

Samuli Reinikainen

Pääkomponenttianalyysi ja parametriset elinaikamallit ohivirtausmoottoreiden jäljellä olevan käyttöiän ennustamisessa

Sisältö

1	Johdanto	3
2	Pääkomponenttianalyysi	4
2.1	Populaatiopääkomponenttien määritelmä	4
2.2	Pääkomponenttien latausten laskeminen	4
2.3	Otospääkomponentit	5
2.4	Pääkomponenttianalyysi korrelaatiomatriisista	5
2.5	Selitetty varianssi	6
2.6	Komponenttien valinta	6
3	Elinaika-analyysi	8
3.1	Elossaoloaika ja kestofunktio	8
3.2	Sensurointi	8
3.3	Uhkafunktio	9
3.4	Parametriset elinaikamallit	9
3.5	Nopeutetun vikaantumisajan malli	9
3.6	Suhteellisten vaarojen malli	10
4	Weibull-jakauma elinaikamalleissa	11
4.1	Weibull-jakauma	11
4.2	Elinaika-, kesto- ja uhkafunktio	11
4.3	Odotusarvo	12
4.4	Weibull-jakauman logaritmuunnos	12
4.5	Weibull nopeutetun vikaantumisajan malli	12
4.5.1	Estimointi ja testaus	13
4.5.2	Ennustaminen	14
5	Aineisto	16
5.1	Muuttuja-avaruuden laajennus	16
5.2	Lopullinen aineisto	17
6	Tulokset	18
6.1	Pääkomponenttianalyysi	18
6.2	Malli	20
6.3	Ennustaminen	21
6.4	Tulosten vertailua	22
	Lähteet	24
	Liitteet	25

1 Johdanto

Laitteiden kunnan ja jäljellä olevan käyttöiän mallintaminen ja ennustaminen ovat yhä tärkeämpi osa erikokoisten teollisuusalan toimijoiden kunnossapitoa. Laitteiden vikaantumisen ennustamisella voidaan saada merkittävää hyötyä niin tuottavuuden ja käyttövarmuuden kuin työturvallisuudenkin kannalta. Tätä ennustavaksi kunnossapidoksi kutsuttua toimintaa on lähitulevaisuudessa viemässä eteenpäin erilaisten helposti Internetiin yhdistettävien mittalaitteiden ja sensorien sekä niiden käyttöön tarvittavan tietoteknisen infrastruktuurin yleistyminen (Li, Wang ja He 2016, s. 43). Laitteista kerättäviä mittauksia on jo pitkään käytetty apuna laitteiden kunnan manuaalisessa valvonnassa, mutta mittauskapasiteetin lisääntyessä kaiken informaation hyödyntämiseen tarvitaan tilastotieteellisiä välineitä.

Elinaika-analyysi viittaa laajaan kirjoon menetelmiä, joilla mallinnetaan jonkin tapahtuman sattumiseen kuluva aikaa. Lääketieteessä elinaika-analyysin käytöllä on pitkä historia, mutta sen soveltaminen teollisuuden tarpeisiin, kuten laitteiden vikaantumisen ennustamiseen, on jäänyt vähemmälle huomiolle (Ma ja Krings 2008, s. 1). Ennustavassa kunnossapidossa on tyypillisesti käytetty tilastollisia menetelmiä, kuten yleistettyjä lineaarisia malleja, sekä laskennallisia menetelmiä, kuten neuroverkkoja (kts. esim. Babu et al. 2016 ja Li et al. 2018). Elinaika-analyysi tarjoaa kuitenkin toimivan ja vahvasti tilastolliseen päättelyyn perustuvan vaihtoehdon jäljellä olevan käyttöiän mallintamiseen ja ennustamiseen.

Sensorimittausten sekä niistä johdettujen johdannaiskovariaattien käyttö mallinnuksessa johtaa helposti dimensioiltaan suuriin aineistoihin. Lisäksi sensorimittaukset voivat olla vahvasti korreloituineita, jolloin useilla tilastollisilla menetelmillä voidaan kohdata sekä multikolinearisuus- että singulaarisuusongelmia. Ratkaisuna näihin ongelmiin voidaan käyttää dimensionvähennystekniikoita, joista käytetyimpiin kuuluu pääkomponenttianalyysi. Pääkomponenttianalyysiä voidaan käyttää yhdessä elinaika-analyysin kanssa luomaan vahva ja laskennallisesti tehokas tapa ennustaa laitteiden vikaantumista.

Tässä tutkielmassa esitellään pääkomponenttianalyysin sekä parametristen elinaikamallien teoriaa ja sovelletaan niitä mittausdatapohjaiseen jäljellä olevan käyttöiän ennustamiseen ohivirtausmootoreille. Kappaleessa kaksi käsitellään pääkomponenttianalyysin teoriaa sekä sen käyttöä dimensio-naalisuuden vähentämisessä. Kappaleessa kolme käydään läpi elinaika-analyysin teoriaa ja käsitteitä. Kappaleessa neljä käsitellään Weibull-jakaumaan pohjautuvien parametristen elinaikamallien teoriaa sekä niiden käyttöä ennustamisessa. Kappaleessa viisi esitellään käytetty aineisto ja käytetyt sensorimittauksista lasketut johdannaiskovariaatit. Kappaleessa kuusi käsitellään menetelmien soveltamista aineistoon sekä vertaillaan niiden tuottamia tuloksia muihin kirjallisuudessa esitettyihin tuloksiin.

2 Pääkomponenttianalyysi

Pääkomponenttianalyysi on menetelmä, jossa useita korreloituneita muuttujia sisältävälle aineistolle tehdään lineaarinen transformaatio, joka vähentää aineiston muuttujien määrää eli dimensionaalisuutta säilyttäen mahdollisimman suuren osan aineiston vaihtelusta. Tuloksena on ortogonaalinen joukko muuttujia, joita kutsutaan pääkomponenteiksi. Pääkomponentit järjestetään usein laskevasti alkupe-
räisessä aineistossa selittämänsä vaihtelun perusteella. (Jolliffe 2002, s. 1)

Pääkomponenttianalyysiä voidaan käyttää itsessäänkin tilastolliseen päättelyyn, mutta usein las-
kettuja pääkomponentteja käytetään kovariaatteina muissa menetelmissä dimensionaalisuuden vähen-
tämiseksi. Tässä kappaleessa esitellään pääkomponenttianalyysin peruseriaatteet keskittyen etenkin
niihin ominaisuuksiin, jotka ovat olennaisia pääkomponenttien käytössä regressiomallinnuksessa.

2.1 Populaatiopääkomponenttien määrittely

Olkoon $X = (X_1, X_2, \dots, X_p)$ p :stä kappaleesta satunnaismuuttujia koostuva satunnaisvektori. Ensim-
mäinen pääkomponentti Z_1 määritellään lineaarikombinaationa

$$(2.1) \quad Z_1 = \mathbf{k}'_1 X = k_{11}X_1 + k_{12}X_2 + \dots + k_{1p}X_p,$$

jolla on suurin varianssi kaikista lineaarikombinaatioista $\alpha'X$, joissa α on yksikkövektori. Toisin
sanoen \mathbf{k}_1 :lle pätee $\mathbf{k}_1 = \arg \max_{|\alpha|=1} \{\text{Var}(\alpha'X)\}$.

Toiseksi pääkomponentiksi Z_2 määritellään vastaavasti lineaarikombinaatio

$$Z_2 = \mathbf{k}'_2 X = k_{21}X_1 + k_{22}X_2 + \dots + k_{2p}X_p,$$

jolla on suurin varianssi lineaarikombinaatioista, jotka ovat ortogonaalisia Z_1 :n kanssa. Siis \mathbf{k}_2 :lle
pätee $\mathbf{k}_2 = \arg \max_{|\alpha|=1} \{\text{Var}(\alpha'X)\}$ ja $\mathbf{k}'_1 X \cdot \mathbf{k}'_2 X = 0$. Kolmas pääkomponentti määritellään samoin

lineaarikombinaatioksi $Z_3 = \mathbf{k}'_3 X$, jolla on suurin varianssi lineaarikombinaatioista, jotka ovat orto-
gonaalisia Z_1 :n ja Z_2 :n kanssa. Yleisesti m :s pääkomponentti määritellään lineaarikombinaatioksi Z_m ,
jolla on suurin varianssi lineaarikombinaatioista, jotka ovat ortogonaalisia aiempien pääkomponenttien
 Z_1, Z_2, \dots, Z_{m-1} kanssa. Pääkomponentteja voidaan määrittellä yhteensä p kappaletta, jolloin muuttujia
on siis yhtä monta kuin alkuperäisessä aineistossa. Kertoimia \mathbf{k}_m kutsutaan m :nnen pääkomponentin
latauksiksi. (Jolliffe 2002, s. 2-3)

