

GPS Sporing som grunnlag for ny fartsmodell for syklist

En vurdering av offentlig tilgjengelig
GPS-data som grunnlag til sykkelmodellering

Christoffer Viena Jørgensen

Bygg- og miljøteknikk

Innlevert: juni 2014

Hovedveileder: Trude Tørset, BAT

Norges teknisk-naturvitenskapelige universitet
Institutt for bygg, anlegg og transport



Oppgavens tittel:	Dato: 10.06.2014		
GPS Sporing som grunnlag for ny fartsmodell for syklist	Antall sider (inkl. bilag): 140		
	Masteroppgave	X	Prosjektoppgave
Navn:	Christoffer Viena Jørgensen		
Faglærer/veileder:	Trude Tørset		
Eventuelle eksterne faglige kontakter/veiledere:	Tomas Levin		

Ekstrakt:

I denne oppgaven skal en prøve å vurdere potensialet til GPS-data som er samlet inn av privatpersoner ved hjelp av treningsloggere og mobilapplikasjoner, og deretter delt på internett. Dette gjøres ved å utvikle en fartsmodell basert på tilgjengelig data, samt diskutere andre potensielle bruksområder for dataene. Litteraturstudiet viser at mange av dagens transportmodeller har enkle tilnærminger til rutevalg og hastigheter for syklist. Det samme gjelder diverse ruteplanleggere som er tilgjengelig på nettet. Det kommer også frem at avstand og reisetid er viktige forklaringsfaktorer for syklisters rutevalg, men andre variabler som stigning, tilgjengelig infrastruktur, antall kryss og kontinuitet i sykkelnettverket er viktige. Datagrunnlaget består av GPS-data som er samlet inn fra nettstedet *Garmin Connect*. Turer ble hentet fra Oslo, Kristiansand og Trondheim, og de var gjennomført i perioden 10. april 2013 til 10. april 2014. Hovedvekten av disse turene var fra Oslo. Modellene som ble utviklet viste middels forklaringsgrad og datasettet viste tegn til dårlig homoskedastisitet og autokollinearitet, men det var til en viss grad forventet da det antas at det er flere viktige variabler som ikke er tilgjengelig i datasettet. Offentlig tilgjengelig GPS-data viser seg å ha mange potensielle bruksområder, men det bør først gjennomføres en kontrollert studie, som kan forbedre filtreringsprosessen og kartlegge personlige og fysiske forklaringsfaktorer som ikke er tilgjengelig i datasettet. Spesielt bør egenskaper som kjønn, fysisk form, erfaring, infrastruktur og trafikkmengde undersøkes. Resultatet bør brukes til å oppdatere fartsmodellen samt utvikle en rutevalgmodell for å tette hull i dagens transportmodeller. GPS-dataene har også andre bruksområder som kan være interessante. Identifisering av konfliktområder og flaskehalsen samt før/etter-analyser er noen bruksområder hvor offentlig tilgjengelig GPS-data kan være en meget enkel, billig og relevant datakilde.

Stikkord:

1. Sykkel
2. GPS
3. Fart
4. Rutevalg

(sign.)

Forord

Denne rapporten er dokumentasjon på mitt siste arbeid som student ved institutt for veg, transport og geomatikk på Norges teknisk-naturvitenskapelige universitet. Arbeidet ble utført våren 2014 og markerer avslutningen på min mastergrad.

Først og fremst vil jeg takke Lars Dag Theisen og Per Qvalbein, samt resten av Asplan Viak AS, avdeling Kristiansand for innspill og tips til tema, samt hyggelige stunder og godt arbeidsmiljø under arbeidet. Tomas Levin og resten av SINTEF transport forskning for gode idéer under hele arbeidsprosessen og tilrettelegging av både data og programmer.

Videre ønsker jeg å rette en helt spesiell takk til Trude Tørset, SINTEF / NTNU, som utvilsomt har fått hennes tålmodighet satt på prøve, men likevel stått på og vært en kontinuerlig inspirasjonskilde og veileder.

Til sist rettes en takk til mine klassekamerater i veg- og transportklassen for en fantastisk tid. Spesielt vil jeg fremheve Halvor Grønlund og Sivert Overå som har bidratt til gode diskusjoner, tips, triks og samvær som har vært meget nyttig i arbeidet.

Kristiansand, juni 2014.

Christoffer Viena Jørgensen

Sammendrag

Bruken av GPS-data i transportforskning er ikke en ny tanke, men data som ligger offentlig tilgjengelig på internett har blitt viet liten til ingen oppmerksomhet. I denne oppgaven skal en prøve å vurdere potensialet til GPS-data som er samlet inn av privatpersoner ved hjelp av treningsloggere og mobilapplikasjoner, og deretter delt på internett. Dette gjøres ved å utvikle en fartsmodell basert på tilgjengelig data, samt diskutere andre potensielle bruksområder for dataene.

Litteraturstudiet viser at mange av dagens transportmodeller har enkle tilnærminger til rutevalg og hastigheter for syklister. Det samme gjelder diverse ruteplanleggere som er tilgjengelig på nettet. Det kommer også frem at avstand og reisetid er viktige forklaringsfaktorer for syklisters rutevalg, men andre variabler som stigning, tilgjengelig infrastruktur, antall kryss og kontinuitet i sykkelnettverket er viktige.

Datagrunnlaget består av GPS-data som er samlet inn fra nettstedet *Garmin Connect*. Turer ble hentet fra Oslo, Kristiansand og Trondheim, og de var gjennomført i perioden 10. april 2013 til 10. april 2014. Hovedvekten av disse turene var fra Oslo. Etter filtreringer og selektering ble det til slutt lagd to datasett som bestod av ca. 1100 turer fordelt på 1 000 000 målepunkt hver. Det ene datasettet ble benyttet til å utvikle en fartsmodell basert på stigninger og stigningslengder, mens det andre datasettet ble brukt til å teste modellene. Modellene som ble utviklet viste middels forklaringsgrad og datasettet viste tegn til dårlig homoskedastisitet og autokollinearitet, men det var til en viss grad forventet da det antas at det er flere viktige variabler som ikke er tilgjengelig i datasettet.

Offentlig tilgjengelig GPS-data viser seg å ha mange potensielle bruksområder, men det bør først gjennomføres en kontrollert studie, som kan forbedre filtreringsprosessen og kartlegge personlige og fysiske forklaringsfaktorer som ikke er tilgjengelig i datasettet. Spesielt bør egenskaper som kjønn, fysisk form, erfaring, infrastruktur og trafikkmengde undersøkes. Resultatet bør brukes til å oppdatere fartsmodellen samt utvikle en rutevalgmodell for å tette hull i dagens transportmodeller.

GPS-dataene har også andre bruksområder som kan være interessante. Både areal- og transportplanlegging kan nytte godt av denne datakilden. Identifisering av konfliktområder og flaskehalsen samt før/etter-analyser er noen bruksområder hvor offentlig tilgjengelig GPS-data kan være en meget enkel, billig og relevant datakilde.

Summary

The use of GPS-data in transport research is not a new idea, but data that are available through online communities for sharing exercises and activities have received little to no attention. This thesis will reveal some of the potential of public accessible GPS-data from sport trackers and mobile applications. To do so, this study will develop a new method for calculating bicycle speeds based on GPS-data collected from such online communities as well as debating other potential uses.

Earlier studies show that many of the currently available route planners and transport models have relatively easy approximations towards bicycle speeds. It is also shown that bicyclist choose their route according to several factors and not just the shortest available one. These factors include elevation, available bicycle facilities, number of intersections and continuity of the bicycle network.

The dataset is collected from *Garmin Connect*, which is a social web page where people can upload, analyse and share their activities. The trips that were collected were from the Norwegian cities of Oslo, Trondheim and Kristiansand, and were carried out between 10th of April 2013, to 10th of April 2014. The majority of these trips were from Oslo. The data was split into two sets after filtering, each set with roughly 1100 trips and 1 000 000 track points. One set was used to create a method for calculating bicycle speeds with regards to slopes and slope lengths, while the other set was used to test the method. The developed method account for roughly 30 to 40 % of the observed variations in bicycle speed, and have some troubles with low homoscedasticity and autocorrelation. This was partly expected due to some important predictors not available in the dataset, as well as non-independent point data.

Public accessible GPS-data show great potential as a source for bicycle data in several areas of transport research, modelling and planning. However, a controlled study should be conducted to both improve the filtering process and identify important predictors that are not directly available in the data. The study should also look into important attributes that could be vital in the development of a route choice model for bicyclists.

Other possible uses, among many, might be before/after studies, identification and localisation of “hot spots”, speed barriers or conflict areas as well as identifying areas that are already well adapted to bicyclist and areas that are not. The GPS-data that are available are already collected and easy to both extract and process, and completely free.

Innholdsfortegnelse

Forord	V
Sammendrag.....	VII
Summary.....	IX
Figurliste	XIV
Tabelliste.....	XVI
Ordforklaringer.....	XVII
1 Innledning	1
1.1 Bakgrunn.....	1
1.2 Problemstilling og omfang.....	2
1.3 Oppgavens oppbygging	3
2 Litteratur om fart og rutevalg	5
2.1 4-steg metodikken	5
2.1.1 Nyttefunksjonen og Level-of-service.....	8
2.2 Forutsetninger for sykkelfart i ruteplanleggingsverktøy og transportmodeller	12
2.2.1 Sykkelfart i ruteplanleggere.....	12
2.2.2 Sykkelfart i transportmodeller	15
2.2.3 Oppsummering av sykkelfart i planleggere og modeller	16
2.3 Viktige faktorer for syklisters rutevalg.....	17
2.3.1 Global Positioning System (GPS), stated preference og revealed preference	17
2.3.2 Viktige forklaringsfaktorer for rutevalg – Stated preference studier	21
2.3.3 Viktige forklaringsfaktorer for rutevalg – Revealed preference studier	24
2.3.4 Oppsummering av syklisters rutevalg	28
2.4 Oppsummering	32
3 Metode og fremgangsmåte.....	33
3.1 Kartlegging, vurdering og datainnsamling	33
3.1.1 Datakilder	33

3.1.2	Tidlig filtrering og innhenting av data.....	36
3.1.3	Parsing, utregninger og videre filtreringer.....	40
3.1.4	Manuell filtrering, værdata og utvalg av turer	43
3.2	Fartsmodellen og beskrivende variabler	46
3.2.1	Statistikk fra datasettet og metode	46
3.2.2	Statistiske antagelser	49
3.2.3	Oppbygging og testing av modellene	51
3.3	Oppsummering	53
4	Resultat.....	55
4.1	Spredninger og andre funn i datasettet	55
4.1.1	Hastighetsfordeling og spredning	55
4.2	Ny metode for å beregne sykkelfart	62
4.2.1	Modell for fart i nedoverbakker (stigning < -2,0 %).....	62
4.2.2	Statistikk for metoden for negative segmenter	63
4.2.3	Modell for tilnærmet flate stigninger (-2,0 % ≤ stigning ≤ 2,0 %)	68
4.2.4	Statistikk for flatmarks-modellen.....	69
4.2.5	Modell for fart i oppoverbakker (Stigning > 2,0 %).	72
4.2.6	Statistikk for oppoverbakke-modellen	73
4.2.7	Utelatte variabler.....	75
4.3	Testing av modellene i alternativt datasett	76
4.4	Implementering i nettverket for DOM Nidaros	81
4.5	Oppsummering	85
5	Diskusjon.....	87
5.1	Tilgjengelige datakilder, bias og usikkerhet	87
5.1.1	Bias i datasettet	87
5.1.2	Usikkerheter i datasettet	88
5.2	Metode og modeller.....	93

5.2.1	Filtreringsprosessen.....	93
5.2.2	Modellene	94
5.2.3	Effekten av å benytte sykkelfart i stedet for avstand i RTM	96
5.2.4	Statistiske svakheter og manglende variabler	97
5.3	Bruk av datakildene til utvikling av rutevalgmodell	100
5.3.1	Andre potensielle bruksområder for bruk av offentlig tilgjengelig GPS-data..	101
5.4	Oppsummering	103
6	Konklusjon og videre arbeid	105
	Referanseliste	109
	Liste over vedlegg	113

Figurliste

Figur 2.1 – Oversikt over 4-steg metodikken	6
Figur 2.2 – Eksempel på O-D matrise.	7
Figur 2.3 – Skjermdump fra Google Maps	14
Figur 2.4 – Trilaterasjon	18
Figur 3.1 – Skjermdump fra Garmin Connect.	36
Figur 3.2 – Filtrert data på grunn av navn	37
Figur 3.3 – Eksempel på filtrering på grunn av dårlig fartsdata	38
Figur 3.4 – Eksempel på filtrering på grunn av dårlig høydedata.....	39
Figur 3.5 – Illustrasjon over segmentinndeling for en tilfeldig tur	42
Figur 3.6 – Eksempel på tur som ble fjernet fra datasettet	43
Figur 3.7 – Eksempel på tur som kanskje kan repareres	44
Figur 3.8 – Plot av fart og stigning for alle målepunkt.....	47
Figur 3.9 – Histogram for stigninger.	48
Figur 3.10 – Hastighetsfordelingen i datasettet.....	48
Figur 4.1 – Boxplot for stigning mellom -35 og 0 %	56
Figur 4.2 – Boxplot for stigninger mellom 0 og 35 %	57
Figur 4.3 – Boxplot for stigninger mellom 4 og 11 %	58
Figur 4.4 – Fordeling av hastigheter ved 4 % stigning.....	59
Figur 4.5 – Hastighetsfordeling ved 7 % stigning.....	60
Figur 4.6 – Hastighetsfordeling for 1 % stigning.....	61
Figur 4.7 – Partial Regression Plot av fart mot stigning	65
Figur 4.8 – Partial Regression Plot, fart mot $\log_{10}(\text{lengde})$	66
Figur 4.9 – Testing av homoskedastisitet	67
Figur 4.10 – Normalfordeling av residualene.	68
Figur 4.11 – Lineær sammenheng mellom lengde på forrige segment og fart.....	70
Figur 4.12 – Partial Regression Plot for stigning.	71
Figur 4.13 – Residualenes fordeling for flatmarks-modellen	72
Figur 4.14 – Lineær sammenheng mellom stigning og fart.....	74
Figur 4.15 – Homoskedastisitet for oppoverbakke-modellen.....	75
Figur 4.16 – Fordeling av feil mellom observert og modellert fart.....	77
Figur 4.17 – Observerte hastigheter i det alternative datasettet.....	78
Figur 4.18 – Boxplot for modellerte hastigheter, negativ gjennomsnittlig stigning.	79

Figur 4.19 – Boxplot for modellerte hastigheter, positiv gjennomsnittlig stigning.	79
Figur 4.20 – Stigninger ved negativt modellerte hastigheter.	80
Figur 4.21 – Stigningslengder ved negativt modellerte hastigheter.	80
Figur 4.22 – Modellert sykkel fart inn mot sentrum fra City Syd, Tyholt og Lade.	82
Figur 4.23 – Modellerte hastigheter ut fra sentrum mot de respektive områdene.	83

Tabelliste

Tabell 1.1 – Oppgavens oppbygging.....	3
Tabell 2.1 – Eksempel på LoS matrise for syklistene.....	10
Tabell 2.2 – Oversikt over sykkel fart i planleggere og modeller.....	16
Tabell 2.3 – Oversikt over den presenterte litteraturen.	30
Tabell 2.4 – Oversikt over variabelers innvirkning på rutevalg presentert i litteraturen	31
Tabell 3.1 – Oversikt over mulige kilder til GPS-data.	35
Tabell 3.2 – Oversikt over filtrering og deling av datasettet.	45
Tabell 3.3 – Statistikk fra datasettet.	46
Tabell 4.1 – Forklaringsgrad, standardavvik og Durbin-Watson for negative segmenter	63
Tabell 4.2 – Standardavvik og signifikansnivå for koeffisientene.....	64
Tabell 4.3 – Statistikk for flatmarks-modellen	70
Tabell 4.4 – Forklaringsgrad, standardavvik og Durbin-Watson for oppoverbakke-modellen	73
Tabell 4.5 – T-test for gjennomsnittlig differanse mellom observert og modellert fart.....	77
Tabell 4.6 – Statistikk fra modellerte hastigheter	78
Tabell 4.7 – Sammenlikning av reisetider mellom metodene.....	84
Tabell 4.8 – Oversikt over modellstatistikk.....	86
Tabell 5.1 – Antall observasjoner for små stigninger.....	89
Tabell 5.2 – Teoretisk eksempel på avrunding av høydedata.	91

Ordforklaringer

Gang- og sykkelveg	Skiltet veg for gående, syklende eller kombinert gang- og sykkeltrafikk. Adskilt fra annen veg med gress, grøft, gjerde eller annen måte (Statens Vegvesen, 2013).
GIS	Geografisk informasjonssystem. GIS er et integrert system som omfatter hardware, software og data som brukes til å prosessere, tolke og visualisere informasjon. I transport benyttes GIS ofte til å behandle informasjon om transportnettverk, bosettinger, arbeidsplasser, kollektivruter med mer. ArcGIS, GISLINE og Idrisi er noen eksempler på GIS.
GNSS	Global Navigation Satellite System. Fellesbetegnelse for navigasjonssystem som benytter seg av satellitter (Forssell, 2009a).
GPS	Global Positioning System. Amerikansk variant av GNSS, utviklet av militæret men nå tilgjengelig for sivil bruk (Forssell, 2009b).
Map matching	Algoritme for å automatisk koble GPS-data opp mot et nettverk et GIS, slik at man kan gjennomføre analyser i GIS.
Python	Programmeringsspråk for å utvikle egne skript og algoritmer.
Sykkelfelt	Eget kjørefelt langs vegen som er bestemt tilegnet syklist (Statens Vegvesen, 2013).
Sykkelveg	Skiltet veg for syklende som er adskilt fra annen veg med grøft, gjerde eller på annen måte. Kan benyttes av gående der dette er tjenlig (Statens Vegvesen, 2013).
ÅDT	Årsdøgntrafikk. Antall kjøretøy som har passert et snitt i vegen over et helt år, dividert på 365 (Statens Vegvesen, 2013).

1 Innledning

Trening og sosiale medier er i vinden som aldri før, og de går i større og større grad hånd i hånd. Bak «likes», «blinkefjes» og statusoppdateringer ligger det store mengder data og svever i «skyen». Hvilke data ligger der og hva kan de brukes til?

På samme tid er klimaet i krise og storbyene skriker etter gode og effektive tiltak for å begrense biltrafikken. Sykkel i hverdagen må opp og frem, men hvordan vet vi hva som er de gode løsningene?

1.1 Bakgrunn

Bakgrunnen for oppgaven ligger definert i den nasjonale transportplanen:

Tilrettelegging for miljøvennlig transport og arealplanlegging som reduserer transportbehovet er nødvendig for å oppfylle viktige deler av Klimaforliket; slik som målet om at veksten i persontransporten i storbyområdene skal tas av kollektivtransport, sykkel og gange. Samferdselsdepartementet (2013), PP. 16

Ambisjonen om at all vekst i persontransport i storbyene skal tas av kollektivtransport, sykkel og gange er ikke liten. Det naturlige oppfølgingsspørsmålet til sitatet over er: Hvordan? Hva må til for å nå målene? Hvilke tiltak skal man gå for?

For samferdselsprosjekter er det naturlig å hengi seg til Statens vegvesen, som opererer med konsekvensanalyser der «*En konsekvensanalyse er en analyse av sammenhengen mellom årsak og virkning, der tiltaket som skal analyseres er definert som årsak.*» Statens Vegvesen (2014a), PP 10. I konsekvensanalyser inngår det også en samfunnsøkonomisk analyse, som skal være en systematisk gjennomgang av alle relevante fordeler og ulemper som påføres samfunnet, som deles i prissatte og ikke-prissatte konsekvenser (Statens Vegvesen, 2014a). I det store og det hele ønsker man fra et samfunnsøkonomisk synspunkt de løsningene som gir størst nytte. Man vil ha igjen flest mulig av de kronene man investerer, og for trafikanter er reduserte reisetider, kostnader og økt sikkerhet noen eksempler på konsekvenser som oppfattes som fordelaktige.

Så hvordan bestemmer man nytten til et foreslått tiltak? Man prøver å forutsi hva som vil skje med trafikken dersom man bygger en ny strekning, setter opp bomring eller bytter ut ferje

med tunnel. Man benytter seg av transportmodeller for å tilnærme seg det som skjer i den virkelige verden, slik at man har en forutsetning for å vurdere konsekvensene. Men vi mangler de gode verktøyene for syklister. Hvilke ruter foretrekker de? Hvor fort sykler en syklist? Er det forskjell på sykkelveg og sykkelfelt? Velger man omveier for å unngå den verste oppoverbakken på veg til jobb, eller er det viktigere å komme raskest mulig frem? Svarene på disse spørsmålene kan vise seg å være essensielle for å planlegge de riktige tiltakene for å øke bruken av sykkelen i hverdagen.

Samtidig blir sosiale nettsteder for å dele, utforske og «skryte» av sine gjennomførte fysiske aktiviteter mer og mer populære. Nettsteder som Endomondo, Strava og RunKeeper eksisterer kun for at folk skal kunne dele sine aktiviteter til omverdenen. Man er alltid «connected», og dataene man har skal være synlig for andre. Det betyr at det er et stort potensial for enkel innhenting av data, der grovarbeidet allerede er gjennomført. Store mengder informasjon om hastigheter og rutevalg ligger og venter på å bli analysert, og kan vise seg å være meget nyttige i arbeidet med å forbedre modellering og planleggingen av sykkeltrafikk.

1.2 Problemstilling og omfang

Denne oppgaven har som hensikt å undersøke potensialet til data som ligger i «skyen», og bruke disse dataene til å prøve å løse noen av utfordringene man har med modellering av sykkeltrafikk:

Kan GPS-sporing av sykkelturer brukes som grunnlag for en fartsmodell og rutevalgmodell for syklende?

Det er ønskelig å identifisere viktige forklaringsfaktorer for syklisters rutevalg gjennom en litteraturstudie, for deretter å se om GPS kan benyttes til å beskrive eller forklare disse faktorene. Samtidig vil det undersøkes hvordan dagens ruteplanleggere og transportmodeller har tilnærmet seg sykkelfart. Videre skal det identifiseres og samles GPS-data fra offentlig tilgjengelige kilder, for så å utvikle en metode for å beregne sykkelfart ved hjelp av disse dataene. Til sammen skal dette utgjøre grunnlaget for en diskusjon rundt potensialet til offentlig tilgjengelig GPS-data.

Denne studien vil legge vekt på å benytte seg av pendlerturer, hovedsakelig utført i Trondheim og Kristiansand. Det gir muligheten til å gjennomføre observasjoner og bruke lokalkunnskapen. Blir det nødvendig å utvide datasettet kan data fra Oslo, Bergen eller Stavanger benyttes. Det ønskes å unngå rene treningsturer som ikke vil være like representative for å modellere pendlerturer. Metoden for fartsberegning skal utvikles ved hjelp av lineær regresjon med den innsamlede GPS-dataen.

1.3 Oppgavens oppbygging

Oppgaven består av seks kapitler, gitt i tabellen under. I kapitlene 2 til 5 er det gitt en oppsummering på slutten av hvert kapittel som gjengir de viktigste punktene.

Tabell 1.1 – Oppgavens oppbygging

Kapittel	Innhold
1 - Innledning	Bakgrunn, problemstilling og avgrensinger.
2 - Litteraturstudie	Sykkelfart i ruteplanleggere og transportmodeller. Viktige faktorer for syklisters rutevalg.
3 - Metode	Gjennomgang av datainnsamling, filtrering og modellutvikling.
4 - Resultat	Resultater fra datasettet og modellarbeidet.
5 - Diskusjon	Drøfting av resultat opp mot problemstilling og litteraturstudie.
6 - Konklusjon og videre arbeid	Svar på problemstilling og forslag til videre arbeid rundt temaet.

2 Litteratur om fart og rutevalg

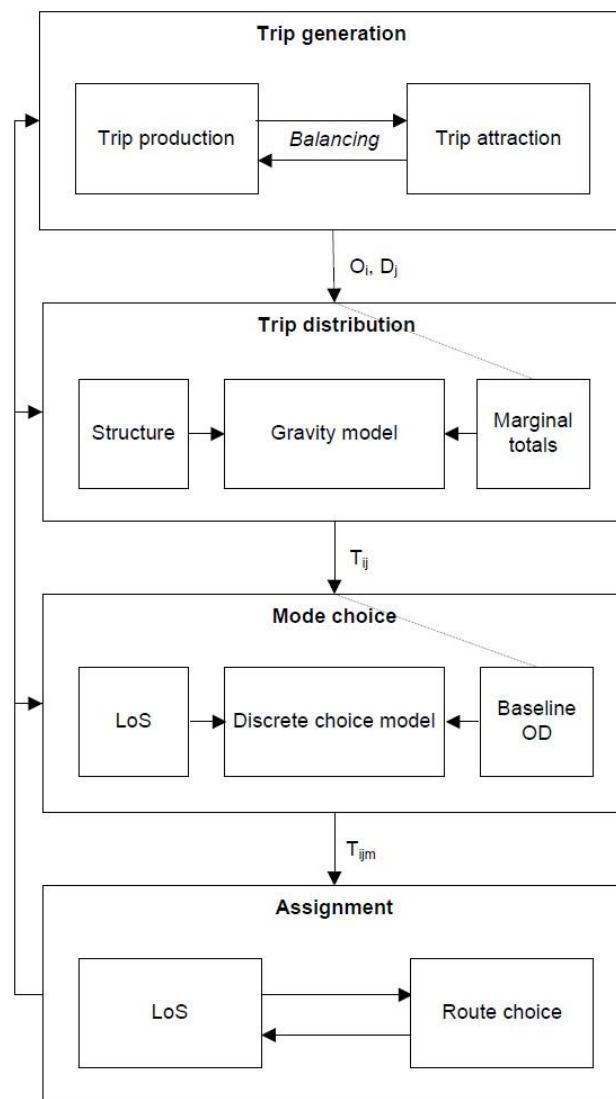
Litteraturkapittelet skal belyse metoder og tidligere forskning innenfor modellering av fart og rutevalg for syklist, og består av tre deler. Ettersom både fart og rutevalg er essensielle parametere i transportmodellering etter 4-steg metodikken, er det gitt en introduksjon til denne i del 2.1. Andre del viser hvordan eksisterende ruteplanleggingsverktøy og transportmodeller beregner farten til syklist, mens tredje del viser viktige forklaringsfaktorer for syklisters rutevalg. Litteraturen er hovedsakelig funnet gjennom søk på internett og biblioteksdata-baser som NTNUs *BIBSYS*, Googles *Scholar*, *Scencedirect* m.fl. Søkene har bestått av ulike sammensetninger av ord som *bicycle*, *speed*, *modeling*, *gradient*, *route choice* og *route planner*, samt forskjellig synonym av disse, også på norsk.

2.1 4-steg metodikken

En metode for å modellere transport er 4-steg metodikken. Metodikken er, som navnet tilsier, en sekvensiell metodikk som består av fire steg (Rich, 2009):

1. Turgenerering (Trip generation)
2. Turdistribusjon (Trip distribution)
3. Reisemiddelvalg (Mode choice)
4. Nettutlegging (Assignment)

I det første steget beregnes det antall turer som gjennomføres i analyseområdet. Dette kan gjøres ved å dele området inn i flere soner for så å analysere sonene hver for seg. Sonedata som antall bosatte, aldersfordelinger, antall skoler og barnehager i sonen, varehandel, bilhold blant de bosatte, samt antall arbeidsplasser kobles opp mot reisevaneundersøkelser, som gir informasjon om nettopp reisevaner for forskjellige deler av befolkningen. Sammen kan dette brukes til å anslå hvor mange reiser en sone genererer og hvor mange reiser den tiltrekker.



Figur 2.1 – Oversikt over 4-steg metodikken (Rich, 2009, pp. 15).

Etter at dette er gjort må turene fordeles mellom sonene og innad i sonene. Etter den tradisjonelle 4-steg metodikken, som er vist i figur 2.1, fordeles turene gjerne ved bruk av en gravitasjonsmodell¹ som er vist i formel (1) (de Dios Ortúzar og Willumsen, 2011):

$$T_{ij} = \alpha O_i D_j f(c_{ij}) \quad (1)$$

Der

T_{ij} = Antall turer som går fra sone i til sone j

α = Proporsjonalsfaktor

¹ Gravitasjonsmodellen er en måte å fordele turer på, og det finnes andre modeller. Andre varianter av 4-steg metodikken benytter andre fremgangsmåter. Det vil ikke bli gått i dybden på noen av dem her.

- O_i = Antall turer som er generert i sone i
- D_j = Antall turer som ender i sone j
- $f(c_{ij})$ = Generalisert funksjon for reisekostnader mellom sone i og sone j.

Reisekostnadene i siste ledd skal prøve å beskrive alle de opplevde kostnadene som skjer i forbindelse med en reise, både direkte og indirekte. Derfor kalles de ofte for de generaliserte kostnadene ved en reise (de Dios Ortúzar og Willumsen, 2011). For bilister vil de generaliserte kostnadene bestå av element som bompenger, reisetid o.l., mens for syklister benyttes det gjerne avstand.

Gravitasjonsmodellen benyttes for alle soner og reisehensikter. Resultatet fra gravitasjonsmodellen presenteres gjerne i en «Origin Destination (O-D) Matrix». O-D matrisen forteller oss hvor mange turer som går fra sone i til sone j for alle sonene i analyseområdet. Et eksempel er vist i figur 2.2, der T_{11} er antall turer som skjer innad i sone 1, mens T_{12} er antall turer som starter i sone 1 og slutter i sone 2, og så videre.

Origins	Destinations					$\sum_i T_{ij}$
	1	2	3	...j	...z	
1	T_{11}	T_{12}	T_{13}	... T_{1j}	... T_{1z}	O_1
2	T_{21}	T_{22}	T_{23}	... T_{2j}	... T_{2z}	O_2
3	T_{31}	T_{32}	T_{33}	... T_{3j}	... T_{3z}	O_3
⋮						
i	T_{i1}	T_{i2}	T_{i3}	... T_{ij}	... T_{iz}	O_i
⋮						
Z	T_{z1}	T_{z2}	T_{z3}	... T_{zj}	... T_{zz}	O_z
$\sum_i T_{ij}$	D_1	D_2	D_3	... D_j	... D_z	$\sum_{ij} T_{ij} = T$

Figur 2.2 – Eksempel på O-D matrise. Hentet fra de Dios Ortúzar og Willumsen (2011), PP. 176.

Når turene er fordelt mellom sonene kan man begynne å analysere hvilket reisemiddel som benyttes for hver reise. I reisemiddelfordelingen benytter man resultatet fra O-D matrisen sammen med *Level of Service*-data (LoS-data) og en *discrete choice model*, altså en modell for diskrete valg².

² I diskrete valg har man et endelig sett med alternativer å velge mellom, og man antar at personer som står ovenfor et diskret valg velger det alternativet som gir dem størst nytte (Hensher og Button, 2008).

Diskrete valg kan modelleres ved hjelp av en *multinomial logit model*, MNL (de Dios Ortúzar og Willumsen, 2011):

$$P_{i,q} = \frac{e^{V_{i,q}}}{\sum_{i'} e^{(V_{i',q})}} \quad (2)$$

Der $P_{i,q}$ er sannsynligheten for at individ q velger alternativ i , mens $V_{i,q}$ er de nytten for individ q når det benytter seg av alternativ i , også kjent som *nyttefunksjonen*. Jo høyere nytten for alternativ i er, jo større er sannsynligheten for at det alternativet blir benyttet. Ut i fra sannsynlighetene som beregnes i MNL får man reisemiddelfordelingen.

LoS-data består av flere matriser som er tilsvarende O-D matrisen, men i stedet for antall turer mellom sonene inneholder matrisene verdier som er relevante for nyttefunksjonen. Ettersom nyttefunksjonen er en meget viktig del av reisemiddelfordelingen gjennomgås denne, sammen med LoS, i detalj i kapittel 2.1.1.

Etter at reisemiddelvalg er gjennomført gjenstår det bare å legge ut de beregnede turene på et nettverk. Det betyr at trafikken «plasseres» på nettverket, slik at man kan analysere trafikkmengder. Det finnes flere metoder for å utføre nettutleggingen på, men en vanlig antagelse er at personer velger den ruten som gir dem de laveste generaliserte kostnadene. Nettverket representerer hovedsakelig det fysiske transportnettverket, men har også informasjon som lengden på en strekning, hastigheter ved fri flyt, kapasiteter, rutetider og holdeplasser for kollektivtransport, med mer (de Dios Ortúzar og Willumsen, 2011).

2.1.1 Nyttefunksjonen og Level-of-service.

Nyttefunksjonen prøver å beskrive hvor mye nytte man har ved å benytte seg av et enkelt alternativ:

$$U_{i,q} = V_{i,q} + \varepsilon_{i,q} \quad (3)$$

Der $U_{i,q}$ er nytten individ q har ved å velge alternativ i . $V_{i,q}$ er den deterministiske komponenten, som ofte kalles for *den observerte nytten* (de Dios Ortúzar og Willumsen, 2011, Hensher og Button, 2008). Den observerte nytten skal beskrive alle målbare variabler

som påvirker nytten. I tillegg er det et feilledd, $\varepsilon_{i,q}$, som er med for å beskrive forskjellen mellom observert og virkelig nytte, og antas å være Gumbel-fordelt³ (Rich, 2009).

Den observerte nytten kan igjen beskrives som:

$$V_{i,q} = \sum_j \beta_j * x_j \quad (4)$$

Der β_j er konstanten og x_j er variabelen til forklaringsfaktor j . Dersom β_j er negativ betyr det at forklaringsfaktor j oppfattes som en kostnad som øker i takt med x_j . Er den positiv oppleves forklaringsfaktoren som nyttig. Et eksempel på nyttefunksjonen for sykkel er gitt i formel (5).

$$V_{i,q} = aL + b\text{\AA} + cK + dA \quad (5)$$

Der

L	=	Avstand fra start til slutt
\AA	=	Årstid
K	=	Kjønn
A	=	Alder
a, b, c, d	=	Konstanter

I nyttefunksjonen er noen variabler direkte avhengig av forholdet mellom sonene, mens andre variabler fungerer som «dummies» eller er ikke avhengig av sonerelasjoner. I eksempelet over vil avstand være en soneavhengig variabel, mens årstid og kjønn vil være dummyvariabler som bare kan ha spesielle verdier.

Level-of-Service-data er en matrise som viser en samlet oversikt over de generaliserte kostnadene som er avhengig av nettverket. LoS-matrisen viser hvordan de forskjellige kostnadene varierer mellom forskjellige soner for et bestemt reisemiddel. For syklist vil LoS-data ofte bare bestå av avstand eller reisetid, og et eksempel på LoS-matrise for sykkel er gitt i tabell 2.1. I tabellen er L_{12} avstanden fra sone 1 til sone 2, mens L_{34} er avstand fra sone 3 til 4, og så videre. LoS-data for andre transportmidler vil ha andre eller flere variabler. For

³ Det fører til at feilleddet forsvinner i utledningen av MNL-modellen, og at man i praksis benytter observert nytte i MNL-modellen.

eksempel vil en LoS-matrise for kollektivtransport bestå av reisedistanse, ombordtid, gangtid, ventetid, byttetid, og billett-kostnad.

Tabell 2.1 – Eksempel på LoS matrise for syklist

Opprinnelse	Destinasjon					
	1	2	3	4	m
1	L_{11}	L_{12}	L_{13}	L_{14}	$L_{1\dots}$	L_{1m}
2	L_{21}	L_{22}	L_{23}	L_{24}	$L_{2\dots}$	L_{2m}
3	L_{31}	L_{32}	L_{33}	L_{34}	$L_{3\dots}$	L_{3m}
4	L_{41}	L_{42}	L_{43}	L_{44}	$L_{4\dots}$	L_{4m}
....	$L_{\dots 1}$	$L_{\dots 2}$	$L_{\dots 3}$	$L_{\dots 4}$	$L_{\dots\dots}$	$L_{\dots m}$
n	L_{n1}	L_{n2}	L_{n3}	L_{n4}	$L_{n\dots}$	L_{nm}

LoS-data blir beregnet ut i fra nettverket som benyttes. Som nevnt inneholder nettverket blant annet informasjon om lengder på strekninger, hastigheter ved fri flyt og kapasitet. Når trafikken legges på nettet vil man kunne bruke informasjonen fra nettet og sammenlikne med transporten som går på dem. Dersom det er trafikkmengder over kapasiteten på en lenke vil hastighetene på denne lenken synke og reisetidene øke, og påvirker både de generaliserte kostnadene og LoS-dataene.

En endring i disse kan altså gjøre det mer attraktivt å reise til en annen sone, benytte et annet reisemiddel eller en annen rute. Det betyr at de fire stegene i metoden ikke bare er avhengig av data fra foregående steg, men også fra neste steg. Derfor er det vanlig å gjennomføre flere iterasjoner i 4-steg metodikken slik at både LoS-data, generaliserte kostnader, reisemiddel-fordelingen, rutevalgene og nettutleggingen stabiliserer seg. Denne tilbakekoblingen av data gjør at det er viktig med både gode forklaringsfaktorer i nyttefunksjonen og de generaliserte kostnadene for å få gode resultat fra modellen.

Som nevnt er avstand ofte brukt i begge uttrykkene for sykkel, men avstanden mellom to soner vil variere med hvilken rute som benyttes. En vanlig tilnærming til syklisters rutevalg er at de benytter seg av den korteste ruten tilgjengelig, men er det virkelig tilfellet? For det første er ikke den korteste ruten nødvendigvis den raskeste. Kan det være at syklist foretrekker ruter som er lengre men raskere? For det andre kan det være at syklist unnviker fra den korteste ruten. Ønsker de å unngå spesielle områder, benytte seg av bestemte fasiliteter, eller er de ikke klar over hva som er den korteste ruten?

Det kan altså være betydelige forskjeller mellom hvilke forklaringsfaktorer som benyttes i modelleringen og hvilke forklaringsfaktorer som faktisk er viktige for syklistene, og det kan få konsekvenser både for turdistribueringen og reisemiddelvalget. I det hele tatt kan det påvirke det totale modellresultatet og gi dårlig tilnærming til virkeligheten for sykkelmodelleringen.

2.2 Forutsetninger for sykkelfart i ruteplanleggingsverktøy og transportmodeller

I dette kapittelet undersøkes det hvordan eksisterende verktøy og modeller beregner sykkelfart. Fart antas å være en viktig forklaringsfaktor for reisemiddel- og rutevalg ettersom den direkte påvirker reisetid. Både ruteplanleggere og transportmodeller kan være nyttige hjelpemidler for å vurdere eksisterende situasjon og planlegge nye tiltak i infrastrukturen. For eksempel kan ruteplanleggere promotere sykling ved at brukere får en bedre oversikt over tilgjengelig infrastruktur og gode sykkelruter, eller hjelpe kommunale myndigheter med å identifisere områder som er lite egnet for sykkel slik at disse kan utbedres (Su et al., 2010). Modellverktøy er i større grad ikke tilgjengelige for publikum men brukes av statlige og kommunale myndigheter som et verktøy i planlegging av transport, for eksempel identifisering av problemområder, samordnet areal- og transportplanlegging og evaluering av tiltak.

2.2.1 Sykkelfart i ruteplanleggere

Med ruteplanleggere menes verktøy, gjerne tilgjengelig på internett, som skal hjelpe med å finne en akseptabel reiserute fra A til B. Akseptabel reiserute kan være veldig individuelt, så derfor presenteres ofte informasjon om reisetid, stigning, tilgjengelige sykkelfasiliteter med mer. Ettersom de fleste reiseplanleggerne presenterer en reisetid er det ønskelig å se på hvilke forutsetninger de har gjort for å beregne reisetiden.

Mange av de tilgjengelige ruteplanleggerne på nett lar deg sette din egen sykkelhastighet: Transport for Londons *Cycle Journey Planner* (CJP) lar deg velge mellom tre sykkelhastigheter og rutevalg: Lett (12 km/t, hovedsakelig rolige gater og stier), moderat (16 km/t, delvis korteste rute med større gater og trafikkmengde) og rask (20 km/t og en direkte rute) (Transport for London, Årstall ikke oppgitt). Cyclestreet.net's *Journey Planner* (JP) lar deg velge mellom behagelig (16 km/t), «cruising» (20 km/t) og hurtig (24 km/t) (Nuttall og Lucas-Smith, 2010). Også Transport Directs *Route Planner* (RP) gir brukere muligheten til å legge inn fart selv, men dersom man ikke legger inn fart benyttes det tilsynelatende en standard fart mellom 17 og 20 km/t⁴. I tillegg kan man velge mellom tre forskjellige ruteprofiler: «Quietest», «Quickest» og «Most Recreational» (Transport Direct, 2010).

