

1. Anskaffelse af data

Et formål med projektet har været at prøve at udnytte eksisterende datakilder for at opnå en omkostningseffektiv dataindsamling. Brugere af flådestyringssystemer (fx pakkeleveringsfirmaer) anvender i forvejen GPS i deres daglige arbejde og indsamler on-line data med en vis frekvens. Disse firmaer er blevet kontaktet og en del af disse har indvilget i at levere deres historiske data under forudsætning om fuld anonymitet både med hensyn til firmaets navn og med hensyn til køretøjets identitet. Indtil nu har køretøjer med specielle hastighedsmønstre fx lastbiler eller busser ikke indgået i dataindsamlingen.

2. Data kvalitet

Flåderne leverer data af forskellig kvalitet. Af vigtighed for os er:

- indsamlingsinterval for GPS punkterne
- retningen (kursen) af køretøjet i hvert punkt
- præcisionen af hvert punkt
- høj stabilitet af udstyret
- korrekte køretøjshastigheder.

Vores analyser viser at der er behov for at data bliver indsamlet med et 15 sekunders interval (eller mindre), især for at detektere ærinder, men også for at de efterfølgende hastighedsberegninger bliver rimeligt præcise. For det andet er kursen på køretøjet vigtigt for at mapmatche køretøjet til det korrekte vejsegment. For det tredje skal nøjagtigheden af GPS-punkterne være mindst 10 meter for at mapmatche korrekt i et finmasket vejnet. For det fjerde må dataopsamlingsenheden ikke "falde ud", men sikre en jævn strøm af data, så reelle trafikale stop kan identificeres. For det femte skal hastigheden måles korrekt inden for nogle kilometer i timens nøjagtighed. Alle fem punkter kunne opfyldes af de kommercielle systemer, der er blevet analyseret.

3. Beregningstrin

De rå GPS-punkter er xy-koordinater, der ikke er relateret til vejnettet. GPS-positionerne skal derfor forbindes til vejnettet ved processen "Map-matching". Når antallet af GPS-punkter bliver meget stort, kræves en effektiv databasestruktur, for at processen kan lykkes indenfor overskuelig tid. Det er tillige valgt at bruge NavTeq's digitale vejnet som grundlag for map-matchingen.

En midling af de mapmatchede hastigheder vil give et forkert mål for hastigheden. Det er vigtigt at undgå, at såkaldte "ærinder" undervejs (f.eks. en levering) bliver opdaget, så tider og hastigheder fra stoppet ikke inddrages i de beregnede hastigheder på vejsegmenterne. Derfor er udviklet et specielt "ærinde-detekterings-værktøj". Alle punkter mellem to ærinder gives efterfølgende et unikt tur-nummer, som benyttes, når middelhastigheden skal beregnes for vejsegmentet. Segmenthastigheden findes derefter ved en ikke-triviell midling, da lave hastigheder vil være overrepræsenteret i rådata. Hvert beregningstrin gennemgås i det efterfølgende (fig. 1).

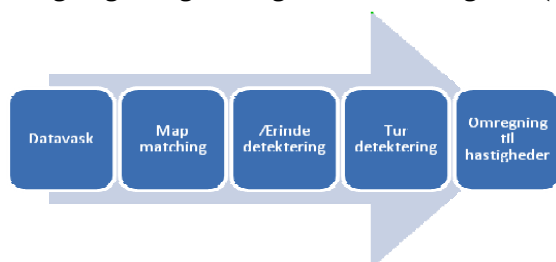


Fig. 1.

4. Map match algoritme

GPS rådata indeholder for hvert punkt en form for koordinater, ofte gemt som længde- og breddegrad. For at kunne bruge data efterfølgende, er det nødvendigt at afgøre, hvilken delstrækning i det digitale vejnet, punktet har befundet sig på ved en såkaldt map-match algoritme.

Algoritmen opererer som tidligere nævnt på hvert enkelt GPS-punkt hver for sig.

Dernæst benytter den en række parametre angående dette punkts relationer til vejnettet:

- afstanden til forskellige strækninger af forskellig type
- retningen af køretøjet registreret af GPS-enheden
- køretøjets hastighed sammenholdt med de mulige vejtypers hastighed.