2.2 Pääkomponenttien latausten laskeminen

Olkoon X määritelty kuten edellä ja olkoon Σ sen tunnettu kovarianssimatriisi. Halutaan löytää \mathbf{k} ,
joka maksimoi varianssin $\text{Var}(\mathbf{k}'X) = \mathbf{k}'\Sigma\mathbf{k}$ ehdolla $|\mathbf{k}| = \mathbf{k}'\mathbf{k} = 1$. Varianssin maksimoinnissa
voidaan käyttää muutakin rajoitetta \mathbf{k} :lle, kuten ehtoa $\sum_{i=1}^p k_i = 1$, josta päädytään kuitenkin vaikeam-
paan optimointiongelmaan. Ensimmäisenä mainittua ehtoa käyttämällä maksimointiongelma voidaan
ratkaista Lagrangen kertoimien tekniikalla. Määritellään aluksi $g(\mathbf{k}) = \mathbf{k}'\mathbf{k} - 1$, jolloin Lagrangen
funktio saa muodon

$$L(X, \mathbf{k}, \lambda) = \text{Var}(\mathbf{k}'X) - \lambda g(\mathbf{k}) = \mathbf{k}'\Sigma\mathbf{k} - \lambda(\mathbf{k}'\mathbf{k} - 1),$$

missä λ on Lagrangen kerroin. Alkuperäisen maksimointiongelman ratkaisu löytyy nyt jostain Lagrangen funktion ääriarvokohdasta (Lucht 2016, s. 11). Kuten Jolliffe (2002, s. 5) osoittaa, ääriarvokohdat toteuttavat yhtälön

$$(\Sigma - \lambda I)\mathbf{k} = \mathbf{0}.$$

Huomataan, että kyseessä on Σ :n kovarianssimatriisin ominaisarvojen määrittely-yhtälö, jossa \mathbf{k} on Σ :n ominaisvektori ja λ on Σ :n ominaisarvo. Ominaisarvon ominaisuuksien mukaan maksimoitava varianssi voidaan nyt kirjoittaa muodossa

$$(2.2) \quad \text{Var}(\mathbf{k}'X) = \mathbf{k}'\Sigma\mathbf{k} = \mathbf{k}'\lambda\mathbf{k} = \lambda\mathbf{k}'\mathbf{k} = \lambda.$$

Pääkomponentin varianssi siis maksimoituu, kun λ on Σ :n suurin ominaisarvo ja \mathbf{k} on sitä vastaava ominaisvektori. Tällöin pääkomponentin varianssin arvoksi saadaan λ .

Samaan tapaan voidaan johtaa toisen pääkomponentin muoto. Pääkomponenttien ortogonaalisuuden johdosta optimointitehtävään tulee lisärajoitus $\mathbf{k}'_1\mathbf{k}_2 = 0$, mikä muuttaa Lagrangen funktion muotoon

$$L(X, \mathbf{k}, \lambda) = \mathbf{k}'\Sigma\mathbf{k} - \lambda(\mathbf{k}'\mathbf{k} - 1) - \phi\mathbf{k}'_1\mathbf{k}_2,$$

missä λ ja ϕ ovat Lagrangen kertoimia. Jolliffe (2002, s. 5) osoittaa, että kyseisen funktion ääriarvot toteuttavat yhtälön

$$(\Sigma - \lambda I)\mathbf{k}_2 = \mathbf{0},$$

joten toisen pääkomponentin osalta päädytään samaan päättelyyn kuin ensimmäisen komponentin tapauksessa. Pääkomponenttien ortogonaalisuuden takia toisen pääkomponentin latauksiksi ei voida valita kovarianssimatriisin suurinta ominaisarvoa vastaavaa ominaisvektoria, vaan valitaan toiseksi suurinta ominaisarvoa vastaava ominaisvektori. Samaa päättelyä voidaan jatkaa seuraaville pääkomponenteille ja yleisesti saadaan, että m :nnen pääkomponentin latausvektori muodostuu kovarianssimatriisin m :nneksi suurinta ominaisarvoa vastaavasta ominaisvektorista λ_m , joka on myös kyseisen pääkomponentin varianssin arvo.

2.3 Otospääkomponentit

Edellä esitellyt määritelmät ja tulokset populaatiopääkomponenteille ovat suoraan yleistettävissä myös otosvastineille eli otospääkomponenteille. Olkoon $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ havaintomatriisi, joka koostuu n :stä kappaleesta havaintoja p :lle satunnaismuuttujalle. Ensimmäinen otospääkomponentti z_1 koostuu tällöin alkioista $z_{i1} = \mathbf{k}'_1\mathbf{x}_i$, missä latausvektori \mathbf{k}_1 on valittu siten, että se on yksikkövektori, joka maksimoi otosvariانسsin

$$\frac{1}{n-1} \sum_{i=1}^n (z_{i1} - \bar{z}_1).$$

Täysin populaatiopääkomponentteja vastaavan päättelyketjun kautta saadaan, että latausvektori \mathbf{k}_i on havaintomatriisista \mathbf{x} estimoidun otoskovarianssimatriisin S suurinta ominaisarvoa vastaava ominaisvektori. Yleisesti saadaan, että m :s otospääkomponentti z_m koostuu alkioista $z_{i1} = \mathbf{k}'_m\mathbf{x}_i$, missä \mathbf{k}_m on otoskovarianssimatriisin S m :nneksi suurinta ominaisarvoa vastaava ominaisvektori. (Jolliffe 2002 s. 30)

2.4 Pääkomponenttianalyysi korrelaatiomatriisista

Pääkomponenttianalyysi on riippuvainen mitta-asteikosta, koska mitta-asteikon muuttaminen vaikuttaa varianssin suuruuteen ja varianssin suuruudella on suora vaikutus kovarianssimatriisin ominaisvek-

toreihin ja siten myös pääkomponenttien latauksiin. Tästä syystä Jolliffe (2002, s. 21) suosittelee laskemaan pääkomponenttien lataukset standardoidusta aineistosta, jossa jokainen muuttuja on keskitetty omalla keskiarvolla ja jaettu omalla keskihajonnallaan. Standardoidun aineiston kovarianssimatriisi on alkuperäisen aineiston korrelaatiomatriisi, joten standardoidun aineiston pääkomponenttien latausvektorit saadaan alkuperäisen aineiston korrelaatiomatriisin ominaisvektoreista.

2.5 Selitetty varianssi

Edellä kaavassa (2.2) saatiin m :nneille pääkomponentille tulos, jonka mukaan

$$\text{Var}(Z_m) = \lambda_m,$$

missä λ_m on Σ :n m :nneksi suurin ominaisarvo. Pääkomponentit ovat siis laskevassa järjestyksessä varianssin suhteen siten, että Z_1 :n varianssi on suurin. Ominaisarvojen ominaisuuksista tiedetään, että $n \times n$ matriisille A pätee

$$\text{tr}(A) = \sum_{i=1}^n \lambda_i,$$

missä $\lambda_1, \lambda_2, \dots, \lambda_n$ ovat matriisin A ominaisarvot. Kovarianssimatriisin Σ diagonaalelementit ovat sitä vastaavan satunnaisvektorin $X = (X_1, X_2, \dots, X_p)$ elementtien varianssit, joten kovarianssimatriisille pätee

$$(2.3) \quad \text{tr}(\Sigma) = \sum_{i=1}^p \text{Var}(X_i) = \sum_{i=1}^p \lambda_i.$$

Pääkomponenttien varianssien summa on siis yhtä suuri kuin alkuperäisten satunnaismuuttujien varianssien summa. Pääkomponentin Z_m alkuperäisten muuttujien kokonaisvarianssista selittämälle osuudelle saadaan edellisten perusteella kaava

$$(2.4) \quad \frac{\text{Var}(Z_m)}{\sum_{i=1}^p \text{Var}(X_i)} = \frac{\lambda_m}{\sum_{i=1}^p \lambda_i}.$$

Kumulatiivinen selitetty varianssi pääkomponentille Z_m määritellään pääkomponenttien Z_1, Z_2, \dots, Z_m selitettyjen varianssien summana

$$(2.5) \quad \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}.$$

2.6 Komponenttien valinta

Käytännön analyysissä käytettävien pääkomponenttien valintaan on useita lähestymistapoja. Valintakriteereihin vaikuttaa se, käytetäänkö pääkomponentteja itsessään tilastolliseen päättelyyn vai esimerkiksi regressiomallinnuksen tukena. Mikäli pääkomponentteja käytetään itsessään analyysissä, kriteerinä pääkomponenttien lukumäärän valinnassa käytetään usein niiden selittämän kumulatiivisen varianssin määrää. Tällöin analyysiin valittaisiin siis m ensimmäistä pääkomponenttia siten, että m on pienin määrä komponentteja, joiden kumulatiivinen varianssi ylittää valitun raja-arvon. Jolliffe (2002, s. 113) ehdottaa, että kumulatiivisen selitetyn varianssin raja-arvo valittaisiin 70-90 %:n väliltä.

Pääkomponenttien valinta regressiomallinnuksessa sisältää selitetyn varianssin kriteerin lisäksi ulkopuolisen kriteerin siitä, miten hyvin pääkomponentit selittävät regressiomallinnuksen kohteena

olevaa muuttujaa. Pääkomponenttien valintaan voidaan soveltaa tavallisia regressiomallien mallinvalintakriteerejä ja näitä voidaan yhdistää selitetyn varianssin kriteeriin.

Akaiken informaatiokriteeri (AIC) on mallinvalintakriteeri, joka ottaa huomioon mallin osuvuuden sekä estimoitujen parametrien määrän. Se lasketaan kaavalla

$$(2.6) \quad AIC = 2k - 2 \log(\hat{L}),$$

missä k on estimoitujen parametrien määrä ja \hat{L} on estimoidun uskottavuusfunktion arvo. Malleja vertailtaessa valitaan se malli, jolla on pienin AIC:n arvo. Pääkomponenttien valintaan AIC:tä voidaan soveltaa lisäämällä komponentteja mallin kovariaateiksi järjestyksessä selitetyn varianssin suhteen, kunnes AIC:n arvo saavuttaa lokaalin minimin. Myös muita informaatiokriteereitä on käytetty mallin valinnassa. Toinen paljon käytetty kriteeri on Bayesin informaatiokriteeri, jonka käyttäminen johtaa usein malleihin, joiden dimensio on pienempi.

