AET Association for European Transport
2021 EUROPEAN TRANSPORT CONFERENCE ONLINE
An even greater opportunity to participate

Using real-time GPS data to compare variability of speed in five European cities

Yan Cheng — Associate Lecturer(Teaching)
Paulo Anciaes — Senior Research Fellow
Nasma Hannawi — PhD. Candidate
All at University College London
UCL

BACKGROUND

1

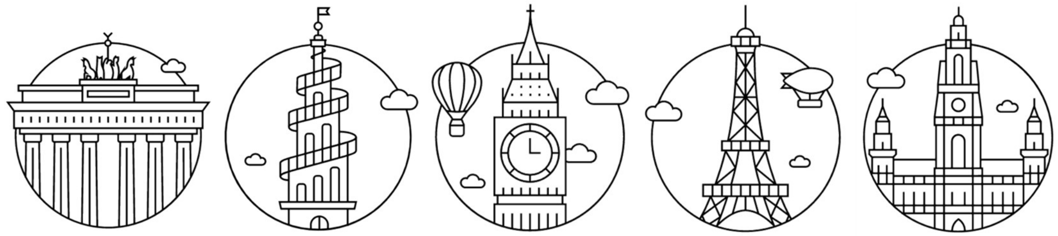
In large cities, traffic speeds are usually lower than what the road design allows for - a problem known as congestion. But cities are now also interested in liveability and the quality of public places. Achieving this requires precisely low traffic speeds, to provide for a safe and amenable environment for pedestrians and street activities. Thus, low speeds may be regarded as a problem only in some roads, in some parts of the city, and at some times, depending on the mix of motorised vs. non-motorised road users.

BACKGROUND

Achieving a good balance of high and low speed for different users requires a good understanding of how speeds vary in space and time. This information can only be provided by detailed data, i.e. covering whole cities (not just commuting corridors) at small intervals (not just hourly averages) and during a large period (not short-term surveys). Having detailed data comparable across cities is even more useful - it allows for city benchmarking. This type of detailed data is now collected by private companies, but it is seldom available for decision-makers and practitioners.

2 DATA

The dataset used in our study contains real-time speeds obtained from GPS probe data from vehicles. The dataset has 2.3 billion data points, covering all 5-minute periods during the whole of 2017 in 47616 road segments (representing 12,229 km) across five cities: Berlin, Copenhagen, London, Paris and Vienna. The dataset was made available by INRIX (a data provider) in the context of a project involving the governments of the five cities, universities, and consultants.



DATA

The location of each segment and its functional classification in a 5-point scale where 1 is the most important for connectivity are also provided.

The road segments in the five cities were divided into three zones: centre and inner and outer functional area, based on the locations of public transport fare zones, ring roads, and circular underground, train lines or bus routes.

DATA

	Berlin	Copenhagen	London	Paris	Vienna
Segments by area					
Centre	169	224	1513	311	161
Inner (not centre)	1220	2240	2693	3775	1608
Outer	5975	2941	16172	10583	3856
Segments by road type					
Level 1	182	228	58	915	221
Level 2	1605	1726	3208	3203	1508
Level 3	4041	1765	15848	10183	2925
Level 4	1390	1328	1213	367	909
Level 5	146	358	51	1	62
Total number of segments (Level 1,2,3)	5828	3719	19114	14301	4654
Total length (km) (Level 1,2,3)	2071.63	1100.78	4685.23	3193.17	1178.65
Average segment length (Level 1,2,3)	0.355	0.296	0.245	0.223	0.253
Total area (km^2)	891	503	1572	763	415

METHODOLOGY

3

We used the data for area-wide modelling of speed within and across cities, using a common typology of areas and road types for the five cities. The analysis of the data had three steps.

METHODOLOGY

Step 1) Decomposition of speed variation

Prophet model

The hourly average speed of each city V(t) is decomposed into five factors: trend, weekly and hourly seasonality, holidays and random error.

Multiplicative model was used because 1) it can avoid the influence of different free-flow speed in five cities, thus adjust the time series to a comparable scale; and 2) the seasonality and random error are assumed to be dependent on trend

$$V(t) = g(t) \times s_w(t) \times s_d(t) \times h(t) \times \varepsilon_t$$

The trend g(t) shows a relatively upward or downward movement over a period of time.

The seasonality is the regular increases and decreases within the year. Multiple seasonality is very common in transport-related data. Fourier series are used to model the periodic effects in Prophet to add seasonality flexibly.

The effect of holidays which occur irregularly in one year and potentially vary among countries.

METHODOLOGY

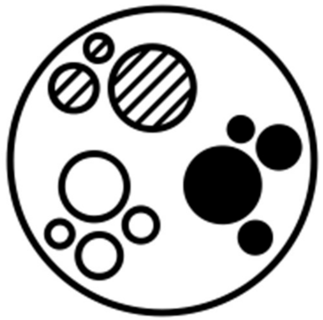
Step 2) Clustering of road segments in each city according to how speeds vary

We used the results of the city-level time series analysis to identify these months, days, and hours. Then we used the means and standard deviations of different time period as input variables. To remove the influence of different scales of variables, all variables were standardised using Z-score normalisation before the cluster analysis.

