



SUOMEN TILASTOSEURAN VUOSIKIRJA 2025

ÅRSBOK FÖR STATISTISKA SAMFUNDET
I FINLAND 2025

THE YEARBOOK OF
THE FINNISH STATISTICAL SOCIETY 2025

2025

Sisällys

Puheenjohtajan palsta: Funktiooni	3
PEKKA PERE	
Sihteerin ja rahastonhoitajan palsta	12
ROOPE RIHTAMO, ARVI TOLVANEN	
Tilastotieteen rooli ilmastonmuutoksen arvioinnissa.	
Suomen Tilastoseuran iltapäiväseminaari 9.9.2025	13
Avauspuheenvuoro	14
JUHA ALHO	
Kasvihuonekaasuinventaario: periaatteet, ohjeistus ja politiikkatoimien vaikutus	20
SINI NIINISTÖ	
Suomen kivennäismaametsien hiilitaseen epävarmuuden arviointi	23
JUHA HEIKKINEN	
Kommentti: Joitakin haja-ajatuksia vaihtelusta, satunnaisuudesta ja epävarmuudesta	34
ELJA ARJAS	
Kommentti: Muutamia huomioita Suomen metsien kasvihuonekaasuinventoinnista	38
ERKKI TOMPPU	
Vastine: Suomen kivennäismaametsien hiilitaseen epävarmuuden arviointi	45
JUHA HEIKKINEN	
Tilastopäivät 9.-10.6.2025	46
Ryväsmuuttujat kausaalipäätelyssä: silta potentiaalisten vasteiden ja graafisten mallien välillä	46
JUHA KARVANEN	
On the Impact of Approximation Errors on Extreme Quantile Estimation with Applications to Functional Data Analysis	49
JAAKKO PERE	
Hyvät priorit ja suolaa päälle — mitä voimme oppia terveysrekisteriaineistosta Markovin piilomalleilla?	51
AAPELI NEVALA	
Tilastojulistekilpailu 2024–2025	53
REIJA HELENIUS	
Vaarallinen tiede ja ammatti	57
PEKKA PERE	

Joviaali kolmen lajin proffa	66
LASSE KOSKINEN, MARKKU SOINTU	
Väitöskirjapalkinto	68
Non-linear state-space methods for Bayesian time series modelling	68
SANTERI KARPPINEN	
Leo Törnqvist -palkinto	80
A Bayesian state-space model for Baltic ringed seal demography	80
MURAT ERSALMAN	
Suomen Tilastoseuran hallitus vuonna 2025	82
Gunnar Modeen -minnesmedaljen	83
Scandinavian Journal of Statistics	85
Palkinnot	86
Leo Törnqvist –palkinnot	86
Väitöskirjapalkinnot	88
Suomen Tilastoseuran julkaisuja	89
Tilastotieteellisiä tutkimuksia	90
Suomen Tilastoseuran vuosikirja	92
Muita julkaisuja	93

Puheenjohtajan palsta: Funktsooni

PEKKA PERE

PEKKA.J.PERE@AALTO.FI

MATEMATIIKAN JA SYSTEEMIANALYYSIN LAITOS
AALTO-YLIOPISTO

Mikä toimii, sitä ei huomata. Tilastoseuran hallituksen jäsen JYRKI MÖTTÖNEN (Helsingin yliopisto) uudisti kotisivumme (<https://tilastoseura.fi/>) sisällönhallintajärjestelmän alkuvuodesta 2025 ja on ylläpitänyt kotisivuamme ajantasaisena. Toiminta sujui seurassa muutenkin menneen vuoden aikana. Silloin ei tule välttämättä ajatelleeksi, että taustalla on tilastotieteelle omistautuneita ihmisiä tekemässä kaiken mahdolliseksi. Hallitus kokoontui yhdeksän kertaa vuoden 2025 aikana. Kiitän niin Jyrkiä kuin kaikkia muita seuran hallituksen jäseniä seuran eteen tehdystä työstä vuonna 2025.

Tieteiden yö 9.1.2025

Tiedevuosi käynnistyi aiempaan tapaan Tieteiden yöllä 9.1. Tilastoseura oli kutsunut JESSICA NISÉNin (Turun yliopisto) luennoimaan suomalaisia askarruttavasta syntyvyyden kehityksestä. Luennon otsikko oli “Mistä tulet, minne menet, syntyvyys?”.

Tilastoseuran hallituksen jäsenet ROOPE RIHTAMO ja ARVI TOLVANEN esittelivät niin toimintaamme kuin tilastotiedettä seuran esittelypisteellä. He havaitsivat kiinnostuneille binomi- ja normaalijakauman kytköstä seuran Galtonin koneella ja otantateoriaa interaktiivisella kyselytutkimusteemaisella Rock 'n Poll -sivulla (<https://rocknpoll.graphics>).

Tilastopäivät 9.–10.6.2025

Tilastoseuran tärkeimpiä toiminnan muotoja ovat Tilastopäivät. Tällä kertaa ne olivat Tampereen yliopistolla 9.–10.6. Päivien ohjelma oli monipuolinen hyvin empiirisestä hyvin teoreettiseen:

Maanantai 9.6.2025

11.00–12.00	Rekisteröityminen Tilastopäiville	
12.00–13.45	Istunto 1	
12.00–12.15	Pekka Pere	Tilastopäivien avaus
12.15–13.15	Steve Pischke	The merits of pre-testing in difference-in-differences models
13.15–13.45	Juha Karvanen	Clustered variables in causal inference: Building on potential outcomes and graphical models
13.45–15.30	Lounas ja posterinäyttely	
15.30–16.30	Istunto 2	
15.30–16.00	Jaakko Pere	On the impact of approximation errors on extreme quantile estimation with applications to functional data analysis
16.00–16.30	Tapio Nummi	On the improved estimation of the normal mixture components for longitudinal data
16.30–17.15	Kahvi ja posterinäyttely	
20.00–23.00	Konferenssi-illallinen ravintola Puistossa	

Tiistai 10.6.2025

9.00–10.30	Istunto 3: Tutkimuksen näyteikkuna (kaksi esitelmää)	
9.00–9.30	Aapeli Nevala	Bayesian hidden Markov model for natural history of colorectal cancer: Handling misclassified observations, varying observation schemes and unobserved data
9.30–10.00	Leena Kalliovirta	Modelling microclimate temperature buffering capacity using mixed models
10.00–10.30	Lauri Viitasaari	On data-driven forecasting of financial time series
10.30–11.15	Kahvi ja posterinäyttely	
11.15–12.15	Istunto 4	
11.15–11.45	Gerardo Iniguez Gonzalez	Universal patterns in egocentric communication networks
11.45–12.15	Jouni Helske	FAN-HMMS for longitudinal causal inference: Workplace dynamics in paternal leave taking
12.15–14.00	Lounas ja posterinäyttely (äänestys päättyy)	
14.00–15.00	Istunto 5	
14.00–14.30	Arto Luoma	Effects of speeding and alcohol usage on fatal crashes: estimating relative risk curves from partial data

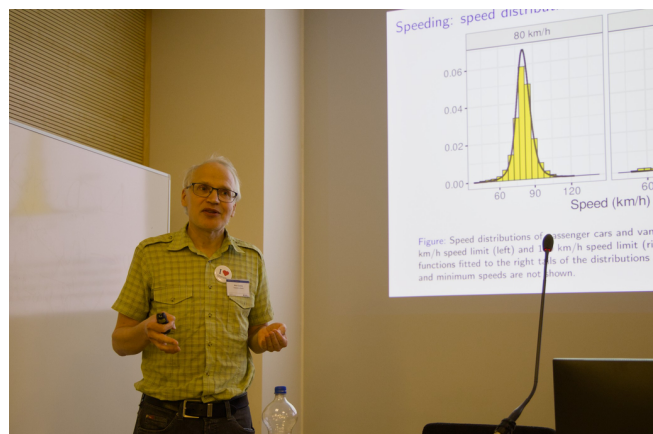
14.30–15.00	Jyrki Möttönen	Posterikilpailun, Leo Törnqvist –palkinnon ja väitöskirjapalkinnon voittajien julkistaminen
15.00	Pekka Pere: Tilastopäivien päättäminen	



Kuva 1: Pääpuhujana oli Steve Pischke.

Pääpuhujaksi olimme saaneet STEVE PISCHKEN (London School of Economics and Political Science, Kuva 1). Pischke on julkaissut syy-seuraus päättelystä kaksi oppikirjaa taloustieteen nobelisti JOSHUA ANGRISTIN kanssa. Pääpuhujan olimme hakeneet tästä aihepiiristä, koska seuran tämänvuotinen kesäkurssi oli niin ikään syy-seuraus päättelystä. Esitelmöinnin lisäksi Pischke kommentoi aktiivisesti muita esitelmiä.

Tutkimuksen näyteikkunassa esiintyivät LEE-NA KALLIOVIRTA (Helsingin yliopisto) ja AAPELI NEVALA (Suomen syöpärekisteri). Päiviä isännöivää Tampereen yliopistoa edusti muun muassa Arto Luoma (Kuva 2). Kaksi esitelmää peruuntui odottamattomien esteiden takia. Erityiskiitos JOUNI HELSKEelle (Turun



Kuva 2: Arto Luoma.

yliopisto), jonka saimme rekrytoitua yleisöstä korvaamaan sairastumisen takia peruuntuneen kolmannen esitelmän.

Posterikilpailussa oli vain kaksi osanottajaa. Voittaja oli OLLI SARALA (Oulun yliopisto) posterilla “Bayesian estimation of sparse partial correlation graphs with smooth temporal evolution”. Tutkimuksen muita tekijöitä olivat M. Kuusimäki, T. Hautamäki, A. E. Korhonen ja M. J. Sillanpää. Leo Törnqvist -palkinto parhaasta pro gradu -tutkielmasta myönnettiin MURAT ERSALMANille (Helsingin yliopisto). Tutkielman aihe oli “A Bayesian state-space model for Baltic ringed seal demography”. SANTERI KARPPINEN (Jyväskylän yliopisto) palkittiin parhaasta väitöskirjasta edeltävien neljän vuoden aikana. Väitöskirjan aihe oli “Non-linear state-space methods for Bayesian time series”. Palkitut saivat kunniakirjan sekä rahapalkinnon (250, 500 tai 1 000 euroa).

Rekisteröityneitä osallistujia oli päivillä 58, ja illallisravintolaan tuli 44 innokasta tilastotieteen ammattilaista tai harrastajaa. Tunnelma Tammerkosken rannalla ravintola Puistossa oli illalla korkealla. Sijoittelumme ravintolassa olisin tosin suonut olleen yhtenäisempi. Muistamme sen seuraaville Tilastopäiville ravintolaa valittaessa.

Kiitän seuran puolesta Tampereen yliopiston tietotekniikan yksikköä tilojen tarjoamisesta ja kahvien sponsoroinnista. Kiitos myös Jenny ja Antti Wihurin rahastolle sekä Suomalaisen Tiedekatemian matematiikan rahastolle avustuksista. Suurin kiitos esitelmöitsijöille, posterikilpailijoille sekä osallistujille, koska te viime kädessä teitte päivät.

Kesäkurssi 9.–13.6.2025

Seuran kesäkurssi järjestettiin 9.–13.6. Tampereen yliopistolla rinnan Tilastopäivien kanssa. Kurssin on tarkoitus olla vuorovuosin tilastollisesta päättelystä ja erityisesti aiheesta. Kesällä 2025 oli vuorossa ensimmäinen kurssi erityisesti aiheesta – nyt syy-seurauspäättelystä. JUHA KARVANEN (Jyväskylän yliopisto) piti kurssin (<https://tilastoseura.fi/causalinference>) englanniksi. Kurssin seurasi alusta loppuun 21 opiskelijaa. Moni oli tohtori, joten yhtä moni ei tenttinyt kurssia. Tentin läpäisi 9 opiskelijaa 11 osallistujasta. Parhaan arvosanan 5 sai 4 opiskelijaa. Juha Karvanen kertoi minulle olleensa hyvin tyytyväinen kurssiin: Opiskelijat olivat olleet hyvin lahjakkaita ja aktiivisia kysymään ja keskustelemaan. Kurssille osallistunut kollegani kehui kurssia erinomaiseksi ja kertoi voivansa hyödyntää sitä tutkimustyössään. Kollegani oli yksi tenttiin osallistumattomista tohtoreista.

Seura on järjestänyt nyt kaksi kesäkurssia. Kokemukset molemmista ovat erittäin kannustavia.

Tilastotieteen sanaston 3. laitos

Yksi vuoden merkkitapahtumista oli Tilastotieteen sanaston 3. laitoksen (Alho ym. 2025) julkistaminen kesällä 2025. Sanaston 1. laitoksen 2021 olivat laatineet JUHA ALHO, ELJA ARJAS, ESA LÄÄRÄ ja allekirjoittanut. Sanastoa päivittävät tätä nykyä lisäksi Juha Karvanen, LASSE LESKELÄ ja Esa Läärä. Sanaston uusin versio on kaikkien edellä mainittujen nimissä.

Kerroin edellisessä Vuosikirjassa, että puheenjohtajan palstaa kirjoittaessani sanastosta oli edeltävän täyden viikon arkipäivinä tehty keskimäärin 1 236

hakua päivässä. Nyt vastaava keskiarvo on 1 637, kasvua on yli 30 %:a. Tätä kirjoittaessani tammikuussa 2026 huomaan, että kuluvalle viikolle on kirjautunut uusi ennätys 7 275 hakua päivässä.

Englanti-suomi-sanastossa on jo 3 700 termiä. Sanastoryhmämme toivoo, että seuran jäsenet ja muut tilastotieteen termien käyttäjät tekevät aktiivisesti ehdotuksia uusista suomennettavista termeistä sanastosivun Ehdota termiä -kuvakkeen kautta. Niin pidämme tilastotieteen suomenkielisen terminologian ajantasaisena ja laadukkaana.

Miksi on tärkeää vaalia tilastotieteen suomenkielistä termistöä? Sitä uhkaa pirstoutuminen, kun tilastotieteen käyttäjäkunta laajenee ja datatieteessä suomennetaan uudelleen termejä, joille on jo olemassa vakiintuneet suomennokset. Termistö uhkaa myös raakalainaistua, kun tilastotiedettä opetetaan Suomessa yhä useammin englannin kielellä suomen sijaan ja opiskelijat käyttävät englanninkielisiä termejä likimain sellaisenaan suomen seassa.

Tilastotieteen opiskelijoiden nykykielessä on yhtäläisyyksiä 1800-luvun ja 1900-luvun alun tieteen kieleen, jossa raakalainat rön sylivät: Puhekielessä rändöm korvaa sattuman. Nykyopiskelija voi puhua samplaamisesta, signifikanssista, buutsträppäämisestä, assumptioneista, mapittamisesta, itsenäisyydestä ja niin edelleen tilastotieteellisinä tai matemaattisina termeinä. Puhe vertautuu matematiikan vanhaan spesiaalsiin suomeen esimerkiksi Polyteknillisen opiston opetusohjelmassa 1901. Siitä löytyy termit ratsionaalinen funktiooni, funktioonein differentisieeraus, infleksioonipiste, integratiooni, rektifikatiooni, partiaaliset ja totaaliset differentiaalit, komplansiooni ja partiaalinen differentiaaliekvatiooni. (Polytekniska institutet i Finland 1901, 9–11.) Ero 1800- ja 2000-luvun opiskelijan välillä on, että nykyopiskelija saattaa lähestyä luennoitsijaa englanniksi, vaikka molempien äidinkieli on suomi. Turun Akatemian ylioppilas ei varmaan kirjoittanut kaksisataa vuotta sitten henkilökohtaista asiaansa professorille latinaksi, vaikka se oli tieteen kieli tuolloin kuten englanti on nyt. On palattu kauemmas kuin menneisyyteen tai menty aivan uuteen ulottuvuuteen. Tämä kinttupolku ei ole kehitystä vaan taantumista.

Yksi huoli tieteen kielestä on sama 1800- ja 2000-luvuilla: Eristäytyvätkö suomalaiset tieteilijät, jos kirjoitamme suomeksi? Ero entiseen on, että tiedämme, mitä historiassa seurasi suomeksi kirjoittamisesta: suomen nousu sivistyskieleksi, itsenäistyminen ja kansainvälistyminen. Oman kielen vaaliminen ei tietenkään tarkoita eristäytymistä. Kansainvälistyä ei voi, jos ei ole omaa, mistä kansainvälistyä.

JOHAN EKLÖFin todenvaihe-laskun eli pienimmän neliösumman menetelmän sovellus 1850 oli ensimmäinen suomenkielinen tilastotieteellinen ja melkein ensimmäinen suomenkielinen tieteellinen artikkeli (Alho ym. 2025, 27). Eklöfin suomennoksista on hienosti edetty tilastotieteen suomenkieliseen rikkaaseen nykitermistöön. Sillä tiellä kannattaa jatkaa.

En tohdi verrata uutta sanastoa moderniin luotijunaan. Ihan sellaisen kyytiin ei olla sentään päästy, mutta vahvaa konetta seura rakentaa.

Iltapäiväseminaari 9.9.2025

Seura järjesti 9.9. iltapäiväseminaarin ”Hiilitaseen arvioinnin tilastolliset ongelmat”. Seminaarin pontimena olivat tiedot, että Suomen metsät eivät olekaan hiilinielu kuten oli ajateltu. Tilastotieteellä olisi varmaan asiasta sanottavaa, sillä onhan epävarmuuden mittaaminen tilastotieteen ydintä. Seminaari oli (viime vuoden iltapäiväseminaarin tapaan) Juha Alhon ideoima, ja hän käytännös-

sä organisoisi sen. Kiitän Juhaa ajankohtaisesta seminaarista tärkeästä aiheesta. Ohjelma löytyy sivulta [13](#).

Tilaisuuteen ilmoittautui 34 kiinnostunutta. Laskujeni mukaan kaikki tulivat paikalle. Kiinnostuneita olisi ollut enemmänkin, sillä minulta toivottiin etäyhteyden järjestämistä seminaariin.

Vuosikirjassa on toisaalla seminaarin esitelmiä. ERKKI TOMPPO (emeritusprofessori, Luke) ei esitelmöinyt mutta kommentoi seminaarin lopussa niin asiantuntevasti, että hänen kommenttinsa ovat Vuosikirjassa artikkelina. Erkki Tomppo kirjoitti perään Helsingin Sanomiin 24.9. hiilinielujen mittaamisen epävarmuudesta (<https://www.hs.fi/mielipide/art-2000011495502.html>). Yle oli 9.4. uutisoinut Erkki Tompon Metsälehteen hiilinieluista kirjoittamasta artikkelista (<https://yle.fi/a/74-20154820>). Seminaarin aihe oli polttava. JUHA HEIKKINEN korosti esitelmässään, kuinka epävarmaa hiilinielun suuruuden mittaaminen on. Loppuvuodesta uutisointiinkin Luonnonvarakeskuksen 15.12. päivitetystä arviosta, että metsät ovat sittenkin hiilinielu (<https://www.luke.fi/fi/uutiset/maankayttosektori-pysyy-suurena-paastolahteena-mutta-met-sat-ovat-pieni-nielu-kasvihuonekaasuinventaarion-2024-ennakkotiedoissa>, <https://www.hs.fi/politiikka/art-2000011697266.html>).

Tieteellisten seurain valtuuskunta 22.9.2025

Viimevuotisella puheenjohtajan palstalla kerroin yrityksestämme saada seuramme edustajan Tieteellisten seurain valtuuskunnan (TSV) hallitukseen, muttemme onnistuneet. TSV:n hallitukseen valittiin 22.9. kolme uutta jäsentä ja heille varajäsenet. Seuramme jäsen ehdokas oli tilastotieteen professori Juha Karvanen varajäsen ehdokkaana sovelletun matematiikan professori Lasse Leskelä (Aalto-yliopisto). Valtuutettuja oli paikalla 36. Kukin seura sai äänestää 1–3 ehdokasta. Ehdokkaita asettaneet seurat saivat kertoa, miksi kunkin ehdokkaat tulisi valita hallitukseen. Puhuin tilastotieteen laaja-alaisuudesta ja osoitin yhteyden kaikkiin ehdokkaita asettaneiden seurojen oppiaineisiin. Kerroin Kotikielen Seura ja Suomalais-Ugrilainen Seura mielessäni, kuinka kesällä julkaisimme Tilastotieteen sanaston uuden laitoksen ja motivaatiosta sille eli huolestamme suomen kielen vaalimisesta tieteen kielenä. Ympäristötutkijoille kerroin toissaviikkosesta iltapäiväseminaaristamme hiilinieluista samassa salissa, jossa olimme nyt. Lakimiehille kerroin oikeuden päätösten algoritmisesta mallittamisesta. Viestini oli, että tilastotiede on upea tiede, jolla on yhteys tieteeseen kuin tieteeseen. Äänestyslappuja jätettiin 37, joista yksi oli tyhjä. Ehdokkaistamme tuli äänikuninkaita 21 äänellä! Seuraavat hallitukseen valitut saivat 19 ja 17 ääntä. Upea tulos niin seurallemme kuin uusille edustajillemme TSV:n hallituksessa. Kiitos erinomaisille ehdokkaillemme ja menestystä heille uudessa tehtävässään!

Kansainvälinen toiminta

SANGITA KULATHINAL (Helsingin yliopisto) pyysi vapautusta tehtävästään Scandinavian Journal of Statistics -säätiön hallituksessa seuran edustajana. Seura haki Statnetin kautta tehtävästä kiinnostuneita. Hallitus nimesi Lasse Leskelän tehtävään edustamaan seuraa. Juha Karvanen jatkaa säätiön hallituksessa toisena suomalaisena seuran edustajana.

Minä ja hallituksemme jäsen REIJA HELENIUS edustimme Tilastoseuraa Euroopan tilastoseurojen liitossa FENStatSissa (The Federation of European National Statistical Societies, <https://www.fenstats.eu>) ja minä ECAS-järjestössä

(European Courses in Advanced Statistics, <https://ecas.fenstats.eu/courses>). ECASin kokouksia oli vuonna 2025 vain yksi.

FENStatS kokousti 26.5. tilastotieteilijän akkreditointi -menettelystä. Yhdysvaltain tilastoseuran ASA:n (American Statistical Association) toiminnanjohtaja RONALD WASSERSTEIN vieraili kokouksessa kertomassa kokemuksista akkreditoinnista Yhdysvalloissa. Akkreditointeja on myönnetty Yhdysvalloissa viisitoista vuotta, ja kokemukset ovat hyviä. Akkreditoitu tilastotieteilijä -nimike kohentaa tilastotieteilijöiden arvostusta erityisesti työpaikoilla, joissa on vain yksi tilastotieteilijä. Akkreditoinnin edistäminen on FENStatSin ensisijaisia hankkeita. Viimeisimmät hankkeeseen liittyneet maat ovat Ranska ja Luxemburg. Keskustelin hankkeen vastuuhenkilön MAGNUS PETERSSONIN kanssa videopuhelussa kesällä, miten seuramme voisi liittyä mukaan. Olemme tiedustelleet jäsenistöltä kiinnostusta akkreditointiin, mutta sitä ei tunnu olevan paljoa. Menettely vaatisi akkreditoijia jäsenistöstämme. Olkaa yhteydessä minuun, jos haluaisitte ryhtyä akkreditoijaksi tai edistämään hanketta muuten seurassa ja Suomessa. Lisätietoja akkreditoinnista on täällä: <https://www.fenstats.eu/accreditation>.

FENStatSin puheenjohtaja LOLA UGARTE viestitti vuoden lopulla FENStatSin sopimuksesta Springer-kustantamon kanssa. Springer tulee julkaisemaan uutta kirjasarjaa yhteistyössä eurooppalaisten tilastoseurojen kanssa. Springerille kannattaa tarjota kirjaa kustannettavaksi. Tarkempia tietoja löytyy linkistä <https://link.springer.com/series/60483>. Puheenjohtaja Ugarte kertoi myös FENStatSin ja Ison-Britannian kuninkaallisen tilastoseuran RSS:n (Royal Statistical Society) sopineen, että ne tunnustavat toistensa vahvistamat tilastotieteilijöiden akkreditoinnit.¹

Tilastoseuran hallitus päätti, että Tilastoseuran järjestämissä tilaisuuksissa FENStatSin ulkomaisten jäsenseurojen jäsenten osallistumismaksu on sama kuin Tilastoseuran jäsenten. FENStatS pyrkii saamaan vastaavan periaatteen käytännöksi kaikille jäsenseuroilleen. Hanke edistää eurooppalaisten tilastoseurojen yhteydenpitoa ja yhteistyötä FENStatSin tavoitteiden mukaisesti.

Kattojärjestömme Kansainvälisen tilastoinstituutin ISI:n (International Statistical Institute) alla toimii tilastojen luku- ja käyttötaitoa yhteiskunnassa edistämään pyrkivä Tilastojen lukutaito -projekti (International Statistical Literacy Project, <https://iase-web.org/islp>). Sillä on lähes kaksisataa maavastaavaa, ja se toimii yhdeksässäkymmenessä maassa. Hallituksemme jäsen Reija Helenius johtaa tätä maailman laajuisia projektia (<https://iase-web.org/islp-executive>). Sen järjestämään kansainväliseen tilastojulistekilpailuun osallistui yli 23 000 nuorta. Kilpailussa nuoret oppivat tilastollista tutkimusprosessia tutkien heitä koskettavia aiheita. Tilastoseuran ja Tilastokeskuksen yhteistyössä järjestämään Suomen kilpailuun otti osaa viisisataa nuorta. Projektilla oli oma kutsuttujen esitysten tilaisuus ISI:n kongressissa Haagissa lokakuussa 2025.

Seura tarjosi edelleen jäsenilleen alennuksia Wileyn, CRC:n ja CUP:n kustantamista tilastotieteellisistä kirjoista. Jäsenet saivat myös tilata alennettuun hintaan ASA:n, RSS:n sekä Australian tilastoseuran SSA:n (Statistical Society of Australia) kustantaman Significance-lehden.

Muuta vuoden varrelta

Tilastotieteen historia Suomessa -kirjaa tekevät kokoontuivat seitsemän kertaa vuonna 2025. Työ jatkuu.

Me Tilastotiedettä

Kuva 3: Pekka Pere ChatGPT:n (Open AI 2026) avustuksella.

Vuosi 2025 oli tilastotoimen riippumattomuuden kannalta huolestuttava monella tapaa. Julkaisimme elokuussa kotisivullamme kannanoton yhdestä huolta herättäneestä tapahtumasta:

Suomen Tilastoseura jakaa Kansainvälisen tilastoinstituutti ISI:n, Ison-Britannian kuninkaallisen tilastoseuran RSS:n sekä Yhdysvaltain tilastoseuran ASA:n toimitusjohtajan huolen Yhdysvaltain työvoimatilastotoimiston (Bureau of Labor Statistics) johtajan Erika McEntarferin erottamisesta. Niiden tapaan Suomen Tilastoseura kannustaa tilastotoimen tilastotieteilijöitä ja tilastoasiantuntijoita edelleen vaalimaan Yhdistyneiden kansakuntien hyväksymiä virallisen tilaston periaatteita ja luottamusta tilastotoimeen. Luotettavat tilastot ovat välttämättömiä demokratian ja talouden toimivuudelle.

Kirjoitan toisaalla Vuosikirjassa lisää vaikeuksista ja vaaroista, joita tilastotoimi ja tilastotieteilijät voivat joutua kohtaamaan.

Seuran varsinainen vuosikokous oli 10.12. Hallitukseen ei tullut uusia jäseniä (<https://tilastoseura.fi/fi/hallitus>). Nykyinen hallitus on erittäin toimiva. Silti ottaisimme mielellään hallitukseen lisää tilastotieteelle sydämensä menettäneitä (Kuva 3), joita kiinnostaa tilastotieteen eteen työskentely vanhassa ja perinteisessä tieteellisessä seurassa. Olkaa mieluusti minuun tai sihteerimme Roope Rihtamoon yhteydessä tässä tarkoituksessa.

Vuosi 2026

Seuraavat Tilastopäivät ovat Kauppakorkeakoulu Hankenilla 28.–29.5.2026. Lähettäkää minulle ehdotuksia esitelmiksi, kiitos!

Päivien yhteydessä Hankenilla on seuran kesäkurssi tällä kertaa monitestauksesta (multiple testing) 23.–27.5.2026. Monitestausteorian tärkeimmät kehittäjät palkittiin toissa vuonna miljoonan dollarin arvoisella joka toinen vuosi jaettavalla Rousseeuw-palkinnolla (<https://www.rousseeuwprize.org/2024>). Tilastotieteen alalla palkinto vertautuu Nobel-palkintoon https://en.wikipedia.org/wiki/Rousseeuw_Prize_for_Statistics). Kurssille kannattaa tulla.

Lähettäkää ihmeessä ehdotuksia myös iltapäiväseminaareiksi. Minä ja sihteerimme olemme jälleen tahot, joihin olla yhteydessä. Kollega kirjoitti vastikään minulle, että Tilastoseura tarjoaa kenties paremmin kuin mikään muu taho Suomessa neutraalin foorumin keskusteluun tutkimusmenetelmistä ja -asetelmista ajankohtaisista teemoista. Niin on!

Kannustan taas jäseniämme ja muitakin kirjoittamaan Vuosikirjaan. Kirjoitus voi olla tavanomaista tieteellistä artikkelia yleistajuisempi, kunhan se on kirjoitettu tilastotieteilijät, tilastotiedettä käyttävät tai tilastotieteestä muuten kiinnostuneet mielessä. Vertaisarvioitavaksi tarjottavat käsikirjoitukset ovat niin ikään erittäin tervetulleita ja eritoten suomeksi tai ruotsiksi. Lista tahoista, joissa voi julkaista englannin kielellä tilastotiedettä, on pitkä. Jäsenillemme tullee ensimmäisenä mieleen Scandinavian Journal of Statistics, jota julkaisemme yhdessä muiden pohjoismaisten tilastoseurojen kanssa. Suomen- tai ruotsinkieli-

sille tilastotieteellisille artikkeleille on sen sijaan paljon vaikeampi löytää julkaisutahoa. En liioittele, kun kirjoitan, että Vuosikirja on ainutlaatuinen foorumi suomen- tai ruotsinkielisille tilastotieteellisille kirjoituksille. Aivan uudenlaisiakin kirjoituksia voi Vuosikirjaan tarjota: katsausartikkeli, kirja-arvostelu, muistokirjoitus, kirje toimitukselle tai tilastotieteellinen mielipide. Siinä muutama idea, ja muunkinlaista kirjoitusta voi tarjota.

Viitteet

Alho J., E. Arjas, J. Karvanen, L. Leskelä, E. Läärä ja P. Pere (2025). *Tilastotieteen sanasto*. Suomen Tilastoseuran julkaisuja no. 8. 3. laitos. Suomen Tilastoseura. Helsinki. <https://sanasto.tilastoseura.fi> (haettu 14.1.2026).

OpenAI (2026). ChatGPT (GPT-5, maaliskuu 2026) [Large language model]. chatgpt.com.

Polytekniska institutet i Finland (1901) Polytekniska institutet i Finland – Program för läsåret 1901–1902. Helsingfors centraltryckeri och bokbinderi. Helsinki. <https://urn.fi/URN:NBN:fi:aalto-201905133101> (haettu 7.5.2025).

Loppuviitteet

¹Ugarten oppikirja Probability and Statistics with R (<https://www.routledge.com/Probability-and-Statistics-with-R/Ugarte-Militino-Arnholt/p/book/9781466504394>) saattaa olla tuttu jäsenille.

Sihteerin ja rahastonhoitajan palsta

ROOPE RIHTAMO
SIHTEERI

ARVI TOLVANEN
RAHASTONHOITAJA

Tilastoseuran sihteerin ja rahastonhoitajan näkökulmasta seuran toimivuodesta 2025 muodostui lopulta varsin eheä kokonaisuus. Erityistä oli se, että edeltävän vuoden sääntöuudistuksen vuoksi jäsenkokouksia järjestettiin nyt kaksin kappalein: yksi keväällä ja yksi syksyllä. Molemmista selvittiin lopulta kunnialla ja uusien sääntöjen mukaisesti, joka lieenee odotusarvoista kun kokousten järjestämisestä vastasi kyseisten sääntöjen laatijat.

Allekirjoittaneet puuhasivat lisäksi taustatukena Tilastopäivillä ja kesän jatkokoulutuskurssilla, joskaan kumpikaan ei osallistunut jälkimmäiseen. Joka tapauksessa voimme olla tyytyväisiä, että seuran hallituksen päätöksestä ja toimesta niin Tilastopäiviä kuin jatkokoulutuskurssia on nyt onnistuneesti järjestetty vuosittain eikä tuleva vuosi ole poikkeus. Vastaavasti myös Vuosikirjan julkaisuväli on onnistuneesti tiivistetty yhteen vuoteen.

Pienemmistä tapahtumista mainittakoon myös Tieteiden yö, jossa allekirjoittaneet olivat menneellä toimikaudella toistamiseen esittelemässä seuran toimintaa. Tieteellisten seurain valtuuskunnan järjestämän ja koordinoiman tapahtuman ohjelmasta vastaavat suomalaiset tieteelliset seurat ja tapahtumatarjonta levittäytyy Helsingin kaupungin alueelle.

Tilastoseuran pisteellä allekirjoittaneet esittelivät seuran toimintaa tilastolisten aktiviteettien, kuten Galtonin koneen esittelyn sekä otantateorian simulointiesimerkin avulla. Jälkimmäinen löytyy verkosta hakemalla “Rock ‘n Poll” ja se demonstroii normaalivaihtelua kyselytutkimuksessa vaalikyselyteemalla. Yliopistotutkija JESSICA NISÄNin (FLUX-hanke, Turun yliopisto) luennoi tapahtumassa Tilastoseuran pyynnöstä otsikolla “Mistä tulet, minne menet, syntyvyys?”. Luento käsitteli syntyvyyden kehitystä Suomessa.

Sihteerin ja rahastonhoitajan kannustavat jäsenistöä olemaan matalalla kynnyksellä yhteydessä seuran hallitukseen. Yhteyttä voi ottaa esimerkiksi palautteenannon, kehitysehdotusten tai vaikkapa Iltapäiväseminaarien teemailtojen aihe-ehdotusten merkeissä. On kunnia saada olla osana yli satavuotiaan tilastotieteellisen seuran toimintaa ja toivomme ja kutsumme erityisesti nuorempia tieteen- ja ammatinharjoittajia mukaan toimintaan runsain joukoin!

Tilastotieteen rooli ilmastonmuutoksen arvioinnissa.

Suomen Tilastoseuran iltapäiväseminaari 9.9.2025

13.00–13.10	Pekka Pere puheenjohtaja Suomen Tilastoseura Juha Alho emeritusprofessori, HY	Tilaisuuden avaus: Mikä on tilastotieteen rooli?
13.10–13.40	Pasi Holm johtava asiantuntija, FCG Finnish Consulting Group	Metsätalouden taloudelliset vaikutukset
13.40–14.10	Sini Niinistö yliaktuaari, Tilastokeskus	Kasvihuonekaasuinventaario: periaatteet, ohjeistus ja politiikkatoimien vaikutus
14.10–14.40	Kahvitauko	
14.40–15.10	Juha Heikkinen tutkimusprofessori, LUKE	Suomen metsämaiden hiilitaseen kehitys 1990–2023 ja sen arvioinnin epävarmuustekijät
15.10–15.30	Elja Arjas emeritusprofessori, HY	Kommenttipuheenvuoro
15.30–16.00	Yleisökeskustelu	



Kuva 1: Juha Alho (vas.) ja Juha Heikkinen.



Kuva 2: Sini Niinistö.

Avauspuheenvuoro

JUHA ALHO

EMERITUSPROFESSORI, HELSINGIN YLIOPISTO
SENIOR FELLOW, ETLA

Tilastotieteen näkökulma

Ihmisen tuottamien (antropogeenisten) kasvihuonekaasujen osuus maapallon ilmaston lämpötilan muutoksessa kuuluu aerosolifysiikan ja sähkömagnetismin tutkimusalaloihin. Kokeellisesti voidaan yleensä tutkia vain osaprosesseja, joiden yhteisvaikutuksia pyritään selvittämään ilmakehästä tehdyillä mittauksilla.

