

Kati Tillander, Anna Matala, Simo Hostikka, Pekka Tiittanen,
Esa Kokki, & Olli Taskinen

Pelastustoimen riskianalyysimallien kehittäminen

VTT TIEDOTTEITA – RESEARCH NOTES 2530

Pelastustoimen riskianalyysimallien kehittäminen

Kati Tillander, Anna Matala, Simo Hostikka
VTT

Pekka Tiittanen, Esa Kokki
Pelastusopisto

Olli Taskinen
Helsingin pelastuslaitos



ISBN 978-951-38-7573-2 (URL: <http://www.vtt.fi/publications/index.jsp>)

ISSN 1455-0865 (URL: <http://www.vtt.fi/publications/index.jsp>)

Copyright © VTT 2010

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 3, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 4374

VTT, Bergsmansvägen 3, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 3, P.O. Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax +358 20 722 4374

Toimitus Mirjami Pullinen

Kati Tillander, Anna Matala, Simo Hostikka, Pekka Tiittanen, Esa Kokki & Olli Taskinen. Pelastustoimen riskianalyyysmallien kehittäminen. Espoo 2010. VTT Tiedotteita – Research Notes 2530. 117 s. + liitt. 9 s.

Avainsanat accident frequency, building fires, rescue services, regression models, spatial regression, self-organizing maps

Tiivistelmä

Tutkimuksessa pyrittiin kolmea eri lähestymistapaa käyttäen luomaan menetelmiä, joiden avulla onnettomuustiheys pystyttäisiin ennustamaan entistä luotettavammin riskiruutuaineiston muuttujia apuna käyttäen. Tämän lisäksi tehtiin suppea yhteenveto vuosina 2005–2008 toteutuneista onnettomuustiheyksistä eri riskialueilla.

Lähestymistavoista lupaavimmaksi osoittautuivat yksinkertaiset regressiomallit, joista tuotettiin helppokäyttöisiä työvälineitä riskinarviointiin. Mallit luotiin erikseen rakennuspaloille ja kiireellisille onnettomuuksille. Rakennuspalomalli kehitettiin koko maalle ruutukoossa $1 \text{ km} \times 1 \text{ km}$. Selittäjinä mallissa ovat asukasluku, kerrosala sekä niiden yhteisvaikutus. Mallin puutteellisuudet huomioon ottaen sitä voidaan käyttää rakennuspalomäärien ennustamiseen.

Kiireellisille onnettomuuksille kehitetyssä mallissa ruutukoko oli $1 \text{ km} \times 1 \text{ km}$ ja selittäjänä pelkästään asukasluku. Aluekohtaiset erot mallin sopivuudessa ja ennustekyvyyssä olivat suuria. Erityisesti pohjoisessa Suomessa malli ei toiminut. Mallia olisikin selvästi parannettava ennen sen laajempaa soveltamista.

Yksinkertaisten regressiomallien rinnalla testattiin myös spatiaalisten mallien kykyä ennustaa onnettomuusmääriä. Mallit osoittautuivat joustaviksi mutta laskennallisesti hankaliksi, ja niiden käytäntöön saattaminen vaatii vielä jatkokehittämistä.

Kolmannessa lähestymistavassa sovellettiin itseorganisoituvia karttoja (*Self-Organizing Maps*, SOM). Lähestymistavan vahvuutena oli se, että mallin muuttujina voitiin käyttää kaikkia riskiruutuaineiston sisältämiä muuttujia toteutuneiden onnettomuusmäärien lisäksi. SOM:n tuottamien onnettomuusmääräennusteiden keskihajonta osoittautui kuitenkin joissakin tapauksissa melko suureksi, eikä absoluuttisia lukumääriä pystytty ennustamaan riittävällä tarkkuudella. Onnettomuusmäärien suuruusluokkia voidaan kuitenkin käyttää riskien arviointiin.

Itseorganisoituvien karttojen osalta testattiin vielä, paraneeko ennustustulos, kun muuttujiksi lisätään sosioekonomista tietoa. Kokeilun tuloksena havaittiin,

että sosioekonominen aineisto saattaisi parantaa ennustustulosta, vaikka tällä aineistolla parannus ei ollutkaan merkittävä. Sosioekonomisen aineiston toimituutta tulisi vielä selvittää ennen kuin voidaan luotettavasti määrittää aineiston vaikutus mallien tuottamiin ennusteisiin.

Alkusanat

Tämä julkaisu on hankkeen ”Pelastustoimen riskianalyysin kehittäminen” loppuraportti. Hanke on toteutettu vuonna 2009 VTT:n, Helsingin pelastuslaitoksen ja Pelastusopiston yhteistyönä. Hanketta ovat rahoittaneet sisäasiainministeriö, VTT, Helsingin pelastuslaitos ja Pelastusopisto.

Hankkeen toteutusta valvoi ohjausryhmä, johon kuuluivat edustajat seuraavilta tahoilta: sisäasiainministeriö, VTT, Helsingin pelastuslaitos, Pelastusopisto, Keski-Uudenmaan pelastuslaitos, Tampereen aluepelastuslaitos ja Jokilaaksojen pelastuslaitos.

Kiitokset Anton von Schantzille ja Perttu Punakalliolle, jotka käyttivät aurinkoiset kesäpäivät aineistoja järjestellen ja spatiaalisia malleja sovittaen. Esitämme kiitokset myös ohjausryhmälle, joka omalla työllään on panostanut hankkeeseen ja edistänyt sen toteutumista merkittävällä tavalla.

Tekijät

Sisällysluettelo

Tiivistelmä	3
Alkusanat	5
Symboliluettelo.....	9
1. Johdanto	11
1.1 Tausta	11
1.2 Tavoite	11
1.3 Käytetty aineisto.....	12
1.3.1 Onnettomuudet.....	12
1.3.2 Riskiruudut	12
1.3.3 Tarkastelualueiden maantieteellinen jako	13
1.4 Riskianalysimallien kehitystyön lähestymistavat.....	13
1.5 Mallien toimivuuden arviointi.....	14
1.5.1 Tausta	14
1.5.2 Vertailu vuonna 2008 toteutuneisiin onnettomuusmääriin	14
1.5.2.1 Yleistä vertailusta.....	14
1.5.2.2 Virhe (ennustettu–havaittu).....	16
1.5.3 Vertailu toimintavalmiusohjeessa esitettyihin malleihin.....	19
2. Toteutunut onnettomuustiheys eri riskiluokan ruuduissa vuosina 2005–2008.....	22
2.1 Koko maa	22
2.2 Eri pelastustoimen alueet.....	23
2.3 Lukumäärien riippuvuus asukasluvusta ja kerrosalasta	25
2.3.1 Kiireelliset onnettomuudet asukaslukuokittain.....	25
2.3.2 Rakennuspalot kerrosalaluokittain	26
3. Yksinkertainen regressiomalli	28
3.1 Aineisto	28
3.1.1 Rakennuspalot	28
3.1.2 Kiireelliset tehtävät	29
3.2 Mallin kuvaus	30
3.2.1 Yhteyden muoto ja mallin termien valinta.....	30
3.2.2 Mallin valinta.....	32
3.3 Tulokset	34

3.3.1	Rakennuspalot	34
3.3.2	Kiireelliset tehtävät	36
3.4	Mallin toimivuuden arviointi.....	37
3.4.1	Rakennuspalot	37
3.4.2	Kiireelliset tehtävät	45
3.5	Johtopäätökset.....	52
3.5.1	Yleistä malleista	52
3.5.2	Rakennuspalot	53
3.5.3	Kiireelliset tehtävät	53
3.6	Kehittämisehdotukset.....	54
4.	Spatiaalinen regressiomalli	56
4.1	Mallinnuksen lähtökohta	56
4.2	Käytetty aineisto.....	57
4.3	Mallin kuvaus	57
4.4	Tulokset	60
4.5	Mallin toimivuuden arviointi.....	71
4.6	Johtopäätökset.....	76
4.7	Kehittämisehdotukset.....	77
5.	Itseorganisoituvat kartat (SOM)	79
5.1	Käytetty aineisto.....	79
5.2	Mallin kuvaus	80
5.2.1	Teoria	80
5.2.2	Sovellus riskiruutuaineistoon.....	81
5.3	Tulokset	82
5.3.1	Rakennuspalot	82
5.3.2	Kiireelliset onnettomuudet	87
5.4	Mallin toimivuuden arviointi.....	92
5.4.1	Vertailu vuonna 2008 toteutuneisiin onnettomuusmääriin	92
5.4.2	Vertailu toimintavalmiusohjeen riskiluokitukseen	96
5.4.2.1	Rakennuspalot	96
5.4.2.2	Kiireelliset onnettomuudet.....	99
5.5	Johtopäätökset.....	101
5.6	Kehittämisehdotukset.....	101
6.	Sosioekonomisen tiedon hyödyntäminen	103
6.1	Käytetty aineisto.....	103
6.2	Tulokset	104
6.3	Johtopäätökset.....	107
6.4	Kehittämisehdotukset.....	108
7.	Mallien välisten erojen vertailu.....	109
7.1	Kiireellisten onnettomuuksien ennustevirhe alueella 1	109
7.2	Kolmen eri mallitustavan tuottamien kiireellisten onnettomuuksien ennusteiden erot....	112
8.	Yhteenveto ja johtopäätökset.....	113

Lähdeluettelo..... 116

Liitteet

- Liite A: Spatiaalisissa malleissa käytetty Winbugs-koodi
- Liite B: Muuttujien järjestäytyminen itseorganisoituvalla kartalla
- Liite C: Yksinkertaisten regressiomallien estimoinnissa käytetty R-koodi

Symboliluettelo

x_1	Asukasluku
x_2	Kerrosala [m ²]
y	Tehtävien keskimääräinen lukumäärä ruudussa
α	Tasoparametri
β	Selittäjän kerroin regressiomallissa
λ	Ruudun autokorreloitunut tekijä
γ	Ruudun autokorreloimaton tekijä
μ	Ruudun naapureiden keskiarvo
τ	Varianssi
<i>Kvartiiliväli</i>	Kvartiiliväli ilmoittaa muuttujan vaihteluvälin, jolle sijoittuu jakauman keskellä olevat havaintoarvot eli 50 % kaikista tapauksista.
<i>Noodi (tai neuroni)</i>	Neuroverkon solmukohtia, joiden referenssi-vektorien avulla etsitään jokaiselle syötevektorille sitä parhaiten vastaava noodi.
<i>Riskiruutu</i>	Riskiruutu on suunnittelun apuväline, joka muodostuu 250 m × 250 m kokoisesta alueesta. Tilastokeskus on tehnyt pelastustointia varten ns. hila-aineiston, jossa koko maa on jaettu 250 × 250 metrin ruutuihin (Sisäasiainministeriö 2003).

<i>Riskiluokka</i>	Riskiluokka riskiruudulle muodostuu, kun toimintavalmiusohjeessa (Sisäasiainministeriö 2003) mainittu asukasluvun tai kerrosalan raja-arvo ruudulla täyttyy.
<i>Riskialue</i>	Riskialue muodostuu, kun vähintään kymmenen samaan tai ylempään riskiluokkaan kuuluvaa riskiruutua koskettavat toisiaan (Sisäasiainministeriö 2003).
<i>SOM</i>	Itseorganisoituva kartta (<i>Self-Organizing Maps</i>), eräs neuroverkkosovellus.

1. Johdanto

1.1 Tausta

Toimintavalmiusohje A:71 on ollut hyvä pohja, johon eri pelastustoimen alueiden riskianalyysit ovat joko kokonaan tai osittain perustuneet. Riskianalyysin perusteella on määritelty pelastustoimen alueella sijaitsevat riskialueet ja palvelutasopäätöksen yhteydessä paloasemaverkosto, jolla on pyritty vastaamaan havaittuihin uhkiin.

Nykyisen toimintavalmiusohjeen käyttöönoton jälkeen on kuitenkin kertynyt paljon tilastotietoa sekä matemaattisia apuvälineitä, ja myös tietokoneiden laskentateho on kasvanut moninkertaiseksi. Niiden avulla voidaan arvioida nykyisten riskiluokitusten toimivuutta ja luoda entistä edistyneempiä työvälineitä alueellisten riskien arviointiin.

Uudet menetelmät pystyvät huomioimaan onnettomuuksien esiintymiseen liittyvien taustamuuttujien vaikutusten alueellista vaihtelua ja muutoksia ajassa. Niiden avulla saadaan myös tarkempi kuva riskianalyysin luotettavuudesta.

1.2 Tavoite

Tavoitteena oli luoda parannettuja menetelmiä, joilla voitaisiin määritellä eri alueiden onnettomuusriskitaso entistä luotettavammin. Lisäksi kartoitettiin tilastotietojen pohjalta viime vuosien aikana toteutuneet onnettomuustiheydet eri riskiluokissa.

1. Johdanto

1.3 Käytetty aineisto

1.3.1 Onnettomuudet

Tarkasteluissa käytettiin PRONTOsta poimittuja rakennuspaloja sekä kiireellisiä onnettomuuksia vuosilta 2005–2008 (Pelastusopisto 2009). Kiireellistä toimintaa vaativiksi onnettomuuksiksi määriteltiin seuraavat onnettomuustyypit:

- rakennuspalo
- liikennevälinepalo
- maastopalo
- muu tulipalo
- räjähdys/räjähdysvaara
- ihmisen pelastaminen
- vaarallisten aineiden onnettomuus
- liikenneonnettomuus
- sortuma/sortumavaara.

1.3.2 Riskiruudut

Aineistona on sisäasiainministeriön Tilastokeskukselta saama pelastustoimen tilastoaineisto vuodelta 2005. Aineiston 250 × 250 metrin kokoinen hilaruudukko kattaa koko Suomen. Tilastokeskuksen ruudukkoon on yhdistetty Väestötietokeskuksen väestötietojärjestelmän 250 × 250 metrin ruutuihin summattua aineistoa, muun muassa asukasluku 31.12.2005 sekä rakennusten lukumäärä ja kerrosala vuonna 2005. Väestötietojärjestelmän rakennustietokantaan ei lueta maatalousrakennuksia, vapaa-ajanrakennuksia eikä ryhmään ”muut rakennukset” kuuluvia rakennuksia, paitsi jos niissä on asuttuja asuntoja tai toimitiloja. Hilaan on viety alueluokitustietoina kuntakoodi sekä kuntakoodin perusteella lääni- ja maakuntakoodi. Lisäksi hilaan on viety pelastustoimen aluekoodi ja hätäkeskusaluekoodi sekä toimintavalmiusohjeen (Sisäasiainministeriö 2003) mukainen ruudun riskiluokan arvo I–IV, joka on määritelty asukasluvun ja kerrosalan perusteella. Aineistossa on yhteensä 6 055 185 ruutua.

1.3.3 Tarkastelualueiden maantieteellinen jako

Laskentateknisistä syistä Suomi jaettiin seitsemään alueeseen, jolle kullekin tulokset laskettiin erikseen. Kunkin alueen muodosti kolmesta viiteen pelastustoimen aluetta taulukon 1 mukaisesti.

Taulukko 1. Käytetty aluejako.

Alue	Pelastustoimen alue
1	Helsinki, Länsi-Uusimaa, Keski-Uusimaa
2a	Varsinais-Suomi, Kanta-Häme, Päijät-Häme, Pirkanmaa, Satakunta
2b	Keski-Suomi, Etelä-Pohjanmaa, Pohjanmaa, Keski-Pohjanmaa
3a	Etelä-Savo, Pohjois-Savo, Pohjois-Karjala
3b	Itä-Uusimaa, Kymenlaakso, Etelä-Karjala
4	Jokilaaksot, Kainuu, Oulu–Koillismaa
5	Lappi

1.4 Riskianalyysimallien kehitystyön lähestymistavat

Riskianalyysimallien kehittämiseen valittiin kolme toisistaan riippumatonta lähestymistapaa:

- yksinkertainen regressiomalli
- spatiaalinen regressiomalli
- itseorganisoiuva kartta (SOM).

Yksinkertaisen regressiomallin idea on nimensä mukaisesti kehittää mahdollisimman yksinkertainen malli, jolla voidaan ennustaa alueen onnettomuusriskit nykyistä tarkemmin. Mallissa selittäjinä toimivat asukasluku ja kerrosala. Malli on myös käyttäjälle hyvin yksinkertainen.

Spatiaalisen regressiomallin selkein ero edelliseen on, että se ottaa huomioon muuttujien alueellisen vaihtelun eli esimerkiksi sen, että asukasmäärän vaikutus onnettomuuksien syntymiseen ei välttämättä ole sama eri puolilla maata. Malli on kuitenkin käyttäjän kannalta selvästi monimutkaisempi kuin yksinkertainen regressiomalli.

1. Johdanto

Itseorganisoitua kartta ”oppii” muuttujien epälineaariset riippuvuussuhteet suoraan havaintoaineistosta. Tässä menetelmässä voidaan käyttää kaikkia mahdollisia ”selittäjiä”, ja määrittää niiden perusteella eri alueiden riskitasot. Menetelmä ei ole käyttäjälle kovinkaan yksinkertainen, ja se toimii käytännössä parhaiten, kun riskitasomäärittäykset tehdään suurehkolle alueelle määrätyn väliajoin (kun uutta tilastoaineistoa onnettomuuksista ja taustamuuttujista on saatu).

1.5 Mallien toimivuuden arviointi

1.5.1 Tausta

Jo työhön ryhtyessä tiedettiin, että käytössä olevan aineiston pohjalta kehitettävien mallien ennustuskyky ei tule olemaan täydellinen. Tämä näkyy erityisesti kiireellisten onnettomuuksien lukumäärien ennustamisessa, jonka osalta mallien ennustamiskykyä heikentää käytössä olevien selittäjien yksipuolisuus. Käytössä ollut aineisto ei esimerkiksi sisällä muuttujia, joiden voitaisiin arvioida selittävän muun muassa liikenneonnettomuuksia. Näin ollen voitiin jo etukäteen olettaa, että absoluuttinen virhe (mallin ennustama-toteutunut onnettomuusmäärä) saattaa olla suurehkokin.

Tutkimuksen tavoitteena oli tuottaa menetelmiä, joiden avulla onnettomuusriskitasot voitaisiin määrittää entistä luotettavammin. Mallien toimivuutta arvioitiin vertaamalla niiden antamia tuloksia vuonna 2008 toteutuneisiin onnettomuusmääriin, toimintavalmiusohjeessa (Sisäasiainministeriö 2003) esitettyjen mallien ennustuskykyyn sekä toisiinsa.

1.5.2 Vertailu vuonna 2008 toteutuneisiin onnettomuusmääriin

1.5.2.1 Yleistä vertailusta

Uudet mallit luotiin käyttämällä onnettomuusaineistoa vuosilta 2005–2007. Mallien toimivuutta arvioitiin muun muassa sen avulla, miten hyvin vuosien 2005–2007 onnettomuusaineistolla kehitetyt mallit onnistuivat ennustamaan vuonna 2008 toteutuneet onnettomuudet. Ennustusvaiheessa mallin muuttujina käytettiin samaa riskiruuuaineistoa (vuodelta 2005) kuin mallia kehitettäessä.

Taulukossa 2 on esitetty kiireellisten onnettomuuksien keskimääräinen lukumäärä vuosina 2005–2008. Taulukon 2 oikeanpuoleisessa sarakkeessa on esitetty vuoden 2008 onnettomuusmäärien prosentuaalinen muutos suhteessa vuosien

1. Johdanto

2005–2007 keskiarvoon. Kuten taulukosta 2 nähdään, onnettomuusmäärä on kasvanut noin 7 %. Rakennuspalot ovat lisääntyneet 15 %.

Tämän toimivuustarkastelun ongelmana on, että tutkimuksessa kehitetyt mallit eivät käsittele onnettomuusmääriä aikasarjoina vaan staattisina ominaisuuksina, eivätkä ne siten ota huomioon onnettomuusmäärien kasvavaa trendiä. Ainoastaan spatiaalinen regressiomalli huomioi trendin, mutta kolmen vuoden perusteella arvio on epätarkka. Kun selittäjinä käytetään samoja riskiruutuaineiston muuttujia (vuodelta 2005) kuin mallien kehittämisessäkin, on odotettavissa, että mallit ennustavat vuosia 2005–2007 vastaavia onnettomuusmääriä myös vuodelle 2008. Ei siis pystytä arvioimaan, miten ajassa tapahtuvat muutokset selittäjissä vaikuttavat onnettomuusmääriin. Tämän lisäksi vain yhden vuoden (2008) kattava vertailuaineisto on tilastollisen tarkkuuden kannalta hyvin pieni eikä riitä mallien toimivuuden kattavan arvioinnin pohjaksi.

1. Johdanto

Taulukko 2. Kiireellisten onnettomuuksien lukumäärät vuosina 2005–2008. Muutosarakkeessa on ilmoitettu, kuinka monta prosenttia suurempi vuoden 2008 lukumäärä oli verrattuna vuosien 2005–2007 keskiarvoon (Tietolähde: PRONTO, parametritilastot, poimittu 28.9.2009).

Onnettomuustyyppi	2005 (kpl)	2006 (kpl)	2007 (kpl)	2008 (kpl)	Muutos (v. 2008 lkm verrattuna v. 2005–2007 keskiarvoon)
Rakennuspalo	3 671	3 998	4 025	4 485	15 %
Liikennevälinepalo	2 630	2 725	2 549	2 379	–10 %
Maastopalo	2 631	6 290	2 813	3 161	–19 %
Muu tulipalo	4 045	4 567	4 769	4 619	4 %
Räjähdykset/räjähdyksvaara	34	23	24	25	–7 %
Sortuma/sortumavaara	0	4	8	16	300%
Ihmisen pelastaminen	2 003	1 964	2 174	2 150	5 %
Vaarallisten aineiden onnettomuus	225	260	217	277	18 %
Liikenneonnettomuus	10 205	10 912	12 874	13 401	18 %
Yhteensä	25 444	30 743	29 453	30 513	7 %

1.5.2.2 Virhe (ennustettu–havaittu)

Ennustetun ja havaitun onnettomuusmäärän välistä eroa tarkastellaan laatikkojanakuvion (*boxplot*) avulla. Laatikkojanakuviokuva esittää luokan arvojen hajonnan selkeässä ja helposti ymmärrettävässä muodossa.

Tarkastelussa ruudut on jaettu kymmeneen luokkaan ennustetun onnettomuusmäärän perusteella. Luokkien rajat on esitetty taulukossa 3. Jokaisen luokan pisteille lasketaan mediaani ja 25 %- ja 75 % -pisteet.

Laatikkojanakuviot koostuvat viivasta, laatikosta, viiksistä ja yksittäisistä pisteistä. Laatikon sisällä oleva viiva näyttää mediaanin paikan. Laatikon yläreuna esittää kohdan, jonka alapuolelle 75 % pisteistä osuu, ja vastaavasti alareuna paikan, jonka alla on 25 % pisteistä. Viiksien pituus on määritelty siten, että se

on maksimissaan 1,5 kertaa kvartiiliväli. Ylimääräiset pisteet kuvaavat sitten niitä ennustevirheen arvoja, jotka sijoittuvat viiksiäkin kauemmaksi laatikosta.

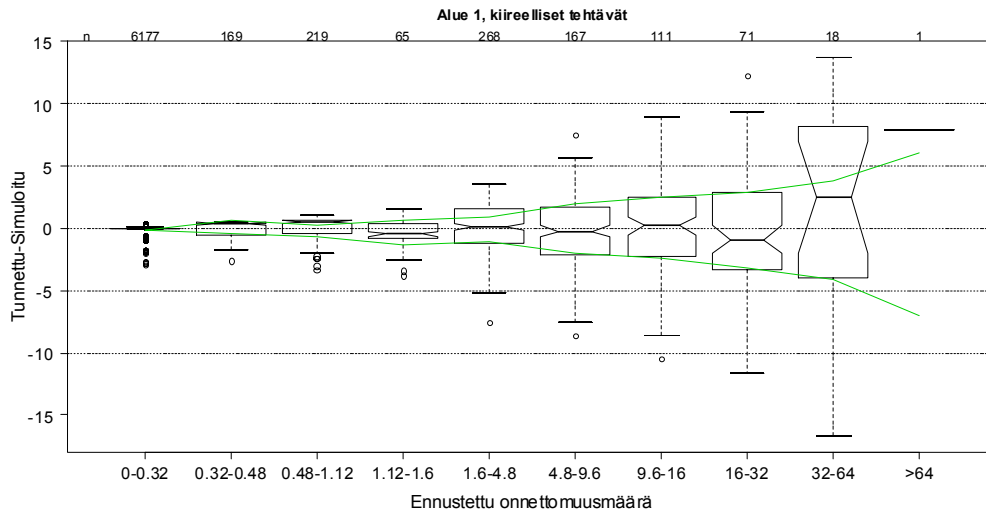
”Ennustevirhe” (ennustettu–havaittu) saadaan vähentämällä ennustettujen onnettomuuksien lukumäärästä (vuonna 2008) havaittujen onnettomuuksien lukumäärä. Tällöin hyvälle mallille mediaanin tulisi olla lähellä nollaa. Mikäli mediaani on suurempi kuin nolla, ennusteet ovat keskimäärin suurempia kuin toteutuneet ja päinvastoin, mikäli mediaani on nollaa pienempi. Jos laatikko on hyvin pieni (tai sitä ei näy), ainakin 50 % virheistä on hyvin lähellä mediaania. Suuri laatikko ja hajallaan olevat pisteet kertovat siitä, että mallin ennustevirheen vaihtelu on hyvin suuri tässä luokassa.

Onnettomuuksien esiintyminen ruudussa on lukumäärämuuttuja, jonka oletetaan syntyvän suureen joukkoon mahdollisia onnettomuuskohteita kohdistuvasta pienestä, riippumattomasta onnettomuustodennäköisyydestä. Tällöin ruudun onnettomuusmäärä vaihtelee vuosittain poissonjakauman mukaisesti. Poissonjakaumassa varianssi ja keskiarvo ovat yhtä suuret. Ennustettaessa yksittäisen vuoden onnettomuusmääriä ei ”täydellisenkään” mallin tule ennustaa virheettä. Kuvassa 1 on havainnollistettu ennustevirheiden laatikkojanakuviolla hyvän mallin ennustevirheiden suuruutta. Jokaiseen alueen 1 ruutuun on ennustettu asukasluvun perusteella kiireellisten tehtävien määrä kertoimen ollessa 0,0047. Jokaiselle ruudulle on sitten arvottu tämän keskimääräisen kiireellisten tehtävien määrän perusteella poissonjakaumasta yksittäisen vuoden tehtävämäärä. ”Tunnetun” ja simuloitun välinen erotus on tämän jälkeen luokiteltu kymmeneen luokkaan taulukon 3 mukaisesti. Vihreällä on esitetty luokkien tehtävämäärien keskiarvon mukainen ala- ja yläkvartiili.

1. Johdanto

Taulukko 3. Laatikkojanakuvioiden (*boxplot*) luokkarajat. Lukumäärät on määritetty yhtä vuotta kohden 1 km × 1 km -ruuduille.

Luokka	Rakennus-paloille	Kiireellisille onnettomuuksille
I	> 6,4	> 64
II	3,2–6,4	32–64
III	1,6–3,2	16–32
IV	0,96–1,6	8–16
V	0,48–0,96	4,8–8
VI	0.16–0.48	1.6–4.8
VII	0.112–0.16	0.96–1.6
VIII	0.048–0.112	0.48–0.96
IX	0.032–0.048	0.32–0.48
X	< 0.032	< 0.32



Kuva 1. Esimerkkikuva hyvän ennustemallin ennustevirheen suuruudesta.

1.5.3 Vertailu toimintavalmiusohjeessa esitettyihin malleihin

Toimintavalmiusohjeeseen vertailun tavoitteena oli selvittää, miten uusien mallien antamat tulokset suhteutuvat toimintavalmiusohjeessa esitettyyn ruutujen riskiluokitteluun. Toimintavalmiusohjeessa (Sisäasiainministeriö 2003) riskiluokkia on neljä. Kunkin riskiruudun (250 m×250 m) riskiluokka muodostuu, kun ohjeessa mainittu asukasluvun tai kerrosalan raja-arvo ruudulla täyttyy. Käytännössä riskiluokkien määrittäminen on tehty pelkästään näiden kahden muuttujan perusteella.

Tässä tutkimuksessa kehitettyjen mallien perusideana on ennustaa käytössä olevien muuttujien pohjalta ruudussa toteutuvien onnettomuuksien lukumäärä. Jotta vertailu toimintavalmiusohjeen riskiluokitteluun voidaan tehdä, on tiedettävä onnettomuusmääräraajat, joiden perusteella tietty ruutu luokitellaan johonkin neljästä riskiluokasta. Toimintavalmiusohjeessa ei kuitenkaan ole yksiselitteisesti ilmoitettu näitä onnettomuusmääriä.

1. Johdanto

Toimintavalmiusohjeen (Sisäasiainministeriö 2003) perustelumuihostiossa esitetään asukastiheyden ja kerrosalan arvojen perusteeksi rakennuspaloja koskevassa tutkimuksessa (Tillander & Keski-Rahkonen 2000) esitetyt riippuvuudet:

$$y = \begin{cases} 0.47x_1 \\ 6.55x_2 \end{cases}, \quad (1)$$

missä y = rakennuspalojen lukumäärä vuodessa, x_1 = asukasluku tuhansina ja x_2 = kerrosala neliökilometreinä. Toimintavalmiusohjeessa (Sisäasiainministeriö 2003) on lisäksi todettu, että yhtä rakennuspaloa kohden Suomessa sattuu vuodessa noin yhdeksänkertainen määrä muita palokunnan kiireellistä toimintaa vaativia tehtäviä.

Taulukossa 4 on esitetty asukas- ja kerrosalaluokkien rajat eri riskiluokissa sekä asukas- ja kerrosalarajoja vastaavat yhtälön (1) perusteella määritetyt onnettomuuksien lukumäärät. Määrät on laskettu rakennuspaloille, ja kiireellisten onnettomuuksien lukumäärän on oletettu olevan ohjeessa mainitusti yhdeksänkertainen. Tämän tutkimuksen aineistossa kiireellisten tehtävien määrä oli 7,3-kertainen vuosina 2005–2007 ja 6,8-kertainen vuonna 2008 rakennuspaloihin nähden (taulukko 2). Taulukossa 4 on esitetty vertailun vuoksi myös toteutunut onnettomuustiheys eri riskiluokan ruuduissa (ruutua ja vuotta kohden) vuosina 2005–2008.

Kuten taulukosta 4 nähdään, toimintavalmiusohjeen (Sisäasiainministeriö 2003) perusteella eri riskiluokissa toteutuneiden onnettomuusmäärien arviointi ei ole yksiselitteistä. Malleja ei siis voi vertailukelpoisesti verrata toimintavalmiusohjeen antamiin tuloksiin. Tarkastelut on esitetty kunkin mallin osalta, mutta tulosten tulokinnassa on otettava huomioon ristiriita onnettomuusmäärärajojen asettamisessa.

Tässä tarkastelussa uusien kehitettyjen mallien antamia ennusteita verrattiin kaavasta (1) asukasluvun perusteella määritettyihin onnettomuusmääriin (keltainen sarake taulukossa 4). Yksinkertaisten regressiomallien osalta vertailu on osittain tehty myös kaavasta (1) kerrosalan perusteella määritettyihin onnettomuusmääriin.

1. Johdanto

Taulukko 4. Asukas- ja kerrosalaluokkien rajat eri riskiluokissa sekä asukas- ja kerrosalalajoja vastaavat yhtälön (1) perusteella määritetyt onnettomuuksien lukumäärät eri riskiluokissa (oletuksena, että kiireellisiä onnettomuuksia tapahtuu yhdeksänkertainen määrä rakennuspaloihin verrattuna). Oikeanpuoleisessa sarakkeessa on lisäksi esitetty vuosien 2005–2008 tietojen perusteella määritetty toteutunut onnettomuustiheys riskiruutua kohden vuodessa. Mallien toimivuuden vertailussa käytettiin keltaisella merkityn sarakkeen lukumäärätietoja. Lukumäärät on määritetty yhtä vuotta kohden 250 m × 250 m:n ruuduille. 1 km:n ruutukoossa rajat ovat 16-kertaiset.

Riski-luokka	Asukasluku- ja kerrosalaluokkien rajat	$0,47 \cdot x_1$	$6,55 \cdot x_2$	Toteutunut lkm riskiruutua kohden 2005–2008
Rakennuspalot				
I	$x_1 > 250$ $x_2 > 10\,000 \text{ m}^2$	$y > 0,12$	$y > 0,07$	0,1
II	$60 < x_1 \leq 250$ $2\,500 < x_2 \leq 10\,000 \text{ m}^2$	$0,03 < y \leq 0,12$	$0,02 < y \leq 0,07$	0,04
III	$10 < x_1 \leq 60$ $250 < x_2 \leq 2\,500 \text{ m}^2$	$0,005 < y \leq 0,03$	$0,002 < y \leq 0,02$	0,006
IV	$x_1 \leq 10$ $x_2 \leq 250 \text{ m}^2$	$y \leq 0,005$	$y \leq 0,002$	0,0003
Kiireelliset onnettomuudet				
I	$x_1 > 250$ $x_2 > 10\,000 \text{ m}^2$	$y > 1$	$y > 0,6$	0,8
II	$60 < x_1 \leq 250$ $2\,500 < x_2 \leq 10\,000 \text{ m}^2$	$0,3 < y \leq 1$	$0,1 < y \leq 0,6$	0,2
III	$10 < x_1 \leq 60$ $250 < x_2 \leq 2\,500 \text{ m}^2$	$0,04 < y \leq 0,3$	$0,01 < y \leq 0,1$	0,05
IV	$x_1 \leq 10$ $x_2 \leq 250 \text{ m}^2$	$y \leq 0,04$	$y \leq 0,01$	0,005

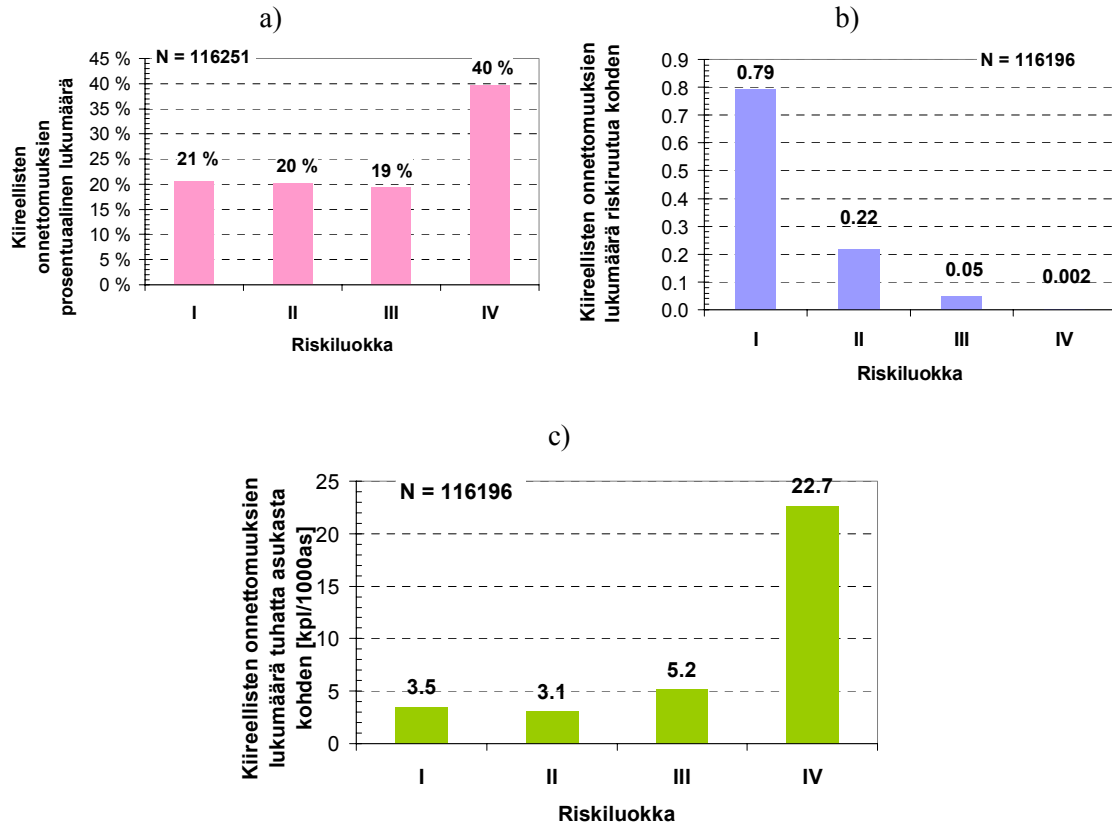
2. Toteutunut onnettomuustiheys eri riskiluokan ruuduissa vuosina 2005–2008

Tässä luvussa tehdään lyhyt yhteenveto siitä, kuinka paljon kiireellisiä onnettomuuksia vuosina 2005–2008 on tapahtunut eri riskiluokan ruuduissa.

2.1 Koko maa

Kuvassa 2a) esitetään kiireellisten onnettomuuksien jakautuminen eri riskiluokkiin merkittyjen riskiruutujen kesken. Kuvassa 2b) esitetään kiireellisten onnettomuuksien määrä riskiruutua ja kuvassa 2c) tuhatta asukasta kohden. Kuvasta 2a) nähdään, että määrällisesti eniten kiireellisiä onnettomuuksia tapahtuu riskialueella IV. Onnettomuustiheys riskiruutua kohden on suurin riskiluokan I ruuduissa, kun taas tuhatta asukasta kohden se on suurin riskiluokan IV ruuduissa.

2. Toteutunut onnettomuustiheys eri riskiluokan ruuduissa vuosina 2005–2008

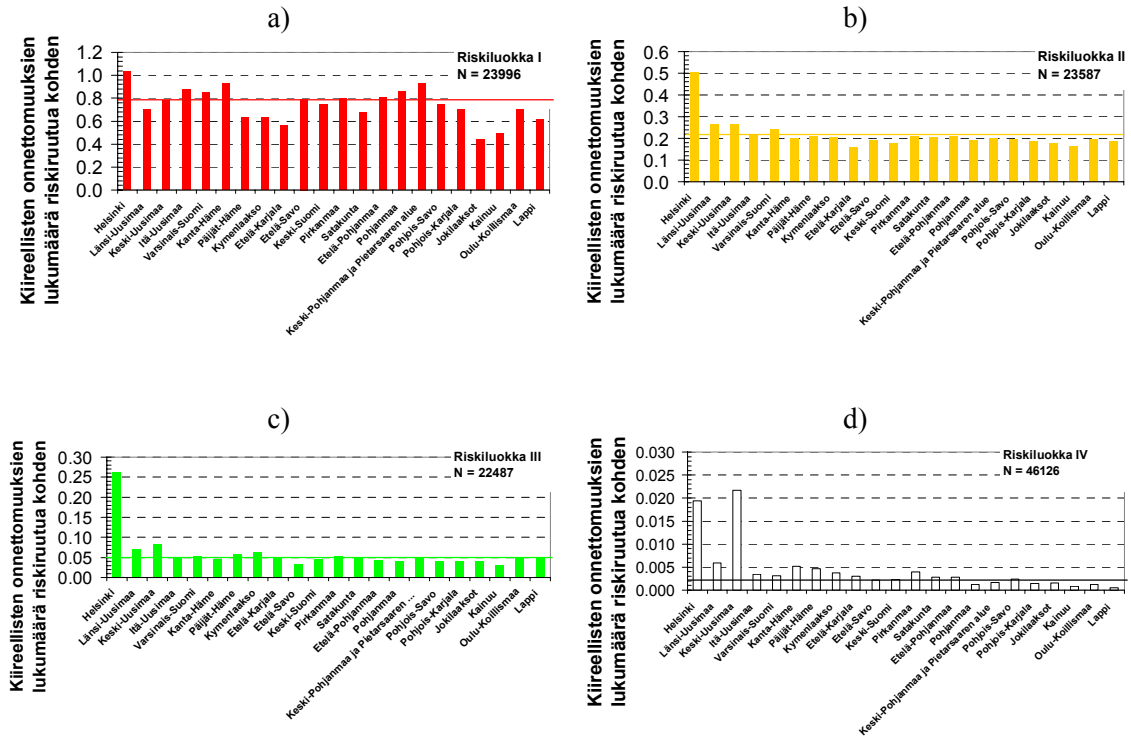


Kuva 2. Kiireellisten onnettomuuksien a) prosentuaalinen lukumäärä, b) lukumäärä riskiruutua kohden vuodessa ja c) lukumäärä tuhatta asukasta kohden vuodessa.

