

JOUNI PAAVILAINEN
ARTTU-MATTI MATINLAURI

Rautatieliikenteen täsmällisyystiedon jalostaminen kehittyneen data-analytiikan keinoin



Jouni Paavilainen, Arttu-Matti Matinlauri

Rautatieliikenteen täsmällisyystiedon jalostaminen kehittyneen data-analytiikan keinoin

Liikenneviraston tutkimuksia ja selvityksiä 55/2011

Liikennevirasto

Helsinki 2011

Kannen kuva: Tommi Mäkelä

Verkojulkaisu pdf (www.liikennevirasto.fi)

ISSN-L 1798-6656

ISSN 1798-6664

ISBN 978-952-255-073-6

Liikennevirasto

PL 33

00521 HELSINKI

Puhelin 020 637 373

Jouni Paavilainen, Arttu-Matti Matinlauri: Rautatieliikenteen täsmällisyystiedon jalostaminen kehittyneen data-analytiikan keinoin. Liikennevirasto, Liikenteenhallinta-toimiala. Helsinki 2011. Liikenneviraston tutkimuksia ja selvityksiä 55/2011. 105 sivua ja 3 liitettä. ISSN-L 1798-6656, ISSN 1798-6664, ISBN 978-952-255-073-6.

Avainsanat: rautatieliikenne, täsmällisyys, kehittynyt data-analytiikka, tiedonlouhinta, toimintatutkimus, päätöksenteon tuki

Tiivistelmä

Epätäsmällisyys heikentää sekä rautatieliikennejärjestelmän suorituskykyä että sen tarjoaman palvelun laatua. Tämä näkyy kustannuksina matkustajille, liikennöitsijöille, radanpitäjälle ja koko yhteiskunnalle. Asian kriittisyydestä huolimatta täsmällisyys on Suomessa laskenut jo useana vuonna, erityisesti kahden viime talven aikana. Ongelmien johdosta rautatietoimijat ovat ryhtyneet tekemään aktiivista työtä täsmällisyyden parantamiseksi. Tässä yhteydessä on huomattu, ettei täsmällisyyteen liittyvä tietopohja ole parhaalla mahdollisella tasolla: päätöksenteon tueksi tarvittaisiin nykyistä syvällisempää ja kattavampaa tietoa täsmällisyyttä heikentävistä ilmiöistä. Tämän tutkimuksen tavoitteena on selvittää, miten täsmällisyystiedon jalostamista päätöksenteon tueksi voitaisiin parantaa kehittyneen data-analytiikan menetelmin.

Kehittyneellä data-analytiikalla tarkoitetaan menetelmiä ja työkaluja, joilla pyritään automatisoidusti löytämään oleellinen tieto suurista tietomassoista. Rautatieliikennejärjestelmä muodostaa monimutkaisen ja vaikeasti hahmotettavan kokonaisuuden, jonka toiminnasta kerätään paljon dataa. Näin ollen kehittynyt data-analytiikka tarjoaa suuren potentiaalin järjestelmän analysointiin. Työn teoreettinen viitekehys tarkastelee analytiikkaa ennen kaikkea päätöksenteon tukemisen näkökulmasta. Konkreettisista menetelmistä pureudutaan assosiaatio- ja sekvenssianalyysiin, klusterointiin, luokitteluun sekä visuaaliseen analytiikkaan.

Tällä hetkellä Suomen rautatieliikenteen toimijoiden täsmällisyysanalyysit ovat hyvällä perustasolla. Ne tarkastelevat järjestelmää melko yleisellä tasolla, tarkastelusyklin ollessa pääsääntäisesti yksi kuukausi. Raportit antavat kuvan järjestelmän yleisestä suoriutumisesta, mutta niiden käyttäminen päätöksenteon tueksi vaatii yleensä lisäselvityksiä. Datan ja sen analysoinnin nykytila ja siihen liittyvät haasteet ja mahdollisuudet selvitettiin osallistumalla aktiivisesti Liikenneviraston ja VR:n täsmällisyystyötä tekeville foorumeille sekä analyysitoimintaan.

Kehittyneen data-analytiikan teoreettisen viitekehysten ja nykyisessä analysointitoiminnassa tunnistettujen haasteiden pohjalta työssä toteutettiin useita erilaisia esimerkkianalyyskejä hyödyntäen kehittyneen data-analytiikan menetelmiä. Analyysien tarkoituksena on osoittaa, mitä mahdollisuuksia menetelmät tarjoavat tiedontuotantoprosessin kehittämiseksi ja edelleen päätöksenteolle. Lisäksi esimerkkianalyysit paljastavat, miten tietojärjestelmiä ja analyysiympäristöjä tulisi kehittää, jotta analysoinnista saataisiin paras mahdollinen hyöty.

Kehittyneitä data-analytiikka-menetelmiä hyödyntämällä olisi mahdollisuus tuottaa huomattavasti nykyistä parempaa ja syvällisempää tietoa päätöksenteon tueksi. Menetelmillä analysointi on tehokasta ja nopeaa. Näin aikaa jää analyysien tulkinnalle. Menetelmien täysipainoinen hyödyntäminen kuitenkin edellyttää kehittämistä niin datan ja dataympäristön laadun, analyysitoiminnan kuin tuotetun tiedon hyödyntämisenkin osalta. Tähän kehitysohjelmaan tämä raportti antaa erinomaiset eväät.

Jouni Paavilainen, Arttu-Matti Matinlauri: Förädling av punktlighetsinformation i järnvägstrafiken genom avancerad dataanalys. Trafikverket, trafikledning. Helsingfors 2011. Trafikverkets undersökningar och utredningar 55/2011. 105 sidor och 3 bilagor. ISSN-L 1798-6656, ISSN 1798-6664, ISBN 978-952-255-073-6.

Nyckelord: järnvägstrafik, punktlighet, avancerad dataanalys, datautvinning, verksamhetsundersökning, stöd för beslutsfattande

Sammanfattning

Opunktlighet försvagar både järnvägstrafiksystemets prestationsförmåga och kvaliteten på den service som det erbjuder. Detta märks i form av kostnader för passagerare, trafikanter, banhållaren och hela samhället. Trots att det här är en kritisk fråga har punktligheten försämrats i Finland redan i flera år, och särskilt under de två senaste vintrarna. På grund av problemen har järnvägsaktörerna börjat satsa aktivt på att förbättra punktligheten. I sammanhanget har det kommit fram att kunskapsunderlaget i anslutning till punktlighet inte håller bästa möjliga nivå: till stöd för beslutsfattandet skulle behövas djupare och mera omfattande information om de fenomen som försvagar punktligheten. Syftet med denna undersökning är att reda ut hur förädlad punktlighetsinformation till stöd för beslutsfattandet skulle kunna förbättras med metoder för avancerad dataanalys.

Med avancerad dataanalys avses metoder och verktyg med vilka man automatiserat försöker hitta den väsentliga informationen i stora datamassor. Järnvägstrafiksystemet bildar en komplicerad och svårgreppad helhet, från vars verksamhet mycket data samlas in. Härmed erbjuder avancerad dataanalys en stor potential för analys av systemet. Arbetets teoretiska referensram granskar analysen framför allt ur perspektivet stöd till beslutsfattandet. Via konkreta metoder sätter man sig in i associations- och sekvensanalys, klusterbildning, klassificering samt visuell analys.

För närvarande håller järnvägsaktörernas punktlighetsanalyser i Finland en god grundnivå. Dessa granskar systemet på en ganska allmän nivå, kontrollcykeln är i regel en månad. Rapporterna ger en bild av hur systemet fungerar generellt, men för att använda dem till stöd för beslutsfattandet krävs i allmänhet extra utredningar. Det nuvarande tillståndet vad gäller data och analyseringen av data samt de relaterade utmaningarna och möjligheterna har utretts genom aktivt deltagande i Trafikverkets och VR:s forum för punktlighetsarbete samt genom analysverksamhet.

Utifrån den teoretiska referensramen för avancerad dataanalys och de svårigheter som har identifierats i den nuvarande analysverksamheten genomfördes i arbetet flera olika exempelanalyser där man drog nytta av metoderna för avancerad dataanalys. Syftet med analyserna är att visa vilka möjligheter metoderna erbjuder för att utveckla dataproduktionsprocessen och för beslutsfattandet. Dessutom avslöjar exempelanalyserna hur datasystemen och analysmiljöerna borde utvecklas för att analysen ska ge bästa möjliga nytta.

Genom att utnyttja metoderna för avancerad dataanalys vore det möjligt att producera betydligt mera och djupare information till stöd för beslutsfattandet. Med dessa metoder blir analyserna effektiva och snabba och tid blir över för tolkning av dem. Ett fullgott utnyttjande av metoderna förutsätter dock att kvaliteten i data och datamiljö, analysverksamheten samt utnyttjandet av den producerade informationen utvecklas. Denna rapport ger utomordentliga möjligheter för det här utvecklingsarbetet.

Jouni Paavilainen, Arttu-Matti Matinlauri: Refining data on rail transport punctuality by means of sophisticated data analysis. Finnish Transport Agency, Traffic Management. Helsinki 2011. Research reports of the Finnish Transport Agency 55/2011. 105 pages and 3 appendices. ISSN-L 1798-6656, ISSN 1798-6664, ISBN 978-952-255-073-6.

Keywords: rail transport, punctuality, sophisticated data analysis, data mining, action research, decision support

Summary

Lack of punctuality is detrimental to the performance of the rail transport system and its quality of service. This will bring additional costs for the passengers, transport operators, rail maintenance organisations, and society at large. While a critical factor, the punctuality of the Finnish rail transport system has been on the decline for several consecutive years, especially over the past two winters. Alarmed by these problems, parties involved in rail transport have engaged in active co-operation to improve punctuality. In this context, it has been widely recognised that the various parties' understanding of punctuality rests on a less than perfect foundation: in fact, information on factors that impair punctuality is required, more comprehensive and more in-depth than what is currently available, to decision support. The purpose of this study is to examine how sophisticated data analysis techniques can be used to improve the refinement of data concerning punctuality, in decision support.

Sophisticated data analysis techniques employ methods and tools that, via automated processes, are aimed at obtaining relevant information from large masses of data. A rail transport network is a complex system, and gaining an overview of such an infrastructure is a difficult task; furthermore, such systems generate large volumes of data. Therefore, sophisticated data analysis holds great potential for examination of the system. The theoretical framework underpinning the present study examines data analysis from the standpoint of the support it offers policymakers. With respect to concrete methods, the study focuses on associative and sequential analyses, clustering, classification, and visual analysis.

At present, the punctuality analyses produced by the stakeholders in the Finnish rail transport system are at basic reporting level. They take a general view of the system, chiefly with a one-month review cycle. The reports provide an overall picture of the system's performance, but they do not form a sufficient basis for decision-making, which generally requires additional surveys. The current state of the data and of the data analysis, with the associated challenges and opportunities, were clarified through active participation in the forums and analysis projects co-ordinated by the Finnish Transport Agency and the VR Group.

Building on the theoretical framework of sophisticated data analysis and the challenges identified in current analysis methods, the present study modelled several sample analyses with the aid of sophisticated data analysis methodology. The sample analyses were intended to identify the opportunities that the analysis methods offer for further development of the data-gathering process and, ultimately, of the decision-making. In addition, these analyses reveal how data systems and analysis environments should be developed in order that the best possible outcomes can be optimised.

Through utilisation of sophisticated data analysis techniques, opportunities would arise to produce significantly higher-quality and more in-depth data than the data currently available, thus facilitating policymaking. Performing data analysis that utilises explicit methods is both quick and inexpensive. This leaves the researchers more time to interpret analysis results. Exploiting data analysis methods to their fullest requires further development, both with respect to the quality of the data and the data environment and in terms of utilisation of the data thus produced. This report provides an excellent basis for such development work.

Esipuhe

Täsmällisyys on rautatieliikennejärjestelmän suorituskyvyn ja palvelun laadun keskeinen osatekijä. Niinpä täsmällisyyden eri osa-alueita seuraamalla voidaan arvioida rautatiejärjestelmän kokonaistoimivuutta, radanpitäjän ja rautatieoperaattorin onnistumista tehtävissään sekä asiakkaille tarjotun palvelun laatua.

Rautatieliikennejärjestelmä on monimutkainen kokonaisuus, minkä vuoksi täsmällisyyden luotettava ja seikkaperäinen analysointi on vaativa tehtävä. Siinä onnistuminen vaatii paljon sekä lähtötiedoilta että analysoinnilta. Viime vuosina on huomattu, ettei tämä tiedonjalostusprosessi ole parhaalla mahdollisella tasolla. Tämä hankaloittaa rautatieliikennejärjestelmän kehittämistä, sillä ilman luotettavia ja kattavia analyysejä on kehitystoimien oikea kohdistaminen hankalaa tai jopa mahdotonta.

Tämän tutkimuksen tavoitteena oli selvittää, miten täsmällisyystiedon jalostamista päätöksenteon tueksi voidaan parantaa kehittyneen data-analytiikan menetelmin. Tuloksena on sekä uusia, tehokkaita analysointitapoja että suosituksia lähtötietoihin liittyen. Tutkimus tarjoaakin tärkeää tietoa sekä tietojärjestelmien että analysointiprosessien kehittäjille.

Tutkimus on osa Liikenneviraston ja Tampereen teknillisen yliopiston välistä tutkimusyhteistyötä, jonka pääteemana on rautatieliikenteen täsmällisyys. Tutkimuksen tekivät ja tämän raportin kirjoittivat Tampereen teknillisen yliopiston projektipäällikkö Jouni Paavilainen ja tutkija Arttu-Matti Matinlauri. Tutkimuksen ohessa Matinlauri laati diplomityönsä samasta aihepiiristä.

Tutkimus tehtiin tiiviissä yhteistyössä Liikenneviraston kanssa hyödyntäen Liikenneviraston asiantuntemusta ja tietovarantoja. Tutkimusta ohjasi Liikenneviraston toiminnan ohjaus ja verkon käyttö -yksikön päällikkö Heli Mattila. Lisäksi Pertti Tapola Liikennevirastosta ja Egon Blomqvist VR:ltä antoivat työhön merkittävän asiantuntijapanoksensa. Työ aloitettiin joulukuussa 2010 ja saatiin päätökseen vuotta myöhemmin.

Helsingissä joulukuussa 2011

Liikennevirasto
Liikenteenhallinta-toimiala

Sisällysluettelo

1	JOHDANTO	9
1.1	Tutkimuksen tausta ja asemointi	9
1.2	Toimintaympäristö	10
1.2.1	Suomen rautatieliikennejärjestelmä	10
1.2.2	Rautatieliikennejärjestelmän täsmällisyys	12
1.2.3	Rautatieliikenteen toimijat Suomessa.....	12
1.3	Tutkimuksen tavoitteet ja rajaukset	13
1.4	Tutkimusmenetelmä.....	14
1.5	Tutkimuksen rakenne	16
2	KEHITTYNYT DATA-ANALYTIikka PÄÄTÖKSENTEON TUKENA	17
2.1	Päätöksentekomenetelmät.....	17
2.2	Tiedon tasot.....	20
2.3	Kehittyneen data-analytiikan menetelmät	21
2.3.1	Assosiaatioanalyysi & sekvenssianalyysi.....	24
2.3.2	Klusterointi	26
2.3.3	Luokittelu	28
2.3.4	Visuaalinen analytiikka	29
2.4	Yhteenveto	30
3	RAUTATIELIIKENTEEN TÄSMÄLLISYYSTIEDON JALOSTAMISEN JA HYÖDYNTÄMISEN NYKYTILA	31
3.1	Täsmällisyyden perusseuranta	31
3.2	Viiveiden luokitteluperiaate	33
3.3	Data ja analyysiympäristö	36
3.3.1	JUSE-data.....	36
3.3.2	GRATU-data.....	38
3.3.3	Arvio datan ja dataympäristön nykytilasta	40
3.4	Analyysit	41
3.4.1	Kuukausittaiset analyysit	41
3.4.2	Vuosittainen analyysi	43
3.4.3	Ad hoc -analyysit	44
3.4.4	Arvio analyysien nykytilasta	44
3.5	Työryhmät.....	44
3.5.1	VR:n Täsmällisyyden ohjausryhmä ja Ryhti-ryhmä	45
3.5.2	VR:n Täsmä-ryhmä	45
3.5.3	Liikenneviraston Infra-täsmäryhmä.....	46
3.5.4	VR:n Liikenteen laatu -ryhmät (LiLa-ryhmät)	46
3.5.5	Liikenneviraston alueelliset infratäsmäryhmät.....	47
3.5.6	VR:n Täsmällisyyden analyysiryhmä (A-ryhmä)	47
3.5.7	Arvio työryhmien nykytilasta	48
4	KEHITTYNEEN DATA-ANALYTIIKAN MENETELMIEN HYÖDYNTÄMINEN TÄSMÄLLISYYSDATAN ANALYSOINNISSA	49
4.1	Yksittäisen junan toistuvat viiveet	49
4.1.1	Väite analyysin taustalla.....	49
4.1.2	Analyysi	49
4.1.3	Päätelmät	62
4.2	Yksittäisen ilmiön vaikutus koko maahan.....	62
4.2.1	Väite analyysin taustalla.....	62

4.2.2	Analyysi.....	62
4.2.3	Päätelmät.....	78
4.3	Matkustajapalvelun viiveiden vaikutus	79
4.3.1	Väite analyysin taustalla	79
4.3.2	Analyysi.....	79
4.3.3	Päätelmät.....	84
4.4	Helsingin alueen ruuhkaisuus lähiliikenteen näkökulmasta	84
4.4.1	Väite analyysin taustalla	84
4.4.2	Analyysi.....	84
4.4.3	Päätelmät.....	88
4.5	Viiveiden visualisointi karttapohjaisella animaatiolla	89
4.6	Vuosiraportin analyysien kehittäminen	90
4.7	Kuukausiraportoinnin automatisointi	91
5	PÄÄTELMÄT JA SUOSITUKSET	93
5.1	Kehittyneen data-analytiikan mahdollisuudet.....	93
5.2	Data- ja analytiikkaympäristön kehittäminen	94
5.2.1	Kulktietojärjestelmän tarkoituksenmukaisuuden parantaminen.....	95
5.2.2	Kulktietojärjestelmään liittyvien kirjaamiskäytäntöjen kehittäminen.....	97
5.3	Analyysitoiminnan kehittäminen	99
5.3.1	Analysointiprosessin systematisointi.....	99
5.3.2	Yleisimpien ongelmatyyppien analysointi.....	101
5.4	Työryhmien toiminnan kehittäminen	103
5.5	Tutkimuksen arviointi ja jatkotutkimusehdotukset.....	103
	LÄHTEET	104
	LIITTEET	
Liite 1	JUSEn aktiivikanta	
Liite 2	JUSEn historiakannan taulut	
Liite 3	Työryhmiin osallistuminen ja esiin nousseet aihealueet	

1 Johdanto

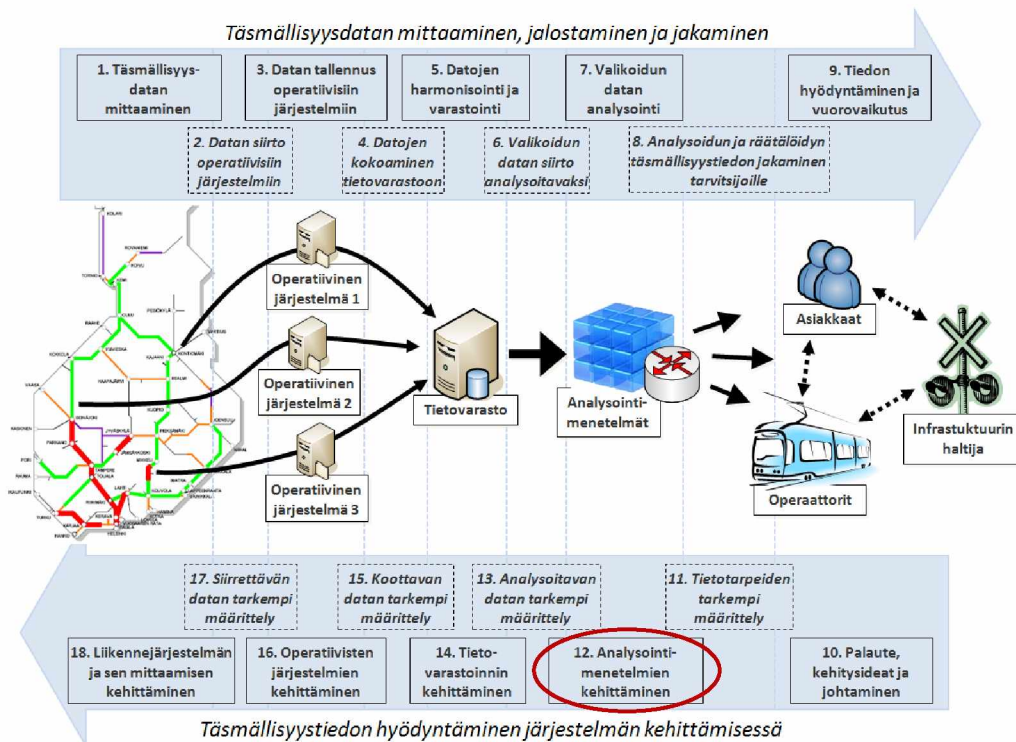
1.1 Tutkimuksen tausta ja asemointi

Rautatieliikenteen täsmällisyyden merkitys on kasvanut viime vuosikymmeninä muun muassa ihmisten vaatimustason noususta, kulkumuotojen aiempaa voimakkaammasta keskinäisestä kilpailusta sekä rataverkon käyttöasteen kasvusta johtuen. Epätasällisyys heikentää sekä rautatieliikennejärjestelmän suorituskykyä että sen tarjoaman palvelun laatua. Tätä kautta se heijastuu kustannuksina matkustajille, liikennöitsijöille, radanpitäjälle ja koko yhteiskunnalle. Asian kriittisyydestä huolimatta Suomen rautateiden täsmällisyys on laskenut jo useana vuonna. Erityisesti ankarien talvien 2009–2010 ja 2010–2011 aikana se painui erittäin huonolle tasolle.

Kielteisen kehityksen vuoksi asiaan on viime vuosina puututtu monin eri tavoin. Yksi näistä tavoista on vuonna 2009 alkunsa saanut Liikenneviraston ja TTY:n välinen, täsmällisyyteen keskittyvä tutkimusyhteistyö. Yhteistyön aikana on selvitetty muun muassa *täsmällisyyteen liittyvän tieteellisen tutkimuksen nykytilaa, eri toimijoiden täsmällisyyteen liittyviä tietotarpeita* sekä *eri maiden konkreettisia käytäntöjä täsmällisyysdatan keräämiseen, analysointiin ja hyödyntämiseen liittyen*.

Edellä mainituissa tutkimuksissa on havaittu, että kehittyneiden data-analytiikkamenetelmien hyödyntäminen täsmällisyysdatan analysoinnissa antaisi mahdollisuuden eri toimijoiden tietotarpeiden parempaan tyydyttämiseen. Syystä tai toisesta näitä menetelmiä ei kuitenkaan tällä alueella ole juuri hyödynnetty, ei Suomessa eikä myöskään kansainvälisesti. Myös asian tieteellinen tutkimus on melko vähäistä. Nyt käsillä oleva tutkimushanke, *Täsmällisyystiedon jalostaminen kehittyneen data-analytiikan keinoin*, pyrkii täyttämään tätä aukkoa. Tutkimus sekä palvelee konkreettista täsmällisyystyötä että luo uutta tietoa alan tieteelliselle kentälle.

Kuvassa 1 on esitetty malli, joka havainnollistaa täsmällisyyteen liittyvän datan keräämiseen, analysointiin ja hyödyntämiseen liittyvää iteratiivista prosessia. Reaali maailman prosessin yleisen jäsentämisen lisäksi mallia voidaan hyödyntää tutkimuksen asemoinnissa. Tämä tutkimushanke tarkastelee ensisijaisesti prosessin vaihetta 12 (*analysointimenetelmien kehittäminen*). Analysointimenetelmien kehittämisessä pyritään ottamaan kantaa sekä datan muokkaukseen että analysointiin liittyviin asioihin. Tutkimus ei kuitenkaan rajaudu vain tähän vaiheeseen, vaan sen kautta pyritään lisäksi hahmottamaan, kuinka koko tiedontuotantoprosessia tulisi kehittää.



Kuva 1. *Prosessimalli, jossa mitattua täsmällisyysdataa jalostetaan tarvitsijoille, ja hyödynnetään heiltä saatua palautetta järjestelmän kehittämisessä. (muokattu lähteestä Paavilainen et al. 2011)*

1.2 Toimintaympäristö

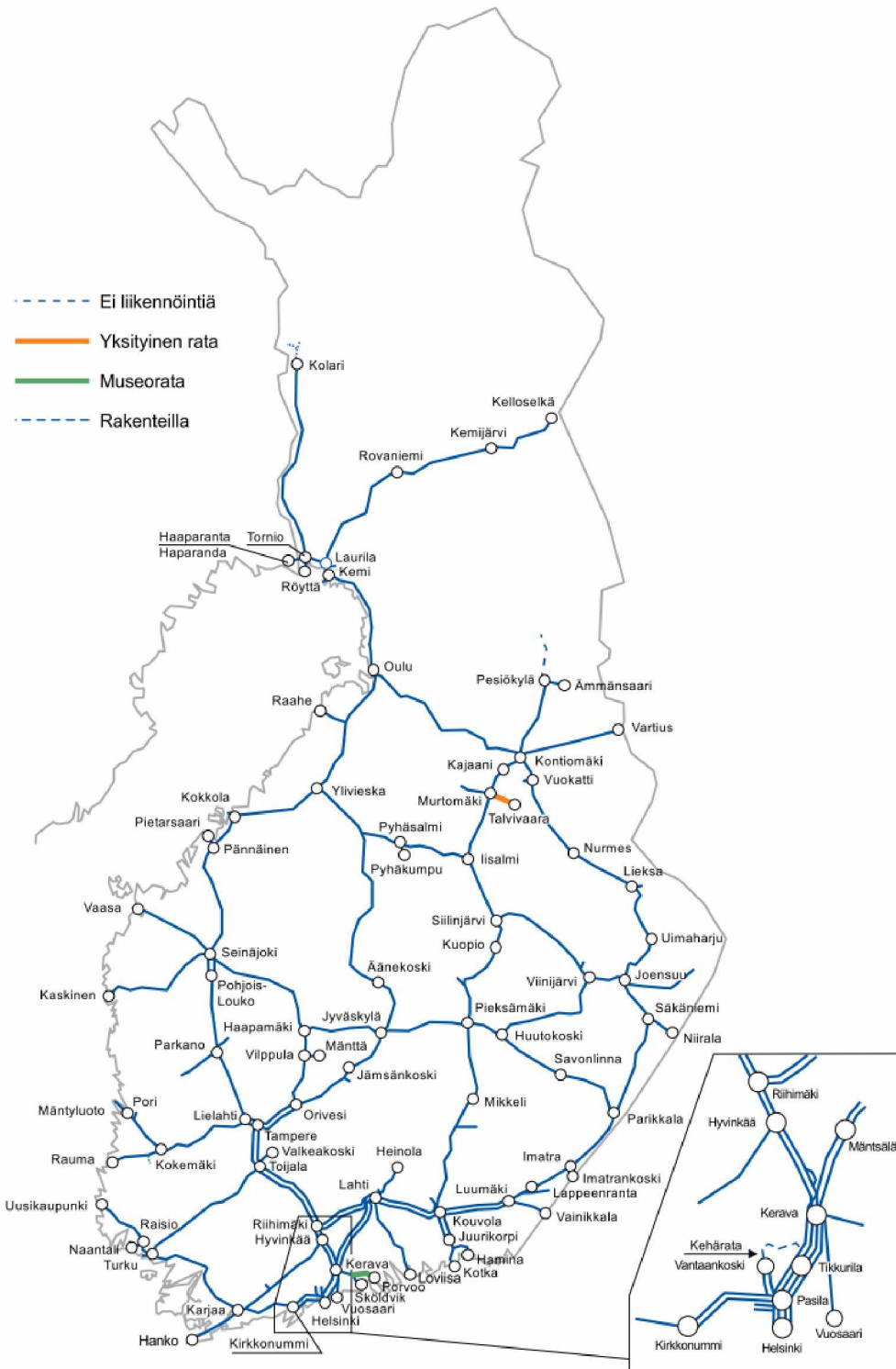
1.2.1 Suomen rautatieliikennejärjestelmä

Rautatieliikenne on luonteeltaan hyvin poikkeava muihin liikennemuotoihin verrattuna. Se on energiatehokas ja ekologinen liikennemuoto: rautateillä kyetään liikuttamaan suuria massoja ihmisiä ja tavaraa pienin resurssein. Toisaalta rautatieliikenne on vahvasti sitoutunut raiteisiin, joka rajoittaa sitä, mihin junat voivat liikkua. Rautateillä liikenteen turvallisuus on ensisijaisen tärkeää ja liikennettä ohjataan ja valvotaan monin erilaisin teknisin laittein.

Raiteet määrittelevät, mihin juna voi mennä. Näin ollen junilla ei ole vaihtoehtoisten reittien käyttämahdollisuuksia ja ohitusmahdollisuuksia samalla tavalla kuin esimerkiksi tieliikenteessä. Tämä ominaisuus tekee rautatieliikenteestä erittäin häiriöherkän liikennemuodon: yksikin hajonnut juna tai jumiutunut vaihde tukkia laajojen alueiden liikenteen. Haaste korostuu Suomessa, jossa rataverkko on yli yhdeksänkymmentäprosenttisesti yksiraiteista. Täten rautatieliikenteessä tulisikin pyrkiä pitämään häiriöt minimissä. Vain tällä tavoin voidaan taata järjestelmän luonteesta johtuvien kerrannaisvaikutusten pysyminen aisoissa.

Suomessa pääosin yksiraiteisen rataverkon pituus on noin 5800 kilometriä. Kuten kuvasta 2 voidaan havaita, useampiraiteisia osuuksia on olemassa ainoastaan Helsingistä Kirkkonummelle, Helsingistä Tampereen kautta Orivedelle, Helsingistä Lahden ja Kouvolan kautta Luumäelle ja Juurikorpeen sekä Riihimäen ja Lahden välillä. Hie-

man yli puolet kaikesta radasta on sähköistettyä. Suomen rataverkolla liikkuu päivittäin noin 300 henkilökaukoliikenteen, 900 Helsingin seudun lähiliikenteen ja noin 450 tavaraliikenteen junaa. Tällä hetkellä ainoalla operaattorilla, VR-Yhtymällä, on käytössään noin 400 veturia, 1000 henkilöliikenteen vaunua ja yli 10 000 tavaravaunua tämän liikennöinnin toteuttamiseen. (Liikennevirasto 2010).



Kuva 2. Suomen rataverkko (Liikennevirasto 2011).

1.2.2 Rautatieliikennejärjestelmän täsmällisyys

Täsmällisyyttä ja siihen tiiviisti liittyviä käsitteitä on käsitelty laajasti aiemmissa TTY:n ja Liikenneviraston yhteistyöprojektien tutkimusraporteissa (esim. Salkonen et al. 2009, Salkonen & Mäkelä 2010, Liikennevirasto 2010, Paavilainen et al. 2011). Näin ollen tähän tutkimukseen on poimittu näistä raporteista vain termien tiiviit määritelmät. Tärkein termi on luonnollisesti täsmällisyys:

Täsmällisyys (punctuality) viittaa junan kykyyn liikennöidä ennalta ilmoitetun aikataulun mukaisesti. Täsmällisyys käsitetään yleensä diskreettinä muuttujana, joka kuvaa, alittaako juna aikataulupoikkeamille asetetut hyväksyttävät maksimiarvot. Jos juna liikennöi maksimiarvojen sisällä, se on täsmällinen, muussa tapauksessa ei. Junan täsmällisyys määritellään yleensä ennalta määritellyissä mittauspisteissä, kuten junan saapuessa tai lähtiessä asemalta. Käytännössä täsmällisyyttä voidaan mitata monin eri kriteerein. Suomen käytäntöjä on valotettu luvussa 3.

Täsmällisyyteen tiiviisti liittyviä muita käsitteitä ovat:

Viive (delay) on aikatauluajan ja toteutuneen ajan välinen ero tietyssä mittauspisteessä. Viivettä mitataan ajan yksiköissä, yleensä minuuteissa. Viivettä voi olla sekä jäljessä kulku (arvo positiivinen) että etuajassa kulku (arvo negatiivinen). Vaikka junalla olisi viivettä, eli se kulkisi aikataulustaan jäljessä, ei se silti välttämättä ole myöhästynyt.

Primääriviive (primary delay) tarkoittaa junan alkuperäistä viivettä, joka ei ole aiheutunut muista junista. Primääriviiveet voivat aiheutua niin kaluston, radanpidon, matkustajista, sääoloista, onnettomuuksista tai muista toisista junista riippumattomista syistä.

Sekundääriviive (secondary delay) on yhden junan myöhästyminen, joka johtuu toisen junan alkuperäisestä myöhästymisestä. Viive voi syntyä esimerkiksi sen takia, että toinen juna varaa tietyn rataosuuden, asemapaikan tai kaluston. Sekundäärisiä viiveitä kutsutaan yleensä ketjuuntuneiksi viiveiksi, koska ne johtuvat toisten junien aiheuttamista viiveistä.

Myöhästyminen tarkoittaa negatiivista poikkeamaa, eli positiivista viivettä, aikatauluajasta tietyssä mittauspisteessä. Tämän viiveen suuruuden tulee ylittää ennakkoon määritellyt raja-arvo, jotta sitä voidaan kutsua myöhästymiseksi.

1.2.3 Rautatieliikenteen toimijat Suomessa

Rautatieliikennejärjestelmästä – rautatieverkosta, sillä liikennöivistä junista sekä näiden ohjaus- ja informaatiojärjestelmistä – vastaavat useat eri toimijat. *Liikennevirasto* on vastuussa Suomen valtion rataverkon rakentamisesta, ylläpitämisestä ja kehittämisestä. Liikennevirasto vastaa myös rataverkon turvallisuudesta sekä ratakapasiteetin jakamisesta ja liikenteenohjauksesta. Virasto ostaa ratojen kunnossapito- ja rakentamistyöt sekä osan ratojen suunnittelusta ja liikenteenohjauksen palveluista ulkopuolisilta yrityksiltä. Rautatieliikenteen sujumista valvoo ja koordinoi ympäri vuorokauden Liikenneviraston Rautatieliikennekeskus. Lisäksi Liikennevirasto vastaa matkustajainformaatiosta asema- ja laiturinäyttöjen ja -kuulutusten osalta. (Liikennevirasto 2010).

Tällä hetkellä Suomen rataverkolla liikennöi yksi rautatieyritys, *VR-Yhtymä Oy*. Kyseessä on Suomen valtion kokonaan omistama osakeyhtiö, joka perustettiin vuonna 1995 jatkamaan Valtionrautateiden toimintaa. Yhtiöittämisen myötä vastuu rataverkosta siirtyi viranomaisille (nykyiset Liikennevirasto ja *Liikenteen turvallisuusvirasto TraFi*) ja vastuu liikenteen harjoittamisesta jäi VR:lle. Rautatieliikennöitsijänä VR-Yhtymä Oy vastaa junakalustosta ja sen kunnossapidosta sekä juna- ja asemahenkilöstöstä. Lisäksi matkustajainformaatio junissa on sen vastuulla. VR-konserniin kuuluu myös mm. *Oy VR-Rata Ab*, joka tarjoaa ratojen kunnossapidon, rakentamisen ja suunnittelun palveluita. (Liikennevirasto 2010).

1.3 Tutkimuksen tavoitteet ja rajaukset

Tutkimuksen tavoitteena on selvittää, miten täsmällisyystiedon jalostamista päätöksenteon tueksi voidaan parantaa kehittyneen data-analytiikan menetelmin.

Kehittyneellä data-analytiikalla tarkoitetaan tässä tutkimuksessa menetelmiä ja työkaluja, joilla pyritään automatisoidusti löytämään oleellinen tieto suurista tietomassoista. Kun perinteisten data-analyysimenetelmien ja -työkalujen avulla on lähinnä mahdollista analysoida pieniä datamääriä muutaman muuttujan suhteen kerrallaan, on kehittyneen data-analytiikan avulla mahdollista tarkastella laajojakin datamassoja useiden satojen muuttujien suhteen kehittyneiden analyysialgoritmien ja -työkalujen avulla. Näin ollen kehittyneempiä menetelmiä hyödynnettäessä ei tarvita valmiita hypoteeseja, vaan ideana on, että työkalut löytävät datamassasta johdonmukaisuuksia, kuten esimerkiksi poikkeamia, säännönmukaisuuksia sekä syy-seuraussuhteita. (Kudyba & Hoptroff 2001, Giudici 2003)

Tutkimuksen tavoitteet voidaan tiivistää seuraaviin osa-alueisiin:

1. Kirjallisuusselvityksen avulla arvioidaan, miten analytiikalla voidaan tukea päätöksentekoa yleisellä tasolla
2. Kirjallisuusselvityksen avulla tarkastellaan, mitkä ovat keskeisimmät kehittyneen data-analytiikan menetelmät ja miten niitä hyödynnetään
3. Toimintatutkimuksena kartoitetaan, miten kehittyneen data-analytiikan menetelmiä voisi hyödyntää rautatieliikennejärjestelmää ja siihen liittyvää analyysitoimintaa kehittäessä
4. Case-tutkimuksena toteutetaan konkreettisia mallianalyysejä kehittyneen data-analytiikan menetelmin, ja arvioidaan niiden kautta menetelmien potentiaalia rautatieorganisaatioiden päivittäisessä analyysityössä.

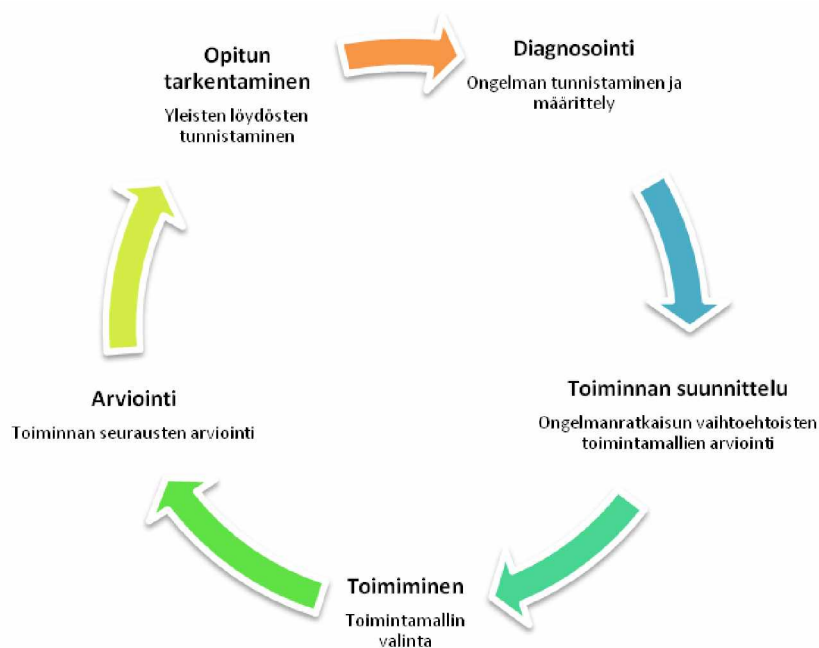
Tutkimuksessa osallistutaan kohdeorganisaatioiden – Liikenneviraston ja VR:n – toimintaan aktiivisesti. Tämä ilmenee muun muassa erinäisten vakioanalyysien kehittämisenä sekä analyysejä hyödyntävien työryhmien toimintaan osallistumisena. Tutkimuksen fokus on kehittyneen data-analytiikan menetelmien hyödyntämisessä päätöksenteon tukena. Näin ollen esimerkiksi päätöksentekijöiden tietotarpeisiin ei puututa kuin siinä määrin, että voidaan arvioida esimerkianalyysien hyödyllisyyttä.

1.4 Tutkimusmenetelmä

Tutkimus edellyttää kohdeorganisaation analytiikkaan liittyvien toimintatapojen ja tarpeiden ymmärtämistä. Tämän vuoksi tutkimus toteutettiin interventionistisena tutkimuksena eli toimintatutkimuksena. Interventionistisen tutkimuksen pääajatus, kohdeorganisaation toiminnasta oppiminen ja mahdollisesti sen toiminnan kehittäminen aktiivisen osallistumisen myötä, esitetään seuraavaksi tarkemmin.

Baskerville & Myers (2004) esittävät, että interventionistinen tutkija on yleensä organisaation ulkopuolinen henkilö, joka tarjoaa asiantuntemustaan kohdeorganisaation kehittämiseen ja systemaattisesti arvioi osallistumista rakentaakseen ymmärrystä organisaation toiminnasta. Levin et al. (2002) mukaan interventionistisessä tutkimuksessa on kuitenkin se haaste, että ulkopuolinen tutkija ei välttämättä täysin nauti kohdeorganisaation luottamusta, jolloin häneltä jää saamatta kaikista luottamuksellisiin tietoihin. Iversen et al. (2004) painottavatkin, että interventionistisessä tutkimuksessa asiakas–tutkija-suhde on yksi suurimmista vaikuttimista tutkimuksen onnistumisen kannalta.

Kuvassa 3 on esitetty Susman & Everedin (1978) kehittämä toimintatutkimuksen syklinen prosessi. Heidän mukaansa toimintatutkimus koostuu viidestä vaiheesta: *diagnosointi*, *toiminnan suunnittelu*, *toimiminen*, *arviointi* ja *opitun tarkentaminen*. *Diagnosointi*-vaiheessa pyritään tunnistamaan ratkaistava ongelma ja määrittelemään se tarkemmin. *Toiminnan suunnittelu* -vaiheen tavoitteena on tuottaa vaihtoehtoisia toimintamalleja ongelmanratkaisuun ja arvioida niitä. Tämä on tärkein vaihe prosessin onnistumisen kannalta. *Toimiminen*-vaiheessa valitaan yksi kehitetty toimintamalli ja ajetaan se organisaation käyttöön. *Arviointi*-vaiheessa tarkastellaan toiminnan vaikutuksia organisaatioon. Viimeinen vaihe, *opitun tarkentaminen*, on tutkimuksen kannalta tärkein. Siinä tarkastellaan yleisiä löydöksiä ja pyritään vetämään johtopäätöksiä siitä, miten mikäkin asia vaikuttaa kohdeorganisaation toimintaan. (Susman & Evered 1978).



Kuva 3. Toimintatutkimuksen syklinen prosessi (mukailtu lähteestä Susman & Evered 1978).

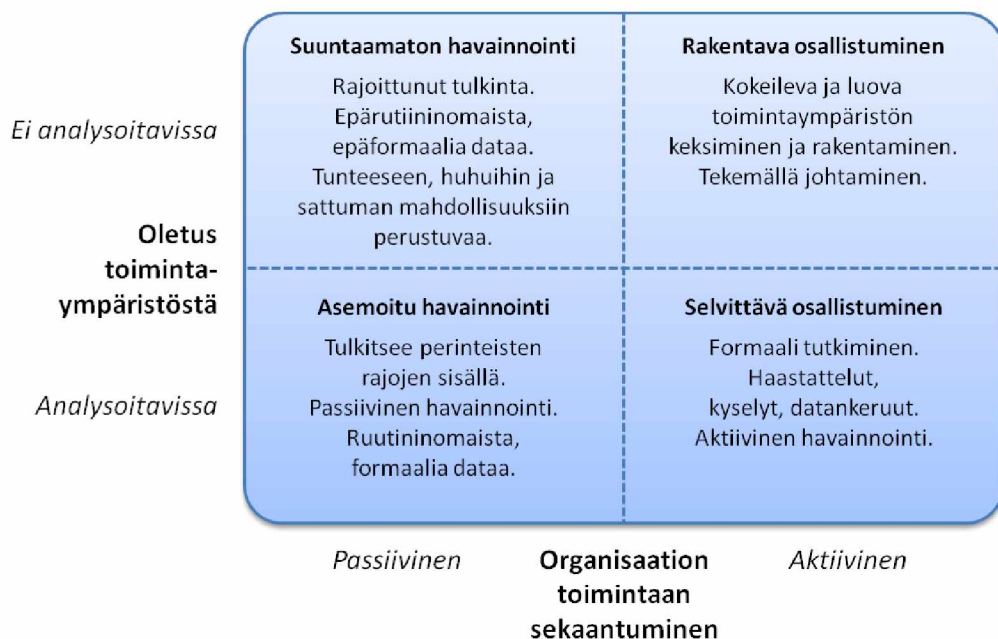
Tämän tutkimuksen voidaan nähdä sijoittuvan toimintatutkimuksen syklisen prosessin eri vaiheisiin seuraavasti:

Diagnosointi. Tutkimuksessa analysoidaan aluksi, miten tällä hetkellä tehdään täsmällisyysanalyysiä. Arvioinnin kohteena on sekä käytettävä data että tehtävät analyysit. Tämän lisäksi arvioidaan analyysien käyttöä päätöksenteon tukena.

Toiminnan suunnittelu & toimiminen. Tutkimuksen painopiste on tehtävissä case-analyyseissä, sillä lähtöoletuksena on, että olemassa olevat vakioanalyysit voisivat olla sisällöllisesti rikkaampia. Vakioanalyysiprosessin kehittämisen ohella arvioidaan, kuinka ad-hoc-analyysien tuottamista voitaisiin edistää. Tutkimuksessa tehtävät esimerkkianalyysit valitaan siten, että ne tarjoavat uutta tietoa konkreettisiin ongelmiin. Tarkasteltavat ongelmat puolestaan valitaan sen perusteella, mitä nousee esille, kun tutkijat osallistuvat kohdeorganisaatioiden toimintaan.

Arviointi & opitun tarkentaminen: Tutkimuksessa tuotetaan sellaisia esimerkkianalysejä, joita kohdeorganisaatiot voisivat kehittyneen data-analytiikan menetelmien avulla tuottaa päivittäisessä työssään. Näitä analyysejä ja niiden hyödyllisyyttä arvioidaan kunkin analyysin kohdalla erikseen. Arviointi suoritetaan subjektiivisesti tutkijoiden ja kohdeorganisaatioiden henkilöstön mielipiteiden mukaan. Arviointiin vaikuttaa ainakin se, kuinka hyödyllisiä analyysejä nähdään olevan ja kuinka yksinkertaisesti niitä saadaan luotua.

Toimintatutkimusta voi tehdä monella eri tavalla. Toimintatutkija voi Daft & Weickin (1984) mukaan osallistua organisaation toimintaan ja nähdä tutkittavan toimintaympäristön eri tavoin. Kuvassa 4 on esitetty malli organisaation tulkinnan eri moodeista. Sen mukaan organisaation toimintaan voidaan sekaantua joko aktiivisesti tai sitten tyytyä vain passiivisesti havainnoimaan toimintaa. Lisäksi toimintaympäristö voidaan nähdä joko helposti analysoitavana tai sitten sellaisena, jota ei kyetä analysoimaan, jolloin täytyy tehdä oletuksia tai jopa rakentaa toimintaympäristö itse.



Kuva 4. Organisaation tulkinnan moodit (mukailtu lähteestä Daft & Weick 1984).

Tässä tutkimuksessa painopiste on selvittävän osallistumisen moodissa johtuen siitä, että osallistuminen kohdeorganisaatioiden toimintaan on aktiivista ja oletuksena on, että toimintaympäristö on analysoitavissa. Toisin sanoen tutkijat nostavat toimintaympäristöstä aktiivisesti esille selvitettäviä asioita ja sen jälkeen analysoivat niitä erilaisin menetelmin.

Tutkimuksen aikana kuitenkin nousee esille mielenkiintoisia aiheita, joista ei ole saatavissa tarkempaa tietoa, eli toisin sanoen toimintaympäristö ei ole niiden osalta analysoitavissa. Näiltä osin tutkimus on jossakin määrin luonteeltaan myös rakentavaa osallistumista. Tästä esimerkkinä toimivat tutkimuksen aikana esiin nousevat asiat, joita olisi hyödyllistä analysoida kehittyneen data-analytiikan menetelmillä, mutta se ei datan puuttuessa tällä hetkellä ole mahdollista.

Aktiivisen osallistumisen kautta on tarkoitus muodostaa käsitys tutkittavan asian tämän hetkisestä tilanteesta ja etsiä ratkaisuja nykytilassa vallitseviin haasteisiin. Haasteiden ratkomiseen sovelletaan kehittyneen data-analytiikan menetelmiä. Näitä menetelmiä ja niiden käyttöä on tarkoitus kehittää tämän tutkimuksen puitteissa ja jatkossa toimintatutkimukselle ominaisella tavalla, eli syklisesti arvioiden menetelmien hyödyllisyyttä kussakin käyttötilanteessa.

1.5 Tutkimuksen rakenne

Työ alkaa johdantoluvulla (tämä luku), jossa esitellään tutkimuksen tausta ja aseointi, toimintaympäristö, johon tutkimus sijoittuu, tutkimuksen tavoitteet ja rajaukset, käytettävä tutkimusmenetelmä sekä tutkimuksen rakenne.

Luvussa 2 keskitytään data-analytiikkaan ja päätöksentekomenetelmiin. Ensin luvussa esitellään eri päätöksentekomenetelmiä aina intuitiosta joukkojen viisauteen, sekä tarkastellaan, miten analytiikka sijoittuu päätöksenteon tutkimuskenttään. Sen jälkeen määritellään kehittyneeseen data-analytiikkaan liittyvää käsitteistöä ja esitellään kehittyneen data-analytiikan menetelmiä.

Luvussa 3 esitellään rautatieliikenteen täsmällisyyteen liittyvän analyysitoiminnan nykytila. Tämä tarkoittaa käytettävissä olevan datan, tällä hetkellä tehtävien analyysien sekä analyysijä käyttävien ja tuottavien työryhmien esittelyä. Luvun tavoitteena on rakentaa ymmärrys siitä, mitä haasteita täsmällisyysanalyysin tekemisessä on tällä hetkellä.

Luvussa 4 tehdään erilaisia analyysijä tutkimuksen aikana esiin nousseista liikennejärjestelmän haasteista. Tavoitteena on esittää, miten kehittyneen data-analytiikan menetelmillä kyetään analysoimaan ja visualisoimaan liikenteestä kerättyä dataa niin, että pystytään saamaan parempi ymmärrys järjestelmän toiminnasta. Jokaisen analyysin osalta esitellään analysoitava ongelma, itse analyysi vaihe vaiheelta sekä arvioidaan tehdyn analyysin ja siinä käytettyjen menetelmien hyödyllisyyttä ja käytökelpoisuutta. Luvun lopuksi esitellään lyhyesti kolme tutkimuksen aikana toteutettua täsmällisyysanalyysiin liittyvää alaprojektia.

Luku 5 sisältää tutkimuksen yhteenvedon ja päätelmät; siinä esitetään, miten kehittyneen data-analytiikan menetelmiä voidaan hyödyntää rautatieliikennejärjestelmää analysoitaessa. Lisäksi annetaan suositukset dataympäristön, analytiikan ja tiedon hyödyntämisen kehittämiseen. Lopuksi arvioidaan tutkimuksen onnistumista.

