

Yksittäisten maatilojen tarkastelu korkeaulotteisessa talousavaruudessa itseorganisoivan kartan avulla

Mika Sulkava¹⁾ ja Maria Yli-Heikkilä²⁾

¹⁾ *Luonnonvarakeskus (Luke), PL 2, 00791 Helsinki, mika.sulkava@luke.fi*

²⁾ *Luonnonvarakeskus (Luke), Tietotie 4, 31600 Jokioinen, maria.yli-heikkila@luke.fi*

Luonnonvarakeskuksen (Luke) Tilastopalvelut-yksikkö tuottaa kannattavuuskirjanpidossa monipuolista tietoa Suomen maa- ja puutarhatalouden taloudellisesta tilanteesta. Kannattavuuskirjanpitoaineistoon kerätään vuosittain tietoa vajaalta 1000 maa- ja puutarhataloutta harjoittavalta yritykseltä. Kannattavuuskirjanpidon aggregoituja tuloksia esitetään Luken Taloustohtori –verkkopalvelussa.

Suomen maa- ja puutarhatilojen kannattavuutta ja muita taloudellisia tekijöitä kuvataan tässä työssä itseorganisoivan kartan avulla. Itseorganisoiva kartta on akateemikko Teuvo Kohosen 1980-luvulla kehittämä neuroverkko menetelmä, jonka avulla suuri, moniulotteinen havaintoaineisto projisoidaan kaksiulotteiselle hilalle, ja voidaan siten esittää havainnollisena kaksiulotteisena karttana. Kartalla samankaltaiset tilat sijaitsevat lähellä toisiaan, ja tilojen ominaisuuksia voidaan tarkastella muuttujakohtaisten komponenttitasojen avulla. Menetelmän avulla aineistoja voidaan visualisoida tehokkaasti, ja visuaalisointien avulla aineistosta voidaan löytää useiden muuttujien välisiä epälineaarisia yhteyksiä.

Tutkimuksessamme opetamme kannattavuuskirjanpitoaineiston pohjalta laskettujen keskeisimpien talousmuuttujien avulla itseorganisoivan kartan. Aikaisemmassa tutkimuksessa olemme löytäneet tällä menetelmällä useita kiinnostavia yhteyksiä kerättyjen taloudellisten muuttujien välillä. Muuttujien suuri määrä hankaloittaa aineiston tarkastelua perinteisiä visualisointimenetelmiä käyttäen.

Tässä työssä esitämme, miten keinotekoiset, mutta todellisia tiloja tilastollisilta ominaisuuksiltaan vastaavat esimerkkitalat sijoittuvat maatalouden tilakohtaisella talousdatalla opetetulle itseorganisoivalle kartalle. Tutkimme kartan tekemän projektion laatua ja erityisesti sitä, miten hyvin kartalle sijoitettujen tilojen tulkinta karttahilan avulla vastaa tilan todellista tilannetta.

Yksittäisten tilojen esittäminen itseorganisoivan kartan avulla on mahdollista saada nopea yleiskatsaus tilan taloudellisen tilanteesta. Lisäksi tällainen esitysmuoto kertoo havainnollisella tavalla siitä, miten tilan taloudellinen tilanne vertautuu muihin tiloihin. Tämän avulla voidaan esimerkiksi arvioida, minkä suuntaiset muutokset maatalon toiminnassa voisivat auttaa parantamaan tilan kannattavuutta.

Luken Taloustohtoriin ollaan rakentamassa järjestelmää, joka mahdollistaa kannattavuuskirjanpito-tilojen lisäksi minkä tahansa maatalon sijoittamisen itseorganisoivan kartan hilalle, kunhan tilalta on saatavilla tarpeelliset taloudelliset tiedot. Näin on siis mahdollista esittää kartalla esimerkiksi tiloja, joiden tiedot verkkopalvelun käyttäjä on syöttänyt selaimellaan.

