



## Machine learning til realtidsforudsigelser af oprindelse-til-destination efterspørgsel for jernbaner med smart card og udbudsdata

Flensburg, Johan Victor; Lima Azevedo, Carlos M.; Pereira Duarte Rodrigues, Filipe M; Cerreto, Fabrizio

*Published in:*

Proceedings from the Annual Transport Conference at Aalborg University

*Publication date:*

2022

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Flensburg, J. V., Lima Azevedo, C. M., Pereira Duarte Rodrigues, F. M., & Cerreto, F. (2022). Machine learning til realtidsforudsigelser af oprindelse-til-destination efterspørgsel for jernbaner med smart card og udbudsdata. In *Proceedings from the Annual Transport Conference at Aalborg University* Danish Journal of Transportation Research - Dansk tidsskrift for transportforskning

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Denne artikel er udgivet i det elektroniske tidsskrift  
**Artikler fra Trafikdage på Aalborg Universitet**  
(Proceedings from the Annual Transport Conference  
at Aalborg University)  
ISSN 1603-9696  
<https://journals.aau.dk/index.php/td>

# Machine learning til realtidsforudsigelser af oprindelse-til-destination efterspørgsel for jernbaner med smart card og udbudsdata

Victor Flensburg, [jvfl@bane.dk](mailto:jvfl@bane.dk) / [jvfl@dtu.dk](mailto:jvfl@dtu.dk), Banedanmark / DTU Management  
Carlos M. Lima Azevedo, [climaz@dtu.dk](mailto:climaz@dtu.dk), DTU Management  
Filipe Rodrigues, [rodr@dtu.dk](mailto:rodr@dtu.dk), DTU Management  
Fabrizio Cerreto, [fceo@bane.dk](mailto:fceo@bane.dk), Banedanmark

---

## Abstrakt

Realtidsforudsigelser af passagerefterspørgsel på jernbanen kan bidrage til smartere trafikstyring og på sigt til at udvikle et offentligt transportsystem som på forskellig vis imødekommer ekstraordinær efterspørgsel. Dette kræver adgang til detaljeret information om efterspørgselsmønstre i form af løbende indsamling af passagertal for hvert par af oprindelses- og destinationsstationer i korte tidsintervaller. I dette studie udvikles en machine learning model til forudsigelser af afvigelser fra det periodiske efterspørgselsmønster på Københavns S-bane i 15 minutters intervaller ved hjælp af realtidsdata fra Rejsekortet på efterspørgselssiden og Banedanmarks driftsstatistikker på udbudssiden. Studiet belyser dels betydningen af udbud for forudsigelse af efterspørgsel og dels udforskes måden hvorpå spatiotemporal data indlejres i modeller fra dyb læring for at opnå nøjagtige forudsigelser for mange-dimensionale og sparsomme data som disse.

