Detecting outlying demand in multi-leg bookings for transportation networks

Lancaster University

N. Rennie, C. Cleophas, A. M. Sykulski, F. Dost

EURO2021

Supporting decision making in RM

- RM systems often allow analysts to make adjustments to forecasts.
- However, judgemental forecasts can be biased and even superfluous (De Baets, S. and Harvey, N., 2020).
- Decision support for analysts is needed to reduce complexity.

Railway network RM

- Bookings are reported on the leg-level.
- Outliers don't affect entire network, nor single leg.
- Partition network using clustering.

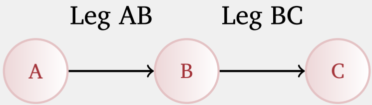


Figure: Railway network with two legs

Clustering railway legs

- Nodes represent stations.
- Edges represent legs connecting stations.

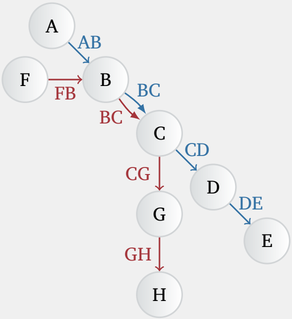


Figure: Railway network graph

Clustering railway legs

- Nodes represent legs.
- Edges define which legs can be in same cluster.

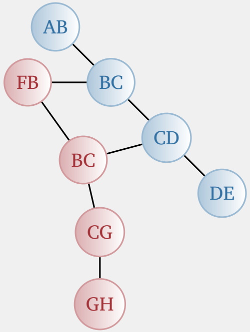


Figure: Inverted graph

Clustering railway legs

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

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

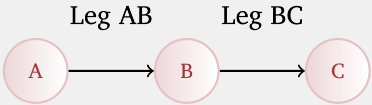


Figure: Railway network with two legs

Clustering railway legs

- Edge weights are $1 - r()$
- Obtain minimum spanning tree (Prim's algorithm)

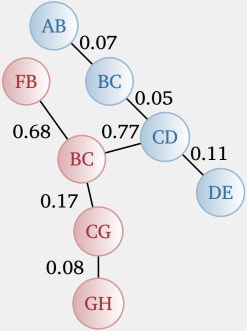


Figure: Minimum spanning tree with edge weights

Clustering railway legs

- Remove edges with weight above some threshold.

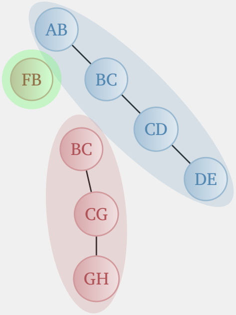
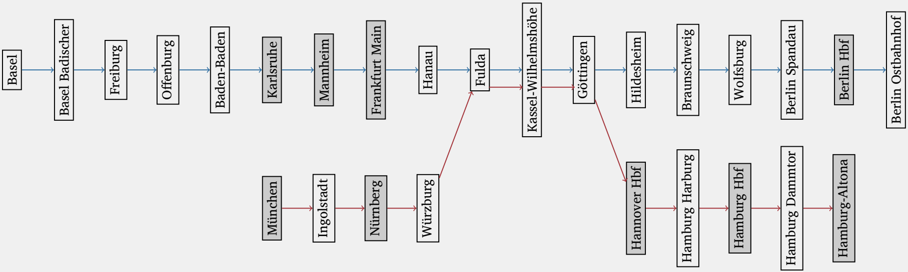
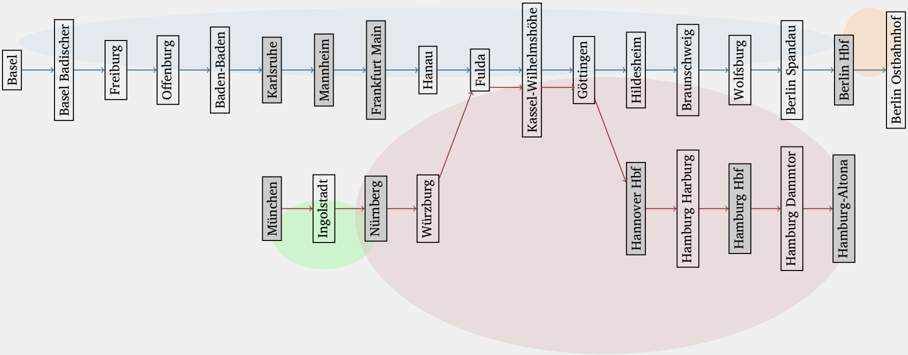