I første trin prøver algoritmen at finde hvilke strækninger, der ligger inden for en given max. afstand af punktet. Jo tættere på, desto højere point gives der for, at det skal matches til strækningen blandt de udvalgte. Dernæst sammenlignes køretøjets hastighed med strækningens vejtype. Jo mere højklasset vejen er (fx motorvej) desto hurtigere må bilen køre. Hvis en grænseværdi overskrides, droppes punktet i relation til den vejtype. Endelig sammenlignes køretøjets retning og de potentielle strækningers "retning". Jo bedre de passer desto højere point gives. Til slut sammenvægtes kriterierne og GPS-punktet mapmatches til den strækning, der har det højeste antal point. For alle kriterier er der dog max.-grænse for hvor store afvigelsen må være.

Resultat-tabellen fra map-matchingen af ca. 100 mio. punkter, opdelt efter om punkterne efterfølgende blev udpeget som en del af en tur eller et ærinde, ser således ud (datavask havde fjernet ca. 20 mill. punkter):

Match Type	Alle punkter (1000)	Heraf i ture	Heraf i Ærinde
Afvist på forhånd	3	2	1
Afstand til vej for stor	13.949	796	13.153
Fart for høj	1.197	741	456
Kurs forkert	5.029	692	4.337
Match OK	77.989	65.528	12.461
I alt	98.167	67.759	30.408

Fig. 2.

Som det fremgår, lykkedes det at matche ca. $\frac{3}{4}$ af punkterne. Langt de fleste af de punkter der **ikke** blev matchet, er efterfølgende udpeget som en del af et ærinde. De ikke-matched punkter som er markeret som en del af en tur, fordeler sig nogenlunde ligeligt mellem følgende 3 muligheder:

- Punktet er for langt fra links i vejkategoriene 1-5
- Farten i punktet er for høj i forhold til vej kategorier i nærheden
- Kursen for punktet passer ikke til vejsegmenterne i nærheden

Disse bliver gennemgået herunder:

A) Afstand fra links

Hvis punktet er for langt fra de nærmeste links, kan det skyldes at:

- Præcisionen (koordinaterne) for punktet er for dårlig. Som sagt er den normale præcision rigtig god, men der kan være enkelte kiks. I ganske få tilfælde viser data i korte perioder koordinater og hastigheder der er helt forkerte.
- Det digitale vejnet er ikke altid placeret 100 % præcist, og nogle steder er det simplificeret.
- Præcisionen er god nok, men køretøjet befinder sig faktisk et sted, der ikke er med i vejnettet, fx på et parkeringsareal.

Fig. 3 viser eksempler på dette: De røde punkter er ikke-matchede. Den store røde klump øverst til højre er en parkeringsplads, så det er ok at de ikke er matchede (den tredje kategori). For vejen mod syd fra parkeringspladsen er det tydeligt, at langt de fleste punkter ligger til højre for vejen, og at en del af dem ligger for langt væk til at de matches, selv om det virker oplagt at de burde blive det. Formentlig en fejl af type to (tolerancen mht. afstand er siden sat lidt op). De helt alenestående punkter kan skyldes fejl af første type.

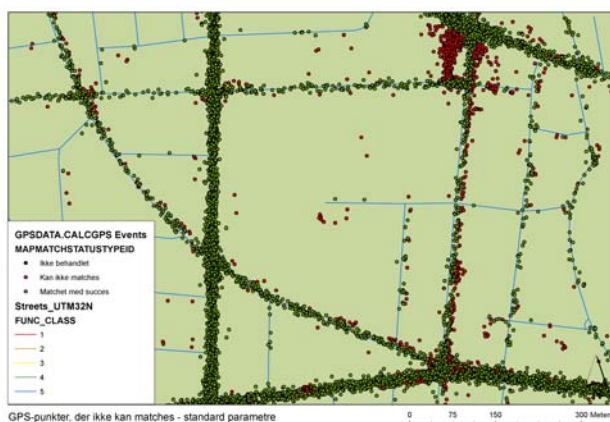


Fig. 3.

