



# Användning av tågpassagedata för estimering av primära störningsfördelningar

Anders Lindfeldt  
Hans Sipilä

Rapport

Stockholm 2016



# Användning av tågpassagedata för estimering av primära störningsfördelningar

Anders Lindfeldt  
anders.lindfeldt@abe.kth.se

Hans Sipilä  
hans.sipila@abe.kth.se

Kungliga Tekniska högskolan  
Transportplanering, ekonomi och teknik (TET)  
10044 Stockholm

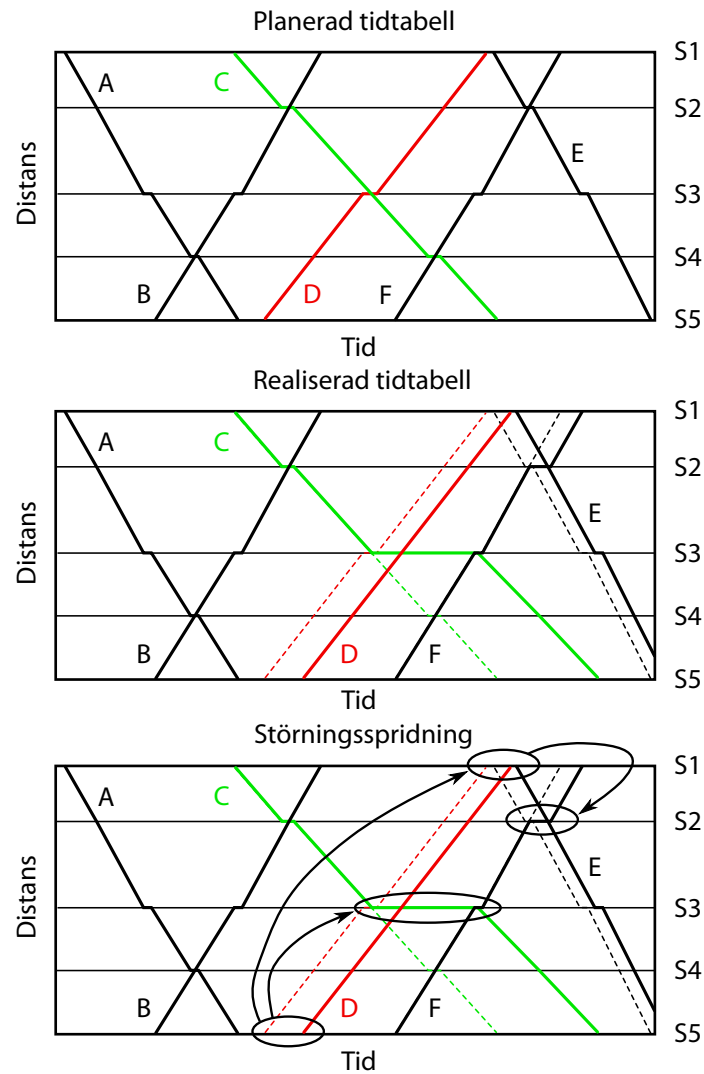
## Inledning

Simulering av tågtrafik erbjuder möjligheter att analysera och utvärdera olika scenarion i syfte att ge en uppfattning om skillnader i förväntat utfall mellan dessa. Vad som menas med utfall varierar, några vanliga mått är t.ex. punktlighet och medelförsening. Oavsett vilken typ av simuleringsmodell och metodik som används behöver ofta någon form av kontroll göras avseende överensstämmelsen mellan modellerat och verkligt utfall. Om en studie är inriktad på befintlig trafik och infrastruktur finns ofta historisk data som kan användas för kalibrering och validering av en simuleringsmodell. Om studien istället avser framtida infrastruktur eller helt nya trafikupplägg kan eventuellt data från befintliga delar i ett nätverk med liknande förutsättningar användas som grund för att skapa indata till simuleringen. Alternativt görs antaganden om framtida förutsättningar. Syftet med detta PM är att presentera metoder för estimering av primära linje- och uppehållsfördelningar (störningar) för användning i simulering. Metoderna bygger på användning av Trafikverkets tågpassagedata.

## Primära och sekundära förseningar

De förseningar som inträffar i ett verkligt järnvägsnät kan delas upp enligt bakomvarande orsaker vilka vanligtvis delas upp i två huvudkategorier, primära och sekundära förseningar. En händelse av primär karaktär initieras av en eller flera yttre orsaker, t.ex. problem med infrastruktur eller fordon och förlängd uppehållstid under passagerarutbyte. Om en primär störning, som leder till en tågförsening, inträffar kan det drabbade tåget senare störa andra tåg så att även dessa drabbas av förseningar. De övriga tågen som påverkas har då fått sekundärförseningar, d.v.s. förseningar som orsakas av att ett eller flera andra tåg redan är försenade och inte följer sin tidtabell. Ett annat sätt att uttrycka det som sker är att en initial händelse som ger en primärförsening för ett tåg sprids till andra tåg, antingen direkt via det första tåget eller indirekt genom att flera tåg påverkar varandra.

Figur 1 illustrerar hur en initial försening för tåg D vid station S5 sprider sig till ytterligare tre tåg. Tåg C och E drabbas direkt av att tåg D inte följer sin tidtabell. Tåg F drabbas indirekt genom att tåg E blivit försenad. I ett stort järnvägsnät eller en linje med tät trafik kan initiala störningar spridas till många tåg och dessutom i flera led. Olika typer av marginaler och tidstillägg i en tidtabell kan ha en motverkande effekt på förseningsspridningen.



Figur 1: Exempel på skillnader i planerad och realiserad tidtabell på en enkelspårig linje. Tåg D startar med en initial försening alternativt medför den från en annan del av järnvägsnätet vid station S5 och orsakar sekundära förseningar på tåg C och E samt indirekt på tåg F.

## Störningstyper i simulering

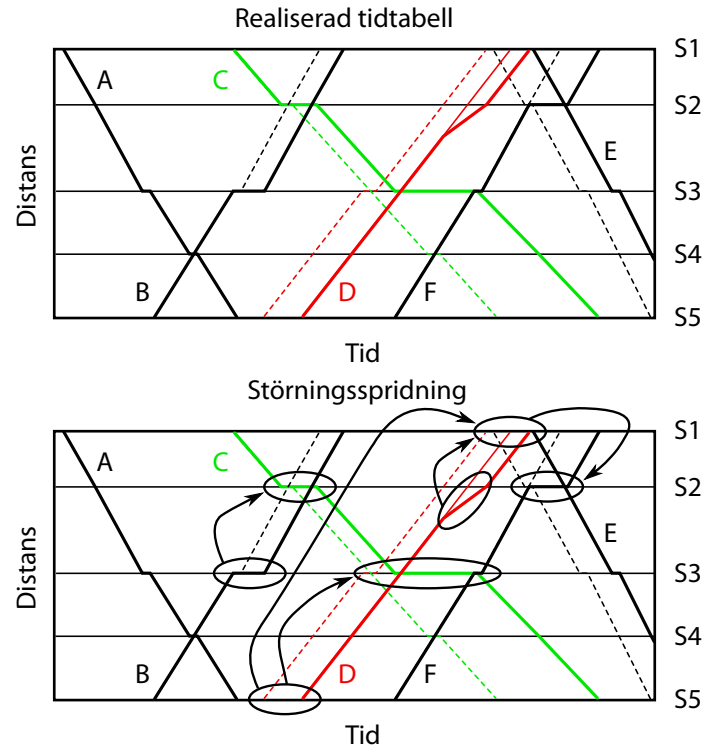
Givet att infrastruktur, fordon och tidtabell är införda tillkommer generellt tre huvudtyper av stokastiska och/eller deterministiska störningar. Syftet med dessa är att skapa variation i förhållandet mellan planerad tidtabell och den trafik som i verkligheten uppnås. Tåg som initieras i en simulering påförs störningar som t.ex. beskriver variationer i tågens ankomst till det modellerade nätverket. Vissa tåggrupper kan uppvisa en låg medelförsening och spridning medan andra grupper kan ha en stor andel högre avvikelser och stor spridning. På liknande sätt kan tåg som i verkligheten startar på någon station som ingår i det simulerade nätverket utsättas för samma typ av störningar.

