

RAKENNUSTEKNISTEN, GEOLOGISTEN JA
ILMASTOLLISTEN TEKIJÖIDEN VAIKUTUS
ASUNTOJEN SISÄILMAN RADONPITOISUUTEEN

Topi Kinnunen

Tilastotieteen pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
kesä 2008

Jyväskylän yliopisto

Matematiikan ja tilastotieteen laitos

KINNUNEN, TOPI: Rakennusteknisten, geologisten ja ilmastollisten tekijöiden vaikutus asuntojen sisäilman radonpitoisuuteen.

Pro gradu -tutkielma, 53 s., 1 liite

Tilastotiede

Kesäkuu 2008

Tiivistelmä

Asuntojen sisäilman radonpitoisuuden tutkiminen kiinnostaa radonin haitallisten terveysvaikutusten johdosta. Erityisesti Suomessa on paljon korkeiden radonpitoisuuksien asuntoja maamme rakennustekniikan, geologian ja ilmaston seurauksena. Tässä tutkimuksessa mallinnetaan Kanta- ja Päijät-Hämeessä sijaitsevien matalaperustaisten omakoti- ja rivitalojen sisäilman radonpitoisuutta käytössä olevien rakennusteknisten, geologisten ja ilmastollisten tekijöiden avulla.

Tämän tutkimuksen empiiriset aineistot ovat peräisin kolmelta taholta. Säteilyturvakeskus tekee radonmittauksia ja kerää tietoa mittaamiensa asuntojen rakennustekniikasta Suomessa. Geologian tutkimuskeskuksen aineistoissa on tietoa maalajeista sekä maaperän uraanipitoisuudesta. Ilmatieteen laitos kerää lämpötila- ja tuulitietoja.

Sisäilman radonpitoisuutta mallinnetaan tässä työssä lineaarisella sekamallilla. Sekamalleilla mallinnettavan havaintojen korreloituneisuuden aiheuttavat maantieteellisesti lähekkäin sijaitseviin asuntoihin samankaltaisesti vaikuttavat havaitut geologiset ja ilmastolliset tekijät. Asuntojen lähekkäisyyden määrittämiseksi työssä esitellään yksinkertainen algoritmi.

Työn tuloksena saadaan estimaatit rakennusteknisille, geologisille ja ilmastokovariaateille. Lopullisessa mallissa asuntojen sisäilman radonpitoisuutta selittävät ilmanvaihtotapa, ulko- ja sisäilman lämpötilaero, kantavien rakenteiden materiaali, perustamistapa, asuttujen kerrosten lukumäärä, radonputkiston olemassa olo, maaperän uraanipitoisuus sekä perustamispaikan maalaji. Erityisen myönteiseksi tulokseksi nousee radontorjunnan radonpitoisuutta merkittävästi alentava vaikutus. Mielenkiintoisiksi ja hieman ristiriitaisiksikin tuloksiksi muodostuvat ilmanvaihtotavan ja lämpötilaeron sekä maalajin tiiviyyden ja maaperän uraanipitoisuuden yhdysvaikutukset.

Avainsanat: sisäilman radonpitoisuus, spatiaalinen interpolointi, sekamalli

Esipuhe

Tämä pro gradu -tutkielma on tehty Säteilyturvakeskuksen toimeksiantamana Jyväskylän yliopistossa keväällä 2008.

Työ sai alkunsa kesällä 2007 ollessani tilastotieteen kesäharjoittelussa Säteilyturvakeskuksessa Helsingissä. Työtehtäväni liittyivät asuntojen sisäilman radonpitoisuuden valtakunnalliseen otantatutkimukseen, ja aiheeseen liittyen heräsi kiinnostus yhdessä erikoistutkija Ilona Mäkeläisen kanssa pro gradu -työtä kohtaan. Säteilyturvakeskuksen mielenkiinnon kohteena oli erityisesti geologisten tekijöiden vaikutus sisäilman radonpitoisuuteen rakennusteknisten tekijöiden ohella. Työssä käytetty radonaineisto on Säteilyturvakeskuksesta, maaperäaineisto Geologian tutkimuskeskuksesta ja säätiedot Ilmatieteen laitokselta.

Työn tekeminen on ollut mielestäni erittäin mielenkiintoista ja antoisaa yhteistyössä Säteilyturvakeskuksen yhteyshenkilöiden ja työn ohjaajien kanssa. Haluankin kiittää Säteilyturvakeskuksen erikoistutkija Ilona Mäkeläistä, tutkija Tuomas Valmaria ja työn ohjaajana toiminutta laboratorion johtaja Hannu Arvelaa väivännööstä aineiston kokoamiseen sekä radontutkimukseen liittyvistä asiantuntijakommenteista. Lisäksi haluan erityisesti kiittää työn toisena ohjaajana toiminnutta Jyväskylän yliopiston matematiikan ja tilastotieteen laitoksen lehtori Harri Högmanderia asiantuntevasta ja kannustavasta ohjauksesta sekä kiinnostuksesta aiheittani kohtaan. Osoitan kiitokseni myös Jyväskylän yliopiston matematiikan ja tilastotieteen laitoksen professori Antti Penttiselle arvokkaista kommenteista työn eri vaiheissa sekä käytännön asioiden hoitamisesta Säteilyturvakeskuksen kanssa.

Topi Kinnunen

Sisältö

1 Johdanto	5
2 Taustaa radonista	7
2.1 Perustietoa radonista	7
2.2 Sisäilman radonpitoisuuden kuvaaminen fysikaalisella mallilla . . .	7
2.2.1 Maaperän huokosilman virtauksesta maaperästä sisätiloihin	8
2.2.2 Maaperän huokosilman radonpitoisuudesta sekä radonin kul-	
keutumisesta diffuusion ja talousveden kautta	9
2.2.3 Rakennuksen ilmanvaihtuvuudesta ja tilavuudesta	9
2.3 Radonpitoisuuden mittaamisesta	10
3 Aineisto	12
3.1 Aineistojen kerääminen	12
3.2 Aineistojen muuttajat	13
4 Tilastollisten mallien teoriaa	15
4.1 Spatiaalinen interpolointi	15
4.1.1 Stationaarisuus ja isotrooppisuus	16
4.1.2 Variogrammi	17
4.1.3 Variogrammin estimoinnista	19
4.1.4 Variogrammimallin sovittamisesta	21
4.1.5 Tavanomainen kriging	21
4.2 Sekamalli	23
4.2.1 Uskottavuusfunktio ja suurimman uskottavuuden estimointi	26
4.2.2 Varianssin harhaton estimointi normaalijakauman tilanteessa	27
4.2.3 REML-menetelmä sekamalleissa	27
4.2.4 Kiinteän osan estimointi ja satunnaisvaikutusten ennusta-	
minen	28
4.2.5 Hypoteesien testaaminen	29
4.2.6 Sekamallin oletusten tutkimisesta	30
5 Radonpitoisuuden mallintaminen	31
5.1 Tutkimusaineiston rajaus	31
5.2 Uraanipitoisuuden interpolointi	31
5.3 Sekamallin sovittaminen aineistoon	33
5.4 Mallin tulkintaa	41
5.5 Radonpitoisuuksissa tapahtuneet muutokset asunnon valmistumis-	
vuoden mukaan	43
5.6 Asunnon radonpitoisuuden ennustaminen	44
6 Yhteenveto	48
Viitteet	50
Liite: Radonmittauslomake	52

1 Johdanto

Tämän pro gradu -työn tarkoituksena on tutkia rakennusteknisten, geologisten ja ilmastollisten tekijöiden yhteyttä asuntojen sisäilman radonpitoisuuden kanssa. Tavoitteena on, että työn valmistumisen jälkeen voitaisiin hyödyntää saatavilla olevaa monipuolista geologista kartoitusaineistoa nykyistä paremmin myös rakentamattomilla alueilla. Lisäksi tavoitteena on hyödyntää rakennusteknisten tekijöiden vaikutusten arviointia uudisrakentamisen radontorjunnan kehittämisessä ja viranomaispäätöksenteossa.

Sisäilman radonpitoisuuden tutkiminen on tärkeää, koska radonin tiedetään aiheuttavan tupakoinnin jälkeen eniten keuhkosyöpää. Radonpitoisuuden tutkiminen on Suomessa erityisen kiinnostavaa ilmaston, rakennustekniikan ja geologian takia. Edellä mainituista tekijöistä johtuen radonpitoisuudet Suomessa ovat kenties maailman suurimpia. Radonilta voidaan suojautua rakentamalla asunnot radonturvallisesti. Radonturvallinen rakentaminen on erityisen tärkeää eteläisessä Suomessa johtuen alueen maaperän geologiasta. Tämän työn tutkimusalueena ovat juuri eteläisen Suomen korkean radonpitoisuuden alueet, joissa on tehty viimeisten 20–30 vuoden aikana paljon mittauksia.

Sisäilman radonin lähteinä pidetään maa- ja kallioperän radonpitoista huokosilmaa, rakennusmateriaaleista erittyvää radonia sekä porakaivoveden mukana asuntoon kulkeutuvaa radonia. Radonpitoisen huokosilman kulkeutuminen maaperästä asuntoon johtuu asunnon sisäilman ja ulkoilman lämpötilojen erosta, joka aiheuttaa asuntoon alipaineen. Alipaine aiheuttaa sen, että korvausilmaa tulee ulkoilman lisäksi myös maaperästä. Huokosilman virtauksen määrään maaperästä asuntoon vaikuttavat muun muassa maalajin tiiviys, perustamistapa, perustuksen materiaali ja talotyyppi. Muiden tekijöiden pysyessä vakioina sisäilman radonpitoisuus on sitä pienempi, mitä suurempia ovat asunnon ilmanvaihtuvuus ja tilavuus. Itse radonmittaus voidaan tehdä monella eri menetelmällä. Tämän aineiston mittaustulokset on saatu käyttämällä niin sanottua alfa jälkimenetelmää, jossa radonmittauspurkkia on pidetty asunnossa kaksi kuukautta.

Asuntojen sisäilman radonpitoisuus ilmiönä sisältää paljon vaihtelua. Vaihtelun mallintamisen tekee haasteelliseksi se, että radonpitoisuuteen vaikuttavat hyvin monet tekijät sekä se, että useimmat tekijät ovat luokitteluasteikollisia. Tästä syystä tutkimusaineistoa rajataan sisällöllisin perustein. Tutkittavat asunnot sijaitsevat Kanta- ja Päijät-Hämeessä, jolloin tutkimusalueen koko on noin 200 km × 150 km. Tällä alueella tarkastelun kohteena ovat matalaperustaiset omakoti- ja rivitalot. Radonmittauksen suorittaminen ja talokohtaisten tietojen kirjaaminen tapahtuvat aina asukkaan toimesta, joten mittauksen laatu ja tietojen oikeellisuus riippuvat jonkin verran myös asukkaasta.

Aikaisemmissa tutkimuksissa ei ole juurikaan käytetty tässä tutkimuksessa käytössä olevaa tietoa maa- ja kallioperän uraanipitoisuudesta radonpitoisuuden mal-

lintamisessa. Maaperänäytteitä, joista uraanipitoisuus on määritetty, on kerätty ympäri Suomea. Uraanipitoisuuden ennustamiseen mitattujen talojen kohdalla käytetään alaluvussa 4.1 esitettyä kriging-menetelmää, jossa hyödynnetään uraanipitoisuuden spatiaalista riippuvuutta.

Asuntojen sisäilman radonpitoisuuden mallintaminen tehdään lineaarisella sekamallilla, joka esitellään alaluvussa 4.2. Sekamallin kiinteän osan muodostavat rakennustekniset, ilmastolliset ja maaperän tiiviyttä kuvaavat tekijät yhdessä maaperän uraanipitoisuuden kanssa. Sekamallissa mallinnettavan havaintojen korreloituneisuuden aiheuttanevat sellaiset ilmastolliset ja geologiset tekijät, jotka vaikuttavat paikallisesti ja joista ei ole mittaustietoa. Tällaisia tekijöitä voivat olla esimerkiksi asuinalueen maalajin raekoko ja maaston peitteisyys. Ensimmäisenä mainittu liittyy maaperän tiiviyteen ja jälkimmäinen tuulen vaikuttavuuteen.

Johdannon jälkeen tutkielman toisessa luvussa kerrotaan perustietoa radonista ja sen kulkeutumisesta asunnon sisäilmaan sekä radonpitoisuuden mittaamisesta. Kolmannessa luvussa esitellään Säteilyturvakeskuksesta ja Geologian tutkimuskeskuksesta saadut tutkimusaineistot. Spatiaalisen interpoloinnin ja sekamallin teoriaa käsitellään luvussa neljä ja viidennessä luvussa esitellään näiden menetelmien soveltamista tutkimusaineistoon.

2 Taustaa radonista

2.1 Perustietoa radonista

Radon on näkymätön, hajuton ja mauton radioaktiivinen kaasu, jonka todettu lisäävän riskiä sairastua keuhkosityöpään. Hengityksen mukana radonkaasu kulkeutuu keuhkoihin ja myös poistuu sieltä. Radonin kiinteät hajoamistuotteet tarttuvat keuhkojen sisäpintaan ja lisäävät siten säteilyannosta. Radonia syntyy maa- ja kallioperässä radioaktiivisen uraanin hajotessa. Suomessa vuosittain todettavista noin 2000 keuhkosityöpätapauksesta 100–600:n arvioidaan aiheutuvan radonista. (Weltner, Arvela, Turtiainen, Mäkeläinen & Valmari 2003.)

Radonpitoisuuden yksikkö on becquereliä kuutiometrissä (Bq/m^3). Becquerel on säteilylähteen aktiivisuuden yksikkö, ja yksi becquerel tarkoittaa yhden atomin hajoamista sekunnissa. Pientaloasuntojen sisäilman radonpitoisuuden vuosikeskiarvo Suomessa on $145 \text{ Bq}/\text{m}^3$ (Arvela, Mäkeläinen & Castrén 1993, 11). Maa- ja kallioperän huokosilman radonpitoisuus vaihtelee maaperän uraanipitoisuuden mukaan, mutta maaperän radonpitoisuuden keskiarvo on noin $50\,000 \text{ Bq}/\text{m}^3$ (Arvela 1995b, 35). Sosiaali- ja terveysministeriön antaman päätöksen perusteella sisäilman radonpitoisuuden enimmäisarvo (vuosikeskiarvo) uusille asunnoille on $200 \text{ Bq}/\text{m}^3$ ja olemassa oleville asunnoille $400 \text{ Bq}/\text{m}^3$ (1.11.1992/944).

2.2 Sisäilman radonpitoisuuden kuvaaminen fysikaalisella mallilla

Sisäilman radonpitoisuutta voidaan kuvata yksinkertaistetulla fysikaalisella mallilla, jossa radonin lähteet ovat osoittajassa ja radonpitoisuutta laimentavat tekijät nimittäjässä:

$$(1) \quad C = \frac{S_s + S_d + S_w}{N \cdot V} = \frac{Q \cdot C_s + S_d + S_w}{N \cdot V},$$

missä

C on sisäilman radonpitoisuus (Bq/m^3),

S_s on radonin virtausnopeus asuntoon becquereleinä tunnissa (Bq/h),

S_d on radonin diffuusio rakennusmateriaaleista (Bq/h),

S_w on talousvedestä vapautuva radon (Bq/h),

N on ilmanvaihtuvuus rakennuksessa (h^{-1}),

V on rakennuksen tilavuus (m^3),

Q on maaperän huokosilman virtaus maaperästä sisätiloihin (m^3/h) ja

C_s on maaperän huokosilman radonpitoisuus (Bq/m^3).

Kaavassa (1) esitetyistä tekijöistä radonpitoisuutta lukuun ottamatta ei ole käytössä mallinnukseen soveltuvaa mittausaineistoa. Kaavan oikealla puolella olevat suureet ovat fysikaalisia ja tässä tutkimuksessa käytössä olevien selittävien muuttujien valossa latentteja, havaitsemattomia. Yksityiskohtaisempaa fysikaalista lähestymistapaa asuntojen sisäilman radonpitoisuuden vaihtelun selittämiseksi käsitellään esimerkiksi artikkeleissa Arvela, Voutilainen, Mäkeläinen, Castrén & Winqvist (1988), Arvela & Winqvist (1989) sekä Arvela (1995b). Seuraavaksi selitetään tarkemmin kaavassa (1) esiintyviä tekijöitä ja mitattujen muuttujien vaikutuksia sisäilman radonpitoisuuteen.

2.2.1 Maaperän huokosilman virtauksesta maaperästä sisätiloihin

Huokosilman virtauksen maaperästä asunnon sisätiloihin aiheuttaa alipaine, joka syntyy lämpötilaerosta sisä- ja ulkoilman välillä. Alipainetta saattaa lisätä koneellinen ilmanvaihto ja erityisesti koneellinen poistoilmanvaihto. Koneellisella poistoilmanvaihdolla ei pystytä vaikuttamaan korvausilman määrään ulkoilmasta yhtä hyvin kuin koneellisella tulo- ja poistoilmanvaihdolla, jossa poistoilmanvirtaus säädetään vain lievästi tuloilmavirtausta suuremmaksi. Pientaloihin virtaa huokosilmaa tyypillisesti 0.2–2 m³/h (Arvela 1995a). Seuraavaksi esitetään maaperägeologiaa ja rakennusteknisiä tekijöitä, joiden tiedetään vaikuttavan huokosilman virtaukseen asunnon sisätiloihin.

Maaperän läpäisevyys kasvattaa asunnon alipaineisuuden aiheuttamaa maaperän radonpitoisen ilman virtausta sisätiloihin. Tiiviinä maalajeina pidetään savea, hiesua ja eloperäisiä maalajeja. Läpäisevinä maalajeina pidetään jäätikkömuodostumia. Muut maalajit, kuten kallio, hiekka ja moreeni, sijoittuvat edellä mainittujen välimaastoon.

Pientalotyyppi vaikuttaa lähinnä huokosilman vuotoreittien määrään. Eri pientalotyyppisiä ovat kellarittomat (matalaperustaiset) talot, kellarilliset talot ja rinnetalot. Pientalotyypeistä rinnetaloissa on keskimäärin suurempia pitoisuuksia kuin muissa talotyypeissä, sillä maanvastaisten seinien kautta asuntoihin pääsee helpommin radonpitoista ilmaa. Kellaritaloissa pitoisuudet ovat pienempiä, koska maaperästä kulkeutuva radonpitoinen ilma pääsee usein helpoimmin kellariin. Kellarin ilmanvaihto ja suora kulkuyhteys ulos poistavat radonpitoista ilmaa ilman, että se kulkeutuu asuttuihin sisätiloihin. Selvästi yleisin pientalotyyppi on matalaperustainen talo. Matalaperustaiset talot sijoittuvat radonpitoisuuksien suhteen edellä mainittujen talotyyppien välimaastoon.

Kolmantena tekijänä esitellään perustamistapa. Yleisin perustamistapa on perusmuuri ja maanvarainen lattialaatta. Muita käytettyjä tapoja ovat reunajäykistetty laatta ja ryömintätilainen perustus sekä edellisten yhdistelmät. Edellä mainituista radonturvallisista ratkaisuista on ryömintätilainen perustus, sillä maaperästä tuleva radonpitoinen ilma sekoittuu ryömintätilassa puhtaaseen ulkoilmaan. Ryömintätilaisten perustusten osuus viimeisen viidenkymmenen vuoden aikana

rakennetuissa taloissa on kuitenkin pieni. Muissa perustamistavoissa lattia tai lattialaatta on suorassa yhteydessä maaperään, jolloin suurimmat radonpitoisen ilman vuotoreitit ovat lattiassa olevat läpiviennit ja maanvaraisen lattialaatan tapauksessa myös laatan reunaraot. Perusmuurin materiaaleista betoni tiiviimpänä on radonturvallisuuden kannalta parempi ratkaisu kuin huokoinen kevytsoraharkko.

Rakennusvaiheen radontorjuntatoimenpiteet sekä radonkorjaustoimenpiteet ole-massa olevaan asuntoon alentavat oikein tehtyinä asunnon radonpitoisuutta. Radonilta voidaan suojautua rakennusvaiheessa esimerkiksi tiivistämällä perustus-ten ja lattialaatan väliset raot ja laatan läpiviennit sekä asentamalla lattialaatan alle imuputkisto, joka otetaan tarvittaessa käyttöön asentamalla siihen imuri. Val-miissa rakennuksessa radonilta voidaan suojautua esimerkiksi tiivistämällä pai-kannetut vuotokohdat, tehostamalla ilmanvaihtoa sekä asentamalla radonimuri tai radonkaivo (ks. Arvela 1995a, 24). Edellä mainituissa toimenpiteissä oleellista on työn laatu ja huonosti tehty torjunta- tai korjaustoimenpide ei välttämättä vaikutakaan radonpitoisuutta alentavasti.

