

TYÖSELOSTUS:

TIEDONLOUHINTAA MICROSOFT EXCELILLÄ ILMATIETEEN LAITOKSEN AVOIMESTA TIETOAINIESTOSTA

Risto Hietala

1503520 / risto.hietala@eng.tamk.fi

Tietotekniset työkalut, 5S00BL76

29.8.2015

Insinööri (AMK),

Sähkö- ja automaatiotekniikka,

Monimuotototeutus

1 Johdanto

Tämä työselostus on tehty Tietotekniset työkalut –kurssille, ja sen tarkoituksena on esitellä avoimen datan ja Microsoft Office –ohjelmien käyttöä.

Tiedonlouhinta (eng. data mining) on tällä hetkellä kasvava ala sekä tieteessä että yritysmaailmassa. Tästä tehtävänannosta sain hyvän syyn selvittää, kuinka hyvin Microsoft Excel soveltuisi tiedonlouhintaan suurille tietomäärille. Valitsin tietoaaineistoksi Ilmatieteen laitoksen tuottaman auringon globaalisäteilyn¹.

2 Tiedon lataus ja muokkaus

Ilmatieteen laitoksen avoimen datan rajapinta ei ole helposti lähestyttävä ja pääasiallinen ohjeistus *Latauspalvelun pikaohje*² on sekava. Alkuun pääsemisessä suureksi avuksi olikin Matias Arpikarin blogikirjoitus *Ilmatieteen laitoksen avoimen datan hyödyntäminen*³.

Palvelu toteuttaa OpenGIS Web Feature Service 2.0 Interface –standardin (OGC 09-025r1 tai ISO/DIS 19142)⁴ WFS Simple –profiilin ja käyttää tiedon siirtoon XML:ää.

Standardinmukaisuus on tietysti hyvä asia, mutta valmiiden työkalujen puuttuessa voi hidastaa tekemistä aluksi koska aikaa kuluu itse asian lisäksi myös standardin tulkitsemiseen.

Ilmatieteen laitoksen ohjeistuksessa mainitaan, että tietoa pystyisi lataamaan suoraan Excelillä, mutta tähän tarvisi todennäköisesti Excelin WEBSERVICE-funktiota, joka ei ole tuettu käyttämässäni Excel 2016 for Mac –versiossa⁵.

Tiedon lataamista ja Excelille muokkaamista varten tarvitsi siis tehdä erillinen ohjelma.

2.1 OHJELMA

Valitsin harjoituksen vuoksi latausohjelman kieleksi Clojuren⁶. Valmis ohjelma on saatavilla Githubissa⁷ sekä liitteessä 1. Esimerkki tuotetusta CSV-tiedostosta on liitteessä 2.

Ohjelma kaivaa datasta eri mittauspaikkojen nimet, koordinaatit sekä mittauksilukokset ja muokkaa näistä Excelille soveltuvan CSV-tiedoston. Ohjelman tekeminen oli mukava sormiharjoitus ja Clojure-muistojen verestäminen, mutta ei varsinaisesti tämän työn pääasia, niin en paneudu siihen tämän enempää. Koodin kommentteista saattaa selvitä sen toimintalogiikka.

¹ <http://ilmatieteenlaitos.fi/avoin-data-saahavainnot#Aurinko>

² <http://ilmatieteenlaitos.fi/latauspalvelun-pikaohje>

³ <http://matias.biz/ilmatieteen-laitoksen-avoimen-datan-hyodyntaminen/>

⁴ http://portal.opengeospatial.org/files/?artifact_id=39967

⁵ <https://support.office.com/en-us/Article/WEBSERVICE-function-0546a35a-ecc6-4739-aed7-c0b7ce1562c4>

⁶ <http://clojure.org/>

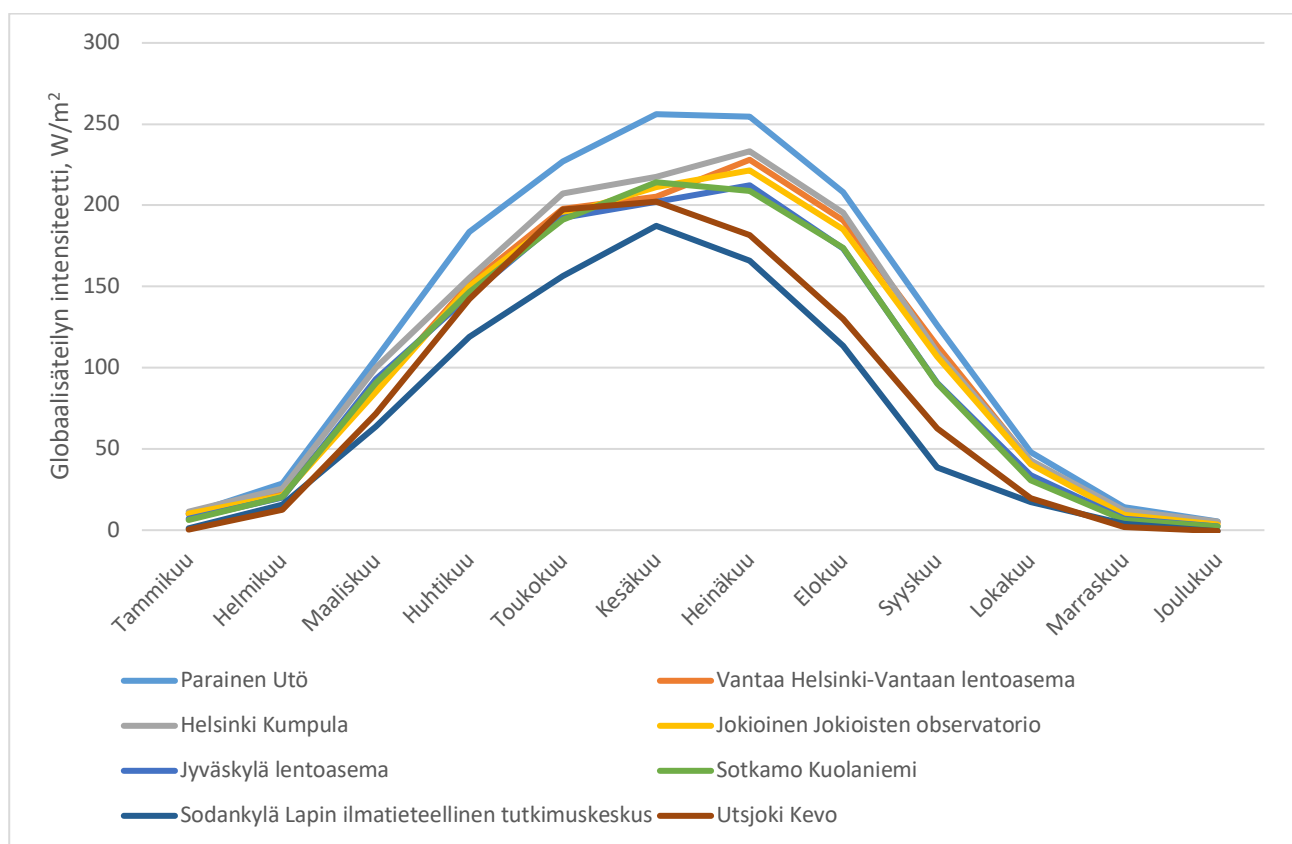
⁷ <https://github.com/rhietala/fmidownload>

Tiedoston viemisessä Exceliin täytyy muuttaa Import-toiminnon viimeisellä välilehdellä desimaalierotin pisteestä pilkuksi ja skandeja en saanut toimimaan oikein vaikka tiedoston merkistöksi valitsi UTF-8:n.

3 Keskiarvoja

Koko tietomäärän piirtäminen ei onnistu Excelillä, sillä mittausarvoja on noin 26000 jokaista mittauspistettä kohti, joten heti aluksi on tehtävä jotain tiedon koostamista. Yksinkertainen tapa on mittaustulosten keskiarvoistaminen jonkin aikaikkunan yli.

3.1 KUUKAUSITTAINEN KESKIARVO



KUVIO 1: Kuukausittainen keskiarvo

Kuviossa 1 esitetyt säteilyn kuukausittaiset keskiarvot saatiin AVERAGEIF-funktiolla:

`=AVERAGEIF(Data!D4:D26367;$A2;Data!H$2:H$26367)`

ja lasketut arvot on esitetty myös taulukossa 1. Lähdetauluun lisättiin kuukaudelle oma sarake MONTH-funktiolla mittauksen ajanhetkestä, jotta kuukausia pystyttäisiin vertailemaan keskenään.

