



HIUKKASSUODIN- JA HIUKKASSILOITINALGORITMIT SEKÄ NIIDEN
SOVELTAMINEN AoA-MENETELMÄÄN PERUSTUVASSA
BLUETOOTH-SISÄTILAPAIKANNUKSESSA

Lasse Rintakumpu

Pro gradu -tutkielma
Syyskuu 2024

Tarkastajat:
Ohjaajan titteli (Prof./Dos./FT) ja nimi
Toisen tarkastajan titteli (Prof./Dos./FT) ja nimi

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä

TURUN YLIOPISTO
Matematiikan ja tilastotieteen laitos

LASSE RINTAKUMPU: Hiukkassuodin- ja hiukkassiloitinalgoritmit sekä niiden soveltaminen AoA-menetelmään perustuvassa Bluetooth-sisätilapaikannuksessa
Pro gradu -tutkielma, X s.

Tilastotiede

Syyskuu 2024

Tutkielmassa esitetään hiukkassuodin- ja hiukkassiloitinalgoritmien teoria Bayesilaisessa tilastotieteellisessä viitekehysessä. Lisäksi tutkielmassa käsitellään hiukkassuodtimien varianssin estimointia.

Empiirisänä esimerkkinä tutkielmassa tarkastellaan hiukkassuodin- ja hiukkassiloitinalgoritmien käyttöä AoA-teknologiaan perustuvassa Bluetooth-sisätilapaikannusratkaisussa.

Asiasanat: SMC-menetelmät, Monte Carlo -menetelmät, sekventiaalinen Monte Carlo, suodinongelma, hiukkassuodin, hiukkassiloitin, SIR-algoritmi, sisätilapaikannus, BLE, AoA, triangulaatio, Bayesilainen päätely

Sisällys

1 Johdanto	3
1.1 Notaatioista	3
1.2 Suodinongelma	4
1.3 Suodin- ja siloitteluongelmien historiaa	7
1.4 Monte Carlo -menetelmistä	9
1.4.1 Monte Carlo -approksimaatio	10
1.4.2 Tärkeytysotanta	10
1.5 Bayesilainen suodin	11
1.6 Kalman-suotimen ja hiukkassuotimen yhteydestä ja eroista	12
2 Hiukkassuotimet	15
2.1 SIR-algoritmi	15
2.1.1 Parametrien valinta	17
2.1.2 Konvergenssituloksia	20
2.1.3 Marginaalijakauma	21
2.1.4 Aikakompleksisuus	21
2.2 Saapasremmisuodin	21
2.3 Varianssin estimoinnista	22
3 Hiukkasilottimet	25
3.1 Bayesilainen siloitin	25
3.2 Offline-algoritmit	26
3.2.1 SIR-siloitin	26
3.2.2 BS-PS-siloitin	27
3.2.3 Uudelleenpainottava hiukkassiloitin	27
3.3 Online-algoritmit	28
3.3.1 Kiinteän viipeen siloitin	28
3.3.2 Mukautuvan viipeen siloitin	29

4 Hiukkassuodin ja -siloitin sisätilapaikannuksessa	31
4.1 Sisätilapaikannuksesta	31
4.2 Teknologian kuvaus	32
4.2.1 AoA-menetelmistä	35
4.2.2 Kalibraatiosta	36
4.3 Datan kuvaus	36
4.3.1 Karttaprojektiosta	39
4.3.2 Muunnettua dataa	40
4.4 Sisätilapaikannusalgoritmi	41
4.4.1 Ongelman kuvaus	41
4.4.2 Uskottavuusmallit	43
4.4.3 Datan valinta	44
4.4.4 Dynaaminen malli	44
4.4.5 Siloittelumalli	47
4.4.6 WB-sisätilapaikannusalgoritmi	47
4.5 Empiirinen esimerkki	51
4.5.1 Koeasetelma	51
4.5.2 Parametrien valinta	54
4.5.3 Tulokset	55
5 Lopuksi	67
Liite A - Karttapolut	69
Vaihe 1	69
Vaihe 2	71
Vaihe 3	74

Luku 1

Johdanto

Hiukkassuotimet ovat joukko Monte Carlo -algoritmeja, joiden avulla voidaan ratkaista ns. suodinongelma, kun ongelma on epälineaarinen ja/tai ongelmaan liittyvä kohina ei noudata normaalijakaumaa. Hiukkassuotimille on lukuisia sovellutuksia esimerkiksi Bayesilaisessa tilastotieteessä, fysiikassa ja robotiikassa.

Tämän tutkielman tavoitteena on esittää pääpiirteittään hiukkassuotimien sekä näihin läheisesti liittyvien hiukkassiloittimien teoria. Lisäksi tutkielman käsittelyssä joitakin menetelmäperheeseen kuuluvia algoritmeja ja sovelletaan näitä sisätilapai-kannuksessa.

Tutkielman ensimmäisessä luvussa kuvataan yleisellä tasolla sekä suodinongelma että sen ratkaisujen historiaa ja esitetään joitakin Monte Carlo -menetelmiin liittyviä yleisiä tuloksia sekä Bayesilainen viitekehys suodinongelmalle. Toisessa luvussa kuvataan kaksi hiukkassuodinalgoritmia, saapasremmisuodin sekä SIR-algoritmi ja perehdytään hiukkassuotimen varianssin estimointiin. Kolmannessa luvussa tar-kaatellaan siloitteluongelmaa ja esitetään hiukkassiloitinalgoritmeja tämän ongelman ratkaisemiseksi. Neljäs luku keskittyy hiukkassuotimen käyttöön empiirisessä AoA/Bluetooth-teknologiaan perustuvassa sisätilapaikanussovelluksessa. Luvussa käsitellään lisäksi lyhyesti sisätilapaikanuksessa hyödynnettävää karttasovitusalgo-ritmia.

Hiukkassuodin- ja hiukkassiloitinalgoritmien osalta tutkielman esitykset seuraavat erityisesti Simo Särkän kirjaa *Bayesian Filtering and Smoothing* (2013) [29], Fredrik Gustafssonin artikkelia “Particle Filter Theory and Practice with Positioning Applications” (2010) [14] sekä Olivier Cappén, Simon J. Godsill ja Eric Moulines’n artikkelia “An overview of existing methods and recent advances in sequential Monte Carlo” (2007) [3]. Hiukkassuotimien varianssin estimointi seuraa erityisesti Nick Whiteley ja Anthony Leen artikkelia “Variance estimation in the particle filter” (2018) [19] sekä Randal Doucet ja Jimmy Olssonin artikkelia “Numerically stable online estimation of variance in particle filters” (2019) [26].

1.1 Notaatioista

Tässä tutkielman käytetään seuraavia yleisiä notaatioita. Hiukkassuotimien vekto-reita merkitään pienellä kursivoidulla kirjaimella x , y ja w . Hiukkassuotimen hiuk-

kaset sisältäviä vektoreita merkitään x_k^i , missä alaindeksi viittaa aika-askeleeseen $k, k = \{1, \dots, T\}$ ja yläindeksi partikkeliin i , missä $i = \{1, \dots, N\}$. Vastaavasti aika-askeleiden $k, k = \{1, \dots, T\}$ havainnot sisältäviä vektoreita merkitään $\{y_1, \dots, y_k\}$ ja painovektoria w_k .

Pienin kursivoiduin kirjaimin viitataan myös funktioihin ja reaaliarvoisiin skaalaareihin. Pienet kirjaimet f, h, g, p ja q on varattu ensisijaisesti funktioille, p ja q erityisesti todennäköisyysfunktioille. Skalaareihin viitataan myös isoilla kursivoidulla kirjaimilla, esimerkiksi T ja N . Joukkoihin viitataan myös isoilla kursivoidulla kirjaimilla, esimerkiksi S . Se, käytetäänkö isoa kursivoitua kirjainta ilmaisemaan skalaaria vai joukkoa ilmenee asiayhteydestä. Matriiseja merkitään isolla lihavoidulla kirjaimella, esimerkiksi \mathbf{X} ja funktiota merkitään pienellä kursivoidulla kirjaimella $f(\cdot)$. Prosesseihin viitataan alaindeksoidulla isolla kirjaimella, esimerkiksi X_k . Taulukossa 1.1 esitetään tarkemmin tutkielman keskeisimmät merkinnät. Taulukossa 1.2 esitetään tutkielmassa käytetyt lyhenteet.

1.2 Suodinongelma

Stokastisten prosessien teoriassa suodinongelmaksi kutsutaan tilannetta, jossa halutaan muodostaa keskineliövirheen mielessä paras mahdollinen estimaatti jonkin järjestelmän tilan arvoille, kun ainoastaan osa tiloista voidaan havaita ja/tai havaintoihin liittyy kohinaa. Tavoitteena on toisin sanoen laskea jonkin prosessin posteriorijakauma kyseisten havaintojen perusteella. Ongelmaa havainnollistaa kaavio (1.1).

$$\begin{array}{ccccccc} x_1 & \longrightarrow & x_2 & \longrightarrow & x_3 & \longrightarrow & \dots & \text{piilossa olevat tilat} \\ \downarrow & & \downarrow & & \downarrow & & \\ y_1 & & y_2 & & y_3 & & \dots & \text{havainnot} \end{array} \quad (1.1)$$

Tässä tutkielmassa keskitytään erityisesti epälineaarisen, ns. Markovin piilomallin posteriorijakauman Bayesilaiseen ratkaisuun. Ongelmassa tiedetään, miten havaitut muuttujat y_k kytkeytyvät ”piilossa oleviin” tilamuuttuijiin x_k sekä osataan sanoa jotain tilamuuttujien todennäköisyysistä. Oletetaan myös, että piilossa oleville tiloille X_k pätee Markov-ominaisuus, jolloin kutakin hetkeä seuraava tila x_{k+1} riippuu menneistä tiloista $x_{1:k}$ ainoastaan tilan x_k välityksellä. Lisäksi havaittu tila y_k riippuu tiloista x_k ainoastaan jonkin x_k :n funktion kautta. Kun aika-avaruus on diskreetti ja aika-askeleella $k = \{1, \dots, T\}$ piilossa olevan prosessin tilaa merkitään x_k ja havaittua prosessia y_k , saadaan mallit

$$x_{k+1} = f(x_k, \nu_k), \quad (1.2)$$

$$y_k = h(x_k) + e_k. \quad (1.3)$$

Lisäksi tiedetään prosessin alkuhetken jakauma $x_0 \sim p_{x_0}$, tähän liittyväni kohinaprosessin jakauma $\nu_k \sim p_{\nu_k}$ sekä malliin y_k liittyvä kohina $e_k \sim p_{e_k}$. Koska hiukkassuodinalgoritmit pyrkivät ratkaisemaan juurikin epälineaarisen, ei-Gaussisen

Taulukko 1.1: Symbolit ja notaatiot

Merkintä	Selitys
$1 : k$	$1, \dots, k$
k	Aika-askel, skalaari, $k = 1, \dots, T$
N	Hiukkassuotimen käytämä partikkalien lukumäärä / otoskoko
x_k^i	Hiukkassuotimen hiukkaset sisältävä vektori, alaindeksi k määrittää aika-askeleen, yläindeksi $i = 1, \dots, N$ partikkelin
y_k	Havainnot sisältävä vektori, k määrittää aika-askeleen
w_k^i	Hiukkassuotimen painovektori
$p(x), q(x)$	x :n tiheysfunktioita
$\hat{p}(x), \hat{q}(x)$	x : tiheysfunktion estimaatteja
$p(x y)$	x :n tiheysfunktio ehdolla y
$x_{k k-1}$	x_k :n arvo ehdollistettuna arvoilla aika-askeleeseen $k - 1$ asti
$\mathbb{E}[x]$	x :n odotusarvo
\mathbf{X}	Matriisi
\mathbf{X}^\top	Matriisin \mathbf{X} transpoosi
$\delta(x)$	Diracin deltafunktio
$\log(x)$	Luonnollinen logaritmifunktio
$\text{Var}(x)$	Varianssifunktio
$\mathcal{O}(\cdot)$	Algoritmin asymptoottisen suoritusajan Ordo-notaatio, suoritusajan mielessä pahin mahdollinen tapaus
$\mathcal{N}(\mu, \sigma^2)$	Normaalijakauma, keskiarvoparametri μ , varianssiparametri σ^2
$\mathcal{U}(a, b)$	Jatkuva tasajakauma, välin rajaavat parametrit a ja b
X_k	Dynaaminen prosessi
\hat{X}_k	Dynaamisen prosessin estimaatti
$\int f(x)dx$	Funktion $f(x)$ Lebesguen integraali yli \mathbb{R}^n
\mathbb{P}	Todennäköisyysfunktio $\mathcal{A} \rightarrow \mathbb{R}$
\mathcal{A}	σ -algebra
\mathbb{R}	Reaalilukujen joukko
\mathbb{R}^n	n-ulotteinen vektoriavaruus
$\mathcal{B}(X)$	Borel-mitta, pienin X :n avoimet joukot sisältävä σ -algebra
\propto	Yhtäsuuruus normalisoivaa vakiota lukuunottamatta
\emptyset	Tyhjä joukko
Ω_k	Painot aika-askelleella k normalisoiva tekijä
$\text{id}(x)$	Identiteettifunktio
$\text{erf}(x)$	Virhefunktio

Taulukko 1.2: Lyhenteet

Lyhenne	Selitys
ALvar	Adaptive-Lag variance
AoA	Angle-of-Arrival
BLE	Bluetooth Low Energy
BS-PS	Backwards simulation particle smoother
CLE	Chan and Lai estimate
CR	Lithium Rechargeable
CSI	Channel State Information
EKF	Extended Kalman filter
GFSK	Gaussian frequency-shift keying
GIS	Geographic information system
GPS	Global Positioning System
IMU	Inertial measurement unit
IoT	Internet of Things
IQ, I/Q	In-phase / Quadrature
Lidar	Light detection and ranging
LPDDR	Low-Power Double Data Rate
MAC	Medium access control address
MC	Monte Carlo
MUSIC, MuSiC	MULTiple SIgnal Classification
OD	Olsson and Douc
OSCU	On-site computing unit
PCB	Printed circuit board
PKF	Position Kalman filter
RSS(I)	Received signal strength (indicator)
RTSS	Rauch-Turn-Striebel smoother
SIR	Sequential Importance Resampling
SIS	Sequential Importance Sampling
SLF	Statistically linearized Kalman filter
SMC	Sequential Monte Carlo
SNR	Signal-to-noise ratio
TDoA	Time Difference of Arrival
ToA	Time of Arrival
ToTal	Three object Triangulation algorithm
UKF	Unscented Kalman filter
WB	Walkbase
WGS	World Geodetic System

suodinongelman, voivat funktiot $f(\cdot)$ ja $h(\cdot)$ olla epälineaarisia eikä kohinan tarvitse olla normaalijakautunutta.

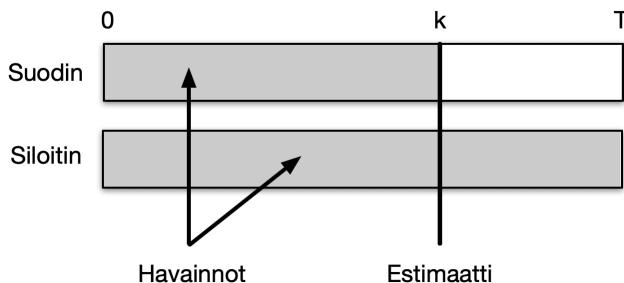
Mallit voidaan esittää myös yleisemmässä jakaumamuodossa

$$x_{k+1} \sim p(x_{k+1}|x_k), \quad (1.4)$$

$$y_k \sim p(y_k|x_k). \quad (1.5)$$

Tutkielman teoriaosassa käytetään ensisijaisesti yhtälöiden (1.4) ja (1.5) muotoilua. Empiirisessä osassa palataan yhtälöiden (1.2) ja (1.3) muotoiluun.

Suodinongelmaa lähellä on myös ns. siloitteluongelma (*smoothing problem*), jossa ollaan kiinnostuneita prosessin x_k posteriorijakaumasta $p(x_k|y_k)$ jokaisella aika-askeleena $\{1, \dots, k\}$ ei ainoastaan tietyllä aika-askeleella k . Hiukkassuodinalgoritmit näyttävät ratkaisevan siloitteluongelman ilmaiseksi, mutta tähän liittyy kuitenkin joidenkin mallien kohdalla mahdollista epätarkkuutta, joten tarvittaessa tasoitusongelma pitää ratkaista erikseen. Tähän ongelmaan palataan tutkielman luvussa 3. Kuva 1.1 selittää suodin- ja siloitteluongelmien eron. Kuva mukaillee Särkkää (2013) [29].



Kuva 1.1: Suodin- ja siloitteluongelma

1.3 Suodin- ja siloitteluongelmien historiaa

Tämä alaluku esittää pääpiirteitäan suodinongelmalle esitettyjen ratkaisujen historian. Lineaарisen suodinongelman osalta alaluku noudattaa Dan Crisanin artikkelia “The stochastic filtering problem: a brief historical account” (2014) [6] sekä Mohinder S. Grewalin ja Angus P. Andrewsin artikkelia “Applications of Kalman Filtering in Aerospace 1960 to the Present” (2010) [12]. Hiukkassuotimien osalta lähteenä toimii Cappé &al (2007) [3].

Suodinongelma nousi esille insinöörityteteiden sekä sotateollisuuden käytännön ongelmista 2. maailmansodan aikana, vaikkakin suodinongelman diskreetin ajan ratkaisut juontavat jo Andrei N. Kolmogorovin 30-luvun artikkeleihin. Jatkuvan ajan tilanteessa ensimmäisen optimaalisen, kohinan sallivan suotimen esitti matemaatikko, kybernetiikan kehittäjä Norbert Wiener. Wiener-suotimena tunnettua ratkaisuaan

varten Wiener muotoili seuraavat kolme ominaisuutta, jotka prosessin X_k estimaatin \hat{X}_k pitää toteuttaa.

1. *Kausaliteetti*: X_k tulee estimoida käyttäen arvoja Y_s , missä $s \leq k$.
2. *Optimaalisuus*: X_k :n estimaatin \hat{X}_k tulee minimoida keskineliövirhe $\mathbb{E}[(X - \hat{X}_k)^2]$.
3. *On-line-estimointi*: Estimaatin \hat{X}_k tulee olla saatavissa jokaisella aika-askeleella k .

Wiener sovelsi ratkaisussaan stationaaristen prosessien spektriteoriaa. Tulokset julkaistiin salaisina Yhdysvaltojen asevoimien tutkimuksesta vastanneen National Defense Research Committeeen (NDRC) raportissa vuonna 1942. Tutkimus tunnettiin sodan aikana lempinimellä ”Keltainen vaara” sekä painopaperinsa värin etä vaikeaselkoisuutensa vuoksi. Myöhemmin Wiener esitti tuloksensa julkisesti kirjassaan *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (1949). Wienerin alkuperäiset kolme perusperiaatetta pätevät edelleen kaikille suodinongelman ratkaisuille, myös hiukkassuotimille.

Kenties tärkein ja varmasti tunnetuin lineaariseen suodinongelman ratkaisu on Kalman-suodin. Suotimen kehittivät R.E. Kalman ja R.S. Bucy 1950- ja 60-lukujen taitteessa Yhdysvaltain kylmän sodan kilpavarustelutarpeisiin perustetussa Research Institute for Advanced Studies -tutkimuslaitoksessa (RIAS). Kalman-suodin on suodinongelman diskreetin ajan ratkaisu, kun taas Kalman-Bucy-suodin on jatkuvan ajan ratkaisu. Kohinan ollessa normaalijakautunutta on Kalman-suodin Wiener-suotimen tavoin lineaarisen suodinongelman optimaalinen ratkaisu. Wiener-suotimella ja Kalman-suotimella on kuitenkin erilaiset oletukset, minkä vuoksi erityisesti säätö- ja paikannusovelluksissa Kalman-suotimen käyttö on luontevampaa. Suotimien oletuksia ja oletusten välisiä eroja ei käsitellä tässä tutkielmassa, mutta alaluvussa [1.6](#) käsitellään Kalman-suotimen formaalia yhteyttä hiukkassuotimiin.

Kalman-suodinta voidaan soveltaa myös epälineaarisessa tapauksessa, kunhan suodinongelman funktiot $f(\cdot)$ ja $h(\cdot)$ ovat derivoituvia ja niihin liittyvä kohina oletetaan normaalijakautuneeksi. Tätä ratkaisua kutsutaan laajennetuksi Kalman-suotimeksi (*extended Kalman filter*, EKF). Suodin kehitettiin 60-luvulla NASA:n Apollo-ohjelman tarpeisiin, vaikkakin itse avaruusalusten laitteistot hyödynsivät lento-ratojen laskennassa Kalman-suotimen perusversiota. Laajennetun Kalman-suotimen toimintaperiaate perustuu epälineaaristen funktioiden linearisointiin Taylorin kehitelmän avulla kulloisenkin estimaatin ympärillä. Laajennettu Kalman-suodin on erityisesti paikannusovellusten *de facto*-suodinstandardi, mutta suodin ei kuitenkaan ole epälineaarisen ongelman optimaalinen estimaattori.

Kalman-suotimesta on lisäksi olemassa lukuisia muita epälineaarisiin ongelmiin soveltuivia laajennuksia, muun muassa paikkaratkaisun Kalman-suodin (*position Kalman filter*, PKF), hajustamaton Kalman-suodin (*unscented Kalman filter*, UKF) sekä tilastollisesti linearisoitu Kalman-suodin (*statistically linearized Kalman filter*, SLF). Kuitenkin jos prosessin X_t mallia ei tunneta tarkasti tai kohinaa ei voida olettaa normaalijakautuneeksi, ovat hiukkassuotimet eli sekventiaaliset Monte Carlo -menetelmät Kalman-suotimen johdannaisia parempia ratkaisuja. Vaikka

tila-avaruuden dimensioiden kasvaessa kasvaa myös hiukkassuotimien vaatima laskentateho, ovat hiukkassuotimet aina sitä parempia mitä epälineaarisempia mallit ovat ja mitä kauempana normaalijakaumasta kohina on. Viimeisten vuosikymmenten aikana myös laskennan teho on kasvanut merkittävästi samalla kun laskennan hinta on vastaavasti romahtanut, mikä puolaa Monte Carlo -menetelmien käyttöä entistä useammissa ongelmissa.

Jotakin suodinongelman rekursiivisia Monte Carlo -ratkaisuja löytyy jo 1950–70-luvulta, erityisesti säätöteoriaan piiristä. Olennainen nykyalgoritmeihin periytynyt oivallus varhaisissa suodinalgoritmeissa oli tärkeytysotannan käyttö halutun jakamaestimaatin laskennassa. Tärkeytysotanta-algoritmiin voidaan turvautua, kun emme pysty suoraan tekemään havaintoja jostakin jakaumasta p ja teemme sen sijaan havaintoja jakaumasta q , jota painotamme niin, että tuloksena saadaan jakauman p harhaton estimaatti. Algoritmi on kuvattu tarkemmin tutkielman luvussa 2.

Tärkeytysontaa käytävä suodinongelman ratkaiseva SIS-algoritmi (*sequential importance sampling*) ei kuitenkaan vielä 70-luvulla löytänyt suurta käytännön suosiota. Osin tämä johti puutteellisesta laskentatehosta, mutta algoritmi kärsi myös otosten ehtymisenä (*sample impoverishment*) tunnetusta ongelasta. Monissa ongelmissa SIS-algoritmia käytettäessä suuri osa painoista päättyy vain tietyille partikkeleille, jolloin vastaavasti suuri osa partikkeleista ei enää estimoi haluttua jakaumaa. Tähän ongelmaan palataan myöhemmin tutkielmassa.

Merkittävän ratkaisun ehtymisongelmaan esittivät Gordon, Salmond ja Smith artikkelissaan “Novel approach to nonlinear/non-Gaussian Bayesian state estimation” (1993) [11]. Artikkelin ratkaisu kulki nimellä *bootstrap filter*, saapasremmisuodin. Saapasremmisuodin vältti ehtymisen uudellenotannalla, jossa matalapainoiset partikkelit korvattiin otoksilla korkeapainoisemmista partikkeleista. Ratkaisussa painot eivät myöskään riippuneet partikkelien aiemmista poluista vaan ainoastaan havaintojen uskottavuusfunktiossa. Vastaavaa ratkaisua käytetään tämän tutkielman uudemmassa SIR-algoritmissa (*sampling importance resampling*), jossa myös uudelleenotantaaan sovelletaan tärkeytysontaa.

Sekventiaalisissa Monte Carlo -menetelmissä stokastisen prosessin posteriorijakauman esittämiseen käytettyjä otoksia kutsutaan myös partikkeleiksi tai hiukkasiksi ja menetelmiä siten hiukkassuotimiksi. Erityisesti myöhemmin esitettyvä SIR-algoritmia kutsutaan usein hiukkassuotimeksi. Termiä hiukkassuodin käytti ensimmäisen kerran Del Moral artikkelissa “Nonlinear Filtering: Interacting Particle Resolution” (1996) [22], SMC-menetelmät termiä Liu ja Chen artikkelissa “Sequential Monte Carlo Methods for Dynamic Systems” (1998) [20]. Tässä tutkielmassa käytetään yleisemmin käytettyä termiä hiukkassuotimet.

1.4 Monte Carlo -menetelmistä

Tässä alaluvussa kuvataan lyhyesti hiukkassuotimissa käytettävien Monte Carlo -menetelmien perusperiaate todennäköisyysjakauman estimoinnissa. Lisäksi esitetään tärkeytysotanta-algoritmi (*importance sampling*), jonka tarkoituksesta on estimoida harhattomasti jakaumaa $p(x|y_{1:k})$, josta emme voi suoraan tehdä otoksia, mutta jota voimme approksimoida toisella jakaumalla q . Esitykset noudattavat Särkkää (2013)

[29].