⁴ Basert på egne tester i RP.

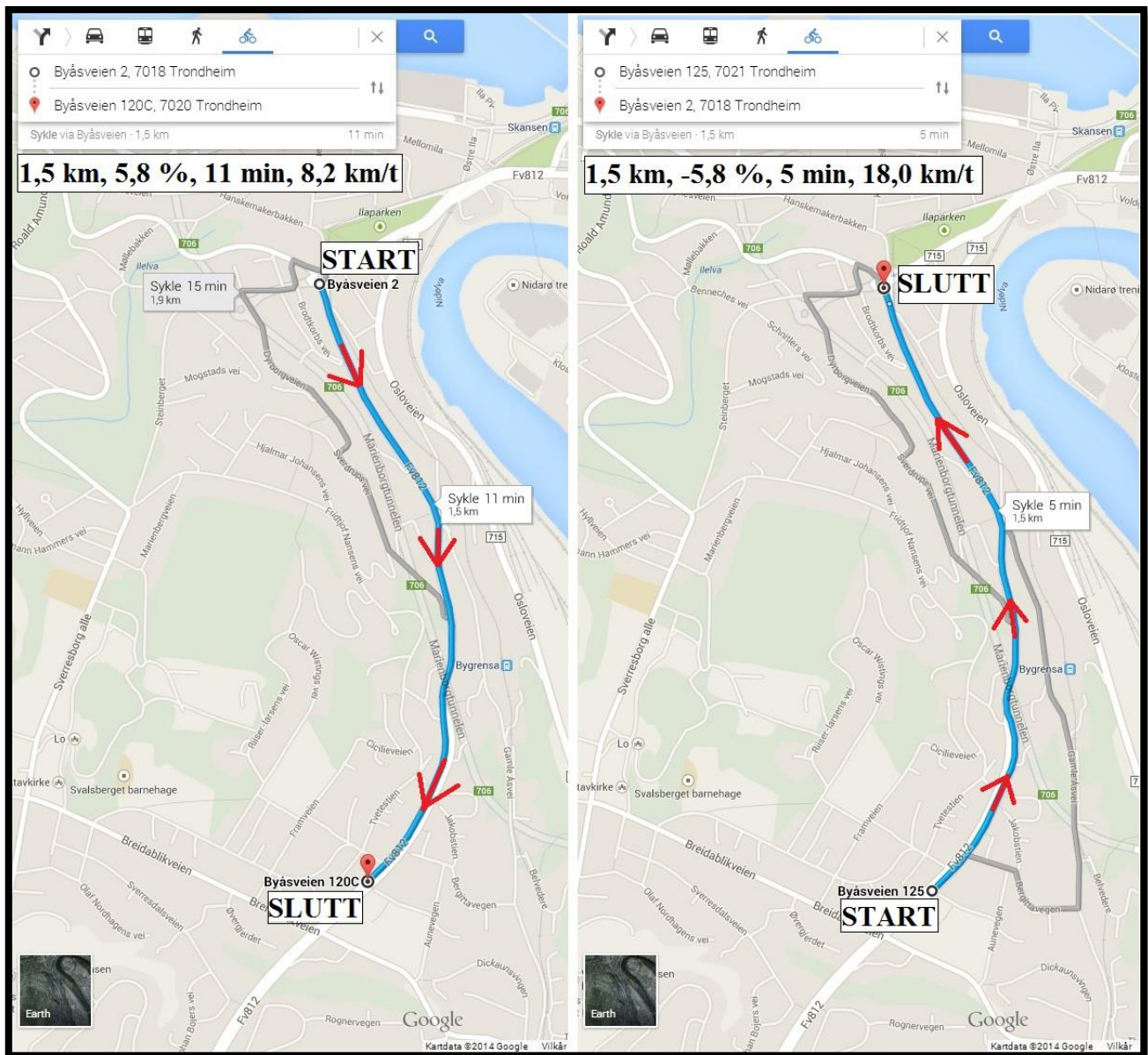
Mens CJP kun benytter seg av den oppgitte hastigheten og ingen andre faktorer spiller inn⁵, har JP tatt med forsinkelser grunnet stigning, lyskryss og kryssing av veger, men ikke oppgitt hvor store disse er eller hvordan de beregnes (Nuttall og Lucas-Smith, 2010). I RP beregnes reisetiden ut i fra den oppgitte hastigheten, men med tillegg for stigning og hva slags infrastruktur du sykler på (Transport Direct, 2010). Hvor stort dette tillegget er for de enkelte stigninger og sykkelanlegg er ikke oppgitt.

Google *Maps* er et verktøy som er mye brukt i Norge, blant annet av Stavanger kommune og *Miljøpakken* i Trondheim (Miljøpakken, 2014, Stavanger Kommune, 2013). Google Maps har en innebygd ruteplanlegger for syklist, men dokumentasjon på hastighetsberegninger har vært vanskelig å oppdrive. I følge en forumpost på Googles produktforum tar rutevelgeren hensyn til stigning, men ikke stopp eller nedbremsinger på grunn av kryss og liknende (Google, 2010). Dette er en upålitelig kilde som må sees på med skeptiske øyne, men egne tester i Google Maps viser at farten varierer mellom to punkter avhengig av hvilken retning man sykler, som indikerer at reisetiden er påvirket av stigningen langs ruten.

Et godt eksempel på dette er gitt i figur 2.3 der reisetiden mellom Byåsveien 120C og Byåsveien 2 i Trondheim er vist i Google Maps. (For returen benyttes Byåsveien 125, som ligger på motsatt side av vegen fra Byåsveien 120C, for å forhindre kluss med kjøreretninger.) Avstanden er ca. 1,5 kilometer i begge tilfellene, og går langs Byåsveien (Fylkesveg 829) som har en gjennomsnittlig stigning på 5,8 % (Statens Vegvesen, 2014b). Dersom man sykler nedover mot Byåsveien 2 blir reisetiden oppgitt til ca. 5 minutter, mens i motsatt retning blir den oppgitt til ca. 11 minutter⁶. Det gir en fart på ca. 7,5 km/t i oppoverbakke mot ca. 18,0 km/t i nedoverbakke.

⁵ Basert på egne tester i CJP.

⁶ Alle reisetider over 1 minutt blir rundet til hele minutter, mens alle avstander over 1 kilometer blir rundet til hele 100 meter, så det er antagelig feilmarginer på ± 30 sekunder og ± 50 meter.



Figur 2.3 – Skjermdump fra Google Maps med reisetid mellom to punkter i forskjellig retning. Venstre rute viser reisetid når man sykler opp Byåsveien, mens høyre rute viser motsatt retning.

Sandnes kommune, via *Sykelbyen Sandnes*, har et interaktivt sykkelkart utviklet av Norkart AS med en innebygd vegviser (Sandnes Kommune, 2010). Larvik kommune benytter seg av samme løsningen (Larvik Kommune, 2013). Heller ikke i dette tilfellet har det vært mulig å oppdrive dokumentasjon på hvordan reisetidene for syklister er beregnet. I skrivende stund er ikke kartet over Sandnes helt operativt, mens planleggeren i Larvik gir veldig høye hastigheter som kan tyde på at det er fart for bilister og ikke syklister (Over 30 km/t i bynære områder).

Sykkelveg er Statens Vegvesens egen sykkelplanlegger, som for tiden er implementert i områdene Hedmarken, Nedre Glomma (Fredrikstad/Sarpsborg) og Gjøvik (Jetlund, 2013). Fra Jetlund (2013) kan vi lese at: «*I ruteplanleggeren er det brukt en avansert effektmodell som tar hensyn til at ulike typer syklist i ulik grad blir påvirket av høydeforskjeller.*» I en epost den 20. Februar sa Jetlund (2014) at rutevalget i sykkelveg er basert på en modell fra Svenske Triona AB, med en effektmodell for beregning av fart og stigning. Effektmodellen bruker en energimodell som blant annet tar hensyn til vindmotstand, friksjon mellom dekk og underlag, hva slags sykkel som benyttes og hvor stor effekt personen som sykler genererer (varierer f.eks. om det er treningstur eller pendling) (Löfås, 2014). Energimodellen brukes som grunnlag for å bestemme hastigheter og rutevalg, og det er i tillegg lagt inn forsinkelser ved veldig bratte stigninger (der man antar at syklisten går av sykkelen) og i krappe svinger der syklisten er nødt til å bremse for å unngå å havne utenfor vegen (Löfås, 2014).

2.2.2 Sykkelfart i transportmodeller

ATP-modellen er mer en tilgjengelighetsmodell enn transportmodell da den hovedsakelig brukes til å beregne tilgjengelighet til eller fra et bestemt område, men det er likevel interessant å se hvordan farten modelleres. Modellen ble utviklet i perioden 1997-1999 av Asplan Viak AS med støtte fra LOKTRA-programmet i Norsk Forskningsråd (Brukernetverket for ATP-modellen, 2005), og er fortsatt i bruk. Modellen kan blant annet beregne reiseruter og reisetider for syklist, og reisetidene baserer seg på en gjennomsnittshastighet satt til 15 km/t som endres i takt med stigningen på strekningen (Norrdal og Ørnes, 2012). Hastighetene er begrenset til minst 5 km/t på strekninger med høy stigning til maksimalt 40 km/t på strekninger med stor helning (Norrdal og Ørnes, 2012). Det er også lagt inn en hastighet på 3 km/t der stigningen er såpass bratt at man antar personer vil gå av sykkelen og leie den opp bakken (Norrdal og Ørnes, 2012).

Den regionale transportmodellen (RTM) er en landsdekkende transportmodell for reiser kortere enn 100 km. Modellen ble utviklet gjennom et samarbeid mellom flere transportetater i prosjektet «NTP Transportanalyser», og er spesifikt tilpasset de fem regionene som tilsvarer Statens Vegvesens regioninndeling (Madslie et al., 2005). RTM bruker blant annet 4-steg metodikken for å beregne trafikkmengder, også for fotgjengere og syklist. I modellen benyttes det ikke fart men distanse i de generaliserte kostnadene (Rekdal et al., 2013), og det er først i trafikantnytte-modulen (kost/nytte-beregninger) at det er lagt inn en gjennomsnittsfart på 15 km/t (Malmin, 2013), som er hentet fra Statens forureningstilsyn (2001).

2.2.3 Oppsummering av sykkelfart i planleggere og modeller

Tabell 2.2 gir en oversikt over fart og hvilke faktorer som påvirker farten i de nevnte ruteplanleggerne og modellene. Stigning brukes i flere, men de færreste har oppgitt detaljer om hvordan farten blir påvirket, eller hvilke data de baserer antagelsene på. Fra tabellen ser vi også at ruteplanleggerne har litt høyere fart enn hva de to modellene presenterer på flat mark. Til sammenlikning målte Parkin og Rotheram (2010) en gjennomsnittlig hastighet på 21,6 km/t på flater, en gjennomsnittlig økning i hastighet på 0,86 km/t for hver 1 % stigningsøkning i nedoverbakke og en gjennomsnittlig reduksjon på 1,44 km/t per 1 % stigningsøkning i oppoverbakke. Den studien ble gjennomført med forsøkspersoner som var erfarne syklister og delaktige i sin lokale sykkelklubb, som kan tyde på at hastighetene er noe høye i forhold til den gjennomsnittlige syklist (Parkin og Rotheram, 2010).

Tabell 2.2 – Oversikt over sykkelfart i planleggere og modeller

Ruteplanlegger / modell	Stigning	Infra- struktur	Trafikk- mengde	Tidsfor- sinkelse i kryss	Fart på flat mark [km/t]	Variasjoner i fart [km/t]
Cycle Journey Planner	-	-	-	-	12, 16 eller 20*	
Journey Planner	X	-	-	X	16, 20 eller 24*	
Route Planner	X	X	-	-	~ 17 – 20**	
Google Maps	X	-	-	.	~ 16 – 20***	~ 6 – 30***
Norkart / Webatlas	Uvisst	Uvisst	Uvisst	Uvisst	Uvisst	Uvisst
Sykkelveg	X	Uvisst	Uvisst	Uvisst	~ 18****	~ 2 – 40****
ATP-modellen	X	-	-	-	15	3 - 40
RTM*****	-	-	-	-	15	-

X = Påvirker farten

- = Påvirker ikke farten

* Avhengig av hva brukeren spesifiserer.

** Basert på egne tester i Route Planner

*** Basert på egne tester i Google Maps.

**** Basert på egne tester i Sykkelveg.

***** Benyttes ikke før kost/nytte-beregningene.

2.3 Viktige faktorer for syklisters rutevalg

I dette underkapittelet skal viktige faktorer for syklistenes valg av ruter mellom to punkter belyses. Noen av studiene har benyttet seg av *Global Positioning System* (GPS) i undersøkelsen, mens andre studier har brukt spørreundersøkelser. Det gjør at en kort introduksjon til GPS og dets virkemåte, samt *stated preference*- og *revealed preference*-studier vil være nyttig for leseren. Dette er gitt i kapittel 2.3.1. Litteratur som benyttet *stated preference* er gitt i kapittel 2.3.2, mens litteratur som benyttet *revealed preference* er gitt i kapittel 2.3.3.

2.3.1 Global Positioning System (GPS), *stated preference* og *revealed preference*

GPS er et verdensomspennende, satellittbasert navigasjonssystem (Global Navigation Satellite System, GNSS) som er utviklet av det Amerikanske forsvarsdepartementet (Forssell, 2009b, Forssell, 2009a). Det finnes flere varianter av GNSS som GLONASS (Russisk) og GALILEO (EU), og disse systemene brukes primært til navigasjon og posisjonering. (Ogaja, 2011, Forssell, 2009a).

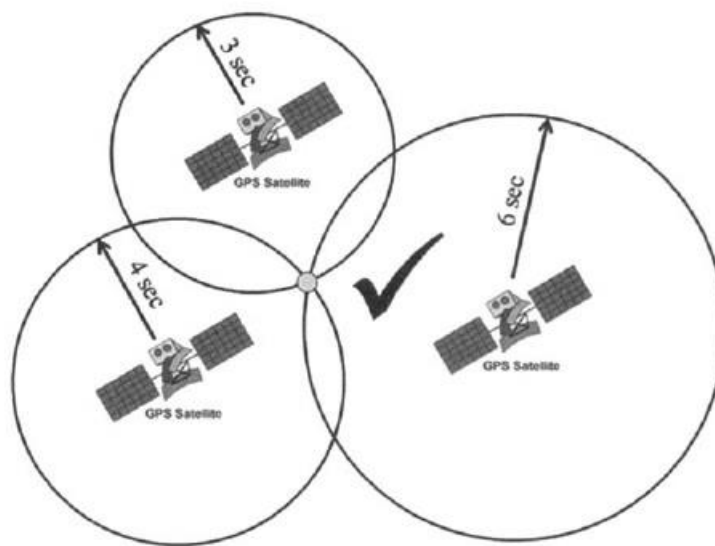
GPS består av tre hovedkomponenter (Ogaja, 2011): Bruker-komponenten, kontroll-komponenten og verdensrom-komponenten. Verdensrom-komponenten består av flere satellitter som går i bane rundt jorden, mens kontroll-komponenten består av en *Master Control Station* og flere *Passive Monitoring Stations*. Disse sørger for at satellittene fungerer som de skal (Ogaja, 2011). Den siste komponenten er den som blir benyttet av brukeren, ofte kalt GPS-mottaker. Til å begynne med var GPS-systemet tiltenkt store fremkomstmidler som skip og fly, men med årene har GPS-mottakeren blitt mindre, mer kompakt og mer energieffektiv slik at det har vært mulig å produsere håndholdte varianter (Ogaja, 2011). Det betyr at bruksområdene har økt kraftig, og i dag finner man GPS-mottakere i både biler, industrimaskiner, smarttelefoner og treningsredskaper som pulsklokker og sykkelcomputere (Garmin Ltd., 2014, Ogaja, 2011).

For å bestemme posisjon benytter GPS seg av en gammel landmåler-teknikk som er kjent som *trilaterasjon* (Ogaja, 2011). Trilaterasjon er en metode som er avhengig av at man har minst tre kjente punkt og en fremgangsmåte for å bestemme avstanden mellom de kjente punktene og den ukjente posisjonen. For GPS er det satellittene som er de kjente punktene, og de sender ut et radiosignal med kjent hastighet som også inneholder tidspunktet signalet ble sendt ut på. Dermed kan GPS-mottakeren bestemme hvor lang tid signalet brukte på å komme frem til

mottakeren, og på den måten bestemme avstanden til satellitten ved hjelp av sammenhengen mellom hastighet, tid og lengde:

$$\text{Hastighet} = \frac{\text{Lengde}}{\text{Tid}} \quad (6)$$

Ved å gjøre dette fra tre eller flere kjente punkter kan man bestemme posisjonen til mottakeren, og prinsippet for trilaterasjon er vist i figur 2.4 (Ogaja, 2011). Dersom signalet fra tre forskjellige satellitter bruker henholdsvis 3,4 og 6 sekunder på å nå GPS-mottakeren er det bare ett sted man kan befinne seg.



Figur 2.4 – Trilaterasjon (Ogaja, 2011)

Resultatet presenteres ofte i tredimensjonalt koordinatsystem som X-, Y- og Z-koordinater. Nøyaktigheten til GPS varierer med flere faktorer og forskjellige kilder oppgir forskjellig nøyaktighet. For håndholdte, sivile mottakere melder Ogaja (2011) at det er det vanlig med ca. fem meters nøyaktighet i det horisontale planet (X- og Y-koordinatene) og ti meter i det vertikale planet (Z-koordinaten). Stopher og Speisser (2011) opplevde en nøyaktighet innenfor ± 2 meter for syklistene, men opplevde også at enheter mistet kontakt med satellittene i opp til to minutter. I en annen studie melte Stopher et al. (2008) om ± 10 meters nøyaktighet for noen enheter, men ned mot $\pm 2,5$ meter for andre, men det er usikkert på om dette var kun i x- og y-koordinater eller om den inkluderte høyde. For mobiltelefoner var nøyaktigheten noe

varierende men innenfor de samme grensene som nevnt over (Friedl, 2012, Shaner, 2013). Dette var riktig nok enkle, praktiske tester som ikke var av vitenskapelig natur.

Nøyaktigheten vil også variere med hvor mange satellitter man har kontakt med (flere satellitter øker nøyaktigheten), ubrutt kontakt mellom GPS og satellitt (høye bygninger eller liknende som blokkerer signalene), atmosfæren og mottakeren (kompliserte enheter kan øke nøyaktigheten, men ofte på bekostning av størrelse og pris). Det finnes eksempelvis noen GPS-mottakere som benytter seg av lokale basestasjoner som gjør at man kan få nøyaktige målinger ned på centimeternivå (Ogaja, 2011). Introduksjonen av GLONASS og GALILEO gjør at mottakere kan benytte seg av alle de tre globale satellittnavigasjonssystemene. Dersom mottakerne i tillegg kan bruke mobilnettverket til å triangulere posisjonen sin slik som mange av dagens smarttelefoner kan, vil man etter hvert få et meget nøyaktig system til en relativt liten kostnad.

Det finnes også forskjellige metoder for å logge GPS-dataene. Noen enheter kan logge data hvert sekund, men dette krever at enheten har tilstrekkelig minne. Derfor er det noen enheter som benytter seg av *smart logging* der det bare logges når det er en endring i fart, retning eller hjerterytme (Garmin Ltd., 2012). Det gir dårligere detaljering men gir brukeren mulighet til å logge flere punkt før enheten blir full.

Utviklingen av håndholdte mottagere har ført til at det har blitt mer populært å benytte GPS til å registrere reisevaner, ettersom de er billige, lette og enkle i bruk. Ved å benytte GPS til datainnsamling får man observert reisevanen til deltagerne, og man kan se hvilke valg de gjør og dokumentere dette i de innsamlede dataene. Dette kalles for revealed preference-studie (RP), hvor observasjonene danner grunnlaget for å forklare de valgene som blir tatt. Andre metoder for å samle inn data til en revealed preference-studie kan være blåtann, WiFi, manuelle tellinger og observasjoner, samt videoopptak.

Stated preference-studier (SP) er studier hvor deltagerne forklarer, beskriver eller begrunner de valgene de tar. Ofte gjøres dette ved hjelp av spørreundersøkelser eller intervjuer med hypotetiske spørsmål, enten på internett eller skjema som sendes ut. Grunnen til at det skilles mellom de to metodene i litteraturkapittelet er fordi de har sine egne fordeler og ulemper. Mens revealed preference er hva vi gjør, er stated preference hva vi påstår at vi ville gjort, og det er ikke alltid det er samsvar mellom de to. Stated preference gir mulighet for å stille oppfølgingsspørsmål i intervjuer og god betenkningstid dersom spørreundersøkelsen gjennomføres på internett eller ved utfylling av skjema. Dessverre er det ikke alltid at

personer snakker helt sant når de svarer på spørsmål, og det kan være flere grunner til det (Meyer, 2013). Det krever også mye arbeid i å lage en god stated preference-undersøkelse (Hensher og Button, 2008). For revealed preference vil observasjonene alltid være «sanne», og det er mulig å foreta datainnsamlingen uten mye forarbeid. Dersom dataene lagres digitalt kan også mye av etterarbeidet utføres automatisk. På den andre siden kan det være unøyaktighet i dataene, brukerne kan taste feil eller glemme å skru av eller på enheten. Det kan føre til tunge og tidkrevende filtreringsprosesser i etterkant, før dataene kan benyttes i en analyse.

2.3.2 Viktige forklaringsfaktorer for rutevalg – Stated preference studier

Aultman-Hall et al. (1997) gjennomførte en studie i Guelph, Canada, basert på to spørreundersøkelser fra 1993. Spørreundersøkelsene inneholdt blant annet kart over byen hvor syklister kunne plote reiserutene sine, samt flere generelle spørsmål om bruk av sykkel. Det ble samlet inn totalt 1500 svar fra syklister og gående, men for den gjennomførte studien ble det kun analysert syklister som pendlet til jobb eller skole. Det førte til 338 respondenter og over 390 forskjellige ruter ble analysert i studien. Analysen ble gjennomført ved hjelp av et geografisk informasjonssystem (GIS)⁷, hvor nettverket inneholdt en rekke informasjon som lenketype, trafikkvolum, gradient, m.m.

Analysen viste at deltagerne valgte ruter som i gjennomsnitt var 0,4 km lengre enn den korteste ruten tilgjengelig. Bare 14,6 prosent av turene som ble gjennomført var på den ruten som faktisk var kortest. Ca. 45 % av den totale reiselengden ble gjennomført på hovedinnfartsårene (2-, 4- eller 6-felts veier) inn eller ut sentrum. Samleveger og gater stod for ca. 50 %, mens grusveger og terrengstier stod bare for 5 prosent. Sammenliknet med den korteste ruten tilgjengelig var det tydelig (statistisk signifikant) at syklister foretrakk å sykle på samleveger, selv om dette betydde lengre reise. Videre viste det seg at syklistene i stor grad benyttet seg av bilbroer, selv om det var tilgjengelig gang- og sykkelbro i nærheten: Av de 264 rutene som benyttet seg av en bro, var 208 av disse på bilbroer. Det blir ikke gitt noen forklaring på hvorfor syklistene benyttet seg såpass mye av bilbroene. Etersom studien fokuserte på syklister som pendlet til og fra jobb og skole kan syklistenes erfaring og et ønske om redusert reisetid være grunnen.

Studien indikerte også at syklister foretrakk ruter med mindre stigning og færre svinger, men benyttet seg av ruter med flere lyskryss enn hva den korteste ruten tilgjengelig hadde. Ruter med lyskryss ble spesielt benyttet i forbindelse med svingebevegelser, men dataene kunne ikke si noe om hvilke typer svingebevegelser (de var ikke detaljerte nok). Forfatterne påpekte også at syklistene unngikk gater hvor det var flere enn to bussruter, men argumenterer for at det ikke var bussene som var årsaken: I Guelph var det få områder som hadde mer enn to bussruter da spørreundersøkelsene ble gjennomført og disse områdene hadde høy aktivitet både av kjøretøy og myke trafikanter. Forklaringen kan derfor være at syklistene ønsket å unngå områder med flere forskjellige typer trafikanter og mange kompliserte bevegelser.

⁷ GIS er et integrert system som omfatter hardware, software og data som brukes til å prosessere, tolke og visualisere informasjon. I transport benyttes GIS ofte til å behandle informasjon om transportnettverk, bosettinger, arbeidsplasser, kollektivruter med mer. ArcGIS, GISLINE og Idrisi er noen eksempler på GIS.

Sener et al. (2009) gjennomførte en analyse av syklisters foretrukne reiserutevalg ved hjelp av en nettbasert spørreundersøkelse. Hovedfokuset var å undersøke hvordan parkering langs reiseruten påvirket rutevalget. Målgruppen for spørreundersøkelsen var syklistene i Texas, USA, og undersøkelsen ble sendt ut til sykkelgrupper, studenter på universiteter i Texas samt delt på nettbaserte forum for syklistene. De fikk også hjelp av lokale mediehus (tv-kanaler, aviser, magasiner m.m.) og departementer for transport og planlegging til å spre spørreundersøkelsen. Spørreundersøkelsen tok for seg elleve forskjellige egenskaper ved en reiserute dersom deltageren benyttet seg av sykkel til pendling, og ti forskjellige ruteegenskaper hvis deltageren drev med aktiviteter som ikke ble beskrevet som pendling (ærend, besøk, trening, shopping o.l.).

Hver respondent fikk fire caser der tre alternative ruter ble presentert. Respondenten måtte velge én av de tre. For å begrense arbeidet for respondenten var bare fem av ruteegenskapene presentert i hver undersøkelse, men parkerte biler langs ruten var alltid en av de fem ettersom det var hovedfokuset til studien. Totalt 1621 individ deltok i undersøkelsen, og disse var fordelt omtrent likt mellom pendlende syklistene (50,2 %) ⁸ og syklistene som kun brukte sykkelen til andre formål (49,8 %).

Resultatene fra spørreundersøkelsen viste at syklistene foretrakk ruter uten parkering langs kantsteinene. Der det var parkering var det bedre at biler stod parkert skrått inn mot kantsteinen enn parallelt med kantsteinen. Dette ga bedre oversikt, kortere tidsrom for konflikter og gjorde at syklistene følte seg tryggere. Syklistene foretrakk ruter hvor det var få parkerte biler, lav utskifting av bilene som stod parkert, samt at strekningen med parkering skulle være så kort som mulig. Det kom også frem at lange strekninger med moderate mengder parkerte biler var bedre enn en middels lang strekning med mange parkerte biler.

Videre viste analysen også at syklistene foretrakk å dele areal med annen trafikk i motsetning til egne, avgrensede gang- og sykkelveger. Forfatterne vurderte om dette kunne ha sammenheng med at de fleste deltagerne var erfarne og entusiastiske syklistene, som foretrakk å ha frihet til å bevege seg slik de selv ønsket i trafikken og hadde en form for «road warrior» mentalitet. Syklistene påpekte at et helhetlig sykkelnett var viktig, spesielt for pendlere, og kantsteinparkering langs sykkelnettet ga inntrykk av et u-kontinuerlig og ufullstendig sykkelnett. Høyt trafikkvolum, høye fartsgrenser og mange kryss (enten lysregulert, skiltet eller uregulert) var egenskaper som virket negativt på rutevalget til syklistene. Videre var

⁸ 98 % av syklistene som pendlet brukte også sykkelen til andre formål.

lettere kupert terreng å foretrekke fremfor flatt eller meget kupert terreng⁹, men dette var kun statistisk signifikant for ikke-pendlere.

Det viktigste for syklistene var likevel kort reisetid og lavt trafikkvolum. Kryssing av veger (stoppskilt, lyskryss eller uregulert), høye fartsgrenser, parkering langs ruten og u-kontinuerlige sykkelveger var egenskaper som var negative for en sykkelrute.

Stinson og Bhat (2003) gjennomførte en internetbasert spørreundersøkelse, der deltagerne måtte velge mellom to forskjellige ruter i hvert spørsmål. Rutene som ble presentert hadde tre til fire egenskaper listet opp, hvor noen av dem var forskjellige. Egenskapene var delt inn i segment-nivå (Link-level) og rute-nivå (Route-level). Spørreundersøkelsen ble distribuert gjennom nyhetsbrev og nettforum for syklistere over hele USA, og nesten 3 150 personer gjennomførte spørreundersøkelsen. Av disse var 91 % erfarne syklistere som ofte benyttet sykkel til pendling.

På segmentnivå viste analysen at syklistene foretrakk oppmerkede arealer langs vegbanen fremfor separerte sykkelveger og utvidede kjørebane uten oppmerking. På broer ønsket de imidlertid egne broer for myke trafikanter eller fysiske barrierer mellom kjøretøy og syklistere fremfor sykkelfelt. De foretrakk også å sykle langs adkomstveger fremfor samleveger og hovedveger. Videre var lettere kupert (Hilly) terreng bedre enn flatmark, mens meget kupert (Mountainous)⁹ var det minst foretrukne. De foretrakk også segmenter hvor det ikke var parkering for kjøretøy og asfalten holdt god standard. Blant egenskapene på rute-nivå var kontinuitet i sykkel fasiliteter en viktig faktor som telte positivt. Syklistene ønsket å unngå ruter med mange stoppskilt, røde lys og ruter som førte til de måtte krysse store tverrgater. Lav reisetid var også en viktig faktor.

⁹ Det ble ikke oppgitt noen verdier på hva som lå i de forskjellige begrepene.

2.3.3 Viktige forklaringsfaktorer for rutevalg – Revealed preference studier

Hood et al. (2011) gjennomførte et prosjekt i San Fransisco der en egen applikasjon for smarttelefoner ble utviklet til å samle inn GPS-data. Formålet med prosjektet var å utvikle en modell for rutevalget til syklistene. Innsamlingen av data begynte med at brukeren av applikasjonen valgte hensikt med reisen, eksempelvis pendling, skole, eller trening. Da startet også GPS-loggingen, og den fortsatte helt frem til brukeren enten kansellerte turen eller ga beskjed om at den var fullført. Det var også mulig å legge inn personlig data som kjønn, alder, adresse(r) og så videre, men dette var frivillig.

Totalt ble det logget 7096 turer, utført av 952 forskjellige brukere. For å forenkle analysen ble alle turer utenfor San Francisco City fjernet, samtidig som en *map matching algorithm* ble benyttet til å knytte GPS-dataene til et nettverk i et GIS. Det førte til at datagrunnlaget ble redusert til 3034 turer gjennomført av 366 forskjellige brukere. Av disse hadde ca. 55 % av turene pendling til og fra jobb eller skole som hensikt, mens private ærend var den nest største aktiviteten med 16 %. Videre kom sosialt, shopping og jobb-relatert med henholdsvis 10, 9 og 5 % av turene. Nettverket inneholdt data som lengde, trafikkvolum, hastigheter ved fri flyt, kvaliteten på sykkelvegene, kriminalitet i området, med mer.

Prosjektet viste at syklistene i San Francisco unngikk kraftige stigninger og svinger, men ønsket ikke å avvike for mye fra den korteste ruten tilgjengelig. Videre foretrakk syklistene i stor grad *bike lanes*, som var egne oppmerkede areal i vegbanen for syklistene. Bike Lanes i San Francisco er omkring 1,5 meter brede, og kjøretøy har kun lov til å bevege seg over dette området i forbindelse med svinger eller parkering. *Bike paths*, som er egne stier adskilt fra vegbanen, hovedsakelig i parker, langs elver, o.l., var også populært blant syklistene, men ikke i like stor grad som bike lanes. En mulig forklaring kan være at bike lanes også er for gående, og/eller at sykkelhastighetene kan være høyere på bike paths. Videre finnes det også *bike routes* i San Francisco, som er skiltede strekninger hvor syklistene deler areal i vegbanen med kjøretøy. Disse var klart mindre populære enn bike lanes og bike paths, men var likevel bedre enn ingen oppmerking. Analysen viste også at en stor andel enveiskjørte gater har negativ innvirkning på syklistenes rutevalg. De fant ingen sammenheng mellom rutevalg og kriminalitet, trafikkvolum eller hastighetene til andre kjøretøy.

Broach et al. (2012) utførte et liknende prosjekt i Portland, Oregon, USA med bakgrunn i data fra Dill og Gliebe (2008). Der ble 164 syklistene utstyrt med GPS for flere dager. Deltagerne plottet inn vær og reisehensikt ved begynnelsen av hver reise, samtidig som de hadde

muligheten til å indikere om de benyttet seg av et annet reisemiddel underveis, eksempelvis buss med sykkelstativ. Analysen ble gjennomført i et GIS, med et nettverk som inneholdt informasjon om sykkelfasiliteter, kryss, enveiskjørte gater med mer. Forfatterne valgte å fokusere på turer som innebar pendling, handling, ærend og liknende, så turer som var merket trening ble fjernet. Datagrunnlaget ble da redusert til 1449 turer, hvor ca. 30 % av disse var pendlerturer mellom hjem, jobb eller skole.

For pendleturer var den gjennomsnittlige avstanden 6 km, mens de andre turene hadde samlet sett en gjennomsnittlig lengde på 3,5 kilometer. Den gjennomsnittlige hastigheten var også høyere for pendlerturene sammenliknet med andre turer, henholdsvis 19 km/t mot 16,1 km/t. Om lag 50 % av turene ble utført på infrastruktur der det var tilrettelagt for sykkel, og mange syklistene unngikk hovedveger som ikke hadde sykkelfelt (Dill og Gliebe, 2008).

Også her viste analysen at syklistene unngikk svinger, bakker og stopp som skyldes lyskryss eller skilt. Likevel ønsket ikke syklistene å vike for langt unna den korteste tilgjengelige ruten. Å velge den korteste ruten var viktigere for pendlerturene enn for de andre turene, og studien tydet på at avviket fra korteste rute var relativt til lengden på turen. Lengre turer viste større avvik enn kortere turer, og i snitt var turene ca. 17 % lengre enn den korteste tilgjengelige ruten. Videre viste denne analysen at trafikkmengden påvirket syklistenes rutevalg dersom det ikke var egne arealer satt av til syklistene i vegbanen. Dette var spesielt tydelig dersom ÅDT oversteg 20 000 eller 30 000.

Menghini et al. (2010) gjorde det samme for Zurich i Sveits. Etter filtrering var det identifisert 2498 turer med unike start- og slutt punkt. Disse var overført til et nettverk i et GIS ved hjelp av map matching, hvor nettverket inneholdt informasjon om stigning, antall trafikkløys, oppmerkede sykkelareal med mer.

Analysen viste i dette tilfellet at nesten 36 % av turene var langs den korteste ruten. De aller fleste syklistene unngikk ruter med stigning, og over 40 % av de valgte rutene hadde en maksimal stigning som var mindre enn 1 %. Studien viste også at syklistene var mer sensitive til maksimal stigning enn gjennomsnittlig stigningen. Det var også høy bruk av oppmerkede sykkelruter: ca. 35 % av alle turene som ble gjennomført var på oppmerkede sykkelruter i sin helhet. Trafikkløys så ut til å ha liten effekt på rutevalget til syklistene, men kryss som ikke er regulert med trafikkløys eller rundkjøringer ble ikke studert.

Landis et al. (1997) gjennomført en studie der formålet var å utvikle en modell som beskrev kvaliteten på sykkelnettet, kalt *Bicycle Level-Of-Service* (BLOS). Level-of-service var derfor ikke et mål på trafikkavviklingen slik som i Amerikanske *Highway Capacity Manual* eller Statens Vegvesens *Håndbok 159 – Kapasitet på vegstrekninger*, men et mål på hvor brukervennlig sykkelstrekningen var (Landis et al., 1997) (Statens Vegvesen, 1991).

Studien baserte seg på en vurdering gjort av nesten 150 sykklister, der samtlige syklet den samme 27 kilometer lange ruten. Ruten var forhåndsbestemt og ble delt i 30 segmenter med omtrentlig samme lengde, men med forskjellige egenskaper. Blant annet varierte ÅDT fra 550 til 36 000, andel tunge kjøretøy fra 0 til 2 %, fartsgrenser fra 40 til 80 km/t, samt store forskjeller i infrastruktur og nærliggende bebyggelse. Deltagerne ble bedt om å gi hvert segment en vurdering fra A til F, basert på hvor trygge og komfortable de følte seg på det aktuelle segmentet, der A var best og F var dårligst.

Basert på disse resultatene ble det utviklet en modell for å beskrive BLOS. Etter å ha eliminert variabler som viste seg å være irrelevante endte de opp med en modell lik den vist i (7).

$$BLOS = a_1[f(V)] + a_2[f(S)] + a_3[f(M)] + a_4[f(X)] + a_5[f(p)] + a_6[f(W)] + C \quad (7)$$

Hvor

<i>BLOS</i>	=	Opplevd brukervennlighet
<i>V</i>	=	Trafikkvolum
<i>S</i>	=	Kjøretøyenes hastighet
<i>M</i>	=	Blandingen av trafikk
<i>X</i>	=	Potensiell trafikk på tvers av sykkelruten
<i>P</i>	=	Asfaltkvalitet
<i>W</i>	=	Tilgjengelig bredde for sykler
<i>C</i>	=	Konstant
a_1, a_2, \dots, a_6	=	Variabelkonstanter

Ettersom BLOS i (7) forteller noe om brukervennligheten langs en strekning, vil variabelkonstantene gi en god indikasjon på hvordan de forskjellige variablene påvirker den

opplevde nytten. Etter videre analyse av dataene ble den endelige BLOS-modellen utviklet, og den er gjengitt nedenfor.

$$\begin{aligned}
 BLOS = & 0,589 \ln\left(\frac{Vol_{15}}{L}\right) + 0,826 \ln[SPD_p(1 + \%HV)] & (8) \\
 & + 0,019 \ln(COM15 * NCA) + 6,406(PC_5)^{-2} \\
 & - 0,005(W_e)^2 - 1,579
 \end{aligned}$$

Hvor

<i>BLOS</i>	=	Opplevd brukervennlighet
<i>Vol₁₅</i>	=	Trafikkvolum over en 15-mintutters periode på samme strekning
<i>L</i>	=	Antall felt langs strekningen
<i>SPD_p</i>	=	Skiltet fartsgrense over strekningen
<i>%HV</i>	=	Andel tunge kjøretøy på strekningen
<i>COM15</i>	=	Turproduksjon-intensitet til områder nærliggende strekningen
<i>NCA</i>	=	Andel uregulerte innkjørsler på strekningen (også kantstein-parkering)
<i>PC₅</i>	=	Asfaltkvalitet etter Federal Highway Assosiations 5-poengs indeks langs strekningen, der 5 er best, 1 er dårligst.
<i>W_e</i>	=	Effektiv bredde på ytterst fil

Den endelige formelen for BLOS er ikke så lett å tolke på stående fot, men desto lavere BLOS-verdi, jo bedre er brukervennligheten. Fra (8) ser man at både økt trafikkvolum og høye hastigheter fører til høyere BLOS-verdi. Det samme gjelder for potensialet for krysstrafikk. Dårlig asfaltkvalitet ga dårlig BLOS, mens økt effektive bredde på ytterste fil bedre brukervennlighet.

Videre analyser viste at syklistene foretrakk at biltrafikken var adskilt fra syklistene ved hjelp av oppmerkede striper i asfalten fremfor ingen oppmerking. Sammenliknet med en nesten identisk strekning uten oppmerking ga syklistene i snitt 50 % bedre karakter til den strekningen som hadde oppmerking, selv om den hadde omtrent dobbelt så høyt trafikkvolum. Det kom også tydelig frem at meget dårlig asfalt ga stor reduksjon i opplevd komfort og sikkerhet, mens meget god asfalt bare ga en liten økning.

Harvey og Krizek (2007) gjennomførte en detaljert analyse av turene til 51 deltagere med hjelp av GPS, spørreundersøkelser og diskusjoner over tre uker i Minneapolis, USA. Dette er derfor en undersøkelse som benytter seg av både stated og revealed preference. Alle syklistene var vant til å pendle med sykkel, og gjorde så mer enn tre dager i uken. Dette var noen av kravene som måtte tilfredsstilles for å bli plukket ut til analysen. I tillegg ble det valgt ut omtrent like mange kvinner som menn og de var jevnt fordelt med tanke på alder. Det var også et krav at man jobbet eller studerte i sentrum av Minneapolis eller ved University of Minnesota, som også ligger nær sentrum. Det var fordi de ønsket at syklistene stor sett syklet i de samme områdene.

