

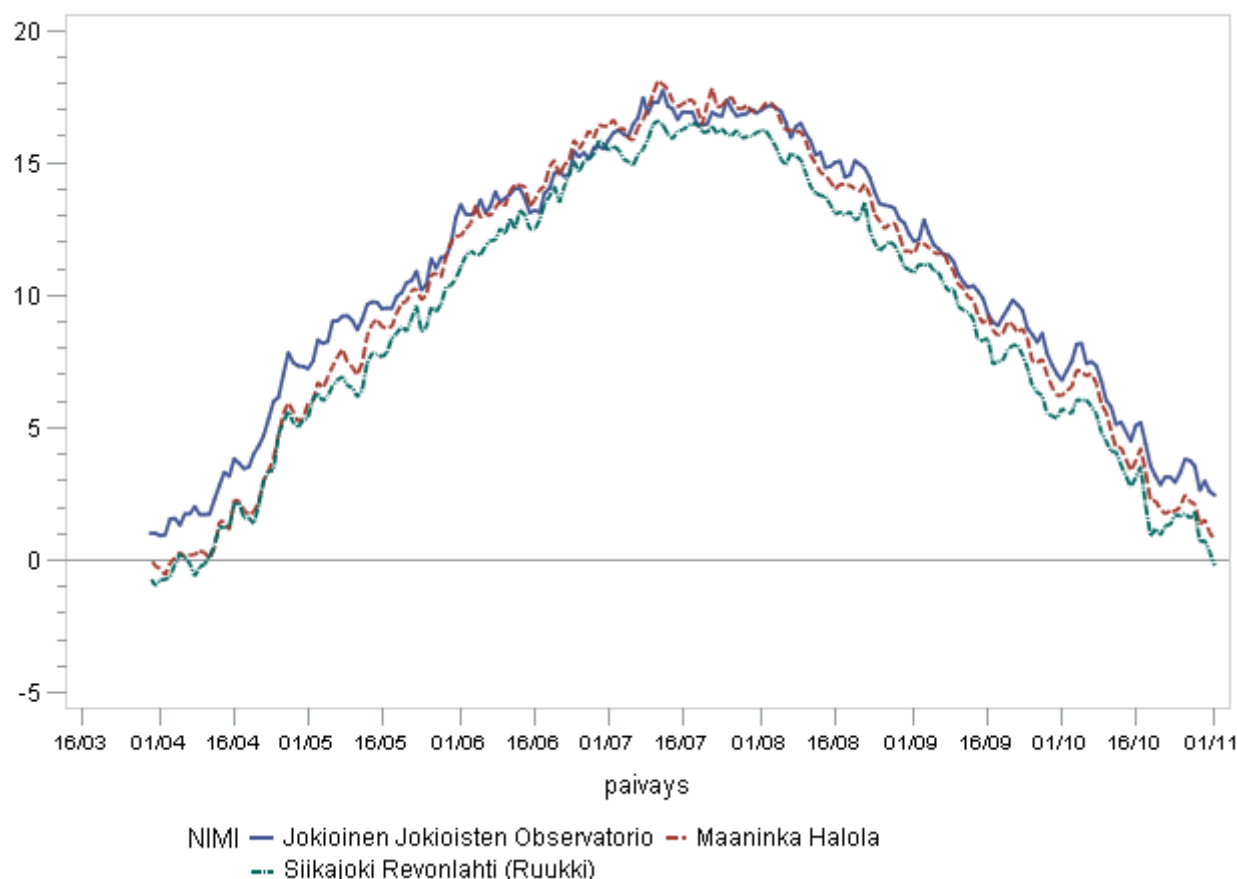
MTT:n sääaineistot ja tietokannat hallintaan SASilla

Tietoiskutilaisuus 21.3.2013.

Versio 1.1

Timo Pitkänen

VRK KESKILAMPOTILA



MTT:n sääaineistot ja tietokannat hallintaan SASilla

Tämän tietoiskumateriaalin tarkoitus on antaa lyhyt kuvaus MTT:lle hankittujen sääaineistojen käyttämisestä SASilla. Kasvukauden sääaineistot haetaan Oracle -tietokannasta ja pitkittäissäätä käsitellään SAS-datasta.

Esimerkit on tehty käyttäen SAS-koodia, joka on mahdollista suorittaa SAS Enterprise Guidessa. Kaikki tämän materiaalin kuvat ja listaukset on poimittu suoraan EG:stä ja ne on tuotettu esitetyillä SAS-koodeilla. SAS-ohjelmista kannattaa poimia omaan käyttöön ideoita ja keinoja omien tutkimusten tarpeisiin.

Tietokoneelle tarvittavat asetukset tietokantayhteyksiä varten ja niiden testaaminen

Tietokoneella täytyy olla tietokantayhteyksien käyttöä varten asennettuna SAS ja siihen *SAS/Access Interface to Oracle* -tuote. Tämän lisäksi on asennettava Oracle tietokantaan pääsyn tarjoava Oracle -client. Tietokantayhteyden käyttäminen vaatii MTT:n verkossa olemisen. Yhteys toimii myös toimikortin kautta. Asennukset hoidetaan IT-tuen kautta.

Tietokantayhteyden toiminnan testaaminen

Tietokantayhteyden toiminnan voi testata ajamalla SASissa rivin

```
libname saa_mtt oracle user=mtt password=mtt path=metis.mtt.fi;
```

Jos kaikki yhteydet ovat kunnossa, tulee logiin seuraavaa:

```
NOTE: Libref SAA_MTT was successfully assigned as follows:
```

```
Engine: ORACLE
```

```
Physical Name: metis.mtt.fi
```

Tarvittavien komponenttien puuttuessa voi tulla esimerkiksi seuraava viesti:

```
ERROR: The ORACLE engine cannot be found.
```

```
ERROR: Error in the LIBNAME statement.
```

SAS/Access Interface to Oracle asennuksen testaaminen

Tuotteen *SAS/Access Interface to Oracle* olemassaolon ja kaikkien muiden asennettujen tuotteiden listan voi tarkistaa ajamalla SASissa koodirivin:

```
proc setinit; run;
```

Jos Access Interface to Oracle on asennettu, löytyy logista rivi:

```
--SAS/ACCESS Interface to Oracle    31JAN2014
```

Oracle -clientin asennuksen varmistaminen

Jos Oracle -client on asennettu, pitäisi käynnistä valikosta löytyä siihen liittyvä kohta. Paina käynnistä -> kaikki ohjelmat. Listassa pitäisi olla "Oracle - OraClient10g...".

Tietokantayhteyden muodostaminen

Yleistietoa

SAS tukee monia eri tietokantamoottoreita, joista MTT:llä on käytössä Oracle. Yhteys tietokantaan muodostetaan kirjastoviitteen kautta samaan tapaan kuin viitattaessa omalla tietokoneella olevaan hakemistoon. Tietokantayhteyden viittauksessa tarvitaan hieman enemmän parametreja kuin hakemistoon viitatessa.

Tietokantaan viitattaessa eroksi tulee se, ettei kaikissa tapauksissa nähdä suoraan SASista minkä nimisiä tietokantatauluja viittauksen osoittamassa paikassa on. Tämä johtuu tietokantaan liittyvistä teknisistä syistä ja sen vuoksi käyttäjän on esimerkiksi sääaineistojen kohdalla tiedettävä tarkalleen haluamansa sääaineiston taulun nimi.

MTTn yhteisten tietovarantojen (mm. sääaineistot) käyttämistä varten on Oracle-tietokantaan tehty näkymiä. Näiden käyttämiseen salasana ja käyttäjätunnus on **mtt**.

Tietokantayhteyden vaatiman kirjastoviitteen tekeminen koodilla

KoodiSASia käyttäessä tietokantayhteys muodostetaan libname -lauseen avulla:

```
libname saa_mtt oracle user=mtt password=mtt path=metis.mtt.fi;
```

Koodi luo kirjastoviitteen nimeltä *saa_mtt*, joka on yhteys oracle -tietokantaan.

Tietokantayhteyden vaatiman kirjastoviitteen tekeminen EG:ssä

Enterprise guidessa valitse **tools -> assign project library**

Assign Project Library

1 of 4 Specify a name and server for the library.

This wizard assigns a library for the current project and the current user. This library will be un-assigned when you exit the application. If you want to create a permanent library on a server, create a server library definition by going to menu Tools->Enterprise Guide Explorer.

Name (enter 8 or fewer characters):
SAA_MTT

1. Kirjastoviitteen nimi

Server:
Local - The SAS server on your machine

<Back Next> Finish Cancel Help

Assign Project Library

2 of 4 Specify the engine for the library.

Engine type:
Database System

1. Valitse Database System

Additional information needed for 'Database System'

Engine:
ORACLE - SAS/ACCESS Interface to Oracle

2. Valitse Oracle - SAS/ACCES interface to Oracle

Database server: Schema Name:

User ID: Password:

mtt

3. Käyttäjätunnus ja salasana (mtt ja mtt)

<Back Next> Finish Cancel Help

Assign Project Library

3 of 4 Specify options for this library.

Enter options as name-value pairs:

Name	Value

Delete

1. path=metis.mtt.fi

Additional Options:

path=metis.mtt.fi

<Back Next> Finish Cancel Help

Paina next ja finish. Kirjastoviite on nyt luotu.

Sääaineistojen hakeminen ja tarkastelu

Tietokannassa olevat sääaineistot

Tietokannassa olevat sääaineistot ja niiden taulujen nimet on esitetty seuraavassa taulukossa

Aineisto	Taulun nimi
Sääasemien tiedot	saata_havaintoasema
Havaintoasemien vuorokausisää kasvukaudelta	saata_kasvusaa_asema_vrk
Havaintoasemien 3h aineisto kasvukaudelta	saata_kasvusaa_asema_3h
Hila-aineisto vuorokausi	saata_kasvusaa_hila_vrk
Hila-aineisto 3h	saata_kasvusaa_hila_3h

Tällä hetkellä tietokannassa ei ole pitkäaikaissäätä. Se tullaan lisäämään sinne myöhemmin.

Tietokannan ulkopuoliset sääaineistot

Pitkäaikaissää ja muita sääaineistoja löytyy SAS -muodossa J -asemalle perustetusta sääkansiosista
J:\MTT_yhteiset\Sääaineistot .

