

PRO GRADU -TUTKIELMA

Hemmo Immonen

**Jyväskylän hoitokalastuksen BACI-asetelman mallinnus
lineaarisen sekamallin avulla**

JYVÄSKYLÄN YLIOPISTO
Matematiikan ja tilastotieteen laitos
Tilastotiede
Joulukuu 2011

Jyväskylän yliopisto

Matematiikan ja tilastotieteen laitos

IMMONEN, HEMMO: Jyväsjärven hoitokalastuksen BACI-asetelman mallin-
nus lineaarisen sekamallin avulla

Pro gradu -tutkielma, 31 s., 2 liitettä.

Tilastotiede

Joulukuu 2011

Tiivistelmä

Hoitokalastus on rehevöityneen järven kunnostusmenetelmä, jonka tavoitteena on parantaa veden laatua ja järven biologista tilaa. Suomessa tehdään paljon hoitokalastuksia, mutta niiden vaikutuksista on vähän tutkittua tietoa. Järvien tilaa kuvaavissa biologisissa muuttujissa on paljon luontaista vaihtelua, mikä asettaa haasteita hoitokalastuksien vaikutusten tutkimiselle. Tässä tutkielmassa tutkittiin hoitokalastuksen vaikutusta Jyväsjärven pääsyvänteen ja Ainolan syvänteen pohjaeläimistöön BACI-asetelman (Before-After-Control-Impact) avulla. BACI-asetelmassa tutkittavat muuttujat mitattiin ennen ja jälkeen hoitokalastuksen sekä vaikutusalueelta että kontrollialueilta. Hoitokalastus toteutettiin Jyväsjärvessä vuosina 2004–2006. Pohjaeläinnäytteet kerättiin vuosina 2000–2008 Jyväsjärvestä ja kolmesta kontrollijärvestä. Aineisto analysoitiin lineaarisen sekamallin avulla. Sekamalli soveltuu riippuvien ja hierarkkisten aineistojen mallintamiseen. Analysoidut muuttujat olivat pohjaeläinten kokonaisyksilötiheys ja pohjanlaatuindeksi. Kummallekin muuttujalle mallinnettiin soveliaain riippuvuus rakenne. Jyväsjärven pääsyvänteessä hoitokalastuksen vaikutusta pohjaeläinten kokonaisyksilötiheyteen ei havaittu, mutta Ainolan syvänteessä yksilötiheys pieneni hoitokalastuksen jälkeen. Pohjanlaatuindeksin arvot kohosivat hoitokalastuksen jälkeen sekä pääsyvänteessä että Ainolan syvänteessä. Tämän tutkimuksen perusteella Jyväsjärven tila on parantunut hoitokalastuksen ansiosta vain vähän. Järven eri alueet näyttävät reagoivan hoitokalastukseen eri tavoin.

Asiasanat: pohjaeläin, pohjanlaatuindeksi, rehevöityminen

Sisältö

1	Johdanto	7
2	Aineisto	8
3	Lineaarinen sekamalli	10
3.1	Yleinen lineaarinen sekamalli	10
3.2	Kovarianssiparametrien estimointi	11
3.3	Kiinteän osan estimointi ja testaus	13
3.4	Satunnaisvaikutusten ennustaminen	14
3.5	Kovarianssirakenteet	14
3.6	Mallin valinta ja oletusten tarkastelu	16
4	Mallin sovitus	17
4.1	Sekamalli BACI-aineistolle	17
4.2	Kokonaisyksilöitiheys	20
4.3	Pohjanlaatuindeksi	24
5	Pohdintaa	28
	Lähdeluettelo	29
	Liite 1: Pohjaeläinaineisto	32
	Liite 2: R-koodi	33

1 Johdanto

Vesistöjen rehevöityminen eli ravinnepitoisuuksien liiallinen kasvu on merkittävä ongelma, joka uhkaa järviä kaikkialla maailmassa. Valuma-alueelta tulevan ravinnekuormituksen vähentäminen on tärkein tekijä pyrittäessä elvyttämään rehevöitynyttä järveä, mutta usein tarvitaan myös kunnostustoimia itse järvessä. Eräs rehevöityneen järven kunnostusmenetelmä on hoitokalastus, jonka aikana tehokkaalla kalastuksella vähennetään voimakkaasti vähäarvoisia kaloja kuten särkiä ja pieniä ahvenia, joiden mukana järvestä poistuu ravinteita. Hoitokalastuksen toivottuja vaikutuksia ovat veden kirkastuminen, kalaston rakenteen parantuminen ja syvänteiden happipitoisuuksien paraneminen.

Vaikka hoitokalastuksia on tehty runsaasti, niiden onnistumisista on vain vähän näyttöä (Syväranta 2008). Usein vaikutus jää lyhytaikaiseksi tai vaikutuksen tasoa ei pystytä luotettavasti mittaamaan. Hoitokalastuksen onnistumista mittaavissa muuttujissa on usein suurta vuosien välistä vaihtelua, mikä hankaloittaa tulosten tulkintaa ja vaatii pitkiä seurantajaksoja (Ruuhijärvi & Olin 2002).

BACI-tutkimus (*Before-After Control-Impact*) on tapa mitata jonkin ekologisen tapahtuman tai käsittelyn aiheuttamaa vaikutusta luonnossa. BACI-asetelmassa vastemuuttujan arvoja mittaavat näytteet otetaan ennen käsitteilyä ja sen jälkeen sekä vaikutusalueelta että kontrollialueelta. Mikäli käsittelyllä on todellista vaikutusta vastemuuttujan arvoihin, pitää vaikutusalueen ennen-jälkeen-vaihtelun poiketa merkitsevästi kontrollialueella havaitusta ennen-jälkeen-vaihtelusta. Pätevä BACI-asetelma vaatii runsaan määrän biologisia mittauksia, jotka ovat hitaita ja työläitä toteuttaa käytännössä. Usein esim. mittausvuosien tai kontrollialueiden määrästä joudutaan tinkimään, jolloin tilastollinen tehokkuus ei riitä tutkittavan vaikutuksen todentamiseen (Underwood 1994).

Tässä pro gradu -tutkielmassa esitellään lineaaristen sekamallien teoriaa ja tutkitaan lineaarisen sekamallin avulla hoitokalastuksen vaikutusta Jyväsjärven syvänteiden pohjaeläimistöön. Pohjaeläinaineisto on kerätty BACI-asetelman mukaisesti. Perinteisissä regressiomalleissa ja varianssianalyyseissä havainnot oletetaan keskenään riippumattomiksi ja perusjoukoittain samavarianssisiksi. Näistä oletuksista voidaan sekamallia käytettäessä luopua. Sekamalli on käyttökelpoinen menetelmä, kun tutkitaan pitkittäisiä, spatiaalisia tai muutoin hierarkkisia aineistoja (Nissinen 2009; Pinheiro & Bates 2002).

2 Aineisto

Jyväskylän keskustan kupeessa sijaitseva Jyväsjärvi oli teollisuuden ja asutusten jätevesien rajusti saastuttama 1960–1970-luvuilla. Viimeisten vuosikymmenten aikana järven kunto on parantunut huomattavasti, mutta Jyväsjärvi täyttää edelleen rehevöityneen järven tunnusmerkit (Meriläinen ym. 2003, Salonen ym. 2005). Syyt järven kunnan parantumiseen ovat ulkoisen ravinnekuormituksen vähentyminen sekä vuonna 1979 aloitettu pääsyvänteen ilmastus. Ilmastus tapahtuu pumppaamalla hapekasta pintavettä järven pohjalle. Ulkoinen ravinnekuormitus on vähentynyt, koska jätevesiä ei enää lasketa järveen puhdistamattomina.

Jyväsjärvi-hankkeen (Päijänne LTER 2011) tavoitteena on ollut parantaa järven tilaa ja samalla muodostaa Jyväsjärvestä ja Pohjois-Päijänteestä monipuolinen ja käytännönläheinen opetus- ja tutkimusympäristö. Jyväsjärven tilan parantamiseksi toteutettiin kolmevuotinen hoitokalastus rysäpyyntinä vuosina 2004–2006. Kolmen vuoden aikana poistettiin 101 tonnia (300 kg/ha) särki- ja ahvenkaloja. Yleisimmät saalislajit olivat lahna (40 %), ahven (27 %) ja särki (24 %).

Pohjaeläinaineisto koostuu Jyväsjärvi-hankkeen vuosina 2000–2008 (Patajärvässä 2001–2003) kerätyistä syvännepohjaeläinnäytteistä. Näytteet otettiin Ekman-noutimella, joka on vaijerin varassa järven pohjaan laskettava noudin. Noutimen pohjassa on sulkumekanismi, joka kahmaisee tietyltä pinta-alalta (256 cm²) pohjalietettä noutimen sisään. Noudin nostetaan veneeseen ja lietteessä olevat pienet eläimet seulotaan talteen.

Hoitokalastuksella käsitellystä Jyväsjärvestä tutkittiin kaksi aluetta, pääsyväntä (24 m) ja Ainolan syväntä (15 m). Ainolan syväntä on laajan rantavyöhykkeen ympäröimä, kun taas pääsyväntä on laajemman ulappa-alueen kohdalla. Kontrollialueina olivat kooltaan ja vedenlaadultaan samankaltaiset Patajärvi, Alvajärvi ja Tuomiojärvi. Alvajärvi ja Tuomiojärvi sijaitsevat samalla valuma-alueella ja laskevat vetensä Palokkajärven ja Tourujoen kautta Jyväsjärveen. Patajärvi sijaitsee erillään muista järvistä Korpilahdella (Kuva 2.1).

Kunakin yhdeksänä vuonna kontrollijärvien pääsyvänteistä otettiin seitsemän ja Jyväsjärven molemmista syvänteistä viisi rinnakkaisnäytettä. Kunkin paikan viisi tai seitsemän rinnakkaisnäytettä yhdistettiin laskemalla niistä keskiarvo. Vuosina 2000–2003 kerätyt näytteet edustavat tilannetta ennen hoitokalastuksia, vuosien 2004–2006 näytteet kalastusten aikaista tilannetta, ja vuodet 2007–2008 tilannetta käsittelyn jälkeen.