3 Elinaika-analyysi

Elinaika-analyysi on joukko tilastollisia menetelmiä, jotka pyrkivät kuvailemaan ja mallintamaan jonkin tapahtuman ilmenemiseen kuluvaan aikaan. Tapahtumaksi voidaan ajatella mikä tahansa tilanne, jossa havaintoyksikkö siirtyy selkeästi määritetystä tilasta toiseen, kuten esimerkiksi ihmisen kuolema tai laitteen hajoaminen (Kleinbaum ja Klein 1996, s. 4). Elinaika-analyysi pitää sisällään laajan kirjon erilaisia menetelmiä, mutta tässä tutkielmassa keskitytään parametrisiin elinaikamalleihin sekä etenkin Weibull-jakaumaan. Seuraavissa kappaleissa määritellään elinaika-analyysin ja parametristen elinaikamallien peruskäsitteitä.

3.1 Elossaoloaika ja kestofunktio

Lääketieteellisestä taustasta johtuen tapahtuman ilmenemiseen kuluvaan aikaan kutsutaan usein selviytymisajaksi tai elossaoloajaksi ja itse tapahtumaa epäonnistumiseksi. Kuvattakoon elossaoloaikaan jatkuvalla ja ei-negatiivisella satunnaismuuttujalla T , joka noudattaa todennäköisyysjakaumaa, jonka tiheysfunktio on $f(t)$ ja kertymäfunktio $F(t)$. T :n noudattamaa jakaumaa kutsutaan tällöin elinaikajakaumaksi ja jakauman tiheysfunktioita elinaikafunktioiksi. T :n kertymäfunktio kuvaa todennäköisyyttä tapahtuman ilmenemiselle ennen ajan t kulumista, mutta elinaikanalyysissä ollaan kuitenkin usein kiinnostuneita tämän vastatapahtumasta, eli tapahtuman ilmenemisestä ajan t jälkeen. Tämän todennäköisyyttä kuvaa T :n kestofunktio

$$(3.1) \quad S(t) = P(T > t) = 1 - F(t).$$

Kestofunktioksi käy mikä tahansa laskeva funktio, jolle pätee $S(0) = 1$ ja $\lim_{t \rightarrow \infty} S(t) = 0$. (Kleinbaum ja Klein 1996, s. 9.)

Kestofunktiolle $S(t)$ voidaan osoittaa ominaisuus

$$(3.2) \quad E(T) = \int_0^{\infty} S(t) dt,$$

missä $E(T)$ on elossaoloajan T odotusarvo (Muldowney 2012). Odotusarvo tulkitaan elinaikajakaumien tilanteessa etenkin väestö- ja lääketieteen sovelluksissa odotetuksi eliniäksi.

3.2 Sensurointi

Elinaika-analyysissä kohdataan usein tilanteita, joissa tapahtumaan kulunut tarkka aika ei ole tiedossa joillekin havaintoyksiköistä. Tätä ilmiötä kutsutaan sensuroinniksi ja havaintoja, joissa tarkka elossaoloaika ei ole tiedossa, kutsutaan sensuroiduiksi havainnoiksi. Jos tapahtumaa ei ole ilmentynyt aineiston tarkasteluvälillä, elossaoloajasta tiedetään ainoastaan, että se on suurempi kuin viimeiseen havaintoon asti kulunut aika. Tällöin havaintoa kutsutaan oikealta sensuroiduksi. Vastaavasti mikäli elossaoloajasta on tiedossa vain yläraja, havaintoa kutsutaan vasemmalta sensuroiduksi. Jos tiedetään sekä elinajan ylä- että alaraja, havaintoa kutsutaan välisensuroiduksi (Kleinbaum ja Klein 1996, s. 8). Tässä tutkielmassa keskitytään tilanteeseen, jossa sensurointia ei esiinny, joten sen vaikutukset analysointiin jätetään jatkossa huomiotta.

3.3 Uhkafunktio

Uhkafunktio kuvaa tapahtuman hetkellistä todennäköisyyttä ajanhetkellä t ehdolla, että tapahtumaa ei ole ilmennyt ennen ajanhetkeä t . Uhkafunktio määritellään (Kleinbaum ja Klein 1996, s.10) raja-arvona

$$(3.3) \quad \lambda(t) = \lim_{dt \rightarrow 0} \frac{P(t < T < t + dt | T > t)}{dt}.$$

Osoittajassa olevalle ehdolliselle todennäköisyydelle pätee Bayesin kaavan mukaan

$$P(t < T < t + dt | T > t) = \frac{P(t < T < t + dt \cap T > t)}{P(T > t)} = \frac{P(t < T < t + dt)}{S(t)}.$$

Lisäksi infinitesimaalisen pienelle aikavälille dt pätee

$$P(t < T < t + dt) = f(t)dt,$$

joten uhkafunktion, elinaikafunktion ja kestofunktion yhteydelle saadaan tulos

$$(3.4) \quad \begin{aligned} \lambda(t) &= \lim_{dt \rightarrow 0} \frac{P(t < T < t + dt | T > t)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{f(t)dt/S(t)}{dt} \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Nämä kolme elinaika-analyysin keskeisintä funktiota ovat yksinkertaisessa yhteydessä toisiinsa siten, että kolmannen funktioista voi aina selvittää, jos kaksi muuta tiedetään.

3.4 Parametriset elinaikamallit

Parametrisissa elinaikamalleissa elossaoloajan oletetaan noudattavan jotain parametrissa todennäköisyysjakaumaa, ja päättely pohjautuu aineistosta lasketuille parametristimaateille. Käytetyimpiä oletuksia elinaikajakaumaksi ovat muun muassa gammajakauma, log-logistinen jakauma, log-normaalijakauma, Weibull-jakauma sekä sen erikoistapaus eksponenttijakauma (Kleinbaum ja Klein 1996, s. 292). Jakaumaoletuksen ollessa voimassa elossaoloajan T tai jonkun elossaoloajan funktion $f(T)$ riippumista kovariaateista voidaan mallintaa elinaikaregressiomalleilla. Näistä kaksi käytetyintä ovat suhteellisten vaarojen malli, jossa kovariaattien oletetaan vaikuttavan multiplikatiivisesti uhkafunktion arvoon, ja nopeutetun vikaantumisaajan malli, jossa kovariaattien multiplikatiivinen vaikutus kohdistuu selviytymisaikaan.

3.5 Nopeutetun vikaantumisaajan malli

Nopeutetun vikaantumisaajan mallissa oletetaan, että elossaoloajan T ja kovariaattien $\mathbf{X} = (X_1, X_2, \dots, X_m)$ välillä pätee log-lineaarinen suhde

$$(3.5) \quad \log(T) = -\beta\mathbf{X} + \sigma\epsilon,$$

missä $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_m]'$ on regressiokerroinvektori, ϵ on virhetermi, joka noudattaa jotain sopivaa jakaumaa ja σ on virhetermin skaalausparametri (Collett 2003, s. 207). Malliin voidaan ottaa mukaan myös vakiotermi β_0 . Mallin mukaan elinaikajakaumalle pätee

$$(3.6) \quad \begin{aligned} T &= \exp(-\boldsymbol{\beta}X + \sigma\epsilon) \\ &= \exp(-\boldsymbol{\beta}X) \exp(\sigma\epsilon) \\ &= \exp(-\boldsymbol{\beta}X)T_0, \end{aligned}$$

missä $T_0 = \exp(-\sigma\epsilon)$ on vertailuyksikön eli yksikön, jonka kovariaattien arvot ovat nolliä, elinaikajakauma ja $\exp(-\boldsymbol{\beta}X)$ on kovariaattien vaikutus. Nopeutetun vikaantumisajan mallissa kovariaattien muutoksella on siis multiplikatiivinen vaikutus elossaoloaikaan.

Vertailuyksikön kestofunktio määräytyy jakaumaoletuksen perusteella ja sille pätee

$$(3.7) \quad S_0(t) = P(T_0 > t) = P(\exp(\sigma\epsilon) > t).$$

Kestofunktion arvo havaituilla kovariaateilla X on

$$(3.8) \quad \begin{aligned} S(t|X) &= P(T > t) \\ &= P(\exp(-\boldsymbol{\beta}X)T_0 > t) \\ &= P(T_0 > t \exp(\boldsymbol{\beta}X)) \\ &= S_0(t \exp(\boldsymbol{\beta}X)). \end{aligned}$$

Havaintoyksikön, jolta on havaittu kovariaatit X , todennäköisyys olla elossa ajan t jälkeen on siis yhtä suuri kuin vertailuyksikön todennäköisyys olla elossa ajan $t \exp(\boldsymbol{\beta}X)$ jälkeen. Kovariaattien vaikutuksen voidaan siis tulkita ikään kuin nopeuttavan tai hidastavan ajan kulkua vertailuyksikköön nähden, mistä mallin nimi onkin peräisin. Yksittäiselle kovariaatille X_j kiihdytyskerroin on $\exp(\beta_j)$. Kiihdytyskerroin kuvaa kiihdyttävää vaikutusta vikaantumisaikaan, kun kovariaatin arvo kasvaa yhdellä.

Samaan tapaan voidaan osoittaa (Collett 2003, s. 208), että nopeutetun vikaantumisajan mallissa uhkafunktiolle pätee

$$(3.9) \quad \lambda(t|X) = \exp(\boldsymbol{\beta}X)\lambda_0(\exp(\boldsymbol{\beta}X)t),$$

missä λ_0 on jakaumaoletuksesta määräytyvä uhkafunktio. Siis tapahtuman hetkellinen todennäköisyys ajanhetkellä t havaintoyksikölle, jonka kovariaattien vaikutus on $\exp(-\boldsymbol{\beta}X)$, on $\exp(\boldsymbol{\beta}X)$ -kertainen vertailuyksikön tapahtuman todennäköisyyteen ajanhetkellä $\exp(\boldsymbol{\beta}X)t$.