---

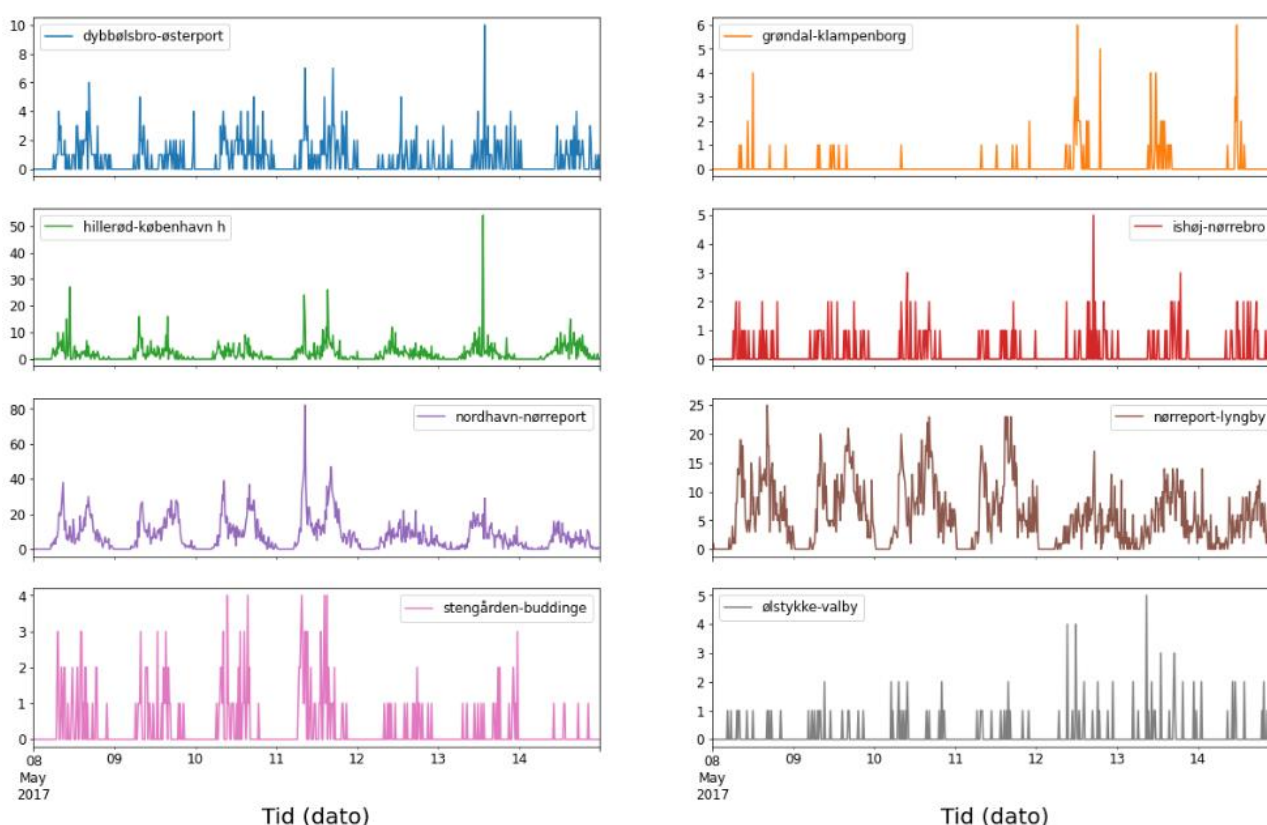
## Baggrund og formål

Forudsigelse af passagerefterspørgsel er essentielt i offentlig transport, idet det faciliterer de socio-økonomiske modeller, der benyttes til udvælgelse af nye infrastrukturprojekter, men også muliggør optimeret design og styring af trafikplaner. Udbuddet af transport styres således sådan at det imødekommer transportefterspørgslen. Selvom transportefterspørgsel generelt følger et periodisk mønster, kan det være ekstremt svært at forudsige afvigelser fra dette mønster i løbet af kortere tidsintervaller, da afvigelser dels kan være forholdsvis store og dels kan skyldes en lang række eksternaliteter, der er svære at monitorere uden løbende adgang til information, såsom forsinkelser, nedbrud, vejr, særlige begivenheder som fodboldkampe, osv. Generelt udgør vekselvirkning mellem udbud og efterspørgsel en betydelig faktor i at fastlægge graden af sårbarhed i et transportsystem (Cats & Jenelius, 2014). I jernbaner kan manglende transportudbud i forhold til variationerne i efterspørgsel resultere i gener for passagerer i form af forsinkelser og ubehag på grund af længere passagerudvekslingstider på stationer som følger af overfyldte perroner og tog, og i værste tilfælde kan nogle passagerer nægtes indstigning. Herudover kan store forsinkelser spredes til større dele af netværket, hvorved det kan være særligt svært at genoprette normal drift på grund af afhængigheder i signalsystem og infrastrukturens indskrænkede natur. Adgang til nøjagtige forudsigelser af

efterspørgsel for den nære fremtid kan muliggøre foregribende trafikstyring og kan medvidere indgå i et design af et højdynamisk intelligent offentligt transportsystem, der kan reagere på variationer i efterspørgsel.