Edellä mainitusta pohjaeläinaineistosta määritettiin kahdeksan yhteisömuut-



Kuva 2.1. Tutkimuspaikkojen sijainti.

tujaa: kokonaisyksilötiheys, kokonaisbiomassa, harvasukasmatojen biomassa, surviaissääskien biomassa, pohjanlaatuindeksi (BQI), Shannonin diversiteetti-indeksi, taksonien määrä sekä harvasukasmatojen ja surviaissääskien suhde. Muuttujien tilastollisesta analysoinnista esitetään tässä pro gradu -työssä sekamallin sovitus kahdelle muuttujalle: kokonaisyksilötiheydelle (yksilöä/m²) ja pohjanlaatuindeksille BQI (*Benthic Quality Index*, Wiederholm 1980).

Järvisyvänteissä suuret pohjaeläinyksilötiheydet ovat usein tyypillisiä rehevöityneille järville, joissa muutama heikenneitä oloja sietävä laji menestyy esiintyen poikkeuksellisen runsaslukuisena. BQI ilmentää järvisyvänteiden biologista tilaa ja perustuu tiettyjen surviaissääskien määriin ja pistearvoihin. Painotettuna keskiarvona laskettavien indeksiarvojen vaihteluväli on 1–5. Suuri arvo kuvaa vähäistä ravinnekuormitusta, korkeaa happipitoisuutta ja siten parempaa syvänteiden laatua. Pienet arvot ovat tyypillisiä matalille ja reheville järvisyvänteille, joissa alusveden happipitoisuus on kerrostuneisuuskauden aikana vähäinen. BQI saa arvon 0, mikäli yhtään indikaattorilajia ei tavata näytteistä.

3 Lineaarinen sekamalli

Tässä luvussa esitellään lineaaristen sekamallien teoriaa. Teoria mukailee Nissisen (2009) esitystä.

3.1 Yleinen lineaarinen sekamalli

Matriisimerkinnöin lineaarinen sekamalli on

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad (3.1)$$

missä

\mathbf{y} on $n \times 1$ -vastevektori,

\mathbf{X} on kiinteän osan $n \times p$ -kerroinmatriisi,

$\boldsymbol{\beta}$ on kiinteiden vaikutusten $p \times 1$ -vektori,

\mathbf{Z} on satunnaisosan $n \times k$ -kerroinmatriisi,

\mathbf{u} on satunnaisvaikutusten $k \times 1$ -vektori ja

$\boldsymbol{\varepsilon}$ on $n \times 1$ -jäännösvektori.

Satunnaisvaikutukset ja jäännökset oletetaan toisistaan riippumattomiksi ja niiden odotusarvot nolliksi, eli

$$E(\mathbf{u}) = \mathbf{0}, \quad E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad \text{ja} \quad Cov(\boldsymbol{\varepsilon}, \mathbf{u}) = \mathbf{0}.$$

Satunnaisvektoreiden kovariansseja merkitään

$$Cov(\mathbf{u}) = \mathbf{D} = \mathbf{D}(\boldsymbol{\theta}) \quad \text{ja} \quad Cov(\boldsymbol{\varepsilon}) = \mathbf{R} = \mathbf{R}(\boldsymbol{\theta}),$$

missä $\boldsymbol{\theta}$ -vektorit sisältävät estimoitavat varianssiparametrit. Edellä mainittujen oletusten toteutuessa odotusarvoiksi tulee

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{ja} \quad E(\mathbf{y}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}.$$

Jäännökset eli satunnaisvirheet ovat

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

Lisäksi sekamallissa yleensä oletetaan, että satunnaisvaikutukset ja jäännökset ovat normaalisti jakautuneita

$$\mathbf{u} \sim N_k(\mathbf{0}, \mathbf{D}) \quad \text{ja} \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \mathbf{R}).$$

Sekamallissa havainnot ovat korreloituneita ja niiden kovarianssimatriisi on

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

3.2 Kovarianssiparametrien estimointi

Lineaarisen sekamallin oletusten mukaan vastevektori noudattaa multinormaalijakaumaa $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, jossa kovarianssimatriisi $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}$ on varianssiparametrien $\boldsymbol{\theta}$ funktio $\mathbf{V}(\boldsymbol{\theta})$. Silloin uskottavuusfunktioiksi tulee

$$L_{SU}(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) = (2\pi)^{-n/2} |\mathbf{V}(\boldsymbol{\theta})|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right],$$

joka logaritmoituna ja kahdella kerrottuna on

$$2l_{SU}(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) = c - \log|\mathbf{V}(\boldsymbol{\theta})| - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (3.2)$$

missä c on $\boldsymbol{\theta}$:sta ja $\boldsymbol{\beta}$:sta riippumaton vakio. Derivoidaan (3.2) $\boldsymbol{\beta}$:n suhteen

$$\frac{\partial 2l_{SU}(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\beta}} = \mathbf{X}' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

josta saadaan GLS-ratkaisu (*Generalized Least Squares*)

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

Sijoittamalla $\boldsymbol{\beta}$:n paikalle $\tilde{\boldsymbol{\beta}}$ uskottavuusfunktioon (3.2) saadaan logaritminen profiiliuskottavuusfunktio

$$2l_{SU}(\boldsymbol{\theta}; \mathbf{y}) = c - \log|\mathbf{V}(\boldsymbol{\theta})| - (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}), \quad (3.3)$$

josta maksimoimalla $\boldsymbol{\theta}$:n suhteen saadaan iteratiivisesti ratkaistua varianssiparametrien suurimman uskottavuuden estimaatti $\hat{\boldsymbol{\theta}}_{SU}$.

SU-menetelmällä on hyvät asymptoottiset ominaisuudet, mutta se tuottaa usein harhaisia varianssiestimaatteja. Harhaisuuden vähentämiseksi käytetään REML-menetelmää (*Restricted Maximum Likelihood* tai *Residual Maximum Likelihood*), jossa maksimoidaan muunnoksen $\mathbf{w} = \mathbf{A}'\mathbf{y}$ avulla saatu uskottavuusfunktio. \mathbf{A} on täysiasteinen $n \times (n - p)$ matriisi, jolle pätee $\mathbf{A}'\mathbf{X} = \mathbf{0}$ ja $\mathbf{A}'\mathbf{A} = \mathbf{I}_{n-p}$. Silloin

$$\mathbf{w} \sim N_{n-p}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A}),$$

josta saatava logaritminen REML-uskottavuusfunktio voidaan Harvillen (1974) mukaan kirjoittaa muotoon

$$2l_R(\boldsymbol{\theta}; \mathbf{y}) = c - \log|\mathbf{V}(\boldsymbol{\theta})| - \log|\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}| - (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$$

$$= 2l_{SU}(\boldsymbol{\theta}; \mathbf{y}) - \log|\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}|.$$

$\hat{\boldsymbol{\theta}}$ on $\boldsymbol{\theta}$:n REML-estimaattori, jos $2l_R(\boldsymbol{\theta}; \mathbf{y})$ saavuttaa maksimin arvolla $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Yleensä ratkaisua ei löydy suljetussa muodossa, vaan estimoinnissa käytetään iteratiivisia keinoja kuten EM-algoritmia tai Newtonin-Raphsonin menetelmää. Matriisi \mathbf{A} voidaan valita mielivaltaisesti, kunhan se toteuttaa em. ehdot.

Tarkastellaan esimerkiksi varianssin estimointia lineaarisessa mallissa $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, missä \mathbf{y} on $n \times 1$ havaintovektori, \mathbf{X} on p -asteinen $n \times p$ kerroinmatriisi ja $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Nyt $\mathbf{V} = \sigma^2\mathbf{I}$, joten

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ &= [\mathbf{X}'(\sigma^2\mathbf{I})^{-1}\mathbf{X}]^{-1}\mathbf{X}'(\sigma^2\mathbf{I})^{-1}\mathbf{y} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\frac{1}{\sigma^2}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \hat{\boldsymbol{\beta}},\end{aligned}$$

joka on pienimmän neliösumman ratkaisu $\boldsymbol{\beta}$:lle. Suurimman uskottavuuden estimaatti σ^2 :lle saadaan maksimoimalla uskottavuusfunktio (3.2)

$$\begin{aligned}2l_{SU}(\sigma^2) &= c - \log|\sigma^2\mathbf{I}_{n \times n}| - (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= c - \log(\sigma^2)^n - (\sigma^2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= c - n\log(\sigma^2) - (\sigma^2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),\end{aligned}$$

joka derivoidaan ja asetetaan derivaatta nolllaksi

$$\frac{\partial 2l_{SU}(\sigma^2)}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\sigma^2)^2} = 0,$$

mistä tulee

$$\hat{\sigma}_{SU}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}.$$

Tiedetään, että $\hat{\sigma}_{SU}^2$ ei ole harhaton. REML-estimaatti tässä tapauksessa saadaan maksimoimalla uskottavuusfunktio (3.3)

$$\begin{aligned}2l_R(\sigma^2) &= c - \log|\sigma^2\mathbf{I}_{n \times n}| - \log|\mathbf{X}'(\sigma^2\mathbf{I}_{n \times n})^{-1}\mathbf{X}| - (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= c - \log(\sigma^2)^n - \log|(\sigma^2)^{-1}\mathbf{X}'\mathbf{X}| - (\sigma^2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= c - n\log(\sigma^2) - \log(\sigma^2)^{-p} - (\sigma^2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= c - n\log(\sigma^2) + p\log(\sigma^2) - (\sigma^2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= c - (n - p)\log(\sigma^2) - (\sigma^2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).\end{aligned}$$

Kun tämä derivoidaan ja asetetaan derivaatta nolllaksi

$$\frac{\partial 2l_R(\sigma^2)}{\partial \sigma^2} = -\frac{n - p}{\sigma^2} + \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\sigma^2)^2} = 0,$$

saadaan

$$\hat{\sigma}_{REML}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - p},$$

joka on harhaton REML-estimaatti σ^2 :lle (Davis 2002).