I den første uken syklet deltagerne den ruten de vanligvis brukte, mens i den andre uken var de nødt til å sykle en rute som ble tildelt dem av forskerne. Den tredje og siste uken kunne de velge hvilken rute de ville sykle. Deltagerne måtte i tillegg føre loggbok hvor alle turer skulle logges, hvor de logget blant annet den opplevde tryggheten langs ruten, hvor vidt de følte seg motivert for sykkelturen, og andre faktorer som kunne påvirke reisetiden (vind, grønn *bølge* i lyskryss). Fra de 51 deltagerne ble det samlet inn 852 turer hvor GSP-data og loggbok stemte overens. Gjennomsnittlig reise var ca. 10,4 kilometer med en gjennomsnittshastighet på ca. 20 kilometer per time.

Fra analysen kom det frem at syklistene først og fremst foretrakk ruter der de følte seg trygge, dernest kom ruter hvor det var komfortabelt og de følte seg selvsikre. Dedikerte sykkel fasiliteter med god kvalitet var det foretrukne rutevalget. Dårlig kvalitet på asfalt, komplekse kryss uten lysregulering, og områder hvor det lett oppstod konflikter med fotgjengere var faktorer som syklistene beskrev som negative.

2.3.4 Oppsummering av syklisters rutevalg

En oppsummering av fokus, metode og deltagere for den presenterte litteraturen er gitt i tabell 2.3. Videre er det gitt en oversikt over relevante variabler for syklisters rutevalg og hva litteraturen har funnet om de enkelte variablene i tabell 2.4. I tabell 2.4 betyr X at variabelen ble beregnet til å være statistisk signifikant, mens – (minus) betyr at det ikke er statistisk signifikant. Signifikans-nivået varierer fra studie til studie, men de fleste er på 0,05- eller 0,10-nivå med mindre noe annet er oppgitt. Videre betyr negativ at variabelen oppfattes som en kostnad og syklist vil vurdere å sykle andre steder for å unngå den nevnte variabelen. Er

variabelen positiv ønsker syklister å benytte seg av ruter som har den variabelen og den oppfattes som en nytte.

Tabell 2.4 viser at flere av studiene konkluderer med mye av det samme, og det finnes noen interessante detaljer og motsetninger:

- Stigning ser ut til å ha større negativ effekt på turer hvor formålet er pendling til jobb eller skole. Det kommer også frem fra at maks stigning og lengden på stigningen kan være viktigere enn gjennomsnittlig stigning. Det er også forskjell mellom kjønn. Menn virker mindre negative til stigning enn kvinner.
- Reisetid og avstand er viktige faktorer!
- Mindre veger (samleveger og gater) er bedre enn større og mindre hovedveger, og kontinuitet i sykkelfelt og sykkelveger er bra.
- Varierende resultat for kryss¹⁰. Lyskryss ser ut til å være bedre enn uregulert i forbindelse med svingebevegelser, men kryss og stoppskilt generelt påvirker negativt.
- Parkering langs sykkelruten oppfattes som negativt av syklistene. Skrå parkering inn mot kantstein er bedre enn parallell parkering langs kantstein. Lav aktivitet (utskifting, antall biler, inn- og utkjøring) og kort strekning gjør at parkeringen oppfattes mindre negativt.
- For å krysse elver og liknende hindringer foretrakk syklistene egne sykkelbroer. Dersom de må benytte samme bro som motoriserte kjøretøy var fysisk barrierer mer attraktivt enn sykkelfelt.
- En rute blir mer negativ dersom trafikkmengden øker, men dette er viktigst dersom det ikke er dedikerte sykkelfasiliteter tilgjengelig. Det er også negativt med høyt trafikkvolum i kryss, spesielt for den kryssende trafikken.
- Syklister foretrekker ruter med lavere fartsgrenser, men mer erfarne syklister er ikke påvirket i like stor grad.

¹⁰ Skyldes trolig detaljgraden til de forskjellige studiene. Noen har tatt alle former for kryss under en paraply, andre skiller mellom kryss etter trafikkmengde mens en tredje har delt mellom lysregulert og ikke-regulert.

Tabell 2.3 – Oversikt over den presenterte litteraturen.

Forfatter(e)	Tittel	Årstall	Metode	Spesielt Fokus	Deltagere
Aultman-Hall, Lisa	Analysis of Bicycle Commuter Routes Using Geographic Information Systems: Implications for Bicycle Planning	1997	Sporreundersøkelse	Generelt	338 personer. Sporreundersøkelsen ble sendt ut til tilfeldigvis personer, og måtte returneres av personen selv
Hall, Fred					
Baetz, Brian					
Sener, Ipek N.	An analysis of bicycle route choice preferences in Texas, US	2009	Sporreundersøkelse	Parkering	1621 personer. Mange erfarne syklistere.
Eluru, Naveen					
Bhat, Chandra R.					
Stinson, Monique A.	Commuter bicyclist route choice: analysis using a stated preference survey	2003	Sporreundersøkelse	Identifisere variabler på segment- og rutenivå	3145 deltagere. Hovedsakelig syklistere med erfaring som pendlere.
Bhat, Chandra R.					
Hood, Jeffrey	A GPS-based bicycle route choice model for San Francisco, California	2011	Applikasjon til smarttelefon med GPS	Generelt	366 personer. Tilgjengelig for alle med smarttelefon i San Francisco.
Sall, Elizabeth					
Charlton, Billy					
Broach, Joseph	Where do cyclists ride? A route choice model developed with revealed preference	2012	GPS	Utvikle rutevalgsmodell.	164 personer. Tilfeldig utvalg blant personer som meldte sin interesse.
Dill, Jennifer					
Gliebe, John					
Menghini, G.	Route choice of cyclists in Zurich	2010	GPS	Utvikle rutevalgsmodell	2438 personer. Originale data var for flere framkomstmidler, som ble filtrert ut til studien
Carrasco, N.					
Schüssler, N.					
Axhausen, K. W.					
Landis, Bruce	Real-Time Human Perceptions: Toward a Bicycle Level of Service	1997	Praktisk forsøk. Syklistere gjennomførte forhåndsbestemt rute.	Utvikle <i>Bicycle level-of-service</i> . Fokus på trygghetsfølelse og komfort.	Nesten 150 frivillige, fordelt på alder, kjønn og sykkel erfaring.
Vattikuti, Venkat					
Brannick, Michael					
Harvey, Francis J.	Commuter Bicyclist Behavior and Facility Disruption	2007	GPS, spørreundersøkelse og diskusjoner	Verdien av helhetlig sykkelnettverk.	51 deltagere. Spesielt utvalgt med hensyn på kjønn, alder og hjem- og arbeidssted.
Krizek, Kevin J.					

Tabell 2.4 – Oversikt over variabelers innvirkning på rutevalg presentert i litteraturen

Forfatter(e)	Titel	Årstall	Stigning	Infrastruktur*	Kryss	Parkering	Tiltak på bilbroer	Trafikkvolum**	Fart eller fartsgrenser	Reisetid eller avstand	Vegtype
Auhlmann-Hall, Lisa	Analysis of Bicycle Commuter Routes Using Geographic Information Systems: Implications for Bicycle Planning	1997	X	X	X	Ikke testet	-	Ikke testet	Ikke testet	Ikke testet	Ikke testet
Hall, Fred			Negativ	Positiv for samleveger	Negativ for antall svinger per kilometer, positiv for antall lyskryss.		Ser ikke ut til å påvirke rutevalget				
Baetz, Brian											
Sener, Ipek N.	An analysis of bicycle route choice preferences in Texas, US	2009	X	X	X	X	Ikke testet	X	X	X	Ikke testet
Eluru, Naveen			Hovedsakelig negativ.	Positiv for kontinuitet	Negativ for antall kryss (Alle typer)	Negativ for lengde, utskifting, og belegg samt parallell og skråparkering		Negativ. Øker i takt med lengden på turen og diskontinuitet i infrastrukturen	Negativ. Øker i takt med økte hastigheter	Reisetid. Negativ, men meget lav verdi	Ikke testet
Bhat, Chandra R.			Positiv for ikke-pendlere og menn								
Stinson, Monique A.	Commuter bicyclist route choice: analysis using a stated preference survey	2003	X	X	X	X	X	Ikke testet	Ikke testet	X	X
Bhat, Chandra R.			Negativ for bratte stigninger (Mountainous)	Positiv for sykkelfelt og fysisk barrierer.	Negativ for antall stoppskilt på en strekning	Negativ for parallell parkering	Positiv for alle type broer som er tilrettelagt for syklistar.			Reisetid. Negativ, høy verdi	Negativ for større og mindre hovedinnsatsår
Hood, Jeffrey	A GPS-based bicycle route choice model for San Francisco, California	2011	X	X	X	Ikke testet	Ikke testet	Ikke testet	Ikke testet	X	X
Sall, Elizabeth			Negativ. Større for pendlere	Positiv. Større for sykkelfelt enn sykkelsti	Negativ for antall svinger per kilometer.					Avstand. Negativ. Middels verdi.	Negativ for enveiskjørt gater. Andre typer ble ikke testet
Charlton, Billy											
Broach, Joseph	Where do cyclists ride? A route choice model developed with revealed preference GPS data	2012	X	X	X	Ikke testet	X	X	Ikke testet	X	Ikke testet
Dill, Jennifer			Negativ. Større for pendlere.	Positiv. Større for sykkelsti enn for bike boulevard***	Negativ. Øker med trafikkmengde.		Positiv. Separert sykkelsti bedre enn sykkelfelt	Negativ. Øker med mengden.		Avstand. Negativ. Lavere for pendlere.	
Gliebe, John											
Menghini, G.	Route choice of cyclists in Zurich	2010	X	X	X	Ikke testet	Ikke testet	Ikke testet	Ikke testet	X	Ikke testet
Carrasco, N.			Negativ. Maks stigning og lengde viktigere enn gjennomsnittlig stigning.	Positiv. Gjelder alle typer.	Positiv, men lav verdi. Gjelder trafikklys.					Avstand. Negativ.	
Schüssler, N.											
Axhausen, K. W.											
Landis, Bruce	Real-Time Human Perceptions: Toward a Bicycle Level of Service	1997	Ikke testet	X	Ikke testet	Ikke testet	Ikke testet	X	X	Ikke testet	Ikke testet
Vattikuti, Venkat				Positiv. Gjelder asfaltkvalitet og tilgjengelig bredde langs vegen for syklistar				Negativ for mengde og for potensiell krystrafikk.	Negativ. Øker dersom det er flere tunge kjøretøy		
Brannick, Michael											
Harvey, Francis J.	Commuter Bicyclist Behavior and Facility Disruption	2007	Ingen verdier presentert	Ingen verdier presentert	Ingen verdier presentert	Ingen verdier presentert	Ingen verdier presentert	Ingen verdier presentert	Ingen verdier presentert	Ingen verdier presentert	Ingen verdier presentert
Krizek, Kevin J.											

X = Statistisk signifikant
 - = Ikke signifikant

- Sykkelfelt, sykkelveg, etc. Gjelder ikke kryss og rundkjøringer
- ** Motoriserte kjøretøy
- *** Mindre gater der det er gjennomført trafikkberoligende tiltak.

2.4 Oppsummering

Hvordan man modellerer rutevalget er en grunnstein i 4-steg metodikken, og endring i rutevalg kan påvirke både fordelingen av turer og reisemidler. Store forskjeller mellom måten man modellerer rutevalget og hvordan rutevalget skjer i virkeligheten kan gi store forskjeller mellom modellert og observert sykkeltrafikk.

Mange av de tilgjengelige ruteplanleggere på nett gir lite dokumentasjon om hvordan sykkelfarten beregnes. Flere benytter seg av en standardfart som kan påvirkes av noen variabler, men det er ofte få forklaringer både på hvorfor og hvordan dette gjøres. Hastighetene på flat mark ser ut til å ligge i området 15-20 km/t for de fleste planleggerne. Modellene er bedre dokumentert, men RTM benytter seg ikke av fart før man kommer til trafikantnytteberegningene, og der er farten konstant, nemlig 15 km/t.

Det er gjennomført flere forsøk for å forklare syklisters rutevalg, og det er benyttet GPS i noen av disse. Stigning, parkering, fartsgrenser og trafikkvolum er noen faktorer som påvirker rutevalget negativt. Kryss kan både være negativt og positivt avhengig av trafikkmengde på den aktuelle strekningen og svingebevegelsene som skal gjennomføres av syklisten. For infrastruktur ser sykkelveger og egne sykkelfelt ut til å være attraktivt, spesielt for erfarne syklistere.

Det er også tydelig at reisetid er en viktig faktor. Syklistere velger ikke nødvendigvis den aller raskeste ruten, men det pleier som regel å være en av de viktigste faktorene for rutevalget. Reisetid kan være avhengig av flere faktorer, men det antas at stigning vil være en direkte indikator på farten til syklistene.

3 Metode og fremgangsmåte

Med bakgrunn i problemstillingen og den fremstilte litteraturen ble det valgt å legge vekt på stigningens betydning for sykkelfart, ettersom både reisetid og stigning viste seg å være viktige forklaringsfaktorer for syklisters rutevalg. Det videre arbeidet ble delt i to, og sammen med litteratursøket vil disse punktene gi et godt grunnlag for å vurdere bruken av GPS-data i utviklingen av en fartsmodell og en rutevalgmodell:

1. Identifisering, vurdering og innhenting av data fra offentlig tilgjengelige kilder.
2. Oppbygging og testing av fartsmodell.

Det ble derfor naturlig å dele dette kapittelet inn i to deler. Den første delen forklarer fremgangsmåten for innhenting av data mens andre del tar for seg arbeidet med å utvikle en fartsmodell.

3.1 Kartlegging, vurdering og datainnsamling

Det var ønskelig å vurdere hvilket potensial offentlig tilgjengelig GPS-data hadde i forbindelse med modellering av sykkeltrafikk. Første steg i arbeidet ble derfor å redegjøre for flest mulig potensielle kilder til GPS-data. Flertallet av sidene var sosiale nettsted med formål for å dele og utforske både ruter og treningsøkter, og derfor en helt annen hensikt enn denne studien. Kapittel 3.1.1 tar for seg arbeidet rundt datakildene mens kapittel 3.1.2 forklarer prosessen med tidlig filtrering og innhenting av data. Parsing, segmentering og mer filtrering er beskrevet i kapittel 3.1.3 og i kapittel 3.1.4 presenteres den siste runden med filtrering og deling av datasettet før analysen.

3.1.1 Datakilder

Det var ønskelig å finne kilder som både hadde rikelig med data for norske byer, muligheter for å benytte en form for kildekode til å effektivisere datainnsamlingen, og hadde stor andel turer hvor pendling eller transport var formålet.

Innledningsvis var det lagt vekt på å få tak i turer i områdene Kristiansand og Trondheim på grunn av muligheten til å supplere med lokal kunnskap og observasjoner under arbeidet. Det ble etter hvert bestemt å utvide området ved å inkludere Oslo for å øke datagrunnlaget. Etter hvert som kilder ble kartlagt, ble det også nødvendig å se på hvilket format dataene var tilgjengelig på og hvorvidt datasettet inneholdt nok informasjon til å kunne benyttes videre i

analysen. Formatene «GPX», «TCX» og «CSV»¹¹ viste seg å være blant de vanligste, men informasjonsmengden varierte mellom de tre formatene. Mens TCX-filer var innholdsrike i noen tilfeller, var GPX- og CSV-filer ofte mer sparsomme. Flere av de potensielle kildene tilbød bare GPX- eller CSV-filer hvor fartsdata eller tidsverdier ikke var oppgitt. Dersom fart og tid ikke var oppgitt ble kilden vurdert som ikke brukbar til utvikling av en fartsmodell. Tabell 3.1 gir en oversikt over hvilke nettsteder som ble vurdert, samt tilgjengelige turer, eksportmuligheter og en total vurdering.

¹¹ GPX = GPS exchange format, TCX = Training Center XML, CSV = comma separated values. Forskjellige metoder for å representere data. GPX og TCX benytter seg av en XML-struktur (Extensive Markup Language), mens CSV er et format der verdiene er separert ved hjelp av komma. Mens GPX og TCX bør konverteres til andre format for videre arbeid, kan CSV gjerne brukes direkte.

Tabell 3.1 – Oversikt over mulige kilder til GPS-data.

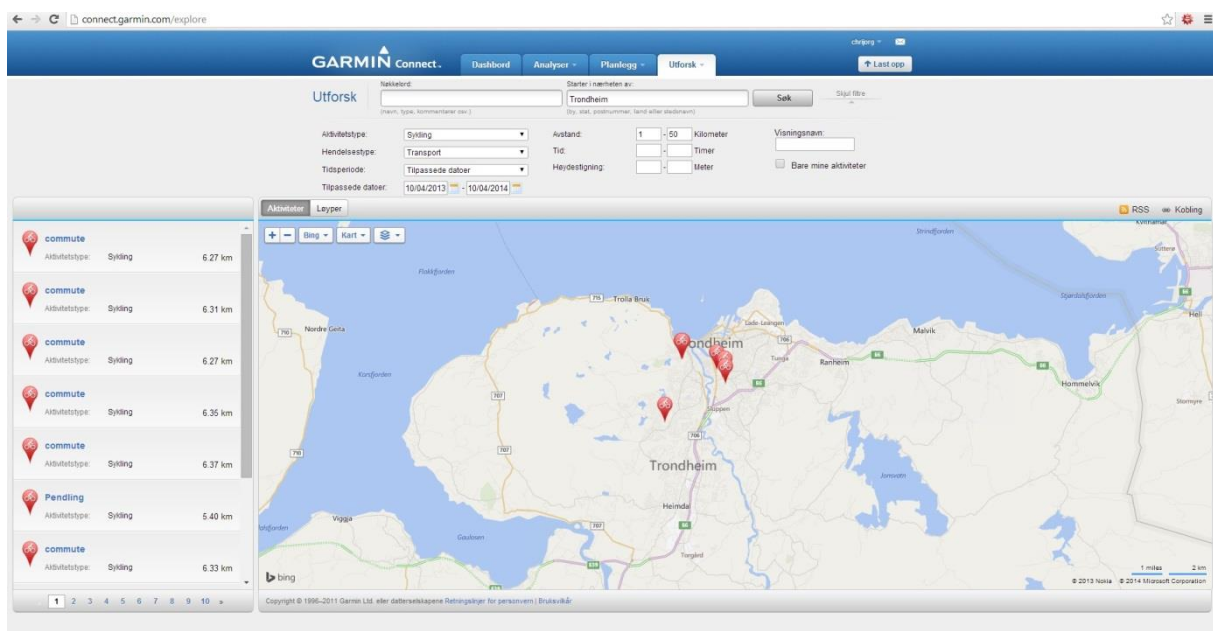
Navn	Nettsted	Tilgjengelig ruter	Eksportmuligheter	Vurdering
Strava	http://www.strava.com/	Mye i Trondheim og Oslo, lite i Kristiansand. Hovedsakelig treningsturer.	GPX-format, men krever premium-konto.	Lavt potensial, usikkerhet på tilgjengelig informasjon i GPX fil.
MapMyRide	http://www.mapmyride.com/	Mange pendle/transport-turer, over hele landet.	GPX-format med lengde- og breddegrad, CSV-format med kumulativ avstand og høyde. Ingen fart, men total tid er tilgjengelig på nettsiden.	Middels potensial. Usikkerhet uten oppgitt fart.
RunKeeper	http://runkeeper.com/index	Flere pendler-turer, både i Oslo, Kristiansand og Trondheim	Ingen eksportmulighet funnet.	Ikke brukbar.
Garmin Connect	http://connect.garmin.com/en-US/	Noen i Kristiansand og Trondheim, mange i Oslo	GPX- og TCX format, meget detaljrike.	Stort potensial
RideWithGPS	http://ridewithgps.com/	Få ruter, hovedsakelig trening. Må skilles manuelt mellom trening eller transport, ingen gode avgrensingsmuligheter på søket.	TCX som inneholder lengdegrad, breddegrad, høyde og tidspunkt.	Lavt potensial. Far er ikke oppgitt, me kan bereges. For få ruter i pendle/transport kategorien.
DailyMile	http://www.dailymile.com/	Hovedsakelig treningsturer, flere transportmidler, og dårlig søkefunksjon. Kan ikke sette avgrensinger på søket.	GPX-format med lengde- og breddegrad. Ikke fart eller tid.	Ikke brukbar.
Runtastic	https://www.runtastic.com/	Få tilgjengelige ruter, mest treningsturer.	GPX-og TCX-format, men ikke tidsstempel eller fart.	Ikke brukbar.
Endomondo	http://www.endomondo.com/	Mest treningsturer, for det meste i/rundt Oslo. Tungvint grensesnitt for søk.	GPX-format uten fart eller tidsstempel.	Ikke brukbar.
Cyclemeter	https://abvio.com/cyclemeter/	Data kun tilgjengelig gjennom en applikasjon på iPhone	GPX/TCX/CSV, usikkerhet rundt detaljgraden.	Ikke brukbar.
NikePlus	http://nikeplus.nike.com/	Ingen relevante ruter.	Ingen eksportmulighet	Ikke brukbar.
2peak	http://2peak.com/index.php	Ingen mulighet for å søke opp ruter. Månedlig abonnement nødvendig.	Ingen eksportmulighet funnet.	Ikke brukbar.
TrainingPeaks	http://home.trainingpeaks.com/	Få turer på sykkel, mest treningsturer.	GPX og TCX, men krever premium medlemskap.	Ikke brukbar.

Kildene *MapMyRide* og *Garmin Connect* kom ut som de beste kandidatene og ble tatt med videre i arbeidet. *MapMyRide* hadde ikke oppgitt fart i dataene som ble lastet ned, men det var en mulighet for at denne kunne beregnes ved hjelp av de tilhørende dataene for kumulativ avstand og høyde som var tilgjengelig på CSV-format. I tillegg var det oppgitt total reisetid på nettsiden. Ved å anta et fast intervall mellom hvert målepunkt i dataene og sammenlikne dette med den oppgitte tidsbruken kunne man til en viss grad beregne en gjennomsnittshastighet

mellom hvert målepunkt. Det viste seg at denne antagelsen ble for svak. Ettersom forskjellige GPS-mottakere kan ha forskjellige målefrekvenser ble det for store sprik i fartsverdiene, og usikkerheten ble vurdert til å være for stor. Derfor ble data fra MapMyRide ekskludert fra analysen, og Garmin Connect stod igjen som eneste kilde. Data som var tilgjengelig på Garmin Connect var meget detaljert, som i tillegg til lengdegrad, breddegrad, høyde, tid og fart inneholdt informasjon som hjerterytme, aktivitet, GPS-modell, og tilbakelagt avstand.

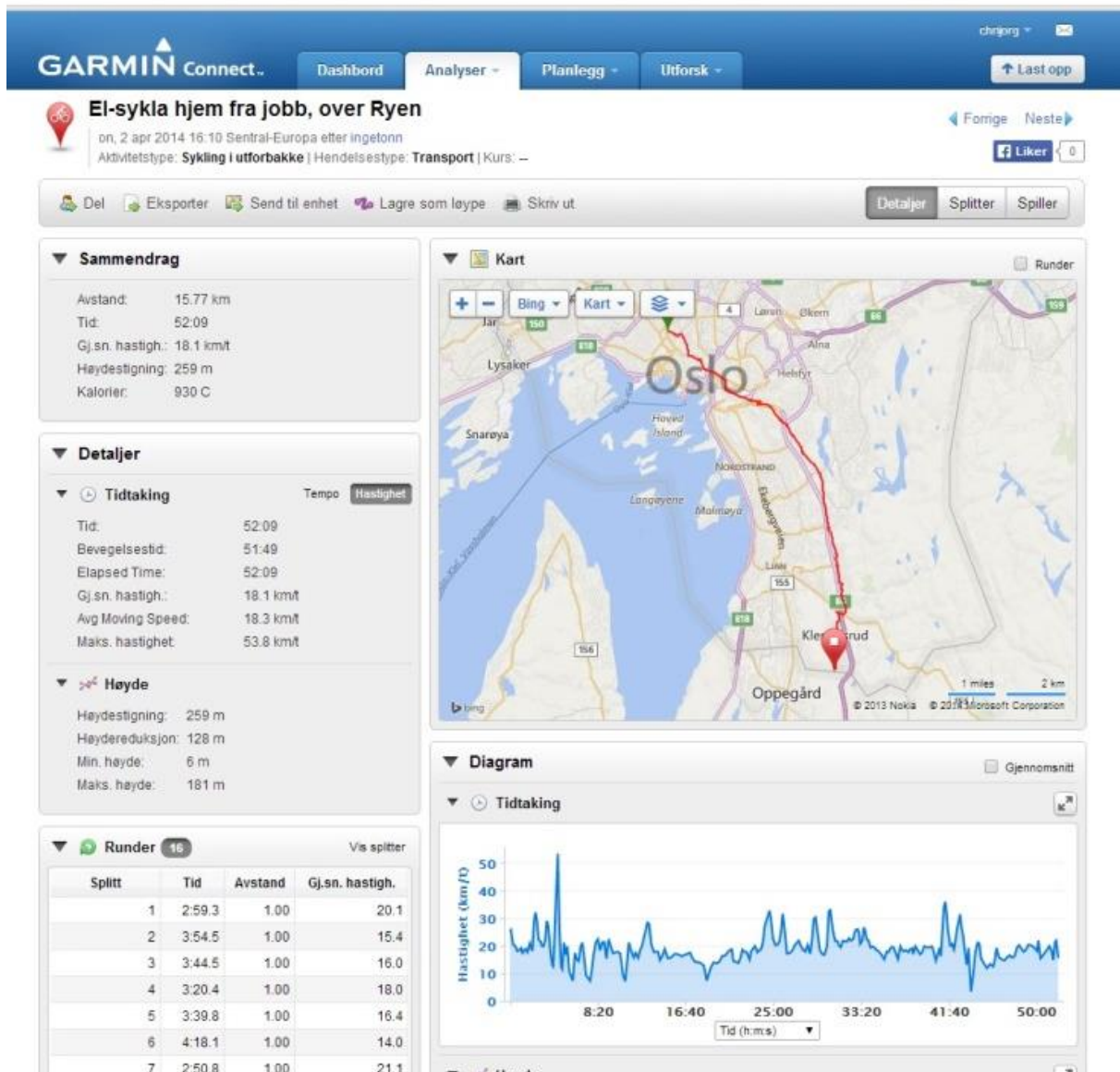
3.1.2 Tidlig filtrering og innhenting av data.

Allerede i den første fasen av datainnsamlingen ble uønskede turer filtrert bort. Det var ønskelig å få turer som representerte typiske transport- eller pendlerturer, som turer mellom jobb, skole, barnehage, butikk og liknende. Garmin Connect hadde et meget oversiktlig og rent brukergrensesnitt som gjorde søk og avgrensning relativt lett. Figur 3.1 viser en skjermdump fra Garmin Connect og er et eksempel på hvordan søket ble utført. Det var flere valg for både aktivitetstype og hendelsestype, men den metoden som ga flest akseptable turer var kombinasjonen «sykkel» som aktivitet og «transport» som hendelse. Det ble også satt en nedre grense på 1 kilometer og en øvre grense på 50 kilometer, da det ble antatt at de fleste pendlerturer lå i dette intervallet. Videre ble det satt samlet inn turer mellom 10. April 2013 og 10. April 2014, altså ett helt år med data. Dette ble gjort bare for å ha en oversikt over hvilke data som var lastet ned, samt at man sikret seg en viss reserve dersom det skulle bli nødvendig å utvide datasettet.



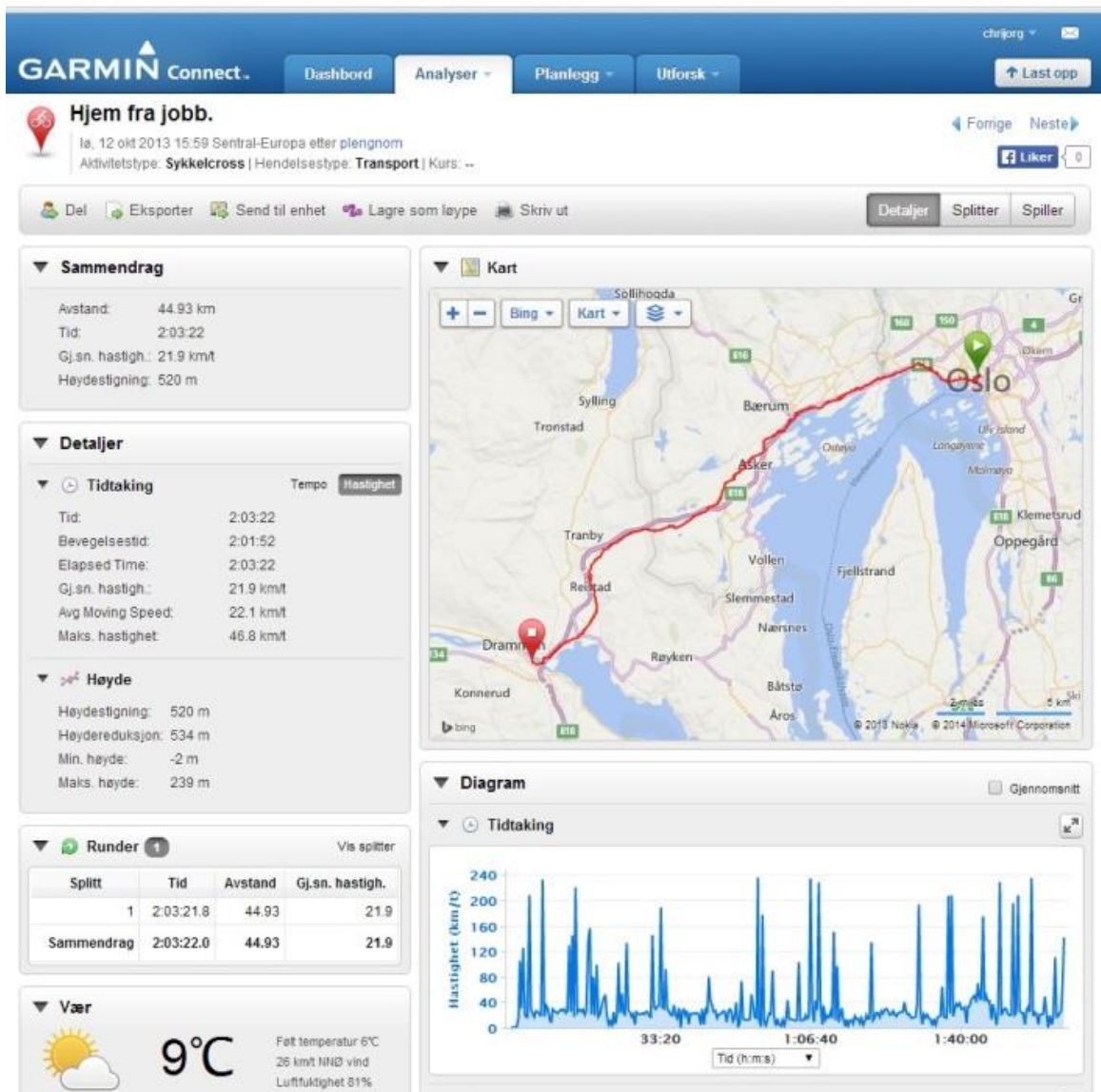
Figur 3.1 – Skjermdump fra Garmin Connect.

Det ble også gjennomført en visuell kontroll før en tur ble lastet ned. Eksempelvis ble data som så ut som en «typisk» treningstur¹², hadde urimelige hastigheter eller høydedata eller navn som hentydet at dette ikke var en pendletur ikke lastet ned. Figur 3.2 til figur 3.4 gir noen eksempler på data som ble filtrert bort.

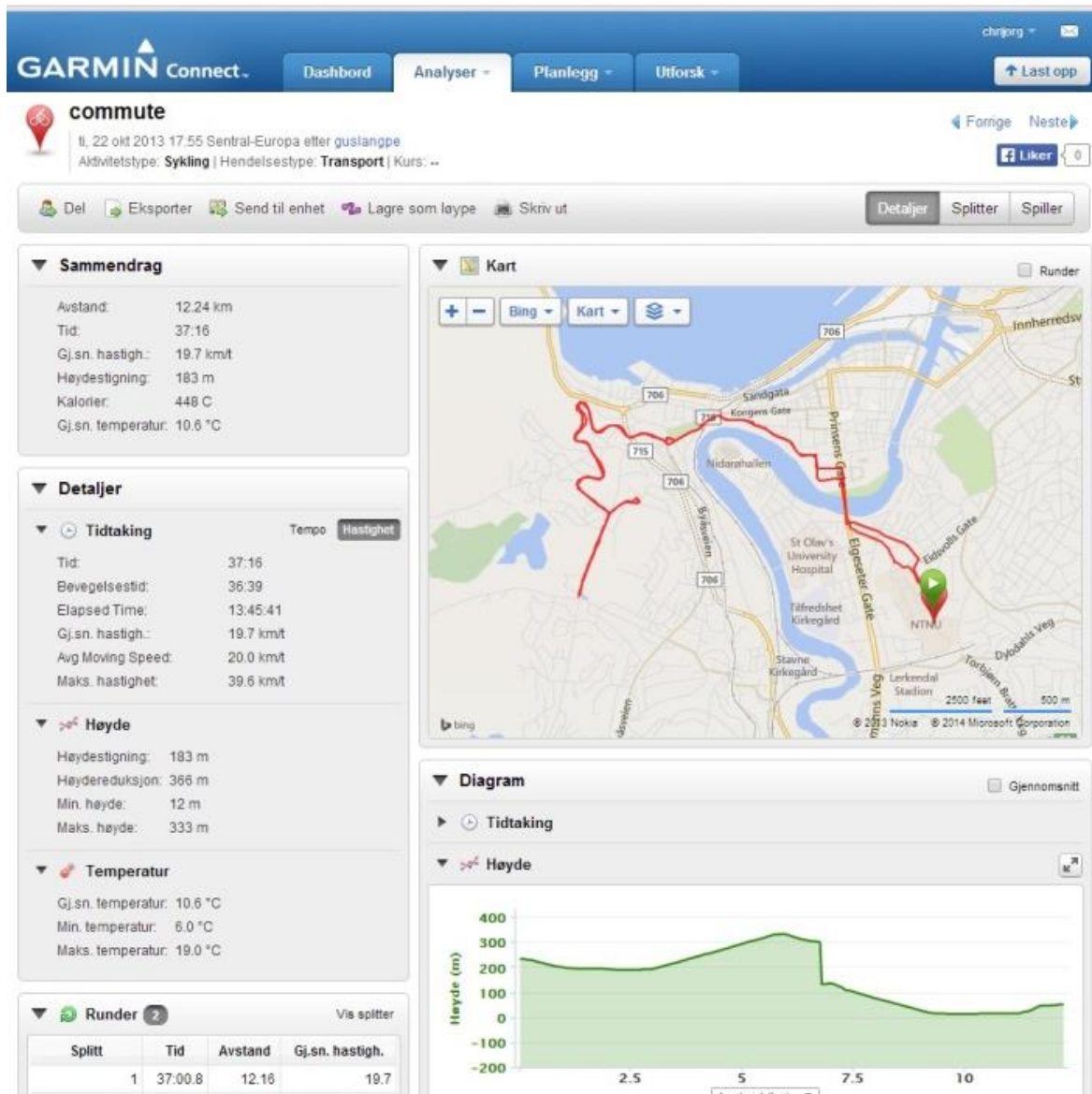


Figur 3.2 – Filtrert data på grunn av navn . (EL-sykel regnes ikke som representativt i dette tilfellet.)

¹² Startet og stoppet på samme sted, opprinnelse eller destinasjon utenfor bebygget område, o.l.



Figur 3.3 – Eksempel på filtrering på grunn av dårlig fartsdata . Ujevn profil med meget høye hastigheter.



Figur 3.4 – Eksempel på filtrering på grunn av dårlig høydedata . Stort hopp rundt 8,6 km.

For å få på plass et stort grunnlag med data ble det laget et skript i programmeringsspråket *Python* som gjorde innsamlingen lettere. Skriptet ble matet med den direkte nettadresse til en ønsket tur og kopierte deretter turdataene automatisk. I tillegg ble nettadressen og en identifikasjonsvariabel mellom nettadressen og de nedlastede turdataene lagret, slik at man hadde en tilbakekobling til de «originale» dataene dersom dette skulle være nødvendig. Etter innledende filtrering og innsamling var det totale grunnlaget på 3868 gjennomførte turer fordelt over 3 911 188 registrerte målepunkt.

3.1.3 Parsing, utregninger og videre filtreringer.

Dataene som ble innsamlet var på TCX-format som ikke var spesielt brukervennlig i andre applikasjoner. Derfor ble det benyttet et script som leste inn alle TCX-filene og konverterte dem til det mer utbredte CSV-formatet, som blant annet fungerer godt med statistikkprogrammet SPSS fra IBM og regneprogrammet Excel fra Microsoft. Metoden kalles ofte for *parsing*¹³. I tillegg manglet dataene noen biter med informasjon som kunne være essensiell for videre analyse, men på grunn av den gode detaljeringsgraden var det mulig å beregne disse med lav usikkerhet. Hovedsakelig dreiet det seg om avstand og stigning mellom hvert målepunkt, som ble beregnet ved hjelp av formel (9) og (10). Avstand ble beregnet i samme script som parset TCX-filene, mens stigningen ble beregnet rett i datasettet ved hjelp av SPSS. I tillegg var fart oppgitt i meter per sekund. Denne ble konvertert til kilometer per time.

$$L_{p_n-p_{n-1}} = L_{p_n} - L_{p_{n-1}} \quad (9)$$

Hvor

$$\begin{aligned} L_{p_n-p_{n-1}} &= \text{Avstand mellom målepunkt n og målepunkt n-1} \\ L_{p_n} &= \text{Akkumulert avstand fra start til punkt n} \\ L_{p_{n-1}} &= \text{Akkumulert avstand fra start til punkt n-1} \end{aligned}$$

$$S_{p_n-p_{n-1}} = \frac{H_{p_n} - H_{p_{n-1}}}{L_{p_n-p_{n-1}}} * 100\% \quad (10)$$

Der

$$\begin{aligned} S_{p_n-p_{n-1}} &= \text{Stigning mellom målepunkt n og målepunkt n-1 [\%]} \\ H_{p_n} &= \text{Høyden i punkt n} \\ H_{p_{n-1}} &= \text{Høyden i punkt n-1} \\ L_{p_n-p_{n-1}} &= \text{Avstand mellom målepunkt n og målepunkt n-1} \end{aligned}$$

Etter at dataene var parset og beregningene fullført ble det tydelig at ytterligere filtreringer var nødvendig. Blant annet var det fortsatt hastigheter som var urimelig høye og høydedata som ga liten mening. Derfor ble det i første omgang fjernet alle hastigheter over 70 kilometer per

¹³ Parse: «Analyse», Parser: «Analysealgoritme» (Clue Dictionary Software, 2011) Metode for å hente ut informasjon fra data som er strukturert på en spesiell måte.

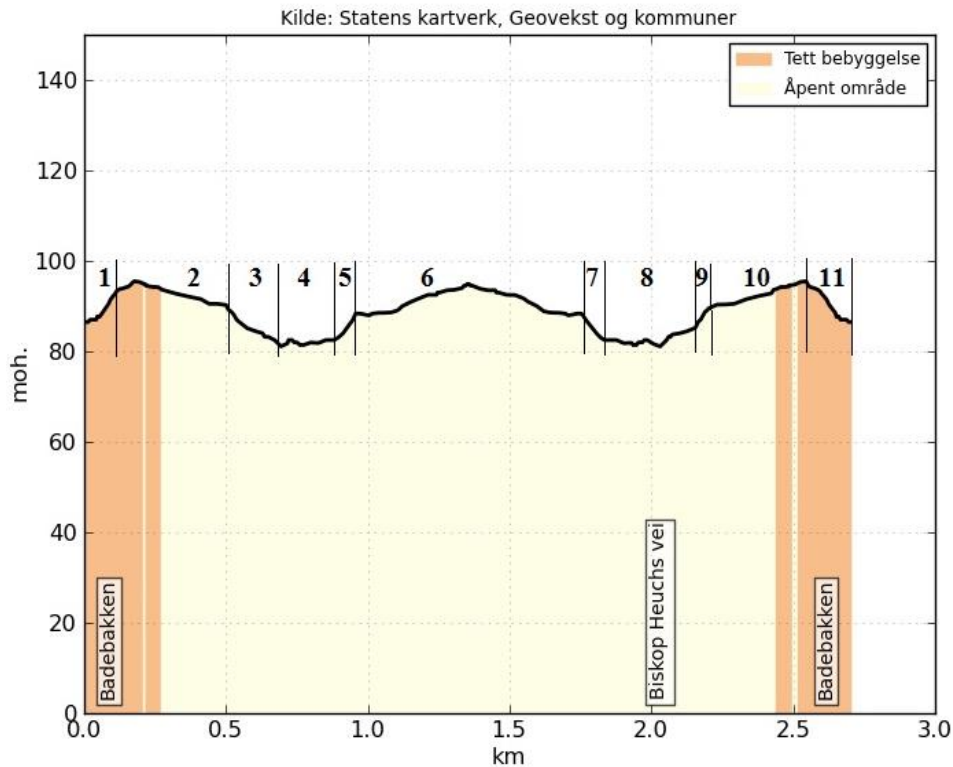
time og alle stigninger brattere enn $\pm 35\%$. Verdier som havnet utenfor disse områdene ble ansett som urimelige og ikke representative, og ble fjernet fra datasettet.