1 Havaintoasemien sääaineiston hakeminen ja perustietojen katselu

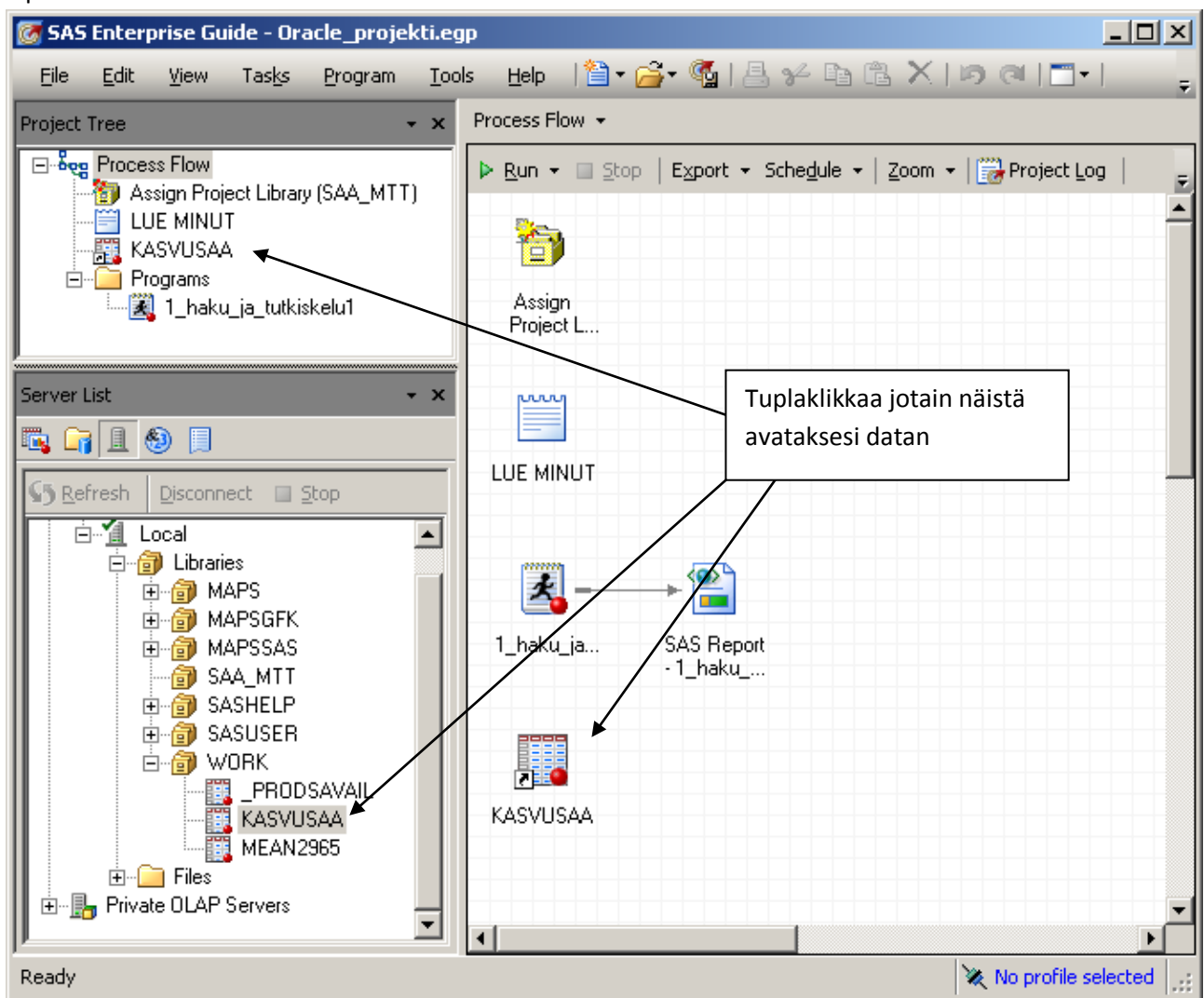
MTT:n sääaineistossa olevien asemien tiedot saa esille seuraavalla tavalla:

```
*katsotaan asematietoa ;  
data asematieto; set saa_mtt.saata_havaintoasema;  
run;  
proc print data=asematieto;  
run;
```

Haetaan havaintoasemien vuorokausisää kasvukaudelta (saata_kasvusaa_asema_vrk) ja katsotaan mitä eri muuttujia siinä on.

```
*aineiston haku kasvusaa nimelle;  
data kasvusaa; set saa_mtt.saata_kasvusaa_asema_vrk;  
run;
```

Edellinen koodi hakee kannasta nimellä **saata_kasvusaa_asema_vrk** olevan taulun ja muodostaa siitä tilapäisen SAS-datan nimeltä kasvusaa. Tämän datan sisältöä voi tutkiskella EG:ssä esimerkiksi tuplaklikkaamalla **Process Flow**:ssa olevaa **kasvusaa** -dataa



Yhteenvedo aineistossa olevista muuttujista saadaan ajamalla koodi:

```
*muuttujat ja muuta informaatiota;
proc contents data=kasvusaa;
run;
```

Alphabetic List of Variables and Attributes					
#Variable	Type	Len	Format	Informat	Label
1ASEMA_TIETO_SEQNO	Num	8			ASEMA_TIETO_SEQNO
26DD	Num	8			DD
28ETS5	Num	88.1		8.1	ETS5
2HAKUASEMA_SEQNO	Num	8			HAKUASEMA_SEQNO
24HUM	Num	8			HUM
25HUMQ	Num	8			HUMQ
10KK	Num	8			KK
4LPNN	Char	4\$4.		\$4.	LPNN
27MS	Num	8			MS
5NIMI	Char	50\$50.		\$50.	NIMI
9PV	Num	8			PV
8PVM	Num	8	DATETIME20.	DATETIME20.	PVM
20RR	Num	87.1		7.1	RR
22RRA	Num	87.1		7.1	RRA
23RRAQ	Num	8			RRAQ
21RRQ	Num	8			RRQ
12TEMP	Num	86.1		6.1	TEMP
13TEMPQ	Num	8			TEMPQ
18TGMIN	Num	86.1		6.1	TGMIN
19TGMINQ	Num	8			TGMINQ
3TIETOASEMA_SEQNO	Num	8			TIETOASEMA_SEQNO
14TMAX	Num	86.1		6.1	TMAX
15TMAXQ	Num	8			TMAXQ
16TMIN	Num	86.1		6.1	TMIN
17TMINQ	Num	8			TMINQ
11VUOSI	Num	8			VUOSI
6X	Num	8			X
7Y	Num	8			Y

Tulostuksen perusteella aineistossa on 28 eri muuttujaa. Huomioitavaa on muuttujan **pvm** formaatti *datetime20*. Datetime formaatti tarkoittaa sitä, että muuttujassa **pvm** oleva tieto sisältää päivämäärän ja kellonajan, esimerkiksi 30MAR2010:00:00:00. Asematunnus **LPNN** on tallennettu merkkimuodossa. Lisäksi aineistossa on muuttujat vuosi, kk ja pv, joita tullaan myöhemmin hyödyntämään tehtäessä valintaa aineistosta. Asemien koordinaatit on tallennettu X ja Y muuttujiin. **X** ja **Y** on tallennettu KKJ yhtenäiskoordinaatistossa. Jatkokäsittelyiden helpottamiseksi erotetaan päivämäärä omaksi muuttujaksi ja luodaan apupäivä muokkaamalla ensimmäistä datavaihetta:

```
*Erotetaan paivamaara muuttujaan paiva;
data kasvusaa; set saa_mtt.saata_kasvusaa_asema_vrk;
```



```

format paiva apupaiva ddmmyy8.; * asetetaan tulostusformaatti
paivalle;
paiva = datepart( pvm );
apupaiva = mdy( kk, pv, 2012);
* kello = timepart( pvm );
run;

```

Muuttuja **paiva** sisältää havainnon päivämäärän SASin päivämäärämuodossa. Muuttuja apupaivassa on muuten sama tieto, mutta vuosi on pakotettu olemaan aina 2012. Tätä muuttujaa tarvitaan myöhemmin kuvien piirtämisen yhteydessä.

Seuraavaksi tarkastellaan miten havaintoja on kertynyt eri asemilta. Katsotaan jokaisen aseman ja vuoden ensimmäisen ja viimeisen havainnon päivämäärä.

```

proc tabulate data=kasvusaa;
var paiva;
class nimi lpnn vuosi;
table nimi*lpnn*vuosi, (min max)*paiva=' '*f=ddmmyy5.;
run;

```

Valittujen asemien tiedot			Min	Max
NIMI	LPNN	VUOSI		
Jokioinen Jokioisten Observatorio	1201	2010	30/03	31/10
		2011	30/03	01/11
		2012	30/03	01/11
Kokemäki Tulkkila	1140	2012	13/04	01/11
Maaninka Halola	3603	2010	30/03	31/10
		2011	30/03	01/11
		2012	30/03	01/11
Siikajoki Revonlahti (Ruukki)	5402	2010	30/03	31/10
		2011	30/03	01/11
		2012	30/03	01/11

Tulostuksen perusteella valtaosalta sääasemista löytyy kolmelta vuodelta tietoa. Yleensä aloituspäivä on maaliskuun 30. ja lopetuspäivä lokakuun 31 tai marraskuun 1. Samasta tulostuksesta näkee jokaisen aseman LPNN -tunnuksen, jonka perusteella voidaan tehdä valintoja.