2.2 Eri pelastustoimen alueet

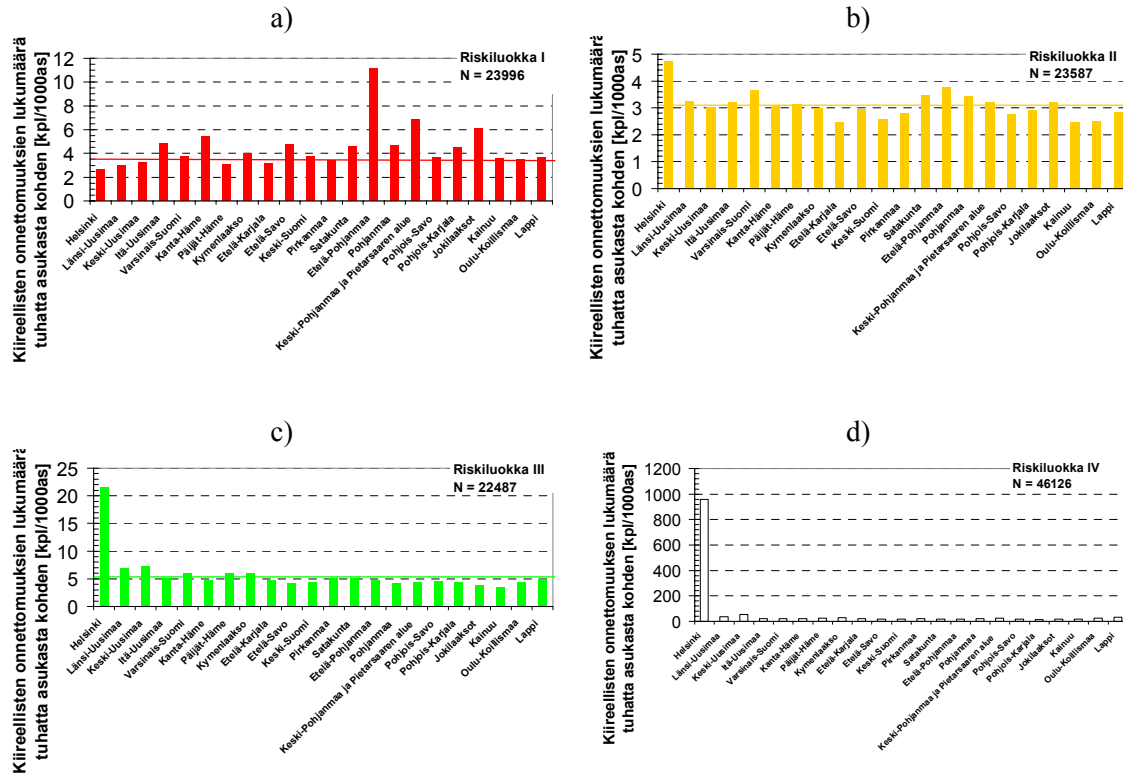
Kuvissa 3 ja 4 on esitetty onnettomuustiheydet riskiruutua (kuva 3) ja tuhatta asukasta (kuva 4) kohden pelastustoimen alueittain. Kuviin piirretty vaakaviiva kuvaa koko maan keskimääräistä onnettomuustiheyttä. Kuten kuvista 3 ja 4 nähdään, eri riskiluokan ruutujen onnettomuustiheyksissä on jonkin verran alueellisia eroja.

2. Toteutunut onnettomuustiheys eri riskiluokan ruuduissa vuosina 2005–2008



Kuva 3. Kiireellisten onnettomuuksien vuotuinen lukumäärä riskiruutua kohden eri pelastustoimen alueilla a) riskiluokan I, b) riskiluokan II, c) riskiluokan III ja d) riskiluokan IV ruuduissa. Vaakaviiva kuvaa koko maan keskiarvoa.

2. Toteutunut onnettomuustiheys eri riskiluokan ruuduissa vuosina 2005–2008



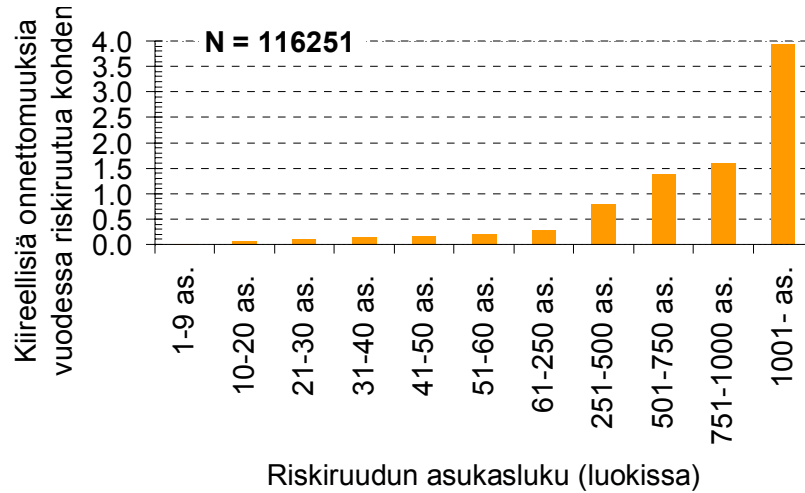
Kuva 4. Kiireellisten onnettomuuksien vuotuinen lukumäärä tuhatta asukasta kohden eri pelastustoimen alueilla a) riskiluokan I, b) riskiluokan II, c) riskiluokan III ja d) riskiluokan IV ruuduissa.

2.3 Lukumäärien riippuvuus asukasluvusta ja kerrosalasta

2.3.1 Kiireelliset onnettomuudet asukasluvukittain

Kuvassa 5 riskiruudut on jaoteltu asukasluvun mukaisesti luokkiin ja kullekin luokalle on määritetty kiireellisten onnettomuuksien lukumäärät riskiruutua kohden. Toisin sanoen siinä on esitetty, kuinka paljon kiireellisiä onnettomuuksia eri luokkien yhdessä yksittäisessä ruudussa on vuosittain tapahtunut (tarkasteluajavälillä 2005–2008). Kuvasta 5 nähdään, että onnettomuustiheys kasvaa asukasluvun kasvaessa.

2. Toteutunut onnettomuustiheys eri riskiluokan ruuduissa vuosina 2005–2008

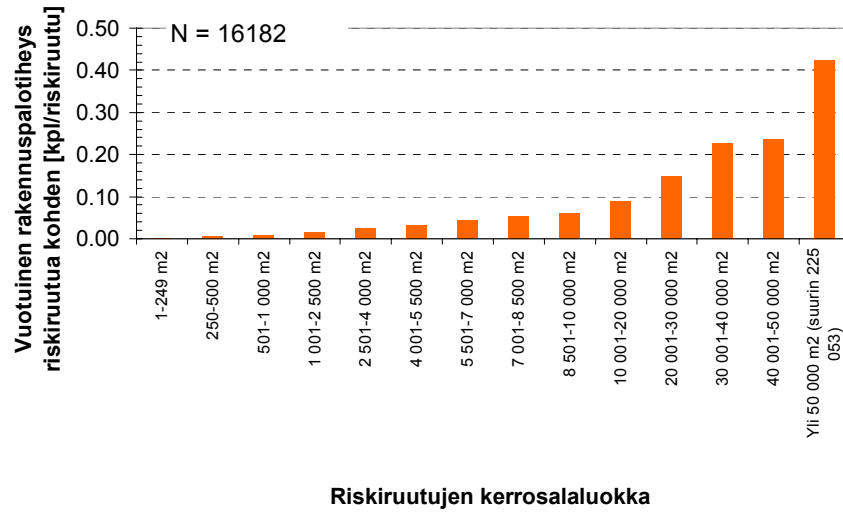


Kuva 5. Kiireellisten onnettomuuksien lukumäärä riskiunitua kohden, kun riskiruudut on jaoteltu luokkiin asukasluvun perusteella.

2.3.2 Rakennuspalot kerrosalaluokittain

Kuvassa 6 riskiruudut on jaoteltu kerrosalan mukaisesti luokkiin ja kullekin luokalle on määritetty rakennuspalojen lukumäärät riskiunitua kohden. Kuvassa 6 on toisin sanoen esitetty se, kuinka paljon rakennuspaloja eri luokkien yhdessä yksittäisessä ruudussa on tapahtunut vuotta kohden (tarkasteluaikavälillä 2005–2008). Kuvasta 6 nähdään, että rakennuspalotiheys kasvaa kerrosalan kasvaessa.

2. Toteutunut onnettomuustiheys eri riskiluokan ruuduissa vuosina 2005–2008



Kuva 6. Rakennuspalojen lukumäärä riskiruutua kohden, kun riskiruudut on jaoteltu kerrosalan perusteella luokkiin.

3. Yksinkertainen regressiomalli

3.1 Aineisto

Regressiomallinnuksessa käytettiin pääasiassa $1 \text{ km} \times 1 \text{ km}$ -ruutuaineistoa. Mallinnus aloitettiin $250 \text{ m} \times 250 \text{ m}$ -aineistolla, mutta analysointi osoittautui ongelmalliseksi aineiston suuren koon ja sen ominaispiirteiden vuoksi. Osa rakennuspaloanalyyseistä kuitenkin onnistuttiin tekemään tällä aineistolla.

3.1.1 Rakennuspalot

Rakennuspalomalliin valittiin potentiaalisiksi selittäjiksi asukasluku ja kerrosala. Malliin otettiin mukaan vain ne ruudut, joissa rakennustietokannan mukaan oli rakennuksia, joiden kerrosala tiedettiin. Tämän seurauksena osa rakennuspaloista jäi pois, koska ne olivat ruudussa, jossa ei rakennustietokannan mukaan ollut rakennuksia lainkaan, tai koska kerrosala ei ollut tiedossa (taulukko 5). Asukasluvun ja kerrosalan tunnuslukuja on esitetty taulukossa 6.

3. Yksinkertainen regressiomalli

Taulukko 5. Rakennuspalomäärät vuosina 2005–2007.

Alue	Koko data		Rakennus olemassa		Kerrosala tiedossa	
	Ruutuja (n)	Paloja (n)	Ruutuja (n)	Paloja (n)	Ruutuja (n)	Paloja (n)
Kaikki alueet	382 197	11 648	119 424	11 253	116 120	11 213
1	10 862	1 905	5 929	1 885	5 790	1 878
2a	59 429	3 551	31 098	3 465	30 290	3 458
2b	60 076	1 893	21 494	1 817	21 085	1 811
3a	61 434	1 700	28 643	1 626	27 541	1 615
3b	20 744	886	9 412	842	9 106	839
4	68 707	1 100	14 420	1 052	14 103	1 048
5	101 665	613	8 428	566	8 205	564

Taulukko 6. Kuvailevia tunnuslukuja asukasluvusta ja kerrosalasta.

Alue	Asukasluku (n)		Kerrosala (m ²)	
	md	min–max	md	min–max
Kaikki alueet	5	0–19 990	329	7–2 178 000
1	13	0–19 990	860	8–2 178 000
2a	6	0–8 779	412	7–1 217 000
2b	6	0–5 632	358	7–675 400
3a	4	0–6 882	245	7–727 700
3b	6	0–5 952	428	7–518 400
4	4	0–5 076	268	7–617 300
5	3	0–4 541	219	7–475 200

3.1.2 Kiireelliset tehtävät

Kiireellisten tehtävien malliin valittiin selittäjäksi vain asukasluku. Taulukossa 7 on tunnuslukuja kiireellisistä tehtävistä.

3. Yksinkertainen regressiomalli

Taulukko 7. Tunnuslukuja kiireellisistä tehtävistä vuosina 2005–2007.

Alue	Ruutuja (n)	Tehtäviä (n)	Tehtäviä (max/ruutu)
Kaikki alueet	382 917	85 133	185
1	10 862	16 717	182
2a	59 429	24 980	185
2b	60 076	12 862	148
3a	61 434	10 851	100
3b	20 744	7 257	71
4	68 707	7 857	104
5	101 665	4 609	106

3.2 Mallin kuvaus

Mallin määrittelyssä jouduttiin niukkojen laskentaresurssien vuoksi käyttämään alueen 2a 1 km×1 km -aineistoa. Alue 2a valittiin, koska sen katsottiin edustavan hyvin sekä hieman tiheämpään että harvempaan asuttua osaa Suomesta. Ensin tarkasteltiin ennalta valittujen selittäjien ja selitettävän lukumäärämuuttujan välisen yhteyden muotoa ns. yleistetyllä additiivisella poissonmallilla (Hastie & Tibshirani 1990). Epälineaariset yhteydet linearisoitiin logaritmuunnoksella ja malliin tulevat termit valittiin yksinkertaisemmalla poissonregressiomallilla. Lopuksi poissonmallia vertailtiin viiteen vaihtoehtoiseen mallityyppiin, joista valittiin tiettyjen kriteerien perusteella paras malli. Malli määriteltiin erikseen rakennuspaloilte ja kiireellisille tehtäville. Mallinnuksessa käytettiin ilmaista R tilasto -ohjelmistoa (R Development Core Team 2009).

3.2.1 Yhteyden muoto ja mallin termien valinta

Regressiomallinnuksessa on tärkeää, että selittävien ja selitettävän muuttujan välinen yhteyden muoto on mallinnettu mahdollisimman oikealla tavalla. Raportissa käytetyt mallit olettavat, että selittäjän yhteys selitettävään muuttujaan on lineaarinen. Tätä lineaarisuusoletusta tarkasteltiin ns. yleistetyllä additiivisella

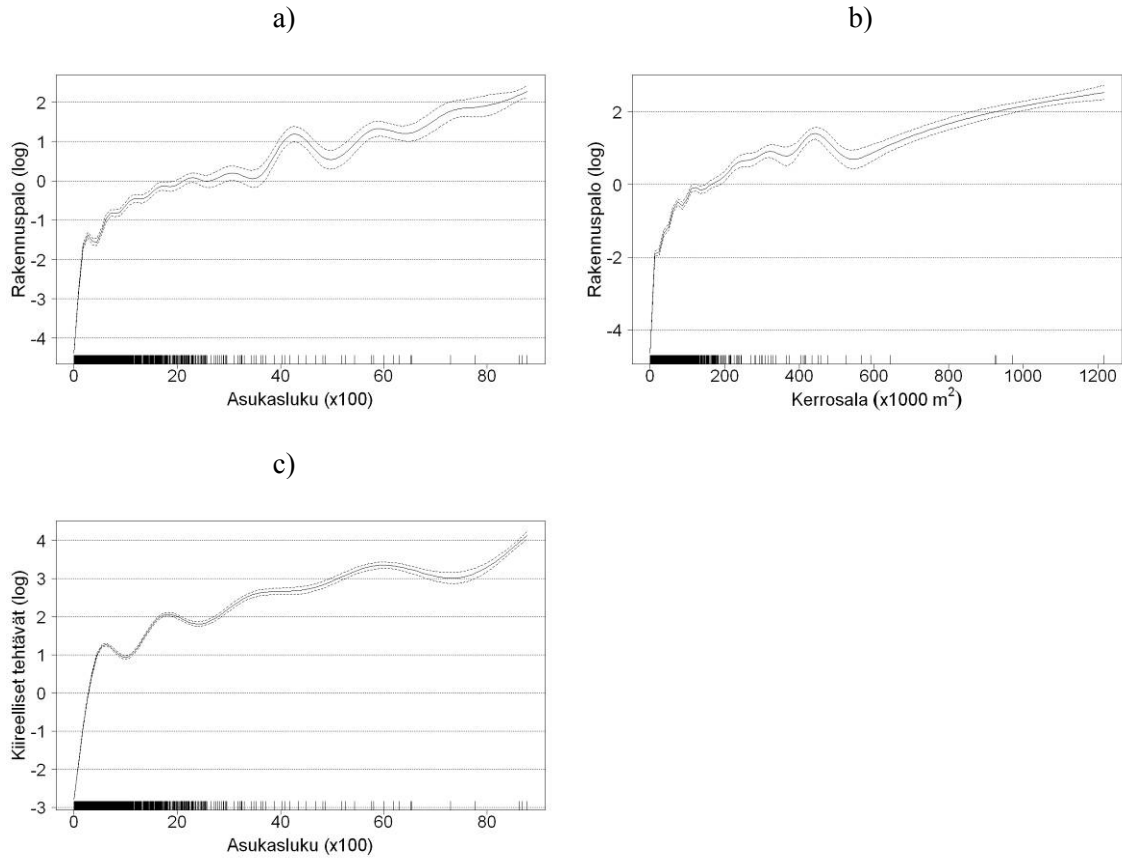
3. Yksinkertainen regressiomalli

poissonmallilla, joka itse joustavasti määrittelee tilanteeseen parhaiten sopivan yhteyden muodon. Tämä tasoite voi olla hyvinkin epälineaarinen, mutta sopivalta muunnoksella se usein pyritään ja pyrittiin nytkin linearisoimaan. Tarkastelut tehtiin R:n mgcv-pakettiin kuuluvalla gam-funktiolla (Wood 2008).

Asukasluvun ja kerrosalan yhteyttä rakennuspaloihin tarkasteltiin ensin erikseen. Sekä asukasluvun että kerrosalan yhteys rakennuspaloihin näytti log-lineariaariselta (kuvat 7a ja 7b). Tämän vuoksi molempiin selittäjiin tehtiin logaritimuunnos (luonnollinen logaritmi). Ennen logaritimuunnosta asukaslukuun lisättiin yksi, jotta myös asumattomat ruudut saatiin pidettyä mallituksessa mukana. Yhdenmukaisuuden vuoksi kerrosalaan tehtiin vastaava muunnos, vaikka se olisikaan muuten ollut tarpeen. Myös kiireellisissä tehtävissä asukaslukuun tehtiin vastaava log-muunnos (kuva 7c).

Rakennuspalomallin termien valinnassa käytettiin perinteistä poissonmallia. Ensinnä verrattiin mallia, jossa oli sekä log-muunnatut asukasluku- että kerrosalamuuttajat, sellaisiin malleihin, joissa nämä selittäjät olivat yksinään. Malleja vertailtiin devianssitestillä toisiinsa, ja malli, jossa molemmat selittäjät olivat mukana, sopi aineistoon selvästi paremmin. Tämän jälkeen malliin lisättiin vielä asukasluvun ja kerrosalan yhteisvaikutustermi. Yhteisvaikutustermi kuvaa sitä, että asukasluvun yhteys rakennuspaloihin riippuu siitä, mitä arvoja kerrosala saa. Tämä malli osoittautui vielä paremmaksi, joten se valittiin jatkotarkasteluun, jossa poissonmallia vertailtiin vaihtoehtoisiin, hieman monimutkaisempiin malleihin.

3. Yksinkertainen regressiomalli



Kuva 7. a) Asukasluvun ja rakennuspalon välinen yhteys, b) kerrosalan ja rakennuspalon välinen yhteys ja c) asukasluvun ja kiireellisten tehtävien välinen yhteys.

3.2.2 Mallin valinta

Koska selitettävä muuttuja oli lukumäärämuotoinen eikä noudattanut normaali-jakaumaa lähimainkaan, mallituksen lähtökohdaksi otettiin poissonjakaumaan perustuva regressiomalli. Klassinen poissonmalli olettaa, että lukumäärämuuttujan keskiarvo ja varianssi ovat samat. Tämä ei kuitenkaan pitänyt paikkaansa tämän aineiston kohdalla. Näin ollen aineistoon sovitettiin negatiiviseen binomijakaumaan perustuva malli R:n MASS-kirjaston `glm.nb`-funktiota käyttäen (Venables & Ripley 2002).

3. Yksinkertainen regressiomalli

Aineistossa oli lisäksi erittäin suuri määrä ruutuja, joissa ei ollut tapahtunut lainkaan onnettomuuksia (rakennuspaloista 97,3 % ja kiireellisistä tehtävistä 96,2 % ruuduista), mikä herätti epäilyjä negatiivisen binomimallin riittävydestä. Niinpä oli mietittävä myös muita vaihtoehtoisia malleja. Kirjallisuuden perusteella vertailtaviksi malleiksi valittiin ns. zero-inflated negatiivinen binomi -malli ja hurdle negatiivinen binomi -malli (Zeileis et al. 2008). Sekä zero-inflated että hurdle-malleista koekiltiin myös poissonjakaumaan perustuvaa mallia.

Näiden kuuden eri mallin sopivuutta aineistoon vertailtiin ns. Vuongin testillä (Vuong 1989) sekä tarkastelemalla ennustettuja kokonaismääriä ja keskimääräistä ennustevirhettä. Lisäksi valittavan mallin tuli olla riittävän yksinkertainen. Rakennuspalojen osalta paras malli oli negatiivinen binomi -malli. Tämä malli on muodoltaan yksinkertaista esitettynä seuraava:

$$\log(y) = \alpha + \beta_1 \log(x_1 + 1) + \beta_2 \log(x_2 + 1) + \beta_3 \log(x_1 + 1) \log(x_2 + 1) \quad (2)$$

missä y = rakennuspalojen lukumäärä, $\log(x_1+1)$ = luonnollinen logaritmi (asukasluku+1):stä ja $\log(x_2+1)$ = luonnollinen logaritmi (kerrosala+1):stä. Kerrosala on alun perin neliömetreinä. α = mallin vakiotermi, β_1 , β_2 , ja β_3 = regressiokertoimet.

Hurdle negatiivinen binomi -malli sopi Vuongin testin mukaan parhaiten kiireellisten tehtävien selittämiseen, mutta se aliarvioi suuremmat onnettomuusmäärät selvästi. Negatiivinen binomi -malli sopi aineistoon Vuongin testin mukaan toiseksi parhaiten: se ennusti kiireellisiä tehtäviä kokonaisuudessaan paremmin kuin hurdle negatiivinen binomi -malli. Keskimääräiset ennustevirheet olivat hyvin lähellä hurdle negatiivisen binomi -mallin ennustevirheitä, ja osin jopa pienempiä. Edellä mainittujen syiden sekä yksinkertaisuuden ja osin yhteneväisyydenkin vuoksi negatiivinen binomi -malli valittiin lopulliseksi malliksi. Yksinkertaista esitettynä se on seuraava:

$$\log(y) = \alpha + \beta_1 \log(x_1 + 1) \quad (3)$$

missä y = kiireellisten tehtävien lukumäärä, $\log(x_1+1)$ = luonnollinen logaritmi (asukasluku+1):stä. α = mallin vakiotermi ja β_1 = regressiokerroin.

Malleissa on huomioitu myös se, että ennustemallin määrittelyssä käytetyt onnettomuusmäärät ovat kolmen vuoden ajalta. Jotta mallin antama ennuste kuvai-

3. Yksinkertainen regressiomalli

si yhtä vuotta, mallissa on ns. offset-termi, joka tässä tapauksessa on vakioarvo ($=\log(\text{vuosien lkm})=\log(3)$).

3.3 Tulokset

Pelastustoimen alueista muodostetuille seitsemälle isommalle alueelle estimoitiin aluekohtaiset mallit ja yhteinen malli. Aluekohtaisilla malleilla pyrittiin tarkastelemaan mahdollisia alueiden välisiä eroja ja arvioimaan yhteisen mallin sopivuutta eri alueille.

3.3.1 Rakennuspalot

Regressiomallin tulokset $1 \text{ km} \times 1 \text{ km}$ ja $250 \text{ m} \times 250 \text{ m}$ -aineistoilla on esitetty taulukossa 8. Tekstissä käsitellään $1 \text{ km} \times 1 \text{ km}$ -aineiston tuloksia, ellei muuta ole mainittu.

Aluekohtaisten mallien regressiokertoimet muistuttavat paljon toisiaan, joten alueiden yhteinen malli näyttäisi siten toimivalta. Kerrosalan ja yhteisvaikutustermin regressiokertoimet ovat kaikissa malleissa positiivisia ja alueiden 4 ja 5 yhteisvaikutustermiä lukuun ottamatta tilastollisesti merkitseviä ($p < 0,05$). Asukasluvun regressiokerroin puolestaan on alueen 4 mallia lukuun ottamatta negatiivinen mutta tilastollisesti merkitsevä ($p < 0,05$) vain alueiden 1, 2a ja 3a malleissa sekä yhteisessä mallissa. Kertoimien suuruus vaihtelee jonkin verran, mutta tämä oli odotettavissa. Alueen 1 malli selittää noin 68 % rakennuspalomäärän vaihtelusta, kun muut aluekohtaiset mallit selittävät 40–50 % rakennuspalomäärän vaihtelusta. Tämä johtuu pitkälti siitä, että alueella 1 on ruututasolla eniten vaihtelua rakennuspalomäärissä ja selittäjissä. Myös ruutuja, joissa paloja ei ole lainkaan, on alueella 1 suhteessa vähemmän kuin muilla alueilla. Mallinnuksessa on täten enemmän voimaa.

3. Yksinkertainen regressiomalli

Taulukko 8. Rakennuspalojen regressiomallien tulokset malleista, joissa alueet yhdessä ja erikseen.

	Vakio		Asukasluku		Kerrosala		Yhteisvaikutus		Selitysaste
	α	se	β_1	se	β_2	se	β_3	se	D %
1 km × 1 km									
Alueet yhdessä	-8,38	0,08	-0,159	0,027	0,65	0,012	0,020	0,002	49,8 %
1	-7,81	0,32	-0,125	0,073	0,54	0,041	0,025	0,006	67,9 %
2a	-7,86	0,16	-0,296	0,051	0,59	0,024	0,034	0,004	49,6 %
2b	-8,25	0,22	-0,110	0,067	0,62	0,033	0,019	0,006	41,2 %
3a	-8,70	0,20	-0,246	0,070	0,73	0,032	0,022	0,006	41,4 %
3b	-7,60	0,32	-0,066	0,103	0,48	0,048	0,025	0,009	42,0 %
4	-9,00	0,25	0,004	0,081	0,73	0,036	0,002	0,007	45,1 %
5	-9,68	0,35	-0,175	0,127	0,86	0,049	0,010	0,012	49,3 %
250 m × 250 m									
Alueet yhdessä	-9,11	0,07	-0,249	0,033	0,67	0,010	0,033	0,004	27,4%

α , β_1 , β_2 , ja β_3 = mallin regressiokertoimet, se = regressiokertoimen hajonta, D % = devianssista selittyvä prosenttiosuus.

Yksittäisen regressiokertoimen tulkinta on hankalaa, koska asukasluvun ja kerrosalan välillä on voimakas korrelaatio ($r \sim 0,9$). Tästä seuraa mallinnukseen ns. kollineaarisuusongelma, eli malli ei kykene erottelemaan voimakkaasti korreloituneiden selittäjien vaikutusta. Sekä asukasluvulla että kerrosalalla on yksin mallissa ollessaan tilastollisesti merkitsevä, vahva positiivinen yhteys rakennuspaloihin. Kun asukasluku ja kerrosala laitetaan malliin yhtä aikaa, kerrosalan regressiokerroin pysyy lähes ennallaan mutta asukasluvun kerroin pienenee selvästi pysyen kuitenkin edelleen positiivisena. Kun malliin lisätään vielä asukasluvun ja kerrosalan yhteisvaikutustermi, asukasluvun kerroin kääntyy negatiiviseksi. Tilastollisesti merkitsevä yhteisvaikutustermi parantaa kuitenkin mallin ennustekykä, joten se on jätetty malliin. Regressiokertoimien tulkintaa hankaloittaa hieman myös se, että kerrosalan yhteys rakennuspaloihin riippuu asukas-

3. Yksinkertainen regressiomalli

luvusta (ja päinvastoin). Positiivinen yhteisvaikutustermi kertoo, että kerrosalan yhteys rakennuspaloihin on voimakkaampi silloin, kun asukasluku on suurempi (ja päinvastoin). Kollineaarisuusongelmasta huolimatta mallin ennusteisiin voi luottaa. Tosin ennusteita laadittaessa on muistettava antaa arvo sekä asukasluvulle ja kerrosalalle.

250 m × 250 m -aineistolla regressiokertoimet ovat hyvin samanlaiset kuin 1 km × 1 km -aineistolla. Mallin selitysaste on kuitenkin selvästi alhaisempi, noin 27 %.

3.3.2 Kiireelliset tehtävät

Regressiomallin tulokset on esitetty taulukossa 9.

Taulukko 9. Kiireellisten tehtävien regressiomallien tulokset malleista, joissa alueet yhdessä ja erikseen.

	Vakio		Asukasluku		Selitysaste
	α	se	β_1	se	D %
Kaikki alueet	-4,39	0,013	0,85	0,005	58,7 %
1	-3,32	0,063	0,69	0,016	65,5 %
2a	-4,02	0,025	0,77	0,008	57,2 %
2b	-4,26	0,032	0,78	0,011	51,4 %
3a	-4,32	0,026	0,80	0,011	45,1 %
3b	-3,80	0,044	0,71	0,016	47,9 %
4	-4,61	0,028	0,85	0,011	51,9 %
5	-5,22	0,046	1,08	0,028	46,1 %

α ja β_1 = mallin regressiokertoimet, se = regressiokertoimen hajonta, D % = devianssista selittyvä prosenttiosuus.

Aluekohtaisten mallien regressiokertoimet muistuttavat paljon toisiaan, joten alueiden yhteinen malli näyttäisi tämän puolesta toimivalta. Asukasluvun regressiokerroin on kaikissa malleissa positiivinen ja tilastollisesti merkitsevä ($p < 0,05$). Kertoimien suuruus vaihtelee jonkin verran, mutta tämä oli odotettavissa. Alueen 1 malli selittää noin 66 % kiireellisten tehtävien määrän vaihtelus-

3. Yksinkertainen regressiomalli

ta, kun muut aluekohtaiset mallit selittävät 45–57 % rakennuspalojen määrän vaihtelusta. Tämä selittyy pitkälti sillä, että alueella 1 on ruututasolla eniten vaihtelua rakennuspalomäärissä ja asukasluvussa. Myös ruutuja, joissa ei ole kiireellisiä tehtäviä lainkaan, on alueella 1 suhteessa vähemmän kuin muilla alueilla. Mallinnuksessa on täten enemmän voimaa.

3.4 Mallin toimivuuden arviointi

Tässä luvussa tarkastellaan alueiden yhteisen mallin toimivuutta. Aluekohtaiset tiedot perustuvat siis alueiden yhteiseen malliin.

3.4.1 Rakennuspalot

Mallin toimivuutta arvioidaan erikseen mallittamiseen käytettyjen vuosien 2005–2007 osalta ja varsinaiseen ennustamiseen käytetyn vuoden 2008 osalta. Vuosien 2005–2007 osalta malli ennustaa rakennuspalojen kokonaismäärää erinomaisesti, mutta vuoden 2008 osalta selvästi alakanttiin (taulukko 10). Tämä selittyy sillä, että vuonna 2008 oli selvästi enemmän rakennuspaloja kuin vuosina 2005–2007 keskimäärin. Malli ei pysty huomioimaan tällaista ajassa tapahtuvaa muutosta vaan olettaa, että rakennuspalojen kokonaismäärä vuonna 2008 vastaa kolmen edellisen vuoden määriä. Alueiden yhteinen malli ennustaa aluekohtaisia kokonaismääriä kohtalaisen hyvin, vaikka pieniä alueellisia eroja onkin havaittavissa. Malli käyttäytyy 250 m × 250 m -aineistolla hyvin samalla tavalla kuin 1 km × 1 km -aineistolla (taulukko 11).

3. Yksinkertainen regressiomalli

Taulukko 10. Rakennuspalojen havaitut ja ennustetut kokonaismäärät (1 km × 1 km).

Alue	Vuodet 2005–2007			Vuosi 2008		
	Havaittu (n)	Ennustettu (n)	Ero (%)	Havaittu (n)	Ennustettu (n)	Ero (%)
Alueet yhdessä	11 213	11 165	–0,4 %	4 349	3 722	–14,4 %
1	1 878	2 096	11,6 %	931	699	–24,9 %
2a	3 548	3 315	–6,6 %	1 264	1 105	–12,6 %
2b	1 811	1 748	–3,5 %	673	583	–13,4 %
3a	1 615	1 514	–6,3 %	595	505	–15,1 %
3b	839	935	11,4 %	315	312	–1,0 %
4	1 048	1 055	0,7 %	347	352	1,4 %
5	564	501	–11,2 %	224	167	–25,4 %

Taulukko 11. Rakennuspalojen havaitut ja ennustetut kokonaismäärät (250 m × 250 m).

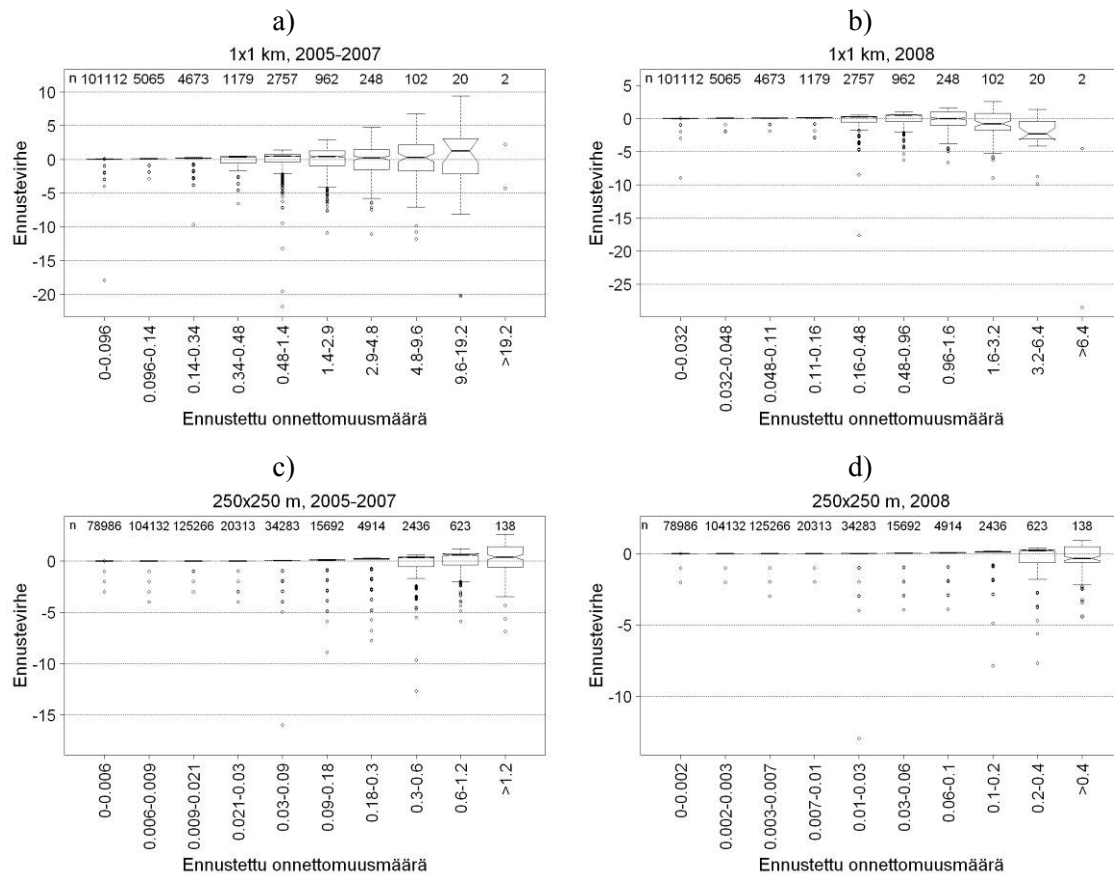
Alue	Vuodet 2005–2007			Vuosi 2008		
	Havaittu (n)	Ennustettu (n)	Ero (%)	Havaittu (n)	Ennustettu (n)	Ero (%)
Alueet yhdessä	9 881	9 873	–0,1 %	3 926	3 290	–16,2 %
1	1 753	1 973	12,5 %	885	658	–25,6 %
2a	3 019	2 962	–1,9 %	1 137	987	–13,2 %
2b	1 621	1 522	–6,1 %	606	507	–16,3 %
3a	1 336	1 256	–6,0 %	513	419	–18,3 %
3b	742	831	12,0 %	278	277	–0,4 %
4	931	909	–2,4 %	305	303	–0,7 %
5	479	420	–12,3 %	202	140	–30,7 %

Mallin ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan luokiteltuna (kuva 8a) paljastaa, että malli ennustaa rakennuspaloja vuosina 2005–

3. Yksinkertainen regressiomalli

2007 keskimäärin melko hyvin. Ennustevirheen hajonta tosin kasvaa siirryttäessä korkeampiin ennustemääriin. Jokaisessa luokassa on lisäksi joitakin ruutuja, joissa ennustettu määrä on selvästi liian alhainen.

Ennustettaessa vuoden 2008 rakennuspaloja keskimääräinen ennustevirhe on neljässä alimmassa ennusteluokassa lähellä nollaa (kuva 8b). Sen sijaan kahdessa keskimmälisessä luokassa ennustetaan keskimäärin hieman liikaa ja neljässä korkeimmassa ennusteluokassa liian vähän paloja. Tämä johtuu pitkälti siitä, että malli ei huomioi sitä, että vuonna 2008 paloja oli selvästi enemmän kuin aikaisemmin. Jokaisessa luokassa on lisäksi joitakin ruutuja, joissa ennustettu määrä on selvästi liian alhainen. 250 m × 250 m -aineistolla tilanne on samanlainen (kuvat 8c ja 8d).



Kuva 8. a) Ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan. Vuodet 2005–2007 (1 km × 1 km), b) ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan, vuosi 2008 (1 km × 1 km), c) ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan, vuodet 2005–2007 (250 m × 250 m), d) ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan, vuosi 2008 (250 m × 250 m).

3. Yksinkertainen regressiomalli

Toimintavalmiusohjeen A:71 ja mallin tapaa luokitella ruutuja riskiluokkiin vertailtiin (taulukko 12). Vertailua varten ruudut jaettiin riskiluokkiin toisaalta mallin antamien ennusteiden ja toisaalta toimintavalmiusohjeen mukaan (jakotapa kuvattu luvussa 1.5.3). Malli luokittelee systemaattisesti ruutuja alempaan riskiluokkaan kuin toimintavalmiusohje. Se luokittelee esimerkiksi riskiluokkaan I vain 86 ruutua, kun toimintavalmiusohjeen mukaan riskiluokkaan I kuuluu 332 ruutua. Jako riippuu kuitenkin pitkälti siitä, millaisia riskiluokkarajoja käytetään. Jos mallin ennusteet luokiteltaisiin riskiluokkiin kerrosalaan perustuvia luokkarajoja käyttäen, riskiluokkaan I kuuluisi 306 ruutua (taulukko 13). Kerrosalaluokkarajoilla malli luokittelee riskiluokkiin II ja III jonkin verran enemmän ruutuja kuin toimintavalmiusohje. 250 m × 250 m -aineistolla tilanne on samanlainen kuin 1 km × 1 km -aineistolla (vastaavia taulukoita ei ole esitetty raportissa).

Vuonna 2008 644 ruudussa (0,55 % kokonaismäärästä) tapahtui vähintään kaksi rakennuspaloa (taulukko 14). Näitä ruutuja voi toimintavalmiusohjeen mukaan pitää riskiluokkaan I kuuluviksi. Toimintavalmiusohjeen mukaan riskiluokkaan I kuuluvia ruutuja piti olla vain 0,28 % (n = 332). Malli puolestaan ennusti asukasluukuun perustuvia riskiluokkarajoja käytettäessä, että vastaavia ruutuja olisi vielä vähemmän, vain 0,07 % (n = 86). Kerrosalaan perustuvia riskiluokkarajoja käytettäessä malli ennusti, että I riskiluokan ruutuja olisi 0,26 % (n = 306).

Vaikka tästä tarkastelusta ei yhden vuoden datan ja luokitteluun liittyvien ongelmien vuoksi voi vetää pitkälle meneviä johtopäätöksiä, sekä toimintavalmiusohjeen että mallin mukaan riskiluokkaan I kuuluvia ruutuja on vain puolet siitä, mitä toteuma vuonna 2008 antaa odottaa. 250 m × 250 m -aineistolla tilanne on samanlainen kuin 1 km × 1 km -aineistolla (vastaavaa taulukkoa ei ole esitetty raportissa).