Etter to runder med filtrering ble datasettet delt opp i segmenter avhengig av stigningen og det ble definert tre segmentgrupper. Det ble antatt alle stigninger i området $\pm 2,0\%$ oppfattes som tilnærmet flate strekninger, mens stigninger utenfor dette området oppfattes som oppoverbakke eller nedoverbakke:

1. Nedoverbakker: Stigning $< -2,0\%$
2. Tilnærmet flatt: $-2,0\% \leq \text{Stigning} \leq 2,0\%$
3. Oppoverbakker: Stigning $> 2,0\%$

Figur 3.5 viser prinsippet, men er kun ment som illustrasjon. Segment 1 i figuren består av den første, lange bakken som er brattere enn 2% , og blir derfor klassifisert som oppoverbakke. Segment 2 består av flate elementer med flere mindre bakker som har en stigning som er mindre eller lik $\pm 2\%$ og blir klassifisert som tilnærmet flatt. Segment 3 er en nedoverbakke med stigning som er mindre enn $-2,0\%$ og er derfor definert som en nedoverbakke. En slik segmentinndeling ble gjennomført for hver tur i datasettet.

Grunnen til denne segmenteringen var i første omgang for å få inn variabler som beskrev maksimal og gjennomsnittlig stigning i en bakke, samt en variabel på stigningens lengde. Segmenteringen ble gjennomført i et skript ved hjelp av Python.



Figur 3.5 – Illustrasjon over segmentinndeling for en tilfeldig tur . (Statens kartverk, 2013)

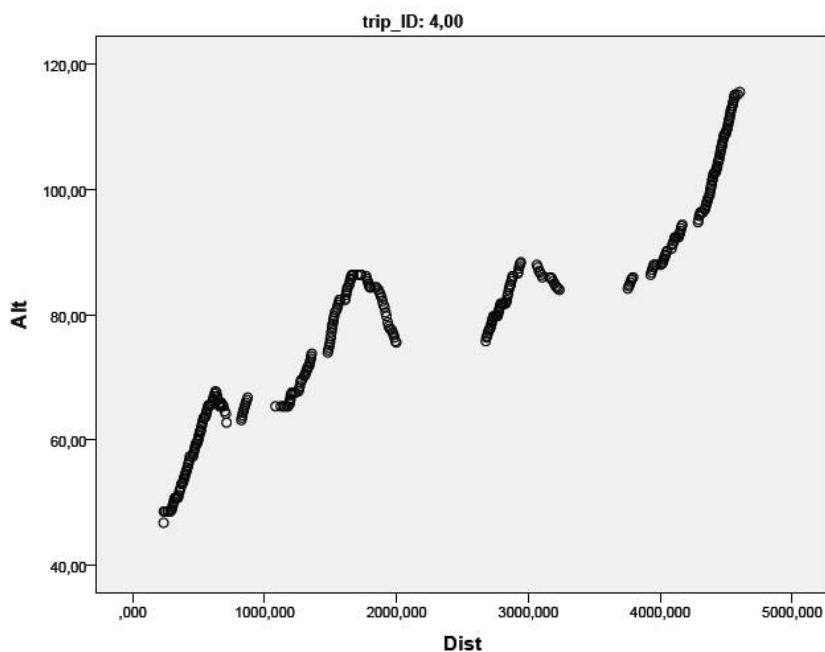
Tross de to første rundene med filtrering viste det seg at datasettet fortsatt inneholdt noen «merkelige» data. Statistiske analyser av stigningen viste et stort flertall av tilfeller hvor stigningen ble beregnet til 0,0 %, altså at det var ingen endring i høyde mellom to målepunkt. Ved å hente data fra Statens vegvesens *Vegkart* kunne man sammenlikne lengden på stigningene. Det ble påvist at flate strekninger i *Vegkart* hadde en lengde som var statistisk signifikant kortere enn 100 meter, se vedlegg B. Derfor ble alle segment med en stigning på 0,0 % og lengde større enn 100 meter antatt til å være urimelige og fjernet fra datasettet.

Det ble også fjernet de første og de siste 100 meterne fra hver tur. Antagelsen var at farten i begynnelsen og slutten av hver tur var lavere enn hva den var midt i en tur. Grunnene til dette var at mange brukte tid på å komme seg ordentlig i gang og oppnå realistiske hastigheter etter at de hadde startet dataloggingen. Det samme gjelder de siste 100 meterne. Man sakker av litt fordi man nærmer seg, man triller kanskje sykkelen litt eller dataloggingen registrerer bevegelse selv om man egentlig låser fast sykkelen eller liknende. Både gjennomsnittlig fart og varians i fart var statistisk forskjellig for de første og de siste 100 meterne enn resten av dataene, som vist i vedlegg C. Det førte til at det ble fjernet ytterligere 71 unike turer og

698 280 målepunkt i den filtreringsrunden. Det utgjorde ca. 2 % av turene og 18 % av målepunktene som var igjen fra forrige filtrering.

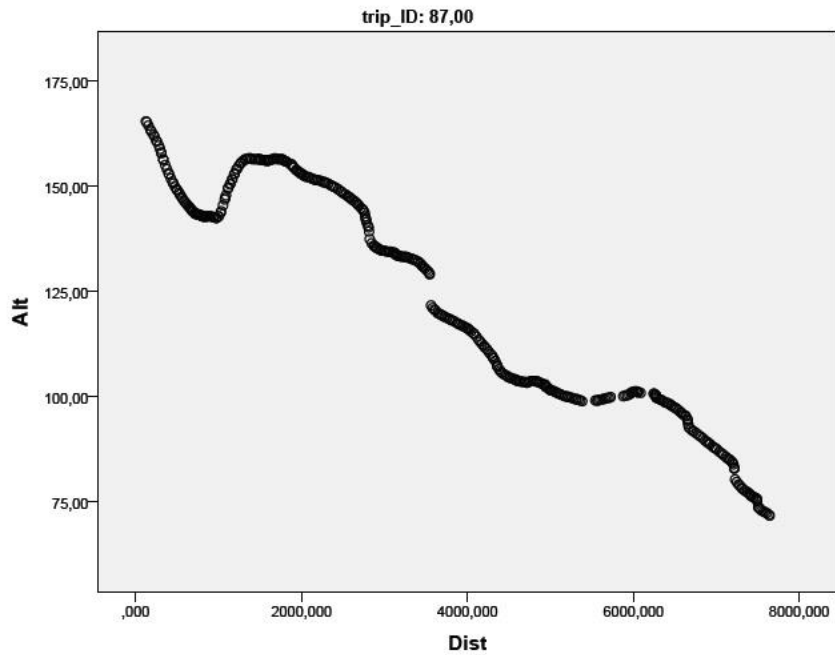
3.1.4 Manuell filtrering, værdata og utvalg av turer

Det viste seg at noe av den innledende filtreringen var noe uheldig. Det var flere tilfeller med rare data der det var store hopp i høyde eller urimelige 0-verdier, som delvis kan skyldes originale GPS-data og delvis den innledende filtreringen der flere målepunkt ble slettet. Derfor ble en siste runde med filtrering gjennomført manuelt. Høydeprofil for alle turene ble plottet ved hjelp av SPSS og gjennomgått. Der ble hver tur klassifisert i en gruppe avhengig om det skulle fjernes helt, beholdes eller bare hadde mindre mangler. Turer som hadde meget store mangler i data ble fjernet helt, og et eksempel på en slik tur er vist i figur 3.6.



Figur 3.6 – Eksempel på tur som ble fjernet fra datasettet . trip_ID er unik ID for akkurat denne turen, Alt står for høyde mens Dist er tilbakelagt distanse.

Turer som hadde mindre mangler kunne eventuelt repareres dersom det var behov for mer data, og ble separert ut til et eget datasett. Fra figur 3.7 ser man at det er bare mindre mangler i høydeprofilen, og disse kan enten repareres ved bruk av nærliggende målepunkt eller en alternativ kilde til høydedata skulle det bli nødvendig med flere turer og målepunkt.



Figur 3.7 – Eksempel på tur som kanskje kan repareres . trip_ID er unik ID for denne turen, Alt er høyde og Dist er tilbakelagt distanse.

Det ble også hentet ned værdata fra Meteorologisk Institutt (MET) for de tre byene, som ble koblet opp mot datasettet. Værdataene inneholdt døgnverdier for temperatur og nedbør¹⁴ fra en eller to stasjoner i eller nær sentrum for byene.

Etter den manuelle filtreringen var det fortsatt over to millioner målepunkt og 2 000 unike turer igjen i datasettet. For å begrense tidsbruken i forbindelse med beregningene og plottingen i SPSS ble datasettet delt i to omtrentlige store sett. I tillegg til å redusere beregningstiden så fikk man også et ferdig filtrert datasett som kunne brukes til testing eller som «back up» dersom det skulle skje noe med det første. Det endelige datasettet inneholdt nesten 1150 turer og overkant av 1 000 000 målepunkt. Tabell 3.2 gir en oversikt over data som ble filtrert ut i de forskjellige prosessene.

¹⁴ Døgnverdi for temperatur går fra 1800 UTC foregående døgn til 1800 UTC inneværende døgn, mens for nedbør beregnes verdien fra 0700 norsk normaltid foregående døgn til 0700 norsk normaltid inneværende døgn.

Tabell 3.2 – Oversikt over filtrering og deling av datasettet.

	Originale data	Filtrering				Deling av datasettet
		Fart > 70 Stigning > 35 & Stigning < -35 Stigning = 0,0 % og lengde > 100 meter Første og siste 100 meter	Ikke brukbar data	Reparerbar data	Totalt	
Antall turer	3 868	3 797	2 748	2 258	2 258	1 129
Fjernede turer		71	1 049	490	1 610	1 129
Fjernede turer [%]		1,84 %	27,63 %	17,83 %	42,40 %	50,00 %
Antall målepunkt	3 911 188	3 212 908	2 570 059	2 137 006	2 137 006	1 046 894
Fjernede målepunkt		698 280	642 849	433 053	1 774 182	1 090 112
Fjernede målepunkt [%]		17,85 %	20,01 %	16,85 %	55,22 %	51,01 %

3.2 Fartsmodellen og beskrivende variabler

Etter at datasettet var på plass ble modelleringsarbeidet satt i gang. For å komme frem til en metode for å beregne sykkelfart ble det benyttet lineær regresjon hvor stigning ble brukt som utgangspunkt. I dette kapittelet skal man først se litt statistikk over viktige variabler fra det endelige datasettet, før fremgangsmåten for modellarbeidet presenteres i kapittel 3.2.3.

3.2.1 Statistikk fra datasettet og metode

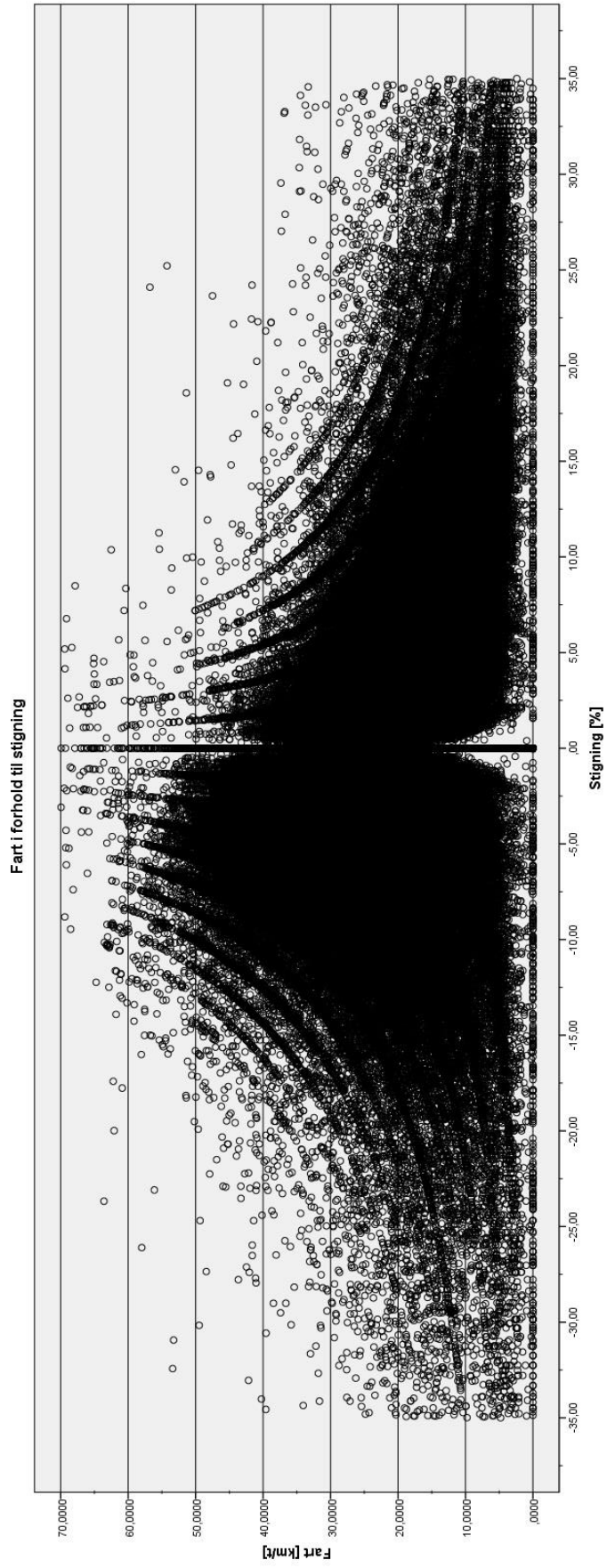
Datasettet som ble benyttet i analysen bestod av 1124 turer og 1 046 786 målepunkt¹⁵. Av disse målepunktene var ca. 91 % hentet fra Oslo, 6 % fra Trondheim og 3 % fra Kristiansand. Tabell 3.3 gir statistikk over noen utvalgte data. Det er flere punkt som tyder på at man ville hatt nytte av en grundigere filtrering, blant annet den store andelen punkter hvor stigningen er beregnet til 0,0 %. Videre ser man også at det finnes flere verdier som ligger på minimums- eller maksimumsgrensen for hva som ble filtrert bort. Spesielt tydelig er dette for stigning og fart.

Tabell 3.3 – Statistikk fra datasettet.

	Gjennom- snitt	Median	Standard- avvik	Minimum	Maksimum
Fart [km/t]	21,99	21,73	8,51	0,00	69,95
Stigning [%]	0,54	0,00	5,37	-35,00	35,00
Lengde på segmenter [m]	101,80	35,02	288,95	0,01	4380,40
Maks stigning på segmenter [%]	0,77	0,00	8,48	-35,00	35,00
Gjen. Stigning på segmenter [%]	0,54	0,00	5,01	-34,99	34,85
Temperatur [grader C]	8,32	8,70	6,95	-13,00	22,00
Nedbør [mm]	2,08	0,00	4,38	0,00	40,00

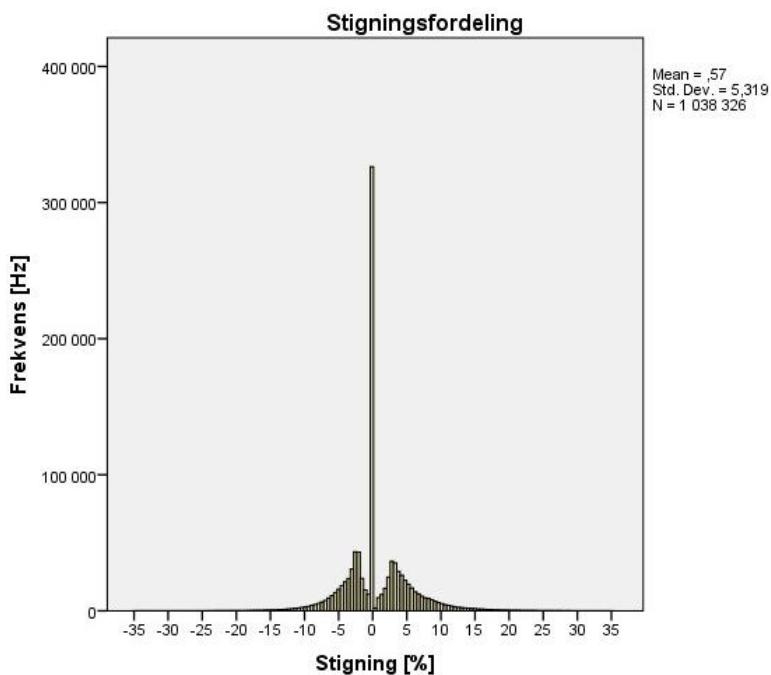
Figur 3.8 gir et plot av fart mot stigning for alle målepunkt, og nok en gang ser man tydelig overvekten av målepunkt med en stigning på 0 % og punkt med en registrert hastighet på 0 km/t. Det er også tydelig at det er store spredninger i datasettet, men dette skal vi se nærmere på i kapittel 4.1.1. Hastighetene ser ut til å være høyest i området -15 % til 0 % stigning, og flate ut på hver sin side. Det tyder på at syklister holder lavere fart i nedoverbakker som er bratte enn nedoverbakker som er slake.

¹⁵ Dette ble redusert til 1 038 328 målepunkt på grunn av negative avstander som ble oppdaget noe sent i analyseprosessen.

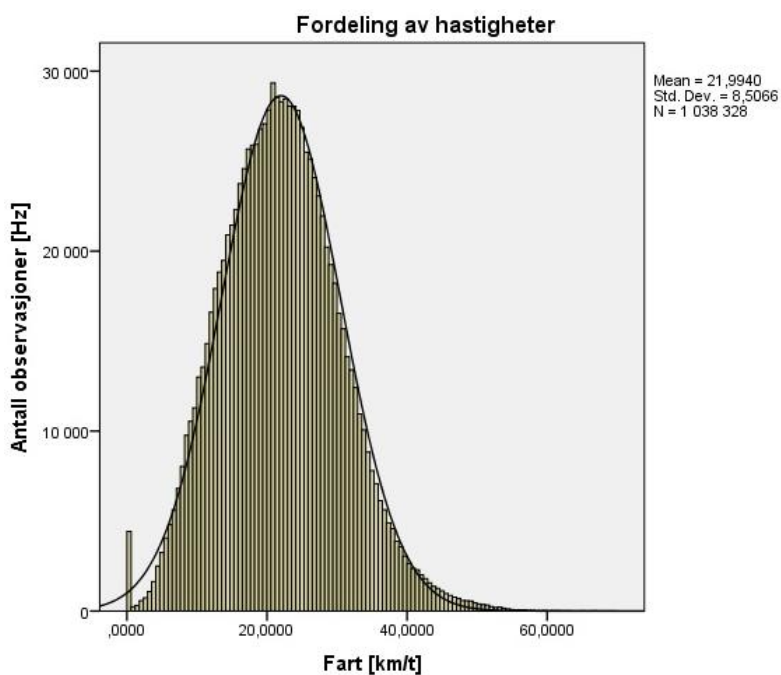


Figur 3.8 – Plot av fart og stigning for alle målepunkt.

Den store andelen målepunkt der stigning er registrert til 0 % kommer enda tydeligere frem i figur 3.9 som viser et histogram over stigningene. Opp mot 350 000 av de nesten 1 040 000 målepunktene har en stigning på 0,0 %. Legg også merke til at antall registrerte stigninger ser ut til å gå ned i området -2 til 2 %. Hastighetene ser ut til å være normalfordelte, med unntak av hastigheter registrert lik 0,0 km/t, se figur 3.10.



Figur 3.9 – Histogram for stigninger.



Figur 3.10 – Hastighetsfordelingen i datasettet

3.2.2 Statistiske antagelser

For å gjøre en lineær regresjon av datasettet var det flere statistiske tester som burde være oppfylt for at modellen skulle bli god (Lund og Lund, 2013):

1. Uavhengige residualer¹⁶ (Ingen autokorrelasjon)
2. Lineær sammenheng mellom beskrivende variabel og målvariabel
3. Likhet i residualenes varians (homoskedastisitet)
4. Ingen multikollinearitet
5. Ingen betydelige utliggere eller innflytelsesrike punkter
6. Normalfordelt residualer

De fleste av disse antagelsene ble løpende vurdert under arbeidet med oppbygging av modellen, og vil bli omtalt i detalj senere i kapittelet. Det bør nevnes at i denne analysen ble det gjort noen antakelser som gjorde at spesielt den første og den tredje testen kunne bli noe vanskelig å tilfredsstille:

1. Flere variabler som antageligvis påvirket farten var ikke tilgjengelige i datasettet (utstyr, fysisk form, infrastruktur, med flere).
2. Datasettet består av punktdata der farten i ett punkt har sammenheng med farten i forrige punkt. Observasjonene er altså ikke uavhengige.
3. Det kan være grupperinger i datasettet, spesielt med tanke på kjønn, fysisk form og erfaring som syklist.

Dersom de to første antagelsene holdt ville det bety at det var store sjanser for autokorrelasjon i datasettet. Det er et typisk problem for data som er tidsserier dersom det som skjer i punkt n er tett knyttet til det som skjedde punkt $n-1$. I tillegg kan sesong-avhengige variabler som vær og vind forsterke en forskjell mellom dataene (Chatterjee og Simonoff, 2013). Autokorrelasjon kan by på flere problemer for lineær regresjon: Minste-kvadraters metode er ikke lenger like effektiv og standard feil for regresjonskoeffisienter og varians inneholder bias. På grunn av denne biasen er konfidensintervall, signifikanstester og predikasjonsintervall ikke lenger nøyaktige (Chatterjee og Simonoff, 2013). Autokorrelasjon kan testes ved å undersøke *Durbin-Watson*-verdien, og skal helst være så nær 2,0 som mulig.

Lineær sammenhengen mellom indentifiserte variabler og målvariabelen er nødvendig for å unngå store feil i enkelte intervaller av modellen. Den lineære sammenhengen ble testet visuelt ved å produsere grafer som viste sammenhengen mellom en variabel og målvariabelen.

¹⁶ Residual = Feilledd, rest, gjenværende, rest-feil (Clue Dictionary Software, 2011). Forskjell mellom observert verdi og modellert verdi i et punkt (Chatterjee og Simonoff, 2013).

Dersom det ikke var en lineær sammenheng var det mulig å transformere variablene. Dette kunne man også se på en variabels skjevhet (skewness) og kurvatur (kurtosis).

Problemet med dårlig homoskedastisitet kan komme av at det finnes forskjellige grupperinger eller kategorier i datasettet som ikke er tatt hensyn til (Chatterjee og Simonoff, 2013). I dette tilfellet kan det være snakk om forskjellige grupper av personer: Personer i god fysisk form som sykler for å trene mot personer som ikke nødvendigvis gjør det for treningen sin skyld men fordi det er praktisk. Man kan da anta at den første gruppen vil holde høyere fart for en gitt stigning i forhold til den andre gruppen. Variansen i den første gruppen kan også bli mindre enn variansen i den andre gruppen, dersom man antar at den første gruppen er mer homogen enn den andre. Det kan være flere grupperinger i datasettet, og dersom disse grupperingene ikke er tatt hensyn til kan det by på problemer i regresjonsanalysen: Koeffisientene som kommer frem fra regresjonsanalysen er ikke så nøyaktige som man ønsker, konfidensintervallene kan være feil og hypotesetestene kan få feil signifikansnivå. Det betyr også at predikasjonene og predikasjonsintervallet kan være feil (Chatterjee og Simonoff, 2013). I praksis betyr dette at modellen bommer mer på enkelte verdier enn det gjør på andre. Homoskedastisitet kan testes ved å benytte en ANOVA-test eller ved å plote residualene mot predikasjonene. Det skal være en spredt fordeling der det er like store variasjoner i residualene for alle predikasjonene.

Multikollinearitet kan sjekkes ved å studere *variance inflation factor* (VIF), og skyldes at to eller flere variabler beskriver eller forklarer den samme variasjonen i målvariabelen. Dersom man har multikollinearitet betyr det ofte to eller flere variabler er kollineære; endring i den ene variabelen fører til endring i den andre variabelen, og en av variablene er derfor ofte overflødig. Kollinearitet skaper problemer ettersom det påvirker t- og F-testene, og små endringer i dataene kan føre til store endringer i modellkoeffisientene (Chatterjee og Simonoff, 2013).

Utliggere og innflytelsesrike punkt er registrerte data som betegnes som uvanlige observasjoner: De kan skyldes feil i registreringen eller tilsvare ekstremtilfeller som i de aller fleste tilfeller ikke er representative for hva man ønsker å modellere. De trenger ikke være feil, men kan påvirke regresjonen på en uheldig måte ettersom minste kvadraters metode kan være meget sensitiv for uvanlige observasjoner (Chatterjee og Simonoff, 2013). Utliggere kan identifiseres ved å se på residualene mens innflytelsesrike identifiseres av *Cooks Distance* eller *leverage value*. Alle disse tre verdiene kan beregnes automatisk under regresjons-

analysen. Residualer som er tre ganger større enn standardavviket eller målepunkt som har Cooks Distance større enn 1,0 bør undersøkes. Kritisk punkt for leverage value vil variere avhengig av antall observasjoner og variabler (Chen et al., 2003).

Siste og endelige test var å undersøke om residualene var normalfordelt, som lett sjekkes ved å plote residualene i et histogram. Dersom residualene ikke var normalfordelt kunne de påvirke påliteligheten til f- og T-testene: Konfidens- og predikasjonsintervallene vil ikke være representative (Chatterjee og Simonoff, 2013).

3.2.3 Oppbygging og testing av modellene

Tross innledende mistanke om problemer med statistiske forutsetninger ble det valgt å benytte lineær regresjon i utledingen av modellen for fart, ved hjelp av statistikkprogrammet *SPSS*. Autokorrelasjonsproblemet kan i noen tilfeller fikses og i andre tilfeller er det mulig å godta denne feilen så lenge man er klar over problemene dette byr på. Dersom autokorrelasjonen i stor grad skyldes mangel på beskrivende variabler vil den forsvinne når flere variabler blir inkludert (Chatterjee og Simonoff, 2013), og i dette tilfellet var det mistanker om at datasettet ikke inneholdt alle variablene som man antok påvirket sykkelhastighetene.

Det ble raskt klart at det var nødvendig å utvikle flere modeller ettersom noen variabler antageligvis ville påvirke farten forskjellig avhengig av stigningen. Eksempelvis var det naturlig å anta at lengre segmenter ville føre til økt fart for negative stigninger men redusert fart for positive stigninger. Det ble også antatt at noen variabler ville være mer forklarende i bakker enn på rette lengder. Derfor ble det valgt å lage tre modeller som samsvarte med segmentinndelingen som ble gjennomført tidligere i arbeidet.

Modellene ble hovedsakelige utarbeidet i *SPSS* ved hjelp av den innebygde funksjonen for lineær regresjon gjennom flere runder. Som nevnt ble det hentet inn værddata for MET, og i tillegg ble det opprettet variabler som beskrev egenskapene til forrige og neste segment (maks stigning, gjennomsnittlig stigning og lengde) slik at det var mulig å se på effekten av dette. Antagelsen bak dette var at farten i et segment kunne påvirkes både av forrige segment og neste segment: Dersom man akkurat hadde syklet opp en meget bratt bakke ville man tro at farten er lavere enn dersom man akkurat kom fra en lang nedoverbakke.

Ettersom datasettet bestod av flere turer utført av flere personer ble det nødvendig å redigere denne koblingen i overgangen mellom turene slik at ikke data fra en tur feilaktig påvirket data for en annen tur. Verdiene ble fjernet i hver overgang og disse punktene ble utelatt i

regresjonsanalysen. (Hadde variablene for forrige og neste segment blitt beregnet *før* de første og de siste for hver tur ble fjernet, hadde ikke dette vært et problem.)

For hver runde ble variabler enten fjernet eller lagt til avhengig av flere faktorer. Hovedsakelig bestod dette av å fjerne variabler som hadde liten innvirkning målvariabelen og på den totale forklaringsgraden, eller viste tegn til høy kollinearitet med andre variabler. Det ble også nødvendig å transformere noen variabler på grunn av dårlig lineær sammenheng, skjevhet eller kurvatur. De fleste av disse hadde skjevhet som førte til at det ofte ble brukt en logaritmisk eller kvadratisk transformasjon. Etersom stigningsvariablene hadde negative verdier ble det lagt til en verdi på 36,0 da de ble transformert. På den måten ble minste verdi som skulle transformeres lik 1,0 og man unngikk problemet med å ta logaritmen av nullverdier og negative tall. Det ble identifisert flere utliggere men ettersom de ikke var innflytelsesrike ble de ikke fjernet. Det ble testet ved flere tilfeller men førte ikke til forbedringer på noen signifikant måte i modellene.

Da modellene var utviklet ble det utført to tester. Først ble modellene testet opp mot den andre halvdel av det originale datasettet. Ved å se på forskjellen mellom observerte og modellerte hastigheter var det mulig å danne et inntrykk av hvor godt modellen fungerte. Dette ble testet ved å gjennomføre en hypotesetest der man sjekket om forskjellen i hastighetene var statistisk signifikant forskjellig fra 0 km/t.

I den andre testen ble modellene knyttet opp mot et nettverk for Trondheim ved hjelp av et GIS (ArcMAP). Nettverket som ble benyttet var det samme som nettverket til delområdemodellen Nidaros (DOM Nidaros), som er en mindre versjon av den regionale transportmodellen. Formålet var å undersøke fart og reisetid dersom man benyttet de nye modellene sammenlignet med dagens metode i RTM. Etersom de nye modellene mest sannsynlig hadde en eller flere variabler som var avhengig av forrige eller neste segment, ble det benyttet en innebygd funksjon i ArcMAP for å bestemme den korteste ruten mellom to punkt. Deretter ble rutedataene eksportert og hastighetene ble så beregnet ved hjelp av et Python-skript. Deretter ble hastighetene importert tilbake til ArcMAP for å visualisere forskjellene.

3.3 Oppsummering

Det ble indentifisert flere kilder som hadde offentlig tilgjengelig GPS-data, men kun Garmin Connect viste seg å være tilfredsstillende til utviklingen av en metode for sykkelfart. Derfor ble data kun hentet fra denne kilden. Både under og etter innhenting ble det nødvendig med filtreringer. Typiske treningsturer og forurenset data ble fjernet for at datasettet i størst mulig grad skulle inneholde pendleturer.

Dataene som samlet inn bestod hovedsakelig av turer som ble gjennomført i Oslo, men data for Trondheim og Kristiansand ble også samlet inn. Det endelige datasettet bestod av ca. 2 million målepunkt fordelt over ca. 2200 turer, men ble delt i to omtrent like store sett. Det ene settet ble brukt til modellutviklingen mens det andre settet ble brukt for å gjennomføre en statistisk test av modellene.

Tross flere runder med filtreringer inneholdt datasettet noen skjevheter. Det var blant annet registrert en stor mengde punkt med 0,0 % stigning. Det var også registrert flere punkter hvor hastigheten var 0,0 km/t.

Modellene ble utviklet ved hjelp av lineær regresjon. Først ble målepunktene samlet i segment, for så å beregne hvert segments lengde, gjennomsnittlige stigning og maksimale stigning. Deretter ble det laget koblinger som beskrev de samme egenskapene til forrige og neste segment, mens til slutt ble det lagt til nedbørs- og temperaturdata fra Meteorologisk Institutt. Variabler ble transformert, lagt til eller fjernet alt ettersom forklaringsgraden økte eller sank. Samtidig måtte man være på utkikk etter faresignaler som multikollinearitet, skjevhet og innflytelsesrike punkter. Dette ble vurdert fortløpende under arbeidet og det ble utviklet en modell for nedoverbakker, en for tilnærmet flate strekninger og en for oppoverbakker.

Modellene ble testet på to måter: Først ble de implementert i den andre halvdel av datasettet og det ble sammenliknet målte mot modellerte hastigheter. Deretter ble det benyttet et nettverk for Trondheim og det ble utført en sammenlikning mellom dagens metode for sykkelfart i RTM opp mot den nye metoden.

4 Resultat

I dette kapitlet vil de viktigste resultatene fra undersøkelsene presenteres. Det er delt i fire hoveddeler der den første delen tar for seg interessante funn i dataene, mens andre del tar for seg modellene og de tilhørende statistiske resultatene. Tredje og fjerde del ser på testing av modellene opp mot et alternativt datasett og dagens metode for sykkelfart i RTM.

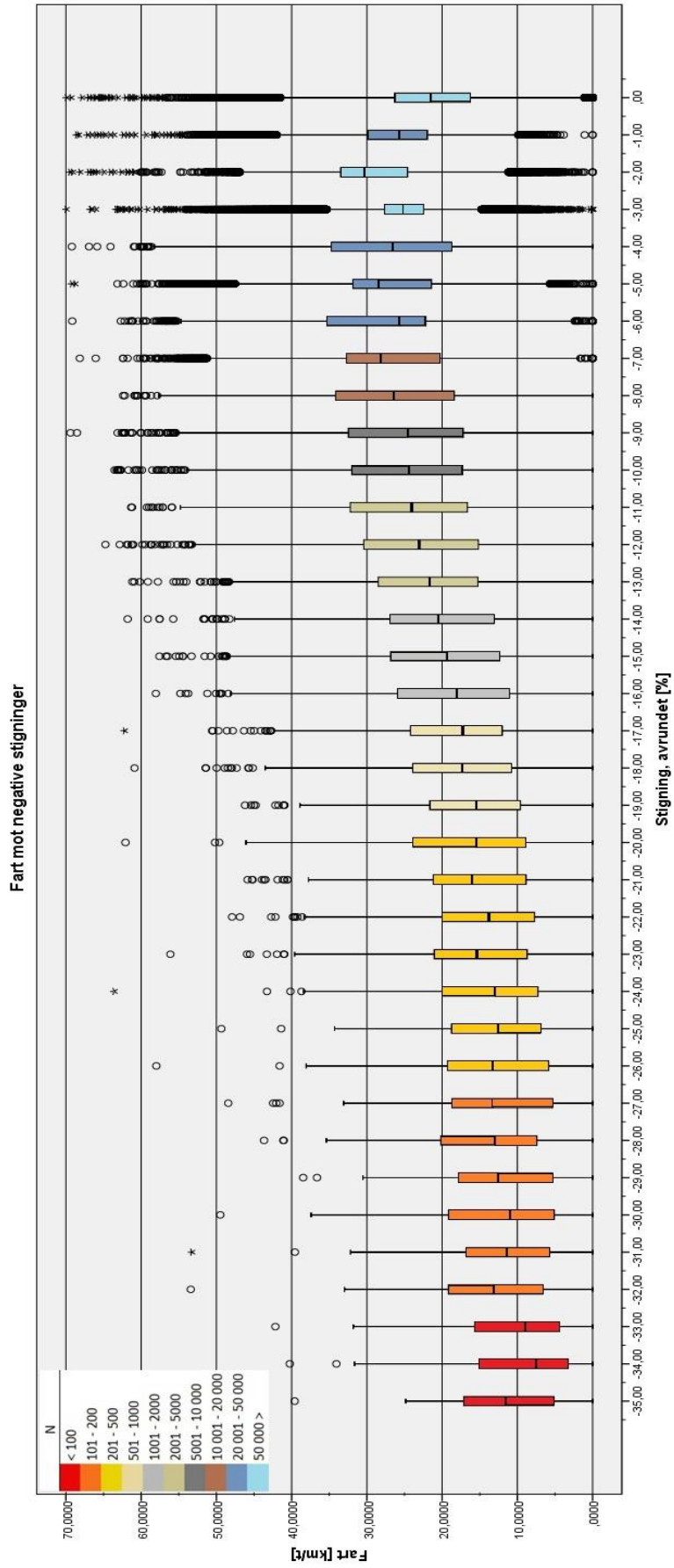
4.1 Spredninger og andre funn i datasettet

Ettersom det har vært lite arbeid med å undersøke potensialet til GPS-data som er tilgjengelig i «skyen», kan det være interessant å se på kvaliteten til dataene. Først og fremst er det interessant å se spredningen i datasettet. Dette kan gi noen indikasjoner på presisjonen til datasettet og danner et viktig grunnlag for videre diskusjoner.

4.1.1 Hastighetsfordeling og spredning

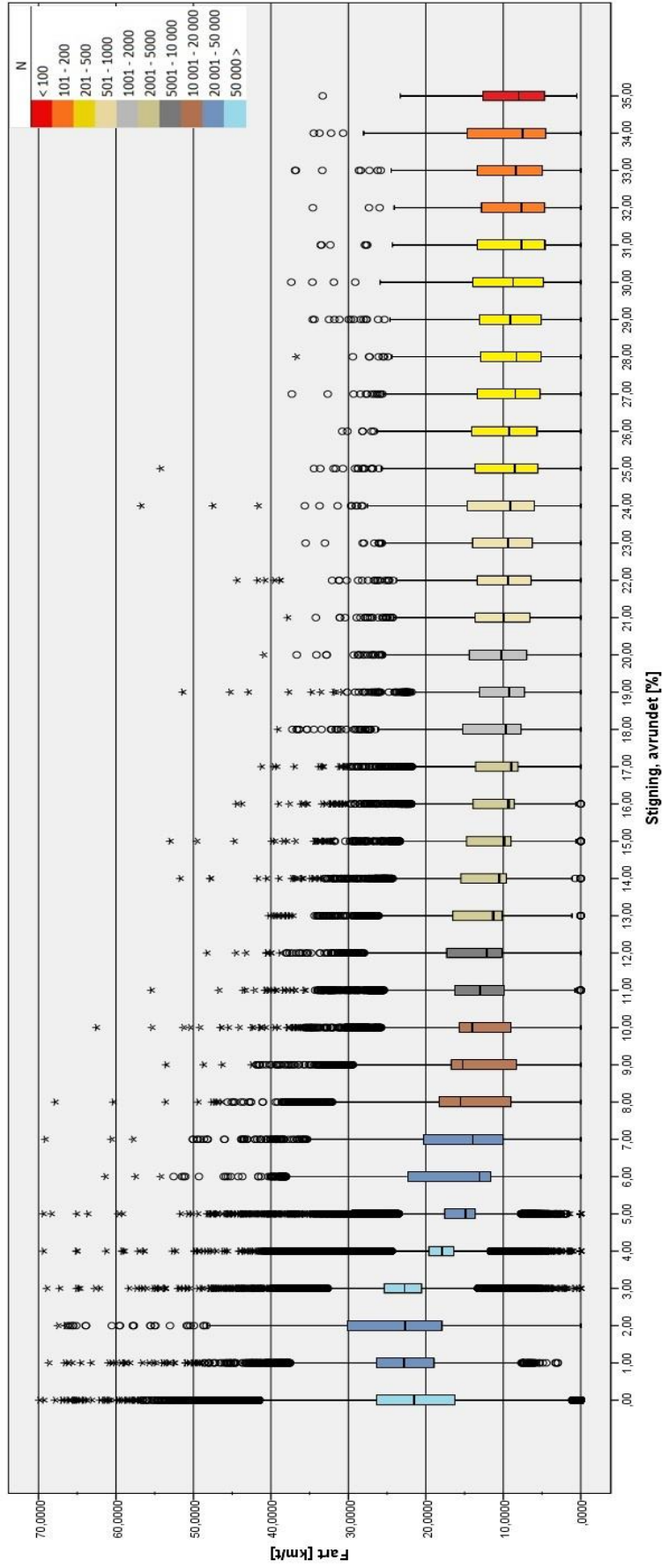
Som nevnt tidligere var det store spredninger i fartsverdier i datasettet. Dette kommer enda tydeligere frem i plottene i figur 4.1 og figur 4.2, som viser boxplot for fart mot stigning. For de fleste stigninger er fartsdataene spredt over et vidt område, med noen få unntak: I området -3 % til 0 % er dataene mindre spredt enn for resten av de negative stigningene, men det er fortsatt mange ekstremverdier. For de positive verdiene ser man at i området 3 til 6 % er det lite spredning, men nærliggende verdier varierer stort. Spesielt interessant er det å studere medianverdien for de positive stigningene: Medianen ser ut til å synke i takt med økende stigning, men for stigninger over 6 % ser det ut som om det faktisk er en liten øking. Dette er vist enda tydeligere i figur 4.3, og verdien ser ut til å øke for stigningene 7 % og 8 %, før den på nytt begynner å synke ved 9 %. Samtidig ser vi tydelige forskjeller i spredningen på dataene: Mens hovedvekten av de registrerte fartsverdiene for 5 % ligger mellom ~7,5 til ~24,0 km/t er tilsvarende intervall mellom 0 og ~38 km/t for 6 %.