3.6 Suhteellisten vaarojen malli

Parametrisessa suhteellisten vaarojen mallissa pyritään mallintamaan uhkafunktiota suoraan käyttäen hyväksi elossaoloajan jakaumaoletusta. Havaintoyksikön uhkafunktion $\lambda(t)$ sekä havaittujen kovariaattien X välillä oletetaan olevan suhde

$$(3.10) \quad \lambda(t|X) = \lambda_0(t) \exp(-\boldsymbol{\beta}X),$$

missä $\lambda_0(t)$ on jakaumaoletuksesta määräytyvä vertailuyksikön uhkafunktio ja $\exp(-\boldsymbol{\beta}X)$ on kovariaattien vaikutus (Collett 2003, s. 57). Kuten nopeutetun vikaantumisajan mallissa mukaan voidaan ottaa myös vakiotermi β_0 . Suhteellisten vaarojen mallissa kovariaateilla on siis multiplikatiivinen vaikutus uhkafunktion arvoon, jolloin kahden havaintoyksikön uhkafunktioiden arvojen eli tapahtuman hetkellisten todennäköisyyksien suhde on havaintoyksiköiden kovariaattien vaikutusten $\exp(-\boldsymbol{\beta}X)$ suhde. Yksittäiselle kovariaatille X_j voidaan määritellä uhkasuhde $\exp(\beta_j)$, joka kuvaa multiplikatiivista vaikutusta tapahtuman hetkelliseen todennäköisyyteen, kun kovariaatin arvo kasvaa yhdellä.

4 Weibull-jakauma elinaikamalleissa

Weibull-jakauma on useissa elinaika-analyysin sovelluskohteissa olennainen jakauma, mutta sen käyttö korostuu etenkin laitteita koskevissa sovelluksissa. Lisäksi sillä on joitain uniikkeja parametrisissa elinaikamalleissa hyödyllisiä ominaisuuksia. Tässä kappaleessa käsitellään Weibull-jakaumaa ja sen käyttöä parametrisissa elinaikamalleissa. Aluksi esitellään Weibull-jakauma ja sen elinaika-analyysiin liittyvät funktiot, minkä jälkeen käsitellään niiden käyttöä nopeutetun vikaantumisaajan mallissa.

4.1 Weibull-jakauma

Satunnaismuuttuja T noudattaa Weibull-jakaumaa parametrein l ja k , jos sen kertymäfunktio on

$$(4.1) \quad F(t; l, k) = P(T < t) = \begin{cases} 1 - \exp(-(\frac{t}{l})^k) & t > 0 \\ 0 & t \leq 0 \end{cases},$$

missä $l > 0$ on jakauman skaalaparametri ja $k > 0$ on muotoparametri (McCool 2012, s. 73). Tällöin merkitään $T \sim W(l, k)$. Weibull-jakauman skaalaparametrille pätee

$$(4.2) \quad T \sim W(l, k) \implies aT \sim W(al, k),$$

jos $a > 0$, koska satunnaismuuttujan aT kertymäfunktiolle pätee

$$\begin{aligned} P(aT < t) &= P(T < \frac{t}{a}) \\ &= F(\frac{t}{a}) \\ &= 1 - \exp(-(\frac{t}{al})^k), \end{aligned}$$

joka on $W(al, k)$ -jakauman kertymäfunktio.

4.2 Elinaika-, kesto- ja uhkafunktio

Weibull-jakauman tiheysfunktio eli elinaikafunktio saadaan derivoimalla kertymäfunktio

$$(4.3) \quad f(t) = \frac{d}{dt}F(t) = \frac{k}{l}(\frac{t}{l})^{k-1} \exp(-(\frac{t}{l})^k).$$

Weibull-jakauman kestofunktio on kaavan (3.1) perusteella

$$(4.4) \quad S(t) = 1 - F(t) = \exp(-(\frac{t}{l})^k).$$

Uhkafunktio saadaan kaavan (3.4) perusteella

$$(4.5) \quad \lambda(t) = \frac{f(t)}{S(t)} = \frac{\frac{k}{l}(\frac{t}{l})^{k-1} \exp(-(\frac{t}{l})^k)}{\exp(-(\frac{t}{l})^k)} = \frac{k}{l}(\frac{t}{l})^{k-1}.$$

Ylläolevasta kaavasta (4.5) nähdään, että uhkafunktio on kasvava, jos $k \geq 1$ ja muulloin vähenevä.

4.3 Odotusarvo

Odotusarvo saadaan kaavan (3.2) sekä sijoituksen $u = (\frac{t}{l})^k$ avulla

$$(4.6) \quad E(T) = \int_0^{\infty} \exp(-(\frac{t}{l})^k) dt = \int_0^{\infty} \exp(-u) \frac{l}{k} u^{\frac{1}{k}-1} du = \frac{l}{k} \int_0^{\infty} \exp(-u) u^{\frac{1}{k}-1} du = \frac{l}{k} \Gamma(\frac{1}{k}),$$

missä Γ on gammafunktio.

4.4 Weibull-jakauman logaritminmuunnos

McCool (2012, s. 84-85) osoittaa, että jos $T \sim W(l, k)$, niin $Y = \log(T)$ noudattaa Gumbel-jakaumaa, jonka kertymäfunktio on

$$(4.7) \quad F(t, l, k) = P(Y < y) = 1 - \exp\{\exp[k(y - \log(l))]\}.$$

Tällöin merkitään $Y \sim G(\log(l), \frac{1}{k})$ ja muunnoksien $a = \log(l)$ ja $b = \frac{1}{k}$ merkitään $Y \sim G(a, b)$. Tiheysfunktio saadaan derivoimalla ja se on

$$(4.8) \quad f_Y(y, l, k) = k \exp[k(y - \log(l))] \exp\{-\exp[k(y - \log(l))]\}.$$

Vastaava tulos pätee myös toiseen suuntaan eli

$$(4.9) \quad Y \sim G(a, b) \implies T = \exp(Y) \sim W(\exp(a), \frac{1}{b}).$$

Gumbel-jakaumalle pätee

$$(4.10) \quad Y \sim G(a, b) \implies \alpha Y \sim G(\alpha a, \alpha b),$$

jos $\alpha > 0$, koska satunnaismuuttujan αY kertymäfunktioille pätee

$$\begin{aligned} P(\alpha Y < y) &= P(Y < \frac{y}{\alpha}) \\ &= F(\frac{y}{\alpha}) \\ &= 1 - \exp\{-\exp[-\frac{1}{b}(\frac{y}{\alpha} - a)]\} \\ &= 1 - \exp\{-\exp[-\frac{1}{\alpha b}(y - \alpha a)]\}, \end{aligned}$$

mikä on $G(\alpha a, \alpha b)$ -jakauman kertymäfunktio.

4.5 Weibull nopeutetun vikaantumisajan malli

Oletetaan, että elossaoloajan T ja kovariaattien X välillä pätee kaavan (3.5) log-lineaarinen suhde. Lisäksi oletetaan, että virhetermi ϵ noudattaa standardoitua Gumbel-jakaumaa, eli $\epsilon \sim G(0, 1)$. Tällöin kaavan (4.10) perusteella $\sigma \epsilon \sim G(0, \sigma)$. Mallin mukaiselle vertailuyksikön elinaikajakaumalle pätee nyt kaavan (4.9) mukaan

$$T_0 = \exp(\sigma \epsilon) \sim W(1, \frac{1}{\sigma}).$$

Elinaikajakaumalle pätee siis yleisesti kaavojen (3.6) ja (4.2) mukaan

$$T = \exp(-\beta X)T_0 \sim W(\exp(-\beta X), \frac{1}{\sigma}).$$

Elin aika on siis Weibull-jakautunut kaikilla kovariaattien arvoilla. Vertailuyksikön kestofunktio $S_0(t)$ on kaavan (3.7) mukaisesti $W(1, \frac{1}{\sigma})$ -jakauman kestofunktio ja yleisesti $S(t|X)$ on $W(\exp(-\beta X), \frac{1}{\sigma})$ -jakauman kestofunktio. Elinajan odotusarvolle pätee kaavan (4.6) mukaisesti

$$E(T_0) = \sigma \Gamma(\sigma + 1)$$

ja

$$E(T|X) = \exp(-\beta X)\sigma \Gamma(\sigma + 1).$$

Uhkafunktiolle pätee kaavan (3.9) perusteella

$$\lambda(t|X) = \exp(\beta X)\lambda_0(\exp(\beta X)t),$$

missä λ_0 on $W(1, \frac{1}{\sigma})$ -jakauman uhkafunktio. Sijoittamalla Weibull-uhkafunktioon (4.6) saadaan

$$\lambda(t|X) = \exp(\beta X) \frac{1/\sigma}{1} \left(\frac{\exp(\beta X)t}{1} \right)^{1/\sigma-1} = \frac{1/\sigma}{\exp(-\beta X)} \left(\frac{t}{\exp(-\beta X)} \right)^{1/\sigma-1},$$

mikä on $W(-\exp(\beta X), \frac{1}{\sigma})$ -jakauman uhkafunktio. Nyt uhkafunktioiden $\lambda_0(t)$ ja $\lambda(t|X)$ suhteelle pätee

$$(4.11) \quad \frac{\lambda(t|X)}{\lambda_0(t)} = \frac{\frac{1/\sigma}{\exp(-\beta X)} \left(\frac{t}{\exp(-\beta X)} \right)^{1/\sigma-1}}{1/\sigma t^{1/\sigma-1}} = \exp(\beta X)^{1/\sigma} = \exp\left(\frac{1}{\sigma}\beta X\right).$$

Weibull nopeutetun vikaantumisajan malli toteuttaa siis kaavan (3.10) suhteellisten vaarojen mallin oletuksen

$$\lambda(t|X) = \lambda_0(t) \exp(\alpha X),$$

jos valitaan $\alpha = \frac{1}{\sigma}\beta$. Weibull-jakauman tapauksessa nopeutetun vikaantumisajan malli ja suhteellisten vaarojen malli ovat siis parametrintia lukuunottamatta yhtäpitäviä malleja, joten Weibull-mallia voidaan tulkita kumpana tahansa. Weibull-jakauma on ainut jakauma, jolle kyseinen ominaisuus pätee (McCool 2012, s. 210).

