

Modellering av tågnät på Arriva DK



LUNDS
UNIVERSITET
Lunds Tekniska Högskola

Författare

Carl Rydholm och Per Vingå

Handledare

Johan Marklund, LTH

Stefan Vidgren, Arriva

Förord

Detta examensarbete har utförts under våren 2014 som avslutning på våra civilingenjörsutbildningar inom Industriell ekonomi på Lunds Tekniska Högskola, LTH. Examensarbetet genomfördes på avdelningen för Produktionsekonomi och det bedrevs på Arriva Danmark i Köpenhamn.

Vår målsättning med examensarbetet var att kunna få tillämpa de kunskaper vi har fått under vår tid på LTH, speciellt från specialiseringen i logistik och produktionsekonomi. Vi hoppas att Arriva kommer använda och bygga vidare på våra modeller och att de blir en naturlig del i analysarbetet inom företaget.

Vi vill tacka vår handledare på Arriva, Stefan Vidgren, för många bra diskussioner och en inblick i hur teori måste anpassas till verkligheten. Vi vill också tacka vår handledare på Produktionsekonomi, Johan Marklund, som alltid tagit sig tid och gett oss nya synsätt på de problem vi har stött på under arbetets gång. För att lösa detaljfrågor med ExtendSim och VBA-programmering i Microsoft Excel har många personer tillfrågats och varit till hjälp. Några av dessa är; Jonas Enhörning på Duke Systems, David Krahl på ExtendSim och Nate Farm på Excel Forum.

Sammanfattning

- Titel:** Modellering av tågnät på Arriva DK
- Författare:** Per Vingå, Carl Rydholm
- Handledare:** Johan Marklund LTH, Stefan Vidgren Arriva
- Syfte:** Syftet med detta examensarbete är ta fram två simuleringsmodeller av Arrivas tågnätverk i Danmark för att underlätta planläggningen av tidtabeller, en deterministisk och en stokastisk. Den deterministiska modellen ska användas för att kontrollera att tidtabeller inte innehåller några konflikter, ett arbete som idag görs manuellt. Den stokastiska modellen ska ge en bättre avspegling av verkligheten och användas till att jämföra olika tidtabeller. Syftet är inte att utveckla en helt färdig produkt utan snarare att ta fram ett tillvägagångssätt för modelluppbyggnad och modeller att bygga vidare på.
- Avgränsningar:** Modellerna ämnar visa en normal vardag. Därför bortser modellen från större störningar som signalfel, olyckor med mera. Modellerna tar heller inte hänsyn till materialplan eller tjänsteplan, det vill säga hur tillgången på tåg och personal påverkar tidtabellen.
- Andra avgränsningar är att modellerna bara tar hänsyn till Arrivas och inte andra företags tåg, samt att modellen använder en förenklad beslutsprocess vid konflikt i systemet.

- Metod:** Vi har använt en modifierad variant av 8-stegsmodell för stora simuleringsprojekt Hillier och Lieberman (2010, pp. 959-963). Processen har genomförts i omgångar, en för varje modell.
- Slutsatser:** Verifiering och validering visar att den deterministiska modellen stämmer väl överens med verkligheten. Modellen anses därför redo att användas som hjälpmedel vid kontroll av tidtabeller.
- För den stokastiska modellen har det varit svårare att klargöra validiteten. Detta på grund av att simuleringsmodellen inte inkluderar material- och tjänsteplan och därför blir det svårt att jämföra utdata med datamaterialet från Arriva. Validitet kan visas för logiken i systemet samt transporttidernas och passagerarantalens fördelningar. Dock visar analysen att stopptidens fördelningar behöver utredas vidare. Det råder även osäkerhet om de beslutsregler som tillämpas för att prioritera tåg vid förseningar.
- Genom valideringsarbetet har det visat sig att datamaterialet från Arriva har bristande kvalitet.
- Rekommendationen till Arriva är att börja med att utreda sin datasinsamling och kontrollera kvalitén. Sedan bör modellen utökas för att inkludera tjänsteplan och materialplan, för att kunna valideras mot det nya datamaterialet.
- Nyckelord:** Modellering, tågnät, simulering

Abstract

Title: Modeling of a large train network

Authors: Per Vingå, Carl Rydholm

Supervisors: Johan Marklund LTH, Stefan Vidgren Arriva

Purpose: The purpose of this thesis is to develop two simulation models of Arrivas train network in Denmark to facilitate the process of timetabling. One of the models assumes deterministic times in the other model these times are stochastic. The purpose of the deterministic model is to evaluate if timetables are free from conflicts, a process that currently involves a lot of manual labor. The main goal of the stochastic model is to be able to compare different timetables, using a more realistic model of the train network that compensates for the inherent uncertainties in travel and stop times.

The goal is not to develop a finished product for immediate practical implementation. The aim has been to develop a platform and process for modelling and serve as first version for further improvement.

Delimitations: The models focus on normal operations. Therefore unexpected events like signal errors, accidents etc. are disregarded. Another delimitation is that the availability of trains and staff, and how it affects the timetable, are disregarded.

Additional delimitations are that the model only accounts for Arriva trains and not for their competitors' trains, and that the decision process for prioritizing trains when conflicts occur is simplified in the simulation models.

Method: The method used in this thesis is a modified version of the 8-step process for large simulation projects in Hillier and Lieberman (2010, pp. 959-963). The process has been implemented in two stages, one for each model.

Conclusions: Verification and validation show that the deterministic model is consistent with reality. The model is therefore considered ready to be used as an aid in the process of making new timetables.

The stochastic model has been more difficult to validate. This is because the material plan and service plan is not yet included in the simulation model, which makes it hard to compare the simulation results to the observed data for the real system. However, validity has been achieved for the logic of the system, and for the distributions of transport times and number of passengers. We can also conclude that the distributions for the time spent on stations needs to be further investigated and validated. There is also an uncertainty about the rules of prioritization of trains in case there are delays.

During the validation of the model, the dataset has shown lack of quality in large parts of the total set. Therefore we recommend Arriva to investigate their data collection. After the quality of the data is improved, we recommend Arriva to implement the service plan and material plan into the simulation model to be able make a fair comparison between observed data and simulation results.

Keywords: Modeling, train networks, simulation

Ordlista

Materialplan	En materialplan avser att se till att rätt resurser finns på plats vid rätt tidpunkt, i denna studie avser materialplan huvudsakligen tåg.
Tjänsteplan	En tjänsteplan syftar till att tillhandahålla en plan för att personal ska finnas på rätt plats i rätt tidpunkt.
Regularitet	Andelen tåg som ankommer i tid eller inom ett visst tidsintervall efter avsatt ankomsttid. Inom Arriva är regularitetsgränsen satt till 5 minuter.
Stationstid	Den totala tiden som ett tåg spenderar på stationen. Stationstiden kan delas in stopptid och väntetid.
Stopptid	I stopptiden ingår inbromsning, acceleration samt tiden för öppning och stängning av dörrar samt relaterade aktiviteter. Tiden för öppning och stängning av dörrar varierar beroende på hur många som går av eller på tåget. Stopptiden är en delmängd av stationstiden.
Väntetid	Väntetid är den tid som tåget måste stå på stationen för att invänta sin avgångstid samt vänta på att spåret framför blir ledigt. Väntetiden är en delmängd av stationstiden.
Teknisk station	En station utan kundaktivitet som endast används för att samla in data om tåglinjen.
VBA	Visual Basic, programmeringsspråk som används i Microsoft Office paketet. Inom detta projekt har VBA-kod speciellt använts för att hantera data i Excel.

Innehållsförteckning

1	Inledning	1
1.1	Bakgrund	1
1.2	Problembeskrivning	2
1.3	Utvärdering av tidtabeller och planering av tågtrafik	3
1.4	Simulering	5
1.5	Syfte	7
1.6	Avgränsningar	7
1.7	Användningsområde	9
1.8	Rapportens disposition	10
2	Metod	11
2.1	Metodval	11
2.1.1	Formulera problem och planera studie	13
2.1.2	Ta fram konceptuell modell	13
2.1.3	Validering av konceptuell modell	14
2.1.4	Konstruera simuleringsmodell och samla data	14
2.1.5	Verifiering av simuleringsmodell	15
2.1.6	Validering av simuleringsmodell	16
2.1.7	Presentera resultat	16
2.2	Källkritik	16
2.3	Använd mjukvara	17
2.3.1	Microsoft Excel	17
2.3.2	ExtendSim	17
2.3.3	Statfit	18
3	Modellbeskrivning	19
3.1	Beskrivning av Arrivas tågnät i Danmark	19
3.2	Beskrivning av data material	20
3.3	Framtagningen av simuleringsmodeller	23

3.3.1	Tåglinjernas struktur	23
3.3.2	Beskrivning av tågets färd	24
3.3.3	Konceptuell modell	26
3.3.4	Deterministisk simuleringsmodell	28
3.3.5	Stokastisk simuleringsmodell	29
3.3.6	Analys av indata	32
3.3.7	Modellförklaring	47
4	Verifiering och Validering av modell	55
4.1	Verifiering av simuleringsmodeller	55
4.1.1	Kontroll av moduler	55
4.1.2	Kontroll av tåglinjer	55
4.1.3	Avstämning mot tredje part	56
4.1.4	Kontroll av import	56
4.2	Validering av simuleringsmodeller	56
4.2.1	Deterministisk modell	56
4.2.2	Stokastisk modell	57
5	Slutsats och rekommendation	73
5.1	Slutsats	73
5.2	Rekommendation	75
6	Litteraturförteckning	76
6.1	Böcker	77
6.2	Artiklar	77
6.3	Internet	78
7	Appendix A - Detaljerad översikt av Arrivas tåglinjer	79
8	Appendix B - Stokastiska fördelningar	80
9	Appendix C - Parametrar för transporttid	85
10	Appendix D - Lista över tekniska stationer	88

1 Inledning

I kapitlet beskrivs bakgrund, syfte och avgränsningar till rapporten. Fokus är att sätta in läsaren i problemet och vilka delar av problemet som projektet ämnar lösa, samt vad som överlämnas till Arriva för vidare arbete. I kapitlet presenteras även rapportens disposition.

1.1 Bakgrund

Arriva är en koncern med fokus på kollektivtrafik i Europa. Koncernen har över 50 000 anställda i 15 olika länder (Arriva, 2014) med en omsättning på över 3 miljarder Euro (Arriva, 2011).

2003 startade Arriva sin verksamhet i Danmark som tågoperatör efter att det danska tågmonopolet upphört två år tidigare. Arriva tillhandahåller idag sju sträckor på Midt- och Väst-Jylland vilket uppskattas till 15 % av den danska tågtrafiken (Arriva, 2010a). De sju sträckorna delas upp i totalt 81 stationer varav 71 som Arriva driver själva, se detaljerad översikt Appendix A (Arriva, 2010b).

Arriva har i hela sin verksamhet stort fokus på punktlighet. Under 2013 kunde Arriva presentera att 97 % av alla tåg avgick i tid, vilket tågdirektör för Arriva



Figur 1 – Sex av Arrivas tågsträckor i Danmark (Arriva, 2010c)

Danmark, Ivan Skødt, nämner i ett uttalande på Arrivas hemsida:

”2013 har været et fantastisk godt år for Arriva. Mere end 97 procent af vores tog er for første gang nogensinde kørt til tiden igennem et helt år, og det er et utrolig flot resultat, som alle medarbejdere fortjener stor ros for. Vi arbejder hver eneste dag på at få vores tog af sted til tiden, og det er en fornøjelse, at det er lykkedes så godt i 2013”
(Arriva, 2013)

1.2 Problembeskrivning

För att upprätthålla hög punktlighet, jobbar Arriva aktivt med schemaläggning av sina tåglinjer. En väl fungerande tåglinje utan förseningar kräver god planering och denna kan delas in i tre delar: tidtabell(tidsplan), materialplan och tjänsteplan. Materialplanen avser se till att rätt resurser finns på plats vid rätt tid, huvudsakligen tågen. Tjänsteplanen syftar till att tillhandahålla en plan för att personal ska finnas på rätt plats i rätt tid.

En bra tidtabell behöver dock inte innebära att den ger en bra materialplan och på samma sätt kan en bra materialplan innebära en dålig tjänsteplan. Det är Arrivas ansvar att generera tidtabell, materialplan och tjänsteplan för sina tåglinjer på Jylland. I dagsläget är detta en helt manuell process med endast ett fåtal enkla hjälpmedel vilket gör det svårt och tidskrävande. En faktor som försvårar planeringen är att majoriteten av spåren är enkelspår vilket innebär att endast ett tåg åt gången kan befinna sig mellan två stationer.

Försenade tåg innebär straffavgifter och missnöjda kunder för Arriva. För att Arriva ska undvika straffavgifter får endast en viss andel av tågen vara högst 5 minuter sena vid ankomst, en regularitetsgrad. Notera att det är viktigt att även ta mindre förseningar i beaktande eftersom de kan upplevas negativt av resenärer. Om ett tåg är försenat kan det även innebära störningar för senare tåg och på så sätt kan tidtabellen sättas ur spel. Då mindre störningar i trafiken hör till vardagen är det viktigt att den tidtabell som tas fram klarar av att återhämta sig i dessa situationer.

I samband med längre avbrott i den normala tågtrafiken såsom underhåll av spår eller stationer kan man bli tvungen att lägga om tidtabellen. Som tidigare nämnts är framtagandet av fungerande tidtabeller komplicerat och tidskrävande. Dessutom kan ett stort antal justeringar behöva göras för att få en fungerande och kostnadseffektiv kombination av tidtabell, materialplan och tjänsteplan.

En bra planläggning leder till att fler tåg avgår i tid, vilket leder till en förbättrad kundupplevelse. Med högre kundnöjdhet kan Arriva attrahera fler kunder samt få en bättre beläggningsgrad, vilket i förlängningen ger en samhällsvinst i form av en miljöbesparing. Miljöbesparingarna realiseras genom att tåg blir ett mer attraktivt substitut av bilen, som ur miljösynpunkt är ett sämre val.

För att undersöka om liknande problemställningar har behandlats tidigare genomförs en litteraturstudie.

1.3 Utvärdering av tidtabeller och planering av tågtrafik

Studien tar sin utgångspunkt i forskningsområdet schemaläggning. Schemaläggning är ett relativt brett forskningsområde som är väl dokumenterat i den vetenskapliga litteraturen, framförallt inom flygindustrin men också i tågindustrin. Författarna har fokuserat på litteratur rörande tåg avsedda för passagertrafik då den skiljer sig markant från annan tågtrafik, till exempel godstransporter.

Olika problem som uppstår för en tågoperatör i samband med planering kan enligt Huisman et al (2005) klassificeras på ett antal olika sätt. Ett sätt är att se på planeringshorisonten, strategiskt, taktiskt, operationellt eller kortsiktigt. Ett annat sätt att klassificera är att se på var problemet uppstår, om det är globalt i tågnätet eller lokalt. Huisman gör följande indelning, se Figur 2.

level	time	central	local
strategic	10-20 years	rolling stock management	
	2-5 years	crew planning	
	every few years	line planning	
tactical	1/year	timetabling (basic)	platform assignment (basic)
	1/year	8 o'clock rolling stock assignment	
operational	6/year	timetabling (details)	platform assignment (details)
	6/year	rolling stock circulation	shunting
	6/year	crew scheduling	crew rostering
short-term	daily	timetabling	platform assignment
	daily	rolling stock circulation	shunting
	daily	crew scheduling	

Figur 2 - Klassificering av planeringsområden

Timetabling, eller framtagning av tidtabeller kan ses som en tågoperatörs huvudprodukt och bestäms i grunden på taktisk nivå (Huisman, et al., 2005). Det är vanligt att tidtabeller är cykliska, det vill säga att avgångar sker på fasta minuttider varje timme eller halvtimme. Cykliska tidtabeller är fördelaktiga eftersom de gör det lättare för passagerare att memorera tidtabellen och höjer därmed kundnöjdheten, men innebär samtidigt att det blir svårare att ta fram en fungerande tidtabell (Cacchiani & Toth, 2011). Det är dock inte bara tidtabellen i sig som är viktig utan också hur pålitlig och robust den är (Huisman, et al., 2005). För en enskild passagerare mäts tidtabellens pålitlighet utifrån passagerarens uppfattning av förseningar, men för att utvärdera en tidtabell behövs mer objektiva mått.

Ett vanligt mått för tidtabellers pålitlighet är punktlighet, som i sin tur kan definieras på flera olika sätt. En vanlig definition är hur många procent av tågen som är mindre än ett visst antal minuter sena vid ankomst (Huisman, et al., 2005). Vid internationella jämförelser är det vanligtvis 5 minuter och detta är även det mått som används på Arriva. Det är också viktigt att se på vilken medelförsening tåg har vid ankomst även i de fall då tåget inte är mer än 5 minuter försenat eftersom även detta upplevs som negativt av passagerarna.

Huisman et al (2005) presenterar tre olika tillvägagångssätt för att förutsäga pålitligheten i en given tidtabell. Dessa tillvägagångssätt är max-plus algebra, stokastisk analys och simulering.

Max-plus algebra

Max-plus algebra är en analytisk metod för att utvärdera tidtabeller. En nackdel med max-plus algebra är dock att det är en statisk metod, ingen litteratur finns som visar att stokastiska störningar använts i samband med max-plus algebra (Huisman, et al., 2005).

Stokastisk analys

Flera olika modeller för stokastisk analys av tidtabeller har tagits fram. Kroon et al (2008) beskriver en stokastisk optimeringsmodell som kan användas för att allokera bufferttider för att få fram en tidtabell som är maximalt robust mot stokastiska störningar i systemet. Denna typ av modeller kan fungera bra på enskilda linjer eller mindre tågnätverk men klarar inte av att hantera mer komplicerade strukturer (Huisman, et al., 2005).

Simulering

Det tredje tillvägagångssättet för att förutsäga pålitligheten i en given tidtabell är simulering. Ett antal olika mjukvaror för att simulera

tågnätverk finns för stora och små tågnätverk. Bergmark (1996) beskriver hur simuleringsmjukvaran SIMON använts på dåvarande Banverket för att analysera kapaciteten och hitta flaskhalsar i delar av Sveriges tågnät. Även i andra delar av Europa används simulering för förbättring av tidtabeller. Kaas (2000) beskriver hur simuleringsmjukvaran UX-SIMU använts i Köpenhamn, och i Nederländerna har simuleringsmjukvaran SIMONE använts för att simulera hela det Nederländska tågnätet (Middelkoop & Bouwman, 2000).

Efter denna genomgång står det klart att simulering är den metod som är bäst lämpad för att lösa problemet detta arbete grundar sig i. Detta eftersom max-plus algebra endast lämpas för statisk analys och att stokastisk analys inte hanterar stora system. De existerande simuleringsmodeller som undersökts har framförallt fokuserat på kapacitetsanalys av tågnät. Dessutom har ingen modell funnits som kan integrera materialplan och tjänsteplan. Andra nackdelar med att använda redan existerande simuleringsmodeller är att det medför kostnader för licenser samt att de inte går att anpassa till det egna tågnätet på samma sätt som en egenutvecklad modell.

Av dessa anledningar har författarna valt att själva ta fram simuleringsmodeller för utvärdering av Arrivas tidtabeller.

1.4 Simulering

Simulering är en imitation av verkliga processer eller system i tiden (Banks, 1998, p. 3). Det är ett kraftfullt verktyg eftersom en simuleringsmodell komprimerar både tid och rum och därmed gör det möjligt att testa teorier och planer som är allt för kostsamma för verkliga experiment (Laguna & Marklund, 2013, p. 254).

Simuleringsmodeller kan enligt Laguna & Marklund (2013, p. 255) klassificeras på tre sätt beroende på dess egenskaper:

Statisk eller dynamisk. Skillnaden mellan en statisk och en dynamisk modell är att i den dynamiska modellen har tiden betydelse men i den statiska modellen är tidsaspekten obetydlig.

Deterministisk eller stokastisk. I en deterministisk modell kan utfallet beräknas så fort inparametrarna är kända och en viss uppsättning av inparametrar ger alltid samma utfall. I en stokastisk modell varierar utfallet beroende på slumpmässiga händelser.

Diskret eller kontinuerlig. I en kontinuerlig simuleringsmodell representeras systemets tillstånd av variabler som förändras kontinuerligt i tiden (Banks, 1998, p. 43). I en diskret modell däremot sker förändringar endast i diskreta steg och mellan dessa steg är tillståndet oförändrat (Banks, 1998, p. 42).

Skillnaden mellan en diskret och en kontinuerlig simuleringsmodell ligger alltså i hur tiden avanceras och systemets tillstånd uppdateras. I en kontinuerlig modell avanceras tiden enligt följande algoritm:

1. Avancera tiden med Δt tidsenheter, där Δt är litet
2. Undersök om någon händelse inträffade under tidsintervallet och i så fall uppdatera systemet.
3. Upprepa från steg 1

I en diskret modell avanceras tiden istället enligt följande algoritm:

1. Sätt klockan till $t = 0$
2. Ta reda på vilken händelse som inträffar närmast
3. Avancera klockan till denna tidpunkt
4. Uppdatera systemets tillstånd
5. Uppdatera händelselistan
6. Upprepa från steg 2

Algoritmen för kontinuerlig avancering av tiden kan även ses som en diskret avancering om Δt är stort. En fördel med diskret simulering enligt algoritm två är att den är snabbare än kontinuerlig simulering eftersom tiden när inget händer hoppas över.

1.5 Syfte

Det övergripande syftet med detta examensarbete är att ta fram två simuleringsmodeller av Arrivas tåglinjer för att underlätta planläggningen av tidtabeller. Den ena modellen är dynamisk, deterministisk och diskret och den andra är dynamisk, stokastisk och diskret.

I den deterministiska modellen antas allt ske utan osäkerhet och variation, det vill säga att samtliga tåg ankommer och avgår enligt given tidtabell.

Den stokastiska modellen ska mer korrekt återspegla verkligheten genom att fånga variationer i transporttid samt uppehållstid på stationer.

Modellerna ska tas fram på ett sätt som gör att de i framtiden kan kompletteras med materialplan och tjänsteplan.

1.6 Avgränsningar

Arbetet fokuserar på framtagningen av simuleringsmodeller för testning av tåg tidtabeller. Modellerna utformas på så sätt att vilken tidtabell som helst kan testas givet att den förs in i rätt format. Dock krävs modifieringar om nya start- eller slutstationer införs. De för närvarande giltiga start- och slutstationerna ges av dagens tidtabeller för de givna tåglinjerna.

Modellerna tar ej hänsyn till större störningar som urspårade tåg, andra olyckor eller underhållsarbeten som kan innebära långa stopp eller omläggningar av trafik till andra spår eller ersättningsbussar. Anledningen till att ingen hänsyn tas till denna typ av oplanerade längre uppehåll är att de oavsett hur tidtabellen är utformad leder till störningar i trafiken. En anledning till att en ny tidtabell måste tas fram kan vara ett i förväg planerat underhållsarbete som förväntas pågå en längre tid. Detta innebär att tåg kan tvingas dirigeras om eller köra med reducerad hastighet. Då det är omöjligt att i förväg förutse var, när och hur detta kommer ske, kan det ej tas hänsyn till i modellen. För att underlätta vid en situation som denna utformas modellen på ett sådant sätt att en person som är bekant med simulerings-programvaran och tågsystemet utan större svårigheter kan göra nödvändiga modifikationer. Stationer där olika linjer möts antas alltid ha lediga spår för ankommande tåg och på så sätt kan varje linje

betraktas var för sig. Linjerna ska kopplas ihop i samband med att material- och tjänsteplan läggs till.

Tåg som körs av andra tågbolag på samma spår som Arriva tas ej hänsyn till i modellerna men det kan lätt göras givet att en tidtabell för dessa tåg kan tas fram och föras in på samma format som Arrivas tåg. Antalet tåg som körs av andra operatörer är få och antas därför inte ha någon större inverkan på systemet.

Trafikledning och tågförarens beslut vid enskilda situationer, till exempel om vem som kör först då två försenade tåg vill köra in på samma spår kan ej återskapas till fullo i modellerna. Anledningen är att osäkerhet råder kring enhetliga beslutsregler. Istället används enklare beslutsregler i simuleringsmodellerna.

