

# Itseorganisoituva kartta (SOM) ja muut koneoppimismenetelmät Lukessa

Luontotietomassat hyötykäyttöön – työpaja 16.3

# SOM Lukessa

- Monissa tutkimuksissa on lähdetty liikkeelle eksploratiivisesti: on olemassa jokin vajavaisesti tunnettu moniulotteinen aineisto → voisi löytyä kiinnostavia ilmiöitä?
- SOM tarjonnut mahdollisuuden helposti tulkittaviin visualisointeihin ja tiivistelmiin datan suurista linjoista. Tyypillisessä SOM-analyysiin päätyneessä datassa on paljon havaintoja, monta muuttujaa ja lisäksi sekä aika- että paikkadimensiot.
- SOM:ia on käytetty esimerkiksi maatalouden kannattavuuskirjanpitoaineiston tutkimiseen, ja metsäntutkimuksen puolella SOM on ollut käytössä CO<sub>2</sub>-vaihdon ja havunneulasten alkuainepitoisuuksien kehityksen tutkimuksessa. Lisäksi menetelmää on käytetty mm. ajoneuvokatsastus- ja rakennusten energiankulutusaineistojen analyysiin.

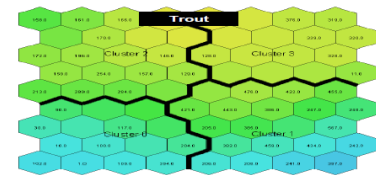
# SOM Lukessa

- Datan esikäsittely → SOM:n parametrien valinta → kartan opetus. SOM:in komponenttitasojen ja U-matriisin tarkastelulla löydetään yhteyksiä muuttujien välillä ja saadaan käsitys datan tiheydestä ja mahdollisesta klusterirakenteesta. Kartan kokoa vaihtelemalla päästään käsiksi eri mittaluokan ominaisuuksiin. Tehtyjen havaintojen perusteella voidaan tehdä esim. SOM-pohjainen klusterointi ja aikasarjatarkasteluja. Näin saadaan uusia hypoteeseja, ja voidaan jatkaa tilastoanalyysin keinoin.
- SOM on tehokas visualisointityökalu moniulotteisten ja suurten aineistojen tutkimiseen. Se tekee samaan aikaan epälineaarisen projektion tasoon ja vektorikvantisoinnin. Tämä tekee visualisoinneista helposti tulkittavia.
- Vaihtoehtoisesti olisi voinut käyttää SOM:in sukulaismenetelmiä tai perinteisempiä projektiomenetelmiä ja tarvittaessa yhdistää tulos esim. klusterointiin (kuten esim. k-Means)

# The impact of groundwater upwelling on the Tornionjoki trout

Atso Romakkaniemi, Pekka Jounela ja Olli van der Meer (2020)

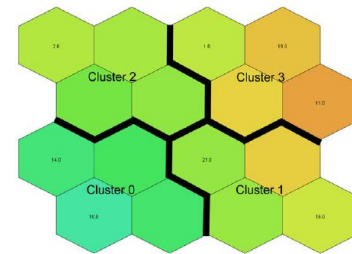
- Pohjavesipurkaumien vaikutusta kalojen (taimen, simput) esiintymiseen tutkittiin Valkeajoen ja Kuerjoen vesistöissä, jotka ovat Pohjois-Suomessa sijaitsevan Tornionjoen vesistöön kuuluvan Äkäs-joen sivuvesistöjä.
- Tutkimuspisteistä kerättiin habitaattidata-muuttujista, joiden ennakoita tiedetään vaikuttavan taimenen esiintymiseen. Lisäksi pohjaveden esiintymistä/purkautumista tutkimuspisteissä arvioitiin pohjautuen sekä lämpökameralla tehtyihin ilmakuvauskuviin, että mittaamalla jokiveden ja pohjasedimentin lämpötilaa.
- Taimenten ja simppejen esiintymistä suhteessa kartoitettuihin muuttujiin analysoitiin itseorganisoituvien karttojen (self-organizing map, SOM) ja Random Forest (RF) -menetelmällä.
- SOM-menetelmä visualisoi selitettävien muuttujien (kalojen esiintyminen) ja selittäjien välisiä yhteyksiä. RF-menetelmällä määritettiin ko. yhteyksien muotoja.
- Analyysitulosten mukaan taimenenpoikaset suosivat tutkimuspisteitä, joissa oli alhainen veden ja sedimentin lämpötila, suhteellisen hidas virtausnopeus, korkea happipitoisuus, lyhyt etäisyys joen rantaviivaan, matala veden syvyys ja vesikasveja paljas pohja.
- Tulokset tukevat kirjallisuudessa esitettyjä yleisiä taimenen habitaattipreferenssejä, mutta uutena tuloksena havaittiin taimenen suosivan myös pohjaveden kumpuamispaikkoja.