3.3 Kiinteän osan estimointi ja testaus

Sekamallin kiinteän osan kertoimet voidaan estimoida pienimmän neliösumman menetelmällä (PNS)

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

mutta menetelmä ei ota huomioon havaintojen korreloituneisuutta. Parempi menetelmä kiinteän osan estimoimiseksi on paras lineaarinen ja harhaton estimaattori BLUE(β)

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

joka voidaan johtaa ratkaisemalla Hendersonin sekamalliyhtälöt (Henderson ym. 1959):

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}. \quad (3.4)$$

Hendersonin sekamalliyhtälöiden avulla saavutetaan laskennallista hyötyä, koska $n \times n$ -kokoista kovarianssimatriisia \mathbf{V} ei tarvitse kääntää. Matriisien $\mathbf{D}_{q \times q}$ ja $\mathbf{R}_{n \times n}$ käänteismatriisit sen sijaan pitää laskea, mutta q ei yleensä ole kovin suuri ja usein \mathbf{R} on diagonaalinen.

Tarkastellaan hypoteesien testausta tilanteessa, jossa \mathbf{V} on tunnettu. Testattavana on nollahypoteesi

$$H_0 : \mathbf{L}'\beta = \mathbf{0}$$

ja vastahypoteesina

$$H_1 : \mathbf{L}'\beta \neq \mathbf{0},$$

missä β on kiinteiden vaikutusten $p \times 1$ vektori ja \mathbf{L} on valittu q -asteinen $p \times q$ matriisi, jossa $q \leq p$. Lineaarikombinaation $\mathbf{L}'\beta$ estimaattorina käytetään estimaattoria $\mathbf{L}'\tilde{\beta}$, jonka kovarianssimatriisi on

$$Cov(\mathbf{L}'\tilde{\beta}) = \mathbf{L}'Cov(\tilde{\beta})\mathbf{L} = \mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}.$$

Kun H_0 on voimassa, Waldin testisuure

$$\chi^2 = \tilde{\beta}'\mathbf{L}[\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}]^{-1}\mathbf{L}'\tilde{\beta}$$

noudattaa χ^2 -jakaumaa vapausasteina kerroinmatriisi \mathbf{L} :n aste $rank(\mathbf{L})$.

Yleensä \mathbf{V} ei ole tunnettu, vaan se korvataan aineistosta lasketulla estimaattilla $\hat{\mathbf{V}}$. Silloin $\mathbf{L}'\tilde{\beta}$:n kovarianssin estimaatiksi tulee

$$Cov(\mathbf{L}'\tilde{\beta}) = \mathbf{L}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{L}$$

ja testisuurena käytetään Waldin F -testiä

$$F = \frac{\chi^2}{rank(\mathbf{L})},$$

joka noudattaa F -jakaumaa, kun H_0 on tosi. Vapausasteina osoittajassa on kerroinmatriisi \mathbf{L} :n aste $\text{rank}(\mathbf{L})$ ja nimittäjän vapausaste approksimoidaan. Usein käytetty approksimointimenetelmä on Satterthwaiten (1946) menetelmä, ks. myös Pinheiro & Bates (2002), s. 91–92. Yksittäisen kiinteän kertoimen testaamisessa käytetään testisuuretta $T = \sqrt{F}$. Kun $\mathbf{L} = \mathbf{l}$ on vektori, testisuureeksi tulee

$$T = \sqrt{F} = \sqrt{\frac{(\mathbf{l}'\tilde{\boldsymbol{\beta}})^2}{\text{Var}(\mathbf{l}'\tilde{\boldsymbol{\beta}})}} = \frac{|\mathbf{l}'\tilde{\boldsymbol{\beta}}|}{s.e.(\mathbf{l}'\tilde{\boldsymbol{\beta}})},$$

jonka jakaumaa approksimoidaan t -jakaumalla ja vapausasteet estimoidaan kuten edellä.

3.4 Satunnaisvaikutusten ennustaminen

Kun kovarianssimatriisi \mathbf{V} on tunnettu, paras lineaarinen harhaton ennuste (*BLUP*, *Best Linear Unbiased Predictor*) satunnaisvaikutusten vektorille \mathbf{u} on

$$\tilde{\mathbf{u}} = \text{BLUP}(\mathbf{u}) = \mathbf{DZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}),$$

missä $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Tavallisesti kovarianssiparametreja ei tunneta, vaan ne estimoidaan havaintoaineistosta. Silloin \mathbf{V} ja \mathbf{D} korvataan estimaateillaan ja saadaan empiirinen *BLUP*

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

joka voidaan ratkaista sekamalliyhtälöistä (3.4), jossa kovarianssimatriisit \mathbf{V} , \mathbf{D} ja \mathbf{R} korvataan estimaateillaan.

3.5 Kovarianssirakenteet

Sekamallissa havaintojen kovarianssimatriisia voidaan mallintaa antamalla satunnaisvaikutusten kovarianssimatriiseille ja jäännösten kovarianssimatriisille erilaisia rakenteita. Satunnaisvaikutuksille voidaan asettaa esimerkiksi tasakorrelaatio-, diagonaalinen tai lohki diagonaalinen rakenne.

Lineaarisen sekamallin perusmuodossa rajoitetaan jäännökset riippumattomiksi, samoin jakautuneiksi ja samavarianssiksi satunnaismuuttujiksi keskiarvolla nolla. Monissa käytännön sovelluksissa jäännökset ovat heteroskedastisia, korreloituneita tai sekä heteroskedastisia että korreloituneita. Lineaarista sekamallia voidaan laajentaa niin, että malli sallii jäännöksille erilaisia riippuvuusrakenteita. (Pinheiro & Bates 2002).

Tarkastellaan jäännösten kovarianssirakennetta

$$\mathbf{R} = \sigma_\varepsilon^2 \boldsymbol{\Lambda}.$$

Kun $\boldsymbol{\Lambda}$ on positiivisesti definiitti, se voidaan esittää muodossa

$$\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}^{1/2})' \boldsymbol{\Lambda}^{1/2}$$

ja käänteismatriisin muoto on

$$\mathbf{\Lambda}^{-1} = \mathbf{\Lambda}^{-1/2}(\mathbf{\Lambda}^{-1/2})'.$$

Tehdään sekamallille $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ muunnos $(\mathbf{\Lambda}^{-1/2})'\mathbf{y}$. Silloin saadaan

$$(\mathbf{\Lambda}^{-1/2})'\mathbf{y} = (\mathbf{\Lambda}^{-1/2})'\mathbf{X}\boldsymbol{\beta} + (\mathbf{\Lambda}^{-1/2})'\mathbf{Z}\mathbf{u} + (\mathbf{\Lambda}^{-1/2})'\boldsymbol{\varepsilon}.$$

Kun merkitään muunnetun mallin osia

$$\begin{cases} \mathbf{y}^* = (\mathbf{\Lambda}^{-1/2})'\mathbf{y} \\ \mathbf{X}^* = (\mathbf{\Lambda}^{-1/2})'\mathbf{X} \\ \mathbf{Z}^* = (\mathbf{\Lambda}^{-1/2})'\mathbf{Z} \\ \boldsymbol{\varepsilon}^* = (\mathbf{\Lambda}^{-1/2})'\boldsymbol{\varepsilon}, \end{cases}$$

saadaan malliksi

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{Z}^*\mathbf{u} + \boldsymbol{\varepsilon}^*.$$

Muunnetun mallin jäännösten kovarianssimatriisi on

$$\begin{aligned} Cov[\boldsymbol{\varepsilon}^*] &= Cov[(\mathbf{\Lambda}^{-1/2})'\boldsymbol{\varepsilon}] \\ &= (\mathbf{\Lambda}^{-1/2})'Cov[\boldsymbol{\varepsilon}](\mathbf{\Lambda}^{-1/2}) \\ &= (\mathbf{\Lambda}^{-1/2})'(\sigma_\varepsilon^2\mathbf{\Lambda})(\mathbf{\Lambda}^{-1/2}) \\ &= \sigma_\varepsilon^2(\mathbf{\Lambda}^{-1/2})'\mathbf{\Lambda}\mathbf{\Lambda}^{-1/2} \\ &= \sigma_\varepsilon^2(\mathbf{\Lambda}^{-1/2})'(\mathbf{\Lambda}^{1/2})'\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{-1/2} \\ &= \sigma_\varepsilon^2\mathbf{I}. \end{aligned}$$

Odotusarvo on

$$E[\boldsymbol{\varepsilon}^*] = E[(\mathbf{\Lambda}^{-1/2})'\boldsymbol{\varepsilon}] = (\mathbf{\Lambda}^{-1/2})'E[\boldsymbol{\varepsilon}] = (\mathbf{\Lambda}^{-1/2})'\mathbf{0} = \mathbf{0}.$$

Muunnetun mallin jäännökset ovat siis riippumattomia normaalijakautuneita

$$\boldsymbol{\varepsilon}^* \sim N(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}).$$

Matriisi $\mathbf{\Lambda}$ voidaan hajottaa komponentteihin

$$\mathbf{\Lambda} = \mathbf{TCT},$$

missä \mathbf{T} on diagonaalimatriisi ja \mathbf{C} on korrelaatiomatriisi. \mathbf{T} :n diagonaalialkiot t_{ii} liittyvät jäännösten variansseihin ja \mathbf{C} kuvaa jäännösten korrelaatioita:

$$Var(\varepsilon_i) = \sigma_\varepsilon^2 t_{ii}^2 \quad \text{ja} \quad Cor(\varepsilon_i, \varepsilon_j) = c_{ij}.$$

Tällä tavoin hajoteltuna $\mathbf{\Lambda}$ voidaan jakaa varianssikomponenttiin ja korrelaatiokomponenttiin. Kahta komponenttia on mahdollista mallintaa erikseen monipuolisilla varianssi-kovarianssi-rakenteilla. (Pinheiro & Bates 2002).