Då arbetet innefattar framtagning av ett stort antal fördelningar för passagerarantal, stopptid på stationer och transporttid finns inte möjlighet att behandla varje fördelning för sig. Fokus läggs på att ta fram standardfördelningar som genom val av parametrar på ett godtagbart sätt kan representera datamaterialet.

1.7 Användningsområde

Modellerna har tre huvudsakliga användningsområden:

Deterministisk modell

- Identifiera om en tidtabell innehåller konflikter och deterministiskt precisera var konflikten uppstår. Skulle tidtabellen innehålla några konflikter upptäcks det i modellen genom att tåg tvingas vänta och därmed missar sin givna avgångstid. Syftet med denna modell är att kontrollera om en given tidtabell är tillämplig.

Stokastisk modell

- Jämföra tidtabeller. Möjlighet till att jämföra olika tabeller för att åskådliggöra hur förändringar ger effekter på det större systemet.
- Identifiera möjliga framtida problem. Detta kan genomföras på två olika sätt. Först är det möjligt att simulera och notera var den stokastiska modellen får konflikter eller var säkerhetsmarginalen är liten i systemet. Denna process kan påvisa var problem kan uppkomma eller uppkommer utan att det detekteras i nuläget. Den andra metoden utgår från att inparametrar ändras för att simulera en framtida förändring.

Notera att modellerna inte är färdiga produkter för direkt implementering, målsättningen är att ta fram ett tillvägagångssätt för modelluppbyggnad och en modellstruktur att bygga vidare på. Därför kommer den tredje punkten, att identifiera möjliga framtida problem, överlämnas till Arriva. Processen kommer även att visa om Arrivas infrastruktur av datainsamling har tillräcklig kvalitet eller behöver utvecklas innan ytterligare arbete med modellering bör göras.

1.8 Rapportens disposition

Kapitel 1: Inledning

I kapitel 1 beskrivs bakgrund, syfte och avgränsningar för examensarbetet. Fokus är att sätta in läsaren i problemet och vilka delar av problemet som projektet ämnar lösa, samt vad som överlämnas för vidare arbete. I kapitlet presenteras även rapportens disposition.

Kapitel 2: Metod

I kapitel 2 beskrivs den metodik som använts vid framtagning av simuleringsmodellerna, samt motivering till denna process. Kapitlet avslutas med diskussion om källkritik, samt en beskrivning av använd mjukvara.

Kapitel 3: Modellbeskrivning

I kapitel 3 ges en detaljerad bild av modellens uppbyggnad och hur den tagits fram. Första delen av kapitlet beskriver hur Arrivas tågnät är uppbyggt. Därefter presenteras hur den konceptuella modellen tagits fram följt av den deterministiska simuleringsmodellen och den stokastiska simuleringsmodellen.

Som avslutning presenteras den stokastiska simuleringsmodellen mer detaljerat, utifrån hur den ser ut i simuleringsprogrammet. Målet är att hjälpa läsaren att förstå modellen genom att belysa ett antal detaljer i simuleringsuppbyggnaden.

Kapitel 4: Verifiering och Validering

Kapitel 4 beskriver verifiering och validering som genomförts av simuleringsmodellerna. I kapitlet belyses likheter och olikheter mellan simuleringen och datamaterialet från Arriva för att påvisa moment när modellerna överensstämmer väl och där avvikelser inträffar. Kapitlet avslutas med förklaringar till avvikelserna.

Kapitel 5: Slutsats och rekommendation

Slutsatser och rekommendationer av arbetet presenteras tillsammans med hur väl syftet för projektet är uppfyllt. Därefter ges förslag på framtida studier.

2 Metod

I kapitlet beskrivs den metodik som använts vid framtagning av simuleringsmodellerna, samt motivering till denna process. Kapitlet avslutas med diskussion om källkritik, simulering samt en beskrivning av använd mjukvara.

2.1 Metodval

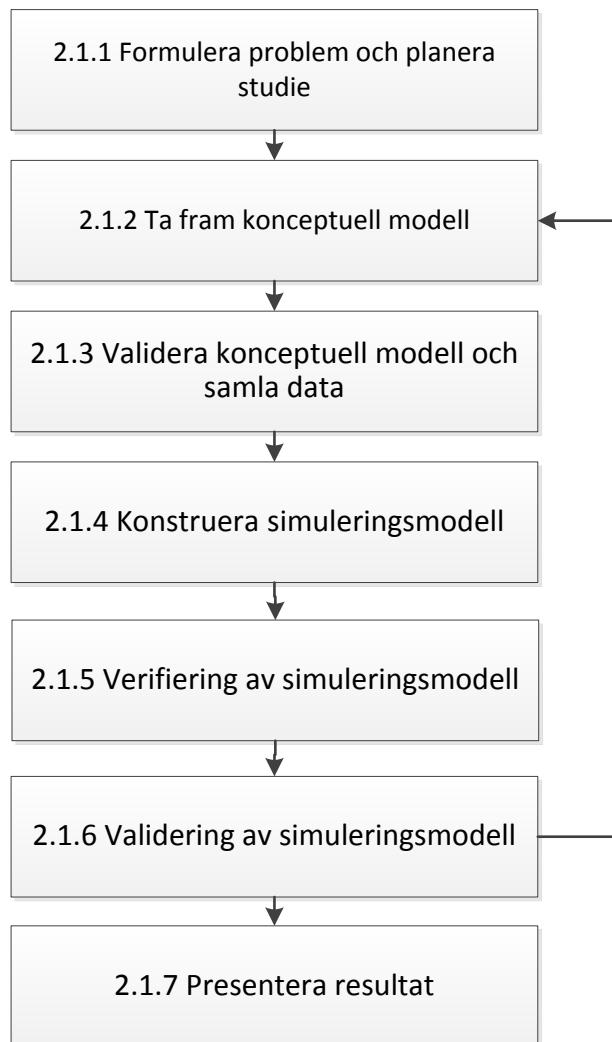
Arbetsmetoden i detta projekt tar sin utgångspunkt i den modell i 8 steg för stora simuleringsprojekt som beskrivs i boken *Introduction to Operations Research* (Hillier & Lieberman, 2010, pp. 959-963):

1. Formulera problemet och planera studien
2. Samla data samt formulera modell (konceptuell)
3. Validera konceptuell modell
4. Välj mjukvara och konstruera en simuleringsmodell
5. Validera resultat med verkligheten
6. Planera simulering
7. Genomför simulering och analysera resultat
8. Presentera rekommendationer

Metoden tar sin utgångspunkt i metoden ovan, dock krävs vissa modifikationer. I denna studie ligger fokus framför allt på själva modelleringen och därför kommer punkt 6 och 7 inte genomföras inom ramen för detta examensarbete. Metoden som används i detta examensarbete sammanfattas i Figur 3, och beskrivs därefter utförligare i avsnitt 2.1.1-2.1.7.

Processen innehåller itereringar av vissa steg. Steg 2-6 i Figur 3 kommer att genomföras minst två gånger. Första gången för uppbyggnad av den deterministiska modellen och andra gången för den stokastiska modellen. Den senare involverar logik för att hantera de stokastiska händelserna.

Metoden har valts för att den ger en bra struktur för fortsatt arbete samt att det är en etablerad metod för operationsanalytiska simuleringsprojekt.



Figur 3 – Processchema examensarbetets arbetsmetod

2.1.1 Formulera problem och planera studie

För att ta fram en detaljerad problembeskrivning genomfördes intervjuer med vår kontaktperson på Arriva. Utfallet stämde sedan av med handledaren för att tillsammans utforma en problemformulering.

Under detta steg formuleras mål, milstolpar, vägskäl, utdata, indata samt begränsningar. Hillier och Lieberman (2010, p. 959) rekommenderar att intervjuer sker med personer inom flera avdelningar och olika bakgrund. Inom detta projekt har intervjuer genomförts med Stefan Vidgren, vår kontaktperson på Arriva. Ytterligare intervjuer har inte ansetts nödvändiga på grund av Stefan Vidgrens breda kunskap inom både modellering och Arrivas verksamhet. Stefan har en bakgrund som universitetslärare och teknisk licentiat vid Lunds tekniska högskola och har stor kunskap om Arrivas verksamhet.

I detta steg diskuteras vilka program som ska användas i projektet. Ett program används för datahantering innan och efter simulering samt ett annat för själva simuleringen. ExtendSim valdes som simuleringsprogram och Microsoft Excel för datahantering. Programmen valdes primärt på grund av författarnas kunskap och vana av programmen.

Microsoft Excel och ExtendSim bedömdes av författare och handledare vara lämpliga för att genomföra projektet.

2.1.2 Ta fram konceptuell modell

En konceptuell modell av tågnätet används som grund till både den deterministiska simuleringsmodellen och den stokastiska. Den konceptuella modellen skapas utifrån insamlad data om tågsystemets struktur och logik. Datainsamlingen är gjord genom intervjuer med Stefan Vidgren. Modellen visualiseras med flödesdiagram som avspeglar sammanlänkningsen mellan stationer och transportsträckor. I detta steg definieras även de antaganden som gäller för den konceptuella modellen, och därmed för samtliga simuleringsmodeller.

2.1.3 Validering av konceptuell modell

Validering av den konceptuella modellen sker genom avstämning av modellen mot vår kontaktperson på Arriva. Under validering diskuteras antaganden, oklarheter och detaljer kopplade till modellen. Validering innebär en detaljkontroll av modellens logiska struktur.

2.1.4 Konstruera simuleringsmodell och samla data

När den konceptuella modellen har validerats översätts den till exekverbara simuleringsmodeller.

Simuleringsmodellerna utformats genom att bygga små moduler i form av stationer och transportsträckor som sedan skalas upp för att avspegla hela tågnätet. Modulhierarkierna är stationer/transportsträckor, linjer och hela systemet. Det vill säga, först byggs en station och en transportsträcka för att testa logiken i en liten skala. Sedan skapas ett flertal stationer och transportsträckor med ursprungsstationen som mall, vilket resulterar i en linje. Innan linjen är klar hanteras eventuella specialfall. Detta kan vara tåg som inte har ändstation i sista station på linjen eller stationer där det endast finns ett spår på stationen. Vid färdigställandet av samtliga linjer, sätts de ihop till ett helt system.

Vid byggandet av simuleringsmodellerna är det av största vikt att alltid beakta att de ska vara lätta att skala upp och modifiera. Modulerna skapas för att vara generella, där ändringar i parametrar eller struktur kan göras utan att ändra själva blockens uppbyggnad eller individuell inmatning. En viktig aspekt är att göra parameter- och strukturändringar genom generell kodning som kan ändras genom att ändra värden i indata. Detta medför möjlighet att göra flera ändringar snabbt, vilket är viktigt då ändringar kan bli tidskrävande i ett system med ett stort antal stationer.

Utöver den data som samlades in vid framtagning av den konceptuella modellen behövs Arrivas tidtabell eftersom den beskriver hur tågen ska skapas och röra sig i simuleringsmodellerna.

2.1.4.1 Deterministisk simuleringsmodell

För framtagningen av den deterministiska simuleringsmodellen används ingen ny data utöver tidtabellen och den data som samlas in i samband med framtagningen av den konceptuella modellen.

2.1.4.2 Stokastisk simuleringsmodell

Den stokastiska simuleringsmodellen grundar sig i den konceptuella modellen men kan betraktas som en påbyggnad och generalisering av den deterministiska modellen. Vid byggandet av den stokastiska modellen används den deterministiska modellen som grund, men med modifieringar för att addera de stokastiska moment som uppkommer på grund av osäkerhet och variation i transporttider mellan stationer och uppehåll på stationer. För att inkorporera dessa osäkerhetsmoment i simuleringsmodellen på ett adekvat sätt måste ytterligare data samlas in.

Datainsamlingen sker i fyra steg: insamling av rådata från Arrivas databas, rensning av data, skattning av sannolikhetsfördelningar för osäkerhetsmomenten samt formatering av data för import till simuleringsprogram.

Arriva har en databas som lagrar information med ett femtiotal parametrar för samtliga avgångar. Exempel på parametrar är: tågnummer, tågriktning, avgångstid från station, ankomsttid till station och antal kontrollanter. Data klassas som primärdata eftersom den inte har bearbetas av annan part med annat syfte efter insamling. En utförlig beskrivning av datamaterialet finns i avsnitt 3.2.

Efter insamlingen av primärdata sker en rensning av densamma. Rensning genomförs för att eliminera uppenbart felaktig och irrelevant data. Rensning innefattar t.ex. att ta bort eventuella extra mellanslag och danska bokstäver som kan innebära problem vid vidare hantering.

Efter rensning skattas statistiska fördelningar utifrån datamaterialet. Fördelningarna ska representera variation i transporttid och tid på stationer som finns i systemet.

För att kunna använda datamaterialet på ett smidigt och automatiserat sätt formateras data för att anpassas för import till simuleringsprogrammet.

2.1.5 Verifiering av simuleringsmodell

Verifiering innebär en detaljkontroll av simuleringsmodellernas logiska struktur. För att säkerställa att simuleringsmodellerna beter sig korrekt rent logiskt och inte innehåller programmeringstekniska fel kontrolleras de genom simulering med animering. Kontroller görs också av att data som importerats är korrekt och hamnar på rätt plats i modellerna. Delar av simuleringsmodellerna verifieras genom genomgång tillsammans med Stefan Vidgren.

2.1.6 Validering av simuleringsmodell

Efter simuleringsmodellerna är skapade är det viktigt att validera att utdata stämmer överens med verkligheten. Hillier och Lieberman (2010, pp. 961-962) rekommenderar att modellresultaten testas mot antingen en matematisk modell, mätning mot verkligheten genom fälttest eller intervjuer med nyckelpersoner om resultatens rimlighet och validitet. Ett annat alternativt sätt är att inspektera animeringen av simuleringen.

2.1.6.1 *Deterministisk modell*

Den deterministiska modellen valideras genom att Arrivas befintliga tidtabell används som indata. Vid testning av denna tidtabell ska inga förseningar uppkomma i det deterministiska fallet. Inspektion av animering sker även för att validera rimligt resultat på lägre nivåer. Stickprov görs för att kontrollera att tåg avgår och ankommer enligt den importerade tidtabellen.

2.1.6.2 *Stokastisk modell*

Den stokastiska modellen valideras genom att jämföra utdata från simuleringsmodellen med datamaterialet rörande ankomster och avgångar i systemet som erhållits från Arriva. Resultatets rimlighet valideras även genom samtal med Stefan Vidgren.

2.1.7 Presentera resultat

Presentation genomförs både skriftligt, i form av denna rapport, samt muntligt till ansvarig inom simuleringsområdet på Arriva. Rapporten och presentationen ska innehålla tillvägagångssätt, rekommendationer och validering av modellen (Hillier & Lieberman, 2010, p. 963).

2.2 Källkritik

Litteraturen i denna rapport är huvudsakligen facklitteratur och akademiska artiklar. För att ytterligare säkerhetsställa rapportens tillförlitlighet har litteraturens kvalité säkerställs genom konsultation av handledare.

Internetkällor har använts för att hämta bilder och bakgrundsdata om Arriva. Källorna kan anses som tillförlitliga eftersom Arriva har ett

anseende att leva upp till och bör därför ge en sådan rättvis bild som möjligt av sin verksamhet i Danmark.

Intervjuer har genomförts med Stefan Vidgren på Arriva. För att validera fakta hade ytterligare intervjuer kunnat genomföras med andra intressenter. Författarna tog dock beslutet att ytterligare intervjuer inte var nödvändigt på grund av Stefan Vidgrens breda kunskap om Arriva.

2.3 Använd mjukvara

Under denna rubrik beskrivs de program som använts för att genomföra projektet.

2.3.1 Microsoft Excel

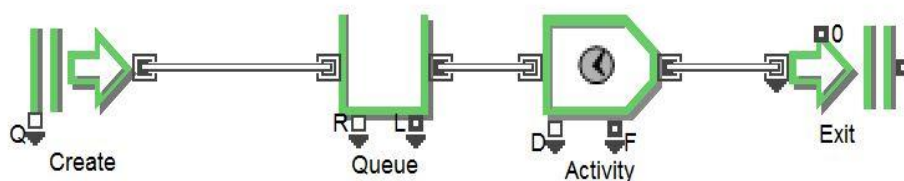
Microsoft Excel har i huvudsak använts för att hantera indata till modellen och utdata från modellen. Vid indatahantering har fokus varit på att formatera det datamaterial som erhållits från Arriva till ett importerbart format till simuleringsprogrammet. Till exempel har Pivot-tabeller använts för att sortera ut relevant data samt VBA-kod för beräkningar och formatering.

För utdata har Microsoft Excel används för att presentera data och förenkla analys. Exempelvis har utdata delats upp efter både tåg och station, vilket gör att det blir lättare att undersöka var och varför förseningar uppstår. Microsoft Excel har även använts för att visualisera data, samt att räkna ut nyckeltal som regularitet.

2.3.2 ExtendSim

ExtendSim är ett kommersiellt simuleringsprogram som utvecklas och säljs av företaget Imagine That Inc (Imagine That Inc, 2014). Programvaran används för att bygga simuleringsmodeller med hjälp av förprogrammerade block. Varje block representerar olika processer, till exempel en aktivitet där tågen måste vänta enligt en sannolikhetsfördelning eller kontroll om tågspåret är ledigt. Blocken binds samman med kontaktlinjer som överför information eller objekt mellan dem. Både kontinuerliga och diskreta simuleringsmodeller kan skapas i Extendsim.

I Figur 4 återfinns ett exempel på en enkel simuleringsmodell byggd i Extendsim. Blocket längst till vänster skapar objekt som sedan hamnar i nästa block märkt "Queue". I köblocket väntar objektet tills nästa block, "Activity-blocket", är ledigt. Där uppehålls objektet en viss tid enligt den sannolikhetsfördelning som specificerats i blocket. När objektet släpps från Activity-blocket hamnar det i "Exit-blocket" och tas därmed ur simuleringsmodellen.



Figur 4 - Exempel på modell i Extendsim

2.3.3 Statfit

För att underlätta analysen av sannolikhetsfördelningar har programvaran Statfit används. Statfit är en programvara som används för att anpassa statistiska fördelningar till ett datamaterial. I Statfit läses datamaterial in och testas mot standardfördelningar genom att ett goodness-of-fit test utförs. Ett goodness-of-fit test är ett statistiskt hypotestest som görs för att undersöka hur väl ett datamaterial kan approximeras med en statistisk fördelning på en viss signifikansnivå. Detta görs genom att på olika sätt jämföra hur observerade värden skiljer sig från motsvarande värden som fås med den föreslagna fördelningen. De goodness-of-fit test som kan utföras i Statfit är Chi-square, Kolmogorov-Smirnov och Anderson Darling. Vilket test som ska utföras väljs manuellt i Statfit av användaren, om inget test väljs används Kolmogorov-Smirnov automatiskt.

Testet ger sedan en inbördes ranking mellan fördelningarna som visas tillsammans med en graf där testade fördelningar plottas mot datamaterialet. Statfit är en fristående programvara, men ingår i programpaketet ExtendSim suite.

3 Modellbeskrivning

I detta kapitel ges en detaljerad bild av modellens uppbyggnad och hur den tagits fram. Första delen av kapitlet beskriver hur Arrivas tågnät är uppbyggt. Därefter presenteras hur den konceptuella modellen tagits fram följt av den deterministiska simuleringsmodellen och den stokastiska simuleringsmodellen.

Som avslutning presenteras den stokastiska simuleringsmodellen mer detaljerat, utifrån hur den ser ut i simuleringsprogrammet. Målet är att hjälpa läsaren att förstå modellen genom att belysa ett antal detaljer i simuleringsuppbyggnaden.

3.1 Beskrivning av Arrivas tågnät i Danmark

Arrivas tågnät är koncentrerat till Jylland och består av 81 stationer fördelat på 7 sträckor. Tågnätet täcker större delen av Jylland och fortsätter även ner till Niebüll i Tyskland. De 7 sträckorna är:

- Tåglinje 75, Aarhus - Skjern
- Tåglinje 82, Esbjerg – Niebüll
- Tåglinje 84, Varde – Norre Nebel
- Tåglinje 85, Skjern - Esbjerg
- Tåglinje 88, Struer - Skjern
- Tåglinje 92, Struer - Thisted
- Tåglinje 95, Aarhus - Struer

Majoriteten av sträckorna körs endast av Arriva men på några delsträckor delas spåren med statliga DSB. Arrivas 41 tåg drivs av diesellok och det är därför viktigt att planera hur och var loken ska tankas och ändå klara av att hålla sin tidtabell, vilket är en del av materialplanen. Alla tågen är av samma typ men man har valt att dela upp avgångarna i ”vanliga tåg” och ”Rex-tåg”. Vanliga tåg stannar på samtliga stationer mellan start och slutstation medan Rex-tågen kör förbi vissa stationer och fungerar därför som ett sorts snabbtåg. Samtliga linjer är i drift från cirka fem på morgonen till cirka tolv på

kvällen alla dagar i veckan och antalet avgångar varierar mellan en och fyra i timmen beroende på linje och tid på dygnet. I regel är det enkelspår på transportsträckor och dubbelspår på stationer, men i undantagsfall vid särskilt belastade partier finns även dubbelspår på transportsträckor och på ett fåtal stationer finns endast enkelspår.

3.2 Beskrivning av datamaterial

Arriva har en databas som lagrar information med ett femtiotal parametrar för samtliga avgångar. Exempel på parametrar är: tågnummer, tågriktning, avgångstid från station, ankomsttid till station, antal kontrollanter med mera. Vi klassificerar denna data som primärdata eftersom den inte har bearbetats efter insamling.

Registrering av händelser i systemet sker automatiskt genom två system: GPS-mätare och kameror. GPS-mätarna är placerad på varje tåg för att mäta ankomst och avgångstider till varje station. Tåget anses vara framme vid en station när det når fram till GPS-kordinaten för en station samt att hastigheten är tillräckligt låg. För att mäta passageraraktivitet används kameror som är monterade vid dörrarna på samtliga tåg. Med hjälp av dessa kameror kan antalet på- och avstigande noteras. Registrering av data köper Arriva in av en extern part och materialet levereras månadsvis.

För att ta fram parametrar för sannolikhetsfördelningar samt validering av simuleringsmodellerna har ett datamaterial med följande information använts:

- Dato
 - Datum för observationen
- Ugedag
 - Vilken veckodag observationen gjordes
- Linie
 - Vilken linje mätningen kommer från
- LinieRetning
 - Vilken riktning tåget åker i
- Tognummer
 - Avgångens nummer
- FraStationsNavn

- Vilken station mätningen är gjord vid
- TilStationsNavn
 - Nästkommande station
- AnkomstPlanlagt
 - Planlagd ankomsttid till "FraStationsNavn"
- AnkomstFaktisk
 - Faktisk ankomsttid till "FraStationsNavn"
- AvgångPlanlagt
 - Planlagd avgångtid från "FraStationsNavn"
- AvgångFaktisk
 - Faktisk avgångtid från "FraStationsNavn"
- PassIn
 - Antal passagerare som går på tåget
- PassOut
 - Antal passagerare som går av tåget
- PassOnboard
 - Antal passagerare ombord på tåget
- HelTime
 - Vilken timme på dygnet mätningen utförts

I Tabell 1 visas vilka vilken information som har använts till att ta fram de sannolikhetsfördelningar som används i den stokastiska simuleringsmodellen för att representera de osäkerhetsmoment som finns i systemet och vilken information som använts för validering av simuleringsmodellerna. En noggrann beskrivning hur fördelningar tagits fram och validering genomförts beskrivs i avsnitt 3.3.6 respektive 4.2. Denna information från perioden 2012-01-01 till 2012-12-31 refereras till i rapporten som datamaterialet.