Asiasanat: itseorganisoiva kartta, neuroverkko, maatila, kannattavuus, kirjanpito, muutos

Johdanto

Maatilayritysten liiketoiminnan keskeinen edellytys pitkällä aikavälillä on taloudellinen kannattavuus, mikä mahdollistaa maatalouden toiminnan osana vakaata ruokaketjua. Maatilojen kannattavuus on vaihdellut runsaasti viime vuosina, mutta yleinen kannattavuustrendi on 2000-luvulla ollut laskeva (Rantala ja Tauriainen, 2017). Tämä saattaa hankaloittaa viljelijöiden tulevaisuuden suunnittelua.

Tässä artikkelissa käytetään itseorganisoivaa karttaa (self-organizing map, SOM) maa- ja puutarhatalousyriyten monivuotisen taloudellisen aineiston analysointiin. Itseorganisoiva kartta on tehokas neuroverkkomenetelmä suurten tietoaaineistojen tutkimiseen. Tässä työssä analysoidun aineiston on kerännyt Luonnonvarakeskus (Luke) kannattavuuskirjanpitoiloista koostuvasta otoksesta. Aineisto on myös monien Suomen maataloutta kuvaavien lukujen lähteenä Luken Taloustohtori-verkkopalvelussa (<http://www.luke.fi/taloustohtori>). Työn tavoitteena on esittää, miten yksittäisten tilojen taloudellista kehitystä voidaan kuvata itseorganisoivan kartan avulla.

Itseorganisoivaa karttaa on aiemmin käytetty menestyksekkäästi monissa maataloustieteen ongelmissa (katso esim. Sulkava ym. 2014 ja 2015). Kuitenkaan maatalouden taloudellisen toiminnan analyysissä menetelmää ei ole aiemmin sovellettu kovinkaan paljon. Tässä artikkelissa esitetään uusia tuloksia aikasarjamuotoisen kannattavuuskirjanpitoaineiston analyysistä SOM:n avulla.

Artikkelin loppuosan rakenne on seuraavanlainen. Seuraavassa osassa esitellään kannattavuuskirjanpitoaineisto ja sen käsittely, Taloustohtori sekä SOM:n peruseriaatteet. Kolmannessa osassa on tuloksia yksittäisen maatilan esittämisestä SOM:n avulla, neljännessä osassa tulosten tarkastelua ja viimeisessä osassa näiden pohjalta tehtyjä johtopäätöksiä.

Materiaali ja menetelmät

Kannattavuuskirjanpitoaineisto

Suomen maa- ja puutarhatalousyriyten kannattavuutta kuvaavat luvut lasketaan vuosittain Luken kannattavuuskirjanpidon perusteella. Nämä kannattavuusluvut kuvaavat noin 50 000 yrityksen keskimääräisiä tuloksia. Tilojen kannattavuutta tarkkaillaan noin 900 tilan otoksen avulla. Tässä työssä käytetään vuosien 2000–2015 aineistoa. Kannattavuuskirjanpidon alkuperäisenä tavoitteena on ollut edustaa 40 000 Suomen suurinta tilaa, ja tämän vuoksi otoksessa on melko vähän pieniä tiloja. Itseorganisoivan kartan opetuksen kannalta tällä on se merkitys, että kartta ei kuvaa välttämättä kaikkein pienimpien tilojen muuttujien välisiä yhteyksiä yhtä tarkasti kuin aineistossa paremmin edustettujen hieman suurempien tilojen.

Kannattavuuskirjanpitoaineistoon kerätään tiloilta tuhansia muuttujia. Tässä työssä käytetyistä muuttujista osa on olemassa kirjanpitoaineistossa jo valmiiksi. Latukka (1998) ja Community Committee for the Farm Accountancy Data Network (2009) tarjoavat tarkempaa tietoa kannattavuuskirjanpidon muuttujista.