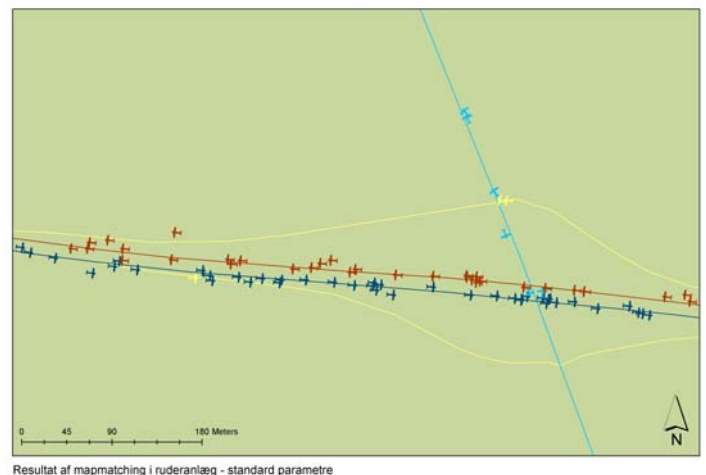


Fig. 4.

B) Hastigheden i forhold til hastighedsgrænsen

Fig. 4 viser match-resultatet opdelt, så farven på punktet passer med farven på den matchede strækning – ”flyver-ikonet” angiver retningen. Generelt ser det ok ud, men det punkt der er matchet til rampen (den gule flyver) til venstre, burde sandsynligvis være matchet til selve motorvejen. Næste punkt for det pågældende køretøj befinder sig nemlig ”stadig” på motorvejen længere imod højre.

Hastighedstolerancen er indført for at undgå, at et vildfarent punkt fra en motorvej (med fx 125 km/t) bliver matchet til en lokalvej. Kriteriet fungerer som regel godt, men specielt problematiske er ramper, som i realiteten skal kunne rumme de samme hastigheder som den tilhørende motorvej. Jo bedre den geografiske præcision af GPS-punkterne bliver, desto mindre skal hastighedsfølsomheden være.

C) Kurs

GPS-kursen har vist sig at være stabil, når køretøjerne er i bevægelse; men når køretøjerne holder stille, er der en del eksempler på, at kursen svinger. Der opereres med en max kursafvigelse på 70 grader, så et punkt skal næsten vende på tværs af køreretningen, før det sorteres fra af den grund. Det kan dog ske f.eks. ved en parkeringsmanøvre. Når køretøjet holder stille skal der kun en smule usikkerhed på koordinaterne til, før at kursen bliver forkert. Dette kan medføre at nogle punkter ikke bliver matchede, og tillige også at nogle punkter matches **forkert** - i nogle tilfælde til den modsatte retning af hvor køretøjet rent faktisk var på vej hen! I eksemplet fig. 5 har adskillige af punkterne med lav hastighed en forkert kurs, hvoraf nogle passer til en forkert strækning eller en forkert retning. Bemærk at farven angiver hastigheden.

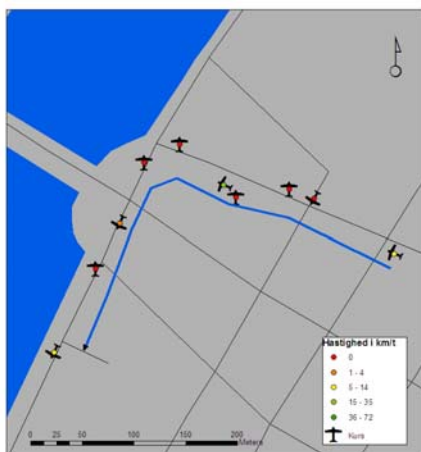


Fig. 5.

Skift i kurs, som **ikke** skyldes GPS-fejl, kan også opstå hvis et køretøj bakker fx under en parkeringsmanøvre i vejkanthen. Kursen angiver altså den retning køretøjet bevæger sig i, og som ikke altid er lig den retning køretøjer "peger". Analyser peger på at det måske er muligt automatisk at detektere og tilrette kursfejl, hvis frekvensen er omkring 1 sekund, men med 15 sekunder imellem punkterne er det meget svært. Det er dog vores vurdering, at det er en meget lille del af punkterne der har denne fejl.



Fig. 6.

D) Kombinationer (svingende bilister)

En speciel variation af problemstillingen med nærtliggende strækninger er svingbevægelser i kryds. I et simpelt kryds er der en vis risiko for at specielt venstresvingende bilister bliver matchet til forkerte strækninger. I fig. 6 et eksempel, hvor koordinaterne for, hvor bilen holder stille, samt selve svingbevægelsen, er placeret **efter** selve krydset ifølge NavTeq. Mange af disse punkter matches til den strækning som fortsætter ligeud (med lav hastighed), selv om den reelt ikke er brugt på turen. I det viste eksempel er centerlinien for sidevejen formentligt ikke præcist placeret. Hvis den havde været præcis, ville selve stoppet måske være blevet matchet til den rigtige strækning, men de første punkter af selve svingbevægelsen ville nok stadig ligge på den "forkerte" strækning (lige ud).