Information	Fördelningar			Validering
	Antal passagerare	Stopptid	Transporttid	
Dato		X	X	X
Ugedag	X	X	X	X
Linie	X	X	X	X
LinieRetning	X	X	X	X
Tognummer		X	X	
FraStationsNavn	X	X	X	X
TilStationsNavn		X	X	X
AnkomstPlanlagt		X	X	X
AnkomstFaktisk		X	X	X
AvgångPlanlagt		X	X	X
AvgångFaktisk		X	X	X
PassIn	X	X		X
PassOut	X	X		X
PassOnboard	X	X		X
HelTime	X			X

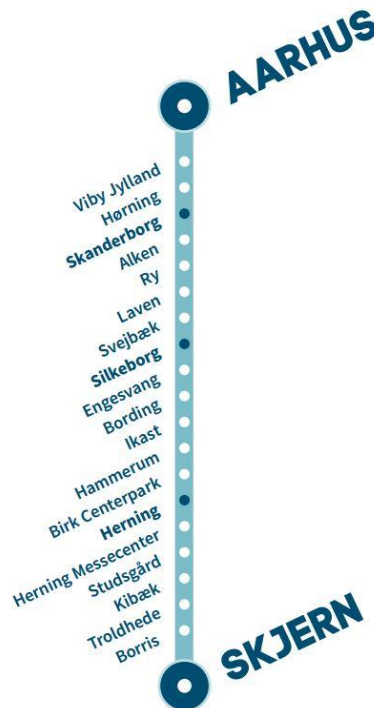
Tabell 1 – Beskrivning av användning av information från Arrivas databas

3.3 Framtagningen av simuleringsmodeller

För att underlätta framtagningen av simuleringsmodellerna gjordes först en konceptuell modell av tågnätet. Den kan ses som ett underlag till de färdiga simuleringsmodellerna och syftet var att få en bättre bild av hur de skulle struktureras, vilken in- och utdata som behövdes och vilka antaganden som bör göras. Den konceptuella modellen har sedan modifierats under arbetets gång i och med att ny information framkommit och simuleringsmodellerna tagit form. För att få en bättre förståelse för hur den konceptuella modellen tas fram beskrivs först tåglinjens struktur och hur ett tåg färdas längs linjerna.

3.3.1 Tåglinjernas struktur

Tågnätet består av 7 sträckor. Dessa sträckor består av ett antal stationer med antingen enkelspår eller dubbelspår mellan sig. Som exempel visas linje 75 Aarhus – Skjern, se Figur 5. Sträckan består av 21 hållplatser samt en ”teknisk” station vilket innebär att den endast används för datainsamling.



Figur 5 - Tåglinje Aarhus - Skjern

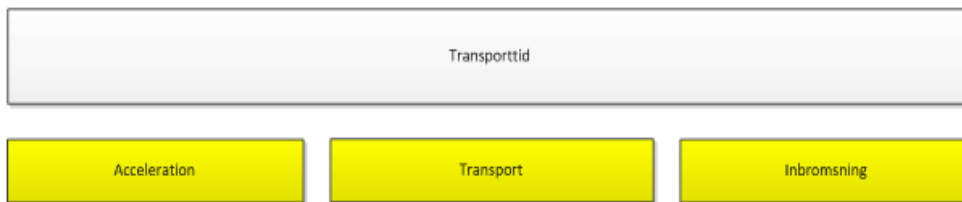
Transportsträckorna mellan Skjern - Skanderborg är enkelspåriga och därmed kan endast ett tåg åt gången befinna sig mellan två stationer oberoende av riktning. Sträckan Skanderborg – Aarhus däremot är dubbelspårig och tillåter fler tåg åt gången. På sträckorna Aarhus – Viby Jylland, Hørning – Skanderborg och Viby Jylland – Hørning tillåts 1 tåg, 3 tåg respektive 5 tåg åt gången i varje riktning. Att antalet tillåtna tåg per spår varierar beror på vilka signalmöjligheter som finns på de olika sträckorna. I normalfallet signaleras det på stationen när spåret fram till nästa station är fritt och tåget tillåts då lämna stationen. På linje 75 Aarhus – Skjern är stationerna Alken, Birk Centerpark, Herning Messecenter, Studsgård och Troldhede enkelspåriga. Detta innebär att tåg ej kan mötas på dessa stationer och eftersom spåren som leder till och från dessa stationer också är enkelspåriga uppkommer en längre sammanhängande sträcka där tåg ej kan mötas. Som exempel antas ett tåg stå på station Herning beredd att åka i riktning mot Kibæk och Skjern. Eftersom de två nästkommande stationerna Herning Messecenter och Studsgård är enkelspåriga måste hela sträckan Herning – Kibæk vara fri för att tåget ska få lämna stationen.

3.3.2 Beskrivning av tågets färd

Tågets färd kan delas upp i två delar, tid mellan stationer och tid på stationer. I följande avsnitt förklaras dessa tiders komponenter mer ingående.

3.3.2.1 Tid mellan stationer - Transporttid

Ett tåg färdas från startstation till slutstation enligt en given tidtabell. Tiden det tar att färdas mellan två stationer, transporttiden, kan delas upp i tre delar: acceleration, transport och inbromsning, se Figur 6. Den första delen, *acceleration*, är den tiden det tar för ett tåg att accelerera till marschfart. Motsvarande tid när tåget kommer fram till stationen är *inbromsning*, vilket är den tiden det tar för tåget att bromsa från marschfart. Tiden mellan är *transport* som alltså är den tid det tar för tåget att färdas i marschfart mellan stationerna. Givet att ett tåg färdas från station A till station B och stannar på båda stationerna består transporttiden av samtliga tre delar. Om tåget ska köra förbi station B behöver den däremot ingen inbromsningstid. Den totala transporttiden mellan A och B blir då acceleration plus transport eftersom inbromsningen är obefintlig. På motsvarande sätt behövs ingen accelerationstid när tåget fortsätter från station B till C.

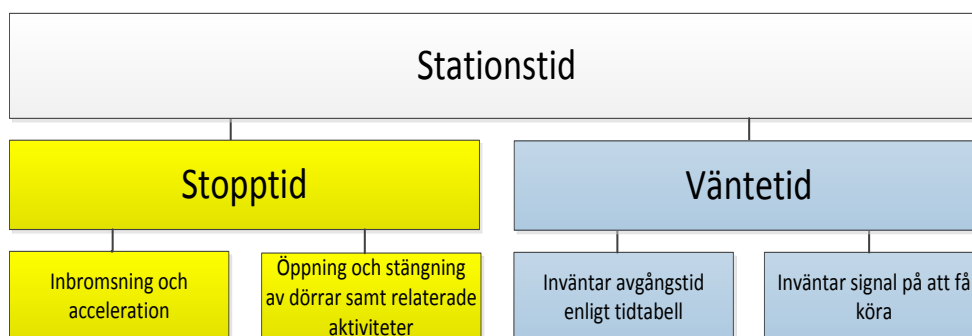


Figur 6 - Transporttidens komponenter

3.3.2.2 Tid på stationer - Stopptid och stationstid

Stationstiden definieras som den totala tiden som ett tåg spenderar på stationen. Stationstiden kan i sin tur delas in *stopptid* och *väntetid*, se Figur 7. *Väntetid* är den tid som tåget måste stå på stationen för att invänta sin avgångstid samt vänta på att spåret framför blir ledigt. Det är inte alltid som ett tåg måste göra detta, till exempel om ett tåg är sent och därför inte behöver invänta sin avgångstid, således är väntetiden ibland obefintlig. I *stopptiden* ingår de moment i stationstiden som alltid genomförs. Inbromsning och acceleration genomförs varje gång ett tåg ankommer till respektive avgår från en station. Observera att denna inbromsning och acceleration inte överlappar med den som beskrivs för transporttiden i Figur 6. Den andra komponenten av stopptiden är öppning och stängning av dörrar samt relaterade aktiviteter. Med relaterade aktiviteter menas den tid det tar för föraren att försäkra sig om att inga fler passagerare skall gå av eller på tåget innan föraren stänger dörrarna igen. Minsta tid för denna del av stopptiden är 20 sekunder eftersom det är den tid det tar att öppna och stänga dörrarna. Observera att tiden för öppning och stängning av dörrar samt relaterade aktiviteter kan variera beroende hur många som går av eller på tåget.

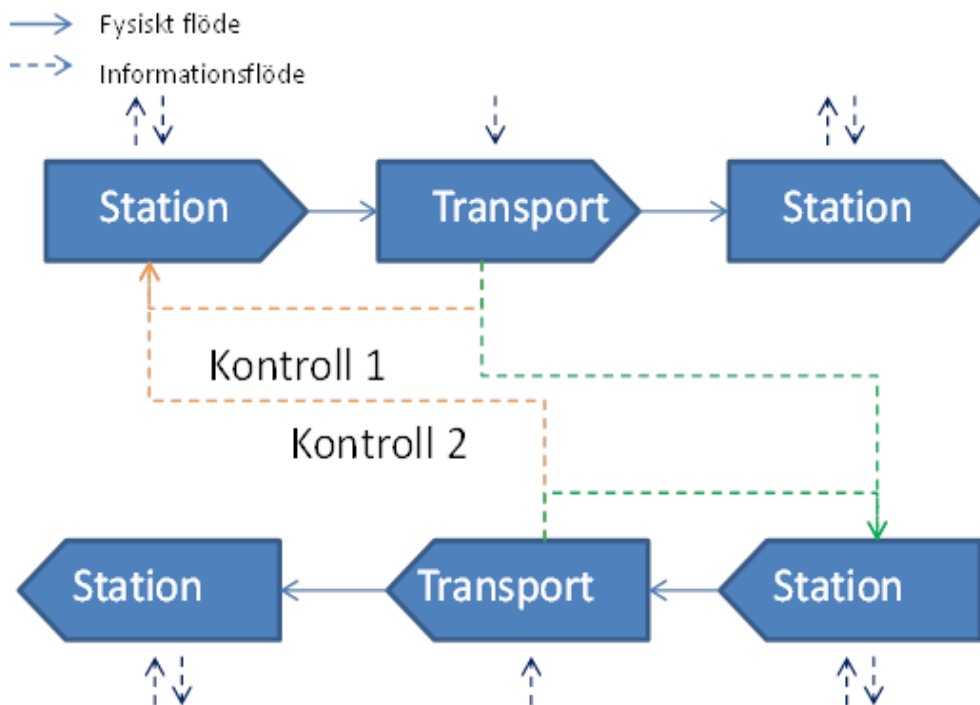
Denna uppdelning av stationstid i stopptid och väntetid är central i framtagningen av den stokastiska modellen och beskrivs mer i detalj vid framtagning av sannolikhetsfördelningar i avsnitt 3.3.6.3.



Figur 7 - Stationstidens komponenter

3.3.3 Konceptuell modell

Den konceptuella modellen är det andra steget i framtagningen av simuleringsmodellerna. Då modellen ska vara generell och lätt att skala upp valde författarna att bygga upp modellen i moduler. De två huvudsakliga modulerna är stationer och transporter. Eftersom en tåglinje har tåg som avgår i båda riktningarna motsvaras en station av två stationsblock och en transportsträcka av två transportblock, se Figur 8. Det vill säga om tåget befinner sig på den undre transportsträckan är det på väg åt vänster i modellen medan om det



Figur 8 - Konceptuell modell

befinner sig på den övre delen är tåget på väg till höger i modellen.

Stationsblocken genererar utdata med information om tåget är försenat, i tid, eller för tidigt. Indata till stationsblocket i den deterministiska modellen är den tid som tåget planeras stå på stationen enligt tidtabellen. Transportblockets indata i den deterministiska modellen är den planerade transporttiden. För den stokastiska simuleringsmodellen tillkommer ytterligare in- och utdata i form av tider som slumpgenereras vilket diskuteras senare i rapporten.

Till varje station skickas kontrollsignaler från nästkommande transportsträcka samt motsvarande sträcka i motsatt riktning för att signalera för tåg på stationen om spåret är fritt eller upptaget. I de fall då en eller flera stationer endast har enkelspår sker även dessa kontroller längre fram. Till exempel om ett tåg står på station 1 och station 2 endast har enkelspår så kontrolleras spåret mellan station 1 och station 2 samt spåret mellan station 2 och station 3. Kontroll sker även av spåret på station 2 eftersom hela sträckan mellan station 1 och 3 måste vara fri för att tåget ska kunna lämna stationen. I grundfallet sker alltså två kontroller, det vill säga spåret mellan nuvarande och nästkommande station i båda riktningar. Antalet kontroller ökar sedan med fyra för varje efterföljande station som har enkelspår.

I simuleringsmodellen har valdes att för varje tåg skapa två parallella objekt. Ett som motsvarar det fysiska tåget och ett som motsvarar tidtabellen. Tidtabellen är helt oberoende av yttre faktorer och finns endast i modellen som en sorts referens för att se om tåget är för tidigt eller försenat och se till att tåget inte avgår före sin planerade avgångstid. Denna uppbyggnad har valts för att underlätta ytterligare påbyggnad av modellen med bland annat materialplan och tjänsteplan. Dessa utbyggnader kan då läggas till som ett ytterligare "lager" genom att skapa nya parallella objekt motsvarande materialplan och tjänsteplan i modellen.

I den konceptuella modellen görs ett antal antaganden. Dessa antaganden gäller således för både den deterministiska simuleringsmodellen och den stokastiska simuleringsmodellen. Antagandena är:

1. Ett tåg avgår aldrig innan angiven avgångstid

Om ett tåg är redo att åka från stationen före avgångstiden i tidtabellen väntar det på stationen tills denna tid är nådd. Detta antagande är rimligt eftersom ett tåg ej bör avgå före sin tidtabell för att alla passagerare som tänkt åka med tåget ska ha en möjlighet att hinna med.

2. Endast ett tåg åt gången kan befinna sig på ett spår mellan två stationer om inget annat anges

Som tidigare nämnts är standardförfarandet att signal endast ges på stationen för att indikera att spåret fram till nästa station med dubbelspår är fritt. Därmed kan endast ett tåg befinna sig på spåret åt gången även om de åker i samma riktning. Ett fåtal undantag till denna regel återfinns i simuleringsmodellen.

3. Ett tåg kan inte köra förbi ett annat tåg

Då avgångar på dessa tåglinjer är så pass utspridda under dagen anses det vara mycket osannolikt att en omkörning blir aktuell. I verkligheten är omkörningar mycket ovanliga och det finns inget standardförfarande för denna situation. En omkörning är dessutom mycket svårt att behandla i modellen då den bygger på en tidtabell med tåg angivna i kronologisk ordning.

4. Stationer som är knytpunkter har alltid tillgängliga spår för inkommande tåg

De stationer där olika tåglinjer möts antas alltid ha lediga spår för inkommande tåg. Antagandet har gjorts i samråd med Stefan Vidgren. Detta innebär att varje tåglinje kan behandlas individuellt.

I följande avsnitt beskrivs den deterministiska simuleringsmodellen och den stokastiska samt vilka antaganden som är unika för respektive modell.

3.3.4 Deterministisk simuleringsmodell

Den deterministiska simuleringsmodellen syftar till att testa en given tidtabell för att se om den är konfliktfri under optimala förhållanden, utan slumpmoment som ger upphov till förseningar. Om en tidtabell inte klarar testet med den deterministiska modellen innebär det att fler tåg är planlagda på samma sträcka under en och samma tidpunkt och måste därför planeras om. När en tidtabell klarar detta test kan man gå vidare och utvärdera den med hjälp av den stokastiska simuleringsmodellen för att se hur den klarar sig under mer verklighetstroga förhållanden.

Antaganden

För att veta hur den deterministiska simuleringsmodellen ska byggas måste ett antal antaganden utöver de som presenteras för den konceptuella modellen i avsnitt 3.3.3 göras. Följande antaganden görs angående den deterministiska simuleringsmodellen:

1. Transporttiden är fast och enligt tidtabell

Genom att avläsa en given tidtabell beräknas den planerade transporttiden mellan varje station för samtliga angivna avgångar. En transporttid mellan samma två stationer kan alltså variera beroende på avgång. Eftersom den deterministiska modellen inte innehåller någon osäkerhet är transporttiden i simuleringsmodellen alltid densamma som den planerade transporttiden enligt tidtabellen.

2. Stoptiden är fast och enligt tidtabell

På samma sätt som transporttiden beräknats utifrån en given tidtabell beräknas även den tid som ett tåg planeras stå på stationen. Även denna tid kan variera mellan avgångar. En utförligare beskrivning av hur tidtabellen behandlas finns i avsnitt 3.3.6.1. Stopptiden i modellen är alltid densamma som den planerade stationstiden enligt tidtabell. Ett tåg kan dock bli tvunget att stå kvar på stationen efter planerad stationstid. Detta sker om spåret framför ej är fritt vid planerad avgångstid. Detta ger då upphov till en försening vilket registreras i simuleringsmodellen och kan avläsas i dess utdata.

In- och utdata

Stationsblocken i den deterministiska simuleringsmodellen genererar utdata med information om tåget är försenat, i tid, eller för tidigt. Indata till stationsblocket är den tid som tåget planeras stå på stationen. Transportblockets indata är den planerade transporttiden. En sammanställning av in- och utdata ses i Tabell 2.

Stationsblock	
<u>Indata:</u>	Planerad stationstid enligt tidtabell
<u>Utdata:</u>	Försening Buffert
Transportblock	
<u>Indata:</u>	Planerad transporttid enligt tidtabell

Tabell 2 - Indata och utdata, deterministisk simuleringsmodell

3.3.5 Stokastisk simuleringsmodell

En tidtabell som klarat testet med den deterministiska modellen och därmed är helt konfliktfri i ideala förhållanden ska sedan testas i den stokastiska simuleringsmodellen för att se hur den klarar sig i mer verklighetstroga förhållanden. För att bygga den stokastiska modellen måste data studeras för att ta fram sannolikhetsfördelningar på grund av osäkerheten som finns i transporttider och stationstider. Logiken i systemets uppbyggnad följer dock den konceptuella modellen, se Figur 8.

Antaganden

För den stokastiska simuleringsmodellen läggs följande antaganden till utöver de som beskrivs för den konceptuella modellen i avsnitt 3.3.3:

1. Transporttiden är stokastisk

Transporttiden mellan två sträckor är inte alltid exakt densamma. Anledningarna till variation kan vara många, alltifrån vardagliga företeelser som skiftande väderförhållanden och smärre förseningar som tvingar tåget att köra snabbare eller långsammare till större missöden som en personolycka på perrongen, urspårningar eller större snöfall med mera.

2. Stopptiden är stokastisk och beror på antalet av- och påstigande

Tiden som ett tåg står still på en station varierar av olika anledningar. Det kan bero på antalet av- och påstigande, tekniska problem med tåget eller att ett tåg är tvunget att vänta på att lämna stationen. Den mest konkreta kopplingen till ”vardaglig” variation i stationstid antas vara antal av- och påstigande. Desto fler som ska av och på ju längre tid behöver tåget på stationen. Genom att utgå från att antal av- och påstigande påverkar stopptiden kan simuleringsmodellen även användas för att undersöka huruvida en tidtabell är bra även vid en eventuell framtida ökning eller minskning av passagerare.

3. Minsta stopptid är 20 sekunder

Den absolut minsta stopptiden för ett tåg är 20 sekunder då detta är den tid det tar att öppna och stänga dörrarna. Denna tid kan varieras i modellen.

4. Vid framtagning av fördelningar tas endast hänsyn till ”vardaglig” variation

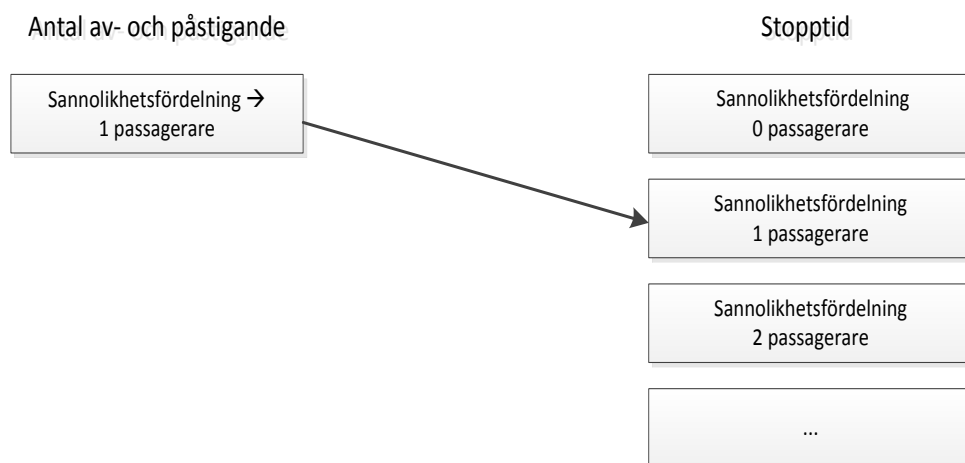
Denna modell syftar till att underlätta planeringen och framtagningen av nya tidtabeller som är robusta och klarar att hålla en hög regularitet under vardagliga förhållanden. Därför tar simuleringsmodellen ej hänsyn till sådana händelser som orsakar större avbrott i trafiken. Exempel på sådana händelser är: urspårningar, olyckor på perrongen, stora snöfall, fallna träd på spåret och större tekniska fel på tåget.

Vill man testa effekten av en större störning av denna typ och till exempel se om systemet klarar av att återhämta sig kan modellen lätt modifieras för att simulera en sådan händelse.

5. Regularitetsbrott inträffar då ett tåg är fem minuter sen eller mer vid ankomst till en station

Detta är definitionen på regularitet enligt Arriva. Denna tid ska i simuleringsmodellen gå att variera för att se hur regulariteten ändras om regularitetskravet sänks eller höjs.

I simuleringsmodellen finns som tidigare nämnts tre olika stokastiska moment. Det är transporttid mellan stationer, antal av- och påstigande vid stationerna och stopptid på stationen. Dessa stokastiska moment representeras av sannolikhetsfördelningar i simuleringsmodellen. Eftersom stopptiden varierar beroende på antal av- och påstigande blir resultatet passagerarantalets sannolikhetsfördelning avgörande för valet av inparametrar till sannolikhetsfördelningen för stopptid, se exempel i Figur 9. En fördel med att ha två fördelningar för att generera stopptiden är att antalet av- och påstigande kan varieras för att simulera olika framtida scenarion. Till exempel hur tidtabellen fungerar om passagerarantalet ökar med 10 %.



Figur 9 – Exempel, val av sannolikhetsfördelning för stopptid utifrån antal av- och påstigande

In- och utdata

Den stora skillnaden mellan den deterministiska simuleringsmodellen och den stokastiska ligger i stations- och transportmodulernas uppbyggnad, framför allt vilken in- och utdata som skickas. En sammanställning av in- och utdata för den stokastiska simuleringsmodellen ges i Tabell 3.

Stationsblock	
<u>Indata:</u>	Planerad stationstid enligt tidtabell Inparametrar till sannolikhetsfördelning för antal av och påstigande Inparametrar till sannolikhetsfördelning för stokastisk stopptid
<u>Utdata:</u>	Buffert medelvärde före/efter station Buffert standardavvikelse före/efter station Försening medelvärde före/efter station Försening standardavvikelse före/efter station Antal passerade tåg medelvärde Antal passerade tåg standardavvikelse Antal regularitetsbrott medelvärde Antal regularitetsbrott tåg standardavvikelse Regularitet medelvärde Regularitet standardavvikelse
Transportblock	
<u>Indata:</u>	Planerad transporttid enligt tidtabell Parametrar till sannolikhetsfördelning för stokastisk transporttid

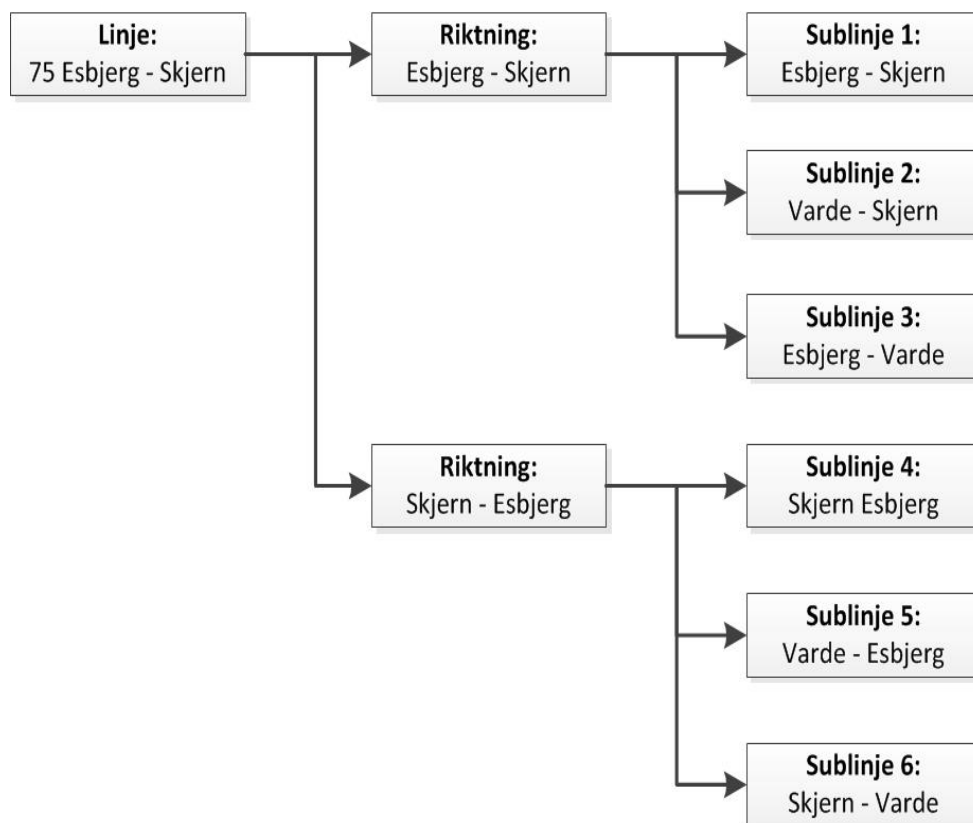
Tabell 3 - Indata och utdata, stokastisk simuleringsmodell

3.3.6 Analys av indata

Indata till den stokastiska simuleringsmodellen kan delas in i fyra delar. Tidtabell, fördelningar för passagerare, fördelningar för stopptid och fördelningar för transporttid. I följande avsnitt beskrivs hur dessa indata har analyserats. Indata till den deterministiska simuleringsmodellen är endast tidtabell och den är gemensam för de två modellerna. I Appendix B återfinns en teoretisk genomgång av de sannolikhetsfördelningar som används i den stokastiska simuleringsmodellen.

