

Automatisk Passageropregning hos DSB S-Tog

Rasmus Dyhr Frederiksen, Rapidis ApS

DSB S-Tog har brug for detaljerede informationer om deres passagerers rejsevaner (hvor mange der rejser, hvorfra og hvortil), både i forbindelse med fordeling af takstpenge men også i forbindelse med både kort- og langsigtet planlægning. Disse oplysninger er hidtil blev skaffet til veje vha. den årlige Østtælling, suppleret med månedlige tællinger på udvalgte banesnit.

Nye datakilder sammen med udviklingen af en række nye metoder på trafikmodel-området har gjort det muligt at afløse det hidtidige system med et nyt som muliggør et kæmpe spring i detaljeringsgrad, nøjagtighed og opdateringshastighed. Resultatet er meget detaljerede daglige beskrivelser af S-Tog passagerernes rejse-aktiviteter.

Dette nye system, S-Tog Automatisk Passager Opregning (APO) er blevet udviklet i et samarbejde mellem DTU Transport, DTU Informatik, Rapidis ApS samt DSB S-Tog. GIS spiller en central rolle i forbindelse med præsentation og analyse af resultater.

Indledning

DSB S-Tog har brug for detaljerede informationer om deres passagerers rejsevaner, ikke blot hvor mange passagerer der rejser i alt, men også på hvilke tidspunkter de rejser, hvor de rejser fra og til, og hvor de evt. skifter undervejs.

Disse oplysninger bruges dels på det overordnede plan i forbindelse med fordeling af takstpenge i hovedstadsområdet, og dels i forbindelse med driftsplanlægning – hvornår er det bedst at lukke et spor for vedligeholdelse, hvor store togstammer skal der køres med, hvem vil blive påvirket af evt. køreplansændringer, osv.

Hidtil er disse oplysninger blevet skaffet til veje med en kombination af data fra en stor årlig stations-tælling (Østtællingen, hvor alle rejsende får udleveret papirlapper der skal afleveres ved rejsens afslutning), suppleret med jævnlige tællinger på udvalgt strækninger. Disse tællinger anvendes derefter til at foretage overordnede justeringer af stations-tællingen. Denne tilgang giver et rimeligt godt overordnet billede af de rejsendes antal og valg af ruter på et overordnet niveau, men mangler detaljer, samt er upræcis med hensyn til uge- og sæsonvariationer. Endelig opdateres data ikke dagligt.

I takt med at DSB S-Togs nye tog-materiel er blevet indført, har der åbnet sig nye mulig-

heder for at indsamle data. Alle de nye S-Tog foretager løbende og med god præcision vejninger af passagerne i hvert togsæt. Disse data anvendes umiddelbart til løbende at kalibrere bremsesystemerne i togsættene. Men ved at indsamle disse vejedata, samt kombinere dem med tælledata fra en mindre del af togsættene, åbner der sig en unik mulighed for at skaffe sig et detaljeret og løbende opdateret datasæt med et stort potentiale.

Med udgangspunkt i DSB S-Togs ønske om et nyt og bedre system til produktion af passager-tal og detaljerede rejse-opgørelser, er der blevet gennemført et projekt i to trin. Første trin har været at undersøge kvaliteten af de tilgængelige data, samt mulighederne for at udnytte dem.

Dernæst er erfaringerne fra denne fase blevet udnyttet til at designe og udvikle et driftssystem der dels kan levere daglige detaljerede rejseopgørelse og dels kan bidrage til den overordnede statistik med hensyn til det samlede antal rejsende.

De gennemgående personer i projektet med hensyn til det faglige arbejde kan ses i listen efter artiklen.

Resten af denne artikel vil gå lidt mere i dybden med de forskellige aspekter af projektet:

- Datagrundlag – datakilder og den statistiske undersøgelse af disse.
- De anvendte metoder og test af disse
- Opbygningen af driftssystemet og hvordan det teknisk hænger sammen
- Hvordan GIS er blevet anvendt i projektet, samt hvordan det indgår i driftssystemet
- Endelig samles der op i en konklusion.

Data-grundlag

Med udskiftningen af DSB S-Togs rullende materiel er nye datakilder blevet tilgængelige. Ved hjælp af det indbyggede ToDat-system er det muligt løbende at indsamle veje- og tælle-data. Der indsamles veje-data fra alle togsæt, det vil sige en daglig stikprøve på 100%.

