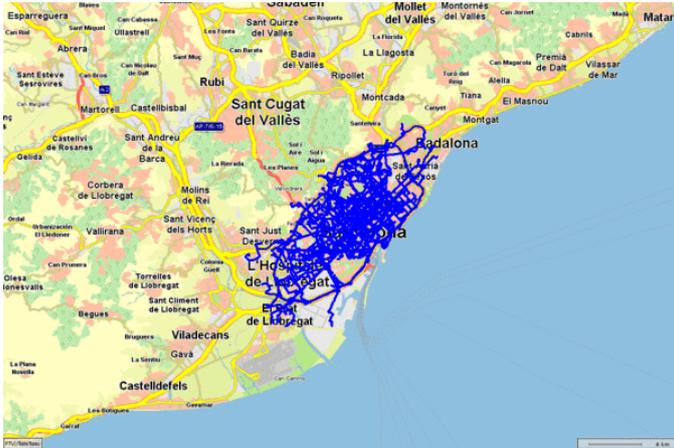
Irrespective of the size of the sample, as it was randomly collected among a very dispersed sample of possible routes, the proportion of segments obtained for each networkclass remains stable (see Tables 5.5 and 5.7).

However, this ratio is not uniform between networkclasses:

- Of networkclass 1, no sample has been obtained in any case.
- Of networkclasses 6 and 7 there is not even 1% of the total sample. And from the collected ones, all of them belong to the same speedclass (see Tables 5.6 and 5.8).
- Of networkclass 2, the proportion is decreased to about 1%.
- Of networkclasses 0 and 5, the proportion is a little low, between 10 and 15%. In the case of networkclass 0, this reduction of coverage is because this category belongs to the main highways. And with this reduced area only samples are collected for the *Rondas* (see Figure D.9 for the NC-0 coverage in BCN area).
- Of networkclasses 3 and 4 a large proportion is found between 30 and 42%.

We can also see a tendency that the collected samples belong to the slower speedclasses, as opposed to using the whole area of Barcelona where the distribution was wider.

However, since we have information for at least half of the speed classes, we do not expect any major problems when estimating their speeds. On the contrary, as occurred with the previous area, it is expected to find the same behaviour on the speeds estimation of speedclasses 1, 6 and 7.



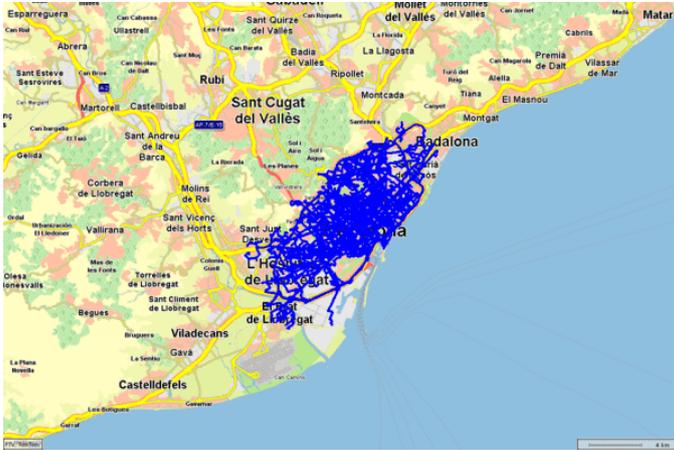
NC	COUNT	%
0	25,756	18.12
1	0	0
2	1,776	1.25
3	42,521	29.91
4	55,752	39.22
5	16,239	11.42
6	33	0.02
7	84	0.06

Figure 5.3: Barcelona reduced area coverage with 1,000 trips

Table 5.5: Networkclass coverage with 1,000 trips on Barcelona reduced area

NC/SC	0	1	2	3	4	5	6	7
0	0	0	0	0	1,797	65	19,840	4,054
1	0	0	0	0	0	0	0	0
2	0	0	0	67	56	109	1,329	215
3	0	147	28	114	335	1,490	2,183	38,224
4	0	3	0	1	14	129	17,909	37,696
5	0	0	0	1	0	9,717	5,991	530
6	0	0	0	0	0	0	33	0
7	0	0	0	0	0	84	0	0

Table 5.6: Networkclass-Speedclass coverage with 1,000 trips on Barcelona reduced area



NC	COUNT	%
0	75,656	17.67
1	0	0
2	5,403	1.26
3	131,266	30.66
4	168,578	39.38
5	46,841	10.94
6	100	0.02
7	264	0.06

Table 5.7: Networkclass coverage with 3,000 trips on Barcelona reduced area

Figure 5.4: Barcelona reduced area coverage with 3,000 trips

NC/SC	0	1	2	3	4	5	6	7
0	0	0	2	0	4,797	211	57,495	13,151
1	0	0	0	0	0	0	0	0
2	0	0	0	185	195	296	3,961	766
3	0	403	77	250	994	4,242	6,440	118,860
4	0	6	1	6	63	387	55,369	112,746
5	0	0	0	19	6	28,926	16,482	1408
6	0	0	0	0	0	0	100	0
7	0	0	0	0	0	264	0	0

Table 5.8: Networkclass-Speedclass coverage with 3,000 trips on Barcelona reduced area

To see the coverage of the segments with the Barcelona area within each networkclass, in Appendix D.2 a visualization for each class can be seen.

### 5.1.3 Data source integration limitations

During the first trials of the system, it was found that a large part of the data was erroneously obtained from the HERE service. Analysing the data, it was detected that the problem came because the system was trying to request information from a path with too many intermediate points.

The following table shows some relevant information on the analysis of requests for each sample. In the case of the BCN area it reaches up to 60% of invalid requests, decreasing to only 16% for the reduced area as the routes are usually shorter and therefore contain fewer intermediate points.

To see the limit between a valid request and an invalid request, the size of the total points requested has been analysed. Reaching up to 203 points in valid cases and being from 204 points for all invalid cases.

	VALID				INVALID			
	COUNT	%	MIN P	MAX P	COUNT	%	MIN P	MAX P
BCN-1000	398	39.80	12	203	602	60.20	204	454
BCN-3000	1,187	39.57	6	203	1,813	60.43	204	527
BCN-R-1000	839	83.90	4	203	161	16.10	204	320
BCN-R-3000	2,546	84.87	4	203	454	15.13	204	336

Table 5.9: Summary of invalid requests for each sample

According to their documentation there is no restriction in the number of points. However, such requests were erroneous in the case of very long routes. For this reason, it was decided to try generating the requests by batches, trying to limit the size of the requests splitting the route in subroutes and then gathering all the splitted information (see Figure 5.5 which includes a schema of the process).

To determine if this would solve our problem, it was decided to implement this method with a limitation of up to 200 points per request. As a result a 100% validity was achieved.

However, this implies that from making one request initially, now multiple requests must be made to get all the segments information. With the consequential penalisation in the execution times for each request and the cost per each request performed.

As a summary, the Table 5.10 shows some information about the number of batches that are produced for each initial request. In the case of the whole area of Barcelona, an average of 1.62 batches are produced, implying almost twice as many initial requests. While in the case of the reduced area, as they are shorter routes on average it is required only 1.2 requests.

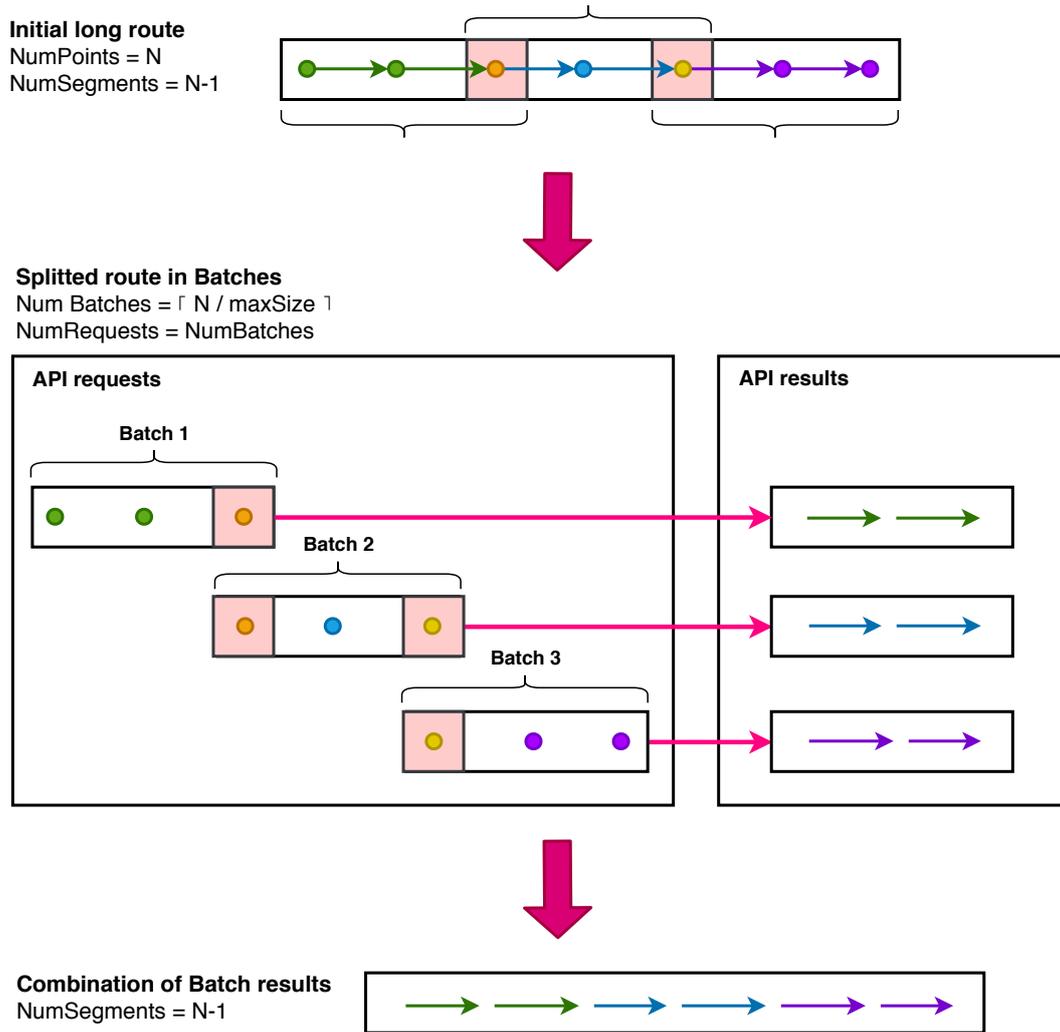


Figure 5.5: Route splitting by Batches

	MIN BATCHES	MEAN BATCHES	MAX BATCHES
BCN-1000	1	1.61	3
BCN-3000	1	1.62	3
BCN-R-1000	1	1.18	2
BCN-R-3000	1	1.16	2

Table 5.10: Summary of Batches for each sample

## 5.2 Distance analysis

When combining data from two traffic networks using PTV xRoute as the base information and HERE as the traveltime reference, the main problem lies in the fact that the routes may not correspond to each other.

Therefore, due to the problems mentioned previously about the errors that can occur when using the coordinates system of PTV xRoute on the HERE service (see Section 3.1.3), it is required to do a filtering of the sample that does not correspond between both services:

- In the case of models based on segment traveltimes, this filtering will be done on the distances of the segments.
- In the case of models based on trip traveltimes, this filtering will be done on the total distances of the trips.

In order to analyse the distances and compare the distributions of both samples (of PTV xRoute and HERE), it was decided to apply the **Two-Sample Kolmogorov-Smirnov test** (KS test) as it allows to analyse the distribution of two continuous samples [13]. Its hypotheses are:

$H_0$ : both samples come from the same distribution.

$H_1$ : the samples come from different distributions.

Assuming that the first sample has a size of  $n$  with an empirical Cumulative Distribution Function (CDF)  $F_1(x)$  and the second sample has a size of  $m$  and an empirical CDF  $F_2(x)$ , the KS test defines the statistic:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \quad (5.1)$$

Code 5.1: KS statistic.

The statistical test D is defined as the absolute maximum distance between the two empirical CDFs. As a result, a lower value on this statistic implies that the distributions resemble each other.

It is decided to use this test because of its reliability when two large samples are used and because it allows the comparison of two samples regardless of the distribution from which they come from [14].

### 5.2.1 Segment distance analysis

#### 5.2.1.1 Initial analysis

A first analysis is carried out on the initial samples of the segment distances collected between the two services to determine whether they are similar or not.

In the Appendix E.1.1 it is shown the distributions of the distances on both samples. First for the area of Barcelona and then for the reduced area. Due to computational costs because of the size of the data (more than 650k segments), the plot of the statistical KS cannot be displayed.

As it can be seen in all the cases, very short distances are obtained because those are segment distances (small portions from a route). It can also be seen how in most cases an error close to 0 is obtained between the distances of both services. However, sometimes there are errors of more than 15 km. These errors are due to what was described in Section 3.1.3 about the integration of traffic data from two different sources.

A summary of the KS test statistics is shown in the following table. It can be seen that within all cases the KS test achieves high values in the D statistic and p-values of 0. As a result, both samples do not seem to match and further filtering is required.

	1,000 TRIPS		3,000 TRIPS	
	D	P-VAL	D	P-VAL
BCN	0.069	0.000	0.070	0.000
BCN-R	0.050	0.000	0.049	0.000

Table 5.11: KS test statistics of segment distances without filter

### 5.2.1.2 Maximum absolute distance error filter (MAXD)

As a first attempt, it is decided to try a filter of maximum absolute error in the distances. In this case, there are many values that could be considered as filtering factors. But as due to resource limitations and computational costs it is not possible to carry out a design of experiments to find the best values for this factor that minimises the deviations in the traveltimes aimed by this system.

Thus, it is decided to perform a sensitivity analysis on this factor to make a selection of values to consider. As this analysis permits to study possible regions in the factor values that maximise the considered criterion. In our case, that maximises the KS test statistic without the need to experiment with multiple values that will not be suitable. Performing a first delimitation of the factor values.

The following images show the variation of the statistic D (in pink), the p-value (in green) and the resulting data size after filtering (in blue) for different factors and samples.

As can be seen, the p-value rapidly drops as the value of the factor increases. And only for values close to 0 there exist a balance maintained between a low D statistic and an approximate high p-value. However, it can also be seen that for these values, a 50% of the data is filtered.

As a possible factor of this filter, it is decided to consider the value of 0.001 km (1 m), since it minimises the statistic and achieves a slightly good p-value.

## 5.2. DISTANCE ANALYSIS

To finish, in Appendix E.1.2 the distributions resulting from filtering with a factor of 0.001 km are shown. Below it is a summary of the KS statistics for the different samples. Although the statistic improves, the p-value is still sometimes low.

	1,000 TRIPS		3,000 TRIPS	
	D	P-VAL	D	P-VAL
BCN	0.003	0.591	0.003	0.059
BCN-R	0.005	0.390	0.004	0.032

Table 5.12: KS test statistics of segment distances with MAXD-0.001 filter

To show the implications of this filter, Tables 5.13 and 5.14 present the proportion of segments per networkclass for each of the samples. As it can be seen, more than 50% of the data is filtered, filtering completely those of networkclasses 6 and 7.

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	31,023	28.85	61.92	94,583	29.14	62.28
1	0	0	-	0	0	-
2	4,390	4.08	56.11	14,196	4.37	55.61
3	30,226	28.11	44.73	90,248	27.80	44.56
4	33,233	30.90	39.87	99,878	30.77	40.01
5	8,670	8.06	43.78	25,723	7.92	43.69
6	0	0.00	100.00	0	0.00	100.00
7	0	0.00	100.00	0	0.00	100.00

Table 5.13: Networkclass coverage after using with MAXD-0.001 filter over segment distances in BCN area

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	9,550	12.23	62.92	28,740	12.18	62.01
1	0	0	-	0	0	-
2	1,212	1.55	31.76	3,557	1.51	34.17
3	23,862	30.55	43.88	73,622	31.19	43.91
4	34,013	43.55	38.99	102,559	43.45	39.16
5	9,464	12.12	41.72	27,547	11.67	41.19
6	0	0.00	100.00	0	0.00	100.00
7	0	0.00	100.00	1	0.00	99.62

Table 5.14: Networkclass coverage after using with MAXD-0.001 filter over segment distances in BCN-R area

Therefore, this filter is discarded as a candidate method because of its high instability and it does not offer good results in the analysis of the distances.

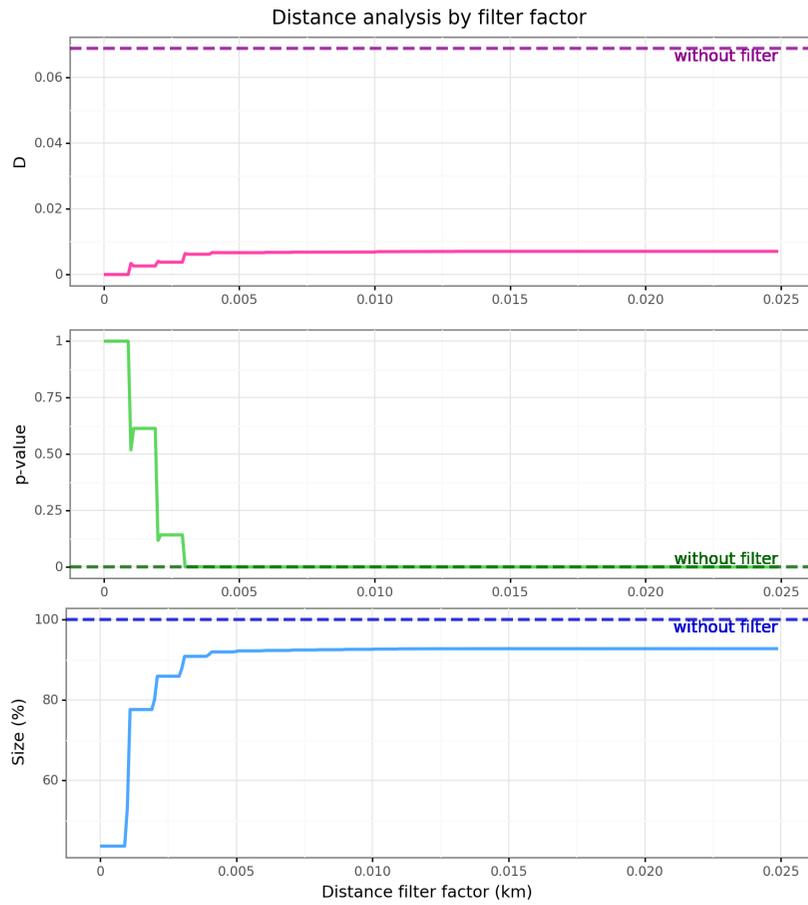


Figure 5.6: Analysis of MAXD filter factor in segment distances of BCN-1000

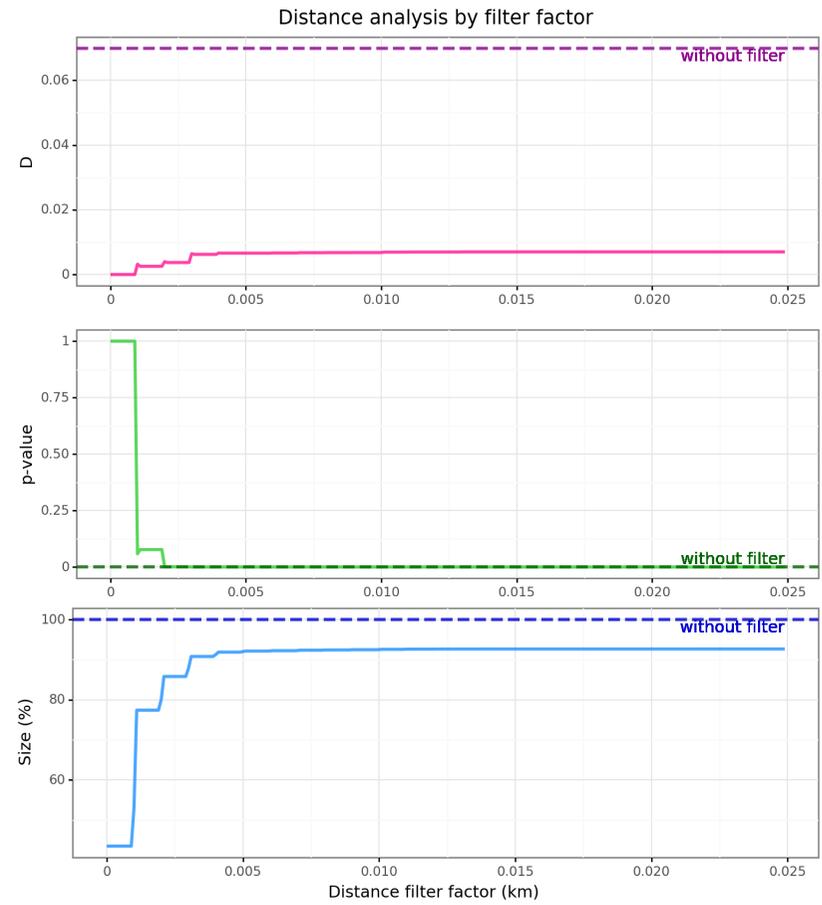


Figure 5.7: Analysis of MAXD filter factor in segment distances of BCN-3000

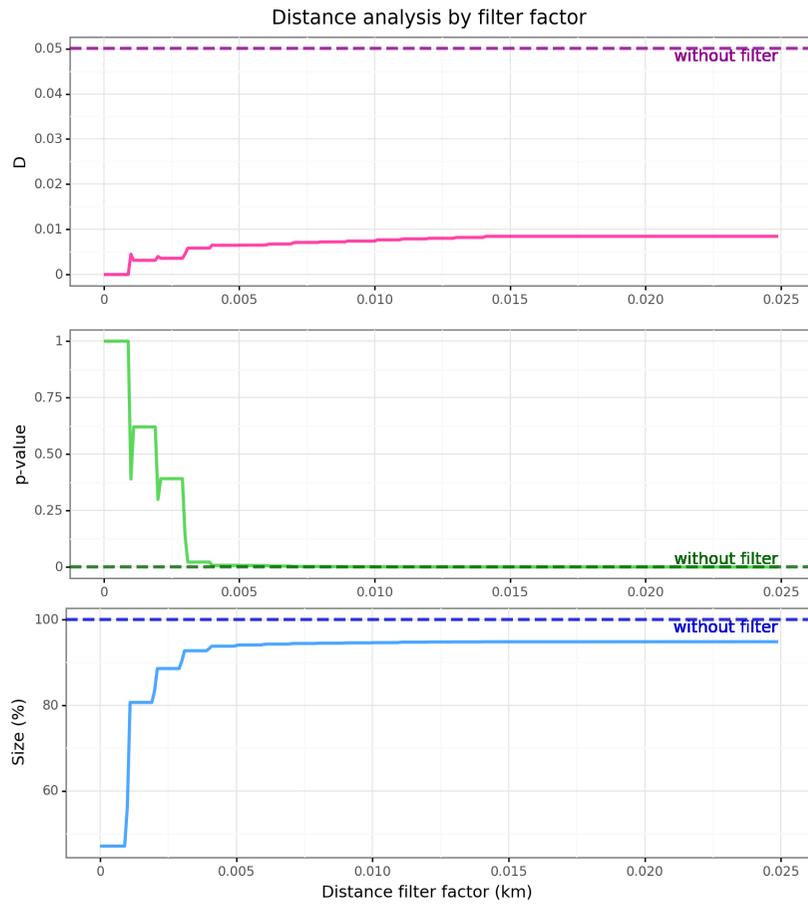


Figure 5.8: Analysis of MAXD filter factor in segment distances of BCN-R-1000

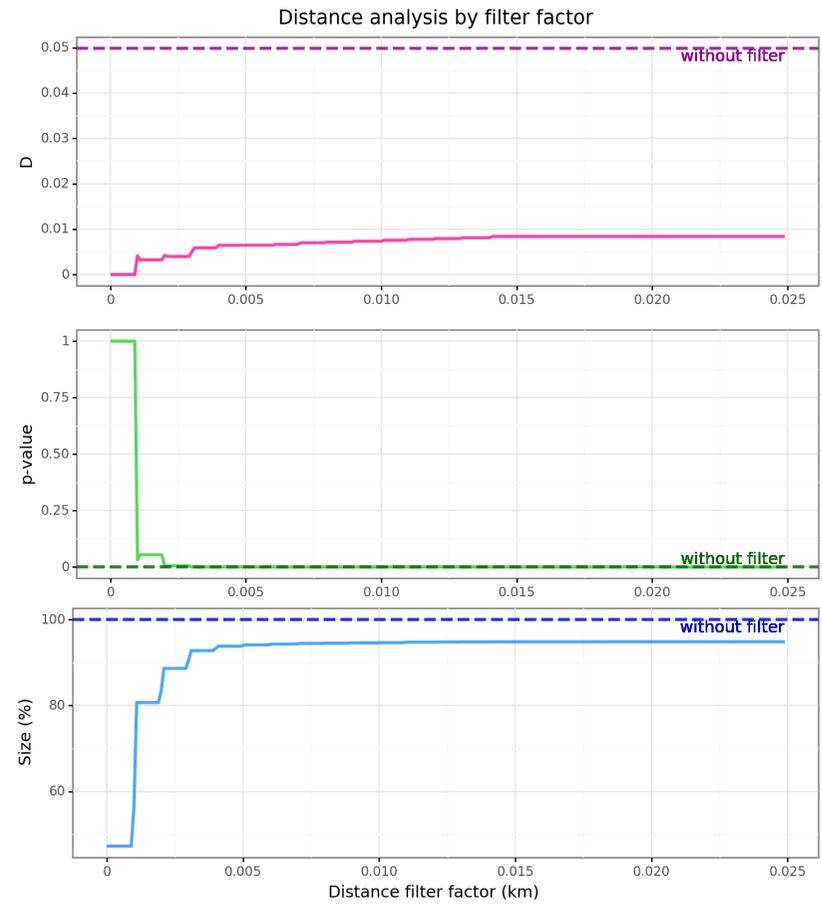


Figure 5.9: Analysis of MAXD filter factor in segment distances of BCN-R-3000

### 5.2.1.3 Maximum absolute distance percentage error filter (MAXDP)

As an attempt to correct the previous filter because of its instability, it can be used a filter of maximum absolute percentage in the distances. As in the previous case, due to resource limitations and computational costs it is not possible to carry out a design of experiments to find the best values for this factor that minimises the deviations in the traveltimes aimed by this system.

Thus, it is decided to perform a sensitivity analysis on this factor to make a selection of values to consider. As this analysis permits to study possible regions in the factor values that maximise the considered criterion. In our case, that maximises the KS test statistic without the need to experiment with multiple values that will not be suitable. Performing a first delimitation of the factor values.

The following images show the variation of the statistic D (in pink), the p-value (in green) and the resulting data size after filtering (in blue) for different factors and the different samples.

As can be seen, it can be appreciated how this filter behaves similar to the previous one, although less accentuated. The p-value rapidly drops as the value of the factor increases for the big sample cases, but a little slower for the small sample cases.

In either case, for low factor values between 0 and 0.5%, high values close to 1 in the p-value and low values in the D statistic are obtained. However, it can also be seen that for these values 50% of the data is filtered.

To finish, in Appendix E.1.3 the distributions resulting from filtering with a factor of 0.005 (0.5%) are shown. Below it is a summary of the KS statistics for the different samples. It can be seen as in all cases the statistic is quite low and very close to 0 with a p-value of 1.

	1,000 TRIPS		3,000 TRIPS	
	D	P-VAL	D	P-VAL
BCN	0.001	1.000	0.002	1.000
BCN-R	0.002	1.000	0.001	1.000

Table 5.15: KS test statistics of segment distances with MAXDP-0.005 filter

To show the implications of this filter, Tables 5.16 and 5.17 present the proportion of segments per networkclass for each of the samples. As it can be seen, more than 50% of the segments of each category are filtered, filtering completely those of networkclasses 6 and 7.

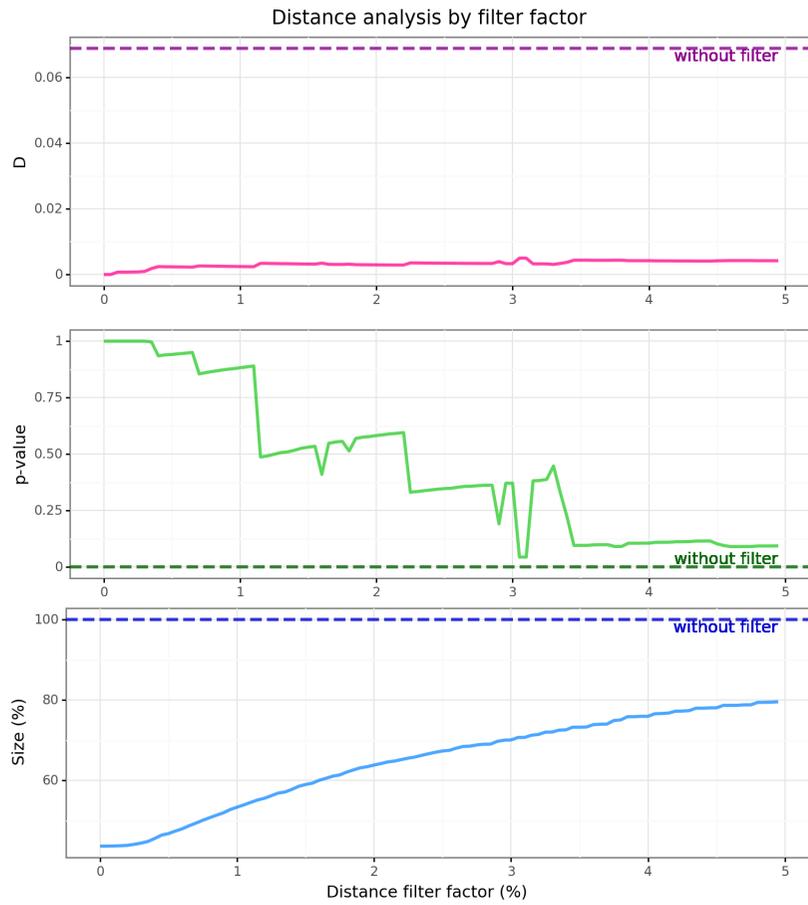


Figure 5.10: Analysis of MAXDP filter factor in segment distances of BCN-1000

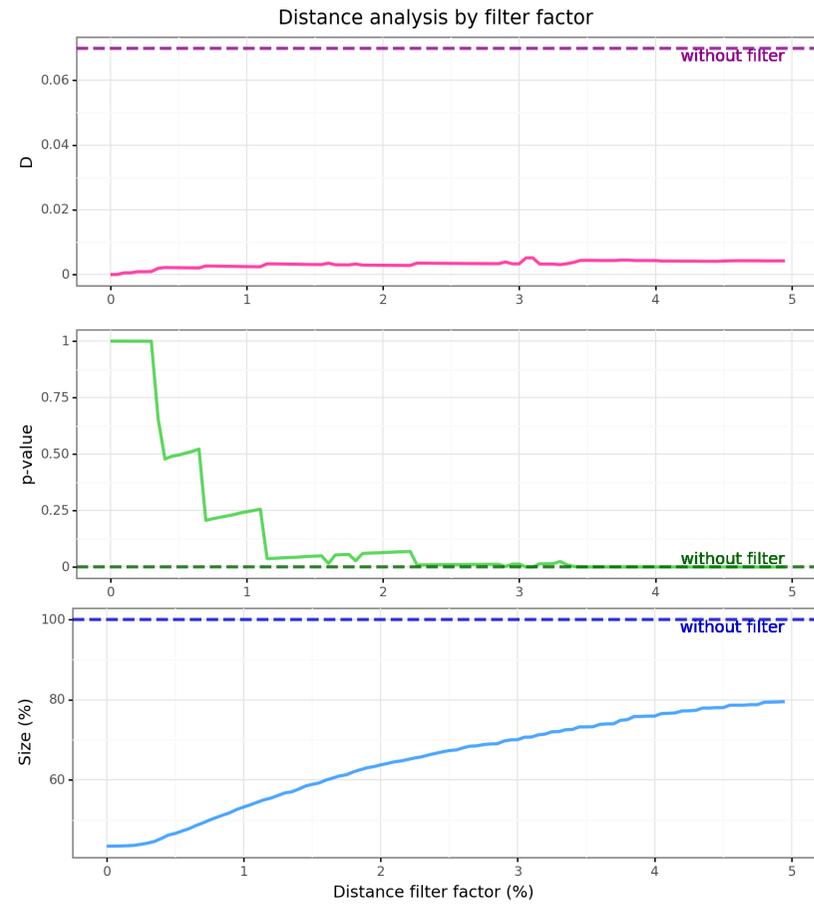


Figure 5.11: Analysis of MAXDP filter factor in segment distances of BCN-3000

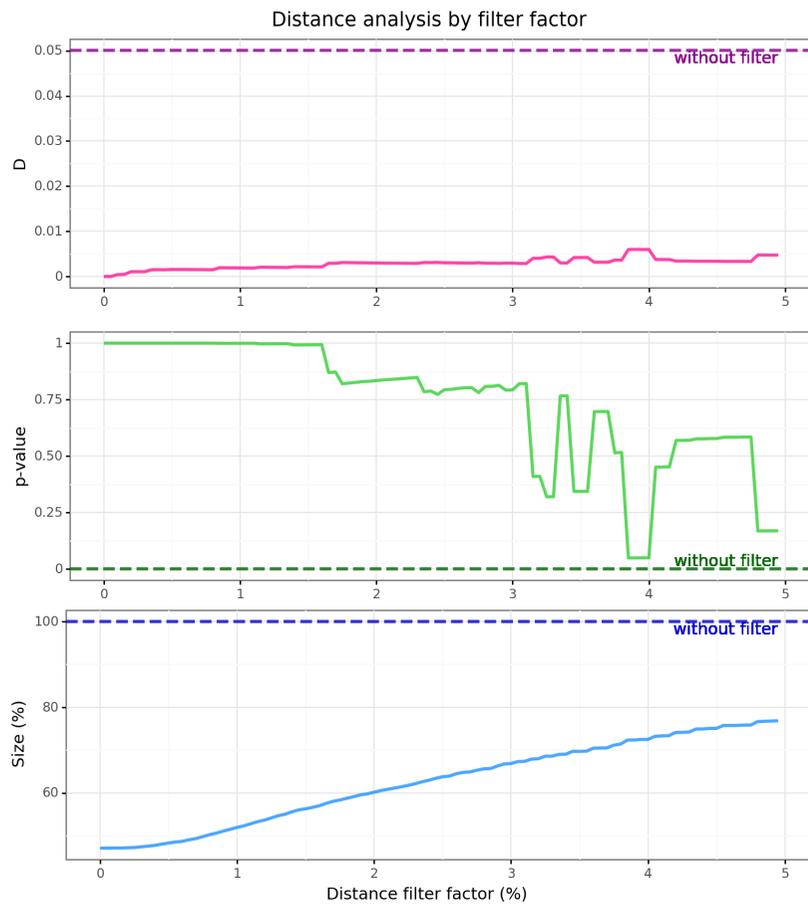


Figure 5.12: Analysis of MAXDP filter factor in segment distances of BCN-R-1000

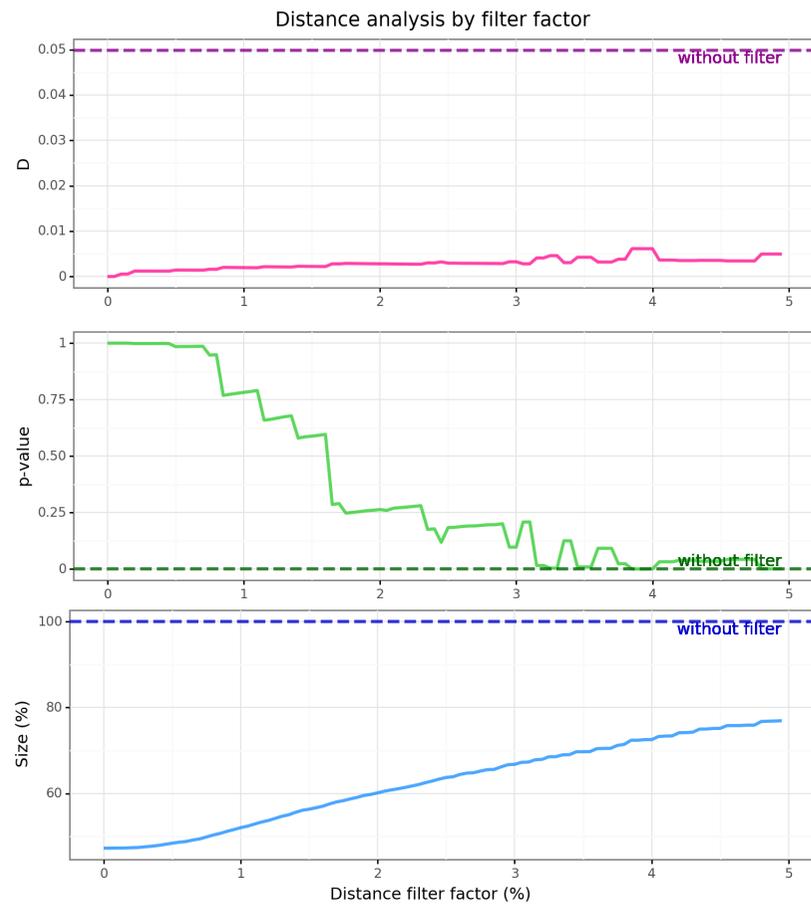


Figure 5.13: Analysis of MAXDP filter factor in segment distances of BCN-R-3000

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	33,959	33.50	58.32	103,577	33.78	58.69
1	0	0	-	0	0	-
2	4,818	4.75	51.83	15,554	5.07	51.36
3	26,443	26.08	51.64	79,018	25.77	51.46
4	28,461	28.08	48.50	85,628	27.93	48.57
5	7,693	7.59	50.12	22,822	7.44	50.04
6	0	0	100.00	0	0	100.00
7	0	0	100.00	0	0	100.00

Table 5.16: Networkclass coverage after using with MAXDP-0.005 filter over segment distances in BCN area

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	10,110	14.76	60.75	30,297	14.65	59.95
1	0	0	-	0	0	-
2	1,098	1.60	38.18	3,223	1.56	40.35
3	20,430	29.83	51.95	63,070	30.49	51.95
4	28,571	41.72	48.75	86,157	41.65	48.89
5	8,275	12.08	49.04	24,121	11.66	48.50
6	0	0	100.00	0	0	100.00
7	0	0	100.00	1	0.00	99.62

Table 5.17: Networkclass coverage after using with MAXDP-0.005 filter over segment distances in BCN-R area

Although it filters a 50% of the sample and two networkclass sample almost completely, it is decided to consider this filter as a possible filtering method as it obtains good results in the distance analysis and carry out a later analysis with the experiments.