1.4.1 Monte Carlo -approksimaatio

Bayesilaisessa päätelyssä ollaan yleisesti kiinnostuttu laskemaan johonkin posteriooritihesjakaumaan p liittyvää odotusarvoa

$$\mathbb{E}[g(x)|y_{1:k}] = \int g(x)p(x|y_{1:k})dx, \quad (1.6)$$

missä g on tila-avaruuden mielivaltainen funktio ja $p(x|y_{1:t})$ on havaintoihin $\{y_1, \dots, y_k\}$ liittyvä x :n posterioritihesjakauma. Odotusarvo on laskettavissa suljetussa muodossa vain harvoissa tapauksissa, suodinongelman kohdalla silloin, kun kyseessä on lineaarinen ja Gaussinen malli. Odotusarvoa voidaan kuitenkin approksimoida niin sanoituilla Monte Carlo -menetelmillä. Menetelmien perusperiaatteet ovat tehdä riippumattomia otoksia estimoitavasta jakaumasta ja laskea haluttu odotusarvo otosten avulla. Jos tehdään N otosta jakaumasta $x^i \sim p(x|y_{1:t})$, missä $i = \{1, \dots, N\}$ saadaan näiden otosten avulla laskettua odotusarvon estimaatti

$$\mathbb{E}[g(x)|y_{1:k}] \simeq \frac{1}{N} \sum_{i=1}^N g(x^i). \quad (1.7)$$

Monte Carlo -estimaatti konvergoi keskeisen raja-arvolauseen nojalla ja sen estiointivirheen voidaan osoittaa olevan luokkaa $\mathcal{O}(\frac{1}{\sqrt{N}})$ riippumatta tilamuuttujan x dimensiosista. Hiukassuotimet hyödyntävät Monte Carlo -estimointia sekventiaalisesti, jolloin estimaatti lasketaan rekursiivisesti kullekin aika-askeleelle $k = \{1, \dots, T\}$. Tähän palataan luvuissa 3 ja 4.

1.4.2 Tärkeytysotanta

Tilanteessa, jossa Monte Carlo -otoksia ei voida tehdä suoraan jakaumasta p , voidaan hyödyntää jakaumaa p approksimoivaa tärkeytys- tai ehdotusjakaumaa $q(x|y_{1:k})$ sekä ns. tärkeytysotantaa. Oletetaan, että tunnetaan priorijakauma $p(x)$ ja on olemassa havaintomalli $p(y_{1:k}|x)$ sekä valittu ehdotusjakauma $q(x|y_{1:k})$, josta voidaan tehdä otoksia. Ehdotusjakaumalta edellytetään lisäksi, että sen kantaja on suurempi tai yhtä suuri kuin jakauman $p(x|y_{1:k})$ ja että se saa nollasta poikkeavia arvoja kaikkialla missä $p(x|y_{1:k})$ saa nollasta poikkeavia arvoja. Kirjoitetaan halutun posteriorijakauman odotusarvo integraalina

$$\int g(x)p(x|y_{1:k})dx = \int g(x) \frac{p(x|y_{1:k})}{q(x|y_{1:k})} q(x|y_{1:k})dx, \quad (1.8)$$

jolle voidaan muodostaa Monte Carlo -approksimaatio tekemällä N otosta jakaumasta $x^i \sim q(x|y_{1:k})$.

Muodostetaan näin odotusarvo

$$\mathbb{E}[g(x)|y_{1:k}] \simeq \frac{1}{N} \sum_{i=1}^N \frac{p(x^i|y_{1:k})}{q(x^i|y_{1:k})} g(x^i) = \sum_{i=1}^N w^i g(x^i), \quad (1.9)$$

missä $g(x)$ on jokin estimoinnissa hyödyllinen, mielivaltainen funktio. Tutkielman käytetty notaatio x_k^i viittaa aika-askeleen k partikkeliin i , missä $i = \{1, \dots, N\}$. Tärkeytysotantaa kuvaaa nyt algoritmi 1. Kun posteriorijakauman estimaatti muodostetaan kyseisellä algoritmillä voidaan tulos kirjoittaa

$$\hat{p}(x|y_{1:k}) = \sum_{i=1}^N w^i \delta(x - x^i), \quad (1.10)$$

missä $\delta(x)$ on Diracin deltafunktio.

Algoritmi 1: Tärkeytysotanta

```

begin
  for  $i = 1, 2, \dots, N$  do
    begin
       $\lfloor$  Otetaan  $N$  otosta ehdotusjakaumasta  $x^i \sim q(x|y_{1:k})$ .
    begin
       $\lfloor$  Lasketaan normalisoimattomat painot  $w_*^i = p(y_{1:k}|x^i)p(x^i)/q(x^i|y_{1:k})$ .
       $\lfloor$  ja normalisoidut painot  $w^i = w_*^i / \sum_{j=1}^N w_*^j$ .
    begin
       $\lfloor$  Estimoidaan  $p$  laskemalla tiheydelle approksimaatio
       $\lfloor$   $\mathbb{E}[g(x)|y_{1:k}] \simeq \sum_{i=1}^N w^i g(x^i)$ .
     $\rfloor$ 
   $\rfloor$ 

```

1.5 Bayesilainen suodin

Suodinongelmassa ollaan kiinnostuttu tilavektorin posteriorijakauman $p(x_k|y_{1:k})$ estimoinnista. Tässä alaluvussa käydään läpi yleinen rekursiivinen, Bayesilainen posteriorijakauman laskenta. Tällaista suodinongelman ratkaisua kutsutaan myös Bayesilaiseksi suotimeksi. Koska epälineaarisessa, ei-normaalijakautuneessa tilanteessa rekursiota ei voida laskea analyyttisesti, pitää estimoinnissa käyttää numeerisia menetelmiä. SMC-menetelmissä tämä tarkoittaa jakauman sekventiaalista Monte Carlo -approksimointia, jonka toteutus esitetään alaluvun 4 algoritmissa. Molemmat esitykset noudattavat Gustafssonia (2010) [14].

Bayesilainen ratkaisu tilavektorin posteriorijakauman estimaatiille $\hat{p}(x_k|y_{1:k})$ saadaan seuraavalla rekursiolla (käydään läpi jokaisella aika-askeleella $k = \{1, \dots, T\}$). Lasketaan ensin

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}, \quad (1.11)$$

joka saadaan suoraan Bayesin kaavasta $\mathbb{P}(A|B) = \mathbb{P}(B|A)\mathbb{P}(A)/\mathbb{P}(B)$. Normalisointivakio lasketaan integraalina

$$p(y_k|y_{1:k-1}) = \int_{\mathbb{R}^{n_x}} p(y_k|x_k)p(x_k|y_{1:k-1}) dx_k, \quad (1.12)$$

joka saadaan kokonaistodennäköisyykskaavasta $\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A|X)] = \int_{-\infty}^{\infty} \mathbb{P}(A|X=x)f_X(x) dx$. Merkintä \mathbb{R}^{n_x} vastaa tässä piilossa olevan tilavektorin dimensiota n .

Lopuksi lasketaan päivitysaskel ajalle, joka saadaan edelleen kokonaistodennäköisyydellä

$$p(x_{k+1}|y_{1:k}) = \int_{\mathbb{R}^{n_x}} p(x_{k+1}|x_k)p(x_k|y_{1:k}) dx_k. \quad (1.13)$$

Rekursion avulla voimme laskea jakauman $p(x_k|y_{1:k})$ estimaatti käymällä rekursion läpi k kertaa.

1.6 Kalman-suotimen ja hiukkassuotimen yhteydestä ja eroista

Tässä alaluvussa käsitellään lyhyesti Kalman-suotimen yhteyttä hiukkassuotimeen edellä esitetyn teorian valossa. Esitys noudattaa Särkkää (2013). [29] Merkitään kuten edellä dynaamista mallia x_k ja havaintomallia y_k ja oletataan toisin kuin edellä, että nämä ovat lineaarisia ja noudattavat normaalijakaumaa. Koska mallit ovat lineaarisia, voidaan ne nyt kirjoittaa muotoon

$$x_k = \mathbf{A}_{k-1}x_{k-1} + q_{k-1}, \quad (1.14)$$

$$y_k = \mathbf{H}_k x_k + r_k, \quad (1.15)$$

missä \mathbf{A}_{k-1} on dynaamisen mallin tilasiirtymään kuvaava matriisi ja \mathbf{H}_k on havaintojen mallimatriisi. Normaalisuusoletuksesta seuraa, että sekä mallin että prosessin kohinavektorit noudattavat normaalijakaumia $q_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$ ja $r_k \sim \mathcal{N}(0, \mathbf{R}_k)$, missä \mathbf{Q}_{k-1} ja \mathbf{R}_k ovat kovarianssimatrissejä. Lisäksi oletetaan, että prosessin priorijakauma on normaali eli $x_0 \sim \mathcal{N}(m_0, \mathbf{P}_0)$. Mallit voidaan nyt kirjoittaa tiheysfunktioimuodossa

$$p(x_k|x_{k-1}) = \mathcal{N}(x_k|\mathbf{A}_{k-1}x_{k-1}, \mathbf{Q}_{k-1}), \quad (1.16)$$

$$p(y_k|x_k) = \mathcal{N}(y_k|\mathbf{H}_k x_k, \mathbf{R}_k), \quad (1.17)$$

joista voidaan edelleen johtaa suodinongelman mallit

$$p(x_k|y_{1:k-1}) = \mathcal{N}(x_k|m_k^*, \mathbf{P}_k^*), \quad (1.18)$$

$$p(x_k|y_{1:k}) = \mathcal{N}(x_k|m_k, \mathbf{P}_k), \quad (1.19)$$

$$p(y_k|y_{1:k-1}) = \mathcal{N}(y_k|\mathbf{H}_k m_k^*, \mathbf{S}_k) \quad (1.20)$$

ja ongelma ratkaista näin algoritmillä 2.

Algoritmi 2: Kalman-suodin

Result: Posteriorijakauman $p(x_{1:k}|y_{1:k})$ estimaatti.

Data: Havainnot y_k . Priorijakauman x_0 keskiarvovektori m_0 ja kovarianssimatriisi \mathbf{P}_0 .

```

begin
  for  $k = \{1, 2, \dots, T\}$  do
    begin
      Ennusteaskel.
       $m_k^* = \mathbf{A}_{k-1}m_{k-1}$ 
      if  $k < t$  then
        begin
          Päivitysaskel.
           $v_k = y_k - \mathbf{H}_k m_k^*$ 
           $\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_k^* \mathbf{H}_k^\top + \mathbf{R}_k$ 
           $\mathbf{K}_k = \mathbf{P}_k^* \mathbf{H}_k^\top \mathbf{S}_k^{-1}$ 
           $m_k = m_k^* \mathbf{K}_k v_k$ 
           $\mathbf{P}_k = \mathbf{P}_k^* - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top;$ 
    
```

Esitetty algoritmi on ns. Kalman-suodin, joka selkeästi toimii suodinongelman ratkaisuna, kun mallit ovat haluttua lineaarista normaalimuotoa. Jos tämä oletus ei täyty, on Kalman-suotimesta kehitetty useita versioita, joissa ei-lineaarinen malli voidaan linearisoida tiettyjen ehtojen vallitessa.

Tämän alaluvun tarkoituksena oli esittää, että Kalman-suotimessa ongelma on samaa muotoa kuin hiukkassuotimessa, joten linearisoituja Kalman-suotimia ei tässä käsitellä. Hiukkassuodin myös ratkaisee ongelman mille hyvänsä epälineariselle mallille.

Luku 2

Hiukkassuotimet

2.1 SIR-algoritmi

Tässä alaluvussa esitetään epälineaarisen suodinongelman ratkaisemiseksi SMC-menetelmiin kuuluva SIR-algoritmi. Algoritmi on numeerinen toteutus luvussa 1.5 kuvatusta Bayesilaisesta suotimesta. Esitetty algoritmi perustuu Gustafssonin (2010). Ilman uudelleenotantavaihetta kyseessä olisi SIS-algoritmi.

Algoritmi alustetaan jakaumasta $x_1^i \sim p_{x_0}$ generoiduilla N kappaleella partikkeleita. Jokaiselle partikkelille annetaan alustuksessa sama paino $w_{1|0}^i = 1/N$. Algoritmi suoritetaan jokaiselle partikkelille $i = \{1, 2, \dots, N\}$ jokaisella aika-askeleella $k = \{1, 2, \dots, T\}$.

Seuraava toistetaan jokaisella aika-askeleella $k = \{1, 2, \dots, T\}$. Algoritmin ensimmäisessä vaiheessa päivitetään painot yhtälön (2.1) mukaan.

$$w_{k|k}^i = \frac{1}{c_k} w_{k|k-1}^i p(y_k | x_k^i). \quad (2.1)$$

Tämä vastaa yllä esitetyn Bayes-suotimen päivitysvaihetta (1.11). Normalisointipaino c_k lasketaan puolestaan yhtälöstä (2.2), mikä vastaa Bayes-suotimen normalisointivaiion laskemista (1.12) ja asettaa painojen summaksi $\sum_{i=1}^N w_{k|k}^i = 1$.

$$c_k = \sum_{i=1}^N w_{k|k-1}^i p(y_k | x_k^i). \quad (2.2)$$

Seuraavassa vaiheessa estimoidaan p laskemalla tiheyden $p(x_{1:k} | y_{1:k})$ Monte Carlo -estimaatti yhtälön (2.3) perusteella

$$\hat{p}(x_{1:k} | y_{1:k}) = \sum_{i=1}^N w_{k|k}^i \delta(x_{1:k} - x_{1:k}^i). \quad (2.3)$$

Tämän jälkeen suoritetaan valinnainen uudelleenotanta. Uudelleenotanta voidaan tehdä jokaisella askeleella tai efektiivisen otoskoon perusteella alla kuvatun

kynnysarvoehdon $\hat{N}_{eff} < N_{th}$ täytessä, jolloin uudelleenotantaa kutsutaan adaptiiviseksi uudelleenotannaksi. Tällaista uudelleenotantaa hyödynnetään esitetyssä algoritmissa (3). Lopuksi päivitetään aika-askeleita (jos $k < T$), luodaan uudet ennusteet partikkeille ehdotusjakaumasta (2.4)

$$x_{k+1}^i \sim q(x_{k+1}|x_k^i, y_{k+1}) \quad (2.4)$$

ja päivitetään partikkelienvaihtoainekset tärkeytysotannalla (2.5), sen mukaan kuinka todennäköisiä partikkelienvaihtoainekset ovat

$$w_{k+1|k}^i = w_{k|k}^i \frac{p(x_{k+1}^i|x_k^i)}{q(x_{k+1}^i|x_k^i, y_{k+1})}. \quad (2.5)$$

Vaiheet 2.4 ja 2.5 vastaavat Bayes-suotimen aikapäivitystä (1.13).

Alla käsitellään algoritmiin liittyvän uudelleenotantamenetelmän, partikkelienvaihtoaineksen ja ehdotusjakauman valinta. Lopuksi esitetään algoritmin konvergenssia, marginaalijakaumaa sekä aikakompleksisuutta koskevia tuloksia ja käsitellään algoritmin tuottaman jakaumaestimaatin varianssin estimointia.

Algoritmi 3: SIR

Result: Posteriorijakauman $p(x_{1:k}|y_{1:k})$ estimaatti.

Data: Havainnot y_k . Generoitut $x_1^i \sim p_{x_0}$ missä $i = \{1, \dots, N\}$ ja jokainen partikkeli saa saman painon $w_{1|0}^i = 1/N$.

```
begin
    for  $k = \{1, 2, \dots, T\}$  do
        for  $i = \{1, 2, \dots, N\}$  do
            begin
                | Päivitetään painot  $w_{k|k}$ .
            begin
                | Estimoidaan  $p$  laskemalla tiheydelle approksimaatio
                |  $\hat{p}(x_{1:k}|y_{1:k}) = \sum_{i=1}^N w_{k|k}^i \delta(x_{1:k} - x_{1:k}^i)$ .
            begin
                | Lasketaan efektiivinen otoskoko  $\hat{N}_{eff}$ .
            if  $\hat{N}_{eff} < N_{th}$  then
                begin
                    | Otetaan uudet  $N$  otosta palauttaen joukosta  $\{x_{1:k}^i\}_{i=1}^N$ , missä
                    | otoksen  $i$  todennäköisyys on  $w_{k|k}^i$ .
                begin
                    | Asetetaan painot  $w_{k|k}^i = 1/N$ .
            if  $k < T$  then
                begin
                    | Aikapäivitys.
                    | Luodaan ennusteet partikkeillel ehdotusjakaumasta
                    |  $x_{k+1}^i \sim q(x_{k+1}|x_k^i, y_{k+1})$ ,
                    | päivitetään partikkeliin painot tärkeytysotannalla.
```

2.1.1 Parametrien valinta

Ennen algoritmin suorittamista valitaan ehdotusjakauma $q(x_{k+1}|x_{1:k}, y_{k+1})$, uudelleenotantamenetelmä sekä partikkeliin määärä N . Ehdotusjakauman ja uudelleenotantamenetelmän valinnassa tärkeimpänä päämääränä on välttää otosten ehtymistä, kun taas partikkeliin määärä säätelee kompromissia algoritmin suorituskyvyn ja tarkkuuden sekä varianssin välillä.

2.1.1.1 Otoskoon N valinta

Yleispätevästi sääntöä otoskoon/partikkeliin lukumäärän N valinnalle on vaikeaa antaa, sillä vaadittava estimointitarkkuus riippuu usein kässillä olevasta ongelmasta. Gordon &al. (1993) [11] esittävät kuitenkin kolme tekijää, jotka vaikuttavat partikkeliin lukumäärän valintaan

- a. tila-avaruuden ulottuvuuksien lukumäärä n_x ,

- b. tyyppillinen päälekäisyys priorin ja uskottavuuden välillä
- c. sekä tarvittava aika-askelten lukumäärä.

Ensimmäisen tekijän vaikutus on selvä. Mitä useammassa ulottuvuudessa otantaa tarvitsee tehdä, sen korkeammaksi on N asetettava, jotta jokainen ulottuvuus pystytään kattamaan. Tekijät (b) ja (c) puolestaan seuraavat uudelleenotannasta. Jos se osa tila-avaruutta, jossa uskottavuus $p(y_k|x_k)$ saa merkittäviä arvoja on pieni verrattuna siihen osaan, jossa priorijakauma $p(x_k|y_{1:k-1})$ saa merkittäviä arvoja, suuri osa partikkeleista saa pieniä painoja eikä näin valikoidu uudelleenotantaan.

Yleisesti ottaen N kannattaa asettaa sellaiseksi, että se paitsi tuottaa riittävän tarkan estimaatin, on se myös käytettävissä olevan laskentatehon sekä vaadittavan laskentanopeuden kannalta järkevä. Tähän palataan tutkielman empiirisessä paikannusesimerkissä.

2.1.1.2 Uudelleenotantamenetelmän valinta

Ilman uudelleenotantaa on mahdollista, että algoritmi alkaa kärsiä SIS-algoritmille ominaisesta otosten ehtymisestä. Toisin sanoen kaikki painot alkavat keskittyä vain muutamalle partikkeliille eikä algoritmi enää approksimoi tehokkaasti haluttua jakaumaa. Uudelleenotanta tarjoaa osittaisen ratkaisun tähän ongelmaan, mutta hävittää samalla informaatiota ja lisää siten satunnaisotantaan liittyvää epävarmuutta. Yleisesti ottaen uudelleenotanta kannattaa aloittaa vasta siinä vaiheessa algoritmin suorittamista, kun siitä on otosten ehtymisen kannalta hyötyä, esimerkiksi efektiivisen otoskoon pudottua jonkin kynnysarvon alapuolelle (adaptiivinen uudelleenotanta). Efektiivinen otoskoko saadaan laskettua variaatiokertoimesta c_ν kaavalla

$$N_{eff} = \frac{N}{1 + c_\nu^2(w_{k|k}^i)} = \frac{N}{1 + \frac{\text{Var}(w_{k|k}^i)}{(\mathbb{E}[w_{k|k}^i])^2}} = \frac{N}{1 + N^2 \text{Var}(w_{k|k}^i)}. \quad (2.6)$$

Näin laskettu efektiivinen otoskoko maksimoituu ($N_{eff} = N$), kun kaikille painoille pätee $w_{k|k}^i = 1/N$ ja minimoituu ($N_{eff} = 1$), kun $w_{k|k}^i = 1$ todennäköisyydellä $1/N$ ja $w_{k|k}^i = 0$ todennäköisyydellä $(N - 1)/N$. Normalisoitujen painojen avulla saadaan effektiiviselle otoskoolle aika-askelleella k laskennallinen approksimaatio

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^N (w_{k|k}^i)^2}. \quad (2.7)$$

Määritelmissä 2.6 ja 2.7 pätee $1 \leq \hat{N}_{eff} \leq N$. Yläraja saavutetaan, kun jokaisen partikkelin paino on sama. Alarajalle päädytään, kun kaikki paino keskittyy yksittäiselle partikkeliille. Tästä saadaan määritettyä algoritmille SIR-uudelleenotantaehto $\hat{N}_{eff} < N_{th}$. Gustafsson (2010) [14] esittää uudelleenotannan kynnysarvoksi esimerkiksi $\hat{N}_{th} = 2N/3$.

Uudelleenotanta ei muuta approksimoitavan jakauma p odotusarvoa, mutta se lisää jakauman Monte Carlo -varianssia. On kuitenkin olemassa esimerkiksi osittamiseen perustuvia uudelleenotantamenetelmiä, jotka pyrkivät minimoimaan varianssin lisäyksen. Varianssin pienennysmenetelmät jätetään tämän tutkielman ulkopuolelle.

2.1.1.3 Ehdotusjakauman valinta

Yksinkertaisin muoto ehdotusjakaumalle on $q(x_{1:k}|y_{1:k})$ eli jokaisella algoritmin suorituskerralla käydään läpi koko aikapolku $1:k$. Tämä ei kuitenkaan ole tarkoituksesta, erityisesti jos kyseessä on reaalialainen sovellutus. Kirjoitetaan sen sijaan ehdotusjakauma muodossa

$$q(x_{1:k}|y_{1:k}) = q(x_k|x_{1:k-1}, y_{1:k})q(x_{1:k-1}|y_{1:k}). \quad (2.8)$$

Jos yhtälöstä (2.8) poimitaan ehdotusjakaumaksi ainoastaan termi $q(x_k|x_{1:k-1}, y_{1:k})$ voidaan tämä kirjoittaa edelleen Markov-ominaisuuden nojalla muotoon $q(x_k|x_{k-1}, y_k)$. Tämä on suodinongelman kannalta riittävää, koska olemme kiinnostuneita posteriorijakaumasta ja arvosta x ainoastaan aika-askeleella k (siloitinongelmassa tarvitsisimme koko polun $x_{1:k}$). Alla tarkastellaan edelleen Gustafssonia (2010) [14] seuraten kahta ehdotusjakauman valintatapaa, prioriotantaa (*prior sampling*) sekä uskottavuusotantaa (*likelihood sampling*).

Ennen ehdotusjakauman tarkastelua määritellään mallille signaali-kohinasuhde uskottavuuden maksimin ja priorin maksimin välisenä suhteena

$$\text{SNR} \propto \frac{\max_{x_k} p(y_k|x_k)}{\max_{x_k} p(x_k|x_{k-1})}. \quad (2.9)$$

Yhdistetään lisäksi ehdotusjakaumia varten yhtälöt (2.1) ja (2.2), jolloin saadaan painojen päivitys muotoon

$$w_{k|k}^i \propto w_{k-1|k-1}^i \frac{p(y_k|x_k^i)p(x_k|x^{k-1})}{q(x_k|x_{k-1}^i, y_k)}. \quad (2.10)$$

Kun suhde (2.9) on matala, on prioriotanta luonnollinen valinta. Tässä käytetään ehdotusjakaumana tilavektorin ehdollista prioria eli

$$q(x_k|x_{1:k-1}, y_k) = p(x_k|x_{k-1}^i). \quad (2.11)$$

Yhtälön (2.11) perusteella saadaan edelleen prioriotannan painoiksi

$$w_{k|k}^i = w_{k-1|k-1}^i p(y_k|x_k^i) = w_{k-1|k-1}^i p(y_k|x_k^i). \quad (2.12)$$

Kun signaali-kohinasuhde on kohtalainen tai korkea, on parempi käyttää ehdotusjakaumana skaalattua uskottavuusfunktiota (2.14). Tarkastellaan ensin tekijöihin jakoa

$$p(x_k|x_{k-1}^i, y_k) = p(y_k|x_k) \frac{p(x_k|x_{k-1}^i)}{p(y_k|x_{k-1}^i)}. \quad (2.13)$$

Kun SNR on korkea ja uskottavuusfunktio on integroituva pätee $p(x_k|x_{k-1}^i, y_k) \propto p(y_k|x_k)$, jolloin voidaan asettaa (2.14)

$$q(x_k|x_{k-1}^i, y_k) \propto p(y_k|x_k). \quad (2.14)$$

Yhtälön (2.14) perusteella saadaan edelleen uskottavuusotannan painoiksi (2.15).

$$w_{k|k}^i = w_{k-1|k-1}^i p(x_k^i|x_{k-1}^i). \quad (2.15)$$

2.1.2 Konvergenssituloksia

Alla esitetään kaksi SIR-algoritmiin liittyvää konvergenssitulosta. Se, kuinka hyvin esitettyllä algoritmillä arvioituna posterioriheys $\hat{p}(x_{1:k}|y_{1:k})$ approksimoi todellista tiheysfunktiota $p(x_{1:k}|y_{1:k})$ sekä mikä on approksimaation keskineliövirhe. Tulokset 1–2 noudattavat Crisanin ja Doucet’n artikkeleita “Convergence of Sequential Monte Carlo Methods” (2000) [7] ja “A Survey of Convergence Results on Particle Filtering Methods for Practitioners” (2002) [8], tulos 3 Chopinin artikkelia “Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference” (2004) [5].

Konvergenssitulos 1: Kun $N \rightarrow \infty$ algoritmilelle pätee $\forall k$ tulos (2.16).

$$\hat{p}(x_{1:k}|y_{1:k}) \xrightarrow{a.s.} p(x_{1:k}|y_{1:k}). \quad (2.16)$$

Konvergenssitulos 2: Keskineliövirheelle pätee asymptoottinen konvergenssi (2.17).

$$\mathbb{E}(\hat{g}(x_k) - \mathbb{E}(g(x_k)))^2 \leq \frac{p_k \|g(x_k)\|}{N}, \quad (2.17)$$

missä g on mikä hyvänsä piilossa olevan tila-avaruuden rajoitettu Borel-mitallinen funktio ($g \in \mathcal{B}(\mathbb{R}^{n_x})$), $\|g(\cdot)\|$ kyseisen funktion supremum-normi ja p_k jokin äärellinen vakio, jolle pätee ajanhetkestä k riippumatta $p_k = p < \infty$.