Tælle-data indsamles fra infra-røde tællebjælker, som er blevet monteret i 18% af togene. Da der altid vil være nogen togsæt der er ude af drift pga. almindelige vedligeholdelse, vil der nok typisk kunne indsamles tælle data fra ca. 15% af togsættene.

Alle ToDat-data logges med tognr, tidspunkt og lokation hvilket gør det muligt at knytte de indsamlede data til den afviklede køreplan for dagen.

A. Egenskaber ved datakilder

De indsamlede data byder på meget forskellige egenskaber. Tælle-dataene er umiddelbart direkte anvendelige, da der er tale om direkte observationer af passagerer i form af antal af- og påstigere. I modsætning hertil er veje-dataene kun anvendelige hvis der kan foretages en omregning fra vægt til antal passagerer. Endelig kan man ud fra veje-data kun fastslå differensen mellem hvor mange der ankom til en station og hvor mange der kørte videre – det er ikke muligt direkte at registrere hvor mange der egentlig steg af og på.

Til gengæld har veje-dataene som beskrevet en meget stor dækningsgrad i forhold til tæl-

le-dataene. Endnu vigtigere, så er veje-data en direkte observation af hvor mange rejsende der sidder i et togsæt, når det kører fra en station til en anden.

Det er et problem ved tælle-data at idet der er tale om af- og på-stiger tællinger, har man kun et indirekte billede af hvor mange rejsende der er i toget mellem hver station (antal påstigere minus antal afstigere, akkumuleret langs et vognløb). Idet det ikke kan undgås at der er en vis usikkerhed forbundet mellem tællingerne, akkumuleres denne usikkerhed langs et vognløb. Derfor kan bud på antal rejsende mellem stationer være ret usikre i slutningen af et vognløb, hvis de baseret på tælle-data alene.

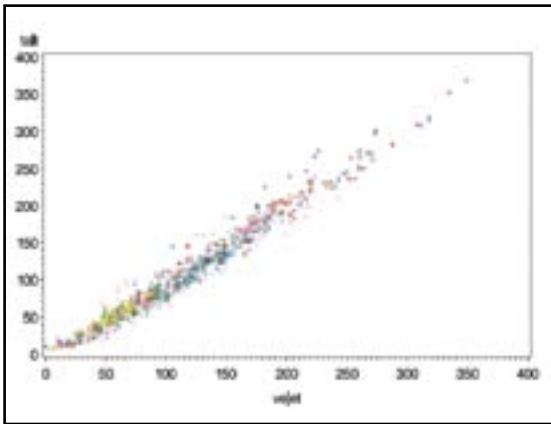
Det er således klart at veje- og tælle-data tilsammen danner et bedre analysegrundlag end hver datakilde for sig.

B. Indledende analyse af data-kvalitet

Der har været udført et stort arbejde for at teste kvaliteten af henholdsvis veje- og tælledata. Dette er sket ved at indsamle veje og tælle-data fra den daglige drift, og sammenholde disse med omfattende manuelle kontrol-tællinger. Der er efterfølgende blevet udført en omfattende statistisk analyse af disse data. De overordnede konklusioner kan her opsummeres kort:

- Veje-dataene har stort set samme usikkerhed som tælle-dataene
- Den gennemsnitlige passager-vægt er overraskende stabil henover døgnnet, samt henover ugen.
- Begge datakilder har en tilstrækkelig høj kvalitet til at de kan anvendes i praksis

Begge datakilder bød på visse systematiske fejl – vejedataenes usikkerhed steg ved meget få passagerer, tælle-dataene havde en tendens til at undertælle når mange rejsende steg af på samme tid. Men i begge tilfælde viste det sig at være muligt systematisk at kompensere for disse.



Figur 1. Eksempel data, omregnet vægt sammenholdt med tællinger for forskellige vognløb med samme tog-sæt

Metoder

Når der skal beregnes passager-tal for kollektive transport-systemer anvendes typisk en af to tilgange:

Justering af fast turmønster på basis af snit-tællinger

Med udgangspunkt i snit-tællinger justeres forskellige dele af et fast tur-mønster (Øst-tællingen) op og ned, således at niveauet passer. Metoden giver et vist indblik i rejsendes tur-mønster og i det totale antal rejsende. Men kun indirekte indblik i kapacitetsudnyttelsen af de enkelte tog.