Standard K-Means minimises the intra-cluster sum of squared Euclidean distances and maximizes the inter-cluster sum of squared Euclidean distances when given the number of clusters K. Each segment was assigned an equal weight, which means all the segments are with the same importance.

Weighted K-Means algorithm

It is extended from standard K-Means by assigning segment length as sample weight. This allow us to take into account the fact that some segments are longer than the others, which implies having higher importance.



METHODOLOGY

Step 3) Regression models to explain how the variability of speeds depends on location (i.e. distance to city centre), road types, and network effects.

Spatial analysis (Multinomial logit model)

We mapped the clusters and analysed the spatial distribution by calculating the mean distance of the observations in each cluster to the city centre and the proportion of each cluster in each road level and region, and estimating models of the probability belonging to each cluster. We used multinomial logit models where the dependent variable is the log odds of a segment belonging to a given cluster, relative to an omitted cluster. The explanatory variables include:

- Dummies representing roads with functional classification levels 1 and 2
- Dummy variables representing the city centre and the inner zone
- Distance to the city centre (for segments in the inner and outer zones)
- The proportion of road segments in each cluster within 1km (with the proportion of the omitted cluster also being omitted).
- Segment length

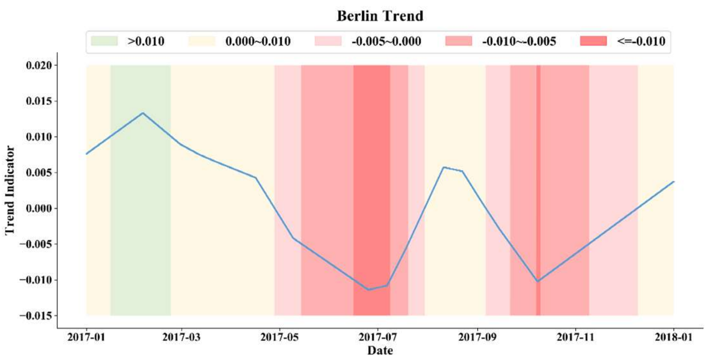
RESULTS

4

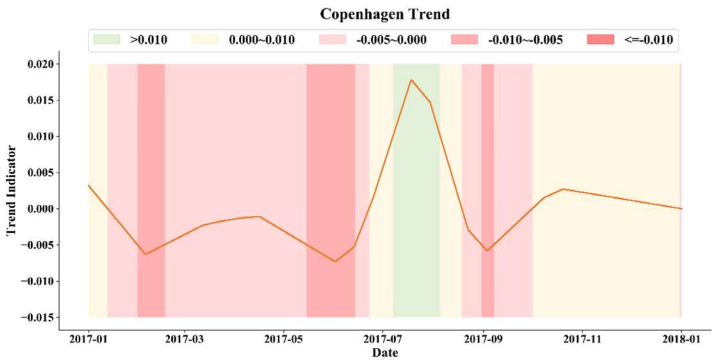
1. Time patterns of congestion variability
2. Types of congestion
3. Spatial patterns of congestion

RESULTS FOR STEP 1

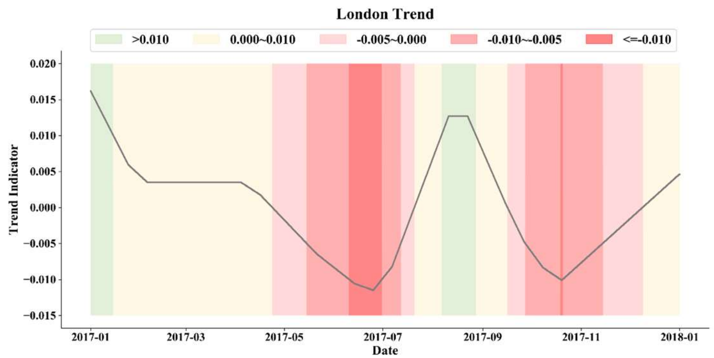
Time patterns of congestion variability - Trend