Muuttujien valinta ja esikäsittely

Tuhansien muuttujien esittäminen sellaisenaan tuottaisi hyvin hitaasti tulkittavan lopputuloksen, joten päädyimme rajaamaan muuttujien määrää pienemmäksi. Taloudellinen kannattavuus on maatalousyriyten liiketoiminnan kannalta keskeinen tekijä, ja valitsimme analyysiin mukaan 85 muuttujaa, joilla on vahvin yhteys kannattavuuskertoimeen Yli-Heikkilän ym. (2018) esittämässä ennustavassa satunnaismetsämallissa. Menetelmään liittyvän mallinvalinnan yksityiskohdat on esitetty em. artikkelissa. Satunnaismetsämallin valitsemat muuttujat liittyvät tuotanto-, työtunti- ja tilinpäätöstietoihin.

Satunnaismetsämallin valitsemasta kahdesta eri resoluution tuotantosuuntamuuttujasta jätettiin vähemmän tarkka pois, koska sen informaatio sisältö oli lähes sama tarkemman muuttujan kanssa. Lisäksi koska jäljelle jäänyt tuotantosuunta on luokkamuuttuja, muutettiin se SOM:n opetusta varten kutakin tuotantosuuntaa vastaaviksi binäärisiksi muuttujiksi.

Taloustohtori

Luken Taloustohtori-verkkopalvelu tarjoaa suuren määrän kannattavuuskirjanpitoiloiden tietojen pohjalta laskettuja maa- ja puutarhatalousyritysten tulostietoja. Taloustohtorissa on myös runsaasti muunkinlaista alkutuotantoon, kuten rannikkokalastukseen sekä poro- ja mehiläistalouteen liittyvää tietoa.

Tässä artikkelissa esiteltävä analyysimenetelmää kehitetään osaksi Taloustohtori-palvelua. Tavotteena on rakentaa uusi verkkopalvelu, joka antaa kannattavuuskirjanpitoilaille ja mahdollisesti muillekin tiloille mahdollisuuden tarkastella oman tilan kehitystä SOM:n avulla, verrata tuloksia eri tilaryhmiin sekä arvioida vaihtoehtoisten toimintatapojen vaikutuksia.

Itseorganisoiva kartta

Itseorganisoiva kartta eli SOM (Kohonen 2001) on tehokas menetelmä eksploraatiiviseen data-analyysiin, eli aineistoon tutustumiseen tilanteissa, joissa tutkittavan aineiston rakenteesta ei ole kovin tarkkaa etukäteistietoa. SOM on yksi monista neuroverkkomenetelmistä. Neuroverkkomenetelmät ovat aivojen toimintaa jäljitteleviä tietokoneohjelmia, jotka oppivat automaattisesti niille syötetyn aineiston piirteitä. SOM oppii sille opetetun aineiston rakenteen ja projisoi sen tasopinnalle, josta on helppo piirtää aineistoa esittäviä kuvia. Itseorganisoivaa karttaa on viime vuosikymmenten aikana käytetty menestyksellisesti monissa sovelluksissa (Kaski ym. 1998, Oja ym. 2003, Pöllä ym. 2009).

SOM koostuu säännöllisestä, yleensä kaksiulotteisesta karttayksiköiden hilasta. Kaksiulotteisella hilalla olevan sijaintinsa lisäksi jokaisella karttayksiköllä on moniulotteinen prototyyppivektori alkuperäisen aineiston muodostamassa avaruudessa. Tällä tavoin SOM määrittelee epälineaarisen projektion moniulotteisesta aineistoavaruudesta kaksiulotteiselle hilalle. SOM opetetaan edustamaan alkuperäistä aineistoa mukauttamalla prototyyppivektoreita aineiston jakauman mukaisesti. Yksittäinen tila projisoidaan karttakilalla siihen karttayksikköön (best-matching unit, BMU), jonka prototyyppivektori on alkuperäisessä korkeaulotteisessa avaruudessa lähimpänä.

Opetuksen jälkeen SOM:n laatua voidaan tarkastella mm. topografisen virheen avulla. Topografinen virhe kertoo niiden havaintojen osuuden, joille BMU ei ole kartta hilalla sen karttayksikön vieressä, jonka prototyyppivektori on toiseksi lähimpänä havaintoa (Kiviluoto 1996).

