

# **Harjoitustyöraportti**

## **Tiedonlouhinta**

Markus Johansson YM3  
Kimmo Johansson TK5  
Mikko Koponen TK3  
29.5.2013

## Sisältö

<b>1</b>	<b>JOHDANTO</b> .....	<b>3</b>
<b>2</b>	<b>AINEISTO JA MENETMÄT</b> .....	<b>4</b>
<b>2.1</b>	<b>AINEISTON KUVAUS</b> .....	<b>4</b>
<b>2.2</b>	<b>MENETELMÄT</b> .....	<b>9</b>
2.2.1	<i>Varianssiskaalaus</i> .....	9
2.2.2	<i>k-means klusterointi</i> .....	10
2.2.3	<i>Davies-Bouldin indeksi</i> .....	10
2.2.4	<i>Itseorganisoituva kartta</i> .....	10
2.2.5	<i>Sammonin kuvaus</i> .....	12
<b>3</b>	<b>ESIPROSESSOINTI</b> .....	<b>12</b>
<b>3.1</b>	<b>AINEISTON SIIVOUS</b> .....	<b>12</b>
<b>3.2</b>	<b>AINEISTOJEN YHDISTÄMINEN</b> .....	<b>13</b>
<b>4</b>	<b>TULOKSET</b> .....	<b>15</b>
<b>4.1</b>	<b>AINEISTON KUVAAJAT</b> .....	<b>15</b>
4.1.1	<i>Aikasarjakuvaajat</i> .....	15
4.1.2	<i>Autokorrelaatiot</i> .....	18
4.1.3	<i>Ristikorrelaatiot</i> .....	21
<b>4.2</b>	<b>HIILIDIOKSIDI JA TVOC PITOISUUKSIEN ANALYSOINTI</b> .....	<b>22</b>
<b>4.3</b>	<b>ITSEORGANISOITUVA KARTTA</b> .....	<b>31</b>
<b>4.4</b>	<b>ITSEORGANISOITUVAN KARTAN KLUSTEROINTI K –MEANS MENETELMÄLLÄ</b> .....	<b>33</b>
<b>4.5</b>	<b>KLUSTERIEN HISTOGRAMMIT</b> .....	<b>34</b>
<b>5</b>	<b>JOHTOPÄÄTÖKSET</b> .....	<b>37</b>

Markus

## 1 JOHDANTO

Tässä harjoitustyöraportissa esitellään sisäilmanlaadun analysointia erilaisilla menetelmillä. Sisäilmanlaadun aineisto on mitattu kymmenestä Kuopion alueen kouluista ja yhdestä omakotitalosta. Sisäilmanlaadusta analysointiin periodiset komponentit ja omakotitalosta tehtiin klusterianalyysi. Sisäilmanlaadun tutkiminen liittyy ympäristöterveyteen, josta on lyhyt esittely alla olevissa kappaleissa.

### Ympäristöterveys

Ympäristöterveydellä tarkoitetaan miten ympäristön altisteet vaikuttavat ihmisen terveyteen. Ympäristöaltisteet jaetaan fysikaalisiin, kemiallisiin ja biologisiin tekijöihin. Fysikaalisia altisteita ovat esimerkiksi melu, lämpötila, ilmankosteus, säteily ja valaisuolosuhteet. Kemialliset altisteet voivat olla hiukkasmaisia tai kaasumaisia, jotka voidaan jakaa orgaanisiin ja epäorgaanisiin yhdisteisiin esimerkiksi savukaasut, liottimet tai sisäilman hiilidioksidi. Biologisiin altisteisiin kuuluvat esimerkiksi virukset, bakteerit ja sienet, joista esimerkiksi bakteerien ja sienien (esimerkiksi homeet) aineenvaihdunnassa vapautuvat orgaaniset yhdisteet tai itiöt ovat terveydelle haitallisia.

### Sisäilma

Nykyään ihminen oleilee sisätiloissa suurimman osan ajasta työpaikoilla ja kotona. Sisäilman fysikaaliset, kemialliset ja biologiset altisteet voivat aiheuttaa vakavia terveyshaittoja, jos sisäilman laatu ei ole hyvä. Sisäilman laatu vaikuttaa työtehokkuuteen ja viihtyisyyteen. Fysikaalisista altisteista lämpötila ja ilmankosteus vaikuttavat terveyteen ja viihtyisyyteen. Liian kylmä tai lämmin sisäilma aiheuttaa työtehon ja viihtyisyyden alenemista. Suositeltava sisäilman lämpötila on noin +21 °C. Ilmankosteudella on suuri merkitys terveyteen ja viihtyisyyteen. Suositeltava suhteellinen ilmankosteus tulisi olla 20 – 60 %. Liian matala ilmankosteus (alle 20 %) kuivattaa limakalvoja ja siten aiheuttaa terveyshaittoja. Liian korkeassa ilmankosteudessa kosteus saattaa tiivistyä rakenteisiin ja siten saattaa nopeuttaa bakteerien ja sienien kasvua. Sisäilman kemiallisia tekijöitä on paljon esimerkiksi rakennusmateriaaleista haittavat yhdisteet, pesuaineet ja ihmisen aineenvaihdunnassa syntyvä hiilidioksidi. Rakennusmateriaaleista ja pesuaineista haittavia yhdisteitä sanotaan yleisesti haihtuvat orgaanisen yhdisteet, josta käytetään lyhennettä VOC (Volatile organic compound). Sisäilmaan haihtuvien yhdisteiden kokonaismäärää sanotaan usein TVOC –pitoisuudeksi (*Total Volatile Organic Compounds*). Suurin osa haihtuvista orgaanisista yhdisteistä ei aiheuta vakavaa terveyshaittaa, mutta ne voivat aiheuttaa esimerkiksi päänsärkyä. Sisäilman biologiset altisteet tulevat yleensä rakennuksen rakenneviasta. Näitä ovat esimerkiksi ilmavuodot ja kosteusvauriot. Kosteusvauriot yleensä huonontavat sisäilman laadun varsin nopeasti. Syynä on bakteeri ja sieni kasvustot, jotka aiheuttavat hengityselintulehduksia ja allergioita. Sisäilman hiilidioksidipitoisuus voi kohota paljon, jos tilassa on riittämätön ilmavaihto. Tällöin suuri hiilidioksidipitoisuus voi aiheuttaa väsymystä ja päänsärkyä. Sisäilman hiilidioksidipitoisuudelle on asetettu sisäilmastoluokituksen mukaiset enimmäispitoisuudet, jotka ovat esitetty alla:

- Sisäilmastoluokka S1 700 ppm
- Sisäilmastoluokka S2 900 ppm
- Sisäilmastoluokka S3 1200 ppm

jossa S1 pidetään hyvänä, S2 tyydyttävänä ja S3 huonona

### **Ympäristöinformatiikan sisäilman mittaukset**

Ympäristöinformatiikan tutkimusryhmä tekee Kuopion alueella koulusta, kerrostaloista ja omakotitaloista sisäilman laadunmittauksia. Tavoitteena on mitata pitkiä aikasarjoja, jotta saadaan selville esimerkiksi vuodenajan vaihteluiden vaikutukset sisäilmanlaatuun.

### **Mittalaitteet**

Mittalaitteiden mittaustiheys on joko 10 sekuntia tai 1 minuutti. Mittaukset ovat olleet toiminnassa yli kaksi vuotta (mittaukset aloitettu 2010/2011). Mittalaitteista suurin osa mittaa sisäilman lämpötilan, suhteellisen ilmankosteuden, hiilidioksidipitoisuuden ja TVOC pitoisuuden. Lisäksi on joitakin yksittäisiä mittalaitteita, jotka mittaavat häkäpitoisuuden, paine-eron ja pienhiukkasmäärän. Mittalaitteet lähettävät mittaustulokset langattomilla tiedonsiirtotekniikoilla, joissa tapahtuu usein tiedonsiirtovirheitä tai koko mittaustulos katoaa kun radioilla ei ole hyvä kuuluvuus. Mittaustulokset tallentuvat ympäristöinformatiikan tutkimusryhmän palvelimen tietokantaan.

Ympäristöinformatiikan tutkimusryhmän yleisemmät mittaukset

- Lämpötila
  - o 0- 50 °C
- ilmankosteus
  - o 0 – 100 %
- hiilidioksidi (CO<sub>2</sub>)
  - o 0 - 2000 ppm
- TVOC
  - o 0 – 30 ppm
  - o mittaa kokonaispitoisuuden: tolueeni (C<sub>6</sub>H<sub>5</sub>CH<sub>3</sub> / C<sub>7</sub>H<sub>8</sub>), ammoniakki (NH<sub>3</sub>) ja rikkivety (H<sub>2</sub>S)
  - o muut: hiki, tärpähti ja muut liuottimet

## **2 AINEISTO JA MENETLMÄT**

Tässä kappaleessa esitellään analysoinnissa käytetyt aineistot ja menetelmät.