Konvergenssitulos 3: Keskeinen raja-arvolause (2.18).

$$\text{Kun } N \rightarrow \infty : \sqrt{N} \left\{ \frac{1}{N} \sum_{i=1}^N \hat{g}(x_k^i) - \mathbb{E}(g(x_k^i)) \right\} \xrightarrow{D} \mathcal{N}(0, \sigma^2 < \infty), \quad (2.18)$$

missä g on jälleen mikä hyvänsä piilossa olevan tila-avaruuden rajoitettu Borel-mitallinen funktio ($g \in \mathcal{B}(\mathbb{R}^{n_x})$). Konvergenssituloksia ei tämän tutkielman puitteissa todisteta.

2.1.3 Marginaalijakauma

Edellä kuvattu algoritmi 3 tuottaa approksimaation koko prosessin posteriorijakaumalle $p(x_{1:k}|y_{1:k})$. Jos halutaan tietää ainoastaan posteriorijakauman $p(x_k|y_{1:k})$ estimaatti, voidaan käyttää yksinkertaisesti viimeisestä tilasta x_k laskettua estimaatia

$$\hat{p}(x_k|y_{1:k}) = \sum_{i=1}^N w_{k|k}^i \delta(x_k - x_k^i). \quad (2.19)$$

Toinen, tarkempi vaihtoehto on käyttää laskennassa tärkeytyspainoa

$$w_{k+1|k}^i = \frac{\sum_{j=1}^N w_{k|k}^j p(x_{k+1}^i | x_k^j)}{q(x_{k+1}^i | x_k^i, y_{k+1})} \quad (2.20)$$

painon (2.5) sijaan. Tällöin jokaisella aikapäivitysaskeleella lasketaan painot kaikkien mahdollisten tila-aika-avaruuspolkujen yli. Samoin kuin uudelleenotanta tämä pienentää painojen varianssia.

2.1.4 Aikakompleksisuus

Algoritmin perusmuodon aikakompleksisuus on $\mathcal{O}(N)$. Uudelleenotantamenetelmän tai ehdotusjakauman valinta ei suoraan vaikuta aikakompleksisuuteen. Sen sijaan marginalisointi tärkeytyspainolla (2.20) lisää algoritmin aikakompleksisuutta $\mathcal{O}(N) \rightarrow \mathcal{O}(N^2)$, koska jokaisen partikkelin kohdalla painot lasketaan jokaisen tila-aika-avaruuspolun yli. On selvää, että erityisesti isoilla otoskoon N arvoilla ei yllä esitetty marginalisointi enää ole mielekästä.

Tällaisia tilanteita varten algoritmista on olemassa $\mathcal{O}(N \log(N))$ -versioita, jotka perustuvat esimerkiksi N:n kappaleen oppimiseen (*N-body learning*). Näiden algoritmien käyttely jäätä tämän tutkielman ulkopuolelle, mutta katsauksen algoritmeista ovat esittäneet esimerkiksi Klaas & al. artikkelissa “Toward Practical N^2 Monte Carlo: the Marginal Particle Filter” (2005) [18].

2.2 Saapasremmisuodin

Saapasremmisuodin 4 eli *bootstrap filter* on SIR-algoritmin muunnelma, jossa tärkeytysotannan (kts. 1) käytetään dynaamista mallia $p(x_k|x_{k-1})$.

Algoritmi 4: Saapasremmisuodin

Result: Posteriorijakauman $p(x_{1:k}|y_{1:k})$ estimaatti.

Data: Havainnot y_k . Generoitu $x_1^i \sim p_{x_0}$ missä $i = \{1, \dots, N\}$ ja jokainen partikkeli saa saman painon $w_{1|0}^i = 1/N$.

begin

```
for k = {1, 2, ..., T} do
    for i = {1, 2, ..., N} do
        begin
            Luodaan uudet estimaatit dynaamisesta mallista  $x_k^i \sim p(x_k|x_{k-1}^i)$ ;
        begin
            Päivitetään hiukkasten painot  $w_k^i$  uskottavuusfunktion  $p(y_k|x_k^i)$ 
            mukaan.
        begin
            Estimoidaan  $p$  laskemalla tiheydelle approksimaatio
             $\hat{p}(x_{1:k}|y_{1:k}) = \sum_{i=1}^N w_{k|k}^i \delta(x_{1:k} - x_{1:k}^i)$ .
        if k < t then
            begin
                Aikapäivitys. Suoritetaan uudelleenotanta kuten SIR-algoritmissa
                3.
```

Saapasremmisuodin on edellä esitettyä SIR-algoritmia yksinkertaisempi toteuttaa, mutta epäinformatiivisen tärkeytysjakauman vuoksi algoritmi saattaa vaatia SIR-algoritmia suuremman määrään hiukkasia. Saapasremmisuodin esitetään tässä sen historiallisen tärkeyden vuoksi, sillä kyseessä oli ensimmäinen uudelleenotataan- taa hyödyntävä hiukkassuodinalgoritmi. Suotimien käytännön toteutukseen palataan luvussa 4.

2.3 Varianssin estimoinnista

Hiukkassuotimen varianssin estimoinnissa ollaan kiinnostuneita jakaumaestimaatin $\hat{p}(x_{1:k}|y_{1:k})$ varianssin estimoinnista. Yksinkertaisin tapa estimoida hiukkassuodinalgoritmin varianssia on ajaa algoritmi $M > 1$ kertaa. Koska ajot ovat toisistaan riippumattomia, voidaan estimaatin varianssi laskea kullekin aika-askeleelle k näiden ajojen k -hetken estimaattien otosvarianssina:

$$\hat{\sigma}_{MC}^2 = \text{Var}_{\text{MC}}(\hat{p}(x_{1:k}|y_{1:k})) = \frac{1}{M-1} \sum_{i=1}^M (x_k^i - \bar{x}_k)^2., \quad (2.21)$$

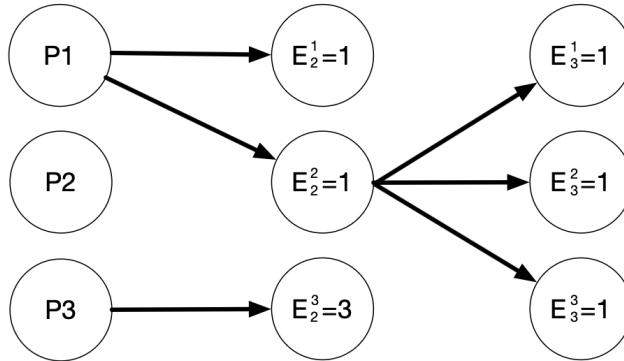
missä x_k^i on k :nen aika-askeleen piste-estimaatti ajolle $i = 1, \dots, M$ ja \bar{x}_k pistee-estimaattien aritmeettinen keskiarvo laskettuna kaikkien M ajojen yli. Tällaisen Monte Carlo -varianssin estimoiminen on kuitenkin laskennallisesti tehotonta. Monissa käytännön sovelluksissa jo yhden hiukkassuodinalgoritmin ajaminen vaatii runsaasti laskentatehoa, jolloin Monte Carlo -varianssin laskeminen ei ole mah-

dollista. Varianssia ei voi myöskään laskea analyttisesti, mutta koska keskeisen raja-arvolauseen 2.18 nojalla tiedetään, että asymptoottinen varianssi

$$\lim_{N \rightarrow \infty} N \text{Var}(\hat{g}(x_k^i)), \quad (2.22)$$

on olemassa, on sen estimointiin kehitetty lähi vuosina joitakin approksimatiivisia menetelmiä. Alla käsitellään Leen ja Whitleyn (2018) ehdottamaa varianssin estimointitapaa [19].

Ajetaan SIR-algoritmi kuten esitetty algoritmissa 3, mutta merkitään kyllakin algoritmin suorituskerralla $k = 1, \dots, T$ ja jokaisen partikkelin $n = 1, \dots, N$ kohdalla indeksillä E_k^n kunkin partikkelin kantaisää eli sitä partikkeilia, josta kyseinen partikkeli on uudelleenotantojen kautta polveutunut aika-askeleesta $k = 1$ lähtien. Kaavio 2.1 havainnollistaa partikkelienvaihdostusta.



Kuva 2.1: Esimerkki Eeva-indekseistä, kun $T = 3$ ja $N = 3$.

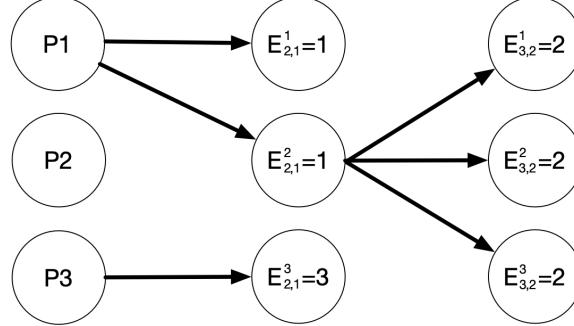
Varianssiestimaatti voidaan laskea näiden, Leen ja Whitleyn Eeva-indekseiksi nimeämien indeksien perusteella seuraavasti:

$$\hat{\sigma}_{CLE}^2 = \frac{1}{N^2} \left[\left(\sum_{i=1}^N \gamma(x_k^i) \right)^2 - \left(\frac{N}{N-1} \right)^{n+1} \left(\sum_{i=1}^N \sum_{j:E_n^j=i} \gamma(x_k^i) \gamma(x_k^j) \right) \right], \quad (2.23)$$

missä $\gamma : \mathbf{X}_n \rightarrow \mathbb{R}$ on rajoitettu, \mathcal{X}_n -mitallinen funktio, SIR-algoritmin kohdalla jakaumaestimaatti $\hat{p}(x_{1:k}|y_{1:k})$. Kyseessä on harhaton ja konsistentti asymptoottisen varianssin estimaatti. Tarkemmin $N\hat{\sigma}_{CLE}^2$ konvergoi asymptoottiseen varianssiin 2.22. Tätä ei tutkielman puitteissa todisteta.

Yllä esitetty varianssiestimaatti kärsii kuitenkin epätarkkuudesta, sillä kun $k \rightarrow \infty$, polveutuvat kaikki hiukkaset lopulta samasta kantaisästä eli indeksit E_k^1, \dots, E_k^N ovat kaikki yhtäsuuria. Tämän vuoksi on mielekästä johtaa indeksit ainoastaan tietystä aiemmasta aika-askeleesta alkaen. Olsson ja Douc (2019) [26] ehdottavat tähän tarkoitukseen Henok-indeksiä $E_{k,m}^n$, jossa m merkitsee aika-askeleen

$m < k$ partikkelia, josta kyseinen partikkeli polveutuu. Partikkelienv kantaisien sukupolvi määritetään viipeellä λ niin, että $m = k - \lambda$. Kaavio 2.2 havainnollistaa partikkelienv polveutumista, kun $\lambda = 1$.



Kuva 2.2: Esimerkki Henok-indeksistä, kun $T = 3$, $N = 3$ ja $\lambda = 1$.

Nyt varianssi saadaan muotoon

$$\hat{\sigma}_{OD}^2 = \frac{1}{N^2} \left[\left(\sum_{i=1}^N \gamma(x_k^i) \right)^2 - \left(\frac{N}{N-1} \right)^{n+1} \left(\sum_{i=1}^N \sum_{j:E_{n,k(\lambda)}^j=i} \gamma(x_k^i) \gamma(x_k^j) \right) \right], \quad (2.24)$$

missä $k(\lambda) := k - \lambda$. Viive λ on varianssiestimaatin suunnitteluparametri. Pienillä viipeillä estimaatti on harhainen, mutta harha laskee viipeen kasvaessa. Olsson ja Douc suosittavat viipeen ylärajaksi arvoa $\lambda = 20$, jolloin estimaattorin harha on käytännössä kokonaan hävitetty. Tämän jälkeen estimaatti voi myös alkaa kärsiä samasta epätarkkuudesta kuin Eeva-indeksihin perustuva CLE-estimaatti.

Mastrototaro ja Olsson (2023) [21] laajentavat tästä estimaattia edelleen niin, että viive λ valitaan mukautuvasti. Mastrototaron ja Olssonin ALvar-estimaatti (*Adaptive-Lag variance*) lasketaan kuten OD-varianssi edellä 2.24, mutta kunkin algoritmin ajokerran jälkeen asetetaan seuraavan ajanhetken λ seuraavasti:

$$\lambda_{k+1} \leftarrow \arg \max_{\lambda \in [0, \lambda_k + 1]} \hat{\sigma}_{k+1, \lambda}^2(\gamma_{k+1}), \quad (2.25)$$

Tämä perustuu havaintoon, jonka mukaan ehtyneille Henok-indeksille on olemassa $\lambda' \in [0, \lambda - 1]$, joka täyttää ehdon $\hat{\sigma}_{k+1, \lambda'}^2(\gamma_{k+1}) < \hat{\sigma}_{k+1, \lambda}^2(\gamma_{k+1})$, koska myös ehtyneiden indeksien kantaisät ovat ehtyneitä. Nyt indeksi λ_{n+1} voidaan valita rekursiivisesti niin, että se tuottaa suurimman varianssiestimaatin, joka on kuitenkin rajoitettu ylhäältä arvoon $\lambda_n + 1$. Tämä indeksi ei ole koskaan ehtynyt, joten se on myös parhaan varianssiestimaatin tuottava valinta.

Esitettyjä varianssiestimaatteja voidaan hyödyntää paitsi algoritmien ja parametrivalintojen vertailussa myös mukautuvan viipeen hiukkassiloittimen viipeen valinnassa (kts. luku 3). Varianssiestimaatti lasketaan myös luvun 4 empiirisessä esimerkissä.

Luku 3

Hiukkassilottimet

Tässä luvussa käsitellään suodinongelman läheisesti liittyvän siloitteluongelman ratkaisemista hiukkassiloitinalgoritmien avulla. Kuten hiukkassuotimien kohdalla, myös tässä luvussa esitetään ongelma ensin yleisessä Bayesilaisessa muodossa, jonka jälkeen siirrytään käsittelemään hiukkasmenetelmiin pohjautuvia siloitinalgoritmeja. Luvussa käsiteltävät algoritmit jaetaan kahteen pääkategoriaan, offline-algoritmeihin, joita sovelletaan hiukkassuodinalgoritmin ajon jälkeen sekä online-algoritmeihin, jotka suoritetaan yhdessä hiukkassuodinalgoritmin rinnalla.

Siloitinongelman esittely seuraa Särkkää (2013) [29]. Algoritmien käsitteily pohjautuu SIR- ja BS-PS-siloittimien osalta Särkkään (2013) [29] sekä SIR-siloittimen osalta Genshiro Kitagawan artikkeliin “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models” (1996) [17]. Kiinteän viipeen silotin seuraa niin ikään Kitagawaa (1996). Uudelleenpainottava silotin perustuu Doucetin & al. artikkeliin “On sequential Monte Carlo sampling methods for Bayesian filtering” (2000) [10]. Mukautuvan viipeen silotin seuraa puolestaan Alenlövin ja Olssonin artikkelia “Particle-Based Adaptive-Lag Online Marginal Smoothing in General State-Space Models” (2019) [1].

3.1 Bayesilainen silotin

Bayesilaisen siloittimen tarkoitus on laskea tilan x_k marginaaliposteriorijakauma $p(x_k|y_{1:T})$ aika-askeleella k , kun käytössä on havaintoja aika-askeleeseen T asti, missä $T > k$. Ero Bayesilaiseen suotimeen (kts. alaluku 1.5) on siinä, että suodinongelmassa havaintoja on saatavilla ainoastaan aika-askeleeseen k asti, kun taas siloitinongelmassa myös tulevat havainnot ovat saatavilla. Ajassa taaksepäin etenevät rekursiiviset yhtälöt ongelman ratkaisemiseksi voidaan esittää muodossa

$$p(x_{k+1}|y_{1:k}) = \int_{\mathbb{R}^{n_x}} p(x_{k+1}|x_k)p(x_k|y_{1:k}) dx_k \quad (3.1)$$

ja

$$p(x_k|y_{1:T}) = p(x_k|y_{1:k}) \int \frac{p(x_{k+1}|x_k)p(x_{k+1}|y_{1:T})}{p(x_{k+1}|y_{1:k})} dx_{k+1}, \quad (3.2)$$

missä $p(x_k|y_{1:k})$ on suodintiheys aika-askeleella k ja $p(x_{k+1}|y_{1:k})$ prediktiivinen jakauma ajanhetkelle $k+1$. Kuten suodinongelman kohdalla, voidaan ongelma ratkaista suljetussa muodossa, kun mallit ovat lineaarisia. Tällöin kyseessä on Rauch-Turn-Striebel-siloitin (RTSS), josta käytetään myös nimitystä Kalman-siloitin. Samoin, kuten Kalman-suotimen kohdalla, ongelma voidaan tiettyjen ehtojen vallitessa linearisoida. Näitä linearisoituja siloittimia ei käsitellä tässä tutkielmassa. hiukkassuotimen tavoin hiukkassiloitin ratkaisee ongelman mille hyvänsä epälineaariselle mallille.

3.2 Offline-algoritmit

Offline-siloittimet estimoivat siloitintiheyttä aika-askeleella $k < T$, kun havaintodata on käytössä koko ajanjaksoalta $1 \dots T$. Alla esitettyt algoritmit siis olettavat, että kaikki mahdollinen tuleva data on jo niiden käytössä. Ohessa käsitellään lyhyesti muutamaa ehdotettua offline-hiukkassiloitinalgoritmia.

3.2.1 SIR-siloitin

Kuten aiemmin mainittua, näyttäävät hiukkassuodinalgoritmit, erityisesti SIR-algoritmi 3, ratkaisevan siloitteluongelman ilmaiseksi, kunhan tallennamme aika-askeleella k koko otoshistorian $x_{0:k}^i$. Tällöin voimme estimoida täytä siloitteluposteriorijakaumaa seuraavasti:

$$p(x_{0:T}|y_{1:T}) \approx \sum_{i=1}^N w_T^i \delta(x_{0:T} - x_{0:T}^i). \quad (3.3)$$

Nyt aika-askeleen k siloitinjakauma saadaan laskettua

$$p(x_k|y_{1:T}) \approx \sum_{i=1}^N w_T^i \delta(x_k - x_k^i), \quad (3.4)$$

missä x_k^i on $x_{0:T}^i$:n k :s elementti. Koska uudelleenotanta hävittää otoshistorian, pitää uudelleenotanta suorittaa koko otoshistoriasta $x_{0:k}^i = (x_{0:k-1}^i, x_k^i)$ pelkän aika-askeleen k otoksen x_k^i sijaan. Koska nyt koko otoshistoria pitää tallentaa, vaatii SIR-siloitin NkT muistia pelkän N sijaan. Vastaavasti myös uudelleenotannan aikakompleksisuus kasvaa [17].

SIR-siloittimen suurin ongelma on kuitenkin sen tuottamien estimaattien epätarkkuus. Kun ajanhetkien määrä kasvaa, johtaa koko otoshistorian uudelleenotanta kaiken painon kasautumiseen historian tietyille otoksiille, jolloin SIR-siloittimen tuottamat estimaatit eivät enää estimoi haluttua (siloitteluposteriorijakaumaa [17].

3.2.2 BS-PS-siloitin

Backward-simulation particle smoother (BS-PS) eli taaksepäin simuloiva hiukkasiloitin estimoii hiukkassuotimen tulosten perusteella paremmin siloitinjakaumaa. Tässä algoritmissa hiukkasten historia simuloidaan aika-askeleesta T taaksepäin ensimmäiseen aika-askeleeseen:

Algoritmi 5: Taaksepäin simuloiva hiukkassiloitin

Result: Posteriorisiloitinjakauman $p(x_k|y_{1:T})$ estimaatti.

Data: Suodinkakaumia edustavat hiukkaset ja näihin liittyvät painot w_k^i, x_k^i , missä $i = 1, \dots, N$ ja $k = 1, \dots, T$

```

begin
  begin
    Valitaan  $\tilde{x}_T = x_T^i$ 
    for  $k = \{T - 1, \dots, 0\}$  do
      begin
        Lasketaan uudet painot
         $w_{k|k+1}^i \propto w_k^i p(\tilde{x}_{k+1}|x_k^i)$ 
      begin
        Valitaan  $\tilde{x}_k = x_k^i$  todennäköisyydellä  $w_{k|k+1}^i$ .
    
```

Nyt siloittelujakaumaa voidaan estimoida seuraavasti:

$$p(x_{0:T}|y_{1:T}) \approx \frac{1}{S} \sum_{i=1}^N \delta(x_{0:T} - \tilde{x}_{0:T}^i), \quad (3.5)$$

missä $S, j = 1, \dots, S$ on algoritmin 5 toistokertojen määrä. Koska $\tilde{x}_{0:T}^j$ pitää sisällään kaikki otospolut, saadaan marginaalijakauma aika-askeleella k yhtälöstä 3.5 yksinkertaisesti valitsemalla sen k :net elementit. Sekä algoritmin aikakompleksisuus että muistivaade on $\mathcal{O}(STN)$.

3.2.3 Uudelleenpainottava hiukkassiloitin

Uudelleenpainottavassa hiukkassiloittimessa (tunnetaan myös nimellä marginaalihiukkassiloitin, kts. mm. Doucet, Godsill & ja Andrieu [10]) siloitinjakaumaa estimoidaan käyttämällä SIR-hiukkassuodinalgoritmista 3 saatuja hiukkasia, mutta ne painotetaan uudelleen käyttäen dataa aika-askeleesta T alkaen, edeten ajassa taaksepäin.

Algoritmi 6: Uudelleenpainottava hiukkassiloitin

Result: Posteriorisiloitinjakauman $p(x_k|y_{1:T})$ estimaatti.

Data: Suodinjakaumia edustavat hiukkaset ja näihin liittyvät painot w_k^i, x_k^i , missä $i = 1, \dots, N$ ja $k = 1, \dots, T$

begin

begin

 Asetetaan $w_{T|T}^i = w_T^i$, jokaiselle $i = 1, \dots, N$;

for $k = \{T - 1, \dots, 0\}$ **do**

begin

 Lasketaan uudet painot

$$w_{k|T}^i = \sum_j w_{k+1|T}^j \frac{w_k^i p(x_{k+1}^j | x_k^i)}{\sum_l w_k^l p(x_{k+1}^l | x_k^i)}$$

Halutun siloitinjakauma estimaatti aika-askeleella k saadaan painotettuna keskiarvona $p(x_k|y_{1:T}) \approx \sum_i w_{k|t}^i \delta(x_k - x_k^i)$. Algoritmin aikakompleksisuus on $\mathcal{O}(N^2)$.

3.3 Online-algoritmit

Yllä esitetyt offline-algoritmit ratkaisevat siloitinongelman niin, että kaikki data ajanhetkeen asti T on saatavilla. Käytännössä siloitin siis ajetaan suodinalgoritmin jälkeen. Käytännön sovellutuksissa tämä ei ole aina mahdollista, jos siloittelujakauman pitää olla saatavilla reaalialkaisesti. Online-siloittimet ratkaisevat siloitinongelman niin, että saatavilla on dataa aika-askeleeseen $k + L \leq T$ asti, missä L on dataan lisätty L :n aika-askeleen viive. Online-algoritmit voidaan edelleen jakaa kiinteän viipeen siloittimiin (*fixed-lag smoother*) ja mukautuvan viipeen siloittimiin (*adaptive-lag smoother*). Nimensä mukaisesta kiinteään viipeen siloitinalgoritmessa viive L valitaan suunnitteluparametrina, kun taas mukautuvan viipeen siloittimet pyrkivät valitsemaan parhaan tai optimaalisen viipeen johonkin laskennalliseen kriteeriin perustuen.

3.3.1 Kiinteän viipeen siloitin

Yksinkertaisin tapa toteuttaa kiinteän viipeen siloitin on käyttää SIR-siloitinta niin, että maksimiaika-askel T korvataan valitulla viipeellä $k + L \leq T$. [17]. Nyt yhtälön 3.3 jakauma saadaan muotoon

$$p(x_{0:(k+L)}|y_{1:(k+L)}) \approx \sum_{i=1}^N w_{k+L}^i \delta(x_{0:(k+L)} - x_{0:(k+L)}^i), \quad (3.6)$$

ja nykyisen aika-askeleen k siloitinjakauma lasketaan tästä jakaumasta kuten SIR-siloittimessa (kts. yhtälö 3.4). Kiinteän viipeen siloitin välttää SIR-siloittimen approksimaatio-ongelmat. Kun viipeelle L pätee $k + L \ll T$ parantaa viipeen pidentäminen tiettyyn pisteesseen asti jakauman approksimaatiota. Kitagawa (1996) suosittelee 10–20 aika-askeleen viivettä ja esittää 50 aika-askelta viipeen ylärajaksi

[17]. Paremman estimaatin vastapainona pidemmän viipeen valinta lisää myös viivettä, joka dataa tuottavaan järjestelmään pitää lisätä. Siloittimien tulokset ovat saatavilla vasta L :n aika-askeleen jälkeen, mikä ei aina ole käytännössä mahdollista tai haluttua. Pidempi viive myös lisää algoritmin muistivaatimuksia, joskin muistivaatimukset pysyvät aina pienempinä kuin SIR-siloittimessa.

Kiinteän viipeen siloitinta (viipeellä $L = 1$) voidaan hyödyntää myös prediktivisenä siloittimenä, jossa siloittelujakaumaa $p(x_{0:(k+1)}|y_{1:(k+1)})$ käytetään suodinjakauman $p(x_{1:(k)}|y_{1:k})$ laskennassa [24]. Ydinajatuksena on muokata SIR-algoritmia 3 niin, että aika-askeleen k painoja w_k^i painotetaan edelleen seuraavasta aika-askeleesta $k + 1$ lasketuilla painoilla. Näin algoritmi painottaa jo nykyhetkessä niitä hiukkasia, joiden uskottavuus on seuraavalla aika-askeleella suurempi. Tämä prediktivinen siloitin voidaan toteuttaa lisäämällä SIR-algoritmiin painotusvaiheen jälkeen seuraava ala-algoritmi:

Algoritmi 7: Prediktivinen siloitin (viive=1)

Result: Prediktivisellä siloittimella lasketut painot painotettu \tilde{w}_k^i .
Data: Viipeen $L = 1$ avulla saadut havainnot y_{k+1} . Partikkelit x_k^i ja niitä vastaavat painot w_k^i

```

begin
  for  $i = \{1, 2, \dots, N\}$  do
    begin
      | Luodaan simuloidut hiukkaset  $\tilde{x}_{k+1}^i$  ehdotusjakaumsta  $q(\tilde{x}_{k+1}|x_k^i, y_{k+1})$ 
      |
      begin
        | Lasketaan simuloiduille hiukkaille painot  $\tilde{w}_{k+1}^i$ 
      begin
        | Päivitetään nykyiset painot  $\tilde{w}_k^i = w_k^i \tilde{w}_{k+1}^i$ 
      |
      begin
        | Korvataan nykyiset painot  $w_k$  siloitetuilla painoilla  $\tilde{w}_k^i$ ;
    
```

Kun hiukkasten määrä N pysyy samana, lisää prediktivinen siloitin suodinjakauman laskemisen tarkkuutta. Vastaavasti prediktivinen siloitin mahdollistaa saman suodinjakauman estimaatin tarkkuuden kuin SIR-algoritmi pienemmällä määrellä hiukkasia, kuitenkin vain tuplalten uskottavuusfunktiota laskettaessa vaadittavan laskentatehon ja muistitarpeen.

