

Myyräkuumeen ja myyrärunsauden välisen suhteen mallintaminen tila-avaruusmalleilla

Olli Leppänen

Tilastotieteen pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
25. elokuuta 2016

JYVÄSKYLÄN YLIOPISTO
Matematiikan ja tilastotieteen laitos

Leppänen, Olli: *Myyräkuumeen ja myyrärunsauden välisen suhteen mallintaminen tila-avaruusmalleilla*

Tilastotieteen pro gradu -tutkielma, 29 s
25. elokuuta 2016

Tiivistelmä

Tutkielma käsittelee myyräkuumeen mallintamista tila-avaruusmalleilla. Tutkin myyrärunsauksien ja myyräkuumetapausten välistä riippuvuutta ja selvitän, voidaanko myyrärunsauksilla ennustaa tulevia myyräkuumetapausten lukumääriä. Sen lisäksi selvitän, kuinka monen kuukauden viiveellä myyräkuumetapaukset ilmenevät suhteessa myyrärunsauksiin. Sekä myyrärunsaudet että myyräkuumetapausten lukumäärät vaihtelevat vuodenajan mukaan.

Myyräkuumetapausten aineisto on kerätty Terveyden ja hyvinvoinnin laitoksen tartuntatautirekisterin tilastotietokannasta ja myyrärunsausaineisto on saatu Jyväskylän yliopiston Bio- ja ympäristötieteiden laitokselta. Molempien aikasarjojen alkamisvuosi on 1995. Myyrärunsauksia on vuoteen 2014 asti ja myyräkuumetapauksia on sen lisäksi vielä vuodelta 2015. Alueellisesti rajoitutaan tarkastelemaan vain Keski-Suomen, Pohjois-Savon ja Etelä-Savon sairaanhoitopiirien tartuntatapauksia.

Aiemmin myyräkuumetapausten ja myyrärunsauksien välistä riippuvuutta on mallinnettu yleistetyillä lineaarisilla sekamalleilla. Tila-avaruusmallit soveltuvat kuitenkin monipuolisemmin ajassa vaihtelevan kehityksen mallintamiseen.

Tutkielmani antaa viitteitä siitä, että myyräkuumetapaukset ilmenevät viiden kuukauden viiveellä metsämyyrärunsauksien suhteen. Käyttämäni mallin ennustekyky on Keski-Suomen osalta epätarkka, mutta Pohjois- ja Etelä-Savon myyräkuumetapausten ennustaminen onnistuu melko hyvin kuusi kuukautta eteenpäin.

Avainsanat: aikasarja-analyysi, tila-avaruusmalli, Poisson-jakauma, tärkeysotanta, interpolointi, myyräkuume

Kiitokset

Haluan kiittää Jyväskylän yliopiston Bio- ja ympäristötieteiden laitoksen tutkijoita Esa Koskelaa ja Tapio Mappesta haastavan myyräaineiston antamisesta käyttööni. Lisäksi haluan kiittää työni ohjaajaa professori Jukka Nyblomia oivallisesta ohjauksesta ja avusta koko prosessin aikana. Myös Jouni Helskelle kuuluu kiitokset KFAS-pakettiin perehdyttämisestä.

Sisältö

1	Johdanto	1
2	Tila-avaruusmallit	2
2.1	Lineaarinen gaussinen tila-avaruusmalli	2
2.2	Kalmanin suodin	3
2.2.1	Suodatus ja tasoitus	3
2.2.2	Puuttuvat havainnot	4
2.2.3	Uusien havaintojen ennustaminen	4
2.3	Selittäjämuuttujan lisääminen tila-avaruusmalliin	5
2.4	Stokastinen sykli	6
2.5	Tuntemattomien parametrien estimointi	7
2.6	Tärkeysotanta	8
2.7	Eksponentiaalinen perhe	10
2.8	Parametrien estimointi Poisson-jakauman tapauksessa	10
3	Aineistot ja ennustaminen	11
3.1	Taustaa ja aineistot	11
3.1.1	Myyräkuumeen taustasta	11
3.1.2	Myyräaineisto	13
3.1.3	Myyräkuumeaineisto	13
3.2	KFAS -paketti R-ympäristössä	16
3.3	Myyrärunsauksien interpolointi	17
3.3.1	Myyrärunsauksien interpolointimalli	17
3.3.2	Myyräaineistoon sovitettu malli	18
3.4	Myyräkuumetapausten ennustaminen	20
3.4.1	Viiveen määrittäminen	20
3.4.2	Mallinnus	22
4	Yhteenveto	27

1 Johdanto

Ihminen voi saada elimistönsä haitallisia viruksia luonnosta. Tällaisiin viruksiin kuuluu muun muassa myyräkuumetta aiheuttava Puumala-virus. Myyräkuumeen ihminen voi saada Puumala-virusta kantavalta metsämyyrältä ja pahimmillaan tartunta voi olla hengenvaarallinen. Myyräkuumetapausten lukumäärän on havaittu riippuvan viisi kuukautta aiemmasta metsämyyräpopulaation koosta. On myös havaittu, että sekä myyräkuumetapausten lukumäärät että metsämyyräpopulaation koko vaihtelevat syklisesti ajan suhteen (Kallio et al., 2009). Tartuntatapausten ymmärtäminen ja tulevien tartuntatapausten ennustaminen vaatii kuitenkin lisätutkimusta.

Tutkielmassani perehdyn myyräkuumetapausten mallintamiseen ja ennustamiseen myyrärunsauksien avulla Keski-Suomen, Pohjois-Savon ja Etelä-Savon sairaanhoitopiireissä. Sekä myyrärunsauksista että myyräkuumetapauksista on kerätty havaintoja vuosilta 1995–2014 ja ennustaminen suoritetaan vuodelle 2015 kuukausitasolla. Myyräaineisto on poimittu Konneveden tutkimusaseman läheisyydestä Keski-Suomesta.

Myyrärunsauksien ja myyräkuumetapausten välisen riippuvuuden mallintaminen tapahtuu tila-avaruusmallien avulla. Tila-avaruusmallit ovat monipuolisia malleja, joilla voidaan mallintaa ajassa tapahtuvaa kehitystä ja interpoloida puuttuvia havaintoja. Tutkielmassani käydään hieman läpi tila-avaruusmallien teoriaa tilanteessa, jossa havainnot noudattavat normaalijakaumaa. Tämän lisäksi esitellään, kuinka menettely eroaa kun havainnot noudattavat Poisson-jakauma. Tila-avaruusmalleja sovelletaan myyräaineistoon, joka sisältää myyrien lukumääriä neljältä kuukaudelta vuodesta. Puuttuvien kuukausien myyrämäärät pystytään interpoloimaan Kalmanin suotimen avulla. Interpoloituja myyrämääriä käytetään myyräkuumetapausten ennustamiseen.

Tila-avaruusmalleilla voidaan ratkaista monia aikasarjoihin liittyviä ongelmia. Tutkielmassani tila-avaruusmalleja käytetään aikasarjojen suodatukseen, tasoitukseen ja ennustamiseen. Oletuksena niissä on, että tutkittava ilmiö muuttuu ajan suhteen havaitsemattomien tilojen kautta. Nämä tilat liittyvät varsinaiseen havaittavaan aikasarjaan. Tilojen ja havaintojen välinen suhde välittyy tila-avaruusmallissa.

2 Tila-avaruusmallit

Aikasarja koostuu ajallisesti järjestyksessä olevista havainnoista y_1, y_2, \dots, y_n , jotka riippuvat toisistaan. Tyypillisesti peräkkäisten havaintojen välinen aikaväli on vakio. Usein kuitenkin aineisto sisältää puuttuvuutta, jolloin näin ei ole. Havaintojen puuttuminen voi aiheuttaa ongelmia, erityisesti jos puuttuvuutta on paljon. Tila-avaruusmallien tapauksessa puuttuvat havainnot voidaan interpoloida Kalmanin suotimella oletetun mallin avulla. Kalmanin suodin on esitelty tarkemmin alaluvussa 2.2. Tässä luvussa pääasiallisena lähteenä on kirja *Time Series Analysis by State Space Methods* (Durbin and Koopman, 2001).

2.1 Lineaarinen gaussinen tila-avaruusmalli

Merkinnällä $\mathcal{N}(\mu, \Sigma)$ tarkoitetaan multinormaalijakaumaa, jonka odotusarvo on μ ja kovarianssimatriisi on Σ . Lineaarinen gaussinen tila-avaruusmalli voidaan yleisesti kirjoittaa muodossa

$$\begin{aligned}y_t &= Z_t \alpha_t + \varepsilon_t, \\ \alpha_{t+1} &= T_t \alpha_t + R_t \nu_t,\end{aligned}\tag{1}$$

missä $t = 1, \dots, n$, $\varepsilon_t \sim \mathcal{N}(0, H_t)$ ja $\nu_t \sim \mathcal{N}(0, Q_t)$. Havaintovektori, y_t , on kokoa $p \times 1$ ja α_t on havaitsematon $m \times 1$ tilavektori. Tiloja, α_t , ei havaita suoraan ja ne ovat näin ollen tuntemattomia, toisin kuin havainnot, jotka ovat puuttuvia havaintoja lukuunottamatta tiedossa. Ensimmäinen yhtälö mallissa (1) on havaintoyhtälö. Se kuvaa sitä, miten tilat kuvautuvat havainnoiksi. Toinen yhtälö on tilayhtälö ja kuvaa tilojen kehittymistä ajan suhteen. Matriisit Z_t, T_t, R_t sekä kovarianssimatriisit H_t ja Q_t riippuvat valitusta mallista. Ne ovat usein myös ajan suhteen riippumattomia. Edellä mainitut matriisit sisältävät usein tuntemattomia parametreja, jotka täytyy estimoida. Virhetermit ε_t ja ν_t oletetaan toisistaan riippumattomiksi kaikilla ajanhetkillä ja niiden oletetaan myös olevan riippumattomia aiemmista virhetermeistä.

Tila-avaruusmalleja voidaan soveltaa laajalti aikasarja-analyysiin. Niitä voidaan käyttää muun muassa suodatukseen, tasoitukseen ja ennustamiseen. Suodatuksella tarkoitetaan tilojen jakaumien estimointia ehdolla aiemmat havainnot. Tasoituksessa tiloja estimoitaessa huomioidaan kaikki havainnot. Tilannetta, jossa halutaan estimoida tilojen jakaumat ajanhetkelle $n + h$, missä $h = 0, 1, 2, \dots$ ja n on aikasarjan viimeinen havainto, sanotaan ennustamiseksi. Tila-avaruusmalleissa oletetaan, että tutkimuksen kohteena olevaa prosessia voidaan kuvata tiloilla, $\alpha_1, \alpha_2, \dots, \alpha_n$. Nämä tilat kuvastavat tutkittavan ilmiön kehittymistä ajan suhteen. Tavoitteena on saada estimoitua tilat havaintojen avulla ja ennustaa aikasarjalle jatkoa.

2.2 Kalmanin suodin

Tila-avaruusmallinnuksessa päätavoitteena on saada tietoa latenteista tiloista ehdolla havainnot. Kun tila-avaruusmallin matriisit on määritelty valitun mallin mukaisiksi, on mahdollista käyttää Kalmanin suodinta. Kalmanin suodin on rekursiivinen algoritmi, jolla voidaan estimoida tila-avaruusmallin tuntemattomat tilat. Suodatuksella saadaan tilojen arvot ehdolla aiemmat havainnot. Ensimmäistä tilaa edeltäviä havaintoja ei ole, joten ensimmäinen tila täytyy alustaa. Tämä voidaan tehdä olettamalla $\alpha_1 \sim \mathcal{N}(a_1, P_1)$, mikä on ensimmäisen tilan a priori-jakauma. Epäinformatiivisuutta voidaan korostaa valitsemalla $P_1 = \kappa I$, missä I on yksikkömatriisi ja $\kappa \rightarrow \infty$. Usein kuitenkin käytännön syistä κ :n tilalle valitaan jokin suuri luku, esimerkiksi 10^7 .