2.2.2 Maaperän huokosilman radonpitoisuudesta sekä radonin kul-keutumisesta diffuusion ja talousveden kautta

Huokosilman radonpitoisuus on keskimäärin noin $50\,000\text{ Bq/m}^3$, joka on laskettu Suomessa olevan moreenin keskimääräisen uraanipitoisuuden, 41 Bq/kg , avulla (ks. Arvela 1995b, 35). Huokosilman radon siirtyy asunnon sisäilmaan konvek-tion kautta. Konvektiolla tarkoitetaan radonin siirtymistä sitä kuljettavan aineen mukana.

Kiviainesta sisältävissä rakennusmateriaaleissa erityisesti betonissa on radonia tuottavaa radiumia. Rakennusmateriaaleista radon kulkeutuu sisäilmaan diffuusion avulla, jolla tarkoitetaan radonin kulkeutumista suuremmasta pitoisuudesta pienempään (ks. Weltner ym. 2003, 157). Puu on rakennusmateriaalina radonin kannalta turvallisempi, koska se ei sisällä radiumia ja ei siten eritä radonia.

Talousvedessä esiintyvä radon siirtyy huoneilmaan vedenkäytön yhteydessä. Asun-noissa, joihin talousvesi tulee porakaivosta, havaitaan suurempia sisäilman radon-pitoisuuksia kuin vesilaitosvettä käyttävissä asunnoissa, joihin veden kautta ei ole-teta kulkeutuvan radonia. Radonpitoinen porakaivovesi saattaa olla merkittävä radonin lähde käytettäessä hyvin radonpitoista vettä.

2.2.3 Rakennuksen ilmanvaihtuvuudesta ja tilavuudesta

Radonpitoisuutta laimentavia tekijöitä ovat rakennuksen ilmanvaihtuvuus (N) ja talon tilavuus (V). Ilmanvaihto laimentaa radonpitoisuutta siirtämällä radonpi-toista sisäilmaa asunnosta ulos ja korvaamalla poistettua ilmaa vähemmän radon-pitoisella ulkoilmalla. Ilmanvaihtoon vaikuttavat myös rakenteiden tiiviys, raitis-

ilmaventtiilit ja ikkunoiden kautta tapahtuva tuuletus. Tuuli vaikuttaa asunnon ilmanvaihtuvuuteen siten, että tuulen voimistuessa ilmanvaihto tehostuu.

Eri ilmanvaihtoratkaisuja ovat painovoimainen, koneellinen poisto- sekä koneellinen tulo- ja poistoilmanvaihto. Pientaloissa käytettävä ilmanvaihto on tehostunut ajan myötä. Aikaisemmin yleisesti käytössä ollutta painovoimaista ilmanvaihtoa on viime vuosina korvattu koneellisilla ilmanvaihtoratkaisuille. Koneellinen poistoilmanvaihto siirtää radonpitoista sisäilmaa ulos, mutta suurentaa samalla asunnon alipainetta, joka lisää maaperän huokosilman virtausta asuntoon. Oikein säädetty koneellinen tulo- ja poistoilmanvaihto tehostaa radonpitoisen ilman siirtymistä asunnosta ulkoilmaan, mutta pitää talon alipaineen pienempänä kuin koneellinen poisto.

Talon tilavuus liittyy radonin laimenemiseen siten, että monikerroksisessa talossa radonpitoisuus laimenee suhteessa enemmän yksikerroksiseen taloon verrattuna. Monikerroksisen asunnon alemmissa kerroksissa on suuremmat pitoisuudet kuin ylemmissä, koska pääsääntöisesti maasta tuleva radonpitoinen ilma ei sekoitu täysin kerrosten välillä.

2.3 Radonpitoisuuden mittaamisesta

Asunnon sisäilman radonpitoisuutta mitataan Säteilyturvakeskuksen radonmittauspurkilla. Mittaus suoritetaan kahden kuukauden mittauksena marras- ja huhtikuun välisenä aikana, sillä tällöin radonpitoisuudet ovat suurimpia sisä- ja ulkoilman lämpötilaerosta johtuen. Asukas on ohjeistettu sijoittamaan purkki asunnon alimpaan asuttuun kerrokseen. Mittausmenetelmä on niin sanottu alfajälkimenetelmä, jossa purkkiin menevä radonkaasu ja sen hajoamistuotteiden lähettämä alfasäteily vaikuttavat purkin pohjalla olevan muovikalvon rakenteeseen. Säteilyturvakeskuksessa muovikalvo etsataan sähkökemiallisesti, jolloin radonin aiheuttamat jäljet muuttuvat näkyviksi. Jäljet lasketaan kuva-analyytilaitteistolla. (Arvela ym. 1993.)

Mittausajan radonpitoisuus saadaan etsauksessa esiin saatujen jälkien, tunnetun taustajälkien määrän ja mittausajan avulla tietyllä muunnoskaavalla. Etsaukseen ja jälkien laskentaan liittyy jonkin verran mittausvirhettä. Suhteellinen mittausvirhe on suuri radonpitoisuuden ollessa pieni, koska tällöin taustajälkiin liittyvä epätarkkuus on merkittävä. Erityisesti asunnossa, jossa on kahden kuukauden aikana ollut suuri radonpitoisuus, muovikalvon jälkien määrä on hyvin suuri. Tällöin jälkien lukumäärän arvioiminen kuva-analyytilaitteistolla muuttuu epätarkaksi. Tästä johtuen muunnoskaava joudutaan vaihtamaan toiseen, kun muovikalvon jälkimäärä ylittää raja-arvon.

Asunnon sisäilman radonpitoisuutta voidaan mitata useilla eri menetelmillä. Säteilyturvakeskuksessa käytössä oleva alfajälkimenetelmä on suhteellisen nopea ja edullinen. Lisäksi mittaustarkkuus vuosikeskiarvoa ajatellen on parempi kuin

esimerkiksi viikon mittaisissa mittauksissa. Edellisessä kappaleessa mainittu radonpitoisuus on kahden kuukauden mittausajan radonpitoisuus, joka on tässä työssä vastemuuttujana. Sosiaali- ja terveysministeriön asettamat enimmäisarvot asuntojen sisäilman radonpitoisuudelle koskevat vuoden keskimääräistä radonpitoisuutta. Radonpitoisuuden vuosikeskiarvoa varten Säteilyturvakeskuksessa on määritetty malli, jolla talviajan mittauksesta saadaan arvioitua vuoden keskimääräinen radonpitoisuus asunnossa. Malli huomioi mittauksen aikana vallinneen lämpötilan ja tuulen nopeuden.

3 Aineisto

Tutkimuksessa käytettävät aineistot ovat peräisin Säteilyturvakeskuksesta (STUK) ja Geologian tutkimuskeskuksesta (GTK) sekä Ilmatieteen laitokselta. Säteilyturvakeskuksesta saatu aineisto koostuu kunkin asunnon radonmittauksen tai radonmittausten lisäksi asunnon osoite- ja rakennusteknisistä tiedoista. Rakennustekniset tiedot kerätään mittauksen yhteydessä asukkaalle lähetetyllä kaksisivuisella lomakkeella (Liite), jonka asukas itse täyttää. Radonaineisto sisältää lisäksi kunkin mittauspurkin mittausajan lämpötila- ja tuulikeskiarvot, jotka ovat peräisin Ilmatieteen laitokselta, Helsinki–Vantaan lentokentän sääasemalta. Radonaineistoon on liitetty Geologian tutkimuskeskuksesta saatu tieto asunnon perustamispaikan maalajista. Maantieteellisesti radonaineisto on eteläisestä Suomesta siten, että pohjoisimmat maakunnat ovat Etelä-Pohjanmaa, Keski-Suomi ja Etelä-Savo. Geologian tutkimuskeskuksesta on saatu valtakunnalliset kallioperän ja moreenin uraaniaineistot.

3.1 Aineistojen kerääminen

Tässä työssä analysoitava radonaineisto on Säteilyturvakeskuksen radonmittaus-tietokannasta, johon aineisto on kerätty kahdella eri tavalla. Osa havainnoista on kuntien tilauksiin ja yksittäisten henkilöiden omaan aktiivisuuteen perustuvia mittauksia, jotka asiakkaat kustantavat itse. Tämän tutkimuksen aineiston ensimmäiset radonmittaukset ovat vuodelta 1991.

Toinen lähde on ns. radontalkooaineisto (Säteilyturvakeskus 2008). Säteilyturvakeskus on järjestänyt kunnissa radontalkoita vuodesta 2003 alkaen erityisesti eteläisessä Suomessa korkean radonpitoisuuden alueella. Kunnat ovat tukeneet näitä mittauksia rahallisesti ja asukkaat ovat voineet teettää mittaukset hieman edullisemmin. Edellä mainitut tekijät ovat vaikuttaneet siihen, että radontalkooaineiston mittaukset on usein saatu maantieteellisesti hyvin pieniltä alueilta. Tämän tutkimuksen radonmittausaineistossa on yhteensä 13 136 radonmittausta taustatietoineen.

Kallioperän uraaniaineisto on peräisin GTK:n Suomen kalliogeokemian tietokannasta, joka sisältää kemiallista tietoa 6544 kallioperänäytteestä eri puolilta Suomea. Mittaukset on suoritettu vuosien 1990 ja 1995 välisenä aikana. Mittausten otantastrategia perustuu ositettuun otantaan, jossa alueelta otettujen näytteiden määrä riippuu alueen kivilajien vaihtelusta. Otantastrategialla on pyritty saamaan edustava otos kivilajeista. (Rasilainen, Lahtinen & Bornhorst 2007.)

Moreenin uraanipitoisuusaineisto on peräisin GTK:n suuralueellisesta moreenin geokemiallisesta kartoituksesta. Moreeninäytteet on kerätty lähes säännöllisen hila-lakuvion mukaisesti koko Suomesta. Aineisto sisältää 1057 näytettä, jotka on kerätty vuoden 1984 aikana. Tavoitteena on ollut saada mahdollisimman hyvin alueen moreenia kuvaava näyte ja näytteet (5 kpl/alue) onkin kerätty suorakaiteen

muotoiselta alueelta, jonka koko on $300 \text{ m} \times 1\,000 \text{ m}$. (Koljonen, Gustavsson, Noras & Tanskanen 1992, 14.)

3.2 Aineistojen muuttujat

Säteilyturvakeskuksen radonaineistossa on kaikkiaan 56 muuttujaa, joista tutkimuksessa käytetään vastemuuttujaa (radonpitoisuus) sekä ennakkotiedon perusteella valikoituja radonpitoisuuteen vaikuttavia muuttujia. Rakennustekniset ja asunnon käyttöön liittyvät tiedot ovat asukkaan ilmoittamia. Säteilyturvakeskus on lisännyt aineistoon radonmittaukseen liittyvät tiedot, tiedon asunnon perustamispaikan maalajista sekä lämpötila- ja tuulitiedot Helsinki–Vantaan lentokentältä. Taulukossa 1 on esitetty käytettävät muuttujat.

Molemmat uraaniaineistot sisältävät uraanimittausten sijaintitiedot (yhtenäiskoordinaatit) ja mittaustuloksen. Kallioperän uraaniaineistossa on myös tieto kilvilajista, josta näyte on otettu.

Taulukko 1: Radonaineiston muuttujat

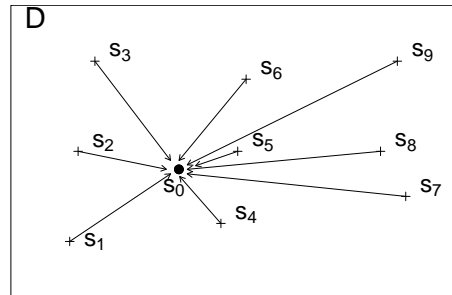
muuttuja	kuvaus	luokat tai mitta-asteikko
harkko	harkko perusmuurin materiaalina	1 = kyllä, 0 = ei
huone	mittaushuone	mak = makuuhuone, olo = olohuone, muu = muu
ilma	ilmanvaihtotapa	A = painovoimainen, B = koneellinen poisto, C = kon. tulo/poisto
kallio	asunto perustettu kalliolle	A = ei, B = kyllä, C = kyllä, kalliota louhittu
kanmat	kantavien rakenteiden materiaali	M = mineraalipohjainen, P = puu
korjaus	radonkorjaus	1 = on tehty toimenpide, 0 = ei ole tehty
maalaji	perustamispaikan maalaji	H = jäätikkömuodostuma(sora) HI = hiekka, karkea hieta, K = kallio, M = moreeni, S = savi, hieno hieta, hiesu ja eloperäiset maalajit
mitker	mittauskerros	0 = maanpinnan alla, 1 = maantaso, 2 = 2. kerros
pintaa	asuinpinta-ala (m ²)	suhdeasteikko
pertapa	perustamistapa	A = perusmuuri ja laatta, B = reunajäkistetty laatta, C = ryömintätilainen, sokkeli, D = ryömintätilainen, kivijalka, E = yhdistelmä edellisistä
pientyu	pientalotyyppi	A = kellariton, B = kellarillinen, C = rinne
ptkerros	kerrosten lkm	1 tai 2
radon	radonpitoisuus (Bq/m ³)	suhdeasteikko
taloty	talotyyppi	A = omakotitalo, B = rivitalo
tdifka	lämpötilaeron keskiarvo (°C)	välimatka-asteikko
torj	radontorjunta	1 = perustuksissa on imuputkisto (imuri on tai ei ole toiminnassa), 0 = ei ole
vesi	talousvesi porakaivosta	1 = kyllä, 0 = ei
vv	valmistumisvuosi	välimatka-asteikko
wspdka	tuulikeskiarvo (m/s)	suhdeasteikko
ykx	yhtenäiskoordinaatti	pohjoiskoordinaatti
yky	yhtenäiskoordinaatti	itäkoordinaatti

4 Tilastollisten mallien teoriaa

Tässä luvussa käsitellään tutkimukseen liittyvää tilastotieteen teoriaa. Ensimmäisessä alaluvussa esitellään spatiaalista interpolointia, jonka avulla saadaan estimaatti uraanipitoisuudelle jokaisen radonmittauksessa mukana olevan talon kohdalla. Toinen alaluku johdattelee sekamallien teoriaan. Tässä tutkimuksessa analyysimenetelmänä käytetään sekamallia, koska toisiaan lähellä olevissa asunnoissa mitattujen radonpitoisuuksien voidaan olettaa korreloivan. Spatiaalisen korrelaation aiheuttanevat asuntoihin samankaltaisesti vaikuttavat sellaiset geologiset ja ilmastolliset tekijät, joista ei ole mittaustietoa.

4.1 Spatiaalinen interpolointi

Spatiaalinen interpolointi perustuu siihen, että lähekkäiset havainnot ovat samankaltaisempia kuin suurella etäisyydellä olevat. Spatiaalisen interpoloinnin ideana on spatiaalisen muuttujan $Z(s)$ arvon estimointi paikassa s_0 , jossa muuttujaa ei ole välttämättä mitattu (Kuva 1). Interpolointimenetelmiä on useita, yksinkertaisemmista esimerkkinä lähinaapurimenetelmä ja käänteiseen etäisyyden neliöön perustuva painotettujen keskiarvojen menetelmä. Tässä työssä käytetään muuttujan $Z(s)$ spatiaalista autokorrelaatiota hyödyntävää krigingiä. Kriging on painotettuun keskiarvoon ja keskineliövirheen minimointiin perustuva spatiaalisen ennustamisen menetelmä. Nimi kriging tulee eteläafrikkalaisen kaivosinsinöörin Daniel G. Krigen mukaan, joka kehitti malminetsintään liittyviä empiirisiä menetelmiä 1950-luvulla.



Kuva 1: Spatiaalinen interpolointi, joka perustuu paikoissa s_1, \dots, s_9 tehtyihin havaintoihin

Oletetaan satunnaiskenttä (tai satunnaisprosessi) $\{Z(s): s \in D\}$, jossa $D \subset \mathbb{R}^2$ on rajoitettu. Merkitään D :n pisteitä, joissa muuttuja havaitaan s_1, \dots, s_n , jolloin aineisto on $Z(s_1), \dots, Z(s_n)$. Prosessilla $\{Z(s): s \in D\}$ tulee olla tiettyjä ominaisuuksia, jotta satunnaiskentän arvo voidaan ennustaa jossain halutussa pisteessä s_0 . Oletetaan, että $\mu(s) \equiv E[Z(s)]$ on olemassa kaikille $s \in D$, jolloin funktiota μ kutsutaan trendipinnaksi tai keskiarvofunktioksi. Oletetaan lisäksi, että prosessin varianssi on olemassa kaikille $s \in D$ eli $E[Z(s) - \mu(s)]^2 < \infty$. Prosessin

kovarianssifunktio on

$$\begin{aligned} C(s_i, s_j) &= \text{Cov}(Z(s_i), Z(s_j)) \\ &= \text{E}[Z(s_i) - \mu(s_i)][Z(s_j) - \mu(s_j)], \quad \text{kaikille } s_i, s_j \in D. \end{aligned}$$

Kovarianssifunktion C tulee olla positiivisesti definiitti. Yleisimpinä valideina kovarianssifunktioina mainittakoon sfäärinen, eksponentiaalinen ja gaussinen (Bailey & Gatrell 1995, 178–179). Seuraavassa esitellään stationaarisuuden ja isotrooppisuuden käsitteet sekä semivariogrammi. Tämän jälkeen tarkastellaan variogrammin estimointia ja variogrammin mallin sovittamista. Lopuksi käsitellään tavanomaisen krigingin ideaa.

4.1.1 Stationaarisuus ja isotrooppisuus

Prosessille $Z(s)$ voidaan määritellä useita stationaarisuuskäsitteitä, jotka eivät ole yhteneviä. Prosessia $Z(s)$ sanotaan vahvasti stationaariseksi (*strong stationary*), jos mille tahansa paikoille $s_1, \dots, s_n \in D$ ja siirrolle $h = s_i - s_j \in \mathbb{R}^2$ ($Z(s_1), \dots, Z(s_n)$):n yhteisjakauma on sama kuin ($Z(s_1 + h), \dots, Z(s_n + h)$):n yhteisjakauma. Prosessi on puolestaan heikosti stationaarinen (*weak stationary*) tai toisen kertaluvun stationaarinen (*second-order stationary*), jos sen toisen kertaluvun ominaisuudet säilyvät siirrossa h . Kolmantena stationaarisuuden muotona mainitaan sisäinen stationaarisuus (*intrinsic stationary*), jolloin riittää, että variogrammi pysyy samana siirrossa h . Koska tarkasteltava prosessi $Z(s)$ on gaussinen, niin edelliset määritelmät ovat yhtäpitäviä, sillä prosessi määräytyy kahdesta ensimmäisestä momentistaan. (Cressie 1991, 53, 60–61.)

Tarkastellaan seuraavaksi toisen kertaluvun stationaarisuutta. Oletetaan, että

$$(2) \quad \text{E}[Z(s)] = \mu, \quad \text{kaikille } s \in D,$$

jolloin siis $\mu(s) \equiv \mu$ (vakio). Oletetaan lisäksi, että

$$(3) \quad \text{Cov}(Z(s_i), Z(s_j)) = C(s_i - s_j), \quad \text{kaikille } s_i, s_j \in D,$$

missä funktiota C kutsutaan kovariogrammiksi tai stationaariseksi kovarianssifunktioksi. Satunnaisprosessia $\{Z(s): s \in D\}$ sanotaan toisen kertaluvun stationaariseksi prosessiksi, jos sille pätevät (2) ja (3). Jos prosessi on stationaarinen, niin sen keskiarvofunktio $\mu(s) \equiv \mu$ ja kovarianssifunktio

$$\begin{aligned} C(s + h, s) &= \text{Cov}(Z(s + h), Z(s)) \\ &= C(s + h - s) = C(h) \end{aligned}$$

ovat siis riippumattomia sijainnista s . Kovarianssifunktio riippuu ainoastaan siirrosta h (siirtoinvarianttisuus). Lisäksi, jos $C(s_i - s_j)$ on funktio vain pisteiden välisestä etäisyydestä $\|h\| = \|s_i - s_j\|$, niin kovarianssifunktiota C sanotaan isotrooppiseksi (*isotropic*) eli $C(s_i, s_j) = C(s, s + h) = C(h) = C(\|h\|)$ (kiertoinvarianttisuus).

Jos satunnaisprosessissa $\{Z(s): s \in D\}$ on trendi, prosessi on epästationaarinen. Tällöin aineistosta on poistettava trendi esimerkiksi sovittamalla malli

$$(4) \quad Z(s) = \beta_0 + \beta_1 x + \beta_2 y + \epsilon(s), \quad s = (x, y)$$

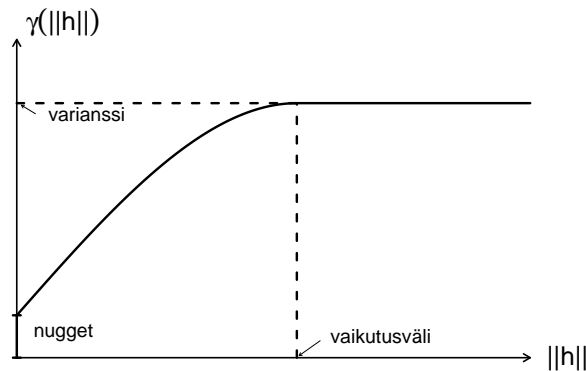
pienimmän neliösumman menetelmällä. Sovitetun mallin (4) jäännökset

$$\epsilon(s) = Z(s) - \hat{\beta}_0 - \hat{\beta}_1 x - \hat{\beta}_2 y$$

muodostavat uuden, mahdollisesti stationaarisen, aineiston.