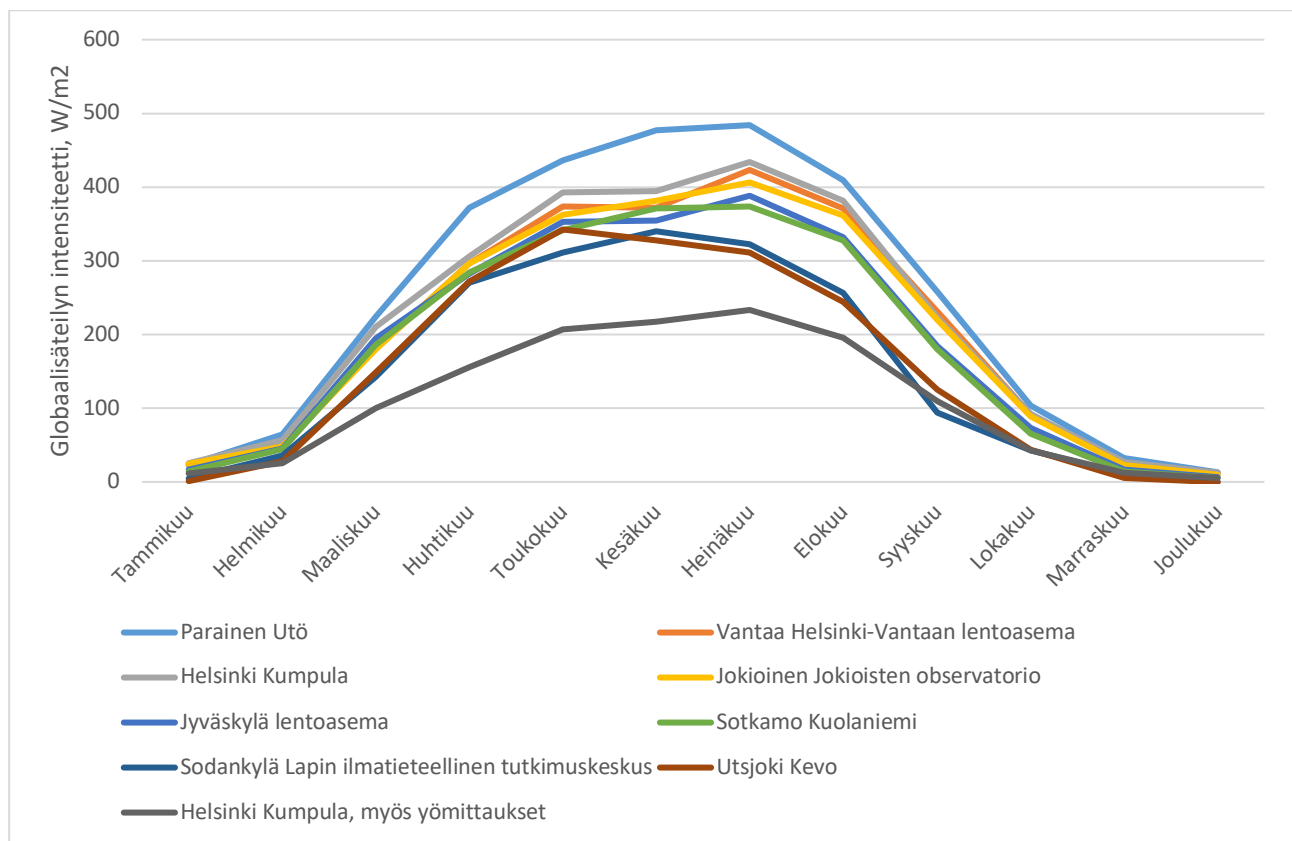
Kuviosta huomataan, että auringon säteilyn intensiteetti on suurin kesäkuukausina ja pienin talvella. Eri mittauspisteiden erot ovat myös havaittavissa: mitä pohjoisemmassa mittauspiste on, sitä pienempi on mitattu intensiteetti.

Keskiarvon laskeminen ei ota huomioon yö- ja päiväaikaa ja näin käyrien vertailu ei kerro koko totuutta: yöllä aurinko ei paista kuukaudesta tai mittauspaikasta riippumatta, mutta silti yölliset mittaustulokset ovat mukana lasketussa keskiarvossa. Pohjois-Suomen yöttömästä yöstä lisää kappaleessa 3.4.

TAULUKKO 1: Kuukausittainen keskiarvo

| | Parainen Utö | Vantaa Helsinki- Vantaan lentoasema | Helsinki Kumpula | Jokioinen | Jyväskylä lentoasema | Sotkamo Kuolaniemi | Sodankylä | Utsjoki Kevo |
|-----------|-----------------|--|---------------------|-----------|-------------------------|-----------------------|-----------|-----------------|
| Tammikuu | 9,94 | 10,02 | 11,38 | 10,20 | 6,87 | 6,21 | 1,26 | 0,33 |
| Helmikuu | 28,77 | 23,57 | 25,66 | 21,08 | 20,13 | 20,42 | 15,68 | 12,58 |
| Maaliskuu | 105,69 | 86,38 | 100,22 | 85,06 | 93,12 | 90,65 | 63,81 | 71,70 |
| Huhtikuu | 183,78 | 151,81 | 155,40 | 150,15 | 144,38 | 146,57 | 118,83 | 142,20 |
| Toukokuu | 227,14 | 197,97 | 207,23 | 193,84 | 192,21 | 191,19 | 156,49 | 197,44 |
| Kesäkuu | 256,11 | 205,41 | 217,65 | 211,21 | 202,34 | 214,20 | 187,36 | 202,26 |
| Heinäkuu | 254,56 | 228,04 | 233,26 | 221,49 | 212,32 | 208,81 | 165,97 | 181,61 |
| Elokuu | 207,88 | 190,75 | 195,35 | 185,29 | 173,22 | 173,92 | 113,53 | 129,84 |
| Syyskuu | 125,96 | 113,54 | 109,40 | 106,64 | 90,79 | 90,20 | 38,43 | 62,48 |
| Lokakuu | 47,90 | 42,26 | 42,94 | 40,61 | 34,03 | 30,75 | 17,20 | 19,84 |
| Marraskuu | 14,26 | 11,11 | 12,11 | 9,43 | 6,91 | 6,28 | 3,93 | 1,94 |
| Joulukuu | 5,37 | 4,74 | 5,25 | 3,46 | 2,28 | 2,24 | -0,67 | -0,54 |

3.2 KUUKAUSITTAINEN KESKIARVO, VAIN VALOISAN AJAN TUNNIT



KUVIO 2: Kuukausittainen keskiarvo, vain valoisan ajan tunnit

Kuukausikeskiarvo vain päivätunneilta laskettuna saatiin AVERAGEIFS-funktiolla:

=AVERAGEIFS(Data!H\$4:H\$26367;Data!\$D\$4:\$D\$26367;\$A16;Data!\$G\$4:\$G\$26367;TRUE)

ja lähdedataan lisättiin sarake päivä/yö-tiedolle funktiolla:

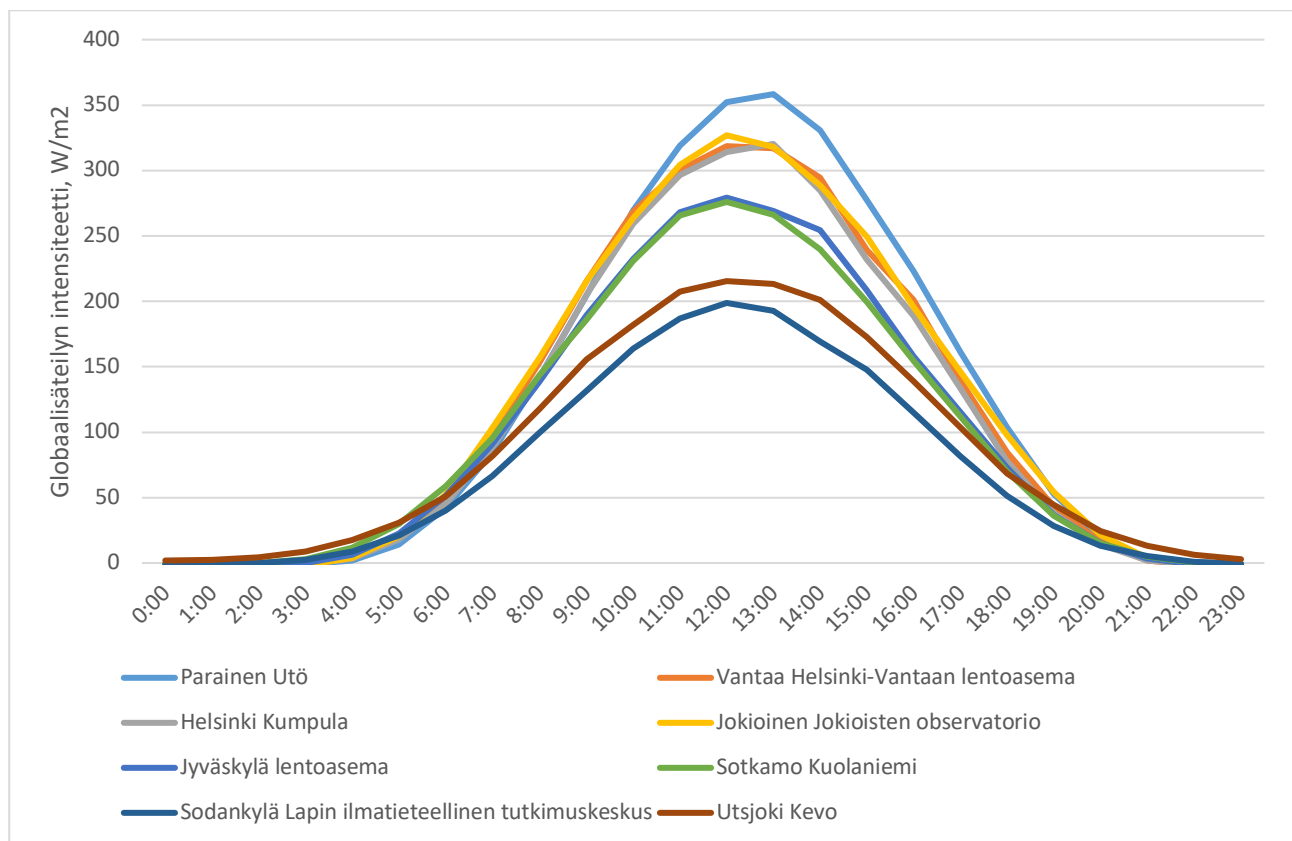
=AND(HOUR(A4)>=7;HOUR(A4)<=17)

Oudosti edellisessä luvussa käytetyn AVERAGEIF-funktion parametrien järjestys on eri kuin AVERAGEIFS (edellisessä ensimmäinen parametri on ehdollisuustarkastuksessa käytettävä data, jälkimmäisessä ensimmäinen parametri on keskiarvon laskentaan käytettävä data).

