

U
n
i
v
e
r
s
i
t
e
t

Heidi Bull-Berg, Nils Olsson og Anette Sørensen

Mobiltelefondata – en potensiell datakilde i evaluering av jernbaneutbygginger?

Arbeidsrapport

Forord

Big Data er et felt i rask utvikling, som åpner for muligheter i forbindelse med analyse, planlegging, drift og evaluering av store statlige investeringsprosjekter. Det kan være mobiltelefondata eller andre bevegelsesrelaterte data, store datamengder fra internettaktivitet, eller fra kommersielle transaksjoner.

I 2014 gjorde Concept-programmet en forstudie om temaet, hvor vi kartla hvilke typer av Big Data som i årene fremover kan være aktuelle i analyse og evaluering av statlige investeringsprosjekter. Studien konkluderte positivt – gitt at en ivaretar personvern og andre hensyn er det et betydelig potensial for å bruke Big Data, både alene og i kombinasjon med annet materiale.

Dette er nybrottsarbeid og krever en trinnvis uttesting ved bruk av piloter. Concept-programmet engasjerer seg vanligvis i mer anvendt forskning, men har i dette tilfellet gjort et unntak og finansiert en studie som kan anses som første trinn i en undersøkelse av hvordan mobiltelefondata kan brukes til å kartlegge passasjertrafikk på jernbane.

Studien viser at potensialet antakelig er stort. Ikke bare vil mobiltelefondata kunne erstatte togoperatørens egne tellinger, som forskerne sjelden får tilgang til grunnet forretningshensyn, men det åpner seg helt nye muligheter for å beskrive hele reisen fra dør til dør inkludert skifte mellom transportformer og andre ærender. Dette er informasjon som en i dag kun får gjennom spørreundersøkelser som baserer seg på befolkningens hukommelse med de begrensningene det innebærer.

Denne studien er bare et lite skritt på veien, men det er all grunn til å henge med på utviklingen videre.

Trondheim, mai 2017

Sammendrag

I tidligfase av store investeringsprosjekter er gjennomføring av samfunnsøkonomiske analyser en godt etablert praksis for å vurdere om et prosjekt skal anbefales gjennomført eller ikke. Og i etterkant bør det gjennomføres evalueringer for å undersøke de faktiske resultatene, herunder om prosjektet nådde sine mål og var samfunnsøkonomisk lønnsomt. Informasjon om passasjertrafikken er vanligvis avgjørende både i ex ante og ex post evaluering. Samtidig har vi erfart at slike data er mangelfulle. I ex ante analyser benyttes transportmodeller som simulerer endringer i trafikken, men data for antall reisende og deres reisevaner som inngår som inputdata i de ulike modellene og verktøyene, er i dag stort sett basert på manuelle tellinger og reisevaneundersøkelser, som har flere svakheter. I ex post evaluering behøver man data om den faktiske trafikken. Togoperatørene gjør gjerne egne tellinger, men også disse har svakheter, og konsulenter og forskere får typisk ikke tilgang til dem.

I 2014 ble det gjennomført en studie finansiert av Concept-programmet kalt "Nye datakilder i evaluering av store statlige investeringer – et potensial for Big Data?" (Bull-Berg og Olsson, 2014). Studien kartla og drøftet potensialet for at nye, store datakilder (såkalte Big Data) kan brukes til evaluering av store statlige investeringer, mer spesifikt innenfor sektorene transport og bygg. Studien konkluderte med at det synes å være store muligheter for bruk av nye (store) data i evaluering, og anbefalte å utføre piloter som ser på mulighetene for utnyttelse av denne typen data i evaluering av store statlige investeringstiltak innenfor transport og bygg.

Big Data er datasett som er så store at de ikke er egnet til å hverken innhente, lagre, prosessere eller analysere ved hjelp av tradisjonelle databaseverktøy. Tradisjonelle kjennetegn er volum, hastighet, variasjon og aktualitet. Bruk av Big Data handler om å hente ut innsikt for å kunne ta kunnskapsbaserte avgjørelser. Det store fortrinnet er muligheten til å koble utallige datakilder for å se nye sammenhenger, mønstre, effekter mv. Det er mulig å "oppdage ting man ikke visste man lette etter". Man kan prosessere nye typer data og utnytte ustrukturert informasjon.

Big Data kan deles inn i følgende kategorier etter måten de samles inn/genereres på:

- Internettaktivitet, inkludert aktivitet på sosiale media og data fra søkemotorer (cookies, tekst, klikk)
- Bevegelsesrelaterte data, inkludert GPS, mobiltrafikk og lokalisering, bomstasjoner, RFID-brikker på gods
- Data om fysiske omgivelser
- Kommersiell aktivitet og bruk av betalingstjenester

Det er flere måter å skaffe data om antall reisende på tog. Manuelle kilder som tellinger og spørreundersøkelser har vært den mest benyttede metoden. Nyere studier indikerer likevel en dreining mot å benytte automatiserte (gjærne elektroniske) teknologier for telling og registrering. Flere studier er blitt gjort for å undersøke bruk av mobiltelefondata for å kartlegge transportmønstre og reisevaner. Litteraturen viser at mobiltelefondata ser ut til å ha et potensiale for å beskrive folks bevegelsesmønstre, som et alternativ (eller supplement) til tradisjonelle transport- og reisevaneundersøkelser.

Studien som nå er gjennomført med finansiering fra Concept-programmet er en pilot, hvor målet har vært å teste tilgjengelighet og bruk av mobiltelefondata til å kartlegge personreiser med jernbane, som vil være relevant i ex post og ex ante evaluering av investeringstiltak i jernbaneinfrastruktur. Mer konkret har vi analysert mobiltelefondata, opp mot manuelle tellinger av togpassasjerer og punktlighetsdata for tog, for å se hvordan disse kildene i kombinasjon kan utnyttes til å estimere antall reisende på tog.

Vi har fått tilgang til mobiltelefondata fra Telenor, basert på en kilde med aggregert data for totalt antall oppkoblinger per basestasjon. Dataene er klassifisert til å ikke være personopplysninger av Telenors personvernombud, og i henhold til veilederen til Norsk senter for forskningsdata (NSD). Data over antall reisende og punktlighet har vi fått fra et jernbaneselskap (ønsker å være anonymt).

Som et første steg i å se på egnetheten i bruk av mobiltelefondata for å beregne antall passasjerer, ønsket vi gjennom en empirisk analyse å se på om hendelsesmønsteret (antall oppkoblinger mot basestasjonene) i mobiltelefondataene var sammenfallende med den faktiske togtrafikken på utvalgte strekninger. Var dette tilfelle ville det indikere at mobiltelefondata også kan være valide som kilde til å kunne si noe om passasjerene som finner seg om bord på disse togene. Studien ser nærmere bestemt på fem basestasjoner som har plassering i umiddelbar nærhet til jernbanelinja. Antall registrerte hendelser koblet med tidspunktet togene passerer basestasjonene ble analysert systematisk med en algoritme som ble utviklet i studien.

Ved å så kombinere tider for faktiske togpasseringer, antall registrerte hendelser på basestasjonene og faktiske passasjerdata (automatiserte tellinger) for de samme strekningene på samme tidspunkt, ønsket vi å se på muligheten for å kunne predikere passasjerantallet ved hjelp av mobiltelefondata. Siden de mobiltelefondata vi hadde tilgjengelig ikke representerer alle reisende (kun én leverandør - Telenor) undersøkte vi på om APC-data kunne fungere som en kilde til validering av *forholdet* mellom antall passasjerer og antall hendelser på de tilstøtende basestasjonene. Slike forholdstall kan være en nøkkel til å bruke mobiltelefondata som mål på antall reisende. Tanken er å finne nøkkeltall av typen «for hver registrert hendelse på basestasjon er det x antall reisende på toget».

Studien viser at det er mulig å kombinere mobiltelefondata med data fra jernbaneselskaper som viser infrastruktur, planlagt og faktisk togtrafikk, samt APC-data med passasjertall. Forutsatt at man velger basestasjoner med omhu og har egnet frekvens for målinger på basestasjoner, så er det god sammenheng mellom tidspunkt for passeringer av tog og endringer i antall registreringer på basestasjonene. Det ser også ut til å være mulig å beregne forholdstall som kan gi indikasjon på antall reisende, men for å gjøre dette bør man sannsynligvis ha data for lengre tidsperioder. I tillegg bør forholdstallene tilpasses hver delstrekning og basestasjon, i tillegg til å tilpasses etter ukedag og tid på døgnet. Grunnlaget er derved lagt for å i senere runder bruke mobiltelefonregistreringer til å estimere antall reisende, og gir også et potensial til å følge passasjerflyten i og utenfor jernbanen. Ideelt sett skulle vi også ville måle antall reisende før og etter større endringer, som ny infrastruktur, ny ruteplan etc. Dette vil derimot være en oppgave for etterfølgende studier, fortrinnsvis basert på denne typen data.

Studien har resultert i to artikler til publisering i vitenskapelige journaler: 1) en empiri-basert artikkel som illustrerer faktisk bruk av mobiltelefondata for å kunne estimere antall reisende på

jernbane, og 2) en artikkel basert på en litteraturstudie over temaet. Denne arbeidsrapporten bygger på funn i de to vitenskapelige artiklene.

Innhold

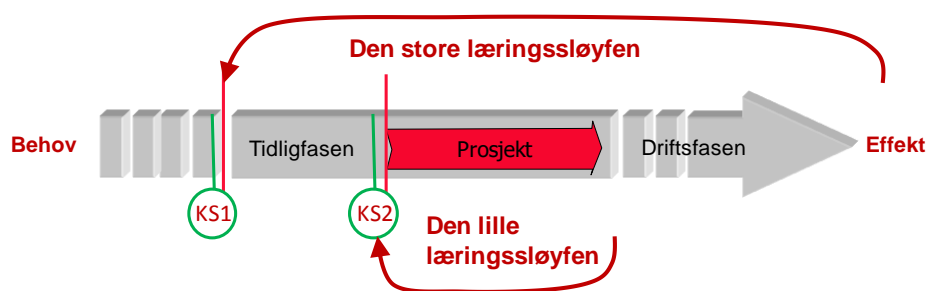
FORORD	1
SAMMENDRAG	2
INNHOOLD	5
1 INNLEDNING	6
1.1 BAKGRUNN OG FORMÅL MED STUDIEN	6
1.2 KORT OM EVALUERING AV JERNBANEUTBYGGINGER	8
2 METODER FOR ESTIMERING AV PERSONREISER MED JERNBANE	10
2.1 ULIKE METODER OG DATAKILDER.....	10
2.2 FLERE STUDIER VISER POTENSIAL FOR BRUK AV MOBILTELEFONDATA	14
2.3 MULIGHETER FOR Å AVDEKKE REISEMØNSTRE	15
3 DATA VI HAR BENYTTET I ANALYSEN	17
3.1 MOBILTELEFONDATA	17
3.2 PUNKTLIGHETSDATA	18
3.3 PASSASJERDATA (APC-DATA).....	19
4 ANALYSER OG RESULTATER	20
4.1 BEREGNING AV TIDSPUNKT FOR NÅR TOGET PASSERER BASESTASJON	20
4.2 KOBLING AV TIDSPUNKT FOR TOGPASSERING OG REGISTRERTE HENDELSER I MOBILTELEFONDATA	21
4.3 ALGORITME FOR Å TREKKE UT MÅLETIDSPUNKT SOM SAMSVARER MED TIDSPUNKT FOR FAKTISKE TOGPASSERINGER	27
4.4 SAMMENLIGNING MED APC-DATA.....	27
4.5 TIDSINTERVALL MELLOM REGISTRERTE HENDELSER – 5 MIN. VS. 1 MIN.	29
5 KONKLUSJON OG VIDERE ARBEID	31
REFERANSER	33

1 Innledning

1.1 Bakgrunn og formål med studien

Tilgang på gode relevante data kan være en utfordring ved evaluering av store statlige investeringer (Volden og Samset 2013). På oppdrag fra Concept evaluerte Nilsson et al. (2012) utbyggingen av dobbeltspor mellom Sandvika og Asker noen år etter ferdigstillelse av prosjektet. Studien påpeker at det var mangel på informasjon om antall reisende før og etter utbyggingen, noe som gjorde at det ikke var mulig å gi en samlet vurdering av verken måloppnåelsen (økt kollektivandel) eller utbyggingens samfunnsøkonomiske lønnsomhet. Ingen av Concepts utførte evalueringer har fått tilgang på antall reisende, til tross for at passasjertall er en vesentlig parameter ved evaluering av et stort jernbanetiltak. Eksempelvis står det i evaluering av nytt dobbeltspor Sandnes-Stavanger at "Det finnes ikke statistikk på antall reisende på dobbeltsporet før 2012" og "Enkle beregninger vi har foretatt basert på antall reisende på Jærbanen i 2008 oppgitt i nyhetsartikler tilsier at dette er et rimelig anslag" (side 28 i rapporten av Oslo Economics, 2015). Tilsvarende skriver Nilsson et al. (2012) om Asker-Sandvika at de har etterspurt data om antall reisende på strekningen Asker-Sandvika i tidsrommet mellom 2000 og 2010, men at de fikk beskjed om at det ikke finnes konsistente data for denne strekningen og tidsrommet. Det bør derfor være sterkt ønskelig å få tilgang til alternative data for antall reisende i fremtidige evalueringer.