4.1.2 Variogrammi

Kovarianssifunktion $C(s_i, s_j)$ lisäksi stationaarisen muuttujan $Z(s)$ spatiaalista riippuvuutta voidaan kuvata variogrammin (tai semivariogrammin) avulla. Semivariogrammista (Kuva 2) nähdään esimerkiksi prosessin varianssi (*sill*), mahdollinen pienen mittakaavan vaihtelun suuruus (*nugget effect*) ja vaikutusväli (*range*). Semivariogrammia tulkitaan siten, että pienillä etäisyyksillä $\|h\|$ muuttujan $Z(s)$ arvot riippuvat toisistaan ja vaikutusväliä suuremmilla etäisyyksillä havaintoja pidetään riippumattomina, jolloin semivarianssi kuvaa muuttujan $Z(s)$ varianssia. Mikäli prosessissa $\{Z(s): s \in D\}$ ei ole lainkaan spatiaalista riippuvuutta, variogrammi on vakio.



Kuva 2: Semivariogrammin muoto

Oletetaan, että

$$(5) \quad \text{Var}(Z(s_i) - Z(s_j)) = 2\gamma(s_i - s_j) = 2\gamma(s_i, s_j), \quad \text{kaikille } s_i, s_j \in D.$$

Suuretta 2γ kutsutaan variogrammiksi ja γ semivariogrammiksi. Tässä vaiheessa semivariogrammin avulla voidaan määrittellä sisäinen stationaarisuus. Satunnaisprosessin $\{Z(s): s \in D\}$ sanotaan olevan sisäisesti stationaarinen, jos sen

ensimmäisen ja toisen kertaluvun ominaisuudet voidaan kirjoittaa

$$\begin{aligned} E[Z(s)] &= \mu, & \text{kaikille } s \text{ ja} \\ \gamma(s, s+h) &= \gamma(h), & \text{kaikille } s \text{ ja } h. \end{aligned}$$

Edelleen, jos variogrammi riippuu ainoastaan pisteiden s_i ja s_j välisestä etäisyydestä, siis $2\gamma(s_i - s_j) = 2\gamma(\|s_i - s_j\|)$, sitä kutsutaan isotrooppiseksi. Kun prosessi $\{Z(s): s \in D\}$ on epäisotrooppinen eli riippuvuus $Z(s)$:n ja $Z(s+h)$:n välillä on vektorin h suunnan ja pituuden funktio, variogrammi ei enää riipu vain kahden pisteen välisestä etäisyydestä. Epäisotrooppisessa tilanteessa muuttujan $Z(s)$ spatiaalista vaihtelua voidaan edelleen kuvata semivariogrammilla, mutta nyt pistepareille, jotka ovat tietyssä suunnassa toisiinsa nähden, esimerkiksi $45^\circ \pm 22.5^\circ$. Pisteparien välisen kulman toleranssi riippuu muun muassa aineiston määrästä. Ilmeistä on, että $\gamma(-h) = \gamma(h)$ ja $\gamma(0) = 0$. Jos $\gamma(h) \rightarrow c_0 > 0$, kun $h \rightarrow 0$, niin c_0 :aa kutsutaan nuggetiksi, joka kuvaa pienen mittakaavan vaihtelua ja mittaamisen epätarkkuutta aiheuttaen epäjatkuvuuden origossa (ks. Cressie 1991, 59–60).

Validin variogrammin 2γ tulee toteuttaa ehdollisen negatiivisen definiittisyyden ehto

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j 2\gamma(s_i - s_j) \leq 0$$

jokaiselle äärelliselle kokoelmalle $\{s_i: i = 1, \dots, n\}$ paikkoja ja reaalilukuja $\{a_i: i = 1, \dots, n\}$, joille $\sum_{i=1}^n a_i = 0$. Negatiivisen definiittisyyden lisäksi validin variogrammin tulee toteuttaa myös muita ehtoja, kuten $\lim_{\|h\| \rightarrow \infty} \gamma(h)/\|h\|^2 = 0$ (ks. Cressie 1991, 86–89).

Seuraavaksi esitellään kolme tasossa \mathbb{R}^2 validia isotrooppista semivariogrammi-mallia:

Lineaarinen malli (*linear model*)

$$\gamma(h) = \begin{cases} 0, & h = 0, \\ c_0 + b_l \|h\|, & h \neq 0, \end{cases}$$

missä $c_0 \geq 0$ ja $b_l \geq 0$.

Sfäärinen malli (*spherical model*)

$$\gamma(h) = \begin{cases} 0, & h = 0, \\ c_0 + c_s \left\{ \frac{3}{2} (\|h\|/a_s) - \frac{1}{2} (\|h\|/a_s)^3 \right\}, & 0 < \|h\| \leq a_s, \\ c_0 + c_s, & \|h\| > a_s, \end{cases}$$

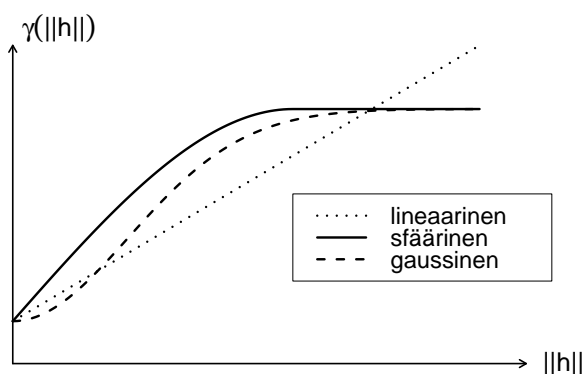
missä $c_0 \geq 0$, $c_s \geq 0$ ja $a_s \geq 0$.

Gaussinen malli (*Gaussian model*)

$$\gamma(h) = \begin{cases} 0, & h = 0, \\ c_0 + c_s \{1 - \exp(-\|h\|^2/a_s^2)\}, & h \neq 0 \end{cases}$$

missä $c_0 \geq 0$, $c_s \geq 0$ ja $a_s \geq 0$.

Malleissa parametri c_0 on nugget, b_l on kulmakerroin, c_s on varianssin (*sill*) ja nuggetin erotus ja a_s on vaikutusväli. Edellä esiteltyjä semivariogrammimalleja havainnollistetaan Kuvassa 3.



Kuva 3: Esimerkkejä semivariogrammimalleista

4.1.3 Variogrammin estimoinnista

Edellisessä kohdassa 4.1.2 esiteltyä variogrammia voidaan estimoida usealla eri menetelmällä. Variogrammiestimaattorit (otosvariogrammit) eroavat toisistaan muun muassa soveltuvuudeltaan eri otanta-asetelmille. Seuraavaksi esitellään kolme estimaattoria. Oletetaan aluksi, että ehdot (2) ja (3) pätevät, jolloin

$$(6) \quad 2\gamma(s_i, s_j) = E[Z(s_i) - Z(s_j)]^2.$$

Momenttimenetelmään perustuva estimaattori on

$$(7) \quad 2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2, \quad h \in \mathbb{R}^d,$$

missä

$$N(h) = \{(s_i, s_j) : s_i - s_j = h; i, j = 1, \dots, n\}$$

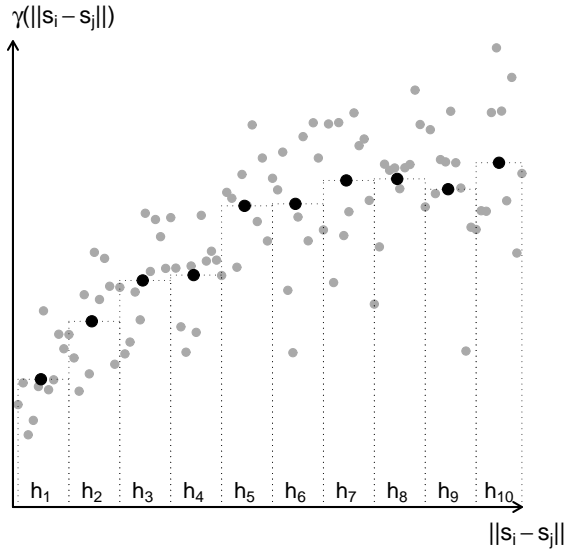
ja $|N(h)|$ on parien määrä. Momenttimenetelmä tarkoittaa sitä, että yhtälön (6) odotusarvo $E(\cdot)^2$ korvataan aineistosta lasketulla keskiarvolla $\frac{1}{|N(h)|} \sum_{N(h)} (\cdot)^2$.

Momenttimenetelmään perustuva estimaattori edellyttää, että havaintopisteet s_1, \dots, s_n sijaitsevat säännöllisesti, jolloin joukossa $N(h)$ on useita toistoja eri h :n arvoilla. Esiteltyä estimaattoria (7) kutsutaan klassiseksi variogrammiestimaattoriksi.

Jos havaintopisteet sijaitsevat epäsäännöllisesti, otetaan käyttöön toleranssialue $T(h(l))$. Toleranssialueeseen hyväksytään pisteparit (s_i, s_j) , joille $s_i - s_j$ on likimain h . Tällöin saadaan histogrammiestimaattori

$$2\gamma^+(h(l)) = \frac{1}{|T(h(l))|} \sum_{(s_i, s_j) \in T(h(l))} (Z(s_i) - Z(s_j))^2,$$

missä $|T(h(l))|$ on summattavien pisteparien (s_i, s_j) määrä, $s_i - s_j \in T(h(l))$ ja $l = 1, \dots, k$ (etäisyys). Toleranssialueiden tulisi olla niin pieniä, että aineiston spatiaalinen tarkkuus säilyy, mutta toisaalta riittävän suuria, jotta estimaattori $2\gamma^+(h(l))$ olisi vakaa. Lisäksi suosituksena on, että $|T(h(l))|$ olisi vähintään 30 (Cressie 1991, 70). Estimaattorin toimintaa isotrooppisessa tilanteessa havainnollistetaan Kuvassa 4, jossa aineiston kaikille pistepareille (s_i, s_j) on laskettu neliöity erotus $\frac{1}{2}(Z(s_i) - Z(s_j))^2$ pisteiden välisen etäisyyden $\|s_i - s_j\|$ funktiona ("variogrammipilvi").



Kuva 4: Esimerkki variogrammin estimoinnista

Kolmantena estimaattorina esitellään robusti variogrammiestimaattori

$$2\bar{\gamma}(h) = \frac{\left\{ \frac{1}{|N(h)|} \sum_{N(h)} |Z(s_i) - Z(s_j)|^{1/2} \right\}^4}{0.457 + 0.494/|N(h)|},$$

missä $|N(h)|$ ja $N(h)$ ovat kuten momenttimenetelmään perustuvassa estimaattorissa.

Epäisotrooppisessa tilanteessa variogrammiestimaattoreita käytetään kuten edellä, mutta pistepareille, jotka ovat tietyssä suunnassa toisiinsa nähden, esimerkiksi $0^\circ, 45^\circ, 90^\circ$ ja $135^\circ \pm 22.5^\circ$. Suunnat voi olla mahdollista valita luontevasti etukäteistiedon tai aineiston perusteella ja toleranssin valintaan vaikuttaa muun muassa aineiston määrä.

4.1.4 Variogrammimallin sovittamisesta

Prosessin $\{Z(s): s \in D\}$ spatiaalinen toisen kertaluvun riippuvuus rakenne tiivistetään aineiston avulla variogrammiestimaatiksi. Variogrammiestimaattiin sovitaan parametrinen variogrammimalli, joka kuvaa parhaiten aineistossa havaittua spatiaalista riippuvuutta. Mallin valinnassa, erityisesti sen parametrien osalta, auttaa ennakkotieto. Esimerkiksi nuggetin arvosta voi olla selkeä käsitys vaikeai analysoitavassa aineistossa olisikaan mittauksia riittävän lyhyillä etäisyyksillä.

Variogrammimallin parametrien estimointi ehdotetaan tehtäväksi esimerkiksi (painotetulla) pienimmän neliösumman menetelmällä (ks. Cressie 1991, 94). Tällöin parametrit, esimerkiksi $\boldsymbol{\theta} = (c_0, c_s, a_s)$, estimoidaan minimoimalla

$$\sum_{j=1}^k \left\{ 2\gamma^\#(h(j)) - 2\gamma(h(j); \boldsymbol{\theta}) \right\}^2,$$

missä $2\gamma^\#(h(j))$ on variogrammiestimaattori ja $2\gamma(h(j); \boldsymbol{\theta})$ valittu variogrammimalli. Estimointi on stabiilimpaa, jos jokin parametreista (tyypillisesti nugget) voidaan kiinnittää tai antaa jokin väli, jolta estimaattia haetaan. Tilastollisissa ohjelmissa on usein mahdollisuus käyrän sovittamiseen myös ”käsini” variogrammimalleja ja parametrien arvoja muuttamalla. Tällöin voidaan vaikuttaa käyrän muotoon ja saadaan myös käsitys eri mallien soveltuvuudesta aineistoon.

4.1.5 Tavanomainen kriging

Sanalla kriging tarkoitetaan optimaalista interpolointia. Tavoitteena on löytää paras harhaton lineaarinen ennuste prosessille $\{Z(s): s \in D\}$ pisteessä s_0 . Krigingmenetelmiä ovat muun muassa yksinkertainen (*simple kriging*), tavanomainen (*ordinary kriging*) ja universaali kriging (*universal kriging*). Krigingmenetelmistä on kehitetty myös robusteja versioita. Kirjoitetaan malli

$$Z(s) = \mu(s) + \delta(s), \quad s \in D,$$

missä μ on keskiarvofunktio ja δ on spatiaalisesti korreloitunut jäännösprosessi. Krigingmenetelmät poikkeavat toisistaan keskiarvofunktioon liittyvien oletusten suhteen. Yksinkertainen kriging olettaa μ :n tunnetuksi funktioksi, tavanomainen kriging olettaa, että $\mu(s) \equiv \mu$ on tuntematon vakio ja universaali kriging sallii

trendin. Koska tässä työssä käytettävä uraaniaineisto osoittautuu stationaariseksi, esitellään seuraavaksi vain tavanomaisen krigingin idea.

Oletetaan, että tiedetään havainnot pisteissä s_1, \dots, s_n ja että semivariogrammi on annettu. Valitaan lineaarinen ennuste, joka on havaintojen $Z(s_i)$, $i = 1, \dots, n$, painotettu summa

$$\hat{Z}(s) = \sum_{i=1}^n \lambda_i(s) Z(s_i),$$

missä $\sum_{i=1}^n \lambda_i(s) = 1$, koska ennusteelta vaaditaan, että $E[\hat{Z}(s)] = \mu$ (harhattomuusehto). Ennuste on harhaton, sillä

$$E[\hat{Z}(s)] = E\left[\sum_{i=1}^n \lambda_i(s) Z(s_i)\right] = \sum_{i=1}^n \lambda_i(s) E[Z(s_i)] = \mu \sum_{i=1}^n \lambda_i(s) = \mu.$$

Koska tavoitteena on löytää keskineliövirheen mielessä paras lineaarinen ennuste, minimoidaan

$$(8) \quad E[\hat{Z}(s) - Z(s)]^2 = \lambda'(s) \mathbf{C} \lambda(s) + \sigma^2 - 2\lambda'(s) \mathbf{c}(s),$$

missä \mathbf{C} on $n \times n$ matriisi kovariansseista $C(s_i, s_j)$, $i, j = 1, \dots, n$, $\mathbf{c}(s)$ on $n \times 1$ sarakevektori kovariansseista $C(s, s_i)$, $i = 1, \dots, n$ ja $\sigma^2 = C(s, s)$ on prosessin $Z(s)$ varianssi. Kirjoitetaan $\lambda = (\lambda_1, \dots, \lambda_n)'$, niin harhattomuusehto voidaan esittää muodossa $\lambda' \mathbf{1} = 1$, kun $\mathbf{1}$ on $n \times 1$ ykkösvektori.

Optimointitehtävään (8) täytyy lisätä harhattomuusehto esimerkiksi Lagrangen menetelmällä. Tällöin minimoitavana on lauseke

$$\lambda'(s) \mathbf{C} \lambda(s) + \sigma^2 - 2\lambda'(s) \mathbf{c}(s) + 2(\lambda' \mathbf{1} - 1) \nu(s)$$

λ :n ja Lagrangen kertoimen $\nu(s)$ suhteen. Lausekkeen ratkaisu on

$$\begin{aligned} \lambda' \mathbf{1} &= 1 \\ \mathbf{C} \lambda(s) + \mathbf{1} \nu(s) &= \mathbf{c}(s). \end{aligned}$$

Tämä voidaan kirjoittaa lineaarisena yhtälöryhmänä

$$\begin{pmatrix} C(s_1, s_1) & \cdots & C(s_1, s_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(s_n, s_1) & \cdots & C(s_n, s_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1(s) \\ \vdots \\ \lambda_n(s) \\ \nu(s) \end{pmatrix} = \begin{pmatrix} C(s, s_1) \\ \vdots \\ C(s, s_n) \\ 1 \end{pmatrix},$$

jota voidaan merkitä myös yhtälönä $\mathbf{C}_+ \lambda_+(s) = \mathbf{c}_+(s)$. Yhtälöryhmän ratkaisuna saadaan painot

$$\lambda_+(s) = \mathbf{C}_+^{-1} \mathbf{c}_+(s)$$

ja niin sanottu kriging-varianssi

$$\sigma_k^2 = \sigma^2 - \mathbf{c}'_+(s) \mathbf{C}_+^{-1} \mathbf{c}_+(s).$$

Kriging voidaan suorittaa yhteen pisteeseen s_0 tai kyseessä olevan aineiston mittakaavan suhteen tiheässä hilassa, jolloin saadaan kriging-kartta. Usein on hyödyllistä tuottaa kartta myös kriging-varianssista, jolloin saadaan käsitys ennustamiseen liittyvästä epävarmuudesta. Huomautetaan, että piirrettäessä kriging-karttaa painot λ_i lasketaan jokaiselle hilapisteelle erikseen. Toisin sanoen painot λ_i riippuvat paikasta, johon interpoloidaan. Jos oletetaan, että prosessi Z on gaus-sinen, niin kriging-varianssin avulla ennusteelle $\hat{Z}(s)$ saadaan 95 %:n ennustevali

$$[\hat{Z}(s) - 1.96\sigma_k, \hat{Z}(s) + 1.96\sigma_k].$$

4.2 Sekamalli

Tässä tutkimuksessa analyysimenetelmänä käytetään sekamallia, koska on syytä olettaa, että lähellä toisiaan mitatut radonpitoisuudet korreloivat vahvemmin kuin suurella etäisyydellä mitatut. Sekamalli antaa muun muassa mahdollisuuden mallintaa jäännösten kovarianssirakenne siten, että se ottaa huomioon havaintopaikkojen väliset etäisyydet. Toisaalta koska asunnot sijaitsevat usein ryhmitäin ja taloryhmät kaukana toisistaan, niistä voidaan myös muodostaa klustereita maantieteellisen sijaintinsa mukaan. Tällöin saadaan estimoitua klustereiden aiheuttama sisäkorrelaatio, kun klusterit ajatellaan mallissa satunnaisvaikutuksiksi.