3. Yksinkertainen regressiomalli

Taulukko 12. Ruutujen luokittelu riskiluokkiin toimintavalmiusohjeen A:71 ja mallin perusteella (asukaslukurajoilla).

	Malli A				
A:71	I	II	III	IV	Yht.
I	86	246	0	0	332
II	0	1 002	982	0	1 984
III	0	0	4 338	2 583	6 921
IV	0	0	0	106 883	106 883
Yht.	86	1 248	5 320	109 466	116 120

Taulukko 13. Ruutujen luokittelu riskiluokkiin toimintavalmiusohjeen A:71 ja mallin perusteella (kerrosalarajoilla).

	Malli B				
A:71	I	II	III	IV	Yht.
I	301	31	0	0	332
II	5	1 960	19	0	1 984
III	0	471	6 450	0	6 921
IV	0	0	9 640	97 243	106 883
Yht.	306	2 462	16 109	97 243	116 120

Taulukko 14. Riskiluokan I ruutujen määrä vuoden 2008 toteuman mukaan (vähintään 2 paloa), toimintavalmiusohjeen A:71 mukaan ja mallin mukaan (Malli A = asukaslukurajoilla, Malli B = kerrosalarajoilla).

	Toteutunut	A:71	Malli A	Malli B
Alueet yhdessä	0.55 %	0.28 %	0.07 %	0.26 %

Kerrosalan ja asukasluvun välisissä sirontakuivissa havainnot on (= ruudut) eroteltu vuonna 2008 toteutuneiden rakennuspalojen mukaan (kuvat 9a ja 9b). Ku-

3. Yksinkertainen regressiomalli

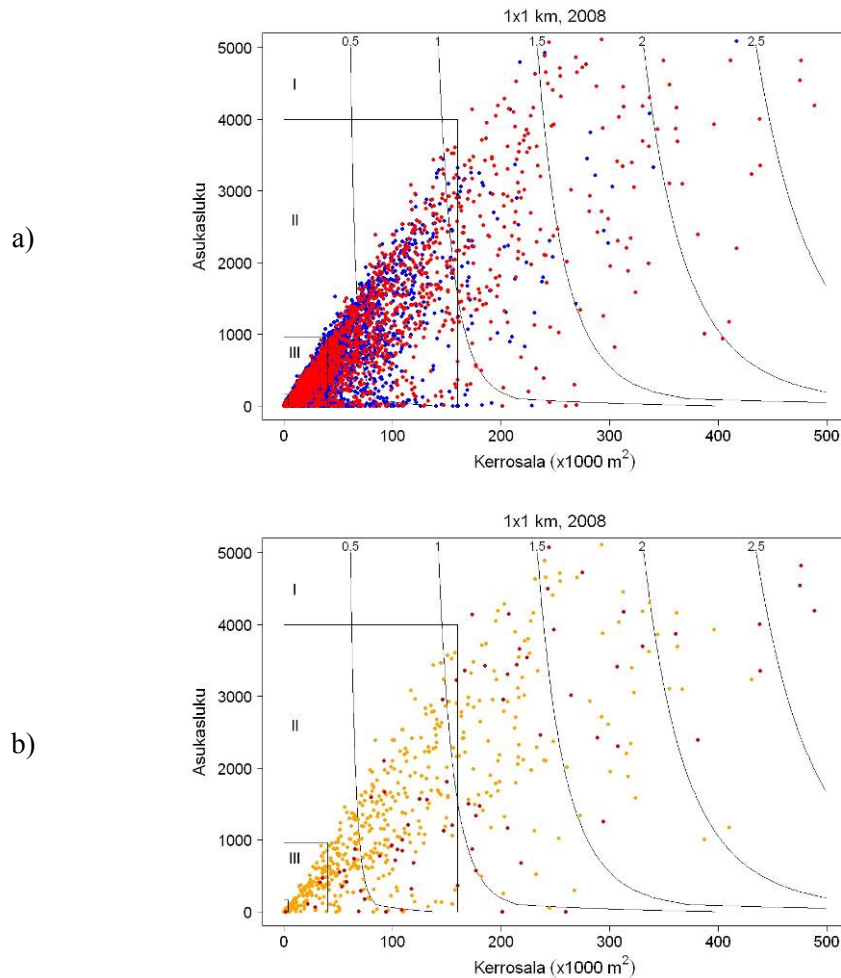
viin on lisäksi piirretty toimintavalmiusohjeen mukaiset riskiluokkarajat ja viisi ns. tasa-arvokäyrää, jotka kertovat, millä kerrosalan ja asukasluvun yhdistelmällä malli ennustaa käyrän osoittaman määrän paloja (0,5; 1, 1,5; 2 ja 2,5 paloa). Kuvat kertovat selvästi ensinnäkin sen, että paloja on suhteellisesti eniten riskiluokan I ruuduissa. Toisaalta ruutuja, joissa on useampi kuin yksi palo, on kaikissa riskiluokissa, eli kerrosala ja asukasluku eivät selitä kaikkia rakennuspaloja.

Kuvasta 9a näkee myös sen, miten huonosti toimintavalmiusohjeen riskiluokkarajojen määrittelyn taustalla olevasta asukaslukuun perustuvasta yhtälöstä lasketut riskiluokkarajat luokittelevat mallin ennusteet. Asukaslukuun perustuvan rajan mukaan riskiluokan I ruudussa on $1 \text{ km} \times 1 \text{ km}$ -ruututasolla vähintään 1,92 rakennuspaloa vuodessa. Tätä rajaa käyttäen mallin mukaan riskiluokkaan I kuuluvat karkeasti ottaen ne ruudut, jotka jäävät tasa-arvokäyrän 2 oikealle puolelle. Toisaalta toimintavalmiusohjeen mukaan riskiluokkaan I kuuluvat ne ruudut, joissa on yli 4 000 asukasta tai kerrosalaa yli $160\,000 \text{ m}^2$ (käytännössä kerrosala on määräävä luokittelukriteeri). Tasa-arvokäyrä 1 osuu melko hyvin yksiin em. kerrosalarajan kanssa. Jos ruutujen riskiluokittelussa haluttaisiin siis etsiä sellaiset rajat, jotka luokittelisivat ruudut vastaavalla tavalla kuin toimintavalmiusohje, karkeasti ottaen riskiluokkaan I kuuluisivat ne ruudut, joihin ennustetaan vähintään yksi rakennuspalo vuodessa (eli kaikki ruudut tasa-arvokäyrän 1 oikealla puolella).

Sirontakuvat kertovat selkeästi myös sen, miten erilaisilla kerrosala- ja asukasmäärillä malli ennustaa saman määrän onnettomuuksia. Esimerkiksi tasa-arvokäyrälle 1 osuu ruutu, jossa kerrosalaa on noin $150\,000 \text{ m}^2$ ja asukkaita noin 3 500. Toisaalta samalle käyrälle osuu myös ruutu, jossa kerrosalaa on noin $240\,000 \text{ m}^2$ ja asukkaita vain noin 100. Tasa-arvokäyriä on syytä tarkastella vain siltä osin, miten ne osuvat havaintopilven kanssa päällekkäin. $250 \times 250 \text{ m}$ -aineistolla tilanne on samanlainen kuin $1 \text{ km} \times 1 \text{ km}$ -aineistolla (vastaavia kuvia ei ole esitetty raportissa). Kuvien tulkinta on kuitenkin hankalampaa havaintopisteiden suuren määrän vuoksi.

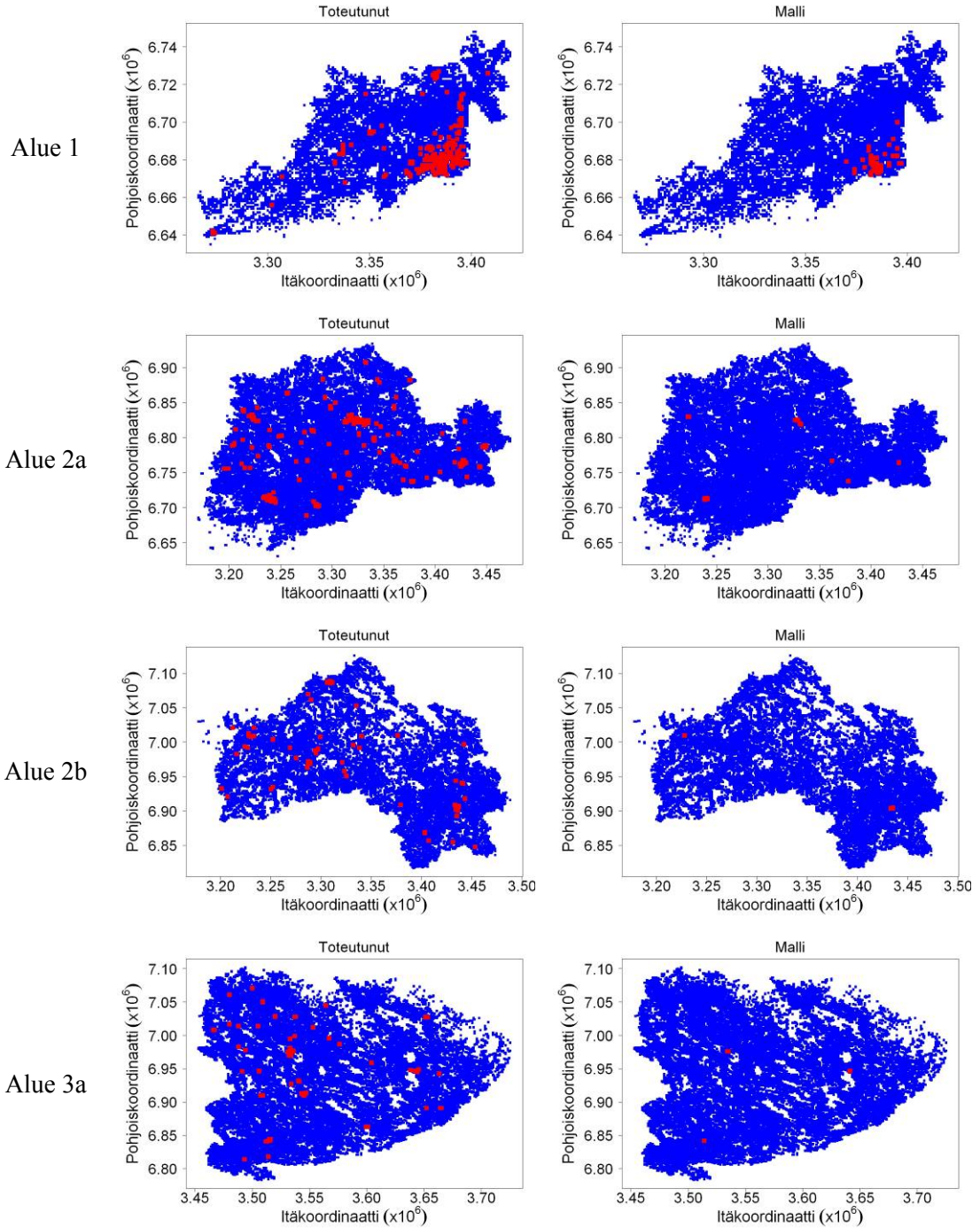
Kuten karttakuvasta 10 hyvin näkee, malli ennustaa selvästi vähemmän yli kaksi rakennuspaloa sisältäviä ruutuja kaikille alueille verrattuna toteutuneeseen tilanteeseen vuonna 2008. Tämä johtuu pitkälti mallin riskiluokkarajojen määrittelyn ongelmallisuudesta, kuten edellä on jo käynyt ilmi. $250 \text{ m} \times 250 \text{ m}$ -aineistolla tilanne on samanlainen kuin $1 \text{ km} \times 1 \text{ km}$ -aineistolla (vastaavia kuvia ei ole esitetty raportissa).

3. Yksinkertainen regressiomalli

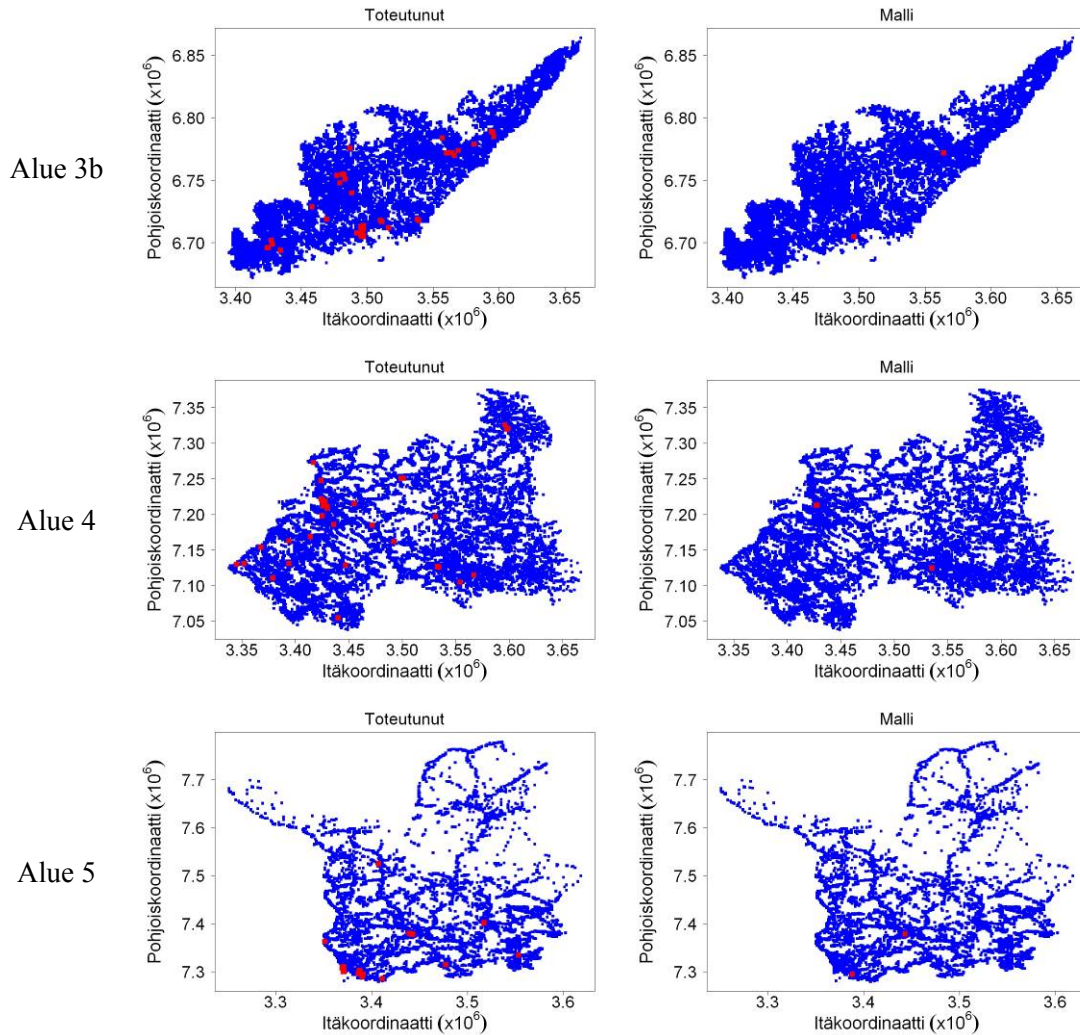


Kuva 9. a) Kerrosalan ja asukasluvun välinen sirontakuva. Sinisellä on erotettu ruudut, joissa ei ole ollut paloja lainkaan, ja punaisella ruudut, joissa on ollut ainakin yksi palo vuonna 2008. Kuvassa näkyy myös toimintavalmiusohjeen mukaiset riskiluokkarajat sekä ns. tasa-arvokäyrät, jotka kuvaavat niitä asukasluvun ja kerrosalan yhdistelmiä, joilla malli ennustaa 0,5, 1, 1,5, 2 ja 2,5 rakennuspaloa. b) Kerrosalan ja asukasluvun välinen sirontakuva. Oranssilla on erotettu ruudut, joissa on ollut kaksi tai kolme paloa, ja punaisella ruudut, joissa on ollut vähintään neljä paloa vuonna 2008. Kuvassa näkyy myös toimintavalmiusohjeen mukaiset riskiluokkarajat sekä ns. tasa-arvokäyrät, jotka kuvaavat niitä asukasluvun ja kerrosalan yhdistelmiä, joilla malli ennustaa 0,5, 1, 1,5, 2 ja 2,5 rakennuspaloa.

3. Yksinkertainen regressiomalli



3. Yksinkertainen regressiomalli



Kuva 10. Suurimmat onnettomuusmäärät (vähintään kaksi onnettomuutta 1 km × 1 km -ruudussa, kartalla punainen väri). Vasemmalla toteutuneet ja oikealla mallin ennustamat onnettomuudet.

3.4.2 Kiireelliset tehtävät

Malli ennustaa rakennuspalojen kokonaismäärää vuosien 2005–2007 osalta heikosti (7,9 % yli toteutuneen määrän) mutta vuoden 2008 osalta erinomaisesti (taulukko 15). Vuoden 2008 ennusteen tarkkuus ei kuitenkaan johdu mallin hy-

3. Yksinkertainen regressiomalli

vyydestä vaan siitä, että mallin ennustama kokonaismäärä käy sattumalta yksiin sen kanssa, että vuonna 2008 kiireellisiä tehtäviä on ollut selvästi enemmän kuin edellisenä kolmena vuonna keskimäärin. Alueiden väliset erot ovat lisäksi suuria sekä vuosina 2005–2007 että vuonna 2008. Malli ennustaa esimerkiksi vuonna 2008 alueelle 5 kiireellisiä tehtäviä yli 45 % toteutunutta määrää enemmän. Alueilla 1 ja 2a ennuste jää puolestaan selvästi liian pieneksi.

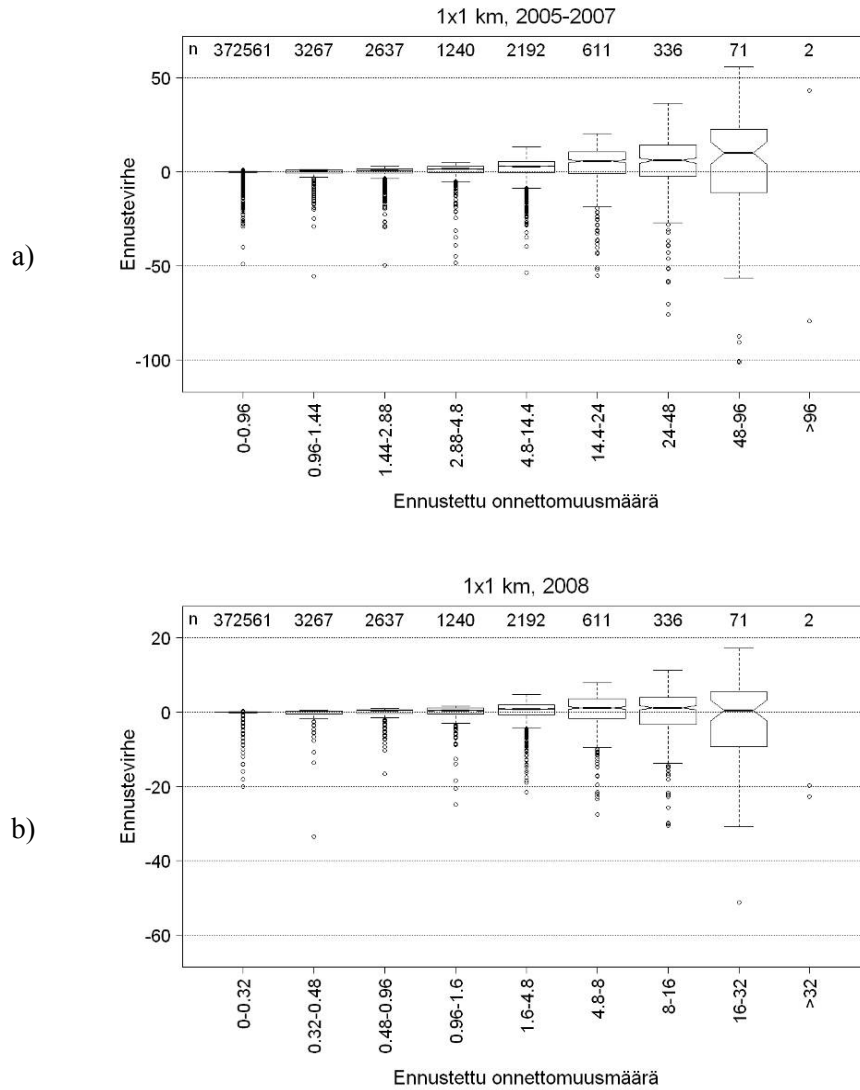
Taulukko 15. Kiireellisten tehtävien havaitut ja ennustetut kokonaismäärät.

Alue	Vuodet 2005–2007			Vuosi 2008		
	Havaittu (n)	Ennustettu (n)	Ero (%)	Havaittu (n)	Ennustettu (n)	Ero (%)
Kaikki alueet	85 133	91 890	7,9 %	30 457	30 630	0,6 %
1	16 717	16 868	0,9 %	6 791	5 623	–17,2 %
2a	24 980	24 905	–0,3 %	9 287	8 302	–10,6 %
2b	12 862	13 964	8,6 %	4 351	4 655	7,0 %
3a	10 851	11 998	10,6 %	3 433	3 999	16,5 %
3b	7 257	7 243	–0,2 %	2 524	2 414	–4,4 %
4	7 857	10 000	27,3 %	2 486	3 333	34,1 %
5	4 609	6 910	49,9 %	1 585	2 303	45,3 %

Mallin ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan eriteltynä kertoo (kuva 11a), että keskimääräinen ennustevirhe kasvaa systemaattisesti siirryttäessä pienemmistä onnettomuusmääristä suurempiin. Samalla ennustevirheen hajonta kasvaa. Jokaisessa luokassa on lisäksi ruutuja, joissa ennustettu määrä on selvästi liian alhainen. Kahdessa ylimmässä luokassa on vain yksi havainto kummassakin, joten ne on syytä jättää tarkastelun ulkopuolelle.

Ennustettaessa vuoden 2008 kiireellisiä tehtäviä keskimääräinen ennustevirhe pysyttelee kaikissa luokissa lähellä nollaa (kuva 11b). Tosin onnettomuusmäärän kasvaessa myös keskimääräinen ennustevirhe kasvaa hieman. Jokaisessa luokassa on lisäksi ruutuja, joissa ennustettu määrä on selvästi liian pieni.

3. Yksinkertainen regressiomalli



Kuva 11. a) Ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan (vuodet 2005–2007), b) ennustevirheen jakauma ennustetun onnettomuusmäärän mukaan (vuosi 2008).

Verrattaessa toimintavalmiusohjeen A:71 ja mallin tapaa luokitella ruutuja riskiluokkiin nähdään, että malli luokittelee ruutuja pääsääntöisesti yhtä pykälää alempan riskiluokkaan kuin toimintavalmiusohje (taulukko 16). Malli luokittelee esimerkiksi riskiluokkaan I 73 ruutua, kun toimintavalmiusohje luokittelee siihen

3. Yksinkertainen regressiomalli

332 ruutua. Malli luokittelee ruutuja myös kahta ja jopa kolmea pykälää alempaan riskiluokkaan. Tämä ei johdu pelkästään käytetyistä riskiluokkarajoista vaan kuvastaa myös mallin heikohkoa erottelu- ja ennustuskykyä.

Vuonna 2008 143 ruudussa (0,037 % kokonaismäärästä) tapahtui vähintään 17 kiireellistä tehtävää (taulukko 17). Näitä ruutuja voi toimintavalmiusohjeen mukaan pitää riskiluokkaan I kuuluviksi. Toimintavalmiusohjeen mukaan riskiluokkaan I kuuluvia ruutuja piti olla 0,086 % (n = 332). Malli puolestaan ennusti asukasluokkaan perustuvia riskiluokkarajoja käytettäessä, että vastaavia ruutuja olisi vähemmän, vain 0,019 % (n = 73). Vaikka tästä tarkastelusta ei yhden vuoden datan ja luokitteluun liittyvien ongelmien vuoksi voi vetää pitkälle meneviä johtopäätöksiä, toimintavalmiusohje luokittelee ruutuja riskiluokkaan I yli kaksinkertaisen määrän kuin vuoden 2008 toteuma antaa odottaa. Malli puolestaan luokittelee niitä yli kaksi kertaa vähemmän.

Taulukko 16. Ruutujen luokittelu riskiluokkiin toimintavalmiusohjeen A:71 ja mallin perusteella.

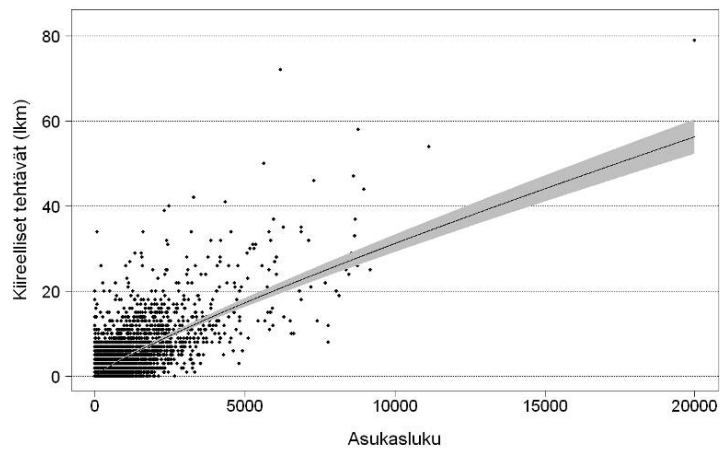
	Malli				Yht.
	A:71	I	II	III	
I	73	224	26	9	332
II	0	723	1 082	179	1 984
III	0	0	3 625	3 296	6 921
IV	0	0	5	373 675	373 680
Yht.	73	947	4 738	377 159	382 917

Taulukko 17. Riskiluokkaruutujen prosentuaalinen määrä vuoden 2008 toteuman, toimintavalmiusohjeen A:71 ja mallin mukaan (malli A = asukasluokarajoilla).

	Luokka	Toteutunut (%)	A:71 (%)	Malli A (%)
Alueet yhdessä	I	0,037	0,087	0,019
	II	0,29	0,52	0,25
	III	3,5	1,8	1,2
	IV	96,2	97,6	98,5

3. Yksinkertainen regressiomalli

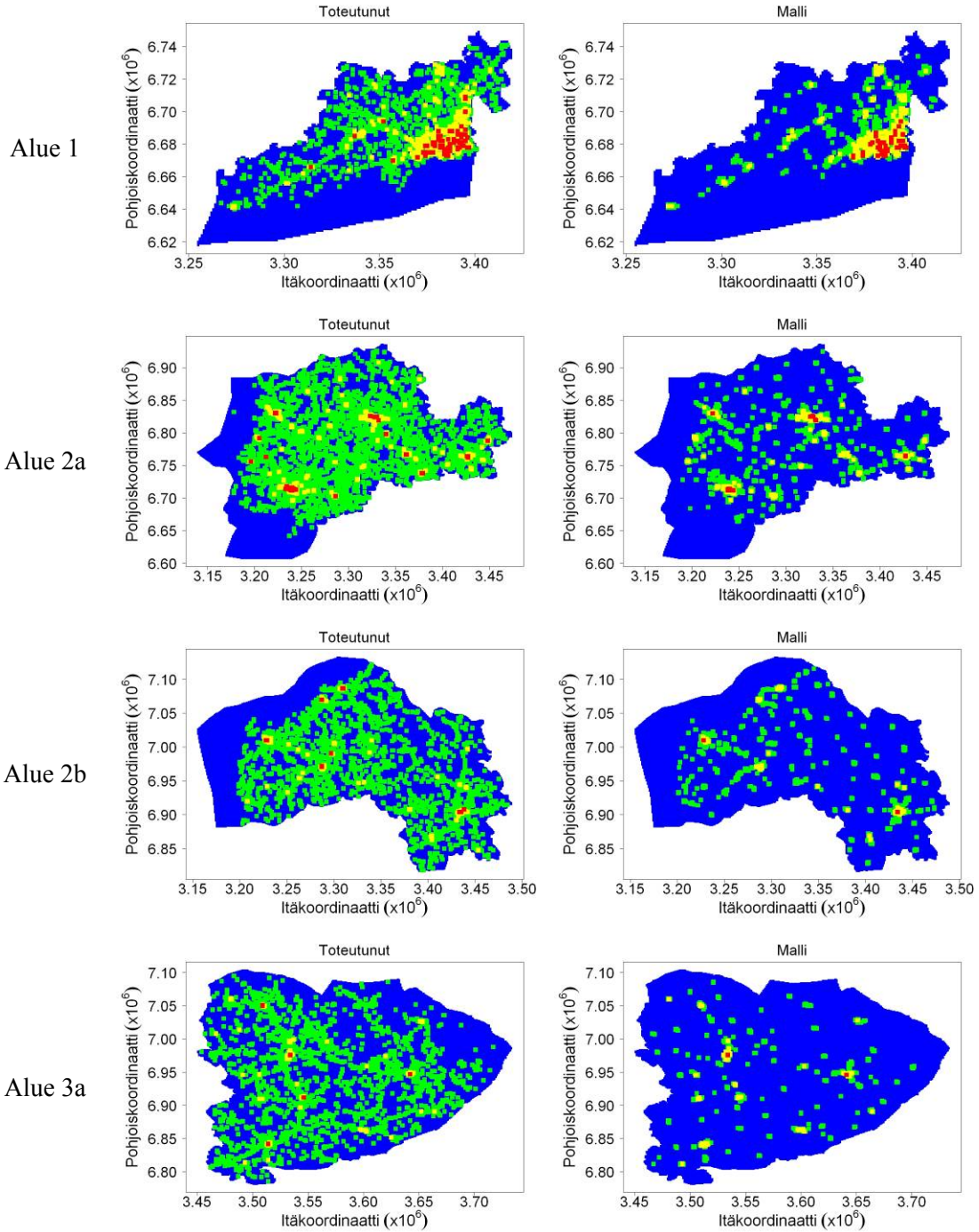
Kuva 12 paljastaa osaltaan, miten heikosti pelkkä asukasluku selittää kiireellisten tehtävien määrää. Havaintopisteiden hajonta ennustekäyrän ympärillä on suuri. Esimerkiksi ruuduissa, joissa on asukkaita noin 6 000, kiireellisten onnettomuuksien määrä on vaihdellut 10 ja 70 onnettomuuden välillä. Malli kuitenkin ennustaa, että 6 000 asukkaan ruudussa onnettomuuksien määrä on 20,2 (95 %:n ennusteväli on 19–21,5). Toisaalta ruuduissa, joissa on ollut 20 kiireellistä onnettomuutta, asukasluku on vaihdellut 0–8 000 asukkaaseen.



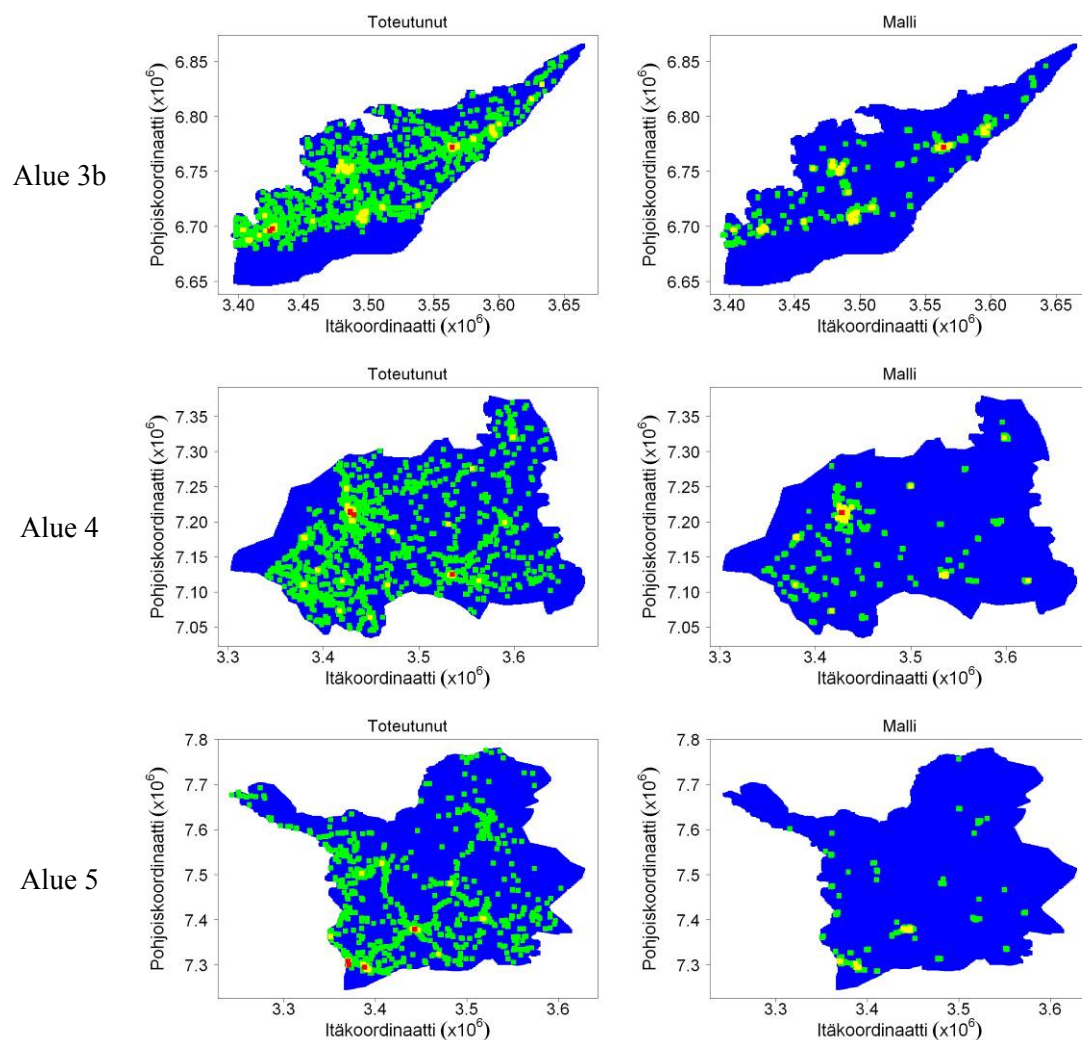
Kuva 12. Kiireelliset tehtävät vuonna 2008 ja asukasluku. Kuvassa myös mallin ennustekäyrä ja 95 %:n ennusteväli.

Kuvassa 13 ruudut on jaettu toteutuneen tai ennustetun onnettomuusmäärän mukaisesti neljään luokkaan. Malli ennustaa ruutuja luokkiin I–III selvästi toteutunutta tilannetta vähemmän. Tämä johtunee pääasiassa vajavaisesta mallista mutta osin myös mallin luokkarajojen määrittelyn ongelmallisuudesta, kuten edellä on jo käynyt ilmi.

3. Yksinkertainen regressiomalli



3. Yksinkertainen regressiomalli



Kuva 13 Riskiruutujen (1 km \times 1 km) sijoittuminen kartalle, kun ruudut on jaettu toteutuneen tai ennustetun onnettomuusmäärän mukaisesti neljään luokkaan. Käytetyt onnettomuusmäärien luokkarajat on esitetty taulukossa 1 ja rajojen valintaperusteet luvussa 1.5. Eri luokkiin sijoittuvia ruutuja on merkitty eri väreillä seuraavasti: luokka I (punainen), luokka II (keltainen), luokka III (vihreä), luokka IV (sininen). Vasemmalla toteutumat ja oikealla ennusteet.

3. Yksinkertainen regressiomalli

3.5 Johtopäätökset

3.5.1 Yleistä malleista

Yksinkertaisella regressiomallilla voidaan yhdistää ruuduissa havaitut onnettomuusmäärät selittävien muuttujien arvoihin ja laatia ruutukohtaisia ennusteita. Mallin regressiokertoimien ja niiden keskihajontojen avulla voidaan helposti laskea ennusteet ruutujen onnettomuusmäärille. Mutta jotta mallilla voidaan ennustaa, kaikille selittäjille on annettava todellisuutta vastaava arvo tai ainakin mahdollisimman hyvä arvio siitä. Rakennuspalomallissa on siis annettava arvo sekä asukasluvulle että kerrosalalle.

Yksinkertaisen regressiomallin eräs rajoite on, että malli olettaa vuoden 2008 onnettomuusmäärän vastaavan vuosien 2005–2007 keskimääräistä tasoa. Tämä ei kuitenkaan pidä paikkaansa, sillä vuonna 2008 oli selvästi enemmän tehtäviä kuin kolmena aiempana vuonna. Tällaisia ajassa tapahtuvia trendejä mallilla ei nyky muodossaan tavoiteta.

Ennustevirheiden tarkastelu paljasti, että asukasluvun ja kerrosalan yhteys onnettomuusmääriin ei ole niin yksiselitteinen. Joissakin ruuduissa onnettomuusmäärä on suuri, vaikka ruudussa on vain vähän asukkaita ja kerrosalaa. Tämä pätee sekä rakennuspaloihin että kiireellisiin tehtäviin. Näissä ruuduissa suuret onnettomuusmäärät johtuvat jostakin muusta. Näitä ruutuja olisikin syytä tarkastella tarkemmin, ja jos syy löytyy, se pitäisi saada lisätyksi malliin.

Mallien todellisesta ennustuskyvystä on vaikea sanoa mitään varmaa. Mallin ennustuskykyä on ensinnäkin tarkasteltu vain vuoden 2008 aineistolla. Kun otetaan huomioon tarkasteltavien onnettomuuksien harvinaisuus (erityisesti rakennuspallo), jo pelkkä satunnaisvaihtelu vaikeuttaa johtopäätösten tekoa. Ruudut, joissa tapahtuu paljon onnettomuuksia, erottuvat hyvin, mutta esimerkiksi riskiluokan II ja III ruutujen erottelu on jo hankalaa. Mallin ennustuskykyä pitäisikin tarkastella useamman vuoden aineistolla.

Ennustuskyvyn arviointia vaikeuttaa myös ajantasaisen selittäjä tiedon puuttuminen. Nyt mallin rakentamiseen ja itse ennustamiseen on jouduttu käyttämään vuoden 2005 asukasluku- ja kerrosalatieitoja. Täten ei ole tavoitettu sitä, miten selittäjissä ajan myötä tapahtuvat muutokset vaikuttavat onnettomuusmääriin.

Mallin vertailu toimintavalmiusohjeeseen osoittautui hankalaksi, koska toimintavalmiusohjeen riskiluokkarajamäärittely ei ole aivan yksiselitteinen. Mallin ennusteita luokiteltiin lähinnä toimintavalmiusohjeen asukaslukuun perustuvia

3. Yksinkertainen regressiomalli

riskiluokkarajoja käyttäen, mutta ne osoittautuivat liian korkeiksi. Sen sijaan kerrosalatietoon perustuvat luokkarajat näyttäisivät toimivan rakennuspalojen tarkastelussa kohtalaisen hyvin. Lisäksi vertailua vaikeuttaa vain yhdeltä vuodelta oleva ennustedata.

3.5.2 Rakennuspalot

Alueiden yhteinen malli 1 x 1 km:n tasolla selitti noin 50 % rakennuspalomäärän vaihtelusta, mitä voi pitää kohtalaisen hyvänä tuloksena, kun otetaan huomioon mallin yksinkertaisuus. Malli ennusti vuosien 2005–2007 rakennuspalojen kokonaisuusmäärää hyvin ja sopi aineistoon muutenkin melko hyvin. Mallin keskimääräinen ennustevirhe oli lähellä nollaa eikä riippunut ennustettavasta onnettomuusmäärästä.

Malli ennusti vuoden 2008 alhaisia rakennuspalomääriä melko hyvin mutta keskisuuria määriä hieman yläkanttiin ja korkeita määriä selvästi alakanttiin. Mallin kyvyttömyys huomioida rakennuspalojen kasvava trendi näkyy nimenomaan ennustettaessa suurempia onnettomuusmääriä. Myös ennustevirheen hajonta tällöin kasvaa, mutta se selittyy pitkälti liian vähäisellä datalla ylemmissä ennusteluokissa. Alueiden yhteinen malli näyttäisi toimivan kohtalaisen hyvin kaikilla alueilla, joten sitä voi soveltaa ennustamiseen, kunhan muistaa sen puutteet.

