

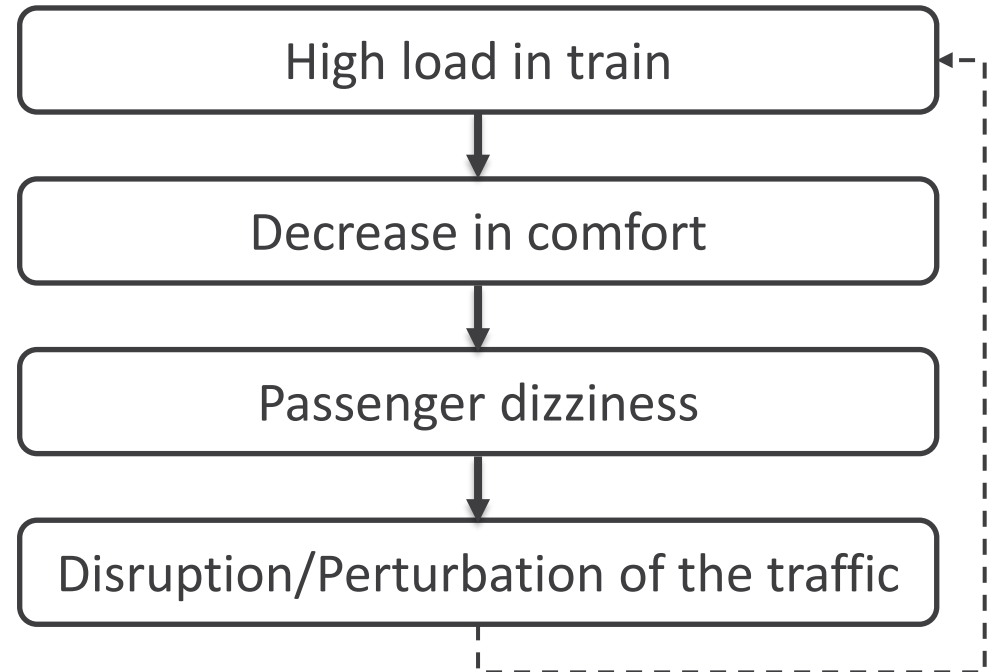
# Impact of passenger load on disruption probability in suburban railway services

Mattéo GUILLEMARD – Mehdi BAALI – Rémi  
COULAUD – Christine BUISSON  
*TransitData, 03/07/2024*



# Context

## Introductory example



# Primary disruption

## Definitions

Disruption  $\neq$  delay (caused by a disruption)

Primary disruption  $\neq$  secondary disruption  
(delay caused by the delay of another train)

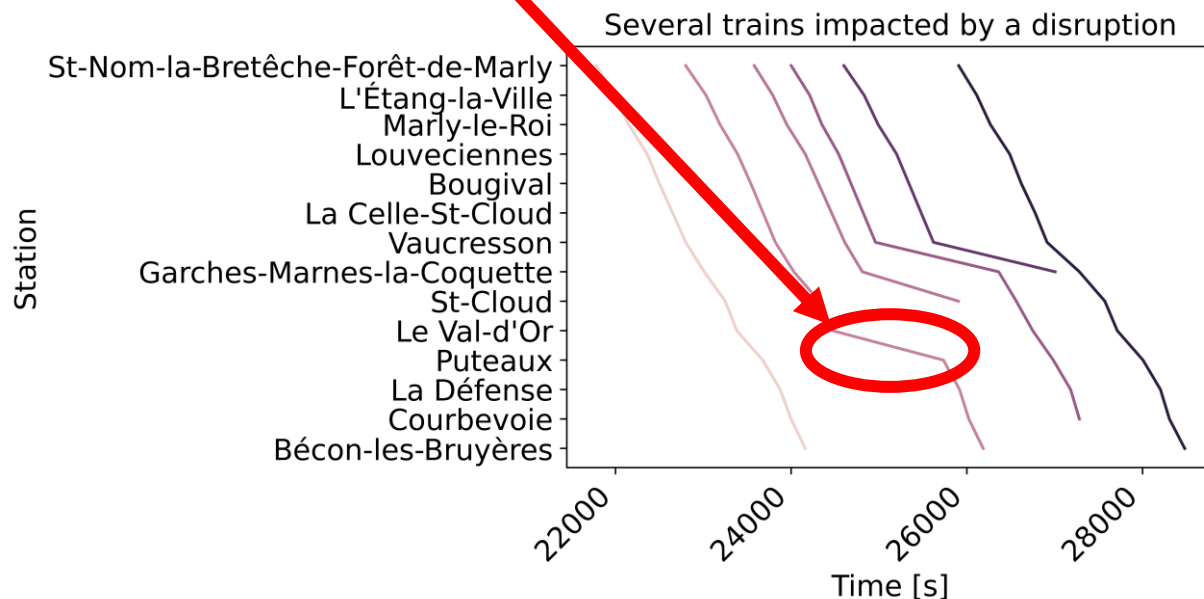
(Palmqvist, 2023)