3.3.6.1 Tidtabell

Tidtabellen hämtas först i Excelformat och är uppdelad på de sju olika tåglinjerna. För att kunna bryta ner och hantera tidtabellen har varje linje delats upp i två riktningar som sedan delats upp i fler ”sublinjer”. En linje har minst två sublinjer men fler sublinjer skapas om det finns tåg som startar eller slutar på en annan station än linjens ändstationer. Ett exempel ses i Figur 10. Totalt finns 7 linjer, 14 riktningar och 36 sublinjer.



Figur 10 - Exempel på tåglinje uppdelad i sublinjer

Ett VBA-script söker igenom den givna tidtabellen för varje linje, beräknar hur länge tåget ska stå på respektive station, hur lång transporttiden är mellan varje station, vilken sublinje tåget tillhör samt vilken starttid tåget har. Resultatet blir en matris för varje linje, se Figur 11.

Tripnummer	Departure	Arrival	From	To
5300	5:12:00	5:17:30	Silkeborg	Svejbæk
5300	5:18:00	5:22:30	Svejbæk	Laven
5300	5:23:00	5:27:30	Laven	Ry
5300	5:30:30	5:35:30	Ry	Alken
5300	5:36:00	5:41:00	Alken	Skanderborg
5300	5:42:30	5:48:00	Skanderborg	Hørning
5300	5:48:30	5:55:00	Hørning	Viby Jylland
5300	5:55:30	6:01:00	Viby Jylland	Århus H
...

Avgångstid	Tripnummer	Sublinje	Transporttid 1	Stationstid 1	Transporttid 2	Stationstid 2	...
5:12:00	5300	Silkeborg Århus	330	30	270	30	...
...
...
...
...
...
...

Figur 11 - Exempel på del av tidtabell före och efter VBA-script

Ur den nya tabellen i Figur 11 kan man nu avläsa att tåg nummer 5300 ska avgå klockan 05.12 från Silkeborg mot Århus, den första transportsträckan tar 330 sekunder, på den första stationen stannar tåget 30 sekunder och så vidare. I tabellen går det ej att avläsa vilken tid som hör till vilken transportsträcka eller station, endast i vilken ordning de kommer. Tågets sublinje avgör var i simuleringsmodellen tåget avgår och den första transportmodulen som tåget passerar läser av Transporttid 1, nästa läser av Transporttid 2 och så vidare. På samma sätt läses stationstiderna av i stationsmodulerna. Omskrivningen av den givna tidtabellen görs för att den ska kunna importeras till en databas i ExtendSim där önskad information kan avläsas automatiserat. Tidtabellerna för de sju tåglinjerna lagras i samma databas men i varsin tabell.

3.3.6.2 Sannolikhetsfördelningar och parametrar för antal passagerare

Eftersom tiden ett tåg står på stationen antas variera med antalet passagerare som går av och på tågen tas sannolikhetsfördelningar fram för antal av- och påstigande vid stationerna. Intensiteten i antalet av- och påstigande passagerare varierar mellan stationer men också mellan olika veckodagar och tider på dygnet och är även beroende av vilken riktning tåget åker i. I samråd med Stefan Vidgren bestämdes att endast studera variationen under måndag - torsdag eftersom de

dagarna följer samma mönster samt täcker den största delen av den totala trafiken.

Sannolikhetsfördelningar tas fram för varje kombination av station, riktning och tid på dygnet. Mängden fördelningar uppgår till cirka 3800 (antal stationer * antal timmar på dygnet * två riktningar). På grund av det stora antalet fördelningar kan de inte behandlas individuellt utan måste representeras av en standardfördelning där parametrarna för varje kombination av station, tidpunkt och riktning kan tas fram på ett mer eller mindre automatiserat sätt utifrån datamaterialet. En aspekt är också att det ska vara lätt att uppdatera parametrarna utifrån ett nytt datamaterial.

Som utgångspunkt för analysen användes datamaterialet som beskrivs i avsnitt 3.2 från perioden januari till december 2012. Till en början analyseras datamaterialet för att rensa bort värden som misstänks vara mätfel. I Figur 12 ses ett utdrag ur datamaterialet. Kolumner finns bland annat för station, passagerare som stiger av, passagerare som stiger på samt hur många passagerare som är på tåget. De markerade kolumnerna är tillagda för att kontrollera om mätningen är korrekt. Kolumnernas värden beräknas på följande sätt:

$$Passtotal = PassIn + PassOut$$

$$PassOnboardBeforeStation = PassOnboard - PassOut + PassIn$$

$$BeforeStation - PassOut = PassOnboardBeforeStation - PassOut$$

Station	PassIn	PassOut	PassOnbo	PassTotal	PassOnboardBeforeStation	BeforeStation-Passout
Varde	6	0	7	6	1	1
Billum	3	4	2	7	3	-1
Boulevarden	18	8	8	26	-2	-10
Dyreby	0	0	6	0	6	6

Figur 12 - Exempel på bearbetning av rådata för passagerarantal

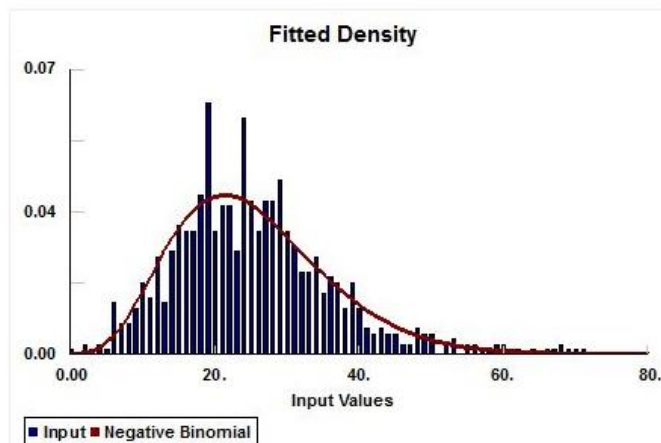
Om något av värdena i kolumnerna PassOnboardBeforeStation eller BeforeStation-PassOut är negativa innebär det att mätvärdet är orimligt och ska därför ej beaktas vid framtagning av sannolikhetsfördelningarna. Datamaterialet rensas även på mätningar

som gjorda på dagar då passagerartrycket inte anses vara representativt för en vanlig vardag, detta görs genom att bara se på mätvärden registrerade måndag till torsdag. Anledningen till att fredagar inte räknas som en vardag är att passagerarmönstret skiljer sig från övriga vardagar.

När bearbetningen av datamaterialet är slutförd tas stickprov för att undersöka vilken sannolikhetsfördelning som kan tänkas vara representativ. Stickprov tas genom att samtliga stationer förutom en filtreras bort och mätvärdena i kolumnen PassTotal studeras. För att på ett enkelt och effektivt sätt kunna analysera många stickprov används programvaran Statfit. Statfit analyserar stickprovsdata och väljer automatiskt vilka fördelningar som ska testas. Maximum Likelihood skattningar för fördelningarna beräknas och goodness-of-fit test utförs och därefter visas en inbördes ranking mellan fördelningarna. Det goodness-of-fit test som utförs är Kolmogorov-Smirnov. Resultatet av goodness-of-fit testet visas tillsammans med en graf där testade fördelningar plottas mot stickprovet från datamaterialet. I Figur 13 ses ett exempel av resultat från Statfit med data taget från stationen Viby Jylland klockan mellan 15.00 och 15.59.

Auto::Fit of Distributions

distribution	rank	acceptance
Negative Binomial[7., 0.216]	100	do not reject
Discrete Uniform[0., 71.]	0.	reject
Geometric[3.79e-002]	0.	reject
Poisson[25.4]	0.	reject



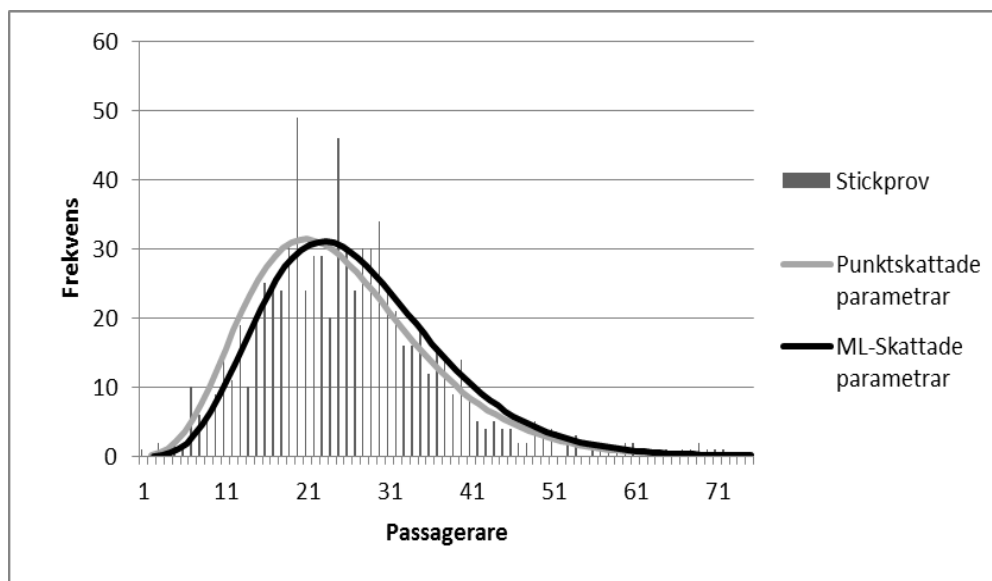
Figur 13 - Resultat från Statfit, staplarna motsvarar stickprovet och kurvan den skattade fördelningen

I Figur 13 ses att i detta fall testas Negativ binomialfördelning, Rektangulärfördelning, Geometrisk fördelning samt Poissonfördelning. Negativ binomialfördelning har rank 100 vilket innebär att det är den fördelning som ger den bästa approximationen av datamaterialet av de testade fördelningarna. För mer information om Negativ binomialfördelning se Appendix B. Notera även att det är den enda fördelningen som inte kan förkastas på 5 procents signifikansnivå. Betraktar man sedan täthetsfunktionen för Negativ binomialfördelning plottad mot datamaterialet ser man att överensstämmelsen är bra. Genom att analysera ett hundratal stickprov på detta sätt kunde Negativ binomialfördelning identifieras som den sannolikhetsfördelning som bäst används för att beskriva antalet av- och påstigande passagerare.

Goodness of fit testen förkastar ibland hypotesen att stickprovet kan approximeras med en Negativ binomialfördelning, men detta kan förklaras av att testerna kan vara känsliga för extremvärden. Om extremvärden tas bort genom att värden större än tre standardavvikelser från medelvärdet rensas bort resulterar det i

hypotesen förkastas mer sällan. Även i de fall då hypotesen förkastas av goodness of fit testerna är Negativ binomialfördelning den fördelningen som inbördes rankas högst eller näst högst och utseendet på täthetsfunktion överensstämmer bra med stickprovet. Denna analys indikerar att Negativ binomialfördelning är en lämplig standardfördelning för att generera antalet av- och påstigande passagerare givet de begränsningar vi har.

De parametrar som anges för Negativ binomialfördelning i Figur 13 är Maximum Likelihood-skattade. För att vi ska få ett system som är lätt att uppdatera utifrån ett nytt datamaterial vill vi skatta parametrar med punktskattningsmetoden som anges i Appendix B. Därför jämfördes även datamaterialet med täthetsfunktionen för Negativ binomialfördelning med punktskattade parametrar enligt avsnitt 8.1 i Appendix B. I Figur 14 återfinns resultatet med stickprovet från Viby Jylland mellan 15.00 och 15.59. I figuren kan man utläsa att täthetsfunktionen med punktskattade fördelningar är något förskjutet åt väster jämfört med fördelningen från Statfit men att den i stort sett är den samma. Detta test upprepades för stickproven och resultaten var konsekvent att skillnaden var mycket liten om parametrarna var skattade med ML-metoden i Statfit eller punktskattade.



Figur 14 - Plot med punktskattade parametrar, staplarna motsvarar stickprovet och kurvorna de skattade fördelningarna

För att använda Negativ binomialfördelningen som standardfördelningen för att generera antalet av och påstigande passagerare räknas medelvärde och standardavvikelse ut för samtliga

stationer. Medelvärde och standardavvikelsen för antal passagerare som stigit av och på, på varje station generas med en pivot-tabell som används för att räkna ut inparametrar r och p till den negativa binomialfördelningen i simuleringsmodellen enligt 8.1 i Appendix B. I ett fåtal av fallen är medelvärdet större än variansen vilket inte är tillåtet i en Negativ binomialfördelning. För att fortsatt kunna hantera det stora antalet fördelningar på ett automatiserat valdes att i dessa fall sätta varians lika med medelvärde. Detta görs eftersom författarna finner det viktigast att behålla ett korrekt medelvärde. Ändringen innebär att Negativ binomialfördelning i detta fall övergår i en Poissonfördelning. Detta hanteras rent praktiskt av det VBA-script som räknar om medelvärde och standardavvikelse för varje kombination av station, riktning och klocklag till motsvarande parametrar r och p . Resultatet blir 14 matriser med parametrar, två för varje tåglinje, som importerats till Extendsim vid varje simulering.

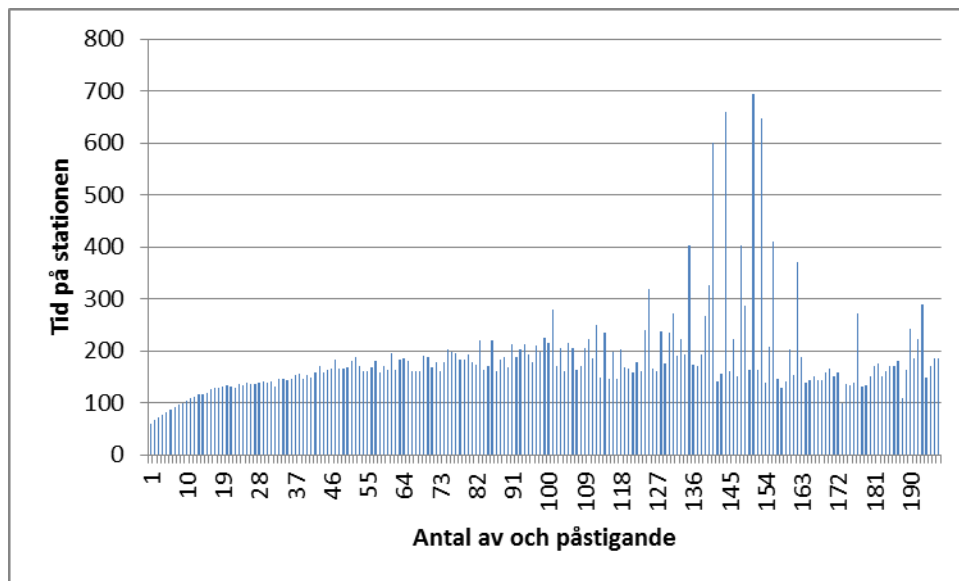
3.3.6.3 Sannolikhetsfördelningar och parametrar för stopptid

Stopptiden definieras i avsnitt 3.3.2.2 som tiden det tar att utföra de moment av stationsden som alltid utförs: acceleration, inbromsning och öppning och stängning av dörrar samt relaterade aktiviteter. Det nämns även att denna tid varierar med antalet av- och påstigande.

Passagerarantalet som genereras vid en station enligt Negativ binomialfördelning används därför som en inparameter vid beräkning av stopptiden.

I simuleringsmodellen genereras endast stopptid och inte väntetid. Väntetiden beror enligt definitionen på hur lång tid tåget måste vänta på sin avgångstid samt hur lång tid tåget väntar på att spåret framför ska bli ledigt, vilket sköts av beslutsreglerna i modellen och därför inte bör inkluderas i sannolikhetsfördelningen.

I datamaterialet från Arriva anges endast total stationstid vilket inte går att dela upp i stopptid och väntetid rakt av. Givet att ett tåg är försenat in till stationen antas det spendera minsta möjliga tid på stationen för att på så sätt försöka ta igen förseningen. Det vill säga, tåget kommer in till stationen, släpper av och på passagerare och åker vidare så fort som möjligt. Detta innebär att den komponenten av stationstiden som beror på att tågen väntar på sin avgångstid försvinner och skillnaden mellan stationstid och stopptid minskar. Ett tåg i detta tillstånd benämns som ”*stressat*”. Eftersom att sannolikhetsfördelningen ska motsvara stopptiden undersöks endast stressade tåg. Till en början plottades antalet av- och påstigande mot medelvärdet på hur lång tid tåget spenderar på stationen enligt datamaterialet. Resultatet ses i Figur 15.



Figur 15 - Plot av medelstationstid beroende på antal av- och påstigande

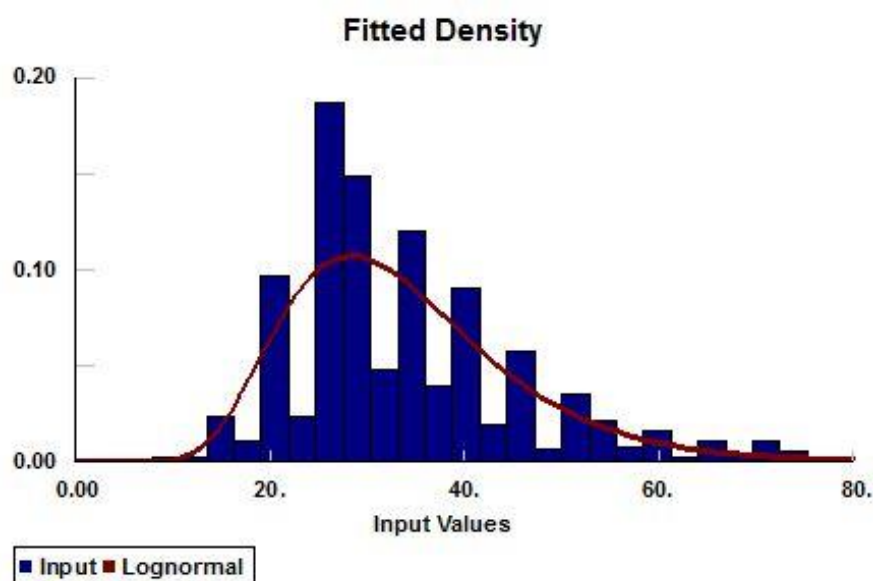
Ur detta avläses att stationstidens medelvärde är relativt konstant när antal av- och påstigande är cirka 60 eller mer. Dessutom minskar antalet mätvärden drastiskt för stigande passagerarantal vilket gör det svårt att behandla dem individuellt. På grund av detta antas att stationstiden för alla passagerarantal över 60 kan representeras av samma sannolikhetsfördelning som för 60 passagerare. Topparna som ses i Figur 15 beror på att endast enstaka mätvärden finns för dessa passagerarantal. Härefter undersöks stationstiden vidare för varje passagerarantal från 0 till 60 individuellt.

Man kan utläsa i Figur 15 att vid 0 passagerare står tåget i snitt cirka 60 sekunder på stationen vilket kan verka som en lång tid om ingen går av eller på tåget, men i detta medelvärde är all hela stationstiden inräknad, stopptid och väntetid. Eftersom fördelningen inte ska avspegla hela stationstiden utan endast stopptiden, måste data för varje passagerarantal 0 till 60 rensas individuellt så att endast komponenterna för stopptid räknas in.

Efter att ha studerat datamaterialet och diskuterat med Stefan Vidgren kom vi fram till att det är rimligt att för varje enskilt passagerarantal från 0 till 60 förkasta de mätvärden på stationstid som är högre än två medelvärden från noll. Detta görs eftersom att de antas bero på att tåget blivit stående på stationen på grund av väntetid, det vill säga av andra anledningar än antalet passagerare. Denna regel är inte baserad på någon statistisk grund utan bygger på vår och framför allt Stefans vetenskap om vad som är en rimlig tid för denna process.

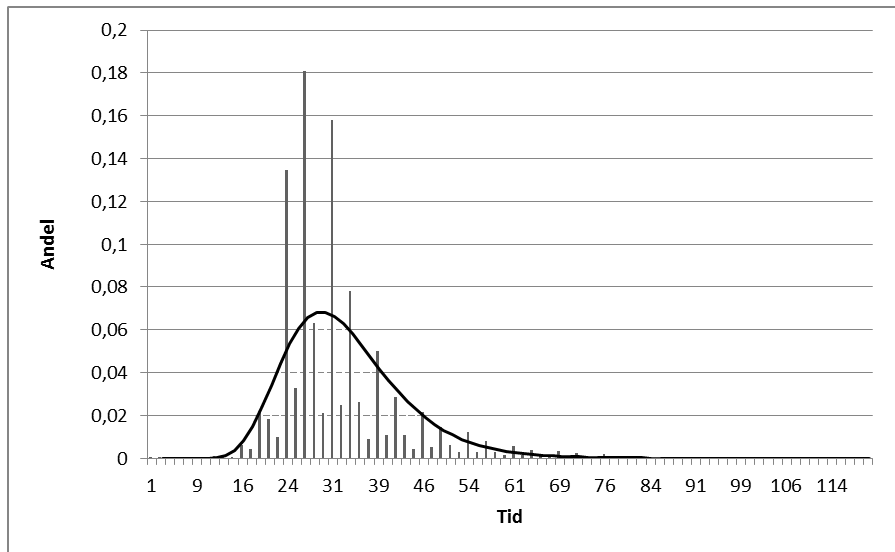
Anledningen till att inte varje kombination av station och riktning analyseras för sig som görs för av- och påstigande är att datamaterialet nu måste bearbetas på ett sätt så att en sådan analys inte skulle bli hanterbar rent tidsmässigt inom ramen för detta examensarbete samt att antalet mätvärden i många fall är för få. Försök till att segmentera stationer på olika sätt för att hitta mönster i stopptiden har även gjorts men utan framgång.

I likhet med framtagningen av fördelningar för stopptiden används nu Statfit för att avgöra vilka fördelningar som bäst överensstämmer med stickproven från datamaterialet, se exempel i Figur 16.



Figur 16 - Fördelning av stopptid i Statfit, staplarna motsvarar stickprovet och kurvan den skattade fördelningen

Återkommande för olika passagerarantal är att Lognormalfördelningen är den som rankas högst bland föreslagna fördelningar, men att hypotesen ofta förkastas i goodness of fit testerna. En genomgång av Lognormalfördelningen och dess egenskaper återfinns i avsnitt 8.3 i Appendix B. Återkommande är också att Lognormalfördelningen fångar stickprovets form men att samtliga föreslagna fördelningar har svårt att fånga de toppar som ofta uppkommer nära medelvärdet. Detta blir ännu tydligare när stickprovet plottas mot en Lognormal fördelning i Excel, se Figur 17.