2 Kehittynyt data-analytiikka päätöksenteon tukena

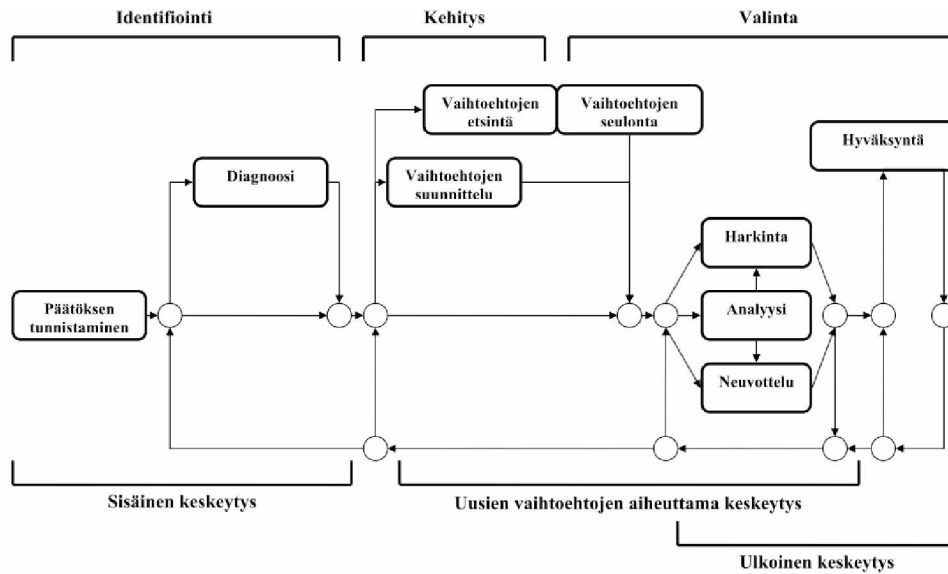
Päätöksiä voidaan tehdä monella eri tavalla. Päätöksenteko voi perustua esimerkiksi aikaisempaan kokemukseen, intuition tai ihmisten mielipiteiden yhdistelyyn. Nämä menetelmät, vaikkakin kaikki varsin toimivia, perustuvat lähinnä mielipiteisiin, näkemyksiin ja tunteisiin. Niinpä niiden faktuaalinen pohja on kyseenalaistettavissa. Analytiikka päätöksentekomenetelmänä perustuu asiayhteydestä kerättyyn dataan, jolloin sen pohjalta tehtyjä päätöksiä voidaan pitää oikeansuuntaisina – kunhan data on luotettavaa ja kuvaa oikeaa asiaa. On olemassa varsin erilaisia analytiikan menetelmiä. Kehittyneen data-analytiikan menetelmien avulla pyritään jalostamaan informaatiota ja ymmärrystä laajoista datajoukoista enemmän tai vähemmän automatisoidusti. Tässä luvussa esitellään, miksi kehittyneen data-analytiikan menetelmiä kannattaa harkita sovellettavaksi rautatieliikennedatan analysointiin.

2.1 Päätöksentekomenetelmät

Päätöksentekoa on tutkittu pitkään ja tutkimuskenttä on varsin kirjava. Jo vuonna 1956 Cyert et al. (1956) tutkivat päätöksentekoa organisaatioissa. Heidän mukaansa päätöksentekotilanteita liiketoiminnassa leimaavat seuraavat ominaispiirteet:

1. Päätöksenteon vaihtoehtoja ei ole annettu, vaan niitä tulee etsiä.
2. Eri päätösten mahdollisia lopputuloksia ei ole annettu, vaan niitä tulee etsiä.
3. Eri vaihtoehtojen vertailua ei voida suorittaa yksinkertaisten päätöksentekokriteerien, kuten liikevoiton perusteella, vaan tulee ottaa huomioon myös muut päätöksenteon vaikutuksen kohteet, kuten aineettomat kriteerit.
4. Ratkaistava ongelma on hyvin harvoin annettu, vaan päätöksentekoa varten täytyy löytää ongelma.

Cyert et al. (1956) olivat edelläkävijöitä esittäessään, että päätöksien tekeminen on myös ongelmien tunnistamista sekä päätöksentekovaihtoehtojen ja niistä seuraavien lopputulosten kehittämistä ja arviointia. Mintzberg et al. kehittivät vuonna 1976 päätöksenteon yleisen mallin, joka on esitetty kuvassa 5. Tässä mallissa he esittävät, että päätökset syntyvät kolmessa vaiheessa. Nämä esitetyt vaiheet ovat *päätöksentekotarpeen identifiointi*, *päätöksentekovaihtoehtojen kehitys* sekä *päätöksentekovaihtoehdon valinta*. He esittävät, että vaihtoehtoja tulisi syvällisesti analysoida ja harkita ennen lopullista päätöksentekoa. Heidän tutkimuksensa jälkeen kiinnostus alkoi suuntautumaan siihen, miten päätöksentekotilanteissa eri vaihtoehtoja analysoidaan ja harkitaan. (Mintzberg et al 1976).

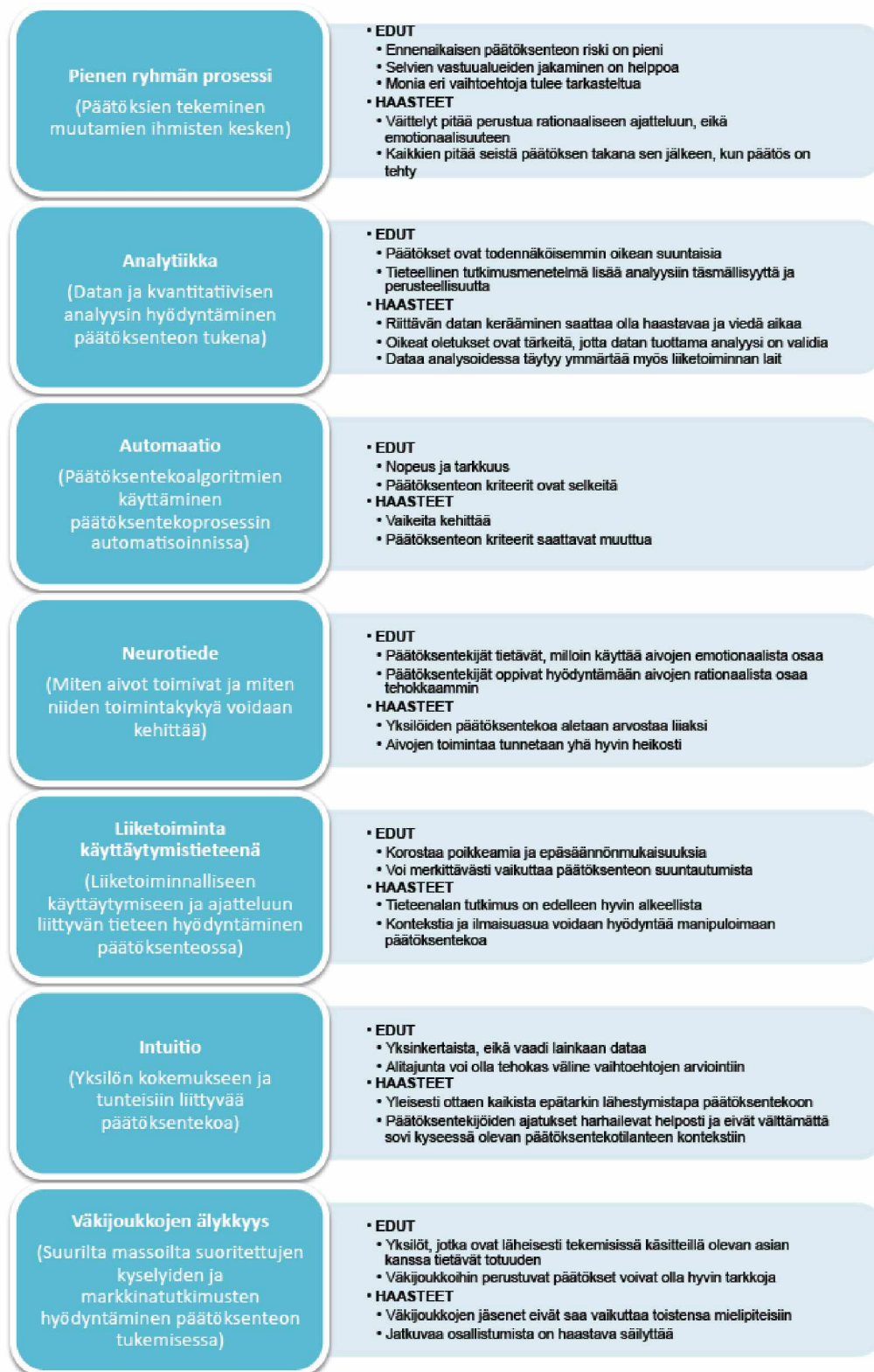


Kuva 5. Strategisen päätöksentekoprosessin yleinen malli (mukailtu lähteestä Mintzberg et al. 1976).

Uusimpia tutkimuksia päätöksenteosta edustaa Davenportin (2009) tekemä tutkimus. Hänen mukaansa päätöksiä voi tehdä seitsemällä eri tavalla. Nämä kuvassa 6 esitetyt tavat tehdä päätöksiä ovat:

- Tekemällä päätöksiä muutamien ihmisten kesken
- Hyödyntämällä dataa ja kvantitatiivista analyysiä päätöksenteon tukena
- Käyttämällä päätöksentekoa algoritmeja päätöksentekoprosessin automatisoinnissa
- Hyväksikäyttämällä liiketoiminnallista käyttäytymistietoa päätöksenteossa
- Hyödyntämällä hallittua aivojen käyttöä päätöksenteossa
- Päättämällä puhtaasti yksilön kokemusten ja tunteiden pohjalta intuitiivisesti
- Kartoittamalla suurten massojen haluja päätöksenteon tukena.

Davenportin mukaan näitä päätöksenteon menetelmiä voi sekä hyödyntää yksinään tai yhdistellä keskenään päätöksentekotilanteissa. Kuitenkin jokainen päätöksentekotilanne sisältää vähintään jonkin menetelmän hyödyntämistä.



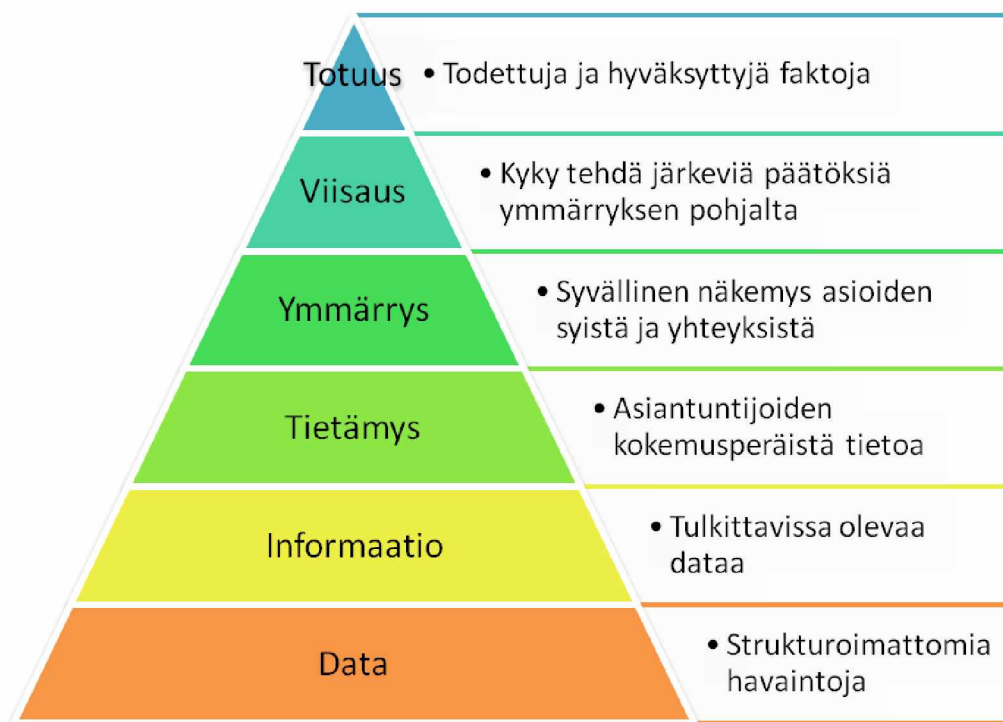
Kuva 6. Päätöksenteon eri menetelmät (mukailtu lähteestä Davenport 2009).

Davenport (2009) on nostanut analytiikan yhdeksi seitsemästä tavasta tehdä päätöksiä. Sen eduksi on esitetty se, että päätökset ovat tällöin todennäköisemmin oikeasuuntaisia, täsmällisempiä ja perusteellisempia. Heikkouksina on taasen nähty datan keruun haastavuus ja aikaavievyyys sekä liiketoiminnan ymmärtämisen tarve luodessa hypoteeseja sekä lopputuloksia. (Davenport 2009). Täten voidaankin todeta, että

vaikka hyvin usein odotetaan, että kehittyneen analytiikan avulla kyetään analysoimaan tietoa automatisoidusti, ei asia ole puhtaasti näin. Hyvin suuria informaatio- ja datamassoja analysoitaessa voidaan hyväksikäyttää erinäisiä automatisoituja analytiikan menetelmiä, mutta taustalla täytyy silti olla ymmärrys sekä analysoitavasta asiakontekstistä että käytettävistä analytiikkatyökaluista. Haasteista huolimatta analytiikkaa on syytä harrastaa, mikäli päätöksenteon tueksi halutaan luotettavaa tietoa.

2.2 Tiedon tasot

Data on yksi tiedon ilmenemismuoto. Thierauf (2001) on määritellyt tiedon koostuvan kuudesta eri tasosta. Nämä tasot, kuten kuva 7 esittää, ovat raakamuotoisimmasta jalostuneimpaan listattuna *data*, *informaatio*, *tietämys*, *ymmärrys*, *viisaus* ja *totuus*.



Kuva 7. Tiedon tasot (mukailtu lähteestä Thierauf 2001).

Data on tiedon muodoista kaikista raakamuotoisinta. Thieraufin (2001) mukaan data on joukko strukturoimattomia faktuaalisia havaintoja. Datasta luodaan informaatiota muokkaamalla ja ryhmittelemällä sitä niin, että datasta ymmärretään mitä se sisältää ja edustaa. **Informaatio** onkin tulkittavissa olevaa dataa, jota kyetään hyödyntämään datan sisällön tulkitsemisessa. Tiedon seuraava taso, **tietämys** on asiantuntijoiden kokemusperäistä tietoa jostakin tietyistä asiasta. Tietämyksen muodostuminen ei tarvitse tuekseen informaatiota, sillä se on kokemusperäisesti syntyntä. **Ymmärrys** puolestaan on syvälistä näkemystä asioiden syistä ja yhteyksistä. Toisin sanoen ymmärrys edellyttää sekä kykyä hahmottaa ja tulkita kerätystä datasta muodostetun informaation sisältöä ja yhdistää tätä kokemusperäisen tiedon kanssa. (Thierauf 2001). Tämän työn kannalta yksi mielenkiintoinen tarkastelunäkökulma onkin, kyetäänkö kehittyneen data-analytiikan menetelmien avulla saavuttamaan ymmärrystä rautatie-liikennejärjestelmän toiminnasta, ilman että varsinaisesti on syvälistä tietämystä siitä, miten järjestelmä tällä hetkellä toimii.

Ymmärrystä korkeammat tiedon tasot ovat viisaus ja totuus. Näistä *viisaus* on määritelty kykynä tehdä järkeviä päätöksiä ymmärryksen pohjalta. Tämä voidaan ymmärtää niin, että vaikkakin päätöksentekijällä on hyvä ymmärrys, ei se vielä tarkoita, että hän osaisi tehdä parhaita mahdollisia päätöksiä tulkiten omaa ymmärrystään oikein tilanteen kannalta. Viisaus voidaankin nähdä enemmän henkilökohtaisena ominaisuutena. Tiedon ylin taso on *totuus*. Totuus on yleisesti todettuja ja hyväksytyjä faktoja. (Thierauf 2001). Rautatieliikenteessä hyvin monia asioita pidetään ”totuuksina”. Saattaa kuitenkin olla, että osa näistä totuuksista perustuu vanhentuneeseen dataan, tai alun perinkin vääriin oletuksiin. Tämän vuoksi uusien totuuksien muodostamiseksi täytyy tietoa jalostaa alimmilta tasoilta asti.

2.3 Kehittyneen data-analytiikan menetelmät

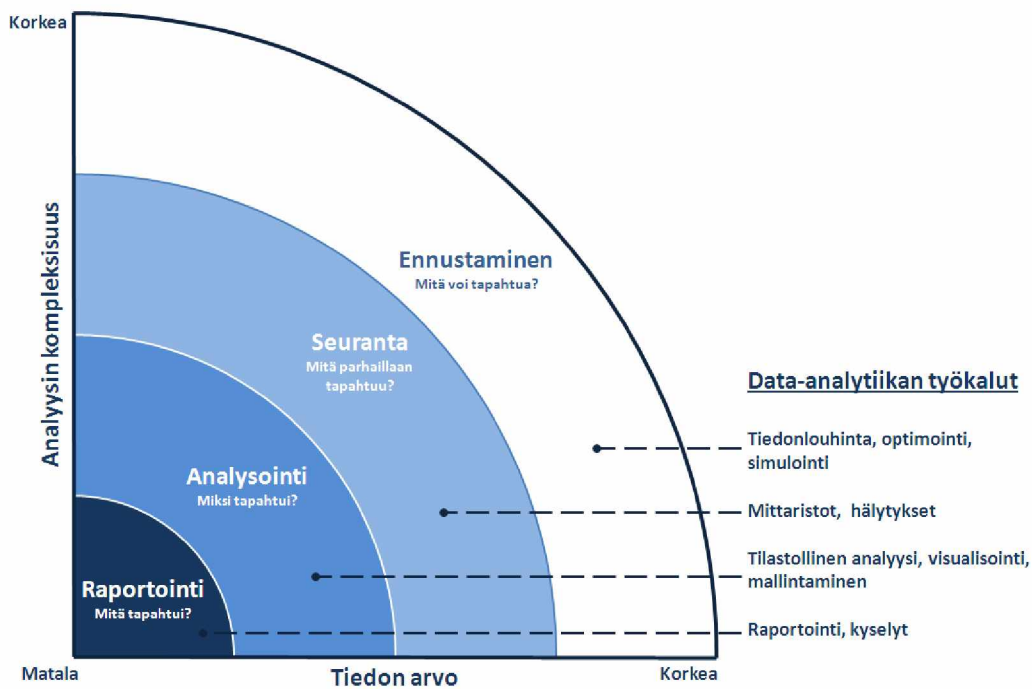
Kehittynyt analytiikka on Bosen (2009) mukaan yleistason käsite, jolla tarkoitetaan kehittyneiden analyttisten menetelmien hyödyntämistä tiedon analysointiin kysymyksiin vastaamiseksi ja ongelmien ratkomiseksi. *Kehittynyt data-analytiikka* puolestaan määriteltiin jo aiemmin olevan joukko menetelmiä ja työkaluja, joilla pyritään automatisoidusti löytämään oleellinen tieto suurista tietomassoista (Kudyba & Hoptroff 2001, Giudici 2003).

Hyvin usein kirjallisuudessa törmää erikseen kehittyneen analytiikan ja kehittyneen data-analytiikan termien käyttöön hyvin samankaltaisissa konteksteissa. Myös tässä tutkimuksessa termeillä tarkoitetaan lähes samaa asiaa kuitenkin sillä erotuksella, että kehittyneessä data-analytiikassa lähtötietona on puhtaasti dataa, kun kehittyneessä analytiikassa lähtötietoa voi olla myös tiedon ylempiltä tasoilta.

Analytiikkaa käytetään Davenport & Harrisin (2007) mukaan siksi, että ymmärrettäisiin paremmin datan sisältämiä piirteitä. Davenport & Harrisin (2007) määrittelevät kehittyneen analytiikan datan, tilastollisen sekä kvantitatiivisen analyysin ja selittävien sekä ennustavien mallien kattavaksi hyödyntämiseksi päätöksenteon ja toiminnan ajamisen tukena. Kehittynyt analytiikka ei Davenport & Harrisin (2007) mukaan vaadi alkuhypoeseja. Bosen (2009) mukaan kehittynyt analytiikka ei itsessään ole teknologia, vaan ryhmä työkaluja, joita hyödynnetään informaation keräämiseen, muokkaamiseen, analysointiin sekä tulosten esittämiseen ja ennustamiseen. Bosen (2009) mukaan kehittynyt analytiikka koostuu seuraavista osa-alueista:

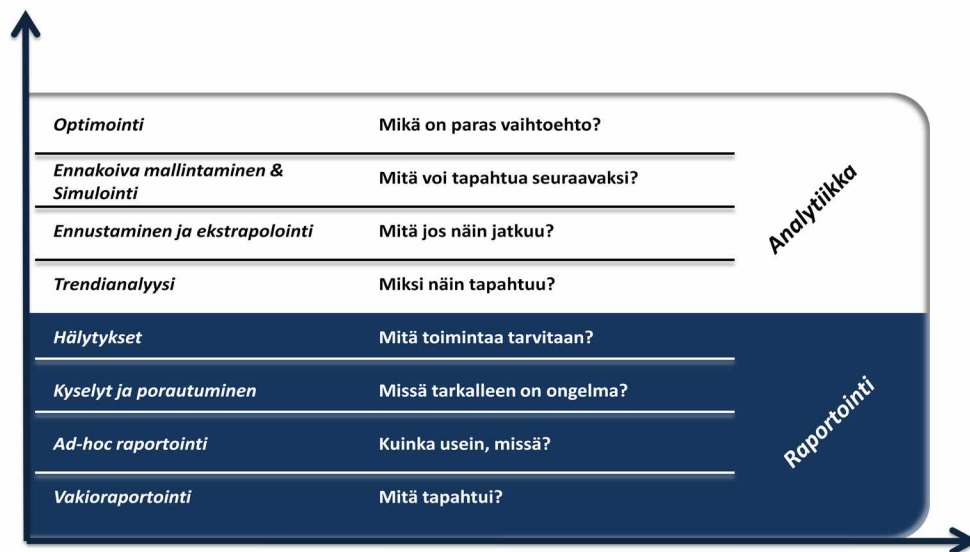
- Dataintegraatio
- Tiedonlouhinta
- Tekstinlouhinta
- Weblouhinta
- Visualisointijärjestelmät
- Ennustava analytiikka
- Suosituksenantojärjestelmät

Eckerson (2007) lajittelee data-analytiikan työkalut raportointi-, analysointi-, seuranta- ja ennustamistyökaluihin. Kuvassa 8 on esitetty Eckersonin (2007) esittämä luokittelu data-analytiikan työkaluista analyysin kompleksisuuden ja tuotetun tiedon arvon suhteen. Näistä ennustamistyökaluihin luokitellut tiedonlouhinta-, optimointi-, simulointi- ja analysointityökalut kuten tilastollinen analyysi, visualisointi ja mallintaminen vastaavat Bosen (2009) sekä Davenport & Harrisin (2007) kehittyneiden työkalujen määritelmiä.



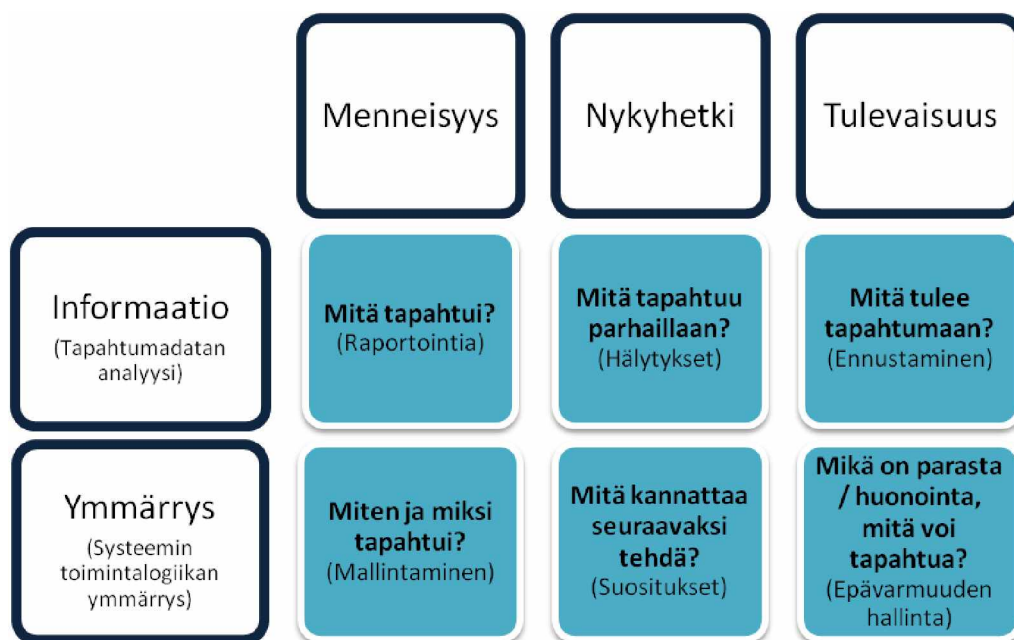
Kuva 8. Data-analytiikan työkalut (mukailtu lähteestä Eckerson 2007).

Davenport & Harris (2007) luokittelevat datan analysoinnin kahdeksaan eri tasoon, jotka on esitetty kuvassa 9. Näistä neljän alimman, vakioraportoinnin, ad-hoc-raportoinnin, kyselyjen ja porautumisen sekä hälytyksien he näkevät olevan raportointityökaluja (Davenport & Harris 2007). Analytiikkaa ovat puolestaan trendianalyysit, ennustaminen ja ekstrapolointi, ennakoiva mallintaminen sekä simulointi ja optimointi. Kysymykset, joihin analytiikan avulla voidaan vastata, ovat Davenport & Harrisin mukaan seuraavat: Miksi näin tapahtuu? Mitä jos näin jatkuu? Mitä voi tapahtua seuraavaksi? Mikä on paras vaihtoehto?



Kuva 9. Datan analysoinnin tasot (mukailtu lähteestä Davenport & Harris 2007).

Davenport et al. (2010) ovat pyrkineet kuvaamaan, mihin kysymyksiin analytiikan menetelmillä voidaan saavuttaa vastaus aikadimension ja analyysin tuottaman tiedon tason suhteen. Davenport et al. (2010) ovat sitä mieltä, että kehittyneen data-analytiikan menetelmien avulla voidaan tuottaa sekä informaatiota, että ymmärrystä tuottavaa tietoa (kuva 10). Kysymykset, joihin kyetään ymmärrystä tuottavien analytiikan työkalujen avulla vastaamaan, ovat: Miten ja miksi tapahtui? Mitä kannattaa tehdä seuraavaksi? Mikä on parasta/huonoita mitä voi tapahtua? Kuten kysymyksistä voidaan todeta, ovat ne aidosti ymmärrystä tuottavia. Tässä tutkimuksessa datan luonteen vuoksi keskitytään lähinnä mallintamaan liikennejärjestelmää, eli vastaamaan kysymykseen: miten ja miksi jokin ilmiö tapahtui?



Kuva 10. Kysymykset, joihin analytiikan eri menetelmillä pyritään vastaamaan lajiteltuna aikadimension ja analyysin tuottaman tiedon suhteen (muokattu lähteestä Davenport et al. 2010).

Edellä esitettyjen määritelmien pohjalta on muodostettu tässä tutkimuksessa käytävä laaja määritelmä **kehittyneelle data-analytiikalle**:

Kehittynyt data-analytiikka koostuu menetelmistä ja työkaluista, joilla pyritään automatisoidusti löytämään oleellinen tieto suurista tietomassoista. Kun perinteisten data-analyysimenetelmien ja -työkalujen avulla on lähinnä mahdollista analysoida pieniä datamääriä muutaman muuttujan suhteen kerrallaan, on kehittyneen data-analytiikan avulla mahdollista tarkastella laajojakin datamassoja useiden satojen muuttujien suhteen sen kehittyneiden analyysialgoritmien ja työkalujen avulla. Näin ollen kehittyneempiä menetelmiä hyödynnettäessä ei tarvita valmiita hypoteeseja, vaan ideana on, että työkalut löytävät datamassasta johdonmukaisuuksia, kuten esimerkiksi poikkeamia, säännönmukaisuuksia sekä syy-seuraussuhteita. Kehittyneen data-analytiikan menetelmien avulla pystytään tuottamaan sekä informaatiota että ymmärrystä datan pohjalta. Kehittyneen data-analytiikan menetelminä voidaan nähdä muun muassa data-integraatio, tiedonlouhinta, visuaalinen analytiikka sekä erinäiset ennustamiseen, ennakointiin ja suosituksenantoon pyrkivät menetelmät.

Tässä tutkimuksessa kehitettyneen data-analytiikan menetelmistä hyödynnetään tiedonlouhinnan osalta assosiaatioanalyysiä, sekvenssianalyysiä, klusterointia ja luokittelua. Muista menetelmistä hyödynnetään visuaalista analytiikkaa. Nämä menetelmät esitetään seuraavissa alaluvuissa. Ohjelmistona tutkimuksessa hyödynnetään SAS:in tuotteita. On kuitenkin huomattava, etteivät kehitettyneen data-analytiikan menetelmät ole millään tapaa sidottuja yksittäiseen ohjelmistoon.

2.3.1 Assosiaatioanalyysi & sekvenssianalyysi

Assosiaatioanalyysin avulla pyritään löytämään tilastollisesti usein toistuvia hahmoja sekä tapahtumien välisiä yhteyksiä ja riippuvuuksia (Han & Kamber 2000). Yleistetyksi tavoitteena on löytää tapahtumia, jotka ilmenevät yleisesti yhdessä. Sekvenssianalyysin ainoa ero assosiaatioanalyysiin on, että sekvenssianalyysillä tutkitaan näiden yhteyksien ja riippuvuuksien tapahtumajärjestystä, mikäli datassa on otettu ajan tai tapahtumien kulkujärjestys huomioon (Han & Kamber 2000). Klassinen esimerkki assosiaatioanalyysistä löytyy kaupan alalta, jossa on jo pidemmän aikaa tutkittu, mitä tuotteita asiakkaat ostavat samaan aikaan. Tätä sovellutusta kutsutaan niin sanotuksi markkinakoriantalyysiksi. (Giudici 2003).

Assosiaatioanalyysiä varten tieto tulee olla tallennettuna tietokantaan, jossa riveinä on transaktioita ja sarakkeina havaintoja, jotka liittyvät kyseiseen transaktioon. Taulukkoon 1 on luotu esimerkkikäyttöä varten datajoukko, johon on tallennettu tapahtumatransaktioita ja se, mitä kyseisten transaktioiden myötä on tapahtunut. Taulukossa ensimmäinen sarake kertoo, mikä transaktio on kyseessä ja loput sarakkeet, mitkä tapahtumista A, B, C ja D on tapahtunut kyseisen transaktion aikana. Tapahtumasarakkeiden koodaus on toteutettu binäärisesti tarkoittaen, että mikäli sarakkeessa on 1, tapahtuma on toteutunut ja 0 mikäli ei.

Taulukko 1. Tapahtumatransaktiodatajoukko assosiaatioanalyysin havainnointia varten.

TRANSAKTIO-ID	A	B	C	D
1	1	0	1	0
2	1	1	1	0
3	1	0	0	0
4	1	1	1	1
5	0	0	1	1
6	0	1	0	0

Assosiaatiosäännöt ilmaistaan muodossa $X \rightarrow Y$, jossa X on edeltäjä (engl. antecedent) ja Y seuraus (engl. consequent) (Giudici 2003, Han & Kamber 2000). Assosiaatio- ja sekvenssianalyysijä voidaan luoda myös useammille kuin kahdelle muuttujalle, vaikkakin esimerkit on toteutettu kahden muuttujan suhteen. Edeltäjän ja seurauksen riippuvuuden voimakkuutta voidaan tarkastella kolmen eri tunnusluvun avulla. Nämä tuki (engl. support), luottamus (engl. confidence) sekä noste (engl. lift).

Tuen avulla saadaan selvitettyä, kuinka suuressa osassa koko dataa ilmenevät haluttu muuttuja tai muuttujien yhdiste. Giudicin sekä Han & Kamberin mukaan tuki määritellään havaintojen jotka sisältävät X ja Y lukumäärän suhteen kaikkien havaintojen lukumäärään, eli toisin sanoen todennäköisyytenä:

$$Supp(X \rightarrow Y) = P(X \cap Y)$$

Ja vastaavasti yhdelle havainnolle X:

$$Supp(X) = P(X)$$

(Giudici 2003, Han & Kamber 2000)

Esimerkkidatan avulla voidaan havainnoida tätä tilannetta. Jos halutaan tutkia missä ilmenee sekä A että B, saadaan tueksi $2/6 = 0,33 = 33\%$. Jos haluttaisiin tutkia vain C:n ilmenemistä, saataisiin tueksi $4/6 = 0,66 = 66\%$.

Luottamusta käytetään selvittäessä, kuinka usein niissä havainnoissa, joissa ilmenee X, ilmenee myös Y. Luottamus määritellään Giudicin sekä Han & Kamberin mukaan X:n ja Y:n sisältävien havaintojen lukumäärän suhteena X:n sisältävien havaintojen lukumäärään. Tämä voidaan tulkita myös Y:n ehdollisena todennäköisyytenä ehdon ollessa X.

$$Conf(X \rightarrow Y) = \frac{Supp(X \rightarrow Y)}{Supp(X)} = \frac{P(X \cap Y)}{P(X)} = P(X|Y)$$

(Giudici 2003, Han & Kamber 2000)

Esimerkkidatan avulla voidaan tutkia luottamusastetta: ”mikäli havainto on sisältänyt A:n, on se myös sisältänyt C:n”-tarkastelussa tulee laskentakaavaksi $3/4 = 0,75 = 75\%$. Vastaava luku $(A \rightarrow D)$ -luottamusasteelle on $1/4 = 0,25 = 25\%$.

Nostetta käytetään määriteltäessä muuttujien keskinäistä riippuvuutta. Giudicin sekä Han & Kamberin mukaan noste määritellään havaitun tuen suhteena odotettuun tukeen, jos X ja Y olisivat riippumattomia. Koska noste on normeerattu muuttuja, ilmaisee se yhdestä poikkeavilla arvoilla assosiaatiota. Yhtä suuremmat arvot ilmaisevat positiivista assosiaatiota ja yhtä pienemmät arvot negatiivista assosiaatiota.

$$Lift(X \rightarrow Y) = \frac{Supp(X \rightarrow Y)}{Supp(X)Supp(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)} = \frac{P(X|Y)}{P(Y)}$$

(Giudici 2003, Han & Kamber 2000)

Tutkiessa esimerkkidatan avulla $(A \rightarrow B)$ -nostetta, saadaan luvuksi $(2/4)/(3/6) = 1$. $(A \rightarrow C)$ -noste on $(3/4)/(4/6) = 1,125$ ja $(A \rightarrow D)$ -noste $(1/4)/(2/6) = 0,75$. Näistä voidaan todeta, että A:lla ja C:llä on positiivinen assosiaatio, kun puolestaan A:lla ja D:llä on negatiivinen assosiaatio.

Assosiaatio- ja sekvenssianalyysejä voidaan hyödyntää, kun halutaan poimia datajoukosta havaintoja, joiden alkiot liittyvät vahvasti toisiinsa. Näitä analyysejä voidaan laatia kahden tai useamman muuttujan suhteen. Analyysejä voidaan myös tehdä tapahtumien ilmenemisjärjestyksen suhteen. Assosiaatio- ja sekvenssianalyyseiden lopputuloksina saadaan molemmissa tuen ja luottamusasteen tilastotiedot havainnoille, mutta nosteen tilastotieto syntyy vain assosiaatioanalyysin yhteydessä.

2.3.2 Klusterointi

Klusterointi on menetelmä, jonka avulla voidaan tunnistaa ja lajitella dataa samankaltaisiin joukkoihin. Kudyba & Hopftroff (2001) mukaan tilastollisten algoritmien avulla voidaan jakaa datajoukkoja ryhmiin niin, että ryhmät muodostuvat havainnoista, joiden muuttujat ovat keskenään samankaltaisia. Klusteroinnin ja luokittelun erottaa toisistaan se, että kun luokittelussa määritellään kohdemuuttuja, jonka suhteen halutaan dataa jakaa ryhmiin, ei klusteroinnissa tällaista valita, vaan algoritmit lajittelevat havainnot klustereihin automatisoidusti (Kudyba & Hopftroff 2001).

Klusterointi vaatii Giudicin mukaan datan käsittelyä etukäteen. Muuttujat, joilla ei ole analyysin kannalta merkitystä, olisi syytä poistaa datasta, sillä nämä saattavat muodostaa omia klustereita, sekä vääristää muita klustereita. Samoin muuttujat tulisi normittaa siten, että eri muuttujien vaikutus klusterointiin olisi yhtäläinen. Muuttujia normittaessa myös klusterit muuttuvat, joten tämä vaatii erityistä tarkkuutta käyttäjältä. (Giudici 2003).

Klusteroinnin toteuttamiseen on olemassa useita eri menetelmiä (Giudici 2003, Han & Kamber 2000). Hierarkisen klusteroinnin tarkoitukseen on tarjolla muun muassa seuraavat kolme eri menetelmää (SAS 2006):

- "Average linkage" eli keskimääräinen linkitys
- "Centroid method" eli keskusmenetelmä
- "Ward's minimum-variance method" eli Wardin menetelmä

Kaikkien näiden menetelmien aloitustilanne on sellainen, jossa jokainen havainto on oma klusterinsa. Tämän jälkeen menetelmät yhdistelevät lähimpiä klustereita toisiinsa siten, kunnes haluttu määrä klustereita saavutetaan. Menetelmien väliset erot ovat siinä, miten ja missä järjestyksessä ne tunnistavat toisiaan lähellä olevat klusterit. (SAS 2006). Havaintojen välisenä mittana käytetään Euklidista normia (Giudici 2003).

Giudicin mukaan *keskimääräisessä linkityksessä* kahden klusterin välinen etäisyys määritellään aritmeettisena keskiarvona havaintoparien N_k ja N_l välillä, joista toinen on yhdessä klusterissa ja toinen toisessa:

$$D_{KL} = \frac{1}{N_K N_L} \sum_{i \in C_K} \sum_{j \in C_L} \|x_i - x_j\|^2.$$

(Giudici 2003)

Keskusmenetelmä on Giudicin mukaan on hyvä sietämään ylimääräisiä, ei-selittäviä, muuttujia paremmin kuin muut menetelmät, koska se ensin laskee klusterin keskuksen, jota se vertailee muihin klustereihin. Keskusmenetelmässä kahden klusterin välinen etäisyys on klusterien keskusten välinen etäisyys. Tällöin

$$D_{KL} = \|\bar{x}_K - \bar{x}_L\|^2$$

(Giudici 2003)

Wardin menetelmä on SAS Enterprise Miner 4.3:ssa oletuksena käytettävä klusterointimenetelmä (SAS 2006). Wardin menetelmän ongelmana on taipumus tuottaa klustereita, joissa on suunnilleen saman verran havaintoja (Giudici 2003). Wardin menetelmän algoritmi pyrkii maksimoimaan klusterien sisäistä yhteneväisyyttä ja maksimoimaan klustereiden välisiä etäisyyksiä seuraavan algoritmin avulla:

$$D_{KL} = \frac{\left\| \bar{x}_K - \bar{x}_L \right\|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$

(Giudici 2003)

Optimaalisessa klusteroinnissa on tarkoitus minimoida klustereiden sisäinen heterogeisuus, tai vaihtoehtoisesti maksimoida klustereiden eristyneisyys muusta datasta. Optimaaliseen klusterointiin on työkaluissa usein tarjolla useita eri menetelmiä (SAS 2006). Menetelmien erona on käytettävä klusterointikriteeri, sekä iteraatioiden lukumäärä (SAS 2006). Kaikki nämä menetelmät perustuvat klustereiden sisäisen heterogeisuuden minimointiin, ja käytettävä klusterointikriteeri on kaikissa p-normi (SAS 2006).

Klusterointi toteutetaan Giudicin (2003) mukaan k-means-algoritmillä riippumatta edellä esitetyn menetelmän valinnasta. K-means etenee seuraavasti:

1. Jaetaan havainnot alustaviin klustereihin
2. Lasketaan kunkin klusterin keskus
3. Liitetään kukin alkio siihen klusteriin, minkä keskusta se on lähinnä
4. Toistetaan 2-3 kunnes muutoksia klustereiden keskuksissa ei enää tapahdu, tai vaihtoehtoisesti iteraatioita on suoritettu asetettu maksimimäärä.

Klusterianalyysin ensimmäinen vaihe on Giudicin mukaan määrittää klustereiden lukumäärä. Käyttäjä voi halutessaan määritellä haluamansa klustereiden määrän tai vaihtoehtoisesti optimaalinen määrä ratkaistaan automaattisesti (Giudici 2003). Määritettyjen arvojen perusteella määritellään tietyille (esimerkiksi 2000 havaintoa) satunnaisesti valitulle havainnolle alustavat klusterit, joita on asetettu klustereiden maksimimäärä. Tämän jälkeen näille alustaville klustereille suoritetaan hierarkinen klusterointi, ja klustereiden määräksi valitaan määrä, joka on suurempi kuin asetettu minimimäärä, ja jossa selitysasteen raja ylittyy. (SAS 2006). Tämän jälkeen suoritetaan valitun hierarkisen klusteroinnin menetelmän proseduuri, joka tekee optimaalisen klusteroinnin datalle.

Klusterointi on varsin kompleksinen ja monivaiheinen prosessi. Klusterointi on kuitenkin hyvä työkalu datan tutkimiseen, kun ei vielä ymmärretä, miten datan sisältämät havainnot käyttäytyvät. ***Klusteroinnin avulla voidaankin löytää ne havainnot datasta, jotka muistuttavat läheisesti toisiaan, tai toisaalta poikkeavat muista.*** Näin on mahdollista löytää potentiaalisia ongelmakohtien ryppäitä esimerkiksi rautatieliikennejärjestelmästä.

2.3.3 Luokittelu

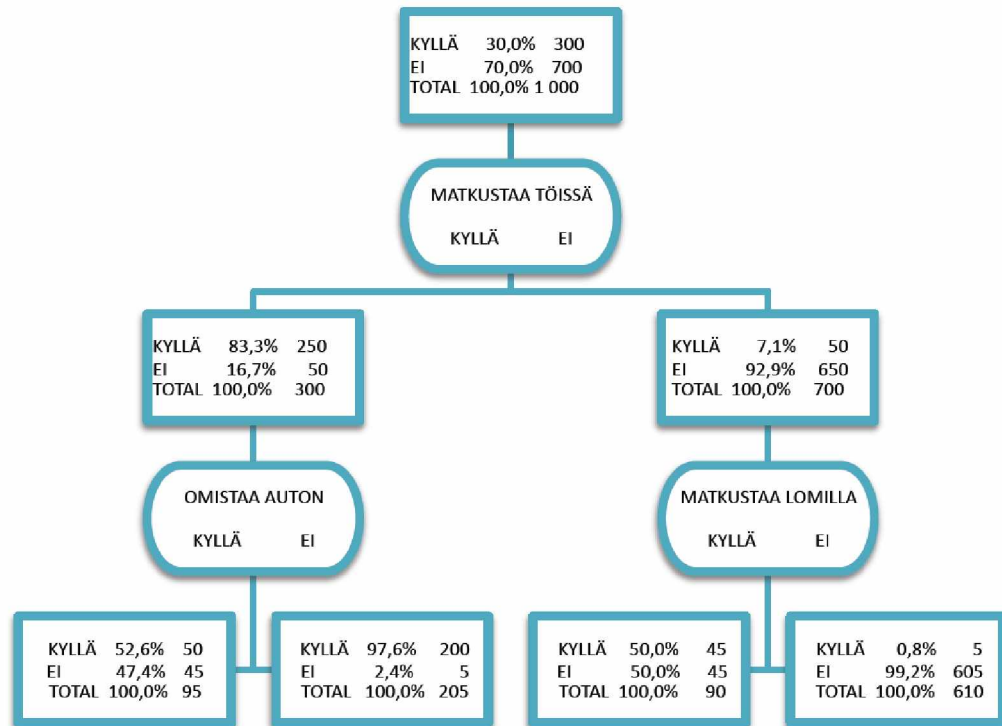
Luokittelun avulla pyritään kategorisoimaan ja segmentoimaan dataa valitun muuttujan suhteen. Luokittelussa valitaan Kudyba & Hoptroffin (2001) mukaan kohdemuuttuja, jonka suhteen valittu analyysialgoritmi jakaa datan pienempiin osajoukkoihin käyttäen datan muita muuttujia selittävinä tekijöinä jaon suhteen. Luokittelun lopputuloksena syntyy yleensä niin sanottu puumalli, jonka avulla on helppo havainnoida eri luokkien suuruuksia ja luokkiin jaossa käytettyjä selittäviä muuttujia.

Luokittelussa käytettäviä jakokriteereitä ovat muun muassa CHAID (Chi-squared automatic interaction detection), CART (Classification and Regression Trees) sekä C4.5 ja C5.0 (Han & Kamber 2000). Näistä CART on käytöltään kaikista suosituin (Giudici 2003). Menetelmien erot liittyvät algoritmeihin, joiden perusteella dataa jaetaan alemmilla tasoilla parhaiten kuvaaviin ryhmiin (Giudici 2003). Esimerkiksi CHAID:issa luokkiin jako perustuu chi-neliötestin tuottamaan itsenäisyysasteeseen, kun CART puolestaan pyrkii löytämään selittäviä monimuuttujia lineaaristen ominaisuuksien joukosta (Han & Kamber 2000). CART:in käyttämä algoritmi on nimeltään *Gini impurity*. Giudicin mukaan Ginin epäpuhtausalgoritmi valitsee jokaisessa jakovaiheessa kaikista vaihtoehtoisista jakovaihtoehtoista sen, jonka tuloksena syntyvä jako on kaikista puhtain, eli toisin sanoen ryhmät ovat mahdollisimman homogeenisiä. Ginin epäpuhtauden algoritmi on seuraava:

$$gini(D) = 1 - \sum_{j=1}^n p_j^2$$

(Giudici 2003)

Kuvassa 11 on esitetty esimerkki luokittelun lopputuloksena syntyvästä puumallista. Esimerkki on mukailtu Giudicin (2003) esittämästä esimerkistä. Mukailtu versio tutkii, ketkä ihmiset ostavat junalippuja ja ketkä eivät. Puumallin ensimmäisenä jakokriteerinä toimii se, matkustavatko ihmiset töiden vuoksi. Niistä ihmisistä, jotka matkustavat töiden vuoksi 83,3 % ostaa junalippuja, kun taas niistä ihmisistä, jotka eivät matkusta töiden vuoksi 92,9 %:n enemmistö ei osta junalippuja. Tämän jälkeen toisen tason jakokriteerit poikkeavat toisistaan. Niistä ihmisistä, jotka matkustavat töiden takia ja jotka eivät omista autoa 97,6 % ostaa junalippuja, kun autollisistakin 52,6 % ostaa lippuja. Puolestaan ne henkilöt, jotka eivät töiden puolesta matkusta, mutta matkustavat lomilla 50,0 % ostaa junalippuja. Ryhmässä, jossa ei matkusteta töissä eikä lomilla, vain 0,8 % ostaa junalippuja. Kuten voidaan todeta, luokittelulla voidaan tämän esimerkin tavoin segmentoida junalippujen ostokäyttäytymistä muiden muuttujien suhteen.



Kuva 11. Esimerkki luokittelun puumallista. Tutkittavana kohdemuuttujana, osataako henkilö junalippuja. (mukailtu lähteestä Giudici 2003).

Luokittelu on varsin tehokas työkalu tietyn kohdemuuttujan eri arvojen segmentointiin muiden selittävien muuttujien suhteen. Luokittelussa voidaan vaihdella valittua kohdemuuttujaa, jolloin voidaan tutkia, mitkä muut muuttujat ovat parhaita selittäviä tekijöitä kulloinkin valitun kohdemuuttujan arvojen suhteen. Luokittelu onkin hyvä valinta datan tutkimiseen, kun *halutaan ymmärtää, miten jokin tietty muuttuja vaikuttaa muihin ja täten kyetään mahdollisesti jopa ennakoimaan tulevaisuuden tapahtumia olettaen, että tulevaisuudessa muuttujien keskinäiset vaikutukset ovat samat.*

2.3.4 Visuaalinen analytiikka

Ihmisen yksi parhaista aisteista asioiden tulkitsemiseen on näköaisti. Paras tapa tehdä datasta helposti analysoitavaa on luoda siitä kontekstiin sopivia visualisointeja, jotka ovat helposti tulkittavissa. Kudyba & Hoptroff (2001) toteavat, että visuaalinen analytiikka on paljon enemmän kuin vain kauniiden kuvaajien piirtämistä. Visuaalisessa analytiikassa tavoitteena on viestiä laaja määrä dataa ihmisavoille lyhyessä ajassa, jotta kyetään hyödyntämään parasta mahdollista hahmontunnistusmoottoria: visuaalista aivolohkoa (Kudyba & Hoptroff 2001).

Visuaalinen analytiikka on Kudyba & Hoptroffin (2001) mukaan silloin parhaimmillaan, kun tuotetaan usein samaan pohjaan perustuvia visualisointeja samasta aiheesta, jolloin ihmisaivojen on helppo vertailla eri kuvia ja kuvien muutosta. Keim et al (2008) totavatkin, että visuaalinen analytiikka on tieteenlaji, joka ei ole puhdasta analytiikkaa, vaan datan mallintamisen ja ihmisen kognitiivisen havainnoinnin väliin sijoittuva tieteenala. Keim et al. (2008) visuaalinen analytiikka on joukko menetelmiä ja työkaluja, joiden avulla ihmiset kykenevät:

- Yhdistelemään tietoa ja rakentamaan ymmärrystä laajoista, muuttuvista ja tulkinnanvaraisista datajoukoista
- Havainnoimaan odotetun ja paljastamaan odottamattoman
- Välittämään ymmärrettäviä arviointeja
- Viestimään arvioita toimintaa varten.

Visualisoinnit tulisivat Kudyba & Hoptroffin (2011) mukaan rakentaa niin, että niistä on helppo havaita poikkeamia. Tarvittaessa visualisointeja voi toteuttaa esimerkiksi animaatioina, jolloin silmä pystyy havaitsemaan muutokset helposti (Kudyba & Hoptroff 2001).

Visualisointien hyödyntämisen prosessi tapahtuu Kudyba & Hoptroffin (2001) mukaan yleensä seuraavasti:

1. Tarkastelija havaitsee selvät poikkeamat visualisoinnista.
2. Tarkastelija käyttää muutaman minuutin ymmärtääkseen mitä muuttujia visualisointi sisältää.
3. Tarkastelija käyttää muutaman lisäminuutin ymmärtääkseen mitä visualisointi kertoo.

Visualisoinnit ovat yksi tehokkaimmista tavoista ymmärtää ja viestiä datan sisältö. Nykyaikaiset tietotekniset työkalut mahdollistavat mitä monimuotoisinten ja tilanteeseen sekä toimintaympäristöön sopivien visualisointien laatimisen, jolloin niitä tulisi hyödyntää parhaan mahdollisen ymmärryksen rakentamiseksi. *Visualisointien tekeminen ei kuitenkaan ole vain kuvien ja kuvaajien laatimista, vaan niissä tulisi miettiä, miten visualisointi on helpoiten tulkittavissa kuitenkin niin, että se tarjoaa kaiken tarpeellisen informaation tarkastelijallensa.*

2.4 Yhteenveto

Kehittyneen data-analytiikan menetelmät mahdollistavat suurten datamassojen muokkaamisen ja analysoinnin niin, että lopputuloksena on ihmisille helposti ymmärrettävissä olevia analyysejä. Näiden analyysien avulla on mahdollista kehittää parempi tietotila analysoitavan kohteen toiminnasta ja täten tehdä päätöksenteosta informoidumpaa kuin mitä se muutoin olisi. Voidaankin todeta, että kehittyneen data-analytiikan menetelmien avulla ei pelkästään luoda datasta informaatiota, vaan jotain paljon enemmän, nimittäin ymmärrystä. Tässä tutkimuksessa hyödynnetään luvuissa 2.3.1–2.3.4 esitettyjä kehittyneen data-analytiikan menetelmiä. Olemassa olevia kehittyneen data-analytiikan menetelmiä on kuitenkin paljon enemmän. Tulevaisuudessa onkin mielenkiintoista pohtia myös muiden menetelmien hyödyntämistä rautatie liikenteen analysoinnissa.