Tåg som gör planerade uppehåll, t.ex. i form av passagerarutbyte på en driftplats, uppvisar tidsvariationer relativt de planerade tiderna. Tillvägagångssättet för att uppnå detta i simuleringar är, liksom med föregående exempel, att införa lämpliga störningsfördelningar och koppla dessa till olika platser och tåg. En vanlig tillämpning är, som redan nämnts, uppehåll för passagerarutbyte men samma princip gäller även för andra typer av uppehåll, t.ex. förarbyte eller aktiviteter för godståg.

Både störningar vid uppehåll och vid initiering av tåg kopplas till specifika platser och tåggrupper, alternativt enstaka tåg. En ytterligare störningstyp som kan krävas i syfte att närma sig ett realistiskt beteende i simuleringar är störningar som inträffar på linjen mellan olika driftplatser. Dessa medför att det uppstår avvikelser relativt tågens planerade körtider och används t.ex. för att till viss del modellera förarbeteende, avvikelser i tågkonfigurationer mellan planerad och utförd trafik, väderrelaterade problem eller andra mindre problem med fordon och infrastruktur.

Ett exempel som bygger på den som visas i figur 1 visas i figur 2. Tåg B drabbas av en uppehållsförsening vid station S3 och påverkar tåg C vid station S2 som får en sekundärförsening. Tåg D startar med en initial försening alternativt medför den från en annan del av järnvägsnätet vid station S5 och orsakar sekundära förseningar på tåg C och E samt indirekt på tåg F. Tåg D drabbas dessutom av en primärförsening mellan station S3 och S2 som, tillsammans med den ursprungliga förseningen vid station S5, bidrar till förseningarna för tåg E och F. Eftersom tåg D redan är försenad vid ankomst till station S3 sker ingen förseningsöverföring från tåg C, som också är försenat, till tåg D.

I en simulering behöver störningarna för tåg B vid station S3 och tåg D vid station S5 samt mellan S3 och S2 ges av indata. De övriga störningarna kan ses som reaktiva händelser på dessa och modelleras i själva simuleringen. Det är med andra ord de primära störningarna som behövs innan en simulering. Idealt kan de fördelningar som krävs för att beskriva indata bestämmas med stor noggrannhet. I praktiken är det dock ofta svårt att uppnå detta eftersom den data som behövs inte finns tillgänglig med tillräckligt stor noggrannhet och kan dessutom bestå av en blandning mellan primär- och sekundärförseningar.



Figur 2: Planerad tidtabell enligt figur 1. Tåg B drabbas av en uppehållsförsening vid station S3 och påverkar tåg C vid station S2 som får en sekundärförsening. Tåg D startar med en initial försening alternativt medför den från en annan del av järnvägsnätet vid station S5 och drabbas av en primärförsening mellan station S3 och S2. Tåg C, E och F påverkas.

Förseningar kan även spridas utan att tågen interagerar med varandra. Antag att ett passagerartåg som skulle ha körts innan ett annat passagerartåg saknas p.g.a. försening eller att den ställts in från en tidigare plats. Det andra passagerartåget kan då förseñas vid ett flertal stationer eftersom den passagerarmängd som skulle ha delats upp på två tåg är hänvisade till ett tåg och det finns en sannolikhet att de planerade uppehållstiderna överskrids. Den eventuella försening som uppstår kan ur ett perspektiv ses som sekundär eftersom orsaken är att ett annat tåg inte följer sin tidtabell. Ur ett simuleringsperspektiv kan man dock behandla de eventuellt överskridna uppehållstiderna som primärstörningar.

## Linjeförseningar

Denna del beskriver ett förslag på metodik som genom att utnyttja tågpassagedata kan användas för att uppskatta de primära linjestörningarna som behövs som indata till simuleringar. Syftet är att relativt snabbt bearbeta stora mängder data och därigenom förenkla och effektivisera processen för kalibrering och validering av en simuleringsmodell. Metoden är utvecklad i Python och MATLAB men de delarna som görs i MATLAB kan i princip ersättas av motsvarande rutiner i Python eller liknande.

### Allmänt om tågpassagedata

Förseningsdata (tågpassagedata) insamlas kontinuerligt när tågen passerar t.ex. driftplatser i järnvägssystemet. Grunden för tågpassagedata är en planerad tidtabell för olika former av tågfordonsrörelser och tidsstämplar för realiserade passager. Båda tidsangivelserna har en upplösning på hela minuter. De realiserade tiderna är dessutom trunkerade vilket t.ex. betyder att tiderna 17:40:05 och 17:40:55 anges som 17:40 i den data som presenteras. Det innebär att vid beräkning av faktiska avvikelser uppkommer en viss underskattning av försenade tåg och på motsvarande en överskattning av tidiga tåg. Dessutom sker en tidskompensation på vissa driftplatser i syfte att reflektera en mer central punkt på platsen i förhållande till var själva registreringen initieras. Vilka spårledningningar som registrerar tågpassager och hur eventuell tidskompensering ser ut på olika driftplatser är inte känt i detalj.

I tågpassagedatan redovisas planerade tidtabeller för tåguppdragen, dessa innehåller bl.a. uppgifter om på vilka platser uppehåll ska ske och tiden för dessa, däremot finns inte information om uppehållstyper. Exempel på dessa är resandeutbyte, annan aktivitet eller av teknisk karaktär som främst avser tågmöte eller förbigång. Ett uppehåll kan samtidigt uppfylla flera funktioner. Vad gäller resandetåg kan information om platser för resandeutbyte inhämtas från andra källor. Om ett planerat uppehåll inte görs p.g.a. trafiksituationen framkommer det oftast genom att den realiserade ankomst- och avgångstiden för en plats har samma tidstämpelvärde. Om däremot ett uppehåll som inte är planerat i tidtabell behöver göras saknas verkliga ankomsttider. Det leder till att det inte fullt ut går att bestämma hur tågen trafikerat en viss bana, mer specifikt uppkommer problem att fördela tiden mellan två på varandra följande avgångsregistreringar på faktiskt körtid och uppehållstid. I teorin är det dock möjligt att göra antaganden för ett tåg genom att betrakta omkringliggande tågs registreringar om sådana finns.