Figur 17 - Fördelning av stopptid i Excel, staplarna motsvarar stickprovet och kurvan den skattade fördelningen

Trots detta anser författarna att Lognormal-fördelning ändå är den bästa möjliga standardfördelning för detta ändamål eftersom den är lätt att jobba med och övriga fördelningar som testats har samma problem med att fånga dessa ”spikar” i datamaterialet. Det ska även nämnas att bedömningen som gjordes att förkasta värden större än två medelvärden gjordes med detta problem i åtanke. En snävare rensning av datamaterialet innebär att dessa spikar bättre kan fångas eftersom en lägre standardavvikelse ger en högre topp för Lognormalfördelningen. Detta avvägande gjordes eftersom författarna anser det viktigt att fånga beteendet runt medelvärdet. Som tidigare nämnts syftar modellen till att återspegla mindre variationer i systemet och inte extremhändelser som innebär större stopp i trafiken. Parametrar till stopptid för samtliga passagerarantal från 1 till 60 beräknas enligt avsnitt 8.3 i Appendix B och importeras till en databas i Extendsim. Antalet passagerare som genereras vid en station när ett tåg kommer in bestämmer sedan vilka av dessa parametrar som ska läsas in och användas vid generering av stopptid för just detta tåg.

3.3.6.4 Sannolikhetsfördelningar och parametrar för transporttid

För att ta fram sannolikhetsfördelningar för transporttiderna valde författarna att undersöka hur mycket snabbare eller långsammare tågen kört jämfört med sin planerade transporttid utifrån datamaterialet. Genom att sortera datamaterialet på datum, tågnummer och planerad ankomst kan man avläsa ett tågs samtliga ankomst och avgångstider ordnade från startstation till slutstation. Genom att sortera mätningarna kan man jämföra avgång och ankomsttider mellan stationer och därmed räkna ut transporttiderna. I Figur 18 ses ett utdrag ur datamaterialet där ett tåg går från Tønder Nord till Dønstrup Sønderjylland.

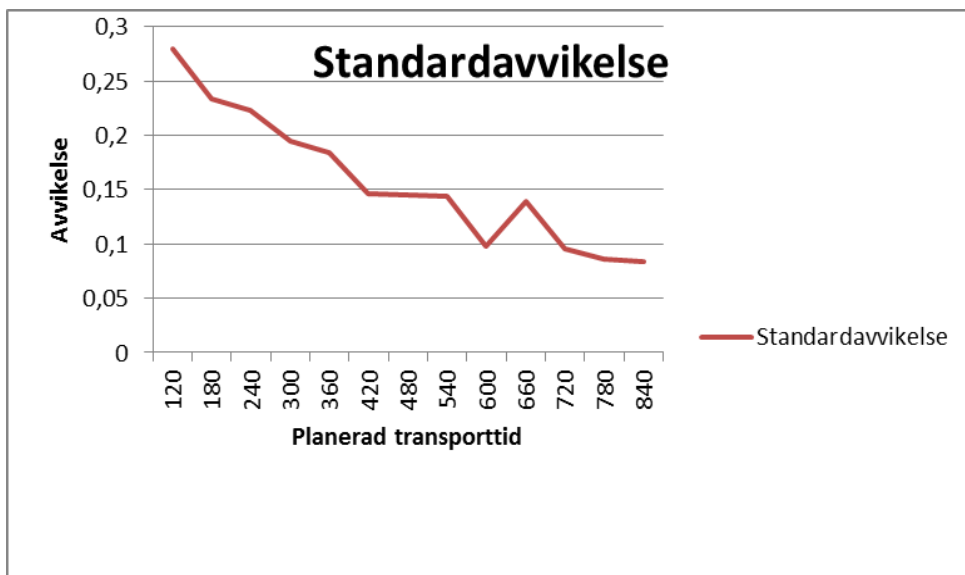
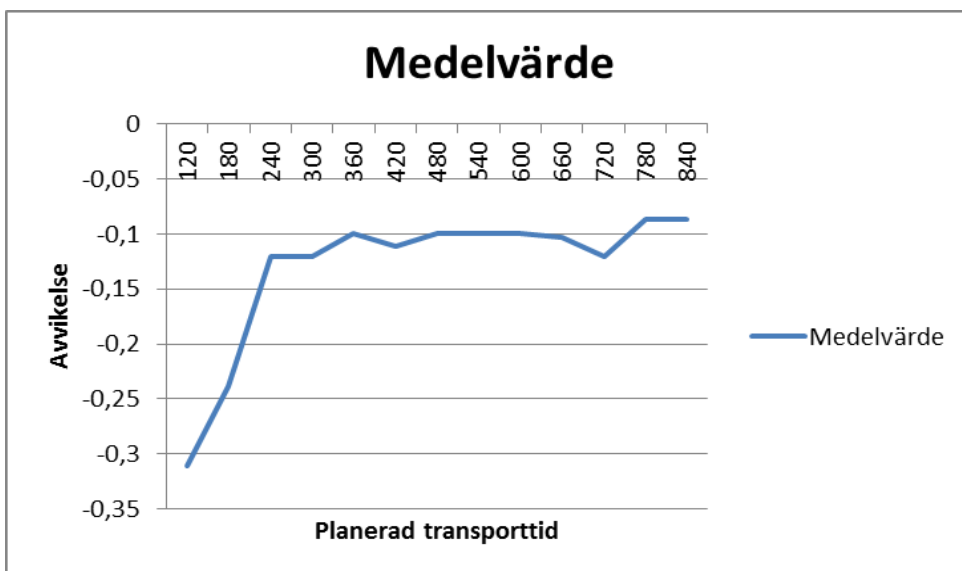
FraStationsNavn	TilStationsNavn	Plan Ank	Faktisk Ank	Plan Avg	Faktisk Avg	Actual time	Planned time	Actual - planned	Diff from planned time
Tønder Nord	Visby	08:37:00	08:40:09	08:37:30	08:41:04	420	480	-60	-0,125
Visby	Bredebro	08:45:30	08:48:04	08:46:00	08:48:49	255	300	-45	-0,15
Bredebro	Dønstrup Sønderjylland	08:51:00	08:53:04	08:51:30	08:54:04	336	420	-84	-0,2

Figur 18 - Exempel på beräkning av transporttid

På första raden i Figur 18 ser man att tåget har en planerad ankomst till Tønder Nord 08.37.00 och en planerad avgång mot Visby 08.37.30. På nästa rad kan man se att den planerade ankomsten till Visby är 08.45.30 och därmed är den planerade transporttiden 8 minuter eller 480 sekunder vilket ses i kolumnen Planned time. På samma sätt beräknas den faktiska transporttiden vilken ses i kolumnen Actual time. Jämför man dessa två tider för sträckan Tønder Nord till Dønstrup Sønderjylland ser man att tåget åkte 60 sekunder snabbare än planerad tid och delas detta tal med den planerade tiden får man hur många procent snabbare än planerad tid tåget körde, i detta fall 12,5% snabbare. Detta görs för varje transport som registrerats i datamaterialet och på så sätt kan man se hur den verkliga transporttiden avviker från den planerade transporttiden för samtliga transporter under 2012.

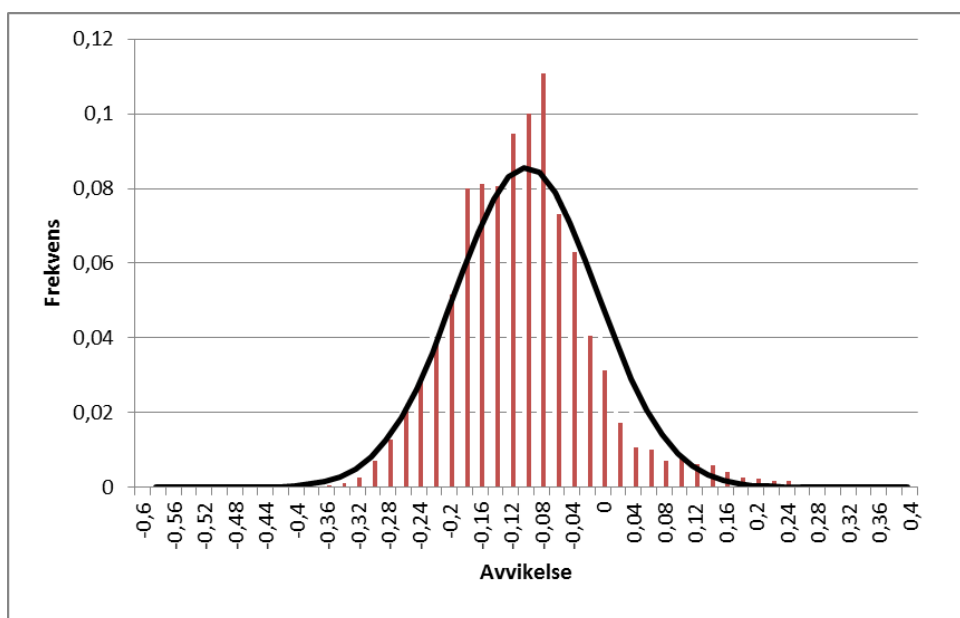
I Figur 19 ses medelvärde respektive standardavvikelse för avvikelsen mellan planerad och verklig transporttid plottad för de olika planerade transporttiderna. Till exempel kan vi se att om den planerade transporttiden är 360 sekunder, eller 6 minuter, så är avvikelsen i medeltal -10 %. Det vill säga, tåget åker sträckan på 90 % av den planerade tiden. I graferna kan vi se att medelvärdet på avvikelsen i samtliga fall är negativ. Detta är väntat och kan ses som att tågen har en bufferttid inbakad i den planerade transporttiden. Man kan även se att avvikelsen tenderar att minska för längre transporttider och samma gäller för standardavvikelsen. Eftersom det inte finns tillräckligt med mätvärden för varje planerad transporttid för att analysera var för sig aggregeras avvikelserna i intervall om två minuter. Det innebär att vi undersöker hur lång transporttiden är i verkligheten om den planerade

transporttiden är 0-2 minuter, 2-4 eller 4-6 minuter och så vidare. Notera att det inte förkommer några stora variationer i avvikelse inom intervallen förutom för 2-4 minuter, valideringen i avsnitt 4.2.2.4 visar att detta tillvägagångssätt ger ett bra resultat.



Figur 19 – Medelvärde och standardavvikelse för avvikelse från planerad transporttid

För att få bort höga mätvärden orsakade av större störningar i trafiken rensas varje intervall på värden som är mer än tre standardavvikelser från medelvärdet och därefter plottas datamaterialet för avvikelserna mot en Normalfördelning med motsvarande medelvärde och standardavvikelse. I Figur 20 ses avvikelserna för intervallet 4-6 minuter plottade med motsvarande skattad Normalfördelning. Parametrar och plottar för samtliga tidsintervall finns i Appendix C. För samtliga intervall ger Normalfördelningen en bra överensstämmelse med datamaterialet. För mer information om Normalfördelningen se avsnitt 8.4 i Appendix B. Med planerad transporttid mellan två stationer som ingångsvärde kan alltså en normalfördelad variabel slumpas som anger hur mycket snabbare eller långsammare än planerad körtid tåget ska åka mellan dessa två stationer i simuleringsmodellen.



Figur 20 - Datamaterial och skattad Normalfördelning för transporttid 4-6 minuter

Vid framtagningen av parametrar till sannolikhetsfördelningen för transporttider upptäcktes oväntade tendenser i datamaterialet. Mätningar när ett tåg åker mellan stationer som har så kallade tekniska stationer mellan sig visade sig vara uppenbart felaktiga. Exakt vad som blir fel och hur felet uppstår är inte bekräftat men förmodligen beror det på att fel tidpunkt från tidtabellen har lästs in som planerad avgångstid för stationen efter den tekniska stationen. Som exempel kan sträckan Silkeborg – Engesvang betraktas. Mellan de två stationerna finns den tekniska stationen Funder som inte visas i den

officiella tidtabellen men återfinns i den interna tidtabellen. I Figur 21 ses ett utdrag ur datamaterialet där den planerade ankomsten, klockan 08.09.00, till Engesvang är markerad. Enligt den interna tidtabellen är detta ankomsttiden för Funder men i det historiska datamaterialet har den alltså registrerats som ankomsttid för Engesvang. Den egentliga planerade ankomsttiden för Engesvang är enligt tidtabellen 08.15.00. Detta innebär att den verkliga transporttiden ser ut att vara betydligt längre än den egentligen är, i många fall nära, eller till och med över, det dubbla. Detta är ett mönster som återkommer vid samtliga sträckor som har tekniska stationer. När transporttiden beräknas kan detta problem lätt undvikas genom att de felaktiga mätvärdena rensas bort men vid validering av modellen innebär detta större svårigheter eftersom felmätningarna ger ogiltiga värden även på efterkommande stationer vid regularitetsberäkningar. Problemet diskuteras mer i kapitel Validering 4.2.

FraStationsNavn	TilStationsNavn	Plan Ank	Faktisk Ank	Plan Avg	Faktisk Avg	Actual time	Planned time	Actual - planned	Diff from planned time
Silkeborg	Engesvang	08:00:30	08:04:06	08:03:00	08:05:28	661	360	301	0,836111111
Engesvang	Bording	08:09:00	08:16:29	08:09:30	08:17:21	288	300	-12	-0,04

Figur 21 - Exempel på fel i datamaterialet

3.3.7 Modellförklaring

För att läsaren ska få en uppfattning av hur simuleringsmodellen fungerar ges här en översiktlig beskrivning av simuleringsförloppet och den stokastiska simuleringsmodellens uppbyggnad. Simuleringsmodellen är uppdelad i tre hierarkiska nivåer: Översikt, Tåglinje, Detalj. Detaljnivån är den lägsta, där i princip alla block är tagna ur ExtendSims bibliotek och kan inte brytas ner som hierarkiska block.

3.3.7.1 Översiktlig nivå

Den högsta nivån i hierarkin är den översiktliga nivån som används för att lätt kunna navigera till rätt del av simuleringsmodellen. En simulering börjar med import av indata, som i detta fall är tidtabeller, parametrar för sannolikhetsfördelningar i transportblocken och parametrar för sannolikhetsfördelningar i stationsblocken. Importen görs med "import-block" som finns samlade i blocken "Import Parameters for stations", "Import Transport time parameters" och "Import Timetables" i Figur 22. Indata lagras sedan i databaser i Extendsim för att vara lättillgängliga i simuleringen. Tåglinjerna är uppbyggda i de sju mindre blocken som återfinns nedanför blocket "Import Parameters for stations". Efter simuleringen är genomförd sker export till Excel med "export-block" som finns samlade i blocket "Export".

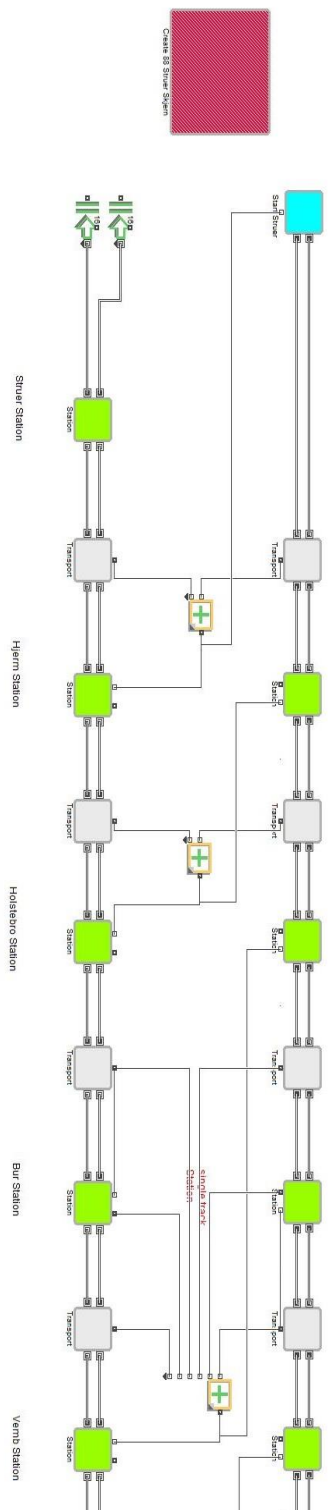


Figur 22 – Bild av den stokastiska simuleringsmodellens översiktliga nivå

3.3.7.2 Tåglinjenivå

Tåglinjenivån ger en översiktlig bild av en tåglinjes uppbyggnad utan att visa detaljerad information om vad som sker på varje station och transportsträcka.

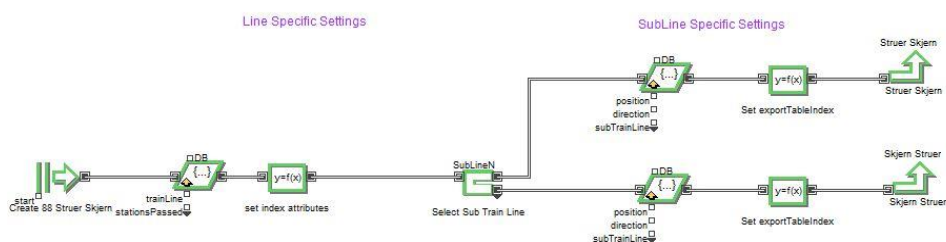
Blocken i Figur 23 med stationsnamn under sig representerar stationer och blocken med texten ”Transport” är transportsträckor. I det större blocket som återfinns längst till vänster i Figur 23 skapas tidtabellsobjekt och blocket längst upp till vänster i figuren är en startstation. För att modellera att tågen kör i båda riktningar finns två uppsättningar av varje station och transportsträcka. Till exempel om ett tåg befinner sig i det vänstra, översta transportblocket är tåget på väg från Struer till Hjerm, medan om det befinner sig på i det vänstra nedersta transportblocket är tåget på väg från Hjerm till Struer. Strukturen gör det enkelt att kontrollera att inte flera tåg befinner sig på en transportsträcka som har enkelspår. Dessa kontroller är linjerna mellan de två parallella tåglinjestrukturerna. Dessa linjer hämtar data från transportblocken till ett additionsblock, blocket med kors, och meddelar stationsblocken om sträckan är upptagen eller inte. Vid stationer med enkelspår krävs att kontroll sker av både sträcka och stationer framför. Denna mer komplicerade struktur visas i den högra delen av Figur 23.



Figur 23 - Uddrag nr fæghjenivå for strækkan Struer - Skjern

3.3.7.3 Detaljnivå

För att få en detaljerad förståelse av modellen följs ett tågs färd i detalj på den lägsta hierarkiska nivån. Först sker import av indata till databaser som sedan skickar data till "create-blocket" i Figur 23. Innehållet i "create-blocket" på detaljnivå ses i Figur 24. När indata importerats och lagrats i databaser skapas tågavgångarna enligt de importerade tidtabellerna i modellen. Varje avgång skapas som ett objekt i respektive tåglinje-block och tilldelas ett antal attribut som kan delas in i tre kategorier: attribut som dirigerar objektet till rätt plats i modellen, attribut för att läsa objektets tidtabell på rätt plats i rätt databas samt attribut som ser till att utdata skrivs i rätt databas. Detta objekt kan ses som avgångens tidtabell. När objektet tilldelats sina attribut skickas det till sin startstation, i Figur 23 återfinns startstationen Struer längst upp till vänster i bilden. I Figur 24 skapas ett objekt i blocket längst till vänster och sedan ges de tre attributen beskrivna ovan innan de skickas till aktuell startstation. Notera att alla objekt skapas här men beroende på vilken sublinje objektet tillhör skickas de härifrån till olika startstationer i simuleringsmodellen. I detta fall skickas objekten till det båda startstationerna för linjen i båda riktningar. För andra tåglinjer med flera förgrenar sig modellen med upp till 8 olika möjligheter.

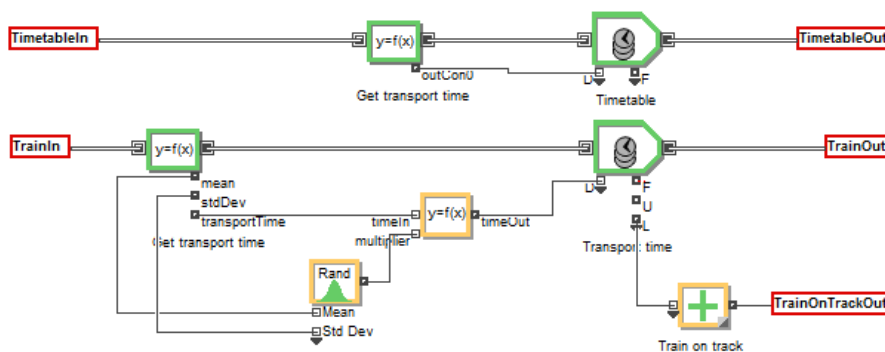


Figur 24 - Detaljerad bild av "Create-blocket"

I startstationen skapas ett nytt objekt som motsvarar själva tåget. Tåget färdas parallellt med tidtabellsobjektet i modellen, detta kan ses i Figur 23, observera att varje station och transportsträcka är kopplad med två anslutningar. Tidtabellsobjektet färdas på transportsträckor och stannar på stationer enligt de tider som finns i tidtabellen för den avgång den representerar och fungerar som en referens för objektet som motsvarar det fysiska tåget.

När det fysiska tåget lämnar sin startstation och kommer in på en transportsträcka läses dess attribut av och den planerade transporttiden hämtas i en databas. I Figur 25 visas innehållet i ett transportblock, där den övre rutten hanterar tidtabeller och den undre faktiska tåg. Utifrån den planerade transporttiden genereras en transporttid enligt de fördelningar som tagits fram under analysen av indata. Den

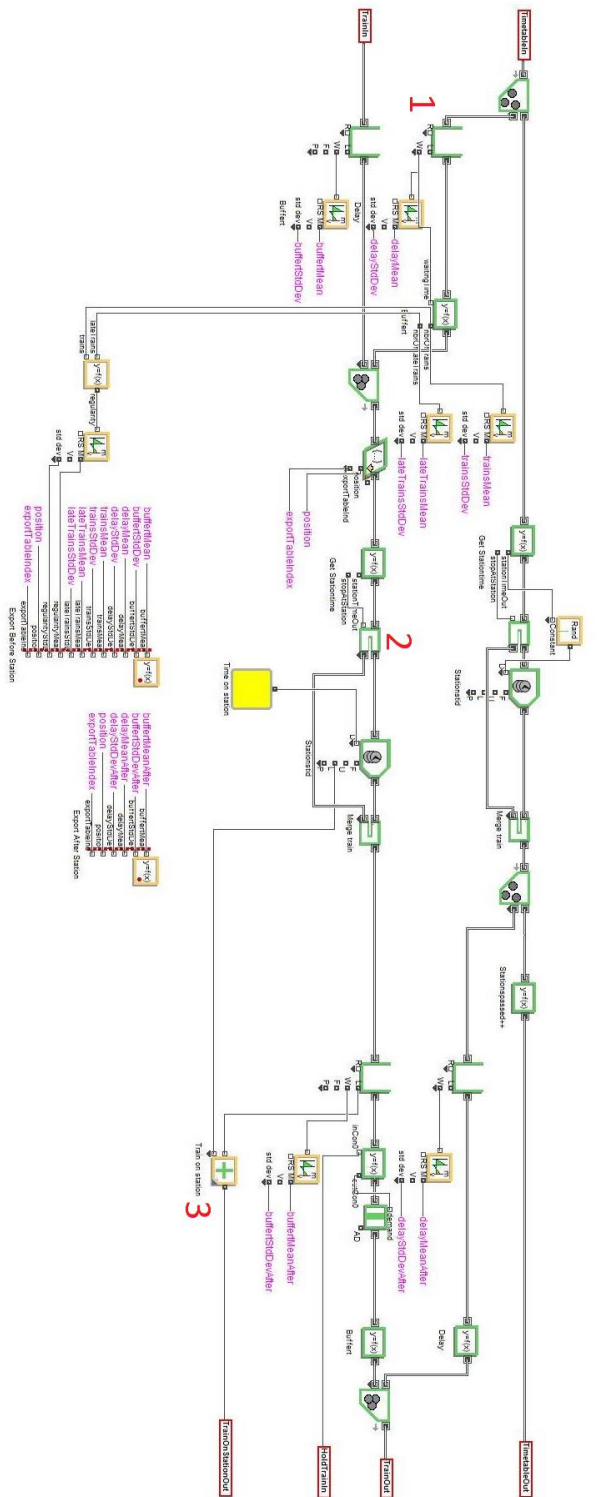
genererade tiden förs in i ett activitydelay-block, blocket benämnt ”Transport time”, som representerar den tid det tar att färdas mellan två stationer. Tidtabellsobjektet färdas på den övre rutten och läser in sin väntetid från ett liknande activitydelay-block, dock helt deterministiskt enligt tidtabell.