Niille ovat tunnusomaisia mittavirheet ja se, että kaikkia vaihtelun syitä ei aina pystytä identifioimaan. Joudutaan tekemään oletuksia, joiden pätevyyttä arvioidaan vertaamalla mallilaskelmia havaintoihin. Vaikka ilmiön tieteellinen ymmärrys perustuu fysiikan lakeihin, päätöksenteossa tarvitaan tilastollista mallinrakennusta, estimointia ja eri tekijöiden suhteellisen merkityksen punnintaa.

Koska kiinnostuksen kohteena ovat nimenomaan antropogeeniset tekijät, joudutaan tarkastelemaan myös niitä teollisia, taloudellisia ja yhteiskunnallisia tekijöitä joiden tuloksena ihmisen vaikutus ilmastoon toteutuu. Joudutaan kysymään, millaisiin toimenpiteisiin kannattaa ryhtyä, kun eri vaihtoehtoihin välttämättä liittyy epävarmuuksia. Koska ilmastotoimien globaalit kustannukset ja hyödyt eivät jakaudu tasaisesti eri alueille ja eri kulttuuripiireihin, päätöksentekoon liittyy intressistiriitoja.

Seuraavassa tarkastellaan iltapäivätilaisuuden taustaksi esimerkkejä, joissa tilastotieteellä on näkyvä rooli.

Ilmastotoimien kokonaistaloudellinen hyöty

Nicholas Sternin lähes 700-sivuinen raportti antropogeenisen ilmastomuutoksen taloudellisista vaikutuksista oli yksi varhaisimmista, joissa pyrittiin luomaan kokonaiskuva aiheesta. Poimin oheen raportin lähtökohdan.

Reippaasti yksinkertaistaen raportissa verrataan nykytilaa ($i = 0$) vaihtoehtoon ($i = 1$). Niitä kuvataan reaaliarvoisilla funktioilla $u_i(t) > 0, i = 0, 1$, jotka kuvaavat kaikkien ihmisten hyvinvointien summaa hetkellä $t \geq 0$. Hyvinvointi voidaan tulkita esimerkiksi hyödykkeiden ja palveluiden kulutukseksi, mutta se voi kuvata myös esimerkiksi terveydentilaa. Eriaikaiset hyvinvoinnin tasot saatetaan yhteismitalliseksi diskonttaamalla nykyhetken $t = 0$: $W_i = \int_0^{+\infty} u_i(t)e^{-\delta t} dt, i = 0, 1$, missä $\delta \geq 0$ on korko. Jos $W_1 > W_0$, vaihtoehtoinen tila kannattaa valita.

Antropogeenisen ilmastomuutoksen kohdalla ajatus on, että lähivuosina uhrattaisiin hyvinvointia ($u_1(t) < u_0(t)$, kun t on pieni), ja sen seurauksena saatettaisiin korkeampi hyvinvointi myöhemmin ($u_1(t) > u_0(t)$, kun t on suuri).¹

Koska kyse on ihmiskunnan nykyisen ja tulevan hyvinvoinnin vertailusta, siihen ei liity samanlaista aikapreferenssiä, joka tulee kyseeseen, kun henkilö lainaa toiselta rahaa, ja maksaa siitä hyvästä korkoa. Tuleva hyvinvointi on yhtä tärkeää kuin nykyinenkin. Tästä näkökulmasta voisi ajatella, että $\delta = 0$. Stern esittää kuitenkin kaksi perustetta diskonttorolle.

Näistä numeerisesti vähämerkityksinen, mutta tilastotieteilijän näkövinkkeistä kiintoisa, liittyy ihmiskunnan elinaikaan. Koska ihmiskunta ei ole olemassa ikuisesti, on olemassa jotkin todennäköisyydet elossaoloon tulevina vuosina t . Stern haarukoi mahdollisuuksia taulukoiden muodossa, ja päätty – tilastotieteen kielelle käännettynä – oletamaan, että ihmiskunnan elinajalla on eksponenttijakauma, jonka vakiohasardi on $\mu = 0.001$. Tällöin todennäköisyys, että olemme vielä olemassa t :n vuoden päästä olisi $e^{-0.001t}$. Toisaalta oletuksesta seuraa, että elinajan odotusarvo on $1/\mu = 1000$ vuotta. Selvästikään valitulla hasardin arvolla ei ole kovin tukevia perusteita.

Tärkeämpi peruste korolle liittyy siihen, että teknologian kehittymisen ja tuotannon tehostumisen seurauksena tulevaisuudessa voidaan tuottaa nykyisiä vastaavat hyödykkeet ja palvelut halvemmalla. Tämä lisää kulutusmahdollisuuksia. Kun siis tulemme olemaan rikkaampia, hyvinvointiero voidaan tasata diskonttaamalla. Simon Dietzin mukaan Sternin arvio on osapuilleen 0.013, joka

vastaa 1.3 %:n kulutuksen kasvua vuodessa, ilman ilmastomuutosta. Yhteensä kahdesta lähteestä saadaan koroksi $0.1 + 1.3 = 1.4\%$.

Sternin raportissa korostetaan korkoarvioiden approksimatiivisuutta. Esiin tuodaan esiin suuri joukko tekijöitä, jotka täytyy käytännössä ottaa huomioon. Näitä ovat väestön koon muutokset, kotitalouksien erilaiset kulutustavat, kulutuksen kasvun aikavaihtelut, pääomamarkkinoiden toiminnan epäsäännöllisyydet ja sellainenkin matemaattinen yksityiskohta kuin millä ehdoilla edellä määritellyt integraalit W_i ovat äärellisinä olemassa.

Kun Sternin raportti julkaistiin runsas 20 vuotta sitten, monet kriitikot kiinnittivät huomiota siihen, että korkoarvot ovat selvästi markkinakorkoja alempia. Ymmärtäisin, että näin on pakko ollakin, ovathan ne “preskriptiivisiä” sosiaalisia korkoja, eivät “deskriptiivisiä”. Ilmastotoimenpiteiden kriitikoille aihe tarjosi kuitenkin vettä myllyyn.

Tarkoitushakuinen kritiikki

Ensi toiminaan presidentiksi 2024 valittu D. Trump irrotti Yhdysvallat kansainvälisistä ilmastositoumuksistaan. Tätä seurasivat ylimpien liittovaltion virkojen täyttö Trumpin kannattajilla. Energia-asioista vastaavaksi nimettiin C. Wright, jonka tausta oli mm. uusissa öljynporausmenetelmissä (“fracking”). Hän nimitti viiden tutkijan ryhmän tarkastelemaan kriittisesti ilmastopolitiikan perusteita. Tekijät ovat kokeneita tutkijoita (kaikilla Ph.D.-tutkinto, ura yliopistoissa, tutkimuslaitoksissa, hallinnossa jne.), mutta myös tunnettuja ilmastotoimien kriitikoita.

Raportti julkaistiin heinäkuun lopulla 2025 (Climate Working Group, 2025). Raportin aiheet korostavat ilmastotutkimuksen epävarmuuslähteitä sekä erityisesti hallitustenvälisen ilmastomuutospaneelin IPCC:n raporteissa tehtyjä oletuksia. Seuraavaan on poimittu lähes 150-sivuisesta raportista kaksi esimerkkiä regressioanalyysin alalta.

(1) Ilmastoherkkyys (equilibrium climate sensitivity) tarkoittaa maapallon ilmakehän lämpötilan muutosta, joka seuraa hiilidioksidipitoisuuden kaksinkertaistumisesta sen esiteollisesta arvosta 280 ppm. Parametri on keskeinen, sillä pääosa antropogeenisestä ilmastomuutoksesta ilmaistaan muuntamalla vaikutukset ekvivalentin hiilidioksidin lisäyksen aiheuttamaan muutokseen. IPCC:n tuoreimman arvion mukaan ilmastoherkkyiden arvoa ei tarkkaan tunneta, mutta se on todennäköisyydellä 2/3 välillä (2.5, 4.0) Celsius-astetta. Aiemmin ajateltiin, että todennäköisyysväli olisi (1.5, 4.5).

Muutos näyttäisi perustuvan tuoreeseen yhteenvedoon, jossa bayesläistä laskentaa käyttäen on estimoitu ilmastoherkkyydelle tiheysfunktio. Menettely perustuu ilmaston fysikaaliseen rakennemalliin yhdessä muinaisilmastoa sekä ilmaston vaihtelua ja sitä selittäviä tekijöitä (merien lämpötila, auringon toiminta, tulivuorten aktiivisuus, ilman aerosolit jne.) kuvaaviin aikasarjoihin. Asian tuntijaharkintaa on tarvittu tarkastelussa olevien sarjojen mallinnuksessa ja mallien parametrien priorien muotoilussa. Keskeisenä tavoitteena on ollut luoda kaikki tietolähteet huomioonottava kuvaus ja sen kanssa sopusoinnussa oleva tiheysfunktio.

Raportissa Climate Working Group (2025) arvostellaan yhteenvedoa mm. informatiivisten priorijakaumien käytöstä ja viitataan lähteisiin, joissa osoitetaan, että käyttämällä vähemmän informatiivisia (Jeffreys-tyyppisiä) prioreja saataisiin jonkin verran poikkeavia tuloksia. Näin tietysti onkin. Ymmärtääkseni prio-

rien rooli oli kuitenkin kvantifioida prosesseja koskeva asiantuntijamielipide, jota pidettiin tärkeänä mm. aineistojen huonon laadun takia. Tämä hyöty esitetyssä vaihtoehtoisessa analyysitavassa menetetään.

(2) Sormenjälkianalyysi (fingerprinting) tarkoittaa havaittujen ilmastomuutosten antropogeenisten syytekijöiden identifiointia ja estimointia erotukseksi luonnon vaihtelusta. Tilastollisena menetelmänä käytetään ns. pienimmän totaalineliösumman menetelmää (Total Least Squares, TLS). Menetelmässä etsitään regressiosuoraa, johon havaintopisteiden kohtisuorat etäisyydet minimoituvat. Tämä on motivoitua, kun selittäjissä on mittausvirhettä. Menetelmällä näyttää olevan käyttäjäkunta kemometrian alalla, vaikkei sitä tilastotieteilijöiden keskuudessa juuri tunneta.

Eräs raportin Climate Working Group (2025) kirjoittajista on erityisesti kirjoittanut siitä, että TLS:n tulokset ovat epästabiileja ilmastomuutokseen liittyvässä sormenjälkianalyysissä, ja ehdottaa tilalle vaihtoehtoja. Mittavirheiden käsittely on tunnetusti hankalaa, kun sellaisia virheettömiä mittauksia ei ole, joita vastaan virheiden suuruutta ja jakautumista voitaisiin arvioida. Oletuksilta on vaikea välttyä, erityisesti ilmastosovelluksissa, joissa selittäjät itse ovat mallitusten tuloksia.

Sen enempää ilmastoherkkyyttä kuin sormenjälkianalyysiäkään koskevat huomautukset eivät ole yllättäviä eivätkä vääriä. Sen sijaan niiden pohjalta vedettyjä johtopäätöksiä on vaikea ymmärtää. Koko raportin tärkein johtopäätös on päätösluvun yhteenvedossa, joka on lyhykäisyydessään seuraava: “U.S. policy actions are expected to have undetectably small direct impacts on the global climate and any effects will emerge only with long delays.”

On vaikea nähdä, miten sinänsä keskustelunarvoisista teknisistä tarkasteluista saadaan tulema, jolla on vain vähän tekemistä perusteluiden kanssa. Ovathan esimerkiksi kohdissa (1) ja (2) esitetyt huomautukset merkitykseltään pieniä. Vaikutelmaksi jää, että kirjoittajilla on jo alussa ollut tiedossaan loppupäätelmät, ja on yritetty löytää niille perustelut. – On lisäksi häkellyttävää, että raportin kirjoittajat onnistuvat näkemään USA:n niin pienenä tekijänä ilmastopolitiikassa, ettei sen teoilla voi olla merkitystä!

Raporttia on jo käytetty USA:n ilmastopolitiikasta luopumisen perusteluna.² On odotettavissa, että sitä tullaan kaupittelemaan myös Eurooppaan ja kenties Suomeenkin samassa tarkoituksessa.

Epävarmuus pitäisi ottaa aidosti huomioon

IPCC on YK:n mandaatilla toimiva tieteellinen elin, joka perustettiin vuonna 1988. Tuorein arviointiraportti IPCC (2023) on huolella editoitu, 115 sivua pitkä. Siinä tunnistetaan empiirisiin arvioihin liittyvä epävarmuus käyttämällä probabilistista kieltä. Päätelmiä kuvataan esimerkiksi termein “virtually certain”, “very likely”, “likely”, “more likely than not” jne. Näille annetaan myös ohjeelliset todennäköisyysarvot. Raportin lukijan on kuitenkin vaikea saada selville, miten arviot on täsmälleen ottaen saatu.

Skenaarion käsite on raportissa keskeinen. Pääsääntöisesti kyseessä ovat deterministiset mallilaskelmat, joiden tulos määräytyy suoraan lähtöoletuksista. “These are quantitative projections and are neither predictions nor forecasts”, kuvataan raportissa. Jo 1990-luvulla on kuitenkin havaittu, että esimerkiksi

väestönkehitystä kuvaavilla skenaarioilla ei voida kuvata tulevan väestönkehityksen epävarmuutta loogisesti oikein. Ilmastomuutosta koskevat laskelmat ovat usein tyyppiä “ $I = PAT$ ”, eli $\text{impakti} = \text{populaatio} \times \text{afluenssi} \times \text{teknologia}$. Nähdään, että väestö yksinään edellyttäisi erilaista lähestymistapaa kuin käytetyt skenaariot.

Ilmastovaikutusten arviointi on monimutkaisempaa kuin väestönkehityksen arviointi. Tästä on esimerkkinä se, että fysikaaliset skenaariot tuotetaan usein perättäisten mallilaskelmin tuloksena, joissa yhden mallin tulos on seuraavan syöte. Mutta kun otetaan huomioon ilmiön globaali tärkeys, tuntuu oudolta, ettei tilastotieteen koneistoa käytetä lähellekään siinä mitassa kuin olisi mahdollista. Vaikutelmaksi jää, että vain osalla mukana olevista tutkijoista on taustaa tilastotieteessä.

Jälkikirjoitus

Iltapäivätalaisuudessa keskityttiin Suomen metsien hyödyntämiseen liittyviin kysymyksiin. Suomen kannalta olisi tärkeää pohtia, onko meillä ilmastolakiin otettu hiilineutraaliustavoite ylipäättään saavutettavissa.

Maapallon pinta-alasta noin 30 % on maata. Tästä vajaa puolet on asuttavaa. Väestön tiheys vaihtelee maasta toiseen tavattoman paljon, samoin maiden sisällä. Esimerkiksi Savukosken asukastiheys on Suomen pienin: asukkaita on 1 000, pinta-ala on 6 500 km². Helsingin Kallion kaupunginosan pinta-ala on 1 km², mutta asukkaita on 19 000. Päästökauppa tarjoaa tavan asettaa alueet samanarvoiseen asemaan, mutta tilastotieteilijä jää ihmettelemään millä hinnoilla. Ja kenelle mahdolliset hiilinelut jyvitetään? – On hyvä muistaa, että Suomi on maailman Savukoski!

Ranskalainen matemaatikko ja fyysikko Joseph Fourier oli ensimmäisiä, jotka laskivat planeetoille auringosta tulevan energian määriä ja niistä aiheutuvaa lämpenemistä. Kun iltapäivätalaisuuden teemana on ollut tilastotieteen rooli ilmastomuutoksen kuvaamisessa, sopinee muistuttaa, että kyseessä on samainen Fourier, jonka nimi liitetään aikasarja-analysissä käytettyihin Fourier’n sarjoihin, ja juuri hän kehitti Fourier’n integraalimuunnoksen, jonka tilastotieteilijät tuntevat todennäköisyysjakauman karakteristisena funktiona.

Viitteet

- Andrieu, C., Doucet, A. ja Holenstein, R. (2010). ”Particle Markov chain Monte Carlo methods”. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72.3, s. 269–342. DOI: <https://doi.org/10.1111/j.1467-9868.2009.00736.x>.
- Chopin, N. ja Singh, S. S. (2015). ”On particle Gibbs sampling”. *Bernoulli* 21.3, s. 1855–1883. DOI: <https://doi.org/10.3150/14-BEJ629>.
- Climate Working Group (heinäkuu 2025). *A Critical Review of Impacts of Greenhouse Gas Emissions on the U.S. Climate*. Tekninen raportti. Washington DC: Department of Energy.
- Del Moral, P. (2004). *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Springer. DOI: <https://doi.org/10.1007/978-1-4684-9393-1>.
- Gerber, M., Chopin, N. ja Whiteley, N. (2019). ”Negative association, ordering and convergence of resampling methods”. *The Annals of Statistics* 47.4, s. 2236–2260. DOI: <https://doi.org/10.1214/18-AOS1746>.

- IPCC (2023). *Climate Change 2023. Sixth Assessment Report (AR6). Synthesis Report, longer version*. Intergovernmental Panel on Climate Change.
- Karppinen, S., Singh, S. S. ja Vihola, M. (2024). "Conditional Particle Filters with Bridge Backward Sampling". *Journal of Computational and Graphical Statistics* 33.2, s. 364–378. DOI: [10.1080/10618600.2023.2231514](https://doi.org/10.1080/10618600.2023.2231514).
- Karppinen, S. (lokakuu 2025). *GenericSSMs.jl: v0.1.1*. Versio v0.1.1. DOI: [10.5281/zenodo.17288723](https://doi.org/10.5281/zenodo.17288723). URL: <https://doi.org/10.5281/zenodo.17288723>.
- Karppinen, S. ja Vihola, M. (lokakuu 2025). *Resamplings.jl: v0.1.2*. Versio v0.1.2. DOI: [10.5281/zenodo.17288764](https://doi.org/10.5281/zenodo.17288764). URL: <https://doi.org/10.5281/zenodo.17288764>.
- Whiteley, N. (2010). "Discussion on "Particle Markov chain Monte Carlo methods"". *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72.3, s. 306–307. DOI: <https://doi.org/10.1111/j.1467-9868.2009.00736.x>.
- Vihola, M., Helske, J. ja Franks, J. (2017). "Importance sampling type estimators based on approximate marginal Markov chain Monte Carlo". *arXiv preprint 1609.02541v3*. DOI: <https://doi.org/10.48550/arXiv.1609.02541>.

Loppuviitteet

¹Argumenttiin kuuluu ajatus, että vaihtoehto on pienehkö muutos nykytilasta siinä mielessä, että u_1 :n arvot saadaan differentiaaliapproksimaationa u_0 :n arvoista. Tämä liittyy hyötyfunktioiden kautta Sternin korkotarkasteluihin, muttei varsinaisesti tilastotieteen rooliin, ja jätetään siksi sivuun.

²"The Trump Administration will not jeopardize our country's economic and national security to pursue vague climate goals". (Valkoisen talon edustaja, The New York Times, 28.8.2025) 30.1.2026 liittovaltion tuomari W. Young totesi raportin tekijöiden valintaprosessin laittomaksi. Perusteena oli, että raportti oli selvästikin tarkoitettu politiikan teon välineeksi, ei informaation levittämiseksi. Tästä huolimatta tuomari ei vaatinut raporttia poistettavaksi EPA:n sivuilta. (NYT 3.2.2026)

Kasvihuonekaasuinventaario: periaatteet, ohjeistus ja politiikkatoimien vaikutus

SINI NIINISTÖ

TILASTOKESKUS

Kansainväliset ilmastosopimukset kasvihuonekaasuinventaarion taustalla

Tarve kasvihuonekaasuinventaarioille eli ihmistoiminnan päästöjen ja poistumien arvioinnille ja raportoinnille on syntynyt yhteisesti sovitusta tavoitteesta vähentää päästöjä. YK:n ilmastosopimus¹ vuodelta 1992 asetti joukolle teollisuusmaita kuten Suomelle velvollisuuden valmistella kasvihuonekaasuinventaario vuosittain. Ilmastosopimuksen osapuolikokouksen päättämässä ohjeissa määritellään raportoitavat kaasut, sektorit, päästöluokat ja hiilivarastot ja aikasarjan aloitusvuosi (1990).

Pariisin sopimuksen solmiminen vuonna 2015 oli merkittävä saavutus. Siinä sovittiin yhtäläisistä velvollisuuksista kaikille osapuolille asettaa kansallisesti määritellyn panoksensa eli ilmastotavoitteensa ja raportoida edistymisestä sen suhteen kahden vuoden välein. Tämä raportointi sisältää myös kasvihuonekaasuinventaarion laajoine raportointitaulukkosetteineen. EU:lla ja sen jäsenmailla on yhteinen päästövähennystavoite Pariisin sopimuksen alla.

Kasvihuonekaasuinventaariotietoja käytetään kunkin osapuolen tavoitteen seurannassa mutta myös Pariisin sopimuksen maailmanlaajuisessa ilmastomuutoksen hillintätavoitteen mukaisessa seurannassa (global stock take). Suomen ensimmäinen kaksivuotisraportti valmistui joulukuussa 2024 ja tarkastettiin kansainvälisen asiantuntijaryhmän toimesta vuonna 2025. YK:n ilmastosopimusten lisäksi myös EU:n lainsäädäntö edellyttää vuosittaista kasvihuonekaasuinventaariota, jonka perusteella seurataan EU:n päästövähennystavoitteita ja jäsenmaakohtaisia velvoitteita.

Kansallinen inventaariojärjestelmä

YK:n ilmastosopimuksen ja Pariisin sopimuksen kansallisena vastuutahona toimii Suomessa ympäristöministeriö ja nimettynä kansallisena kontaktina eli focal pointina Marjo Nummelin. Tilastokeskus toimii kasvihuonekaasuinventaarion kansallisena vastuuyksikkönä ilmastolain mukaisesti. Laskenta, raportointi ja kehittäminen tehdään yhteistyössä Luonnonvarakeskuksen ja Suomen ympäristökeskuksen kanssa sektorikohtaisten vastuiden ja sopimusten perusteella. Tilastokeskus julkistaa inventaariotiedot, vastaa inventaarioraportoinnin lähetyksistä YK:lle ja EU:lle ja koordinoi YK:n ilmastosopimusten ja EU:n suorittamat inventaariotarkastukset. Ministeriöiden, asiantuntijalaitosten ja Tilastokeskuksen edustajista koottu neuvottelukunta tukee inventaariotyötä.

Kasvihuonekaasuinventaarion tietoja tarvitaan kansallisesti mm. hallituksen eduskunnalle vuosittain antamaan ilmastovuosikertomukseen, kansallisen il-

mastolain ja EU-tason päästövähennystavoitteiden seurantaan sekä politiikka-toimien suunnitteluun ja niiden vaikutuksen seurantaan.

Inventaarion säännöt, ohjeistukset ja vaatimukset

Ensimmäisenä sääntöjen, ohjeistusten ja vaatimusten hierarkiassa tulevat ilmastositimusten osapuolikokousten päätöksissään hyväksymät inventaarioreportointiohjeet ja -taulukot, jotka ovat samat YK:n ilmastositimukselle ja Pariisin sopimukselle. IPCC:n eli hallitustenvälisen ilmastopaneelin hyväksymät menetelmäohjeet maakohtaisille kasvihuonekaasuinventaarioille antavat ohjeita, kuinka laskea päästöt ja poistumat ja oletusmenetelmän suurimmalle osalle päästöluokista. Oletusmenetelmien sijaan mailla on mahdollisuus, ja inventaarioreportointilukien osalta velvollisuus, kehittää olosuhteisiinsa sopivia kansallisia päästökertoimia, seurantoja tai laskentamalleja. Mailla on velvollisuus raportoida päästöt kaikille luokille, joille IPCC:n ohjeista löytyy oletusmenetelmä ja joille mailla on tarvittavat lähtötiedot.

Keskeisiä IPCC:n ohjeiden periaatteita kasvihuonekaasuinventaarioille ovat päästöjen raportointi kansallisten rajojen määräämältä alueelta, keskittyminen ihmisen toiminnasta aiheutuviin päästöihin ja poistumiin, kansallisen inventaarioreportointijärjestelmän luominen ja ylläpito, päästöjen allokointi sektoreille, päästö- tai maankäyttöluokkiin tai hiilivarastoihin sekä tietylle vuodelle. Lisäksi päästöjen ja poistumien arviointi tehdään vuosittain, ja aikasarjan yhteneväisyys eli eri vuosien välinen vertailukelpoisuus säilytetään laskemalla päästöt uudelleen kaikille vuosille, joita mahdollinen muutos lähtötiedoissa ja laskentamenetelmässä uusimmassa inventaariossa koskettaa. Jatkuva kehittäminen on myös yksi periaate.

Myös Suomen kasvihuonekaasuinventaariota kehitetään jatkuvasti laadun parantamiseksi ja uusiin tietotarpeisiin vastaamiseksi. Lisäksi Suomen inventaariota koskevien kansainvälisten tarkastusten suositukset ja uusien inventaario-ohjeiden ja tutkimustulosten valmistuminen luovat tarpeen kehittää laskentaa. Myös uudet tiedot kuten esimerkiksi puustomittaustiedot valtakunnan metsien inventoinnista aiheuttavat inventaarioreportointilukien lähimenneisyyden vuosien päästöjen ja poistumien uudelleen laskennan. Muuttuneet käytännöt ja uudet politiikkatoimet lisäävät osaltaan tarvetta kehittää inventaariota niin, että niiden vaikutus saadaan näkyviin mahdollisuuksien mukaan. Osalle toimista on haasteellista kehittää laskentamenetelmiä koko maan kattavaan, erilaisia olosuhteita kuvaavaan inventaarioon. Tästä esimerkkinä voi mainita erilaiset turvepohjaisille maatalousmaille suunnitellut ilmastotoimet.

IPCC:n arviointiraportit *versus* kasvihuonekaasuinventaario

Kansainvälisen tutkijajoukon säännöllisin välein valmistelemaat IPCC:n arviointiraportit² ovat paras mahdollinen luonnontieteellinen näkemys ilmakehän kasvihuonekaasupitoisuuksien ja ilmastomuutoksen tähänastisesta ja tulevasta kehityksestä. Tulevalle kehitykselle esitetään arviointiraporteissa monta eri skenaarioria, joille monilukuiset mallit antavat arvionsa vaihteluväleineen.

Kasvihuonekaasuinventaario on puolestaan käytännöllinen, päästövähennystavoitteiden seurantaan tarkoitettu järjestelmä: Se on oltava kaikkien maiden tehtävissä, perustuen parhaaseen mahdolliseen tietämykseen, mutta myös käy-

tännön toteutettavuus vuosittaisena, jatkuvana tehtävänä on huomioitava inventaariomenetelmäohjeissa ja toteutuksen suunnittelussa.

IPCC:n arviointiraporteissa huomioidaan myös valtameret ja ihmisen toiminnan ulkopuolella olevat alueet ja ilmiöt kuten jäätikköalueet ja tulivuoren purkaukset, kun taas kasvihuonekaasuinventaario keskittyy ihmisen toiminnan aiheuttamiin päästöihin ja poistumiin ja ihmistoiminnan alaisten alueiden³ päästöihin ja poistumiin.

IPCC:n arviointiraporteissa on viime vuosina tuotu esiin raportin tulosten ja johtopäätösten todennäköisyyttä tai epävarmuutta: Luokituksena on käytetty esimerkiksi High confidence/medium confidence – very likely/extremely likely-limited evidence/ -luokittelua. Vastaavasti IPCC:n menetelmäohjeet kasvihuonekaasuinventaariolle tuovat nykyään aiempaa selkeämmin esiin esitettyjen parametrien tai päästökertoimien epävarmuuden esimerkiksi kertomalla luottamusvälit, käytettyjen havaintojen lukumäärän ja käytetyn tutkimuskirjallisuuden. IPCC:n inventaariomenetelmäohjeet ovat kansainvälisen tutkija- ja inventaarioasiantuntijajoukon laatimat ja vastaavasti kuin IPCC:n arviointiraportit vertaisarviodut avoimilla, maailmanlaajuisilla kommentointikierroksilla. IPCC kehittää inventaariomenetelmäohjeita joko päivittämällä aika ajoin aikaisempia ohjeita tai laatimalla uusia, täydentäviä ohjeistuksia.

Loppuviitteet

¹United Nations Framework Convention on Climate Change (UNFCCC)

²Assessment Reports

³nk. IPCC:n kasvihuonekaasuinventaariomenetelmäohjeiden ”managed land proxy”

Suomen kivennäismaametsien hiilitaseen epävarmuuden arviointi

JUHA HEIKKINEN

JUHA.HEIKKINEN@LUKE.FI

LUONNONVARAKESKUS (LUKE)

Tiivistelmä

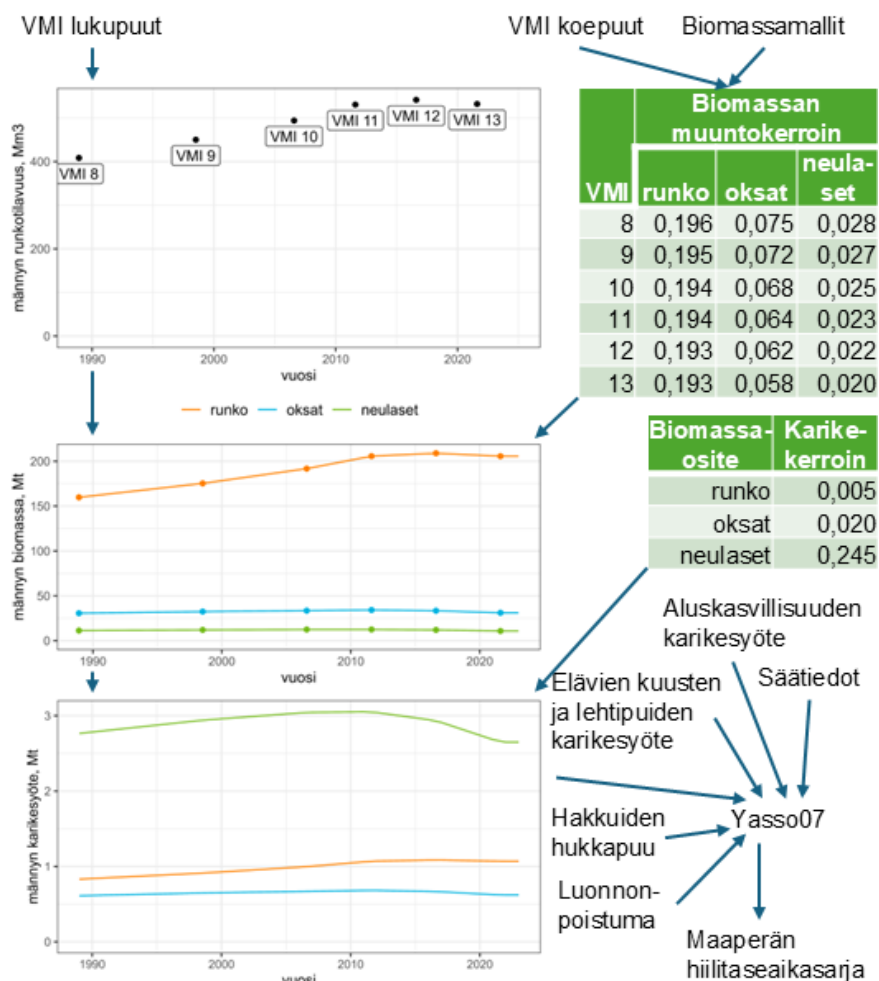
Suomen kasvihuonekaasuinventaarion mukaan kivennäismaametsien maaperät olivat 1990-luvulla merkittävä ja kasvava hiilinielu. Tämän vuosituhannen aikana nielu on kuitenkin vähentynyt jyrkästi ja viimeisimpien vuosien arviot viittaavat siihen, että kivennäismaaperät ovat voineet kääntyä jo hiilidioksidin nettopäästöksi. Tähän saakka käytetty hiilitaseen epävarmuuden arviointitapa ei ole enää perusteltu, kun tasearvio on lähellä nollaa. Tässä kirjoituksessa esitellään tähän uuteen tilanteeseen päivitetty epävarmuusanalyysi. Tärkeimmät analyysiin sisältyvät epävarmuuden lähteet ovat puuston biomassamallien ja inventaariossa käytettävän maahiilimallin parametristimaatteihin liittyvä epävarmuus, malliennusteisiin liittyvä jännösvaihtelu ja tasearvioiden tuottamisessa käytettävien aineistojen otannasta johtuva epävarmuus.

Johdanto

Suomen metsät ovat olleet vakaa ja merkittävä hiilinielu noin vuoteen 2012 saakka, jonka jälkeen nielu on nopeasti pienentynyt (esim. Luke, 2025b). Vuoden 2023 inventaarion mukaan metsämaa on ollut kasvihuonekaasujen päästölähde vuodesta 2021 lähtien (Luke, 2025b, kuva 2). Metsämaan kasvihuonekaasutaseiden arviot ovat kuitenkin varsin epävarmoja. Luken Luonnonvaratietosivuston kirjoituksessa (Luke, 2025c) käsitellään yleisellä tasolla kaikkien metsämaan päästöluokkien epävarmuuksia. Tämän kirjoituksen tavoitteena on esitellä epävarmuuksien arvioinnin perusteita hiukan syvällisemmin.

Esittelyssä keskitytään esimerkinomaisesti metsämaiden kivennäismaaperän hiilitaseeseen. Metsämaan kasvihuonekaasutaseeseen sisältyvät näiden lisäksi elävä puusto ja orgaaninen maaperä, joista jälkimmäisen osalta raportoidaan hiilivaraston muutosten lisäksi myös metaani- ja typpioksiduulipäästöt. Elävän puuston hiilitaseen arviointimenetelmä on huomattavasti suoravii- vaisempi kuin maaperätaseiden arviointimenetelmät ja sen epävarmuuden arviointiin käytetään samankaltaista menetelmää kuin tähän kirjoitukseen sisältyvään puustobiomassa-aikasarjojen epävarmuuden arviointiin (ks. esim. Ståhl ym., 2014). Orgaanisten maiden hiilitaseeseen liittyvä epävarmuusarviointi kuvataan seikkaperäisesti julkaisun Alm ym. (2023) liitteessä A ja muiden kaasujen osalta käytetään hallitustenvälisen ilmastomuutospaneelin ohjeiden (IPCC, 2006) mukaista oletusmenetelmää. Kattava kuvaus niin metsämaan kuin muidenkin raportointiluokkien ja sektoreiden inventaariomenetelmistä löytyy kansallisesta inventaarioraportista (Forsell ym., 2025).

Suomen kasvihuonekaasuinventaariossa (KHKI) metsien ”maaperän” hiilivarastoon sisältyvät kuolleet puut, elävästä puustosta ja aluskasvillisuudesta tuleva karike sekä sen hajoamistuotteena syntynyt maaperän orgaaninen aines. Hiilitase, eli hiilivaraston vuosittainen muutos, on varastoa kasvattavan karike-



Kuva 1: Kivennäismaametsien maaperän hiilitaseen arviointi KHKI:ssa, esimerkkinä Etelä-Suomen laskenta; numeeriset havainnollistukset elävien männytien maanpäällisiä biomassaositteita koskevilla tuloksilla. Runkotilavuusarvot perustuvat VMI:n lukupuihin, biomassan muuntokertoimet VMI:n koepuihin ja biomassamalleihin. Runkotilavuusarvioista päästään muuntokertoimilla elävän puuston biomassa-arvioihin joista elävän puuston karikesyöte johdetaan karikekertoimien avulla. Yasso07-mallin syötteeksi tarvitaan lisäksi hakkuista ja luonnonpoistumasta syntyvä karikesyöte ja alueittaiset säätiedot.

syötteen hiilen määrän ja varastoa pienentävän orgaanisen aineksen hajotuksen seurauksena ilmakehään vapautuvan hiilidioksidin hiilen määrän erotus. ”Karikesyöte” muodostuu elävän puuston ja aluskasvillisuuden karikkeesta, hakkuissa korjaamatta jääneestä ”hukkapuusta” ja ”luonnonpoistumasta” eli vuoden aikana kuolleista puista.