3 Rautatieliikenteen täsmällisyystiedon jalostamisen ja hyödyntämisen nykytila

Tässä luvussa esitellään tutkimuksen empiirinen tausta-aineisto: Suomen rautatieliikenteen täsmällisyysanalyysiin liittyvä data, tällä hetkellä tehtävät analyysit sekä työryhmät, joissa analyysijä käytetään. Toimintatutkimuksen osallistuvan luonteen vuoksi esitellään myös, miten näihin aineistoihin vaikutettu tutkimuksen tulosten saavuttamiseksi. Pääasiassa tämä tarkoittaa datan vaatimien muokkausten esittämistä. Myös analyysistä sekä työryhmistä osa on sellaisia, joiden tekemiseen ja toimintaan tutkimuksen aikana on osallistuttu toimintatutkimusmenetelmän esittämällä tavalla.

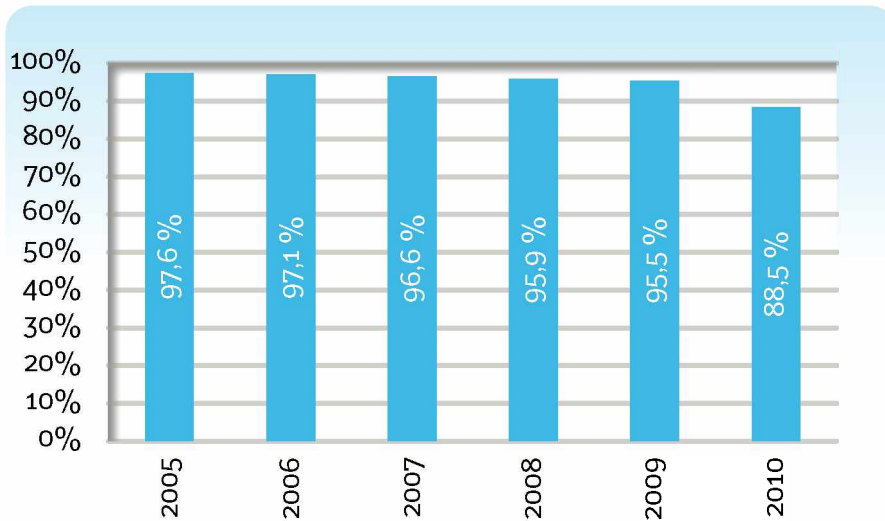
3.1 Täsmällisyyden perusseuranta

Suomessa seurataan erikseen Helsingin seudun lähiliikenteen, kaukoliikenteen ja tavaraliikenteen täsmällisyyttä, kutakin eri kriteerein ja tavoittein. Seurannan pohjana toimii JUSE-kulikutietojärjestelmä, jota on käsitelty luvussa 3.3. Täsmällisyysseurannan kriteerit ja tavoitteet liikennelajeittain on koottu taulukkoon 2.

Taulukko 2. Koonti täsmällisyyden tavoitteista Suomessa (muokattu lähteestä Liikennevirasto 2011)

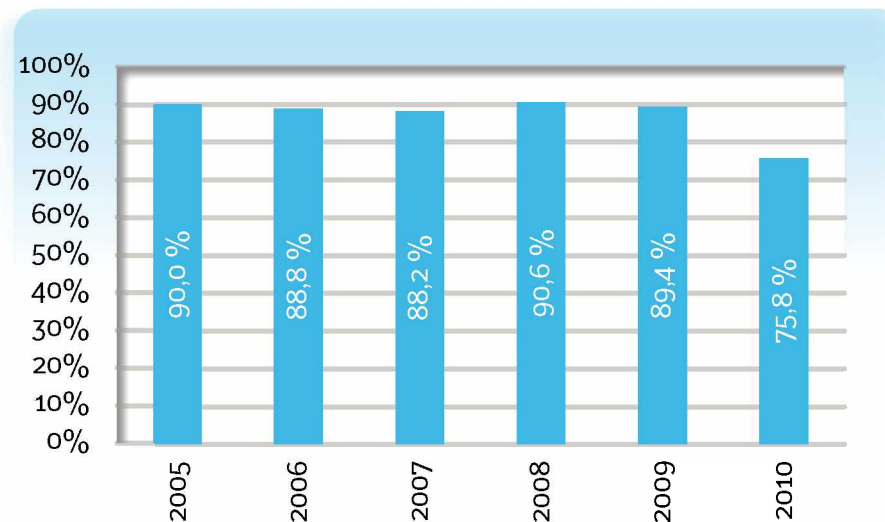
Liikennelaji	Sallittu viive	Mittauspaikka	Kokonais-täsmällisyyden tavoiteprosentti	Mittauksen ajanjaksot
Henkilökaukoliikenne	≤ 5 min	Määräasema	90 %	Kuukausittain & vuosittain
Helsingin seudun lähiliikenne	< 3 min	Lähtö- ja määräasema	97,5 %	Kuukausittain & vuosittain
Tavaraliikenne	≤ 15 min	Määräasema	90 %	Kuukausittain & vuosittain

Helsingin seudun lähiliikenteessä täsmälliseksi on määritelty juna, joka poikkeaa aikataulustaan lähtö- ja määräasemallaan alle kolme minuuttia (mikäli juna on ajoissa vain toisella näistä, on se 50 %:sesti ajoissa). Näin mitattuna lähiliikenteen täsmällisyystavoite on 97,5 prosenttia. Tavoitteen toteutumista seurataan kuukausittain ja vuosittain. Vuosittainen toteuma on esitetty kuvassa 12.



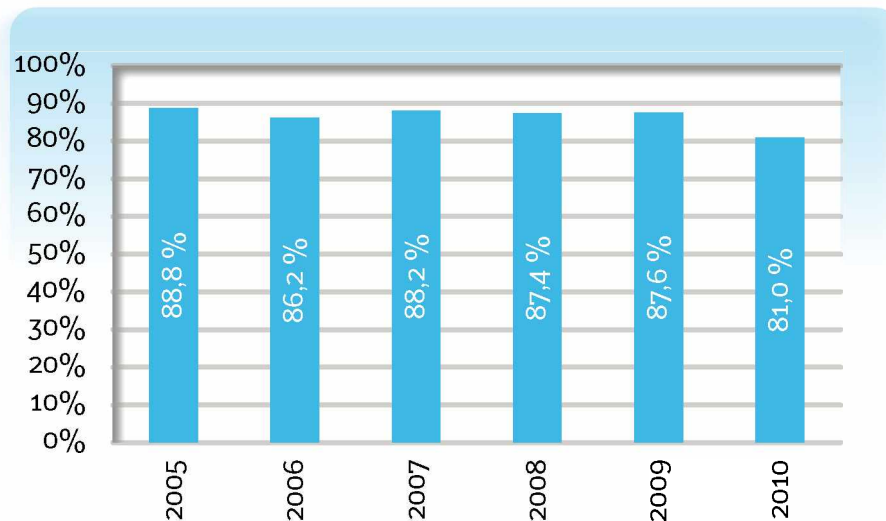
Kuva 12. Helsingin seudun lähiliikenteen täsmällisyys vuodesta 2005, jolloin nykyinen junien seurantajärjestelmä otettiin käyttöön (muokattu lähteestä Liikennevirasto 2011).

Kaukoliikenteessä täsmälliseksi on määritelty juna, joka poikkeaa aikataulustaan määräasemallaan korkeintaan viisi minuuttia. Näin mitattuna kaukoliikenteen täsmällisyystavoite on 90,0 prosenttia. Tavoitteen toteutumista seurataan kuukausittain ja vuosittain. Vuosittainen toteuma on esitetty kuvassa 13.



Kuva 13. Henkilökaukoliikenteen täsmällisyys vuodesta 2005, jolloin nykyinen junien seurantajärjestelmä otettiin käyttöön (muokattu lähteestä Liikennevirasto 2011).

Tavaraliikenteessä täsmälliseksi on määritelty juna, joka poikkeaa aikataulustaan määräasemallaan korkeintaan viisitoista minuuttia. Näin mitattuna tavaraliikenteen täsmällisyystavoite on 90,0 prosenttia. Tavoitteen toteutumista seurataan kuukausittain ja vuosittain. Muista poiketen tavaraliikenteessä seurataan myös etuajassa kulua, joka on jopa suurempi ongelma kuin myöhästymiset.



Kuva 14. Tavaraliikenteen täsmällisyys vuodesta 2005, jolloin nykyinen junien seurantajärjestelmä otettiin käyttöön (muokattu lähteestä Liikennevirasto 2011).

Euroopassa käytössä olevat täsmällisyyden mittarit eivät ole täysin vertailukelpoisia keskenään. Joka tapauksessa Suomi on perinteisesti ollut eurooppalaisittain suhteellisen täsmällinen rautatiemaa. Viimeisten parin vuoden aikana tilanne on kuitenkin heikentynyt merkittävästi: kuten kuvista 12–14 voidaan havaita, on täsmällisyystrendi ollut useampana vuonna laskeva. Tämä osaltaan osoittaa, että täsmällisyyteen liittyvää työtä tulee tehdä.

3.2 Viiveiden luokitteluperiaate

Vaikka täsmällisyyden perusseurannassa keskitytään junien määräasemiin (lähiliikenteessä myös lähtöasemiin), viiveitä ja niiden syitä kirjataan myös junan matkan aikana. Kulktieto tallennetaan jokaisessa seurantapisteessä, sekä saavuttaessa että lähdetäessä. Mikäli juna on seurantapisteessä myöhästymistoleranssin ulkopuolella, kirjataan viiveelle myös syy. Tämän jälkeen kirjataan uusi syy, mikäli juna jää minuutinkin lisää myöhään. Junan seurantapisteillä – alku- väli- ja määräasemilla – saamia viiveitä kutsutaan yleisesti *lisämyöhästymisiksi*. On tärkeää huomata, että juna voi saada matkan aikana paljonkin lisämyöhästymisminuutteja, ja olla silti perillä aikataulussaan. Täsmällisyyden perusseuranta ja lisämyöhästymisiä ei siis tule sekoittaa keskenään.

Suomessa viiveet luokitellaan niiden syyn mukaan. Syykoodeja on yhteensä 60 kappaletta, 12 eri luokassa. Osa näistä syykoodeista (ts. syistä) on määritelty primäärisiksi, osa puolestaan sekundäärisiksi. Lisäksi tavarajunien etuajassa kululle on olemassa 7 eri syykoodia. Syyt kirjataan JUSE-kulktietojärjestelmään, jota on käsitelty luvussa 3.3. Syykoodit on esitetty taulukossa 3.

Taulukko 3. Suomessa käytettävän tapahtumaperustaisen myöhästymisten kirjauksen syy-luokat ja -koodit (mukailtu lähteestä Blomqvist 2011).

AIKATAULUPOIKKEAMISSA KÄYTETTÄVÄT SYYKOODIT

Liikenneonnettomuudet

- O1 ALLEJÄÄNTI (IHMINEN)
- O2 ALLEJÄÄNTI (ELÄIN)
- O3 TASORISTEYSONNETTOMUUS
- O4 MUUT ONNETTOMUUKSET JA VAURIOT

Matkustajapalvelu

- M1 MATKUSTAJARUUHKA
- M2 MATKUSTAJIEN AIHEUTTAMIA HÄIRIÖITÄ
- M3 AIKATAULUSTA POIKKEAVA PYSÄHDYS
- M4 PASSI- JA TULLITARKASTUS
- M5 VANKIEN KUORMAUS JA PURKU
- M6 PYSÄHTYMISAJAN YLITYS

Tavarapalvelu

- T1 TAVARAN KUORMAUS JA PURKU
- T2 POSTIN KUORMAUS JA PURKU
- T3 RAVINTOLAVAUNUN KUORMAUS
- T4 AIKATAULUSTA POIKKEAVA PYSÄHDYS

Liikennetekniset syyt

- L1 YHTEYSLIIKENTEEEN ODOTUS
- L2 JUNAKOHTAUS, EDELLÄ KULKEVA JUNA TAI SIVUUTUS
- L3 AHTAUSRATAPIHALLA
- L4 RISTEÄVÄT KULKUTIET
- L5 MYÖHÄSTYMINEN ULKOMAILTA
- L6 HÄIRIÖ PÄIVYSTYSTYÖSSÄ
- L7 TULOJUNA MYÖHÄSSÄ
- L8 LIIKENTEEHOITOVIRHE

Henkilökunta

- H1 HENKILÖKUNNAN VAIHTO
- H2 HENKILÖKUNNAN ODOTUS
- H3 MUUT SYYT

Junakokoonpano

- J1 VAUNUJEN OTTO TAI JÄTTÖ
- J2 VAUNUJEN TAI VAUNURYHMIEN ODOTUS
- J3 JARRUJEN KOETTELU
- J4 ERIKOISKULJETUS
- J5 ALENNETTU SN

Veturit

- V1 VETURIN ODOTUS
- V2 VETURIVIKA JA JKV-VIKA VETURILAITTEESSA
- V3 VETOVOIMAN PUUTE
- V4 VETURIN VAIHTO, LISÄYS TAI POISTO

Moottorijunat ja vaunut

- K1 JARRUVIKA
- K2 LAAKERIVIKA
- K3 JUNAN KATKEAMINEN
- K4 KYTKENTÄ TAI IRROITUS (Sm / Dm)
- K5 KALLISTUSVIKA (Sm3)
- K6 LOVIPYÖRÄ
- K7 MUU VIKA

Rata

- R1 TILAPÄISET NOPEUSRAJOITUKSET
- R2 ESTE RADALLA
- R3 RADAN KUNNOSSAPITO- JA RAKENNUSTYÖT
- R4 RATATYÖN SOVITUN AJAN YLITYS

Sähköistys

- S1 JÄNNITEKATKO
- S2 TEKNISET VIAT
- S3 SÄHKÖRADAN KUNNOSSAPITO- JA RAKENNUSTYÖT
- S4 HÄIRIÖ VALTAKUNNAN VERKOSSA

Turva-, valvonta- ja viestilaitteet

- P1 TURVALAITEVIKA
- P2 OPASTINVIKA
- P3 VAIHDEVIKA
- P4 JKV-VIKA RATALAITTEISSA
- P5 LINJA- TAI RATAPIHARADIOVIKA TAI PUHELINVIKA
- P6 GSM-R -VIKA
- P7 VALVONTALAITEVIKA TAI AIHEETON HÄLYTYS

Muut syyt

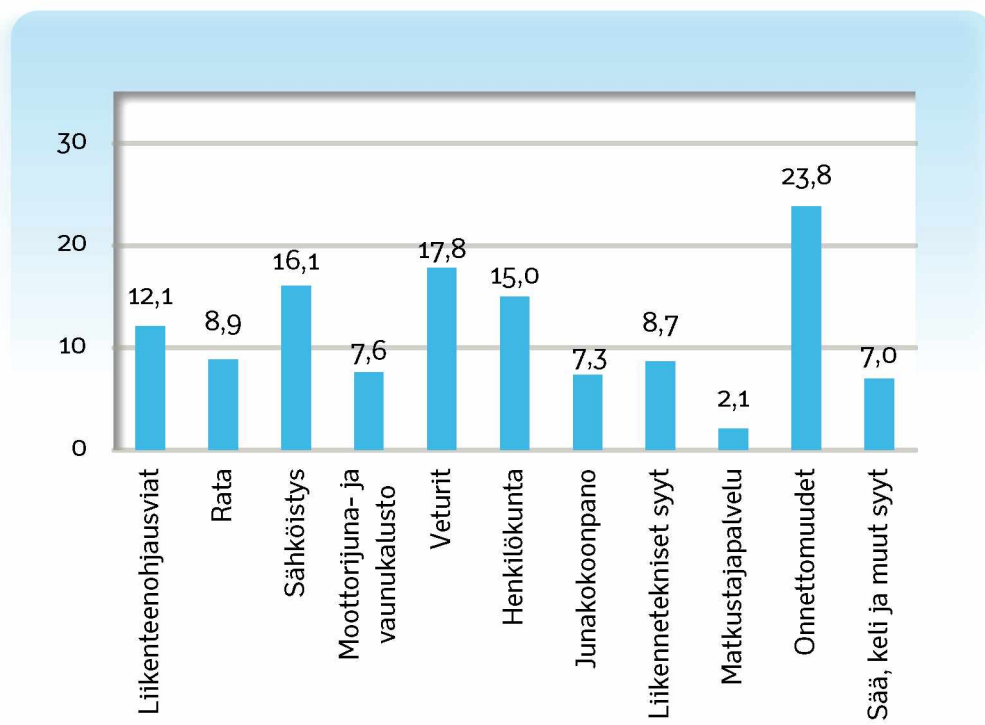
- I1 SÄÄ (SUMU, HUONO KELI TMS)
- I2 LUMIESTEET
- I3 ASIATTOMAT RADALLA LIIKKUJAT TAI ILKIVALTA
- I4 MUUT HÄIRIÖT

Etuajassakulku (vain tavaraliikenteessä)

- E1 TULOJUNA ETUAJASSA
- E2 VAIHTOTÖITÄ VÄHÄN TAI EI OLLENKAAN
- E3 PIENI JUNAKOKO
- E4 KÄÄNTÖAJAN ALITUS
- E5 VETURINAKULKU
- E6 AJOAJAN ALITUS / LIIKENNETEKNISET SYYT
- E7 MUU SYY

Viiveitä aiheuttavat syyt voidaan jakaa liikennöitsijästä johtuviin, radanpitäjästä johtuviin sekä ulkopuolisiin syihin. Liikennöitsijästä johtuvat viiveet liittyvät usein junakaluston toimimattomuuteen, henkilökunta- ja kalustokiertoihin sekä junakokoonpanoihin. Radanpitäjästä johtuvat viiveitä ovat puolestaan radan huonosta kunnosta johtuvat tilapäiset nopeusrajoitukset, radan kunnossapito- ja rakennustyöt, liikenneohjausjärjestelmien ja turvalaitteiden viat sekä sähkövauriot. Ulkopuolisia syitä ovat sää, keli, onnettomuudet allejäännit ja ilkivalta. (Liikennevirasto 2011)

Vaikka primääriset viiveet usein lajitellaan rataverkosta johtuviin viiveisiin, liikennöinnistä johtuviin viiveisiin sekä ulkopuolisista tekijöistä johtuviin viiveisiin, ovat nämä tekijät kuitenkin myös osittain keskenään riippuvaisia. Esimerkiksi turvalaitteet ovat herkkiä ukkoselle ja vaihteet tuiskuavalle lumelle. Myös viallisella kalustolla ajo saattaa rikkoo rataa tai rata saattaa rikkoo kalustoa. Primääriset viiveet ovat keskenään myös hyvin erilaisia vaikuttavuudeltaan. Osa viiveistä on hyvin lyhytkestoisia, kuten matkustajaruuhkasta aiheutuneet viiveet. Osa taas on hyvin pitkäkestoisia, kuten radan rikkoutuminen tai veturin rikkoutuminen yksiraiteiselle osuudelle, jolloin koko liikenne tukkiutuu. (Liikennevirasto 2011). Kuvassa 15 on esitetty primääristen viiveiden keskimääräinen kesto vuonna 2010.



Kuva 15. Eri syyryhmien aiheuttamien myöhästymisten keskimääräinen kesto (minuuttia) henkilökauliikenteessä (muokattu lähteestä Liikennevirasto 2011).

Tällä hetkellä primäärisiksi viiveiksi katsotaan kaikki muut paitsi myöhästymiskoodit L1, L2 ja L7. (Blomqvist 2011). L1 tarkoittaa yhteysliikenteen odotusta, jolloin juna odottaa toisesta junasta tai muusta yhteysliikennemuodosta saapuvia matkustajia. L2-koodi tarkoittaa junakohtausta, edellä kulkevaa junaa tai sivuutusta, joka tapahtuu tyypillisimmin, kun yksiraiteisella osuudella toinen junista on viivästynyt aikataulustaan ja tästä johtuen toinen joutuu odottamaan kohtaauspaikalla, että alun perin myöhässä oleva juna saapuu kohtaauspaikalle. Tulojuna myöhässä, syykoodiltaan L7, puolestaan kuvaa sitä, että vaadittu kalusto on jäänyt jonain toisena junana myöhään. Tällöin se juna, jona kaluston pitäisi lähteä uudelle reitille, viivästyy.

L2-viiveet, eli viiveet joiden syynä ovat junakohtaukset, edellä kulkevan hitaamman junan perässä kulku sekä sivuutukset, johtuvat hyvin pitkälti Suomen rataverkon yksiraiteisista osuuksista, joiden osuus on yli 90 % koko verkosta. Näillä rataosuuksilla on olemassa vain rajoitettu määrä lyhyitä kohtaustaikoja, joissa junat pääsevät toi-

sistaan ohi. Tämä tarkoittaa sitä, että mikäli junat eivät osu kohtauspaikalle silloin, kun aikatauluissa on suunniteltu, aiheuttaa se vääjäämättä viivettä junalle, joka joutuu odottamaan kohtausta.

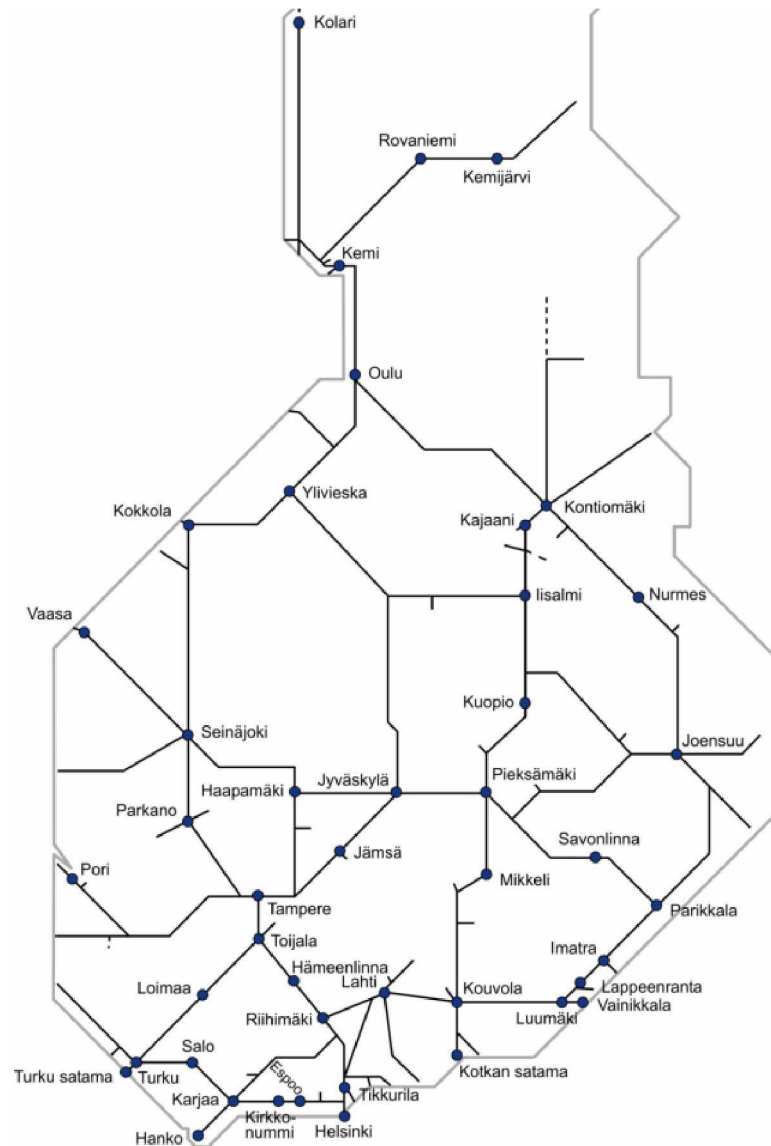
L1-, L2- ja L7-viiveiden lisäksi on myös monia muita viiveitä, jotka ovat luonteeltaan sekundäärisiä. Esimerkiksi L3 (ahtaus ratapihalla) ja L4 (risteävät kulkutiet) ovat hyvin vahvasti sekundäärisluontoisia. Lisäksi junakokoonpanoon liittyvät syyt J1–J5, Henkilökuntaan liittyvät H1–H3 ja pysähtymisajan ylitykseen asemilla liittyvät M1 ja M6 sekä liikennetekniset syyt L5, L6 ja L8 ovat osittain sekundäärisluontoisia. (Matti-la 2011). Tällä hetkellä raportoinnissa otetaan kuitenkin vain syykoodit L1, L2 ja L7 huomioon sekundäärisinä syinä. Lähitulevaisuudessa Suomen rautatietoimijoiden tulisi tarkentaa omaa kirjaamis- ja raportointikäytäntöjensä ymmärtääkseen paremmin, ovatko viiveet aidosti primäärisiä vai sekundäärisiä, jolloin on helpompi päästä viiveiden juurisyihin käsiksi.

3.3 Data ja analyysiympäristö

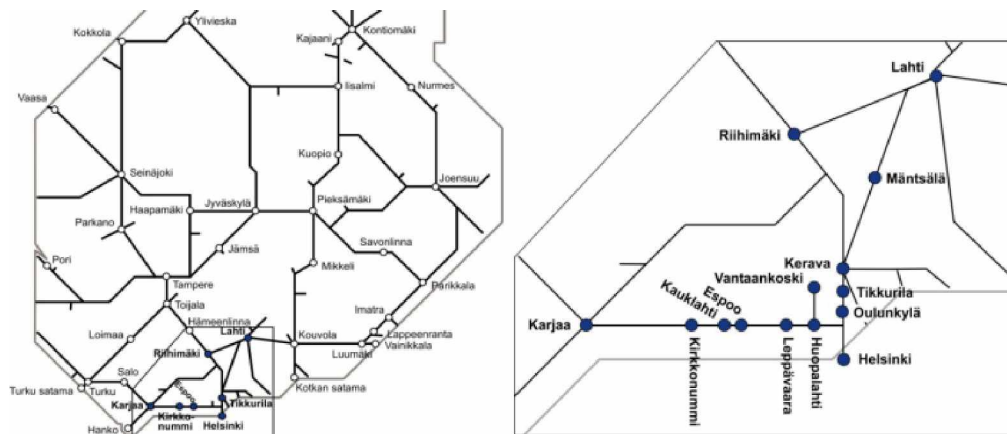
3.3.1 JUSE-data

JUSE, eli junien seurantajärjestelmä, on Liikenneviraston junien kulkutiedon järjestelmä. Kulkutiedot kirjautuvat seurantajärjestelmään liikenteenohjausjärjestelmistä tai tarvittaessa liikenteenohjaajien syöttäminä. Lisäksi liikenteenohjaaja kirjaa järjestelmään junien kulkuun liittyvää tietoa, kuten myöhästymissyöt, syyn lisätietoja sekä muita mahdollisia tietoja (Liikennevirasto 2011). Myöhästymissyöiden kirjaamisessa käytetään taulukossa 3 esitettyä koodistoa.

Kuvassa 16 on esitetty henkilökaukoliikenteen JUSE:n seuranta-asemat joita on noin 50 kappaletta ja kuvassa 17 puolestaan Helsingin seudun lähiliikenteen seuranta-asemat, joita on JUSE:ssa 15 kappaletta. Tavaraliikennettä varten JUSE:ssa on lisäksi joitakin lisäseuranta-asemia, esimerkiksi tavara-asemien yhteydessä. Vuodesta 2005 lähtien käytössä ollut JUSE:sta liikenteenohjaus pystyy seuraamaan reaaliajassa, mikä on junien täsmällisyyden tilanne suhteessa seuranta-asemiin, joka helpottaa muun muassa häiriötilanteiden hallintaa (Liikennevirasto 2011). JUSE:sta välittyy myös VR:n Internet-sivustolla esitettävä junien kulkutieto, josta matkustajat voivat seurata junien kulun tilannetta.



Kuva 16. Henkilökaukoliikenteen täsmällisyyden seuranta-asetat junien seurantajärjestelmässä vuonna 2010 (Liikennevirasto 2011a).



Kuva 17. Helsingin seudun lähiliikenteen täsmällisyyden seuranta-asetat junien seurantajärjestelmässä vuonna 2010 (Liikennevirasto 2011a).

JUSE koostuu aktiivikannasta sekä historiakannasta. Aktiivikannan tiedot ovat varsin kattavia, kuten liitteessä 1 esitetystä aktiivikannan tietokantakuvasta käy ilmi. Aktiivikannassa tietoja tallennetaan kuluvaan päivän ja kahden edeltävän päivän ajalta. Tämän jälkeen tiedot siirretään soveltuvin osin historiakantaan. (Tapola 2011). **Tämän tutkimuksen pääasiallisena lähtödatana oli historiakanta vuodelta 2010**; tutkimuksessa käytetyt historiakannan taulut tietosisältöineen on esitetty liitteessä 2. Lisäksi tähän dataan rakennettiin uusia tietokenttiä seuraavasti:

- *LIIKENNEPAIKKA_LAHTO_RECODED = Alkuperäisessä datassa junan liikennepaikkakohtaisissa tulo- ja lähtötiedoissa käytettiin samaa liikennepaikkainumeroa. Tämä koodattiin sekvenssianalyysiä varten siten, että lähtö on myöhemmin kuin tulo.*
- *LISÄMYÖHÄSTYMINEN = Kuinka monta minuuttia juna on viivästynyt lisää verrattuna edelliseen seurantapisteeseen ja tapahtumakoodiin. Laskenta on toteutettu järjestelemällä junat kulkujärjestykseen ja laskemalla erotukset.*
- *RATAOSA = Liikennepaikkatieto uudelleenkoodattuna niin, että lähdöt (mahdollinen viive tapahtunut asemalla) erottuvat saapumisista (mahdollinen viive tapahtunut rataosalla).*
- *SYYNLISATIETO_EXTRACTED = Syyn lisätieto -kentästä eriytetty junanumerot.*
- *LKP_SYY_JOINED = RATAOSA ja SYYKOODI -kenttien tiedot yhdistetty peräkkäisiksi.*

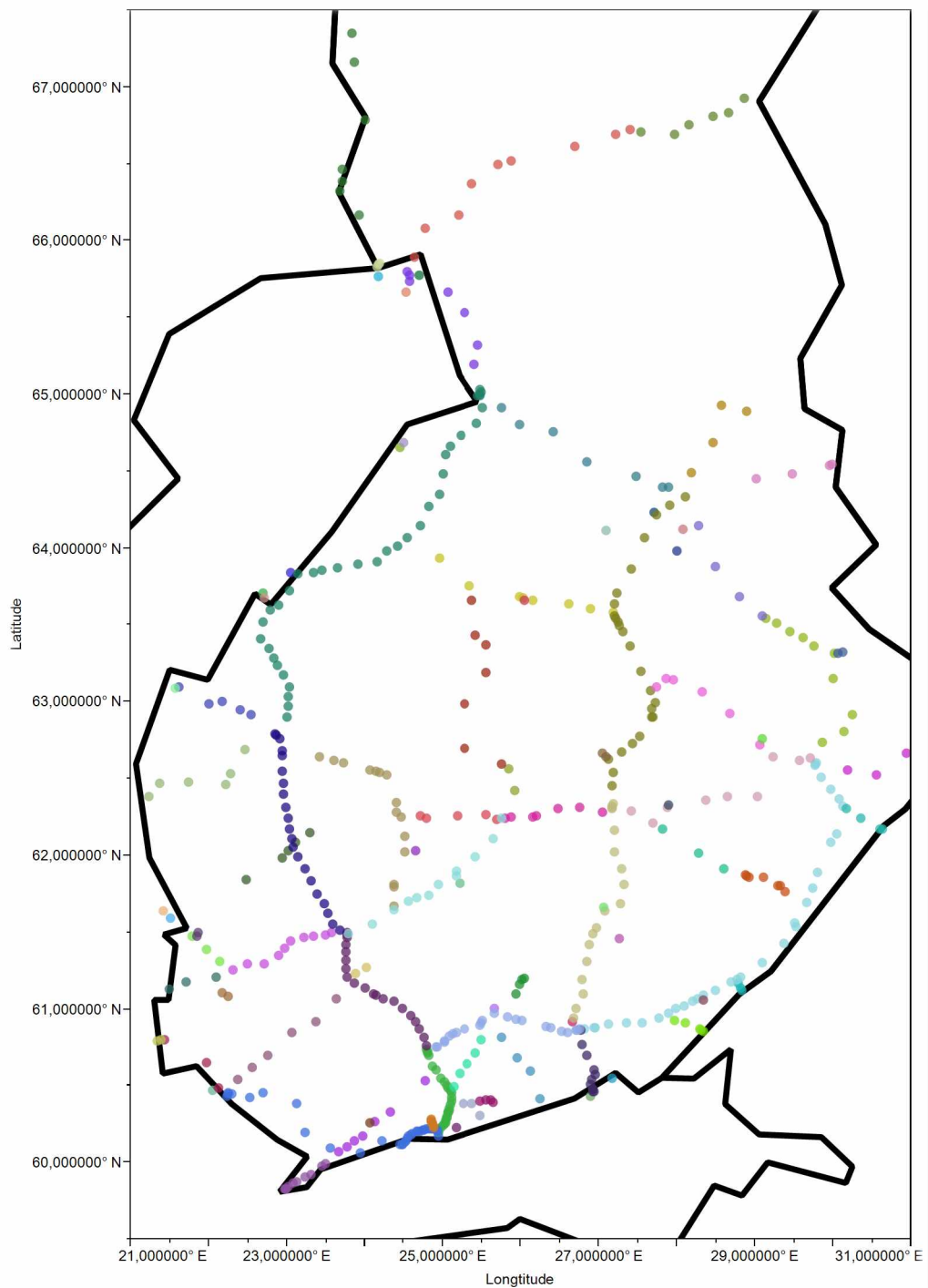
JUSE-datassa voidaan todeta olevan analyttistä käyttöä silmällä pitäen useita puutteita, joita on käsitelty luvussa 3.3.3. Näistä puutteista huolimatta tässä tutkimuksessa käytettiin pääasiassa nimenomaan JUSE-dataa. Näin saavutettiin vertailukelpoisuus aikaisemmin tehtyihin analyyseihin.

3.3.2 GRATU-data

GRATU on graafisiin aikataulujärjestelmiin liittyvä järjestelmä, johon kerätään dataa junien kulusta. GRATU-datan seuranta-asemien määrä, 324 kappaletta, on huomattavasti kattavampi kuin JUSE:ssa. Nämä seuranta-asetat on esitetty kuvassa 19.

Analyseissä GRATU-datan tarkkuutta ei päästä hyödyntämään täysimääräisesti. Ongelmana on, että syykoodit kirjautuvat GRATU:un JUSE:n kautta. Näin ollen syytietoon liittyvä analysointi pystytään tekemään vain JUSE-datan mahdollistamalla tarkkuudella. Niinpä useimmat tämän tutkimuksen analyysit tehtiin JUSE-datalla sen helppokäyttöisyyden ja vertailukelpoisuuden vuoksi.

Tarkkuuden ohella GRATU-datan vahvuutena on se, että se sisältää myös kellonaikatieidot. Niinpä tutkimuksessa hyödynnettiin GRATU-dataa sellaisissa analyyseissä, joissa junien kulkuajankohdalla oli merkitystä. Tällainen oli esimerkiksi Helsingin ratapihan ruuhkaisuustarkastelu.



Kuva 18. GRATU-datan täsmällisyyden seuranta-asemat.

GRATU-datasta tähän tutkimukseen ajettiin seuraava osadata:

- *PÄIVÄ* = Päivä, jolle juna kirjataan (Identifioi junat toisistaan)
- *JUNATYYPPI* = Minkä tyypin juna on kyseessä
- *JUNANUMERO* = Junan numero (Eri päivinä samaan aikaan ja samaa reittiä kulkevat junat menevät samalla numerolla)
- *SN* = Junan nopeusrajoitus
- *LÄHTÖ* = Asema, josta juna alun perin lähti
- *TULO* = Asema, johon juna lopulta saapuu
- *LIIKENNEPAIKKA* = Paikka, jossa kyseinen havainto tehti

- *AIKAT TULO = Aikataulun mukainen tuloaika*
- *TOTEUTUNUT TULO = Toteuman mukainen tuloaika*
- *AIKAT LÄHTÖ = Aikataulun mukainen lähtöaika*
- *TOTEUTUNUT LÄHTÖ = Toteuman mukainen lähtöaika*
- *MYÖH TULO = Toteuman minuuttimääräinen poikkeama tulon aikatauluajasta*
- *MYÖH LÄHTÖ = Toteuman minuuttimääräinen poikkeama lähdön aikatauluajasta*
- *SYKKOODI = Syykoodi, mikäli juna on enemmän, kuin 5 minuuttia myöhässä. Kirjautuu JUSEn mukaisten seuranta-asemien mukaan.*

3.3.3 Arvio datan ja dataympäristön nykytilasta

Kun käytössä olevaa dataa ja laajemmin dataympäristöä tarkastellaan analytiikan näkökulmasta, nousee esiin joitakin haasteita. Ensinnäkin täsmällisyyteen liittyvät tietojärjestelmät ovat aikanaan tehty hyvin suppeaa käyttötarkoitusta ajatellen, eikä analytiikkaa ole tällöin juuri otettu huomioon. Näin ollen erityisesti **JUSE-järjestelmän tarkoituksenmukaisuus ei tällä hetkellä ole paras mahdollinen**. Kirjautusjärjestelmä ja siihen liittyvä syykoodisto (ks. taulukko 3) on puutteellinen monin osin. Tutkimuksen aikana esille nousseita ongelmia ovat:

- *Seurantapisteitä on liian harvassa.* Ne eivät esimerkiksi anna ei mahdollisuutta tarkastella, missä kohtaa yhteysväliä viive on tapahtunut, tai varsinkaan, millä raiteella. Ongelma korostuu niillä verkon osilla, joissa asemia (ja siten seurantapisteitä) on harvassa.
- *Eri junalajeilla on käytössä erilainen seurantapisteverkosto.* Esimerkiksi Venäjän lisäjunien kulkua ei seurata yhtä tarkasti kuin muiden junien, jolloin niistä ei voida tehdä yhtä tarkkoja analyysejä kuin muista junista. Lisäksi junalajien välinen vertailu on vaikeaa. Asia olisi korjattavissa parametreja muuttamalla.
- *Tietojen muokkaaminen suunnitelmien muuttuessa on hankalaa tai mahdotonta.* Esimerkiksi supistamissuunnitelman mukaisen toimenpiteen, K- ja I-junien muuttaminen N-juniksi, kirjaaminen JUSE:en ei ole mahdollista, mikä vääristää tietoja. Tähän on kuitenkin tulossa parannus.
- *Aikaleimojen tarkkuus on melko huono.* Minuuttipohjaisuuden lisäksi epätarkkuutta aiheuttaa se, että seurantapisteisiin on jouduttu lisäämään erilaisia korjausaikoja, koska ne eivät sijaitse asemien välittömässä läheisyydessä.
- *Havaintojen yhteyteen ei kirjaudu aikataulunmukainen aika, ainoastaan poikkeama siitä.* Tämä hankaloittaa analysointia merkittävästi.
- *Viiveen syy merkitään vasta, kun se ylittää myöhästymisen rajan* (kaukoliikenteessä 3 min, lähiliikenteessä 2 min), *vaikka tämän jälkeen minuutinkin lisämyöhästymiselle merkitään syy.* Tämä vääristää syykohtaisia analyysejä.
- *Ensimmäisten viiveiden erottaminen lisäviiveistä ei ole mahdollista.* Usein lisäviiveet ovat ainakin osittain seurausta ensimmäisestä viiveestä. Näin ollen näiden viiveiden tarkastelu samanarvoisina hukkaa arvokasta tietoa.
- *Kullekin viiveelle voi olla vain yksi syy.* Monissa tapauksissa yksittäinen viive voi olla monen tekijän summa. Näissä tilanteissa tietyille syykoodeille lankeaa todellista enemmän viiveminuutteja.
- *Kaikki syykoodit ovat järjestelmässä erillisiä ja samantasoisia,* vaikka osa niistä on todellisuudessa päällekkäisiä tai alisteisia toisilleen (esim. R1: tilapäiset nopeusrajoitukset ja R3: radan kunnossapito- ja rakennustyöt). Tämän vuoksi samankaltaisia tilanteita kirjataan eri tavoin.
- *Myös primääriset ja sekundääriset koodit ovat järjestelmässä samantasoisia* (vasta myöhemmin on päätetty, mitkä koodeista on primäärisiä, mitkä sekundäärisiä). Todellisuudessa osa koodeista voi kuitenkin tapauksesta riippuen

olla joko primäärisiä tai sekundäärisiä (esim. H2: henkilökunnan odotus). Tämä vaikeuttaa viiveiden ketjuuntumisen analysointia.

- *Sekundäärisissä viiveissä aiheuttavan junan numero merkitään vapaamuotoiseen kenttään.* Koska kenttään voi kirjoittaa minkä tahansa merkkijonon, ei junanumeroiden koneellinen luku onnistu. Kentän voi myös jättää tyhjäksi. Näin ollen sekundääristen viiveiden analysointi on vaikeaa.
- *Syykirjauksiin ei merkitä vastuutahoa,* vaan se päätellään koodista. Kaikissa tapauksissa koodi ei kuitenkaan kerro todellista vastuutahoa. Ongelma korostuu niissä syissä, jotka voivat olla joko primäärisiä tai sekundäärisiä.
- *Junilla ei ole yksilöllisiä ID-numeroita.* Tämä hidastaa tietojen käsittelyä ja yhdistämistä muihin tietoihin.

Edellä mainittujen JUSE-järjestelmän teknisten puutteiden lisäksi ongelmana on, että ***kirjaamiskäytännöt ovat monin osin melko kirjavia.*** Toisin sanoen samankaltaisia asioita kirjataan monin eri tavoin. Lisäksi vapaaehtoiset kentät jätetään usein täyttämättä. Tästä seuraava datan epätarkkuus ja puutteellisuus heijastuu luonnollisesti analyyseihin. Tällä hetkellä ongelma koetaan erittäin suureksi: monia analyysejä jätetään jopa tekemättä tai hyödyntämättä sen takia, että data oletetaan virheelliseksi.

JUSE-järjestelmään liittyvien ongelmien lisäksi ***haasteena on myös, ettei analytiikka tähän asti ole ollut organisoitu toiminto.*** Analyysejä ovat tehneet muutamat asiantuntijat omiin ja muiden tarpeisiin sillä datalla, mitä on ollut saatavilla, käyttäen olemassa olevia Oracle- ja Office-työkaluja. Ongelmana siis on:

- *Analytiikkaa varten ei ole olemassa omaa, tehokasta datan tallennus- ja prosessointiympäristöä.* Nykyisellä tavalla jo perusraporttien teko on hidasta, kankeaa ja manuaalista. Joustaviin ad-hoc-analyyseihin ei näin ollen useinkaan riitä resursseja.
- *Tämänhetkiset analyysityökalut ovat sellaisia, että niiden avulla kyetään luomaan klassisia analyysejä vain suppeasta datajoukosta.* Kaikki datan muokkaus ja analysointi tapahtuu Oraclen ja MS Excelin perustyökaluilla.

3.4 Analyysit

3.4.1 Kuukausittaiset analyysit

Tällä hetkellä täsmällisyyden analysoinnin ja täsmällisyyteen liittyvän tiedonjakamisen kulmakivinä toimivat Liikenneviraston ja VR:n kuukausittain laatimat sisäiset kuukausiraportit rautatieliikenteen täsmällisyydestä. Tässä luvussa esitellään näiden kuukausiraporttien keskeinen sisältö ja raportointimenetelmät. Esimerkkiraportteina toimivat Mattilan & Matinlaurin (2011) sekä Kaijaan (2011a, 2011b, 2011c) laatimat kuukausiraportit huhtikuun 2011 osalta.

Liikenneviraston täsmällisyyden kuukausiraportointi

Liikennevirasto laatii jokaisen kuukauden alkupuolella edellisen kuukauden täsmällisyyttä käsittelevän MS PowerPoint -muotoisen raportin. Monet raportin tiedoista ovat suuntaa-antavia, johtuen edellisessä luvussa käsitellyistä lähtödatan virheistä ja puutteista. Raportti jaetaan Liikenneviraston sisäisesti erikseen määritellyille henkilöille sekä lisäksi VR:n yhteyshenkilöille.

Raportin alussa esitellään menneen kuukauden täsmällisyyden yleistila kaikkien kolmen eri rautatieliikennelajin osalta täsmällisyysprosenttien muodossa. Tämän jälkeen esitellään henkilökaukoliikenteen tilaa tarkemmin. Henkilökaukoliikenteen tilan kuvauksessa hyödynnetään karttavisualisointeja: rataosittain ja liikennepaikoittain syntyviä viiveitä on tarkasteltu kolmen eri kartan avulla:

- primääriset viiveet
- sekundääriset viiveet (L1, L2 ja L7 -syyt) ja
- radanpidosta johtuvat viiveet (R, S ja P -syyluokat).

Muutoin henkilökaukoliikenteen analyysi koostuu pääasiassa pylväsdiagrammeista, joilla on esitetty mm.

- täsmällisyysprosentin kehittyminen kuukausittain
- kyseisen kuukauden päivittäiset täsmällisyysprosentit
- eri kunnossapitoalueille aiheutuneet viiveet kuukausittain ja
- kunnossapitoalueille radanpidosta aiheutuneet viiveet päivätasolla.

Henkilökaukoliikenteen esittelyn jälkeen raportti siirtyy käsittelemään Helsingin seudun lähiliikenteen täsmällisyyttä. Lähiliikenteen osuus on huomattavasti suppeampi kuin kaukoliikenteen johtuen lähinnä siitä, että analysoitavaa rataverkkoa on huomattavasti vähemmän. Lisäksi lähiliikenteeseen liittyvän datan tarkkuus on melko karkealla tasolla verrattuna järjestelmän kompleksisuuteen. Analyysissä painottuu pylväsdiagrammeihin, esitellen menneiden kuukausien täsmällisyysprosentteja, viivemiinuuttien muodostumista syylokittain sekä radanpidon aiheuttamia viiveitä Helsingin seudun lähiliikenteessä.

Tavaraliikenteen suhteen analyysi on vieläkin suppeampi: analyysi koostuu kahdesta pylväsdiagrammista. Ensimmäinen esittää tavaraliikenteen täsmällisyysprosentin kehittymistä kuukausittain, toisen esittäessä menneen kuukauden tavaraliikenteen päivittäisen täsmällisyyden.

Kokonaisuutena voidaan todeta Liikenneviraston kuukausiraportin olevan hyvässä perustilassa. Tähän asti analyysiä on tehty pääasiallisesti manuaalisesti MS Excelillä, työn viedessä reilun työpäivän verran aikaa kuukausittain.

VR-Yhtymä Oy:n täsmällisyyden kuukausiraportointi

VR-Yhtymä Oy laatii kuukausittain kolme eri sisäistä kuukausiraporttia liikenteen täsmällisyydestä: kullekin liikennelajille, eli henkilökaukoliikenteelle, Helsingin seudun lähiliikenteelle sekä tavaraliikenteelle tehdään oma MS PowerPoint -muotoinen kuukausiraporttinsa. Raportit sijoitetaan VR:n intranettiin, jossa ne ovat koko henkilöstön saatavilla. Uusista raporteista lähetetään sähköpostitiedote erikseen määritellyille VR:n henkilöille sekä Liikenneviraston yhteyshenkilöille.

VR:n henkilökaukoliikennettä esittelevä raportti alkaa menneen kuun täsmällisyysprosenttien yhteenvedolla. Tietyt priorisoidut junatyyppit on esitelty erikseen. Tämän jälkeen raportissa esitellään yhteenvedot menneen kuukauden merkittävimmistä viiveistä suhteessa ennusteeseen. Nämä viivemäärien ennusteet ovat merkittävä askel kohti kehittynyttä data-analytiikkaa, joskin kyseessä ovat pikemminkin tavoitteet, eivät ennusteet: ennusteet perustuvat kokemukseräiseen tietoon siitä, millä viivemiinuuttimäärillä täsmällisyystavoitteet ovat vielä saavutettavissa. Muutoin VR:n raportti esittelee kaukoliikenteen täsmällisyyttä hyvin samaan tyyliin kuin Liikennevirasto, painopisteen ollessa radanpidon sijaan junatyypeissä. Mainitsemisen arvoista on, et-

tä VR varsin ansiokkaasti luokittelee syntyneitä viiveitä minuuttiluokkiin, jolloin tarkastelu on taulukkojen yhteydessä selkeämpää.

Helsingin seudun lähiliikennettä koskeva raportointi muistuttaa hyvin paljon henkilökaukoliikenteen raportointia. Esiin nousee kuitenkin muutama ansiokas analyysi. Helsingin seudun lähiliikennettä on analysoitu erikseen liikennelinjoittain sekä kaupunkiradoittain ja sekakäyttöraiteittain, jolloin saadaan hyvin laaja-alainen käsitys täsmällisyyden tilasta, niin raiteiston kuin matkustajankin näkökulmasta.

Tavaraliikenteen täsmällisyyttä ei ole aikaisemmin pidetty liikennejärjestelmän kannalta kovinkaan tärkeänä osa-alueena. Tämän vuoksi myös sen analysointi on huomattavasti kevyempää kuin mitä henkilöliikenteen puolella, joskin Liikennevirasto on monipuolisempaa. Perusanalyysin lisäksi VR tutkii kuukausittain, mitkä junat ovat myöhästelleet eniten, missä ja mistä syystä. Myös tavarajunien etuajassakulkua raportoidaan kuukausittain.

VR:llä tehdään perusraportointia varsin perusteellisesti, mutta perinteisin menetelmin ja kuvaajin. Perusraportoinnin voidaankin katsoa olevan jopa kattavampaa kuin Liikennevirastolla. Tämä selittyy pitkälti sillä, että VR:ää kiinnostaa myös junatyypin täsmällisyys Liikenneviraston keskittyessä pääasiassa rataan liittyviin seikkoihin.

3.4.2 Vuosittainen analyysi

Liikennevirasto on laatinut vuosittain rautatieliikenteen edellisen vuoden täsmällisyyttä käsittelevän raportin. Tämä julkinen raportti on toiminut viestintävälineenä niin sisäisesti kuin myös ulkoisille sidosryhmille.

Rautatieliikenteen täsmällisyys -vuosiraportti alkaa yleisellä esittelyllä rautatieliikenteen täsmällisyydestä, Suomen rautatieverkosta ja -liikenteestä, eri viiveistä sekä rautatieliikenteen täsmällisyydestä Suomessa ja Euroopassa. Tämä antaa perehtymätömällekin lukijalle kattavan kuvan, millaisesta toimintaympäristöstä on kyse.

Vuosiraportin seuraavassa luvussa esitellään kyseisenä vuonna eniten epätäsmällisyyttä aiheuttaneet yksittäiset häiriöt. Tämän tyylinen yksittäisten merkittävien ilmiöiden esittäminen on varsin toivottavaa, koska nostamalla niitä aktiivisesti esille on mahdollista suunnata kehitystoimia oikeaan suuntaan. Vaikkakin luku nostaa esiin yksittäisiä merkittäviä täsmällisyyteen vaikuttavia ilmiöitä ja tapahtumia, ei näitä ole analysoitu kokonaisjärjestelmän kannalta.

Vuosiraportin neljäs, viides ja kuudes luku käsittelevät täsmällisyyttä henkilökaukoliikenteessä, Helsingin seudun lähiliikenteessä ja tavaraliikenteessä. Analyysit ovat varsin kattavia. Erityisesti luvussa 4 esitetyt aikasarja-analyysit primääristen ja sekundääristen viiveiden korrelaatiosta sekä karttavisuaalisointi rataosakohtaisesta täsmällisyydestä ovat varsin informatiivisia ja tarkastelevat järjestelmää muutoinkin kuin vain syykoodien ja minuuttien valossa. Muutoin analyysit noudattelevat perinteistä pylväsdiagrammi- ja piirakkakuvaaja-linjaa.

Kokonaisuutena Rautatieliikenteen täsmällisyys -vuosiraportti on ollut hyvin informatiivinen paketti. Siinä on nostettu esiin liikennettä kohdanneet suurimmat haasteet koko vuoden ajalta sekä tuotettu kattava perustason analyysi eri liikennelajeista.