Den samme problemstilling gør sig i princippet gældende hver gang et køretøj passerer et sted hvor to strækninger mødes i forlængelse af hinanden, men det er kun i kryds, at det for alvor er kritisk, hvilken af strækningerne punkterne matches til.

I kryds kan problemet nemt betyde, at en del punkter med lave hastigheder kommer til at ligge på den forkerte strækning. Det er kritisk for metoden til at beregne hastigheder, både fordi at det kan generere en tur på en strækning som faktisk ikke er brugt, og fordi der kun registreres de lave hastigheder i forbindelse med svingmanøvren (så hastigheden på den forkerte strækning bliver lav). Tiltag for at forbedre map-matchingen er under overvejelse.

5. Ærinde detektering

De rådata, Vejdirektoratet har modtaget fra dataleverandørerne, stammer fra køretøjer i mange forskellige situationer. GPS-projektet er interesseret i hastigheder på vejnettet, og kun i trafikale situationer, men GPS-modtagerne samler ofte data ind, selv om at motoren er slukket. Således indeholder rådata en masse observationer fra f.eks. parkeringspladser og industriområder, som det er

vigtigt ikke får indflydelse på hastighedsanalyserne. Også data fra selve vejene, f.eks. under parkeringsmanøvrer, bør sorteres fra, idet de lave hastigheder ikke skyldes trafikale problemer. I de modtagne data har op imod 25 % af punkterne en hastighed på 0 km/t, men det er kun en mindre del af disse der foregår i trafikale situationer, i forbindelse med kryds og alvorlig trængsel. Det er derfor vigtigt at få detekteret de stop der skyldes ærinder stop o.l.

Begrebet ærinder bruges i denne sammenhæng både om deciderede ærinder (stop) og om andre situationer, hvor der er afbrud i data. Det sidste kan ske når GPS-udstyret er slukket eller ikke virker. Med så mange data som der opereres med her, er det nødvendigt at sætte algoritmer op til automatisk frasortering af ærinde-punkter. Det er vanskeligt at lave en frasortering, som er 100 % præcis. Specielt køretøjer, som bruges til f.eks. levering, har en del meget korte stop, og det kan være umuligt at skelne en parkering i vejsiden på 30 sekunder fra et stop for rødt lys.

Detektering af ærinder er en balancegang mellem at finde flest mulige reelle ærinder, samtidigt med at undgå at så få **trafikale** stop som muligt bliver markeret forkert som ærinder. Fordi det er besluttet at beregne hastigheder direkte på de enkelte segmenter, er det ikke så kritisk hvis en tur bliver delt op i mindre dele – så længe frasorteringen af punkter så vidt muligt sker uden bias mht. hastigheden.

Algoritmen er delt op i tre faser, hvor den første detekterer om et stop har fundet sted, den anden definerer hvilke GPS-punkter, der markeres som en del af stoppet og den tredje afgør om det er en reel tur, der er fundet. I det følgende betragtes ét køretøj, hvis GPS-punkter er sorteret efter dato og klokkeslæt (fig.7):

I første fase er der 5 forskellige kriterier for, hvornår et ærinde udpeges som stop i datamaterialet.

1. Et stop sættes automatisk ind for den første observation for et køretøj.
2. Ærinde-stop: Hvis køretøjet bevæger sig for lidt (<40 meter) for længe (over 2½ minut). De 40 meter beregnes ud fra koordinaterne - selve punkthastigheden bruges ikke.
3. Slukning/udfald af GPS-enheden i en periode (50 sekunder). Dette kan være en indikation af at et stop har fundet sted, eller det kan skyldes tekniske problemer med udstyret, eller at køretøjet kører i et område med mange høje bygninger eller lign.. Under alle omstændigheder betyder fænomenet, at vi ikke ved hvad køretøjet foretager sig i den periode, og dermed ikke ved, om der har været et stop/ærinde eller ej.
4. Statusvariabel er sat til stop og hastighed under 50 km/t. En dataleverandør har en statusvariabel som ofte bruges, når en vareudlevering slutter. Variablen bruges ikke helt konsistent, og er derfor kombineret med et hastighedscheck.
5. Ingen map-matched punkter i 65 sekunder: En periode uden at observationer er matchet til vejnettet, kan være en indikation af, at køretøjet er ”på afveje”, dvs. uden for det offentlige vejnet og det sker ofte forbindelse med et ærinde.