## Criteria

**Disruption** = when a train undergoes a **delay increase of more than 5min** between two stops

**Primary disruption** = when the **train right before underwent no disruption** during its whole travel

## Primary disruption



# Context

## Goals

Prediction of the **probability of passenger-linked primary disruption**

Study the specific relationship with **passenger load**

## Literature review

Authors(s)	Year	Context	Prediction Scale	
Goverde, Hansen, Hooghiemstra & Lopuhaä	2001	Train delays	No	Stop
Yuan	2006	Train delays	No	Stop
Spanninger, Trivella, Büchel & Corman	2022	Train delays	Yes	
Vere-Jones	1995	Earthquakes	Yes	
Berthe	2022	Injuries in sport	Yes	
Zou & Yue	2017	Road	Yes	
Fink, Zio & Weidmann	2014	Infrastructure-linked train disruption	Yes	Equipment
Yap & Cats	2019, 2021	Train disruption	Yes	Station, Time period

# Data

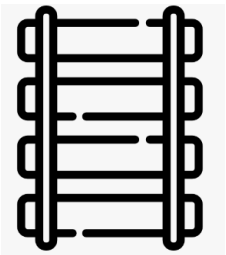


**Stop scale** = 1 direction, 1 date, 1 train, 1 station

3 specific stop-scale data sources:

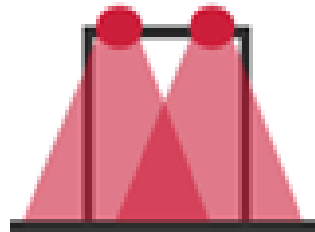
## Transport plan

From on-rail tag systems



## Load

From passenger counting  
infrared sensors

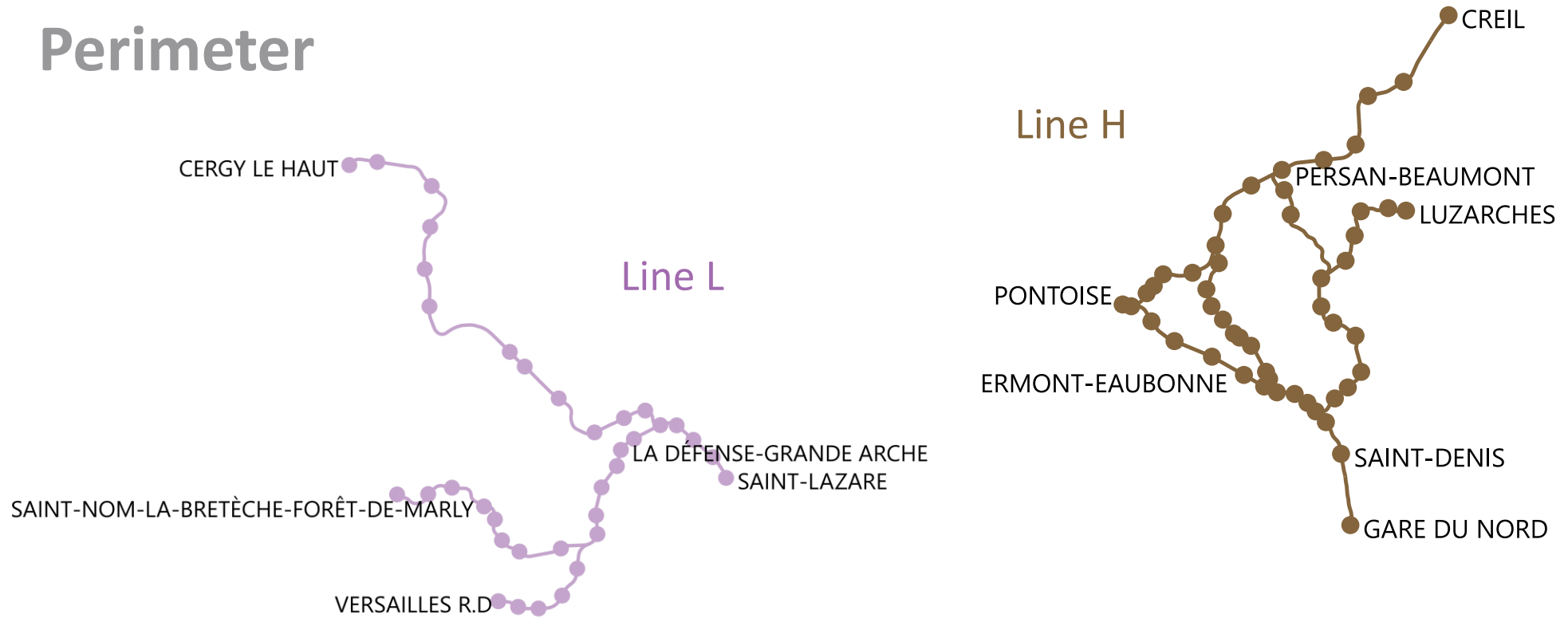


## Disruption details

Manually reported



# Perimeter

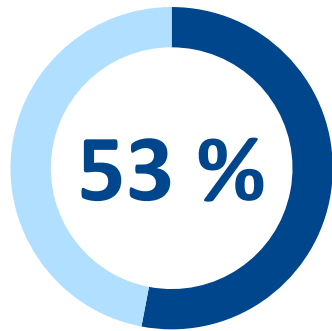


*6 months of data: September 2022 – February 2023*

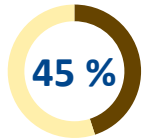
In this perimeter: **1,200,087 observations**  
1 observation = 1 direction, 1 date, 1 train, 1 station