## 5.2.2 Trip distance analysis

### 5.2.2.1 Initial analysis

A first analysis is carried out on the initial samples of the trip distances collected between the two services to determine whether they are similar or not.

In the Appendix E.2.1 it is shown the distributions of the distances and the KS test on both samples first for the area of Barcelona and then for the reduced area.

As it is logical the distances are larger for the area of Barcelona. And also the shape of the distributions differs between both scenarios, being in the reduced area one similar to a long tail distribution.

As a summary, the following table shows the results of the KS test statistics. It can be seen that the KS test achieves high values in the D statistic for all cases and with a quite low p-value. As a result, both samples do not seem to match and further filtering is required.

	1,000 TRIPS		3,000 TRIPS	
	D	P-VAL	D	P-VAL
BCN	0.027	0.855	0.024	0.333
BCN-R	0.043	0.307	0.041	0.012

Table 5.18: KS test statistics of trip distances without filter

### 5.2.2.2 Maximum absolute distance error filter (MAXD)

As a first attempt, it is decided to try a filter of maximum absolute error in the distances. In this case, there are many values that could be considered as filtering factors. But due to resource limitations and computational costs, it is not possible to carry out a design of experiments to find the best values for this factor that minimises the deviations in the traveltimes aimed by this system.

Thus, it is decided to perform a sensitivity analysis on this factor to make a selection of values to consider. As this analysis permits to study possible regions in the factor values that maximise the considered criterion. In our case, that maximises the KS test statistic without the need to experiment with multiple values that will not be suitable. Performing a first delimitation of the factor values.

The following images show the variation of the statistic D (in pink), the p-value (in green) and the resulting data size after filtering (in blue) for different factors and the different samples.

As can be observed for all samples, using low values between 1 and 4 km, the statistic D decreases by more than half compared to the initial sample and having a p-value close to 1. Therefore, the single factor of 1 km could be considered as a feasible value. Since for all samples the best values in the statistic D are obtained and retaining even more than 65% of the initial data in the worst case.

To finish, in Appendix E.2.2 the distributions resulting from filtering with a factor of 1 km are shown. The distributions remain stable, but the errors are reduced a lot. Below is the summary table with the KS test statistics that have decreased a lot.

	1,000 TRIPS		3,000 TRIPS	
	D	P-VAL	D	P-VAL
BCN	0.011	1.000	0.007	1.000
BCN-R	0.019	0.998	0.013	0.975

Table 5.19: KS test statistics of trip distances with MAXD-1 filter

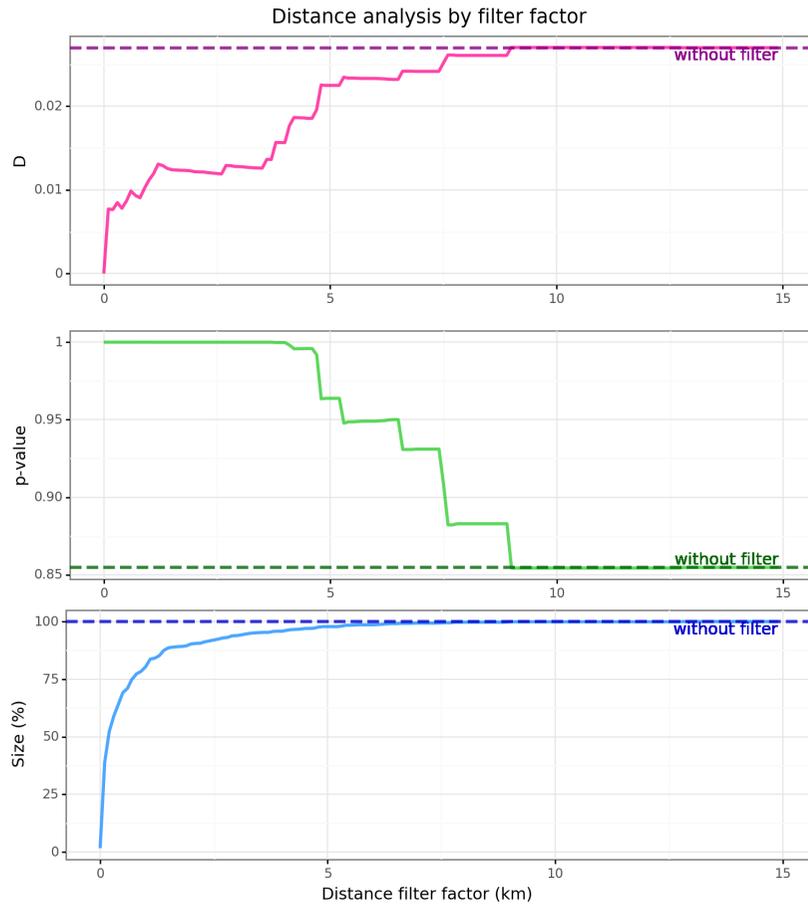


Figure 5.14: Analysis of MAXD filter factor in trip distances from BCN-1000

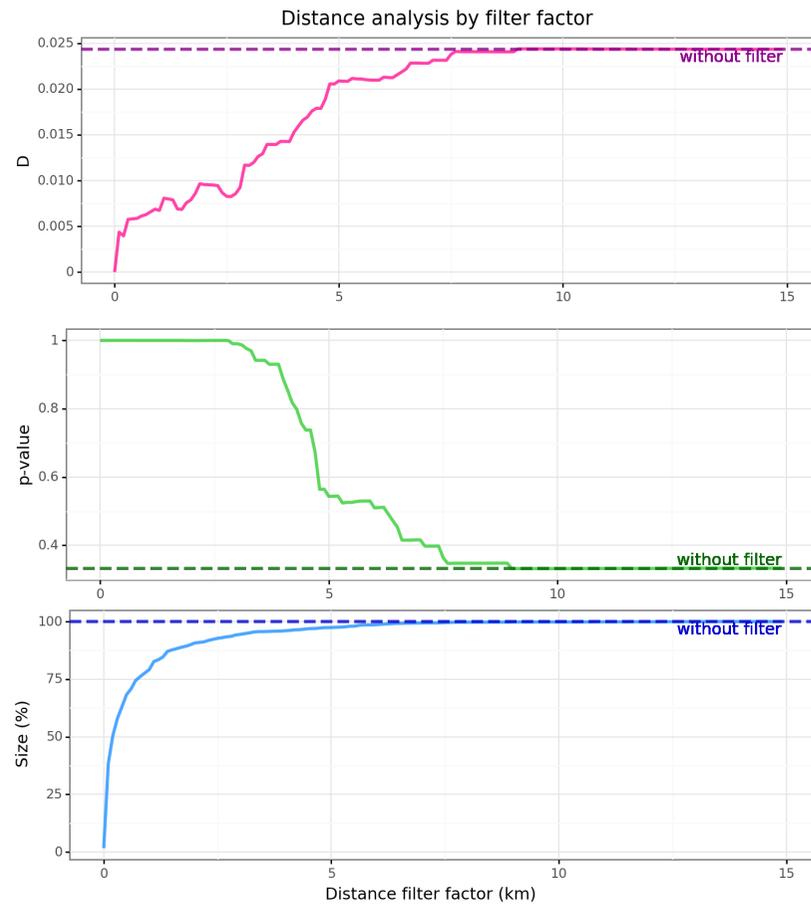


Figure 5.15: Analysis of MAXD filter factor in trip distances from BCN-3000

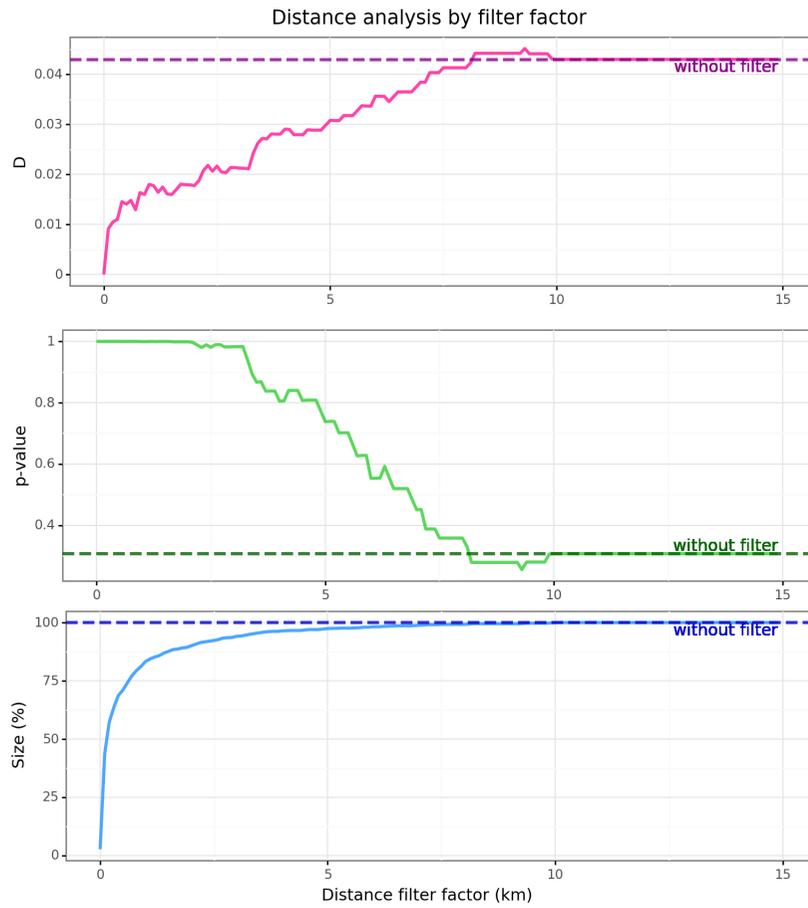


Figure 5.16: Analysis of MAXD filter factor in trip distances from BCN-R-1000

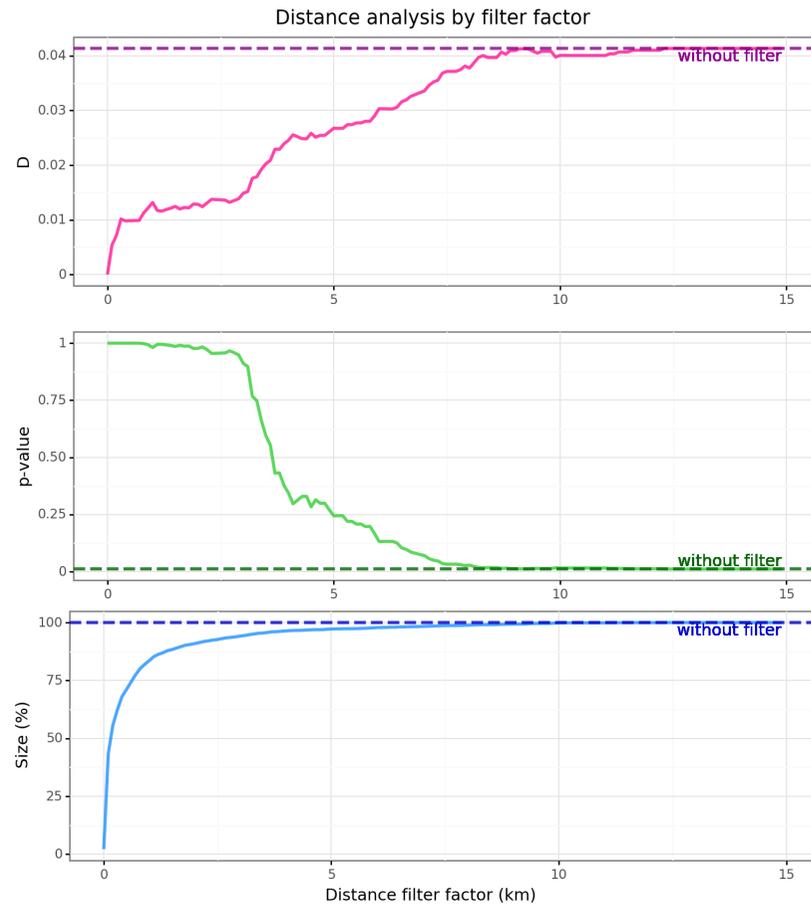


Figure 5.17: Analysis of MAXD filter factor in trip distances from BCN-R-3000

To show the implications of this filter, Tables 5.20 and 5.21 present the proportion of segments per networkclass for each of the samples. As opposed to the segment distance filtering, in this case it can be seen how a smaller portion of the data is filtered, still keeping part of the segments of networkclasses 6 and 7.

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	58,426	35.83	28.28	181,283	36.61	27.70
1	0	0	-	0	0	-
2	7,502	4.60	25.00	24,418	4.93	23.64
3	44,041	27.01	19.46	129,556	26.17	20.42
4	40,983	25.13	25.84	123,797	25.00	25.64
5	11,868	7.28	23.05	35,183	7.11	22.98
6	18	0.01	21.74	55	0.01	23.61
7	225	0.14	24.75	817	0.17	22.56

Table 5.20: Networkclass coverage after using with MAXD-1 filter over trip distances in BCN area

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	16,035	14.63	37.74	48,326	14.33	36.12
1	0	0	-	0	0	-
2	785	0.72	55.80	2,622	0.78	51.47
3	34,481	31.45	18.91	108,885	32.29	17.05
4	44,822	40.89	19.60	138,389	41.04	17.91
5	13,403	12.23	17.46	38,628	11.46	17.53
6	29	0.03	12.12	88	0.03	12.00
7	72	0.07	14.29	229	0.07	13.26

Table 5.21: Networkclass coverage after using with MAXD-1 filter over trip distances in BCN-R area

Therefore, this filter factor seems to be working properly and can be taken into account to be used in the system.

### 5.2.2.3 Maximum absolute distance percentage error filter (MAXDP)

As an attempt to improve the previous filter to better extend its application to multiple areas, it can be used a filter of maximum absolute percentage in the distances. As in the previous case, due to resource limitations and computational costs it is not possible to carry out a design of experiments to find the best values for this factor that minimises the deviations in the traveltimes aimed by this system.

Thus, it is decided to perform a sensitivity analysis on this factor to make a selection of values to consider. As this analysis permits to study possible regions in the factor values that maximise the considered criterion. In our case, that maximises the KS test statistic without the need to experiment with multiple values that will not be suitable. Performing a first delimitation of the factor values.

The following images show the variation of the statistic D (in pink), the p-value (in green) and the resulting data size after filtering (in blue) for different factors and the different samples.

As can be observed for all samples, using low values between 1 and 20%, the statistic D decreases by more than half compared to the initial sample and having a p-value close to 1. Therefore, the single factor of 10% could be considered as a feasible value. Since for all samples the best values in the statistic D are obtained and retaining even more than 65% of the initial data in the worst case.

To finish, in Appendix E.2.3 the distributions resulting from filtering with a factor of 10% are shown. The distributions remain stable, but the errors are reduced a lot. Below is the summary table with the KS test statistics that have decreased a lot.

	1,000 TRIPS		3,000 TRIPS	
	D	P-VAL	D	P-VAL
BCN	0.011	1.000	0.005	1.000
BCN-R	0.013	1.000	0.004	1.000

Table 5.22: KS test statistics of trip distances with MAXDP-0.1 filter

To show the implications of this filter, Tables 5.23 and 5.24 present the proportion of segments per networkclass for each of the samples. As it can be seen, this filter preserves more part of the samples, keeping also some portion of networkclass 6 and 7.

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	70,825	37.51	13.07	220,183	38.63	12.19
1	0	0	-	0	0	-
2	8,633	4.57	13.69	28,280	4.96	11.56
3	47,901	25.37	12.41	140,691	24.68	13.58
4	47,814	25.32	13.48	140,988	24.74	15.31
5	13,379	7.09	13.25	38,869	6.82	14.91
6	21	0.01	8.70	64	0.01	11.11
7	240	0.13	19.73	887	0.16	15.92

Table 5.23: Networkclass coverage after using with MAXDP-0.1 filter over trip distances in BCN area

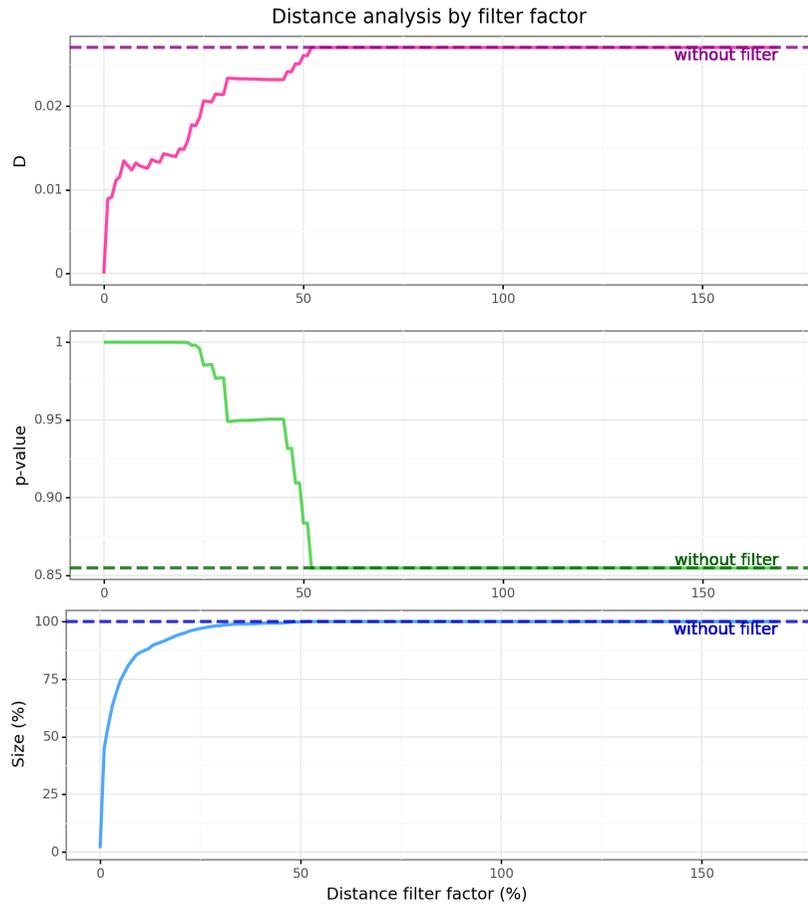


Figure 5.18: Analysis of MAXDP filter factor in trip distances from BCN-1000

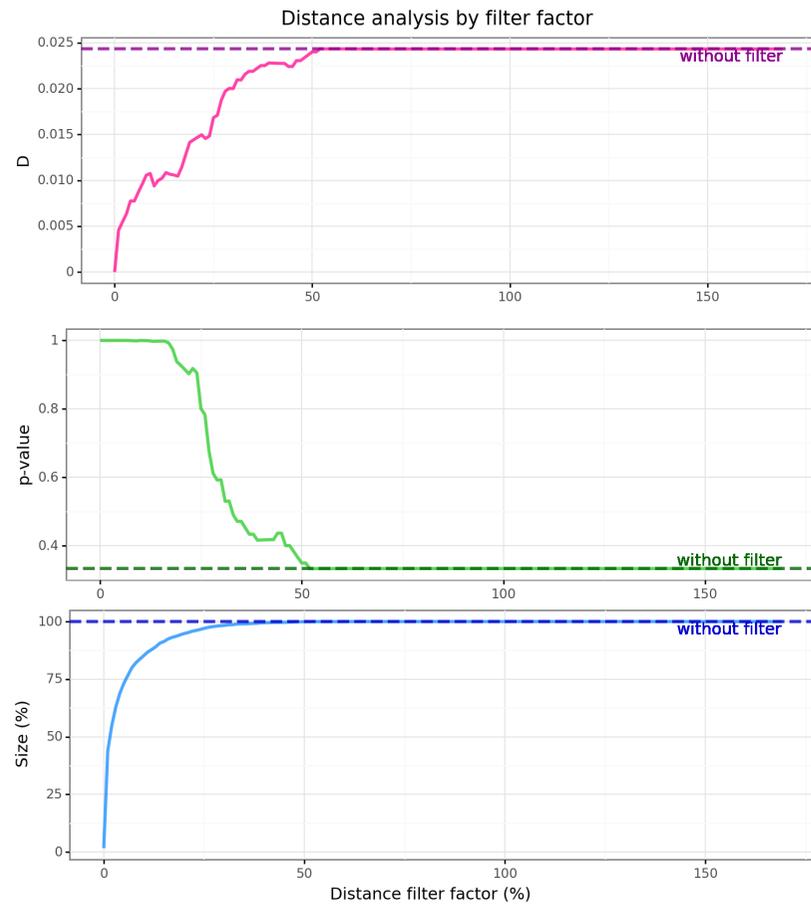


Figure 5.19: Analysis of MAXDP filter factor in trip distances from BCN-3000

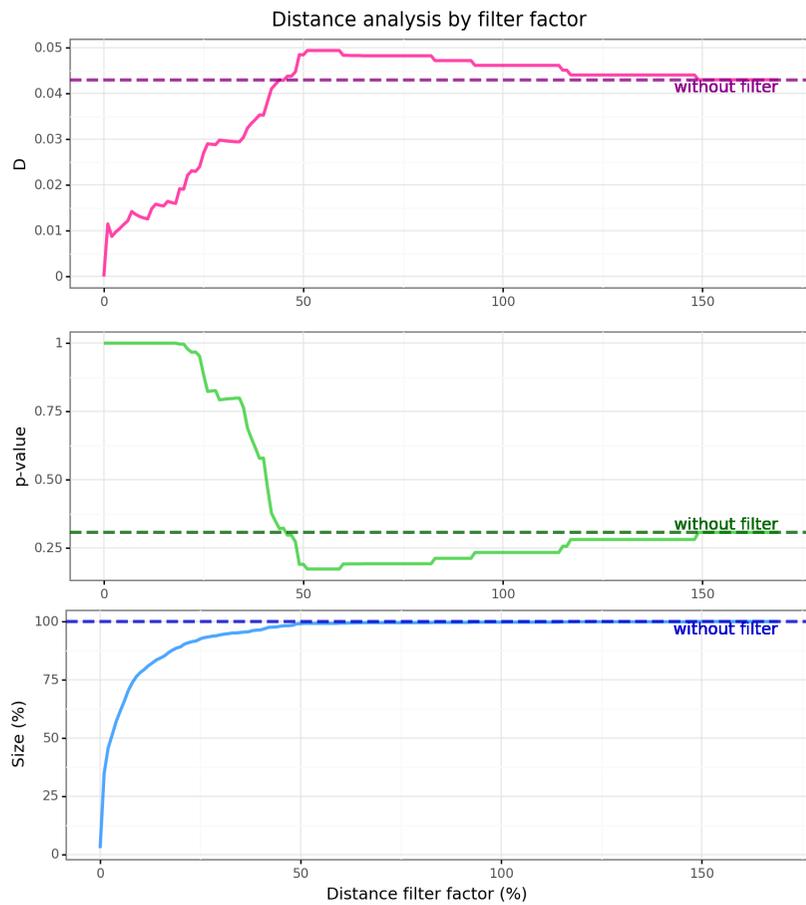


Figure 5.20: Analysis of MAXDP filter factor in trip distances from BCN-R-1000

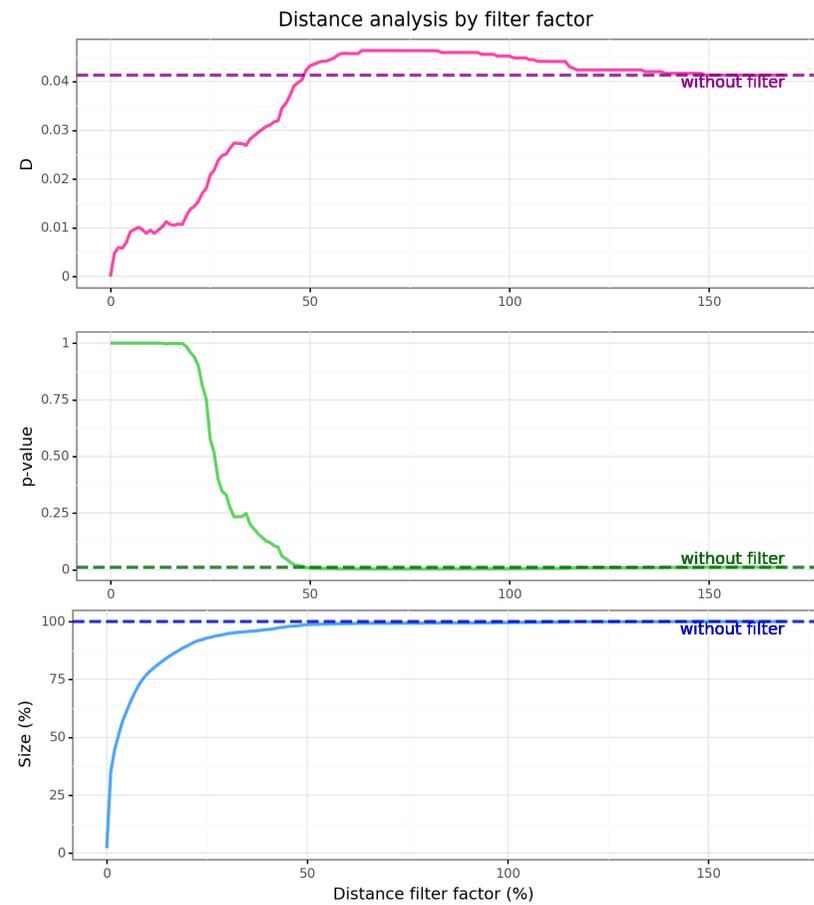


Figure 5.21: Analysis of MAXDP filter factor in trip distances from BCN-R-3000

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	16,719	15.92	35.09	52,051	16.22	31.20
1	0	0	-	0	0	-
2	814	0.77	54.17	2632	0.82	51.29
3	32,848	31.27	22.75	102,305	31.88	22.06
4	42,274	40.25	24.17	127,918	39.87	24.12
5	12,290	11.70	24.32	35,677	11.12	23.83
6	25	0.02	24.24	81	0.03	19.00
7	63	0.06	25.00	195	0.06	26.14

Table 5.24: Networkclass coverage after using with MAXDP-0.1 filter over trip distances in BCN-R area

Therefore, this filter factor seems to be working properly and can be taken into account to be used in the system.

#### 5.2.2.4 Euclidean distance filter (EUCL)

As previously mentioned in the specification of this filter (see Section 3.2 and also the Appendix C), according to several studies on distance distributions, a typical filter on Euclidean route distances is often used. In the case of Barcelona, the filter factor that usually taken in Spain is 1.269. Therefore, it is decided to test its effectiveness in our case.

As can be seen in the Appendix E.2.4, using this filter factor the shape of the distributions of the distances is completely shattered.

Below is the summary of the statistics from the KS test using this filter. In all cases the ks test statistic increases and a high p-value is obtained. As a result, the samples do not come from the same distribution.

	1,000 TRIPS		3,000 TRIPS	
	D	P-VAL	D	P-VAL
BCN	0.055	0.999	0.030	1.000
BCN-R	0.066	0.916	0.028	0.999

Table 5.25: KS test statistics of trip distances with EUCL-1.269 filter

To show the implications of this filter, Tables 5.26 and 5.27 present the proportion of segments per networkclass for each of the samples. In this case, it is easily observable how this filter removes most of the data. Filtering a minimum of 85% of all the segments from each networkclass.

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	3,384	21.70	95.85	12,208	25.29	95.13
1	0	0	-	0	0	-
2	279	1.79	97.21	849	1.76	97.35
3	6,130	39.30	88.79	18,345	38.00	88.73
4	4,661	29.88	91.57	13,687	28.36	91.78
5	1,102	7.07	92.85	3,057	6.33	93.31
6	1	0.01	95.65	4	0.01	94.44
7	40	0.26	86.62	120	0.25	88.63

Table 5.26: Networkclass coverage after using with EUCL-1.269 filter over trip distances in BCN area

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	1,289	6.78	95.00	3,250	6.46	95.70
1	0	0	-	0	0	-
2	70	0.37	96.06	135	0.27	97.50
3	7,051	37.08	83.42	19,813	39.39	84.91
4	8,855	46.57	84.12	22,659	45.05	86.56
5	1,739	9.15	89.29	4,397	8.74	90.61
6	2	0.01	93.94	11	0.02	89.00
7	9	0.05	89.29	30	0.06	88.64

Table 5.27: Networkclass coverage after using with EUCL-1.269 filter over trip distances in BCN-R area

Given these poor results in the distance analysis and the large proportion of the filtered sample, this filter does not seem to be a good option for our system. Therefore, it is decided to discard this filter.

### 5.2.2.5 Dynamic euclidean distance filter (DYN-EUCL)

To rectify the previous Euclidean distance filter with an incorrect factor for our analysed area, it is decided to apply this Euclidean distance filter with a dynamic factor specific to each area (see specification in Section 3.2.3 and Appendix C).

The first step is to calculate this factor for each area. For a proper estimation of this factor, a sufficiently large and representative sample has to be used. In this case, it is decided to use a sample of 100,000 trips randomly distributed throughout the area for this process.

This factor is estimated as the average value between the Euclidean distance and the shortest path between all trips. For the calculation of the shortest path any navigation service can

be used with a minimum reliability. In this case, it is decided to use the PTV xRoute service itself and not HERE because of its unlimited quota of requests.

In Figures 5.22 and 5.23 the ratio between the Euclidean and shortest path distances is shown. In either case it is possible to observe the typical shape (similar to a triangle shape) in the relationship between the two distances.

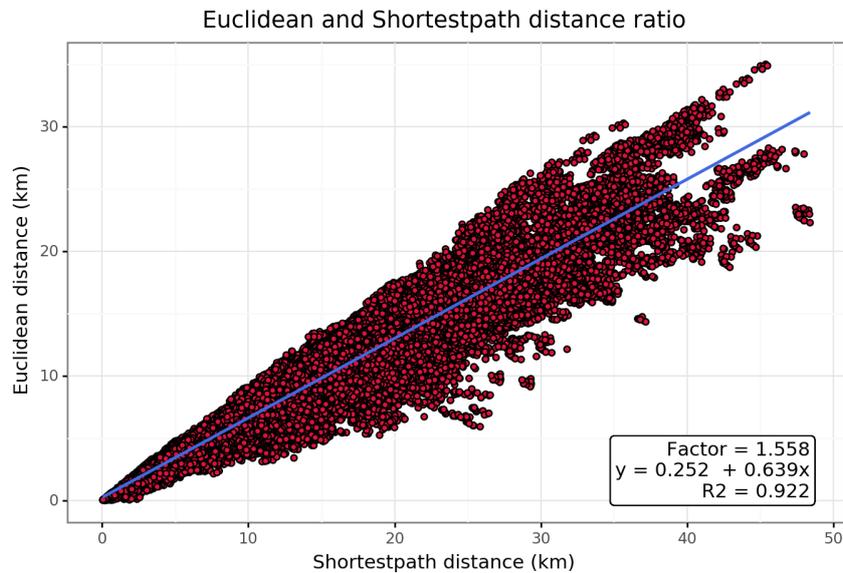


Figure 5.22: Euclidean and shortest path distance ratio on the Barcelona area

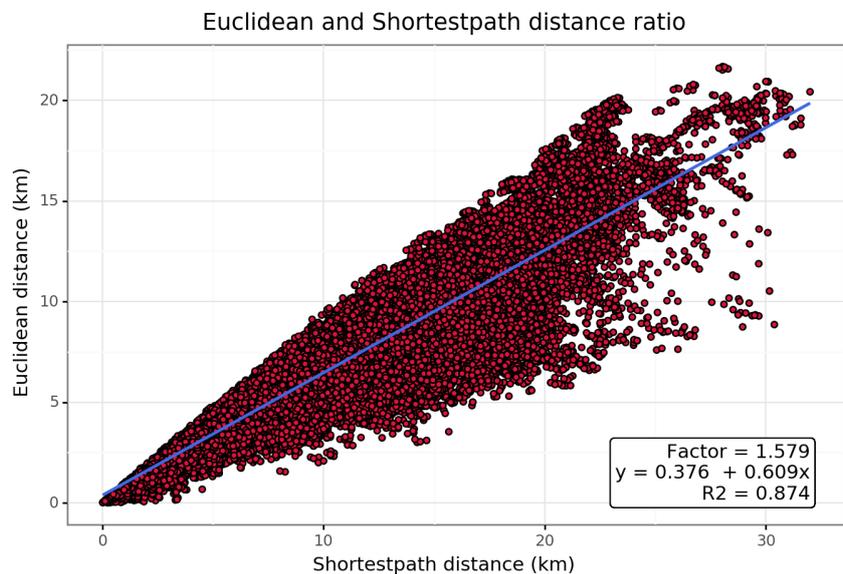


Figure 5.23: Euclidean and shortest path distance ratio on the Barcelona reduced area

The result obtained are factor values of 1.558 and 1.579 for the Barcelona and reduced Barcelona areas respectively. Differing quite a lot from the 1.269 factor value commonly used in Spain.

To finish, in Appendix E.2.5 the distributions resulting from filtering with those dynamic euclidean factors are shown. The distributions remain stable, but the errors are considerably reduced.

As a summary of the results, in the following table it is shown the KS test statistics and the p-values obtained. In this case, using these new values as a factor for the Euclidean distance filter it can be observed how the statistics of the KS test decrease by about a 50% of the initial value having a very high p-value.