Opetuksen seurauksena aineistossa toisiaan lähellä olevat havainnot ovat tyypillisesti lähellä toisiaan myös kaksiulotteisella hilalla. Itseorganisoivan kartan opetusta voidaan verrata joustavan verkon venyttämiseen aineiston jakauman muotoiseksi moniulotteisessa avaruudessa. SOM:n avulla muodostettua esitystä voidaan käyttää informaation visualisointiin, ryvästyksen ja aineistoon tutustumiseen. Tässä työssä SOM:n opetus ja analysointi toteutettiin käyttäen MATLAB-ohjelmiston SOM Toolbox -lisäosaa (www.cis.hut.fi/projects/somtoolbox/package/docs2/somtoolbox.html). SOM:sta voi tehdä visualisointeja mm. käyttäen komponenttitasoja. Yksi komponenttitalaso näyttää yhden muuttujan arvot värisävyinä karttakilalla.

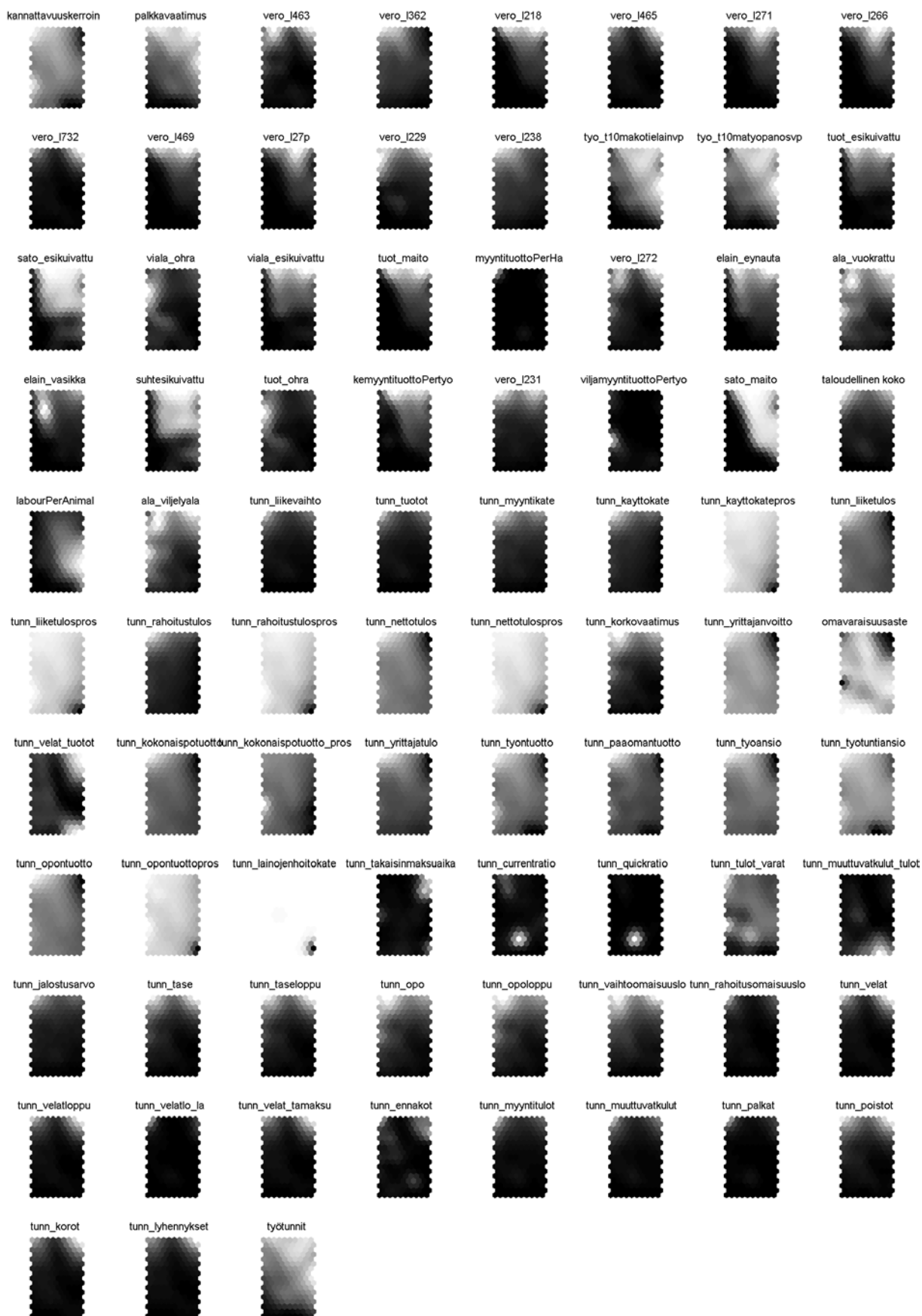
SOM:n opetuksessa käytettävät havainnot muodostuvat yhden tilan yhden vuoden muuttujista. Vuosiluku ei ole opetuksessa käytettävä muuttuja. Saman tilan eri vuosien tiedot ovat siis kaikki mukana opetuksessa erillisinä havaintoina.

