

Informaatioarvo ja sekamallit – sovellus järvien ekologiseen tilaluokitteluun

Tilastotieteen pro gradu -tutkielma

11. heinäkuuta 2018

Vilja Koski

Matematiikan ja tilastotieteen laitos

Jyväskylän yliopisto

JYVÄSKYLÄN YLIOPISTO

Matematiikan ja tilastotieteen laitos

Koski, Vilja: *Informaatioarvo ja sekamallit – sovellus järvien ekologiseen tilaluokitteluun*

Tilastotieteen pro gradu -tutkielma (39 sivua)

11. heinäkuuta 2018

Tiivistelmä

Suomessa toteutetaan vesistöjen suojelemiseksi ekologisen tilan seurantaa, jonka perusteella tehdään päätöksiä mahdollisista kunnostustoimenpiteistä. Tässä työssä käytetään informaatioarvon käsitettä (*value of information, VOI*) selvittämään, mikä on järven seurannan arvo. Informaatioarvo on päätösteorian käsite lisäaineiston arvon arvioimiseksi. Se tarkoittaa maksimihintaa, joka päätöksentekijän kannattaa vielä maksaa lisäinformaatiosta päätöksen tekemiseksi. Käsitettä sovelletaan Hiidenveden klorofyllipitoisuuden seuranta-aineistoon ja selvitetään, kuinka paljon lisäaineistosta kannattaa maksaa, kun Hiidenveden ekologisesta tilasta on jo ennakkokäsitys. Aineiston on kerännyt Suomen ympäristökeskus (SYKE).

Informaatioarvoja lasketaan sekä täydelliselle että epätäydelliselle informaatiolle. Täydellisellä informaatiolla tarkoitetaan sellaista tietoa, jolla saadaan täysin varma käsitys ekologisesta tilasta. Tässä työssä tutkitaan myös, miten epätäydellinen informaatio, eli toisin sanoen kerättävän informaation luotettavuus, vaikuttaa informaatioarvoon. Se tehdään lineaarisen sekamallin avulla. Hiidenveden havaintoihin sovitetaan lineaarinen sekamalli, josta saadaan havaintojen odotusarvon estimaattien jakauma. Saatua jakaumaa verrataan Hiidenveden vertailujärvien klorofyllihavaintojen keskiarvojen empiiriseen jakaumaan, jolloin saadaan ehdolliset todennäköisyydet ennustaa järven ekologinen tila ehdolla oikea tila. Lisäksi tehdään sensitiivisyystarkasteluja, miten ennakkokäsitys ja ehdolliset todennäköisyydet vaikuttavat informaatioarvoon.

Tulosten perusteella nähdään, että mitä luotettavampaa kerättävä tieto järvestä on, sitä enemmän siitä kannattaa maksaa. Resursseja ei kannata kohdentaa sellaisiin järviin, joissa ennakkotieto järven ekologisesta tilasta on jo vahva. Sen sijaan tietoa järven ekologisesta tilasta kannattaa hankkia sellaisista järvistä, joista ennakkotietoa ekologisesta tilasta ei vielä ole tai se on hyvin epävarmaa. Erityisesti informaatio on arvokkaampaa silloin, kun ollaan ennakkoon varmempia siitä, että järvi on paremmassa kuin että se on huonommassa tilassa. Tällöin kannattaa kerätä informaatiota ja varmistua siitä, että järvi on hyvässä tilassa, jolloin kalliit kunnostustoimenpiteet järvelle voidaan jättää tekemättä.

Avainsanat: ekologinen tila, epätäydellinen informaatio, Hiidenvesi, informaatioarvo, lineaarinen sekamalli, päätöksenteko

Sisältö

1	Johdanto	1
2	Aineisto ja tutkimusongelma	4
3	Informaatioarvon teoriaa	7
3.1	Päätöksenteko	7
3.2	Informaatioarvo	10
3.2.1	Informaatioarvo täydelliselle informaatiolle	10
3.2.2	Informaatioarvo epätäydelliselle informaatiolle	12
3.3	Informaatioarvo vesistöjen seurannoissa	13
3.4	Informaatioarvon soveltaminen Hiidenveden seuranta-aineistoon	15
4	Sekamallien teoriaa	18
4.1	Lineaarinen sekamalli	18
4.2	Monitasomalli	19
4.3	Mallin valinta	22
5	Tulokset	24
5.1	Informaation luotettavuuden arvioiminen	24
5.2	Informaatioarvot Hiidenvedelle	26
5.3	Informaatioarvoon vaikuttavat tekijät	30
6	Yhteenveto	33
	Viitteet	37

1 Johdanto

Huoli vesistöjen ekologisesta tilasta on aiheellista. Vesistöjen tilaa uhkaa esimerkiksi rehevöityminen, joka tarkoittaa vesistön liiallisesta ravinteiden saannista johtuvaa kasvillisuuden liikakasvua. Rehevöitymistä aiheuttavat muun muassa fosfori ja typi, jotka kulkeutuvat vesistöihin laajoilta alueilta kaupungeista ja maataloudesta. Vesiekosysteemeissä nämä ravinteet aiheuttavat monenlaisia ongelmia, kuten myrkyllisiä leväkukintoja, happikatoa, kalojen kuolemia, biologisen monimuotoisuuden vähenemistä sekä vesikasvien ja koralliriuttojen vähenemistä. Liialliset ravintoaineet vaurioittavat siten vakavasti vesiekosysteemejä ja heikentävät mahdollisuutta käyttää vettä esimerkiksi juomavetenä, teollisuudessa, maanviljelyssä ja vapaa-aikana. (Carpenter et al., 1998.)

Euroopan Unionin tasolla sisävesistöjen suojeluun on tartuttu vesipuidedirektiivin (VPD) avulla. Vesipuidedirektiivi on Euroopan parlamentin direktiivi EU:n vesipolitiikan suuntaviivoista. Jo vuonna 2000 voimaan tulleen direktiivin tavoitteena on ollut suojella ja tarvittaessa kunnostaa vesistöjä niin, että vesistöjen tila ei heikkene ja että niiden tila on vähintään hyvä koko EU-alueella vuoteen 2015 mennessä. Osatavoitteiden saavuttamista voidaan jatkaa vuoteen 2027 asti. Direktiivin tavoitteiden saavuttamiseen tarvittavat toimet on jätetty jäsenmaiden itsensä suunniteltavaksi ja hoidettavaksi. (Euroopan parlamentti, 2000.)

Suomessa sisävesistöjen ekologisen ja kemiallisen tilan seuraamiseksi toteutetaan vesipuidedirektiivin mukaista seurantaa. Ympäristön seuranta on säännöllistä ja pitkäaikaista tietojen keräämistä ympäristöstä. Sillä pyritään erottamaan ihmisen toimien aiheuttamat ympäristömuutokset luonnollisesta vaihtelusta. Seurannan avulla toteutetaan vesistöjen luokittelua ekologiin tilaluokkiin. Ensimmäinen seuranta- ja luokittelujärjestelmä valmistui vuonna 2008 ensimmäistä suunnittelukautta (2010–2015) varten. Luokittelu toteutettiin toisen kerran vuoden 2012 aikana toista suunnittelukautta (2016–2021) varten (Aroviita et al., 2012). Tässä työssä käytetty seuranta-aineisto on vuosilta 2006–2012, eli aineisto, jonka perusteella toisen kierroksen luokittelu tehtiin.

Luokittelujärjestelmän avulla vesistöt luokitellaan viiteen laatuluokkaan. Luokat ovat erinomainen, hyvä, tyydyttävä, välttävä ja huono. Indikaattoreina vesistön laadun selvittämiseksi käytetään kasviplanktonia, vesikasveja, piileviä, pohjaeläimiä, kalakantoja, fysikaalis-kemiallisia tekijöitä sekä hydrologis-morfologisia tekijöitä. Tässä työssä rajoitutaan järvien luokitteluun ainoastaan kasviplanktonin, ja tarkemmin sen yhden indikaattorimuuttujan, eli a-klorofyllin osalta. Järvien eko-

logista tilaa pystytään karkeasti arvioimaan klorofyllipitoisuuden avulla. Perusteet erityyppisten järvien luokitteluun klorofyllipitoisuuden avulla ovat erilaiset, koska esimerkiksi runsashumuksisissa järvissä pitoisuus on luontaisesti korkeampi kuin vähähumuksisissa järvissä. (Vuori et al., 2009.)

Nykyinen järvien seuranta perustuu kertaluontoisiin näytteenottoihin ennalta määritetyistä havaintopaikoista sekä laboratorioissa tehtäviin analyyseihin. Viime vuosina on kuitenkin ollut painetta havaintopaikkojen vähentämiseen (Kotamäki, 2015). Tämän työn tarkoituksena on selvittää, kuinka arvokasta seurannoista kerättävä tieto on kunnostuspäätösten tekemiseksi. Tähän vastataan informaatioarvon käsitteen avulla (*value of information, VOI*) (Eidsvik et al., 2015).

Informaatioarvo on päätösteorian käsite hankitun lisätiedon arvon analysoimiseksi. Oikean tiedon kerääminen ja sen riittävä, optimaalinen määrä ovat ratkaisevan tärkeitä kaikissa päätöksentekoprosesseissa. Monissa ympäristösovelluksissa harkitaan lisätietojen hankkimista ennen päätösten tekemistä. Keskeinen kysymys tiedon tuottamisessa on, kuinka paljon tietoja pitäisi hankkia ja mihin hintaan. Lisätiedot voivat auttaa vähentämään epävarmuutta ongelman ratkaisussa, mutta jos lisätiedolla ei ole vaikutusta lopputulokseen eikä sitä kautta syntyvään päätökseen, tiedon hankkiminen ei ole taloudellisesti järkevää. Informaatioarvo on tapa arvioida aineiston arvo ennen sen hankkimista ja käsittelemistä. Käsite on jo tunnettu ja paljon käytetty lääketieteessä sekä talous- ja rahoitusalueilla, mutta menetelmä on vasta tulossa suosituksi ekologiassa ja biologian aloilla. (Eidsvik et al., 2015.)

Informaation arvoa arvioiville tutkimuksille vaikuttaisi olevan tarvetta niin ympäristön kuin muidenkin alojen tutkimuksessa. Sopiva lähestymistapa, joka olisi sekä teoreettisesti järkevä että empiirisesti toteuttamiskelpoinen, näyttää kuitenkin puuttuvan (Bouma et al., 2009). Bouma et al. (2009) ehdottavat tällaista uutta lähestymistapaa. He käyttävät informaatioarvon teoriaa Pohjanmeren veden laadunhallinnan tietokantojen arvon arvioimisessa. Heidän tavoitteenaan on arvioida, parantaako satelliittitietojen käyttö meriveden laaduntarkkailujärjestelmää, ja tätä varten he kehittävät kehikon metodille, jolla voidaan arvioida informaation merkitystä päätöksenteossa epävarmuuden vallitessa.

Nygård et al. (2016) taas tarkastelevat Suomen merivesiseurannan kustannuksia ja käyttävät niiden arviointiin informaatioarvon käsitettä. Heidän tavoitteenaan on arvioida seurannan kokonaiskustannuksia käyttämällä hyväksi merivesiseurannasta koituvia taloudellisia kuluja sekä ottamalla huomioon myös seurannasta saatavat taloudelliset hyödyt. Tutkimus osoittaa, että merien seurantaan käytettävät taloudel-

liset resurssit ovat suuruusluokaltaan pieniä sen saavuttamiin taloudellisiin hyötyihin nähden. Heidän kehittämänsä käsitteellistä mallia vesistöistä kerättävän informaation arvon arvioimiseksi käytetään hyväksi myös tässä työssä, ja siitä kerrotaan tarkemmin luvussa 3.3.

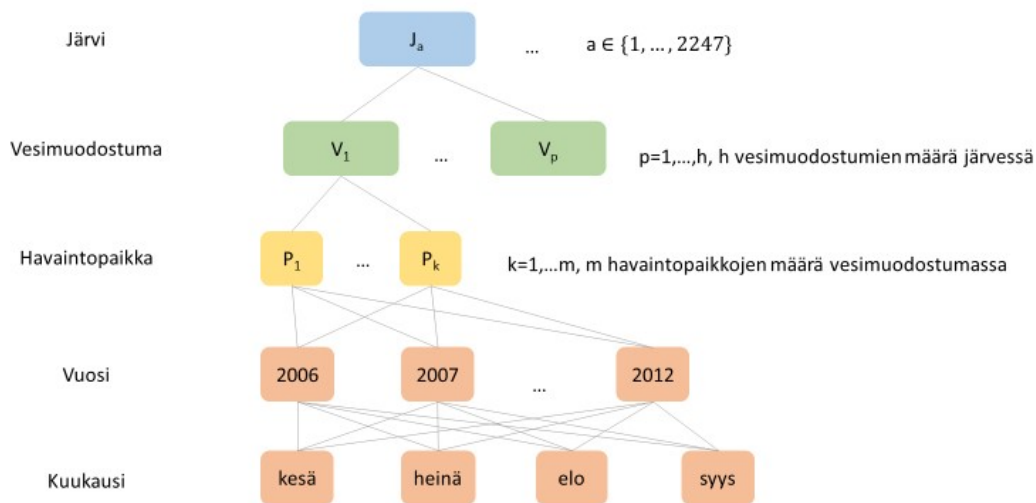
Nygård et al. (2016) tarkastelevat informaation arvoa ainoastaan tapauksessa, jossa kerättävä informaatio antaa täysin varman tiedon meren tilasta. Informaatiota kutsutaan tällöin täydelliseksi. Kuitenkin myös epätäydellisestä informaatiosta voi olla hyötyä päätöksenteossa, eli kun kerättävään informaatioon liittyy epävarmuutta. Tässä työssä tutkitaan täydellisen informaation lisäksi, miten epätäydellistä informaatiota voidaan käyttää hyödyksi informaatioarvon laskemisessa. Epätäydellisen informaation hyödyntämiseksi on arvioitava kerättävän informaation luotettavuutta. Se tehdään lineaarisen sekamallin avulla. Lineaarinen sekamalli on lineaarisen mallin laajennus (Demidenko, 2013). Se soveltuu riippuvien ja hierarkkisten aineistojen mallintamiseen, kuten tässä työssä käytettävässä seuranta-aineistossa, jonka rakenne on pitkittäisaineiston kaltainen. Kerättävän tiedon luotettavuudella tarkoitetaan tässä työssä ehdollisia todennäköisyyksiä ennustaa järven ekologinen tila klorofyllimittauksen avulla ehdolla järven oikea ekologinen tila. Luvussa 5.1 on esitetty menetelmä ehdollisten todennäköisyyksien arvioimiseksi lineaarisen sekamallin avulla. Järven klorofyllihavaintoihin sovitetaan lineaarinen sekamalli, josta saadaan havaintojen odotusarvon estimaattien jakauma. Saatua jakaumaa verrataan järven vertailujärvien ekologisen tilan luokkarajoihin. Lisäksi luvussa 5.3 on tehty sensitiivisyystarkasteluja, miten sekä prioritieto että tiedon luotettavuus eli ehdolliset todennäköisyydet vaikuttavat informaatioarvoon.