Sekamallissa osa parametreista ajatellaan satunnaismuuttujiksi. Lineaarinen sekamalli voidaan kirjoittaa muodossa

$$(9) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon},$$

missä \mathbf{y} on $n \times 1$ havaintovektori, \mathbf{X} on selittävien kiinteiden muuttujien $n \times p$ asetelmamatriisi (design-matriisi), $\boldsymbol{\beta}$ on $p \times 1$ parametrivektori, \mathbf{Z} on satunnaisvaikutusten $n \times c$ asetelmamatriisi, \mathbf{u} on $c \times 1$ satunnaisvaikutusten vektori ja $\boldsymbol{\epsilon}$ on satunnaisvirheiden $n \times 1$ vektori. Mallin ensimmäistä osaa $\mathbf{X}\boldsymbol{\beta}$ kutsutaan kiinteäksi osaksi ja jälkimmäistä $\mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$ satunnaiseksi. Sekamallissa oletetaan, että satunnaisvaikutukset ja satunnaisvirheet ovat normaalisti jakautuneita

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$$

ja

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}),$$

missä \mathbf{D} on satunnaisvaikutusten $c \times c$ kovarianssimatriisi ja \mathbf{R} on satunnaisvirheiden $n \times n$ kovarianssimatriisi. Sekamallissa oletetaan lisäksi, että satunnaisvaikutukset ja satunnaisvirheet ovat riippumattomia $\text{Cov}(\mathbf{u}, \boldsymbol{\epsilon}) = \mathbf{0}$. Havaintojen \mathbf{y} odotusarvo ehdolla satunnaisvaikutukset on

$$E(\mathbf{y}|\mathbf{u}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$$

ja havaintojen odotusarvovektori on

$$E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}) = \mathbf{X}\boldsymbol{\beta}.$$

Lisäksi satunnaisvirheet $\boldsymbol{\epsilon}$ määritellään

$$\boldsymbol{\epsilon} = \mathbf{y} - E(\mathbf{y}|\mathbf{u}).$$

Edellisten perusteella havaintojen \mathbf{y} kovarianssimatriisi on

$$\begin{aligned} \mathbf{V} &= \text{Cov}(\mathbf{y}) = \text{Cov}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}) \\ &= \text{Cov}(\mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}) = \mathbf{Z}\text{Cov}(\mathbf{u})\mathbf{Z}' + \text{Cov}(\boldsymbol{\epsilon}) \\ &= \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}. \end{aligned}$$

Tässä tutkimuksessa havaintojen riippuvuuden aiheuttavat läheisiin asuntoihin samankaltaisesti vaikuttavat ilmastolliset ja geologiset tekijät. Havainnollistetaan seuraavaksi havaintojen kovarianssimatriisin rakennetta kuvitteellisilla esimerkeillä. Havaintojen riippuvuusrakenne voidaan mallintaa pelkästään satunnaisvirheiden kovarianssimatriisin \mathbf{R} avulla. Tällöin mallissa ei ole lainkaan satunnaisvaikutusosaa $\mathbf{Z}\mathbf{u}$ ja mallia voidaan pitää sekamallin erikoistapauksena. Alaluvussa 4.1.2 on esitelty tasossa \mathbb{R}^2 valideja isotrooppisia semivariogrammillemalleja, joiden tyyppisiä rakenteita voidaan käyttää nyt sekamallin satunnaisvirheiden korrelaatioiden mallintamisessa. Havaintojen kovarianssimatriisi \mathbf{V} voitaisiin kirjoittaa kolmen havainnon ja gaussisen mallin tapauksessa esimerkiksi

$$\mathbf{V} = \mathbf{R} = \sigma_{\epsilon}^2 \begin{pmatrix} 1 & \exp(-d_{12}^2/a_s^2) & \exp(-d_{13}^2/a_s^2) \\ \exp(-d_{21}^2/a_s^2) & 1 & \exp(-d_{23}^2/a_s^2) \\ \exp(-d_{31}^2/a_s^2) & \exp(-d_{32}^2/a_s^2) & 1 \end{pmatrix}.$$

Usein asunnot sijaitsevat maantieteellisesti ryhmittäin, jolloin jäännösten korreloituneisuutta voidaan approksimoida klusteroimalla asunnot ja suorittaa mallintaminen sekamallilla. Tällöin sekamallissa satunnaisvaikutus liittyy radonpitoisuuden tasoon.

Havainnollistetaan kovarianssimatriisin \mathbf{V} rakennetta esimerkillä, jossa on kolmen ja kahden asunnon klusterit eli yhteensä viisi asuntoa (mittausta) ja kaksi klusteria. Esimerkkitapauksessa voidaan (9) kirjoittaa asuntokohtaisesti

$$y_{ji} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{p-1} x_{(p-1)i} + u_j + \epsilon_{ji},$$

missä $i = 1, \dots, 5$ on asuntoindeksi ja $j = 1, 2$ klusteri-indeksi. Tarkastellaan seuraavaksi asuntojen varianssia ja kovarianssia. Klusterissa j sijaitsevalle asunnolle i varianssi on

$$\text{Var}(y_{ji}) = \text{Var}(u_j) + \text{Var}(\epsilon_{ji}) = \sigma_u^2 + \sigma_{\epsilon}^2.$$

Saman klusterin asunnoille i ja i' kovarianssi on

$$\begin{aligned} \text{Cov}(y_{ji}, y_{ji'}) &= \text{Cov}(u_j + \epsilon_{ji}, u_j + \epsilon_{ji'}) = \text{Cov}(u_j, u_j) \\ &= \text{Var}(u_j) = \sigma_u^2 + \sigma_{\epsilon}^2 h(d_{ii'}, \boldsymbol{\theta}), \end{aligned}$$

missä $h(d_{ii'}, \boldsymbol{\theta})$ on jokin havaintojen y_{ji} ja $y_{j'i'}$ välisen euklidisen etäisyyden $d_{ii'}$ parametrinen funktio parametrein $\boldsymbol{\theta}$. $h(d_{ii'}, \boldsymbol{\theta})$ voi olla esimerkiksi gaussinen, kuten ensimmäisessä esimerkissä. Eri klustereiden asuntojen kovarianssi on

$$\text{Cov}(y_{ji}, y_{j'i'}) = \text{Cov}(u_j + \epsilon_{ji}, u_{j'} + \epsilon_{j'i'}) = \sigma_\epsilon^2 h(d_{ii'}, \boldsymbol{\theta}).$$

Saman klusterin asuntojen korrelaatio on siten

$$\rho = \text{Corr}(y_{ji}, y_{j'i'}) = \frac{\text{Cov}(y_{ji}, y_{j'i'})}{\sqrt{\text{Var}(y_{ji})\text{Var}(y_{j'i'})}} = \frac{\sigma_u^2 + \sigma_\epsilon^2 h(d_{ii'}, \boldsymbol{\theta})}{\sigma_u^2 + \sigma_\epsilon^2},$$

jota kutsutaan sisäkorrelaatioksi. Jos satunnaisvirheiden kovarianssimatriisiksi \mathbf{R} valitaan diagonaalimatriisi eli funktiota h ei estimoida, niin $\text{Cov}(y_{ji}, y_{j'i'}) = \sigma_u^2$, $\text{Cov}(y_{ji}, y_{j'i'}) = 0$ ja sisäkorrelaatio on muotoa

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}.$$

Satunnaisvaikutusten asetelmamatriisi \mathbf{Z} on

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}'$$

ja satunnaisvaikutusten kovarianssimatriisi $\mathbf{D} = \sigma_u^2 \mathbf{I}_2$ eli

$$\mathbf{D} = \sigma_u^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Havaintojen kovarianssimatriisi on nyt

$$\mathbf{V} = \mathbf{ZDZ}' + \mathbf{R} = \sigma_u^2 \mathbf{Z}\mathbf{Z}' + \mathbf{R}.$$

Ensimmäinen summattava osa on

$$\sigma_u^2 \mathbf{Z}\mathbf{Z}' = \begin{pmatrix} \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & 0 & 0 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & 0 & 0 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_u^2 & \sigma_u^2 \\ 0 & 0 & 0 & \sigma_u^2 & \sigma_u^2 \end{pmatrix}$$

ja jälkimmäinen on

$$\mathbf{R} = \sigma_\epsilon^2 \begin{pmatrix} 1 & h(d_{12}, \boldsymbol{\theta}) & h(d_{13}, \boldsymbol{\theta}) & h(d_{14}, \boldsymbol{\theta}) & h(d_{15}, \boldsymbol{\theta}) \\ h(d_{21}, \boldsymbol{\theta}) & 1 & h(d_{23}, \boldsymbol{\theta}) & h(d_{24}, \boldsymbol{\theta}) & h(d_{25}, \boldsymbol{\theta}) \\ h(d_{31}, \boldsymbol{\theta}) & h(d_{32}, \boldsymbol{\theta}) & 1 & h(d_{34}, \boldsymbol{\theta}) & h(d_{35}, \boldsymbol{\theta}) \\ h(d_{41}, \boldsymbol{\theta}) & h(d_{42}, \boldsymbol{\theta}) & h(d_{43}, \boldsymbol{\theta}) & 1 & h(d_{45}, \boldsymbol{\theta}) \\ h(d_{51}, \boldsymbol{\theta}) & h(d_{52}, \boldsymbol{\theta}) & h(d_{53}, \boldsymbol{\theta}) & h(d_{54}, \boldsymbol{\theta}) & 1 \end{pmatrix}.$$

Havaintojen kovarianssimatriisi \mathbf{V} on siten

$$\mathbf{V} = \begin{pmatrix} \sigma_u^2 + \sigma_\epsilon^2 & \sigma_u^2 + \sigma_\epsilon^2 h & \sigma_u^2 + \sigma_\epsilon^2 h & \sigma_\epsilon^2 h & \sigma_\epsilon^2 h \\ \sigma_u^2 + \sigma_\epsilon^2 h & \sigma_u^2 + \sigma_\epsilon^2 & \sigma_u^2 + \sigma_\epsilon^2 h & \sigma_\epsilon^2 h & \sigma_\epsilon^2 h \\ \sigma_u^2 + \sigma_\epsilon^2 h & \sigma_u^2 + \sigma_\epsilon^2 h & \sigma_u^2 + \sigma_\epsilon^2 & \sigma_\epsilon^2 h & \sigma_\epsilon^2 h \\ \sigma_\epsilon^2 h & \sigma_\epsilon^2 h & \sigma_\epsilon^2 h & \sigma_u^2 + \sigma_\epsilon^2 & \sigma_u^2 + \sigma_\epsilon^2 h \\ \sigma_\epsilon^2 h & \sigma_\epsilon^2 h & \sigma_\epsilon^2 h & \sigma_u^2 + \sigma_\epsilon^2 h & \sigma_u^2 + \sigma_\epsilon^2 \end{pmatrix},$$

missä funktio h on kuten edellä \mathbf{R} :ssä. Mikäli klusterit sijaitsevat kaukana toisistaan, voidaan eri klustereiden havaintojen korrelaatiot käytännössä olettaa nolliksi. Tällöin estimoinnissa otetaan huomioon havaintopaikkojen etäisyys vain klustereiden sisällä. Toisaalta koska klusteroinnilla jo approksimoidaan korreloituneita jäännöksiä, jäännösten korrelaatiomatriisi \mathbf{R} voidaan asettaa diagonaalimatriisiksi $\mathbf{R} = \sigma_\epsilon^2 \mathbf{I}_5$.

Seuraavaksi tarkastellaan menetelmiä, joilla voidaan suorittaa sekamallin varianssiparametrien ja kiinteiden β -parametrien estimointi sekä satunnaisvaikutusten ennustaminen. Lopuksi esitellään sekamalliin liittyvien hypoteesien testausta ja mallin oletusten tutkimista.

4.2.1 Uskottavuusfunktio ja suurimman uskottavuuden estimointi

Tässä esitellään suurimman uskottavuuden (*maximum likelihood*, ML) estimointia lineaarisen sekamallin tilanteessa, jossa vastemuuttuja \mathbf{y} noudattaa multinormaali-jakaamaa

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{ZDZ}' + \mathbf{R}).$$

Sekamallitilanteessa uskottavuusfunktio on

$$(10) \quad L_{ML}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\boldsymbol{\alpha})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}}{(2\pi)^{\frac{n}{2}}|\mathbf{V}(\boldsymbol{\alpha})|^{\frac{1}{2}}},$$

missä $\boldsymbol{\alpha} = (\sigma_u^2, \sigma_\epsilon^2, \boldsymbol{\theta})$ on \mathbf{V} :n sisältämien varianssiparametrien vektori ja n on havaintojen lukumäärä. Kun $\boldsymbol{\alpha}$ on tunnettu varianssiparametrien vektori, suurimman uskottavuuden estimaattori $\boldsymbol{\beta}$:lle voidaan ratkaista logaritmisesta uskottavuusfunktioista

$$\begin{aligned} \log(L_{ML}(\boldsymbol{\beta}, \boldsymbol{\alpha})) &= l_{ML}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\ &= c - \frac{1}{2}\log|\mathbf{V}(\boldsymbol{\alpha})| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\boldsymbol{\alpha})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \end{aligned}$$

missä c on parametreista $\boldsymbol{\beta}$ ja $\boldsymbol{\alpha}$ riippumaton vakio. Asettamalla osittaisderivaatta $\boldsymbol{\beta}$:n suhteen nolllaksi saadaan

$$\begin{aligned} \mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= 0 \\ \Leftrightarrow \mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{y} \\ (11) \quad \Leftrightarrow \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{y}, \end{aligned}$$

missä $\mathbf{W} = \mathbf{V}^{-1}$. $\hat{\boldsymbol{\beta}}$ on $\boldsymbol{\beta}$:n suurimman uskottavuuden estimaattori ja se on olemassa, kun käänteismatriisi $(\mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{X})^{-1}$ on olemassa. Estimaattorin $\hat{\boldsymbol{\beta}}$ varianssi on

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}.$$

Kun $\boldsymbol{\alpha}$ on tuntematon, kuten yleensä, mutta estimaatti $\hat{\boldsymbol{\alpha}}$ on käytössä, voidaan sijoittaa $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\alpha}}) = \hat{\mathbf{W}}^{-1}$ ja estimoida $\boldsymbol{\beta}$ yhtälön (11) avulla sijoittamalla \mathbf{W} :n paikalle $\hat{\mathbf{W}}$. Kaksi paljon käytettyä menetelmää $\boldsymbol{\alpha}$:n estimoimiseksi ovat suurimman uskottavuuden menetelmä ja jäännösten (tai rajoitettu) suurimman uskottavuuden (*residual maximum likelihood*, REML) menetelmä. Ensin mainittu tuottaa harhaisen estimaattorin $\boldsymbol{\alpha}$:lle ja jälkimmäinen vähemmän harhaisen, koska siinä otetaan kiinteiden parametrien estimointi huomioon.

4.2.2 Varianssin harhaton estimointi normaalijakauman tilanteessa

Seuraavaksi esitellään sellainen varianssin σ^2 estimointitekniikka, joka ei vaadi \mathbf{y} :n odotusarvorakenteen estimointia. Merkitään havaintoja $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ ja n -ulotteista ykkösvektoria $\mathbf{1}_n = (1, 1, \dots, 1)'$. Tällöin \mathbf{y} noudattaa jakaumaa $N(\mathbf{1}_n\mu, \sigma^2\mathbf{I}_n)$ ja harhaton otosvarianssi on $s^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / (n-1)$, missä \bar{y} on otoskeskiarvo. Olkoon \mathbf{A} jokin täysiasteinen $n \times (n-1)$ matriisi, jolle $\mathbf{A}'\mathbf{1}_n = \mathbf{0}$. Määritellään vektori $\mathbf{U} = \mathbf{A}'\mathbf{y}$, josta seuraa, että \mathbf{U} noudattaa normaalijakaumaa $N_{n-1}(\mathbf{0}, \sigma^2\mathbf{A}'\mathbf{A})$. Maksimoimalla vastaava uskottavuusfunktio ainoan parametrin σ^2 suhteen, saadaan REML-estimaattori

$$\hat{\sigma}^2 = \frac{\mathbf{y}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}}{n-1},$$

jonka voidaan osoittaa vastaavan otosvarianssia s^2 . (Verbeke & Molenberghs 2000, 43.) Todetaan lisäksi, että REML-estimaattori ei riipu matriisiin \mathbf{A} valinnasta.

4.2.3 REML-menetelmä sekamalleissa

Yleensä sekamalleihin sisältyy kiinteä osa, johon liittyy useita estimoitavia parametreja. Tällaisissa tilanteissa tulee ottaa huomioon kiinteän osan parametrien estimoinnista aiheutuva vapausasteiden menetys, kun estimoidaan varianssiparametreja $\boldsymbol{\alpha}$. REML-estimaattori $\boldsymbol{\alpha}$:lle saadaan sekamallissa maksimoimalla uskottavuusfunktio vektorille $\mathbf{U} = \mathbf{A}'\mathbf{y}$, missä \mathbf{A} on jokin täysiasteinen $n \times (n-p)$ matriisi, jolle $\mathbf{A}'\mathbf{X} = \mathbf{0}$. \mathbf{U} noudattaa normaalijakaumaa keskiarvolla $\mathbf{0}$ ja kovarianssilla $\mathbf{A}'\mathbf{V}(\boldsymbol{\alpha})\mathbf{A}$, eikä siis riipu kiinteistä β -parametreista. Uskottavuusfunktio muunnokselle \mathbf{U} voidaan kirjoittaa muodossa

$$(12) \quad L(\boldsymbol{\alpha}) = (2\pi)^{-\frac{n-p}{2}} |\mathbf{X}'\mathbf{X}|^{\frac{1}{2}} \times |\mathbf{X}'\mathbf{V}^{-1}(\boldsymbol{\alpha})\mathbf{X}|^{-\frac{1}{2}} \times |\mathbf{V}(\boldsymbol{\alpha})|^{-\frac{1}{2}} \\ \times \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))'\mathbf{V}^{-1}(\boldsymbol{\alpha})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))\right\},$$

missä $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ on kuten kaavassa (11) (Harville 1974, Verbeken & Molenberghsin 2000, 45 mukaan). Uskottavuusfunktio (12) voidaan kirjoittaa myös

$$L(\boldsymbol{\alpha}) = C|\mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{X}|^{-\frac{1}{2}}L_{ML}(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}), \boldsymbol{\alpha}),$$

missä C on $\boldsymbol{\alpha}$:sta riippumaton vakio ja kuten aikaisemmin $\mathbf{W}(\boldsymbol{\alpha}) = \mathbf{V}^{-1}(\boldsymbol{\alpha})$ sekä $L_{ML}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ on uskottavuusfunktio (10). Koska $|\mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{X}|$ ei riipu $\boldsymbol{\beta}$:sta, niin siitä seuraa, että REML-estimaattorit varianssiparametreille $\boldsymbol{\alpha} = (\sigma_u^2, \sigma_\epsilon^2, \boldsymbol{\theta})$ saadaan maksimoimalla niin sanottu REML-uskottavuusfunktio

$$L_{REML}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = |\mathbf{X}'\mathbf{W}(\boldsymbol{\alpha})\mathbf{X}|^{-\frac{1}{2}} L_{ML}(\boldsymbol{\beta}, \boldsymbol{\alpha}).$$

Uskottavuusfunktioiden L_{ML} tai L_{REML} numeerinen maksimointi parametrien suhteen voidaan tehdä usealla eri iteratiivisella menetelmällä, joista käytetyimpiä ovat EM-menetelmä ja Newton-Raphsonin menetelmä. EM-algoritmi on suosittu iteratiivinen menetelmä suurimman uskottavuuden estimointiin sellaisten mallien tapauksessa, joilla mallinnetaan epätäydellistä aineistoa. Lineaaristen sekamallien tapauksessa epätäydellisyyden aiheuttavat satunnaisvaikutukset u_1, \dots, u_c . Menetelmässä pyritään iteroimalla löytämään \mathbf{u} :n ehdollisen jakauman kautta suurimman uskottavuuden tuottama varianssiparametrien estimaatti $\hat{\boldsymbol{\alpha}}$. Newton-Raphsonin algoritmilla saadaan logaritmisesta uskottavuusfunktioista lasketun pistemääräfunktion (*score function*) ja sen derivaatan (Hessin matriisi) avulla uskottavuusfunktion maksimoiva $\hat{\boldsymbol{\alpha}}$. (Pinheiro & Bates 2000, 79.)

4.2.4 Kiinteän osan estimointi ja satunnaisvaikutusten ennustaminen

Yleensä aineiston mallinnuksessa mielenkiinnon kohteena on kiinteiden β -kertoimien estimointi. Voi kuitenkin olla mielekästä ennustaa myös satunnaisvaikutukset. Edellinen tarkoittaa tässä tutkimuksessa asuntoklustereiden keskimääräisen radonpitoisuuden poikkeamaa kaikkien asuntojen keskiarvosta. Yleisesti $\boldsymbol{\beta}$ ja \mathbf{u} ratkaistaan vastemuuttujan \mathbf{y} ja satunnaisvaikutusten \mathbf{u} yhteisjakauman

$$(13) \quad f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u})f(\mathbf{u}) \\ = \frac{\exp\left\{-\frac{1}{2}((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}'\mathbf{D}^{-1}\mathbf{u})\right\}}{(2\pi)^{\frac{1}{2}(n+c)}|\mathbf{R}|^{\frac{1}{2}}|\mathbf{D}|^{\frac{1}{2}}}$$

avulla (McCulloch & Searle 2001, 258). Henderson, Kempthorne, Searle ja Von Krosig (1959) ovat osoittaneet, että maksimoimalla (13) $\boldsymbol{\beta}$:n ja \mathbf{u} :n suhteen saadaan niin sanotut sekamalliyhtälöt

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

(Verbeke & Molenberghs 2000, 79–80). Sekamalliyhtälöiden ratkaisuna saadaan

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

ja

$$\tilde{\mathbf{u}} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}).$$

$\tilde{\boldsymbol{\beta}} = BLUE(\boldsymbol{\beta})$ on $\boldsymbol{\beta}$:n paras lineaarinen ja harhaton estimaattori (*best linear and unbiased estimator*) ja $\tilde{\mathbf{u}} = BLUP(\mathbf{u})$ on \mathbf{u} :n paras lineaarinen ja harhaton ennuste (*best linear and unbiased predictor*), kun varianssiparametrit ovat tunnettuja. Käytännössä varianssiparametrit estimoidaan aineistosta ja sijoitetaan edellä oleviin yhtälöihin.