Myös VR laatii vuosittaisia analyysejä, mutta vain sisäiseen käyttöön.

3.4.3 Ad hoc -analyysit

Vakioraporttien lisäksi sekä VR:n että Liikenneviraston analyytikot tuottavat tarpeen mukaan erilaisia ad hoc -raportteja. Yleensä tarve liittyy erilaisiin sisäisiin tai ulkoiisiin selvityspyyntöihin, joiden taustalla on jokin vakava täsmällisyysongelma. Muita ad hoc -analyysijä näiden käytännössä pakollisten analyysien lisäksi ei ole resursseja tehdä. Näin ollen toiminta on hyvin reaktiivista.

3.4.4 Arvio analyysien nykytilasta

Suomen rautatieliikenteen täsmällisyyden analysointi on tällä hetkellä raportoinnin tasolla: täsmällisyyttä raportoidaan pääasiassa kuukausittain ja vuosittain. Analytiikan näkökulmasta voidaan nostaa esille seuraavat haasteet:

1. Raportointi ja siihen liittyvät visualisoinnit suoritetaan tällä hetkellä pääasiassa manuaalisesti MS Excelillä. Tähän rutiinityöhön kuulu tarpeettoman paljon henkilötyöresursseja.
2. Raporttien hyödyntäminen vaikuttaa olevan vähäistä: suuri osa raporttien sisällöstä näyttää jäävän vaille huomiota ja vain merkittävimpiin haasteisiin tartutaan. Toisaalta tämä johtuu siitä, että toimintaympäristö on viime aikoina kohdannut poikkeuksellisen runsaasti merkittäviä haasteita, jolloin ne ovat varastaneet suurimman huomion.
3. Raporttien sisältö ei kaikilta osin vastaa todellisia tietotarpeita. Vakiomuotoisten raporttien tuottaminen kuukaudesta toiseen ei auta, jos sisältö ei ole sellaista, että siitä olisi hyötyä päätöksentekijöille.
4. Nykyisissä ad hoc -raporteissa keskitytään lähinnä yksittäisten ilmiöiden analysointiin sen sijaan, että pyrittäisiin tarkastelemaan näiden ilmiöiden mahdollisia toistuvuuksia tai laajempia vaikutuksia kokonaisjärjestelmälle.
5. Organisaatioissa ei ole riittävästi resursseja (Liikennevirastolla yksi henkilö sivutoimisesti, VR:llä kaksi henkilöä) tuottamaan ad-hoc-analyysijä vakioraportoinnin lisäksi. Tämän vuoksi ajankohtaisista haasteista ei aina saada luotua dataan pohjautuvaa tilannekuvaa ja ymmärrystä.
6. Kaikki analyytikot ovat tällä hetkellä ”yleisanalyytikoita”. On kuitenkin kyseenalaista, että yksittäisellä olisi riittävän kattava rautatieliikenteen eri osa-alueiden osaaminen, että hän pystyisi poimimaan kunkin toiminnon kannalta oleellisen esiin. Toisin sanoen on ilmeistä, ettei tämänhetkinen analyysityö palvele kaikkia osa-alueita parhaalla mahdollisella tavalla.

3.5 Työryhmät

Tässä luvussa esitellään rautatieliikenteen täsmällisyyden parissa toimivia työryhmiä. Työryhmien toimintaa on kartoitettu osallistumalla osan niiden toimintaan aktiivisesti ja havainnoimalla sekä keskustelemalla työryhmiin osallistuvien henkilöiden kanssa heidän näkemyksistään työryhmien toiminnasta. Osan toimintaa puolestaan on kartoitettu sähköpostikyselyin. Työryhmien esittelyt perustuvat siis ihmisten näkemyksiin asioista, eivätkä virallisiin linjauksiin tai kuvauksiin ryhmien tavoitteista ja toiminnasta. Liitteessä 3 on esitetty ne työryhmäkerrat, joihin on tutkimuksen aikana osallistuttu. Liitteen tarkoitus on esittää keskeisimmät esille nousseet ajatukset täsmällisyysanalyysiin liittyen.

3.5.1 VR:n Täsmällisyyden ohjausryhmä ja Ryhti-ryhmä

”Täsmällisyyden ohjausryhmä ja Ryhti-ryhmä ovat korkeimmat täsmällisyyteen ja sen kehittämiseen vaikuttavat ryhmät”

Täsmällisyyden ohjausryhmä on pidemmän tähtäimen jatkuvaa täsmällisyytyötä ohjaava ryhmä, Ryhti-ryhmä puolestaan hoitaa akuutteja, valtakunnallisesti merkittäviä häiriötilanteita. Säännöllisten kokousten lisäksi ryhmät ovat kokoontuneet liikennetilanteen niin vaatiessa. Ryhmien kokoonpanoihin kuuluvat kaikkien keskeisten toimintayksiköiden johdot sekä VR:ltä että Liikennevirastolta.

Vaikka kyseessä on kaksi eri ryhmää, on niiden kokoukset käytännössä varsin usein yhdistetty, koska ryhmiin ovat kuuluneet pitkälti samat henkilöt. Niinpä ryhmät käsitellään tässä yhteydessä yhtenä kokonaisuutena.

Työryhmien rooli ja toiminta

VR:n vetämien ryhmien päätehtävänä on tehdä strategisen tason täsmällisyyttä koskevia päätöksiä ja toisaalta toimia päätöksentekofoorumina laajoissa liikenteen häiriötilanteissa. Ryhmillä on valtuudet tehdä päätöksiä laajoista liikennejärjestelyistä ja toisaalta edistää laajoja kehityskohteita. Ryhmien ansiosta on viime aikoina saatu laajoja poikkeustilanteita hoidettua hallitummin, koska ne pystyvät tekemään nopeasti suuriakin päätöksiä.

Työryhmien vaikuttaminen täsmällisyyteen

Ryhmät pyrkivät edistämään korkean tason täsmällisyyteen liittyviä linjauksia ja täten parantamaan liikennöinnin täsmällisyyttä. Toiminnassaan ryhmät hyödyntävät VR:n ja Liikenneviraston kuukausittain laatimia analyyssejä. Vaikkakin laaja edustusjoukko on ollut ryhmän parhaita ominaisuuksia, on sillä ollut myös käänköpuolensa: lukuisat erilaiset intressit ovat tehneet toiminnasta välillä haastavaa.

3.5.2 VR:n Täsmä-ryhmä

”Täsmä-ryhmä analysoi liikenteen laatu -ryhmistä tulleita ehdotuksia ja pyrkii täten edistämään täsmällisyyttä, joka on ryhmän päätarkoitus”

VR:n Täsmä-ryhmä pyrkii ratkomaan junien operointiin liittyviä täsmällisyshaasteita. Työryhmä kokoontuu kerran kuussa. Työryhmään osallistuu laajasti päällikkötason henkilöitä sekä VR:n että Liikenneviraston keskeisimmiltä toiminta-alueilta.

Työryhmän rooli ja toiminta

Täsmä-ryhmän tehtävänä on analysoida Liikenteen laatu -ryhmistä tulleita ehdotuksia, tehdä niistä päätöksiä ja tarvittaessa edistää asioiden etenemistä. Lisäksi ryhmä analysoi liikennetäsmällisyyttä valtakunnan tasolla. Työryhmä raportoi täsmällisyyden ohjausryhmälle ja Ryhti-ryhmälle. Toiminnassaan työryhmä hyödyntää Liikenneviraston ja VR:n kuukausittaisia täsmällisyysanalyyssejä.

Työryhmän vaikuttaminen täsmällisyyteen

Ryhmän pääasiallinen tavoite on täsmällisyyden parantaminen ja tämän myötä sen keskeisenä roolina onkin ohjata alueellista täsmällisyytyötä. Koska työryhmän toimintaan osallistuu edustajia useilta eri tahoilta, on sen toiminta koettu hyödylliseksi

niin tiedonjakokanavana kuin moninäkökantaisena kehitystoimijana. Työryhmä on jo saanut aikaiseksi pieniä edistysaskelia täsmällisyyttä kehitettäessä. Täsmä-ryhmän toimintaa tulisi kuitenkin tarkastella uudelleen VR:n käynnissä olevan täsmällisyysjohtamisuudistuksen myötä.

3.5.3 Liikenneviraston Infra-täsmäryhmä

”Infra-täsmäryhmä pyrkii vähentämään ratainfraan johtuvia liikennehäiriöitä ja täten kehittämään rautatieliikenteen täsmällisyyttä”

Liikenneviraston vetämä Infra-täsmäryhmä kokoontuu noin joka toinen kuukausi. Ryhmään osallistuu henkilöitä laajalti Liikenneviraston ja VR:n eri toimialueilta. Liikenneviraston organisaation uudistuessa Infra-täsmäryhmä on kuitenkin tällä hetkellä muutosvaiheessa ja sen rakennetta mietitään parhaillaan uusiksi.

Työryhmän rooli ja toiminta

Työryhmän tavoitteena on pyrkiä vähentämään ratainfraan johtuvia liikennehäiriöitä. Infra-täsmäryhmä raportoi toiminnastaan Täsmällisyyden ohjausryhmälle, mutta yhteys on todellisuudessa ollut varsin löyhä. Työryhmä keskittyy rataan lähinnä operatiivisella ja taktisella tasolla, strategisen tason suunnittelun ja toimenpiteiden päätösvallan ollessa toisaalla Liikenneviraston organisaatiossa.

Työryhmän vaikuttaminen täsmällisyyteen

Täsmällisyyteen Infra-täsmäryhmä pyrkii etsimään parantavia ratkaisuja mahdollisimman kustannustehokkaalla tavalla. Täsmällisyyden parantaminen onkin työryhmän pääasia. Ryhmän toiminta on kuitenkin ollut haastavaa, koska työryhmän tavoitteita, vastuita tai valtuuksia ei ole määritelty riittävällä tarkkuudella. Työryhmän suurimmat saavutukset lienevätkin tiedon jakamisessa ja täsmällisyyteen liittyvä tietoisuuden lisäämisessä varsinaisten kehitystoimenpiteiden sijaan.

3.5.4 VR:n Liikenteen laatu -ryhmät (LiLa-ryhmät)

”LiLa-ryhmät ovat paikallisen tason ihmisistä koostuvia asiantuntijaryhmiä, joiden tarkoituksena on nostaa epäkohtia esiin ja esittää niitä ylemmille toimijoille”

VR:n Liikenteen laatu -ryhmät ovat alueellisesti toimivia alueelliseen täsmällisyyteen ja liikenteen laatuun keskittyviä ryhmiä. Ryhmät kokoontuvat kerran kuussa, pääasiallisesti kuukauden toisella viikolla. LiLa-ryhmiä johtavat paikallisten ohjauskeskusten päälliköt. Työryhmissä on henkilöedustus ohjauspalvelusta, tuotannosta, matkustajakaukoliikenteestä (Helsingissä myös lähiliikenteestä), tavaraliikenteestä, vetopalvelusta ja kunnossapidosta mahdollisimman laaja-alaisen arvioinnin ja tiedonkulun varmistamiseksi.

Työryhmien rooli ja toiminta

LiLa-ryhmissä käsitellään matkustajaliikenteen ja tavaraliikenteen operointia niin toteutuneiden kuljetusten kuin tulevien nähtävissä olevien näkymienkin valossa. Työryhmissä arvioidaan erinäisten toimenpiteiden, kuten ratatöiden vaikutusta liikenteeseen ja varaudutaan niihin muun muassa liikennejärjestelyjen sekä sisäisen että ulkoisen tiedottamisen tarpeet huomioon ottaen. Liikenteen laatu -ryhmissä tarkastellaan myös kalustoon liittyviä toimia, kuten kalustokunnossapidon riittävyttä. Lisäksi

LiLa-ryhmissä pyritään tarkastelemaan oman alueen toimintaa kriittisesti ja tekemään kehitysehdotuksia havaittuihin ongelma-kohtiin. Työryhmät raportoivat valtakunnalliselle Täsmä-ryhmälle.

Työryhmien vaikuttaminen täsmällisyyteen

LiLa-ryhmät kokoontuvat täsmällisyyden vuoksi. Koska LiLa-ryhmät ovat ne varsinaiset ”kentällä” sijaitsevat toimijat, on niillä kaikista läheisin kosketusrajapinta paikalliseen toimintaan, toimijoihin ja asiakkaisiin ja täten yleensä myös konkreettisimmat näkemykset kehittämistoimista. Tälläkin kolikolla on kuitenkin kääntöpuolensa. LiLa-ryhmien sijaitessa organisaatiohierarkiassa ruohonjuuritasolla, ovat niiden suorat vaikuttamismahdollisuudet vähäiset.

3.5.5 Liikenneviraston alueelliset infratäsmäryhmät

Liikenneviraston kunnossapitoalueittaiset infratäsmäryhmät toimivat samalla logiikalla kuin VR:n LiLa-ryhmät. Ne käsittelevät oman alueensa infrastruktuuriin liittyviä kysymyksiä ja raportoivat valtakunnalliselle Infra-täsmäryhmälle. Näitä ryhmiä ei tässä tutkimuksessa analysoitu tämän tarkemmin.

3.5.6 VR:n Täsmällisyyden analyysiryhmä (A-ryhmä)

”A-ryhmä tuottaa analyysejä menneen kuun täsmällisyydestä ottaen huomioon kaikki liikennöitsijän toimijaosuudet”

VR:n Täsmällisyyden analyysiryhmä (A-ryhmä) kokoontuu kerran kuukaudessa. A-ryhmään muut osallistuu analyytikoiden lisäksi henkilöitä tavaraliikenteestä, henkilö-kaukoliikenteestä, Helsingin seudun lähiliikenteestä, kaluston kunnossapidosta, Helsingin varikolta sekä henkilöstön ohjauksesta.

Työryhmän rooli ja toiminta

A-ryhmä pyrkii rakentamaan ymmärrystä ja vaikuttamaan koko organisaation läpi. Ryhmän kokoontumisen aikana käydään läpi työryhmän laatimaa PowerPointesitystä, jossa analysoidaan edellisen kuun täsmällisyyteen vaikuttaneita syitä. Analyysi laaditaan kaikista eri junalajeista ja kaikkien edustettujen organisaatioyksiköiden näkökulmista. Lisäksi työryhmä keskustelee tulevista täsmällisyyteen vaikuttavista asioista, kuten kalustokunnossapitoon ja ratatöihin liittyvistä seikoista. Työryhmän suurin haaste on saada kaikki ihmiset ja osa-alueiden analyysit paikan päälle.

Työryhmän vaikuttaminen täsmällisyyteen

A-ryhmän tärkein tavoite on ratkoa täsmällisyyden ongelmia ja esittää mahdollisia toimenpiteitä täsmällisyyden kehittämiseen. A-ryhmä raportoi suoraan divisioonien johdolle ja VR:n ylimmälle johdolle. Kaikki ryhmän esitykset käsitellään VR:n täsmäryhmässä. Koska ryhmä on suhteellisen uusi, eivät kaikki sen toimintatavat ole vielä vakiintuneet. Tavoitteena on, että tuotettavat analyysit saataisiin vakioitua, täsmällisyyteen vaikuttaviin asioihin puututtaisiin merkittävästi ja tehtyjen toimenpiteiden vaikutusta täsmällisyyteen seurattaisiin tarkemmin.

3.5.7 Arvio työryhmien nykytilasta

Täsmällisyyteen liittyviä työryhmiä on olemassa varsin runsaasti, erityisesti VR:llä. Työryhmissä on hyvin integroitu eri toimijoita (VR & Liikennevirasto) samoihin työryhmiin, jonka myötä tiedonkulku ja päätöksenteko helpottuvat. Työryhmien toiminnassa on kuitenkin tunnistettavissa seuraavia analytiikkaan liittyviä haasteita:

1. Työryhmät eivät hyödynnä käytettävissä olevia analyysejä parhaalla mahdollisella tavalla, vaan usein keskustelu ja päätöksenteko perustuvat kokemuksiin ja oletuksiin.
2. Työryhmillä ei useinkaan ole käytössään tarvittavia, yksittäiseen tilanteeseen liittyviä analyysejä.

Lisäksi työn aikana tunnistettiin seuraavat ryhmien toimintaan liittyvät haasteet, jotka eivät suoranaisesti liity analytiikkaan:

3. Useimmat työryhmät keskittyvät sekä strategisten että operatiivisten täsmällisyshaasteiden ratkomiseen. Tällöin toiminnan fokus saattaa kärsiä.
4. Työryhmät (pl. Täsmällisyyden ohjausryhmä ja Ryhti-ryhmä) toimivat vailla nimettyjä taloudellisia resursseja, jolloin niiden vaikuttamismenetelmä on luoda esityksiä ylemmille päätöksentekijöille. Tämä tekee toiminnasta usein hidasta ja kankeaa.
5. Työryhmiin nimetyt henkilöt eivät aina ehdi osallistua ryhmien toimintaan. Tällöin työryhmien kokonaisvaltainen näkemys kaikista eri toiminnoista jää puutteelliseksi ja toisaalta tällöin työryhmien toiminnalla ei ole kaikkien organisaatioyksiköiden tukea.
6. Työryhmien toiminnan seuranta on puutteellista. Työn tuloksia lyhyellä ja pitkällä aikajänteellä ei mitata riittävän selkeästi, jolloin ei ole todellisuudessa tiedossa, mitkä toimenpiteet ovat olleet tehokkaita ja mitkä eivät.

4 Kehittyneen data-analytiikan menetelmien hyödyntäminen täsmällisyysdatan analysoinnissa

Rautateiltä kerätään päivittäin merkittävä määrä junien kulkutietodataa, jota analysoidaan tällä hetkellä pääasiassa perinteisen analytiikan menetelmin. Tässä luvussa tutkitaan esimerkianalyysien avulla, miten kehittyneen data-analytiikan menetelmien avulla voidaan tuottaa datasta nykyistä kattavampaa ja syvällisempää informaatiota. Analysoitavien kohteiden valintaan ja analyysien näkökulmiin ovat vaikuttaneet toimintatutkimuksen aikana esiin nousseet ajatukset.

4.1 Yksittäisen junan toistuvat viiveet

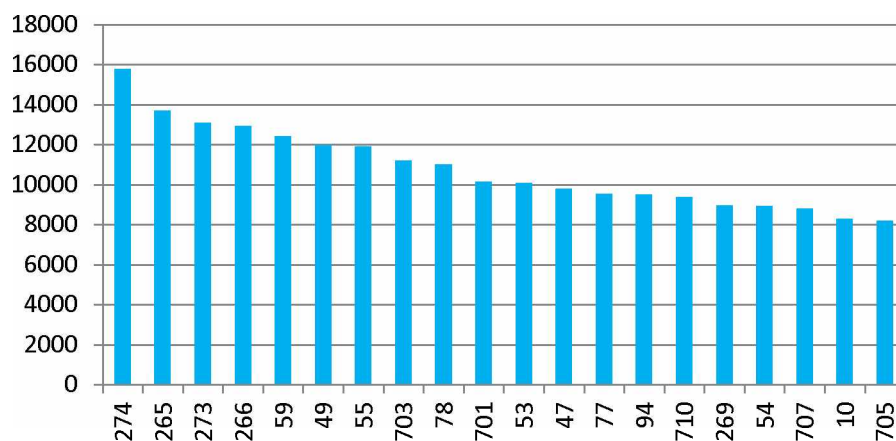
4.1.1 Väite analyysin taustalla

Tietylle junalle eri päivinä koituvat viiveet ovat poikkeavat yleensä merkittävästi toisistaan. Näin ollen seurantatiedon avulla tuskin paljastuu sellaisia systemaattisia ongelmia, jotka eivät jo olisi tiedossa.”

4.1.2 Analyysi

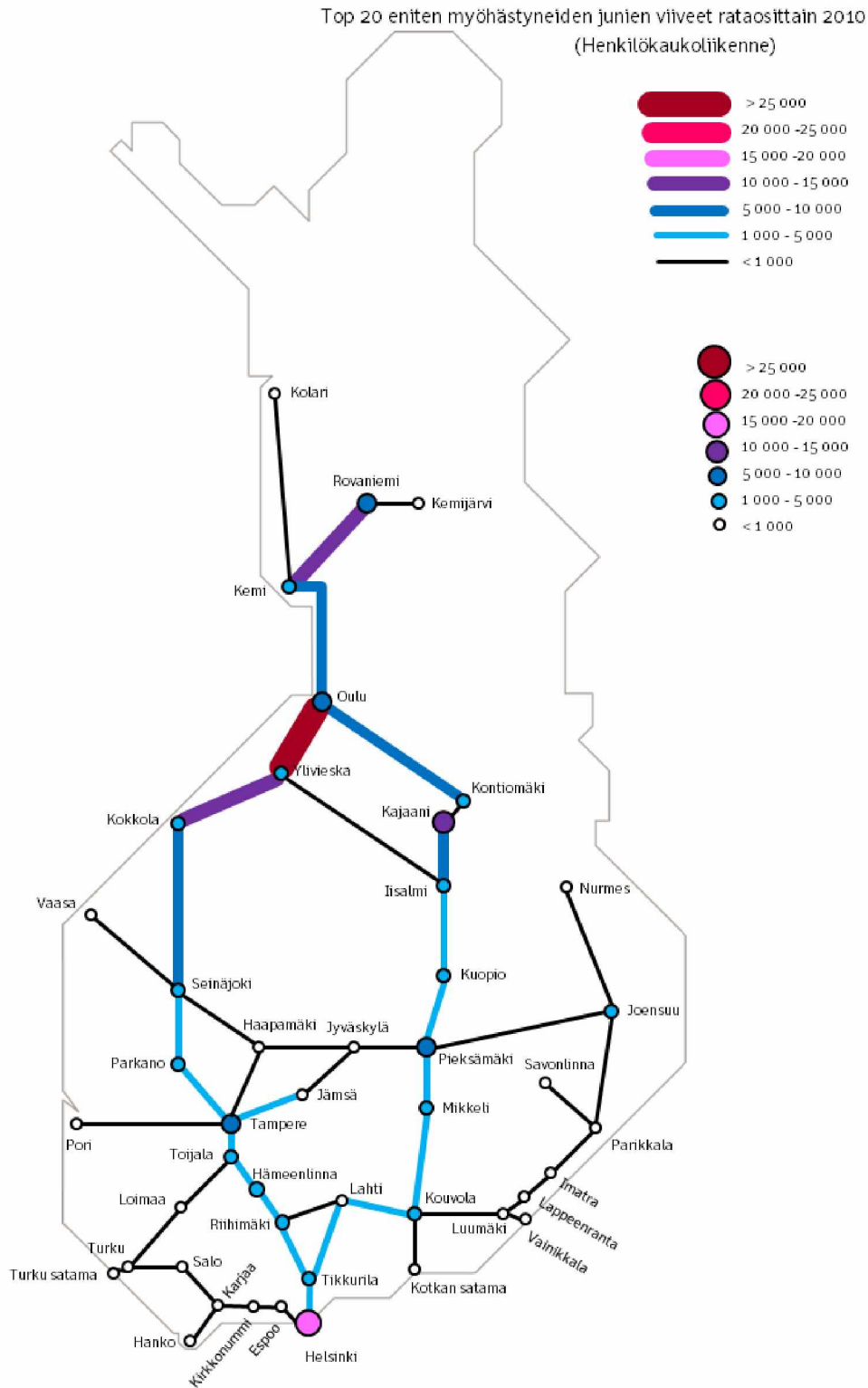
Varsin usein esitetään, että junille koituvat yllättävät ongelmat poikkeavat siinä määrin toisistaan, ettei niitä voida ennakoida eikä siten ottaa huomioon esimerkiksi aikatauluja laadittaessa. Tämän analyysin tavoitteena on osoittaa, että satunnaisilta vaikuttavien ongelmien taustalla voi olla jokin systemaattinen virhe, ja että kehittyneillä data-analytiikkamenetelmillä tämä ilmiö on mahdollista tunnistaa ja analysoida.

Analyysin lähtökohdaksi otettiin junat, jotka myöhästivät vuoden 2010 aikana minuuttimääräisesti eniten (kuva 19). Kuten havaitaan, puolet junista on joko 200-sarjan tai 700-sarjan junia, jotka ovat pohjoisen yöjunia. Näiden junien suuret viiveet johtuvat siitä, että ne matkaavat pitkiä matkoja läpi Suomen, eikä niitä priorisoida yhtä korkealle kuin muita junia. Niinpä niiden tarkastelu tässä yhteydessä ei ole niin kiinnostavaa kuin sellaisten junien, joiden täsmällisyysvaatimus on suurempi. Näistä normaaleista päiväjunista eniten viiveitä keräsi juna 59.



Kuva 19. Vuoden 2010 aikana kaksikymmentä eniten myöhästynyttä junaa.

Kuvassa 21 on kuvattu kahdenkymmenen eniten myöhästyneen junan viiveet rataosittain. Sen perusteella Pohjanmaan rata Kokkolasta Ouluun, rata Rovaniemeltä Kemiin sekä liikennepaikat Helsinki ja Kajaani ovat aiheuttaneet näille junille suurimmat myöhästymiset.



Kuva 20. Kahdenkymmenen eniten myöhästyneen junan viiveet rataosittain, vuosi 2010.

Kuten edellä mainittiin, on juna 59 normaaleista päiväjunista epätasaisesti. Se liikennöi Helsingistä Ouluun, eli useimpien äsken mainittujen ongelmapaikkojen halki. Näiden seikkojen vuoksi analyysin kohteeksi valittiin juna 59 ja sen viiveet vuonna 2010. Tästä eteenpäin tämä analyysi tutkii vain ja ainoastaan junan 59 viiveitä.

Juna 59: perustiedot

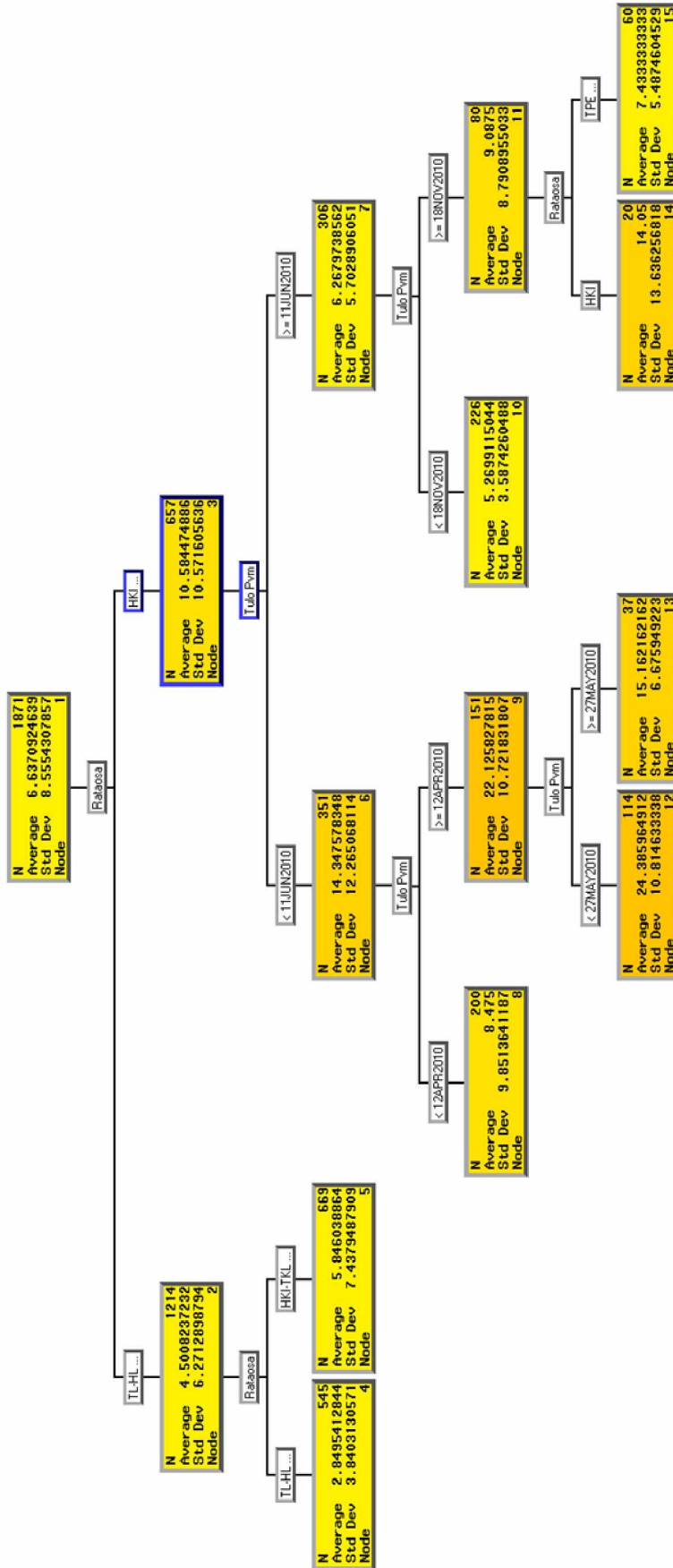
Juna 59 on Pendolino, joka kulkee jokaisena viikonpäivänä Helsingistä Ouluun. Sen aikataulun mukainen lähtöaika Helsingistä on 18:30. Vuoden 2010 aikana juna koki yhteensä 1881 lisämyöhästymistä, joista kertyi yhteensä 12 402 myöhästymisminuuttia. Olettaen, että juna on ajettu jokaisena 365 päivänä vuoden 2010 aikana, yksittäisellä junalla on ollut keskimäärin 34 lisämyöhästymisminuuttia.

Juna 59: luokittelu

Junan 59 analysointi aloitettiin myöhästymisminuuttien luokittelulla luokittelukriteerien ollessa liikennepaikat, rataosat ja päivämäärät. Kuvassa 21 on esitetty tämän luokittelun lopputuloksena syntynyt puumalli neljällä alalehtitasolla. Ensimmäiseksi jakokriteeriksi on valikoitunut liikennepaikka/rataosa. Kuvassa sinisellä korostettu alkio 3 (engl. node) sisältää selvästi suurempia myöhästymisiä saaneet havainnot. Alkioon kuuluvat liikennepaikat Helsinki ja Tampere sekä rataosat ja Kokkola–Ylivieska ja Ylivieska–Oulu. Muut, alkion 2 alle sijoittuvat, liikennepaikat ovat saaneet huomattavasti pienempiä lisämyöhästymisiä.

Alkion 3 alta voidaan selvänä poikkeamana poimia alkio 9. Se sisältää kaikki junan lisämyöhästymiset Helsingissä ja Tampereella sekä Kokkola–Ylivieska- ja Ylivieska–Oulu-väleillä aikavälillä 12. huhtikuuta – 10. kesäkuuta. Tällä kuudenkymmenen päivän ajanjaksolla juna on saanut näillä neljällä paikalla yhteensä 151 lisämyöhästymistä, joiden keskimääräinen suuruus on ollut reilu 22 minuuttia. Näiden myöhästymisten keskihajonta on vajaa 11 minuuttia. Tämä tarkoittaa sitä, että valtaosa (tässä tapauksessa 80 %) lisämyöhästymisistä on 11 ja 33 minuutin välillä. Tarkempi tarkastelu paljastaa, että juna 59 on kärsinyt sekä Kokkola–Ylivieska- että Ylivieska–Oulu-rataosilla lisämyöhästymisen jokaisena päivänä kyseisenä ajanjaksona. Kokkola–Ylivieska-välillä keskimääräinen myöhästymisen on ollut noin 19,9 minuuttia myöhästymisten keskihajonnan ollessa 7,0 minuuttia. Ylivieska–Oulu-välillä vastaavat luvut ovat 25,1 minuuttia (keskimääräinen myöhästymisen) ja 6,9 minuuttia (keskihajonta).

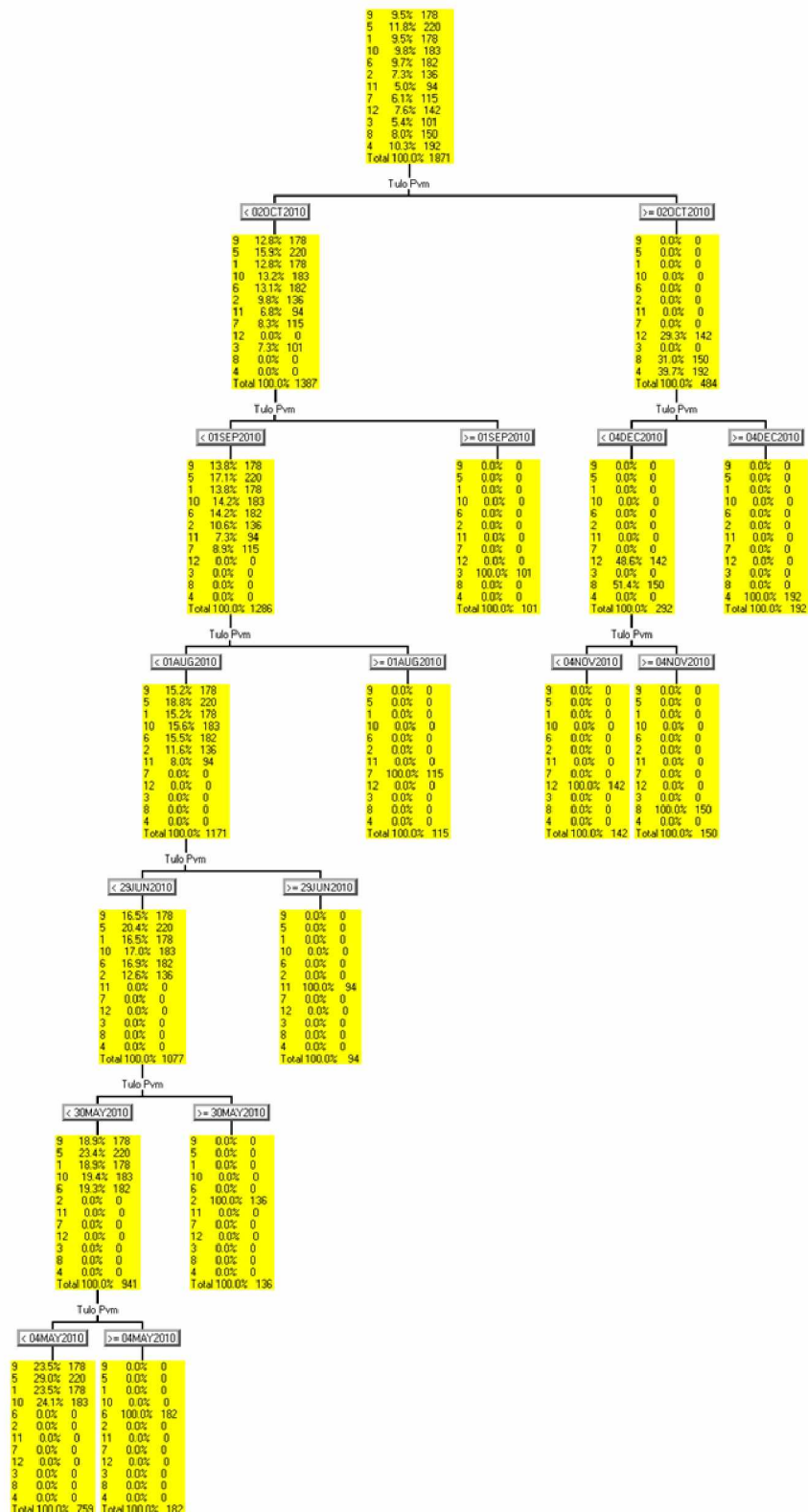
Edellisen perusteella voidaan todeta, että 12. huhtikuuta - 10. kesäkuuta ajetuilla vuoroilla Kokkola–Ylivieska-välin aikataulun mukaisen ajoajan olisi pitänyt olla vähintään 13 minuuttia suurempi, Ylivieska–Oulu-välillä puolestaan 18 minuuttia suurempi. Tällöin ainoastaan 10 %:lla junista olisi ollut liian löysä aikataulu. Loput 90 % olisivat olleet täsmällisiä tai jonkin verran jäljessä aikataulustaan.



Kuva 21. Junan 59 vuoden 2010 viiveiden luokittelu. Kohdemuuttujana myöhästymisminuutit, luokittelukriteereinä liikennepaikat, rataosat ja päivämäärät.

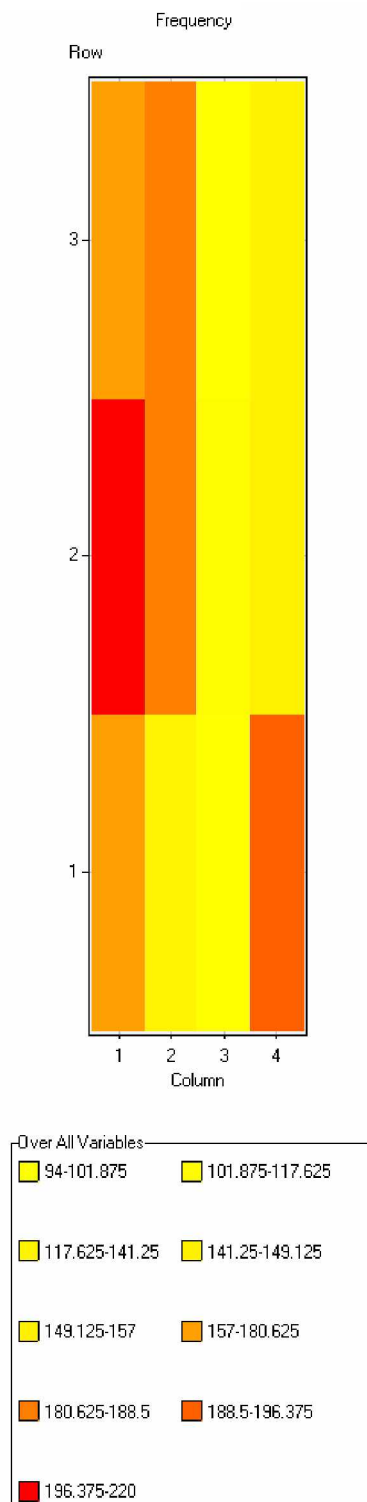
Juna 59: klusterointi kahdellatoista klusterilla

Junan 59 dataa tutkittiin myös klusteroinnin avulla. Kohosen itseorganisoituvien karttojen avulla saatiin informatiivisia tuloksia. Muuttujien ollessa rataosa, myöhästymisminuutit, syyluokka, alue sekä päivämäärä, jakaa klusterointialgoritmi datan kahdeksatoista eri klusteriin päivämäärien mukaan (kuva 22).



Kuva 22. Junan 59 klusteriprofiili puumallina, vuosi 2010.

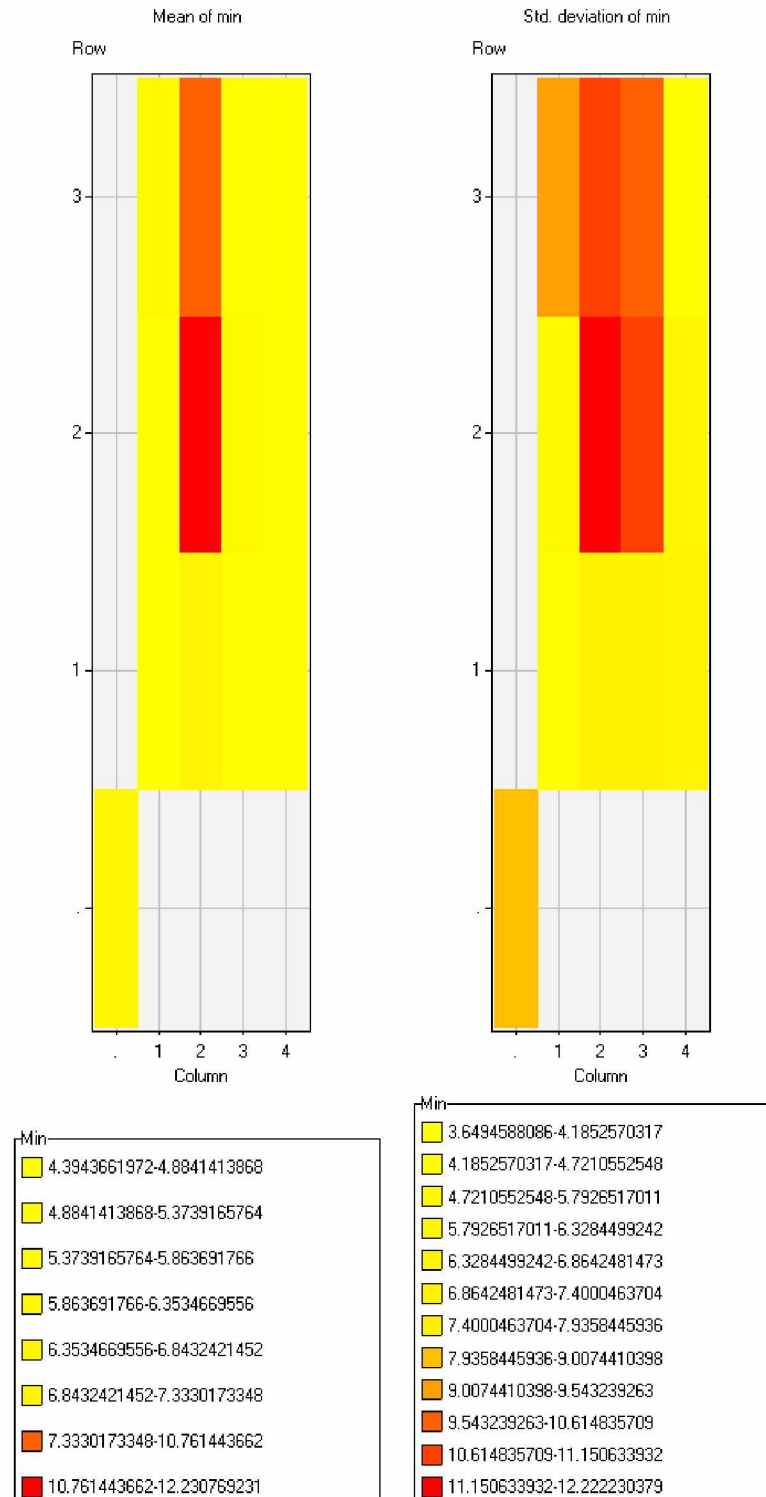
Havaintojen frekvensseittäin jaotellussa klusterikartassa klusteri 2:1 (rivi:sarake) on kaikista suurin (kuva 23).



Kuva 23. Junan 59 viiveiden klusterikartta frekvenssein kahdellatoista klusterilla, vuosi 2010.

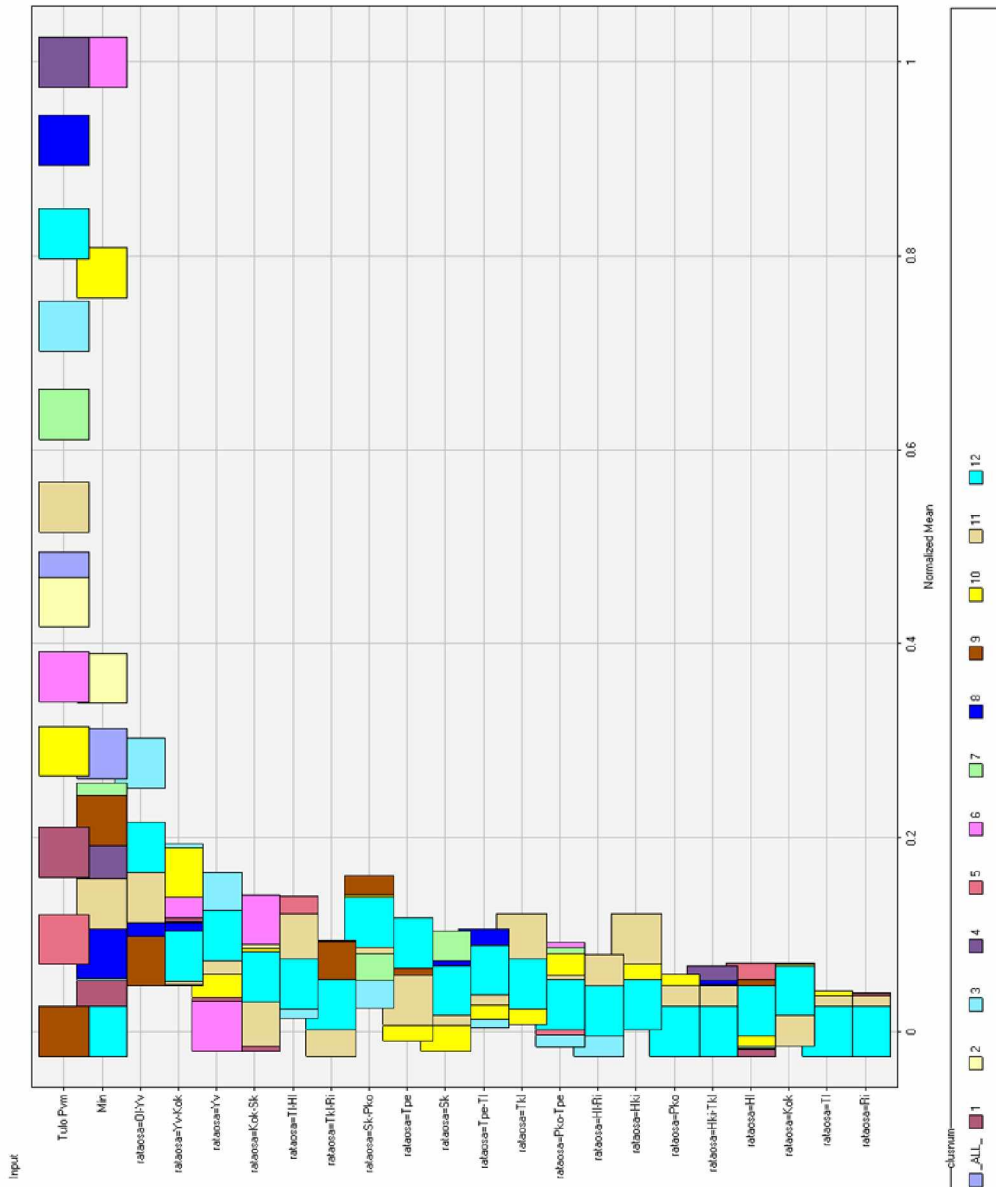
Tässä tilanteessa ei kuitenkaan välttämättä kiinnosta, mikä klustereista on suurin vaan se, mitkä klusterit eroavat massasta ja miksi. Kuvassa 24 onkin esitetty junan 59 viiveiden klusterikartat vuodelta 2010. Vasemmalla on esitetty klusterit viiveminuuttien keskiarvon mukaan ja oikealla viiveminuuttien keskihajonnan mukaan väritetty-

nä. Klusterit 2:2 ja 3:2 sisältävät suurimmat keskiarvot kaikkien myöhästymishavaintojen minuuttimäärille. Tämän vuoksi tutkimme näitä klustereita tarkemmin. Klusteri 2:2 pitää sisällään kaikki havainnot 3. toukokuuta – 30. toukokuuta väliseltä ajalta (kuvassa 25 numerolla 6). Klusteri 3:2 puolestaan pitää sisällään havainnot juuri tätä edeltävältä noin kuukauden ajalta (kuvassa 25 numerolla 6).



Kuva 24. Junan 59 viiveiden klusterikartat kahdellatoista klusterilla, vuosi 2010. Vasen kuva esittää klusterit viiveminuuttien keskiarvon mukaan ja oikea kuva viiveminuuttien keskihajonnan mukaan väritettynä.

Tarkastellessa kaikkien klusterien klusterikuvaajaa (kuva 25), voidaan todeta muun muassa seuraavaa. Kaikki klusterit ovat jakautuneet tasaisesti vuodenaikojen mukaan. Klusterit 6 sekä klusterit 10 keräävät kuitenkin selvästi suurimmat myöhästymisminuutit. Tämä oli havaittavissa jo klusterikartoista, joissa kuvattiin minuuttien keskiarvot eri klustereille. Kaikista rataosista Oulu–Ylivieska-rataosaan liittyviä havaintoja on eniten, kun puolestaan Riihimäkeen liitettäviä myöhästymisiä on vähiten.



Kuva 25. Junan 59 viiveiden klusterikuvaaja kahdentoista klusterin jaolla, vuosi 2010.

Yhteenvedon voidaan todeta, että tässä tilanteessa klusteroinnin suurin lisäarvo on se, että sen tuottamat tilastot ja kuvaajat esittävät varsin hyvin, minä ajanjaksoina ja millä rataosilla myöhästymisiä syntyy. Edellä esitellyn luokittelun pohjalta tämä kuitenkin oli jo tiedossa suurimpien myöhästymisten aiheuttajien suhteen. Näin ollen klusteroinnin suurin lisäarvo tässä analyysissä oli asian havainnollistaminen erilaisien visualisointien muodossa.

Junan 59: sekvenssianalyysi

Junan 59 liikennepaikoittain tapahtuneita viiveitä analysoitiin myös sekvenssianalyysin menetelmin. Kuvassa 26 on esitetty junalle 59 vuonna 2010 sattuneet viiveet syykoodeittain ja liikennepaikoittain ryhmiteltynä. Kuten voidaan havaita, eniten viiveitä syntyi junakohtauksista sekä radan huonosta kunnosta ja ratatöistä johtuen. Myös matkustajapalveluun liittyvät syyt, matkustajaruuhka ja pysähtymisajan ylitys, korosivat viiveiden määrissä mitattuna.

	Count	Item
1	118	L2 YV
2	108	R1 YV-OL
3	98	L2 PKO-SK
4	97	L2 KOK-YV
5	84	R1 KOK-YV
6	80	L2 YV-OL
7	77	K5 HL-TL
8	64	R3 YV-OL
9	63	L2 TPE-PKO
10	47	R1 PKO-SK
11	45	M6 KOK
12	42	M6 TKL
13	39	L2 TL-TPE
14	39	L2 TKL-RI
15	38	L1 TPE
16	32	M6 SK
17	30	M1 TKL
18	29	K5 TL-TPE
19	25	R1 SK-KOK
20	25	L2 SK-KOK
21	24	K7 HKI
22	22	L2 HL-TL
23	21	K5 RI-HL
24	16	M6 TPE
25	16	L3 HKI
26	15	L7 HKI
27	15	L2 TKL
28	15	L2 RI-HL
29	14	R3 SK-KOK
30	14	M3 PKO
31	14	L2 HKI-TKL
32	14	L1 TKL
33	12	J5 PKO-SK
34	12	J5 HL-TL
35	11	M3 SK-KOK
36	10	K4 SK
37	10	I4 PKO-SK
38	9	M3 RI
39	9	M3 HL
40	8	M3 HL-TL
41	8	K5 PKO-SK
42	7	M3 TL
43	7	L4 HKI
44	7	L2 PKO
45	7	L2 HL
46	6	L4 TPE
47	6	L4 KOK-YV
48	6	K7 RI-HL

Kuva 26. Junan 59 myöhästymissyiden määrät liikennepaikoittain vuonna 2010.

Kuvassa 27 on esitetty merkittävimmät kahden viiveen myöhästymisketjut, kuvassa 28 puolestaan kolmen viiveen myöhästymisketjut. Lajittelu on tehty tuen (support) mukaan, jolloin korostuu kokonaistapahtumien määrä. Ketjuja tulkitessa tulee olla varovainen, minkälaisia päätelmiä niistä tekee. Vaikka tietyt tapahtumat esiintyvät peräkkäin, ei niillä välttämättä ole syy-seuraus-suhdetta:

- Tietyillä paikoilla radan huonosta kunnosta johtuen viiveitä on syntynyt peräkkäisillä rataosilla (esim. kuva 27, rivi 1). Vaikka tapahtumat esiintyvät peräkkäin, niillä ei ole syy-seuraussuhdetta.

- Sekundääriset viiveet eivät aiheuta primäärisiä viiveitä (ainoastaan päinvas-
toin), vaikka tiettyä sekundääristä viivettä seuraisikin usein jokin tietty pri-
määrinen viive (esim. kuva 27, rivi 5).