# Primary disruptions in this perimeter

2439 events of primary disruption

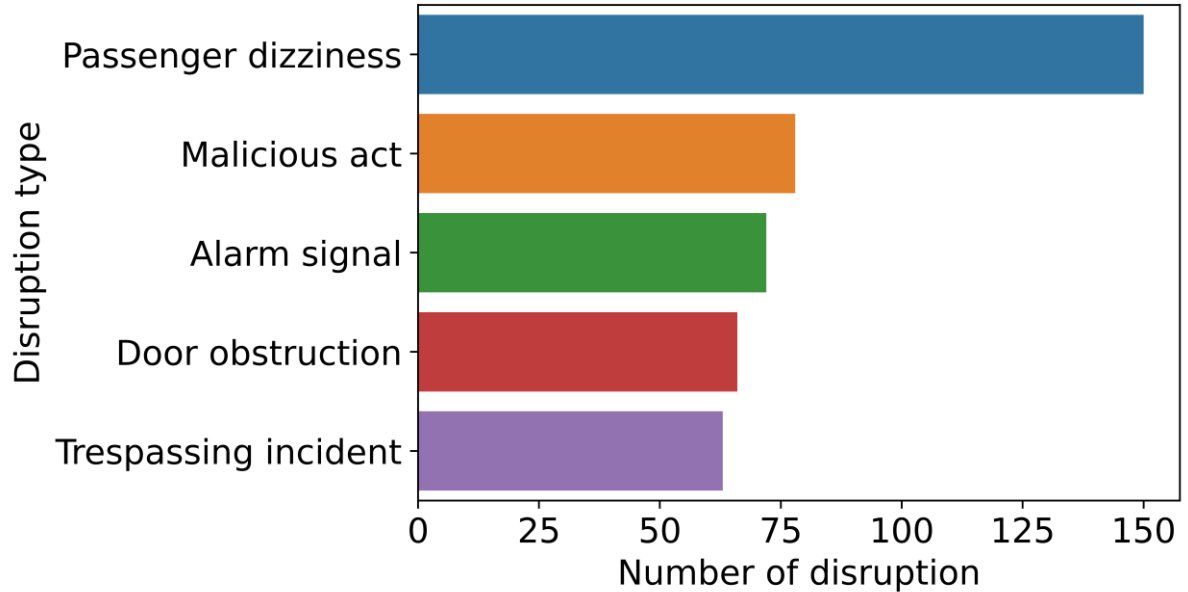


53% of the delays are **caused by primary disruptions**



45% in a case study in the Netherlands (*Weeda, 2006*)

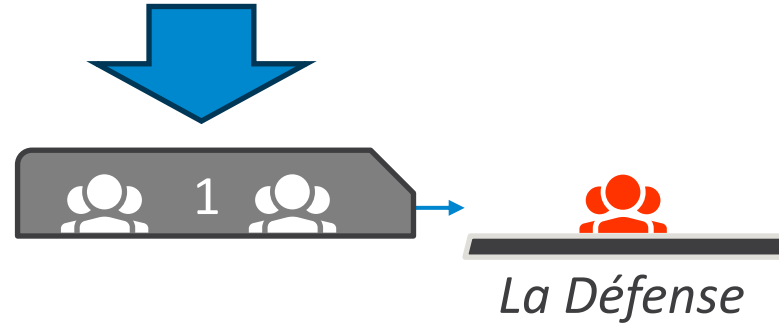
Distribution of primary disruption type by occurrence