Mangel på data kan synes som et paradoks, når omfanget av data generelt sett øker. Dette har vært vårt utgangspunkt for å se på potensialet for å bruke alternative datakilder i evaluering av store statlige investeringer. Med statlig investeringsvirksomhet menes de investeringer som finansieres, planlegges og gjennomføres av sentralforvaltningen. For statlige investeringsprosjekter med kostnadsramme over 750 millioner kroner gjelder Finansdepartementets kvalitetssikringsordning (KS) hvor prosjektene skal gjennomgå ekstern kvalitetssikring i to formelle beslutningspunkter, KS1 (før vedtak i regjering) og KS2 (før vedtak i storting) som vist i Figur 1. Concept-programmet driver følgeforskning på prosjektene som er en del av ordningen, samt på selve ordningen (metodeutvikling, effekten av den etc.). Programmet gjennomfører blant annet evaluering av de statlige investeringene (ex post og ex ante) for å styrke den store læringsløyfen som vist i Figur 1.



Figur 1 Kvalitetssikringsregimet (Volden og Samset 2013)

I 2014 ble det gjennomført en studie finansiert av Concept-programmet kalt "Nye datakilder i evaluering av store statlige investeringer – et potensial for Big Data?" (Bull-Berg og Olsson, 2014). Studien kartla og drøftet potensialet for at nye datakilder (såkalte Big Data) kan brukes til evaluering av store statlige investeringer, mer spesifikt innenfor sektorene transport og bygg.

Stadig større mengder data blir generert gjennom bruk av nye teknologiske systemer og løsninger, såkalte "store data", ofte betegnet som Big Data. Big Data har både i Norge og internasjonalt blitt det nye begrepet på denne typen data. Denne typen data kjennetegnes ved at de er så store at de ikke er egnet til å hverken innhente, lagre, prosessere eller analysere ved hjelp av tradisjonelle databaseverktøy. Tradisjonelle kjennetegn er volum, hastighet, variasjon og aktualitet. Bruk av Big Data handler om å hente ut innsikt for å kunne ta kunnskapsbaserte avgjørelser. Det store fortrinnet er muligheten til å koble utallige datakilder for å se nye sammenhenger, mønstre, effekter mv. Det er mulig å "oppdage ting man ikke visste man lette etter". Man kan prosessere nye typer data og utnytte ustrukturert informasjon. Big Data kan deles inn i følgende kategorier etter måten de samles inn/genereres på:

- Internettaktivitet, inkludert aktivitet på sosiale media og data fra søkemotorer (cookies, tekst, klikk)
- Bevegelsesrelaterte data, inkludert GPS, mobiltrafikk og lokalisering, bomstasjoner, RFID-brikker på gods
- Data om fysiske omgivelser
- Kommersiell aktivitet og bruk av betalingstjenester

Evalueringer av store statlige tiltak kan adressere oppnåelse av prosjektets mål på ulike nivåer, i tillegg til samfunnsøkonomisk lønnsomhet. Resultatmål omfatter typisk ferdigstillestid, kostnad og oppfyllelse av spesifikasjonen. Effektmål relateres til bruk av prosjektet og samfunns mål angir de langsiktige og overordnede virkningene av prosjektet. Vedrørende evaluering av tiltak, og store statlige investeringer spesielt, har vi i studien fokusert på muligheter til å bruke Big Data som en (av flere) kilde(r) til evaluering av effektmål, og der det synes hensiktsmessig også samfunns mål. Bakgrunnen for dette fokuset er at vi vurderer at det er for disse typer av mål som det har vist seg utfordrende å finne gode data til bruk i evaluering. Data om resultatmål er i de fleste tilfeller intern informasjon som finnes i den utførende etat. Vi kjenner likevel til pågående forskning som undersøker hvordan Big Data mer generelt kan brukes til operativ oppfølging av tiltak med hensyn på resultatmål, men da typisk for å gi prosjektlederen tidlig varsling om eventuelle avvik. I tillegg til evaluering av måloppfyllelse mener vi Big Data kan ha et potensiale når det gjelder å måle ulike tilsiktede og utilsiktede eksterne virkninger av en investering, og eventuelt sammenstille alle typer virkninger og beregne den samfunnsøkonomiske lønnsomheten av tiltaket. En kan også tenke seg at nye datakilder kan avsløre eksterne virkninger man før kanskje ikke hadde muligheten til å observere.

Det finnes eksempler på bruk av Big Data i evaluering, men det er fortsatt et felt i utvikling, preget av casestudier for å illustrere muligheter. Flere eksempler inkluderer transport (Ferris et al. 2010) og folks atferd i byer (MIT 2013; Arikawa et al. 2008).

Bull-Berg og Olsson (2014) sin studie konkluderte med at det synes å være store muligheter for bruk av nye (store) data i evaluering, og anbefalte å utføre piloter som ser på mulighetene for

utnyttelse denne typen data i evaluering av store statlige investeringsiltak innenfor transport og bygg.

Studien som nå er gjennomført med finansiering fra Concept-programmet er en pilot, hvor målet har vært å teste tilgjengelighet og bruk av mobiltelefondata til å kartlegge personreiser med jernbane, som vil være relevant i ex post og ex ante evaluering av investeringsiltak i jernbaneinfrastruktur. Mer konkret har vi analysert mobiltelefondata, manuelle tellinger av togpassasjerer og punktlighetsdata for tog, for å se hvordan disse kildene i kombinasjon kan utnyttes til å estimere antall reisende på tog. Vi har fått tilgang til mobiltelefondata fra Telenor, basert på en kilde med aggregerte data per basestasjon (totalt antall oppkoblinger mot stasjonen). Disse dataene er klassifisert til å ikke være personopplysninger. Data over antall reisende har vi fått fra et jernbaneselskap. Punktlighetsdata og rutetabeller kommer fra Jernbaneverket (nå Bane NOR).

Studien har resultert i to artikler til publisering i vitenskapelige journaler: 1) en empiri-basert artikkel som illustrerer faktisk bruk av mobiltelefondata for å kunne estimere antall reisende på jernbane, og 2) en artikkel basert på en litteraturstudie over temaet. Denne arbeidsrapporten bygger på funn i de to vitenskapelige artiklene.

1.2 Kort om evaluering av jernbaneutbygginger

I tidligfase av store investeringsprosjekter er gjennomføring av samfunnsøkonomiske analyser en godt etablert praksis for å vurdere om et prosjekt skal anbefales gjennomført eller ikke. Bakgrunnen for å bruke slike analyser er at samfunnet rår over begrensede ressurser, og det er viktig at disse fordeles på riktige prosjekter til rett tid. Analysene benyttes som en del av beslutningsgrunnlaget både for investering og for prioritering mellom prosjekter. Den mest utbredte analysen er nyttekostnadsanalysen. Her framstilles nytte og kostnader basert på nåverdiprinsippet, og resultatet gir et tilsynelatende klart svar på hvorvidt et prosjekt eller tiltak er samfunnsøkonomisk lønnsomt. Dersom en følger lønnsomhetsprinsippet i beslutningstakingen, skal en prioritere alle tiltak som har positiv lønnsomhet. Hovedformålet med analysen er å synliggjøre alle relevante konsekvenser som kan oppstå ved gjennomføringen av et tiltak eller et prosjekt, og vurdere nytten av prosjektet mot de kostnadene prosjektet medfører.

Som beskrevet i forrige avsnitt skal det for statlige investeringsprosjekter med kostnadsramme over 750 millioner kroner gjennomføres en ekstern kvalitetssikring av konseptvalget. Ordningen med ekstern kvalitetssikring av konseptvalget (KS1) krever at det skal gjennomføres en alternativanalyse av nullalternativet og minst to andre konseptuelt ulike alternativer. Alternativanalysen skal være bearbeidet i en samfunnsøkonomisk analyse. I denne analysen skal alle virkninger av de alternative konseptene identifiseres, kvantifiseres og verdsettes i kroner så langt det er mulig og gir meningsfull informasjon. I ettertid bør prosjektene evalueres med tanke på både måloppnåelse og lønnsomhet. Dette gjøres i dag av Concept-programmet. Samferdselsdepartementet krever også etterprøving av de prissatte virkningene av et utvalg store prosjekter fem år etter at de er ferdigstilte.

I de aller fleste tilfeller vil et tiltak som gir billigere eller bedre transport, også føre til at flere benytter seg av transporttilbudet, eller benytter det mer enn de gjør fra før. For jernbanetransport og annen rutegående transport vil det da være et marked med en tilbudsside (togoperatøren) og en etterspørselsside (kundene), som begge tilpasser seg slik at det blir en ny likevekt.

I praksis er det vanskelig å vite hvordan et fornyings- eller utbedringstiltak for en jernbanestrekning vil bli utnyttet på sikt. Hvilke tiltak som skjer på relaterte strekninger vil også ha betydning. Resultatet kan bli en ny ruteplan med en kombinasjon av kortere reisetid, flere avganger og økt pålitelighet. Hvilke antakelser en gjør om hva som ville blitt situasjonen med og uten det aktuelle tiltaket, kan ha stor betydning for hvor lønnsomt tiltaket framstår (Eliasson og Börjesson 2014).

I tillegg til at en er opptatt av lønnsomhet, så har jernbaneutbygginger effektmål som typisk er relatert til tid og volum. Tid er ofte relatert til reisetidsreduksjoner, men kan også være koblet til reduksjon av variasjon i reisetid, det vil si økt forutsigbarhet, fremkommelighet eller punktlighet. Volum innebærer eksempelvis antall reisende på toget eller godsvolum som transporteres. I tillegg kan utbyggingen ha mål relatert til sikkerhet, miljø og økonomisk utvikling.

Investeringer i et bedret jernbanetilbud får normalt positive konsekvenser for trafikantene på jernbanen. Kortere reisetid, økt punktlighet og et mer attraktivt rutetilbud representerer nytte for de som reiser eller sender varer med jernbanen. Disse kvalitetsfaktorene veier tungt i nyttekostnadsanalysene. De vanligste konsekvensene som påvirker trafikantnyttene er:

- Redusert reisetid
- Redusert ventetider
- Redusert gangtid
- Økt trafikkvolum
- Endrede priser
- Økt tilgjengelighet
- Økt punktlighet
- Redusert ulykkesfrekvens

Trafikantnyttene utgjør vanligvis størstedelen av nytten ved investering i ny infrastruktur. Som hovedregel skal det gjennomføres trafikkberegninger med en transportmodell ved investeringsprosjekter i jernbanenettet. De viktigste grunnlagsdata for nytte-kostnadsanalysene hentes som resultater fra transportmodeller (trafikkberegningsmodeller). Det betyr at kvaliteten på resultatene fra transportmodellen i stor grad er styrende for resultatene av nytte-kostnadsanalysene. Eksisterende nasjonale og regionale transportmodeller har begrensninger i forhold til jernbane som infrastruktur. Jernbaneverket hadde også egne modeller (Jernbaneverket 2015), nå videreført hos Jernbanedirektoratet (Jernbanedirektoratet 2017). Data for antall reisende og deres reisevaner inngår som inputdata i de ulike modellene og verktøyene, hvor kildene i dag stort sett er basert på manuelle tellinger og reisevaneundersøkelser. Dataene er lite tilgjengelige og de har flere svakheter. Vi vil i neste kapittel drøfte svakheter knyttet til bruk av disse tradisjonelle kildene og argumenter for bruk av nye kilder som mobiltelefondata.

2 Metoder for estimering av personreiser med jernbane

Flere studier har pekt på vanskelighetene med å skaffe gode data på antall reisende med tog. Det er minst to utfordringer med hensyn til denne typen data. For det første vurderes slike data som virksomhetskonfidensiell informasjon, spesielt i høye oppløsninger. For det andre, dataene som faktisk er tilgjengelig er av varierende kvalitet og dekning. De store togselskapene i Norge teller antall passasjerer manuelt og rapporterer antall reisende på et aggregert nivå (årlig per togprodukt og strekning). Dette finnes tilgjengelig på www.ssb.no (SSB, 2017) fordelt på strekninger. Jernbaneverket har gjentatt ganger etterspurt mer detaljerte data. Togselskapene ønsker derimot ikke å levere ut detaljerte data da de oppfatter det som en konkurransefordel ovenfor potensielle konkurrenter å ha detaljert oversikt over antall reisende på ulike togavganger.

2.1 Ulike metoder og datakilder

Det er flere måter å skaffe data om antall reisende på tog. Vi gir her en kort oversikt over ulike kilder og metoder som i større eller mindre grad benyttes for å anslå antall reisende på tog og deres reisevaner.

Tabell 1 lister opp en rekke identifiserte kilder for identifikasjon av antall reisende på tog og andre transportmodi. Det er en tradisjon for at jernbaneselskaper har benyttet reisevaneundersøkelser og manuelle tellinger for å samle inn data om togreiser og antall reisende. Jernbaneverket rapporterer årlig aggregerte data om antall reisende, samt passasjerkilometer, og antall solgte enkeltbilletter og månedsbilletter til den offisielle jernbanestatistikken (Jernbaneverket – nå Bane NOR). Kilden er manuelle tellinger på utvalgte stasjoner på hver jernbanelinje. Nyere studier indikerer likevel en dreining mot å benytte automatiserte (gjerne elektroniske) teknologier for telling og registrering.