2.2.1 Suodatus ja tasoitus

Suodatuksessa on tavoitteena saada tilojen α_t , $t = 1, \dots, n$, jakaumat ajanhetkellä t käyttäen hyväksi aiempia havaintoja. Tilojen ennusteet ja ennustevirheet lasketaan yksi kerrallaan. Käytetään hyväksi tietoa, että ehdollinen odotusarvo minimoi keskineliövirheen ja on näin ollen paras ennuste (Rao, 1973). Ennusteet ja ennustevirheet, ajanhetkellä $t = 1, \dots, n$, ovat

$$\begin{aligned} a_{t+1} &= \mathbb{E}(\alpha_{t+1} \mid y_t, \dots, y_1) = T_t(a_t + K_t F_t^{-1} v_t), \\ v_t &= y_t - Z_t a_t. \end{aligned}$$

Vastaavat kovarianssimatriisit ovat

$$\begin{aligned} P_{t+1} &= \text{Var}(\alpha_{t+1} \mid y_t, \dots, y_1), \\ F_t &= \text{Var}(v_t) = Z_t P_t Z_t' + H_t. \end{aligned}$$

Muut yleisen tila-avaruusmallin suodatuksen yhtälöt ovat

$$\begin{aligned} K_t &= P_t Z_t', \\ P_{t+1} &= T_t(P_t - K_t F_t^{-1} K_t') T_t' + R_t Q_t R_t. \end{aligned}$$

Kun suodatus on käyty loppuun, voidaan käyttää tasoitusta. Tilojen tasoi-tukset ja niiden varianssit

$$\begin{aligned} \hat{\alpha}_t &= \mathbb{E}(\alpha_t \mid y_n, \dots, y_1), \\ V_t &= \text{Var}(\alpha_t \mid y_n, \dots, y_1), \end{aligned}$$

saadaan käyttämällä Kalmanin suotimella saatuja arvoja a_t ja v_t . Tasoitetut arvot sopivat aineistoon suodatettuja paremmin, koska niitä estimoidaessa käytetään enemmän informaatiota. Tasoituksiin liittyvät rekursioalgoritmit löytyvät teoksesta Durbin and Koopman (2001).

2.2.2 Puuttuvat havainnot

Aineisto voi sisältää puuttuvia havaintoja, mikä voi johtua useasta eri syystä. Koska aikasarjoissa havainnot riippuvat toisistaan ajan suhteen, niin puuttuvien havaintojen korvaaminen tapahtuu estimoidun mallin avulla ja mallin oletuksien puitteissa. Tutkielmassani käyttämäni myyräaineisto koostuu kahdenkymmenen vuoden aikasarjasta, jossa kultakin vuodelta on havainnot neljältä kuukaudelta. Havainnot ovat aina samoilta neljältä kuukaudelta, jolloin loppuilta kuukausilta ei ole lainkaan tietoa. Erityisesti loppuvuoden ja loppukevään välillä ei myyrärunsauksia tunneta, mikä tarkoittaa että joka vuodelta on puoli vuotta yhtäjaksoisesti aikaa, joilta myyrärunsauksista ei ole tietoa. Myyräpopulaation koon suhteen oletetaan, että se vaihtelee vuoden sisällä suuresti siten, että myyriä on eniten loppukesällä ja syksyllä ja vähiten alkuvuodesta. Puuttuvat havainnot interpoloidaan käyttäen kyseistä oletusta myyräpopulaation koon syklistä vaihtelusta.

Oletetaan, että havainnot y_t , missä $t = \tau, \dots, \tau^*$ puuttuvat. Kalmanin suotimessa vektori v_t ja matriisi K_t asetetaan nolliksi puuttuvien havaintojen kohdalla. Tällöin Kalmanin suotimen päivitykset muuttuvat seuraavanlaisiksi

$$a_{t+1} = T_t a_t, \quad P_{t+1} = T_t P_t T_t' + R_t Q_t R_t, \quad t = \tau, \dots, \tau^* - 1.$$

Tasoituksen rekursiokaavojen suhteen on vastaavanlaisia muutoksia. Voi myös olla, että vain osa havainnoista puuttuu jollakin ajanhetkellä. Oletetaan, että ajanhetkellä t osa y_t :n alkioista puuttuu. Tällöin merkitään havaintovektoria $y_t^* = W_t y_t$, missä matriisi W_t on tunnettu. Sen rivit ovat osajoukko yksikkömatriisista I_p . Havaintoyhtälö (1) muuttuu puuttuvien havaintojen ajakohtien suhteen seuraavasti

$$y_t^* = Z_t^* \alpha_t + \epsilon_t^*, \quad \epsilon_t^* \sim \mathcal{N}(0, H_t),$$

missä $Z_t^* = W_t Z_t$, virhetermit sisältävä vektori $\epsilon_t^* = W_t \epsilon_t$ ja kovarianssimatriisi $H_t^* = W_t H_t W_t'$. Tämän jälkeen Kalmanin suodin ja tasointo toimivat samoin kuin ilman puuttuvia havaintoja, kunhan havaintovektori y_t , matriisi Z_t ja kovarianssimatriisi H_t korvataan vektorilla y_t^* , matriisilla Z_t^* ja H_t^* . Puuttuvien havaintojen helppo käsittely tekee tila-avaruusmallien käyttämisestä houkuttelevaa. (Durbin and Koopman, 2001, s.92.)

2.2.3 Uusien havaintojen ennustaminen

Aikasarjoissa on usein mielenkiinnonkohteena prosessin ja tulevien havaintojen ennustaminen. Olkoon havainnot y_1, \dots, y_n aikasarja, joka voidaan mal-

lintaa tila-avaruusmallilla (1). Tarkoituksena on ennustaa seuraavat J kappaletta havaintoja y_{t+j} , missä $j = 1, \dots, J$. Olkoon estimaatti \bar{y}_{n+j} sellainen, jolla on pienin keskineliövirhematriisi ehdolla havainnot Y_n . Kuten edellä alaluvussa 2.2.1 on mainittu, havainnoilla ehdollistettu kovarianssimatriisi $F_t = \mathbb{E}[(\bar{y}_{n+j} - y_{n+j})(\bar{y}_{n+j} - y_{n+j})' | y]$ on minimissään, kun valitaan $\bar{y}_{n+j} = \mathbb{E}(y_{n+j} | y)$. Tästä saadaan ensimmäinen ennuste ajanhetkelle $t = n + 1$ suoraviivaisesti. Nyt uusi havainto on $y_{n+1} = Z_{n+1}\alpha_{n+1} + \epsilon_{n+1}$, joten sen ehdollinen keskiarvo on

$$\bar{y}_{n+1} = Z_{n+1}\mathbb{E}(\alpha_{n+1}|y), = Z_{n+1}a_{n+1}. \quad (2)$$

ja ensimmäisen askeleen ennusteiden virhevarianssimatriisi on

$$\begin{aligned} \bar{F}_{n+1} &= \mathbb{E}(\bar{y}_{n+1} - y_{n+1})(\bar{y}_{n+1} - y_{n+1})' \\ &= Z_{n+1}P_{n+1}Z_{n+1}' + H_{n+1}. \end{aligned} \quad (3)$$

Kaavan (2) odotusarvo a_{n+1} ja kaavan (3) virhevarianssimatriisi saadaan Kalmanin suotimen avulla kuten alaluvussa 2.2.1 sanotaan. Loput ennustetut arvot saadaan siten, että niitä käsitellään puuttuvina havaintoina. (Durbin and Koopman, 2001, s.93-94.)

2.3 Selittäjämuuttujan lisääminen tila-avaruusmalliin

Tila-avaruusmalliin voidaan sisällyttää prediktorimuuttujia, joiden arvot ovat matriiseissa X_t . Tällöin nähdään miten aikasarjan keskimääräinen taso muuttuu selittäjämuuttujan arvo muuttuessa. Tämänkaltaisessa regressiomallissa havaintojen, Y_t , oletetaan tällöin korreloivan näiden prediktoreiden kanssa. Tila-avaruusmalli voidaan muotoilla siten, että prediktorien kulmakerroin, β , ei riipu ajasta. Yksiulotteisessa tapauksessa havaintoyhtälö on muotoa

$$\begin{aligned} y_t &= [Z_t \quad X_t] \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + \epsilon_t \\ &= Z_t\alpha_t + X_t\beta_t + \epsilon_t \end{aligned}$$

ja tilayhtälöksi saadaan

$$\begin{pmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{pmatrix} = \begin{bmatrix} T_t & 0 \\ 0 & I_k \end{bmatrix} \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + \begin{bmatrix} R_t \\ 0 \end{bmatrix} \eta_t. \quad (4)$$

Molemmissa yhtälöissä $t = 1, \dots, n$. Tilayhtälöstä (4) nähdään, että koska regressiokertoimien (β) kohinatermiä kerrotaan nolllalla, pysyy kerroin kulakin ajanhetkellä samana ja ei näin ollen riipu ajasta. Alkutilanteessa β_1

voidaan valita diffuusina tai sen jakaumaksi voidaan valita jokin normaali-jakauma. Diffuusissa tapauksessa komponenteille β_1 ja α_1 alkujakaumiksi voidaan asettaa

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} \sim \mathcal{N} \left\{ \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}, P_1 \right\},$$

missä $P_1 = \begin{bmatrix} \kappa P_\infty + P_* & 0 \\ 0 & \kappa I_k \end{bmatrix}$ ja $\kappa \rightarrow \infty$. Usein vektori a_1 asetetaan nollavektoriksi, mikäli ei ole muuta informaatiota sen arvosta. Matriisin P_1 sisällä olevat matriisit P_∞ ja P_* määräytyvät käytettyjen tilakomponenttien mukaan. Tässä tutkielmassa käytetään stokastista sykli-komponenttia ja yksittäiselle syklille kyseiset matriisit ovat

$$P_\infty = \text{diag}[1, 1] \quad \text{ja} \quad P_* = \text{diag}[0, 0].$$

Kalmanin suotimen alustuksesta voi lukea tarkemmin teoksesta Durbin and Koopman (2001, Luku 5).

2.4 Stokastinen sykli

Tila-avaruusmalliin voidaan lisätä kausivaihtelun kaltainen syklikomponentti. Sykli kuvaa aikasarjassa olevaa jaksottaista vaihtelua. Syklistä vaihtelua on esimerkiksi lämpötila vuodenajan suhteen, kun talvella on kylmintä ja kesällä lämpimintä. Tällöin syklin pituus, s , on 12. Sykli joka säilyy ajan suhteen samanlaisena on deterministinen. Deterministinen sykli on muotoa

$$c_t = \alpha \cos(\lambda t) + \beta \sin(\lambda t). \quad (5)$$

Kaavassa (5) esiintyvät termit α ja β ovat vakioita ja frekvenssi $\lambda = 2\pi/s$, missä s on syklin pituus. Kyse on syklistä, jos $c_t = c_{t+s}$, sillä kosini- ja sini-funktioiden syklin pituus on 2π , mistä seuraa

$$\begin{aligned} c_{t+s} &= \alpha \cos\left(\frac{2\pi t}{s} + 2\pi\right) + \beta \sin\left(\frac{2\pi t}{s} + 2\pi\right) \\ &= \alpha \cos\left(\frac{2\pi t}{s}\right) + \beta \sin\left(\frac{2\pi t}{s}\right) = c_t. \end{aligned}$$

Syklikomponentteja tulee tila-avaruusmalliin yhtä sykliä kohden vain yksi varsinainen parametri. Tämän lisäksi syklille tarvitaan pseudosyklikomponentti, jota käytetään vain varsinaisen syklikomponentin saamiseksi. Kaavan (5) sykli toteuttaa yhtälön

$$\begin{bmatrix} c_{t+1} \\ c_{t+1}^* \end{bmatrix} = \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} \begin{bmatrix} c_t \\ c_t^* \end{bmatrix}, \quad (6)$$

missä c_t^* on lisätty pseudosyklikomponentti. Kaava (6) ei vielä sisällä satunnaista vaihtelua. Lisäämällä sykliin kohina voidaan syklikomponentin ja pseudosyklikomponentin sisältävä vektori esittää muodossa

$$\begin{bmatrix} c_{t+1} \\ c_{t+1}^* \end{bmatrix} = \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} \begin{bmatrix} c_t \\ c_t^* \end{bmatrix} + \begin{bmatrix} \omega_t \\ \omega_t^* \end{bmatrix}. \quad (7)$$

Pseudoseudokomponentti c_t^* on vain syklikomponentin c_t rakennuksessa mukana. Viimeisessä vektorissa oleville satunnaistermeille ω_t :lle ja ω_t^* :lle oletetaan yleensä sama varianssi tai korreloimattomuus. Usein niiden osalta tehdään molemmat oletukset. (Harvey, 1993, s. 182, 183 ja 227.)