Disse parametre forudsætter et samplingsinterval på højst 15 sekunder. Det har vist sig problematisk at detektere ærinder, hvis samplingsintervallet når op på 1 minut.

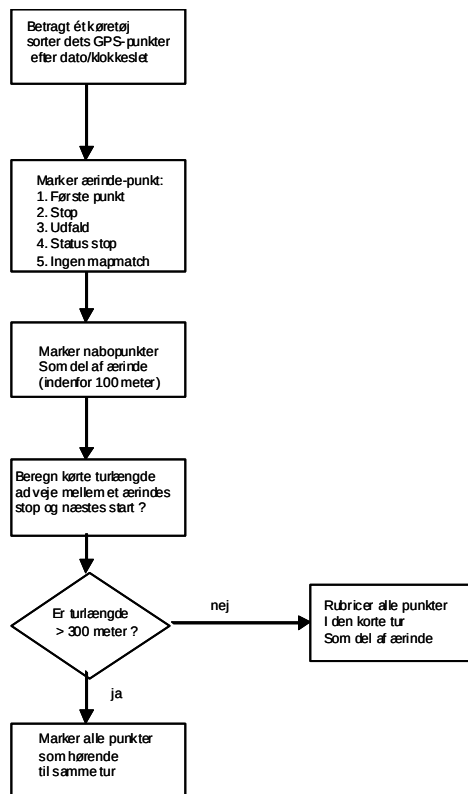


Fig. 7



Fig. 8.

Derefter træder den anden fase af algoritmen i kraft, og den forsøger at identificere de punkter, som menes at høre til ærindet. Tilstødende punkter markeres som en del af ærindet, hvis:

- Afstanden til ærindemarkeringen er under 100 meter (i fugleflugtslinie) både før og efter ærindet.
- Første punkt i en tur skal være matchet til vejnettet.

Kernen i udpegningen af punkter er altså, at alle punkter før og efter et detekteret ærinde markeres som en del af ærindet. Dermed markeres ofte flere punkter, end der strengt taget hører med til selve ærindet, men det sker uden at punkthastigheden er kriterie. Det er valgt, fordi at det sikrer at sving- og parkeringsmanøvrer direkte i forbindelse med ærindet, ikke kommer til at tælle med i efterfølgende analyser af hastigheder på de nærliggende vejstrækninger.

Som den tredje del af algoritmen defineres en "tur" som værende alle observationer mellem to ærinder, og turens samlede længde beregnes ud fra afstanden fra punkt til punkt (akkumuleret over turen). Definitionen af en tur benyttes i den efterfølgende hastighedsberegning. Punkter fra en sammenhængende tur på **under** 300 meter vurderes til at være problematiske, og bliver udeladt og markeret som del af "nabo"-ærindet.

I eksemplet fig. 8 kommer køretøjet fra nord, og tætheden af punkterne antyder at det har haft lavere fart i det store kryds, lige der hvor markeringen skifter fra tur (grøn) til ærinde (rød). Men ærinde-markeringen afhænger altså ikke af hastighederne.

6. Beregning af rejsetider

Et væsentligt formål med GPS-projektet er at beregne middelrejsetiden for en trafikant i specifikke tidsperioder på typiske hverdage, for eksempel mellem kl. 8.00 og kl. 8.30. Derfor opbygges en database, der samler rejsetiden i hvert segment i det digitale Navteq kort, opdelt på halvtimes tidsperioder på typiske hverdage (=ikke-ferieperioder).

Normalt leveres data for hastighed og position hvert 15 sekund. Det er en udfordring i beregningen af rejsetiden, fordi sandsynligheden for at indsamle en observation på et vejsegment afhænger af køretøjets hastighed. En simpel middeltalsdannelse af rejsetiderne på et givent vejsegment vil overestimere rejsetiden. Derfor må denne tendens justeres. Metoden er udviklet på basis af 1 sekunds interval data. 1 sekunds data giver et præcist estimat af segment-rejsetiden og kan bruges til at sammenligne med estimater baseret på 15 sekund interval data. Proceduren inkluderer også en metode til at behandle hastighedsdata med 0 km/t i rejsetidsberegningen.