Markus, Mikko

### **2.1 Aineiston kuvaus**

Mittausaineistot ovat ajalta 11.10.2011 - 13.7.2012. Joihinkin kohteisiin on asennettu lisää mittalaitteita edellä mainitun ajanjakson aikana, joten aineistossa on puuttuvia arvoja. Puuttuvat arvot ovat merkitty ”NaN” merkkeinä, jota matlab –ohjelma käsittelee arvona not a number. Aineistossa on joissakin mittauksissa aika paljon puuttuvia arvoja, koska mittalaitteet eivät välillä

toimineet kunnolla. Aineiston ensimmäinen sarake on aikaleima sekuntiresoluutiolla Unix aikamuodossa eli joka alkanut 1.1.1970 arvosta 0 ja kasvaa sekunnin välein eli on kumuloituva.

Koulumittausaineistossa on eri koulujen mittaukset siten, että aikaleiman jälkeen on Paloahon koulu. Koulut ovat eroteltu sarakkeilla, joissa on numerot 1-10. Siten, että numeron jälkeen alkaa toisen koulun mittaukset ja päättyvät seuraavaan numerosarakkeeseen.

Koulut ovat järjestyksessä:

1 Koulu 1

2 Koulu 2

3 Koulu 3

4 Koulu 4

5 Koulu 5

6 Koulu 6

7 Koulu 7

8 Koulu 8

9 Koulu 9

10 Koulu 10

Liitteessä 1 on kouluaineiston muuttujien tilastolliset ominaisuudet. Omakotitalon muuttujien tilastolliset ominaisuudet on esitetty taulukossa 1.

Taulukko 1. Omakotitalon mittauksien tilastolliset ominaisuudet

	yksikkö	minimi	maksimi	keskiarvo	mediaani	keskihajonta
Temperature Fireplace	°C	27,7	81,9	34,0	29,4	9,3
CO OH	ppm	0,0	0,4	0,0	0,0	0,0
Temperature ET	°C	21,6	25,6	23,3	23,2	0,5
Humidity ET	%RH	11,3	57,2	27,3	26,5	7,8
CO2 ET	ppm	404,2	1117,4	521,8	513,8	73,3
Diff Pressure	Pa	-1,0	0,5	-0,3	-0,4	0,2
Temperature PH	°C	17,4	29,1	22,3	22,4	1,2
Humidity PH	%RH	13,4	67,9	30,9	28,8	9,5
CO2 PH	ppm	324,0	936,5	407,4	397,6	55,3
MH1 Humidity	%RH	14,6	48,7	28,7	27,6	6,3
MH1 Temperature	°C	21,8	27,4	23,5	23,2	0,9
MH1 CO2	ppm	436,9	1980,5	765,9	674,6	287,5
MH2 Humidity	%RH	11,4	50,1	27,4	27,1	6,6
MH2 CO2	ppm	365,9	1971,1	663,7	559,2	316,5
MH2 Temperature	°C	19,5	26,3	22,8	22,7	1,1
MH3 Temperature	°C	20,7	25,8	23,6	23,6	0,5
MH3 Humidity	%RH	13,1	53,9	26,6	25,5	7,0
MH3 CO2	ppm	319,3	1229,2	513,6	506,7	116,6
MH3 VOC	ppm	1,7	27,3	14,3	14,3	2,5
OH PIR	on/off	0,0	1,0	0,4	0,0	0,5
OH Humidity	%RH	10,9	63,8	25,5	23,9	8,1
OH Temperature	°C	20,0	37,7	23,1	22,9	1,0
OH CO2	ppm	460,1	1423,0	629,5	623,3	105,2
Electricity (Har1)	Wh	2176503,1	8672516,1	6038356,1	6378812,4	1972293,5
Water (Har1)	m <sup>3</sup>	136,9	1115,3	717,2	769,0	297,0
Ulkolämpötila	°C	-29,7	24,6	1,8	1,3	9,6
L1-menovesi	°C	22,0	41,9	30,4	30,8	3,3
L1-sisälämpötila, lähtö	°C	0,0	0,0	0,0	0,0	0,0
L1-paluu	°C	21,4	35,0	27,3	27,7	2,3
Käyttövesi (LV)	°C	53,7	61,5	54,9	54,9	0,4
Käyttöveden ennakointi (LV)	°C	37,8	58,8	52,9	53,3	1,3
L2-menovesi	°C	21,9	62,4	37,9	39,3	8,5
mittaus 6	°C	0,0	0,0	0,0	0,0	0,0
L1 venttiilin asentotieto	%	0,0	0,0	0,0	0,0	0,0
L2 venttiilin asentotieto	%	0,0	0,0	0,0	0,0	0,0
LV venttiilin asentotieto	%	0,2	18,1	7,6	7,3	2,0
L2-sisälämpötila, tulo	°C	0,0	0,0	0,0	0,0	0,0
KL tulolämpötila	°C	64,3	107,0	77,7	76,8	8,7
KL paluulämpötila	°C	25,1	52,5	34,3	33,3	3,9
KL energia	kWh	68387,1	98509,6	85795,5	88281,9	10090,5
KL vesi	m <sup>3</sup>	0,0	1342130576,0	1253419814,9	1337837575,5	325872516,6
air quality monitor (0.5 µm)	count/ft <sup>3</sup>	0,0	6677,0	447,9	333,3	418,1
air quality monitor (2.5 µm)	count/ft <sup>3</sup>	0,0	1097,0	34,3	22,5	45,3

Kouluaineistoa analysoitiin CO<sub>2</sub> pitoisuuden osalta. Tätä varten aineistosta poistettiin summittaisesti sellaiset ajat jolloin koulussa ei ole oppilaita, eli vuorokaudenajoista on poistettu kaikki muut kuin kello 8-16 kohdalle osuneet mittaukset ja viikonloput. Kesälomatkin näyttäisivät olevan suuremmalti osin poistuneet, mutta datassa saattaa olla siltä osin vielä saastumia. Muita lomiamia ei ole otettu huomioon.

Muuttujia oli datassa 51, eli kaikki CO<sub>2</sub>-mittaukset.

Rivejä jäi dataan siivouksen jälkeen 1695 kappaletta.

Puuttuvia arvoja oli 48578 kappaletta, eli 56 % koko datasta.

Mittauspisteiden maksimiarvot CO<sub>2</sub> analysointiaineistossa:

<b>CO2.115</b>	<b>CO2.130</b>	<b>CO2.137</b>
998.58	711.91	1014.80
<b>CO2.146</b>	<b>X108.CO2</b>	<b>X221.CO2</b>
1098.70	933.04	970.59
<b>X305.CO2</b>	<b>X310.CO2</b>	<b>A112.CO2</b>
1052.80	1214.60	995.23
<b>A108.CO2</b>	<b>A209.CO2</b>	<b>A207.CO2</b>
845.47	1289.50	655.68
<b>A107.CO2</b>	<b>B109.CO2</b>	<b>B112.CO2</b>
1015.40	700.31	769.06
<b>B224.CO2</b>	<b>B306.CO2</b>	<b>B309.CO2</b>
931.72	975.70	883.36
<b>X102.CO2</b>	<b>X149.CO2</b>	<b>X222.CO2</b>
1160.40	855.66	1514.60
<b>X219.CO2</b>	<b>X216.CO2</b>	<b>X1212.CO2</b>
1201.40	956.25	780.98

<b>X1213.CO2</b>	<b>X2101.CO2</b>	<b>X1203.CO2</b>
719.14	1534.50	787.38

<b>X2102.CO2</b>	<b>X2119.CO2</b>	<b>X2105.CO2</b>
900.60	959.10	733.55

<b>X2043.CO2</b>	<b>X2027.CO2</b>	<b>X2049.CO2</b>
873.91	978.52	856.76

<b>X2052.CO2</b>	<b>X2040.CO2</b>	<b>X109.CO2</b>
819.26	899.69	964.01

<b>X202.CO2</b>	<b>X206.CO2</b>	<b>X306.CO2</b>
1054.60	1281.40	1351.40

<b>X508.CO2</b>	<b>luokka.1.CO2</b>	<b>luokka.2.CO2</b>
1445.40	569.41	578.01

<b>pienluokka.CO2</b>	<b>luokka.4.CO2</b>	<b>luokka.5.CO2</b>
581.25	683.71	674.92

<b>X2043..CO2</b>	<b>X2060..CO2</b>	<b>X2059.CO2</b>
695.20	647.85	764.34

<b>X2026.CO2</b>	<b>X2019.CO2</b>
1033.90	758.36



Markus

## 2.2 Menetelmät

Mittausaineistoja analysoitiin aikasarjakuvaajilla, periodogrammilla, autokorrelaatiolla, ristikorrelaatiolla, itseorganisoiduilla kartoilla ja sammonin kuvauksella, sekä k-means klusteroinilla. Aikasarjakuvaajat piirrettiin mielenkiintoisimmista muuttujista ajan suhteen, jolloin puuttuvat arvot näkyvät kuvaajassa. Aikasarjakuvaajista näkee varsin hyvin esimerkiksi kausivaihtelut. Periodogrammilla voidaan tarkastella aineiston periodisia komponentteja, esimerkiksi vuorokausirytmien. Autokorrelaatiolla voidaan kuvata aikasarjan havaintopisteiden välistä riippuvuutta. Aikasarjan autokorrelaatiosta voidaan tarkastella johtuuko havaintopisteiden välinen riippuvuus jostain havainnosta esimerkiksi vuorokausivaihtelusta, eli ei satunnaisuudesta. Ristikorrelaatiolla voidaan tarkastella kahden muuttujan välistä riippuvuutta ajan suhteen siten, että toisen muuttujan datarivejä siirretään ajan esimerkiksi tunnin verran. Sammonin kuvaus (Sammon's mapping) on algoritmi, jolla n-ulotteinen datajoukko voidaan kuvata k-ulotteiseksi (n>k) siten, että datapisteiden väliset etäisyydet ovat mahdollisimman lähellä alkuperäisiä etäisyyksiä. Tässä harjoitustyössä Sammonin kuvausta käytettiin tarkasteltaessa klusteroidun itseorganisoidun kartan klustereiden samankaltaisuutta. Itseorganisoiduilla kartoilla voi analysoida moniulotteista dataa varsin helposti, koska menetelmä pyrkii esittämään samankaltaiset datajoukot kaksiulotteisella pinnalla siten, että kartan hilapisteisiin ...