I takt med, at "big data" og automatiske billetteringssystemer, også kendt som "smart cards", har vundet indpas, har forskning i storskalaanalyser af forskellige problemer i offentlig transport i stigende grad fundet sted (Faroqi, Mesbah, & Kim, 2018), såsom estimering af oprindelse-til-destination (OD) efterspørgsel (Hussain, Bhaskar, & Chung, 2021), heriblandt for jernbanen. Den særlige fordel ved smart card data er at de indeholder detaljeret information om hver enkelt rejse, som kan bruges til at lave realtidsforudsigelser ved hjælp af løbende indsamlet efterspørgselsdata, både med hensyn til hvor og hvornår, dvs. præcis angivelse af oprindelses-, destinations- og skiftestop såvel som præcise tidsangivelser. Følgelig har megen forskning de seneste år fokuseret på at forudsige OD efterspørgsel for korte intervaller, eksempelvis 15 minutter, ved hjælp af datasultne metoder som dyb læring (Zhang, Chen, Zijia, & Liu, 2019) (Liu, et al., 2021) (Liu, et al., 2020) (Noursalehi, Koutsopoulos, & Zhao, 2021) (Jiang, Ma, & Koutsopoulos, 2022), selvom andre typer af modeller ligeledes har været udforsket (Cheng, Trepanier, & Sun, 2021) (Gong, Li, Zhang, Liu, & Zheng, 2020) (Yao, Zhao, & Yu, 2015).

### En uges OD efterspørgsel (antal passagerer) i 15 minutters intervaller for otte udvalgte OD par

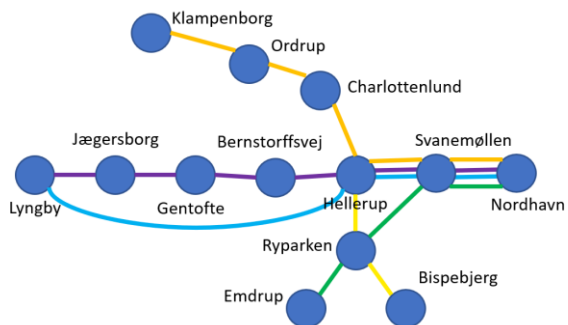


Figur 1: Udviklingen af efterspørgsel over en bestemt uge i maj 2017 for otte udvalgte OD stationspar på Københavns S-bane. Her ses nogle af de pludselige variationer som ikke gentages periodisk samt efterspørgsel i forskellige størrelsesordener. Bemærk desuden, at fredag den 12. maj 2017 var Store Bededag i Danmark, en helligdag. Graferne er baseret på data fra Rejsekort A/S.

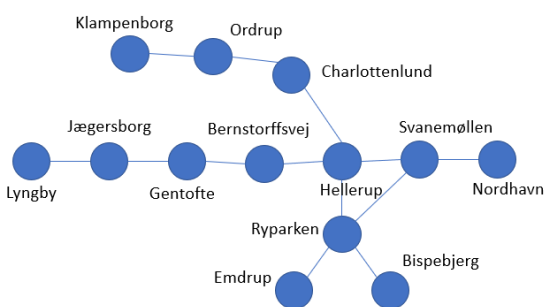
### Problemformulering og metode

Dette arbejde har som formål at opnå nøjagtige forudsigelser af OD passagerefterspørgsel i et urbant jernbanenetværk i løbet af intervaller af 15 minutter ved hjælp af efterspørgselsdata fra en smart card udbyder. I studiet foreslås en model til at nå dette mål. Udover efterspørgselsdata benyttes også parametre for variation og (u)pålidelighed af transportudbud for at tage højde for sammenhængen mellem udbud og efterspørgsel, hvilket har været overset i litteraturen for efterspørgselsforudsigelser. Udbudsparametrene i