Tämän tutkimuksen rakenne on pääpiirteissään seuraava. Luvussa 2 esitellään tutkimuksessa käytettävää Suomen ympäristökeskuksen (SYKE) aineistoa ja muotoillaan tarkemmin tutkimuskysymys. Luvuissa 3 ja 4 perehdytään käytettäviin menetelmiin ja niiden teoriaan, eli luvussa 3 informaatioarvon teoriaan ja luvussa 4 sekamallien teoriaan. Luvussa 5 edetään tulosten esittelyyn ja luku 6 on varattu yhteenvedolle.

2 Aineisto ja tutkimusongelma

Tässä työssä käytetyn aineiston on kerännyt Suomen ympäristökeskus (SYKE). Aineisto kuuluu Suomen kansalliseen vesien laadun seurantaan, ja tietoja tallennetaan jatkuvasti Suomen ympäristökeskuksen tietokantaan (<http://www.syke.fi/avointieto>). Seuranta toteutetaan ennalta määrättyissä paikoissa koko Suomen alueen vesistöistä. Tässä työssä rajoitutaan tarkastelemaan vain järviä. Järvet lohkotaan vielä eri vesimuodostumiin, koska yksittäinen monimutkaisen muotoinen tai suuri järvi voi sisältää hyvin erityyppisiä ympäristöjä. Vesimuodostuma on siis koko järvi tai rajattu, vedenlaadultaan homogeenisempi osa järveä. Useimmilla vesimuodostumilla on muutama havaintopaikka, josta näytteitä on kerätty, tyypillisimmin yksi tai kaksi paikkaa. Havaintopaikkoja saattaa kuitenkin olla vesimuodostumassa enemmänkin, esimerkiksi suurilla järvillä tai jos havaintopaikkoja on vaihdeltu vuosittain. Näytteenotto ei ole kaikissa järvissä jatkuvaa, vaan seurantoja toteutetaan kertaluontoisesti eri järvissä asiantuntijoiden näkemysten mukaisesti. Seuranta-aineiston rakennetta on havainnollistettu kuvassa 1.

Järvet jaetaan 13 järvityyppiin, jotka perustuvat EU:n vesipuitedirektiivissä määriteltäviin periaatteisiin. Ensimmäisenä omiksi tyypeikseen erotetaan luontaisesti runsaravinteiset järvet, luontaisesti runsaskalkkiset järvet sekä Pohjois-Lapin järvet. Loput jäljelle jäävät järvet luokitellaan ryhmiin niiden pinta-alan, keskisyvyyden, veden värin ja retentioajan mukaan (Kotamäki et al., 2015). Järven retentioajalla

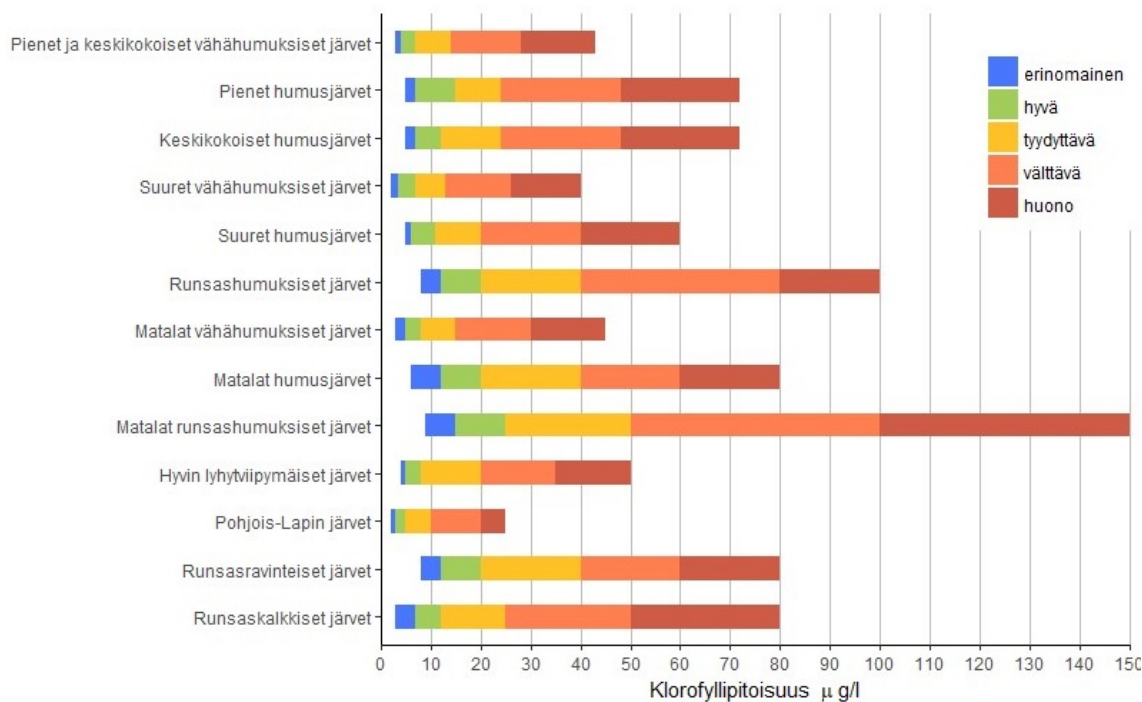


Kuva 1: Seuranta-aineiston hierarkkinen rakenne.

tarkoitetaan, kuinka kauan kuluu aikaa ennen kuin järven vesi vaihtuu kokonaan.

Seurannoissa järvistä mitataan muun muassa veden a-klorofyllipitoisuus. Se kuvaa kasviplanktonin kokonaismäärää eli biomassaa. Pitoisuus ilmoitetaan aineistossa yksikössä $\mu\text{g/l}$. Klorofyllipitoisuuden avulla saadaan arvioitua järven ekologista tilaa, suurempi klorofyllipitoisuus viittaa rehevöitymiseen ja siten huonompaan ekologiseen tilaan (Aroviita et al., 2012; Vuori et al., 2009).

Järvet luokitellaan kuuden vuoden välein muun muassa klorofyllipitoisuuden avulla vesipuitedirektiivin mukaisiin tilaluokkiin, jotka ovat erinomainen, hyvä, tyydyttävä, välttävä ja huono. Eri järvityypeille tilaluokkien rajat ovat erilaiset, koska järvityypit eroavat ominaisuuksiltaan. Esimerkiksi runsashumuksisissa järvissä pitoisuus on luontaisesti korkeampi kuin vähähumuksisissa järvissä. Järvityypit sekä niiden luokkarajat klorofyllipitoisuudelle on esitetty kuvassa 2 (Vuori et al., 2009). Esimerkiksi runsasravinteisille järville erinomaisen tilan rajat ovat 8–12 $\mu\text{g/l}$, hyvän rajat 12–20 $\mu\text{g/l}$, tyydyttävän rajat 20–40 $\mu\text{g/l}$, välttävän rajat 40–60 $\mu\text{g/l}$ ja huonon rajat 60–80 $\mu\text{g/l}$. Järvien ekologisen tilan luokittelua tehdään myös muiden muuttujien perusteella, esimerkiksi pohjaeläimiä ja kalakantoja käyttäen (Aroviita et al., 2012). Tässä tutkimuksessa keskitytään järvien luokitteluun ainoastaan klorofyllin avulla. Tässä työssä informaatioarvon tarkastelu rajoitetaan vain yhdelle Suomen järvelle. Esimerkijärveksi valikoitui Hiidenvesi Etelä-Suomessa, Uudenmaan maakunnan



Kuva 2: Vuoden 2008 ekologisen luokituksen luokkarajat klorofyllipitoisuudelle järvityypeittäin.

länsiosassa, Vihdin ja Lohjan kuntien alueilla. Se kuuluu järvityypiltään runsasravinteisiin järviin. Ensimmäisellä luokittelukierroksella vuonna 2008 Hiidenveden tila luokiteltiin tyydyttäväksi. Toisella luokittelukierroksella, joka valmistui vuonna 2012, tila ei ollut parantunut, ja Hiidenvesi luokiteltiin yhä tyydyttäväksi (Ahtiainen, 2008). Hiidenveden ekologista tilaa on tutkittu paljon ja siitä löytyy runsaasti erilaista kirjallisuutta (Hyytiäinen, 2008). Tässä työssä hyödynnetään Hiidenveden klorofyllin seuranta-aineistoa vuosilta 2006–2012, eli aineistoa, jonka perusteella toisen kierroksen luokittelu tehtiin. Klorofyllihavainnot on kasvukauden ajalta kesäkuusta syyskuuhun.

Tämän työn tarkoituksena on avata informaatioarvon käsitettä ja soveltaa sitä Hiidenveden seuranta-aineistoon. Informaatioarvon avulla voidaan selvittää, kuinka paljon kannattaa maksaa lisätiedosta järven tilan selvittämiseksi, kun tilasta on jo jonkinlainen ennakkokäsitys. Lisäksi halutaan tutkia, miten kerättävän informaation luotettavuus vaikuttaa informaatioarvoon. Tiedon luotettavuutta arvioidaan lineaarisen sekamallin avulla, joka sovitetaan Hiidenveden klorofyllin seuranta-aineistoon. Tämän työn tarkoituksena ei siis ole vielä suunnitella optimaalista, kattavaa näytteenottojärjestelmää, jolla järvien ekologista tilaa voitaisiin seurata.

3 Informaatioarvon teoriaa

Tässä luvussa käydään läpi päätöksentekoanalyysin sekä informaatioarvon teoriaa. Luvussa 3.1 käsitellään ensin päätöksentekoa yleisesti sekä päätöksentekoon liittyvää teoriaa, joka johdattaa luvussa 3.2 esitettyyn informaatioarvon teoriaan. Luvussa 3.3 esitellään, miten informaatioarvoa on aiemmin käytetty vesistöistä kerättävän informaation arvon määrittämiseen, minkä jälkeen luvussa 3.4 sovelletaan esiteltyä teoriaa aineistoon. Teoria ja merkinnät ovat teoksesta *Value of Information in the Earth Sciences* (Eidsvik et al., 2015).

3.1 Päätöksenteko

Päätöksenteko on valitsemista – usein toisensa poissulkevien – vaihtoehtojen välillä. Päätöksenteolla tarkoitetaan usein vain hallinnollista päätöksentekoa, vaikka termin ei tarvitse rajoittua koskemaan pelkästään hallintoa. Päätöksentekoanalyysi on tutkimusala, joka noudattaa päätöksentekoteoriaa ja jolla pyritään auttamaan päätöksentekijää tärkeissä käytännön päätöksentekotilanteissa. Termin päätösanalyysi (*decision analysis*) otti käyttöön Ronald Howard (1964). Termi antoi täsmällisen määritelmän analyysille, joka arvioi loogisesti käytettävissä olevia erilaisia vaihtoehtoja päätöksentekotilanteissa, ottaen samalla huomioon mahdolliset epävarmuustekijät sekä päätöksentekijän omat mieltymykset. Tarkoituksena on siis ohjata päätöksentekijää tekemään parempia päätöksiä. Päätöksentekoa tutkitaan monilla eri aloilla, esimerkiksi psykologiassa ja politiikan tutkimuksessa. Siksi samasta asiasta on käytössä useita eri nimiä; puhutaan esimerkiksi maksimaalisesta odotetusta hyödystä (*maximum expected utility*). (Eidsvik et al., 2015, s. 64-67.)

Seuraavaksi käydään läpi päätöksentekotilanteisiin liittyviä merkintöjä. Ensimmäiseksi luokitellaan päätöksentekotilanteeseen liittyvät muuttujat sen mukaan, pystyykö päätöksentekijä vaikuttamaan muuttujan arvoon. Muuttujien oletetaan olevan tässä diskreettejä. Sellaista muuttujaa, jonka arvoa päätöksentekijä pystyy kontrolloimaan, kutsutaan päätökseksi. Nimitetään päätöksen saamia arvoja vaihtoehtoiksi tai toimiksi (*alternative, action*), ja merkitään niitä kirjaimella a . Päätöksentekijä voi valita minkä tahansa vaihtoehdon joukosta $a \in A$. Vastaavasti sellaista muuttujaa, johon päätöksentekijä ei pysty vaikuttamaan, kutsutaan epävarmuustekijäksi. Epävarmuustekijän saamia arvoja kutsutaan tiloiksi tai realisaatioiksi (*state, realization*). Merkitään epävarmuustekijän arvoja kirjaimella $x \in \Omega$. Epävarmuustekijä voi realisoitua miksi tahansa otosavaruuden tilaksi todennäköisyydellä $p(x)$. Epä-

varmuustekijät ovat siten kiinnostuksen kohteena olevia satunnaismuuttujia.

Kun kaikki tilanteeseen liittyvät muuttujat ovat tunnistettu, voidaan alkaa tarkastella niistä aiheutuvia erilaisia skenaarioita (*scenario*). Skenaario tarkoittaa jokaisen muuttujan ilmentymää päätöksentekotilanteessa. Tilanteeseen liittyy aina $|\Omega| \times |A|$ kappaletta eri skenaarioita. Jokaiseen erilaiseen skenaarioon liittyy omanlaisensa lopputulos (*outcome*). Jokaiseen lopputulokseen yhdistetään vielä päätöksentekijän asettama arvo (*value*). Arvo tarkoittaa lopputuloksen hyötyä päätöksentekijälleen sen toteutuessa. Helpointa on ymmärtää arvo rahana, mutta muitakin yksiköitä voidaan käyttää, jos yksikkö on selkeästi ymmärrettävissä arvona. Päätöksentekijän asettamia arvoja eri lopputuloksille merkitään arvofunktiolla, $v(\cdot)$. Arvo määräytyy siis päätöksestä a sekä epävarmuustekijöistä x , jolloin voidaan merkitä $v(x, a)$. Eri arvofunktion saamat arvot voidaan taulukoida. Arvotaulukon avulla nähdään, miten päätöksentekijä arvottaa jokaisen lopputuloksen.