Itseorganisoiduva kartta pyrkii esittämään samankaltaiset datarivit samassa hilassa. Esimerkiksi jos kartta on 10\*10 hilainen, kartalla pystyy esittämään 100 aineiston tyypillisimmät tapaukset eli yhteen hilaan voi kuulua useampi aineiston rivi. Itseorganisoiduvalla kartalla analysoidessa usein data esikäsitellään siten, että datan skaalat ovat samassa skaalassa. Tähän voi käyttää monia eri menetelmiä, kuten min-max skaalausta tai varianssiskaalausta. Itseorganisoiduva kartta muodostetaan ennalta määrätyn hilarakenteen muotoon. Usein käytetään neliömäistä karttaa, jossa voi olla esimerkiksi 10\*10 hilaa tai jos tarkastellaan tarkemmalla tasolla aineistoa, niin käytetään suurempaa karttaa. Itseorganisoiduvasta kartasta voi tehdä klusterianalyysin, tällöin kartan komponenttitaso klusteroidaan sopivalla klusterointimenetelmällä. Klusteroinnissa samankaltaiset hilat muodostavat klusterin. Klustereiden ominaisuuksia voidaan esittää esimerkiksi histogrammeina, jossa usein klusterin ominaisuuksista esitetään tapauksien lukumäärä ja muuttujan arvot.

### 2.2.1 Varianssiskaalaus

Varianssiskaalaus taas skaalaa aineiston jokaisen muuttujan keskiarvoksi nollan ja varianssiksi 1. Varianssiskaalauksen avulla aineistosta saadaan esille myös mahdollinen normaalivaihtelu. Varianssiskaalaus voidaan esittää seuraavasti:

$$c'_i = \frac{c_i - u_c}{\sigma_c} \quad (1)$$

missä,

$u_c$  on vektorin  $\mathbf{c}$  arvojen keskiarvo ja

$\sigma_c$  vektorin  $\mathbf{c}$  arvojen keskihajonta

### 2.2.2 k-means klusterointi

K-means (McQueen, 1967) on yksi tunnetuimmista klusterointimenetelmistä ja se kuuluu optimointimenetelmien luokkaan. Optimointimenetelmät vaativat syötteenään klustereiden lukumäärän. Klustereiden lukumäärän tulee olla tiedossa tai se on pystyttävä arvioimaan etukäteen. Minimoitava tavoitefunktio (yleensä keskineliövirhe, MSE) voidaan esittää seuraavasti:

$$MSE = \frac{1}{N} \sum_{i=1}^N d(x_i, c_{pi})^2 \quad (2)$$

missä,

$N$  on aineiston alkioden lukumäärä,

$x_i$  aineiston  $i$ :s alkio,

$c_{pi}$  klusterin keskipiste, johon alkio  $x_i$  kuuluu ja

$d(x_i, c_{pi})$  on aineiston  $i$ :nnen alkion ja klusterinsa keskipisteen välinen euklidinen etäisyys.

Menetelmässä käytetään kahta optimaalisuuskriteeriä, joita vuorottelemalla optimoidaan lokaalisti kaavan 2 tavoitefunktio.

### 2.2.3 Davies-Bouldin indeksi

Davies-Bouldin –indeksi (Davies & Bouldin 1979) kuvaa klustereiden samankaltaisuutta, joten sen mukaan parhain mahdollinen klusterointi on se, jolla indeksi saa mahdollisimman pienen arvon. Davies-Bouldin –indeksi voidaan määritellä seuraavasti:

$$DB_{nc} = \frac{1}{n_c} \sum_{i=1}^{n_c} R_i \quad (3)$$

missä,

$$R_i = \max_{j=1, \dots, n_c, i \neq j} R_{ij}, i = 1, \dots, n_c \text{ ja}$$

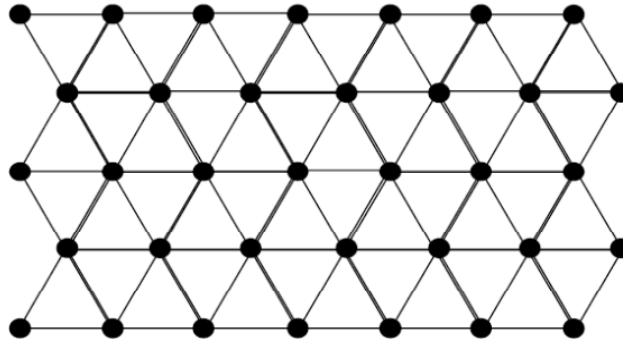
$n_c$  on klustereiden lukumäärä

### 2.2.4 Itseorganisoituva kartta

Itseorganisoituva kartta eli SOM (engl. Self-Organizing Map) on Teuvo Kohosen kehittämä neuraalilaskennan menetelmä suurten moniulotteisten datajoukkojen visualisointia ja tulkintaa varten. SOM pyrkii säilyttämään käsiteltävän moniulotteisen datajoukon keskeiset riippuvuussuhteet myös kaksiulotteisessa visualisoinnissa. Visualisointi tapahtuu yleensä kaksiulotteisesti. Kartta muodostetaan muuttujien välisien yhteyksien perusteella. Samankaltaiset datarivit hakeutuvat opetuksessa samoihin neuroneihin. (Kohonen 1995)

Itseorganisoituva kartta on yleensä yksi- tai kaksiulotteisen hilan muotoinen neuroverkko, jossa jokainen karttayksikkö edustaa yhtä prosessointiyksikköä eli neuronua. Kuvassa 1 on esimerkki

SOM- kartan rakenteesta. Kartalle jakautuneet neuronit ovat vuorovaikutuksessa lähistöllä sijaitsevien neuronien kanssa. (Kohonen 1995)



Kuva 1. SOM verkon rakenne

Kartan opetus perustuu ohjaamattomaan opetukseen, mikä tarkoittaa että verkolle ei anneta tavoitearvoja, vaan ne syntyvät oppimisprosessin tuloksena. Oppimisprosessin aikana kartalla olevat neuronit tulevat herkiksi erityyppisille syötevektoreille tai syötevektoreiden muodostamille ryhmille. (Kohonen 1995)

SOM- algoritmista on olemassa useita variaatioita. Perusalgoritmi on seuraava (Kohonen 1995): Algoritmin jokaisella opetusaskeleella  $t = 1, 2, 3, \dots, T$ . Näytevaruudesta valitaan satunnaisesti datavektoreita. Datavektorille haetaan syötevaruudessa lähin mallivektori  $m_c$ , jolle:

$$\|x - m_c\| = \min_i \{\|x - m_i\|\} \quad (4)$$

missä,

$x$  on datavektori

$m_c$  on mallivektori

Mallivektoria  $m_c$  vastaavaa karttayksikköä päivitetään seuraavasti:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \quad (5)$$

missä  $h_{ci}$  on naapurustokernel (vähenevä ajan suhteen, jotta algoritmi konvergoituisi)

$$h_{ci} = \alpha(t)h(c, i; t) \quad (6)$$

missä,

$\alpha$  on oppimisnopeuskerroin,

$h$  on naapurustofunktio

Oppimisnopeuskerroin  $\alpha$  on yleensä lineaarisesti vähenevä tai funktion

$$\alpha(t) = \frac{A}{t+B} \quad (7)$$

mukaisesti ajan suhteen vähenevä. Lisäksi  $0 < \alpha < 1$ .

Valitaan naapurustofunktio  $h$  (esim. kuplanaapurusto):

$$h(c, i; t) = \begin{cases} 1, & \|n_c - n_i\| \leq \sigma(t) \\ 0, & \text{muulloin} \end{cases} \quad (8)$$

Opetusjoukko käydään läpi useita kertoja, jotta saadaan riittävä määrä opetusaskelia.

Kartan erottelukykyä mitataan kvantisointivirheellä  $e_q$ :

$$e_q = \frac{1}{N} \sum_{i=1}^N \|x_i - m_c\| \quad (9)$$

Kvantisointivirheen avulla voidaan tarkastella kuinka opetusparametrien muuttaminen vaikuttaa kartan erottelukykyyn (Kohonen 1995).

### 2.2.5 Sammonin kuvaus

Sammonin kuvaus on algoritmi, jolla  $n$ -ulotteinen datajoukko voidaan kuvata  $k$ -ulotteiseksi ( $n > k$ ) siten, että datapisteiden väliset etäisyydet ovat mahdollisimman lähellä alkuperäisiä etäisyyksiä. (Sammon, 1969) Sammonin kuvausta käytetään tutkiessa vektoreiden keskinäistä sijaintia data-avaruudessa. Menetelmää voidaan käyttää esimerkiksi arvioitaessa asuntilojen samankaltaisuutta.