Opregning af antal ture ud fra antal påstigere samt omstigningsfaktor

Antal påstigere tælles. Der justeres for den forventede andel af påstigere som foretager skift. Metoden giver et indblik i kapacitetsudnyttelsen af det rullende materiel, og det totale antal ture, men kun indirekte indblik i de rejsendes turmønster. Typisk indsamles data for en vis daglig stikprøve, hvorefter der laves en samlet opregning, når der er indsamlet data fra alle linier.

Begge metoder opregner det samlede antal ture direkte ud fra tællinger, og evt. manglende tællinger skal derfor kunne håndteres på en rimelig måde.

På grund af det gode datagrundlag samt ønsket om mere detaljeret daglige opgørelser blev det i forbindelse med APO-projektet besluttet, at anvende en anden tilgang, baseret på matematiske værktøjer fra trafikmodel og simuleringsområdet.

I stedet for at opregne passager-tal mere eller mindre direkte fra tællinger, så foretages en justering af tur-mønster-data således at de simulerede passager-mængder passer med de observerede passager-mængder. Dette gøres ved hjælp af en *rutevalgsmodel* og en *tur-mønster justering*, som beskrives nedenfor.

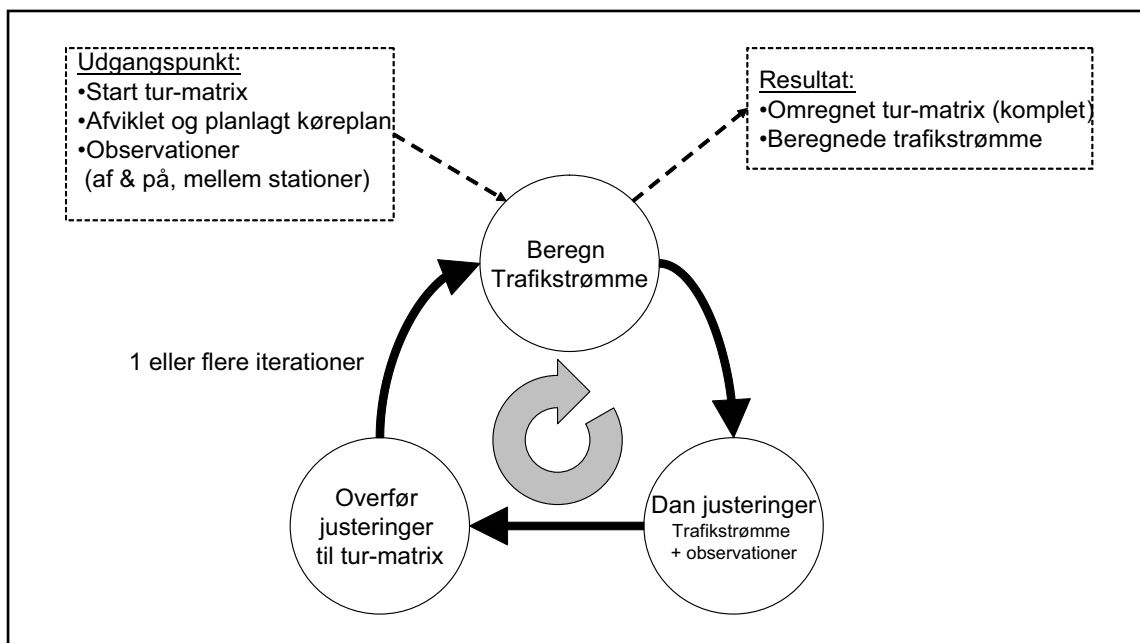
A. Rutevalgsmodel

I forbindelse med trafikmodeller anvendes typisk *rutevalgsmodeller*, som er matematiske modeller for rejsendes valg af rute igennem et transport-system. Input til en rutevalgsmodel er overordnet set data der beskriver:

- Et tur-mønster, en beskrivelse af hvor mange der rejser, hvornår, og hvor fra og hvor til.
- Et transport-netværk, i denne forbindelse S-Tog systemets køreplan
- En model beskrivelse af de rejsendes præferencer – hvordan vægter de rejsende de forskellige oplevede omkostninger (ventetid, rejsetid, skift, penge-omkostninger).