Yötuntien poistaminen vaikutti oleellisesti keskiarvoihin, kuten kuviosta 2 voidaan huomata.

Vertailukohtana käytetty Helsingin Kumpulan mittauskeskiarvot ovat noin kaksi kertaa isommat kun keskiarvo lasketaan vain päivätunneilta, mikä on tietysti ymmärrettävää.

3.3 TUNTIKESKJARVO

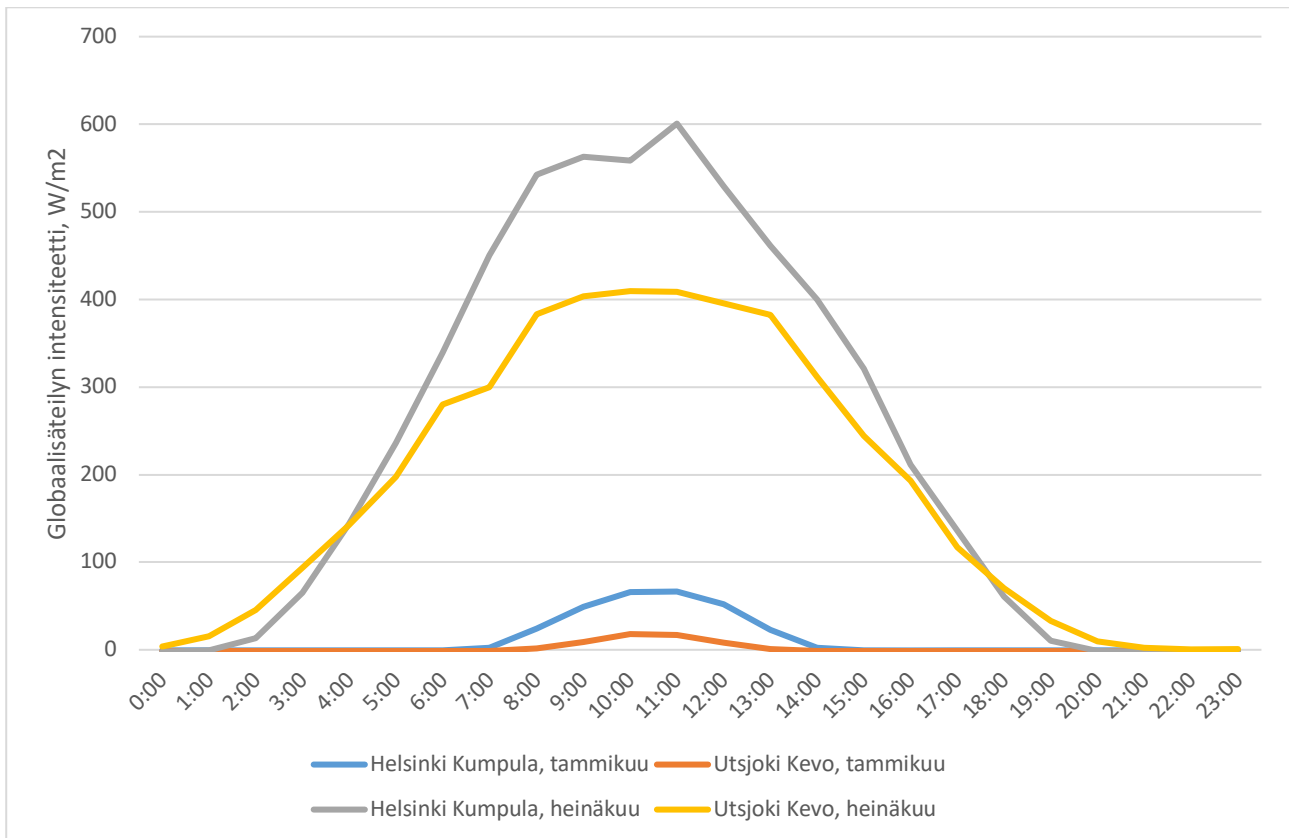


KUVIO 3: Tuntikeskiarvo

Kuviossa 3 on esitetty säteilyintensiteetin keskiarvot tunneittain eri mittauspisteille. Luonnollisesti keskipäivällä intensiteetti on suurimmillaan ja keskiyöllä pienimmillään. Laskentaan on käytetty AVERAGEIF-funktiota:

=AVERAGEIF(Data!\$F\$4:\$F\$26367;\$B2;Data!H\$2:H\$26367)

3.4 TUNTIKESKIARVO TAMMI- JA HEINÄKUUSSA



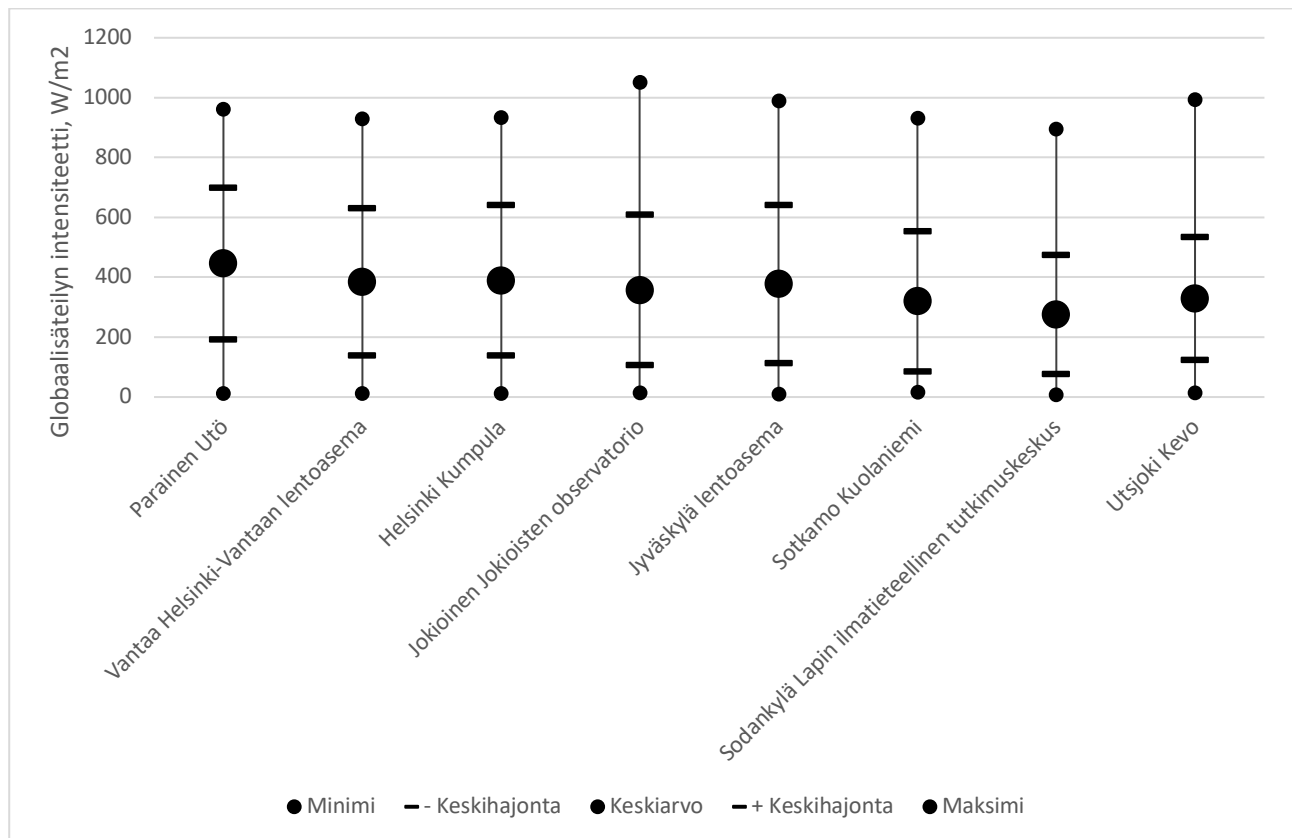
KUVIO 4: Tuntikeskiarvo tammi- ja heinäkuussa

Edellisen luvun tuntikeskiarvot laskettiin kaikista mittauksista, kuvio 4 esittää kesä- ja talvituntien eron. Kuviosta näkee myös Pohjois-Suomen yöttömän yön: illalla ja aamuyöstä mitattu intensiteetti on heinäkuussa suurempi Utsjoella kuin Helsingissä. Auringon säteilyintensiteetti on kesällä noin kuusinkertainen talveen verrattuna.