2.5 Tuntemattomien parametrien estimointi

Tila-avaruusmallit sisältävät tuntemattomia parametreja, jotka täytyy estimoida aineistosta. Estimoinnissa käytetään suurimman uskottavuuden menetelmää. Estimoitavia parametreja on erityisesti matriiseissa H_t ja Q_t .

Ensimmäisille tiloille α_1 oletetaan alkujakaumaksi normaalijakauma $\mathcal{N}(a_1, P_1)$, missä odotusarvo a_1 on tunnettu. Merkitään havaintoa y_t edeltäviä havaintoja $Y_{t-1} = \{y_1, \dots, y_{t-1}\}$, sekä tuntemattomia parametreja sisältävää vektoria Ψ :llä. Havaintojen ehdollinen uskottavuusfunktio ehdolla aiemmat havainnot on

$$L_g(y; \Psi) = p(y_1, y_2, \dots, y_n; \Psi) = p(y_1; \Psi) \prod_{t=2}^n p(y_t | Y_{t-1}; \Psi).$$

Yleisesti mallissa (1) havainnon y_t odotusarvo ehdolla aiemmat havainnot on

$$\mathbb{E}(y_t | Y_{t-1}; \Psi) = Z_t a_t.$$

Logaritmiseksi uskottavuudeksi gaussisessa tapauksessa saadaan

$$\begin{aligned} \log L_g(y; \Psi) &= \frac{np}{2} \log(2\pi) + \frac{1}{2} \sum_{t=1}^n \log |F_t|^{-1} \\ &\quad + \log \left[\exp \left[-\frac{n}{2} \sum_{t=1}^n ((y_t - Z_t a_t)' F_t^{-1} (y_t - Z_t a_t)) \right] \right] \\ &= \frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n (\log |F_t| + v_t' F_t^{-1} v_t), \end{aligned}$$

missä v_t ja F_t riippuvat parametreista Ψ . Havainnon ja sen odotusarvon erotus, v_t , ja havaintojen varianssi ehdolla aiemmat havainnot, F_t , saadaan

kuten luvussa 2.2.1 on esitelty. Matriisi F_t oletetaan kääntyväksi kaikilla ajanhetkillä $t = 1, \dots, n$. Suurimman uskottavuuden estimaattori on vektori, joka maksimoi funktion $\log L_g(y; \Psi)$. Uskottavuusfunktion maksimointiin on olemassa useita numeerisia algoritmeja. Tässä tutkielmassa käytetään R-ohjelmiston (R Core Team, 2016) KFAS-pakettia (Helske, 2015) tila-avaruusmallin rakentamiseen ja mallinnukseen.

2.6 Tärkeysotanta

Edellä on esitelty uskottavuusfunktio lineaarisessa gaussisessa tapauksessa. Mikäli havaintojen jakauma ei ole gaussinen, ei tuntemattomien parametrien estimointi onnistu yhtä suoraviivaisesti. Tällöin uskottavuusfunktiota approksimoidaan edellä olevan gaussisen uskottavuuden avulla. Merkinnöillä $g(\cdot)$, $g(\cdot | \cdot)$ ja $g(\cdot, \cdot)$ tarkoitetaan normaalijakauman tiheysfunktiota, yhteisjakauman tiheysfunktiota ja ehdollisen jakauman tiheysfunktiota. Vastavasti diskreetin funktion p tapauksessa tarkoitetaan pistetodennäköisyys-, ehdollista pistetodennäköisyys- ja yhteisjakauman pistetodennäköisyysfunktioita. Funktio p voi tärkeysotannassa olla myös jatkuva.

Kiinnostuksen kohteena on estimoida tiloihin $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ liittyvä ehdollinen odotusarvo

$$\bar{x} = \mathbb{E}(x(\boldsymbol{\alpha}) | y) = \int x(\boldsymbol{\alpha})p(\boldsymbol{\alpha} | y)d\boldsymbol{\alpha}. \quad (8)$$

Odotusarvoa \bar{x} voitaisiin estimoida poimimalla satunnaisotos jakaumasta, jonka tiheys on $p(\boldsymbol{\alpha} | y)$ ja laskea niiden avulla otoskeskiarvo tilojen funktiosta $x(\boldsymbol{\alpha})$. Käytännössä se on hankalaa, sillä tiheysfunktiota $p(\boldsymbol{\alpha} | y)$ ei ole mahdollista kirjoittaa eksplisiittisessä muodossa epälineaarille tila-avaruusmalleille. Jatkossa käytetään hyväksi yhtälöä

$$g(\boldsymbol{\alpha} | y) = \frac{g(\boldsymbol{\alpha}, y)}{g(y)}. \quad (9)$$

Odotusarvoa (8) voidaan approksimoida käyttämällä normaalijakauman tiheysfunktiota hyväksi. Valitaan gaussinen tärkeystiheys $g(\boldsymbol{\alpha} | y)$ annetulle vektorille Ψ siten, että se muistuttaa mahdollisimman läheisesti tiheys-/pistetodennäköisyysfunktiota $p(\boldsymbol{\alpha} | y)$.

Kerrotaan ja jaetaan funktiolla $g(\boldsymbol{\alpha} | y)$ integraalin (8) integroitava ja käytetään hyväksi yhtälöä (9). Ehdollinen jakauma $p(\boldsymbol{\alpha} | y)$ on laskennallisesti monimutkainen epälineaarille malleille toisin kuin vastaava yhteistiheys $p(\boldsymbol{\alpha}, y)$. Alla olevan yhtälön (10) toisen yhtäsuuruuden kohdalla käytetään

hyväksi yhtälöä (9). Tekemällä edellä esitetyt muokkaukset yhtälölle (8) saadaan se muotoon

$$\begin{aligned}\bar{x} &= \int x(\boldsymbol{\alpha}) \frac{p(\boldsymbol{\alpha} | y)}{g(\boldsymbol{\alpha} | y)} g(\boldsymbol{\alpha} | y) d\boldsymbol{\alpha} \\ &= \int x(\boldsymbol{\alpha}) \frac{p(\boldsymbol{\alpha}, y) g(y)}{g(\boldsymbol{\alpha}, y) p(y)} g(\boldsymbol{\alpha} | y) d\boldsymbol{\alpha} \\ &= \frac{g(y)}{p(y)} \mathbb{E}_g \left[x(\boldsymbol{\alpha}) \frac{p(\boldsymbol{\alpha}, y)}{g(\boldsymbol{\alpha}, y)} \right].\end{aligned}\tag{10}$$

Yhtälön (10) odotusarvo \mathbb{E}_g on tärkeystiheyden $g(\boldsymbol{\alpha} | y)$ suhteen. Asettamalla tilojen funktio vakioksi $x(\boldsymbol{\alpha}) = 1$ saadaan

$$1 = \frac{g(y)}{p(y)} \mathbb{E}_g \left[\frac{p(\boldsymbol{\alpha}, y)}{g(\boldsymbol{\alpha}, y)} \right].\tag{11}$$

Merkitsemällä tiheyksien osamäärää $w(\boldsymbol{\alpha}, y) = \frac{p(\boldsymbol{\alpha}, y)}{g(\boldsymbol{\alpha}, y)}$ ja ottamalla osamääräyllä olevista yhtälöistä (10) ja (11) saadaan

$$\bar{x} = \frac{\mathbb{E}_g [x(\boldsymbol{\alpha}) w(\boldsymbol{\alpha}, y)]}{\mathbb{E}_g [w(\boldsymbol{\alpha}, y)]}.\tag{12}$$

Tilanteessa, jossa havainnot noudattavat jotain muuta jakaumaa kuin normaalijakaumaa, mutta tilavektori generoidaan yhtälön (1) avulla, jolloin $p(\boldsymbol{\alpha}) = g(\boldsymbol{\alpha})$, saadaan painofunktioksi

$$w(\boldsymbol{\alpha}, y) = \frac{p(\boldsymbol{\alpha}, y)}{g(\boldsymbol{\alpha}, y)} = \frac{p(\boldsymbol{\alpha}) p(y | \boldsymbol{\alpha})}{g(\boldsymbol{\alpha}) g(y | \boldsymbol{\alpha})} = \frac{p(y | \boldsymbol{\alpha})}{g(y | \boldsymbol{\alpha})}.\tag{13}$$

Keskiarvoa \bar{x} voidaan estimoida Monte Carlo -menetelmällä seuraavasti, jolloin kaavan (12) nimittäjäksi saadaan

$$\mathbb{E}_g \left[\frac{p(y | \boldsymbol{\alpha})}{g(y | \boldsymbol{\alpha})} \right] \approx \frac{1}{N} \sum_{i=1}^N w_i = \frac{1}{N} \sum_{i=1}^N \frac{p(y | \alpha^i)}{g(y | \alpha^i)},\tag{14}$$

ja osoittajaksi

$$\mathbb{E}_g \left[x(\boldsymbol{\alpha}) \frac{p(y | \boldsymbol{\alpha})}{g(y | \boldsymbol{\alpha})} \right] \approx \frac{1}{N} \sum_{i=1}^N x(\alpha^i) w_i = \frac{1}{N} \sum_{i=1}^N x(\alpha^i) \frac{p(y | \alpha^i)}{g(y | \alpha^i)}.\tag{15}$$

Yhtälöissä (14) ja (15) esiintyvät tilat α^i arvotaan jakaumasta $g(\boldsymbol{\alpha} | y)$. Odotusarvon \bar{x} Monte Carlo estimaatin tarkkuuteen voidaan vaikuttaa simulointien määrällä sekä tärkeysjakauman $g(\cdot)$ valinnalla.

2.7 Eksponentiaalinen perhe

Tila-avaruusmallit voidaan laajentaa koskemaan myös eksponentiaalista perhettä, jolloin havaintojen jakauman tiheys on muotoa

$$p(y_t|\theta_t) = \exp [y_t\theta_t - b_t(\theta_t) + c_t(y_t)], \quad (16)$$

missä $b_t(\theta_t)$ on kahdesti derivoituva ja $c_t(y_t)$ on vain havaintojen y_t funktio. Eksponentiaalisen perheen tapauksessa havaintovektori ja tilavektori muistuttavat gaussista tapausta (1). Nyt yhtälön (1) havainto, y_t , korvataan signaalilla $\theta_t = Z_t\alpha_t$, joka on lineaarinen prediktori. Signaali, θ_t , on yhteydessä havaintojen odotusarvoon $\mathbb{E}(y_t) = \lambda_t$ linkkifunktion $l(\lambda_t) = \theta_t$ kautta. Koska linkkifunktio ei ole lineaarinen, niin suora muunnos aiheuttaisi harhaa. Tämän vuoksi käytetään tärkeysotantaan perustuvaa simulointimenetelmää. Yleensä ollaan kiinnostuneempia odotusarvosta λ_t kuin signaalista θ_t . Tilavektorit eivät riipu aiemmista havainnoista ja ne generoidaan vastaavanlaisella yhtälöllä kuin lineaarisen gaussisen mallin tapauksessa

$$\alpha_{t+1} = T_t\alpha_t + R_t\nu_t, \quad \nu_t \sim g(\nu) \quad (17)$$

missä $t = 1, 2, \dots, n$ ja virhetermit ν_t ovat ajallisesti riippumattomia. Virhetermien jakauman oletetaan useimmiten olevan normaalijakauma.