2 Havaintoasemien säähavaintohistorian piirtäminen

Piirretään muutamalle asemalle eri vuosien sade ja lämpösumman kertyminen ja päivittäiset havainnot. Ensin luodaan apudata **valitut**, johon valitaan muutama asema ja niiltä havainnot syyskuun loppuun saakka. Huomaa, että asemien numerot on laitettava ”hypsuihin”, koska LPNN on tallennettu tekstimuodossa.

```

data valitut; set kasvusaa;
where LPNN IN( '3603', '1201', '5402' ) AND kk < 10;

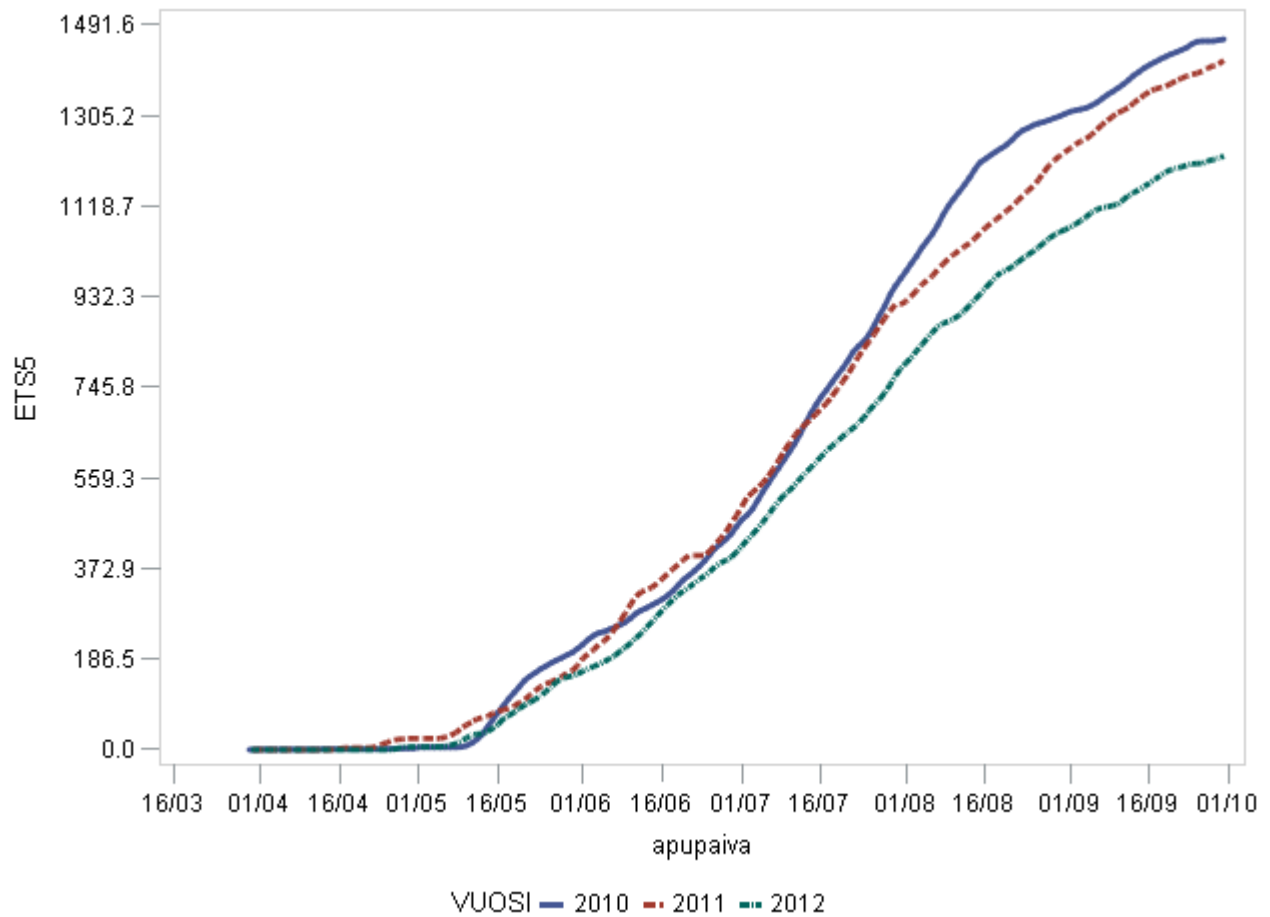
```

```
run;
```

Lämpösummakuva valituille asemille saadaan koodilla:

```
axis1 label=(angle=90) *Määritellään akseli, jonka otsikko on  
pystysuorassa ;  
proc gplot data= valitut;  
  by LPNN nimi;  
  format apupaiva ddmmyy5.;  
  plot ets5 * apupaiva = vuosi /vaxis=axis1 ;  
  symbol i=join v=none w=3;  
run;
```

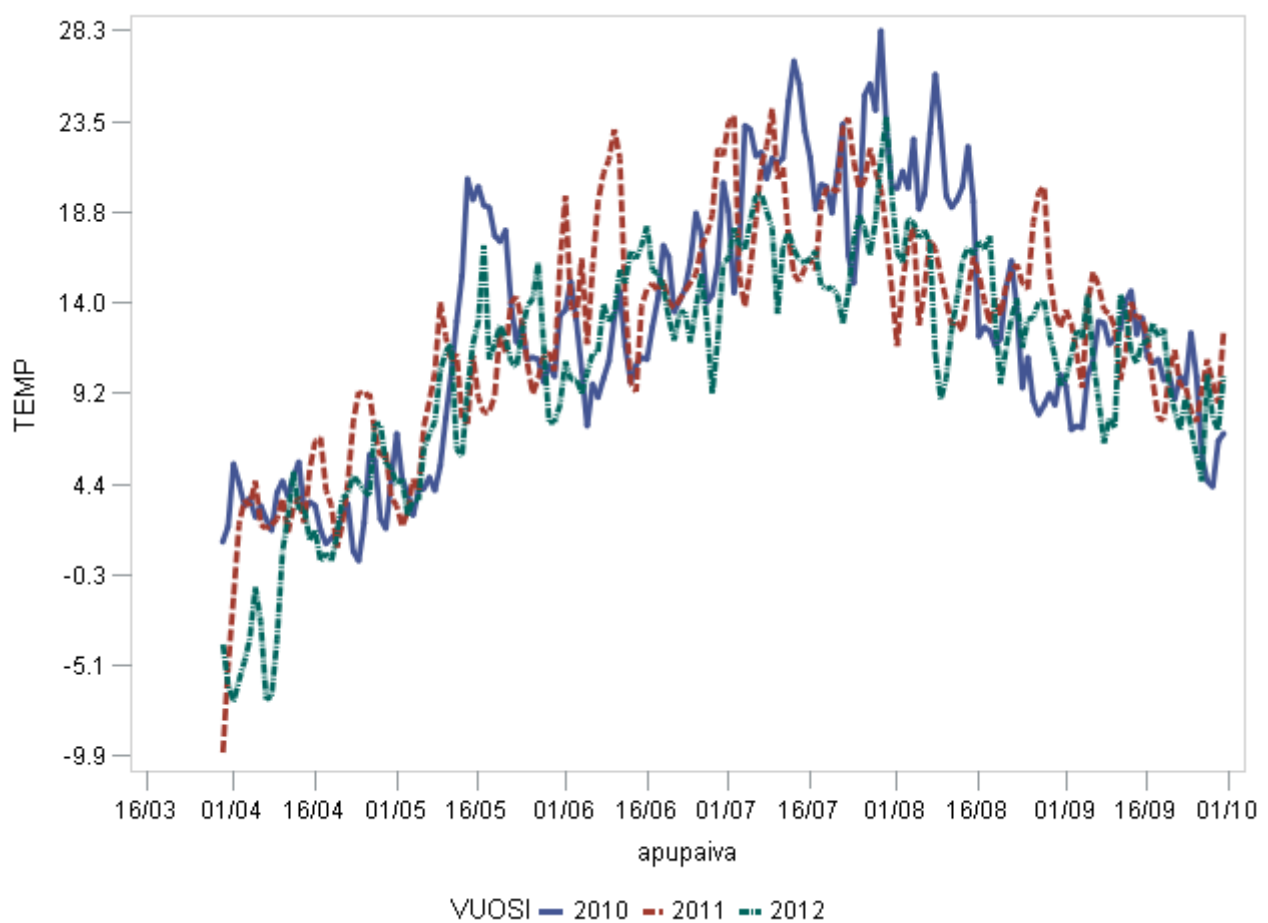
LPNN=3603 NIMI=Maaninka Halola



Vuorokausien keskilämpötila saadaan piirretyksi koodilla:

```
title "Vuorokauden keskilämpötila";  
proc gplot data=valitut uniform;  
  by LPNN nimi;  
  format apupaiva ddmmyy5.;  
  plot temp * apupaiva = vuosi /vaxis=axis1 ;  
  symbol i=none v=none w=3;  
run;
```

LPNN=3603 NIMI=Maaninka Halola



Vuoden 2010 korkein lämpötila, 37,2C, mitattiin 29.7 Liperissä. Samana päivänä näkyy piikki keskilämpötiloissa.

Sadesummaa ei tässä aineistossa suoraan ole laskettu, joten se pitää ensin itse muodostaa.

Tulevaisuudessa kasvukauden sadesumma tullaan lisäämään aineistoon. Kun aineiston päivittäistä sademuuttujaa RR katselee, havaitaan että siellä on negatiivisia arvoja, yleensä -1.0. Nämä arvot täytyy muistaa poistaa tai muuttaa nolaksi ennen kumulatiivisen summan laskemista, muuten laskennat menevät pieleen. Koodi yksinkertaiseen sadesumman laskentaan on tässä:

```
proc sort data= valitut;
  by lpnn vuosi kk pv;
run;

data valitut; set valitut;
  by lpnn vuosi;
  retain ssumma;
  * jos vuoden 1. havainto, on sadesumma 0;
  if first.vuosi then ssumma=0;
  * jos lämpösumma on > 0, on kasvukausi alkanut ja sadesummaa
lasketaan;
  * tarkistetaan myös, että RR > 0 eli jotain sadetta on tullut;
  if ( ets5 > 0 )and ( RR > 0 ) then ssumma + RR;
run;
```