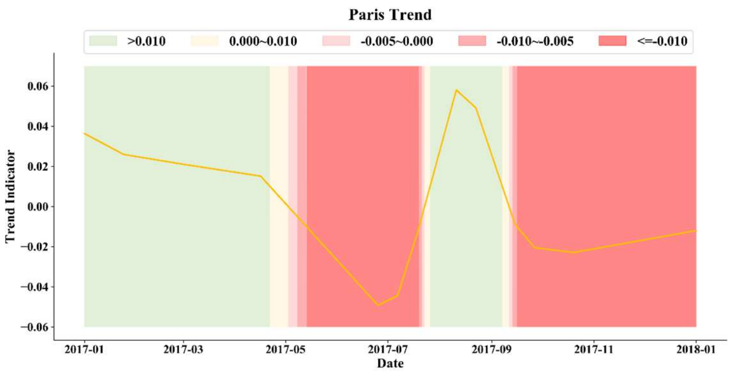
Berlin



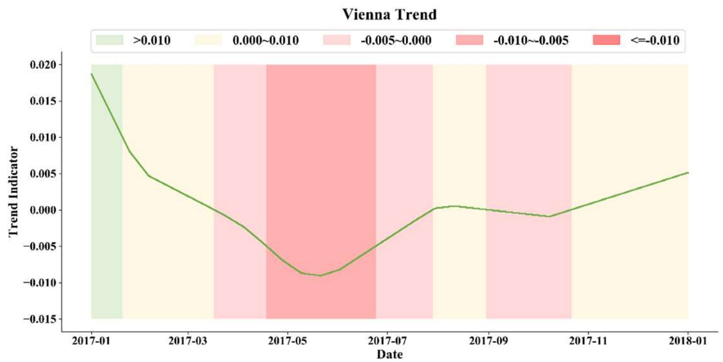
Copenhagen



London



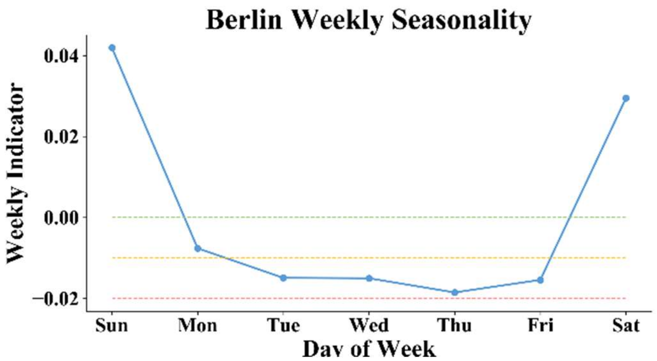
Paris



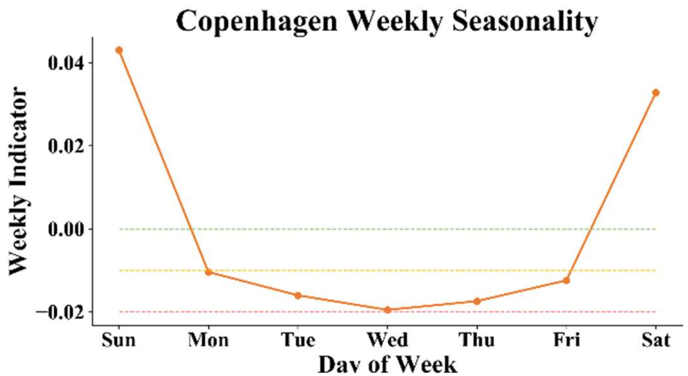
Vienna

RESULTS FOR STEP 1

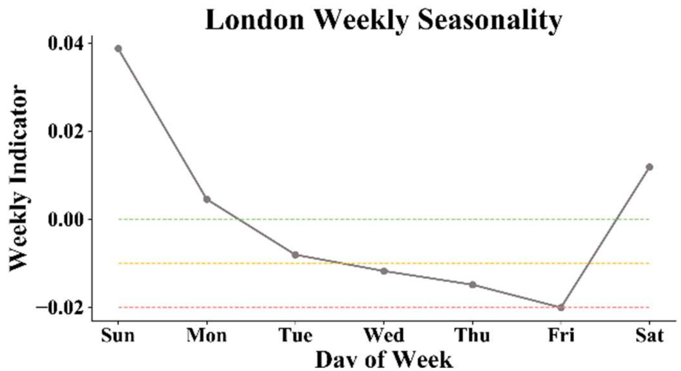
Time patterns of congestion variability - Weekly Seasonality