Kivennäismaaperän hiilitasetta arvioidaan KHKI:ssa erikseen Etelä- ja Pohjois-Suomen laskenta-alueilla Yasso07-maahiilimallilla (Tuomi ym., 2011, ks. Kuva 1). Mallin syötteeksi tarvitaan vuosittaiset karikesyötteen määrät laskenta-alueittain. Karikesyöte on lisäksi jaettava aineksen karkeuden ja

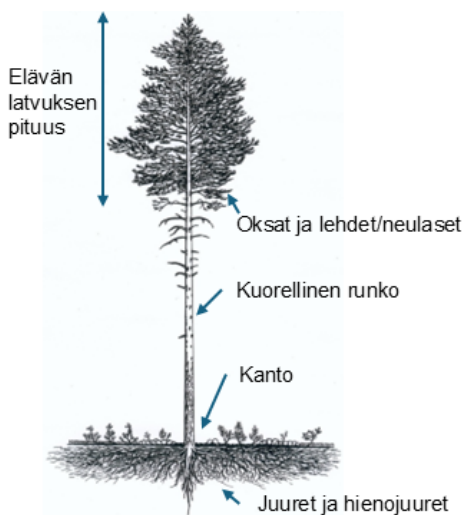
liukoisuuden mukaisiin osiin, mutta se ei ole tämän kirjoituksen kannalta olennaista. Karikkeen lisäksi Yasso07:n syötteenä annetaan alueittaisia säätietoja (orgaanisen aineksen hajoamisnopeus riippuu vahvasti lämpötilasta ja kosteudesta), mutta nämä ovat varsin tarkkoja, joten niitäkään ei tässä kirjoituksessa käsitellä tämän enempää.

Elävästä puustosta peräisin olevan karikkeen vuosittaisia määriä arvioidaan puusto-ositteiden biomassojen ja niiden karike(tuotanto)kerrointen tuloina. Vuosittaiset biomassat arviot perustuvat valtakunnan metsien inventointiin (VMI; Korhonen ym., 2024) ja puutason biomassamalleihin (ks. luku ”Elävän puuston biomassa”). Vuodesta toiseen käytetään samoja karikekertoimia, jotka perustuvat erilaisiin kokeisiin ja tutkimuksiin (ks. Lehtonen ja Heikkinen, 2016).

Puuston poistuman, eli hakkuiden ja luonnonpoistuman, vuosittaiset runkotilavuusarviot saadaan Luken tilastojulkaisuista (Luke, 2025a). Nämä arviot muunnetaan biomassoiksi VMI-aineistoihin perustuvilla kertoimilla.

Seuraavissa kolmessa luvussa kuvataan KHKI:ssä käytettävä hiilitaseen arviointimenetelmän yksityiskohtia siinä määrin kuin on tarpeellista epävarmuuden arviointimenetelmän perusteiden ymmärtämiseksi. Sen jälkeen esitellään kokonaisuutena epävarmuuden arvioinnin perusperiaate ja yksityiskohtaisemmin elävästä puustosta peräisin olevan karikesyötteen epävarmuuden arviointimenetelmä. Lopuksi hahmotellaan vielä epävarmuusarvioiden rajoitteita.

Tässä kirjoituksessa esitettävät tulokset perustuvat biomassavarastojen, karikesyötteen arvioiden osalta (Kuva 1) vuodet 1990 – 2023 sisältävään KHKI-raporttiin (Forsell ym., 2025) ja epävarmuusarvioiden (Kuvat 3 – 5) osalta uusiin, aiemmin raporttoimattomiin simulointeihin, joiden tavoitteena oli päivittää Lehtosen ja Heikkisen (2016) julkaisun tulokset. Tässä kirjoituksessa esiteltäviä elävän puuston biomassojen epävarmuusarvioita lukuunottamatta menetelmät olivat samoja kuin tuossa julkaisussa. VMI-aineistoista käytettiin viimeisimmän kokonaan mitatun kierroksen, VMI13:n (2019 – 2023), ja sitä edeltävien kierrosten aineistoja.



Kuva 2: Biomassaositteet.

Elävän puuston biomassa

Elävän puuston kariketuotannon (kuten myös puuston poistumasta peräisin olevan karikesyötteen; seur. luku) määrän arviointia varten puustobiomassaa arvioidaan puulajiryhmittäin (mänty, kuusi, lehtipuut) ositteissa (kuorellinen runkopuu, oksat, lehdet, kanto, juuret ja hienijuuret (Kuva 2); esimerkiksi leh-

tipuiden lehtiositteen karikkekerroin on tietysti suurempi kuin havupuiden ja molemmat monta kertaluokkaa suurempia kuin runkopuuositteiden karikkekertoimet. Biomassavarastojen arviointia varten on kehitetty puutason ositteittaisia biomassamalleja (Repola, 2008; Repola, 2009).

Valtakunnan metsien inventoinneissa poimitaan satunnaisotoksia kaikista metsämailla kasvavista puista. Otokseen valittuja puita kutsutaan ”lukupuiksi”. Kaikille lukupuista määritetään puulaji, mitataan rinnankorkeusläpimitta ja ennustetaan runkotilavuus (Korhonen ym., 2024, S1). ”Koepuista”, jotka ovat osaotos lukupuista, mitataan lisäksi mm. pituus ja elävän latvuksen pituus, jotka ovat tärkeitä selittäviä muuttujia Repolan (2008, 2009) biomassamalleissa. Koepuille voidaan näin ollen ennustaa runkotilavuuden lisäksi myös ositteittaiset biomassat.

Puukohtaisista tilavuus- ja biomassaennusteista laaditaan elävän puuston (lb) biomassa-arviot erikseen kaikille inventaarioon sisältyville laskenta-alueiden r (Etelä-Suomi, Pohjois-Suomi), puulajiryhmien s , biomassaositteiden c ja VMI-kierrosten t yhdistelmille:

$$\hat{B}_{lb,rst} = \widehat{BCEF}_{lb,rst} \hat{V}_{lb,rst},$$

missä $\hat{V}_{lb,rst}$ on lukupuiden tilavuusennusteisiin perustuva runkotilavuusarvio (Korhonen ym., 2024, kappale 2.3.2) ja biomassan muuntokerroin $\widehat{BCEF}_{lb,rst}$ on koepuista saatujen biomassa- ja runkotilavuusestimaattorien suhde. Tarkemmat yksityiskohdat löytyvät Forsellin ym. (2025) liitteestä 6c.

Poistuman muuntokertoimet ja karikesyöte

Hakkuiden ja luontaisen kuolemisen seurauksena poistuneiden puiden biomassan muuntokerrointen estimointi perustuu VMI:n pysyviin koealoihin, joilla samoja puita mitataan n. viiden vuoden välein. Hakkuupoistuman kertoimet $\widehat{BCEF}_{\logg,rst}$ estimoidaan VMI-kierroksen $t-1$ mittauksista niistä puista, jotka on hakattu kierrosten $t-1$ ja t välisenä aikana, ja luonnonpoistuman kertoimet $\widehat{BCEF}_{nm,rst}$ niistä puista, jotka ovat kuolleet edellisen mittauksen jälkeen.

Luonnonpoistuma on kokonaisuudessaan karikesyötettä, joten siitä tarvitaan vain kokonaisbiomassa. Hakkuupoistuman osalta ositekohtaiset kertoimet ovat kuitenkin olennaisia, koska hakatusta runkopuusta suurin osa korjataan pois metsästä ja muut biomassaositteet jäävät pääosin karikesyötteeksi.

Karikesarjojen interpolointi ja vuositaseet

Kukin elävän puuston biomassa-arvio $\hat{B}_{lb,rst}$ kohdennetaan sen pohjana olevien VMI-mittausten keskimääräiseen ajankohtaan ja kullekin alueelle r , puulajiryhmälle s ja biomassaositteelle c muodostetaan ajassa jatkuva biomassasarja näiden ajankohtien välisellä lineaarisella interpolaatiolla; viimeisen VMI-kierroksen keskimääräisen mittausajan jälkeistä biomassaa ekstrapoloidaan vakiona (Kuva 1, keskimäinen kuvio vasemmalla). Vuosittaisia elävän puuston karikesyötteiden määriä (Kuva 1, alavasen) arvioidaan poimimalla aikasarjoista vuoden ensimmäiseen päivään liittyvät biomassa-arviot ja kertomalla nämä kyseisen puulajiryhmän ja biomassaositteiden karikkekertoimella.

Poistuman muuntokertoimet interpoloidaan vastaavalla tavalla ja kerrotaan sitten vuosittaisilla runkotilavuusarvioilla. Edellä mainitun mukaisesti karikesyötteeseen lisätään luonnonpoistuman kaikkien ositteiden biomassa-arviot.

Hakkuupoistumasta siihen lisätään muiden kuin runko-ositteen biomassat kokonaisuudessaan ja runkomassasta arvioitu hukkapuuosuus. Syötteestä kuitenkin poistetaan vuosittain kerättävän energiapuun arvioitu biomassa.

Näin muodostetut karikeaikasarjat sekä tarvittavat vuosittaiset säätiedot annetaan syötteeksi Yasso07-mallille, jolla saadaan arviot kivennäismaametsien maaperän orgaanisen aineksen, kuolleen puun ja karikkeen hiilivaraston vuosittaisista muutoksista.

Kokonaisepävarmuuden arvioinnin perusperiaate

Kivennäismaaperän hiilitaseaikasarjojen epävarmuusarviointi toteutetaan Monte Carlo -menetelmällä. Siinä tuotetaan suuri määrä satunnaisia realisaatioita (taseaikasarjoja) todennäköisyysjakaumasta, jolla pyritään kuvaamaan aikasarjaestimaattiin liittyviä epävarmuuksia.

Epävarmuuden mittarina käytetään hallitustenvälisen ilmastomuutospaneelin ohjeistuksen (IPCC, 2006, vol. 1, luku 3) mukaista suuretta

$$U = \frac{100 \left(\frac{D}{2}\right)}{|X|},$$

missä X on tasearvio ja D sen 95%:n luottamusvälin pituus. Kun arvioidaan esimerkiksi vuoden 2023 tasearvion epävarmuutta, X on tähän vuoteen liittyvä arvo taseaikasarjan piste-estimaatista ja

$$\hat{D} = 2 \bullet 1.96\sigma,$$

missä σ on vuoteen 2023 liittyvien arvojen keskihajonta yli Monte Carlo -menetelmällä tuotettujen aikasarjojen.

Kukin realisaatio tuotetaan

1. lisäämällä estimoituun karikesyöteaikasarjaan sen epävarmuutta jäljittelevät satunnaisvirheet,
2. lisäämällä Yasso07-mallin parametreihin näiden epävarmuutta jäljittelevät satunnaisvirheet ja
3. toteuttamalla malliajo kohtien 1 ja 2 arvoilla.

Kohdan 1 satunnaisvirheiden simulointia käsitellään seuraavassa luvussa tarkemmin elävästä puustosta peräisin olevan karikesyötteen osalta. Kohdan 2 satunnaisvirheet simuloidaan Yasso07-parametrien posteriorijakaumasta (Tuomi ym., 2011).

Elävän puuston karikesyötteen epävarmuus

Repolan (2008, 2009) mallit ovat muotoa

$$\log(b_{ic}) = \beta'_{s(i),c} \mathbf{x}_{ic} + \varepsilon_{ic},$$

missä b_{ic} on puun i ositteen c biomassa, vektori β_{sc} sisältää puulajiryhmän s ositteen c malliparametrit ja vektori $\mathbf{x}_{ic}$ tämän mallin selittävät muuttujat. Enustevirheet ε_{ic} muodostuvat metsikkökuviokohtaisten satunnaiskomponenttien

ja puukohtaisten jäännösten summana, $\varepsilon_{ic} = \alpha_{s(i),m(i),c} + \varepsilon_{0,ic}$. Saman puulajiryhmän eri ositteisiin liittyvät satunnaiskomponentit ja jäännökset ovat keskenään korreloituneita eli $\alpha_{sm} \sim N(\mathbf{0}, \Sigma_{\alpha,s})$ ja $\varepsilon_{0,i} \sim N(\mathbf{0}, \Sigma_{\varepsilon,s})$, missä vektori $\alpha_{sm} = [\alpha_{smc}]$ sisältää kaikki yksittäiseen lajiryhmään s ja kuvioon m liittyvät satunnaiskomponentit ja vektori $\varepsilon_{0,i} = [\varepsilon_{0,ic}]$ kaikki yksittäiseen puuhun i liittyvät jäännökset.

Näin ollen sovitettut mallit sisältävät parametriestimaattien $\hat{\beta}_{sc}$ ja niiden kovarianssimatriisien estimaattien $\hat{\Sigma}_{\beta,sc}$ lisäksi ennustevirheisiin liittyvien kovarianssimatriisien estimaatit $\hat{\Sigma}_{\alpha,s}$ ja $\hat{\Sigma}_{\varepsilon,s}$. Kaikkien Repolan (2008, 2009) mallien parametriestimaatit on koottu Ståhlin ym. (2014) taulukkoon A1.1 ja kovarianssimatriisien estimaatit taulukoihin A1.2 – A1.4.

Biomassaestimaattorien $\hat{B}_{lb,rsc} = \widehat{BCEF}_{lb,rsc} \hat{V}_{lb,rsc}$ ”tilastollisen” epävarmuuden arviointi perustuu olennaisesti siihen oletukseen, että mallit kuvaavat biomassan vaihtelun kaikkien Suomen metsämaiden puustoissa. Parametriestimaattien epävarmuuden ja jäännösvaihtelun lisäksi otetaan huomioon VMI-aineiston otantaan liittyvä epävarmuus. VMI:n otantavarianssi estimointia on kuvattu yleisellä tasolla mm. julkaisun Korhonen ym. (2024) kappaleessa 2.3.3. Nyt käsillä olevaan tapaukseen liittyy kaksi seikkaa, joiden takia tarvitaan tavallista mutkikkaampaa menetelmää.

Ensinnäkin VMI:n ”tavallisessa” otantavarianssin estimoinnissa ei oteta huomioon vastemuuttujan arvoihin liittyvää epävarmuutta. Normaalaa satunnaista ja korreloimatonta mittausvirhettä ei tarvitsekaan ottaa huomioon, sillä se kasvattaa havaintojen välistä vaihtelua ja tulee tätä kautta automaattisesti huomioitua otantavarianssia estimoitaessa. Tilanne on kuitenkin toinen, kun vastemuuttujan arvo on ennustettu vaihtelua pienentävällä regressiomallilla, kuten käsillä olevassa tilanteessa. Näin on myös VMI:n runkotilavuuksien laita, mutta niihin liittyvä ennustevirhe on oletettu mitättömäksi. Tässä kirjoituksessa raportoitavat tulokset tukevat tätä oletusta. Muiden ositteiden kuin rungon biomassan ennusteisiin liittyvän virheen vaikutusta otantavarianssin estimointiin ei kuitenkaan voi jättää huomiotta, kuten jäljempänä esitettävät tulokset (Kuva 5, vas.) osoittavat. Monte Carlo -menetelmään perustuvassa epävarmuusarvioinnissa ennustevirheen vaikutus otantavarianssin estimointiin voidaan huomioida lisäämällä ennusteisiin jäännös(ko)varianssin mukaisia satunnaisvirheitä ennen otantavarianssin estimointia (allaolevan algoritmin kohta 3).

Toiseksi, estimaattorien $\widehat{BCEF}_{lb,rsc}$ ja $\widehat{BCEF}_{lb,rsc't}$, $c' \neq c$, otantavirheet ovat välttämättä vahvasti korreloituneita, koska ne perustuvat samojen puiden samoihin mittauksiin. Myös estimaattorit $\widehat{BCEF}_{lb,rsc}$ ja $\hat{V}_{lb,rsc}$ perustuvat osin samoihin puihin, joten niidenkin otantavirheiden mahdolliset korrelaatiot on huomioitava. VMI:n tavallisen otantavarianssin yleistys kovarianssien estimointiin esitetään esimerkiksi julkaisun Haakana ym. (2020) liitteessä A.

Summa summarum, luvussa ”Kokonaisepävarmuuden arvioinnin peruseriaate” hahmotellun Monte Carlo -algoritmin kohdassa 1 toimittiin seuraavasti, kun tuotettiin yksi realisaatio k epävarmuusjakauman mukaiset satunnaisvirheet sisältävästä vuosittaisesta elävän puuston karikesyöteajaksarjasta.

1. Simuloidaan ”epävarmuuden sisältävät” biomassamallien parametriarvot $\beta_{sc,k} \sim N(\hat{\beta}_{sc}, \Sigma_{\beta,sc})$ jokaiselle puulajiryhmälle s ja ositteelle c , puulajiryhmittäisten satunnaiskomponenttien realisaatiot $\alpha_{sm,k} \sim N(\mathbf{0}, \Sigma_{\alpha,s})$ kullekin käytettävien VMI-aineistojen kuviolle m ja jäännösvektorit $\varepsilon_{0,i,k} \sim N(\mathbf{0}, \Sigma_{\varepsilon,s})$ kullekin käytettävälle VMI:n koepuulle i .

2. Simuloidaan puu- ja komponenttikohtaiset epävarmuuden sisältävät biomassasaennusteet, joissa on mukana kohdassa 1 simuloidut parametrien estimointivirheet, satunnaiskomponentit ja jäännökset:

$$b_{ic,k} = \exp \left(\beta'_{s(i),c,k} \mathbf{x}_{ic} + \alpha_{s(i),m(i),k,c} + \varepsilon_{0,ic,k} \right).$$

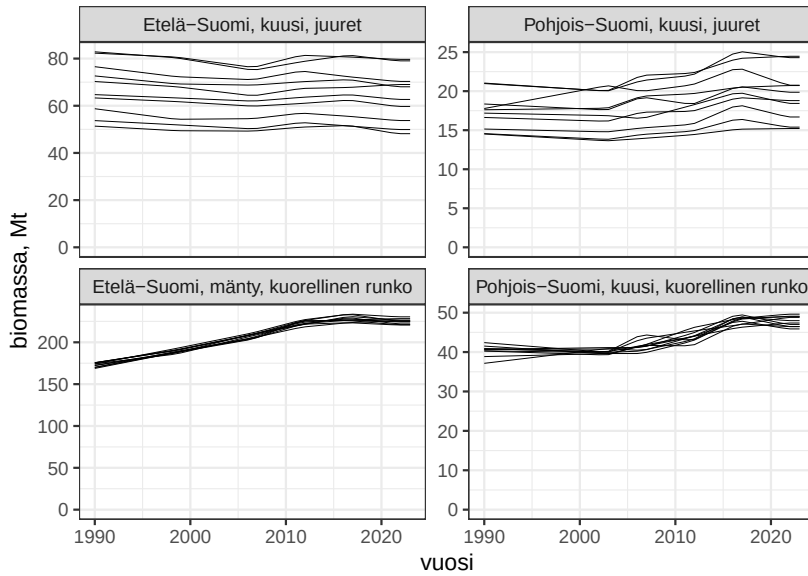
3. Lasketaan kohdassa 2 simuloituihin biomassoihin $b_{ic,k}$ perustuvat muunto-kerrointen estimaatit $\widehat{BCE\widehat{F}}_{lb,rsct,k}$ ja estimoidaan kullekin alueelle r , puulajiryhmälle s ja VMI-kierrokselle t erikseen näiden sekä runkotilavuus-estimaattorien $\widehat{V}_{lb,rsct}$ otantakovarianssit perustuen simuloitujen biomassojen $b_{ic,k}$ ja runkotilavuuden VMI-ennusteiden vaihteluun.
4. Lisätään kohdan 3 kovarianssimatriiseista simuloidut otantavirhevektorit muuntokerrointen ja runkotilavuuksien estimaatteihin ja lasketaan näin saatujen, epävarmuuden sisältävien, muuntokerrointen ja runkotilavuuksien tuloina simuloidut epävarmuuden sisältävät biomassat $B_{lb,rsct,k}$.
5. Interpoloidaan kohdassa 4 saadut VMI-kierros -kohtaiset biomassat vuosittaisiksi, kuten luvussa ”Karikesarjojen interpolointi . . .” kuvattiin.
6. Kerrotaan kohdassa 5 saadut vuosittaiset biomassaa-arviot karikekertoimilla, joihin on lisätty niiden epävarmuusjakaumasta simuloitu estimointivirhe.

Biomassamallien parametriestimaattorien variansseista johtuvaa epävarmuutta arvioitiin laskemalla tulokset algoritmin sellaisella versiolla, jossa kohdan 1 satunnaiskomponenttien ja jäännösten arvoksi asetettiin aina nolla eikä kohdassa 4 lisätty simuloituja otantavirheitä, otantavirheestä johtuvaa epävarmuutta puolestaan sellaisella versiolla, jossa biomassamallien parametreihin ei simuloitu vaihtelua.

Tuloksia

Biomassamallien parametriestimaattorien variansseista johtuva epävarmuus kokonaisbiomassan estimaateissa $\widehat{B}_{lb,rsct}$ vaihteli runkomassalle välillä 2,2 – 3,8 %, lehtimassalle välillä 3,8 – 11 % ja juurimassalle välillä 12 – 32 %. Otantavirheestä johtuva epävarmuus oli Etelä-Suomessa 1,0 – 3,0 %, Pohjois-Suomessa 1,4 – 4,2 %. Näiden kahden epävarmuuden lähteen vaikutus biomassasarjoihin on erilaatuinen. Parametrien estimointivirheet ovat samat läpi koko sarjan, jolloin ne aiheuttavat systemaattisen tasomuutoksen. Tämä näkyy parhaiten Kuvan 3 vasemmanpuoleisissa paneeleissa, joissa otantavirhe on pieni. Eri VMI-kierrosten otantavirheet ovat puolestaan toisistaan riippumattomia¹, joten ajallista satunnaisvaihtelua on sitä enemmän, mitä suurempi on otantavirheistä johtuvan epävarmuuden osuus (esim. Kuvan 3 oikeanpuoleiset paneelit).

Samaan alueeseen ja puulajiryhmään liittyvät oksa- ja lehtibiomassan arviot olivat vahvasti korreloituneita (Kuva 4, vas.), koska sekä niihin liittyvät ennustemallin jäännökset että niiden otantavirheet ovat korreloituneita. Jälkimmäinen korrelaatio on seurausta siitä, että saman alueen ja puulajiryhmän oksa- ja lehtibiomassojen arviointiin käytetään samoja VMI-mittauksia. Samoin samaan puulajiryhmään ja biomassaositteeseen liittyvät arviot olivat korreloituneita (Kuva 4, kesk.), koska niissä käytetään samoja malleja samoilla parametrialvoilla ja

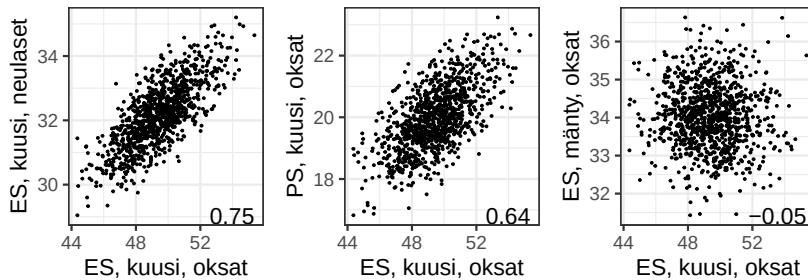


Kuva 3: Esimerkkejä epävarmuusjakaumasta simuloituista biomassa-aikasarjoista. Alemmissa paneeleissa biomassamallin parametriestimaattorien epävarmuudesta johtuva vaihtelu on vähäisempää, ylemmissä se on suurempaa. Vasemmanpuoleisissa paneeleissa otantavirhe on pienempi, oikeanpuoleisissa suurempi.

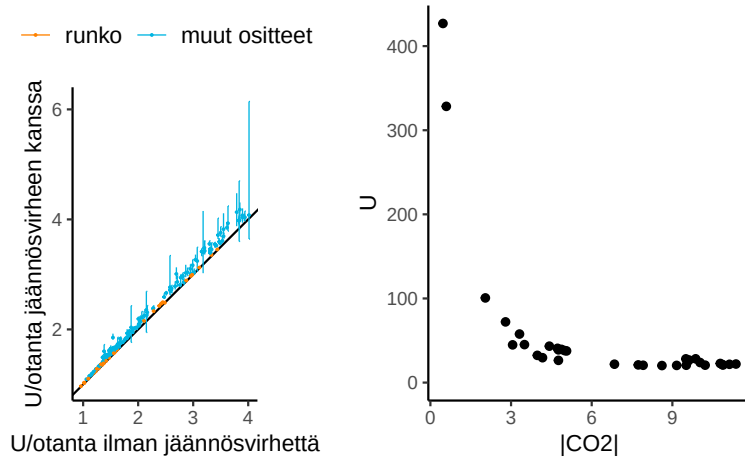
näin ollen myös samoilla parametrivirheillä. Toisaalta eri puulajiryhmiin liittyvät arviot ovat toisistaan riippumattomia (Kuva 4, oik.), koska niiden perustana on eri VMI-mittaukset ja eri mallit.

Ennustevirhe ei juuri vaikuttanut runkomassojen otantavirheisiin (Kuva 5, vas.). Joissakin muissa ositteissa otantavirhe tulisi kuitenkin aliarvioitua, jos sitä arvioitaisiin ennen biomassamallin jäännösvarianssin mukaisen satunnaisvirheen lisäämistä ennusteisiin.

KHKI-raportoinnissa kivennäismaaperän hiilitasearvioiden suhteellinen epävarmuus on vuodet 1990 – 2023 sisältävään raporttiin (Forsell ym., 2025) saakka oletettu vakioksi $U = 31,5\%$, joka perustuu Lehtosen ja Heikkisen (2016) tulok-



Kuva 4: Vuodelle 2023 simuloituja, epävarmuuden sisältäviä biomassa-arvioita, Mt. ES =Etelä-Suomi, PS=Pohjois-Suomi, oikean alakulman luku = Pearsonin korrelaatiokerroin.



Kuva 5: Vas.: Otannasta johtuvan epävarmuuden arvioita ilman jäännösvirheen simulointia ja jäännösvirheen simuloinnin jälkeen estimoituina. Pystyviivat näytävät vaihteluvälin yli simuloitujen realisaatioiden ja pisteiden y-arvot mediaanin. Oik.: Koko maan metsien kivennäismaaperän hiilitasearvioiden (varaston muutos / vuosi; Mt CO₂-ekv.) epävarmuuden suhde arvioiden iteisarvoihin.

siin. Sellaisilla nieluvarvioilla, joita saatiin n. vuoteen 2006 saakka (< -6 Mt CO₂-ekv; Luke, 2025b), tämä oletus osoittautuikin kohtuullisen hyväksi, mutta viime vuosien osalta, kun tasearvio on lähestynyt nollaa, suhteellinen epävarmuus on luonnollisesti kasvanut jyrkästi (Kuva 5, oik.). Luonnonvaratieto-sivuston epävarmuuskirjoituksessa (Luke, 2025c) arvioitiin vuoden 2023 päästön, 0,4 Mt CO₂-ekv, epävarmuudeksi 524,3 % perustuen Lehtosen ja Heikkisen (2016) tuloksiin ja olettaen vakio U :n sijasta vakio D . Kuvan 5 nojalla tämä arvio oli huomattavasti realistisempi.

Lopuksi

Uusimmat arviot Suomen kasvihuonekaasuinventaarion epävarmuuksista on metsiä koskevien hiilitasetulosten osalta päivitetty tässä kirjoituksessa esitetyjä laskelmia hyödyntäen (Luke, 2025d). Erityisesti arvio kivennäismaaperää koskevista epävarmuuksista on merkittävästi aiempaa realistisempi. Perusteellisempaan pohdintaan tulosten merkityksestä hiilitaseiden arvioiden tutkimukselle ja yhteiskunnalliseen päätöksentekoon sekä vertailuun kansainväliseen tutkimukseen syvennyttään aiheesta tekeillä olevassa kansainväliseen julkaisusarjaan lähetettävässä tutkimusartikkelissa.

Tässä kirjoituksessa käsitellyt, mallinnus- ja sovellusaineistoista laskettavissa olevat ”tilastolliset” epävarmuusarviot eivät sisällä käytettävien mallien rakenteeseen tai soveltuvuuteen tai mallien sovitukseen käytettyjen aineistojen edustavuuteen liittyvää epävarmuutta. Esimerkiksi biomassamallien sovitusaineistoa ei ole poimittu satunnaisotannalla kaikista Suomen metsistä, joten aineiston valikoitumisharhan mahdollisuutta ei voi helposti sulkea pois. Lisäksi voidaan kyseenalaistaa esimerkiksi se, voiko samoja malleja käyttää vuodesta ja vuosikymmenestä toiseen vai onko biomassa-allokaatio mahdollisesti muuttunut. Tai se, ovatko kariketuotantokertoimet erilaisia nyt kuin 1990-luvulla.

Tämän kaltaisten epävarmuuden lähteiden vaikutusta voidaan arvioida riippumattomiin aineistoihin perustuvilla validoinneilla. Esimerkiksi Yasso07-mallin soveltuvuutta on validoitu julkaisussa Rantakari ym. (2012).

Viitteet

- Alm, J. ym. (2023). "A new method for estimating carbon dioxide emissions from drained peatland forest soils for the greenhouse gas inventory of Finland". *Biogeosciences* 20.18, s. 3827–3855. DOI: [10.5194/bg-20-3827-2023](https://doi.org/10.5194/bg-20-3827-2023).
- Forsell, P. ym. (2025). *Greenhouse gas emissions in Finland 1990 to 2023 – National inventory report under the UNFCCC and Paris Agreement*. Tekninen raportti. Statistics Finland. URL: https://stat.fi/media/uploads/tup/khkinv/fi_nid_un_2023_2025-04-15.pdf.
- Haakana, H. ym. (2020). "Precision of exogenous post-stratification in small-area estimation based on a continuous national forest inventory". *Canadian Journal of Forest Research* 50.4, s. 359–370. DOI: [10.1139/cjfr-2019-0139](https://doi.org/10.1139/cjfr-2019-0139).
- IPCC (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Toim. H. S. Eggleston ym. IGES, Japan. URL: <https://www.ipcc-nggip.iges.or.jp/public/2006gl/>.
- Korhonen, K. T. ym. (2024). "Forests of Finland 2019–2023 and their development 1921–2023". *Silva Fennica* 58.5. DOI: [10.14214/sf.24045](https://doi.org/10.14214/sf.24045).
- Lehtonen, A. ja Heikkinen, J. (2016). "Uncertainty of upland soil carbon sink estimate for Finland". *Canadian Journal of Forest Research* 46.3, s. 310–322. DOI: [10.1139/cjfr-2015-0171](https://doi.org/10.1139/cjfr-2015-0171).
- Luke (2025a). *Hakkuukertymä ja puuston poistuma*. <https://www.luke.fi/fi/tilastot/hakkuukertyma-ja-puuston-poistuma>.
- (2025b). *Kasvihuonekaasuinventaario 2023: maataloussektorin ja maankäyttösektorin lopulliset tulokset hyvin lähellä ennakkotuloksia*. <https://www.luke.fi/fi/uutiset/kasvihuonekaasuinventaario-2023-maataloussektorin-ja-maankayttosektorin-lopulliset-tulokset-hyvin-lahella-ennakkotuloksia>.
- (2025c). *Kasvihuonekaasuinventaarion epävarmuudet*. <https://www.luke.fi/fi/luonnonvaratieto/tiedetta-ja-tietoa-luonnonvaroista-luonnonvaroista/kasvihuonekaasuinventaario/yleista-kasvihuonekaasuinventariosta/kasvihuonekaasuinventaarion-epavarmuudet>.
- (2025d). *Maankäyttösektori pysyy suurena päästölähteenä mutta metsät ovat pieni nielu – kasvihuonekaasuinventaarion 2024 ennakkotiedoissa*. <https://www.luke.fi/fi/uutiset/maankayttosektori-pysyy-suurena-paastolahteenamutta-metsat-ovat-pieni-nielu-kasvihuonekaasuinventaarion-2024-ennakkotiedoissa>.
- Rantakari, M. ym. (2012). "The Yasso07 soil carbon model – Testing against repeated soil carbon inventory". *Forest Ecology and Management* 286, s. 137–147. DOI: [10.1016/j.foreco.2012.08.041](https://doi.org/10.1016/j.foreco.2012.08.041).
- Repola, J. (2008). "Biomass equations for birch in Finland". *Silva Fennica* 42.4, s. 236. DOI: [10.14214/sf.236](https://doi.org/10.14214/sf.236).
- (2009). "Biomass equations for Scots pine and Norway spruce in Finland". *Silva Fennica* 43.4, s. 625–647. URL: <http://www.metla.fi/silvafennica/full/sf43/sf434625.pdf>.
- Ståhl, G. ym. (2014). "Sample-based estimation of greenhouse gas emissions from forests - a new approach to account for both sampling and model errors". *Forest Science* 60.1, s. 3–13. DOI: [10.5849/forsci.13-005](https://doi.org/10.5849/forsci.13-005).

Tuomi, M. ym. (2011). "Soil carbon model Yasso07 graphical user interface". *Environmental Modelling & Software* 26.11, s. 1358–1362. DOI: [10.1016/j.envsoft.2011.05.009](https://doi.org/10.1016/j.envsoft.2011.05.009).

Loppuviitteet

¹Näissä tarkasteluissa ei ole otettu huomioon pysyvien koealojen toistomittausten korreloituneisuutta.

Kommentti: Joitakin haja-ajatuksia vaihtelusta, satunnaisuudesta ja epävarmuudesta

ELJA ARJAS

ELJA.ARJAS@HELSINKI.FI

MATEMATIIKAN JA TILASTOTIETEEN OSASTO

HELSINGIN YLIOPISTO

Kysymys metsien hiilitaseesta on viime aikoina noussut julkiseen keskusteluun ja muodostaa tärkeän osan niistä pohdinnoista, joilla Suomessa pyritään varautumaan ilmastomuutokseen. Erityistä huomiota on kiinnitetty metsämaiden kivennäismaaperää koskeviin arvioihin, joissa on tapahtunut ilmeisesti useimpia asiantuntijoitakin yllättäneitä muutoksia. Juha Heikkisen kirjoitus, joka käsittelee Suomessa käytettyjä arviointimenetelmiä sekä tuloksiin liittyvää epävarmuutta, on sen vuoksi tärkeä ja ajankohtainen.

Maallikkona on ollut hieman hämmentävää seurata, kuinka tiedotusvälineissä raportoidut tulokset ovat vaihdelleet suhteellisen lyhyenä aikana. Aiemmin pidettiin selvänä, että metsät muodostavat merkittävän hiilinielun, jonka varassa Suomen olisi melko vaivatonta selviytyä kansainvälisistä sitoumuksistaan ilmakehän hiilidioksidipitoisuuden hillitsemiseksi. Sitten kerrottiin, että ne ovatkin päästölähde. Tuoreimpien tietojen mukaan ne ovat jälleen nettonielu, tosin tuota aiempaa arviota huomattavasti pienempi. Vaikka itse kohde, kivennäismaaperään sitoutuneen hiilen määrä, varmasti vaihtelee ajan mukana, on sen hyvin nopeisiin suuriin muutoksiin vaikea uskoa. Lieneekin tunnistettava, että tässä mittausongelmassa on hyvin vaikeaa erottaa signaalia ja kohinaa toisistaan: hiilinielujen ja -päästöjen suuruuta koko maan alueelta ei voi millään menetelmällä mitata tarkasti.