3.3.2 Mukautuvan viipeen siloitin

Yllä esitetyssä kiinteän viipeen siloittimessa on valittu viive L suunnitteluparametri. Valittu viive on aina kompromissi: liian suuri viive kasvattaa siloitinjakauman estimaatin epätarkkuutta ja hidastaa laskentaa, kun taas liian pieni viive saattaa johtaa niin ikään epätarkkuuteen. Lisäksi valittu viive ei välttämättä johda jokaisella aika-askeleella optimaaliseen tai edes hyvään laskentatulokseen. Mukautuvan viipeen siloittimet yrittävät ratkaista tämän ongelman mukauttamalla kunakin ajanhetkenä valittua viivettä johonkin laskennalliseen kriteeriin perustuen. Eräään version mu-

kautuvan viipeen siloittimesta esittävät Johan Alenlöv ja Jimmy Olsson artikkelissa “Particle-Based Adaptive-Lag Online Marginal Smoothing in General State-Space Models” (2019) [1]. Siloitin hyödyntää hiukkassuotimen varianssiestimaattia viipeen valinnassa.

Yksinkertaisin versio siloittimesta on esitetty algoritmissa 8. Perusidea on viivästyttää siloitinjakauman luomista aika-askeleella k , kunnes tarjolla on viipeet $S = 1, \dots, s$, joiden varianssi

$$\sigma_{s|t}^2 = \sum_{i=1}^N \frac{w_t^i}{\Omega_t} \left\{ \tilde{x}_{s|t} - \sum_{j=1}^N \frac{w_t^j}{\Omega_t} \tilde{x}_{s|t} \right\}^2, \quad (3.7)$$

pysyy tietyn valitun rajan ϵ yläpuolella, missä $\tilde{x}_{s|t}$ on kyseiselle viipeellä laskettu marginaalisiloitinjakauman painovektori (kts. algoritmi 6). Kun tämä ehto ei enää täyty, käytetään suurimmalle kriteerin $\sigma^2 s|t < \epsilon$ täyttämälle viipeelle laskettuja painoja siloitinjakauman estimointiin kaikilla $t' \geq t$. Varianssin estimoinnista katso alaluku 2.3.

Algoritmi 8: Mukautuvan viipeen siloitin

Result: Siloittelujakauman estimaatti viipeellä s , tarkemmin $\sum_i^N w_t^i \tilde{x}_{s|t}^i \Omega_t$.

Data: Olkoon S joukko kullakin viipeellä s laskettuja painoja $\tilde{x}_{s|t}^i$. Alustetaan

```

 $S \leftarrow \emptyset$ 
begin
  for  $t = \{1, 2, \dots, T\}$  do
    begin
       $\quad \mid$  Ajetaan SIR-algoritmi 3 ajanhetkenä  $t$ 
      begin
         $\quad \mid$  Jokaiselle  $s \in S$  lasketaan painovektori kuten algoritmissa 6.
      begin
         $\quad \mid$   $S \leftarrow S \cup \{s\}$  Jokaiselle  $s \in S$  lasketaan painovektori kuten
          algoritmissa 6.
      begin
         $\quad \mid$  Jokaiselle  $s$  lasketaan varianssi  $\sigma_{s|t}^2$  kuten yhtälössä 3.7. Jos  $\sigma_{s|t}^2 < \epsilon$ 
          poistetaan  $s$  joukosta  $S$  ja käytetään siloitinjakauman estimaattia
           $\sum_i^N w_t^i \tilde{x}_{s|t}^i \Omega_t$  kaikille ajanhetkille  $t' \geq t$ .
    
```

Myös tähän siloittimeen liittyy suunnitteluparametrien valinta. Vaikka itse viivettä L ei valita, pitää parametri ϵ valita. Pienempi ϵ tuottaa suurempia viipeitä ja tätten parempia estimaatteja, mutta on myös laskennallisesti sekä muistin käytöltään raskaampi. Alenlöv ja Olsson ehdottavat ϵ -arvoja väliltä $(.5, 10^{-3})$.

Luku 4

Hiukkassuodin ja -siloitin sisätilapaikannuksessa

Sisätilapaikannus tarkoittaa nimensä mukaisesti ihmisten tai esineiden automaattista paikantamista sisätiloissa. Koska GPS-järjestelmät toimivat sisätiloissa huonosti tai eivät lainkaan, tarvitaan rakennusympäristöihin muita paikannusratkaisuja. Tässä luvussa käydään läpi sisätilapaikannuksen yleisiä periaatteita sekä joitakin ehdotettuja ratkaisuja. Tämän jälkeen keskitytään Walkbase-ohjelmistoyrityksen sisätilapaikan-nusteknologian ympärille kehitettyyn hiukkassuodinalgoritmiin, jonka tarkkuutta ja tehokkuutta testataan koeympäristössä.

4.1 Sisätilapaikannuksesta

Sisätilapaikannus tarkoittaa teknisiä ratkaisuja sekä menetelmiä, joilla paikannetaan ihmisiä, laitteita tai esineitä sisätiloissa, joissa perinteinen GPS-signaali ei ole riittävä tai saatavilla. Sisätilapaikannuksessa hyödynnetään useita erilaisia teknologioita, kuten radiotaajuksia, magneettikenttiä, ultraääntä ja optisia menetelmiä. Kattava esitys sisätilapaikannukseen käytettävistä teknologioista löytyy George Ogunatalan &al. artikkelista “Indoor location identification technologies for real-time IoT-based applications: An inclusive survey” (2018) [25], johon myös tämä alaluku perustuu.

Optisessa paikannuksessa hyödynnetään kameroita tai muita optisia/valoon perustuvia sensoreita, kuten esimerkiksi Lidar-valotutkia, halutun kohteen sijainnin määrittämiseen. Magneettikenttiin perustuvat paikannusratkaisut hyödyntävät rakennusten sisäisten metallirakenteiden maapallon magneettikenttään luomian pakkallisia muutoksia. Näitä muutoksia voidaan käyttää vertailutietona paikannuksessa, mittaan malla ympäröivän magneettikentän vahvuutta ja vertaan malla sitä ennalta tallennettuihin karttoihin. Ultraäänipohjainen paikannus puolestaan hyödyntää ääni-altojen kulkuaikaa lähettimen ja vastaanottimien välillä.

Yleinen teknologiavalinta sisätilapaikannuksessa ovat erilaiset Bluetooth-standardiin tai muuhun radioteknologiaan (kuten esimerkiksi millimetriaaltoihin) perustuvat lähetin-vastaanotinratkaisut. Näistä suurin osa hyödyntää laitteiden välisiä radiosignaaleja laitteiden välisen etäisyyden tai kulman mittauiseen. Radioteknologiaan perustuvilla järjestelmillä voidaan kotuullisin kustannuksin

saavuttaa jopa senttimetritason paikannustarkkuus sovelluksissa. Radioteknologiaan perustuvat järjestelmät ovat myös yksityisyyden näkökulmasta yksinkertaisempia toteuttaa kuin esimerkiksi kameroiden avulla tapahtuvaan paikannukseen perustuvat järjestelmät.

4.2 Teknologian kuvaus

Turkulainen teknologia- ja analytiikkayritys Walkbase käyttää Bluetooth/BLE-sisätilapaikannusta asiakkaiden käyttäytymistä koskevan datan keräämiseen erityisesti ruokakaupoissa sekä tavarataloissa. Tyypillisessä asennusskenarioissa lähettimet (tagit) kiinnitetään liikkeen ostoskärryihin ja paikantimet kiinnitetään liiketilan kattoripustuksiin.

Walkbase on kehittänyt sisätilapaikannukseen oman laitteisto- ja ohjelmistoratkaisunsa, jonka tavoitteena on tarjota kaikissa ympäristöissä alle metrin paikannustarkkuus. Koska tagit kiinnitetään tunnetulle korkeudelle ostoskärryihin, riittää paikannustarkkuutta laskettaessa tarkastella ainoastaan leveys- ja pituusasteita. Kun tagin todellinen sijainti $P_k = (P_{\text{lon}_k}, P_{\text{lat}_k})$ tiedetään diskreetillä aika-askeleella k , voidaan sijaintiestimaatin $\hat{x}_k = (\hat{x}_{\text{lon}_k}, \hat{x}_{\text{lat}_k})$ kyseisen aika-askeleen paikannusvirhe ϵ_{pos_k} laskea yksinkertaisesti

$$\epsilon_{\text{pos}_k} = d(P_k, \hat{x}_k) = \sqrt{(P_{\text{lon}_k} - \hat{x}_{\text{lon}_k})^2 + (P_{\text{lat}_k} - \hat{x}_{\text{lat}_k})^2}, \quad (4.1)$$

eli käytetään paikannusvirheenä yksinkertaisesti euklidista etäisyyttä. Kun hyödynnetään myöhemmin alaluvussa 4.3.1 esitettävä lineaarista interpolatiota, antaa tämä paikannusvirheen suoraan metreinä. Haluttu, alle metrin paikannustarkkuus saavutetaan, kun missä hyvänsä testiasetelmassa kaikkien asetelmaan liittyvien paikannusvirheiden 50. persentiili on $\leq 1\text{m}$.

Walkbasen paikannusratkaisu koostuu kolmesta eri laitteistokomponentista, AT-2-Bluetooth-lähetin-vastaanottimesta, jotka kiinnitetään ostoskärryihin, XR-2-Bluetooth-lähetin-vastaanottimesta, jotka kiinnitetään tilan kattoripustuksiin sekä OSCU-laskentayksiköstä, joka luo paikkadataa XR-2.1-vastaanotinten perusteella ja lähetää paikkadatan edelleen palvelinkeskukseen.

AT-2 on Bluetooth 5.1 (BLE) -strandardin mukainen lähetin-vastaanotin, joka toimii 2.4Ghz taajuusalueella. Walkbasen suunnittelemän laitteen PCB-kehäantenni kykenee lähetämään GFSK-moduloitua dataa 2Mbps nopeudella. Laite saa virtansa yhdestä CR-2477-paristosta.

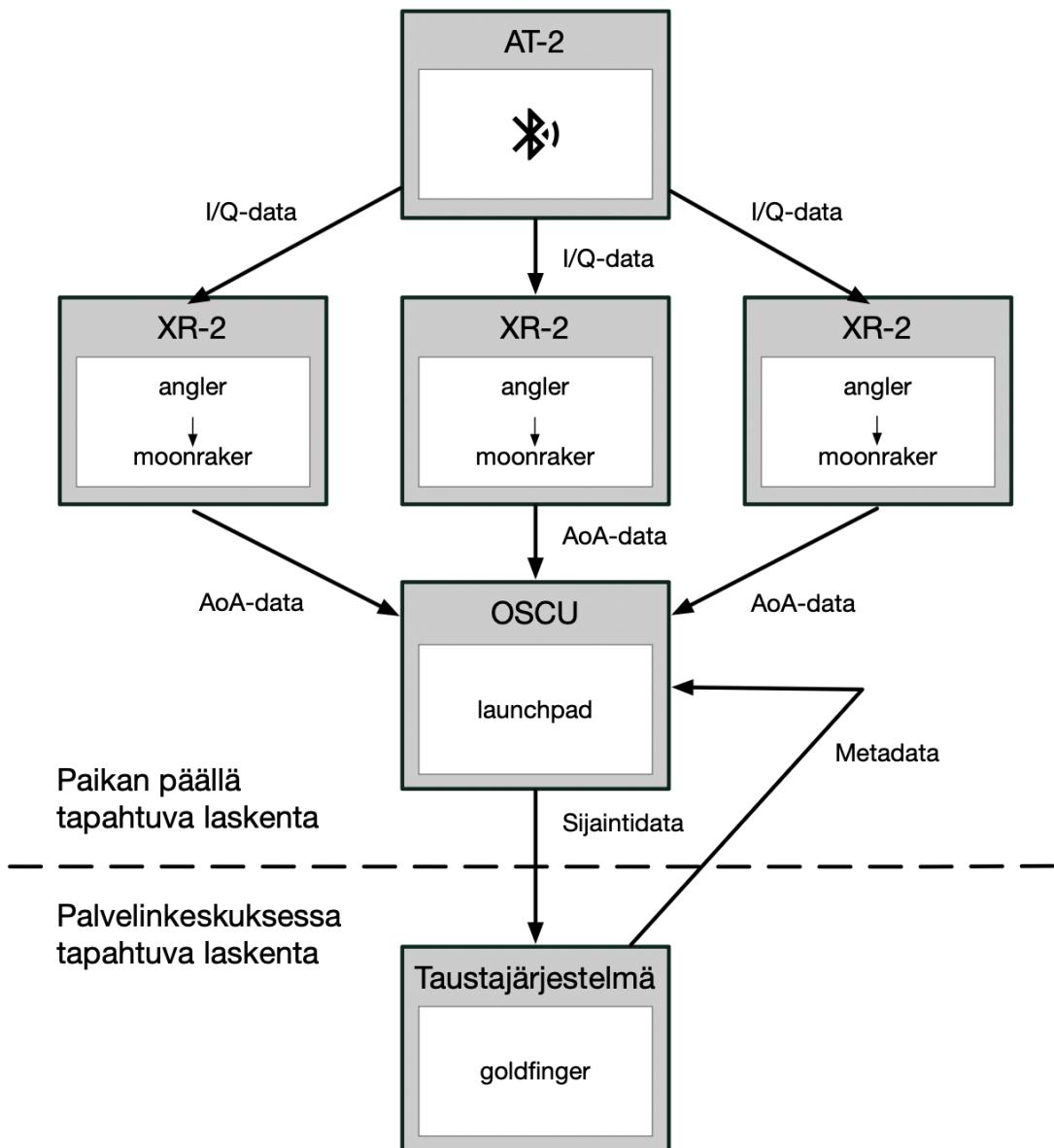
XR-2.1 on Bluetooth 5.1 (BLE) -strandardin mukainen lähetin-vastaanotin, joka toimii 2.4Ghz taajuusalueella. Walkbasen suunnittelemän laitteen PCB-kehäantenni kykenee lähetämään GFSK-moduloitua dataa 2Mbps nopeudella. Laite saa virtansa ethernet-lähiverkosta 802.3af-standardin mukaisesti. Laitteen vaatima laskenta tapahtuu Raspberry Pi Compute Module 4 -piirilevytietokoneella. Lisäksi laite sisältää inertiamittausyksikön, jota voidaan käyttää asennetun laitteen kallistumis- ja nyökkäämiskulman (*roll* ja *pitch*) arviomiseen.

OSCU-laskentayksikkönä käytetään Ubuntu-käyttöjärjestelmällä toimivaa paikannusjärjestelmää varten rakennettua tietokonetta. Koska OSCU-laskentayksikön

laskentateho on rajallista, on paikannusalgoritmin aikakompleksisuus yksi käytettävän algoritmin ydinkriteereistä.

Tarvittavasta laskennasta vastaava ohjelmistojärjestelmä koostuu neljästä ohjelmistokomponentista. C-ohjelmointikielellä toteutettu *angler* laskee AT-2-tagin lähetämän I/Q-datan perusteella signaalien tulokulman, Go-ohjelmointikielellä toteutettu *moonraker* lähetää tulokulmadatan paikallisverkon yli OSCU-laskentayksikölle, jossa Go-ohjelmointikielellä toteutettu *launchpad* luo siitä sijaintidataa, jonka se lähetää edelleen palvelinkeskuksen taustajärjestelmään.

Palvelinkeskuksen taustajärjestelmässä Go-ohjelmointikielellä toteutettu *goldfinger* prosessoi sijaintidatan Walkbasen analytiikka-alustan käyttämään muotoon. *Goldfinger* vastaa myös siitä, että kaikki *launchpad*-sovelluksen vaatima metadatta on *launchpad*-sovelluksen käytössä. Kaavio 4.1 kuvailee järjestelmän laitteisto- ja ohjelmistoarkkitehtuurit.

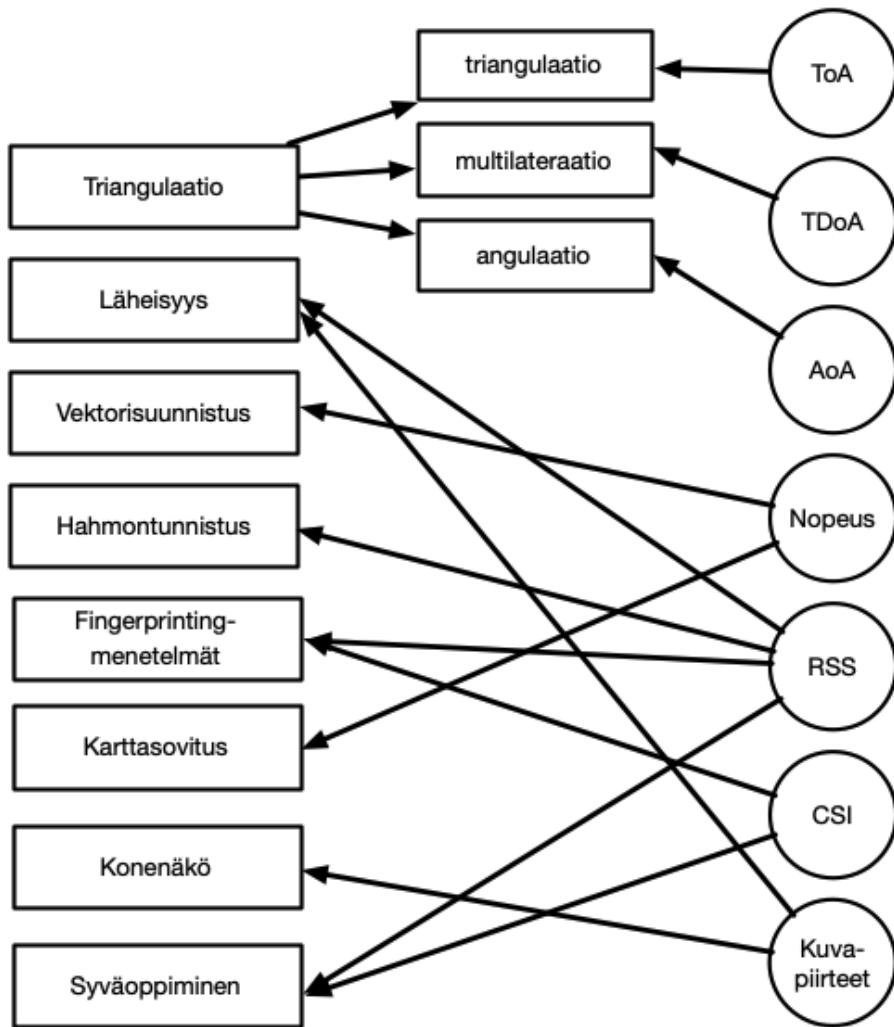


Kuva 4.1: Järjestelmäarkkitehtuuri

Tässä luvussa keskitytään *launchpad*-sovelluksen hyödyntämään paikannusalgoritmiin, mutta sitä ennen käsitellään lyhyesti tulokulman laskentamenetelmiä sekä *angler*-sovelluksen toimintaa. Immateriaalinoikeusyistä tutkielmanassa ei hyödynnetä Go-ohjelmointikielellä toteutettua ohjelmaa. Sen sijaan paikannusalgoritmi on toteutettu R-ohjelmointikielellä.

4.2.1 AoA-menetelmistä

Riippuen käytetystä teknologiasta ja teknologian tuottamasta datasta, voidaan sisätilapaikannuksessa soveltaa lukuisia eri paikannustekniikoja ja -algoritmeja. Oguntalaan (2018) [25] mukailevassa kaaviossa 4.2 on esitetty pääpiirteittään sisätilapaikannusmenetelmät käytetyn datan (kaaviossa oikealla) jaoteltuna.



Kuva 4.2: Paikannusmenetelmien luokittelu

Walkbasen toteuttama paikannusratkaisu perustuu AoA-menetelmään. Tämä on paikannusteknologia, joissa lähettimen ja vastaanottimen välinen kulma estimoidaan signaalin saapumiskulman perusteella. Tässä estimoinnissa hyödynnetään signaalin vaihe-eroja, kun sama signaali vastaanotetaan usealla eri antennilla.

Walkbase hyödyntää radiosignaalien tulokulman estimoinnissa MUSIC-algoritmia (*Multiple Signal Classification*), joka arvioi tulokulmia analysoimalla signaalien autokorrelatiomatriisia ja etsimällä sen ominaisvektoreita. Esitys MUSIC-algoritmista löytyy esimerkiksi Monson H. Hayesin kirjasta *Statistical*

Digital Signal Processing and Modeling (1996) [15]. Marco Guniam & al (2023) [13] puolestaan esittävä artikkelissa “Analysis and Design of a MuSiC-Based Angle of Arrival Positioning System” tapoja optimoida MUSIC-algoritmia soveltumaan monimutkaisiin, heijastuksia sisältäviin sisätilaympäristöihin.

Angler-sovellus hyödyntää MUSIC-algoritmin ohella omaa hiukkassuodinalgoritmiaan signaalisketrin analysointiin sekä tulokulmien estimointiin. MUSIC-algoritmia tai *angler*-sovelluksen hiukkassuodinalgoritmia ei käsitellä tarkemmin tämän tutkielman puitteissa.

4.2.2 Kalibraatiosta

Koska *angler*-sovellus laskee suuntimakulman aina antennielementin määräytystä reunaasta nähdien, on tärkeää, että laitteen asemointi karttapohjoiseen nähdien on tiedossa. Koska laitteen tarkan kulman arvioiminen asennusympäristössä on haastavaa eikä laite sisällä kompassia, sovelletaan kunkin laitteen kalibraatioon omaa SIR-algoritmiaan, jonka on mahdollista huomioon myös laitteen IMU-intertiamittausyksiköstä saatavat kallistumis- ja nyökkäämiskulmat.

Jos IMU-dataa hyödynnetään, tuottaa kalibraatioalgoritmi jokaiselle XR-2-laitteelle rotaatiomatriisiin R_l . Muussa tapauksessa kalibraatioalgoritmi tuottaa jokaiselle laitteelle atsimuuttiakulman kohdistuksen η . Kalibraatioalgoritmia ei käsitellä tämän tutkielman puitteissa. Paikannusongelman yksinkertaistamiseksi myöskään IMU-yksikön tuottamaa dataa ei sisällytetä kalibraatioon ja jokaisen XR-2-laitteen oletetaan olevan lattiaan nähdien vaakatasossa. Atsimuuttipulman kalibraatiokohdista lisää luvussa 4.3 alla.

4.3 Datat kuvaus

Koeasetelmaa varten AT-2-tagit on asetettu lähettämään IQ-dataotoksia 6hz taajuudella, joka vastaa järjestelmän tuotantokäytöä. Parempi sijaintitarkkuus saavutettaisiin korkeammalla taajuudella, mutta käytännön sovellutuksissa ei tagien akun kesto salli 6hz korkeampaa lähetystaajuutta.

Angler-sovellus koostaa jokaisesta tagin lähettämästä dataotoksesta s atsimuutteli suuntimakulman θ_s ja korkeuskulman γ_s , jotka noudattavat suuntimakulman osalta von Mises -jakaumaa ja korkeuskulman osalta katkaistua normaalijakaumaa:

$$\theta_s \sim \text{von Mises}(\mu_{\theta_s}, \kappa_k), \quad (4.2)$$

$$\gamma_s \sim \mathcal{N}_{\text{katkaistu}}(\mu_{\gamma_s}, \sigma_s^2). \quad (4.3)$$

Nämä on kuvattu tarkemmin alaluvussa 4.4.2. Jokaisen XR-2-laitteen *angler*-sovellus lisää dataan myös sekvenssinumeron, jonka perusteella samasta AT-2-tagin lähettämästä IQ-dataoksesta lasketut usean eri vastaanottimen laskemat tulokulmat voidaan yhdistää samaan IQ-dataotokseen. XR-2-laitteen lähettämä tulokulmadata on kuvattu alla.

Muuttuja	Kuvaus	Esimerkkiarvo
id	havainnon yksilöivä tunniste	317092
ts	havainnon aikaleima	2024-04-08 21:38:20.998+00
locator_mac	XR-2-laitteen MAC-osoite	2c:e3:10:00:07:a6
asset_tag_mac	AT-2-tagin MAC-osoite	2c:e3:10:00:63:89
sequence_nr	kulmadatan IQ-dataotokseen yhdistäävä juokseva numerointi	2066
azimuth_location	atsimiuuttikulman θ jaukauman sijaintiparametri μ_θ (rad)	0.39
azimuth_scale	atsimiuuttikulman θ jaukauman skaalaparametri κ	80.98
elevation_location	jaukauman sijaintiparametri μ_γ (rad) korkeuskulman γ	0.13
elevation_scale	jaukauman skaalaparametri σ^2	0.012
quality_sndr	signaali-kohinasuhde	22.0
rssi	signaalin vahvuus (dBm)	-81
distance	arvioitu etäisyys lähettimeen (m)	18.6

Taulukko 4.1: Tulokulmamuuttujat

Etäisyys on estimoitu signaalin vahvuudesta käyttäen propagaatiomallia. Etäisyyttä tai signaalin vahvuutta ei käytetä paikantamiseen, joten tämän mallin käsittely jätetään tutkielman ulkopuolelle. Munoz (2009) luku 2 sisältää yleiskatsauksen propagaatiomalleista [23].

Launchpad-sovelluksessa tulokulmadataan yhdistetään XR-2-laitteen MAC-osoitteiden perusteella tarvittava, XR-2-laitteita koskevat metadatamuuttujat. Näihin kuuluvat laitteen korkeus, laitteen suuntimakulma ja karttakoordinaatit. Metadata on kuvattu taulukossa 4.2.