Toisaalta ketjuista voidaan päätellä joitakin *mahdollisia* syy-seuraus-suhteita (esim. kuva 27, rivit 2 ja 8). Näissäkin yhteyksissä kyseessä voi kuitenkin olla puhdas sattuma.

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
1	2	19.89	84.52	71	R1 KOK-YV ==> R1 YV-OL
2	2	8.68	31.96	31	L2 KOK-YV ==> L2 YV-OL
3	2	8.68	31.63	31	L2 PKO-SK ==> L2 YV-OL
4	2	8.40	30.61	30	L2 PKO-SK ==> L2 YV
5	2	8.12	24.58	29	L2 YV ==> R1 YV-OL
6	2	7.84	23.73	28	L2 YV ==> R3 YV-OL
7	2	7.84	44.44	28	L2 TPE-PKO ==> R1 KOK-YV
8	2	7.56	35.06	27	K5 HL-TL ==> L2 PKO-SK
9	2	7.28	26.80	26	L2 KOK-YV ==> R3 YV-OL
10	2	7.28	41.27	26	L2 TPE-PKO ==> R1 YV-OL
11	2	7.00	25.51	25	L2 PKO-SK ==> R1 YV-OL
12	2	7.00	25.51	25	L2 PKO-SK ==> R1 KOK-YV
13	2	7.00	53.19	25	R1 PKO-SK ==> R1 YV-OL
14	2	6.72	31.17	24	K5 HL-TL ==> R1 YV-OL
15	2	6.72	31.17	24	K5 HL-TL ==> L2 YV-OL
16	2	6.72	31.17	24	K5 HL-TL ==> L2 YV
17	2	6.72	38.10	24	L2 TPE-PKO ==> L2 PKO-SK
18	2	5.88	84.00	21	R1 SK-KOK ==> R1 KOK-YV
19	2	5.88	100.00	21	K5 RI-HL ==> K5 HL-TL
20	2	5.88	33.33	21	L2 TPE-PKO ==> L2 YV
21	2	5.60	20.41	20	L2 PKO-SK ==> L2 KOK-YV
22	2	5.60	80.00	20	R1 SK-KOK ==> R1 YV-OL
23	2	5.32	40.43	19	R1 PKO-SK ==> R1 KOK-YV
24	2	5.32	16.10	19	L2 YV ==> L2 YV-OL
25	2	5.32	24.68	19	K5 HL-TL ==> L2 KOK-YV
26	2	5.32	90.48	19	L2 YV & R1 KOK-YV ==> R1 YV-OL
27	2	5.32	24.68	19	K5 HL-TL ==> K5 TL-TPE
28	2	5.32	48.72	19	L2 TL-TPE ==> L2 PKO-SK
29	2	5.04	46.15	18	L2 TKL-RI ==> K5 HL-TL
30	2	5.04	42.86	18	M6 TKL ==> L2 KOK-YV
31	2	5.04	18.37	18	L2 PKO-SK ==> M6 KOK
32	2	4.76	77.27	17	L2 HL-TL ==> L2 TL-TPE
33	2	4.48	25.40	16	L2 TPE-PKO ==> L2 YV-OL
34	2	4.48	35.56	16	M6 KOK ==> L2 YV
35	2	4.48	20.78	16	K5 HL-TL ==> L2 TPE-PKO
36	2	4.48	20.78	16	K5 HL-TL ==> R1 KOK-YV
37	2	4.20	19.48	15	K5 HL-TL ==> R1 PKO-SK
38	2	4.20	39.47	15	L1 TPE ==> L2 PKO-SK
39	2	4.20	60.00	15	L2 SK-KOK ==> R1 YV-OL
40	2	4.20	19.48	15	K5 HL-TL ==> M6 KOK
41	2	4.20	50.00	15	M1 TKL ==> L2 PKO-SK
42	2	4.20	15.46	15	L2 KOK-YV ==> R1 YV-OL
43	2	3.92	36.84	14	L1 TPE ==> L2 YV
44	2	3.92	63.64	14	L2 HL-TL ==> L2 PKO-SK
45	2	3.92	22.22	14	L2 TPE-PKO ==> R1 PKO-SK
46	2	3.92	35.90	14	L2 TKL-RI ==> L2 PKO-SK
47	2	3.92	35.90	14	L2 TL-TPE ==> L2 KOK-YV
48	2	3.92	22.22	14	L2 TPE-PKO ==> M6 KOK
49	2	3.92	35.90	14	L2 TL-TPE ==> R1 YV-OL
50	2	3.92	56.00	14	L2 SK-KOK ==> R1 KOK-YV
51	2	3.92	31.11	14	M6 KOK ==> L2 KOK-YV
52	2	3.92	29.79	14	R1 PKO-SK ==> L2 YV

Kuva 27. Junan 59 viivesekvenssit kahden viiveen ketjuina tuen mukaan lajiteltu-
na.

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
134	3	6.72	85.71	24	L2 TPE-PKD ==> R1 KOK-YV ==> R1 YV-OL
135	3	5.60	95.24	20	R1 SK-KOK ==> R1 KOK-YV ==> R1 YV-OL
136	3	5.60	80.00	20	L2 PKO-SK ==> R1 KOK-YV ==> R1 YV-OL
137	3	4.76	89.47	17	R1 PKO-SK ==> R1 KOK-YV ==> R1 YV-OL
138	3	3.64	100.00	13	L2 TPE-PKD ==> R1 SK-KOK ==> R1 KOK-YV
139	3	3.64	81.25	13	K5 HL-TL ==> R1 KOK-YV ==> R1 YV-OL
140	3	3.64	92.86	13	L2 SK-KOK ==> R1 KOK-YV ==> R1 YV-OL
141	3	3.36	85.71	12	L2 TPE-PKD ==> R1 PKO-SK ==> R1 KOK-YV
142	3	3.36	92.31	12	L2 TPE-PKD ==> R1 SK-KOK ==> R1 YV-OL
143	3	3.36	70.59	12	L2 HL-TL ==> L2 TL-TPE ==> L2 PKO-SK
144	3	3.36	92.31	12	L7 HKI ==> R1 KOK-YV ==> R1 YV-OL
145	3	3.36	60.00	12	L2 PKO-SK ==> L2 KOK-YV ==> L2 YV-OL
146	3	3.36	85.71	12	L2 TPE-PKD ==> R1 PKO-SK ==> R1 YV-OL
147	3	3.08	45.83	11	L2 TPE-PKD ==> L2 PKO-SK ==> L2 YV-OL
148	3	3.08	40.74	11	K5 HL-TL ==> L2 PKO-SK ==> L2 YV-OL
149	3	2.52	60.00	9	K5 HL-TL ==> R1 PKO-SK ==> R1 YV-OL
150	3	2.52	50.00	9	L2 PKO-SK ==> M6 KOK ==> L2 YV-OL
151	3	2.52	33.33	9	K5 HL-TL ==> L2 PKO-SK ==> M6 KOK
152	3	2.52	52.94	9	L2 HL-TL ==> L2 TL-TPE ==> L2 YV-OL
153	3	2.52	47.37	9	L2 TL-TPE ==> L2 PKO-SK ==> L2 YV-OL
154	3	2.52	100.00	9	L2 RI-HL ==> L2 HL-TL ==> L2 TL-TPE
155	3	2.24	42.11	8	K5 HL-TL ==> L2 KOK-YV ==> L2 YV-OL
156	3	2.24	38.10	8	L2 TPE-PKD ==> L2 YV ==> R1 YV-OL
157	3	2.24	100.00	8	L2 PKO-SK ==> R1 SK-KOK ==> R1 KOK-YV
158	3	2.24	29.63	8	K5 HL-TL ==> L2 PKO-SK ==> L2 YV
159	3	2.24	33.33	8	L2 TPE-PKD ==> L2 PKO-SK ==> L2 YV
160	3	2.24	57.14	8	L2 HL-TL ==> L2 PKO-SK ==> L2 YV-OL
161	3	2.24	33.33	8	K5 HL-TL ==> L2 YV ==> R1 YV-OL
162	3	2.24	100.00	8	L2 TKL-RI ==> L2 HL-TL ==> L2 TL-TPE
163	3	2.24	66.67	8	M1 TKL ==> K5 HL-TL ==> L2 PKO-SK
164	3	1.96	25.93	7	K5 HL-TL ==> L2 PKO-SK ==> L2 KOK-YV
165	3	1.96	50.00	7	R1 PKO-SK ==> L2 YV ==> R1 YV-OL
166	3	1.96	50.00	7	L2 TPE-PKD ==> M6 KOK ==> L2 YV
167	3	1.96	87.50	7	L2 TL-TPE ==> R1 KOK-YV ==> R1 YV-OL
168	3	1.96	38.89	7	L2 TKL-RI ==> K5 HL-TL ==> L2 YV
169	3	1.96	38.89	7	L2 TKL-RI ==> K5 HL-TL ==> L2 PKO-SK
170	3	1.96	46.67	7	M1 TKL ==> L2 PKO-SK ==> L2 KOK-YV
171	3	1.96	77.78	7	M1 TKL ==> R1 KOK-YV ==> R1 YV-OL
172	3	1.96	53.85	7	M6 TKL ==> L2 PKO-SK ==> L2 YV-OL
173	3	1.96	38.89	7	M6 TKL ==> L2 KOK-YV ==> L2 YV-OL
174	3	1.96	87.50	7	L2 PKO-SK ==> R1 SK-KOK ==> R1 YV-OL
175	3	1.96	41.18	7	L2 HL-TL ==> L2 TL-TPE ==> L2 KOK-YV
176	3	1.96	33.33	7	K5 RI-HL ==> K5 HL-TL ==> L2 PKO-SK
177	3	1.96	33.33	7	K5 RI-HL ==> K5 HL-TL ==> K5 TL-TPE
178	3	1.96	77.78	7	M6 TKL ==> R1 KOK-YV ==> R1 YV-OL
179	3	1.96	77.78	7	L2 RI-HL ==> L2 TL-TPE ==> L2 PKO-SK
180	3	1.96	77.78	7	L2 RI-HL ==> L2 HL-TL ==> L2 PKO-SK
181	3	1.96	50.00	7	L2 TKL-RI ==> L2 PKO-SK ==> L2 YV-OL
182	3	1.96	100.00	7	L2 HKI-TKL ==> L2 KOK-YV ==> L2 YV-OL

Kuva 28. Junan 59 viivesekvenssit kolmen viiveen ketjuina tuen mukaan lajiteltuina.

Kun lajitellaan viivesekvenssit luottamusasteen mukaan, antaa analyysi huomattavasti mielenkiintoisempia tuloksia. Kuvassa 29 viivesekvenssit on lajiteltu luottamusasteen mukaan. Sen perusteella voidaan todeta mielenkiintoisia havaintoja, jotka eivät ole selitettävissä Pohjanmaan radan huonolla kunnolla.

Seitsemännen havainnon mukaan aina (luottamusaste 100 %), kun juna 59 on viivästynyt Riihimäki–Hämeenlinna-välillä junakohtauksen (L2) vuoksi, on se viivästynyt myös Hämeenlinna–Toijala- sekä Toijala–Tampere-väleillä junakohtauksen vuoksi. Näitä kolmen viiveen ketjuja on tapahtunut yhdeksän kertaa vuodessa. Edelleen voidaan havaita (rivi 34), että kun juna 59 on saanut junakohtauksesta johtuvan viiveen Hämeenlinna–Toijala-välillä, on se yleensä (luottamusaste 77 %) saanut L2-viiveen

myös Tampere–Toijala-välillä. Näissä tapahtumissa luottamusaste on erittäin korkea, joten kyseessä on mitä ilmeisimmin systemaattinen ongelma aikataulurakenteessa.

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
1	3	3.64	100.00	13	L2 TPE-PKO ==> R1 SK-KOK ==> R1 KOK-YV
2	4	3.36	100.00	12	L2 TPE-PKO ==> R1 PKO-SK ==> R1 KOK-YV ==> R1 YV-OL
3	3	1.96	100.00	7	L2 HKI-TKL ==> L2 KOK-YV ==> L2 YV-OL
4	3	2.24	100.00	8	L2 TKL-RI ==> L2 HL-TL ==> L2 TL-TPE
5	3	2.24	100.00	8	L2 PKO-SK ==> R1 SK-KOK ==> R1 KOK-YV
6	2	5.88	100.00	21	K5 RI-HL ==> K5 HL-TL
7	3	2.52	100.00	9	L2 RI-HL ==> L2 HL-TL ==> L2 TL-TPE
8	3	5.60	95.24	20	R1 SK-KOK ==> R1 KOK-YV ==> R1 YV-OL
9	3	3.64	92.86	13	L2 SK-KOK ==> R1 KOK-YV ==> R1 YV-OL
10	3	3.36	92.31	12	L7 HKI ==> R1 KOK-YV ==> R1 YV-OL
11	3	3.36	92.31	12	L2 TPE-PKO ==> R1 SK-KOK ==> R1 YV-OL
12	4	3.36	92.31	12	L2 TPE-PKO ==> R1 SK-KOK ==> R1 KOK-YV ==> R1 YV-OL
13	2	5.32	90.48	19	L2 YV & R1 KOK-YV ==> R1 YV-OL
14	3	4.76	89.47	17	R1 PKO-SK ==> R1 KOK-YV ==> R1 YV-OL
15	3	1.96	87.50	7	L2 TL-TPE ==> R1 KOK-YV ==> R1 YV-OL
16	3	1.96	87.50	7	L2 PKO-SK ==> R1 SK-KOK ==> R1 YV-OL
17	4	1.96	87.50	7	L2 PKO-SK ==> R1 SK-KOK ==> R1 KOK-YV ==> R1 YV-OL
18	2	3.64	86.67	13	L7 HKI ==> R1 YV-OL
19	2	3.64	86.67	13	L7 HKI ==> R1 KOK-YV
20	3	3.36	85.71	12	L2 TPE-PKO ==> R1 PKO-SK ==> R1 YV-OL
21	3	6.72	85.71	24	L2 TPE-PKO ==> R1 KOK-YV ==> R1 YV-OL
22	3	3.36	85.71	12	L2 TPE-PKO ==> R1 PKO-SK ==> R1 KOK-YV
23	2	19.89	84.52	71	R1 KOK-YV ==> R1 YV-OL
24	2	5.88	84.00	21	R1 SK-KOK ==> R1 KOK-YV
25	3	3.64	81.25	13	K5 HL-TL ==> R1 KOK-YV ==> R1 YV-OL
26	2	5.60	80.00	20	R1 SK-KOK ==> R1 YV-OL
27	3	5.60	80.00	20	L2 PKO-SK ==> R1 KOK-YV ==> R1 YV-OL
28	4	1.96	77.78	7	L2 RI-HL ==> L2 HL-TL ==> L2 TL-TPE ==> L2 PKO-SK
29	3	1.96	77.78	7	L2 RI-HL ==> L2 TL-TPE ==> L2 PKO-SK
30	2	1.96	77.78	7	M3 HL ==> M3 HL-TL
31	3	1.96	77.78	7	L2 RI-HL ==> L2 HL-TL ==> L2 PKO-SK
32	3	1.96	77.78	7	M6 TKL ==> R1 KOK-YV ==> R1 YV-OL
33	3	1.96	77.78	7	M1 TKL ==> R1 KOK-YV ==> R1 YV-OL
34	2	4.76	77.27	17	L2 HL-TL ==> L2 TL-TPE
35	3	3.36	70.59	12	L2 HL-TL ==> L2 TL-TPE ==> L2 PKO-SK
36	3	2.24	66.67	8	M1 TKL ==> K5 HL-TL ==> L2 PKO-SK
37	4	2.24	66.67	8	L2 HL-TL ==> L2 TL-TPE ==> L2 PKO-SK ==> L2 YV-OL
38	2	2.52	64.29	9	R3 SK-KOK ==> R3 YV-OL
39	2	2.52	64.29	9	L2 HKI-TKL ==> L2 YV-OL
40	2	3.92	63.64	14	L2 HL-TL ==> L2 PKO-SK
41	2	2.52	60.00	9	L2 RI-HL ==> L2 HL-TL
42	3	2.52	60.00	9	K5 HL-TL ==> R1 PKO-SK ==> R1 YV-OL
43	2	4.20	60.00	15	L2 SK-KOK ==> R1 YV-OL
44	3	3.36	60.00	12	L2 PKO-SK ==> L2 KOK-YV ==> L2 YV-OL
45	2	2.52	60.00	9	L2 RI-HL ==> L2 TL-TPE
46	3	2.24	57.14	8	L2 HL-TL ==> L2 PKO-SK ==> L2 YV-OL
47	2	2.52	56.25	9	M6 TPE ==> L2 YV-OL
48	2	3.92	56.00	14	L2 SK-KOK ==> R1 KOK-YV
49	3	1.96	53.85	7	M6 TKL ==> L2 PKO-SK ==> L2 YV-OL
50	2	2.24	53.33	8	L2 RI-HL ==> L2 YV-OL
51	2	2.24	53.33	8	L2 RI-HL ==> L2 PKO-SK
52	2	7.00	53.19	25	R1 PKO-SK ==> R1 YV-OL
53	2	2.52	53.04	9	L2 HL-TL ==> L2 TL-TPE ==> L2 YV-OL

Kuva 29. Junan 59 viivesekvenssit kahden, kolmen ja neljän viiveen ketjuina luottamusasteen mukaan lajiteltuna.

Suodattaessa raakadatasta *syyn lisätieto* -kentästä junanumerot, saadaan dataan uusi sarake, jossa näkyy kunkin havainnon kohdalla se junanumero, joka on ollut osallisena kyseisessä viiveessä. Tämä liikenteenohjaajien syöttämä tieto ei ole kovin kattavaa eikä luotettavaa, koska kirjauksia tehdään puutteellisesti ja niissä esiintyy virheitä. Se kuitenkin toimii hyvin suuntaa-antavana tietona siitä, mitkä junat aiheuttavat toisil-

leen viiveitä. Lisäksi sitä hyödyntämällä voidaan osoittaa, mitä datalla voitaisiin tehdä, mikäli se olisi kattavampaa ja luotettavampaa.

Kuvassa 30 on esitetty sekvenssianalyysi siitä, mitä viivesyitä milläkin liikennepaikoilla ja millekin junille juna 59 on aiheuttanut huhti-toukokuussa 2010. Tänä 61 päivän ajanjaksolla juna 59 on esimerkiksi aiheuttanut 36 yhteysliikenteen odotusta Seinäjoelta lähteville junille (rivi 1). Tarkempi analyysi paljastaa, että:

- 20 kertaa viivästyneenä on ollut juna 407. Minuuttimääräisesti tämä tarkoittaa 308 minuuttia. Keskimääräinen viive on ollut 15,4 minuuttia.
- 16 kertaa viivästyneenä on ollut juna 453. Minuuttimääräisesti tämä tarkoittaa 256 minuuttia. Keskimääräinen viive on ollut 15,4 minuuttia.

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
1	2	1.75	22.36	36	59 ==> L1 Sk
2	2	0.97	12.42	20	59 ==> 407
3	2	0.83	10.56	17	59 ==> L1 Tpe
4	2	0.83	10.56	17	59 ==> 453
5	2	0.78	9.94	16	59 ==> L2 Yv
6	2	0.49	6.21	10	59 ==> 927
7	2	0.34	4.35	7	59 ==> 266
8	2	0.24	3.11	5	59 ==> L2 Yv-Kok
9	2	0.19	2.48	4	59 ==> 5014
10	2	0.19	2.48	4	59 ==> L2 Kok-Sk
11	2	0.15	1.86	3	59 ==> L2 Pko-Tpe
12	2	0.15	1.86	3	59 ==> 56
13	2	0.15	1.86	3	59 ==> L2 OI
14	2	0.15	1.86	3	59 ==> 5012
15	2	0.15	1.86	3	59 ==> L2 OI:Yv
16	2	0.10	1.24	2	59 ==> L2 Sk-Pko
17	2	0.10	1.24	2	59 ==> L2 Pko
18	2	0.10	1.24	2	59 ==> 5016
19	2	0.10	1.24	2	59 ==> 58
20	3	0.97	55.56	20	59 ==> L1 Sk ==> 407
21	3	0.78	44.44	16	59 ==> L1 Sk ==> 453
22	3	0.49	58.82	10	59 ==> L1 Tpe ==> 927
23	3	0.29	37.50	6	59 ==> L2 Yv ==> 266
24	3	0.19	25.00	4	59 ==> L2 Yv ==> 5014
25	3	0.15	75.00	3	59 ==> L2 Kok-Sk ==> 5012
26	3	0.15	100.00	3	59 ==> L2 Pko-Tpe ==> 56
27	3	0.10	40.00	2	59 ==> L2 Yv-Kok ==> 5016

Kuva 30. Junan 59 aiheuttamat viiveet muille junille syykoodi-liikennepaikka-yhdistelmittäin ja junanumeroittain.

Aikaisemmin todettiin, että juna 59 olisi tarvinnut merkittäviä löysennyksiä aikatauluihin Kokkola–Ylivieska- sekä Ylivieska–Oulu-väleille huhti-toukokuun aikana. Sekvenssianalyysin perusteella se kuitenkin aiheutti merkittäviä viiveitä muille junille jo ennen Seinäjoelta pohjoiseen kulkevaa Pohjanmaan rataosaa. Näin ollen ei yksiselitteisesti voida rajata junan tarvitsevan lisää ajoaikaa vain Pohjanmaan rataosille.

Kuten voidaan havaita, sekvenssianalyysistä on merkittävää hyötyä analysoitaessa yksittäisten junien viiveitä. Junan 59 tapauksessa esiin nousseet viiveketjut olivat toki melko triviaaleja ja Pohjanmaan radan huonon kunnon aiheuttamien ongelmien sävyttämiä. Kun mukaan otettiin *syyn lisätieto* -kentän tarjoama tieto siitä, mikä juna on liikenteenohjaajien mukaan ollut kunkin viiveen taustalla, datasta saatiin jo huomattavasti enemmän irti. Analyysin perusteella kyettiin esimerkiksi toteamaan, että junan 59 ainoat ongelmat eivät olleet Pohjanmaan heikkokuntoisilla rataosuksilla.

4.1.3 Päätelmät

Kehittyneen data-analytiikan menetelmät tarjoavat selviä hyötyjä yksittäistä juna-analysoitaessa. Yleisellä tasolla eduiksi havaittiin ennen kaikkea nopeus, monipuolisuus ja havainnollisuus.

Luokittelumenetelmillä voidaan tarkastella esimerkiksi, mitkä tekijät ovat merkittävimpiä viiveiden aiheuttajia. Tässä tapauksessa osoittautui, että ajanjakso ja rataosa selittivät parhaiten tapahtuneita viiveitä: yksi ajanjakso tietyillä rataosilla nousi muita merkittävämmäksi. Tässä tapauksessa kyseessä oli melko itsestään selvä ilmiö, Pohjanmaan radan routavauriot, mutta mikään ei estä käyttämästä menetelmää myös piilevämpien ilmiöiden tunnistamiseen.

Klusterointi antaa mahdollisuuden datan automaattiseen jäsentämiseen ilman ennakkotietoja. Tehdyssä klusteroinnissa merkittävimäksi selittäväksi tekijäksi nousi ajanjakso. Tämän myötä saatiin selville muun muassa eri ajanjaksojen keskimääräisten viiveminuuttien suuruudet ja keskihajonnat. Tekemällä klusterointia suuremmilla ja pienemmällä klusterimäärillä sekä integroimalla dataan esimerkiksi kalustotietoja, toisi menetelmä esille huomattavasti enemmän selittäviä eroavaisuuksia.

Sekvenssianalyysin potentiaali osoittautui erittäin suureksi, sillä rautatiejärjestelmä sisältää paljon sisäisiä riippuvuuksia ja tapahtumaketjuja – sekvenssejä. Näiden ketjujen tunnistamiseen ja analysointiin sekvenssianalyysi on paras mahdollinen työkalu. Sen avulla päästään käsiksi esimerkiksi aikataulurakenteen ongelmakohtiin. Tutkimuksessa esiin nousseet sekvenssit olivat pääasiassa melko triviaaleja, mutta jo niiden perusteella kävi ilmi, että menetelmän avulla esimerkiksi ongelmien esiintymistiheyksien vertailu ja siten priorisointi on helppoa. Mikään ei myöskään jatkossa estä monimutkaisempien ja piilevämpien sekvenssien tarkastelua. Sekvenssianalyysin täysimääräinen hyödyntäminen kuitenkin edellyttäisi, että kulkutietodata sisältäisi myös kellonajat, että dataa kerättäisiin nykyistä tiheämmin ja että *syyn lisätieto* -kentän data olisi kattavampaa ja luotettavampaa.

Oikean datan valinnan jälkeen itse analyysiprosessi oli varsin tehokas eikä vienyt juurikaan aikaa. Näin aikaa jäi itse analyysien tulkinneille. Tämän perusteella yksittäiseen junaan liittyvä analyysiprosessi kannattaakin systematisoida. Näin analyysien laatiminen yksittäisistä junista on nopeaa ja helppoa.

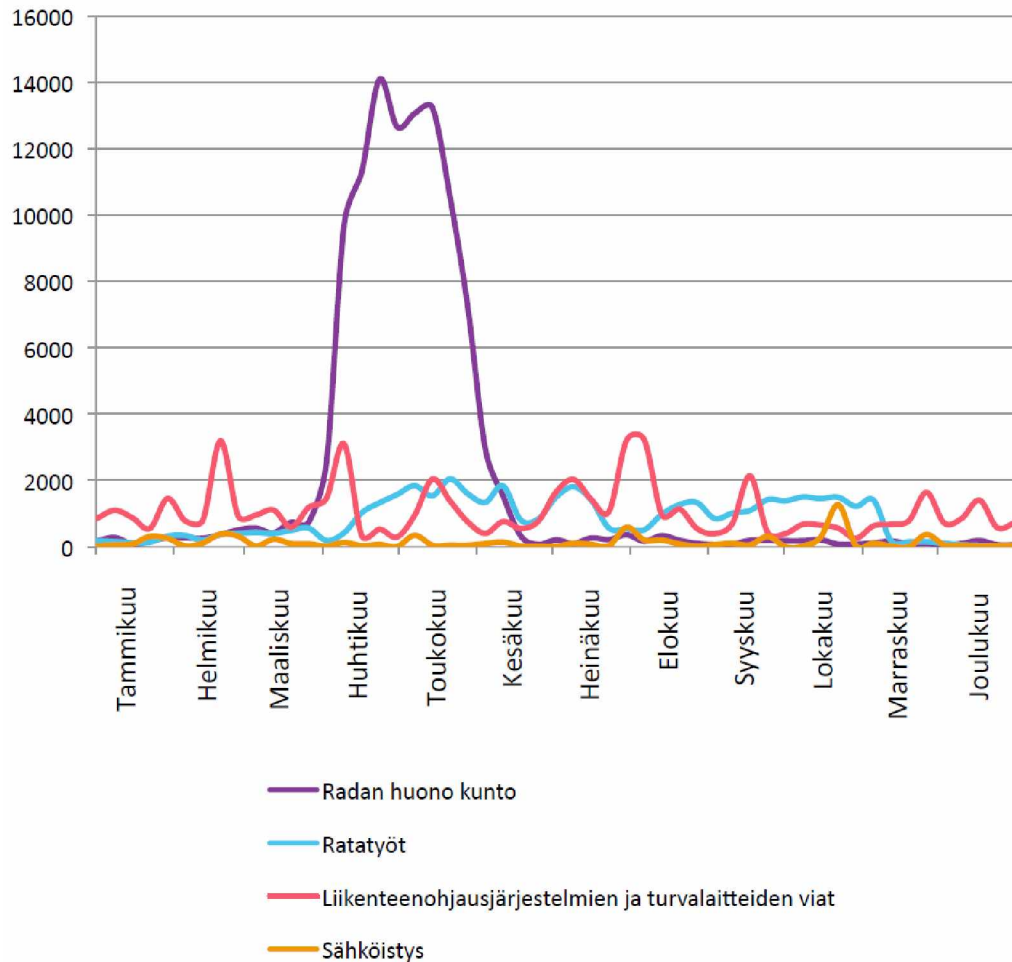
4.2 Yksittäisen ilmiön vaikutus koko maahan

4.2.1 Väite analyysin taustalla

Pohjanmaan radan heikon kunnan vuoksi junat myöhästelevät merkittävästi kevään routa-aikana ja tämän vuoksi myös muu valtakunnallinen liikenne myöhästelee.

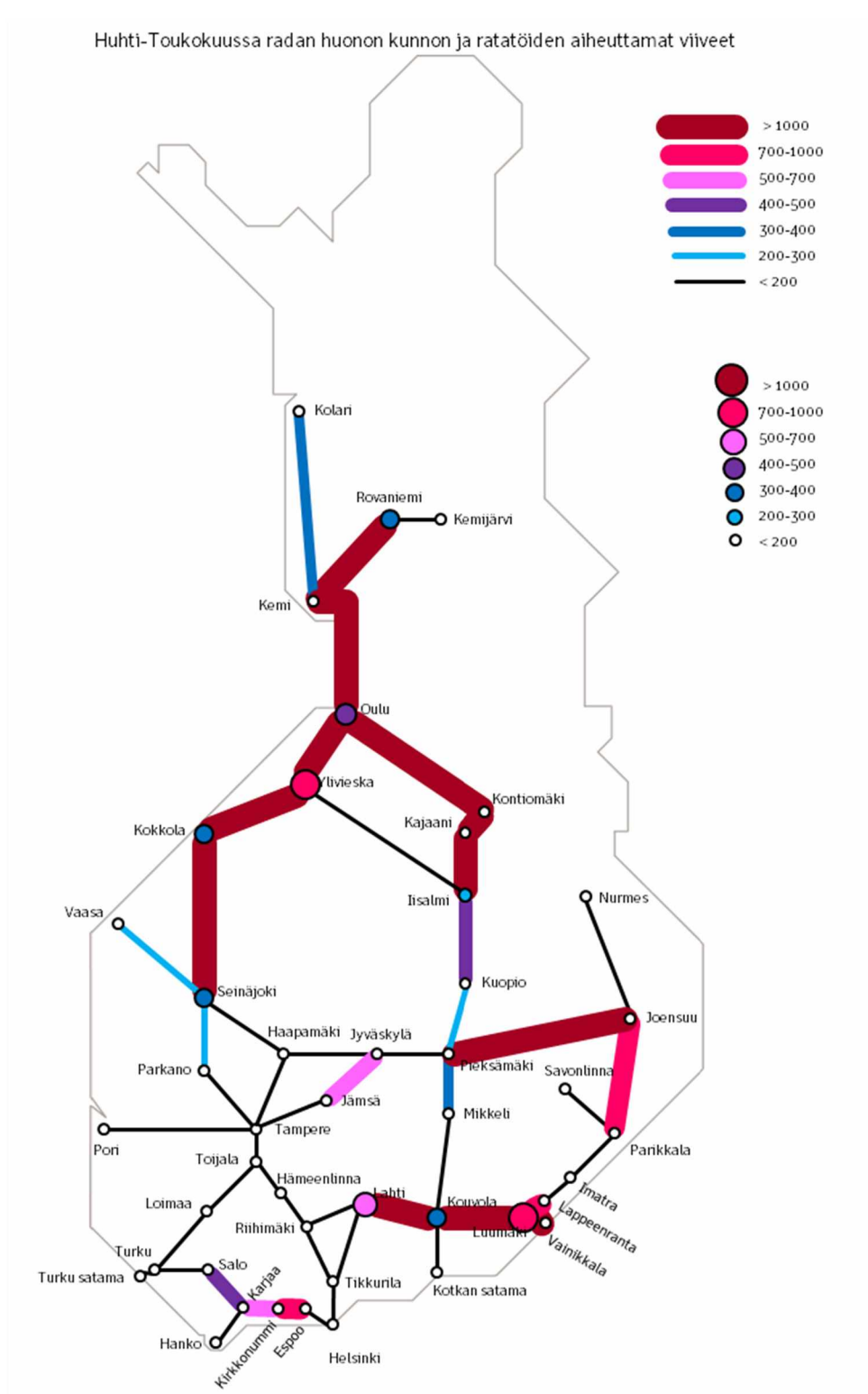
4.2.2 Analyysi

Vuonna 2010 radan huono kunto aiheutti vuoden aikana syntyneistä henkilökaukoliikenteen primäärisistä myöhästymisminuuteista noin 23 %. Tämä tarkoittaa 106 228 lisämyöhästymisminuuttia. Kuvassa 31 on tarkasteltu radanpitäjän vastuulla olevia viivesyitä aikasarjana seitsemän päivän summina. Kuten kuvasta voidaan havaita, on kevään routa-aika merkittävin syy radan huonolle kunnolle.

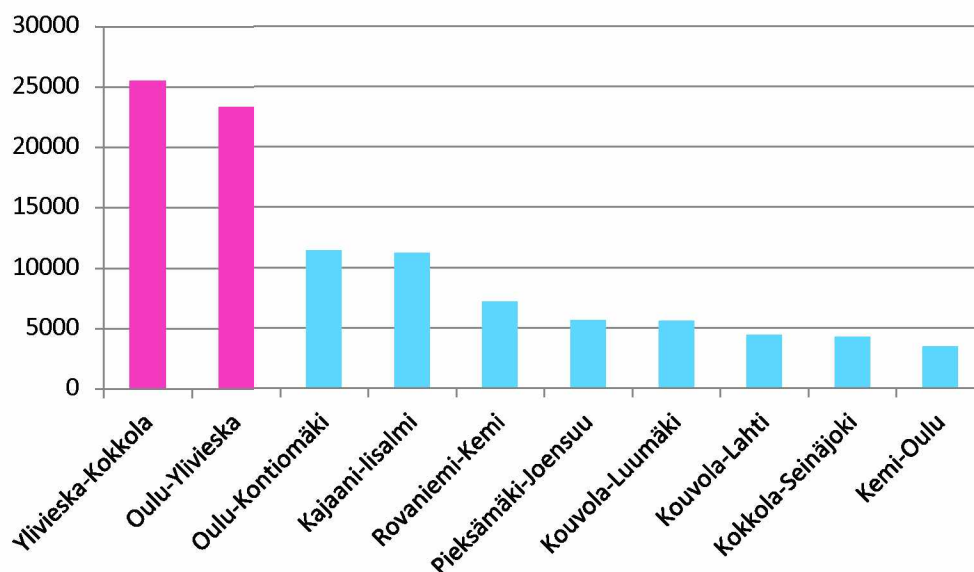


Kuva 31. Liikenneinfrastruktuuriin liittyvät viikoittain summatut myöhästymisminuutit kaukoliikenteessä vuonna 2010.

Kuvasta 32 voidaan todeta, että radan huonosta kunnosta johtuvat myöhästymiset sijoittuvat pääasiassa Pohjanmaan radalle, joskin viiveitä on merkittävästi myös muualla Suomessa. Kuva 33 kuitenkin paljastaa, että Oulu–Ylivieska- ja Ylivieska–Kokkola-rataosien viivesummat kahden kuukauden ajalta ovat merkittävästi suuremmat kuin muualla Suomessa. Datan tarkempi tutkinta paljastaa, että huhtikuun ja toukokuun aikana Oulu–Ylivieska-välillä lähes 90 % junista jäi yli viisi minuuttia lisää myöhään. Ylivieska–Kokkola-välillä vastaava suhteellinen lisämyöhästymisosuus oli lähes 80 %.



Kuva 32. Radan huonosta kunnosta ja ratatöistä johtuvat viiveet rataosittain vuonna huhti-toukokuussa 2010.



Kuva 33. Radan huonosta kunnosta ja ratatöistä johtuvat viiveminuutit rataosittain vuonna huhti-toukokuussa 2010.

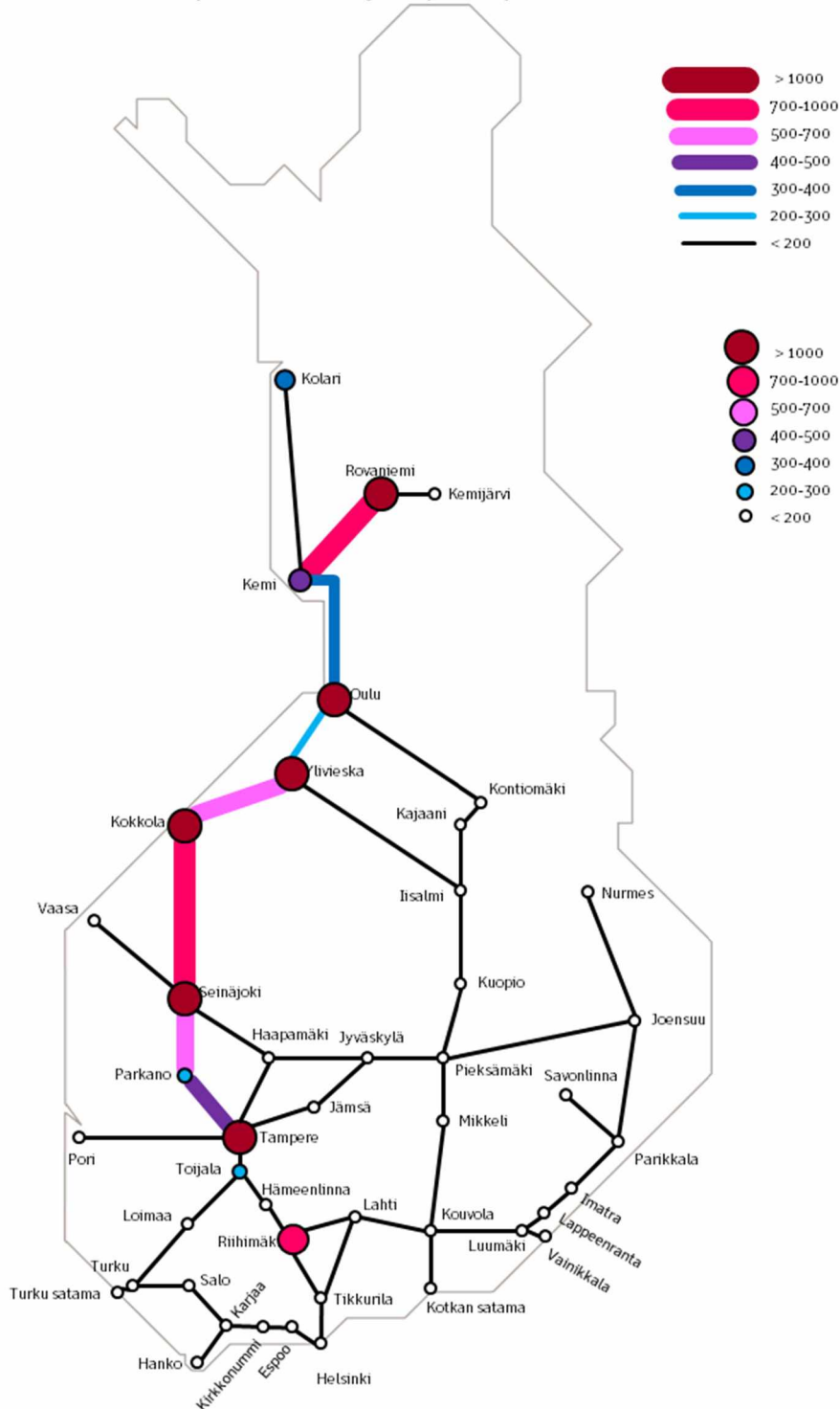
Liikennejärjestelmän näkökulmasta on mielenkiintoista ymmärtää, miten routa-aikana radan huonon kunnan ja tästä seuranneiden ratatöiden takia myöhästyneet junat vaikuttivat liikennöinnin täsmällisyyteen muussa valtakunnassa. Tätä analyysiä varten datasta poimittiin vuoden 2010 huhtikuun ja toukokuun aikana kulkeneiden junien viivetiedot. Tästä 52 570 havainnon datajoukosta poimittiin tämän jälkeen ne havainnot, joissa viivesyynä oli radan huono kunto ja ratatyöt sekä tapahtumapaikkana Kokkolan ja Oulun välinen rataosuus. Näiden 2745 havainnon perusteella poimittiin datasta ne viivehavainnot, joiden *syyn lisätieto* -kentässä on viite vastaavana päivänä Pohjanmaan radalla radan vuoksi myöhästyneeseen junaan. Näistä havainnoista karsittiin ne, jotka ilmenivät jo em. primäärimyöhästymishavainnoissa. Aidosti sekundäärisiä havaintoja nousi esiin 1894 kappaletta. Tätä vastaavaa iteraatiota toistettiin etsiessä sekundääristen havaintojen avulla tertiärisiä havaintoja (532 kpl) ja edelleen tertiärisien avulla kvartiaarisia havaintoja (365 kpl). Johtuen *syyn lisätieto* -kentän data puutteellisuudesta ei analyysin tuloksia voida pitää 100 % luotettavina. Sen sijaan ne toimivat lähinnä suuntaa-antavina ja yhtenä esimerkkinä, miten kokonaisliikennejärjestelmää voisi analysoida.

Toisen asteen viiveet

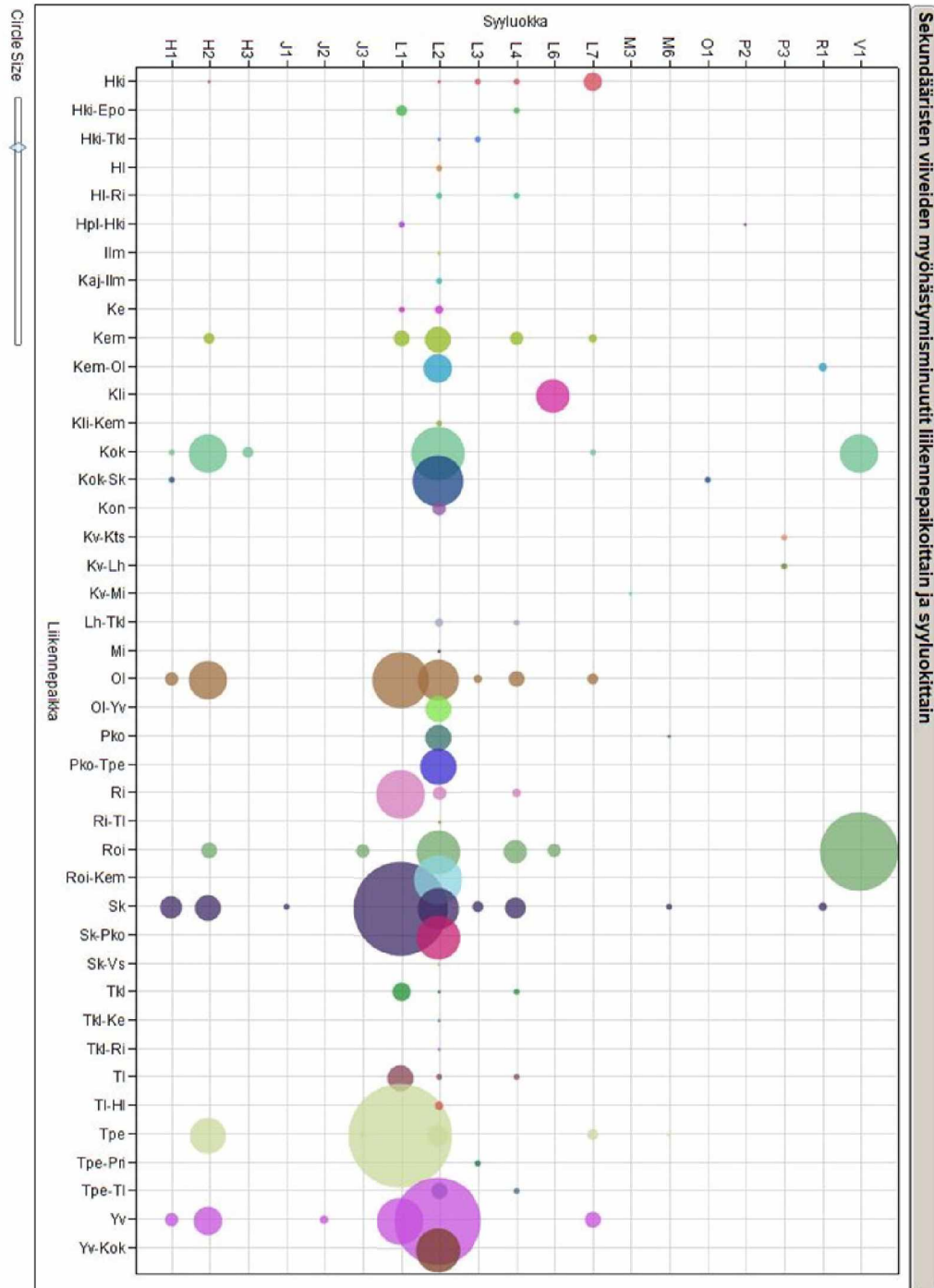
Kuvassa 34 on esitetty Pohjanmaan radan huonon kunnan ja ratatöiden vuoksi myöhästyneiden junien aiheuttamat toisen asteen viiveet. Radan huono kunto aiheuttaa merkittäviä viiveitä Tampereen, Seinäjoen, Kokkolan, Ylivieskan, Oulun, Rovaniemen sekä Riihimäen asemilla. Lisäksi Seinäjoki-Kokkola- sekä Rovaniemi-Kemi-rataosilla viiveet ovat huomattavia. Nämä sekundääriset myöhästymisminuutit on esitetty liikennepaikan ja syyluokan suhteen kuvassa 35 niin, että kutakin viivettä kuvaavan ympyrän koko kertoo myöhästymisminuuttien summan. Kuten kuvaajasta voidaan havaita, syyluokka L2 (junakohtaukset) ilmenee lähes kaikilla sekundäärisiä viiveitä kokeneilla liikennepaikoilla. Syyluokka L1 tarkoittaa yhteysliikenteen odotusta, tilannetta, jossa juna odottaa matkustajia toisesta junasta. Tämä syyluokka on kerryttänyt merkittäviä viiveitä Tampereella, Seinäjoella sekä Oulussa. Rovaniemellä ja Kokkolassa on tullut merkittävät sekundääriset viiveet V1-syyluokan, eli veturin odotuksen

vuoksi. Tämä tarkoittaa sitä, että kyseisillä liikennepaikoilla tavarajunat ovat odottaneet veturia toisesta junasta. Myös henkilökunnan odottamisesta on aiheutunut mainitsemisen arvoisia viiveitä Kokkolassa, Oulussa, Seinäjoella, Tampereella sekä Ylivieskassa.

Huhti-Toukokuussa Pohjanmaan radalla myöhästyneiden junien aiheuttamat sekundääriset viiveet



Kuva 34. Kokkola-Oulu välillä radan huonon kunnon ja ratatöiden vuoksi myöhästyneiden junien aiheuttamat toisen asteen viiveet huhti-toukokuussa 2010.



Kuva 35. *Kokkola-Oulu välillä radan huonon kunnon ja ratatöiden vuoksi myöhästymisten aiheuttamat toisen asteen viiveiden myöhästymisminuutit liikennepaikoittain ja syyluokittain huhti-toukokuussa 2010.*

Tampere on rautatieliikennejärjestelmän keskeisimpiä solmukohtia. Näin ollen siellä tapahtuneiden sekundääristen viiveiden tutkiminen on järjestelmän kannalta keskeistä. Kuten kuvasta 35 voidaan havaita, Tampereelle kohdistuvat viiveet ovat pääasiassa yhteysliikenteen odotusta. Tämän yhteysliikenteen odotuksen takia on syntynyt sekundäärisiä viiveitä 274 junalle 3265 minuutin verran. Tämä tarkoittaa keskimäärin

hieman vajaan kahdentoista minuutin odotusta per juna. Suunnittain viiveminuutit jakautuvat seuraavasti:

- Turku 34,85 %
- Jyväskylä 30,10 %
- Pori 23,48 %
- Helsinki 5,97 %
- Haapamäki 5,44 %
- Seinäjoki 0,16 %

Tampereelle kirjattuja lisämyöhästymisminuutteja on tarkasteluajanjaksolla ollut 9901 minuutin verran. Tämä tarkoittaa sitä, että Pohjanmaan suunnasta saapuvat junat ovat aiheuttaneet noin 33 % kaikista Tampereen sekundäärisistä viiveistä.

Riihimäen sijaitessa verrattain kaukana Kokkolan ja Oulu välisestä radasta, on sitä mielenkiintoista tutkia sekundäärysten viiveiden osalta. Kuvasta 35 voidaan havaita, että Riihimäelle Pohjanmaan radan huonosta kunnosta seuranneet sekundääriset viiveet ovat pääasiassa yhteysliikenteen odotusta. 86:lle eri junalle aiheutuneet viiveet (759 minuuttia, keskimäärin 8,8 minuuttia) muodostivat 10 % kaikista Riihimäen lisämyöhästymisminuuteista. Viiveistä vastasivat kymmenellä eri junanumerolla pohjoisesta Helsinkiin kulkevat junat. Nämä junat olivat Riihimäen kohdalla keskimäärin 39,9 minuuttia myöhässä huhti-toukokuun aikana. Tämä on merkittävästi enemmän kuin yhteysliikenteen odottamisen raja-arvo 20 minuuttia. Olisikin mielenkiintoista selvittää, minkä vuoksi Riihimäellä on odotettu yhteysliikennettä näin suurissa myöhästymistapauksissa.

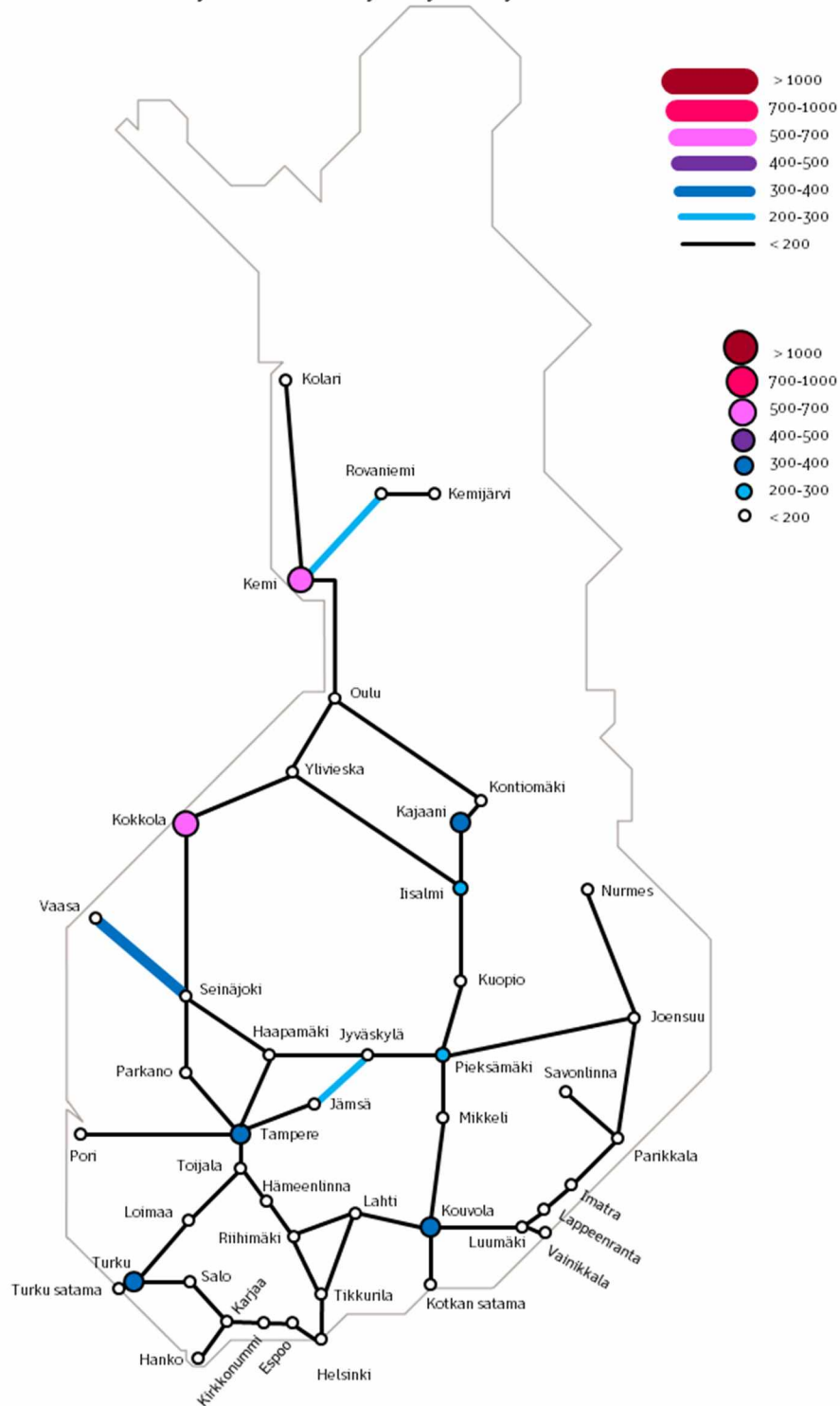
Pohjanmaan radan huonon kunto ja ratatyöt aiheuttivat ennen kaikkea yhteysliikenteen odotuksesta johtuvia sekundäärisiä viiveitä. Näiden viiveiden voidaan ajatella olevan erityisen haastavia liikennejärjestelmälle, koska ne heijastunevat edelleen muihin juniin. Kolmannen asteen viiveitä tarkastellessa onkin mielenkiintoista nähdä, miten esimerkiksi Tampereen ja Riihimäen odottamaan joutuneiden junien viivästy-miset näkyvät edelleen.