Man ser også at antallet observasjoner er meget lavt for de bratte stigningene sammenliknet med de slake stigningene. Det er først i stigningsområdet -16 % til +20 % at vi har over 1000 observasjoner per stigning. Til sammenlikning er det over 20 000 observasjoner for hver stigning i området -6 % til +7 %, men det er en merkelig «dupp» i antall observasjoner for stigningene -1 %, 1 % og 2 %.



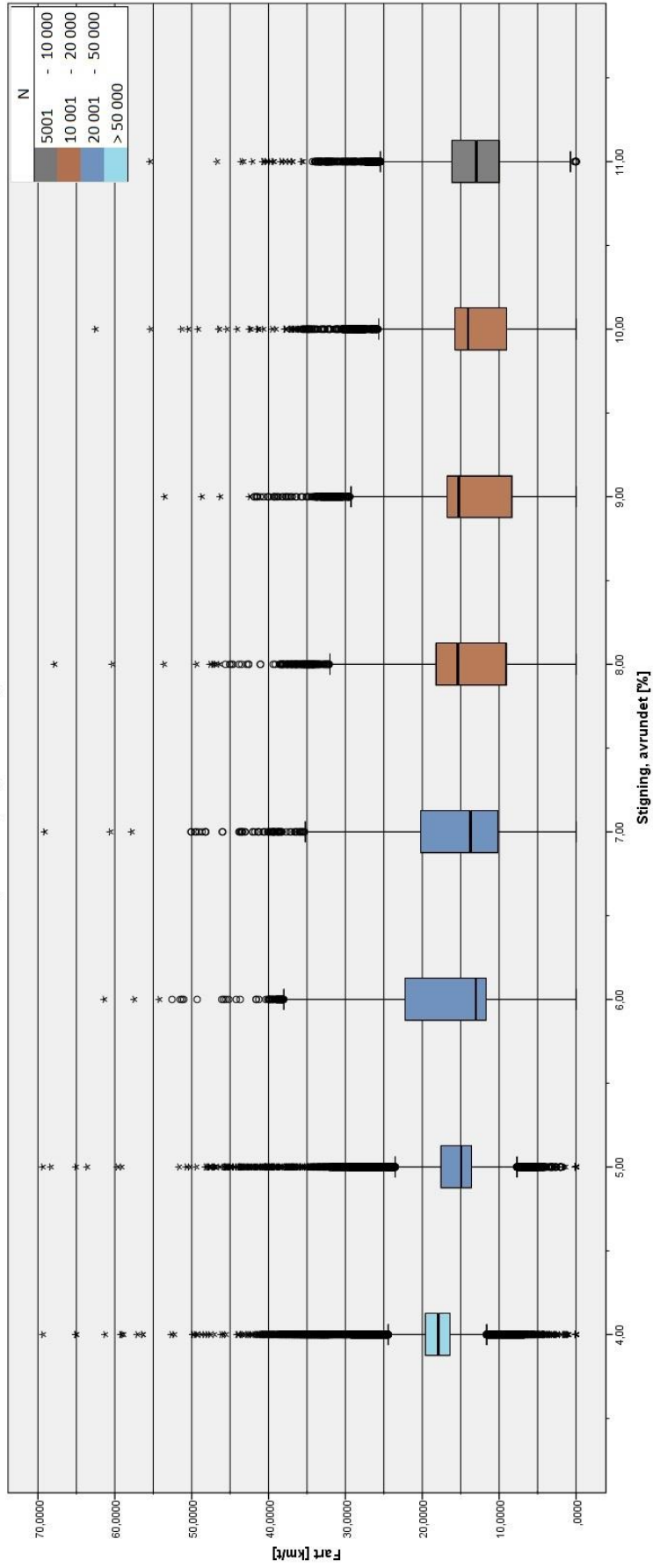
Figur 4.1 – Boxplot for stigning mellom -35 og 0 % . N er antall observasjoner, mens sirklar er utliggere. Stjerne tilsvarer en ekstrem utligger.

Fart i forhold til positive stigninger



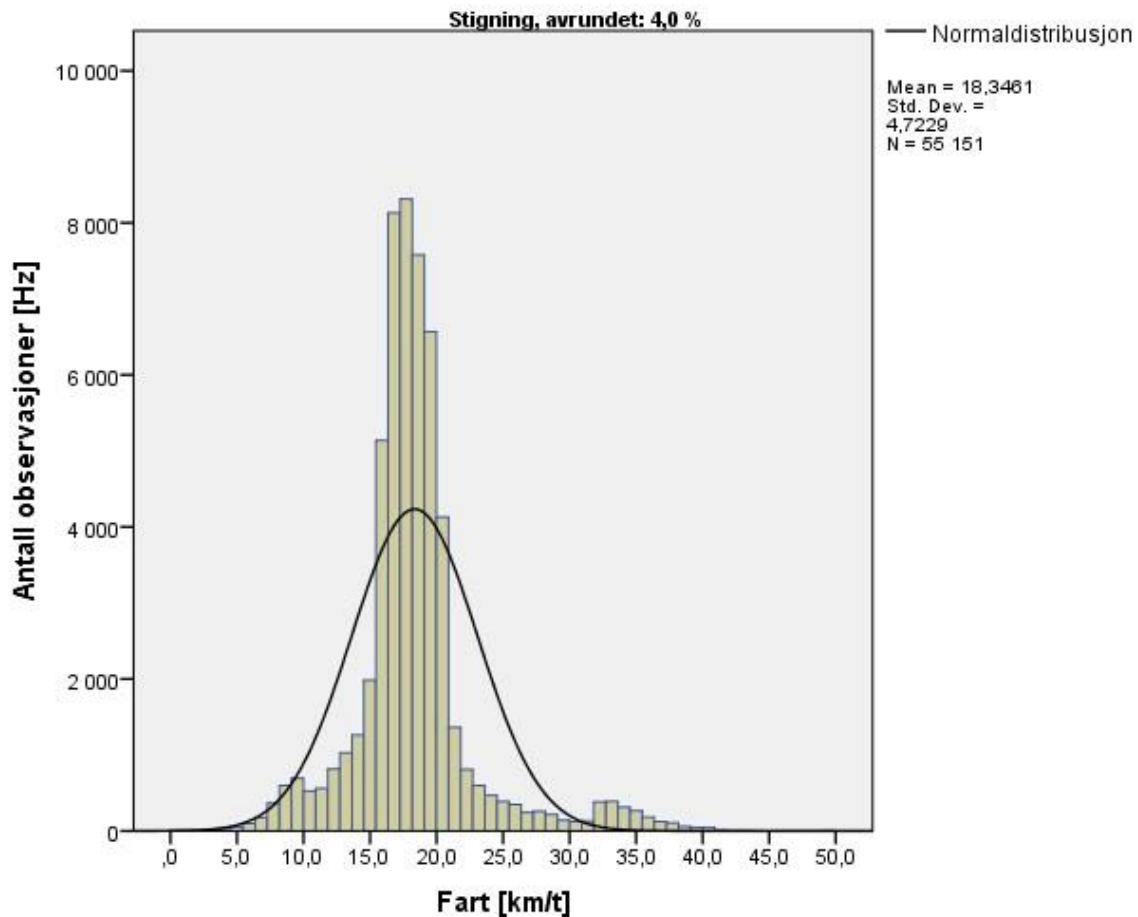
Figur 4.2 – Boxplot for stigninger mellom 0 og 35 % . N = Antall observasjoner, mens sirkler er utliggere. Stjerne tilsvarer en ekstrem utligger.

Boxplot for stigning mellom 4 og 11 %



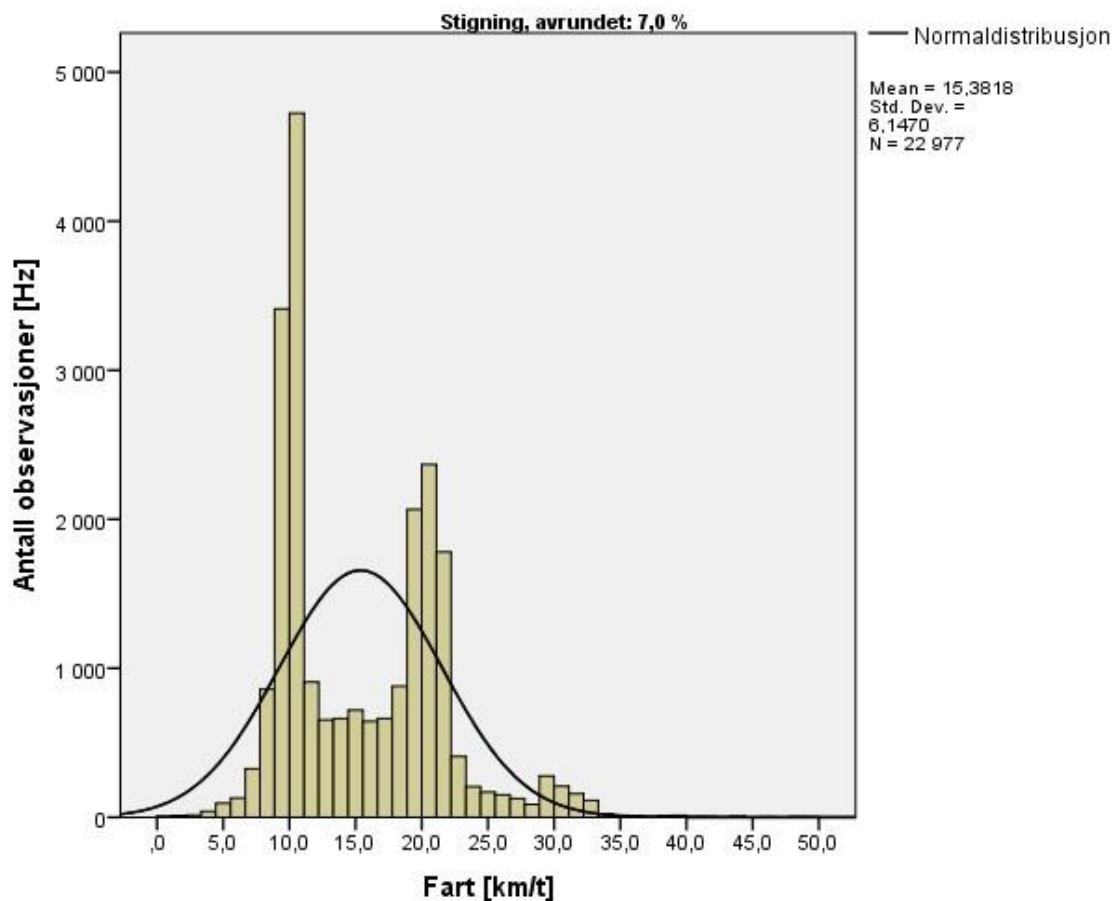
Figur 4.3 – Boxplot for stigninger mellom 4 og 11 % . Økt spredning i data fra 6 % og utover sammenliknet med 4 og 5 % . N = Antall observasjoner, mens sirkler er utliggere. Stjerne tilsvarer en ekstrem utligger.

For å undersøke hva økningen i medianverdi og økt spredning skyldes kan man se på histogram med tilhørende normalkurve for stigningene. Dette er gitt i figur 4.4 for en stigning på 4 %, og i figur 4.5 for stigning på 7 %. Det kommer tydelig frem av figurene at datasettet er mer samlet for 4 % enn for 7 %. De fleste verdiene befinner seg innenfor et intervall på ca. 5 km/t.



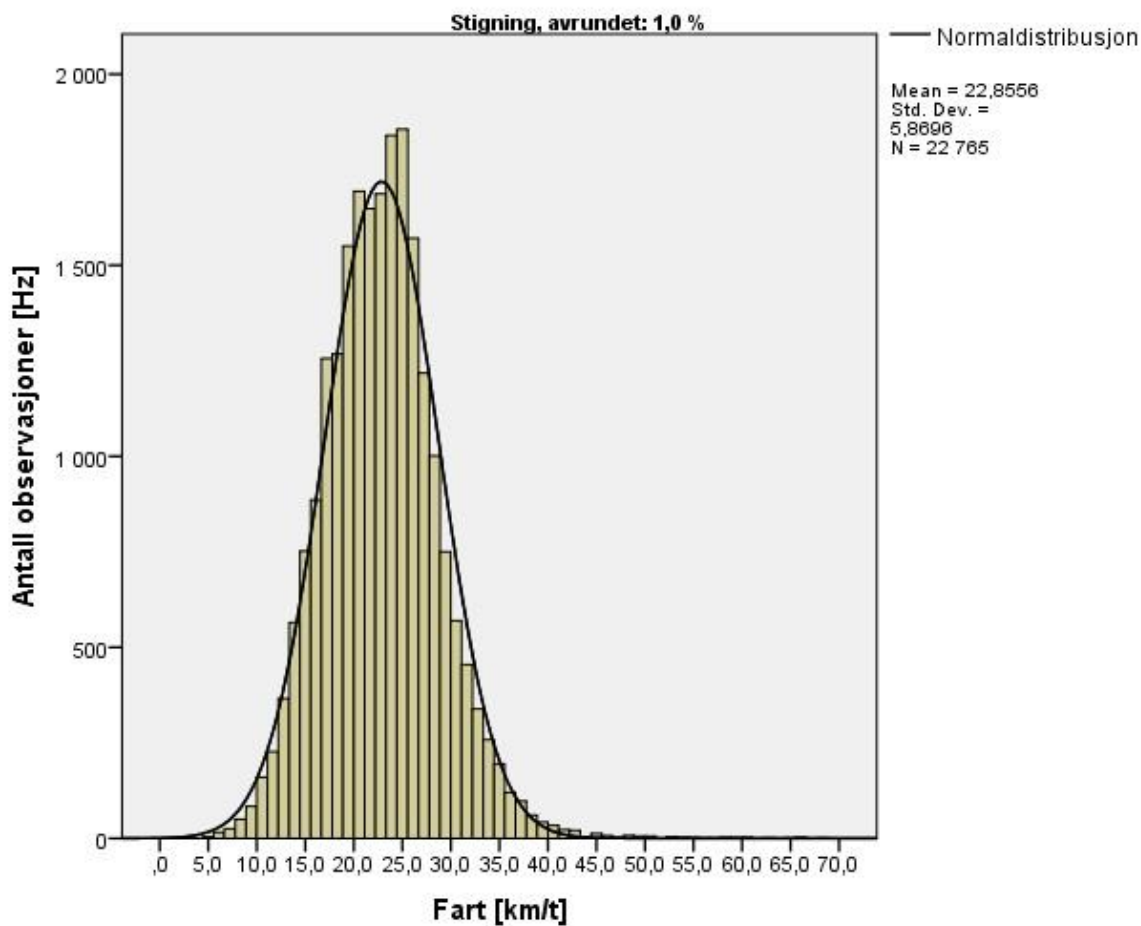
Figur 4.4 – Fordeling av hastigheter ved 4 % stigning , med normalfordelingskurve. Mean = gjennomsnitt, Std. Dev = standardavvik, N = antall observasjoner

For 7 % ser man at dataene ligger hovedsakelig i en av to «grupper»: Den ene gruppen har en hastighet på $10 \text{ km/t} \pm 2 \text{ km/t}$, mens den andre gruppen ligger nærmere $21 \text{ km/t} \pm 2 \text{ km/t}$, altså en forskjell på omtrent 10 km/t.



Figur 4.5 – Hastighetsfordeling ved 7 % stigning med normalfordelingskurve. Mean = gjennomsnitt, Std. Dev = standardavvik, N = antall observasjoner.

Denne «grupperingen» som er vist i figur 4.5 er synlig for flere andre stigninger. For noen stigninger er det flere grupperinger som ikke er like dominerende, mens andre stigninger har enda mer dominerende grupperinger enn den som er vist over. Det finnes også noen stigninger der hastighetene er tilnærmet normalfordelt, som vist i figur 4.6.



Figur 4.6 – Hastighetsfordeling for 1 % stigning med normalfordelingskurve. Mean = gjennomsnitt, Std. Dev = standardavvik, N = antall observasjoner.

Det er helt tydelige spredninger i datasettet. Det ser man klart fra boxplottene, spesielt med tanke på den store mengden utliggere og ekstremverdier. For de mindre stigningene ser man at det er så mange verdier som er registrert som utliggere at det er vanskelig å telle dem ut i fra plottet.

4.2 Ny metode for å beregne sykkelfart

Det ble valgt å utvikle tre modeller for å beregne sykkelfart som samsvarte med segmentinndelingen gjort i kapittel 3.1.3. Metoden er en forenkling av virkeligheten ettersom man regner på segmentverdiene og ikke de direkte stigningsverdiene. De tre modellene vil omtales hver for seg i sitt eget underkapittel. I tillegg til å presentere selve metoden vil resultatene fra de tilhørende statistiske testene nevnt i tidligere kapittel også presenteres for hver modell.

4.2.1 Modell for fart i nedoverbakker (stigning < -2,0 %)

Modellen er presentert i formel (11) og gjelder for nedoverbakker som har en stigning som er brattere enn -2,0 %. Hver koeffisient forteller oss hvor stor endring det er i fart dersom den tilhørende variabelen endres med 1,0 og alle andre variabler holdes konstante.

$$V = 66,73 + 0,86\bar{s} + 6,85 * \log(l_s) + 0,16t - 31,1 * \log(\bar{s}_f + 36) \quad (11)$$

Der

- V = Hastighet [km/t]
- \bar{s} = Gjennomsnittlig stigning, gjeldende segment [%]
- l_s = Lengde på stigningen, gjeldende segment [m]
- t = Døgnmiddeltemperatur [grader C]
- \bar{s}_f = Gjennomsnittlig stigning, forrige segment [%]

Det første man legger merke til er den relativt høye konstanten, men ettersom denne formelen kun gjelder for negative stigninger vil \bar{s} ligge intervallet [-35, -2), som gjør at andre ledd alltid vil være negativt. Det er interessant å legge merke til at dette leddet indikerer at man holder lavere fart ved brattere stigninger, og at farten øker når stigningen blir flatere. Tredje ledd indikerer at farten øker med den logaritmiske funksjonen av lengden på stigningene. Det betyr at en endring i lengde gir større utslag for korte bakker enn lengre bakker.

Vi ser også at temperatur har en innvirkning på farten. Dersom temperaturen kryper under 0 grader vil det føre til lavere hastigheter, mens temperaturer over 0 grader fører til økte hastigheter. Det er urimelig å anta at det er så rett frem som dette, og det er mer trolig at temperatur er en indikator på generelt vær eller sykkelføre.

Siste ledd i funksjonen tilsier at farten synker logaritmisk med stigningen på forrige segment. Jo brattere forrige segment var, jo lavere fart har man i nedoverbakken. Dette er en rimelig antagelse, ettersom jo brattere den forrige bakken jo lavere fart vil man ha når man begynner på nedoverbakken. Koeffisienten til dette leddet er veldig høy, men på grunn segmentinndelingen som ble gjennomført vil \bar{s}_f ligge i området [-2, 35]. Sammen med den logaritmiske transformasjonen fører det til at dette leddet vil variere mellom -47,6 til -56,7 km/t, og balansere ut den høye konstanten. Det er riktignok noen urimeligheter i dette, ettersom leddet alltid vil gi et negativt bidrag til farten. Det betyr dersom forrige stigning var i området -2,0 % til 0 % vil det fortsatt bety at det påvirker farten negativt, selv om det ikke virker like naturlig. En svak nedoverbakke burde kanskje føre til økte hastigheter, men avhenger av noen fysiske variabler som vindmotstand, vekt og rullemotstand.

Slik koeffisientene er satt opp kan man i ekstremtilfeller (meget lav gjennomsnittlig stigning, meget lav temperatur, kort stigningslengde og meget høy stigning i forrige segment) få negative verdier for fart. Dette er ikke rimelig, og en nedre verdi for fart for negative stigninger bør derfor legges inn for å unngå rare resultat.

4.2.2 Statistikk for metoden for negative segmenter

Resultatet for forklaringsgrad (R^2), standardavvik og Durbin-Watson (autokorrelasjon) er gitt i tabellen under. Forklaringsgraden forteller oss at vi gjør rede for om lag 33 % av variasjonene i fart ved hjelp av modellen, som ved første øyekast er ganske svakt. Ideelt skal forklaringsgraden være så nær 1,0 slik at alle variasjonene i fart kan forklare av modellen. Man ser også at standardavviket er ganske stort, samt en meget bekymrende Durbin-Watson verdi.

Tabell 4.1 – Forklaringsgrad, standardavvik og Durbin-Watson for negative segmenter

Modell	R	R^2	Justert R^2	Standardavvik til predikasjonene	Durbin-Watson
Stigning < - 2,0 %	0,575	0,331	0,331	7,189	0,268

Den lave forklaringsgraden kan bety at det er flere variabler som mangler i analysen. Flere av variablene som var tilgjengelige i datasettet har blitt utelatt fra denne modellen ettersom tilføyingen eller fjerningen av disse variablene ikke førte til økt forklaringsverdi eller på noen annen måte bedre modell.

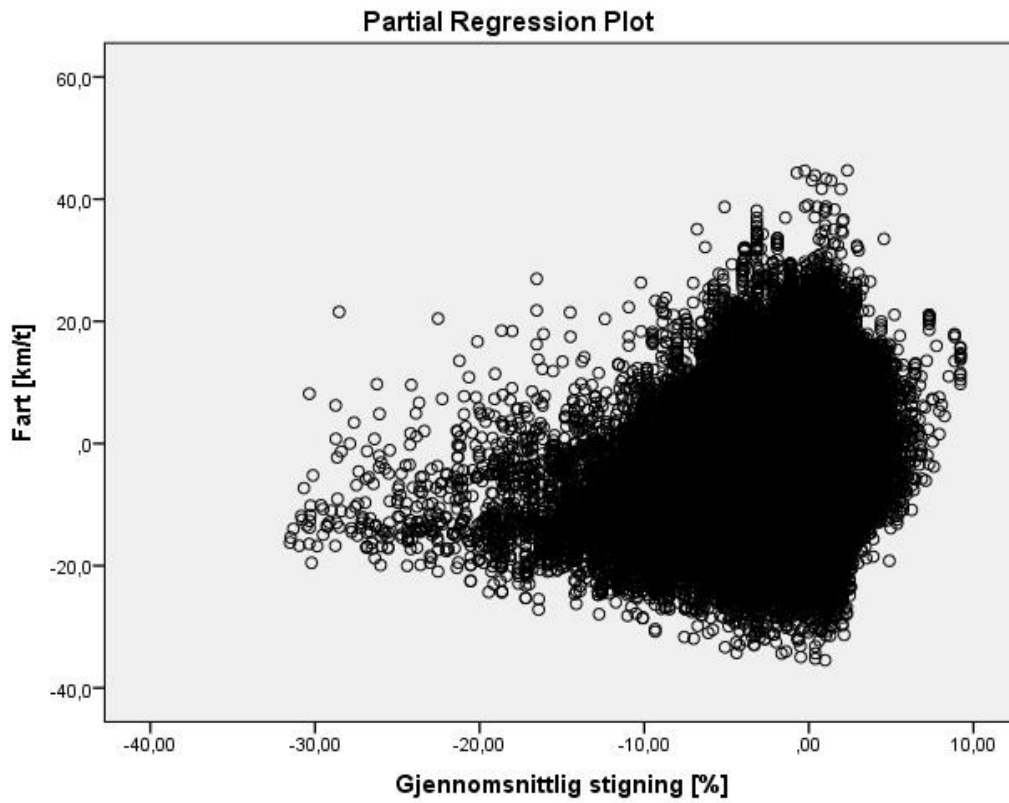
Det store problemet er i midlertid det tydelige tegnet på autokorrelasjon. Durbin-Watson verdien ligger alltid i området $[0, 4]$ og jo nærmere et av disse ytterpunktene man befinner seg jo større er sjansen for at man har autokorrelasjon. Ideelt sett skal verdien være så nærme 2,0 som mulig. Fra kapittel 3.2.1 ble det gitt at dette er et typisk problem for tidsserie-data slik som dette, og at det finnes struktur i datasettet som ikke blir beskrevet av modellen. Det betyr også at standardavviket i tabell 4.2 inneholder bias og signifikansnivået ikke er gyldig.

Tabell 4.2 – Standardavvik og signifikansnivå for koeffisientene

Variabel	Koeffisient	Standardavvik	Signifikansnivå
Konstant	66,729	0,759	0,000
Gjennomsnittlig stigning	0,863	0,005	0,000
log10 stigningslengde	6,854	0,023	0,000
Temperatur	0,157	0,002	0,000
log10 gjennomsnittlig stigning + 36 %, forrige segment	-31,100	0,485	0,000

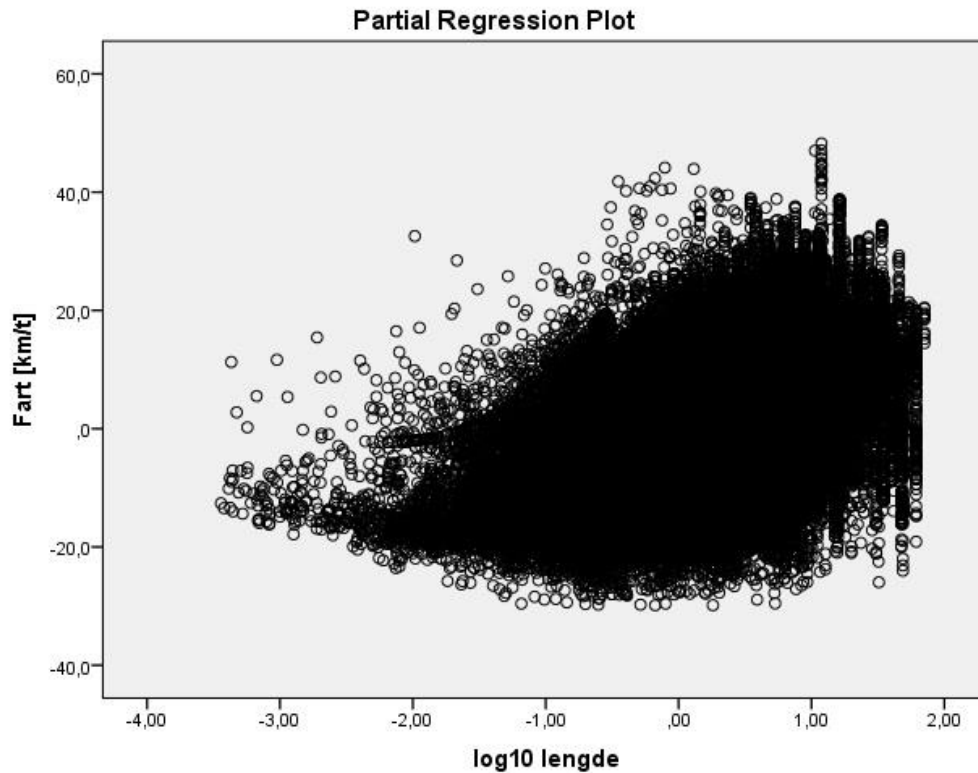
Den lineære sammenhengen mellom beskrivende variabel og målvariabel er varierende. Figur 4.7 viser et *partial regression plot*¹⁷ som viser sammenhengen mellom endring i stigning og fart, og illustrerer nok en gang hvor stor spredning samt mengden av målepunkt i datasettet.

¹⁷ Viser sammenheng mellom valgt variabel og målvariabel. Gir et mer korrekt bilde på sammenheng når det er mer enn én variabel i analysen.



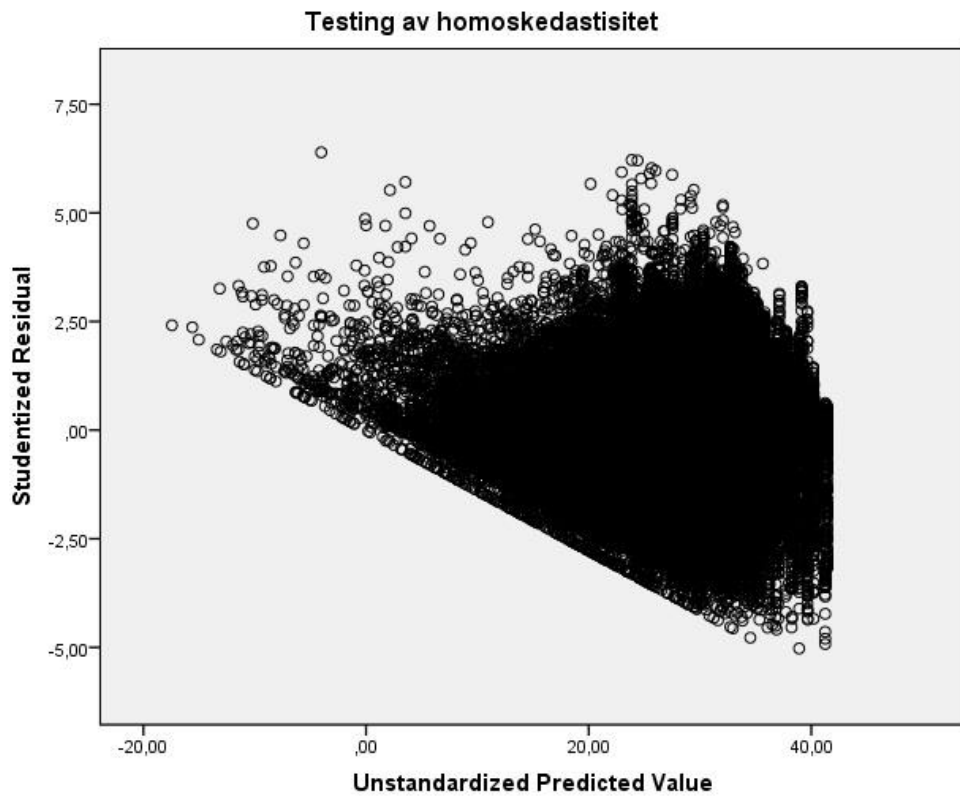
Figur 4.7 – Partial Regression Plot av fart mot stigning

For den logaritmiske transformasjonen av stigningens lengde ser plottet litt bedre ut. Dette er vist i figur 4.8.



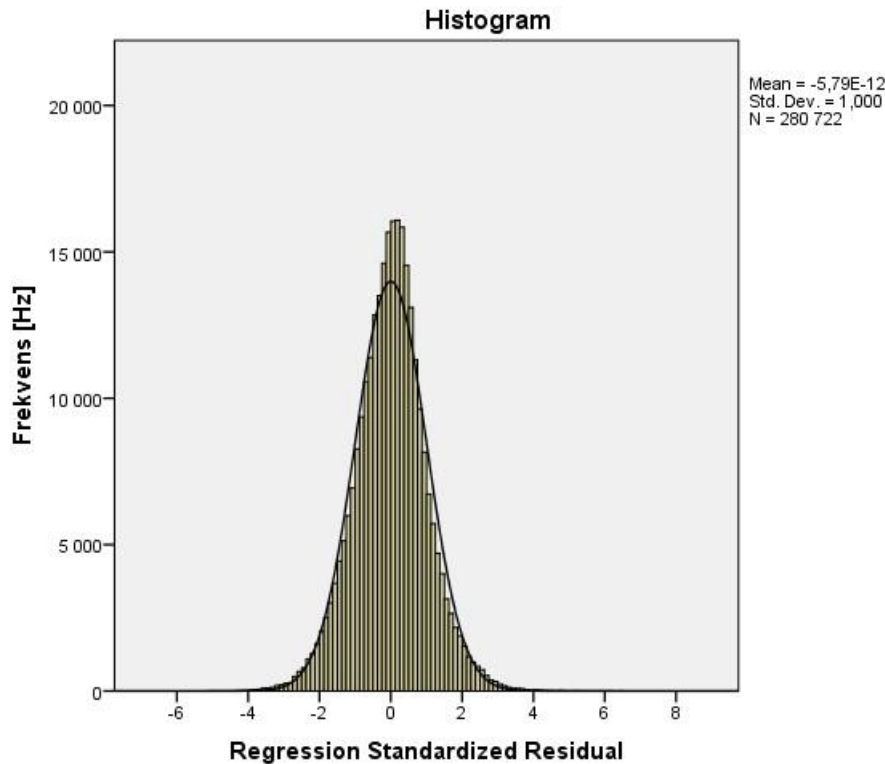
Figur 4.8 – Partial Regression Plot, fart mot log10(lengde)

Neste test for modellen er å vurdere homoskedastisiteten. Figur 4.9 viser et plott av residualene mot predikasjonsverdiene. Her var det ønskelig med en jevn spredning av residualene, men vi ser at dette ikke er tilfellet. Det er tydelig forskjell i spredningen for lave verdier i motsetning til høye verdier, og ser en tydelig nedre grense. Man bør også legge merke til at det er predikert negative verdier, som ikke er reelt. Multikollinearitet var ikke et problem ettersom alle variabler som viste tegn dette ble luket ut i tidligere faser. VIF-verdiene for den endelige modellen er gitt i vedlegg D og ingen av dem er i nærheten av den kritiske verdien på 10,0. Som nevnt ble det identifisert en rekke utliggere men ingen av dem hadde stor innflytelse. Sletting av disse målepunktene ga ingen forbedringer på modellen, så ingen av dem ble fjernet.



Figur 4.9 – Testing av homoskedastisitet i datasettet for nedoverbakker. Studentized residual = residualer, Unstandardized predicted value = predikert verdi.

Siste og avsluttende test var å se på fordelingen av residualene. Denne måtte være tilnærmet normalfordelt, og fra figur 4.10 ser vi at det viser seg å være tilfellet.



Figur 4.10 – Normalfordeling av residualene.

4.2.3 Modell for tilnærmet flate stigninger ($-2,0 \% \leq \text{stigning} \leq 2,0 \%$)

Modellen er presentert i formel (12). Sammenliknet med modellen for nedoverbakker er det første man legger merke til at denne modellen er avhengig av hva som skjer i neste segment, og ikke bare forrige segment. Man ser også at lengdene til disse tilsluttede segmentene er med i modellen.

$$V = 12,6 - 2,4\bar{s} + 2,44 * \log(l_s) - 0,2\bar{s}_f + 2,22 * \log(l_{s_f}) - 0,18\bar{s}_n \quad (12)$$

$$+ 1,47 * \log(l_{s_n}) + 0,2t$$

Der

- V = Hastighet [km/t]
- \bar{s} = Gjennomsnittlig stigning, gjeldende segment [%]
- l_s = Lengde på stigningen, gjeldende segment [m]
- l_{s_n} = Lengde på stigning, neste segment [m]
- \bar{s}_f = Gjennomsnittlig stigning, forrige segment [%]
- l_{s_f} = Lengde på stigningen, forrige segment [m]
- \bar{s}_n = Gjennomsnittlig stigning, neste segment [%]

t = Døgnmiddeltemperatur [grader C]

De to første leddene beskriver egenskapene til gjeldende segment. Dersom stigningen er negativ vil farten øke, mens den vil synke for positive stigningsverdier. \bar{s} ligger i intervallet $[-2, 2]$ og bidraget fra denne variabelen vil derfor variere fra $-4,8$ km/t til $4,8$ km/t. Til sammenlikning er stigningen til forrige og neste segment, \bar{s}_f og \bar{s}_n , i intervallene $[-35, -2)$ og $(2, 35]$ som potensielt kan gi en endring i fart på ca. ± 6 km/t. Dette er selvsagt i ytterpunktene, og bidraget vil sjeldent være i den størrelsesordenen. Likevel er det interessant å se at for denne modellen det ikke er en logaritmisk sammenheng mellom fart og stigning i forrige segment, slik som det er i forrige modell.

I modellen for nedoverbakker ga bratte nedoverbakker lavere fart enn slake nedoverbakker. For denne modellen kan vi se at det er motsatt: Er det slak, negativ helning vil farten være større enn dersom det er helt flatt eller svak positiv stigning. Dette gjelder for \bar{s}_f og \bar{s}_n også, ettersom koeffisientene er negative. Her er resultatene altså litt motstridende. I forrige modell førte brattere nedoverbakker til lavere fart, men i denne modellen vil bratte nedoverbakker i forrige segment bety økt fart.

Det er også interessant å se at mens forrige og neste segments stigning ser ut til å påvirke farten negativt, påvirker lengden på forrige og neste segment farten positivt. Det betyr, i følge modellen, at man har høyere fart dersom neste bakke er lang enn dersom den er kort, selv om det er en bratt oppoverbakke. De logaritmiske transformasjonene av lengdene tyder på at det er størst forskjell for de korte stigningene. Denne modellen kan også ende opp med negative hastigheter som gjør at en nedre grense bør innføres.

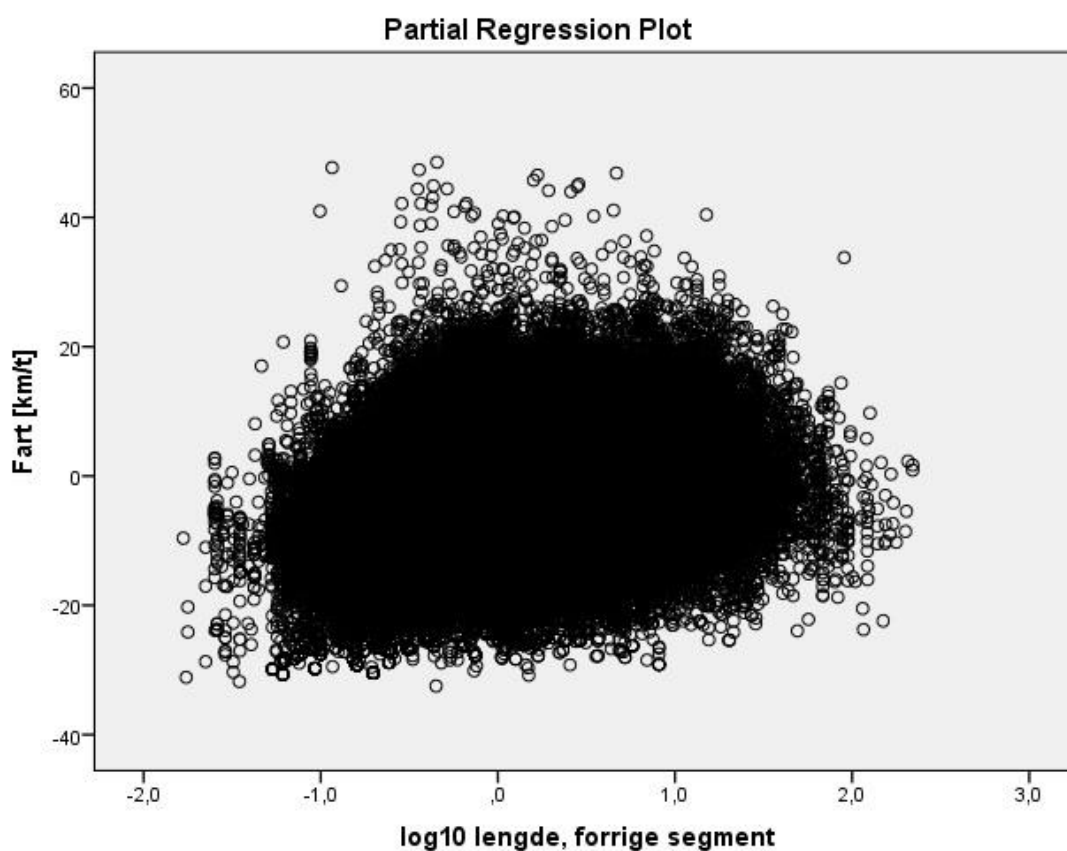
4.2.4 Statistikk for flatmarks-modellen

Modellen for flat mark inneholdt flere variabler uten at det så ut til å gi bedre forklaringsgrad sammenliknet med modellen for nedoverbakker. Tabell 4.3 viser forklaringsgrad, standardavvik og Durbin-Watson verdi for flatmarks-modellen, og vi ser at forklaringsgrad er nede på ca. 27 %, som er dårligere enn for nedoverbakke-modellen. Standardavviket er noe lavere, men fortsatt høyt. Nok en gang ser vi en lav Durbin-Watson verdi som gir sterke indikasjoner på at det er autokollinearitet i datasettet.