Resultatet af en rutevalgsberegning er en simulering af hvilke ruter de rejsende (beskrevet i tur-mønsteret) vælger i transport-netværket, samt detaljerede oplysninger om de forventede rejsetider. Der kan læses mere om rutevalgsmodeller for kollektiv trafik i Nielsen & Frederiksen (2003).

I forbindelse med S-Tog APO projektet har det været en stor fordel at der tidligere er blevet opstillet en rutevalgsmodel baseret på et matematisk netværk (en graf) specielt for S-Togs nettet. En nærmere beskrivelse kan læses i Seest, Nielsen & Frederiksen (2005).



Figur 2. Tur-mønstret justeres iterativt indtil simulerede passager-strømme matcher observationer

B. Tur-mønster justering

Med udgangspunkt i rutevalgsmodellen beskrevet ovenfor er det mulighed at opbygge et værktøj til Tur-mønster justering. Fremgangsmåden kan kort beskrives således:

Med udgangspunkt i et turmønster og en køreplan foretages justeringer i tur-mønstret, således at trafikstrømmene der simuleres ved hjælp af rutevalgsmodellen afviger så lidt som muligt fra de observerede passagertal.

Den anvendte metode kaldes MPME-metoden (Multiple Path Matrix Estimation). En mere detaljeret beskrivelse kan findes i Nielsen (1998). Metoden vil ikke blive beskrevet nærmere her, da den er ret omfattende.

Det aspekt ved metoden der er relevant i denne forbindelse er, at hver enkelt turmønster-element (fra-station, til-station, antal rejsende, tidsinterval) justeres på basis af et stort antal observationer (tællinger og vejninger), og at denne justering ikke er følsom

O	D	Tidsbaandstart	Passagerer
624	634	16	26.40863
624	634	16	29.5277
624	634	16	19.345488
624	634	16	24.53867
624	634	17	6.318844
624	634	17	7.0488
624	634	17	3.901896
624	634	17	5.651135
624	634	18	15.008832
624	634	18	16.741
624	634	18	10.677285
624	634	18	17.728343
624	634	19	10.378368
624	634	19	11.5758
624	634	19	5.244053
624	634	19	8.377077
624	634	20	2.150822
624	634	20	2.3994
624	634	20	1.080992
624	634	20	1.738407
624	634	21	6.290577
624	634	21	7.0178

Figur 3. Eksempel på beregningsoutput fra MPME-værktøj – antal rejsende fra station til station i et bestemt tidsinterval

overfor evt. manglende enkelt-observationer. I stedet gælder det overordnet at jo flere observationer justeringen baseres på, jo mere pålidelig er justeringen.

Umiddelbart er resultatet et tur-mønster, hvor rejsende mellem alle stations-par er opgjort på 42 tids-intervaller.

Ved hjælp af den ovenfor beskrevne rutevalgsmodel kan dette tur-mønster omdannes til en komplet beskrivelse af S-Togs systemet – hvor mange rejsende der er i hvert togsæt, hvor mange der stiger af og på ved hver eneste afgang.

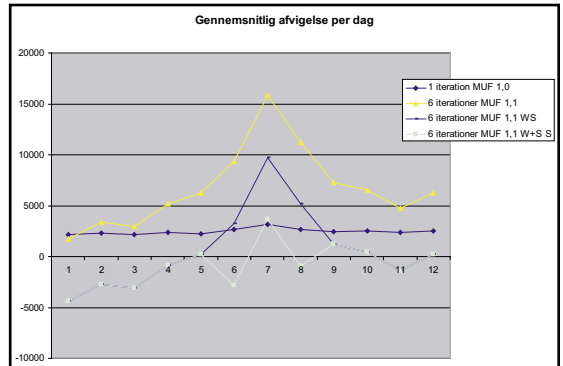
C. Test af metoder og data

Der har været udført omfattende test af den foreslåede fremgangsmåde. Fremgangsmåden har helt overordnet været at opstille et test-eksempel: et eksisterende tur-mønster data-sæt (Øst-tællingen, opstillet ved hjælp af en stations-tælling) og en køreplan indsamlet for en typisk driftsdag blev ved hjælp af Rutevalgsmodellen for S-Togs rejsende omdannet til et sæt data for af- og påstigere, samt passagertal for togsættene.