Rakennuspalojen mallittaminen 250 x 250 m:n tasolla jäi vaillinaiseksi, sillä siirtyminen pienempään ruutukokoon aiheuttaa joitakin lisäongelmia. Ensinnäkin datan harveneminen sekä palomäärän ja selittäjien vaihtelun väheneminen korostavat satunnaisvaihtelun merkitystä. Toiseksi 250 x 250 m:llä mallin rakentamisvaiheessa putoaa peräti 8,5 % rakennuspaloista pois, koska rakennuspaloruudussa ei aineiston mukaan ole rakennuksia lainkaan. Aggregoitaessa aineisto 1 x 1 km -tasolle rakennuspaloja jää mallittamisen ulkopuolelle 3,7 %. Mallin estimoimisessa on myös teknisiä ongelmia, jotka tosin tällä kertaa onnistuttiin kiertämään. Vaikka tulokset 250 x 250 m:n aineistolla ovat samansuuntaisia kuin 1 x 1 km:n aineistolla, mallin sopivuutta ja toimivuutta pitäisi tarkastella lähemmin.

3.5.3 Kiireelliset tehtävät

Alueiden yhteinen malli selitti noin 60 % kiireellisten tehtävien vaihtelusta, mitä voi pitää kohtalaisen hyvänä tuloksena, kun otetaan huomioon mallin yksinkertaisuus. Malli ennusti kuitenkin vuosien 2005–2007 kiireellisten tehtävien koko-

3. Yksinkertainen regressiomalli

naismäärän lähes 8 % toteutumaa suuremmaksi. Mallin keskimääräinen ennustevirhe oli pieniä määriä ennustettaessa lähellä nollaa, mutta siinä oli selvä nouseva trendi. On siis todettava, ettei malli sopinut tähän aineistoon hyvin.

Malli ennusti vuoden 2008 pieniä kiireellisten tehtävien määriä melko hyvin ja keskisuuria ja suuria määriäkin vain hieman yläkanttiin. Tämä ei kuitenkaan ollut mallin ansiota, vaan kiireellisten tehtävien kasvutrendi kompensoi mallin kyvyttömyyttä ennustaa suuria määriä. Ennustevirheen hajonta kasvaa ennustettaessa suurempia määriä, mutta se selittyy lähinnä datan vähäisyydellä ylemissä ennusteluokissa. Myös aluekohtaiset erot mallin sopivuudessa ja ennustekyvyyssä olivat suuria: erityisesti alueilla 4 ja 5 malli ei toiminut. Mallia olisikin selvästi parannettava ennen sen laajempaa soveltamista.

3.6 Kehittämisehdotukset

Datan tarkkuutta ja oikeellisuutta olisi hyvä tarkastella kattavammin. Mikä esimerkiksi selittää rakennuspalot sellaisessa ruudussa, jossa ei ole rakennuksia lainkaan? Rakennuspalo on saatettu paikantaa puutteellisesti tai rakennustiedot ovat vanhentuneet tai muuten puutteelliset. Jos syy on puutteellisissa rakennustiedoissa eikä saatavilla ole parempaa dataa, puuttuvien rakennustietojen paikkaamisen menetelmää olisi hyvä miettiä. Ajantasainen data ei vain helpottaisi mallin rakentamista vaan myös tarkentaisi ennusteita.

Mallin sopivuutta ja toimivuutta on vielä pohdittava useasta eri näkökulmasta. Valitun mallin sensitiivisyyttä käytetyn aineiston suhteen voisi tutkia muun muassa rakentamalla malli uudestaan ilman äärihavaintoja. Äärihavaintoja ovat tässä tapauksessa ruudut, joissa on poikkeuksellisen suuri onnettomuustiheys tai suuri (tai pieni) asukasluku tai kerrosala. Tällaiset jakauman jommassakummassa ääripäässä olevat havainnot saattavat joskus vaikuttaa malliin paljonkin. Samoin voisi pohtia, millainen merkitys vuoden 2006 maastopalojen poikkeuksellisen suurella määrällä on kiireellisten tehtävien kokonaismäärän ylimallittamisessa.

Mallityyppejä ja selittäjien muunnoksia olisi hyvä tarkastella lisää. Negatiivinen binomi -mallia voisi vertailla tarkemmin malleihin, jotka mallittavat suuren nollien määrän selitettävässä muuttujassa hieman eri tavalla (ns. zero-inflated ja hurdle-mallit). Yhteys selitettävään muuttujaan on määritelty log-lineaariseksi, mutta muitakin vaihtoehtoja on. Mallin jäännöstarkasteluihin voisi panostaa enemmän, jolloin voisi löytyä esimerkiksi selitys sille, miksi joissain ruuduissa on poikkeuksellisen suuri ennustevirhe. Lisäselittäjillä voisi yrittää parantaa mallia (mm. työpaikkojen ja tieliikenteen määrä). Aluetarkastelujen pitäisi olla

3. Yksinkertainen regressiomalli

kattavammat, jotta saataisiin parempi varmuus yhteisen mallin toimivuudesta eri alueilla. Ennen kaikkea vaativan, suuren ja melko kompleksisen aineiston mallittamiseen pitäisi varata huomattavasti enemmän aikaa.

Kiireelliset tehtävät koostuvat hyvin erityyppisistä onnettomuuksista, joiden yhtäaikainen mallittaminen melko puutteellisilla selittäjillä on hankalaa. Esimerkiksi kiireellisistä tehtävistä noin kolmannes tapahtuu ruuduissa, joissa ei ole asukkaita tai rakennuksia lainkaan. Lisäksi liikenneonnettomuuksista peräti 40 prosenttia tapahtuu näissä ruuduissa. On myös vaikea kuvitella, että yksinään asukasluku selittäisi maastopaloja.

Kiireellisten tehtävien osalta koko mallinnuksen lähestymistapaa olisikin hyvä miettiä. Eräs vaihtoehto voisi olla, että ensin mallitettaisiin kukin yksittäinen onnettomuustyyppi relevanteilla selittäjillä ja lopuksi yhdistettäisiin onnettomuustyyppikohtaisten mallien tulokset. Rakennuspalojen osalta olisi hyvä miettiä, miten rakennuspalovaarat huomioidaan.

Kaiken kaikkiaan tulisi vielä pohtia, kuinka hyvään ennusteeseen mallilla on päästävää, jotta se täyttäisi tarkoituksensa. Ennusteen 95 %:n ennusteväli on suuren aineiston vuoksi melko kapea, joten siinä mielessä ennuste on tarkka jo nykyisellään. Mutta kun mallien tulosten taakse kurkistaa tarkemmin, selittämätöntä vaihtelua on vielä paljon, eikä malli todellakaan ennusta kaikkien ruutujen onnettomuusmääriä oikein. Miettimisen arvoista olisi lopulta myös se, miten mallia kohdennettaisiin siten, että se hyödyttäisi pelastustoimen suunnittelua parhaiten.

4. Spatiaalinen regressiomalli

4.1 Mallinnuksen lähtökohta

Spatiaalisella regressiomallilla tarkoitetaan tässä yhteydessä mallia, jossa ennustettavan ruudun sijainti vaikuttaa ennusteeseen suoraan eikä vain ruudun selittäjien kautta. Yksinkertaiset regressiomallit lähtevät oletuksesta, että mallin residuaalit (havaittu–ennustettu) ovat riippumattomia toisistaan, eli esimerkiksi viereisten ruutujen residuaalit eivät muistuta toisiaan sen enempää kuin toisistaan kaukana sijaitsevien ruutujen residuaalitkaan. Tämä on spatiaalisissa regressiomalleissa varmistettu sisällyttämällä malliin alueellinen ja paikallinen termi tavanomaisten selittäjien vaikutusta kuvaavien regressiotermien lisäksi.

Kiireelliset tehtävät ovat verrattain harvinaisia suurimmassa osassa maata tarkasteltaessa niiden esiintymistä 1 km:n ja erityisesti 250 m:n ruudukossa. Tyyppillinen esiintymisintensiteetti on alle yksi kiireellistä tehtävää vuodessa 250 m:n ruudukossa. Näin alhaisilla tehtäväintensiteeteillä on sattumaa, havaitaanko vuosittain ruudussa yksi tai ei yhtään tehtävää. Jos kuitenkin oletetaan, että tehtäväintensiteetti on vierekkäisissä ruuduissa samankaltainen, voidaan tietoa hyödyntää estimoinnissa ”lainaamalla voimaa naapuriruuduista”. Ainoaksi ongelmaksi jää arvioida, miten paljon estimoinnissa tulisi painottaa naapureita ja miten paljon ruudun omaa havaintoa. Käytetyt spatiaaliset regressiomallit antavat tähän yhden johdonmukaisen tavan. Spatiaalisuus voidaan tuoda malliin monilla eri tavoilla. Voidaan joko olettaa mallin regressioparametrien olevan spatiaalisesti autokorreloituneita tai sitten kuvata riippuvuutta spatiaalisella termillä.

Käytettävissä ei ollut vuosittaista tietoa selittäjien arvoista. Lähtökohtaisesti vuosittaista vaihtelua ei siis voitu ennustaa muuten kuin trendinä ajassa eli joko valtakunnallisesti yhtenäisenä kasvuna tai sitten alueellisesti vaihtelevana.

4. Spatiaalinen regressiomalli

Hankkeessa kehitettiin malleja riskianalyysin tueksi. Tähän tarkoitukseen tarvitaan malleja, jotka pystyvät antamaan mahdollisimman luotettavia ennusteita tehtävämääristä ja ”palvelutarpeen” sijainnista. Yksittäisellä ruudulla tai yksittäisten ruutujen luokittelulla ei niinkään ole merkitystä; olennaista on mahdollisuus laskea yhteen ruutuja ja saada palvelutarpeesta alueellisia arvioita. Tärkeää on myös pystyä arvioimaan ennusteen tarkkuutta. Käytetty spatiaalisten mallien estimointitapa on joustava kaikenlaisten mallista johdettujen suureiden laskemiseen osana mallin parametrien estimointia.

4.2 Käytetty aineisto

Mallien sovitukseen käytettiin luvussa 1.3 kuvattua aineistoa sillä erotuksella, että meriruudut leikattiin pois karkealla rantaviivalla (Maanmittaushallituksen kuntarajakartta 2006). Meriruutujen yhteydessä rajautui pois tehtäviä, jotka sijaitsivat saarissa.

Kerrosala- ja asukasmäärätietoja käytettiin siinä muodossa kuin ne Tilastokeskuksen ruutuaineistossa olivat. Onnettomuuksista tarkasteltiin kaikkia kiireellisiä tehtäviä, rakennuspaloja ja liikenneonnettomuuksia. Tilastokeskuksen ruututieto yhdistettiin 1 km×1 km -ruutuihin Helsingin, Keski-Uudenmaan ja Länsi-Uudenmaan pelastusalueiden osalta ja niistä muodostettiin osa-alue 1. Kokomaan mallitus toteutettiin 5 km ×5 km -ruudukossa. Näiden kahden aineiston ja vastaavien mallien testiaineistojen sisältämät ruudut ja tehtävät on esitetty taulukossa 18.

Taulukko 18. Spatiaalisten regressiomallien aineisto. Vuosien 2005–2007 osalta taulukossa on esitetty kolmen vuoden yhteenlaskettu määrä.

Malli	Vuodet	Ruutuja	Kiireellisiä tehtäviä	Rakennuspaloja	Liikenneonnettomuuksia
5 km:n ruudut	2005–2007	13 898	84 780	11 567	33 752
Testi	2008	13 898	30 293	4 446	13 330
1 km:n ruudut	2005–2007	7 266	16 837	1 910	6 057
Testi	2008	7 266	6 805	932	2 766

4.3 Mallin kuvaus

Käytetyt mallit perustuvat bayesilaiseen tilastotieteeseen. Siinä kaikkia parametreja pidetään satunnaismuuttujina, joita kuvataan todennäköisyysjakaumilla.

4. Spatiaalinen regressiomalli

Käytettyä mallia kutsutaan Besag-York-Mollie-malliksi sitä ensimmäisenä soveltaneiden tutkijoiden mukaan (Besag et al. 1991). Mallia on käytetty laajasti muun muassa epidemiologiassa harvinaisten sairauksien alueellisen vaihtelun mallintamiseen (Best et al. 2005). Mallin erityispiirre on autokorrelaatorakenne, jolla yksittäiset ruutuestimaatit sidotaan toisiinsa rajanaapureidensa kautta.

Käytetyissä malleissa oletettiin, että vuosittain ruudussa havaittu tehtävämäärä, n_i , noudattaa poissonjakaumaa (kaava 4). Yksinkertaisista regressiomalleista poiketen ruuduissa havaittujen tehtävämäärän ei oletettu noudattavan negatiivista binomijakaumaa, sillä spatiaalinen malli sisälsi ruututermiä γ_i , jonka katsottiin varmistavan poissonoletuksen täyttyminen.

$$n_i \sim \text{poisson}(y_i), \quad (4)$$

Mallin estimoinnin kannalta oli hyödyllistä olettaa, että selittäjien vaikutukset olisivat log-asteikolla lineaarisia. Tällöin myös selittäjille tehtiin log-muunnos. Malli oli siis rakenteeltaan lähellä luvussa 3 esiteltyjä yksinkertaisia regressiomalleja. Ruutukohtaisen intensiteetin luonnollisen logaritmin $\log(y_i)$:n oletettiin riippuvan selittäjistä: kerrosalasta, asukasmäärästä ja vuodesta kaavan 5 mukaisesti:

$$\begin{aligned} \log(y_i) = & \alpha + \beta_1 \log(x_2 + 1) + \beta_2 \log(x_1 + 1) \\ & + \beta_3 \log(x_1 + 1) \log(x_2 + 1) + \beta_4 x_3 + \lambda_i + \gamma_i \end{aligned} \quad (5)$$

jossa α on tasoparametri, $\beta_1 \dots \beta_4$ ovat regressiokertoimia, x_1 on asukasluku, x_2 kerrosala neliömetreinä, x_3 vuosi ja λ_i ruudun i alueellinen termi, joka on sidottu autokorrelaatorakenteella ruutuihin, joiden kanssa sillä on yhteinen raja tai kulma. γ_i on ruudun i ruututermi.

Tarkasteltaessa intensiteetin riippuvuutta lineaarisella asteikolla (kaava 6) huomataan, että vaikutukset ovat multiplikatiivisia – esimerkiksi yhden yksikön muutos λ :ssa muuttaa intensiteettiä e^λ -kertaiseksi.

$$y_i = e^\alpha e^{\beta_1 \log(x_1+1)} e^{\beta_2 \log(x_2+1)} e^{\beta_3 \log(x_1+1) \log(x_2+1)} e^{\beta_4 x_3} e^{\lambda_i} e^{\gamma_i}, \quad (6)$$

Selittäjien merkityksen arvioimiseksi sovitettiin myös malli, joka ei huomioinut ruudun asukas- tai kerrosalatietoja (kaava 7).

4. Spatiaalinen regressiomalli

$$\log(\mu_i) = \alpha + \beta_4 x_3 + \lambda_i + \gamma_i \quad (7)$$

Mallien parametreille annettiin epäinformatiiviset priorit (kaava 8). Priorilla tarkoitetaan ennakkokäsitystä parametrin arvojen todennäköisyysjakaumasta. Priorit ovat välttämättömiä käytetyssä bayesilaisessa lähestymistavassa. Analyysin tuloksena saadaan ns. posteriorijakauma, joka yhdistää prioritiedon ja aineiston. Käytetyt priorit olivat

$$\begin{aligned} \alpha &\sim \text{tasajakauma}(-\infty, \infty) \\ \beta_1 \dots \beta_4 &\sim \text{norm}(0, 10000) \\ \lambda_i &\sim \text{norm}(\mu_{\text{naapurit}}, \tau_1 / n_{\text{naapurit}}) \\ \gamma_i &\sim \text{norm}(0, \tau_2) \\ \sqrt{\tau_1} &\sim \text{tasajakauma}(0, 1000) \\ \sqrt{\tau_2} &\sim \text{tasajakauma}(0, 1000) \end{aligned} \quad (8)$$

Erityisen merkittävä on λ :lle annettu autokorreloitu prior, joka sitoo alueellisen tekijän naapureihinsa niiden keskiarvon ja määrän kautta, ja γ :lle annettu prior, joka sallii ruudun poiketa naapureistaan. Käytännössä vain nämä erottavat mallin yksinkertaisesta regressiomallista.

Alueellinen termi ja ruututermi sisältävät sekä ”havainnoimattomien” selittäjien vaikutuksia että paikallista virhettä, joka syntyy mallityypin mahdollisesti virheellisesti kuvaaman selitettävien ja kerrosalan ja asukasmäärän välisestä riippuvuudesta.

Mallin parametrit estimoitiin käyttämällä Winbugs 1.4 -ohjelmaa (Imperial College and MRC, Lunn et al. 2000), joka on saatavissa ilmaiseksi osoitteesta <http://www.mrc-bsu.cam.ac.uk/bugs/>. Ajossa käytetty mallikoodi on esitetty liitteessä A. Estimointi perustuu MCMC-menetelmään, jolla poimitaan posteriorijakaumasta otos. Ohjelman alkuarvoina käytettiin regressioparametreille nolliä ja variansseille ykkösiä. Otoksesta hylättiin ensimmäiset 10 000 arvoa, jotta alkuarvojen vaikutus häviäisi, ja tunnusluvut laskettiin seuraavista 10 000 arvosta. Estimaatin tarkkuutta kuvattiin HDI:llä (*High Density Interval*, uskottavuusväli). Se ilmaisee jakauman lyhimmän välin, jolla 95 % todennäköisyydestä sijaitsee.

4. Spatiaalinen regressiomalli

4.4 Tulokset

Tarkasteltaessa valtakunnallisesti kiireellisten tehtävien ja rakennuspalojen esiintymistä $5 \text{ km} \times 5 \text{ km}$ -ruudukossa asukasmäärän vaikutus (β_1) ei ollut merkitsevä, mutta liikenneonnettomuuksien mallissa vaikutus oli positiivinen (taulukot 19, 20 ja 21). Myös kerrosalan vaikutus (β_2) kaikkiin kolmeen tarkasteltuun onnettomuusluokkaan oli positiivinen. Kerrosalan ja asukasmäärän yhteisvaikutus oli kaikissa merkitsevä ja positiivinen. Kertoimien suhteet ja selittävien muuttujien vaihtelu huomioiden estimaattien vaihtelu syntyi pääosin kerrosalan vaihtelusta.

Alueellisen termin keskihajonta oli kiireellisten tehtävien ja liikenneonnettomuuksien malleissa samansuuruista mutta rakennuspaloissa vain puolet tästä. Rakennuspalojen selittäjät siis pystyivät kuvaamaan vaihtelua niin, ettei jäännöshajonnassa ollut vastaavaa alueellisuutta kuin kiireellisissä tehtävissä tai liikenneonnettomuuksissa. Ruututekijän keskihajonta taas oli suurinta liikenneonnettomuuksissa, samansuuruista kiireellisissä tehtävissä ja rakennuspaloissa. Tämä kuvastaa käytettyjen selittäjien kyvyttömyyttä selittää liikenneonnettomuuksien määrää. Kaikissa malleissa kuitenkin havaittiin merkittävää alueellista vaihtelua (τ_2 taulukoissa 19, 20 ja 21). Yksittäisistäkin ruuduista oli löydettävissä klustereita, joissa intensiteetti y poikkesi yli 95 %:n todennäköisyydellä keskimääräisestä selittäjien antamasta tasosta (kuvat 14d–20d). Vertailtaessa kuvia 14d–17d ja 19d–20d huomataan selvästi, miten spatiaalinen malli pystyy löytämään liikenneonnettomuuksien kasaumat pääliikenneväylien ruuduista, vaikeivät selittäjät niitä edustakaan.

Eri resoluutiolla olevien mallien regressiokertoimien vertailua ei valitettavasti pystytty tekemään mielekkäästi, sillä sovitetut mallit kohdistuivat eri alueisiin. Laskentateknisistä syistä malleja ei pystytty sovittamaan 1 km :n ruudukossa kuin alueella 1. Kerrosalan vaikutus (β_2) oli kaikissa alueen 1 malleissa positiivinen, mutta asukasmäärän vaikutus (β_1) kiireellisiin tehtäviin ja rakennuspaloihin oli negatiivinen. Rakennuspalojen ja kiireellisten tehtävien valtakunnallisissa malleissa asukasmäärän kerroin ei erottunut nolosta tarkasteltaessa parametrin HDI:tä. Kerrosalan ja asukasmäärän välinen yhteisvaikutus oli kaikissa malleissa positiivinen. Se oli kuitenkin niin suuri, ettei aineiston sisällä aiheudu tilanteita, joissa onnettomuusmäärien estimaatit laskisivat asukasmäärän kasvaessa. Regressiokertoimien tulkinta malleissa, joissa muuttujat eivät ole riippumattomia toisistaan, on vaikeaa. Kertoimien suuruudet onkin hyvä ymmärtää lähinnä tekniseksi välivaiheeksi estimaattien laskemisessa. Alueella 1 alueellisen termin ja ruutu-

4. Spatiaalinen regressiomalli

termin keskihajonnan suhteet mallien välillä olivat paljolti samankaltaiset kuin valtakunnallisessa mallissakin. Poikkeuksena kiireellisten tehtävien malli, jossa ruututermi oli merkityksetön (kuva 17c). Hajontaparametrien estimaatit eivät kuitenkaan ole niin tarkkoja, että pieniä eroja voisi pitää merkittävänä.

Taulukko 19. Kiireellisten tehtävien määrän spatiaalisten mallien parametrit.

Malli	Parametri	Keskiarvo	95% HDI
Koko Suomi, 5km:n ruudut	α	-1,423	-1,386...-1,444
	β_1	-0,00032	-0,01926...0,03435
	β_2	0,3722	0,346...0,3855
	β_3	0,04681	0,04437...0,04854
	β_4	0,07293	0,06454...0,08125
	$\sqrt{\tau_1}$	0,2846	0,1938...0,3269
	$\sqrt{\tau_2}$	0,9555	0,8815...1,081
Alue 1, 1 km:n ruudut	α	-3,713	-3,838...-3,582
	β_1	-0,3944	-0,4585...-0,3351
	β_2	0,2733	0,2483...0,299
	β_3	0,04846	0,04353...0,05337
	β_4	0,1409	0,1221...0,1593
	$\sqrt{\tau_1}$	0,000638	0,0006376...0,01745
	$\sqrt{\tau_2}$	1,815	1,815...2,006
Alue 1, 1 km:n ruudut, ilman asukaslukua ja kerrosalaa	α	-2,42362	-2,487...-2,357
	β_4	0,147514	0,1283...0,167
	$\sqrt{\tau_1}$	0,078637	0,01089...0,1556
	$\sqrt{\tau_2}$	2,937103	2,822...3,057

4. Spatiaalinen regressiomalli

Taulukko 20. Rakennuspalojen määrän spatiaalisten mallien parametrit.

Malli	Parametri	Keskiarvo	95 % HDI
Koko Suomi, 5 km:n ruudut	α	-3,521	-3,608...-3,443
	β_1	0,02892	-0,03013...0,07972
	β_2	0,5708	0,5259...0,6301
	β_3	0,01914	0,0147...0,0241
	β_4	0,04931	0,02694...0,07136
	$\sqrt{t_1}$	0,222	0,1806...0,2596
	$\sqrt{t_2}$	0,5532	0,4809...0,6334
Alue 1, 1 km:n ruudut	α	-6,443	-6,854...-6,052
	β_1	-0,3363	-0,4696...-0,2163
	β_2	0,3678	0,3135...0,4272
	β_3	0,04913	0,03852...0,05734
	β_4	0,09131	0,0374...0,1474
	$\sqrt{t_1}$	0,2167	0,08247...0,3312
	$\sqrt{t_2}$	0,6814	0,4682...0,8945

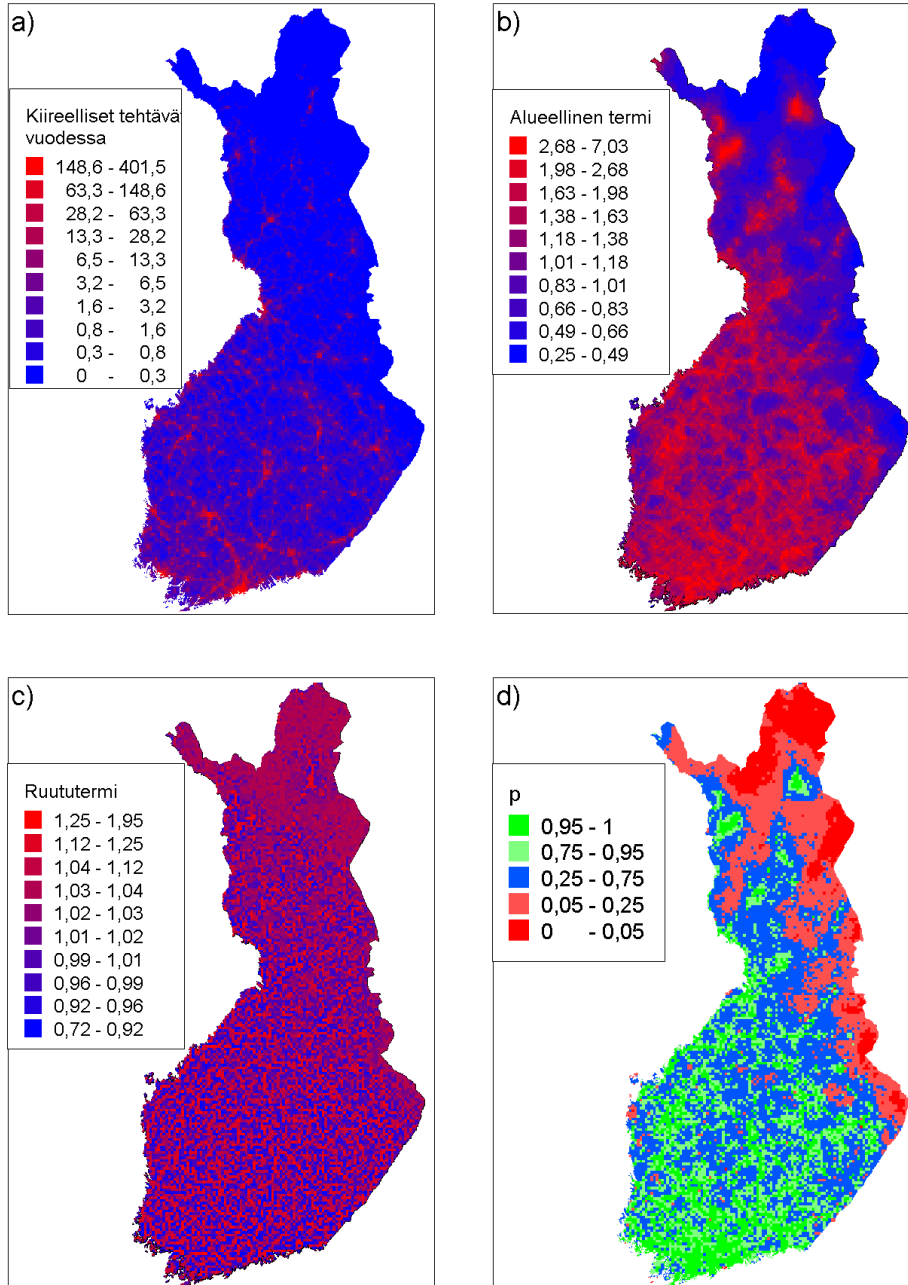
4. Spatiaalinen regressiomalli

Taulukko 21. Liikenneonnettomuuksien määrän spatiaalisen mallin parametrit

Malli	Parametri	Keskiarvo	95 % HDI
Koko Suomi, 5 km:n ruudut	α	-2,577	-2,629...-2,517
	β_1	0,1685	0,1213...0,2248
	β_2	0,3806	0,334...0,4202
	β_3	0,02793	0,02301...0,03078
	β_4	0,1204	0,1076...0,1334
	$\sqrt{\tau_1}$	0,5421	0,4575...0,5984
	$\sqrt{\tau_2}$	1,104	0,9577...1,337
Alue 1, 1 km:n ruudut	α	-5,392	-5,692...-5,144
	β_1	-0,2824	-0,456...-0,1418
	β_2	0,3296	0,2877...0,3791
	β_3	0,03295	0,02011...0,04508
	β_4	0,267	0,2353...0,2978
	$\sqrt{\tau_1}$	0,2709	0,152...0,5901
	$\sqrt{\tau_2}$	2,75	2,425...2,927

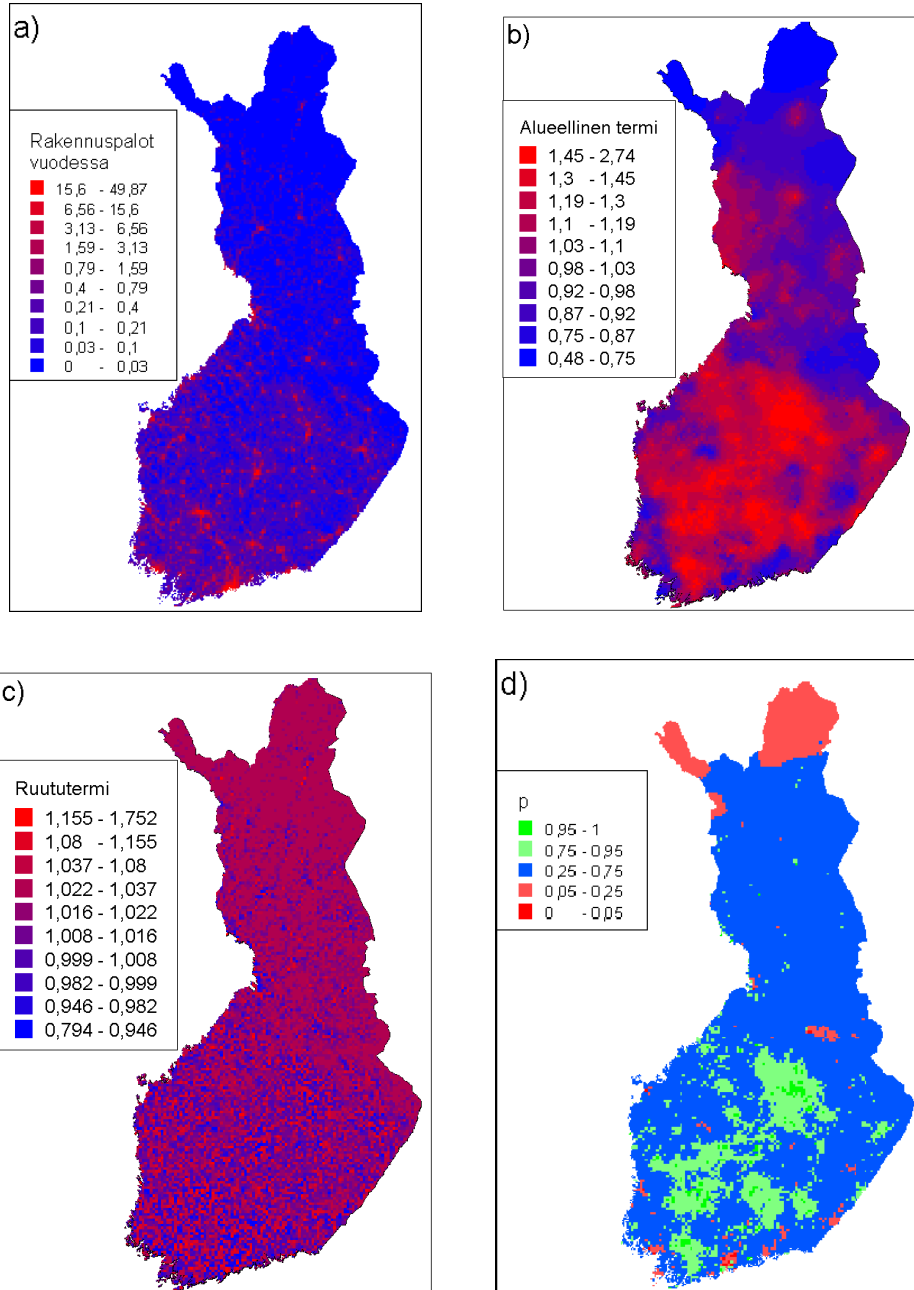
Karttoja tarkasteltaessa on tärkeää huomata, etteivät väriskaalat ole tasavälisiä. Luokat on valittu niin, että kartta tuo hyvin esille kaiken sen vaihtelun, jota ruutujen välillä on. Alueellisten ja ruututermien kartat on esitetty muunnettuna e^λ ja e^γ tulkinnan helpottamiseksi.

4. Spatiaalinen regressiomalli



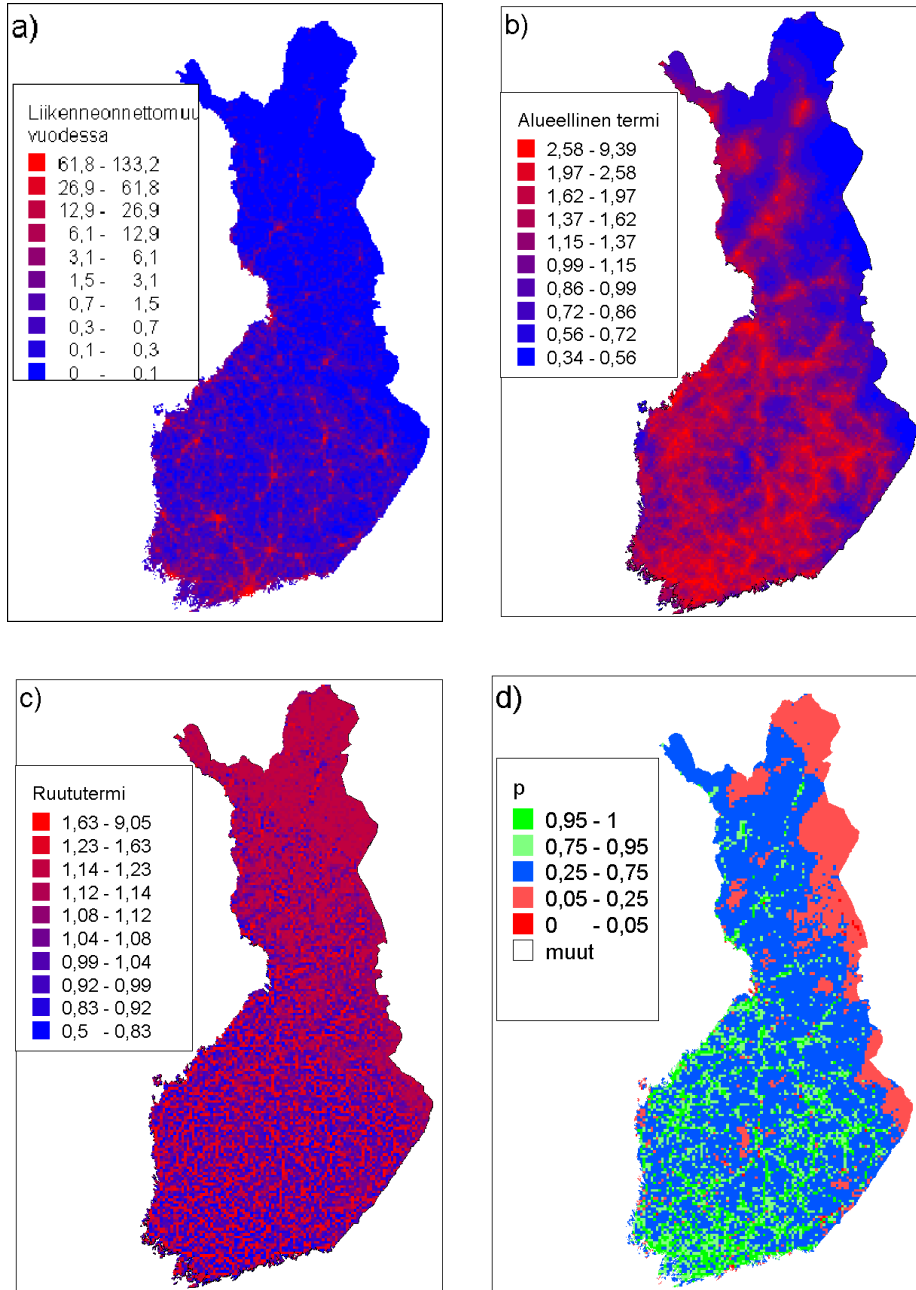
Kuva 14. Kiireellisten tehtävien spatiaalisen mallin a) ennuste vuodelle 2008, b) alueellinen termi, e_l, c) ruututermi, e_y, ja d) todennäköisyys sille, että ruudun ennuste on suurempi kuin pelkän asukasmäärän, kerrosalan ja vuoden perusteella voitaisiin ennustaa.

4. Spatiaalinen regressiomalli



Kuva 15. Rakennuspalojen spatiaalisen mallin a) ennuste vuodelle 2008, b) alueellinen termi, e^{λ} , c) ruututermi, e^{ν} , ja d) todennäköisyys sille, että ruudun ennuste on suurempi kuin pelkän asukasmäärän, kerrosalan ja vuoden perusteella voitaisiin ennustaa.

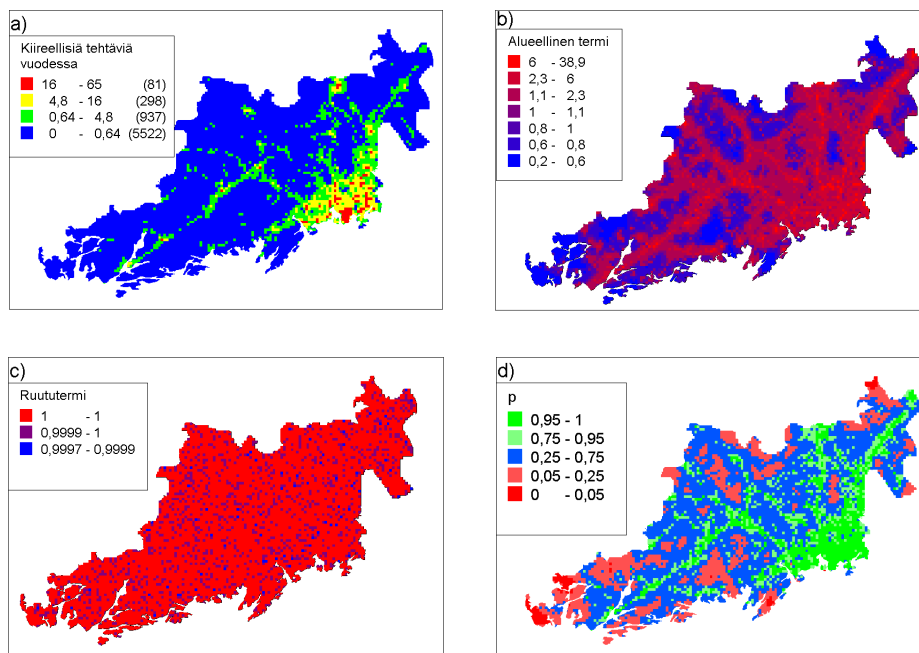
4. Spatiaalinen regressiomalli



Kuva 16. Liikenneonnettomuuksien spatiaalisen mallin a) ennuste vuodelle 2008, b) alueellinen termi, e^{λ} , c) ruututermi, e^{ν} , ja d) todennäköisyys sille, että ruudun ennuste on suurempi kuin pelkän asukasmäärän, kerrosalan ja vuoden perusteella voitaisiin ennustaa.