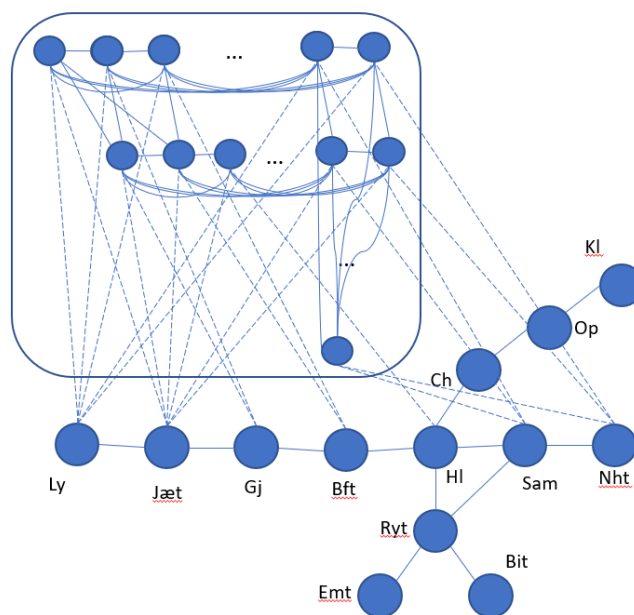
modellen inkluderer metrikker for frekvens, forsinkelser ved togankomster og togaflysninger. Modellens forudsigelser er baseret på netværkets nylige og nuværende tilstande med hensyn til efterspørgsel og driftsniveau. Herudover bruges forskellige kontekstuelle parametre såsom ugedag, klokkeslæt og ferieinformation til at identificere periodiske mønstre. For at lade modellen fokusere på mønstrene i variationerne af efterspørgsel, trækkes det historiske periodiske mønster fra i inputtet for efterspørgsel således at modellen forudsiger afvigelser fra den historiske gennemsnitlige efterspørgsel. Metoder fra dyb læring er udvalgt til at løse dette problem på grund af de komplekse spatiotemporale sammenhænge i data, for hvilke dyb læring generelt opnår bedre resultater end traditionelle machine learning metoder (Wang, Cao, & Yu, 2020).



Figur 2: Et udsnit af Københavns S-bane omkring Hellerup med linjeforbindelser i farver.



Figur 3: Simple grafrepræsentation i henhold til direkte sporforbindelser.



Figur 4: Udvidet grafrepræsentation på OD niveau med knuder, der repræsenterer OD par med indbyrdes forbindelser. Her fremgår både den simple repræsentation på stationsniveau nederst og OD grafrepræsentationen i den markerede kasse. Forbindelsen mellem OD grafen og den simple netværksrepræsentation er givet ved de stiplede forbindelser, hvor hver OD knude er forbundet til to stationsknuder.

Nogle af de mest fremtrædende udfordringer i forbindelse med realtidsforudsigelser af OD efterspørgsel er problemerne med mange-dimensioneret og sparsomt input- og outputdata, som betyder at forudsigelsen består af mange punktestimater med høj varians på tværs af både rum og tid i domænet, se eksempelvis variationerne i **Figur 1**. Dette studie betragter forudsigelse af hele OD matricen som sandsynligvis omfatter et stort antal OD par med lav efterspørgsel. På grund af dette baseres den foreslåede model på metoder fra dyb læring som specifikt er designet til spatiotemporale forudsigelser, navnlig det såkaldte "Graph Convolution Network" (GCN), som hører til familien af "Graph Neural Networks" (GNN) og udfører såkaldte "convolutions" eller "foldninger" på grafindlejret data (Kipf & Welling, 2017). Dette bruges til at samle og videregående information i det neurale netværk for at fange eksisterende sammenhænge mellem forskellige

områder i netværket over tid, i dette tilfælde forskellige stationer eller OD par i jernbanenetværket. Vi udfører undersøgelser af vigtigheden af grafstrukturen for at identificere hvordan den a priori viden om sammenhænge i tid og rum, som udgøres af forbindelserne i grafen, bør fastlægges. I nærværende problem kan en graf, der repræsenterer et givet jernbanenetværk (se **Figur 2**) fx bestå af knuder og kanter på stationsniveau (se **Figur 3**) eller på OD niveau (se **Figur 4**), hvilket ændrer indlejringen af data samt måden hvorpå information spredes.

Vi sammenligner resultater mellem forskellige grafrepræsentationer samt med og uden variable for udbuddet af transport. Studiets vigtigste videnskabelige bidrag er at indikere vigtigheden af at tage højde for transportudbudsvariable i realtidsforudsigelser af passagerefterspørgsel på jernbanen samt at udforske grafdatastrukturer for GNN-modeller hvori variable indlejres med henblik på at opnå større nøjagtighed i mange-dimensionale og sparsomme forudsigelser.