4.5.1 Estimointi ja testaus

Weibull nopeutetun vikaantumisajan mallin parametrit $\sigma, \beta_1, \beta_2, \dots, \beta_p$ voidaan estimoida suurimman uskottavuuden menetelmällä. Uskottavuusfunktio havaitulla log-elinajalla t ja kovariaateilla X käyttäen kaksiparametrissa Gumbel-tiheysfunktioita on

$$(4.12) \quad L(\sigma, \beta|t) = f_G(\beta, \sigma|t) = \frac{1}{\sigma} \exp\left(\frac{t - \beta X}{\sigma}\right) \exp\left[-\exp\left(\frac{t - \beta X}{\sigma}\right)\right].$$

Uskottavuusfunktio n :lle riippumattomalle havainnolle on

$$(4.13) \quad L(\sigma, \beta|t) = \prod_{i=1}^n f_G(\beta, \sigma|t_i) = \prod_{i=1}^n \frac{1}{\sigma} \exp\left(\frac{t_i - \beta X}{\sigma}\right) \exp\left[-\exp\left(\frac{t_i - \beta X}{\sigma}\right)\right].$$

Log-uskottavuusfunktio on

$$\begin{aligned}
 l(\sigma, \boldsymbol{\beta}, t) &= \log(L(\sigma, \boldsymbol{\beta}, t)) \\
 &= \log\left(\prod_{i=1}^n \frac{1}{\sigma} \exp\left(\frac{t_i - \boldsymbol{\beta}X}{\sigma}\right) \exp\left[-\exp\left(\frac{t_i - \boldsymbol{\beta}X}{\sigma}\right)\right]\right) \\
 &= \sum_{i=1}^n \log\left(\frac{1}{\sigma} \exp\left(\frac{t_i - \boldsymbol{\beta}X}{\sigma}\right) \exp\left[-\exp\left(\frac{t_i - \boldsymbol{\beta}X}{\sigma}\right)\right]\right).
 \end{aligned}
 \tag{4.14}$$

Suurimman uskottavuuden estimaatit $\hat{\sigma}$ ja $\hat{\boldsymbol{\beta}}$ voidaan laskea log-uskottavuusfunktion ääriarvokohdistta Newtonin-Rhapsonin menetelmällä. Estimointi voidaan toteuttaa R-ohjelmiston survival-paketin WeibullReg-funktiolla.

Uskottavuusteorian mukaan suurella otoskoolla suurimman uskottavuuden estimaatit noudattavat normaalijakaumaa. Parametrien tilastollista merkittävyyttä voidaan testata Waldin testillä, jossa parametrille β nollahypoteesiä $H_0 : \beta = 0$ verrataan vaihtoehtoiseen hypoteesiin $H_A : \beta \neq 0$. Testisuure $\frac{\hat{\beta}}{\hat{s}(\beta)}$, missä $\hat{s}(\beta)$ on parametrin β estimoitu keskivirhe, noudattaa asympoottisesti standardoitua normaalijakaumaa, jos H_0 on tosi. Testistä laskettu p-arvo on $p = P(|Z| > \frac{\hat{\beta}}{\hat{s}(\beta)})$, missä $Z \sim N(0, 1)$ ja se kuvaa todennäköisyyttä, että laskettu tai sitä harvinaisempi testisuureen arvo havaittaisiin satunnaisesti. Jos p-arvo on pienempi kuin valittu riskitaso α , niin H_0 voidaan hylätä kyseisellä riskitasolla. Tällöin todetaan, että parametri on tilastollisesti merkitsevä ja kyseinen parametri on mallissa tarpeellinen. (McCool 2012, s. 69.)

4.5.2 Ennustaminen

Piste-ennusteena havaintoyksikön, jolta on havaittu kovariaatit X , elinajalle voidaan käyttää kaavan (4.16) ehdollista odotusarvoa. Jos $\hat{\sigma}$ ja $\hat{\boldsymbol{\beta}}$ ovat suurimman uskottavuuden estimaatit parametreille σ ja $\boldsymbol{\beta}$, niin suurimman uskottavuuden estimaatin invarianssiperiaatteen mukaan suurimman uskottavuuden estimaatti ehdolliselle odotusarvolle on

$$\hat{E}(T|X) = \exp(-\hat{\boldsymbol{\beta}}X) \hat{\sigma} \Gamma(\hat{\sigma}).
 \tag{4.15}$$

Henderson, Jones ja Stare (2001, s. 7) osoittavat, että jos havaitulla elinajalla t ja ennustetulla elinajalla p ennustetta pidetään tarkkana, kun $\frac{p}{k} < t < pk$ jollain kertoimen $k > 1$ arvolla, niin optimaalinen ennuste elinajalle on

$$p = \left[\frac{\log(k) \frac{2}{\hat{\sigma}}}{k^{\frac{1}{\hat{\sigma}}} - k^{-\frac{1}{\hat{\sigma}}}} \right]^{\hat{\sigma}} \exp(-\hat{\boldsymbol{\beta}}X).
 \tag{4.16}$$

Tällöin p minimoi yli k :n suuruisten multiplikatiivisten ennustevirheiden määrän. Kyseistä ennustemetelmää voidaan kutsua pienimmän ennustevirheen menetelmäksi. Henderson suosittelee käyttämään ennusteissa k :n arvoa kaksi tai kolme.

Ennusteiden osuvuutta voidaan tutkia RMSE-pisteytyksellä eli keskineliövirheen neliöjuurella, jonka pienempi arvo kuvaa parempaa ennustetarkkuutta. Keskineliövirheen neliöjuuri lasketaan ennusteista p_1, p_2, \dots, p_n ja toteutuneista arvoista t_1, t_2, \dots, t_n kaavalla

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - p_i)^2}.$$

Yksittäisen havainnon vaikutus RMSE-arvoon kasvaa suhteessa virheen neliöön, joten suuremmilla virheillä on suhteellisesti isompi vaikutus RMSE-arvoon. Tästä syystä se voi olla herkkä poikkeaville arvoille. Toinen tapa tutkia ennusteiden osuvuutta on keskimääräinen absoluuttinen virhe, MAE, jossa yksittäisen havainnon vaikutus kasvaa suoraan suhteessa virheeseen. Keskimääräinen absoluuttinen virhe lasketaan kaavalla

$$\frac{1}{n} \sum_{i=1}^n |t_i - p_i|.$$

Keskimääräinen absoluuttinen virhe on tulkittavissa suoraan kohdemuuttujan mittayksiköissä, kun taas RMSE-arvo voidaan tulkita ennustevirheen varianssin estimaattina.

5 Aineisto

Aineisto (Saxena ja Goebel, 2008) koostuu C-MAPPS-ohjelmalla simuloituista sensorimittausten monimuuttujaisista aikasarjoista Turbofan-ohivirtausmoottoreille. Aineiston kovariaatit koostuvat 21 sensorimittapistestä sekä kolmesta moottorin käyttöasetuksia kuvaavasta muuttujasta. Aikasarjat on simuloitu siten, että ne alkavat satunnaisesta moottorin elinkaaren kohdasta, jossa moottori toimii normaalisti ja kehittää kokoluokaltaan suurenevan vian aikasarjan aikana. Aikasarjoissa kahden peräkkäisen havainnon välinen aika on yksi moottorin toiminnallinen kierros eli sykli. Saxena et al. (2008) kuvaavat aineiston sekä simulointiprosessin yksityiskohtaisesti. Aineisto on jaettu neljään osa-aineistoon, joissa sensorimittauksia on simuloitu erilaisissa olosuhteissa ja eri määrillä erilaisia vikaantumismalleja.

Tässä tutkielmassa on käytössä ainoastaan ensimmäinen osa-aineisto. Kyseinen osa-aineisto jakautuu opetusaineistoon ja testiaineistoon, jotka molemmat sisältävät sata aikasarjaa, jotka ovat peräisin eri moottoreista. Opetusaineistossa aikasarjat loppuvat aina moottorin vikaantumiseen, joten moottorin jäljellä oleva elinikä tiedetään jokaisen mittaushetken kohdalla. Aineistossa ei siis esiinny ollenkaan sensurointia. Jäljellä oleva käyttöikä sykleissä i :nulle havainnolle voidaan laskea kaavalla $t_i = n_i - i$, missä n_i on kyseisen aikasarjan pituus. Aikasarjojen pituuden vaihtelevat välillä 128 – 362. Testiaineistossa aikasarja loppuu jokin tiedetty aika ennen laitteen vikaantumista.

5.1 Muuttuja-avaruuden laajennus

Aineiston sensorimittaukset sisältävät kiinnitetyn mitattavan suureen todellisen arvon k_i lisäksi satunnaisen häiriökomponentin ϵ_i , joten yksittäiset sensorimittausten arvot eivät ole välttämättä täysin luotettavia (Saxena et al. 2008, s. 5). Jos satunnaisten häiriöiden oletetaan olevan riippumaton mitattavasta arvosta ja toisistaan ja lisäksi oletetaan, että $E(\epsilon_i) = 0$ sekä $\text{Var}(\epsilon_i) = \sigma^2$, niin satunnaisen häiriön vaikutusta voidaan vähentää laskemalla kyseisen mittauksen liukuva keskiarvo. Jos on havaittu mittaukset $a_1 = k_1 + \epsilon_1, a_2 = k_2 + \epsilon_2, \dots, a_n = k_n + \epsilon_n$, niin liukuva keskiarvo määritellään korkeintaan n :n peräkkäisen havainnon otoskeskiarvona, eli

$$(5.1) \quad \hat{a}_i = \begin{cases} \frac{1}{n} \sum_{j=i-n}^i a_j & i > n \\ \frac{1}{i} \sum_{j=1}^i a_j & i \leq n. \end{cases}$$

Varianssin ominaisuuksien mukaan

$$\text{Var}(\hat{a}_i) = \text{Var}\left(\frac{1}{n} \sum_{j=i-n}^i a_j\right) = \frac{1}{n^2} \sum_{j=i-n}^i \text{Var}(a_j) = \frac{1}{n^2} n \text{Var}(\epsilon) = \frac{\sigma^2}{n},$$

jos $i > n$. Liukuvan keskiarvon satunnaisesta kohinasta johtuva varianssi on siis pelkän mittauksen varianssia pienempi. Liukuvan keskiarvon laskeminen jokaiselle sensorimittaukselle lisää aineistoon 21 uutta muuttujaa.

