



Aalto-yliopisto  
Perustieteiden  
korkeakoulu

# Lajittelumenetelmät ilmakehän kaukokartoituksen laadun tarkkailussa (valmiin työn esittely)

*Viivi Halla-aho*

*30.9.2013*

*Ohjaaja: Dos. Johanna Tamminen*

*Valvoja: Prof. Harri Ehtamo*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

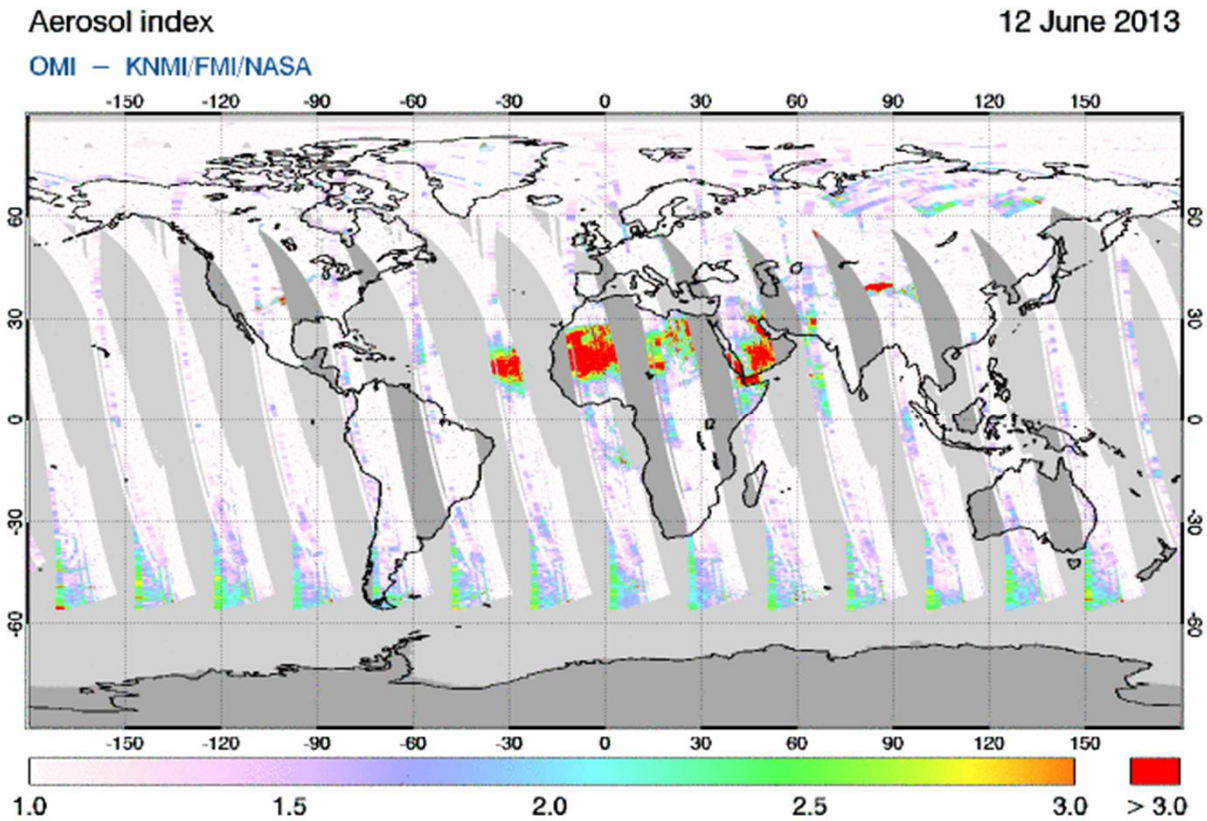
# Sisältö

- Tausta
- Tavoitteet ja rajaukset
- Käytetyt menetelmät
  - Hierarkkinen ryhmittely
  - Reunojen tunnistus Sobel-operaattorilla
  - Autokorrelaatioiden analysointi
- Tulokset
- Tuloksien yhteenveto ja päätelmiä
- Lopuksi

# Tausta

- OMI (Ozone Monitoring Instrument) on Aura-satelliitin instrumentti, josta saadaan pääasiassa otsonituotteita, mutta myös tietoa mm. NO<sub>2</sub>- ja SO<sub>2</sub>-kaasuista, pilvistä, aerosoleista ja UV-valosta.
- Aerosolituotteessa vuodenaikojen mukaan paikkaa vaihtava anomalia etelässä tai pohjoisessa, joita ei ole poistettu Level 2 -tasoista tuotteista.
- Ongelma johtuu vaikeista mittausolosuhteista, joissa auringon tulokulma (*atsimuuttikulma*) kasvaa liian suureksi.