	<b>1,000 TRIPS</b>		<b>3,000 TRIPS</b>	
	<b>D</b>	<b>P-VAL</b>	<b>D</b>	<b>P-VAL</b>
BCN	0.016	1.000	0.011	1.000
BCN-R	0.023	0.996	0.018	0.921

Table 5.28: KS test statistics of trip distances with DYN-EUCL filter

To show the implications of this filter, Tables 5.29 and 5.30 present the proportion of segments per networkclass for each of the samples. In this case, a large part of the data is still filtered, but in a smaller proportion. Also keeping a minimum presence of all network classes.

<b>NC</b>	<b>1,000 TRIPS</b>			<b>3,000 TRIPS</b>		
	<b>COUNT</b>	<b>%</b>	<b>% FILT.</b>	<b>COUNT</b>	<b>%</b>	<b>% FILT.</b>
0	40,920	34.53	49.77	130,122	36.24	48.11
1	0	0	-	0	0	-
2	3,867	3.26	61.34	12,177	3.39	61.92
3	33,264	28.07	39.17	97,154	27.06	40.32
4	31,898	26.91	42.28	94,547	26.33	43.21
5	8,431	7.11	45.33	24,514	6.83	46.34
6	12	0.01	47.83	39	0.01	45.83
7	130	0.11	56.52	505	0.14	52.13

Table 5.29: Networkclass coverage after using with DYN-EUCL filter over trip distances in BCN area

As a result, both samples appear to come from the same distribution and this filter can be considered as a good distance filter for our system.

NC	1,000 TRIPS			3,000 TRIPS		
	COUNT	%	% FILT.	COUNT	%	% FILT.
0	9,529	11.34	63.00	25,772	10.28	65.94
1	0	0	-	0	0	-
2	669	0.80	62.33	1,789	0.71	66.89
3	28,622	34.06	32.69	89,185	35.56	32.06
4	35,542	42.29	36.25	106,502	42.46	36.82
5	9,619	11.44	40.77	27,340	10.90	41.63
6	21	0.02	36.36	63	0.03	37.00
7	44	0.05	47.62	152	0.06	42.42

Table 5.30: Networkclass coverage after using with DYN-EUCL filter over trip distances in BCN-R area

### 5.2.3 Conclusions

Taking into account all the previous results of the distance analysis, it can be concluded that initially the data from the external routing service is not valid for using without any filtering in neither case with segment or trip distances. As mentioned in Section 3.1.3, by using different coordinate systems, errors can be introduced in the routes and not coincide between them.

For the segment distance filters, due to these matching errors at each of the intermediate points required, many errors are introduced in more than 50% of the size of the samples. In this case, only the filter of the error percentages with a 0.5% factor (MAXDP-0.005) can be considered as valid.

In contrast, for trip distances, it can be noticed that the information collected is more stable as errors are only introduced at the locations close to the origin and destination. For this reason the filters perform better than with the segment distances.

About the filtering of the errors in the distances with factor of 1 m (MAXD-0.001) despite filtering less amount of data, it obtains worse results with the analysis of the distances that conclude that both samples do not come from the same distribution. Therefore it is decided to discard it.

From the analysed filters, it can be seen that the Euclidean distance with 1.269 (EUCL-1.269) cannot be considered. In this case, although it is a factor value usually used in Spain, it does not improve the results and it filters almost the totality of the sample. Therefore it is decided to discard this filter.

From the other possibilities, the maximum absolute distance error of 1 km (MAXD-1), the maximum absolute distance percentage error of 10% (MAXDP-0.1) and the euclidean dynamic factor of 1.558 or 1.579 (DYN-EUCL), it can be seen that the results are promising.

However, among all these filters, it has to be emphasised that only the case of the DYN-EUCL is adaptable to multiple areas since its factor is unique to each area. In either case, the other filters will continue to be considered for further analysis of the calculated speed profiles.

### 5.3 Speed calibration results

This section presents the design of experiments that have been planned and their results. The objective of this is to assess how the different proposed models behave under the different areas and filters.

In order to distinguish the experiments each of them has been labelled with a unique name of the form *AREA-SIZE\_FILTER\_MODEL* composed of three components:

- *AREA*: identifies the analyzed area. *BCN* for the Metropolitan Area of Barcelona or *BCN-R* for the reduced area of Barcelona between *Rondas*.
- *SIZE*: identifies the sample size. *1000* for 1,000 number of trips or *3000* for 3,000 number of trips.
- *FILTER*: identifies the distance filter. *MAXD-X* for the maximum absolute distance error filter with factor *X* km, *MAXDP-X* for the maximum absolute distance percentage error filter with factor *X*, *DYN-EUCL* for the Euclidean distance filter with a dynamic exclusive factor for each area and *NO-FILTER* to use the unfiltered sample.
- *MODEL*: identifies the model. *TT-SE* for the optimisation model based on total traveltimes or *ST-SE* for the optimisation model based on traveltimes in the segments.

The idea to also include the *NO-FILTER* as an option in the design of experiments is to analyse the effect of using a sample with erroneous information when calibrating the speed profiles.

With the filters previously selected, in the following table a summary of all the experiments considered is shown. Being a total of 24 experiments.

Moreover, since when using navigation services with real time data, the traveltime constantly varies according to the instant of the execution time, it implies that the results of the experiments would not be comparable among them when executed sequentially. For this reason, it is decided to set all HERE service requests with the same specific day and hour.

In addition, since the process of calculating the speed profiles is iterative and may not converge, it is decided to set the maximum number of iterations at 10. This is because it is a costly process, both in terms of execution time for each iteration (see further description in Section 5.4) and in terms of the cost of external service requests.

EXPERIMENT	AREA	SIZE	FILTER	MODEL
BCN-1000_MAXD-1_TT-SE	BCN	1000	MAXD-1	TT-SE
BCN-1000_MAXDP-0.005_ST-SE			MAXDP-0.005	ST-SE
BCN-1000_MAXDP-0.1_TT-SE			MAXDP-0.1	TT-SE
BCN-1000_DYN-EUCL_TT-SE			DYN-EUCL	
BCN-1000_NO-FILTER_ST-SE			NO-FILTER	ST-SE
BCN-1000_NO-FILTER_TT-SE				TT-SE
BCN-3000_MAXD-1_TT-SE		3000	MAXD-1	TT-SE
BCN-3000_MAXDP-0.005_ST-SE			MAXDP-0.005	ST-SE
BCN-3000_MAXDP-0.1_TT-SE			MAXDP-0.1	TT-SE
BCN-3000_DYN-EUCL_TT-SE			DYN-EUCL	
BCN-3000_NO-FILTER_ST-SE			NO-FILTER	ST-SE
BCN-3000_NO-FILTER_TT-SE				TT-SE
BCN-R-1000_MAXD-1_TT-SE	BCN-R	1000	MAXD-1	TT-SE
BCN-R-1000_MAXDP-0.005_ST-SE			MAXDP-0.005	ST-SE
BCN-R-1000_MAXDP-0.1_TT-SE			MAXDP-0.1	TT-SE
BCN-R-1000_DYN-EUCL_TT-SE			DYN-EUCL	
BCN-R-1000_NO-FILTER_ST-SE			NO-FILTER	ST-SE
BCN-R-1000_NO-FILTER_TT-SE				TT-SE
BCN-R-3000_MAXD-1_TT-SE		3000	MAXD-1	TT-SE
BCN-R-3000_MAXDP-0.005_ST-SE			MAXDP-0.005	ST-SE
BCN-R-3000_MAXDP-0.1_TT-SE			MAXDP-0.1	TT-SE
BCN-R-3000_DYN-EUCL_TT-SE			DYN-EUCL	
BCN-R-3000_NO-FILTER_ST-SE			NO-FILTER	ST-SE
BCN-R-3000_NO-FILTER_TT-SE				TT-SE

Table 5.31: Design of experiments considered

To estimate the effectiveness of the models, the MAE (Mean Absolute Error) and MAXAE (Maximum Absolute Error) in the traveltimes will be calculated. Being computed by comparing the traveltimes obtained from PTV xRoute service and the reference traveltimes obtained from HERE service.

The improvement of the calibration in terms of the MAE estimator will also be analysed. Comparing the MAE computed with the traveltimes calculated using the default speed profiles and the MAE computed from the traveltimes calculated using the calibrated speed profiles.

NC	MAX (KM/H)	MIN (KM/H)
0	134.4	70.0
1	122.0	38.0
2	105.0	30.0
3	100.0	25.0
4	95.0	20.0
5	50.0	10.0
6	19.2	3.8
7	10.0	3.0

Table 5.32: Initial speed profiles per networkclass in Barcelona

### 5.3.1 Area 1: Barcelona Metropolitan Area (BCN)

The first area to be analysed is the Barcelona Metropolitan Area (BCN). Using the two samples considered of 1,000 or 3,000 trips and the speed profiles currently used in the PTV xRoute service (see Table 5.32), the following statistics are obtained:

AREA	MAE (MIN)	MAXAE (MIN)
BCN-1000	2.99	10.52
BCN-3000	3.00	10.52

Table 5.33: Initial MAE and MAXAE with default speed profiles in BCN area

Now the mean deviation in traveltimes is 3 min for both cases and a deviation of 10.5 min in the worst case.

The tables 5.34 and 5.35 show the MAE and MAXAE estimator values for each experiment on each of its iterations where it can be seen:

- In all cases using a filtering method there has been a significant reduction by half in the mean deviation (from 3 to 1.5 min). And even the maximum deviation has also been decreased on average more than 1 min.

This largest reduction is achieved at the first iteration.

- It is clearly seen the importance of using a filter in the ST models as the deviations increase dramatically since many more errors are introduced in the generation of the data. In the opposite case, in the TT models there is not so much influence, since this error is only introduced at the origin or destination.

In addition, for the experiment of the unfiltered ST model there was a memory failure after the 2nd iteration. Since with the calibrated speed profiles, the sample increased to almost 1 million segments.

- The ST model achieves the worst results in the MAE. It obtains the minimum values in the MAXAE, although it worsens at the second iteration with the large sample.

Moreover, at each iteration normally no improvement is achieved, instead it gets worse.

- The DYN-EUCL filter with the TT model achieves the worst improvements in the MAXAE.
- The MAXD and MAXDP filters with the TT model get the best results in both the MAE and MAXAE.

To better compare all options, Appendix F.1.3 presents a comparison of the total traveltime errors before and after the calibration where it can be seen:

- Overall it appears as any of the options (with the exception of the unfiltered ST model) generate very good results.

With no calibration, the deviations was centered in -4 min. Indicating the PTV xRoute service less total traveltime as compared to the reference. And after a calibration it can be seen how the average becomes more centered in the 0 min of error. Also sharpening the error distribution.

- And on the other hand, it is also easily seen as the unfiltered ST model causes large errors in traveltimes. With deviations up to 90 more minutes.

As a final analysis of the calibration, the speed profiles calculated with each of the experiments at the first iteration can be seen in the Tables 5.36 and 5.37. If these are compared with the initial speeds of Barcelona (see Table 5.32), the following can be concluded:

- In the case of networkclass 1, this class had no representation because there was no segment of it. In the case of networkclasses 6 and 7, either there was no representation due to a total filtering of the segments, or there was a very minority representation for lack of data (less than 0.5%). For this reason, the calculated speeds can never be meaningful.
- In the contrary, in the case of the rest of networkclasses, it can be seen how a minimum representation of the segments provides good results. Although for networkclasses 0, 2, 3 and 4 the speeds have decreased according to the initial speed profiles.
- And clearly the speeds calculated with the ST models without using a filter do not make any sense. Since the speeds are very low.

For the full results over all iterations, Figures F.1 and F.2 show the variation of the calculated speed profiles. From which it can be deduced:

- The speeds remain approximately constant since the first iteration. As it was expected, because the MAE results were also constant.
- The speeds from networkclass 1, 6 and 7 are not very significant as they are not highly represented.

To finish and better explain the bad results of the ST models with the unfiltered samples, in Appendix F.1.2 the speeds distribution obtained with the HERE service depending on the networkclass and speedclass can be seen. From which it can be deduced:

- Initially the distributions of the speeds for each networkclass look like they show some differentiated classes. Being 0 the one that reaches higher speeds and 7 the one with lower speeds. Result that coincides with the hierarchy of importance of the networkclasses.
- However, in the distribution of the speeds for each speedclass depending on the networkclass it can be seen how these speeds do not satisfy the speedclass behaviour. In some cases the speeds are decreasing being 0 the fastest and 7 the slowest. But in most cases, these speeds vary quite a bit and do not follow that decreasing linear relationship.
- In addition, the speeds are quite scattered even after filtering. Implying that using all the traveltimes associated with these speeds produces worse results.  
Although it can also be seen how filtering reduces this dispersion. Therefore, the filtration is required if using ST models.

EXPERIMENT	MAE (MIN)									
	IT-1	IT-2	IT-3	IT-4	IT-5	IT-6	IT-7	IT-8	IT-9	IT-10
BCN-1000_MAXD-1_TT-SE	1.47	1.49	1.45	1.48	1.44	1.48	1.42	1.43	-	-
BCN-1000_MAXDP-0.005_ST-SE	1.64	1.91	1.81	1.78	1.79	-	-	-	-	-
BCN-1000_MAXDP-0.1_TT-SE	1.48	1.48	-	-	-	-	-	-	-	-
BCN-1000_DYN-EUCL_TT-SE	1.52	1.54	1.53	-	-	-	-	-	-	-
BCN-1000_NO-FILTER_ST-SE	23.35	24.38	22.09	23.99	23.11	25.55	21.71	24.44	23.76	22.76
BCN-1000_NO-FILTER_TT-SE	1.47	1.46	-	-	-	-	-	-	-	-
BCN-3000_MAXD-1_TT-SE	1.52	1.46	1.46	-	-	-	-	-	-	-
BCN-3000_MAXDP-0.005_ST-SE	1.59	1.79	1.79	-	-	-	-	-	-	-
BCN-3000_MAXDP-0.1_TT-SE	1.46	1.43	1.44	-	-	-	-	-	-	-
BCN-3000_DYN-EUCL_TT-SE	1.50	1.47	1.48	-	-	-	-	-	-	-
BCN-3000_NO-FILTER_ST-SE	23.39	x	x	x	x	x	x	x	x	x
BCN-3000_NO-FILTER_TT-SE	1.45	1.44	-	-	-	-	-	-	-	-

Table 5.34: MAE iterative results in BCN area

EXPERIMENT	MAXAE (MIN)									
	IT-1	IT-2	IT-3	IT-4	IT-5	IT-6	IT-7	IT-8	IT-9	IT-10
BCN-1000_MAXD-1_TT-SE	8.20	8.83	8.79	9.41	9.03	8.68	8.47	9.24	-	-
BCN-1000_MAXDP-0.005_ST-SE	7.92	8.05	8.00	7.91	7.93	-	-	-	-	-
BCN-1000_MAXDP-0.1_TT-SE	8.45	8.96	-	-	-	-	-	-	-	-
BCN-1000_DYN-EUCL_TT-SE	9.69	10.25	10.14	-	-	-	-	-	-	-
BCN-1000_NO-FILTER_ST-SE	89.22	95.98	84.72	106.61	88.93	107.10	93.00	103.97	92.38	99.05
BCN-1000_NO-FILTER_TT-SE	8.42	8.12	-	-	-	-	-	-	-	-
BCN-3000_MAXD-1_TT-SE	8.76	8.65	8.64	-	-	-	-	-	-	-
BCN-3000_MAXDP-0.005_ST-SE	8.65	8.99	8.99	-	-	-	-	-	-	-
BCN-3000_MAXDP-0.1_TT-SE	8.71	8.65	8.68	-	-	-	-	-	-	-
BCN-3000_DYN-EUCL_TT-SE	9.34	9.19	9.19	-	-	-	-	-	-	-
BCN-3000_NO-FILTER_ST-SE	96.57	x	x	x	x	x	x	x	x	x
BCN-3000_NO-FILTER_TT-SE	8.61	8.62	-	-	-	-	-	-	-	-

Table 5.35: MAXAE iterative results in BCN area

NC	TT_SE_MAXD1		ST_SE_MAXDP0.005		TT_SE_MAXDP0.1		TT_SE_DYN-EUCL		ST_SE_NO-FILTER		TT_SE_NO-FILTER	
	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN
0	115	52	111	60	118	53	127	53	27	9	117	53
1	12	5	12	5	12	5	12	5	12	5	12	5
2	74	66	78	65	74	66	78	70	35	6	74	66
3	65	24	78	24	58	24	66	24	29	13	60	24
4	75	18	68	19	75	18	71	18	53	17	75	18
5	51	10	39	15	53	9	52	6	36	15	53	9
6	150	143	12	5	37	29	150	143	12	5	150	143
7	150	143	12	5	150	143	25	17	12	5	150	143

Table 5.36: Speed profiles per networkclass computed from BCN-1000 sample

NC	TT_SE_MAXD1		ST_SE_MAXDP0.005		TT_SE_MAXDP0.1		TT_SE_DYN-EUCL		ST_SE_NO-FILTER		TT_SE_NO-FILTER	
	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN
0	119	51	111	60	119	53	131	51	26	8	119	52
1	12	5	12	5	12	5	12	5	12	5	12	5
2	74	66	79	61	73	65	76	68	37	5	75	67
3	60	24	78	24	53	24	56	24	31	13	56	24
4	73	18	68	19	76	18	67	18	53	17	75	18
5	49	9	37	15	52	9	52	7	44	12	50	9
6	12	5	12	5	12	5	150	142	12	5	12	5
7	150	142	12	5	61	53	150	142	12	5	150	143

Table 5.37: Speed profiles per networkclass computed from BCN-3000 sample

### 5.3.2 Area 2: Reduced Barcelona Metropolitan Area (BCN-R)

The second area to be analysed is the reduced Barcelona Metropolitan Area between *Rondas* (BCN-R). Using the two samples considered of 1,000 or 3,000 trips and the speed profiles currently used in the PTV xRoute service (see Table 5.32), the following statistics are obtained:

AREA	MAE (MIN)	MAXAE (MIN)
BCN-R-1000	2.42	8.49
BCN-R-3000	2.49	10.64

Table 5.38: Initial MAE and MAXAE with default speed profiles in BCN-R area

Now the mean deviation in traveltimes is of 2.5 min for both cases and a deviation between 8.5 and 10.5 min in the worst case.

It is possible to observe how initially the traveltime deviations have been reduced compared to those of the whole area, being that these speed profiles adapt better to the situation of the city of Barcelona.

The tables 5.39 and 5.40 show the MAE and MAXAE estimator values for each experiment on each of its iterations where it can be seen:

- In all cases there has been a significant reduction by half in the mean deviation (from 2.5 to 1.25 min). But maximum deviations do not always achieve a reduction. In the small sample this deviation is increased by about 2 min in most of the cases. And in the case of the large sample, it fluctuates by 1 min above or below.  
It can also be seen how the best results are obtained at the first iteration. Moreover, at each iteration normally they get worse.
- It is clearly seen the importance of using a filter in the ST models as the deviations increase dramatically since many more errors are introduced in the generation of the data. In the opposite case, in the TT models there is not so much influence, since this error is only introduced at the origin or destination.
- The ST model achieves the worst results in the MAE. However, it obtains the minimum values in the MAXAE. Improving it at each iteration performed, but worsening also a bit the MAE.
- The MAXD, MAXDP and DYN-EUCL filters with the TT model achieve the best results in the MAE. However, worsening the MAXAE. In the case of the small sample, it increases on average 2 min from the initial situation (from 8.5 to 10.5 min).

To better compare all options, Appendix F.2.3 presents a comparison of the total traveltime errors before and after the calibration where it can be seen:

- Overall it appears as any of the options (with the exception of the unfiltered ST model) generate very good results.  
With no calibration, the deviations were centered in -4 min. Indicating the PTV xRoute service less total traveltime as compared to the reference. And after a calibration it can be seen how the average becomes more centered in the 0 min of error. Also sharpening the error distribution.
- And on the other hand, it is also easily seen as the unfiltered ST model causes large errors in traveltimes. With deviations up to 40 more minutes.
- Compared to the results of the BCN area, it can be seen how the error distributions have become more sharpened. Implying that with a reduced area, the deviations in traveltimes can be even more minimized.

As a final analysis of the calibration, the speed profiles calculated with each of the experiments at the first iteration can be seen in the Tables 5.41 and 5.42. If these are compared with the initial speeds of Barcelona (see Table 5.32), the following can be concluded:

- In the case of networkclass 1, this class had no representation because there was no segment of it. In the case of networkclasses 6 and 7, either there was no representation due to a total filtering of the segments, or there was a very minority representation for lack of data (less than 0.5%). For this reason, the calculated speeds can never be meaningful.
- On the other hand, in the case of the rest of networkclasses, it can be seen how a minimum representation of the segments provides good results. Although for networkclasses 0, 2, 3 and 4 the speeds have decreased according to the initial speed profiles.
- And clearly the speeds calculated with the ST models without using a filter do not make any sense. Since the speeds are very low.

For the full results over all iterations, Figures F.13 and F.14 show the variation of the calculated speed profiles. From which it can be deduced:

- The speeds remain approximately constant since the first iteration. As was to be expected, because the MAE results were also constant.
- The speeds from networkclass 1, 6 and 7 are not very significant as all are not highly represented.

To finish and better explain the bad results of the ST models with the unfiltered samples, in Appendix F.2.2 the speeds distribution obtained with the HERE service depending on the networkclass and speedclass can be seen. From which it can be deduced:

- Initially the distributions of the speeds for each networkclass look like they show some differentiated classes. Being 0 the one that reaches higher speeds and 7 the one with

lower speeds. Result that coincides with the hierarchy of importance of the network-classes.

- However, in the distribution of the speeds for each speedclass depending on the networkclass it can be seen how these speeds do not satisfy the speedclass behaviour. In some cases the speeds are decreasing being 0 the fastest and 7 the slowest. But in most cases, these speeds vary quite a bit and do not follow that decreasing linear relationship.
- In addition, the speeds are quite scattered even after filtering. Implying that using all the traveltimes associated with these speeds produces worse results. Although it can also be seen how filtering reduces this dispersion. Therefore, the filtration is required if using ST models.

EXPERIMENT	MAE (MIN)									
	IT-1	IT-2	IT-3	IT-4	IT-5	IT-6	IT-7	IT-8	IT-9	IT-10
BCN-R-1000_MAXD-1_TT-SE	1.22	1.20	1.19	-	-	-	-	-	-	-
BCN-R-1000_MAXDP-0.005_ST-SE	1.20	1.25	1.25	-	-	-	-	-	-	-
BCN-R-1000_MAXDP-0.1_TT-SE	1.18	1.19	-	-	-	-	-	-	-	-
BCN-R-1000_DYN-EUCL_TT-SE	1.22	1.22	-	-	-	-	-	-	-	-
BCN-R-1000_NO-FILTER_ST-SE	6.35	6.18	7.00	7.58	7.45	7.32	7.51	7.36	7.59	7.37
BCN-R-1000_NO-FILTER_TT-SE	1.18	1.18	-	-	-	-	-	-	-	-
BCN-R-3000_MAXD-1_TT-SE	1.19	1.26	1.23	1.21	1.23	1.23	-	-	-	-
BCN-R-3000_MAXDP-0.005_ST-SE	1.21	1.28	1.28	-	-	-	-	-	-	-
BCN-R-3000_MAXDP-0.1_TT-SE	1.20	1.23	1.19	1.18	-	-	-	-	-	-
BCN-R-3000_DYN-EUCL_TT-SE	1.21	1.22	-	-	-	-	-	-	-	-
BCN-R-3000_NO-FILTER_ST-SE	6.70	5.96	7.24	6.94	7.42	6.89	7.11	6.95	7.63	7.90
BCN-R-3000_NO-FILTER_TT-SE	1.20	1.19	-	-	-	-	-	-	-	-

Table 5.39: MAE iterative results in BCN-R area

EXPERIMENT	MAXAE (MIN)									
	IT-1	IT-2	IT-3	IT-4	IT-5	IT-6	IT-7	IT-8	IT-9	IT-10
BCN-R-1000_MAXD-1_TT-SE	10.61	10.57	10.57	-	-	-	-	-	-	-
BCN-R-1000_MAXDP-0.005_ST-SE	9.17	8.72	8.72	-	-	-	-	-	-	-
BCN-R-1000_MAXDP-0.1_TT-SE	10.57	10.60	-	-	-	-	-	-	-	-
BCN-R-1000_DYN-EUCL_TT-SE	9.92	10.63	-	-	-	-	-	-	-	-
BCN-R-1000_NO-FILTER_ST-SE	43.71	28.01	50.74	45.80	50.59	53.02	49.98	52.53	50.66	49.31
BCN-R-1000_NO-FILTER_TT-SE	8.38	8.52	-	-	-	-	-	-	-	-
BCN-R-3000_MAXD-1_TT-SE	9.27	10.66	10.71	11.28	11.40	11.40	-	-	-	-
BCN-R-3000_MAXDP-0.005_ST-SE	9.23	8.74	8.69	-	-	-	-	-	-	-
BCN-R-3000_MAXDP-0.1_TT-SE	10.43	11.51	11.54	11.49	-	-	-	-	-	-
BCN-R-3000_DYN-EUCL_TT-SE	10.32	11.50	-	-	-	-	-	-	-	-
BCN-R-3000_NO-FILTER_ST-SE	45.98	31.62	50.34	53.21	61.07	53.21	58.41	56.56	59.12	57.18
BCN-R-3000_NO-FILTER_TT-SE	9.77	9.64	-	-	-	-	-	-	-	-

Table 5.40: MAXAE iterative results in BCN-R area

NC	TT_SE_MAXD1		ST_SE_MAXDP0.005		TT_SE_MAXDP0.1		TT_SE_DYN-EUCL		ST_SE_NO-FILTER		TT_SE_NO-FILTER	
	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN
0	150	53	125	53	124	57	144	53	49	5	150	51
1	12	5	12	5	12	5	12	5	12	5	12	5
2	31	23	107	16	27	19	35	27	40	10	66	33
3	122	21	122	21	122	21	122	21	22	13	122	21
4	27	18	27	19	30	18	31	18	24	16	27	19
5	36	11	38	14	38	11	39	11	36	14	37	10
6	150	142	12	5	14	6	12	5	12	5	150	143
7	25	17	12	5	15	7	12	5	12	5	23	15

Table 5.41: Speed profiles per networkclass computed from BCN-R-1000 sample

NC	TT_SE_MAXD1		ST_SE_MAXDP0.005		TT_SE_MAXDP0.1		TT_SE_DYN-EUCL		ST_SE_NO-FILTER		TT_SE_NO-FILTER	
	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN
0	147	52	126	54	138	53	138	54	50	5	150	51
1	12	5	12	5	12	5	12	5	12	5	12	5
2	119	18	105	17	94	19	33	25	32	10	98	25
3	122	21	122	21	122	21	122	21	21	13	122	21
4	29	19	27	19	29	19	33	18	24	16	27	19
5	32	12	39	13	31	12	32	13	32	14	33	12
6	13	5	12	5	16	8	12	5	12	5	13	5
7	47	39	12	5	35	27	16	8	12	5	33	25

Table 5.42: Speed profiles per networkclass computed from BCN-R-3000 sample

### 5.3.3 Conclusions

Taking into account all the previous results of the analysis of the calibration results, it can be concluded that for models based on segment traveltimes (ST models) the filtering method is essential.

It is seen that using the complete sample obtained directly from the reference service, the results are a disaster. Because when using many intermediate waypoints, a large number of errors are introduced in between.

In contrast, it can be observed that if these errors are filtered out, the results obtained can be comparable to those of models based on total traveltimes (TT models). Even with a slight maximum deviation values lower than in TT models.

On the other hand, it is also seen as for the TT models, the filtering is not crucial since similar results are obtained in almost all their experiments. This is because with this method, it is only used as a reference the total traveltimes of the shortest path between the origin and destination. And therefore, error are only introduced at these two points.

Although the analysis of the distances suggested that without a filtering method, the samples would not be comparable. These results show that this is not entirely true.

Despite that, it is recommended to carry out some kind of filtering, at least a very slight one. Because although in the evaluated area there have not been obtained very high deviations in the total distances, it is not possible to predict if in a new area the same situation will occur.

## 5.4 Time performance

One concern of this system was its performance. Since it is a calibration of speed profiles with real-time traffic data, it is required to make multiple requests to both services (PTV xRoute and HERE) with the consequent performance penalty for each request.

Therefore, this consideration has been taken into account throughout the development to minimise as possible this execution time.

Initially all phases were implemented iteratively, but there was a big penalty in the execution times when all the requests were made, sequentially one after the other. Therefore, it have been decided to implement in parallel the processes detected as critical in performance, in order to maximise the machine potential.

Moreover, in order to obtain performance results comparable between them, a single machine has been set for all experiments as specified below:

- Processor Intel Core i5-8250U CPU
  - 4 cores, 8 threads
  - 1.60-3.40GHz

- RAM Memory of 8 GB DDR3
- SSD Hard disk of 256 GB

Furthermore, when dealing with WEB services, these critical factors can be identified that can affect the measurements of execution times:

- *Internet connection*: executing experiments in different locations with a different bandwidth can make it slower the multiple requests to be executed. Implying the measurements not comparable. For this reason, all these executions have been carried out under the same conditions (location and bandwidth).
- *Totality of CPU*: the machine selected to perform the experiments may not assign the totality of the CPU for the execution of our system, assigning a portion of it to some other processes in background related to the operative system. In this case, as it cannot be avoided, if some executions with large differences between the mean values are detected they will be discarded for the estimations of the execution times.
- *Routing services*: both PTV xRoute and HERE are shared services. That implies if their servers have more peak load when our experiments are being run, our measurements of execution times may also vary. In this case, it is not avoidable and if large differences are detected between the mean values, those experiments will be discarded for the estimations of execution times.

The following tables show a summary of the measured execution times for each of the main processes depending on the sample size, obtained with the mean value observed between all the valid executions of the experiments.

PROCEDURE	EXECUTION TIME (MIN)			
	1,000 TRIPS		3,000 TRIPS	
	SEQ	PAR	SEQ	PAR
initial xRoute data generation	12.56	2.59	39.97	6.20
initial HERE data generation	4.89	0.60	13.82	2.53
speed calibration data processing per iteration	1.30	1.17	2.81	3.16
speed calibration computation per iteration	0.03	0.08	0.64	0.65
xRoute data re-generation per iteration	12.65	2.58	41.25	6.36
HERE data re-generation per iteration	-	-	-	-
initial execution time	17.45	3.19	53.79	8.73
execution time per iteration	13.98	3.83	44.70	10.17

Table 5.43: Time performance with TT models depending on the sample size

PROCEDURE	EXECUTION TIME (MIN)			
	1,000 TRIPS		3,000 TRIPS	
	SEQ	PAR	SEQ	PAR
initial xRoute data generation	12.74	2.71	38.29	6.18
initial HERE data generation	22.38	2.29	68.13	7.02
speed calibration data processing per iteration	2.55	2.94	5.02	5.60
speed calibration computation per iteration	0.12	0.14	0.37	0.34
xRoute data re-generation per iteration	12.59	2.82	37.88	6.43
HERE data re-generation per iteration	6.71	0.86	13.63	1.76
initial execution time	35.12	5.00	106.42	13.20
execution time per iteration	21.97	6.76	56.90	14.13

Table 5.44: Time performance with ST models depending on the sample size

The first table shows the results using a model based on the total traveltimes of trips, while the second one is from a model based on the traveltimes in the segments. From these the following conclusions can be drawn:

- The PTV xRoute data generation time is the same in any case as it requires the same information at trip-level (coordinates, total traveltime and total distance) and segment-level (coordinates, traveltime, distance, networkclass and speed).  
In this case, these times are higher than HERE's because it is necessary to request all the information from the trips and also computing the speedclasses from the speeds with the used speedstable (see methodology described in the Appendix B).
- The HERE data generation time is reduced using total traveltime-based models. Since in this case, it is not required the information segment-related and implies much less requests to this API. As with long routes, multiple calls to the API have to be made to obtain all the information of the segments (see Section 5.1.3).  
In the case of trip-based models, the measurement of data re-generation times is not applicable because the total traveltime for the fixed day and hour is the same. And only for segment-based models it is necessary to re-calculate this process for the trips with some variation in their route at each new iteration.
- The speed calibration data processing time is also increased in the case of segment traveltime-based models. Since the data size is much larger using the segments information than with the trips only (3,000 trips contain approximately more than 650,000 segments).
- The speed calibration computation time varies differently depending on the model used. It can be seen how with few data the model based on totaltravel times is faster. However, with the larger dataset, these times increase up to twice compared to the other model.

Overall it can be seen how implementing all data generation processes has significantly reduced the execution times. To measure this improvement the Speedup estimator can be used which measures the ratio between sequential and parallel execution time:

$$S_p = \frac{T_1}{T_p} \quad (5.2)$$

Code 5.2: Speedup formula.

Below is a summary table with the speedup achieved using eight threads on each process of the data generation (the parallel processes) for the different samples and models used. Resulting in the worst case a speedup of 4.5.

	SPEEDUP			
	TT MODELS		ST MODELS	
	1,000 TRIPS	3,000 TRIPS	1,000 TRIPS	3,000 TRIPS
initial xRoute data generation	4.85	6.45	4.70	6.20
initial HERE data generation	8.15	5.46	9.77	9.71
xRoute data re-generation per iteration	4.9	6.48	4.46	5.89
HERE data re-generation per iteration	-	-	7.80	7.88

Table 5.45: Speedup per procedure depending on the sample size and model

## 5.5 Validation

At this moment the PTV xRoute service uses unique static speed profiles for each area. This means that independently of the instant of the trip, the computed traveltimes are always the same.

With this calibration system, it is attempted to replicate the existing variability throughout the day. Previously it has been seen how the mean error between PTV xRoute and HERE has been minimised given the same instant of time.

Therefore, as a validation method it is decided to verify whether with this calibration it is really possible to reproduce this daily variability.

For this purpose, the calibration system for different hours as the departure hours of the trips will be executed and the mean traveltime for those executions will be calculated. In addition, to verify that the variability is accurate, it will be compared with the mean traveltimes obtained from HERE and will be used as a reference.

Furthermore, due to resource limitations and computational costs, this experiment will be only performed for each 2h interval with a sample size of 1,000 trips.

As a result, the following figures show the mean traveltimes from HERE and the initial and calibrated PTV xRoute service for each of the tested areas. As it can be seen, without any calibration the traveltime is always the same giving also values below the real situation. By contrast, with the calibration procedure it replicates the existing variability quite well.