### Inläsning av tågpassagedata

I arbetet med att föreslå en metod för att effektivisera processen för att erhålla kalibrerade och alternativt även validerade simuleringsmodeller av olika storlek i RailSys är det nödvändigt att hantera stora mängder tågpassagedata. För att i ett

senare skede möjliggöra sökningar i syfte att bilda fördelningar och annan data för olika tåggrupper och sträckor görs en inläsning med hjälp av Python-script och Pandas (tilläggsmodul för dataanalys) till ett komprimerat HDF-format. Den erhållna datan för år 2014 är uppdelad på månadsfiler och består totalt av knappt 35 miljoner rader uppdelade på 37 kolumner med en blandning av numerisk data och text. I tabell 1 ges några exempel på vilken typ av information som kolumnerna i den ursprungliga datan innehåller.

Tabell 1: Exempel på information i tågpassagedata.

Tåguppdrag	Numeriskt värde, kopplar vid behov ihop flera tågnummer till ett uppdrag. Tågnummerbyten sker t.ex. vid ändringar i riktningdefinitioner vid övergång mellan olika banor.
Tågnummer	Numeriskt värde, vanligtvis samma nummer som återfinns i de grafiska tidtabellerna eller den så kallade tidtabellsboken.
Datum-PAU	Datumstämpel för varje tåguppdrag baserat på avgång från första station. Samma tåguppdragsnummer har unika Datum-PAU stämplat.
Tågslag	Bokstavskombination som anger om tåget är t.ex. ett godståg, resandetåg, tjänstetåg, växlingsrörelse eller spärrfärd.
Inställtflagga	Anger om och på vilka stationer ett tåg är inställt.
Uppehållstyp	Uppdelat på ankomst och avgång. Anger främst om ett tåg har planerat uppehåll eller inte på en plats.
Delsträckanummer	Anger platsordning för tågnummer. Kan användas för att bestämma platsordning om datan t.ex. är sorterad på platser, alternativt görs en sortering på Datum-PAU, tåguppdrag och planerade avgångstider för respektive uppdrag.
Avgångs/ankomstplats	Varje rad i datan beskriver en förflyttning mellan två platser i järnvägsnätet. Finns med både platsnamn och platssignatur.
Planerad avgångs/ankomsttid	Information uppdelat på flera kolumner, görs vid inläsning om till tidstämplat av formen <code>åååå:mm:dd tt:mm:ss</code> .
Avgångs/ankomsttid	Realiserade tider, i övrigt se ovan.
Dragfordon ID	Textsträng som anger fordonstyp, ofta kombinerat med ett nummer, t.ex. <code>X2K2032</code> eller <code>RC6-EP1329</code> .



Ytterligare exempel på data är uppgifter om rapporterad tåglängd, antal hjulaxlar och dragfordon. Informationen om dragfordonsidentitet är inte konsekvent angivet och samma fordonstyp kan kodas på olika sätt. Uppgifter om fordonstyp är av intresse eftersom det kan användas för att t.ex. söka data för tåg där åtminstone en av söksträngarna matchar delar av strängen för dragfordonsidentitet. Metoden kan delvis ersätta sökningar baserade på tåguppdrag eller tågnummer.

I samband med inläsningen görs en mappning av viss textdata till numerisk data och i vissa fall till boolesk data, syftet är att minska på datastorleken och i teorin åstadkomma snabbare sökningar. Realiserade tidstämplarna saknas helt vid hållplatser och hållställen för tåg utan planerade uppehåll. Dessutom saknas dessa vid vissa driftplatser för samtliga tåg under vissa perioder vilket förmodligen beror på att delar av systemet varit ur bruk eller inte fungerat som avsett.

## Bearbetning av tågpassagedata

För att bilda de störningsfördelningar som normalt används i en simulering krävs någon form av historisk data och ibland antaganden om framtida förhållanden. Om syftet är att endast jämföra olika scenarion sinsemellan, utan ambition att göra jämförelser mot verkligt utfall, kan det räcka med enbart antaganden om fördelningarnas medelvärden och spridning. Även i ett sånt fall bör dock någon form av rimlighetsbedömning göras. Ofta jämförs resultatet från en simulering med verkligt utfall även om trafikperioderna inte överensstämmer. Vanliga mått för jämförelser är medelförseningar, försenings-spridning och punktlighet där den sistnämnda avser andelen tåg som ligger under ett visst förseningsgränsvärde. Medelvärden är inte alltid representativa om spridningen är stor, t.ex. med flertal stora värden, och det kan vara nödvändigt med kompletterande mått och villkor.

Initialstörningar, d.v.s. störningar som påverkar tåg när de skapas, kan relativt lätt bildas från data genom att söka registreringar för en viss grupp av tåg vid en specifik plats och riktning. Dessa kan i teorin också kräva en anpassning eftersom avgångsregistreringar är påverkade av både primär- och sekundärhändelser och det är en uppgift för själva simuleringen att generera realistiska sekundärstörningar givet primärstörningar och annan indata. Det gäller t.ex. för tåg som initieras på andra stationer än de yttersta i ett simuleringsnätverk och där eventuella försenade avgångar dels kan orsakas av andra tåg eller av verkliga primärhändelser. Ett annat ställningstagande är huruvida begränsningar behövs avseende maximala förseningar.

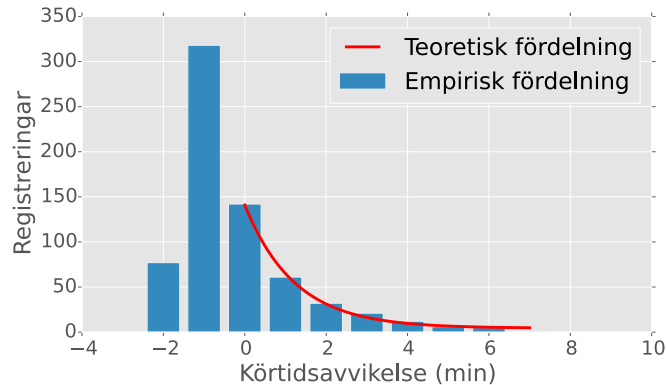
Primära linjestörningar kan ses som en delmängd av den fördelning som ges av skillnaden mellan planerade och realiserade körtider. Begränsningen för en sträcka kan dels utgöras av två på varandra följande driftplatser eller bestå av flera sådana sträckor i följd. Naturliga gränspunkter utgörs av större stationer eller andra platser där tåg startar eller avslutar sitt uppdrag, alternativt att två eller flera linjer korsar eller ansluter. Om en alltför kort sträcka används innebär det att tidstämplarnas upplösning och trunkering på hela minuter relativt sett kan

påverka mer och ge en skev fördelning. Idén är att utgå från fördelningar som beskriver körtidsavvikelser och anpassa dessa genom att antingen ändra parametrar i en empirisk fördelning eller anpassa den empiriska till en teoretisk fördelning och därefter ändra de parametrar som beskriver den.

