

Resestatistik för långväga resor baserat på aggregerade och anonymiserade mobilnätsdata

Pilotprojekt inom projektet Nya lösningar för framtida resvaneundersökningar

David Gundlegård och Clas Rydergren, Linköpings universitet
Version 2018-02-05

Dokumentinformation

Titel: Resestatistik för långväga resor baserat på aggregerade och anonymiserade mobilnätsdata

Författare: David Gundlegård, david.gundlegard@liu.se och Clas Rydergren, clas.rydergren@liu.se, Avdelningen för kommunikations- och transportsystem, Institutionen för teknik och naturvetenskap, Linköpings universitet, Campus Norrköping.

Beställare: Trafikanalys. Kontaktperson: Eva Lindborg, 010-4144244, 072-5202460, Eva.Lindborg@trafa.se

Dokumenthistorik:

Version	Datum	Förändring	Distribution
0.7	2017-12-15	Första rapport	Beställare
0.8	2017-12-22	Utkast slutrapport	Beställare
1.0	2018-02-05	Slutrapport	Beställare

Sammanfattning

Denna rapport presenterar resultatet av en pilotstudie om användning av aggregerad och anonymiserad mobilnätdata för att skapa statistik om det långväga resandet i Sverige. Pilotstudien har genomförts inom projektet "Nya lösningar för framtida resvaneundersökningar", som är den del av regeringens samverkansprogram "Nästa generations resor och transporter". Syftet med pilotprojektet är att utföra en första utvärdering av möjligheter och utmaningar med att skatta resestatistik för långväga resor baserat på aggregerad och anonymiserad mobilnätdata.

Telia Sverige har levererat anonymiserad och aggregerad mobilnätdata som beskriver resandet i Sverige för resor över 100 kilometer. Dessa data har i studien jämförts med data från RVU Sverige utifrån antal resor, resornas geografiska spridning, dess fördelning över tid, och dess längd. Jämförelserna av antal resor, resornas spridning och dess fördelning över tid, görs med geografisk upplösning på länsnivå inom Sverige. Jämförelse av resornas längdfördelning görs i form av histogram med specifika längdintervall.

Jämförelserna visar på potential att skapa statistik för långväga resande med hjälp av mobilnätdata med hög tidsupplösning, exempelvis per dag. Den automatiserade processen för beräkning av statistiken gör att aktualiteten och punktligheten blir väldigt hög, och skulle i dagsläget kunna levereras exempelvis på månadsbasis. Resultatet av jämförelserna visar på möjligheten att få rimliga uppskattningar på långväga totalresande och geografisk fördelning. Den främsta utmaningen med att skatta resestatistik baserat på mobilnätdata ser i dagsläget ut att vara kopplad till kunskap om hur beräkning av statistiken sker, både gällande möjlighet att förstå tillförlitligheten i data, exempelvis kunna beräkna konfidensintervall, men även att möjliggöra jämförbarhet över tid med olika mönster i användning av dagens mobilnät, nya typer av nät eller data från olika operatörer.

Abstract

This report presents the results of a pilot study on the use of aggregated and anonymized mobile network data to create statistics about long-distance travel in Sweden. The pilot study has been conducted within the project "Nya lösningar för framtida resvaneundersökningar", which is part of the government's cooperation program "Nästa generations resor och transporter". The purpose of the pilot project is to carry out an initial evaluation of the opportunities and challenges of estimating travel statistics for long-distance travel based on aggregated and anonymized mobile network data.

Telia Sverige has delivered anonymous and aggregated mobile network data describing the travel in Sweden for journeys over 100 kilometers. These data have been compared with data from RVU Sverige based on the number of journeys, the geographic spread of the travel, its distribution over time, and its length. The comparisons of the number of trips, resonance distribution and its distribution over time are done with a geographic resolution of county level within Sweden. Comparison of the travel lengths is made in the form of histograms with specific length ranges.

The comparisons show the potential for creating statistics for long-distance trips with reasonable time-resolution, such as per day. The automated process of calculating the statistics makes the actuality and punctuality very high, at present it could be delivered on a monthly basis. The results from the comparisons shows that reasonable estimations of the number of trips and their distribution. The primary challenge with obtaining travel statistics based on mobile network data is currently linked to knowledge of how calculation of the statistics occurs, both regarding the ability to understand the reliability of data, for example, to calculate confidence intervals, but also to allow comparability over time with different patterns in usage of the network, new developments in the mobile phone network, and data from different mobile phone network providers.

Förord

Denna rapport är resultatet av en pilotstudie inom projektet "Nya lösningar för framtida resvaneundersökningar" inom regeringens samverkansprogram "Nästa generations resor och transporter". Projektet "Nya lösningar för framtida resvaneundersökningar" leds av Trafikanalys. Inom detta projekt genomfördes en intressentanalys och en behovsanalys för att kartlägga verktyg och applikationer för insamling av resvanor under våren 2017. Dessa kartläggningar resulterade i förslag på ett antal pilotstudier och denna rapport är resultat av en av dessa pilotstudier.

I denna pilotstudie analyserar vi vilken möjlighet som finns med att använda aggregerad och anonymiserad mobilnätdata som del av eller komplement till offentlig statistik för långväga resor. Pilotstudien har finansierats av Trafikanalys, och rapporten har författats David Gundlegård och Clas Rydergren vid Linköpings universitet. Telia Sverige har levererat de aggregerade och anonymiserade mobilnätdata som har använts i studien. Trafikanalys har bidragit med data för långväga resor i RVU Sverige, och information runt dessa enkätdata.

Från Trafikanalys har vi under projektets gång haft kontakt med Andreas Holmström, Eva Lindborg och Mats Wiklund. Från Telia Sverige har vi haft kontakt med Petter Bohman, Christian Fredriksson och Madeleine Körling under projektets gång. Urval av diagram och jämförelser, analys och slutsatser har gjorts av författarna.

Norrköping i februari 2018

David Gundlegård och Clas Rydergren

Innehållsförteckning

1. Introduktion.....	1
1.1. Bakgrund.....	1
1.2. Syfte.....	2
1.3. Metod.....	2
1.4. Disposition	3
2. Enkätdata.....	3
2.1. Statistiska egenskaper.....	4
3. Mobilnätsdata.....	4
3.1. Statistiska egenskaper.....	6
4. Dataunderlag	7
4.1. Urval från RVU	7
4.2. Data från mobilnätet.....	11
5. Resultat	12
5.1. Totalresande	12
5.2. Resors spridning mellan län.....	14
5.3. Resors spridning från utvalda län	22
5.4. Resors fördelning över dygn.....	27
5.5. Resors längdfördelning.....	30
6. Analys	37
7. Slutsatser	38
Referenslista.....	39

1. Introduktion

Regeringen presenterade den 1 juni 2016 fem strategiska samverkansprogram. I ett av dessa program, "Nästa generations resor och transporter", har ett behov av att finna och utveckla nya lösningar för resvaneundersökningar identifierats. Inom ramen för detta program leder Trafikanalys projektet "Nya lösningar för framtida resvaneundersökningar". Projektets övergripande syfte är att undersöka och testa nya tekniska lösningar för datainsamling, samt att undersöka hur olika datainsamlingsmetoder skulle kunna användas, enskilt eller i kombination, i framtida resvaneundersökningar, som underlag för offentlig statistik. Under våren 2017 genomfördes inom ramen för projektet en intressent- och behovsanalys för att kartlägga de behov av resvanor som finns, samt en kartläggning av verktyg och applikationer för insamling av resvanor (Trivector, 2017). Resultaten från denna kartläggning har resulterat i ett antal pilotstudier som har genomförts under hösten 2017 och även kommer fortsätta under våren 2018. Denna rapport beskriver resultatet från en pilotstudie där möjligheterna att använda aggregerad och anonymiserad mobilnätsdata som del av eller komplement till offentlig statistik för resor i Sverige analyseras.

Uppgifter om resvanor utgör ett viktigt kunskapsunderlag för utformningen av den nationella och regionala trafikpolitiken. Det finns en generell trend i samhället med sjunkande svarsfrekvenser i stickprovsundersökningar. Även om det inte behöver betyda att resultaten inte längre är representativa för befolkningen i stort, ökar risken för skeva resultat. Undersökningarna är dessutom kostsamma att genomföra. Teknikutvecklingen har lett till nya möjligheter att följa resmönster och det har under en längre tid spekulerats i hur data från mobilnätet skulle kunna komplettera eller ersätta delar av de data som samlas in med resvaneundersökningar. De senaste 10 åren har mycket teknisk forskning inom skattning av mobilitet baserat på data från mobilnätet utförts och lyckade pilotprojekt har på senare tid rapporterats från olika delar av världen (se t.ex. ONS, 2017).

1.1. Bakgrund

Det har länge funnits ett intresse av att skatta resor och mobilitet med nya typer av datakällor. Mycket forskning har behandlat möjligheterna att skatta exempelvis restider, OD-matriser, och rörelsemönster från mobilnätsdata, se exempelvis Gundlegård och Karlsson (2009), Calabrese m.fl. (2011) och González m.fl. (2008).

Den första utmaningen med att skatta resmönster från mobilnätsdata innefattar att förstå vilken typ av data som är tillgänglig i mobilnätet för detta syfte. Tidigare arbete som gjorts för att identifiera vilken typ av data som är tillgänglig, och hur den kan användas för att estimeras resor finns beskrivet i exempelvis Allström m.fl. (2015). I samma källa finns också en beskrivning av möjlig användning av mobilnätsdata för prognosmodeller med fokus på svenska förhållanden. Processer för att skapa beskrivningar av efterfrågan på resor, och hur resor kan identifieras utifrån mobilnätsdata finns i publikationerna Gundlegård m.fl. (2016) respektive Breyer m.fl. (2017) vilka är skapade inom ramen för de Vinnova-finansierade forskningsprojekten "Mobile network Origin Destination matrix Estimation" (MODE, 2013) samt fortsättningsprojektet "Mobilnätsdata för Framtidens Transportsystem" (MOFT, 2017) som pågår 2017-2020.

Mobilnätsdata skiljer sig markant som mätmetod jämfört med dagens resvaneundersökningar. Grunden i resvaneundersökningen är att fråga ett statistiskt kontrollerat urval mycket detaljer om resandet för en specifik dag, eller som i fallet med långväga resor om resandet under en månads tid, medan mobilnätsdata samlar in observationer om förflyttningar för en stor andel av Sveriges befolkning, i dagsläget utan exakt kunskap om urvalets sammansättning, kontinuerligt över en längre tid. I resvaneundersökningen ingår c:a 12-14 000 noggrant utvalda personer varje år i detalj om deras resebeteende (se exempelvis Trafikanalys, 2017b). Antalet varierar mellan åren beroende på hur stora extraurval som ingått. Svarsfrekvensen har de flesta åren efter 2011 legat runt 40%, men runt 30% det senaste året. Gällande mobilnätsdata så finns det enligt Trafikanalys (2017a) c:a 14.6 miljoner mobilabonnemang i Sverige. Hypotesen är att mobilnätsdata kan användas för att ersätta visa delar av resvaneundersökningen samt att de två mätmetoderna har god potential att komplettera varandra.

1.2. Syfte

Syftet med denna studie är att utföra en första utvärdering av möjligheter och utmaningar med att skatta resestatistik för långväga resor baserat på aggregerad och anonymiserad mobilnätsdata. I denna studie kommer följande frågor att besvaras:

- Hur ska enkätdata från RVU Sverige jämföras med statistik från mobilnätsdata gällande långväga resor för att tydliggöra skillnader och likheter i de olika mätmetoderna?
- Hur förhåller sig statistik om långväga resor baserad på mobilnätsdata till de statistiska kvalitetskriterier som är framtagna för att utvärdera officiell statistik i Sverige?
- Hur ska resultaten från jämförelsen utvärderas och tolkas i termer av potential att på sikt använda mobilnätsdata för att ersätta delar av eller komplettera data från RVU Sverige?
- Vad är potentialen att i nuläget ersätta delar av eller komplettera data från RVU Sverige med aggregerad och anonymiserad mobilnätsdata?
- Vilka delar av RVU Sverige är i nuläget inte lämpliga att ersätta med aggregerad och anonymiserad mobilnätsdata?

1.3. Metod

Telia Sverige har levererat anonymiserad och aggregerad mobilnätsdata till Linköpings universitet (LiU) enligt en överenskommen specifikation (se Kapitel 4). Trafikanalys har tillhandahållit data för långväga resande från RVU Sverige. LiU har genomfört en jämförelse av dessa data utifrån antal resor, resornas geografiska spridning, dess fördelning över tid, och dess längd.

Jämförelsen har baserats på två månaders aggregerad och anonymiserad data från mobilnätet och på ett urval av data från RVU Sverige. Data från mobilnätet är inhämtad för september och oktober 2017. Data för långväga resor i RVU Sverige är tillgänglig för 2011 till 2016, och i denna pilotstudie har data från åren 2013, 2014 och 2016 använts. Jämförelsen är till stor del baserad på kartbilder, diagram och histogram skapade från de två datakällorna. Från dessa har

slutsatser dragits om, bland annat, likheter och skillnader i det totala resandet i Sverige, reandet mellan länen i Sverige, dess spridning över tid, och längden på resor.

1.4. Disposition

Rapporten har följande struktur. Kapitel 2 beskriver de enkätdata, från RVU Sverige, som använts vid jämförelsen. Kapitel 3 beskriver hur information från ett mobilnät kan användas för att skapa information om resor, samt beskriver vilka förväntningar man bör ha när det gäller kvalitet på denna information. I Kapitel 4 beskrivs det urval som gjorts från RVU Sverige och de aggregerade och anonymiserade mobilnätsdata som använts. I Kapitel 5 redovisas jämförelseunderlagen i form av kartbilder och diagram. I Kapitel 6 görs en analys av de likheter och skillnader som har hittats och i Kapitel 7 redovisas slutsatserna från studien.

2. Enkätdata

Under åren 2011-2016 har det årligen genomförts en nationell resvaneundersökning (RVU) som, baserat på statistiskt säkerställt urval, innehåller information om hur ett antal individer gör sina resor. Från dessa enkätsvar finns framtagen statistik på exempelvis färdssättsval, reslängder i genomsnitt från olika regioner, genomsnittligt antal resor per person, och liknande, som ger en god bild av hur resor genomförs.

För resor längre än 100 km enkel väg utförs kompletterande frågor i RVU som behandlar dessa. Då långväga resor utförs mer sällan så har dessa frågor en längre referensperiod, en månad för resor längre än 100 km enkel väg och ytterligare en månad för resor längre än 300 km, jämfört med mättdagsresor som bara samlas in för en specifik dag. Även för långväga resor samlas uppgifter om till exempel färdssättsval, reslängder, restider och antal resor in.

I RVU sammanställs statistik för både långväga huvudresor och långväga delresor. Långväga huvudresor startar och slutar i en huvudresepunkt, vilken utgörs av respondentens folkbokförda bostad, annan permanent bostad eller fritidsbostad. Långväga delresor börjar och slutar i en delresepunkt, vilket är definierat som en plats där respondenten har ett mycket viktigt ärende. Ärendet ska vara så viktigt att det är styrande för huvudresans resväg och vanligtvis finns bara en delresepunkt på en långväga huvudresa, nämligen målpunkten för resan. Långväga resor innehåller också data om resor som passerar landsgränsen och resan kan då vara kortare än 100 km om den startar närmare än 100 km från landsgränsen.

En långväga huvudresa som startar och slutar i samma adress och har minst en målpunkt på resan kallas tur och returresa och de måste vara minst 200 kilometer för att räknas som en 100-kilometers huvudresa. Långväga huvudresor som startar och slutar i olika adresser, exempelvis resor mellan huvudbostad och fritidsbostad, kallas för enkelresa och dessa måste uppgå till minst 100 km för att utgöra en långväga resa. Långväga delresor kan vara hur korta som helst, men i och med att den långväga delresans ärende måste vara mycket viktigt och påverka huvudresans resväg så är i de flesta fall delresorna lika långa som de långväga enkelresorna eller hälften så långa som de långväga tur och returresorna, dvs minst 100 km. En mer utförlig beskrivning av definitionen av långväga resor finns i Trafikanalys (2012).

Då definitionen av långväga delresor stämmer bättre överens med den definition av resor som görs baserat på mobilnätdata som använts i denna studie (se kapitel 3), har fokus varit på analys av långväga delresor i kontrast till långväga huvudresor.

2.1. Statistiska egenskaper

Kvaliteten för den statistik som återfinns i RVU Sverige – den nationella resvaneundersökningen, beskrivs närmare i Trafikanalys (2017b).

3. Mobilnätdata

Detta avsnitt syftar till att översiktligt beskriva den data som är tillgänglig i mobilnät i syfte att skatta resestatistik. Då informationen om egenskaper för Telias underliggande data och processerna för att bearbeta data är en affärshemlighet kommer detta avsnitt inte beskriva egenskaperna för de data som används i utvärderingen specifikt, utan beskrivningen är en generell beskrivning av vilka data som finns tillgängliga i ett mobilnät.

Då en stor andel av Sveriges befolkning använder sig av en mobiltelefon och denna telefon ofta tas med vid resor är det möjligt att använda mobiltelefoners rörelser som approximation för personers rörelser. Denna möjlighet gör att man på ett effektivt sätt kan utnyttja mobiloperatörernas omfattande infrastruktur för att samla in data om resor.

Mobilnät är uppbyggda som yttäckande radionät där kommunikationen i avgränsade geografiska områden hanteras av en basstation. Varje basstation består av en eller flera antenner som kan vara rundstrålande eller riktade. Täckningsområdet för varje antenn beror främst på antennens egenskaper och placering, exempelvis frekvens, utsänd effekt, höjd över marken och riktning. Även andra faktorer påverkar dock täckningsområdet, exempelvis mängden trafik i olika delar av nätet, väder och förändringar i utbredningsmiljön. Täckningsområdet tillsammans med algoritmer för radioresurshantering i mobilnätet samt i mobiltelefonen kommer sedan avgöra vilken antenn som används av en specifik telefon vid varje tidpunkt.

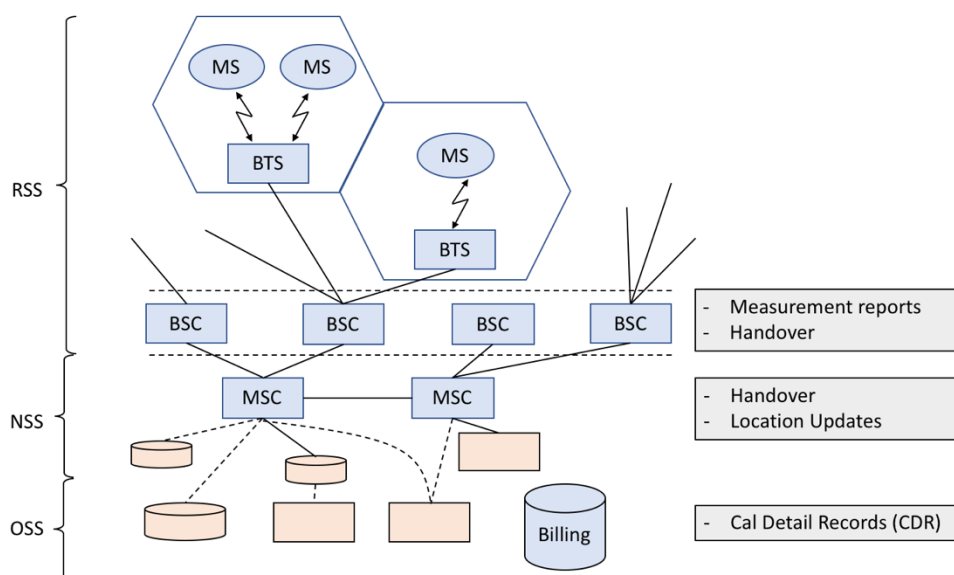
Det finns två huvudsakliga anledningar till att mobilnätet håller reda på rörelser av mobiltelefoner i mobilnätet, det ena är att möjliggöra kontinuerlig uppkoppling trots att en mobiltelefon rör sig mellan täckningsområdet för olika antenner och det andra är att möjliggöra uppkoppling av inkommande samtal eller datasessioner, dvs att mobiltelefonen är nåbar trots att den rör sig mellan täckningsområdet för olika antenner. Beroende på typen av mobilnät, typen av mobiltelefon samt hur mobiltelefonen används kommer olika information finnas tillgänglig om mobiltelefonens lokalisering i mobilnätet.