## 3 ESIPROSESSOINTI

Tässä kappaleessa esitellään alkuperäisen mittausaineiston esikäsittelyä. Esikäsittelyllä yleensä poistetaan poikkeavat arvot ja muodostetaan uusia muuttujia.

Markus

### 3.1 Aineiston siivous

Alkuperäisessä mittausaineistossa oli poikkeavia arvoja johtuen mittalaitteiden tiedonsiirtovirheistä. Mittalaitteet asettavat tiedonsiirtovirheen tapahtuessa mitatun suureen arvoksi -999. Tiedonsiirrossa voi tapahtua myös duplikaattirivien muodostuminen, jos esimerkiksi mittalaite lähettää saman mittaustuloksen tietokantaa. Nämä poikkeavat arvot ja duplikaattirivit poistettiin aineistosta siten, että kaikki arvot jotka olivat mittalaitteen valmistajan ilmoittamien rajojen ylä- tai alapuolella poistettiin. Duplikaattirivit poistettiin siten, että etsittiin samalla aikaleimalla olevat rivit ja duplikaattiriveistä jätettiin aineistoon vain yksi uniikkirivi. Tämän jälkeen

aineisto muodostettiin jatkuvaksi ajan suhteen, koska välillä puuttui pitkiäkin aikoja jolloin mittalaitteet eivät olleet lähettäneet mittaustuloksia. Aineisto muodostettiin jatkuvaksi ajan suhteen siten, että ensimmäisen ja viimeisen aikaleimojen välinen aika oli jatkuvana yhden sekunnin resoluutiolla. Tässä tapauksessa kun mittausväli oli joko 10 sekuntia tai yksi minuutti niin näiden mittausväliin jäävien rivien muuttujien arvoksi asetettiin arvo 'NaN'. Näin muodostetusta aineistosta laskettiin keskiarvot tunneittain eli lopullisen esikäsitellyn aineiston aikaresoluutio oli yksi tunti.

Markus

### 3.2 Aineistojen yhdistäminen

Tuntikeskiarvoistettu aineisto yhdistettiin, koska eri mittalaitteet ja kohteet olivat tallennettu tietokantaan omiksi tauluiksi ja aineiston analysoinnin kannalta oli helpompaa, kun aineisto oli yhtenäisessä taulukossa. Aineiston yhdistäminen tehtiin siten, että jokaisen tuntikeskiarvoistettujen mittaustulosten aikaleima sarakkeelta etsittiin samaa aikaleimaa vastaavat rivit. Koska eri mittalaitteet olivat asennettu eri aikoina eli samalta ajan hetkeltä ei ollut mittaustuloksia niin aineiston taulukko alustettiin 'NaN' arvoilla ensin siten, että rivien lukumäärä vastasi eri tauluissa olevien mittaustulosten lukumäärää sitä missä oli eniten rivejä ja sarakkeiden lukumäärä vastasi eri mittaustulosten muuttujien lukumäärien summaa. Koulu ja omakotitalon mittaukset jätettiin eri taulukoihin.

Koulujen mittaustulokset

1 Koulu 1 = 2964529 riviä, 16 muuttujaa

2 Koulu 2 = 515053 riviä, 16 muuttujaa

3 Koulu 3 = 697928 riviä, 17 muuttujaa

4 Koulu 4 = 457200 riviä, 16 muuttujaa

5 Koulu 5 = 450176 riviä, 17 muuttujaa

6 Koulu 6 = 335047 riviä, 24 muuttujaa

7 Koulu 7 = 240873 riviä, 17 muuttujaa

8 Koulu 8 = 1215906 riviä, 16 muuttujaa

9 Koulu 9 = 352446 riviä, 17 muuttujaa

10 Koulu 10 = 145115 riviä, 17 muuttujaa

Lopullisessa kouluaineistossa oli 6623 riviä ja 173 muuttujaa + aikaleima ja eri koulujen mittauksien loppuissa numerot 1...10 eli yhteensä 183 saraketta.

Omakotitalon mittaustulokset eri mittalaitteilta (numerot 11, 74 jne ovat mittalaitteiden ID numeroita)

11 = 3210423 riviä, 9 muuttujaa

74 = 2074900 riviä, 14 muuttujaa

75 = 1997214 riviä, 2 muuttujaa

85 = 2190649 riviä, 16 muuttujaa

93 = 5480 riviä, 2 muuttujaa

Lopullisessa omakotitaloaineistossa oli 6623 riviä ja 43 muuttujaa + aikaleima eli 44 saraketta.

## 4 TULOKSET

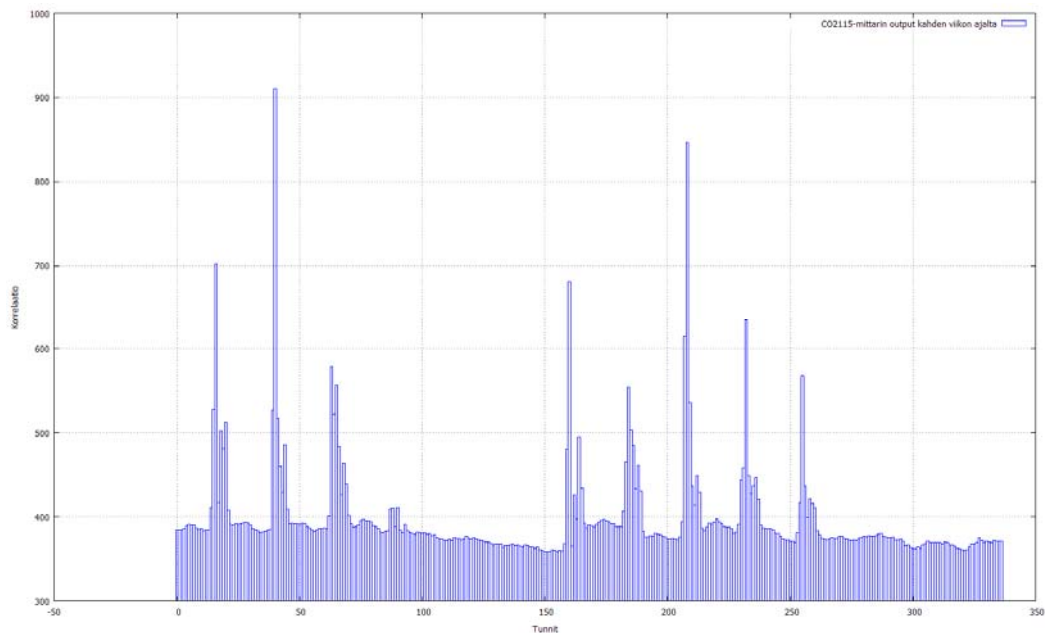
### 4.1 Aineiston kuvaajat

Koulujen ja omakotitalon mittausaineistoista piirrettiin aikasarjakuvaajia, joista näkee eri muuttujien sisältöä.

Markus, Mikko, Kimmo

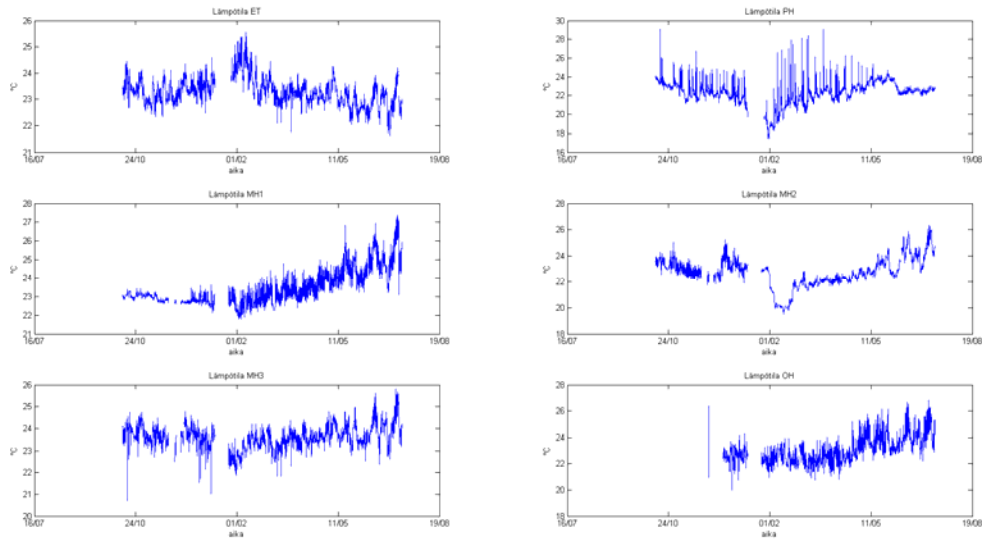
#### 4.1.1 Aikasarjakuvaajat

Kuvassa 2 on esitetty yhden koulun yhden luokan CO<sub>2</sub> pitoisuus kahden viikon ajalta. Kuvasta näkee koulupäivien ja viikonloppujen erot. Koulupäivät näkyvät selvästi piikkeinä. Ensimmäiset kolme piikkiä ovat loppuviikon koulupäivät keskiviikko, torstai ja perjantai. Sen jälkeen on viikonloppu ja sitten viikonlopun jälkeen oleva ensimmäinen piikki on maanantai. Kuvasta näkee, että sisäilmanlaatu CO<sub>2</sub> osalta pysyy hyvin sisäilmastoluokan S1 rajoissa eli hyvä ilmanlaatu, mutta joinakin päivinä CO<sub>2</sub> pitoisuus nousee yli sisäilmastoluokan S1 rajan yli luokkaan S2 eli tyydyttävään sisäilmalaatuun.