I nuläget medger simuleringsprogrammet RailSys via gränssnittet stöd för empiriska och exponentialfördelningar. Andra teoretiska fördelningar kan i princip användas men det krävs att dessa beskrivs i empirisk form. Om hantering av processen som tilldelar störningsvärden till tåg och genererar den data (XML-format) som krävs av RailSys finns inga egentliga begränsningar avseende fördelningstyper. Fördelen med att använda teoretiska fördelningar är att dessa oftast kan justeras enkelt genom att ändra ett fåtal parametrar, till skillnad mot en empirisk fördelning där det inte är entydigt på vilket sätt fördelningen ska ändras för att t.ex. uppnå ett visst medelvärde. Det måste dock betonas att formen på empiriska data kan vara sådan att den inte beskrivs tillräckligt väl av någon teoretisk fördelning, i de fall krävs rutiner för att anpassa empiriska fördelningar. I en simuleringsstudie på Södra stambanan ändras empiriska fördelningar genom att behålla en viss procentuell andel registreringar på varje minutnivå och flytta resten till nollnivån [1]. Någon koppling till medelvärden eller standardavvikelser för de empiriska fördelningarna görs inte.

I fallstudien som redovisas anpassas den delen av körtidsavvikelsefördelningar som representeras av längre körtider än planerat till exponentialfördelningar med ett visst medelvärde. I det här läget tas ingen hänsyn till om tåget ligger före eller efter sin tidtabell, enbart den faktiska avvikelsen i körtid. Andelen registreringar med en längre körtid än planerat bestämmer med vilken frekvens, d.v.s. hur ofta, värden dras slumpmässigt ur aktuell fördelning när störningarna genereras innan en simulering. Även en delmängd av registreringar med värdet noll kan inkluderas i denna andel med hänvisning till trunkeringen av tidstämplar. Figur 3 visar hur den positiva delen av en empirisk fördelning kan anpassas till en exponentialfördelning.

Utgångspunkten i den ansatsen som presenteras är att den verkliga fördelningen (som inte är känd) för primärstörningar på linjen inte har högre medelvärde eller spridning än den anpassade fördelningen. Eftersom fördelningen över körtidsavvikelser representerar den sammanlagda effekten av primära och sekundära störningar är det inte orimligt med antagandet att nivån på de verkliga primärstörningarna ligger under den ursprungliga fördelningen. Om det finns möjligheter för betydande förseningsåterhämtning på en sträcka, t.ex. tidstillägg i tidtabellen eller att möten respektive förbigångar frekvent uteblir av olika orsaker, kan den verkliga fördelningen för primärstörningar ligga högre än motsvarande totala fördelning. I de situationerna blir det alltså nödvändigt att t.ex. öka medelvärdet istället för att sänka den.



Figur 3: Skillnad mellan realiserad och planerad körtid, empirisk och anpassad teoretisk fördelning.

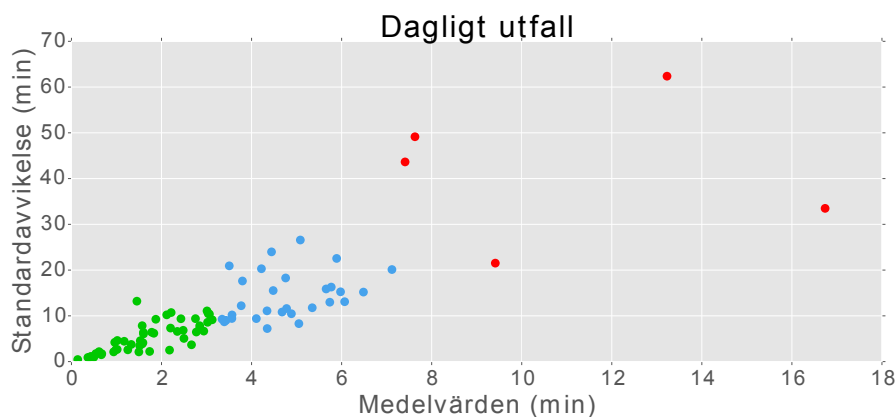
I tidigare studier har linjestörningarna hanterats genom att dela in simuleringsnätverket i ”lagom” långa sektioner som typiskt sträckt sig över 3–6 driftplatser beroende på avstånd. Anledningen till det har delvis varit att en tätare uppdelning medför mer arbete med att färdigställa indata samt att en alltför tät uppdelning inte nödvändigtvis motsvarar verkligheten bättre om hänsyn tas till att störningarna fördelas oberoende av varandra. Dessutom innebär upplösningen i hela minuter att felet procentuellt sett kan bli högre om för korta sträckor används vid bearbetning av tågpassagedata. Det finns dock inget hinder att använda kortare sträckor jämfört med tidigare eftersom en del av idén med metoden är att olika parametrar som styr hur indata tas fram lätt ska kunna ändras

## Fallstudie

Metodiken används i en fallstudie på Ostkustbanan mellan Gävle och Sundsvall. En förenkling görs genom att använda Strömsbro (strax norr om Gävle C), vidare modelleras inte korsande trafik vid gränsplatserna. En typtidtabell skapas för en torsdag i oktober 2014 baserat på Trafikverkets tågplan T14. Natten innan och efter används som ”warm-up” respektive ”cool-down” perioder och utvärderas inte. Inga andra parametrar varierar förutom nivån på medelvärdet för respektive exponentialfördelning. Tågen delas in i fyra grupper: snabbtåg, regionaltåg, övriga resandetåg och godståg. Det är dessa som används för att söka efter tåg i passagedatan, både fordonstyp och tågslag utnyttjas. Vardagar under perioden 17 augusti till 13 december används som underlag.

För att jämföra simuleringsutfall och modellera en kalibrering respektive validering delas dagarna upp i en grupp med lägre medelförseningar och standardavvikelser samt en grupp med högre enligt figur 4. En liknande uppdelning görs även i [2]. Fem dagar med tydligt avvikande värden används inte i någon av grupperna. Förseningar för södergående tåg vid Strömsbro samt för norrgående tåg vid Sundsvall används. Alla fyra tåggrupper och två riktningar är här ihopslagna.

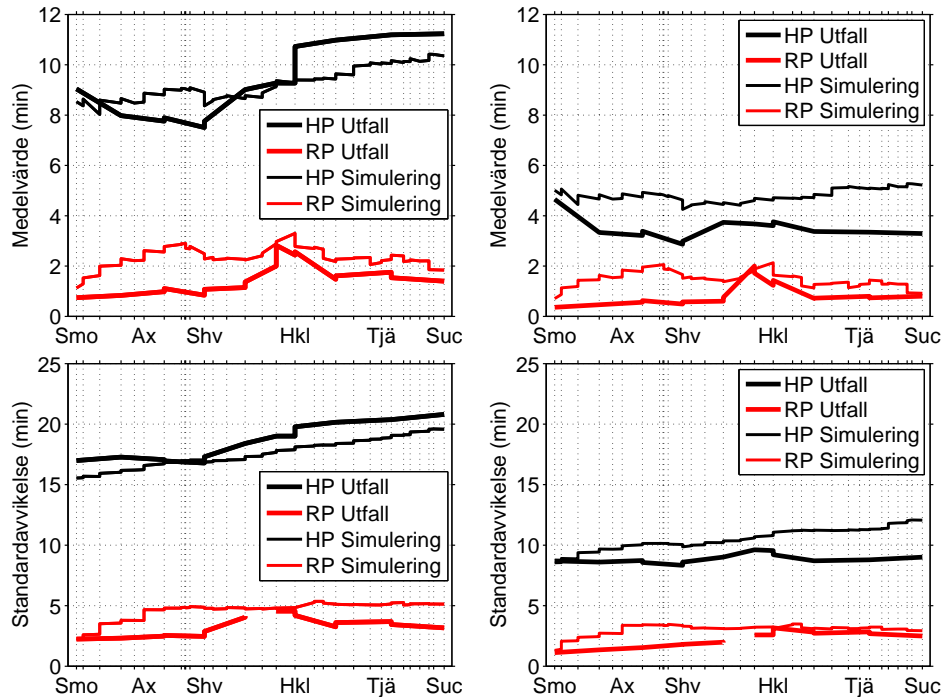
Andra sätt att göra en uppdelning är t.ex. att utgå från förseningar som tågen har när de kommer i det studerade området eller skillnaden mellan in- och utförsening.