4.2.5 Hypoteesien testaaminen

Esitellään seuraavaksi kiinteiden vaikutusten ja kovarianssirakenteen testaamiseen käytettyjä menetelmiä. Kiinteän osan testaamiseen käytetään yleistä lineaarista hypoteesia

$$H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{0} \quad \text{vs.} \quad H_1 : \mathbf{K}\boldsymbol{\beta} \neq \mathbf{0},$$

missä \mathbf{K} on jokin tunnettu matriisi. Testisuurena käytetään F -testisuuretta

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{K}' \left[\mathbf{K} (\mathbf{X}' \mathbf{V}^{-1} (\hat{\boldsymbol{\alpha}}) \mathbf{X})^{-1} \mathbf{K}' \right]^{-1} \mathbf{K} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\text{rank}(\mathbf{K})}.$$

Testisuureen jakaumaa voidaan approksimoida F -jakaumalla, jonka osoittajan vapausasteet ovat $\text{rank}(\mathbf{K})$ ja nimittäjän vapausasteet täytyy estimoida aineistosta (ks. Verbeke & Molenberghs 2000, 57). Yksittäistä kiinteää vaikutusta testattaessa ($H_0 : \beta_i = 0$) valitaan $\mathbf{K} = \mathbf{k}$, joka on vektori ($\text{rank}(\mathbf{k}) = 1$) ja $T = \sqrt{F}$, jolloin testisuureen

$$T = \sqrt{F} = \sqrt{\frac{\hat{\beta}_i^2}{\text{Var}(\hat{\beta}_i)}} = \frac{|\hat{\beta}_i|}{\text{s.e.}(\hat{\beta}_i)}$$

jakaumaa voidaan approksimoida t -jakaumalla.

Edellisen lisäksi testattaessa mallin kiinteää osaa, sisäkkäisten mallien testaamiseen soveltuu uskottavuussuhteen testi (*likelihood ratio test*), kun estimointimenetelmänä käytetään suurimman uskottavuuden menetelmää. REML-menetelmässä matriisi \mathbf{A} ja samalla ”havaintoaineisto” $\mathbf{U} = \mathbf{A}\mathbf{y}$ vaihtuvat, kun asetelmamatriisi \mathbf{X} muuttuu, joten testaus ei näin ollen ole mielekäästä. Uskottavuussuhteen testi määritellään

$$-2\log\lambda = -2\log\left(\frac{L_{ML}(\hat{\boldsymbol{\theta}}_{ML,0})}{L_{ML}(\hat{\boldsymbol{\theta}}_{ML})}\right),$$

missä L on uskottavuusfunktio, $\hat{\boldsymbol{\theta}}_{ML}$ on $\boldsymbol{\theta}$:n rajoittamaton ja $\hat{\boldsymbol{\theta}}_{ML,0}$ jonkin rajoitteen vallitessa estimoitu estimaatti. H_0 :n ollessa voimassa testisuure $-2\log\lambda$ noudattaa likimain χ_r^2 -jakaumaa, missä r on rajoitteiden lukumäärä.

Kovarianssirakennetta voidaan testata edellä mainitulla uskottavuussuhteen testillä. REML-menetelmä on nyt käytössä, koska kiinteiden vaikutusten asetelmamatriisi \mathbf{X} pysyy samana. Eri kovarianssirakenteita, kuten myös kiinteitä osia, voidaan tutkia lisäksi muun muassa informaatiokriteereiden avulla, jolloin mallien ei tarvitse olla sisäkkäisiä. Informaatiokriteerit ovat käyttökelpoisia etenkin silloin, kun käsitellään suurta havaintomäärää, koska tällöin edellä mainitut testausmenetelmät saattavat antaa liian merkitseviä tuloksia. *AIC* (*Akaike information criterion*) määritellään

$$AIC = -2l(\hat{\boldsymbol{\theta}}) + 2n_{par},$$

missä $l(\hat{\boldsymbol{\theta}})$ on logaritmoidun uskottavuusfunktion arvo ja n_{par} on estimoitavien parametrien lukumäärä. *BIC* (*Bayesian information criterion*) määritellään

$$BIC = -2l(\hat{\boldsymbol{\theta}}) + n_{par}\log(n_{obs}),$$

missä $l(\hat{\boldsymbol{\theta}})$ ja n_{par} kuten *AIC*:ssa ja lisäksi n_{obs} on havaintojen lukumäärä. Näin määritellyissä kriteereissä pienimmän arvon saanut malli tulkitaan parhaaksi. Verbeke ja Molenberghs (2000, 75) korostavat, että informaatiokriteereitä ei tulisi tulkita tilastollisen merkitsevyydestin tapaan, vaan ne ovat vain suuntaa antavia menetelmiä tilastollisten mallien vertailuun.

4.2.6 Sekamallin oletusten tutkimisesta

Sekamalliin liittyy kaksi tärkeää oletusta. Kuten lineaarisessa regressiomallissa, satunnaisvirheille oletetaan normaalijakauma keskiarvolla $\mathbf{0}$ ja kovarianssilla \mathbf{R} . Satunnaisvaikutuksille oletetaan myös normaalijakauma keskiarvolla $\mathbf{0}$ ja kovarianssilla \mathbf{D} . Tässä tutkimuksessa satunnaisvaikutusten kovarianssimatriisi on muotoa $\mathbf{D} = \sigma_u^2\mathbf{I}$. Oletusten voimassaolon tutkiminen lienee havainnollisinta graafisesti.

Satunnaisvaikutusten jakaumaa voidaan tutkia esimerkiksi piirtämällä satunnaisvaikutusten ennusteet suhteessa normaalijakaumaan (*quantile-quantile plot*) tai ennusteet luokittelevien muuttujien eri tasoilla. Satunnaisvirheiden jakaumaa voidaan tarkastella esimerkiksi

- satunnaisvaikutuksittain (asuntoklusterit),
- erikseen luokittelevien muuttujien tasoilla,
- piirtämällä jäännökset jatkuvan muuttujan suhteen,
- piirtämällä jäännökset sovitettujen arvojen suhteen,
- piirtämällä havaitut arvot sovitettujen arvojen suhteen ja
- piirtämällä jäännökset suhteessa normaalijakaumaan.

5 Radonpitoisuuden mallintaminen

Tässä luvussa esitetään aineiston analyysia kahdessa vaiheessa. Ensin interpoloidaan uraanipitoisuus mitattujen talojen kohdalle, jotta saadaan uraanipitoisuus radonpitoisuuden selittäjäksi. Tämän jälkeen sovitetaan aineistoon sekamalli. Havaintojen korreloituneisuuden aiheuttavat lähekkäisiin asuntoihin samankaltaisesti vaikuttavat maaperä- ja ilmastotekijät. Luvussa esitellään myös asuntojen klusteroinnissa käytetty yksinkertainen menetelmä. Lopuksi ennustetaan radonpitoisuus kolmeen asuntoon. Tilastollinen analyysi suoritetaan R-ohjelman versiolla 2.4.1.

5.1 Tutkimusaineiston raja

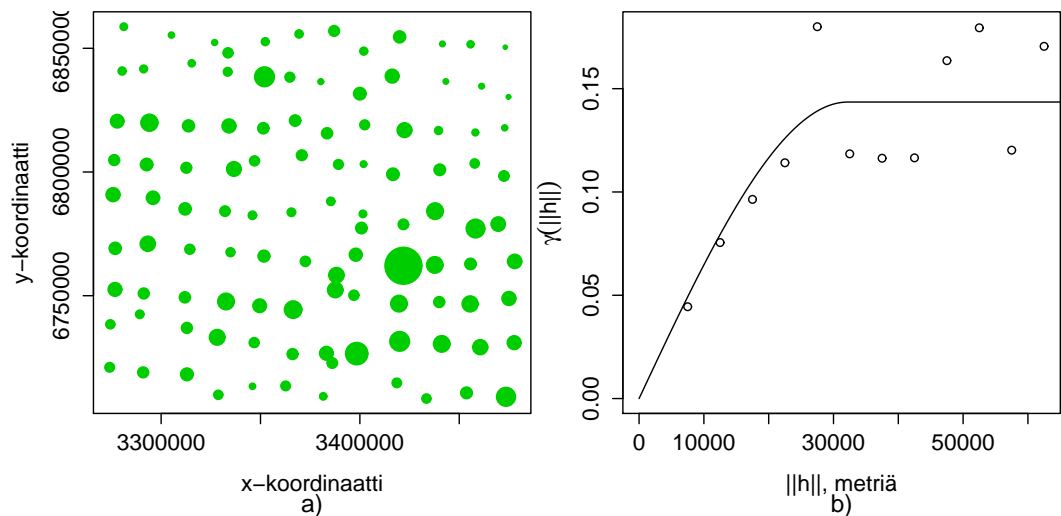
Ennen analyysia tutkimusaineistosta poistetaan sellaiset havaintoyksiköt, jotka sisältävät puuttuvaa tietoa analyysin kannalta välttämättömissä muuttujissa. Radonaineisto rajataan maantieteellisesti Kanta- ja Päijät-Hämeeseen, koska alueella on suoritettu paljon radonmittauksia ja on helpompi tarkastella suhteellisen pientä aluetta. Toisaalta kyseinen alue sijaitsee lähellä Helsinki–Vantaan lentokenttää, josta lämpötila- ja tuulitiedot on mitattu. Lisäksi edellä mainitulla alueella mitattuja asuntoja rajataan muun muassa rakennusteknisten tekijöiden suhteen, jotta saadaan mahdollisimman homogeeninen tutkimusjoukko. Rajaukset tehdään sisältötiedon perusteella. Tutkimukseen otetaan mukaan vain matalaperustaiset omakoti- ja rivitalot, koska rinne- ja kellaritaloissa radonpitoisuudet vaihtelevat sellaisten tekijöiden suhteen, joista ei tämän tutkimuksen aineistossa ole mittaustietoa. Sellaisia asuntoja, jotka saavat talousvetensä porakaivosta, ei hyväksytä mukaan, sillä porakaivoista tulevaa radonin määrää ei voida tämän tutkimuksen aineiston avulla arvioida. Tutkittavien asuntojen perustamistapa on perusmuuri ja maanvarainen lattialaatta, reunajäkistetty laatta tai ryömintätilainen perustus. Edellisten perustamistapojen yhdistelmiä ei sallita. Valmiiseen asuntoon tehtävien korjaustoimenpiteiden vaikutukset riippuvat merkittävästi työn laadusta eikä tällaisia asuntoja myöskään oteta mukaan tutkimukseen. Edellä mainittujen rajausten jälkeen tutkimusaineistossa on 908 havaintoyksikköä alkuperäisen aineiston koon ollessa 13 136 havaintoyksikköä.

5.2 Uraanipitoisuuden interpolointi

Tässä tutkimuksessa on käytössä moreenin ja kallioperän uraanipitoisuusaineistot. Kallioperän uraanipitoisuus näyttää riippuvan hyvin paljon kivilajista, josta näyte on otettu. Variogrammi osoittaa, että kallioperän uraanipitoisuudella ei näytä olevan lainkaan spatiaalista korrelaatiota. Jatkossa keskitymme tarkastelemaan moreenin uraanipitoisuutta.

Moreeni on peräisin jääkaudelta, jolloin mannerjäätikkö irroitti kallioperästä kiviainesta. Jäätikkö kuljetti, kasasi ja kerrosti maa-aineksia, joista syntyi tämä sekalajitteinen maalaji. Koljosen ym. (1992, 14) mukaan Suomessa moreeni ja

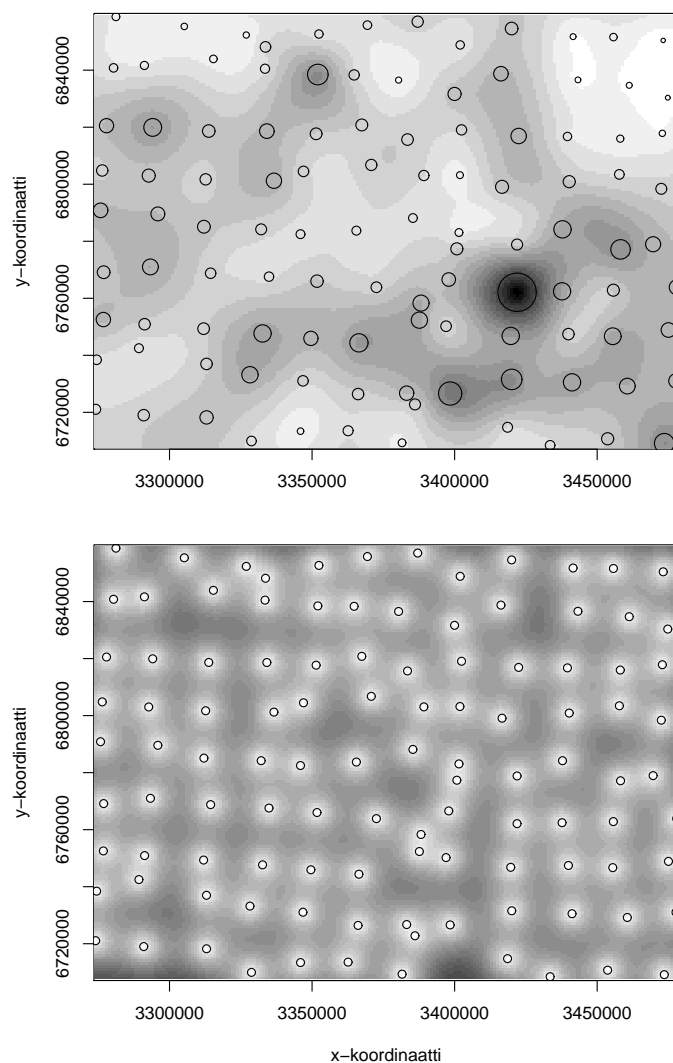
sen hienoaines ovat osoittautuneet parhaiten kallioperää kuvaavaksi ja helposti saatavissa olevaksi aineeksi, koska se on sekoitus paikallisesta ja jonkin matkaa kulkeutuneesta aineksestä. Moreenin uraanipitoisuutta pidetään log-normaalisti jakautuneena (Koljonen ym. 1992, 22). Kuvassa 5a on esitetty moreenin uraanipitoisuus hieman tutkimusaluetta suuremmalla alueella. Kuvassa ympyrän koko on suhteessa uraanipitoisuuden logaritmiin, jota käytetään tämän työn malleissa selittäjänä. Huomataan, että kuvan oikeassa alareunassa olevalla Lahden–Hollolan alueella on suuria uraanipitoisuuksia. Tällä alueella havaitaan myös suuria sisäilman radonpitoisuuksia, joten näiden syiden johdosta mitattuja asuntoja on paljon.



Kuva 5: Moreenin uraanipitoisuus ja sfäärinen semivariogrammisovite

Uraanipitoisuuden logaritmillemme estimoitu suuntaisvariogrammi osoittaa, että aineistossa ei ole epäisotrooppisuutta. Kuvassa 5b on isotrooppinen semivariogrammiestimaattori ja siihen sovitettu sfäärinen semivariogrammimalli. Nugget on kiinnitetty nolaksi osittain siksi, että variogrammiestimaattorista näin voidaan päätellä, vaikka lyhimmän pisteparin etäisyys on 4.9 kilometriä. Säteilyturvakeskuksen tutkijan Heikki Reisbackan (henkilökohtainen tiedonanto 14.2.2008) mukaan voidaan olettaa, että samalta alueelta otetuista näytteistä saadaan hyvin samankaltaisia tuloksia ja siten voidaan nugget olettaa nolaksi. Vaikutusvälin estimaatti on 32.2 kilometriä ja varianssin 0.144.

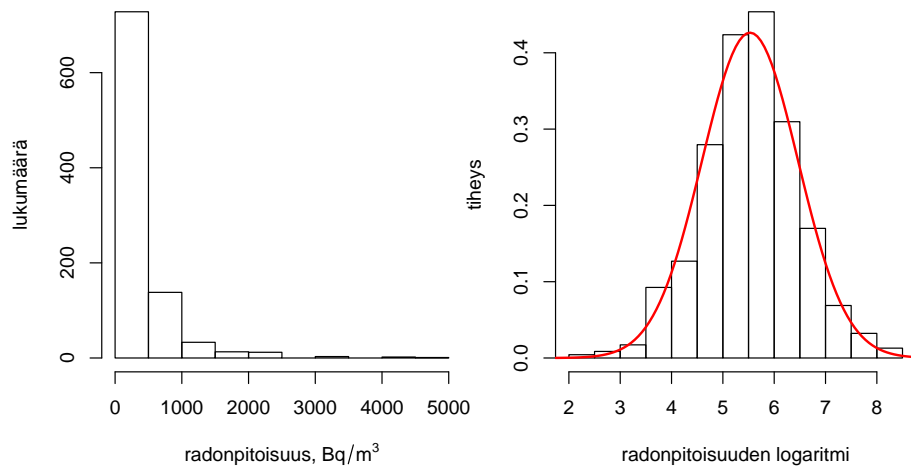
Semivariogrammin sovittamisen jälkeen muodostetaan kriging-kartta tutkimusalueelle. Kuvassa 6 on kriging-ennuste ja -varianssi sekä uraanimittauspisteet. Uraanipitoisuus vaihtelee tutkimusalueella 2.8 ja 48.2 ppm:n välillä. Yksikkö ppm (*parts per million*) tarkoittaa miljoonasosaa eli tässä yhteydessä yhtä milligrammaa uraania kilossa maa-ainesta. Koska nugget on asetettu nolaksi, on huomion arvoista, että interpoloitu pinta kulkee mittaushavaintojen kautta.



Kuva 6: Kriging-ennuste (ylhäällä) ja -varianssi (alhaalla) sekä uranimittauspisteet

5.3 Sekamallin sovittaminen aineistoon

Tarkastellaan aluksi radonpitoisuuden jakaumaa (Kuva 7). Asuntojen sisäilman radonpitoisuus näyttää noudattavan log-normaalijakaumaa ja analyysissa käytetäänkin vastemuuttujana radonpitoisuuden logaritmia. Kuvaan on punaisella piirretty normaalijakauman ($N(5.53, 0.88)$) tiheysfunktio. Tässä työssä tutkittavat selittävät muuttujat ovat Taulukossa 1 esitellyt *harkko*, *huone*, *ilma*, *kanmat*, *maalaji*, *mitker*, *pintaa*, *pertapa*, *ptkerros*, *tdifka*, *torj*, *vv* ja *wspdka* sekä edellisessä alaluvussa interpoloitu *uraanisph*.



Kuva 7: Radonpitoisuuden jakauma

Ennen mallin sovittamista mainittakoon, että perusmuurin materiaalia kuvaava muuttuja *harkko* esiintyy vain perustamistavan luokassa ”A” (perusmuuri ja maanvarainen lattialaatta). Tästä syystä edellä mainittu perustamistavan luokka on jaettu muuttujan *harkko* mukaan kevytsoraharkollisiin ja betonisiin siten, että se saa arvon ”Ahar” harkon tapauksessa ja ”A” betonin.

Jos käytetään klusterointia mittaushavaintojen korreloituneisuuden mallintamisessa, tulisi hyväksikäyttää tietoa maaperän geologiasta ja karttoja, joista maaston muodot ja avonaisuuden pystyisi päättämään. Tälle työlle asetettujen vaatimusten perusteella tyydytään muodostamaan algoritmi, joka kokoaa tutkimusjoukosta taloklustereita tietyn säännön mukaan. Algoritmiin voi antaa maksimietäisyyden, jolta asunnot hyväksytään samaan klusteriin sekä taloklusterin minimikoon. Talot, joita ei edellä mainituilla kriteereillä hyväksytä mihinkään klusteriin, sijoitetaan yhteen ”perusklusteriin”. Esimerkiksi hyvin useita etäisyyksiä ja muutamia klusterikokoja vertailemalla parhaimman sovituksen informaatiokriteerien mielessä antavat 200 metrin maksimietäisyys ja kolmen klusterin minimikoko. Kyseisillä kriteereillä tehty klusterointi tuottaa 64 asuntoklusteria ja lisäksi perusklusterin. Klustereiden koot ovat kolmen ja 53:n asunnon välillä perusklusterin koon ollessa 574 asuntoa. Tällaisen algoritmin puutteena voidaan pitää sitä, että se voi määritellä klusteriksi esimerkiksi pitkän tien varrella sijaitsevat asunnot. Tällöin kauimmaisat asunnot saattavat olla etäällä toisistaan ja korrelaatioiden asuntojen välillä voi olla pieni.