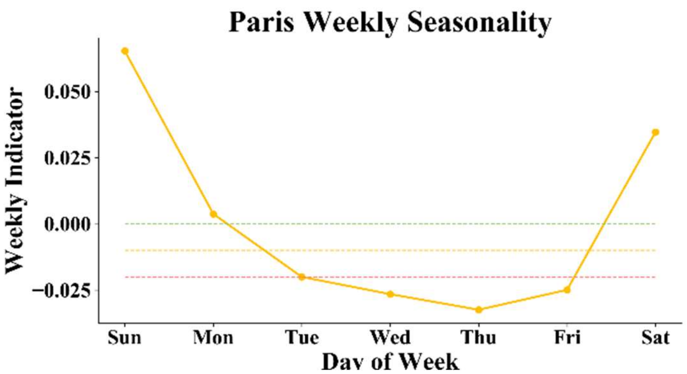
Berlin



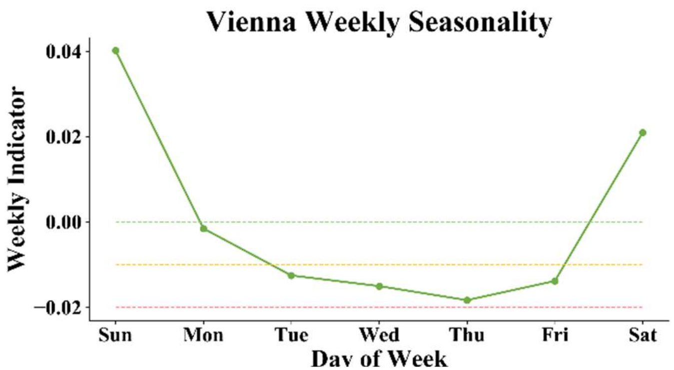
Copenhagen



London



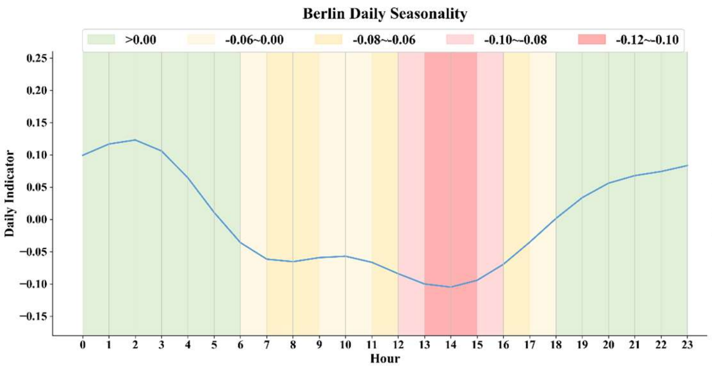
Paris



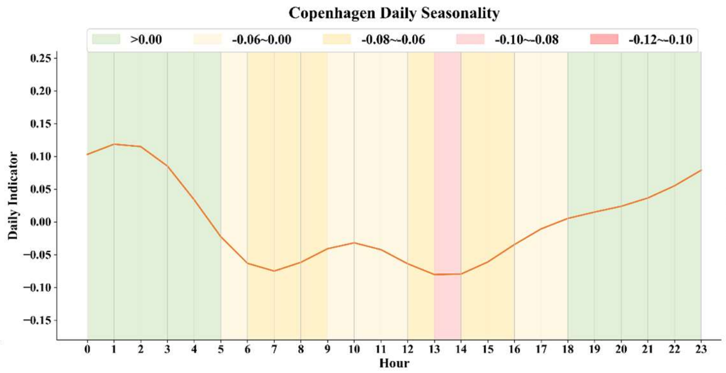
Vienna

RESULTS FOR STEP 1

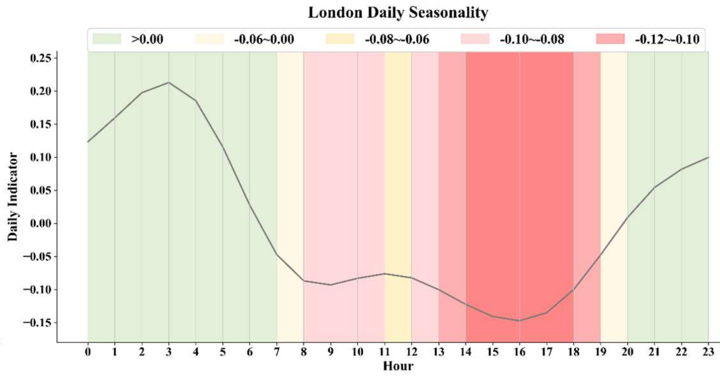
Time patterns of congestion variability - Daily Seasonality