Tulokset

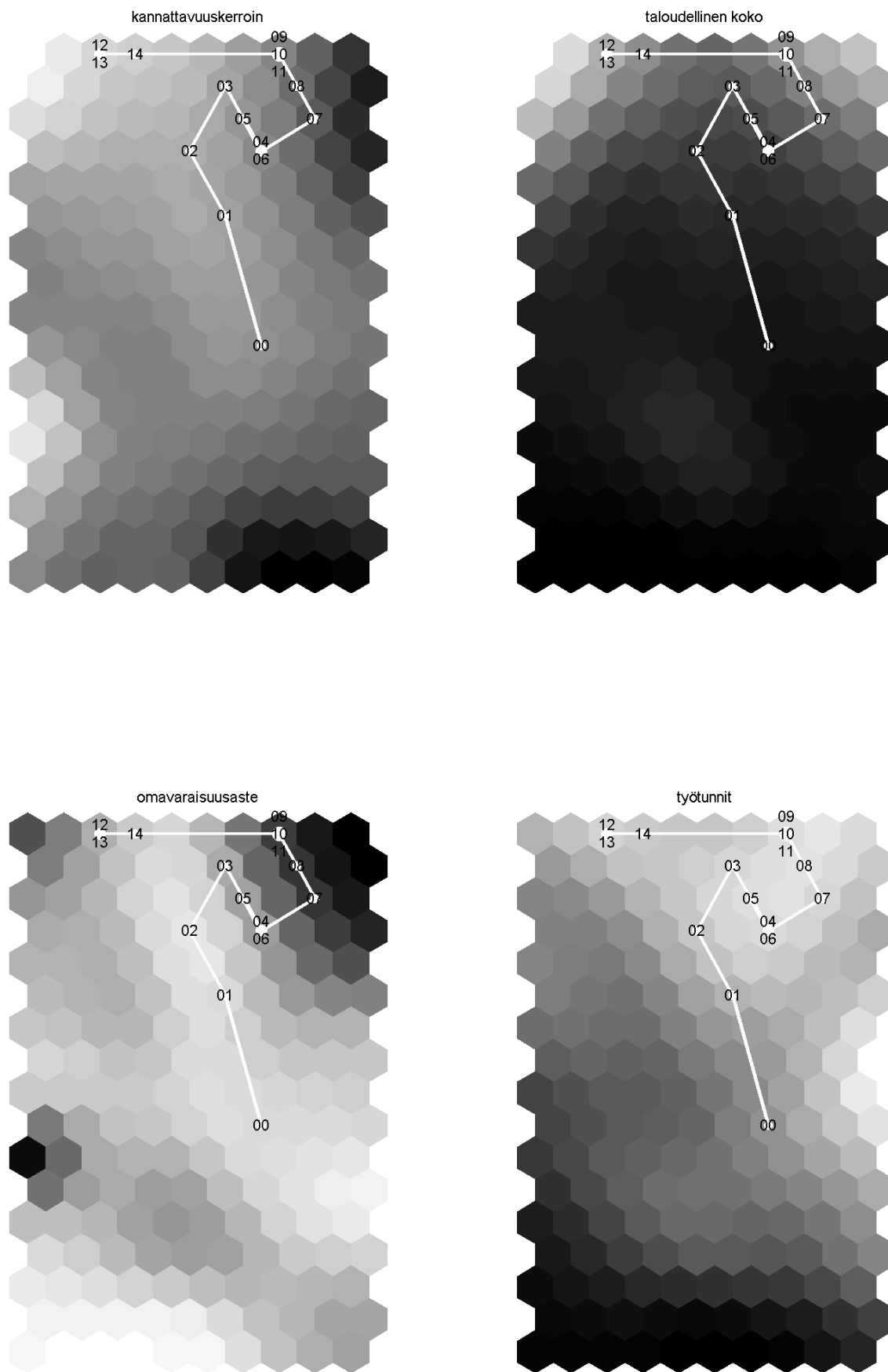
Kannattavuuskirjanpitoiloina kuvaava taloudellinen kartta muodostettiin opettamalla SOM kannattavuuskirjanpitoaineiston edellä kuvatuilla muuttujilla. Kuvassa 1 on esitetty näin saadun SOM:n komponenttitasot. Kuvassa ei ole binäärisiä tuotantosuuntamuuttujia niiden suuren määrän vuoksi. Kunkin tuotantosuunnan tilat ovat tyypillisesti keskittyneet omille yhtenäisille alueilleen kartalla. Kuvasta voidaan nähdä, että kartan kummassakin yläkulmassa ja oikeassa alakulmassa on tiloja, joille on tyypillistä kannattavuuskertoimen äärimmäiset arvot.