I olika delar av mobilnätet finns stora mängder information om mobiltelefonens aktiviteter, men vilken information som används för att skatta mobilitet beror på mobiloperatörens plattform för att hantera signaleringsdata och göra om den till skattningar av mobiltelefonernas position över tid. Frekvensen på hur ofta signaleringsdata skickas från mobiltelefoner kan variera från millisekund-nivå till flertal timmar beroende på hur telefonen används och vilken typ av mobilnät den är uppkopplad mot. Man kan förvänta sig att en modern telefon i rörelse genererar information som är användbar med några minuters

mellanrum, men beroende på vilken data som används internt av operatören kan samplingstiden också uppgå till ett antal tiotal minuter mellan varje uppdatering.

Den rumsliga upplösningen för varje mätning av mobiltelefonens position beror på var telefonen befinner sig samt vilken typ av signaleringsdata som används. Ofta används ett uppskattat täckningsområde för varje antenn som skattning av var telefonen befinner sig och täckningsområdet för en antenn kan variera från några hundra meter i radie i tätbefolkade områden till tiotals kilometer i radie i glesbefolkade områden.

Figur 1 visar en översiktsbild över ett mobilnät och några exempel på data som är tillgänglig i olika delar i nätet. I centrala databaser (del av Operations and Support Subsystem, OSS) finns information lagrad för fakturering och dessa innehåller vanligtvis information om vilken antenn som använts vid samtal, SMS och i vissa fall även vid datauppkoppling (eng. Call Data Records, CDR). I kärnnätet (Network Switching Subsystem, NSS) finns information om uppdateringar av mobiltelefonen även vid byte av större geografiskt område samt vid ett specifikt tidsintervall (eng. location updates), exempelvis en gång per timme samt även om byte av antenn har skett under pågående samtal (eng. handover). I radionätet (Radio Subsystem, RSS) finns information om vilken cell som mobiltelefonen är uppkopplad mot samt även signalstyrka på nuvarande och närliggande celler (eng. measurement reports). Denna typ av mättrapporter kan skapas flera gånger per sekund för mobiltelefoner med pågående samtal; denna information används i dagsläget dock sällan för storskaliga mobilitetsanalyser.



Figur 1: Översiktsbild över data som genereras i mobilnätet. Här står MS för mobiltelefon, BTS för basstation, BSC för basstationsstyrenhet, MSC för central växlingsenhet och Billing den databas där CDR lagras.

För att ta fram resestatistik och mobilitetsmönster görs i ett första steg signalerings-informationen om till en sekvens av skattade områden där en mobiltelefon befunnit sig, vilken i nästa steg ska göras om till resor. Till skillnad mot RVU-data finns ingen koppling till ärende för resor, men däremot antas ofta någon form av aktivitet eller ärende utföras då mobiltelefonen inte rört sig under en viss tid och resor definieras som förflyttningar mellan

dessas aktiviteter. Resedetektering baserat på mobilnätdata kan göras på flera olika sätt och exakt vilken metod som används kommer också påverka slutresultatet i form av hur många samt vilken typ av resor som skattas (se Breyer m.fl., 2017).

Resor från mobilnätdata motsvarar i första hand ett urval av populationen baserat på alla som använder den specifika operatören. Därefter kan också operatören välja att filtrera bort vissa typer av abonnemang som inte ska användas för beräkningen av statistik, det kan vara att exempelvis abonnemang som är tydligt kopplade till maskin-till-maskin-kommunikation eller andra abonnemang som av någon anledning inte bör vara med i analysen. För att statistiken ska vara representativ för hela populationen behöver en uppskalning av de resor som uppmätts utföras. Vanligtvis används operatörens marknadsandel i olika geografiska områden för att skala upp resandet. Om mer information om varje abonnent finns tillgänglig, exempelvis kön eller ålder, kan denna information användas för att göra bättre skattningar av marknadsandelen inom dessa kategorier och även kontrollerade urval kan göras baserat på denna information. Uppskalningen innebär en osäkerhetsfaktor i resultaten, speciellt om ingen information är tillgänglig om abonnenterna, men den stora fördelen med mobilnätdata jämfört med RVU-data är att uppskalningsfaktorerna är betydligt mindre samt att mätdata finns mer eller mindre kontinuerligt tillgängliga över tid.

3.1. Statistiska egenskaper

Detta avsnitt diskuterar översiktligt de statistiska kvalitetskriterier som behöver bedömas för att mobilnätdata ska kunna fungera som officiell statistik för långväga resor.

Relevans

Målpopulationen för de långväga resor som estimerats från mobilnätdata består av alla resor om minst 100 kilometer (enkelresa) genomförda av samtliga personer som är folkbokförda i Sverige. Variabler är reslängd, start- och slutområde samt starttid och sluttid för långväga resor. Data från mobilnätet ges normalt i aggregerad form och presenteras för respektive variabel som antal, summanvärden och medelvärden. Både strukturen hos mobilnätet och hur aggregeringen genomförs påverkar den geografiska upplösningen för start- och slutområdet för resan.

Tillförlitlighet

En osäkerhetskälla i mobilnätdata är urvalet och hur uppskalningen gör för att matcha målpopulationen. Om ingen information delges om abonnenterna eller urvalet från operatören så minskar det tillförlitligheten för uppskalningen till målpopulationen. Å andra sidan är antalet abonnenter väldigt stort, så uppskalningsfaktorerna blir relativt små. Ytterligare faktorer som påverkar tillförlitligheten är eventuella abonnemang som är med i datamängden men som inte används aktivt och att personer kan ha fler än ett abonnemang med sig. Mätfel kan uppstå i processen för resegenerering i form av att resor startar eller delas upp på fel ställe. Även fel i start- och slutpunkt kan uppstå på grund av den relativt grova rumsliga upplösningen, speciellt i gleset befolkade områden. Vidare kan fel i referensdata för antenner finnas, vilket också kan påverka resultatet. Algoritmer för att bestämma reslängd utifrån mobilnätdata är komplexa och hur dessa implementeras påverkar tillförlitligheten. För att beräkna statistiken utförs automatisk bearbetning av stora mängder data och fel kan uppstå i mjukvaran som orsakar systematiska fel i statistiken.

Aktualitet och punktlighet

Statistik från mobilnätdata genereras automatiskt och kan, teoretiskt, tas fram från och med att processen att hantera data startas. Sammanställning av statistik kan ske snabbt, och kan sammanställas kontinuerligt, t.ex. från månad till månad. Tillgång till historisk data är mycket begränsad.

Tillgänglighet och tydlighet

Skapandet av statistik från mobilnätdata har här baserats på att data aggregeras och anonymiseras hos mobilnätoperatören. Tillgängligheten är därför beroende av den operatör som förväntas leverera data. Vilken typ av statistik som kommer vara aktuell att beräkna i framtiden kommer behöva utredas i detalj när förståelsen för den data som är tillgänglig har blivit bättre.

Jämförbarhet och sam användbarhet

För att åstadkomma jämförbarhet över tid behöver samma typ av rådata processas. Vid större förändringar i teknologin för mobiltelefoni eller vid justeringar av de algoritmer som används för hanteringen av data så behöver validering genomföras för att bibehålla jämförbarheten över tid. Hög tidsupplösning, och relativt god rumslig upplösning, gör att resestatistik från mobilnätdata blir relativt enkel att använda tillsammans med andra typer av data. Samanvändbarheten begränsas dock av att data anonymiseras och aggregeras.

4. Dataunderlag

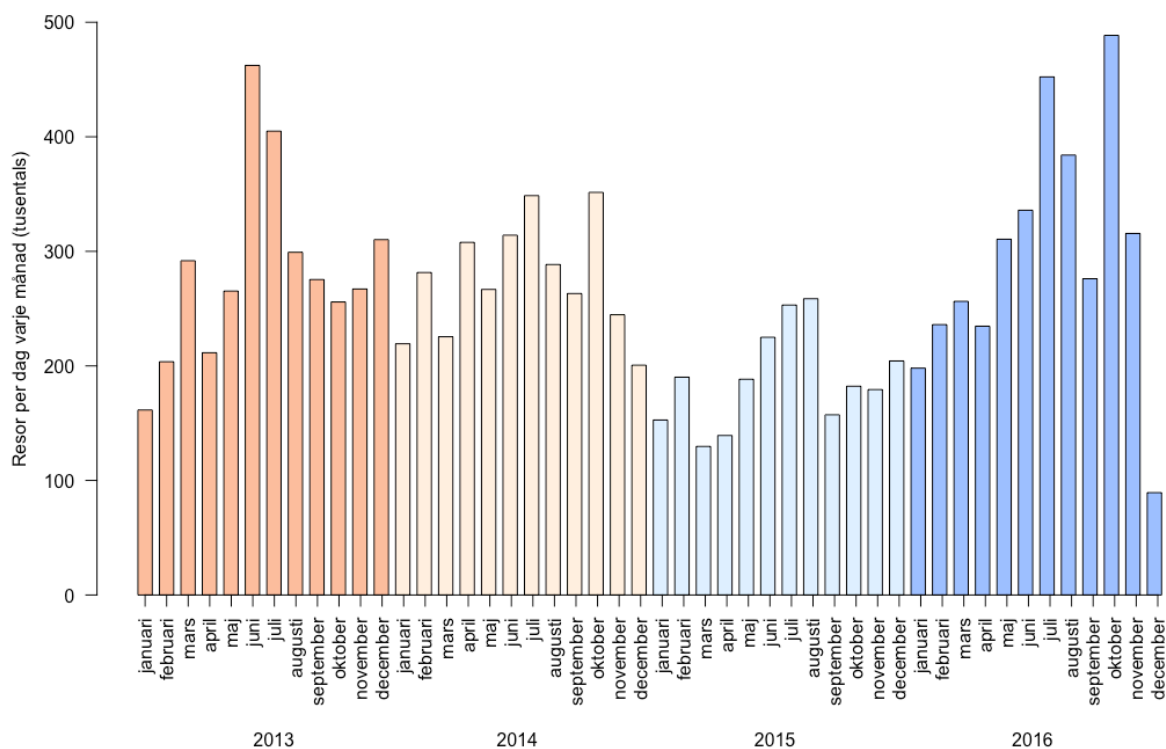
I detta avsnitt beskrivs vilken typ av data som är levererad av Telia Sverige samt vilken data från RVU Sverige som har använts.

4.1. Urval från RVU

I detta avsnitt görs en inledande analys av storleksordning på det genomsnittliga antalet resor per månad som finns i RVU Sverige åren 2013 till 2016, och jämförs med antal resor identifierade från mobilnätdata för september och oktober 2017. Då variationen mellan år, och även mellan månader i RVU Sverige är stor, så kan endast begränsade slutsatser dras av jämförelsen med två månader (september och oktober) av data från mobilnätet. Alla data presenteras som genomsnittligt antal (i tusental) resor per dygn, över alla dygn i månaden. Vi separerar alltså inte helgdagar och veckodagar i denna del av analysen.

Enkätdata ger inte alltid fullständig information om de genomförda resorna. För vissa resor så saknas information om startpunkt eller slutpunkt, starttid och/eller resans längd. Då analys av bl.a. fördelning över tid och mellan län kräver denna information, så måste svar som saknar tidsangivelse, startlän eller slutlän för resan utelämnas från analysen. Effekten av detta diskuteras i slutet av detta delkapitel.

Figur 2 visar genomsnittligt antal resor per dag i respektive månad för åren 2013 till och med 2016 från RVU Sverige. Som synes i figuren så är antalet resor per år för 2015 lägre än övriga år.



Figur 2: Genomsnittligt antal resor per dag för årets månader under 2013, 2014, 2015 och 2016 från RVU Sverige.

Tittar vi på genomsnitt för respektive år så får vi de data som presenteras i Tabell 1.

Tabell 1: Genomsnittligt antal resor per dag för år 2013, 2014, 2015 och 2016.

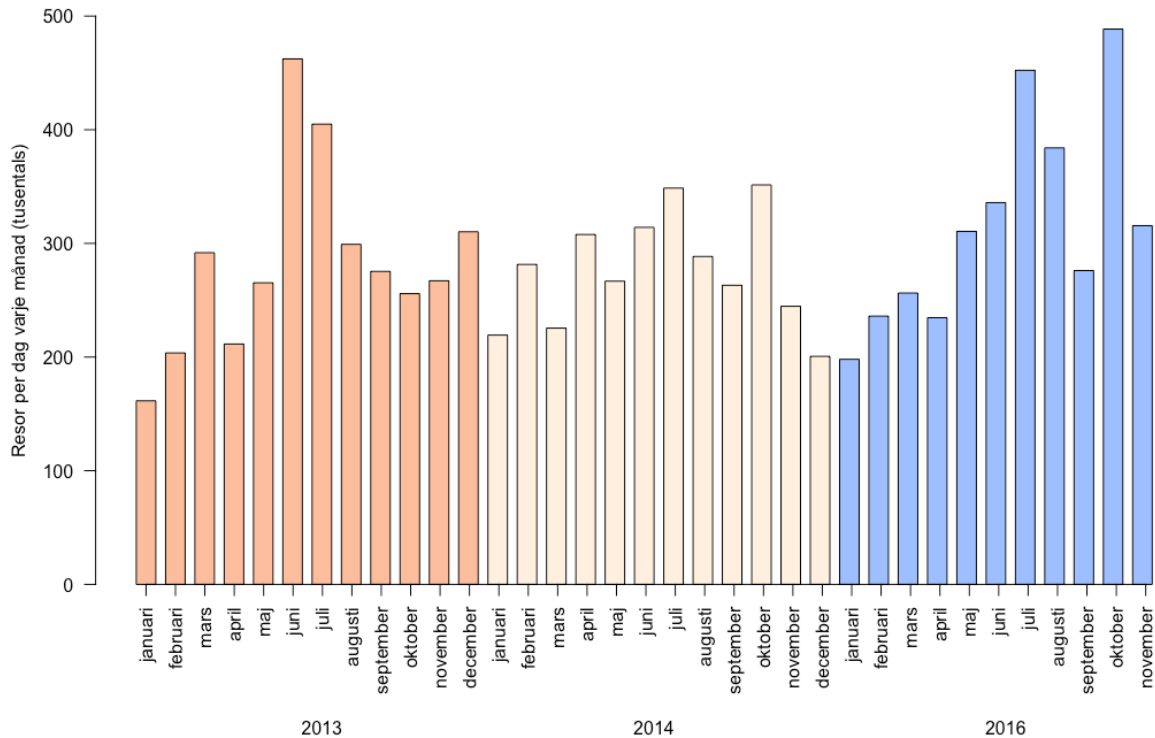
År	Resor per dag (tusentals)
2013	284.1
2014	276.0
2015	188.3
2016	298.1

På grund av det avvikande antalet resor för år 2015 så utlämnas data för hela 2015 från vidare analys. December 2016 innehåller heller ej komplett data ty eftersom information om långväga resor samlas in för den 30 dagarsperiod som föregår mät dagen så saknas svar från personer som har mät dag efter 31 december 2016. På grund av detta så utlämnas även december 2016 från vidare analys. Det ger det genomsnittliga antalet resor per månad enligt Tabell 2.

Tabell 2: Genomsnittligt antal resor per dag för år 2013, 2014 och 2016 (jan-nov).

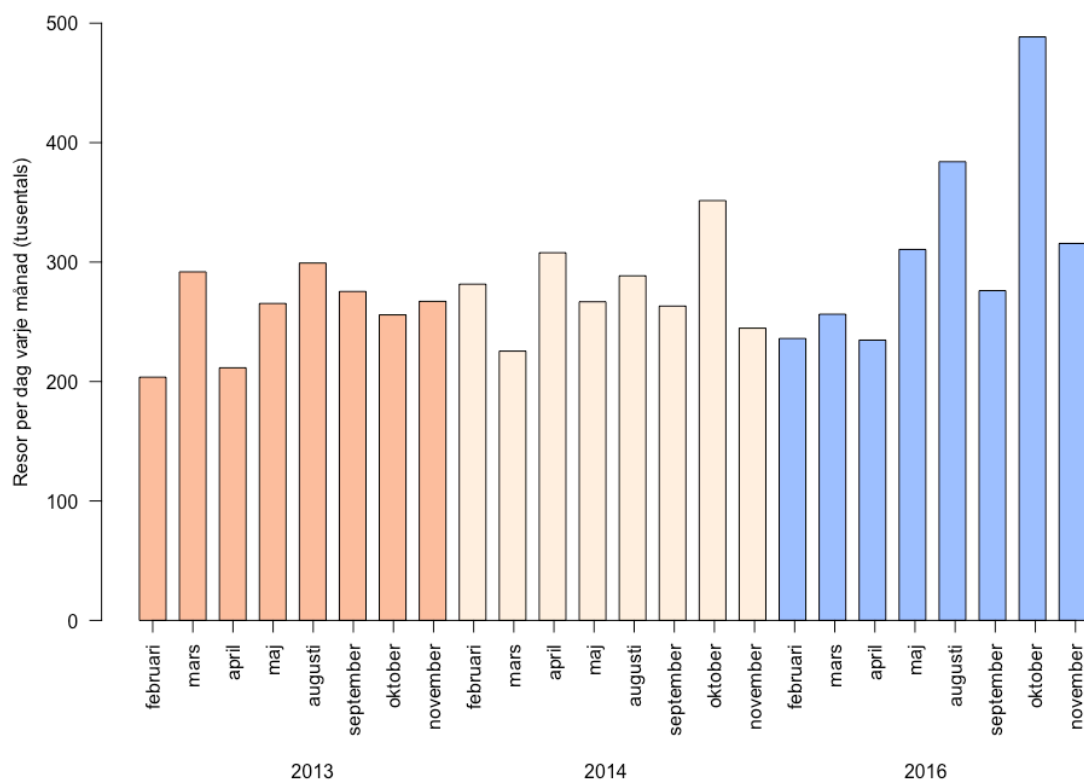
År	Resor per dag (tusentals)
2013	284.1
2014	276.0

Resor 2013, 2014 och januari till november 2016 ser ut enligt diagrammet i Figur 3.



Figur 3: Genomsnittligt antal resor per dag i årets månader 2013, 2014 och 2016 (jan-nov) från RVU.