Muuttuja	Kuvaus	Esimerkkiarvo
locator_mac	XR-2-laitteen MAC-osoite	b8:27:eb:66:0d:2a
lat	vastaanottimen sijainti (leveyspiiri)	60.448265
lon	vastaanottimen sijainti (pituuspiirit)	22.294823
direction	suuntimakulma η (astetta)	34
height	vastaanottimen korkeus (m)	2.22

Taulukko 4.2: Metadata

Atsimiuuttikulma θ lasketaan aina vastaanottimen tietyltä sivulta, joten se vastaa napapohjoista ainoastaan siinä tapauksessa, että vastaanottimen kyseinen sivu on asetettu kohtisuoraan napapohjoiseen nähden. Käytännössä vastaanottimien asettaminen tiettyyn kulmaan ei ole aina mahdollista eikä vaihe-erojen mittaan

kannalta edes suotavaa. Tämän vuoksi jokaiselle vastaanottimelle on tietokantaan tallennettu oma suuntimakulma η . Toisin kuin tulokulmadatan kulmat, on tämä tallennettu tietokantaan astineina. Kokeessa käytetään napapohjoisesta laskettuja kulmia Φ , jotka lasketaan jokaiselle havainnolle havainnon vastaanottimen suuntimakulman avulla

$$\Phi = (\theta + \eta \times \frac{\pi}{180^\circ}) \mod 2\pi. \quad (4.4)$$

Suuntimakulma Φ kertoo vastaanottimen ja lähettimen välisen kulman. Lisäksi saatavilla on PostGIS-muotoon tallennettua polygonidataa, joka vastaa koeympäristön pohjapiirrustusta sekä koeympäristössä esiintyviä liikkumisen estäviä kohteita, kuten hyllyjä tai pöytää. Näitä hyödynnetään karttasovitusalgoritmissa (kts. alaluku 4.4.4.2).

Havaintomuuttujien ohella koetilanteesta on tallennettu testipolku, jota pitkin AT-2-tagia liikutetaan koetilanteessa. Testipolkudata pitää sisällään karttaan piirrettyn janan päät- ja sisäpisteet. Tallennettu sijainti perustuu koeympäristön lattiaan pohjapiirrustusten sekä laser-mittausten avulla tehtyihin merkintöihin. Näin saadut testimuuttujat on kuvattu taulukossa (4.3).

Muuttuja	Kuvaus	Esimerkkiarvo
path_lat	polkupisteen sijainti (leveyspiiri)	60.44819
path_lon	polkupisteen sijainti (pituuspiiri)	22.29493

Taulukko 4.3: Testimuuttujat

Edelleen alaluvussa 4.4.4.2 kuvatulla interpolatiolla metreiksi muutettuja testimuuttuja käytetään paikannusalgoritmien paikannusvirheen laskemisessa. Koska koeasetelmassa (kts. alaluku 4.5.1) ei laitteen sijaintia tallenneta mittauksia tehtäessä, laskennetaan paikannusvirhe janojen päättepisteiden joukosta tiivistämällä joukkoa algoritilla 9.

Algoritmi 9: Paikannusvirheen laskeminen

Result: Aika-askeleen k paikannusvirhe ϵ_{pos_k} (m);
Data: Metreiksi muutetut testimuuttujat sisältävä pistejoukko P . Aika-askeleen k sijaintiestimaatti \hat{x}_k . Tihennysparametri s , vakioarvo $s = 1000$.

```
begin
  begin
    Tihennetään pistejoukko  $P$ . Lisätään joukosta  $P$  luodulle janoille pisteitä lineaarisella interpolatiolla, kunnes joukossa on  $s$  pistettä. Saadaan joukko  $S$ ;
  begin
    Asetetaan  $\hat{\epsilon}_{\text{pos}_k} = \infty.$ ;
  for  $i = \{1, 2, \dots, s\}$  do
    begin
      Lasketaan sijaintiestimaatin  $\hat{x}_k$  ja pisteen  $S_i$  välinen Euklidinen etäisyys  $d_i$ .
      if  $d_i < \hat{\epsilon}_{\text{pos}_k}$  then
        begin
           $\hat{\epsilon}_{\text{pos}_k} = d_i$ ;
    begin
      Palautetaan aika-askeleen  $k$  paikannusvirhe  $\epsilon_{\text{pos}_k} = \hat{\epsilon}_{\text{pos}_k}$ ;
```

Algoritmiin palataan alaluvun 4.5 empiirisessä esimerkissä.

4.3.1 Karttaprojektiosta

Kaikki yllä esitetyssä datassa esiintyvät sijaintikoordinaatit on tallennettu tietokantaan WGS 84 -tasokoordinaattijärjestelmässä. Hiukkassuotimiin perustuva paikanusalgoritmia sovellettaessa on kuitenkin monin paikoin tarve syöttää parametereja metrijärjestelmässä. Tästä syystä leveys- ja pituusasteisiin perustuvat koordinaatit muunnetaan laskentaa varten metreiksi ja metreinä esitetyt sijaintitulokset muutetaan tulosten esittämistä varten takaisin WGS 84 -koordinaattijärjestelmään.

Muunnos tapahtuu lineaarisella interpolatiolla. Määritellään ensin kerrospolygonin rajausalue. Koska kaikki käytetyt koordinaatit ovat tämän rajausalueen sisällä, voidaan tästä rajausaluesta käyttää metrikonversiossa. Rajausalue koostuu neljästä kulmapisteestä. Poimitaan näistä pisteistä minimit ja maksimit sekä pituus- että leveyskoordinaateille. Näin saadaan neljä arvoa $B_{\text{lonlat}} = \{\text{lon}_{\min}, \text{lon}_{\max}, \text{lat}_{\min}, \text{lat}_{\max}\}$. Määritellään rajausalueen sivujen pituus metreinä geodeettisen etäisyyden avulla, jolloin saadaan kaksi metreissä laskettua etäisyyttä $D_m = \{d_{\text{lon}}, d_{\text{lat}}\}$. Käytetään näitä kulmapisteitä sekä metreinä laskettuja etäisyyksiä interpoloimaan koordinaatit WGS 84 -koordinaattijärjestelmästä metreissä esitetyille arvoalueille $[0, \max(x_m)]$ ja $[0, \max(y_m)]$ seuraavasti:

$$x_m = f(x_{\text{lon}}; B_{\text{lonlat}}, D_m) = \frac{x_{\text{lon}} - \text{lon}_{\min}}{\text{lon}_{\max} - \text{lon}_{\min}} \times d_{\text{lon}} \quad (4.5)$$

$$y_m = f(y_{\text{lat}}; B_{\text{lonlat}}, D_m) = \frac{y_{\text{lat}} - \text{lat}_{\min}}{\text{lat}_{\max} - \text{lat}_{\min}} \times d_{\text{lat}}, \quad (4.6)$$

missä x_m vastaa leveyskoordinaatteja metreissä ja y_m pituuskoordinaatteja metreissä. Vastaavasti käänös takaisin WGS 84 -koordinaattijärjestelmään tapahtuu seuraavasti:

$$x_{\text{lon}} = f(x_m; B_{\text{lonlat}}, D_m) = \frac{x_m}{d_{\text{lon}}} \times (\text{lon}_{\max} - \text{lon}_{\min}) + \text{lon}_{\min} \quad (4.7)$$

$$y_{\text{lat}} = f(y_m; B_{\text{lonlat}}, D_m) = \frac{y_m}{d_{\text{lat}}} \times (\text{lat}_{\max} - \text{lat}_{\min}) + \text{lat}_{\min}. \quad (4.8)$$

4.3.2 Muunnettua dataa

Kun dataan on tehty yllä esitetyt konversiot ja metadata on liitetty jokaisen aika-askeleen k dataan, saadaan data lopulliseen, empiirisessä esimerkissä käytettävä muotoon, jossa jokaisen aika-askeleen kohdalla k on käytettävissä seuraava data:

Muuttuja	Kuvaus	Esimerkkiarvo
id	havainnon yksilöivä tunniste	317092
ts	havainnon aikaleima	2024-04-08 21:38:20.998+00
locator_mac	XR-2-laitteen MAC-osoite	2c:e3:10:00:07:a6
asset_tag_mac	AT-2-tagin MAC-osoite	2c:e3:10:00:63:89
sequence_nr	kulmadatan IQ-dataotokseen yhdistäävä juokseva numerointi atsimuuttikulman θ	2066
azimuth_location_mdf	jaukauman muunnettua sijaintiparametri Φ (rad)	0.39
azimuth_scale	atsimutikulman θ	80.98
elevation_location	jaukauman skaalaparametri κ	0.13
elevation_scale	korkeuskulman γ	0.012
quality_snrd	jaukauman sijaintiparametri μ_γ (rad)	22.0
rssi	signaali-kohinasuhde	-81
distance	arvioitu etäisyys lähettimeen (m)	18.6
x_m	vastaanottimen sijainti (x -koordinaatti, metreinä)	1.24
y_m	vastaanottimen sijainti (y -koordinaatti, metreinä)	0.78
height	vastaanottimen korkeus (m)	2.22

Taulukko 4.4: Muunnettua data

4.4 Sisätilapaikanusalgoritmi

4.4.1 Ongelman kuvaus

Tarkoituksena on estimoida liikkuvan AT-2-tagin sijaintia. Merkitään estimoitavaa tilasarjaa $x_{1:k} = \{x_1, \dots, x_k\}$, missä sijainnit on estimoitu sekunnin tarkkuudella. Lisäksi merkitään x_0 testilaitteen lähtösijaintia. Jokainen tilasarjan havainto koostuu suuntimakulmasta sekä pituus- että leveyskoordinaateista (x_k^x, x_k^y) . Määritellään tilalle liikkuvan AT-2-tagin kulkua kuvaava vektorisuunnistukseen (*dead reckoning*) perustuva dynaaminen malli (4.9)

$$x_{k+1} = f(x_k, \nu_k) = x_k + D_k \begin{bmatrix} \cos \psi_k \\ \sin \psi_k \end{bmatrix} + \nu_k, \quad (4.9)$$

missä D_k on AT-2-tagin sekunnin aika-askeleen $k \rightarrow k+1$ aikana kuljettama matka ja ψ_k AT-2-tagin liikkeen suuntimakulma kyseisellä aika-askeleella. ν_k on kohinaa,

joka syntyy mittausvirheestä ja jolle voidaan olettaa $\sim \mathcal{N}(\mu_x, \sigma_x^2)$. Jos laite on paikallaan, yksinkertaistuu malli muotoon $x_{k+1} = f(x_k) = \text{id}(x_k) = x_k$, missä $\text{id}(x)$ on identiteettifunktio.

Vastaavasti $y_{k_{i,j}} = \{y_{k_{1,1}}, \dots, y_{k_{n_s, h_z}}\}$ kuvaa AT-2-tagin ja XR-2-laitteiden välillä yhden sekuntin aika-askeleen aikana laskettuja kulmahavaintoja, missä $i = 1, \dots, n_s$ käy läpi n_s XR-2-laitetta ja $j = 1, \dots, h_z$ kustakin laitteesta vastaanotetut kulma-havainnot. h_z ilmaisee tässä AT-2-tagin lähetystaajuutta, joka koeasetelmassa on $h_z = 6$. Näin ollen jokainen havainto koostuu (maksimissaan) $6 \times n_s$ kulmahavaintoja. Lisäksi tunnetaan XR-2-laitteisiin $\{s^1, \dots, s^{n_s}\}$ liittyvät pituus- ja leveyskoordinaatit (λ, ϕ) , jotka on muutettu alaluvussa 4.3.1 esitettyllä interpolatiolla metreiksi.

$$u = \begin{bmatrix} \lambda^1 & \phi^1 \\ \lambda^1 & \phi^1 \\ \vdots & \vdots \\ \lambda^{n_s} & \phi^{n_s} \end{bmatrix}, \quad (4.10)$$

missä jokainen koordinaattipari on toistettu h_z kertaa. Määritellään nyt havaintomalli

$$y_{k_{i,j}} = h(x_k, u) + e_k = \text{atan2}\left(\begin{bmatrix} \phi^1 - x_k^y \\ \phi^1 - x_k^y \\ \vdots \\ \phi^{n_s} - x_k^y \end{bmatrix}, \begin{bmatrix} \lambda^1 - x_k^x \\ \lambda^1 - x_k^x \\ \vdots \\ \lambda^{n_s} - x_k^x \end{bmatrix}\right) + e_k, \quad (4.11)$$

missä

$$\text{atan2}(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{jos } x > 0, \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{jos } x < 0 \text{ ja } y \geq 0, \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{jos } x > 0 \text{ ja } y < 0, \\ +\frac{\pi}{2} & \text{jos } x = 0 \text{ ja } y > 0, \\ -\frac{\pi}{2} & \text{jos } x = 0 \text{ ja } y < 0, \\ \text{ei määritelty} & \text{jos } x = 0 \text{ ja } y = 0 \end{cases} \quad (4.12)$$

ja kohina noudattaa moniulotteista normaalijakaumaa $e_k \sim \mathcal{N}(0, \Sigma)$.

Määrittelemätön atan2-tapaus, jossa $x = 0$ ja $y = 0$ on käytettyllä mittaustarkkuudella käytännössä mahdoton. Jos tapaus halutaan välttää, voidaan nolla-arvot tarpeen vaatiessa korvata joillakin hyvin lähellä nollaa olevalla arvolla.

Kumpikaan funktiosta $h(\cdot)$ ja $f(\cdot)$ ei ole lineaarinen, joten SIR-algoritmi on sopiva valinta ongelman ratkaisemiseksi. Koetuloksia arvioidaan ensisijaisesti paikan-nusvirheen avulla. Paikannusvirhe ϵ_{pos_k} lasketaan jokaisen aika-askeleen k posteriori-jakaumaestimaatista \hat{p}_k painotettuna keskiarvona

$$\epsilon_{\text{pos}} = d(P, \sum_{i=1}^N w_i^k x_k^i), \quad (4.13)$$

missä w_i^k on ajanketken k partikkeliensä normalisoitu paino ja $d(x_k^i, y_k)$ partikkeliensä ja testilaitteen todellisen sijainnin P välisen etäisyyden laskeva algoritmi 9.

4.4.2 Uskottavuusmallit

Jokainen yhtä tulokulmaa vastaava *angler*-sovelluksen tuottama havaintodatarivi pitää sisällään neljä parametrimuuttuja, **azimuth_location_mdf**, **azimuth_scale**, **elevation_location** ja **elevation_scale**. Koska *angler*-sovellus on kirjoitettu varta vasten tuottamaan dataa hiukkassuodinpaikannusalgoritmia varten, ovat nämä suoraan hiukkassuotimen uskottavuusmallin parametreja.

Angler-sovelluksessa XR-2-laitteen ja AT-2-tagin välinen atsimuuttikulma simuloidaan von Mises -jakaumasta, jolloin muuttujat **azimuth_location_mdf** ja **azimuth_scale** vastaavat tämän jakauman sijainti- ja skaalaparametreja μ ja κ ja näistä edellistä voidaan pitää itse atsimuuttikulman estimaattina. Määritellään siis jokaiselle tulokulmalle h sekä XR-2-laitteelle l seuraava atsimuuttikulman uskottavuusmalli:

$$L_{\theta_{h,l}}(y_{h,l}|x_{h,l}; \mu_{h,l}, \kappa_{h,l}) = \frac{e^{\kappa_{h,l} \cos(x_{h,l} - \mu_{h,l})}}{2\pi I_0(\kappa_{h,l})}, \quad (4.14)$$

missä I_0 on 0:s ensimmäisen lajin Bessel-funktio ja $x_{h,l}$ on jokaisen hiukkasen $n = 1, \dots, N$ sekä XR-2-laitteen l välinen suuntimakulma.

Vastaavasti *angler*-sovelluksessa XR-2-laitteen ja AT-2-tagin välinen korkeuskulma simuloidaan katkaistusta normaalijakaumasta, jolle $a = 0$ ja $b = 2\pi$. Nyt muuttujat **elevation_location** ja **elevation_scale** vastaavat tämän jakauman sijainti- ja skaalaparametreja μ ja σ^2 ja näistä edellistä voidaan pitää itse korkeuskulman estimaattina. Määritellään jokaiselle tulokulmalle h sekä XR-2-laitteelle l seuraava korkeuskulman uskottavuusmalli:

$$L_{\gamma_{h,l}}(y_{h,l}|x_{h,l}; \mu_{h,l}, \sigma_{h,l}^2, a = 0, b = 2\pi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_{h,l} - \mu_{h,l})^2}{2\sigma^2}\right) / \sqrt{\sigma^2} Z_{h,l}, \quad (4.15)$$

missä $Z_{h,l} = \frac{1}{2}(1 + \operatorname{erf}(\frac{b - \mu_{h,l}}{\sqrt{\sigma^2}})) - \frac{1}{2}(1 + \operatorname{erf}(\frac{a - \mu_{h,l}}{\sqrt{\sigma^2}}))$ ja $x_{h,l}$ on jokaisen hiukkasen $n = 1, \dots, N$ sekä XR-2-laitteen l välinen korkeuskulma.

Uskottavuusmallit 4.14 ja 4.15 kertomalla saadaan yhdistetty uskottavuusmalli hiukkassuotimen aika-askeleelle k sekä XR-2-laitteelle l

$$L_{k,l} = \prod_{h=1}^{hz} L_{\theta_{h,l}} \times \prod_{h=1}^{hz} L_{\gamma_{h,l}}, \quad (4.16)$$

josta voidaan edelleen laskea jokaisen hiukkasen uskottavuus aika-askeleella k

$$L_k = \prod_i^{n_s} L_{k,i}. \quad (4.17)$$

Koska yllä esitettyjen mallien uskottavuudet ovat käytännössä erittäin pieniä, käytetään numeerisista syistä itse algoritmissa logaritmoituja uskottavuusmalleja, jolloin $l_{k,l} = \log(\prod L_{\theta_{k_l}}) + \log(\prod L_{\gamma_{k_l}})$ ja $l_k = \sum_i^{n_s} l_{k,i}$.

4.4.3 Datan valinta

Koska Walkbasen paikannusjärjestelmä toimii monimutkaisessa, heijastuksia sisältävässä radioympäristössä, suoritetaan jokaisella aika-askeleella k vielä ylimääriäinen datan valinta. Valinnan tarkoituksesta on jättää mahdollisia heijastuksia sisältävät tai muuten selkeästi virheelliset tulokulmat paikannusdatan ulkopuolelle.

Yksinkertaisin tapa valita käytettävät kulmat on asettaa alaraja signaalin vahvuudelle, jolloin ainoastaan tiettyä kynnys-dBm-arvoa suuremman signaalivahvuuden omaavat tulokulmat valitaan paikannukseen. Datan valinta suoritetaan kunkin aika-askeleen alussa ja tyypillisesti kynnysarvoja ovat esimerkiksi -100 dBm, -90 dBm ja -80 dBm. Jos kynnysarvo ei ole käytössä, käytetään ohjelmakoodissa arvoa -120 dBm, joka jättää kaikki tulokulmat dataan. Datan valintaan palataan parametrien valinnan yhteydessä alaluvussa 4.5.2.

4.4.4 Dynaaminen malli

Dynaamisen mallin tehtäväänä on liikuttaa hiukkasia aika-askeleiden k ja $k+1$ välillä. Dynaaminen malli perustuu vektorisuunnistusmalliin 4.9. Vastaavaa mallia hyödyntää muun muassa Solin (2016) [28]. Koska AT-2-tagi ei kuitenkaan lähetä nopeus- tai suuntadataa, vektorisuunnistusmallin nopeutta ilmaiseva skalaarimuuttuja $D_k = 0$, jolloin malli yksinkertaistuu satunnaiskävelymalliksi

$$x_{k+1} = x_k + v_k, \quad (4.18)$$

missä v_k on kohinaa. Kohina luodaan jokaiselle aika-askeleelle seuraavasti. Ensinnäkin luodaan etäisyysvektori d_k katkaistusta normaalijakaumasta, jonka sijaintiparametri on 0, katkaisukohdat $a = 0$ ja $b = 10$ ja jonka keskijointa q on paikannusalgoritmin suunnitteluparametri:

$$d_k \sim \mathcal{N}_{\text{katkaistu}}(\mu = 0, \sigma = q, a = 0, b = 10). \quad (4.19)$$

Tämän jälkeen luodaan suuntavektori p tasajakaumasta

$$p_k \sim \mathcal{U}(0, 2\pi) \quad (4.20)$$

ja lopulta kohinavektori

$$v_k = (\cos(p_k) \times d_k, \sin(p_k) \times d_k), \quad (4.21)$$

missä vektorin ensimmäinen elementti vastaa sijaintien leveysasteiden kohinaa ja toinen elementti pituusasteiden kohinaa.

4.4.4.1 Paikallaanolon havaitseminen

Virransäästyösyistä AT-2-tagi lähetää dataa ainoastaan liikkueessaan. Jos laite ei ole 10 sekunnin aikana havainnut IMU-yksikkönsä perusteella kiihtyvyyttä, lopettaa laite datan lähetämisen. Laite on tällöin valmismoodissa. Tätä tietoa voidaan hyödyntää poistamalla dynaamisesta mallista kohina, kun laitteen tiedetään olevan paikallaan. Jos paikannusalgoritmiin lisätään tarvittava 10 sekunnin viive, voidaan tätä viivettä hyödyntää paikallaolon tehokkaampaan havaitsemiseen.

Jos yksikään XR-2-laite ei ole havainnut tagia aika-askeleen aikana, voimme olettaa laitteen olevan valmismoodissa ja siten myös paikoillaan. Huomattavaa on kuitenkin, että pääinvastainen ei päde. Paikoillaan oleva laite ei välttämättä ole valmisumoodissa, sillä valmismoodiin siirtyminen kestää 10 sekuntia. Tähän oletukseen perustuen voidaan havaita paikallaolon ja vaimentaa dynaamista mallia algoritmin 10 avulla. Algoritmi ajetaan jokaisella aika-askeleella k ennen hiukkasten siirtoa dynaamisen mallin avulla.

Algoritmi 10: Paikallaanolon havaitsemisalgoritmi

Result: Positiivinen kokonaislukumuuttuja m , joka osoittaa kuinka monaksi aikahetkeksi dynaamista mallia tulee vaimentaa. Jos $m = 0$ mallia ei vaimenneta.;

Data: Tagin tulokulmadata $L + 1$ aika-askeleelle (nykyinen aika-askel + L aika-askelta tulevaisuuteen). L tulee asettaa niin, että paikallaan oleva tagi ehtii valmiustilaan. Jos saatavilla, edellinen muuttujan m arvo. Algoritmin ensimmäisellä ajokerralla asetetaan $m = 0$;

```

begin
  begin
    if  $m > 0$ , asetetaan  $m = f(m) = m - 1$  ja pysytetään algoritmi. Jos
       $m = 0$  jatketaan algoritmin suorittamista.;

    for  $l = \{k + 1, \dots, k + L\}$  do
      begin
        if  $o_l$  kulmahavaintojen määrää aika-askeleella  $l$ ;
        if  $o_l = 0$  ja  $m = 0$  then
          begin
            Asetetaan  $m = l - k$ ;;
        else
          if  $o_l = 0$ ,  $m > 0$  ja  $m + 1 = l - k$  then
            begin
              Asetetaan  $m = l - k$ ;;

```

Yllä esitetty algoritmi etsii ensin ensimmäisen aika-askeleen, jonka aikana ei ole tallennettu lainkaan kulmahavaintoja ja asettaa muuttujan m arvon vastaamaan tätä aika-askelta. Jos algoritmi löytää useita peräkkäisiä aika-askeleita, joiden aikana ei ole tallennettu lainkaan kulmahavaintoja, valitsee se näistä suurimman. Koska koeasetelmassa AT-2-laitteen valmiustila aktivoituu 10 sekunnin kohdalla, valitaan $L = 10$.

Jos paikallaanolon havaitsemisalgoritmin nojalla dynaamista mallia päädystään vaimentamaan (so. $m > 0$), asetetaan dynaamisen mallin keskihajonta-arvo $q_{vaimennettu} = \frac{q}{10}$. Dynaamista mallia ei siis täysin poisteta käytöstä, jotta paikannusalgoritmi pystyy tehokkaammin hyödyntämään myös vaimennettujen aika-askeleiden dataa ja sijaintiestimaatti konvergoi mahdollisimman lähelle todellista sijaintia, johon tagi on pysähtynyt.

4.4.4.2 Karttasovitusalgoritmi

Dynaamista mallia sovellettaessa, voimme myös höydyntää saatavilla olevaa sisätilan karttadataa. Määritellään kaksi polygonityyppiä, lattiapolygoni F sekä estepolygonit $E = \{E_1, \dots, E_n\}$. Lattiapolygonit määrittävät alueen, jonka sisällä hiukkassuodinalgoritmien hiukkasten pitää pysyä. Lattiapolygoneja on vain yksi. Vastaavasti estepolygonien joukko määrittää lattiapolygonien sisällä alueet, joiden sisälle hiukkassuodinalgoritmin hiukkaset eivät voi siirtyä ja joiden läpi tagi ei voi kulkea. Käytännössä estepolygonit kuvaavat tilan seiniä sekä tiedossa olevia esteitä, kuten hyllyjä ja pöytää.

Tarkastellaan siis ensin jokaisen $i = 1, \dots, N$ hiukkasen kohdalla, siirtääkö dynaaminen malli hiukkasen polygonin F ulkopuolelle tai jonkin polygoneista E sisäpuolelle ja asetetaan näitä hiukkasia vastaava paino nollaan.

$$w_{k+1}^i = \begin{cases} w_{k+1}^i & \text{jos } x_{k+1}^i \in F, \\ 0 & \text{jos } x_{k+1}^i \notin F \vee x_{k+1}^i \in E, \end{cases} \quad (4.22)$$

jonka jälkeen tarkastellaan jokaisen jäljellä olevan ($w_{k+1}^i \neq 0$) mallin siirtämän hiukkasen polku $x_{k+1}^i - x_k^i$. Jos tämä polku ylittää yhden tai useamman estepolygonin E asetetaan kyseiselle hiukkaselle rangaistus seuraavasti:

$$w_{k+1}^i = \begin{cases} \frac{1}{P} \times w_{k+1}^i & \text{jos } x_{k+1}^i \text{ ylittää estepolygonin,} \\ w_{k+1}^i & \text{muulloin} \end{cases}, \quad (4.23)$$

jossa rangaistus P on algoritmin suunnitteluparametri. Vastaavaa toteutusta ovat hyödyntäneet muun muassa Davidson & al (2010) [9], jotka ehdottavat rangaistusarvoa $P = \frac{1}{1000}$. Rangaistuksen valintaan palataan parametrien valinnan yhteydessä alaluvussa 4.5.2.