Dette sæt data er herefter blevet påtrykt "støj", dvs. simuleret variation svarende til de usikkerheder som det indledende statistiske data-arbejde viste, kunne forventes på indsamlede observationer. Dette simulerede sæt data svarer til et fuldt sæt observationer (tællinger og vejninger). Ved at anvende dette sæt data som observationer, svarende til de indsamlede veje- og tælle-data, var det muligt at afvikle beregninger med MPME-metoden, og sammenholde resultatet af disse med det "rigtige", kendte resultater.

Ved hjælp af denne fremgangsmåde kunne forskellige test-scenarier afprøves, eksempelvis:

- Sammenligning af MPME-metoden i forhold til det eksisterende system?
- Hvad vil den forventede nøjagtighed være for en typisk driftsdag?



Figur 4. Afvigelse mellem estimerede og sande antal passagerer pr. dag, gennemsnit for hver måned

- Hvilken nøjagtighed giver forskellige stikprøve-størrelser for hhv. vejninger og tællinger?
- Hvor godt håndteres den forventede uge- og sæson-variation i passager-tal? Eller f.eks. begivenheder såsom koncerter der giver atypiske rejsemønstre?
- Hvor meget falder nøjagtigheden hvis dele af data-indsamlingen falder ud?
- Hvor godt håndteres metoden større usikkerheder på observationerne.

I forbindelse med hvert scenarie er et helt driftsår blevet simuleret, hvor der for hver driftsdag i simulationen er udført:

- opstilling af et "sandt" tur-mønster for hvert driftsdag
- dannelse af "sande" observationer (på basis af rutevalgsberegning)
- påtrykning af støj på observationer for at simulere usikkerheden på de rigtige data
- MPME-beregning og opregning af overordnede resultater.

Generelt har simulationsberegningerne været udført med meget konservative estimater for usikkerheden af vejninger og tællinger. Disse

usikkerheder har siden hen vist sig at være mere nøjagtige.

Nedenfor vises et eksempel på et test-resultat. Figuren viser forskellen mellem det "sande" antal passagerer pr. dag, og det antal passagerer som beregnes på basis af observationerne på trykt støj. Det viste tal er det månedlige gennemsnit for den daglige afvigelse (Figur 4):

På figuren vises oversigtsresultater for 4 forskellige simulationer, som har anvendt forskellige opsætninger af systemet, samt forskellige forudsætninger om data-kvalitet. For at sætte tallene i perspektiv kan det nævnes at der på en typisk hverdag rejser ca. 320.000 – 330.000 passagerer med S-Togene.

APO – System, design og implementering

Det samlede APO-system kan deles op i flere separate trin, skitseret nedenfor:

- De grundlæggende data hentes fra et S-Tog datawarehouse
- Der udføres en "datavask" som behandler de grundlæggende tælle- og veje-data (dette beskrives nedenfor). Det grundlæggende resultat er en omregning af vejninger til antal passagerer
- Endelig udføres en MPME-beregning, således at produceres et nyt tur-mønster, og dermed et bud på det samlede antal rejsende med S-Tog.

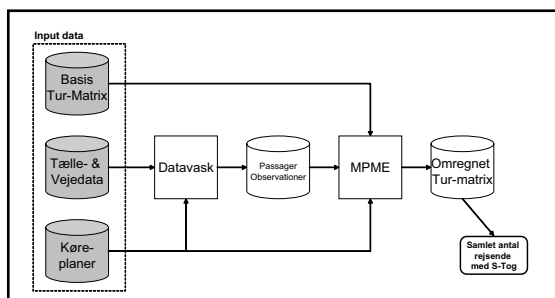
Denne samlede proces udføres hver dag, når dagens driftsdata er indsamlet.