Då vi vill jämföra med resor uppmätta under september och oktober 2017, så vill vi identifiera månader som historiskt inte har stor avvikelse från september och oktober. För år 2013 är det tydligt att juni och juli har ett avvikande resande. Även januari ser ut att ha mindre resande än genomsnittet. Detta gör att vi eliminerar januari, juni, juli alla år, och för symmetriens skull, även december 2013 och december 2014. Resor per dag under de kvarvarande åtta månaderna åren 2013, 2014 och 2016 illustreras i Figur 4.



Figur 4: Genomsnittligt antal resor per dag för 8 utvalda månader 2013, 2014 och 2016.

Genomsnittligt antal resande under dessa 8 månader, under respektive år, ges i Tabell 3.

Tabell 3: Antal resande per dag baserat på utvalda månader 2013, 2014 och 2016.

År	Resor per dag (tusentals)
2013	258.7
2014	278.6
2016	312.7

I ovanstående analys har en filtrering av data i RVU Sverige utförts där resor kortare än 100 kilometer, utlandsresor och enkätsvar som saknar startlän eller slutlän utelämnats. Resultatet av filtreringen är att för de 24 ovan utpekade månaderna under 2013, 2014 och 2016 så går antal resor från i genomsnitt 365 000 resor per dag, till 283 300 resor per dag. Av de 365 000 resorna så finns 325 000 resor per dag kvar när kravet att reslängden är mer än 100 kilometer, vilket också ställer krav på att en reslängd finns i datamaterialet för resan. Adderar vi dessutom krav på att resan ska ha en geografisk angivelse för startpunkten (län) och en geografisk angivelse för destinationspunkt, samt att det ska finnas ett datum och (ev. approximativt) klockslag då resan påbörjades så sker 283 300 resor per dag, i genomsnitt. Resor med destination utomlands är bland de resor som saknar geografisk angivelse för destinationspunkt. Av de 365 000 resor per dag som finns i datamaterialet för de 24 månaderna under 2013, 2014 och 2016 så är 37 800 av dessa, i genomsnitt, en resa med startpunkt eller destination utomlands. Jämförelsen mellan RVU Sverige och mobilnätdata baseras på det dataunderlag med 283 300 resor per dag, i genomsnitt.

För det genomsnittliga resandet per dag under de 24 månaderna 2013, 2014 och 2016, så har Trafikanalys tagit fram konfidensintervall för skattningen på 365 000 resor per dag. Konfidensintervallet, som är framtaget för 95% konfidens, är i storleksordningen plus minus 20 000 resor per dag. Motsvarande konfidensintervall, om man gör ett urval för 2013 specifikt så är intervallet plus minus 25 000, för 2014, plus minus 30 000 och för 2016, plus minus 35 000 resor per dag, i genomsnitt. Begränsar man urvalet från RVU Sverige till enskilda månader så ökar konfidensintervallets storlek. Baserat enbart på data för september 2016, vilken har c:a 350 000 resor per dag i genomsnitt, så är motsvarande konfidensintervall plus minus 60 000. Motsvarande för oktober 2016 är plus minus 170 000, för ett genomsnittligt resande på 545 000 resor per dag. Begränsas urvalet mer till en specifik dag så ökar konfidensintervallets storlek ytterligare. Per den 5:e september 2016, när estimatet från RVU är 401 000 resor, så är konfidensintervallet plus minus 226 000 resor. Den osäkerhet som fångas in av konfidensintervallen är den som beror på att det är en urvalsundersökning och eventuella mätfel. Eventuella systematiska fel fångas inte in i dessa konfidensintervall.

4.2. Data från mobilnätet

Följande data är levererad av Telia Sverige och används för jämförelse mellan mobilnätsdata och RVU Sverige:

1. Startpunkt och målpunkt (OD-matris) för alla resor längre än 100 km mellan alla par av län.
 - a. OD-matris på dygnsnivå under september och oktober 2017.
 - b. OD-matris för genomsnittlig måndag-torsdag, fredag och lördag-söndag, uppdelade på timnivå
2. Reslängdsfördelning för resor över 100 km inom Sverige.
 - a. Fördelningen av alla resornas längd i kilometer för genomsnittliga måndag-torsdag, fredag, lördag-söndag.
 - b. Fördelning av längden på alla resor mellan par av län.

Data för punkt 1a har givits i form av matriser av storlek 21x21 län, en för varje dygn för de två månaderna. Data för 1b är tre gånger 24 matriser av storlek 21x21 län. Data för 2a ges i form av antal resor i intervallen 100-120 km, 120-140 km, etc., för varje dag-kategori. Data för 2b ges på samma form som för 2a.

Då informationen om egenskaper för Telia Sveriges underliggande data och processerna för att bearbeta data är en affärshemlighet i Telia Crown Insights (Telia Sverige, 2017) kan inte information om detta redovisas eller studeras i denna pilotstudie. Från analys av data av ovanstående typ så har vi identifierat att t.ex. antal resande per dygn i genomsnitt, skiljer sig något mellan data enligt 1a och 1b ovan. Denna skillnad beror på den anonymiserings- och aggregeringsprocess som data går igenom. Om man ökar granulariteten i tid eller rum så innebär det att fler resor blir så pass unika att de tas bort i anonymiseringsprocessen. Datamängd 1b innehåller högre temporal granularitet eftersom man där har både och timme och veckodag, jämfört med datamängd 1a där det endast finns individuella datum. För att få en känsla för storleksordningen om underliggande data så kan vi notera att enligt Trafikanalys (2017a) fanns det c:a 14.6 miljoner mobilabonnemang i Sverige 2016 och Telia Sverige uppskattas ha en marknadsandel på c:a 36% (PTS, 2017).

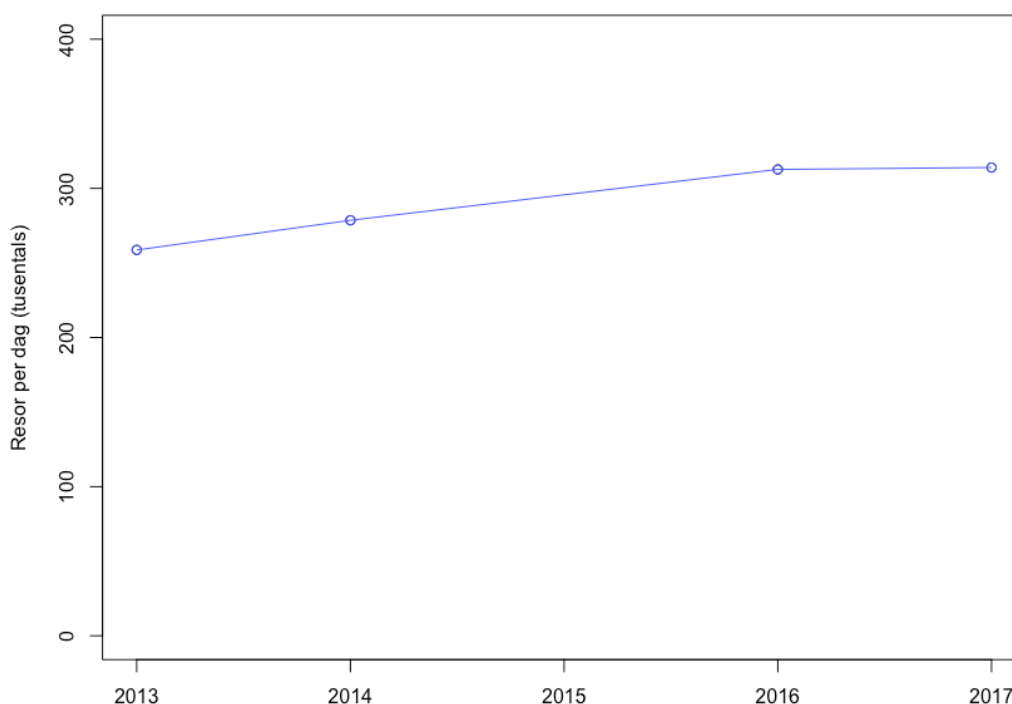
5. Resultat

I detta kapitel så redovisar vi resultatet av jämförelsen mellan urvalet av långväga resor från RVU Sverige och de långväga resorna skattade från aggregerad och anonymiserad mobilnätdata.

5.1. Totalresande

Från mobilnätdata så uppskattas det totala antalet resor under september 2017 till 326 000 resor per dag. För oktober är motsvarande siffra 301 000 resor. Det är ett något högre antal resor än vad som är skattat från RVU Sverige för 24 månader 2013, 2014 och 2016, vilket är 283 300 resor per dag i genomsnitt.

Tittar man på genomsnittliga antalet resor per dag under 8 månaderna för 2013, 2014 respektive 2016 (från Tabell 3), och jämfört med september och oktober 2017 från mobilnätdata fås grafen i Figur 5.

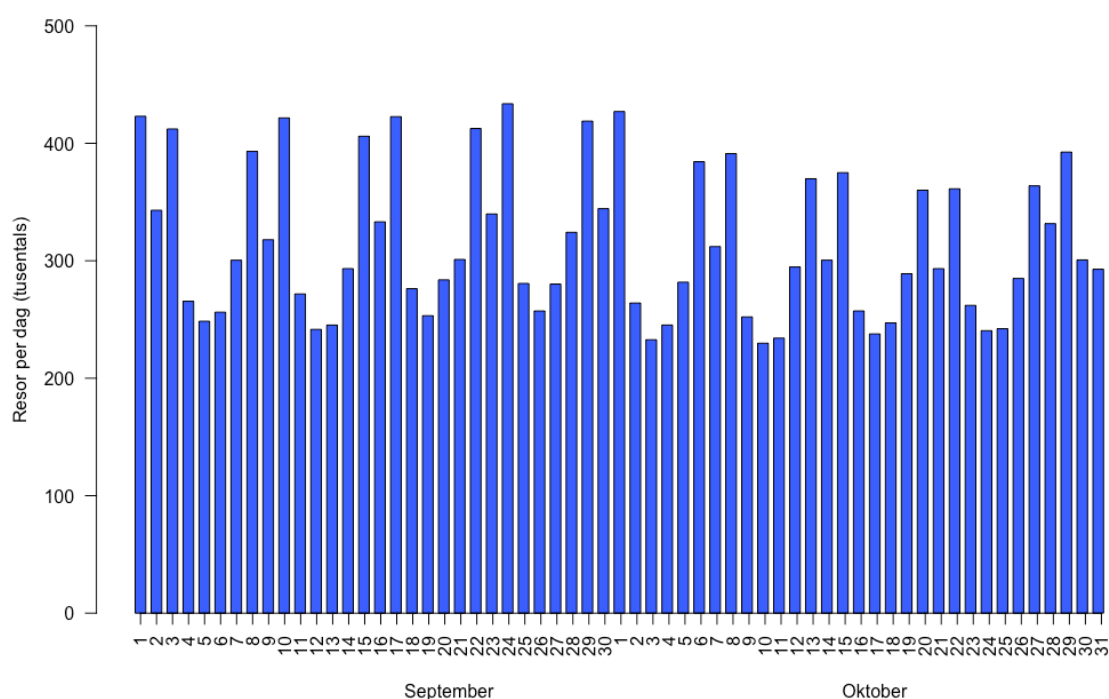


Figur 5: Genomsnittligt antal resor per dag (tusentals) under åren 2013, 2014 och 2016 från RVU och för september och oktober 2017 från mobilnätdata.

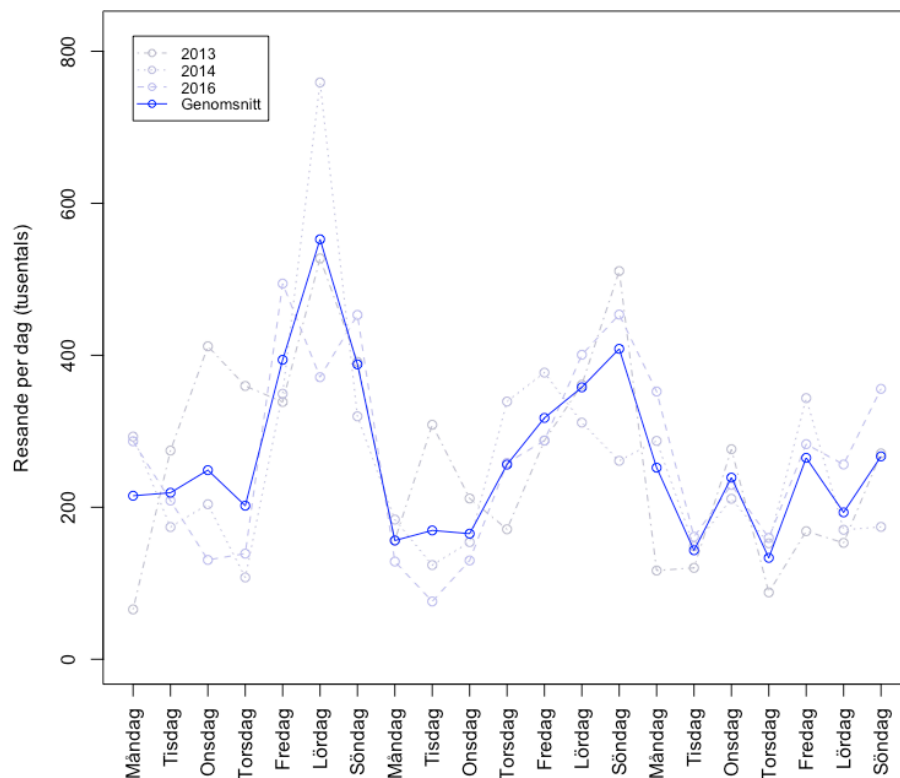
Vid jämförelser där data från 8 månader under 2013, och 2014 och 2016 används för att identifiera ett genomsnittligt resande, så är kvoten mellan RVU-data och mobilnätdata 283 000/314 000, dvs RVU under denna period innehåller 90.2% av det genomsnittliga antalet resor som fångats i mobilnätdata för september och oktober 2017. I de utvalda RVU-data ser det ut att vara en trend med en årlig ökning på knappt 20 000 resor per år, om denna

används för att skriva upp resultatet för 2017 så får vi medelantalet resor per dag till c:a 331 000 resor. Motsvarande kvot mellan RVU-data och mobilnätsdata blir då 105.4%.

Antalet resor per dag för september och oktober 2017 från mobilnätsdata illustreras i Figur 6. I figuren ser vi att det finns systematiskt fler resor under fredagar och söndagar. Tittar vi på antal resor per dag estimerat från RVU så är resultatet mer svårtolkat. I Figur 7 återfinns antalet resor per veckodag baserat på RVU-data från september 2013, september 2014 och september 2016. De ljusare kurvorna visar respektive år, och den heldragna mörkare blåa kurvan visar medelvärdet. Här kan ingen systematik ses när det gäller t.ex. fredagar och söndagar. Vår hypotes är att de mer slumpmässiga variationerna i RVU-data beror av att urvalet är för litet för att göra denna typ av analys. Detta gäller variationen mellan månadens alla dagar; om variationen mellan veckodagar hade analyserats så eventuellt ett systematiskt mönster kunnat identifieras.



Figur 6: Antal resande per dag (tusentals) under september och oktober 2017 skattat från mobilnätsdata. Fredagar och söndagar har systematiskt högre antal resor.

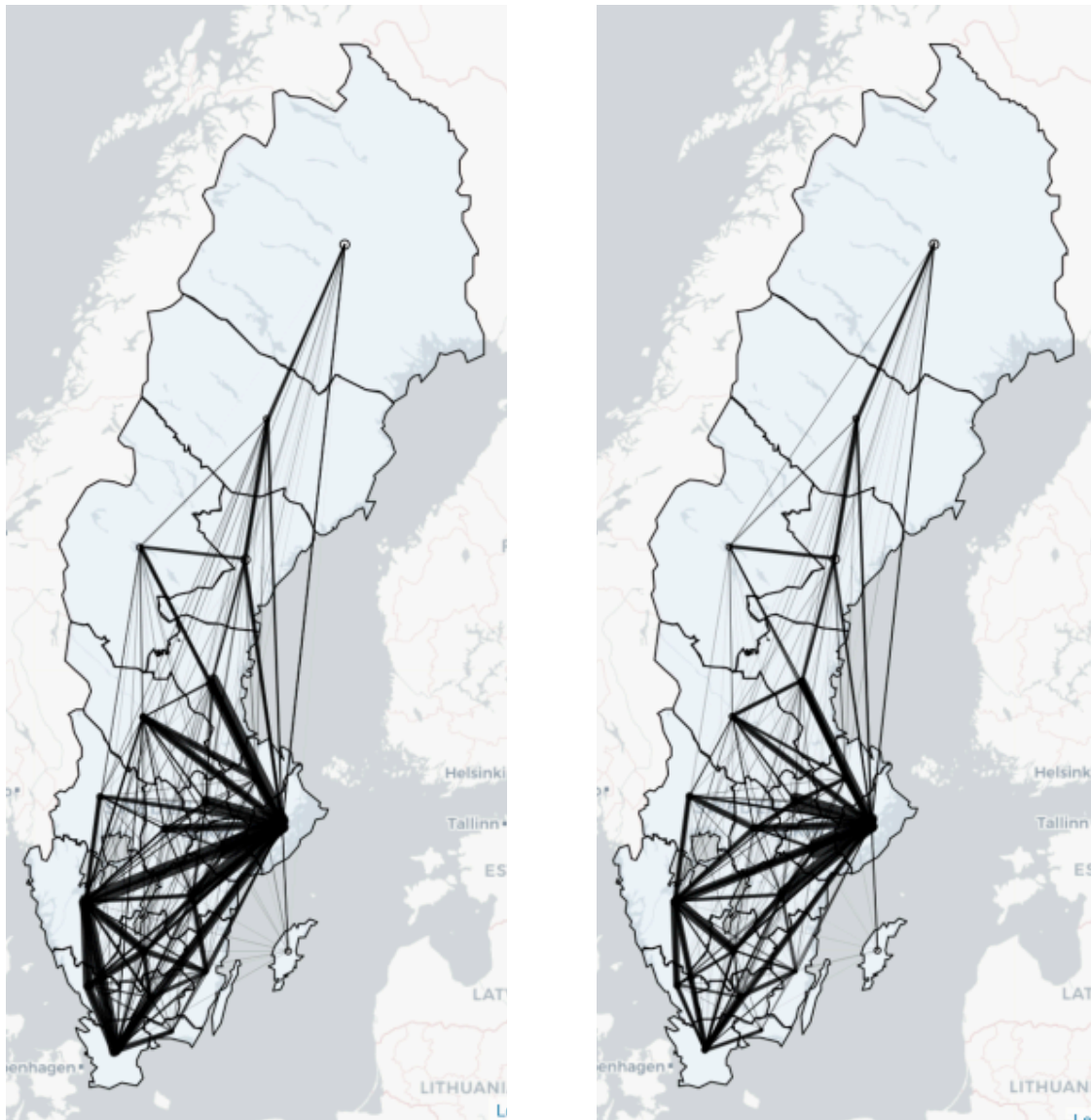


Figur 7: Genomsnittligt antal resor per dag från första måndagen i september 2013, 2014 och 2016 till sista söndagen.

5.2. Resors spridning mellan län

Denna sektion baseras på data från RVU (8 månader 2013, 8 månader 2014 och 8 månader 2016) och mobilnätdata (september och oktober 2017) i data-mängden 1a (se Kapitel 4.2.). Data från RVU är genomsnitt beräknade från hela datamängden. Data från mobilnätet har aggregerats till en genomsnittlig dag för varje par av län genom att summera antal resor för alla dagar i september och oktober och dividerat med antalet dagar.