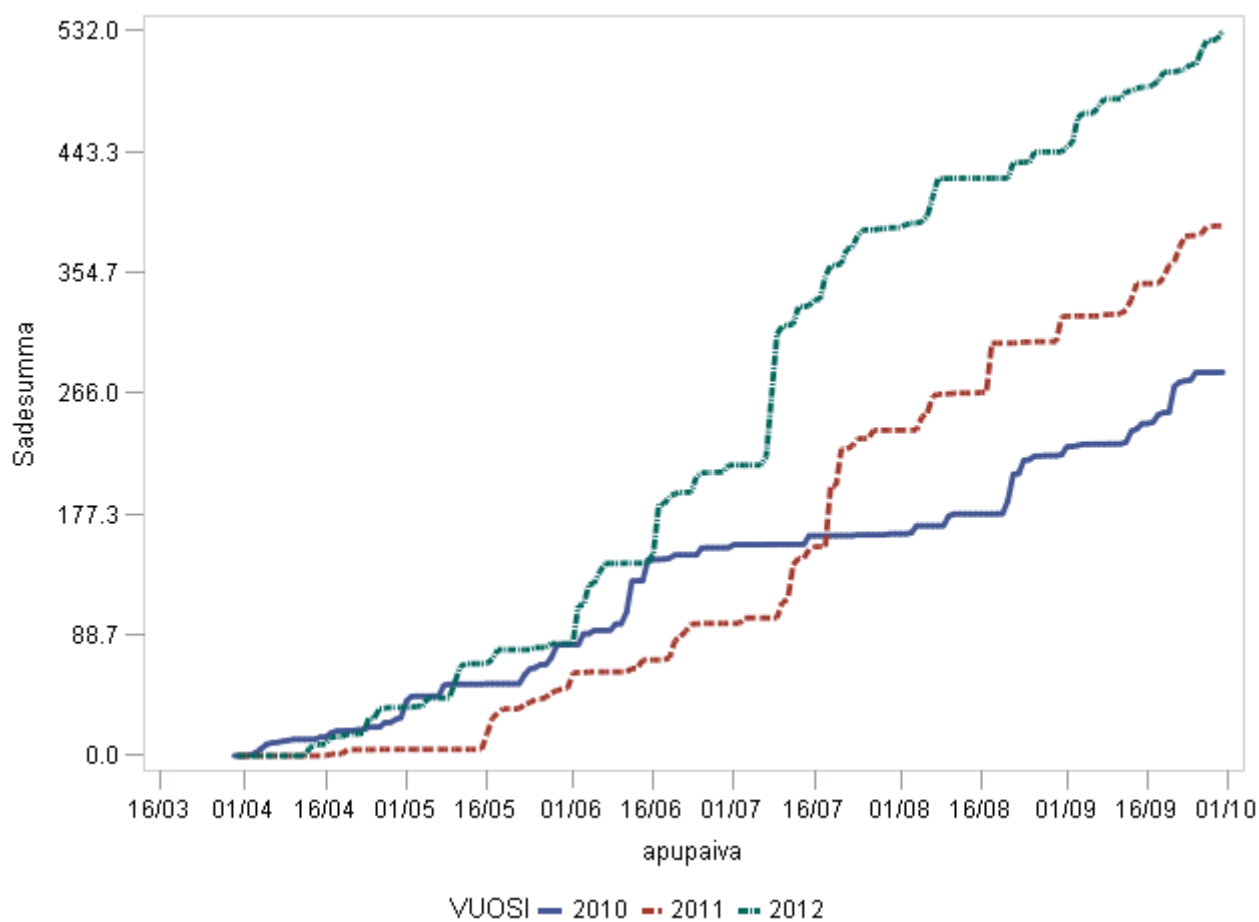
Sadesumman laskennassa käytetään SAS-peruskurssin aihealueen ohittavia ”kikkoja”. Aineiston järjestyksen tarkoituksena on ensinnäkin varmistaa, että aineisto on varmasti päivämäärän mukaan nousevassa järjestyksessä ja mahdollistaa by -lauseen käyttäminen data -vaiheessa. By -lause puolestaan mahdollistaa vuoden ensimmäisen havainnon löytämisen SASin apumuuttujalla **first.vuosi**, joka saa arvon yksi, kun aineistossa peräkkäisten datarivien vuosi vaihtuu. Tällöin kasvukausi vaihtuu ja sadesumman on oltava nolla. Retain lause pyytää SASia pitämään ssumma -muuttujan arvon samana kuin edellisellä havainnolla ellei sitä muuteta.

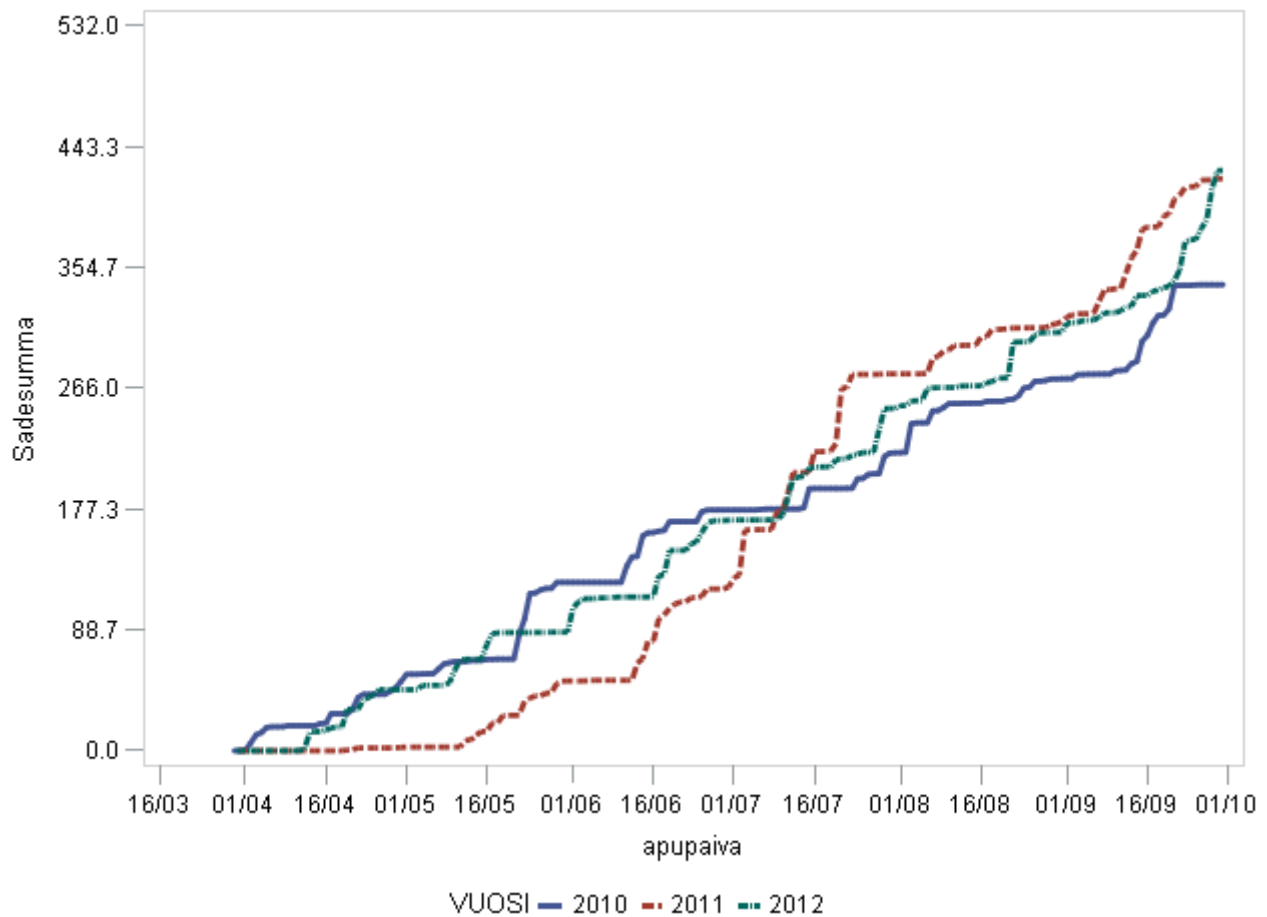
Kuva piirretään samaan tapaan kuin edellisetkin kuvat, nyt muuttujana on **ssumma**:

```
title "Sadesumma";
axis2 label=(angle=90 'Sadesumma') ;
proc gplot data=valitut uniform;
  by LPNN nimi;
  format apupaiva ddmmyy5.;
  plot ssumma * apupaiva = vuosi /vaxis=axis2 ;
  symbol i=join v=none w=3;
run;
```

Jokioisilla ei kovin suurta vuosien välistä vaihtelua ole ollut viime vuosina, kun taas Maaningalla on hyvin erilaiset sadesummaprofiilit eri vuosina.

LPNN=3603 NIMI=Maaninka Halola





3. 3h sääasema-aineiston tarkastelu

Tutkitaan seuraavassa hieman sääasemakohtaista 3h dataa. Haetaan aineisto Oracle -kannasta ja katsotaan sen sisältöä.

```
data kasvusaa3h; set saa_mtt.saata_kasvusaa_asema_3h;
run;

*muuttujat;
proc contents data=kasvusaa3h;
run;
```

Alphabetic List of Variables and Attributes					
#Variable	Type	Len	Format	Informat	Label
1ASEMA_TIETO_SEQNO	Num	8			ASEMA_TIETO_SEQNO
19DD	Num	8			DD
2HAKUASEMA_SEQNO	Num	8			HAKUASEMA_SEQNO
17HUM	Num	8			HUM
18HUMQ	Num	8			HUMQ
10KK	Num	8			KK
4LPNN	Char	4	\$4.	\$4.	LPNN
20MS	Num	8			MS
5NIMI	Char	50	\$50.	\$50.	NIMI
9PV	Num	8			PV
8PVM	Num	8	DATE TIME20.	DATE TIME20.	PVM
15RR	Num	8	7.1	7.1	RR
16RRQ	Num	8			RRQ
13TEMP	Num	8	6.1	6.1	TEMP
14TEMPQ	Num	8			TEMPQ
3TIETOASEMA_SEQNO	Num	8			TIETOASEMA_SEQNO
12UTC	Char	2	\$2.	\$2.	UTC
11VUOSI	Num	8			VUOSI
6X	Num	8			X
7Y	Num	8			Y

Tässäkin aineistossa **pvm** on tallennettu datetime -muodossa ja päivämäärän erottelu on tarpeen. Erityisen tärkeää on huomata **UTC** muuttujan kellonajan olevan UTC aikaa, joka on käytännössä sama kuin **GMT** eli **Greenwich Mean Time**. Jos mittaus on tehty kesäaikana, lisätään UTC-aikaan 3 tuntia ja talviaikana 2, jolloin saadaan mittauksen ajankohta Suomen aikavyöhykkeellä. Kellonaikaan on tallennettu tunti kahden merkin pituisena. Seuraavassa koodissa lisätään pvmkello muuttujan kellonaikaan 3h. Tästä seuraa, että UTC aikaa 21 oleva aika siirtyy seuraavan vuorokauden puolelle kello 0:00.