Jäännösten korrelaatorakennetta voidaan sovelluksesta riippuen mallintaa esim. autokorrelaatorakenteilla (AR, MA, ARMA) tai spatiaalisilla variogrammeilla. Jäännösten varianssirakennetta voidaan mallintaa muodostamalla varianssille malli, jonka arvot riippuvat jostain taustamuuttujasta \mathbf{x} jonkin funktion $f(\mathbf{x}, \boldsymbol{\beta})$ kautta. Tilanteesta riippuen käytetään erilaisia varianssifunktioita ja erilaisia taustamuuttujia varianssin mallintamiseen. Usein varsinkin biologisissa tutkimuksissa varianssi on suhteessa vastemuuttujan tasoon, jolloin taustamuuttujana voidaan käyttää sovitettuja arvoja.

3.6 Mallin valinta ja oletusten tarkastelu

Yleinen menetelmä sisäkkäisten SU-estimoitujen mallien vertailuun on uskottavuussuhteen testi eli LR -testi (*Likelihood Ratio Test*). Kun L_2 on laajemman mallin uskottavuus, L_1 on rajoitetun mallin uskottavuus ja nollahypoteesi pitää rajoitettua mallia riittävänä, testisuure

$$LR = 2\log(L_2/L_1) = 2[\log(L_2) - \log(L_1)]$$

noudattaa asympotoottisesti χ^2 -jakaumaa vapausastein $k_2 - k_1$, missä k_2 on laajemman mallin parametrien määrä ja k_1 on rajoitetumman mallin parametrien määrä. LR -testiä voidaan käyttää myös REML-estimoitujen mallien vertailuun esim. erilaisten kovarianssirakenteiden vertailussa, mikäli molemmissa vertailtavissa malleissa on samanlainen kiinteä osa (eli sama \mathbf{X} -matriisi yhtälössä (3.1)).

Testaamisen ohessa mallien vertailumenetelmänä käytetään myös informaatiokriteereitä. AIC (*Akaike Information Criterion*) on mallien vertailuun käytetty kriteeri, joka määritellään

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

missä l voi olla joko SU- tai REML-uskottavuusfunktio ja k on estimoitujen parametrien lukumäärä. Kriteeri tulkitaan siten, että pienimmän arvon tuottava malli tulkitaan paremmaksi.

Sekamallin satunnaisosaan liittyvien oletusten tarkastelua voidaan tehdä graafisesti. Piirtämällä hajontakuvio sovitettujen arvojen ja (standardoitujen) jäännösten suhteen voidaan havainnoida \mathbf{R} :n rakennetta (esim. heteroskedastisuus tai autokorrelaatio). Jäännösten ja satunnaisvaikutusten normaalisuusole- tusten ($\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$ ja $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$) voimassaoloa voidaan tarkastella piirtämällä jäännöskvantiilit suhteessa normaalijakauman teoreettisiin kvantiileihin ja vastaavasti piirtämällä satunnaisvaikutusten ennusteet suhteessa normaali- jakaumaan.

4 Mallin sovitus

4.1 Sekamalli BACI-aineistolle

Pohjaeläinaineisto analysoitiin R 2.10.1 -ohjelmistolla (R Development Core Team 2009). Aineistoon sovitettiin lineaarinen sekamalli R:n nlme-kirjaston nlme-funktiolla. Mallin muoto on

$$y_{ijt} = \mu + \alpha_t + \beta_i + (\alpha\beta)_{ti} + u_j + \varepsilon_{ijt},$$

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{D}) \quad \text{ja} \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R}),$$

missä

y_{ijt} on rinnakkaisnäytteiden keskiarvo alueesta $i = 1, 2, 3$, järvestä $j = 1, 2, 3, 4$ ja vuodelta $t = 2000, \dots, 2008$,

μ on yleinen keskitaso,

α_t on kiinteä aikafaktori,

β_i on kiinteä aluefaktori (kontrolli, Pääsyväne ja Ainola),

u_j on satunnainen järvfaktori (Jyväsjärvi, Alvajärvi, Patajärvi ja Tuomiojärvi), ε_{ijt} on jäännösvirhe alueesta i , järvestä j ja vuodelta t .

Mallin estimoinnissa käytettiin treatment-kontrastia, jonka avulla kunkin kiinteän faktorin ykköstatot asetetaan vertailutasoksi. Ennen-vuodet 2000–2003 ja kontrollialueet merkittiin ykkösellä, joten $\alpha_{2000} = \alpha_{2001} = \alpha_{2002} = \alpha_{2003} = 0$ ja $\beta_1 = 0$. Vuosien 2004–2008 kertoimet tulkitaan erotuksena ennen-vuosien yhteisestä keskimääräisestä arvosta ja Pääsyvänteen sekä Ainolan kertoimet ovat vastaavasti erotuksia kontrollialueiden keskitasosta. Matriisimerkinnöin $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ mallissa on silloin

$$\mathbf{y}_{44 \times 1} = \begin{bmatrix} y_{2;1;2000} \\ y_{2;1;2001} \\ \vdots \\ y_{1;4;2008} \end{bmatrix}, \quad \mathbf{X}_{44 \times 18} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \\ \mathbf{X}_5 \end{bmatrix}, \quad \boldsymbol{\beta}_{18 \times 1} = \begin{bmatrix} \mu \\ \alpha_{2004} \\ \vdots \\ \alpha_{2008} \\ \beta_2 \\ \beta_3 \\ \alpha_{2004}\beta_2 \\ \vdots \\ \alpha_{2008}\beta_2 \\ \alpha_{2004}\beta_3 \\ \vdots \\ \alpha_{2008}\beta_3 \end{bmatrix},$$

$$\mathbf{Z}_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ & & \vdots & \\ 0 & 0 & 1 & 0 \end{bmatrix}_{9 \times 4} \quad \text{ja} \quad \mathbf{Z}_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ & & & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}_{9 \times 4}.$$

Aluksi sovitetussa mallissa on $\mathbf{D} = \sigma_u^2 \mathbf{I}$ ja $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$. Sen jälkeen mallia laajennettiin tarvittaessa erilaisilla kovarianssirakenteilla. Molemmille vastemuuttujille valittiin erikseen omat kovarianssirakenteet, joiden sopivuutta arvioitiin uskottavuussuhteen testillä, informaatiokriteereiden avulla sekä graafisilla jäännöstarkasteluilla. Lopullisten mallien kiinteät osat testattiin ”marginal”-tyypin F -testillä, jossa käytetään tyypin kolme neliösummia (Pinheiro & Bates 2002).

Hoitokalastuksen vaikutusta kuvaavaa aika- aluetekijöiden yhdysvaikutusta testattiin BACI-kontrastilla eli lineaarikombinaatiolla, joka muodostaa nollahypoteesin

$$H_0 : \mu(BC) - \mu(AC) = \mu(BI) - \mu(AI)$$

eli

$$H_0 : \mu(BC) - \mu(AC) - \mu(BI) + \mu(AI) = 0,$$

jossa $\mu(BC)$ on kontrollialueiden keskimääräinen arvo ennen käsittelyä olevana aikana, $\mu(AC)$ on kontrollialueiden keskimääräinen arvo käsittelyn jälkeisenä aikana, $\mu(BI)$ on käsittelyalueiden keskimääräinen arvo ennen käsittelyä ja $\mu(AI)$ on käsittelyalueiden keskimääräinen arvo käsittelyn jälkeisenä aikana. Vastahypoteesina on, että käsittelyalueen ennen–jälkeen-muutos ($\mu(BI) - \mu(AI)$) ei ole yhtäsuuri kuin kontrollialueen ennen–jälkeen-muutos ($\mu(BC) - \mu(AC)$).

BACI-kontrastin kertoimien muoto β -vektorille on

	Ennen	Jälkeen
Kontrollialue	1	-1
Käsittelyalue	-1	1

Kun testataan hoitokalastuksen vaikutusta koko Jyväsjärven, asetetaan käsittelyalueeksi Jyväsjärvi. Silloin BACI-kontrastin kertoimet ovat

	α_{2000}	α_{2001}	α_{2002}	α_{2003}	α_{2004}	α_{2005}	α_{2006}	α_{2007}	α_{2008}
β_1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{5}$	$-\frac{1}{5}$	$-\frac{1}{5}$	$-\frac{1}{5}$	$-\frac{1}{5}$
β_2	$-\frac{1}{8}$	$-\frac{1}{8}$	$-\frac{1}{8}$	$-\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
β_3	$-\frac{1}{8}$	$-\frac{1}{8}$	$-\frac{1}{8}$	$-\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