## Case study og data

Vi bruger Københavns S-banenetværk som case på grund af den famøse fingerstruktur med intuitive efterspørgselsmønstre, størrelse og tilgængelighed af data i høj kvalitet. Efterspørgselsdata er stillet til rådighed gennem Rejsekortets automatiske billetteringssystem, som bruges i stort omfang i Hovedstadsområdet. Disse data er ekstremt værdifulde, da de giver indsigt i detaljeret information om de enkelte rejsers på- og afstigningslokationer, skift mellem transportformer og præcise tidsangivelser for alle transaktioner (såsom check-ind og check-ud). Efterspørgselsdata aggregeres for intervaller på 15 minutter for hvert par af oprindelses- og destinationsstationer i S-banenetværket, se **Figur 1**. Data for driftsplan og pålidelighed er trukket fra databasen for regularitets- og driftsstatistikker i Banedanmark, også kendt som RDS.

## Anerkendelser

Dette studie er tilknyttet det europæiske multidisciplinære projekt SORTEDMOBILITY<sup>1</sup>, understøttet af Europa-Kommissionen og finansieret under Horizon 2020 ERA-NET Cofund skemaet under bevillingsaftale N. 875022 med intern finansiering i Danmark i fællesskab mellem Danmarks Innovationsfond og Banedanmark. Særlig tak til Rejsekort A/S for at stille data til rådighed.

## Referencer

- Cats, O., & Jenelius, E. (2014). Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information. *Netw Spat Econ*, pp. 435–463. doi:<https://doi.org/10.1007/s11067-014-9237-7>
- Cheng, Z., Trepanier, M., & Sun, L. (2021). *Real-time forecasting of metro origin-destination matrices with high-order weighted dynamic mode decomposition*. Retrieved from arXiv 2101.00466: <https://arxiv.org/abs/2101.00466>
- Faroqi, H., Mesbah, M., & Kim, J. (2018). Applications of transit smart cards beyond a fare collection tool: A literature review. *Advances in Transportation Studies*, pp. 107-122.
- Gong, Y., Li, Z., Zhang, J., Liu, W., & Zheng, Y. (2020). Online Spatio-temporal Crowd Flow Distribution Prediction for Complex Metro System. *IEEE Transactions on Knowledge and Data Engineering*, p. 1. doi:10.1109/TKDE.2020.2985952
- Hussain, E., Bhaskar, A., & Chung, E. (2021). ransit od matrix estimation using smartcard data: Recent developments and future research challenges. *Transportation Research Part C: Emerging Technologies*, p. 103044.
- Jiang, W., Ma, Z., & Koutsopoulos, H. N. (2022). Deep learning for short-term origin–destination passenger flow prediction under partial observability in urban railway systems. *Neural Comput & Applic*, pp. 4813-4830. doi:10.1007/s00521-021-06669-1

---

<sup>1</sup> For mere information om dette projekt, se <https://www.sortedmobility.eu/about/>

- Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. *5th International Conference on Learning Representations, {ICLR} 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. Retrieved from <https://openreview.net/forum?id=SJU4ayYgl>
- Liu, L., Chen, J., Wu, H., Zhen, J., Li, G., & Lin, L. (2020). Physical-Virtual Collaboration Modeling for Intra- and Inter-Station Metro Ridership Prediction. *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-15. doi:10.1109/TITS.2020.3036057
- Liu, L., Zhu, Y., Li, G., Wu, Z., Bai, L., & Lin, L. (2021). *Online Metro Origin-Destination Prediction via Heterogeneous Information Aggregation*. Retrieved from arXiv: 2107.00946: <https://arxiv.org/abs/2107.00946>
- Noursalehi, P., Koutsopoulos, H. N., & Zhao, J. (2021). Dynamic Origin-Destination Prediction in Urban Rail Systems: A Multi-Resolution Spatio-Temporal Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-10. doi:10.1109/TITS.2020.3047047
- Wang, S., Cao, J., & Yu, P. (2020). Deep Learning for Spatio-Temporal Data Mining: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, p. 1. doi:10.1109/TKDE.2020.3025580
- Yao, X., Zhao, P., & Yu, D.-d. (2015). Real-time origin-destination matrices estimation for urban rail transit network based on structural state-space model. *Journal of Central South University*, pp. 4498-4506.
- Zhang, J., Chen, F., Zijia, W., & Liu, H. (2019). Short-Term Origin-Destination Forecasting in Urban Rail Transit Based on Attraction Degree. *IEEE Access*, pp. 133452-133462. doi:10.1109/ACCESS.2019.2941177