Figur 25 - Detaljerad bild av Transportblock

Tåget kommer nu in det första stationsblocket men innan det stannar på själva stationen sker en jämförelse mellan tidtabell och fysiskt tåg. Kontrollen sker genom att tidtabellsobjektet delar på sig i ett ”unbatch-block”, där den ena halvan fortsätter på den övre rutten och den andra går vidare till en kö, se nummer 1 i Figur 26. Om tåget kommit in före tidtabellen uppehålls det till det att tidtabellen anländer och väntetiden registreras som bufferttid. Om tidtabellen anländer före tåget registreras istället väntetiden som en försening. Är förseningen dessutom 5 minuter eller mer registreras det även som ett regularitetsbrott.

Tåget är nu redo att komma in på stationen. Innan stationen sker en kontroll om tåget ska stanna på station och därefter kör tåget antingen runt stationen eller stannar, se nummer 2 i Figur 26. För att generera stopptiden på stationen genereras först hur många passagerare som ska gå av och på tåget. Detta görs genom att en klocka läses av för att avgöra vilken tid på dygnet det är och på så sätt välja rätt parametrar från databasen till den Negativa binomialfördelning som används för att generera antalet personer enligt analysen i avsnitt 3.3.6.2. Antalet passagerare används sedan för att välja parametrar till Lognormalfördelningen som används för att generera stopptiden på stationen enligt analysen i avsnitt 3.3.6.3. Den genererade stopptiden skickas sedan till ett activitydelay-block. Genereringen sker i blocket benämnt ”Time on station” i mitten av Figur 26.



Figur 26 - Detaljerad bild av station. 1: kö för tidtabell där data om ankomstförsering registreras. 2: Uppdelning om ett objekt ska stanna på en station eller inte. 3: Kontroll för vidare färd.

När tåget lämnat det activitydelay-block som motsvarar stopptiden finns ett liknande system som det innan stationen. Där tåget jämförs med sin tidtabell för att se till att tåget inte lämnar stationen för tidigt och se till att eventuella bufferttider eller förseningar registreras. Här finns även en beslutsmekanism som säger till när tåget kan få lämna stationen, se ”3” i Figur 26. Denna beslutsmekanism läser av kommande transportblock för att se om det är något tåg på spåret och när spåren är fria släpps tåget från stationen.

Modellen är byggd så att den ska vara lätt att skala upp och modifiera, därför har alla stationsblock samma struktur med några få undantag. En tåglinje har alltid två ändstationer men kan också ha mellanliggande start- och slutstationer som bildar sublinjer. I mellanliggande stationer som också är slutstationer i en sublinje finns därför en beslutsmekanism som läser av tågets och tidtabellernas attribut för att avgöra om de ska skickas vidare till nästa station eller tas ut ur systemet. Transportblocken har alla samma struktur oavsett var i modellen de finns eftersom indata till blocket läses in endast med hjälp av attributen från de objekt som passerar. På så sätt blir modellen lätt att modifiera och förändringar i indata kräver inte nödvändigtvis förändringar av blocken i sig.

4 Verifiering och Validering av modell

Kapitlet beskriver validering och verifiering som genomförts av simuleringsmodellerna. I kapitlet belyses likheter och olikheter mellan simuleringen och datamaterialet från Arriva för att påvisa moment när modellen överensstämmer väl och där avvikelser inträffar. Kapitlet avslutas med förklaringar till avvikelserna.

4.1 Verifiering av simuleringsmodeller

Verifiering innebär en detaljkontroll av simuleringsmodellernas logiska struktur. För att säkerhetsställa att modellerna beter sig som planerat har ett antal test genomförts under modellernas skapande.

4.1.1 Kontroll av moduler

Uppbyggnaden av simuleringsmodellerna har gjorts med antal moduler, transportsträckor och stationer. Efter uppbyggnad av varje modul har logiken testats genom att kontrollera animeringen. I animeringen är det tidtabell och faktiska tåg som är möjliga att följa.

Efter uppskalningen har animeringen speciellt kontrolleras i de fall där modulerna avviker i beteende, till exempel vid dubbelspår på en sträcka eller enkelspår på en station.

4.1.2 Kontroll av tåglinjer

För att kontrollera att logiken för en linje fungerade, kontrollerades beteendet då det uppstår en konflikt i systemet. Kontrollen genomfördes genom att en konflikt skapades på ett bestämt parti under tåglinjen och sedan observerades animeringen. Konflikter kan skapas genom att öka transporttid eller stopptid.

4.1.3 Avstämning mot tredje part

Kontroll av att modulens uppbyggnad kontrolleras dels internt, det vill säga att den person som inte bygger ett avsnitt kontrollerar att det avsnittet är korrekt. Delar av modellen verifieras även från externt håll av Stefan Vidgren på Arriva som har lång erfarenhet av att bygga modeller i Extendsim.

4.1.4 Kontroll av import

Importen sker först till en ExtendSim databas och sedan hämtats vidare till blocken via kod. Vid samtliga importmoment av parametrar har kontroller utförts för att säkerställa att rätt data hamnar på rätt plats i både databas och block i modellerna.

4.2 Validering av simuleringsmodeller

För att säkerhetsställa att simuleringsmodellerna överensstämmer med verkligheten genomförs en valideringsprocess. Den huvudsakliga metoden är att jämföra verklig data för en bestämd tidtabell, med data från simuleringsmodellen för samma tidtabell. Om de överensstämmer kan modellen anses valid.

4.2.1 Deterministisk modell

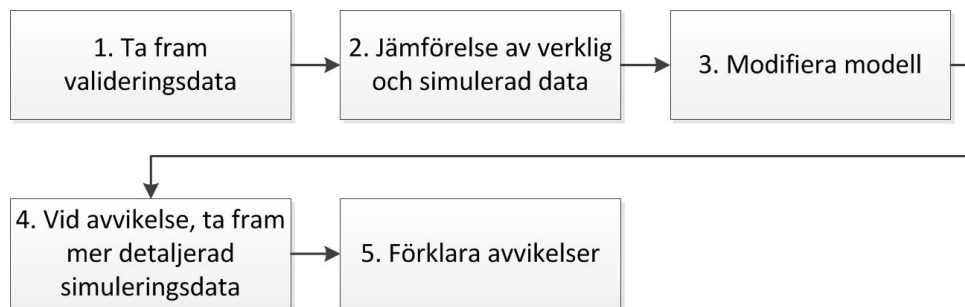
För den deterministiska modellen genomfördes simuleringar med befintlig tidtabell. Simuleringen gav inga förseningar och därmed finns inga konflikter i tidtabellen. Avgångstider och ankomsttider i modellen jämfördes också med den importerade tidtabellen för att säkerställa att modellen hanterar indata korrekt. Att tidtabellen inte innehåller några konflikter var väntat eftersom denna kontroll redan genomförts manuellt då denna tabell redan är i bruk. För att se hur simuleringsmodellen beter sig när konflikter finns i den tidtabell som testas, förs konflikter medvetet in i tidtabellen varpå den testas i simuleringsmodellen. Resultatet blir att simuleringsmodellen upptäcker konflikterna och därpå varnar användaren. Därmed antas den deterministiska modellen vara valid. För att ytterligare säkerställa validiteten av simuleringsmodellen kan fler tidtabeller testas.

4.2.2 Stokastisk modell

I den stokastiska simuleringsmodellen körs 20 simuleringar av en dag, vilket tar cirka 20 minuter att genomföra. För att säkerhetsställa att inga problem uppstår på grund av att för få simuleringar körs har längre simuleringar genomförs, 100 simuleringar av en dag. Resultatet visar inga större skillnader mot de kortare simuleringarna och därför anses antalet simuleringar vara i bra balans mellan tidsåtgång och precision.

Stefan Vidgren har kontrollerat att regularitetsnivåerna från den stokastiska simuleringsmodellen ligger inom ett rimligt intervall. Stefan Vidgren har även löpande kontrollerat logiken och indatahantering med återkoppling om förändringar och förbättringar.

För att validera den stokastiska modellen har följande process används, se Figur 27:



Figur 27 - Process för validering

De fem stegen i valideringsprocessen är beskrivna i detalj i avsnitt 4.2.2.1 - 4.2.2.5.

4.2.2.1 Steg 1 - Ta fram valideringsdata

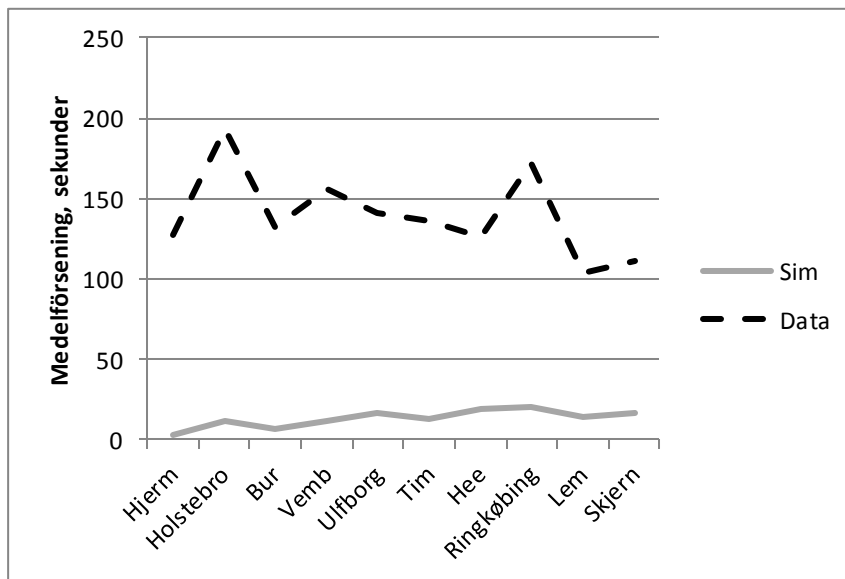
Valideringsdata består av medelförseningar och regularitet stationsvis för att kunna jämföra med utdata från simuleringen. Detta material har tagits fram ur datamaterialet från Arriva som beskrivs i avsnitt 3.2, samma datamaterial som använts för framtagning av parametrar till sannolikhetsfördelningar. Data har rensats från följande ogiltiga datapunkter:

1. Avgångar som går till ogiltiga stationer, det vill säga ankommer till stationer som inte överensstämmer med tidtabell. Exempelvis finns det flera datapunkter där start och slutstation är samma station.
2. Stationer där planerad eller faktisk stationstid är noll. Undantaget är slutstationer.
3. Datapunkter för startstationer rensas bort eftersom regularitet mäts utifrån ankomsttider och startstationer endast har avgångstider.

Valideringsdata togs fram för samtliga linjer och stationer. Då datamaterialet från Arriva innehåller många brister, framför allt kopplat till mätvärden i anslutning till tekniska stationer, har det inte varit möjligt att ta fram tillförlitlig valideringsdata till samtliga linjer. Observera också att valideringsdata avspeglar verkligheten och innehåller förseningar och regularitetsbrott orsakade av både vardagliga förseningar och större avbrott. Förseningarna från simuleringssmodellen förväntas således inte uppnå riktigt samma nivåer som valideringsdata eftersom modellen är utformad för att endast fånga vardagliga förseningar.

4.2.2.2 Steg 2 - Jämförelse verklig och simulerad data

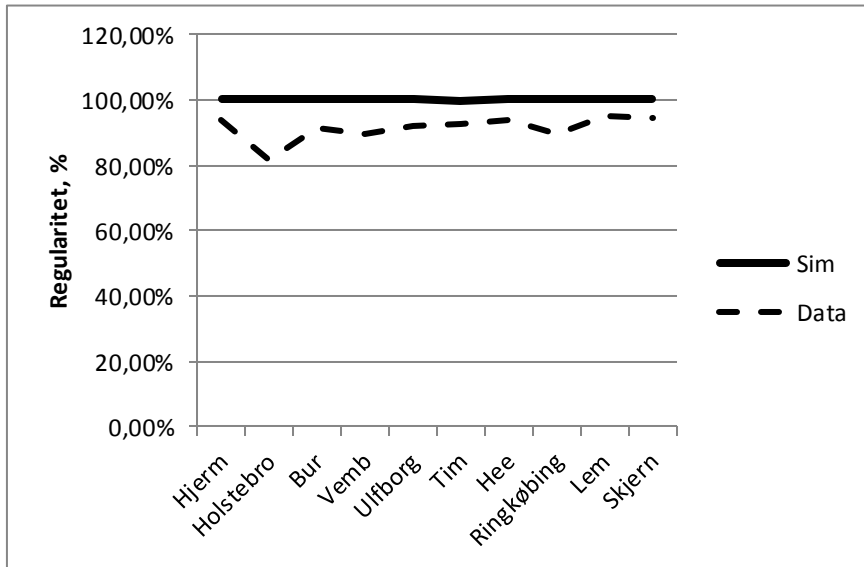
Vid jämförelse av valideringsdata och utdata från simuleringen kunde avvikelser noteras. Den största avvikelsen är att det uppkommer en stor medelförsening i början på varje linje i valideringsdata. Medelförseningen uppgår ofta till över 100 sekunder i medeltal och vilket gör att det blir en stor förskjutning jämfört med den simulerade medelförseningen, se Figur 28.



Figur 28 - Medelförsening, 88 Struer-Skjern

Den stora medelförseningen gör att regulariteten har stora avvikelser. Eftersom tågen i simuleringen antas avgå enligt tidtabell har de inte någon stor startförsening och därför händer det väldigt sällan att en försening överstiger 5 minuter. Det vill säga ett regularitetsbrott. Detta medför att regulariteten i simuleringen nästan uteslutande hamnar på 100 % medan valideringsdata på visar betydligt lägre regularitet, se Figur 29. Liknande mönster finns på samtliga linjer.

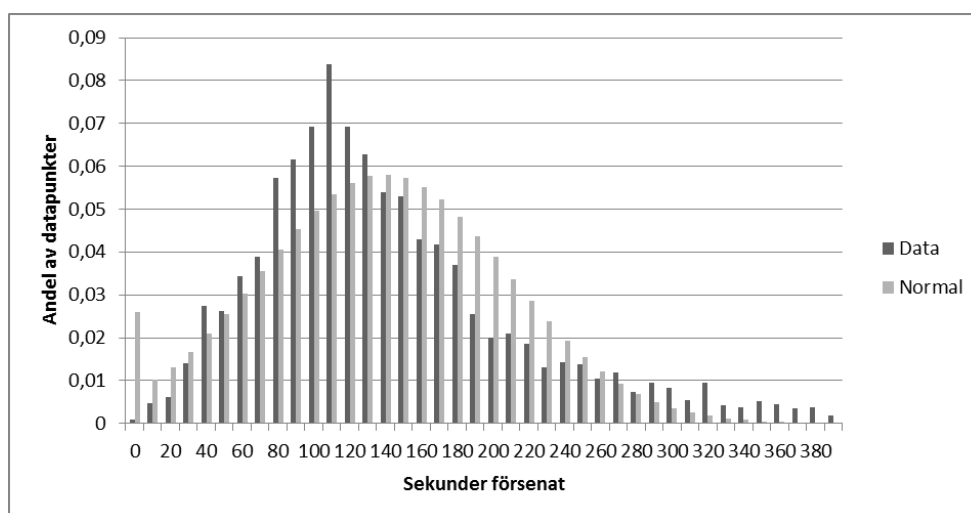
Den stora medelförseningen vid startstationer som påvisas i valideringsdata beror på att tjänsteplan och materialplan inverkar. Till exempel om en avgång är förseiad till sin slutstation och samma tåg ska användas för en annan avgång kommer den nya avgången att börja med en viss försening. Meningen är att materialplan och tjänsteplan ska läggas till modellen i framtida arbete, dock är det viktigt att göra en korrigering i modellen för att ha möjlighet att validera hur rimlig simuleringsmodellens utdata är i nuläget. Utan förändring blir skillnaden för stor att kunna göra några bra jämförelser eftersom ett tåg beter sig olika om det är förseiad eller i tid. När tåget är förseiad kommer varje buffert på station och transport konsumeras för hinna ikapp tidtabell. Dessutom hamnar tåget ur fas med tidtabellen och därför är det större sannolikhet konflikt uppstår i systemet. I nästa steg justeras modellen för att kompensera för avsaknaden av material- och tjänsteplan.



Figur 29 -Regularitet, 88 Struer - Skjern

4.2.2.3 Steg 3 - Modifiera modell

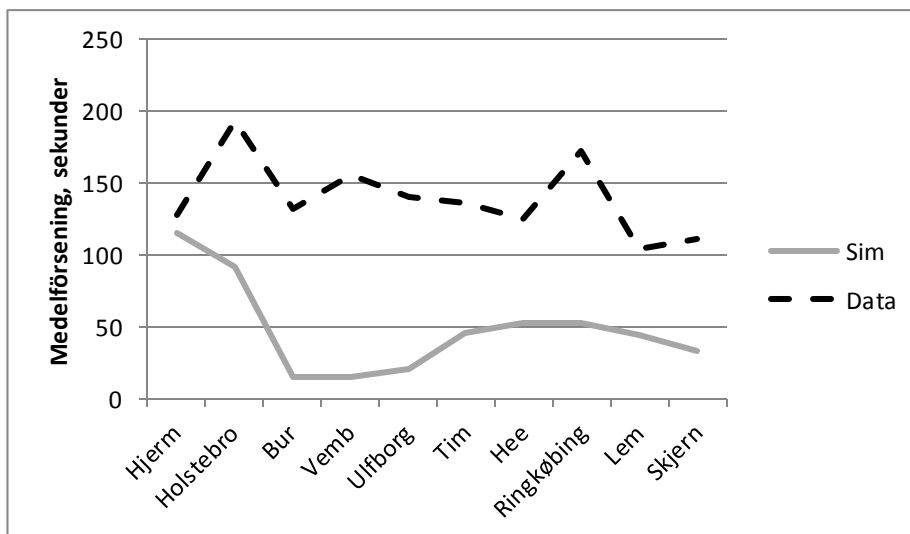
För att kunna göra en rättvisare jämförelse, läggs en startförsening till vid varje startstation i modellen. Denna försening läggs alltså till för att kompensera för konsekvenser orsakade av material- och tjänsteplan. Startförseningen är tagen ur förseningens medelvärde och standardavvikelse för respektive station. Fördelning sattes till Normalfördelning, efter ett antal stickprov genomförs, till exempel för startstationen Thisted till Sjørring, se Figur 30. Fördelningen följer stickprovet i stora drag men passar inte perfekt. En möjlighet är att göra en mer utförlig analys av startförseningens fördelning för varje station men i den begränsade tidsram som examensarbetet utgör har inte detta prioriterats.



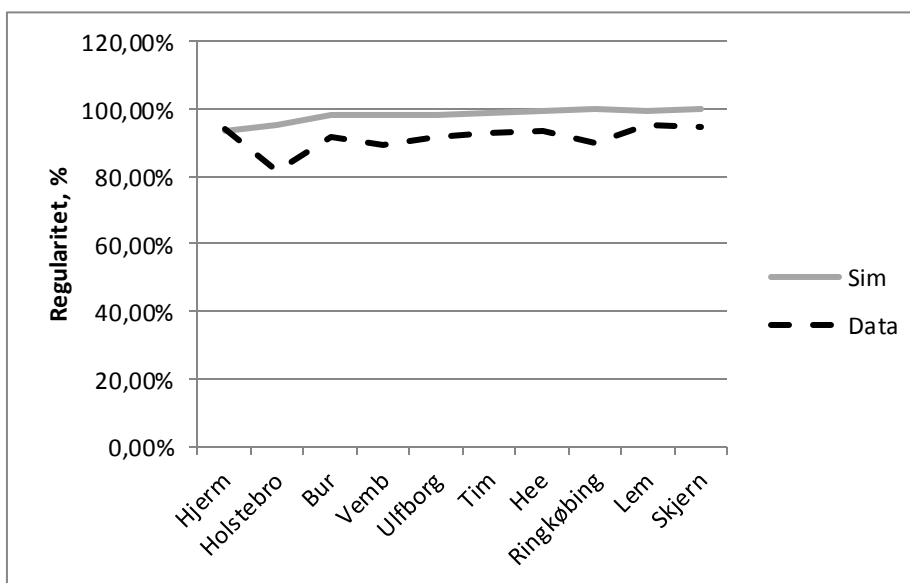
Figur 30 - Histogram av startförsening: Struer – Hjern (startstation)

4.2.2.4 Steg 4 - Jämföra med data och ta fram noggrannare data för simulering

Genom att lägga till en startförsening för att ta i beaktande avsaknaden av material och tjänsteplan kan valideringsdata och simulering jämföras rättvisare. Dock uppkommer fortfarande avvikelser i medelförsening och regularitet, se Figur 31 och Figur 32 och jämför med Figur 28 och Figur 29.



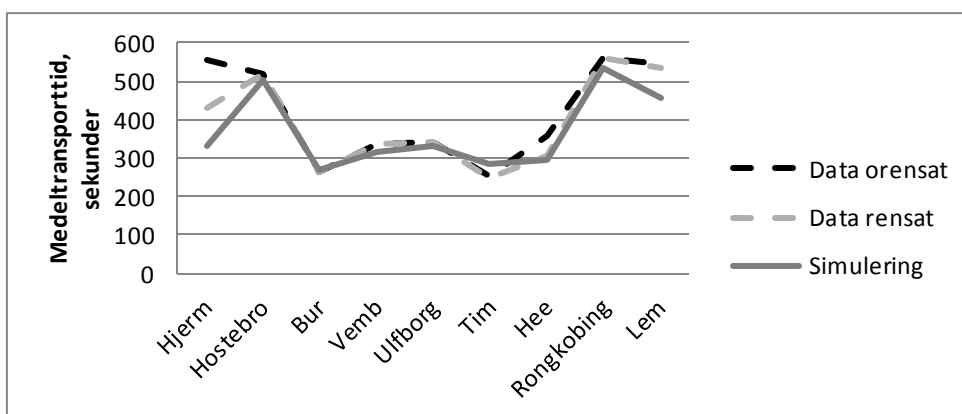
Figur 31 - Medelförsening, 88 Struer - Skjern. Startförsening tillagd



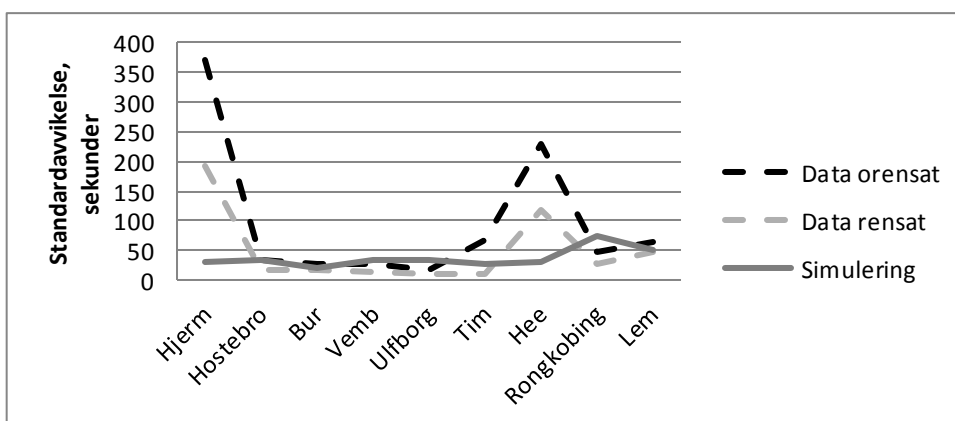
Figur 32 - Regularitet, 88 Struer – Skjern. Startförsening tillagd

För att förstå varför kvarvarande avvikelser uppstår bryts simuleringstiderna ner i sina beståndsdelar: transporttider, passagerarantal och stationstider. Detta möjliggjordes genom att ett antal nya utdatapunkter temporärt adderades till simuleringsmodellen.