# Tausta



Esimerkkikuva anomaliasta. Kuvassa on esitetty aerosolimittaukset päivältä 12.6.2013. Kuvan lähde: temis.nl

# Tavoitteet ja rajaukset

- Poistaa **jälkikäteen** selkeät virheet niin, että mahdollisimman vähän 'oikeaa' dataa menee hukkaan.
- Aerosolituotteiden lisäksi muiden tuotteiden (kuten SO<sub>2</sub>) vastaavien ongelmien korjaus. Menetelmän tulee siis sopia monenlaisen datan käsittelyyn.
- Tutkitaan erilaisia mahdollisia menetelmiä ja kokeillaan niitä aineistoon.
- Tarkoituksena on nimenomaan poistaa anomalia jo valmiista kuvasta, eikä puuttua satelliitti-instrumentin antamien mittausten käsittelyyn.
- Riittää, että kuvasta saadaan vähemmän harhaanjohtava.

# Testatut menetelmät

- Hierarkkinen ryhmittelyanalyysi
- Reunojen tunnistus (*edge detection*) Sobel-operaattorilla
- Autokorrelaatioiden analysointi

# Hierarkkinen ryhmittely

- Etsitään havaintojen ominaisuuksien perusteella ryhmät ja jaetaan havainnot niihin.
- Tavoite: samassa ryhmässä olevat havainnot mahdollisimman samanlaisia keskenään ja vastaavasti eri ryhmissä olevat havainnot mahdollisimman erilaisia keskenään.
- Erilaisuutta kuvataan **etäisyysmitalla**.
- Ryhmittelyn vaiheet
  - Aluksi kaikki havainnot omia ryhmiään
  - “Lähimmät” havainnot yhdistetään
  - Toistetaan kunnes kaikki samaa ryhmää
  - Katkaistaan ryhmittely vaiheessa, jolloin ryhmiä on haluttu määrä
  - Voidaan tehdä myös toisin päin, eli ensin kaikki samaa ryhmää ja lopuksi kaikki erillään
- Hyvän ryhmittelyn tunnistaa siitä, että klusterit erottuvat selvästi toisistaan ja niille voidaan keksiä mielekäs tulkinta.

# Reunojen tunnistus Sobel-operaattorilla

- Yritetään löytää kuvasta reunoja, eli intensiteetin epäjatkuvuuskohtia.
- Sobel-operaattori pohjautuu gradientin approksimaatioiden laskemiseen kullekin pisteelle.
- Kynnysarvon ylittävät gradientin magnitudin arvot tulkitaan reunoiksi.