Nyt kun $\alpha_{2000}, \dots, \alpha_{2003}$ ja β_1 ovat nollia, saadaan

$$\begin{aligned} & \frac{1}{4}(4\mu) - \frac{1}{5}(5\mu + \alpha_{2004} + \dots + \alpha_{2008}) \\ & - \frac{1}{8}(4\mu + 4\beta_2) - \frac{1}{8}(4\mu + 4\beta_3) \\ & + \frac{1}{10}(5\mu + \alpha_{2004} + \dots + \alpha_{2008} + 5\beta_2 + \alpha_{2004}\beta_2 + \dots + \alpha_{2008}\beta_2) \\ & + \frac{1}{10}(5\mu + \alpha_{2004} + \dots + \alpha_{2008} + 5\beta_3 + \alpha_{2004}\beta_3 + \dots + \alpha_{2008}\beta_3) \end{aligned}$$

$$\begin{aligned}
&= \mu - \mu - \frac{1}{5}(\alpha_{2004} + \dots + \alpha_{2008}) \\
&\quad - \frac{1}{2}\mu - \frac{1}{2}\beta_2 - \frac{1}{2}\mu - \frac{1}{2}\beta_3 \\
&\quad \frac{1}{2}\mu + \frac{1}{10}(\alpha_{2004} + \dots + \alpha_{2008}) + \frac{1}{2}\beta_2 + \frac{1}{10}(\alpha_{2004}\beta_2 + \dots + \alpha_{2008}\beta_2) \\
&\quad \frac{1}{2}\mu + \frac{1}{10}(\alpha_{2004} + \dots + \alpha_{2008}) + \frac{1}{2}\beta_3 + \frac{1}{10}(\alpha_{2004}\beta_3 + \dots + \alpha_{2008}\beta_3) \\
&= \frac{1}{10}(\alpha_{2004}\beta_2 + \dots + \alpha_{2008}\beta_2) + \frac{1}{10}(\alpha_{2004}\beta_3 + \dots + \alpha_{2008}\beta_3),
\end{aligned}$$

josta saadaan nollahypoteesi

$$H_0 : \mathbf{L}'\boldsymbol{\beta} = \frac{\alpha_{2004}\beta_2 + \dots + \alpha_{2008}\beta_2 + \alpha_{2004}\beta_3 + \dots + \alpha_{2008}\beta_3}{10} = 0.$$

Testattaessa hoitokalastuksen vaikutusta erikseen pääsyvänteessä ja Ainolan syvänteessä tulee nollahypoteesiksi pääsyvänteelle

$$H_0 : \mathbf{L}'\boldsymbol{\beta} = \frac{\alpha_{2004}\beta_2 + \dots + \alpha_{2008}\beta_2}{5} = 0$$

ja Ainolan syvänteelle saadaan

$$H_0 : \mathbf{L}'\boldsymbol{\beta} = \frac{\alpha_{2004}\beta_3 + \dots + \alpha_{2008}\beta_3}{5} = 0.$$

4.2 Kokonaisyksilöitiheys

Kokonaisyksilöitiheydelle kokeilluista kovarianssirakenteista malli, jossa on jäännösten potenssifunktio, antaa parhaan informaatiokriteerin arvon (Taulukko 4.1). Myös LR-testin perusteella tämä malli on ensimmäistä mallia parempi ($LR = 35.99; p < 0.0001$). Mallissa satunnaisvaikutusten kovarianssirakenne on $\mathbf{D} = \sigma_u^2 \mathbf{I}$ ja varianssin estimaatti $\hat{\sigma}_u^2 = 3045$. Jäännösten kovarianssirakenne on diagonaalinen, jossa jäännösvariانسsit ovat suhteessa vasteen tasoon ja samalla mallin sovitettuihin arvoihin $E[y|u]$. Estimoitu funktio on

$$\hat{V}ar(\varepsilon_{ijt}) = 0.0897[E(y_{ijt}|u_j)]^{2.153},$$

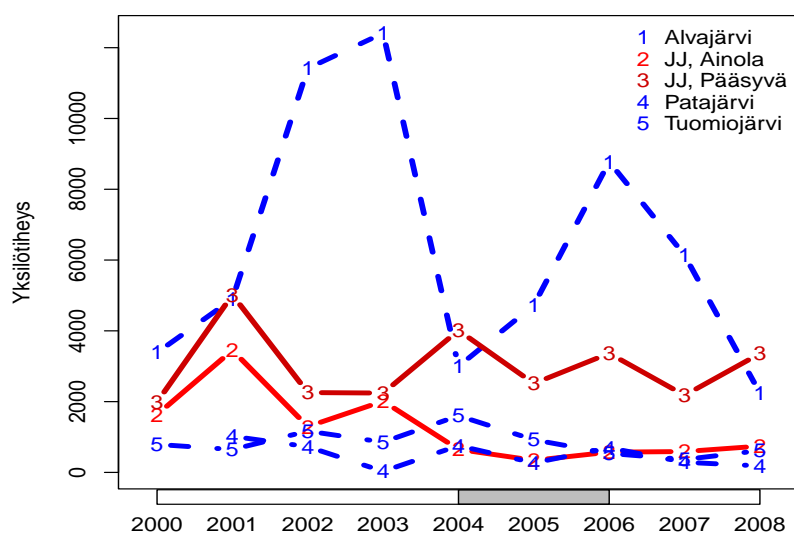
missä $i = 1, \dots, 9$, $j = 1, \dots, 4$ ja $t = 2000, \dots, 2008$.

Pohjaeläinten kokonaisyksilöitiheyden vuosittainen vaihtelu oli runsasta tutkimusjärjässä (Kuva 4.1). Jälkeen-vuosien keskiarvo on pienentynyt ennenvuosiin verrattuna kontrollijärjässä ja Ainolassa, mutta kohonnut Pääsyvänteessä (Taulukko 4.2). Ainolan vuosittaisten muutosten yhdysvaikutukset olivat tilastollisesti merkitseviä hoitokalastusvuosina, mutta eivät enää 2007–2008 (Taulukko 4.3). Kiinteitä kertoimia yhdessä testattaessa ne eivät osoittautuneet merkitseviksi (Taulukko 4.4). BACI-kontrastien testien perusteella hoitokalastus on pienentänyt pohjaeläinten kokonaisyksilöitiheyttä Ainolassa, mutta koko Jyväsjärjässä ja Pääsyvänteessä vaikutusta ei havaittu (Taulukko 4.4).

Kokonaisyksilöitiheyden mallin jäännökset jakaantuvat kohtuullisen tasaisesti sovitteiden funktiona (Kuva 4.2). Mallissa on hiukan enemmän pieniä jäännöksiä kuin mitä normaalisuus edellyttäisi (Kuva 4.3).

Taulukko 4.1. Informaatiokriteerin arvot kokonaisyksilötiheyden mallin erilaisille kovarianssirakenteille.

Malli	Kovarianssirakenne	df	AIC
1	$D = \sigma_u^2 \mathbf{I}, \mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$	20	527,9
2	D rajoittamaton, $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$	29	542,6
3	D diagonaalinen, $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$	23	530,6
4	$D = \sigma_u^2 \mathbf{I}$, järvillä omat jäännösvarianssit	23	495,6
5	$D = \sigma_u^2 \mathbf{I}$, jäännöksillä AR(1)-rakenne	21	528,1
6	$D = \sigma_u^2 \mathbf{I}$, jäännöksillä potenssifunktio	21	493,9
7	$D = \sigma_u^2 \mathbf{I}$, järvillä omat jäännösvar. ja AR(1)-rakenne	24	497,2
8	D diagonaalinen, jäännöksillä potenssifunktio	24	494,6



Kuva 4.1. Pohjaeläinten kokonaisyksilötiheyden (yksilöä/m²) keskiarvot alueittain tutkimusvuosina. Hoitokalastusvuodet on merkitty harmaalla palkilla.

Taulukko 4.2. Pohjaeläinten kokonaisyksilöitiheyden alueittaiset havaitut keskiarvot. Ennen = vuodet 2000–2003, jälkeen = vuodet 2004–2008.

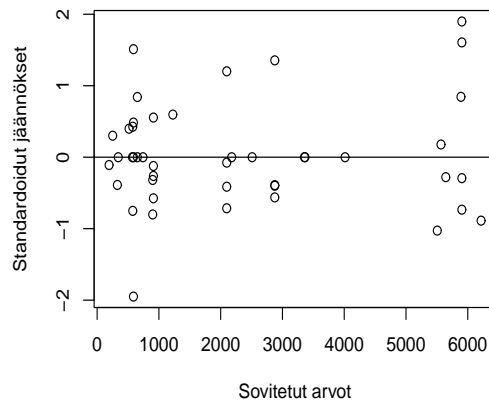
	Ennen	Jälkeen	Yht. keskiarvo
Kontrollit yht.	3395	2074	2633
Pääsyväne	2875	3085	2992
Ainola	2098	579	1254
Jyväsjärvi yht.	2487	1832	2123
Yht. keskiarvo	3010	1977	2424

Taulukko 4.3. Pohjaeläinten kokonaisyksilöitiheyden kiinteät kertoimet keskivirheineen ja testituloksineen.

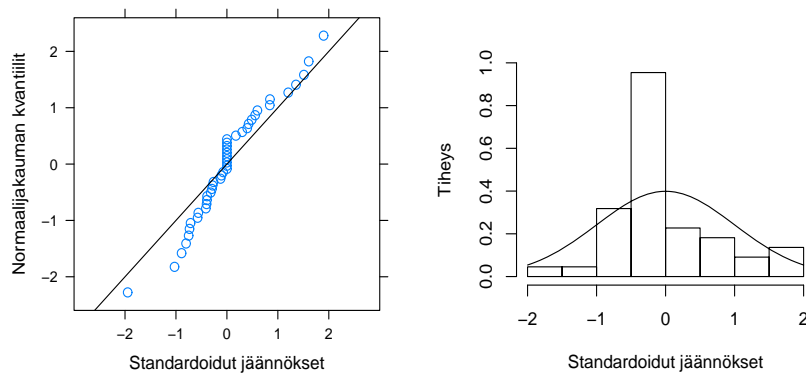
Kerroin	Estim.	SE	df	t	p
Vakio	2468	1798	23	1,372	0,183
2004	313	340	23	0,802	0,431
2005	-260	193	23	-1,346	0,191
2006	-11,9	272	23	-0,043	0,966
2007	-336	173	23	-1,938	0,065
2008	-395	160	23	-2,469	0,021
Pääsyväne	408	3623	23	0,113	0,911
Ainola	-369	3581	23	-0,103	0,919
2004×Pääsyvä	826	2429	23	0,340	0,737
2005×Pääsyvä	-104	1590	23	-0,066	0,948
2006×Pääsyvä	501	2051	23	0,244	0,809
2007×Pääsyvä	-362	1424	23	-0,254	0,802
2008×Pääsyvä	876	2034	23	0,431	0,671
2004×Ainola	-1761	756	23	-2,330	0,029
2005×Ainola	-1497	616	23	-2,428	0,023
2006×Ainola	-1514	685	23	-2,211	0,037
2007×Ainola	-1174	655	23	-1,790	0,087
2008×Ainola	-960	692	23	-1,388	0,178

Taulukko 4.4. F -testien tulokset pohjaeläinten kokonaisyksilöitiheyden kiinteille kertoimille ja niiden kontrasteille.