Figur 4: Dagligt utfall på Ostkustbanan mellan Gävle–Sundsvall, vardagar kl. 4–24 under perioden 17 augusti till 13 december 2014.

Exempel på resultat visas som diagram över medelvärden och standardavvikelser för tre tåggrupper. Överensstämmelsen mellan simulering och verkligt utfall är bäst vid nivån 40 %, som i det fallet betyder att exponentialfördelningarnas medelvärden är 40 % av den ursprungliga nivån. De verkliga utfallen plottas för nio platser längs sträckan medan simuleringsresultaten plottas för varje plats. Både ankomst- och avgångsvärden visas, för verkligt utfall sätts ankomstregistrering lika med avgångsregistrering i de fall inget planerat uppehåll ligger inlagt.

Figur 5 visar resultat för norrgående snabb- och regionaltåg. Platserna som anges i diagrammen är Strömsbro (Smo), Axmarby (Ax), Söderhamns västra (Shv), Hudiksvall (Hkl), Tjärnvik (Tjä) och Sundsvall C (Suc). De simulerade värdena ligger mestadels något högre än verkligt utfall. Diskontinuiteten i två av kurvorna för verkligt utfall är kopplat till det faktum att när standardavvikelsen är noll tyder det på att väldigt få värden finns registrerade vid någon plats med hänsyn till antalet dagar och tåg i aktuell grupp. Det sker också en kontroll av antalet registreringar vid genereringen av fördelningar och annan data. En plats med få registreringar bör inte utgöra gräns för en körtidsavvikelsefördelning. Skillnaderna mellan simulerat och verkligt utfall på nivån 40 % är i samma storleksordning för södergående tåg.

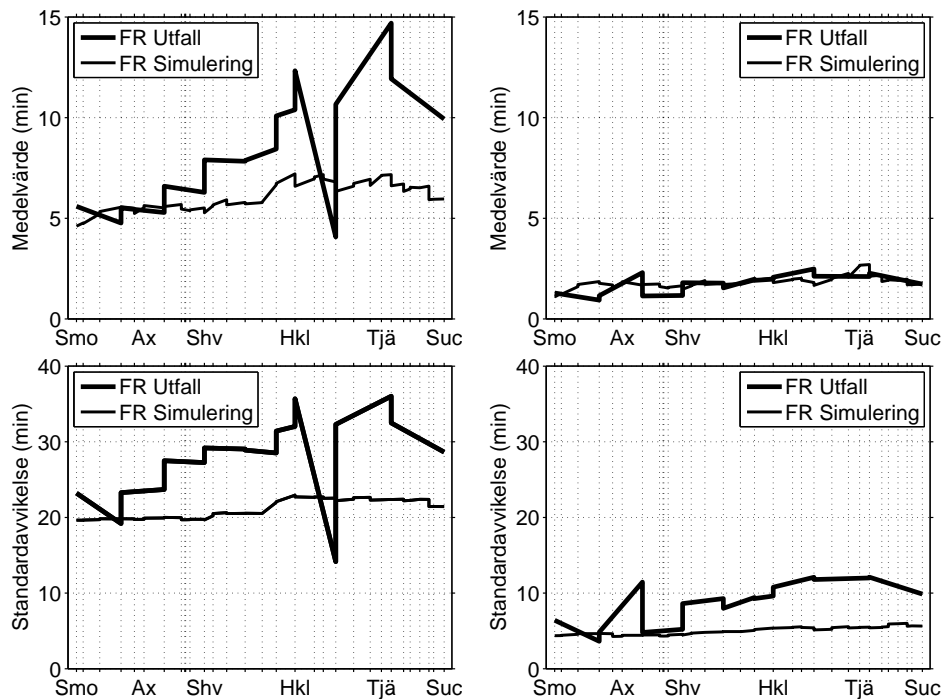


Figur 5: Medelvärden och standardavvikelser för snabbtåg (HP) och regionaltåg (RP) från Strömsbro till Sundsvall. Urvalsgruppen med höga förseningar till vänster och låga förseningar till höger.

Resultaten för godståg visas i figur 6. Godstågens beteende är generellt svårare att efterlikna än resandetågens i simuleringar. I grunden beror det dels på att i princip varje godståg har en individuell tidtabell och inte ingår i system som de flesta resandetåg, dessutom uppvisar godståg betydligt större avvikelser mot planerad tidtabell. Godståg kan ha stora förseningar men också ligga betydligt före tidtabellen. Scenariot för dagar med höga förseningar uppvisar större avvikelser mellan simulering och verkligt utfall än scenariot för dagar med låga förseningar. Detta är också vad som kan förväntas eftersom skillnaden mellan simulering kan antas växa ju större förseningar och varians tågen uppvisar.

I scenariot för dagar med höga förseningar uppvisar kurvorna för verkligt utfall en anmärkningsvärd avvikelse vid driftplatsen Harmånger. Denna är svårförklarad enbart genom att studera diagrammen och kräver en djupare analys av tågpassagedatan med tanke på att avvikelsen är mellan ankomst- och avgångsvärden vid samma plats. Eftersom enbart planerade uppehåll genererar verkliga ankomstvärden och övriga tåg ges samma värde för ankomst som för avgång, kan det möjligtvis förklaras av att flera godståg i så fall har ett planerat uppehåll just vid den platsen. Detta bör dock kontrolleras i tågpassagedatan. Med tanke på att omkringliggande värden ligger på liknande nivåer, kan det även röra sig om

felangivelser i tågpassagedatan eller fel vid inläsningen av datan. Beteendet syns dock inte i scenariot för dagar med låga förseningar.



Figur 6: Medelvärden och standardavvikelse för godståg (FR) från Strömsbro till Sundsvall. Urvalsgruppen med höga förseningar till vänster och låga förseningar till höger.