Laskemiseen käytettiin AVERAGEIFS-funktiota:

=AVERAGEIFS(Data!J\$4:J\$26367; Data!\$D\$4:\$D\$26367; 1; Data!\$F\$4:\$F\$26367; \$B3)

4 Tilastollisia suureita



KUVIO 5: Intensiteetin keskiarvo heinäkuussa 2015 päivällä (klo 7-17)

Tähän asti kaikki kuvioden laskennat ovat käyttäneet pelkkää keskiarvoa ja sen laskemiseen käytettävien mittaustulosten valintaa. Tuloksista voi toki laskea muitakin tilastotieteellisiä suureita, kuten minimin, maksimin, keskihajonnan ja varianssin. Kuviossa 5 on esitetty heinäkuun 2015 valoisian ajan mittaustuloksista keskiarvo, keskihajonta keskiarvon ympärillä sekä minimi ja maksimi eri mittauspisteille. Mitään oleellisia johtopäätöksiä kuvioista on ei löydy.

Aikaisemmin käytetyn AVERAGEIF-funktion perusteella oletin, että muillekin vastaaville funktioille olisi ehdolliset versiot, mutta olin väärässä. Pitkän pätkäilyn jälkeen löysin tietokantafunktiot (database functions)⁸, joista löytyy keskiarvolle DAVERAGE, minimille DMIN, maksimille DMAX, varianssille DVAR, ja niin edelleen. Funktioiden käyttölogiikka tosin on täysin erilainen kuin AVERAGEIF:llä, niin toinen hyvä tovi vierähti yrittäessä, erehtyessä ja ohjetta lukiessa. Tietokantafunktioiden laskentaan valittavien alkioden ehdot kirjoitetaan taulukkoon eikä funktiokutsuun, mikä sopii hyvin eri ehtojen kokeiluun myös funktiokutsuja osaamattomille. Kuvion 5 tuloksien laskennassa käytetty ehtotaulu on esitelty taulukossa 2. Valittu toteutus tosin rajoittaa paljon käyttöä: esimerkiksi jos kuvion 5 luvut haluaisi yhdestä mittauspisteestä joka kuukaudelle samaan kuvioon, niin taulukon 2 kaltaisia ehtotauluja tarvisi yhden jokaiselle kuukaudelle, sillä yhdet ehdot tuottavat vain yhden tulokset.

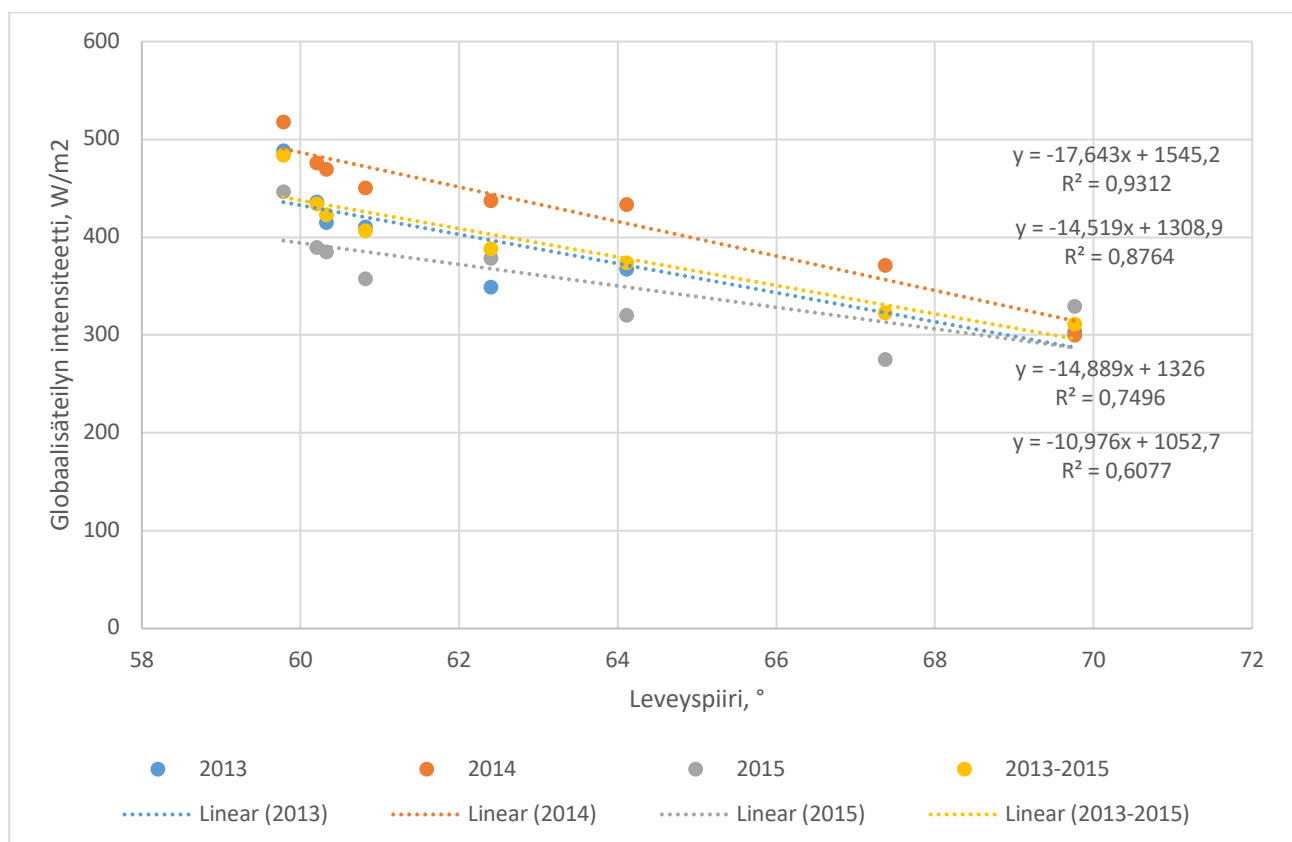
⁸ <https://support.office.com/en-au/article/Database-functions-reference-ad87e69b-fc20-4d3d-9d52-d7dc023f5c23>

TAULUKKO 2: Kuvion 5 tuloksien laskennassa käytetty ehtotaulu

| Paiva | Kuukausi | Vuosi |
|-------|----------|-------|
| =TRUE | =7 | =2015 |

Keskustelupalstojen perusteella saman ongelman voi ratkaista myös Excelin taulukkokaavoilla (array formula)⁹, joilla edellisessä kappaleessa esitettyä rajoitetta ei ehkä ole, mutta niiden tarkastelu ei mahtunut tähän selostukseen.

5 Lineaarinen regressio



KUVIO 6: Leveyspiirin vaikutus säteilyintensiteettiin

Kuviossa 6 on esitetty leveyspiirin vaikutus heinäkuun aurinkoisen ajan keskimääräiseen säteilyintensiteettiin. Kuviossa on mittauskeskiarvojen lisäksi myös lineaariset regressiokäyrät, niiden yhtälöt ja determinatiokertoimet (R^2). Determinatiokertoimen arvo 1 tarkoittaa että regressiomalli selittää täysin tulosten vaihtelun, ja arvo 0 taas että pisteiden välillä ei ole lineaarista yhteyttä.¹⁰

⁹ <https://support.office.com/en-nz/article/create-an-array-formula-e43e12e0-afc6-4a12-bc7f-48361075954d>

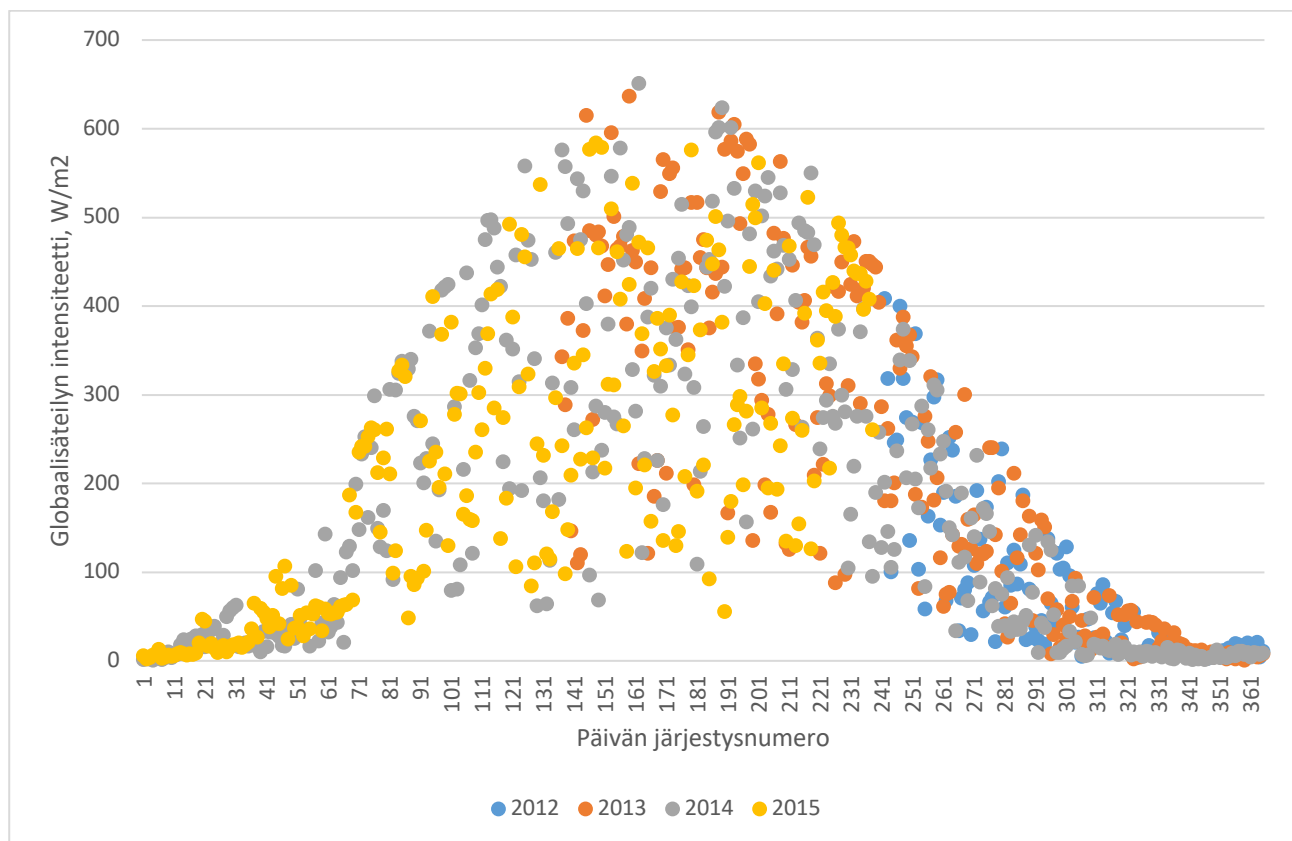
¹⁰ https://en.wikipedia.org/wiki/Coefficient_of_determination