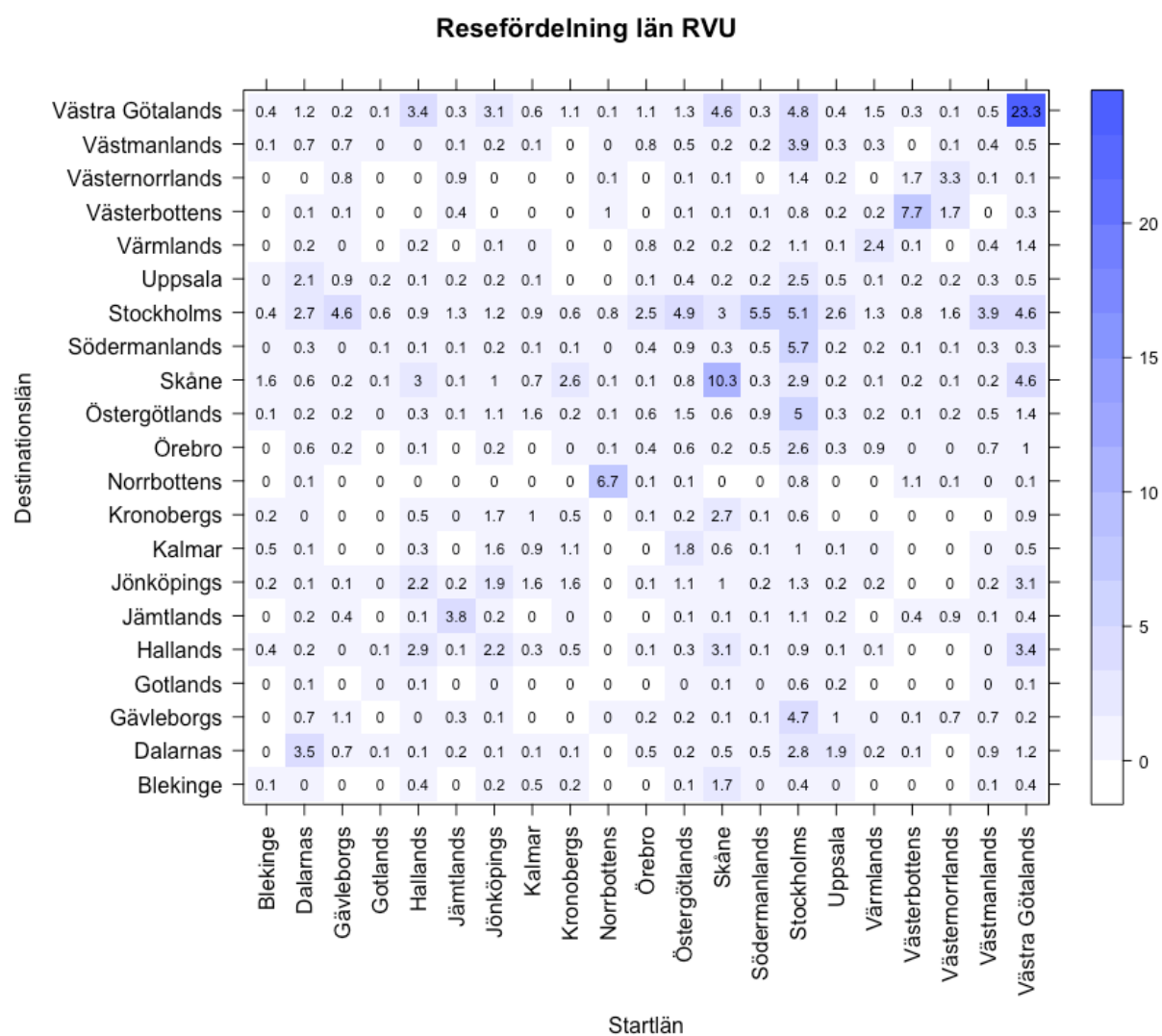
Figur 8 (vänster) ger en övergripande bild över hur spridningen av resorna från RVU ser ut mellan länen i Sverige. Bredare linje indikerar fler resor. Här kan noteras att mycket av resandet kopplas till Stockholms län, Västra Götalands län, Skåne län och Hallands län. Motsvarande kartbild över spridningen av resor i september och oktober 2017 från mobilnätdata ges i Figur 8 (höger). Notera att skalningen av linjerna i Figur 8 är sådan att den relation som har flest resor har samma bredd i den vänstra respektive högra figuren.



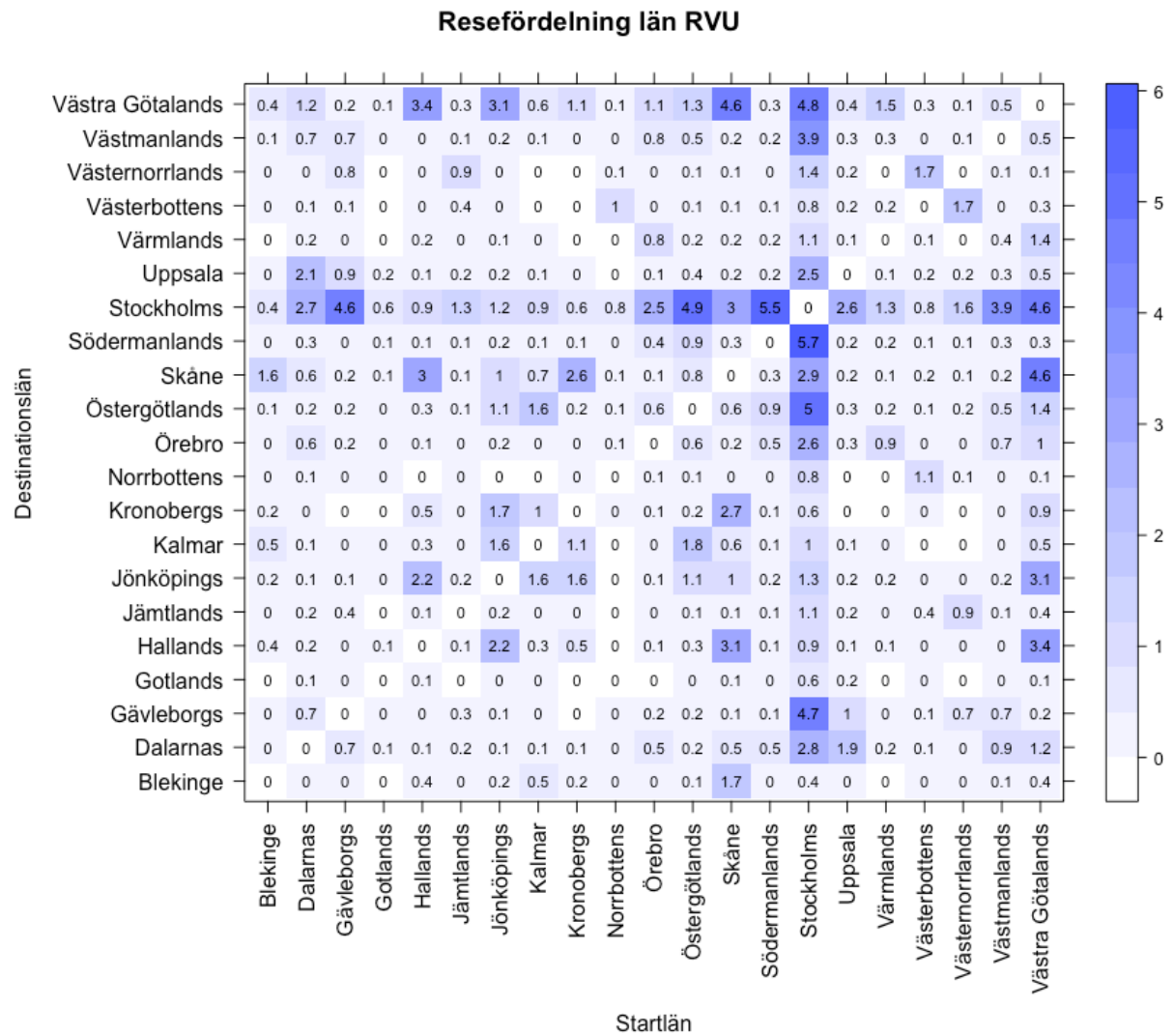
Figur 8: Vänster: Illustration av spridningen av resor mellan län baserat på 24 månaders RVU-data. Höger: Illustration av spridning av resor mellan län baserat på mobilnätdata för september och oktober 2017. Bredare linje indikerar fler resor. Bredden på linjerna är normerade så att bredaste linjen har samma bredd i vänstra och högra kartbilden.

Den OD-matris som återspeglas i Figur 8 (vänster) ges i Figur 9. OD-matrisen i Figur 9 visar antal resor per dygn i tusental. I Figur 10 ges samma bakomliggande OD-matris, men här är inomlänresor eliminerade (OD-elementen är satta till 0).

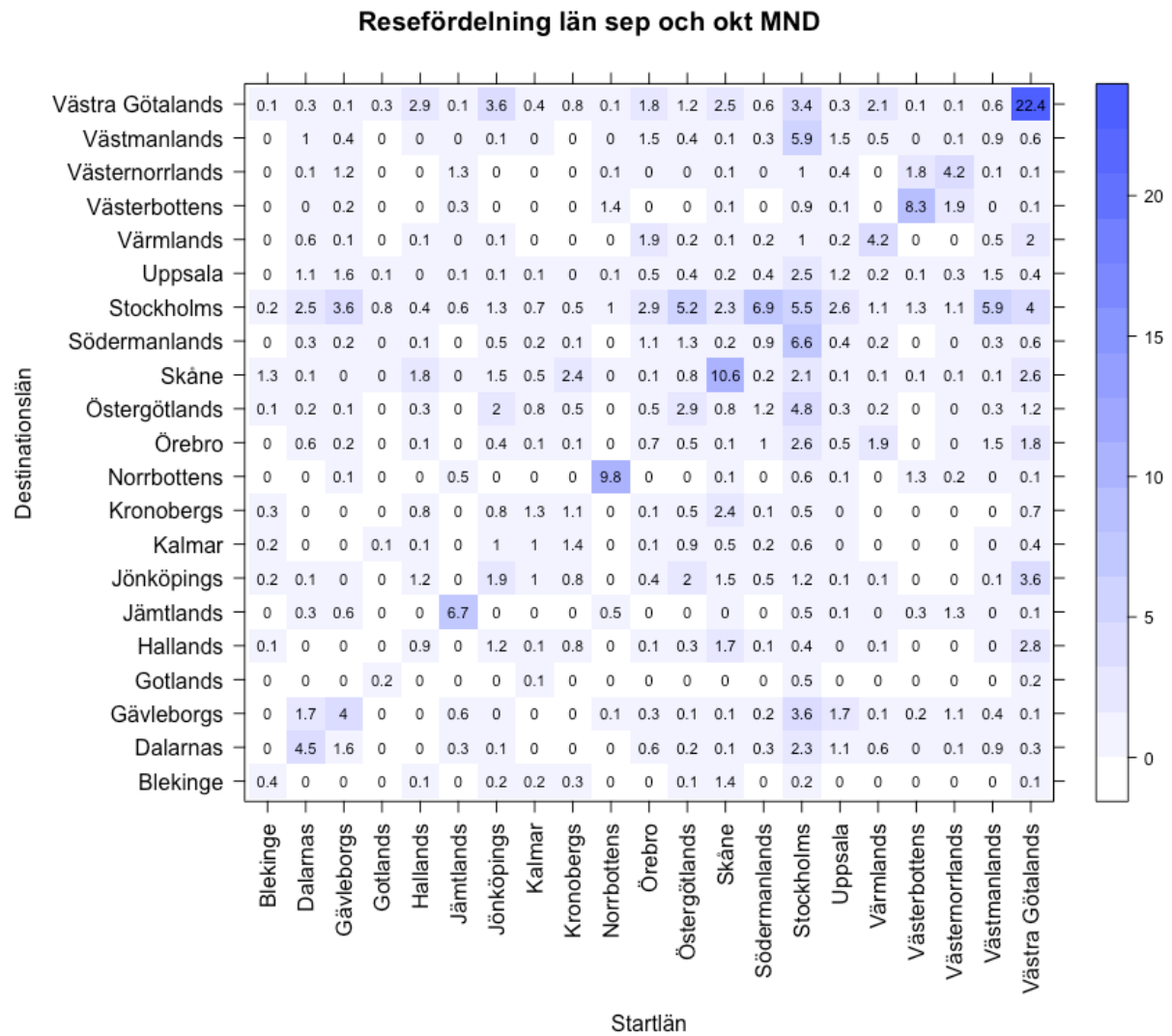
Den bakomliggande OD-matrisen för resor som illustreras i Figur 8 (Höger) ges i Figur 11 och Figur 12 med och utan inomlänresorna. OD-matrisen i Figur 12 visar resor per dygn i tusental. Vi noterar att skalan är mellan 0 och drygt 7 tusen resor per dygn i Figur 12, och mellan 0 och 6 tusen resor per dygn i Figur 10.



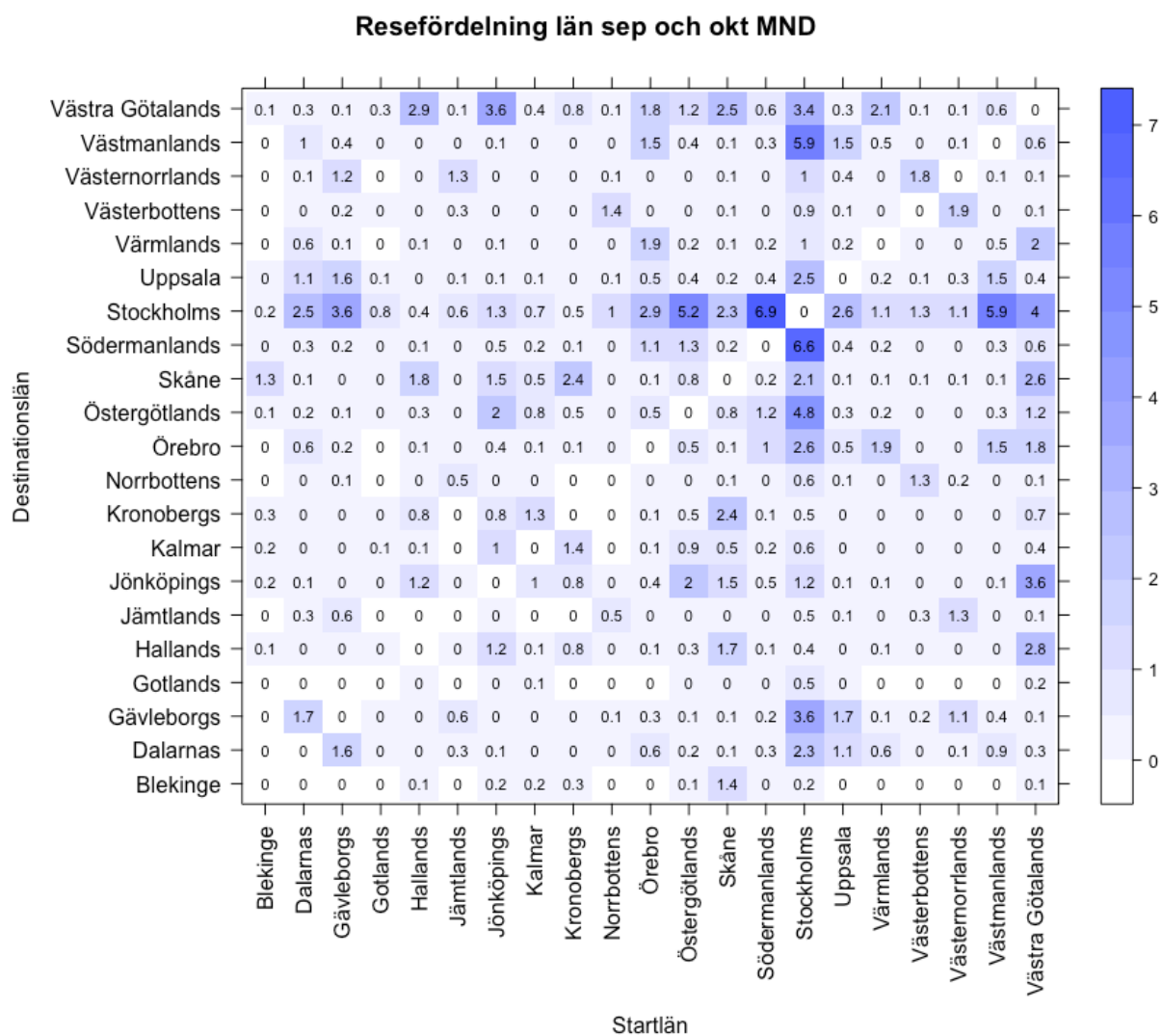
Figur 9: OD-matris med antal resor per dag (tusentals) mellan par av län, inklusive inomlännsresor.



Figur 10: OD-matris med antal resor per dag (tusentals) mellan par av län, exklusive inomlännsresor.

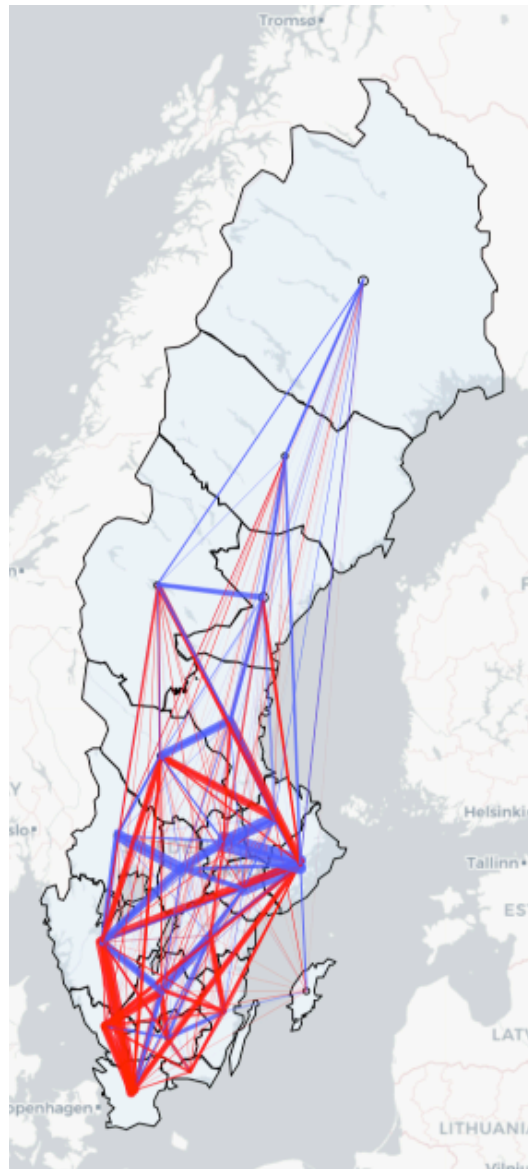


Figur 11: OD-matris med antal resor per dag (tusentals) mellan par av län för mobilnätsdata, inklusive inomlännsresor.