Der skal ske en sammenvejning af hastigheder logget under gennemkørsel af den givne delstrækning. Der tages udgangspunkt i den gængse formel for beregning af strækningsmiddelhastighed, det vil sige, at der midles over rejsetiderne. Udtrykket i formel (1) indeholder to bidrag. Det første bidrag beregnes på basis af observationer, hvor der er registreret en hastighed større end nul. Andet bidrag er tiden fra observationer med 0 km/t. Logges eksempelvis med 15 sekunder adderes antallet af 0'er (m) ganget med 15 sekunder. Betragtes et bestemt segment i ét bestemt tidsinterval, kan middelrejsetiden estimeres som:

$$(1) \quad \hat{t} = \frac{L}{n} \sum_{i=1}^n \frac{f(v_i)}{v_i} + m \tau \quad \text{hvor } v_i > 0$$

hvor n = antal målinger hvor hastighed er større end nul

m = antal målinger hvor hastighed er nul

v_i = målt hastighed

$f(v_i)$ = vægtningsfunktion

L = længde af delstrækning

τ = logningsinterval

Den gennemsnitlige hastighed (strækningsmiddelhastighed) på dette segment og i dette tidsinterval er efterfølgende givet ved:

$$(2) \quad \hat{v} = \frac{L}{\hat{t}}$$

I udtrykket (1) indgår vægtningsfunktionen $f(v)$, da sandsynligheden for en observation afhænger af den kørte hastighed. Hvis bilisten kører langsomt, er der eksempelvis større sandsynlighed for, at den næste hastighed logges på samme strækning, Hvis bilisten derimod sætter hastigheden betydeligt op, så falder sandsynligheden for, at næste hastighed logges på samme strækning. Logninger af lave hastigheder på en given strækning må derfor formodes at være overrepræsenteret.

Det er ikke muligt at bestemme sandsynlighedsfordelingen teoretisk til vægtning af observationer. På basis af data fra udvalgte strækninger er forskellige funktioner undersøgt og vurderet. Generelt opnås de bedste resultater ved brug af følgende vægtningsfunktion:

$$(3) \quad f(v) = 1 - e^{-\beta v}$$

hvor β er en kalibreret konstant.

Denne funktion er fundet ved kalibrering med 1 sek's data, hvor hver 15'tende observation er udtrukket. Vægtningsfunktionen benyttes når man har data i 15 sekunders intervaller. Hvis man har andre samplingsintervaller, må man kalibrere andre vægtningsfunktioner.

Bemærk at grundenheden for beregning af rejsehastigheden på et delsegment er en "gennemkørsel" af delstrækningen. Det er valgt, da vi har køretøjer med forskellige samplingsintervaller (1 sek og 15 sekunder) på samme delstrækning. Deres "gennemkørsler" skal helst veje lige meget. Fremgangsmåden forudsætter at de punkthastigheder der er målt som indgående i "gennemkørslen" er repræsentative for en komplet "gennemkørsel" af delstrækningen. Det behøver ikke at være tilfældet. Specielle "gennemkørsler" skal derfor frasorteres, inden rejsehastigheden på delsegmentet beregnes ved formlerne (1) og (2).

Populært sagt beregner man det forventede antal punktobservationer, man kan forvente på en delstrækning af den givne længde, med det givne samplingsinterval og med den hastighed, der er på segmentet. Hvis ikke en rimelig procentdel af delstrækningen "gennemkøres" forkastes observationen fra denne "gennemkørsel". Det er et kompliceret regelsæt, der tager hensyn til mange specialtilfælde, fx rene 0-hastigheder, rene 1 punkts "gennemkørsler", delstrækningslængder som kan gå fra 20 meter til 2 km. Slutresultatet er dog, at man har en database over rejsehastighederne på delsegmenter opdelt i ½-times intervaller for typiske hverdage på det overordnede vejnet. I fig. 10 er vist et tid-sted diagram for Køge Bugt motorvejen. Y-aksen har ½-times intervaller øverst fra kl. 6:00 til nederst kl. 20:00. X-aksen har et felt for hver delstrækning på ruten fra Torvet i Køge via Køge Bugt motorvejen til Rådhuspladsen i København. Hvert lille felt i dette diagram har fået en rejsehastighed estimeret ved metoden ovenfor som udtryk for den "gennemsnitlige" rejsehastighed på et typisk hverdagsdøgn.