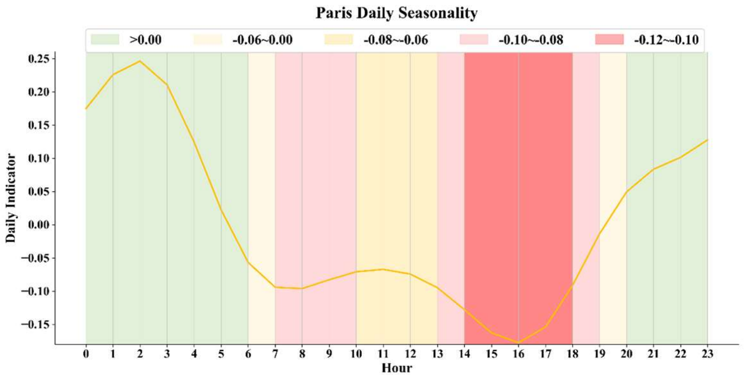
Berlin



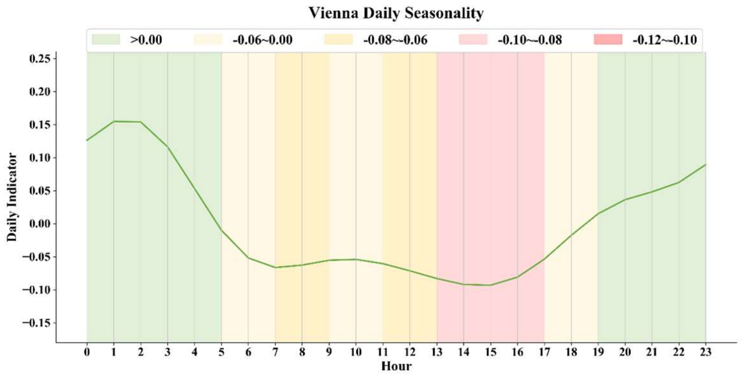
Copenhagen



London



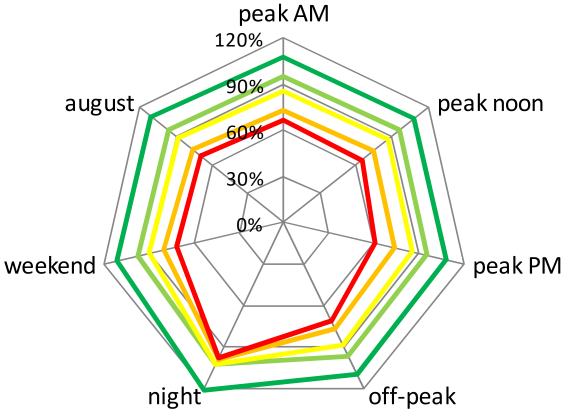
Paris



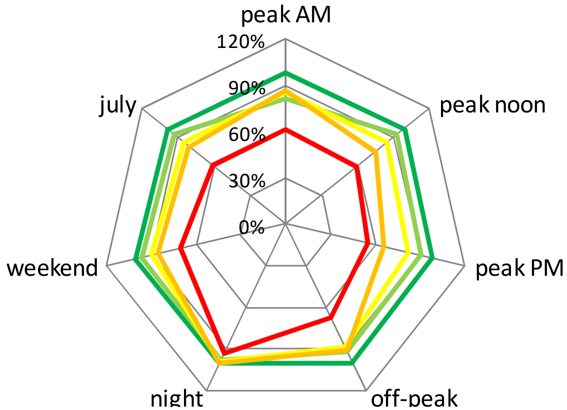
Vienna

RESULTS FOR STEP 2

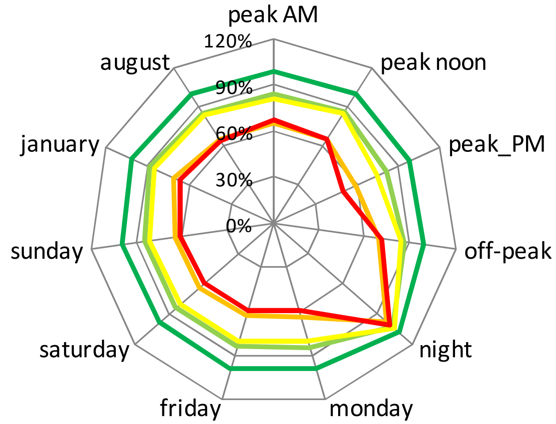
Types of congestion - Mean values among the observations in each cluster