Kerroin	$df1$	$df2$	F	p
Vakio	1	23	1,883	0,183
Aika	5	23	2,005	0,116
Alue	2	23	0,321	0,729
Aika \times Alue	10	23	0,877	0,567
BACI-kontrasti:				
Jyväsjärvi	1	23	0,635	0,434
Pääsyväne	1	23	0,095	0,760
Ainola	1	23	5,277	0,031



Kuva 4.2. Pohjaeläinten kokonaisyksilöitiheyden mallin standardoidut jäännökset suhteessa sovitettuihin arvoihin.



Kuva 4.3. Pohjaeläinten kokonaisyksilöitiheyden mallin jäännösten jakauma suhteessa normaalijakaumaan.

4.3 Pohjanlaatuindeksi

Pohjanlaatuindeksin (BQI) mallissa jäännösten kovarianssirakenteena käytettiin varianssifunktiota, jossa järvifaktorin kullakin tasolla on oma jäännösvarianssi. AIC-arvojen perusteella hiukan paremmalta vaikutti malli, jossa oli lisäksi vielä AR(1)-rakenne (Taulukko 4.5), mutta LR-testin mukaan malli ilman AR(1)-rakennetta oli yhtä hyvä ($LR = 3.335$; $p = 0.068$). Lopullisessa mallissa satunnaisvaikutusten estimoitu kovarianssimatriisi $\hat{\mathbf{D}} = 0.124\mathbf{I}$ ja jäännösten estimoitu kovarianssimatriisi on

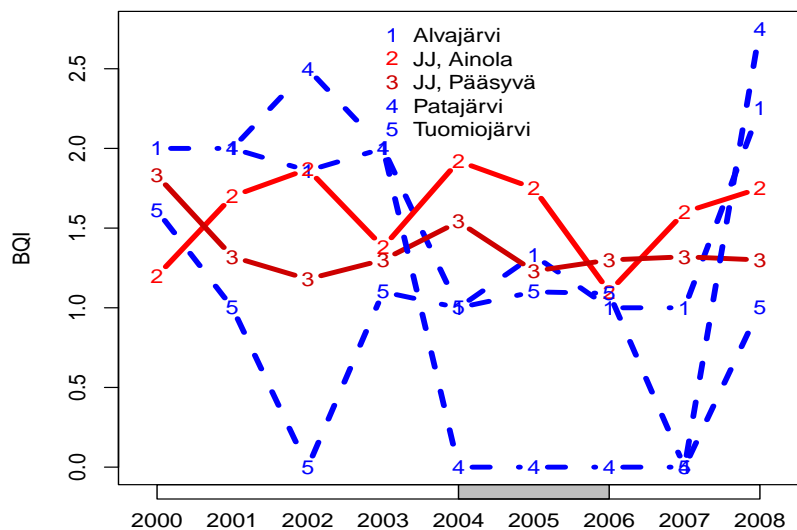
$$\hat{\mathbf{R}} = \begin{bmatrix} 0.088\mathbf{I}_{18 \times 18} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0.005\mathbf{I}_{9 \times 9} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0.554\mathbf{I}_{8 \times 8} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0.448\mathbf{I}_{9 \times 9} \end{bmatrix}.$$

Taulukko 4.5. Informaatiokriteerin arvot pohjanlaatuindeksin mallin erilaisille kovarianssirakenteille.

Malli	Kovarianssirakenne	df	AIC
1	$\mathbf{D} = \sigma_u^2\mathbf{I}, \mathbf{R} = \sigma_\varepsilon^2\mathbf{I}$	20	97,4
2	\mathbf{D} rajoittamaton, $\mathbf{R} = \sigma_\varepsilon^2\mathbf{I}$	29	114,7
3	\mathbf{D} diagonaalinen, $\mathbf{R} = \sigma_\varepsilon^2\mathbf{I}$	23	102,7
4	$\mathbf{D} = \sigma_u^2\mathbf{I}$, järvillä omat jäännösvarianssit	23	84,6
5	$\mathbf{D} = \sigma_u^2\mathbf{I}$, jäännöksillä AR(1)-rakenne	21	91,4
6	\mathbf{D} diagonaalinen, järvillä omat jäännösvarianssit	26	89,9
7	$\mathbf{D} = \sigma_u^2\mathbf{I}$, järvillä omat jäännösvar. ja AR(1)-rakenne	24	83,2

Pohjanlaatuindeksin arvot vaihtelivat kontrollijärvissä enemmän kuin Jyväsjärvessä (Kuva 4.4). Ainolassa indeksi on noussut jälkeen-vuosina verrattuna ennen-vuosiin. Pääsyvänteessä keskiarvo on hieman laskenut ja kontrollijärvissä pienentyminen on huomattava (Taulukko 4.6). Aika- ja aluetekijöiden yhdysvaikutustermit ovat tilastollisesti merkitseviä F -testin perusteella (Taulukko 4.8). Voimakkaimmat kohoamiset olivat molemmilla käsittelyalueilla vuonna 2004 (Taulukko 4.7). BACI-kontrastien testien perusteella hoitokalastus on nostanut pohjanlaatuindeksin arvoa Jyväsjärvessä (Taulukko 4.8). Merkitsevä vaikutus löytyy sekä Pääsyvänteestä että Ainolan syvänteestä.

Pohjanlaatuindeksin mallissa jäännökset jakaantuvat melko tasaisesti (Kuva 4.5). Suuria jäännöksiä näyttää olevan enemmän, mutta normalisuus toteutuu likimäärin (Kuva 4.6).



Kuva 4.4. Pohjanlaatuindeksin alueittaiset keskiarvot tutkimusvuosina. Hoitokalastusvuodet on merkitty harmaalla palkilla.

Taulukko 4.6. Pohjanlaatuindeksin alueittaiset havaitut keskiarvot. Ennen = vuodet 2000–2003, jälkeen = vuodet 2004–2008.

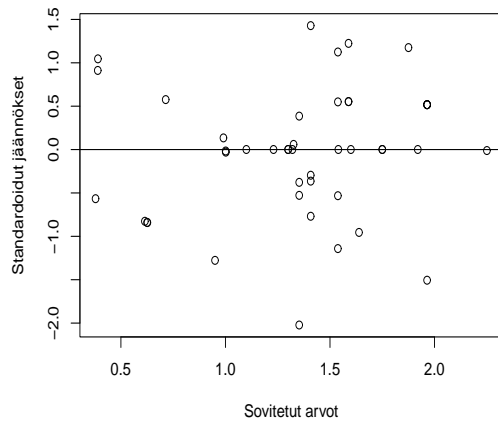
	Ennen	Jälkeen	Yht. keskiarvo
Kontrollit yht.	1,64	0,90	1,22
Pääsyväne	1,41	1,34	1,37
Ainola	1,54	1,62	1,59
Jyväsjärvi yht.	1,47	1,48	1,48
Yht. keskiarvo	1,57	1,13	1,32

Taulukko 4.7. Pohjanlaatuindeksin kiinteät kertoimet keskivirheeseen ja testituloksineen.

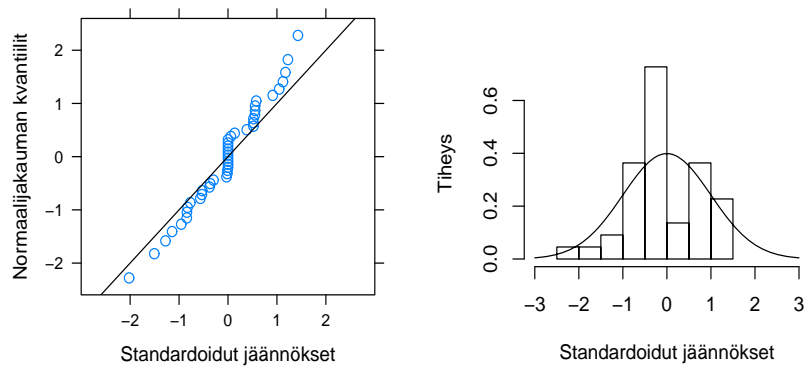
Kerroin	Estim.	SE	<i>df</i>	<i>t</i>	<i>p</i>
Vakio	1,635	0,231	23	7,077	0,000
2004	-0,963	0,077	23	-12,570	0,000
2005	-0,638	0,077	23	-8,331	0,000
2006	-0,962	0,077	23	-12,558	0,000
2007	-0,974	0,077	23	-12,707	0,000
2008	0,287	0,077	23	3,740	0,001
Pääsyväanne	-0,228	0,446	23	-0,510	0,615
Ainola	-0,098	0,446	23	-0,219	0,829
2004×Pääsyvä	1,096	0,340	23	3,227	0,004
2005×Pääsyvä	0,461	0,340	23	1,357	0,188
2006×Pääsyvä	0,855	0,340	23	2,517	0,019
2007×Pääsyvä	0,886	0,340	23	2,610	0,016
2008×Pääsyvä	-0,394	0,340	23	-1,161	0,258
2004×Ainola	1,346	0,340	23	3,963	0,001
2005×Ainola	0,851	0,340	23	2,506	0,020
2006×Ainola	0,525	0,340	23	1,545	0,136
2007×Ainola	1,036	0,340	23	3,052	0,006
2008×Ainola	-0,074	0,340	23	-0,218	0,829

Taulukko 4.8. *F*-testien tulokset pohjanlaatuindeksin kiinteille kertoimille ja niiden kontrasteille.