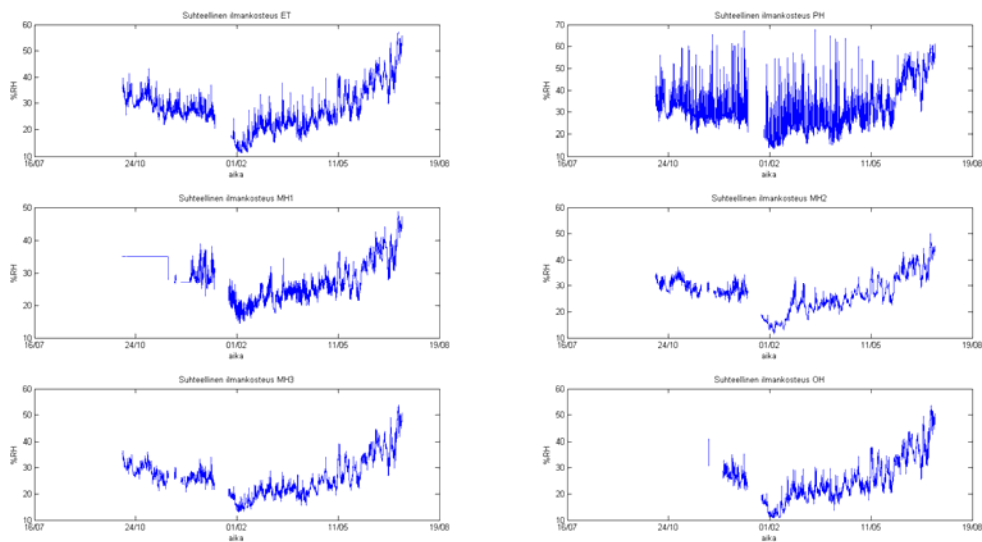
Kuva 2. Yhden koulun yhden luokan CO<sub>2</sub> pitoisuus noin kahden viikon ajalta

Kuvassa 3 on esitetty omakotitalon lämpötilat koko aineistosta. Kuvasta näkee eri vuodenaikojen vaikutukset eri huoneiden lämpötiloihin. Makuuhuoneiden ja olohuoneen lämpötilojen kasvu näyttää johtuvan kevät- ja kesäkaudesta. Pesuhuoneen lämpötilassa näyttää olevan suuria vaihteluita. Pesuhuone on saunan yhteydessä, joten lämpötilan nousu johtuu saunan ja suihkun käytöstä. Pesuhuone sijaitsee kellarikerroksessa, joka on kiviseinäinen, joten talviaikaan tapahtuva lämpötilan lasku johtuu siitä, että pakkaspäivinä kellari viilenee huomattavasti. Lämpötilat näyttäisivät olevan hieman liian korkeat suositellusta +21 °C lämpötilasta.



Kuva 3. Aikasarjakuvaajat eri huoneiden lämpötiloista koko aineistosta

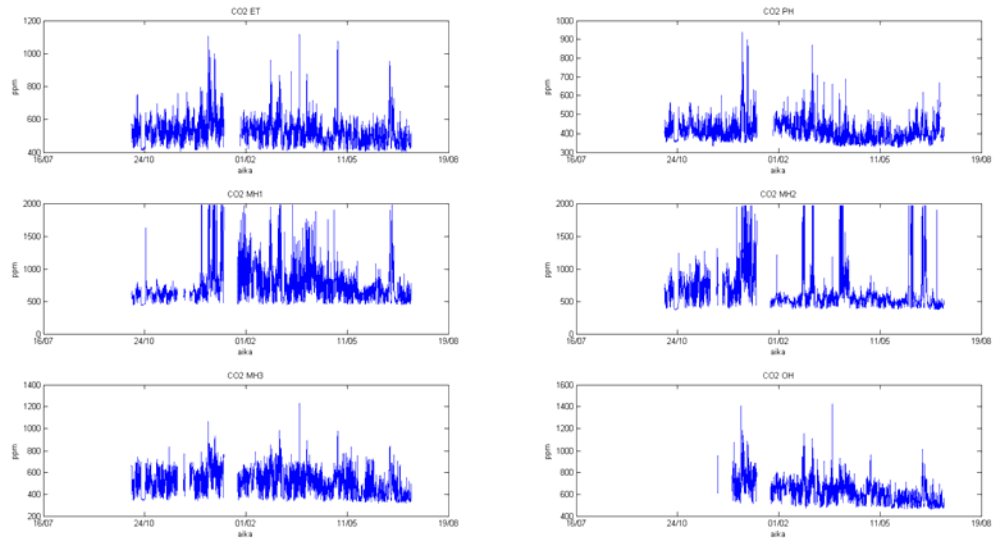
Kuvassa 4 on esitetty omakotitalon ilmankosteudet koko aineistosta. Kuvasta näkee eri vuodenaikojen vaikutukset eri huoneiden ilmankosteuksiin. Eri huoneissa näyttäisi olevan varsin samanlaiset ilmankosteudet, jotka johtuvat eri vuoden ajoista. Kesällä on kosteampaa ilmaa kuin talvella. Pesuhuoneen ilmankosteudessa näyttää olevan suuria vaihteluita, jotka johtuvat saunan ja suihkun käytöstä. Ilmankosteudet näyttäisivät olevan suositetuissa rajoissa eli 20- 60 %, mutta talvikaudella ilmankosteus on liian matala eli alle 20 %.



Kuva 4. Aikasarjakuvaajat eri huoneiden suhteellisista ilmankosteuksista koko aineistosta

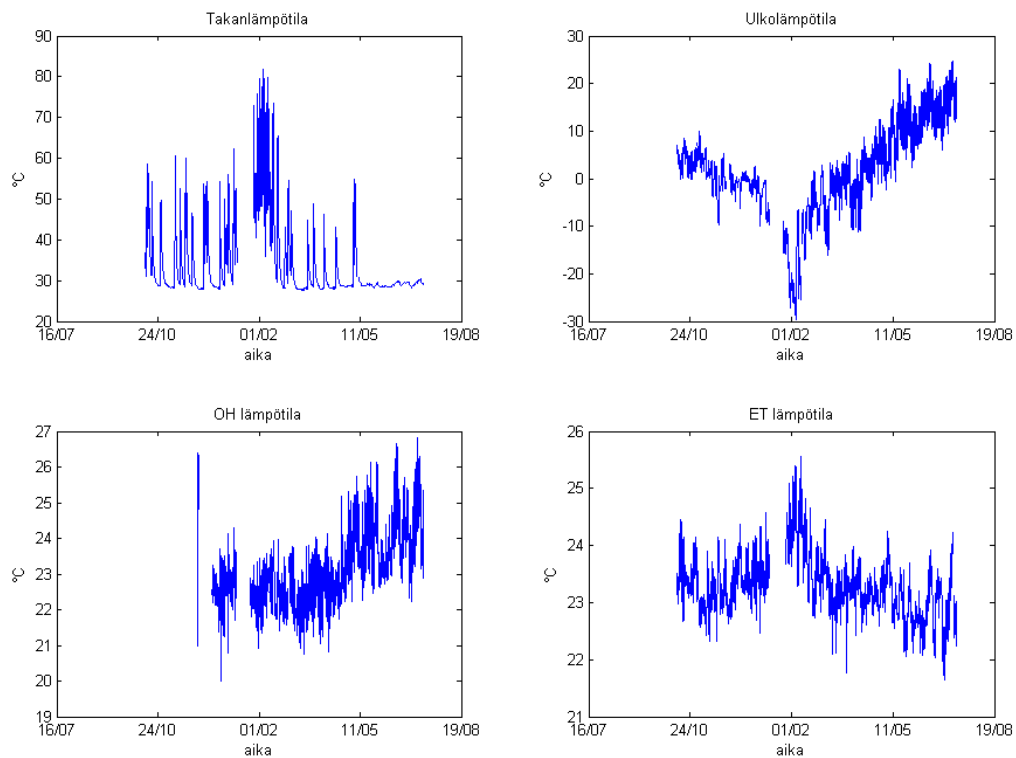
Kuvassa 5 on esitetty omakotitalon CO<sub>2</sub> pitoisuudet koko aineistosta.