Tutkimusaineisto sisältää sekä jatkuvia että luokittelevia selittäviä muuttujia. Jos kaikki selittävät muuttujat olisivat luokittelevia (faktoreita), niin asetelmamatriisi \mathbf{X} koostuisi nolista ja ykkösistä. Jotta asetelmamatriisista ei tulisi yliparametrisoitu, sen ensimmäiseksi sarakkeeksi asetetaan ykkösvektori $\mathbf{1}_n$ kuvaamaan

jokaisen muuttujan vertailuluokkaa. Tarkastellaan esimerkinomaisesti ilmanvaihtotapaa (*ilma*), jolla on kolme tasoa ("A", "B" ja "C") eli se jakaa aineiston kolmeen ryhmään ja kerrosten lukumäärää (*ptkerros*), jonka tasot ovat "1" ja "2". Valitaan yksi ryhmä vertailuluokaksi muuttujittain, esimerkiksi ilmanvaihtotavan luokka "A" ja kerrosten lukumäärän luokka "1". Tällöin muodostetaan dummy-muuttujat *ilmaB*, *ilmaC* ja *ptkerros2*, jotka kuvaavat muutosta vertailuryhmään. Jos aineistossa on myös jatkuvia muuttujia, niin ne lisätään asetelmamatriisiin dummy-muuttujien lisäksi.

Taulukko 2: Mallin 1 parametrien estimaatit

	estim.	keskiv.	t-arvo	p
vakio	5.032	0.262	19.22	0.0000
ilmanvaihto, kon.poisto	0.050	0.069	0.71	0.4755
ilmanvaihto, kon.tulo-poisto	-0.286	0.072	-3.98	0.0001
kantavien rak. mat., puu	-0.328	0.070	-4.68	0.0000
maalaji, hiekka	-0.252	0.085	-2.96	0.0032
maalaji, kallio	-0.255	0.116	-2.20	0.0280
maalaji, louh.kallio	0.124	0.168	0.74	0.4593
maalaji, moreeni	-0.285	0.089	-3.19	0.0015
maalaji, savi	-0.484	0.094	-5.16	0.0000
per.tapa, perusmuuri(betoni)	-0.204	0.062	-3.29	0.0010
per.tapa, reunaj.laatta	-0.443	0.091	-4.89	0.0000
per.tapa, r.tilainen(sokkeli)	-1.035	0.138	-7.51	0.0000
per.tapa, r.tilainen(kivijalka)	-1.202	0.171	-7.01	0.0000
kerrosten lkm, 2	-0.236	0.065	-3.65	0.0003
lämpötilaero	0.023	0.008	2.83	0.0048
imuputkisto, on	-0.331	0.092	-3.58	0.0004
uraanipitoisuus(log)	0.323	0.058	5.53	0.0000
$\hat{\sigma}_u^2 = 0.142, \hat{\sigma}_\epsilon^2 = 0.561$				

Aluksi sovitaan radonpitoisuuden logaritmille sekamalli, jossa ovat mukana kaikki kovariaatit, asuntoklusterit satunnaisvaikutuksina ja havaintojen kovarianssimatriisilla blokki-diagonaalirakenne ("klusterimalli"). Toisin sanoen mittauspaikkojen välisiä etäisyyksiä käytetään vain klusteroinnissa ja jäännösten korrelaatiomatriisi $\mathbf{R} = \sigma_\epsilon^2 \mathbf{I}_n$. Näin tehdään osittain siitä syystä, että korreloituneiden jäännösten mallin sovittaminen on laskennallisesti hyvin vaativaa. Lopuksi vertaillaan eri kovarianssirakenteita.

Selvästi merkityksettömiksi havaitut mittaushuonetta, mittauserrosta, pinta-alaa ja tuulen nopeutta kuvaavat muuttujat poistetaan. Radonpitoisuus voi kas-

vaa erityisen suureksi sellaisessa tilassa, jossa ilmanvaihto ei toimi. Esimerkiksi saman asunnon suljetussa makuuhuoneessa ja suuressa avoimessa olohuoneessa samanaikaisesti mitatut radonpitoisuudet voivat erota huomattavasti. Ihmisten erilainen huoneiden käytön aktiivisuus johtanee siihen, että mittaushuoneille ei saada yksiselitteisiä vaikutuksia. Asunnon alimmissa kerroksissa mitatut radonpitoisuudet ovat lähes poikkeuksetta suurempia kuin ylemmissä kerroksissa. Tämä johtuu asunnon sisäilman vähäisestä sekoittumisesta kerrosten välillä ja näin ollen mittaukset on ohjeistettu tehtäväksi alimmassa asutussa kerroksessa. Tutkimusaineisto sisältää vain alimman kerroksen mittauksia. Asuinpinta-alaa kuvaavalle muuttujalle ei löydy jatkuvana eikä myöskään luokiteltuna muuttujana tilastollisesti merkitsevää vaikutusta radonpitoisuuteen tämän aineiston perusteella. Tuulen vaikutus puolestaan jää merkityksettömäksi mahdollisesti sen vuoksi, että tuulen voimakkuus on mitattu Helsinki–Vantaan lentokentältä ja mitattavat talot ovat siten kaukana mittauspisteestä.

Taulukko 3: Mallin 2 parametrien estimaatit

	estim.	keskiv.	t-arvo	p
vakio	5.074	0.256	19.83	0.0000
ilmanvaihto, kon. tulo-poisto	-0.305	0.064	-4.77	0.0000
kantavien rak. mat., puu	-0.327	0.070	-4.70	0.0000
maalaji, keskim.	-0.291	0.070	-4.14	0.0000
maalaji, tiivis	-0.507	0.089	-5.69	0.0000
per.tapa, perusmuuri(betoni)	-0.212	0.061	-3.47	0.0005
per.tapa, reunaj. laatta	-0.451	0.090	-5.01	0.0000
per.tapa, r.tilaiset	-1.114	0.114	-9.78	0.0000
kerrosten lkm, 2	-0.242	0.064	-3.77	0.0002
lämpötilaero	0.023	0.008	2.86	0.0043
imuputkisto, on	-0.330	0.092	-3.60	0.0003
uraanipitoisuus(log)	0.325	0.057	5.71	0.0000
$\hat{\sigma}_u^2 = 0.139, \hat{\sigma}_\epsilon^2 = 0.560$				

Merkityksettömien muuttujien poistamisen jälkeen sovitettu Malli 1 tuottaa Taulukon 2 mukaiset estimaatit. Jatkuvien selittäjien, lämpötilaero ja uraanipitoisuus, kertoimet eroavat tilastollisesti merkitsevästi nolasta, kuten myös fakto-
reiden, kantavien rakenteiden materiaali, kerrosten lukumäärä ja radontorjunta, dummy-muuttujien kertoimet. Sen sijaan koneellisen poistoilmanvaihdon kerroin näyttää olevan nolla ($p = 0.48$) eli koneellisen poistoilmanvaihdon ja painovoimaisen ilmanvaihdon vaikutus radonpitoisuuteen on samaa luokkaa. Maalajien hiekka, kallio ja moreeni kertoimet näyttävät olevan samaa tasoa. Louhitun kallion kerroin ei eroa tilastollisesti merkitsevästi nolasta ($p = 0.46$) eli se on samaa tasoa kuin vertailuluokaksi valitulla jäätikkömuodostumilla. Myös ryömintätilais-

ten perustamistapojen vaikutukset radonpitoisuuteen näyttävät samankaltaisilta. Testataan edellämainittuja yhtäsuuruuksia samanaikaisesti lineaarisella hypoteesilla

$$H_0 : \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_{\text{ilmanvaihto, kon. poisto}} \\ \beta_{\text{maalaji, hiekka}} \\ \beta_{\text{maalaji, kallio}} \\ \beta_{\text{maalaji, louh. kallio}} \\ \beta_{\text{maalaji, moreeni}} \\ \beta_{\text{per. tapa, r. tila (sokkeli)}} \\ \beta_{\text{per. tapa, r. tila (kivijalka)}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Leftrightarrow H_0 : \begin{pmatrix} \beta_{\text{ilmanvaihto, kon. poisto}} \\ \beta_{\text{maalaji, hiekka}} - \beta_{\text{maalaji, kallio}} \\ \beta_{\text{maalaji, kallio}} - \beta_{\text{maalaji, moreeni}} \\ \beta_{\text{maalaji, louh. kallio}} \\ \beta_{\text{per. tapa, r. tila (sokkeli)}} - \beta_{\text{per. tapa, r. tila (kivijalka)}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Testi tuottaa testisuureen $F = 0.413$, jolloin vapausastein 5 ja 827 $p = 0.84$. Nollahypoteesi $\mathbf{K}\boldsymbol{\beta} = \mathbf{0}$ saa tukea.

Muodostetaan edellä esitetyn lineaarisen hypoteesin perusteella uudet muuttujat *ilma2*, *pertapa2* ja *maalaji2* seuraavasti:

$$\begin{aligned} ilma2 &= \begin{cases} \text{"A"}, \text{ kun ilmanvaihtotapa on painovoimainen tai kon. poisto,} \\ \text{"C"}, \text{ koneellinen tulo- ja poisto,} \end{cases} \\ pertapa2 &= \begin{cases} \text{"Ahar"}, \text{ kun perustamistapa on perusmuuri kevytsoraharkosta,} \\ \text{"A"}, \text{ perusmuuri betonista,} \\ \text{"B"}, \text{ reunajäkistetty laatta,} \\ \text{"C"}, \text{ ryömintätilainen perustus,} \end{cases} \\ maalaji2 &= \begin{cases} \text{"läpäisevä"}, \text{ kun maalaji on jäätikkömuodostuma tai louh. kallio,} \\ \text{"keskim."}, \text{ hiekka, kallio tai moreeni,} \\ \text{"tiivis"}, \text{ savi.} \end{cases} \end{aligned}$$

Uutta mallia (Malli 2), jossa on uudelleen luokitellut muuttujat, voidaan verrata ensimmäiseen lisäksi uskottavuussuhteen testillä käyttäen estimoinnissa ML-menetelmää, koska mallit ovat sisäkkäiset. Uskottavuussuhteen testi tuottaa testisuureen

$$-2\log\lambda = -2(l_{ML}(\hat{\boldsymbol{\theta}}_{ML,0}) - l_{ML}(\hat{\boldsymbol{\theta}}_{ML})) = -2(-1045.534 + 1044.489) = 2.089.$$

Rajoitteiden määrän ollessa viisi ($r = 5$), $p = 0.84$ eli muuttujien aggregoinnissa ei menetetä oleellista informaatiota tämän testin perusteella.

Taulukko 4: Mallin 3 parametrien estimaatit

	estim.	keskiv.	t-arvo	p
vakio	5.434	0.334	16.29	0.0000
ilmanvaihto, kon. tulo-poisto	-2.005	0.420	-4.78	0.0000
kantavien rak. mat., puu	-0.340	0.069	-4.92	0.0000
maalaji, keskim.	0.014	0.246	0.06	0.9558
maalaji, tiivis	-0.104	0.312	-0.33	0.7394
per.tapa, perusmuuri(betoni)	-0.216	0.061	-3.58	0.0004
per.tapa, reunaj. laatta	-0.454	0.089	-5.08	0.0000
per.tapa, r.tilaiset	-1.125	0.113	-9.93	0.0000
kerrosten lkm, 2	-0.247	0.064	-3.87	0.0001
imuputkisto, on	-0.321	0.091	-3.54	0.0004
ilmanvaihto, pv&k.p:l.tilaero	0.000	0.010	0.03	0.9746
ilmanvaihto, k.t/p:l.tilaero	0.068	0.014	5.01	0.0000
maalaji, läp.:uraanipit.(log)	0.415	0.085	4.86	0.0000
maalaji, kes.:uraanipit.(log)	0.273	0.081	3.37	0.0008
maalaji, tii.:uraanipit.(log)	0.220	0.123	1.80	0.0726
$\hat{\sigma}_u^2 = 0.118, \hat{\sigma}_\epsilon^2 = 0.553$				

Toisen mallin estimaatit on esitetty Taulukossa 3. Kaikkien jatkuvien ja dummy-muuttujien kertoimet ovat tilastollisesti merkitseviä. Seuraavaksi tarkastellaan mallissa olevien muuttujien yhdysvaikutuksia, joita merkitään Taulukoissa 4 ja 5 symbolilla ”:”. Sisällöllisesti mielenkiintoisia ovat ilmanvaihtotavan ja lämpötilaeron sekä maalajin ja uraanipitoisuuden yhdysvaikutukset. Tiedetään muun muassa, että ilmanvaihtotavan, erityisesti painovoimaisen ilmanvaihdon, tehokkuus riippuu sisä- ja ulkoilman lämpötilaerosta. Lisäksi voi olla mahdollista, että maaperän uraanipitoisuus vaikuttaa radonpitoisuuteen erilailla maaperän läpäisevyyden mukaan. Yhdysvaikutuksella tarkoitetaan faktorin ja jatkuvan muuttujan tapauksessa sitä, että jatkuvalle muuttujalle estimoidaan regressiokertoimet faktorin eri tasoilla.

Sovitetaan Malli 3 (Taulukko 4), jossa uudelle ilmanvaihtofaktorille estimoidaan vakiotermit sekä lämpötilan regressiokertoimet kullekin tasolle. Lisäksi uudelle maalajifaktorille estimoidaan vakiotermit ja uraanipitoisuuden regressiokertoimet kullekin tasolle. Taulukosta 4 nähdään, että painovoimaisen ja koneellisen poistoilmanvaihdon asunnoille estimoitu lämpötilaeron regressiokerroin on nolla ($p = 0.97$), kuten myös se, että maalajien vakiotermit näyttävät olevan identtiset.

Taulukko 5: Mallin 4 parametrien estimaatit

	estim.	keskiv.	t-arvo	p
vakio	5.425	0.145	37.53	0.0000
ilmanvaihto, kon. tulo-poisto	-2.003	0.346	-5.79	0.0000
ilmanvaihto, k.t-p:lämpötilaero	0.068	0.014	5.01	0.0000
kantavien rak. mat., puu	-0.340	0.069	-4.93	0.0000
per.tapa, perusmuuri(betoni)	-0.216	0.060	-3.58	0.0004
per.tapa, reunajäyk. laatta	-0.453	0.089	-5.09	0.0000
per.tapa, ryömintätilaiset	-1.123	0.113	-9.94	0.0000
kerrosten lkm, 2	-0.248	0.064	-3.89	0.0001
imuputkisto, on	-0.321	0.090	-3.55	0.0004
maalaji, läpäisevä:uraanipit.(log)	0.421	0.055	7.63	0.0000
maalaji, keskim.:uraanipit.(log)	0.287	0.060	4.77	0.0000
maalaji, tiivis:uraanipit.(log)	0.177	0.064	2.77	0.0057
$\hat{\sigma}_u^2 = 0.117, \hat{\sigma}_\epsilon^2 = 0.551$				

Sovitetaan seuraavaksi Malli 4 (Taulukko 5), jossa vakiotermit $\beta_{maalaji,keskim.}$ ja $\beta_{maalaji,tiivis}$ on asetettu nolaksi ja uuden ilmanvaihtomuuttujan ja lämpötilaeron yhdysvaikutus korvattu uudella lämpötilamuuttujalla *ilmatdifka*. Uusi lämpötilamuuttuja on muodostettu seuraavasti

$$ilmatdifka = \begin{cases} 0 & , \text{ kun ilmanvaihtotapa on painovoimainen tai kon. poisto,} \\ tdifka & , \text{ kun ilmanvaihtotapa on koneellinen tulo- ja poisto.} \end{cases}$$

Malli 4 on nyt Mallin 3 erikoistapaus, joten tehtyjä nollarajoituksia voidaan testata uskottavuussuhteen testillä, kun estimoinnissa käytetään ML-menetelmää. Testisuure on

$$-2\log\lambda = -2(-1035.983 + 1035.896) = 0.175$$

ja rajoitteiden lukumäärä on kolme, joten $p = 0.98$ ja nollarajoitukset jäävät voimaan.

Testataan lopuksi uraanipitoisuuden regressiokertoimia keskimääräisen ja tiiviin maalajin välillä. Lineaarinen hypoteesi kirjoitetaan

$$H_0 : \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_{maalaji,kes.:uraanipit.(log)} \\ \beta_{maalaji,tii.:uraanipit.(log)} \end{pmatrix} = 0$$

$$\Leftrightarrow H_0 : \left(\beta_{maalaji,kes.:uraanipit.(log)} - \beta_{maalaji,tii.:uraanipit.(log)} \right) = 0$$

ja F -testisuureen arvo on $F = 9.133$ (vapausasteet 1 ja 832), joten $p = 0.0026$. Regressiokertoimet eroavat tilastollisesti merkitsevästi toisistaan.

Kaikkien tarkastelun kohteena olleiden faktoreiden yhdysvaikutukset tarkistettiin eivätkä mitkään muut pareittaiset yhdysvaikutukset nousseet tilastollisesti merkitseviksi, joten Mallin 4 sisältämät kiinteät tekijät on syytä sisällyttää lo-pulliseen malliin. Verrataan seuraavaksi erilaisia korrelaatorakenteita. Informaatiokriteereiden mielessä paras sovite saadaan korreloituneiden jäännösten mallilla (Malli 5), jossa jäännöksille sovitetaan sfäärinen korrelaatorakenne. Malli, jossa estimoidaan nugget (c_0) ja vaikutusväli (a_s), tuottaa Taulukon 6 mukaiset estimaatit.

Taulukko 6: Mallin 5 parametrien estimaatit

	estim.	keskiv.	t-arvo	p
vakio	5.297	0.142	37.23	0.0000
ilmanvaihto, kon. tulo-poisto	-2.033	0.348	-5.85	0.0000
ilmanvaihto, k.t-p:lämpötilaero	0.069	0.014	5.06	0.0000
kantavien rak. mat., puu	-0.336	0.072	-4.64	0.0000
per.tapa, perusmuuri(betoni)	-0.207	0.060	-3.46	0.0006
per.tapa, reunajäyk. laatta	-0.462	0.089	-5.20	0.0000
per.tapa, ryömintätilaiset	-1.134	0.115	-9.89	0.0000
kerrosten lkm, 2	-0.247	0.063	-3.88	0.0001
imuputkisto, on	-0.325	0.092	-3.55	0.0004
maalaji, läpäisevä:uraanipit.(log)	0.480	0.066	7.32	0.0000
maalaji, keskim.:uraanipit.(log)	0.346	0.070	4.92	0.0000
maalaji, tiivis:uraanipit.(log)	0.236	0.073	3.22	0.0013
$\hat{\sigma}_\epsilon^2 = 0.595$, $\hat{c}_0 = 0.49$, $\hat{a}_s = 1852$ m				

Informaatiokriteereiden arvot Mallille 4 (AIC = 2148.56, BIC = 2215.7) ovat vain hieman suurempia kuin parhaan korreloituneiden jäännösten mallin (AIC = 2142.9, BIC = 2214.9) ja jälkimmäinen tulkitaan siten paremmaksi informaatiokriteereiden mielessä. Erityisesti BIC:n mukaan malleilla ei ole suurtakaan eroa, joka tarkoittaa sitä, että klusterointi on korreloituneiden jäännösten hyvä approksimaatio. Monimutkaisempi korrelaatorakenne, jossa asuntojen etäisyydet huomioidaan klustereiden sisällä, tuottaa selvästi suurempia informaatiokriteereiden arvoja kuin edelliset.

Edellä mainitut kolme mallia ovat tulkinnaltaan hieman erilaisia. Malli 4 on siinä mielessä hyödyllinen, että klustereiden ennustetut satunnaisvaikutukset antavat informaatiota kyseisen alueen radonpitoisuuksista. Toisaalta sekamalli olettaa, että klustereiden sisällä havainnot korreloivat ja tämä ei ehkä päde peruskluste-

rin osalta, joka on selvästi suurin ryhmä. Edellistä monimutkaisempi malli, jossa jäännösten korrelaatio mallinnetaan klustereiden sisällä, ei toimi tässä tapauksessa hyvin. Syynä voi olla se, että asuntojen välisiä etäisyyksiä hyödynnetään sekä klusteroinnissa että jäännösten korrelaation mallintamisessa, joka aiheuttaa ongelmia mallinnukseen. Tilastollisesti parhaiten sopiva Malli 5 on oletusten osalta kunnossa ja tulkinnaltaan selkeä. Siinä läheisten mittausten korrelaatio mallinnetaan ilman tarvetta muodostaa keinotekoisia asuntoryhmiä.