Kerroin	<i>df</i> 1	<i>df</i> 2	<i>F</i>	<i>p</i>
Vakio	1	23	50,08	< 0,001
Aika	5	23	90,13	< 0,001
Alue	2	23	0,26	0,771
Aika×Alue	10	23	4,69	0,001
BACI-kontrasti:				
Jyväsjärvi	1	23	19,89	< 0,001
Pääsyväanne	1	23	8,12	0,009
Ainola	1	23	13,07	0,002



Kuva 4.5. Pohjanlaatuindeksin mallin standardoidut jäännökset suhteessa sovitettuihin arvoihin.



Kuva 4.6. Pohjanlaatuindeksin (BQI) mallin jäännösten jakauma suhteessa normaalijakaumaan.

5 Pohdintaa

Hoitokalastuksen halutaan parantavan järvisyvänteiden pohjaeläinten elinympäristöä parantamalla alusveden ja pohjan olosuhteita muuttamalla niitä karumpaan ja hapekkaampaan tilaan. Silloin pohjanlaatuindeksin (BQI) pitäisi nousta ja pohjaeläinten yksilötiheyden pienentyä, koska hapekkaalla pohjalla viihtyvien lajien menestyminen paranisi ja likaisella pohjalla viihtyvät runsaslukuiset pohjaeläimet korvautuisivat kookkaammilla lajeilla.

Jyväsjärnessä havaittiin tilastollisesti merkitsevä BQI-indeksin kohoaminen hoitokalastuksen jälkeen. Pohjaeläinten kokonaisyksilötiheys laski merkitsevästi vain Ainolan syvänteessä, mutta ei Pääsyvänteessä. Hoitokalastus näyttää siten parantaneen Jyväsjärveä BQI-indeksin perusteella. Yksilötiheyden perusteella vaikutus havaittiin vain Ainolan syvänteessä.

BQI:n mallinnus on hieman epäluotettava, koska aineisto sisältää useita nollahavaintoja. Pohjanlaatuindeksi on nolla, kun yhtään indikaattorilajia ei löydetä tutkittavalta paikalta. Jos BQI-lajien puuttuminen johtuu näytteiden huonosta edustavuudesta, pitäisi nolla-arvot tulkita mallinnuksessa puuttuvina havaintoina. Nolla-arvoja havaittiin kontrollijärvissä, Tuomiojärnessä kahtena vuonna ja Patajärnessä neljänä vuonna. Patajärven seitsemän rinnakkaisnäytettä on enemmän kuin pohjaeläintutkimuksissa yleensä, eli näytteiden huonoa edustavuutta ei voi pitää indikaattorilajien puuttumisen syynä. Patajärven tila on siis BQI:n perusteella heikko. Mallinnuksessa BQI:n arvojen vaihteluvälinä oli siten käytännössä 0–5, vaikka vaihteluvälin yleensä sanotaan olevan 1–5.

Orgaanisen aineksen sedimentaationopeus ja pohjaeläinyhteisöt ovat eri syvänteissä erilaiset, joten Jyväsjärven kahden tutkimuspaikan erottelu analyysissä on biologisesti perusteltua. Pääsyvänteen ilmastus on saattanut olla merkittävä syy, miksi vaikutusta kokonaisyksilötiheyteen ei siellä havaittu. Ilmastuksen takia Pääsyvänteen lienee suhteellisen riippumaton hoitokalastuksesta, koska olosuhteet siellä ovat olleet jo valmiiksi paremmat. Toisaalta Ainolan alueen matalien rantojen suurempi osuus saattaa kertoa siitä, että hoitokalastus vaikuttaakin paremmin matalissa, pienissä järvissä. Tämän tutkimiseksi pitäisi rakentaa BACI-tutkimus, jossa käsittelyjärviä olisi useita.

Käytetyissä sekamalleissa järvi oli satunnainen vaikutus eli järvet ajateltiin satunnaisotokseksi suuremmasta järvien joukosta. Tämä ei ole välttämättä paras oletus. Tuloksia ei ehkä kannata yleistää muihin järviin, vaan tarkastelun kohteena on mieluummin pelkästään Jyväsjärvi. Siinä mielessä järvi vaikutus olisi voitu pitää mallissa kiinteänä faktorina, mutta järvien välisen suuren vaihtelun takia järvi-faktori pidettiin mukana mallin satunnaisosassa. Mikäli

haluttaisiin tutkia hoitokalastuksen vaikutusta esim. pienissä keskisuomalaisissa järvissä, tulisi BACI-asetelmaan arpoa useita hoitokalastettavia järviä em. järvijoukosta.

Tämä tutkimus vahvistaa käsitystä, että biologisten muuttujien suuri vaihtelu asettaa haasteita BACI-asetelmien tutkimiselle. Vaihtelu oli runsasta ennen-vuosien välillä sekä kontrollijärvien välillä. Mahdollisimman monen kontrollialueen liittäminen BACI-tutkimuksiin on tärkeää. Tässä työssä analysoitu Jyväsjärvi-hankeen tuottama BACI-aineisto on biologian näkökulmasta kattava: ennen-vuosia on neljä ja vertailujärviä kolme. BACI-tutkimuksena Jyväsjärven tapaus on harvinainen, koska käsittelyalueita on kaksi. Mittaamisen työläyden takia on usein tyydyttävä vaatimattomampaan aineistoon. Mallintamisen ja testaamisen kannalta havaintomäärä on kuitenkin vielä tässäkin tutkimuksessa suhteellisen pieni. Esimerkiksi autokorrelaattiorakenteen ja heteroskedastisuuden tunnistaminen sekä normaalisuusoletusten voimassaolon tarkastelu olisi suuremmalla havaintomäärällä helpompaa.

Lähdeluettelo

- Davis C. S. 2002: *Statistical Methods for the Analysis of Repeated Measurements*. Springer, New York.
- 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 Krosing C. N. 1959: Estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15: 192–218.
- Meriläinen J. J., Hynynen J., Palomäki A., Mäntykoski K. & Witick A. 2003: Environmental history of an urban lake: A palaeolimnological study of Lake Jyväsjärvi, Finland. *Journal of Paleolimnology* 30: 387–406.
- Nissinen K. 2009: *Small area estimation with linear mixed models from unit-level panel and rotating panel data*. University of Jyväskylä, Department of Mathematics and Statistics, Report 117.
- Pinheiro J. C. & Bates D. M. 2002: *Mixed-Effects Models in S and S-PLUS*. Springer, New York.
- Päijänne LTER - Lake Päijänne LTER. Saatavilla internetistä: <http://www.paijanne.org/>. Viitattu 14.12.2011.
- R Development Core Team 2009: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Saatavilla internetistä: <http://www.R-project.org>.
- Ruuhijärvi J. & Olin M. 2002: Onnistuuko ja vaikuttaako hoitokalastus? *Vesitalous* 6: 38–41.
- Salonen K., Karjalainen J., Högmander P., Keskinen T., Huttula T. & Palomäki A. 2005: Recovery of Lake Jyväsjärvi after pollution by municipal and industrial waste waters. *Verhandlungen des Internationalen Verein Limnologie* 29: 619–622.
- Satterthwaite F. E. 1946: An approximate distribution of estimates of variance components. *Biometrics bulletin* 2: 110–114.
- Syväranta J. 2008: Impacts of biomanipulation on lake ecosystem structure revealed by stable isotope analysis. Saatavilla internetistä: <http://urn.fi/URN:ISBN:978-951-39-3101-8>. Viitattu 14.12.2011.

- Underwood A. J. 1994: On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecological Applications* 4(1): 3–15.
- Wiederholm T. 1980: Use of benthos in lake monitoring. *Journal of the Water Pollution Control Federation* 52: 537–547.

Liite 1: Pohjaeläinaineisto

lake	year	jarvi	aika	alue	paikka	Totden	BQI
JJMB	2000	1	1	2	1	1986.04	1.83
JJMB	2001	1	1	2	1	5017.00	1.32
JJMB	2002	1	1	2	1	2255.92	1.18
JJMB	2003	1	1	2	1	2242.08	1.30
JJMB	2004	1	2004	2	1	4013.60	1.54
JJMB	2005	1	2005	2	1	2510.76	1.23
JJMB	2006	1	2006	2	1	3364.76	1.30
JJMB	2007	1	2007	2	1	2177.70	1.32
JJMB	2008	1	2008	2	1	3356.22	1.30
JJAB	2000	1	1	3	2	1633.12	1.20
JJAB	2001	1	1	3	2	3453.08	1.70
JJAB	2002	1	1	3	2	1294.04	1.87
JJAB	2003	1	1	3	2	2013.72	1.38
JJAB	2004	1	2004	3	2	650.48	1.92
JJAB	2005	1	2005	3	2	341.60	1.75
JJAB	2006	1	2006	3	2	572.18	1.10
JJAB	2007	1	2007	3	2	589.26	1.60
JJAB	2008	1	2008	3	2	742.98	1.75
ALVA	2000	2	1	1	3	3385.86	2.00
ALVA	2001	2	1	1	3	4898.37	2.00
ALVA	2002	2	1	1	3	11413.06	1.86
ALVA	2003	2	1	1	3	12411.51	2.00
ALVA	2004	2	2004	1	3	2995.37	1.00
ALVA	2005	2	2005	1	3	4727.50	1.33
ALVA	2006	2	2006	1	3	8777.90	1.00
ALVA	2007	2	2007	1	3	6136.60	1.00
ALVA	2008	2	2008	1	3	2238.70	2.25
PATA	2001	3	1	1	4	1020.70	2.00
PATA	2002	3	1	1	4	726.60	2.50
PATA	2003	3	1	1	4	29.66	2.00
PATA	2004	3	2004	1	4	756.26	0.00
PATA	2005	3	2005	1	4	268.40	0.00
PATA	2006	3	2006	1	4	695.40	0.00
PATA	2007	3	2007	1	4	286.70	0.00
PATA	2008	3	2008	1	4	183.00	2.75
TUOMIO	2000	4	1	1	5	790.86	1.61
TUOMIO	2001	4	1	1	5	647.51	1.00
TUOMIO	2002	4	1	1	5	1166.51	0.00
TUOMIO	2003	4	1	1	5	855.11	1.10
TUOMIO	2004	4	2004	1	5	1601.49	1.00
TUOMIO	2005	4	2005	1	5	921.10	1.10
TUOMIO	2006	4	2006	1	5	536.80	1.09
TUOMIO	2007	4	2007	1	5	366.00	0.00
TUOMIO	2008	4	2008	1	5	616.10	1.00