2.8 Parametrien estimointi Poisson-jakauman tapauksessa

Luvussa 2.1 on esitelty lineaarinen gaussinen tila-avaruusmalli. Havaintojen ollessa lukumääriä, voidaan käyttää Poisson-mallia. Tällöin havaintojen y_t odotusarvo λ_t on myös niiden varianssi. Tila-avaruusmalleissa Poisson-jakautuneelle vasteelle havaintoyhtälö ja tilayhtälö eroavat hieman yleisestä gaussisesta muodosta (1). Oletetaan havaintojen noudattavan Poisson-jakaumaa, jonka odotusarvo on $\lambda_t = e^{\theta_t} = e^{Z_t\alpha_t}$. Tiheysfunktio on tässä tapauksessa muotoa

$$\begin{aligned} p(y_t|\lambda_t) &= \frac{\lambda_t^{y_t}}{y_t!} e^{-\lambda_t} \\ &= \exp [y_t Z_t \alpha_t - \exp(Z_t \alpha_t) - \log y_t!]. \end{aligned} \quad (18)$$

Huomataan, että Poisson-jakauman tiheys (18) saadaan kaavasta (16) valitsemalla $\theta_t = Z_t\alpha$, $b_t(\theta_t) = \exp(Z_t\alpha_t)$ ja $c_t(y_t) = -\log(y_t!)$.

Alla olevan yhtälön (19) toisessa yhtäsuuruudessa käytetään hyväksi yhtälöä (9). Merkitään havaintojen uskottavuutta ehdolla tuntemattomat parametrit

$L(\Psi) = p(y|\Psi)$, jolloin

$$\begin{aligned}
L(\Psi) &= \int p(\boldsymbol{\alpha}, y) d\boldsymbol{\alpha} \\
&= \int \frac{p(\boldsymbol{\alpha}, y)}{g(\boldsymbol{\alpha} | y)} g(\boldsymbol{\alpha} | y) d\boldsymbol{\alpha} \\
&= g(y) \int \frac{p(\boldsymbol{\alpha}, y)}{g(\boldsymbol{\alpha}, y)} g(\boldsymbol{\alpha} | y) d\boldsymbol{\alpha} \\
&= L_g(\Psi) \mathbb{E}_g[w(\boldsymbol{\alpha}, y)],
\end{aligned} \tag{19}$$

missä $L_g(\Psi) = g(y)$ on gaussinen approksimaatio Poisson-mallin uskottavuudesta. Sitä käytetään tärkeystiheyden $g(\boldsymbol{\alpha} | y)$ saamiseksi. Odotusarvo \mathbb{E}_g otetaan ehdollisen normaalijakauman $g(\boldsymbol{\alpha} | y)$ suhteen ja painot w ovat kuten kaavassa (13). Tuntematonta parametria Ψ estimoidaan maksimoimalla uskottavuusfunktio $L(\Psi)$, jolloin saadaan suurimman uskottavuuden estimaatti $\hat{\Psi}$. Käytännössä kuitenkin logaritmisen uskottavuuden $\log L(\Psi)$ maksimointi on stabiilimpaa, sillä uskottavuusfunktio voi saada suuria arvoja. Lisäksi arvo, Ψ , joka maksimoi logaritmisen uskottavuuden $\log L(\Psi)$ maksimoi myös uskottavuusfunktion $L(\Psi)$. Logaritmiseksi uskottavuudeksi saadaan

$$\begin{aligned}
\log L(\Psi) &= \log \int p(\boldsymbol{\alpha}, y) d\boldsymbol{\alpha} \\
&= \log L_g(\Psi) + \log \mathbb{E}_g[w(\boldsymbol{\alpha}, y)].
\end{aligned} \tag{20}$$

Yhtälössä (20) olevat termit $L_g(\Psi)$ ja $\mathbb{E}_g[w(\boldsymbol{\alpha}, y)]$ saadaan kuten luvuissa 2.5 ja 2.6 on esitelty.

3 Aineistot ja ennustaminen

3.1 Taustaa ja aineistot

3.1.1 Myyräkuumeen taustasta

Myyräkuumetapauksia esiintyy vuosittain koko Suomen laajuisesti. Sen aiheuttaa Puumala-virus (PUUV), jonka ihminen voi saada elimistönsä metsämyyrän (*Myodes glareolus*) eritteistä. Myyräkuume (*nephropathia epidemica*) on munuaisoireinen verenvuotokuume, josta seuraa korkea kuume, päänsärkyä, pahoinvointia sekä vatsa- ja selkäkipuja. Kuolemantapaukset ovat kuitenkin erittäin harvinaisia. Myyräkuumetapausten on havaittu heijastavan metsämyyräpopulaatioiden aaltoilua. Kun metsämyyriä on runsaasti niin on myyräkuumetapauksiakin. Suomessa kuukausittaisia tartuntatapauksia

tilastoi Terveyden ja hyvinvoinnin laitos. (Terveyden ja hyvinvoinnin laitos; Heyman et al., 2001; Niklasson et al., 1995)

Myyräkuumeen aiheuttava Puumala-virus kuuluu hantavirusten muodostamaan sukuun, jotka kuuluvat *Bunyaviridae* -virusten heimoon. Hantavirusta esiintyy useassa eri nisäkäslahkossa, mutta vain jyrsijältä saatuna ne aiheuttavat ihmisille tauteja. Hantavirukset edustavat jyrsijöiden kautta kulkevaa virusryhmää ja ne ovat levittäytyneet laajemmalle alueelle kuin mikään muu luonnonvarainen virus. Hantavirukset kiertävät luonnossa jyrsijässä ja kulkeutuvat vahingon kautta ihmisiin usein aiheuttaen vakavuudeltaan eritasoisia tauteja. Urosmetsämyyrien on havaittu aiheuttavan enemmän hantavirustartuntoja kuin naaraiden. PUUV-tartunta on sitä kantavalle metsämyyrälle oireeton ja pitkäaikainen. (McCaughey and Hart, 2000; Kallio et al., 2013, 2009; Brummer-Korvenkontio et al., 1980; Bernshtein et al., 1999)

Ihminen voi saada virustartunnan metsämyyrän eritteistä hengitysteitse. Kolmasosalla sairastuneista esiintyy ohimeneviä näköhäiriöitä. Ensimmäisellä sairastuneella voi ilmetä munuaisten vajaatoimintaa kuten virtsanmäärän vähenemistä ja veren kreatiinipitoisuuden suurentumista. Toipumisvaiheessa, joka on noin 1–2 viikkoa munuaisten vajaatoiminnan jälkeen, virtsanmäärät lisääntyvät normaalia suuremmiksi. Jälkitauteja myyräkuumeeseen liittyy harvoin. Taudin sairastettuaan henkilö saa suurella todennäköisyydellä elinikäisen vastustuskyvyn sitä vastaan. Myyräkuume voidaan todentaa tutkimalla Puumala-viruksen vasta-aineita verinäytteestä. Puumalavirus ei tartu ihmisistä toiseen ja varsinaista lääkettä myyräkuumeeseen ei ole. (Terveyden ja hyvinvoinnin laitos).

Kuten edellä on todettu, ihminen voi saada myyräkuumeen metsämyyrän eritteistä. Metsämyyrä on jyrsijä, jonka ruumis on 8–13 cm pitkä ja paino on 15–40 g. Suomessa metsämyyrää tavataan kaikkialla Lappia lukuunottamatta. Se viihtyy metsissä, metsien ja soiden reunoilla sekä hakkuualueilla. Metsämyyrät saavat 2–4 pesuetta vuodessa ja pesueissa on 3–9 poikasta. Sukukypsyyden ne saavuttavat 45–60 päivän ikäisinä. (Luontoportti, 2015.)

Euroopassa yleinen metsämyyrä on Puumala hantaviruksen kantaja. Pohjoisessa Fennoskandiassa metsämyyräpopulaatiossa on vuosittaista jaksollista vaihtelua. Kyseiset vaihtelut ovat kolmesta viiteen vuoteen. Korkean jyrsijärunsauden oletetaan ja ennustetaan usein lisäävän jyrsijöiltä saatavien tautitartuntojen tapauksia ihmisissä. Suoria todisteita tästä suhteesta on vähän. Suhdetta selitetään oletuksella, että suurempi isäntätiheys kasvattaa sekä tartunnan saaneiden yksilöiden määrää että tartunnan levinneisyyttä. (Olsson et al., 2005; Kallio et al., 2009; Adler et al., 2008)

3.1.2 Myyräaineisto

Tutkielmassani käytetty metsämyyräaineisto on saatu Jyväskylän yliopiston Bio- ja ympäristötieteiden laitokselta. Aineisto on kerätty Keski-Suomesta Konneveden tutkimusaseman läheisyydestä. Alue josta myyrät on pyydystetty on pinta-alaltaan noin 100 km². Myyrämääriä aineistossa on vuosilta 1995–2014. Myyriä kerättiin kahdestakymmenestä eri paikasta, joissa kussakin oli neljä ansaa, joista kuhunkin voi joutua useampia myyriä. Ansat asetettiin samoille paikoille neljänä ajankohtana vuodessa. Niitä pidettiin samoilla paikoilla kaksi peräkkäistä yötä. Näin ollen pyyntiöitä tuli yhteensä 160 kappaletta/pyyntikerta. Pyyntiajankohdat ovat alkukesä, keskikesä, loppukesä ja syksy. Alkukesänä suoritettu pyynti tapahtui lisääntymiskauden alkuaikana toukokuussa. Keskikesän pyynnin ajankohta oli myyrien lisääntymiskauden puolivälissä, tyypillisesti kesäkuun ensimmäisellä viikolla. Loppukesän pyynti tapahtui elokuun lopusta. Viimeisen eli syksyn pyynnin ajankohta sijoittui lokakuun lopusta marraskuun alkuun. Pyydetty myyrät kuljetettiin Konneveden tutkimusasemalle, jossa kirjaukset on suoritettu. Myyriltä on otettu toukokuusta 2002 lähtien laboratoriossa verinäytteet ja kesäkuusta 2006 lähtien niistä on tutkittu PUUV:n liittyviä vasta-aineita.

Myyrämerkintöjen kirjanpidossa olevat pyyntiajankohdat sisälsivät useiden päivien vaihtelua yksittäisen pyyntiajankohdan sisällä. Aineiston keränneiden tutkijoiden mukaan myyrät kuitenkin pyydettiin kahdelta peräkkäiseltä päivältä. Näin ollen tässä työssä oletetaan kaikkien alkukesän pyyntien tapahtuvan toukokuussa. Keskikesän pyynnit oletetaan heinäkuulle. Loppukesän pyynnit merkitään elokuulle ja syksyn pyynnit lokakuulle. Tämä oletus tehdään sen vuoksi, että myyrämäärien interpolointi puuttuville kuukausille olisi helpompaa.

3.1.3 Myyräkuumeaineisto

Myyräkuumetaapusten aineisto on saatu Terveiden ja hyvinvoinnin laitoksen verkkosivuilta (2016). Tapaukset ovat kuukausikohtaisia lukumääriä tammi-kuusta 1995 lähtien. Kaikki tartunnat on varmennettu laboratorioissa. Tutkielmassani käytän Keski-Suomen, Pohjois-Savon ja Etelä-Savon sairaanhoidopiirin myyräkuumetartuntoja. Nämä kolme maakuntaa ovat sijainniltaan melko lähellä Konneveden tutkimusasemaa, jonka läheisyydestä myyräaineisto on kerätty. Alla olevassa taulukossa 1 on tartuntatapaukset maakunnittain vuosille 1995–2014.