## Uppehållsförseningar

Tåguppehåll är en viktig källa till förseningar. Tåg som stannar för uppehåll för passagerarutbyte, omlastning av gods, personalbyte etc. riskerar att bli försenade om uppehållet tar längre tid än vad som är planerat i tidtabellen. Dock medger uppehåll också en möjlighet att hämta igen förseningar. Ett tåg som ankommer till stationen sent och lyckas göra ett uppehåll som är kortare än vad som är planerat, lämnar stationen med en mindre försening än vad det hade när det ankom. Tågens uppehåll står således för en stor del av tågens dynamik relativt tidtabellen och det är viktigt att beakta och modellera dessa på ett realistiskt sätt i de modeller som används när järnvägstrafik och förseningar analyseras.

I RailSys används i huvudsak tre parameterar för att modellera planerade uppehåll: planerad uppehållstid ("scheduled stop time"), minsta uppehållstid ("minimum stop time") och förlängd uppehållstid ("dwell time extension"). Planerad

uppehållstid är den tid som det är planerat att tåget ska stå still vid en viss punkt (plats), d.v.s. skillnaden mellan avgångstid och ankomsttid. Minsta uppehållstid är den kortaste möjliga tid det är antaget att ett uppehåll kan ta. För ett passagerartåg kan det t.ex. motsvara den tid det tar att öppna och stänga dörrarna, utan någon tid för passagerare att gå av eller på, och den tid det tar innan alla tekniska system på tåget ger klartecken till föraren att åka. Planerad uppehållstid ges av tidtabellen och minsta uppehållstid specificeras för varje enskilt uppehåll i RailSys. Vilken tid som bör användas som minsta uppehållstid kan variera beroende på bl.a. tågtyp och typ av uppehåll. Dessutom beror det även av hur en uppehållsfördelning är konfigurerad. En vanlig minsta uppehållstid är 30 sekunder och härrör från manuella mätningar av uppehållstider för X2000 utförda under år 2007 [4].

Om inte någon förlängd uppehållstid definierats i RailSys, är tåget redo att avgå efter att den minsta uppehållstiden har passerat. Den minsta uppehållstiden är på förhand bestämd och varierar inte i RailSys-simuleringar. I verkligheten tar inte alla uppehåll lika lång tid, utan de varierar i längd beroende på till exempel antal av- och påstigande. I RailSys modelleras denna varians i uppehållstid genom att addera en förlängd uppehållstid till den minsta uppehållstiden, vilket innebär att tåget måste göra ett uppehåll som är minst lika långt som summan av minsta uppehållstid och förlängd uppehållstid. Den förlängda uppehållstiden modelleras som en stokastisk sannolikhetsfördelning. Fördelningen bestämmer hur sannolikt det är att få förlängda uppehållstider av olika längd. Tidigare forskning vid KTH [4] har funnit att lognomala fördelningar passar uppmätt data väl.

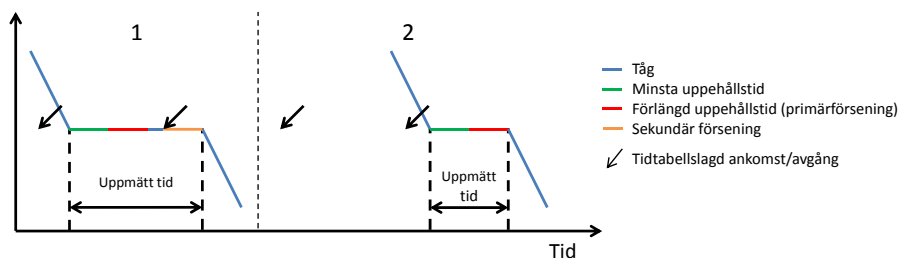
Detta avsnitt syftar till att presentera en metod för att estimeras dessa fördelningar ur tågpassagedata från Trafikverket. Ett alternativ är som tidigare nämnts att utföra manuella mätningar genom att ha personal på plats som mäter längden på uppehållen. Detta är dock ett tids- och resurskrävande alternativ till att använda den data som samlas in automatiskt av Trafikverkets system. En ytterligare möjlighet är att bearbeta olika typer av information från fordonsloggar.

## Metod

Det som i idealfallet uppmäts av signalanläggningen är tågets totala stopptid, d.v.s. hur länge det står still vid plattformen. Hur länge tåget blir stående vid plattformen beror på hur lång tid det tar innan tåget är redo att avgå, tågets ankomsttid relativt tågets planerade avgångstid (tåget får inte avgå före tidtabellen) och eventuella sekundärförseningar som gör att tåget blir stående vid plattformen trots att tåget är redo att avgå och den planerade avgångstiden har passerats (figur 7. I RailSys ser trafikledningsalgoritmen till att tåget inte avgår före tidtabellen (givet att den uppehållstypen inte tillåter det) och räknar ut eventuella sekundärförseningar. Det är därför inte önskvärt att den tiden av uppehållet som beror på att tåget inväntar avgångstid eller på sekundära förseningar inkluderas när den förlängda uppehållstiden skattas då detta skulle medföra att primärförseningarna skulle bli för stora och tågen mer försenade i modellen rela-

tivt verkligheten. För att reducera inverkan av sekundärförseningar och att tåg inväntar avgångstid, ställs följande krav för att en registrering i tågpassagedatan ska användas i analysen:

1. Det aktuella tåget ska ankomma den aktuella platsen samtidigt som, eller efter tågets planerade avgångstid från den aktuella platsen.
2. Inga andra tåg får ha registrerade ankomsttider eller avgångstider på den aktuella platsen för tätt inpå det aktuella tågets registrerade avgångstid.



Figur 7: Modellering av uppehåll.

Genom att kräva att tåget ankommer minst lika sent till stationen som uppehållsplanens planerade längd, elimineras observationer där tåg kan ha inväntat planerad avgångstid trots att de varit klara att avgå. Genom att kräva att inga andra tåg ankommer eller avgår från stationen inom ett visst tidsintervall runt det aktuella tågets registrerade avgångstid undviks sekundärförseningar p.g.a. möten, förbigångar och avsaknad av samtidigheter. Principen visas i figur 7. Utöver de redan nämnda kraven ställs även följande krav för att observationen ska användas:

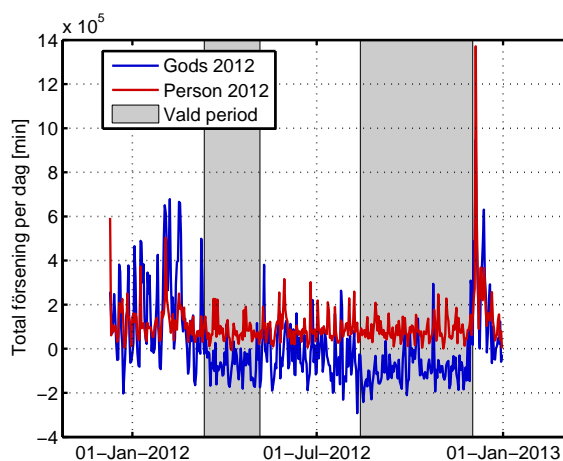
3. Observationen är fullständig, d.v.s. inga data saknas om planerad eller registrerad ankomst/avgångstid.
4. Tåget är av specificerad tågtyp, t.ex. RST (resandetåg) eller GT (godståg).
5. Tåget tillhör specificerat tågnummerintervall, t.ex. 400–499.
6. Observationen är gjord på någon av de stationer som specificerats, t.ex. LSE, SHV.
7. Observationen har den planerade uppehållslängden som angivits, t.ex. två minuter.
8. Observationen har gjorts under någon av de veckodagar som angivits, t.ex. måndag–fredag.
9. Registrerad uppehållstid överskrider inte specificerad maxtid, t.ex. fem minuter.