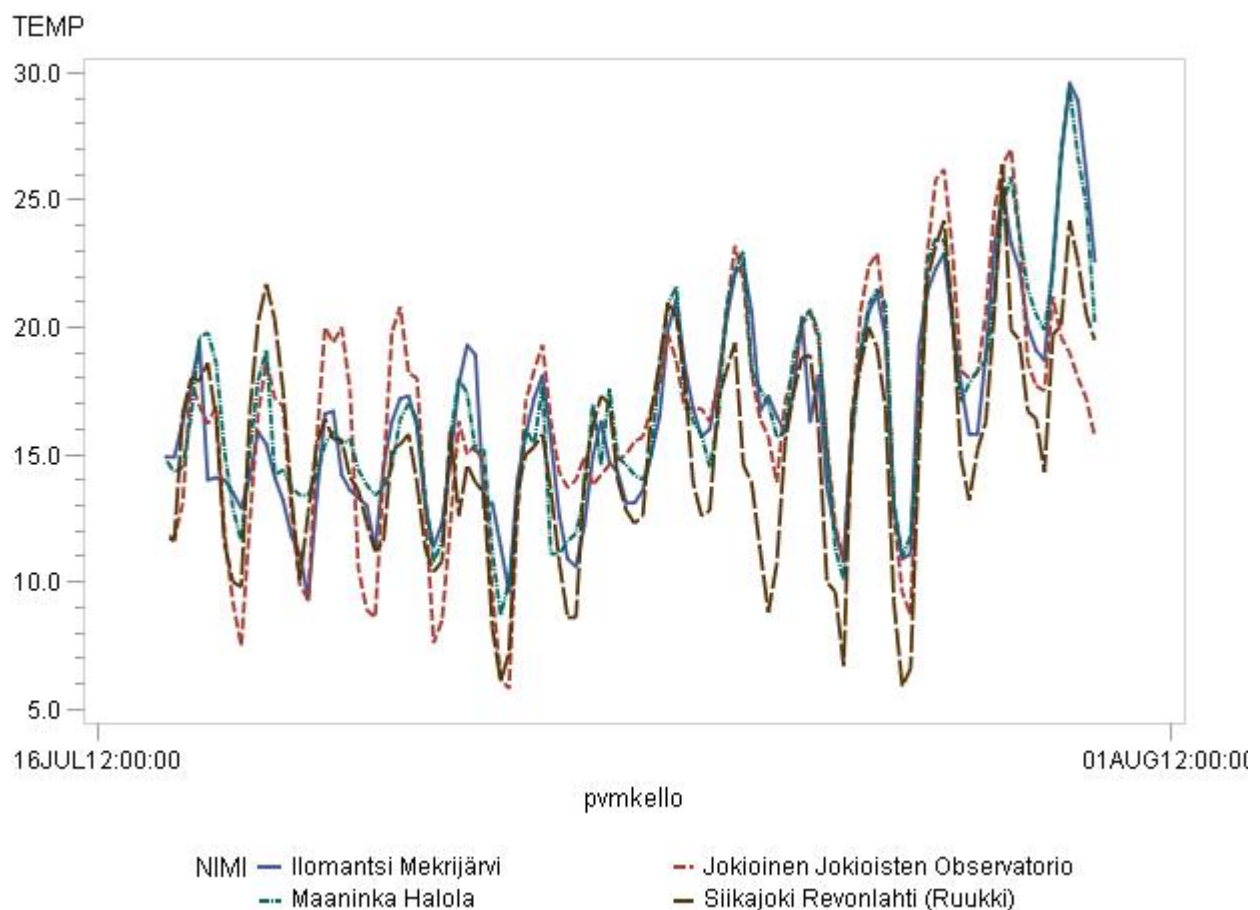
Muokataan aineistoon kellonaika ja päivämäärä SAS-muotoisiksi ja järjestetään aineisto, koska se ei välttämättä ole järjestyksessä.

```
data kasvusaa3h; set kasvusaa3h;
  format utckello hmmm5.;
  format paiva ddmmyy8.; * asetetaan tulostusformaatti paivalle;
  format pvmkello datetime13.;
  paiva = datepart( pvm );
  utckello= hms( inputn(UTC,2.), 0 , 0); *muutetaan kellonaika
  sasin time -muotoon;
  * muodostetaan datetime -muotoa oleva aikamuuttuja;
  pvmkello = dhms( paiva, hour(utckello) +3 ,0,0);
run;
proc sort data=kasvusaa3h;
  by lpnn pvm;
run;
```

Seuraavaksi katsotaan valituilta asemilta tietyn ajanjakson havaintoja. Huomaa, kuinka asemien ja ajanjakson valinta tehdään käyttäen **where** -lausetta suoraan **gplot** -proseduurissa. Ajanjakson valintaa varten päivämäärä syötetään SASin 'date constat' -muodossa.

```
title "Lämpötilahavainnot 3h välein";  
axis1 interval=even;  
proc gplot data= kasvusaa3h;  
  where LPNN IN ('3603', '1201', '5402', '3919') AND '17JUL2012'd <=  
  paiva <= '30JUL2012'd ;  
  plot temp * pvmkello = NIMI /haxis=axis1;  
  symbol i=join v=none w=2;  
run;
```

Lämpötilahavainnot 3h välein



Sääaineiston liittäminen omaan aineistoon

Esimerkkiaineisto

Seuraavassa luodaan kolme eri mittauspistettä sisältävä aineisto, joista jokaiselta on kaksi mittausta kesällä 2012.

```
data aineisto;
input koepaikka y x pv kk vuosi mittaus;
pvm = mdy(kk,pv,vuosi);
format pvm ddmmyy10.;
datalines;
1 6712879 3269388 10 7 2012 10
1 6712879 3269388 18 7 2012 15
2 7243778 3504193 25 7 2012 1
2 7243778 3504193 26 7 2012 2
3 6963365 3522947 1 6 2012 5
3 6963365 3522947 4 6 2012 19
;
```

Lähimmän sääaseman laskeminen ja tietojen hakeminen

Sääaineiston liittämistä varten on tiedettävä mitä halutaan liittää ja minkä aseman tietoja milloinkin käytetään. Yksittäisen koepaikan kohdalla sääaseman valinta on mahdollista tehdä käsin, mutta jos aineistossa on paljon mittauksia eri puolilta Suomea, on parempi turvautua automaattiseen mittauspaikan valintaan.

Ensimmäiseksi tarvitaan aineistosta jokaisen paikan sijaintitiedot ja vastaava tieto myös sääasemista.

```
* otetaan koepaikat ja niiden sijainnit talteen;
proc sort data=aineisto(keep=koepaikka y x) out=koepaikat nodupkey;
by koepaikka y x ;
run;
* otetaan sääaineistosta asematunnukset ja niiden sijainnit;
proc sort data=saa_mtt.saata_kasvusaa_asema_vrk(keep=lpnn nimi x y)
out=asemat nodupkey;
by lpnn x y;
run;
```

Sort proseduurit luovat SAS-datat **koepaikat** ja **asemat**. Seuraavassa vaiheessa lasketaan koepaikkojen etäisyys kaikkiin sääasemiin ja poimitaan jokaisesta lyhin. Koodi

```
* lasketaan jokaiselle pisteelle lähin asema;
proc sql;
```

```

create table lahin AS(
  select distinct a.koepaikka, b.lpnn,b.nimi,
    min( sqrt( (a.x - b.x)**2 + (a.y - b.y)**2 ) / 1000 )as etaisyys
  from work.koepaikat a, work.asetat b
  group by a.koepaikka
  having sqrt((a.x - b.x)**2 + (a.y - b.y)**2 )/1000 = etaisyys )
  order by a.koepaikka;
quit;

```

Tämä SQL -koodi luo SAS-datan **lahin**, jonka sisältö on:

Obs	koepaikka	LPNN	NIMI	etaisyys
1	10103	Kaarina	Yltöinen (Piikkiö)	16.1629
2	25502	Vaala	Pelso	93.6124
3	33603	Maaninka	Halola	41.8010

Tätä dataa editoimalla voi muuttaa koepaikkaan liitettävää asemaa, jos lähin asema ei edusta koepaikalla vallitsevia ilmasto-olosuhteita. Näin voi käydä esimerkiksi, jos koepaikka sijaitsee sisämaassa ja lähin asema on rannikolla tai päinvastoin. Tässä esimerkissä koepaikkaa 2 lähinnä oleva asema on 93 kilometrin päässä. Aseman ja koepaikan säätilojen yhdenmukaisuus on vähintäänkin kyseenalaista.

Seuraavaksi liitetään lähimmän aseman LPNN tunnus aineistoon.

```

proc sort data= aineisto;
  by koepaikka;
run;
data aineisto; merge aineisto lahin;
  by koepaikka;
run;

```

Haetaan LPNN tunnuksen ja päivämäärän perusteella sääaineistosta lämpösumma, keskilämpötila, sade ja sateen laatuluokitus.

```

proc sql ;
  create table aineisto1 AS(
    select a.*, b.nimi, b.ets5, b.temp, b.rr, b.rrq
    from work.aineisto a LEFT JOIN
    saa_mtt.saata_kasvusaa_asema_vrk b ON( a.lpnn = b.lpnn AND a.pvm =
    datepart(b.pvm) )
  );
quit;

```

Tämän jälkeen **aineisto1** näyttää tältä (lähimmän aseman nimi poistettu)

koepaikka	y	xpv	kk	vuosi	mittaus	pvm	LPNN	etaisyys	ETS5	TEMP	RR	RRQ
16712879326938810	7	2012			1010/07/2012	20103	16.1629556.7	17.0	5.4	0		
16712879326938818	7	2012			1518/07/2012	20103	16.1629641.5	14.1	3.5	0		
369633653522947	1	6	2012		501/06/2012	23603	41.8010161.5	10.9	-1.0	0		
369633653522947	4	6	2012		1904/06/2012	23603	41.8010175.6	9.215.0		0		
27243778350419325	7	2012			125/07/2012	25502	93.6124562.7	14.8	4.0	5		
27243778350419326	7	2012			226/07/2012	25502	93.6124572.3	14.6	-1.0	0		

Pitkäaikaissää

Pitkäaikaissääksi kutsutaan sääasemilta kerättyä sääaineistoa, joka yltää jollain asemilla aina 1970 vuoden alkuun saakka. Pitkäaikaissää ei vielä ole tietokannassa, vaan se joudutaan hakemaan manuaalisesti linux-palvelimelta apollon.mtt.fi. Tunnuksia palvelimeen ja apua sääaineiston hakuun saa IT-tuesta. Aineistosta on kaksi versiota, koko aineisto ja viimeisimmän 8 vuoden aineisto. Molemmat aineistot sekä paljon muita sääaineistoja löytyy valmiiksi tehtynä SAS-datana verkkolevyltä polusta: **J:\MTT_yhteiset\Sääaineistot**. Apollon -palvelimella pitkäaikassäät löytyvät hakemistosta **/data/saa/aineistot**. Koko aineisto on nimellä **saa19702011.sas7bdat** ja kahdeksan edellistä vuotta sisältävä aineisto on nimeltään **saakanta.sas7bdat**.