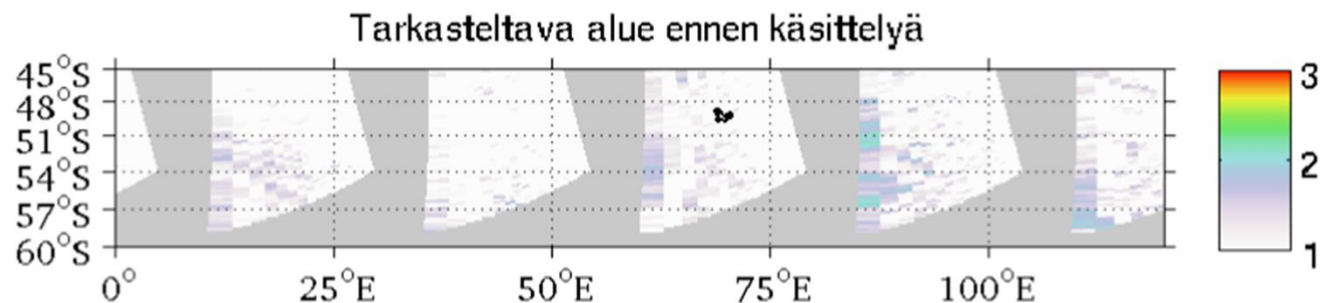
# Autokorrelaatioiden analysointi 1/2

- Autokorrelaatioita käytetään usein aikasarja-analyysissä kuvaamaan havaintojen riippuvuutta toisistaan havaintojen välisen aikaeron funktiona.
- Kuinka soveltaa ajatusta satelliittidataan?
  - Ajatellaan dataa leveysasteiden suuntaisina riveinä
  - Anomalia toistuu joka *orbitilla* tietyssä kohtaa riviä
  - Voidaan laskea autokorrelaatiot rivin havainnoille. Aikaeron sijaan käytetään havaintopikselien indeksien erotusta muuttujana.

# Autokorrelaatioiden analysointi 2/2

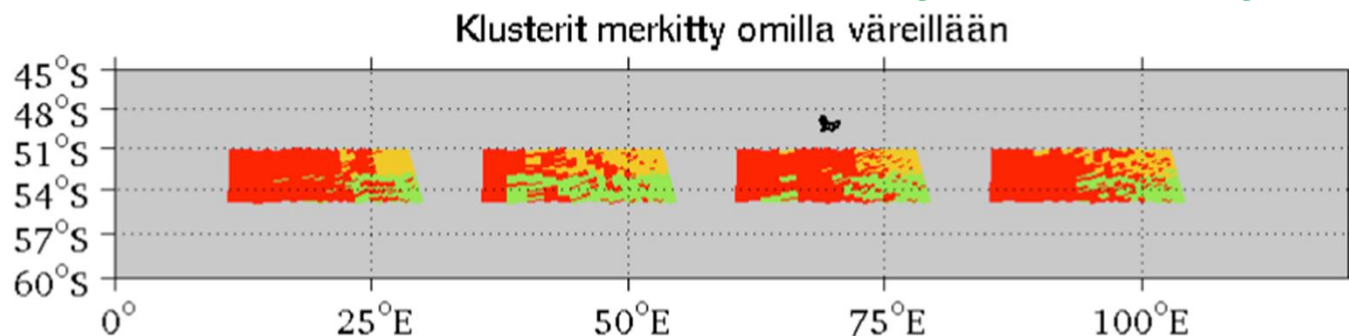
- Mitä autokorrelaatiot kertovat?
  - Autokorrelaatio kertoo, kuinka rivin havainnot riippuvat toisistaan indeksien erotuksen funktiona.
  - Mikäli autokorrelaatiot ovat vahvimmissaan aina *orbitin* leveyden välein, on kyse etsitystä anomaliasta.
- Merkitään virheellisiksi rivit jotka täyttävät seuraavat ehdot:
  - Rivillä on tarpeeksi suuria, kynnysarvon ylittäviä autokorrelaatioita tarpeeksi monta kappaletta
  - Autokorrelaatioiden tulee ilmetä indeksin arvoilla, jotka ovat jaollisia *orbitin* leveydellä (60 pikseliä)