Tabell 1 Kilder for identifikasjon av antall reisende på tog og andre transportmodi

Indikator på antall reisende	Kilde/metode	Del av reise	Dataeier	Styrker	Svakheter
Faktisk antall reisende på et gitt tidspunkt	Manuelle tellinger utføres av togpersonalet	Ett gitt tidspunkt i løpet av reisen.	Transportoperatør	Vel etablert	Krever tilgjengelig personell, varierende nøyaktighet
Informasjon fra reisevaneundersøkelser	Spørreskjema på et utvalg av antall reisende	Utformet for ulike deler	De som utfører undersøkelsen	Kan utformes til å dekke alle ulike deler og former av reiser	Krever bruk av manuelle ressurser. Utvalg, ikke kontinuerlig måling.
Dørpasseringer	Automatisk registrering av dørpasseringer	På- og avstigning	Transportoperatør	Nøyaktig	Kostnader ved innstallering. Bare en del av flåten har utstyret.
Vekt på kjøretøy	System for kontroll av rullende materiell	Antall personer til enhver tid	Transportoperatør	Ikke behov for ekstra utstyr. Identifiserer ikke individer	Vektmålingssystemer må være installert
Observerte passasjerer med videokamera	Videoovervåkingskamera ombord	Antall personer til enhver tid	Transportoperatør	Utstyret kan være der fra før, ikke behov for noe mer	Krevende å analysere. Utstyr må være installert.
Billettsalg	Salgsdata	Avreisested for enkeltreiser. På- og avstigning for periodebilletter	Transportoperatørens salgsavdeling el.	Ikke behov for ekstra utstyr	Forretningssensitive opplysninger. Business sensitive. Vanskelig å identifisere enkeltreiser for periodebilletter.
Billettbruk	Billettregistreringssystemer	Inn- og utgang på stasjoner	Transportoperatørens salgsavdeling el.	Identifiserer individuelle reiser, trolig både på- og avstigning	Krever et billettregistreringssystemer
Bruk av WiFi	Antall innlogginger på togets trådløse nettverk	Antall enheter innlogget til en hver tid.	Transportoperatør	Ikke behov for ekstra utstyr hvis WiFi er installert	Krever at WiFi installeres. Måler bare de som er logget inn.
Nærhetsregistrering	Bluetooth gjenkjenning	Antall enheter som passerer	De som plassere utstyr for registrering	Fleksibel lokalisering	Identifiserer bare enheter med Bluetooth
Bruk av (tog)reiseapp på smarttelefoner	Reiserute-gjenkjenning (GPS etc.)	Lokalisering av enhet	Transportoperatør	Ikke behov for ekstra utstyr	Måler bare de som bruker appen
Registrert bruk av mobiltelefon ved basestasjoner	Teleoperatør som eier nettverket	Enheter som registreres på de enkelte basestasjoner hos operatør. Bevegelser i nettverket.	Teleoperatør	Trenger ingen direkte tilgang til tog eller jernbaneinfrastruktur	Liten erfaring med bruk av denne type data. Personvern. Måler bare kunder av teleoperatører.
Antall betalings-transaksjoner i forbindelse med betaling av reiser	Betalingsstatistikk	Bevegelser mellom punkter med kommersiell aktivitet	Eier av betalingsstatistikken	Trenger ingen direkte tilgang til tog eller jernbaneinfrastruktur	Indirekte måling, fanger bare opp kunder som betaler med kort.

Av de ulike datakildene i tabell 1 kan alle unntatt reisevaneundersøkelsene omfatte datasett som kan karakteriseres som Big Data, da de har mulighet for å generere store datamengder med høy oppløsning. Samtidig er det hittil vært vanlig å bruke sammenstilte data, som da typisk ikke er å betrakte som Big Data.

Manuelle tellinger

Manuelle kilder som tellinger og spørreundersøkelser har lenge vært den dominerende metoden for å finne antall reisende (Lawrence, 2003). Vuchic (2005) nevner at brukerundersøkelser og manuelle passasjertellinger er godt etablert når det gjelder å få informasjon om antall passasjerer på jernbane. Ifølge Boyle (1998) brukte mer enn to tredjedeler av de studerte amerikanske transportetatene manuelle tellinger og bruk av papir og blyant for å innhente data om reisende. Manuelle tellinger er brukt på alle system-, rute-, tur- og segmentnivå, og hyppigheten av disse tellingene varierer på alle nivåer. Hvilken type analyse som skal utføres bestemmer nivået på hvilke data som samles inn. Praktiske rutiner og metoder for å gjennomføre manuelle tellinger, lagring og analyse er godt etablert (beskrevet i Vuchic, 2005). Denne typen data har også vært hovedkilde for de norske jernbanevirksomhetene, i tillegg til reisevaneundersøkelser. Informasjonen som er mest fordelaktig er den som er registrert med henvisning til bestemt tid og sted. Den største fordelen med denne metoden er at den er enkel, fleksibel, og krever ikke spesielle tekniske ferdigheter eller utstyr. Baksiden er at den krever bruk av personressurser som er kostbart, har stabilitetsproblemer, har risiko for feil både i innsamlings- og registrering og kan ha dårlig nøyaktighet grunnet skjev representativitet i utvalg og skaleringsutfordringer.

Reisevaneundersøkelser

Tradisjonelt er det flere metoder for å beregne antall reiser mellom en opprinnelse og et bestemmelsessted. Den vanligste er bruk av OD-matriser (Origin - Destination) som oppgir antall reiser mellom ulike geografiske soner som representerer avreisestedet (O) og reisemålet (D) på en rute (Frias-Martinez et al. 2012). Den mest brukte metoden for å innhente data til disse matrisene er reisevaneundersøkelser. Disse undersøkelsene gjennomføres gjerne årlig. Dataene er kun representative for et begrenset delmengde av hele populasjonen. Følgelig gir denne metoden lav frekvens og høye kostnader. Dataene måler det vi kaller "stated preferences", altså det folk oppgir som preferanser om reisevaner. Dette til forskjell fra de øvrige datakildene vi omtaler her, som er "revealed preference", som viser faktiske reisevalg.

Automatiserte passasjertellere (APC)

Automatiserte passasjertellere (APC) er enheter som automatisk registrerer passasjerenes ombordstigning eller avstigning på toget. Det finnes to hovedtyper av APC-teknologi. Den vanligste er direkte måling av passasjertellinger med infrarøde sensorer eller videobilder (i mindre grad), mens den andre er basert på en indirekte beregning ut fra forholdet mellom den totale vekten av passasjerer ombord og gjennomsnittlig passasjervekt. Barabino et al. (2014) lister opp en rekke utfordringer knyttet til APC, noe som også Oslo Sporveier har erfart ved innføring av automatisk billettkontroll. I 2013 startet man i Norge testing av APC-teknologi fra det tyske selskapet Dilax (Teknisk Ukeblad 2014). Norsk Regnesentral har utviklet en algoritme basert på APC-data, som bruker en statistisk modell for å beregne det totale antall passasjerer.

Vektbaserte estimater

Nielsen et al. (2013) presenterer en metode for å beregne passasjerantall i Københavns jernbanesystem basert på elektronisk veiing (Electronic Weighing (EWE)) Slike veiingssystemer er installert i mange moderne tog da de leverer data til bremsesystemene. Informasjonen kan brukes til å anslå antall passasjerer i togene, basert på at vekten av et tog er en funksjon av antall passasjerer i toget, til enhver tid. Vektestimatene kalibreres med automatiske passasjertellinger, som samlet sett gir et godt estimat på antallet reisende. Gitt at denne teknologien allerede er installert på togene, kan dette være en kostnadseffektiv måte å beregne antall reisende. Nielsen et al. (2013) viser at EWE-basert overvåking kan gi estimater med høyere nøyaktighet enn infrarød sensor teknologi (APC). Ifølge Nielsen et al. (2013) kartlegges distribusjonen av passasjerer på jernbanenettet i København med en kombinasjon av EWE og APC. De to systemene gir i kombinasjon både informasjon om det totale trafikkvolumet (EWE) og passasjerflyten (APC). De to systemene kan også brukes til å validere hverandre.

Billettssystemer

Billettsalg kan være en kilde til informasjon om antall reisende, spesielt for togprodukter der hver reise kjøpes enkeltvis. For månedskort og lignende billetter sier billettsalget mindre om faktisk reiseaktivitet.

En annen måte å telle antall reisende er gjennom registrering av billetter og reisekort elektronisk. Denne kategorien inkluderer både registreringsenheter på stasjonene, samt om bord på togene. Metoden er for det meste brukt til å samle data på aggregert nivå. Manuelle tellinger er vanligvis korrelert med denne typen data for valideringsformål. Boyle (1998) fremhever fordeler og ulemper med slike enheter. Fordelene inkluderer betydelig større utvalg, bedre pålitelighet og mulighet for detaljer i form av tidspunkt og sted for registrering. Ulempen er feil på utstyr og undervurdering av passasjerdata på grunn av sniking. Furmin (2010) bruker data fra automatiske billettregistreringssystemer ved inngang og utgang på stasjoner, for å utvikle en metode for å estimere reisemønstre på undergrunnene i London.

Sporingsteknologi

Det finnes flere måter å samle inn elektroniske spor på, gjerne generert av smarttelefoner. Dette inkluderer GPS, Bluetooth og WiFi-oppkoblinger. Ulike apper kan brukes til slik sporing. En metode for å beregne reisetid innebærer registrering av GPS-spor for å avdekke reisetider. Disse metodene er til en viss grad innovative og fanger i detalj individuelle reisevaner, men er begrenset av sine utvalgsstørrelser (for eksempel antall frivillige), og gir utfordringer med hensyn til skalering (Holleczek et al. 2014). Lav bruk av GPS på smarttelefoner på verdensbasis, og begrenset tilgang til GPS-relatert informasjon fra teleoperatørene på grunn av personvern hensyn, er også et hinder for at dette skal være en effektiv måte for å beregne reisetider fremover. Elektroniske spor kan også legges igjen fra betalinger, der man ser hvor og når betalingskort blir brukt som beskrevet ovenfor.

Mobiltelefondata

Mobiltelefondata inneholder registrerte "hendelser" på en gitt basestasjon på et gitt tidspunkt. En basestasjon er i praksis en mobilmast som ved bruk av mobilnettet registrerer og lagrer data om oppringning eller mottak av samtaler, sending eller mottak av tekstmeldinger, bruk av mobiltelefon ved nedlastinger mv., det som her omtales som hendelser. Dataene er basert på

utvalg, og omfatter kun antall enheter som er koblet til de forskjellige basestasjoner. Anonymiseringskrav knyttet til disse data er svært viktig, og håndteres gjerne gjennom aggregering. Disse tallene er en indikator på det antall personer som bruker mobilnettet i dekningsområdet til basestasjonen, og kan dermed gi en indikasjon på antall reisende. Vi vil i de to neste avsnittene beskrive mobiltelefondata som kilde nærmere.

Oppsummering av metoder og datakilder

Som en oppsummering kan de ulike teknologiene og metodene for å få informasjon om antall reisende på tog kategoriseres som følgende:

- Manuelle metoder, som tellinger og reisevaneundersøkelser (basert på intervjuer, spørreskjemaer etc.)
- Sensorer på toget, som dørpassering, vektmåling, TV-overvåkning og logger av WiFi-bruk
- Data fra billettsystemer, billettsalg eller billettvalidering
- Registeringer av hele eller større deler av reisen, som kan baseres på data om betalinger eller fra mobiltelefoner

De fleste tilnærminger er avhengig av tilgang til jernbanestasjonen eller rullende materiell utstyr. I tillegg finner vi at de fleste teknologier måler enten antall reisende kontinuerlig, eller ved innganger og utganger på togstasjoner. Cottrill og Derrible (2015) oppsummerer begrensninger med dagens metoder som sporadiske og ineffektive, for små utvalgsstørrelser i forhold til målet og et dårlig nivå av nøyaktighet. Som vi skal se kan målinger basert på mobiltelefon brukes uavhengig av trafikk operatører, og de har potensial til å måle komplette reiser, blant annet med bruk av flere transportmodi. Disse forholdene er to argumenter som taler for å undersøke bruk av mobiltelefon som kilde, selv om det allerede finnes flere teknologier og metoder for å måle antall reisende.

2.2 Flere studier viser potensial for bruk av mobiltelefondata

Flere, men ikke mange, studier er blitt gjort for å undersøke bruk av mobiltelefondata for å kartlegge transportmønster og reisevaner. Litteraturen viser at mobiltelefondata ser ut til å ha et potensial for å beskrive folks bevegelsesmønster, som et alternativ til de tradisjonelle transport- og reisevaneundersøkelser.

En av studiene er Calabrese et al. (2012). De brukte mobiltelefonregistreringer fra en million brukere i Boston i løpet av tre måneder, for å beskrive transportbehovet. Tilsvarende studie bør være mulig å gjøre før og etter ferdigstillelse av større investeringer i infrastruktur. Calabrese peker på tre utfordringer ved bruk av mobiltelefon: (1) økonomiske og demografiske opplysninger om enkeltpersoner er ikke tilgjengelig grunnet personvern hensyn (2) mobilbrukere er ikke nødvendigvis representative for hele befolkningen og (3) data er ikke formaterte for denne type analyser. For å adressere den første utfordringen brukte de aggregerte data der brukerne ble samlet i grupper som tilsvarer det mest detaljerte nivået som økonomiske og demografiske data var tilgjengelige. Til å kalibrere og kvalitetssikre dataene ble det brukt informasjon fra sikkerhetsinspeksjoner av kjøretøyene, som inkluderte km-stand.