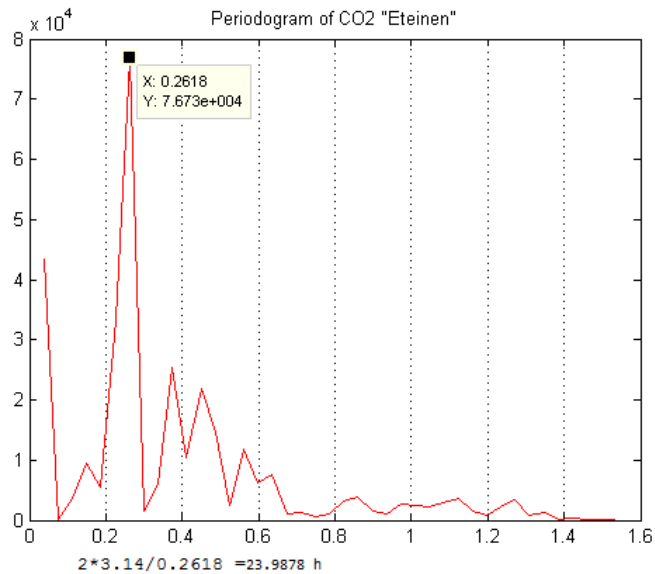
Kuva 5. Aikasarjakuvaajat eri huoneiden CO<sub>2</sub> pitoisuuksista koko aineistosta

Kuvassa 6 on esitetty kuvaajat omakotitalon takanlämpöttilasta, ulkolämpöttilasta, sekä olohuoneen ja eteisen lämpöttiloista koko aineistosta. Kuvasta nähdään, että takkaa lämmitetään talvikaudella varsin useasti kun ulkona on kylmä. Takka on olohuoneen ja eteisen välisessä seinässä, joten takan lämmitys olohuoneessa nostaa myös eteisen lämpöttilaa.



Kuva 6. Aikasarjakuvaajat takanlämpöttilasta, ulkolämpöttilasta, sekä olohuoneen ja eteisen lämpöttiloista koko aineistosta

Kuvassa 7 on esitetty periodogrammi omakotitalon eteisen CO<sub>2</sub> pitoisuudesta. Periodogrammista näkee, että CO<sub>2</sub> periodinen komponentti on 24 h välein toistuva. Periodogrammissa on myös muita periodisia komponentteja. Eri periodiset komponentit voivat johtua siitä, että eteisessä käydään varsin säännöllisesti 24 h välein, mutta myös muina aikoina.

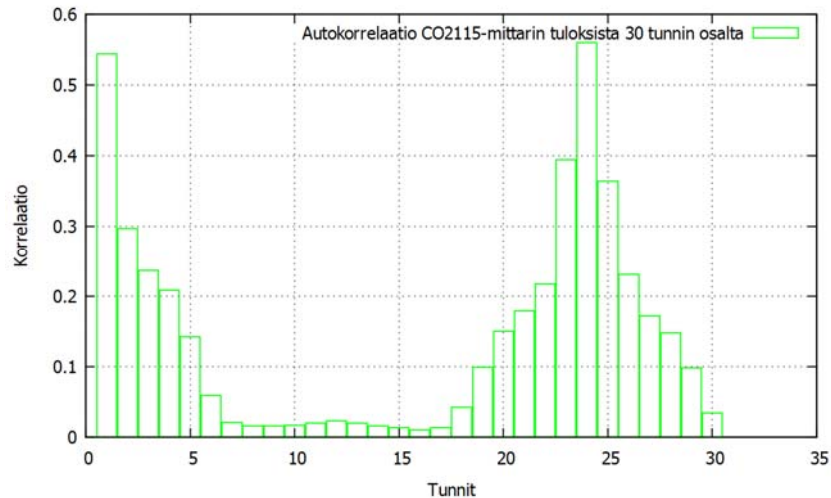


Kuva 7. Periodogrammi eteisen CO<sub>2</sub> pitoisuudesta

Markus, Mikko

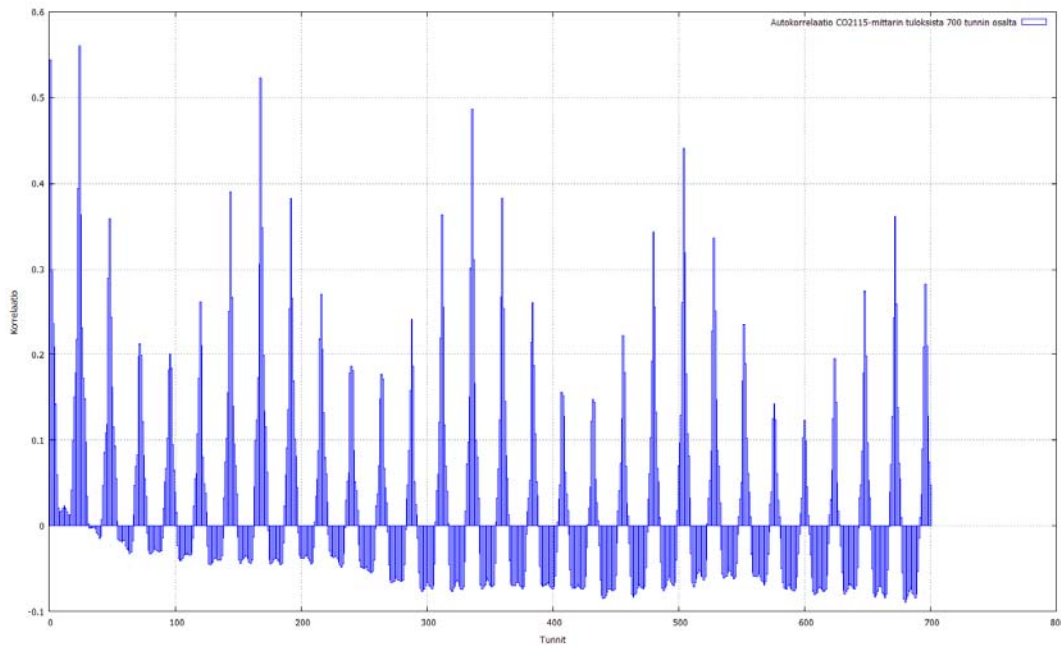
#### 4.1.2 Autokorrelaatiot

Kuvassa 8 on esitetty yhden koulun yhden luokan CO<sub>2</sub> pitoisuuden autokorrelaatio 30 tunnilta. Autokorrelaatiokuvaajasta nähdään, miten luokassa CO<sub>2</sub> pitoisuus korreloi hyvin 24 h välein. Kuvasta nähdään hyvin, miten koulupäivän jälkeen luokassa ei ole henkilöitä, koska korrelaatio laskee jyrkästi koulupäivän jälkeen.



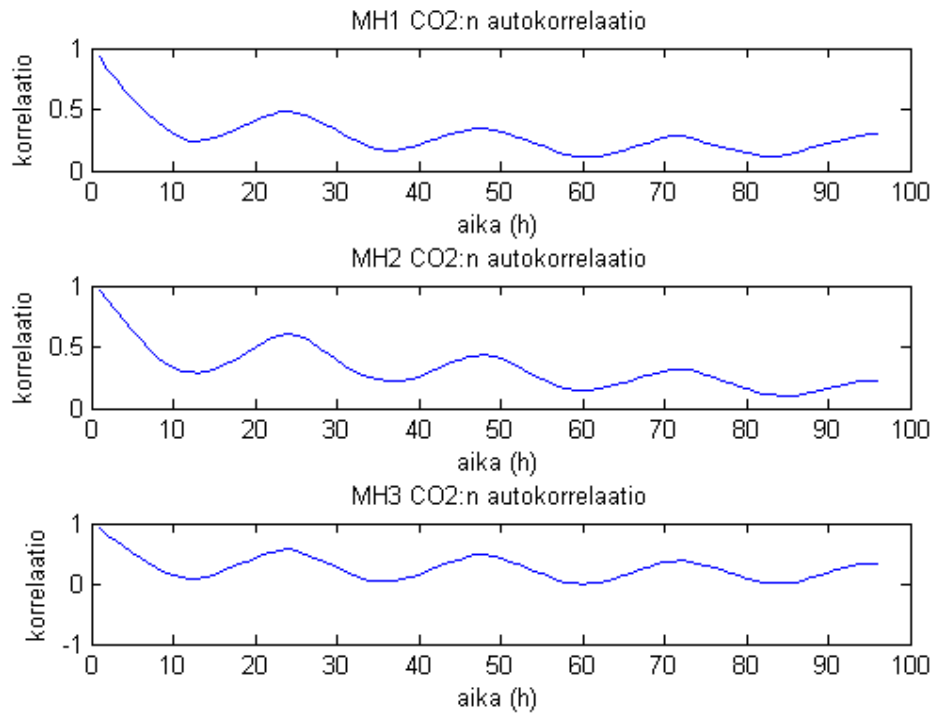
Kuva 8. Yhden koulun yhden luokan CO<sub>2</sub> pitoisuuden autokorrelaatio 30 tunnilta

Kuvassa 9 on esitetty yhden koulun yhden luokan CO<sub>2</sub> pitoisuuden autokorrelaatio kuukauden ajalta. Autokorrelaatiokuvaajasta nähdään, miten luokassa CO<sub>2</sub> pitoisuus korreloi hyvin 24 h välein, mutta myös viikkojakso näkyy hyvin.