Figure: Clusters obtained in inverted graph

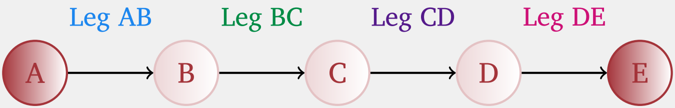
Clustering Deutsche Bahn network



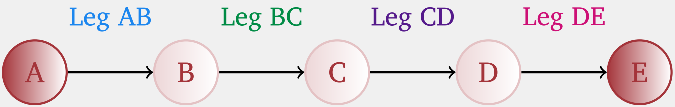
Clustering Deutsche Bahn network



Detecting outliers within clusters



Detecting outliers within clusters



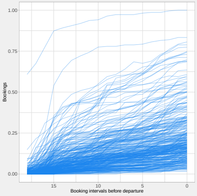


Figure: Leg AB

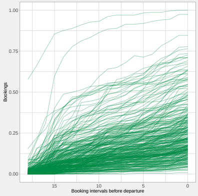


Figure: Leg BC

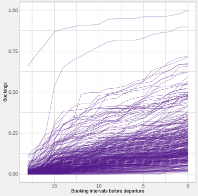


Figure: Leg CD

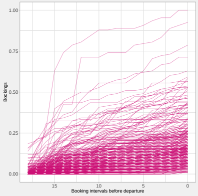
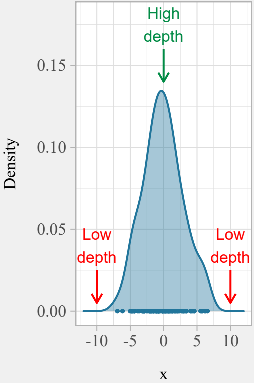


Figure: Leg DE

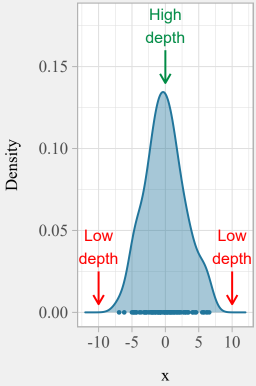
Detecting outliers within clusters

- Univariate depth: provides an ordering of the data

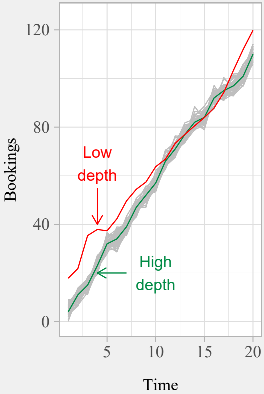


Detecting outliers within clusters

- Univariate depth: provides an ordering of the data

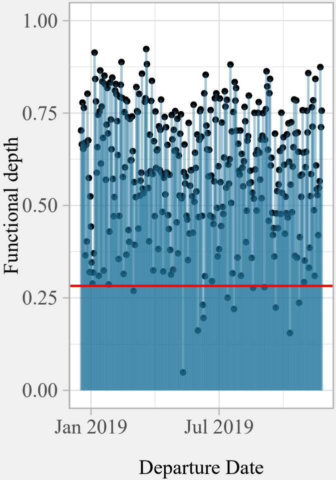


- Functional depth: measure of how central a trajectory is.



Detecting outliers within clusters

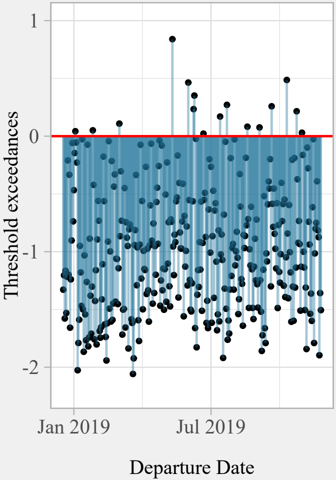
- Define a threshold for the functional depth on each leg.
- Departures with depth below threshold are outliers.



Detecting outliers within clusters

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

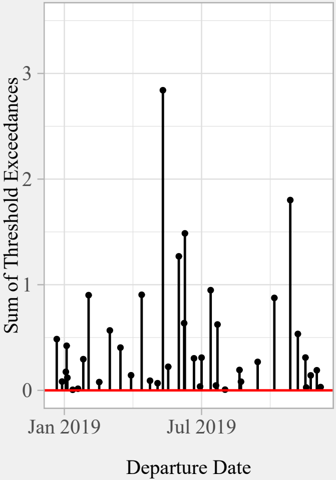
$$z_{nl} = \frac{C_l - d_{nl}}{C_l}.$$



Detecting outliers within clusters

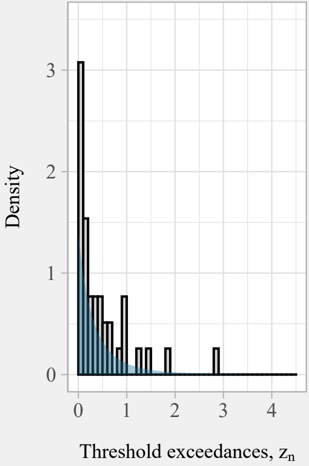
- Next we define the sums of threshold exceedances across legs:

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



Detecting outliers within clusters

- We want to measure outlier severity.
- Fit a generalised Pareto distribution (GPD) to the threshold exceedances.