Alexander et al. (2015) presenterer metoder for å beregne gjennomsnittlig daglige "til-fra" reiser fra triangulert mobiltelefon registreringer (Cellphone call Detail Records - CDR) av millioner av anonymiserte brukere. CDR inneholder tidsstemplede koordinater til de anonymiserte kundene. Registreringene blir aggregert til klynger av ulike typer lokasjoner og videre tolket til å defineres som hjemme, på jobb eller andre avhengig av observasjon frekvens, ukedag og tid på døgnet. Metoden bruker også data fra undersøkelser av reiser i store amerikanske byer og folkeregistre, for å skille ut gjennomsnittlige daglige turer. Anvendelsen av metoden er støttet av validering gjennom bruk av tidsmessige og romlige fordelinger av reiser rapportert i lokale og nasjonale undersøkelser. Alexander et al. (2015) sammenligner denne metoden med bruk av reisevaneundersøkelser. Det konkluderes med at fordelene med mobiltelefondata er at siden de blir automatisk registrert er de lettere (og mer hyppig) tilgjengelig, og mindre kostbare å samle inn. I motsetning til tradisjonelle reisevaneundersøkelser som spør respondenten om detaljer om reisevaner fra en, eller noen, av de siste dagene, fanger bruk av CDR variasjoner i daglige reisevaner over en lengre tidsperiode. Ulempene med metoden er mangelen på annen informasjon om den reisende, som alder, inntekt eller formålet med turen. I tillegg vil registreringene i CDR være noe unøyaktige med hensyn til lokasjon, da selve registreringene skjer i nærheten av basestasjonene, men man vet ikke nøyaktig hvor.

Holleczek et al. (2014) viser at urbane mobilitetsmønstre og valg av transportmodi kan avledes fra detaljerte registreringer av mobiltelefonbruk, kombinert med reisedata for kollektivreiser. Reisedataene består av registrerte reiser gjort av 4,4 millioner anonymiserte brukere av Singapores offentlige transportsystem. De reisende har benyttet digitale reisekort som skal registreres ved av- og påstigning på busser og tog. Dataene omfatter stasjon og tidspunkt for avreise og ankomst for hver reise.

Leo et al. (2016), Bianchi et al. (2015) og Zhao et al. (2016) studerte brukervaner og reisemønstre til såkalte "mobile brukere". Frias-Martinez et al. (2012) brukte CDR data for å fylle ut matriser for reisers opprinnelse og destinasjonssted (OD-matriser), og sammenlignet de to tilnærmingene. De konkluderte med at CDR kan brukes til å beregne pendlermatriser med stor grad av nøyaktighet. Järv et al. (2012) viste at mobiltelefondata også kan benyttes som et verktøy for sanntids trafikkovervåking. Kujala et al. (2016) ser eksplisitt på bruk av mobiltelefon til å estimere reisetid mellom større byer. Xu et al. (2016) brukte et stort mobiltelefondatasett til å estimere etterspørselen for sykkelturer i en by. Studien identifiserte reiseruter som potensielt kan benyttes med bruk av sykkel, og estimerte det potensielle antallet sykkelturer på forskjellige steder i byen og ulike tider av dagen.

2.3 Muligheter for å avdekke reisemønstre

Nye typer data som mobiltelefondata åpner mange muligheter for analyse av transporttiltak. Det mest spennende er muligheten til å kunne se på transportmønstre og ikke bare måle volumet av trafikk på disse punktene der en teller.

Begrepet reiser og reisemønstre i forbindelse med transportundersøkelser må tolkes bredere enn registrering av enkeltturer. Den komplette reisen fra opprinnelsessted til destinasjon er interessant, da den kan bestå av flere enkeltturer med bruk av ulike transportmodi. F.eks. vil en

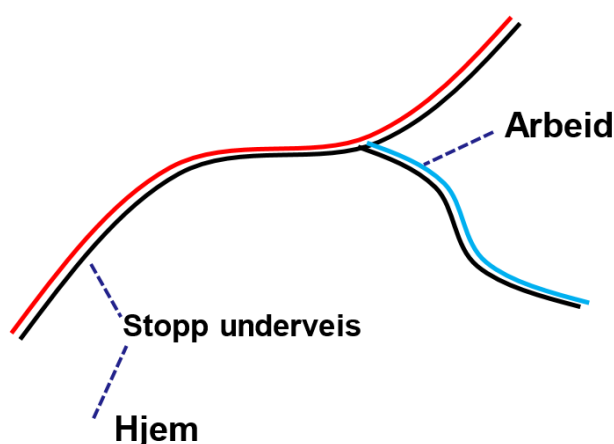
småbarnsforelders hverdag kunne bestå av følgende enkeltreiser (med ulike transportmodi) som til sammen utgjør et komplett reisemønster:

Hjem til jobb:

Hjem – Barnehage (bil), Barnehage – Pendlerparkering (bil), Pendlerparkering – Sentralstasjon (tog), Sentralstasjon – T-banestasjon (t-bane), T-banestasjon – jobb (gange)

Jobb til hjem:

Jobb - T-banestasjon (gange), T-banestasjon - Sentralstasjon (t-bane), Sentralstasjon - Pendlerparkering (tog), Pendlerparkering - Barnehage (bil), Barnehage – Butikk (bil), Butikk - Hjem (bil)



Figur 2 Illustrasjon av reiser som inkluderer jernbanen, men også annen transport. Rød og blå er ulike toglinjer.

Reisetiden og dens mulige optimalisering er med andre ord basert på moduser av reiser, hvis hver enkelt bruksmønster og tetthet må bli undersøkt. Det å skulle kartlegge og analysere dette reisemønstret, og f.eks. anslå samlet reisetid, er svært krevende med hensyn til tilgjengelige datakilder (for alle ulike transportmodi), representativitet og kostnader ved datainnhenting.

Den fremste grunnen til å vurdere reisemønster og enkeltreiser er å undersøke om det finnes flere alternativer tilgjengelig for reisende, spesielt når det gjelder fremkomstmiddel, og hvordan disse vil påvirke effekten av og måloppnåelsen for infrastrukturtiltaket hvis de utnyttes. Er det et potensiale for å korte ned reisetiden, unngå kødannelse eller redusere miljøbelastningen ved å endre reisemønstre? Analysene vil være viktige for beslutningstakere. Det å forstå hvordan ulike faktorer påvirker de komplette reisemønstrene er avgjørende for utviklingen av en robust politikk som vil oppfylle mål om fremtidige effektive og bærekraftige transportløsninger.

Det pågår et arbeid i Telenor for å gjøre avklaringer om hvordan mobiltelefondata for bevegelser fra en basestasjon til en annen kan lagres og gjøres tilgjengelige på en måte som er akseptable i forhold til personvern. Når det er avklart åpnes større muligheter til å kartlegge reisemønster, overganger mellom ulike transportformer og total reisetid fra hjem til jobb. Vi fikk ikke tilgang til slike data i denne studien. Dette vil derfor være en oppgave for etterfølgende studier basert på disse data.

3 Data vi har benyttet i analysen

Tre ulike datasett ble fremskaffet for å utføre analysene (de to siste primært for validering av mobiltelefondataene):

- 1) Mobiltelefondata i et bestemt format fra tilstøtende basestasjoner
- 2) Punktlighetsdata for utvalgte togstasjoner
- 3) Passasjertellinger fra togreiser

Både fra et etisk og juridisk synspunkt, er det viktig å beskytte personlig informasjon og respektere folks privatliv. Data som ikke inneholder personopplysninger er i utgangspunktet uproblematisk, både som individuelle datakilder og kombinasjonen av flere kilder. Kombinasjon av ulike datakilder, der det er personer som er bindeleddet mellom de ulike dataene er mer problematisk. Data fra ulike kilder med personopplysninger kan kombineres uten å avsløre personlig informasjon, men dette kan være utfordrende. Anonymitet i datasett oppnås vanligvis ved aggregering, hvor hver gruppe har så mange personer at enkeltpersoner ikke kan identifiseres. For å sikre at vi ikke bryter med personvern hensynene, er data som brukes i studien godkjent av Telenors personvernombud, og i henhold til veilederen til Norsk senter for forskningsdata (NSD).

3.1 Mobiltelefondata

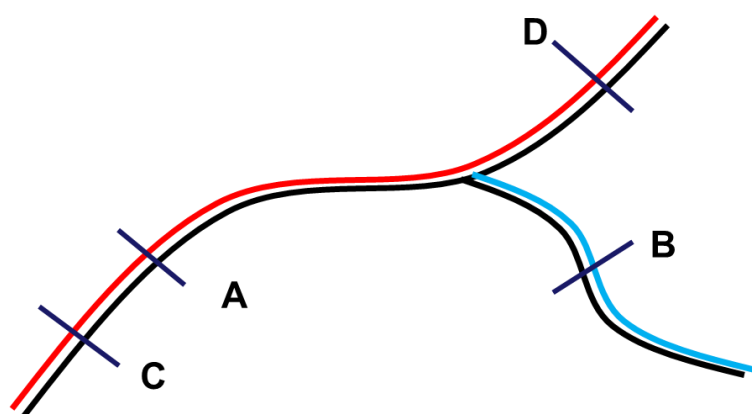
Det første datasettet vi har benyttet er data over registrerte hendelser (bruk av mobiltelefon) fra utvalgte basestasjoner. Dataene er gitt av Telenor, og er basert på en kilde med aggregerte data pr. basestasjon (totalt antall oppkoblinger mot stasjonen). De valgte basestasjonene er plassert slik at det er sannsynlig at en stor del av mobiltelefontrafikken er generert av togpassasjerer. Dette er rene tellinger av hendelser på et gitt tidspunkt. Dataene er basert på utvalg, og omfatter kun antall enheter som er koblet til de forskjellige basestasjoner. Dataene er anonyme. Disse tallene er en indikator på det antall personer som bruker mobilnettet i dekningsområdet til basestasjonen, og kan dermed gi en indikasjon på antall reisende.

Det første steget var å trekke ut basestasjonene som ligger i umiddelbar nærhet til jernbanen, basert på plasseringene til basestasjonene og beskrivelser av deknningen. En basestasjon er en radiosender som fungerer som bindeledd mellom mobiltelefoner og telefonnettet. Vi brukte et GIS-basert kartprogram (QGIS) til å importere koordinatene til basestasjonene, sammen med et kart over området og jernbanelinjene. Et kriterium var at basestasjonen ikke skulle ligge mer enn 2 km fra jernbanelinjen. Vi ekskluderte basestasjoner som lå mer enn 1 km fra endestasjonen i motsatt retning og basestasjoner med innendørs dekning. Med disse kriteriene endte vi opp med ca. 100 basestasjoner. Deretter valgte vi ut de basestasjonene som med størst sikkerhet mottar signaler fra togreisende; dvs. basestasjoner som mottar signaler fra mobiltelefoner på jernbanen, togstasjonene og i jernbanetunnelene. Vi snevret ytterligere ned ved å velge basestasjoner som (1) har plassering mellom togstasjoner, (2) ikke har plassering i nærheten av en hovedvei, og (3) har plassering i mer øde deler, utenfor de mer tettbygde områdene. Til slutt endte vi opp med et utvalg på 5 aktuelle basestasjoner.

Selve dataene fra basestasjonene er registrerte hendelser på en gitt basestasjon på et gitt tidspunkt. Med hendelser menes her hver gang en basestasjon har registrert et signal fra en mobiltelefon gjennom følgende aktiviteter:

- Motta/sendte en tekstmelding
- Motta/starte en telefonsamtale
- Starte en data sesjon på nettverket
- Registrering hos nettverket/IMSI tilkobling (Slå på telefonen)
- Avmelding med Network /koble fra IMSI (Slå av telefonen)
- Overlevering med basestasjonen

Antall hendelser vil være en indikasjon på antall personer som bruker mobilnettverket i dekningsområdet til basestasjonen. To forskjellige datasett med mobiltelefondata fra de utvalgte basestasjonene var tilgjengelig og ble brukt i denne analysen. For datasett 1 er tidsintervallet mellom hver måling på 5 minutter, fra en tidsperiode på åtte dager på våren 2016. For datasett 2 er tidsintervallet mellom hver måling på 1 minutt, fra en tidsperiode på tre dager på høsten 2016.



Figur 3 Tilnærming i denne studien. Tellinger ved utvalgte basestasjoner.

3.2 Punktlighetsdata

Norske jernbaneselskaper har i løpet av 2000-tallet bygd opp betydelig kapasitet for automatisk datafangst fra jernbanedrift. Det er plassert ut sensorer som fanger data automatisk både i infrastrukturen og om bord på togene. Nyere utvikling innen digitalisering og datamining åpner spennende muligheter for bruk av jernbanens store datamengder. Eksempler på data som registreres automatisk inkluderer tidspunkt for passering av hovedsignal for innkjøring og utkjøring på stasjoner, lokalisering (GPS), tidspunkt for åpning og lukking av dører, strømforbruk og passasjertellinger. I tillegg dokumenteres en rekke forhold manuelt. Forsinkelser som overstiger gitte grenser, tilegnes årsaker, feil i infrastruktur registreres i BaneData, og en rekke forhold med konsekvens for drift registreres i en hendelseslogg.