Hyötyfunktio $u(\cdot)$ on arvofunktion laajennus, joka ottaa huomioon myös päätöksentekijän kyvyn sietää riskiä. Toisin sanoen arvofunktio kuvaa varmaan lopputulokseen liittyvää arvoa, ja hyötyfunktio kuvaa arvoa, kun lopputulokseen liittyy epävarmuutta. Päätöksentekijä saattaa olla joko riskiä karttava (*risk averse*), riskiä hakeva (*risk seeking*) tai jotain siltä väliltä. Usein oletetaan riskineutraali (*risk neutral*) päätöksentekijä. Riskineutraalille päätöksentekijälle hyötyfunktio on lineaarinen, ja voidaan siis kirjoittaa muodossa

$$u(v) = \alpha + \beta v, \quad (1)$$

jossa α ja β ovat vakioita ja $\beta > 0$. Riski on siis päätöksentekijän ominaisuus, ei päätöksentekotilanteen. Hyötyfunktio $u(v)$ kuvaa päätöksentekijän saavuttamaa hyötyä arvosta $v(a, x)$.

Hyötyfunktiolle voidaan ottaa käyttöön vastaava, mutta konkreettisempi muuttuja. Merkitään tätä muuttujaa CE (*certain equivalent* tai *certainty equivalent*). Taloustieteessä siitä käytetään nimitystä varmuusekvivalentti. Varmuusekvivalentti tarkoittaa sitä hintaa, jolla päätöksentekijä on valmis luopumaan päätöksentekotilanteestaan. Toisin sanoen se on minimihinta, jolla päätöksentekijä on halukas myymään mahdollisesti saavuttamansa hyödyn päätöksentekotilanteestaan eteenpäin. Sitä suuremmalla rahamäärällä päätöksentekijän kannattaisi ehdottomasti luopua tilanteesta, koska arvioi saavuttamansa hyödyn pienemmäksi, ja sitä pienemmällä hän ei suostuisi myymään, koska arvioi saavuttamansa hyödyn suuremmaksi. Se on siis päätöksentekijän henkilökohtainen ominaisuus.

Tarvittavat merkinnät ja terminologia on nyt käyty läpi, jotta voidaan määrittää päätöksentekijän paras vaihtoehto eri päätöksistä. Päätöksentekoteorian periaatteiden mukaan päätöksentekijä valitsee aina vaihtoehdon, jolla hyöty maksimoituu. Olkoon nyt yksinkertainen päätöksentekotilanne, jossa pitää tehdä yksi päätös a ja siihen liittyy yksi epävarmuustekijä x . Oletetaan, että päätöksentekijä valitsee vaihtoehdon a . Tällöin hänen saavuttamansa arvo on $v(x, a) + w$, jossa $v(x, a)$ on päätöksen a arvo ja w on päätöksentekijän aikaisempi varallisuus. Päätöksentekijän saavuttama hyöty on nyt $u(v(x, a) + w)$. Maksimaalinen odotettu hyöty (*maximum expected utility*) on

$$\begin{aligned} MEU &= \max_{a \in A} \left\{ \sum_{x \in \Omega} u(v(x, a) + w) p(x) \right\} \\ &= \max_{a \in A} \{ E(u(v(x, a) + w)) \} \\ &= E(u(v(x, a^*) + w)), \end{aligned} \quad (2)$$

jossa a^* on optimaalinen vaihtoehto, eli se vaihtoehto, joka maksimoi odotetun hyödyn:

$$a^* = \arg \max_{a \in A} \{ E(u(v(x, a) + w)) \}.$$

Johdetaan seuraavaksi lauseke päätöksentekijän varmuusekvivalentille päätöstilanteessa. Merkitään CE :llä sitä hintaa, jolla päätöksentekijä on valmis myymään päätöksentekotilanteensa. Koska päätöksentekijä saa saman hinnan silloin, jos myisi tilanteensa hintaan CE , kuin jos pysyisi tilanteessaan ja saisi siitä saatavan hinnan, voidaan merkitä näistä tilanteista saatavat hyödyt yhtä suuriksi:

$$\begin{aligned} u(w + CE) &= MEU = E(u(v(x, a^*) + w)) \\ \Rightarrow CE &= u^{-1}(E(u(v(x, a^*) + w))) - w. \end{aligned}$$

Varmuusekvivalentin lauseke voidaan vielä yksinkertaistaa niin, että saadaan lauseke riippumattomaksi päätöksentekijän alkuperäisestä varallisuudesta w . Sanotaan, että hyötyfunktio $u(v)$ täyttää niin kutsutun delta-ominaisuuden, jos seuraava ehto on totta:

$$u^{-1}(E(u(v(x, a) + \Delta))) = u^{-1}(E(u(v(x, a)))) + \Delta.$$

Voidaan osoittaa, että vain lineaarinen ja eksponentiaalinen hyötyfunktio voivat täyttää delta-ominaisuuden ehdon (Eidsvik et al., 2015, s. 72). Jos päätöksentekijän hyötyfunktio täyttää nyt delta-ominaisuuden, varmuusekvivalentin lauseke yksinkertaistuu muotoon

$$CE = u^{-1}(E(u(v(x, a^*))).$$

Myöhemmin informaatioarvoa laskiessa samaa delta-ominaisuutta voidaan käyttää, ja se osoittautuu hyvin hyödylliseksi. Oletetaan vielä päätöksentekijän olevan riskineutraali, jolloin hyötyfunktio on lineaarinen eli muotoa $u(v) = \alpha + \beta v$, $\beta > 0$. Lauseke yksinkertaistuu tällöin edelleen muotoon, jossa varmuusekvivalentti on yhtä suuri kuin odotettu arvo:

$$CE = E(v(x, a^*)).$$

(Eidsvik et al., 2015, s. 68-75.)

3.2 Informaatioarvo

Päätösanalyysin tavoite on luonnollisesti auttaa päätöksentekijää valitsemaan vaihtoehtoja, joilla päästään parempaan lopputulokseen kuin ilman analyysiä olisi päästy. Päätöksentekoon liittyy kuitenkin usein edellä mainittuja epävarmuustekijöitä. Vaikka edellisen luvun päättelyllä oltaisiinkin päästy lopputulokseen parhaasta mahdollisesta vaihtoehdosta a , epävarmuustekijöiden takia lopputulos ei silti ole täysin varmasti paras mahdollinen. Epävarmuuden mallintaminen on informaatioarvon keskeinen ajatus.

Tarkastellaan seuraavaksi tilannetta, jossa päätöksentekijän täytyy ennen varsinaista päätöksentekoa päättää, hankkiiko hän lisätietoja päätöksenteon tueksi vai ei. Lisätiedot liittyvät epävarmuustekijään. On siis kyse alkuperäiseen päätöksentekotilanteeseen liittyvästä lisäpäätöksestä, joka täytyy tehdä ennen varsinaista päätöksentekoa. Tarkoituksena on löytää maksimaalinen hinta lisätiedolle, joka sen hankkimisesta kannattaa maksaa.

3.2.1 Informaatioarvo täydelliselle informaatiolle

Informaatioarvo (*value of information*, *VOI*) tarkoittaa sitä hintaa, jolla päätöksentekijä on kahden vaiheilla, hankkiiko päätöksenteossa auttavan lisätiedon vai ei. Toisin sanoen informaatioarvo on maksimihinta, jolla päätöksentekijän vielä kannattaa ostaa haluamansa lisäinformaatio päätöksen tekemiseksi. Jos haluttu informaatio maksaa enemmän kuin maksimihinta, päätöksentekijä häviäisi rahaa hankkiessaan tiedon. Jos haluttu informaatio maksaa vähemmän kuin maksimihinta, tieto kannattaa hankkia.

Tällainen informaation maksimaalinen hinta voidaan laskea käyttäen edellisen luvun merkintöjä. Tarkastellaan jälleen yksinkertaista päätöksentekotilannetta, jossa

on yksi epävarmuustekijä x ja yksi päätös a . Samalla tavalla kuin varmuusekvivalentin laskennassa edellisessä luvussa, nytkin pitää asettaa odotetut hyödyt kahdesta eri tilanteesta yhtä suuriksi; hyöty siinä tilanteessa, että lisäinformaatio päätetään hankkia sekä siinä tilanteessa, että sitä ei päätetä hankkia. Olkoon nyt x havaittu ja vaihtoehto a valittu. Päätöksentekijä päättää maksaa hinnan P lisäinformaatiosta. Päätöksentekijällä on alkuperäinen varallisuus w ja hän saa päätöstilanteesta arvon $v(x, a)$. Oletetaan päätöksentekijän olevan riskineutraali, jolloin hyötyfunktio $u(v(x, a))$ on lineaarinen eli muotoa $u(v) = \alpha + \beta v$, $\beta > 0$. Optimaalinen vaihtoehto, jos x on havaittu, on

$$a^*(x) = \arg \max_{a \in A} \{u(v(x, a) + w - P)\}.$$

Epävarmuustekijä x on havaittu todennäköisyydellä $p(x)$. Odotettu hyöty tilanteesta on nyt

$$MEU' = \sum_x \max_{a \in A} \{u(v(x, a) + w - P)\} p(x). \quad (3)$$

Informaatioarvo on sellainen hinta P , jolla odotetut hyödyt MEU' ja MEU (2) ovat samat:

$$\begin{aligned} MEU' &= MEU \\ \Leftrightarrow \sum_x \max_{a \in A} \{u(v(x, a) + w - P)\} p(x) &= \max_{a \in A} \{E(u(v(x, a) + w))\}. \end{aligned} \quad (4)$$

Yhtälö (4) saadaan ratkaistua iteratiivisesti muuttamalla hintaa P kunnes se toteuttaa yhtälön. Informaatioarvo saadaan kuitenkin laskettua käytännössä myös helpommin käyttämällä edellisestä luvusta tuttua delta-ominaisuutta. Lasketaan ensin päätöksentekijän varmuusekvivalentti CE tilanteelle, jossa informaatio päätetään hankkia, yhtälöstä (3):

$$\begin{aligned} u(CE' + w) &= MEU' \\ u(CE' + w) &= \sum_x \max_{a \in A} \{u(v(x, a) + w - P)\} p(x) \\ \Rightarrow CE' &= u^{-1} \left(\sum_x \max_{a \in A} \{u(v(x, a) + w - P)\} p(x) \right) - w. \end{aligned}$$

Delta-ominaisuutta hyödyntämällä alkuperäinen varallisuus w voidaan hävittää ja lauseke yksinkertaistuu muotoon

$$CE' = u^{-1} \left(\sum_x \max_{a \in A} \{u(v(x, a))\} p(x) \right) - P.$$

Kun yhtälöä (4) muokataan ottamalla puolittain hyötyfunktion käänteisfunktio ja käyttämällä delta-ominaisuutta, saadaan informaatioarvolle helpompi lauseke:

$$VOI(x) = u^{-1} \left(\sum_x \max_{a \in A} \{u(v(x, a))\} p(x) \right) - u^{-1} \left(\max_{a \in A} \{E(u(v(x, a)))\} \right).$$

Kun päätöksentekijän oletetaan olevan lisäksi riskineutraali, hyötyfunktion käänteisfunktioista päästään eroon. Informaatioarvo voidaan tällöin kirjoittaa

$$VOI(x) = PoV(x) - PV, \quad (5)$$

jossa

$$\begin{aligned} PV &= \max_{a \in A} \{E(u(v(x, a)))\} \\ &= \max_{a \in A} \left\{ \sum_x u(v(x, a)) p(x) \right\} \end{aligned} \quad (6)$$

ja

$$PoV(x) = \sum_x \max_{a \in A} \{u(v(x, a))\} p(x). \quad (7)$$

PV tarkoittaa varmuusekvivalenttia siinä tilanteessa, että lisäinformaatiota ei hankita (*prior value*) ja $PoV(x)$ tarkoittaa varmuusekvivalenttia siinä tilanteessa, että lisäinformaatio on hankittu (*posterior value*). Toisin sanoen PV on suurin odotettu tuotto, kun otetaan huomioon odotetut tuotot kaikista saatavilla olevista vaihtoehdoista. $PoV(x)$ taas on saatavilla olevien vaihtoehtojen keskimääräinen odotettavissa oleva tuotto. Informaatioarvo on näiden tuottojen erotus. Merkintää $VOI(x)$ käytetään selventämään sitä, että informaatioarvo lasketaan epävarmuustekijälle x . Jos informaatioarvo ylittää informaation hinnan, päätöksentekijän tulee hankkia informaatio. (Eidsvik et al., 2015, s. 93-95.)

3.2.2 Informaatioarvo epätäydelliselle informaatiolle

Kaikki tähänastiset päätelmät on tehty olettaen, että lisäinformaatio, jota ollaan hankkimassa, antaisi täysin varman tiedon epävarmuustekijän tilasta. Tällöin sanotaan, että informaatio on täydellistä. Jos tieto epävarmuustekijästä pitää paikkansa vain tietyllä todennäköisyydellä, sanotaan, että informaatio on epätäydellistä. Myös tällaista tietoa voidaan käyttää hyödyksi päätöksenteossa.