7. Verifikation af rejsetiderne

De resulterende rejsetider, snithastigheder og rejsehastigheder er blevet sammenlignet med andre dataindsamlingsmetoder i Vejdirektoratet. Vejdirektoratets tællesystem Mastra registrerer hastighederne i et givet snit fortrinsvis ved spolemålinger af samtlige køretøjer. Fig. 9 viser hastighedsprofilen over døgnet for en Køge Bugt motorvejen for en typisk hverdag i 2007.

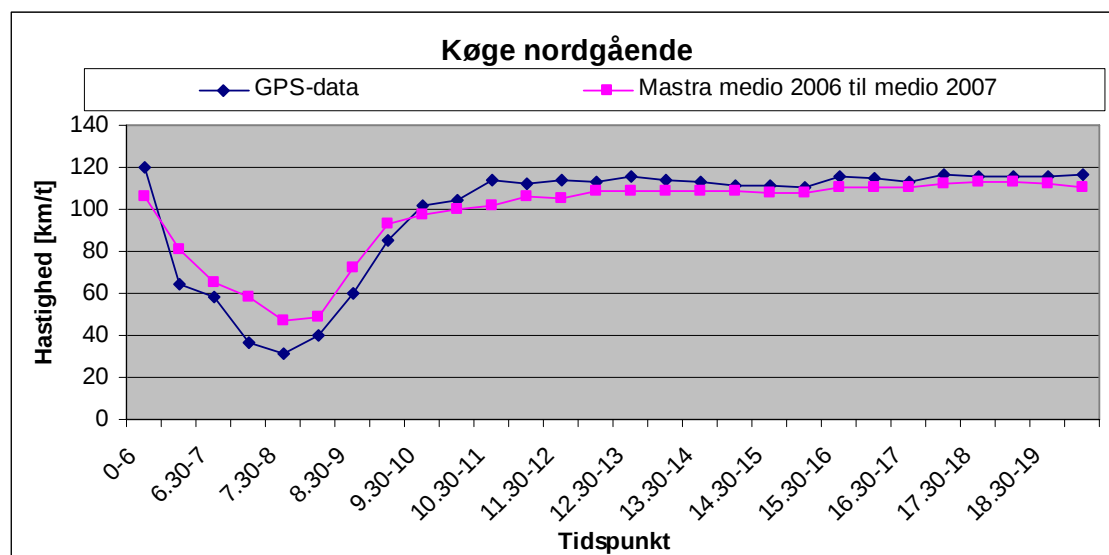


Fig. 9.

Begge kurver viser, at der er køproblemer i morgenmyldretiden. Derudover ses, at hastighederne fra Køge mod nord for de to datasæt er meget ens. Det er kun i morgenmyldretiden, at GPS-data ligger

væsentligt lavere (op til 22 km/t lavere). Forskellen kan skyldes, at Mastra's spolebaserede system ikke er i stand til at måle hastigheder under ca. 15 km/t. Netop i myldretiden er der mange køretøjer med sådanne lave hastigheder. Desuden indeholder GPS-data på nuværende tidspunkt ikke lastbiler hvilket forklarer forskellen ved fri hastighed. I alt er der foretaget sammenligninger på en snes strækninger på denne måde med tilfredsstillende resultat.

Udover dette har man også sammenlignet med rejsetiderne fundet ved kamerasystemerne med nummerplade-genkendelse og rejsehastigheder med kørebilmålinger i myldretiderne. Alle med en tilfredsstillende verifikation.

8. Anvendelser

Middelrejsehastighederne på vejsegmenterne på hele det overordnede vejnet i Danmark giver et godt billede af trængselsforholdene i de daglige myldretider og visualiserer, hvor der er flaskehalse på typiske hverdage. Især kan man se, hvor der er problemstrækninger uden for de motorveje, der normalt følges tæt. Der er således udarbejdet kort der visualiserer trængselsforholdene i Danmark.