Tilastotieteen kursseilla opetetaan tavallisesti, että *satunnaisia* ovat asiat, jotka *vaihtelevat*, esimerkkinä termit *satunnaismuuttuja* ja *satunnaisotos*. Käytännössä kumpikaan niistä ei tosin enää vaihtelee sen jälkeen, kun niiden arvot on havaittu ja niistä on muodostunut tilastotieteilijälle *dataa*. Kyse on enemmänkin ajatuskokeesta: ”Jos samanlainen mittaus toistettaisiin vastaavissa olosuhteissa, niin silloin ...” Myös tilastotieteessä tavanomainen frekvenssitulkintaan nojaava todennäköisyyskäsite ripustetaan tähän useimmiten vain kuvitteelliseen toistokokeeseen. Samalla kuitenkin luovutaan tavasta määritellä käsite, joka vastaisi merkitykseltään sanalle *todennäköisyys* normaalissa kielenkäytössä annettua sisältöä.

Tilastotieteessä yleensä vältetään myös *epävarmuuden* käsitteen suoraa käyttöä, koska sille ei tilastotieteen teoriassa ole määriteltä västinettä. *Todennäköisyyden* käsitesisältö on jo kiinnitetty niin ahtaaksi, ettei se enää soveltuisi tähän tarkoitukseen. Pidän kuitenkin ongelmallisina kaikkia tilastotieteessä vakiintuneita tapoja ilmaista epävarmuutta havaitusta otoksesta laskettujen tunnuslukujen avulla. Ongelma heijastuu tilastotieteen termeissä, kuten kiertoilmaisujen *luottamus* ja *riski* runsaana käytönä.

Pohdin seuraavassa hieman sitä, voitaisiinko hiilitaseongelmaankin liittyvää epävarmuutta edes periaatteessa käsitellä niin, että siinä vedottaisiin suoraan

todennäköisyyden käsitteeseen. Todennäköisyyskäsite täytyy silloin kuitenkin tulkita perinteistä toistokoekehikkoa väljemmin.

Juha Heikkisen kuvaama arviointiprosessi jakaantuu useaan eri vaiheeseen, joista ensimmäisen, ja samalla koko tutkimuksen lähtökohdan, muodostavat määrääjoin tehtävät elävän puuston VMI-inventointiohjelmat. Niiden taustana on vuosikymmeniä jatkunut vahva suomalainen metsänarvioinnin tutkimusperinne. Elävän puuston laji- ja ositekohtaisista biomassaa-arvioista siirrytään niitä vastaaviin karikemääriin, joista lopuksi päädytään Yasso07-malliin avulla kasvihuonekaasujen määriä koskeviin arvioihin. Tutkimuksen viimeistä vaihetta ei Heikkisen kirjoituksessa käsitellä tarkemmin.

Kaikille näille tutkimuksille on tunnusomaista, että suurta kokonaisuutta koskevat johtopäätökset joudutaan perustamaan vain hyvin rajoitettuun havaintomäärään. Inventoinnissa käytetyt menetelmät noudattavat asetelmaperusteista (design-based) tutkimusperinnettä, jossa otantatodennäköisyydet huomioon ottavia painokertoimia käyttämällä konstruoidaan harhattomia estimaattoreita. Arvioihin liittyvä epävarmuus ilmaistaan sitten varianssien tai luottamusvälien avulla.

Heikkisen esittelemässä tilanteessa ei käytetä, kirjoituksessa mainituista perustelluista syistä, havaintokohteista tehtyjä suoria mittauksia. Ne korvataan sovitettuna regressiomallin avulla lasketuilla ennustearvoilla, joita Monte Carlo -simulaationa tuotetaan hyvin suuri määrä. Samalla ennusteisiin sisällytetään 3-tasoisesta hierarkkisen mallin mukaisesti termejä, joita kirjoituksessa luonnehditaan sanoilla ”epävarmuutta jäljittelevä satunnaisvirhe”. Täten siinä siirrytään – ainakin puoliksi – malliperusteisten (model-based) menetelmien piiriin.

Menetelmä on läheistä sukua puuttuvan datan käsittelyssä laajasti käytetylle monipaikkaukselle (multiple imputation). Monte Carlo -simulaation tuotoksena, kun toistojen määrä on suuri, yksittäisen havainnon korvaakin käytännössä kokonainen mallin perustuva todennäköisyysjakauma. Tällä laskenta-algoritmista johtuvalla teknisellä muutoksella voi nähdä olevan merkittävä seuraus myös tulosten tulkinnassa: jakauma ei vain ”jäljittele epävarmuutta” vaan se on määrittelyllisesti epävarmuuden kvantitatiivinen ilmaus suhteessa simuloinnissa käytettyyn generoivaan malliin. Tarkasteltua ja mielenkiinnon kohteena olevaa suureta, kuten puulajiryhmän ja ositteen mukaista biomassaa, voidaan siten käsitellä todennäköisyyslaskennan sääntöjä noudattaen satunnaismuuttujana silloinkin, kun sen arvoihin ei liity esimerkiksi ajan mukana tapahtuvaa vaihtelua; tähän riittää, ettei tarkkaa arvoa tunneta.

Malliperusteisella lähestymistavalla yhdistettynä tähän satunnaisuuden ja todennäköisyyskäsitteen liberaaliin tulkintaan voisi – ainakin periaatteessa – olla seurauksia myös siihen, miten metsien hiilitaseongelmaa käsiteltäisiin.

Ensimmäinen mieleen tuleva mahdollisuus olisi luopua Monte Carlo -simuloinneissa käytetyistä kiinteistä, aiempien tutkimusten tuloksina saaduista pistoke-estimaateista (plug-in estimate) regressiokerrointen ja kovarianssimatriisien arvoille, ja siten ottaa huomioon myös niihin liittyvä epävarmuus. Teknisesti tämä tarkoittaisi simulointimallin laajentamista siten, että siinä käytettäisiin alkuperäisen mittausdatan lisäksi hyväksi myös generoivan mallin parametreja koskevaa asiantuntijatietoa määrittämällä niitä koskeva ennakkojakauma (prior distribution). Todennäköisyyslaskennan teoria antaa tällaisten, hierarkkisen mallin eri kerroksiin sisältyvien epävarmuuksien käsittelyyn yksinkertaisen reseptin: kysymys on sisäkkäisten ehdollisten odotusarvojen käsittelystä. Niihin liittyvät numeeriset laskut ovat myös periaatteessa suoraviivaisesti toteutetta-

vissa algoritmisilla menetelmillä.

Monipaikkauksen tekniikkaa voisi ehkä olla mahdollista kehittää eteenpäinkin, luopuen silloin asetelmaperusteisen otantavarianssin tarkastelusta, ja korvaten se esimerkiksi kunkin alueen, puulajiryhmän ja ositteen kohdalla, vastaavalla biomassan jakaumamuotoisella ennusteella. Näin syntyvä ennuste olisi jo sovelletun Monte Carlo -menetelmän jatke, ja edustaisi toteutettuna muuttujalavennuksen (data augmentation) periaatetta. Menetelmän toteutus edellyttäisi kuitenkin nykyisten regressiomallien laajentamista siten, että myös niissä käytettyjen, yksittäisiä puita koskevien selittävien muuttujien arvoja voitaisiin simuloida mallin avulla.

En toki tarkoita, että luotaisiin järjestelmä, jossa kaikille Suomen puille olisi virtuaalinen vastineensa, ja jota sitten pyöritettäisiin Monte Carlo -simulaationa. Jonkinlainen mallihierarkia olisi tietenkin välttämätön jo puhtaasti laskennallisista syistä. Mutta periaatteessa tällaisia malleja voitaisiin laatia ja kytkeä ne sitten peräkkäin, alkaen metsän puista ja jatkuen kariketuotannosta aina kasvihuonekaasuihin saakka. Kunkin osamallin tulokset toimoisivat seuraavan syötteenä. (Tuntematta tarkemmin Yasso07-mallin periaatteita ja toimintaa, yksi tällainen osamalli voi ollakin valmiina.) Eri osamallien latenttien muuttujien, eli malliparametrien ja virtuaalidatan, jakaumia uusinnettaisiin laskennan edetessä aina paikallisesti, poimimalla niille arvoja ehdollisesta jakaumasta, jossa kaikkien muiden muuttujien arvot on sillä hetkellä kiinnitetty.

Menetelmän suuri etu olisi, että se tuottaisi numeerisen approksimaation siihen sisältyvien latenttien muuttujien yhteisjakaumasta. Mikä tahansa näistä muuttujista voitaisiin ottaa tarkasteltavaksi lähemmin ja määrittää sille laskennan tuloksena syntyvä todennäköisyysjakauma. Jakauma olisi määritelmän mukainen ilmaus siitä epävarmuudesta, joka kyseisen muuttujan todellisesta arvosta vallitsisi mallin, kaiken siihen liitetyn ennakkotiedon sekä tehtyjen havaintojen perusteella. Jakaumasta voitaisiin sitten poimia havainnollisesti tulkittavia tunnuslukuja kuten kvanttileja tai todennäköisyysvälejä. Todennäköisyysjakaumiin pohjautuva tarkastelu avaisi myös jatkomahdollisuuden systemaattiselle päätösanalyysille, jossa eri päätösvaihtoehtojen seurauksia arvioitaisiin niiden odotusarvoisten hyötyjen ja haittojen perusteella.

On jo etukäteen selvää, että epävarmuuden aito huomioon ottaminen päätelyketjun eri vaiheissa, oletettujen vakioarvojen tai pistoke-estimaattien sijaan, pikemminkin kasvattaisi kuin pienentäisi lopputuloksiin sisältyvää epävarmuuden astetta. Epävarmuuden aito määrällinen arviointi on kuitenkin olennainen osa tieteellisen tutkimuksen laadunvarmistusta, ja todennäköisyyslaskenta toisaalta ainoa matemaattisesti ongelmaton ja käsitteellisesti tyydyttävä tapa käsitellä sitä systemaattisesti.

P.S. Kiitoset Juhalle tästä mielenkiintoisesta kirjoituksesta, jonka lukeminen johti ajatukseni edellä sanoitetuille harhapoluille. Toisaalta lukiessa mieleeni palautui tutkimusyhteistyömme lähes 30 vuotta sitten, jonka tuloksena syntyi kolme julkaisua. Yhdessä niistä (Heikkinen ja Arjas, 1999) tavoitteena oli tuottaa pienen metsäalueen topografiakartta, täydentäen silloin hilapisteissä tehtyjä maaston korkeusmittauksia karttaan merkityillä puiden sijaintitiedoilla. Tässä pienessä esimerkkitapauksessa onnistuimme toteuttamaan tehtävän käyttämällä vain parametrittomasti määriteltäviä todennäköisyysmallia ja sopivasti määriteltäviä Monte Carlo -laskenta-algoritmia.

Viitteet

Heikkinen, J. ja Arjas, E. (1999). "Modeling a Poisson forest in variable elevations: a nonparametric Bayesian approach". *Biometrics* 55.3, s. 738–745.
DOI: [10.1111/j.0006-341X.1999.00738.x](https://doi.org/10.1111/j.0006-341X.1999.00738.x).

Kommentti: Muutamia huomioita Suomen metsien kasvihuonekaasuinventoinnista

ERKKI TOMPPO

ERKKI.TOMPPO@HELSINKI.FI

METSÄTIETEIDEN OSASTO

HELSINGIN YLIOPISTO

Johdanto

Suomen ja muiden sopimuksiin kuuluvien maiden on raportoitava YK:n Ilmastopöytäkirjan mukaan vuosittain ja Pariisin sopimuksen mukaan joka toinen vuosi ihmistoiminnasta aiheutuvat kasvihuonekaasujen (KHK) päästöt ja poistot eli nielut. Kasvihuonekaasuinventaariotietoja käytetään kunkin jäsenen päästövähennystavoitteen ja -velvoitteen seurantaan (taakanjakosektori) sekä maankäyttösektorilla syntyvien päästöjen ja poistumien seurantaan EU:ssa LULUCF-asetusten (Land Use, Land Use Change and Forestry) mukaisesti. Asetukset määrittävät, mitkä maankäytön, maankäytön muutoksen ja metsänhoidon päästöt ja nielut huomioidaan EU:n ilmastotavoitteissa esimerkiksi vuosina 2021–2030. EU:n ilmastotavoitteet perustuvat eurooppalaiseen ilmastolakiin, jonka tavoitteena on tehdä EU:sta ilmastoneutraali vuoteen 2050 mennessä.

Maaluokat ovat: metsämaa, viljelysmaa, ruohikkoalueet, kosteikot, rakennetut alueet ja muu maa. Ne kattavat koko Suomen maa-alan ja sisävedet. Maankäyttöluokat jaetaan edelleen niihin, jotka ovat pysyneet edeltävät 20 vuotta samassa maankäyttöluokassa ja niihin, jotka ovat muuttuneet muusta maankäytöstä nykyiseen viimeisen 20 vuoden aikana. Raportoinnissa on siis 36 maankäyttöluokkaa ja tai maaluokkien välistä muutosta.

IPCC-ohjeiden (Intergovernmental Panel on Climate Change) mukaan hiilidioksidin (CO_2) päästöistä ja poistumista johtuvat hiilivarastojen muutokset näissä 36 maaluokassa raportoidaan viidessä varastossa: maanpäällinen ja maanalainen elävä kasvibiomassa, kuollut puuaines, karike ja maaperän orgaaninen hiili. Lisäksi raportoidaan metsälannoituksen dityppioksidin (N_2O) päästöt, ojitettujen metsämaiden ja turvetuotantoalueiden metaani- (CH_4) ja N_2O -päästöt, voimakkaasti muutettujen sisävesien CH_4 -päästöt, CH_4 -päästöt orgaanisen aineksen hajotuksesta kivennäismailla (alle 20 vuotta sitten metsämaaksi, viljelysmaaksi, ruohikkoalueeksi tai rakennetuksi alueeksi muuttuneet alueet) sekä maastopalojen, metsäkulutuksen ja ennallistamispolttojen päästöt.

IPCC-ohjeistus (2006) ohjeistus sisältää inventaariossa käytettäviä menetelmiä ja määritelmiä, mutta jättää paljon valinnan varaa maille. Yksinkertaisimmillaan, eli tason 1 mukaan, voidaan käyttää ilmastovyöhykkeittäisiä kertoimia. Seuraava taso 2 edellyttää kansallisia kertoimia ja vaativin taso, taso 3, seurantamittauksia ja/tai kansallisiin oloihin kehitettyjä malleja. Tasot 1 ja 2 eivät anna riittävän luotettavia päästö- ja nieluestimateja esimerkiksi taloudellisten päätösten tekoon. Tavoite on, että kaikki sopimusmaat raportoisivat tulevaisuudessa tason 3 mukaan. On huomattava, että tason 3 mukainen raportointikaan ei välttämättä ole tae tietojen luotettavuudesta.

Tason 3 mukainen tehtävä on erittäin vaativa, monen eurooppalaisen asian-tuntijan mielestä mahdoton, jos pidetään kiinni edes kohtalaisesta luotettavuus-vaatimuksesta ja erityisesti vertailtavuudesta. Euroopan tutkimuslaitokset eivät ole osanneet varautua raportointiin samalla pitkäjänteisellä tutkimus- ja kehitystyöllä kuin millä aikanaan kehitettiin metsävarojen inventointimenetelmiä eli valtakunnan metsien inventointeja (VMI). KHK-laskenta on toki monilta osin vaativampi tehtävä kuin VMI.

Vaativuutta lisäävät sekä mukana olevat KHK-hiilivarastot eli poolit, mukaan lukien maaperä, niiden vaikea mitattavuus ja hiilivarastojen suuri alueellinen vaihtelu sekä varastojen muutosten alueellinen ja ajallinenkin vaihtelu, ja maaperän hiilivarastojen ajallisen vaihtelun hitaus (esim. Ilvesniemi, 2026). Tuloksena on kirjava joukko eri maissa käytettäviä menetelmiä ja ei-vertailukelpoisia tuloksia. Metsien kasvihuonekaasuihin liittyviä kansainvälisiä tutkimusjulkaisuja on varovaisinkin arvion mukaan kymmeniä tuhansia, mutta ne eivät välttämättä muodosta kasvihuonekaasuinventaarion kannalta kokonaista järjestelmää edes yhden maan inventaarioon. Useimpien maiden laskentamenetelmät onkin koottu raportointivaatimusten pakottamina irrallisista, ei välttämättä toisiinsa sopivista palasista. Valtakunnan metsien inventointien tiedot toki täyttävät hyvän tutkimuksen kriteerit siltä osin kuin VMI:n tietoja käytetään.

Tämän kirjoituksen tarkoitus on kuvata muutamien esimerkkien avulla niitä vaikeuksia, joita liittyy kasvihuonekaasujen inventaarioihin niinkin edistyneessä metsien inventoinnin maassa kuin Suomessa. Tilanne on monin verroin vaikeampi maissa, joissa metsien inventointi tehdään määrävälein ja paljon harvemmin kuin Suomessa. Myös KHK-inventointeihin sisältyvät maaluokat vaihtelevat maiden välillä.

Professori Juha Heikkisen artikkeli tässä numerossa antaa jo hyvän kuvan inventaarion vaikeuksista mukaan lukien epävarmuuksien tilastotieteellinen arviointi.

Suomen KHK-inventaarion tietolähteistä ja menetelmistä

Suomen metsissä on noin 77 miljardia elävää, vähintään 1,3 metrin pituista puuta. Lyhyempiä puuta on arviolta 10–100-kertainen määrä riippuen siitä, lasketaanko mukaan esimerkiksi vaihtuva taimiaines vai ei. On huomattava, että elävien puiden lukumäärä vaihtuu jonkin verran lähes päivittäin hakkuiden ja luontaisen kuoleman sekä kasvukauden aikana myös kasvun mukaan.

Yhden valtakunnan metsien inventointikierron aikana eli viidessä vuodessa mitataan koko maassa noin miljoona todennäköisyysotannalla valittua elävää, vähintään 1,3 metrin pituista puuta. Näistä pieni osaotos, runsaat 60 000 puuta, myös todennäköisyysotannalla valittua puuta mitataan tarkemmin. Rinnankorkeusläpimitan (maanpinnan tasosta 1,3 metrin korkeudella olevan läpimitan), puulajin ja muutaman muun tunnuksen lisäksi näistä "koepuista" mitataan muun muassa puun pituus ja elävän latvuksen pituus. Suoraviivaisesti ennustettava puukohtainen tunnus on puun runkotilavuus, jolle on kehitetty puulaji-kohtaiset mallit (Laasasenaho, 1982; Kangas, Pitkänen ym., 2022).

KHK-inventaario edellyttää biomassan ja sen muutoksen arviointia myös muiden puuositteiden kuin runkopuun osalta. Suomen nykyisessä järjestelmässä ne ennustetaan ensin malleilla koepuille. Samalla saadaan laskentaositteittaiset biomassakertoimet, jotka ovat kunkin biomassan ja runkotilavuuden suhteita. Näiden avulla ennustetaan biomassat ja sen muutokset muille puille kuin koe-

puille, niin sanotuille lukupuille (Repola ym., 2025; Heikkinen, 2026a; Heikkinen, 2026b). Ositteet ovat kuoreton runkotilavuus, kuori, elävät oksat, kuolleet oksat, neulas- ja lehdet, kanto, vähintään yhden cm:n paksuiset juuret ja ohuemmat juuret, mukaan lukien hienoitujuuret, jotka kuuluvat puuston karikkeeseen (Repola, 2008; Repola, 2009; Repola ym., 2025). Kariketuotanto ennustetaan biomassaositteista erikseen mallinnetuilla kertoimilla (Forsell ym., 2025; Lehtonen, Sievänen ym., 2004; Muukkonen ja Lehtonen, 2004; Muukkonen, 2005; Starr ym., 2005; Liski ym., 2006). Ojitettujen metsämaan soiden puuston kariketuotanto on erikseen kuvattu julkaisussa Ojanen ym. (2014).

Puustobiomassamallit biomassaositteittain ja biomassojen muuntokertoimet on kuvattu alkuperäisten artikkeleiden lisäksi (Repola, 2008; Repola, 2009; Repola ym., 2025) kansallisessa raportissa (Forsell ym., 2025; Luke, 2025b; Luke, 2025c; Luke, 2025a), samoin biomassan kasvun estimaattien laskenta. Biomassaositteiden ja muuntokertoimien epävarmuuden arvioinnissa otetaan huomioon VMI:n otantavirhe sekä biomassamallien parametriestimaattien kovarianssimatriisien estimaatit (Ståhl ym., 2014).

Biomassamittaukset ovat erittäin työläitä, minkä vuoksi näytteitä oli aikaisemmin mitattu vain pieni määrä. Esimerkiksi männyn biomassaositteet on mitattu koko maassa 44 metsiköstä, kuusen vastaavasti 34 ja koivun 24 metsiköstä. Puukohtaisista ositteista runkopuun biomassa on mitattu vastaavasti 626, 366 ja 127 puusta, jotkut vaativimmat ositteet vähän pienemmästä määrästä puita. Männyn oksakarikkeen ja kuusen neulas- ja oksakarikedatan mallit perustuivat 1988–1990 mitattuun aineistoon, ns. valtakunnalliseen puututkimusdataan (VAPU-data) (Korhonen ja Maltamo, 1990; Lehtonen, Sievänen ym., 2004; Muukkonen ja Lehtonen, 2004) ja kattoivat vain Suomen eteläosan. Etelä-Suomen VAPU-dataa käytettiin vain latvuksen muodon arviointiin ja oksakariketuotannon laskennassa. Siihen liittyvä latvuksen nousemisarvio laskettiin koko maan VMI-datasta. (Tarkkaan ottaen Etelä-Suomen VAPU-otos ei täyttänyt todennäköisyysotannon kriteereitä.) Uusimmissa malleissa mittauksia on lisätty huomattavasti. Tiedot on koottu erillisistä tutkimuksista. Kohdemetsiköt edustavat lähes koko Suomea ja eri kasvupaikkoja sekä kivennäismailla että soilla maan pohjoisinta osaa lukuun ottamatta (Repola ym., 2025). Biomassapuita on yhteensä 2482, joista mäntyjä 1328, kuusia 834 ja koivuja 320 kappaletta. Puiden runkotilavuudet on ennustettu Laasasenahon (1982) ja Kankaan ym. (2022) malleilla ja laskettu kunkin biomassaositteen ja runkotilavuuden suhteet. Näitä sovelletaan kaikkiin VMI:n lukupuihin.

Kivennäismaametsien aluskasvillisuuden biomassalle, mukaan lukien varvut ja sammaleet, on omat mallit (Mäkipää ja Heikkinen, 2003; Muukkonen, Mäkipää ym., 2006), samoin niiden kariketuotannolle (Liski ym., 2006). Orgaanisen aineksen hajoamista arvioidaan maahiilimallilla Yasso07 (Tuomi, Rasinmäki ym., 2011; Tuomi, Laiho ym., 2011), jossa käytetään lähtötietoina karikesyötettä sekä Ilmatieteen laitoksen lämpötila- ja sadanta-aineistoja.

Suomen kasvihuonekaasuinventaariossa (KHKI) karike, kuollut puu ja metsien maaperän orgaaninen hiili käsitellään yhtenä kokonaisuutena laskentamenetelmistä, kuten mm. Yasso07-maaperämallista johtuen. Raportointiluokkien yläjako on siis: 1) elävän puuston hiilidioksidi, 2) kivennäismaan metsien maaperän hiilidioksidi, 3) orgaanisten maiden maaperän hiilidioksidi, 4) orgaanisten maiden maaperän metaani, 5) orgaanisten maiden maaperän typpioksiduuli ja metsämaa yhteensä.

Ojitettujen soiden maaperän päästöjen ja poistojen arvioinnit poikkeavat jonkin verran kivennäismaiden vastaavista. Elävien puiden karikesyöte estimoi-

daan kuitenkin kuten kivennäismaiden tapauksessa. Aluskasvillisuuden karikesyöte arvioidaan regressiomalleilla (Ojanen ym., 2014). Hienojuurten karikesyöte on kuvattu Almin ym. (2023) julkaisussa. Turpeen, puiden hienojuurten ja aluskasvillisuuden karikkeen hajoamiseen (lukuun ottamatta hakkuutähteiden kariketta) käytetään Ojasen ym. (2014) johtamia regressiomalleja. Vastemuuttuja on 68 ”häiriöttömästä” suokohteesta mitatut KHK-kaasujen vuomittaukset, jotka edustavat 4,7 miljoonan hehtaarin ojitettuja metsämaan soita. Selittäjinä on karikesyötteen lisäksi kasvukauden lämpötila ja puuston pohjapinta-ala. Jälkimmäisen tarkoitus on ottaa huomioon myös suon vedenpinnan korkeus, joka vaikuttaa lämpötilan lisäksi turpeen hajoamisnopeuteen. Hakkuutähteiden ja luonnonpoistumapuiden karikkeen hajoaminen estimoidaan Yasso07-maaperämallilla.

Estimaattien epävarmuuden arviointi

Kivennäismaan maaperän hiilivarannon muutosten epävarmuutta on arvioitu Lehtosen ja Heikkisen (2016) esittämällä Monte Carlo menetelmällä, joka käyttää Yasso07-maaperämallin syöttötietoina edellä mainittuja kariketietoja sekä lämpötila- ja kosteustietoja ja tuottaa orgaanisen aineksen hajoamisesta vapautuvan hiilidioksidin Yasso07-mallin ennusteen mukaisesti. Lehtosen ja Heikkisen menetelmä ottaa huomioon syöttötietojen otantavirheen lisäksi mallien parametrien estimaattien epävarmuuden (ks. myös Heikkinen, 2026a; Heikkinen, 2026b).

Heikkisen (2026a, 2026b) esittämä uusi menetelmä ottaa huomioon lisäksi otannan epävarmuuden, joka aiheutuu siitä, että vastemuuttajat eivät ole mitattuja vaan ennustettu vaihtelua pienentävällä regressiomallilla. Se toteutetaan lisäämällä malliin parametrien estimointivirheiden lisäksi malliin mallin jäännösvirheet. Jäännösvirheet muodostuvat kuvioittaisen vaihtelun huomioon ottavasta satunnaiskomponenteista ja puukohtaisista jäännöksistä. Näiden estimaattoireiden multinormaalista jakaumista simuloidaan edellä mainitut virheet, jotka siis lisätään malliin ja simuloidaan uudet biomassaennusteet. Heikkinen analysoi myös jäännösvarianssin mukaisen satunnaisvirheen lisäämisen vaikutusta otantavirheisiin ja kokonaisvirheisiin. Vaikutus vaihtelee odotetusti biomassao-sitteittain ja on pieni runkomassan osalta, mutta suuri muiden ositteiden osalta, kuten voidaan odottaakin. Heikkisen esittämä uusi menetelmä edustaa merkittävää metodologista edistysaskelta kivennäismaiden maaperään liittyvien KHK-taseiden epävarmuuksien kvantifioinnissa ja arvioinnissa. Menetelmä tarkentaa analyysien luotettavuutta verrattuna aikaisemmin käytössä olleeseen lähestymistapaan. Tieteellisen uutuusarvonsa ja käytännöllisen merkityksensä vuoksi menetelmä ansaitsee tulla julkaistuksi kansainvälisissä vertaisarvioituissa julkaisusarjoissa.

Alm ym. (2023) kuvaa ojitettujen soiden maaperän KHK-taseiden laskennan menetelmän yksityiskohdat luotettavuusarvioineen sekä mallin rajoitukset ja kehittämistarpeet.

Kuten em. julkaisuissa ja Heikkinen (2026a, 2026b) on esitetty, näinkään edistyneet tilastotieteen menetelmät epävarmuusanalyyseineen eivät voi ottaa huomioon malleissa käytetyn datan alueellisen ja/tai ajallisen vaihtelun puutteita. Toinen ongelma, myös Heikkisen (2026a, 2026b) julkaisussa mainittu, on mahdolliset mallien rakenteelliset ongelmat, esimerkiksi soiden vedenpinnan korkeuden vaihtelun huomioon ottaminen malleissa.

Datan saatavuudesta riippumaton tarkennus voitaneen tehdä kuolleen puun

tuottamassa karikkeessa ja kuolleen puun varastomuutoksessa. Kuollut puu otetaan huomioon vain kariketuotossa ja siitäkin VMI-inventointikierrosten välissä VMI-koealoilta mitattu luonnonpoistumapuu. Pysty- ja maapuita ei erotella, vaikka niiden hajoamisnopeuksissa on eroja. Lisäksi kuolleen puun varastomuutokset vaikuttavat joko päästönä tai poistona ja jäävät nykyisessä inventoinnissa vaille huomiota (Kangas, Heikkinen ym., 2026). Olisi siten perusteltua arvioida, missä määrin VMI-koealoilta mitattua kuolleen puun määrää sekä siihen liittyviä rakenteellisia ja ekologisia muuttujia voitaisiin hyödyntää nykyistä tehokkaammin kasvihuonekaasutaseiden laskennassa. Tällainen kehitystyö parantaisi laskennan kattavuutta ja vähentäisi siihen liittyvää rakenteellista epävarmuutta. Tietääkseni kehitystyö on meneillään Luonnonvarakeskuksessa.

Nykyiset KHK-seurantajärjestelmät osoittavat kuitenkin metsien osalta maankäyttösektorin KHK-taseiden muutosten suunnan melko luotettavasti (Heikkinen, 2026a; Heikkinen, 2026b).

Loppusanat

Metsien kasvihuonekaasutaseiden laskenta on erittäin vaativa tehtävä. Vaatimusten lisääntyttyä ja tarvittavien maastomittausten puuttuessa on jouduttu turvautumaan erillisissä tutkimuksissa kerättyihin, joskus pieniin aineistoihin, joista osa on kerätty muihin tarkoituksiin kuin KHK-taseiden seurantaan.

Näin mittavan järjestelmän suunnitteluun ja kehittämiseen pitäisi varata riittävästi aikaa ja resursseja etenkin, jos tuloksia halutaan käyttää ohjaamaan päätöksentekoa ja jos tuloksilla on taloudellisia vaikutuksia. Suunnittelun lähtökohtana pitäisi olla tavoitellut tarkkuusvaatimukset ja käytettävissä olevat resurssit myös jatkuviin mittauksiin ja tulosten analysointiin. Niiden pohjalta olisi suunniteltava otanta-asetelmat ja mittaukset. Tilastollinen otannan suunnittelu edellyttäisi tietoa mitattavien muuttujien alueellisesta ja ajallisesta vaihtelusta, jota tietoa ei toistaiseksi ole riittävästi olemassa. Tietojen lisääntyessä mittauksia voitaisiin tehostaa, kuten on tehty metsien inventointien tapauksissa.

Tällä hetkellä eikä lähitulevaisuudessa ole kaukokartoitusinstrumentteja tarvittavien muuttujien mittaamiseen eikä ilmakehän kaasuja mittaavien instrumenttien havaintoja pystytäkään kohdentamaan raportoinneissa tarkoitetuille alueille. Luotettavasti voidaan siten sanoa, että maastomittauksia tarvitaan lisää ja erityisesti kariketuotannon ja hajoamisen mallinnukseen sekä ojitettujen soiden turpeen hajoamisen mallinnukseen, jos halutaan jatkaa mallipohjaista menetelmää. Se koskee elävän puuston, kuolleen puuston, hakkuupoistuman ja luonnon poistuman kariketta ja lisäksi aluskasvillisuuden tuottamaa kariketta. Mittauksia tulisi tehdä käytettävien metsä- ja maaperämuuttujien aikamuutosten vuoksi vähintään määrävälein, ellei jatkuvasti jollakin intensiteetillä, kuten VMI-mittauksia tehdään.

Tilanne on monin verroin vaikeampi maissa, joissa metsien inventointi tehdään määrävälein ja paljon harvemmin kuin Suomessa, yhtenä esimerkkinä Saksa. Myös inventointeihin sisältyvät maaluokat vaihtelevat maiden välillä. Ongelmaksi jää siten tulosten kansainvälinen vertailtavuus.

Viitteet

Alm, J. ym. (2023). "A new method for estimating carbon dioxide emissions from drained peatland forest soils for the greenhouse gas inventory of Finland". *Biogeosciences* 20.18, s. 3827–3855. DOI: [10.5194/bg-20-3827-2023](https://doi.org/10.5194/bg-20-3827-2023).

- Forsell, P. ym. (2025). *Greenhouse gas emissions in Finland 1990 to 2023 – National inventory report under the UNFCCC and Paris Agreement*. Tekninen raportti. Statistics Finland. URL: https://stat.fi/media/uploads/tup/khkinv/fi_nid_un_2023_2025-04-15.pdf.
- Heikkinen, J. (2026a). ”Maaperän hiilitaseiden epävarmuusanalyysi”. Teoksessa: *Hiilinielulaskennan epävarmuusanalyysi*. Toim. A. Kangas ym. Luonnonvara- ja biotalouden tutkimus 6/2026. Helsinki: Luonnonvarakeskus, s. 67.
- (2026b). ”Suomen kivennäismaametsien hiilitaseen epävarmuuden arviointi”. *Tilastoseuran Vuosikirja 2025*.
- Iivesniemi, H. (2026). ”Maaperämallit ja –mittaukset”. Teoksessa: *Hiilinielulaskennan epävarmuusanalyysi*. Toim. A. Kangas ym. Luonnonvara- ja biotalouden tutkimus 6/2026. Helsinki: Luonnonvarakeskus.
- IPCC (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Toim. H. S. Eggleston ym. IGES, Japan. URL: <https://www.ipcc-nggip.iges.or.jp/public/2006gl/>.
- Kangas, A., Heikkinen, J. ym. (2026). *Hiilinielulaskennan epävarmuusanalyysi*. 6/2026. Helsinki: Luonnonvarakeskus, s. 67.
- Kangas, A., Pitkänen, T. P. ym. (2022). ”Mixed linear and non-linear tree volume models with regional parameters to main tree species in Finland”. *Forestry: An International Journal of Forest Research* 96.2, s. 188–206. DOI: [10.1093/forestry/cpac038](https://doi.org/10.1093/forestry/cpac038).
- Korhonen, K. T. ja Maltamo, M. (1990). *Männyn maanpäällisten osien kuivamassat Etelä-Suomessa*. Tekninen raportti 371. Metsäntutkimuslaitos, s. 1–29.
- Laasasenaho, J. (1982). ”Taper curve and volume functions for pine, spruce and birch”. *Communicationes Instituti Forestalia Fennica* 108, s. 1–74.
- Lehtonen, A. ja Heikkinen, J. (2016). ”Uncertainty of upland soil carbon sink estimate for Finland”. *Canadian Journal of Forest Research* 46.3, s. 310–322. DOI: [10.1139/cjfr-2015-0171](https://doi.org/10.1139/cjfr-2015-0171).
- Lehtonen, A., Sievänen, R. ym. (2004). ”Potential litterfall of Scots pine branches in southern Finland”. *Ecological Modelling* 180, s. 305–315.
- Liski, J. ym. (2006). ”Carbon accumulation in Finland’s forests 1922–2004 – an estimate obtained by combination of forest inventory data with modelling of biomass, litter and soil”. *Annals of Forest Science* 63, s. 687–697.
- Luke (2025a). *Hakkuukertymä ja puuston poistuma*. <https://www.luke.fi/fi/tilastot/hakkuukertyma-ja-puuston-poistuma>.
- (2025b). *Kasvihuonekaasuinventaario 2023: maataloussektorin ja maankäyttösektorin lopulliset tulokset hyvin lähellä ennakkotuloksia*. <https://www.luke.fi/fi/uutiset/kasvihuonekaasuinventaario-2023-maataloussektorin-ja-maankayttosektorin-lopulliset-tulokset-hyvin-lahella-ennakkotuloksia>.
- (2025c). *Kasvihuonekaasuinventaarion epävarmuudet*. <https://www.luke.fi/fi/luonnonvaratieto/tiedetta-ja-tietoa-luonnonvaroista-luonnonvaroista/kasvihuonekaasuinventaario/yleista-kasvihuonekaasuinventariosta/kasvihuonekaasuinventaarion-epavarmuudet>.
- Muukkonen, P. (2005). ”Needle biomass turnover rates of Scots pine (*Pinus sylvestris* L.) derived from the needle-shed dynamics”. *Trees* 19, s. 273–279. DOI: [10.1007/s00468-004-0381-4](https://doi.org/10.1007/s00468-004-0381-4).
- Muukkonen, P. ja Lehtonen, A. (2004). ”Needle and branch biomass turnover rates of Norway spruce (*Picea abies*)”. *Canadian Journal of Forest Research* 34, s. 2517–2527.