Koska erityisesti 4.22 asettaa hiukkasten painoja nollaan, eivät karttasoviteet hiukkaset välttämättä enää estimoi suodinkakaumaa tehokkaasti. Bojja & al [2]

ehdottavat suoritettavaksi uudelleenotantaa karttasovituksen jälkeen. Uudellenotanta suoritetaan seuraavasti. Merkitään nollapainoisten hiukkasten lukumäärää N_0 . Nyt otetaan uudet N_0 otosta palauttaen joukosta $\{x_{1:k}^i\}_{i=1}^N$, missä otoksen i todennäköisyys on $w_{k|k}^i$. Korvataan nollapainoiset hiukkaset näin saadulla otoksella.

4.4.5 Siloittelumalli

Koska haluttu sisätilapaikannusalgoritmi on online-algoritmi ja käytössä olevat laskennalliset resurssit ovat rajallisia, käytetään siloitteluun algoritmissa 7 esitettyä prediktivistä siloitinta. Alaluvussa 4.5 esitettävässä empiirisessä esimerkissä on tästä huolimatta kaikki data on algoritmin saatavilla, ei viivettä tarvitse lisätä algoritmiin. Siloittelu saavutetaan yksinkertaisesti päivittämällä aika-askeleella $k < T$ painot seuraavan lauseen mukaan

$$w_k = w_k^i \bar{w}_{k+1}^i, \quad \text{missä } \bar{w}_{k+1}^i = \frac{\sum_{j=1}^N w_k^j p(x_{k+1}^i | x_k^j)}{q(x_{k+1}^i | x_k^i, y_{k+1})} \quad (4.24)$$

Partikkelen siirtämiseen käytetty satunnaiskulkumalli ei ole myöskään siloittelualgoritmien kannalta optimaalinen, joten empiirisessä esimerkissä alla hyödynnetään koetarkoitussessa vain tätä yksinkertaista siloittelumallia.

4.4.6 WB-sisätilapaikannusalgoritmi

Alla esitetään luvun 3 SIR-algoritmiin (3) ja yllä esitettyihin algoritmeihin perustuva sisätilapaikannusalgoritmi kokonaisuudessaan. Algoritmin priorijakaumana p_{x_0} käytetään kahta toisistaan riippumatonta otosta tasajakaumista, joista toinen vastaa leveys- ja toinen pituusasteita. Jakaumien alkupisteet valitaan niin, että ne vastasivat pienimpiä paikantimien leveys- ja pituusasteista. Vastaavasti päätepisteet valitaan niin, että ne vastasivat suurimpia paikantimien leveys- ja pituusasteita.

$$p_{x_{0_{\text{lon}}}} \sim \mathcal{U}(\min \lambda, \max \lambda), \quad (4.25)$$

$$p_{x_{0_{\text{lat}}}} \sim \mathcal{U}(\min \phi, \max \phi). \quad (4.26)$$

Koska järjestelmän on tarkoitus toimia ainoastaan paikantimien muodostaman suorakaiteen sisäpuolella, ovat valitut jakaumien päätepisteet riittävät. Kummastakin jakaumasta otetaan \sqrt{N} otosta, jolloin N partikkelia x_0^i saatiin näiden otosten permuutatioina. Alaluvussa 2.1.1.2 määritellyn signaali-kohinasuhteen 2.9 voi todeta olevan hyvin matala, sillä koe-esimerkissä dynaaminen malli on pelkkää kohinaa. Tämän perusteella WB-sisätilapaikannusalgoritmi käyttää ehdotusjakaumanaan tilavektorin ehdollista prioria $p(x_k | x_{k-1})$. Hiukkasten määrän N sekä muiden suunnitteluparametrien valinnasta kts. alaluku 4.5.2 alla.

Yllä esitetyn algoritmin suoritusnopeus on perusmuodossaan luokkaa $\mathcal{O}(N)$. Varianssin estimoinnissa käytetään ALvar-varianssia (kts. lauseet 2.24 ja \ref{ALvar-lambda}). Taulukossa 4.5 on esitetty algoritmin suunnitteluparametrit, joiden valintaan palataan empiirisessä esimerkissä alla.

Taulukko 4.5: WB-sisätilapaikannusalgoritmin suunnitteluparametrit

Parametri	Selitys
q	Dynaamisen mallin kohina-arvo
map_matching	Käytetäänkö karttasovitusalgoritmia, T/F
N	Hiukkasten määrä, kokonaisluku
P	Jos karttasovitusalgoritmi on käytössä, rangaistusparametri P , liukuluku
resampling	Adaptiivisen uudelleenotannat kynnysarvo, liukuluku välillä [0, 1]. Jos 0, uudelleenotantaa ei käytetä.
rssi_threshold	Datan valinnassa käytettävä signaalin vahvuuden kynnysarvo, kokonaisluku
smoothing	Käytetäänkö prediktivistä siloitinta, T/F

Algoritmi 11: WB-sisätilapaikannusalgoritmi

Result: Tagin sijaintiestimaatti \hat{x} kullekin aika-askeleelle $k = 1, \dots, T$;

Data: Taulukossa 4.4 esitetty data y_k kullekin aika-askeleelle $k = 1, \dots, T$
alaluvussa 4.4.4.2 esitetty polygonimetadata;

```
begin
    begin
        Luodaan priorijakauma  $p_{x_0}$  otantana jakaumista 4.25, asetetaan painot
         $w_0 = 1/N.$ ;
    for  $k = \{1, \dots, T\}$  do
        begin
            Ajetaan paikallaanolon havaitsemisalgoritmi 10. Jos havaitaan
            paikallaanolo, päivitetään kohina-arvo  $q_k = \frac{q}{10}$ , muussa tapauksessa
             $q_k = q$ ;
        begin
            Sovitetaan partikkeleihin dynaaminen malli 4.18  $x_{k+1} = x_k + v_k$ ,
            missä  $v_k$  on luotu alaluvussa 4.4.4 esitettyllä menetelmällä sekä
            kohina-arvolla  $q_k$ ;
        begin
            Jos karttasovitusalgoritmi on käytössä, päivitetään painot  $w_k$ 
            alaluvun 4.4.4.2 karttasovitusalgoritmilla ja rangaistusarvolla  $P$ ;
        for  $i = \{1, 2, \dots, N\}$  do
            begin
                Päivitetään painot  $w_{k|k}$  alaluvun 4.4.2 uskottavuusmalleilla.
            begin
                Estimoidaan  $p$  laskemalla tiheydelle approksimaatio
                 $\hat{p}(x_{1:k}|y_{1:k}) = \sum_{i=1}^N w_{k|k}^i \delta(x_{1:k} - x_{1:k}^i)$ .
            if  $k < T$  then
                begin
                    Päivitetään paino  $w_k^i$  prediktioivisellä siloittimella  $w_k^i = w_k^i \bar{w}_{k+1}^i$ .
            begin
                Luodaan sijaintiestimaatti  $\hat{x}_k = \sum_{i=1}^N w_i^k x_k^i$ . Estimaatti lasketaan
                erikseen pituus- ja leveyskoordinaateille.
            begin
                Lasketaan ALvar-varianssi  $\hat{\sigma}_{\text{ALvar}_k}^2$ .
            begin
                Lasketaan efektiivinen otoskoko  $\hat{N}_{eff}$ .
            if  $\hat{N}_{eff} < N_{th}$  then
                begin
                    Otetaan uudet  $N$  otosta palauttaen joukosta  $\{x_{1:k}^i\}_{i=1}^N$ , missä
                    otoksen  $i$  todennäköisyys on  $w_{k|k}^i$ .
            begin
                Asetetaan painot  $w_{k|k}^i = 1/N$ .
            begin
                Päivitetään Henok-indeksit  $E$  uudelleenotannan perusteella.
```

Koeasetelmassa käytetty algoritmin [11](#) toteutus on ohjelmoitu R-kielellä. Algoritmin toteutus on pääosin vektorisoitu ja tehokas. For-silmukkaa on käytetty ainoastaan aika-askeleiden läpikäyntiin. Koska tämän silmukan muuntaminen vektorisoituuun muotoon ei ole mahdollista, voidaan toteutusta pitää näiltä osin hyvin optimoituna. Kaikki algoritmin datan käsittely on toteutettu suorituskyvyltään erinomaisella `data.table`-kirjastolla.

Koska algoritmin tuottamat uskottavuudet sekä painot ovat pieniä, on laskentatarkkuusongelmien välttämiseksi toteutuksessa käytetty logaritmoituja painoja ja uskottavuusfunktiota. Tämä ei vaikuta itse algoritmin toimintaan, mutta estää numeroisten ongelmien syntymisen.

Paikannusvirheen laskemisessa on etäisyysfunktiona käytetty `raster`-kirjaston `pointDistance()`-funktiota. Algoritmissa [9](#) kuvattu testipolun tihennyt on tehty ennen algoritmin ajoa ja tihennettyä testipolkua käytetään algoritmin syötteenä. Tihennys on tehty `smoothr`-kirjaston `smooth_densify()`-funktiolla. Koodissa on koodistettu tiiviyden sijaan luettavuutta ja koodi on kommentoitu kattavasti. Algoritmin R-koodi löytyy osoitteesta https://github.com/rintakumpu/progradu/analyysi/R/pf_positioning.R.

4.5 Empiirinen esimerkki

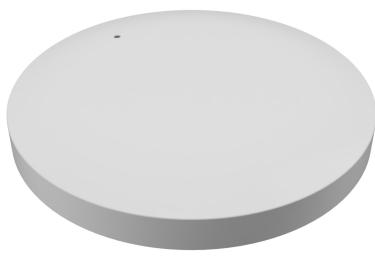
4.5.1 Koeasetelma

Esimerkissä käytetään hiukkasalgoritmia Bluetooth-paikannusovelluksessa lähettimen sijainnin laskemiseen. Paikannukseen käytettävä data kerätään ruokakauppaympäristössä Bluetooth Low Energy (BLE) -lähettimen sekä kattoon sijoitettujen vastaanottimien avulla. Havainnot koostuvat vastaanottimien lähettimien signaalien perusteella laskemista, BLE5.1-standardin mukaisista signaalil tulokulmista eli AoA-havainnoista (*angle of arrival*). Esimerkissä analysoidaan ja vertaillaan algoritmin eri versioiden suorituskykyä sekä suorituskyvyn että paikannustarkkuuden näkökulmasta. Vertailuarvona käytetään perinteistä triangulaatio-algoritmia.

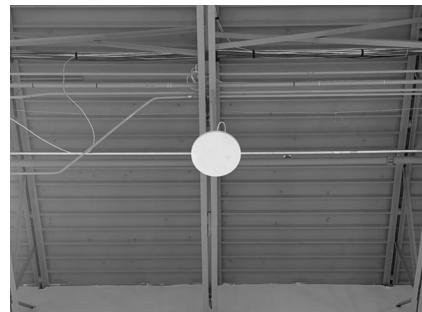
Paikaunnusesimerkissä lähettimenä toimi 6hz taajuudella havaintoja lähetävä AT-2 paikannustagi (kuva [4.3](#)), vastaanottimena testiympäristöön asennetut $n_s = 33$ Walkbase XR-2 -vastaanotinta (kuvat [4.4](#) ja [4.5](#)). Jokainen vastaanotin sisältää kuusitoista antennia, joiden vastaanottamien lähetinsignaalien perusteella vastaanottimet laskevat signaalil tulokulman suhteessa vastaanottimeen. Havaintoja on datassa yhteensä $n_{obs} = 12379$ kappaletta ja ne kattavat $T = 432$ sekuntia. Havaintojen aikaleimat on tallennettu sekunnin tuhannesosan tarkkuudella.



Kuva 4.3: Walkbase AT-2



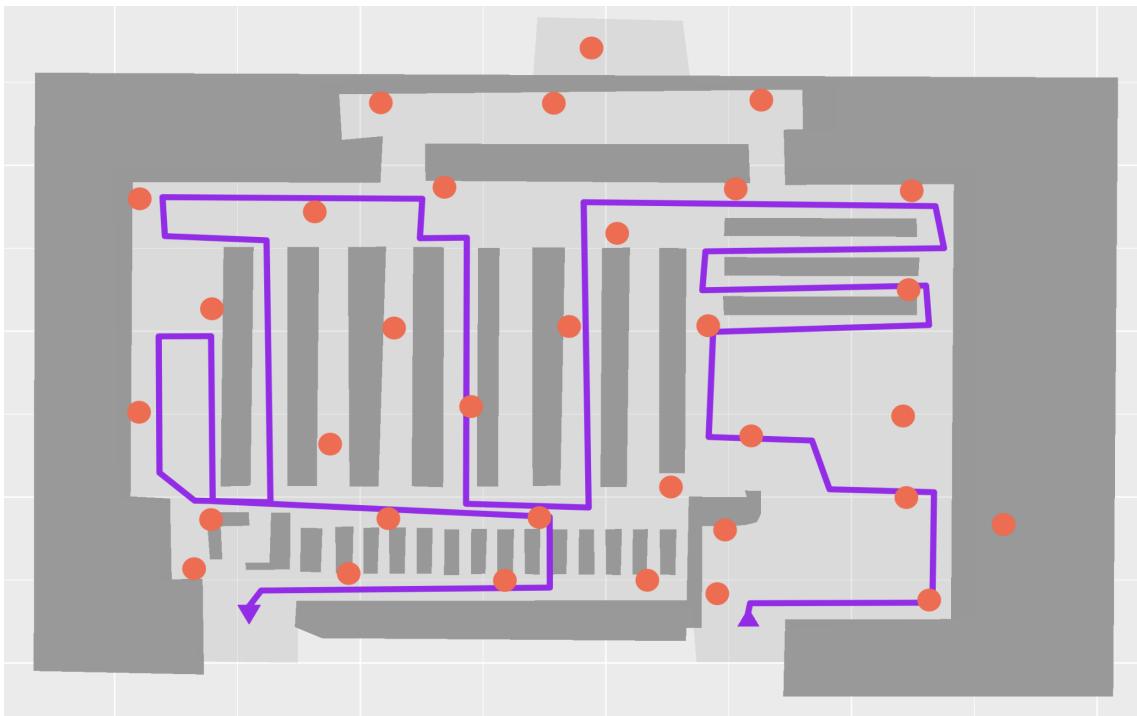
Kuva 4.4: Walkbase XR-2



Kuva 4.5: Walkbase XR-2 asennettuna

Koejärven pohjapiirustus on esitetty kuvassa 4.6. Piirustuksessa XR-2-paikantimet on kuvattu punaisilla ympyröillä ja kuljettu testipolku on merkitty violetilla janalla. Piirustus on luotu käyttäen `ggplot2`-kirjastoa ja se löytyy RDS-muotoon tallennettuna osoitteesta

<https://github.com/rintakumpu/progradu/analyysi/R/data/sitemap.RDS>.



Kuva 4.6: Koeasetelman pohjapiirustus

Tagi kiinnitettiin ostoskärryyn 1.2 metrin korkeudelle (kts. kuvat [4.7](#) ja [4.8](#)) ja testipolku käveltiin mahdollisimman tasaisella nopeudella. Data kerättiin aikaan, jolloin testiympäristön käyttöaste oli alhainen. Tällä minimoitiin radiosignaalien tielle osuvien ihmisten vaikutus signaaleihin. Kerätty tulokulmadata löytyy osoitteesta <https://github.com/rintakumpu/progradu/analyysi/data/y.csv>. Pohjapiirustus- ja polkudata löytyy osoitteista https://github.com/rintakumpu/progradu/analyysi/data/exclusion_polygons.csv (ekskluusiopolygonit), https://github.com/rintakumpu/progradu/analyysi/data/inclusion_polygons.csv (lattiapolygoni) ja https://github.com/rintakumpu/progradu/analyysi/data/test_path.csv (testipolku). Datassa koordinaatit on valmiiksi interpoloitu metreiksi, jotta testiympäristön sijaintia ei voi paikallistaa koordinaattien perusteella. Interpolointiin käytetty ohjel makoodi löytyy osoitteesta https://github.com/rintakumpu/progradu/analyysi/R/interpolate_coordinates.R. Osoitteesta <https://github.com/rintakumpu/progradu/analyysi/analyysi.Rmd> löytyy itse analyysikoodin sisältävä R Markdown -muistikirja.





Kuva 4.8: Walkbase AT-2 kiinnitettyynä

4.5.2 Parametrien valinta

Kokeessa oli tarkoituksena testata kunkin taulukossa 4.5 esitetyn parametrin vaiutusta paikannusvirheeseen, ajoaan sekä varianssiin. Koska kaikkien parametrikombinaatioiden testaaminen ei ollut mahdollista (eikä mielekästä), suoritettiin yllä kerättyyn dataan perustuva paikannus kolmessa vaiheessa. Kussakin vaiheessa paikannusalgoritmi ajettiin jokaisella vaiheeseen liittyvällä suunnitteluparametrikombinaatiolla $r = 30$ kertaa ja tulokset laskettiin näiden 30 ajon aritmeettisena keskiarvona.

Ensimmäisessä vaiheessa tarkasteltiin partikkeli määrä $N = 100, 1000, 10000$ sekä uudelleenotannan kynnysarvon $resampling = 0, 1/10, 2/3, 1$ vaikutusta paikannuskeskivirheeseen. Karttasovitusalgoritmia ei käytetty, kuten ei myöskään prediktivistä siloitinta eikä signaalin vahvuuden kynnysarvoa. Dynaamisen mallin kohina-arvo $q = 2$ pidettiin vakiona. Ensimmäisessä vaiheessa tarkasteltiin siis 12 eri suunnitteluparametrikombinaatiota.

Toisessa vaiheessa valittiin edellisen vaiheen tulosten perusteella parhaimman paikannusvirheen suhteessa suoritusaikaan ja varianssiin tuottava uudelleenotannan kynnysarvo $resampling$ sekä partikkeli määriä N . Nämä pidettiin vakioarvoisina ja testattiin karttasovitusalgoritmia $map_matching = T, F$, karttasovitusalgoritmin rangaistusarvoa $P = 1, 100, 1000$ sekä dynaamisen mallin kohina-arvoa $q = 0.75, 1.5, 2, 2.5$. Signaalin vahvuuden kynnysarvoa ei käytetty, kuten ei käytetty myöskään prediktivistä siloitinta. Koska rangaistusarvo P oli käytössä ainoastaan karttasovitusalgoritmia käytettäessä, tarkasteltiin toisessa vaiheessa 16 eri suunnitteluparametrikombinaatiota.

Viimeisessä vaiheessa valittiin edellisten vaihdeiden tulosten perusteella parhaimman paikannusvirheen suhteessa suoritusaikaan ja varianssiin tuottavat parametrit testattujen joukosta ja testattiin datan valinnassa käytettävä signaalin vahvuuden kynnysarvoa $rssi_threshold = -100, -90, -80$ sekä prediktivistä siloitinta $smoothing = T, F$ eli kuutta eri suunnitteluparametrikombinaatiota.

Tulosten vertailukohtana käytettiin Pierlot & al. artikkelissa “A New Three Object Triangulation Algorithm Based on the Power Center of Three Circles” (2011) esittämää ToTal-triangulaatioalgoritmia. [27] Triangulaatio-algoritmia ei käsitellä tässä tarkemmin, mutta se on esitetty algoritmissa 12. Algoritmia varten valittiin kunakin aika-askeleella k ne kolme paikanninta ja kulmahavaintoa, joiden RSSI-arvo

oli korkein. Käytetty algoritmin toteutus löytyy osoitteesta <https://github.com/rintakumpu/progradu/analyysi/R/total.R>.

Paikannusalgoritmi ajettiin RStudion versiossa 2023.09.0+463 R-ohjelmointikielen versiolla 4.2.0. Tietokoneena käytettiin vuoden 2021 mallia olevaa MacBook Pro -kannettavaa, jossa oli Apple M1 Pro -prosessori sekä 32 gigatavua LPDDR5-muistia. Suoritusnopeuden mittaanmiseen käytettiin `Sys.time()`-funktiota.

Algoritmi 12: ToTal (Three object Triangulation algorithm)

Result: Testilaitteen sijaintiestimaatti (x_R, y_R) .

Data: Kolmen paikantimen koordinaatit (x_i, y_i) , $i = \{1, 2, 3\}$ ja näitä vastaavat vastakkaiset kulmahavainnot $\Phi'_1, \Phi'_2, \Phi'_3$.

begin

 Lasketaan muokatut koordinaatit

$$x'_1 = x_1 - x_2, \quad y'_1 = y_1 - y_2, \quad x'_3 = x_3 - x_2, \quad y'_3 = y_3 - y_2.$$

begin

 Lasketaan kotangentit

$$T_{12} = \cot(\Phi'_2 - \Phi'_1), \quad T_{23} = \cot(\Phi'_3 - \Phi'_2), \quad T_{31} = \frac{1 - T_{12}T_{23}}{T_{12} + T_{23}}.$$

begin

 Lasketaan muokatut ympyröiden keskipisteet (x'_{ij}, y'_{ij})

$$x'_{12} = x'_1 + T_{12}y'_1, \quad y'_{12} = y'_1 - T_{12}x'_1$$

$$x'_{23} = x'_3 - T_{23}y'_3, \quad y'_{23} = y'_3 + T_{23}x'_3$$

$$x'_{31} = (x'_3 + x'_1) + T_{31}(y'_3 - y'_1), \quad y'_{31} = (y'_3 + y'_1) - T_{31}(x'_3 - x'_1).$$

begin

$$\quad \text{Lasketaan } k'_{31} = x'_1x'_3 + y'_1y'_3 + T_{31}(x'_1y'_3 - x'_3y'_1).$$

begin

 Lasketaan nimittääjä D (jos $D = 0$ palautetaan virhe).

$$\quad D = (x'_{12} - x'_{23})(y'_{23} - y'_{31}) - (y'_{12} - y'_{23})(x'_{23} - x'_{31}).$$

begin

 Lasketaan ja palautetaan sijaintiestimaatti (x_R, y_R) .

$$\quad x_R = x_2 + \frac{k'_{31}(y'_{12} - y'_{23})}{D} \quad y_R = y_2 + \frac{k'_{31}(x'_{23} - x'_{12})}{D}.$$

4.5.3 Tulokset

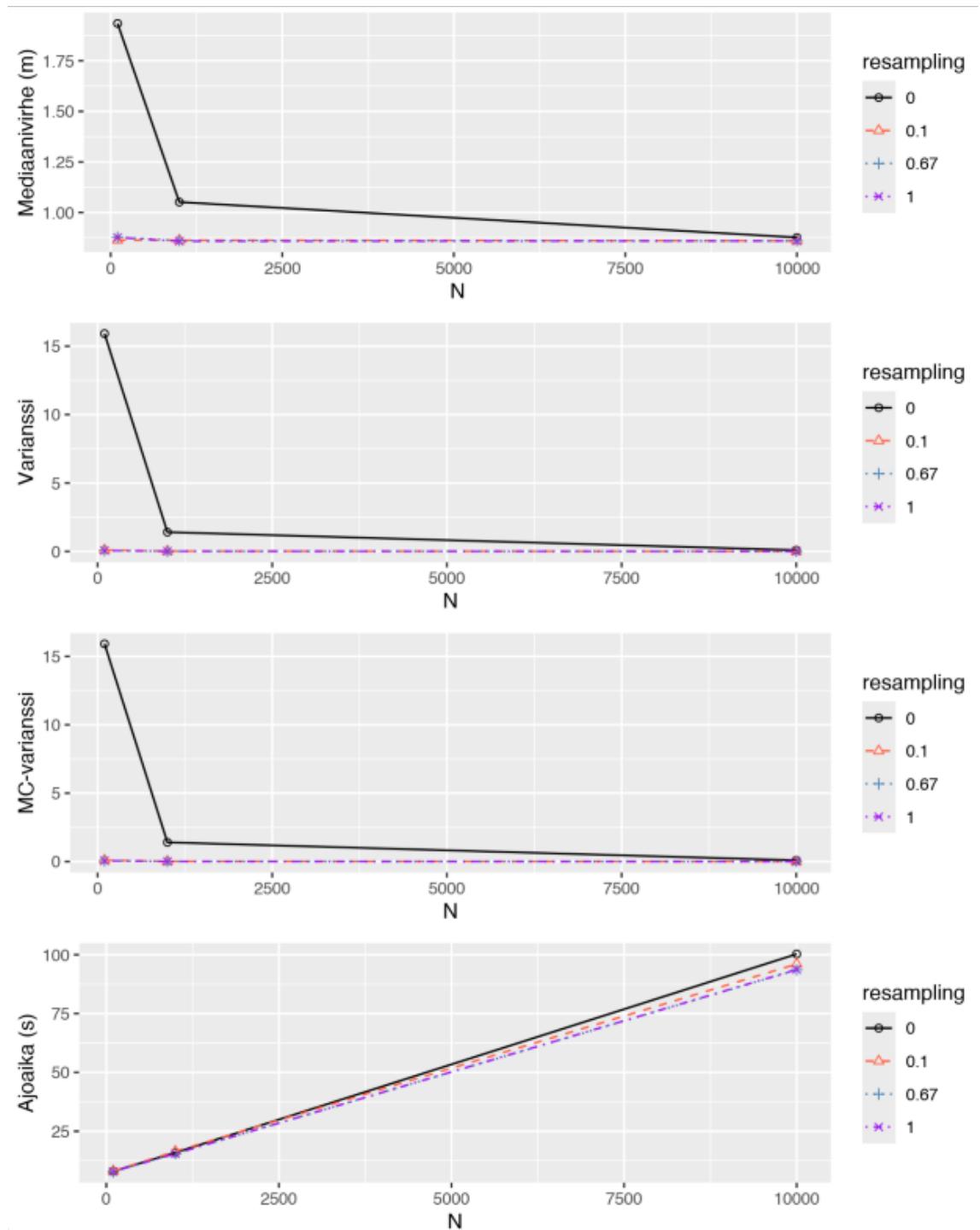
Ensimmäisessä vaiheessa suoritettiin paikannus partikkeliin määräällä $N = 100, 1000, 10000$ sekä uudelleenotannan kynnysarvolla $resampling = 0, 1/10, 2/3, 1$. Kun kynnysarvo oli 0, uudelleenotantaa ei käytetty, jolloin SIR-algoritmin sijaan paikannus suoritettiin SIS-algoritmilla. Kun kynnysarvo oli 1, uudelleenotanta suoritettiin jokaisella aika-askeleella. Arvoilla 1/10 ja 2/3 käytettiin adaptiivista uudelleenotantaa. Tulokset on esitetty kuvassa 4.9 sekä taulukoissa 4.6 ja 4.7. Ajojen tulokset on esitetty karttapolkuina liitteen A alaluvussa 5.