Tarkastellaan miltä pitkäaikaissää näyttää. Aluksi on määriteltävä kirjastoviite hakemistoon, jossa ko. tiedostot ovat. Sen jälkeen korjataan aineiston päivämäärämuuttuja oikeaksi ja liitetään asemien nimet toiseseta tiedostosta.

```
libname pitka "J:\MTT_yhteiset\Sääaineistot";
* pitkäaikaissää;
data psaa; set pitka.saa19702011;
    * tehdään pvm -muuttuja kuntoon;
    pvm = mdy( kk, pv, vuosi ) ;
run;

* asemien nimet;
data asemat; set pitka.asemat;
    asema = inputn(lpnn, 4.);
    keep asema nimi;
run;

proc sort data=psaa ;
    by asema vuosi kk pv;
run;
proc sort data= asemat;
    by asema;
run;
```

```

* liitetään aseman nimi pitkäaikaaisdataan;
data psaa; merge psaa(in=A) asemat(in=B);
  by asema;
  if(A);
run;
* katsotaan sisältoa;
proc contents data=psaa ;
run;

* katsotaan havaintohistorian kesto;
proc tabulate data=psaa;
  var pvm;
  class asema asemannimi ;
  table asema*asemannimi, (min max)*pvm=' '*f=ddmmyy9.;
run;

```

Alphabetic List of Variables and Attributes				
#Variable	Type	Len	Format	Informat Label
1ASEMA	Num	84.	15.2	HAVAINTOASEMA
19AURINKO	Num	8	4.1	AURINGONPAISTETUNNIT VRK:SSA
22KASVUAL	Num	81.		KASVUKAUDEN ALK.PAIVA
				LAMPOSUMMA TERM. KEVAAN
8KELAMPOS	Num	8		ALKAMISESTA
7KESKIL	Num	8	4.1	VRK KESKILAMPOTILA
5KK	Num	82.		KUUKAUSI
2KUNTA	Num	8		KUNTA
				TEHOISA LAMPOTILASUMMA
9LAMPOS	Num	8		KASVUK.ALUSTA
12MAAMIN	Num	8	4.1	MAANPINNAN MIN.LAMPOT.
10MAX1	Num	8	4.1	VRK MAKSIMILAMPOT.
11MIN1	Num	8	4.1	VRK MINIMILAMPOT.
6PV	Num	82.		PAIVA
		DDMMYY10		
24PVM	Num	8.		
13SADE	Num	8	4.1	VRK SADEMAARA.
14SADESUM	Num	8		SADESUMMA VUODEN ALUSTA
20SATEILY	Num	8	4.	KOKONAISSATEILY kJ/NELIOM
23SUHTKKA	Num	8	3.	suhtkka
				SUHTEELLINEN KOSTEUS KLO 12
15SUHTKOST	Num	8	3.	UTC
21TERMIAL	Num	81.		TERMISEN KEVAAN ALK.PAIVA
16TUULIKA	Num	8	2.	TUULEN NOPEUS VRK K-ARVO
18TUULINOP	Num	8	2.	TUULEN NOPEUS KLO 12 UTC
17TUULISUU	Num	8	3.	TUULEN SUUNTA KLO 12 UTC
4VUOSI	Num	84.	8.2	VUOSI
3YKS	Num	8		TUTKIMUSYKSIKKO
26_FREQ_	Num	8		
25_TYPE_	Num	8		

Tässä aineistossa **ASEMA** vastaa kasvukauden sään **LPNN** -muuttujaa. **ASEMA** tallennettu numeromuodossa toisin kuin kasvukauden aineistossa. Sadesummaa kasvukauden alusta Ei ole aineistossa valmiiksi laskettuna, vaan se täytyy laskea itse samaan tapaan kuin kasvukauden aineistossakin tehtiin.

Tarkastellaan seuraavaksi maaliskuun keskilämpötiloja tietyiltä asemilta.

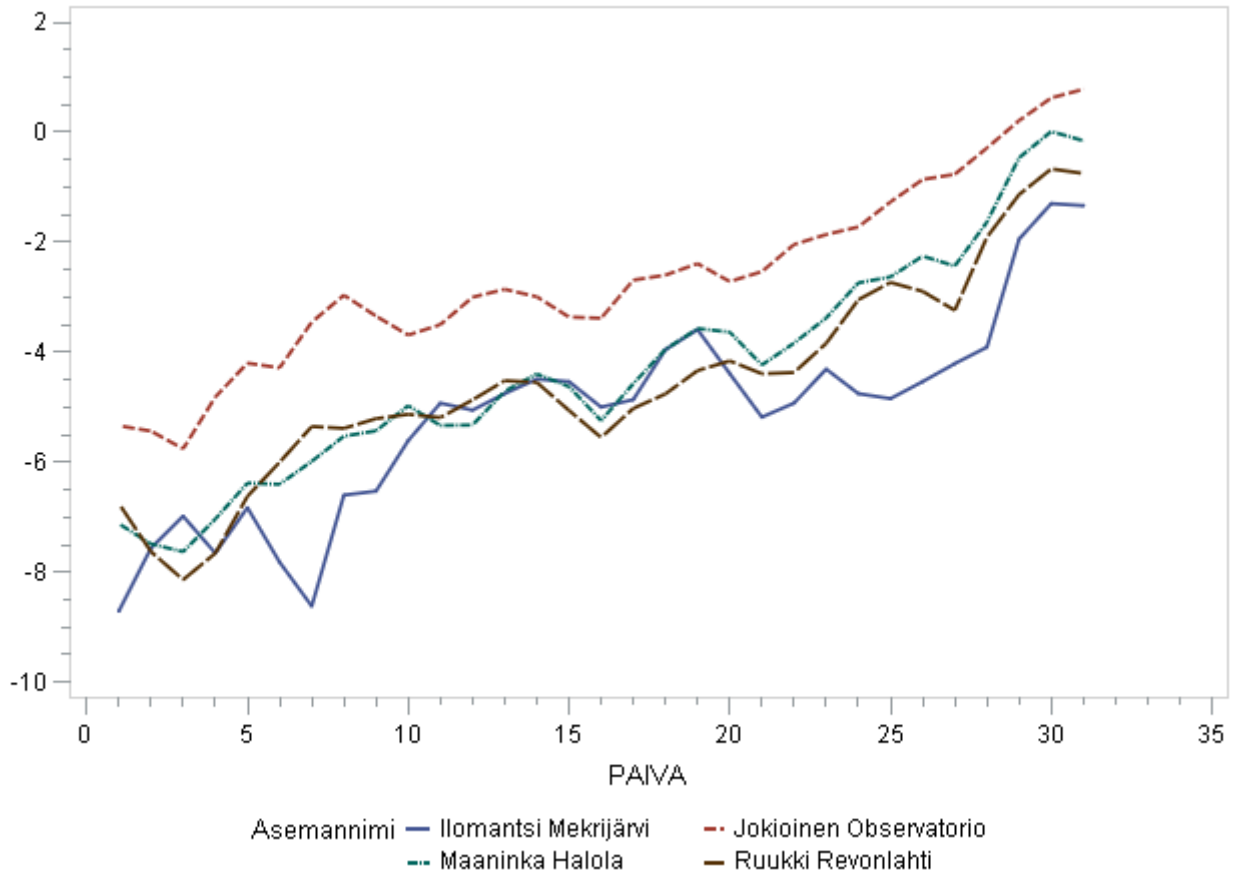
```
* Lasketaan joidenkin asemien maaliskuiden keskilämpötilat. ;  
proc means data=psaa nway;  
  where asema IN( 3603, 1201,5402, 3919 ) AND KK=3;  
  class asema asemannimi pv;  
  var keskil;  
  output out=maaliska mean=;  
run;
```

Piirretään asemien keskilämpötilojen kehitys kuvaan

```
proc gplot data=maaliska;  
  plot min1 * pv = asemannimi;  
  symbol i=join v=none w=2 ;  
run;
```