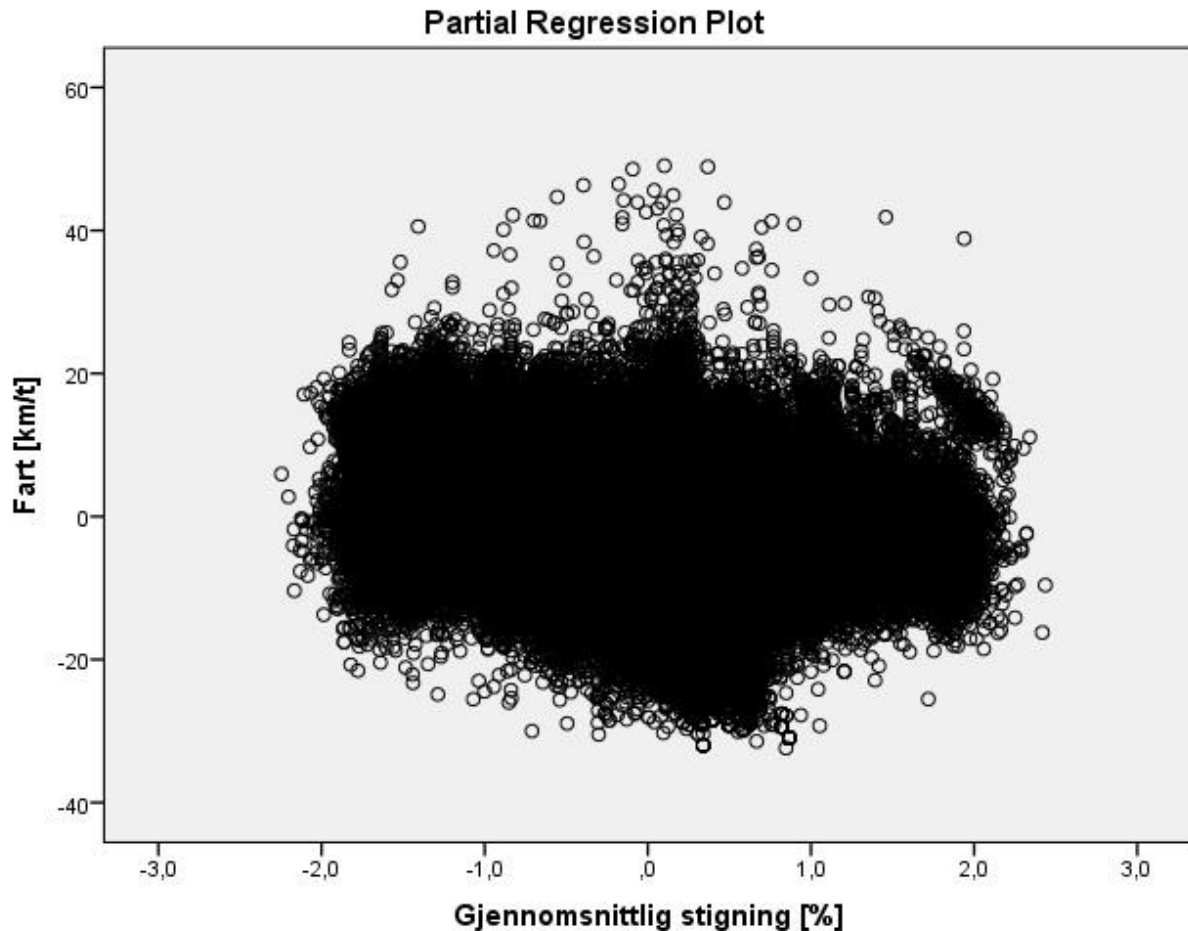
Tabell 4.3 – Statistikk for flatmarks-modellen

Modell	R	R ²	Justert R ²	Standardavvik til predikasjonene	Durbin-Watson
-2,0 % <= s <= 2,0 %	0,519	0,269	0,269	6,672	0,247

Det betyr at signifikansnivået gitt i neste tabell ikke er gyldig ettersom det er sterkt påvirket av autokollineariteten. Den lineære sammenhengen mellom variabler og fart er som for forrige modell, noe varierende og ikke alltid like tydelig. Figur 4.11 og figur 4.12 viser noen eksempler.

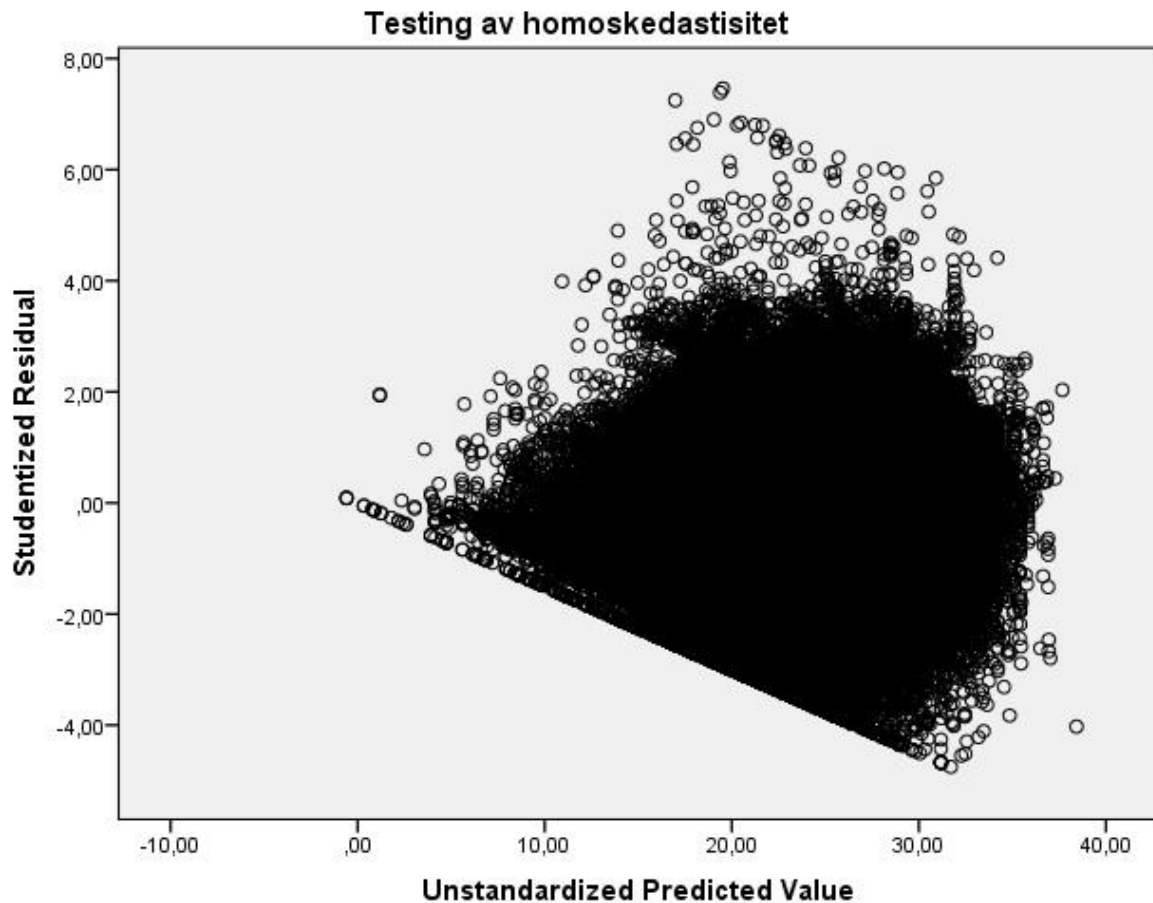


Figur 4.11 – Lineær sammenheng mellom lengde på forrige segment og fart



Figur 4.12 – Partial Regression Plot for stigning.

Figur 4.13 viser residualer mot predikert verdi og også for denne modellen er det problemer med lite homoskedastisitet. Dette er tydelig for de laveste verdiene for fart, og plottet har den samme, karakteristiske nedre grensen som forrige modell. I motsetning er det for dette tilfellet mindre predikerte verdier som er negative. Også her ble det identifisert flere utliggere uten at de hadde bemerkelsesverdige høye Cooks-verdier. Det ble eksperimentert med å fjerne noen av disse, men det ga ingen forbedringer i modellen og ble derfor ikke fjernet fra datasettet. Det var ingen multikollinearitet i variablene da alle VIF-verdiene lå nær 1,0. Residualene var normalfordelt, og dette, sammen med mer statistikk for datasettet, er gitt i vedlegg D.



Figur 4.13 – Residualenes fordeling for flatmarks-modellen . Studentized residual = residualer, Unstandardized predicted value = predikert verdi.

4.2.5 Modell for fart i oppoverbakker (Stigning > 2,0 %).

Den siste modellen tar for seg fart i oppoverbakker der stigningen er større enn 2,0 %. Sammenliknet med modellen for flat mark ser vi at de har samme variablene men med noe forskjellig koeffisienter. Modellen er gitt i formel (13).

$$V = 20,25 - \bar{s} - 1,1 * \log(l_s) - 0,27\bar{s}_f + 2,5 * \log(l_{s_f}) - 0,25\bar{s}_n + 0,87 * \log(l_{s_n}) + 0,18t \quad (13)$$

Der

V = Hastighet [km/t]

\bar{s} = Gjennomsnittlig stigning, gjeldende segment [%]

l_s = Lengde på stigningen, gjeldende segment [m]

l_{s_n} = Lengde på stigning, neste segment [m]

\bar{s}_f = Gjennomsnittlig stigning, forrige segment [%]

- l_{sf} = Lengde på stigningen, forrige segment [m]
 \bar{s}_n = Gjennomsnittlig stigning, neste segment [%]
 t = Døgnmiddeltemperatur [grader C]

Nok en gang ser vi at lengden på stigningene (både forrige, inneværende og neste) er logaritmisk transformert som betyr at en endring i lengde har størst effekt ved mindre verdier. Vi ser også at stigningslengden påvirker farten negativt, i motsetning til hva som var tilfellet for nedoverbakker og flat mark, som høres fornuftig ut. Modellen vil for meget bratte bakker gi negativ fart, som betyr at også for denne modellen bør det legges inn en nedre grense for fart.

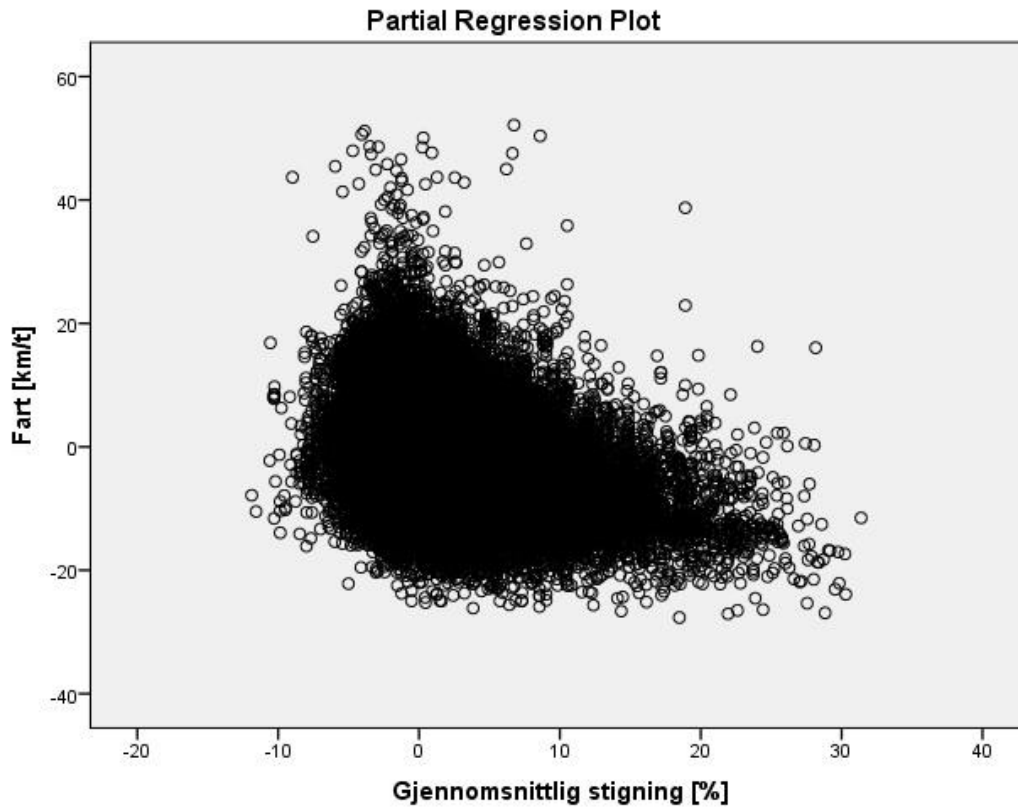
4.2.6 Statistikk for oppoverbakke-modellen

Statistiske nøkkeltall for oppoverbakke-modellen er gitt i tabell 4.4. Denne modellen ser ut til å være den med best forklaringsgrad, opp mot 0,40, men sliter med tydelige tegn på problemer med autokorrelasjon, slik som de andre. Durbin-Watson verdien er helt nede på 0,34 mens standardavviket til predikasjonene ligger på ca. $\pm 5,3$ km/t.

Tabell 4.4 – Forklaringsgrad, standardavvik og Durbin-Watson for oppoverbakke-modellen

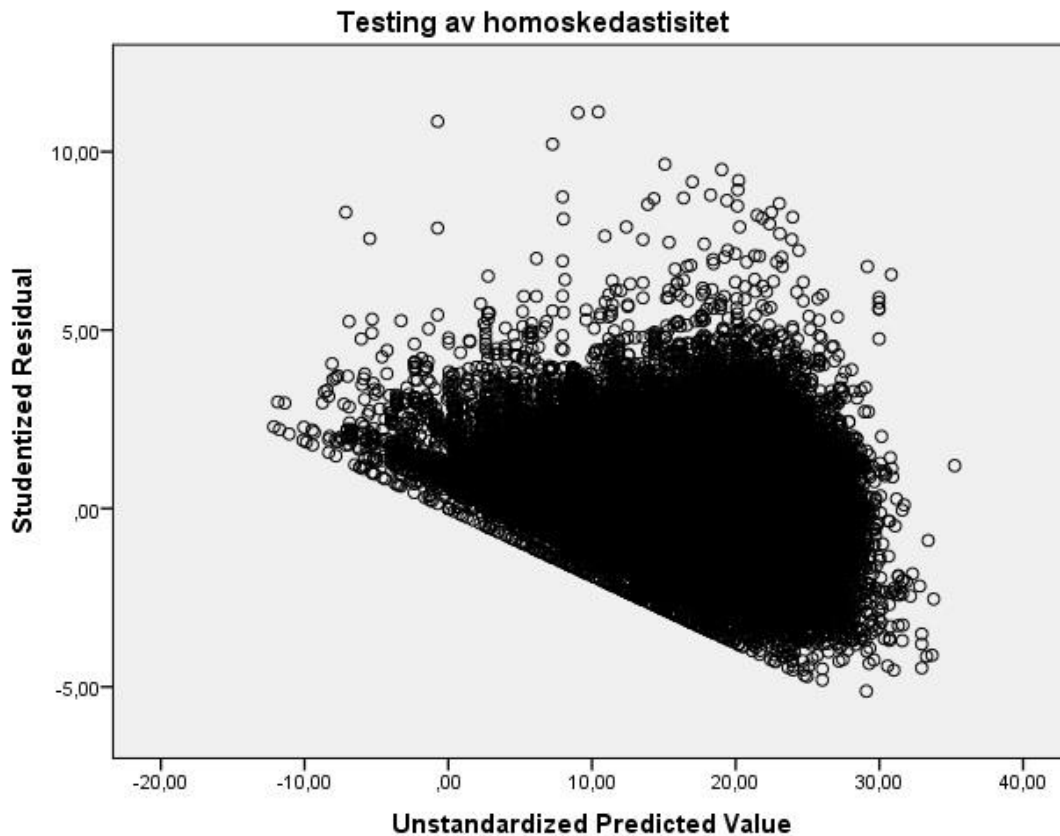
Modell	R	R ²	Justert R ²	Standardavvik til predikasjonene	Durbin-Watson
s > 2,0 %	0,633	0,400	0,400	5,302	0,342

Videre har modellen grei lineær sammenheng for de fleste variablene, men med noe variasjon. Et eksempel er gitt i figuren under der man ser et partial regression plot for den gjennomsnittlige stigningen.



Figur 4.14 – Lineær sammenheng mellom stigning og fart.

Figur 4.15 viser sammenhengen mellom residualene og den beregnede verdien, og som med de andre modellene er det også her tydelige tegn på dårlig homoskedastisitet og negative verdier. Modellen viste ingen tegn til multikollinearitet mellom de beskrivende variablene, og fjerning av utliggere og influenspunkt ga ingen forbedring. Residualene ser ut til å være greit normalfordelt, og er vist i vedlegg D, sammen med flere statistiske verdier fra regresjonsanalysen.



Figur 4.15 – Homoskedastisitet for oppoverbakke-modellen . Studentized residual = residualer, Unstandardized predicted value = predikert verdi.

4.2.7 Utelatte variabler

Det var andre variabler i datasettet som ikke ble med i noen av modellene. De fleste av disse var variabler som viste lav forklaringsgrad og ga liten til ingen forbedring til modellene dersom de ble fjernet eller lagt til. Slike variabler er det liten vits i å ha med for å begrense kompleksiteten til modellene. Eksempler på slike variabler var for eksempel tid, både på døgnet og totalt for turen. Man kunne spekulere i om man pendlere, som det mest sannsynlig er flest av i tidsrommet 07-09 og 15-17 ville hatt større fart enn sykkelturer utenfor disse tidsrommene, men det var lite som viste tegn til dette. Total tid eller tid siden start baserer seg på en antagelse om at man blir sliten jo lengre man sykler og at farten etter hvert vil synke. Det var få tegn som tydet på at dette var spesielt relevant for fartsmodellene, og det samme gjaldt total avstand og avstand fra start, der den samme antagelsen var lagt til grunn.

Maksimal stigning så ut til å være sterkt multikollinear med den gjennomsnittlige stigningen, som ikke er veldig overraskende. Den gjennomsnittlige stigningen ga bedre forklaringsgrad, og maksimal stigning ble derfor utelatt. Nedbør ble innført i modellen, men på grunn av en feiltolkning ble denne ikke brukbar.

4.3 Testing av modellene i alternativt datasett

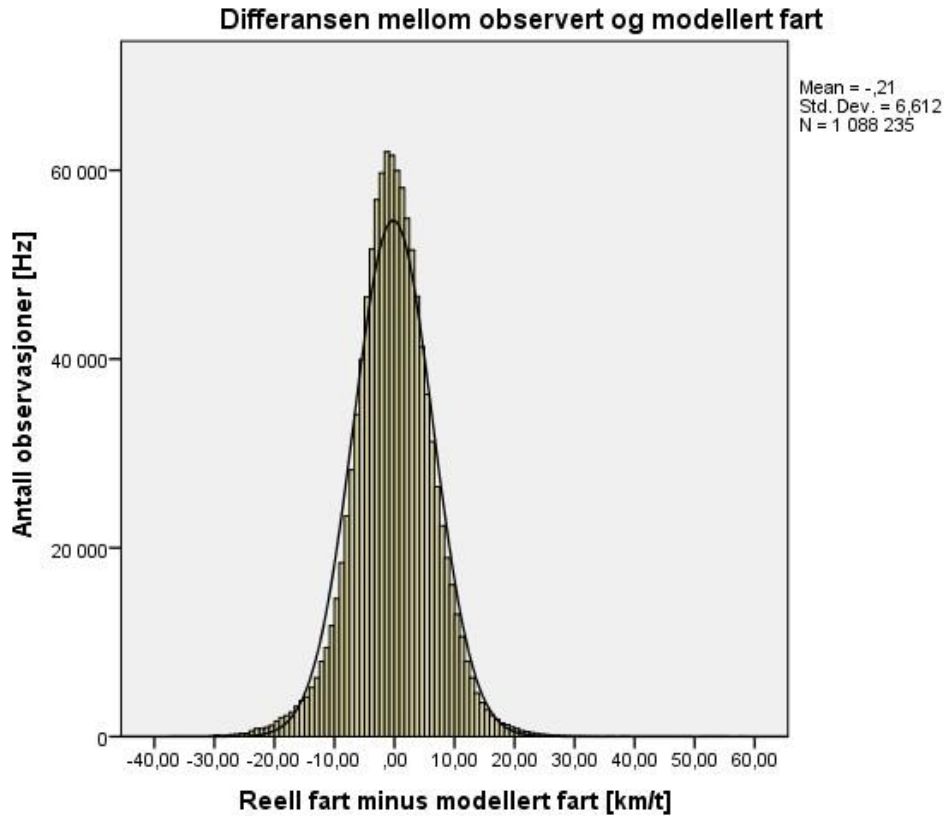
Ettersom datasettet ble delt i to etter filtreringen var det mulig å teste modellen på det ubrukte datasettet. Ved å legge modellene inn i datagrunnlaget kan man vurdere den målte farten opp mot den modellerte. For noen tilfeller ble farten beregnet til å være negativ. I slike tilfeller ble det i stedet antatt at man var på et punkt hvor det var naturlig å gå av sykkelen og leie den i stedet, eventuelt sykle meget sakte. Her ble det satt en fart på 3 km/t i stedet for de negative verdiene. Verdier som var lavere enn 3 km/t men større enn null ble ikke redigert.

Ved å gjennomføre en hypotesetest kan man se om den gjennomsnittlige forskjellen mellom modellert fart og observert fart er forskjellig fra 0,0 km/t. Ettersom det er problemer med autokorrelasjoner og dårlig homoskedastisitet er ikke t-testen helt til å stole på. Her benyttes t-testen på differansen mellom modellert og observert fart, men det er uvisst i hvor stor grad dette påvirker resultatet, men det er trolig at den beregnede t-verdien er overdrevet lav. Hypotesen er gitt under.

$$H_0: \mu_{\Delta V} = 0$$

$$H_1: \mu_{\Delta V} \neq 0$$

Der ΔV er forskjellen mellom observert fart og modellert fart, mens $\mu_{\Delta V}$ gjennomsnittet av denne forskjellen. Figur 4.16 viser fordelingen av ΔV sammen med normalfordelingskurven, og man ser at fordelingen ligger litt på den negative siden av 0 km/t.



Figur 4.16 – Fordeling av feil mellom observert og modellert fart

Tabell 4.5 viser resultatene og som en ser er $\mu_{\Delta V}$ ca. -0,2 km/t, som viser seg å være statistisk signifikant forskjellig (mindre) fra 0 km/t på 5 % signifikansnivå. Det betyr at modellen ikke greier å forutsi farten helt perfekt, men beregner den til å være litt lavere enn den observerte farten.

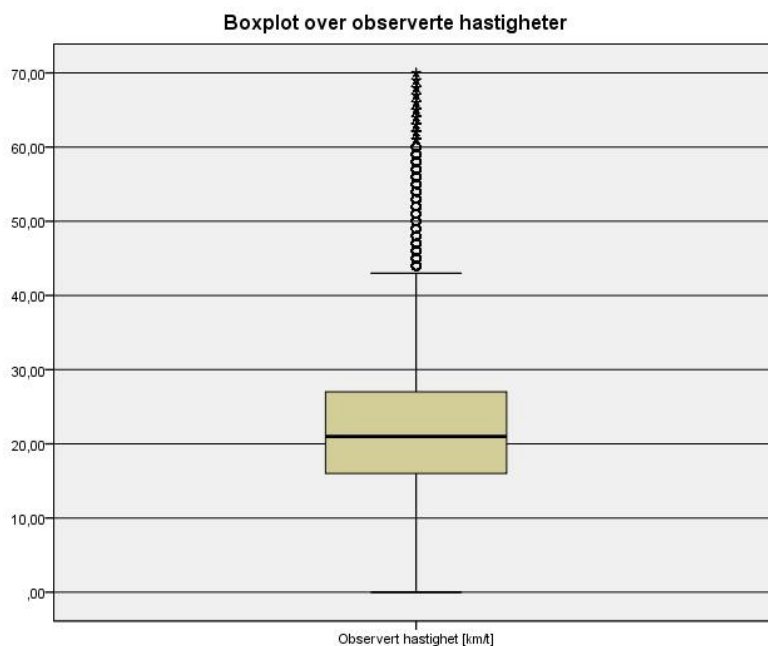
Tabell 4.5 – T-test for gjennomsnittlig differanse mellom observert og modellert fart

	Test verdi = 0		Sig. (2- halet)	Gjennomsnittlig forskjell	95 % Konfidensintervall	
	t-verdi	Frihetsgrader			Nedre grense	Øvre grense
Observed speed minus modeled speed	-32,911	1 088 234	0,000	-0,209	-0,221	-0,196

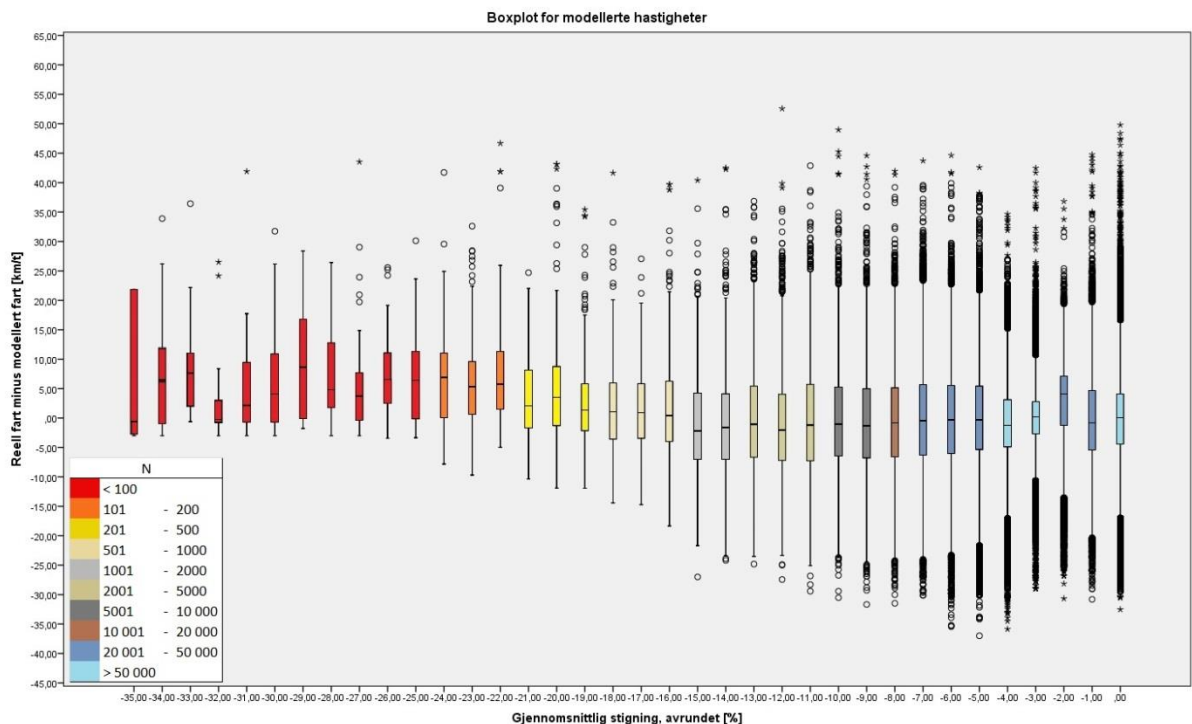
De modellerte hastighetene ligger i området 0 til ca. 41 km/t mens gjennomsnittet ligger på ca. 22 km/t. Til sammenlikning har det aktuelle datasettet en gjennomsnittsfart på 21,7 km/t og omtrent 99 % av verdiene er under ca. 43 km/t, se figur 4.17.

Tabell 4.6 – Statistikk fra modellerte hastigheter

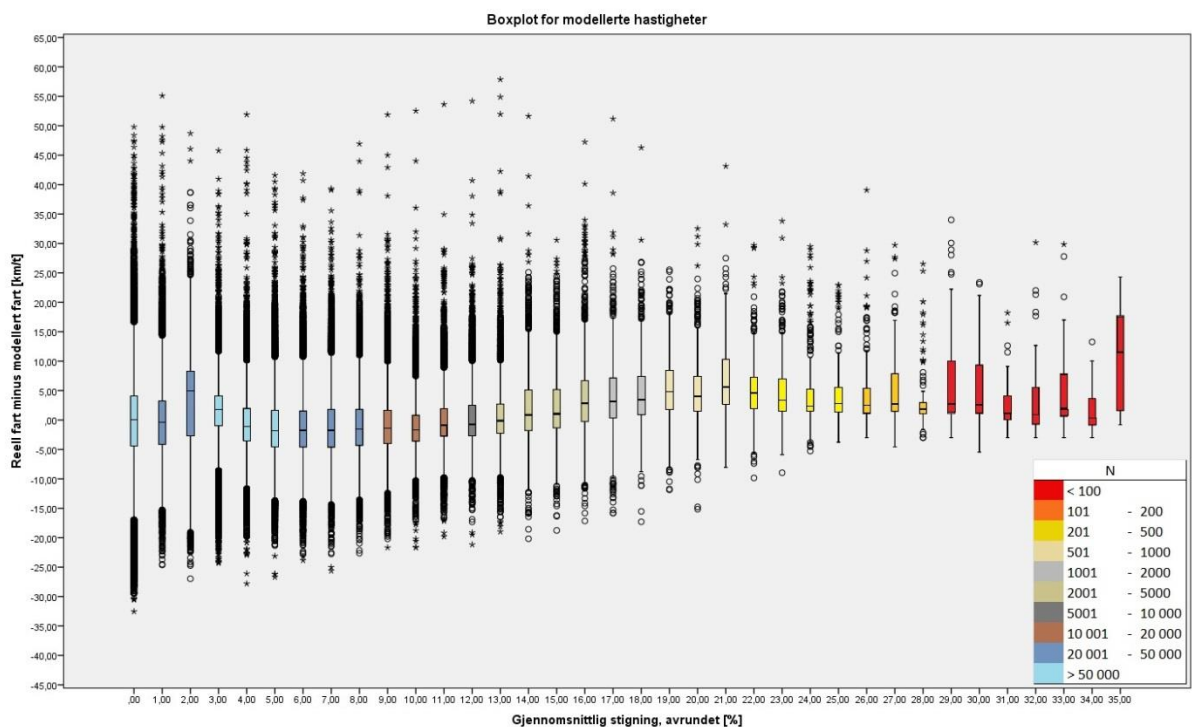
	Antall observasjoner	Minimum	Maksimum	Gjennomsnitt	Standardavvik
Modellert fart [km/t]	1 088 235	,00	41,46	21,905	5,599

**Figur 4.17 – Observerte hastigheter i det alternative datasettet.**

De høyeste modellerte hastighetene er for stigninger i området -6 til -3 %. Registrerte forskjeller mellom modellert og observert fart er gitt figur 4.18 og figur 4.19. Nok en gang ser man også at det er få observasjoner for verdier utenfor området -15 til 18 %. Ettersom det benyttes gjennomsnittlig stigning er denne effekten litt forsterket. I tillegg ser man at spredningen ser ut til å minske jo nærmere 0 % stigning man kommer, men antall utliggere ser ut til å øke. Det skyldes nok økt antall observasjoner, og ikke presisjon. Det er en plutselig spredning i differansen mellom modellert og observert fart for 2 % stigning. Dette kan kanskje ha noe med segmentinndelingen å gjøre. Man kan også legge merke til at det ser ut til å være mindre spredning for oppoverbakker enn nedoverbakker, som forventet da oppoverbakkemodellen hadde høyere forklaringsgrad. For omtrent alle stigningene ligger 50 % av de predikerte verdiene innenfor ± 5 km/t av de observerte.



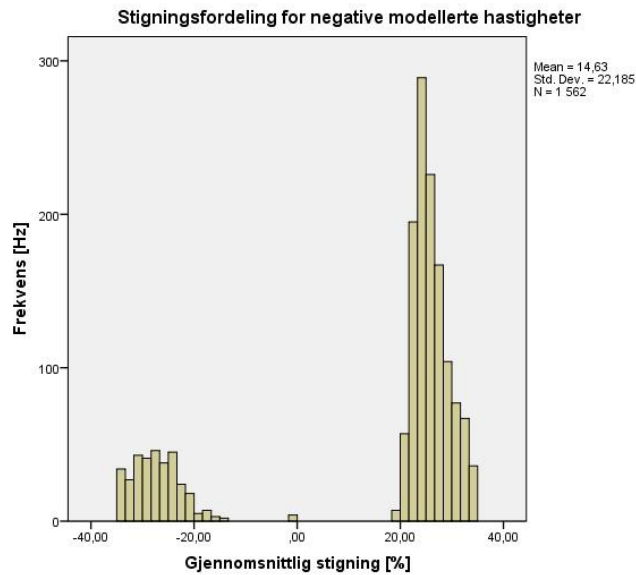
Figur 4.18 – Boxplot for modellerte hastigheter, negativ gjennomsnittlig stigning.



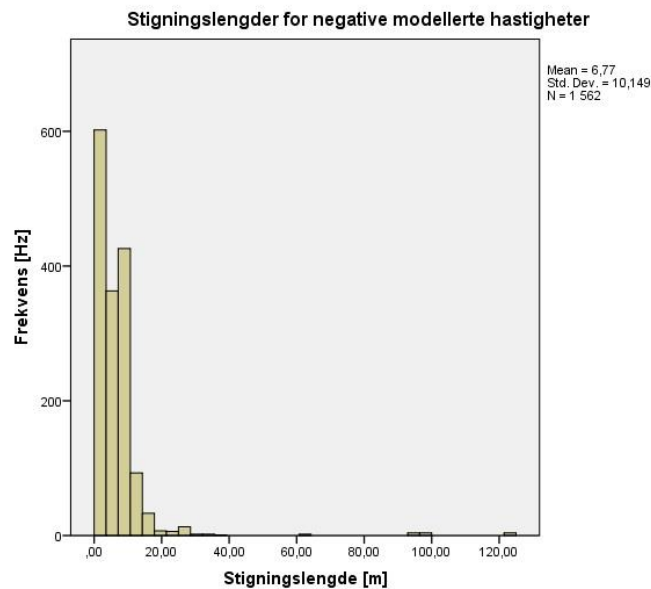
Figur 4.19 – Boxplot for modellerte hastigheter, positiv gjennomsnittlig stigning.

Det ble nevnt at modellene i noen tilfeller kunne beregne negative hastigheter. Da modellene ble testet var omtrent 1600 (av nesten 1,1 millioner) hastigheter beregnet til å være mindre

enn 0,0 km/t. Nærmere analyser av disse viste at dette i hovedsak skjedde ved meget bratte stigninger og korte segment, se figur 4.20 og figur 4.21. Der ser man at de fleste tilfellene av negativt modellert fart skjer ved $\pm 20\%$ stigning, og segmenter under 100 meters lengde, men med noen unntak.



Figur 4.20 – Stigninger ved negativt modellerte hastigheter.



Figur 4.21 – Stigningslengder ved negativt modellerte hastigheter

4.4 Implementering i nettverket for DOM Nidaros

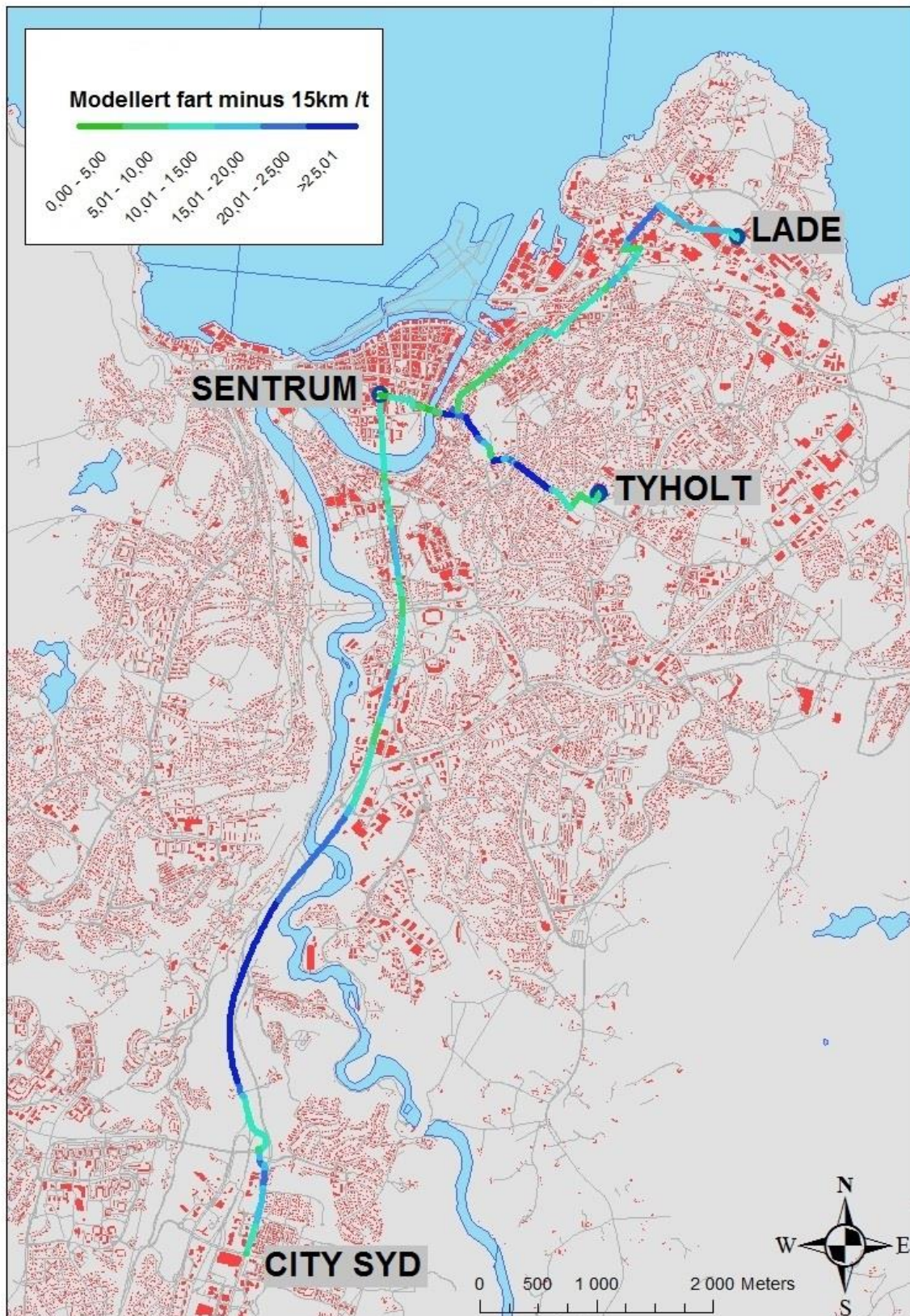
De nye modellene for sykkelfart ble testet i et nettverk for Trondheim, og det ble bestemt å se på tre forskjellige ruter: Den første gikk fra City Syd til Trondheim sentrum, den andre gikk fra Tyholt til Trondheim sentrum mens den tredje gikk fra Lade til Trondheim sentrum. Alle rutene ble analysert inn mot sentrum og tilbake igjen, og sammenliknet med en konstant fart på 15 km/t slik som benyttes i dagens trafikantnyttmodul i RTM.

De tre rutene inn mot sentrum er vist i figur 4.22 med fargekoder over differansen mellom modellert fart og en fast fart på 15 km/t. De lokalkjente vil fort se at for ruten fra City Syd og inn mot sentrum skjer deler av trafikken på E6, hvor sykler egentlig ikke har tilgang. Det skyldes i hovedsak at det ble benyttet en algoritme for å finne korteste rute avhengig av avstand, uten noen restriksjoner. Det fører til at reisetidene er noe lavere enn hva de vil være i realiteten.