# Tulokset: hierarkkinen ryhmittely 1/2



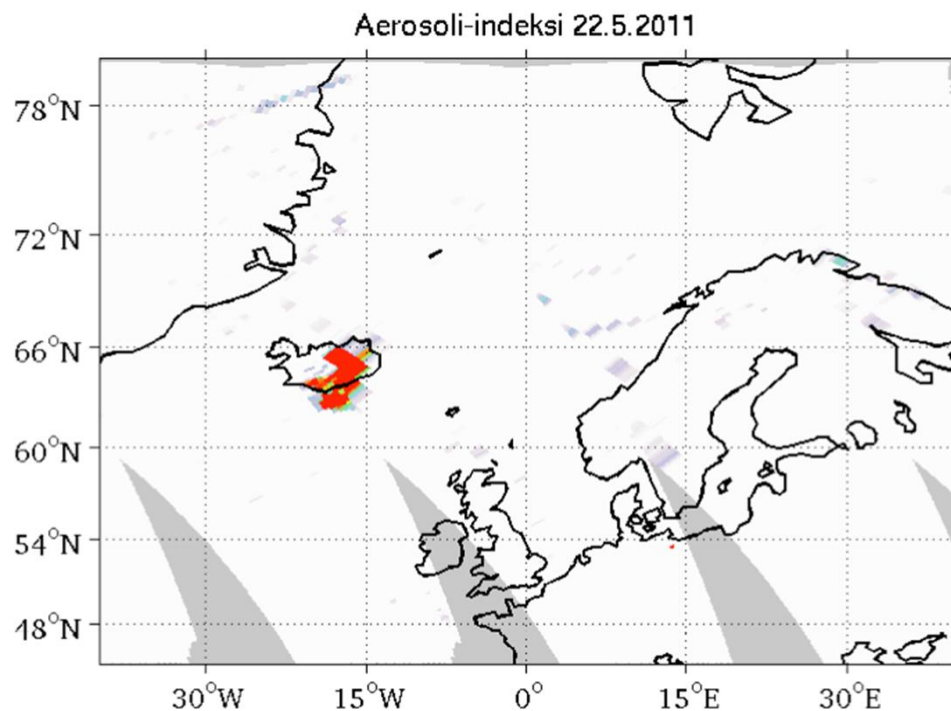
- Ryhmitellään OMAERUV-tuotetta päivältä 15.4.2013. Valitaan testihavainnoiksi pieni otos (noin 8300 havaintoa)
- Ryhmittely tehdään seuraavien ominaisuuksien perusteella:
  - Aerosoli-indeksi
  - Auringon atsimuuttikulma
  - Latitudi
  - Havainnon rivi-indeksi
- Käytetään R-ohjelmiston *fastcluster*-paketin *hclust*-funktiota ryhmittelyyn. Havaintoarvot standardoidaan ennen toimitusta vääristymien välttämiseksi. Valitaan ryhmien lukumääräksi 5.

# Tulokset: hierarkkinen ryhmittely 2/2



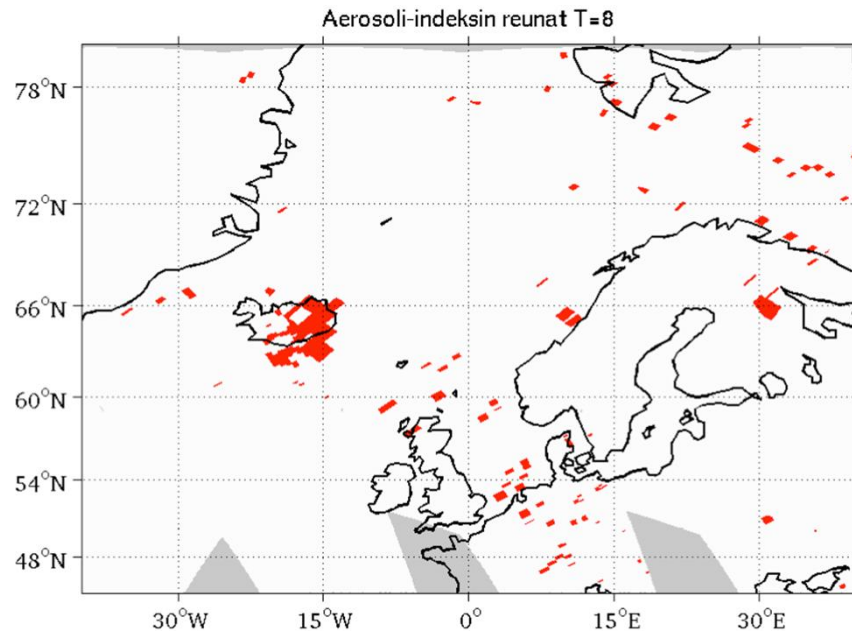
- Saadaan 5 klusteria, jotka on esitetty kartalla omilla väreillään. Punainen vastaa klusteria 1, oranssi klusteria 2, vihreä klusteria 3, turkoosi klusteria 4 ja vaaleansininen klusteria 5. Vain kolme ensimmäistä erottuvat kuvassa.
- Graafien lisäksi tutkittiin klustereiden ominaisuuksia tilastollisin tunnusluvuin.
- Klusteri 1: Suuret aerosoli-indeksin arvot, suurehko auringon atsimuuttikulma, mahdollinen 'virheryhmä'?
- Klusterit 2 ja 3: Pienet tai puuttuvat aerosoli-indeksit, hieman pienemmät auringon atsimuuttikulmat.
- Klusterit 4 ja 5: Kaikki aerosoli-indeksit puuttuvat, suuret auringon atsimuuttikulmat.
- Menetelmä vaatii kuitenkin aina jonkun tekemään tulkinnot ryhmille ja tekemään päätöksiä, eikä tee tätä itsenäisesti. Kattava päätöksentekoaalyysi, joka toimisi kaikenlaiselle datalle voisi olla haastavaa tehdä.