Yksi sensorimittausten havaintorivi sisältää tietoa moottorista kyseisen syklin aikana. Moottorin käyttöiän kannalta olennaista on kuitenkin sensorimittausten arvojen muutos ajan suhteen. Mittauksissa tapahtuvaa muutosta voidaan laajentaa uusiksi muuttujiksi usein eri tavoin, mutta tässä tutkielmassa laajennukseen käytetään liukuvaa otosvarianssia sekä liukuvaa vaihteluväliä. Liukuva otosvarianssi

määritellään liukuvan otoskeskiarvon tavoin korkeintaan $n:n$ peräkkäisen havainnon otosvarianssina

$$(5.2) \quad \sigma_i^2 = \begin{cases} \frac{1}{n-1} \sum_{j=i-n}^i (a_j - \hat{a}_i)^2 & i > n \\ \frac{1}{i-1} \sum_{j=1}^i (a_j - \hat{a}_i)^2 & 1 < i \leq n \\ 0 & i = 1, \end{cases}$$

missä \hat{a}_i on liukuva otoskeskiarvo. Liukuva vaihteluväli määritellään korkeintaan $n:n$ edellisen havainnon maksimi- ja minimiarvojen erotuksena

$$(5.3) \quad r_i = \begin{cases} \max(y_{i-n}, \dots, y_i) - \min(y_{i-n}, \dots, y_i) & i > n \\ \max(y_1, \dots, y_i) - \min(y_1, \dots, y_i) & i \leq n. \end{cases}$$

Liukuvan varianssin sekä liukuvan vaihteluvälin laskeminen jokaiselle sensorimittaukselle lisää aineistoon $2 \times 21 = 42$ uutta muuttujaa.

5.2 Lopullinen aineisto

Edellisessä kappaleessa tehtyjen muutosten jälkeen aineisto sisältää yhteensä 88 muuttujaa, joihin kuuluu 21 alkuperäistä sensorimittausmuuttujaa, 63 sensorimittauksista laskettua liukuvaa tunnuslukua, kolme moottorin käyttöasetuksia kuvaavaa muuttujaa sekä yksi jäljellä olevaa elinikää kuvaava muuttuja. Jatkossa käyttöasetusmuuttujiin viitataan lyhenteillä OP_1, OP_2, OP_3 , sensorimittauksiin lyhenteillä S_1, S_2, \dots, S_{21} , niiden liukuviin keskiarvoihin lyhenteillä $SM_1, SM_2, \dots, SM_{21}$, variansseihin lyhenteillä $SV_1, SV_2, \dots, SV_{21}$ ja vaihteluväleihin lyhenteillä $SR_1, SR_2, \dots, SR_{21}$.

Opetusaineiston jokaisesta aikasarjasta otettiin yhden havainnon satunnaisotos, jolloin aineistoon saadaan sadan havainnon satunnaisotos laitteiden elinkaarien eri vaiheilta. Tämä vastaa johdannaiskovariaatteja lukuunottamatta tosielämän tilannetta, jossa sadalle eri laitteelle tehdään yksittäiset mittaukset jossain niiden elinkaarien vaiheissa. Näin päädytään tilanteeseen, jossa lopullisessa opetusaineistossa on sata havaintoa 88 muuttujasta ja jokainen havainto on peräisin eri aikasarjasta.

6 Tulokset

Edellisessä kappaleessa kuvatulla tavalla saatu opetusaineisto standardoitiin ja sille tehtiin pääkomponenttianalyysi. Laskettujen otospääkomponenttien pistemääriin sovitettiin Weibull -nopeutetun viikaantumisajan malli käyttäen komponenttien valintaan kappaleessa 2.16 esitettyä Akaiken informaatiokriteeriin perustuvaa menettelyä. Mallissa estimoitavia parametrejä on kunkin pääkomponentin regressiokerroin, vakio, sekä skaalausparametri.

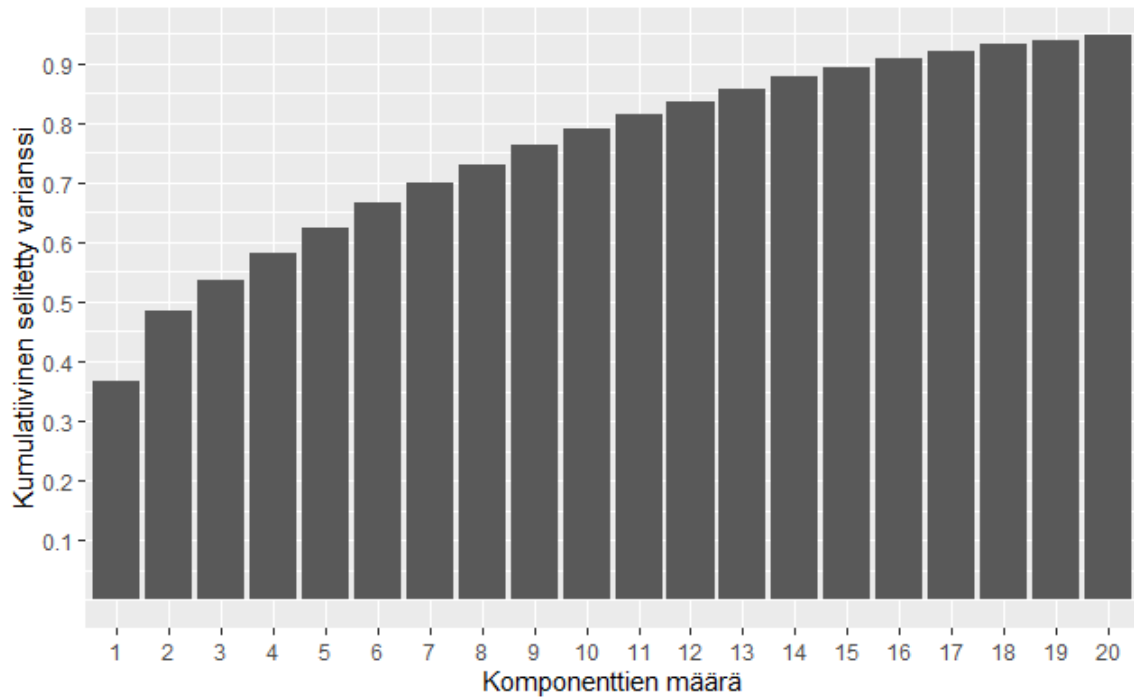
Testiaineistojen aikasarjojen viimeiselle havainnolle tehtiin samat muunnokset kuin opetusaineistojen aikasarjoille. Havainnot standardoitiin opetusaineiston keskiarvolla ja keskihajonnalla, jonka jälkeen laskettiin kyseisen osa-aineiston pääkomponenttianalyysin latauksia käyttäen pääkomponenttipistemäärät. Testiaineiston pääkomponenttipistemäärien pohjalta tehtiin kaavojen (4.15) ja (4.16) mukaiset piste-ennusteet laitteen jäljellä olevasta elinajasta ja niitä verrattiin todelliseen arvoon.

6.1 Pääkomponenttianalyysi

Tutkielmassa käytössä olleessa ensimmäisessä osa-aineistossa muuttujat $OP_3, S_1, S_5, S_6, S_{10}, S_{16}, S_{18}$ ja S_{19} ovat vakioita kaikissa aikasarjoissa, joten kyseisiä muuttujia tai niiden johdannaissuureita ei käytetä analyysissä. Käyttöön jää siis 14 sensorimittausta, $3 \times 14 = 42$ johdannaissuuretta sekä kaksi moottorin käyttöasetusmuuttujaa. Yhteensä käytettäviä kovariaatteja on 58. Kuvassa 6.1 on 20 ensimmäisen pääkomponentin kumulatiivisen selitetyn varianssin pylväsdiagrammi. Ensimmäinen pääkomponentti selittää lähes 37 % aineiston vaihtelusta ja toinen lähes 12 %. Yli 90 %:n selittävyys tarviin 16 ensimmäistä pääkomponenttia.