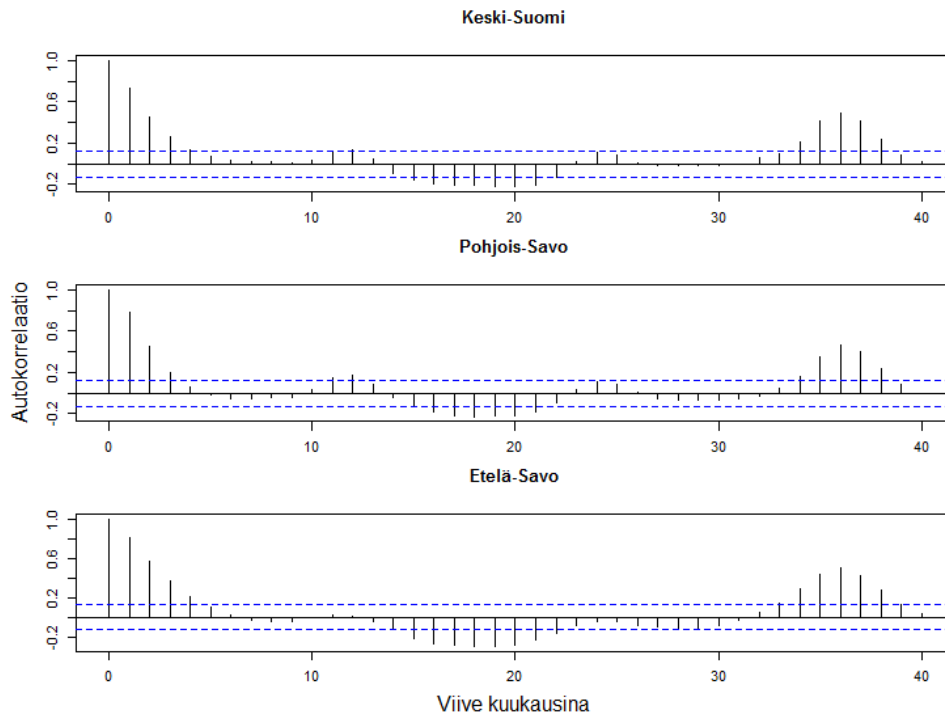
Taulukko 1: Myyräkuumetapaukset vuosittain ja sairaanhoitopiireittäin. Hyppäykset myyräkuumetapauksissa vuodesta 1999 alkaen kolmen vuoden välein on lihavoitu.

Vuosi	Keski-Suomi	Pohjois-Savo	Etelä-Savo
1995	66	130	85
1996	92	86	51
1997	91	46	44
1998	135	240	139
1999	212	235	220
2000	75	39	26
2001	155	176	87
2002	381	275	248
2003	230	139	117
2004	178	187	132
2005	253	292	292
2006	137	180	128
2007	170	223	65
2008	305	406	323
2009	163	175	197
2010	140	197	98
2011	194	192	114
2012	92	134	61
2013	189	330	158
2014	259	341	106
2015	133	158	68
Yhteensä	3517	4023	2691

Taulukosta havaitaan, että myyräkuumetapauksia on eniten Pohjois-Savossa ja vähiten Etelä-Savossa. Siitä nähdään myös, että vuodesta 1999 lähtien kaikilla alueilla myyräkuumetapauksissa on hyppäys noin kolmen vuoden välein. Suurimmat myyräkuumetapaukset eivät satu aina aivan kolmen vuoden syklille. Muun muassa Pohjois-Savossa vuosina 1998 ja 2010 on hie- man enemmän myyräkuumetapauksia kuin vuosina 1999 ja 2011, sekä Etelä-Savossa on vuonna 2013 huomattavasti enemmän tapauksia kuin 2014. Pää- sääntöisesti kolmen vuoden välein tapahtuvat hyppäykset tapahtuvat samoina vuosina kuin myyräkannat ovat runsaimmillaan.

Kuviossa 1 näkyy kaikkien kolmen alueen myyräkuumetapausten autokor- relaatiofunktiot. Viiveet ovat kuukauden tarkkuudella, jolloin vaaka-akselilla kohdassa 12 on vuoden viiveellä olevien havaintojen autokorrelaatio. Jokaisel- la alueella näyttäisi olevan myyräkuumetapausten kohdalla selkeää syklistä

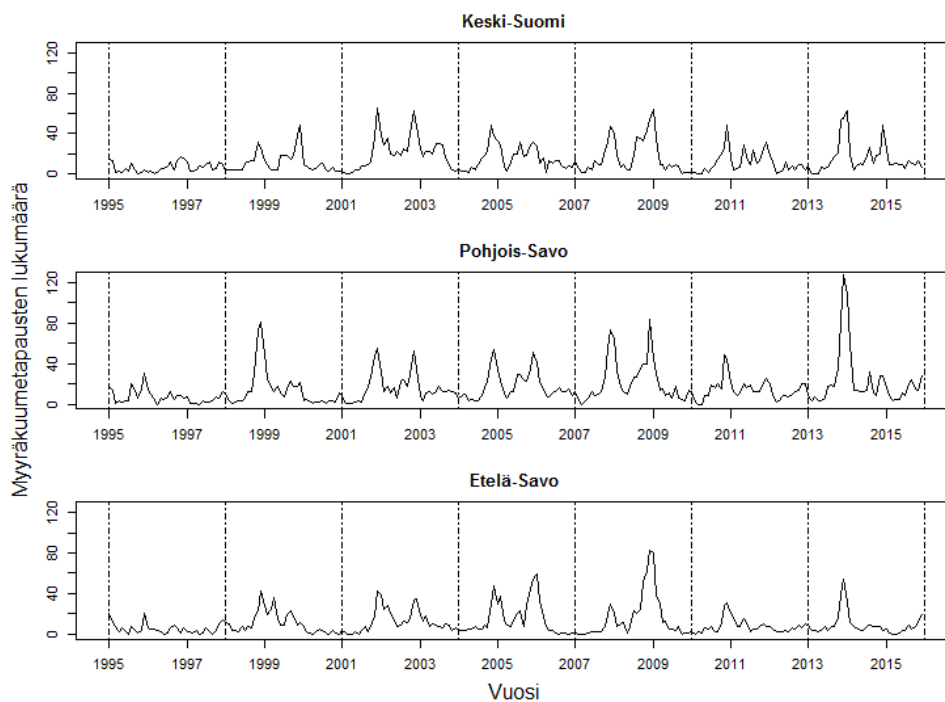
vaihtelua ja vaihtelu näyttää olevan pääpiirteisesti samanlaista jokaisella alueella. Kullakin alueella on selkeä positiivinen autokorrelaatio 36 kuukauden välein ja Keski-Suomen ja Pohjois-Savon tapauksissa lievä 12 kuukauden kausivaihtelu. Jokaisella alueella on myös lievä negatiivinen autokorrelaatio noin 18 kuukauden välein.



Kuvio 1: Autokorrelaatiofunktiot Keski-Suomen, Pohjois-Savon ja Etelä-Savon sairaanhoitopiirien myyräkuumemäärille kuukauden viiveestä lähtien. Sinisten katkoviivojen ulkopuolella olevat arvot ovat tilastollisesti merkitseviä riskirajalla $\alpha = 0.05$.

Kuviossa 2 on kuukausittaiset aikasarjat Keski-Suomen, Pohjois-Savon ja Etelä-Savon myyräkuumetapauksista. Kuvasta nähdään, että pääpiirteittäin suurimmat piikit tapauksissa ovat alueittain samaan aikaan, mutta suuriaakin poikkeamia on. Pohjois-Savossa on loppuvuosina 1998 sekä 2013 huomattavasti suuremmat määrät myyräkuumetapauksia kuin Keski-Suomessa tai Etelä-Savossa. Keski-Suomessa on vuosien 2000 ja 2014 lopussa selkeät piikit myyräkuumetapauksissa, mutta samaan aikaan Pohjois- ja Etelä-Savossa tapauksia on suhteellisen vähän. Kaikilla alueilla on aikasarjan alkupäässä vuosina 1995-1997 melko vähän tapauksia ja varsinainen sekä vuosittainen että

kolmen vuoden välinen sykliisyys myyräkuumetapauksissa näkyy selkeämmin vuodesta 1999 alkaen. Tosin kuten aiemminkin jo on todettu, silloinkaan suurimmat piikit eivät aina osu täsmälleen kolmen vuoden välein.



Kuvio 2: Myyräkuumetapausten lukumäärät kuukausittain Keski-Suomen, Pohjois-Savon ja Etelä-Savon sairaanhoitopiireissä.

Seuraavaksi on tarkoitus tutkia kuinka myyrärunsauksilla voidaan ennustaa myyräkuumetapauksia vuodelle 2015. Oikeat myyräkuumemäärät vuodelle 2015 tiedetään, joten estimoituja määriä voidaan verrata niihin ja näin ollen mallin todellista ennustamiskykyä voidaan testata. Ennen kuin ennustemalli voidaan sovittaa, täytyy myyrärunsaudet interpoloida puuttuville kuukausille.

3.2 KFAS -paketti R-ympäristössä

Tila-avaruusmallien rakentamiseen löytyy R-ohjelmistossa useita kirjastoja, kuten **d1m**, **dse** ja **KFAS**. Tässä työssä myyrätiheyksien interpolointi puuttuville kuukausille sekä myyräkuumeen ja myyrätiheyksien välinen mal-

lintaminen suoritetaan KFAS-paketin avulla. KFAS-paketilla voidaan tila-avaruusmallien lisäksi mallintaa myös muun muassa yksinkertaista regressiota, yleistettyjä lineaarisia sekamalleja ja autoregressiivisiä integroituvia liukuvan keskiarvon (ARIMA) aikasarjoja. Tila-avaruusmallien tapauksessa sillä voidaan mallintaa eksponentiaalisen perheeseen kuuluvia jakaumia: normaali-, Poisson-, binomi, negatiivinen binomi ja gamma-jakaumaa. Kun haluttu malli on valittu, tulee kaavassa (1) määrittellä matriisit Z, H, T, R ja Q sopivasti. Tämän lisäksi ensimmäisille tiloille ja niiden variansseille voidaan antaa alkuarvot. Tässä työssä KFAS-pakettia käytetään tila-avaruusmallien mallinnukseen. (Helske, 2015)

3.3 Myyrärunsauksien interpolointi

Tutkielmani päätavoitteena on tutkia myyrätiheyksien ja myyräkuumeta-
pausten välistä riippuvuutta ja selvittää voidaanko myyrärunsauksilla ennustaa tulevia myyräkuumeta-
pausten lukumääriä. Kyseisen riippuvuuden tutkiminen ei kuitenkaan ole täysin suoraviivaista, sillä myyräkuumeaineiston ollessa kuukausittaista, on myyrärunsaudet sen sijaan vain neljältä kuukaudelta vuodessa. Myyrrien populaatiokoko on prosessi, joka vaihtelee ajan suhteen.