Datavask trinnet (se Figur 6) består i virkeligheden af en række mindre trin. Input til datavasken er

- Køreplaner (planlagt og afviklet)
- Veje og tælle-data

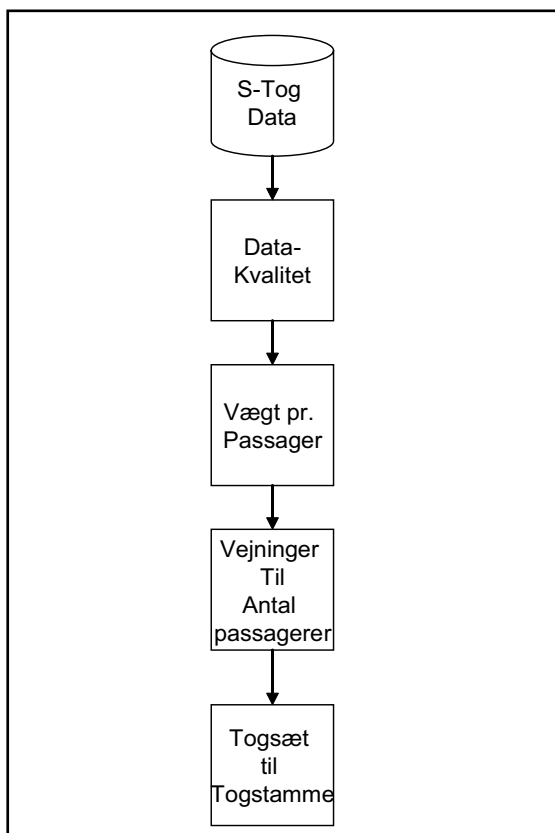
Formålet med datavasken er:

- Konvertering af vejedata til passagertal
- Konvertering af tælledata til konsistente tællinger



Figur 5. Overordnet opbygning af APO-system

- Overvågning af datakvalitet og indsamling
- Overvågning af regularitet – er der tale om en driftsdag med mange aflysninger og store forsinkelser er beregningen ikke nødvendigvis brugbar



Figur 6. Overstruktur af Datavask

Eksempelvis justeres der for tilfælde hvor der kun er foretaget vejninger eller tællinger for nogen af togsættene i en togstamme. Der opsamles en detaljeret statistik over den gennemsnitlige passager-vægt fordelt på linie, gren, retning og tidspunkt (beregnet ved at sammenholde passager-tællinger og vejninger for togsættene med tælle-bjælker). Der advares hvis der fx er tale om en driftsdag med så store irregulariteter at resultatet formodentligt vil være upålideligt.

Slutresultatet, om alt går vel, er at input-data er konverteret til observationer i form af antal af- og påstigere, samt antal rejsende i togsæt, som kan anvendes direkte i MPME-beregningen.

F. Opbygning og teknik

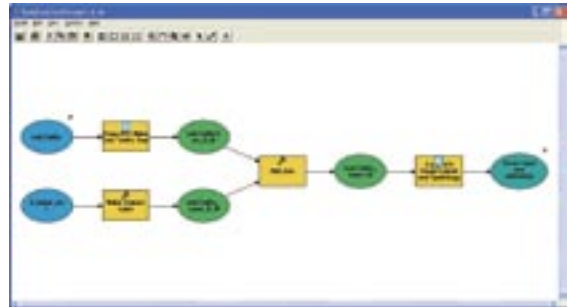
Rent teknisk er systemet opbygget som en række moduler, der udveksler data vha. nogen fælles database-formater. Modulerne er opbygget med udgangspunkt i eksisterende software udviklet af Rapidis, hvilket har nedbragt udviklingstiden betragteligt.

I sin grundlæggende form er APO-systemet et server-baseret system, som hver nat afvikler en samlet beregning. Men for at lette test af systemet, samt give mulighed for at lave forskellige scenarie beregninger, er alle moduler tilgængelige i to forskellige udgaver – en konsol udgave til server brug og som et beregningsmodul, der kan anvendes i ArcGIS, fra firmaet ESRI. ArcGIS understøtter ved hjælp af GeoProcessing muligheden for ved hjælp af diagrammer at sammenstykke og afvikle komplicerede beregnings processer. Dette beskrives mere detaljeret nedenfor.

De GIS-baserede udgaver kan således udvikles direkte fra ArcGIS. Desuden er brugeren heller ikke bundet til det overordnede data-warehouse, men kan f.eks. hente data ned til en lokal Access-database, og arbejde videre med dem der.