# Tulokset: reunojen tunnistus 1/2



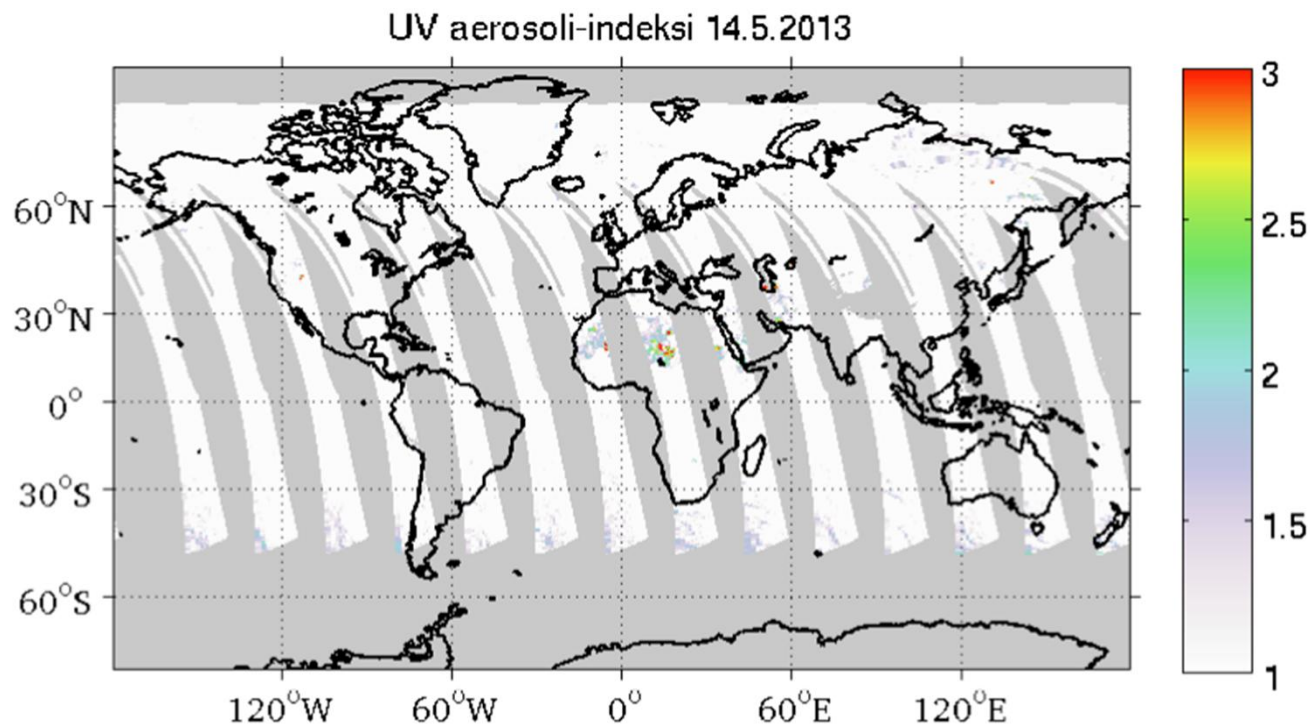
- Aineistona päivän 22.5.2011 aerosolidata
- Etsitään kuvasta tulivuorenpurkauksen aiheuttama tuhkapilvi .
- Valitaan kynnyksarvoksi  $T=8$ 
  - Liian pieni kynnyksarvo päästää läpi pienetkin havainnot
  - Liian suuri kynnyksarvo hävittää kaiken
  - Testien perusteella  $T=8$  aika hyvä tähän tarkoitukseen