3.3.1 Myyrärunsauksien interpolointimalli

Havaittuja myyrärunsauksia mallinnetaan Poisson-jakaumalla

$$p(y_t|\mu_t) = \frac{\lambda_t^{y_t}}{y_t!} e^{-\lambda_t},$$

jonka odotusarvo ja varianssi on $\lambda_t = \exp(\theta_t) = \exp(Z_t\alpha_t)$. Ennusteet myyräkuumeta-
pauksille halutaan kuukausitasolla, joten myyrärunsaudetkin tulee olla joka kuukaudelta. Tämä saadaan interpoloimalla myyrärunsaudet puuttuville kuukausille. Interpoloinnissa lasketaan estimoidut arvot kullekin ajanhetkelle ja käytetään niitä todellisina arvoina. Interpolointi suoritetaan tila-
avaruusmallina siten että tiloina ovat 12 ja 36 kuukauden syklikkomponentit sekä satunnaiskulku. Kohinatermit η_t ovat keskenään riippumattomia ja samoinjakautuneita satunnaismuuttujia odotusarvolla nolla ja varianssilla σ_η^2 (Durbin and Koopman, 2001, s. 9). Interpolointimallina myyrärunsauksille

käytetään Poisson-jakaumaa, joka on muotoa

$$\begin{aligned}
p(y_t|\theta_t) &= \exp [y_t\theta_t - \exp(\theta_t) - \log y_t!], \\
\theta_t &= \mu_t + c_{12,t} + c_{36,t}, \\
\mu_{t+1} &= \mu_t + \eta_t, \\
c_{12,t+1} &= c_{12,t} \cos\left(\frac{\pi}{6}\right) + c_{12,t}^* \sin\left(\frac{\pi}{6}\right) + \omega_{12,t}, \\
c_{12,t+1}^* &= -c_{12,t} \sin\left(\frac{\pi}{6}\right) + c_{12,t}^* \cos\left(\frac{\pi}{6}\right) + \omega_{12,t}^*, \\
c_{36,t+1} &= c_{36,t} \cos\left(\frac{\pi}{18}\right) + c_{36,t}^* \sin\left(\frac{\pi}{18}\right) + \omega_{36,t}, \\
c_{36,t+1}^* &= -c_{36,t} \sin\left(\frac{\pi}{18}\right) + c_{36,t}^* \cos\left(\frac{\pi}{18}\right) + \omega_{36,t}^*,
\end{aligned} \tag{21}$$

missä $c_{12,t}$ on 12 kuukauden syklikomponentti ja vastaavasti $c_{36,t}$ on 36 kuukauden syklikomponentti. Interpolointimallin tilavektori on tällöin

$$\alpha_t' = [\mu_t \quad c_{12} \quad c_{12}^* \quad c_{36} \quad c_{36}^*]$$

ja kovarianssimatriisi

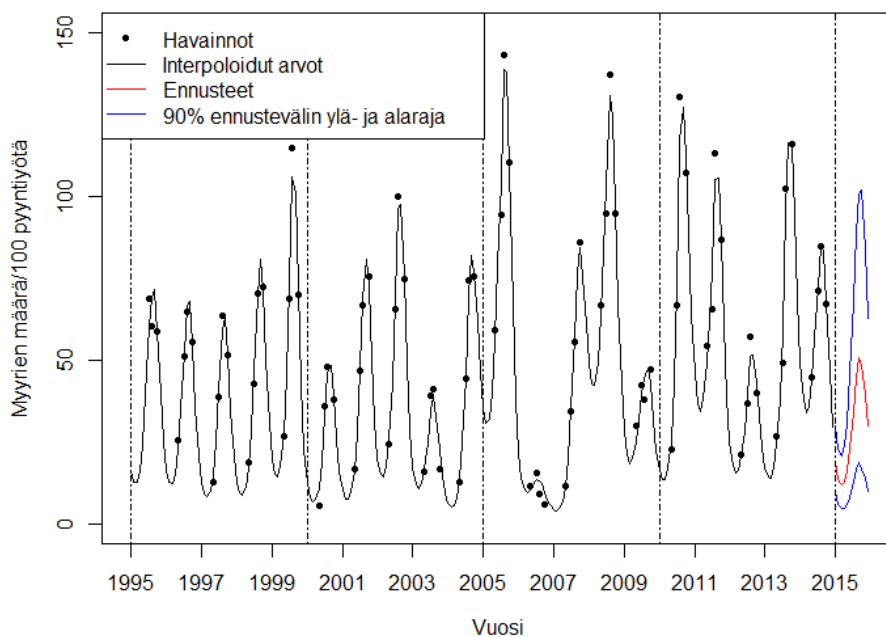
$$Q_t = \text{diag} [\sigma_{\eta_t}^2 \quad \sigma_{\omega_{12}}^2 \quad \sigma_{\omega_{12}}^2 \quad \sigma_{\omega_{36}}^2 \quad \sigma_{\omega_{36}}^2].$$

3.3.2 Myyräaineistoon sovitettu malli

Seuraavaksi tarkastellaan interpolointimallin sopivuutta graafisesti tutkimala tasoituksia ja mallin ennusteita. Kuviossa 3 on kuukausittaiset havaitut myyrämäärät (myyrien lukumäärä/100 pyyntiyötä). Sen lisäksi kuvasta nähdään interpolointimallilla (21) saadut interpoloidut arvot, ennusteet ja ennusteiden 90% ylä- ja alaraja. Havaitaan, että pyydetyissä myyrämäärissä on suuri vuosittainen vaihtelu. Vuonna 2005 myyriä on pyydetty ennätysmäärin, kun taas sitä seuraavana vuonna määrä on alimmillaan koko seuranta-aikana. Vuoden 2006 myyrärunsaudet ovat olleen myös kesällä todella alhaiset. Tämä voi selittyä myyrien poimimiseen liittyvällä satunnaisvaihtelulla. Taulukosta 1 havaittiin, että myyräkuumetapaukset ovat huipussaan noin kolmen vuoden välein alkaen vuodesta 1999.

Kuviosta 3 nähdään, että samankaltaisia piikkejä on myös myyrärunsauksissa noin kolmen vuoden välein vuodesta 1999 alkaen. Vuonna 2011 tutkijoiden ansoihin oli jäänyt kuitenkin hieman vähemmän kuin sitä edeltävänä vuonna, kuten myös viimeisimpänä pyyntivuotena 2014. Molempina vuosina 2011 ja 2014 on myyrärunsauksissa ollut kuitenkin selkeät piikit. Kuviosta nähdään

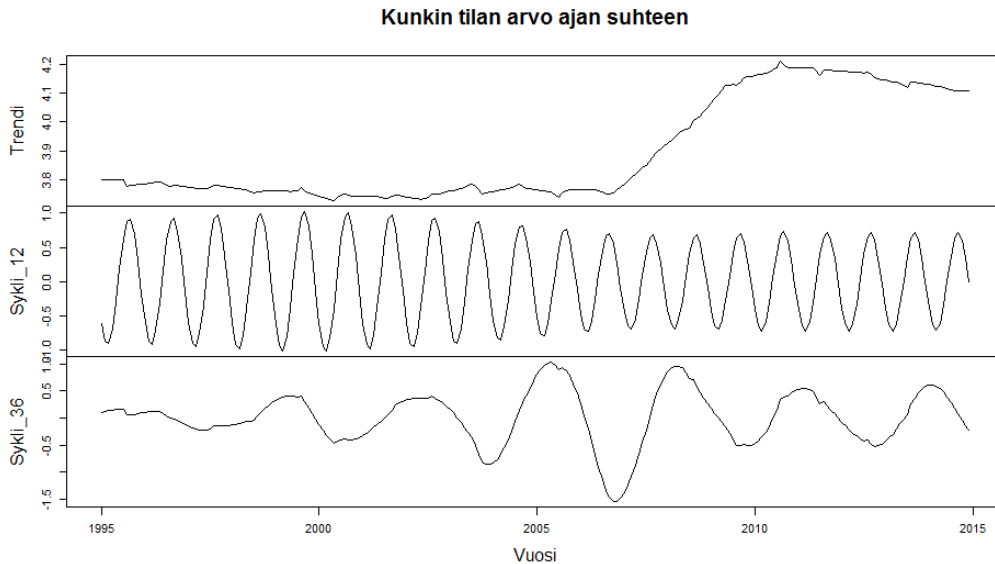
myös selkeä vuosittainen sykli, jossa alku- ja loppuvuodesta myyrärunsaudet ovat alimmillaan ja kesällä ne ovat huipussaan. Interpolointimallilla on ennustettu vuodelle 2015 myyrärunsaudet, jotta myyräkuumetapauksia voidaan ennustaa kyseiselle vuodelle. Ennusteet näyttävät ainakin silmämääräisesti uskottavilta. Vaikka ennusteväli on melko leveä, pysyy se maltillisena myyrämäärien ollessa alimmillaan alkukeväällä ja ylimmillään syys-lokakuussa. Malli ennustaa koko vuodelle noin 350 myyrää/100 pyyntiyötä.



Kuvio 3: Havaitut ja interpoloidut myyrärunsaudet kuukausittain aikaväliltä 1995-2014 sekä ennustetut arvot vuodelle 2015. Mustat pisteet ovat havaittuja arvoja. Punainen viiva on myyrärunsauksien ennustekäyrä vuodelle 2015 ja siniset viivat ovat ennusteiden 90 % luottamusväli.

Kuviossa 4 on tilojen tasoitettut estimaatit $\mathbb{E}(\alpha|y_1, \dots, y_n)$. Kuvioista nähdään, että tasossa on pieni nousu vuodesta 2006 vuoteen 2010. Huomataan myös, että 12 kuukauden sykli estimoituu hyvin, mutta sen sijaan kolmen vuoden välinen vaihtelu ei täysin muistuta säännöllistä sini -sykliä. Siinä sykleissä on välillä kaksi huippua. Kolmen vuoden syklin estimoinnin haasteellisuus voi johtua puuttuvista havainnoista. Lisähaasteen estimointiin aiheuttaa se, että 20 vuoden aikasarjalle mahtuu vain seitsemän kolmen vuoden sykliä.

Kuten aiemmin on todettu kaikki huippukohdat myyrämäärissä eivät satu täsmälleen kolmen vuoden välein, joten sekin vaikeuttaa estimointia. Tilojen variansseiksi tuli $\sigma_\eta^2 = 1.6 * 10^{-3}$, $\sigma_{\omega_{12}}^2 = 4.0 * 10^{-4}$ ja $\sigma_{\omega_{36}}^2 = 7.9 * 10^{-3}$.



Kuvio 4: Interpolointimallin (21) tasoitetut tilat. Ylimmässä kuvassa on trendi, keskimmaisessä kuvassa on 12 kuukauden ja alemmassa 36 kuukauden syklin tasoitukset.

Seuraavaksi käytetään interpolointimallia (21) myyräkuumemäärien ennustamiseksi vuodelle 2015. Edellä käytetyn interpolointimallin avulla saatuja tasoitettuja arvoja käytetään nyt todellisina myyrämäärinä. Koska kyse on lukumääristä, on ne pyöristetty lähimpään kokonaislukuun.

3.4 Myyräkuumetapausten ennustaminen

3.4.1 Viiveen määrittäminen

On oletettavaa, että myyrärunsauksien ja myyräkuumetapausten välillä olevassa riippuvuudessa on viivettä. Toisin sanoen kun tietyssä kuukauteen on havaittu suuri määrä myyriä, on tietyn ajan kuluttua luvassa keskimääräistä suurempi määrä myyräkuumetapauksia ja päinvastoin. Mallin valinnan kannalta on tärkeää selvittää kuinka monen kuukauden viiveestä on kyse. Tämä saadaan selville mallintamalla myyräkuumetapausten määrää

myyrärunsauksilla ja muilla mallissa olevilla parametreillä, siten että myyrärunsauksien ajankohtaa siirretään k määrällä kuukausia eteenpäin, missä $k = 1, 2, 3, 4, 5, 6$. Näin ollen vertailtavien mallien osalta pienin viive myyrärunsauksien ja myyräkuumetapausten välillä on yksi kuukausi ja suurin viive kuusi kuukautta. Tyypillisesti tilastollisia mallien sopivuutta aineistoon vertaillaan Akaiken informaatiokriteerillä tai Bayesilaisella informaatiokriteerillä. Koska vertailtavissa malleissa on yhtä monta parametria, riittää vertailla mallien logaritmisia uskottavuuksia.