Kolmannen asteen viiveet

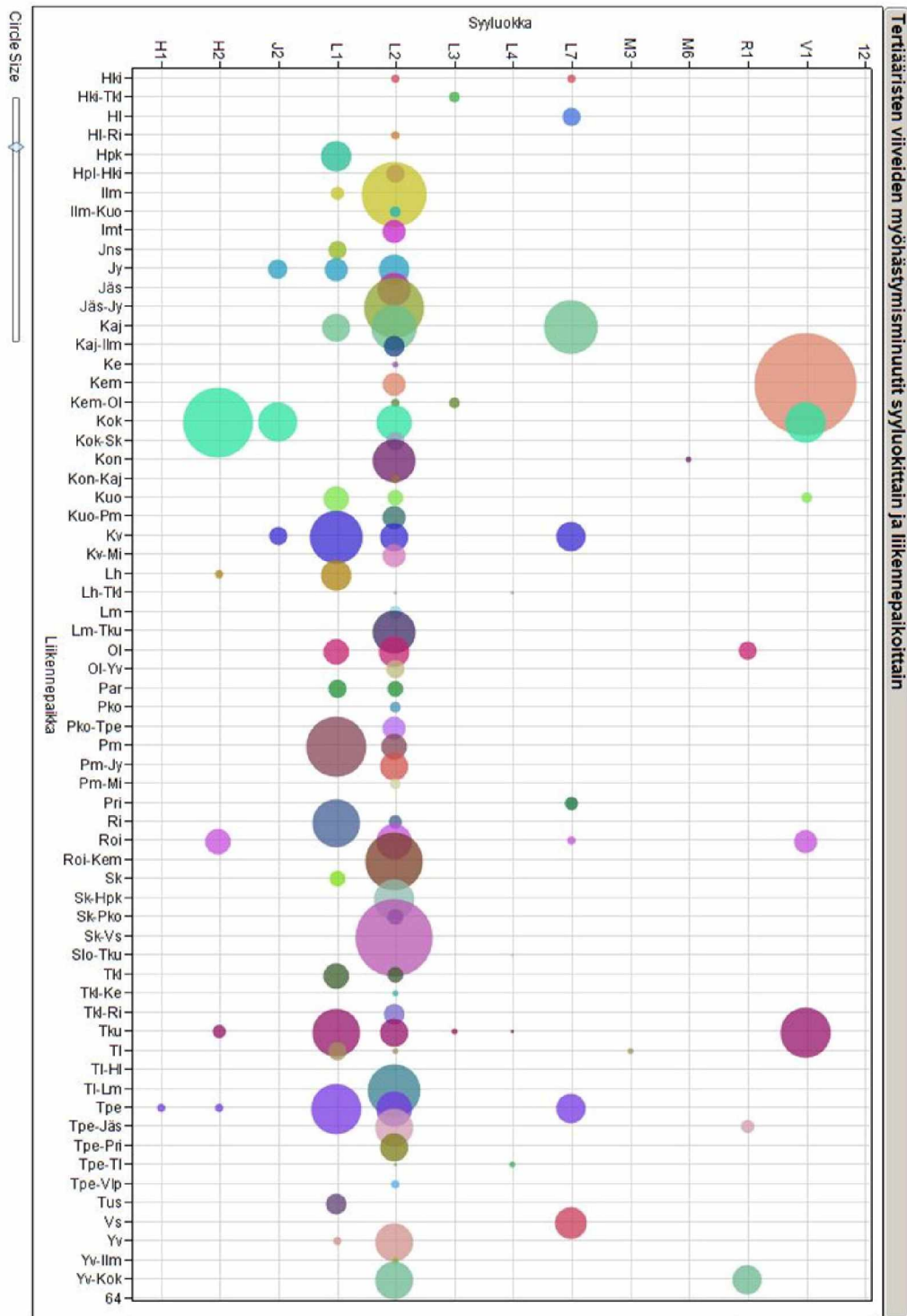
Kolmannen asteen viiveiden tarkastelussa mielenkiintoisena analysoitavana kohteena on tutkia, miten runsaasti toisen asteen viiveitä keränneiden liikennepaikkojen kuten Seinäjoen, Tampereen ja Riihimäen saamat viiveet näkyvät kolmannelle asteelle ketjuuntuneina viiveinä. Kuvassa 36 on Pohjanmaan radan huonon kunnan vuoksi kolmannelle asteelle ketjuuntuneiden viiveiden suuruus. Kuten karttavisualisoinnista voidaan havaita, toisella asteella Tampereella yhteysliikenteen odotuksesta johtuvat viiveet ovat kertaantuneet Jämsä–Jyväskylä-rataosalla sekä Pieksämäen asemalla idän suuntaan ajettaessa, sekä Turussa mentäessä länteen päin.

Sen sijaan Porin suuntaan ajettaessa ei ole Tampereen vuoksi syntynyt lisämyöhästymisiä. Tämä selittyy hyvin pitkälti sillä, että Porin radalla ei ole muuta henkilöliikennettä kuin Pori–Tampere-välin harvakseltaan kulkeva liikenne. Visualisoinnista voidaan myös havaita, että Riihimäellä yhteysliikenteen odotuksen vuoksi alun perin viiveitä saaneet junat aiheuttavat kolmannen asteen lisämyöhästymisiä Kouvolassa. Seinäjoella toisella asteella viivästyneet junat aiheuttavat kolmannen asteen viiveitä Seinäjoki–Vaasa-rataosalla.

Huhti-Toukokuussa Pohjanmaan radalla myöhästyneiden junien aiheuttamat tertiääriset viiveet



Kuva 36. Kokkola-Oulu välillä radan huonon kunnon ja ratatöiden vuoksi myöhästyneiden junien aiheuttamat kolmannen asteen viiveet huhti-toukokuussa 2010.



Kuva 37. Kokkola-Oulu välillä radan huonon kunnon ja ratatöiden vuoksi myöhästymisten aiheuttamat kolmannen asteen viiveiden myöhästymisminuutit liikennepaikoittain ja syylokkittain huhti-toukokuussa 2010.

Viiveiden ketjuuntumisen osalta kuva 37 paljastaa seuraavaa: Tampereelle, Turulle, Pieksämäelle sekä Kouvolalle aiheutuu kolmannen asteen viiveinä suhteellisesti eniten myöhästymisiä yhteysliikenteen odotuksen vuoksi. Huomioimisen arvoista on L7-viiveiden (Tulojana myöhässä) suhteellisen osuuden merkittävä kasvu niin Tampereella, Kajaanissa, Kouvolassa kuin Ylivieskassakin. Tulojan myöhästymisen takia huhti-toukokuussa lähti 18 junaa kukin keskimäärin reilu 26 minuuttia myöhässä.

Tampereelta eri suuntiin lähteneille junille merkittäviä lisämyöhästymisiä syntyy Jämsä–Jyväskylä-rataosalla junakohtausten vuoksi sekä Pieksämäellä ja Turussa yhteysliikenteen odotuksen vuoksi. Ottaen huomioon, että nämä liikennepaikat sijaitsevat keskenään eri suunnissa ja verraten kaukana Pohjanmaan radalta, voidaan todeta järjestelmän sisältävän suuria keskinäisiä riippuvuuksia.

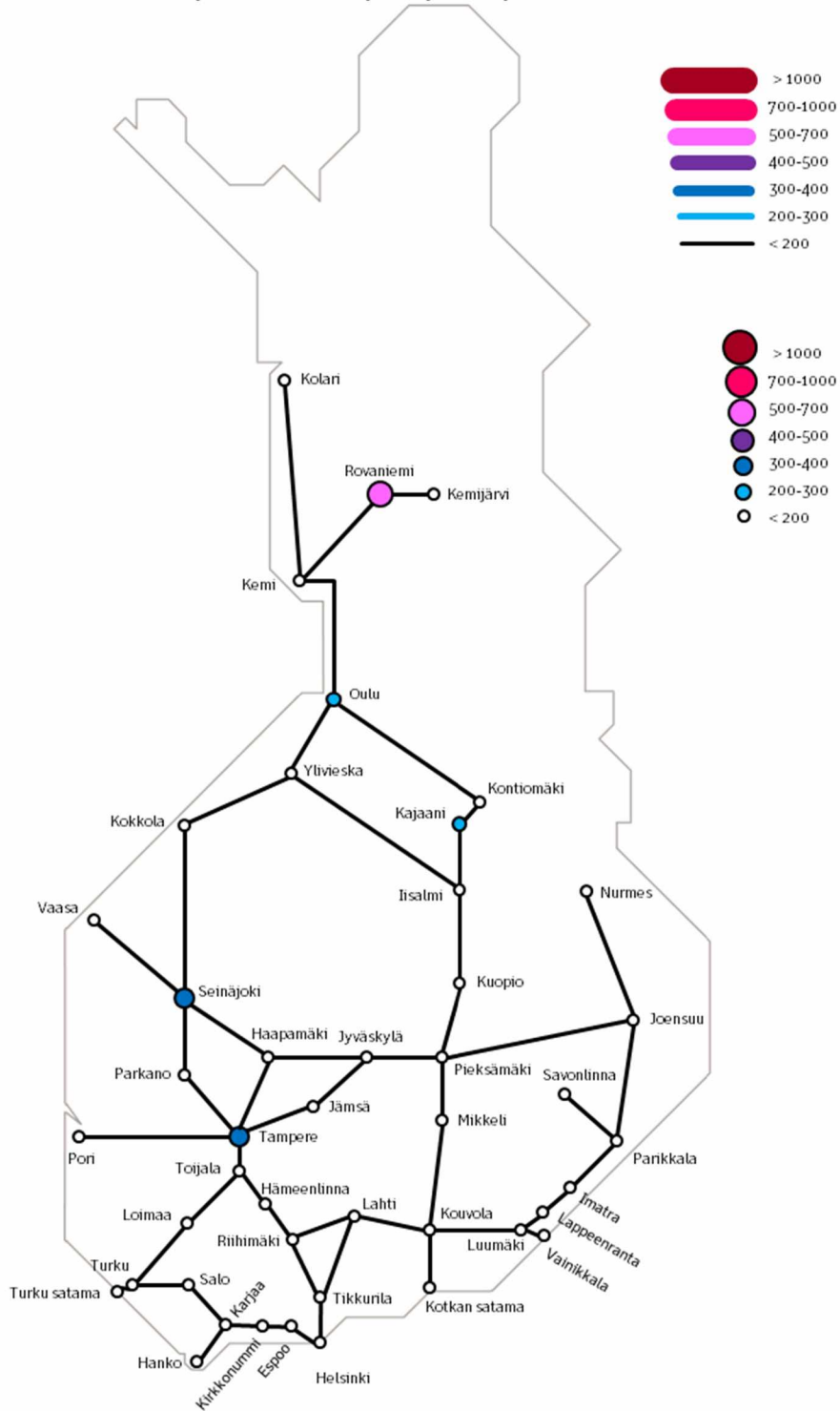
Neljännän asteen viiveet

Neljännän asteen viiveiden tarkastelu ei enää tarjoa juurikaan uutta (kuvat 38 ja 39). Viiveet noudattelevat hyvin pitkälti samaa kaavaa kuin toisen ja kolmannen asteen viiveissäkin, merkittävimpien viiveiden johtuessa junakohtauksista ja yhteysliikenteen odotuksesta. Lisäksi henkilökunnan ja veturien odotuksesta johtuvat viiveet ovat mainitsemisen arvoisia.

Vertaillen kolmannen ja neljännen asteen viiveitä nousee kuitenkin esille yksi mielenkiintoinen piirre järjestelmässä. Kun kolmannella asteella Kemin seuranta-asema keräsi veturin odotuksen (V1) vuoksi merkittävästi viiveitä (626 minuuttia), ei neljännellä asteella tätä viivettä enää ole. Sen sijaan Rovaniemelle on kertynyt veturin odotuksen vuoksi 363 minuuttia viiveitä. Tutkiessa dataa tarkemmin voidaan todeta, että näillä kahdella tapahtumalla ei ole junien suhteen mitään yhteyttä keskenään: Kemissä kolmannella asteella tapahtuvat veturin odotukset liittyvät lähinnä tavarajunien viiveisiin, kun taas neljännellä asteella Rovaniemellä liittyvät myös henkilökaukojuniin. Tästä pohjoisen runsaasta V1-viiveiden määrästä voidaan kuitenkin vetää se johtopäätös, että pohjoisessa veturien käyttöön liittyvä häiriöherkkyyks melko suurta.

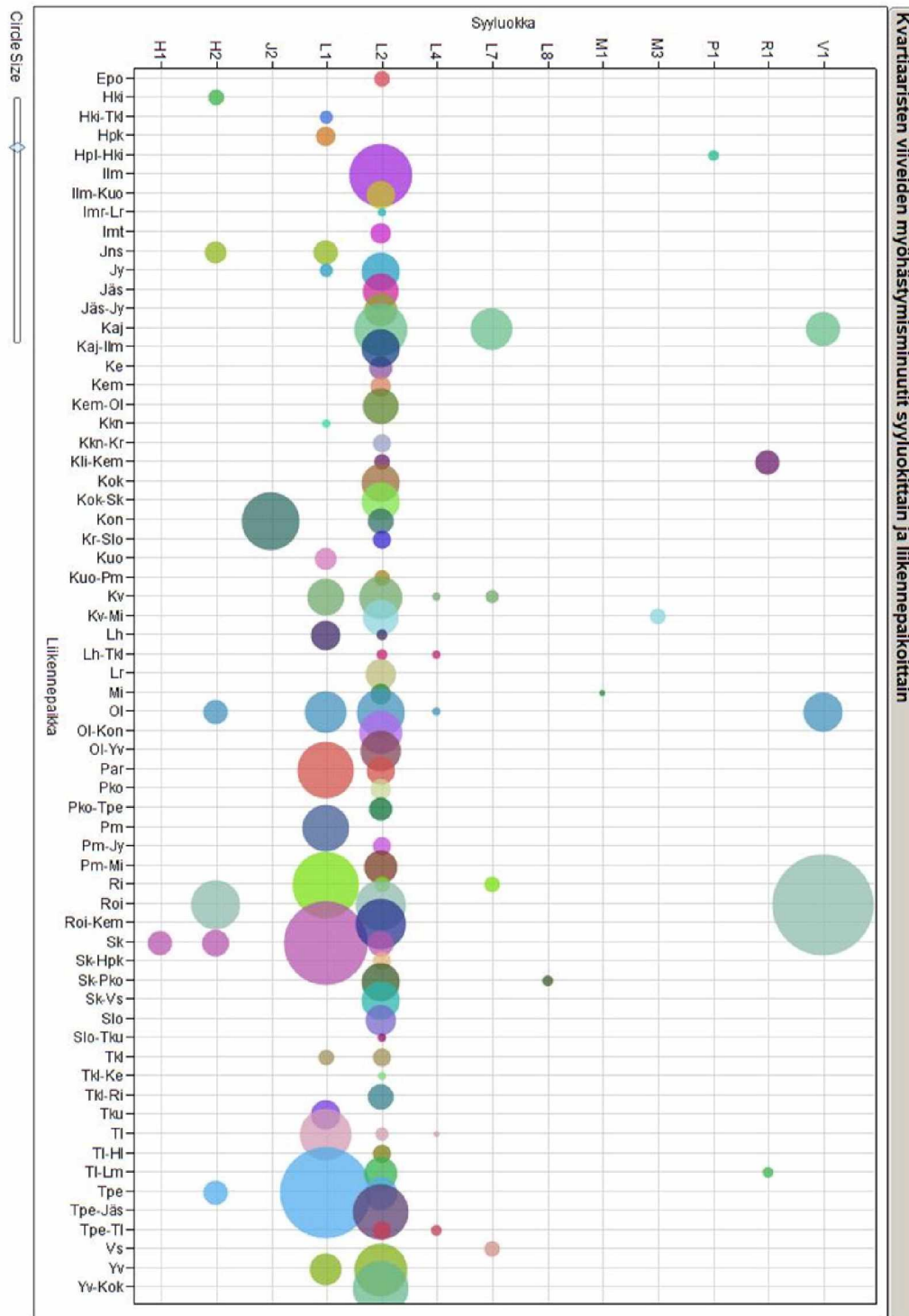
Tarkastelun pohjalta voidaan todeta, että valtakunnallista viiveiden ketjuuntumista on mahdollista tutkia jo olemassa olevan datan pohjalta. Analyysiprosessin aikana laaditut visualisoinnit olivat omiaan tehostamaan nopean ymmärryksen saavuttamista. Kunhan visualisointityökalu toimii datan pohjalta automaattisesti, ovat analyysien tulokset nopeasti nähtävissä ja tulkittavissa.

Huhti-Toukokuussa Pohjanmaan radalla myöhästyneiden junien aiheuttamat kvartaariset viiveet



Kuva 38.

Kokkola-Oulu välillä radan huonon kunnon ja ratatöiden vuoksi myöhästyneiden junien aiheuttamat neljännen asteen viiveet huhti-toukokuussa 2010.

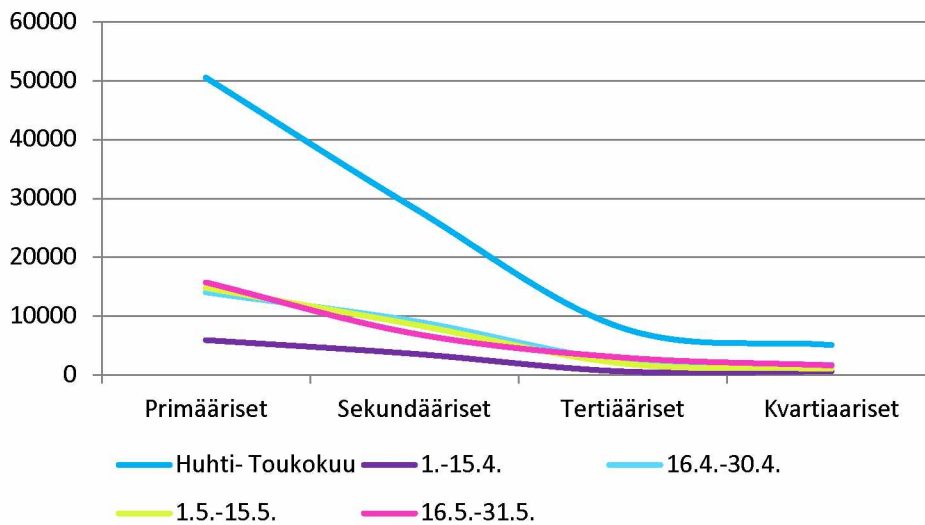


Kuva 39. Kokkola-Oulu välillä radan huonon kunnon ja ratatöiden vuoksi myöhästymisten aiheuttamat neljännen asteen viiveiden myöhästymisminuutit liikennepaikoittain ja syyloukittain huhti-toukokuussa 2010.

Viiveiden vaimeneminen

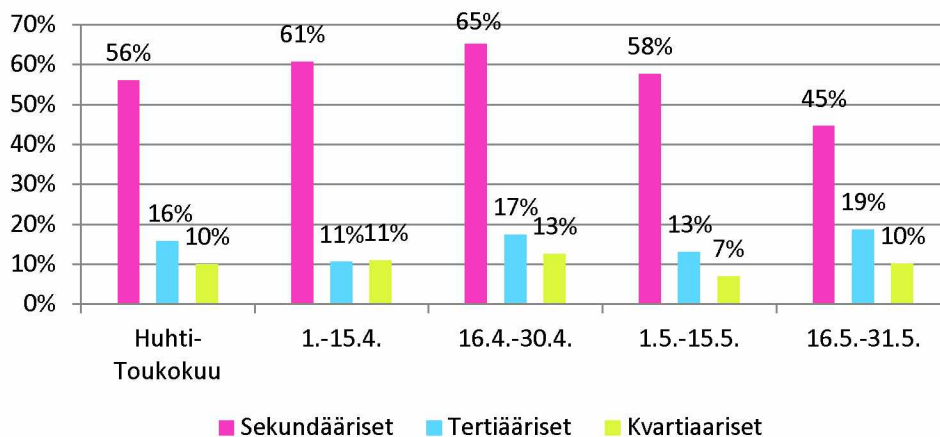
Viiveiden kokonaisvaikutuksen ymmärtämiseksi on mielenkiintoista tutkia viiveiden vaimenemista. Viiveiden vaimenemisella tarkoitetaan tässä asiayhteydessä sitä, miten viiveminuuttien määrä vähenee viiveasteittain alkuperäisistä viiveminuuteista. Kuvassa 40 on esitetty viiveminuuttien määrän vaimeneminen primäärisistä kvartiaa-

risiin niin huhti-toukokuun kokonaisjaksona kuin myös tämän kahden kuukauden ollessa jaettuna neljään eri jaksoon. Eri ajanjaksot noudattelevat toisiaan.



Kuva 40. Viiveminuuttien määrän vaimeneminen ensimmäisen asteen viiveistä neljännen asteen viiveisiin.

Kuvassa 41 on esitetty kuvaaja eri viiveasteiden viiveminuuttien prosentiosuuksista suhteutettuna primääristen viiveiden määrään. Prosentuaalinen kuvaus on toteutettu siksi, että pystytään vertailemaan vaimenemista riippumatta primääristen viiveiden alkuperäisestä suuruudesta. Kuten kuvaajasta voidaan havaita, eri viiveasteiden suhteellisten osuuksien varianssi eri ajanjaksojen välillä on pientä. Sekundäärisiä viiveminuutteja on keskimäärin 56 % primäärisistä, tertiäärisiä 16 % primäärisistä ja kvartiaarisia 10 % primäärisistä. Tästä saadaan yhteensä 82 % ketjuuntunutta viiveminuuttia suhteessa primäärisiin viiveisiin. Toisin sanoen yksi primäärinen viiveminuutti aiheuttaa vajaan yhden minuutin sekundäärisiä viiveitä. Tämä vastaa olemassa olevaa nyrkkisääntöä siitä, että yksi primäärinen viiveminuutti tuottaa yhden minuutin ketjuuntuneita viiveitä. Hiukan pienempi osuus selittyy sillä, että datakirjaukset ovat osin puutteellisia.



Kuva 41. Viiveminuuttien prosentuaalinen osuus alkuperäisistä primääri viiveminuuteista viiveasteittain.

Tämän tarkastelun pohjalta ei voida vetää yksiselitteisiä johtopäätöksiä viiveiden vaimenemisen käyttäytymisestä. Tarkastelu vaatii laajemman otoksen niin ajanjaksojen kuin rataosien suhteen. Viiveiden vaimenemisen suhteen olisikin mielenkiintoista tarkastella, miten tietyille rataosalle tai seuranta-asetalle aiheutuneet viiveet vaimenevat alkuperäisen ongelman ratkettua. Tällaisella tarkastelulla olisi mahdollista todeta, mitkä rataosat ovat häiriöherkkiä ja mitkä selviytyvät häiriöistä nopeasti.

Pohjanmaan junien aiheuttamien viiveiden sekvenssianalyysi

Pohjanmaan junien viiveiden ketjuuntumista voidaan tarkastella sekvenssianalyysin avulla: mitkä junat aiheuttavat viiveitä millekin junille ja missä. Tätä analyysiä varten valikoitiin kaikki ne havainnot, joiden *syyn lisätieto* -kentässä esiintyy huhti-toukokuussa 2010 Pohjanmaan rataa pitkin kulkenut juna.

Kuvassa 42 on esitetty näiden Pohjanmaan junien muille junille aiheuttamien viiveiden ketjuuntuminen junanumeroin. Kuten kuvasta voidaan havaita, junat 49, 54, 50 ja 58 aiheuttavat eniten viiveitä muille junille. Tapahtumamäärien perusteella merkittävimmät viiveketjut ovat junan 49 aiheuttamat viiveet junalle 5168 (29 kertaa), junan 53 aiheuttamat viiveet junalle 339 (27 kertaa) sekä junan 54 aiheuttamat viiveet junalle 53 (26 kertaa). Näitä viiveitä on tapahtunut lähes joka toinen kerta, kun kyseiset junat ovat olleet liikenteessä huhti-toukokuussa 2010.

Luottamusasteen avulla voidaan tarkastella, kuinka suuri osa junan aiheuttamista viiveistä on aiheutunut juuri tarkasteltavalle junalle. Esimerkiksi juna 5432 aiheutti 58,33 % kaikista aiheuttamista viiveistään junalle 5401. Juna 410 puolestaan aiheutti 36,84 % kaikista aiheuttamista viiveistään junalle 53. Tämä tieto toimii erinomaisena pohjana, kun lähdetään arvioimaan, miten tietyn junan viiveitä kannattaisi lähteä vähentämään.

	Count	Item
1	216	49
2	197	54
3	183	50
4	180	48
5	165	59
6	130	53
7	126	58
8	117	60
9	117	55
10	111	52
11	107	46
12	103	56
13	102	274
14	97	266
15	93	45
16	84	43
17	78	41
18	77	273
19	62	47
20	59	265
21	53	5016
22	38	5400
23	38	410
24	33	453
25	31	407
26	30	272
27	29	5168
28	29	3067
29	28	5014
30	28	269
31	27	449
32	27	267
33	26	91
34	25	5012
35	24	5432
36	23	934
37	22	911
38	22	62
39	22	494
40	21	916
41	20	5006
42	20	465
43	19	3053
44	19	3051
45	18	928
46	18	5401
47	18	452
48	18	447

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
1	2	1.37	13.43	29 49 ==>	5168
2	2	1.28	20.77	27 53 ==>	449
3	2	1.23	13.20	26 54 ==>	53
4	2	0.99	11.67	21 48 ==>	911
5	2	0.99	11.67	21 48 ==>	916
6	2	0.95	11.11	20 48 ==>	465
7	2	0.95	12.12	20 59 ==>	407
8	2	0.95	18.69	20 46 ==>	43
9	2	0.85	9.14	18 54 ==>	928
10	2	0.85	9.84	18 50 ==>	49
11	2	0.85	14.29	18 58 ==>	59
12	2	0.81	10.30	17 59 ==>	453
13	2	0.81	13.49	17 58 ==>	934
14	2	0.76	8.12	16 54 ==>	55
15	2	0.76	20.51	16 41 ==>	405
16	2	0.76	8.12	16 54 ==>	471
17	2	0.76	8.89	16 48 ==>	47
18	2	0.71	12.82	15 60 ==>	453
19	2	0.71	14.56	15 56 ==>	55
20	2	0.71	8.20	15 50 ==>	706
21	2	0.66	36.84	14 410 ==>	53
22	2	0.66	7.78	14 48 ==>	443
23	2	0.66	58.33	14 5432 ==>	5401
24	2	0.66	7.78	14 48 ==>	315
25	2	0.66	22.58	14 47 ==>	60
26	2	0.66	7.11	14 54 ==>	45
27	2	0.62	11.11	13 55 ==>	5016
28	2	0.62	6.02	13 49 ==>	62
29	2	0.62	6.02	13 49 ==>	60
30	2	0.62	6.60	13 54 ==>	494
31	2	0.62	7.10	13 50 ==>	45
32	2	0.62	12.15	13 46 ==>	41
33	2	0.62	13.98	13 45 ==>	54
34	2	0.62	11.11	13 55 ==>	91
35	2	0.62	6.60	13 54 ==>	91
36	2	0.57	9.23	12 53 ==>	410
37	2	0.57	10.81	12 52 ==>	924
38	2	0.57	6.67	12 48 ==>	45
39	2	0.57	12.37	12 266 ==>	59
40	2	0.57	14.29	12 43 ==>	52
41	2	0.57	10.26	12 60 ==>	59
42	2	0.57	9.52	12 58 ==>	451
43	2	0.52	10.28	11 46 ==>	441
44	2	0.52	10.68	11 56 ==>	49
45	2	0.52	5.09	11 49 ==>	274
46	2	0.52	10.28	11 46 ==>	912
47	2	0.52	6.01	11 50 ==>	41
48	2	0.52	9.91	11 52 ==>	43
49	2	0.47	5.46	10 50 ==>	273
50	2	0.47	8.55	10 60 ==>	407
51	2	0.47	7.69	10 53 ==>	921
52	2	0.47	6.06	10 59 ==>	927

Kuva 42. Pohjanmaan junien viiveiden ketjuuntuminen. Vasemmassa taulukossa on esitetty datassa ilmenevien junanumeroiden lukumäärät ja oikeassa taulukossa sekvenssit. Sekvenssiketjussa vasemmat junat ovat syyn lisätietokentästä poimitut junanumerot, eli ”viiveen aiheuttajat” ja oikealla puolella junanumerokentästä poimitut junanumerot eli ”viiveen kokee- neet”.

Toinen tapa tarkastella viiveiden ketjuuntumista on tutkia, mitkä junat aiheuttavat viiveitä milläkin liikennepaikoilla ja mistä syystä. Kuvassa 43 on esitetty viiveitä aiheuttaneet junat viiveen kokeneiden syykoodi-liikennepaikka-yhdistelmien suhteen. Mielenkiintoista on esimerkiksi se, että juna 48 on aiheuttanut 59 kertaa yhteysliikenteen odotuksesta johtuvan viiveen Tampereelta lähtevälle toiselle junalle ja nämä viiveet ovat vain 35,98 % kaikista sen aiheuttamista viiveistä. Kuten kuvan vasemmanpuoleisesta taulukosta voidaan havaita, kokevat Tampereelta lähtevät junat 288 kertaa tällaisen yhteysliikenteen odotuksesta johtuvan viiveen.

	Count	Item		Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
1	288	L1 Tpe	1	2	2.98	35.98	59 48 ==>	L1 Tpe
2	268	L2 Yv	2	2	1.85	22.35	38 54 ==>	L1 Tpe
3	187	L1 Sk	3	2	1.75	36.36	36 59 ==>	L1 Sk
4	170	54	4	2	1.46	33.71	30 52 ==>	L1 Tpe
5	164	48	5	2	1.41	18.83	29 49 ==>	V1 Roi
6	163	50	6	2	1.36	37.84	28 53 ==>	L1 Sk
7	154	49	7	2	1.27	24.76	26 58 ==>	L1 Tpe
8	114	L2 Sk-Pko	8	2	1.12	14.11	23 50 ==>	L1 Tpe
9	109	L2 Pko-Tpe	9	2	1.07	24.18	22 56 ==>	L1 Tpe
10	105	58	10	2	1.07	13.41	22 48 ==>	L1 Ri
11	99	59	11	2	1.07	26.51	22 60 ==>	L1 Sk
12	98	46	12	2	1.02	12.35	21 54 ==>	L2 Pko-Tpe
13	97	L2 Kok	13	2	0.97	20.41	20 46 ==>	L1 Tpe
14	93	L2 Kok-Sk	14	2	0.97	12.20	20 48 ==>	L2 Pko-Tpe
15	91	56	15	2	0.97	12.20	20 48 ==>	L1 Sk
16	89	52	16	2	0.97	19.05	20 58 ==>	L1 Sk
17	88	L1 Ri	17	2	0.97	12.27	20 50 ==>	L2 Pko-Tpe
18	83	60	18	2	0.83	11.04	17 49 ==>	L2 Roi-Kem
19	76	274	19	2	0.83	17.17	17 59 ==>	L1 Tpe
20	74	53	20	2	0.83	10.43	17 50 ==>	L2 Kok-Sk
21	68	266	21	2	0.83	11.04	17 49 ==>	L2 Sk-Pko
22	64	55	22	2	0.83	58.62	17 5400 ==>	L2 Yv
23	53	L2 Yv-Kok	23	2	0.78	72.73	16 410 ==>	L2 Yv
24	52	45	24	2	0.78	10.39	16 49 ==>	L1 Sk
25	48	L2 OI	25	2	0.78	16.16	16 59 ==>	L2 Yv
26	47	41	26	2	0.78	10.39	16 49 ==>	L2 Yv
27	43	L2 Pko	27	2	0.73	23.44	15 55 ==>	L2 Yv
28	43	L1 OI	28	2	0.73	28.85	15 45 ==>	L2 Yv
29	40	L2 Roi-Kem	29	2	0.68	8.24	14 54 ==>	L2 Yv
30	38	273	30	2	0.68	8.24	14 54 ==>	L2 Sk-Pko
31	35	265	31	2	0.68	8.59	14 50 ==>	L1 OI
32	33	43	32	2	0.68	18.42	14 274 ==>	L1 Tpe
33	32	L2 Sk	33	2	0.68	8.59	14 50 ==>	L1 Ri
34	30	V1 Roi	34	2	0.68	20.59	14 266 ==>	L2 Yv
35	30	L1 Yv	35	2	0.63	7.65	13 54 ==>	L1 Sk
36	29	5400	36	2	0.63	7.65	13 54 ==>	L1 Yv
37	28	L2 Kem-OI	37	2	0.63	7.98	13 50 ==>	L2 Kem-OI
38	28	47	38	2	0.63	34.21	13 273 ==>	L2 Sk-Pko
39	24	L2 Tpe	39	2	0.63	15.66	13 60 ==>	L2 Kok-Sk
40	24	L2 OI-Yv	40	2	0.58	16.22	12 53 ==>	L2 Yv
41	22	410	41	2	0.58	60.00	12 5432 ==>	V1 Kok
42	21	L2 Ri	42	2	0.58	36.36	12 43 ==>	L2 Yv
43	21	L2 Kem	43	2	0.54	6.47	11 54 ==>	L1 Ri
44	20	V1 Kok	44	2	0.54	10.48	11 58 ==>	L2 Sk-Pko
45	20	5432	45	2	0.54	11.22	11 46 ==>	L1 Sk
46	18	H2 Sk	46	2	0.54	6.75	11 50 ==>	L2 Yv
47	18	H2 OI	47	2	0.54	12.09	11 56 ==>	L2 Kok
48	17	5016	48	2	0.54	13.25	11 60 ==>	L1 Tpe
			49	2	0.49	11.24	10 52 ==>	L2 Yv
			50	2	0.49	12.05	10 60 ==>	L2 Yv
			51	2	0.49	6.49	10 49 ==>	L2 Kem
			52	2	0.49	5.88	10 54 ==>	L2 Kok

Kuva 43. Pohjanmaan junien viiveiden ketjuuntuminen. Vasemmassa taulukossa on esitetty datassa ilmenevien junanumeroiden ja syykoodi-rataosatiето-yhdisteiden lukumäärät ja oikeassa taulukossa on esitetty sekvenssit. Sekvenssiketjussa vasemmat junat ovat syyn lisätietokentästä poimitut junanumerot, eli ”viiveen aiheuttajat”, oikealla puolella on yhdistetty syykoodi ja rataosatiето.

Sekvenssejä voidaan ketjuttaa myös pidemmissä osissa. Kuvan 44 sekvenssianalyysin pohjalla oleva data on muokattu niin, että sekvenssin ensimmäisenä vaiheena on syyn lisätieto -kentästä poimittu junan numero, toisena vaiheena syykoodi-liikennepaikka-yhdistelmä ja kolmantena viiveen kokenut juna. Tällä tavoin tarkastellessa voidaan tutkia koko ketjua: mitkä junat aiheuttavat, mitä viiveitä, missä ja mille junille. Kuvasta voidaan tulkita esimerkiksi seuraavaa (rivi 1): juna 49 on aiheuttanut 29 kertaa V1-viiveen Rovaniemellä junalle 5168. Tämän tapahtuman luottamusaste on 100 %. Toisin sanoen aina kun juna 49 on aiheuttanut Rovaniemellä toiselle junalle viiveen veturin odotuksesta, on kärsijänä ollut juna 5168. Tällaisen monta osaa sisältävän ketjun sekvenssianalyysin avulla on mahdollista tutkia ilmiöitä tarkemmin.

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
1	3	1.41	100.00	29	49 ==> V1 Roi ==> 5168
2	3	1.32	96.43	27	53 ==> L1 Sk ==> 449
3	3	1.02	35.59	21	48 ==> L1 Tpe ==> 911
4	3	0.97	33.90	20	48 ==> L1 Tpe ==> 465
5	3	0.97	95.56	20	59 ==> L1 Sk ==> 407
6	3	0.83	95.00	17	50 ==> L2 Pko-Tpe ==> 49
7	3	0.78	90.00	16	48 ==> L2 Pko-Tpe ==> 47
8	3	0.78	44.44	16	59 ==> L1 Sk ==> 453
9	3	0.78	61.54	16	58 ==> L1 Tpe ==> 934
10	3	0.78	42.11	16	54 ==> L1 Tpe ==> 471
11	3	0.73	25.42	15	48 ==> L1 Tpe ==> 916
12	3	0.73	71.43	15	54 ==> L2 Pko-Tpe ==> 55
13	3	0.73	68.18	15	60 ==> L1 Sk ==> 453
14	3	0.68	100.00	14	54 ==> L2 Yv ==> 45
15	3	0.68	87.50	14	410 ==> L2 Yv ==> 53
16	3	0.68	70.00	14	48 ==> L1 Sk ==> 443
17	3	0.68	100.00	14	54 ==> L2 Sk-Pko ==> 53
18	3	0.68	100.00	14	50 ==> L1 OI ==> 706
19	3	0.63	59.09	13	48 ==> L1 Ri ==> 315
20	3	0.63	76.47	13	50 ==> L2 Kok-Sk ==> 45
21	3	0.63	76.47	13	49 ==> L2 Sk-Pko ==> 62
22	3	0.63	81.25	13	49 ==> L2 Yv ==> 60
23	3	0.63	34.21	13	54 ==> L1 Tpe ==> 91
24	3	0.63	86.67	13	45 ==> L2 Yv ==> 54
25	3	0.63	100.00	13	54 ==> L1 Yv ==> 494
26	3	0.58	100.00	12	43 ==> L2 Yv ==> 52
27	3	0.58	100.00	12	5432 ==> V1 Kok ==> 5401
28	3	0.58	90.00	12	55 ==> L2 Yv ==> 5016
29	3	0.58	100.00	12	53 ==> L2 Yv ==> 410
30	3	0.54	64.71	11	49 ==> L2 Roi-Kem ==> 274
31	3	0.54	78.57	11	266 ==> L2 Yv ==> 59
32	3	0.54	100.00	11	50 ==> L2 Yv ==> 41
33	3	0.54	100.00	11	56 ==> L2 Kok ==> 49
34	3	0.54	36.67	11	52 ==> L1 Tpe ==> 924
35	3	0.54	95.00	11	58 ==> L1 Sk ==> 451
36	3	0.49	100.00	10	46 ==> L2 Sk-Pko ==> 43
37	3	0.49	100.00	10	52 ==> L2 Yv ==> 43
38	3	0.49	100.00	10	41 ==> L1 OI ==> 405
39	3	0.49	76.92	10	50 ==> L2 Kem-OI ==> 273
40	3	0.49	76.92	10	54 ==> L1 Sk ==> 447
41	3	0.49	58.82	10	59 ==> L1 Tpe ==> 927
42	3	0.44	100.00	9	45 ==> L1 Yv ==> 494
43	3	0.39	57.14	8	274 ==> L1 Tpe ==> 904
44	3	0.39	50.00	8	49 ==> L1 Sk ==> 447
45	3	0.39	80.00	8	60 ==> L2 Yv ==> 49
46	3	0.39	61.54	8	273 ==> L2 Sk-Pko ==> 5016
47	3	0.39	40.00	8	58 ==> L1 Sk ==> 452
48	3	0.39	21.05	8	54 ==> L1 Tpe ==> 928
49	3	0.39	100.00	8	47 ==> H2 OI ==> 60
50	3	0.34	100.00	7	58 ==> L2 Pko-Tpe ==> 59
51	3	0.34	97.50	7	56 ==> H2 Sk ==> 55
52	3	0.34	100.00	7	48 ==> L2 Sk-Pko ==> 45

Kuva 44. Pohjanmaan junien viiveiden ketjuuntuminen. Sekvenssiketjussa vasemmanpuoleiset junat ovat syyn lisätieto -kentästä poimitut junanumerot, eli "viiveen aiheuttajat", keskellä on syykoodi-rataosa-yhdistelmät, jotka kertovat, mitä on tapahtunut ja missä, sekä oikealla puolella juna-numerot, jotka viittaavat viiveen kärsineisiin juniin.

4.2.3 Päätelmät

Tämä analyysi oli vain pintaraapaisu siihen, miten kehittyneen data-analytiikan menetelmiä voidaan hyödyntää luotaessa kokonaisymmärrystä jostakin ilmiöstä. Analyysissä keskityttiin tutkimaan, miten yksittäisen ilmiön aiheuttamat viiveet ketjuuntuvat eri puolille Suomen rautatieliikennejärjestelmää ja mitkä viiveet ylipäättään ovat ketjuuntuneita. Tämän lisäksi tutkittiin, miten viiveminuuttien suuruudet vaimenevat viiveasteen mukaan sekä miten, miksi ja missä eri junat aiheuttavat viiveitä toisilleen.

Toisen asteen viiveiden analyysiä tarkastellessa havaittiin, että perinteisesti ketjuuntuneina viiveinä pidetyt yhteysliikenteen odotus (L1) sekä junakohtaus (L2) ovat vahvasti läsnä. Tämän lisäksi havaittiin, että perinteisesti primäärisiksi luokiteltuja viiveitä, kuten veturin odotus (V1), henkilökunnan vaihto (H1) ja henkilökunnan odotus (H2), ilmeni runsain määrin myös sekundäärisinä. Tämä lisäsi ymmärrystä siitä, miten viiveet ketjuuntuvat.

Viiveiden ketjuuntuminen ja viivevaikutusten leviäminen oli hyvin tulkittavissa eri vaiheiden visualisoinneista. Analyysin mielenkiintoisinta antia olivat kolmannen ja neljännen asteen viiveiden tarkastelut. Ne osoittivat, kuinka paikallisen ilmiön viiveet heijastuvat koko maahan. Jatkossa tulisi tarkastella yksittäisten liikennepaikkojen ja -rataosien käyttäytymistä ja toipumiskykyä ketjuuntuneiden viiveiden syntymisen ja niistä toipumisen suhteen. Tähän olisi hyvä saada liitettyä myös kellonaikatieta.

Myös tässä analyysissä sekvenssitarkastelu näytti voimansa. Kun viiveen aiheuttaja laitettiin sekvenssijärjestykseen ensimmäiseksi, voitiin tarkastella, mille junille ja syykoodi-liikennepaikka-yhdistelmille kukin juna aiheutti viiveitä. Jatkossa olisi mielenkiintoista kääntää sekvenssianalyysi toisin päin. Silloin pystyttäisiin esimerkiksi tutkimaan, mitkä junat ovat olleet aiheuttamassa tarkasteltavalle junalle viiveitä ja millaisen osuuden nämä viiveet ovat aiheuttaneet kaikista viiveistä.

Kokonaisuutena tehty analyysi oli varsin rohkaiseva. Kokonaisjärjestelmää koskettavia ilmiöitä on suhteellisen yksinkertaista analysoida, kun on olemassa laadukkaat työkalut niin datan muokkaukseen kuin myös sen analysointiin. Menetelmien avulla on mahdollista analysoida yksittäisten ilmiöiden vaikutuksia laajemmin ja useista eri näkökulmista. Näin analyysin myötä voidaan muodostaa entistä enemmän ymmärrystä liikennejärjestelmästä kokonaisuutena.

4.3 Matkustajapalvelun viiveiden vaikutus

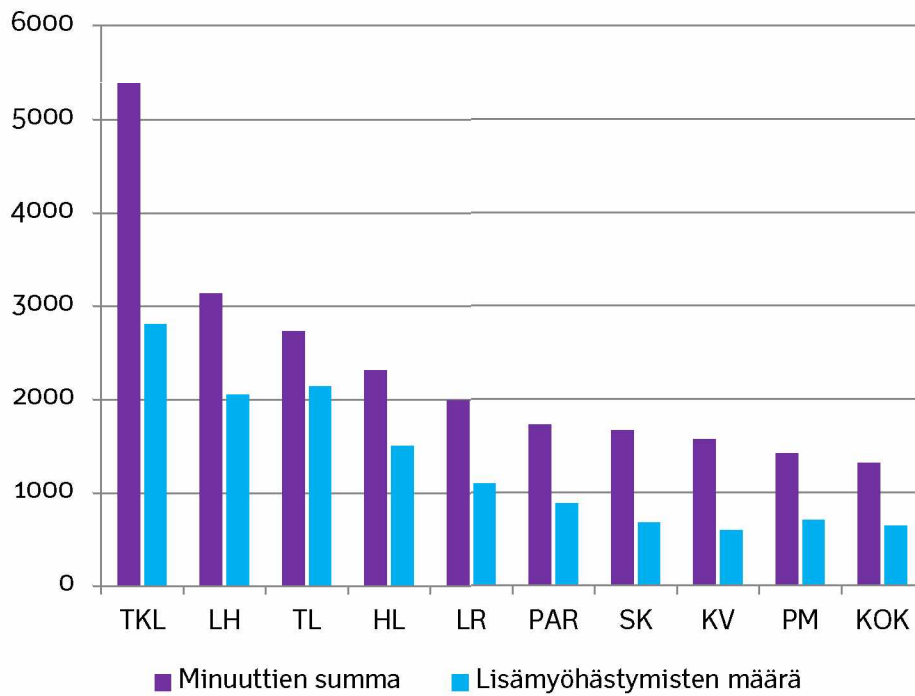
4.3.1 Väite analyysin taustalla

”M-syistä johtuvat viiveet ovat niin lyhyitä, ettei niillä ole vaikutusta täsmällisyyteen: nämä minuutin-parin viiveet saadaan ajettua kiinni.”

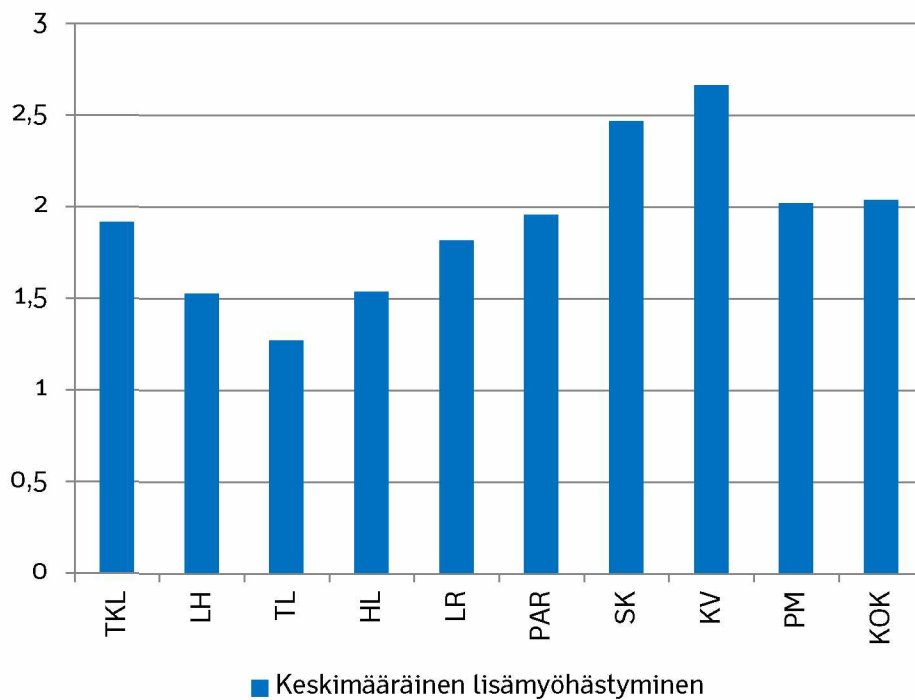
4.3.2 Analyysi

Vuoden 2010 aikana Suomen rautatieverkolla koettiin 119 598 lisämyöhästymistä, joista kertyi yhteensä 912 989 minuuttia. Matkustajaruuhkan (M1) ja pysähtymisajan ylityksen (M6) takia olevia lisämyöhästymisiä oli näistä 22 802 kappaletta, joista muodostui 44 108 myöhästymisminuuttia. Näin ollen matkustajaruuhkasta ja pysähtymisajan ylityksestä johtuvat lisämyöhästymiset vastasivat 19,1 % lisämyöhästymistapauksista, mutta vain 4,8 % minuuteista. Viiveet ovat siis keskimäärin suhteellisen lyhyitä.

Kuvassa 45 on esitetty kymmenen eniten viiveminuutteja matkustajaruuhkan ja pysähtymisajan ylityksen vuoksi kokenutta liikennepaikkaa. Kuvassa 46 on puolestaan esitetty näiden liikennepaikkojen M1- ja M6-syistä johtuvien viiveiden keskimääräiset kestot. Viiveet ovat kestoltaan muutamien minuuttien luokkaa.



Kuva 45. Eniten M1- ja M6-syistä viiveminuutteja saaneet liikennepaikat vuonna 2010 (TKL=Tikkurila, LH=Lahti, TL=Toijala, HL=Hämeenlinna, LR=Lappeenranta, PAR=Parikkala, SK=Seinäjoki, KV=Kouvola, PM=Pieksämäki, KOK=Kokkola).



Kuva 46. M1- ja M6-viiveiden keskimääräiset kestot liikennepaikoittain.

Yllä on esitetty ne liikennepaikat, jotka kokivat eniten viiveitä M1- ja M6-syistä. Mutta ovatko nämä viiveet merkittäviä kokonaistasmällisyyden suhteen? Jotta tähän voitaisiin vastata, on tarpeen tutkia, saadaanko näitä viiveitä kiinni seuraavalla rataosalla. Tätä varten on dataksi valittu vuoden 2010 ensimmäisen kvartaalin viivedata.

Vuoden 2010 ensimmäisen kvartaalin aikana tapahtui 79 720 yksittäistä lisämyöhästymistä, joista kertyi 478 444 minuuttia. 13 354 lisämyöhästymistä johtui matkustajaruuhkasta ja pysähtymisajan ylityksestä, kerryttäen yhteensä 23 621 lisämyöhästymisminuuttia. Näistä viivekerroista 8 045 tapausta olivat sellaisia, joissa seuraavalla osuudella junat eivät saaneet syntynyttä viivettä kiinni, vaan jäivät jopa lisää myöhään (yhteensä 14 713 minuutin edestä). Toisin sanoen 60,2 % tapauksista juna ei saanut kärsimäänsä M1- tai M6-viivettä kiinni, vaan saapui seuraavalle asemalle yhtä myöhässä tai jätään jopa lisää myöhään.

5309 junaa kuitenkin ottivat seuraavalla rataosalla viivettään kiinni, yhteensä 14 106 minuutin edestä. Näin ollen on totta, että *keskimäärin* junat eivät M-syyn jälkeen jää lisää myöhään (14 713 – 14 106 = 607 minuuttia). Tämä ei kuitenkaan kerro, onko tilanne sama kaikilla rataosilla. On tarpeen tutkia liikennepaikka- ja rataosakohtaisesti, missä viiveitä syntyy lisää ja missä niitä saadaan kiinni. Tätä varten on laadittu taulukko 4.