# Tulokset: reunojen tunnistus 2/2



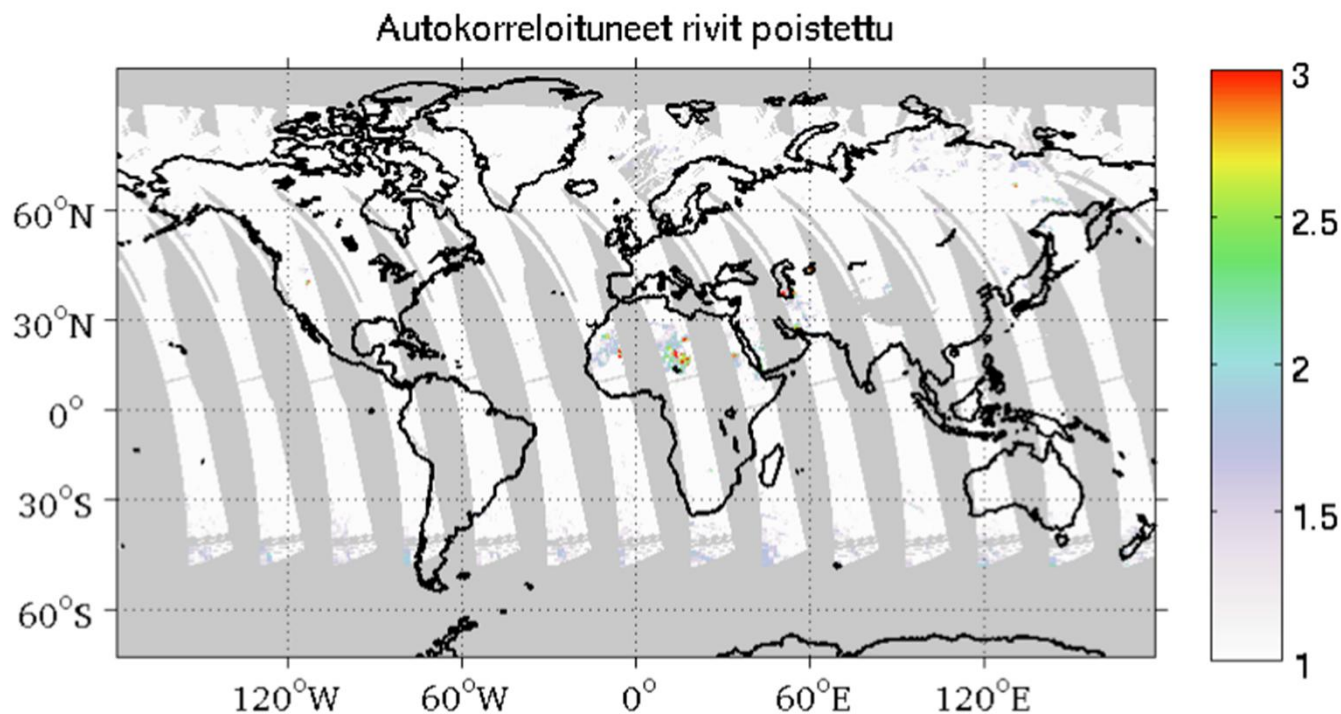
- Kuvassa esitetty löydetyt aerosoli-indeksin reunat.
- Menetelmä ei ota kantaa siihen, onko löydetty alue anomalia vai ei.
- Ei tuo ratkaisua ongelmaan.

# Tulokset: autokorrelaatioanalyysi 1/2



- Analysoidaan autokorrelaatioita päivän 14.5.2013 UV aerosoli-indeksidatalle.
- Vaaditaan vähintään 4 kynnysarvon 0,4 ylittävää autokorrelaation arvoa anomaliaa sisältävältä riviltä.

# Tulokset: autokorrelaatioanalyysi



- Autokorreloituneet rivit poistettu kuvasta.
- Tulos ei ole täydellinen eli menetelmä ei löydä kaikkia selkeitä virheitä.
- Menetelmä on herkkä epäsäännöllisyydelle.



# Tuloksien yhteenveto ja päätelmiä

- Hierarkkinen ryhmittely
  - Liian hidas! Vaatii paljon raskaita laskentaoperaatioita.
  - Tuloksien hyödyntäminen vaatii päätöksenteon automatisointia.
- Reunojen etsintä
  - Ei tee sitä mitä halutaan, voi toki olla muuten hyödyllinen.
- Autokorrelaatioiden analysointi
  - Tulokset ovat lupaavia!
  - Täyttää useimmat vaatimuksista.
  - Ei vielä kovin hienovarainen keino käyttää ak:ta hyväksi.

# Lopuksi

- Kysymyksiä, kommentteja?