# Passenger load

## Load

Focus on the **load** before arriving at a stop

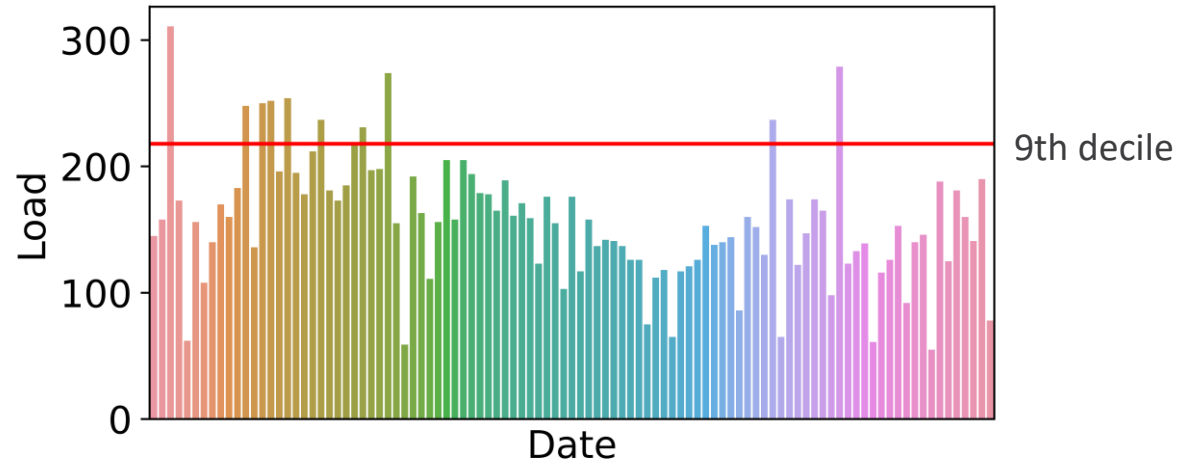


## Relative load

**Relative load:** decile of the load everything else being equal but the date (direction, train, station, day type)

**≠ Absolute load**

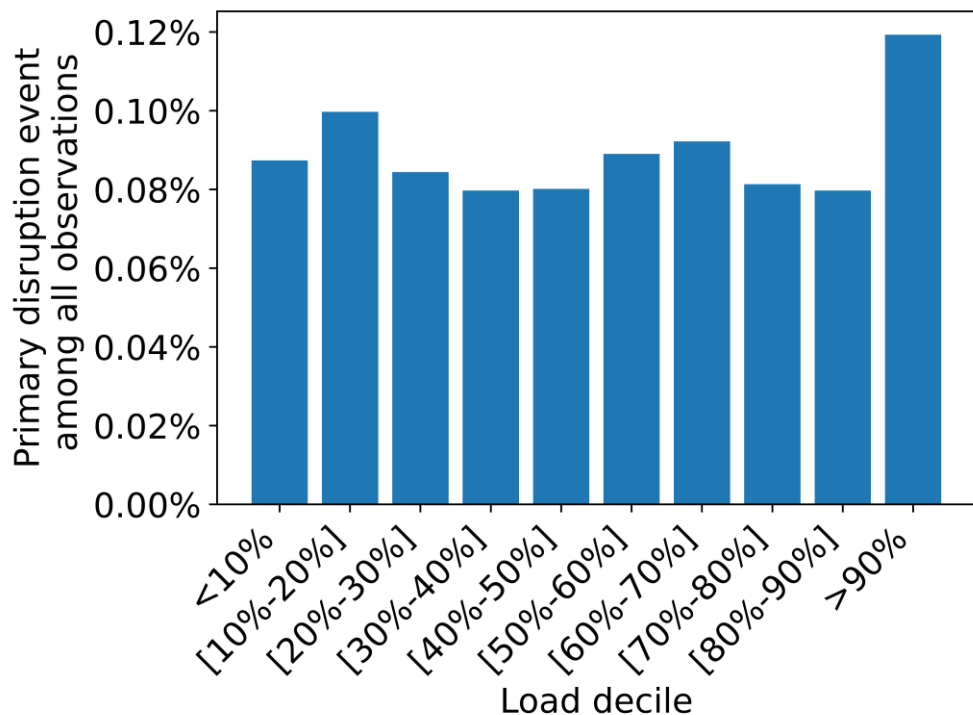
Load arriving at La Défense, weekdays, train 134509





# Preliminary observation

## Correlation



	<9th decile	>9th decile
No disruption	89 %	11 %
Passenger-linked primary disruption	82 %	<b>18 %</b>

→ Disruption more frequent over the 9th decile in load

→ **Chi-squared test**: reject the independence at the 5% threshold

→ **Varying** with disruption type: **24% for passenger dizziness**

# Logistic model

- Specific focus on passenger-linked disruptions
- **Isolate** the specific impact of the load
- Control variables **from the literature** (*Yap & Cats, 2021*)

$$\ln \left( \frac{P(y = 1|X)}{1 - P(y = 1|X)} \right) = \text{logit}(P(y = 1|X)) = \beta_0 + \beta_1 * x_1 + \sum_{j=2}^9 \beta_j * x_j + \varepsilon$$

where  $y$ : disruption dummy,  $x_1$ : absolute passengers load,  $x_i$ : control variables

**Significant:**  $\text{pseudo} - R_{MCF}^2 = 0.31$ ; high according to *Domenichich & McFadden, (1975)*.

# Logistic model

## Parameters

$$\text{logit}(P(y = 1|X)) = \beta_0 + \beta_1 * x_1 + \sum_{j=2}^9 \beta_j * x_j + \varepsilon$$