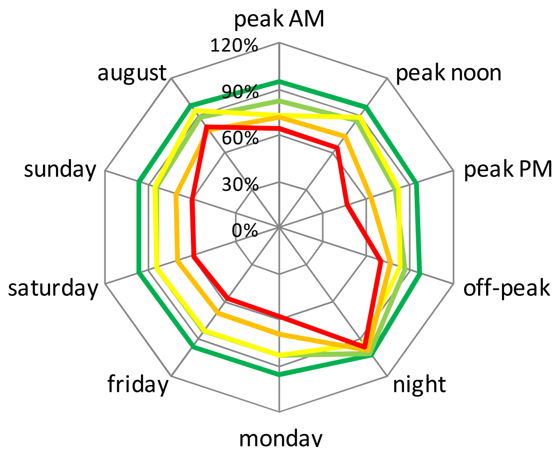
Berlin



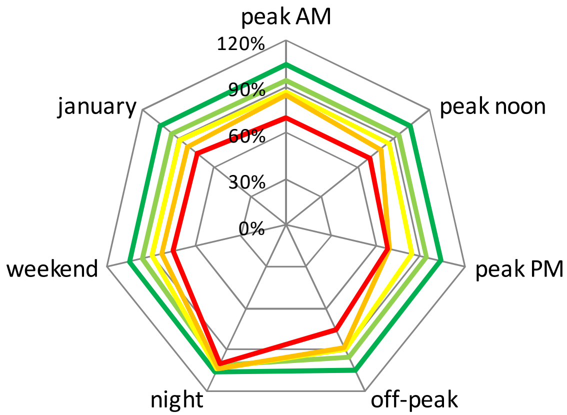
Copenhagen



London



Paris

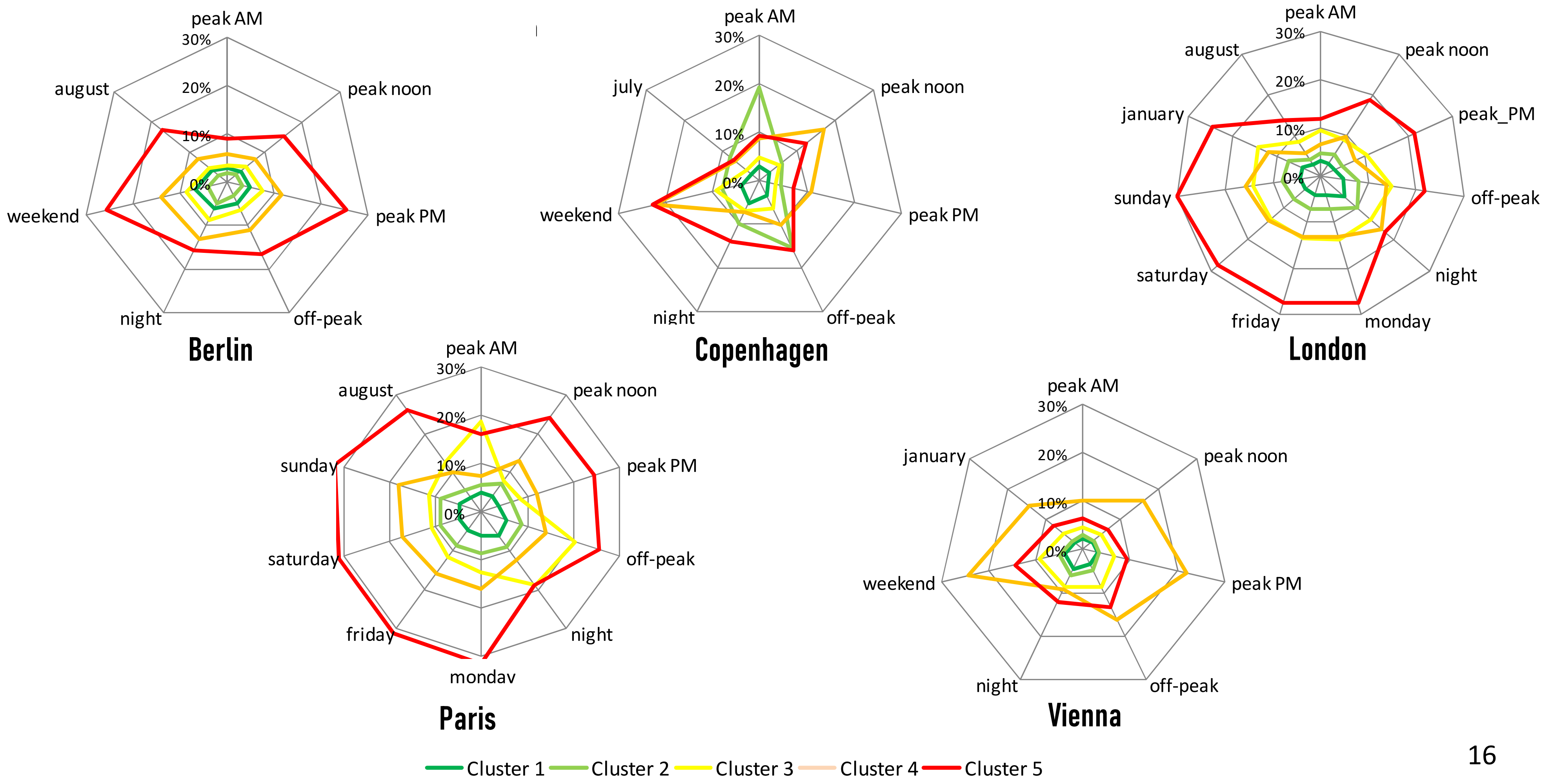


Vienna

Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

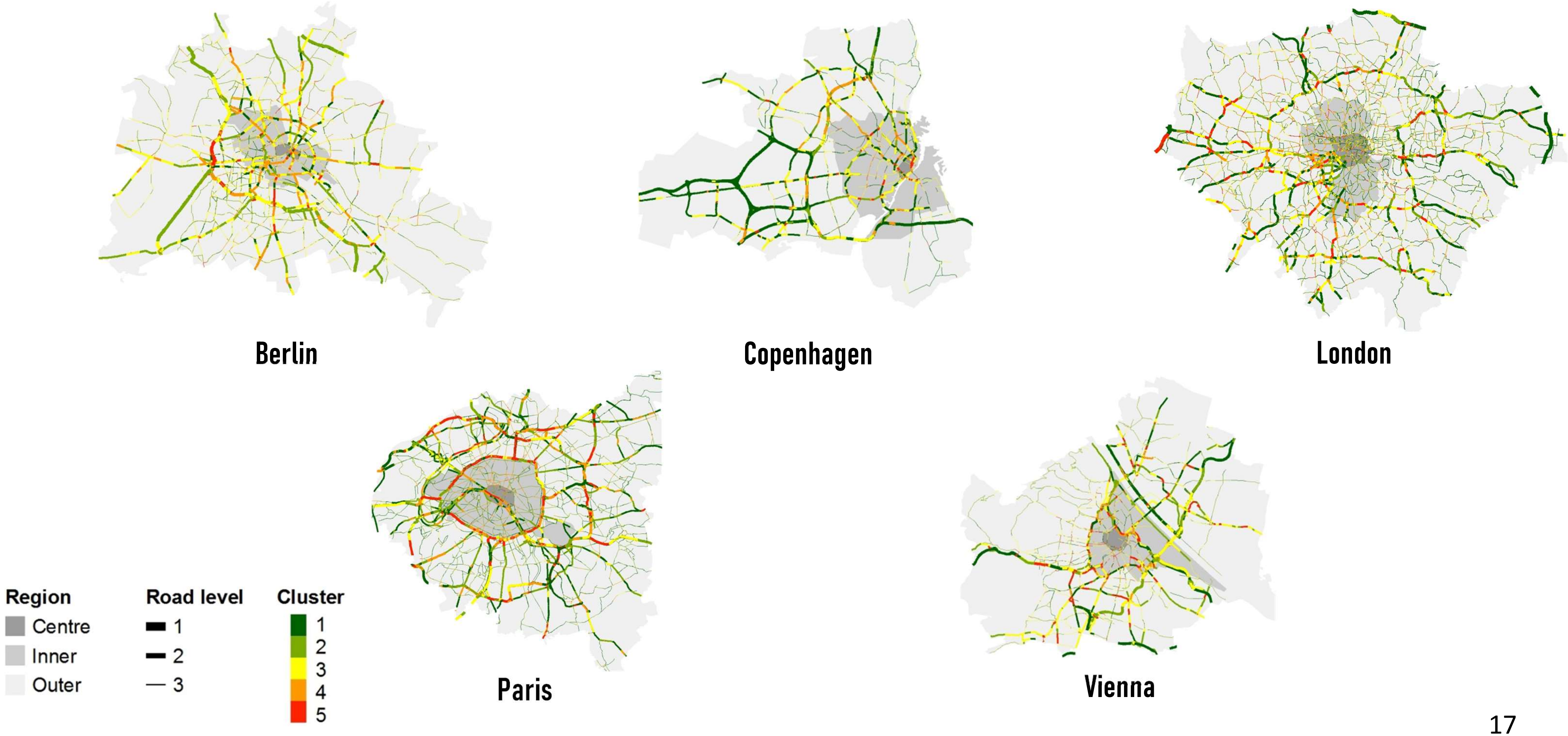
RESULTS FOR STEP 2

Types of congestion – Standard deviations among the observations in each cluster



RESULTS FOR STEP 3

Spatial patterns of congestion – Maps



RESULTS FOR STEP 3

Spatial patterns of congestion – Regression models

	Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Coeff.	p	Coeff.	p	Coeff.	p	Coeff.	p
Berlin								
Level 1	0.62	0.28	0.40	0.50	0.66	0.30	2.33	<0.01***
Level 2	-0.24	0.23	-0.32	0.10	0.30	0.16	0.56	0.08*
Centre	-0.32	0.70	0.17	0.79	0.69	0.26	0.67	0.45
Inner	-0.26	0.36	0.07	0.80	0.13	0.66	-0.45	0.36
Distance to centre	0.17	<0.01***	0.11	<0.01***	0.10	<0.01***	0.10	0.04**
Neighbours								
% cluster 2	2.53	<0.01***	1.77	<0.01***	0.99	0.03**	0.58	0.57
% cluster 3	1.75	<0.01***	2.84	<0.01***	2.45	<0.01***	2.75	<0.01***
% cluster 4	1.12	0.01**	2.43	<0.01***	2.71	<0.01***	3.35	<0.01***
% cluster 5	0.50	0.53	2.29	<0.01***	2.91	<0.01***	3.90	<0.01***
Length	0.34	0.17	0.11	0.66	-0.77	0.01***	-1.13	0.01***
Constant	-1.67	<0.01***	-1.56	<0.01***	-1.66	<0.01***	-3.43	<0.01***

RESULTS FOR STEP 3

Spatial patterns of congestion – Regression models

	Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Coeff.	p	Coeff.	p	Coeff.	p	Coeff.	p
Copenhagen								
Level 1	2.06	<0.01***	-1.69	<0.01***	1.61	<0.01***	-13.63	0.99