Erilaisia taloudellisia tilatyyppisiä on helppo löytää komponenttitasojen avulla. Esimerkiksi kartan vasemman yläkulman tilojen tase (tunn_tase), oma pääoma (tunn_opo), taloudellinen koko ja kannattavuuskerroin ovat suuria. Sulkava ym. (2012, 2015) ovat aiemmin analysoineet yhden vuoden kannattavuuskirjanpitoaineistoa yksityiskohtaisemmin komponenttitasojen avulla – tosin käyttäen hieman erilaista taloudellisten muuttujien joukkoa.

Kuvassa 2 näkyy, miten yksittäinen keinotekoinen esimerkkitala liikkuu karttakilalla ajan kuluessa, kun tilan taloudellinen tilanne muuttuu. Tähän kuvaan on valittu tilan rajallisuuden vuoksi ja tarkastelun helpottamiseksi vain neljä keskeistä muuttujaa.



Kuva 1. Kannattavuuskirjanpitoaineistolla opetetun SOM:n komponenttitasot. Vaaleat harmaasävyt merkitsevät komponenttitasoilla suuria muuttujien arvoja ja tummat pieniä. Yksittäinen tila sijaitsee yhtenä vuonna aina kartan samassa kohdassa kaikilla komponenttitasoilla.



Kuva 2. Esimerkkitalan neljän muuttujan kehitys SOM:n komponenttitasoilla vuosina 2000–2014 (00–14). Vaaleat harmaasävyt kuvaavat muuttujien suuria arvoja ja tummat sävyt pieniä.

Tulosten tarkastelu

SOM:n tekemän projektion laatu oli hyvä. Kartan topografinen virhe on 0,06, mikä tarkoittaa sitä, että aineiston suurelle enemmistölle (94%) kaksi havaintoa lähinnä olevaa karttayksikköä ovat naapuriyksiköitä karttahilalla. Kartta on siis järjestynyt varsin hyvin opetuksen aikana.