Jernbaneverket (nå Bane NOR) registrerer punktlighetsdata (for hvert tog hver dag), og består av registreringer fra ankomst og avgangssignaler. Dataene blir registrert i en database som vi har fått en kopi av. Dataene beskriver togenes bevegelser i nettverket, og inkluderer planlagte og faktiske ankomst- og avgangstider for hvert tog på hver stasjon, tognummer, driftsselskap, og klasseinformasjon (f.eks. frakt, kjører tom, eller passasjertog).

Den fysiske utformingen av det norske jernbanenettet er beskrevet i flere formater, mange av dem er tilgjengelig på internett. Vi har brukt denne informasjonen til å kombinere punktlighetsdataene med mobiltelefondata. Mer konkret har vi beregnet når tog passerte nær basestasjoner.

3.3 Passasjerdata (APC-data)

Antall reisende registreres tradisjonelt ved manuelle tellinger. På strekninger med høy andel enkeltbilletter kan billettstatistikk også benyttes. Billettstatistikk er vanskeligere å bruke på strekninger med stor andel periodebilletter (som månedskort), da det ikke er validering av billetten på stasjoner eller på toget i Norge. På enkelte av de nyere togene finnes det automatisk registrering av antallet på- og avstigende reisende. Systemet kalles automatisk passasjer telling (Automatic Passenger Count, APC), hvor det er sensorer ved dørene som registrerer antall på- og avstigende passasjerer på hver stasjon. Norsk Regnesentral har utviklet en matematisk metode, som beregner antall passasjerer basert på APC dataene med en statistisk modell. Vi har fått tilgang til slike data på den strekningen vi har studert i en av de to tidsperiodene som er aktuelle. Disse data er benyttet til å undersøke muligheter for å beregne et forhåndstall mellom observert antall passasjerer og antall hendelser på de tilstøtende basestasjonene.

4 Analyser og resultater

Utgangspunktet for denne studien var som nevnt tidligere å teste tilgjengelighet og bruk av mobiltelefondata til å kartlegge personreiser med jernbane.

Som et første steg ønsket vi gjennom en empirisk analyse å se på om hendelsesmønsteret (antall signaler registrert på basestasjonene) i mobiltelefondataene var sammenfallende med den faktiske togtrafikken på utvalgte strekninger. Var dette tilfelle ville det indikere at mobiltelefondata også kan være valide som kilde til å kunne si noe om passasjerene som finner seg om bord på disse togene. Studien ser nærmere bestemt på fem basestasjoner som har plassering i umiddelbar nærhet til jernbanelinja og i tilknytning til fem av togstasjonene på en jernbanelinje som går til en by. For å anonymisere dataene har vi angitt togstasjonene som U, V, W, X og Y, hvor stasjon U ligger lengst unna og stasjon Y ligger nærmest byen. Antall registrerte hendelser koblet med tidspunktet togene passerer basestasjonene ble analysert systematisk. En algoritme ble utviklet for en ensartet evaluering av de forskjellige trender, og dataene ble kategorisert i flere kategorier basert på tid og retningsinformasjon.

Ved å så kombinere disse togtrafikkanalysen, samt hendelsesdata, med faktiske passasjerdata (APC-data) for de samme strekningene på samme tidspunkt, ønsket vi å se på muligheten for å kunne predikere passasjerantallet ved hjelp av mobiltelefondata. Siden de mobiltelefondata vi hadde tilgjengelig ikke på noen som helst måte var representative for alle reisende (kun en leverandør - Telenor), så vi på om APC-data kunne fungere som en kilde til validering av forholdet mellom antall passasjerer og antall hendelser på de tilstøtende basestasjonene. Slike forholdstall kan være en nøkkel til å bruke mobiltelefondata som mål på antall reisende. Tanken er å finne nøkkeltall av typen «for hver registrert hendelse er det x antall reisende».

Neste skritt vil kunne være å skaffe data som viser flyt mellom base stasjoner, for eksempel fra A til B i Figur 3, men det foreligger ikke tilgjengelige mobiltelefondata for dette. Dette er interessant generelt, og spesielt hvis det er et overgangspunkt mellom A og B. Enda mer interessant er det å se på strømmer mellom "hjemmeområder" og sannsynlige "arbeidsområder". Dette er interessant i seg selv, men er ekstra interessant i et evalueringsperspektiv, hvis en kan sammenligne total reisetid før og etter en investering. Dessverre er ikke slike data tilgjengelig pr i dag.

I det følgende vil vi presentere resultater fra analysene, samt metodiske og tekniske utfordringer med en slik tilnærming.

4.1 Beregning av tidspunkt for når toget passerer basestasjon

Det ble avgjørende å finne ut tidspunktene for når togene passerte basestasjonene. Vi så på to forskjellige tilnærminger for å finne den tilnærmede tiden for når toget passerer basestasjonen.

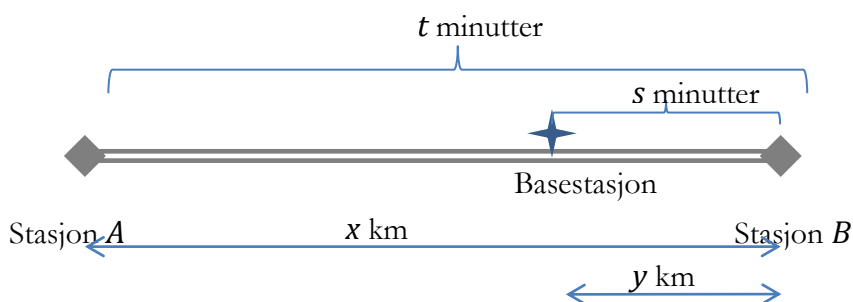
Den første metoden er basert på avstandene og tiden mellom togstasjonene, med antakelsen om konstant hastighet.

Med antakelsen at toget har en konstant hastighet fra stasjon A til stasjon B , så kan tiden s uttrykkes som $s = \frac{y}{x} t$ minutter, der x og y defineres som vist i figur 4.

Den andre metoden er basert på den tillatte hastigheten på jernbanelinja.

Basert på de tillatte hastighetene på de ulike delene av jernbanelinja, som sjelden er konstant mellom to stasjoner, så beregnet vi gjennomsnittlig hastighet v fra den nærmeste togstasjonen til basestasjonen. Reisetiden er da gitt av $s = \frac{y}{v} 60$ minutter.

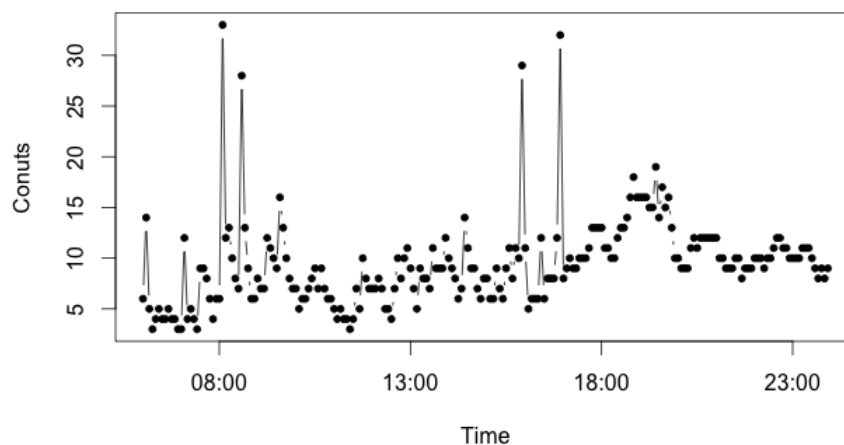
Metodene gir reisetiden s fra nærmeste togstasjon til basestasjonen. Basert på en av metodene kan vi beregne tidspunktet for når tog passerer basestasjonen, ved bruk av dataene for faktiske togbevegelser. Reisetiden t fra stasjon A til stasjon B (se Figur 4) kan enten være den ruteplanbestemte reisetiden, eller den faktiske reisetiden for et gitt tog på et gitt tidspunkt. Vi brukte faktisk reisetid når det var tilgjengelig i togtrafikk dataene. Hvis toget reiser fra stasjon A til stasjon B , så er tidspunktet for når toget passerer basestasjonen gitt av $t_{ankomst}(B) - s$. Mens dersom toget reiser fra stasjon B til stasjon A , er tidspunktet for når toget passerer basestasjonen gitt av $t_{avgang}(B) + s$.



Figur 4 Illustrasjon av plassering av en basestasjon mellom to togstasjoner, og notasjonen brukt i beregningen av reisetid fra stasjon B til basestasjonen.

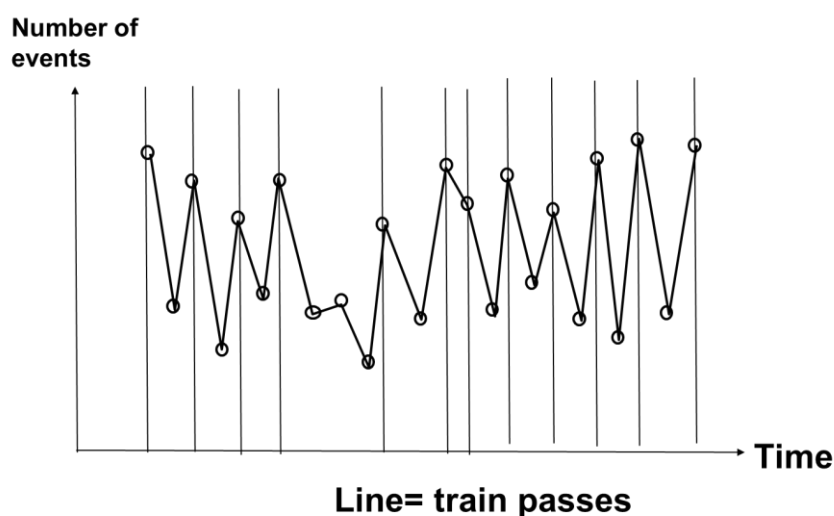
4.2 Kobling av tidspunkt for togpassering og registrerte hendelser i mobiltelefondata

Figur 5 viser tellinger av registrerte hendelser samlet på én dag på basestasjonen med plassering etter stasjon X. Tellingene har tydelige hopp på noen tidspunkt, som vi ønsker å sammenligne med tilnærmet tidspunktet for når togene passerer basestasjonen.

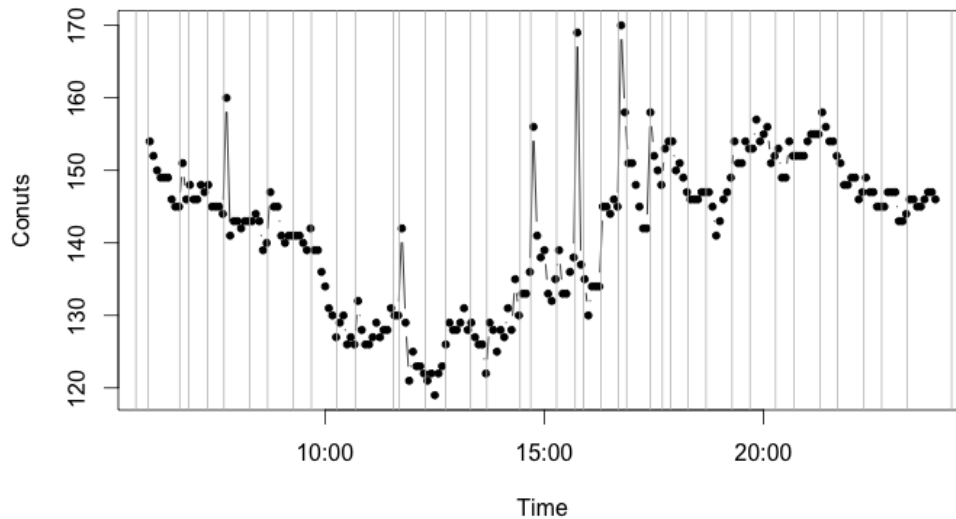


Figur 5 Registrerte hendelser på basestasjonen med plassering etter stasjon X.

Først ser vi på hopp i tellingene når togene passerer, uavhengig av hvilken retning togene kjøre. Basert på beregningene av når togene passerer basestasjonene og registrerte hendelser i mobiltelefondata, plottet vi telling med passerende tog, begge som en funksjon av tid. Det generelle formatet på grafen er illustrert i Figur 6, hvor de vertikale linjene representerer tidspunktene når togene passerer og linjegrafen viser hendelser på gitte tidspunkt i tid med fem minutters intervaller mellom måletidspunktene gjennom en dag. Det forventes at når togene passerer vil antall brukere koblet til den bestemte basestasjonen øke med tilsvarende antall reisende på toget.



Figur 6 Illustrasjon av forventet variasjon i registrerte hendelser knyttet til tog som passerer basestasjonen.



Figur 7 Registrerte hendelser på en basestasjon med plassering mellom stasjon U og stasjon V. Vertikale linker representerer tidspunkt hvor tog passerer basestasjonen.