Maaliskuun keskilämpötilojen keskiarvo

VRK KESKILAMPOTILA

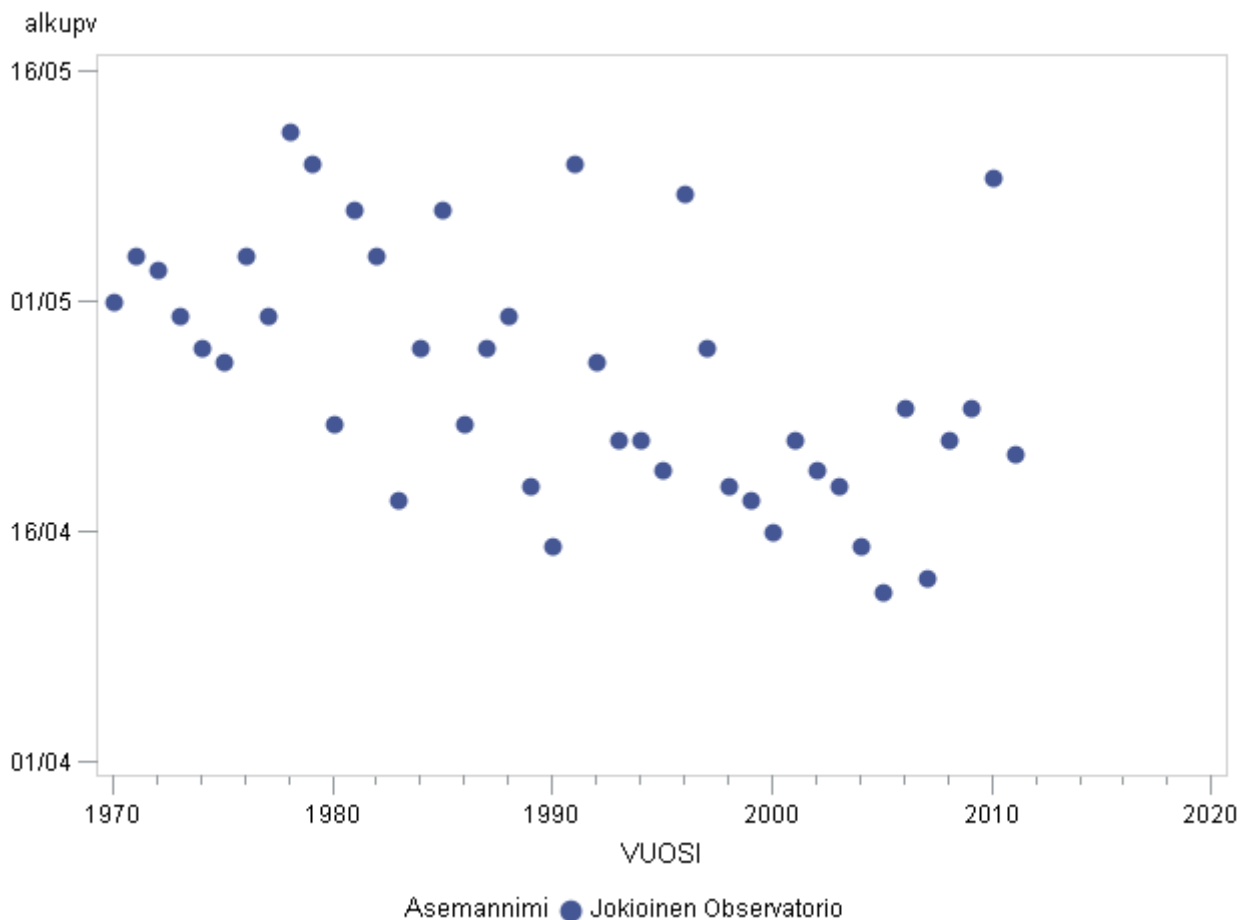


Katsotaan seuraavassa kasvukauden alkamispäivämäärät vuosittain. Kasvukauden alkamispäivänä muuttuja **KASVUAL** saa arvon 1, muulloin arvo on puuttuva.

```
data kasvualku; set psaa;  
where asema IN( 3603, 1201, 5402, 3919 ) AND kasvual=1;  
alkupv = mdy(kk, pv, 2010);  
format alkupv ddmmyy5.;  
run;  
title "Kasvukauden alkupäivämäärä";  
proc gplot data = kasvualku;  
by asema asemannimi;  
plot alkupv*vuosi=asemannimi;  
symbol i=none v=dot;  
run;
```

Kasvukauden alkupäivämäärä

HAVAINTOASEMA=1201 Asemannimi=Jokioinen Observatorio



Keskimääräinen kasvukausi vs nykyinen

Joskus voi olla tarvetta verrata kuluva kasvukautta edellisiin kasvukausiin. Katsotaan miten vuoden 2012 kasvukausi suhteutui lämmön ja sateen suhteen pitkäaikaiseen keskiarvoon.

Pitkäaikaissääaineistosta saadaan lasketuksi haluttuja keskiarvoja. Lasketaan vertailuksi 30 vuoden keskiarvot lämpösummasta, sadesummasta, vuorokauden keskilämpötilasta ja keskimääräisestä sateesta.

```
* pitkäaikaissäästä 30v keskiarvo huhti, touko, kesä, heinä, elo ja syys;  
* vuorokauden keskilämpötila, keskimääräinen sade, lämpösumma ja sadesumma;  
proc sort data=psaa;  
  by asema vuosi kk pv;  
run;  
  
*lasketaan sadesumma kasvukauden alusta;  
data psaa; set psaa;  
  by asema vuosi;  
  retain ssumma 0;  
  * jos vuoden 1. havainto, on sadesumma 0;  
  if first.vuosi then ssumma=0;  
  * jos lämpösumma on > 0, on kasvukausi alkanut ja sadesummaa lasketaan;  
  * tarkistetaan myös, että sade > 0 eli jotain sadetta on tullut;  
  if ( lampos > 0 )and ( sade > 0 ) then ssumma = ssumma + sade;  
run;
```

Seuraavassa oleva means proseduuri laskee huhtikuusta syyskuuhun jokaiselle päivälle keskiarvon vuosien 1982 - 2011 havainnoista.

```
* otetaan vuosien yli keskiarvo;  
proc means data=psaa noprint nway ;  
  where 1982 <= vuosi <= 2011 and 4 <= kk <= 9;  
  class asema asemannimi kk pv;  
  var keskil sade lampos ssumma;  
  output out=pka mean=lampoka sadeka lska sska;  
run;  
  
* muutetaan keskiarvodatan päivämäärät vuodelle 2012;  
data pka; set pka;  
  vuosi = 2012;  
  paivays = mdy( kk, pv, vuosi);  
run;  
  
proc sort data=pka;  
  by asema paivays;  
run;
```

```

* yhdistetään vuoden 2012 kasvukauden dataan;
data kasvu2012; set saa_mtt.saata_kasvusaa_asema_vrk ;
  where vuosi = 2012;
  *muutetaan lpnn merkkijono numeroksi;
  asema = inputn( lpnn, 4.);
  *muutetaan pvm sas-pvm:ksi;
  paivays = datepart(pvm) ;
run;
proc sort data= kasvu2012;
  by asema paivays;
run;

* lasketaan sadesumma;
data kasvu2012; set kasvu2012;
by asema ;
retain ssumma 0;
if first.asema OR lag(vuosi) NE vuosi then ssumma=0;
* jos lämpösumma on > 0, on kasvukausi alkanut ja sadesummaa
lasketaan;
* tarkistetaan myös, että RR > 0 eli jotain sadetta on tullut;
if ( ets5 > 0 ) and ( RR > 0 ) then ssumma = ssumma + RR;
run;

data kasvu2012; merge kasvu2012(in=a) pka(in=b);
  by asema paivays;
  if a;
run;

```

Nyt datassa **kasvu2012** on kasvukauden 2012 havaintojen lisäksi 30 vuoden keskiarvot lämpösummasta, sadesummasta, vuorokauden keskilämpötilasta ja vuorokauden keskimääräisestä sademäärästä kullekin päivälle. Näitä voidaan visualisoida eri tavoilla, seuraavassa muutamia.

```

title "Lämpösumma 2012 ja pitkäaikaikeskiarvo";
proc gplot data= kasvu2012 uniform;
  format paivays ddmmyy5.;
  where asema IN(1201, 5402, 3603 ) and 5 <= kk <= 9;
  by asema nimi;
  plot ets5*paivays = 1 lska*paivays=2 /overlay;
  symbol1 i=join v=none w=2 color=black ;
  symbol2 i=join v=none w=2 color=green ;
run;

```

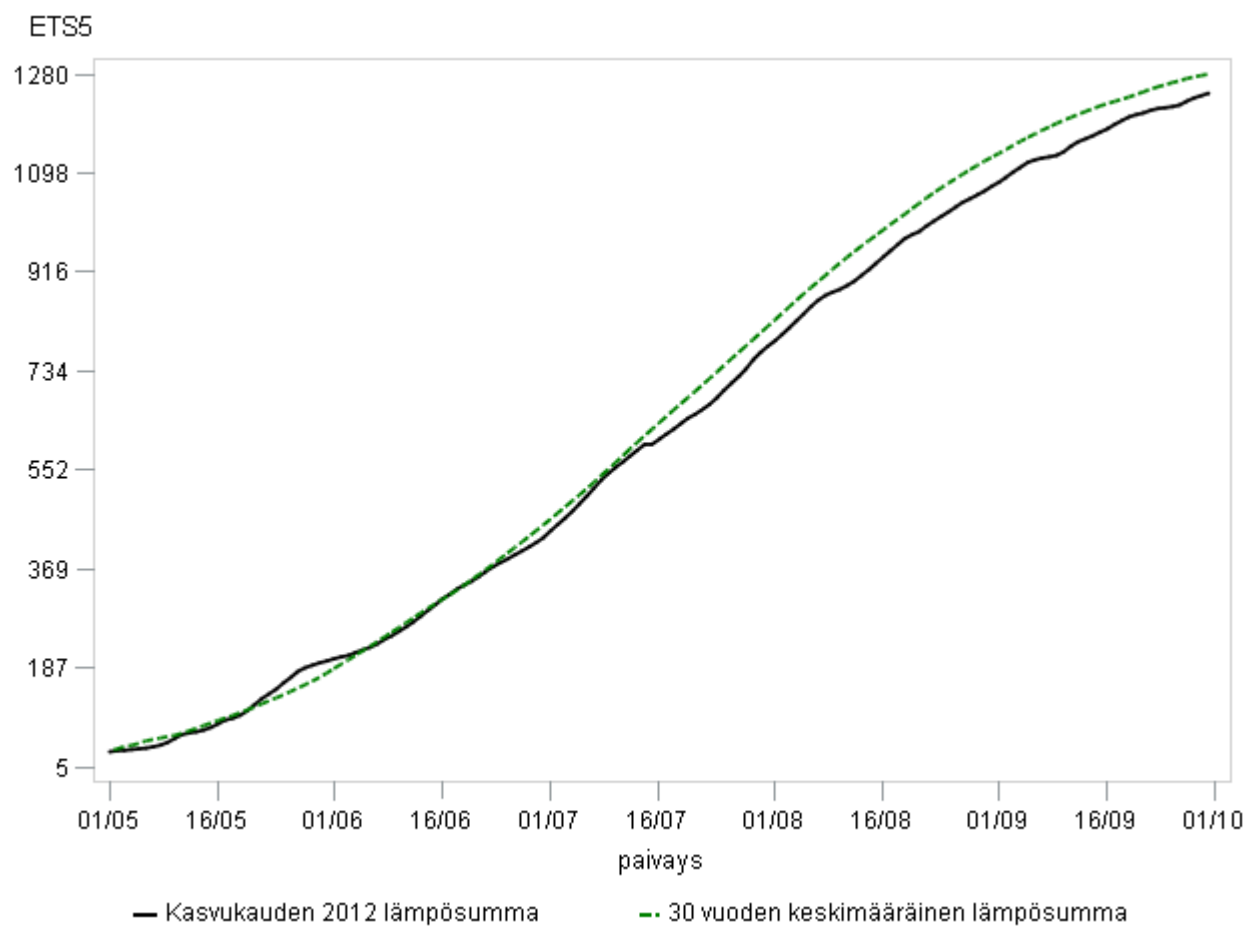
```

title "Sadesumma 2012 ja pitkäaikaikeskiarvo";
proc gplot data= kasvu2012 uniform;
  format paivays ddmmyy5.;
  where asema IN(1201, 5402, 3603) and 4 <= kk <= 9;
  by asema nimi;
  plot ssumma*paivays = 1 sska*paivays=2 /overlay;
  symbol1 i=join v=none w=2 color=black ;
  symbol2 i=join v=none w=2 color=green ;
run;