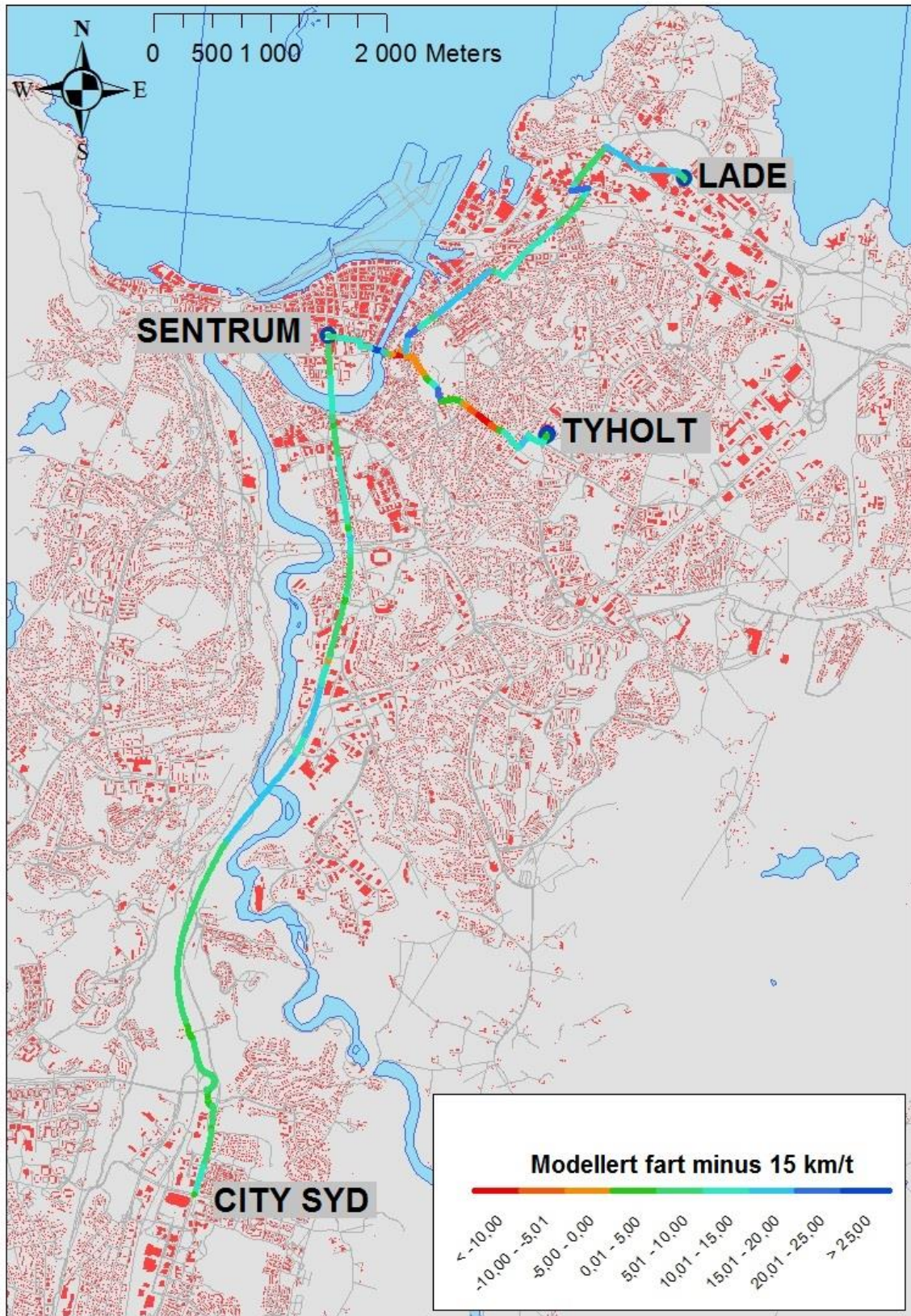
Områder merket mørkeblått er områder for den modellerte hastigheten er mye høyere enn 15 km/t. Fra figuren ser man at dette i hovedsak gjelder segmenter der det er bratte nedoverbakker. For slakere segmenter er det ikke like store forskjeller i hastigheter, som for eksempel fra Lade. Der er det ganske flatt på store deler av ruten, men den modellerte hastigheten er alltid noe høyere.

De samme rutene er vist i figur 4.23 men i dette tilfellet sykler man ut av sentrum og mot de respektive områdene. Her ser det ut til å være mindre spredning mellom hastighetene ut mot City Syd. Man ser også at det nå er negative forskjeller, der modellert hastighet er lavere enn 15 km/t. Dette gjelder spesielt opp mot Tyholt fra sentrum, hvor det er noen meget bratte bakker (Brubakken fra Bakklandet opp mot Kristiansten festning, samt Blusuvoldbakken fra Eidsvollveien opp mot Kong Inges gate). I modellene ble det også lagt inn en nedre grense på fart på 3,0 km/t, og denne ble benyttet ved noen tilfeller i de bratteste bakkene.

Tabell 4.7 viser en oversikt over reisetidene mellom de forskjellige rutene og metodene. For noen turer ser man at den gamle metoden beregner opp mot dobbelt så lang reisetid sammenliknet med den nye. Fra City Syd er det modellert en gjennomsnittshastighet på nesten 30 km/t som gir en reisetid på ca. 15 minutter. Tilbake er hastighetene noe lavere. Mot Tyholt er det tydelig at det er stigningen har stor innvirkning på den modellerte hastigheten, ettersom det er redusert reisetid mot sentrum men økt reisetid på returen. Ruten mellom Lade og sentrum har mindre forskjell i reisetid, men her er også reisetiden meget lav fra før av.



Figur 4.22 – Modellert sykkelfart inn mot sentrum fra City Syd, Tyholt og Lade.



Figur 4.23 – Modellerte hastigheter ut fra sentrum mot de respektive områdene.

Tabell 4.7 – Sammenlikning av reisetider mellom metodene.

	Lengde [m]	Gjennomsnittsfart [km/t]	Tid, ny metode [min]	Tid, gammel metode [min]	Forskjell [min]	Forskjell [%]
City Syd	Til sentrum	7937	15,0	31,7	-16,8	-52,8 %
	Fra sentrum	7937	19,9	31,7	-11,9	-37,4 %
Tyholt	Til sentrum	2394	4,8	9,6	-4,8	-49,9 %
	Fra sentrum	2394	12,1	9,6	2,5	25,8 %
Lade	Til sentrum	4415	9,7	17,7	-8,0	-45,0 %
	Fra sentrum	4415	11,1	17,7	-6,6	-37,3 %

4.5 Oppsummering

I dette kapitlet har vi lest at datasettet har store spredninger i fart. I de fleste tilfeller er 50 prosent av de målte verdiene (25 % på hver side av medianen) innenfor et intervall på ca. 10 km/t. Hastighetene ser ikke ut til å være normalfordelte for en gitt stigning, med noen få unntak. For flere stigninger er det tydelige grupperinger av data, men antall grupperinger varierer også.

Det ble utviklet tre modeller, hvor den ene modellen tok for seg nedoverbakker, den andre var for tilnærmet flate strekninger og den siste modellen var for oppoverbakker. Inndelingene tilsvarte segmentinndelingen gjort tidligere i arbeidet og modellene er presentert under.

$$V = 66,73 + 0,86\bar{s} + 6,85 * \log(l_s) + 0,16t - 31,1 * \log(\bar{s}_f + 36) \quad s < -2,0\%$$

$$V = 12,6 - 2,4\bar{s} + 2,44 * \log(l_s) - 0,2\bar{s}_f + 2,22 * \log(l_{s_f}) - 0,18\bar{s}_n + 1,47 * \log(l_{s_n}) + 0,2t \quad -2,0\% \leq s \leq 2,0\%$$

$$V = 20,25 - \bar{s} - 1,1 * \log(l_s) - 0,27\bar{s}_f + 2,5 * \log(l_{s_f}) - 0,25\bar{s}_n + 0,87 * \log(l_{s_n}) + 0,18t \quad s > 2,0\%$$

Der

V	=	Hastighet [km/t]
\bar{s}	=	Gjennomsnittlig stigning, gjeldende segment [%]
l_s	=	Lengde på stigningen, gjeldende segment [m]
l_{s_n}	=	Lengde på stigning, neste segment [m]
\bar{s}_f	=	Gjennomsnittlig stigning, forrige segment [%]
l_{s_f}	=	Lengde på stigningen, forrige segment [m]
\bar{s}_n	=	Gjennomsnittlig stigning, neste segment [%]
t	=	Døgnmiddeltemperatur [grader C]

Alle tre modellene har problemer med autokorrelasjon og homoskedastisitet, samt relativt lav forklaringsgrad. Tabell 4.8 gir noen av de viktigste verdiene fra regresjonsanalysen.

Tabell 4.8 – Oversikt over modellstatistikk

Modell	R	R ²	Justert R ²	Standardavvik til predikasjonene	Durbin-Watson
Nedoverbakke	0,575	0,331	0,331	7,189	0,268
Flatmark	0,519	0,269	0,269	6,672	0,247
Oppoverbakke	0,633	0,400	0,400	5,302	0,342

Modellene ble implementert i den andre halvdel av datasettet. Testing av modellert fart opp mot den registrerte farten viser at den gjennomsnittlige forskjellen er statistisk signifikant forskjellig fra 0 på 0,5 % nivå og den beregnede farten viste seg å være noe lavere enn det observerte. De ble også knyttet opp mot et nettverk for Trondheim for å sammenlikne reisetider opp mot den eksisterende metoden for sykkelfart i RTM. Modellene beregnet for det meste høyere hastigheter og lavere reisetider, men fra sentrum og opp mot Tyholt beregnet modellene lavere hastighet.

5 Diskusjon

I dette kapittelet skal resultatene fra arbeidet gjennomgås mer i sømmene. I den første delen er de tilgjengelige datakildene omtalt og det diskuteres blant annet usikkerheter og bias. Andre del ser på de utviklede modellene og hvordan datasettet påvirket modellene mens siste del knytter de første delene sammen ved å drøfte hvorvidt datakildene kunne blitt brukt i en rutevalgmodell.

5.1 Tilgjengelige datakilder, bias og usikkerhet

Det ble identifisert flere potensielle kilder, men til bruk i en fartsmodell var det flere av dem som ikke hadde de nødvendige dataene. Hovedsakelig hadde de fleste kildene bare tilgjengelig lengde- og breddegrader, og i noen tilfeller høydedata, men manglet enten hastigheter eller tidsverdier. Det gjorde det umulig å gjøre beregninger på farten. Garmin Connect var enestående i hvilke data og mengden data som var tilgjengelig, mens andre kilder som MapMyRide og RideWithGps bare hadde deler av den ønskede informasjonen. Andre kilder som Strava, Endomondo og RunKeeper var ikke brukbare.

5.1.1 Bias i datasettet

Det er utvilsomt bias i datasettet og det er tre hovedgrunner til det. For det første har man ingen personlige data på de som har gjennomført turene. Det kan være godt være at datasettet hovedsakelig består av mennesker innenfor et snevert utvalg av alder, kjønn, erfaring og fysisk form. Det er faktisk naturlig å anta at de fleste som har lastet opp sine sykkelturer har gjort det fordi de enten er interessert i sykling, opptatt av trening eller begge deler. Derfor er det også trolig at de fleste turene er gjennomført av erfarne syklister, som betyr at det er stor sjanse for at hastighetene som er registrert i datasettet er høyere enn for «den gjennomsnittlige syklist».

Den andre grunnen til at vi har bias i datasettet er at det er hovedsakelig turer hentet fra Oslo. Noen få er hentet fra Kristiansand og litt flere fra Trondheim, men fortsatt over 90 % av turene gjennomført i Oslo. Det er vanskelig å anslå hvor mye lokale effekter har å si, men det kan være forskjeller i både kultur og skikk (blant annet i forhold til andre trafikanter og kjøretøy), samt lokale variasjonen i vær, topografi og tilgjengelig infrastruktur som gjør at hastigheter i en del av landet kan være statistisk forskjellig fra et annet.

Den siste grunnen til bias skyldes liten kontroll på antall brukere og gjentakelser. Da dataene ble lastet ned ble det ikke registret hvem som hadde utført turen, selv om dette var tilgjengelig

på nettsiden, på grunn av tekniske vanskeligheter og tidsbruk. Det betyr at det ikke er noen oversikt over antall unike brukere eller hvilke turer som ble utført av hvilken bruker, slik at samme bruker kan ha gjennomført samme tur flere ganger uten at det er tatt spesiell hensyn til dette.

Spesielt den første grunnen gjør at hastighetene trolig er noe høyere enn hva som er antatt til å være gjennomsnittet for alle syklistene. Ruteplanleggerne som ble presentert innledningsvis i oppgaven opererte med hastigheter rundt 15-20 km/t på flatmark. Til sammenlikning gir modellen for flatmark en hastighet på ca. 23,1 km/t dersom man antar at stigningen er 0 %, segmentet er langt (1000 meter) og de tilsluttede segmentene er korte og slake (50 meter, $\pm 2,1$ % stigning). Dette er også noe høyere enn hva Parkin og Rotheram (2010) kom fram til, som også var noe høyere enn hva ruteplanleggerne presenterte. Hvis man i tillegg husker at den gjennomsnittlige, modellerte hastigheten var statistisk signifikant lavere enn den observerte, er det tydelig at hastighetene i datasettet er høyere enn hva det er for den gjennomsnittlige syklist.

Dette er noe vi har sett tydelig i datasettet. For det første er det registrert meget store hastigheter i flere tilfeller som vi ser godt fra boxplottene i kapittel 3.2 og 4.3. For det andre er det tydelige grupperinger når man ser på fartsdata for hver enkelt stigning. Disse grupperingene kan skyldes mange ting men det mistenkes at fysisk form og erfaring er to av de største grunnene. Som nevnt var det i noen tilfeller få grupper, i andre tilfeller mange grupper, som tyder på at det er flere grupperinger og flere faktorer som kan spille inn. Andre faktorer som kan forklare grupperingene kan være kjønn, hva slags infrastruktur som ble benyttet, vær og føre, utstyr eller noe så enkelt som tilgang på dusj på arbeidsplassen.

5.1.2 Usikkerheter i datasettet

I tillegg til å ha høyere hastigheter enn hva man antar er gjennomsnittet så er det også flere usikkerheter i datasettet. En har ingen kontroll over at de opplastede dataene «holder det de lover», ettersom dataene er samlet inn og lastet opp av syklisten på eget initiativ. Om den oppgitte hensikten («pendling til jobb», «hjem fra butikken») stemmer er ikke nødvendigvis tilfellet. Noen brukere har kanskje gjennomført flere turer før de laster opp alle sammen samlet, og har glemt detaljer eller lastet opp under feil kategori. Det er også noen problemer når personer kombinerer pendlingen med trening. Er det en treningstur? Er det en pendletur? Hvordan påvirker dette farten? Det er naturlig å tro at flere av turene som er brukt i analysen ligger i en gråsoner eller rett og slett ikke er representativ som en pendletur, selv om den er

lastet opp under den kategorien. Det fører til at mange observerte fartsverdier kan være høyere enn det man forventer fra en pendletur. Det er heller ikke tatt hensyn til hvilken dag turene er gjennomført på. Det betyr at flere turer kan være fra lørdager, søndager eller helligdager, uten at dette er undersøkt nærmere.

Det er også usikkerheter rundt nøyaktigheten til de registrerte posisjonene i datasettet. Langt i fra alle deltagerne har benyttet samme type mottager, og forskjellige mottagere kan ha forskjellig nøyaktighet. I litteraturkapittelet kunne vi lese at det er snakk om en nøyaktighet på ± 2 til ± 10 meter, som er ganske vidt område. Høydedata i datasettet ble oppgitt i 0,2 meters intervaller som tyder på at det skjer en avrunding til nærmeste 20 centimeter. Om dette skjer når mottakeren samler inn data eller når data hentes ut fra mottakeren er uvisst, men det er naturlig å anta at det skjer under dataloggingen for at enheten kan spare minne. Derfor er det grunnlag for å anta at det er store usikkerheter rundt de målte høydedataene.

Dette problemet med nøyaktighet ser man tydelig for lave stigninger. Stigningene ser ut til å være normalfordelt helt til man nærmer seg området -1,5 til 1,5 %. Antall observasjoner går ned mens det er en klar overvekt av tilfeller hvor det ikke er registrert noen endring i høyde, som fører til at stigningen blir 0 %. En grunn til dette kan være at endringen i høydedata ikke er stor nok til å bli registrert på grunn av det 0,2 meters avrundingsintervallet. Dette blir tydelig dersom man ser på antall observasjoner for hver stigning i figur 4.1 og i figur 4.2. Tallene er gjengitt i tabell 5.1. Der ser vi en kraftig nedgang i antall observasjoner for -1, 1, og 2 %, mens 0 % har meget mange observasjoner.

Tabell 5.1 – Antall observasjoner for små stigninger

Stigning [%]	Antall observasjoner
-3,0	73 086
-2,0	71 927
-1,0	28 913
0,0	329 594
1,0	22 765
2,0	44 658
3,0	73 393

Problemet blir enda tydeligere dersom man ser for seg en dataenhet som logger posisjonen hvert sekund, slik noen enheter gjør. Sykler man med konstant fart på 5,56 meter per sekund (20 km/t) i en svak oppoverbakke med 1,5 % stigning vil man klatre 0,083 meter per sekund.

Denne økningen i høyde er alt for liten til å bli registrert i dette 0,2 meters intervallet, og blir derfor avrundet. Tabell 5.2 viser et teoretisk eksempel på hvordan dette kan påvirke registrerte stigninger mot de virkelige. Forskjellen i gjennomsnittlig stigning er ikke stor, og for noen tilfeller (jevn fart, jevn stigning) vil den bli mindre. Det finnes sikkert tilfeller hvor forskjellen også er større, men det er trolig denne avrundingen som er årsaken til at det er registrert langt flere punkt med 0 % stigning enn punkt med ca. -2 % til 2 % stigning.

Man kan også se at denne avrundingen kan føre til problemer med segmentinndelingen som ble gjennomført tidligere i oppgaven. De ekte stigningene for målepunkt 1 til målepunkt 10 er definert som tilnærmet flat strekning, men på grunn avrundingen vil det i stedet bli definert som oppoverbakke og flat mark om hverandre i segmentinndelingen. Dette er vist med fargekoder, der grønn betyr at målepunktet blir registrert som oppoverbakke mens rød betyr flat mark. Dette er et uheldig problem som kan oppstå ved flere situasjoner, men er mest tydelig for små stigninger og høye hastigheter.

Segmenteringen som ble gjennomført var for å ha en forenklet kobling mot det som skjedde før og etter den inneværende stigningen. Denne segmenteringen var kanskje i enkleste laget: Dersom en oppoverbakke (over 2 % stigning) startet ganske slakt, men ble brattere etter hvert, ble det fortsatt definert som ett segment. En bedre metode kunne vært å dele oppover- og nedoverbakker inn etter hvor bratte stigningene var, som slake, middels og bratte bakker. Det hadde ført til større nøyaktighet på gjennomsnittlig stigning og stigningslengde.

Tabell 5.2 – Teoretisk eksempel på avrunding av høydedata. Rød farge = registrert som flatt, grønn farge = registrert som oppoverbakke.

Tid [s]	Stigning [%]	Fart [m/s]	Endring i høyde [m]	Høyde [moh]	Høyde, avrundet [moh]	Endring i avrundet høyde [m]	Differanse	Stigning, avrundet [%]
0				0,000	0,00			
1	1,50 %	5,6	0,084	0,084	0,00	0,00	0,084	0,00 %
2	1,80 %	5,4	0,097	0,181	0,20	0,20	-0,103	3,70 %
3	2,00 %	5,3	0,106	0,287	0,20	0,00	0,106	0,00 %
4	2,00 %	5,2	0,104	0,391	0,40	0,20	-0,096	3,85 %
5	2,00 %	5,1	0,102	0,493	0,40	0,00	0,102	0,00 %
6	1,70 %	5,0	0,085	0,578	0,60	0,20	-0,115	4,00 %
7	1,70 %	5,0	0,085	0,663	0,60	0,00	0,085	0,00 %
8	1,50 %	5,0	0,075	0,738	0,80	0,20	-0,125	4,00 %
9	1,50 %	4,8	0,072	0,810	0,80	0,00	0,072	0,00 %
10	2,00 %	4,7	0,094	0,904	1,00	0,20	-0,106	4,26 %
11	2,50 %	4,6	0,115	1,019	1,00	0,00	0,115	0,00 %
12	3,00 %	4,5	0,135	1,154	1,20	0,20	-0,065	4,44 %
Gjennomsnitt	1,93 %							2,02 %

En måte man kunne prøvd å redusere denne usikkerheten ville vært og koblet datasettet opp mot applikasjonen «Norgeskart», som drives av Statens kartverk. Der er det mulig å legge inn GPS-data for lengde- og breddegrad og få ut høydedata for det aktuelle punktet, og på den måten har man en metode for å sjekke høydedata. Ulempen er selvsagt at dersom det er feil i lengde- og breddegradene vil den verdien man får ut fra Norgeskart også være feil. Det kunne også ha vært mulig å benytte en form for filtrering eller algoritme for å forutsi hva den virkelige stigningen er, kanskje noe i samme kategori som et Kalman-filter. Et Kalman-filter benytter input fra en sensor samt forrige observerte verdi for å komme med et anslag på neste verdi. På den måten har man en alternativ verdi man kan sammenlikne den oppgitte stigningen med, og kanskje få glattet ut høydedataene mer. Det er uvisst om det allerede foregår en slik filtrering eller glatting i dataloggingen allerede.

Det er også usikkerhet rundt fartsverdiene. Allerede i innledende filtreringer var det flere tilfeller av urealistiske hastigheter, og dette fortsatte videre i analysen. Fartsverdiene antas å være omtrent like nøyaktige som posisjonsbestemmelse i lengde- og breddegradene, så det er trolig flere verdier som ikke er helt riktige. Det er også flere hastigheter som er registrert til 0,0 km/t som burde vært analysert. Det antas at disse lave hastighetene skyldes stopp, mest sannsynlig på grunn av trafikk. Andre årsaker kan være at de har stoppet for å handle, prate

med kjente, sjekke ting på mobiltelefon, sykkel eller annet utstyr, tatt en liten pause på toppen av bakken for å drikke eller liknende.

Usikkerheten rundt slike verdier ville vært mulig å redusere dersom man koblet datasettet opp mot vegnettet. Da kunne man for det første enkelt ha sett dersom stoppet skjedde i eller nær et kryss, eventuelt identifisert andre grunner til hvorfor syklisten stopper opp.

5.2 Metode og modeller

Både arbeidsmetoden og modellene består av flere steg hvor det er gjort antagelser eller vurderinger som bør diskuteres. Noen ganger har avgjørelsen blitt tatt i det problemet oppstod, i andre tilfeller har det blitt gjort mer arbeid rundt antagelsen før arbeidet gikk videre.

5.2.1 Filtreringsprosessen

Filtreringsprosessen bød tidlig på utfordringer. Ønsket om å vurdere hovedsakelig pendlerturer og unngå de typiske treningsturene gjorde at man først burde ha en idé om hva en typisk pendlertur er. Ser man til store sykkelnasjoner som Danmark, Tyskland og Nederland er en pendlertur noe for alle og enhver, fra minstemann på fem år som akkurat har tatt av støttehjulene, til bestefar på sytti som skal til butikken. Det brukes enkle, klassiske «bysykler», hvor folk sykler i alt fra dress til tradisjonelle nasjonaldrakter, gjerne uten hjelm. I Norge er det nok noe annerledes. Om morgenen er sykkelstiene fulle av personer i trange sykkelshorts med hjelm, briller og sykkelcomputer, gjerne på en relativt ny «racersykkel». Likevel er det et ønske om at alle skal benytte seg mer av sykkelen til hverdagslige formål. Det hele gjør det vanskelig å definere hva som skal være «den representative gjennomsnittssyklist». Norske sykkelhastigheter som de er i dag er nok noe høyere sammenliknet med landene nevnt over, nettopp på grunn av den trolige overvekten av «elitesyklister».

Det har ført til at det i filtreringsprosessen ble lagt inn ganske liberale grenser for fart. Et tak på 70 km/t er nok for høyt, og de fleste av de identifiserte utliggerne i regresjonsanalysen var målepunkt der det var registrert meget høye hastigheter (over 60 km/t). Men det er fortsatt vanskelig å si, ettersom man ikke har noen gode data på hva naturlige minimums og maksimumshastigheter er for norske gjennomsnittssyklister. Ideelt sett skulle det vært lagt inn øvre og nedre grenser for fart for hver enkelt (avrundet) stigningsverdi. For eksempel vil kanskje 30 km/t i 9 % stigning vise seg å være meget urealistisk, men slike verdier finnes i datasettet. Grunnen er rett og slett at farten ikke bare er avhengig av stigningen der og da, men også hva som har skjedd i forrige element, og hva som skjer i neste element. Å skulle gå igjennom alle hastighetene og vurdere for hver enkelt om den er «representativ» er vanskelig når man har lite materiale å støtte seg på, samt at det ville krevd utrolig mye arbeid på grunn av den store datamengden.

Stigninger på ± 35 % er heller ikke en spesielt streng grense. Som nevnt og vist kan avrundingsmetoden til mottakerne føre til at noen stigninger blir registrert til å være større enn det de egentlig er. I det endelige datasettet var det registrert om lag 3700 målepunkt som hadde en stigning der absoluttverdien var større enn 25 % men mindre enn 35 %. Det utgjør om lag 0,3 % av det endelige datasettet som ble brukt i regresjonsanalysen. Setter man i stedet grensen på 20 % vil ytterligere 5000 målepunkt forsvinne, men man har fortsatt igjen nesten 99,2 % av datasettet. Det er lite trolig at fjerning av disse ville utgjort store forskjeller på regresjonsanalysen, men det bør undersøkes.

5.2.2 Modellene

Forenklingen som ble gjort i segmentinndelingen ble også gjort i modellutviklingen. Det betyr at de fleste beskrivende variablene er gjennomsnittsverdier eller totalverdier for segmentet, og ikke den direkte stigningen man befinner seg på. Denne forenklingen passer greit inn i RTM der høydeverdiene er lagret i nodene (endepunktene til hver linje), som gjør at stigningen er konstant for hver lenke i nettverket. I den virkelige verden er ikke dette tilfellet, hvor en stigning neppe vil være helt jevn, men med tanke på usikkerheten i høydedataene nevnt i kapittel 5.1.2 ser man at den gjennomsnittlige stigningen vil være mer representativ enn direkte stigning.

Modellene har også en annen grov tilnærming og det er at farten for et segment antas å være konstant. Slik vil det heller ikke være i virkeligheten. Både i nedoverbakker, flat mark og oppoverbakker vil det være akselerasjon eller retardasjon nesten hele tiden. For nedoverbakker vil farten sannsynligvis øke jo lengre ned i bakken man kommer, mens for oppoverbakker vil farten synke jo lengre opp i bakken man kommer. Samtidig er det uvisst hvor lenge «effekten» av å ha en nedoverbakke i forrige segment vil ha ut i inneværende segment. En mulighet for å se på dette hadde vært å innføre en variabel som fortalte hvor langt inn i et segment man var. Eventuelt kunne man sett på en akselerasjonsmodell eller variabel som gjorde samme nytten.

For modellen for nedoverbakker ser vi at det er en meget høy verdi på konstanten og koeffisienten for gjennomsnittlig stigning for forrige segment. Dette er også den eneste modellen der det ble gjennomført en logaritmisk transformasjon av gjennomsnittlig stigning for forrige segment. I tillegg er det den eneste modellen der hverken lengden på forrige segment, lengden på neste segment eller stigningen til neste segment er med. Dette kan nok

skyldes segmentinndelingen som betyr at forrige segment enten er tilnærmet flatt eller oppoverbakke, som gjør at farten inn i nedoverbakken trolig er lav i utgangspunktet.

Kanskje det mest interessante er at den gjennomsnittlige stigningen har positivt fortegn men variabelen vil alltid være negativ ettersom det er snakk om nedoverbakker. Det betyr at man holder en lavere fart i bratte nedoverbakker sammenliknet med slake nedoverbakker. Det samme ser vi fra figurene som viser fart og spredningen av denne. Grunnen til dette er trolig sikkerhet og trygghet. Mer erfarne mennesker eller folk som har mer «guts» vil kanskje holde høyere fart. For bratte bakker vil man ha en høy akselerasjon som gjør at man ønsker å holde lavere fart for å være mer i kontroll. For slakere bakker vil ikke akselerasjonen være like høy og det føles kanskje tryggere å sykle i slike bakker. Stigningslengden ser vi er logaritmisk, som betyr at endringer for lave verdier gir større utslag enn endring for større verdier. Det betyr også at etter et visst punkt vil ikke farten øke nevneverdig selv om stigningslengden gjør det. Det skyldes trolig balansen mellom rullemotstand, vindmotstand og akselerasjon i forhold til hvor fort man sykler. Både rullemotstand, akselerasjon og vindmotstand vil være delvis unik for hver syklist, ettersom de vil variere med kroppslige egenskaper som vekt og fasong, og tekniske egenskaper som selve sykkelen, underlag og føre. Det er også trolig at erfaring vil være relevant i for fart i nedoverbakker, der kanskje de mest erfarne holder høyere hastigheter enn nybegynnere.

Som nevnt ser vi at temperatur ser ut til å ha en effekt. Det er riktignok døgnmiddeltemperaturen, som i tillegg er målt på en værstasjon i nærheten av det respektive sentrum. Det betyr at denne ikke er nøyaktig for hver tur, ettersom temperaturen 0730 på morgenen kan være betraktelig forskjellig fra døgnmiddeltemperaturen. Det kan i tillegg være lokale variasjoner som gjør at dette ikke er en nøyaktig variabel i det hele tatt, og at den fungerer heller som en indikasjon på føre, forhold eller vær generelt. Det ble også lagt inn nedbør som en variabel, men her ble det begått en feil. Den oppgitte verdien som ble hentet fra MET for en spesiell dag var i praksis nedbøren for forrige dag. Dersom dette hadde vært oppdaget tidligere og fikset på kunne kanskje nedbør vært en mer presis variabel enn temperatur, eventuelt gitt en bedre indikasjon på føret. Likevel er det mer til føre enn bare temperatur og nedbør, som for eksempel vind og statusen på underlaget. Er det snø, er det brøytet, is, strødd, vått og glatt eller tørt og bart? Motvind vil definitivt være en viktig faktor, men vil i praksis være vanskelig å modellere på grunn av store lokale variasjoner (Bygninger, retning, kast, etc.). Skal dette implementeres i RTM bør det heller vurderes å legge inn en dummy for årstid, som gir en reduksjon eller økning i reisetid.

Modellene for flatmark og oppoverbakker er ganske like, men med noen få unntak. Variablene er de samme og koeffisientenes fortegn er like med unntak av stigningens lengde. For oppoverbakker er stigningslengde negativ men for flatmark er den positiv. Det vil si at man sykler saktere og saktere dersom lengden øker i oppoverbakker, men raskere og raskere dersom lengden øker på de tilnærmet flate strekningene. Dette er en naturlig antagelse og det gir god mening. Konstanten og koeffisientene varierer litt i verdi mellom de to modellene, men alt i alt har man større hastigheter dersom forrige segment var en nedoverbakke, slik at man har mulighet til å bygge opp litt fart. Begge disse modellene har en tydelig svakhet i det at både forrige segments lengde og neste segments lengde har positiv koeffisient. Det betyr dersom forrige og/eller neste segment er en bratt oppoverbakke vil hastighetene øke i takt med lengden på segmentet. Det er motstridende med hva modellen for oppoverbakker viser, der stigningslengden påvirker hastighetene negativ. Dette blir delvis balansert ved koeffisientene for de tilhørende segmentenes stigning, men er likevel merkelig. Det kan være at syklister «gir på» litt ekstra dersom de ser at neste bakke er en lang motbakke for å ha med seg mest mulig fart inn i bakken, men dette må undersøkes nærmere. Man skulle tro at de var villige til å gi på dersom det var en kort bakke også, slik at det ble minst mulig fysisk arbeid. I oppoverbakker vil det også være en stor grad av personlige forskjeller. For eksempel vil trolig fysisk form og muskelmasse, samt kjønn og også kanskje reisehensikt ha mye å si for hastighetene. Om man er villig til å bli svett eller ikke, kan være en stor faktor (om ikke en av de største?) for farten opp en bratt bakke.

Det er uvisst hvordan akkurat dette problemet med koeffisientene kunne løses, men man kunne undersøkt muligheten for å innføre en slags dummyvariabel for forrige og neste segment. Variabelen forteller noe om de tilhørende segmentene er negative eller positive, slik at modellen kunne justert seg etter dette.

5.2.3 Effekten av å benytte sykkelbart i stedet for avstand i RTM

Fra implementeringen i nettverket til DOM Nidaros ser modellene ut til å gi ganske høye hastigheter. Snittfartene for de tre rutene lå i området 23 - 30 km/t med ett unntak, som var fra sentrum opp mot Tyholt. Dette er en strekning med meget bratte bakker (20 % stigning flere steder), men likevel er det modellert en snittfart på nesten 20 km/t som virker lite trolig. Dette kommer trolig av noen av forenklingene som er gjort, hvor man antar konstant fart på en strekning.

Likevel bør man legge merke til de store forskjellene i reisetid dersom man bruker modellert fart i stedet for 15 km/t. Selv om den modellerte farten antas å være noe høyere enn hva som er representativt er det fortsatt områder hvor det skiller over 10 km/t mellom hastigheten. Det fører til at reisetidene i noen tilfeller er nesten halvert, som gjør at det er betydelige forskjeller mellom de to metodene. Dersom man benytter reisetid i stedet for avstand i de generaliserte kostnadene, samt de nye modellene for fart, vil det bli mye mer attraktivt å reise mellom City Syd og Trondheim sentrum på sykkel., ettersom reisetiden viser seg å være lavere. Det samme gjelder mellom Lade og sentrum, men ikke for Tyholt. Der beregner modellen at man bruker under halvparten så lang tid fra Tyholt til sentrum som fra sentrum til Tyholt. Det er mer realistisk enn hva den gamle modellen antar. Ettersom Tyholt ligger omtrent 110 meter over havet mens sentrum ligger ca. 5-10 meter over havet er det meget lite trolig at man vil bruke like lang tid på å sykle ned mot sentrum som opp mot Tyholt. Dette tar ikke den gamle metoden hensyn til i det hele tatt, og dette er en tydelig svakhet for dagens metode.

Det er tydelig at dersom de nye modellene for sykkelfart implementeres, samt at man benytter reisetid og ikke avstand i de generaliserte kostnadene, nyttefunksjonen og rutevalget vil det også påvirke reisemiddelvalget. Til hvilken grad dette kommer til å skje er vanskelig å si, ettersom det alltid vil være behov for kalibrering og validering når man oppdaterer RTM, men noen områder kommer til å bli mer attraktive, mens andre kommer til å bli mindre attraktive.

Som nevnt er de modellerte hastighetene noe høye. Det er også ikke tatt hensyn til rutevalget, men kun benyttet den korteste mulige ruten. Det er trolig at syklister ville valgt andre ruter, spesielt ut mot City Syd og kanskje opp mot Tyholt. Dersom reisetid skal være en faktor i de generaliserte kostnadene eller LoS-dataene er det derfor viktig at rutevalget er riktig modellert. Det er fordi reisetid ikke bare er avhengig av avstand, men også stigning. Avstand og stigning vil være avhengig av hvilken rute man velger, men rutevalget kan igjen være avhengig av reisetiden på de forskjellige rutene. Derfor må det trolig itereres mellom reisetid og rutevalg for å finne det mest akseptable ruten for syklister.

5.2.4 Statistiske svakheter og manglende variabler

Alle de tre modellene har svakheter ettersom de sliter både med autokollinearitet og dårlig homoskedastisitet. Autokollineariteten skyldes trolig at det er benyttet punktdata som ikke er uavhengige. Det er fordi målepunktene er samlet inn kontinuerlig når en person samler data,

og da er farten i målepunkt n sterkt avhengig av hva som skjedde i målepunkt n-1, selv om egenskapene til målepunktene (stigning, osv.) kan være forskjellige.

Autokollineariteten kan også skyldes manglende variabler. Det betyr at mye av den variasjonen man ser i fart ikke kan beskrives ved hjelp av de variablene som er i datasettet. Med en forklaringsgrad på 0,27 til 0,40 er det tydelig at det mangler vesentlige faktorer for å kunne beskrive farten. Dette kan også være årsaken til den dårlige homoskedastisiteten. Grupperinger som kan ligge skjult i datasettet, som blant annet kjønn, alder og fysisk form er ikke gjort rede for, og kan være avgjørende for farten. Eventuelt kan det komme av at de lineære sammenhengene i noen tilfeller er tynne. Bedre transformasjoner med færre utliggere kan også være en mulighet. Trolig vil både homoskedastisiteten og autokollineariteten bedres ved introduksjonen av flere beskrivende variabler.

Begge disse problemene fører til problemer for de andre statistiske testene som F-testen og t-testen. Det gjør at man ikke har noe godt statistisk grunnlag til å kunne forsvare modellene. Selv om testene viser at alle de indentifiserte variablene er signifikante, er det stor fare for at testene viser overdrevet gode resultater. Vi ser også fra testingen opp mot det andre datasettet at modellen ikke er helt perfekt. Differansen mellom modellert fart og observert fart er forskjellig fra null.

Til sammen tyder dette på at det mangler informasjon i datasettet. Mange av disse variablene har allerede blitt lett omtalt, men for videre diskusjoner kan det være naturlig å dele de manglende variablene inn i to grupper:

1. Variabler som forteller mer om syklisten.
2. Variabler som forteller mer om ruten de har syklet på.

Variabler om syklistene kan omfatte mye. Både alder og kjønn kan være tydelig indikatorer, det samme vil kanskje fysisk form og sykkelerfaringen. Andre interessante punkt kan være om de har barn og sykler med vogn, tilgang på dusj og skifterom på jobben, noen form for godtgjørelse eller kampanje fra arbeidsgiver for å øke sykkelbruken, om de har medlemskap i sykkel- eller treningsklubb, hva slags utstyr de benytter seg av, hvor de bor (er det for eksempel tydelige forskjeller mellom byer?) eller hvorfor de sykler (Er det en forskjell mellom personer som sykler for å være miljøbevisst mot personer som sykler for å få fysisk aktivitet?).

Variabler som handler om ruteegenskapene kan også være interessante. Kryss er et naturlig stoppested, som gjør at man ofte må bremse inn mot krysset for så å bruke tid på å bygge opp fart igjen, men finnes det flere? Er det forskjell mellom kryss og rundkjøringer? Holder man høyere hastigheter i noen kryss på grunn av bedre utforming, lavere trafikkmengde eller mer passende avviklingsplan? Er det forskjell på uregulerte og lysregulerte kryss? Det kan også være forskjeller avhengig av hva slags infrastruktur man sykler på. På gang- og sykkelsti er det naturlig å anta at farten blir noe lavere fordi man må vise hensyn til fotgjengere, mens på sykkelveg kan man kanskje holde høyere hastigheter fordi de er forbeholdt sykler. Sykler man på vegen kan det være hastigheten synker noe sammenliknet med sykkelveg for sikkerheten, eller det kan være den øker for at syklisten skal holde følge med annen trafikk. Andre egenskaper som vedlikehold på vinterstid og gatelys kan kanskje ha en innvirkning.

Skal man ha tak i personlig informasjon nytter det ikke bare å benytte seg av offentlige tilgjengelige kilder, men man må utvide analysen med spørreundersøkelser eller intervjuer. På grunn av personvern og nettsikkerhet er det tvilsomt at alle er like interessert i å dele slike typer data. Informasjon om ruteegenskapene er delvis tilgjengelig i datasettet men krever at man knytter alle turene opp mot et detaljrikt nettverk. Dette har flere av studiene som ble nevnt i litteraturstudiet gjort med gode resultater, for eksempel ved hjelp av GIS.

5.3 Bruk av datakildene til utvikling av rutevalgmodell

For å benytte datasettet til å modellere rutevalg er man ikke avhengig av hastighetsverdier eller tidsverdier slik som man er for en fartsmodell. Ettersom man i tillegg kan benytte seg av Kartverkets applikasjon for å hente ut høydedata trenger man egentlig bare et datasett som inneholder lengde- og breddegrader, så lenge datasettet består av representative turer. Det betyr at man potensielt har et enda større datagrunnlag enn det man har for fartsmodellen.