Kuvion 6 kaikkien vuosien determinaatiokertoimen 0,876 voi tulkita siten, että 88% säteilyintensiteetin keskiarvojen vaihtelusta selittyy mittauspisteiden leveyspiirillä ja 12% muilla tekijöillä. Silmämääräisesti eteläisimmän mittauspisteen (Utö, Parainen) tulokset sopivat lineaariseen regressioon huonosti (tulokset ovat jokaisena vuonna regressiokäyrän yläpuolella), ja yksi mahdollinen selitys tälle voisi olla että ulkosaaristossa olisi heinäkuisin vähemmän pilviä kuin mantereella.

Regressiokäyrän yhtälöstä voidaan laskea arvio auringon säteilylle niille paikkakunnille, joilla ei ole omaa mittauspistettä. Tampereen koordinaatit ovat 61.4981503° N, 23.7610254° S, joten tämän yksinkertaisen mallin perusteella Tampereella pitäisi heinäkuinen auringon säteilyintensiteetti olla keskimäärin:

$$y = -14,519 \frac{W}{m^2} * 61,4981^{\circ} + 1308,9 \frac{W}{m^2} = 416,01 \frac{W}{m^2}$$

6 Epälineaarisen käyrän sovitus



KUVIO 7: Päiväkeskiarvo vuosittain Helsingin mittauspisteessä

Kuviossa 7 on esitetty päiväkeskiarvo vuosittain Helsingin mittauspisteessä valoisan ajan tunneilta. Vaikka kuviossa on tunnistettava hahmo, niin suuren päivästä toiseen vaihtelun takia kuvasta on vaikea tehdä ennusteita, tai edes katsoa kuinka suuri intensiteetti oli keskimäärin.

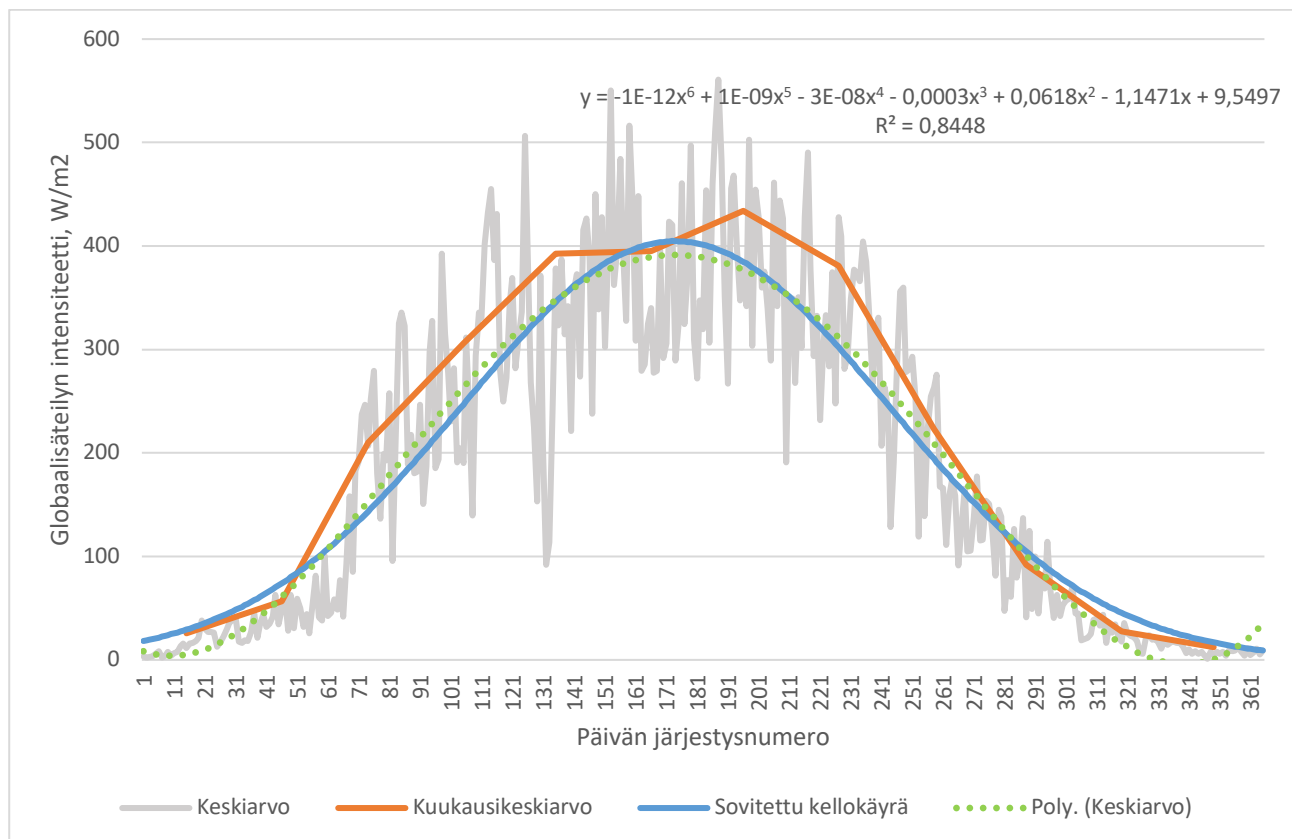
Aikaisemmin näistä mittaustuloksista onkin tehty kuukausittaiset keskiarvot, jotka ovat kuviossa 2. Kuviossa 6 taas käytettiin lineaarista regressiota, joka ei suoraan käy tähän ongelmaan, mutta Excelistä löytyy myös polynomi-regressiokäyrä, jolla voi mallintaa tämän tyyppisiä tuloksia. Kolmas mielenkiintoinen vaihtoehto

on Solver-lisäosan käyttö sovittamaan mallintava funktio tuloksiin.¹¹ Vaikka kyseessä eivät ole todennäköisyydet, niin valitsin mallintavaksi funktioksi Gaussin kellokäyrän¹² sen muodon perusteella:

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

missä μ on jakauman odotusarvo ja

σ on keskihajonta (ja σ^2 varianssi).



KUVIO 8: Eri vaihtoehtoja päiväkeskiarvon mallinnukseen

Kuviossa 8 on esitetty eri malleja kuvion 7 mittaustuloksille. Keskiarvo on eri vuosien keskiarvo jokaiselle päivälle, ja edelleen todella kohinainen. Kuukausikeskiarvo on laskettu kunkin kuukauden kaikista mittaustuloksista. Polynominen regressiokäyrä on laskettu käyttäen 6-asteista polynomia (Excelin maksimi), ja sen yhtälö on kuvio 8:n yläkulmassa, mutta selvemmin vielä tässä:

$$y = -1 * 10^{-12}x^6 + 1 * 10^{-9}x^5 - 3 * 10^{-8}x^4 - 0,0003x^3 + 0,0618x^2 - 1,1471x + 9,5497$$

missä x on päivän järjestysnumero ja

y on globaalisäteilyn intensiteetti.

¹¹ <http://sharpstatistics.co.uk/excel/non-linear-curve-fitting-in-exce/>

¹² https://fi.wikipedia.org/wiki/Normaali_jakauma

Sovitettu kellokäyrä on laskettu Solverin avulla siten, että jokaiselle päivälle lasketaan kellokäyrän parametreilla arvo, ja neliövirhe¹³ tästä arvosta ja mittausten keskiarvosta. Kellokäyrään täytyy lisätä vakiokerroin C jotta tulokset saa skaalattua oikeaan mittaluokkaan. Neliövirheistä lasketaan vielä summa koko vuodelle. Solverille annetaan tehtäväksi minimoida neliövirhe muuttamalla kellokäyrän parametreja. Kellokäyrän arvon saa laskettua Excelin funktiolla:

$$=NORM.DIST(C16; \$M\$16; \$M\$17; 0)*\$M\$18$$

ja neliövirheen vastaavasti:

$$=(J16-H16)^2$$

Jälkikäteen ajateltuna neliövirhe olisi ollut parempi laskea vuosikohtaisista keskiarvoista, eikä niistä johdetusta keskiarvosta, mutta tulos tuskin olisi ollut oleellisesti erilainen.