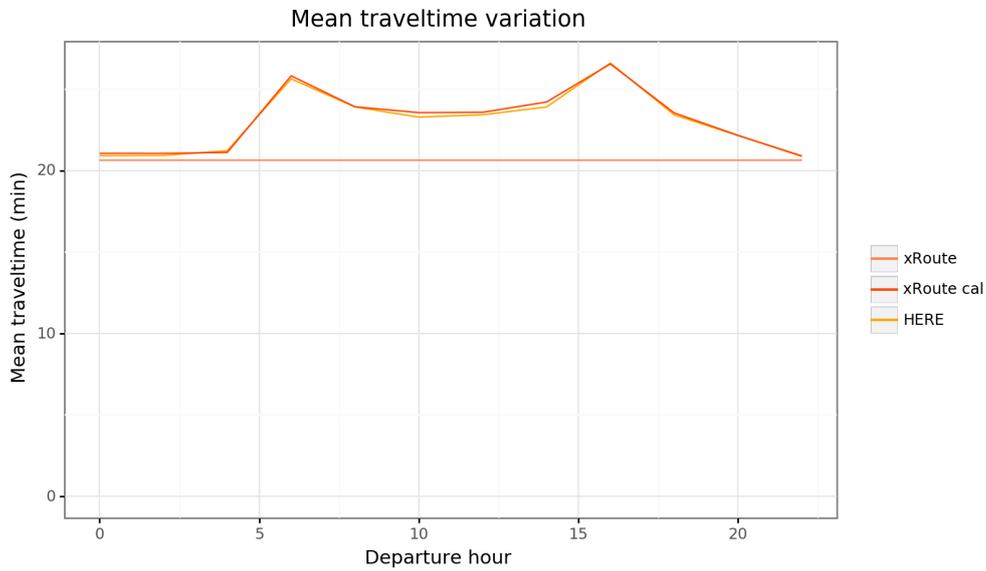


Figure 5.24: Comparison of mean traveltime variation throughout the day with PTV xRoute and HERE services in BCN area

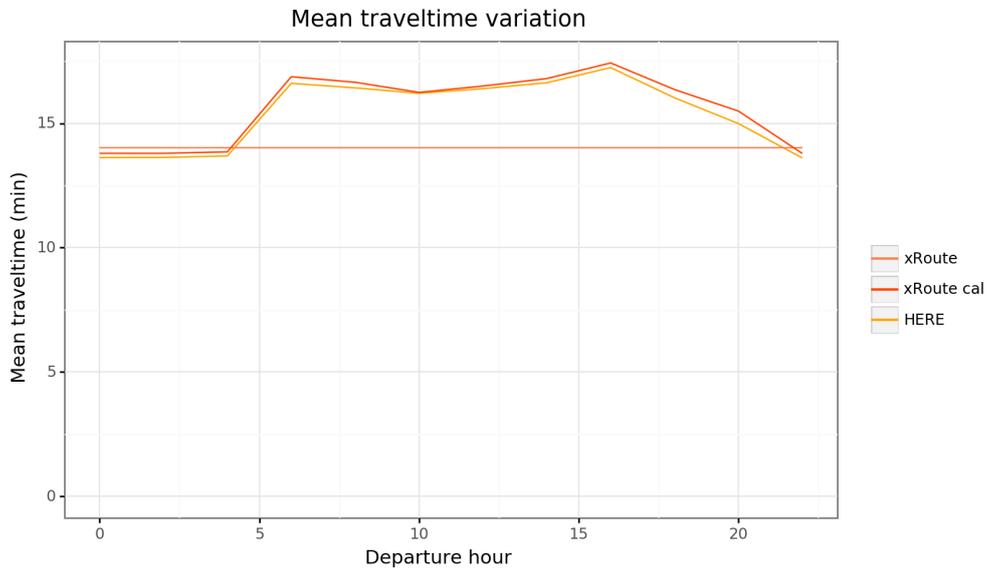


Figure 5.25: Comparison of mean traveltime variation throughout the day with PTV xRoute and HERE services in BCN-R area

Specifically, this mean traveltime is increased during peak hours. Since it is when the traffic

## 5.5. VALIDATION

---

is more dense due to the increased number of trips going to work in the morning and from work in the afternoon.

Furthermore, in the reduced area as they are shorter routes, these mean traveltimes are decreased by about 8 min compared to the entire area of Barcelona.

As a summary of this variability, the following tables show the MAE depending on the departure hour. Before, it can be seen as the mean error in the traveltimes at peak hours increases a lot as this variability could not be replicated. And after the calibration, although the errors in peak hours slightly increase, it manages to replicate the variability much better.

	DEPARTURE HOUR											
	0	2	4	6	8	10	12	14	16	18	20	22
Before	1.58	1.60	1.80	4.76	3.29	2.98	3.76	4.52	4.11	2.76	1.97	1.58
After	1.03	1.12	1.11	2.10	1.54	1.52	1.92	2.14	1.95	1.48	1.21	1.03

Table 5.46: MAE comparison before and after calibration at different departure hours in BCN area

	DEPARTURE HOUR											
	0	2	4	6	8	10	12	14	16	18	20	22
Before	1.20	1.12	1.12	2.65	2.53	2.44	2.90	3.33	3.00	2.06	1.38	1.11
After	1.01	0.89	0.88	1.26	1.19	1.22	1.41	1.54	1.43	1.20	1.12	0.86

Table 5.47: MAE comparison before and after calibration at different departure hours in BCN-R area

With such results, it can be determined that the existing daily variability on these samples has been replicated a bit. However, it is still necessary to carry out independent tests with real routes data. However, due to difficulties in accessing this kind of data because of related costs or of confidentiality issues with PTV Group clients, this validation could not be performed.

# Chapter 6

## Conclusions

### 6.1 Achievement of objectives

The main objective of this project was to develop an automatic calibration system for the speed profiles of the PTV xRoute service. This system was intended to facilitate the work currently being done manually and to improve the service provided to customers.

Once this project has been completed, this objective has been satisfactorily achieved. In order to quantify it, it will be evaluated in terms of the objectives stated in Section 1.3.2:

- An automated system has been developed to generate reference data from an external data provider and to integrate it with the PTV xRoute service data. Obtaining the required input data for the calibration component of the speed profiles.
- Different optimisation models have been implemented for minimising the deviations in traveltimes. Two types of models have been proposed, based on a different level of detail of the reference data: one based on trip traveltimes (TT models) and the other based on segment traveltimes (ST models).

In addition, an in-depth analysis of the implications of using each of the two sample formats has been covered in Section 5.2. Being problematic the process to obtain the segment traveltimes as errors are introduced at each intermediate point.

However, as it has been demonstrated, filtering such deviations properly allows to obtain similar results as those of the TT models.

However, not all the proposed objectives could be fully tested:

- From the beginning it has been specified that the prototype experiments would be carried out in the Barcelona area. With which good results have been obtained. However, as it has not been possible to study other areas, it cannot be estimated how effective its application to multiple areas is.
- This system has been designed in such a way that the external data provider can easily be changed if the need arises. Using basic KPIs (traveltime and distance) as reference data that could be obtained from any data provider. However, it has not

been tested with any other provider other than HERE due to a lack of resources or lack of accessibility to other data sources.

- In addition, due to the difficulty of accessibility to real trip data either due to cost issues or due to confidentiality issues of PTV Group clients. It has not been possible to carry out a more in-depth validation to estimate the effectiveness of the proposed solution.

## 6.2 Contributions

The most significant contribution is providing a system for the automatic calibration of PTV xRoute service speed profiles. As explained at the beginning of the document (Section 1.3.1), a such tool developed in this project was not yet available and it required a manual calibration of this parameter.

In addition, it studies the applicability of using data from historical traffic data sources as the reference data for the calibration process. Usually this problem is solved using Floating Car Data (FCD). However, the intention of this system was to evaluate whether this new type of reference data which is more easily accessible worldwide can be used to calibrate the speed profiles.

In addition it has been proved the importance of a filtering of the generated data between both routing services. However, this influence is only significant for ST models as more errors are introduced at each intermediate point. And in TT models these are only introduced at the origin or destination.

## 6.3 Further research

Although this project has satisfactorily fulfilled the initial objectives, new research lines have also been opened to improve the results of this system.

The refinement is one of the most important processes in the entire system. The errors introduced by integrating data from two different sources are not avoidable and therefore require a proper refinement to filter out those differences.

By analysing the results it is possible to identify the existing relationship between these filters and the effectiveness of the optimisation models. That is, if the information is not comparable, the added errors will worsen the results. Consequently, a new design of experiments could be used to explore the best factors to further minimise the deviations obtained.

Some other questions have also been opened about the nature of the sample used. In the current case, the routes are evenly distributed over a specific area, containing all kinds of routes. Although an imbalance in the distribution over the hierarchy of segments to be

studied, the networkclasses, has been detected.

It has been encountered how there was a lacking in the samples of an entire networkclass and others in a very minority compared to the rest. This obviously had implications on the results since these classes were not properly represented and thus could not be correctly calculated.

This raises interesting concerns about alternative methods of data generation for avoiding this imbalance in networkclass proportions. As well as to study how a minimum representation of all networkclasses could be obtained. And to study the impact if a more balanced-networkclass data is used in our system.

## 6.4 Personal assessment

This has been the first big project that I have had to develop from scratch in a business environment and it has allowed me to learn the whole process that has to be followed to make an initial proposal and its subsequent development.

In addition, it allowed me to design a complete system from the data generation from different providers and their integration to its application in optimisation models to calculate speed profiles that minimise the deviations between the traveltimes from both data sources.

Furthermore, it has allowed me to study the implications of using data from different networks and the possible errors introduced when using different coordinate systems. Being able to study also different approaches in order to detect these errors on the samples and refine them.

Lastly, if I had more time and resources, I would have liked to study the impact of more types of data filters and to implement more optimisation models to explore other possibilities. Improvements that have been explained in Section 6.3.



# References

- [1] (Jan. 2017). The rise of mobility as a service. Reshaping how urbanites get around, [Online]. Available at: <https://www2.deloitte.com/insights/us/en/deloitte-review/issue-20/smart-transportation-technology-mobility-as-a-service.html> (Last accessed: 04/30/2019).
- [2] A. TAYLOR. (Jun. 2018). What are the Benefits of Mobility as a Solution for Commuters?, [Online]. Available at: <https://www.businessmaas.com/apps/what-are-the-benefits-of-mobility-as-a-solution-for-commuters/> (Last accessed: 04/30/2019).
- [3] The MaaS Accelerator Program, [Online]. Available at: <https://www.ptvgroup.com/en/mobilitynext/accelerator-program/> (Last accessed: 04/30/2019).
- [4] PTV Visum, [Online]. Available at: <https://www.ptvgroup.com/en/solutions/products/ptv-visum/> (Last accessed: 04/30/2019).
- [5] PTV xServer, [Online]. Available at: <https://www.ptvgroup.com/en/solutions/products/ptv-xserver/> (Last accessed: 04/30/2019).
- [6] PTV xServer portafolio, [Online]. Available at: <http://xserver.ptvgroup.com/au/products/portfolio/> (Last accessed: 04/30/2019).
- [7] PTV Maps&Data, [Online]. Available at: <https://www.ptvgroup.com/en/solutions/products/ptv-maps-data/> (Last accessed: 04/30/2019).
- [8] PTV Map&Guide, [Online]. Available at: <https://www.ptvgroup.com/en/solutions/products/ptv-mapandguide/> (Last accessed: 04/30/2019).
- [9] R. SEVLIAN and R. RAJAGOPAL, “Travel time estimation using floating car data”, Dec. 2010.
- [10] B. Y. CHEN, H. YUAN, Q. LI, W. H. LAM, S.-L. SHAW, and K. YAN, “Map-matching algorithm for large-scale low-frequency floating car data”, *International Journal of Geographical Information Science*, vol. 28, no. 1, pp. 22–38, 2014. [Online]. Available at: <https://doi.org/10.1080/13658816.2013.816427>.
- [11] N. COHN and H. BISCHOFF. (2012). Floating car data for transportation planning. Explorative study to technique and applications and sample properties of GPS data.
- [12] T. JESKE, “Floating Car Data from Smartphones: What Google and Waze Know About You and How Hackers Can Control Traffic”,

## REFERENCES

---

- [13] “Kolmogorov-Smirnov Test”, in *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, pp. 283–287. [Online]. Available at: [https://doi.org/10.1007/978-0-387-32833-1\\_214](https://doi.org/10.1007/978-0-387-32833-1_214).
- [14] E. FEIGELSON and G. J. BABU. Beware the Kolmogorov-Smirnov test!, [Online]. Available at: <https://asaip.psu.edu/Articles/beware-the-kolmogorov-smirnov-test> (Last accessed: 05/21/2019).
- [15] J. M. del CASTILLO GRANADOS, “Modelos de distancias”, in. *Curso Básico de Ingeniería del Transporte*, 2006, ch. 1, pp. 17–21.
- [16] “Minkowski Distance”, [Online]. Available at: <https://www.sciencedirect.com/topics/computer-science/minkowski-distance>.
- [17] Minkowski distance, [Online]. Available at: [https://en.wikipedia.org/wiki/Minkowski\\_distance](https://en.wikipedia.org/wiki/Minkowski_distance) (Last accessed: 05/21/2019).
- [18] R. F. LOVE and J. H. WALKER, “Facilities Location: Models and Methods”, 1988.

# Appendix

## A Routing APIs examples

This section will show an example of the routing APIs that have been used in this project. First with the PTV xRoute API for an specific origin-destination coordinates, and then with the HERE API reusing the response coordinates of the previous one to match the same route as close as possible.

### A.1 PTV xRoute API - Calculate Route

Code A.1: PTV xRoute API - Calculate Route request example

```
1 {
2   "waypoints": [
3     {
4       "$type": "OffRoadWaypoint",
5       "location": {
6         "offRoadCoordinate": {
7           "x": 2.117159,
8           "y": 41.389025
9         }
10      }
11    },
12    {
13      "$type": "OffRoadWaypoint",
14      "location": {
15        "offRoadCoordinate": {
16          "x": 2.116186,
17          "y": 41.390327
18        }
19      }
20    }
21  ],
22  "resultFields": {
23    "segments": {
24      "enabled": True,
25      "descriptors": True
```

## A. ROUTING APIS EXAMPLES

---

```
26     },
27     "nodes": True
28 },
29 "storedProfile": "car.xml",
30 "requestProfile": {
31     "vehicleProfile": {
32         "speeds": {
33             "maximumSpeed": max_speed,
34             "speedRangesByNetworkClass": {
35                 "maximumSpeeds": maximumSpeeds,
36                 "minimumSpeeds": minimumSpeeds
37             }
38         }
39     }
40 }
41 }
```

Code A.2: PTV xRoute API - Calculate Route response example

```
1 {
2     "$type": "RouteResponse",
3     "distance": 315.0,
4     "travelTime": 58.71,
5     "nodes": [
6         {
7             "coordinate": {"x": 2.1171589782, "y": 41.389025018},
8             "type": "INPUT_NODE"
9         },
10        {
11            "coordinate": {"x": 2.1171510372, "y": 41.389036933},
12            "type": "LINK_NODE"
13        },
14        {
15            "coordinate": {"x": 2.1171545853, "y": 41.389038201},
16            "type": "NETWORK_NODE"
17        },
18        ...
19        {
20            "coordinate": {"x": 2.1166787995, "y": 41.390336197},
21            "type": "NETWORK_NODE"
22        },
23        {
24            "coordinate": {"x": 2.1161866247, "y": 41.390317691},
25            "type": "LINK_NODE"
26        },
27    ]
28 }
```

```
27     {
28         "coordinate": {"x": 2.1161859489, "y": 41.390326944},
29         "type": "INPUT_NODE"
30     }
31 ],
32 "segments": [
33     {
34         "distance": 1.0,
35         "travelSpeed": 15.0,
36         "travelTime": 0.24,
37         "startNodeIndex": 0,
38         "endNodeIndex": 1,
39         "type": "LINK_SEGMENT"
40     },
41     {
42         "attributes": {
43             "descriptors": {
44                 "country": "ES",
45                 "networkClass": 4,
46                 "roadName": "Carrer de Jordi Girona"
47             }
48         },
49         "distance": 0.0,
50         "travelTime": 0.0,
51         "travelTime": 3.78,
52         "startNodeIndex": 1,
53         "endNodeIndex": 2,
54         "type": "NETWORK_SEGMENT"
55     },
56     ...
57     {
58         "attributes": {
59             "descriptors": {
60                 "country": "ES",
61                 "networkClass": 5,
62                 "roadName": "Carrer de Dulcet"
63             }
64         },
65         "distance": 42.0,
66         "travelSpeed": 16.0,
67         "travelTime": 9.45,
68         "startNodeIndex": 8,
69         "endNodeIndex": 9,
70         "type": "NETWORK_SEGMENT"
71     },
```

## A. ROUTING APIS EXAMPLES

---

```
72     {
73         "distance": 1.0,
74         "travelSpeed": 15.0,
75         "travelTime": 0.24,
76         "startNodeIndex": 9,
77         "endNodeIndex": 10,
78         "type": "LINK_SEGMENT"
79     }
80 ]
81 }
```

## A.2 HERE Routing API - Calculate Route

Code A.3: HERE Routing API - Calculate Route request example

```
1 {
2     "app_id": HERE_APP_ID,
3     "app_code": HERE_APP_CODE,
4     "waypoint0": "41.389036933,2.1171510372",
5     ...
6     "waypoint8": "41.390317691,2.1161866247"
7     "mode": "fastest;car;traffic:enabled",
8     "departure": "now",
9     "metricSystem": "metric",
10    "routeAttributes": "sh"
11 }
```

Code A.4: HERE Routing API - Calculate Route response example

```
1 {
2     "response": {
3         "route": [
4             {
5                 "leg": [
6                     {
7                         "length": 0,
8                         "travelTime": 0,
9                         "end": {
10                            "label": "Carrer de Jordi Girona",
11                            "mappedPosition": {
12                                "latitude": 41.3891029,
13                                "longitude": 2.1173787
14                            },
15                            "mappedRoadName": "Carrer de Jordi Girona",
16                            "originalPosition": {
```

```
17         "latitude": 41.389103,
18         "longitude": 2.1173708
19     }
20 },
21 "start": {
22     "label": "Carrer de Jordi Girona",
23     "mappedPosition": {
24         "latitude": 41.3890278,
25         "longitude": 2.1171427
26     },
27     "mappedRoadName": "Carrer de Jordi Girona",
28     "originalPosition": {
29         "latitude": 41.3890381,
30         "longitude": 2.1171546
31     }
32 }
33 },
34 ...
35 {
36     "length": 42,
37     "travelTime": 18,
38     "end": {
39         "label": "Carrer de Dulcet",
40         "mappedPosition": {
41             "latitude": 41.3903181,
42             "longitude": 2.1161865
43         },
44         "mappedRoadName": "Carrer de Dulcet",
45         "originalPosition": {
46             "latitude": 41.3903176,
47             "longitude": 2.1161866
48         }
49     },
50     "start": {
51         "label": "Passeig dels Til.lers",
52         "mappedPosition": {
53             "latitude": 41.3903368,
54             "longitude": 2.1166813
55         },
56         "mappedRoadName": "Passeig dels Til.lers",
57         "originalPosition": {
58             "latitude": 41.3903362,
59             "longitude": 2.1166787
60         }
61     }
62 }
```

## A. ROUTING APIS EXAMPLES

---

```
62     }
63   ],
64   "shape": [
65     "41.3890278,2.1171427",
66     "41.3891029,2.1173787",
67     ...
68     "41.3903368,2.1166813",
69     "41.3903181,2.1161865"
70   ],
71   "summary": {
72     "baseTime": 72,
73     "distance": 317,
74     "trafficTime": 92,
75     "travelTime": 92
76   },
77 }
78 ]
79 }
80 }
```

## B Speedclass computation

The speedclass associated to each segment of the PTV network is not provided by the PTV xRoute service. Thus, this must be obtained in another way such as through the distribution of the speeds within the area analysed (see Table B.1).

NC	MIN SC	MAX SC	MIN SPEED	MAX SPEED
0	2	7	70	116
1	2	7	38	98
2	0	7	30	105
3	0	7	25	100
4	0	7	20	95
5	0	7	10	50
6	1	6	6	17
7	1	6	4	9

Table B.1: Default speeds distribution per networkclass example in Barcelona.

With this table that indicates the speedclass and the speed ranges assigned to each networkclass, it can be computed the speedstable of the area itself with a linear interpolation (see Table B.2).

NC/SC	0	1	2	3	4	5	6	7
0	134.4	125.2	116.0	106.8	97.6	88.4	79.2	70.0
1	122.0	110.0	98.0	86.0	74.0	62.0	50.0	38.0
2	105.0	94.3	83.6	72.9	62.1	51.4	40.7	30.0
3	100.0	89.3	78.6	67.9	57.1	46.4	35.7	25.0
4	95.0	84.3	73.6	62.9	52.1	41.4	30.7	20.0
5	50.0	44.3	38.6	32.9	27.1	21.4	15.7	10.0
6	19.2	17.0	14.8	12.6	10.4	8.2	6.0	3.8
7	10.0	9.0	8.0	7.0	6.0	5.0	4.0	3.0

Table B.2: Default speeds table example in Barcelona.

This way the speedclass associated to a segment given its networkclass and speed can be computed directly from the speedstable.

## C Euclidean distance factor

This section will explain the basis behind the usage of the Euclidean distance metric in an urban network. As well as the description of the approximation factor on this distance metric to detect if the distances collected disagree.

Many different types of metrics can be defined for distances. However, only those that approximate the real situation and topography of the network to be evaluated have any practical usage [15].

But these metrics for representing an approximation of the real distance measures, they need to satisfy the following mathematical properties [16]:

- Non-negativity: the distance of two distinct points is positive.
- Identity: the distance of a point to itself is 0.
- Symmetry: the distance between two points  $P_1$  and  $P_2$  is the same as the distance between  $P_2$  and  $P_1$ .
- Triangular inequality: the distance between two points  $P_1$  and  $P_2$  is less or equal to the distance between  $P_1$  and  $P_2$  via a third point  $P_3$ .

In this manner, the generalisation of Minkowski distances (known as  $L_p$ ) allows to model different distance metrics between two points [15, 17]. In the case of a space of an order 2, the distance generalisation  $d(P_1, P_2)$  between two points  $P_1$  and  $P_2$  with coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$  can be defined with the following formula. Being the parameter  $p$  that allows modelling different types of distances, where  $p \geq 1$  must meet to satisfy the triangular inequality property.

This way, when  $p = 1$  the Manhattan metric is obtained that models distances in an area with grid-shaped streets (as in the Barcelona's Eixample). And when  $p = 2$  the so-called Euclidean metric is obtained, which models the distances as a straight line (the diagonal).

$$d(P_1, P_2) = (|x_1 - x_2|^p + |y_1 - y_2|^p)^{1/p}$$

Code C.1: Distance metric generalisation.

However, the main problem in the field of logistics and transportation is that it is not always possible to have the distance matrices between all locations in an area. Either because the cartography is not yet digitalised or because, although it is available, the required data involves a certain budget or computational costs that are not admissible in the project.

In these cases, the actual distances in the area are usually approximated by an inflated  $L_p$  metric with a  $\beta$  factor [15]. Which can be seen below.

$$d(P_1, P_2) = \beta (|x_1 - x_2|^p + |y_1 - y_2|^p)^{1/p}$$

Code C.2: Inflated distance metric generalisation.

As a consequence, this metric along with the  $\beta$  inflation factor takes into account the deviation between the actual distances and the distances from the metric. Being in the case of the euclidean distance metric, this  $\beta$  factor will take into account the existing deviation from the straight line between two points.

With this distance model it is possible to approximate the real distances only by knowing the coordinates. Being this information easier to obtain as opposed to the real distances where it is required to know the cartography of the area (e.g. direction of circulation, shortest path...).

However, for this approximation to be effective and cover the real distances as closely as possible, it should be calculated on a representative set of locations for which the real distances are known. Being the value of  $\beta$  the one that minimises the error between the euclidean distances and the real distances of the set:

$$d(P_1, P_2) = \beta \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Code C.3: Inflated euclidean distance metric.

Accordingly to some studies that have been carried out to determine the  $\beta$  factors for different areas, the following table shows a summary of some values found [15, 18]. As seen, in the case of Spain this factor takes as value 1.269 implying that the real distances are usually 26.9% greater than the straight line distance.

<b>COUNTRY</b>	$\beta$	<b>CITY</b>	$\beta$
Austria	1.353	Brussels	1.0744
Belgium	1.195	London	1.1168
Canada	1.332	Los Angeles	1.1237
USA	1.181	New York	1.1248
Spain	1.269	Paris	1.1118
France	1.218	Sydney	1.2314
UK	1.144	Tokyo	1.1094
Italy	1.310	Toronto	1.2028

Table C.1: Estimated  $\beta$  values for the euclidean distance metric model

However, it should be highlighted that these values have not been obtained with a digitalised cartography and may therefore differ from the real values. For this reason, Section 5.2.2.5: Dynamic euclidean distance filter (DYN-EUCL) presents a method for calculating these values more accurately using digital maps through navigation services. And therefore, obtaining more reliable values for each area evaluated.

## D Area coverage

This section shows a visualization of the segments coverage on each networkclass for both analyzed areas.

### D.1 Area 1: Barcelona Metropolitan Area (BCN)

#### D.1.1 Sample BCN-1000: BCN area with 1,000 trips

Segments coverage with a small sample of 1,000 trips on Barcelona area.



Figure D.1: Barcelona area coverage of Networkclass-0 with 1,000 trips



Figure D.2: Barcelona area coverage of Networkclass-1 with 1,000 trips



Figure D.3: Barcelona area coverage of Networkclass-2 with 1,000 trips



Figure D.4: Barcelona area coverage of Networkclass-3 with 1,000 trips

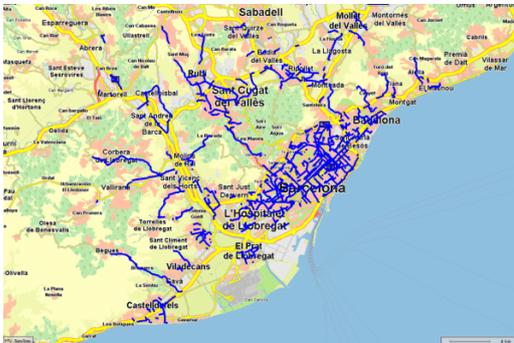


Figure D.5: Barcelona area coverage of Networkclass-4 with 1,000 trips

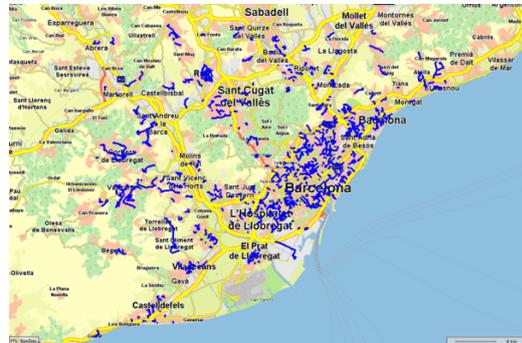


Figure D.6: Barcelona area coverage of Networkclass-5 with 1,000 trips



Figure D.7: Barcelona area coverage of Networkclass-6 with 1,000 trips



Figure D.8: Barcelona area coverage of Networkclass-7 with 1,000 trips

### D.1.2 Sample BCN-3000: BCN area with 3,000 trips

Segments coverage with a big sample of 3,000 trips on Barcelona area.



Figure D.9: Barcelona area coverage of Networkclass-0 with 3,000 trips



Figure D.10: Barcelona area coverage of Networkclass-1 with 3,000 trips



Figure D.11: Barcelona area coverage of Networkclass-2 with 3,000 trips



Figure D.12: Barcelona area coverage of Networkclass-3 with 3,000 trips

## D. AREA COVERAGE

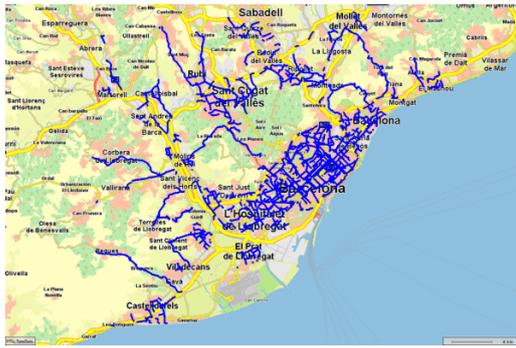


Figure D.13: Barcelona area coverage of Networkclass-4 with 3,000 trips

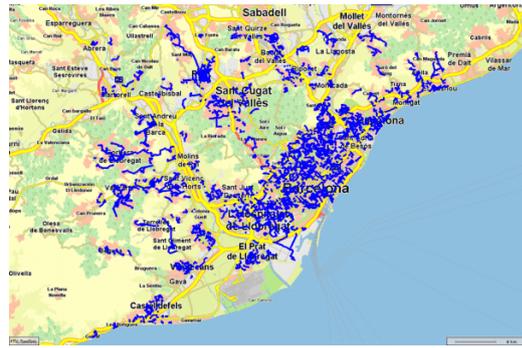


Figure D.14: Barcelona area coverage of Networkclass-5 with 3,000 trips



Figure D.15: Barcelona area coverage of Networkclass-6 with 3,000 trips



Figure D.16: Barcelona area coverage of Networkclass-7 with 3,000 trips

## D.2 Area 2: Reduced Barcelona Metropolitan Area (BCN-R)

### D.2.1 Sample BCN-R-1000: BCN-R area with 1,000 trips

Segments coverage with a small sample of 1,000 trips on reduced Barcelona area.



Figure D.17: Barcelona reduced area coverage of Networkclass-0 with 1,000 trips



Figure D.18: Barcelona reduced area coverage of Networkclass-1 with 1,000 trips



Figure D.19: Barcelona reduced area coverage of Networkclass-2 with 1,000 trips



Figure D.20: Barcelona reduced area coverage of Networkclass-3 with 1,000 trips

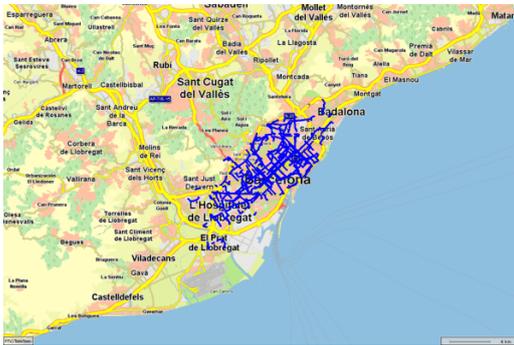


Figure D.21: Barcelona reduced area coverage of Networkclass-4 with 1,000 trips

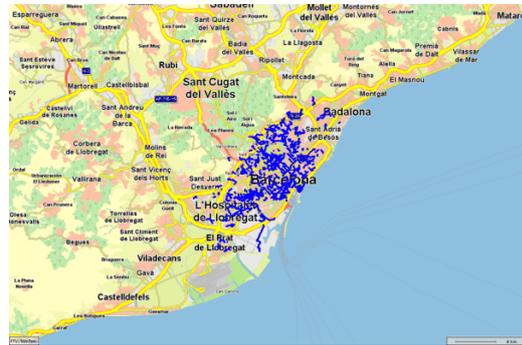


Figure D.22: Barcelona reduced area coverage of Networkclass-5 with 1,000 trips



Figure D.23: Barcelona reduced area coverage of Networkclass-6 with 1,000 trips



Figure D.24: Barcelona reduced area coverage of Networkclass-7 with 1,000 trips

### D.2.2 Sample BCN-R-3000: BCN-R area with 3,000 trips

Segments coverage with a big sample of 3,000 trips on reduced Barcelona area.

## D. AREA COVERAGE



Figure D.25: Barcelona reduced area coverage of Networkclass-0 with 3,000 trips



Figure D.26: Barcelona reduced area coverage of Networkclass-1 with 3,000 trips



Figure D.27: Barcelona reduced area coverage of Networkclass-2 with 3,000 trips

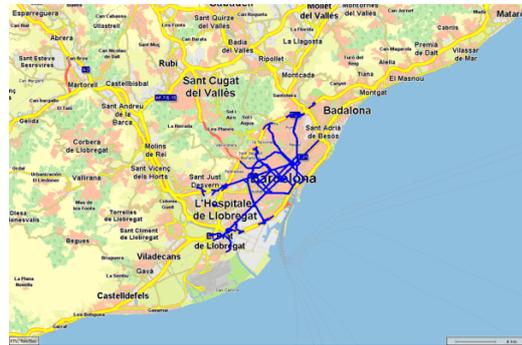


Figure D.28: Barcelona reduced area coverage of Networkclass-3 with 3,000 trips

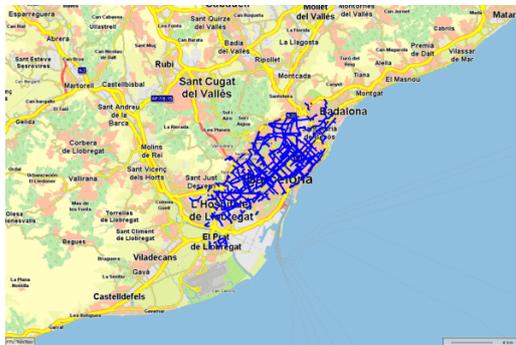


Figure D.29: Barcelona reduced area coverage of Networkclass-4 with 3,000 trips

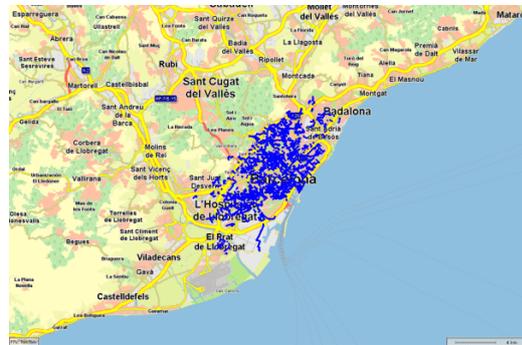


Figure D.30: Barcelona reduced area coverage of Networkclass-5 with 3,000 trips



Figure D.31: Barcelona reduced area coverage of Networkclass-6 with 3,000 trips



Figure D.32: Barcelona reduced area coverage of Networkclass-7 with 3,000 trips

## **E Distance analysis**

This section presents the visualizations of the results of the analysis of the distances obtained from the PTV xRoute and HERE services. First an initial analysis of the raw distances and then for each filter used in the system.

For the segment analysis, the distance distributions and the distance error distributions are shown.

For the trip analysis, the distance distributions and the corresponding KS tests are shown.