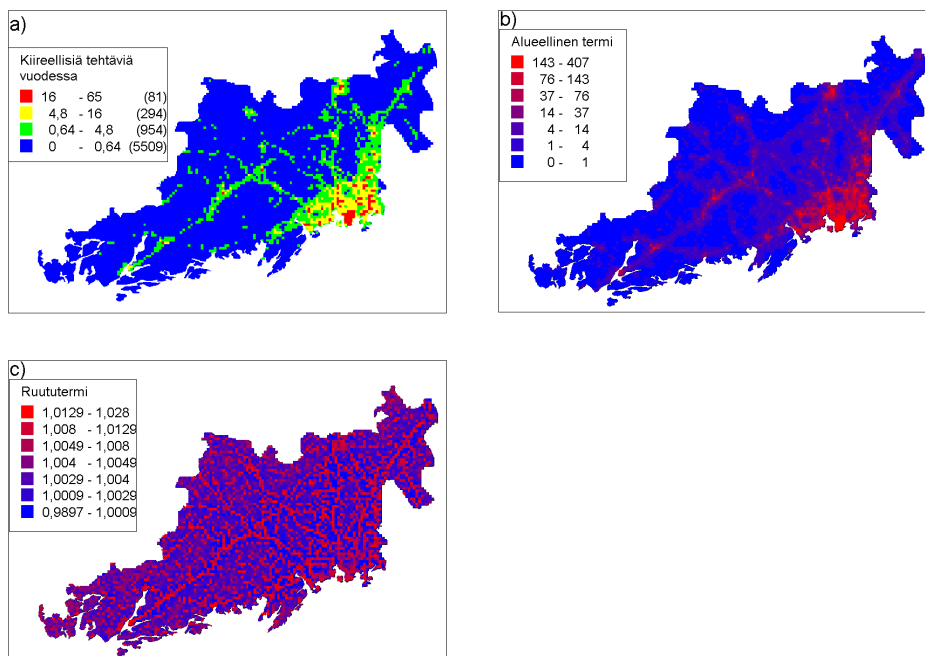
4. Spatiaalinen regressiomalli

Aluetta 1 tarkasteltaessa malli löytää 1 km:n ruudukosta pitkälti samat rakenteet niin kiireellisissä tehtävissä, rakennuspaloissa kuin liikenneonnettomuuksissakin. Tämä on luonnollista, sillä liikenneonnettomuudet ja rakennuspaloit lukeutuvat kiireellisiin onnettomuuksiin. Mielenkiintoista on kuitenkin, miten alueellinen tekijä löytää laajoilta alueilta $\pm 20\%$:n rakennuspalojen määrän vaihtelua suhteessa asukasluvun ja kerrosalan ennustamaan tasoon (kuva 19b)



Kuva 17. Helsingin, Keski-Uudenmaan ja Länsi-Uudenmaan pelastuslaitosten kiireellisten tehtävien spatiaalisen mallin a) ennuste vuodelle 2008, b) alueellinen termi, e^{λ} , c) ruutu-termi, e^{γ} , ja d) todennäköisyys sille, että ruudun ennuste on suurempi kuin pelkän asukasmäärän, kerrosalan ja vuoden perusteella voitaisiin ennustaa.

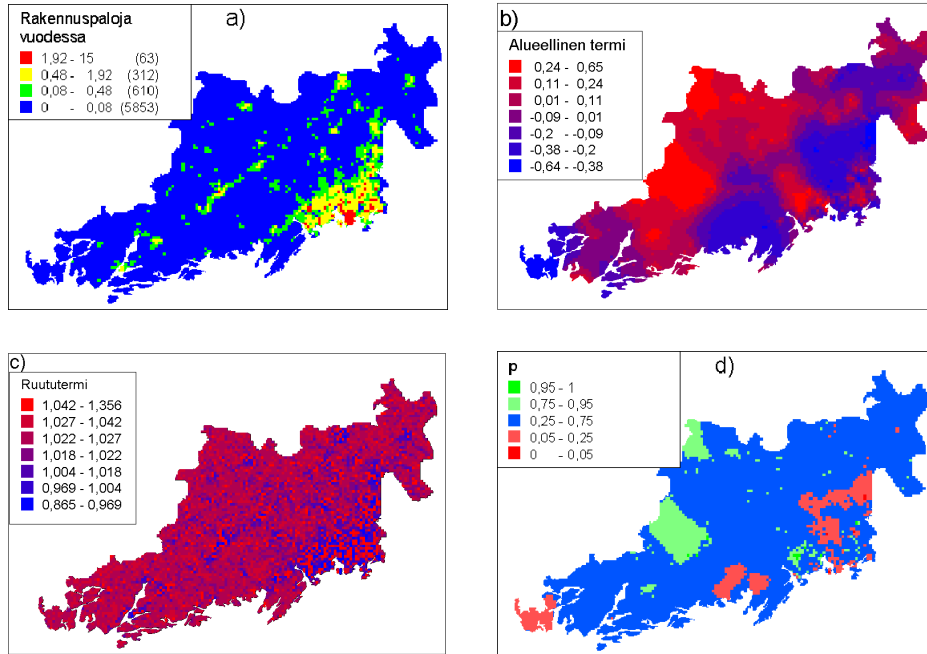
4. Spatiaalinen regressiomalli



Kuva 18. Helsingin, Keski-Uudenmaan ja Länsi-Uudenmaan pelastuslaitosten kiireellisten tehtävien spatiaalisen mallin ilman selittäjiä a) ennuste vuodelle 2008, b) alueellinen termi, e^{λ} , c) ruututermi, e^{ν} .

Alueella 1 kokeiltiin myös mallia, jossa kiireellisiä tehtäviä selitettiin ilman asukasmäärää ja kerrosalaa. Vaikka selittäjien regressiokertoimet olivat selvästi merkitseviä, ei niillä ollut suurtakaan vaikutusta kiireellisten tehtävien estimaatteihin (kuva 18). Alueellinen ja ruututermi pystyivät imemään itseensä niiden vaikutukset. Tämä regressiomalleille tyypillinen piirre onkin tärkeä muistaa kertoimia tulkittaessa. Sama ilmiö nähdään liikenneonnettomuuksien esiintymistä selittävässä mallissa.

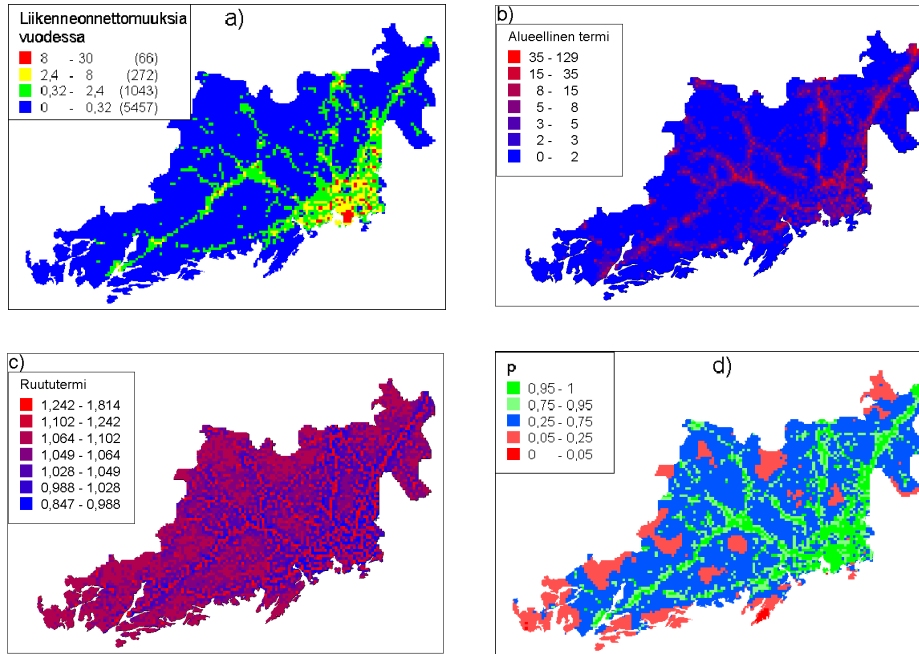
4. Spatiaalinen regressiomalli



Kuva 19. Helsingin, Keski-Uudenmaan ja Länsi-Uudenmaan pelastuslaitosten rakennuspalojen spatiaalisen mallin a) ennuste vuodelle 2008, b) alueellinen termi, e^{λ} , c) ruututermi, d) todennäköisyys, että ruudun ennuste on suurempi kuin pelkän asukasmäärän, kerrosalan ja vuoden perusteella voitaisiin ennustaa.

Liikenneonnettomuuksien alueellisen termin ja ruututermin kartoissa mallin toiminnalle on kuvaavaa, että alueellinen termi on tasaisempaa ja levittäytyy liikenneväylien ympäristöön. Spatiaalisessa mallissahan ei ole mitenkään huomioitu, että ruutujen samankaltaisuus saattaisi olla orientoitunutta. Ruututermi taas saa väylillä positiivisia ja välittömästi niiden ulkopuolella negatiivisia arvoja. Ruutujen välisen vaihtelun jakaminen autokorreloituneeseen ja autokorreloitamattomaan onkin lähinnä laskentatekninen seikka eikä kuvasta suoraan mitään todellista jakoa.

4. Spatiaalinen regressiomalli



Kuva 20. Helsingin, Keski-Uudenmaan ja Länsi-Uudenmaan pelastuslaitosten liikenneonnettomuuksien spatiaalisen mallin a) ennuste vuodelle 2008, b) alueellinen termi, e^{λ} , c) ruututermi, e^{ν} , ja d) todennäköisyys sille, että ruudun ennuste on suurempi kuin pelkän asukasmäärän, kerrosalan ja vuoden perusteella voitaisiin ennustaa.

Alueen 1 ruuduista 2,15 % luokitui rakennuspalojen esiintymisen mukaisesti I riskiluokkaan. Spatiaalisen mallin ennusteen mukaisesti I riskiluokkaan sijoittui 0,87 % ruuduista. Taulukoissa 22 ja 23 on esitetty, miten toimintavalmiusohjeen mukainen riskiluokitus ja selittäjät sisältänyt spatiaalinen malli luokittelivat ruudut riskiluokkiin. Spatiaalinen malli aliarvioi selvästi I riskiluokan ruutujen määrän. Tämä oli täysin oletettavaa, sillä käytetyssä aineistossa kiireellisten tehtävien ja rakennuspalojen välinen suhde oli pienempi kuin toimintavalmiusohjeessa mainittu.

4. Spatiaalinen regressiomalli

Taulukko 22. Ruutujen luokittelu riskiluokkiin toimintavalmiusohjeen A:71 ja spatiaalisen mallin perusteella alueella 1.

	Spatiaalinen malli				
A:71	I	II	III	IV	Yht.
I	70	84	2	0	156
II	11	200	181	2	394
III	0	17	506	221	744
IV	0	2	290	5 680	5 972
Yht.	81	303	979	5 903	7 266

Taulukko 23. Alueen 1 ruutujen jakaantuminen riskiluokkiin toimintavalmiusohjeen kerrosalan ja asukasmäärärajojen sekä spatiaalisen mallin mukaisesti käytettäessä toimintavalmiusohjeen mukaisia kiireellisten tehtävien esiintymisrajoja.

Riskiluokka	A:71	Spatiaalinen malli
1	2,1 %	1,1 %
2	5,4 %	4,2 %
3	10,2 %	13,5 %
4	82,2 %	81,2 %

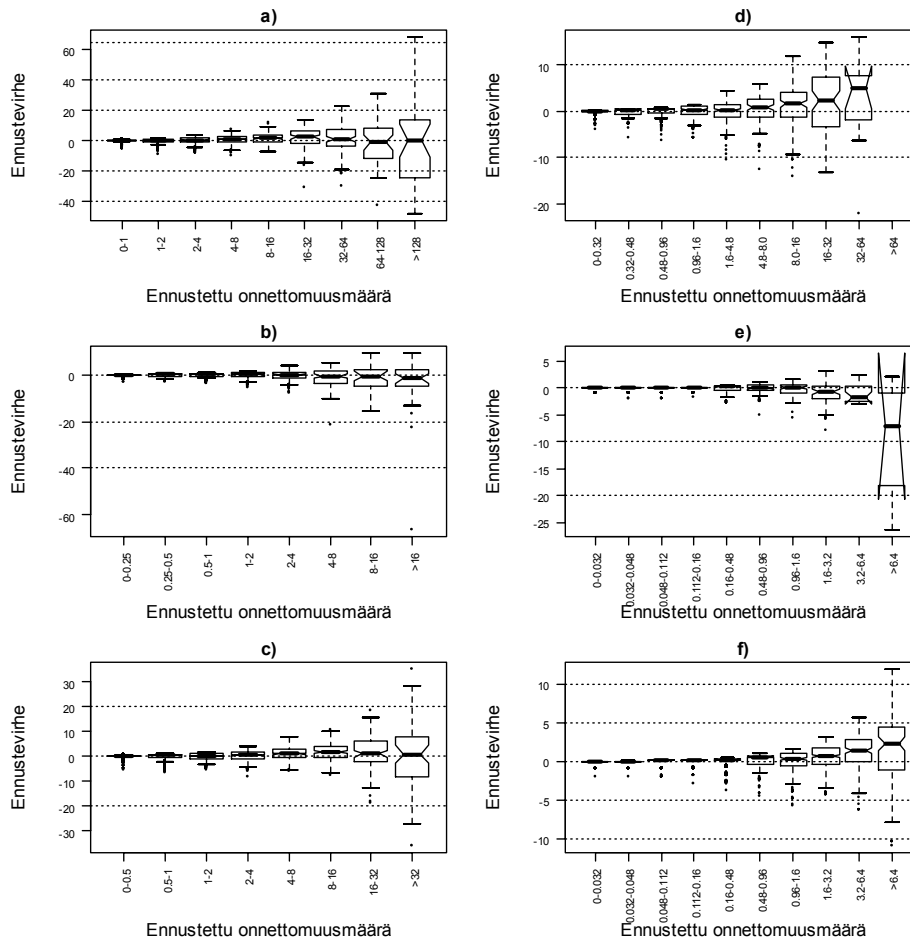
4.5 Mallin toimivuuden arviointi

Mallin käyttökelpoisuutta ennusteiden laatimisessa tarkasteltiin sen kyvyllä ennustaa vuoden 2008 toteutuneiden tehtävien määrää. Tarkkuutta kuvattiin ennusteen ja havaitun eron mediaanilla ennusteluokittain. Vertailuksi laskettiin myös toimintavalmiusohjeen taustalla olevien mallien sekä kolmen viimeisen vuoden keskiarvoon perustuvan ennusteen tarkkuus.

Ennustettaessa vuotta 2008 valtakunnallinen malli 5 km:n ruudukossa yliarvioi kiireellisten tehtävien määrän 7,8 %:lla, rakennuspalojen 4,3 %:lla ja liikennohnettomuuksien 6,9 %:lla. Alueelle 1 ennusteet laskettiin 1 km:n ruuduk-

4. Spatiaalinen regressiomalli

koon. Kiireelliset tehtävät yliarvioitiin 8,6 %:lla, rakennuspalot aliarvioitiin 18,2 %:lla ja liikenneonnettomuudet yliarvioitiin 21,6 %:lla.

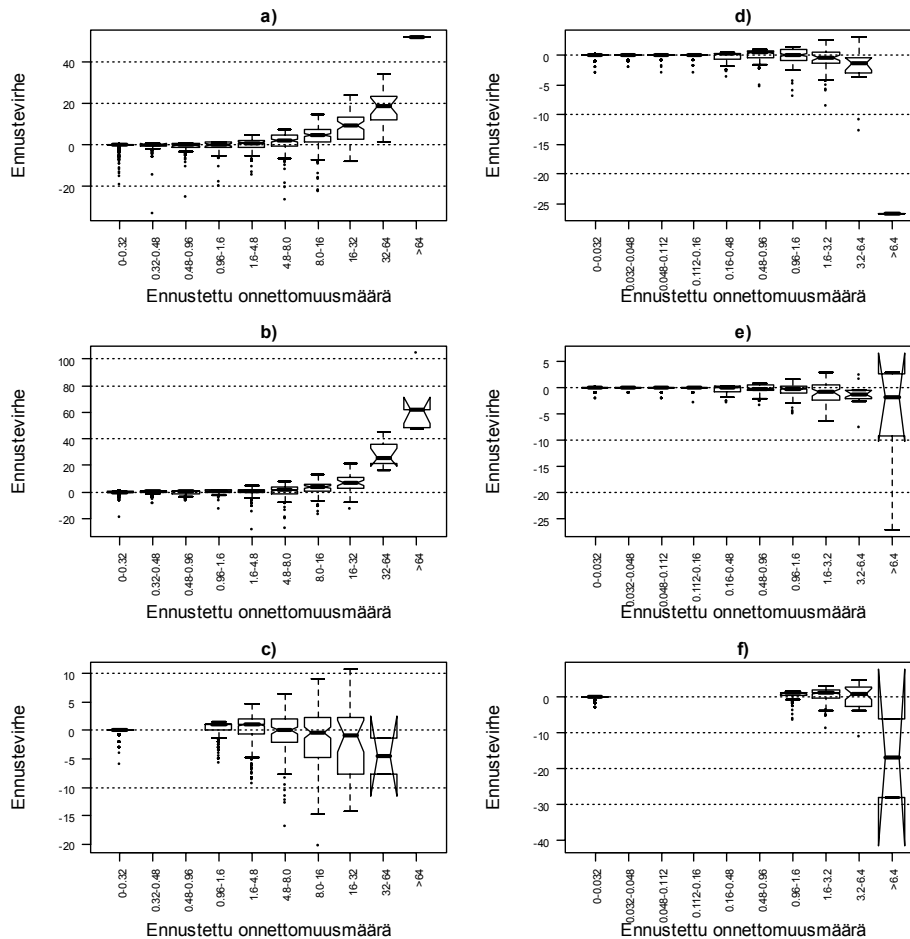


Kuva 21. Ennustevirheet vuodelle 2008 ennusteluokittain 5 km ruudukossa a) kiireelliset tehtävät, b) rakennuspalot ja c) liikenneonnettomuudet. 1 km ruudukossa alueelle 1 d) kiireelliset tehtävät, e) rakennuspalot ja f) liikenneonnettomuudet.

Kaikkien mallien ennustevirheet kasvavat, kun ennuste kasvaa. Tämä on luonnollista, jos pidämme ilmiötä poissonprosessina. Ennustevirhevarianssin tulisi olla yhtä suuri kuin ennusteen. Valtakunnallisissa malleissa (kuvat 21a–c) ennustevirhevarianssit ylittävät keskiarvon selvästi neljässä suurimmassa estimaattiluokassa. Suuret ennusteet on yliarvioitu varsinkin alueella 1. Systemaattiset

4. Spatiaalinen regressiomalli

virheet ennusteissa syntyvät paljolti vuosien 2005–2007 perusteella arvioidusta trendistä, johon vuoden 2008 havainnot eivät soineet.

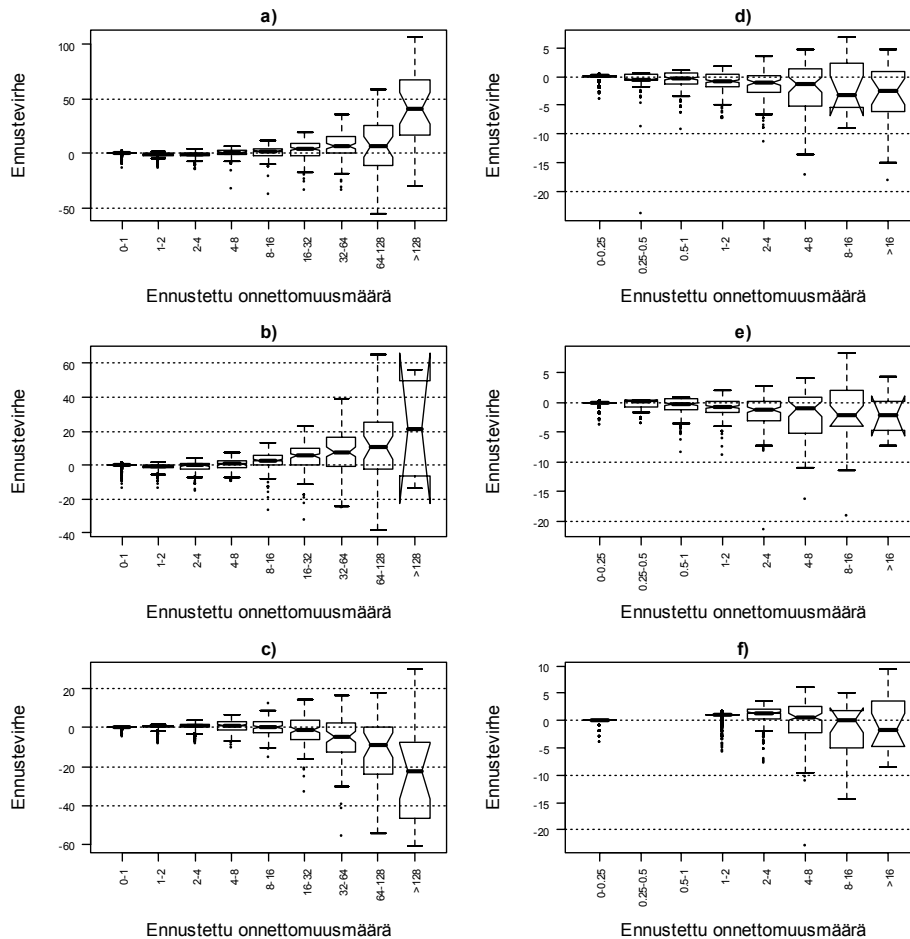


Kuva 22. Kiireellisten tehtävien ennustevirheet vuodelle 2008 ennusteluokittain 1 km × 1 km -ruudukossa käytettäessä ennusteena a) asukasluku*0,0047, b) kerrosala*0,0000655 c) vuosien 2005–2007 keskiarvoa. Rakennuspalojen ennustevirheet käytettäessä ennusteena d) asukasluku*0,00047, e) kerrosala*0,0000655 ja vuosien 2005–2007 keskiarvoa.

Toimintavalmiusohjeen taustalla olevat yksinkertaiset suhdelukuun perustuvat regressiomallit yliarvioivat selvästi kiireellisten tehtävien määriä mutta ennustavat rakennuspaloja samalla tavalla kuin spatiaaliset mallit (kuvat 21–23). Kolmen edellisen vuoden keskiarvo aliarvioi kiireellisten tehtävien ja rakennuspalojen määrän, sillä tehtävien määrät ovat kasvaneet ajanjakson aikana. Valtakun-

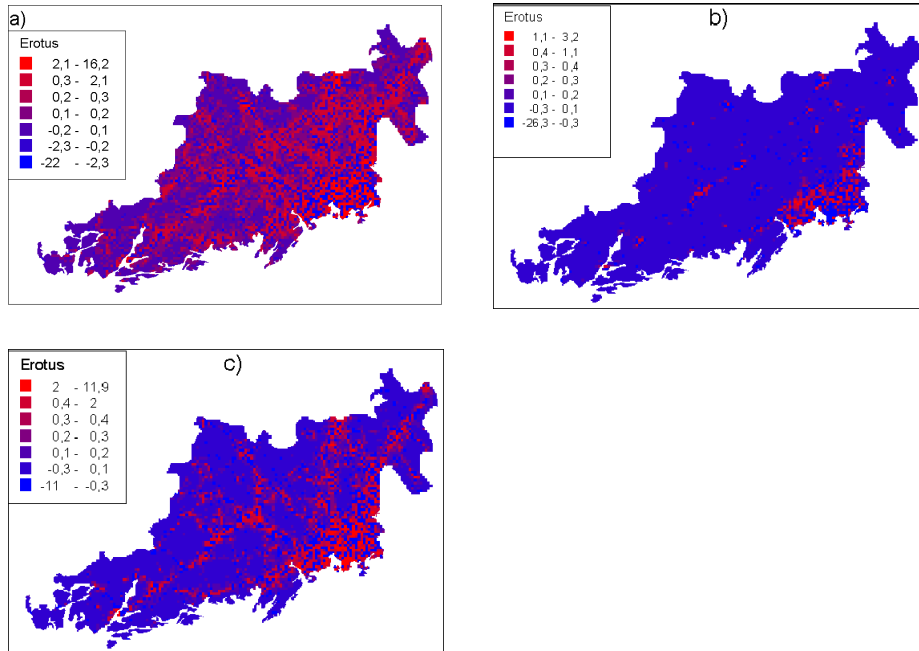
4. Spatiaalinen regressiomalli

nalliset mallit yliarvioivat kiireellisten tehtävien määrän systemaattisesti mutta näyttävät aliarvioivan rakennuspalojen määrää, mikä viittaa siihen, että kertoimen arvo on muuttunut aiemmin määritellystä.



Kuva 23. Kiireellisten tehtävien ennustevirheet vuodelle 2008 ennusteluokittain 5 km × 5 km -ruudukossa käytettäessä ennusteena a) asukasluku*0.0047, b) kerrosala*0,0000655 c) vuosien 2005–2007. Rakennuspalojen ennustevirheet käytettäessä d) asukasluku*0,00047, e) kerrosala*0,0000655 ja vuosien 2005–2007 keskiarvoa.

4. Spatiaalinen regressiomalli

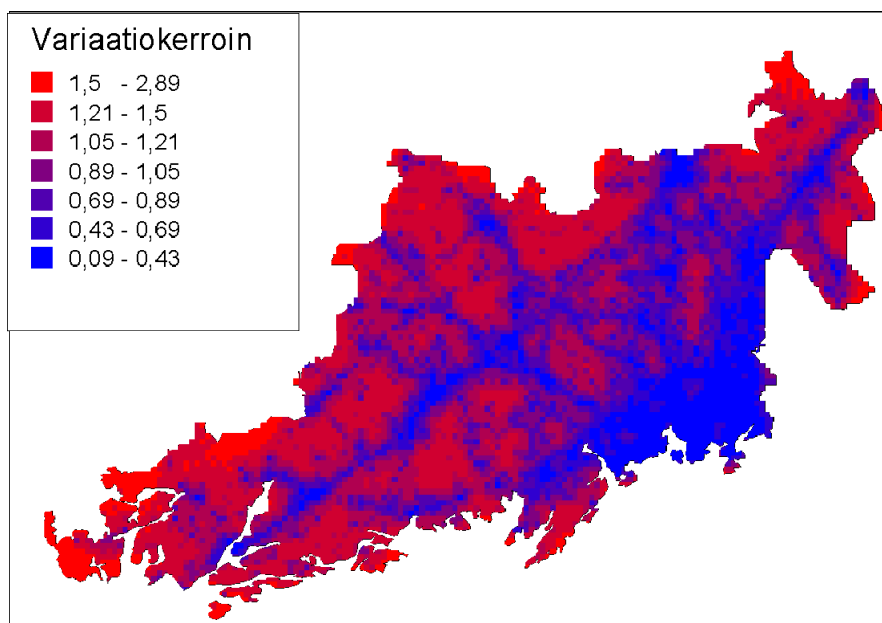


Kuva 24. Vuoden 2008 havaittujen ja spatiaalisella mallilla ennustettujen erotuskartta (estimointivirhe) 1 km × 1 km -ruudukossa alueelle 1 a) kiireelliset tehtävät, b) rakennuspalot ja c) liikenneonnettomuudet.

Alueella 1 ennustevirheiden sijoittumisessa ei ole havaittavissa mitään erityistä alueellisuutta: pienet ja suuret ennustevirheet sijoittuvat lähekkäin suuren riskin alueille (kuva 24).

Spaatialisen regressiomallin sovitusten menetelmä tuottaa intensiteetin todennäköisyysjakauman jokaiselle ruudulle. Kuvassa 25 on esitetty ennustettujen kiireellisten tehtävien määrän variaatiokerroin (keskihajonta/keskiarvo). Estimaatin tarkkuus mallin näkökulmasta kuvastaa paljolti intensiteetin vaihtelua.

4. Spatiaalinen regressiomalli



Kuva 25. Kiireellisten tehtävien spatiaalisen mallin estimaattien variaatiokerroin.

4.6 Johtopäätökset

Esitetyt spatiaalisen mallin antamat kiireellisten tehtävämäärien ennusteet ovat systemaattisesti suurempia kuin vuoden 2008 toteutuma. Pääasiassa ero selittyy sillä, että vuosien 2005–2007 perusteella estimoitu tehtävien määrän kasvu ei toteutunut vuonna 2008. Jotta trendien estimointi olisi luotettavaa, tulisi arvion perustua pidempään ajanjaksoon ja jos mahdollista ajassa muuttuviin selittäjiin. Spatiaaliset mallit mahdollistavat myös spatiaalisesti autokorreloituneen trendin huomioimisen, mutta alueellisen trenditekijän estimointi on mielekäästä vasta pidemmästä aikasarjasta. Osa ajallisista muutoksista heijastaa myös asukasmäärässä ja kerrosalassa tapahtuvia muutoksia. Ilman vuosittaisia selittäjiä näiden huomioiminen on tehotonta. Epäonnistuminen yksittäisen vuoden tehtävämäärän ennustamisessa ei kuitenkaan tarkoita, ettei alueellisia eroja olisi tavoitettu.

Spatiaalisen mallin ennustevirhe yksittäistä vuotta ennustettaessa ei ole merkittävästi pienempi käytettäessä yksinkertaisesti viimeisen kolmen vuoden keskiarvoa. Mallin edut tulevat esiin lähinnä hyvin alhaisissa intensiteeteissä, jolloin ruutukohtaisten tehtävien määrä on vähäinen. Malli ei kuitenkaan ole laskenta-

4. Spatiaalinen regressiomalli

teknisistä syistä sovellettavissa laajemmille alueille pienissä ruuduissa, joten nykyisellään hyöty jää vähäiseksi.

Yksinkertaisiin regressiomalleihin nähden spatiaalisilla malleilla ei päästy parempiin ennusteisiin. Spatiaalisen mallin kiistaton etu kuitenkin on, että se toimii ilman selittäjiä mutta pystyy toisaalta hyödyntämään selittäjiä ennustamiseen, jos niitä on saatavilla. Malli on myös joustava puuttuvien tietojen suhteen, jos niille vain voidaan muodostaa mielekäs priorijakauma. Vastaava ei ole yhtä helpposti toteutettavissa yksinkertaisten regressiomallien estimoinnissa.

Alueellista termiä ja ruututermiä ei voida tulkita täysin yksiselitteisesti, mutta ne antavat kuitenkin viitteitä siitä, missä ruuduissa selittäjien ja selitettävän (tehtävät) välinen riippuvuus poikkeaa keskimääräisestä. Näiden erojen ymmärtäminen voi avata uusia näkökulmia alueellisten erojen tarkasteluun.

Mallin puutteet tulevat sen laskennallisesta raskaudesta ja mallinnuksen vaativuudesta. Spatiaaliset regressiomallit ovat joustava tilastollinen apuväline riskianalyysin tueksi, mutta ne vaativat käyttäjältään tilastotieteellistä osaamista.

Toimintavalmiusohjeen taustalla olevat suhdelukuun perustuvat mallit toimivat yllättäen hyvin siihen nähden, että rakennuspalojen osuus tässä kiireellisiksi määritellyistä tehtävistä oli suurempi kuin mihin toimintavalmiusohjeen perustelumuiotiossa viitataan.

4.7 Kehittämisehdotukset

Spatiaalisten mallien keskeisin rajoite on nykyisellään niiden laskennallinen raskaus. Laskennallisesti keveämpiä spatiaalisia menetelmiä (kuten *Poisson kring*) tulisikin kokeilla riskianalyysin apuvälineinä. Laskentakapasiteetin kasvassa ja menetelmien kehittyessä bayesilaiset spatiaaliset mallit tulevat laajemmin käyttöön, joten niiden hyödyntämistapoja olisi syytä kehittää myös pelastustoimen riskianalyysissä.

Tässä tutkimuksessa esitetyt mallit eivät varmastikaan ole parhaita. Mallien systemaattinen virhe korkeissa ennusteissa viittaa mallin virheelliseen muotoon. Additiivisuus log-asteikolla helpottaa mallien laskentaa, mutta myös vaihtoehtoisia mallityyppejä tulisi testata.

Käytettävissä oleva aineisto oli tehoton mallien selittäjien vaikutusten testaamiseen, sillä samojen ruutujen selittäjistä ei ollut tietoa useammalta eri vuodelta. Olisikin tärkeää testata, miten hyvin spatiaaliset mallit toimivat, jos myös ennustettavan vuoden selittäjien arvot ovat ennustettuja, mutta oletamme kuitenkin

4. Spatiaalinen regressiomalli

tuntevamme niihin liittyvän epävarmuuden. Tällainen sovellustilanne tulee olemaan tulevaisuuden riskianalyysissä todennäköisin.

Spatiaalisten mallien kehittämistä tulisi jatkaa hyödyntämällä pidempiä aikasarjoja ja vuotuisia selittäjien arvoja sekä lisäämällä selittäjiin ainakin ruutujen sisältämien tieosuuksien pituudet. Menetelmä on siinä määrin lupaava, että olisi syytä mallintaa myös muiden onnettomuustyyppien alueellisia jakautumia ja kehittää malleja, joilla ennustetaan yhtä aikaa useiden eri onnettomuustyyppien esiintymistä ruuduissa. Estimointia kun voidaan tehostaa korreloituneiden selittävienkin kautta.

Spatiaalisten mallien estimointi MCMC-menetelmällä mahdollistaa erilaisten malliparametreista johdettujen suureiden laskemisen ilman, että kohdataan erityisiä vaikeuksia estimaattien hajontaominaisuuksien arvioinnissa. Laskentaan voidaan siten yhdistää ongelmitta esimerkiksi spatiaalisesti mallinnettu toimintavalmiusaika ja toimintavalmiuden hyötyfunktio epävarmuuksineen. Tällöin voidaan arvioida vaikkapa hyödyn alueellista jakautumista ja laskea estimaateille mielekkäitä tarkkuuksia.

5. Itseorganisoituvat kartat (SOM)

5.1 Käytetty aineisto

SOM-sovelluksessa käytettiin riskiruuuaineistoa, Pronton onnettomuustietoja sekä yhdessä alueessa Experian Oy:n Mosaic-dataa. Mosaic esitellään tässä raportissa myöhemmin.

Mallia testattiin erikokoisilla ruuduilla ja lopulta 16 kappaletta 250 x 250 m²:n kokoisia ruutuja yhdistettiin yhdeksi 1 x 1 km²:n ruuduksi. Tällöin onnettomuuksien määrissä on jo merkittäviä eroja ruutujen välillä, ja ennustaminen on järkevämpää. Taulukossa 24 on esitetty ruutujen luku- ja onnettomuusmäärät kaikilla alueilla. Pronton onnettomuustiedoista käytettiin yksin rakennuspaloja sekä kaikkia kiireellisiä onnettomuuksia (lueteltu luvussa 1.3.1).

Mallia kehitettiin ja testattiin useilla eri muuttujaryhmillä, alkaen vain muuttujan malleista ja päätyen jopa 41 muuttujan malliin. Parhaiten toimivat mallit, joissa oli eniten muuttujia. Käytettäväksi valittiin muuttujaryhmä, jossa olivat mukana

- koordinaatit (2 kpl)
- asukkaiden määrät ikäryhmittäin (3 kpl)
- rakennusten lkm eri käyttötarkoituksiluokissa (13 kpl)
- kerrosala (m²) eri käyttötarkoituksiluokissa (13 kpl)
- työpaikkojen lkm (1 kpl)
- onnettomuustiedot (rakennuspalo (1 kpl) / kiireelliset onnettomuudet (9 kpl)).

5. Itseorganisoituvat kartat (SOM)

Taulukko 24. Alueiden 1x1 km:n ruutujen ja onnettomuuksien lukumäärät. Vuosien 2005–2007 osalta on esitetty kolmen vuoden keskiarvo.

Malli	Vuodet	Ruutuja	Rakennus-paloja	Kiireellisiä onnettomuuksia
Alue 1	2005–2007	10 862	637	5 610
	2008	10 862	939	6 828
Alue 2a	2005–2007	59 429	1 185	8 337
	2008	59 429	1 296	9 295
Alue 2b	2005–2007	60 076	632	4 295
	2008	60 076	691	4 362
Alue 3a	2005–2007	61 434	567	3 620
	2008	61 434	625	3 438
Alue 3b	2005–2007	20 744	296	2 427
	2008	20 744	332	2 529
Alue 4	2005–2007	68 707	367	2 623
	2008	68 707	368	2 489
Alue 5	2005–2007	101 665	198	1 475
	2008	101 665	229	1 524

5.2 Mallin kuvaus

5.2.1 Teoria

Malli rakennettiin käyttäen neuroverkkojen itseorganisoituvia karttoja (SOM), jotka ovat yksi neuroverkkojen tyyppi. Neuroverkot ovat epälineaarinen estimointimenetelmä, joiden perusajatus tulee luonnollisista neuroverkoista. Neuroverkkoa opetetaan esimerkkien avulla, ja se oppii muuttujien epälineaariset riip-

5. Itseorganisoituvat kartat (SOM)

puvuussuhteet suoraan havaintoaineiston perusteella. Neuroverkot rakentuvat neuroneista, jotka koostuvat kolmesta osasta:

- synapsi – kahden neuronin välinen kytkentä, paino $w_{k,j}$
- summaaaja – summaa kaikilta neuroneilta tulleet signaalit
- aktivaatiofunktio – laskee summaajan epälineaarisen kuvauksen.

Itseorganisoituvat kartat (Kohonen 1995) ovat neuroverkon eräs erikoistapaus. Yhteydet neuronien välillä muunnetaan yksinkertaisiksi geometrisiksi suhteiksi, joiden avulla voidaan helposti visualisoida muuttujat esimerkiksi kaksiulotteista karttaa käyttäen. Karttaa muodostettaessa verkko järjestäytyy siten, että samankaltaiset syötöt ovat lähellä toisiaan klustereissa. Kartan muodostamisen jälkeen sitä voidaan käyttää myös epätäydellisten syötevektoreiden avulla ennustamaan puuttuvia muuttujien arvoja.

Kartan klusterit nähdään U-matriisista, jossa neuronien topologiset etäisyydet esitetään värien avulla. Tässä työssä käytetyn väriskaalan mukaisesti sininen tarkoittaa lyhyttä etäisyyttä ja punainen pitkää. Ominaisuuksiltaan toisiaan lähellä olevat neuronit ovat siis sinisiä, jolloin alueiden (klustereiden) rajat ovat punaisia. Muuttujia kuvataan U-matriisin tapaisilla kaksiulotteisilla kartoilla, mutta niiden tulkinta on erilainen. Muuttujien väriyksessä sininen tarkoittaa pientä muuttujan arvoa ja punainen suurta.

5.2.2 Sovellus riskiruutuaineistoon

Itseorganisoituvat kartat toteutettiin Matlabilla käyttäen SOM-toolboxia (Vesanto et al. 2000). Toolbox on ilmainen, ja se on saatavilla Internetistä osoitteesta <http://www.cis.hut.fi/somtoolbox/>.

Itseorganisoituvia karttoja käytetään yleensä luokittelemaan syötteitä. Tässä työssä niitä sovelletaan hieman eri tavalla, sillä tavoitteena on saada jokaiselle ruudulle tai alueelle onnettomuusmääräennuste selittävien muuttujien perusteella. SOM ryhmittelee muuttujat siten, että samanlaiset syötöt (eli tässä muuttujien osalta samantapaiset ruudut) hakeutuvat samaan verkon noodiin. Kutakin noodia vastaava onnettomuusmääräennuste on noodiin hakeutuneiden ruutujen onnettomuusmäärien keskiarvo. On mahdollista, että syötötietojen perusteella samanlaisissa ruuduissa tapahtuu täysin eri määrä onnettomuuksia. Tällöin noodiin liittyvien onnettomuusmäärien keskihajonta kasvaa suureksi, ja ennusteeksi saatua keskiarvoa on pidettävä epätarkkana. Tämä viittaa osaltaan siihen, ettei kaikkia merkitseviä muuttujia ole otettu huomioon.

5. Itseorganisoituvat kartat (SOM)

Karttaa muodostettaessa käytetään onnettomuustietoja vuosilta 2005–2007. Kun etsitään ennustetta vuodelle 2008, onnettomuustiedot eivät ole mukana, eikä SOM ota niitä huomioon ruutujen ja noodien välisten etäisyyksien laskennassa. Tällöin muiden muuttujien osuus painottuu parhaiten vastaavan noodin etsinnässä, ja samat ruudut saattavat päätyä eri noodiin. Ennuste ei siis ota huomioon juuri tämän ruudun onnettomuushistoriaa vaan vertaa ruutua muiden muuttujien osalta samanlaisten ruutujen historiaan. Tämä on oikea lähestymistapa, mikäli todella kaikki onnettomuuksiin johtavat selittäjät on otettu huomioon karttaa muodostettaessa. Mikäli tärkeitä selittäjiä ei ole huomioitu, ennuste saattaa olla virheellinen.