Ratkaistut parametrit ovat:

$$C = 70378,7104,$$

$$\mu = 173,984321 \text{ ja}$$

$$\sigma = 69,3747294,$$

ja näin saatu yhtälö on:

$$f(x) = 70378,7104 \frac{1}{69,3747294\sqrt{2\pi}} e^{-\frac{(x-173,984321)^2}{2*69,3747294^2}}$$

Polynominen regressiokäyrä ja sovitettu kellokäyrä ovat huomattava parannus päivittäisen keskiarvon mallintamiseen. Ne eivät poikkea toisistaan niin paljon että solverin käyttö olisi ollut välttämätöntä, mutta sen käyttö antaa mahdollisuuden sovittaa mitä tahansa yhtälöitä tuloksiin jos regressiokäyrät eivät enää riitä.

¹³ https://en.wikipedia.org/wiki/Mean_squared_error

Liitteet

LIITE 1: FMIDOWNLOAD/CORE.CLJ

```
(ns fmidownload.core
  (:require [clj-http.client :as client]
            [clj-time.core :as t]
            [clj-time.format :as f]
            [clojure.zip :as zip]
            [clojure.data.zip.xml :as zx]
            [clojure.data.xml :as xml]
            [clojure.data.csv :as csv]
            [clojure.java.io :as io]))

(def APIKEY (clojure.string/trim-newline (slurp "apikey.txt")))
(def BASEURL "http://data.fmi.fi/fmi-apikey/")
(def PATH "/wfs/eng")
(def PARAMS {:request "getFeature"
             :storedquery_id "fmi::observations::radiation::timevaluepair"
             :parameters "glob_1min"
             :timestep "60"})

;; XML-related functions

(defn parse-position
  "Given a string (s) in format '59.78423 21.36783 ', parse them to two floats
  and return them as a sequence (59.78423 21.36783). Data will be lat-long
  coordinates."
  [s]
  (map read-string (rest (re-find #"([0-9]+\.[0-9]+) ([0-9]+\.[0-9]+)" s))))

(defn get-time-and-value
  "Parse timestamp and value from XML node (node), return a map"
  [node]
  {:timestamp (f/parse
               (f/formatters :date-time-no-ms)
               (zx/xml1->
                (zip/xml-zip node)
                first
                :time
                zx/text))
   :value (read-string (zx/xml1->
                       (zip/xml-zip node)
                       first
                       :value
                       zx/text)))})

(defn get-results
  "Traverse XML from (node) to result nodes and get their timestamps and values,
  return a sequence of maps."
  [node]
  (zx/xml1->
   (zip/xml-zip node)
   first
   :PointTimeSeriesObservation
   :result
   :MeasurementTimeseries
   :point
   :MeasurementTVP
   get-time-and-value))
```

```

(defn get-member-location
  "Given an XML node for measurement location (node), parse its name and position,
  return a map"
  [node]
  (let [name (zx/xml1->
              (zip/xml-zip node)
              first
              :PointTimeSeriesObservation
              :featureOfInterest
              :SF_SpatialSamplingFeature
              :shape
              :Point
              :name
              zx/text)
        pos (zx/xml1->
              (zip/xml-zip node)
              first
              :PointTimeSeriesObservation
              :featureOfInterest
              :SF_SpatialSamplingFeature
              :shape
              :Point
              :pos
              zx/text)]
    {:name name
     :pos (parse-position pos)}))

(defn parse-member
  "Given a measurement XML node (node), parse its name, position and measurement
  results. Return a map."
  [node]
  (merge (get-member-location node) {:results (get-results node)}))

(defn get-time
  "Return :starttime and :endtime parameters as one week periods counting
  backwards from today midnight with (week-offset), where 0 is previous week,
  1 is the one before that, etc."
  [week-offset]
  (let [n (t/now)
        now (t/date-time (t/year n) (t/month n) (t/day n))]
    {:starttime (f/unparse (f/formatters :date-time-no-ms)
                           (t/minus now (t/weeks (+ week-offset 1))))
     :endtime (f/unparse (f/formatters :date-time-no-ms)
                          (t/minus now (t/weeks week-offset))))))

(defn get-radiation
  "Given (week-offset), query for FMI data with constant parameters
  BASEURL, APIKEY, PATH and PARAMS."
  [week-offset]
  (client/get (str BASEURL APIKEY PATH)
              {:query-params (merge PARAMS (get-time week-offset))}))

(defn get-zip
  "Given (week-offset), return xml-zip root node of FMI data."
  [week-offset]
  (zip/xml-zip (xml/parse-str ((get-radiation week-offset) :body))))

(defn download-fmi-data
  "Given (week-offset), return a vector of measurement locations (as maps)
  with :name, :pos and :results keys."
  [week-offset]
  (map parse-member (zx/xml->
                     (get-zip week-offset)
                     :member)))

(defn format-timestamp
  "Given a map (x) with :timestamp as clj-time object, return formatted time
  as string."
  [x]
  (f/unparse (f/formatter "yyyy-MM-dd HH:mm:ss")
             (:timestamp x)))

(defn format-timestamps
  "Given a location data map (loc-data), return all its measurement times as
  a vector of formatted time strings"
  [loc-data]
  (vec (sort (map format-timestamp
                  (:results loc-data)))))

```

```

(defn format-for-excel
  "Given a vector of location data sets (data), format data for excel import.
  Write an UTF-8 formatted CSV file (filename) so that measurement locations
  are in columns and results are in rows with first column being the measurement
  time"
  [data filename]
  (let [timestamps (format-timestamps (first data))
        parsed (mapv (fn [loc]
                        { :name (:name loc)
                          :lat (:lat loc)
                          :lon (:lon loc)
                          :results (map (fn [x] {:timestamp (format-timestamp x)
                                                  :value (:value x)})
                                         (:results loc)))
                        data)
                    data)
        csvheader (cons "" (mapv #(:name %) parsed))
        csvlat (cons "latitude" (mapv #(:lat %) parsed))
        csvlon (cons "longitude" (mapv #(:lon %) parsed))
        csvresults (mapv (fn [timestamp]
                           (cons timestamp
                                   (mapv (fn [loc]
                                           (:value
                                            (first
                                             (filter #(= timestamp (:timestamp %))
                                                       (:results loc))))
                                           parsed)))
                                   timestamps))
                           data)
        out-file (io/writer filename)]
    (with-open [out-file (io/writer filename)]
      (csv/write-csv
        out-file
        (vec (concat [csvheader csvlat csvlon] csvresults))))))

(defn merge-data
  "Given a vector of polled fmi-data sets (data), as only one week can be queried at
  a time, merge the data sets so that all data for each measurement location is
  in the same result vector, sorted by time"
  [data]
  (let [names (map #(:name %) (first data))]
    (map (fn [name]
           (let [pos (:pos (first (filter #(= name (:name %)) (first data))))]
             { :name name
               :lat (str (first pos))
               :lon (str (second pos))
               :results (sort-by :timestamp t/before?
                                (flatten (map #(:results %)
                                              (filter #(= name (:name %))
                                                      (flatten data))))))
           names))))

(defn -main
  "Entry point, query FMI for data from (num-weeks) and write measurement results
  to (filename)"
  [num-weeks filename]
  (let [weeks (range (read-string num-weeks))
        weekly-data (map download-fmi-data weeks)
        merged (merge-data weekly-data)]
    (format-for-excel merged filename)))

