# Detecting outlier demand in railway networks

**Nicola Rennie**[1], Catherine Cleophas[2], Adam M. Sykulski[1], Florian Dost[3]

[1]Lancaster University, [2]CAU Universität zu Kiel, [3]Brandenburg University of Technology

STOR-i
excellence with impact

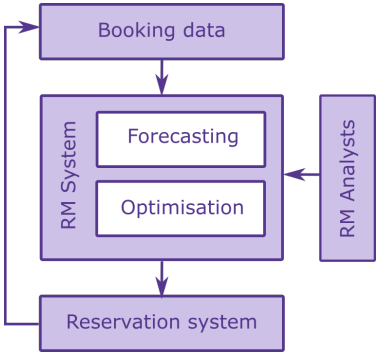
Lancaster University

UKRI Engineering and Physical Sciences Research Council

DB

# Revenue management

- Commonly used by transport service providers to control ticket availability.
- Combines forecasting with optimisation.
- Decision support for RM analysts is needed.



# Network revenue management: Deutsche Bahn

- Focus on long-distance trains.
- Over 1000 stations.
- Over 110,000 possible origin-destinations (O-D).
- Bookings are reported on the leg-level, not O-D.

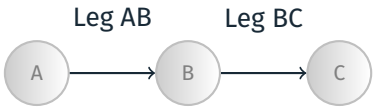
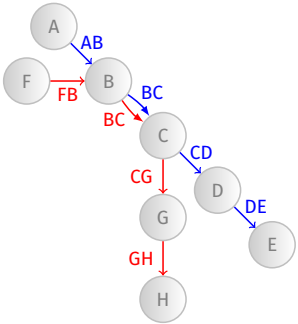


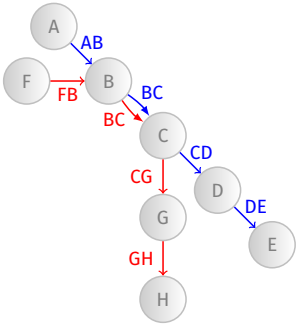

Figure: Railway network with two legs

# Clustering: Graphical representation

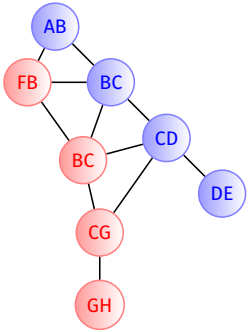



(a) Railway network graph

# Clustering: Graphical representation




(a) Railway network graph




(b) Inverted graph

# Clustering: Leg similarity

The **common traffic ratio** of legs AB and BC is:

$$r(AB, BC) = \frac{D_{AC}}{D_{AB} + D_{BC} + D_{AC}},$$

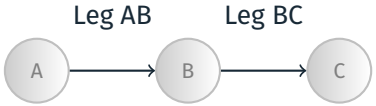



Figure: Railway network with two legs

# Clustering: Leg similarity

The **common traffic ratio** of legs AB and BC is:

$$r(AB, BC) = \frac{D_{AC}}{D_{AB} + D_{BC} + D_{AC}},$$

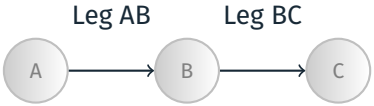
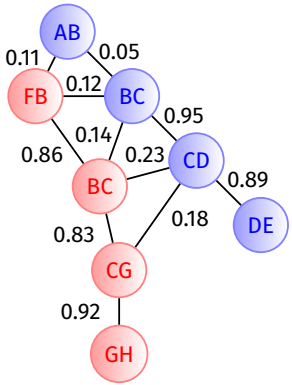



Figure: Railway network with two legs

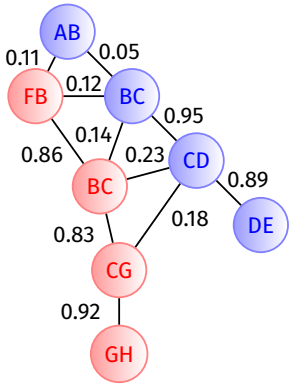
- If $r(AB, BC) = 1$, then the number of bookings on leg AB and leg BC are identical, and the correlation between them is 1.
- Conversely, if $r(AB, BC) = 0$, then the bookings on leg AB and leg BC are independent with correlation 0.
- Estimate $r(AB, BC)$ with correlation.