De simulerade transporttiderna överensstämmer väl med de verkliga, se Figur 33 och Figur 34. Jämförelse har gjorts med dels orensad valideringsdata och rensad valideringsdata där värden som överstiger tre standardavvikelser från medelvärdet av transporttiden tagits bort. Simulering följer valideringsdata väl för medelvärde och standardavvikelse förutom vid den första stationen Hjerm. Avvikelsen vid Hjerm kan dock förklaras med att det finns en teknisk station mellan Hjerm och Holstebro som orsakar mätfel i Arrivas datamaterial.

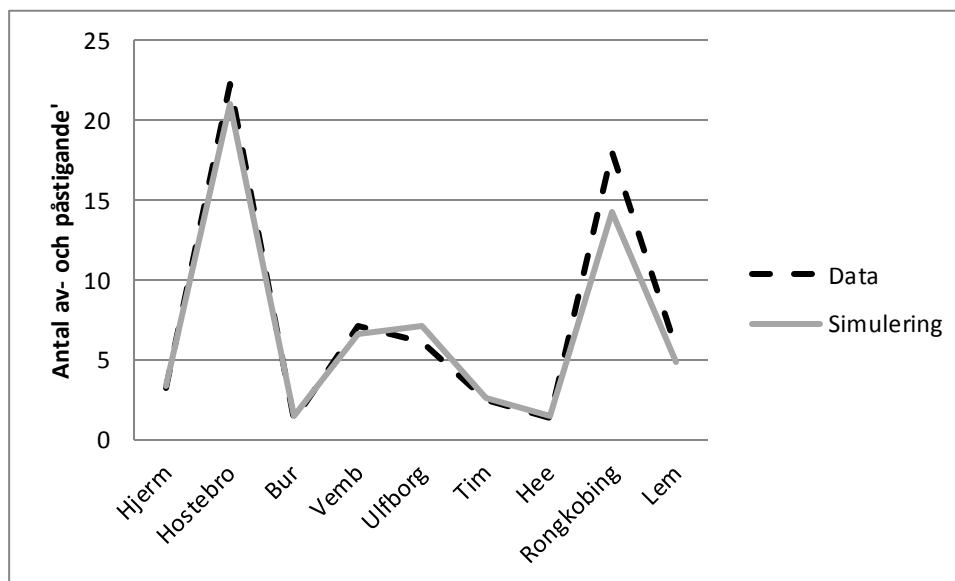


Figur 33 - Medeltransporttid för linje 88. Jämförelse mellan simulerad transporttid och transporttider från valideringsdata



Figur 34 - Standardavvikelse för linje 88. Jämförelse mellan simulerad transporttid och transporttider från data

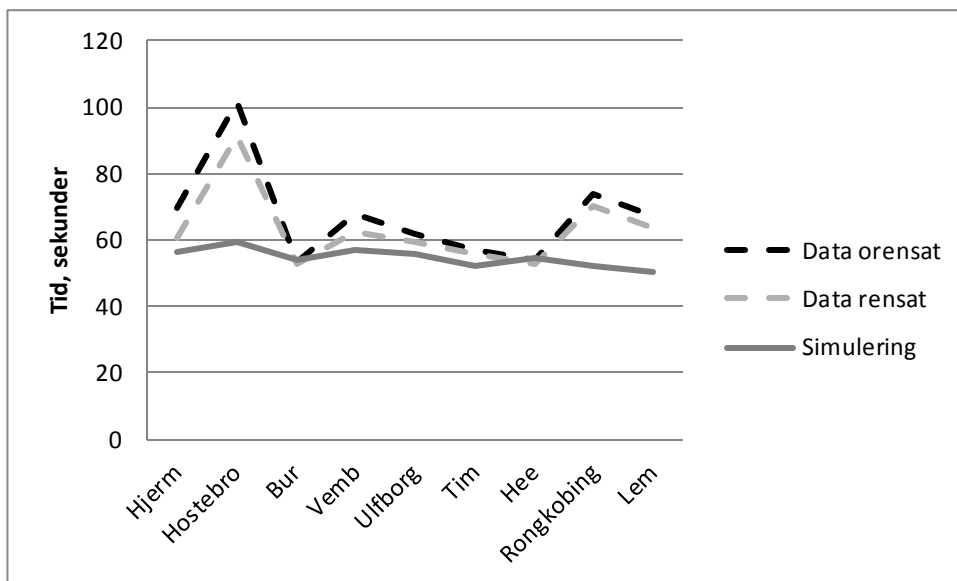
Stationstiderna består av två fördelningar: antalet passagerare och stopptiden som beror på antalet passagerare samt väntetiden som tas hand om av logiken i simuleringsmodellen. Först jämfördes antalet passagerare från simuleringen mot valideringsdata. Skattningen av antalet passagerare stämmer väl överens med datamaterialet, se Figur 35. Liknande resultat erhålls från samtliga linjer och därför anses den fördelningen ha en god validitet.



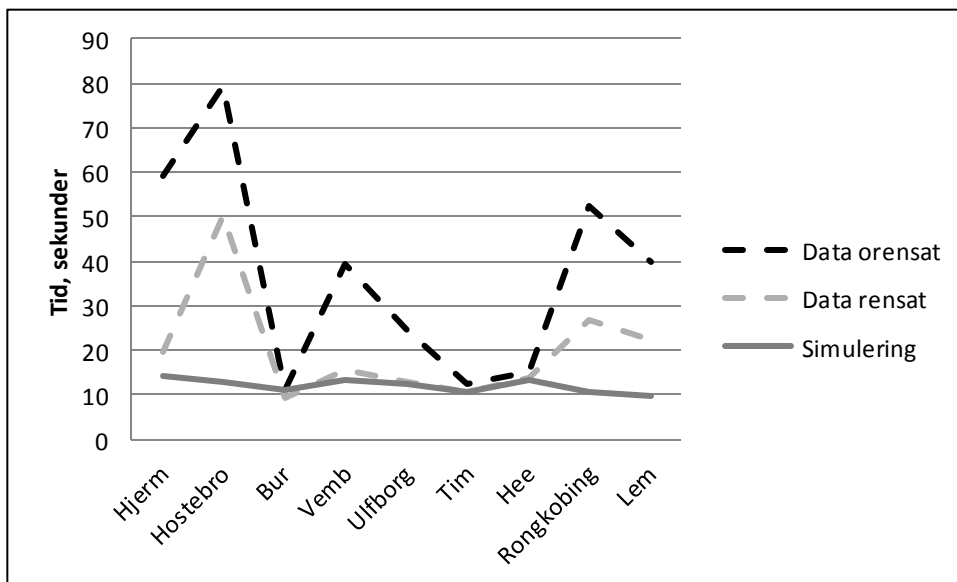
Figur 35 - Medelvärde av passagerarantal. Jämförelse data mot simulering

Sedan observerades den totala stationstiden för varje station från simuleringen och jämfördes med valideringsdata. Jämförelsen visade att det uppkom avvikelser, speciellt uppvisade valideringsdata vissa längre stationstider som inte återfinns i simuleringsresultaten, se Figur 36. Det finns två uppenbara förklaringar till detta. Första anledningen är att fördelningen för stopptid inte avspeglar hur lång tid tåget står på stationen beroende på antalet av- och påstigande. En annan möjlig anledning är att det finns andra faktorer som inte beror på passagerarantalet som påverkar, detta kan vara till exempel vilka beslutsregler som används i fall det blir en konflikt på spåret, vilket påverkar hur länge ett tåg får vänta på att få lämna stationen. Resterande delen av detta avsnitt fokuserar på analys av sannolikhetsfördelningarna för stopptid, beslutsreglernas påverkan diskuteras i avsnitt 4.2.2.5.

För att avgöra hur väl skattningen av stopptid avspeglar verkligheten har standardavvikelse plottas. Jämförelsen visar att standardavvikelsen är mycket låg i modellen och kan tyda på att rensningen av data har varit för strikt, se Figur 37. Notera att rensad data i jämförelsen är tre standardavvikelser från medelvärdet, medan indata till simuleringen är rensat på värden större än två medelvärden från 0 vilket i detta fall innebär en kraftigare rensning.

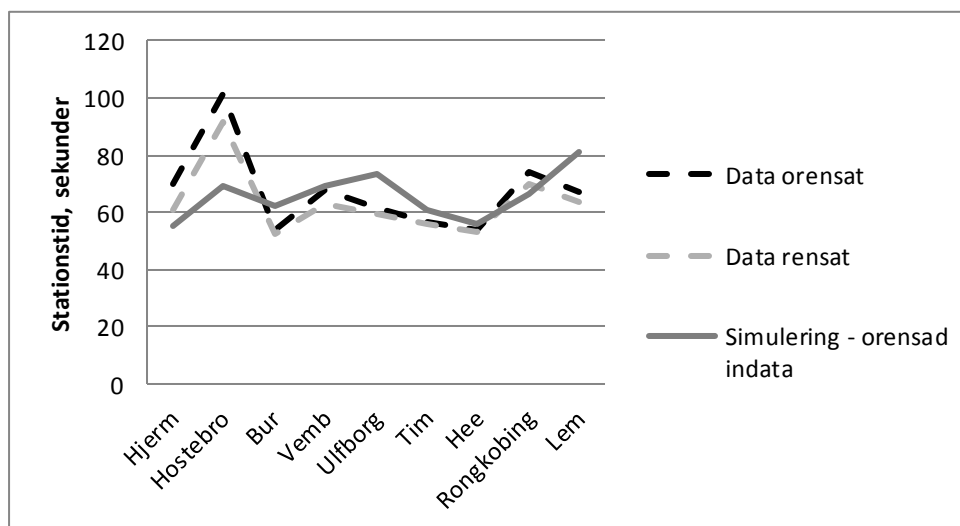


Figur 36 - Jämförelse medeltid för stationstid, 88 Struer – Skjern



Figur 37 - Jämförelse standardavvikelse stationstid, 88 Struer - Skjern

För att undersöka detta vidare gjordes nya simuleringar med helt orensad data som indata till stationstiderna i simuleringsmodellen. Jämförelsen visar att simuleringen får högre stationstider än tidigare, se Figur 38.



Figur 38 - Jämförelse medeltid på station med orensad indata till simulering, linje 88 Struer - Skjern

Simuleringens toppar når inte upp till topparna från valideringsdata vid till exempel Holstebro men i andra fall, som stationerna Bur, Ulfborg och Tim är medeltiden högre i simuleringen än i datamaterialet. Om man ser på medelvärdet över samtliga stationer på linjen hamnar dock utdata från simuleringen nära datamaterialet, se Tabell 4.

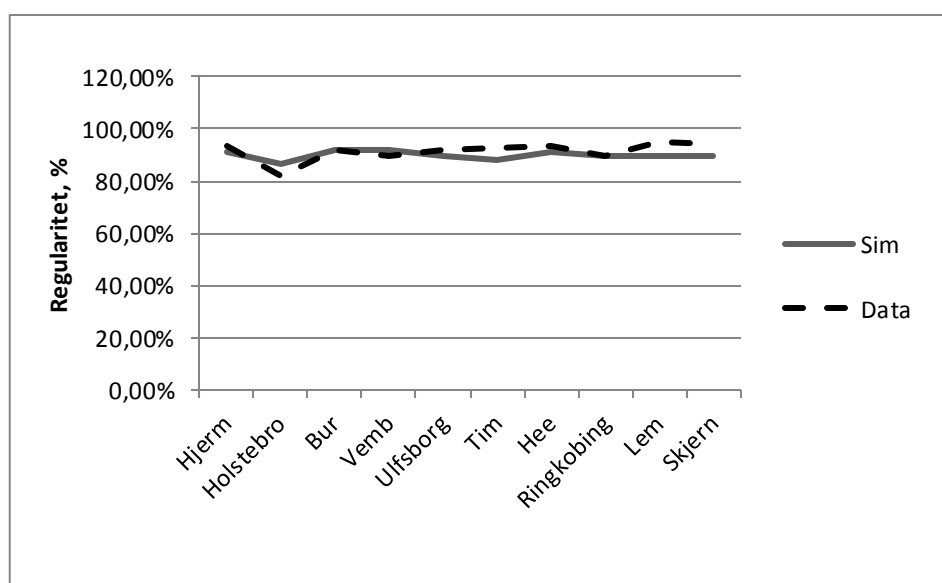
Stationstid				
	Orensad data	Rensad data	Simulering med rensad indata	Simulering orensad indata
Medelvärde	67,320615	63,235309	54,673333	65,816667
Standardavvikelse	37,041846	20,176301	12,051806	36,804652

Tabell 4 - Medelvärde och standardavvikelse för stationstid

Att stationstiden vid simulering med orensade inparametrar är högre vid somliga stationer och lägre vid andra beror troligen på att fördelningarna för stopptid inte tagits fram individuellt för varje kombination av station, sträcka och passagerarantal. Eftersom fördelningarna i stället tagits fram som utifrån mätvärden från samtliga stationer för olika antal av och påstigande mister vi dessa individuella variationer mellan stationer men medelvärdet behålls. Att

analysera stopptiden individuellt för varje station, sträcka och passagerarantal är givetvis önskvärt men har dessvärre inte varit möjligt inom tidsramen för detta examensarbete och med nuvarande datakvalité.

Efter att en startförsening lagts till i simuleringsmodellen och orensad indata för stopptider används närmar sig ses en minskad avvikelse mellan simuleringens utdata och valideringsdata för regulariteten, se Figur 39. Detta tyder på att modellen har god validitet trots vissa svagheter i framför allt stopptidernas sannolikhetsfördelningar.



Figur 39 - Regularitet, 88 Struer – Skjern. Orensade inparametrar för stationstid.

I denna sektion har en tåglinje används som exempel. Liknande beteende kan påvisas på samtliga linjer, dock är flera linjers valideringsdata av låg kvalité, vilket diskuteras mer i sektion 4.2.2.5.

4.2.2.5 Steg 5 - Förklaring av avvikelser

Ett genomgående problem i valideringen och stora delar av projektet har varit låg datakvalité. Att stora delar av datamaterialet från Arriva har visat sig vara bristfälligt påverkar både modellen och vår möjlighet att validera den. De brister som upptäckts i datamaterialet har i största möjliga mån rensats bort för att få en så bra bild av verkligheten som möjligt. Men osäkerhet råder vilka vidare konsekvenser dessa brister har för resten av datamaterialet och om den

bild av systemet som tas fram i form av valideringsdata påverkas nämnvärt.

Den allvarligaste bristen som upptäckts i datamaterialet är de felaktiga mätningarna kopplat till tekniska stationer. För komplett lista över tekniska stationer se Appendix D. Dessa mätfel gör det omöjligt att ta fram pålitlig valideringsdata för flera av de linjer där tekniska stationer återfinns i den interna tidtabellen.

Utöver detta upptäcktes även att det för en del stationer finns mycket få mätvärden i datamaterialet. Ett exempel är stationen Uphusum som har 37 datapunkter medan medelantalet för linjen ligger runt 2500 stycken. Speciellt tåglinje 84 är drabbad av detta fenomen där 6 av 18 stationer har mindre än 20 giltiga mätvärden. För komplett lista över stationer med få mätvärden, se Appendix D. Detta gör att man måste ifrågasätta om datamaterialet verkligen avspeglar verkligheten.

Den jämförelse som gjorts på linje 88 Struer – Skjern tros dock vara relativt opåverkad av fel i datamaterialet och differenserna bör därför ha andra orsaker.

Under valideringen har framför allt tre avvikelser mellan datamaterialet och utdata från simuleringen påvisats:

1. Differens i hur försenade tåg är vid start, se Figur 28
2. Stora skillnader i uppehållstid vid somliga stationer mellan datamaterialet och utdata från simuleringen, se Figur 36
3. Olika beteenden i medelförsening. I simuleringsmodellen minskar medelförsening när tåget lämnar startstation medan datamaterialet visar en ökad medelförsening på samma sträcka, se Figur 31

Avvikelse 1 beror framförallt på avsaknaden av material och tjänsteplan och modellerades genom att lägga till en försening i början av varje startstation i modellen men modellen har byggts så att en komplett material- och tjänsteplan ska kunna adderas vid senare tillfälle. Denna modifiering ger en bättre överensstämmelse mellan simuleringsmodellens resultat och det verkliga datamaterialet.

Avvikelse 2 beror i huvudsak på hur parametrarna för stopptid tagits fram. En förbättring ses om inparametrarna inte rensas lika hårt som tidigare gjorts. Dock bör denna indataanalys göras på nytt när bättre data finns tillgänglig. Den andra anledningen till de höga stationstiderna är att det sker en kraftig försening vid knytpunkter. Detta beror på att vår modell inte tar i beaktande materialplan eller tjänsteplan, vilket till exempel innebär att tåg inte måste vänta på

andra tåg vid knytpunkter. Avvikelse 2 är även sammankopplad med avvikelse 3 eftersom förseningar innebär ökade uppehållstider på stationer.

Avvikelse 3 kan förklaras med att beslutsreglerna ser olika ut i modell och verklighet. För simuleringsmodellen, med en sjunkande medelförsening, är en förklaring att eftersom en startförsening införs i början på varje linje kommer tåget att försöka konsumera alla buffertar på stationer och transportsträckor till förseningen är eliminerad. Beteendet tyder på att modellen agerar korrekt efter uppsatta antaganden. Problemet är att samma beteende borde ses i datamaterialet för förseningar i systemet. Istället ser vi där en ökning i medelförsening från startstation, se Figur 31. För att hitta anledningen till de olika beteendena analyserades simuleringsmodellens antaganden. Analysen visar att den troligaste anledningen till avvikelserna är skillnaden i hantering av beslutsregler. I simuleringsmodellen används beslutsregeln att det tåg som är redo att köra ut på en viss transportsträcka först får köra in på transportsträckan och andra tåg får vänta tills sträckan är ledig. Enligt Stefan Vidgren används ett flertal olika beslutsregler i verkligheten och de varierar efter vem som jobbar den dagen på trafikledningen. Vanligast är att det tåg som är mest rättidigt prioriteras och sena tåg får vänta. Med rättidigt menas det tåg som är minst försenat. Genom denna prioritering minimeras antalet tåg med regularitetsfel medan medelförseningen och stationstiden borde bli högre i genomsnitt än för beslutsreglerna i simuleringsmodellen.

Anledningen till varför beslutsreglerna i simuleringsmodellen borde ge lägre genomsnittsförsening är att resursförbrukningen av transportsträckor optimeras, det vill säga ett tåg som har möjlighet att köra in på en ledig sträcka kommer göra det. Med beslutsregler för prioritering av rättidighet kan ett tåg bli stillastående på en station eftersom det måste vänta på ett mötande tåg som är mer rättidigt i den motsatta riktningen.

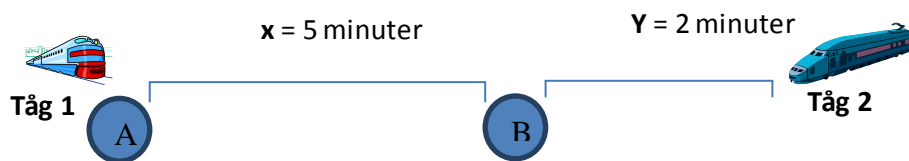
Som exempel antar vi att finns två stationer A och B med en mellanliggande transporttid på 5 minuter. Tåg 1 står på station A och är 2 minuter sent, men är redo att köra in på transportsträckan mot station B, se det vänstra tåget i Figur 40. Tåg 2, se högra tåget i Figur 40, är i tid och på väg till station B med 2 minuter transporttid kvar. För enkelhetens skull antar vi att stationstiden på station B är noll. Beroende på vilka beslutsregler som används blir utfallet av denna situation olika.

Används simuleringsmodellens beslutsregler åker tåg 1 mot station B och behåller sin försening på 2 minuter. När tåg 2 ankommer till

station B efter 2 minuter måste den dock vänta 3 minuter på att få köra vidare mot station A eftersom spåret nu är upptaget. Konsekvensen blir att tåg 2 blir försenat med 3 minuter. Utfallet blir alltså två försenade tåg med en total försening på 5 minuter.

Om beslutsregler där rättidiga tåg prioriteras används sker följande. Tåg 1 är redo att köra mot station B men eftersom tåg 2 är i tid och snart redo att köra in på samma transportsträcka måste tåg 1 vänta. Konsekvensen blir att tåg 1 blir ytterligare 7 minuter försenat men tåg 2 är fortfarande i tid. Utfallet blir alltså endast ett försenat tåg men en total försening på 9 minuter.

I det generella fallet där körsträckan mellan station A och B är x minuter och tiden till ankomst vid station B för det rättidiga tåget är y minuter blir väntetiden för tåg 2 om simuleringsmodellens beslutsregler används, $x-y$ vid konflikt i systemet. Konflikt uppkommer om $x > y$. För beslutsregler där rättidiga tåg prioriteras blir väntetiden $x + y$, det vill säga alltid längre.



Figur 40 – Exempel på konflikt mellan tåg

En möjlig situation där simuleringsbeslutsmekanismen skulle få en högre medelförsening är om ett långsamt tåg skulle köra först och det fanns snabbare tåg som körde ifatt och fick vänta. Denna situation skulle ge väldigt långa väntetider, men är inte trolig eftersom alla tåg betar sig snarligt på transportsträckorna och i kappkörningar är orealistiska. Det vill säga det är mycket troligt att beslutsprocessen som används i verkligheten i regel ger en större medelförsening och stationstid.

Det är därför rimligt att anta att annorlunda beslutsregler bidrar till avvikelse 2 och 3, dock är det svårt att säga till vilken grad. Ett möjligt test för att kontrollera hur stor påverkan beslutsreglerna har är att modifiera simuleringsmodellen efter att prioritera tåg utifrån rättidighet. Beslutsregeln har dock inte implementerats på grund av tidsbrist. Vid framtida arbete bör olika beslutsregler diskuteras.

4.2.2.6 *Resultat av validering*

På grund av avsaknaden av materialplan och tjänsteplan förväntades inte att simuleringsmodellen skulle ge ett resultat som överensstämde fullt ut med verklig data utan att modellen modifierades. Dock har vi kunnat påvisa att modellen ger rimliga resultat i form av transporttid och antal av och påstigande. Vi har även påvisat att modellen ger rimliga slutresultat i form av regularitet då startförseningar temporärt läggs till och inparametrar för stopptid justeras. Detta tyder på att modellen med dess givna avgränsningar är valid. För att fullt ut beskriva tågnätets dynamik behöver dock den nuvarande modellen kompletteras med materialplan och tjänsteplan samt nya inparametrar baserad på bättre datamaterial.

5 Slutsats och rekommendation

Slutsatser och rekommendationer av arbetet presenteras tillsammans med hur väl syftet för projektet är uppfyllt. Därefter ges förslag på framtida studier.

5.1 Slutsats

Författarna har genom olika valideringsmetoder kommit fram till att den deterministiska modellen är väl fungerande och därmed redo att användas för att utvärdera nya tidtabeller. Modellen är generell och kan modifieras för att även innefatta material och tjänsteplan genom att lägga till fler parallella flöden i simuleringsmodellen.

Den stokastiska modellens validitet har varit svårare att säkerställa på grund av dess beroende av material, tjänsteplan och kvalitet på indata. Validiteten går inte att utreda till fullo innan modellen är komplett men genom att bedöma validiteten i modellens nuvarande form har vi kommit fram till att modellen är väl fungerande rent logiskt samt att sannolikhetsfördelningar för antal av och påstigande och transporttider är tillfredställande men att sannolikhetsfördelningar för stopptid bör utredas vidare. Resultaten har diskuterats med Stefan Vidgren och bedöms vara rimliga.

Genom att ha arbetat med datamaterial från Arriva har vi upptäckt brister i hur data registreras. Detta är något som påverkar möjligheterna till att ta fram en bra simuleringsmodell eftersom det är detta datamaterial som ligger till grund för skattningarna av de sannolikhetsfördelningar som används i simuleringsmodellen. Det är viktigt att detta utreds eftersom bra indata är en förutsättning för ett givande analysarbete. Problemet har påpekats för Arriva och arbetet med att förbättra kvalitén på datamaterialet har inletts av Stefan Vidgren och kommer fortgå under det närmsta året.

Vi anser att den största osäkerheten med simuleringsmodellen är datamaterialet som den bygger på. Även om en simuleringsmodell lyckas avspegla datamaterialet bra är det inte mycket värt om datamaterialet i grunden inte avspeglar verkligheten. Denna brist är något vi inte kunnat påverka, men genom att undvika att använda data

som är bristfällig tror vi att modellen ändå överensstämmer relativt bra med verkligheten.