Level 2	1.27	<0.01***	-0.08	0.66	0.86	0.01***	0.19	0.59
Centre	-1.88	0.22	0.40	0.58	0.84	0.35	0.68	0.52
Inner	-0.57	0.18	0.20	0.41	0.32	0.45	0.42	0.46
Distance to centre	-0.15	<0.01***	-0.02	0.34	-0.11	0.02**	-0.03	0.56
Neighbours								
% cluster 2	2.05	<0.01***	1.71	<0.01***	1.73	0.01***	3.46	0.01***
% cluster 3	1.17	0.02**	1.61	<0.01***	0.90	0.07*	2.25	0.01**
% cluster 4	1.57	0.01**	1.30	0.01***	2.27	<0.01***	3.57	<0.01***
% cluster 5	2.87	<0.01***	1.73	<0.01***	2.80	<0.01***	4.63	<0.01***
Length	-0.88	0.02**	-0.87	<0.01***	-0.77	0.04**	-5.68	<0.01***
Constant	-1.85	0.01***	-0.45	0.19	-2.16	<0.01***	-1.74	0.08*

RESULTS FOR STEP 3

Spatial patterns of congestion – Regression models

	Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Coeff.	p	Coeff.	p	Coeff.	p	Coeff.	p
London								
Level 1	-2.21	<0.01***	-0.94	0.09*	-2.04	0.28	0.24	0.69
Level 2	-0.90	<0.01***	-0.06	0.58	-0.48	<0.01***	0.37	0.01**
Centre	-0.55	0.01	-0.08	0.77	-0.74	0.01	-0.27	0.45
Inner	-0.36	0.01**	-0.81	<0.01***	-0.50	0.01***	-0.90	<0.01***
Distance to centre	-0.03	<0.01***	-0.01	0.21	-0.06	<0.01***	-0.01	0.29
Neighbours								
% cluster 2	1.48	<0.01***	1.11	<0.01***	2.88	<0.01***	1.32	<0.01***
% cluster 3	1.12	<0.01***	1.44	<0.01***	2.22	<0.01***	1.59	<0.01***
% cluster 4	2.36	<0.01***	1.76	<0.01***	5.38	<0.01***	3.11	<0.01***
% cluster 5	1.31	<0.01***	1.65	<0.01***	3.56	<0.01***	2.16	<0.01***
Length	-0.29	0.01***	-0.34	0.01**	-1.32	<0.01***	-0.60	<0.01***
Constant	-0.10	0.53	-1.07	<0.01***	-1.82	<0.01***	-2.23	<0.01***