Lasketaan seuraavaksi arvo epätäydelliselle informaatiolle samanlaisessa, yksinkertaisessa päätöksentekotilanteessa kuten edellä. Olkoon nyt y epävarmuustekijä, joka

liittyy aiempaan epävarmuustekijään x . Havaitaan y todennäköisyydellä $p(y)$, joka saadaan kokonaistodennäköisyytenä $p(y) = \sum_x p(x)p(y | x)$. Odotettu hyöty tilanteesta on

$$\begin{aligned} MEU'' &= \sum_y \max_{a \in A} \{E(u(v(x, a) + w - P) | y)\} p(y) \\ &= \sum_y \max_{a \in A} \left\{ \sum_x u(v(x, a) + w - P)p(x | y) \right\} p(y), \end{aligned} \quad (8)$$

jossa posteriorijakauma $p(x | y)$ lasketaan Bayesin kaavalla. Informaatioarvo on taas hinta P , jolla odotetut hyödyt MEU'' (8) ja MEU (2) ovat samat:

$$\begin{aligned} MEU'' &= MEU \\ \Leftrightarrow \sum_y \max_{a \in A} \{E(u(v(x, a) + w - P) | y)\} p(y) &= \max_{a \in A} \{E(u(v(x, a) + w))\}. \end{aligned}$$

Samalla tavalla kuin täydellisen informaation tapauksessa, hinta P saadaan laskettua iteratiivisesti lausekkeesta. Loputkin päättelyt voidaan tehdä samaan tapaan, ja informaatioarvolle saadaan yksinkertaisempi lauseke:

$$VOI(y) = u^{-1} \left(\sum_y \max_{a \in A} \{E(u(v(x, a)) | y)\} p(y) \right) - u^{-1} \left(\max_{a \in A} \{E(u(v(x, a)))\} \right).$$

Kun päätöksentekijän oletetaan olevan riskineutraali, lauseke tyypistyy muotoon

$$VOI(y) = PoV(y) - PV \quad (9)$$

jossa

$$\begin{aligned} PoV(y) &= \sum_y \max_{a \in A} \{E(u(v(x, a)) | y)\} p(y) \\ &= \sum_y \max_{a \in A} \left\{ \sum_x u(v(x, a))p(x | y) \right\} p(y) \end{aligned} \quad (10)$$

ja PV (6) on suurin odotettu tuotto kuten täydellisen informaation tapauksessa. Merkintää $VOI(y)$ käytetään selventämään, että informaatioarvo lasketaan epävarmuustekijälle y . (Eidsvik et al., 2015, s. 95-97.)

3.3 Informaatioarvo vesistöjen seurannoissa

Tässä luvussa tarkastellaan, miten informaatioarvon käsitettä voisi soveltaa vesistöjen seurantaan, tässä tapauksessa Suomen järvivesien seurantaan. Vaikka tutkimuksille, jotka arvioivat luonnosta kerätyn informaation arvoa, vaikuttaisi olevan tarve,

tutkimus näyttäisi olevan vielä alkuvaiheessa. Joitain aikaisempia tutkimuksia kuitenkin on jo olemassa (Nygård et al., 2016; Bouma et al., 2009).

Nygård et al. (2016) tarkastelevat Suomen merivesiseurannan kustannuksia ja käyttävät niiden arviointiin informaatioarvon käsitettä. Heidän tavoitteenaan on arvioida seurannan kokonaiskustannuksia käyttämällä hyväksi merivesiseurannasta koituvia taloudellisia kuluja sekä ottamalla huomioon myös seurannasta saatavat taloudelliset hyödyt. Heidän aineistonsa on saatu Suomen kansallisesta merellisen monimuotoisuuden seurantaohjelmasta. Aineisto seurannan kustannuksista on saatu seurannoista vastaavista suomalaisista laitoksista sekä niissä työskenteleviä asiantuntijoita haastatteleamalla, ja ne on sitten muutettu yhden näytteen kustannuksiksi.

Nygård et al. (2016) luovat käsitteellisen mallin, jota seuraamalla merivesien seurannoista saatavan informaation arvoa pystytään arvioimaan. Toimintamalli on seuraava:

1. Ensimmäisenä määritetään paras mahdollinen arvio meriveden ekologisesta tilasta, eli diskreetistä epävarmuustekijästä $\mathbf{x} = \{x_1, \dots, x_k\}$, jossa k on luokkien määrä, ennen kuin lisäinformaation keräystä on alettu toteuttamaan.
2. Määritetään erilaiset vaihtoehtoiset tavat, joilla informaatiota tilanteesta voidaan alkaa kerätä. Yksi vaihtoehdoista on, että informaatiota ei kerätä lainkaan.
3. Arvioidaan kustannukset näille erilaisille vaihtoehtoisille tavoille kerätä tietoa.
4. Kun informaation keräystapa on valittu ja informaatio meren tilasta kerätty, arvioidaan meren ekologinen tila uudelleen (kohdan 1 tapaan), mutta käytetään kerättyä tietoa nyt apuna. Toisin sanoen syvennetään ymmärrystä tilasta.
5. Määritetään erilaiset vaihtoehtoiset toimintatavat, eli päätökset $\mathbf{a} = \{a_1, \dots, a_p\}$, p on vaihtoehtojen lukumäärä, joilla meren tilaa voidaan alkaa parantaa perustuen kerättyyn tietoon. Yksi vaihtoehdoista on, ettei mitään toimia tehdä.
6. Arvioidaan kustannukset näille erilaisille vaihtoehtoisille toimintatavoille.
7. Arvioidaan ympäristön tilan muutos, mikäli joku toimintavaihtoehdoista on toteutettu.
8. Arvioidaan lopuksi eri meriveden ekologisiin tiloihin liittyvät arvot $v(x, a)$. Esimerkiksi taloudellinen hyöty, joka saadaan saavuttamalla hyvä ekologinen tila meressä.

Informaatioarvon laskemiseksi tarvitaan vielä erilaisten vaihtoehtoisten tilojen prioritodennäköisyydet kohdissa 1, 4 ja 7. Esimerkiksi kohdassa 1 voitaisiin sanoa, että ilman tarkempaa taustatietoa arvioidaan hyvän ekologisen tilan todennäköisyydeksi meressä $p(\text{hyvä}) = 0,3$.

3.4 Informaatioarvon soveltaminen Hiidenveden seuranta-aineistoon

Tässä työssä sovelletaan edellisen luvun käsitteellistä mallia informaatioarvon laskemiseksi Suomen järvivesien seurannassa, ja analyysi päätettiin toteuttaa yhdelle järvelle. Tarkoituksena on siis ennen varsinaista kunnostuspäätöksen tekemistä selvittää, paljonko kannattaa maksaa lisäinformaatiosta järven tilan selvittämiseksi. Esimerkijärveksi valikoitui Hiidenvesi. Ennakkokäsitys Hiidenveden ekologisesta tilasta on, että se on hyvää huonompi. Kaikkia toimintamallin kohtia ei toteuteta sellaisenaan, vaan mallin oletetaan olevan vain suuntaa antava. Esimerkiksi vaihtoehtoisia tapoja kerätä informaatiota järven tilasta tai informaation hankinnan kustannuksia ei tässä työssä käsitellä.

Nygård et al. (2016) mukaista käsitteellistä mallia seuraten ensimmäiseksi määritellään epävarmuutta aiheuttava tekijä x . Se on tässä vesistön todellinen ekologinen tila. Klorofyllimittaukseen perustuvaa ekologista tilaa taas merkitään muuttujalla y . Tilaluokkia on vesipuitedirektiivin mukaisesti viisi; erinomainen, hyvä, tyydyttävä, välttävä ja huono. Ne järvet, joiden tila on erinomainen, tiedetään pelkästään määrittelyn perusteella. Näistä järvistä käytetään nimitystä vertailu- tai referenssijärvet (Stoddard et al., 2006).

Muiden tilojen määrittämisen tarkkuuteen saatiin asiantuntija-apua. Asiantuntija-arvion mukaan vesistön huono ja välttävä tila tiedetään lähes aina varmasti. Jos taas järven tila on jotain siltä väliltä (hyvä tai tyydyttävä), oikean ekologisen tilan määrittäminen vaikeutuu huomattavasti. Lisäksi hyvän ja tyydyttävän rajan erottamisen tärkeyttä lisää se, että lain mukaan toimenpiteisiin on ryhdyttävä, kun vesistön tila on tyydyttävä tai sitä huonompi. Järven ekologisen tilan määrittäminen rajoittuu siis siihen, onko tila sellainen, että kunnostuksiin olisi ryhdyttävä vai ei. Tämän perusteella tässä työssä ekologisten tilojen asteikko yksinkertaistetaan dikotomiseksi, hyvä tai huono tila. Hyvällä tilalla tarkoitetaan sellaisen järven tilaa, joka ei vaadi kunnostusta ja se vastaa vesipuitedirektiivin mukaisia tiloja erinomainen tai hyvä. Huonolla tarkoitetaan sellaisen järven tilaa, joka tarvitsee kunnostuksen ja se vastaa vesipuitedirektiivin mukaisia tiloja tyydyttävä, välttävä ja huono.

Erilaisia vaihtoehtoisia tapoja kerätä informaatiota järven tilasta tai näiden kustannuksia ei tässä työssä käsitellä. Toinen ja kolmas kohta käsitteellisestä mallista jätetään siis tarkastelematta.

Neljännessä kohdassa arvioidaan ekologisten tilojen ehdolliset todennäköisyydet klorofyllimittauksen tekemisen jälkeen. Taulukossa 1 on koottu asiantuntija-arvioon perustuen eräs ehdotus ehdollisten todennäköisyyksien matriisista mitata järven ekologiseksi tilaksi hyvä tai huono ehdolla järven oikea ekologinen tila. Ehdollisia todennäköisyyksiä voidaan pitää myös kerättävän informaation luotettavuutena tai mittauksen tarkkuutena; mitä lähempänä ykköstä diagonaalilla olevat todennäköisyydet ovat, sitä parempaa klorofyllimittauksista saatava informaatio on. Jos diagonaalilla olevat todennäköisyydet olisivat ykkösiä, informaatio olisi täydellistä. Luvussa 5.1 on esitetty menetelmä informaation luotettavuuden arvioimiseksi lineaarisen sekamallin avulla. Lisäksi luvussa 5.3 on tehty sensitiivisyystarkasteluja, miten prioritodennäköisyydet $p(x = \text{järven oikea tila})$ sekä ehdolliset todennäköisyydet $p(y = \text{järven mitattu tila} \mid x = \text{järven oikea tila})$ vaikuttavat informaatioarvoon.

Viidennessä kohdassa määritetään päätökset a . Niitä ovat erilaiset vaihtoehtoiset vesiensuojelutoimenpiteet tai vesistön kunnostukset. Kunnostuksilla pyritään hillitsemään ulkoista kuormitusta, eli asutuksen, maa- ja metsätalouden päästöjä, sekä järven sisäistä kuormitusta eli pohjasta tulevaa ravinnekuormitusta. Keskeisiä kunnostustoimenpiteitä Hiidenvedelle ovat olleet altaiden, kosteikkojen ja suojavyyhykkeiden rakentaminen ja hoitokalastus (Ahtiainen, 2008). Yksi päätösvaihtoehtoista on, ettei järvelle tehdä minkäänlaisia toimenpiteitä. Tässä työssä rajataan vaihtoehdot siten, että päätös a saa arvot $\mathbf{a} = \{\text{ei tehdä mitään, tehdään kunnostus}\}$. Hiidenveden kunnostukseen on budjetoitu varoja 1,4 miljoonaa euroa (Ahtiainen, 2008). Tästä saadaan kohdassa kuusi tarvittavat toimenpiteiden kustannukset.

Taulukko 1: Esimerkki ehdollisista todennäköisyyksistä $p(y \mid x)$ mitata järven ekologinen tila (sarakkeet) ehdolla oikea ekologinen tila (rivit) asiantuntija-arvioon perustuen. Diagonaalilla olevat todennäköisyydet kertovat, kuinka todennäköisesti järvi luokitellaan oikein. Ei-diagonaalilla olevat todennäköisyydet kertovat luokitteluvirheistä.

	$p(y \mid x)$	
	Mitattu tila y	
Oikea tila x	Huono	Hyvä
Huono	0,5	0,5
Hyvä	0,1	0,9

Lopuksi arvioidaan ekologisiin tiloihin liittyvät arvot. Arvojen arviointi osoittautui haastavaksi, koska ympäristön arvottaminen on monimutkaista. Vesistöjen ekologisen tilan parantumisen taloudellisten hyötyjen arviointiin on vain vähän tietoa ja menetelmiä, ja tutkimus on ollut vähäistä (Ahtiainen, 2008). Ahtiainen (2008) on tutkinut järviveden rehevöitymisen vähentämisen taloudellisia hyötyjä, esimerkiksi Hiidenveden arvon kasvaminen. Hän arvioi Hiidenveden tilan parantamisen rahallisia hyötyjä kysymällä ihmisten maksuhalukkuutta järven rehevyyden vähentämiseksi. Tutkimuksessa maksuhalukkuutta selvitettiin lähettämällä Hiidenveden lähikuntien ja -kaupunkien asukkaille sekä vapaa-ajanasuntojen omistajille kysely. Maksuhalukkuus viidelle vuodelle arvioitiin olevan 3 miljoonan ja 5,7 miljoonan välillä. Tämän perusteella on koottu taulukon 2 jokaisen skenaarion arvot Hiidenveden tapauksessa. Arvoista on vähennetty vielä riveittäin toimenpiteen hinta, jolloin on saatu hyödyt jokaisen skenaarion tapauksessa. Jos Hiidenveden tila on huono, kunnostus tehdään ja jos tila myös parantuu hyväksi, järven arvo nousee 5,7 miljoonaa euroa. Kunnostuksesta saadaan kuitenkin vain 4,3 miljoonan euron hyöty kustannusten vähentämisen jälkeen.

Toinen tapa arvottaa järven ekologinen tila voisi olla järven kalakantojen määrän muuttaminen rahalliseksi arvoksi. Lisäksi järviin liittyvän virkistystoiminnan arvoja voisi tutkia. (Tolonen et al., 2014.)

Taulukko 2: Hiidenveden informaatioarvon laskemiseksi tarvittavat muuttujat ja skenaarioiden hyödyt. Hyöty $u(v(x, a))$ on saatu vähentämällä jokaisen skenaarion arvosta $v(x, a)$ päätöksen hinta riveittäin.

Päätös a	Päätöksen hinta (milj. €)	Hyöty $u(v(x, a))$ (milj. €)	
		Järven tila x	
		Huono	Hyvä
Ei tehdä mitään	0	0	5,7
Tehdään kunnostus	1,4	4,3	4,3

4 Sekamallien teoriaa

Sekamallit ovat yleistettyjen lineaaristen mallien laajennuksia. Tavallisesti tilastotieteen oppikirjoissa oletus on, että havainnot ovat riippumattomia ja identtisesti jakautuneita, toisin sanoen havainnot ovat peräisin samasta todennäköisyysjakoumasta. Joissain tapauksissa tämä oletus ei kuitenkaan päde, vaan aineiston havainnot ovatkin riippuvia ja siten aineistossa onkin monimutkaisempi, monitasoinen rakenne. Eri ryhmien havainnot oletetaan keskenään riippumattomiksi, mutta saman ryhmän havainnot ovat riippuvaisia. Sekamalleilla voidaan mallintaa myös pitkätaaisaineistoja, jolloin kunkin yksilön oma aikasarja muodostaa yksittäisen klusterin. Sekamalleista onkin tullut yksi tilastotieteen tärkeimmistä menetelmistä mallintaa korreloituneita aineistoja. (Demidenko, 2013, s. 1-2.)