Om en observation uppfyller samtliga krav ovan, beräknas tågets uppehållstid som skillnaden mellan realiserad avgångstid och realiserad ankomsttid till platsen. För att erhålla den förlängda uppehållstiden subtraheras den på förhand definierade minsta uppehålltiden från den uppmätta uppehållstiden. Därefter anpassas en lognormal sannolikhetsfördelning efter de de uppmätta förlängda uppehållstiderna.

## Indata

Tågpassagedata som används är ett utdrag som täcker hela år 2012 och inkluderar ankomsttider och avgångstider för samtliga tåg på samtliga driftplatser och i viss mån hållplatser i landet. Data är densamma som användes i [5]. All data används inte i analysen och perioder med extrema förseningar har valts bort. Total försening per dag under 2012 visas i figur 8 tillsammans med de valda mätperioderna 12 mars till 6 maj respektive 13 augusti till 2 december.



Figur 8: Total försening per dag under 2012 för valda mätperioder.

## Uppsättning

Metoden demonstreras på Ostkustbanan. Uppehåll för två olika grupper av tåg analyseras. Grupperna har olika uppehållsmönster och de planerade uppehållstiderna är olika långa, tabell 2. I övrigt är inställningarna desamma. För varje tåggrupp, görs fyra olika körningar för att demonstrera effekten av punkt 1 och 2 i avsnittet *Metod*:

1. Använda alla tåg
2. Eliminera tåg som inväntar avgångstid (punkt 1)
3. Eliminera tåg som kan ha fått sekundärförse-  
ningar (punkt 2)
4. Eliminera både tåg som inväntar avgångstid och kan ha fått sekundärförse-  
ningar (punkt 1–2)

Tabell 2: Inställningar för estimering av förlängda uppehållstider, tider i minuter.

Tågnummerintervall	500–599	8000–8999
Stationer	SHV, HKL	LSE, SHV, ID HKL, GNP
Planerad uppehållslängd	2	1
Minsta uppehållslängd	0.5	0.5
Max uppehållslängd	5	5
Minsta tid till närmast innan avg. tåg	5	5
Minsta tid till närmast innan ank. tåg	2	2
Använda veckodagar	Mån–Fre	Mån–Fre

## Resultat

I tabell 3 visas resultaten från beräkningarna. Resultaten är uppdelade för de två olika tåggrupperna var för sig med grupp 500–599 överst och 8000–8999 underst. Fyra kolumner med resultat visas som motsvarar de fyra fallen redovisade i föregående avsnitt. Första raden visar antal observationer som uppfyller de krav som ställs i de fyra fallen och därmed blivit godkända att användas i estimeringen av förseningsfördelningarna. Rad två och tre visar medelvärde och standardavvikelse av de observerade uppehållslängderna. Rad fem och sex visar medelvärde och standardavvikelse för de estimerade lognormalfördelningarna. Observera att fördelningarna representerar uppehållstidsförlängningen och att den minsta uppehållstiden, 30 sekunder, därför har subtraherats från de uppmätta uppehållslängderna innan fördelningarna skattats. Det är därför naturligt att värdena på rad två och tre är större än de på rad fyra och fem.

För båda tåggrupperna blir uppehållslängden kortare om endast försenade tåg eller tåg som inte riskerar att ha utsatts för någon sekundärförse-  
ning används. Detta är väntat och visar på behovet av att beakta detta i analysen. Elimineringen av observationer där tåg inväntar avgångstid har större effekt än den av sekundärförse-  
ningar. Kortast uppehåll blir det om både väntan på avgångstid och sekundärförse-  
ningar tas bort. Utan någon filtrering alls, kolumnen *Original*, gör båda tåggrupperna i snitt längre uppehåll än vad som är planerat, trots att realiserade uppehåll längre än fem minuter exkluderas. Tågen i 8000-serien har en minuts planerat uppehåll och tågen i 500-serien två minuter. I verkligheten gör också tågen i 8000-serien kortare uppehåll än tågen i 500-serien. Uppehållen är i

snitt 40 sekunder längre än planerat för tågen i 500-serien, motsvarande siffra för 8000-serien är 50 sekunder. Om sekundära förseningar och inväntan av avgångstid tas bort, gör 500-serien i snitt lika långa uppehåll som planerat och 8000-serien gör uppehåll som är 18 sekunder för långa, vilket skulle kunna visa på att planerad uppehållstid på en minut är för kort.

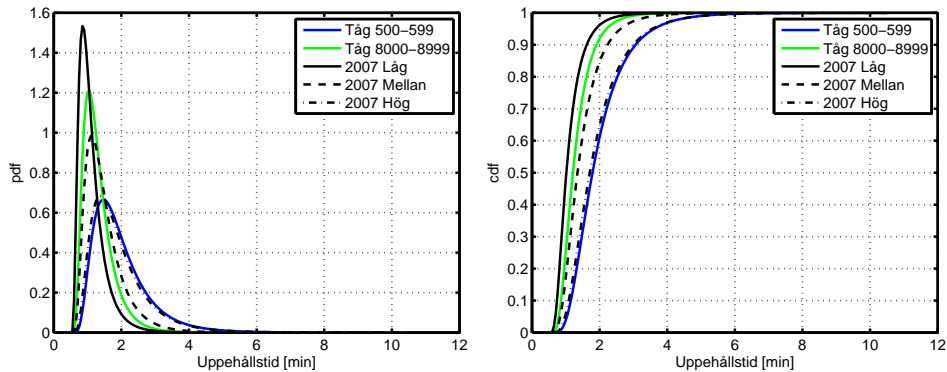
Tabell 4 visar egenskaperna hos de tre fördelningarna för X2000 som togs fram år 2007. En jämförelse med fördelningarna i denna studie visar att de ligger i samma område, både med avseende på medelvärde och på standardavvikelse. Figur 9 visar täthetsfunktionerna och de kumulativa täthetsfunktionerna för de framtagna fördelningarna tillsammans med de tre fördelningarna från 2007. Fördelningen för tågen i 500-gruppen är i princip samma som den höga fördelningen från 2007 och ungefär 60 % av tågen kan göra uppehåll som är två minuter eller kortare. Fördelningen för 8000-gruppen ligger mellan fördelning *Låg* och *Mellan* från 2007 och endast cirka 25 % av tågen klarar av att göra uppehåll som är en minut långa. Fördelningarna som skattats på data som inte filtrerats visar högre medelvärde och standardavvikelse än någon av fördelningarna från 2007.

Tabell 3: Resultat för tågnr 500–599 (överst) och tågnr 8000–8999 underst. Observera att rad två och tre i tabellerna visar mätt uppehållslängd och rad fyra och fem visar förlängd uppehållstid. Tider anges i minuter.