Taulukossa on esitetty:

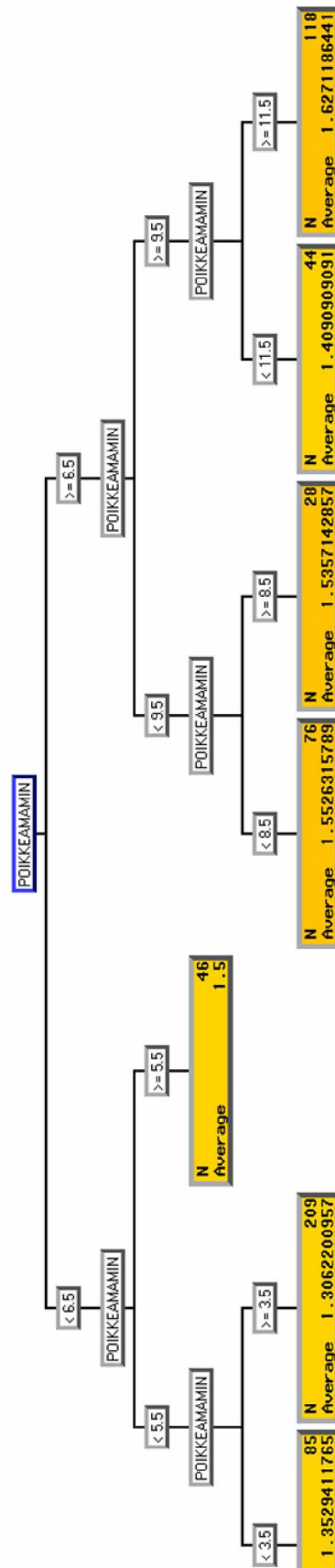
- M-viivettä seurannut rataosa
- Niiden junien määrä, jotka ylipäätään ovat saaneet liikennepaikalla M-viiveen
- Niiden junien määrä, jotka eivät saaneet syntynyttä M-viivettä kiinni
- Prosentuaalinen osuus junista, jotka eivät saaneet M-viivettä kiinni
- Osuus kaikista M-viiveistä saaneista junista
- Noste (merkittävyyskerroin), joka ottaa huomioon prosenttiosuuden niistä junista, jotka eivät saa viivettä kiinni sekä suhteellisen osuuden kaikista M-viiveistä saaneista junista. Taulukko on järjestetty nosteen mukaan.

Taulukko 4. Rataosakohtainen tarkastelu matkustajaruuhkan jälkeisten rataosien lisäviiveistä.

M-viivettä seurannut rataosa	M-viiveitä yhteensä	Ei saa kiinni syntynyttä M-viivettä	Ei saa kiinni M-viivettä / M-viiveet yht. (%)	Paljonko kaikista M-viiveistä (%)	Noste
TKL-KE	606	400	66,01 %	3,00 %	0,01977
LPV-HPL	499	282	56,51 %	2,11 %	0,01193
EPO-LPV	266	197	74,06 %	1,48 %	0,01093
LH-TKL	408	240	58,82 %	1,80 %	0,01057
EPO-KKN	148	142	95,95 %	1,06 %	0,0102
TKL-HKI	362	215	59,39 %	1,61 %	0,00956
OLK-HKI	221	166	75,11 %	1,24 %	0,00934
OLK-TKL	187	152	81,28 %	1,14 %	0,00925
TKL-OLK	182	134	73,63 %	1,00 %	0,00739
LPV-EPO	278	157	56,47 %	1,18 %	0,00664
TKL-LH	254	143	56,30 %	1,07 %	0,00603
TKL-RI	312	156	50,00 %	1,17 %	0,00584
HL-TL	445	185	41,57 %	1,39 %	0,00576
HL-RI	160	106	66,25 %	0,79 %	0,00526
KV-LÄ	96	78	81,25 %	0,58 %	0,00475
PKO-SK	215	111	51,63 %	0,83 %	0,00429
KKN-EPO	68	61	89,71 %	0,46 %	0,0041
LH-KV	234	104	44,44 %	0,78 %	0,00346
TL-TPE	446	142	31,84 %	1,06 %	0,00339
TL-HL	276	107	38,77 %	0,80 %	0,00311
KE-MLÄ	133	74	55,64 %	0,55 %	0,00308

SLO-TKU	89	60	67,42 %	0,45 %	0,00303
KE-TKL	388	122	31,44 %	0,91 %	0,00287
KE-RI	80	54	67,50 %	0,40 %	0,00273
EPO-KLH	43	38	88,37 %	0,28 %	0,00251
HPL-VKS	87	54	62,07 %	0,40 %	0,00251
SK-PKO	106	57	53,77 %	0,43 %	0,0023
RI-TKL	113	58	51,33 %	0,43 %	0,00223
PAR-JNS	181	73	40,33 %	0,55 %	0,0022
HPL-HKI	84	49	58,33 %	0,37 %	0,00214
MLÄ-LH	261	86	32,95 %	0,64 %	0,00212
LM-TKU	60	40	66,67 %	0,30 %	0,002
LR-IMR	132	58	43,94 %	0,43 %	0,00191
HPL-LPV	35	29	82,86 %	0,22 %	0,0018
MI-KV	93	46	49,46 %	0,34 %	0,0017
YV-OL	64	38	59,38 %	0,28 %	0,00169
KV-LH	110	48	43,64 %	0,36 %	0,00157
JÄS-TPE	39	28	71,79 %	0,21 %	0,00151
KOK-YV	81	39	48,15 %	0,29 %	0,00141
LM-TL	58	32	55,17 %	0,24 %	0,00132
HKI-HPL	21	19	90,48 %	0,14 %	0,00129
TL-LM	64	32	50,00 %	0,24 %	0,0012
JY-JÄS	17	16	94,12 %	0,12 %	0,00113
MLÄ-KE	162	47	29,01 %	0,35 %	0,00102
JY-PM	47	25	53,19 %	0,19 %	0,001
KR-SLO	44	24	54,55 %	0,18 %	0,00098
TPE-TL	37	22	59,46 %	0,16 %	0,00098
KR-KKN	16	14	87,50 %	0,10 %	0,00092
KKN-KR	51	24	47,06 %	0,18 %	0,00085
IMR-LR	120	36	30,00 %	0,27 %	0,00081
ILM-KUO	21	15	71,43 %	0,11 %	0,0008
LM-LM	16	13	81,25 %	0,10 %	0,00079
EPO-HKI	28	17	60,71 %	0,13 %	0,00077
MI-PM	115	34	29,57 %	0,25 %	0,00075
RI-HL	113	33	29,20 %	0,25 %	0,00072
KEM-KLJ	35	18	51,43 %	0,13 %	0,00069
PAR-IMR	59	23	38,98 %	0,17 %	0,00067
IMR-PAR	82	27	32,93 %	0,20 %	0,00067
LR-LÄ	193	41	21,24 %	0,31 %	0,00065
KOK-SK	97	28	28,87 %	0,21 %	0,00061
KV-MI	25	14	56,00 %	0,10 %	0,00059
LH-RI	42	17	40,48 %	0,13 %	0,00052
TPE-PKO	42	17	40,48 %	0,13 %	0,00052
OL-KEM	64	20	31,25 %	0,15 %	0,00047
ILM-KAJ	66	20	30,30 %	0,15 %	0,00045
HKI-TKL	28	13	46,43 %	0,10 %	0,00045
KON-KAJ	17	10	58,82 %	0,07 %	0,00044
TKU-TUS	11	8	72,73 %	0,06 %	0,00044
TPE-JÄS	8	6	75,00 %	0,04 %	0,00034
HKI-OLK	18	9	50,00 %	0,07 %	0,00034
PM-KUO	141	25	17,73 %	0,19 %	0,00033
PKO-TPE	55	15	27,27 %	0,11 %	0,00031
KEM-ROI	48	14	29,17 %	0,10 %	0,00031
SK-KOK	92	19	20,65 %	0,14 %	0,00029
SLO-KR	8	5	62,50 %	0,04 %	0,00023
KUO-ILM	16	7	43,75 %	0,05 %	0,00023
YV-KOK	35	10	28,57 %	0,07 %	0,00021
KEM-OL	30	9	30,00 %	0,07 %	0,0002
OL-YV	14	6	42,86 %	0,04 %	0,00019
PM-MI	41	10	24,39 %	0,07 %	0,00018
KON-OL	46	10	21,74 %	0,07 %	0,00016
HPK-SK	2	2	100,00 %	0,01 %	0,00015
RI-KE	5	3	60,00 %	0,02 %	0,00013

Nosteen perusteella vaikuttaa siltä, että mielenkiintoisimman tarkastelukohteen muodostavat junat, jotka lähtevät Tikkurilasta kohti Keravaa: M-viiveitä syntyy paljon, eikä niitä saada seuraavalla rataosalla kiinni. 606 tapauksesta 400 (66 %) on sellaisia, joissa juna ei saa kurottua tapahtunutta viivettä minuuttiakaan kiinni.



Kuva 47.

Tikkurilan matkustajapalvelusta johtuvien lisäyöhästymisten suuruuden luokittelu asemalle saapuneen junan kokonaisviiveen suuruuden suhteen Tikkurilasta Keravalle matkaavilla junilla.

Jotta Tikkurila-Kerava-välin tilanteeseen päästään paremmin kiinni, tarkasteltiin sitä luokitteluanalyysin keinoin (kuva 47). Analyysissä tutkitaan, kuinka paljon Tikkurilaan saapuva juna saa M-viiveminuutteja sen suhteen, kuinka paljon juna on ollut myöhässä saapuessaan asemalle. Kuvasta havaitaan, että mikäli saapuva juna on 6 minuuttia tai enemmän myöhässä, kasvaa matkustajapalvelusta johtuvien viiveiden kesto. Tämä johtuu varmasti siitä, että liikenteen ollessa sekaisin omaa junaa on vaikeampi löytää ja lisäksi osa junista ruuhkautuu tavallista enemmän. Poikkeuksena 10 ja 11 minuutin kokonaisviiveet, jolloin matkustajapalvelusta johtuvien viiveiden keskimääräinen kesto hieman pienenee. Tämä johtunee ainakin osin lähijunien 10 minuutin vuorovälistä (liikenne on kokonaisen ”vaiheen” jäljessä, jolloin tilanne on matkustajien kannalta lähes normaali), mutta asiaa olisi mielenkiintoista tutkia tarkemmin.

Tikkurila-Kerava-välillä liikennöivät sekä kauko- että lähijunat. Lisäksi liikenne on intensiivistä ja välit lyhyitä. Nämä seikat varmasti vaikuttivat M-viiveisiin liittyviin ilmiöihin ja siten analyysin tuloksiin. Jatkossa olisikin mielenkiintoista tarkastella jotakin puhdasta kaukoliikenneväliä. Taulukon 4 perusteella esimerkiksi Jyväskylä-Jämsä-Tampere-välillä M-viiveitä ei pääsääntöisesti saada ajettua kiinni.

Taulukosta 4 kannattaa myös huomioida ne yhteysvälit, joilla pienet M-viiveet onnistutaan pääsääntöisesti ajamaan kiinni. Esimerkiksi Pieksämäeltä Kuopioon päin lähettäessä M-viiveitä on usein (141 kertaa), mutta Kuopioon saavuttaessa viive on useimmiten (82 % tapauksista) saatu ajettua kiinni. Tätä selittää yhteisvälin pituus ja siten suuri pelivara. Tästä tosin herää kysymys, että pitäisikö lyhyemmällä väleillä olla suhteessa suuremmat pelivarat?

4.3.3 Päätelmät

Matkustajaruuhkasta ja pysähtymisajan ylityksestä johtuvat lyhyet viiveet jäävät useammin elämään tai jopa kertaantuvat sen sijaan että ne saataisiin kiinni seuraavalla rataosalla. Datan tarkemman tarkastelun avulla on mahdollista tutkia, millä rataosilla tämä viiveiden kertaantuminen korostuu ja millä rataosilla tilanne ei ole niin paha. Tätä analyysiä voidaan suorittaa myös minuuttimääräisesti.

Analyysissä tehtyä datanmuokkausta voidaan hyödyntää muihinkin tarkoituksiin. Datan pohjalta voidaan tarkastella esimerkiksi sitä, millä rataosilla junien viiveet lisääntyvät, millä niitä puolestaan saadaan aidosti ajettua kiinni. Saatuja tietoja voidaan hyödyntää esimerkiksi aikataulurakennetta tai ennakoivaa toimintaa kehitettäessä.

4.4 Helsingin alueen ruuhkaisuus lähiliikenteen näkökulmasta

4.4.1 Väite analyysin taustalla

”Suuri osa junaliikenteen ongelmista syntyy Helsingin aseman läheisyydessä, erityisesti ruuhka-aikaan. Kärsijänä on yleensä lähiliikenne.”

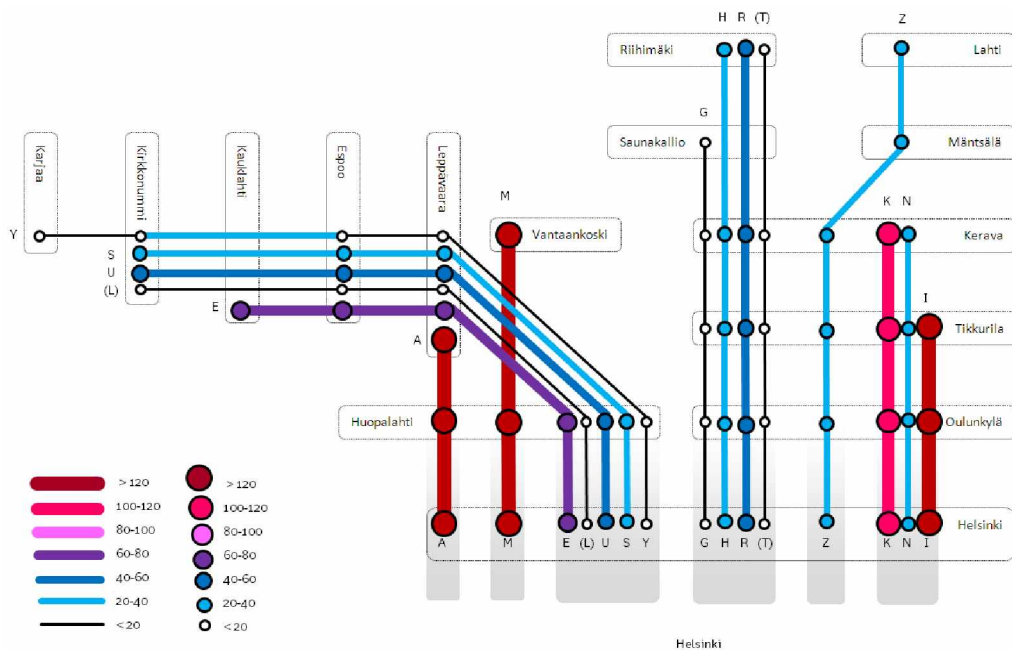
4.4.2 Analyysi

Helsingin seudun lähijunaliikenne muodostaa kompleksisen kokonaisuuden, jonka häiriöherkkyyttä lisää se, että se hyödyntää osin samaa rataverkkoa kuin muu junaliikenne. Ongelmallisin paikka vaikuttaa olevan Helsingin ratapiha Helsingin päärauta-

tieaseman ja Pasilan aseman välillä. Kaikki junat käyvät kääntymässä Helsinkiin päättyvällä ratapihalla, minkä seurauksena ratapihalla on valtavasti liikennettä. Tämän liikenteen analysointi nykydatalla on kuitenkin vaikeaa, koska se on melko epätarkkaa, eikä esimerkiksi Pasilan asemalta saada lainkaan seurantatietoja. Tästä huolimatta tässä analyysissä on pyritty nostamaan esille liikenteen säännönmukaisuuksia.

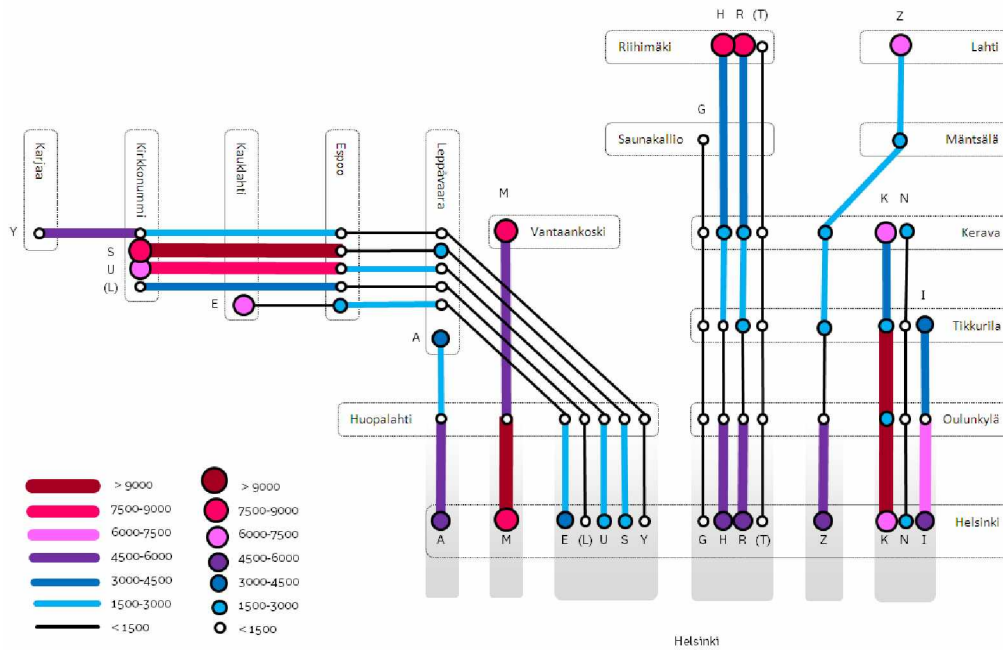
Analyysissä käytetty data kattaa vuoden 2010. Analyysin erikoisuutena on, että siinä hyödynnettiin JUSE-datan sijasta GRATU-dattaa. Näin junien kulkutietoihin päästiin liittämään kellonaikatiedot, mikä antoi mahdollisuuden tarkastella erikseen esimerkiksi ruuhka-ajan liikennettä. Analyysissä tehdyt datanmuokkaukset ovat samankaltaisia kuin aiemmissa analyyseissä, eikä niitä siksi esitellä sen tarkemmin.

Analyysin ensimmäisessä vaiheessa luotiin karttapohja, jolla havainnollistetaan eri linjoja ja niiden liikennemääriä (kuva 48). Rantaradalla erityisesti Helsingin ja Leppävaaran välillä liikennöinti on vilkasta. Pääradalla liikenne on vilkasta erityisesti Tikkurilaan asti. Myös Helsinki–Vantaankoski-rataosalla on paljon liikennettä. Näillä vilkkaimmilla yhteysväleillä merkittävä osa liikenteestä kulkee erillisiä kaupunkiratoja.



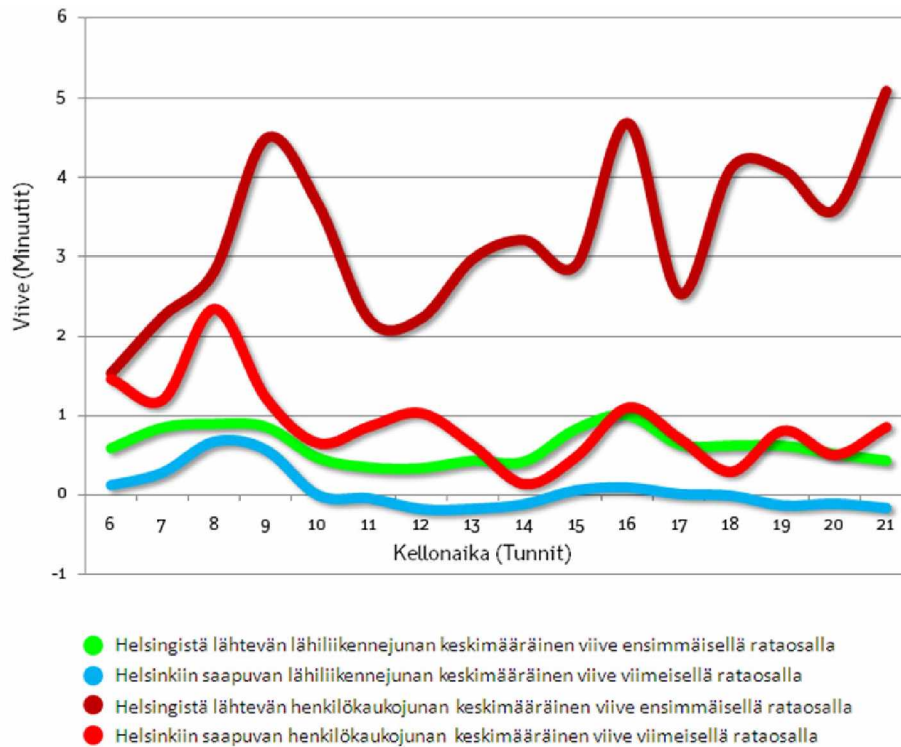
Kuva 48. Helsingin seudun lähijunaliikenteen linjat ja niiden arkipäivien juna-
määrät keväällä 2011. G-linja on lopetettu elokuussa 2011.

Analyysin seuraavassa vaiheessa tarkasteltiin, missä yhteysväleillä ja liikennepaikoilla syntyy eniten lähijunaliikenteen myöhästymisiä. Nämä tiedot mallinnettiin edellisessä vaiheessa luodulle karttapohjalle (kuva 49). Kuvan perusteella voidaan todeta, ettei Helsinki tai sen lähiympäristö ole missään nimessä ainoa ongelma-alue. Viiveitä syntyy paljon esimerkiksi Riihimäen lähtöjen ja kääntöjen yhteydessä. Lisäksi Espoon ja Kirkkonummen välillä syntyy paljon myöhästymisiä. Näyttää jopa siltä, että liikennemääriin verrattuna Helsingin päässä syntyy suhteellisen vähän viiveitä.



Kuva 49. Helsingin seudun lähijunaliikenteen lisämyöhästymiset (minuuttia) eri linjoilla asemittain ja asemaväleittäin vuonna 2010.

Seuraavaksi siirryttiin tarkastelemaan Helsingin asemalta lähtevän ja sinne saapuvan lähijunaliikenteen viiveiden muodostumista Helsingin ratapihalla eri kellonaikoina (kuva 50). Vertailun vuoksi kuvassa on mukana myös kaukoliikenne. Koska Pasilasta ei kulkutietoja ole, on tarkastelu tehty Oulunkylän ja Huopalahden kulkutiedoilla.



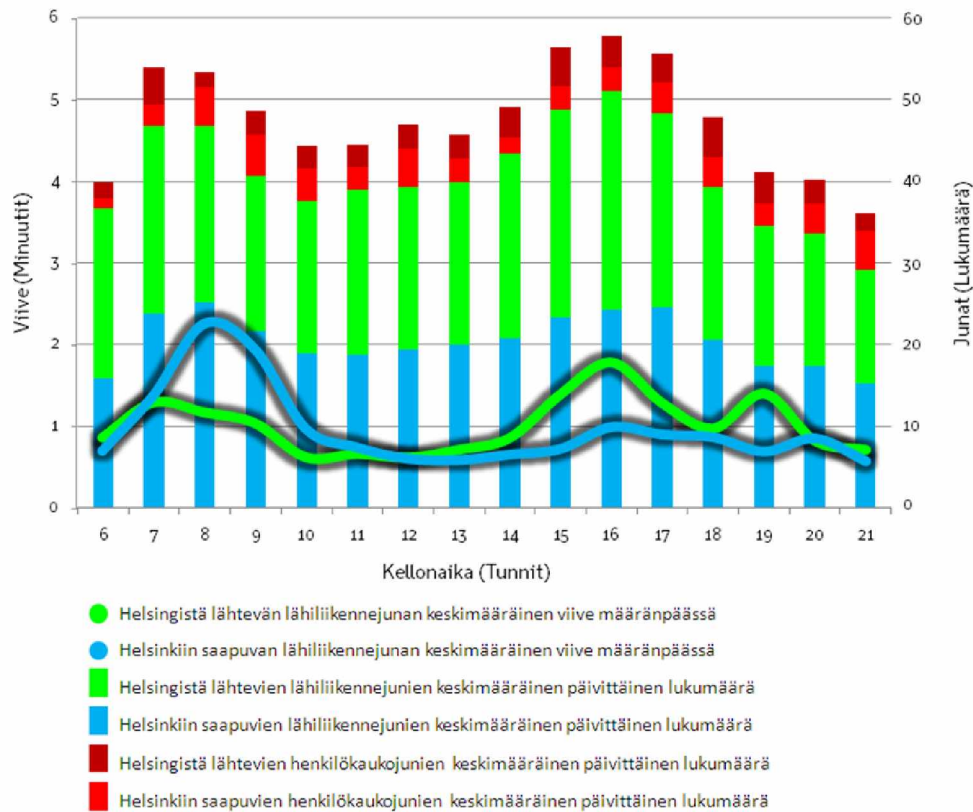
Kuva 50. Helsingistä lähtevien ja sinne saapuvien junien keskimääräiset viiveet vuonna 2010 rataosilla Helsinki–Oulunkylä ja Helsinki–Huopalahti.

Kuvasta voidaan päätellä seuraavaa:

- Helsinkiin saapuvat lähijunat eivät keskimäärin myöhästy yhtään lisää Oulunkylän/Huopalahden jälkeen. Poikkeuksen tekee klo 8-9 aamuruuhka, joskin silloinkin lisämyöhästymisen tällä välillä on keskimäärin alle minuutin.
- Helsingistä lähtevät lähijunat myöhästyvät keskimäärin alle minuutin ensimmäisellä rataosallaan. Suurimmat viiveet, keskimäärin minuutin suuruiset, syntyivät kello 16 aikaan.
- Helsinkiin saapuvat kaukojunat myöhästyvät Oulunkylän/Huopalahden jälkeen keskimäärin noin minuutin lisää. Ehdottomasti suurin keskimääräinen lisäviive, reilut 2 minuuttia, syntyy klo 8 aamuruuhkan aikaan.
- Suurimmat viiveet näyttävät syntyvän Helsingistä lähteville kaukojunille. Junat myöhästyvät Helsingin ja Oulunkylän/Huopalahden välillä kellonajasta riippuen keskimäärin kahdesta viiteen minuuttia. Suurimmat myöhästymiset syntyvät aamu- ja iltapäiväruuhkien aikaan sekä myöhään illalla.

Yhteenvedon voidaan todeta, että minuuttimääräisesti kaukoliikenne myöhästelee tarkasteluvälillä huomattavasti lähiliikennettä enemmän. Tämä tosin johtuu ainakin osittain siitä, että lähiliikenteessä juna perutaan huomattavasti herkemmin, mikäli se on jäämässä myöhään. Näin ollen ajettujen junien viiveet pysyvät matalina. On myös muistettava, että kaukoliikenteellä täsmällisyysvaatimus ei ole niin suuri kuin lähiliikenteessä. Erityisesti Helsingistä lähtevien kaukojunat myöhästely on kuitenkin siinä määrin suurta, että sen syitä ja vaikutuksia tulisi selvittää tarkemmin.

Lopuksi tarkasteltiin, kuinka eri kellonaikojen kokonaisliikennemäärät heijastuvat lähijunaliikenteen täsmällisyyteen (kuva 51). Edellisen tarkastelun perusteella lähijunien viiveet Helsinki–Oulunkylä- ja Helsinki–Huopalahti-väleillä olivat todella pieniä. Näin ollen tarkasteluun otettiin junien määräasemakohtaiset viiveet.



Kuva 51. Helsingistä lähtevien ja sinne saapuvien junien keskimääräiset päivittäiset lukumäärät sekä Helsingin seudun lähiliikenteen junien viiveiden keskimääräiset suuruudet määränpäässä.

Kuvasta voidaan päätellä seuraavaa:

- Liikennemäärät ovat luonnollisesti suurimmillaan klo 7–9 (aamuruuhka) ja klo 15–17 (iltapäiväruuhka).
- Helsinkiin saapuvien junien keskimääräinen viive on noin minuutin suuruisen. Kello kahdeksan ja yhdeksän välillä, se on huomattavasti suurempi, noin yli kaksi minuuttia. On tärkeä huomata, että viiveitä syntyy vasta noin tunnin sen jälkeen, kun liikennemäärät ovat nousseet huippuunsa.
- Helsingistä lähtevien junien keskimääräinen viive on myös noin minuutin suuruisen. Iltapäiväruuhkan aikaan se nousee vajaaseen kahteen minuuttiin.

Yhteenvedon voidaan todeta, että junan keskimääräinen viive on sitä suurempi, mitä ruuhkaisempaan aikaan se liikennöi. Ruuhkajunissa matkustaa eniten ihmisiä. Näin ollen matkustajakohtainen täsmällisyys lienee huonompi kuin täsmällisyyden peruseurannassa käytettävä junien kokonaistäsmällisyys. Vaikka keskimääräiset viiveet, alle minuutista kahteen, voivat vaikuttaa vähäisiltä, ovat ne kuitenkin merkittäviä liikenteen täsmällisyysvaatimukset huomioon ottaen: Suomessa tavoitteena on, että 97,5 % junista olisi alle 3 minuuttia myöhässä. Lisäksi tulee ottaa huomioon, ettei keskimääräinen viive kerro koko totuutta: valtaosa junista on minuutilleen aikataulussa, ja vastaavasti ongelmallisimmat junat useita minuutteja myöhässä. Jatkossa tarkastelu olisikin hyvä kohdistaa näihin häiriikköjuniin.

4.4.3 Päätelmät

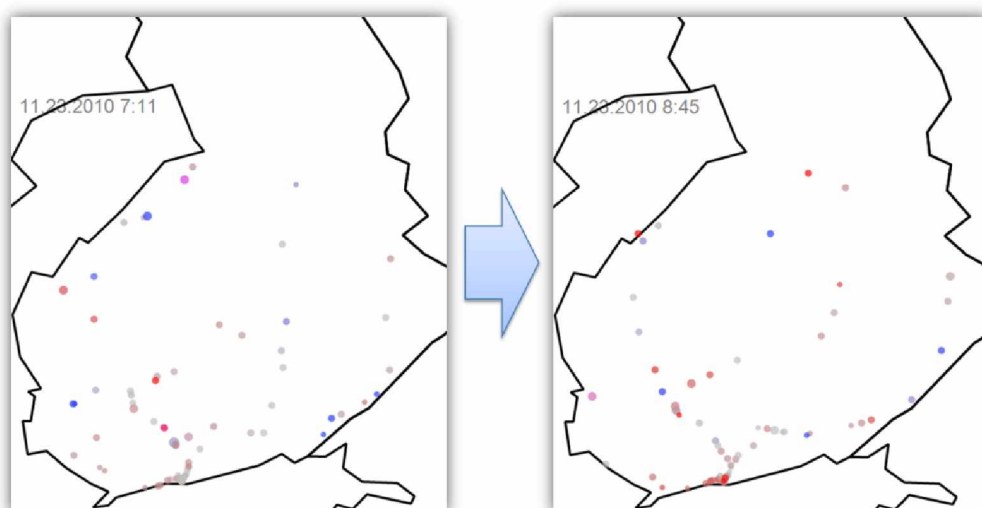
Tehty analyysi antoi mielenkiintoisen, uuden näkökulman Helsingin seudun lähijunaliikenteen ongelmiin. Ensinnäkin, Helsingin ratapihan lisäksi myös monet muut

alueet näyttävät aiheuttavan liikenteelle ongelmia. Toiseksi, kaukoliikenne vaikuttaa kärsivän alueella jopa suuremmista ongelmista kuin lähiliikenne. Tätä tulosta tosin saattaa vääristää se, että analyysissä ei otettu huomioon peruttuja lähijunia: lähiliikenteessä ongelmat johtavat junan perumiseen huomattavasti kaukoliikennettä helpommin. Kolmanneksi, lähiliikenteen viiveet näyttävät kytkeytyvän vahvasti liikennemääriin. Neljänneksi, viiveminuuttien keskiarvot saattavat esittää tilanteen turhan hyvässä valossa. Niinpä analyysi kannattaisi kohdistaa ongelmallisimpiin juniin.

Analyyysi osoitti, että itsestäänselvyyksiäkin kannattaa analysoida kriittisesti ja ennen kaikkea laajemmassa kontekstissa, jotta asiat saavat oikeat mittasuhteet. Lisäksi GRATU-datan hyödyntäminen antoi mahdollisuuden nostaa esille kaksi uutta viiveitä selittävää tekijää: liikennemäärät ja ajankohdan. Tähänastisissa analyyseissä niitä ei ole juurikaan huomioitu, vaikka ainakin tämän analyysin perusteella syytä olisi.

4.5 Viiveiden visualisointi karttapohjaisella animaatiolla

Voimakkaat syy-seuraussuhteet ovat luonteenomaisia rautatieliikennejärjestelmälle. Toisin sanoen yksittäisen junan kulku vaikuttaa voimakkaasti toisiin juniin. Lisäksi järjestelmän tapahtumat ovat luonnollisesti erittäin paikkasidonnaisia. Näistä ominaisuuksista seuraa se, että järjestelmän tapahtumia olisi havainnollista tarkastella sekä *kronologisesti* että *karttapohjaisesti*. Tämän ajatuksen pohjalta toteutettiin demo, jossa yksittäisen päivän junien kulku esitettiin kartta-animaationa. Junat esitettiin palloina, joiden väri muuttui sen mukaan, kulkivatko ne etuajassa, ajallaan vai myöhässä. Animaatiota varten junien seurantapisteiden koordinaatit sijoitettiin karttapohjalle. Itse animaatio tehtiin sijoittamalla junien kulkutietoaineisto tälle karttapohjalle SAS:in JMP-työkalua hyödyntäen. Pohjatyön jälkeen animaation tekeminen oli hyvin nopea toimenpide. Animaatiota on havainnollistettu kuvassa 52.



Kuva 52. Junien kulun ja viiveiden visualisointi karttapohjaisella animaatiolla. Vasen kuvakaappaus on ajanhetkeltä 7.11, oikea 8.45. Kuvista näkyy, kuinka viiveet kehittyvät esim. Helsingin ja Tampereen seudulla.

Animaatio esitettiin useassa eri tilaisuudessa saaden hyvän vastaanoton. Esitystavan vahvuuksiksi koettiin ainakin:

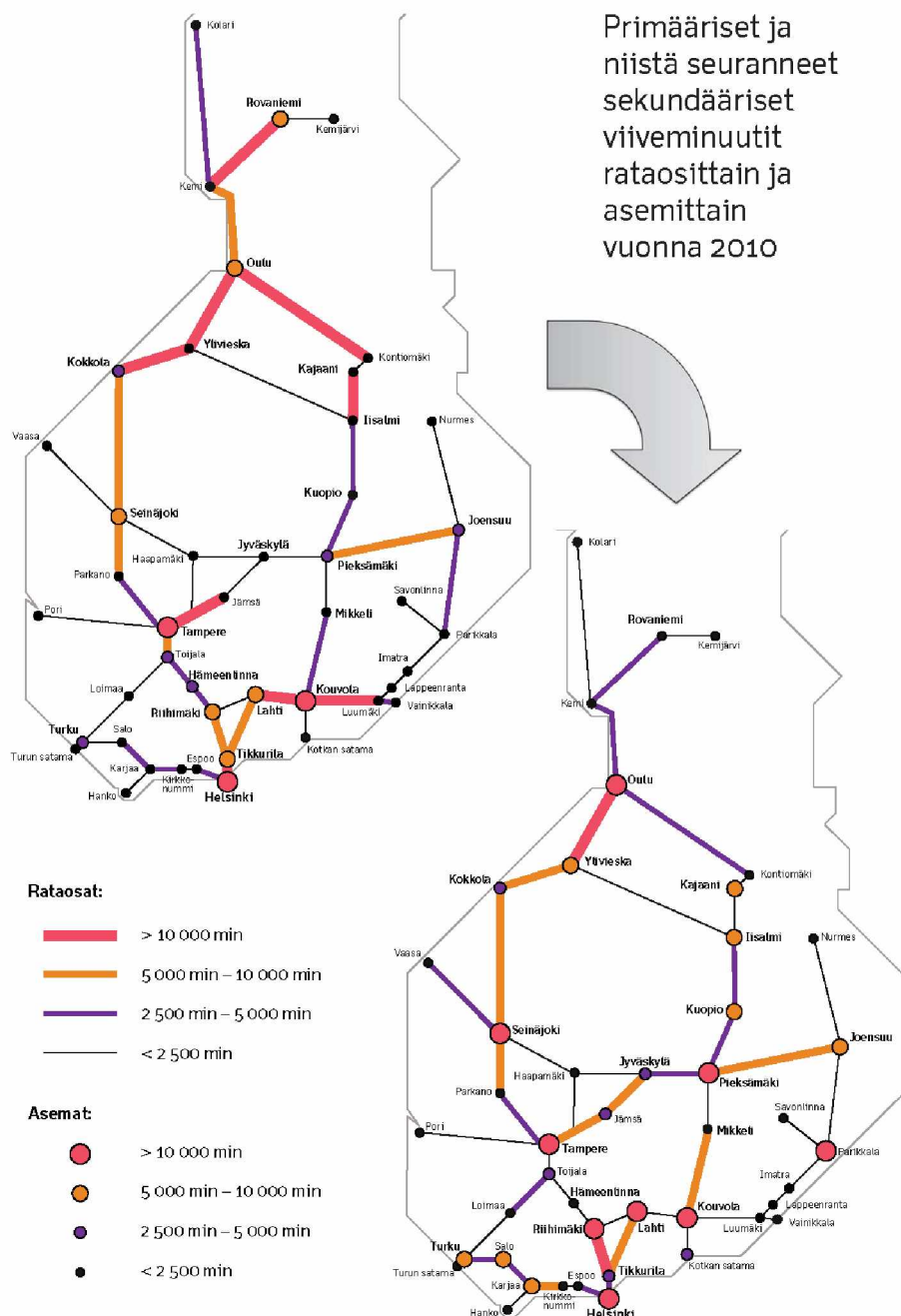
- Häiriöiden maantieteellinen sijoittuminen on helppo havaita
- Häiriöiden laajeneminen ja vaimeneminen (ts. viiveiden ketjuuntuminen) käyvät konkreettisesti ilmi
- Häiriöiden ajankohdat tulevat havainnollisesti esiin
- Yleiskuvan saaminen on helppoa ja nopeaa.

Näiden kokemusten perusteella karttapohjaisia animaatioita kannattaisi jatkossa hyödyntää erityisesti silloin, kun tarvitaan nopea kokonaiskuva tietyn alueen tietyn ajanjakson tapahtumista.

4.6 Vuosiraportin analyysien kehittäminen

Tutkimukseen liittyi myös Liikenneviraston *Rautatieliikenteen täsmällisyys 2010* -julkaisun (Liikennevirasto 2011) analyysiosuuksien laadinta. Pääasiassa osuus toteutettiin edellisen vuoden esimerkkejä hyödyntäen. Joitakin parannuksia analyysiin kuitenkin tehtiin. Erityisesti primääristen ja sekundääristen viiveiden välistä suhdetta pyrittiin avaamaan eri tavoin. Valtakunnallista viiveiden ketjuuntumista havainnollistettiin muun muassa kuvalla 53. Lisäksi tuotiin esille kaupunkiratojen sekaliikenneraiteita parempi täsmällisyys.

Vaikkei vuosiraportin analyysiosuudet tarjonneet analysoinnin kannalta mitään erityisen mielenkiintoista, oli niiden tekeminen tutkimuksen kannalta tärkeää. Niiden avulla oli hyvä perehtyä tarkasteltavaan dataan ja siitä tällä hetkellä tehtäviin perusanalyysiin.



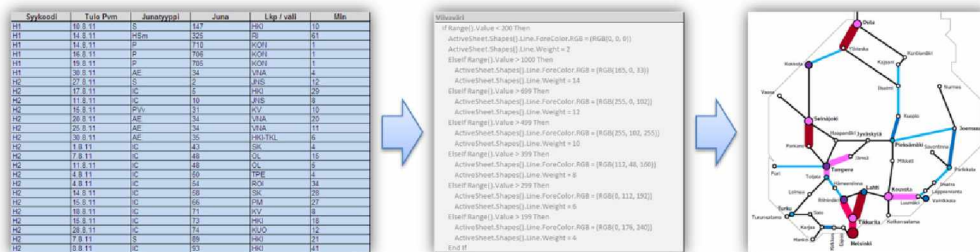
Kuva 53. Henkilökaukoliikenteen primääriset (vasemmalla ylhäällä) ja niistä seuranneet sekundääriset (oikealla alhaalla) viiveminuutit rataosittain ja asemittain vuonna 2010 (Liikennevirasto 2011).

4.7 Kuukausiraportoinnin automatisointi

Tutkimuksen aikana havaittiin, että täsmällisyysdatan analysoijilta kuluu paljon aikaa ns. rutiiniraportointiin. Toisin sanoen samankaltaisina toistuvia kuukausiraportteja varten joudutaan tekemään kerta toisensa jälkeen samat manuaaliset rutiinitoimenpiteet, kuten datanmuokkaukset ja kuvaajien laadinnat. Tämän vuoksi tutkimuksessa toteutettiin Microsoft Excel -pohjainen työkalu (kuva 54), joka tuottaa kuukausirapor-

tin automaattisesti kulkutietodatan perusteella. Työkalu tuottaa mm. seuraavat kuvaajat ja visualisoinnit:

- Kuukauden yhteenvetotaulukot eri junalajien osalta
- Karttavisualisoinnit primäärisistä, sekundäärisistä ja radanpidollisista myöhästymisminuuteista rataosittain ja liikennepaikoittain
- Kunnossapitoaluekohtaiset viiveet



Kuva 54. Kuukausiraportoinnin automatisointityökalu tuottaa JUSE-datan pohjalta erilaisia kuvaajia ja visualisointeja.

Työkalun ansiosta rutiiniraportoinnilla vapautuu aikaa ad hoc -tarpeisiin ja erilaisille syvällisemmille analyyseille. Lisäksi työkalun myötä rutiiniraportteihin oli mahdollista sisällyttää sellaista sisältöä (viiveiden karttapohjainen visualisointi), jonka laadinta manuaalisesti on ollut liian työlästä.

Työkalu osoittaa, ettei kehittyneen data-analytiikka välttämättä edellytä uusien työkalujen hankintaa, vaan jo perinteisillä toimisto-ohjelmilla voidaan tehdä perinteistä monipuolisempia analyyseja ja automatisointeja.

5 Päätelmät ja suositukset

Tällä hetkellä rautatietojen tekijät tekevät systemaattista täsmällisyystyötä useilla eri foorumeilla. Analytiikan hyödyntäminen päätöksenteossa on kuitenkin keskittynyt kuukausittain laadittaviin perusanalyysihin. Kehittyneitä data-analytiikka-menetelmiä hyödyntämällä olisi mahdollisuus tuottaa huomattavasti nykyistä parempaa ja syvällisempää tietoa päätöksenteon tueksi. Tämä kuitenkin edellyttää kehittämistä niin datan ja dataympäristön laadun, analyysitoiminnan kuin tuotetun tiedon hyödyntämisenkin osalta.

5.1 Kehittyneen data-analytiikan mahdollisuudet

Kehittyneen data-analytiikan menetelmin laadittujen analyysien tulokset olivat varsin lupaavia. Menetelmiä monipuolisesti hyödyntämällä rautatieliikennejärjestelmän ilmiöitä on mahdollista tunnistaa ja analysoida nopeasti, havainnollisesti ja monipuolisesti.

Jo yksittäistä junaa tarkastellessa (luku 4.1.) päästiin hyödyntämään monia eri menetelmiä. *Luokittelun* avulla tarkasteltiin, mitkä tekijät olivat valitun junan tapauksessa merkittävimpiä viiveiden aiheuttajia. *Klusterointi* puolestaan antoi mahdollisuuden datan automaattiseen jäsentämiseen ilman ennakkotietoja. Menetelmä nosti esille esimerkiksi eri ajanjaksojen keskimääräisten viiveminuuttien suuruudet. *Sekvenssi-analyysin* potentiaali osoittautui erittäin suureksi, sillä rautatiejärjestelmä sisältää paljon sisäisiä riippuvuuksia ja tapahtumaketjuja – sekvenssejä. Näiden ketjujen tunnistamiseen ja analysointiin sekvenssianalyysi on paras mahdollinen työkalu.

Pohjanmaan radan routavaurioiden myötä päästiin tutkimaan (luku 4.2.), miten yksittäisen ilmiön aiheuttamat viiveet ketjuuntuvat eri puolille Suomen rautatieliikennejärjestelmää ja mitkä viiveet ylipäättään ovat ketjuuntuneita. Tämän lisäksi tutkittiin, miten viiveminuuttien suuruudet vaimenevat viiveasteen mukaan sekä miten, miksi ja missä eri junat aiheuttavat viiveitä toisilleen. Myös tässä analyysissä sekvenssitarkastelu näytti voimansa: sen avulla viiveiden erilaiset ketjuttamiset onnistuvat helposti. Tehty analyysi osoitti, kuinka menetelmien avulla on mahdollista tuottaa nopeasti syvällistä tietoa kokonaisjärjestelmän toiminnasta.

M-syistä johtuvien pienien viiveiden vaikutusten tarkastelu (luku 4.3.) osoitti, kuinka helposti ja nopeasti erilaisten oletusten todenperäisyys pystytään osoittamaan. Analyysi osoitti, että vaikka lyhyet M-syistä johtuneet viiveet keskimäärin saadaankin ajettua kiinni, ei tämä onnistu kaikilla rataosilla. Tällaista tietoa voidaan hyödyntää esimerkiksi aikataulurakennetta tai ennakoivaa toimintaa kehitettäessä.

Myös Helsingin seudun lähiliikennettä käsittelevä analyysi (luku 4.4.) osoitti, että itsestäänselvyyksiäkin kannattaa analysoida kriittisesti ja ennen kaikkea laajemmassa kontekstissa, jotta asiat saavat oikeat mittasuhteet. Tärkeä havainto oli myös se, että erilaiset viiveminuuttien keskiarvot eivät välttämättä kuvaa tarkasteltavia ilmiöitä riittävällä tavalla, vaan saattavat jopa johdattaa harhaan. GRATU-datan hyödyntäminen nosti esille kaksi uutta viiveitä selittävää tekijää: liikennemäärät ja ajankohdan.

Erilaiset visualisoinnit osoittautuivat varsin tehokkaaksi tavaksi havainnoida ja vertailla datan sisältöä. Tulevaisuudessa tulisikin kiinnittää kaikissa visualisoinneissa huomiota siihen, että ne ovat tarkoitukseen sopivia sekä mahdollistavat vertailun eri tarkastelujaksojen välillä niin, että poikkeamat ovat helposti havaittavissa. Myös erilaisten animointien hyödyntämistä on syytä lisätä.

Menetelmiä hyödyntämällä analysointi on tehokasta ja nopeaa. Näin aikaa jää analyysien tulkinnaalle. Tähän tutkimukseen valittiin analysoitavaksi loppujen lopuksi melko selkeitä, helposti tunnistettavia ongelmia: esimerkiksi Pohjanmaan radan rautavauriot aiheuttivat siinä määrin suuria liikenteellisiä ongelmia, että niiden tunnistaminen ja perustason analysointi olisi onnistunut ilman kehittyneitä data-analytiikkaakin. Tällaisia ongelmia valitsemalla varmistuttiin siitä, että tulokset muodostuvat mielekkäiksi. Nyt, kun menetelmien toimivuus on osoitettu, voidaan niitä ryhtyä hyödyntämään myös piilevämpien ilmiöiden tunnistamiseen ja analysointiin. ***Tutkimuksen tärkein ansio ei siis ole tehtyjen analyysien varsinaiset tulokset, vaan tapa, jolla niihin päästiin.***

5.2 Data- ja analytiikkaympäristön kehittäminen

Data- ja analytiikkaympäristöä analytiikan näkökulmasta tarkastellessa nousee esiin joitakin kehityskohteita. Analytiikka ei tähän asti ole ollut organisoitu toiminto, vaan analyysiä ovat tehneet muutamat asiantuntijat omiin ja muiden tarpeisiin sillä datalla, mitä on ollut saatavilla; tietojärjestelmät eivät tehokasta analytiikkaa juuri tue.

Määriteltäessä uusia järjestelmiä tulisi niissä analytiikan näkökulmasta ottaa seuraavat asiat huomioon:

- Mitä kokonaisvaltaisempaa ja kattavampaa tietoa järjestelmistä saadaan, sitä laadukkaampia analyysejä on mahdollista tehdä. Tällöin voidaan analyysin keinoin etsiä ja mallintaa ongelmakohtia ja niiden toimintaa hyvinkin tarkasti. (ks. luku 5.2.1)
- Kerättävän datan laatuun pitäisi kiinnittää huomiota. Epäluotettavasta datasta on mahdotonta tehdä luotettavia analyysejä. (ks. luku 5.2.1)
- Analytiikkaa varten tulisi luoda oma tietojärjestelmäympäristö, joka koostuisi omasta tietokantajärjestelmästä sekä dataintegraatio- että analytiikkatyökaluista. Tämä järjestelmä toimisi toisaalta analytiikkojen tietovarastona, johon historiatieto on tallennettuna ja toisaalta käyttöympäristönä, johon analytiikot voivat dataintegraatio-työkalujen avulla luoda sekä vakiomuotoisia että ad-hoc -tyyppisiä datan haku ja -muokkausproseduureja. Tällöin esimerkiksi vakiomuotoisten analyysien tekeminen onnistuu miltei tahansa ajanjaksolta helposti ja nopeasti. Uusia analyysejä on helppo luoda ja tarvittaessa säilyttää ne tulevaisuutta varten. Historiadata säilyisi koko ajan validina, koska siihen ei tehtäisi lainkaan datan muokkauksia.
- Analytiikkaa varten tulisi hankkia laadukkaat analyysityökalut, joiden avulla voitaisiin luoda nykyistä syvällisempiä analyysejä. Täten esimerkiksi täsmällisyyden kehittämisen ja uusien investointien suunnittelun käyttöön saataisiin analyysejä, joista kyettäisiin tulkitsemaan vaihtoehtoisten toimintamallien vaikutukset järjestelmään.

Analytiikkaympäristön ja siihen soveltuvan dataympäristön luomisen puolesta puhuu se, että niiden avulla on mahdollista tuottaa nykyistä informatiivisempia analyysejä päätöksenteon tueksi. Analytiikkaympäristön rakentamisen puolesta puhuu myös se tosiasia, että tällä tavoin analyyseiden laatimiseen vaadittuja resurssien määrää ei merkittävästi kasva automatiikan hoitaessa ison osan työstä.