Alla olevassa taulukossa 2 on logaritmiset uskottavuudet kullekin viiveelle alueittain. Huomataan, että Keski-Suomen osalta logaritmiset uskottavuudet viiveillä kaksi ja viisi ovat samat kymmenesosan tarkkuudella. Pohjois-Savon tapauksessa suurin logaritminen uskottavuus on mallilla, jossa viive on kaksi kuukautta. Viiden kuukauden viiveen mallin logaritminen uskottavuus on Pohjois-Savon mallissa myös melko suuri. Etelä-Savon aineistoon sovitetussa mallissa logaritminen uskottavuus on suurin viiden kuukauden viiveen mallissa ja toiseksi suurin kahden kuukauden viiveen mallissa. Taulukon 2 logaritmisten uskottavuuksien vertailun perusteella näyttäisi siltä, että olisi yhtä hyvät perustelut valita joko mallit, joissa viive on kaksi kuukautta tai viisi kuukautta. Koska jompi kumpi malli tulee valita jatkokatkeluita varten, niin käytetään jokaisella alueella mallia, jossa on viiden kuukauden viive. Lievänä perusteluna valinnalle on se, että Pohjois-Savon mallin tapauksessa viiden kuukauden viiveen logaritminen uskottavuus on hieman lähempänä kahden kuukauden viiveen mallin logaritmistä uskottavuutta kuin Etelä-Savon mallin tapauksessa kahden kuukauden viiveen malli on viiden kuukauden viiveen mallin logaritmistä uskottavuutta.

Taulukko 2: Eri viiveellä olevien mallien logaritmiset uskottavuudet.

Viive (kk)	Keski-Suomi	Pohjois-Savo	Etelä-Savo
1	782,4	790,2	696,4
2	798,9	794,5	695,7
3	770,9	780,7	690,4
4	778,6	787,6	695,3
5	798,9	793,8	697,9
6	746,1	760,1	687,9

Tutkimuksessa Kallio et al. (2009) käytettiin samaa myyräaineistoa kuin tässä tutkielmassa, mutta siinä viimeisenä tutkittavana vuotena oli 2008. Myös kyseisessä tutkimuksessa selvitettiin myyrärunsauksien ja myyräkuumeen välistä viivettä ja tulokseksi saatiin, että malli joka sisälsi viiden kuu-

kauden viiveen oli paras Akaiken informaatiokriteerin perusteella. Siinä mallinnus ja kuukausittainen interpolointi suoritettiin käyttäen lineaarista sekamallia. Sekamallissa vasteena oli myyräkuumetapaukset ja prediktoreina olivat päävaikutuksina kuukausi, syklin vaihe ja myyriä määrä viiveellä $t = 0, 1, 3, 5, 6$ ja kuukauden ja syklin vaiheen interaktio. Saamani tulokset ovat näin ollen samansuuntaisia kuin edellä mainitussa tutkimuksessa, joskin olen käyttänyt eri menetelmää.

3.4.2 Mallinnus

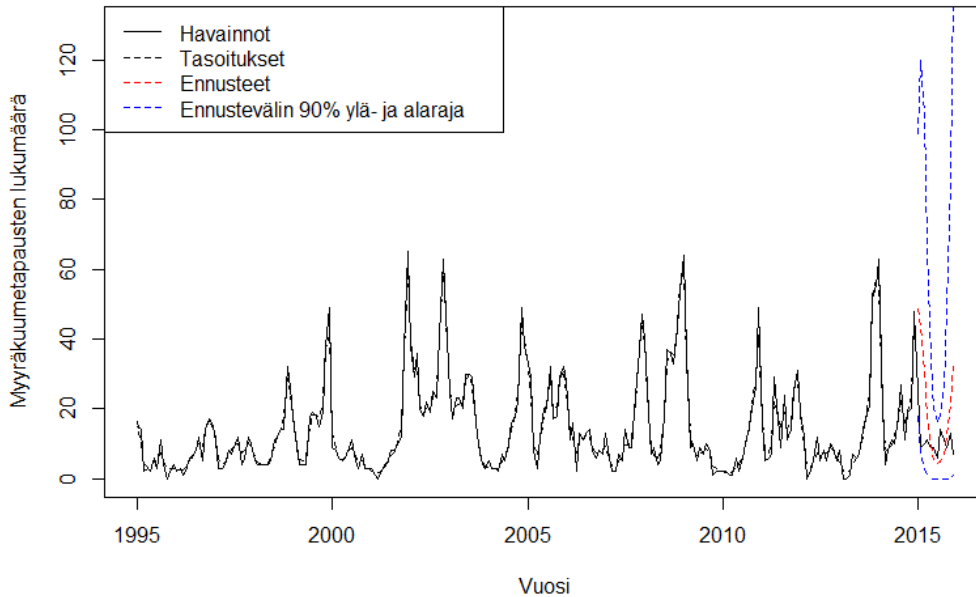
Tarkastellaan seuraavaksi luvussa 3.3.2 sovitetulla interpolointimallilla saatujen myyrärunsauksien ennustekykyä Keski-Suomen, Pohjois-Savon ja Etelä-Savon sairaanhoitopiirien myyräkuumetapauksille. Kullekin alueelle käytetään omaa tila-avaruusmallia, jossa prediktorina on myyrärunsaudet viisi kuukautta aiemmalla ajanhetkellä. Tämän lisäksi malleissa on sekä satunnaiskulkua että vuoden ja kolmen vuoden syklikomponentit. Edellä mainitulla interpolointimallilla (21) saaduilla myyrärunsauksilla myyräkuumemäärien ennustamiseen käytetty aluekohtainen malli on

$$\begin{aligned}
p(y_{t,alue}|\theta_t) &= \exp [y_{t,alue}\theta_t - \exp(\theta_t) - \log y_{t,alue!}], \\
\theta_t &= \mu_t + X_{t-5}\beta_t + c_{12,t} + c_{36,t}, \\
\mu_{t+1} &= \mu_t + \eta_t, \\
\beta_{t+1} &= \beta_t, \\
c_{12,t+1} &= c_{12,t} \cos\left(\frac{\pi}{6}\right) + c_{12,t}^* \sin\left(\frac{\pi}{6}\right) + \omega_{12,t}, \\
c_{12,t+1}^* &= -c_{12,t} \sin\left(\frac{\pi}{6}\right) + c_{12,t}^* \cos\left(\frac{\pi}{6}\right) + \omega_{12,t}^*, \\
c_{36,t+1} &= c_{36,t} \cos\left(\frac{\pi}{18}\right) + c_{36,t}^* \sin\left(\frac{\pi}{18}\right) + \omega_{36,t}, \\
c_{36,t+1}^* &= -c_{36,t} \sin\left(\frac{\pi}{18}\right) + c_{36,t}^* \cos\left(\frac{\pi}{18}\right) + \omega_{36,t}^*,
\end{aligned} \tag{22}$$

missä $alue = \text{Keski-Suomi, Pohjois-Savo, Etelä-Savo}$, X_{t-5} on myyrärunsaus 5 kuukautta aiemmalla ajanhetkellä ja muut termit kuten mallissa (21).

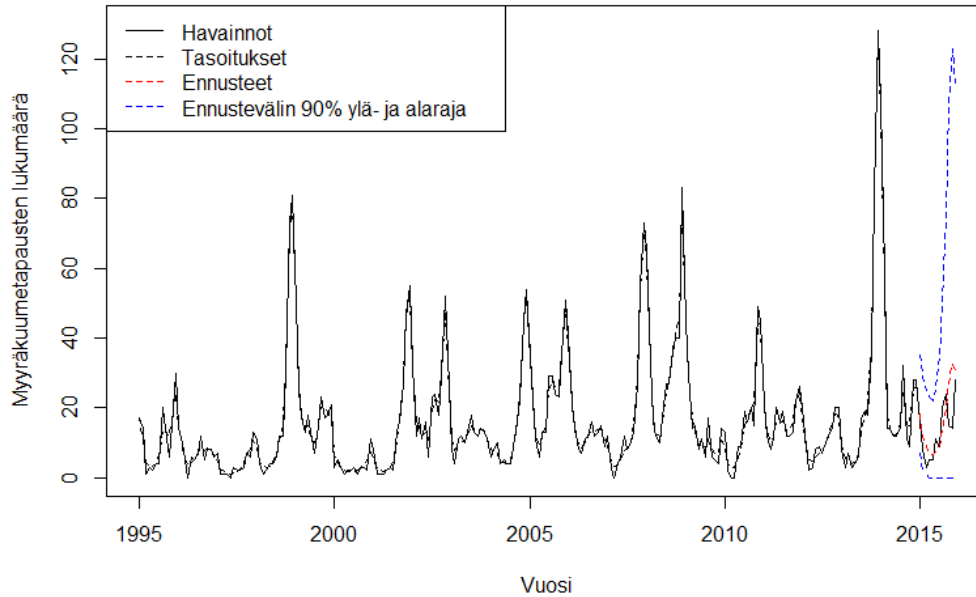
Keski-Suomen myyräkuume-ennusteet nähdään kuvasta 5. Malli ennustaa vuodelle 2015 syklistä vaihtelua myyräkuumetapauksille siten, että alku- ja loppuvuodelle ennustetaan suurempaa määrää kuin kesälle. Malli ennustaa yhteensä koko vuodelle 240 myyräkuumetapausta. Todellinen myyräkuumetapausten lukumäärä Keski-Suomen sairaanhoitopiirissä kyseiselle vuodelle oli 133. Ennusteiden 90% ennusteväli on todella leveä alarajan ollessa lähellä

nollaa ja ylärajan ollessa moninkertaisesti aiempien vuosien suurimpia piikkejä kookkaampi alku- ja loppuvuonna. Kesällä 90% ennusteväli on melko kapea. Mallin tuottamat ennusteet eivät ole kovin käyttökelpoisia.



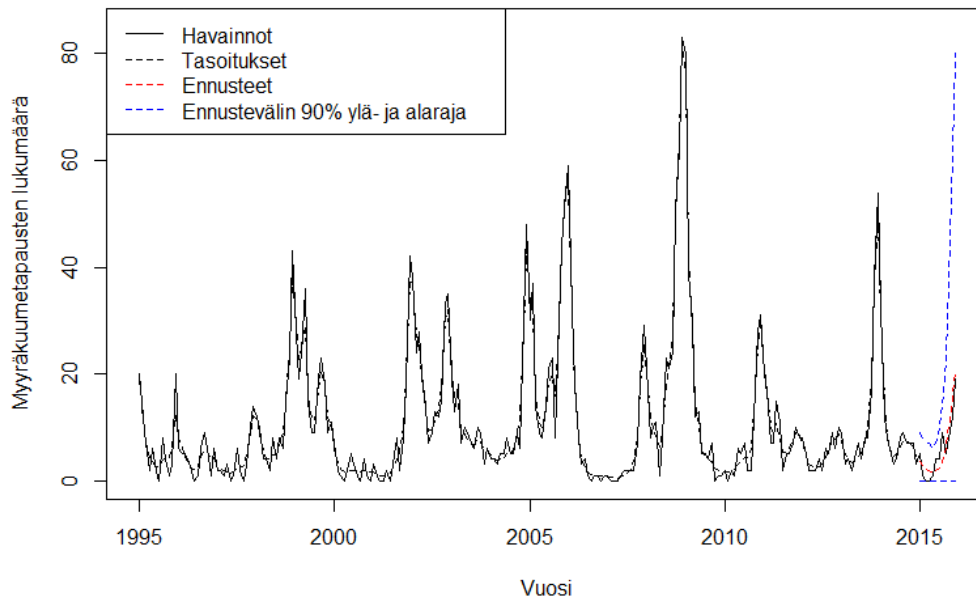
Kuvio 5: Keski-Suomen sairaanhoitopiirin myyräkuumemäärät, mallinnetut tasoiukset ja ennustetut myyräkuumemäärät vuodelle 2015.