- Muukkonen, P., Mäkipää, R. ym. (2006). "Relationship between biomass and percentage cover in understory vegetation of boreal coniferous forests". *Silva Fennica* 40.2, s. 231–245.
- Mäkipää, R. ja Heikkinen, J. (2003). "Large-scale changes in abundance of terricolous bryophytes and macrolichens in Finland". *Journal of Vegetation Science* 14, s. 497–508.
- Ojanen, P. ym. (2014). "Soil CO₂ balance and its uncertainty in forestry-drained peatlands in Finland". *Forest Ecology and Management* 325, s. 60–73. ISSN: 0378-1127.
- Repola, J. (2008). "Biomass equations for birch in Finland". *Silva Fennica* 42.4, s. 236. DOI: [10.14214/sf.236](https://doi.org/10.14214/sf.236).
- (2009). "Biomass equations for Scots pine and Norway spruce in Finland". *Silva Fennica* 43.4, s. 625–647. URL: <http://www.metla.fi/silvafennica/full/sf43/sf434625.pdf>.
- Repola, J. ym. (2025). "Tree level biomass models for Finland". Forestry: An International Journal of Forest Research (submitted manuscript).
- Starr, M. ym. (2005). "Models of litterfall production for Scots pine (*Pinus sylvestris* L.) in Finland using stand, site and climate factors". *Forest Ecology and Management* 205, s. 215–225.
- Ståhl, G. ym. (2014). "Sample-based estimation of greenhouse gas emissions from forests - a new approach to account for both sampling and model errors". *Forest Science* 60.1, s. 3–13. DOI: [10.5849/forsci.13-005](https://doi.org/10.5849/forsci.13-005).
- Tuomi, M., Laiho, R. ym. (2011). "Wood decomposition model for boreal forests". *Ecological Modelling* 222.3, s. 709–718.
- Tuomi, M., Rasinmäki, J. ym. (2011). "Soil carbon model Yasso07 graphical user interface". *Environmental Modelling & Software* 26.11, s. 1358–1362. DOI: [10.1016/j.envsoft.2011.05.009](https://doi.org/10.1016/j.envsoft.2011.05.009).

Vastine: Suomen kivennäismaametsien hiilitaseen epävarmuuden arviointi

JUHA HEIKKINEN

JUHA.HEIKKINEN@LUKE.FI

LUONNONVARAKESKUS (LUKE)

Kiitokset Elja Arjakselle ja Erkki Tompolle mainioista kommenttikirjoituksista, jotka ansiokkaasti ja asiantuntevasti laajensivat aihepiirin käsittelyä omiin suuntiinsa. Molemmat kirjoitukset sisälsivät ehdotuksia, joita tullaan hyödyntämään kasvihuonekaasuinventaarion kehittämistyössä.

Tilastopäivät 9.-10.6.2025

Ryväsmuuttujat kausaalipäättelyssä: silta potentiaalisten vasteiden ja graafisten mallien välillä

JUHA KARVANEN

JUHA.T.KARVANEN@JYU.FI

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

JYVÄSKYLÄN YLIOPISTO

Tiivistelmä

Kirjoitus on yhteenveto Tilastopäivien 2025 esitelmästä ”Clustered variables in causal inference: Building on potential outcomes and graphical models”.

Kausaalipäättelyn koulukunnat

Syy-seuraussuhteita koskevat oletukset voidaan esittää joko graafisten mallien (Pearl, 1995; Pearl, 2009) tai potentiaalisten vasteiden (Neyman, 1923; Rubin, 1974; Holland, 1986) avulla. Yleensä nämä lähestymistavat johtavat samoihin tuloksiin, mutta joissakin tapauksissa toinen vaihtoehtoista voi olla toista käytännöllisempi (Gorbach ym., 2025). Käytettäessä graafisia malleja, syy-seuraussuhteet määritetään parametrittömästi muuttujakohtaisesti (esimerkiksi särmä $X \rightarrow Y$ kertoo, että muuttuja X vaikuttaa muuttujaan Y), kun taas potentiaalisten vasteiden lähestymistavassa oletukset esitetään ehdollisina riippumattomuuksina, jotka voivat koskea myös muuttujajoukkoja (esimerkiksi käsittely X ja potentiaalinen vaste Y^X voivat olla riippumattomia, kun ehdollistetaan sekoittajajoukolla $\mathbf{Z}$).

Ryväsmuuttujat

Muuttujien ryvästäminen (clustering) luo sillan graafien ja muuttujajoukkojen välille. Ryvästämistä voi hyödyntää kahdesta erilaisesta lähtökohdasta: Suurta graafia voi tiivistää esittämällä muuttujajoukon yhtenä solmuna, joka kuvaa kaikkia joukon muuttujia. Toisaalta mallinnus voi lähteä liikkeelle pienestä graafista, jossa kausaalisuhteet on määritelty vain muuttujajoukkojen välille, joita voivat olla esimerkiksi geneettiset tekijät, työhistoria, sosioekonominen asema ja terveys. Tällaisen ylhäältä-alas-mallinnuksen (top-down modeling) etuna on se, että rypäiden sisäisten kausaalisuhteiden määrittely ei ole aina välttämätöntä, ja graafisissa malleissa voidaan näin ollen käsitellä muuttujajoukkoja samaan tapaan kuin potentiaalisten vasteiden lähestymistavassa.

Kausaalipäätelyn kannalta on keskeistä, että ryvästys ei saa muuttaa kausaalivaikutusten identifioitu-
vuutta eli sitä, onko kausaalivaikutus selvitettävissä tehtyjen oletusten perusteella vai ei. Ryvästys on identifioitu-
vuusinvariantti, jos kausaalivaikutus on identifioitavissa ryvästetystä graafista silloin ja vain silloin, kun se on identifioitavissa alkuperäisestä graafista.

Teoreettisia tuloksia identifioitu-
vuusinvarianttisuudesta on johdettu siirtymärypäälle (transit cluster), jossa informaatio kulkee saapumissolmuista (receivers) poistumissolmuihin (emitters) (Tikka, Helske ja Karvanen, 2023; Tabell, Tikka ja Karvanen, 2025). Esimerkkejä siirtymärypäistä on esitetty kuvassa 2. Siirtymärypäitä voidaan käyttää

1. graafin tiivistämiseen,
2. identifioituvuusalgoritmien nopeuttamiseen,
3. identifioituvuusalgoritmien tuottamien lausekkeiden sieventämiseen,
4. kausaalivaikutusten estimoinnin yksinkertaistamiseen ja
5. perusteluna ylhäältä-alas-mallinnukselle.

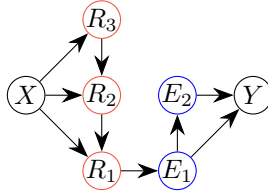
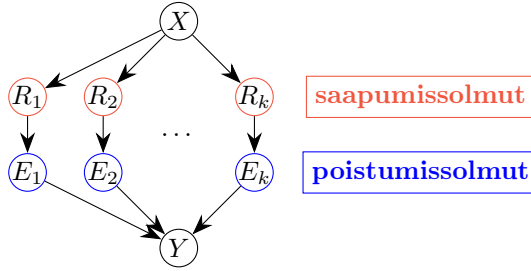
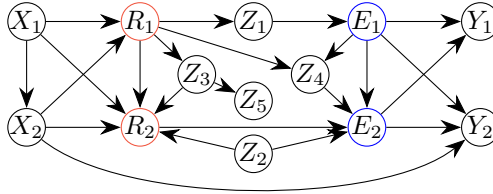
Muuttujien ryvästämiseen liittyvä tutkimus jatkuu edelleen.

Viitteet

Gorbach, T. ym. (2025). "Complementary strengths of the Neyman-Rubin and graphical causal frameworks". *arXiv:2512.09130*.



Kuva 1: Juha Karvanen.

(a) Siirtymäryväs $\{R_1, R_2, R_3, E_1, E_2\}$.(b) Siirtymäryväs $\{R_1, R_2, \dots, R_k, E_1, E_2, \dots, E_k\}$.(c) Siirtymäryväs $\{R_1, R_2, E_1, E_2, Z_1, Z_2, Z_3, Z_4, Z_5\}$.

Kuva 2: Esimerkkejä siirtymärypäistä (Tikka, Helske ja Karvanen, 2023).

Holland, P. W. (1986). "Statistics and causal inference". *Journal of the American statistical Association* 81.396, s. 945–960.

Neyman, J. (1923). "Sur les applications de la théorie des probabilités aux expériences agricoles: Essai des principes". *Roczniki Nauk Rolniczych* 10.1. English translation by D. Dabrowska and T. Speed in *Statistical Science* 5 463 - 472, 1990., s. 1–51.

Pearl, J. (1995). "Causal diagrams for empirical research". *Biometrika* 82.4, s. 669–688. DOI: [10.1093/biomet/82.4.669](https://doi.org/10.1093/biomet/82.4.669).

— (2009). *Causality: Models, Reasoning, and Inference*. 2nd. Cambridge University Press.

Rubin, D. B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66.5, s. 688.

Tabell, O., Tikka, S. ja Karvanen, J. (2025). "Clustering and pruning in causal data fusion". *arXiv:2505.15215*.

Tikka, S., Helske, J. ja Karvanen, J. (2023). "Clustering and structural robustness in causal diagrams". *Journal of Machine Learning Research* 24.195, s. 1–32.



Kuva 1: Jaakko Pere.

On the Impact of Approximation Errors on Extreme Quantile Estimation with Applications to Functional Data Analysis

JAAKKO PERE

JAAKKO.PERE@AALTO.FI

KAUPPAKORKEAKOULU, AALTO-YLIOPISTO

Consider estimation of an extreme quantile, that is, a $(1 - p)$ -quantile corresponding to a small p . In practice, the probability p is usually so small that extrapolation outside the sample is required. Estimation of such quantiles is crucial in many risk assessment scenarios, as realized extreme events can have calamitous consequences.

The field of extreme value theory provides tools for such extrapolation, for example, consider the extreme quantile estimator introduced by Weissman (1978). The asymptotics of Weismann-type extreme quantile estimators are well studied under the true observations, e.g., see Haan ja Ferreira (2006, Chapter 4). However, in many multivariate and infinite-dimensional extreme value theory settings, only approximated observations are available (Virta ym., 2024; Kim ja Kokoszka, 2019).

In the presentation, we consider a Weissmann-type extreme quantile estimator computed with approximated heavy-tailed observations. We give sufficient conditions for the approximation error such that the standard asymptotic results hold. Namely, given the true i.i.d. observations $Y_1, \dots, Y_n$ and the corresponding approximations $\hat{Y}_1, \dots, \hat{Y}_n$, it is sufficient that the maximum approximation error $\max_i |Y_i - \hat{Y}_i|$ grows sufficiently slowly with respect to the tail behavior of

the underlying distribution.

The developed general setting under approximations is applied in the context of functional data. That is, we estimate an extreme quantile corresponding to the L^p -norm $\|X\|_p$ of a random function X , when only approximations of the norms $Y_i = \|X_i\|_p$ are available. In the context of the functional data, the approximation error arises from the fact that, in practice, instead of the true sample paths $X_1, \dots, X_n$, only discretizations are observed. In the case of Riemann sum approximated norms $\hat{Y}_i$, we present the general sufficient condition for the approximation error in terms of continuity of the underlying random function X , heavy-tailedness of the true norm $\|X\|_p$, and density of the observed discretizations.

Bibliography

- Haan, L. de ja Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. 1. painos. Springer Series in Operations Research and Financial Engineering. New York: Springer. DOI: <https://doi.org/10.1007/0-387-34471-3>.
- Kim, M. ja Kokoszka, P. (2019). "Hill estimator of projections of functional data on principal components". *Statistics* 53.4, s. 699–720. DOI: <https://doi.org/10.1080/02331888.2019.1609476>.
- Weissman, I. (1978). "Estimation of Parameters and Large Quantiles Based on the k Largest Observations". *Journal of the American Statistical Association* 73.364, s. 812–815. DOI: <https://doi.org/10.1080/01621459.1978.10480104>.
- Virta, J. ym. (2024). "Latent model extreme value index estimation". *Journal of Multivariate Analysis* 202, s. 105300. DOI: <https://doi.org/10.1016/j.jmva.2024.105300>.



Kuva 1: Aapeli Nevala.

Hyvät priorit ja suolaa päälle — mitä voimme oppia terveysrekisteriaineistosta Markovin piilomalleilla?

AAPELI NEVALA

AAPELI.NEVALA@CANCER.FI

SYÖPÄREKISTERI, HELSINGIN YLIOPISTO

Eräässä tilastotieteen alan iltapäiväkahviseminaarissa pienimuotoisen kohahduksen yleisössä aiheutti esitys otsikolla: “Data ei kiinnosta”. Ajatus provosoivan otsikon taustalla oli, että riittävän laadukkaalla informaatiolla “a-priori”, pelkkä mallin kalibrointi osalla ideaalista datasta saattaa tietyissä tapauksissa riittää, eli laaja lisätiedonkeruu ei ole välttämätöntä. Tämä esitys inspiroi allekirjoittaneen kuuntelijan bayesilaisten Markovin piilomallien sovellukseen syöpäseulonassa Nevala ym., 2024.

Terveysrekisteriaineistojen tilastollisissa analyysissä törmätään usein joukkoon aineiston luonteesta johtuvia ongelmia. Havaintopisteet voivat olla väärin luokiteltuja, tai varsinainen kiinnostuksen kohde kuten oireettoman taudin olemassaolo puuttuu aineistosta kokonaan. Hyvä esimerkki tällaisesta prosessista on syöpäseulonta, jossa jo mainittujen ongelmien lisäksi aineiston taustalla vaikuttaa erilaiset havaintoprosessit. Syöpäseulontaa kuvaavat datapisteet havaitaan osana ennalta päätettyä — syövän olemassaolosta riippumatonta — seulontaprosessia. Seulontatestien ulkopuolella löydetty syöpä taas havaitaan oireiden takia, jolloin toteamisajat antavat informaatiota itse taudista. Lisäksi seulontaan ei tule kukaan kotoa hakemaan, vaan osallistuvien joukko on vahvasti valikoitunut siten, että osallistumattomilla syövän taustailmaantuvuus on

tyypillisesti korkeampi.

Markovin piilomallissa (ks. esim. Cook ja Lawless, 2018) mallinnetaan piilevää tilaa $X(t)$ havaitun aineiston $Y_{t_1:t_k}$ avulla. Suomen syöpäseulontaohjelmien kehittämistä varten rakensimme eri datalähteitä hyödyntävän mallin (2), jossa yhdistellään syöpäseulonnoista kertyvää dataa (epäinformatiiviset havaitsemisajat) ja syöpärekisteridataa (informatiiviset havaitsemisajat). Mallissa $X(t)$ oli viiden tilan malli: ei tautia (1), syövän esiaste (2-3), toteamaton syöpä (4) tai kliinisesti todettu syöpä (5). Tilasiirtymät olivat mahdollisia vain eteenpäin ja yhden tilan kerrallaan, eli kyseessä on progressiivinen malli, jossa siirtymäuhka $\lambda_{ij} \neq 0 \iff j = i + 1$. Tilat 1–4 havaittiin paneelidatana, paitsi tilasiirtymä $4 \rightarrow 5$ eksaktisti. Ajassa t seulonnan ulkopuolella todettu syöpä tarkoittaa, että juuri tätä hetkeä ajassa $t-$ $X(t)$ oli tilassa 4: toteamaton syöpä.

Malli hyödyntää elossaoloanalyysistä tuttua ajatusta. Seulonnassa käyvältä väestöosalta havaitsemme eri seulontaprosessin vaiheista havaittua dataa Y_{t_k} , jonka avulla voidaan estimoida luonnollisen kulun parametrit tavallisen Markovin piilomallin tapaan seulonnassa käyneille. Tilasiirtymä $4 \rightarrow 5$ mallinnetaan kertomalla siirtymän uhkafunktiota tilan (4) tilatodennäköisyysfunktion vasemmalla raja-arvolla: $\lambda_{45} \times p(X(t-) = 4)$. Hyödyntämällä tätä uskottavuusfunktioalaa ja tutkimuskirjallisuutta soveltamalla saatuja informatiivisia prioreja, voimme estimoida luonnollisen kulun parametrit myös seulontaan osallistumattomalle joukolle hyödyntämällä näiden syöpäilmaantuvuusdataa, eli tilasiirtymiä $4 \rightarrow 5$. Tämä askel on tärkeä, sillä osallistumattomien suuremman syöpäilmaantuvuuden takia myös esiasteiden ja oireettomien syöpien vallitsevuus on korkeampi.

Samalla metodiikalla mallia voisi hyödyntää myös kokonaan ilman piileviä tiloja koskevaa dataa muidenkin progressiivisesti etenevien tautien tutkimisessa, kunhan kirjallisuudesta löytyy riittävän luotettavaa informaatiota *a priori*. Jos tiedossa on tutkimuskirjallisuudesta esimerkiksi ruumiinavaustutkimuksiin perustuvia vallitsevuusestimaatteja ja luonnollisen kulun parametreja, voidaan tiedot kalibroida suomalaiseseen aineistoon käyttämällä kuvatulla tavalla todetun taudin ilmaantuvuutta — siis data ei ihan liikaa kiinnosta, paitsi se kalibroinnissa käytettävä todetun taudin data. Tällä tavoin malli voidaan laajentaa kuvaamaan myös piilevien tilojen vallitsevuuksia väestössä, jota ei seulontaan koskaan saada, tai taudeissa joiden piileviä tiloja ei Suomen väestössä tällä hetkellä systemaattisesti havaita.

Viitteet

- Cook, R. J. ja Lawless, J. F. (2018). *Multistate Models for the Analysis of Life History Data*. 1. painos. Chapman ja Hall/CRC. DOI: [10.1201/97813151119731](https://doi.org/10.1201/97813151119731). URL: <https://doi.org/10.1201/97813151119731>.
- Nevala, A. ym. (joulukuu 2024). "Bayesian hidden Markov model for natural history of colorectal cancer: Handling misclassified observations, varying observation schemes and unobserved data". *The Annals of Applied Statistics* 18.4, s. 3050–3070. DOI: [10.1214/24-AOAS1922](https://doi.org/10.1214/24-AOAS1922). URL: <https://doi.org/10.1214/24-AOAS1922>.

Tilastojulistekilpailu 2024–2025

REIJA HELENIUS
REIJA.HELENIUS@STAT.FI
ISLP DIRECTOR
RYHMÄPÄÄLLIKKÖ, TILASTOKESKUS

Tiivistelmä

Suomen tilastoseura ry järjesti yhdessä Tilastokeskuksen ja MAOL ry:n kanssa jo 8. kerran kansallisen nuorten tilastojulistekilpailun. Kilpailuun osallistui 500 nuorta.

Mistä tilastojulistekilpailussa on kyse?

Kansallinen tilastojulistekilpailu on osa Kansainvälistä tilastojulistekilpailua, joka järjestetään joka toinen vuosi International Statistical Literacy Projectin (ISLP) toimesta. Kilpailua on järjestetty vuodesta 2010 lähtien sekä kansainvälisesti ja kansallisesti. Tilastojulistekilpailu on oivaltava tapa tutustua tilastoihin ja tilastolliseen ajatteluun. Kilpailussa tutkitaan arjen kysymyksiä tilastoaineistojen avulla. Löydöksistä laaditaan tarina, joka kerrotaan posterissa. Kilpailuun voivat osallistua alakoulun ja yläkoulun oppilaat, toisen asteen (lukio, ammatikoulu) sekä alempaa korkeakoulututkintoa suorittavat (amk, yliopisto) opiskelijat.

Kilpailu käydään kahdessa osassa. Suomen kilpailun voittajat kustakin sarjasta jatkavat kansainväliseen kilpailuun. Kansallista kilpailua koordinoi Suomen Tilastoseura ry yhteistyökumppaneinaan Tilastokeskus ja MAOL ry. Kansainvälisestä kilpailusta vastaa ISLP-projekti eli International Statistical Literacy Project.

Suomessa sadat nuoret työستävät postereita kymmenissä joukkueissa. Kaikista kilpailukategorioista palkitaan kolme parasta joukkuetta. Sen jälkeen kunkin sarjan voittajatytöt lähtevät edustamaan Suomea kansainväliseen kilpailuun, jonka voittajat palkitaan Kansainvälisen tilastoinstituutin (ISI) WSC-kongressin yhteydessä (World Statistics Congress). Vuosien 2024–2025 kilpailuun osallistui maailmanlaajuisesti yli 23 000 nuorta lähes 30 maasta.

Mitä kilpailussa oppii?

Kilpailun tavoitteena on johdattaa nuoria tilastolliseen ja kriittiseen ajatteluun sekä opettaa tilastollinen tutkimusprosessi pähkinänkuoressa. Kilpailussa nuoret laativat tilastoposterin jostakin ajankohtaisesta, heille tärkeästä ilmiöstä. Kilpailu opettaa nuoria tilastolliseen ajatteluun tilastoposterin laatimisen kautta aina tutkimusongelman miettimisestä, tutkimusmenetelmän valinnasta, aineiston keruusta ja analyysistä sekä tulosten esittämisestä johtopäätösten tekoon. Tutkimus, sen vaiheet ja tulokset kuvataan tiivistetysti posteriin.

Tilastoposterin laatiminen voi toimia opetusmenetelmänä tilasto-opetuksen jaksoilla. Varsinkin Suomessa tämä on melko tuore opetustapa, joka tukee myös luovuutta ja opiskelijoiden itsenäistä ajattelua. Kilpailulla tuetaan tiimipohjaista työskentelyä ja oppilaiden eri vahvuuksien esille tuomista. Yksi oppilas voi olla lahjakas matemaattisesti, toinen visuaalisesti, kolmas tutkittavan ilmiön kautta. Tilastojulisteen laatiminen tukee myös ilmiöpohjaista oppimista ja postereiden tekoon perustuvaa opetusta voidaan hyödyntää yhteisprojektina eri opiaineiden kanssa (esim. matematiikka, historia, kuvaamataito, tietotekniikka, biologia). Kilpailulle ominaista on myös kriittisen lukutaidon edistäminen, luotettavien tietolähteiden tunnistaminen ja yhteiskunnallisen ja numeerisen tiedon hallinta ja ymmärtäminen. Näitä taitoja tarvitaan niin arki- kuin työelämässä. Tilastoposterin laatimisen osaamisesta on hyötyä myös tieteellisten tulosten raportoinnissa ja esittämisessä esimerkiksi eri konferensseissa, joihin opiskelijat saattavat osallistua myöhemmin elämässään.

Vuoden 2024–2025 kilpailun teemana oli ”Talous eri muodoissaan.” Kilpailun järjestämistä tuettiin Oikeusministeriön valtion avustuksella. Kilpailu oli osa Ymmärrä taloutta -hanketta sekä Kansallista talousosaamisen verkoston toimintaa. Verkosto vie eteenpäin Suomen kansallista talousosaamisen strategiaa, jonka tavoitteena on nostaa suomalaisten taloustaidot maailman huipputasolle vuoteen 2030 mennessä. Talousteemojen lisäksi postereissa sai olla myös muita aiheita.

Vuoden 2024–2025 kilpailun voittajat palkittiin juhlagaalassa

Vuoden 2024–2025 yläkoulu- ja toisen asteen oppilaitosten voittajat palkittiin Tilastokeskuksessa järjestetyssä juhlatilaisuudessa. Yliopistosarjasta ei tällä kertaa palkittu voittajaa ja alakoulusarjasta ei lähetetty arviointiin kilpailutoitää. Alakoulusarja oli ensimmäistä kertaa mukana kilpailussa.

Yläkoulu-sarjan voittajat

1. sija: [Kuka syö kasvisruokaa? \(png\)](#)
Koivukylän koulu, Vantaa. Ohjaava opettaja Eeli Tamminen.
2. sija: [Suomen talouskehitys 2014–2030 \(pdf\)](#)
Maatullin peruskoulu, Helsinki.
3. sija: [Kahvin hinta vaihtelee – kulutus säilyy \(png\)](#)
Nummenpakan koulu, Aurajoen yksikkö, Turku. Ohjaava opettaja Taija Peilimö.



Kuva 1: Palkintotilaisuuden osallistujien juhlahuumaa (kuva: Aki Harju).

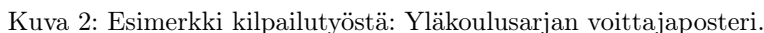
Lukioiden ja ammatillisten oppilaitosten voittajat

1. sija: [Taloudellinen ja psykososiaalinen hyvinvointi lukiolaisilla \(png\)](#)
Ressun lukio, Helsinki. Ohjaava opettaja Janne Karisto.
2. sija: [Talouden muutos kriisiaikoina / Miten finanssikriisi vaikutti Suomen talouteen? \(pdf\)](#)
Otaniemen lukio, Espoo. Ohjaava opettaja Matti Heikkinen.
3. sija: [Suomalaisten matkailu 2023 \(png\)](#)
Savon ammattiopisto, Kuopio. Ohjaava opettaja Miia Moilanen.

ISLP-projekti pähkinän kuoressa

ISLP (eli International Statistical Literacy Project), kansainvälisen tilastollisen luku- ja käyttötaidon edistämiprojekti, on osa ISI:a (International Statistical Institute) ja sen alajaostoa IASE:a (International Association for Statistical Education). ISLP-projektin toimintaan osallistutaan kaikista maanosista ja sillä on lähes 200 maavastaavaa. Toiminta perustuu pitkälti vapaaehtoistoimintaan yhteistyössä yhdistysten, yliopistojen, tutkimuslaitosten, tilastovirastojen ja oppilaitosten kanssa. Projektin tavoitteena on edistää tilastojen luku- ja käyttötaitoa kaikilla elämän osa-alueilla. Erityisessä keskiössä ovat nuoret ja lapset. Tilastojen luku- ja käyttötaidolla edistetään demokraattista yhteiskuntaa, jossa niin kansalaiset kuin päättäjätkin perustavat päätöksensä tietoon, ei olettamuksiin.

International Statistical Literacy Project (aiemmin World Numeracy Project) perustettiin vuonna 1994 ISI:n alaisuuteen tehtävänäään kvantitatiivisten taitojen edistäminen. Vuodesta 2010 projektia on johdettu Suomesta käsin. Projekti tuo omalta osaltaan suomalaista osaamista esille niin projektin kuin sen sidosryhmien, kuten Suomen kouluverkoston kautta. Suomen johtamalla kaudella on muodostettu maailmanlaajuinen maavastaavaverkosto, uudistettu web-sivusto, käynnistetty posterikilpailut, toimitettu uutislehteä ja tuotu aktiivisesti esille tilastojen luku- ja käyttötaidon edistämisen merkitystä niin konferenssie-



- ISLP Director, kehittämispäällikkö Reija Helenius, Tilastokeskus (Suomi)
- Deputy ISLP Director Prof. Pedro Campos, Statistics Portugal, University of Porto (Portugali)
- Deputy ISLP Director Prof. Adriana D'Amelio, Universidad Nacional de Cuyo (Argentiina)
- Deputy ISLP Director Prof. Irena Ograjenšek, University of Ljubljana (Slovenia).
- Deputy Directos Saleha Naghmi Habibullah, Kinnaird College for Women (Pakistan).

Lisätietoa [kilpailusta](#)
Lisätietoa [ISLP-projektista](#)

Vaarallinen tiede ja ammatti

PEKKA PERE

PEKKA.J.PERE@AALTO.FI

MATEMATIIKAN JA SYSTEEMIANALYYSIN LAITOS
AALTO-YLIOPISTO

Johdanto

Tilastot ovat niin tärkeitä, että ne voidaan kokea vaarallisina tai ne todella ovat vaarallisia – ainakin jollekulle. Ongelma on yleensä pohjimmiltaan, että tilastot paljastavat asioita, joista joku tai jotkut eivät tahtoisi kerrottavan. Joskus tilastot koetaan niin vaarallisiksi, että niitä halutaan vääristellä tai salata tai ne peräti salataan.

Osoitan monelle ehkä uuden näkökulman tilastotieteilijän työhön: yhteiskunnallisten ilmiöiden kuvaaminen tai tutkiminen tilastotieteen ja matematiikan keinoin voi olla vaarallista, jos yhteiskunnallinen ilmapiiri on tarpeeksi rajoittavan ideologian sävyttämä tai ilmapiiriä pyritään rajoittamaan. Myös tilastojen ja kyselytutkimusten laatijat, jotka luen tässä tilastotieteilijöiksi, voivat joutua paitsi ankaran kritiikin kohteiksi myös perättömästi syytetyiksi ja monenlaiseen vaaraan.

Mikä on vaarallista, on tärkeää. Julkisia virallisia tilastoja on kuvattu demokratian välttämättömäksi edellytykseksi ja tilastotieteilijöitä demokratian vahvimaksi tukipilariksi. Yksi ponnin artikkelille on osoittaa, kuinka tärkeää ja joskus vaarallista työtä tilastotieteilijät – muun muassa virallisessa tilastotomissa – tekevät ja kuinka altis modernikin yhteiskunta voi olla asiattomiin toimiin tilastotieteilijöitä vastaan. Tilaston laatijan painostaja ei ole aina vallanpitäjä, ja painostuksella voi joskus ajatella olevan vilpittömän hyviä tarkoituksia. Myös silloin on syytä kysyä, onko yhteiskunnan etu piilotella tilastoja.

Tilastojen manipuloinnille on esitetty ryhmitystä neljään luokkaan:¹

1. tilastojen vääristäminen (outright manipulation)
2. poliittisesti motivoituneet epäluotettavat tilastot (politically motivated guesstimating)
3. tarkoitushakuinen metodologisen sallivuuden hyödyntäminen (opportunistic use of methodology space)
4. indikaattoreiden epäsuora hallinta (indicator-management through indirect means).

Luokka 1 on merkitykseltään selvä ja täysin tuomittava. Luokka 2: Tilastoimella – esimerkiksi kehittyvässä maassa – ei aina välttämättä ole tarkkaa tietoa tilastoitavasta ilmiöstä. Tällöin tilastotoimi voi altistua poliittiselle paineelle julkaista toivottuun suuntaan kallellaan olevia “arvattuja” tilastoja. Luokka 3 on edellisiä hämyisempi: Kansainväliset suositukset tilastojen laatimisessa voivat sallia kansallisten erityispiirteiden huomioimisen tilaston laadinnassa. Se tarjoaa poliittiselle painostukselle mahdollisuuden rukata tilastoja toivottuun suuntaan ilman, että kansainvälisiä tilastokäytäntöjä rikottaisiin. Luokka 4 on hienovaraisin tilastojen manipuloinnin muoto: Idea on vaikuttaa tilaston laskuun käytettävään aineistoon. Tilanne vertautuu yrityksen tilinpäätökseen. Tilinpäätös saatetaan saada täysin laillisilla tavoilla näyttämään hyvin erilaiselta.

Esitän itse 5. luokan: Tilaston tekemättä tai julkaisematta jättäminen. Asiaa ei joko tahdota todentaa numeerisesti tai siitä jo laskettua tilastoa paljastaa.

Osoitan esimerkkejä tilastojen vaarallisuudesta (ainakin vallanpitäjien mielestä), mahdollisista tilastojen manipuloinneista (varmuutta manipuloinnista ei aina ole) ja vaaroista, joita totuudenmukaisia tilastoja ja kyselyitä laativat tilastotieteilijät saattavat kohdata. Useimmat lähteeni ovat tieteellisiä aikakauskirjoja tai tutkijoita, mutta käytän myös medialähteitä. Esimerkeistä voi tunnistaa piirteitä kaikista tilastojen manipuloinnin luokista 1–5.

Esimerkit ovat valtaosin ulkomailta tai historiastamme likimääräisessä aikajärjestyksessä. Aikajänne on ensimmäisistä virallisista tilastoista nykypäivään. Tilastotieteen ja tilastotieteilijän ammatin vaarallisuus ei rajoitu paikkaan eikä aikaan.

Esimerkit

Ruotsi ² Ruotsin taulustolaitos perustettiin 1748. Siitä tuli maailman ensimmäinen itsenäinen tilastoviranomainen. Ruotsin väestötilasto on niin ikään maailman vanhin. Ruotsi salasi 1749 aloittamansa ensimmäisen väestölaskentansa tiedot. Jos ulkovallat olisivat ymmärtäneet Ruotsin väkiluvun pienuuden, Ruotsin suurvaltauskottavuus olisi vaarantunut ja riski joutua hyökkäyksen kohteeksi kasvanut. □

Autonomian ajan Suomi I ³ Väinö Fritiof Johanson julkaisi 1940 yhden ensimmäistä suomenkielisistä tilastotieteen oppikirjoista. Hän arvioi siinä, että vallanpitäjä tappoi maamme luvuin kuvanneen ja tutkineen “yliopistotilaston”:

– “yliopistotilasto” lamaantui Suomen Venäjään liittymisen jälkeen. Maamme olot olivat sellaiset, ettei “valtion merkisyyksiä” voitu tieteen objektiivisuudella tutkia ja selvittää. Arvidssonin, Nordströmin ym. kohtalot osoittivat, että valtiollisten olojemme tutkijoille tuli vapaaehtoinen tai pakollinen maanpako ja karkotus työn palkaksi. Venäjän viranomaisten ja heidän kotimaisten apumiestensä harjoittamaan toimintaan kuoli tämä tieteenhaara maassamme.

Adolf Ivar Arwidsson (1791–1858) ja Johan Jacob Nordström (1801–1874) eivät edustaneet yliopistotilastoa vaan kohtaloa, jonka Johanson hahmotti Venäjän hallinnon kritisoijille. Suomalaismielinen Arwidsson erotettiin Turun Akatemiasta ja venäjäkriittistä Nordströmiä ilmeisesti uhkasi professorin virasta pidättäminen Aleksanterin yliopistosta. Molemmat päätyivät tekemään työuransa lopun Ruotsissa. □

Autonomian ajan Suomi II ⁴ Suomeen perustettiin 1865 tilastollinen virkakunta eli väliaikainen tilastollinen toimisto, joka vakinaistettiin 1870. Sen toisena päällikkönä toimi Karl Ferdinand Ignatius 1868–1885. Valtiotieteiden käsikirjan summeeraus Ignatiuksen merkityksestä tilastotoimelle on ylistävä:

Hän ryhtyi suurella tarmolla kehittämään maamme tilastollista tutkimusta, varsinkin virallista tilastoa, kohottaen tämän vähitellen samalle tasolle, kuin millä se oli länsimaiden sivistysvaltioissa. Hänen johdollaan järjestettiin tilastoaineistojen keräily ja kehitettiin tarvittavat käyttelymenetelmät – – .

Silti Ignatiuksen syytettiin tahallaan väärentäneen tilastoja. Lähteestä ei ilmene peruste tai onko Ignatiusta erityisesti painostettu asiassa. Syytöksen koivuus on itsessään painostusta. □

Neuvostoliitto I ⁵ Tilastotieteen käsitteitä ja teorioita kuten satunnaisvaihtelua ja suurten lukujen lakia voitiin pitää Neuvostoliiton kommunistisen puolueen ensimmäisen pääsihteerin Josif Stalinin ajan Neuvostoliitossa “porvarillisina” ja yhteiskunnallisesti sisällyksettöminä. Tilastotiedettä saatettiin soveltaa tekniikassa ja luonnontieteissä, mutta samojen menetelmien käyttäminen yhteiskunnan tutkimiseen “porvarillisten tilastotieteilijöiden” tapaan oli epäasianmukaista. Tällaisessa ajatusmaailmassa tilastotiedekin saatettiin mieltää vaaralliseksi.

Nikolai Kondratiev perusti 1920 vaikutusvaltaisen Moskovan suhdanneinstituutin ja johti sitä. Instituutti julkaisi tilastoja ja niiden analyyseja. Kondratiev erotettiin 1928, pidätettiin 1930 ja teloitettiin 1938. Hänen suuri virheensä oli ollut esittää, että pitkät syklit kuvaavat kapitalistisen yhteiskunnan kehitystä. Se oli ristiriidassa marxilaisen näkemyksen kanssa, jonka mukaan yhteiskunta eteni deterministisesti kohti kommunismia.