Tässä työssä käytetään lineaarista sekamallia ehdollisten todennäköisyyksien $p(y = \text{järven mitattu tila} \mid x = \text{järven oikea tila})$ arvioimiseksi. Seuraavaksi keskitytään lineaarisen sekamallin teoriaan. Luvussa 4.1 muotoillaan malliyhtälö kahden tason tapauksessa, ja luvussa 4.2 pohditaan, mikä muuttuu, jos tasoja onkin monia. Luvussa 4.3 keskitytään mallien vertailuun.

4.1 Lineaarinen sekamalli

Sekamallien tapauksessa vasteeseen liittyvät vaikutukset jaetaan kiinteisiin ja satunnaisiin. Kiinteillä vaikutuksilla tarkoitetaan selittäjiä, joilla on sama vaikutus koko populaatioon. Satunnaiset vaikutukset taas ilmaisevat ryhmän eroa keskimääräisestä ennusteesta. Yleinen lineaarinen sekamalli on malli, jossa yhdistetään sekä kiinteitä että satunnaisia vaikutuksia. Tavallista lineaarista mallia siis laajennetaan lisäämällä malliyhtälöön satunnaisia muuttujia. Lisäämällä satunnaiset vaikutukset malliin saadaan luotua riippuvuusrakenne kunkin yksikön havaintojen välille, olkoon se sitten yksi yksilö tai ryhmä. (Pinheiro & Bates, 2002, s. 3.)

Oletetaan vasteen \mathbf{y}_i olevan jatkuva muuttuja, i viittaa ryhmään. On huomioitava, että mallin vaste ei ole sama satunnaismuuttuja kuin luvussa 3.2.2 esiintyvä epävarmuustekijä y . Yleinen lineaarinen sekamalli voidaan tällöin kirjoittaa matriisimuodossa

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i, \quad (11)$$

jossa $\boldsymbol{\beta}$ on p -ulotteinen kiinteiden vaikutusten regressiokertoimet sisältävä vektori, \mathbf{u}_i on q -ulotteinen satunnaisvaikutusten vektori, \mathbf{X}_i on $(n \times p)$ -kokoinen kovariaat-

timatriisi eli design-matriisi kiinteille vaikutuksille, \mathbf{Z}_i on $(n \times q)$ -kokoinen design-matriisi satunnaistekijöille, sekä $\boldsymbol{\epsilon}_i$ on n_i -ulotteinen satunnaisvirheiden vektori.

Mallin keskeisimmät oletukset ovat satunnaisvirheiden riippumattomuus ja normaalijakautuneisuus, eli $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, sekä satunnaisvaikutusten normaalijakautuneisuus, $\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$. (Pinheiro & Bates, 2002, s. 58.)

Nyt vasteen odotusarvo on

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

ja marginaalinen odotusarvo vastaavasti

$$\mathbf{E}(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\beta}.$$

(McCulloch & Searle, 2001, s. 157.)

4.2 Monitasomalli

Edellisen luvun formulointi koskee aineistoja, joissa on vain kaksitasoinen luokittelu, toisin sanoen vain kaksi muuttujaa, jonka mukaan aineisto saa luokitellun rakenteen. Tällainen aineiston rakenne voisi olla esimerkiksi joukko yksilöitä (i), joilta on otettu mittauksia eri ajanhetkinä (j). Kirjallisuudessa käytetään samasta rakenteesta myös nimitystä yksitasoinen luokittelu. Tällöin tasot lasketaan satunnaisten vaikutusten sisäkkäisten tasojen määränä (Pinheiro & Bates, 2002, s. 61).

Sekamallien aineistossa voi kuitenkin olla myös monitasoisempi luokittelu. Luokittelun rakenne voidaan jakaa ainakin kahteen erilaiseen tyyppiin; sisäkkäiseen (*nested*) ja ristikkäiseen (*crossed*) luokitteluun. Sisäkkäinen luokittelu tarkoittaa, että luokkien sisällä on alaluokkia, joita ei ole muiden luokkien alla. Esimerkiksi järvien ja niistä valittujen havaintopaikkojen välillä on sisäkkäinen rakenne, koska mikään havaintopaikka ei voi kuulua kahteen eri järveen. Ristikkäisessä luokittelussa taas luokkia ei voida järjestää samalla tavalla hierarkkisesti, vaan havainto, joka kuuluu yhdessä luokittelussa yhteen ryhmään, voi kuulua toisessa luokittelussa mihin tahansa ryhmään. Esimerkkinä ristikkäisestä luokittelusta voisi olla järvet, joista on otettu mittauksia samoina ajanhetkinä. Luokitteluiden eroa havainnollistaa taulukko 3. Aineistossa nämä kaksi tyyppiä voivat sekoittua monin eri tavoin. (de Leeuw et al., 2008, s. 301-304.)

Seuraavaksi laajennetaan edellisen luvun malliyhtälö (11) koskemaan myös monitasoisempaa aineistoa. Erityisesti kyseessä ovat nyt mallit, joissa satunnaisvaikutuksissa on sisäkkäinen, monitasoinen rakenne.

Taulukko 3: Aineiston luokittelun rakenne voi olla joko sisäkkäistä (vasen taulukko) tai ristikkäistä (oikea taulukko).

		Tason 1 ryhmät					Tason 1 ryhmät		
		1	2	3			1	2	3
Tason 2 ryhmät	1	X			Tason 2 ryhmät	1	X	X	X
	2	X				2	X	X	X
	3		X			3	X	X	X
	4		X			4	X	X	X
	5		X			5	X	X	X
	6			X		6	X	X	X

Esitetään ensimmäiseksi malli kolmitasoiselle aineistolle. Satunnaisvaikutuksissa on tällöin kaksi sisäkkäistä tasoa. Olkoon nyt jatkuva vaste sisimmällä tasolla \mathbf{y}_{ij} , $i = 1, \dots, M$, $j = 1, \dots, M_i$, jossa M on ensimmäisen tason ryhmien lukumäärä ja M_i on toisen tason ryhmien lukumäärä ensimmäisen tason ryhmässä i . Vektorin \mathbf{y}_{ij} pituus on n_{ij} . Malliyhtälö on silloin muotoa

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{i,j}\mathbf{u}_i + \mathbf{Z}_{ij}\mathbf{u}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad (12)$$

$$i = 1, \dots, M, \quad j = 1, \dots, M_i,$$

jossa $\boldsymbol{\beta}$ on p -ulotteinen kiinteiden vaikutusten vektori kuten kaksitasoisessa mallissa (11), \mathbf{u}_i on q_1 -ulotteinen satunnaisvaikutusten vektori ensimmäisellä luokittelutasolla ja \mathbf{u}_{ij} q_2 -ulotteinen satunnaisvaikutusten vektori toisella tasolla, \mathbf{X}_{ij} on $(n_{ij} \times p)$ -kokoinen design-matriisi kiinteille vaikutuksille sekä $\mathbf{Z}_{i,j}$ ja \mathbf{Z}_{ij} ovat $(n_i \times q_1)$ - ja $(n_i \times q_2)$ -kokoiset design-matriisit niitä vastaaville satunnaisvaikutuksille. Oletetaan, että

$$\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}_1), \mathbf{u}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Psi}_2), \boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Ensimmäisen tason satunnaisvaikutukset \mathbf{u}_i oletetaan riippumattomiksi eri ryhmässä i ja toisen tason satunnaisvaikutukset \mathbf{u}_{ij} oletetaan riippumattomiksi eri ryhmässä i ja j . Lisäksi toisen tason satunnaisvaikutukset \mathbf{u}_{ij} oletetaan riippumattomiksi ensimmäisen tason satunnaisvaikutusten \mathbf{u}_i kanssa. Ryhmien väliset satunnaisvirheet $\boldsymbol{\epsilon}_{ij}$ oletetaan riippumattomiksi eri ryhmässä i ja j sekä riippumattomiksi satunnaisvaikutusten kanssa.

Kuten kolmitasoiselle aineistolle, malli voidaan vastaavasti laajentaa nelitasoiselle aineistolle ja niin edelleen. Esimerkiksi malliyhtälö nelitasoiselle aineistolle on muo-

toa

$$\mathbf{y}_{ijk} = \mathbf{X}_{ijk}\boldsymbol{\beta} + \mathbf{Z}_{i,jk}\mathbf{u}_i + \mathbf{Z}_{ij,k}\mathbf{u}_{ij} + \mathbf{Z}_{ijk}\mathbf{u}_{ijk} + \boldsymbol{\epsilon}_{ijk},$$

$$i = 1, \dots, M, \quad j = 1, \dots, M_i, \quad k = 1, \dots, M_{ij},$$

jossa

$$\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}_1), \mathbf{u}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Psi}_2), \mathbf{u}_{ijk} \sim N(\mathbf{0}, \boldsymbol{\Psi}_3), \boldsymbol{\epsilon}_{ijk} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Tällöin satunnaisvaikutuksissa on siis kolme sisäkkäistä tasoa. (Pinheiro & Bates, 2002, s. 60.)

Jos satunnaisvaikutusten monitasoinen rakenne ei ole sisäkkäistä, malli on hieman erilaista muotoa. Edellisen kaltainen formulointi ei yleisty suoraan ristikkäiselle rakenteelle, koska malli kirjoitetaan alemman tason yksilölle j . Jos rakenne on ristikkäinen, niin sanotusti samanarvoisia tasoja on kaksi. Tällöin ei ole selvää, kumpaa tasoa indeksillä j tarkoitettaisiin, joten malliyhtälö on kirjoitettava uudestaan. Tässä työssä käytettävässä Hiidenveden aineistossa rakenne on ristikkäinen, sillä eri havaintopaikoista on kerätty mittauksia samoina ajankohtina.

Olkoon nyt satunnaisvaikutuksissa kaksi tasoa, joiden rakenne on ristikkäinen. Olkoon ensimmäisellä luokittelutasolla N kappaletta ja toisen tason luokittelevissa muuttujissa M_1 ja M_2 kappaletta ryhmiä. Malliyhtälö voidaan tällöin kirjoittaa

$$\mathbf{y}_{i(j_1, j_2)} = \mathbf{X}_{i(j_1, j_2)}\boldsymbol{\beta} + \mathbf{Z}_{1i j_1}\mathbf{u}_{1j_1} + \mathbf{Z}_{2i j_2}\mathbf{u}_{2j_2} + \boldsymbol{\epsilon}_{i(j_1, j_2)}, \quad (13)$$

$$i = 1, \dots, N, \quad j_1 = 1, \dots, M_1, \quad j_2 = 1, \dots, M_2,$$

jossa \mathbf{u}_{1j_1} ja \mathbf{u}_{2j_2} ovat q_1 - ja q_2 -ulotteiset satunnaisvaikutusten vektorit ensimmäisessä ja toisessa luokittelussa sekä $\mathbf{Z}_{1i j_1}$ ja $\mathbf{Z}_{2i j_2}$ niitä vastaavat design-matriisit, ja muut kuten mallissa (11). Malli rakentuu samaan tapaan kuin jos satunnaisvaikutuksissa olisi vain yksi luokittelun taso. Ensimmäisen muuttujan satunnaisvaikutukset \mathbf{u}_{1j_1} muodostetaan kuten mallissa (11) välittämättä toisesta luokittelevasta muuttujasta. Vastaavasti toisetkin satunnaisvaikutukset \mathbf{u}_{2j_2} muodostetaan huomioimatta ensimmäistä luokittelua. (de Leeuw et al., 2008, s. 305-306.)

Edellä esitettyjä sekamalleja voidaan sovittaa R-ohjelmistolla (R Core Team, 2017), esimerkiksi paketeilla `n1me` (Pinheiro et al., 2017) ja `lme4` (Bates et al., 2015). Satunnaisvaikutusten ristikkäisten rakenteiden käsittely paketilla `n1me` on vaikeampaa verrattuna pakettiin `lme4`; mallin määrittely on monimutkaista ja sovitus aikaa vievää. Paketissa `lme4` ristikkäisten satunnaisvaikutusten sovitus on tehty yksinkertaisemmaksi. Tässä työssä mallin sovitus tehdään paketin `lme4` funktion `lmer` avulla aineiston ristikkäisten rakenteiden vuoksi.

4.3 Mallin valinta

Kun malli on saatu sovitettua, halutaan usein saada selville, miten hyvin malli sopii aineistoon verrattuna muihin mahdollisiin malleihin. Seuraavaksi käydään läpi mallien vertailua.

Sisäkkäisiä malleja, jotka on sovitettu suurimman uskottavuuden menetelmällä, voidaan vertailla uskottavuusosamäärän testillä. Testiä voi käyttää myös REML-menetelmällä sovitettuihin malleihin, jos kiinteät vaikutukset ovat samat kummassakin mallissa. Mallin sanotaan olevan sisäkkäinen toisen mallin kanssa, jos se on erityistapaus toisesta mallista. Mallia kutsutaan tässä rajoitetuksi, jos siinä on vähemmän parametreja, ja yleisemmäksi malliksi, jos siinä on enemmän parametreja verrattuna rajoitettuun malliin. Hypoteesit ovat

H_0 : rajoitettu malli on riittävä,

H_1 : yleisempi malli on rajoitettua selvästi parempi.

Uskottavuusosamäärän testin testisuure on muotoa

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

jossa L_2 on yleisemmän mallin ja L_1 rajoitetun mallin uskottavuusfunktio. Oletetaan, että rajoitetussa mallissa on r parametria vähemmän kuin yleisemmässä mallissa. Testisuure noudattaa χ_r^2 -jakaumaa nollahypoteesin vallitessa:

$$LRT \sim \chi_r^2.$$

(Pinheiro & Bates, 2002, s. 83-84.)