SOM:n antama ennuste koostuu siis keskiarvosta ja keskihajonnasta. Tästä johtuen tuloksien ennustettu–havaittu–virhe on laskettu virherajojen ylä- ja alarajalta. Muut tarkastelut on tehty ainoastaan keskiarvon perusteella, jolloin virherajat on jätetty huomiotta.

5.3 Tulokset

5.3.1 Rakennuspalot

Rakennuspaloja koskevat analyysit tehtiin kullekin taulukon 1 alueelle erikseen. Käytetyt muuttujat on esitetty luvussa 5.1. Muuttujien perusteella SOM määrittä kullekin riskiruudulle ennusteen ruudussa esiintyvien rakennuspalojen keskimääräisestä lukumäärästä keskihajontoineen. Mallin toimivuutta arvioitiin vertaamalla sen antamia tuloksia vuonna 2008 tapahtuneisiin rakennuspalomääriin. Aluksi tutkittiin SOM:n kykyä tunnistaa joukosta ruudut (1 km × 1 km), joissa on sattunut vähintään kaksi rakennuspaloa. Nämä ruudut on merkitty vertailukuviin 26 punaisella. Sama vertailu on esitetty numeroina taulukossa 25. Esimerkki muuttujien sijoittumisesta kartalla on esitetty liitteessä B.

Kuten kuvasta 26 ja taulukosta 25 nähdään, SOM:n ennustama yli kaksi rakennuspaloa sisältävien ruutujen määrä on liian alhainen, kun sitä verrataan vuonna 2008 toteutuneeseen tilanteeseen. Alueille 2b, 3a, 3b, 4 ja 5 SOM ei löytänyt yhtään I-riskiluokkaa vastaavaa ruutua. Kuvasta 26 nähdään suoraan, miten alueet sijoittuvat kartalle toisiinsa nähden. Niillä alueilla, joille SOM ennusti suuriakin onnettomuusmääriä, punaiset ruudut sijoittuvat likimain oikeisiin kohtiin, mutta niitä on liian vähän.

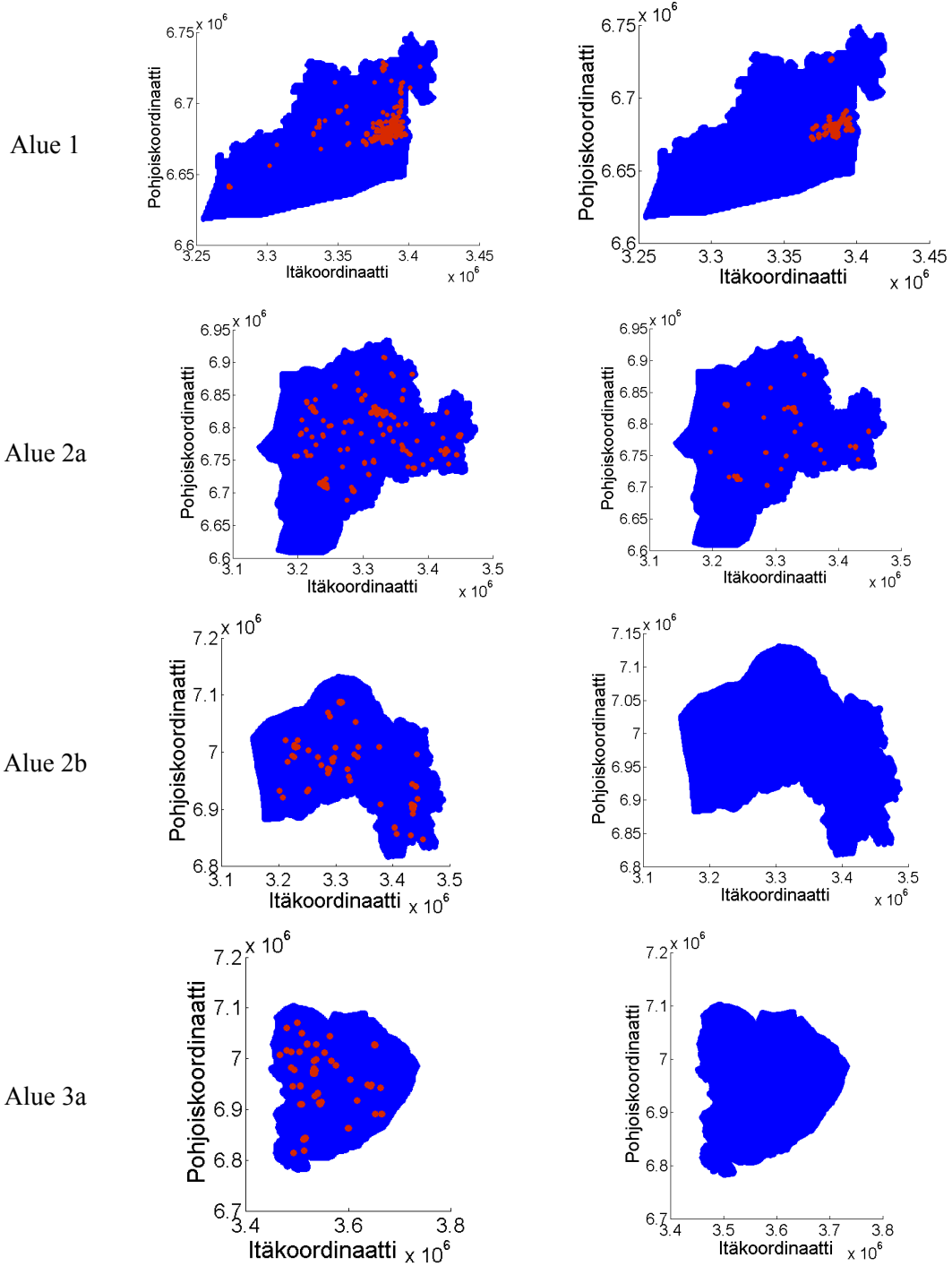
5. Itseorganisoituvat kartat (SOM)

Kuvassa 27 on esitetty rakennuspalojen lukumäärän ennustettu–havaittu–virhe. Tulokset saatiin jakamalla ruudut kymmeneen luokkaan ennustetun rakennuspalomäärän mukaisesti ja laskemalla virhe kullekin luokalle erikseen. Kuvasta nähdään, että tyypillisesti ennustevirhe kasvaa onnettomuusmäärän kasvaessa. Useimmissa tapauksissa ennustevirheen mediaani on positiivinen tai nolla, eli SOM on ennustanut keskisuuria lukumääriä todellista enemmän. Lähes kaikissa tapauksissa toteutunut onnettomuusmäärä on ennusteeseen liittyvän keskihajonnan sisällä, joten tilastollisessa mielessä virheitä ei voi pitää kovin merkittävinä.

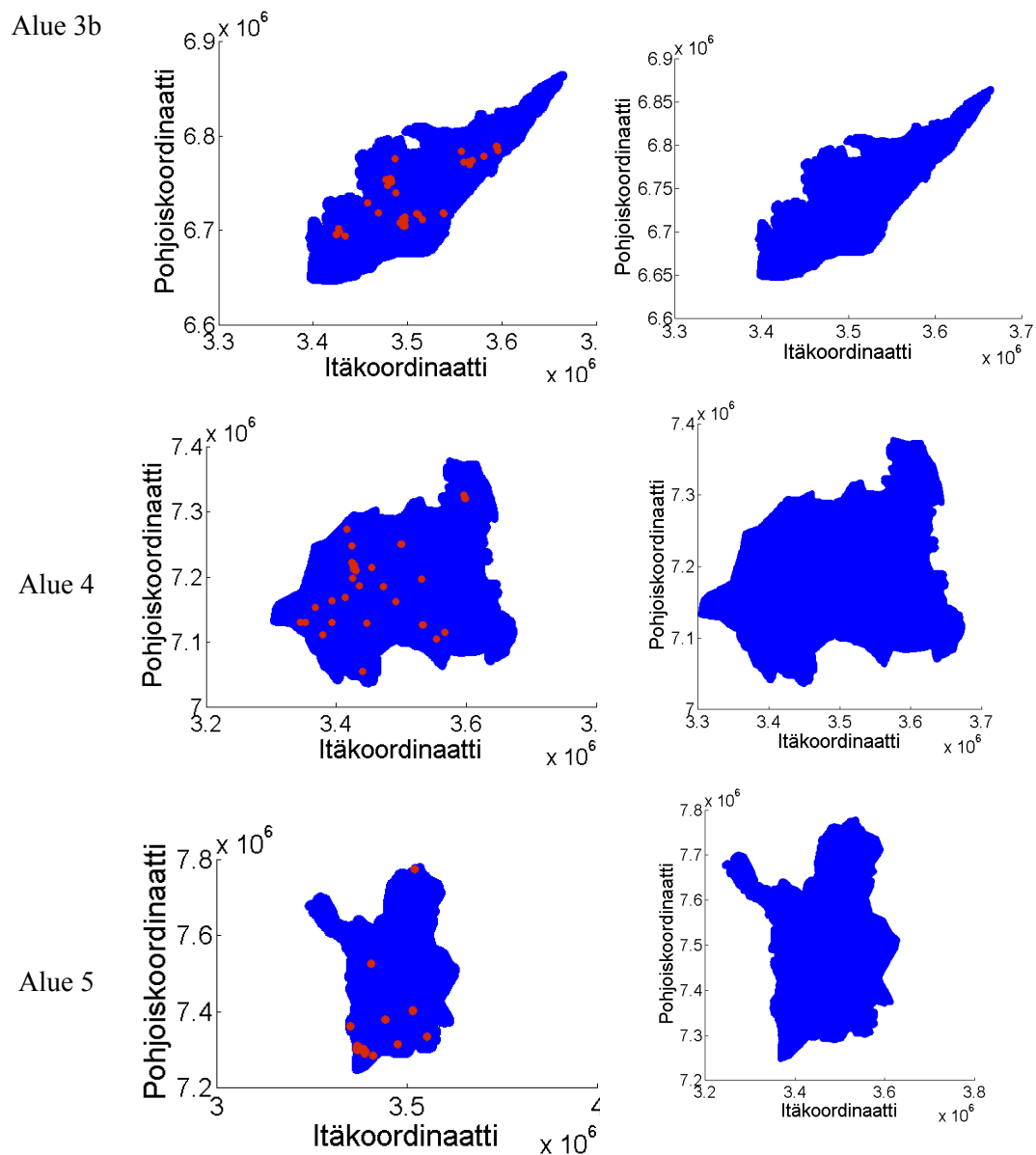
Taulukko 25. Niiden ruutujen prosentuaalinen osuus eri alueilla, joissa esiintyy vähintään kaksi rakennuspaloa vuodessa (1 km × 1 km ruudut). Vasemmalla vuonna 2008 toteutuneet onnettomuudet ja oikealla SOM:n ennusteet.

Alue	Toteutuneet (%)	SOM (%)
1	1,7	0,55
2a	0,35	0,13
2b	0,12	0
3a	0,12	0
3b	0,19	0
4	0,06	0
5	0,03	0

5. Itseorganisoituvat kartat (SOM)

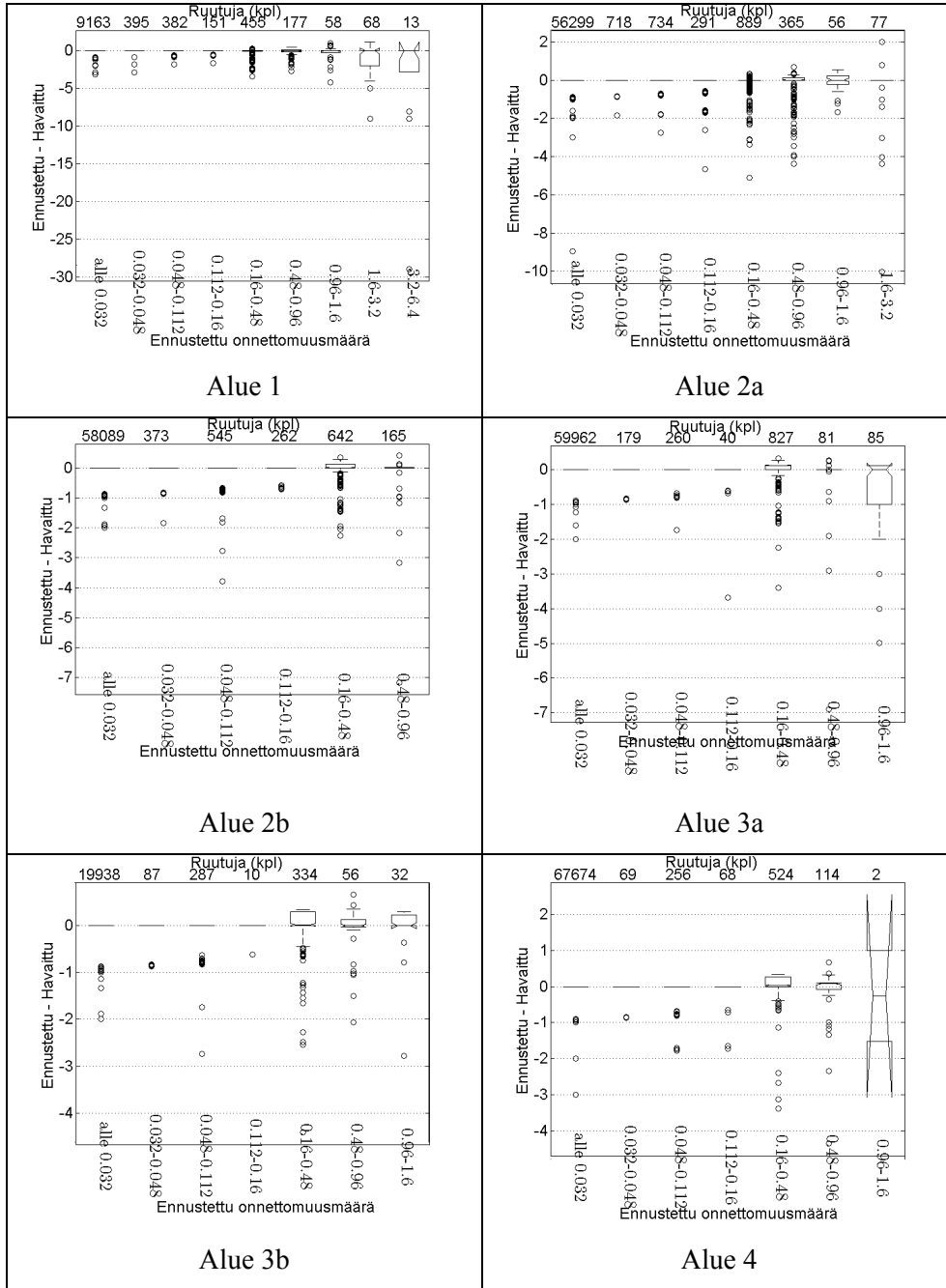


5. Itseorganisoituvat kartat (SOM)

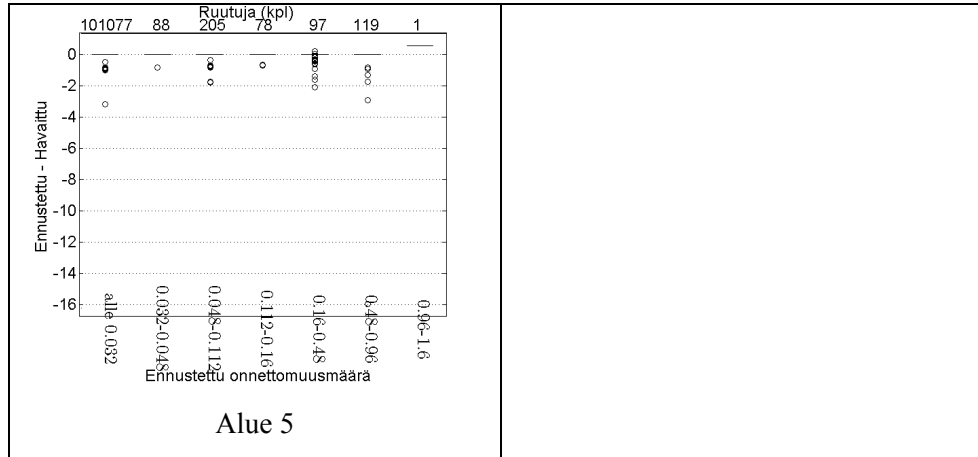


Kuva 26. Suurimmat onnettomuusmäärät (vähintään kaksi onnettomuutta 1 km \times 1 km -ruudussa, kartalla punainen väri). Vasemmalla toteutuneet onnettomuudet, oikealla SOM:n ennusteet.

5. Itseorganisoituvat kartat (SOM)



5. Itseorganisoituvat kartat (SOM)



Kuva 27. Ennustevirhe (ennustettu–havaittu), kun ruudut (1 km × 1 km) on jaettu ennustetun rakennuspalolukumäärän mukaisesti kymmeneen luokkaan.

5.3.2 Kiireelliset onnettomuudet

Samoin kuin rakennuspalotarkastelut myös kiireellisiä onnettomuuksia koskevat analyysit tehtiin kullekin taulukon 1 alueelle erikseen. Käytetyt muuttujat on esitetty luvussa 5.1. Muuttujien perusteella SOM määrittä kullekin riskiruudulle ennusteen ruudussa esiintyvien kiireellisten onnettomuuksien keskimääräisestä lukumäärästä keskihajontoineen. Mallin toimivuutta arvioitiin vertaamalla sen antamia tuloksia vuonna 2008 tapahtuneisiin kiireellisten onnettomuuksien määriin ja arvioimalla sitä, miten SOM tunnistaa joukosta ruudut (1 km × 1 km), joissa on sattunut tietty määrä kiireellisiä onnettomuuksia.

Kuvassa 28 ruudut on jaettu toteutuneen tai ennustetun onnettomuusmäärän mukaisesti neljään luokkaan. Käytetyt onnettomuusmäärien luokkarajat on esitetty taulukossa 4 ja rajojen valintaperusteet luvussa 1.5. Eri luokkiin sijoittuvat ruudut on merkitty väreillä seuraavasti: luokka I (punainen), luokka II (keltainen), luokka III (vihreä), luokka IV (sininen). Prosentuaaliset osuudet on esitetty taulukossa 26.

Kuvasta 28 ja taulukosta 26 nähdään, että SOM ei ennustanut lainkaan ruutuja luokkaan I muualla kuin alueella 1. SOM ennusti toteutunutta vähemmän ruutuja myös luokkaan II kaikilla muilla alueilla paitsi 2b:ssä, 3a:ssa ja 4:ssä, joissa ennustus osui hyvin lähelle toteutunutta. Luokan III ruutuja SOM ennusti vähemmän ja luokan IV ruutuja enemmän kaikilla alueilla toteutuneeseen tilantee-

5. Itseorganisoituvat kartat (SOM)

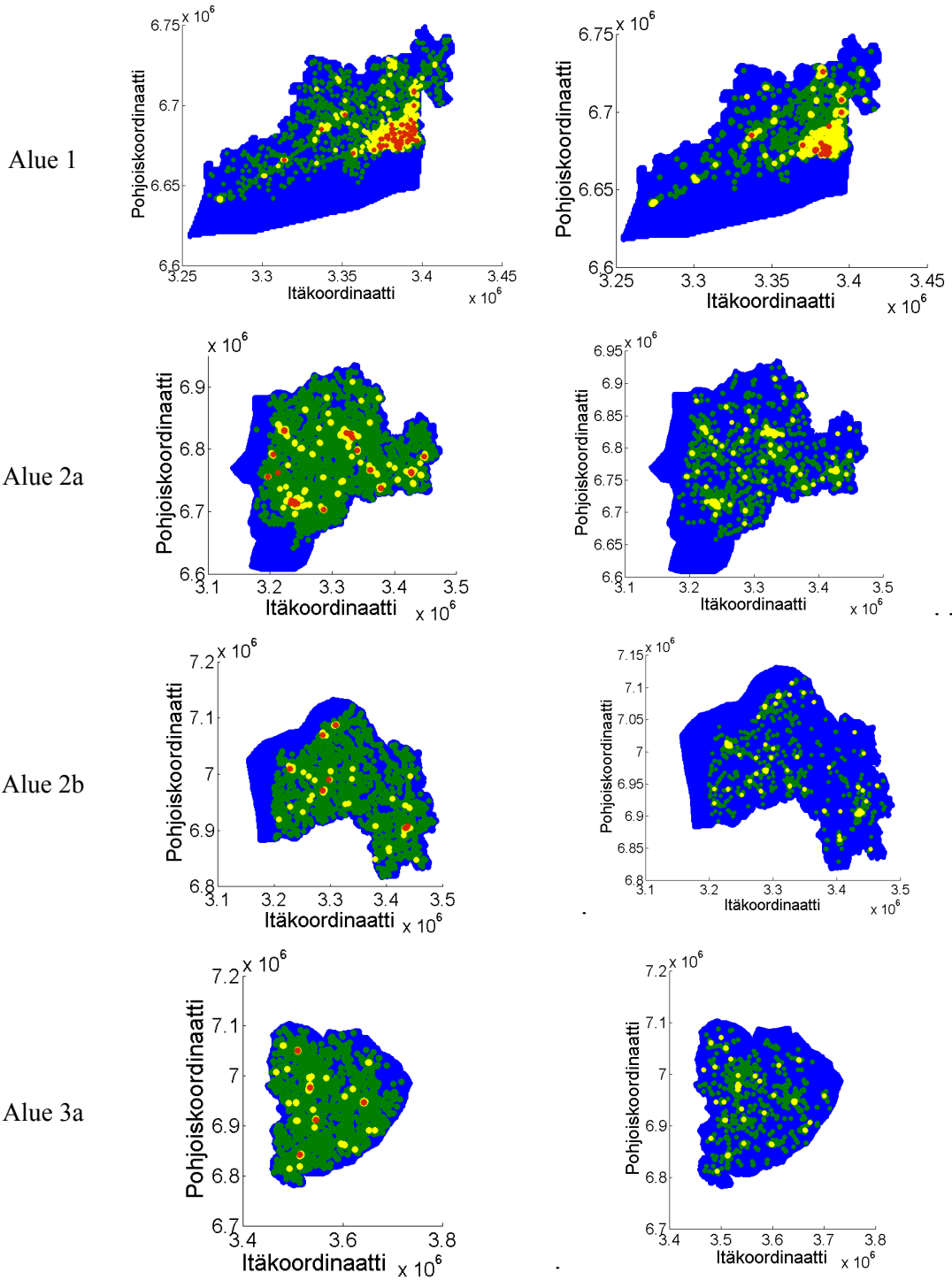
seen verrattuna. Erot toteutuneisiin ovat kuitenkin melko pieniä luokissa II–IV. Vain luokassa I virheitä voi pitää merkittävänä. Kuvan 28 perusteella näyttää siltä, että ainakin luokan II ruudut vastaavat maantieteellisesti melko hyvin toteutuneiden onnettomuuksien riskiluokkia I ja II vierekkäisillä kartoilla. Edellä kuvatut suhteet ruutujen lukumäärissä ovat nähtävissä myös kuvasta 28.

Kuvassa 29 ruudut on jaettu kymmeneen luokkaan ennustetun rakennuspalomäärän mukaisesti, ja kullekin luokalle on määritetty ennustettu–havaittu-virhe. Kuvasta 29 nähdään, että useimmilla tarkastelualueilla ennustevirheen mediaani pysyy nollan tienoilla onnettomuusmäärän kasvaessa.

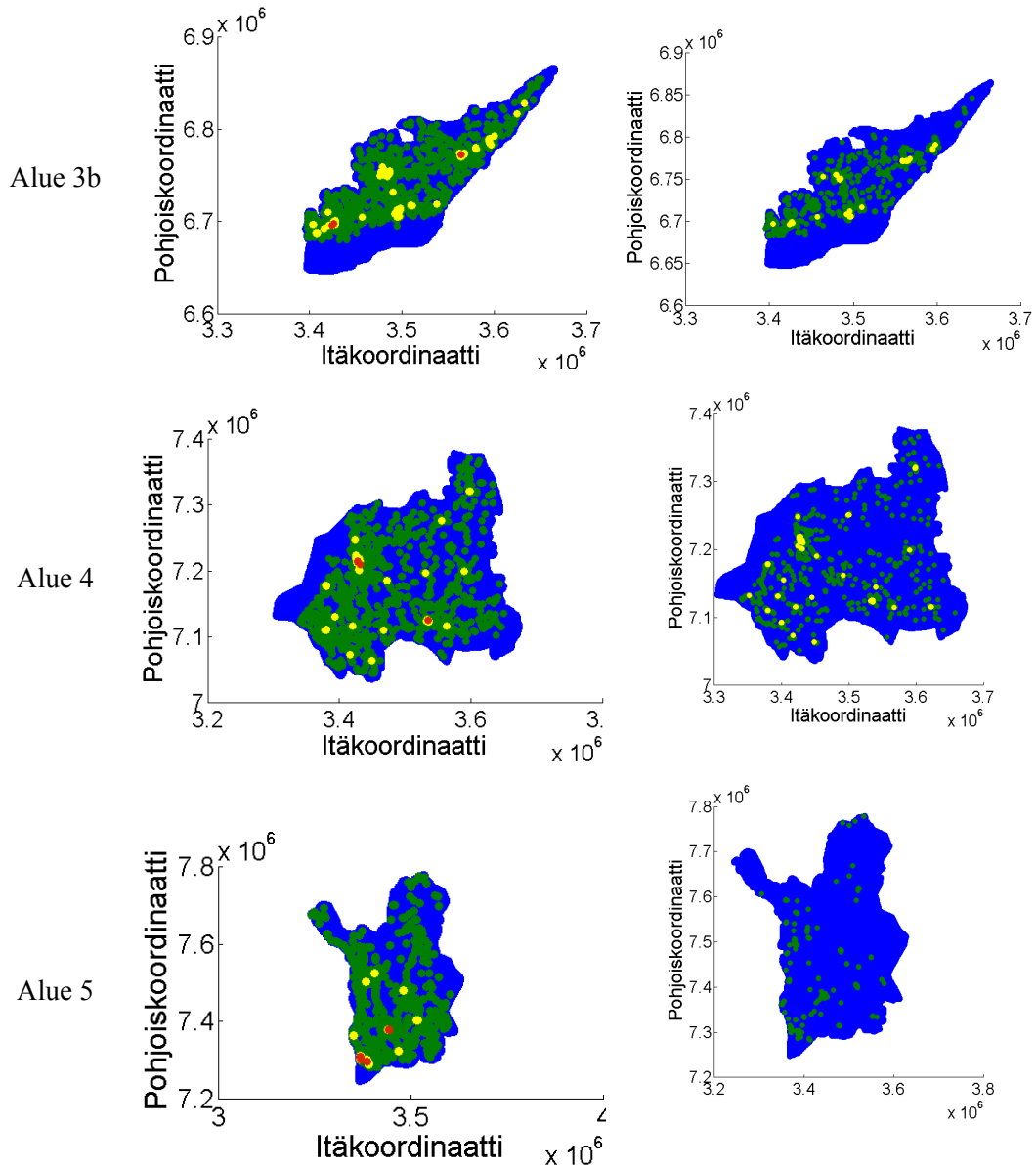
Taulukko 26. Eri luokkiin jaettujen ruutujen prosentuaalinen osuus eri alueilla. Ruudut on jaettu luokkiin onnettomuusmäärän perusteella. Käytetyt luokkarajat on esitetty taulukossa 4. Otsikko Tot. (%) tarkoittaa 2008 toteutuneiden onnettomuuksien mukaan määritettyä luokkajakoa.

Alue	Luokka I		Luokka II		Luokka III		Luokka IV	
	Tot (%)	SOM (%)	Tot (%)	SOM (%)	Tot (%)	SOM (%)	Tot (%)	SOM (%)
1	0,7	0,3	3,1	2,7	12	8,2	85	89
2a	0,1	0	0,6	0,4	6,5	3,6	93	96
2b	0,02	0	0,2	0,2	4,0	1,6	96	98
3a	0,01	0	0,2	0,1	3,4	1,5	96	98
3b	0,01	0	0,5	0,2	6,1	3,5	93	96
4	0,01	0	0,1	0,1	2,1	1,1	98	99
5	0,004	0	0,03	0	0,93	0,27	99	99,7

5. Itseorganisoituvat kartat (SOM)

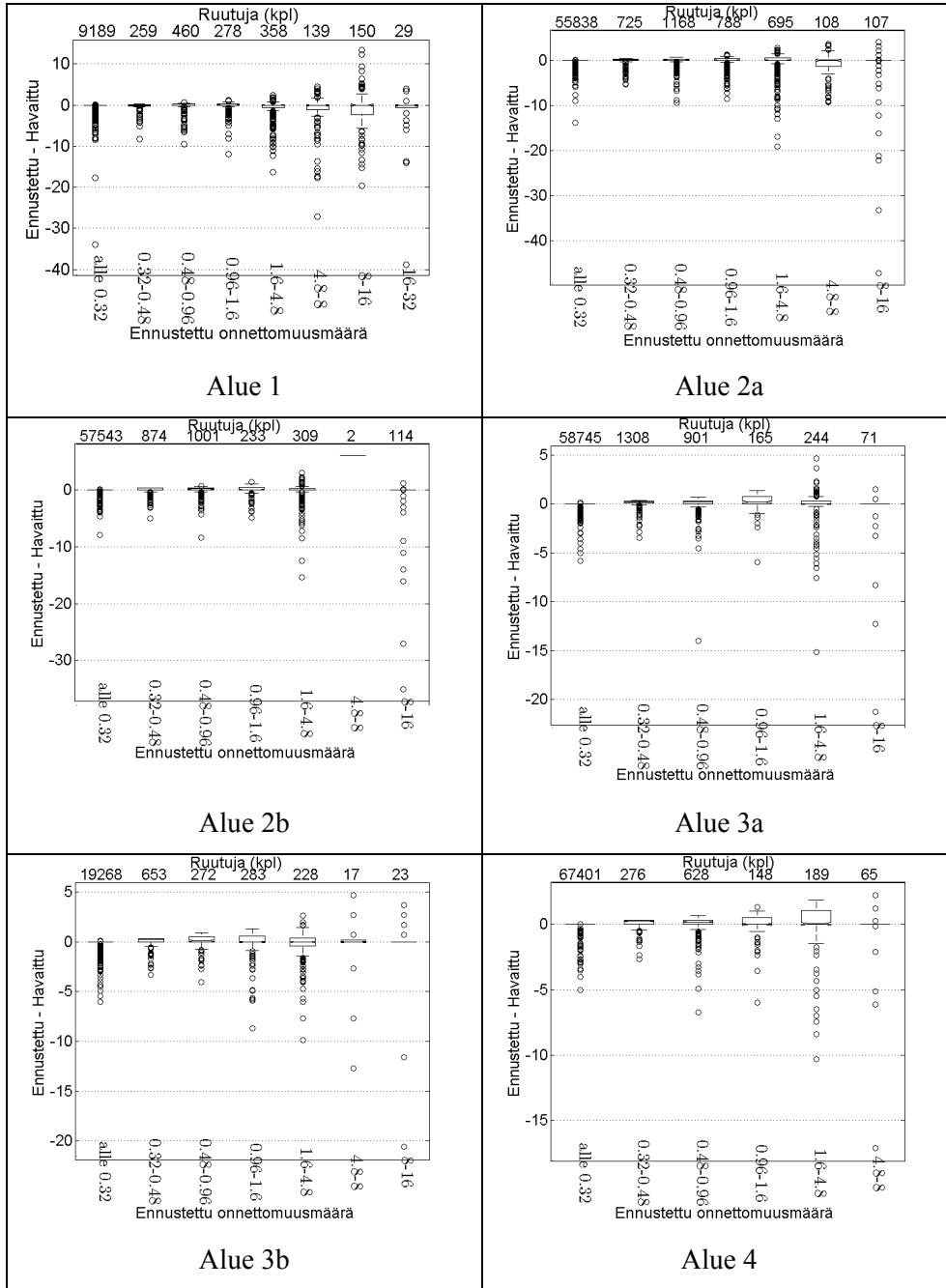


5. Itseorganisoituvat kartat (SOM)

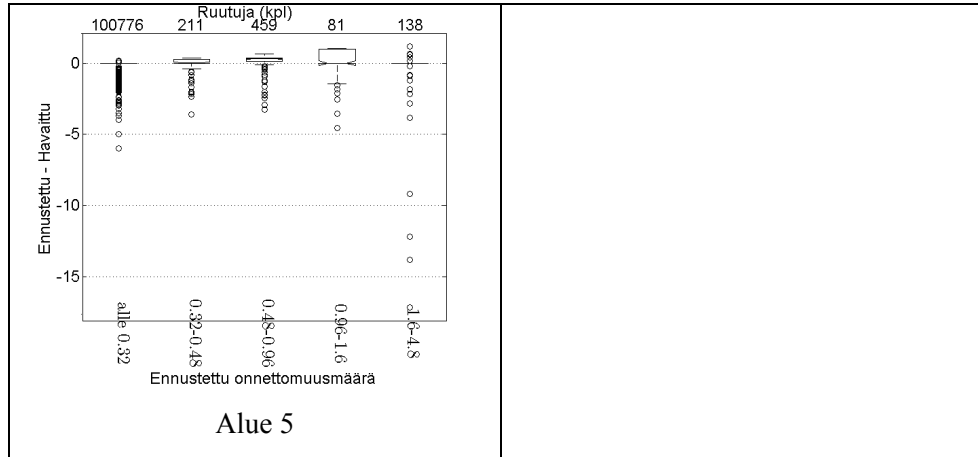


Kuva 28. Riskiruutujen ($1 \text{ km} \times 1 \text{ km}$) sijoittuminen kartalle, kun ruudut on jaettu toteutuneen tai ennustetun onnettomuusmäärän mukaisesti neljään luokkaan. Käytetyt onnettomuusmäärien luokkarajat on esitetty taulukossa 1 ja rajojen valintaperusteet luvussa 1.5. Eri luokkiin sijoittuvat ruudut on merkitty eri väreillä seuraavasti: luokka I (punainen), luokka II (keltainen), luokka III (vihreä), luokka IV (sininen). Vasemmalla toteutumat ja oikealla ennusteet.

5. Itseorganisoituvat kartat (SOM)



5. Itseorganisoituvat kartat (SOM)



Kuva 29. Ennustevirhe (ennustettu-havaittu), kun ruudut (1 km× 1 km) on jaettu ennustetun rakennuspalolukumäärän mukaisesti 10 luokkaan.

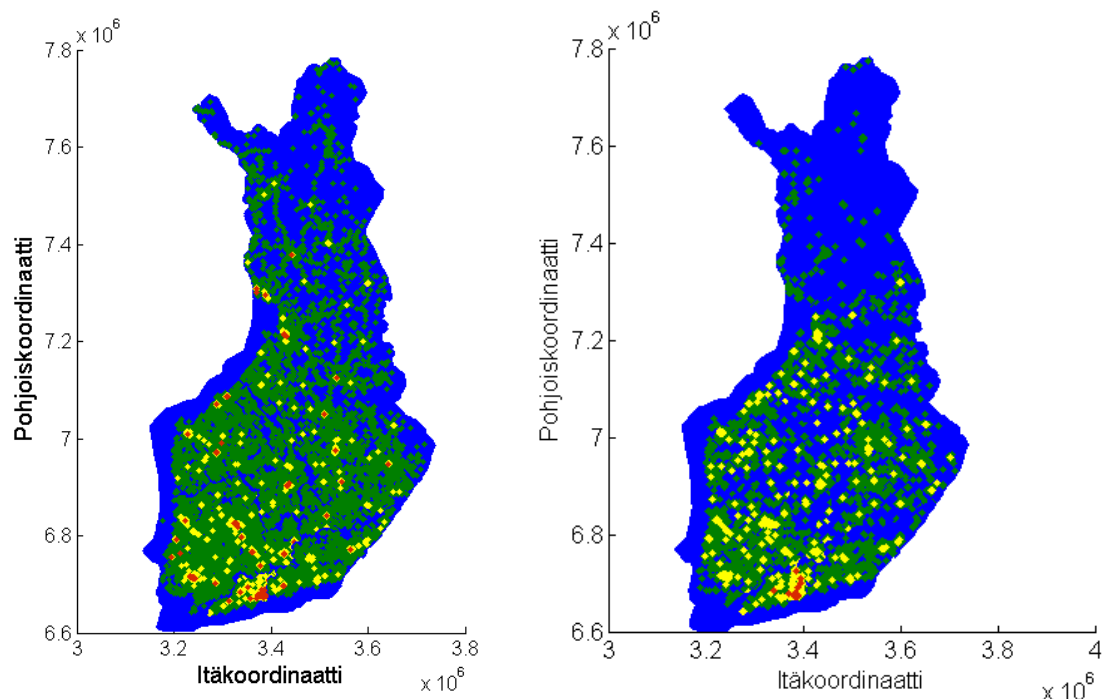
5.4 Mallin toimivuuden arviointi

5.4.1 Vertailu vuonna 2008 toteutuneisiin onnettomuusmääriin

Edellä kuvatut tarkastelut osoittivat, että kaikilla tarkastelualueilla SOM aliarvioi niiden ruutujen määrän, joissa esiintyy useita rakennuspaloja. Kiireellisten onnettomuuksien osalta ennusteet olivat samankaltaisia kaikilla alueilla.

Kuvassa 30 on esitetty ruutujen jakautuminen onnettomuusmäärän perusteella neljään luokkaan koko Suomen alueella. Kuva 30 on yhdistelmä kuvasarjasta 28. SOM on siis muodostettu jokaiselle alueelle erikseen mutta tulokset on piirretty samaan kuvaan. Tämä kartta kertoo saman kuin edellä, eli suuria onnettomuusmääriä ennustetaan kaikille alueille liian vähän tai ei ollenkaan.

5. Itseorganisoituvat kartat (SOM)



Kuva 30. Riskiruutujen (1 km × 1 km) sijoittuminen kartalle, kun ruudut on jaettu toteutuneen tai ennustetun onnettomuusmäärän mukaisesti neljään luokkaan. Käytetyt onnettomuusmäärien luokkarajat on esitetty taulukossa 1 ja rajojen valintaperusteet luvussa 1.5. Eri luokkiin sijoittuvat ruudut on merkitty eri väreillä seuraavasti: luokka I (punainen), luokka II (keltainen), luokka III (vihreä), luokka IV (sininen). Vasemmalla toteutumat ja oikealla ennusteet.

Kuten luvussa 5.2.2 jo todettiin, kartan muodostusvaiheessa muuttujissa oli mukana toteutuneet onnettomuudet vuosilta 2005–2007. Tämän jälkeen testattiin, miten tapahtuneita onnettomuuksia onnistutaan ennustamaan aineistolla, josta on poistettu tieto aiemmin toteutuneista onnettomuusmääristä. Tätä tulosta verrattiin aiemmin vuonna 2008 toteutuneisiin onnettomuusmääriin. Taulukossa 27 on esitetty prosentuaalinen lukumäärä ruuduista, jotka hakeutuvat eri noodiin (eli ryhmittyvät eri ruutujen kanssa), kun aineistossa on mukana onnettomuustiedot ja kun ne on poistettu. Rakennuspaloanalyseissä eri noodeihin hakeutuneiden ruutujen lukumäärä oli noin 3 % vaihdellen välillä 0,4–6 %.

Kiireellisten onnettomuuksien tapauksessa onnettomuustietojen poistamisella oli suurempi vaikutus: eri noodeihin hakeutuneiden ruutujen lukumäärä vaihteli välillä 7–17 %. Eräs selitys suuremmalle osuudelle on se, että rakennuspaloissa

5. Itseorganisoituvat kartat (SOM)

onnettomuuslukumäärätiedot kattoivat vain noin 3 % kaikista muuttujista, mutta yhdeksän kiireellistä onnettomuustyyppiä kattoi muuttujista lähes 22 %.

Taulukko 27. Niiden ruutujen lukumäärä ja osuus, jotka ilman onnettomuustietoja hakeutuivat eri noodiin kuin onnettomuustietojen kanssa.