Esimerkkitalan kehitys näkyy kuvassa 2, ja pelkästään komponenttitasoja tarkastelemalla kuvan kehitystä voi kuvailla seuraavasti. Tilan alkupiste vuonna 2000 on kartan keskiosassa, missä kannattavuuskerroin, omavaraisuusaste ja työtuntien määrä ovat keskitasoa ja taloudellinen koko pieni. Tila siirtyy vuosina 2001–2002 hilalla ylöspäin kohti aluetta, jossa kaikkien neljän muuttujan arvot ovat korkeampia. Vuonna 2003 kehityksen suunta muuttuu hieman, jolloin kannattavuuskerroin ja omavaraisuusaste eivät enää kasva. Vuosina 2004–2006 tilan muutokset ovat pieniä. Vuonna 2007 tila siirtyy kohti huonon kannattavuuden leimaamaa kartan oikeaa yläkulmaa kohti. Samalla taloudellinen koko kasvaa, mutta omavaraisuusaste laskee. Vuosina 2008–2011 tilan muutokset ovat pieniä kohti hieman parempaa kannattavuutta ja suurempaa kokoa. Vuonna 2012 tila siirtyy kohti kartan vasenta yläkulmaa, jossa kannattavuus on hyvä. Taloudellinen koko ja työtunnit pysyvät samalla tasolla, mutta omavaraisuusaste paranee. Vuosina 2013–2014 tila pysyy samalla alueella.

Edellä kuvattu kehitys voisi kuvata kasvavaa maatilaa, joka tekee suuren investoinnin vuonna 2007, ja velkaantuu. Velka ja heikko taloudellinen suhdanne pitävät kannattavuuden heikkona, kunnes paremmat ajat koittavat vuonna 2012.

Esimerkkitalan arvot kartalla vastaavat tilan havaintoarvoja melko hyvin, ja tilan kehitys etenee havaintojenkin perusteella yllä kuvatulla tavalla. Edellä tarkasteltujen neljän muuttujan kohdalla havaittiin seuraavanlaisia eroja havaintojen ja SOM:n karttahilatulokintojen välillä. Kannattavuuskertoimen ja työtuntien vaihtelu vuosien välillä on hieman voimakkaampaa ja taloudellisen koon muutos tasaisempaa kuin, miltä SOM:n komponenttitasojen perusteella näyttää. Lisäksi omavaraisuusasteen ja työtuntien havainnot ovat tämän esimerkkitalan tapauksessa keskimäärin pienempiä kuin SOM:n komponenttitasojen osoittamat arvot.

Edellä esitetyn kaltaisia havaintoja tehtiin myös muilla keinotekoisilla esimerkkitaloilla. Tilan kehityksen suunta näkyy tilan siirtymisissä karttahilalla. Havaintojen ja SOM:n komponenttitasojen osoittamien arvojen välillä ei ollut suuria eroja esimerkkitalojen välillä.

Johtopäätökset

Itseorganisoivan kartan käyttö mahdollisti tässä työssä Suomen maatalousyrityksiä vuosina 2000–2015 kuvaavien taloudellisen aineiston tehokkaan analyysin. Tuotetun kartan avulla on mahdollista erottaa erilaisia maataloustyyppejä ja tutkia muuttujien välisiä yhteyksiä. Itseorganisoivan kartan avulla on helppo tuottaa hypoteeseja jatkotutkimusta varten (vrt. Sulkava ym. 2015).

Esimerkkitalan kehityksen esittäminen kartalla näyttää, miten minkä tahansa maatilain kehitystä voidaan tarkastella itseorganisoivaa karttaa käyttäen. Kannattavuuskirjanpitoaineistolla opetetun kartan tuottama projektio oli laadultaan hyvä ja esimerkkitalan kehitys SOM:n avulla tarkasteltuna vastasi tilan todellista kehitystä varsin luotettavasti. Lopullisten johtopäätösten tekeminen edellyttää aina alkuperäisiin havaintoihin tutustumista, mutta SOM:n avulla on mahdollista saada nopea ja varsin käyttökelpoinen yleiskatsaus yksittäisten tilojen kehityksestä korkealotteisessa talousavaruudessa.