Detecting outliers within clusters

Define θ_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 within clusters

Construct an alert list to send to analysts:

Ranking	Departure	Severity	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 = $\{AB, BC, CD, DE\}$

Detecting outliers within clusters

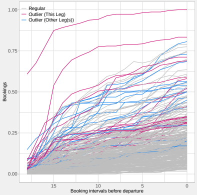


Figure: Leg AB

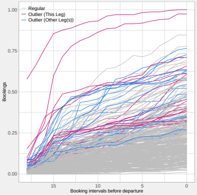


Figure: Leg BC

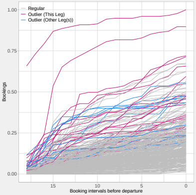


Figure: Leg CD

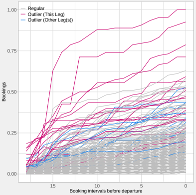


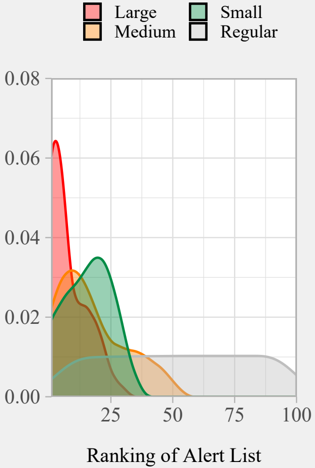
Figure: Leg DE

Detecting outliers within clusters

- Of the 40 outliers detected, 23 (58%) 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.

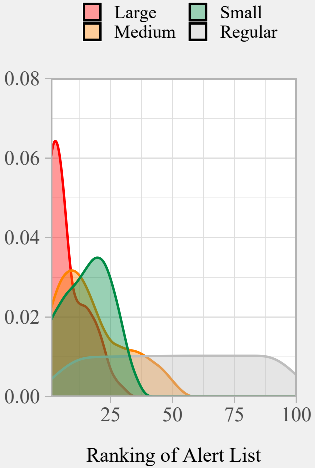
Simulation study results

- Use simulation to evaluate detection and ranking of outliers.



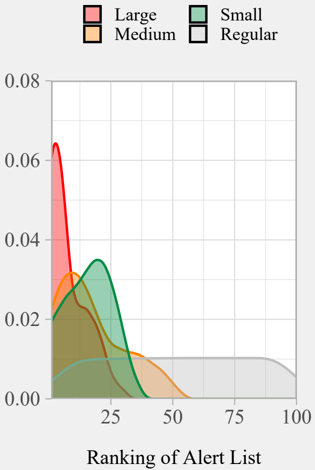
Simulation study results

- Use simulation to evaluate detection and ranking of outliers.
- Larger outliers are ranked higher.



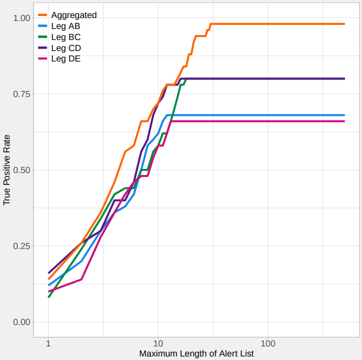
Simulation study results

- Use simulation to evaluate detection and ranking of outliers.
- Larger outliers are ranked higher.
- Ranking of medium outliers depends on sizes of other outliers.

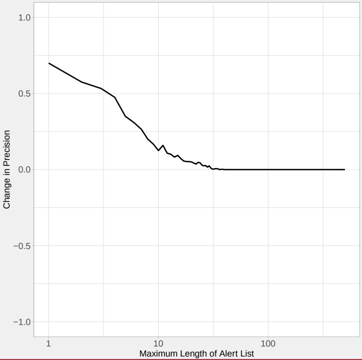


Simulation study results

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



- Δ **Precision:** fraction of detected outliers that are genuine outliers.



Conclusion

- Functional depth correctly identifies and ranks outliers for analysts.
- Aggregating information across similar legs improves performance.

References

- N. Rennie, C. Cleophas, A.M. Sykulski et al. *Identifying and responding to outlier demand in revenue management*. European Journal of Operational Research. Volume 293, Issue 3, 16 September 2021, 1015-1030.

- N. Rennie, C. Cleophas, A.M. Sykulski et al. *Detecting outlying demand in multi-leg bookings for transportation networks*. arXiv. 2021.

n.rennie@lancaster.ac.uk