Kuvassa 6 näkyy Pohjois-Savon sairaanhoitopiirin myyräkuumetapausten aikasarja sekä mallista saadut tasoiukset ja ennusteet vuodelle 2015. Huomataan, että verrattuna Keski-Suomen sairaanhoitopiirin myyräkuumemääriin, malli antaa paremmat ennusteet. Tässä tapauksessa vuoden 2015 ennusteet ovat hyvinkin lähellä havaittuja vuoden 2015 myyräkuumemääriä. Malli ennustaa vuodelle 2015 myyräkuumetapauksia yhteensä 193, kun luku todellisuudessa kyseiselle vuodelle oli 158. Todellisissa vuoden 2015 myyräkuumetapausten lukumäärissä näyttäisi olevan kaksi huippua, mutta koska kyse on hyvin pienistä määristä, niin se voi selittyä satunnaisvaihtelulla. Lisäksi 90% ennusteväli on huomattavasti kapeampi ylärajan ollessa lähellä vuosien 2013-2014 vaihteen suurta piikkiä ja alarajan ollessa lähellä nollaa. Pohjois-Savon tapauksessa mallilla vaikuttaisi olevan kohtalainen ennustekyky. Melko luotettavan ennusteen sillä saa noin puoli vuotta eteenpäin.



Kuvio 6: Pohjois-Savon sairaanhoitopiirin myyräkuumemäärät, mallinnetut tasoitukset ja ennustetut myyräkuumemäärät vuodelle 2015.

Etelä-Savon sairaanhoitopiirin myyräkuumetapausten aikasarja ja mallin antamat ennusteet näkyvät kuviosta 7. Kuviosta nähdään, että ennusteet ovat hyvin lähellä todellisia arvoja. Ennustevalit ovat kuitenkin melko leveät, alarajan ollessa koko ajan nollassa ja ylärajankin käydessä ylimillään joulukuussa 80 tapauksessa. Yhteensä malli ennustaa vuodelle 2015 myyräkuumetapauksia 73, joka on hyvin lähellä todellista 68 myyräkuumetapausta. Malli näyttäisi toimivan melko hyvin ennustamiseen tuleville myyräkuumetapauksille noin puoli vuotta eteenpäin.



Kuvio 7: Etelä-Savon sairaanhoitopiirin myyräkuumemäärät, mallinnetut tasoiukset ja ennustetut myyräkuumemäärät vuodelle 2015.

Alla olevassa taulukossa 3 on sekä havaitut myyräkuumetapaukset vuodelle 2015, että mallin (22) antamat ennusteet ja niiden 90% ennustevälit. Taulukosta nähdään, että Keski-Suomen sairaanhoitopiirille malli ennustaa alkuvuodesta (tammikuu-huhtikuu) ja loppuvuodesta (lokakuu-joulukuu) jonkin verran havaittua enemmän myyräkuumetapauksia. Sen sijaan kesälle malli ennustaa hieman vähemmän kuin havaittiin. Kuten myös kuvasta 5 havaitaan, ennustevälit ovat huomattavasti kapeammat kesällä kuin alkua ja loppuvuodesta. Pohjois-Savon sairaanhoitopiirille malli antaa alkuvuodesta heinäkuulle asti melko hyvät ennusteet vaikkakin ennusteiden 90% yläraja on melko suuri. Elokuusta vuoden loppuun asti malli antaa jonkin verran suurempia ennusteita kuin havaitut määrät ovat. Ennusteiden ylärajan räjähdysmäisen nousun vuoksi loppuvuodesta niiden luotettavuus laskee rajusti. Etelä-Savon sairaanhoitopiirin myyräkuumetapauksille malli onnistuu ennustamaan melko hyvin tautitapausten lukumäärät. Ennusteiden ennustevälit kasvavat jälleen rajusti elokuusta lähtien, mutta niiden ennustuskyky vaikuttaa luotettavilta ainakin parin kuukauden päähän.

Taulukko 3: Havaitut myyräkuumetautit ja mallin (22) ennusteet vuodelle 2015 sairaanhoitopiireittäin.
Kuukausi

	Keski-Suomi			Pohjois-Savo			Etelä-Savo		
	Havaittu arvo	Ennuste	Ennusteen 90% ennusteväli	Havaittu arvo	Ennuste	Ennusteen 90% ennusteväli	Havaittu arvo	Ennuste	Ennusteen 90% ennusteväli
Tammikuu	25	49	(18 ; 99)	16	18	(7 ; 35)	5	4	(0 ; 9)
Helmi	9	44	(7 ; 120)	8	12	(2 ; 30)	1	3	(0 ; 8)
Maalis	10	32	(3 ; 105)	3	9	(1 ; 25)	0	2	(0 ; 7)
Huhti	11	18	(1 ; 65)	5	7	(0 ; 23)	0	2	(0 ; 7)
Touko	9	10	(0 ; 35)	5	7	(0 ; 22)	1	2	(0 ; 6)
Kesä	9	6	(0 ; 20)	11	7	(0 ; 25)	4	2	(0 ; 7)
Heinä	6	4	(0 ; 16)	9	9	(0 ; 34)	4	2	(0 ; 10)
Elo	14	5	(0 ; 18)	20	13	(0 ; 51)	9	4	(0 ; 14)
Syys	11	7	(0 ; 26)	24	20	(0 ; 74)	5	6	(0 ; 23)
Loka	9	13	(0 ; 49)	15	27	(0 ; 107)	8	10	(0 ; 38)
Marras	13	22	(0 ; 86)	14	32	(0 ; 123)	12	16	(0 ; 61)
Joulu	7	32	(1 ; 140)	28	31	(0 ; 113)	19	21	(0 ; 80)
Yhteensä	133	240		158	193		68	73	

4 Yhteenveto

Tutkin Keski-Suomessa pyydettyjen myyrärunsauksien ja Keski-Suomessa, Pohjois-Savossa ja Etelä-Savossa havaittujen myyräkuumetapausten lukumäärän välistä riippuvuutta tila-avaruusmalleilla. Tavoitteena oli selvittää, kuinka monen kuukauden viiveellä myyräkuumetapaukset ilmenivät suhteessa myyrärunsauksiin ja pyrkiä ennustamaan myyräkuumetapauksia vuodeksi eteenpäin. Ennen ennustamista myyrärunsaudet täytyi interpoloida puuttuvilta kuukausilta. Siihen käytetty Kalmanin suodin osoittautui tehokkaaksi menetelmäksi.

Paras ennustemalli selvitettiin vertailemalla mallien logaritmisia uskottavuuksia viiveen vaihdellessa kuukauden välein yhdestä kuuteen kuukautta. Päädyin malliin, jossa myyräkuumetapauksia ennustettiin viisi kuukautta aiemmin havaittujen myyrärunsauksien avulla. Tämän lisäksi malli sisälsi sekä trendin että vuoden ja kolmen vuoden syklikomponentit.

Vuoden 2015 myyräkuumetapausten ennustaminen suoritettiin kullekin alueelle erikseen. Keski-Suomen tapauksessa ennusteiden 90 % ennustevälit olivat huomattavasti suuremmat kuin Pohjois- ja Etelä-Savon tapauksissa. Keski-Suomen mallilla ei myöskään havaittu olevan kovin hyvää ennustamiskykyä mallin ennustaessa yhteensä lähes kaksinkertainen määrä myyräkuumetapauksia verrattuna todelliseen määrään. Pohjois- ja Etelä-Savon ennusteet olivat suhteellisen lähellä todellisia havaittuja arvoja ainakin noin kuusi kuukautta eteenpäin. Pohjois-Savon tapauksessa malli ennusti yhteensä koko vuodelle noin kolmasosan enemmän tautitapauksia havaittuun määrään verrattuna. Etelä-Savoon ennustettiin lähes oikea määrä myyräkuumetapauksia ennustusvuodelle. Siellä ennustettu määrä poikkesi havaitusta arvosta vain viidellä tartuntatapauksella. Kaikkien alueiden mallien 90 % ennustevälien alaraja oli lähes koko ennustusvälillä lähellä nollaa ja yläraja kasvoi vuoden loppua kohden hyvin suureksi.

Ennustettu vuosi 2015 ei sattunut myyräpopulaatiokoon kolmen vuoden syklin huippukohtaan, mikä teki estimoinnista hieman tarkempaa. Olisi mielenkiintoista nähdä, kuinka tarkasti malli onnistuisi ennustamaan suuren piikin esimerkiksi vuodelle 2017, joka on seuraava arvioitu huippukohta myyrärunsauksille.

Metsämyyrät ovat myös useiden muiden eri tartuntatautien kantajia. Tässä tutkielmassa keskityttiin vain myyräkuumeeseen. Käyttämäni tila-avaruusmallia voisi myös kokeilla muihin metsämyyrien kuljettamien tautien mallintamiseen. Se edellyttäisi vain aineiston toisesta sairaudesta ja implementointi olisi näin ollen suoraviivaista.

Lähteet

- Adler, F. R., Pearce-Duvel, J. M., and Dearing, M. D. (2008). How host population dynamics translate into time-lagged prevalence: an investigation of sin nombre virus in deer mice. *Bulletin of mathematical biology*, 70(1):236–252.
- Bernshtein, A., Apekina, N., Mikhailova, T., Myasnikov, Y. A., Khlyap, L., Korotkov, Y. S., and Gavrilovskaya, I. (1999). Dynamics of puumala hantavirus infection in naturally infected bank voles (*Clethrionomys glareolus*). *Archives of virology*, 144(12):2415–2428.
- Brummer-Korvenkontio, M., Vaheiri, A., Hovi, T., Von Bonsdorff, C.-H., Vuorimies, J., Manni, T., Penttinen, K., Oker-Blom, N., and Lähdevirta, J. (1980). Nephropathia epidemica: detection of antigen in bank voles and serologic diagnosis of human infection. *Journal of infectious diseases*, 141(2):131–134.
- Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.
- Harvey, A. C. (1993). *Time Series Models*. Harvester Wheatsheaf.
- Helske (2015). *KFAS: Kalman Filter and Smoother for Exponential Family State Space Models*. R package version 1.1.2.
- Heyman, P., Vervoort, T., Escutenaire, S., Degraeve, E., Konings, J., Vandenvelde, C., and Verhagen, R. (2001). Incidence of hantavirus infections in Belgium. *Virus Research*, 77(1):71–80.
- Kallio, E. R., Begon, M., Henttonen, H., Koskela, E., Mappes, T., Vaheiri, A., and Vapalahti, O. (2009). Cyclic hantavirus epidemics in humans—predicted by rodent host dynamics. *Epidemics*, 1(2):101–107.
- Kallio, E. R., Henttonen, H., Koskela, E., Lundkvist, Å., Mappes, T., and Vapalahti, O. (2013). Maternal antibodies contribute to sex-based difference in hantavirus transmission dynamics. *Biology letters*, 9(6):20130887.
- Luontoportti. Saatavilla internetistä: <http://www.luontoportti.com/suomi/fi/nisakkaat/metsamyyra>, luettu 1.5.2015.
- McCaughey, C. and Hart, C. (2000). Hantaviruses. *Journal of medical microbiology*, 49(7):587–599.

- Niklasson, B., Hornfeldt, B., Lundkvist, A., Bjorsten, S., and Leduc, J. (1995). Temporal dynamics of puumala virus antibody prevalence in voles and of nephropathia epidemica incidence in humans. *The American journal of tropical medicine and hygiene*, 53(2):134–140.
- Olsson, G. E., White, N., Hjältén, J., and Ahlm, C. (2005). Habitat factors associated with bank voles (*Clethrionomys glareolus*) and concomitant hantavirus in northern Sweden. *Vector-Borne & Zoonotic Diseases*, 5(4):315–323.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C. R. (1973). *Linear Statistical Inference*. John Wiley & Sons.
- Terveiden ja hyvinvoinnin laitos. Saatavilla internetistä:
<https://www.thl.fi/fi/web/infektiotaudit/taudit-ja-mikrobit/virustaudit/puumalavirus>, luettu 24.4.2015.