Taulukko 4.6: Vaiheen 1 tulokset, paikannusvirhe

N	resampling	Mediaani (m)	Mediaanin 95%-luottamusväli	<1m
100	0.00	1.93	[1.86, 2.01]	0.28
1000	0.00	1.05	[1.03, 1.07]	0.48
10000	0.00	0.88	[0.87, 0.88]	0.55
100	0.10	0.86	[0.86, 0.87]	0.57
1000	0.10	0.86	[0.86, 0.87]	0.57
10000	0.10	0.86	[0.86, 0.86]	0.58
100	0.67	0.88	[0.87, 0.89]	0.56
1000	0.67	0.86	[0.86, 0.87]	0.58
10000	0.67	0.86	[0.86, 0.86]	0.58
100	1.00	0.88	[0.87, 0.89]	0.56
1000	1.00	0.86	[0.86, 0.86]	0.58
10000	1.00	0.86	[0.86, 0.86]	0.58

Taulukko 4.7: Vaiheen 1 tulokset, varianssi ja ajoaika

N	resampling	Varianssi	MC-varianssi	Ajoaika (s)
100	0.00	NA	15.92	7.99
1000	0.00	NA	1.41	15.92
10000	0.00	NA	0.09	1.67
100	0.10	3.28	0.10	8.11
1000	0.10	1.36	0.02	16.45
10000	0.10	0.62	0.00	1.60
100	0.67	1.09	0.06	7.79
1000	0.67	0.37	0.01	15.49
10000	0.67	0.26	0.00	1.56
100	1.00	1.10	0.06	7.89
1000	1.00	0.38	0.01	15.43
10000	1.00	0.26	0.00	1.56



Kuva 4.9: Vaiheen 1 tulokset

Ensimmäisen vaiheen tuloksista huomataan, ettei hiukkasten määrällä ole juuri-kan vaikutusta mediaanipaikannusvirheeseen, kun uudelleenotanta on käytössä (so. käytämme SIR-algoritmia). Samoin adaptiivisen uudelleenotannan kynnysarvolla ei ole juurikaan vaikutusta mediaanipaikannusvirheeseen tai ajoaikaan.

Se, ettei partikkelien määän kasvattaminen automaattisesti paranna paikan-nustarkkuutta viittaa siihen, että koeasetelma on herkkä priorijakauman valinnalle. Tuloksista huomataan lisäksi, että algoritmin aikakompleksisuus on uudelleenotan-nasta riipumatta luokkaa $\mathcal{O}(N)$, kuten tukielman teoriaosassa todettiin. Samoin hiukkasten määän kasvattaminen pienentää variansseja, kuten teoriaosassa todettiin. MC-varianssia pienentää myös uudelleenotannan käyttäminen.

Koska MC-varianssi on laskettu itse sijaintiestimaateista ja ALvar-varianssi kaikista partikkeleista, on näiden kahden varianssin suuruusluokka eri. Huomataan kuitenkin, että nämä kaksi varianstestimittauksia ovat hyvin korreloituneita, korrelatio-kertoimilla $\rho_{\text{Pearson}} \approx 0.86$ ja $\rho_{\text{Spearman}} \approx 0.88$. Voidaan siis olettaa ALvar-varianssin estimoivan hyvin hiukkassuotimen todellista varianssia.

Ensimmäisen vaiheen tulosten perusteella valitaan seuraavaan vaiheeseen ne N -ja *resampling*-parametriarvot, jotka tuottavat parhaimman mediaanipaikannusvir-heen. Jos kahden eri paikannusvirheen 95%-luottamusvälit ovat päällekkäiset, valitaan arvoista ensin se, joka tuottaa paremman ALvar-varianssin. Näin päädytään pa-rametriyhdistelmään $N = 10000$ ja *resampling* = 2/3. Koska algoritmin ajoaika 10000 partikkelilla on kuitenkin epäkäytännöllinen, valitaan $N = 1000$, vaikka tämä lisääkin hieman estimaatin varianssia.

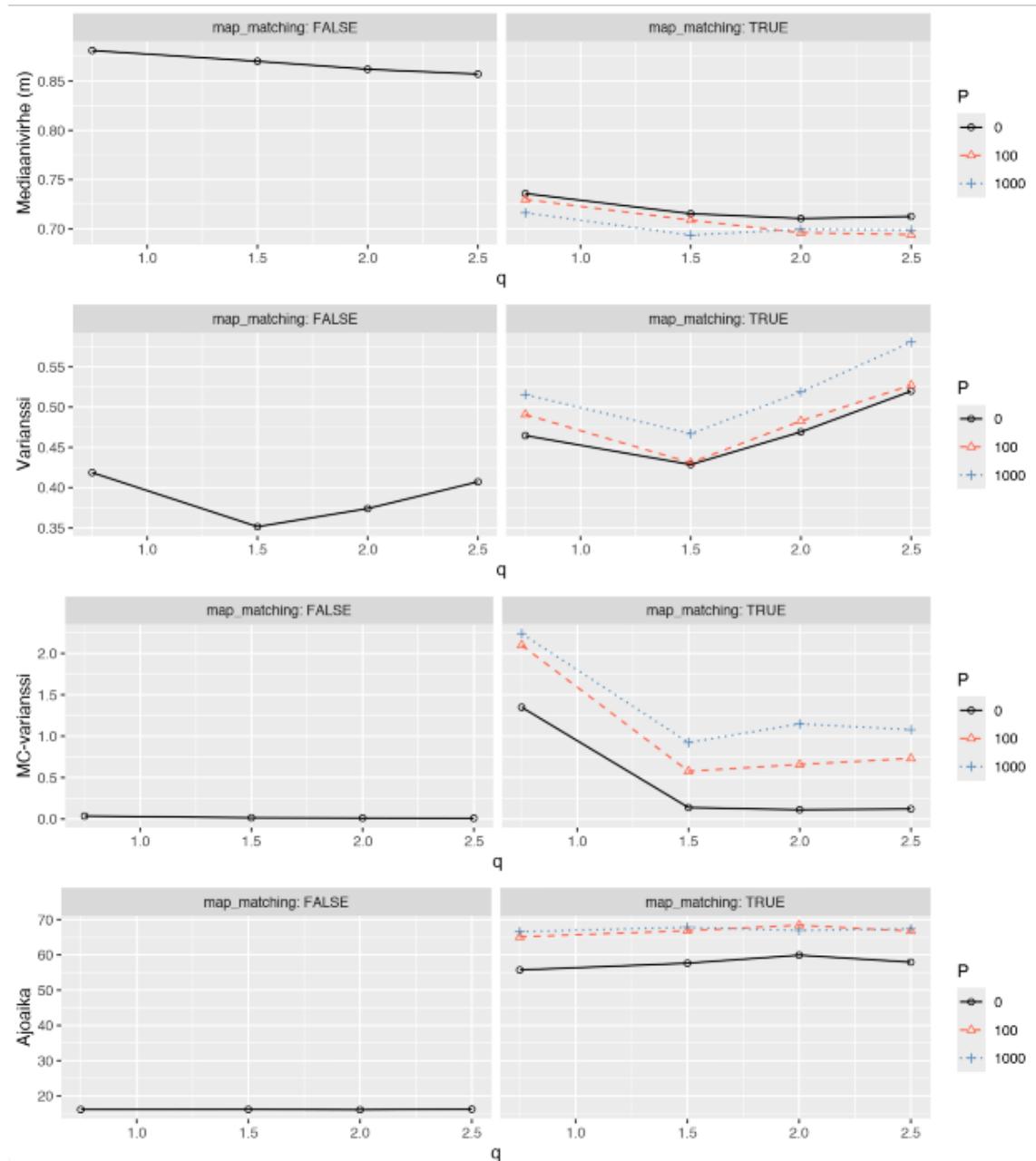
Toisessa vaiheessa $N = 1000$ ja *resampling*=2/3 pidettiin vakioarvoisina ja testattiin karttasovitusalgoritmia *map_matching* = T, F , karttasovitusalgoritmin rangais-tusarvoa $P = 1, 100, 1000$ sekä dynaamisen mallin kohina-arvoa $q = 0.75, 1.5, 2, 2.5$. Signaalin vahvuuden kynnysarvoa ei käytetty, kuten ei käytetty myöskään prediktiivis-tä siloitinta. Koska rangaistusarvo P oli käytössä ainoastaan karttasovitusalgoritmia käytettäessä, tarkasteltiin toisessa vaiheessa 16 eri suunnitteluparametrikombinaatio-ta. Tulokset on esitetty kuvassa 4.10 sekä taulukoissa 4.8 ja 4.9. Ajojen tulokset on esitetty karttapolkuina liitteen A alaluvussa 5.

Taulukko 4.8: Vaiheen 2 tulokset, paikannusvirhe

map_matching	P	q	Mediaani (m)	Mediaanin 95%-luottamusväli	<1m
TRUE	0	0.75	0.74	[0.73, 0.74]	0.66
TRUE	100	0.75	0.73	[0.72, 0.74]	0.66
TRUE	1000	0.75	0.72	[0.71, 0.72]	0.67
TRUE	0	1.50	0.72	[0.71, 0.72]	0.68
TRUE	100	1.50	0.71	[0.7, 0.72]	0.68
TRUE	1000	1.50	0.69	[0.69, 0.7]	0.68
TRUE	0	2.00	0.71	[0.7, 0.72]	0.68
TRUE	100	2.00	0.70	[0.69, 0.7]	0.68
TRUE	1000	2.00	0.70	[0.69, 0.71]	0.68
TRUE	0	2.50	0.71	[0.71, 0.72]	0.68
TRUE	100	2.50	0.69	[0.69, 0.7]	0.68
TRUE	1000	2.50	0.70	[0.69, 0.71]	0.67
FALSE	0	0.75	0.88	[0.88, 0.89]	0.57
FALSE	0	1.50	0.87	[0.87, 0.87]	0.57
FALSE	0	2.00	0.86	[0.86, 0.87]	0.58
FALSE	0	2.50	0.86	[0.85, 0.86]	0.57

Taulukko 4.9: Vaiheen 2 tulokset, varianssi ja ajoaika

map_matching	P	q	Varianssi	MC-varianssi	Ajoaika (s)
TRUE	0	0.75	0.46	1.35	16.22
TRUE	100	0.75	0.49	2.10	16.22
TRUE	1000	0.75	0.52	2.24	16.22
TRUE	0	1.50	0.43	0.14	16.22
TRUE	100	1.50	0.43	0.58	16.22
TRUE	1000	1.50	0.47	0.92	16.22
TRUE	0	2.00	0.47	0.11	16.22
TRUE	100	2.00	0.48	0.66	16.22
TRUE	1000	2.00	0.52	1.15	16.22
TRUE	0	2.50	0.52	0.12	16.22
TRUE	100	2.50	0.53	0.73	16.22
TRUE	1000	2.50	0.58	1.08	16.22
FALSE	0	0.75	0.42	0.04	16.22
FALSE	0	1.50	0.35	0.01	16.22
FALSE	0	2.00	0.37	0.01	16.22
FALSE	0	2.50	0.41	0.01	16.22



Kuva 4.10: Vaiheen 2 tulokset

Taulukko 4.10: Vaiheen 3 tulokset, paikannusvirhe

rssi_threshold	smoothing	Mediaani (m)	Mediaanin 95%-luottamusväli	<1m
-100	TRUE	0.72	[0.72, 0.73]	0.65
-90	TRUE	0.73	[0.72, 0.74]	0.65
-80	TRUE	0.70	[0.69, 0.71]	0.62
-100	FALSE	0.69	[0.69, 0.7]	0.68
-90	FALSE	0.70	[0.7, 0.71]	0.68
-80	FALSE	0.70	[0.7, 0.71]	0.65

Toisen vaiheen tuloksista huomataan, että karttasovituksen käyttäminen parantaa paikannusvirhettä. Syy tähän on helppo havaita liitteenä olevista karttapoluista. Paikannus ottaa nyt huomioon sisätilaympäristön, eikä enää luo estimaatteja sijainteihin, joihin tagin on fyysisesti mahdotonta päästä. Samoin rangaistusarvo P :n lisääminen parantaa paikannusvirhettä. Tämä on odotettua, sillä isomman rangaistusarvon käyttäminen ei ainoastaan estää fyysisesti mahdottomia sijainteja vaan estää myös fyysisesti mahdottomat siirtymät kahden peräkkäisen sijaintiestimaatin välillä.

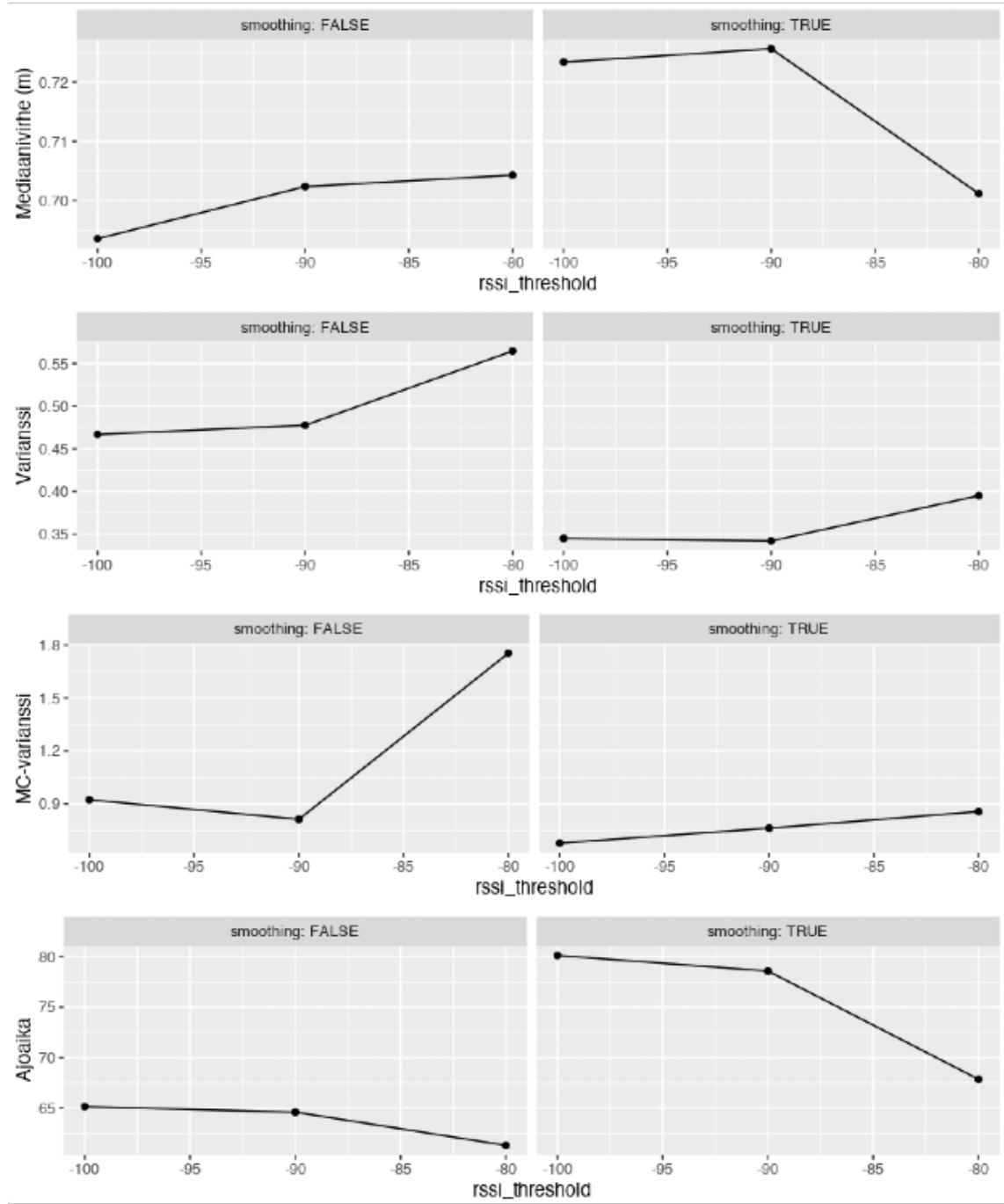
Vastaavasti liikemallin kohina-arvon q pienentäminen parantaa paikannusvirhettä. Liian pienellä q -arvolla algoritmi ei kuitenkaan enää tutki signaaliympäristöä tarpeeksi tehokkaasti ja paikannusvirhe kasvaa, samoin kasvaa estimaattien MC-varianssi. Optimaalinen kohina-arvo on tulosten perusteella $q = 1.5$. Tämä vastaa myös hyvin kirjallisudessa esitettyjä keskimääräisiä kävelynopeuksia (kts. esim. [16]).

Huomataan lisäksi, että parhaimman paikannusvirheen tuottava rangaistusarvo $P = 1000$ on tutkitun parametriavaruuden reunalla, joten mahdollisesti isommalla P -arvolla voitaisiin vielä parantaa paikannusvirhettä. Pienempien P -arvojen paikannusvirheiden luottamusvälit ovat kuitenkin päälekäisiä arvon $P = 1000$ kanssa, joten paikannusvirheeseen saatu lisähyöty ei todennäköisesti olisi tilastollisesti merkityksellistä. Näiden tulosten perusteella valitaan viimeiseen vaiheeseen siis kiinteät arvot $map_matching = T$, $P = 1000$ ja $q = 1.5$.

Viimeisessä vaiheessa testattiin datan valinnassa käytettävää signaalin vahvuuden kynnysarvoa $rssi_threshold = -100, -90, -80$ sekä prediktiivistä siloitinta $smoothing = T, F$ eli kuutta eri suunnitteluparametrikombinaatiota. Tulokset on esitetty kuvassa 4.11 sekä taulukoissa 4.10 ja 4.11. Ajojen tulokset on esitetty karttapolkuina liitteen A alaluvussa 5.

Taulukko 4.11: Vaiheen 3 tulokset, varianssi ja ajoaika

rssi_threshold	smoothing	Varianssi	MC-varianssi	Ajoaika (s)
-100	TRUE	0.35	0.68	61.3
-90	TRUE	0.34	0.76	61.3
-80	TRUE	0.40	0.86	61.3
-100	FALSE	0.47	0.92	61.3
-90	FALSE	0.48	0.81	61.3
-80	FALSE	0.56	1.75	61.3



Kuva 4.11: Vaiheen 3 tulokset

Taulukko 4.12: Tulosten perusteella valitut suunnitteluparametrit

Suunnitteluparametri	Arvo
N	1000
resampling	0.67
map_matching	TRUE
P	1000
q	1.5
smoothing	FALSE
rssi_threshold	-120

Taulukko 4.13: ToTaL-triangulaatioalgoritmin tuottamia paikannusvirheen tunnuslukuja

	Mediaani	Aritmeettinen keskiarvo	Otoskeskihajonta	<1m persentili
ToTaL	1.61	2.35	4.15	0.34

Tuloksista huomataan, ettei siloittelun tai signaalin vahvuuden kynnysarvon käyttäminen paranna paikannusvirhetä. Siloittelun osalta tämä on odotettua, kun liikemallina on käytetty satunnaiskulkua. Liitteenä olevista karttapoluista kuitenkin huomataan, että siloittelu tuottaa odotetusti sileämpiä polkuja, mikä saattaa olla käytännössä haluttu ominaisuus. Taulukossa 4.12 on esitetty vielä tulosten perusteella valitut suunnitteluparametrit.

Tulosten perusteella voidaan todeta, että WB-sisättilapaikannusalgoritmi tuottaa halutun paikannusvirheen. Algoritmin testaamisessa on myös käytetty todelliseen liiketilaan luotua testiasetelemaa, joka vastaa hyvin järjestelmän todellista käyttötarkoitusta. Tämän perusteella voidaan algoritmin myös olettaa yleistyttyä hyvin monimutkaisiin sisätilaympäristöihin. Taulukossa 4.13 on esitetty ToTaL-triangulaatioalgoritmilla 12 luoduista sijaintiestimaateista laskettujen paikannusvirheiden tunnuslukuja.

Triangulaatioon verrattuna WB-sisättilapaikannusalgoritmi tuottaa siis huomattavasti parempia tuloksia. Tämän voi havaita myös vertaamalla kuvan 4.12 ToTaL-karttapolkuja liitteen A WB-sisättilapaikannusalgoritmin tuottamiin karttapolkuihin.



Kuva 4.12: ToTaL-triangulaatioalgoritmin tuottamat sijaintiestimaatit

Algoritmia ja järjestelmää voitaisiin mahdollisesti edelleen parantaa esimerkiksi hyödyntämällä paremmin tagin kiihtyvyysmittarin tuottamaa dataa informatiivisen liikemallin luomisessa. Tällöin olisi myös mielekkäämpää toteuttaa varianssiestimaattiin perustuva adaptiivisen viipeen suodin osana paikannusalgoritmia. Lisäksi datan valinta voitaisiin suorittaa esimerkiksi niin, että datasta poistettaisiin kyllakin aika-askeleella ne kulmahavainnot, jotka poikkeavat kulmahavaintojen suuntakonsensuksesta.

Liitteenä olevia polkuja tarkastelemalla huomataan, että nyt toteutetun algoritmin sijaintestimaatilla on taipumus jäädä osassa testiympäristöä jälkeen itse tagin sijainnista. Tätä ongelmaa voitaisiin mahdollisesti lieventää käyttämällä esimerkiksi Yi Chenging & al artikkelissa “Improved Particle Filter Algorithm for Multi-Target Detection and Tracking” (2024) [4] esittämää menetelmää, jossa partikkeliit jaetaan ns. seurantahiukkasiin sekä etsintähiukkasiin, joista ainoastaan edellisiä käytetään sijaintestimaatin luomisessa ja jälkimmäisten annetaan liikkua suuremmilla kohinaarvoilla. Näin mahdolistetaan satunnaiskulkumallilla laajempi signaalivaruuden tutkinta ja nopeampi ongelmatilanteista toipuminen ilman, että sijaintestimaatit kärsivät liikemalliin lisätystä kohinasta.

Luku 5

Lopuksi

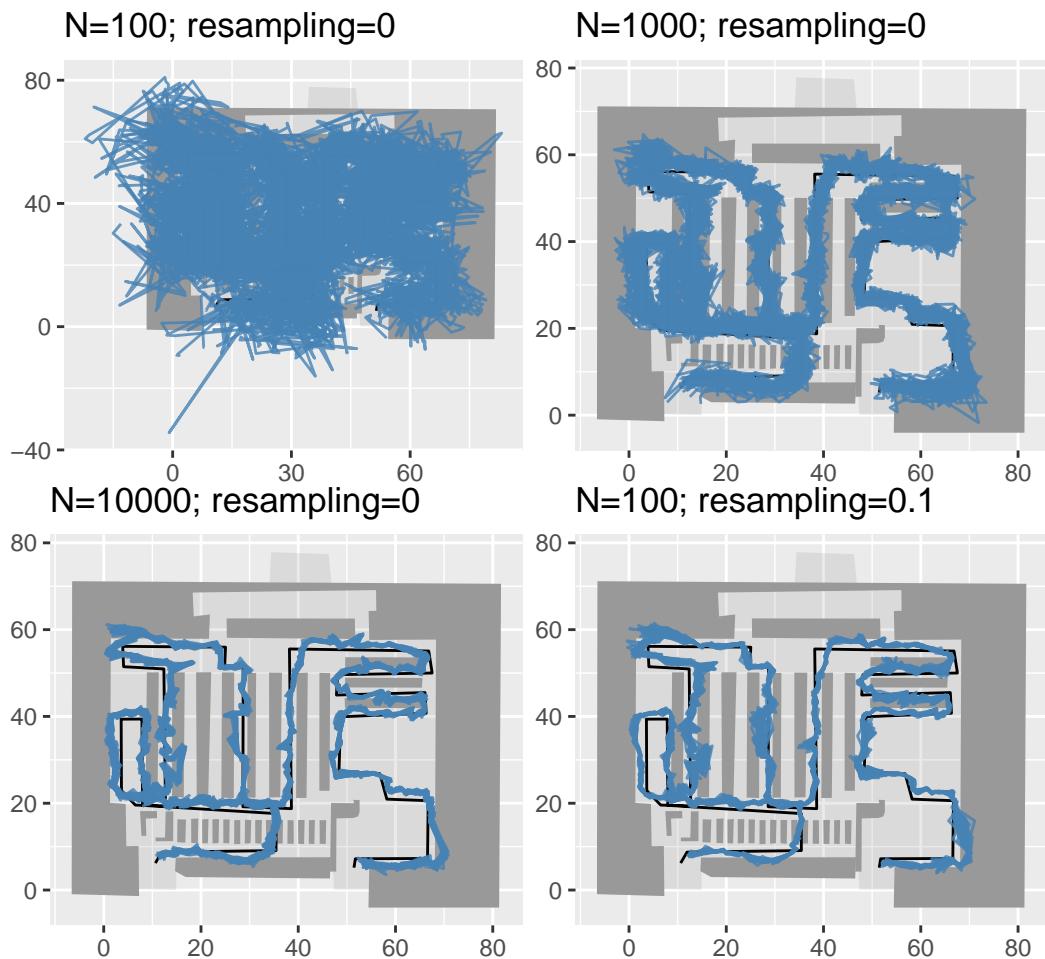
Tässä tutkielmassa on esitetty pääpiirteittäin hiukkassuodin- ja hiukkassiloitinalgoritmien teoria Bayesilaisessa tilastotieteellisessä viitekehysessä. Tutkielmassa on lisäksi käyty läpi uudelleenotantaa efektiivisen otoskoon perusteella hyödyntävä SIR-suodinalgoritmi sekä käsitetty algoritmin varianssin estimointia. Tutkielmassa on myös esitetty SIR-algoritmin parametrien valintaan, suorituskykyyn sekä konvergenssiin liittyviä tuloksia.