Alue	Rakennuspalot	Kiireelliset tehtävät
1	421 (3,9 %)	1 791 (16,5 %)
2a	3 803 (6,4 %)	8 885 (15,0 %)
2b	2 103 (3,5 %)	6 500 (10,8 %)
3a	2 273 (3,7 %)	7 268 (11,8 %)
3b	670 (3,2 %)	3 460 (16,7 %)
4	1 326 (1,9 %)	5 435 (7,9 %)
5	386 (0,4 %)	7 127 (7,0 %)

Ennustettujen ja toteutuneiden onnettomuuksien kokonaismäärien vertailu on esitetty taulukossa 28. SOM ennustaa alueella tapahtuvien onnettomuuksien kokonaismäärän hieman liian pieneksi vuoden 2008 onnettomuuksiin verrattuna. Vuoteen 2005–2007 verrattuna kokonaismäärät ovat lähempänä toisiaan. Rakennuspalloilla ennustettujen ja toteutuneiden onnettomuusmäärien suhde vaihteli välillä 0,62–0,79. Kiireellisten onnettomuuksien suhdeluku vaihteli välillä 0,71–0,85. Ennustettujen ja toteutuneiden onnettomuusmäärien suhde oli parhaimmillaan samaa suuruusluokkaa kuin vuosien välinen ero (2005–2007 keskiarvo ja 2008) oikeastikin oli. Niillä alueilla, joilla poikkeama oli suurempi, selityksenä voi olla edellä mainittu ruutujen hakeutuminen eri noodeihin riippuen siitä, ovatko onnettomuustiedot mukana vai eivät.

Mikäli kaikki ruudut osuisivat aina samoihin noodeihin, kokonaismäärän tulisi olla lähellä oikeaa (ottaen huomioon vuosien välisen vaihtelun). Nyt osa ruuduista kuitenkin liitetään eri noodiin, koska onnettomuuksista ei ole tietoa. Näiden ruutujen osuus ei ole suuren suuri mutta lukumääräisesti kuitenkin merkittävä. Ennusteen kokonaismäärien erot verrattuna vuosiin 2005–2007 selittyvät täysin tällä. Mikäli verrataan eri noodeihin menevien ruutujen ennusteita onnettomuustiedoilla ja ilman, erotus on täsmälleen sama kuin toteutuneiden ja ennusteen kokonaismäärien ero. Esimerkiksi sellaiset ruudut, joissa tapahtuu hyvin paljon onnettomuuksia, kerääntyvät samoihin noodeihin, jolloin noodin ennuste

5. Itseorganisoituvat kartat (SOM)

olisi keskiarvojen mukaisesti suuri. Kun onnettomuusmäärät poistetaan, nämä ruudut saattavat hakeutua noodeihin, joiden antama ennuste on paljon pienempi. Tällöin ennustettujen onnettomuuksien kokonaismäärä laskee. SOM ei pysty kontrolloimaan ennustusten kokonaismäärää.

Taulukko 28. Ennustettujen ja toteutuneiden onnettomuuksien kokonaismäärien vertailu. Suluissa on ilmoitettu suhde ennustettu–toteutunut. Vuosien 2005–2007 onnettomuusmäärät ovat kolmen vuoden keskiarvo.

Rakennuspalot			
Alue	Ennustettu	Toteutunut 2005–2007	Toteutunut 2008
1	602	637 (0,95)	939 (0,64)
2a	958	1 185 (0,81)	1 296 (0,74)
2b	439	632 (0,69)	691 (0,64)
3a	463	567 (0,82)	625 (0,74)
3b	207	296 (0,70)	332 (0,62)
4	290	367 (0,79)	368 (0,79)
5	156	198 (0,79)	229 (0,68)
Kiireelliset tehtävät			
	Ennustettu	Toteutunut 2005–2007	Toteutunut 2008
1	5 370	5 610 (0,96)	6 828 (0,79)
2a	7 290	8 337 (0,87)	9 295 (0,78)
2b	3 552	4 295 (0,83)	4 362 (0,81)
3a	2 911	3 620 (0,80)	3 438 (0,85)
3b	1 900	2 427 (0,78)	2 529 (0,75)
4	1 839	2 623 (0,70)	2 489 (0,74)
5	1 087	1 475 (0,74)	1 524 (0,71)

SOM:n vahvuus – ja tässä myös heikkous – on syötevektoreiden luokittelu selittäjien avulla. Vahvuus se on siksi, että SOM näyttää helposti ymmärrettävässä muodossa muuttujien väliset riippuvuussuhteet. Mikäli kaikki merkittävät selittäjät olisivat käytössä, SOM pystyisi ennustamaan onnettomuuksien määrää

5. Itseorganisoituvat kartat (SOM)

paremmin pelkkien taustamuuttujien avulla, onnettomuushistoriasta välittämättä. Tällöin SOM:n voitaisiin ajatella löytävän muuttujien perusteella myös sellaiset alueet, joilla voisi tapahtua paljon onnettomuuksia mutta joilla syystä tai toisesta näitä ei ole tarkastelujakson aikana tapahtunut. Lisäksi SOM pystyy tehokkaasti käsittelemään useita kymmeniä muuttujia yhtä aikaa. Sen heikkous on kuitenkin siinä, ettei sitä ole kehitetty antamaan absoluuttisia ennusteita syötteille. Vajaille selittäjä tiedoilla tulokselta ei voida edes odottaa suurta tarkkuutta.

5.4.2 Vertailu toimintavalmiusohjeen riskiluokitukseen

Tässä luvussa verrataan toimintavalmiusohjeen ja itseorganisoituvien karttojen antamia riskiluokitusennusteita. Tarkastelu tehdään erikseen rakennuspaloilta ja kiireellisille onnettomuuksille.

Toimintavalmiusohjeen mukaan riskiluokat määräytyvät joko kerrosalan tai asukasluvun mukaan, mutta käytännössä kerrosala on useimmiten määräävä. Jotta tässä esitetty vertailu voidaan tehdä, kullekin riskiluokalle pitää antaa arvio siitä, kuinka paljon onnettomuuksia kunkin luokan ruuduissa keskimäärin tapahtuu. Tätä on pohdittu enemmän luvussa 1.5.3.

Tuloksia tulkittaessa on kuitenkin otettava huomioon, että toimintavalmiusohjeen ennustuskyvyn vertaaminen kiinnitettyihin, asukasluvun perusteella lasketuihin rajoihin on ongelmallista monestakin syystä. Ensinnäkin toimintavalmiusohjeen luokkajako riippuu useimmiten kerrosalasta eikä niinkään asukasluvusta. Toisekseen riskiluokkiin jako tapahtuu selittäjien eikä onnettomuusmäärien mukaan, kuten on tehty toteutuneiden ja mallin ennustuksen tapauksessa.

5.4.2.1 Rakennuspalot

Rakennuspalojen vähäisemmän lukumäärän vuoksi niihin liittyvät vertailut tehtiin vain riskiluokan I osalta. Taulukossa 29 on esitetty, kuinka suuri osa ruuduista sijoittuu riskiluokkaan I valituilla onnettomuusmäärärajoilla vuoden 2008 toteutuneiden onnettomuuksien sekä SOM:n antaman ennusteen perusteella. Lisäksi taulukossa 29 on esitetty myös se osuus ruuduista, joka sijoittuu riskiluokkaan I toimintavalmiusohjeessa annettujen asukasluvu- ja kerrosalaehtoien perusteella. Ruutujen sijoittuminen kartalle on esitetty kuvassa 31.

Kuten tuloksista nähdään, toimintavalmiusohje luokittelee I riskiluokkaan jokaisella alueella liian vähän ruutuja, jos tilannetta verrataan vuoden 2008 toteu-

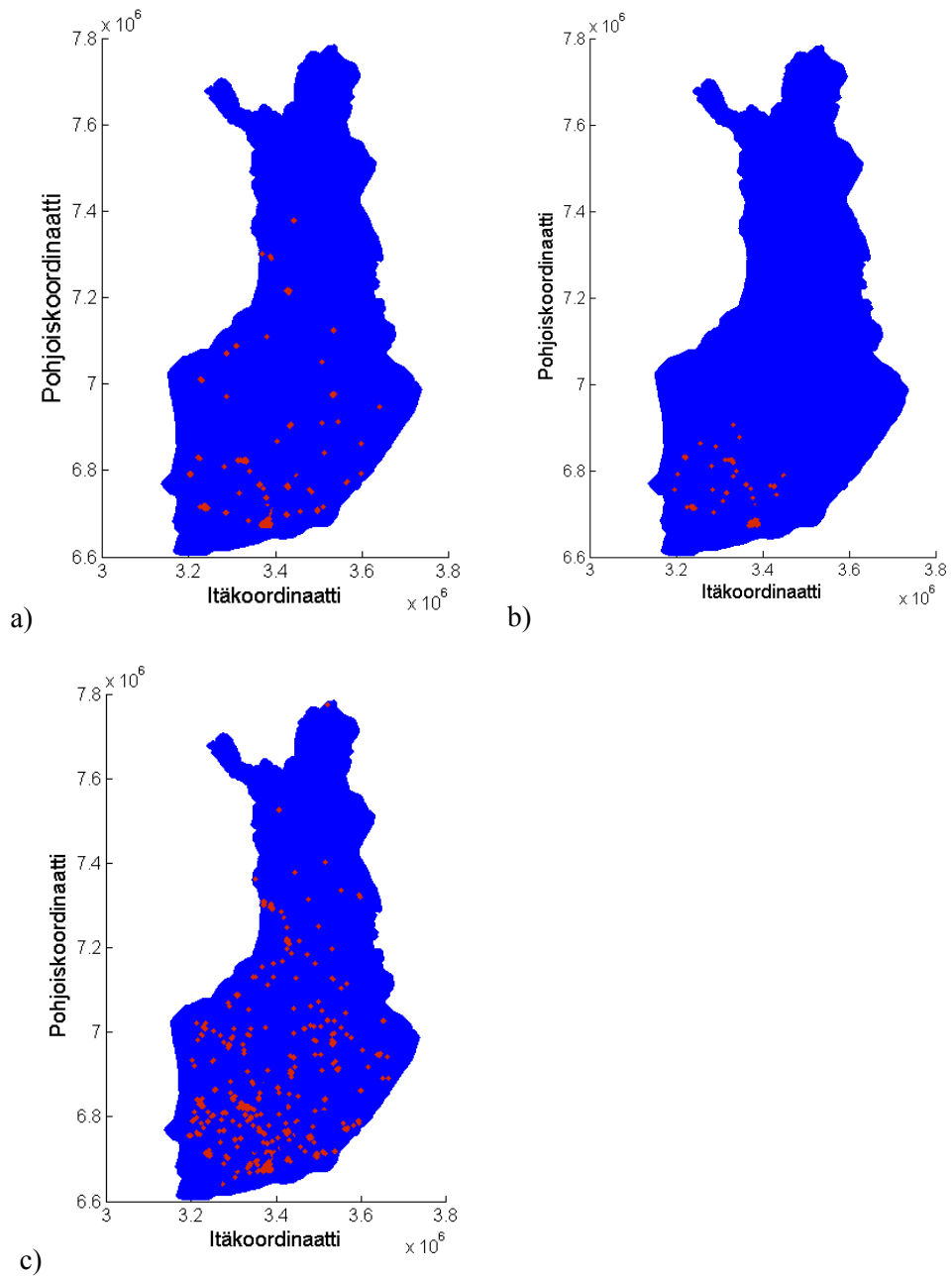
5. Itseorganisoituvat kartat (SOM)

tuneisiin onnettomuusmääriin. Toisaalta SOM luokittelee tähän samaan luokkaan vieläkin vähemmän ruutuja kaikilla alueilla.

Taulukko 29. Riskiluokkaan I sijoittuvien ruutujen osuus kaikista ruuduista eri alueilla, kun onnettomuusmäärinä on käytetty rakennuspaloille määritettyjä onnettomuusrajoja (Tot. 2008 ja SOM). A:71-luokittelu perustuu toimintavalmiusohjeessa annettuihin asukasluoja kerrosalaehtoihin.

Alue	Luokka I		
	Tot. 2008(%)	SOM (%)	A:71 (%)
1	1,7	0,6	1,4
2a	0,4	0,1	0,1
2b	0,12	0	0,04
3a	0,12	0	0,04
3b	0,19	0	0,09
4	0,06	0	0,02
5	0,03	0	0,01

5. Itseorganisoituvat kartat (SOM)



Kuva 31. I riskiluokan ruutujen sijoittuminen kartalla a) toimintavalmiusohjeen mukaan, b) SOM:n rakennuspalomääräennusteen mukaan ja c) toteutuneiden rakennuspalomäärien mukaan määritettynä.

5. Itseorganisoituvat kartat (SOM)

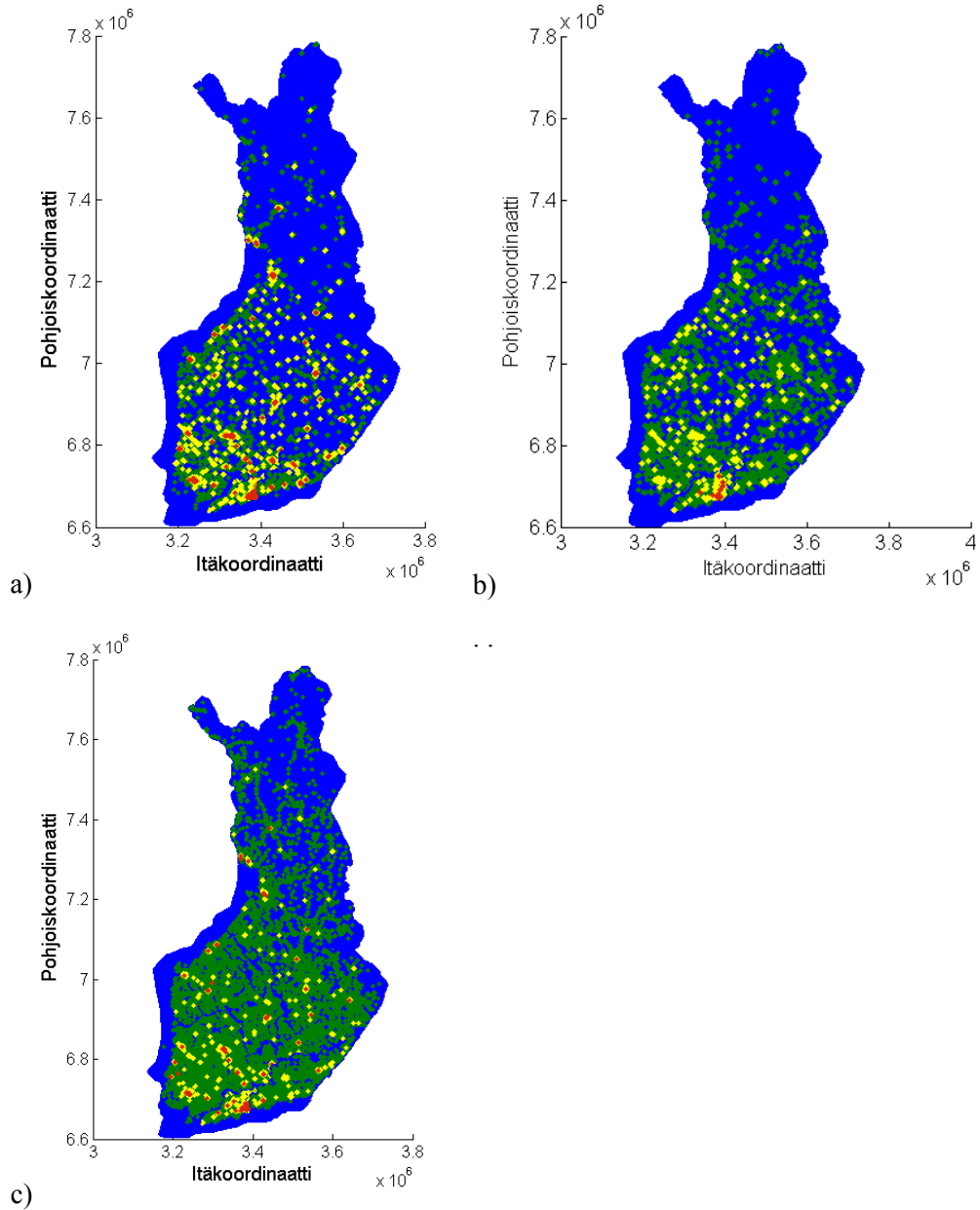
5.4.2.2 Kiireelliset onnettomuudet

Vertailun tulokset kiireellisten onnettomuuksien osalta näkyvät taulukossa 30 ja kuvassa 32. Kuten jo aiemmin havaittiin, SOM ennustaa luokkaan I toteutunutta vähemmän ruutuja. Toisaalta taas toimintavalmiusohje sijoittaa sinne jonkin verran enemmän ruutuja kuin toteutuneiden onnettomuuksien perusteella tulisi tehdä. Luokan II osalta lukumäärät ovat useimmilla alueilla hieman lähempänä toteutunutta tilannetta kuin toimintavalmiusohjeen ennustamat määrät. III-riskiluokan osuudet sekä toimintavalmiusohje että SOM sen sijaan aliarvioivat selvästi. Toimintavalmiusohje ennustaa paremmin kuin SOM etelän alueilla, kun taas SOM toimii paremmin pohjoisessa.

Taulukko 30. Kiireellisten onnettomuuksien jakautuminen riskiluokkiin.

Alue	Luokka I			Luokka II		
	Tot. (%)	SOM (%)	A:71 (%)	Tot. (%)	SOM (%)	A:71 (%)
1	0.7	0.3	1.4	3.1	2.7	3.7
2a	0.1	0	0.1	0.6	0.4	1.1
2b	0.02	0	0.04	0.2	0.2	0.5
3a	0.01	0	0.04	0.2	0.1	0.4
3b	0.01	0	0.09	0.5	0.2	0.9
4	0.01	0	0.03	0.1	0.1	0.3
5	0.004	0	0.01	0.03	0	0.07
Alue	Luokka III			Luokka IV		
	Tot. (%)	SOM (%)	A:71 (%)	Tot. (%)	SOM (%)	A:71 (%)
1	12	8,2	7	85	88,9	88
2a	7	3,6	4	93	96,0	95
2b	4	1,6	2	96	98,2	97
3a	3	1,5	1	96	98,4	98
3b	6	3,5	3	93	96,4	96
4	2	1,1	1	98	98,8	99
5	0,9	0,27	0,4	99	99,7	99,5

5. Itseorganisoituvat kartat (SOM)



Kuva 32. Eri riskiluokkien ruutujen sijoittuminen kartalla a) toimintavalmiusohjeen mukaan, b) SOM:n kiireellisten onnettomuuksien perusteella tehdyn ennusteen mukaan ja c) toteutuneiden kiireellisten onnettomuuksien mukaan määritettynä. Riskiluokka I (punainen), riskiluokka II (keltainen), riskiluokka III (vihreä), riskiluokka IV (sininen).

5.5 Johtopäätökset

Luokittelutyökaluna yleensä käytettäviä itseorganisoituvia karttoja (SOM) sovellettiin onnettomuusmäärien ennustamiseen riskiruutuaineiston pohjalta. Sovellustapa on varsin poikkeava perinteisiin SOM-sovelluksiin verrattuna. Näiden tarkastelujen perusteella voidaan päätellä, että SOM pystyy jossain määrin tunnistamaan kartalta riskin kannalta toisistaan poikkeavat alueet ja antamaan onnettomuusmäärillä mitattuna useimmiten oikean suuntaisia tuloksia. Onnettomuuksien määrät arvioitiin hieman alakanttiin, mutta ottaen huomioon vuosien välisen vaihtelun, ero on hyväksyttävällä tasolla.

SOM ei kuitenkaan sovellu absoluuttisten onnettomuuksien ennustamiseen, jos etenkin suurten onnettomuusmäärien tarkka ennustaminen on tärkeää. Sekä rakennuspalojen että kiireellisten onnettomuuksien osalta suuria onnettomuusmääriä sisältävien ruutujen osuus kaikista ruuduista aliarvioitiin kaikilla tarkastelualueilla. Keskisuurten onnettomuusmäärien osalta absoluuttiset ennusteet voivat olla hyväksyttäviä. Ongelmana on, ettei SOM tunnistanut juuri lainkaan I riskiluokkaan kuuluvia ruutuja. Ennustuksen epätarkkuus viittaa siihen, että käytettävissä olleet selittäjät eivät riittäneet ruutujen välisten erojen tunnistamiseen ja sen selittämiseen, miksi toisaalla tapahtuu paljon onnettomuuksia ja toisaalla ei.

SOM-mallin etuna on, että kaikki käytettävissä olevat selittäjät voidaan ottaa huomioon ilman, että laskentakustannukset huomattavasti kasvavat. Toisaalta SOM-malli pystyy antamaan ennusteita, vaikka joidenkin selittävien muuttujien arvoja ei tiedettäisi lainkaan. Tällöin muiden muuttujien suhteellinen merkitys ruudun luokittelussa korostuu ja tarkkuus luonnollisesti huononee.

5.6 Kehittämisehdotukset

SOM-sovelluksessa on paljon kehittämistä, mikäli sitä halutaan käyttää onnettomuuksien ennustamiseen. Tärkein tekijä ennustuskyvyn parantamisessa olisi eri onnettomuustyyppien tärkeimpien taustamuuttujien tunnistaminen ja niiden käyttäminen verkon rakentamisessa. Esimerkiksi rakennuspalossa sosiaaliset tekijät olisivat mielenkiintoinen lisä. Toisaalta käytetyt rakennustiedot ja asukasmäärät eivät ainakaan suoraan vaikuta tieliikenneonnettomuus- tai maastopalomääriin. Tämän takia merkittävimpiä onnettomuustyyppisiä pitäisi tarkastella erikseen, ja näiden kokemusten perusteella muokata malli vastaamaan kaikkia kiireellisiä onnettomuuksia.

5. Itseorganisoituvat kartat (SOM)

SOM ennustaa onnettomuuksien kokonaismäärät hieman alakanttiin, mikä johtuu syötteiden liittymisestä eri noodeihin ilman onnettomuustietoja. Tämän voisi estää ottamalla historiatiedot tapahtuneista onnettomuuksista mukaan myös ennustusvaiheessa.

Suuria onnettomuusmääriä voitaisiin mahdollisesti saada tunnistettua enemmän, mikäli keskiarvon sijasta käytettäisiin jotain muuta tilastollista suuretta ennusteen kuvaamiseen. Näissä tarkasteluissa ennustetut kokonaismäärät jäävät systemaattisesti toteutuneita määriä pienemmiksi, joten voisi olla hyödyllistä kokeilla esimerkiksi 75 % kvartiilin käyttämistä ennusteen määrittämisessä.

Vuosien välisten onnettomuusmäärien vaihtelun tasoittamiseksi kannattaisi ottaa mukaan entistä enemmän onnettomuushistoriaa. Lisäksi muuttujiin tulisi lisätä aikaa kuvaavia tai ajan mukana muuttuvia suureita.

SOM-mallin käyttöön liittyy joukko käyttäjän määriteltävissä olevia parametreja, esimerkkinä verkon koko, jonka vaikutusta tulosten tarkkuuteen ei tässä työssä ehditty systemaattisesti tutkia. Myös tapa, jolla ruudun luokittelusta johdettiin sitä vastaava onnettomuusmääräennuste, oli varsin ainutlaatuinen.

Muista neuroverkkosovelluksista kannattaisi kokeilla tavallista *feed-forward*-verkkoa. Se on SOM:ia yksinkertaisempi neuroverkko, jota voidaan pitää epälineaarisen regressiomallina. *Feed-forward*-verkoilla verkon koon ja funktiotyyppien valinnan sekä tulosten väliset yhteydet ovat usein helpommin ymmärrettävissä ja mitattavissa kuin SOM-mallilla.

6. Sosioekonomisen tiedon hyödyntäminen

6.1 Käytetty aineisto

Luvussa 5 kuvatun SOM-sovelluksen lisäksi tehtiin suppea kokeilu, jossa käytössä olevien muuttujien lisäksi hyödynnettiin sosioekonomista tietoa. Kokeilualueena oli Keski-Uudenmaan pelastuslaitoksen toimialue, jolta Experian Oy toimitti hankkeen käyttöön Mosaic-tietokannan tietoja.

Mosaic-aineistossa jokaiselle 250 m×250 m:n ruudulle on määritetty yhdeksän tekijää, jotka kuvaavat erilaisia asumiseen, koulutukseen, työllisyyteen ja varakkuuteen liittyviä ominaisuuksia. Experian Oy:n materiaalissa näitä tekijöitä kutsutaan *faktoreiksi*. Tarkastelualueella noin 30 % ruuduista oli vailla Mosaic-tekijöitä. Jokainen tekijä (faktori) on yhdistelmä monesta taustamuuttujasta, joten tekijöiden summaaminen suoraan ruutuja yhdistettäessä on vähintäänkin kyseenalaista. Tästä sekä alueen pienestä koosta johtuen tarkasteluissa päädyttiin käyttämään 250 m×250 m:n ruutukokoa. Tällöin ennustettavat onnettomuusmäärät ovat hyvin pieniä.

Mosaic-aineiston toimivuutta SOM:n kanssa testattiin kolmen eri muuttujaryhmän avulla. Ryhmät olivat

1. riskiruutuaineisto, onnettomuustiedot
2. riskiruutuaineisto, onnettomuustiedot, Mosaic-tekijät
3. onnettomuustiedot, Mosaic-tekijät.

Mosaic-aineiston toimivuutta tarkasteltiin myös ilman riskiruutuaineistoa, sillä molemmat aineistot sisältävät paljon samoja muuttujia. Näiden muuttujien sijoittuminen SOM-kartalle on esitetty liitteessä B.

6. Sosioekonomisen tiedon hyödyntäminen

6.2 Tulokset

Koska onnettomuusmäärät ruutua kohti olivat erittäin pieniä, tässä tarkastellaan ainoastaan kiireellisiä onnettomuuksia. Kuten luvussa 5 esitetyissä tarkasteluissa myös tässä ruudut on jaettu neljään ryhmään onnettomuusmäärien perusteella. Onnettomuusmäärien luokkarajat on esitetty taulukossa 4. Luokkarajojen valinnan vuoksi luokkaan III ei osu yhtään ruutua. Käytännössä luokkaan IV kuuluvat ruudut, joissa ei ole tapahtunut onnettomuuksia, luokkaan II ruudut, joissa on tapahtunut yksi onnettomuus, ja luokkaan I ruudut, joissa on tapahtunut enemmän kuin yksi onnettomuus.

Ruutujen jakautuminen näihin neljään luokkaan on esitetty taulukossa 31. Kuvassa 34 ruutujen sijoittuminen on esitetty kartalla. Kuvan 34 kartoissa luokkaan I sijoittuneet ruudut on merkitty punaisella, luokka II keltaisella ja luokat III ja IV sinisellä värillä. Ennustevirhe tilanteessa, jossa ruudut on jaettu kymmeneen luokkaan ennustetun onnettomuusmäärän mukaan, on esitetty kuvassa 33.

Tulokset osoittavat, että kaikilla muuttujaryhmillä suuria onnettomuusmääriä ennustettiin suurin piirtein oikein. Kun riskiruutuaineistoon ja onnettomuustietoihin (muuttujaryhmä 1) lisättiin Mosaic-tekijät (muuttujaryhmä 2), I luokan ruutujen määrä kasvoi ennusteessa. Määrä lisääntyi edelleen, kun muuttujina käytettiin vain onnettomuustietoja ja Mosaic-tekijöitä (muuttujaryhmä 3). Luokan II ruutujen määrä osui lähimmäs vuonna 2008 toteutunutta muuttujaryhmillä 2. Luokkiin III ja IV sijoittuvien ruutujen määrät ennustettiin lähimmäksi toteutunutta muuttujaryhmillä 3. Nämä tulokset näkyvät myös kartalla. Ennustevirheen mediaani pysyy nollassa kaikilla muuttujaryhmillä muutamaa luokkaa lukuun ottamatta.

Kiireellisten onnettomuuksien ennustetut kokonaismäärät on esitetty taulukossa 32. Tämän perusteella lähimmäksi toteutuneita ennustavat muuttujaryhmät 1 ja 2; selvästi kauemmas jää ryhmän 3 ennuste. Tämä saattaa johtua siitä, että onnettomuusmäärät kattavat ryhmän 3 tapauksessa yli puolet muuttujista. Kun nämä tiedot poistetaan testivaiheessa, verkon on vaikeaa löytää samankaltaisia ruutuja tällä aineistomäärällä.

Liitteissä on esitetty muuttujien klusteroituminen kaikilla muuttujaryhmillä. Näistä kuvista huomataan, että Mosaic-tekijät eivät klusteroidu samalla tavalla kuin muut muuttujat. Tähän saattaa vaikuttaa aineiston pieni koko.

6. Sosioekonomisen tiedon hyödyntäminen

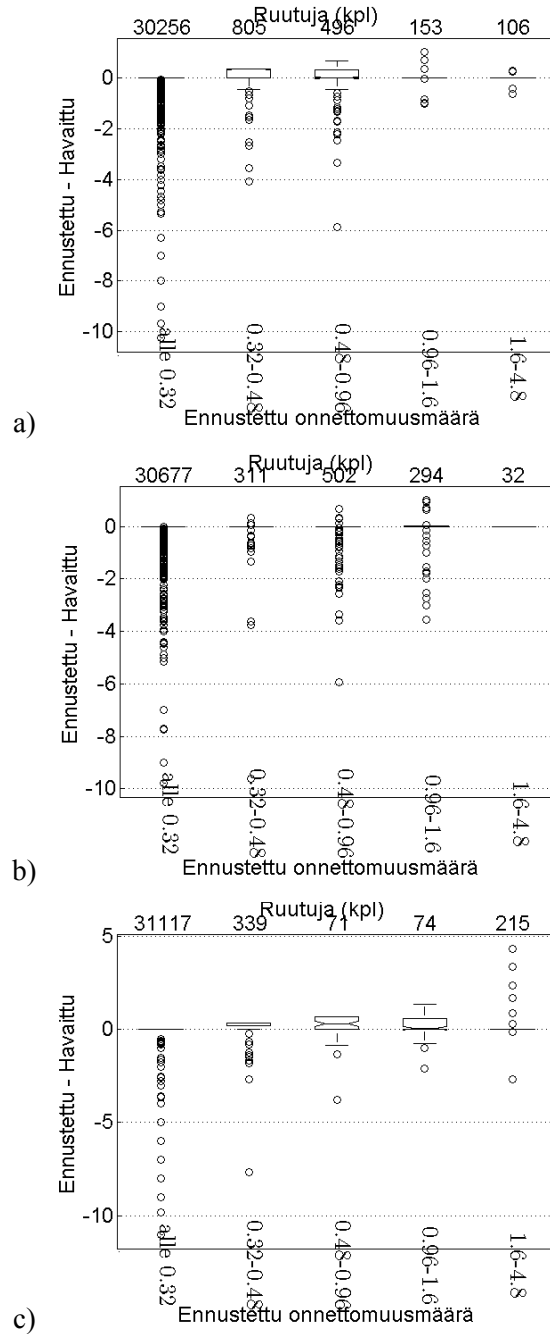
Taulukko 31. Ruutujen jakautuminen riskiluokkiin.

Luokka	Toteutunut v. 2008	Ryhmä 1	Ryhmä 2	Ryhmä 3
I	1 %	0,7 %	0,8 %	0,9 %
II	3 %	4,3 %	3,0 %	1,3 %
III	0 %	5,8 %	5,8 %	0,4 %
IV	95 %	89,2 %	90,4 %	97,4 %

Taulukko 32. Kiireellisten onnettomuuksien kokonaismäärät eri muuttujaryhmillä. Vuosien 2005-2007 osalta on ilmoitettu kolmen vuoden keskiarvo.

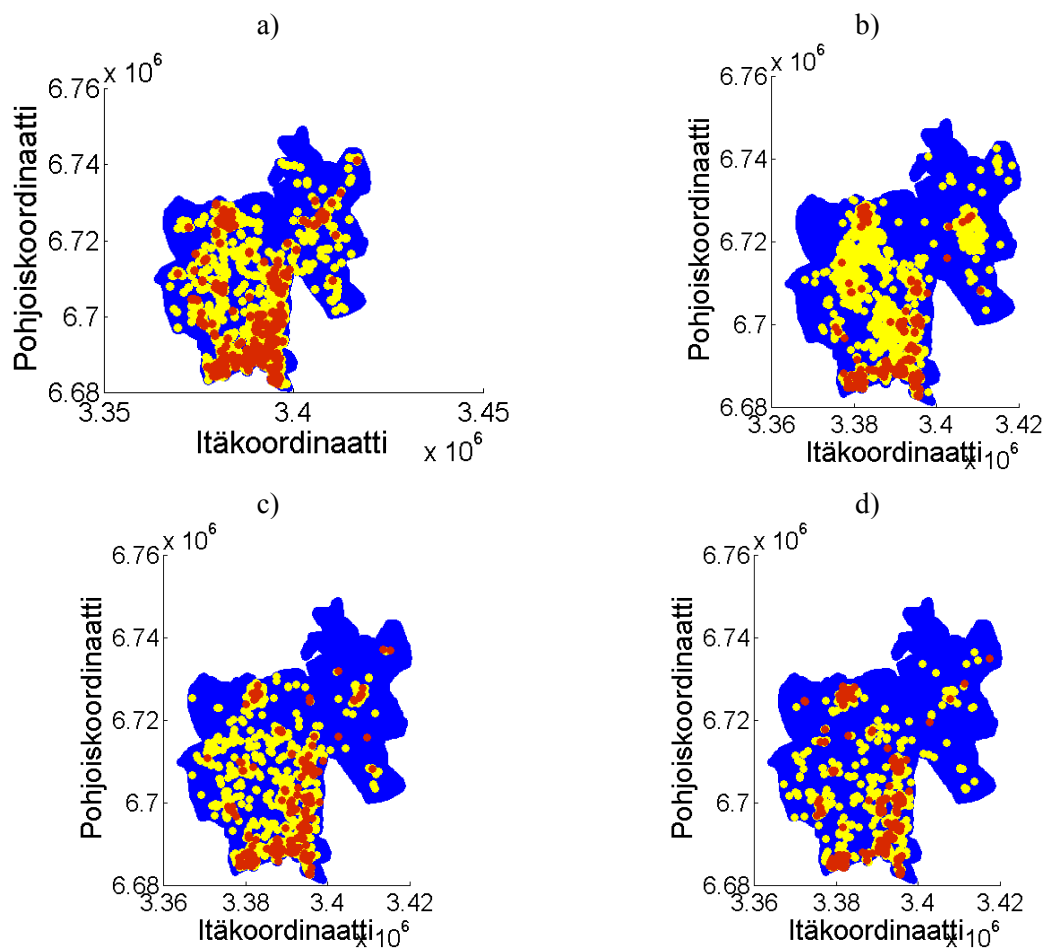
Luokka	Ennustettu	Suhdetoteutuneisiin 2005–2007 (1 813 kpl)	Suhdetoteutuneisiin 2008 (2 149 kpl)
Ryhmä 1	1 277	0,7	0,6
Ryhmä 2	1 223	0,7	0,6
Ryhmä 3	838	0,5	0,4

6. Sosioekonomisen tiedon hyödyntäminen



Kuva 33. Kiireellisten onnettomuuksien ennustevirhe Keski-Uudenmaan pelastustoimen alueella. a) Riskiruutuaineisto ja onnettomuustiedot. b) Riskiruutuaineisto, onnettomuustiedot ja Mosaic. c) Onnettomuustiedot ja Mosaic.

6. Sosioekonomisen tiedon hyödyntäminen



Kuva 34. Suurempien onnettomuusmäärien sijoittuminen kartalla, kun ruudut on jaettu luokkiin onnettomuusmäärän perusteella. Käytetyt luokkarajat on esitetty taulukossa 4. Luokan I ruutuja on merkitty punaisella, luokan II ruutuja keltaisella ja luokkien III ja IV ruutuja sinisellä värillä. a) Toteutuneet onnettomuudet. b) Ennuste muuttujilla 1 (riskiruutuaineisto ja onnettomuustiedot). c) Ennuste muuttujilla 2 (riskiruutuaineisto, onnettomuustiedot ja Mosaic). d) Ennuste muuttujilla 3 (onnettomuustiedot ja Mosaic).

6.3 Johtopäätökset

Mosaic-tekijöitä käytettäessä tulos ei parane dramaattisesti. Pitkälle meneviä johtopäätöksiä ei kuitenkaan tämän analyysin perusteella voida tehdä seuraavista syistä:

6. Sosioekonomisen tiedon hyödyntäminen

- Kukin Mosaic-tekijä muodostuu usean muuttujan yhteisvaikutuksesta, jolloin yksittäisen selittäjän vaikutusta on vaikea havaita. Myös päällekkäiset selittäjät riskiruutu- ja Mosaic-aineistossa voitaisiin poistaa, mikäli käytettäisiin yksittäisiä selittäjiä Mosaic-tekijöiden sijaan.
- Keski-Uudenmaan alue on liian pieni. Se ei sisällä riittävää määrää ruutuja, jotta tarkastelu olisi järkevä.
- $250\text{ m} \times 250\text{ m}$ on liian pieni resoluutio. Ruutuja ei voida yhdistää, sillä useamman ruudun yhdistelmälle ei käytettävissä olleilla tiedoilla voitu yksiselitteisesti määrittää uusia Mosaic-tekijöitä. Pienissä ruuduissa satunnaisvaihtelulla on suurempi merkitys onnettomuusmäärissä.
- 30 % alueen ruuduista oli vailla Mosaic-tekijää.

6.4 Kehittämisehdotukset

Sosioekonomisilla selittäjillä on varmasti suuri merkitys ainakin rakennuspaloja ennustettaessa. SOM pystyisi hyödyntämään tätä tietoa parhaiten yksittäisten selittäjien muodossa. Lisäksi Mosaic-aineistoa pitäisi kokeilla myös muilla maaleilla ja suuremmilla alueilla.

7. Mallien välisten erojen vertailu

7.1 Kiireellisten onnettomuuksien ennustevirhe alueella 1

Kuva 35 esittää alueelle 1 (taulukko 1) eri malleilla määritetyt kiireellisten onnettomuuksien ennustevirheiden (ennustettu–havaittu) keskiarvot kymmenessä luokassa (taulukko 3). Mukana ovat myös toteutuneet onnettomuudet sekä toimintavalmiusohjeen määrät laskettuna kaavasta (1).

Toteutuneiden ”virhe” kuvaa muutosta, joka on tapahtunut vuosista 2005–2007 vuoteen 2008. Tässä myös SOM:n virhe on laskettu pelkän keskiarvon avulla, jättämällä keskihajonnat huomiotta. Suurimpaan luokkaan (onnettomuusmäärä > 64) osui pisteitä vain käytettäessä spatiaalisia regressiomalleja. Yksikään SOM:n ennusteista ei ylittänyt rajaa 32.

Kuvasta 35a nähdään, että kaikilla malleilla keskiarvo on nollan tuntumassa luokkaan 4,8–8 saakka. Tämän jälkeen alkaa esiintyä pientä hajontaa. Spatiaalisen regressiomallin virhe kasvaa onnettomuusmäärien kasvaessa, samoin kuin toimintavalmiusohjeen perustelumuiustiossa esitettyjen molempien mallien (kaava (1) mukaan lasketun ennustevirhe. Nämä mallit siis ennustavat onnettomuuksien määrät liian suuriksi. SOM ja toteutuneiden onnettomuuksien virheiden keskiarvot sen sijaan laskevat nollan alapuolelle onnettomuusmäärien kasvaessa. Yksinkertaisen regressiomallin virhe pysyy hyvin lähellä nollaa lukuun ottamatta suurinta luokkaa, jossa keskiarvon virhe tipahtaa alle –20:een. Tähän luokkaan kuului tosin vain kaksi ruutua.