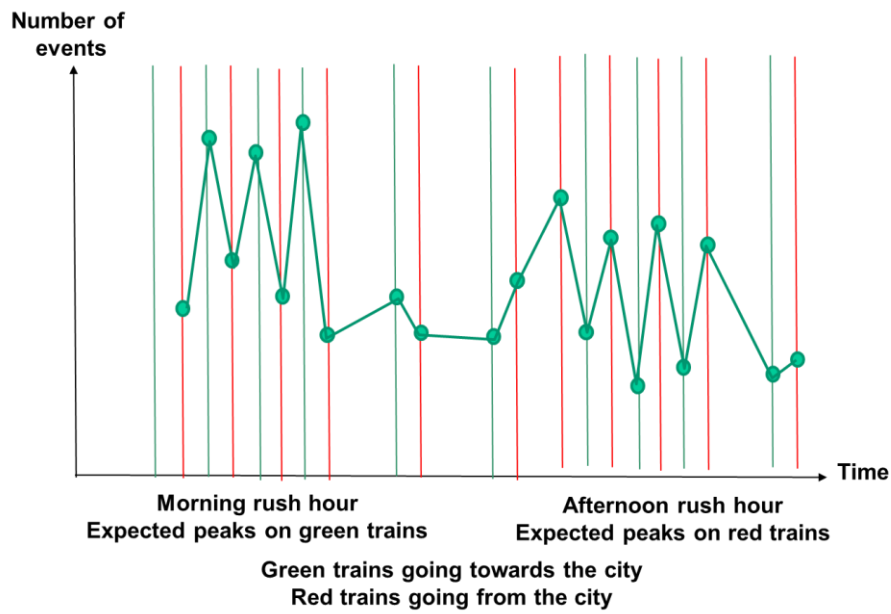
Figur 7 viser faktisk registrerte hendelser (fra mobiltelefonene) på basestasjonene plottet mot tid, og sammenlignet med beregnet tid for når tog passerer basestasjonen. Tidsperioden er én dag fra seks om morgenen til midnatt. Tellingene har noen tydelige hopp som ser ut til å sammenfalle med faktisk togtrafikk.

Vi kan så inkludere hvilken retning togene kjører i analysen. Basert på tognumrene kan togene deles inn i kategoriene:

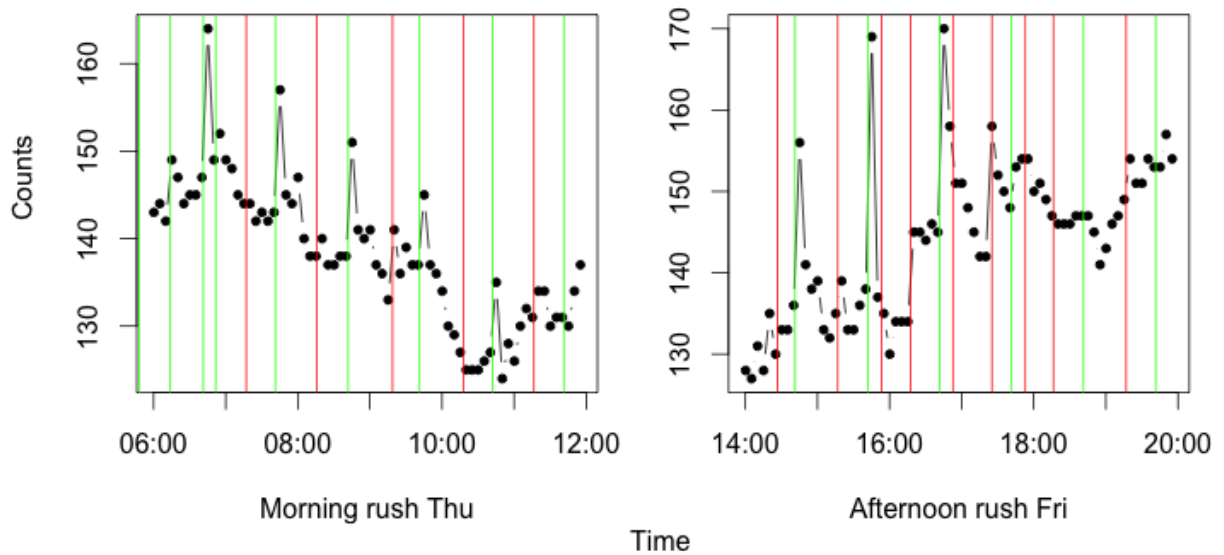
- Toget går inn mot byen
- Toget går ut av byen

I morgenrushet forventer vi at togene som går inn til en bykjerne vil ha flere passasjerer enn togene som går i motsatt retning, spesielt i typiske forsteder og landlige områder. Vi forventet derfor toppen i diagrammene når togene som går inn til byen passerer en basestasjon, som illustrert i Figur 8. I ettermiddagsrushet forventet vi det motsatte. Vi forventet også mindre variasjon i periodene hvor ingen tog passerer, sammenlignet med når tog passerer basestasjonen.

Figur 9 viser eksempel på morgenrush på en torsdag og ettermiddagsrush på en fredag, på basestasjonen mellom togstasjon U og togstasjon V. I dette eksempelet samsvarer toppene hovedsakelig med tog som er på vei inn til byen, uavhengig av tiden på dagen.

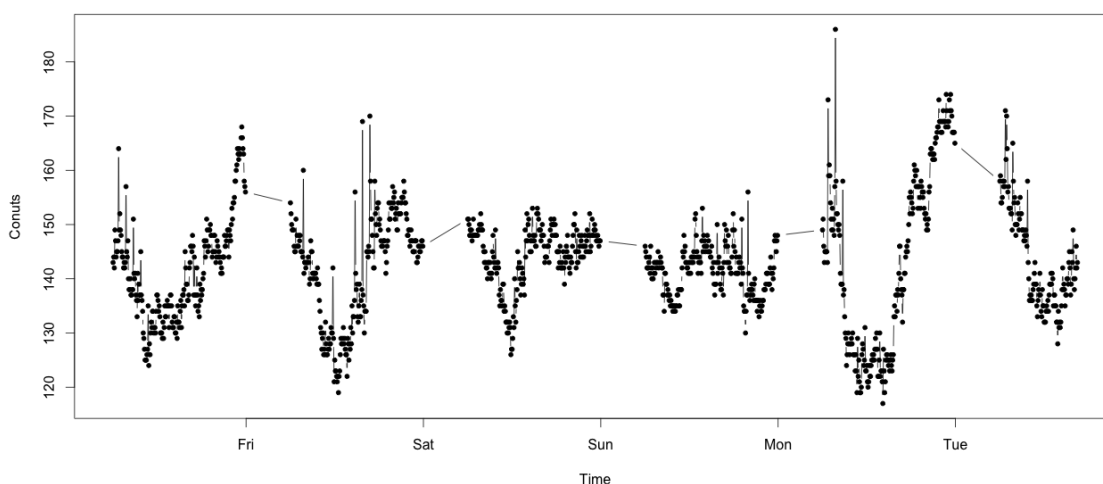


Figur 8 Forventet mønster på antall hendelser og tider hvor togenes morgenrush og ettermiddagsrush.



Figur 9 Antall hendelser i morgenrush og ettermiddagsrush for basestasjonen mellom stasjon U og stasjon V, sammen med angivelse av togpasseringer. Togpasseringene har samme fargekoding som i figur 8.

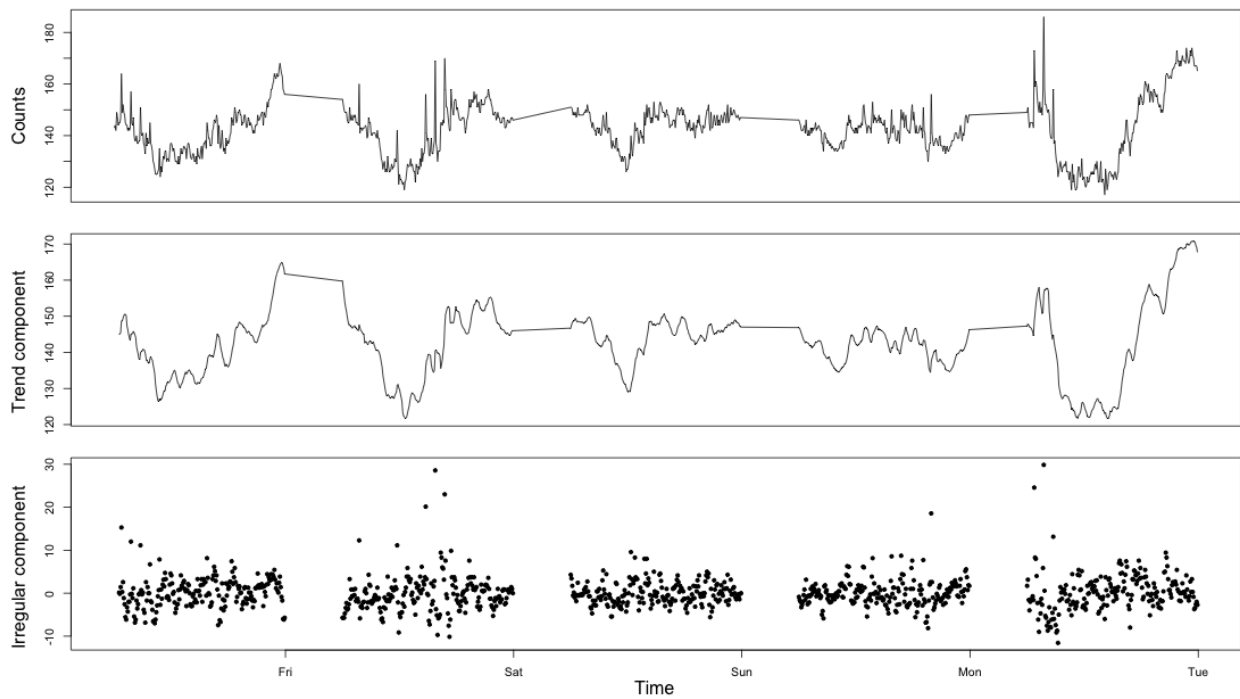
Som Figur 10 viser så svinger tellingene gjennom en dag, og med mindre svingninger i helgene. Vi ønsker derfor å finne størrelsen på toppene, uavhengig av den daglige variasjonen. Vi har testet to ulike fremgangsmåter. Den første måten er å bruke metoden “simple moving average” (SMA) i R. Metoden vil ikke gi en eksakt verdi av toppene, så en mer nøyaktig metode er å foretrekke for fremtidige analyser, men for denne forstudien viser vi resultatet fra analysen med SMA. Figur 11 viser tellinger, trend linjen og de irregulære verdiene, hvor de irregulære verdiene enkelt er tellingene minus trendverdiene. De irregulære verdiene viser topper med verdier mellom 5 og 30, som antagelig er når togene passerer.



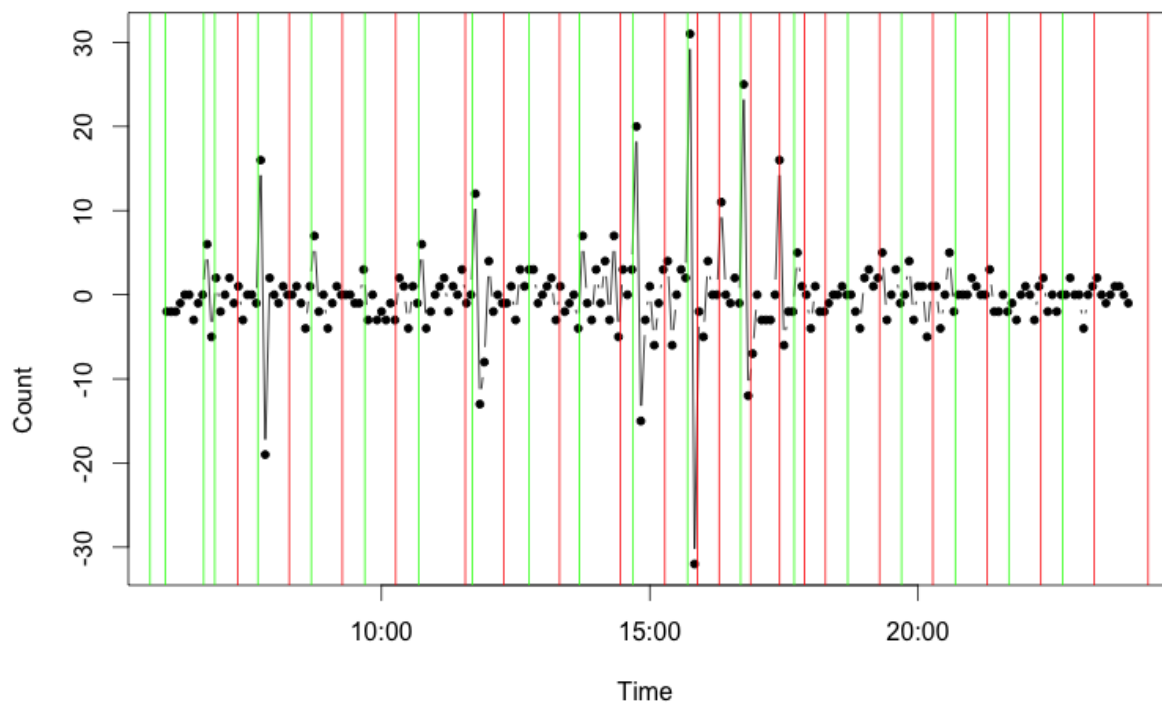
Figur 10 Registrerte hendelser i mobiltelefondata med plassering mellom stasjon U og stasjon V, for en seks-dagers periode i 2016, med fem minutters intervall mellom målingene.