```

LIITE 2: ESIMERKKI-CSV

,Parainen Utö,Vantaa Helsinki-Vantaan lentoasema,Helsinki Kumpula,Jokioinen Jokioisten observatorio,Jyväskylä
lentoasema,Sotkamo Kuolaniemi,Sodankylä Lapin ilmatieteellinen tutkimuskeskus,Utsjoki Kevo
latitude,59.78423,60.3267,60.20307,60.81401,62.39758,64.11197,67.36663,69.75637
longitude,21.36783,24.95675,24.9613,23.49761,25.67087,28.33639,26.62901,27.00678
2015-08-24 00:00:00,-2.8,-2.9,-2.7,-2.7,-2.0,-1.4,-1.4,-0.6
2015-08-24 01:00:00,-2.7,-2.7,-2.3,-1.9,-2.1,-0.9,-1.6,-0.6
2015-08-24 02:00:00,-1.5,-2.4,-2.8,-2.3,-1.6,-0.2,0.6,0.6
2015-08-24 03:00:00,-0.1,3.7,3.6,3.1,8.6,19.5,29.2,12.8
2015-08-24 04:00:00,53.6,70.8,69.5,74.2,87.8,113.9,104.1,41.1
2015-08-24 05:00:00,169.3,156.2,180.5,194.8,199.1,230.1,200.4,51.8
2015-08-24 06:00:00,302.9,317.7,320.7,322.4,317.5,342.8,300.0,53.5
2015-08-24 07:00:00,431.7,436.1,436.2,439.4,433.5,450.8,391.2,93.1
2015-08-24 08:00:00,537.1,535.8,538.9,535.8,530.4,536.9,469.0,154.2
2015-08-24 09:00:00,612.2,608.1,607.9,610.8,592.2,602.6,520.7,480.5
2015-08-24 10:00:00,653.3,652.8,645.2,736.4,639.2,614.2,559.8,683.8
2015-08-24 11:00:00,654.0,638.8,649.9,754.0,645.5,605.9,150.9,261.2
2015-08-24 12:00:00,619.7,603.1,613.1,673.7,568.5,561.1,231.9,180.4
2015-08-24 13:00:00,544.6,537.9,530.8,481.8,494.8,483.3,145.1,364.0
2015-08-24 14:00:00,439.7,421.3,418.9,491.7,394.1,384.0,264.4,345.8
2015-08-24 15:00:00,320.1,296.9,292.2,314.3,278.5,268.1,283.3,283.9
2015-08-24 16:00:00,192.6,166.6,164.5,180.6,159.1,151.5,155.4,94.2
2015-08-24 17:00:00,27.7,53.1,49.8,62.1,51.1,47.9,64.7,85.7
2015-08-24 18:00:00,2.2,0.5,-0.6,1.3,1.2,2.1,6.9,13.7
2015-08-24 19:00:00,-1.7,-2.3,-2.5,-3.0,-2.2,-1.0,-1.3,-1.3
2015-08-24 20:00:00,-1.7,-2.6,-2.0,-2.5,-2.4,-0.6,-2.0,-1.8
2015-08-24 21:00:00,-2.2,-2.2,-2.3,-2.3,-2.0,-0.7,-1.6,-2.3
2015-08-24 22:00:00,-1.8,-2.3,-1.9,-1.9,-2.0,-0.9,-1.3,-1.0
2015-08-24 23:00:00,-2.4,-2.9,-2.2,-2.2,-1.9,-1.1,-1.5,-0.7
2015-08-25 00:00:00,-1.8,-1.9,-2.1,-2.1,-1.5,-1.0,-1.2,-0.4
2015-08-25 01:00:00,-1.8,-1.9,-2.5,-1.7,-1.6,-0.8,-1.3,0.0
2015-08-25 02:00:00,-1.5,-2.2,-1.9,-2.4,-1.4,0.2,0.7,3.4
2015-08-25 03:00:00,-0.3,6.4,8.1,4.8,7.2,22.8,14.5,13.8
2015-08-25 04:00:00,52.2,108.4,88.2,41.9,83.7,115.9,82.9,21.0
2015-08-25 05:00:00,91.4,46.8,48.2,69.3,246.6,225.4,48.7,38.7
2015-08-25 06:00:00,203.1,100.3,120.0,304.4,359.3,338.2,144.7,77.3
2015-08-25 07:00:00,397.4,214.5,518.0,86.6,422.2,440.0,142.8,64.7
2015-08-25 08:00:00,499.0,347.7,317.8,516.4,520.1,524.2,494.0,161.0
2015-08-25 09:00:00,639.0,468.4,342.5,582.9,578.8,577.6,198.9,126.8
2015-08-25 10:00:00,618.1,337.8,487.8,619.2,207.4,609.9,211.2,132.4
2015-08-25 11:00:00,575.7,385.5,776.3,658.3,318.4,596.8,173.9,249.5
2015-08-25 12:00:00,551.5,581.5,607.6,594.6,593.1,546.5,220.1,199.3
2015-08-25 13:00:00,517.8,506.4,494.8,528.0,485.3,484.4,165.6,119.4
2015-08-25 14:00:00,402.0,439.1,457.8,336.6,374.2,368.6,176.4,118.9
2015-08-25 15:00:00,177.0,202.0,276.5,155.5,128.5,255.1,74.8,68.9
2015-08-25 16:00:00,139.8,186.9,155.8,182.4,137.8,51.6,50.5,30.7
2015-08-25 17:00:00,37.7,50.8,33.1,29.4,36.9,29.5,36.0,15.5
2015-08-25 18:00:00,0.4,-1.0,-1.8,-0.5,-0.1,0.9,4.2,2.2
2015-08-25 19:00:00,-1.2,-2.7,-2.8,-2.0,-2.1,-0.3,-1.1,-0.6
2015-08-25 20:00:00,-0.9,-2.6,-2.7,-2.1,-2.0,-1.2,-0.5,0.2
2015-08-25 21:00:00,-0.9,-1.9,-2.7,-2.3,-2.0,-1.3,-0.5,-0.5
2015-08-25 22:00:00,-1.3,-2.3,-2.3,-1.7,-1.8,-0.6,-0.8,-0.4
2015-08-25 23:00:00,-0.6,-2.5,-2.4,-1.4,-2.3,-0.8,-0.9,-0.9
2015-08-26 00:00:00,-0.7,-1.7,-1.6,-1.5,-1.7,-1.1,-0.5,-0.6
2015-08-26 01:00:00,-0.8,-1.9,-1.5,-1.5,-1.4,-0.8,-0.7,-0.6
2015-08-26 02:00:00,-0.9,-1.4,-1.5,-1.1,-1.6,0.4,-0.2,0.6
2015-08-26 03:00:00,-0.3,0.0,0.1,0.7,5.7,12.7,5.9,8.3
2015-08-26 04:00:00,6.9,11.7,16.3,8.6,77.1,100.7,22.6,27.9
2015-08-26 05:00:00,12.6,14.1,20.5,12.8,194.9,148.8,10.3,65.5
2015-08-26 06:00:00,91.5,28.0,20.1,24.1,293.0,343.5,35.4,76.2
2015-08-26 07:00:00,216.2,33.6,102.1,19.8,293.2,450.9,23.9,104.0
2015-08-26 08:00:00,532.0,34.2,73.7,47.9,505.1,286.6,58.2,151.6
2015-08-26 09:00:00,632.5,95.6,91.8,60.9,253.1,698.8,98.7,188.5
2015-08-26 10:00:00,299.2,124.9,58.7,80.2,173.7,297.8,245.8,237.3
2015-08-26 11:00:00,681.0,180.3,511.3,87.5,91.7,343.7,156.8,256.8
2015-08-26 12:00:00,109.4,155.3,239.8,325.5,65.1,176.8,137.8,584.3
2015-08-26 13:00:00,384.1,338.0,108.7,272.2,45.5,121.8,183.7,469.8
2015-08-26 14:00:00,234.2,401.8,441.4,516.6,23.5,92.0,83.8,367.1
2015-08-26 15:00:00,327.1,207.2,291.2,128.5,28.1,269.9,93.6,314.8
2015-08-26 16:00:00,180.2,173.8,69.3,185.7,26.3,79.0,43.4,210.9
2015-08-26 17:00:00,75.5,62.2,60.3,24.3,5.8,46.4,23.7,69.4
2015-08-26 18:00:00,1.3,0.9,1.1,0.3,-1.4,1.2,3.6,8.0
2015-08-26 19:00:00,-2.6,-2.4,-2.3,-1.6,-1.9,-1.0,-0.4,-0.4
2015-08-26 20:00:00,-1.8,-2.3,-3.3,-1.0,-1.6,-0.8,-0.8,-0.9
2015-08-26 21:00:00,-2.1,-2.1,-2.7,-1.3,-1.5,-0.5,-1.1,-1.1
2015-08-26 22:00:00,-1.8,-2.4,-2.1,-1.1,-1.2,-0.3,-0.1,-0.5
2015-08-26 23:00:00,-2.2,-2.4,-2.5,-2.1,-1.7,0.2,-0.5,-0.8