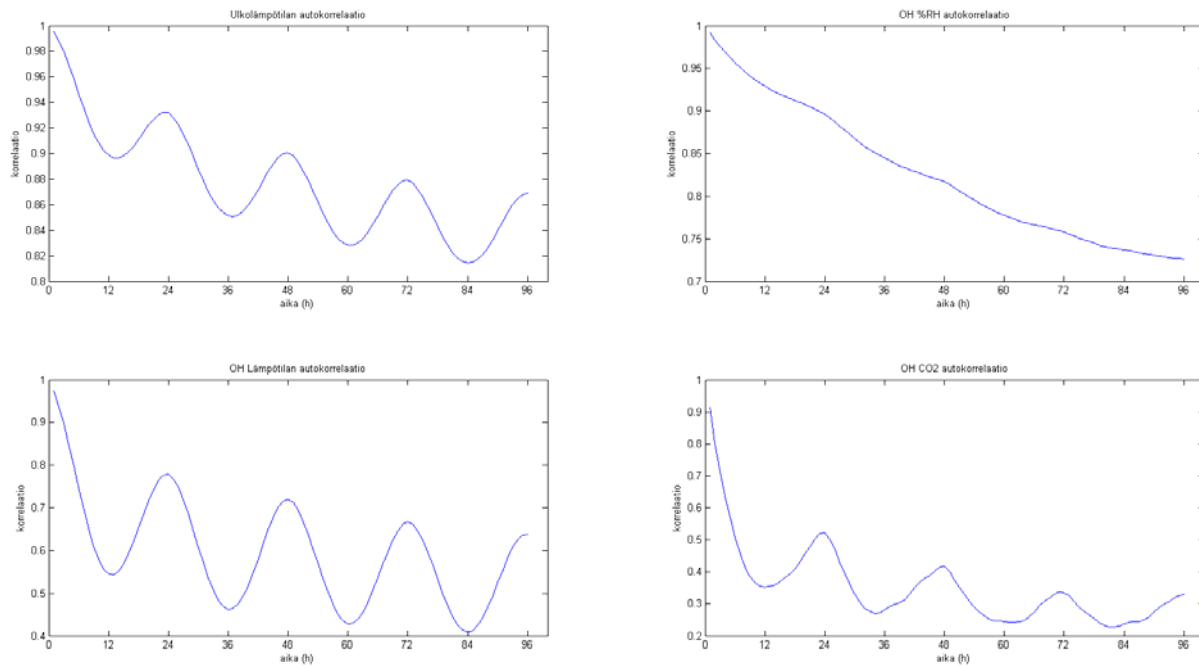
Kuva 9. Yhden koulun yhden luokan CO<sub>2</sub> pitoisuuden autokorrelaatio kuukauden ajalta

Kuvassa 10 on esitetty omakotitalon makuuhuoneiden CO<sub>2</sub> pitoisuuden autokorrelaatio neljältä vuorokaudelta. Autokorrelaatiokuvaajasta nähdään, miten makuuhuoneiden CO<sub>2</sub> pitoisuus korreloi hyvin 24 h välein. Vertaamalla koulujen CO<sub>2</sub> pitoisuuden autokorrelaatioita nähdään, että omakotitalossa on myös 24 h autokorrelaatio, mutta korrelaatio ei laske niin jyrkästi kuin kouluissa. Tämä voi johtua siitä, että omakotitalossa henkilöt ovat pidemmän aikaa sisällä kuin oppilaat luokissa. Vertaamalla kuvaa 7 ja kuvaa 10 huomataan, että periodogrammista näkee myös muut periodiset komponentit kuin autokorrelaatiosta.



Kuva 10. Omakotitalon makuuhuoneiden CO<sub>2</sub> pitoisuuden autokorrelaatiot neljältä vuorokaudelta

Kuvassa 11 on esitetty omakotitalon ulkolämpötilan ja olohuoneen ilmankosteuden, lämpötilan ja CO<sub>2</sub> pitoisuuden autokorrelaatiot neljältä vuorokaudelta. Autokorrelaatiokuvaajasta nähdään, miten ulkolämpötilassa, olohuoneen lämpötilassa ja CO<sub>2</sub> pitoisuudessa on selvä 24 h jakso, mutta ilmankosteudessa ei ole. Ilmankosteus ei näytä vaihtelevan suuresti neljässä vuorokaudessa, mikä on kuitenkin hyvä asia ajatellen asumisterveyttä. Ilmankosteus on helpommin hallittavissa kun se ei vaihtele nopeaan tahtiin.

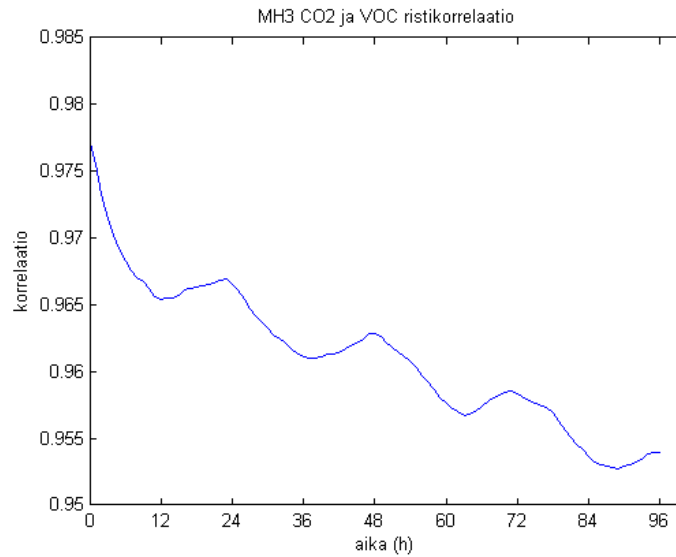


Kuva 11. Omakotitalon ulkolämpötilan ja olohuoneen ilmankosteuden, lämpötilan ja CO2 pitoisuuden autokorrelaatiot neljältä vuorokaudelta

Markus

#### 4.1.3 Ristikorrelaatiot

Kuvassa 12 on esitetty omakotitalon makuuhuoneen CO2 ja TVOC pitoisuuksien ristikorrelaatio neljältä vuorokaudelta. Kuvaajasta nähdään, että CO2 ja TVOC ristikorrelaatio näyttäisi kasvavan vain hieman 24 h jaksoissa, mikä voi johtua henkilöistä.

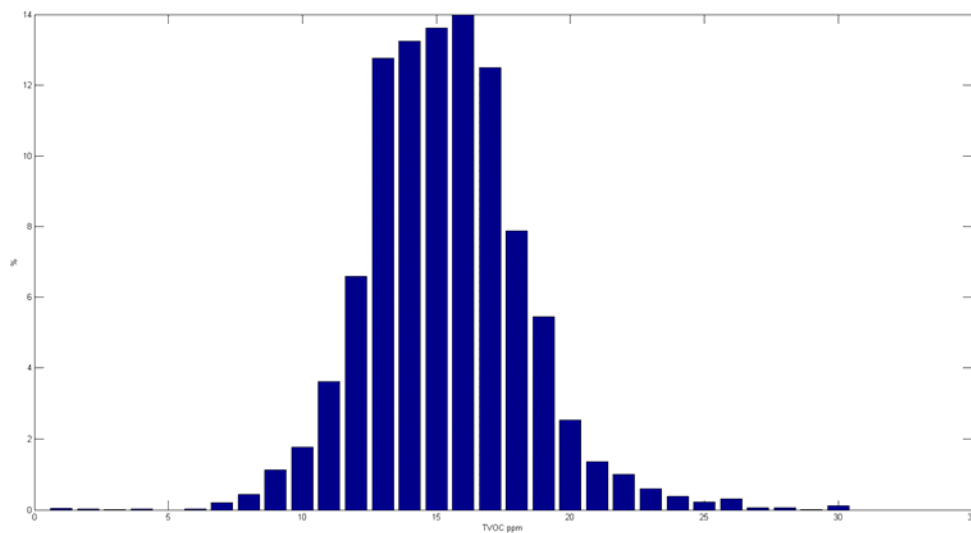


Kuva 12. Omakotitalon makuuhuoneen CO2 ja TVOC pitoisuuksien ristikorrelaatio neljältä vuorokaudelta

Markus, Mikko, Kimmo

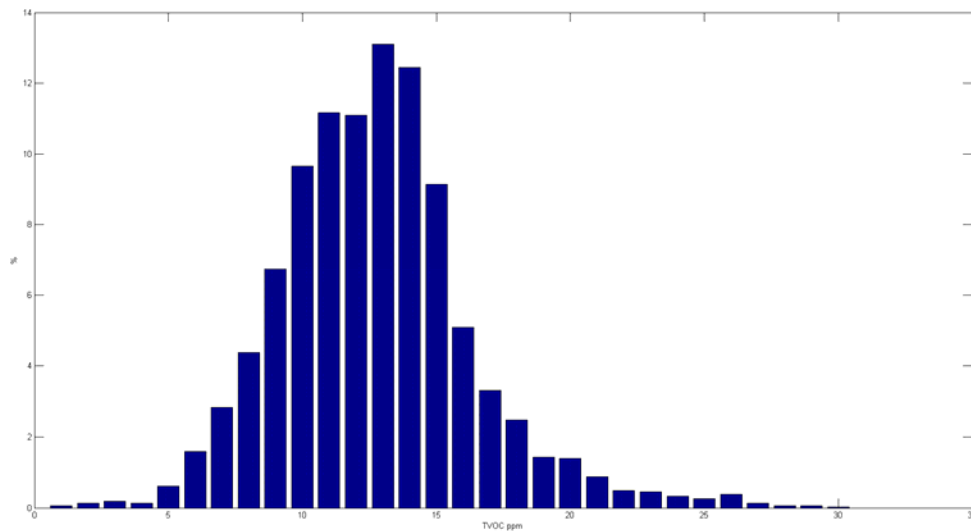
## 4.2 Hiilidioksidi ja TVOC pitoisuuksien analysointi

Hiilidioksidi ja TVOC pitoisuuksista lasketettiin prosenttiosuudet, joiden tuloksia esitellään histogrammeilla. Kuvassa 13 on esitetty omakotitalon TVOC pitoisuuden histogrammi koko aineistosta. Histogrammista nähdään, että pieniä ja suuria pitoisuuksia ei ole ollut paljoa, vaan suurin osa pitoisuuksista keskiarvon eli 15 ppm lähellä. Kuvassa y akselilla on prosenttiosuus ja x akselilla TVOC pitoisuus.