Eugen Slutsky teki urauurtavaa ekonometrista tutkimusta ja työskenteli muun muassa edellä mainitussa suhdanneinstituutissa. Slutskyn teoreettiset saavutukset ovat tuttuja taloustieteen, aikasarja-analyysin ja stokastiikan opiskelijoille. Poliittisen paineen takia Slutsky ja moni muu tutkija jättivät tilastotieteen. Slutsky siirtyi tutkimaan geofysiikkaa, tähtitiedettä ja matematiikkaa. □

Neuvostoliitto II ⁶ Olympiy Kvitkin johti Neuvostoliiton väestönlaskentaa 1937. Hänet teloitettiin samana vuonna, koska laskennan mukaan Neuvostoliiton väestö oli paljon pienempi kuin pääsihteeri Stalin katsoi hyväksi kertoa.

Edvard Gylling oli ensimmäisiä suomalaisia tilastotieteilijäksi tituleerattuja. Hän toimi tilastotieteen dosenttina 1910–1918 Aleksanterin yliopistossa ja kirjoitti tilastotieteestä artikkelin 1909–1922 julkaistuun Tietosanakirjaan. Gylling päätyi Neuvostoliittoon, jossa hän nousi Karjalan neuvostotasavallan kansankomissaarien neuvoston puheenjohtajaksi (pääministeriksi). Gylling julkaisi kollegojensa kanssa 1934 kartaston, jossa kartoilla ja monipuolisella tilastografiikalla havainnollistettiin muun muassa Leningradin (nykyisen Pietarin) alueen ja Karjalan autonomisen sosialistisen neuvostotasavallan kansallisuuksien sijoitumista ja määrää. Kartasto määrättiin pian tuhottavaksi. Gylling teloitettiin “kansanvihollisena” 1938.

Historioitsija Hannu Rautkallio arvioi, että Karjalan neuvostotasavallan väestötilastot “olivat varmastikin aitoja”, kun ne esiteltiin talvisodan 1939–1940 aikaiselle Neuvostoliiton tunnustamalle Terijoen nukkehallitukselle, “sillä tilastojen laatijat oli tapettu”. □

Iso-Britannia ⁷ Margaret Thatcher oli politiikan voimahahmo pääministerikaudellaan 1979–1990 Iso-Britanniassa. On esitetty, että Thatcherin hallitus ja virkamiehistö muuttivat tilastointitapoja, manipuloivat tilastoja poistamalla oudokkeja, harvensivat tilastojen julkaisua ja muutenkin viivästyttivät epämieluisien tilastojen julkistamista. □

Argentina, Venezuela ja Brasilia ^{1,8} Argentiinan arvioidaan keplotelleen tilastoidun inflaation 2007–2011 kolmasosaan todellisesta. The Billion Prices Projectin mukaan Argentiina peukaloi inflaatiotilastoja 2007–2016. Projekti oli Harvardin yliopiston ja Massachusettsin teknillisen korkeakoulun MIT:n (Massachusetts Institute of Technology) hanke, jossa arvioitiin inflaatiota lukuisissa maissa. Peukalointia edelsi Argentiinan tilastokeskuksen INDECin (Instituto Nacional de Estadística y Censos) kuluttajahintaindeksiä laatineen yksikön johtajan Graciela Bevacquan ja hänen alaistensa erottaminen. Argentiinan hallitus määräsi Bevacqualle ja hänen alaisilleen 150 000 dollarin edestä sakkoja epävirallisten hintaindeksien laatimisesta ja julkaisemisesta. Heitä vastaan nostettiin myös rikossyytteitä. Kansainvälisen tilastoinstituutin ISI:n (International Statistical Institute) mukaan argentiinalaisia tilastotieteilijöitä uhkasi yli 110 000 Yhdysvaltain dollarin kulut oikeusprosesseista, joihin heidät haastettiin. Yhdysvaltain tilastoseura (American Statistical Association, ASA) tuki aktiivisesti Bevacquaa ja hänen kollegoitaan. Presidentiksi 2016 valittu Mauricio Macri julisti Argentiinaan “tilastollisen hätätilan”. Hallinnon vaihduttua Bevacqua palasi töihin INDECiin mutta toimii nykyään itsenäisenä konsulttina.

Suomen kuvalehti välitti 19.10.2023 suomalaisille hyvin raadollisen kuvan Argentiinan tilastotieteilijöiden tilanteesta:

Vuonna 2015 tilanne oli jälleen kerran paha. [Kansainvälinen valuuttarahasto] IMF lainasi [Argentiinalle] rahaa ja takaisinmaksuerien suuruus oli sidottu virallisiin inflaatio- ja työttömyyslukuihin. Lukuja oli laskemassa maan tilastokeskuksessa Indecissä tutkija Julieta. – – “Olin tullut aamulla työpaikalleni ja vähän jälkeeni toimistoon saapui esimieheni. Hän pisti pistoolin pöydälle ja sanoi, että tilastot tehdään nyt minun tavallani, kertoo Julieta. “IMF:n tarkastajat olivat tulossa Buenos Airesiin ja meitä estettiin tapaamasta heitä. – – Emme saaneet poistua kahteen päivään toimistolta. Ovella seisivat aseistetut vartijat” – – . IMF:lle syötettiin alaspäin vääristeltyjä lukuja inflaatiosta, työttömyydestä ja köyhyydestä. Julieta erosi Indecin palveluksesta. “Olen taustaltani tutkija. En voinut hyväksyä huijausta.”

Nicolás Maduron vuodesta 2013 vuoteen 2026 autoritaarisesti johtaman Venezuelan talous ajautui hyperinflaatioon ja romahti 2010-luvun lopulla. Venezuela ei julkaissut inflaatiotilastoja 2017–2019. Venezuelan on arvioitu vääristelleen tilastoja koronaviruksen aiheuttamista kuolemista 2020–2021. Venezuelalaiset lääkärit ja sairaanhoitajat keräsivät siksi itse kuolintilastoja. Lääkäreiden kollektiivi raportoi ajoittain suurempia viikottaisia kuolleiden lukumääriä terveydenhoitohenkilökunnan keskuudessa kuin Venezuela koko maasta.

Brasilian presidentti Dilma Rousseffin hallinto pyrki 2012–2015 saamaan valtion julkisen velan näyttämään todellista pienemmältä monin eri keinoin. Toimiiin ryhtymisen taustalla oli muun muassa edessä olleet vaalit 2014. Rousseff asetettiin virkasyytteeseen, todettiin rikkoneen budjettilakeja ja erotettiin virastaan 2016. □

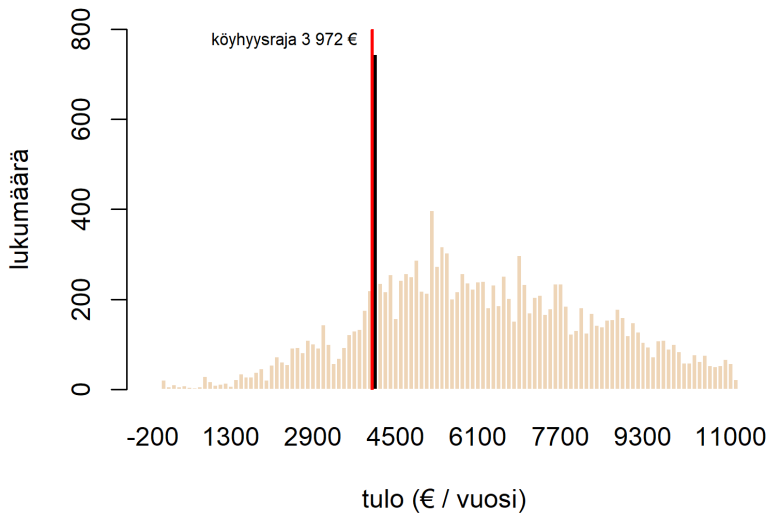
Kreikka ⁹ Euroopan unionin tilastotoimisto Eurostat raportoi tammikuussa 2010, että Kreikan tilastoviranomaiset olivat vuosia vääristelleet tilastoja Kreikan julkisesta velasta ja budjetin alijäämästä. Elokuussa 2010 Kreikan tilasto-

keskuksen ELSTATin presidentiksi nimitettiin Andreas Georgiou. Hän toimitti lokakuussa 2010 Eurostatille aiempia huomattavasti suuremmat tiedot alijäämästä. Eurostat arvioi marraskuussa 2010 uudet tilastot luotettaviksi — kuten monet muut asiantuntijatahot myöhemmin. Syyskuussa 2011 Georgiou asetettiin Kreikassa syytteeseen liioiteltujen alijäämätietojen luovuttamisesta Eurostatille. Häntä uhkasi elinkautinen vankilatuomio. Georgiou joutui eroamaan tehtävästään 2015. Häntä syytettiin 171 miljardin euron vahingon aiheuttamisesta Kreikalle, ja hänet on tuomittu kahdeksi vuodeksi vankeuteen. Euroopan ihmisoikeustuomioistuimien katsoi 2023, että Georgiou ei ollut saanut oikeudenmukaista oikeudenkäyntiä Kreikassa. Oikeudenkäynnit Georgioun ja Kreikan välillä ilmeisesti jatkuvat. Lukuisat tilastotieteelliset yhdistykset ja toimijat — muun muassa Yhdysvaltain tilastoseura ASA, Kansainvälinen tilastoinstituutti (International Statistical Institute, ISI), Tilastoseurojen presidenttien toimikunta (Committee of Presidents of Statistical Societies) sekä Suomen Tilastoseura — ovat vaatineet, että oikeudenkäynnit Georgiouta vastaan tulee lopettaa. Georgioun mukaan tilastotoimi on Kreikassa taistelulaji. □

Kiina ¹⁰ Yle Uutiset arvioi 28.9.2018, että “on yleisesti tiedossa, että Kiina on valmis manipuloimaan tilastojaan”. Helsingin Sanomat 20.7.2019: “Kiinan viralliset tilastot yliarvioivat maan talouskasvua, sanoo Nordean pääekonomisti ja Kiinan talouteen erikoistunut Tuuli Koivu. Hänen mukaansa maan viimeisimmät kasvuluvut ovat ihmetyttäneet kaikkia. — Kiinalla on aina ollut ongelmia tilastojen kanssa, hän lisää.” Helsingin Sanomien 15.8.2023 mukaan “Kiinan virallisella työttömyysluvulla ei tunnetusti ole tekemistä todellisuuden kanssa”. Kiinassa nuorisotyöttömyys oli kesällä 2023 ennätyksellisesti 21 %. Kiinan tilastoviranomaiset olivat juuri kertoneet keskeyttävänsä nuorisotyöttömyyden julkaisemisen ja ryhtyvänsä parantamaan tilaston keruumenetelmiä. Suomen kuvalehdessä 2.11.2025 Koivu kuvasi tapaamistaan Kiinan bruttokansantuotteen laskemisesta vastaavan virkamiehen kanssa: “Kysyin, että ovatko nämä tilastot luotettavia. — hän nauroi, että ‘eivät tietenkään, älä ole hölmö’. — lukuihin vaikutti myös maan johto — Hän kertoi siitä aivan avoimesti.” □

Suomi ¹¹ Etnisen ryhmän edustaja kritisoi 2019 Helsingin yliopiston Kriminologian ja oikeuspolitiikan instituutin julkaisemia tilastoja henkikirikollisuudesta. Henkikirikollisuus oli kyseisen etnisen ryhmän piirissä kuusitoistakertaista muihin suomalaisiin verrattuna. Ryhmän edustaja katsoi, että tilasto luo haitallisia ennakkoluuloja. □

Unkari ¹² Euroopan Unionissa köyhiksi määritellään kotitaloudet, joiden tulot ovat alle 60 % kotitalouksien mediaanituloista. Oppositio kritisoi Unkarin hallintoa lisääntyvästä köyhyydestä 2010-luvun alkupuoliskolla. Kritiikki vaimeni vuosikymmenen loppupuoliskolla tilastoidun köyhyyden vähetessä jyrkästi. Annamária Tátrai ja András Gábos osoittavat poikkeamia Unkarin tilastokeskuksen köyhyystilastoissa 2018–2021, jotka saattavat selittää tilastoidun köyhyyden vähenemistä. Heistä poikkeamat voivat johtua toistuvista metodologisista virheistä tilastoinnissa tai tilastojen manipuloinnista. Kuvassa 1 on unkarilaisten vuositulot (annual equalised household income) 90:nenteen prosentiiliin asti luokiteltuna 100 euron tarkkuudella 2021. Punainen pystysuora viiva osoittaa köyhyysrajan 2021 (3 972 euroa/vuosi). Köyhyysrajan niukasti ylittävä mustalla korostettu tuloluokka on erityisen suuri. Tátrain ja Gábosin mukaan talous-



Kuva 1: Kotitalouksien tulojen jakauma Unkarissa 2021, **punaisella merkitty** köhyysraja 3 972 euroa/vuosi ja **mustalla korostettu** köhyysrajaa seuraava tulo luokkapylväs.

politiikka ei selitä poikkeamaa eikä edeltävinä vuosina ole vastaavia poikkeamia tulojakaumassa. Vastauksessaan poikkeamista kertovaan uutiseen Unkarin tilastokeskus (Központi Statisztikai Hivatal) kiistää jyrkästi syytökset tahallista manipuloinnista ja kertoo voivansa laillisesti tutkia tilastojen ilkeästä tulkintaa, vihanlietsontaa ja laitoksen maineen vahingoittamista. □

Turkki ¹³ Turkin presidentti Recep Tayyip Erdoğan erotti Turkin tilastokeskuksen TurkStatin johtajan Sait Erdal Dincerin, kun TurkStat oli raportoinut vuoden 2021 inflaatiiovauhdiksi 36 prosenttia. Dincer oli aiemmin kertonut, että hän ei suostuisi peukaloimaan inflaatiotietoja, sillä hän on vastuussa 84 miljoonalle ihmiselle, joiden palkankorotukset riippuvat inflaatiiovauhdista. Yhdysvaltain tilastoseura ASA, Ison-Britannian kuninkaallinen tilastoseura RSS ja Kansainvälinen tilastoinstituutti ISI ovat arvostelleet erottamista ja vaatineet, että TurkStatin tilastotieteilijöiden tulee voida tuottaa objektiivista eikä vain hallituksen tarpeita palvelevaa tietoa. □

Venäjä ¹⁴ Venäjän tilastokeskus Rosstat lopetti kuukausittaisten syntymä- ja kuolintilastojen julkaisemisen keväällä 2025. Mahdollinen selitys on tarve piilotella vähäistä syntyvyyttä ja Ukrainan sodan suuria miestappioita. □

Yhdysvallat I ¹⁵ Yhdysvaltain presidentti Donald Trumpin uusi hallinto ryhtyi ensitöikseen 2025 karsimaan julkista hallintoa kovakätisesti. Hallituksen kauppaministeri ehdotti, että julkinen kulutus poistettaisiin Yhdysvaltain bruttokansantuotteen laskentakaavasta. Tällöin julkisen sektorin leikkaukset eivät

näkyisi bruttokansantuotteen pienenemisenä. Kansainvälinen tilastoinstituutti ISI reagoi kannanotolla: virallisia tilastoja tulee voida laatia ilman poliittista häirintää.

Trump erotti Työvoimatilastotoimiston (Bureau of Labor Statistics) johtajan Erika McEntarferin välittömästi, kun toimisto oli korjannut poikkeuksellisen paljon alaspäin aiempia ennakkotietoja työllisyydestä ja julkaissut uusia odotettua huonompia työllisyyslukuja 1.8.2025. Trumpin mukaan työllisyystilastoa oli toistuvasti peukaloitu häntä vastaan. Kansainvälinen tilastoinstituutti ISI, Ison-Britannian kuninkaallinen tilastoseura RSS sekä Yhdysvaltain tilastoseuran ASA:n toiminnanjohtaja Ronald Wasserstein huolestuivat McEntarferin erottamisesta ja välittömästi kannustivat tilastotoimen tilastotieteilijöitä vaalimaan edelleen Yhdistyneiden kansakuntien tilastokomission hyväksymiä virallisen tilaston periaatteita ja luottamusta tilastotoimeen. Wasserstein katsoi Yhdysvaltojen seuraavan Kreikan ja Argentiinan esimerkkiä tilastotoimen johtajien vainoamisessa. Suomen Tilastoseura yhtyi sisarseurojensa huoleen ja julkaisi pian samanhenkisen kannanoton McEntarferin erottamisesta.

Yhdysvallat II ¹⁶ J. Ann Selzerin vaalikyselyjä on kehuttu tarkkuudestaan. Selzer ennusti oikein Donald Trumpin presidentinvaalien 2016 ja 2020 voittajakaksi Iowan osavaltiossa Yhdysvalloissa. Juuri ennen presidentinvaaleja 2024 Selzer ennusti Trumpin saavan 44 % ja Kamala Harrisin 47 % äänistä Iowassa. Trump voitti Iowan vaalit murskaavalla 56 %:n ääniosuudella Harrisin kannatuksen jäädessä 43 %:iin. Trump sekä lakitoimisto Center for American Rights haastoivat Selzerin ja ennusteen julkaisseen lehden oikeuteen. Haasteiden mukaan ennuste oli julkeaa vaalihäirintää, vaalikyselyllä pyrittiin vaikuttamaan äänestäjiin ja lehti petti luotettavaa raportointia odottaneet tilaajansa.

New York Times julkaisi 22.1.2026 kyselytuloksia, jotka olivat epäsuotuisia Trumpille. Esimerkiksi kyselyn mukaan yhdysvaltalaisista asiat olivat maassa 32 %:n mielestä paremmin ja 49 %:n mielestä huonommin kuin vuotta aiemmin ja 40 % puolsi ja 56 % paheksui Trumpin tapaa toimia presidenttinä. Kysely perustui 1 625 äänestäjän vastauksiin. Raportoitu virhemarginaali oli 2.8 %-yksikköä. Lehti oli tavoitellut 86 000 äänestäjää, joten vastauskato oli huomattava. Trump ilmoitti kyselyn julkaisupäivänä, että kysely tullaan liittämään hänen aiempaan oikeushaasteeseen New York Timesia vastaan. Trump kirjoitti Truth Social -palvelussaan, että New York Times julkaisee tietoisesti vääriä kyselytuloksia ja että valheellisten kyselytulosten julkaisemisesta tulee tehdä rikollista. □

Yhdysvallat III ¹⁷ Yhdysvaltalainen Gallup-yhtiö ilmoitti 11.2.2026 lopettavansa kyselytutkimukset kansalaisten tuesta virassa olevalle presidentille (*approval rating*). Yhtiö oli arvioinut tukea vuodesta 1938 lähtien. Yhtiön viimeiseksi jääneessä 28.11.2025 julkaistussa kyselytutkimuksessa tuki Trumpille oli laskenut 36 %:iin eli heikoimmaksi hänen nykyisellä virkakaudellaan. Yhtiö kertoi lopettamispäätöksen perustuvan yksinomaan yhtiön tutkimuspäämääriin ja -prioriteetteihin ja yhtiön keskittyvän jatkossa asioiden ja politiikan kannatuksen arviointiin. □

Lopuksi

Tästä lähdettiin, ja tähän on tultu. Ennen ja nyt, meillä ja muualla, halutaan peitellä tai vääristää tilastoja. Tilastotieteilijä voi joutua haastetuksi oikeuteen.

Oikeudenkäyntien lopputuloksesta riippumatta haasteet oikeuteen voivat vaikuttaa tilastojen ja tutkimusten tekijöihin ja niiden julkaisijoihin. Tieteen ja ammatin vaarat eivät rajoitu oikeudenkäynteihin ja niistä mahdollisesti seuraaviin tuomioihin. Tilastotieteilijät saattavat päätyä myös muunlaisen painostuksen kohteeksi ja joutua vaarantamaan elämänsä monessa mielessä.

Yhteiskunnissa tulisi laajasti ymmärtää, että on yhteinen etu, että tilastot laaditaan riippumattomasti vallanpitäjien ja muiden tahojen yksipuolisista intresseistä ja että tilastotieteilijöille tulee turvata edellytykset muun muassa luotettavien tilastojen ja kyselyjen tekemiseen. Muuten vaarana ovat ”poliittisesti korrekti todennäköisyyslaskenta” ja ”poliittisesti korrekti tilastotiede”, jotka ovat totuuden vaarantavaa sopeutumista ideologian ja politiikan vaatimuksiin.¹⁸ Niiden seurauksia ilmeni jo edellä, ja lisää selviää perehtymällä kyseisten esimerkkien maiden historiaan ja kehitykseen.

Loppuviitteet

¹R. Aragão ja L. Linsi (2022) Many Shades of Wrong: What Governments Do When They Manipulate Statistics. *Review of International Political Economy*, 29, 88–113.

²G. Luther (1993) *Suomen tilastotoimen historia vuoteen 1970*. Tilastokeskus. WSOY. T. Teräsvirta (1987) How We Got the Data. Teoksessa T. Pukkila ja S. Puntanen (toim.) *Proceedings of the Second international Tampere Conference in Statistics*. Department of Mathematical Sciences/Statistics. University of Tampere.

³V. F. Johanson (1940) *Suomen tilaston pääkohdat*. Kirjapaino ABC. S:t 15–16. M. Klinge, R. Knapas, A. Leikola ja J. Strömberg (1989) *Keisarillinen Aleksanterin yliopisto 1808–1917*. Otava. S. 164. K. Tarkiainen (1997–) Arwidsson, Adolf Ivar. Kansallisbiografia-verkkajulkaisu. Studia Biographica 4. Helsinki: Suomalaisen Kirjallisuuden Seura. <https://kansallisbiografia.fi/kansallisbiografia/henkilo/3122> (haettu 19.1.2026).

⁴E. G. Palmén (1886) Kirje Helsingistä. *Valvoja*, 6, 98–103. L. Harmaja, A. E. Tudeer, N. Kärki, E. Böök, R. Erich, J. Forsman, R. Holsti, O. K. Kilpi ja A. Tulenheimo (1921) Ignatius, Kaarle Eemil Ferdinand. Teoksessa *Valtiotieteiden käsikirja. Ensimmäinen osa*. Tietosanakirja.

⁵M. Morgan (1990) *The History of Econometric Ideas*. CUP. S:t 66–67, 80. S. Kotz (1965) Statistical Terminology – Russian vs. English – in the Light of the Development of Statistics in the U.S.S.R. *American Statistician*, 19, 16–22. Esipuhe käännökseen O. Sheynin (2009) E. E. Slutsky. Theory of Correlation and Elements of the Doctrine of the Curves of Distribution. Manual for Studying Some Most Important Methods of Contemporary Statistics. http://www.probabilityandfinance.com/sheynin/038_slutsky.pdf. https://en.wikipedia.org/wiki/Nikolai_Kondratiev. (Haut 28.1.2026.)

⁶E. Fredrikson (1993) *Suomi 500 vuotta Euroopan kartalla*. Gummerus. S:t 114–115. M. Uola (2020) Gylling, Edvard. Kansallisbiografia-verkkajulkaisu. Studia Biographica 4. Helsinki: Suomalaisen Kirjallisuuden Seura. <http://urn.fi/urn:nbn:fi:sk:skbg-005241>. <https://www.bbc.com/news/magazine-20157891>. (Haut 28.1.2026.) H. Rautkallio (1995) Suuri viha. WSOY. S. 39.

⁷F. Römer (2022) Poverty, Inequality Statistics and Knowledge Politics Under Thatcher. *English Historical Review*, 137, 513–551.

⁸A. Cavallo (2013) Online and Official Price Indexes: Measuring Argentina’s Inflation. *Journal of Monetary Economics*, 60, 152–165. A. Carriquiry (2012) Graciela Bevacqua. *Significance*, 9(6), 34–36. <http://www.thebillionpricesproject.com/>. <https://magazine.amstat.org/blog/2012/12/01/argentinastatistics/>. <https://magazine.amstat.org/blog/2016/02/01/argentina/>. <https://isi-web.org/statement/update-situation-argentine-cpi>. <https://www.linkedin.com/in/graciela-bevacqua-09b72815/>. Argentina calls an end to its ‘statistical emergency’. *Financial Times* 15.6.2016. <https://suomenkuvalehti.fi/ulkomaat/argentiina-oli-yksi-maailman-rikkaimmista-maista-mutta-sitten-se-romahti-nyt-pelastusta-lupaa-mies-joka-nimesi-kloonatut-koiransa-ekonomistien-mukaan/>. L. Taylor (2021) The Venezuelan Health-Care Workers Secretly Collecting Covid Stats. *Nature*, 597, 20–21. https://en.wikipedia.org/wiki/Impeachment_of_Dilma_Rousseff. (Haut 28.1.2026.)

⁹B. Tarran (2018) Andreas Georgiou’s Legal Troubles Show No Sign of Ending. *Significance*, 16, 4. <https://isi-web.org/images/news/Court/proceedings/against/Andreas/Georgiou.pdf>

(vanhentunut linkki). https://www.amstat.org/asa/files/pdfs/pressreleases/2018-Sign_on_Letter_Andreas_Georgiou.pdf. <https://iaos-isi.org/2021/08/12/isi-and-iaos-letter-of-support-for-andreas-georgiou/>. <https://isi-web.org/statement/isi-statement-about-persisting-legal-action-against-andreas-georgiou>. <https://magazine.amstat.org/blog/2023/12/01/georgiousecase/>. <https://isi-web.org/statement/isi-statement-reopening-case-against-andreas-georgiou>. (Haut 28.1.2026.) Helsingin Sanomat 16.8.2023. Financial Times 27.11.2011. R. Mink (2022) *Official Statistics – A Plaything of Politics?*. Springer. Luku 7.

¹⁰<https://yle.fi/a/3-10423307>. <https://www.hs.fi/talous/art-2000006179435.html>. <https://www.hs.fi/talous/art-2000009784157.html>. <https://www.reuters.com/world/china/china-stop-releasing-youth-jobless-rate-data-aug-says-stats-bureau-2023-08-15/>. <https://suomenkuvalehti.fi/paajutut/pitaako-kiinalaisista-olla-huolissaan-tuuli-koivu-nyt-on-vain-yksi-lohikaarme-eivat-he-sellaista-halunneet>. (Haut 28.1.2026.)

¹¹Yle Uutiset 1.11.2019 <https://yle.fi/uutiset/3-11042210> (haettu 28.1.2026).

¹²<https://www.valaszonline.hu/2025/04/11/reliability-eu-silc-income-data-in-hungary/>. <https://www.valaszonline.hu/2025/05/06/hcsos-poverty-data-hungary/>. https://www.ksh.hu/sajtoszoba_kozlemenyek_tajekoztatok_2025_04_02_2. https://www.researchgate.net/publication/392362865_Data_Quality_Issues_in_Hungarian_Income_Data_in_EU-SILC_2014-2023_V3. <https://www.valaszonline.hu/2025/06/03/systemic-data-quality-issues-in-hungarian-eu-silc-income-data/>. (Haut 28.1.2026.) Kiitän Annamária Tátraita aggregoidun aineiston luovuttamisesta 17.5.2025 ja selvennyksistä 8.6.2025.

¹³Financial Times 29.1.2020. Helsingin Sanomat 29.1.2020 <https://www.hs.fi/maailma/art-2000008575513.html>. <https://www.amstat.org/news-listing/2022/02/09/statement-from-asa-and-rss-on-the-dismissal-of-turkish-statistics-chair>. <https://rss.org.uk/news-publication/news-publications/2022/general-news/a-statement-from-the-royal-statistical-society-and/>. <https://isi-web.org/statement/isi-and-iaos-statement-dismissal-head-turkish-statistical-office>. (Haut 28.1.2026.)

¹⁴Helsingin Sanomat 8.7.2025 <https://www.hs.fi/maailma/art-2000011354232.html> (haettu 28.1.2026).

¹⁵<https://www.nytimes.com/2025/03/04/business/economy/gdp-data-lutnick-commerce-secretary.html>. <https://isi-web.org/article/isi-advisory-board-ethics-releases-statement-united-states-federal-statistics>. <https://truthsocial.com/@realDonaldTrump/posts/114954846612623858>. <https://truthsocial.com/@realDonaldTrump/posts/114955222046259464>. <https://www.ft.com/content/023b5ad5-b946-486d-a55f-36a9a9f62443>. <https://www.economist.com/united-states/2025/08/04/donald-trump-escalates-his-war-on-numbers>. <https://isi-web.org/statement/statement-about-dismissal-bureau-labor-statistics-bls-commissioner-erika-mcentarfer>. <https://rss.org.uk/news-publication/news-publications/2025/general-news/a-statement-from-the-royal-statistical-society-on/>. https://www.amstat.org/docs/default-source/amstat-documents/communications-from-the-executive-director/august_2_1_i-post.pdf. https://unstats.un.org/unsd/dnss/hb/E-fundamentalprinciples_A4-WEB.pdf. <https://stat.fi/org/tilastotoimi/tilastoperiaatteet-ja-laatu/virallisen-tilaston-periaatteet.html>. (Haut 28.1.2026.)

¹⁶<https://www.theguardian.com/us-news/2024/nov/17/iowa-pollster-j-ann-selzer-quits>. <https://www.washingtonpost.com/style/media/2024/12/17/trump-des-moines-register-ann-selzer-lawsuit/>. <https://iowacapitaldispatch.com/2025/01/08/second-lawsuit-filed-against-des-moines-register-selzer-over-iowa-poll/>. https://en.wikipedia.org/wiki/Ann_Selzer. <https://www.nytimes.com/2026/01/22/us/politics/trump-poll-second-term.html>. <https://truthsocial.com/@realDonaldTrump/posts/115939791935000343>. <https://truthsocial.com/@realDonaldTrump/posts/115939823471731565>. <https://truthsocial.com/@realDonaldTrump/posts/115939860152214301>. (Haut 28.1.2026.)

¹⁷<https://www.nytimes.com/2026/02/11/us/politics/gallup-poll-presidential-approval-ratings-trump.html>. <https://www.theguardian.com/us-news/2026/feb/11/gallup-stop-tracking-presidential-approval-ratings>. (Haut 16.2.2026.)

¹⁸E. Senata (2004) Mathematics, Religion, and Marxism in the Soviet Union in the 1930s. *Historia Mathematica*, 31, 337–367.

Joviaali kolmen lajin proffa

LASSE KOSKINEN, MARKKU SOINTU



Suomalainen emeritusprofessori ja tilastotieteilijä ANTTI JAAKKO KANTO kuoli 2. marraskuuta 2025.

Antti valmistui tilastotieteestä Tampereen yliopistosta filosofian kandidaatiksi 1977 ja tohtoriksi 1983. Suurimmiksi kiinnostukseen kohteikseen tutkimuksessa hän nimesi tilastolliset menetelmät rahoituksessa, raviurheilussa ja urheiluedonlyönissä.

Professori Antti Kanto työskenteli nuoruudessaan erilaisissa akateemisissa tehtävissä Vaasan ja Tampereen yliopistossa. Merkittävimmän osan uraansa hän oli Helsingin kauppakorkeakoulun tilastotieteen ja talousmatematiikan professori. Tuolloin Antti sai maineen erinomaisena luennoitsijana vaikeiden asioiden

den selkeään esittämisen ja humoristien tyyhinsä takia. Uransa viimeiset vuodet hän toimi Tampereen yliopiston vakuutustieteen professorina. Yliopiston lisäksi Antti ansaitsi nimityksen proffa ravipiireissä ja tietovisoissa. Erityislaatusena persoonallisuutena hän jäi pysyvästi tapaamiensa ihmisten mieleen.

Tutkijana Kanto oli monipuolinen sekä tutkimusaiheiden että yhteistyökumppanien suhteen. Hän tutkimuksensa kohteet vaihtelivat kauppatieteiden aihepiiristä lääketieteeseen ja tekniikkaan. Tutkijana hänellä oli poikkeuksellisen kekseliäs teoreettisten menetelmien soveltamisessa moninaisiin tutkimustehäviin ja -aineistoihin.

Henkilöä, joka on iloinen, hilpeä, hyväntahtoinen ja sosiaalinen, jota on helppo lähestyä ja jonka seurassa on miellyttävä olla, sanotaan joviaaliksi. Professori Kannon ystäväpiiri piti häntä joviaalina. Antti toi paljon iloa nokkelalla tilanehuumorillaan tutkimusyhteistyössä ja ohjatessaan opiskelijoita sekä nuoruudessaan pelatessaan lentopalloa, jalkapalloa tai osallistuessaan yleisurheilumoniotteluihin.

Sosiaalinen luonne ohjasi Kannon moniin harrastuksiin. Oli luonnollista, että hän tilastotieteilijänä kiinnostui raveista. Näkökulma ei jäänyt hänellä pelkäästään peliteoreettiseksi, vaan Kanto suoritti raviohjastajan tutkinnon. Hän ajoi kilpaa ainakin Killerjärven raviradalla.

Kiinnostus harrastaa asioita itse ja pohtia niitä teoreettisesti teki Kannosta erinomaisen tietovisailijan. Hän osallistui aikoinaan television suosittuun ”Suomen tietoviisas”-kilpailuun ja lukuisiin pubivisoihin. Hän oli myös luonteva valinta kysymysten laatijaksi ja tuomariksi. Ominaista Antin laatimille kysymyksille oli oivaltavuus. Hyvä tietokilpailukysymys ei vaatinut vain päntättyä tietoa, vaan siihen liitettävää kykyä päätellä. Esimerkiksi: miksi kuvitteellisen HAL 9000 (älykäs tekoälytietokone Avaruusseikkailu-sarjan pääantagonisti) nimi on HAL?

Monelle tutulle Antin mainio taito sanaleikkeihin toi eniten iloa. Maistiaisena: ”Kuuluisat Rossin veljekset ovat Tino, Pape ja Motoc” ja ”The world’s thinnest math book is ’Even prime numbers’, sen on kirjoittanut kiinalainen matemaatikko Oun Lee Wan.”

Rakkaimpia ihmisiä Antille olivat puoliso Virve, tytär Katariina ja sisko Anneli. Katariinan puoliso sai myös kiittävät lausunnot. Katariinan edesmennyt äiti herätti Antissa paljon lämpimiä muistoja.

Kirjoittaja LASSE KOSKINEN on Antin seuraaja vakuutustieteen professorina ja MARKKU SOINTU on julkaisuja Antin kanssa kirjoittanut eläkkeellä oleva matematiikan lehtori.

Antti Jaakko Kanto (s. 23.3.1953 Pirkkala, k. 2.11.2025 Tampere)

Väitöskirjapalkinto

Non-linear state-space methods for Bayesian time series modelling

SANTERI KARPPINEN
SKARPPINEN@IKI.FI
GARMIN JYVÄSKYLÄ OY

Tiivistelmä

Tila-avaruusmenetelmiä käytetään useilla tieteenaloilla parametripäätelyyn sekä niin kutsuttujen suodatus-, tasoitus- ja ennustusongelmien ratkaisuun. Näihin tilastollisen päätelyn ongelmiin on olemassa analyttiset ratkaisut pääasiassa lineaaristen sekä diskreettien tila-avaruusmallien tapauksessa. Näiden erikoistapausten ulkopuolisten yleisten tila-avaruusmallien tilastollisessa päätelyssä hyödynnetään usein simulointimenetelmiä, kuten hiukkassuotimia.

Kehitin väitöskirjassani uusia Bayes-päätelyn menetelmiä yleisille tila-avaruusmalleille, ja sovelsin tila-avaruusmalleja epälineaarisin mallinuongelmiin ja moniulotteisiin aikasarja-aineistoihin. Väitöskirjassa kehittämäni uudet menetelmät soveltuvat erityisesti tila-avaruusmalleille, joiden alkutilan todennäköisyysjakauma on epäinformatiivinen, ja tila-avaruusmalleille, joiden tiladynamiikka on hitaasti sekoittuva tai joiden havaintoprosessit ovat epäinformatiivisia.