Brugen af GIS

GIS har spillet en vigtig rolle i løbet af projektet. GIS har i hovedsagen været anvendt i forbindelse med:



Figur 7. Eksempel GeoProcessing model - Ellipser repræsenterer Data-tabeller, rektangler er processer, f.eks. SQL-forespørgsler eller tilknytning til GIS-lag

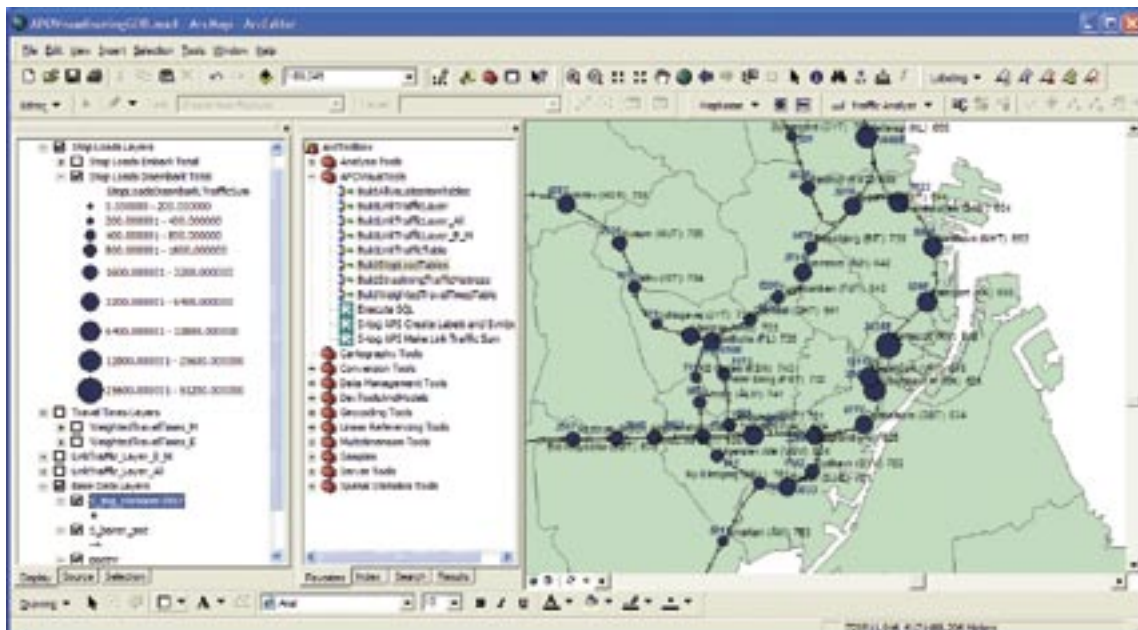
- Arbejdet med at undersøge og kvalitetssikre input-data
- Afvikle test- og scenarie-beregninger vha. GIS-udgaven af APO-komponenterne
- Samt, ikke mindst, være platform for afvikling af databaseforespørgsler og visualiseringer af resultater

GIS (i form af ArcGIS fra ESRI) vil fremover i den daglige drift især blive brugt til automatisk at generere oversigtsrapporter for dagens beregningsresultater. Dette er en vigtig funktionalitet da det grundlæggende beregningsoutput er meget detaljeret, og det typisk er ret tidskrævende manuelt at opstille visualiseringer og oversigts-forespørgsler.

Denne rapport-generering er opstillet ved hjælp af en kombination af ArcGIS GeoProcessing og database-forespørgsler. Dette betyder også at det er relativt simpelt for brugeren hen ad vejen at tilpasse og udvide funktionaliteten.

Et eksempel på et direkte output fra beregningsmodellen er en tabel over hver tog-bevægelse og de tilhørende passagertal, i form af:

- TogNr
- FraStation
- TilStation
- Start-Tidspunkt
- Slut-Tidspunkt
- Antal påstigere, fra-station
- Antal afstigere, til-station



Figur 8. Oversigtsvisualisering, antal afstignere pr station pr døgn

En sådan output-tabel indeholder typisk små 30.000 rækker data, hvilket umiddelbart kan være lidt svært at overskue. Ved at aggregere data og visualisere dem i GIS bliver det dels lettere at overskue resultaterne, men det bliver også lettere at formidle dem til andre brugere.

Eksempelvis vises i Figur 8 en opsummering af antal afstignere pr. station, dannet ud fra den ovenfor beskrevne resultat-tabel.

Ved at knytte andre beskrivende tabeller til resultaterne er det også muligt at granske dele af resultatet mere detaljeret, for eksempel opdelt pr. gren, pr. tidsperiode eller pr. linje, eller kombinationer heraf. Eksempelvis vises i Figur 9 mængden af passager der rejser med Linje B i morgen-perioden (kl. 6.00 – 9.00).