## E.1 Segment distance analysis

### E.1.1 Initial analysis

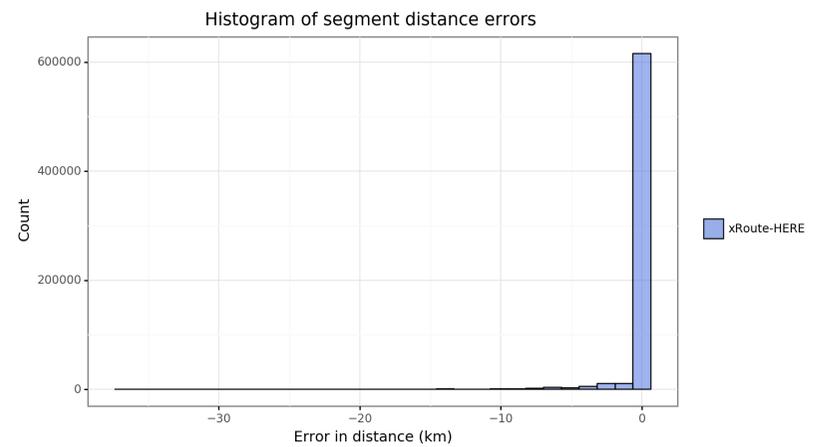
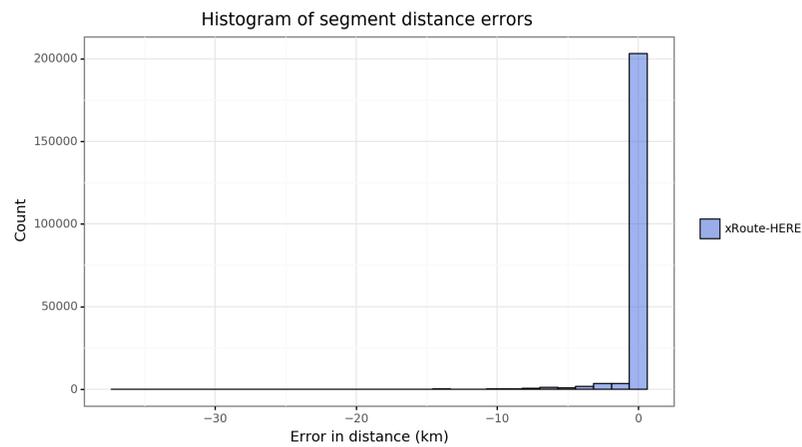
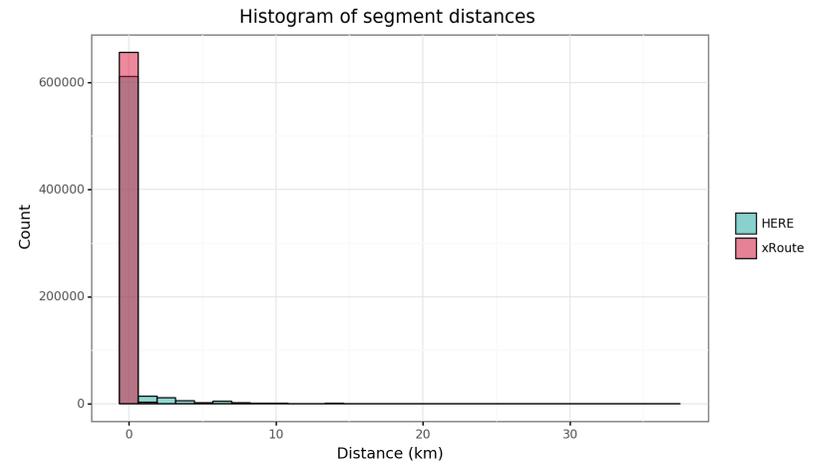
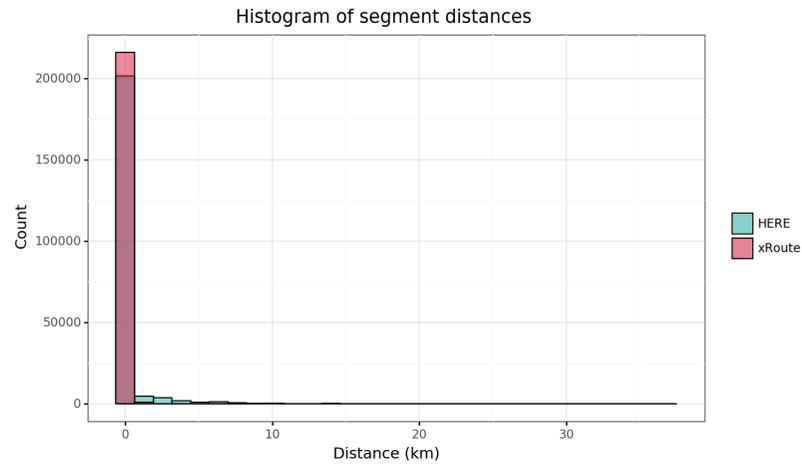


Figure E.1: Segment distance distr. and distance error distr. for BCN-1000 without filter

Figure E.2: Segment distance distr. and distance error distr. for BCN-3000 without filter

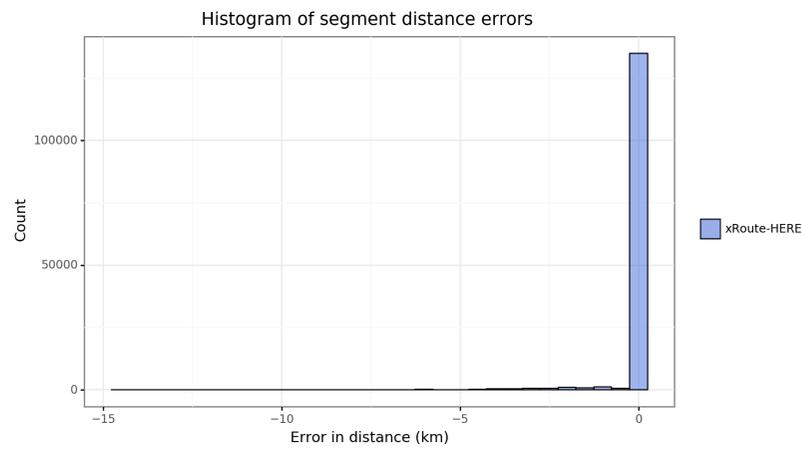
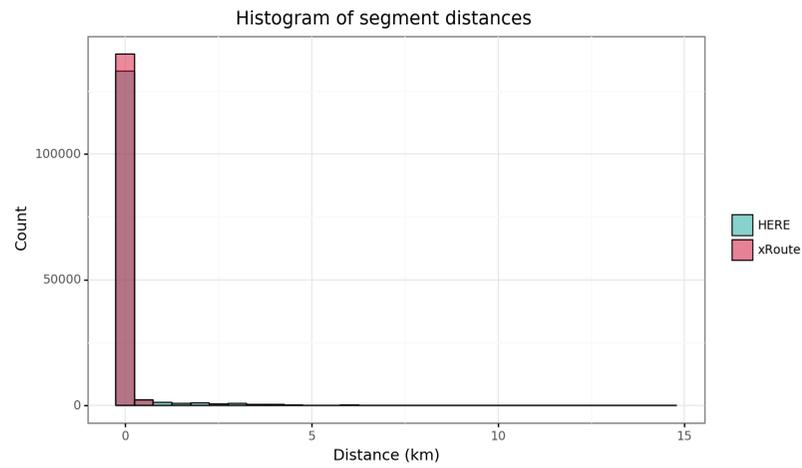


Figure E.3: Segment distance distr. and distance error distr. for BCN-1000 without filter

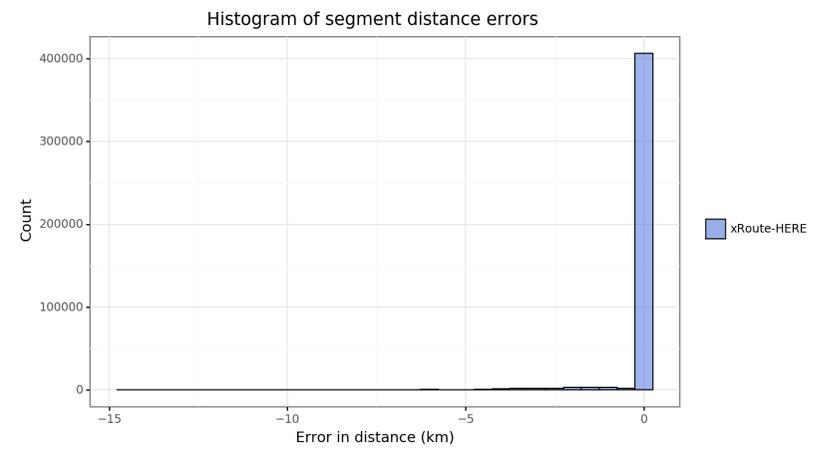
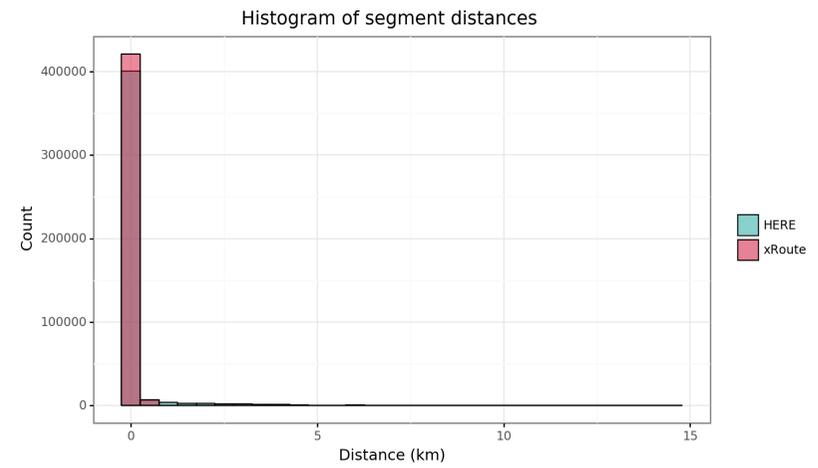
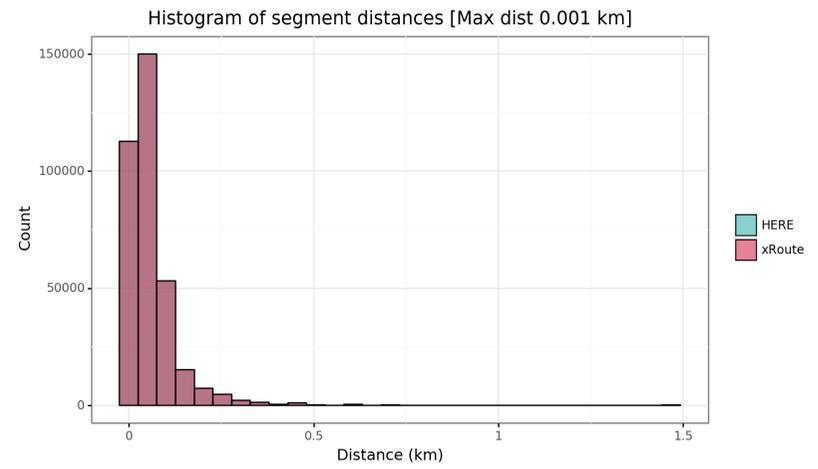
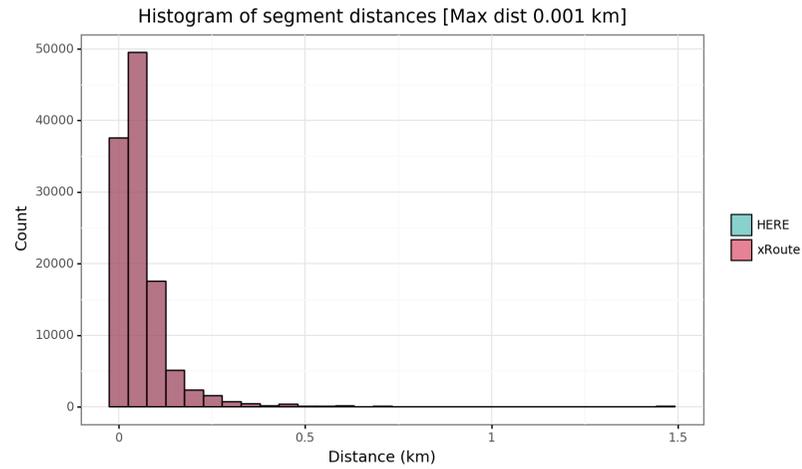


Figure E.4: Segment distance distr. and distance error distr. for BCN-R-3000 without filter

### E.1.2 Maximum absolute distance error filter (MAXD)



FOI

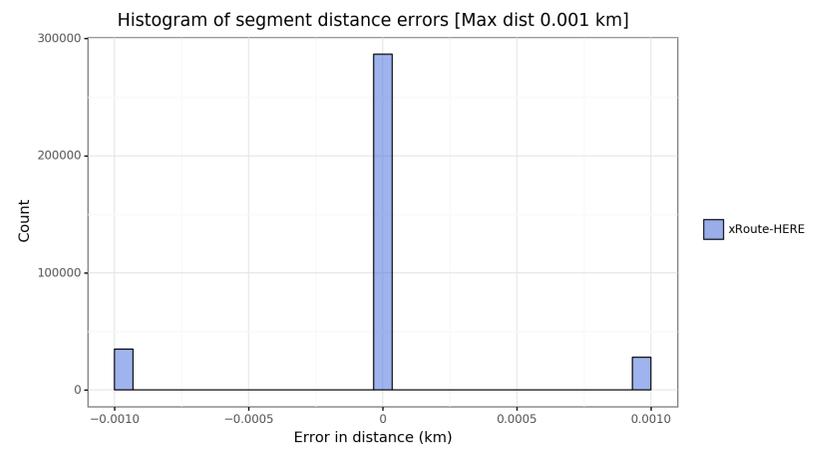
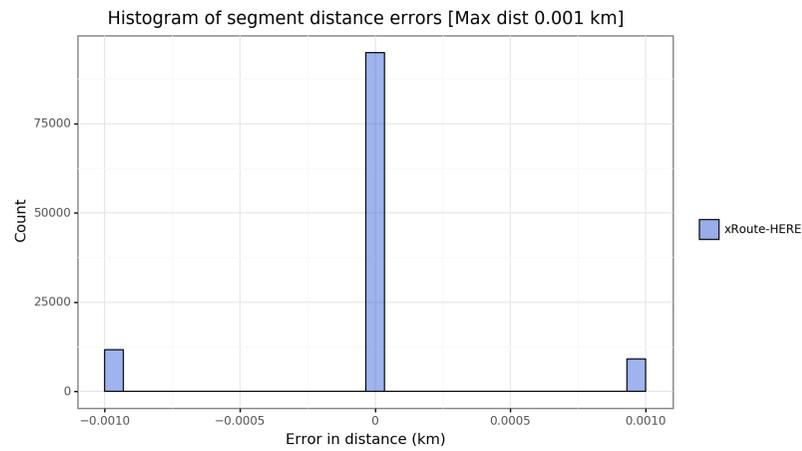


Figure E.5: Segment distance distr. and distance error distr. for BCN-1000 with MAXD-0.001 filter

Figure E.6: Segment distance distr. and distance error distr. for BCN-3000 with MAXD-0.001 filter

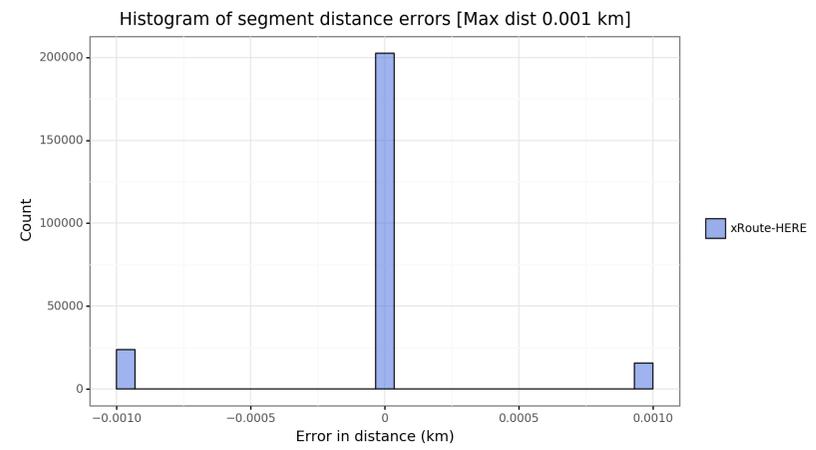
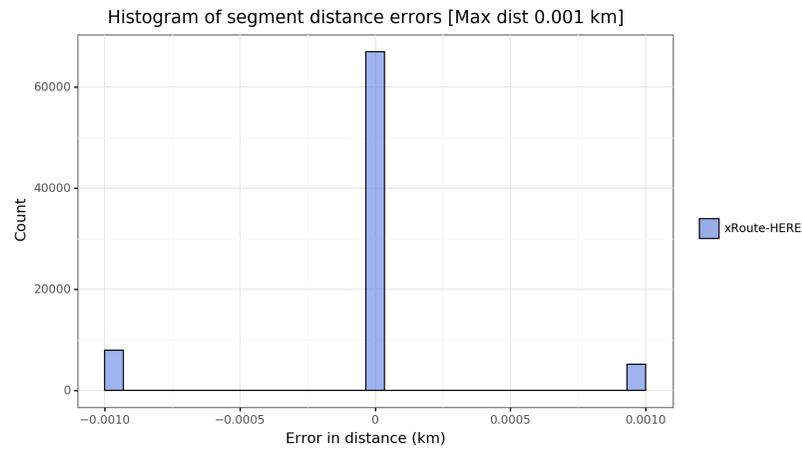
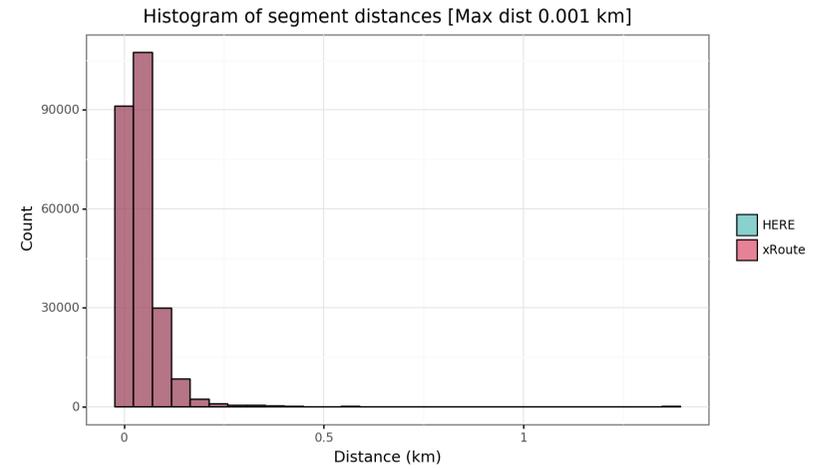
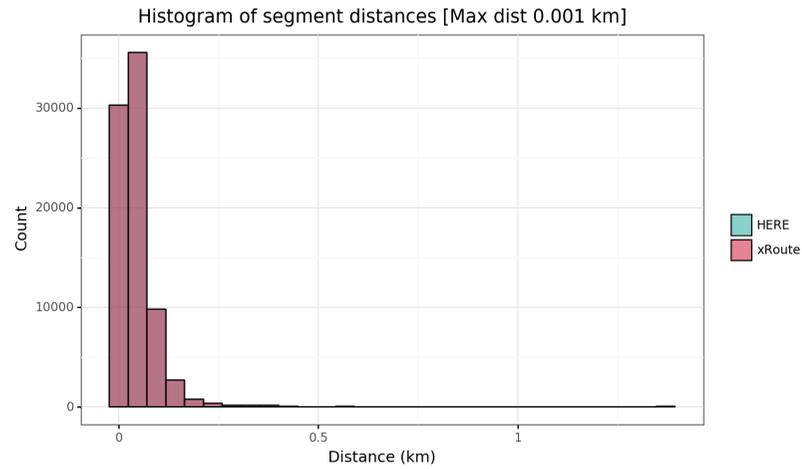


Figure E.7: Segment distance distr. and distance error distr. for BCN-R-1000 with MAXD-0.001 filter

Figure E.8: Segment distance distr. and distance error distr. for BCN-R-3000 with MAXD-0.001 filter

**E.1.3 Maximum absolute distance percentage error filter (MAXDP)**

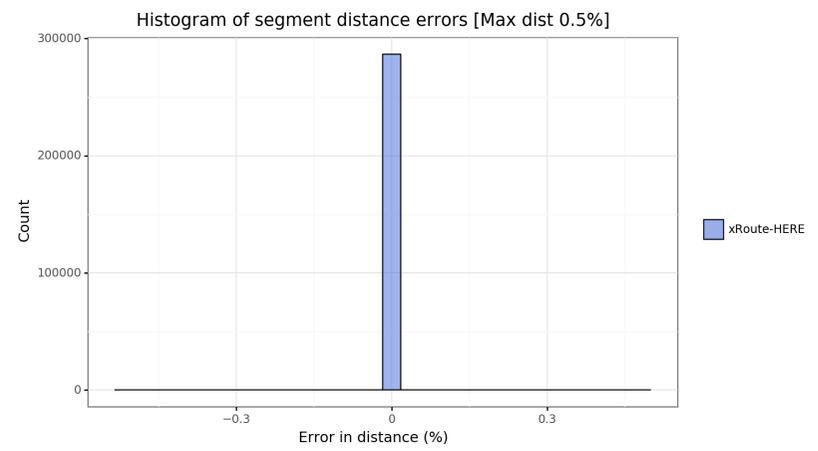
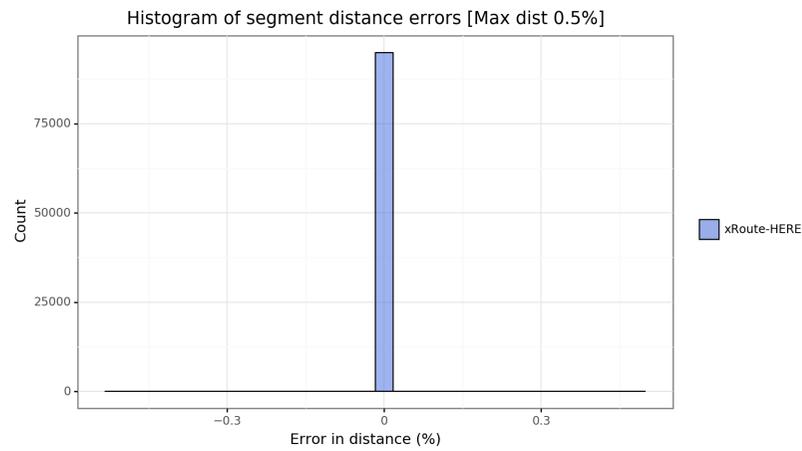
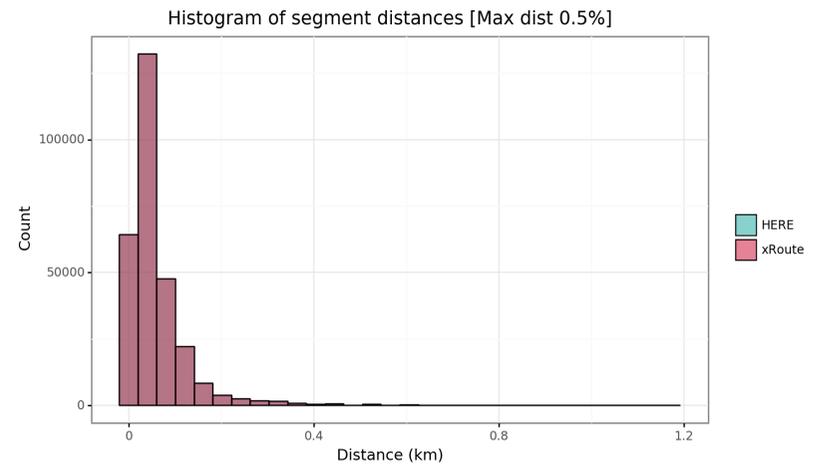
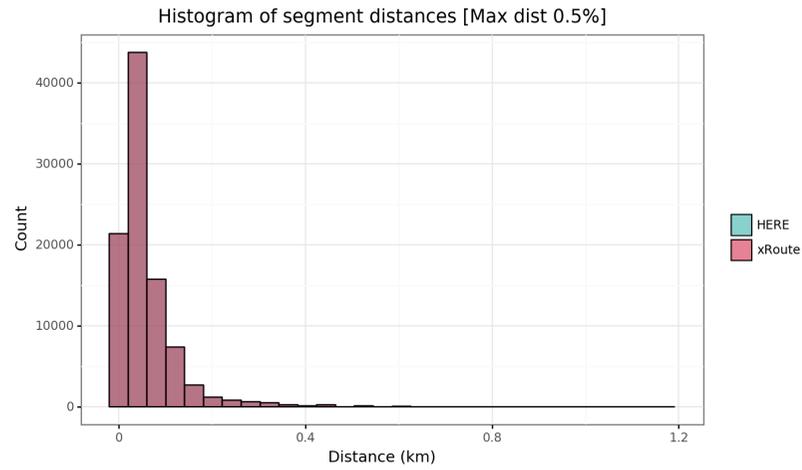


Figure E.9: Segment distance distr. and distance percentage error distr. for BCN-1000 with MAXDP-0.5 filter

Figure E.10: Segment distance distr. and distance percentage error distr. for BCN-3000 with MAXDP-0.5 filter

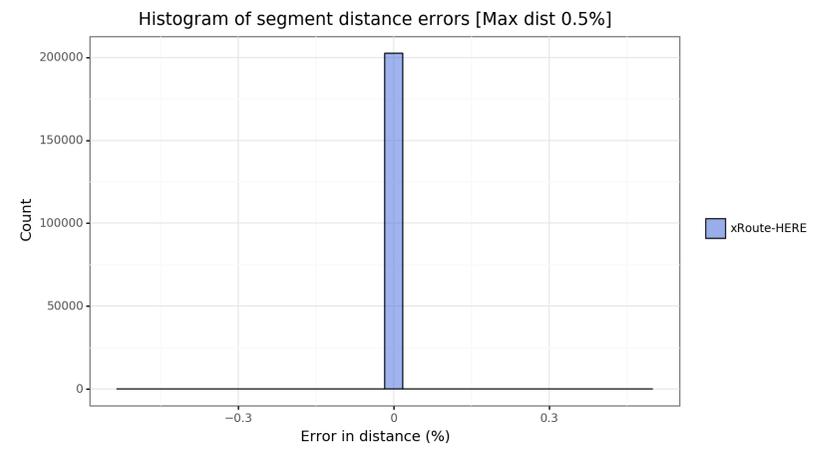
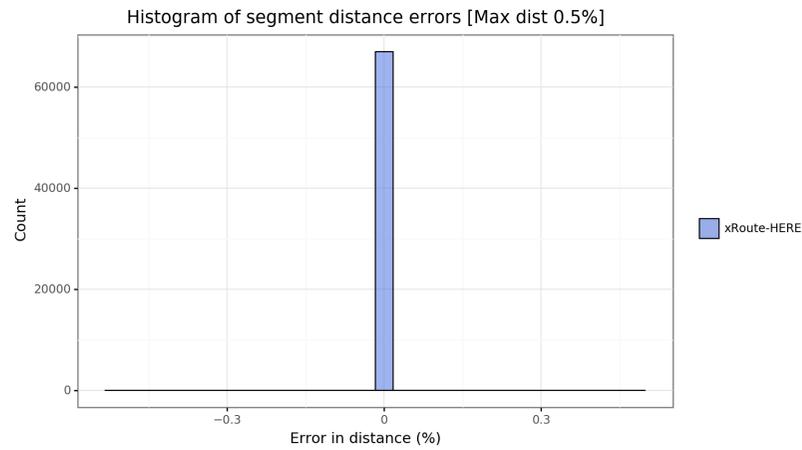
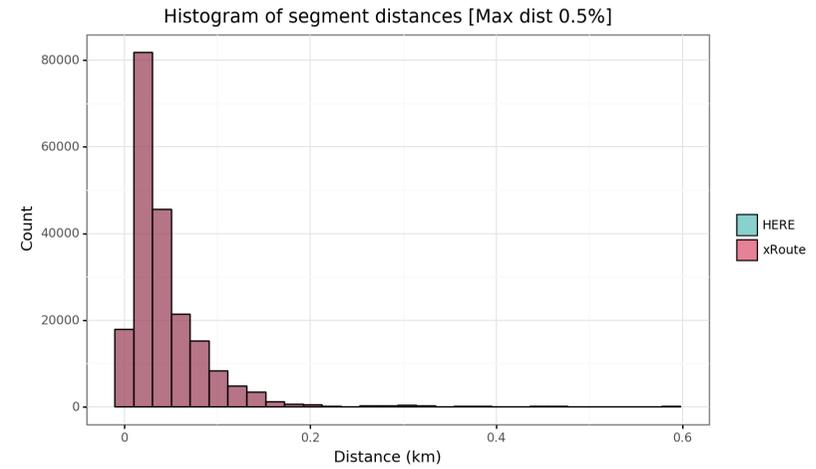
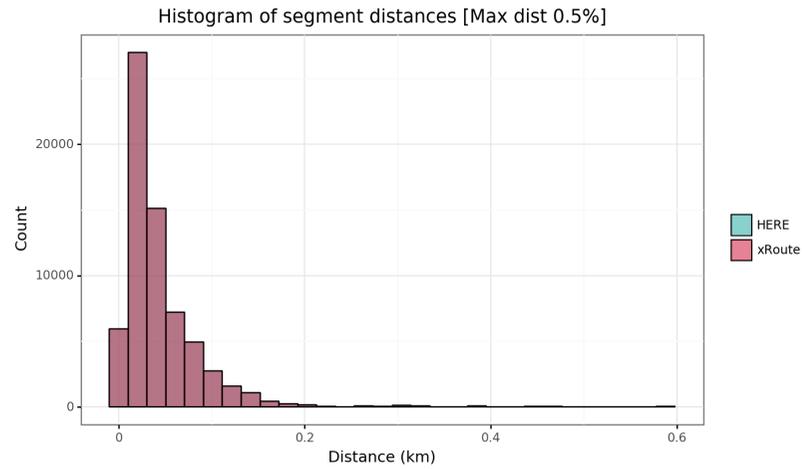


Figure E.11: Segment distance distr. and distance percentage error distr. for BCN-R-1000 with MAXDP-0.5 filter

Figure E.12: Segment distance distr. and distance percentage error distr. for BCN-R-3000 with MAXDP-0.5 filter

## E.2 Trip distance analysis

### E.2.1 Initial analysis

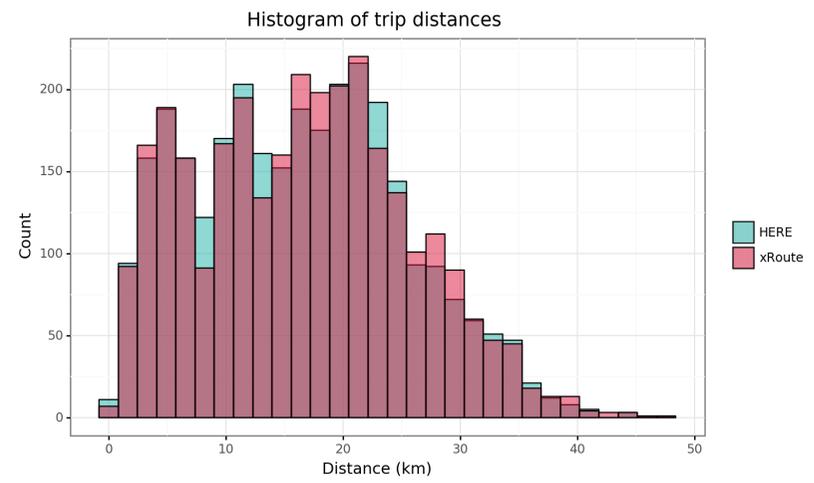
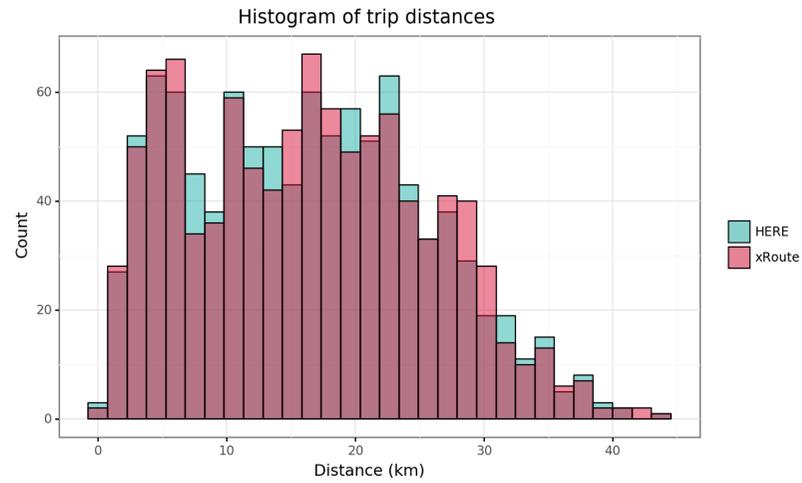
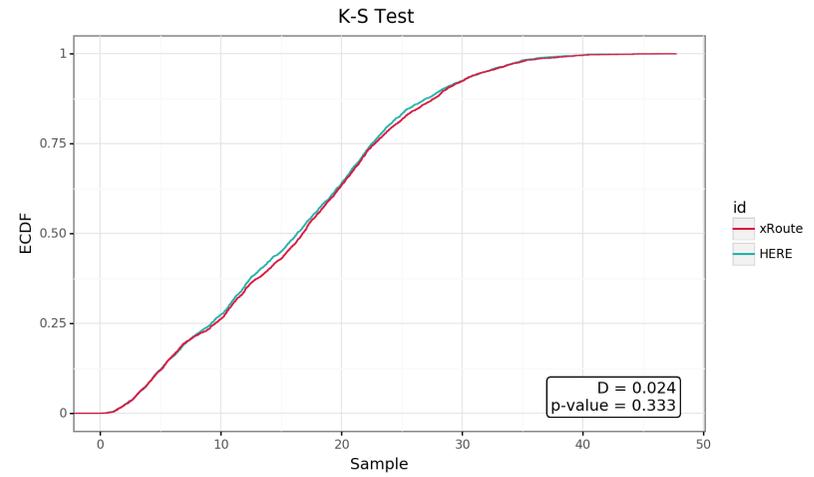
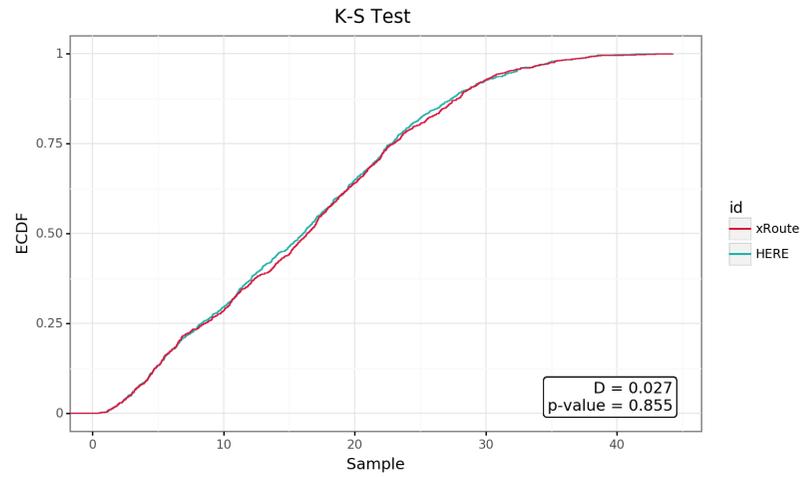


Figure E.13: KS test and trip distance distr. for BCN-1000 without filter

Figure E.14: KS test and trip distance distr. for BCN-3000 without filter

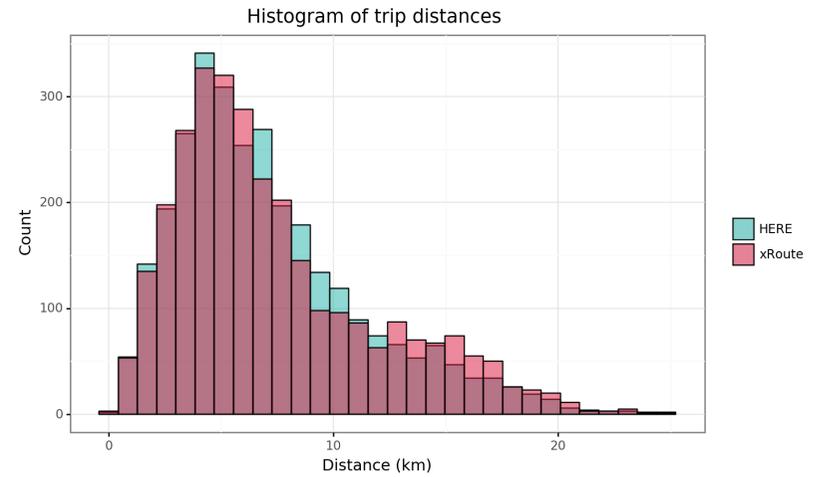
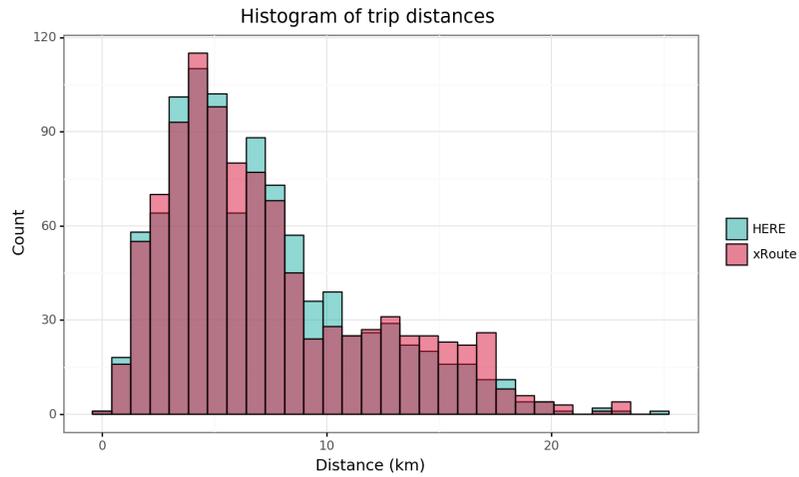
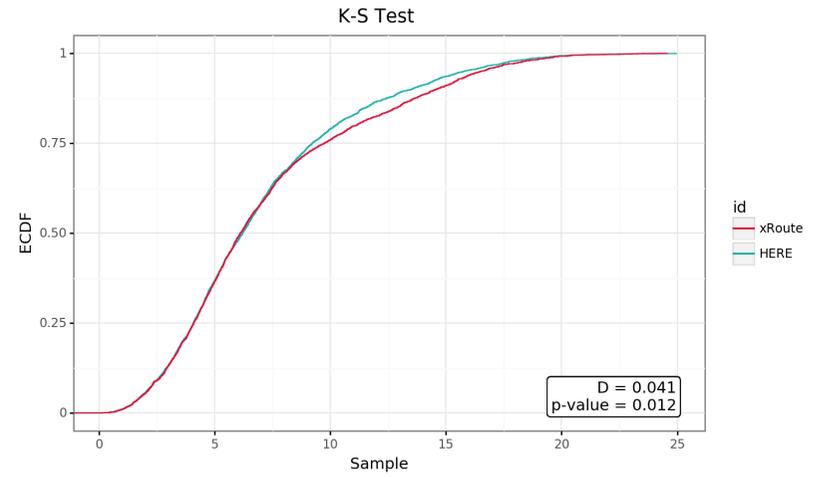
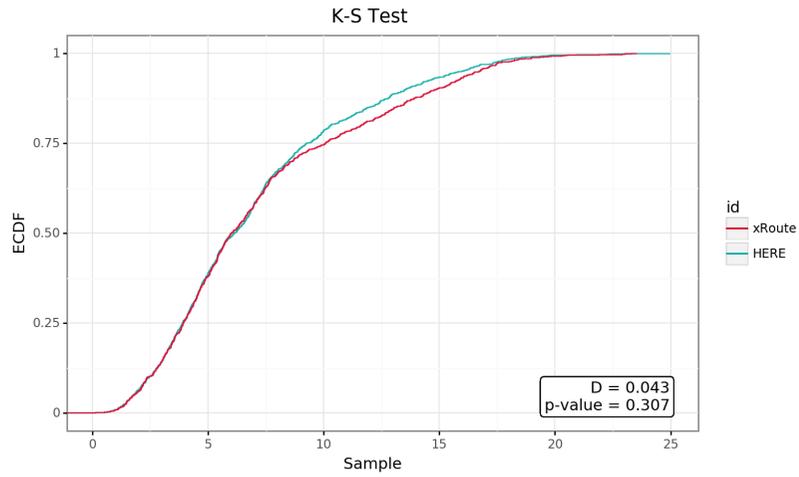