```

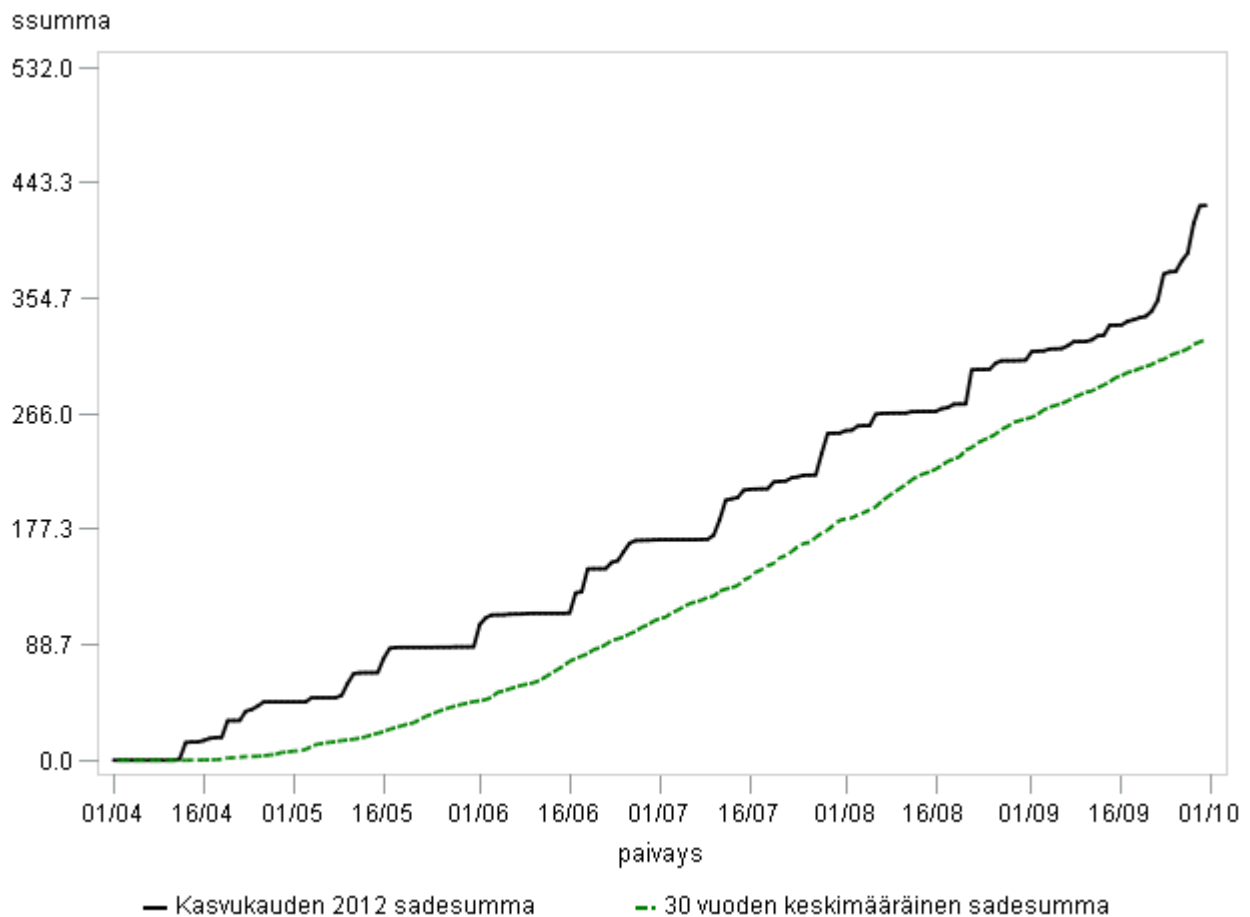

Lämpösumma 2012 ja pitkäaikaiskeskiarvo

HAVAINTOASEMA=1201 NIMI=Jokioinen Jokioisten Observatorio



Sadesumma 2012 ja pitkäaikaiskeskiarvo

HAVAINTOASEMA=1201 NIMI=Jokioinen Jokioisten Observatorio



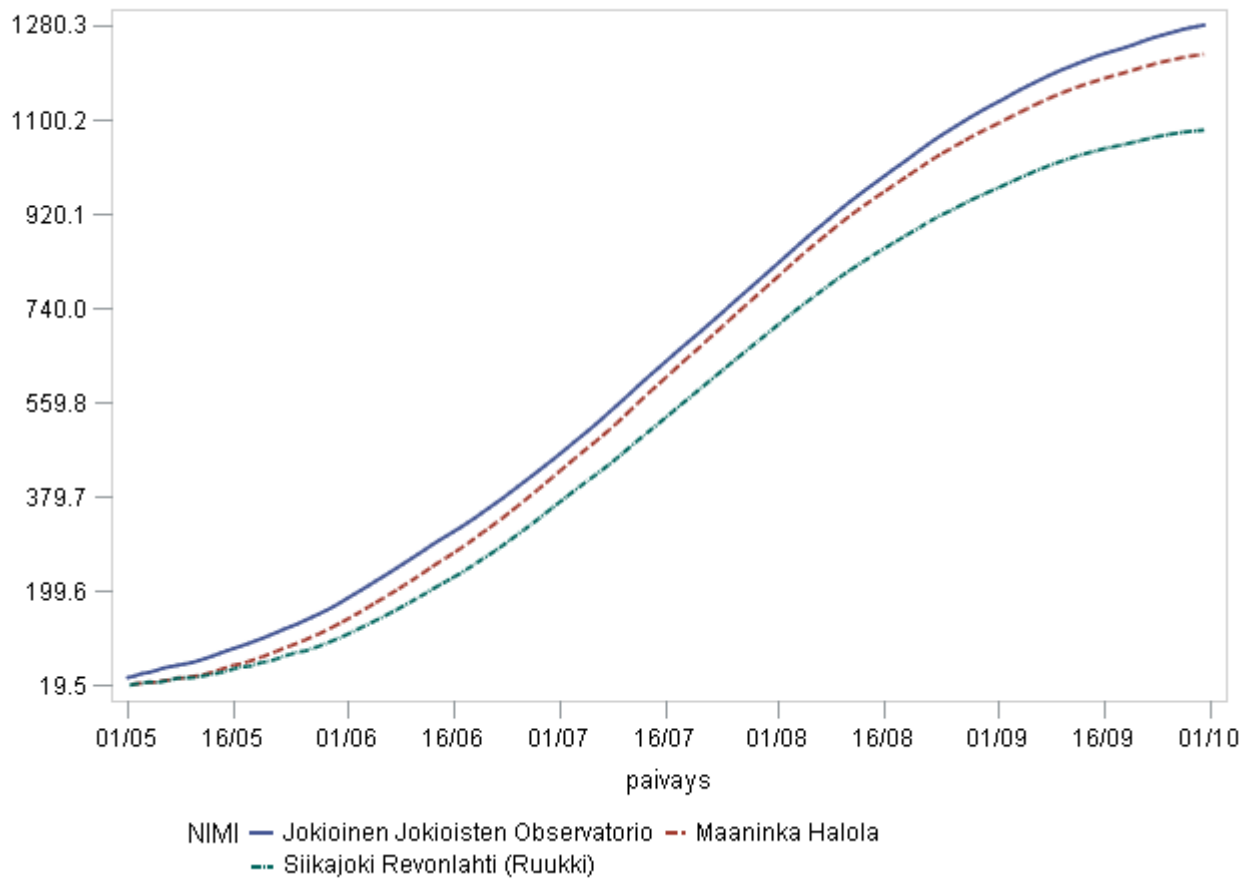
Eri havaintoasemien käyrät samaan kuvaan

```
* Eri asemien keskiarvot samaan kuvaan;
goptions reset=symbol;
title "Havaintoasemien pitkäaikaiskeskiarvot";
proc gplot data= kasvu2012 uniform;
  format paivays ddmmyy5.;
  where asema IN(1201, 5402, 3603 ) and 5 <= kk <= 9;
  plot lska*paivays=nimi ;
  symbol i=join v=none w=2;
run;

proc gplot data= kasvu2012 uniform;
  format paivays ddmmyy5.;
  where asema IN(1201, 5402, 3603 ) and 5 <= kk <= 9;
  plot sska*paivays=nimi ;
  symbol i=join v=none w=2;
run;
```

Havaintoasemien pitkäaikaikeskiarvot

TEHOISA LAMPOTILASUMMA KASVUK.ALUSTA



Piirretään vielä keskilämpötilat kesän ajalta.

```
proc gplot data= kasvu2012;  
  format paivays ddmmyy5.;  
  where asema IN(1201, 5402, 3603 );  
  plot lampoka * paivays=nimi /vref=0;  
  symbol i=join;  
run;
```

Havaintoasemien pitkäaikaikeskiarvot

VRK KESKILAMPOTILA