Stor vikt har lagts vid att modellen ska vara praktisk och lätt att uppdatera när ett nytt datamaterial blir tillgängligt. Det är meningen att modellen ska användas löpande och det är därför viktigt att den bygger på aktuell data och att den är lätt att använda. En konsekvens av detta är att vi har använt oss av standardfördelningar vid skattning av stokastiska fördelningar i modellen. Det hade förstås varit fördelaktigt att behandla varje fördelning individuellt men det hade inneburit stor tidsåtgång och uppdatering av modellen hade blivit osmidig och långsam. Vi bedömer att standardfördelningar kan användas och fortfarande ge tillfredställande resultat vilket styrks i valideringen av antal av och påstigande samt transporttider. Undantaget är fördelningarna för stopptid där avvikelser finns mellan simuleringsresultaten och. Detta beror dock inte på att det är standardfördelningar som används, snarare beror det på den segmentering vi varit tvungna att göra på grund av tidsåtgång och kvalitet på data. Med en bättre struktur och kvalitet på data från tågsystemet är det möjligt att behandla samtliga stationer individuellt, vilket bör ge en förbättring i modelleringen av stopptider.

Syftet med projektet har varit att ta fram två simuleringsmodeller för att underlätta planläggningen på Arriva. Detta anses vara uppfyllt. Simuleringsmodellerna har tre huvudsakliga användningsområden:

1. Identifiera om en tidtabell innehåller konflikter och deterministiskt precisera var konflikten uppstår.
2. Jämföra tidtabeller. Möjlighet till att jämföra olika tabellen för att åskådliggöra hur förändringar ger effekter på det större systemet.
3. Identifiera möjliga framtida problem.

Punkt 1 uppfylls av den deterministiska modellen fullt ut eftersom vi kommit fram till att den är fullt valid. Punkt 2, kan genomföras trots att den stokastiska simuleringsmodellen har brister. Det huvudsakliga användningsområdet för den stokastiska modellen är att jämföra tidtabeller med varandra. Eftersom båda tidtabellerna jämförs utifrån samma förutsättningar kan man få en uppfattning om vilken som är lämpligast även om modellen inte ännu är helt komplett. Punkt 3 är svårare att genomföra i nuläget eftersom den kräver en mer komplett modell som innehåller material- och tjänsteplan och kommer som planerat att överlämnas till Arriva för framtida analys.

5.2 Rekommendation

Vårt förslag till Arriva är att de implementerar den deterministiska simuleringsmodellen och att de fortsätter arbeta med den stokastiska simuleringsmodellen enligt följande schema:

1. Utred och förbättra processen för datainsamling
2. Uppdatera samtliga parametrar utifrån det nya datamaterialet
3. Validera modellen mot det nya datamaterialet
4. Modifiera modellen med beslutsregler utifrån rättidighet och jämför validitet med föregående modell
5. Komplettera modellen med material och tjänsteplan genom att lägga in parallella spår för material och personal

6 Litteraturförteckning

6.1 Böcker

- Axsäter, S., 2007. *Inventory Control*. 2nd Edition. Boston: Springer Science+Business Media, LLC.
- Banks, J., 1998. *Handbook of Simulation: Principles, methodology, advances, applications and practice*. New York: Wiley.
- Everitt, B. & Skrondal, A., 2010. *The Cambridge Dictionary of Statistics*. 4th Edition. Cambridge: Cambridge University Press.
- Feller, W., 1968. *An Introduction to probability Theory and Its Applications*. 3rd Edition. New York: Wiley.
- Hillier, F. & Lieberman, G., 2010. *Introduction to Operation Research*. 9th Edition. Boston: McGraw-Hill.
- Hopp, W. & Spearman, M., 2001. *Factory Physics*. 2nd Edition. New York: Irwin/McGraw-Hill.
- Laguna, M. & Marklund, J., 2013. *Business process modeling, simulation and design*. 2nd Edition. Boca Raton: CRC Press Taylor & Francis Group.

6.2 Artiklar

- Aristizabal, R., 2012. Estimating the Parameters of the Three-Parameter Lognormal Distribution. *FIU Electronic Theses and Dissertations*, Issue <http://digitalcommons.fiu.edu/etd/575/>.
- Bergmark, R., 1996. Railroad capacity and traffic analysis using SIMON. i: e. a. B. Mellitt, red. *Computers in Railways V*. Southampton: WIT Press, p. 183–191.
- Cacchiani, V. & Toth, P., 2011. Nominal and robust train timetabling problems. *European Journal of Operational Research*, 219(3), p. 727–737.
- Huisman, D., Kroon, L. G., Lentink, R. M. & Vromans, M. J., 2005. Operations Research in Passenger Railway Transportation. *Statistica Neerlandica*, 59(4), pp. 467–497.
- Kaas, A., 2000. Punctuality model for railways. i: e. a. J. Allen, red. *Computers in Railways VII*. Southampton: WIT Press, p. 809–816.

Kroon, L. o.a., 2008. Stochastic improvement of cyclic railway timetables. *Transportation Research Part B: Methodological*, 42(6), p. 553–570.

Middelkoop, D. & Bouwman, M., 2000. Train network simulator for support of network wide planning of infrastructure and timetables. i: e. a. J. Allen, red. *Computers in Railways VII*. Southampton: WIT Press, p. 267–276.

6.3 Internet

Arriva, 2010a. *Mit Arriva Danmark*. [Online]
Available at: <http://www.mitarriiva.dk/fakta-om-arriva-tog>
[Använd 11 02 2014].

Arriva, 2010b. *Mit Arriva*. [Online]
Available at: <http://www.mitarriiva.dk/stationer>
[Använd 11 02 2014].

Arriva, 2010c. *Mit Arriva*. [Online]
Available at: <http://www.mitarriiva.dk/straekningsoversigt>
[Använd 11 02 2014].

Arriva, 2011. *Arriva-Corporate-Brochure-2011*. [Online]
Available at: <http://www.arriva.co.uk/~media/Files/A/Arriva-V2/en/pdf/docs/Arriva-Corporate-Brochure-2011.pdf>
[Använd 11 02 2014].

Arriva, 2013. *Arriva*. [Online]
Available at: <http://www.arriva.dk/index.php/39-punktliighed2013>
[Använd 11 02 2014].

Arriva, 2014. *Arriva Sverige*. [Online]
Available at: <http://arriva.se/sv/content/arrivagruppen-fran-liten-till-stor>
[Använd 11 2 2014].

Imagine That Inc, 2014. *Imagine That Inc*. [Online]
Available at: <http://www.extendsim.com/>
[Använd 03 Juni 2014].

7 Appendix A

Detaljerad översikt av Arrivas tåglinjer.



■ Sträckningar hvor Arriva kører



8 Appendix B

Appendix B beskriver de stokastiska fördelningar som används inom projektet. Kapitlet ämnar ge läsaren en snabbare referens för att kontrollera fördelningarnas egenskaper.

8.1 Negativ binomialfördelning

Negativ binomialfördelning används vårt arbete för antal passagerare som förväntas stiga av och på vid varje station. Fördelning består av sekventiella Bernoulli-försök där varje försök att lyckats har sannolikhet p , och sannolikhet för misslyckande är $q = 1 - p$. Om D är en stokastisk variabel som är negativt binomialfördelad med parametrar r och p gäller:

$$P(D = k) = \frac{r(r+1) \dots (r+k-1)}{k!} p^k q^r \quad k = 1, 2, \dots \quad 0 < p < 1$$

$$\text{medelvärde} = rp/q$$

$$\text{varians} = rp/q^2$$

(Axsäter, 2007, p. 81)

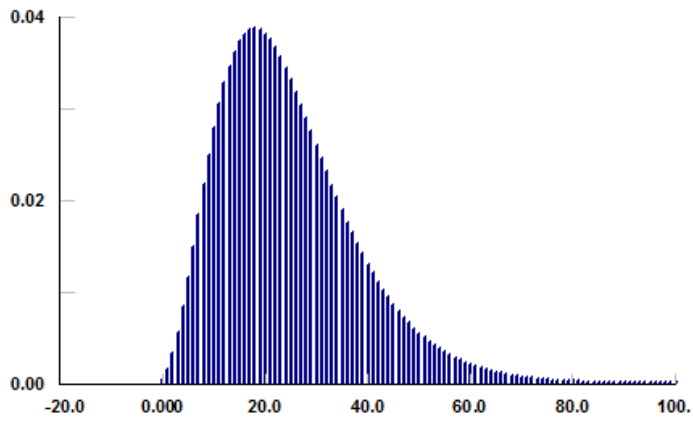
Medelvärde (μ) och standardavvikelse (σ) beräknas utifrån datamaterialet som erhållits från Arriva. Parametrarna kan då beräknas enligt:

$$p = 1 - \frac{\mu}{\sigma^2}$$

$$r = \mu \frac{(1-p)}{p}$$

(Axsäter, 2007, p. 82)

Notera att σ^2 måste vara större än μ annars går p mot noll eller negativa värden, vilket inte är möjligt. Detta är ett specialfall då fördelningen får samma egenskaper som en Poissonfördelning (Feller, 1968, p. 281). I Figur 41 återfinns ett exempel på en täthetsfunktion för en negativt binomialfördelad stokastisk variabel.



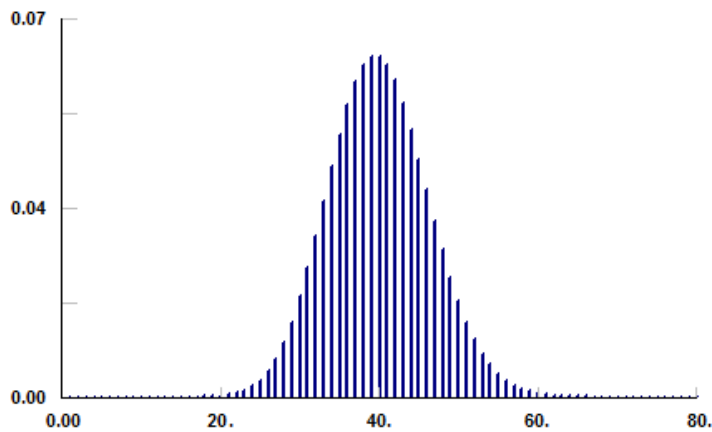
Figur 41 – T hetskfunktion, negativ binomialf rdelning

8.2 Poissonfördelning

Poissonfördelning används i specialfallet när medelvärdet är större än variansen för antal passagerare som förväntas stiga av och på vid varje station. Fördelningen sannolikhetsfunktion är given enligt nedan:

$$p(k) = \frac{e^{-\mu} \mu^k}{k!} \quad k = 0, 1, 2, \dots$$

Poissonfördelningen är en ”en variabel”-fördelning, där medelvärde och varians antar samma värde, μ . Vid uträkning av parametrar används det observerade medelvärdet som inparameter. (Hopp & Spearman, 2001, p. 95) I Figur 42 ses ett exempel på en täthetsfunktion för en poissonfördelad stokastisk variabel.



Figur 42 – Täthetsfunktion, poissonfördelning

8.3 Lognormalfördelning

Vi använder lognormalfördelning för att beskriva tiden som tåget stannar på stationen. Fördelning har en sannolikhetsfunktion som beror på Normalfördelningens medelvärde (μ) och varians (σ^2). Fördelningen är enligt nedan:

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{(-\frac{1}{2\sigma^2}(\ln x - \mu)^2)} \quad 0 < x$$

$$\text{medelvärde } (m) = e^{(\mu + \frac{1}{2}\sigma^2)}$$

$$\text{varians } (v) = (e^{\sigma^2} - 1)e^{(2\mu + \sigma^2)}$$

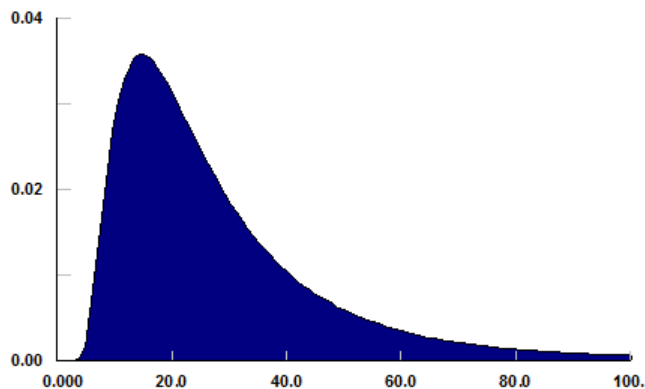
(Everitt & Skron dal, 2010, p. 259)

Parametrarna till lognormalfördelningen beräknas enligt följande:

$$\mu = \ln\left(\frac{m^2}{\sqrt{v + m^2}}\right) \quad \sigma = \sqrt{\ln\left(1 + \frac{v}{m^2}\right)}$$

(Aristizabal, 2012)

Ett exempel på täthetsfunktion för en Lognormalfördelad stokastisk variabel ses i Figur 43.



Figur 43 – Täthetsfunktion, lognormalfördelning

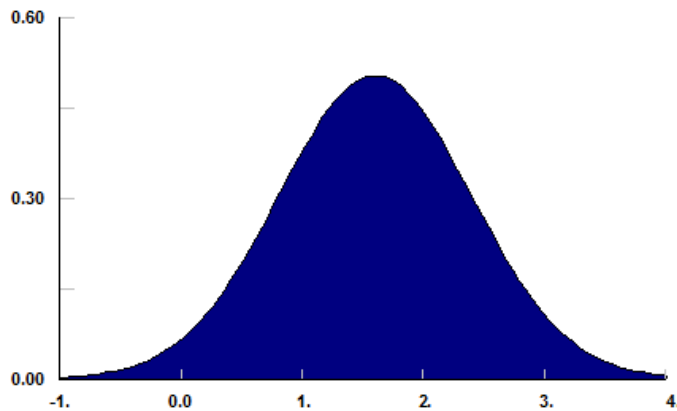
8.4 Normalfördelning

Normalfördelning används i vårt arbete för att i simuleringen beskriva hur ett tågs transporttid avviker från den planerade transporttiden. Inparametrar till fördelningen är medelvärde (μ) och standardavvikelse (σ) som beräknas utifrån datamaterialet som erhållits från Arriva. Sannolikhetsfunktionen är enligt nedan:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$$

(Everitt & Skron dal, 2010, p. 305)

Ett exempel på täthetsfunktion för en Normalfördelad stokastisk variabel ses i Figur 44.



Figur 44 – Täthetsfunktion, normalfördelning

9 Appendix C

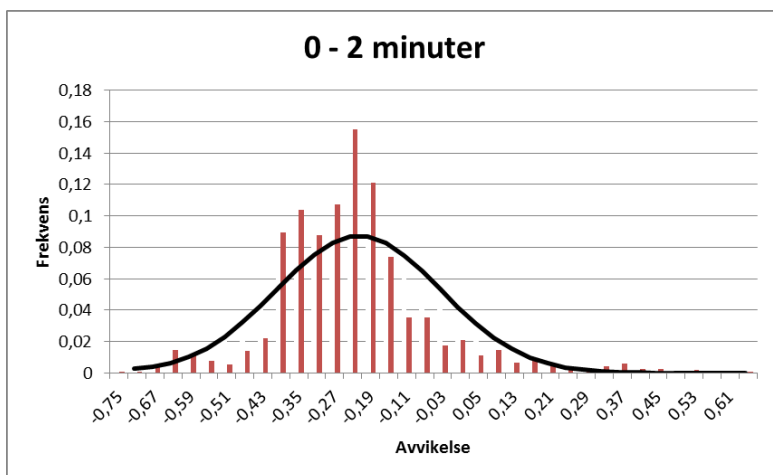
9.1 Parametrar transporttid

Appendix B visar medelvärde och standaravvikelse för den verkliga transporttidens avvikelse från den planerade enligt datamaterialet.

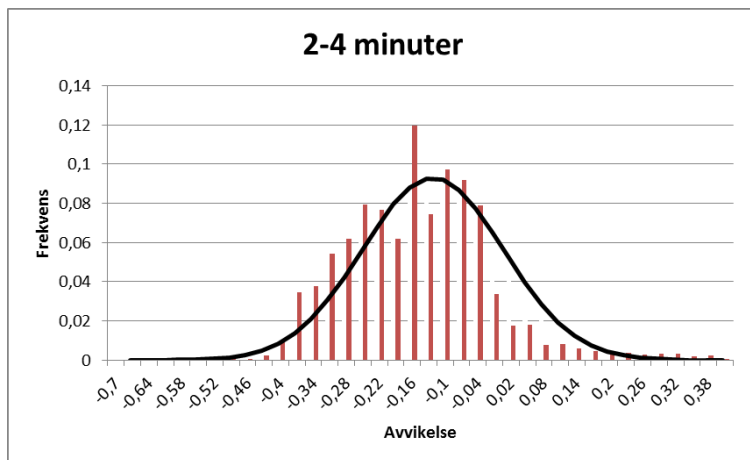
Tabell 5 visar de medelvärden och standardavvikelser som beräknats och importerats till Extendsim. Figur 45 - 49 visar datamaterialet plottade med motsvarande skattade Normalfördelningar.

Intervall	Mean	Std dev
0-2 min	-0,25135	0,181168
2-4 min	-0,14806	0,127632
4-6 min	-0,11725	0,092396
6-8 min	-0,13751	0,094257
>8 min	-0,09983	0,079257

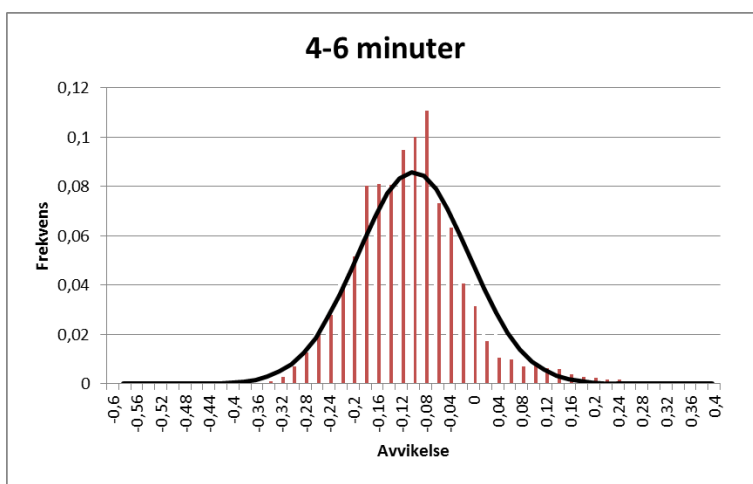
Tabell 5 - Medelvärde och standardavvikelse av olika transportsegment. Segmenten är uppdelade efter planerad transporttid i 2 minuters intervall.



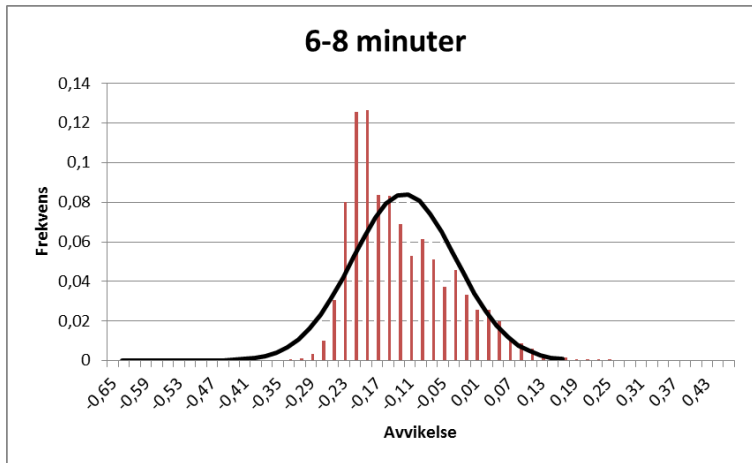
Figur 45 - Jämförelse mellan datamaterial och Normalfördelning för avvikelsen från planerade transporttiden. Grafen visar transporttider i intervallet 0 till 2 minuter.



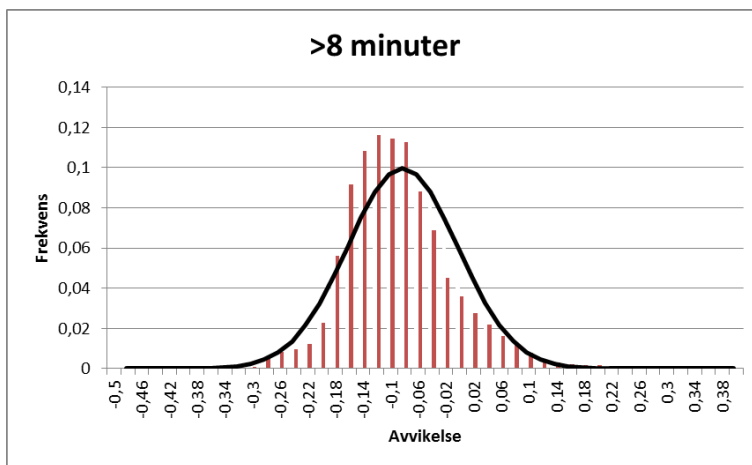
Figur 46 - Jämförelse mellan datamaterial och Normalfördelning för avvikelsen från planerade transporttiden. Grafen visar transporttider i intervallet 2 till 4 minuter.



Figur 47 - Jämförelse mellan datamaterial och Normalfördelning för avvikelsen från planerade transporttiden. Grafen visar transporttider i intervallet 4 till 6 minuter.



Figur 48- Jämförelse mellan datamaterial och Normalfördelning för avvikelsen från planerade transporttiden. Grafen visar transporttider i intervallet 6 till 8 minuter.



Figur 49 - Jämförelse mellan datamaterial och Normalfördelning för avvikelsen från planerade transporttiden. Grafen visar transporttider i intervallet 8 minuter och längre.

10 Appendix D

Appendix C innehåller lista över samtliga tekniska stationer samt stationer med få mätvärden

10.1 Lista tekniska stationer

Förteckning av alla tekniska stationer vi har identifierat, sorterade efter tåglinjer. Stationer som är benämnda med namn återfinns i Arrivas interna tidtabeller och de som är benämnda med förkortningar återfinns i tabeller för minsta transporttider.

- 75 Aarhus - Skjern: Funder, HC (mellan Viby Jylland och Hørning)
- 82 Esbjerg – Niebüll: Inga tekniska stationer
- 84 Varde – Norre Nebel: Inga tekniska stationer
- 85 Skjern – Esbjerg: ESN (mellan Spangbjerg och Gjesing), KJE (mellan Gjesing och Guldager), ARN (mellan Varde och Varde Kaserne)
- 88 Struer – Skjern: SGS (mellan Holsterbro och Hjerm)
- 92 Thisted – Struer: Inga tekniska stationer
- 95 Aarhus – Struer: Rindholm, Ravnstrup, Rønbjerg, EST (mellan Skive och Rønbjerg), MU och BR (båda mellan Hinnerup och Aarhus)

10.2 Lista över stationer med få mätvärden

Förteckning över alla stationer med mindre än 100 mätvärden efter rensning. Rensning bestod av följande steg:

1. Avgångar som går till ogiltiga stationer, det vill säga ankommer till stationer som inte överensstämmer med tidtabell. Exempelvis finns det flera datapunkter där start och slutstation är samma station.

2. Stationer där planerad eller faktisk stationstid är noll. Undantaget är slutstationer.
3. Startstationer rensas bort eftersom jämförelsen är baserad på ankomsttider.

För att kunna relatera till siffrorna kan nämnas att tåglinje 75 har i medelvärde 2000 datapunkter per station. Följande stationer identifierades med mindre än 100 giltiga datapunkter i någon av riktningarna:

- 75 Aarhus - Skjern: Inga stationer
- 82 Esbjerg – Niebüll: Süderlügum, Uphusum
- 84 Varde – Norre Nebel: Baunhøj, Billum, Boulevarden, Dyreby, Frisvadvej, Henne, Hyllerslev, Janderup, Jegum, Lunde, Løftgård, Nr. Nebel, Oksbøl, Outrup (endast Varde har mer än 100 mätvärden)
- 85 Skjern – Esbjerg: Inga stationer
- 88 Struer – Skjern: Inga stationer
- 92 Thisted – Struer: Humlum
- 95 Aarhus – Struer: Inga stationer