Kuva 35b esittää kuvan 35a sisällön suhteellisen virheen muodossa. Tässä ennustevirheen keskiarvo on skaalattu luokan päätepisteiden välin keskiarvolla. Tällöin virheen systemaattisuus tulee paremmin ilmi. Kuvan perusteella onnettomuusmäärät ennustettiin yleisesti ottaen parhaiten luokassa 1,6–4.8. Pienillä onnettomuusmäärillä spatiaalisten mallien virheet ovat suuria – tarkastelluista

7. Mallien välisten erojen vertailu

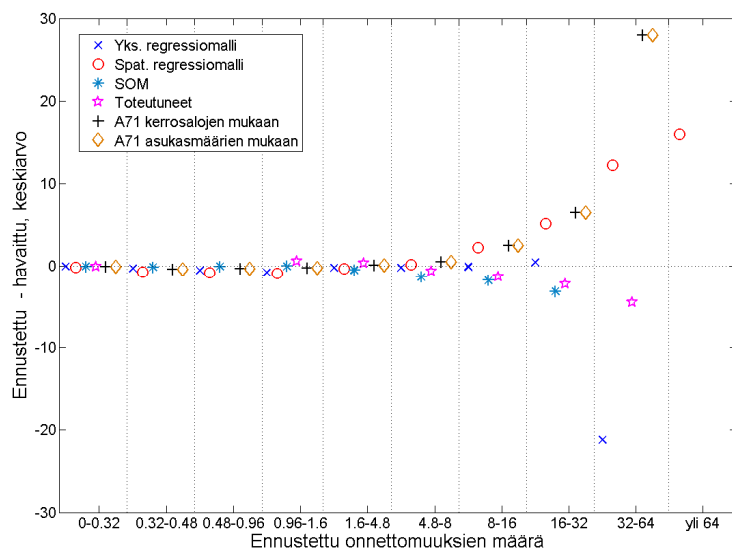
menetelmistä suurimpia. Suuremmissa luokissa virheen suuruus pienenee ja pysyy melko tasaisena, noin 20 %:n molemmin puolin. Yksinkertaisen regressiomallin suhteellinen virhe oli samaan tapaan melko suuri ennustettaessa pieniä onnettomuusmääriä (alle 1,6 onnettomuutta/ruutu) mutta pieni ennustettaessa suuria onnettomuusmääriä (pois lukien ylin luokka). Suuret negatiiviset ennustevirheet vaikuttavat huomattavasti ennustevirheen keskiarvoon erityisesti alimmissa luokissa, minkä vuoksi keskiarvo ei ole paras mahdollinen tunnusluku kuvaamaan keskimääräistä ennustevirhettä. Tämä näkyy erityisesti tarkasteltaessa suhteellista ennustevirhettä, joka antaa liian pessimistisen kuvan mallin ennustekyvystä alimmissa luokissa. SOM:n ennustevirheen keskiarvo on samaa suuruusluokkaa tai pienempi kuin muilla malleilla pienissä luokissa. Suuremmissa luokissa ennustevirhe on toteutuneiden luokkien virheen tasoa.

Kuvasta 35b nähdään, että toimintavalmiusohjeen kerrosalan mukaan lasketun ennusteen virhe vaihtelee 100 % liian pienestä arviosta pienillä onnettomuusmäärillä aina 50 % liian suureen arvioon suurilla onnettomuusmäärillä. Asukasluvun mukaan lasketun ennusteen suhteellinen virhe näyttää hyvin samanlaiselta, paitsi että virheet ovat vielä suuremmat.

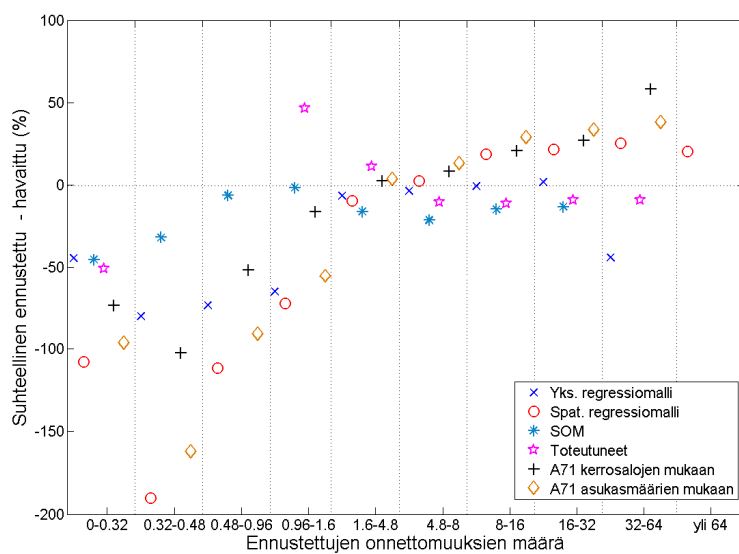
Toteutuneiden ennustevirheestä voidaan todeta, että vuosista 2005–2007 vuoteen 2008 on tapahtunut eniten muutosta pienissä onnettomuusmäärissä. Suurissa onnettomuusmäärissä eroa on vähemmän, mutta sielläkin virhe on noin – 10 %. Toteutuneiden onnettomuuksien suhteellinen virhe vaihtelee suuresti nollan molemmin puolin pienillä onnettomuusmäärillä.

Toteutuneiden onnettomuuksien mukaisia ”virheitä” voidaan käyttää indikaattorina tilastoaineiston sisäisestä vaihtelusta. Koska mikään käytetyistä malleista ei ottanut huomioon riskiruutujen ominaisuuksien muuttumista ajan myötä, ei mallien virheitä voida pitää tilastollisesti kovin merkittävinä, jos ne ovat pieniä kuin toteutuneen tilanteen ”virhe”. Tästä näkökulmasta kaikki käytetyt mallit ovat toimineet hyvin, lukuun ottamatta spatiaalista regressiomallia pienimpien onnettomuusmäärien kohdalla. Toimintavalmiusohjeen ennusteet sen sijaan voivat sisältää merkittäviä virheitä.

7. Mallien välisten erojen vertailu



a)



b)

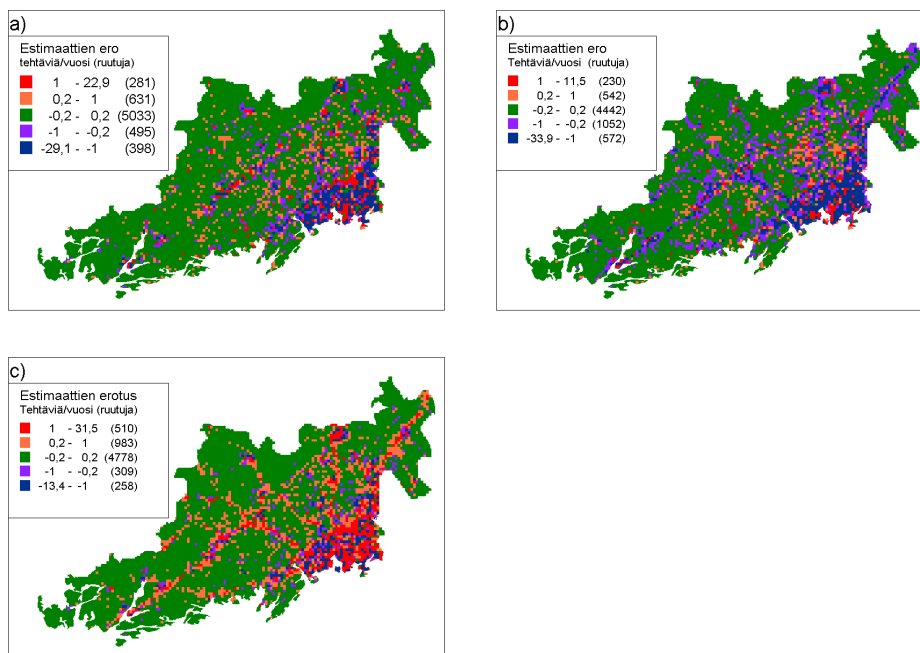
Kuva 35. a) Kiireellisten onnettomuuksien ennustevirheen (ennustettu–havaittu) keskiarvot eri malleilla alueelle 1. b) Ennustettu–havaittu, suhteellinen virhe.

7. Mallien välisten erojen vertailu

7.2 Kolmen eri mallitustavan tuottamien kiireellisten onnettomuuksien ennusteiden erot

Kuvassa 36 on esitetty kolmen mallitustavan keskinäiset erot ennustettaessa kiireellisten onnettomuuksien määrää. Systemaattinen ero itseorganisoituvien karttojen ja yksinkertaisten regressiomallien välillä on pieni. SOM ennusti 240 kiireellistä tehtävää enemmän kuin yksinkertaiset regressiomallit. Spatiaalinen regressiomalli sen sijaan ennusti huomattavasti enemmän kiireellisiä tehtäviä – 2 065 tehtävää enemmän kuin yksinkertainen regressiomalli. Ainakin osittain eron selittää spatiaaliseen malliin sisältyvä trenditermi, joka kuitenkin yliarvioi selvästi toteutuneiden tehtävien määrän kasvun.

Alueellisesti kiireellisten tehtävien ennusteet poikkesivat voimakkaimmin spatiaalisen regressiomallin ja yksinkertaisen regressiomallin välillä. Spatiaalinen malli ”löysi” pääliikenneväylät. Itseorganisoituvien karttojen ja yksinkertaisten regressiomallien merkittävimmät erot löytyivät taajaan asutuilta alueilta Helsingistä, Vantaalta ja Espoosta. Eroista on kuitenkin vaikea löytää mitään erityistä rakennetta.



Kuva 36. Kolmen eri mallitustavan tuottamien kiireellisten tehtävien ennusteiden erot a) itseorganisoituvat kartat – yksinkertaiset regressiomallit, b) itseorganisoituvat kartat – spatiaalinen regressiomalli ja c) spatiaalinen regressiomalli – yksinkertainen regressiomalli.

8. Yhteenveto ja johtopäätökset

Tutkimuksessa pyrittiin kolmea eri lähestymistapaa käyttäen luomaan menetelmiä, joiden avulla onnettomuustiheys pystyttäisiin ennustamaan entistä luotettavammin riskiruutuaineiston muuttujia apuna käyttäen. Tämän lisäksi tehtiin suppea yhteenveto vuosina 2005–2008 toteutuneista onnettomuustiheyksistä eri riskialueilla.

Lähestymistavoista lupaavimmaksi osoittautuivat yksinkertaiset regressiomallit, joista tuotettiin yksinkertaisia työväliteitä riskinarvioinnissa käytettäväksi. Mallit luotiin erikseen rakennuspaloille ja kiireellisille onnettomuuksille.

Rakennuspalomalli kehitettiin koko maalle ruutukoossa $1 \text{ km} \times 1 \text{ km}$. Malli selitti noin 50 % rakennuspalomäärän vaihtelusta, mitä voi pitää kohtalaisen hyvänä tuloksena, kun otetaan huomioon mallin yksinkertaisuus. Selittäjinä mallissa ovat asukasluku, kerrosala sekä niiden yhteisvaikutus. Mallin toimivuustarkasteluissa havaittiin, että se ennustaa alhaisia rakennuspalomääriä melko hyvin, keskisuuria hieman yläkanttiin ja suuria selvästi alakanttiin. Mallin kyvyttömyys huomioda rakennuspalomäärän nouseva trendi näkyy nimenomaan ennustettaessa korkeampia määriä. Mallia voi soveltaa ennustamiseen, kunhan muistaa sen puutteet.

Kiireellisille onnettomuuksille kehitetyssä mallissa ruutukoko oli $1 \text{ km} \times 1 \text{ km}$ ja selittäjänä pelkästään asukasluku. Koko Suomen yhteinen malli selitti noin 60 % kiireellisten tehtävien määrän vaihtelusta, mitä voi pitää kohtalaisen hyvänä tuloksena, kun otetaan huomioon, että malli on hyvin yksinkertainen. Toimivuustarkasteluissa havaittiin, että malli ennusti alhaisia onnettomuusmääriä melko hyvin ja keskisuuria ja korkeita määriä vain hieman yläkanttiin. Tämä ei kuitenkaan ollut mallin ansiota, vaan kiireellisten tehtävien kasvava trendi kompensoi mallin kyvyttömyyttä ennustaa suuria onnettomuusmääriä. Aluekohtaiset erot mallin sopivuudessa ja ennustekyvyyssä olivat suuria. Erityisesti Pohjois-

8. Yhteenveto ja johtopäätökset

Suomea tarkasteltaessa malli ei toiminut. Mallia olisikin selvästi parannettava ennen sen laajempaa soveltamista.

Yksinkertaisten regressiomallien rinnalla testattiin myös spatiaalisten mallien kykyä ennustaa onnettomuusmääriä. Näissä malleissa selittäjinä olivat asukaslu-ku ja kerrosala. Mallien soveltaminen onnistui vain pienissä aineistoissa (15 000 ruutua), joten vertailu muihin malleihin jäi vaillinaiseksi. Spatiaalisten mallien vahvuudet tulevat esille tilanteessa, jossa mukana on joitakin sellaisia muuttujia, joiden vaikutus näkyy onnettomuusmäärissä mutta jotka eivät ole mukana mallin selittäjinä. Spatiaaliset mallit ennustivat toteutuneita tehtävämääriä lähes yhtä hyvin ilman kerrosalaa ja asukasmäärää kuin niiden kanssa. Laskennallisesti ne ovat kuitenkin joustavia hyödyntämään selittäjiä esimerkiksi ennustettaessa tulevaisuuden tehtävämääriä, jos ennusteita selittäjille on saatavissa. Spatiaaliset mallit jakavat havaittua tehtävämäärän vaihtelua selittäjien avulla selitettävään osaan, alueellisesti autokorreloituneeseen osaan ja yksittäisten ruutujen poikkeamiseen. Jako komponentteihin ei ole yksiselitteisesti tulkittavissa, mutta sitä voidaan käyttää avuksi päätellessä, mitä selittäjiä mallista vielä ehkä puuttuu. Spatiaaliset mallit ovat joustavia mutta laskennallisesti hankalia, ja niiden käytäntöön saattaminen vaatii vielä jatkokehittämistä.

Kolmannessa lähestymistavassa sovellettiin itseorganisoituvia karttoja (SOM). Lähestymistavan vahvuutena oli se, että mallin muuttujina voitiin käyttää kaikkia riskiruutuaineiston sisältämiä muuttujia toteutuneiden onnettomuusmäärien lisäksi. SOM:n tuottamien onnettomuusmääräennusteiden keskihajonta osoittautui kuitenkin joissakin tapauksissa melko suureksi, eikä absoluuttisia lukumääriä pystytty ennustamaan riittävällä tarkkuudella. SOM ennusti liian vähän suuria onnettomuusmääriä sisältäviä ruutuja sekä rakennuspalojen että kaikkien kiireellisten onnettomuuksien osalta. Menetelmän käyttö vaatii käyttäjältä jonkin verran perehtymistä, jotta tuloksia osataan tulkita oikein. Jos lähestymistapaa halutaan käyttää riskitasojen arviointiin, pitäisi SOM:n suorittaman ryhmittelyn ja onnettomuusmäärien välistä yhteyttä vielä tarkastella.

Itseorganisoituvien karttojen osalta testattiin vielä, paranisiko ennustustulos, jos muuttujiksi lisätään sosioekonomista tietoa. Kokeilu tehtiin Keski-Uudenmaan pelastuslaitoksen alueelle 250 m × 250 m:n ruutukoossa. Kokeilun tuloksena havaittiin, että sosioekonominen aineisto saattaisi parantaa ennustustulosta, vaikkakaan tällä aineistolla havaittu parannus ei ollut merkittävä. Kovin pitkälle meneviä johtopäätöksiä tämän analyysin perusteella ei kuitenkaan voida tehdä, sillä kokeilualue oli tämän tyyppiseen tarkasteluun liian pieni. Lisäksi käytössä olleet sosioekonomiseen aineistoon pohjautuvat selittäjät muodostuivat

8. Yhteenveto ja johtopäätökset

usean muuttujan yhteisvaikutuksesta, mikä tuotti ongelmia analyyseihin. Hedelmällisintä lähestymistavan kannalta olisi ollut, jos käyttöön olisi saatu alkuperäistä tietoa näistä yksittäisistä muuttujista, joiden pohjalta selittäjät oli määritetty. Sosioekonomisen aineiston toimivuutta tulisi selvittää vielä tarkemmin ja yksityiskohtaisemmin, ennen kuin voidaan luotettavasti määrittää aineiston vaikutus mallien tuottamiin ennusteisiin.

Mallien toimivuutta arvioitiin vertailemalla niiden antamia ennusteita vuonna 2008 toteutuneisiin onnettomuusmääriin, toimintavalmiusohjeessa esitettyjen mallien ennustuskykyyn sekä toisiinsa. Tarkasteltaessa toteutuneiden onnettomuusmäärien muutosta vuosista 2005–2007 vuoteen 2008, suurimmat muutokset olivat tapahtuneet ruuduissa, joiden onnettomuusmäärä oli pieni. Suurilla onnettomuusmäärillä eroa oli vähemmän, mutta niissäkin muutos oli noin –10 %.

Toteutuneiden onnettomuusmäärien muutosta voidaan käyttää tilastoaineiston sisäisen vaihtelun indikaattorina. Koska mikään käytetyistä malleista ei voinut ottaa huomioon riskiruutujen ominaisuuksien muuttumista ajan myötä, ei mallien virheitä voida pitää tilastollisesti kovin merkittävinä, jos ne ovat suuruusluokaltaan pienempiä kuin ero toteutuneessa tilanteessa vuosien 2005–2007 ja vuoden 2008 välillä. Tästä näkökulmasta katsottuna kaikki käytetyt mallit toimivat pääpiirteissään hyvin. Kullakin mallilla oli kuitenkin omat heikkoutensa, eikä mikään niistä pystynyt ennustamaan absoluuttisia lukumääriä riittävän tarkasti kaikilla eri alueilla ja eri onnettomuusmäärillä. Vertailu osoitti, että myös toimintavalmiusohjeessa esitetyillä malleilla määritetyt ennusteet ovat epävarmoja.

Lähdeluettelo

- Besag J., York J. & Mollie A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 1991; 43: 1–21.
- Best N., Richardson S. & Thomson A. 2005 A comparison of Bayesian spatial models for diseasemapping. *Statistical Methods in Medical Research*. 2005; 14: 35–59.
- Hastie, T.J. & Tibshirani, R.J. 1990. *R: Generalized additive models*. Chapman and Hall, London.
- Imperial college and MRC, UK. Winbugs 1.4. <http://www.mrc-bsu.cam.ac.uk/bugs/>.
- Kohonen, T. 1995. *Self-Organizing Maps*. Springer Series in Information Sciences 30. Springer, Heidelberg.
- Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D. 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10: 325–337.
- Pelastusopisto 2009. Pelastustoimen resurssi- ja onnettomuustilasto PRONTO. <http://www.prontonet.fi>.
- R Development Core Team 2009. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Sisäasiainministeriö 2003. Toimintavalmiusohje A:71. Sisäasiainministeriön pelastusosaston julkaisusarja A. Dnro SM-2002-00018/Tu-35. 12 s.
- Tillander, K. & Keski-Rahkonen, O. 2000. Palokunnan saatavuuden merkitys rakennuksen paloriskitarkastelussa. Espoo, VTT. 213 s. + liitt. 55 s. VTT Tiedotteita – Meddelanden – Research Notes; 2013. ISBN 951-38-5634-8; 951-38-5635-6. <http://www.vtt.fi/inf/pdf/tiedotteet/2000/T2013.pdf>.

- Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankangas, J. 2000. SOM Toolbox for Malab 5, Report A57. Libella Oy.
- Venables, W. N. & Ripley, B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York.
- Vuong, Q.H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307–333.
- Wood, S.N. 2008. Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society (B)* 70(3): 495–518.
- Zeileis A., Kleiber, C. & Jackman, S. 2008. Regression models for count data in R. *Journal of Statistical Software* 27(8): 1–25.

Liite A: Spatiaalisissa malleissa käytetty Winbugs-koodi

Malli ilman asukaslukua ja kerrosalaa selittäjinä

```
model {
  for( i in 1 : nruudut) {
    gamma[i] ~ dnorm(0, tau1)
    for( j in 1:3) {
      N[i,j] ~ dpois(mu[i,j])
      log(mu[i,j]) <- alpha+lambda[i]+gamma[i]+beta4*(j-2)
    }
  }
  #naapureiden painot
  for(k in 1 : nlength) {
    w[k] <- 1
  }
  for(l in 1:nruudut){
    #estimaatit vuodelle 2008
    intens[l] <- exp(alpha+lambda[l]+gamma[l]+beta4*2)
    #yksittäisen ruudun erottuminen perustasosta
    pruuu[l]<- step(lambda[l]+gamma[l])
  }
  # Priorit
  lambda0[1 : nruudut] ~ car.normal(map[], w[], off[], tau2)
  alpha~dflat()
  beta1 ~ dnorm(0, 0.0001)
  tau1 <- 1/(sigma1*sigma1)
  sigma1 ~ dunif(0, 1000)
  tau2 <- 1/(sigma2*sigma2)
  sigma2 ~ dunif(0, 1000)
}
```

Asukasmäärän ja kerrosalan sisältävä malli

```
model {
  for( i in 1 : nruudut)
  {
    gamma[i] ~ dnorm(0, tau1)
    for( j in 1:3){
      N[i,j] ~ dpois(mu[i,j])
      log(mu[i,j]) <-
      al-
      pha+beta1*log(asukas[i]+1)+beta2*log(krl[i]+1)+beta3*log(krl[i]+1)*log(a
      sukas[i]+1) +lambda[i]+gamma[i]+beta4*(j-2)
    }
    #naapureiden painot
    for(k in 1 : nlength)
    {
      w[k] <- 1
    }

    for(l in 1:nruudut)
    {
      # estimaatit vuodelle 2008
      intens[l] <- exp(alpha+beta1*log(asukas[i]+1)+beta2*log(krl[i]+1)
      +beta3*log(krl[i]+1)*log(asukas[i]+1)+lambda[l]+gamma[l]+beta4*2)
      #yksittäisen ruudun erottuminen perustasosta
      pruuu[l]<- step(lambda[l]+gamma[l])

    }
  }
  #priorit
  lambda[1 : nruudut] ~ car.normal(map[], w[], off[], tau2)
  alpha~dflat()
  beta1 ~ dnorm(0, 0.0001)
  beta2 ~ dnorm(0, 0.0001)
  beta3 ~ dnorm(0, 0.0001)
  beta4 ~ dnorm(0, 0.0001)
  tau1 <- 1/(sigma1*sigma1)
  sigma1 ~ dunif(0, 1000)
  tau2 <- 1/(sigma2*sigma2)
  sigma2 ~ dunif(0, 1000)
}
```

Liite A: Spatiaalisissa malleissa käytetty Winbugs-koodi

Termi	Kuvaus
alpha	tasokerroin
asukas	ruudun asukasmäärä
beta1	kerrosalan kerroin
beta2	asukasmäärän kerroin
beta3	asukasmäärän ja kerrosalan kerroin
beta4	vuosikerroin
gamma	ruututermi
krl	ruudun kerrosala
lambda	alueellinen termi
mu	ruudun intensiteetti
N	havaittu määrä
nlength	naapuruuksien lukumäärä
nruudur	ruutujen lukumäärä
tau1	gamman varianssin käänteisluku
tau2	lambdan naapurustovarianssin käänteisluku

Liite B: Muuttujien järjestäytyminen itse-organisoidulla kartalla

SOM järjestää selittäjät kartalla siten, että kaikkien muuttujien suhteen samantyyppiset ruudut ovat lähekkäin. Usein voidaan tehdä joitain johtopäätöksiä selittäjien korreloitumisesta sen perusteella, mihin ne sijoittuvat SOM-kartalla. Kuvassa B1 näkyvät kaikki alueen 1 rakennuspalojen selittäjät SOM:n järjestäminä. Muuttujat ovat

- IKOORD ja PKOORD (ruudun vasemman alareunan itä- ja pohjoiskoordinaatit)
- I0_14V, I15_64V, I65_V (asukkaiden määrät ikäryhmissä 0–14, 15–64 ja yli 64-vuotiaat)
- A1 jne. (rakennusten lukumäärä käyttötarkoitukseluokassa A)
- A3 jne. (kerrosala m², luokassa A)
- TYOPAIK04 (työpaikkojen määrä 2004)
- 21 rakennuspalo (keskimääräinen rakennuspalojen määrä vuosina 2005–2007).

Rakennusten käyttötarkoitukseluokat ovat

- A = asuinrakennukset
- B = vapaa-ajan asuinrakennukset
- C = liikerakennukset
- D = toimistorakennukset
- E = liikenteen rakennukset
- F = hoitoalan rakennukset
- G = kokoontumisrakennukset
- H = opetusrakennukset
- J = teollisuusrakennukset
- K = varastorakennukset

Liite B: Muuttujien järjestäytyminen itseorganisoituvalla kartalla

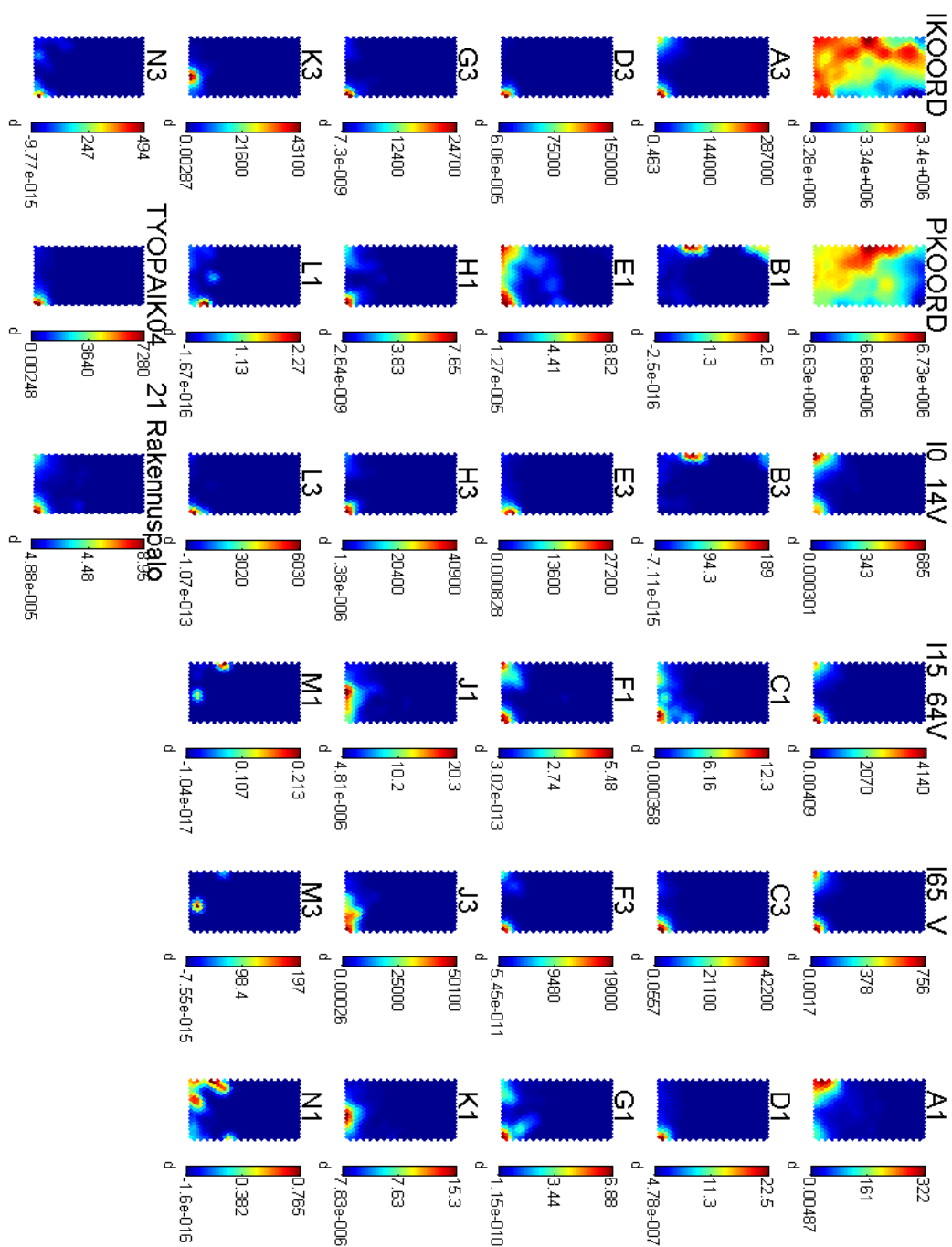
- L = palo- ja pelastustustoimen rakennukset
- M = maatalousrakennukset
- N = muut rakennukset.

Suurimmat rakennuspalomäärät näyttävät sijoittuneen kartan oikeaan alanurkkaan. Täsmälleen samaan kohdassa asuu paljon ihmisiä kaikista ikäryhmistä (0–14-vuotiaita hieman vähemmän kuin muita ikäryhmiä). Samalla lailla rakennuspalojen kanssa korreloituneita selittäjiä ovat A3, C1, C3, D1, D3, E1, E3, F1, F3, G1, G3, H1, H3, J3, L3, N3 sekä työpaikkojen lukumäärät. Lähes kaikki rakennustyypit siis lisäävät rakennuspaloja, mikä ei liene yllätys.

Eri alueilla selittäjäkartat näyttävät hieman erilaisilta, mutta yhteistä on se, että lähes kaikkien muuttujien suuret arvot kerääntyvät pienelle alueelle. Kiireelliset onnettomuudet sijoittuvat kartalle rakennuspalojen tapaan pienelle alueelle (eivät tietenkään aina samaan paikkaan rakennuspalojen kanssa).

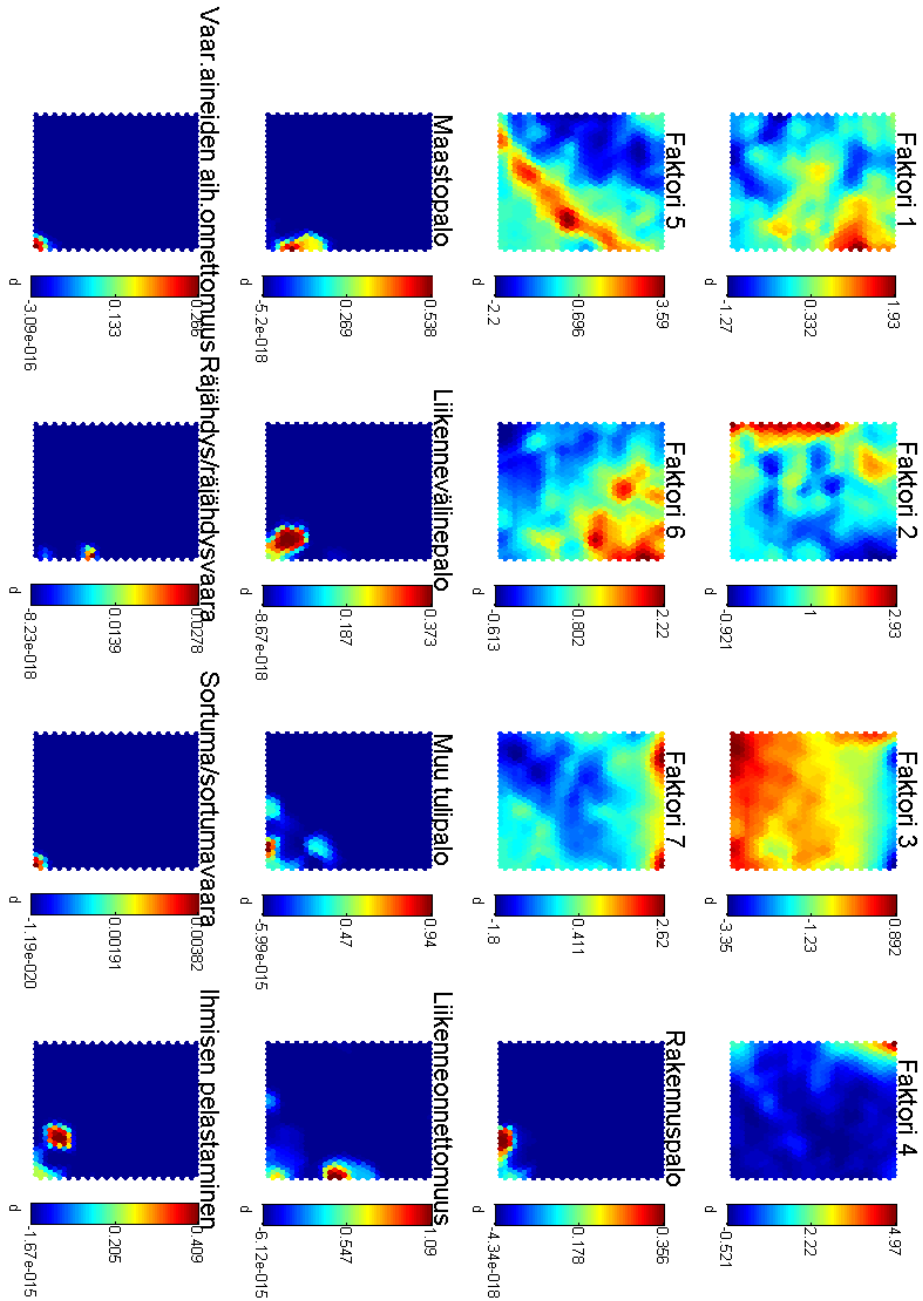
Ainoan poikkeuksen muodostavan Mosaic-tekijät, jotka levittäytyvät koko kartan alueelle. Kuvassa B2 näkyvät Mosaic-muuttujat ja kiireelliset onnettomuudet Keski-Uudellamaalla. Tämä luultavasti johtuu luvussa 6 esitetyistä ongelmista eli siitä, että tekijät ovat yhdistelmiä monesta selittäjästä ja että tarkasteltava alue oli liian pieni.

Liite B: Muuttujien järjestäytyminen itseorganisoituvalla kartalla



Kuva B1. Rakennuspaljo-SOM:n muuttujien sijoittuminen kartalla alueella 1.

Liite B: Muuttujien järjestäytyminen itseorganisoidulla kartalla



Kuva B2. Mosaic-faktoreiden sijoittuminen kartalla Keski-Uudellamaalla kun ennustettiin kiireellisiä onnettomuuksia.

Liite C: Yksinkertaisten regressiomallien estimoinnissa käytetty R-koodi

```
# Tarvittavat kirjastot.
library(foreign)
library(MASS)

# RAKENNUSPALOT
# Ennustemallin estimointi vuosien 2005-2007 datalla.
data.suomi.rakon<-
  + read.spss("C:/riski/data/analysis/rr1km_suomi_05_07_rakon.sav",
  + to.data.frame=T)
attach(data.suomi.rakon)
fm.nbin.suomi<-glm.nb(rakpalo~I(log(aslkmyht05+1)) + I(log(krsalayht+1))
  + I(log(aslkmyht05+1))*I(log(krsalayht+1)) + offset(logyears),
  + subset=(krsalayht>0))
summary(fm.nbin.suomi)

# Vuoden 2008 rakennuspalojen ennustaminen.
# Poistetaan ensin vuoden 2008 aineistosta ruudut, joiden kerrosalaa ei tiedetä.
data.suomi.rakon.2008<-
  + read.spss("C:/riski/data/analysis/rr1km_suomi_08_rakon.sav",
  + to.data.frame=T)
data.08.suomi<-
  + subset(data.suomi.rakon.2008,data.suomi.rakon.2008$krsalayht>0)
attach(data.08.suomi)
pred.08.suomi<-predict(fm.nbin.suomi,newdata=data.08.suomi,
  + type="response")
summary(pred.08.suomi)
```

Liite C: Yksinkertaisten regressiomallien estimoinnissa käytetty R-koodi

```
# KIIREELLISET TEHTÄVÄT
```

```
# Ennustemallin estimointi vuosien 2005-2007 datalla.
```

```
data.suomi<-read.spss("C:/riski/data/analysis/rr1km_suomi_05_07.sav",  
  + to.data.frame=T)
```

```
fm.nbin.suomi<-glm.nb(kiireteht~I(log(aslkmmyht05+1)) + offset(logyears))  
summary(fm.nbin.suomi)
```

```
# Vuoden 2008 kiireellisten tehtävien ennustaminen.
```

```
data.suomi.08<-read.spss("C:/riski/data/analysis/rr1km_suomi_08.sav",  
  + to.data.frame=T)
```

```
attach(data.suomi.08)
```

```
pred.08.suomi<-predict(fm.nbin.suomi,newdata=data.suomi.08,  
  + type="response")
```

```
summary(pred.08.suomi)
```



Julkaisun sarja, numero ja
raporttikoodi

VTT Tiedotteita 2530
VTT-TIED-2530

Tekijä(t) Kati Tillander, Anna Matala, Simo Hostikka, Pekka Tiittanen, Esa Kokki & Olli Taskinen		
Nimeke Pelastustoimen riskianalyysimallien kehittäminen		
Tiivistelmä Tutkimuksessa pyrittiin kolmea eri lähestymistapaa käyttäen luomaan menetelmiä, joiden avulla onnettomuustiheys pystyttäisiin ennustamaan entistä luotettavammin riskiruutuaineiston muuttujia apuna käyttäen. Tämän lisäksi tehtiin suppea yhteenveto vuosina 2005–2008 toteutuneista onnettomuustiheyksistä eri riskialueilla. Lähestymistavoista lupaavimmiksi osoittautuivat yksinkertaiset regressiomallit, joista tuotettiin helppokäyttöisiä työvälineitä riskinarviointiin. Mallit luotiin erikseen rakennuspaikoille ja kiireellisille onnettomuuksille. Rakennuspalomalli kehitettiin koko maalle ruutu-koossa 1 km × 1 km. Selittäjinä mallissa ovat asukasluku, kerrosala sekä niiden yhteisvaikutus. Mallin puutteellisuudet huomioon ottaen sitä voidaan käyttää rakennuspalomäärien ennustamiseen. Kiireellisille onnettomuuksille kehitetyssä mallissa ruutukoko oli 1 km × 1 km ja selittäjänä pelkästään asukasluku. Aluekohtaiset erot mallin sopivuudessa ja ennustekyvyyssä olivat suuria. Erityisesti pohjoisessa Suomessa malli ei toiminut. Mallia olisikin selvästi parannettava ennen sen laajempaa soveltamista. Yksinkertaisten regressiomallien rinnalla testattiin myös spatiaalisten mallien kykyä ennustaa onnettomuusmääriä. Mallit osoittautuivat joustaviksi mutta laskennallisesti hankaliksi, ja niiden käytäntöön saattaminen vaatii vielä jatkokehittämistä. Kolmannessa lähestymistavassa sovellettiin itseorganisoituvia karttoja (SOM). Lähestymistavan vahvuutena oli se, että mallin muuttujina voitiin käyttää kaikkia riskiruutuaineiston sisältämiä muuttujia toteutuneiden onnettomuusmäärien lisäksi. SOM:n tuottamisen onnettomuusmääräennusteiden keskihajonta osoittautui kuitenkin joissakin tapauksissa melko suureksi, eikä absoluuttisia lukumääriä pystytty ennustamaan riittäväällä tarkkuudella. Onnettomuusmäärien suuruus-luokkia voidaan kuitenkin käyttää riskien arviointiin.		
ISBN 978-951-38-7573-2 (URL: http://www.vtt.fi/publications/index.jsp)		
Avainnimeke ja ISSN VTT Tiedotteita – Research Notes 1455-0865 (URL: http://www.vtt.fi/publications/index.jsp)		Projektinumero 33445
Julkaisuaika Helmikuu 2010	Kieli Suomi, engl. tiiv.	Sivuja 117 s. + liitt. 9 s.
Projektin nimi Riskiruutu		Toimeksiantaja(t) Pelastusopisto, Helsingin pelastuslaitos
Avainsanat Accident frequency, building fires, rescue services, regression models, spatial regression, self-organizing maps		Julkaisija VTT PL 1000, 02044 VTT Puh. 020 722 4404 Faksi 020 722 4374

Tutkimuksessa pyrittiin kolmea eri lähestymistapaa käyttäen luomaan menetelmiä, joiden avulla onnettomuustiheys pystyttäisiin ennustamaan entistä luotettavammin riskiruutuaineiston muuttujia apuna käyttäen. Tämän lisäksi tehtiin suppea yhteenveto vuosina 2005–2008 toteutuneista onnettomuustiheyksistä eri riskialueilla.