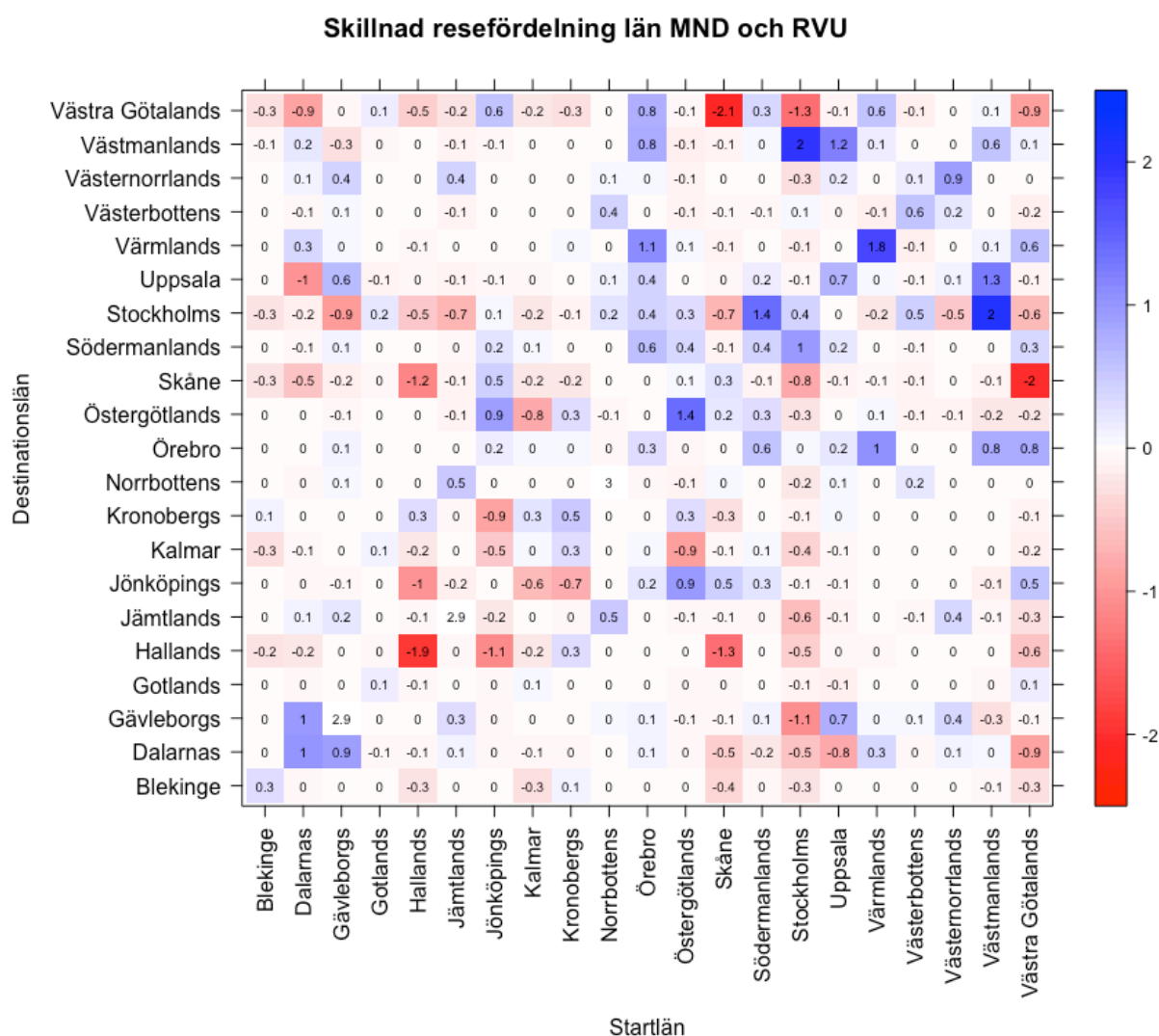
Figur 12: OD-matris med antal resor per dag (tusentals) mellan par av län för mobilnätdata, exklusive inomlännsresor.

I Figur 13 illustreras skillnaden mellan genomsnittligt antal resande per dag mellan länen. En blå linje indikerar att mobilnätdata ger ett högre antal resor än RVU. En röd linje indikerar att mobilnätdata har gett ett lägre antal resor än RVU.



Figur 13: Illustration av flödesskillnad mellan antal resor från mobilnäts data och från RVU. Bredden indikerar differensstorlek, blå anger att mobilnätsdata ger större antal resor än RVU, röd färg indikerar mindre antal resor.

Figur 14 visar skillnaden mellan OD-matriserna i Figur 9 och Figur 11.



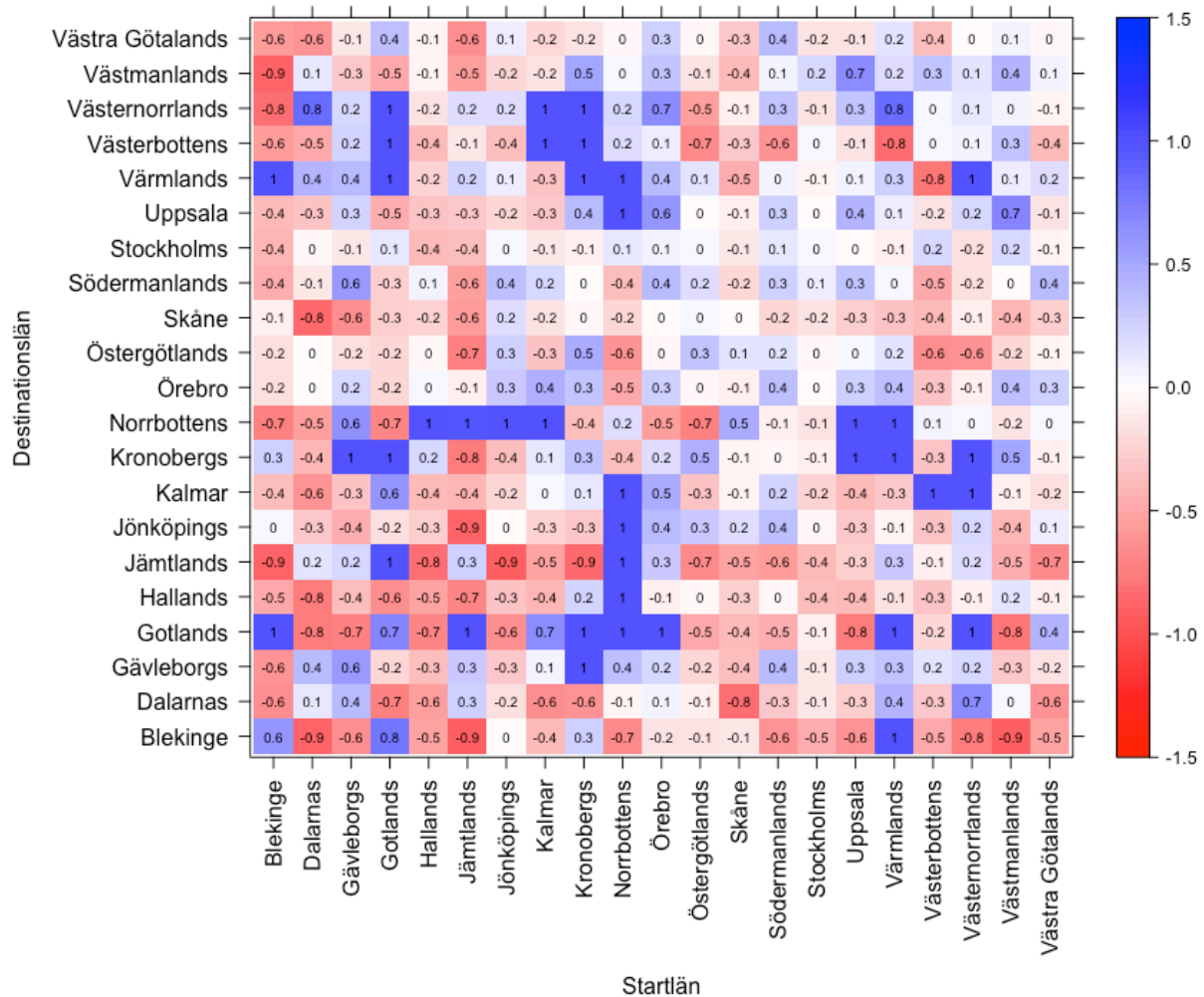
Figur 14: Skillnad i antal resor estimerat från mobilnätdata och från RVU mellan varje par av län.

I Figur 15 är den symmetriska relativa skillnaden mellan genomsnittligt antal resande per dag från RVU och antalet estimerat från mobilnätdata för september och oktober illustrerad. För varje element i denna matris har kvoten

$$r_{OD} = \frac{x_{OD} - y_{OD}}{x_{OD} + y_{OD}}$$

beräknats, där x_{OD} är antalet resande per dag från mobilnätdata och y_{OD} är antal resande per dag estimerat från RVU. För element där $x_{OD} = y_{OD} = 0$ är kvoten definierad som noll.

Symmetrisk relativ skillnad resefördelning län MND och RVU



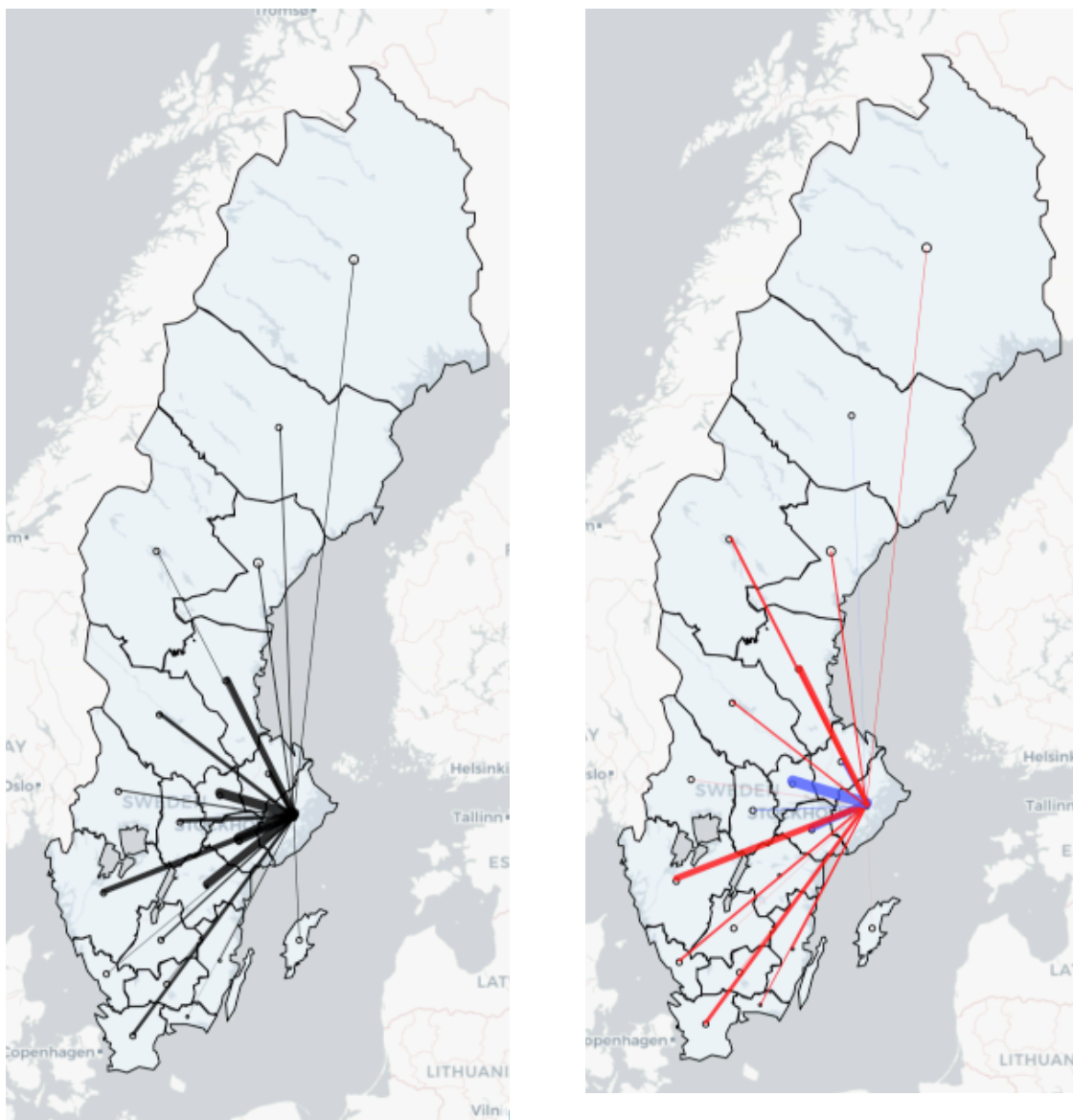
Figur 15: Symmetrisk relativ skillnad i antal resande per dag (tusentals) mellan RVU och mobilnätdata för september och oktober.

5.3. Resors spridning från utvalda län

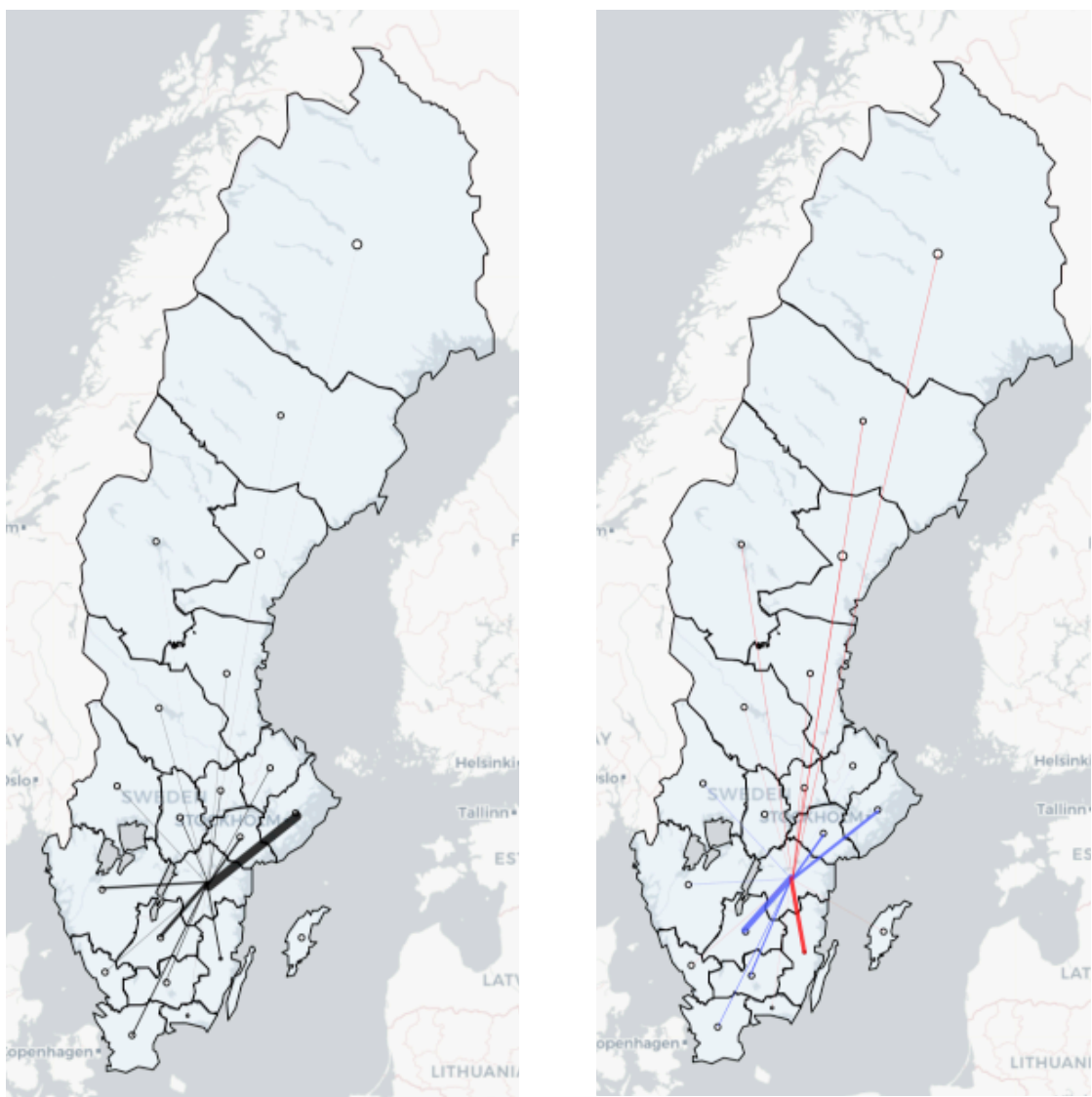
I denna sektion beskrivs spridningen av resor ut från specifika län. Spridningskartor ges för mobilnätdata och den skillnad som finns gentemot RVU data. Spridningen ut från utvalda län illustreras också i histogram där destinationslänen är sorterade i storleksordning.

Beskrivningen görs för Stockholms län, Västra Götalands län och Östergötlands län. Till vänster i Figur 16 illustreras spridningen av resor ut från Stockholms län baserat på mobilnätdata för september och oktober 2017. Till höger i samma figur illustreras skillnaden i spridning ut från Stockholms län mellan estimering från mobilnätdata och från RVU.

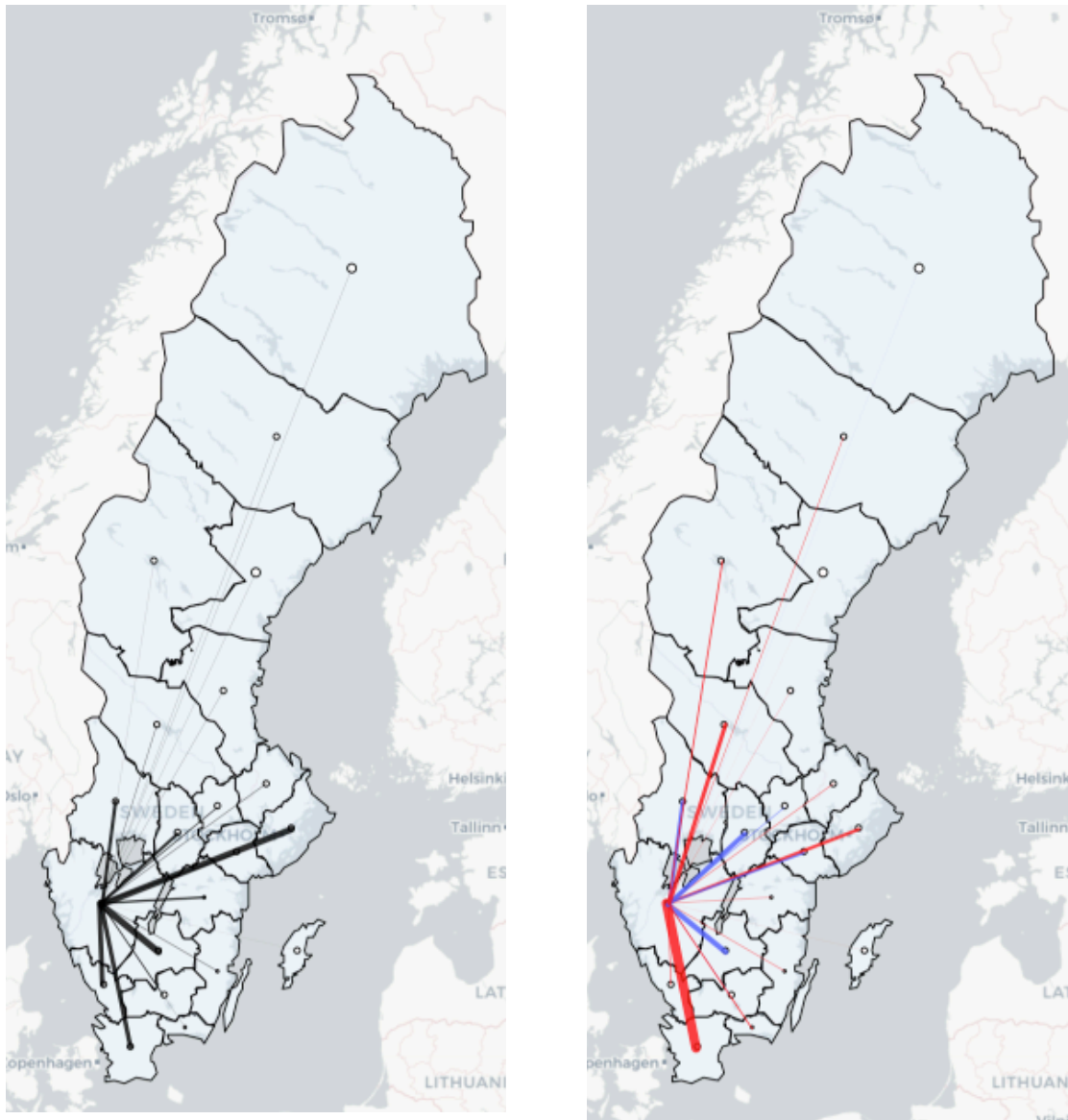
Linjernas bredd i figurerna indikerar storleken på flödet respektive flödesskillnaden och blå linje indikerar att mobilnätdata ger större antal resor än RVU; röd linje lägre antal resor. Notera att linjens bredd i den vänstra och högra är skalade så att bredaste linjen i de två figurerna är av samma bredd. På liknande sätt illustreras spridningen av resor ut från Östergötlands län i Figur 17 och ut från Västra Götalands län i Figur 18.