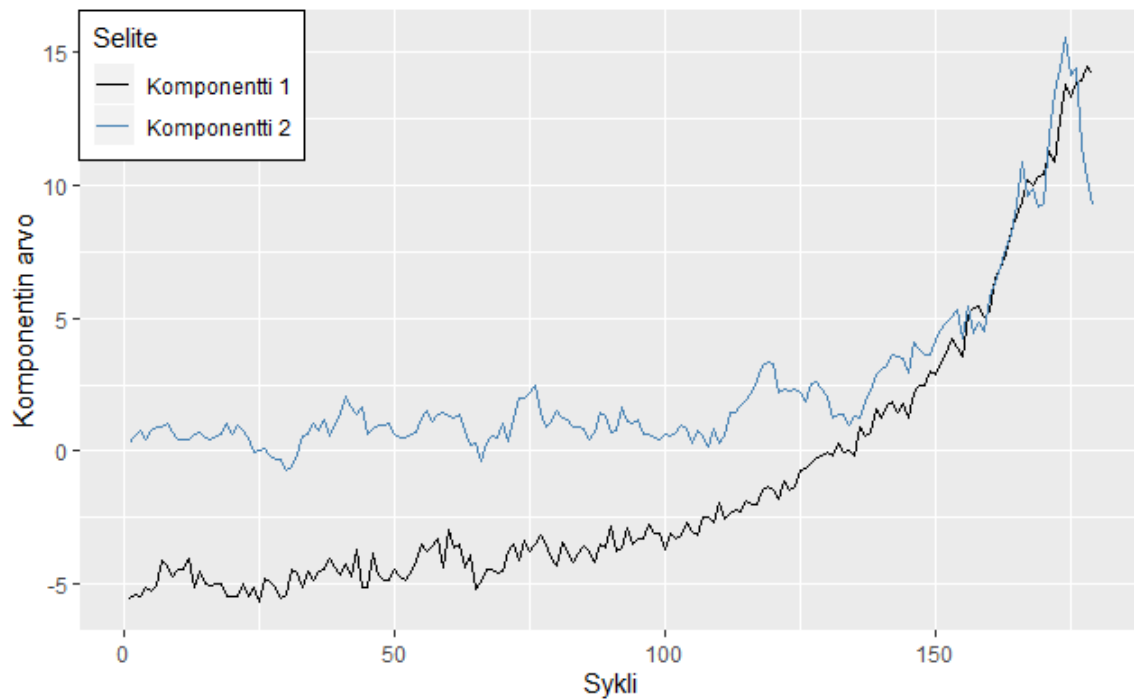
Kahden ensimmäisen pääkomponentin lataukset ovat liitteessä A. Liitteessä on taulukoitu vain lataukset, jotka ovat suurempia kuin 0.15. Muuttujat, joiden lataukset ovat molemmilla pääkomponenteilla alle 0.15 on jätetty pois taulukosta. Ensimmäisessä pääkomponentissa on lautoitunut vahvasti suuri osa sensorimuuttujista sekä niiden keskiarvoista. Toisessa pääkomponentissa vahvimmat lataukset ovat niillä sensoreilla, jotka eivät ole vahvasti latautuneita ensimmäisessä pääkomponentissa. Sama pätee sensoreiden keskiarvoihin, tosin SM_8 ja SM_{13} ovat lautoituneet vahvasti mutta päinvastaisesti molemmissa. Toisessa pääkomponentissa vahvasti ovat latautuneet myös osa sensoreiden liukuvista variansseista sekä liukuvista vaihteluväleistä.

Kuva 6.1. Kahdenkymmenen ensimmäisen pääkomponentin kumulatiivisen selitetyn varianssin pylväsdiagrammi



Kuvassa 6.2 on kahden ensimmäisen pääkomponentin viivakuvaaja yhdessä aikasarjoista. Viivakuvaajasta nähdään pääkomponenttien arvon alkavan kasvamaan jossain laitteen elinkaaren kohdassa, minkä jälkeen kasvu on likimain lineaarista laitteen hajoamiseen asti. Molemmat pääkomponentit ovat siis jollain tavalla sidoksissa laitteen kuntoon ja niiden pohjalta tehty regressiomallinnus on järkevää.

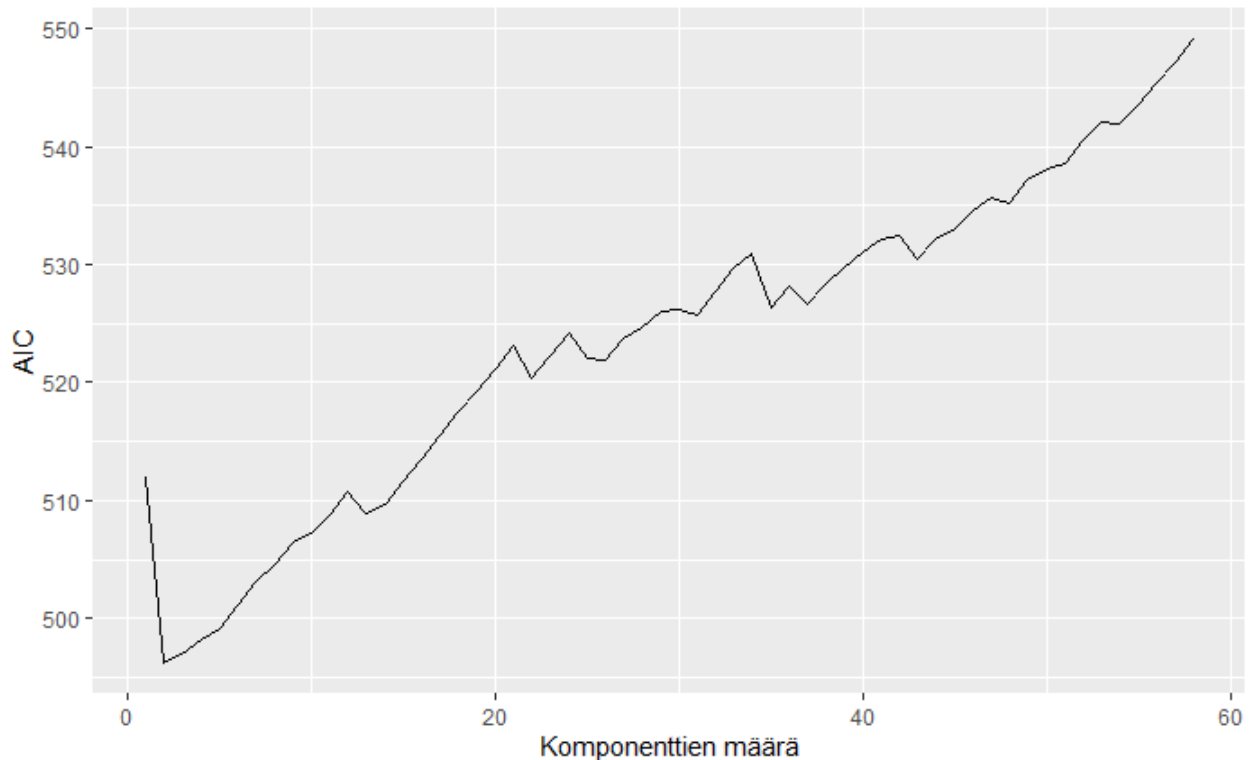
Kuva 6.2. Kahden ensimmäisen pääkomponentin viivakuvaaja yhdessä aikasarjassa



6.2 Malli

Akaiken informaatiokriteerin mukainen paras malli saavutettiin kahdella ensimmäisellä pääkomponentilla. AIC:n arvon ja pääkomponenttien määrän viivakuvaaja on kuvassa 6.3. Akaiken informaatiokriteerin arvo laskee merkittävästi, kun malliin lisätään toinen pääkomponentti, mutta seuraavien pääkomponenttien lisääminen ei kasvata uskottavuusfunktion arvoa merkittävästi, jolloin Akaiken informaatiokriteerin arvo nousee.

Kuva 6.3. Akaiken informaatiokriteerin arvo pääkomponenttien määrän suhteen



Mallin parametrien estimaatit, keskivirheet, Wald-testisuureet sekä testin p-arvot ovat taulukossa 6.1. Taulukossa skaalaparametrin σ sijasta vastaavat luvut on ilmoitettu sen logaritmille käytännön estimointisyyistä. Taulukosta nähdään, että kaikkien parametrien p-arvo on äärimmäisen pieni, joten kaikki parametrit mallissa ovat tilastollisesti erittäin merkitseviä.

Taulukko 6.1. Mallin parametriestimaatit, keskivirheet, Wald-testisuureet ja p-arvot

Parametri	Estimaatti	Keskivirhe	z	p-arvo
Vakio	4.611	0.0373	123.6	< 0.001
Komponentti 1	-0.1556	0.0098	-15.9	< 0.001
Komponentti 2	-0.0993	0.0127	-7.8	< 0.001
$\log \sigma$	-1.0459	0.0742	-14.1	< 0.001

Taulukossa 6.2 on regressiokertoimien kiihdytyskertoimet sekä uhkasuhteet. Yksikkömuutos ensimmäisessä pääkomponentissa kertoo hajoamisen hetkellistä todennäköisyyttä siis arvolla 1.5571 ja kiihdyttää laitteen jäljellä olevaa elinaikaa kertoimella 0.8559. Vastaavasti yksikkömuutos toisessa

pääkomponentissa kertoo hajoamisen hetkellistä todennäköisyyttä arvolla 1.3266 ja kiihdyttää laitteen elinikää kertoimella 0.9055.