Jos halutaan vertailla sisäkkäisiä malleja, joilla kiinteät vaikutukset ovat erilaiset, suositellumpaa on käyttää t - tai F -testejä. Tällaisessa tilanteessa uskottavuusosamäärän testiä voidaan käyttää vain, jos se on sovitettu käyttäen suurimman uskottavuuden menetelmää, koska REML-menetelmää käytettäessä uskottavuusfunktio muuttuu kiinteiden vaikutusten muuttuessa. Uskottavuusosamäärän testin käyttämistä ei kuitenkaan suositella, vaikka testisuure voidaankin laskea. Testi saattaa olla niin sanotusti antikonservatiivinen, eli se hylkää nollahypoteesin liian helposti. (Pinheiro & Bates, 2002, s. 87-92.)

Jos mallit eivät ole sisäkkäisiä, niitä voidaan vertailla esimerkiksi informaatiokriteerien avulla. Akaiken informaatiokriteeri on muotoa

$$AIC = -2 \log(L) + 2n_{par},$$

jossa $\log(L)$ on joko suurimman uskottavuuden menetelmällä tai REML-menetelmällä saatu uskottavuusfunktion logaritmi, ja n_{par} on parametrien lukumäärä. (Pinheiro & Bates, 2002, s. 10 ja 84.)

5 Tulokset

Tässä luvussa käydään läpi tutkimuksen tuloksia. Tavoitteena on selvittää, mikä on havaintopaikoista kerättävän klorofyllimittausten informaation arvo. Analyysi päätettiin toteuttaa yhdelle Suomen järvelle, joksi valikoitui Hiidenvesi. Hiidenvesi valittiin, koska sille on aikaisemmin toteutettu sen tilan parantamisen hyödyistä arvio (Ahtiainen, 2008). Arviosta saatiin hyödyt, joita tarvitaan informaatioarvon laskemiseksi.

Ensimmäiseksi esitetään, miten informaatioarvon laskemisessa tarvittavia ehdollisia todennäköisyyksiä $p(y = \text{järven mitattu tila} \mid x = \text{järven oikea tila})$ on arvioitu lineaarisella sekamallilla verrattuna luvussa 3.4 esitettyihin asiantuntijan esittämiin. Tätä käsitellään luvussa 5.1. Seuraavaksi lasketaan informaatioarvoja erilaisilla prioritodennäköisyyksillä $p(x = \text{järven oikea tila})$ sekä ehdollisilla todennäköisyyksillä Hiidenveden tapauksessa. Tuloksia esitellään luvussa 5.2. Luvussa 5.3 toteutetaan sensitiivisyysanalyysi eli tarkastellaan, miten prioritieto ja tiedon luotettavuus, eli ehdolliset todennäköisyydet, vaikuttavat Hiidenveden informaatioarvoon.

5.1 Informaation luotettavuuden arvioiminen

Hiidenveden seuranta-aineistoon sovitettiin lineaarinen sekamalli. Vasteena on yksittäinen klorofyllin pitoisuusmittaus $chla_{ijk}$. Malli on muotoa

$$chla_{ijk} = \mu + u_{1i} + u_{2j} + \epsilon_{ijk}, \quad (14)$$

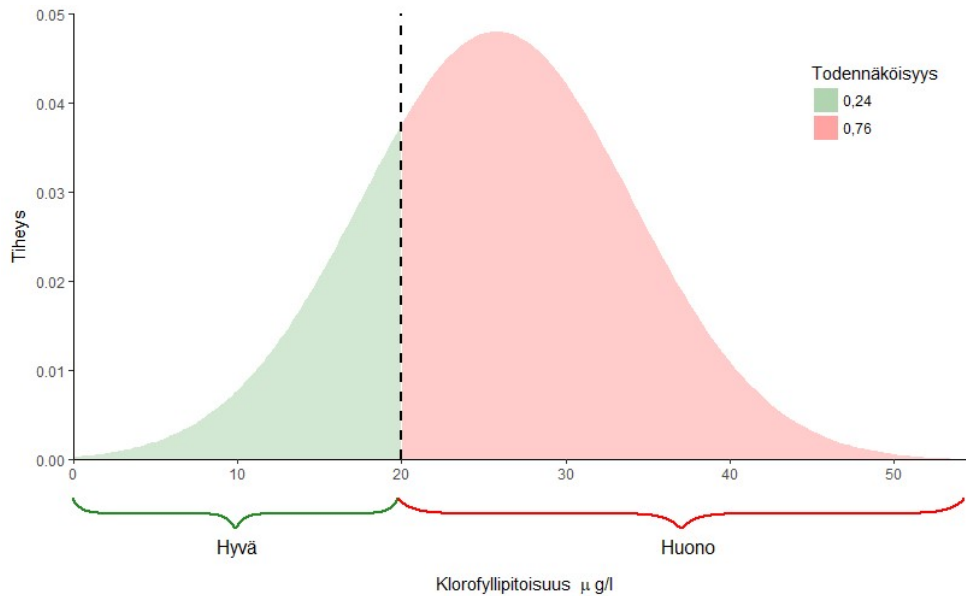
jossa μ on Hiidenveden klorofyllipitoisuuden odotusarvo koko tarkastelujaksolla, u_{1i} on vuosikohtainen satunnaistekijä, $i = 2006, \dots, 2012$, u_{2j} on havaintopaikkakohdainen satunnaistekijä, $j = 1, \dots, m$, kun m on havaintopaikkojen määrä, ja ϵ_{ijk} on jäännösvirhe, $k = 1, \dots, n$, kun n on kokonaishavaintomäärä. Satunnaistekijät siis huomioivat pitoisuuden vaihtelun vuosittain ja havaintopaikoittain. Kyseessä on sekamalli, jonka rakenne on ristikkäinen (kaava (13), katso myös kuva 1). Mallin estimaatit on esitetty taulukossa 4. Mallin diagnostiikkatarkasteluissa huomattiin, että jäännösten normaalisuus toteutunee mallissa likimäärin. Kun piirretään jäännökset sovitetun suhteen, jäännökset kasvavat hieman sovitetun kasvaessa. Seuranta-aineiston vaihtelun lähteitä on tarkastellut tarkemmin Carstensen & Lindgarth (2016).

Hiidenveden klorofyllipitoisuuden odotusarvon estimaattien jakaumasta saadaan arvio ehdollisille todennäköisyyksille. Estimaattien jakauma on sekamallin mukaan

Taulukko 4: Sekamallin parametriestimaatit.

Kiinteä osa	Satunnaisosa			
	$\hat{\mu}$ (s.e.)	$\hat{\sigma}_{u_{1i}}$	$\hat{\sigma}_{u_{2j}}$	$\hat{\sigma}_{\epsilon}$
25,8 (8,3)	8,9	12,6	22,0	

$N(25, 8; 8, 3^2)$ (taulukko 4). Samantapaista lähestymistapaa klorofyllipitoisuuden jakauman määrittämisessä on käyttänyt Kotamäki (2015) hieman erilaisella mallilla. Saatua jakaumaa verrataan Hiidenveden järvityyppiä vastaavien vertailujärvien luokkarajoihin. Hiidenvesi kuuluu järvityyppiin runsasravinteiset järvet, jolle hyvän ja tyydyttävän tilan välinen raja vesipuitedirektiivin mukaisella viisiluokkaisella asteikolla on $20 \mu\text{g/l}$ (kuva 2). Uudella kaksitasoisella asteikolla on siis kyse hyvän ja huonon tilan välisestä rajasta. Jos havaittu pitoisuus on alle rajan, järven tila on hyvä, ja jos pitoisuus on yli rajan, järven tila on huono. Kuvassa 3 on esitetty mallista estimoitu Hiidenveden klorofyllipitoisuuden odotusarvon estimaattien jakauma sekä vertailu vertailujärvien luokkarajaan. Viimeisimmällä luokittelukaudella 2006–2012 Hiidenveden tilaksi arvioitiin tyydyttävä, eli yksinkertaistetulla asteikolla huono (Ahtiainen, 2008). Tästä jakaumasta saadaan siis todennäköisyydet mitata Hiidenveden ekologinen tila ehdolla, että se on huonossa tilassa. Nämä to-



Kuva 3: Sekamallin avulla saatu klorofyllipitoisuuden odotusarvon estimaattien jakauma Hiidenvedelle, $N(25, 8; 8, 3^2)$. Jakaumaa verrataan Hiidenveden järvityyppiä vastaavien vertailujärvien luokkarajoihin (kuva 2), jolloin saadaan laskettua ehdolliset todennäköisyydet ennustaa ekologinen tila ehdolla oikea ekologinen tila, $p(y | x)$.

dennäköisyydet ovat $p(y = \text{huono} \mid x = \text{huono}) = 0,76$ ja $p(y = \text{hyvä} \mid x = \text{huono}) = 0,24$.

Vastaavia Hiidenveden klorofyllipitoisuuden otoskeskiarvoja simuloitiin myös parametrisella bootstrapilla (Efron & Tibshirani, 1993). Jakauma on lähes sama kuin suoraan mallista saatu. Tuloksia ei tässä näytetä.

Lisäksi tarvitaan vielä todennäköisyydet mitata ekologinen tila ehdolla, että oikea tila on hyvä. Ne saadaan vertailujärvien otoskeskiarvojen empiirisestä jakaumasta ja siitä saadusta tilaluokkien rajojen määritelmästä. Ekologisen tilan luokkarajat eri järvityypeille saadaan siten, että tietyn järvityypin vertailujärvien havaintojen vuosimediaanien tai vuosittaisten keskiarvojen jakaumasta otetaan kvantiilipisteitä tilaluokkien rajoiksi (Vuori et al., 2009). Nämä todennäköisyydet ovat $p(y = \text{huono} \mid x = \text{hyvä}) = 0,25$ ja $p(y = \text{hyvä} \mid x = \text{hyvä}) = 0,75$. Tällöin sekä vertailujärvien jakaumasta, että Hiidenveden pitoisuuskeskiarvojen jakaumasta saatujen todennäköisyyksien voidaan katsoa olevan vertailukelpoiset. Vertailurajana voidaan käyttää muitakin kvantiilipisteitä (Hämäläinen et al., 2018). Esimerkiksi taulukossa 1 on ajateltu hyvän ja huonon tilan rajan olevan 90 prosentin kvantiilipisteessä. Tässä taas käytetään yläkvartiilia eli ylintä neljänneistä vertailujärvien jakaumasta.

Sekamallista saadut ehdolliset todennäköisyydet on koottu taulukossa 5. Sekä sekamallin avulla arvioitua todennäköisyysmatriisia että asiantuntijan arvioon perustuvaa todennäköisyysmatriisia käytettiin lopulta informaatioarvon laskemiseen ja vertailtiin, miten erilaiset arviot vaikuttavat informaatioarvoon.

Taulukko 5: Sekamallin avulla arvioidut ehdolliset todennäköisyydet $p(y \mid x)$ mitata ekologinen tila (sarakkeet) ehdolla oikea ekologinen tila (rivit).

	$p(y \mid x)$	
	Mitattu tila y	
Oikea tila x	Huono	Hyvä
Huono	0,76	0,24
Hyvä	0,25	0,75

5.2 Informaatioarvot Hiidenvedelle

Taulukoissa 6 ja 7 on esitetty informaatioarvoja lisätiedon hankkimiselle Hiidenveden ekologisesta tilasta erilaisten priorien arvoilla. Taulukossa 6 on esitetty infor-

Taulukko 6: Hiidenveden klorofyllimittausten informaatioarvoja täydelliselle informaatiolle, kun käytössä on erilaista ennakkotietoa Hiidenveden tilasta. Odotettu tuotto saadaan, kun valitaan päätöksistä vaihtoehto, joka maksimoi tuoton: joko kunnostus päätetään olla tekemättä (e) tai kunnostus päätetään tehdä (k).

Prioritiedon laatu	Ekologisen tilan todennäköisyys $p(x)$		Valitun vaihtoehdon odotettu tuotto, PV (milj. €)	$PoV(x)$ (milj. €)	Informaatioarvo $VOI(x)$ (milj. €)
	Huono	Hyvä			
Huonoa tietoa	0,5	0,5	4,30 (k)	5,00	0,70
Viitteellistä tietoa	0,6	0,4	4,30 (k)	4,86	0,56
	0,4	0,6	4,30 (k)	5,14	0,84
Viitteellistä parempaa tietoa	0,8	0,2	4,30 (k)	4,58	0,28
	0,2	0,8	4,56 (e)	5,42	0,86

maatioarvoja täydelliselle informaatiolle, eli kun kerätyllä lisätiedolla saadaan täysin varmasti selville Hiidenveden ekologinen tila. Informaatioarvoksi saadaan taulukossa esitetystä ennakkotiedosta riippuen 0,28–0,86 miljoonaa euroa. Odotetut tuotot (PV) saadaan, kun valitaan toimintavaihtoehto, joka maksimoi tuoton. Suurimmas- sa osassa skenaarioita tuotto maksimoituu, kun valitaan, että Hiidenvedelle tehdään kunnostus (k). Vain yhdessä skenaariossa on kannattavampaa olla tekemättä kunnostusta (e).

Kun ollaan hyvin epävarmoja Hiidenveden todellisesta ekologisesta tilasta, eli $p(x = \text{hyvä}) = p(x = \text{huono}) = 0,5$, suurin odotettu tuotto saadaan, kun valitaan, että Hiidenvedelle tehdään kunnostus: odotettu tuotto on tällöin 4,3 miljoonaa euroa. Se saadaan sijoittamalla taulukon 2 hyödyt kaavaan (6):

$$PV = \max_{a \in A} \{0 \cdot 0,5 + 5,7 \cdot 0,5; 4,3 \cdot 0,5 + 4,3 \cdot 0,5\} = 4,3.$$

Informaatioarvoa varten lasketaan vielä molempien vaihtoehtojen keskimääräinen odotettavissa oleva tuotto kaavalla (7):

$$PoV(x) = \max_{a \in A} \{0; 4,3\} \cdot 0,5 + \max_{a \in A} \{5,7; 4,3\} \cdot 0,5 = 5,0.$$

Täydellisestä informaatiosta oikean tilan selvittämiseksi kannattaa maksaa tällöin 0,7 miljoonaa euroa. Se lasketaan kaavalla (5).