5.2.1 Kulutietojärjestelmän tarkoituksenmukaisuuden parantaminen

Analyyseiden luotettavuus ja käyttökelpoisuus riippuu suoraan datan kattavuudesta ja oikeellisuudesta. Tähän vaikuttavat asiat voidaan jakaa kahteen osaan: järjestelmän tarkoituksenmukaisuus (mm. oikeanlaiset kentät) ja kirjaamiskäytännöt. JUSE-järjestelmän tarkoituksenmukaisuus ei tällä hetkellä ole paras mahdollinen. Kirjausjärjestelmä ja siihen liittyvä syykoodisto (ks. taulukko 3) on puutteellinen monin osin. Ongelmien korjaamiseksi JUSE:a (tai sen seuraajaa) ehdotetaan kehitettävän seuraavalla tavalla:

- *Seurantatietoa tulisi tarkentaa.* Toisin sanoen esimerkiksi junien kulkutiedot voisivat sisältää myös käytetyt raiteet ja kulkutiet, jolloin kyettäisiin tällä tarkkuudella tarkastelemaan, missä junat saavat viiveitä ja miten esimerkiksi kulkuteiden varaukset vaikuttavat muihin juniin. Tähän liittyvä kehitystyö onkin jo käynnissä, kun kauko-ohjausjärjestelmiä asteittain uusitaan. Vaikka seurantatietoa tarkennetaan, voitaisiin viivesyyt nykyiseen tapaan kirjata vasta asemille, jottei liikenteenohjaajien työmäärä kasva kohtuuttomasti.
 - *Häiriöiden analysoinnin kannalta seurantapistepohjaista järjestelmää parempi ratkaisu olisi häiriöpohjainen järjestelmä, jossa viive kytketyisi automaattisesti suoraan häiriöön, ei seuraavaan seurantapisteseen. Tämä kuitenkin edellyttää GPS-pohjaista junien seuranta ja nykyistä enemmän automatiikkaa. Lisäksi on huomattava, että myös seurantapistettä tarvitaan, esimerkiksi asematäsmällisyyden seurantaan.*
- *Eri juniin liittyvä seurantapisteverkosto tulisi yhdenmukaistaa.* Tämän myötä kaikkien junien analysointi voitaisiin tehdä yhtä tarkasti, ja niiden tiedot olisivat vertailukelpoisia.
- *Tietojen muokkaamismahdollisuuksia tulisi laajentaa ja automatisoida.* Esimerkiksi supistamissuunnitelmien yhteydessä tapahtuvat linjatunnusten muutokset tulisi voida syöttää järjestelmään ja vieläpä mahdollisimman automatisoidusti.
- *Aikaleimojen tarkkuutta tulisi parantaa.* Minuuttipohjaisesta järjestelmästä tulisi siirtyä sekuntipohjaisuuteen. Tämä yhdessä seurantapisteverkoston tarkentamisen myötä antaisi mahdollisuuden huomattavasti syvällisempiin analyyseihin. Sekuntipohjaisuus edellyttää GPS-pohjaista järjestelmää ja seurantapisteen sijoittamista junien todellisille pysähtymispaikoille. Asiaan liittyvä kehitystyö on jo käynnissä.
- *Havaintojen yhteyteen tulisi automaattisesti kirjautua myös aikataulunmukainen aika.* Tämä helpottaisi tapahtumaketjujen analysointia merkittävästi.
- *Ensimmäisen viiveen ja lisäviiveen suuruus tulisi saattaa samalle tasolle.* Pie-nille, alle kolmen minuutin myöhästymisille tuskin on kuitenkaan mielekästä ryhtyä kirjaamaan syitä. Voisikin harkita sitä, että ensimmäiselle viivesyylle kirjautuisi minuiteja vain siltä osin, kuin se ylittää myöhästymisen rajan (ts. myöhästymisrajan ollessa 3 min., kirjautuisi 5 min. viiveestä ko. syyllä vain 2 min.). Tämä lisäisi viivesyiden vertailukelpoisuutta.
- *Ensimmäiselle viiveelle tulisi lisätä automaattinen merkki,* koskien myös kiinniajamisen jälkeen tapahtuvaa ”uutta” ensimmäistä viivettä. Usein lisäviiveet ovat ainakin osittain seurausta ensimmäisestä viiveestä. Näin ollen ensimmäisten viiveiden poiminta analysointia varten tulisi olla helppoa.

- *Tulisi harkita, että kullekin viiveelle olisi mahdollista kirjata useampia syitä.* Tämä lisäisi viivesyiden minuuttimäärien luotettavuutta, mutta toisaalta lisäisi kirjauksiin liittyvää työtä ja harkintaa.
- *Samantasoisten syykoodien päällekkäisyys tulisi poistaa,* jotta kullekin häiriötyypille olisi vain yksi mahdollinen kirjaustapa. Paljon voidaan tehdä jo nykyisten koodien kirjausohjeistuksia tarkentamalla, ja tätä työtä tehdäänkin parasta aikaa. *Lisäksi tulisi harkita sitä, että koodistosta rakennettaisiin hierarkkinen.* Toisin sanoen esimerkiksi radasta johtuville syille olisi yksi ylätasoon koodi ja useampia tarkentavia alatasoon koodeja. Tällöin, mikäli tarkkaa syytä ei tiedetä, voitaisiin häiriölle antaa vain ylätasoon koodi. Näin tarkoille syille lankeaisi vain ne viiveminuutit, jotka niille varmasti kuuluvat.
- *Viiveen primäärisyyttä/sekundäärisyyttä ei tulisi päätellä suoraan viivekoodista,* sillä tietyt viiveet voivat olla tapauksesta riippuen kumpia tahansa. Sen sijaan viiveen primäärisyys/sekundäärisyys tulisi määritellä omalla kentällänsä. Vaihtoehtoisesti primäärisyys voisi olla oletusarvo, jolloin sekundäärisiksi tulkittaisiin ne viiveet, joille on merkitty aiheuttajaksi toinen juna. Kummassakin tapauksessa on syytä huolehtia, ettei puhtaasti primäärisluontoisia syitä pysty kirjaamaan sekundäärisiksi. Vastaavasti puhtaasti sekundäärisluontoisten viiveiden osalta aiheuttavan junan merkintä tulisi olla pakollista. Järjestelmää tältä osin kehittämällä helpotetaan merkittävästi viiveketjujen analysointia ja myös viiveminuuttien allokointia todellisille syilleen.
- *Sekundääristen viiveiden osalta tarvitaan oma, määrämuotoinen kenttensä aiheuttavien junien numeroille.* Tämä mahdollistaisi kentän automatisoidun lukemisen ja siten viiveketjujen analysoinnin sekä sekundääristen viiveminuuttien allokoinnin primäärisyilleen. Kuten jo aikaisemmin todettiin, tulisi kenttä syykoodista riippuen olla joko lukittu, vapaaehtoinen tai pakollinen. *Myös tietuille muille tiedoille, kuten joidenkin koodien edellyttämälle liikennepaikalle, tulisi harkita määrämuotoista kenttää.*
- *Syykirjauksiin tulisi lisätä oma, määrämuotoinen kenttensä vastuutahoa varten.* Tämä antaisi merkittävästi uusia mahdollisuuksia täsmällisyysjohtamiseen (viiveminuutteihin sidotut kannustimet ja sanktiot). Tietyillä syykoodeilla vastuutaho voisi olla valmiiksi määriteltä, toisilla se valittaisiin esimääritelystä valikosta. Lähtökohtana on, että kaikille viiveille kirjattaisiin vastuutaho. Erikseen on kuitenkin vielä tarkasteltava, onko tämä kaikissa tapauksissa järkevää, vai tulisiko kirjaus tiettyjen koodien osalta olla vapaaehtoista.
- *Junille tulisi antaa yksilölliset ID-numerot.* Näin tietojen käsittely ja eri tietojen yhdistely olisi nykyistä helpompaa ja tehokkaampaa. Nyt junat yksilöidään junanumeron ja lähtöpäivän avulla, mikä tarkoittaa ylimääräistä työvaihetta.

Edellisten havaintojen mukainen kirjausjärjestelmä voisi yksittäisen junan osalta olla taulukon 5 kaltainen. Taulukkoa tulkitessa tulee ottaa seuraavat seikat huomioon:

- Osa kentistä (mm. poikkeama) on laskettavissa muista tiedoista, eikä tämä siten liene paras mahdollinen master-tili. Kaikki taulukossa esitetyt tiedot kuitenkin tarvitaan perusanalyysessä tehtäessä, eli viimeistään vakionäkymissä ne tulee esiintyä.
- Vihreällä on merkitty oleelliset parannukset nykyiseen järjestelmään verrattuna.
- Tietueen laadinnan tavoitteena on ollut määritellä, kuinka nykyistä JUSE-järjestelmää tulisi kehittää nimenomaan analytiikan näkökulmasta. Lopullista järjestelmää rakennettaessa tulee myös muiden funktioiden tarpeet ottaa huomioon.

- Tietue ei juuri ota kantaa järjestelmän toiminnallisuustarpeisiin. Esimerkiksi poikkeustilanteissa tapahtuviin linjatunnusten ja junanumeroiden muutoksiin liittyy paljon haasteita.

Taulukko 5. Ajatuksia uudesta kirjausjärjestelmästä (yksittäisen junan tietue).

	Kenttä	Tyyppi	Arvot	Automaattisuus	Pakollisuus	Huom.
Perustiedot (1)	Junan ID	Kok. luku	1...n	Automaattinen	Pakollinen	ID:llä juna voidaan linkata muihin tauluihin, kuten kalustotietoihin
	Junanumero	Kok. luku	1...n	Automaattinen	Pakollinen	
	Lähtöpvm	Pvm	-	Automaattinen	Pakollinen	-
Tarkentavat tiedot (1); linkkaus muista tauluista	<i>Junalaji</i>	<i>Merkkijono</i>	<i>LL, KL, TL</i>	<i>Automaattinen</i>	<i>Pakollinen</i>	
	<i>Junatyyppi</i>	<i>Merkkijono</i>	<i>Esimäär. lista</i>	<i>Automaattinen</i>	<i>Pakollinen</i>	<i>Junalajista riippuen, esim. P, IC, ...</i>
	<i>Linjatunnus</i>	<i>Merkkijono</i>	<i>Esimäär. lista</i>	<i>Automaattinen</i>	<i>(Pakollinen)</i>	<i>LL-junien osalta, esim. A, E, ...</i>
	<i>Operaattori</i>	<i>Merkkijono</i>	<i>Esimäär. lista</i>	<i>Automaattinen</i>	<i>Pakollinen</i>	<i>Oleellinen kilpailun avauduttua</i>
Kulktieto (1...n)	Havaintopiste	Merkkijono	Esimäär. lista	Automaattinen	Pakollinen	HKI, TPE, jne.
	Havaintopisteen tarkennus	Merkkijono	A, M, T, L	Automaattinen	Pakollinen	Alkuas. (A), määräas. (M), tulo välias. (T), lähtö välias. (L)
	Havainnon nro	Kok. luku	1...n	Automaattinen	Pakollinen	Helpottaa havaintojen järjestämistä
	Aikataulu aika	PvmKlo	-	Automaattinen	Pakollinen	Sekunnin tarkkuus
	Toteutunut aika	PvmKlo	-	Automaattinen	Pakollinen	Sekunnin tarkkuus
	Poikkeama	Klo	-	Automaattinen	Pakollinen	Sekunnin tarkkuus
	Ensimmäinen viive	Boolean	Tosi/epätosi	Automaattinen	Pakollinen	Tosi, jos ensimmäinen myöhästymisrajan ylittänyt viive; myös kiinniajon jälkeinen "uusi" ensimmäinen viive
Syytieto (1...n)	Päätason syy	Merkkijono	Esimäär. lista	Manuaalinen	(Pakollinen)	Kirjataan, jos ja vain jos ylittää myöhästymisrajan; <i>koodisto määriteltävä erikseen</i>
	Tarkennettu syy	Merkkijono	Esimäär. lista (riippuu edell. kentästä)	Manuaalinen	Vapaaehtoinen	<i>Koodisto määriteltävä erikseen</i>
	Sekundäärinen	Boolean	Tosi/epätosi	Autom./Manuaal.	Pakollinen	Osa koodeista automaattisesti
	Aiheuttanut juna	Kok. luku	Esimäär. lista	Manuaalinen	<i>Useita</i>	Pakollisuus riippuu edell. kentästä
	Vastuutaho	Merkkijono	Esimäär. lista	Autom./Manuaal.	<i>Useita</i>	Automaattisuus ja pakollisuus riippuvat syykoodista; HUOM. sekundääristen viiveiden automaattinen vastuutus
	Häiriökoodi	Kok. luku	Vapaa	Manuaalinen	Vapaaehtoinen	<i>Myöhemmin automaattinen?</i>
	Lisätieto	Merkkijono	Vapaa	Manuaalinen	Vapaaehtoinen	Vapaaehtoinen kommentointi

5.2.2 Kulktietojärjestelmään liittyvien kirjaamiskäytäntöjen kehittäminen

Datan kattavuuteen ja luotettavuuteen vaikuttaa järjestelmän tarkoituksenmukaisuuden lisäksi vallitsevat kirjaamiskäytännöt. Tällä hetkellä JUSE-järjestelmään liittyvät kirjaamiskäytännöt ovat monin osin melko kirjavia. Toisin sanoen samankaltaisia asioita kirjataan monin eri tavoin. Lisäksi vapaaehtoiset kentät jätetään usein täyttämättä. Tästä seuraava datan epätarkkuus ja puutteellisuus heijastuu luonnollisesti ana-

lyyseihiin. Tällä hetkellä ongelma koetaan erittäin suureksi, ja siihen tulisi puuttua mahdollisimman pian.

Edellä käsiteltyjen teknisten keinojen (mm. kenttien automatisointi ja pakolliseksi asettaminen) lisäksi kirjaamiskäytäntöjä voidaan kehittää pääasiassa ohjeistamalla ja kouluttamalla. Samalla on myös oleellista viestiä, minkä takia kirjausten luotettavuus ja kattavuus on niin tärkeää: analytiikan ohella tietoja käytetään esimerkiksi sanktio- ja kannustejärjestelmissä.

Ongelmana kirjaamiskäytäntöjen kehittämisessä on, että systemaattisten virheiden löytäminen on hankalaa: yksittäistä tietuetta tarkastelemalla on yleensä mahdotonta sanoa, onko se oikein vai väärin kirjattu. Näin ollen pelkän datan perusteella on hankala päätellä, mihin ohjeistusta ja koulutusta tulisi kohdistaa. Nyström (2008) on käsitellyt samaa ongelmaa Ruotsin rautateillä. Hän tarkasteli liikenteenohjaajien kirjaamiskäytäntöjä yksinkertaisen *vignette*-menetelmän avulla. Siinä liikenteenohjaajien tuli osoittaa, kuinka he kirjaavat erilaisia kuvitteellisia tapahtumia kulkutietojärjestelmään. Kukin tapaus esitettiin omana korttinaan, *vignettenä* (kuva 55).

20. Train 01 waits for train 03 because of their scheduled exchange of drivers. Train 03 is 20 minutes late due to a signalling fault (this has previously been coded into TFÖR), causing 01 to leave 18 minutes late from the exchange. How do you code this?

Train 01:

Code: _____ Causing train number: _____ Event report: _____

How important do you judge the correctness of this coding? Mark the most appropriate box.

Not important at all Rather important Important Very important Don't know

Comment: _____

Kuva 55. *Esimerkki vignetteistä. Vignettejä käyttäen voidaan selvittää, kuinka liikenteenohjaajat kirjaavat erilaisia tilanteita kulkutietojärjestelmään. (Nyström 2008)*

Vignette-menetelmää käyttäen saadaan selvitettyä yksinkertaisesti, minkälaisien tapauksien kirjaamisessa on eniten virheellisyttä ja varianssia. Esimerkiksi Nyströmin tutkimuksessa kävi ilmi, että tietynlaisia tapauksia kirjattiin hyvinkin monin eri tavoin. Tällaisen tiedon avulla ohjeistusta ja koulutusta – ja mahdollisesti myös järjestelmän kehitystä – on mahdollista kohdistaa juuri oikeisiin asioihin. Toisaalta menetelmä osoittaa myös ne tapaukset, joissa kirjaukset tehdään jo luotettavasti. Esimerkiksi Nyströmin tutkimuksessa kävi ilmi, että tietynkaltaiset tapaukset kirjattiin jo valmiiksi oikein. Näin ollen tämänkaltaisissa tapauksissa voidaan datasta tehtyihin analyysiin luottaa huomattavasti nykyistä enemmän.

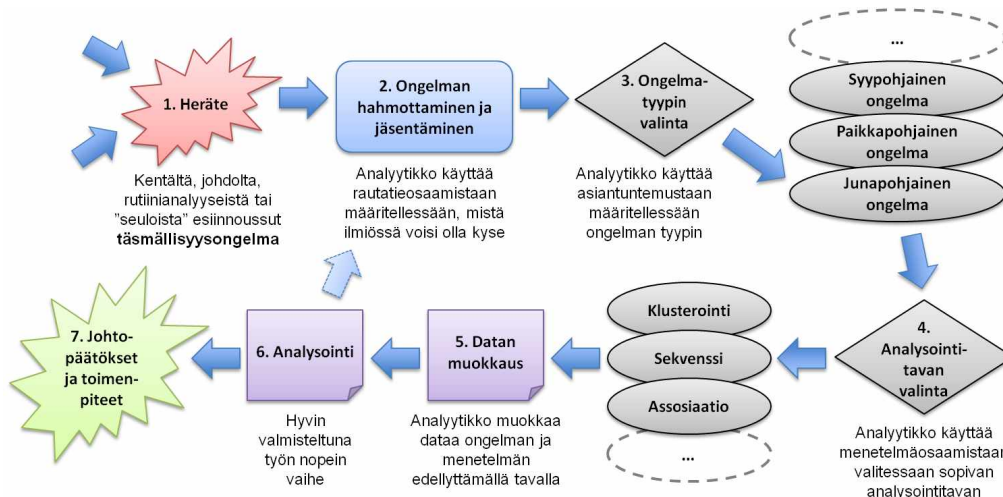
5.3 Analyysitoiminnan kehittäminen

Kunnollisen data- ja analytiikkaympäristön rakentamisen jälkeen on mahdollista keskittyä analyysitoiminnan kehittämiseen. Suomen rautatieliikenteen täsmällisyyden analysointi on tällä hetkellä raportoinnin tasolla ja sitä tulisi kehittää niin, että analyysejä laaditaan muutoinkin kuin vain kuukausittaisten vakioraporttien muodossa. Seuraavassa on esitetty kehitysehdotukset analyysitoiminnan kehittämiseen:

- Vakioraportoinnin automatisointia tulisi kehittää. Tämä mahdollistaisi sen, että raporteja voitaisiin ajaa nopeammin ja myös erilaisilta ajanjaksoilta, kuten esimerkiksi viikoittain. Lisäksi aikaa jäisi tällöin jokaisen kuukauden kohdennetumpaan analysointiin vakiomuotoisten analyysien laatimisen sijaan.
- Tehtyjen raporttien hyödyntämistä tulisi lisätä. Tämä onnistuu toisaalta varmistamalla, että raportit vastaavat todellisiin tietotarpeisiin, toisaalta opastamalla ja kouluttamalla eri tahoja tietojen seurantaan, tulkintaan ja hyödyntämiseen.
- Raportointia tulisi kehittää siten, että se vastaisi paremmin päätöksentekijöiden tietotarpeisiin. Apuna kannattaa hyödyntää *Rautatieliikenteen täsmällisyyteen liittyvät tietotarpeet* -raporttia (Paavilainen et al. 2011). Lisäksi päätöksentekijät tulisi perehdyttää siihen, millaisia analyysejä ylipäättään on mahdollista tehdä.
- Erilaisten ilmiöiden kertaluonteisen, rajatun tarkastelun lisäksi tulisi tarkastelemaan näiden ilmiöiden mahdollisia toistuvuuksia tai laajempia vaikutuksia kokonaisjärjestelmälle.
- Uusien raporttipohjien ja ad hoc -raporttien laadintaa tulisi systematisoida (ks. luvut 5.3.1. ja 5.3.2.)
- Analytikoiden määrää tulisi lisätä, ja heitä tulisi kouluttaa kehittyneen data-analytiikan menetelmien käyttöön. Paras mahdollinen tilanne on perustaa oma analytiikkaorganisaatio, jonka päävastuuna on tuottaa analyysejä niin vakioraporttien muodossa kuin ad hoc -analyysien muodossa.
 - *On huomattava, että ratkaisuna ei voi olla Liikenneviraston ja VR:n analyysivoimavarojen yhdistäminen. Kilpailun avauduttua Liikenneviraston tulee olla ylin organisaatio, joka tuottaa analyysejä koko järjestelmän kehittämiseksi. Liikennöitsijöiden puolestaan tulisi tuottaa analyysejä pääasiassa oman toimintansa kehittämiseen.*
- ”Yleisanalytikoiden” lisäksi tulee varmistaa, että eri toiminta-alueilla on minimissään jonkinlainen yhteyshenkilö, joka nostaa esille analyysejä vaativia kysymyksiä ja hypoteeseja. Optimitilanteessa nämä toiminta-alueiden henkilöt kuuluisivat em. analytiikkaorganisaatioon, osallistuen analysointiin, oman toiminta-alueensa näkökulmasta. Tämä takaisi, että kultakin toiminta-alueelta saadaan tuotettua juuri sille oleellista tietoa.

5.3.1 Analysointiprosessin systematisointi

Vakioraportointi toimii jo tällä hetkellä melko kattavasti ja systemaattisesti. Sen sijaan erilaisista herätteistä kumpuavien kertaluonteisten, ns. ad hoc -analyysien tuottamisessa on kehittämisen varaa. Prosessissa on paljon varianssia (samankaltaisia ongelmia ratkotaan monin eri tavoin) eikä aikaisemmissa analyyseissä kertynyttä tietoa hyödynnetä parhaalla mahdollisella tavalla. Resurssien riittämättömyyden vuoksi monet tarpeelliset analyysit jäävät kokonaan tekemättä.



Kuva 56. Ad hoc -analyysiprosessin systematisointi

Kuvassa 56 on esitetty työn aikana muodostunut käsitys siitä, kuinka ad hoc -analysointiprosessia tulisi systematisoida. Prosessin vaiheet ovat:

- 1. Heräte.** Herätteellä tarkoitetaan kentältä (asiakkailta, työntekijöiltä), johdolta, rutiinianalyseista tai "seuloista" (täsmällisyysdataan kytketyt automaattiset hälytyspisteet) esiinnoussutta *täsmällisyysongelmaa*. Täsmällisyysongelma voi olla esimerkiksi tietyn junan toistuva epätäsmällisyys tai tietyn viivesyyn aikaisempaa runsaampi esiintyminen. Herätteiden määrittelemistä, keräämistä ja tulkintaa tulisi kehittää ja monipuolistaa nykyisestä tasosta.
- 2. Ongelman hahmottaminen ja jäsentäminen.** Seuraavaksi tulee hahmottaa ja jäsentää, mistä herätteen kautta esiinnoussessa ongelmassa itse asiassa on kyse: seurauksien sijaan tulisi etsiä perimmäistä syytä ongelmalle. Tässä vaiheessa on oleellista, että analytikolla on riittävä ymmärrys rautatieliikennejärjestelmän toiminnasta. Mikäli tilanne tulkitaan väärin, ajaudutaan helposti analysoimaan väärää asiaa.
- 3. Ongelmatyyppin valinta.** Ongelman muotoilun jälkeen tutkitaan, onko se jonkin valmiiksi määritellyn *ongelmatyyppin* kaltainen. Valmis ongelmatyyppi voi olla esimerkiksi tietyssä paikassa tapahtuneet säännölliset viiveet (paikkapohjainen ongelma). Valmiin ongelmatyyppin tunnistaminen helpottaa prosessia sen myöhemmissä vaiheissa. Seuraavassa alaluvussa on määritelty yleisimpiä ongelmatyypppejä.
- 4. Analysointitavan valinta.** Seuraavaksi valitaan ongelmalle sopiva *analysointitapa/-tavat*. Valmiiksi määritellyille ongelmatyypeille myös analysointitavat voivat olla valmiiksi määritellyt (ks. seuraava alaluku). Tässä vaiheessa korostuu analytikon menetelmäosaaminen: hänen pitää tietää, mitä milläkin menetelmällä voi tehdä. Lisäksi analytikon tulee tietää, mitä kaikkea tietoa olemassa oleva data ylipäätään voi tarjota, mitä ei. Data tulee siis olla kuvailtuna mahdollisimman hyvin. Poikkeuksellisten ongelmien tapauksessa analytikon asiantuntemus korostuu entisestään.
- 5. Datan muokkaus.** Kun ongelma on jäsenetty ja analysointimenetelmä(t) valittu, tulee käytössä oleva data muokata sopivaan muotoon. Tämä voi tarkoittaa esimerkiksi erilaisten tietokantojen yhdistämistä. Selkeimpien ongelmatyyppien ja niissä käytettävien menetelmien osalta tarvittavat muokkaukset kannattaa määritellä – ja jopa tehdä – etukäteen (ks. seuraava alaluku).

6. *Analysointi.* Varsinainen analysointi on ammattilaiselle työn nopein vaihe. Selkeimpien ongelmatyyppien ja niissä käytettävien menetelmien osalta analysointiprosessit voidaan karkealla tasolla määritellä jopa etukäteen (ks. seuraava alaluku). *On huomattava, että analysointi on iteratiivista työtä. Toisin sanoen on tavallista, että analysointivaihe nostaa esille jonkin tärkeän tai mielenkiintoisen seikan, jota on syytä tutkia tarkemmin. Niinpä analyysoija muotoilee tästä seikasta uuden ongelman, jota hän ryhtyy tarkastelemaan siihen sopivin menetelmin. Prosessissa siis palataan vaiheeseen 2.*
7. *Johtopäätökset ja toimenpiteet.* Tässä tutkimuksessa tarkasteltiin, kuinka analytiikkaa voisi ja tulisi kehittää. Pitää kuitenkin muistaa, että tarkimmatkin analyysit jäävät turhiksi, elleivät ne johda johtopäätöksiin ja edelleen toimenpiteisiin. Tässä vaiheessa korostuu rautatieliikennejärjestelmän tuntemus (ettei tehdä vääriä johtopäätöksiä). Analysoijien ja päätöksentekijöiden yhteistyö puolestaan on avainasemassa siinä, että johtopäätökset päätyvät konkreettisiksi, täsmällisyyttä edistäviksi toimenpiteiksi.

5.3.2 Yleisimpien ongelmatyyppien analysointi

JUSE-datan yksinkertaisuudesta huolimatta sen avulla voidaan analysoida hyvinkin monenlaisia asioita. Valtaosa näistä analyyseistä voidaan kuitenkin jakaa kahteen perustyyppiin: *paikkapohjaiset ongelmat* ja *junapohjaiset ongelmat*. Muitakin ongelmatyyppäjä voidaan tunnistaa, mutta useimmat niistä palautuvat näihin kahteen perustyyppiin. Seuraavaksi käydään yleisellä tasolla läpi, kuinka analyysiprosessi etenee em. kahden perustyyppin osalta. Lisäksi osoitetaan, kuinka muut ongelmatyyppit palautuvat näihin perustyyppihin. Esimerkkinä toimii *syypohjainen ongelma*.

Paikkapohjainen ongelma

Paikkapohjainen ongelma liittyy nimensä mukaisesti johonkin paikkaan, ts. rataosaan tai asemaan. Herätteenä toimii siis havainto siitä, että ko. paikassa viiveiden määrä (joko yleisesti tai tietyn syyryhmän osalta) ylittää tavanomaisen rajan.

Paikkapohjaista ongelmaa voidaan analysoida monella eri tavalla. Analyysissä voi kiinnostaa muun muassa, (1) mitä ongelmia tietyllä paikalla tapahtuu, (2) mitä ongelmia tietyn paikan haasteet aiheuttavat toisaalla sekä (3) miten toimintaympäristö vaikuttaa analysoitavaan paikkaan ja sen ongelmiin. Kukin näistä tapauksista on datan muokkauksen suhteen yksilöllinen.

Ensimmäisessä tapauksessa dataa muokataan siten, että siinä keskitytään vain tutkimaan ko. paikkaan liittyviä havaintoja. Näiden havaintojen tutkimuksessa mielenkiinnon kohteena voi olla vaikkapa, miten tietyn viiveen jälkeen seuraa muita viiveitä (sekvenssianalyysi), mille ajanjaksoille viiveet ryhmittyvät (klusterointi) tai miten kyseessä olevan paikan viiveet voidaan luokitella esimerkiksi vaikuttavuuden suhteen (segmentointi).

Toisessa tapauksessa datan muokkaaminen vaatii jo huomattavasti enemmän. Datan täytyy sisältää jokin tietokenttä siitä, miten eri havainnot linkittyvät keskenään. Lähökohdaksi otetaan tutkittavan paikan viivehavainnot ja niitä ketjutetaan olemassa olevien linkityksien avulla mahdollisimman pitkälle. Vain tällä tavoin voidaan luotettavasti linkittää ko. paikan aiheuttamien viiveiden ketjuuntuminen toisaalle. Mikäli datahavaintoja ei voida luotettavasti linkittää keskenään, voidaan yrittää tutkia, mikä muiden paikkojen viiveet korreloivat tutkittavan paikan viiveiden suuruuden muu-

toksen kanssa. Tässä tavassa ongelmana on se, että syy-seuraus-suhteiden todentaminen on vaikeaa.

Kolmas tapaus on hyvin samankaltainen kuin toinen. Erona on se, että toisessa tapauksessa tutkitaan eteenpäin ketjuuntuneita viiveitä, kun puolestaan kolmannessa viiveitä ketjutetaan taaksepäin. Aivan kuten toisessakin, myös kolmannessa vaaditaan muuttuja, jonka avulla havaintoja voidaan linkittää toisiinsa.

Paikkapohjaisia ongelmia on raportissa käsitelty luvussa 4.2. Kuten analyysistä näkyy, voidaan paikkapohjaisen analyysin avulla tarkastella hyvin monenlaisia asioita. Analyysin laadinta on monivaiheinen ja etenee sen mukaan, mitä edellisen vaiheen datasta on saatu ilmenemään.

Junapohjainen ongelma

Junapohjainen ongelma liittyy tiettyyn junaan, valitulla ajanjaksolla. Herätteenä toimii havainto siitä, että ko. junan kulussa on tavanomaista enemmän poikkeuksia, ts. sen viiveiden määrä (joko yleisesti tai tietyn syyryhmän osalta) on poikkeuksellisen suuri tarkastelujaksolla.

Junapohjaisessa analyysissä tutkitaan aluksi, miten viivehavainnot ryhmittyvät (klusteroituvat) junilla, jolloin voidaan havaita tapahtumia, jotka ovat tyypiltään samanlaisia. Tämän jälkeen voidaan assosiaatioanalyysin avulla tutkia näiden tapahtumien ilmenemisyhteyksien keskinäistä riippuvuutta. Lisäksi sekvenssianalyysin avulla voidaan tarkastella, mitkä viiveet edeltävät ja seuraavat viiveitä. Luokittelun avulla, asetamalla viiveminuutit kohdemuuttujiksi, voidaan puolestaan tarkastella, mitkä viiveet missäkin ovat suuruudeltaan merkittävämpiä ja useammin ilmeneviä kuin toiset.

Junapohjaisessa analyysissä dataa muokataan siten, että dataksi voidaan valita tietyn junanumeron/tiettyjen junanumeroiden mukaiset havainnot. Toisaalta voidaan valita tiettyjen liikennepaikkojen havainnot, jolloin saadaan kaikki junat, jotka käyvät liikennevälin asemilla. Tällöin voidaan mahdollisuuksien mukaan pureutua viiveiden ketjuuntumiseen.

Analyysin avulla voidaan tutkia, millaisia toistuvia viivetyyppejä tietyn liikennöintivälin junat kokevat ja miten ne ketjuuntuvat ja näkyvät koko liikennöintivälillä. Tällä tavoin on mahdollista päästä käsiksi sekä junien kulun rakenteellisiin haasteisiin, kuten käytössä olevien aikataulujen sisältämiin ongelmiin, kuin myös liikennevälin sisältämiin ongelmiin aina yksittäisistä toistuvista ongelmakohdista pullonkauloihin. Esimerkki tästä juna-analyysistä löytyy raportin luvusta 4.1.

Syypohjainen ongelma

Syypohjaisessa ongelmassa herätteenä toimii havainto siitä, että ko. syyryhmän aiheuttamia viiveitä esiintyy tarkastelujaksolla tavanomaista enemmän. Tarkastelualue voi olla joko koko rataverkko tai sen osa.

Syypohjaisen ongelman analysoinnin ehkäpä yksi informatiivisimmista tavoista on assosiaatioanalyysi. Assosiaatioanalyysin avulla voidaan tutkia, mitkä viivetyypit yhdistyvät vahvimmin millekin junille ja liikennepaikoille. Analyysiprosessin perusteella muotoillaan uusi ongelma, joka on joko paikkapohjainen (viiveet ilmenevät tietyillä paikoilla) tai junapohjainen ongelma (viiveet liittyvät tiettyihin juniin).

5.4 Työryhmien toiminnan kehittäminen

Hyvin organisoitu analyysitoiminto, jolla on vakioraportoinnin lisäksi kyvykkyyttä ja resursseja ideoida ja toteuttaa erilaisia ad hoc -raportteja, antaisi merkittävästi lisää mahdollisuuksia työryhmien toimintaan. Työryhmät voisivat teettää analyytikoilla erilaisia analyysejä ennen jälkeen kokouksien. Analyytikot voisivat osallistua kokouksiin esitellen tekemänsä analyysit ja osallistuen uusien analyysitehtävien muotoiluun. Parhaassa tilanteessa työryhmien toimintaan osallistuisi analyytikkoja, joilla on osaaminen ja työkalut laatia analyysejä reaaliaikaisesti työryhmän työskentelyn aikana. Ottamalla analytiikkaan erikoistunutta henkilökuntaa työryhmätoimintaan mukaan, on mahdollista saada lisää tietopohjaa päätöksentekoon.

Työryhmien toiminnassa on tunnistettavissa muitakin kehityskohteita, liittyen esimerkiksi ryhmien organisoitumiseen, rooleihin, vastuisiin ja valtuuksiin. Näitä ei kuitenkaan käsitellä tässä raportissa.

5.5 Tutkimuksen arviointi ja jatkotutkimusehdotukset

Tutkimusprosessin, tulosten ja näistä tehtyjen päätelmien perusteella voidaan todeta tutkimuksen olleen onnistunut. Tutkimuksen avulla kyettiin osoittamaan, että rautatieliikennejärjestelmän analysointia on mahdollista syventää merkittävästi kehittyneen data-analytiikan menetelmien avulla.

On kuitenkin huomattava, että tutkimuksessa toteutetut analyysit olivat lähinnä pintaraapaisu siihen, mitä kaikkea menetelmillä on mahdollista tehdä. Analysoituja aiheita olisi voinut syventää tarkastelemalla niitä useammasta näkökulmasta. Mikäli käytössä olisi ollut tarkempaa ja luotettavampaa kulkutietodataa, olisivat tulokset olleet vieläkin vaikuttavampia.

Tutkimuksen pohjalta syntyi valtavasti uusia analyysikohteita: rautatieliikennejärjestelmässä on lukemattomasti kertaluontoisia ja säännöllisiä ilmiöitä, joita tähän asti ei ole analysoitu. Esimerkkinä voidaan mainita seuraavia tutkimusaiheita:

- Miten viiveet ketjuuntuvat ja vaimenevat kokonaisjärjestelmässä; voidaanko asiasta luoda yleisemmän tason malli?
- Missä ovat järjestelmän ongelmakohdat liikennöinnin kannalta, missä viiveketjut puolestaan saadaan katkeamaan?
- Kuinka viiveet ketjuuntuvat Kytömaan solmukohdassa (Lahden oikoradan ja pääradan risteyksessä)?
- Kuinka henkilö- ja tavaraliikenne aiheuttaa toisilleen häiriöitä?
- Kuinka lähi- ja kaukoliikenne aiheuttaa toisilleen häiriöitä?

Muun muassa tällaisista asioista olisi mahdollista jalostaa tietoa päätöksentekijöiden tueksi. Tiedon avulla niin strategiseen, taktiseen kuin operatiiviseenkin kehittämiseen budjetoidut resurssit saadaan käytettyä siellä, missä niillä on suurin vaikutus järjestelmän kokonaistäsmällisyyden kannalta.

Lähteet

Baskerville, R. & Myers, M.D. 2004. Special issue on action research in information systems: Making IS research relevant to practice – foreword. *MIS Quarterly*. 28(3). ss. 329-335.

Blomqvist, E. 2011. Henkilöliikenteen täsmällisyys kaukoliikenteessä, Henkilöliikenteen täsmällisyys lähiliikenteessä sekä Tavaraliikenteen täsmällisyys - kuukausiraportit. VR-Yhtymä OY.

Bose, R. 2009. Advanced analytics: Opportunities and challenges. *Industrial Management & Data Systems*. 109(2).. ss. 155-172.

Cyert, R. M., Herbert A. S., Donald B. T. 1956. Observation of a business decision. *Journal of business* (29). ss. 237-248.

Daft, R.L. & Weick, K.E. 1984. Toward a model of organizations as interpretive systems. *Academy of Management Review*. 9(2). ss. 284-295.

Davenport, T. H. 2009. Make better decisions. *Harvard Business Review* (November). ss. 117-123.

Davenport, T. H. & Harris, J. G. 2007. *Competing on analytics: the new science of winning*. Boston, Harvard Business School Press. 215 s.

Davenport, T. H., Harris, J. G., Morison, R. 2010. *Analytics at work: smarter decisions, better results*. Boston, Harvard Business Press. 214 s.

Eckerson, W.W. 2007. Predictive Analytics – Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report, First Quarter 2007. 32 s.

Giudici, P. 2003. *Applied data mining: statistical methods for business and industry*. John Wiley & Sons. 364 s.

Han, J. & Kamber, M. 2000. *Data Mining: Concepts and techniques*. Morgan Kaufmann Publishers. 550 s.

Hansen, I.A. 2001. Improving railway punctuality by automatic piloting. 2001. *IEEE Intelligent Transportation Systems Proceedings*, ss. 792-797.

Iversen, J.H., Mathiassen, L., Nielsen, P.A. 2004. Managing risk in software process improvement: An action research approach. *MIS Quarterly*. 28(3). ss. 395-434.

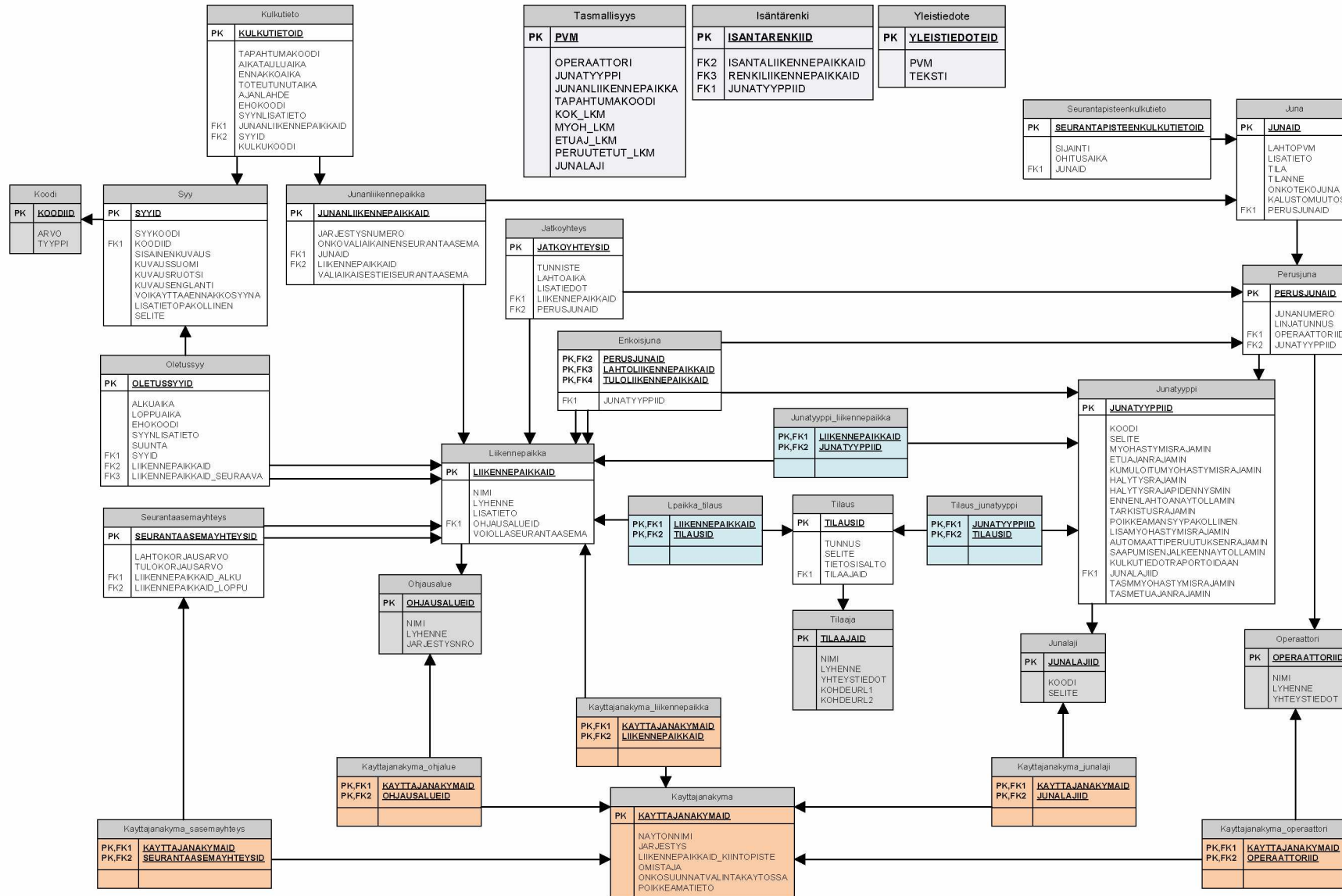
Keim, D., Andrienko, G., Fekete, J-D., Görg, C., Kohlhammer, J., Melancon, G. 2008. *Visual Analytics: Definition, Processes and Challenges*. Kerren, A., Stasko, J., Fekete, J-D., North, C. *Information Visualization: Human-Centered Issues and Perspectives*. Berlin, Springer. ss. 154-175.

Kaijaa, M. 2011a. *Matkustajaliikenne täsmällisyys – kaukojunat huhtikuu 2011*. Helsinki, VR-Yhtymä Oy. Organisaation sisäisesti julkaistu raportti. 15 s.

- Kaijaa, M. 2011b. Helsingin lähiliikenne täsmällisyys – huhtikuu 2011. Helsinki, VR-Yhtymä Oy. Organisaation sisäisesti julkaistu raportti. 10 s.
- Kaijaa, M. 2011c. Tavaraliikenne täsmällisyys – huhtikuu 2011. Helsinki, VR-Yhtymä Oy. Organisaation sisäisesti julkaistu raportti. 10 s.
- Kudyba, S. & Hoptroff, R. 2001. Data Mining and Business Intelligence: A Guide to Productivity. Lontoo UK, Idea Group Publishing. 166 s.
- Levin, D.Z., Cross, R., Abrams, L.C., Lesser, E.L. 2002. Trust and knowledge sharing: A critical combination. IBM Institute for Knowledge-based Organizations.
- Liikennevirasto. 2010. Talvi 2009-2010 Suomen rautateillä – tapahtumat ja johtopäätökset. Helsinki, Liikennevirasto. Liikenneviraston tutkimuksia ja selvityksiä 15/2010. 55 s.
- Liikennevirasto. 2011. Rautatieliikenteen täsmällisyys 2010. Helsinki, Liikennevirasto. Liikenneviraston tutkimuksia ja selvityksiä 18/2011. 67 s.
- Mattila, H. 2011. Myöhästymissyöt. 21.1.2011. Helsinki, Liikennevirasto. Julkaisematon sähköposti. 1 s.
- Mattila, H., Matinlauri, A-M. 2011. Rautatieliikenteen täsmällisyys huhtikuu 2011. Helsinki, Liikennevirasto. Organisaation sisäisesti julkaistu raportti. 42 s.
- Nyström, Birre 2008. A methodology for measuring the quality of deviation reporting: Applied to railway delay attribution. Luleå University of Technology, 2008.
- Paavilainen, J., Salkonen R., Rantala, T. 2011. Rautatieliikenteen täsmällisyyteen liittyvät tietotarpeet. Helsinki, Liikennevirasto. 67 s.
- Profillidis, V. A. 2006. Railway Management and Engineering. 3. pianos. Burlington USA, Ashgate Publishing Company. 469 s.
- Salkonen, R., Mäkelä, T. 2010. Rautatieliikenteen täsmällisyyden mittaamisen ja seurannan käytännöt eri maissa. Helsinki. Liikennevirasto. 76 s.
- Salkonen, R., Paavilainen, J., Mäkelä, T. 2009. Rautatieliikenteen täsmällisyydestutkimuksen kirjallisuuskatsaus. Helsinki, Ratahallintokeskus. 272 s.
- SAS. 2006. SAS Enterprise Miner 4.3 Reference. Cary USA, SAS Institute Inc.
- Susman, G., Evered, R. 1978. An Assessment of the Scientific Merits of Action Research. Administrative Science Quarterly. 23(4). ss. 582-603.
- Tapola, P. 2011. JUSE-aktiivikantaan liittyviä kysymyksiä. 23.5.2011. Helsinki, Liikennevirasto. Julkaisematon sähköposti. 1 s.
- Thierauf, R. J. 2001. Effective Business Intelligence System. West Port, Quorum Books. 370 s.

JUSEn aktiivikanta

LIITE 1



JUSEn historiakannan taulut

- *JUNA (Päivälle merkityn junan tiedot)*
 - JUNANUMERO = Junan numero (Eri päivinä samaan aikaan ja samaa reittiä kulkevat junat menevät samalla numerolla)
 - LAHTOPVM = Päivämäärä, jolloin juna on lähtenyt
 - LINJATUNNUS = Lähiliikenteen junien linjatunnus
 - OPERAATTORI = Operaattorinnumero (VR Tavaraliikenne, Henkilökauko ja Helsingin seudun lähiliikenne eri numeroilla)
 - JUNALAJI = Tavaraliikenne, Henkilökaukoliikenne, Helsingin seudun lähiliikenne ja siirtojunat
 - JUNATYYPPI = Junantyyppi (Esim. IC, S, P, IC2, Dm, Ta, jne.)
 - MYOHASTYMISRAJAMIN = Myöhästymiskirjausten raja
 - ETUJANARAJAMIN = Etuakakirjausten raja
 - LISATIETO = Vapaa lisätietokenttä liikenteenohjaajia varten
 - KALUSTOMUUTOS = Kalustoon tehtyjen muutosten vapaa kirjaus
 - TILANNE = Junan toteuman tilanne (Ajettu, Peruttu, Etuajassa, Myöhässä)
 - TASMMYOHASTYMISRAJAMIN = Täsmällisyyden myöhästymisten raja
 - TASMETUJANRAJAMIN = Täsmällisyyden etuajassakulun raja

- *HISTORIALOKI (Junien kulkutiedot päivittäin ja asemittain)*
 - JUNANUMERO = Junan numero (Eri päivinä samaan aikaan ja samaa reittiä kulkevat junat menevät samalla numerolla)
 - LAHTOPVM = Päivämäärä, jolloin juna on lähtenyt
 - JUNANLIIKENNEPAIKKA = Junan liikennepaikka (Esim. Tpe, Hki, Epo, jne.)
 - JUNANLIIKENNEPAIKKANUMERO = Numero, joka ilmaisee liikennepaikan järjestyksen junan kulun suhteen. Ts. Alkuasema on 1 ja loput kasvavat tästä eteenpäin
 - TAPAHTUMAKOODI = Mikä junan tapahtuma (Alku, Tulo, Lähtö, Määränpää)
 - KULKUKOODI = Miten juna on kulkenut (Normaalisti, Peruttu)
 - POIKKEAMAMIN = Monta minuuttia junan kulku on poikennut aikataulusta
 - SYYKOODI = Myöhästymisen syykoodi
 - SYYNLISATIETO = Lisätietokirjaus myöhästymiseen liittyen, esim. toisen junan junanumero
 - LUOKITTELUID = Kulun luokittelu (Ei käytetä)
 - MUUTTAJA = Kirjauksen muuttaja (Ei käytetä)
 - MUUTOSPAIVA = Kirjauksen muutospäivä (Ei käytetä)

- *TASMALLISYYS (Edellä esitetyistä tauluista laskettua tietoa)*
 - PVM = Päivämäärä
 - OPERAATTORI = Operaattorinnumero (VR Tavaraliikenne, Henkilökauko ja Helsingin seudun lähiliikenne eri numeroilla)
 - JUNATYYPPI = Junantyyppi (Esim. IC, S, P, IC2, Dm, Ta, jne.)
 - JUNANLIIKENNEPAIKKA = Junan liikennepaikka (Esim. Tpe, Hki, Epo, jne.)
 - TAPAHTUMAKOODI = Mikä junan tapahtuma (Alku, Tulo, Lähtö, Määränpää)
 - KOK_LKM = Yhteenlaskettu junien lukumäärä
 - MYOH_LKM = Myöhästyneiden junien lukumäärä
 - ETUAJ_LKM = Etuajassa kulkeneiden junien lukumäärä
 - PERUUTETUT_LKM = Peruutettujen junien lukumäärä
 - LUOKITTELUID = Kulun luokittelu (Ei käytetä)
 - JUNALAJI = Tavaraliikenne, Henkilökaukoliikenne, Helsingin seudun lähiliikenne ja siirtojunat

Työryhmiin osallistuminen ja esiin nousseet aihealueet

Tässä liitteessä on nostettu esiin toimintatutkimuksen aikana eri työryhmistä esiin nousseita keskeisimpiä tutkittavia aihe-alueita. Liitteestä on tietoisesti jätetty pois Mattilan, Tapolan, Blomqvistin ja Paavilaisen kanssa käydyt projektitapaamiset, joissa ideoita nousi myös merkittävästi esiin ja jossa ideoita jalostettiin. Kuten liitteestä voidaan todeta, nousee ryhmissä merkittävästi analyysi-ideoita esille, mutta toteutetut analyysit jäävät usein henkilöiden kiireellisyyden vuoksi tekemättä.

Täsmä-ryhmä:

21.1.2011

- Mitkä ovat radalla pullonkaulat?
- Miksi kaikki myöhästymiset näkyvät Pieksämäen ympäristössä?
- Ruuhka-aikana minuutin myöhästymisen näkyminen pitkänä ketjuna, miten tätä voisi mallintaa?
- Oikoradan haarautuminen Kytömaalla, miten näkyvät täsmällisyyksissä?
- Tavaraliikenteen osalta syykoodit ovat osittain arvontaa
- Mitä yhteysliikenne on odottanut itä-suomessa?
- Lähtötäsmällisyydet jokaiselta seuranta-asemalta mallinnettava
- Yksittäisen junan kulun täsmällisyyden visualisointi kartalle.
- Ruuhka-aika 7:00-9:59 ja 15:00-17:59, miten vaikuttaa täsmällisyyteen?

Infratäsmä-ryhmä:

17.12.2010

- Miten täsmällisyysdatasta saisi irti korrelaatioita, regressioita, lisätietoa?
- Miten visualisoitavissa paremmin?
- Jos kauko-ohjauksjärjestelmä kaatuu, niin saako live-dataa?
- Voisiko esim. Hki ASTL jumitilanteen vaikutuksen nopeutta esim. Tampereella havainnollistaa tällä?
- Voisiko pikavoittoja löytää tämän avulla ("missä aina alkaa myöhästymiset")?

16.2.2011

- Miten kyettäisiin ennakoimaan, milloin turvalaitteet yms. laitteet ovat menossa rikki korjaustiedoista, yms.
- Onko havaittavissa malleja, jotka edeltävät turvalaitteiden rikkimienemisiä?
- Miten pohjanmaan radan routavauriot näkyvät muualla radalla? Miten eroaa viimevuodesta?

Analyysi-ryhmä:

14.2.2011

- Kalustokierron ja kunnossapidon vaikutus täsmällisyyteen?
- Kun ajetaan vajailla Pendojen määrillä, näkyykö tämä jossain?
- Henkilöstökierron ja työläinsäädännön vaikutus vuorojen suunnitteluun.

VR:n kehitysjohtaja Antti Korhosen tapaaminen:

8.3.2011

- Helsingin ratapihan vaihtotyöt ovat merkittävä ongelma. Onko oikeasti näin? Miten tätä voitaisiin tutkia?
- Voidaanko todeta, että jokin kalustoryhmä tai rataosa on merkittävä täsmällisyyshaitta?
- Miten kalustokierto, henkilöstökierto ja kunnossapito vaikuttavat täsmällisyyteen? Dataa mahdollisesti saatavissa.