5.4 Mallin tulkintaa

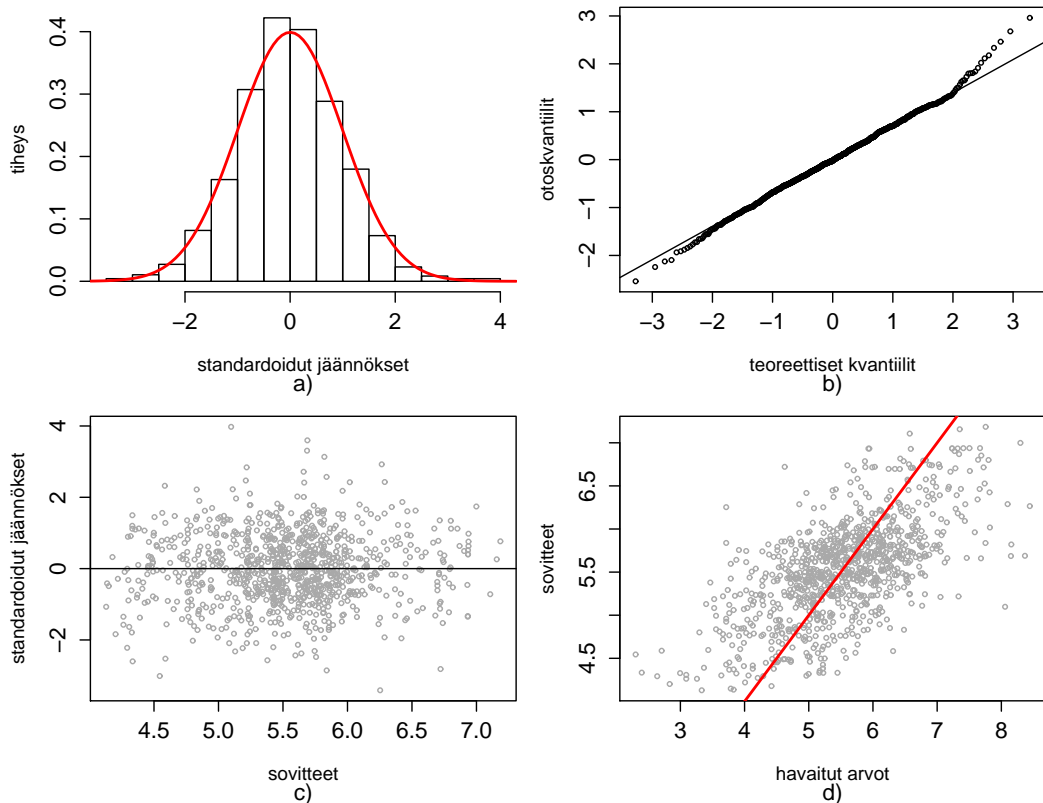
Lopullinen malli on additiivinen vastemuuttujan ollessa radonpitoisuuden luonnollinen logaritmi. Useimpien selittäjien vaikutukset radonpitoisuuteen nähdään tulkinnallisemmin korottamalla kertoimet eksponenttiin (Taulukko 7), jolloin faktoreiden kerrannainen vaikutus nähdään suhteessa vertailuluokkaan. Faktoreiden vertailuluokat on valittu siten, että niissä radonpitoisuudet ovat suurimpia. Toisaalta ne kuvaavat tässä aineistossa yleisimpiä rakennusteknisiä ratkaisuja lukuun ottamatta perustamistapaa, jossa yleisimpänä on betonista rakennettu perusmuuri ja maanvarainen lattialaatta.

Taulukko 7: Mallin 5 β -parametrien estimaatit korotettuna eksponenttiin ja luottamusvälit

	estim.	$e^{\text{estim.}}$	95 %:n luottamusväli
vakio	5.297	199.732	[151.12, 263.98]
ilmanvaihto, kon. tulo-poisto	-2.033	0.131	[0.07, 0.26]
ilmanvaihto, k.t-p:lämpötilaero	0.069	1.072	[1.04, 1.10]
kantavien rak. mat., puu	-0.336	0.715	[0.62, 0.82]
per.tapa, perusmuuri(betoni)	-0.207	0.813	[0.72, 0.91]
per.tapa, reunajäyk. laatta	-0.462	0.630	[0.53, 0.75]
per.tapa, ryömintätilaiset	-1.134	0.322	[0.26, 0.40]
kerrosten lkm, 2	-0.247	0.781	[0.69, 0.88]
imuputkisto, on	-0.325	0.723	[0.60, 0.86]
maalaji, läp.:uraanipit.(log)	0.480		
maalaji, kes.:uraanipit.(log)	0.346		
maalaji, tii.:uraanipit.(log)	0.236		

Esimerkiksi asunnoissa, joissa imuputkisto on asennettu talon perustusten alle, radonpitoisuudet ovat 0.72-kertaisia verrattuna asuntoihin, joista imuputkisto puuttuu. Perustamistavalla näyttää olevan suurin merkitys, sillä ryömintätilaisten perustusten asunnoissa radonpitoisuudet ovat 0.32-kertaisia verrattuna asuntoihin, joiden perustamistapana on kevytsoraharkollinen perusmuuri ja maanvarainen lattialaatta. Muiden faktoreiden estimaatit tulkitaan samalla tavalla.

On huomioitava, että koneellisen tulo- ja poistoilmanvaihdon asunnoissa radonpitoisuudet ovat 0.131-kertaisia vain silloin, kun ulko- ja sisäilman lämpötilaero on nolla astetta. Aineiston pienin lämpötilaero on 15.8 astetta, jolloin radonpitoisuudet koneellisen tulo- ja poistoilmanvaihdon asunnoissa ovat 0.40-kertaisia ($\exp(-2.032 \dots + 0.069 \dots \times 15.8) = 0.391$) vertailuluokkaan nähden.



Kuva 8: Mallin 5 diagnostiikkaa. Kuvassa d suora kuvaa sovitteen ja havainnon yhtäsuuruutta.

Jatkuvan lämpötilaeromuuttujan vaikutus radonpitoisuuteen tulkitaan koneellisen tulo- ja poistoilmanvaihdon asunnoissa siten, että lämpötilaeron kasvaessa b :llä asteella radonpitoisuus muuttuu $\exp(\beta b)$ -kertaiseksi, kun β on lämpötilaeromuuttujan regressiokerroin. Lämpötilaeron kasvaessa yhdellä asteella radonpitoisuus kasvaa siis 7.2 prosenttia eli 1.072-kertaiseksi. Lämpötilaeron kasvaessa puolesta kymmenellä asteella radonpitoisuus kasvaa noin kaksinkertaiseksi ($\exp(0.069 \dots \times 10) = 1.997$). Koska uraanipitoisuus on mallissa logaritmoituna, sen vaikutus radonpitoisuuteen tulkitaan eri tavalla kuin lämpötilaeron. Uraanipitoisuuden kasvaessa c %:lla radonpitoisuus muuttuu $(1 + \frac{c}{100})^\beta$ -kertaiseksi, kun β on logaritmoidun uraanipitoisuusmuuttujan regressiokerroin. Uraanipitoisuuden kasvaessa esimerkiksi 100 %:lla radonpitoisuus kasvaa läpäiseville maalajeille perustetuissa asunnoissa 1.40-kertaiseksi ($2^{0.480 \dots} = 1.40$), keskimääräisille perus-

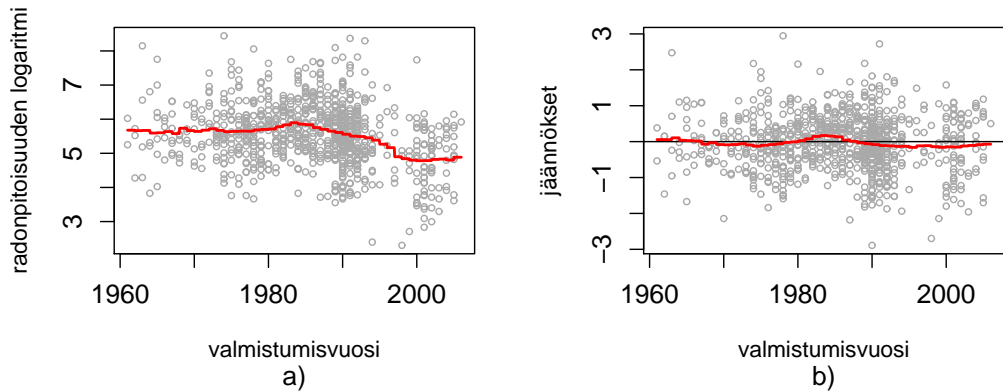
tetuissa 1.27-kertaiseksi ja tiiviille perustetuissa 1.18-kertaiseksi.

Kuvasta 8a nähdään, että jäännökset ovat likimain normaalisti jakautuneita. Kuva 8b osoittaa, että malli tuottaa joitakin erityisen suuria jäännöksiä. Suurimpiin jäännöksiin liittyvien havaintojen poistaminen ei vaikuta juurikaan estimaatteihin eikä keskivirheisiin, joten havainnot pidetään mukana tutkimusaineistossa. Jäännöksissä ei ole havaittavissa heteroskedastisuutta, mikä näkyy Kuvasta 8c, vaan ne ovat jakautuneet melko tasaisesti sovitteiden arvoalueelle. Kuvasta 8d havaitaan, että malli tuottaa konservatiivisen sovitteen eli suurilla radonpitoisuuksilla sovitteet ovat pienempiä ja pienillä pitoisuuksilla vastaavasti suurempia kuin havaittu radonpitoisuus. Kuvassa oleva punainen viiva kuvaa havainnon ja sovitteen yhtäsuuruutta. Tällainen sovitteiden käyttäytyminen johtuu siitä, että radonpitoisuudessa on suurta satunnaisvaihtelua.

5.5 Radonpitoisuuksissa tapahtuneet muutokset asunnon valmistumisvuoden mukaan

Sisällöllisesti erityisen kiinnostavaa on radonpitoisuuksissa tapahtunut lasku, kun tietoisuus radonista on lisääntynyt (Kuva 9a). Erityisen kiinnostavaa on se, onko lasku selitettävissä ilmanvaihdon tehostumisella (koneellinen tulo- ja poistoilmanvaihto) ja radonputkistojen asennuksella rakennusten perustuksiin vai voiko syynä edellisten lisäksi olla jokin rakennusmateriaaleihin tai -tekniikkaan liittyvä kehitys. Kuvasta nähdään selvästi kuinka radonpitoisuudet ovat kääntyneet laskuun 80-luvun lopulta lähtien ja ovat pysyneet tasaisena 90-luvun lopulta alkaen. Kuvissa 9a ja 9b punainen viiva on eräänlainen lokaali keskiarvo. Kuvassa 9b on Mallin 4 tuottamat jäännökset suhteessa valmistumisvuoteen. Lokaalin keskiarvon avulla nähdään, että 80-luvun puolessa välissä jäännökset ovat hieman keskimääräistä suurempia ja 90-luvun loppupuolelta eteenpäin pienempiä. Voitaneen kuitenkin sanoa, että käytössä olevat kovariaatit selittävät melko hyvin radonpitoisuuksissa viimeisen kahdenkymmenen vuoden aikana tapahtuneita muutoksia.

Sisällöllisesti tulee vielä korostaa malleissa olevan radontorjuntamuuttujan moninaista vaikutusta asuntojen radonpitoisuuksiin. Radonpitoisuuksien on havaittu olevan alhaisempia asunnoissa, joiden perustuksissa on passiivinen imuputkisto, eli putkistoon liitetty imuri ei ole käynnissä. Jos putkistoon liitetty imuri on käynnissä, niin radonpitoisuuksien on havaittu alenevan entisestään. Toisaalta, jos talon rakennusvaiheessa perustuksiin on asennettu imuputkisto, on radonilta mahdollisesti suojauduttu muutenkin, esimerkiksi perusrakenteita tiivistämällä. Tutkimusaineistossa on aktiivisen imurin asuntoja yhdeksän ja passiivisen putkiston asuntoja 124.



Kuva 9: Radonpitoisuus ja Mallin 5 jäännökset valmistumisvuoden suhteen

5.6 Asunnon radonpitoisuuden ennustaminen

Tässä alaluvussa ennustetaan radonpitoisuus kolmeen asuntoon. Ennustaminen suoritetaan klusterimallilla (Malli 4) ja korreloituneiden jäännösten mallilla (Malli 5), jolloin malleja voidaan vertailla keskenään myös ennusteiden suhteen. Ennusteet saadaan piste-estimaatteina logaritmiasteikolla. Ennustevariansseille, joiden avulla voidaan muodostaa 95 %:n ennustevalit, esitellään analyttisesti johdetut kaavat.

Tarkastellaan ensin ennustamista Mallilla 4. Uudelle havainnolle klusterissa j saadaan ennuste \hat{y}^* estimaattien $\hat{\beta}$ ja ennustettujen satunnaisvaikutusten \hat{u} avulla. Ennuste voidaan kirjoittaa

$$\hat{y}^* = \mathbf{x}^* \hat{\beta} + \mathbf{z}_j \hat{u},$$

missä \mathbf{x}^* on ennustettavan asunnon (mittauksen) kovariaattien arvot sisältävä $1 \times p$ vaakavektori ja \mathbf{z}_j on asunnon klusterin osoittava $1 \times c$ vaakavektori. Voneshin ja Chinchillinin (1997, 252–253) mukaan ennustevirheen varianssi voidaan kirjoittaa

$$(14) \quad \begin{aligned} \sigma_{*4}^2 &= \text{Var}(\hat{y}^* - y | \mathbf{y}) \\ &= \mathbf{x}^* \text{Var}(\hat{\beta} - \beta) \mathbf{x}^{*'} + \mathbf{z}_j \boldsymbol{\Omega}_j \mathbf{z}_j' + \mathbf{x}^* \boldsymbol{\Psi}_j \mathbf{z}_j' + \mathbf{z}_j \boldsymbol{\Psi}_j \mathbf{x}^{*'} + \sigma_\epsilon^2, \end{aligned}$$

missä

$$\text{Var}(\hat{\beta} - \beta) = [\mathbf{X}'\mathbf{W}\mathbf{X}]^{-1},$$

ennustevirheen $\hat{u}_j - u_j$ varianssi on

$$\begin{aligned} \boldsymbol{\Omega}_j &= \text{Var}(\hat{u}_j - u_j) \\ &= \mathbf{D} - \mathbf{D}\mathbf{Z}_j' \mathbf{W}_j \mathbf{Z}_j \mathbf{D} \\ &\quad + \mathbf{D}\mathbf{Z}_j' \mathbf{W}_j \mathbf{X}_j [\mathbf{X}'\mathbf{W}\mathbf{X}]^{-1} \mathbf{X}_j' \mathbf{W}_j \mathbf{Z}_j' \mathbf{D} \end{aligned}$$

ja

$$\begin{aligned}\Psi_j &= \text{Cov}[(\hat{\beta} - \beta), (\hat{u}_j - u_j)] \\ &= -\text{Var}(\hat{\beta} - \beta)\mathbf{X}'_j\mathbf{W}_j\mathbf{Z}_j\mathbf{D}.\end{aligned}$$

Kaavoissa esiintyvä matriisi \mathbf{Z}_j on klusterin j asuntoihin liittyvä $n_j \times c$ asetelmamatriisi. Kaavassa (14) oletetaan, että varianssikomponentit $\alpha = (\sigma_u^2, \sigma_\epsilon^2)$ tunnetaan. Käytännössä α korvataan aineistosta lasketulla estimaatillaan $\hat{\alpha}$, jolloin approksimatiivinen 95 %:n ennusteväli on

$$\hat{y}_j^* \pm t_{c-p, 975} \hat{\sigma}_{*4}.$$

Korreloituneiden jäännösten mallin (Malli 5) tilanteessa ennusteet voidaan kirjoittaa

$$\hat{y}^* = \mathbf{x}^* \hat{\beta} + \hat{\epsilon}^*,$$

missä $\hat{\epsilon}^*$ on interpoloitu jäännöksistä kriging-menetelmällä. Interpolointi on mielekästä, sillä lähekkäisiin havaintoihin liittyvät jäännökset korreloivat. Ennusteen varianssi koostuu tällöin regressiokertoimien estimointiin liittyvästä epävarmuudesta ja kriging-variانسista

$$\begin{aligned}\sigma_{*5}^2 &= \text{Var}(\hat{y} - y|\mathbf{y}) = \mathbf{x}^* \text{Var}(\hat{\beta} - \beta) \mathbf{x}^{*'} + \text{Var}(\hat{\epsilon} - \epsilon) \\ &= \mathbf{x}^* [\mathbf{X}'\mathbf{W}\mathbf{X}]^{-1} \mathbf{x}^{*'} + \sigma_k^2.\end{aligned}$$

Olettamalla vastemuuttuja y normaalijakautuneeksi, saadaan approksimatiiviseksi 95 %:n ennusteväliksi

$$\hat{y}^* \pm 1.96 \hat{\sigma}_{*5}.$$

Molemmat ennusteet ottavat huomioon aineistossa esiintyvän havaintojen korreloituneisuuden. Klusterimallissa käytetään hyväksi asuntoklusterille j ennustettua satunnaisvaikutusta \hat{u}_j . Korreloituneiden jäännösten mallissa puolestaan hyödynnetään kovarianssimatriisia $\hat{\mathbf{R}}$, joka sisältää informaatiota jäännöksistä estimoiduista nuggetista ja vaikutusvälistä.

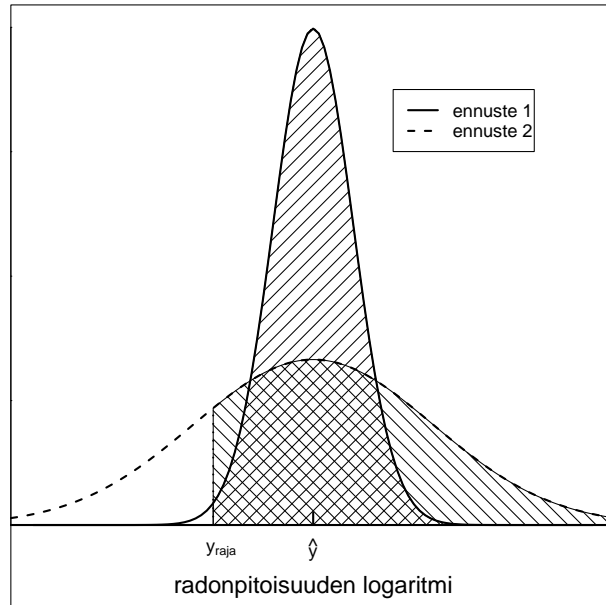
Taulukossa 8 esitellään kolmen tutkimusalueelta valitun asunnon muuttujien arvoja ja ennustamisen tuloksia. Asunnot on valittu siten, että kovariaattien arvot poikkeavat toisistaan ja mitatut radonpitoisuudet ovat selvästi erisuuruiset. Piste-estimaatit ja ennustevälit on tuotettu kunkin ennusteen kohdalla siten, että mallinnuksessa käytetty aineisto ei ole sisältänyt kyseistä havaintoa. Näin on saatu todellinen ennustamistilanne, jolloin parametriestimaatit ja ennuste eivät korreloi.

Taulukko 8: Tietoja ja estimaatteja liittyen kolmeen valittuun asuntoon, joihin radonpitoisuus ennustetaan Malleilla 4 ja 5

	asunto 1	asunto 2	asunto 3
ilmanvaihto	k. poisto	k. tulo-poisto	k. poisto
lämpötilaero	30.2 °C	20.6 °C	25.9 °C
kantavien rak. mat.	min. pohj.	puu	puu
perustamistapa	perusm.(har.)	perusm.(har.)	perusm.(har.)
kerrosten lkm	1	2	1
imuputkisto	ei	on	on
maalaji	jäätikkom.	moreeni	jäätikkom.
uraanipitoisuus (log)	26.1 ppm	4.6 ppm	35.6 ppm
klusteri	peruskl.	peruskl.	3
hav. radonpitoisuus	617 Bq/m ³	148 Bq/m ³	2259 Bq/m ³
piste-estimaatti (Malli 4)	880 Bq/m ³	76 Bq/m ³	890 Bq/m ³
appr. 95 % ennusteväli	[195, 3962]	[17, 343]	[169, 4691]
piste-estimaatti (Malli 5)	815 Bq/m ³	69 Bq/m ³	558 Bq/m ³
appr. 95 % ennusteväli	[387, 1713]	[32, 151]	[264, 1181]

Approksimatiiviset ennustevälit molemmissa malleissa ovat erityisesti ensimmäisen ja kolmannen ennusteen osalta hyvin leveät, joka johtuu radonpitoisuuden suuresta satunnaisvaihtelusta. Ensimmäisen ja toisen asunnon osalta ennusteet ovat melko lähellä havaittuja arvoja ottaen huomioon radonpitoisuuden satunnaisvaihtelun. Havaitut arvot ovat approksimatiivisten ennustevälien sisällä lukuun ottamatta Mallin 5 ennusteväliä kolmannelle asunnolle. On odotettua, että kolmannen asunnon ennuste jää selvästi havaitun radonpitoisuuden alapuolelle, sillä se edustaa hyvin korkeaa radonpitoisuutta. Kuva 8d havainnollistaa hyvin radonpitoisuuden satunnaisvaihtelun suuruutta sekä sovitteiden ja havaittujen arvojen yhteyttä. Kuten estimoidun mallin sovitteet, niin myös ennusteet ovat konservatiivisia, mikä näkyy kolmannen asunnon ennusteen alhaisessa tasossa.

Selvästi suurin osa ennustevirheen varianssista tulee jäännösvarienssista σ_ϵ^2 tai kriging-varienssista σ_k^2 . Esimerkiksi kolmannen ennusteen ennustevirheen varianssin estimaatti Mallissa 4 on $\hat{\sigma}_{*4}^2 = 0.69$, josta jäännösvarienssin osuus on 80 %. Mallissa 5 ennustevirheen varianssin estimaatti on $\hat{\sigma}_{*4}^2 = 0.56$, josta kriging-varienssin osuus on 96 %. Klusterimalli tuottaa siis suuremman ennustevarianssin ja siten leveämmän ennustevälin, joka johtunee satunnaisvaikutusten ennustamiseen liittyvästä epävarmuudesta.