RESULTS FOR STEP 3

Spatial patterns of congestion – Regression models

	Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Coeff.	p	Coeff.	p	Coeff.	p	Coeff.	p
Paris								
Level 1	0.14	0.46	2.21	<0.01***	1.16	<0.01***	3.79	<0.01***
Level 2	-0.16	0.16	0.86	<0.01***	0.25	0.08*	1.62	<0.01***
Centre	-0.22	0.67	-1.53	0.31	0.73	0.16	1.41	0.04
Inner	-0.10	0.48	-0.69	0.02**	0.15	0.43	0.10	0.74
Distance to centre	-0.06	<0.01***	-0.05	0.07*	-0.07	<0.01***	-0.12	<0.01***
Neighbours								
% cluster 2	1.89	<0.01***	1.06	<0.01***	2.53	<0.01***	0.87	0.04**
% cluster 3	1.06	<0.01***	1.68	<0.01***	2.47	<0.01***	1.06	0.06*
% cluster 4	2.36	<0.01***	2.34	<0.01***	4.21	<0.01***	2.09	<0.01***
% cluster 5	0.58	0.19	1.05	0.05**	2.15	<0.01***	1.19	0.03**
Length	-0.09	0.49	-0.61	0.01***	-1.05	<0.01***	-0.91	<0.01***
Constant	-0.50	0.01***	-2.42	<0.01***	-2.03	<0.01***	-2.92	<0.01***

RESULTS FOR STEP 3

Spatial patterns of congestion – Regression models

	Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Coeff.	p	Coeff.	p	Coeff.	p	Coeff.	p
Vienna								
Level 1	-1.88	<0.01***	-1.01	<0.01***	2.50	<0.01***	-1.40	0.09*
Level 2	-0.66	<0.01***	-0.37	0.10	2.20	<0.01***	0.24	0.42
Centre	0.25	0.77	-0.24	0.77	1.02	0.42	-0.43	0.66
Inner	0.12	0.69	0.01	0.98	-0.04	0.94	-0.09	0.82
Distance to centre	0.10	0.04**	0.01	0.77	-0.35	0.01***	-0.08	0.24
Neighbours								
% cluster 2	2.14	<0.01***	1.95	<0.01***	1.81	0.04**	1.51	0.07*
% cluster 3	1.72	<0.01***	3.01	<0.01***	2.16	0.02**	3.39	<0.01***
% cluster 4	1.70	0.05*	2.03	0.02**	1.70	0.17	1.18	0.50
% cluster 5	0.40	0.54	2.76	<0.01***	1.24	0.38	3.68	<0.01***
Length	-0.05	0.85	-0.57	0.05**	-0.57	0.33	-2.98	<0.01***
Constant	-0.60	0.16	-0.87	0.06*	-2.97	0.01***	-1.09	0.18

DISCUSSION 5

Which months have more congestion?

Jun-Jul and Sep-Oct in all cities, but with different level of congestion. Copenhagen also had an extra congestion period in Feb.

Particularly, the congestion is more severe and lasts longer in Paris; while the congestion from Sep-Oct in Vienna is less severe than the others.

There was also a smaller decrease in December and January, especially in Berlin and London.

DISCUSSION 5

Which days have more congestion?

Basically, weekdays are more congested than weekends.

Specifically speaking, Berlin, Paris and Vienna had more congestion on Thursdays, while London is more congested on Fridays and Copenhagen on Wednesdays.

This may be due to the different lifestyles in the five cities.

DISCUSSION

5

When does congestion start and how long it lasts?

London and Paris had shorter(3 hours) congestion in the morning, but the one in London(8 am) is later than Paris (7 am). Both cities had longer (4 hours) in the early evening, starting from 2 pm. Vienna only had evening congestion but with lower level of congestion.

Berlin only had a 2-hour congestion around lunchtime, from 1 -2 pm, while Copenhagen followed the similar pattern, but with lower level of congestion.

DISCUSSION 5

How does congestion vary in space?

In Paris, it is more concentrated in ring roads. In London, it is more dispersed. Congestion tends to be closer to the centre in Paris, Vienna, and Copenhagen.

Distance to centre are significant and negative for almost all clusters in all cities except Berlin.

Data also revealed network effects. The higher proportions of neighbouring segments belonging to clusters 2, 3, 4, and 5, the higher probability of not belonging to cluster 1, in most cases.

CONCLUSIONS 6

The analysis of real-time data from the five cities revealed patterns that can aid decision-makers designing future transport strategies. By showing the exact location and timing of congestion, detailed data facilitates:

1) The design of targeted interventions (e.g. variable road pricing applying differently by road segment and time of day)

2) Comparison of traffic speeds with data on walking and street activities in each segment and time - informing the definition of solutions balancing the needs of motorised and non-motorised users

3) Direct comparison with other cities, pinpointing how cities fare against others, in similar areas and road types, and at the same times of day

None of these aspects could be fully addressed with the aggregate, unstandardized data that decision-makers and practitioners usually have available. The main lesson learnt from the data analysis is therefore that a closer integration of strategies and operation of data providers with governments, universities, and consultants could benefit all parties.

AET Association for European Transport

2021 EUROPEAN TRANSPORT CONFERENCE ONLINE

An even greater opportunity to participate

Thank you for your time and attention.

Any questions/comments/suggestions?

Should you have any follow-up questions, please feel free to contact me:

Yan Cheng   yan.cheng@ucl.ac.uk