# Beliefs in Conflict: The Management of Teno Atlantic Salmon in the Sámi Homeland in Finland

Juha Hiedanpää, Joni Sajjets, Pekka Jounela, Mikko Jokinen ja Simo Sarkki (2020)

- Tenojoki on yksi merkittävimmistä kutujoista Atlantin lohelle Euroopassa. Vuonna 2009 tutkimus ilmoitti, että Tenon lohikanta oli heikossa tilassa, ja huoli Atlantin lohen tulevaisuudesta Tenojossa nousi joen molemmille puolille, Suomeen ja Norjaan.
- Vuonna 2017 hallitukset hyväksyivät uuden Tenon kalastuksen sopimus (Teno Fishing Act 2017). Sopimuksella pyrittiin vähentämään kalastusmääriä 30% ja uudet määräykset koskee kaikkia käyttäjiä, mukaan lukien alkuperäiskansalaiset saamelaiset, muut paikalliset, turistit ja kalastusyrittäjät.
- Tämä laukaisi huolta ja vihaa saamelais- ja muissa paikallisissa yhteisöissä. Kiista nosti esiin kysymyksen Tenojoen lohen hoidosta.
- Tehtiin Q-kysely, jossa oli 43 lausuntoa kattaen Tenojoen loheen liittyvän kiinnostuksen, tiedon, hallinnan ja poliittiset tarpeet.
- Käyttämällä itseorganisointia karttoja (SOM), tutkimusaineistoista löydettiin neljä ihmisryhmää, jotka painottivat eri asioita: perinteinen saamelaiskalastus; lohen suojelu; taloudellinen tilaisuus; ja todistusaineistoon perustuva päätöksenteko.

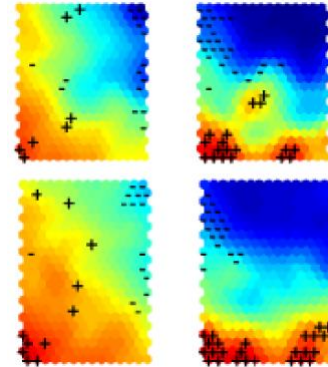


5. The scientific salmon knowledge is unbiased.

# Photosynthesis drives anomalies in net carbon-exchange of pine forests at different latitudes

S. Luyssaert, M. Sulkava, et al. (2007)

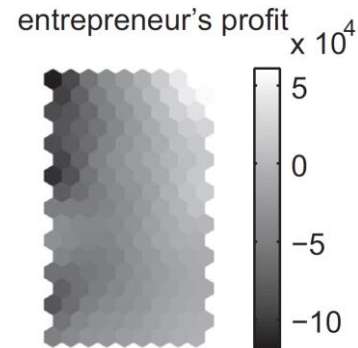
- Ilmakehän hiilidioksidin kasvunopeudella on suuria ajallisia vaihteluita, joka määrittynyt suurelta osin maan ja ilmakehän hiilidioksidi-virtausten vuotuisten vaihteluiden avulla. Tätä maa-ilmakehän hiilidioksidivirtausta ohjaa laajamittainen biomassan palaminen ja vaihtelut nettoekosysteemin vaihdossa (net ecosystem exchange). Vuosien kuluessa ja vuosien sisällä NEE vaihtelee ilmaston vaihtelujen vuoksi.
- Tutkimukset ilmastovaikutuksista bruttofotosynteesin (gross photosynthesis) vuosien välillä ja sisällä tapahtuvaan vaihteluun ja nettohiilenottoon maanpäällisissä ekosysteemeissä ovat osoittaneet ristiriitaisia tuloksia. Nämä ristiriidat liittyvät osittain käytettävien menetelmien eroihin ja joidenkin tutkimusten rajoitettuihin kestoisiin.
- Tässä työssä esitellään havainnointiin perustuva menetelmä, joka tarjoaa käsityksen hiilidioksidivirtausten poikkeavuuksien riippuvuudesta ilmasto-olosuhteista.
- Menetelmää sovellettiin boreaalisen ja kahden lauhkean mäntymetsän virtauksiin.
- Tutkimuksemme korosti selvästi tarvetta käyttää säämalleja (weather pattern) yksittäisten ilmastomuuttujien sijaan hiilidioksidin virtausten poikkeamien ymmärtämisessä.
- Kesän ilmastolliset ääriolosuhteet esiintyivät useammin ja voimakkaammin etelässä kuin pohjoisessa, ja niillä oli suuremmat vaikutukset etelässä, koska fotosynteesin estämisen kriteerit täyttyivät useammin.