# Clustering: Minimum spanning tree clustering

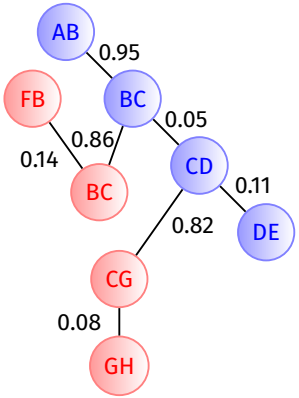



(a) Inverted graph with common traffic ratio

# Clustering: Minimum spanning tree clustering

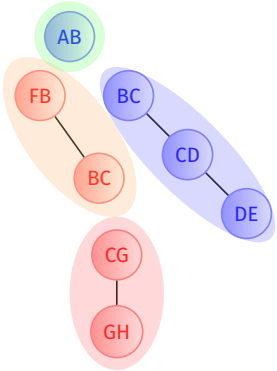



(a) Inverted graph with common traffic ratio



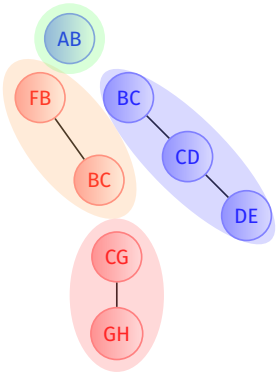

(b) Minimum spanning tree with edge weights

# Clustering: Minimum spanning tree clustering

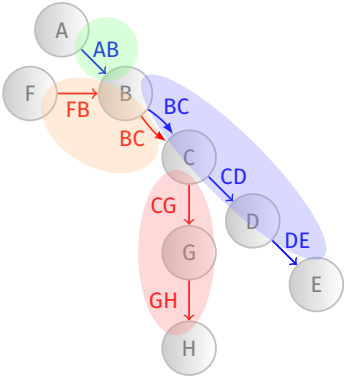



(a) Clusters obtained in inverted

# Clustering: Minimum spanning tree clustering

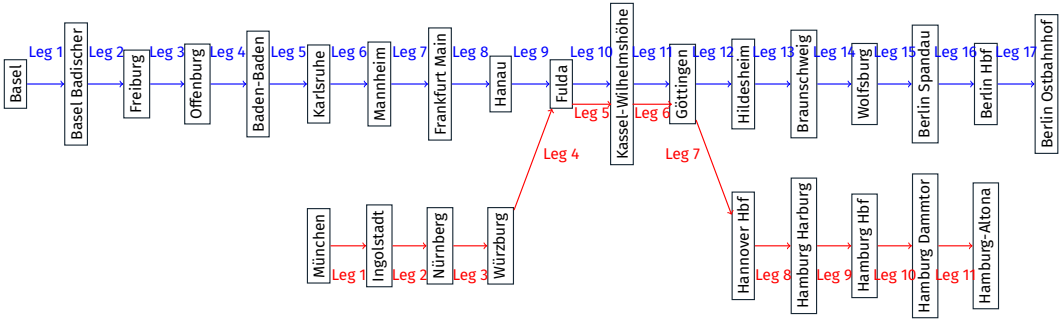



(a) Clusters obtained in inverted

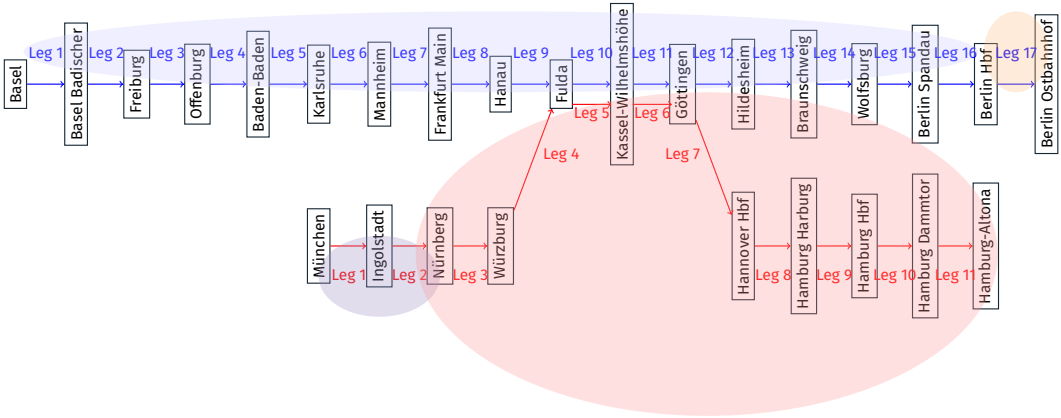



(b) Translation of clusters to original graph

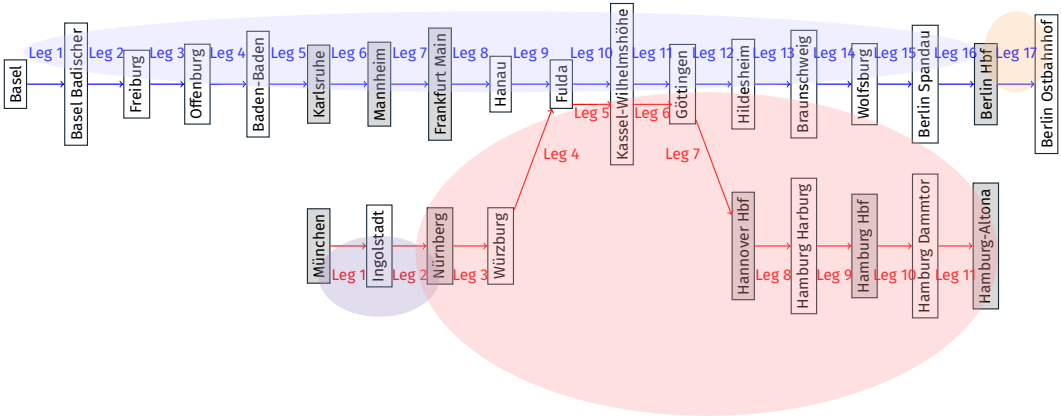
# Clustering: Deutsche Bahn network



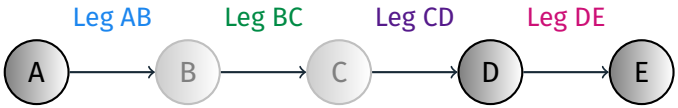