Tämä teksti kertoo, kuinka yleisen tila-avaruusmallin Bayes-päätelyä on mahdollista lähestyä (ehdollisin) hiukkassuotimin. Lopuksi kerron, kuinka väitöstutkimukseni liittyy tähän aihepiiriin.

Väitöskirjani on vapaasti luettavissa osoitteessa https://jyx.jyu.fi/jyx/Record/jyx_123456789_83847.

Johdanto

Kutsun tässä tekstissä *yleiseksi tila-avaruusmalliksi* tilastollista mallia, jossa p -ulotteisen, n -pituisen aikasarjan $(Y_k)_{1 \leq k \leq n}$ oletetaan riippuvan latentista d -ulotteisesta Markov-prosessista $(X_k)_{0 \leq k \leq n}$ siten, että:

$$X_k \mid (X_{k-1} = x_{k-1}) \sim f_k^{(\theta)}(\cdot \mid x_{k-1}), \text{ kun } k \geq 1 \text{ ja } X_0 \sim f_0^{(\theta)}(\cdot) \quad (1a)$$

$$Y_k \mid (X_k = x_k) \sim g_k^{(\theta)}(\cdot \mid x_k), \text{ kun } k \geq 1, \quad (1b)$$



Kuva 1: Jyrki Möttönen (vas.) ja Santeri Karppinen.

jossa i) $f_0^{(\theta)}$ on prioritiheys alkutilalle X_0 , ii) $(f_k^{(\theta)})_{k \geq 1}$ ovat tilamuuttujien X_k Markov-siirtymätiheyksiä, iii) $(g_k^{(\theta)})_{k \geq 1}$ ovat havaintojen Y_k ehdollisia tiheysfunktioita ehdolla $X_k = x_k$ ja iv) vektori θ kuvaa mallin tuntemattomia parametreja, joista f_k :t ja g_k :t voivat riippua.

Yhtälöä (1a) kutsutaan usein (tila-avaruusmallin) *tilayhtälöksi* tai *dynaamiseksi malliksi* ja yhtälöä (1b) tila-avaruusmallin *havaintoyhtälöksi* tai *havaintomalliksi*. Niin ikään latenteja muuttujia X_k kutsutaan mallin *tilamuuttujiksi* ja satunnaismuuttujia Y_k *havainnoiksi*.

Yleinen tila-avaruusmalli (1) taipuu monenlaiseen: väitöskirja-artikkelini (Karppinen ja Vihola, 2021) esimerkissä tilayhtälö (1a) kuvaa koronaepidemian dynamiikkaa, josta havaitaan negatiivista binomijakaumaa noudattavia tartuntalukumääriä. Artikkelissa (Karppinen, Lohi ja M., 2019) tilayhtälö sen sijaan mallintaa lääkityksen vaikutusta akuutin lymfoblastileukemian hoidossa, ja havainnot ovat potilaan leukosyyttipitoisuuksia.

Tila-avaruusmalliin (1) liittyvän Bayes-päätelyn tavoitteena on laskea erilaisia latenttien tilamuuttujien X_k posteriorijakaumia. Nämä estimointiongelmien jaetaan yleensä kohdejakauman mukaan ns. *suodatusongelmiin* (kohteena $p(x_k | y_{1:k})$, kun $k \geq 1$), *tasointuongelmiin* (kohteena $p(x_{0:n} | y_{1:n})$) ja erilaisiin ennusteongelmiin (kohteena esim. $p(x_{k+h} | y_{1:k})$ tai $p(y_{k+h} | y_{1:k})$ jollakin $k, h \geq 1$) (Särkkä, 2013). Lisäksi myös parametrit θ on usein estimoitava samanaikaisesti.

Suodatus-, tasointu- ja ennusteongelmiin tunnetaan suljetun muodon ratkaisut pääasiassa lineaaristen tila-avaruusmallien (ks. esim. Durbin ja Koopman, 2012; Harvey, 1990) ja diskreettien tila-avaruusmallien (ts. piilo-Markov mallien) (Baum ja Petrie, 1966; Rabiner, 1989) tapauksessa. Nämä ovat mallin (1) erikoistapauksia, joissa tiheyksille f_k ja g_k asetetaan lisäoletuksia.

Kun sovelluskohde ei istu näihin tunnetuihin tapauksiin, on tilastollisessa päätelyssä turvauduttava joko mallia tai päätelyä approksimoiviin menetelmiin (ks. esim. Särkkä, 2013) tai stokastiseen simulointiin perustuviin menetelmiin, kuten *hiukkassuotimiin* (*particle filters*) (ks. esim. Doucet, De Freitas ja Gordon, 2001).

Bayes-päättelyä yleiselle tila-avaruusmallille hiukkassuotimin

On hyvä pitää mielessä, että yleiseen tila-avaruusmalliin (1) liittyvässä Bayes-päätelyssä ei välttämättä tarvita hiukkassuodinta lainkaan. Päättelyä on nimittäin mahdollista lähestyä myös hierarkkisista Bayes-malleista tuttuun tapaan, kohdistamalla esimerkiksi Metropolisin-Hastingsin algoritmiin pohjautuva MCMC-menetelmä kohti mallin (1) posterioritiheyttä

$$p(\theta, x_{0:n} \mid y_{1:n}) \propto p(\theta) f_0(x_0) \prod_{k=1}^n f_k(x_k \mid x_{k-1}) g_k(y_k \mid x_k), \quad (2)$$

parametreja θ ja tilamuuttujia $x_{0:n}$ simuloiden. Tämä lähestymistapa on varsin houkutteleva, sillä usein (3) on helppo esittää (Metropolisin-Hastingsin algoritmin johdannaisia hyödyntävissä) hyvin dokumentoiduissa, tunnetuissa ja laajasti saatavilla olevissa probabilistisissa ohjelmointikielissä.

Aikasarjan pituuden, mallin epälineaarisuuden ja/tai ajallisen riippuvuuden kasvaessa tämä lähestymistapa voi kuitenkin jäädä tehottomaksi (ks. esim. Fearnhead, 2011). Tämä näkyy käytännössä simulointialgoritmin heikkona suorituskäytännä.

Lisää tehoa voikin lähteä etsimään esim. hiukkassuotimista, joita voi yksinkertaistaen ajatella simulointimenetelminä, jotka soveltuvat hyvin tila-avaruusmallinnukseen. Hiukkassuotimia on kuitenkin mahdollista käyttää myös muiden kuin aikasarjamallien Bayes-päätelyyn. Tällöin puhutaan yleensä *sekventiaalisista Monte Carlo -menetelmistä* (*sequential Monte Carlo methods*) (ks. esim. Chopin ja Papaspiliopoulos, 2020, kappale 17).

Esitän seuraavaksi kaksi perushiukkassuodinta: tavallisen hiukkassuotimen ja ehdollisen hiukkassuotimen.¹ Kuvaan menetelmät tilanteessa, jossa parametrit θ ovat kiinnitettyinä. Tällöin kiinnostuksen kohteena oleva todennäköisyys tiheys on siis $p(x_{0:n} \mid y_{1:n}) \propto \gamma(x_{0:n}, y_{1:n})$, missä

$$\gamma(x_{0:n}, y_{1:n}) := f_0(x_0) \prod_{k=1}^n f_k(x_k \mid x_{k-1}) g_k(y_k \mid x_k), \quad (3)$$

ja jossa parametrin θ arvo on jätetty merkitsemättä merkintöjen yksinkertaistamiseksi. Jos myös parametri θ on estimoitava, voidaan (tavallinen tai ehdollinen) hiukkassuodin sisällyttää osaksi MCMC-algoritmia; kerron myös tästä lyhyesti tekstin loppupuolella.

Kaikissa tämän tekstin merkinnöissä alaindeksit viittaavat ajanhetkeen k , $0 \leq k \leq n$, ja yläindeksit hiukkasen (joista lisää pian) indeksiin i , $1 \leq i \leq N$. Merkinnällä $i:j$ tarkoitan indeksijoukkoa $\{i, i+1, i+2, \dots, j\}$.

Hiukkassuodin

Hiukkassuotimet ovat tärkeysotantaan (Geweke, 1989) ja erityisesti sekventiaaliseseen tärkeysotantaan (ks. esim. Doucet, De Freitas ja Gordon, 2001, kappale 1.3.2) perustuvia menetelmiä, joiden avulla on käytännön Bayes-päätelyn näkökulmasta mahdollista i) approksimoida posteriorijakaumiin $p(x_{0:k} \mid y_{1:k})$, $k \geq 1$ liittyviä odotusarvoja ja ii) approksimoida tila-avaruusmallin marginaaliusko-tavuuksia $p(y_{1:k})$, $k \geq 1$.

Algoritmi 1 BOOTSTRAP-HIUKKASSUODIN($((f_k)_{k \geq 0}, (g_k)_{k \geq 1}, r, N)$)

```

1: Simuloi  $X_0^{(i)} \sim f_0$  kaikilla  $i = 1, \dots, N$ .
2: Aseta  $\mathbf{X}_0^{(i)} = X_0^{(i)}$  kaikilla  $i = 1, \dots, N$ .
3: Aseta  $W_0^{(i)} = g_0(\mathbf{X}_0^{(i)}) = g_0(X_0^{(i)}) = 1$  kaikilla  $i = 1, \dots, N$ .
4: for  $k = 1, 2, \dots, n$  do
5:   Simuloi  $A_{k-1}^{(i)} \sim r(\cdot \mid W_{k-1}^{(1:N)})$  kaikilla  $i = 1, \dots, N$ .
6:   Simuloi  $X_k^{(i)} \sim f_k(\cdot \mid X_{k-1}^{(A_{k-1}^{(i)})})$  kaikilla  $i = 1, \dots, N$ .
7:   Aseta  $\mathbf{X}_k^{(i)} = (\mathbf{X}_{k-1}^{(A_{k-1}^{(i)})}, X_k^{(i)})$  kaikilla  $i = 1, \dots, N$ .
8:   Laske  $W_k^{(i)} = g_k(y_k \mid \mathbf{X}_k^{(i)}) = g_k(y_k \mid X_k^{(i)})$  kaikilla  $i = 1, \dots, N$ .
9: end for
10: return  $(X_{0:n}^{(1:N)}, W_{0:n}^{(1:N)}, A_{0:n-1}^{(1:N)})$ 
    
```

Hiukkassuotimessa lasketaan N k -pituista *hiukkassujonoa* $\mathbf{X}_k^{(i)}$, joilla on *painot* $\bar{W}_k^{(i)} = W_k^{(i)} / \sum_{j=1}^N W_k^{(j)}$ (jossa $W_k^{(i)}$ on *normalisoimaton paino*) siten, että $\sum_{i=1}^N \bar{W}_k^{(i)} = 1$. Näin saadaan estimaattorit²

$$\mathbb{E}[h(x_{0:k}) \mid y_{1:k}] \approx \sum_{i=1}^N h(\mathbf{X}_k^{(i)}) \bar{W}_k^{(i)} \quad (4a)$$

$$p(y_{1:k}) \approx \prod_{j=0}^k \left(\frac{1}{N} \sum_{i=1}^N W_j^{(i)} \right), \quad (4b)$$

jossa h on lähes mielivaltainen funktio. Lukumäärää N kutsutaan *hiukkaskukumääräksi*; korkea hiukkaskukumäärä on yhteydessä tarkempiin approksimaatioihin, mutta toisaalta kasvattaa hiukkassuotimen laskennallista vaativuutta.

Hiukkassujonot muodostetaan simuloimalla *hiukkasia* $X_k^{(i)}$, $k \in 0:n$, $i \in 1:N$ ja *jälkeläisindeksejä* $A_k^{(i)}$, $k \in 0:n-1$, $i \in 1:N$ siten, että hiukkassujono i määrittellään k :n aika-askelen jälkeen rekursiivisesti:

$$\mathbf{X}_k^{(i)} := (\mathbf{X}_{k-1}^{(A_{k-1}^{(i)})}, X_k^{(i)}). \quad (5)$$

Tässä sulut tarkoittavat hiukkasen $X_k^{(i)}$ liittämistä $(k-1)$ -pituisen hiukkassujonon $\mathbf{X}_{k-1}^{(A_{k-1}^{(i)})}$ jatkoksi, ja jälkeläisindeksi $A_{k-1}^{(i)}$ siis kertoo, missä indeksissä sijaitsee se ajanhetken $k-1$ hiukkanen (tai hiukkassujono) joka on ajanhetken k hiukkasen $X_k^{(i)}$ vanhempi. Hiukkassujonon (normalisoimattomat) painot $W_k^{(i)}$ lasketaan sen perustella, miten hyvin ne ‘selittävät’ aikasarjahavainnot.

Ensimmäisenä hiukkassuotimena pidetään useimmiten (Gordon, Salmond ja Smith, 1993) kehittämää *bootstrap-hiukkassuodinta*. Hieman laajennettu versio tästä menetelmästä on esitetty Algoritmissa 1. Algoritmin syöte sisältää yleisen tila-avaruusmallin (1) muodostavat tiheydet, toisto-otanta algoritmin r (josta lisää alempana), sekä valitun hiukkaskukumäärän N .

Algoritmin logiikka koostuu kolmesta ydinoperaatiosta, joita toistetaan kulakin ajanhetkellä $k \geq 0$. Ensimmäisessä operaatiossa N uutta hiukkasta *simuloidaan* jakaumasta f_k , jo olemassa olevien hiukkassujonon jatkeeksi (rivit 1–2 ja 6–7). Toisessa operaatiossa näin saaduille hiukkassujonoille *lasketaan (normalisoimattomat) painot* (rivit 3 ja 8). Kolmas operaatio simuloi jälkeläisindeksit

$A_{k-1}^{(1:N)}$ toisto-otannalla r . Tämän operaation voi tulkita heuristisesti siten, että N ‘kyvykkäintä’ (eli suuren painon omaavaa) hiukkasjonoa valitaan satunnaisesti ja palauttaen ‘jatkoon’ seuraavalle ajanhetkelle (rivi 5).

Algoritmin 1 rivit 2 ja 7 eivät tarkkaan ottaen ole tarpeellisia (myöhemmät askeleet eivät riipu niistä), sillä bootstrap-hiukkassuotimessa hiukkasjonon $\mathbf{X}_k^{(i)}$ painon määrittää havaintotiheys $g_k(y_k | X_k^{(i)})$, joka ei riipu koko hiukkasjonosta, vain jonon uusimmasta jäsenestä. On kuitenkin merkittävää pitää mielessä, että paino lasketaan nimenomaan hiukkasjonolle, ei hiukkaselle.

Hiukkasjonojen painot ovat syötteenä toisto-otanta algoritmille r , joka voidaan valita monella tavalla. Tunnetuin valinta lienee Gordon, Salmond ja Smith, 1993 käyttämä multinomiaalinen toisto-otanta, jossa jälkeläisindeksit valitaan palauttaen diskreetistä jakaumasta, joka asettaa todennäköisyyden $\bar{W}_k^{(i)}$ kokonaisluvulle i . Yleisesti ottaen on teoreettisesti mielekästä käyttää toisto-otantaa, joka on harhaton, eli jonka valintaprosessille pätee

$$E \left[\sum_{i=1}^N 1(A_{k-1}^{(i)} = j) \right] = N \bar{W}_{k-1}^{(j)}$$

kaikilla $j \in 1:N$. Tällaisia toisto-otanta algoritmeja ovat esimerkiksi mainittu multinomiaalinen toisto-otanta, mutta myös ositettu (Kitagawa, 1996) ja ns. jäännöstöisto-otanta (Baker, 1985; Higuchi, 1997).³ Toisto-otannan valinnalla on usein vaikutusta estimaattorien (4a) ja (4b) tehokkuuteen.

Algoritmin 1 paluuarvo koostuu *hiukkasmatriisista* $X_{0:n}^{(1:N)}$ (joka sisältää jokaisen simuloidun hiukkasen jokaiselta ajanhetkeltä), normalisoimattomasta painomatriisista $W_{0:n}^{(1:N)}$ sekä jälkeläisindeksimatriisista $A_{0:n-1}^{(1:N)}$. Huomaa, että polut $\mathbf{X}_k^{(i)}$ kaikille i ja k voidaan muodostaa ‘jälkikäteen’ jälkeläisindeksien ja hiukkasmatriisin avulla rekursiota (5) hyödyntäen.

Ehdotusjakauman vaihto ja Feynmanin-Kacin esitystapa Algoritmin 1 heikkoutena voi olla, että jakaumasta f_k simulointi tuottaa liian usein hiukkasia, joiden (normalisoimattomat) painot jäävät rivillä 8 pieniksi. Tällöin havainnoilla ehdollistettua tila-avaruutta ei tutkita tehokkaasti ja estimaattorien (4a) ja (4b) tehokkuus heikkenee. ‘Ehdotusjakauma’ f_k on kuitenkin mahdollista vaihtaa toiseen ehdotusjakaumaan q_k , mikä saattaa helpottaa tilannetta. Tällöin Algoritmin 1 rivillä 6 $X_k^{(i)}$ simuloidaan q_k :sta ja lisäksi painojen laskenta riveillä 3 ja 8 korvataan kaavalla

$$W_k^{(i)} = \frac{g_k(y_k | X_k^{(i)}) f_k(X_k^{(i)} | X_{k-1}^{(A_{k-1}^{(i)})})}{q_k(X_k^{(i)} | \mathbf{X}_{k-1}^{(A_{k-1}^{(i)})}, y_{1:k})},$$

algoritmin pysyessä muutoin samana. Tässä ehdotusjakauman q_k on annettu riippua koko hiukkasjonosta sekä k :sta ensimmäisestä havainnosta, joka sekun on mahdollista.

‘Ehdotusjakauman valintaa’ on mahdollista yleistää tätäkin enemmän. Tämä perustuu *Feynmanin-Kacin (FK) esitystapaan* (Del Moral, 2004), (Chopin ja Papaspiliopoulos, 2020, kappale 5), jossa yleiselle tila-avaruusmallille (1) muodostetaan vaihtoehtoinen alkujakauma M_0 , Markov-siirtymät $M_k, k \geq 1$ ja *potentiaalfunktiot* (tai *potentiaalit*) $G_{0:n}$.

Näistä FK-esitystavan *komponenteista* muodostetulta Feynman-Kac-tiheydeltä⁴

$$\kappa(x_{0:n}) := M_0(x_0)G_0(x_0) \prod_{k=1}^n M_k(x_k \mid x_{k-1})G_k(x_{k-1}, x_k), \quad (6)$$

vaaditaan $\kappa(x_{0:n}) = \gamma(x_{0:n}, y_{1:n})$ kaikilla jonoilla $x_{0:n}$ (jossa γ määriteltiin kaavassa (3)). Tällöin sanotaan, että komponentit $M_{0:n}$ ja $G_{0:n}$ muodostavat *erään* yleisen tila-avaruusmallin (1) Feynmanin-Kacin esitystavan. Oleellista on, että tässä tilanteessa FK-esitystavan käyttäminen johtaa samaan Bayes-päätelyyn jonoille $x_{0:n}$ kuin alkuperäinen malli (3).

FK-esitystapaa käyttäen voidaanakin kirjoittaa hyvin yleinen hiukkassuodin kuten Algoritmissa 1, mutta ‘sijoituksilla’ $g_k = G_k$ ja $f_k = M_k$ (ks. Chopin ja Papaspiliopoulos, 2020, kappale 10.1). FK-esitystavassa ‘valittu ehdotusjakautuma’ siis esitetään suoraan ‘mallin esitystavassa’, eikä algoritmia tarvitse enää muuttaa ehdotusjakautuman muuttuessa. Lisäksi painofunktioihin G_k on myös mahdollista sisällyttää ‘lisäfunktioita’, jotka supistuvat juuri sopivasti pois, jotta ehto $\kappa(x_{0:n}) = \gamma(x_{0:n}, y_{1:n})$ pätee. Näiden lisäfunktioiden käyttö voi tehostaa hiukkassuodimen toimintaa tietyissä tilanteissa (ks. esim. Guarniero, Johansen ja Lee, 2017). Lisää FK-esitystavan ‘motivointia’ ja esimerkkejä löytyy väitöskirjani yhteenvedon kappaleesta 4.2.

Ehdollinen hiukkassuodin ja takeneva simulointi

Estimaattori (4a) mahdollistaa hyvin monenlaisten odotusarvojen estimoimisen hiukkassuodimen avulla. Valitetettavasti estimaattori ei kuitenkaan yleensä toimi tehokkaasti sellaisille funktioille h , jotka riippuvat aikasarjan alkupään tilamuuttujista; siis tilamuuttujista x_k , joille $k \ll n$. Syy tälle on toisto-otannassa, jonka käytöstä seuraa helposti, että lähes kaikilla aikasarjan lopun hiukkasjonoilla on aikasarjan alkupuolella yhteinen ‘kantahiukkanen’. Siksi lopullisten hiukkasjonojen alkupään hiukkasissa ei usein ole lähes ollenkaan vaihtelua.

Vaihtoehtoinen, erityisesti tasoitusongelmien mielessä tehokkaampi estimaattori saadaan hyödyntämällä (Andrieu, Doucet ja Holenstein, 2010) esittelemaa *ehdollista hiukkassuodinta* (*conditional particle filter*) (CPF). CPF on algoritmisesti hyvin lähellä tavallista hiukkassuodinta (esim. algoritmi (1)). Keskeisenä erona on se, että CPF:ssä paluuarvona oleva hiukkasmatriisi $X_{0:n}^{(1:N)}$ (ja paino- sekä jälkeläisindeksimatriisi) ehdollistetaan yhdellä, syötteenä annettavalla hiukkasjonoilla, jota kutsutaan *referenssijonoksi*.

Referenssijono $(X_k^*)_{k \geq 0}$ asetetaan CPF:ssä ‘paikalleen’ hiukkasmatriisin indekseihin $B_{0:n}$ (jotka ovat myös algoritmin syöte), ts. hiukkasmatriisissa $X_k^{(B_k)} = X_k^*$ kaikilla k . Tämän jälkeen hiukkasmatriisin ei-referenssijonoon kuuluvien hiukkasten (ja hiukkasjonojen) simulointi suoritetaan kuten tavallisessa hiukkassuodimessa, mutta varmistaen, että referenssijonoa ‘ei kirjoiteta yli’ hiukkasia simuloitaessa ja toisto-otantaa suoritettaessa. Tämä saa aikaan sen, että kussakin toisto-otannassa valitsemattomat hiukkasjonot voivat ‘syntyä uudestaan’ referenssijonosta, josta seuraa hiukkasmatriisin ‘ehdollistuminen’ referenssijonolla.

CPF FK-esitystavalla $\mathcal{FK}_{0:n} := (M_{0:n}, G_{0:n})$, hiukkaslukumäärällä N , referenssipolun $X_{0:n}^*$ sijaitessa indekseissä $B_{0:n}$, on esitetty Algoritmissa 2. Algoritmissa 2 on kolme eroavaisuutta verrattuna bootstrap-hiukkassuodimeen

(Algoritmi 1): i) algoritmi on kirjoitettu kiinnostuksen kohteena olevan tila-avaruusmallin FK-esitystavalle $(M_{0:n}, G_{0:n})^5$, ii) hiukkasen simulointi jätetään tekemättä referenssijonon kohdalla (rivit 1 ja 6) ja referenssijonoon kuuluvat hiukkaset asetetaan paikalleen hiukkasmatriisiin kullakin aika-askeleella (rivit 2 ja 7). Lisäksi, iii) tavallisen toisto-otannan sijaan suoritetaan *ehdollinen toisto-otanta* (jota on merkitty samalla symbolilla r).

Ehdollinen toisto-otanta $r^{(B_{k-1}, B_k)}$ simuloi jälkeläisindeksit siten, että $A_{k-1}^{(B_k)} = B_{k-1}$, ja $A_{k-1}^{(i)}$ kaikilla $i \neq B_k$ on simuloidaan ehdollistamalla tapahtumalla $A_{k-1}^{(B_k)} = B_{k-1}$. Tämä yhdessä yllä olevan kohdan ii) kanssa takaa, ettei referenssijonoa ylikirjoiteta. Yksinkertaisimmillaan ehdollisen toisto-otannan voi toteuttaa ehdollisena multinomiaalisena toisto-otantana, jossa i) suoritetaan tavallinen multinomiaalinen toisto-otanta, ii) sotketaan näin saatu jälkeläisindeksivektorin järjestys sattumanvaraisesti, ja iii) ylikirjoitetaan lopuksi $A_{k-1}^{(B_k)} = B_{k-1}$.⁶

Tarkalleen ottaen Algoritmi 2 on pieni yleistys (Andrieu, Doucet ja Holenstein, 2010) versiosta, jossa toisto-otantaa ei käsitelty ‘yleisenä ehdollisena toisto-otantana’, kuten yllä. Väitöskirjani artikkelissa (Karppinen, Singh ja Vihola, 2024) esitetään riittävä ehto sille, että yleinen (yllä kuvatulla tavalla käyttäytyvä) ehdollinen toisto-otanta toimii Algoritmissa 2. Artikkelissa on esitetty myös kaksi tällaista ehdollista toisto-otantaa.

Algoritmi 2 EHDOLLINEN HIUKKASSUODIN($\mathcal{FK}_{0:n}, N, X_{0:n}^*, B_{0:n}$)

- 1: Simuloi $X_0^{(i)} \sim M_0(\cdot)$ kaikilla $i = 1, 2, \dots, N, i \neq B_0$.
 - 2: Aseta $X_0^{(B_0)} = X_0^*$ ja $\mathbf{X}_0^{(i)} = X_0^{(i)}$ kaikilla $i = 1, 2, \dots, N$.
 - 3: Laske $W_0^{(i)} = G_0(\mathbf{X}_0^{(i)})$ kaikilla $i = 1, 2, \dots, N$.
 - 4: **for** $k = 1, 2, \dots, n$ **do**
 - 5: Simuloi $A_{k-1}^{(1:N)} \sim r^{(B_{k-1}, B_k)}(\cdot \mid W_{k-1}^{(1:N)})$.
 - 6: Simuloi $X_k^{(i)} \sim M_k(\cdot \mid X_{k-1}^{(A_{k-1}^{(i)})})$ kaikilla $i = 1, 2, \dots, N, i \neq B_k$.
 - 7: Aseta $X_k^{(B_k)} = X_k^*$.
 - 8: Aseta $\mathbf{X}_k^{(i)} = (X_{k-1}^{(A_{k-1}^{(i)})}, X_k^{(i)})$ for $i = 1, 2, \dots, N$.
 - 9: Laske $W_k^{(i)} = G_k(\mathbf{X}_k^{(i)})$ for $i = 1, 2, \dots, N$.
 - 10: **end for**
 - 11: **return** $(X_{0:n}^{(1:N)}, W_{0:n}^{(1:N)}, A_{0:n-1}^{(1:N)})$
-

CPF voidaan yhdistää niin sanottuun *jäljitysvaiheeseen* (*traceback*), jossa CPF:n tuottamasta hiukkasmatriisista muodostetaan uusi hiukkasjono. Tehokas tapa toteuttaa jäljitysvaihe on *takeneva simulointi* (*backward sampling*) (Whitley, 2010), joka on kuvattu Algoritmissa 3. Algoritmin ulostulona on indeksijono $\tilde{B}_{0:n}$, joka merkitsee uuden hiukkasjonon $X_{0:n}^{(\tilde{B}_{0:n})}$ CPF:n palauttamassa hiukkasmatriisissa. Merkitty hiukkasjono saattaa olla sellainenkin, jota ei muodostettu CPF:n aikana. Merkintä Categ tarkoittaa algoritmissa kategorista jakaumaa, joka normeeraa syötteenä annettavat painot ja simuloi yhden indekseistä $1:N$ näillä painoilla.

Hiukkasjono $X_{0:n}^{(\tilde{B}_{0:n})}$ voidaan nyt syöttää CPF:ään uudelleen ‘seuraavana referenssijonona’, jonka jälkeen takeneva simulointi ajetaan taas; tuloksena on taas uusi hiukkasjono. Jatkamalla näin iteratiivisesti (yksi askel kuvattu Algoritmissa 4), muodostuu *hiukkasjonon Markovin-ketju* jonka tasapainojakaumana on $p(x_{0:n} \mid y_{1:n})$.⁷ Tästä seuraa, että tasoitusergelmaan liittyvien odotusarvojen

Algoritmi 3 TAKENEVA SIMULOINTI($X_{0:n}^{(1:N)}, A_{0:n-1}^{(1:N)}, W_n^{(1:N)}$)

```

1: Simuloi  $\tilde{B}_n \sim \text{Categ}(W_n^{(1:N)})$ 
2: for  $k = n-1, n-2, \dots, 0$  do
3:   for  $i = 1, 2, \dots, N$  do
4:     Jos  $k \geq 1$ , aseta  $\mathbf{X}_k^{(i)} = (X_{k-1}^{(A_{k-1}^{(i)})}, X_k^{(i)})$ ; muutoin aseta  $\mathbf{X}_k^{(i)} = X_k^{(i)}$ .
5:     Laske  $\omega_k^{(i)} = G_k(\mathbf{X}_k^{(i)})G_{k+1}(X_k^{(i)}, X_{k+1}^{(\tilde{B}_{k+1})})M_{k+1}(X_{k+1}^{(\tilde{B}_{k+1})} \mid X_k^{(i)})$ .
6:   end for
7:   Simuloi  $\tilde{B}_k \sim \text{Categ}(\omega_k^{(1:N)})$ .
8: end for
9: return  $\tilde{B}_{0:n}$ 
    
```

estimoimiseen voidaan siis käyttää tehokasta estimaattoria

$$\frac{1}{M} \sum_{i=1}^M h(X_{0:n}^{*,i}) \xrightarrow{M \rightarrow \infty} \mathbb{E}[h(x_{0:n}) \mid y_{1:n}], \quad (7)$$

joka pätee *kaikilla* $N \geq 2$ (Algoritmissa 2), kun $X_{0:n}^{*,i}, i = 1, 2, \dots, M$ on simuloitu Algoritmeilla 4. Takenevan simuloinnin käyttöä yhdessä CPF:n kanssa kutsutaan ehdolliseksi hiukkassuotimeksi takenevalla simuloinnilla (*conditional particle filter with backward sampling*) (CPF-BS).

Toinen tapa toteuttaa jäljitys vaihe on valita CPF:n tuloksesta sattumanvarainen hiukkasjono painoilla $\bar{W}_n^{(1:N)}$; tätä kutsutaan *jälkeläisjäljitykseksi* (*ancestor tracing*). Toisin kuin takenevassa simuloinnissa, jälkeläisjäljityksessä hiukasmatriisista ei kuitenkaan voida poimia sellaista hiukkasjonoa, jota ei olisi muodostettu CPF:n aikana, mikä vihjaa takenevan simuloinnin olevan tehokkaampi. Chopin ja Singh, 2015 tuloksien perusteella takenevaa simulointia kannattaakin käyttää oikeastaan aina jälkeläisjäljityksen sijaan. Poikkeuksen muodostavat mallit, jossa tiheyksiä $M_k(x_k \mid x_{k-1})$ ei ole mahdollista laskea (jolloin takenevaa simulointia ei voida toteuttaa).

Algoritmi 4 HIUKKASJONON PÄIVITYS($\mathcal{FK}_{0:n}, N, X_{0:n}^*, B_{0:n}$)

```

1:  $(X_{0:n}^{(1:N)}, W_{0:n}^{(1:N)}, A_{0:n-1}^{(1:N)}) \leftarrow \text{EHD. HIUKKASSUODIN}(\mathcal{FK}_{0:n}, N, X_{0:n}^*, B_{0:n})$ 
2:  $(X_{0:n}^{(\tilde{B}_{0:n})}, \tilde{B}_{0:n}) \leftarrow \text{TAKENEVA SIMULOINTI}(X_{0:n}^{(1:N)}, A_{0:n-1}^{(1:N)}, W_n^{(1:N)})$ 
3: return  $(X_{0:n}^{(\tilde{B}_{0:n})}, \tilde{B}_{0:n})$ 
    
```

Tilamuuttujien ja parametrien θ yhteispäättelystä

Mikäli kiinnostuksen kohteena oleva tila-avaruusmalli riippuu myös tuntemattomista parametreista θ , (eli kohdejakaumana on (2)) on Algoritmeja 1 ja 4 (tai näiden muunnelmia) mahdollista hyödyntää osana ns. hiukkas-MCMC menetelmiä (*particle Markov chain Monte Carlo methods*) (Andrieu, Doucet ja Holenstein, 2010).

Näistä merkittäviä ovat erityisesti hiukkas-Metropolisin-Hastingsin (*particle marginal Metropolis-Hastings, PMMH*) ja hiukkas-Gibbsin (*particle Gibbs*) algoritmit, jotka muistuttavat tavallista Metropolisin-Hastingsin -algoritmia ja Gibbsin-otantaa. Kiinnostunut lukija löytää näiden algoritmien tarkemman kuvauksen (tämän tekstin merkinnöin) väitöskirjani yhteenvedon kappaleesta 4.4.

Väitöskirjassani esitettyjä uusia ehdollisia hiukkassuotimia

Kerron lopuksi lyhyesti väitöskirjani artikkeleista (Karppinen ja Vihola, 2021; Karppinen, Singh ja Vihola, 2024). Näissä artikkeleissa on kehitetty uusia ehdollisia hiukkassuotimia, jotka tehostavat erityisesti edellä esitellyn CPF-BS:n toimintaa tietyissä erikoistapauksissa.

Tasoitusergelma epäinformatiivisen alkujakauman tapauksessa

Artikkeli (Karppinen ja Vihola, 2021) keskittyy yleisen tila-avaruusmallin tasoitusergelmaan tilanteessa, jossa tila-avaruusmallin (tai yleisemmin, FK-esitystavan) alkujakauma M_0 on epäinformatiivinen.⁸ Tällaista ‘diffuusin alustuksen’ tilannetta on tutkittu aiemmin aikasarja-analyysin kontekstissa lähinnä lineaaristen tila-avaruusmallien tapauksessa (ks. Durbin ja Koopman, 2012, kappale 5).

Alkujakauman ollessa kovin epäinformatiivinen, CPF-BS:n käyttö johtaa helposti alkutilojen heikkoon sekoittumiseen simuloitavassa Markovin ketjussa. Tämä johtuu siitä, että alkujakauman ollessa epäinformatiivinen, suurin osa CPF:ssa simuloitavista ensimmäisistä hiukkasista päätyy tila-avaruuden pisteisiin, joilla on pieni potentiaali. Tästä taas seuraa, että nämä hiukkaset valitaan harvoin takenevassa simuloinnissa valittuihin hiukkasjonoihin, jonka vuoksi alkutilojen sekoittuminen on hidasta. Osoitamme empiirisesti, kuinka tämä ilmiö voi tapahtua jo yksilutoteisissa, yksinkertaisissa tila-avaruusmalleissa.

Esitämme artikkelissa ongelmalle ratkaisun menetelmän muodossa, joka muuntaa CPF:ää siten, että simulointia alkujakaumasta M_0 ei tarvita. Alkujakauman sijaan menetelmämme simuloikin ensimmäiset hiukkaset M_0 -reversiibelistä Markov-siirtymästä Q . Keskitymme erityisesti tapaukseen, jossa M_0 on epäaito (tasajakauma $\mathbb{R}^d$:ssä), jolloin Q voidaan valita normaalijakautuneeksi satunnaiskävelyksi.

Hyödyntämällä adaptiivista MCMC-kirjallisuutta (Andrieu ja Thoms, 2008; Haario, Saksman ja Tamminen, 2001) kehitämme menetelmästä version, joka valitsee siirtymän Q hyperparametrit automaattisesti, sopivasti valitun tavoitteena olevan hyväksymistodennäköisyyden perusteella. Esitämme myös heuristiikan tämän tavoitetodennäköisyyden valitsemiseksi; näin menetelmämme käyttö ei vaadi yhdenkään hyperparametrin asettamista.

Kehittämäämme menetelmää on mahdollista hyödyntää myös osana hiukkas-Gibbs -algoritmia, kun tila-avaruusmalli sisältää tuntemattomia parametreja. Jos alkutilojen ja malliparametrien välinen riippuvuus on merkittävää, myös parametrien simulointi voi tehostua tehokkaamman alkutilojen simuloinnin seurauksena.