Figure E.15: KS test and trip distance distr. for BCN-R-1000 without filter

Figure E.16: KS test and trip distance distr. for BCN-R-3000 without filter

**E.2.2 Maximum absolute distance error (MAXD)**

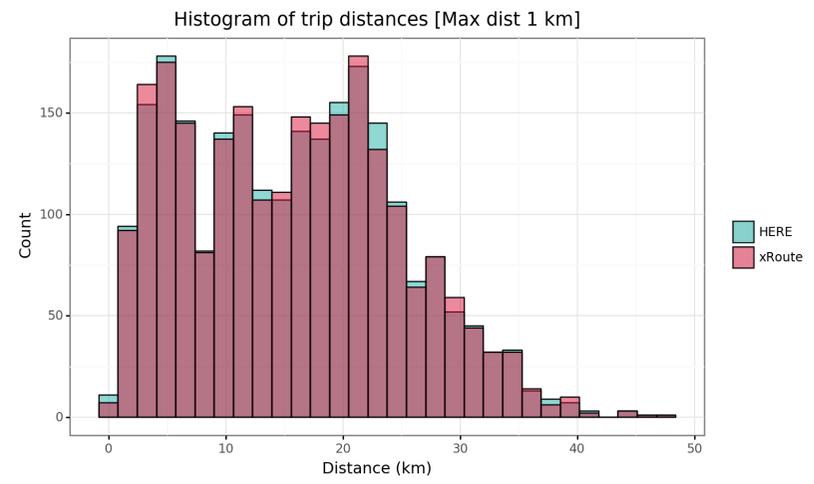
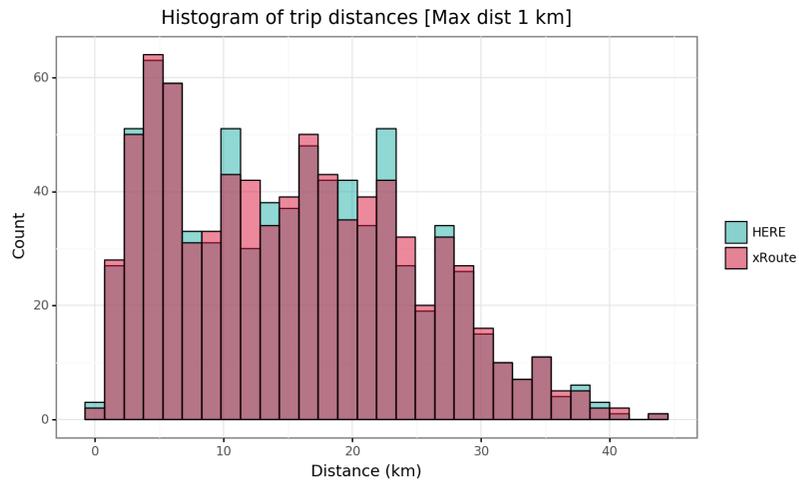
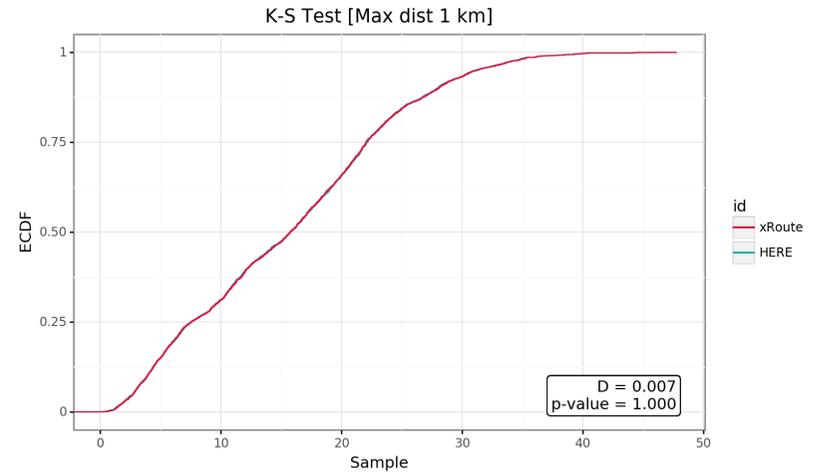
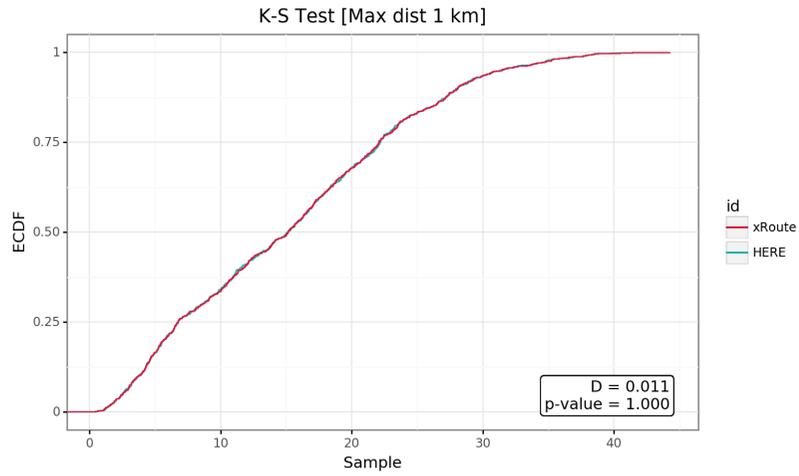


Figure E.17: KS test and trip distance distr. for BCN-1000 with MAXD-1 filter

Figure E.18: KS test and trip distance distr. for BCN-3000 with MAXD-1 filter

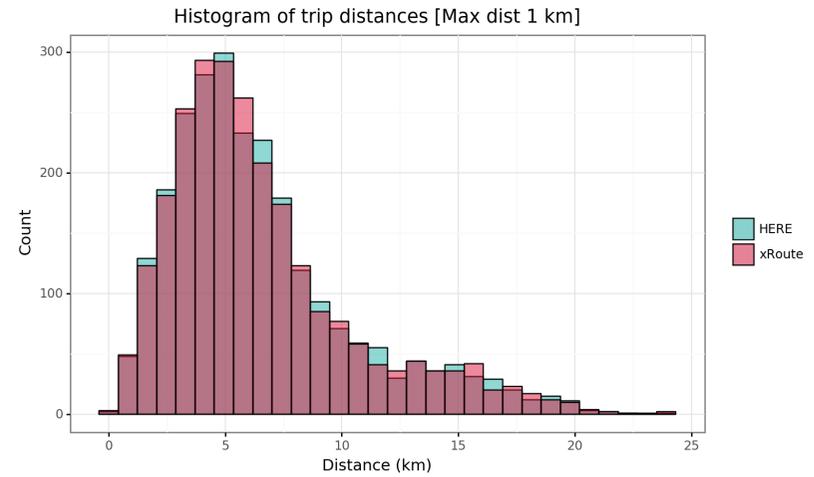
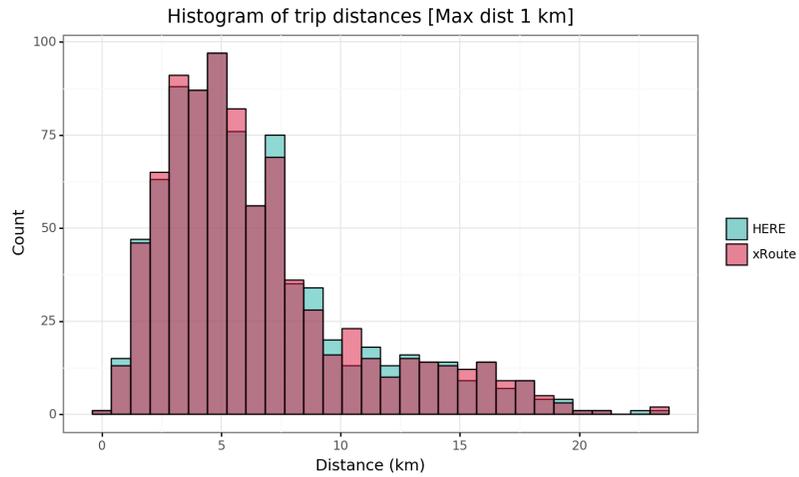
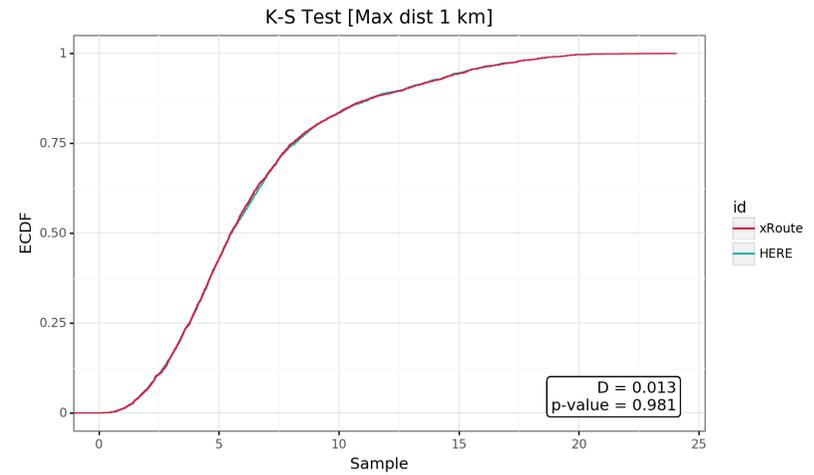
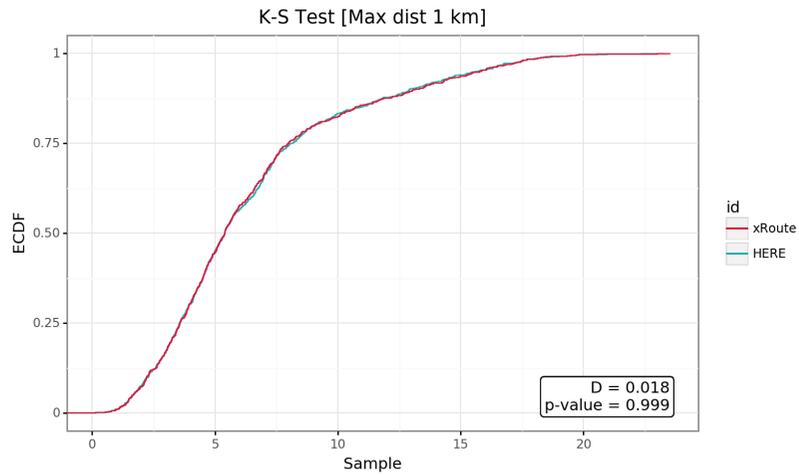


Figure E.19: KS test and trip distance distr. for BCN-R-1000 with MAXD-1 filter

Figure E.20: KS test and trip distance distr. for BCN-R-3000 with MAXD-1 filter

**E.2.3 Maximum absolute distance percentage error filter (MAXDP)**

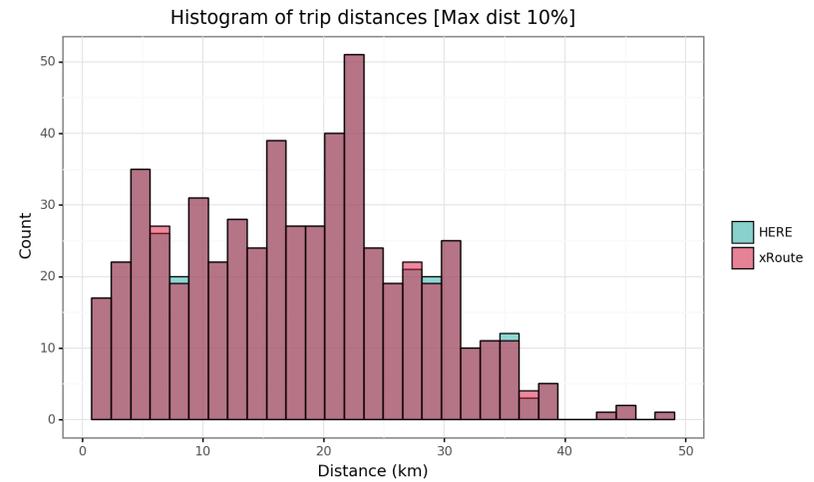
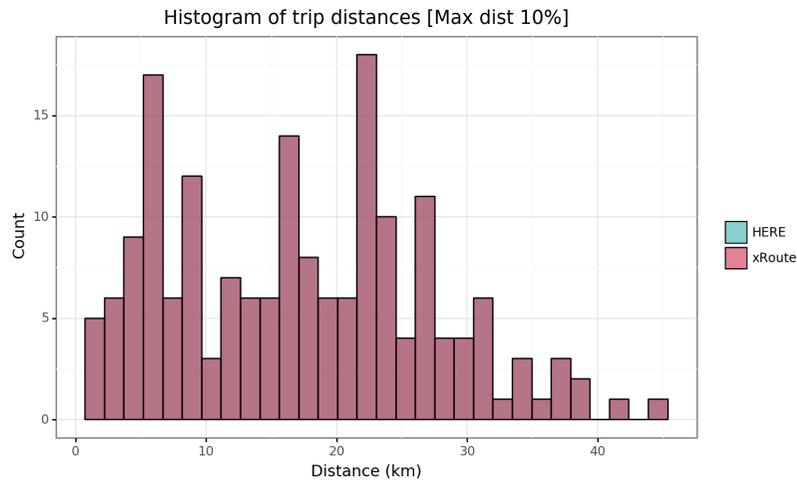
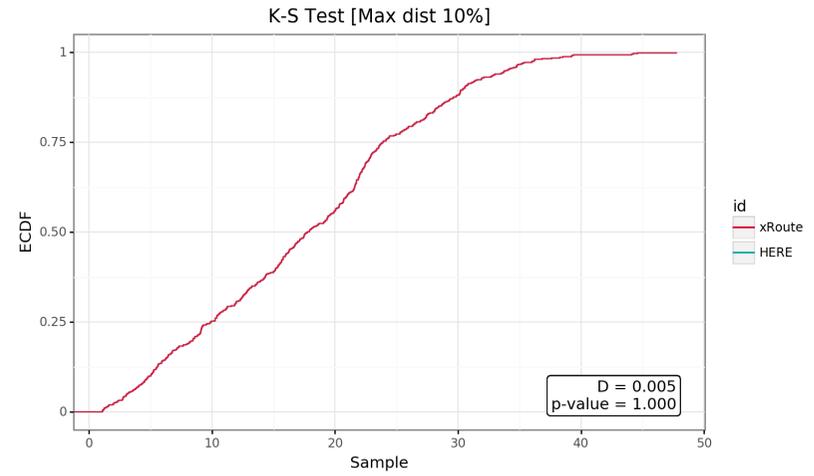
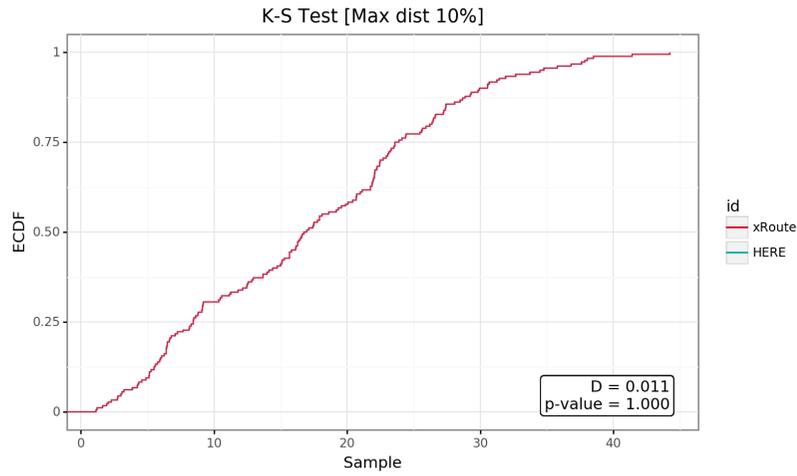


Figure E.21: KS test and trip distance distr. for BCN-1000 with MAXDP-0.1 filter

Figure E.22: KS test and trip distance distr. for BCN-3000 with MAXDP-0.1 filter

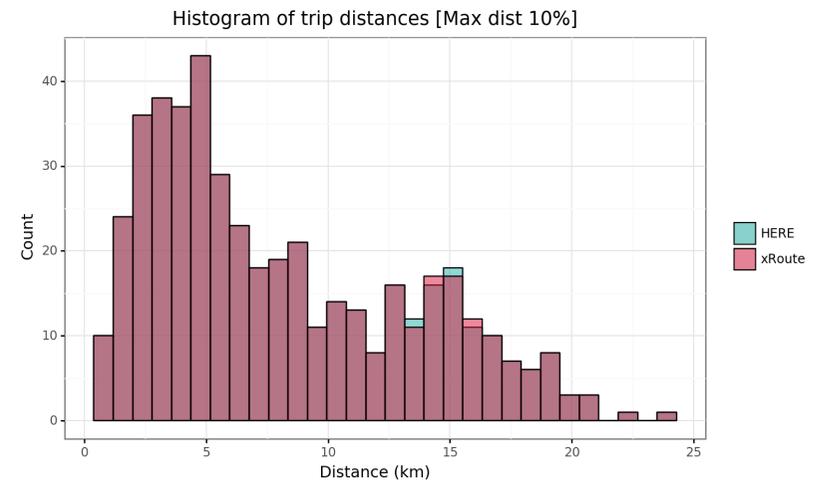
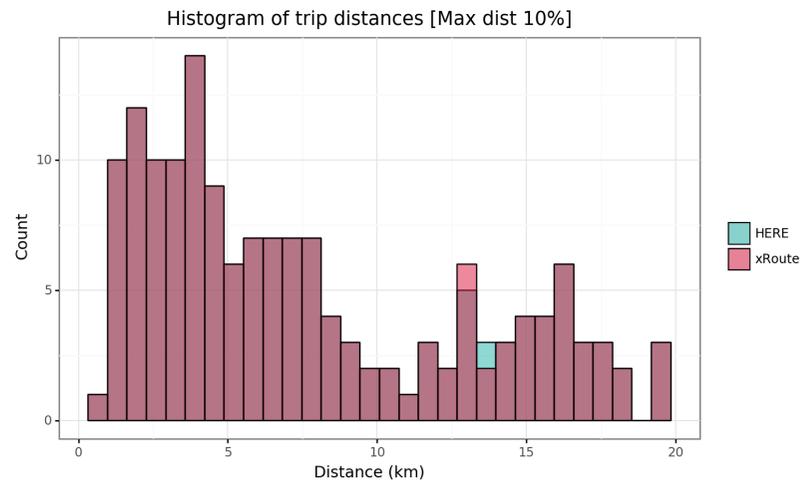
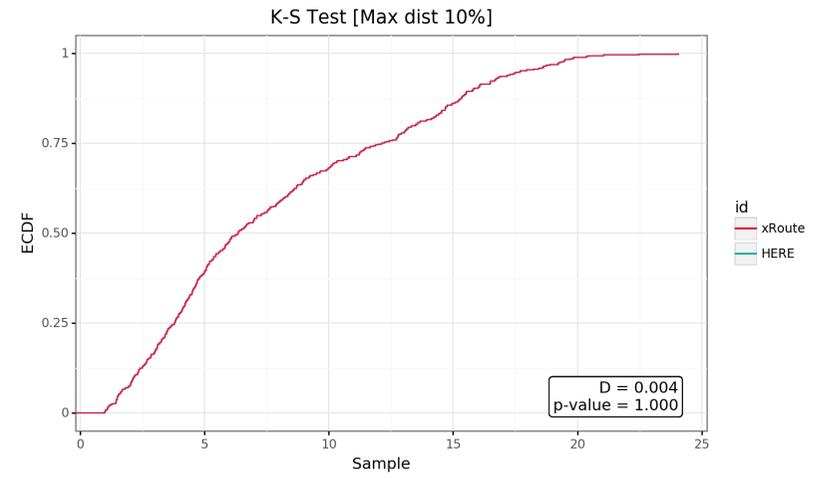
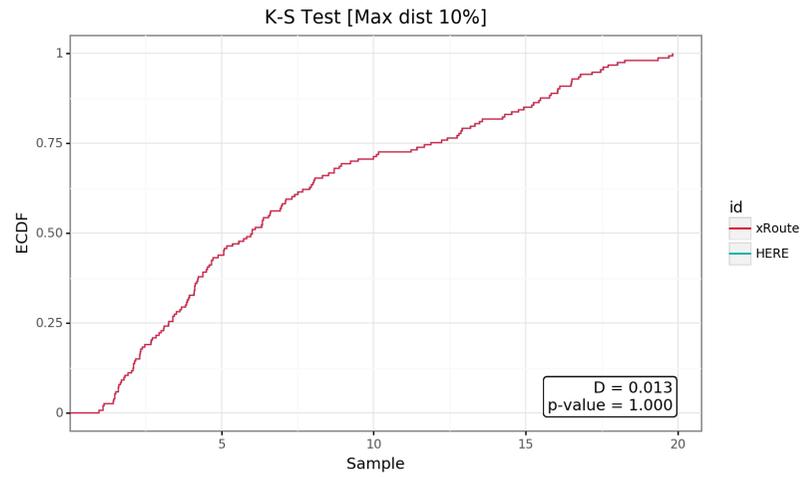


Figure E.23: KS test and trip distance distr. for BCN-R-1000 with MAXDP-0.1 filter

Figure E.24: KS test and trip distance distr. for BCN-R-3000 with MAXDP-0.1 filter

### E.2.4 Euclidean distance filter (EUCL)

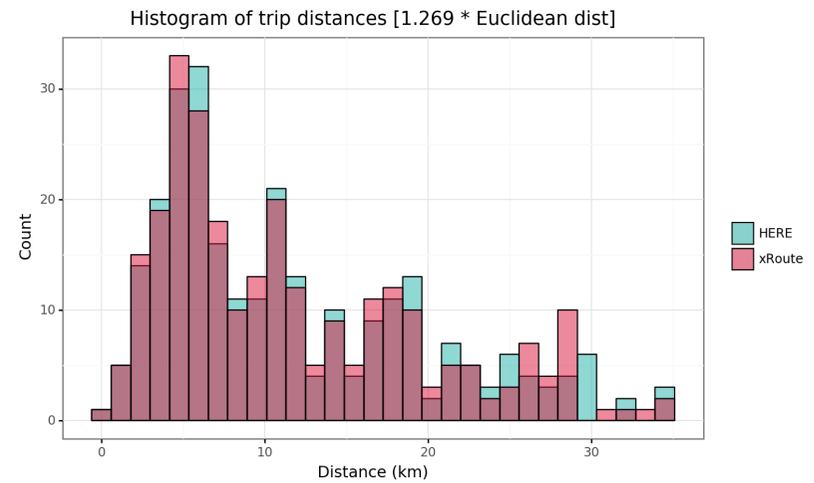
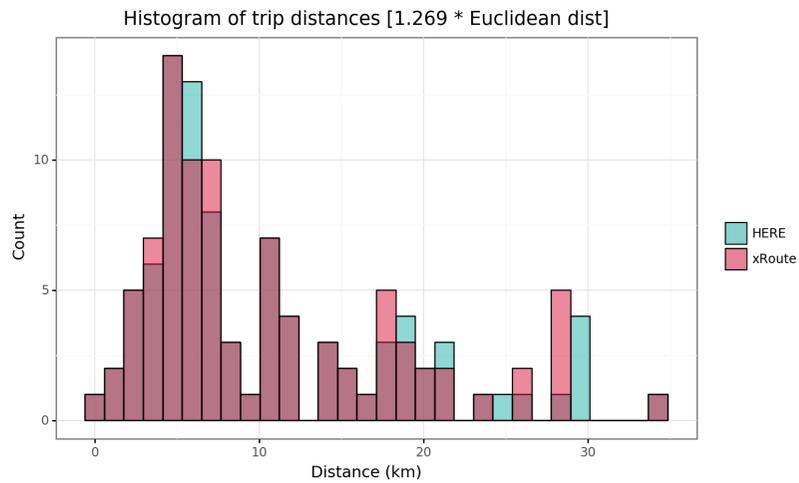
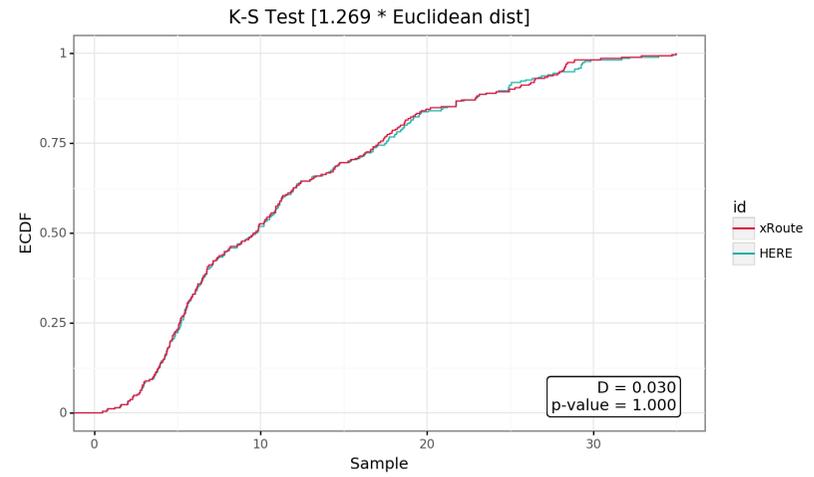
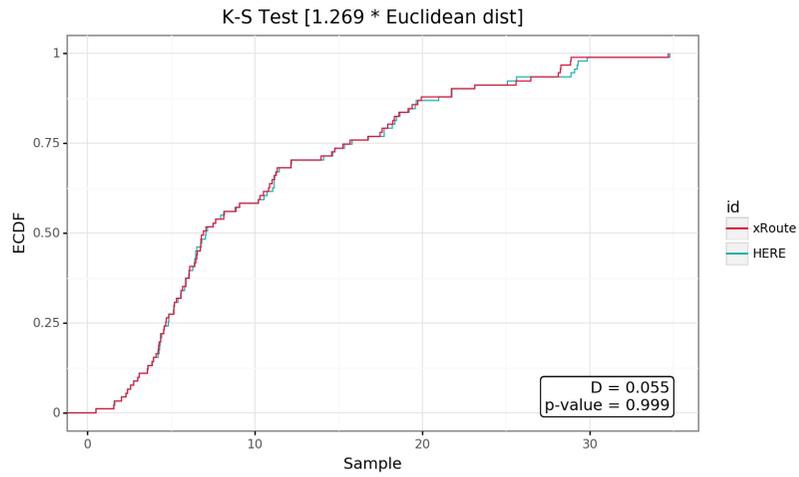


Figure E.25: KS test and trip distance distr. for BCN-1000 with EUCL-1.269 filter

Figure E.26: KS test and trip distance distr. for BCN-3000 with EUCL-1.269 filter

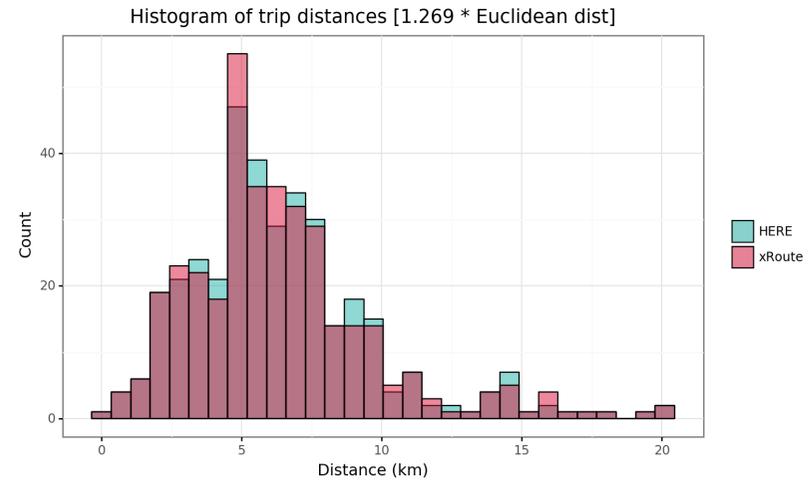
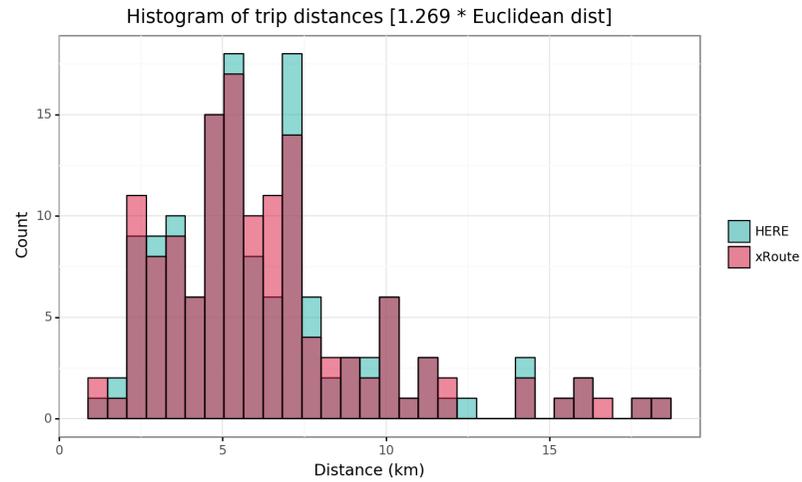
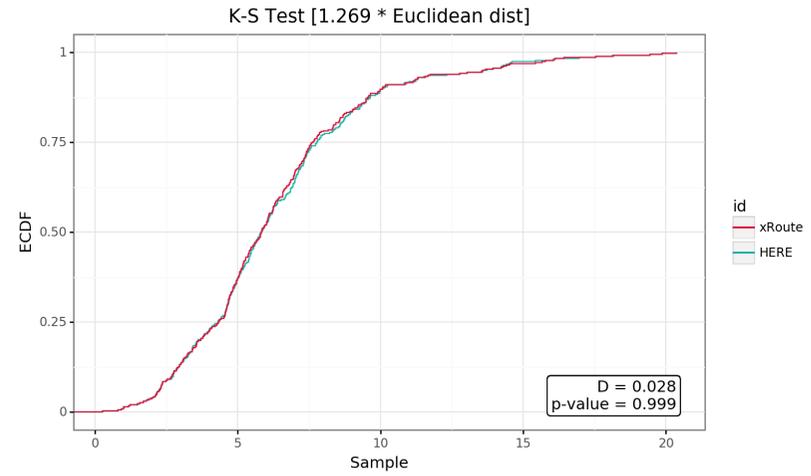
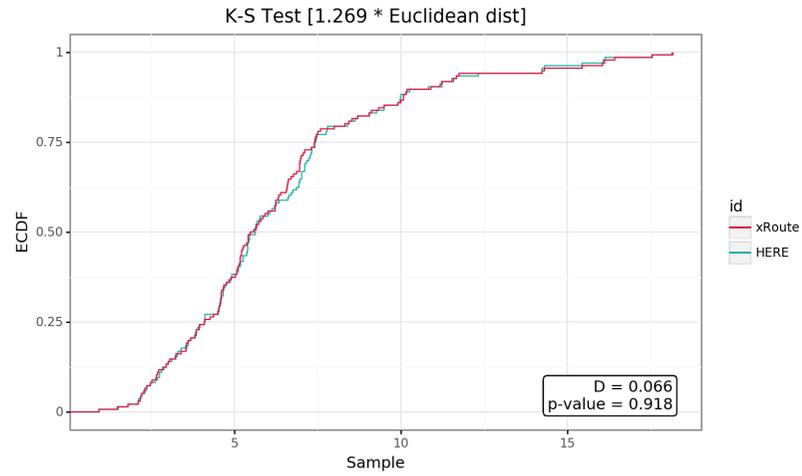


Figure E.27: KS test and trip distance distr. for BCN-R-1000 with EUCL-1.269 filter

Figure E.28: KS test and trip distance distr. for BCN-R-3000 with EUCL-1.269 filter