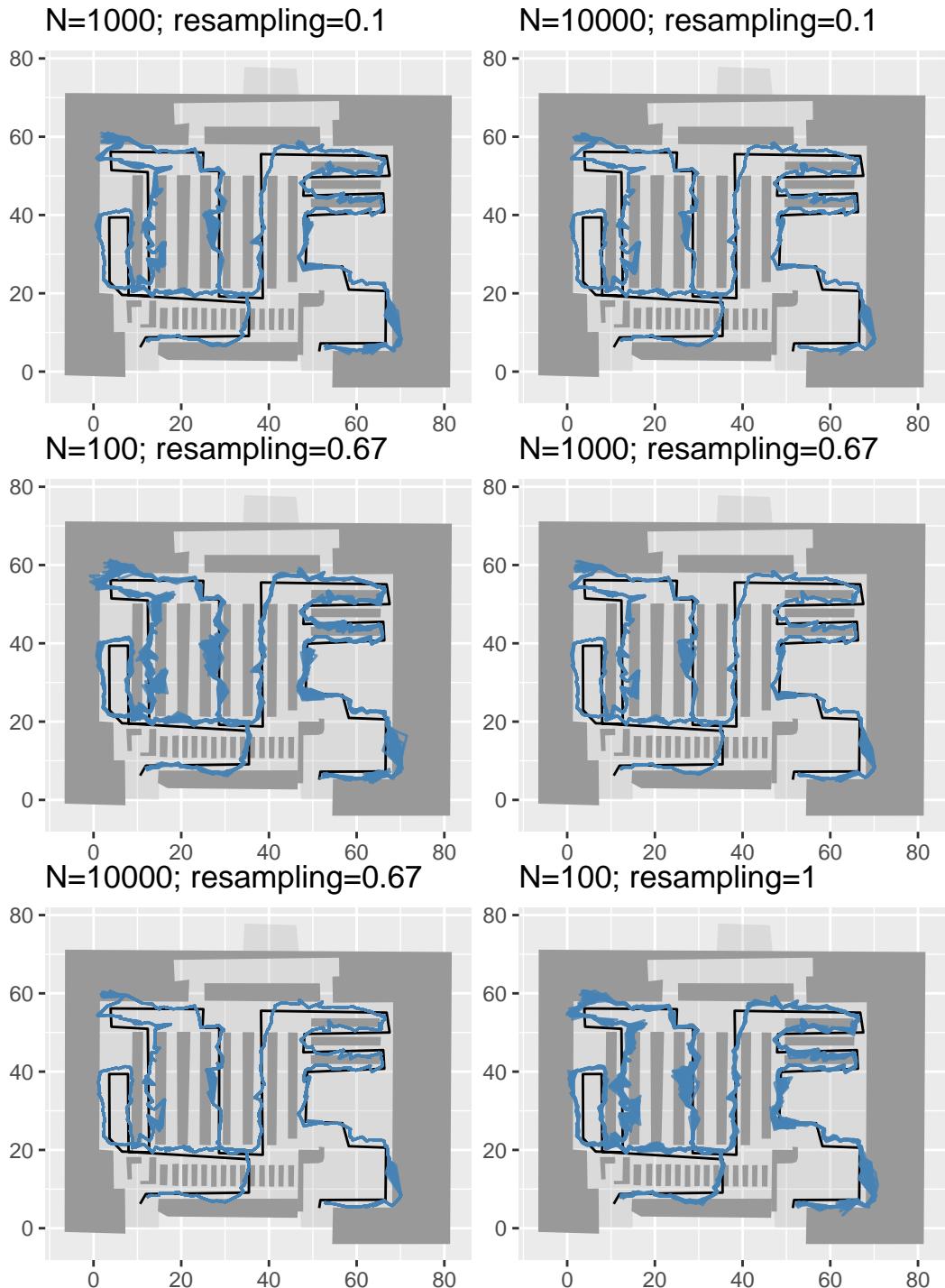
Tutkielmassa on lisäksi esitetty WB-sisätilapaikannusalgoritmi, joka toteuttaa SIR-algoritmin, varianssin estimoinnin sekä hyödyntää sisätilapaikannuksen karttasovitusalgoritmia. Tutkielmassa on lopuksi tarkasteltu miten eri suunniteluparametrien valinnat vaikuttavat tämän algoritmin suorituskykyyn kattavan ja todelliseen ongelmaan sekä dataan perustuvan paikannusesimerkin avulla.

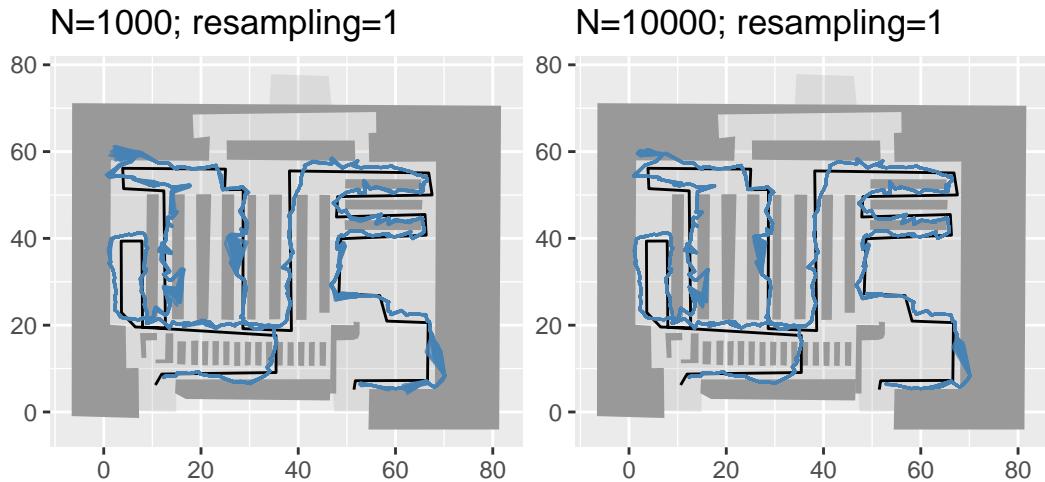
Liite A - Karttapolut

Liite sisältää tutkielman tulososioon liittyvät karttapolut. Kukin kartta käsittää $r = 30$ ajoa kullekin suunnitteluparametrikombinaatiolla. Karttojen otsikossa on mainittu ainoastaan testatut suunnitteluparametrit, vakioarvoiset parametrit on esitettty luvussa 4.5.3.

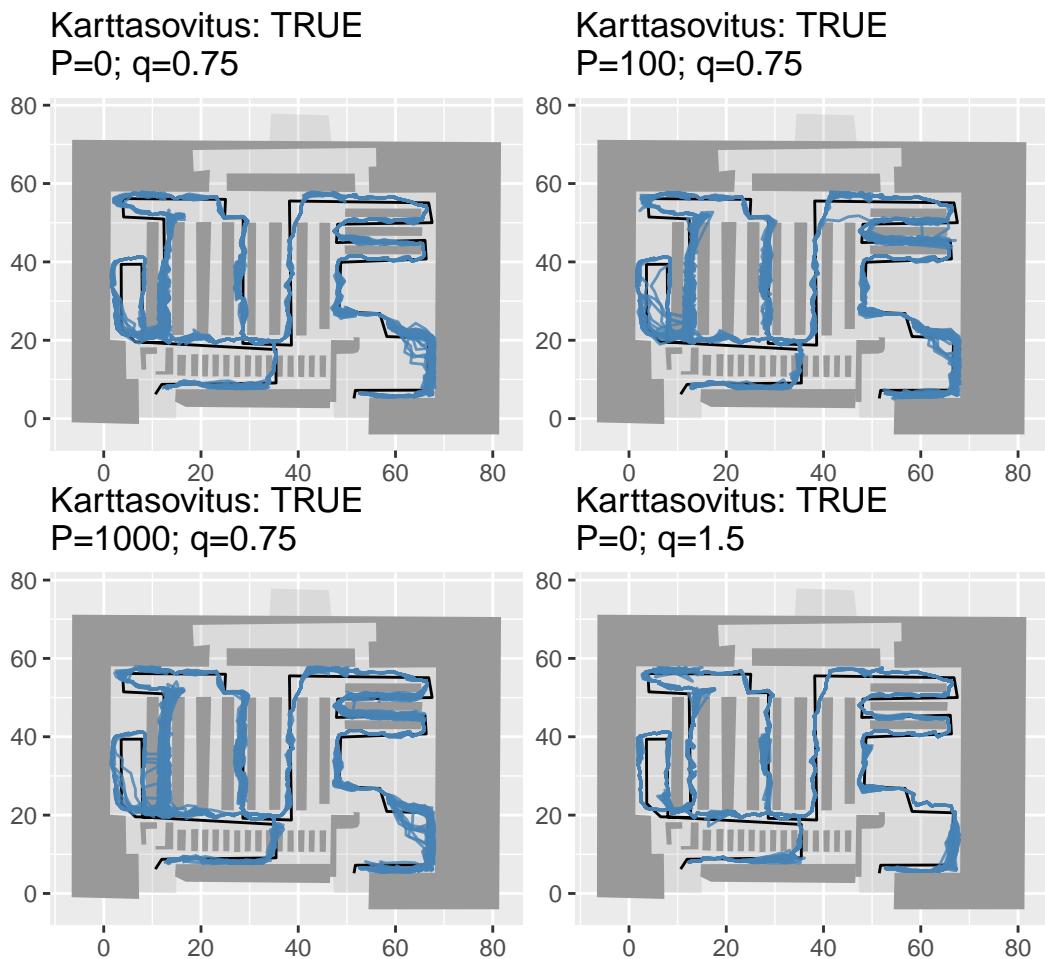
Vaihe 1



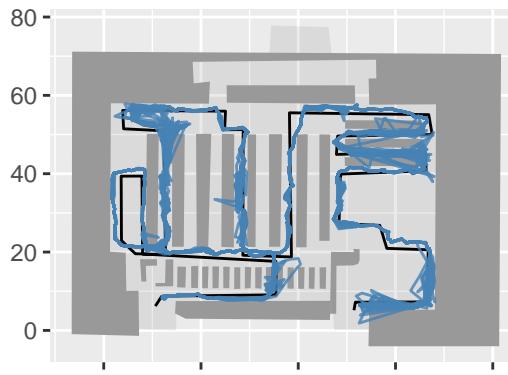




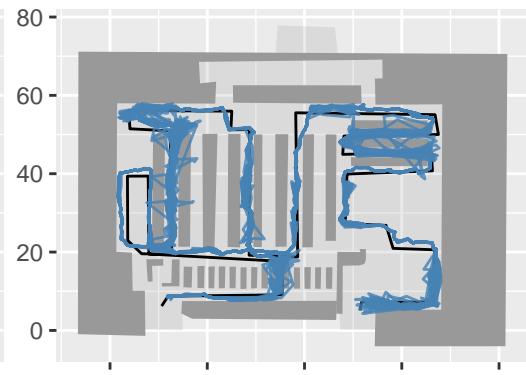
Vaihe 2



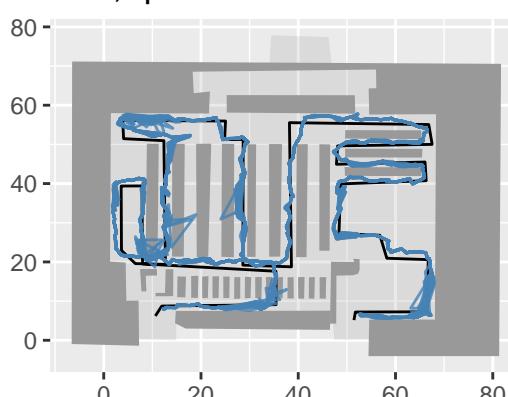
Karttasovitus: TRUE
 $P=100$; $q=1.5$



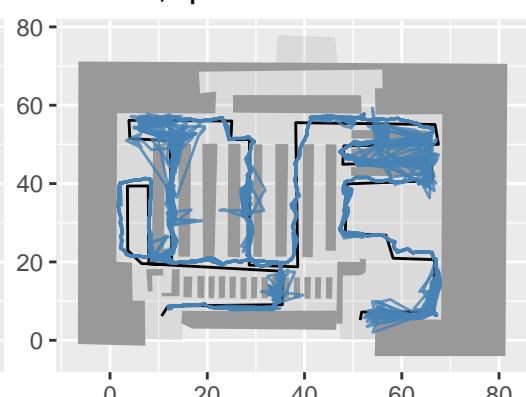
Karttasovitus: TRUE
 $P=1000$; $q=1.5$



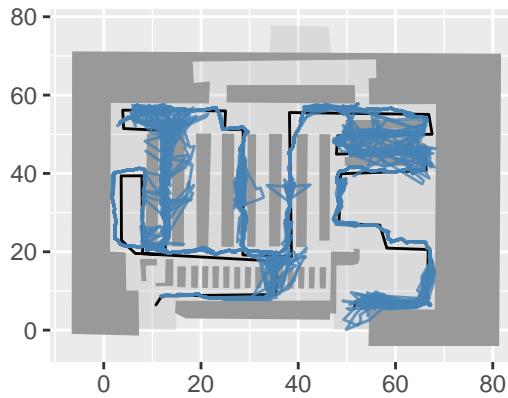
Karttasovitus: TRUE
 $P=0$; $q=2$



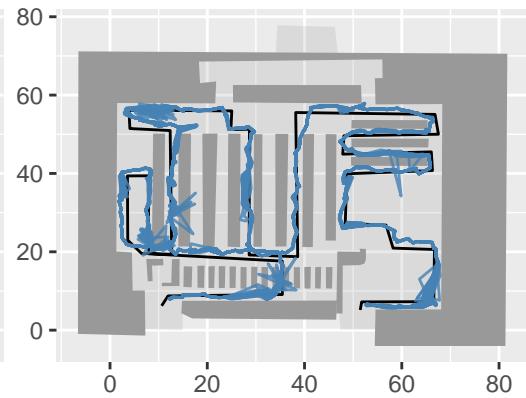
Karttasovitus: TRUE
 $P=100$; $q=2$



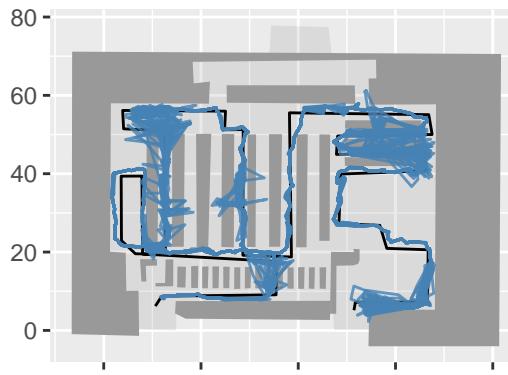
Karttasovitus: TRUE
 $P=1000$; $q=2$



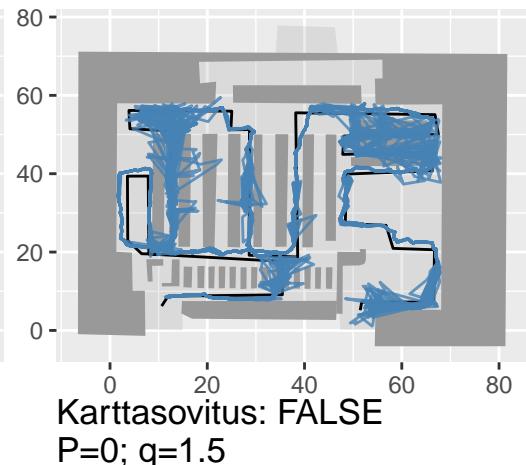
Karttasovitus: TRUE
 $P=0$; $q=2.5$



Karttasovitus: TRUE
 $P=100$; $q=2.5$



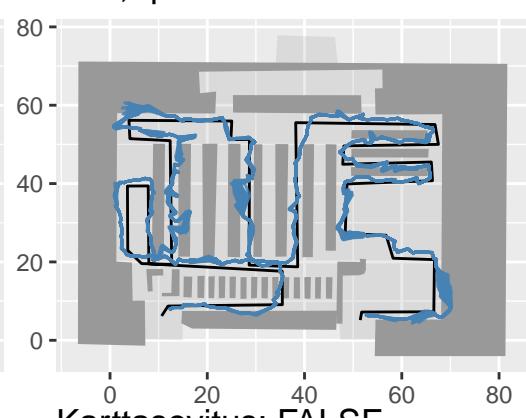
Karttasovitus: TRUE
 $P=1000$; $q=2.5$



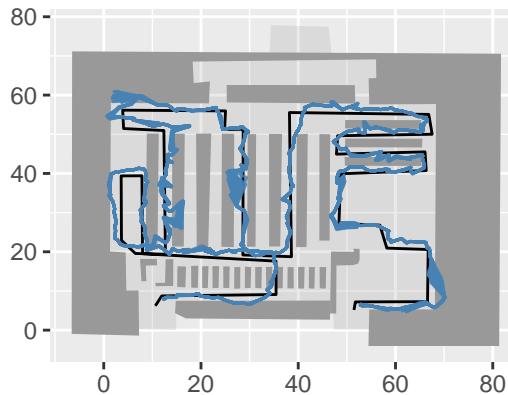
Karttasovitus: FALSE
 $P=0$; $q=0.75$



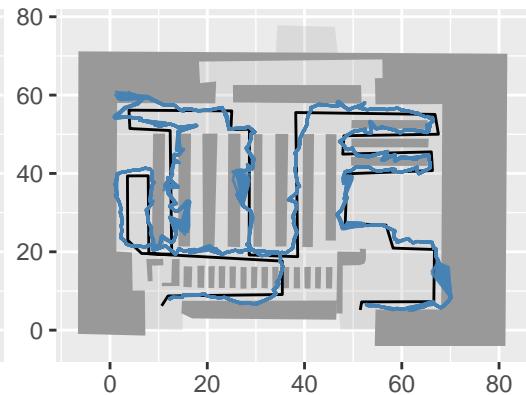
Karttasovitus: FALSE
 $P=0$; $q=1.5$



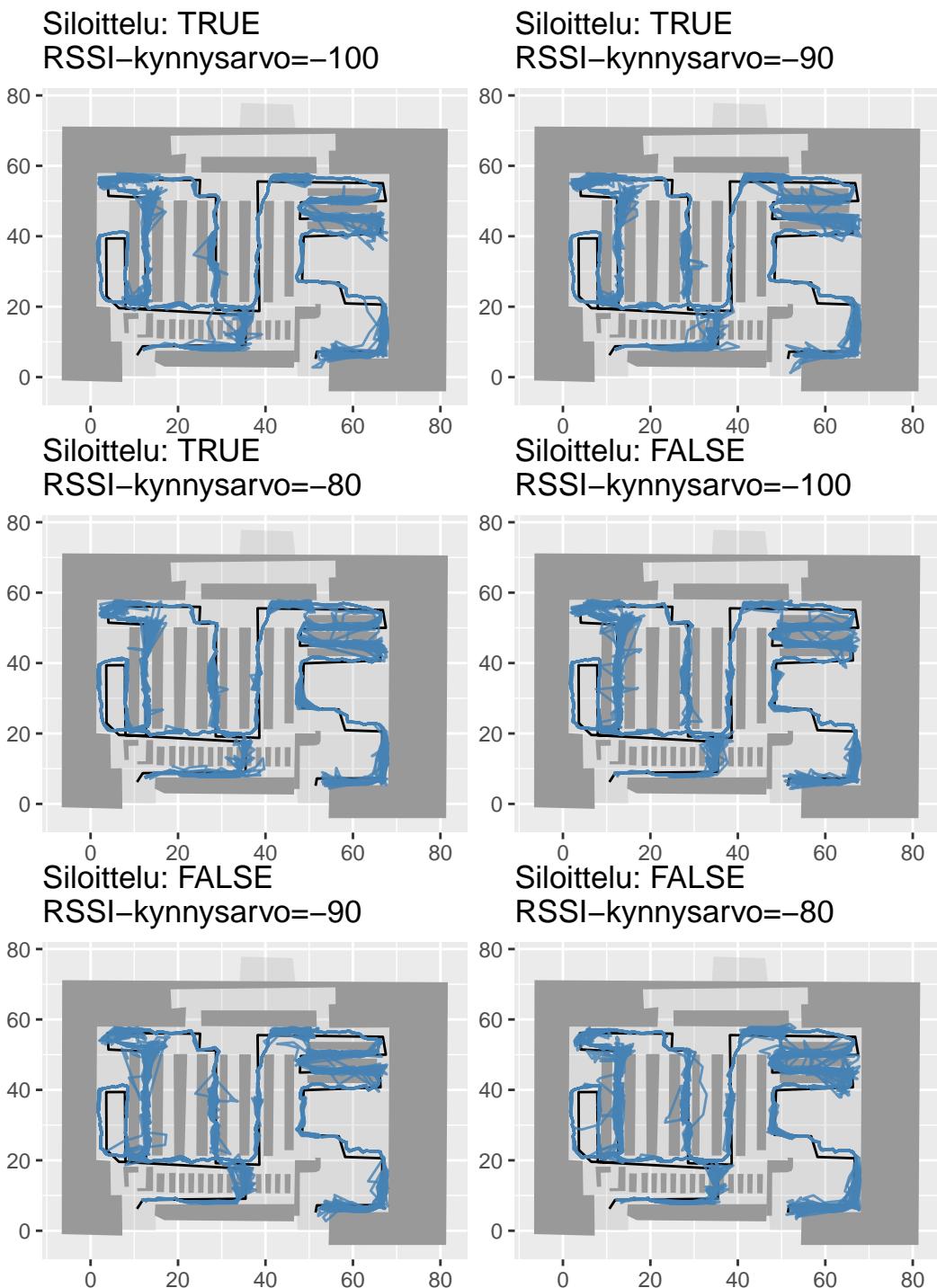
Karttasovitus: FALSE
 $P=0$; $q=2$



Karttasovitus: FALSE
 $P=0$; $q=2.5$



Vaihe 3



Lähteet

- [1] Johan Alenlöv and Jimmy Olsson. Particle-based adaptive-lag online marginal smoothing in general state-space models. *IEEE Transactions on Signal Processing*, 67(21), 2019.
- [2] Jayaprasad Bojja, Jussi Collin, Simo Särkkä, and Jarmo Takala. Pedestrian localization in moving platforms using dead reckoning, particle filtering and map matching. pages 1116–1120, 2015. doi: 10.1109/ICASSP.2015.7178143.
- [3] Olivier Cappé, Simon J. Godsill, and Eric Moulines. An overview of existing methods and recent advances in sequential monte carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.
- [4] Yi Cheng, Wenbo Ren, Chunbo Xiu, and Yiyang Li. Improved particle filter algorithm for multi-target detection and tracking. *Sensors*, 24(14), 2024.
- [5] Nicolas Chopin. Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *The Annals of Statistics*, 32(6):2385–2411, 2004.
- [6] Dan Crisan. The stochastic filtering problem: A brief historical account. *Journal of Applied Probability*, 51A:13–22, 2014.
- [7] Dan Crisan and Arnaud Doucet. Convergence of sequential monte carlo methods. URL https://www.stats.ox.ac.uk/~doucet/crisain_doucet_convergenceofSMC2000.pdf.
- [8] Dan Crisan and Arnaud Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, 2002.
- [9] Pavel Davidsonl, Jussi Collin, and Jarmo Takala. Application of particle filters for indoor positioning using floor plans. pages 1–4, 2010. doi: 10.1109/UPINLB.2010.5653830.
- [10] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.
- [11] N.J. Gordon, D.H. Salmond, and A.F.M Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F (Radar Signal Process)*, 140(2):107–113, 1993.

- [12] Mohinder S. Grewal and Angus P. Andrews. Applications of kalman filtering in aerospace 1960 to the present. *IEEE Control Systems Magazine*, 30(3):69–78, 2010.
- [13] Marco Gunia, Adrian Zinke, and Niko Joram. Analysis and design of a music-based angle of arrival positioning system. *ACM Transactions on Sensor Networks*, 19(3):66, 1–41.
- [14] Fredrik Gustafsson. Particle filter theory and practice with positioning applications. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):53–82, 2010.
- [15] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1996.
- [16] Ngoc-Huynh Ho, Phuc Huu Truong, and Gu-Min Jeong. Step-detection and adaptive step-length estimation for pedestrian dead-reckoning at various walking speeds using a smartphone. *Sensors*, 16(9), 2016.
- [17] Genshiro Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.
- [18] Mike Klaas, Nando de Freitas, and Arnaud Doucet. Toward practical n2 monte carlo: The marginal particle filter. *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, page 308–315, 2005. URL <https://arxiv.org/pdf/1207.1396.pdf>.
- [19] Anthony Lee and Nick Whiteley. Variance estimation in the particle filter. *Biometrika*, 105(3):609–625, 2018.
- [20] J. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998.
- [21] Alessandro Mastrototaro and Jimmy Olsson. Adaptive online variance estimation in particle filters: the alvar estimator. *Statistics and Computing*, 33(77), 2022.
- [22] Pierre Del Moral. Nonlinear filtering: Interacting particle resolution. *Markov Processes and Related Fields*, 2(4):555–580.
- [23] D. Munoz, R. Enriquez-Caldera, and C. Vargas. *Position Location Techniques and Applications*. Elsevier, 2009.
- [24] Samuel Nyobe, Fabien F. Campillo, and Serge Moto. The one step fixed-lag particle smoother as a strategy to improve the prediction step of particle filtering. *hal-03464987*, 2021. URL <https://inria.hal.science/hal-03464987v2/file/papier.pdf>.
- [25] George Oguntala, Raed Abd-Alhameed, Stephen Jones, James Noras, Mohammad Patwary, and Jonathan Rodriguez. Indoor location identification technologies for real-time iot-based applications: An inclusive survey. *Computer Science Review*, 30:55–79, 2018. ISSN 1574-0137. doi: <https://doi.org/10.1016/j.cosrev.2018.09.001>. URL <https://www.sciencedirect.com/science/article/pii/S1574013718301163>.

- [26] Jimmy Olsson and Randal Douc. Numerically stable online estimation of variance in particle filters. *Bernoulli*, 25(2):1504–1535, 2019.
- [27] V. Pierlot, M. Urbin-Choffray, and M. Van Droogenbroeck. A new three object triangulation algorithm based on the power center of three circles. *Research and Education in Robotics - EUROBOT 2011*, page 248–262.
- [28] Arno Solin, Simo Särkkä, Juho Kannala, and Esa Rahtu. Terrain navigation in the magnetic landscape: Particle filtering for indoor positioning. *2016 European Navigation Conference (ENC)*, pages 1–9, 2016. URL <https://api.semanticscholar.org/CorpusID:23211261>.
- [29] Simo Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013. URL https://users.aalto.fi/~ssarkka/pub/cup_book_online_20131111.pdf.