Ehdollinen hiukkasuodin takenevalla siltasimuloinnilla

Artikkeli (Karppinen, Singh ja Vihola, 2024) käsittelee tasoitusergelmaa tilanteessa, jossa tila-avaruusmallin havainnot (tai FK-esitystavan potentiaalit) ovat heikosti informatiivisia, tai dynaaminen malli sekoittuu hitaasti, ollen ‘hidasliikkeinen’. Tällaisia malleja ovat esimerkiksi tila-avaruusmallit, joiden dynaaminen malli on lineaarisen stokastisen differentiaaliyhtälön hieno diskretaatio. Myös log-normaaleissa Cox-prosesseissa (Møller, Syversveen ja Waagepetersen, 1998) tai erilaisissa polkuintegraalimalleissa ks. Del Moral ja Miclo, 2000 voi tulla tällaisia tilanteita vastaan.

Tässä tilanteessa CPF-BS:n käyttö multinomiaalisella toisto-otannalla johtaa kahteen ongelmaan: i) toisto-otannassa syntyy liiallista varianssia jälkeläisindekseihin, ja ii) hidashiikkeisen dynaamisen mallin tapauksessa takeneva simulointi degeneroituu käytännössä jälkeläisjäljitykseksi, johtaen hitaaseen sekoittumiseen.

Esitämme artikkelissa Chopin, Singh ym., 2022 tulosten innoittamana kaksi uutta ehdollista toisto-otantaa, jotka välttävät multinomiaalisen toisto-otannan ongelmat heikosti informatiivisten havaintojen tapauksessa.

Toisen ongelman ratkaisemme kehittämälle uudenlaisen ehdollisen hiukkassuotimen, jota kutsun ehdolliseksi hiukkassuotimeksi takenevalla siltasimuloinnilla (*conditional particle filter with bridge backward sampling*) (CPF-BBS). Menetelmässä jäljitysvaiheen tilamuuttujien päivitys tehdäänkin takenevasta simuloinnista poiketen tilamuuttujien lohkoissa $x_{T_{i-1}:T_i}$, jossa $1 \leq i \leq K$ on lohkon indeksi. Lohkot määritellään menetelmän syötteenä annettavan lohkojen reunajonon $0 = T_0 < T_1 < T_2 < \dots < T_K = n$ perusteella. Lohkopäivitysten toteuttaminen edellyttää, että dynaamisen mallin ehdolliset siltajakaumat lohkojen yli ovat sellaisia, että niistä on mahdollista simuloida ja niiden todennäköisyystiheydet on mahdollista laskea. Käytännössä tämä rajaa kiinnostuksen kohteena olevan mallin FK-esitystapaa jonkin verran. CPF-BBS voidaan nähdä CPF-BS:n yleistyksenä, sillä valitsemalla lohkojen reunajonoksi ‘triviaali reunajono’ $T_{0:n} = 0:n$ menetelmä yksinkertaistuu CPF-BS:ksi.

Lohkojen reunajono on menetelmämme hyperparametri, joka vaikuttaa simuloitavan Markovin ketjun sekoittumiseen. Kehitämme reunajonon valitsemiseksi laskennallisesti kevyen heuristisen menetelmän, joka pyrkii maksimoimaan todennäköisyyden, jolla CPF-BBS päivittää tilamuuttujien arvoja lohkojen alareunoilla. Päivitystodennäköisyydellä on empiiristen kokeidemme perusteella yhteys simuloitavan Markovin ketjun integroituun autokorrelaatio aikaan. Heuristinen menetelmä hyödyntää päivitystodennäköisyyden arvioimiseen approksimatiivista estimaattoria, joka voidaan laskea ilman CPF-BBS:n iterointia.

Viitteet

- Andrieu, C., Doucet, A. ja Holenstein, R. (2010). ”Particle Markov chain Monte Carlo methods”. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72.3, s. 269–342. DOI: <https://doi.org/10.1111/j.1467-9868.2009.00736.x>.
- Andrieu, C. ja Thoms, J. (2008). ”A tutorial on adaptive MCMC”. *Statistics and Computing* 18.4, s. 343–373. DOI: <https://doi.org/10.1007/s11222-008-9110-y>.
- Baker, J. E. (1985). ”Adaptive selection methods for genetic algorithms”. *Proceedings of the First International Conference on Genetic Algorithms and their Applications* 1, s. 101–111.
- Baum, L. E. ja Petrie, T. (1966). ”Statistical inference for probabilistic functions of finite state Markov chains”. *The annals of mathematical statistics* 37.6, s. 1554–1563. URL: <https://www.jstor.org/stable/2238772>.
- Chopin, N. ja Papaspiliopoulos, O. (2020). *An introduction to sequential Monte Carlo*. Springer. DOI: <https://doi.org/10.1007/978-3-030-47845-2>.
- Chopin, N., Singh, S. S. ym. (2022). ”On resampling schemes for particle filters with weakly informative observations”. *The Annals of Statistics* 50.6, s. 3197–3222. DOI: [10.1214/22-AOS222](https://doi.org/10.1214/22-AOS222). URL: <https://doi.org/10.1214/22-AOS222>.

- Chopin, N. ja Singh, S. S. (2015). "On particle Gibbs sampling". *Bernoulli* 21.3, s. 1855–1883. DOI: <https://doi.org/10.3150/14-BEJ629>.
- Del Moral, P. (2004). *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Springer. DOI: <https://doi.org/10.1007/978-1-4684-9393-1>.
- Del Moral, P. ja Miclo, L. (2000). "Branching and interacting particle systems. Approximations of Feynman-Kac formulae with applications to non-linear filtering". *Séminaire de probabilités de Strasbourg* 34, s. 1–145. URL: <http://eudml.org/doc/114038>.
- Doucet, A., De Freitas, N. ja Gordon, N. J. (2001). *Sequential Monte Carlo methods in practice*. Springer. DOI: <https://doi.org/10.1007/978-1-4757-3437-9>.
- Durbin, J. ja Koopman, S. J. (2012). *Time series analysis by state space methods*. 2nd edition. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199641178.001.0001>.
- Fearnhead, P. (2011). "MCMC for state-space models". Teoksessa: *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC. Luku 21. URL: <https://eprints.lancs.ac.uk/id/eprint/8846/1/StateSpaceModels.pdf>.
- Geweke, J. (1989). "Bayesian inference in econometric models using Monte Carlo integration". *Econometrica* 57.6, s. 1317–1339. DOI: <https://doi.org/10.2307/1913710>.
- Gordon, N. J., Salmond, D. J. ja Smith, A. F. (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". 140.2, s. 107–113. DOI: <https://doi.org/10.1049/ip-f-2.1993.0015>.
- Guarniero, P., Johansen, A. M. ja Lee, A. (2017). "The iterated auxiliary particle filter". *Journal of the American Statistical Association* 112.520, s. 1636–1647. DOI: [10.1080/01621459.2016.1222291](https://doi.org/10.1080/01621459.2016.1222291).
- Haario, H., Saksman, E. ja Tamminen, J. (2001). "An adaptive Metropolis algorithm". *Bernoulli* 7.2, s. 223–242. DOI: <https://doi.org/10.2307/3318737>.
- Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO978107049994>.
- Higuchi, T. (1997). "Monte Carlo filter using the genetic algorithm operators". *Journal of Statistical Computation and Simulation* 59.1, s. 1–23. DOI: [10.1080/00949659708811843](https://doi.org/10.1080/00949659708811843).
- Karppinen, S., Lohi, O. ja M., V. (2019). "Prediction of leukocyte counts during paediatric acute lymphoblastic leukaemia maintenance therapy". *Scientific Reports* 9.18076. DOI: <https://doi.org/10.1038/s41598-019-54492-5>.
- Karppinen, S., Singh, S. S. ja Vihola, M. (2024). "Conditional Particle Filters with Bridge Backward Sampling". *Journal of Computational and Graphical Statistics* 33.2, s. 364–378. DOI: [10.1080/10618600.2023.2231514](https://doi.org/10.1080/10618600.2023.2231514).
- Karppinen, S. ja Vihola, M. (2021). "Conditional particle filters with diffuse initial distributions". *Statistics and Computing* 31.24. DOI: <https://doi.org/10.1007/s11222-020-09975-1>.
- Kitagawa, G. (1996). "Monte Carlo filter and smoother for non-Gaussian non-linear state space models". *Journal of computational and graphical statistics* 5.1, s. 1–25. DOI: <https://doi.org/10.1080/10618600.1996.10474692>.
- Møller, J., Syversveen, A. R. ja Waagepetersen, R. P. (1998). "Log Gaussian Cox processes". *Scandinavian Journal of Statistics* 25.3, s. 451–482. DOI: <https://doi.org/10.1111/1467-9469.00115>.

-
- Rabiner, L. R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition". *Proceedings of the IEEE* 77.2, s. 257–286. DOI: [10.1109/5.18626](https://doi.org/10.1109/5.18626).
- Särkkä, S. (2013). *Bayesian filtering and smoothing*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139344203>.
- Whiteley, N. (2010). "Discussion on "Particle Markov chain Monte Carlo methods"". *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72.3, s. 306–307. DOI: <https://doi.org/10.1111/j.1467-9868.2009.00736.x>.

Loppuviitteet

¹Väitöstutkimukseni osana toteutin GenericSSMs.jl Julia-paketin, joka sisältää geneeriset toteutukset näistä perusalgoritmeista (Karppinen, 2025).

²Estimaattori (4a) on tarkentuva eräin (myös käytännössä toteutuvien) ehdoin (Del Moral, 2004; Gerber, Chopin ja Whiteley, 2019) ja estimaattori (4b) on harhaton (Del Moral, 2004), (Vihola, Helske ja Franks, 2017, Proposition 21, (i)), kun hiukkassuotimessa käytetään harhatonta toisto-otantaa (ks. teksti).

³Nämä toisto-otanta algoritmit ja useat muut on toteutettu väitöskirjatyöni ohessa syntyneeseen Julia -kieliseen Resamplings.jl pakettiin (Karppinen ja Vihola, 2025).

⁴FK-esitystavassa riippuvuus havainnoista $y_{1:n}$ jätetään yleensä merkitsemättä.

⁵Algoritmiin 1 voi verrata tekemällä mielessään 'sijoituksen' $M_k = f_k$ ja $G_k = g_k$, joka on eräs FK-esitystapa.

⁶Jo mainittu Resamplings.jl Julia-paketti sisältää tehokkaita toteutuksia myös useista muista ehdollisista toisto-otannoista.

⁷Tämän merkittävän tuloksen oletuksia on laajennettu pikkuhiljaa: (Andrieu, Doucet ja Holenstein, 2010) (multinomiaalinen toisto-otanta ja jälkeläisjäljitys); (Whiteley, 2010) (laajennus takenevaan simulointiin); (Chopin ja Singh, 2015) (eräät ehdolliset toisto-otannat). Artikkelin (Karppinen, Singh ja Vihola, 2024) lauseet 2 ja 8 laajentavat tuloksen 'yleisen ehdollisen toisto-otannan' (ks. Karppinen, Singh ja Vihola, 2024, Määritelmä 1) tapaukseen.

⁸Tarkoitan epäinformatiivisella tässä tilannetta, jossa alkujakauman varianssi on suuri (suhteessa ensimmäisten havaintojen varianssiin ehdolla ensimmäiset tilamuuttujat).

Leo Törnqvist -palkinto

A Bayesian state-space model for Baltic ringed seal demography

MURAT ERSALMAN
MURATERSALMAN@GMAIL.COM
UNIVERSITY OF HELSINKI

Abstract

Integrated population models (IPMs) are a promising approach to assess and manage wildlife populations in dynamic and uncertain conditions. By combining multiple data sources into a unified model, they enable the parametrization of versatile, mechanistic population models that can predict population dynamics in novel circumstances (Besbeas et al., 2002; Schaub and Abadi, 2011). This is in contrast to traditional approaches where independent empirical estimates for demographic parameters are typically incorporated into a population projection matrix such as a Leslie matrix (Caswell, 2001). A major limitation of conventional methods is their inability to fully utilize all available information, as the synergies between different data sources are not exploited.

The Baltic ringed seal (*Pusa hispida botnica*) presents an example illustrating the limitation of conventional monitoring approaches. Despite the availability of long-term monitoring data, population assessment is hindered by dynamic environmental conditions, varying reproductive rates, and the recently re-introduced hunting, thus limiting the quality of information available to managers regarding, for example, hunting quotas. In particular, population counts of ringed seals from aerial surveys have exhibited unexpected trends and large fluctuations during the last decade, making it impossible to obtain reliable estimates of population growth from survey data alone.

This thesis presents a Bayesian IPM for the ringed seal population inhabiting the Bothnian Bay in the Baltic Sea (Ersalman, 2024; Ersalman et al., 2025). The central aim of this work is to outline an approach that can overcome some of the challenges that have crippled Baltic ringed seal monitoring efforts during the last decade, and support science-based management decisions. The thesis broadly consists of three parts. First, a state-space model is presented for the Bothnian Bay ringed seal population. Demographic processes are described through a stochastic age and sex structured population model that includes both hunting mortality and the hypothesized effects of environmental variables such as pollution and sea ice cover on demographic parameters and seal behaviour. Next, the model is fit to census and various demographic and reproductive data, as well as hunting statistics, from 1988 to 2023 under a Bayesian framework where posterior samples of model parameters are obtained using Markov Chain Monte Carlo methods. Finally, posterior estimates of model parameters are used

to construct a Leslie matrix, and model behavior is analyzed using methods developed for matrix projection models (Caswell, 2001). Future population dynamics are also simulated under alternative management scenarios to inform ringed seal management decisions. In general, this thesis demonstrates the value of mechanistic IPMs for monitoring and managing natural populations under changing environments, and supporting science-based management decisions.

Bibliography

- Besbeas, P. ym. (2002). "Integrating Mark–Recapture–Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters". *Biometrics* 58.3, s. 540–547. DOI: [10.1111/j.0006-341X.2002.00540.x](https://doi.org/10.1111/j.0006-341X.2002.00540.x).
- Caswell, H. (2001). *Matrix Population Models: Construction, Analysis, and Interpretation*. Sinauer Associates.
- Ersalman, M. (2024). "A Bayesian state-space model for Baltic ringed seal demography". MSc thesis. Helsingin yliopisto. URL: <https://helda.helsinki.fi/items/a72b0b03-2eba-4123-94fd-14af87fcbdc1>.
- Ersalman, M. ym. (2025). "Integrated population model reveals human- and environment-driven changes in Baltic ringed seal *Pusa hispida botnica* demography and behavior". *Marine Ecology Progress Series* 764, s. 213–236. DOI: [10.3354/meps14886](https://doi.org/10.3354/meps14886).
- Schaub, M. ja Abadi, F. (2011). "Integrated population models: A novel analysis framework for deeper insights into population dynamics". *Journal of Ornithology* 152.1, s. 227–237. DOI: [10.1007/s10336-010-0632-7](https://doi.org/10.1007/s10336-010-0632-7).

Suomen Tilastoseuran hallitus vuonna 2025

BOARD MEMBERS OF THE FINNISH STATISTICAL SOCIETY IN 2025

Puheenjohtaja Chair	Pekka Pere	Doctor of Philosophy DPhil
Varapuheenjohtaja Vice Chair	Ari Jaakola	Filosofian maisteri MSc
Sihteeri Secretary	Roope Rihtamo	Valtiotieteiden maisteri MSocSci
Rahastonhoitaja Treasurer	Arvi Tolvanen	Luonnontieteiden kandidaatti BSc
Jäsen Member	Mika Gissler	Filosofian tohtori PhD
Jäsen Member	Reija Helenius	Yhteiskuntatieteiden maisteri MSocSci
Jäsen Member	Tommi Härkänen	Filosofian tohtori PhD
Jäsen Member	Jyrki Möttönen	Filosofian tohtori PhD
Jäsen Member	Pekka Nieminen	Filosofian tohtori PhD
Varajäsen Deputy member	Kari Djerf	Valtiotieteiden lisensiaatti LicSocSci
Varajäsen Deputy member	Elisa Falck	Valtiotieteiden kandidaatti BSocSci
Klubimestari Club steward	Arvi Tolvanen	Luonnontieteiden kandidaatti BSc

Gunnar Modeen -minnesmedaljen

JUKKA HOFFRÉN
TILASTOKESKUS

Statistiska Samfundet i Finland r.f. har i samband med de nordiska statistikdagarna traditionsenligt delat ut Gunnar Modeen -minnesmedaljen till särskilt meriterade statistiker. Praxisen har varit att dela ut medaljen till en representant för det land där statistikdagarna hålls.

Gunnar Modeen -minnesmedaljen beviljas för en betydande livsgärning inom statistikbranschen. Meningen är att den person som belönas är en framstående senior expert inom statistikbranschen, som uttryckligen utmärkt sig i det praktiska statistik-arbetet och som uppskattas av sina kolleger.

Styrelsen för Statistiska Samfundet väljer den person som får medaljen och medaljen överläts i samband med ett nordiskt statistikermöte. Enligt fondens stadga överläts medaljen till en betydande nordisk statistiker från det land som respektive år arrangerar mötet. Den första medaljen överläts vid det nordiska statistikermöte som hölls i Finland år 1989.

Bakgrunden till och kriterier för GM-minnesmedaljen

Efter Gunnar Modeens bortgång år 1988 grundades en medaljfond till hans minne. Medaljen utarbetades på basis av den medaljong som Gunnar Modeens familj gett konstnären Matti Haupt i uppdrag att utforma till Modeens 70-årsdag år 1965. Mot-tagaren av medaljen väljs av styrelsen för Statistiska Samfundet i Finland och medaljen överläts i samband med ett nordiskt statistikermöte. Enligt fondens stadga överläts medaljen till en betydande nordisk statistiker från det land som respektive år arrangerar mötet. Den första medaljen överläts vid Nordiska Statistikermötet i Finland år 1989. Priset utdelas vart tredje år till en meriterad statistiker från det land där Nordiska Statistikermötet anordnas.

Allmänna kriterier för Gunnar Modeen -minnesmedaljen:

- priset beviljas för en betydande livsgärning inom statistikbranschen.

Den person som tilldelas medaljen:

-
- är en expert inom statistikbranschen, som uttryckligen utmärkt sig i det praktiska statistikarbetet
 - är en nordisk, framstående senior expert som uppskattas av sina kolleger,
 - har akademisk examen (magister, licentiat eller doktor) och
 - är villig att ta emot GM-medaljen

Mottagare av GM-minnesmedaljen

Den första medaljen tilldelades Mauno Koivisto, Finlands dåvarande president, som en särskild hedersbetygelse. År 1989 var han beskyddare av Nordiska Statistikermö-tet i Finland som firade 100-årsjubileum för nordisk statistik. Ytterligare en medalj delades ut på mötet och mottagare var professor Eino H. Laurila. Övriga mottagare av medaljen:

År 1992 tilldelades medaljen inte.

År 1995 direktör Poul Jensen, Danmarks Statistik.

År 1998 professor Sven Nordbotten, Universitetet i Bergen.

År 2001 professor Emeritus Gunnar Kulldorf, Umeå universitet.

År 2004 direktör Asta Manninen, Helsingfors stads faktacentral.

År 2007 generaldirektör Hallgrímur Snorrason, Hagstofa, Island.

År 2010 direktör Lars Thygesen, Danmarks Statistik.

År 2013 Liv Hobbestad Simpson, pensionerad från Statistisk sentralbyrå (SSB) som Head of National accounts och past chair of IARIW

År 2016 Eva Elvers, PhD, pensionerad från Design and Plan & Build and Test som Process owner

År 2019 professor Risto Lehtonen, Helsingfors universitet

År 2022 tilldelades medaljen inte.

Scandinavian Journal of Statistics

Recognised as a leading journal in its field, the Scandinavian Journal of Statistics is an international publication devoted to reporting significant and innovative original contributions to statistical methodology — both theory and applications. The journal specializes in statistical modelling, showing particular appreciation of the underlying substantive research problems. Scandinavian Journal of Statistics is published on behalf of the Danish Society for Theoretical Statistics, the Finnish Statistical Society, the Norwegian Statistical Society, and the Swedish Statistical Society. The journal is currently edited by professors Peter Dalggaard and Niels Richard Hansen. The national editor for Finland is Jukka Corander (University of Helsinki, Finland), and the other national editors are Jacob von Bornemann Hjelmberg (University of Southern Denmark, Denmark), Geir Olve Storvik (University of Oslo, Norway), and Jimmy Olsson (KTH Royal Institute of Technology, Sweden). The chairman of the board is Thomas Scheike (University of Copenhagen, Denmark) and the board members are Juha Karvanen (University of Jyväskylä, Finland), Hans Karlson (University of Bergen, Norway), and Sara Sjöstedt de Luna (Umeå University, Sweden).

Scandinavian Journal of Statistics is published quarterly in March, June, September and December by Wiley-Blackwell Publishers, 108, Cowley Road, Oxford OX4, 1JF, UK or 238 Main Street, Cambridge, MA 02142, USA.

Members of the Finnish Statistical Society are entitled to discount prices when ordering the Scandinavian Journal of Statistics. For further information, please see the webpage at <http://www.wiley.com/bw/subs.asp?ref=0303-6898&site=1>

ISI Journal Citation Reports® Ranking: 2022: 97/125 (Statistics & Probability).

Impact Factor: 1.040 (2021).

Online ISSN: 1467-9469.

Palkinnot

Leo Törnqvist –palkinnot

- 1978 Rene Tigerstedt, Helsingin yliopisto. En modell för valbeteende i trafiken.
- 1979 Pirkko Kirjavainen, Turun kauppakorkeakoulu. Mallin rakentaminen ja ennusteen laatiminen Suomen sähkön kulutukselle kahta aikasarja-analyysimenetelmää käyttäen.
- 1980 Esa Läärä, Helsingin yliopisto. Ikä-, aika- ja kohorttitekijöiden vaikutukset Suomen miesten keuhkosityöpäsairastavuudessa vuosina 1953–76.
- 1981 Arvi Suvanto, Tampereen yliopisto. Kausivaihtelu aikasarjamalleissa.
- 1982 Maija Salo, Helsingin yliopisto. Yritys prioritetiedon käytöstä alkoholi-juomien kulutusta selittävän kysyntämallin tukena. Jamel Boucelham, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1983 Vesa Vihriälä, Helsingin yliopisto. Aikasarjojen välisen riippuvuuden mitaus ja testaus: sovellus suomalaisiin rahatalouden sarjoihin. Pirkko Welin, Tampereen yliopisto: Tunnustuspalkinto.
- 1984 Jari Palsio, Turun kauppakorkeakoulu. Skenaarioiden rakentaminen ristivaikutusanalyysimallia käyttäen.
- 1985 Kenneth Nordström, Helsingin yliopisto. Gauss-Markov-mallien erikoisongelmista.
- 1986 Tapio Nummi, Tampereen yliopisto. APL-pohjainen ohjelmisto GMANOVA-mallille.
- 1987 Ari Veijanen, Helsingin yliopisto. Pickardin kentän soveltamisesta kuva-analyysissä. Kari Nissinen, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1988 Jaason Haapakoski, Helsingin yliopisto. Binomijakautuneiden muuttujien muutospisteongelma.
- 1989 Pasi Korhonen, Helsingin yliopisto. Kemometrian tilastollisista menetelmistä.
- 1990 Päivi Partanen, Jyväskylän yliopisto. Suljetun populaation koon estimointi merkintä-takaisinpyynti-menetelmällä: log-lineaarinen lähestymistapa. Markku Nurhonen, Tampereen yliopisto: Tunnustuspalkinto.

-
- 1991 Elina Järvinen, Helsingin yliopisto. Rajoitettujen, stokastisten ja konveksien estimaattoreiden käytöstä polynomisen viipymämallin parametrien estimoinnissa simulointikokeiden valossa.
- 1992 Jouni Kuha, Helsingin yliopisto. Binääristen regressiomallien selittäjien mittausvirheet ja parametriestimaattien mittausvirhekorjaukset. Juha Heikkinen, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1993 Palkintoa ei jaettu (yhtään ehdotusta ei saatu).
- 1994 Ilkka Taskinen, Jyväskylän yliopisto. Äärelliset Markovin ketjut ja anelointi.
- 1995 Mika Rautakorpi, Teknillinen korkeakoulu. Application of Markov chain techniques in certification of software. Tuija Jäppilä, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1996 Veli-Matti Suppola, Jyväskylän yliopisto. Robustit menetelmät. Jakaumien vinouden vaikutuksesta korrelaatiomatriisin estimointiin.
- 1997 Albert Höglund, Teknillinen korkeakoulu. An Anomaly Detection System for Computer Networks.
- 1998 Samuli Visuri, Oulun yliopisto. Robustista kovarianssimatriisin estimoinnista ja sen sovelluksista signaalinkäsittelyssä.
- 1999 Jani Raitanen, Tampereen yliopisto. Jalkapallo-ottelun lopputuloksen tilastollinen mallintaminen.
- 2000 Reijo Sund, Helsingin yliopisto. Tilastollisia menetelmiä dynaamisten potilaspopulaatioiden mallintamiseen. Tapahtumahistoria-analyysia hoitoilmoitusrekisterin skitsofreenikoille.
- 2001 Samu Mäntyniemi, Oulun yliopisto. A Hierarchial Bayes Model for Assessing Salmon (Salmo salar L.) Parr and Smolt Populations.
- 2002 Ilmari Juutilainen, Oulun yliopisto. Teräslevyjen lujuuden ennustaminen regressio- ja neuroverkkomalleilla.
- 2003 Leena Kalliovirta, Helsingin yliopisto. Mar-malli.
- 2004 Mikko Myrskylä, Jyväskylän yliopisto. Estimation of Class Frequencies with Micro Level Auxiliary Information.
- 2005 Antti Liski, Tampereen yliopisto. Lonkkamurtumapotilaiden hoitokustannusten vertailu vastaavuuspistemäärään perustuvalla menetelmällä.
- 2006 Karri Seppä, Oulun yliopisto. Suomalaisten paksusuolisyöpäpotilaiden ennusteen analyysi suhteellisen elossapysymisen ja syykohtaisen kuolleisuuden malleilla käyttämällä suurimman uskottavuuden ja Bayesin menetelmiä.
- 2006 Jukka Siren, Helsingin yliopisto. Populaatioiden geneettisen rakenteen spatiaalinen mallintaminen.
- 2007 Outi Ahti-Miettinen, Helsingin yliopisto. Kaksivaiheisen potenssiikiintiöinnin käyttö otoksen tehostamisessa - Esimerkkinä otoksen suunnittelu työvoimakustannusindeksin tietojen keruulle.

-
- 2008 Paul Catani, Svenska handelshögskolan. Enhetsrotttest och initialvärdet Tillämpning på arbetslösheten i Finland
- 2009 Elina Ahola, Jyväskylän yliopisto. Eksponenttisen perheen tila-avaruusmallien sovellus alkoholikuolleisuusaineistoon Matias Leppisaari, Aalto yliopiston teknillinen korkeakoulu: Tunnustuspalkinto.
- 2010 Sanna Peltomäki, Tampereen yliopisto. Estimation of Below Threshold Intra-EU Trade.
- 2011–2012 Tytti Pasanen, Tampereen yliopisto. Two-Level Structural Equation Modeling with Non-Normal Observed Variables for Assessing Poverty in Laos.
- 2013–2014 Joni Virta, Turun yliopisto. Some tools for linear dimension reduction.
- 2015–2016 Niko Lietzén, Aalto-yliopisto. New Approach to Complex Valued ICA: From FOBI to AMUSE
- 2015–2016 Santtu Tikka, Jyväskylän yliopisto. Kausaalivaikutusten identifiointi algoritmisesti.
- 2017–2018 Savi Virolainen, Helsingin yliopisto. GMAR- ja StMAR-mallin yleistäminen G-StMAR-malliksi.
- 2021–2022 Topi Halme, Quickest Detection under False Discovery Rate and Communication Constraints.
- 2023 Antti Yläjärvi, Transformer -neuroverkon robustisuuden parantaminen Gaussisen prosessin ja neuroverkon yhdistävillä menetelmillä.
- 2024 Murat Ersalman, Helsingin yliopisto. A Bayesian state-space model for Baltic ringed seal demography.

Väitöskirjapalkinnot

- 2009–2012 Jukka Sirén, Helsingin yliopisto. Statistical models for inferring the structure and history of populations from genetic data.
- 2013–2016 Johan Pensar, Åbo Akademi. Structure Learning of Context-Specific Graphical Models.
- 2017–2020 Anna-Kaisa Ylätaló, Jyväskylän yliopisto. Statistical inference for eye movement sequences using spatial and spatiotemporal point processes.
- 2017–2020 Santtu Tikka, Jyväskylän yliopisto. Improving identification algorithms in causal inference.
- 2021–2024 Santeri Karppinen, Jyväskylän yliopisto. Non-linear state-space methods for Bayesian time series modelling.

Suomen Tilastoseuran julkaisuja

PUBLIKATIONER UTGIVNA AV STATISTISKA SAMMANFUNDET

PUBLICATIONS ISSUED BY THE FINNISH STATISTICAL SOCIETY

1. *Monikielinen väestötieteen sanakirja*, suomenkielinen laitos, Helsinki 1962.
Multilingual Demographic Dictionary, Finnish section, Helsinki 1962.
2. *Suomen Tilastoseura – Statistiska Sammanfundet i Finland 1920-1970*, Porvoo - Borgå 1970.
3. *Pohjoismainen tilastosanasto*, toinen tarkistettu laitos.
Nordisk statistik nomenklatur, andra reviderade upplagan.
Nordic statistical nomenclature, 2nd revised edition. Jyväskylä 1975
4. *Aikasarja-analyysin menetelmiä*, Helsinki 1977.
5. Pekka Tavai: *Leo Törnqvist Posti- ja lennätinhallituksen liiketaloudellisen tutkimuslaitoksen esimiehenä 1949-1977*, Helsinki 1982.
6. Vesa Kuusela ja Leif Nordberg (toim.): *Otanta teoriassa ja käytännössä*. Helsinki 1986.
7. *Suomen Tilastoseura 70 vuotta. Statistiska Sammanfundet i Finland 70 år. The Finnish Statistical Society 70 years*. Helsinki 1991.
8.
 - Juha Alho, Elja Arjas, Esa Läärä ja Pekka Pere: *Tilastotieteen sanasto*. Suomen Tilastoseuran julkaisuja no. 8. Suomen Tilastoseura. 1. laitos, Helsinki 2021.
 - Juha Alho, Elja Arjas, Esa Läärä ja Pekka Pere: *Tilastotieteen sanasto*. Suomen Tilastoseuran julkaisuja no. 8. Suomen Tilastoseura. 2. laitos, Helsinki 2023.
 - Juha Alho, Elja Arjas, Juha Karvanen, Lasse Leskelä, Esa Läärä ja Pekka Pere: *Tilastotieteen sanasto*. Suomen Tilastoseuran julkaisuja no. 8. Suomen Tilastoseura. Verkossa <https://sanasto.tilastoseura.fi/>.

Tilastotieteellisiä tutkimuksia

STATISTISKA UNDERSÖKNINGAR

STATISTICAL RESEARCH REPORTS

ISSN 0356-3499

1. Pentti Manninen: Puolueiden kannatusosuuksien estimoinnin tarkkuus Demingin vyöhykepoiminnassa. (The Accuracy of Party Support Estimation in Deming Zone Selection. In Finnish with English Summary.) Helsinki 1976.
2. Timo Hakulinen: On Competing Risks of Death. Helsinki 1977.
3. Lars-Erik Öller: Time Series Analysis of Finnish Foreign Trade. Helsinki 1978.
4. Pekka Laippala: The Empirical Bayes Two-Action Rules with Floating Optimal Sample Size and Exponential Conditional Distributions. Helsinki 1980.
5. Markku Nurminen: Some Developments in Quantitative Methods of Epidemiology. Helsinki 1982.
6. Pentti Saikkonen: Comparing Asymptotic Properties of Some Tests Used in the Specification of Time Series Models. Helsinki 1985.
7. Lauri Tarkkonen: On Reliability of Composite Scales. Helsinki 1987.
8. Juni Palmgren: Models for Categorical Data with Errors of Observation. Helsinki 1987.
9. Ari Veijanen: On Estimation of Parameters of Partially Observed Random Fields and Mixing Processes. Helsinki 1989.
10. Ritva Luukkonen: On Linearity Testing and Model Estimation in Non-Linear Time Series Analysis. Helsinki 1990.
11. Hely Salomaa: Factor Analysis of Dichotomous Data. Helsinki 1990.
12. Kenneth Nordström: Contributions to the Comparison of Linear Models and to the Löwner-Ordering Antitonicity of Generalized Inverses. Helsinki 1990.
13. Seppo Laaksonen: Handling Household Survey Nonresponse Data. Helsinki 1992.
14. Mervi Eerola: On Predictive Causality in the Statistical Analysis of a Series of Events. Helsinki 1993.
15. Mikael Linden: Studies in Integrated and Co-Integrated Economic Time Series. Helsinki 1995.
16. Tadeusz Dyba: Precision of Cancer Incidence Predictions Based on Poisson Distributed Observations. Helsinki 2000.
17. Kimmo Vehkalahti: Reliability of Measurement Scales. Helsinki 2000.

-
18. Sirpa Heinävaara: Modelling survival of patients with multiple cancers. Helsinki 2003.

Suomen Tilastoseuran vuosikirja

ÅRSBOK FÖR STATISTISKA SAMMANFUNDET I FINLAND

THE YEARBOOK OF THE FINNISH STATISTICAL SOCIETY

ISBN 0355-5941

1975 Helsinki 1976	1997 Helsinki 1998
1976 Helsinki 1977	1998 Helsinki 1999
1977 Helsinki 1978	1999–2000 Helsinki 2000
1978 Helsinki 1979	2001 Helsinki 2002
1979 Helsinki 1980	2002 Helsinki 2003
1980 Helsinki 1981	2003 Helsinki 2004
1981 Helsinki 1982	2004 Helsinki 2005
1982 Helsinki 1983	2005 Helsinki 2006
1983 Helsinki 1984	2006 Helsinki 2007
1984 Helsinki 1985	2007 Helsinki 2008
1985 Helsinki 1986	2008 Helsinki 2009
1986 Helsinki 1987	2009 Helsinki 2010
1987 Helsinki 1988	2010 Helsinki 2011
1988–1989 Helsinki 1990	2011–2012 Helsinki 2012
1990 Helsinki 1991	2013–2014 Helsinki 2014
1991 Helsinki 1992	2015–2016 Helsinki 2017
1992 Helsinki 1993	2017–2018 Helsinki 2018
1993 Helsinki 1994	2019–2022 Helsinki 2023
1994 Helsinki 1995	2023 Helsinki 2024
1995 Helsinki 1996	2024 Helsinki 2025
1996 Helsinki 1997	2025 Helsinki 2026

Tilastoseuran julkaisuja voi tiedustella sihteeriltä sähköpostitse osoitteesta suomentilastoseura@gmail.com. Joidenkin julkaisujen painokset ovat tosin jo loppuneet.

Muita julkaisuja

ANDRA PUBLIKATIONER

OTHER PUBLICATIONS

Suomen Tilastoseura 1920–1945, Helsinki 1946

Statistiska Sammanfundet i Finland 1920–1945, Helsingfors 1946.

Pohjoismainen tilastosanasto – Nordisk statistisk nomenklatur, Kööpenhamina 1954.

13:e Nordiska statistikermötet i Helsingfors 14–16 juni 1973, Jyväskylä 1974

The 13th Joint Meeting of the Nordic Statistical Societies in Helsinki June 1973, Jyväskylä 1974.

Det 18:e nordiska statistikmötet i Esbo, Hundraårsjubileum, Helsingfors 1990.

The Joint Conference of the Nordic Statisticians in Espoo, Finland 1989, Helsinki 1990.



ISSN 0355 – 5941 (painettu)
ISSN 2953 – 9307 (verkossa)