Variable	Coefficient	Standard error
<b>Absolute passenger load</b>	<b>0.0035***</b>	<b>0</b>
Train frequency	-0.0796***	0.011
Terminus station (dummy)	6.6939***	0.314
Transfer station (dummy)	-0.5137***	0.010
Saturday (dummy)	-0.9285***	0.254
Sunday (dummy)	-1.6990***	0.387
Morning (dummy)	-0.4903***	0.132
Afternoon (dummy)	-0.4206**	0.206
Constant	-9.7706***	0.328

\*\*\*p<0.01; \*\*p<0.05  
1,200,087 observations

# Logistic model

## Summary

**100 more people inside the train lead to an increase in +42% of passenger-linked primary disruption probability (everything else being equal)**

For the specific case of **passenger dizziness**, **+100 people** lead to an increase in **+47%**

→ Similar results with a Poisson regression

→ Impact about **half as much** taking all types of primary disruptions and the model is less well fitted

# Conclusion & Discussion

## Outcomes

---

- ✓ New data enables a study of **primary disruptions** at the stop scale
- ✓ Primary disruption is **more frequent** over the **9th decile in load**
- ✓ A quantitative projection is made thanks to a **logit model**

## Future works

---

- Develop a nested logit model taking into account disruption types
- Extend the study to a larger sample (including more crowded lines)
- Add other relevant control variables (weather...)
- In addition to primary disruptions probability, quantify their impacts (delay, number of secondary disruptions...)

# Bibliography

Weeda, V. A. (2006). Analyse dispunctualiteit, verstoringsregistratie en rij-en halteertijden: Resultaten casestudy Rotterdam-Dordrecht (in Dutch), Report T&P 2006.007, Department Transport & Planning. *Delft University of Technology, Delft*. Goverde, Hansen, Hooghiemstra & Lopuhaä 2001

Yuan, J. (2006). *Stochastic modelling of train delays and delay propagation in stations* (Vol. 2006). Eburon Uitgeverij BV.

Spanninger, T., Trivella, A., Büchel, B., & Corman, F. (2022). A review of train delay prediction approaches. *Journal of Rail Transport Planning & Management*, 22, 100312.

Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *International Journal of Forecasting*, 11(4), 503-538.

Berthe, M. (2021). *Modèles de prédiction d'événements rares en suivi longitudinal. Application au risque de blessure chez les sportifs professionnels* (Doctoral dissertation, Université Clermont Auvergne).

Zou, X., & Yue, W. L. (2017). A bayesian network approach to causation analysis of road accidents using netica. *Journal of advanced transportation*, 2017(1), 2525481.

Fink, O., Zio, E., & Weidmann, U. (2013, September). Extreme learning machines for predicting operation disruption events in railway systems. In *Proceedings of the European Safety and Reliability Conference* (pp. 1-8).

Yap, M., & Cats, O. (2019, June). Analysis and prediction of disruptions in metro networks. In *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 1-7). IEEE.

Yap, M., & Cats, O. (2021). Predicting disruptions and their passenger delay impacts for public transport stops. *Transportation*, 48(4), 1703-1731.

Domenichich, T., & McFadden, D. (1975). A Theory of individual travel demand. *Urban Travel Demand*, 33-46.

# Appendix: other models

Model	Disruption type	Log-likelihood	Pseudo-R2	RMSE	MAE
Logit	Passenger-linked	-1846	0.31	0.16	0.04
Logit	All	-5909	0.33	0.33	0.13
Logit	Non passenger-linked	-4636	0.31	0.28	0.10
Poisson	Passenger-linked	-1850	0,31	0.16	0.04
Random forest	Passenger-linked	NA	NA	0.18	0.03
Random forest	All	NA	NA	0.31	0.10

# Appendix: discussion on the pseudo-R2

	pseudo-R2	R2
Used for	Logit regression	Linear regression
Principle	Proportion of deviation	Proportion of variance
Formula	$1 - \frac{\ln(L_M)}{\ln(L_0)}$	$1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$
Good value	>0.3	0.9 - 1
Reference	<i>McFadden, (1973)</i>	<i>Wright, (1921)</i>