Taulukko 6.2. Regressiokertoimien kiihdytyskerroimet ja uhkasuhteet

Parametri	Kiihdytyskerroin	Uhkasuhde
Komponentti 1	0.8559	1.5571
Komponentti 2	0.9055	1.3266

6.3 Ennustaminen

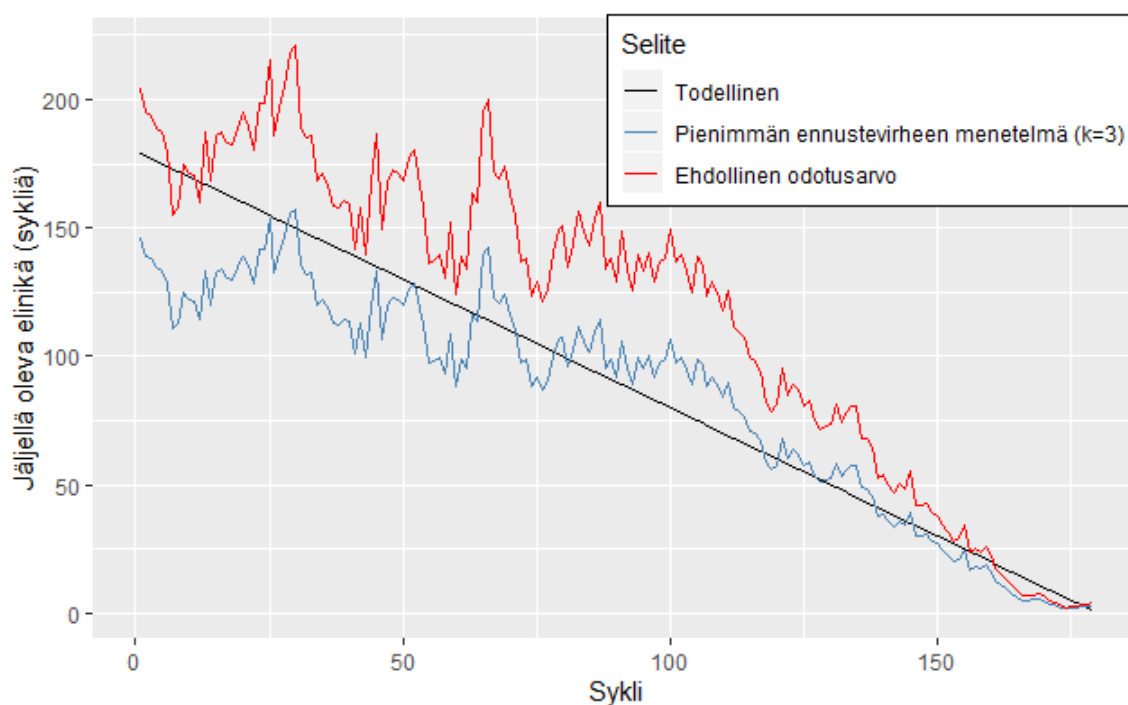
Taulukossa 6.3 on ennusteiden tuloksien RMSE- ja MAE-arvot kappaleessa 4.5.2 esitellyillä ennustemenetelmillä. Parhaat ennusteet saatiin pienimmän ennustevirheen menetelmällä käyttäen $k:n$ arvoa kolme. Tällöin saavutettiin hiukan yli 20 syklin keskimääräinen absoluuttinen virhe ja hiukan yli 26 keskineliövirhe.

Taulukko 6.3. Ennusteiden RMSE-arvot

Menetelmä	Kaava	k	RMSE	MAE
Ehdollinen odotusarvo	4.15		43.28	33.20
Pienimmän ennustevirheen piste-ennuste	4.16	2	35.87	27.07
		3	26.06	20.25

Kuvassa 6.4 on yhden opetusaineiston aikasarjan jokaiselle mittauspisteelle laskettujen ennusteiden viivakuvaaja. Kuvaaja on peräisin samasta aikasarjasta kuin kuvan 6.2 viivakuvaaja. Viivakuvaajassa on myös todellisen jäljellä olevan eliniän suora. Viivakuvaajasta nähdään pienimmän ennustevirheen menetelmän tuottavan keskimäärin tarkempia ennusteita kyseiselle aikasarjalle. Kuvasta on myös havaittavissa, että ennusteiden tarkkuus kasvaa laitteen lähestyessä elinkaarensa loppua.

Kuva 6.4. Ennusteiden viivakuvaaja yhdelle aikasarjalle



Taulukossa 6.4 on sadalle testiaikasarjalle tehdyn ennusteen RMSE-arvo ja varianssi todellisen jäljellä olevan eliniän luokissa. Taulukosta havaitaan, että pienillä todellisen jäljellä olevan eliniän arvoilla ennusteet ovat huomattavasti tarkempia kuin suurilla. Myös ennustevirheen varianssi kasvaa merkittävästi kun jäljellä oleva elinikä kasvaa. Malli kykenee siis ennustamaan lähitulevaisuudessa tapahtuvat viat paremmin kuin pidemmän ajan päästä tapahtuvat. Syyksi tähän arvellaan sitä, että vian täytyy olla suurudeltaan jo merkittävä ennen kuin se käy ilmi kahden ensimmäisen pääkomponentin pistemääristä. Tämä on havaittavissa myös kuvassa 6.2 olevasta pääkomponenttien arvon viivakuvaajasta.

Taulukko 6.4. Pienimmän ennustevirheen menetelmän ennusteiden RMSE-arvot todellisen eliniän luokissa

Todellinen arvo	Havaintoja	RMSE	Virheen varianssi
0–30	25	17.96	113.9
31–60	14	13.33	152.3
61–90	13	24.52	631.9
91–120	35	31.71	947.8
121–150	13	33.24	641.5

6.4 Tulosten vertailua

Tutkimuksessa käytettyä aineistoa on käytetty myös useissa lähteissä kirjallisuudessa etenkin neuroverkkojen ennustetarkkuuden tutkimiseen. Malli tuotti yksinkertaisuutensa nähden vertailukelpoisia tuloksia ennusteissa kirjallisuuteen nähden. Ennusteet olivat tarkkoja pienillä jäljellä olevan eliniän

arvoilla, mutta ennustetarkkuus pienenee, kun vikaan on pidempi aika. Osittain tästä syystä mallin ennustetarkkuus ei pääse täysin samalle tasolle useiden laskennallisesti raskaiden menetelmien kanssa. Babu et. al. (2016) saavuttivat konvulaationeuroverkolla 18.45 RMSE-arvon käyttäen ennusteisiinsa koko aikasarjaa satunnaisotoksen sijaan. Bektas et. al (2017) saivat tulokseksi 18.237 RMSE-arvon käyttäen vikaindeksiregressiotekniikoita. Zheng et. al. (2017) pääsivät LSTM-neuroverkolla 16.14 RMSE-arvoon ja Li et. al (2018) jopa 12.61 RMSE-arvoon.

Esitettyä mallinnustapaa voisi parantaa ottamalla käyttöön ajassa muuttuvien kovariaattien elinai-
kamalleja, jolloin aikasarjoja pystyttäisiin hyödyntämään mallinnuksessa kokonaan satunnaisotoksen
ottamisen sijaan. Myös uusien ja kuvaavimpien johdannaiskovariaattien tai vikaindeksien käyttäminen
mallin kovariaatteina voisi parantaa mallin osuvuutta ja ennustetarkkuutta.

Lähteet

- Babu G. S., Zhao P. ja Li X. (2016). *Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life*. Springer International Publishing.
- Bektas, O., Alfudail A. ja Jones J.A (2017). *Reducing Dimensionality of Multi-regime Data for Failure Prognostics*. Journal of Failure Analysis and Prevention 17, s. 1268–1275.
- Collett D. (2003). *Modelling Survival Data in Medical Research*. 2. painos. Texts in Statistical Science. CRC Press.
- Henderson R., Jones M. ja Stare J. (2001). *Accuracy of Point Predictions in Survival Analysis*. Statistics in Medicine 20. s. 3083-3096.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. 2. painos. Springer Series in Statistics, Springer-Verlag New York.
- Kleinbaum, D. ja Klein M. (2012). *Survival Analysis: A Self-Learning Text*. 3. painos. Springer New York.
- Lucht P. (2016). *The Method of Lagrange Multipliers* Rimrock Digital Technology, Technical Report.
- Li Z., Wang K. ja He Y. (2016). *Industry 4.0 - Potentials for Predictive Maintenance*. Atlantis Press. Advances in Economics, Business and Management Research 24. s. 42-46.
- Li X., Ding Q. ja Sun J. Q. (2018). *Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks*. Reliability Engineering & System Safety 172, s. 1-11.
- Ma, Z. S. ja Krings, A. W. (2008). *Survival Analysis Approach to Reliability, Survivability and Prognostics and Health Management (PHM)*. 29th IEEE-AIAA AeroSpace Conference, s. 20.
- McCool, J. I. (2012). *Using the Weibull Distribution. Reliability, Modeling, and Inference*. Wiley Series in Probability and Statistics, Wiley.
- Muldowney, P., Ostaszewski, K. ja Wojdowski, W. (2012). *The Darth Vader Rule*. Tatra Mountain Mathematical Publications.
- Saxena A. ja Goebel K. (2008). *Turbofan Engine Degradation Simulation Data Set*. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA. URL: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository>, viitattu 9.4.2020.
- Saxena A. et al. (2008). *Damage propagation modeling for aircraft engine run-to-failure simulation*. 2008 International Conference on Prognostics and Health Management.
- Zheng S., Ristovski K., Farahat A. ja Gupta C. (2017) *Long short-term memory network for remaining useful life estimation*. 2017 IEEE International Conference on Prognostics and Health Management, s. 88-95.

Liitteet

A Kahden ensimmäisen pääkomponentin lataukset

Latauksista on näkyvillä ne, joiden arvo on suurempi kuin 0.15. Muuttujia, joiden lataukset ovat pienempiä kuin 0.15 molemmille pääkomponenteille ei ole otettu mukaan taulukkoon.

Muuttuja	Komponentti1	Komponentti 2
S2	0.181	
S4	0.195	
S7	-0.185	
S8	0.171	-0.171
S9		0.302
S11	0.199	
S12	-0.191	
S13	0.164	-0.166
S14		0.312
S15	0.186	
S17	0.173	
S20	-0.177	
S21	-0.189	
SM2	0.209	
SM3	0.209	
SM4	0.212	
SM7	-0.209	
SM8	0.185	-0.180
SM9		0.301
SM11	0.212	
SM12	-0.210	
SM13	0.186	0.175
SM14		0.318
SM15	0.212	
SM17	0.210	
SM20	-0.210	
SM21	-0.212	
SV9		0.273
SV11		0.212
SV14		0.259
SR9		0.248
SR11		0.181
SR14		0.224