**E.2.5 Dynamic euclidean distance filter (DYN-EUCL)**

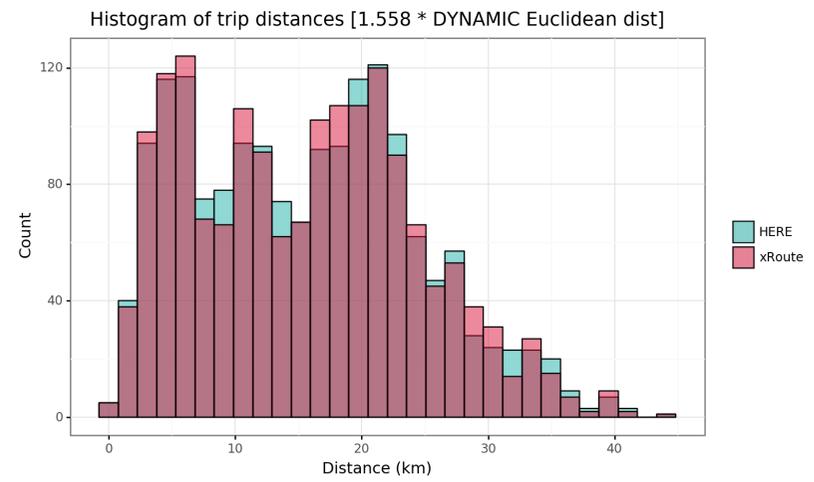
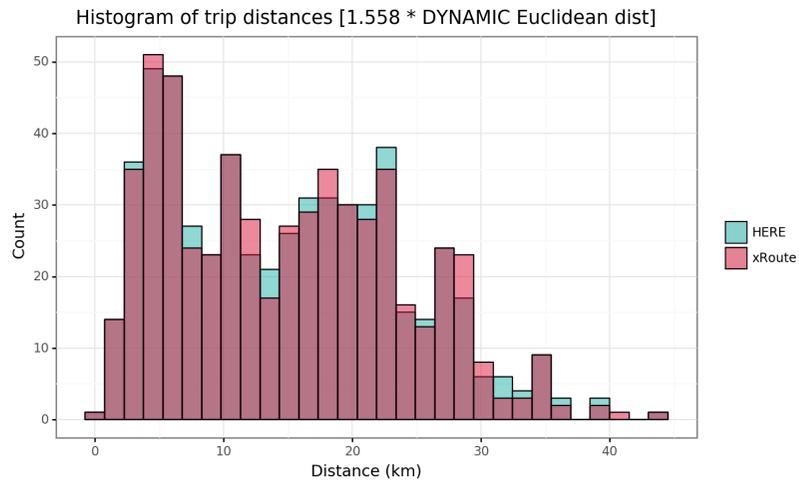
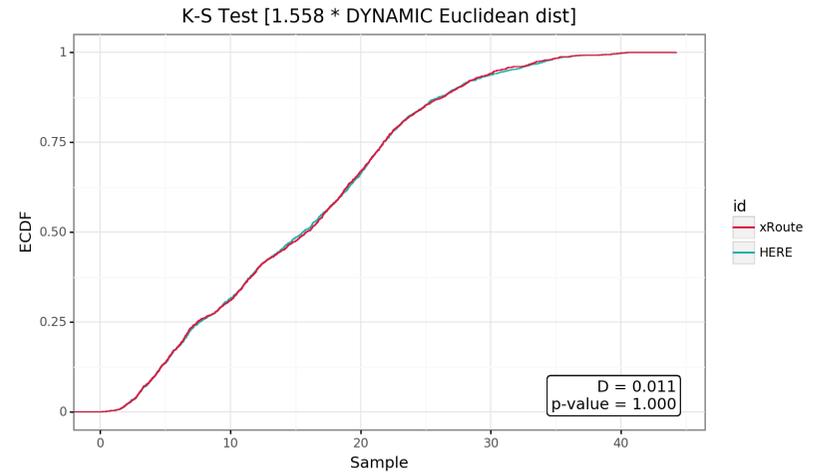
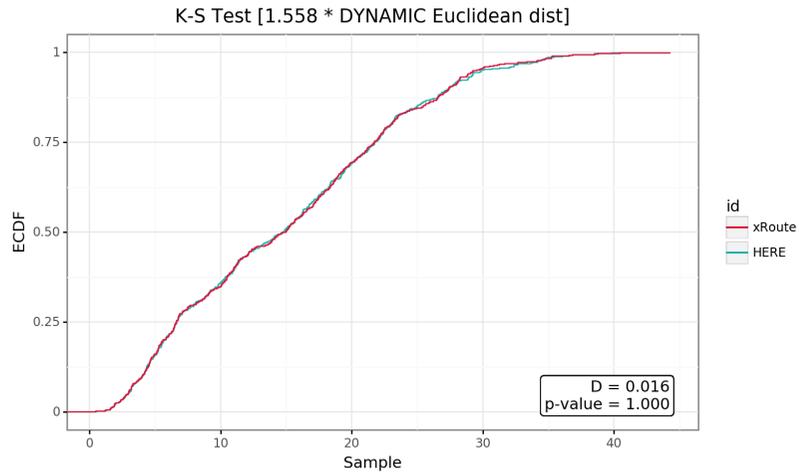


Figure E.29: KS test and trip distance distr. for BCN-1000 with DYN-EUCL filter

Figure E.30: KS test and trip distance distr. for BCN-3000 with DYN-EUCL filter

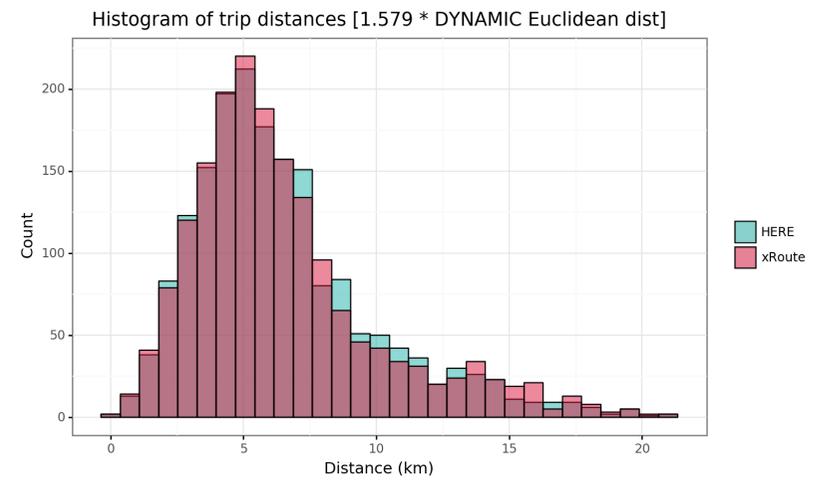
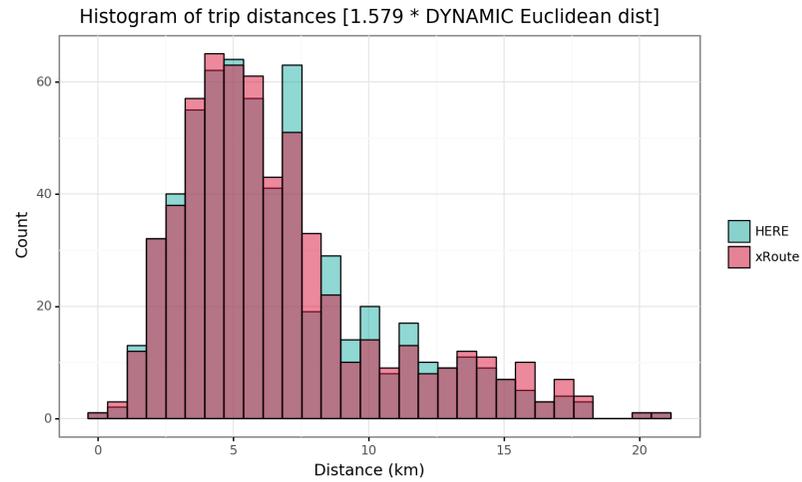
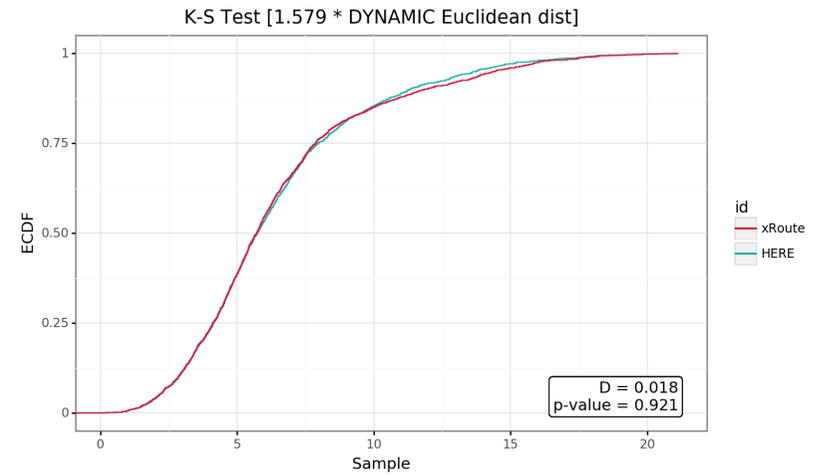
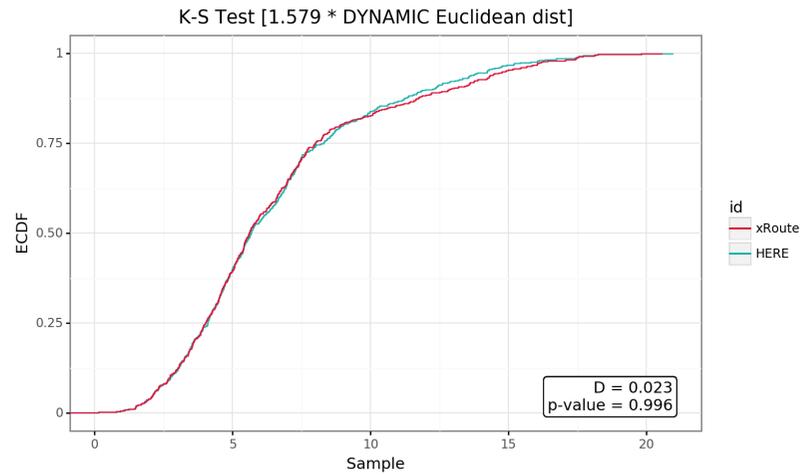


Figure E.31: KS test and trip distance distr. for BCN-R-1000 with DYN-EUCL filter

Figure E.32: KS test and trip distance distr. for BCN-R-3000 with DYN-EUCL filter

## F Speed calibration results

### F.1 Area 1: Barcelona Metropolitan Area (BCN)

#### F.1.1 Speed profiles variation

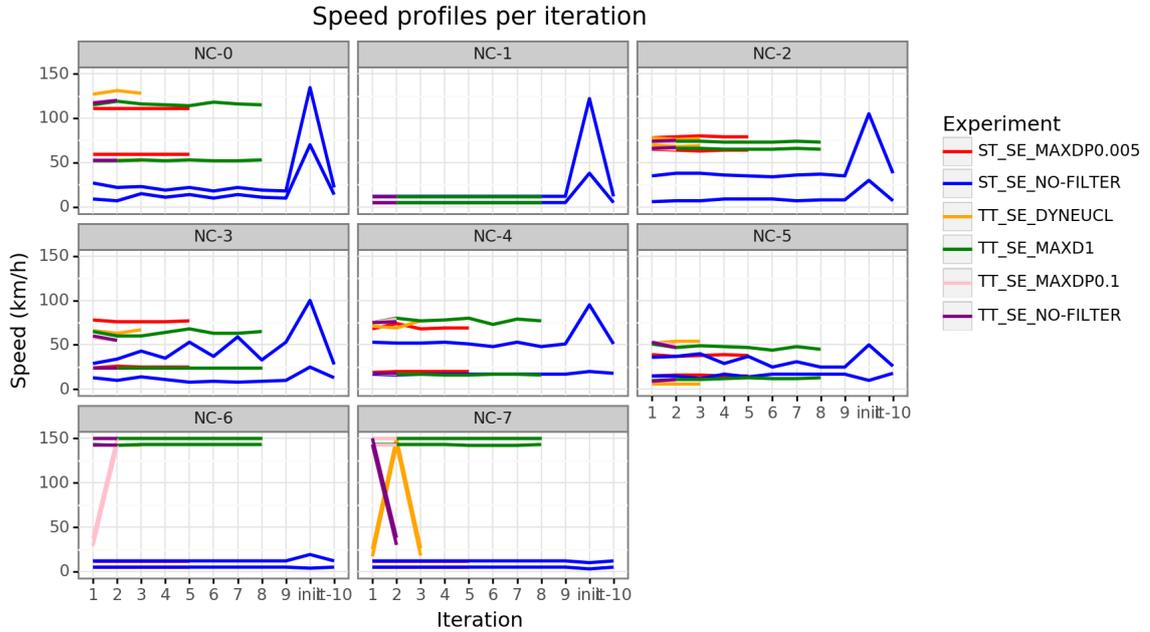


Figure F.1: Speed profiles variation with BCN-1000 sample

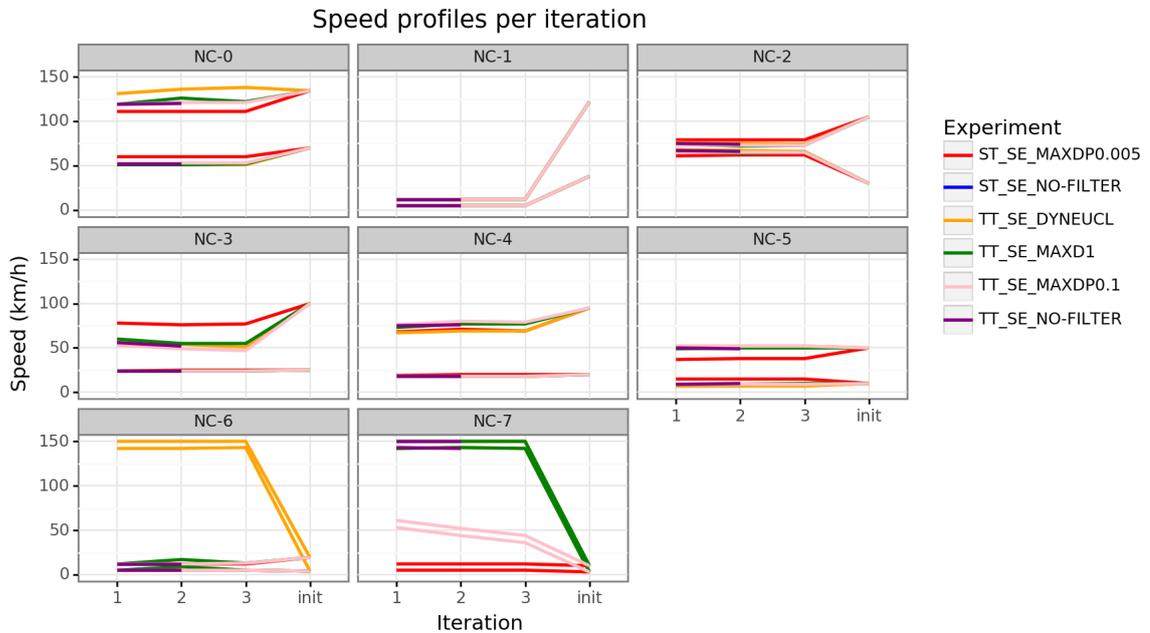
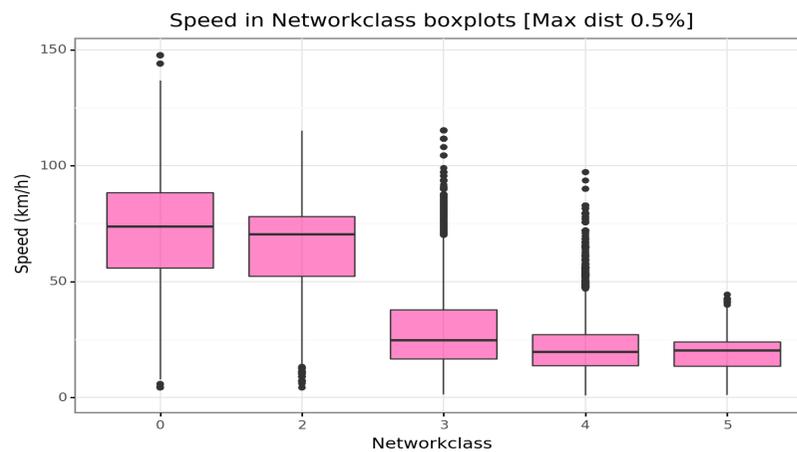
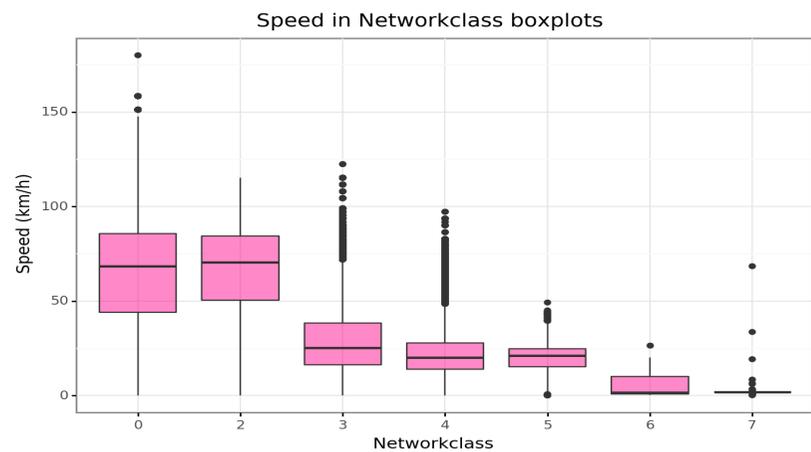


Figure F.2: Speed profiles variation with BCN-3000 sample

### F.1.2 Speeds distribution with ST models



119

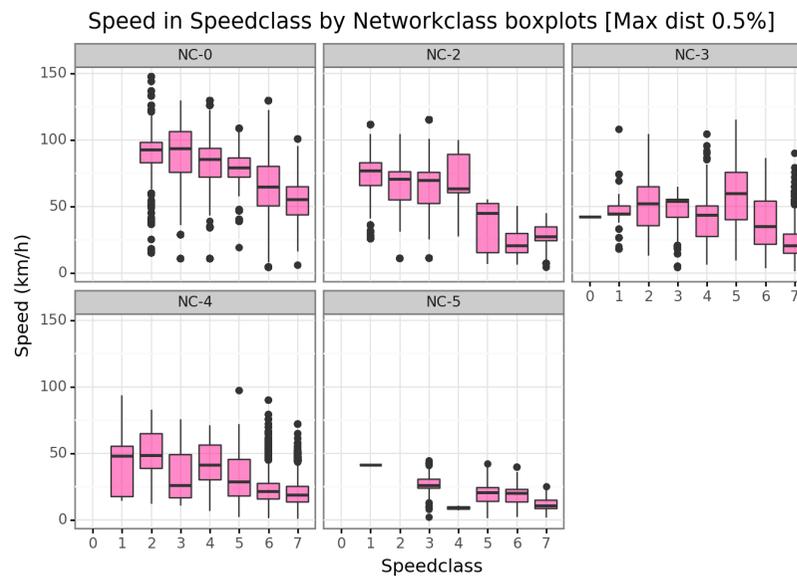
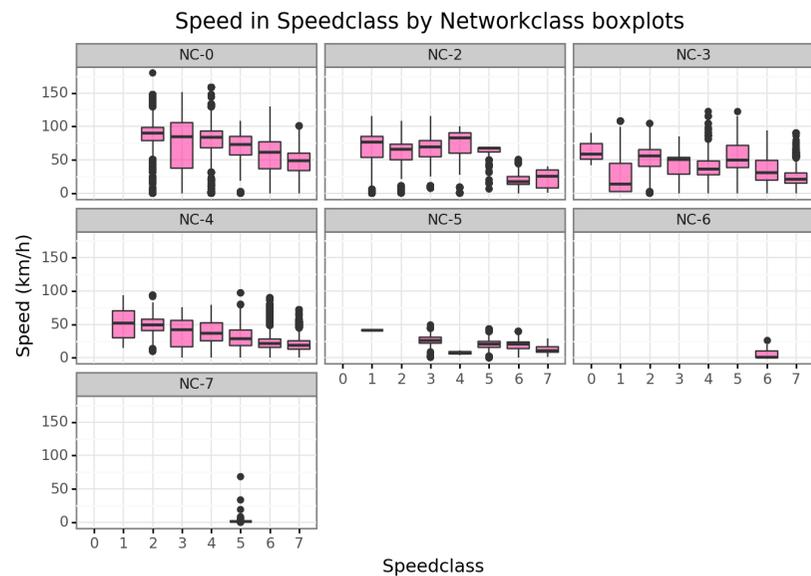
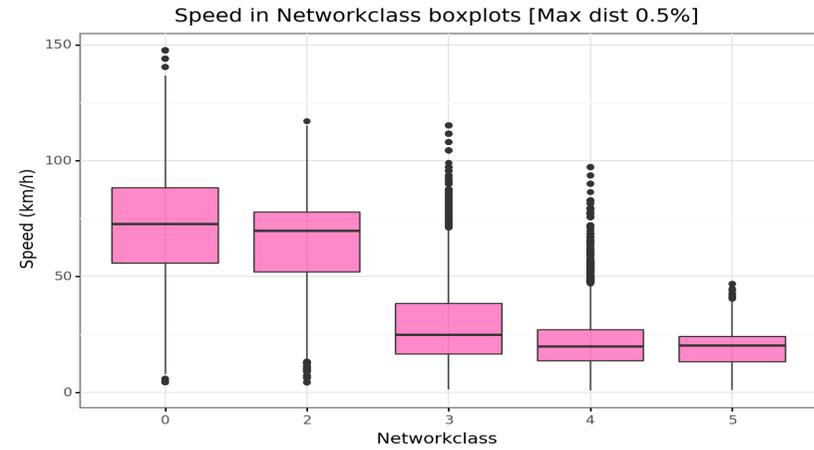
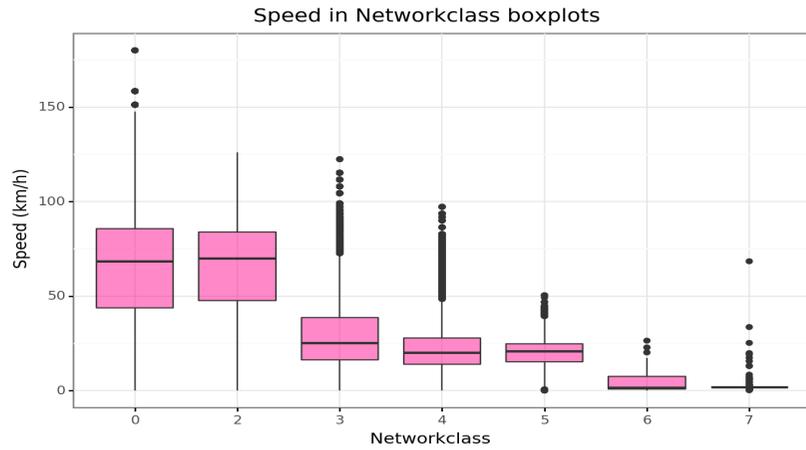


Figure F.3: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-1000 without filter

Figure F.4: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-1000 with MAXDP-0.005 filter



120

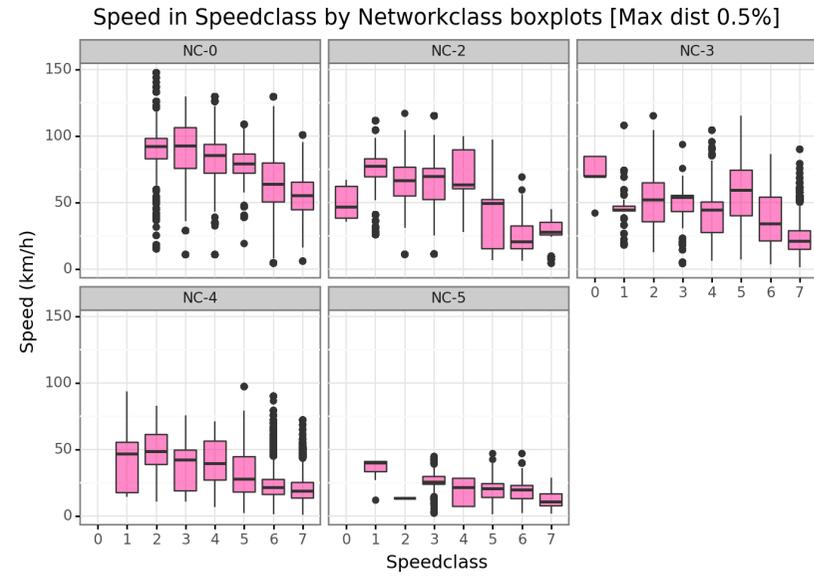
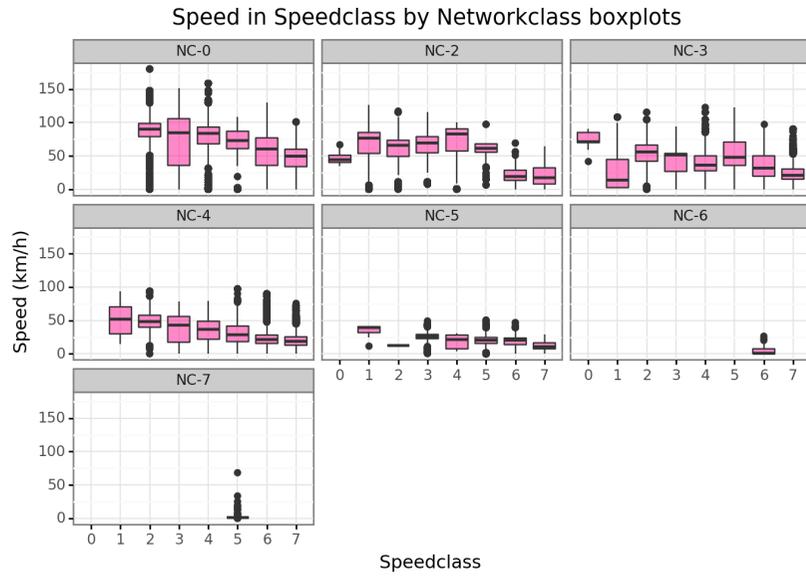


Figure F.5: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-3000 without filter

Figure F.6: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-3000 with MAXDP-0.005 filter

### F.1.3 Traveltime improvements

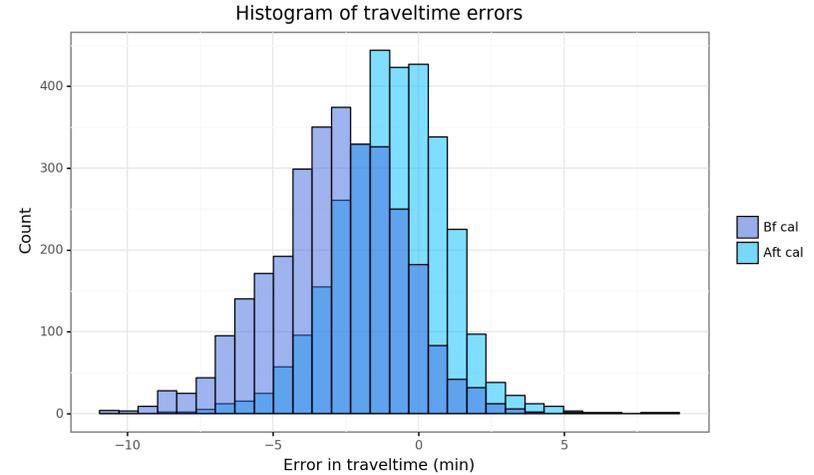
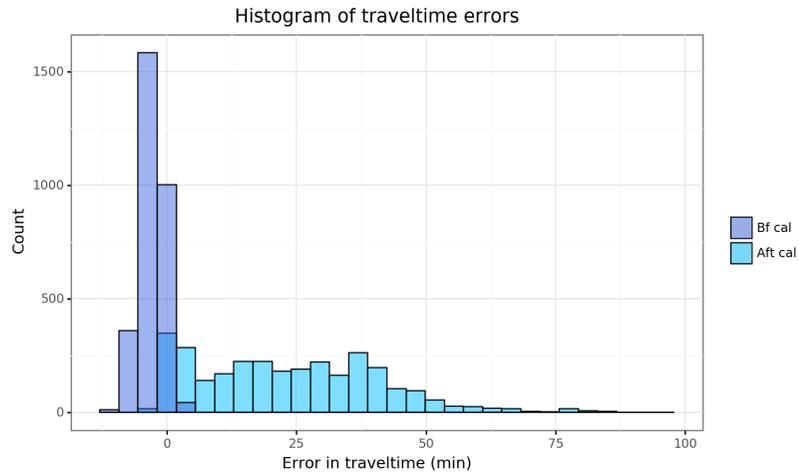
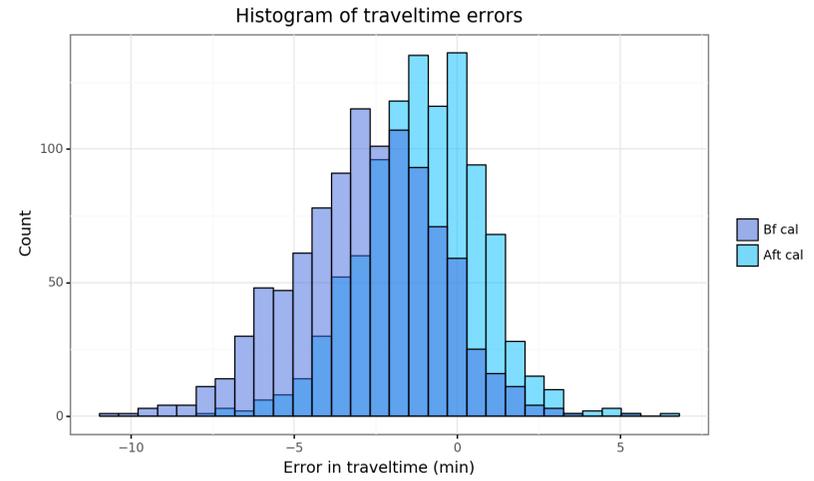
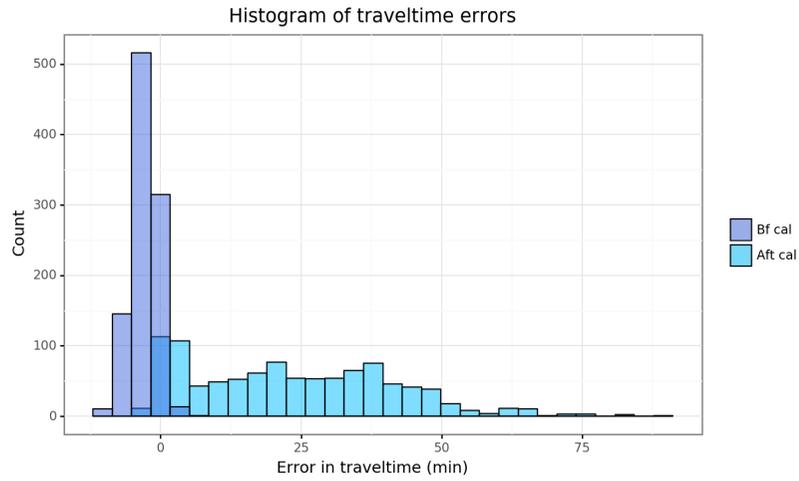
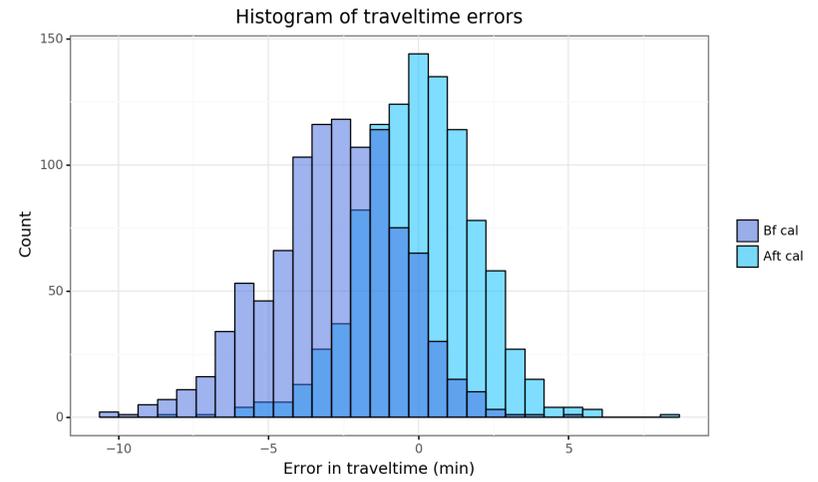
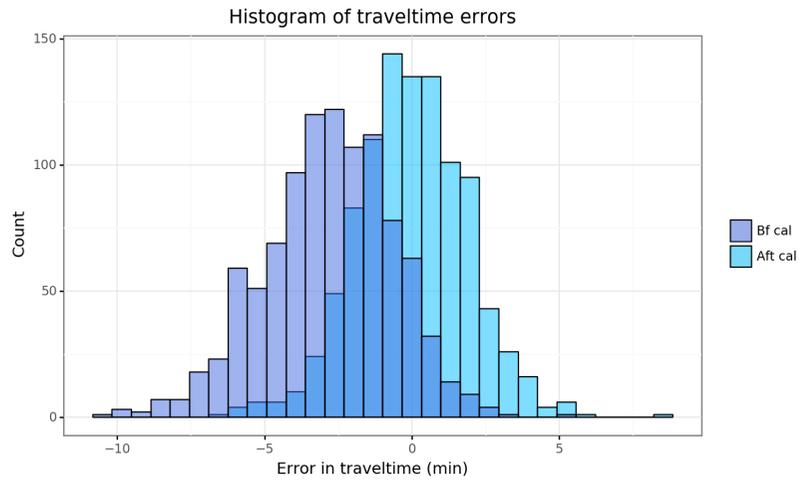


Figure F.7: Total traveltime error distr. for BCN-1000 and BCN-3000 with ST-SE model calibration without filter

Figure F.8: Total traveltime error distr. for BCN-1000 and BCN-3000 with ST-SE model calibration with MAXDP-0.005 filter



122

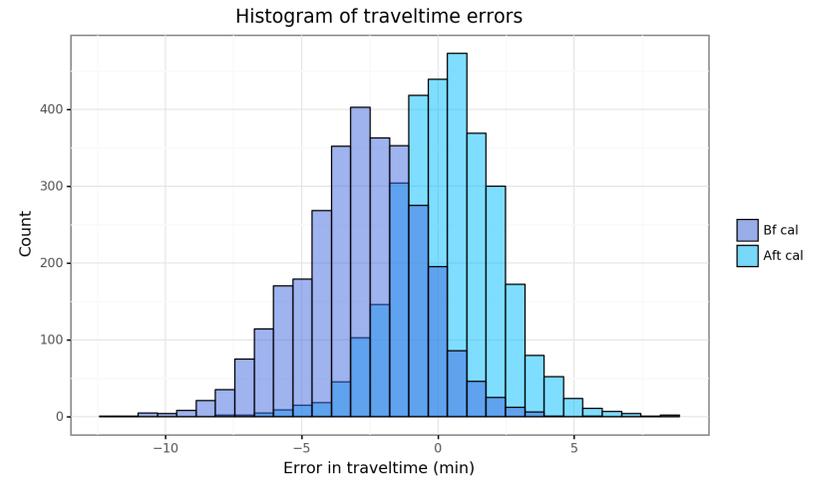
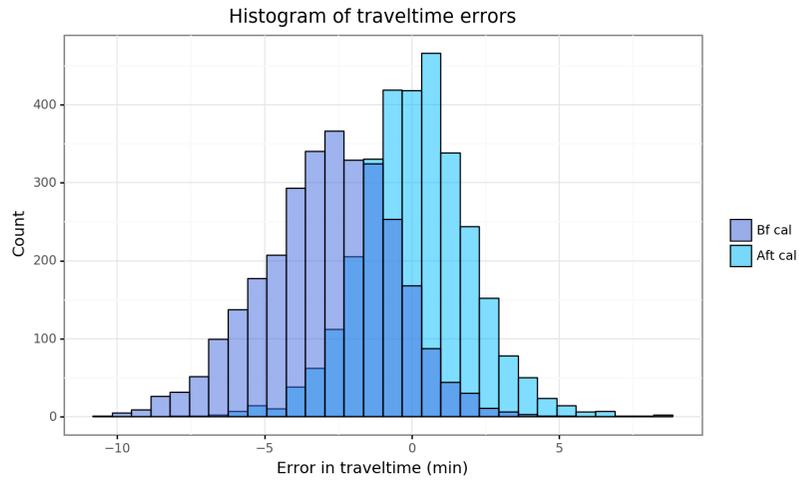


Figure F.9: Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration without filter

Figure F.10: Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration with MAXD-1 filter

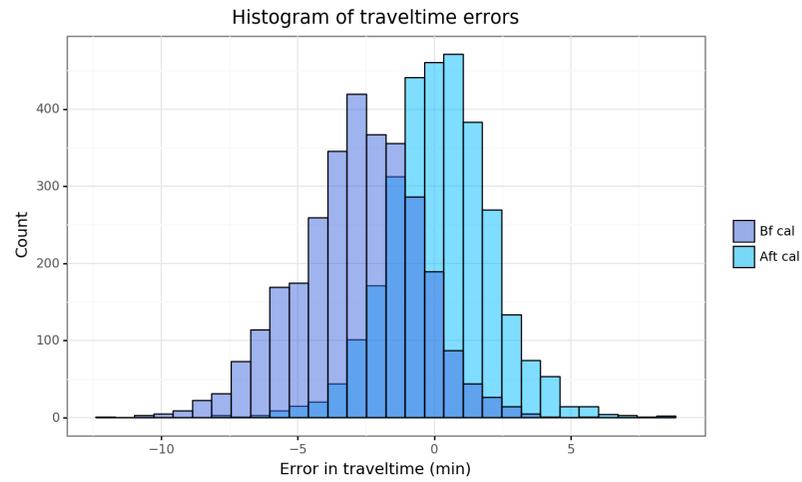
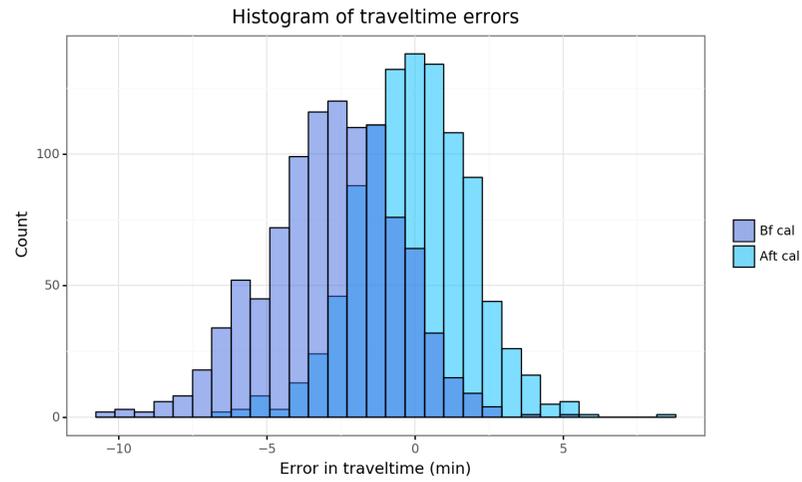