Kun käytössä on viitteellistä tietoa Hiidenveden ekologisesta tilasta, eli $p(x = \text{hyvä}) = 0,6$ tai $p(x = \text{huono}) = 0,6$, kannattavinta on tehdä kunnostus sekä siinä tilanteessa, että tilan oletetaan olevan huono, että tilanteessa, että sen oletetaan olevan hyvä. Odotettu tuotto on tällöinkin 4,3 miljoonaa euroa. Informaatioarvo täydelliselle informaatiolle on suurempi silloin, kun oletetaan tilan olevan hyvä, koska kannattaa varmistua järven hyvästä tilasta, ettei kunnostusta tehtäisi turhaan.

Kun ollaan suhteellisen varmoja Hiidenveden oikeasta tilasta, eli $p(x = \text{hyvä}) = 0,8$ tai $p(x = \text{huono}) = 0,8$, kunnostuspäätöksen tekeminen muuttuu helpommaksi. Kun ollaan aika varmoja, että Hiidenveden tila on huono, kunnostus kannattaa tehdä ja odotettu tuotto on tällöin 4,3 miljoonaa euroa. Kun ollaan aika varmoja, että Hiidenveden tila on hyvä, kannattavampaa on olla tekemättä kunnostusta. Odotettu tuotto on tällöin 4,56 miljoonaa euroa. Lisätiedolla on silti vielä tällöinkin arvoa, vaikka tilanne jo melkein tiedetäänkin. Jos ollaan melkein varmoja, että tila on huono, kannattaa täydellisestä informaatiosta maksaa 0,28 miljoonaa euroa, ja jos ollaan melkein varmoja, että tila on hyvä, kannattaa maksaa 0,86 miljoonaa euroa. Hyvän tilan varmistamisesta kannattaa taas maksaa enemmän kuin huonon tilan varmistamisesta. Jos järven tila varmistuu hyväksi, toimenpiteitä ei tarvitse tehdä ja toimenpiteisiin tarvittut resurssit säästyvät. Jos taas tila varmistuu huonoksi, resurssit lisäinformaation hankkimiseksi käytetään turhaan. Kunnostustoimenpiteet täytyy vielä toteuttaa ja ne olisi voitu toteuttaa myös ilman tilan varmistamista huonoksi.

Oletus informaation täydellisyydestä ei ole järven ekologisen tilan selvittämisen tapauksessa realistista. Siksi Hiidenvedelle haluttiin laskea myös informaatioarvoja epätäydelliselle informaatiolle samoilla prioreiden arvoilla kuin täydellisen informaation tapauksessa. Taulukossa 7 on esitetty informaatioarvoja epätäydelliselle informaatiolle, eli kun kerätyllä lisätiedolla ei saada täysin varmaa tietoa Hiidenveden tilasta, vaan tietoon liittyy epävarmuutta. Odotetut tuotot (PV) ovat samoja kuin taulukossa 6, sillä niihin informaation epätäydellisyys ei vaikuta.

Ensin taulukossa on esitetty informaatioarvoja epätäydelliselle informaatiolle asian tuntija-arvioon perustuen (taulukko 1). Informaatioarvoksi saadaan tällöin 0–0,32 miljoonaa euroa. Esimerkiksi kun saatavissa on huonoa ennakkotietoa Hiidenveden ekologisesta tilasta, vaihtoehtojen keskimääräinen odotettavissa oleva tuotto laske-

Taulukko 7: Hiidenveden klorofyllimittausten informaatioarvoja epätäydelliselle informaatiolle, kun käytössä on erilaista ennakkotietoa Hiidenveden tilasta. Odotetut tuotot ovat samoja kuin taulukossa 6. Vaihtoehdot ovat, että kunnostus päätetään olla tekemättä (e) tai kunnostus päätetään tehdä (k). Informaatioarvojen laskemiseen on käytetty sekä asiantuntija-arvioon perustuvia ehdollisia todennäköisyyksiä, että sekamallilla arvioituja todennäköisyyksiä $p(y = \text{järven mitattu tila} \mid x = \text{järven oikea tila})$.

Prioritiedon laatu	Ekologisen tilan todennäköisyys $p(x)$		Valitun vaihtoehdon odotettu tuotto, PV (milj. €)	Asiantuntija-arvioon perustuva epätäydellinen tieto		Malliin perustuva epätäydellinen tieto	
	Huono	Hyvä		$PoV(y)$ (milj. €)	$VOI(y)$ (milj. €)	$PoV(y)$ (milj. €)	$VOI(y)$ (milj. €)
Huonoa tietoa	0,5	0,5	4,30 (k)	4,30	0	4,31	0,01
Viitteellistä tietoa	0,6	0,4	4,30 (k)	4,30	0	4,30	0
	0,4	0,6	4,30 (k)	4,30	0	4,52	0,22
Viitteellistä parempaa tietoa	0,8	0,2	4,30 (k)	4,30	0	4,30	0
	0,2	0,8	4,56 (e)	4,88	0,32	4,93	0,37

taan kaavalla (10):

$$\begin{aligned}
 PoV(y) &= \max_{a \in A} \{0 \cdot 0,36 + 5,7 \cdot 0,64; 4,3 \cdot 0,36 + 4,3 \cdot 0,64\} \cdot 0,7 \\
 &\quad + \max_{a \in A} \{0 \cdot 0,83 + 5,7 \cdot 0,17; 4,3 \cdot 0,83 + 4,3 \cdot 0,17\} \cdot 0,3 \\
 &= 4,3,
 \end{aligned}$$

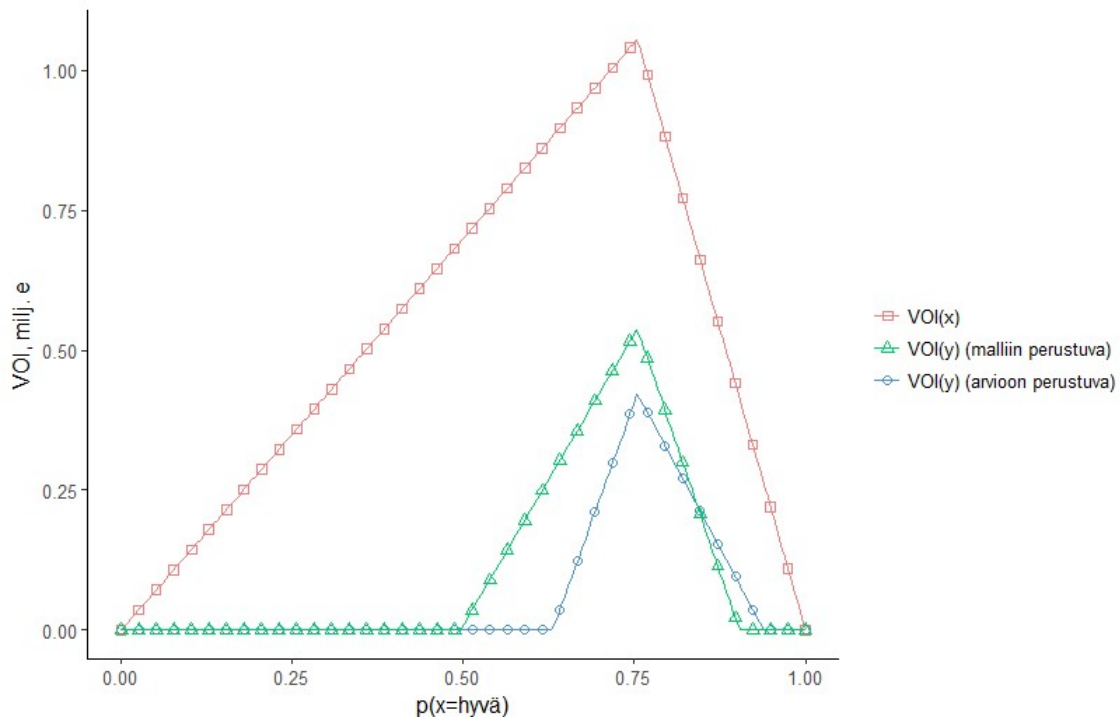
jossa prioritodennäköisyydet $p(y)$ saadaan kokonaistodennäköisyytenä $p(y) = \sum_x p(x)p(y \mid x)$ ja ehdolliset todennäköisyydet $p(x \mid y)$ Bayesin kaavalla. Informaatioarvo lasketaan kaavalla (9).

Sen jälkeen on esitetty informaatioarvoja epätäydelliselle informaatiolle, kun informaation luotettavuutta arvioitiin sekamallin avulla (taulukko 5). Informaatioarvoksi saadaan 0–0,37 miljoonaa euroa. Informaatioarvo epätäydelliselle informaatiolle on pienempi kuin täydelliselle informaatiolle, koska epätäydellisestä tiedosta ei kannata maksaa niin paljon kuin täydellisestä. Lisäksi sekamallin avulla arvioituja ehdollisia todennäköisyyksiä käytettäessä informaatioarvoista saatiin suurempia.

5.3 Informaatioarvoon vaikuttavat tekijät

Lopuksi haluttiin tarkastella, miten ennakkotieto $p(x = \text{järven oikea tila})$ ja ehdolliset todennäköisyydet $p(y = \text{järven mitattu tila} \mid x = \text{järven oikea tila})$ vaikuttavat Hiidenveden klorofyllimittausten informaatioarvoon. Kuvassa 4 on kuvattu informaatioarvo täydellisen ja epätäydellisen informaation tapauksessa, kun ennakkotieto järven ekologisesta tilasta muuttuu. Punainen viiva kuvaa informaatioarvoa täydelliselle informaatiolle, kun ennakkotieto $p(x = \text{järven oikea tila})$ muuttuu. Informaatioarvo täydelliselle informaatiolle on lähes aina suurempi kuin nolla. Ainoastaan silloin, kun on täysin varmaa, että järven tila on hyvä tai huono, lisäinformaatiosta ei enää kannata maksaa mitään. Informaatioarvo on suurin silloin, kun ennakkotiedon mukaan Hiidenveden hyvän tilan todennäköisyys on $p(x = \text{hyvä}) = 0,75$. Informaatioarvo on tällöin 1,06 miljoonaa euroa.

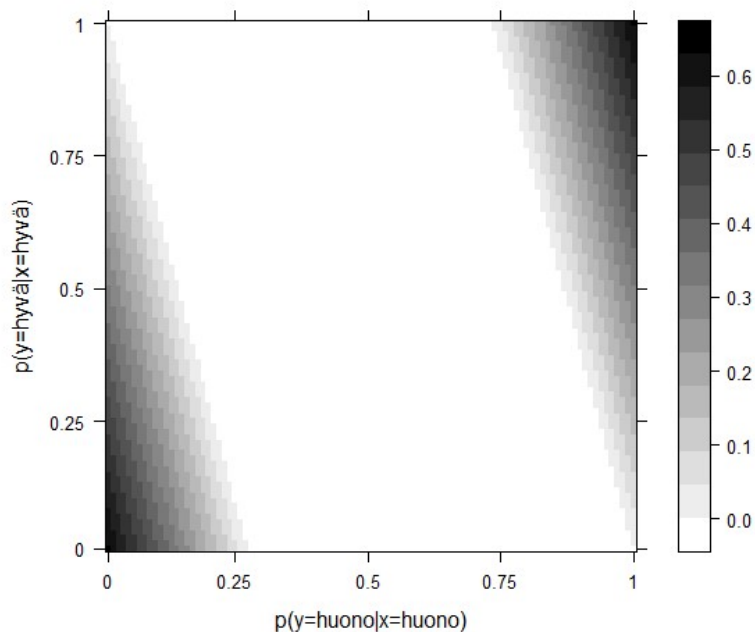
Kuvassa 4 vihreä viiva kuvaa informaatioarvoa epätäydelliselle informaatiolle, kun tiedon luotettavuus on arvioitu sekamallilla (taulukko 5), ja sininen viiva kuvaa informaatioarvoa epätäydelliselle informaatiolle, kun luotettavuus perustuu asiantuntija-



Kuva 4: Prioritodennäköisyyden $p(x = \text{järven oikea tila})$ vaikutus Hiidenveden lisäinformaation arvoon. Punainen viiva (neliö) kuvaa informaatioarvoa täydelliselle informaatiolle. Sininen viiva (ympyrä) kuvaa informaatioarvoa epätäydelliselle informaatiolle, kun epätäydellinen tieto perustuu arvioon (taulukko 1) ja vihreä viiva (kolmio) kuvaa informaatioarvoa epätäydelliselle informaatiolle, kun tieto perustuu sekamalliin (taulukko 5).

arvioon (taulukko 1). Informaatioarvo epätäydelliselle informaatiolle on paljon pienempi kuin täydelliselle. Lisäksi epätäydellinen informaatio on arvotonta myös muulloin kuin täysin varmassa tilanteessa, toisin kuin täydellinen informaatio. Edelleen huonomman epätäydellisen informaation arvo on pienempi kuin paremman epätäydellisen informaation. Informaatioarvo saa suurempia arvoja silloin, kun järven tilan oletetaan olevan hyvä, kuin silloin kun sen oletetaan olevan huono.

Kuvassa 5 tarkastellaan ehdollisen todennäköisyyden $p(y = \text{järven mitattu tila} \mid x = \text{järven oikea tila})$ vaikutusta informaatioarvoon. Ennakkotiedoksi on kiinnitetty $p(x = \text{hyvä}) = p(x = \text{huono}) = 0,5$. Vaaka-akselilla on todennäköisyys $p(y = \text{huono} \mid x = \text{huono})$ (esimerkiksi taulukon 5 ensimmäinen rivi) ja pystyakselilla todennäköisyys $p(y = \text{hyvä} \mid x = \text{hyvä})$ (taulukon 5 toinen rivi). Värin tummuus kertoo informaatioarvon suuruuden; mitä tummempi väri, sitä suurempi on informaatioarvo. Informaatioarvo on suurimmillaan silloin, kun mittaus antaa täysin varman tiedon järven tilasta ehdolla järven oikea tila. Kerättävä informaatio on tällöin täydellistä, ja informaatioarvo jo laskettu 0,7 miljoonaa euroa. Informaatioarvo alkaa nopeasti laskea, mitä huonommaksi mittauksen tarkkuus muuttuu. Informaatioarvo on 0, kun mittaus on hyvin epäluotettava, eli klorofyllimittauksella



Kuva 5: Ehdollisen todennäköisyyden $p(y = \text{mitattu ekologinen tila} \mid x = \text{oikea ekologinen tila})$, eli tiedon epätäydellisyyden vaikutus Hiidenveden informaatioarvoon. Vaaka-akselilla todennäköisyys $p(y = \text{huono} \mid x = \text{huono})$ ja pystyakselilla $p(y = \text{hyvä} \mid x = \text{hyvä})$. Väri kertoo informaatioarvon suuruuden; mitä tummempi väri, sitä suurempi on informaatioarvo.

ei onnistuta selvittämään järven ekologista tilaa oikein.