Den andre måten er å beregne forskjellen mellom nåværende og forrige verdi, det vil si $y_i - y_{i-1}$. Dette er vist i Figur 12 for én dag fra samme tidsperiode som i Figur 10 og Figur 11.

Tidspunktene for toppene er sammenlignet med tidene for når togene passerer basestasjonen. Toppene ser ut til å korrelere godt med de forventede tidene for togenes passeringer, men en god del av togtidene ser ut til å ikke føre til noe reaksjon i tellingene. Noen av toppene har måletidspunkt mellom to tider for passeringer av tog. I disse tilfellene er det mest sannsynlig det foregående toget som er i relasjon med toppen, fordi mobilene er registrert på basestasjonen fram til de blir koblet til en ny basestasjon, så mobilen er sannsynligvis registrert på den valgte basestasjonene en liten tid etter at toget har passert.



Figur 11 Registrerte hendelser, trendkomponent og irregulær komponent for basestasjonen mellom stasjon U og stasjon V, for en seksdagers periode i 2016, men fem minutters intervall mellom målingene. Trendkomponenten er beregnet med "simple moving average" (SMA) i R.



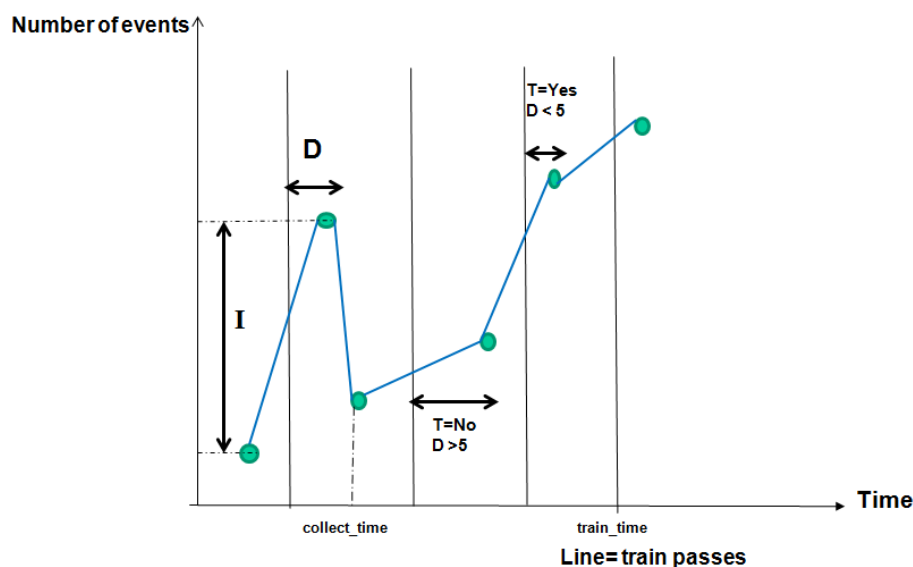
Figur 12 Forandring i antall hendelser hvor y-aksen viser $y_i - y_{i-1}$, for basestasjonen mellom stasjon U og stasjon V, for dagen i samme datasett som Figur 8.

4.3 Algoritme for å trekke ut måletidspunkt som samsvarer med tidspunkt for faktiske togpasseringer

Algoritmen sammenligner måletidspunktene med beregnede tidspunkt for når togene passerer basestasjonen. Input verdiene i algoritmen er antall hendelser, måletidspunkt, og tidspunkt for tog-passering. Algoritmen beregner tre verdier,

- Differansen D mellom måletidspunkt og tidspunkt for tog-passering,
- Avgjør, ved et gitt måletidspunkt, hvorvidt et tog har passert siden det forrige måletidspunktet. Verdien er betegnet T , og én av følgende tekststrenger; 'YES' hvis toget har passert innenfor de siste 5 minuttene, 'NO' hvis ingen tog passerte innenfor de siste 5 minuttene.
- Prosentvis variasjon fra den forrige tellingen på et gitt måletidspunkt, betegnet I .

Algoritmen ble implementert i Matlab. Output variablene fra algoritmen er illustrert i Figur 13.



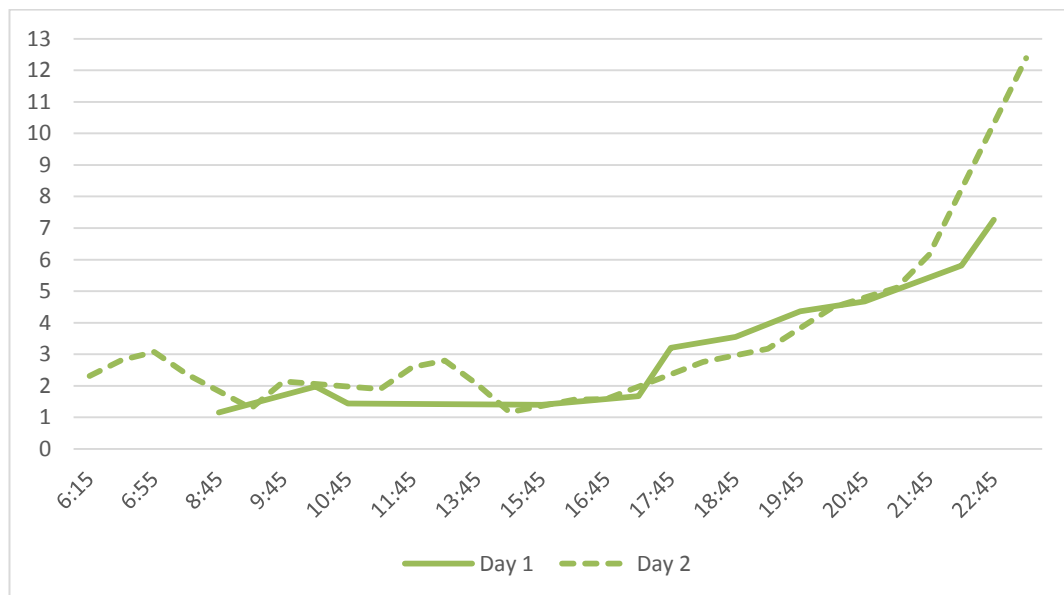
Figur 13 Illustrasjon av output variablene fra algoritmen beskrevet over.

Vi finner at antall enheter som er koblet til en basestasjon tilsvarer relativt godt med den planlagte passering av tog i nærheten av basestasjonene.

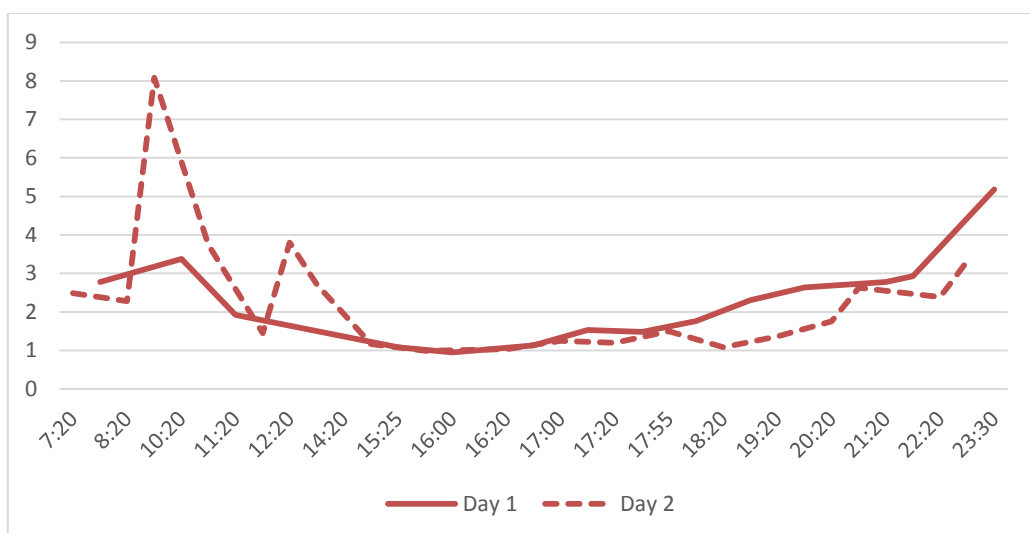
4.4 Sammenligning med APC-data

Ved å kombinere tider for faktiske togpasseringer, antall registrerte hendelser på basestasjonene og faktiske passasjerdata (APC-data) for de samme strekningene på samme tidspunkt, ønsket vi å se på muligheten for å kunne predikere passasjerantallet ved hjelp av mobiltelefondata. Siden de mobiltelfondata vi hadde tilgjengelig ikke representerer alle reisende (kun én leverandør - Telenor), undersøkte vi på om APC-data kunne fungere som en kilde til validering av forholdet mellom antall passasjerer og antall hendelser på de tilstøtende basestasjonene.

Vi har fått tilgang til faktisk antall reisende (APC-data) på den aktuelle strekningen i den tidsperioden som vi har 5-minutters hendelsesregistreringer for. Vi har beregnet forholdet mellom antall hendelser og antall reisende. Slike forholdstall kan være en nøkkel til å bruke mobiltelefondata som mål på antall reisende. Tanken er å finne nøkkeltall av typen "for hver registrert hendelse i en basestasjon er det x antall reisende". Figur 14 og Figur 15 viser indeksert forhold mellom antall registrerte hendelser og antall reisende fra APC systemet, for henholdsvis tog som går mot byen og tog som går ut av byen. Figurene viser at forholdet varierer over døgnet, men at utviklingen for de to påfølgende dagene er relativt like.



Figur 14 Indeksert forhold mellom antall registrerte hendelser og antall reisende beregnet fra APC-data, for tog som går mot byen forbi basestasjonen før togstasjon W på to påfølgende dager.



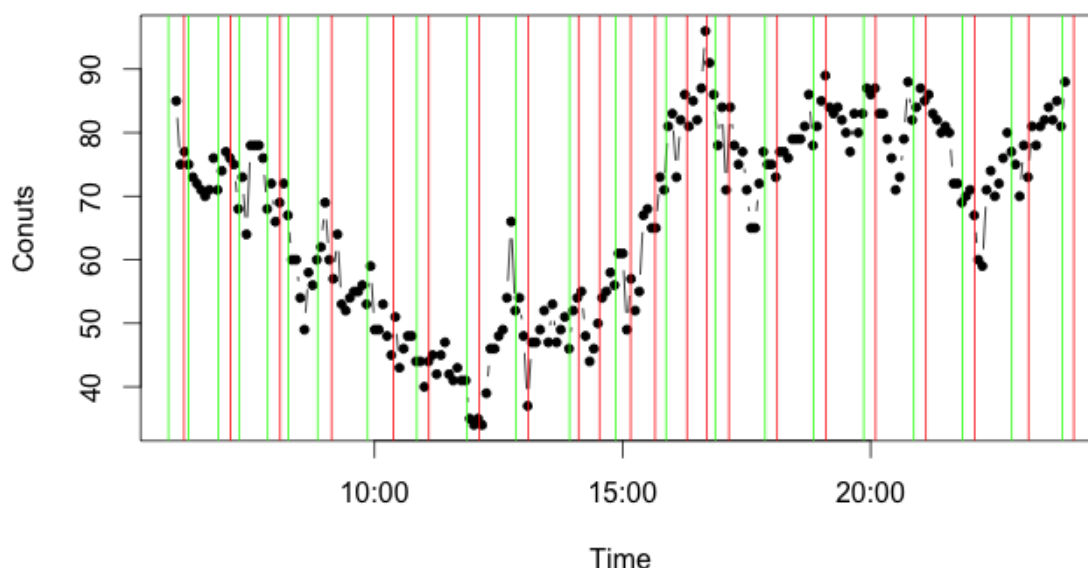
Figur 15 Indeksert forhold mellom antall registrerte hendelser og antall reisende beregnet fra APC-data, for tog som går ut fra byen forbi basestasjonen før togstasjon W på to påfølgende dager.

Forholdstallet er den enkleste måten å konvertere registreringer av hendelser til antall reisende på toget. Det kan være mulig å beregne forholdstall som kan gi indikasjon på antall reisende, men det krever sannsynligvis data for lengre tidsperioder, og forholdstall som er tilpasset hver delstrekning, basestasjon, ukedag og tid på døgnet. Dette er en konkret måte å videreføre og bruke resultatene i denne studie.