Foruden visualiseringer kan det også være relevant med almindelige oversigtstabeller. I Figur 10 vises et eksempel på en oversigt over selve tur-mønsteret (som i den grund-

læggende form er vist i Figur 3), summeret op på rejsende fra gren til gren i S-Togs nettet i morgen-periode.

Konklusion

Udviklingsdelen af APO-projektet er nu afsluttet, og systemet sættes pr. 1. september 2008 i parallel drift ved siden af det tidligere passager-opregnings-system. I takt med at der akkumuleres driftsdata og indsamles erfaringer vil systemet løbende blive justeret. Det forventes at opfølgning og justering afsluttes i løbet af det kommende år.

Med det udviklede APO system har DSB S-Tog fået et system, der er i stand til at udnytte det gode data-grundlag, der er tilgængeligt. Ved at vælge en utraditionel tilgang er det blevet muligt ikke blot at beregne de overordnede passagerantal, men også skabe en detaljeret indsigt i de rejsendes vaner og aktiviteter, som vil være overordentligt nyttig, både med hensyn til kort- og langsiget planlægning.



Figur 9. Visualisering af passager-strømme, Linie B i morgen-perioden (6.00 – 9.00)

Stationen	Tidspunkt	Antal	Retning	Station 2	Antal	Retning	Station 3	Antal	Retning	Station 4	Antal	Retning
Århus	06:00	100	Øst	Århus	100	Vest	Århus	100	Øst	Århus	100	Vest
Århus	06:00	100	Vest	Århus	100	Øst	Århus	100	Vest	Århus	100	Øst
Århus	06:00	100	Øst	Århus	100	Vest	Århus	100	Øst	Århus	100	Vest
Århus	06:00	100	Vest	Århus	100	Øst	Århus	100	Vest	Århus	100	Øst
Århus	06:00	100	Øst	Århus	100	Vest	Århus	100	Øst	Århus	100	Vest
Århus	06:00	100	Vest	Århus	100	Øst	Århus	100	Vest	Århus	100	Øst
Århus	06:00	100	Øst	Århus	100	Vest	Århus	100	Øst	Århus	100	Vest
Århus	06:00	100	Vest	Århus	100	Øst	Århus	100	Vest	Århus	100	Øst
Århus	06:00	100	Øst	Århus	100	Vest	Århus	100	Øst	Århus	100	Vest
Århus	06:00	100	Vest	Århus	100	Øst	Århus	100	Vest	Århus	100	Øst

Figur 10. Oversigtstabel, trafik fra gren til gren i morgen-perioden

Referencer

Nielsen, Otto Anker & Frederiksen, Rasmus Dyhr. Optimisation of timetable-based, stochastic transit assignment models based on MSA. Special Issue of Annals of Operations Research on Optimisation in Transportation. Elsevier. 2003-01-15

Seest, Elsebet; Nielsen, Otto Anker & Frederiksen, Rasmus Dyhr: Opgørelse af passagerregularitet i S-tog. Trafikdage på AUC (2005).

Nielsen, Otto Anker (1998). Two new methods for estimating Trip Matrices from Traffic Counts. Chapter in Travel Behaviour Research: Updating the state of play. Edited by Ortúzar, H. D., Hensher, D & Jara-Díaz, D. Elsevier Science Ltd. Oxford, UK. 1998. pp. 221-250

Om forfatteren:

Rasmus Dyhr Frederiksen, rdf@rapidis.com. Rapidis ApS

De gennemgående personer i projektet med hensyn til det faglige arbejde:

Elsebet Seest og Dorte Filges, DSB S-Tog , Metode-udvikling, integration med eksisterende systemer, formulering af S-Togs krav og behov, test og kvalitetssikring

Professor Otto Anker Nielsen, DTU Transport – Datakvalitet, Metode-udvikling

Lektor Bo Friis Nielsen, DTU Informatik – Datakvalitet, statistisk undersøgelse af data

Bjarke Brun, Philip Bagger, Rasmus Dyhr Frederiksen, Jakob Skriver, Rapidis ApS – Metode-udvikling, GIS, Design og udvikling af driftssystem, test og kvalitetssikring