Fra litteraturstudien ser man at mange av faktorene for rutevalg ser ut til å være fysiske egenskaper til sykkelnettverket: Stigning, infrastruktur, kryss og vegtype er alle viktige faktorer. Andre faktorer som fartsgrenser og trafikkvolum er ikke direkte fysiske egenskaper, men har en tilknytning til nettverket. Alle disse egenskapene kan altså beskrives i et GIS. Det gjør at man kan koble GPS-dataene opp mot vegnettverket og analysere sammenhengene. Reisehensikten har også i noen tilfeller hatt innvirkning på rutevalget. Dette bør man ta hensyn til når dersom man benytter seg av 4-steg metodikken der O-D matrisen varierer for reisehensikt, og det bør også tas hensyn til i utviklingen av rutevalgmodellen.

Det kan riktignok oppstå problemer med tanke på nøyaktigheten til dataene. En nøyaktighet på ± 2 meter kan føre til problemer dersom man har tenkt å benytte seg av en Map matchmaking metode for å koble GPS-dataene opp mot et veg- og sykkelnettverk i et GIS. Dersom sykkelfeltet eller sykkelvegen ligger inntil vegen er det viktig at dataene er nøyaktige nok til at man greier å se hvor syklisten faktisk sykler. Det krever også at man har et detaljert nettverk som inneholder alle stier og snarveier som syklistene kan benytte seg av¹⁸. Likevel har flere av studiene som ble presentert i litteraturdelen gjort nettopp dette uten at de har hatt problemer med resultatene.

Bruken av GPS-logging under datainnsamlingen har flere fordeler. Det er en relativt enkel og billig metode for å samle inn data, sammenliknet med andre metoder. For det første har mange moderne telefoner har både GPS-mottakere og trianguleringsalgoritmer fra basestasjoner innebygd, som gjør at store deler av befolkningen allerede har det tekniske utstyret på plass¹⁹. Ved å utvikle eller benytte seg av allerede utviklede applikasjoner til smarttelefoner har man en meget billig og effektiv plattform for innsamling av data. Dersom

¹⁸ I Norge er det ikke sikkert at dette kommer til å bli et problem, ettersom Statens Vegvesen har brukt mye ressurser på å oppdatere den Nasjonale Vegdatabanken med nettverket for gang- og sykkelveg (Jetlund, 2013).

¹⁹ (Medienorge (2014)) melder at om lag 80 % av nordmenn over 15 år hadde en smarttelefon i 2013. Smarttelefon var definert som en telefon som hadde mulighet til å koble til internett, og innebærer derfor ikke at den har mulighet for GPS-sporing. Likevel er det rimelig å anta at en stor andel av Norges befolkning har denne funksjonen på sin telefon.

man velger dedikerte GPS-mottakere blir disse billigere og bedre etter hvert som utviklingen beveger seg fremover. Med tanke på innføringen av GLONASS og GALILEO vil man ikke bare få mottakere for GPS, men også disse nye systemene, som vil øke nøyaktigheten. Man får treningsklokker med innebygd GPS ned mot 1500 kroner, mens håndholdte mottakere finnes i enda billigere varianter.

Bruk av GPS til datainnsamlingen betyr i tillegg at man får revealed preference data, som man kan supplere med stated preference data dersom det er ønskelig. Man kan for eksempel utføre datainnsamling med GPS-mottakere men i tillegg gjennomføre spørreundersøkelser. Det øker sannsynligheten for at man får nøyaktige data med gode muligheter for oppfølging under arbeidet, og man blir sittende på et godt datasett.

Det er også en meget effektiv metode for å samle inn data, ettersom de aller fleste mottakere allerede har programvare som støtter eksport av data fra enheten. Det er også fullt mulig å benytte seg av programmeringskode, slik som i denne oppgaven, for å formatere dataene til et akseptabelt format som er kompatibelt med populære databehandlingsprogram som Microsoft Excel og IBM SPSS. Dersom man i tillegg får på plass gode algoritmer for filtreringsprosessen og oppkoblingen mot vegnettet kan man i teorien automatisere veldig mye av arbeidet.

Om man bruker en dedikert GPS-mottaker eller smarttelefoner har begge sine svakheter og styrker. Nøyaktigheten ser ut til å være omtrent den samme, men en GPS-mottaker vil i de aller fleste tilfeller ha lengre batteritid og mye av strukturen med å eksportere og dele (samle inn) turene er allerede på plass. Ser man derimot på andelen som har smarttelefoner i Norge er det et meget stort potensial. Ideelt sett kan datainnsamling ved hjelp av en applikasjon på en smarttelefon gi utrolig mye data for veldig liten innsats. Dersom brukeren i tillegg til å logge turen kan tilføye informasjon om alder, kjønn, og så videre, samt turens hensikt rett inn i applikasjonen, har man nesten ett komplett sett med data for en enkelt tur.

5.3.1 Andre potensielle bruksområder for bruk av offentlig tilgjengelig GPS-data

På grunn av den gode tilgjengeligheten og store datamengden som er tilgjengelig kan det være andre bruksområder for GPS enn fart- og rutevalgmodellering. Ved å samle en mengde GPS-data i et GIS får man fort en oversikt over hvor folk sykler. Man kan se hvilke ruter som benyttes, hvilke områder som flest folk sykler til eller sykler fra, hvilke områder ingen sykler,

og så videre. Dersom mange syklistere benytter seg av en strekning og i tillegg unnviker langt fra den korteste ruten for å benytte seg av nettopp denne strekningen kan det være et godt tegn på at denne er en attraktiv strekning. På samme måte kan man se på områder som syklistere unngår. Det gir GPS-data gode muligheter til å benyttes som indikatorer på «hot-spots», områder som sårt trenger oppgraderinger eller strekninger som er populære blant syklistene. Dette kan også være et godt alternativ til å foreta en evaluering etter at et tiltak er utført. Man kan se på GPS-dataene før og etter et tiltak er gjennomført og se om det er endringer i bruksmønsteret.

På samme måte kan dette brukes inn mot arealplanlegging. Ser man at mange turer starter eller ender i samme område kan det være tegn på det er godt tilrettelagt for sykling i eller inn mot disse områdene. Da kan man gå videre i analysen og prøve å finne ut hvorfor det er slikt og hvordan man kan benytte seg av dette i andre områder.

Man kan også benytte seg av hastighetsdataene for å se etter områder som fører til økt reisetid. Det kan være spesielle kryss som syklistene må vente lenge i, eller det kan være områder hvor det er mange konflikter med andre trafikanter. Er det et spesielt område hvor flere av de gjennomførte sykkelturene har lave hastigheter kan dette være et tegn på at det er dårlig tilrettelagt i området.

5.4 Oppsummering

De tilgjengelige datasettene inneholdt mange relevante turer, men flere nettsted mangler nødvendige variabler for bruk til en fartsmodell. For å lage en fartsmodell der stigning er en beskrivende variabel trenger man posisjonen i tre plan samt en tidsverdi eller en registrert fart for hvert målepunkt, og det var kun én side som hadde dette.

Det samlede datasettet inneholder bias med tanke på at det er trolig et flertall av aktive og interesserte syklister som har logget og delt turene sine. Det er også bias ettersom samme person kan ha syklet samme rute flere ganger, uten at det er tatt spesielle hensyn til det. Selv om antallet unike brukere i datasettet kan være stort, kan det være et fåtall som står for flesteparten av turene. Turene er i stor grad hentet fra Oslo som kan føre til ytterligere bias, da det er mistanke om at forskjellige byer kan ha ulik «sykkelkultur».

Det er også store usikkerheter rundt nøyaktighetene til dataene, spesielt for høydedata. På grunn av avrundinger gjort av mottakeren er det overdrevet mange segmenter hvor stigningen er registrert til 0,0 %, som ikke er den virkelige stigningen. Dette jevner seg delvis ut når man regner på gjennomsnittlig stigning for et segment, men kan fortsatt være plaget av feil. Det er også naturlig å anta at det kan være feil i de registrerte hastighetene. Det kan også være at noen mottakere har bedre nøyaktighet enn andre, uten at det er gjort noen tiltak for å identifisere dette.

GPS-data har stort potensial til å benyttes i modellering av rutevalg. Mulighetene for enkel og billig innsamling av data, hvor store deler av grovarbeidet kan gjøres automatisk ved hjelp av programmeringskode, gjør det til et effektivt verktøy. I tillegg ligger det store deler tilgjengelig data på internett som allerede er samlet inn og som kan være brukbart. Mange av de antatt viktigste faktorene for rutevalg lar seg enkelt analyseres ved hjelp av GPS og GIS, men det kan være personlige egenskaper som også vil være viktige. Slike data er ikke alltid frigjort i de offentlig tilgjengelige kildene.

Det finnes også andre scenarioer hvor bruken av offentlig tilgjengelig GPS-data kan være nyttig datakilde for både transport- og arealplanleggere. Identifikasjonen av områder med lave hastigheter, få syklister eller lenker som mange syklister unngår kan være tegn på at det er dårlig tilrettelagt for syklister. På den andre siden kan man se etter strekninger med høye hastigheter, mange syklister eller områder som det er populært å sykle til for å identifisere gode løsninger både på teknisk og planleggingsnivå.

6 Konklusjon og videre arbeid

Oppgavens siste kapittel er konklusjon og forslag til videre arbeid. Konklusjonen vil være en avslutning på de gjennomgående temaene som har vært omtalt i oppgaven, mens videre arbeid vil være anbefalinger på hvordan man kan fortsette det arbeidet som er gjort til nå. Dette kapittelet er ikke delt inn i mindre deler, men er heller flettet sammen og temaene flyter over i hverandre.

Reisetid viser seg å være en viktig faktor for syklistenes rutevalg, men mange av dagens ruteplanleggere har ikke god dokumentasjon på hvorfor de gjør de antagelsene for sykkelfart som de gjør og hvordan det gjøres. For den regionale transportmodellen er de generaliserte kostnadene for syklist kun avhengig av avstand, og fart blir først relevant i trafikantnytteberegningene. Det er tydelig at det er store forskjeller mellom reisetidene avhengig om man benytter de utviklede modellene eller en konstant fart på 15 km/t, selv om de modellerte hastighetene antas å være noe høye. Det er også tydelige forskjeller mellom hvordan rutevalget er modellert i dagens regionale transportmodell sammenliknet med hva studiene som er presentert i denne oppgaven har funnet. Mange viktige forklaringsfaktorer for rutevalg er ikke tatt hensyn til, og antagelsen om at syklist benytter den korteste ruten er ikke en god løsning dersom man ønsker realistiske resultat.

Det betyr at det er et hull i RTM hvor fart, reisetid og rutevalg ikke påvirker de generaliserte kostnadene eller reisemiddelvalget på en god måte. Med tanke på hvor viktig reisetiden er for rutevalget, samt hvor viktig de generaliserte kostnadene og nyttefunksjonen er i 4-steg metodikken, er man helt klart avhengig av en oppdatert metode for å modellere sykkeltrafikk i den regionale transportmodellen. Det vil gi bedre tilnærminger til virkeligheten og derfor gi bedre resultat i turfordelingen og reisemiddel-fordelingen. Derfor bør den regionale transportmodellen ikke bare få en oppdatert metode for å beregne sykkelfart, men også en oppdatert modell for rutevalget.

De tre modellene hadde lav forklaringsgrad og noen problemer med de statistiske testene, som tyder på at datasettet ikke har nok variabler til å skape en fullgod modell for sykkelfarten. Noen av de manglende variablene kan identifiseres ved å benytte seg av GIS og map matching for å koble datasettet opp mot et veg- og sykkelstinettnettverk. På den måten kan kryss og annen infrastruktur inkluderes i modellen. Andre variabler som personlig informasjon om syklisten er ikke tilgjengelig i datasettet eller på nettsiden, og dette må samles inn på en annen måte. I tillegg lider filtreringsprosessen på grunn av usikkerhet rundt representative verdier,

spesielt for hastigheter og stigninger. Samtidig er det usikkerheter og unøyaktigheter i rådataene som gjør dette enda vanskeligere. Grensene som ble brukt i denne oppgaven var nok for liberale, og burde vært strengere. En øvre grense på ca. 50-60 km/t for hastigheter og $\pm 20,0$ % stigning er trolig et bedre alternativ. De modellerte hastighetene virker også noe høye enn hva som er representativt, spesielt da de ble implementert i nettverket til DOM Nidaros. Det skyldes trolig biasen i datasettet, hvor det antas at det er flertall av erfarne og ivrige syklister.

For å utvikle en bedre modell for sykkelfart bør det derfor gjennomføres en kontrollert studie. Studien bør for det første forbedre filtreringsprosessen, blant annet ved å identifisere representative hastigheter og stigninger for «gjennomsnittssyklisten». For det andre bør studien også samle inn personlige egenskaper om deltagerne. På den måten kan noen av grupperingene som ble nevnt i tidligere kapittel forklares og tas med i modellen. Til sist bør studien koble GPS-dataene opp mot et nettverk for å identifisere ruteegenskaper som ikke er beskrevet i datasettet. Til sammen vil disse forbedringene føre til et mer representativt datasett som inneholder flere beskrivende variabler, som ikke bare vil gi økt forklaringsgrad for modellene, men også mulighet for å få bukt på problemene med autokollinearitet og dårlig homoskedastisitet.

Mange av faktorene for rutevalg som ble funnet i litteraturstudie lar seg beskrive ved hjelp av GPS og GIS. Det betyr at man i utviklingen av en rutevalgmodell kan benytte seg av de samme dataene som ble brukt i denne studien. Men, ettersom man mangler blant annet personlige data som kan være relevant, anbefales det heller at utviklingen av modellen gjennomføres med data fra den kontrollerte studien. Da er man sikrere på datakvaliteten, samtidig som arbeidet til en viss grad er likt (blant annet oppkoblingen mot nettverket i et GIS). For å få større tilgang på personlig data på deltagerne bør den kontrollerte studien også inneholde en spørreundersøkelse som kan supplere datasettet. Informasjon om hvorfor syklistene velger den ruten de gjør kan gi store bidrag til den endelige rutevalgmodellen, og lar seg ikke nødvendigvis beskrives av GPS-data alene. Det er litt usikkerhet rundt nøyaktigheten til GPS, men andre studier tyder på at dette ikke er et problem. Mulighetene for billig utstyr og mange potensielle brukere, samt automatisering av databehandling og filtrering gir GPS-data mange fordeler over konvensjonelle innsamlingsmetoder.

Skulle det være nødvendig med supplerende data for rutevalgmodellen er offentlig tilgjengelig GPS-data en meget relevant datakilde. For rutevalget er det ikke nødvendig med

hastighets- eller tidsdata, og man kan i tillegg benytte seg av Kartverkets applikasjon for å tilføye høydedata. Det gjør at man strengt tatt trenger et rådatasett som kun inneholder lengde- og breddegrader, som igjen fører til at mengden tilgjengelig GPS-data på nettet øker.

Offentlig tilgjengelig data har også et potensial som datakilde for andre undersøkelser. Identifisering av flaskehalser, populære eller upopulære strekninger, områder med lav eller høy sykkelbruk og før/etter-analyser er noen eksempler. I disse tilfellene er ikke personlige data like viktig ettersom man ønsker å se populasjonen under ett. Dersom den kontrollerte studien klarer å avdekke representative verdier for filtrerings-prosessen, kan man utvikle gode algoritmer på å automatisere mye av arbeidet. Dersom man kan hente, parse, filtrere og «map matche» mye av dataen automatisk sitter man plutselig på et verktøy med utrolige kapasiteter. I løpet av kort tid vil man ha muligheten til å få store mengder informasjon om en bys sykkeltrafikk rett inn i et GIS. Det gir et meget godt grunnlag for flere typer analyser, der dataene samles inn av seg selv. I tillegg er det tilnærmet kostnadsfritt.

Referanseliste

- AULTMAN-HALL, L., HALL, F. & BAETZ, B. 1997. Analysis of Bicycle Commuter Routes Using Geographic Information Systems: Implications for Bicycle Planning. *Transportation Research Record: Journal of the Transportation Research Board*, 1578, 102-110.
- BROACH, J., DILL, J. & GLIEBE, J. 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46, 1730-1740.
- BRUKERNETTVERKET FOR ATP-MODELLEN 2005. ATP-modellen. www.atpmodell.no.
- CHATTERJEE, S. & SIMONOFF, J. S. 2013. *Wiley Handbooks in Applied Statistics : Handbook of Regression Analysis*, Somerset, NJ, USA, Wiley.
- CHEN, X., ENDER, P., MITCHELL, M. & WELLS, C. 2003. *Regression with SPSS*, Institute for digital research and education, UCLA.
- CLUE DICTIONARY SOFTWARE 2011. *Clue Dictionary Software*.
- DE DIOS ORTÚZAR, J. & WILLUMSEN, L. G. 2011. *Modelling Transport*, Wiley.
- DILL, J. & GLIEBE, J. P. 2008. Understanding and measuring bicycling behavior: A focus on travel time and route choice.
- FORSSELL, B. 2009a. *GNSS - Det Store Norske Leksikon* [Online]. Tilgjengelig: <http://snl.no/GNSS> [Hentet 17.02 2014].
- FORSSELL, B. 2009b. *GPS - Store Norske Leksikon* [Online]. Tilgjengelig: <http://snl.no/GPS> [Hentet 17.02 2014].
- FRIEDL, J. 2012. *Jeffrey Friedl's Blog >> Informal GPS Logger Test: iPhone 4s GPS is Shockingly Good* [Online]. Tilgjengelig: <http://regex.info/blog/2012-01-15/1922> [Hentet 31.05 2014].
- GARMIN LTD. 2012. Brukerveiledning for Forerunner 310XT. Available: http://static.garmincdn.com/pumac/Forerunner310XT_OM_NO.pdf.
- GARMIN LTD. 2014. *Into Sports | Garmin* [Online]. Tilgjengelig: <http://www.garmin.com/en-US/explore/intosports/> [Hentet 18.02 2014].
- GOOGLE. 2010. *How are walking and bicycling times calculated? - Google produktforumer* [Online]. Tilgjengelig: <https://productforums.google.com/forum/#!topic/maps/PHwS75mwUiY> [Hentet 19.02 2014].
- HARVEY, F. J. & KRIZEK, K. J. 2007. Commuter Bicyclist Behavior and Facility Disruption.
- HENSHER, D. A. & BUTTON, K. J. 2008. *Handbook of Transport Modelling (2nd Edition)*. Emerald, Inc.
- HOOD, J., SALL, E. & CHARLTON, B. 2011. A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters*, 3, 63-75.
- JETLUND, K. 2013. *Ny versjon av sykkelveg.no - Ruteplanlegger for syklist* [Online]. Geomatikkbloggen.blogspot.no. Tilgjengelig: <http://geomatikkbloggen.blogspot.no/2013/04/ny-versjon-av-sykkelvegno.html> [Hentet 19.02 2014].
- JETLUND, K. 2014. Personlig kommunikasjon. *RE: Spørsmål ang. sykkelveg.no*. Skrevet til JØRGENSEN, C. V.
- LANDIS, B., VATTIKUTI, V. & BRANNICK, M. 1997. Real-Time Human Perceptions: Toward a Bicycle Level of Service. *Transportation Research Record: Journal of the Transportation Research Board*, 1578, 119-126.

- LARVIK KOMMUNE. 2013. *Sykkel - Larvik Kommune* [Online]. Tilgjengelig: <http://www.larvik.kommune.no/no/Los-struktur/Emneord-a-a/Sykkel/> [Hentet 19.02 2014].
- LUND, A. & LUND, M. 2013. *Multiple regression in SPSS* [Online]. <https://statistics.laerd.com/>: Lund Research Ltd. Tilgjengelig: <https://statistics.laerd.com/premium/mr/multiple-regression-in-spss.php> [Hentet 17.05 2014].
- LÖFÅS, P. 23.02 2014. Personlig kommunikasjon. RE: *Spørsmål angående sykkelveg.no - Statens vegvesens ruteplanlegger for sykkel*. Skrevet til JØRGENSEN, C. V.
- MADSLIEN, A., REKDAL, J. & LARSEN, O. I. 2005. Utvikling av regionale modeller for persontransport i Norge. *TØI rapport*.
- MALMIN, O. K. 2013. CUBE - Teknisk dokumentasjon av Regional transportmodell. A24718. SINTEF Teknologi og samfunn - Transportforskning, Trondheim.
- MEDIENORGE. 2014. *Andel som har smarttelefon* [Online]. Tilgjengelig: <http://www.medienorge.uib.no/statistikk/medium/ikt/379> [Hentet 28.05 2014].
- MENGHINI, G., CARRASCO, N., SCHÜSSLER, N. & AXHAUSEN, K. W. 2010. Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice*, 44, 754-765.
- MEYER, J. L. 2013. Big Bike Data: How GPS Route Data Collected from Smartphones Can Benefit Bicycle Planning.
- MILJØPAKKEN. 2014. *Sykkelbyen Trondheim* [Online]. Tilgjengelig: <http://sykkelbyentrondheim.no/kart/> [Hentet 19.02 2014].
- NORDDAL, K. S. & ØRNES, E. 2012. Kurs i konstruksjon av nettverk i ATP 10.1. 2 ed.
- NUTTALL, S. & LUCAS-SMITH, M. 2010. *CycleStreets >> About CycleStreets* [Online]. Tilgjengelig: <http://www.cyclestreets.net/about/> [Hentet 12.02 2014].
- OGAJA, C. A. 2011. Applied GPS for Engineers and Project Managers. American Society of Civil Engineers (ASCE).
- PARKIN, J. & ROTHERAM, J. 2010. Design speeds and acceleration characteristics of bicycle traffic for use in planning, design and appraisal. *Transport Policy*, 17, 335-341.
- REKDAL, J., LARSEN, O. I., LØKKETANGEN, A. & HAMRE, T. N. 2013. TRAMOD_BY DEL 1: Etablering av nytt modellsystem. *Rapport 1313*. Møreforskning Molde AS, Molde.
- RICH, J. 2009. *Introduction to Transport Models*, Lulu.com.
- SAMFERDSELSDEPARTEMENTET 2013. Meld. St. 26 (2012-2013) Nasjonal transportplan 2014-2023. Samferdselsdepartementet, Oslo.
- SANDNES KOMMUNE. 2010. *Sykkelbyen Sandnes - Sunnby* [Online]. Tilgjengelig: <http://www.sunnby.no/aktiviteter/sunn-livsstil/sykkelbyen-sandnes/> [Hentet 19.02 2014].
- SENER, I. N., ELURU, N. & BHAT, C. R. 2009. An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36, 511-539.
- SHANER, J. 2013. *Smartphones, Tablets and GPS Accuracy | ArcGIS Blog* [Online]. Tilgjengelig: <http://blogs.esri.com/esri/arcgis/2013/07/15/smartphones-tablets-and-gps-accuracy/> [Hentet 31.05 2014].
- STATENS FORURENINGSTILSYN 2001. Samordning av arealbruk og transport - verktøy for planlegging. Utprøving av ATP-modellen i fire fylker - erfaring og eksempler. *TA-1834/2001*. Miljødirektoratet, Oslo.
- STATENS KARTVERK. 2013. *Norgeskart* [Online]. Kartverket. [Hentet 8.5.2014 2014].
- STATENS VEGVESEN 1991. Kapasitet på vegstrekninger. *Håndbok 159*. Vegdirektoratet, Oslo.

- STATENS VEGVESEN 2013. Sykkelhåndboka. *Håndbok 233*. Vegdirektoratet, Oslo.
- STATENS VEGVESEN 2014a. Konsekvensanalyser. *Håndbok V712*. Vegdirektoratet, Oslo.
- STATENS VEGVESEN. 2014b. *Vegkart* [Online]. Tilgjengelig: <https://www.vegvesen.no/vegkart/vegkart/> [Hentet 24.02 2014].
- STAVANGER KOMMUNE. 2013. *Sykkelkart - Stavanger kommune* [Online]. Tilgjengelig: <http://www.stavanger.kommune.no/Tilbud-tjenester-og-skjema/Vei-og-trafikk/stavanger-pa-sykkel/Kart-til-sykkelturen/> [Hentet 19.02 2014].
- STINSON, M. A. & BHAT, C. R. 2003. Commuter bicyclist route choice: analysis using a stated preference survey. *Transportation research record: journal of the transportation research board*, 1828, 107-115.
- STOPHER, P., FITZGERALD, C. & ZHANG, J. 2008. Search for a global positioning system device to measure person travel. *Transportation Research Part C: Emerging Technologies*, 16, 350-369.
- STOPHER, P. R. & SPEISSER, N. 2011. Evaluation of GPS Device Properties for a Possible Use in Future Household Travel Surveys.
- SU, J. G., WINTERS, M., NUNES, M. & BRAUER, M. 2010. Designing a route planner to facilitate and promote cycling in Metro Vancouver, Canada. *Transportation Research Part A: Policy and Practice*, 44, 495-505.
- TRANSPORT DIRECT. 2010. *Cycle Planning Help | Transport Direct* [Online]. Tilgjengelig: <http://www.transportdirect.info/Web2/staticnoprnt.aspx?id= web2 help helpcycle> [Hentet 13.02 2014].
- TRANSPORT FOR LONDON. Årstell ikke oppgitt. *English - Cycle Planner - Transport for London* [Online]. tfl.gov.uk: Transport for London. Tilgjengelig: http://cyclejourneyplanner.tfl.gov.uk/cycle/XSLT_TRIP_REQUEST2?language=en [Hentet 12.02 2014].

Liste over vedlegg

Vedlegg A Oppgavetekst

Vedlegg B Statistiske tester for flate strekninger

Vedlegg C Statistiske tester for de første og de siste

Vedlegg D Utvidede resultat fra modellarbeidet

Vedlegg A – Oppgavetekst

BAKGRUNN

Veksten i persontransporten framover skal tas med kollektivtransport, sykkel og gange. Det er ordlyden i Nasjonal Transportplan. St. 26, kapittel 1.3.1. Det er satt av betydelige investeringsmidler for å sikre måloppnåelse. De regionale transportmodellene er pr i dag ikke egnet til å beregne endringer i etterspørselen etter sykkeltransport, og dermed mangler det et analyseverktøy for evaluering av tiltak. Reisetid og høydeforskjell er to forklaringsfaktorer som sannsynligvis betyr mye for om folk velger sykkel og hvilken rute de velger. Det er derfor behov for en blant fartsmodell for sykkel, hvor stigning og fall inngår i beregningene.

OPPGAVE

Kan GPS sporing av sykkelturer brukes som grunnlag for en fartsmodell og en rutevalgmodell for syklende?

Det er to tilgjengelige datakilder som skal kobles sammen og brukes i prosjektet. Den ene datakilden er sykkelturer som er "delt" på internett, med informasjon om rute og hastighetsprofil. Den andre datakilden er beskrivelse av veglenkene fra Nasjonal Vegdatabank NVDB, med blant annet høydeprofil.

Kandidaten skal i dette arbeidet se på mulighetene til å benytte offentlig tilgjengelig data som grunnlag for en ny metode for å beregne fart for sykkelturer. Kandidaten skal:

- Gjennomføre et todelt litteraturstudium: Første del skal identifisere metoder og modeller brukt for å beskrive fartsprofil på sykkelturer i ruteplanleggingsverktøy og evt. transportmodeller. Andre del skal vise hvilke forklaringsfaktorer som er viktige for beregning av rutevalg med sykkel.
- Finne offentlige tilgjengelige datakilder som kan være nyttige i etableringen av modellen, herunder om gjennomførte turer fra f.eks Strava, Garmin, RunKeeper. Data om vegens utforming skaffes via et TNext-uttak fra NVDB.
- Bruke de offentlige datakildene til å beskrive en ny metode for å beregne sykkelfart
- Vise effekten av å bruke sykkelfart i stedet for avstand som brukes i nettfordelingen av sykkeltransporten i RTM.
- Diskutere om turene fra GPS sporinger også kan brukes til å lage en rutevalgmodell for syklende

Vedlegg B – t-test av flate strekninger

Statistisk test av lengder til segmenter hvor stigningen er 0,0 %. Dataene er hentet fra vegvesenets kartapplikasjon *Vegkart*. Resultatet tyder på at de fleste registrerte vegstrekningene der stigningen er 0,0 % er kortere enn 100 meter.

Tabell A.1 - Statistiske tester for stigninger med stigning 0,0 %.

	Kristiansand	Trondheim	Oslo
Testverdi	100	100	100
Signifikansnivå	0,05	0,05	0,05
Gjennomsnitt	44,920	65,713	57,302
Standardavvik	60,291	114,428	54,752
Standardfeil	12,058	6,188	6,898
Frihetsgrader	24	341	62
t-verdi	-4,568	-5,541	-6,190
Beregnet signifikans	0,000	0,000	0,000
Kritisk t-verdi	-1,711	-1,649	-1,670
Signifikant forskjell	Ja	Ja	Ja

Vedlegg C – F- og t-test av første og siste 100 meter

Tester for å se om det er en statistisk signifikant forskjell mellom de første og de siste sammenliknet med resten av turen. Testene viser at det er statistisk forskjell i hastighetenes varians og gjennomsnittsverdi.

Tabell B.1 – F-test for varians for første 100 meterne mot resten av turen

	Første 100 meterne	100 meter og utover
Gjennomsnitt	13,284	22,506
Varians	51,078	74,954
Observasjoner	43 345	1 048 573
Frihetsgrad	43 344	1 048 572
F-verdi	0,681	
P(F<=f) En-halet	0,00	
Kritisk F-verdi	0,984	

Tabell B.2 – t-test for forskjell i fart mellom første 100 meterne og resten av turen.

	Første 100 meterne	100 meter og utover
Gjennomsnitt	13,284	22,506
Varians	51,078	74,954
Observasjoner	43 345	1 048 573
Hypotetisk forskjell i gjennomsnitt	0,000	
Frihetsgrader	48 755	
t-verdi	-260,860	
P(T<=t) En-halet	0,000	
Kritisk t-verdi, en-halet	2,326	
P(T<=t) To-halet	0,000	
Kritisk t-verdi, to-halet	2,576	

Tabell B.3 F-test for varians. De siste mot første del av turen

	De siste	Opp til de siste 100 meterne
Gjennomsnitt	12,215	22,487
Varians	55,979	75,625
Observasjoner	58 745	1 048 573
Frihetsgrad	58 744	1 048 572
F-verdi	0,740	
P(F<=f) En-halet	0,000	
Kritisk F-verdi	0,986	

	De siste	Opp til de siste 100 meterne
Gjennomsnitt	12,215	22,487
Varians	55,979	75,625
Observasjoner	58 745	1 048 573
Hypotetisk forskjell i gjennomsnitt	0,000	
Frihetsgrader	67 951	
t-verdi	-320,819	
P(T<=t) En-halet	0,000	
Kritisk t-verdi, en-halet	2,326	
P(T<=t) To-halet	0,000	
Kritisk t-verdi, to-halet	2,576	

Vedlegg D – Resultat fra modellarbeidet

Tabell C.1 - Statistiske verdier for nedoverbakkemodellen

Variabel	Koeffisienter		t-verdi	Signifikans	Konfidensintervall for B			Korrelasjoner			Kollinearitetstester			
	B	Standard feil			Nedre grense	Øvre grense	Zero-order	Partial	Part	Toleranse	VIF			
												95,0%		
Konstant	66,729	,759	87,944	0,000	65,242	68,216								
Gjennomsnittlig stigning [%]	,863	,005	167,141	0,000	,853	,873	,219	,301	,259	,895	1,117			
log10 lengde	6,854	,023	300,414	0,000	6,809	6,899	,456	,493	,466	,931	1,075			
Temperatur [grader C]	,157	,002	79,277	0,000	,153	,161	,148	,148	,123	,997	1,004			
log10 gjennomsnittlig stigning, forrige segment	-31,100	,485	-64,174	0,000	-32,050	-30,150	-,239	-,120	-,100	,904	1,106			

Tabell C.2 - Statistiske verdier for variablene i nedoberbakkemodellen

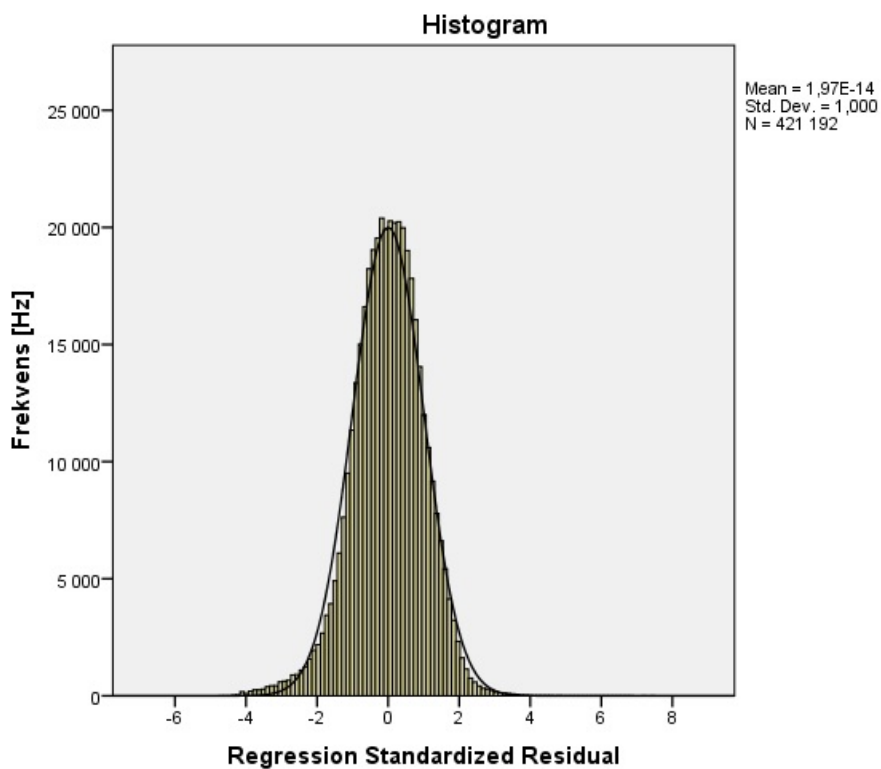
Statistikk for variablene			
Variabel	Gjennomsnitt	Standardavvik	N
Fart [km/t]	26,747	8,786	280 722
Gjennomsnittlig stigning [%]	-5,083	2,791	280 722
log10 lengde	1,702	0,620	280 722
Temperatur [grader C]	8,595	6,897	280 722
log10 gjennomsnittlig stigning, forrige segment	1,563	0,030	280 722

Tabell C.3 - Statistikk for flatmarksmodellen

Variabel	Koeffisienter		t-verdi	Signifikans	Konfidensintervall for B			Korrelasjoner			Kollinearitetstester		
	B	Standard feil			Nedre grense	Øvre grense	Zero- order	Partial	Part	Toleranse	VIF		
												95,0%	
(Constant)	12,602	,035	355,962	0,000	12,533	12,671							
Gjennomsnittlig stigning [%]	-2,401	,022	-110,832	0,000	-2,443	-2,359	-2,35	-1,68	-1,46	,937	1,067		
log10 Stigningslengde	2,442	,019	128,883	0,000	2,405	2,479	,322	,195	,170	,829	1,206		
Gjennomsnittlig stigning, forrige segment [%]	-,214	,002	-107,153	0,000	-,218	-,210	-,260	-,163	-,141	,889	1,125		
log10 lengde, forrige segment	2,222	,024	91,891	0,000	2,175	2,270	,292	,140	,121	,723	1,383		
Gjennomsnittlig stigning, neste segment [%]	-,183	,002	-88,422	0,000	-,187	-,179	-,235	-,135	-,116	,893	1,120		
log10 lengde, neste segment	1,468	,025	59,807	0,000	1,420	1,516	,267	,092	,079	,719	1,390		
Temperatur [grader C]	,195	,001	133,061	0,000	,192	,198	,222	,201	,175	,986	1,014		

Tabell C.4 - Oversikt over variablene for flatmarksmodellen

Statistikk for variablene			
Variabel	Gjennomsnitt	Standardavvik	N
Fart [km/t]	22,039	7,804	421 192
Gjennomsnittlig stigning [%]	-0,051	0,490	421 192
log10 Stigningslengde	1,498	0,596	421 192
Gjennomsnittlig stigning, forrige segment [%]	0,487	5,457	421 192
log10 lengde, forrige segment	1,155	0,500	421 192
Gjennomsnittlig stigning, neste segment [%]	0,565	5,249	421 192
log10 lengde, neste segment	1,161	0,494	421 192
Temperatur [grader C]	8,157	7,065	421 192



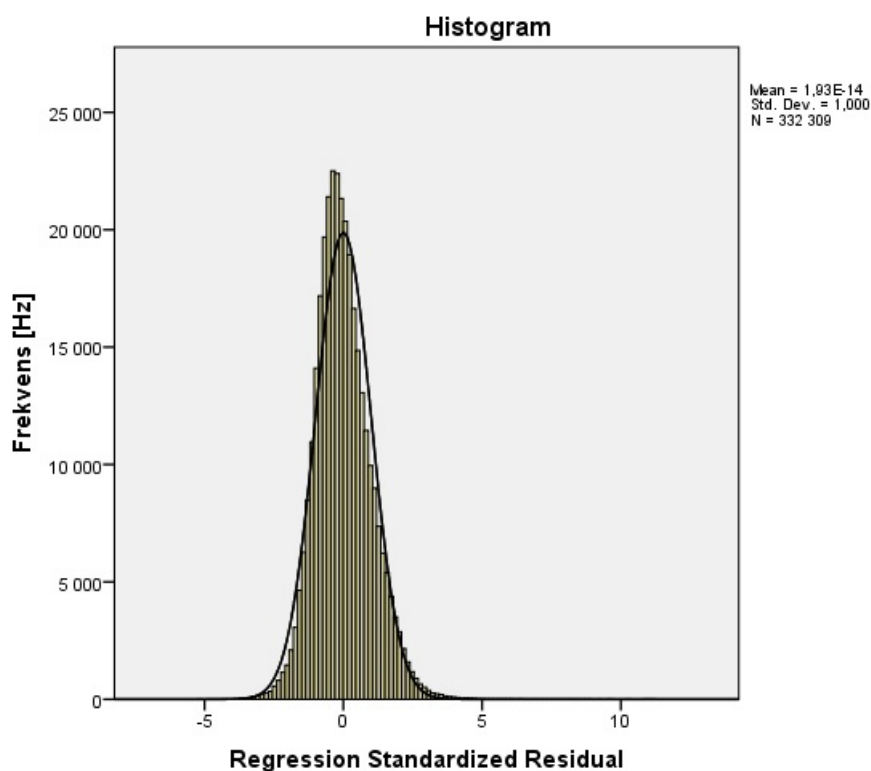
Figur C.1 – Fordeling av residualer for flatmarksmodellen

Tabell C.5 - Statistikk fra oppoverbakkemodellen

Variabel	Koeffisienter			95,0% Konfidensintervall for B				Korrelasjoner			Kollinearitetstester	
	B	Standard feil	t-verdi	Signifikans	Nedre grense	Øvre grense	Zero- order	Partial	Part	Tolerance	VIF	
Konstant	20,246	,037	546,833	0,000	20,174	20,319						
Gjennomsnittlig stigning [%]	-,993	,003	-339,495	0,000	-,999	-,987	-,538	-,507	-,456	,848	1,180	
log10 Stigningslengde	-1,076	,018	-59,461	0,000	-1,112	-1,041	-,067	-,103	-,080	,834	1,200	
Gjennomsnittlig stigning, forrige segment [%]	-,270	,004	-67,571	0,000	-,278	-,263	-,074	-,116	-,091	,916	1,092	
log10 lengde, forrige segment	2,489	,022	111,310	0,000	2,445	2,533	,319	,190	,150	,661	1,514	
Gjennomsnittlig stigning, neste segment [%]	-,251	,004	-62,243	0,000	-,259	-,243	-,054	-,107	-,084	,903	1,107	
log10 lengde, neste segment	,869	,022	39,705	0,000	,826	,912	,274	,069	,053	,667	1,499	
Temperatur [grader C]	,178	,001	131,442	0,000	,176	,181	,221	,222	,177	,985	1,015	

Tabell C.6 - Statistikk for variablene til oppoverbakkemodellen

Statistikk for variablene			
Variabel	Gjennomsnitt	Standardavvik	N
Fart [km/t]	17,901	6,847	332 309
Gjennomsnittlig stigning [%]	6,132	3,416	332 309
log10 Stigningslengde	1,531	0,557	332 309
Gjennomsnittlig stigning, forrige segment [%]	-0,640	2,402	332 309
log10 lengde, forrige segment	1,074	0,506	332 309
Gjennomsnittlig stigning, neste segment [%]	-0,639	2,404	332 309
log10 lengde, neste segment	1,049	0,514	332 309
Temperatur [grader C]	8,261	6,829	332 309



Figur C.2 – Histogram for residualene for oppoverbakkemodellen