Rejsehastighedsdatabasen på basis af GPS data vil i 2009 blive brugt af Rejseplanlæggeren A/S på deres hjemmeside til en mere pålidelig, trængselsafhængig bestemmelse af bilisters rejsetid i myldretimerne.

Desuden vil man ultimo 2009 udvikle et brugervenligt I/F der muliggør at vejadministrationerne til brug for deres planlægning kan udtrække data som ovenstående for alle strækninger på det overordnede vejnet i Danmark.

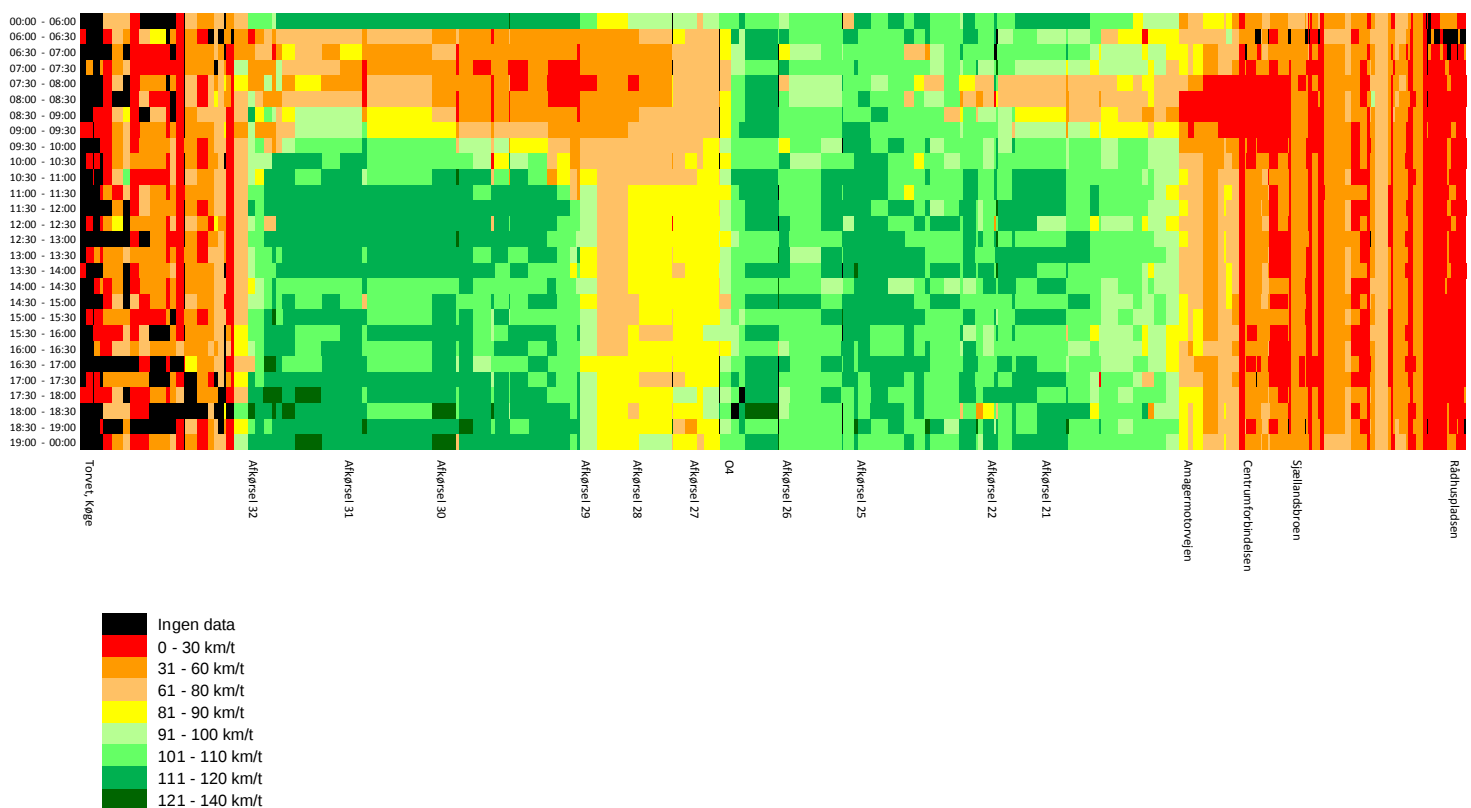


Fig. 10.