Liite 2: R-koodi

```
pohjis <- read.table("U:/gradu/datat/pohjisdata.txt",
  header = TRUE, sep = "\t") # Ladataan aineisto
pohjis$jarvi <- factor(pohjis$jarvi) # Faktoroidaan muuttujat
pohjis$aika <- factor(pohjis$aika)
pohjis$alue <- factor(pohjis$alue)
pohjis$paikka <- factor(pohjis$paikka)
attach(pohjis) # Kiinnitetään objekti pohjis
library(nlme) # Ladataan sekamallikirjasto

# Vastemuuttujana pohjaeläinten kokonaisuksilöitiheys (Totden)

# Piirretään profiilikuva

interaction.plot(x.factor=pohjis$year,trace.factor=pohjis$lake,
  response=pohjis$Totden,fun=mean,type="b",
  col=c("blue","red","red3","blue","blue"),lty=c(2,1,1,2,2),
  xlab="",ylab="Yksilöitiheys",lwd=4,fixed=F,legend=F)
legend("topright",
  c("Alvajärvi","JJ, Ainola","JJ, Pääsyvä","Patajärvi","Tuomiojärvi"),
  pch=c("1","2","3","4","5"), bty="n",
  col=c("blue","red","red3","blue","blue"), title=NULL)
polygon(c(5,7,7,5), c(-900,-900,-490,-490),col="grey",xpd=NA,)
polygon(c(1,9,9,1), c(-900,-900,-490,-490),xpd=NA)

# Lasketaan keskiarvoja

kka <- round(tapply(Totden[alue==1], aika[alue==1], mean),0)
kka[1] # Kontrollit ennen
round(mean(Totden[aika!=1 & alue==1]),0) # Kontrollit jälkeen
round(mean(Totden[alue==1]),0) # Kontrollit yht.

pska <- round(tapply(Totden[alue==2], aika[alue==2], mean),0)
pska[1] # Pääsyväne ennen
round(mean(Totden[aika!=1 & alue==2]),0) # Pääsyväne jälkeen
round(mean(Totden[alue==2]),0) # Pääsyväne yht.

aka <- round(tapply(Totden[alue==3], aika[alue==3], mean),0)
aka[1] # Ainola ennen
round(mean(Totden[aika!=1 & alue==3]),0) # Ainola jälkeen
round(mean(Totden[alue==3]),0) # Ainola yhteensä

jjka <- round(tapply(Totden[alue==2 | alue==3],
  aika[alue==2 | alue==3],mean),0)
jjka[1] # JJ ennen
round(mean(Totden[aika!=1 & (alue==2 | alue==3)]),0) # JJ jälkeen
round(mean(Totden[alue==2 | alue==3]),0) # JJ yhteensä

round(mean(Totden[aika==1]),0) # Ennen yht.
round(mean(Totden[aika!=1]),0) # Jälkeen yht.
round(mean(Totden),0) # Kaikki yhteensä

# Sovitetaan erilaisia malleja

td1 <- lme(Totden ~ aika*alue, data=pohjis,random=~1|jarvi)
td2 <- lme(Totden ~ aika*alue, data=pohjis,
```

```

        random = list(jarvi = pdSymm(~jarvi-1)) )
td3 <- lme(Totden ~ aika*alue, data=pohjis,
        random = list(jarvi = pdDiag(~jarvi-1)) )
td4 <- update(td1, weights=varIdent(form=~1|jarvi))
td5 <- update(td1, corr=corAR1(form=~1|jarvi/paikka))
td6 <- update(td1, weights=varPower())
td7 <- update(td4, corr=corAR1(form=~1|jarvi/paikka))
td8 <- update(td3, weights=varPower())

# Mallien vertailua

anova(td1,td2,td3,td4,td5,td6,td7,td8,test=F)
anova(td1,td6)
plot(td1)
plot(td6)

summary(td6) # Mallin kertoimet ym.

# F-testit

k.jj <- c(0,0,0,0,0,0,0,0, 0.1,0.1,0.1,0.1,0.1,
        0.1,0.1,0.1,0.1,0.1)
k.ps <- c(0,0,0,0,0,0,0,0, 1/5,1/5,1/5,1/5,1/5, 0,0,0,0,0)
k.ai <- c(0,0,0,0,0,0,0,0, 0,0,0,0,0, 1/5,1/5,1/5,1/5,1/5)
anova(td6,type="marginal") # kiinteille kertoimille
anova(td6, L=k.jj) # Jyväsjärven kontrastille
anova(td6, L=k.ps) # Pääsyvänteen kontrastille
anova(td6, L=k.ai) # Ainolan kontrastille

# Jäännöskuvaajia

plot(fitted(td6),resid(td6,type="p"),
xlab="Sovitetut arvot",ylab="Standardoidut jäännökset")
abline(h=0)

qqnorm(td6, abline=c(0,1),xlab="Standardoidut jäännökset",
ylab="Normaalijakauman kvantiilit", xlim=c(-3,3))

hist(resid(td6, type="p"),probability=T, ylim=c(0,1.0),
xlab="Standardoidut jäännökset",ylab="Tiheys",main="")
curve(dnorm,add=T)

# Vastemuuttujana pohjanlaatuindeksi (BQI)

# Piirretään profiilikuva

interaction.plot(x.factor=pohjis$year,trace.factor=pohjis$lake,
response=pohjis$BQI,fun=mean,type="b",
col=c("blue","red","red3","blue","blue"),
lty=c(2,1,1,2,2), xlab="",ylab="BQI",lwd=4,fixed=F, legend=F)
legend("top",
c("Alvajärvi","JJ, Ainola","JJ, Pääsyvä","Patajärvi","Tuomiojärvi"),
pch=c("1","2","3","4","5"), bty="n",
col=c("blue","red","red3","blue","blue"), title=NULL)
polygon(c(5,7,7,5), c(-0.20,-0.20,-0.11,-0.11),col="grey",xpd=NA,)
polygon(c(1,9,9,1), c(-0.20,-0.20,-0.11,-0.11),xpd=NA)

# Lasketaan keskiarvoja

kka <- round(tapply(BQI[alue==1], aika[alue==1], mean),0)
kka[1] # Kontrollit ennen
round(mean(BQI[aika!=1 & alue==1]),0) # Kontrollit jälkeen
round(mean(BQI[alue==1]),0) # Kontrollit yht.

pska <- round(tapply(BQI[alue==2], aika[alue==2], mean),0)
pska[1] # Pääsyvänteen ennen
round(mean(BQI[aika!=1 & alue==2]),0) # Pääsyvänteen jälkeen

```

```

round(mean(BQI[alue==2]),0) # Pääsyväne yht.

aka <- round(tapply(BQI[alue==3], aika[alue==3], mean),0)
aka[1] # Ainola ennen
round(mean(BQI[aika!=1 & alue==3]),0) # Ainola jälkeen
round(mean(BQI[alue==3]),0) # Ainola yhteensä

jjka <- round(tapply(BQI[alue==2 | alue==3],
  aika[alue==2 | alue==3],mean),0)
jjka[1] # JJ ennen
round(mean(BQI[aika!=1 & (alue==2 | alue==3)]),0) # JJ jälkeen
round(mean(BQI[alue==2 | alue==3]),0) # JJ yhteensä

round(mean(BQI[aika==1]),0) # Ennen yht.
round(mean(BQI[aika!=1]),0) # Jälkeen yht.
round(mean(BQI),0) # Kaikki yhteensä

# Sovitetaan erilaisia malleja

bqi1 <- lme(BQI ~ aika*alue, data=pohjis,random=~1|jarvi)
bqi2 <- update(bqi1, random = list(jarvi = pdSymm(~jarvi-1)) )
bqi3 <- update(bqi1, random = list(jarvi = pdDiag(~jarvi-1)) )
bqi4 <- update(bqi1, weights=varIdent(form=~1|jarvi))
bqi5 <- update(bqi1, corr=corAR1(form=~1|jarvi/paikka))
bqi6 <- update(bqi3, weights=varIdent(form=~1|jarvi))
bqi7 <- update(bqi4, corr=corAR1(form=~1|jarvi/paikka))

anova(bqi1,bqi2,bqi3,bqi4,bqi5,bqi6,
  bqi7,test=F) # Mallien vertailua
anova(bqi4, bqi7)
plot(bqi1)
plot(bqi4)
plot(bqi7)

summary(bqi4) # Mallin kertoimet ym.

anova(bqi4,type="marginal") # F-testit kiint. kertoimille
anova(bqi4, L=k.jj) # ja kontrasteille
anova(bqi4, L=k.ps)
anova(bqi4, L=k.ai)

# Jäännösten kovarianssimatriisi

getVarCov(bqi4,individual = 1,
  type="conditional") # Jyväsjärvi
getVarCov(bqi4, 2, type="c") # Alvajärvi
getVarCov(bqi4, 3, type="c") # Patajärvi
getVarCov(bqi4, 4, type="c") # Tuomiojärvi

# Jäännöskuvaajia

plot(fitted(bqi4),resid(bqi4,type="p"),
  xlab="Sovitetut arvot",
  ylab="Standardoidut jäännökset")
abline(h=0)

qqnorm(bqi4, abline=c(0,1),xlab="Standardoidut jäännökset",
  ylab="Normaalijakauman kvantiilit", xlim=c(-3,3))

hist(resid(bqi4, type="p"),probability=T, xlim=c(-3,3),
  xlab="Standardoidut jäännökset",ylab="Tiheys",main="")
curve(dnorm,add=T)

```