Figur 16: Vänster: Spridningen av resor ut från Stockholms län, baserat på mobilnätdata för september och oktober 2017. Höger: Skillnad i spridning från Stockholms län mellan estimering från mobilnätdata och från RVU. Linjens bredd indikerar flödesskillnad. Blå linje indikerar att mobilnätdata ger större antal resor än RVU; röd linje lägre antal resor.

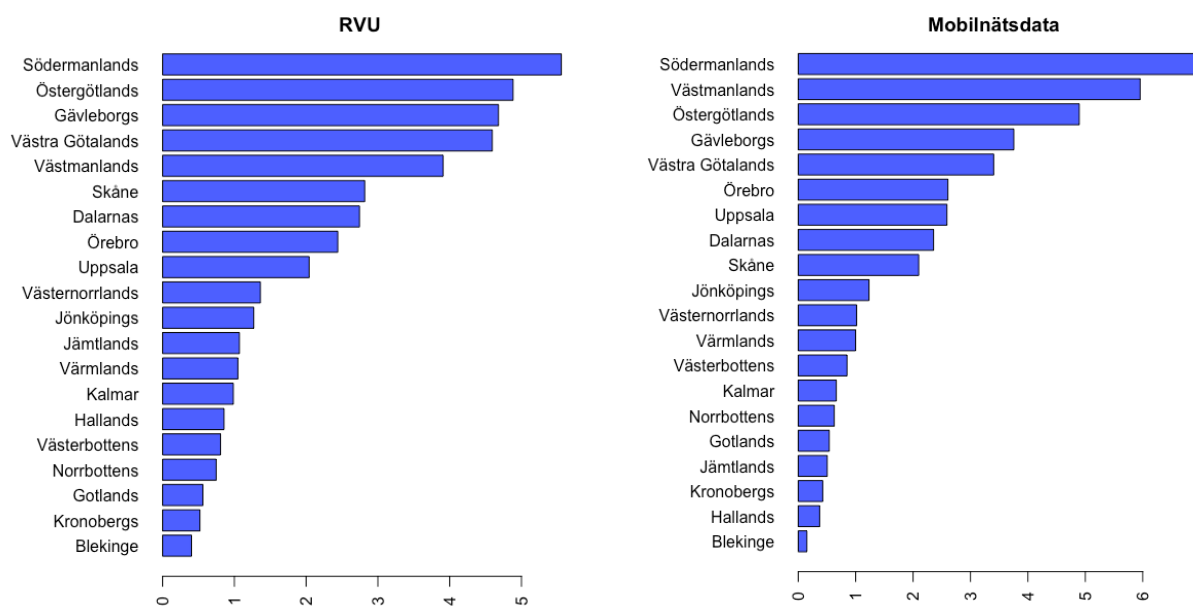


Figur 17: Vänster: Spridningen av resor ut från Östergötlands län, baserat på mobilnätdata för september och oktober 2017. Höger: Skillnad i spridning från Östergötlands län mellan estimering från mobilnätdata och från RVU. Linjens bredd indikerar flödesskillnad. Blå linje indikerar att mobilnätdata ger större antal resor än RVU; röd linje lägre antal resor.



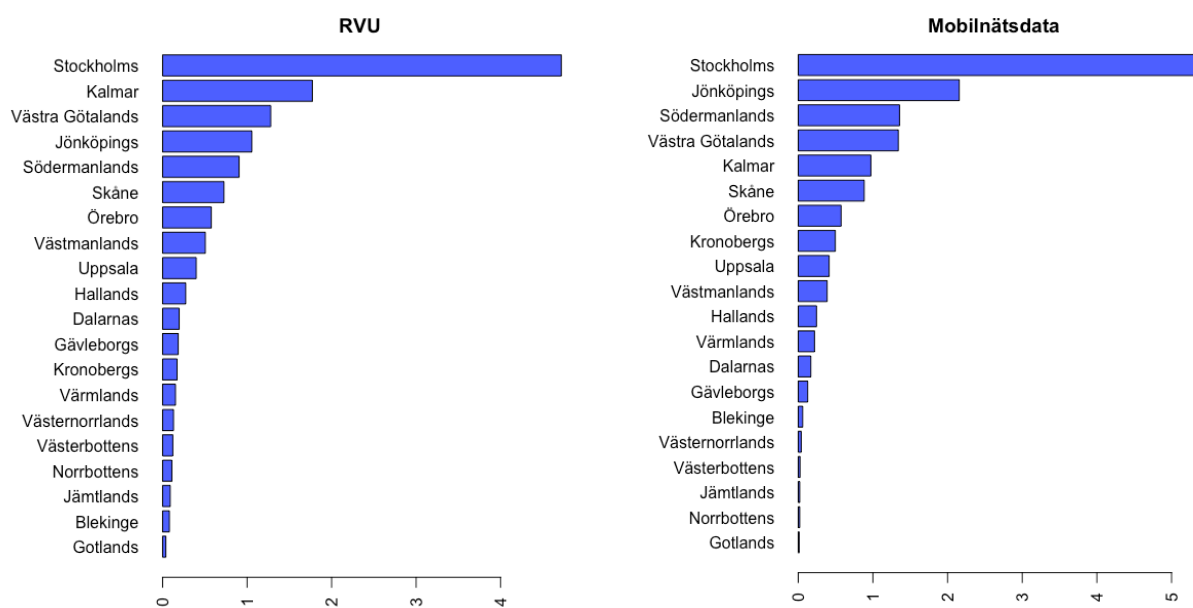
Figur 18: Vänster: Spridningen av resor ut från Västra Götalands län, baserat på mobilnätdata för september och oktober 2017. Höger: Skillnad i spridning från Västra Götalands län mellan estimering från mobilnätdata och från RVU. Linjens bredd indikerar flödesskillnad. Blå linje indikerar att mobilnätdata ger större antal resor än RVU; röd linje lägre antal resor.

Figur 19 visar listor med destinationslän för resor ut från Stockholms län ordnade i avtagande antal resor per dygn för RVU respektive för mobilnätdata.



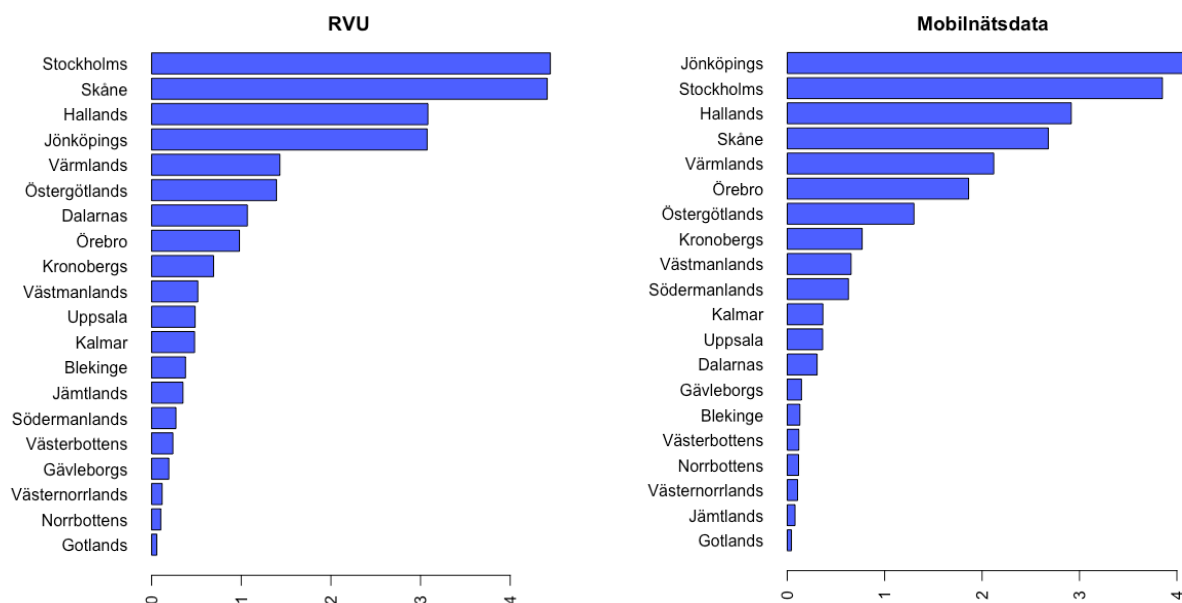
Figur 19: Rangordnad lista över destinationslän, från Stockholms län, ordnade i avtagande antal resor per dygn (tusentals). Resor inom länet är inte med i rangordningen.

Figur 20 visar listor med destinationslän för resor ut från Östergötlands län ordnade i avtagande antal resor per dygn.



Figur 20: Rangordnad lista över destinationslän, från Östergötlands län, ordnade i avtagande antal resor per dygn (tusentals). Resor inom länet är inte med i rangordningen.

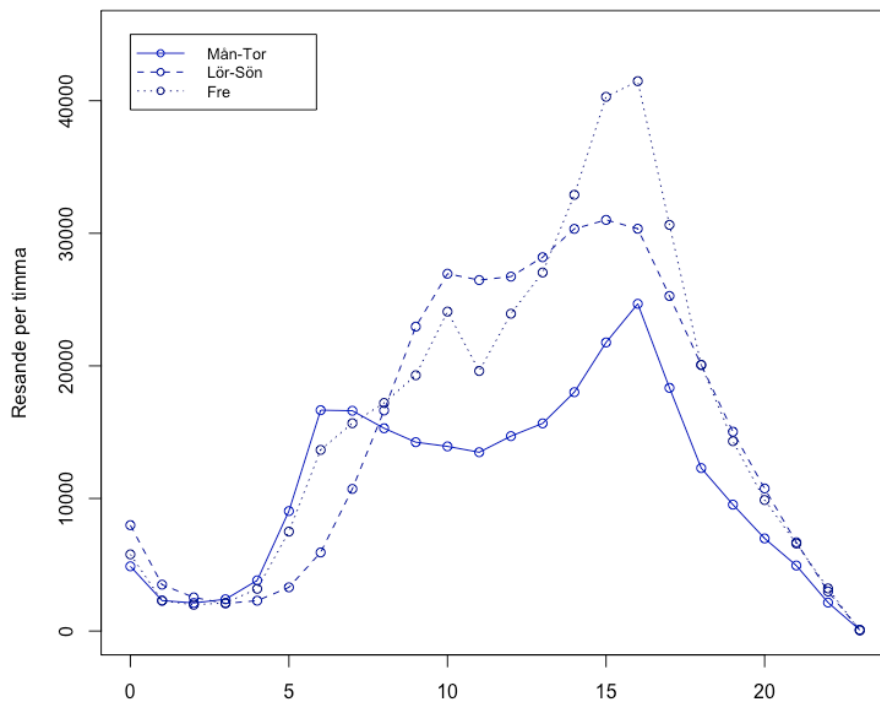
Figur 21 visar listor med destinationslän för resor ut från Västra Götalands län ordnade i avtagande antal resor per dygn.



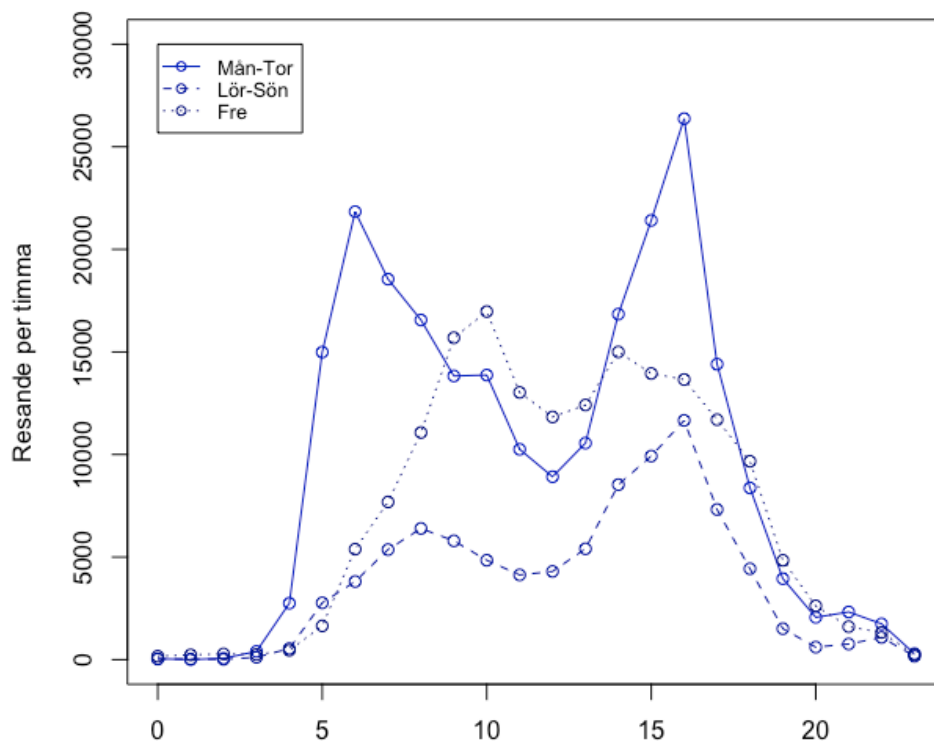
Figur 21: Rangordnad lista över destinationslän, från Västra Götalands län, ordnade i avtagande antal resor per dygn (tusentals). Resor inom länet är inte med i rangordningen.

5.4. Resors fördelning över dygn

I detta delkapitel visas resornas fördelning över dygnet. Samma urval av RVU-data som i tidigare delkapitel har använts. Mobilnätsdata för september och oktober 2017 har använts. I diagrammen är det tidpunkten för start av resan som har använts. I Figur 22 så visas starttidpunktens fördelning över dygnets timmar från mobilnätsdata för september 2017. I figuren särredovisas antalet för tre olika kategorier av dagar, måndag till torsdag, fredag, samt lördag och söndag. Vid x-axelns 0-punkt anges antalet resor under timmen mellan 00:00 och 01:00. Vi kan notera från figuren att antalet resor som startar mellan 23:00 och 24:00 är uppseendeväckande lågt, alternativt att antalet resor första timmarna på dygnet är högt. Orsaken till detta är oklar. (Liknande fenomen kan också ses i Figur 24.) I Figur 23 visar motsvarande fördelning för de tre kategorierna av dagar över dygnets timmar, baserat på estimerat antal resor per timma från RVU-data.

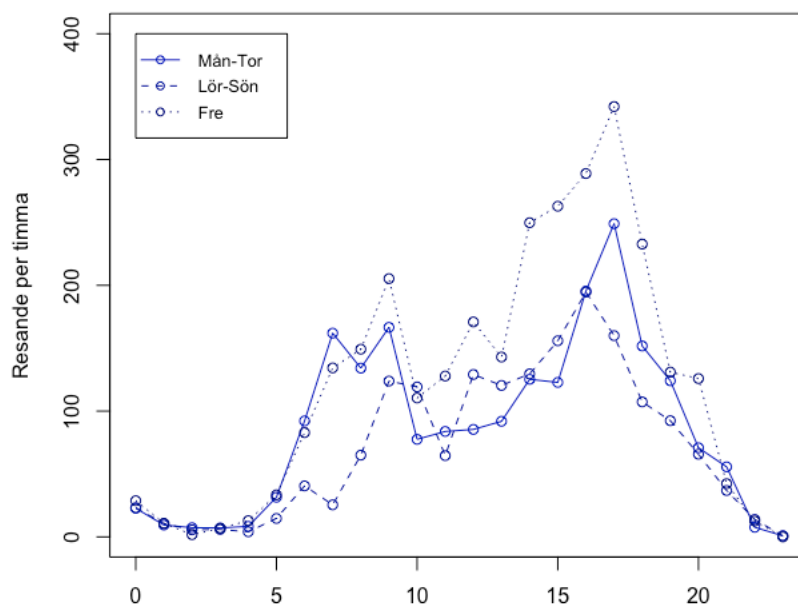


Figur 22: Fördelning av resor per dag över dygnets timmar från mobilnätdata. Heldragen linje visar måndag-torsdag, prickad linje fredag och streckad linje lördag-söndag.

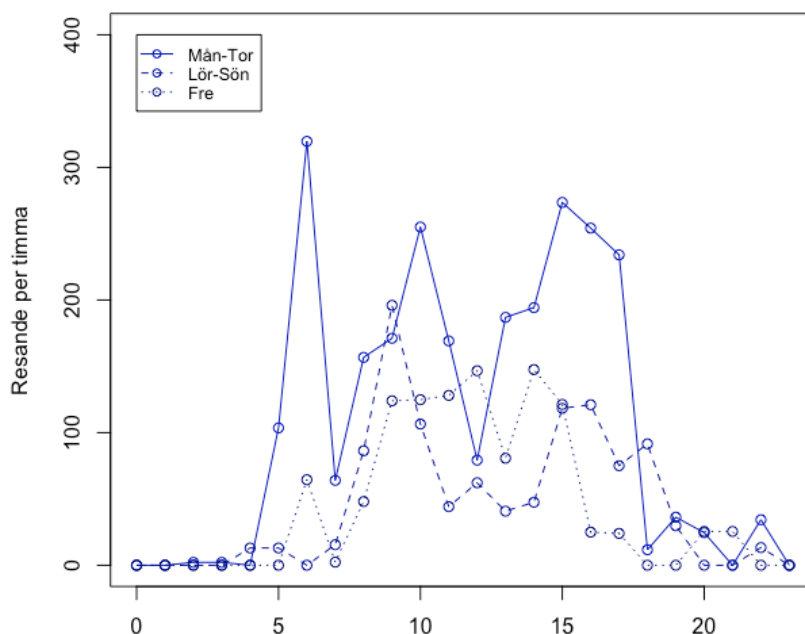


Figur 23: Fördelning av resor per dag över dygnets timmar från RVU. Heldragen linje visar måndag-torsdag, prickad linje fredag och streckad linje lördag-söndag.