Figure F.11: Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration with MAXDP-0.1 filter

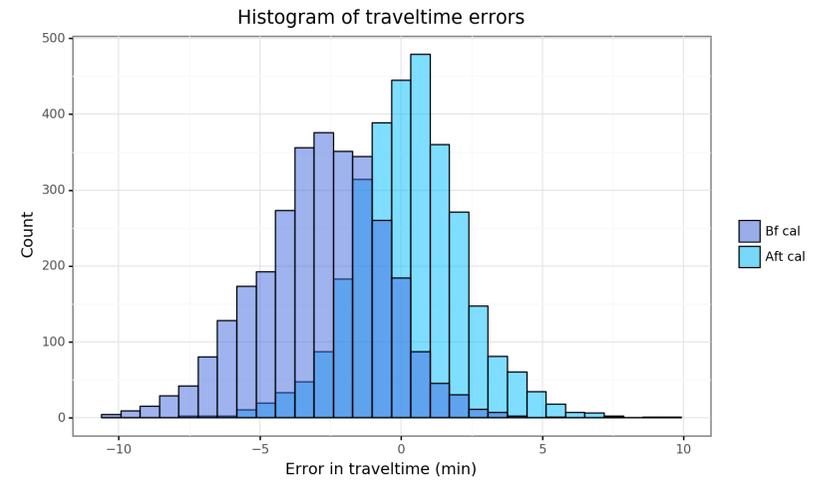
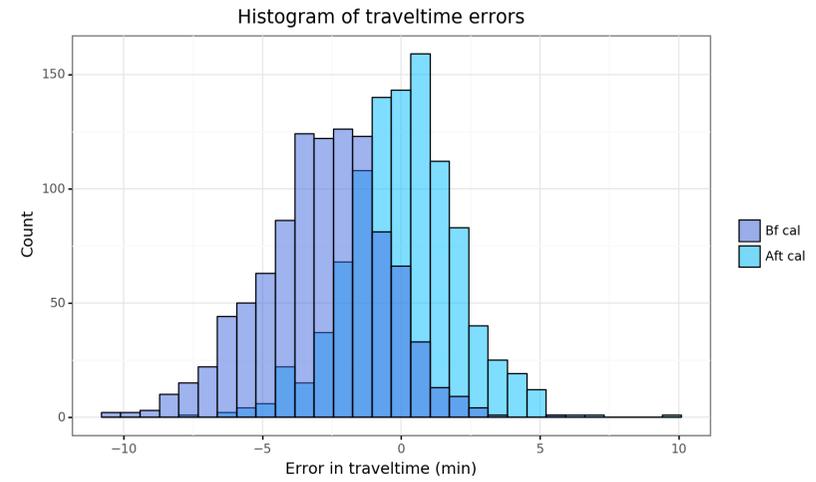


Figure F.12: Total traveltime error distr. for BCN-1000 and BCN-3000 with TT-SE model calibration with DYN-EUCL filter

## F.2 Area 2: Reduced Barcelona Metropolitan Area (BCN-R)

### F.2.1 Speed profiles variation

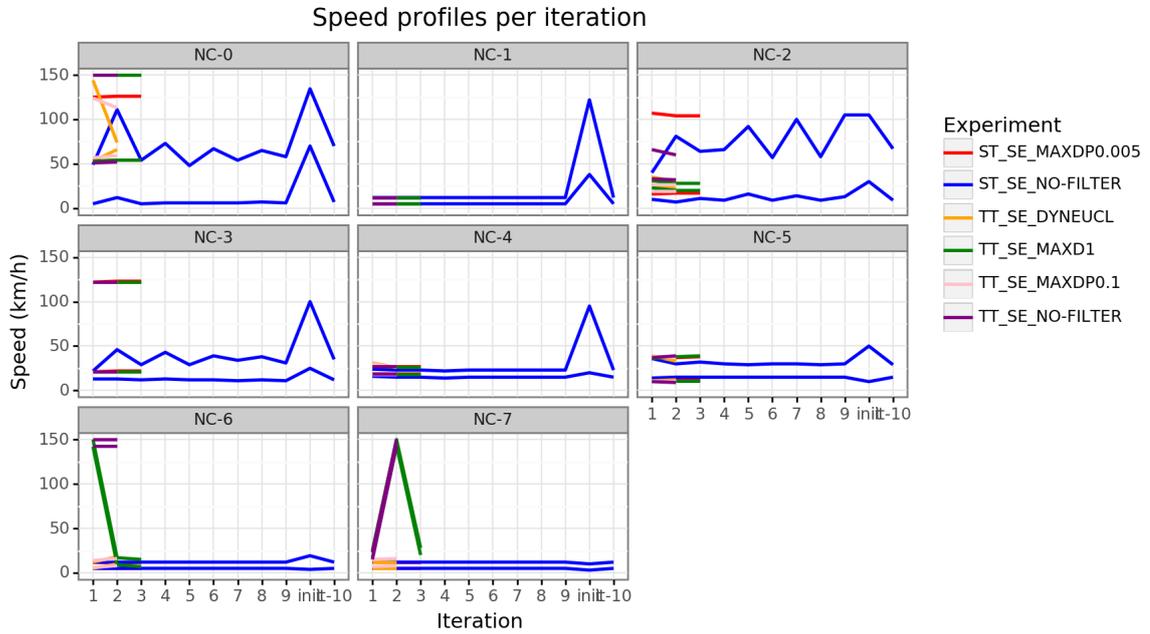


Figure F.13: Speed profiles variation with BCN-R-1000 sample

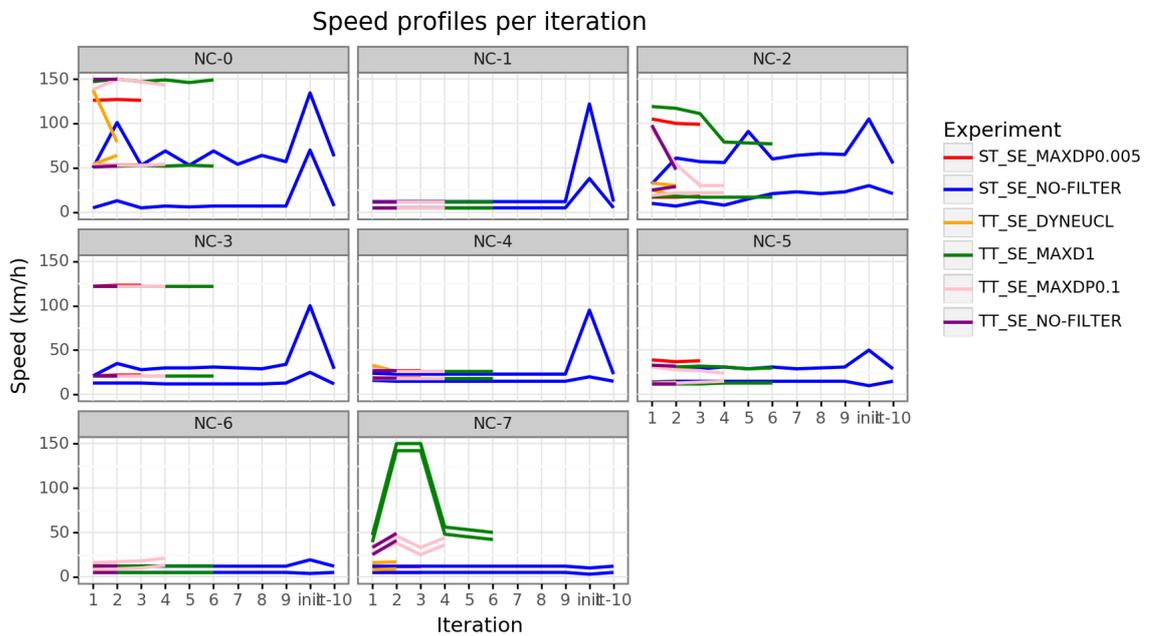


Figure F.14: Speed profiles variation with BCN-R-3000 sample

## F.2.2 Speeds distribution with ST models

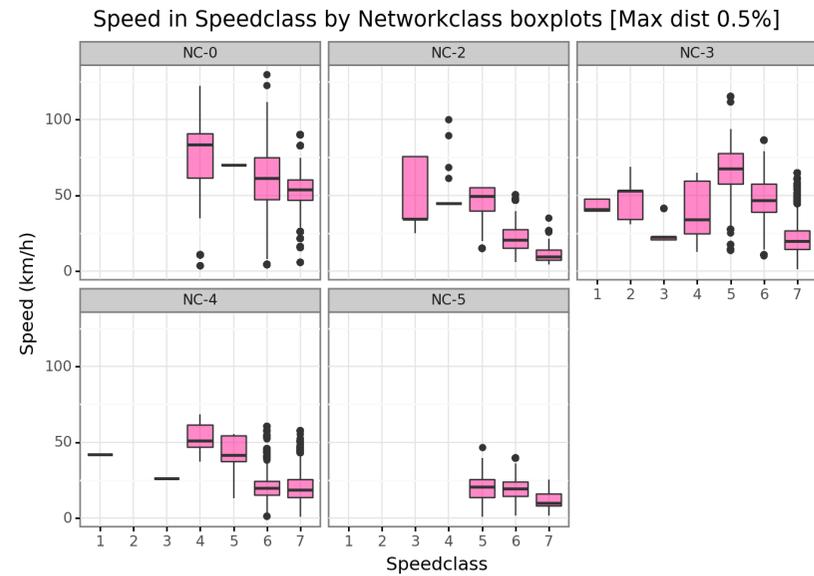
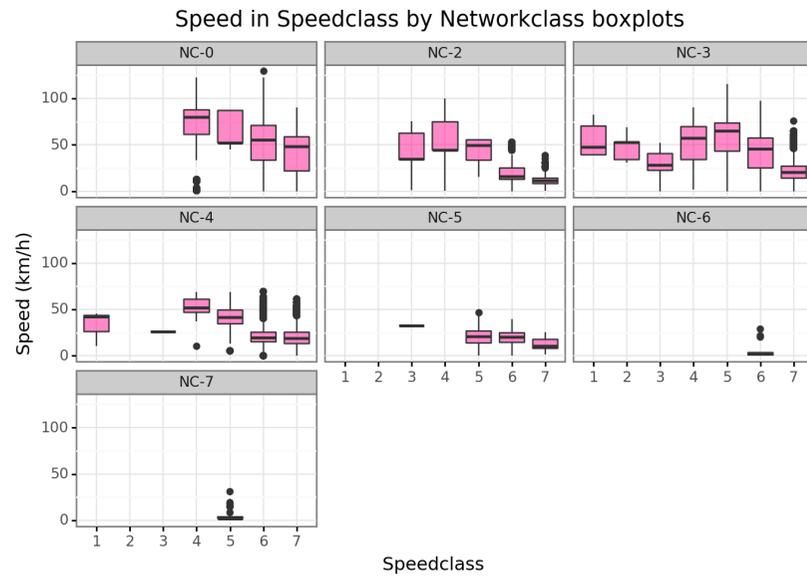
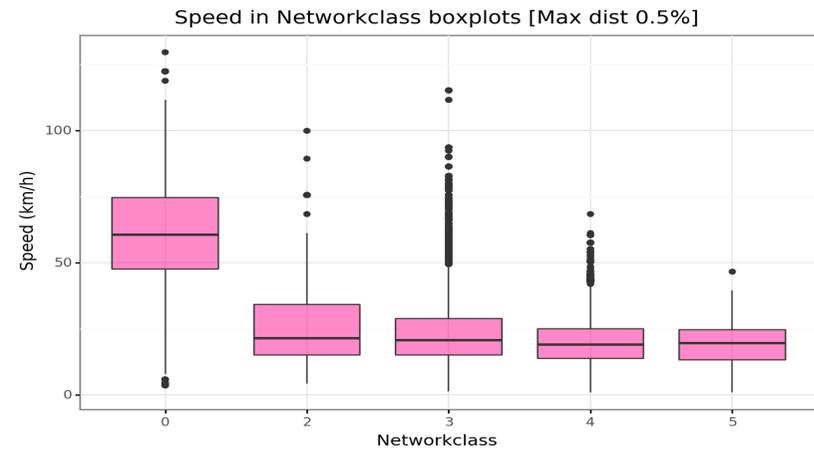
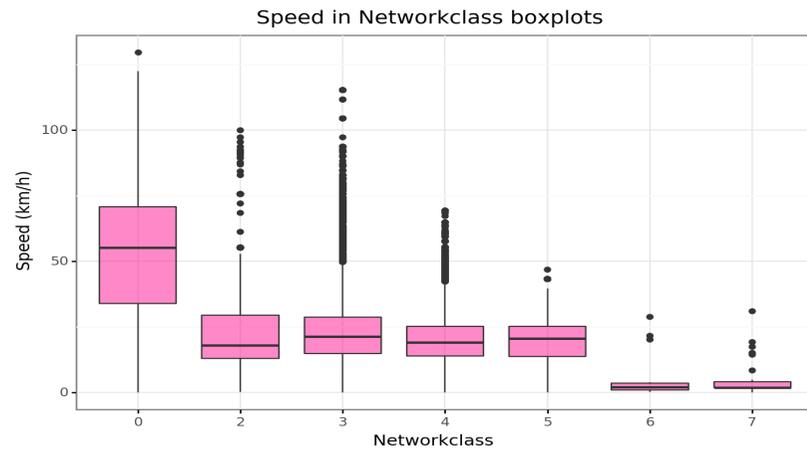


Figure F.15: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-1000 without filter

Figure F.16: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-1000 with MAXDP-0.005 filter

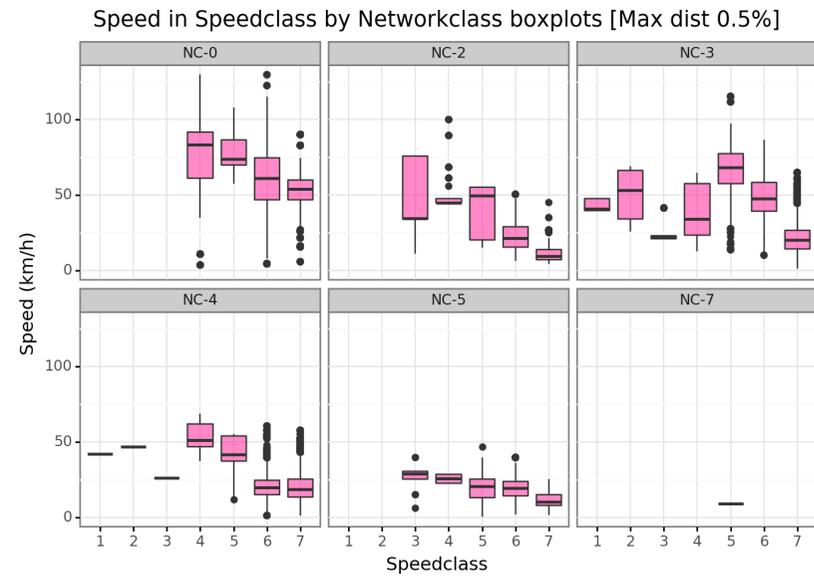
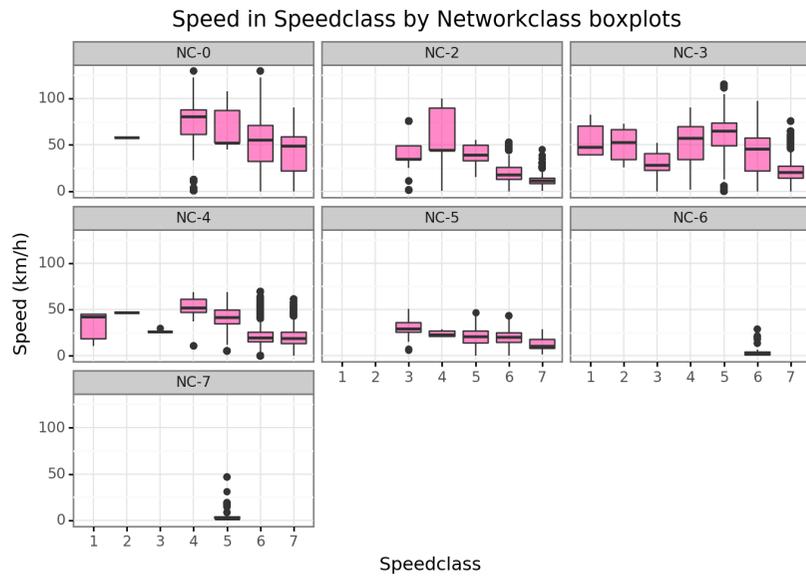
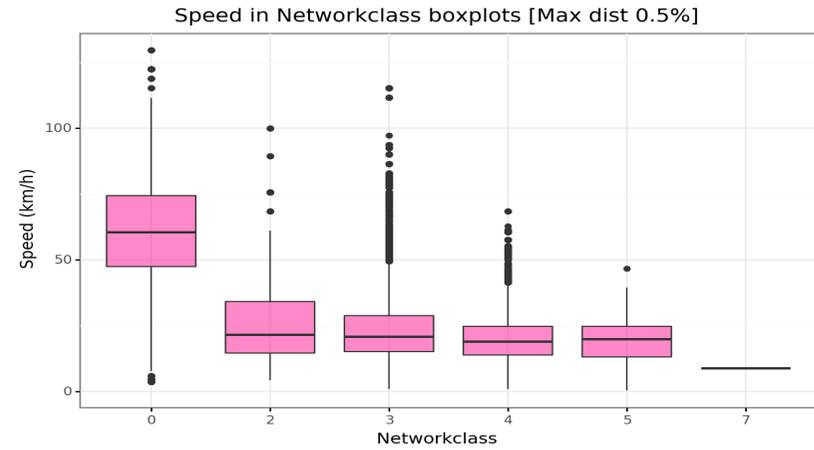
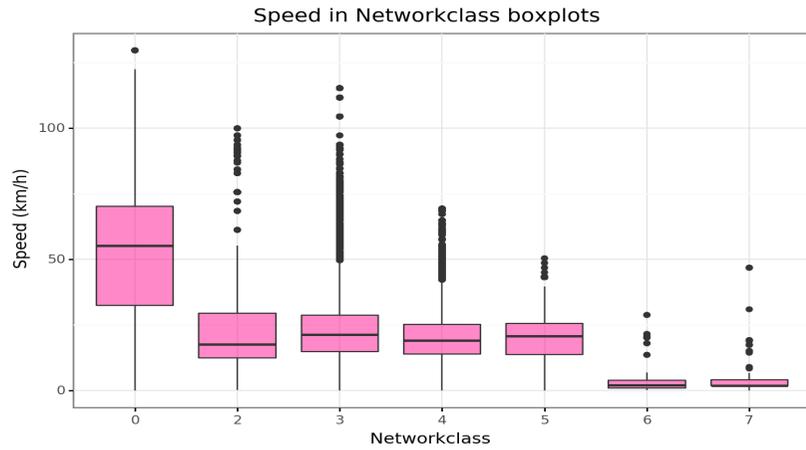


Figure F.17: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-3000 without filter

Figure F.18: Segment speeds distribution per networkclass and per networkclass-speedclass for BCN-R-3000 with MAXDP-0.005 filter

### F.2.3 Traveltime improvements

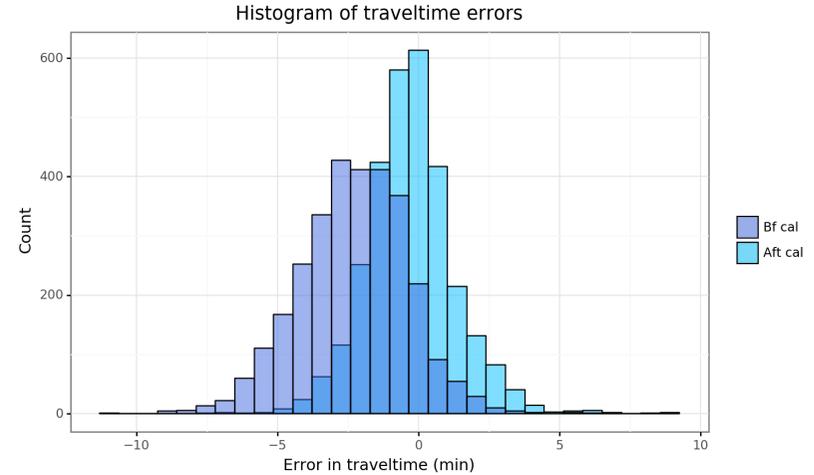
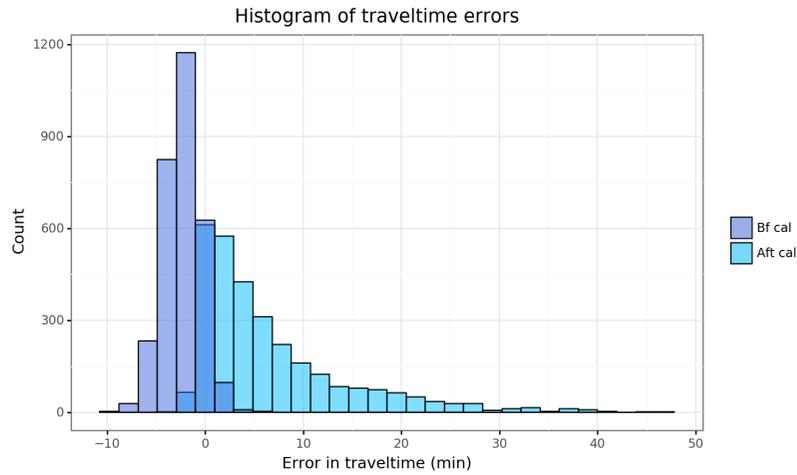
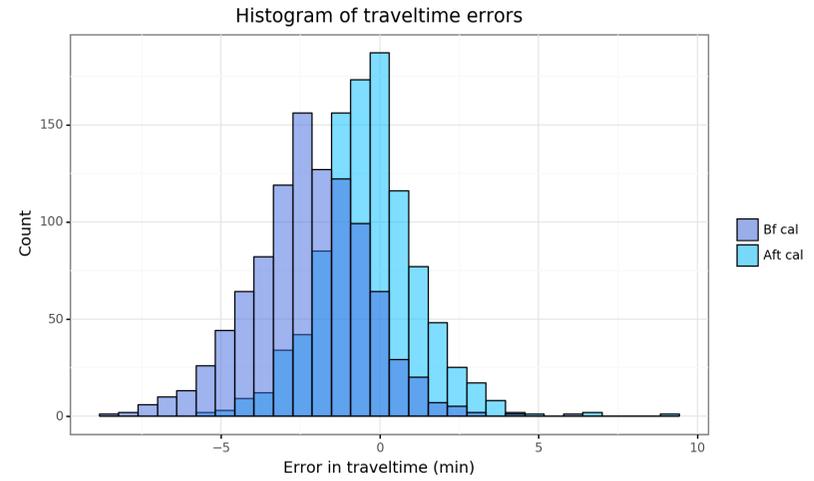
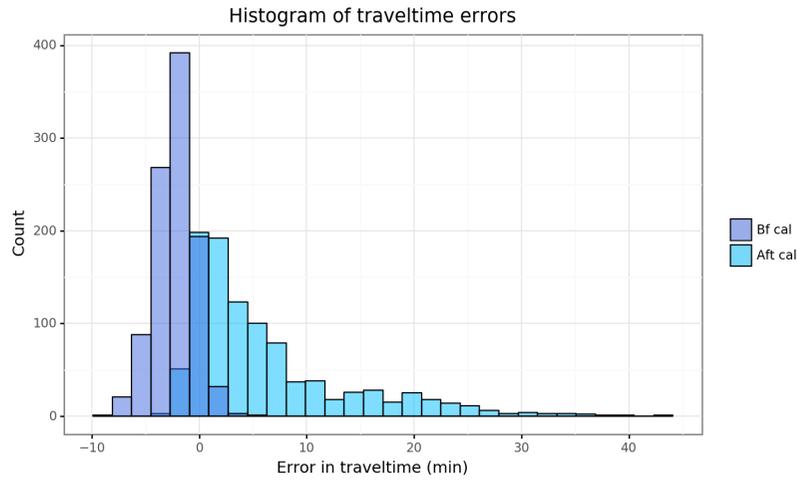


Figure F.19: Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with ST-SE model calibration without filter

Figure F.20: Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with ST-SE model calibration with MAXDP-0.005 filter

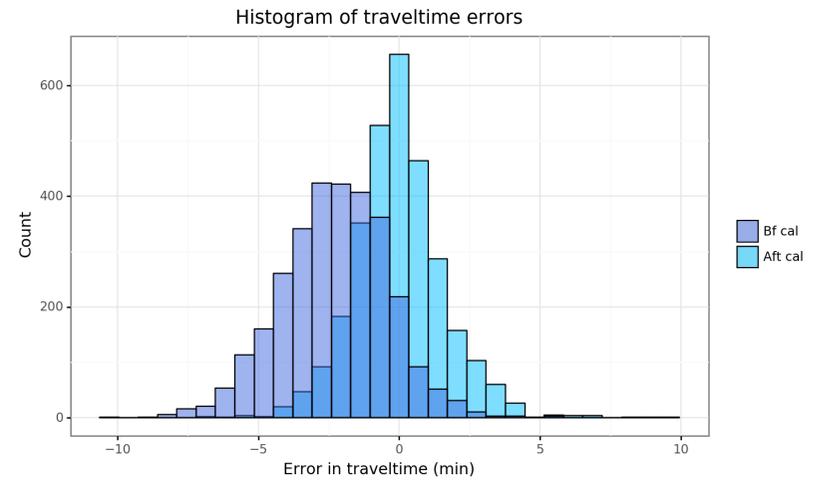
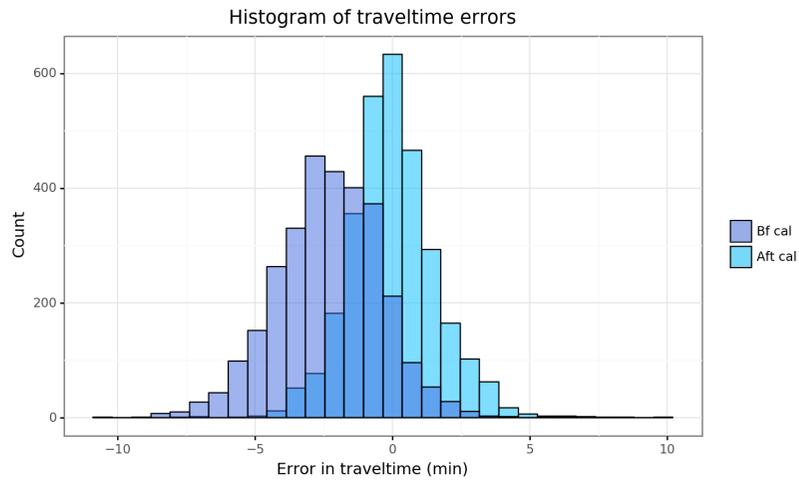
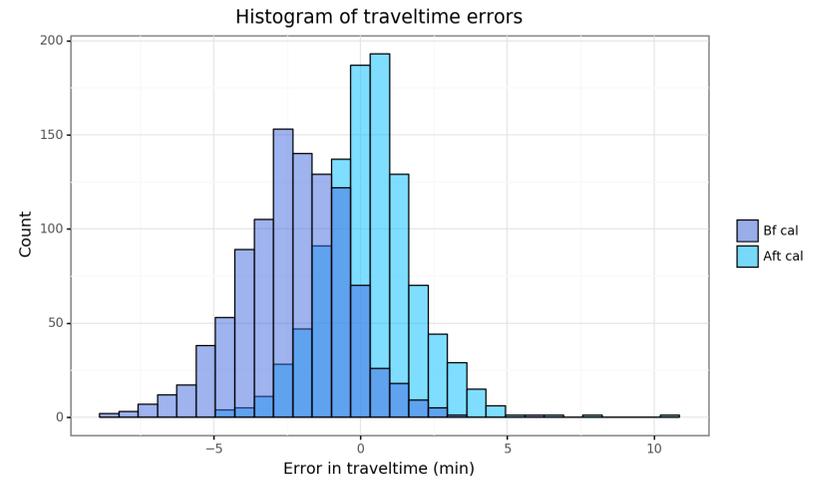
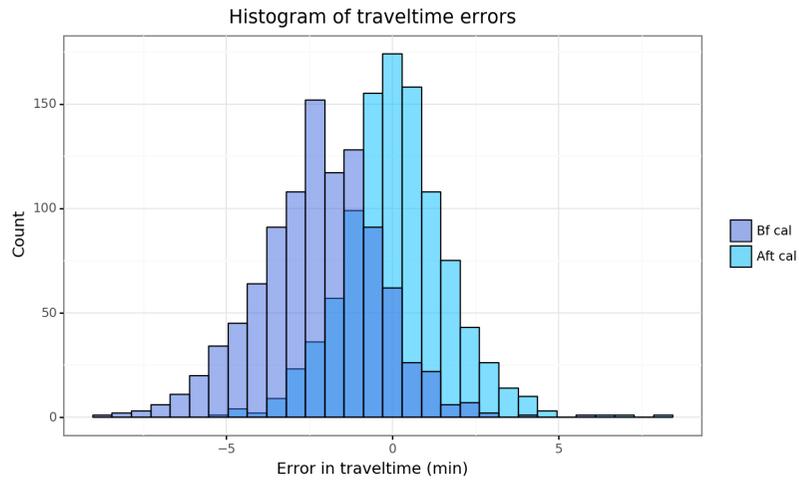


Figure F.21: Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE model calibration without filter

Figure F.22: Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE model calibration with MAXD-1 filter

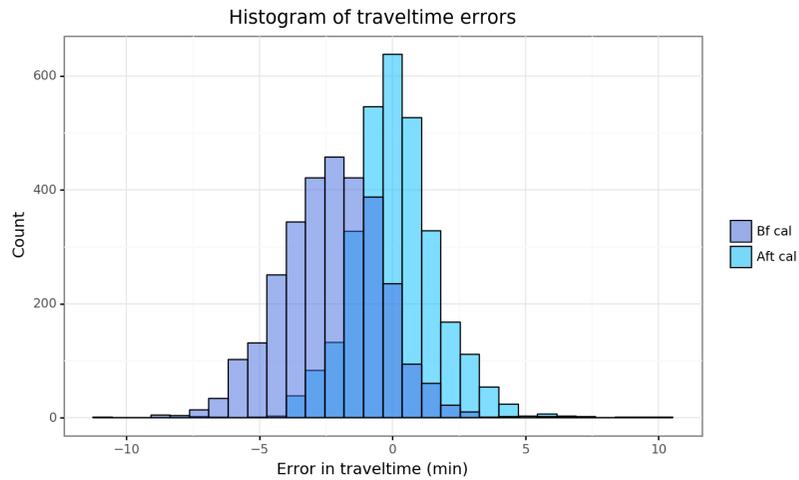
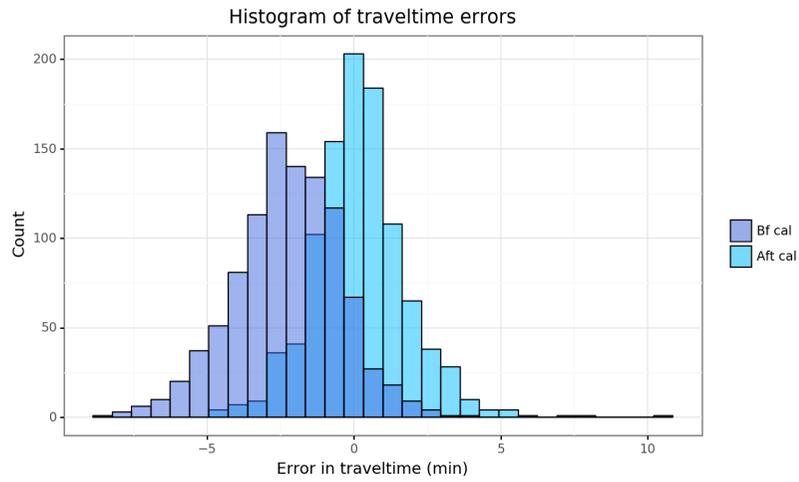


Figure F.23: Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE model calibration with MAXDP-0.1 filter

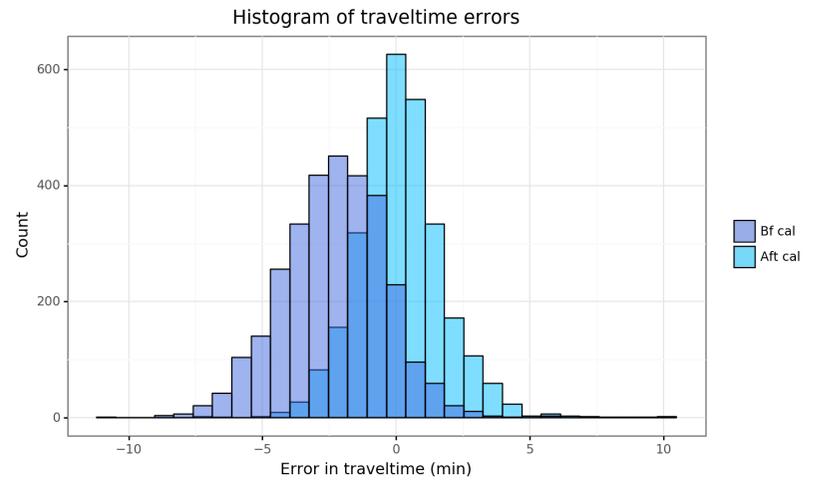
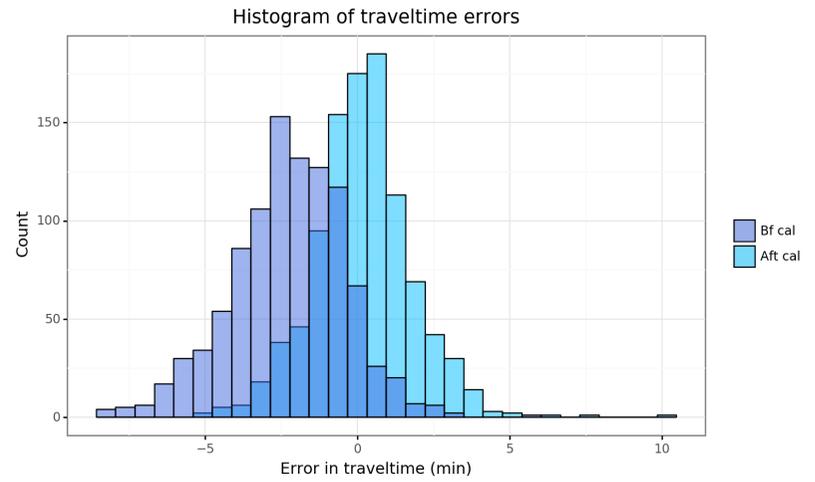


Figure F.24: Total traveltime error distr. for BCN-R-1000 and BCN-R-3000 with TT-SE model calibration with DYN-EUCL filter