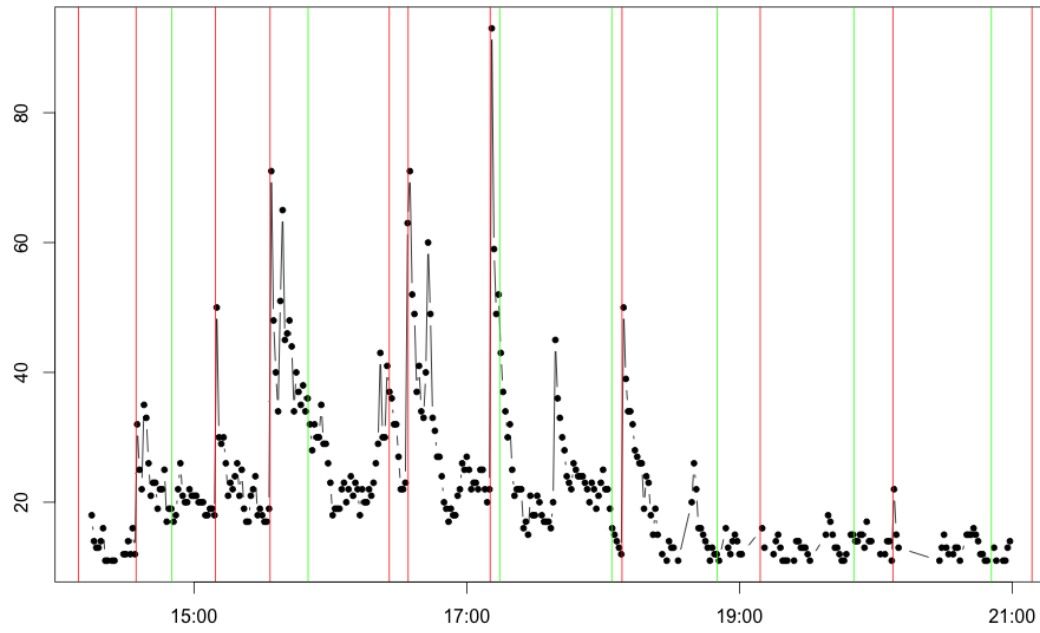
Vi har basert oss på målinger av antall hendelser. Antall hendelser registrert på en basestasjon er både en funksjon av antallet mobilenheter som er oppkoblet mot basestasjonen, og av aktivitetsnivået for de aktuelle mobilenhetene. Registrering av antall hendelser kan derfor variere med mobiladferden til de reisende på toget. Mobiladferden til folk varierer over døgnet og i løpet av uken, noe som, man bør korrigere for ved bruk av den type hendelsesdata som vi har hatt tilgang til.

4.5 Tidsintervall mellom registrerte hendelser – 5 min. vs. 1 min.

Vi fikk også tilgang til et datasett med tidsintervall på 1 minutt mellom målingene. Figur 16 viser tellinger på basestasjonen før togstasjon Y fra datasettet med 5 minutter tidsintervall, mens Figur 17 viser tellinger på samme basestasjon med 1 minutt tidsintervall. Datasettene er ikke fra samme tidsperiode, så en direkte sammenligning er ikke mulig, men figurene viser likevel en tydelig forskjell mellom hvorvidt målingene blir gjort hvert minutt eller hvert femte minutt. I Figur 16, med tidsintervall på 5 minutter, er det få topper og lavt samsvar med når togene passerer. I Figur 17 derimot, med tidsintervall på 1 minutt, er det tydelige topper som samsvarer godt med togpasseringer. Vi merker oss også at for denne basestasjonen, som ligger nærmere byen, faller toppene sammen med passering av tog som reiser ut fra byen i ettermiddagsrushet. Det samsvarer med forventningene vi presenterte tidligere i avsnitt 4.2.



Figur 16 Registrerte hendelser på basestasjonen før togstasjon Y, fra datasettet med tidsintervall på 5 minutter mellom målingene.



Figur 17 Registrerte hendelser på basestasjonen før togstasjon Y, fra datasettet med tidsintervall på 1 minutt mellom målingene.

5 Konklusjon og videre arbeid

Vi har undersøkt muligheten for å bruke mobiltelefondata for å beskrive antall reisende på tog. Studien viser at det er mulig å kombinere mobiltelefondata med data fra jernbanen som viser bruk av infrastrukturen, og planlagt og faktisk togtrafikk. Forutsatt at man velger basestasjoner med omhu og har egnet frekvens (dvs. hyppig nok) for målinger på basestasjoner, så er det god sammenheng mellom tidspunkt for passeringer av tog og endringer i antall registreringer på basestasjonene. Det ser også ut til å være mulig å beregne forholdstall som kan gi indikasjon på antall reisende. Men for å gjøre dette bør man sannsynligvis ha data for lengre tidsperioder. I tillegg bør forholdstallene tilpasses hver delstrekning og basestasjon, i tillegg til å tilpasses etter ukedag og tid på døgnet. Dette innebærer at grunnlaget er lagt for å i senere runder bruke mobiltelefondata til å estimere antall reisende, og dernest til å følge passasjerflyten i og utenfor jernbanen.

I forskningssammenheng er den kombinerte analysen av infrastrukturmodell, trafikkdata, mobiltelefonidata og rutetabell interessant. De fleste analyser på området begrenser seg til å gjøre én-til-én koblinger av slike datasett. Ikke mange forsøker å koble flere, noe som er utført i denne studien.

Et spørsmål var i hvilken grad formatet på tilgjengelige mobiltelefondata er egnet for måling av antall mobile enheter som passerer nær jernbanelinjen. Vi har brukt registrerte hendelser med henholdsvis 5 og 1 minuts oppløsning. Spesielt der det er tett togtrafikk er det en stor fordel å ha sampling hvert minutt. Våre testdata var komplette for 5-minutters-sampling, men ikke for 1-minutt-samplingene. Dette kan være tilfældighet, eller det kan indikere at sampling hvert minutt er en påkjenning på systemene, og kan medføre uregelmessig registrering. For evalueringsformål, er det et behov for data som dekker en forholdsvis lang tidsperiode, typisk noen få år. Vi anbefaler at data er lagret med høyest mulig oppløsning og at det er tydelig beskrevet hvordan data er samlet inn og behandlet. I tillegg bør det gjøres tiltak for å lagre data over lengre tidsperioder.

Ideelt sett skulle vi ville måle antall reisende før og etter større endringer, som ny infrastruktur, ny ruteplan etc. Dette vil være relevant i evalueringssammenheng, men krever sammenliknbare data over tid. Gode case hadde vært å studert situasjonen før og etter ruteendringene i desember 2012, 2013 og 2014 på Østlandet, som samlet sett tar ut effekten av flere større investeringer i infrastruktur og togmateriell. Vurderingen er at dette pr i dag er et for langt tidsperspektiv både når det gjelder bruk av punktlighetsdata og mobiltelefondata, men vi skulle gjerne få undersøkt hva som skal til for å kunne gjøre noe lignende neste gang det kommer en større ruteendring, som f.eks. når Follo-tunellen er ferdig.

Vi har også vurdert mulighetene for på sikt å bruke mobiltelefondata for å beskrive helhetlige reisemønstre som inkluderer flere reisemodi. Et spørsmål vi har berørt er; hva er betingelsene for utvidet bruk av mobiltelefondata for å spore reiser som inkluderer bytte av tog, og gjerne deler av reisen som ikke inkluderer bare tog? Det pågår et arbeid i Telenor for å gjøre avklaringer om hvordan mobiltelefondata for bevegelser fra en basestasjon til en annen kan lagres og gjøres tilgjengelige for evaluatorene og forskere på en måte som er akseptable i forhold til personvern. Når det er avklart, åpnes større muligheter til å kartlegge reisemønstre, overganger mellom ulike

transportformer og total reisetid fra hjem til jobb. Dette har ingen gjort i Norge enda. Vi fikk ikke tilgang til slike data i denne studien. Dette vil derfor være en oppgave for etterfølgende studier.

De data og analysene som er beskrevet over er likevel i seg selv verdifulle, og har ikke hittil vært tilgjengelige for forskere. Et prosjekt med analyser av de data som i dag foreligger vil i tillegg være et viktig utgangspunkt for videre studier. Dette er en type data som også internasjonalt er vanskelig tilgjengelig, grunnet samme forretningshensyn som hittil har hindret tilgang på norske data.

Referanser

Alexander, L., Jiang, S., Murga, M. og González, MC. (2015) Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, Part B, s. 240-250.

Arikawa, M., Konomi, S. og Ohnishi, K. (2007) Navitime: Supporting pedestrian navigation in the real world, *IEEE Pervasive Computing*, 6, s. 21–29

Barabino, B., Di Francesco, M., og Mozzoni, S. (2014) An Offline Framework for Handling Automatic Passenger Counting Raw Data. *IEEE Transactions on Intelligent Transportation Systems*, 15(6), s.2443-2456.

Bianchi, F. M., Rizzi, A., Sadeghian, A., og Moiso, C. (2016) Identifying user habits through data mining on call data records. *Engineering Applications of Artificial Intelligence*, 54, s.49-61.

Boyle, D. B. (1998) *Passenger counting technologies and procedures. TCRP Synthesis of Transit Practice 29*. Washington, DC: Transportation Research Board.

Bull-Berg, H., Olsson, N. (2014) *Nye datakilder i evaluering av store statlige investeringer – et potensial for Big Data?* SINTEF Teknologi og samfunn.

http://www.ntnu.no/documents/1261860271/1262021752/054_Rapportutkast_Endelig%20versjon.pdf lastet ned 1.3.2017

Calabrese, F., Diao, M., Lorenzo, G., Ferreira, J., og Ratti, C. (2012) Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. *Transportation Research Part C*

Cottrill, C. D. og Derrible, S. (2015) Leveraging Big Data for the Development of Transport Sustainability Indicators. *Journal of Urban Technology*, 22(1), s.45-64.
doi:10.1080/10630732.2014.942094

Eliasson, J., og Börjesson, M. (2014) On timetable assumptions in railway investment appraisal. *Transport Policy*, 36, s.118-126.

Ferris, B., K. Watkins, og Borning, A. (2010) Location-aware tools for improving public transit usability, *IEEE Pervasive Computing*, 9(1), s. 13 - 19

Frias-Martinez, V., Soguero, C., og Frias-Martinez, E. (2012) Estimation of urban commuting patterns using cellphone network data. *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, s.9-16.

Frumin, M. S. (2010) *Automatic Data for Applied Railway Management: Passenger Demand, Service Quality Measurement, and Tactical Planning on the London Overground Network*. Department of Civil and Environmental Engineering and the Operations Research Center. Massachusetts Institute of Technology

Holleczek, T, Yu, L., Lee, JK, Senn, O., Ratti, C. og Jaillet, P, (2014) Detecting weak public transport connections from cellphone and public transport data. *Proceedings of the 2014 International Conference on Big Data Science and Computing*, 2014. ACM, 9.

Jernbanedirektoratet (2017) *Analys og metodeutvikling*

<https://www.jernbanedirektoratet.no/no/strategier-og-utredninger/analyse-og-metodeutvikling/> lastet ned 21.3.2017

Järv, O., Ahas, R., Saluveer, E., Derudder, B., og Witlox, F. (2012) Mobile phones in a traffic flow: a geographical perspective to evening rush hour traffic analysis using call detail records. *PLoS One*, 7(11).

Jernbaneverket (2015) *Metodehåndbok. Samfunnsøkonomiske analyser for jernbanen.*

<http://www.banenor.no/contentassets/a3e2c5d145d04eaa95c04e6ebe036e44/metodehandbok-2015.pdf> lastet ned 1.3.2017

Kujala, R., Aledavood, T., og Saramäki, J. (2016) Estimation and monitoring of city-to-city travel times using call detail records. *EPJ Data Science*, 5(1).

Lawrence, G. (2003) Reuter, *A worldwide survey of transportation agency practices: passenger counting and service monitoring*. New York City Transit.

Leo, Y., Busson, A., Sarraute, C., og Fleury, E. (2016). Call detail records to characterize usages and mobility events of phone users. *Computer Communications*, 95, s. 43-53.

Lee, J. Kang (2015) Geospatial Big Data: Challenges and Opportunities. *Big Data Research.*, 2, s.74-81.

MIT (2013) *Senseable city lab*. <http://senseable.mit.edu/> lastet ned 30.7.2013

Nielsen, BF., Frølich, L., Nielsen, O.A. og Filges, D. (2014) Estimating passenger numbers in trains using existing weighing capabilities. *Transportmetrica A: Transport Science*, 10, 502-517.

Nilsson, J.E., Nyström, J. og Pyddoke, R., (2012) *Före och efter - uppföljning av en jærnvägsinvestering i Oslo*, VII noteat 38-2012.

Oslo Economics (2015) *Evaluering av dobbeltspor Sandnes – Stavanger*. Oslo Economics Rapportnr. 2015-22.

<http://www.ntnu.no/documents/1261860271/1262012574/Evaluering+dobbeltspor+Sandnes+Stavanger+juni+2015+-+ENDELIG+RAPPORT.pdf/e850a03b-ff47-401f-a70a-28cb89be4a82> lastet ned 1.3.2017

SSB (2017) *Persontransport med jernbane, etter strekningstype*. <https://www.ssb.no/transport-og-reiseliv/statistikker/jernbane/aar/2016-06-13?fane=tabell&sort=nummer&tabell=269018> lastet ned 1.3.2017

Teknisk Ukeblad (2014) *Matematiske verktøy skal gi kortere stopp på togstasjonene*, January 29 2014, <http://www.tu.no/artikler/matematisk-verktoy-skal-gi-kortere-stopp-pa-togstasjonene/224974> lastet ned 1.3.2017

Volden, G.H., Samset, K. (2013) *Etterevaluering av statlige investeringsprosjekter. Konklusjoner, erfaringer og råd basert på pilotevaluering av fire prosjekter*. Concept Report No. 30 Norwegian University of Science and Technology University.

Vuchic, R.V. (2005) *Urban transit. Operations, planning and economics*. John Wiley and Sons, Hoboken, NJ.

Xu, Y., Shaw, S-L., Fang, Z. og Yin, L. (2016) Estimating Potential Demand of Bicycle Trips from Mobile Phone Data – An Anchor-Point Based Approach. *ISPRS International Journal of Geo-Information*5(8), s.131;