I Figur 24 visas resor per timma baserat på ett genomsnitt över hela september och oktober från Stockholms län till Västra Götalands län, från mobilnätdata. Motsvarande fördelning illustreras för RVU-data i Figur 25.



Figur 24: Fördelning av resor per dag från Stockholms län till Västra Götalands län över dygnets timmar för mobilnätdata. Heldragen linje visar måndag-torsdag, prickad linje fredag och streckad linje lördag-söndag.



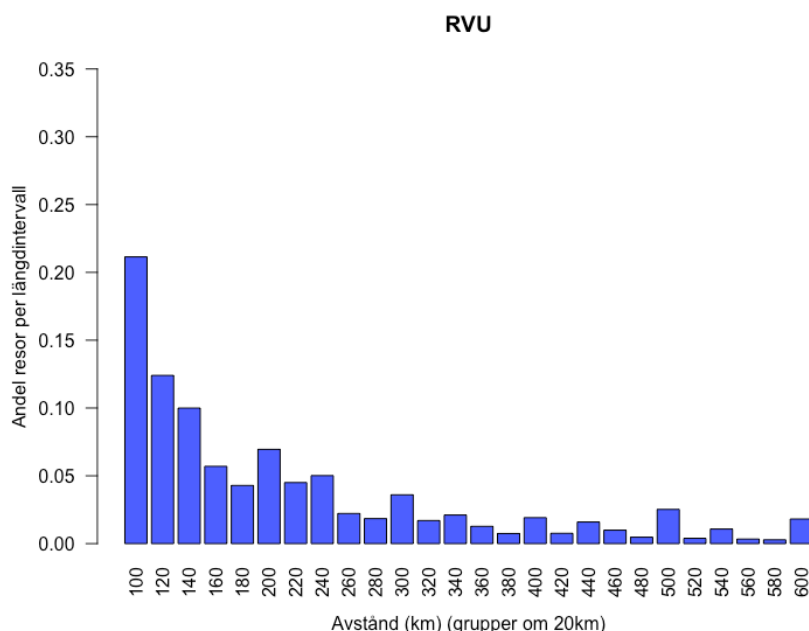
Figur 25: Resors fördelning över dygnet från Stockholms län till Västra Götalands län baserat på ett genomsnitt av 24 månaders data ur RVU. Heldragen linje visar måndag-torsdag, prickad linje fredag och streckad linje lördag-söndag.

5.5. Resors längdfördelning

I denna sektion redovisas fördelningen av resors längd identifierade från mobilnätsdata, och jämförs med motsvarande spridning från 24 månaders av RVU-data. Data från mobilnätet (se Sektion 4.2.) är givna på ett sådant format så att antal resor inte kan utläsas från denna datamängd, utan bara fördelningen. Baserat på totalresande, redovisat i Sektion 5.1. , skulle en uppskattning av antal resor av olika längd kunna tas fram. Fördelningar av resornas längd från mobilnätsdata är ett genomsnitt baserat på september och oktober 2017, därför redovisas ingen analys för enskilda månader.

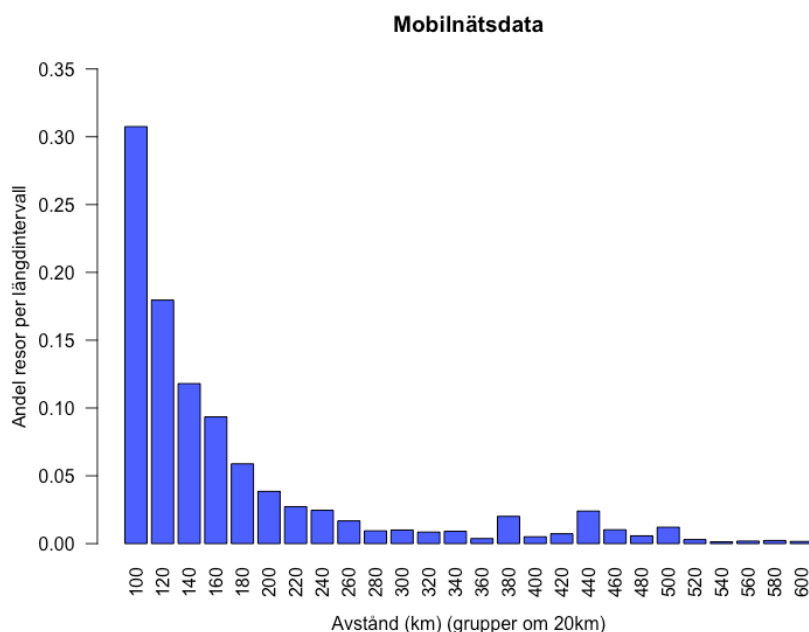
Vid jämförelse av resors längd mellan RVU och mobilnätsdata så ska det noteras att vad som avses med ”reslängd i kilometer” är olika i de två datamängderna. För RVU så är reslängden det avstånd som uppges i enkäten. Det innebär att den dels kan påverkas av individens uppfattning om längden, dels kan en resa mellan en punkt A och en punkt B vara av olika längd beroende på vilket färdstätt som används. Det är också så att det rimligen är olika svårt att ge en korrekt längd t.ex. för en bilresa, då du har en trippmätare i instrumentpanelen på bilen, och för en tågresor, då resenären rimligen har mer fokus på restiden än sträckan som färdats. För mobilnätsdata så är resans längd definierad som det euklidiska avståndet (”fågelvägen”) mellan startpunkten på resan och slutpunkten för resan. Det gör att reslängden för en resa från A till B är lägre i mobilnätsdata jämfört med RVU, under förutsättning att individen har angivit resans längd korrekt i RVU.

I Figur 26 illustreras fördelningen av resornas längd baserat på RVU. Alla resor längre än 100 kilometer är indelade i grupper om 20 kilometer. Andelen av varje sådan grupp illustreras i figuren. Grupperna sträcker sig från mellan 100 kilometer och 120 kilometer upptill 600 kilometer och 620 kilometer i figuren. Andelen är normerad så att alla resor utgör andelen ett. I Figur 26 saknas 4% av resorna, vilka är längre än 620 kilometer.



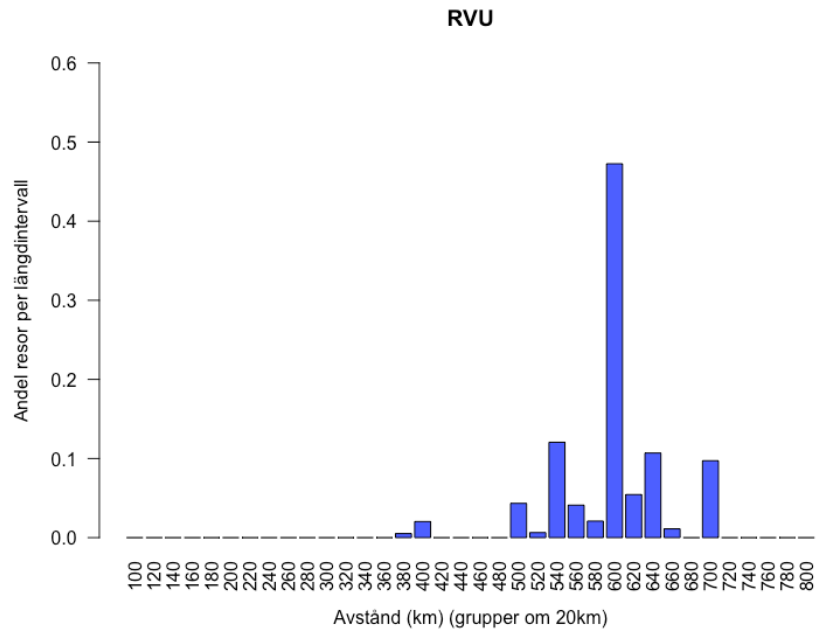
Figur 26: Längdfördelning för resor över 100km från RVU, genomsnitt över 24 månader 2013, 2014, och 2016. Stapeln under 100 innehåller resor i intervallet 100km till 120km, etc. I diagrammet är 4% av resorna trunkerade i och med att histogrammet är kapat vid 620km.

I Figur 27 ges fördelningen av resornas längd baserat på mobilnätdata. Även här är alla resorna grupperade i grupper om 20 kilometer, och omfattar resor mellan 100 kilometer och 620 kilometer. I figuren saknas 2% av resorna, vilka är längre än 620 kilometer.



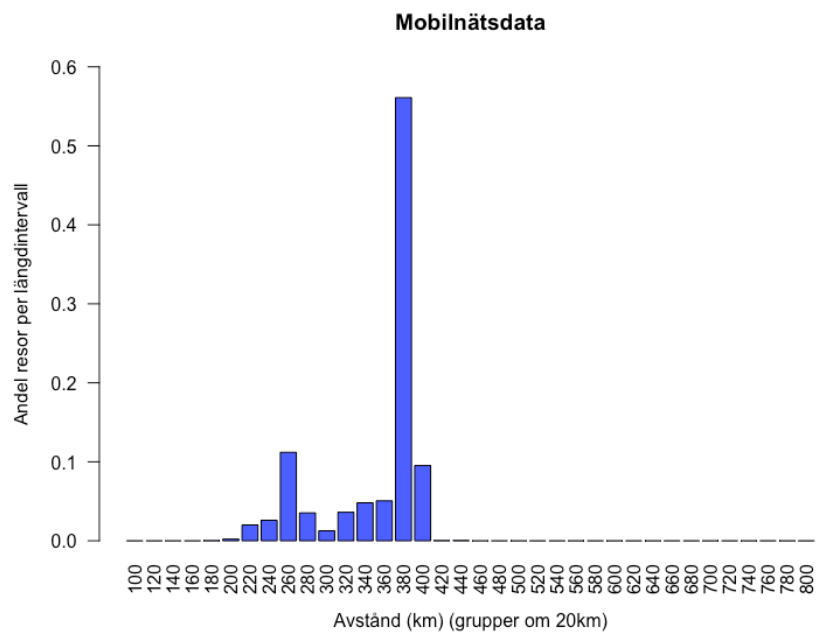
Figur 27: Längdfördelning för resor över 100km från mobilnätdata baserat på resande i september och oktober 2017. Stapeln under 100 innehåller resor i intervallet 100km till 120km, etc. I diagrammet är 2% av resorna trunkerade i och med att histogrammet är kapat vid 620km.

I Figur 28 illustreras längden på resor som går från Stockholms län till Västra Götalands län baserat på RVU-data. Resornas längd redovisas som andel i varje 20 kilometersintervall från 100 kilometer till 820 kilometer. Från figuren kan vi notera att det är en stor andel av resorna mellan dessa två län som har en längd i intervallet 600 kilometer till 620 kilometer.



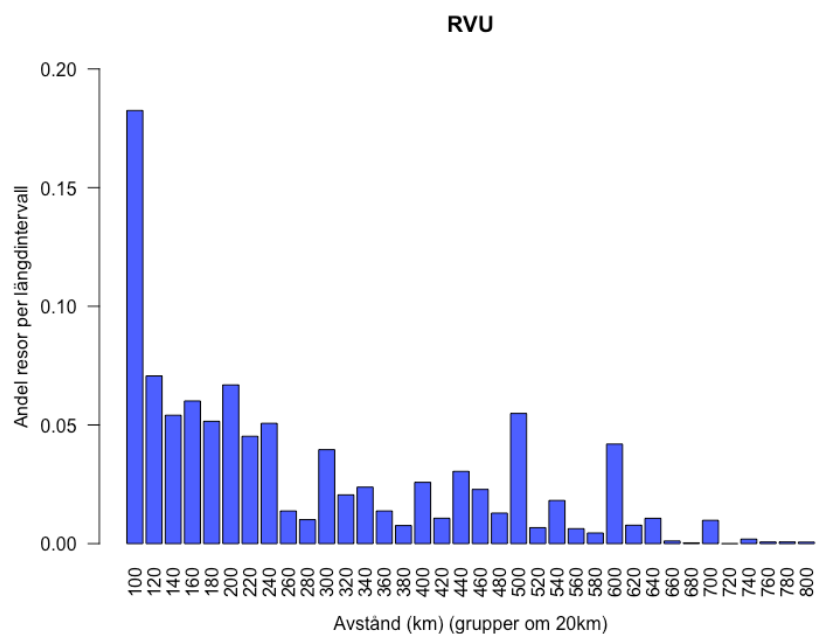
Figur 28: Längdfördelning för resor från Stockholms län till Västra Götalands län från RVU, baserat på 24 månader 2013, 2014 och 2016. Inga resor elimineras på grund av trunkering av histogrammet.

I Figur 29 illustreras längden på resor som går från Stockholms län till Västra Götalands län baserat på mobilnätdata. Även här redovisas resornas längd som andel i varje 20 kilometersintervall från 100 kilometer till 820 kilometer. Från figuren kan vi notera att det är en stor andel av resorna mellan dessa två län som har en längd i intervallet 380 kilometer till 400 kilometer. Skillnaden mellan vilket längdintervall som har det största antal resor i Figur 28 och Figur 29 är uppenbar vid jämförelse av figurerna. Förklaringen till skillnaden bör till mycket stor del härröras till skillnad i hur resans längd definieras i de två datamängderna.



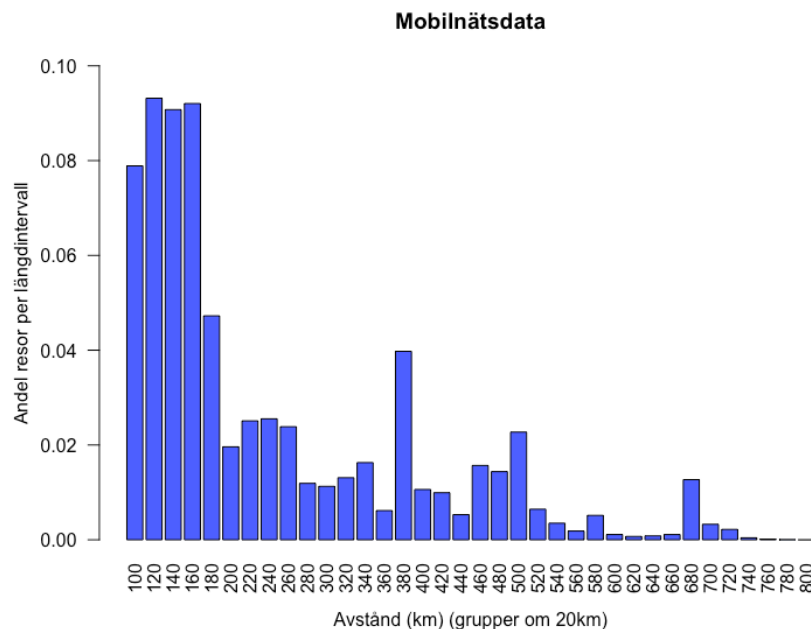
Figur 29: Längdfördelning för resor från Stockholms län till Västra Götalands län baserat på mobilnätetsdata för september och oktober 2017. Inga resor elimineras på grund av trunkering av histogrammet.

I Figur 30 Illustreras längden på de resor som startar i Stockholms län baserat på data i RVU i intervall om 20 kilometer mellan 100 kilometer och 820 kilometer.



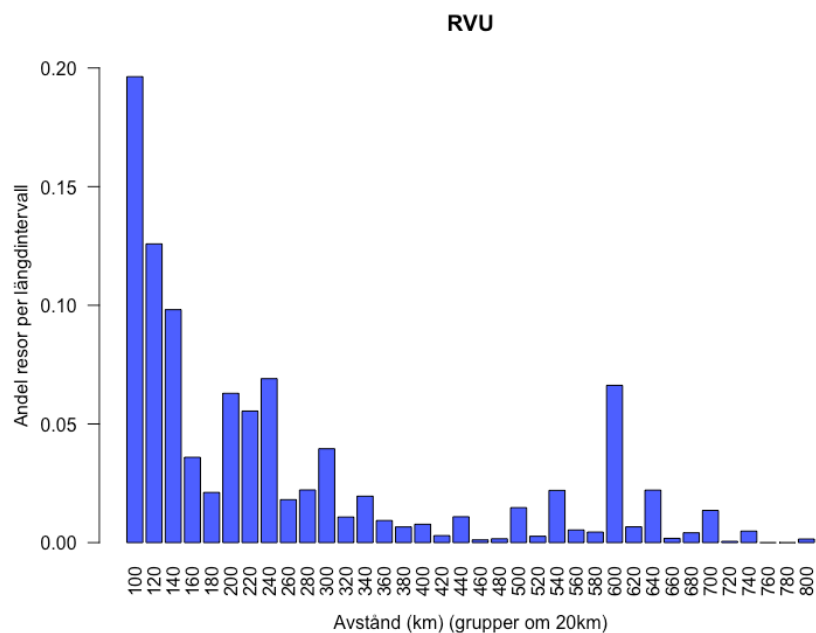
Figur 30: Reslängdfördelning för resor ut från Stockholms län, baserat på RVU för 24 månader under 2013, 2014 och 2016. Inga resor elimineras på grund av trunkering av histogrammet.

I Figur 31 ges fördelningen av resorna längd utifrån Stockholms län från mobilnätdata, jämförbar med fördelningen i Figur 30. Notera att skalan som används för andel i Figur 30 och i Figur 31 skiljer sig. I Figur 30 och Figur 31 kan vi notera en mindre topp i längdfördelningen för intervallet som startar vid 600 kilometer respektive 380 kilometer. Dessa två mindre toppar härrör delvis från resorna från Stockholms län till Västra Götalands län, vilka illustrerades i Figur 28 respektive Figur 29.

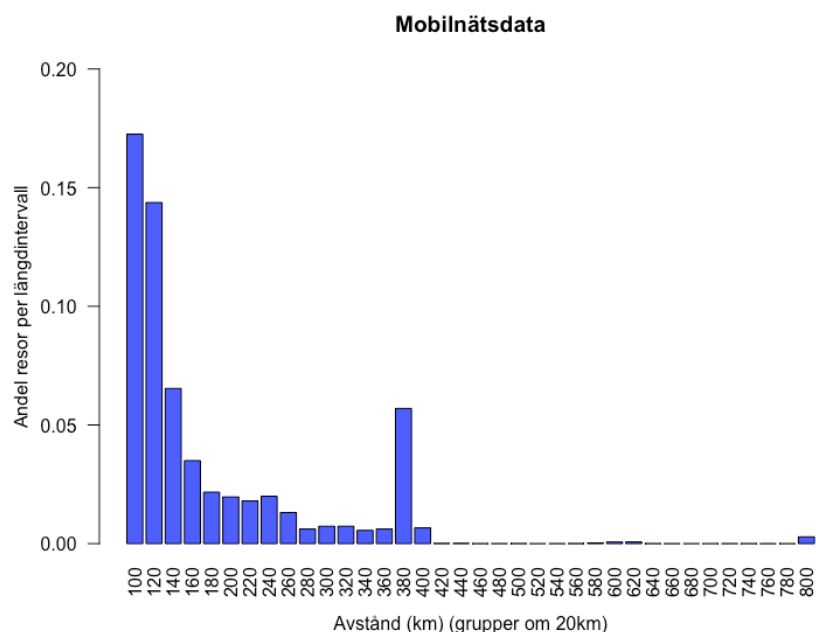


Figur 31: Reslängdfördelning för resor ut från Stockholms län, baserat på mobilnätdata för september och oktober 2017. Mindre än 0.1% av resorna elimineras på grund av trunkering av histogrammet.