Tågnr 500–599 (2 minuters uppehåll)	Original	Endast sena tåg	Utan sekundära förseningar	Endast sena tåg utan sekundära förseningar
Antal observationer	25207	8924	2014	650
Uppmätt uppehållslängd medel	2.66	2.24	2.48	1.96
Uppmätt uppehållslängd std	1.07	0.95	0.92	0.68
Förlängd uppehållstid lognormal medel	2.23	1.80	2.02	1.49
Förlängd uppehållstid lognormal std	1.43	1.22	1.17	0.86
Tågnr 8000–8999 (1 minuts uppehåll)				
Antal observationer	48016	16384	1964	875
Uppmätt uppehållslängd medel	1.88	1.34	1.67	1.30
Uppmätt uppehållslängd std	1.16	1.00	0.85	0.59
Förlängd uppehållstid lognormal medel	1.48	1.02	1.19	0.82
Förlängd uppehållstid lognormal std	1.24	0.72	0.86	0.47

Tabell 4: Uppehållsförseningar från 2007 [4], tider i minuter.

	Manuell mätning från 2007		
	X2 Liten	X2 Mellan	X2 Stor
Driftplatser	SÖÖ	SÖÖ, K	SK
Förlängd uppehållstid medel	0.62	0.99	1.42
Förlängd uppehållstid lognormal std	0.41	0.60	0.87



Figur 9: Beräknade primärstörningsfördelningar. Notera att minsta uppehållstid har adapterats och att fördelningarna därför startar på 30 sekunder. Fördelningar från 2007 i svart.

## Begränsningar med metoden

Försekningsregistreringarna i tågpassagedatan kommer från signalsystemet och trafikledningssystemet i Trafikverkets järnvägsanläggning. Spårledningarna registrerar när ett tåg ankommer en plats/spår och när tåget lämnat platsen/spåret. Det finns vissa begränsningar i noggrannheten i systemet, bl.a. annat ligger inte alltid de spårledningarna som används för att mäta försekningsarna exakt vid plattformen där tåget stannar och deras exakta position kan variera från plats till plats, men även mellan olika spår på en och samma plats.

Dessa skillnader kompenseras till viss del genom att det i systemet finns möjlighet att justera de tidangivelser som registreras med en tidsfaktor. Tidsfaktorn gör det möjligt att få en bättre uppskattning av den tidpunkt då tåget faktiskt stannade/startade vid plattformen och syftet är därmed att öka jämförbarheten mellan registreringarna och den planerade tidtabellen. Hur väl denna kompensation fungerar är dock svårt att veta och de tar t.ex. inte hänsyn till tågets faktiska hastighet eller tågtyp. Osäkerheter i mätdata som introduceras av dessa tidsfaktorer är inte beaktade i det presenterade arbetet. En annan källa till problem är att alla tidangivelser trunckas/avrundas till hela minuter. Detta medför problem när uppehållstider mäts då dessa typiskt är små, 1–3 minuter, relativt den trun-

kering/avrundning som sker i systemet. Följden är att uppehållens längd mäts i hela minuter, t.ex. 0, 1, 2, 3 minuter o.s.v.

För att undvika observationer där tåg inväntar avgångstid eller blir påverkade av sekundärförseningar görs urval av vilka observationer som används i analysen. I tabell 3 framgår att endast runt 2 % av alla observationer används. Detta hårda urval kan påverka resultatet på andra sätt än vad som är önskvärt. Till exempel kan som en följd av att bara använda tåg som är försenade innebära att uppehållsförseningar överskattas då dessa i sig kan vara en av orsakerna till att tågen är försenade. Detta problem är dock troligen litet då en försening på några få minuter inte är ovanlig i sammanhanget.

I syfte att undvika att sekundärförseningar tolkas som primärförseningar slängs observationer där andra tågrörelser registrerats på samma plats för nära inpå avgångsregistreringen för det aktuella tåget. Detta är en grov metod som t.ex. inte tar hänsyn till spårlayouten på driftplatserna vilket medför att många användbara observationer slängs i onödan. I tabell 3 framgår att detta urval innebär att cirka 95 % av observationerna slängs. Att en så liten andel av observationer används kan medföra problem om metoden appliceras på större driftplatser med många tågrörelser. Det finns också en risk för ett skevt urval där t.ex. trafik under lågtrafik får oproportionellt stor vikt.

## Slutsatser

Metoden som presenteras för att uppskatta primära linjestörningar gör det möjligt att processa en stor mängd tågpassagedata och bilda empiriska fördelningar av körtidsavvikelser mellan valfria platser och för olika grupper av tåg. Dessa kan därefter anpassas till lämpliga teoretiska fördelningar. Ett antagande som görs är att formen på de primära linjefördelningarna i någon mån liknar den positiva delen i fördelningar för körtidsavvikelser. I fallstudien används exponentialfördelningar, dels för att de beskrivs enkelt av ett fåtal parametrar och dels för att gränssnittet i RailSys har stöd för dessa. En tanke med metoden är även att effektivisera processen för kalibrering och validering genom att det på relativt kort tid går att få fram en mängd olika uppsättningar som kan simuleras och därefter analyseras.

Simuleringarna som görs använder samma reduktionsfaktor för fördelningarnas medelvärden för alla tåggrupper och sträckor. En första utvidgning är att t.ex. använda olika reduktionsfaktorer för resande- och godståg. Dessutom är det av intresse att jämföra skillnader i utfall med olika sträckindelningar. Vidare kan t.ex. jämförelser göras genom att modellera godståg som initieras före sin tidtabell jämfört med att de inte tillåts starta för tidigt [3].

Den utvecklade metoden för att estimeras uppehållsstörningar ur tågpassagedata är ett effektivt alternativ till att använda sig av kostsamma och tidskrävande manuella mätningar. Metoden gör det möjligt att skatta uppehållsförseningar som inte är influerade av sekundärförseningar och tåg som inväntar avgångstid, vilket

är en förutsättning för att de ska kunna användas i RailSys. Genom att använda data från Trafikverkets databas är det möjligt att på ett enkelt sätt analysera t.ex. olika linjer/driftplatser, tågtyper eller tidsperioder. Vidare arbete kan bestå i att validera estimerade uppehållsstörningar genom att utföra simuleringar i RailSys och om önskvärt därefter använda metoden för att skapa ett bibliotek av fördelningar för uppehållsstörningar under olika förutsättningar.

## Referenser

- [1] Sipilä H.: Calibration of simulation model on the Southern Main Line in Sweden. Railway Engineering – 2011, 11th International Conference, 2011.
- [2] Lindfeldt O.; Sipilä H.: Validation of a simulation model for mixed traffic on a Swedish double-track railway line. Railway Engineering – 2009, 10th International Conference, 2009.
- [3] Lindfeldt A.; Sipilä H.: Simulation of freight train operations with departures ahead of schedule. Computers in Railways XIV, WIT Press, 2014.
- [4] Nellidal B-L.; Lindfeldt O.; Sipilä H.; Wolfmaier J.: Förbättrad punktlighet på X2000 – analys med hjälp av simulering. KTH, 2008.
- [5] Lindfeldt A.: Kapacitetsutnyttjande i det svenska järnvägsnätet – Uppdatering och analys av utvecklingen 2008–2012. KTH, 2014.