2015-08-27 00:00:00,-1.9,-2.0,-2.2,-1.4,-1.4,-0.6,-0.5,-1.1
2015-08-27 01:00:00,-0.9,-2.2,-1.7,-2.0,-1.2,-0.2,-0.3,-0.8
2015-08-27 02:00:00,-1.3,-2.4,-1.8,-2.2,-1.2,0.2,0.0,0.9
2015-08-27 03:00:00,-0.9,0.4,0.4,0.8,1.2,2.1,4.1,13.2
2015-08-27 04:00:00,45.4,43.7,46.4,33.9,19.7,4.8,15.1,46.0
2015-08-27 05:00:00,138.7,149.0,144.4,123.6,79.4,24.8,25.8,88.4
2015-08-27 06:00:00,264.0,304.9,139.7,168.3,111.5,53.1,36.6,165.1
2015-08-27 07:00:00,417.9,434.3,204.4,291.2,113.0,42.9,95.7,214.0
2015-08-27 08:00:00,153.0,618.2,604.4,190.5,261.0,25.4,71.8,180.8
2015-08-27 09:00:00,389.1,712.5,544.4,267.5,279.2,87.0,77.8,227.8
2015-08-27 10:00:00,242.1,476.8,640.3,170.5,217.2,107.1,84.3,215.9
2015-08-27 11:00:00,665.9,331.3,668.1,303.2,68.8,74.2,110.6,348.5
2015-08-27 12:00:00,511.1,422.2,377.8,78.7,83.6,52.8,173.2,465.8
2015-08-27 13:00:00,378.5,153.1,199.7,244.2,65.3,25.7,42.4,238.2
2015-08-27 14:00:00,198.6,348.6,410.5,171.2,98.8,164.7,50.3,353.1
2015-08-27 15:00:00,97.4,266.5,248.8,160.1,97.0,67.0,25.7,239.5
2015-08-27 16:00:00,47.7,161.9,102.8,111.4,116.8,210.1,21.9,63.0
2015-08-27 17:00:00,20.6,23.6,22.6,14.2,16.6,16.7,8.3,20.8
2015-08-27 18:00:00,0.2,-1.5,-1.4,-1.2,-0.1,1.2,-0.2,3.4
2015-08-27 19:00:00,-2.2,-1.5,-1.6,-1.8,-1.1,-1.0,-0.9,-1.5
2015-08-27 20:00:00,-1.0,-1.9,-1.2,-3.7,-0.8,0.5,-1.7,-1.8
2015-08-27 21:00:00,-0.3,-3.6,-2.5,-1.4,-0.7,-0.3,0.3,-0.2
2015-08-27 22:00:00,-1.3,-3.7,-4.3,-3.0,-1.0,-0.3,0.2,-1.2
2015-08-27 23:00:00,-0.7,-4.0,-3.1,-2.4,-1.2,-0.4,-0.2,-1.1
2015-08-28 00:00:00,-0.5,-2.4,-2.0,-2.3,-1.4,0.2,-0.8,0.0
2015-08-28 01:00:00,-0.9,-2.2,-1.0,-1.9,-1.8,-0.1,-1.4,-0.1
2015-08-28 02:00:00,-0.8,-2.0,-1.0,-1.4,-1.0,0.4,-0.7,0.5
2015-08-28 03:00:00,-1.1,-0.6,0.5,-0.7,-1.4,7.4,1.9,5.6
2015-08-28 04:00:00,4.2,7.7,2.7,13.3,4.3,33.3,11.7,24.2
2015-08-28 05:00:00,10.9,4.0,17.1,35.4,24.5,11.1,25.2,25.6
2015-08-28 06:00:00,11.8,39.5,29.7,40.6,127.8,31.2,52.2,40.1
2015-08-28 07:00:00,63.0,59.5,45.8,25.2,166.3,34.7,61.0,48.0
2015-08-28 08:00:00,166.1,82.9,50.5,94.5,60.3,63.6,97.5,36.0
2015-08-28 09:00:00,183.9,82.2,109.8,98.8,116.9,53.8,180.4,60.7
2015-08-28 10:00:00,762.1,225.1,104.2,137.4,213.1,38.6,220.7,69.9
2015-08-28 11:00:00,660.4,139.1,106.8,219.9,104.3,43.0,140.7,54.7
2015-08-28 12:00:00,614.7,156.4,29.0,407.9,115.7,38.6,166.9,245.7
2015-08-28 13:00:00,522.2,269.9,274.4,574.9,129.1,50.7,115.0,245.3
2015-08-28 14:00:00,442.1,182.3,167.7,64.5,100.1,25.0,59.4,104.0
2015-08-28 15:00:00,315.4,294.7,206.1,297.0,298.7,12.7,31.5,85.7
2015-08-28 16:00:00,64.5,188.6,35.8,167.7,137.4,11.1,18.1,53.4
2015-08-28 17:00:00,61.0,18.3,17.3,48.4,14.7,3.4,2.4,24.2
2015-08-28 18:00:00,0.8,-1.3,-1.5,-1.2,-1.1,0.5,-0.6,5.8
2015-08-28 19:00:00,-1.5,-2.8,-2.7,-2.1,-2.3,-0.2,-1.0,-1.8
2015-08-28 20:00:00,-1.7,-2.5,-2.1,-2.6,-2.1,0.2,-0.7,-2.4
2015-08-28 21:00:00,-2.2,-2.6,-2.9,-2.2,-2.3,-0.5,-0.8,-2.4
2015-08-28 22:00:00,-1.8,-2.5,-1.8,-2.1,-2.1,0.6,-0.4,-2.0
2015-08-28 23:00:00,-1.7,-2.1,-2.2,-2.4,-2.3,0.1,-0.7,-0.8
2015-08-29 00:00:00,-2.0,-2.8,-2.3,-1.8,-1.9,0.3,-0.8,-0.5
2015-08-29 01:00:00,-2.4,-2.7,-3.4,-2.2,-1.1,0.0,-0.5,-0.8
2015-08-29 02:00:00,-1.9,-2.8,-2.1,-2.0,-1.4,-0.2,-0.7,3.0
2015-08-29 03:00:00,-1.7,0.3,-0.1,0.1,0.9,5.3,0.8,5.5
2015-08-29 04:00:00,44.6,60.9,59.2,57.4,28.5,38.8,9.9,11.3
2015-08-29 05:00:00,166.8,180.8,179.8,173.6,107.7,31.6,13.4,24.4
2015-08-29 06:00:00,148.3,308.6,281.3,307.6,130.7,57.6,50.9,44.0
2015-08-29 07:00:00,140.8,426.4,426.9,304.4,260.3,29.7,73.0,80.5
2015-08-29 08:00:00,523.3,149.9,182.0,632.3,706.7,99.8,101.9,152.2
2015-08-29 09:00:00,602.4,755.6,727.0,660.4,282.4,187.5,114.6,53.9
2015-08-29 10:00:00,653.7,686.0,742.6,727.7,445.9,435.8,194.8,67.1
2015-08-29 11:00:00,580.7,686.3,441.4,131.7,576.9,341.0,156.5,64.6
2015-08-29 12:00:00,612.2,198.7,668.8,613.8,189.4,551.1,156.3,59.5
2015-08-29 13:00:00,535.7,622.0,571.7,547.4,638.0,488.1,118.7,62.0
2015-08-29 14:00:00,430.2,445.4,404.2,478.8,318.0,303.2,454.4,175.2
2015-08-29 15:00:00,300.3,278.0,278.8,300.0,137.9,217.5,123.4,48.3
2015-08-29 16:00:00,170.8,149.4,145.1,162.5,138.2,141.5,138.1,34.7
2015-08-29 17:00:00,59.0,35.5,34.0,45.8,31.8,30.2,55.2,15.3
2015-08-29 18:00:00,-0.1,-1.9,-2.3,-1.2,-1.1,-0.5,5.5,0.6
2015-08-29 19:00:00,-2.3,-2.7,-2.4,-2.5,-2.4,-1.4,-0.1,-0.2
2015-08-29 20:00:00,-1.5,-2.7,-3.1,-2.2,-1.7,-1.4,-0.8,-0.5
2015-08-29 21:00:00,-1.5,-2.4,-2.8,-2.1,-1.5,-1.1,-0.5,-1.0
2015-08-29 22:00:00,-2.7,-2.6,-2.8,-2.5,-2.1,-0.8,-0.4,-0.9
2015-08-29 23:00:00,-2.2,-2.8,-2.1,-2.3,-2.2,-0.7,-0.7,-0.4
2015-08-30 00:00:00,-2.4,-1.9,-2.7,-2.3,-2.1,-0.5,-0.1,-0.1
2015-08-30 01:00:00,-2.7,-2.6,-2.4,-1.7,-1.7,-0.7,-0.6,-0.4
2015-08-30 02:00:00,-2.3,-2.3,-1.8,-2.5,-1.9,-1.2,-0.1,-0.2
2015-08-30 03:00:00,-1.5,-0.2,0.4,0.0,3.7,7.6,4.9,4.2
2015-08-30 04:00:00,25.8,76.3,73.6,54.2,53.2,89.8,19.5,19.3
2015-08-30 05:00:00,96.9,121.9,102.5,187.3,192.9,203.8,72.5,88.5
2015-08-30 06:00:00,252.4,120.8,266.6,313.6,274.5,233.9,104.7,114.5

2015-08-30 07:00:00,328.0,439.9,428.8,342.8,308.0,439.2,172.7,100.6
2015-08-30 08:00:00,120.8,150.7,642.3,337.2,578.9,232.4,475.3,141.1
2015-08-30 09:00:00,120.7,642.4,542.5,360.9,719.8,563.6,356.4,328.3
2015-08-30 10:00:00,638.3,141.7,240.0,179.3,659.8,590.7,567.9,537.4
2015-08-30 11:00:00,641.5,723.0,293.3,24.0,508.1,575.9,581.4,122.9
2015-08-30 12:00:00,603.2,114.2,249.5,150.9,284.9,534.7,158.4,528.3
2015-08-30 13:00:00,536.3,24.3,5.0,62.1,243.4,457.3,104.1,469.6
2015-08-30 14:00:00,423.8,168.7,199.2,131.8,388.4,424.7,135.5,290.1
2015-08-30 15:00:00,291.8,69.2,262.3,343.9,200.1,304.6,309.9,79.0
2015-08-30 16:00:00,164.0,141.6,38.8,186.5,54.6,149.4,184.6,47.3
2015-08-30 17:00:00,54.8,33.6,13.7,45.5,18.2,30.1,54.5,26.2
2015-08-30 18:00:00,-0.9,-2.1,-2.2,-1.5,-1.4,0.0,1.3,3.0
2015-08-30 19:00:00,-2.6,-2.7,-2.3,-2.2,-1.9,-1.0,-1.9,-2.0
2015-08-30 20:00:00,-1.6,-2.5,-1.9,-2.4,-2.0,-0.4,-2.1,-1.8
2015-08-30 21:00:00,-1.4,-2.3,-2.8,-1.8,-2.1,-0.8,-1.5,-1.7
2015-08-30 22:00:00,-2.1,-2.2,-2.2,-2.0,-2.0,-0.5,-2.0,-1.1
2015-08-30 23:00:00,-1.6,-2.5,-1.9,-2.0,-2.0,-1.0,-1.8,-1.5
2015-08-31 00:00:00,-1.9,-2.3,-1.8,-2.2,-1.8,-1.6,-1.9,-2.1