I Figur 32 och Figur 33 anges resornas längdfördelning utifrån Västra Götalands län från RVU respektive mobilnätdata. Även i dessa histogram återfinns mindre toppar vid intervallen som startar vid 600 kilometer respektive 380 kilometer. Dessa förklaras av en relativt stor andel resor till Stockholms län.

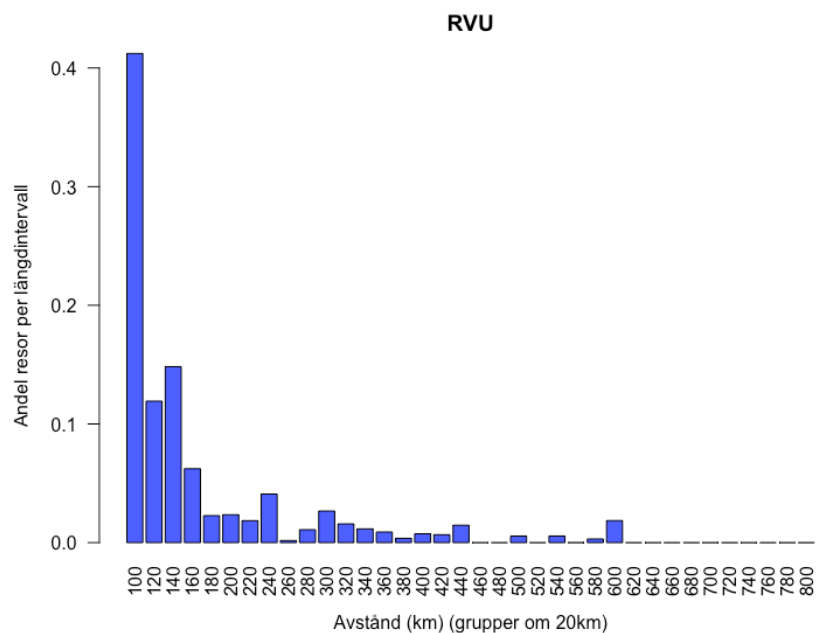


Figur 32: Reslängdfördelning för resor ut från Västra Götalands län, baserat på RVU för 24 månader under 2013, 2014 och 2016. Inga resor elimineras på grund av trunkering av histogrammet.

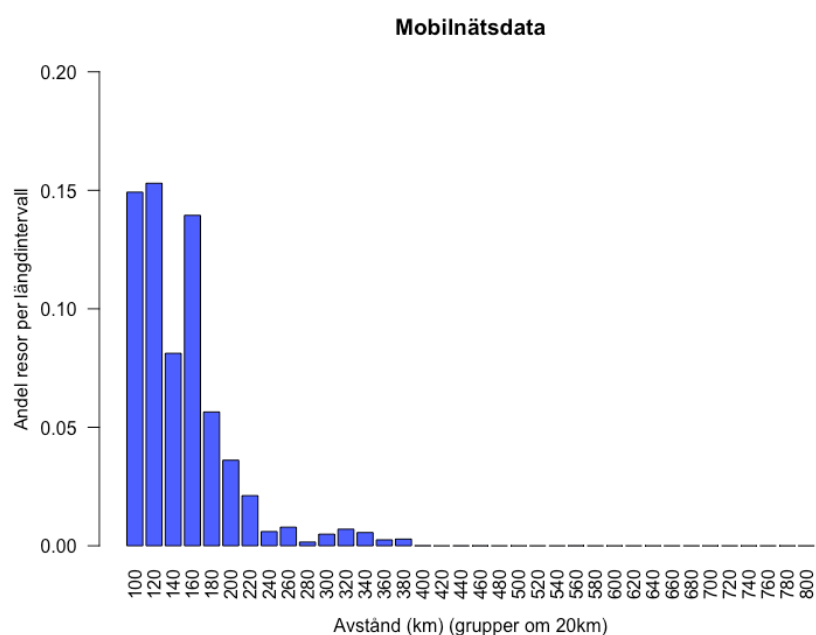


Figur 33: Reslängdfördelning för resor ut från Västra Götalands län, baserat på mobilnätsdata för september och oktober 2017. Mindre än 0.1% av resorna elimineras på grund av trunkering av histogrammet.

I Figur 34 och Figur 35 anges resornas längdfördelning utifrån Östergötlands län från RVU respektive mobilnätsdata. Notera att skalan för andelen i dessa två figurer är olika. Från Figur 17 vet vi att de flesta resor ut från Östergötlands län över 100 kilometer går till Stockholms län.



Figur 34: Reslängdfördelning för resor ut från Östergötlands län, baserat på RVU för 24 månader under 2013, 2014 och 2016. Inga resor elimineras på grund av trunkering av histogrammet.



Figur 35: Reslängdfördelning för resor ut från Östergötlands län, baserat på mobilnätsdata för september och oktober 2017. Mindre än 0.5% av resorna elimineras på grund av trunkering av histogrammet.

6. Analys

För att kunna göra jämförelsen mellan statistik om långväga resor från RVU Sverige med mobilnätsdata från en operatör behöver data från flera år för RVU Sverige analyseras i aggregerad form. I denna studie har 24 månaders RVU-svar från 2013, 2014 och 2016 används som jämförelseunderlag när jämförelser har gjorts mot data för september och oktober 2017 från mobilnätet.

Båda datamängderna har skapats genom att ett urval av Sveriges befolkning används som underlag för estimatet av vilka resor som görs, och när. Även om urvalet i den mobilnätsdata som använts i denna pilotstudie inte har exakt specificerats, så är underlaget flera tiopotenser större. Processen som används för att gå från urvalet till estimatet för den totala befolkningen skiljer sig mellan de två datamängderna. När det gäller RVU Sverige så är det t.ex. information om åldersgrupp, kön, region som har använts vid denna uppskalning. För mobilnätsdata så bygger uppskalningsprocessen på data om penetrationsgrad i olika geografiska områden. Denna data är enbart tillgänglig hos den enskilda operatören, och exakt hur uppskalningen görs samt penetrationsgraden är en affärshemlighet.

Den inledande analysen av antalet resor som estimerats från RVU Sverige, samt de konfidensintervall som redovisats för dessa estimat, ledde till att 24 månaders RVU-svar har använts. Trots denna förhållandevis långa tidsperiod så finns relativt stor osäkerhet i de estimat som tas fram. Tittar man t.ex. på urval för en given månad och dag blir feluppskattningen ofta över 50% av resandet (med 95% konfidensintervall). Därför är det svårt att göra jämförelser över kortare tidsperioder med god kvalitet. De inledande jämförelserna för totalresandet (presenterat som genomsnittligt resande per dag) mellan RVU 2013, 2014 och 2016, med mobilnätsdata för 2017, samt de skattningar av konfidensintervall som gjorts för RVU, ger en indikation på att estimatet av antal resor baserat på mobilnätsdata är i rätt storleksordning. Då det exakta antalet resor som faktiskt genomförts är okänt är det inte möjligt att veta om mobilnätsdata ger ett bättre eller sämre estimat än det baserat på RVU. Det är också oklart hur stor påverkan är av att resor i RVU identifieras som en förflyttning med ett ärende i destinationspunkten medan en resa i hanteringen av mobilnätsdata avslutas när den algoritm som använts anser att ingen förflyttning skett under en fördefinierad tidsperiod.

Spridningen av antal resor mellan länen uppvisar liknande mönster utifrån RVU och mobilnätsdata. I somliga par av län är dock skillnaden relativt stor, även om man tar i beaktande att spridningen som jämförts härrör från olika tidsperioder. Men även här är det svårt att påvisa vilken av källorna som visar det antal som stämmer bäst med verkligheten. För att avgöra detta så skulle en ytterligare källa till information om antalet resor mellan länen behöva användas vid jämförelsen, företrädesvis en källa som kan producera estimat av hög noggrannhet.

I jämförelsen av den relativa spridningen av resorna från ett specifikt län så kan viss skillnad identifieras. Baserat på en rangordning av antalet resor från ett specifikt län, till övriga län så kan vi notera att rangordningen uppvisar liknande mönster, men att skillnader förekommer. Vi noterar att antalet resor till län som ligger närmare startlänet för resorna ofta är högre, vilket är naturligt, men att tolkningen kompliceras av att vi enbart har med resor som är 100 kilometer eller längre. Ytterligare en komplikation, som t.ex. bör inverka på antalet resor som ges i

fördelningen för angränsande län, t.ex. för resor ut från Stockholms län till Södermanlands län, är den att resornas längd bygger på olika definitioner i de två dataunderlagen.

I RVU Sverige så är resans längd svaret på den fråga som ställs om just resans längd. Olika individer har rimligen olika sätt att uppskatta längden, och rimligen är det olika svårt beroende på det färdssätt som använts för resan. Görs resan med bil, och speciellt om det är en resa som görs ofta, så har resenären rimligen god koll på längden tack vare bilens trippmätare. När det gäller en resa där det huvudsakliga färdmedlet är tåg eller flygresor så är det rimligen svårare att ge en uppskattning på längden. De avstånd som ges från mobilnätdata har varit baserade på längden av den rätta linjen mellan starten och slutet på resan. Detta gör att avståndet systematiskt är kortare i mobilnätdata. Även det faktum att det i enkätdata rimligen har gjorts viss avrundning till närmaste heltal kilometer, mil eller tiotal mil, vilket inte görs vid estimeringen av avståndet från mobilnätdata påverkar reslängdsstatistiken. Man kan också notera att insamlandet av reslängder vid insamling av enkätdata har pekats ut som ett problem i Trafikanalys (2016).

De olika definitionerna som används för reslängden i RVU Sverige respektive mobilnätdata påverkar vilka slutsatser som kan dras från direkta jämförelser mellan de längdfördelningar som har presenterats. Båda datamängderna uppvisar en tydlig systematik i att kortare resor är vanligare än längre resor. Vid analys av längden på resor mellan specifika län så kan ett par intressanta observationer göras. Från resultaten av reslängdfördelningen på resor mellan Stockholms län och Västra Götalands län observerar vi att över 40% av resorna har en längd som ligger i intervallet 600-620 kilometer i RVU-data. I mobilnätdata har mer än 50% av resorna en längd som ligger i intervallet 380-400 kilometer. Avståndet, fågelvägen, mellan T-centralen i Stockholm, Stockholms län, och Nils Ericsson Terminalen i Göteborg, Västra Götalands län är c:a 396 kilometer. Avståndet bilvägen mellan dessa två resecentrum är, enligt Google Maps, 470 kilometer (eller längre, beroende på vald väg). Vid analys av enskilda par av län har den process som används vid aggregering och anonymisering haft inverkan på resultatet, vilket i praktiken inneburit att inte 100% av resorna har använts vid framtagande av längdfördelningen. Effekten av detta är dock okänd, men det är troligt att andelen långa resor underskattas något på grund av detta trots att antalet datapunkter är fortfarande stort relativt det dataunderlag som återfinns i RVU-data.

När det gäller analys av resornas fördelning över dygnet så uppvisar data liknande mönster överlag, men skiljer sig i vissa aspekter. Fördelningen över dygnets timmar estimerat från RVU-data visar på toppar i resandet morgon och kväll, för alla kategorier av dagar, måndag till torsdag, fredag respektive lördag och söndag. Fördelningen från mobilnätdata ger att fördelningen för fredagen saknar topp på förmiddagen, och att toppen i resandet på fredagseftermiddagen är avsevärt större. Noteras bör att antalet resor vid enskilda timmar från RVU-data bygger på ett relativt litet datamaterial.

7. Slutsatser

Syftet med pilotprojektet är att utföra en första utvärdering av möjligheter och utmaningar med att skatta resestatistik för långväga resor baserat på aggregerad och anonymiserad mobilnätdata. Möjligheten att få rimliga uppskattningar på långväga totalresande och geografisk fördelning är redan i dagsläget mycket god baserat på resultaten av jämförelsen mellan RVU-data och mobilnätdata. Jämförbarheten skulle dock vara bättre om samma

definition på reslängd hade använts i de två datamängderna eftersom även det skattade totalresandet för resor över 100 kilometer påverkas av denna definition. Liknande problem kan uppstå på grund av skillnader i definition av start och slut på en resa. Dessutom visar projektet på potential att skapa denna statistik med rimliga egenskaper med hög tidsupplösning, exempelvis per dag, samtidigt som den automatiserade processen för beräkning av statistiken gör att aktualiteten och punktligheten blir väldigt hög, den skulle i dagsläget kunna levereras exempelvis på månadsbasis.

Enligt den behovsanalys som har utförts kring data från framtida resvaneundersökningar (Trivector, 2017) så är en av de viktigaste frågeställningarna hur resandet förändras över tid. Baserat på de mobilnätsdata som gjorts tillgänglig för detta pilotprojekt så verkar möjligheterna att skapa statistik om denna aspekt av långväga resande från mobilnätsdata god.

Den främsta utmaningen med att skatta resestatistik baserat på mobilnätsdata i dagsläget är antagligen kopplad till kunskap om hur beräkning av statistiken sker, både gällande möjlighet att förstå tillförlitligheten i data, exempelvis kunna beräkna konfidensintervall, men även att möjliggöra jämförbarhet över tid med olika mönster i användning av kommunikationssystem, nya typer av nät eller data från olika operatörer.

I dagsläget finns i mobilnätsdata ingen koppling till resornas ärende eller någon information om resenärerna, så det måste samlas in separat. Separat insamling av till exempel ärende och information om resenären behöver då kopplas ihop med resorna som identifierats från mobilnätsdata. Även data om färd sätt kopplat till resorna från mobilnätsdata skulle öka användningsmöjligheterna betydligt. I dagsläget är reslängd (och i förlängningen transportarbete) inte lämpligt att ersätta rakt av på grund av definition och beräkning av reslängd på det sätt som det gjorts för denna pilotstudie. Att använda en reslängd som är det euklidiska avståndet ("fågelvägen") i skattning av t.ex. det totala trafikarbetet kommer rimligen att underskatta det totalt trafikarbete mycket.

Innan det är möjligt att skapa officiell statistik från mobilnätsdata så behöver tillförlitligheten analyseras vidare. Hur urvalet och hur uppskalningen görs för att matcha målpopulationen behöver dokumenteras. En framtida högre rumslig upplösning av mobilnätsdata kan ge möjlighet till högre sammanvändbarhet med andra typer av data. För att åstadkomma jämförbarhet över tid behöver definitioner och rutiner anpassas för hantering av statistik baserat på mobilnätsdata både vid behandling av data hos operatören och för statistisk analys hos Trafikanalys. En intressant möjlighet kan vara att den ansvariga statistikmyndigheten tar fram hur resor baserade på mobilnätsdata definieras (i samarbete med operatörerna) och att därefter operatörerna försöker anpassa sina algoritmer eller algoritmparametrar för att få maximal överensstämmelse. Därefter bör en validering göras, t.ex. i en region där enkätunderlaget är mycket gott, eller med hjälp av även andra typer av data.

Referenslista

Allström A., Fransson, M., Kristoffersson, I. och Gundlegård, D. (2015) Mobilnätsdata som indata till prognosmodeller. Tillgänglig på https://www.trafikverket.se/contentassets/773857bcf506430a880a79f76195a080/forskningsresultat/mobilnatsdata_prognosmodeller_final_rad56_webb.pdf [Hämtad 2017-12-18]

Breyer, N., Gundlegård, D., Rydergren, C. och Bäckman, J. (2017) Trip extraction for traffic analysis using cellular network data. I 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, sidorna 321–326. Tillgänglig på <https://doi.org/10.1109/MTITS.2017.8005688>

Calabrese, F., Di Lorenzo, G. Liu, L. och Ratti, C. (2011) Estimating origin-destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area, IEEE Pervasive Computing, 10 (4), sidorna 36-44

González, M., Hidalgo, C.A. och Barabási, A-L. (2008) Understanding individual human mobility patterns, Nature, 453, sidorna 779–782

Gundlegård, D. och Karlsson, J.M. (2009) Handover location accuracy for travel time estimation in GSM and UMTS, IET Intelligent Transport Systems, 3(1), sidorna 87-94.

Gundlegård, D., Rydergren, C., Breyer, N. och Rajna, B. (2016). Travel demand estimation and network assignment based on cellular network data. Computer Communications, 95, sidorna 29–42. Tillgänglig på <https://doi.org/10.1016/j.comcom.2016.04.015>

MODE (2013), Mobile network origin-destination estimation. Projektwebbsida. Tillgänglig på <http://transportanalyticslab.se/mode.html> [Hämtad 2017-12-21]

MOFT (2017), Mobilnätdata i framtidens transportsystem. Projektwebbsida. Tillgänglig på <http://transportanalyticslab.se/moft.html> [Hämtad 2017-12-21]

ONS (2017), Research Outputs: Using mobile phone data to estimate commuting flows, Office for national statistics, United Kingdom. Tillgänglig på <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics/researchoutputsusingmobilephonedatatoestimatecommutingflows> [Hämtad 2018-01-16]

Trafikanalys (2012), Metodrapport RVU Sverige 2011, PM: 2012:8. https://www.trafa.se/globalassets/pm/2011-2015/2012/pm_2012_8_metodrapport_rvu_sverige_2011.pdf [Hämtad 2017-12-15]

Trafikanalys (2016), Resvaneundersökningar som indata till persontransportmodeller – problem, möjligheter och framtida behov i Sverige och Norge. Tillgänglig på https://www.trafa.se/globalassets/rapporter/2016/rapport-2016_21-resvaneundersokningar-som-indata-till-persontransportmodeller--problem-mojligheter-och-framtida-behov-i-sverige-och-norge.pdf [Hämtad 2017-12-20]

Trafikanalys (2017a), Televerksamhet 2016. Tillgänglig på www.trafa.se/globalassets/statistik/televerksamhet/2016/televerksamhet-2016.pdf [Hämtad 2017-12-18]

Trafikanalys (2017b), RVU Sverige – den nationella resvaneundersökningen 2015-2016: Kvalitetsdeklaration. Tillgänglig på <https://www.trafa.se/globalassets/statistik/resvanor/2016/rvu-sverige-2016-kvalitetsdeklaration.pdf> [Hämtad 2017-12-01]

Telia Sverige (2017), Telia crowd insights, Webbsida. Tillgänglig på: <https://www.teliacompany.com/sv/om-foretaget/internet-of-things/crowd-insights/> [Hämtad 2017-12-21]

Trivector (2017), Detaljerad kartläggning av verktyg och applikationer för resvaneundersökningar. Tillgänglig på <https://www.trafa.se/globalassets/rapporter/underlagsrapporter/2017/detaljerad-kartlaggning-av-verktyg-och-applikationer-for-resvaneundersokningar.pdf> [Hämtad 2017-12-17]

PTS (2017), Svensk telekommarknad 2016, Rapportnummer PTS-ER-2017:10, Diarienummer 16-11655, ISSN 1650-9862. Tillgänglig på http://statistik.pts.se/media/1254/svensk-telekommarknad-2016-_slutgiltig.pdf [Hämtad 2017-12-18]