# Clustering: Deutsche Bahn network



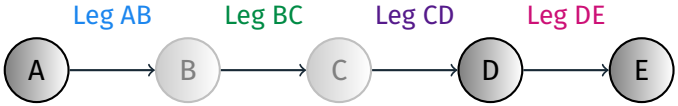
# Clustering: Deutsche Bahn network



# Detecting outliers



# Detecting outliers



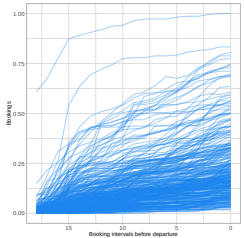




Figure: Leg AB

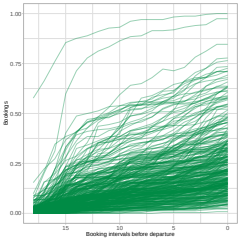



Figure: Leg BC

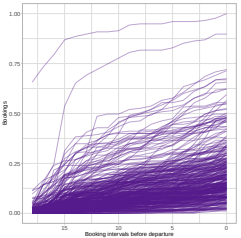



Figure: Leg CD

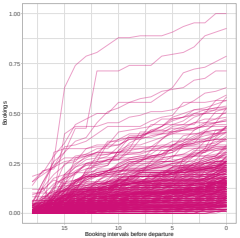
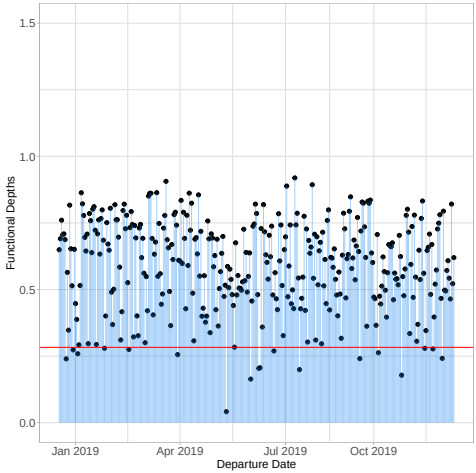



Figure: Leg DE

# Detecting outliers: Functional depth

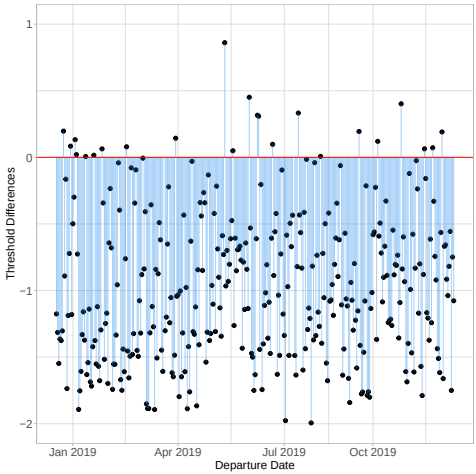
- Calculate the **functional depth** of the booking pattern for each train, separately for each leg.
- Depth measures provide an **ordering** of observations.
- Functional depth tells us how different the magnitude and shape is from the average booking pattern.