Kuva 10: Ennusteiden jakaumat suhteessa raja-arvoon. Viivoitettu alue kuvaa todennäköisyyttä sille, että radonpitoisuuden vuosikeskiarvo ylittää enimmäisarvon.

Ennusteita tulkittaessa täytyy muistaa, että ne ovat marras-huhtikuussa tehtävien kahden kuukauden mittauksen ennusteita. Sosiaali- ja terveysministeriön päätöksen mukaiset enimmäisarvot asuntojen sisäilman radonpitoisuudelle koskevat kuitenkin koko vuoden keskimääräistä radonpitoisuutta. Säteilyturvakeskuksessa määritetyn mallin avulla kahden kuukauden mittaus voidaan laajentaa kuvaamaan asunnon radonpitoisuuden vuosikeskiarvoa, joten myös kahden kuukauden mittauksen ennuste voidaan laajentaa vuosikeskiarvon ennusteeksi. Edellä mainittu muunnosmalli ei ole käytössä tässä tutkimuksessa. Koska piste-estimaatin, ennustevarianssin ja muunnosmallin avulla vuosikeskiarvon ennusteen jakauma tunnetaan, voitaisiin laskea todennäköisyys sille, että asunnon radonpitoisuuden vuosikeskiarvo ylittää asetetun enimmäisarvon (Kuva 10). Kuvassa kahden ennusteen piste-estimaatit ovat samat, mutta ennustevarianssit eroavat. Ennuste 1 on tarkempi kuin ennuste 2, jolloin todennäköisyys enimmäisarvon (y_{raja}) ylitykselle on suurempi ensimmäisen ennusteen kohdalla.

6 Yhteenveto

Tämän pro gradu -työn tarkoituksena oli tarkastella rakennusteknisten, geologisten sekä ilmastollisten tekijöiden yhteyttä matalaperustaisten asuntojen sisäilman radonpitoisuuden kanssa Kanta- ja Päijät-Hämeen alueella. Työssä esiteltiin spatiaalisen uraanimuuttujan ennustaminen mitattujen talojen kohdalle ja sekamallin sovittaminen aineistoon, jossa lähekkäisten asuntojen radonpitoisuuksien oletettiin korreloivan. Työssä mallinnettiin havaintojen korreloituneisuutta erilaisilla kovarianssirakenteilla, joita myös vertailtiin keskenään. Aikaisempiin Suomessa tehtyihin tutkimuksiin verrattuna uuden lähestymistavan toi sekamalli, jolla huomioitiin lähekkäisiin asuntoihin samankaltaisesti vaikuttavat paikalliset geologiset ja ilmastolliset tekijät.

Sisä- ja ulkoilman lämpötilaeron vaikutus radonpitoisuuteen riippuu asunnon ilmanvaihtoratkaisusta. Mallin rakentamisen yhteydessä havaittiin, että lämpötilaero näkyi vain koneellisen tulo- ja poistoilmanvaihdon asunnoissa. Jostain syystä painovoimaisen ja koneellisen poistoilmanvaihdon asunnoissa lämpötilaero ei näyttänyt vaikuttavan radonpitoisuuteen. Arvela ym. (1988) osoittavat fysikaalisen mallilaskennan antavan tulokseksi sen, että lämpötilaeron kasvaessa radonpitoisuus kasvaa. Mallilaskenta siis huomioi sekä ilmanvaihdon että radonin kulkeutumisen lämpötilariippuvuuden. Painovoimaisen ja koneellisen poistoilmanvaihdon osalta tämän tutkimuksen tulos poikkeaa aikaisemmista fysikaalisiin malleihin pohjautuvista tutkimuksista.

Tutkimusalueella moreenin uraanipitoisuus vaihtelee 2.8 ja 48.2 ppm:n välillä. Koska radonia syntyy radioaktiivisen uraanin hajotessa, on ilmeistä, että uraanipitoisuuden vaikutus sisäilman radonpitoisuuteen on tilastollisesti merkitsevä. Tilastollisen mallinrakennuksen perusteella päädyttiin malliin, jossa maalajien vakiotermit ovat identtiset ja uraanipitoisuuden regressiokerroin vaihtelee maalajin läpäisevyyden mukaan. Vaihtoehtona edelliselle oli uraanipitoisuuden yksi regressiokerroin ja vakiotermien estimointi maalajeille. Valinta näiden kahden mallin välillä tehtiin informaatiokriteereiden perusteella. Lisäksi on otettava huomioon, että uraanipitoisuus jokaisen talon kohdalle on interpoloitu ja se sisältää siten satunnaisvirhettä. Tämä aiheuttanee jonkin verran harhaa uraanipitoisuuden regressiokertoimiin.

Ennen tätä tutkimusta oli ennakkokäsityksiä maalajin läpäisevyyden luokittelusta. Kallion ajateltiin olevan suhteellisen tiivis maalaji ja sen ei oletettu poikkeavan kovinkaan suuresti louhituista kallioista. Tämän aineiston ja analyysin perusteella voidaan kuitenkin sanoa, että louhitun kallion läpäisevyys on lähempänä jäätikkömuodostumien kuin kallion läpäisevyyttä. Toisaalta kallion läpäisevyys eroaa tilastollisesti merkitsevästi tiiviinä pidetyn saven läpäisevyydestä. On kuitenkin huomioitava, että läpäiseville kallioille perustettuja asuntoja on aineistossa vain 26 ja kalliolle perustettuja 69, kun muille maalajeille perustettujen asuntojen määrät vaihtelevat 149:n ja 258:n välillä.

Tutkimuksessa käytetyt ilmastokovariaatit, mittausajan lämpötila- ja tuulikeskiarvot, on saatu Helsinki–Vantaan lentoaseman sääasemalta. Etäisyydet sääasemalta mittausasuntoihin vaihtelevat 40 kilometrin ja 160 kilometrin välillä. Koska ilmastokovariaatteja on mahdollista saada useilta sääasemilta, tulevaisuudessa näitä tarkempia tietoja lienee syytä käyttää. Esimerkiksi sovitettaessa Mallin 2 kiinteät selittäjät sisältävä lineaarinen regressiomalli Etelä-Karjalan alueelle, lämpötilaeron estimaatiksi saadaan tilastollisesti merkitsevä negatiivinen kerroin. Tämä tulkitaan siten, että lämpötilaeron kasvaessa radonpitoisuudet pienenevät. Paremmin ilmiötä kuvaaviin estimaatteihin päästäneen käyttämällä mahdollisimman läheltä asuntoja mitattuja säätietoja.

Tutkimusaineistossa on joitakin havaintoyksiköitä, joissa rakennuksen sijaintikoordinaatit ovat virheelliset. Sijainti voi näissä tapauksissa poiketa joitakin satoja metrejä, jolloin on mahdollista, että maalaji on virheellisesti määritelty. Tämä aiheuttaa epätarkkuutta estimointituloksiin. Ongelma esiintyy joissakin vanhemmissa mittauksissa, joita ei voida yksilöidä.

Luontevana jatkotutkimuksena voisi olla tutkimusalueen laajentaminen Kanta- ja Päijät-Hämettä suuremmalle alueelle. Tällöin olisi välttämätöntä saada käyttöön tarkempia ilmastokovariaatteja. On todennäköistä, että joidenkin selittäjien vaikutukset muuttuvat uuden alueen geologian, maaston ja ilmaston johdosta, jolloin tämän tutkimuksen mallia joudutaan modifioimaan. Matalaperustaisten talojen lisäksi olisi syytä ottaa tarkastelun kohteiksi myös kellari- ja rinnetalot.

Mikäli klusterimallia halutaan jatkossa hyödyntää, tulisi ottaa tarkemmin kantaa läheisten asuntojen määrittämiseen ja vähintäänkin harkita maaston muotojen ottamista huomioon. Työssä esiteltyyn algoritmiin voitaisiin lisätä klustereiden kauimmaisten asuntojen etäisyyttä rajoittava ehto. Tällöin esimerkiksi tien varrelle määritetty klusteri, jonka kauimmaisten asuntojen välinen etäisyys on suuri, jaettaisiin osiin.

Viitteet

- Arvela, H. 1995a. *Residential Radon in Finland: Sources, Variation, Modelling and Dose Comparisons*. STUK-A124. Helsinki: Säteilyturvakeskus.
- Arvela, H. 1995b. Seasonal variation in radon concentration of 3000 dwellings with model comparisons. *Radiation Protection Dosimetry*, **59** (1), 33–42.
- Arvela, H., Mäkeläinen, I. & Castrén, O. 1993. *Otantatutkimus asuntojen radonista Suomessa*. STUK-A108. Helsinki: Säteilyturvakeskus.
- Arvela, H., Voutilainen, A., Mäkeläinen, I., Castrén, O. & Winqvist, K. 1988. Comparison of predicted and measured variations of indoor radon concentration. *Radiation Protection Dosimetry*, **24** (1), 231–235.
- Arvela, H. & Winqvist, K. 1989. A model for indoor radon variations. *Environment International*, **15**, 239–246.
- Bailey, T. C. & Gatrell, A. C. 1995. *Interactive Spatial Data Analysis*. Essex: Longman.
- Cressie, N. A. C. 1991. *Statistics for Spatial Data*. New York: Wiley.
- Harville, D. A. 1974. Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.
- Henderson, C. R., Kempthorne, O., Searle, S. R. & Von Krosig, C. N. 1959. Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15**, 192–218.
- Koljonen, T., Gustavsson, N., Noras, P. & Tanskanen, H. 1992. Näytteet, kemialliset analyysit ja tulosten käsittely. Teoksessa T. Koljonen (toim.) *Suomen geokemian atlas, osa2: moreeni*. Espoo: Geologian tutkimuskeskus, 14–27.
- McCulloch, C. E. & Searle, S. R. 2001. *Generalized, Linear, and Mixed Models*. New York: Wiley.
- Pinheiro, J. C. & Bates, D. M. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Rasilainen, K., Lahtinen, R. & Bornhorst, T. J. 2007. The Rock Geochemical Database of Finland Manual. *Geological Survey of Finland, Report of Investigation*, 164.
- STMP 1.11.1992/944. Sosiaali- ja terveysministeriön päätös asuntojen huoneilman radonpitoisuuden enimmäisarvoista.

Säteilyturvakeskus. Radontalkoot. Viitattu 18.4.2008.

http://www.stuk.fi/sateilytietoa/sateily_ymparistossa/radon/radonkampanjat

Verbeke, G. & Molenberghs, G. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Vonesh, E. F. & Chinchilli, V. M. 1997. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

Weltner, A., Arvela, H., Turtiainen, T., Mäkeläinen, I. & Valmari, T. 2003. Radon sisäilmassa. Teoksessa R. Pöllänen (toim.) *Säteily ympäristössä*. Helsinki: Säteilyturvakeskus, 111–161.



PL 14, 00881 HELSINKI
Puh. (09) 759881

STUK täyttää

--	--	--	--

Etinimi _____

Sukunimi _____

Lähiosoite _____

Postinumero _____ Postitoimipaikka _____

Kunta _____

Puhelinnumero _____ - _____

1. Purkin sijainti asunnossa

Huone Kerros (1=maantaso)
0 1 2 3

Olohuone

Makuuhuone

Muu, mikä _____

1. Purkin numero

S	
T	
U	
K	
t	
ä	
y	
t	
t	
ä	

2. Purkin sijainti asunnossa

Huone Kerros (1=maantaso)
0 1 2 3

Olohuone

Makuuhuone

Muu, mikä _____

2. Purkin numero

t	
ä	
y	
t	
t	
ä	

Aloituspäivämäärä (pp.kk.vv)

--	--	--	--

Lopetuspäivämäärä (pp.kk.vv)

--	--	--	--

Onko tämä ensimmäinen radonmittaus asunnossanne ?

 Kyllä Ei

Aiemman mittauspurkin numero tai mittauksen ajankohta, jos tiedossa _____

Asukkaiden lukumäärä _____ Nykyiset asukkaat asuneet vuodesta _____

Kiinteistötunnus (katso kiinteistöverolipusta tai kysy isännöitsijältä)
 _____ - _____ - _____ - _____ - _____

TALO

Talotyyppi

- Omakotitalo Vapaa-ajan asunto
- Rivi- tai paritalo Muu, mikä _____
- Kerrostalo _____

Valmistumisvuosi

--	--	--	--

Perustamispaikka

Onko talo perustettu kalliolle?

 Ei Kyllä Kalliota louhittu

Lisätietoja: esim. rakennuspaikan maalajit ja täytemaan paksuus (jos tiedossa) _____

Saadaanko talousvesi porakaivosta?

 Ei Kyllä

Porakaivoveden radonpitoisuus on tutkittu

 Kyllä

Asuinhuoneiden lattioiden pintamateriaali

(valitse yksi tai useampia vaihtoehtoja)

 Muovimatto Lautalattia

 Parketti Muu, mikä _____

Pääasiallinen julkisivumateriaali (vain yksi rasti)

 Puu Kevytbetoni

 Tiili En tiedä

 Betoni Muu, mikä _____

Kantavien rakenteiden pääasiallinen rakennusmateriaali

(vain yksi rasti)

 Puu Kevytbetoni

 Tiili En tiedä

 Betoni Muu, mikä _____

Lämmitystapa (merkitse ensisijainen 1:llä ja toissijainen 2:lla)

 Vesikeskuslämmitys

 Uunilämmitys

 Suora sähkölämmitys

 Muu, mikä _____

 Ilmakeskuslämmitys (kiertoilmalämmitys)

Raitisilmaventtiilien lukumäärä _____ kpl

 Venttiilit olivat mittauksen aikana auki Kyllä Ei

Lämmönlähde (merkitse ensisijainen 1:llä ja toissijainen 2:lla)

 Kauko- ja aluelämpö

 Sähkö

 Polttoöljy

 En tiedä

 Puu tai turve

 Muu, mikä _____

 Kaasu _____

Ilmanvaihto

 Painovoimainen eli luonnollinen

Liesituuletin toiminnassa

--	--

tuntia vuoro-

 Koneellinen poistoilmajärjestelmä, toiminnassa

--	--

tuntia vuoro-

 Koneellinen tulo- ja poistoilmajärjestelmä, toiminnassa

--	--

tuntia vuoro-

 En tiedä

--	--

tuntia vuoro-

Ilmanvaihdon tehokkuus on mielestäni

 Hyvä Kohtuullinen Huono

TUULETUSTAVAT

Asunnossa tuuletetaan lämmityskauden aikana (loka-toukokuu)

avoimen ikkunan kautta (1 tai 2 rastia)

 Satunnaisesti

 Lähes päivittäin yli 30 min

 Lähes päivittäin alle 10 min

 Öisin

 Lähes päivittäin 10 - 30 min

 Jatkuvasti

RAKENNUSVAIHEEN RADONTORJUNTA

Suoritetut toimenpiteet: (1 tai useampia rasteja)

 Perusrakenteita on tiivistetty Muu toimenpide

 Laatan alle on rakennettu imuputkisto Ei toimenpiteitä

 Imuputkisto ja imuri toiminnassa

RADONKORJAUSTOIMENPITEET

Onko asunnossanne tehty korjauksia radonpitoisuuden alentamiseksi?

Lyhyt kuvaus toimenpiteestä kohtaan lisäselvityksiä.

 Ei Kyllä

KERROSTALOASUNTO (täytetään vain jos kyseessä on kerrostaloasunto)

Asunto _____ kerroksessa (maan tasalla oleva kerros on 1. kerros)

Talossa on _____ kerrosta (kellaria ei huomioida)

Onko talossa kellarikerros

 Ei Kyllä

 Asuinpinta-ala on _____ m²

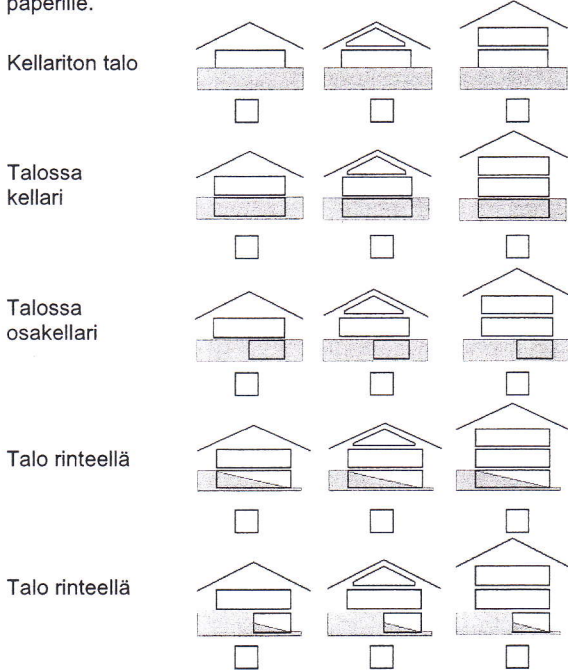
Asunnon alapuolella on

 Toinen asunto tai muita tiloja Maa tai kallio

PIENTALOASUNTO

Asuinpinta-ala on m²

Pientalotyyppi Laita rasti taloasi parhaiten kuvaavan kuvan alla olevaan ruutuun. Voitte myös piirtää selvittäviä kuvia kohtaan lisäselvityksiä tai erilliselle paperille.



KELLARIKUOPPA (käynti luukun kautta)

- Talon alla ei ole kellarikuoppaa
 Talon alla on kellarikuoppa, jonka pinta-ala on m²

Kellarikuopan lattiamateriaali

- Betoni
 Maa tai kallio
 Muu, mikä

Kellarikuopan seinien materiaali

- Betoni
 Kevytsoraharkko
 Muu, mikä

KELLARI, OSAKELLARI TAI RINNETALON ALIN KERROS

Alimman kerroksen käyttötarkoitus

(merkitse kaikki sopivat vaihtoehdot)

- | | |
|------------------------------------------------|-------------------------------------|
| <input type="checkbox"/> Lämmityslaitteet | <input type="checkbox"/> Makuuhuone |
| <input type="checkbox"/> Autotalli | <input type="checkbox"/> Työhuone |
| <input type="checkbox"/> Talouskellari/varasto | <input type="checkbox"/> Takkahuone |
| <input type="checkbox"/> Sauna | <input type="checkbox"/> Muu, mikä |
| <input type="checkbox"/> Pesutilat | <input type="text"/> |

Alimman kerroksen tai kellarin pinta-ala on m²

Kulkuyhteys alimman kerroksen ja muun asunnon välillä

- Avoin portaikko Portaat ja ovi Vain ulkokautta

Alimman kerroksen tai kellarin lattiamateriaali

- Betoni Maa tai kallio Muu, mikä

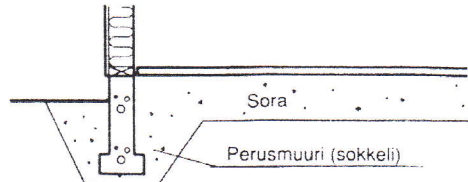
Maanvastaisten seinien materiaali

- Betoni Kevytsoraharkko Muu, mikä

PIENTALON PERUSTAMISTAPA

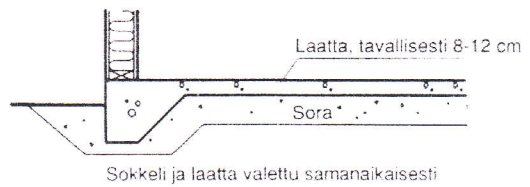
Kellarittoman talon tai kellarittoman talon osan perustus, merkitse rasti oikeaan vaihtoehtoon A-D

- A Perusmuuri ja maanvarainen lattia-laatta

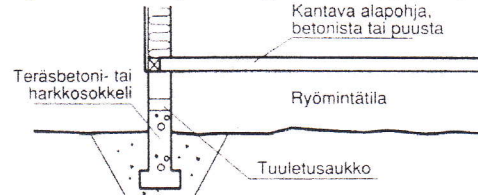


- Perusmuuri (sokkeli)
 Betonista Kevytsoraharkoista En tiedä

- B Reunajäykistetty laatta

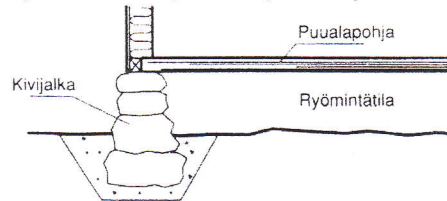


- C Ryömintätilainen perustus, jossa teräsbetoni- tai harkkosokkeli



- Alapohja Betonista Puusta

- D Ryömintätilainen perustus, jossa kivijalka



- Yhdistelmä perustustavoista A-D, mistä
 Muu, mikä
 En tiedä

LISÄSELVITYKSIÄ (tarvittaessa eri paperille)

KIITOS VAIVANNÄÖSTÄ!
 Tarkistathan vielä aloitus- ja lopetuspäivämäärät.