# Clustering of the self-organizing map reveals profiles of farm profitability and upscaling weights

Mika Sulkava et al. (2014)

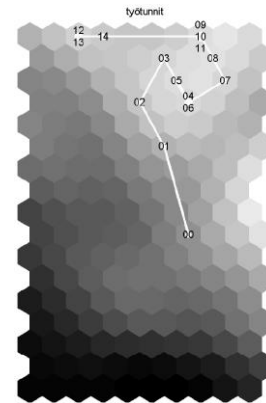
- Analysoidaan maanviljelyn kannattavuutta ja muita taloudellisia näkökulmia Suomessa itseorganisoituvan kartan avulla.
- Kannattavuuden kirjanpidon analyysi paljastaa useita mielenkiintoisia suhteita taloudellisten muuttujien välillä.
- Maatilojen taloudelliset profiilit esitetään klusterointitulosten perusteella ja löydökset on vahvistettu tilastollisilla testeilla.
- Painojen optimointiprosessi esitetään työssä, jolla maatilojen taloudelliset kannattavuusluvut saadaan yleistettyä näytteiden tasolta koko maan tasolle.
- Paino-optimointiprosessin avaintekijäksi osoittautui viljelymaan käyttöaste.



# Yksittäisten maatilojen tarkastelu korkeaulotteisessa talousavaruudessa itseorganisoivan kartan avulla

Mika Sulkava ja Maria Yli-Heikkilä (2018)

- Luonnonvarakeskuksen (Luke) Tilastopalvelut-yksikkö tuottaa kannattavuuskirjanpidossa monipuolista tietoa Suomen maa- ja puutarhatalouden taloudellisesta tilanteesta.
- Suomen maa- ja puutarhatilojen kannattavuutta ja muita taloudellisia tekijöitä kuvataan tässä työssä itseorganisoivan kartan avulla.
- Tutkimuksessamme opetamme kannattavuuskirjanpitoaineiston pohjalta laskettujen keskeisimpien talousmuuttujien avulla itseorganisoivan kartan.
- Aikaisemmassa tutkimuksessa olemme löytäneet tällä menetelmällä useita kiinnostavia yhteyksiä kerättyjen taloudellisten muuttujien välillä.
- 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.





# Loppusanat

- Lukessa hyvin laaja kirjo itseorganisoituvan kartan sovelluksista  
→ kätevä työkalu tiivistämään ja klusteroimaan ilmiöitä moniulotteisissa aineistoissa.
- Itseorganisoituvan kartan lisäksi Lukessa käytetään monia muitakin koneoppimismenetelmiä, kuten random foresteja, tukivektorikoneita ja k-lähimmän naapurin menetelmää.
- Syväoppimismalit (konvoluutioneuroverkot) nousevassa roolissa esimerkiksi kohteiden tunnistuksessa riistaeläinaineistoissa → "Replacing human eye with AI"-projekti (2021-2022) syväoppimispainoiteinen projekti, joka keskittyy luonnonvara-aineistojen automaattiseen analysointiin.

# Kiitos!