Jos kiinnitetään todennäköisyys mitata järven tila varmasti huonoksi ehdolla että se on oikeasti huono, eli $p(y = \text{huono} \mid x = \text{huono}) = 1$, informaatioarvo on suurin silloin, kun saadaan mitattua myös täysin varmasti hyvä tila hyväksi. Kun hyvän tilan mittaustodennäköisyys pienenee, informaatioarvo pienenee, mutta ei saavuta nollaa. Jos kiinnitetään todennäköisyys mitata järven tila varmasti hyväksi ehdolla että se on oikeasti hyvä, eli $p(y = \text{hyvä} \mid x = \text{hyvä}) = 1$, informaatioarvo on suurin taas, kun saadaan mitattua täysin varmasti huono tila huonoksi. Kun huonon tilan mittaustodennäköisyys pienenee, informaatioarvo pienenee nopeasti nollaan.

6 Yhteenveto

Tämän työn tavoitteena oli avata informaatioarvon käsitettä ja soveltaa sitä järvien ekologisen tilan määrittämiseksi kerättävään seuranta-aineistoon. Tarkoituksena oli myös tutkia, miten kerättävän tiedon luotettavuus eli epätäydellinen informaatio vaikuttaa informaatioarvoon. Vastaavaa menetelmää ei ollut aiemmin käytetty järvien seuranta-aineistojen arvon arviointiin. Saatiin osoitettua, että informaatioarvon käsitettä voidaan soveltaa järvistä kerättävien klorofyllimittausten arvon arviointiin ja arvoja saatiin laskettua. Aiemmin vastaavaa ovat tehneet Nygård et al. (2016), kun he sovelsivat informaatioarvon käsitettä merien seuranta-aineistoon Suomessa. Tässä laskentaa vietiin pidemmälle kuin esimerkiksi heidän tutkimuksessaan: informaatioarvoja laskettiin myös epätäydelliselle informaatiolle täydellisen informaation lisäksi.

Soveltamalla informaatioarvoa Hiidenveden seuranta-aineistoon saatiin näytettyä, että mitä parempaa ja luotettavampaa kerättävä tieto järvestä on, sitä enemmän siitä kannattaa maksaa. Lisäksi saatiin selville, että resursseja ei kannata kohdentaa sellaisiin järviin, joissa ennakkotieto järven ekologisesta tilasta on jo vahva. Sen sijaan suurimmat resurssit tiedon hankkimiseksi kannattaa suunnata sellaisiin järviin, joista ennakkotietoa ekologisesta tilasta ei vielä ole tai tila on hyvin epävarma. Erityisesti informaatio on arvokkaampaa silloin, kun ollaan ennakkoon varmempia siitä, että järvi on hyvässä kuin että se on huonossa tilassa. Tällöin kannattaa kerätä informaatiota ja varmistua siitä, että järvi on hyvässä tilassa, jolloin kalliit kunnostustoimenpiteet järvelle voidaan jättää tekemättä. Jos sen sijaan ollaan miltei varmoja, että tila on huono, kunnostustoimenpiteet kannattaa toteuttaa ilman lisätietojen keräystä.

Myös informaation luotettavuuden arviointi lineaarisen sekamallin avulla onnistui, ja saatiin näytettyä, miten informaation luotettavuuden parantaminen vaikutti informaatioarvoon. Mallin hyvyttä ei kuitenkaan tässä työssä tarkemmin tarkasteltu, ja malli onkin hyvin yksinkertainen. Mallia voisi parantaa esimerkiksi ottamalla mukaan kovariaatteja. Toisaalta nyt työssä käytetty malli on yksinkertaisuudessaan vertailukelpoinen vertailujärvien jakaumasta saatavien luokkarajojen kanssa.

Vaikka osoitettiin, että informaatioarvon käsitettä voidaan soveltaa järvien seurantaan, sovelluksessa oli myös omat haasteensa. Osoittautui, että informaatioarvon käsitteen soveltaminen ei onnistu, mikäli järvestä ei ole minkäänlaisia ennakkotietoja. Haastetta toi erityisesti järven hyvän tilan arvon määrittäminen, joka osoittautui työn miltei vaikeimmaksi tehtäväksi. Luonnon arvon muuntaminen rahalliseen ar-

voon on tutkimuskysymys, johon ei vielä ole saatu kunnollista toimintakehikkoa; tutkimus on ollut vähäistä ja menetelmät kehittämissvaiheessa. Suomessa vesien-suojelutoimenpiteiden ja vesien kunnostuksilla aikaansaatuisten hyötyjen arvojen arvioinnit ovat jääneet kuvailevalle tasolle ja niiden suhteuttaminen toimenpiteiden taloudellisiin kustannuksiin ei ole ollut mahdollista (Ahtiainen, 2008). Informaatioarvon soveltamiseksi jonkinlaisen arvon määrittäminen oli kuitenkin välttämätöntä. Koska arvon määrittäminen järven hyvälle tilalle oli vaikeaa, myös informaatioarvon soveltaminen tähän sovellukseen oli siltä osin haastavaa.

Toinen haaste informaatioarvon soveltamisessa järvien seurantaan liittyy ekologisen tilan käsitteeseen. Järven ekologisen tilan määritelmään liittyy jo itsessään epävarmuutta. Ekologian tutkimuksissa indikaattoria, jonka avulla päätelmiä luonnon tilasta tehdään, pidetään hyvänä indikaattorina jo silloin, kun se määrittää oikein 80 prosentin varmuudella (Nygård et al., 2016). Ymmärryksemme ympäristöstä on siten vajavaista ja epävarmuus on aina suurta. Lisäksi tässä tutkimuksessa rajoituttiin ekologisen tilan määrittämiseen vain yhden indikaattorin, klorofyllipitoisuuden, avulla. Todellisuudessa pelkän klorofyllipitoisuuden avulla ei tehdä vielä päätelmiä vesistön tilasta, vaan ekologinen luokittelu tehdään käyttäen hyväksi useista indikaattoreista kerättyjä tietoja.

Tässä työssä informaatioarvo määrättiin diskreetille muuttujalle mitata ekologinen tila hyväksi tai huonoksi klorofyllimittauksen avulla. Tällä tavalla menetetään informaatiota, sillä klorofyllipitoisuuden suuruudella on väliä: mitä suurempi klorofyllipitoisuuden arvo mitataan, sitä vahvemmin se indikoi huonoa tilaa. Luultavasti informaatiolla olisi suurempi arvo, jos käsiteltäisiin jatkuvia klorofyllimittauksen arvoja sellaisenaan muuttamatta muuttujaa luokittelevaksi muuttujaksi, ja laskettaisiin informaatioarvo suoraan klorofyllimittaukselle.

Lisäksi epävarmuutta liittyy järven kunnostukseen ja sen seurauksiin. Vaikka kunnostuksiin käytetään varoja, ne eivät välttämättä tuo haluttua tulosta ja tavoiteltua jäätä saavuttamatta. Esimerkiksi Hiidenveden tilan parantamiseksi on tehty pitkään töitä, mutta havaittavaa muutosta järven tilassa ei ole tapahtunut (Ahtiainen, 2008; Alaja et al., 2012).

Työn ulkopuolelle rajautui vielä monta aiheeseen liittyvää tutkimuskysymystä. Tässä työssä ei esimerkiksi tutkittu, kuinka paljon informaatiota pitäisi kerätä tai kuinka usein mittauksia pitäisi kerätä järvistä ekologisen tilan selvittämiseksi. Tämä on oma tutkimusaiheensa. Näytteenottomäärien vaikutusta informaatioarvoon voisi tutkia esimerkiksi mallin avulla ja selvittää, onko olemassa optimaalista näytteen-

ottotiheyttä resurssien optimoimiseksi. Myöskään informaation hankinnan kustannuksia ei tässä työssä käsitelty eikä siten voitu selvittää, mitä ovat suuruudeltaan informaation hankinnan kustannukset verrattuna informaatiosta saatavaan hyötyyn. Tätä esimerkiksi Nygård et al. (2016) ovat tarkastelleet. Tämän työn tarkoituksena ei myöskään ollut kattavan näytteenottojärjestelmän suunnitteleminen. Tarkastelu oli tällä erää rajattava sellaiseen esimerkijärveen, jolle oli toteutettu hyvän ekologisen tilan hyödyn arvottamista. Voisi olla mielenkiintoista laajentaa tarkastelu koskemaan esimerkiksi yhtä järvityyppiä.

Kiitokset

Tahtoisin kiittää työhöni myötävaikuttaneita henkilöitä. Kiitokset FT Niina Kotamäelle työni alkuun panemisesta, asiantuntevasta ohjauksesta ja selvitystyöstä. Kiitokset FT Heikki Hämäläiselle arvokkaista neuvoista ja materiaaleista. Kiitokset FT Kristian Meissnerille ja dosentti Laura Uusitalolle, jotka jaksoivat kuulla työni etenemisestä ja antaa neuvoja ja tukea. Kiitokset professori Juha Karvaselle työn kommentoinnista. Erityiskiitos kuuluu ohjaajalleni FT Salme Kärkkäiselle, jota ilman työni olisi moneen kertaan päättynyt umpikujaan.

Viitteet

- Ahtiainen, H. (2008). *Järven tilan parantamisen hyödyt. Esimerkkinä Hiidenvesi*. Helsinki: Suomen ympäristökeskus (SYKE).
- Alaja, H., Sundell, P., Palomäki, A., & Hynynen, J. (2012). *Hiidenveden kunnostus- ja hoitosuunnitelma – Osa II ravintoketjukunnostus*. Jyväskylän yliopisto, Ympäristöntutkimuskeskus. Tutkimusraportti 194/2012.
- Aroviita, J., Hellsten, S., Jyväsjärvi, J., Järvenpää, L., Järvinen, M., Karjalainen, S. M., Kauppila, P., Keto, A., Kuoppala, M., Manni, K., Mannio, J., Mitikka, S., Olin, M., Perus, J., Pilke, A., Rask, M., Riihimäki, J., Ruuskanen, A., Siimes, K., Sutela, T., Vehanen, T., & Vuori, K.-M. (2012). *Ohje pintavesien ekologisen ja kemiallisen tilan luokitteluun vuosille 2012–2013 – päivitetty arviointiperusteet ja niiden soveltaminen*. Helsinki: Suomen ympäristökeskus (SYKE).
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, 67(1), 1–48.
- Bouma, J., van der Woerd, H., & Kuik, O. (2009). Assessing the value of information for water quality management in the North Sea. *Journal of Environmental Management*, 90(2), 1280 – 1288.
- Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., & Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, 8(3), 559–568.
- Carstensen, J. & Lindegarth, M. (2016). Confidence in ecological indicators: A framework for quantifying uncertainty components from monitoring data. *Ecological Indicators*, 67, 306 – 317.
- de Leeuw, J., Meijer, E., Goldstein, H., & de Deleeuw, J. (2008). *Handbook of Multilevel Analysis*. Springer, New York, NY.
- Demidenko, E. (2013). *Mixed Models : Theory and Applications with R*. Somerset: Wiley.
- Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Boca Raton, Florida, USA: Chapman & Hall/CRC.

- Eidsvik, J., Mukerji, T., & Bhattacharjya, D. (2015). *Value of Information in the Earth Sciences : Integrating Spatial Modeling and Decision Analysis*. Cambridge: Cambridge University Press.
- Euroopan parlamentti (2000). Euroopan parlamentin ja neuvoston direktiivi 2000/60/EY, annettu 23. päivänä lokakuuta 2000 yhteisön vesipolitiikan puitteista.
- Hämäläinen, H., Aroviita, J., Jyväsjärvi, J., & Kärkkäinen, S. (2018). Dangerous relationships: biases in freshwater bioassessment based on observed to expected ratios. *Ecological Applications*, 28(5), 1260 – 1272.
- Howard, R. (1964). Decision analysis: Applied decision theory. *Proceedings of the 4th International Conference on Operational Research*, Wiley-Interscience, (pp. 55 – 71).
- Hyytiäinen, U.-M. (2008). *Hiidenveden hoito ja kunnostus 2005-2007*. Helsinki: Uudenmaan ympäristökeskus.
- Kotamäki, N. (2015). Näytteenottotiheyden vaikutus tila-arvion tarkkuuteen – OPTIMI-hankkeen työraportti. http://www.syke.fi/fi-FI/Tutkimus__kehittaminen/Tutkimus_ja_kehittamishankkeet/Hankkeet/Ymparistoseurannoista_informaation_yhdistamiseen_MONITOR_2020_laatu_palvelukyky_ja_vaikuttavuutta.
- Kotamäki, N., Pätynen, A., Taskinen, A., Huttula, T., & Malve, O. (2015). Statistical dimensioning of nutrient loading reduction: LLR assessment tool for lake managers. *Environmental Management*, 1(56), 480 – 491.
- McCulloch, C. E. & Searle, S. R. (2001). *Generalized, Linear and Mixed Models*. John Wiley & Sons.
- Nygård, H., Oinonen, S., Hällfors, H. A., Lehtiniemi, M., Rantajärvi, E., & Uusitalo, L. (2016). Price vs. value of marine monitoring. *Frontiers in Marine Science*, 3(205).
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2017). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-131.
- Pinheiro, J. C. & Bates, D. M. (2002). *Mixed-Effects Models in S and S-PLUS*. Springer, New York, NY.

- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Stoddard, J. L., Larsen, D. P., Hawkins, C. P., Johnson, R. K., & Norris, R. H. (2006). Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications*, 16(4), 1267–1276.
- Tolonen, K. T., Hämäläinen, H., Lensu, A., Meriläinen, J. J., Palomäki, A., & Karjalainen, J. (2014). The relevance of ecological status to ecosystem functions and services in a large boreal lake. *Journal of Applied Ecology*, 51, 560–571.
- Vuori, K.-M., Mitikka, S., & Vuoristo, H. (2009). *Pintavesien ekologisen tilan luokittelu*. Helsinki: Suomen ympäristökeskus (SYKE).