Kartan luotettavuus riippuu luonnollisesti opetuksessa käytetyn aineiston laadusta, opetuksen parametreista ja muuttujien valinnasta. Nämä ovat keskeisiä aiheita, joihin tulemme keskittymään kehittäessämme uutta SOM-pohjaista tila-analyysijärjestelmää osaksi Taloustohtori-verkkopalvelua.

Kirjallisuus

- Community Committee for the Farm Accountancy Data Network 2009.** Typology handbook, Tech. Rep. RI/CC 1500 rev. 3, European Commission – Directorate-General for Agriculture and Rural Development.
- Kaski, S., Kangas, J. & Kohonen, T. 1998.** Bibliography of self-organizing map (SOM) papers: 1981–1997. *Neural Computing Surveys* 1: 102–350.
- Kohonen, T. 2001.** Self-Organizing Maps, 3rd Edition. *Springer Series in Information Sciences* 30.
- Kiviluoto, K. 1996.** Topology preservation in self-organizing maps, Teoksessa: Proceedings of the International Conference on Neural Networks (ICNN'96), vol.1, IEEE Neural Networks Council, Piscataway, NJ, USA. s. 294–299.
- Latukka, A. 1998.** Maatalousyritysten tulorahoituksen riittävyyden ennustaminen neuroverkkomenetelmällä, lisensiaatintutkimus, Helsingin yliopisto, Taloustieteen laitos. Maatalouden liiketaloustiede. *Julkaisuja* 22.
- Oja, M., Kaski, S. & Kohonen, T. 2003.** Bibliography of self-organizing map (SOM) papers: 1998–2001 addendum. *Neural Computing Surveys* 3: 1–156.
- Pöllä, M., Honkela, T. & Kohonen, T. 2009.** Bibliography of self-organizing map (SOM) papers: 2002–2005 addendum. *TKK Reports in Information and Computer Science*, TKK-ICS-R23. Helsinki University of Technology.
- Rantala, O. & Tauriainen, J. 2017.** Maatalous- ja puutarhayritysten tulos- ja kannattavuuskehitys. Teoksessa: J. Niemi ja M. Väre (toim.) Suomen maa- ja elintarviketalous 2016/2017. s. 56–61.
- Sulkava, M., Sepponen, A.-M., Yli-Heikkilä, M. & Latukka, A. 2014.** Itseorganisoivan kartan ryvästys paljastaa maatilojen kannattavuusprofiilit. Teoksessa: Hakojärvi, M. & Schulman, N. (toim.). Maataloustieteen päivät 2014. *Suomen Maataloustieteellisen Seuran julkaisuja* nro 30.
- Sulkava, M., Sepponen, A.-M., Yli-Heikkilä, M. & Latukka, A. 2015.** Clustering of the self-organizing map reveals profiles of farm profitability and upscaling weights. *Neurocomputing* 147: 197–206.
- Sulkava, M., Yli-Heikkilä, M. & Latukka, A. 2012.** Analysis of farm profitability and the weighted upscaling system using the self-organizing map. Teoksessa: Estévez, P.A., Principe, J.C. & Zegers, P. (toim.). Advances in Self-Organizing Maps: 9th International Workshop, WSOM 2012, Proceedings. *Advances in Intelligent Systems and Computing* 198: 285–294.
- Yli-Heikkilä, M., Tauriainen, J. & Sulkava, M. 2018.** Prediktiivinen analyysimenetelmä tilan kannattavuuden laskemiseksi Taloustohtorissa. Teoksessa: Hakojärvi, M. & Puhakainen, T. (toim.). Maataloustieteen päivät 2018, *Suomen Maataloustieteellisen Seuran julkaisuja* nro 35.