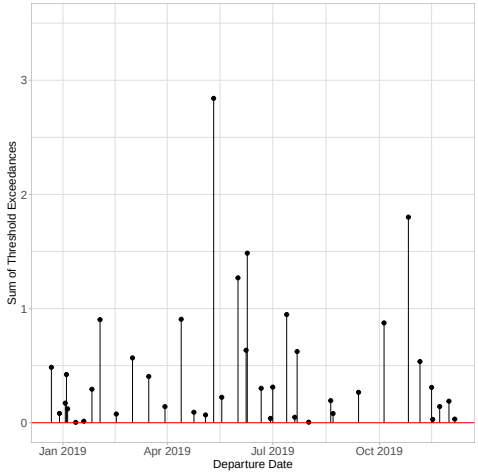
# Detecting outliers: Threshold exceedances

Define $z_{nl}$ to be the normalised difference between the functional depth and the threshold:

$$z_{nl} = \frac{C_l - d_{nl}}{C_l}.$$



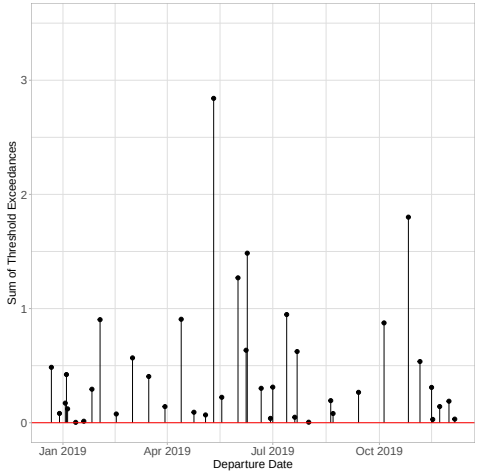
# Detecting outliers: Aggregation across legs




Next we define the sums of threshold exceedances across legs:

$$z_n = \sum_{l=1}^{L} z_{nl} \mathbb{1}_{\{z_{nl}>0\}}.$$

# Detecting outliers: Aggregation across legs



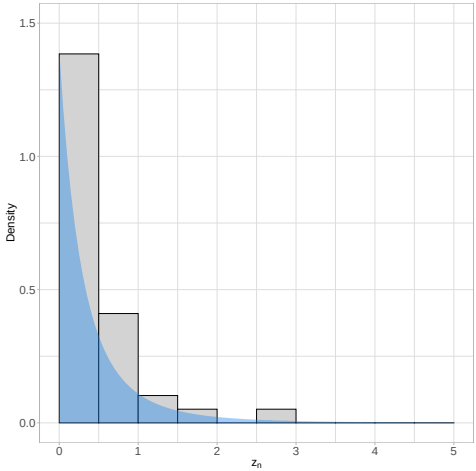

Next we define the sums of threshold exceedances across legs:

$$z_n = \sum_{l=1}^{L} z_{nl} \mathbb{1}_{\{z_{nl}>0\}}.$$

- Outlier in more legs $\Rightarrow$ Larger $z_n$
- **Larger** outlier $\Rightarrow$ Larger $z_n$

# Detecting outliers: Fitting a distribution

- Fit a generalised Pareto distribution (GPD) to the aggregated threshold exceedances.



# Detecting outliers: Outlier probabilities

Define $\theta_n$ to be the non-exceedance probability from the GPD. The non-exceedance probability is given by the CDF:

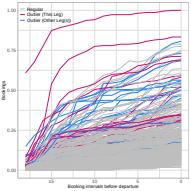
$$\theta_n = F_{(\mu,\sigma,\xi)}(z_n) = \begin{cases} 1 - \left(1 + \frac{\xi(z_n - \mu)}{\sigma}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - exp\left(-\frac{(z_n - \mu)}{\sigma}\right) & \xi = 0 \end{cases}$$

# Detecting outliers: Constructing a ranked alert list

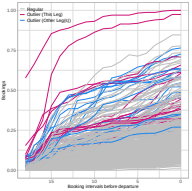
| Ranking | Departure | Probability | Legs Detected In |
|---|---|---|---|
| 1 | 11/05/2019 | 0.985 | AB, BC, CD, DE |
| 2 | 26/10/2019 | 0.960 | AB, BC, CD, DE |
| 3 | 09/06/2019 | 0.942 | AB, BC, CD, DE |
| 4 | 01/06/2019 | 0.874 | AB, BC, CD, DE |
| 5 | 13/07/2019 | 0.865 | AB, BC, CD, DE |
| ⋮ | ⋮ | ⋮ | ⋮ |

Table: Ranked alert list for cluster $= \{AB, BC, CD, DE\}$

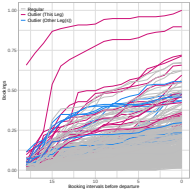
# Results: Deutsche Bahn



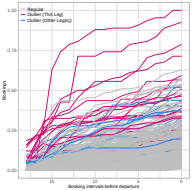

(a) Leg AB




(b) Leg BC




(c) Leg CD




(d) Leg DE

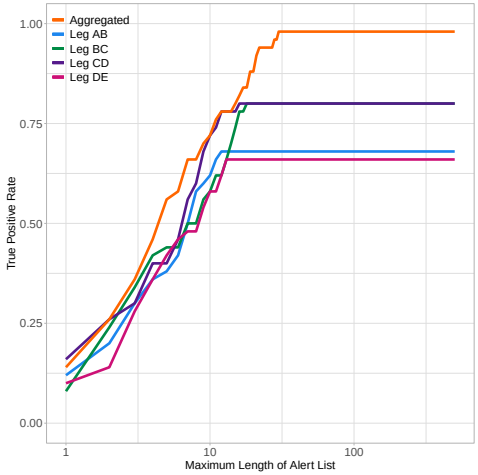
- Of the 40 outliers detected, 23 could be attributed to known events or holidays.
- When considering only the top 10 outliers, this rose to 70%.
- One of the detected outliers had been previously flagged.

# Results: Simulation

- **True positive rate**: fraction of genuine outliers which have been detected,

# Results: Simulation

- **True positive rate**: fraction of genuine outliers which have been detected,

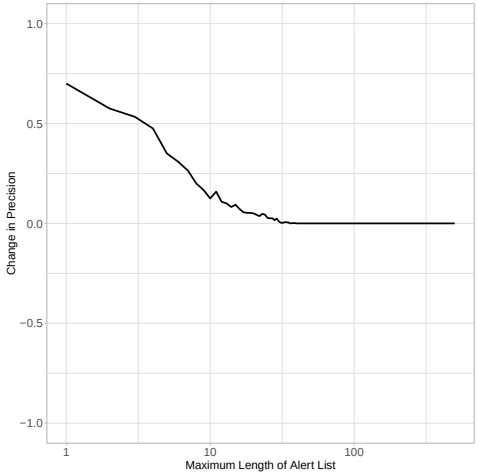


# Results: Simulation

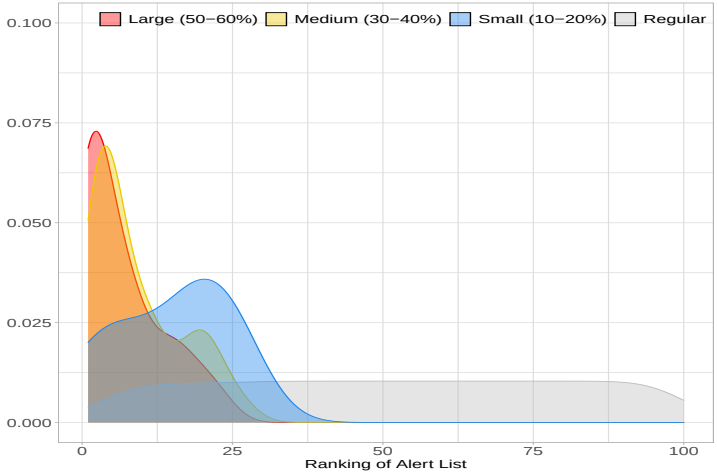
- **Precision**: fraction of trains classified as outliers that are genuine outliers.

# Results: Simulation

- **Precision**: fraction of trains classified as outliers that are genuine outliers.



# Results: Simulation



# References


- N. Rennie, C. Cleophas, A.M. Sykulski et al., Identifying and responding to outlier demand in revenue management. European Journal of Operational Research. (2021)
- M. Hubert, G. Claeskens, B. De Ketelaere, et al. A new depth-based approach for detecting outlying curves. Proceedings of CompStat 2012 (2012), pp. 329-340
- C. Cleophas, D. Kadatz, S. Vock. A literature survey of recent theoretical advances. Journal of Revenue and Pricing Management, 16 (5) (2017), pp. 483-49

# References


- N. Rennie, C. Cleophas, A.M. Sykulski et al., Identifying and responding to outlier demand in revenue management. European Journal of Operational Research. (2021)
- M. Hubert, G. Claeskens, B. De Ketelaere, et al. A new depth-based approach for detecting outlying curves. Proceedings of CompStat 2012 (2012), pp. 329-340
- C. Cleophas, D. Kadatz, S. Vock. A literature survey of recent theoretical advances. Journal of Revenue and Pricing Management, 16 (5) (2017), pp. 483-49



Email: n.rennie@lancaster.ac.uk