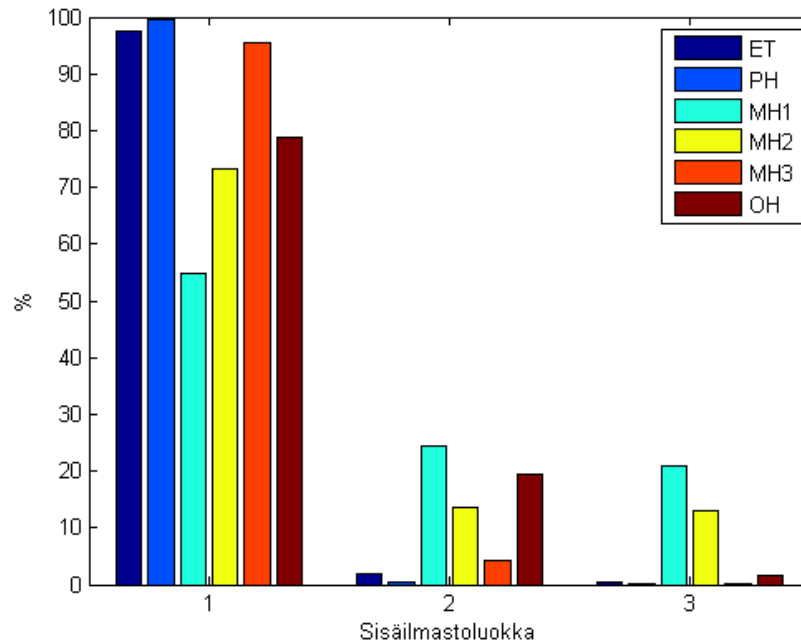
Kuva 13. Histogrammi omakotitalon TVOC pitoisuudesta

Kuvassa 14 on esitetty omakotitalon TVOC pitoisuuden histogrammi. Kuvassa y akselilla on prosenttiosuus ja x akselilla TVOC pitoisuus. Histogrammi piirrettiin aineiston niistä riveistä, joista etsittiin samassa huoneessa tehdystä CO<sub>2</sub> pitoisuusmittauksista ne rivit joiden CO<sub>2</sub> pitoisuuden arvo oli yli 450 ppm. Eli yritettiin arvioida sitä, milloin talossa on henkilöitä paikalla. Histogrammista nähdään, että suuria pitoisuuksia ei ole ollut paljoa, mutta tulos poikkeaa kuvasta 13 siten, että suurin osa pitoisuuksista on arvon 12 ppm lähellä ja pienempiä arvoja on hieman enemmän. Tämä voi johtua siitä, että TVOC pitoisuusmittari mittaa kokonaispitoisuuden ja ei pysty erottelemaan eri haihtuvia orgaanisia yhdisteitä. Henkilöiden ollessa kotona voi esimerkiksi vaatteisiin jääneet pesuainejäämät nostaa TVOC pitoisuutta tai TVOC pitoisuus kasvaa siivouksen aikana.



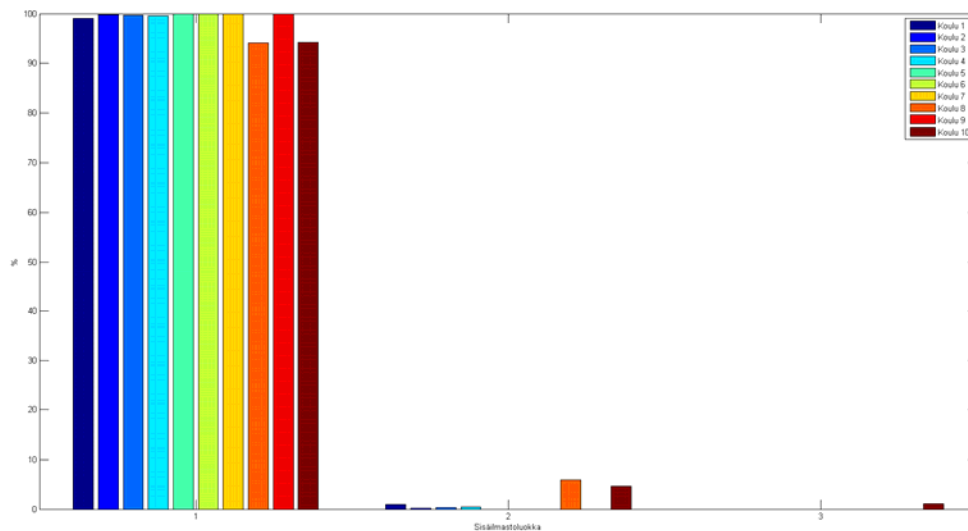
Kuva 14. Histogrammi omakotitalon TVOC pitoisuudesta

Kuvassa 15 on esitetty omakotitalon CO<sub>2</sub> pitoisuuksien histogrammi sisäilmastoluokittain. Kuvassa y akselilla on prosenttiosuus ja x akselilla sisäilmastoluokka. Histogrammista nähdään, että suurimmaksi osaksi sisäilmalaatu on S1 sisäilmastoluokituksen mukainen eli hyvä ilmanlaatu, mutta makuuhuoneen MH1, MH2 ja MH3 (10 -20 % ajasta) S2 sisäilmastoluokituksessa eli tyydyttävä ilmanlaatu. Makuuhuoneissa MH1 ja MH2 ilmanlaatu on ollut myös sisäilmastoluokassa S3 eli huono ilmanlaatu (10 - 20% ajasta). Eteisessä ja pesuhuoneessa on ollut melkein kokoajan hyvä ilmanlaatu.



Kuva 15. Omakotitalon CO<sub>2</sub> pitoisuuksien prosenttiosuudet sisäilmaluokituksilla S1, S2 ja S3

Kuvassa 16 on esitetty koulujen CO<sub>2</sub> pitoisuuksien histogrammi sisäilmastoluokittain. Kuvassa y akselilla on prosenttiosuus ja x akselilla sisäilmastoluokka. Histogrammista nähdään, että suurimmaksi osaksi sisäilmalaatu on S1 sisäilmastoluokituksen mukainen eli hyvä ilmanlaatu, mutta joissakin kouluissa on ollut tyydyttävä ilmanlaatu, kuitenkin alla 10 % ajasta.

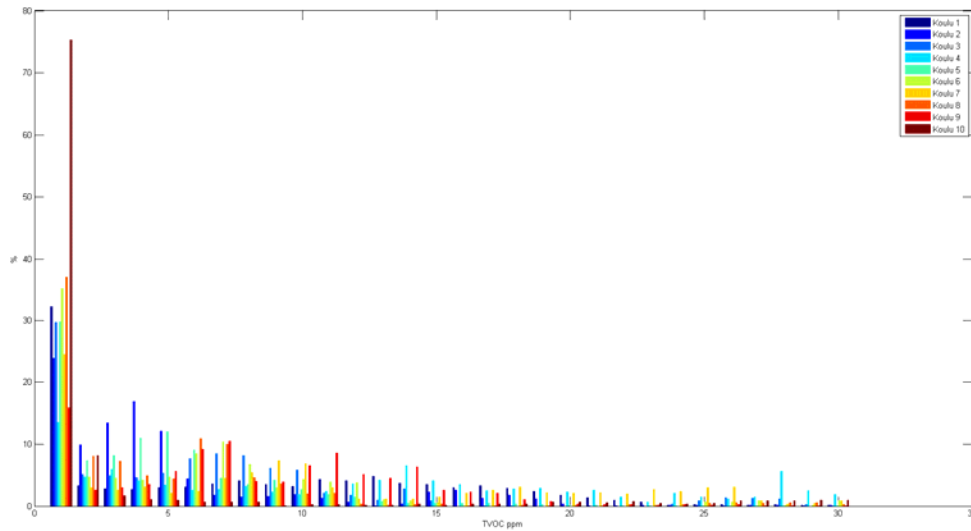


Kuva 16. Koulujen CO<sub>2</sub> pitoisuuksien prosenttiosuudet sisäilmaluokituksilla S1, S2 ja S3

Kuvassa 17 on esitetty koulujen TVOC pitoisuuksien histogrammi koko aineistosta. Kuvassa y akselilla on prosenttiosuus ja x akselilla TVOC pitoisuus. Histogrammista nähdään, että TVOC

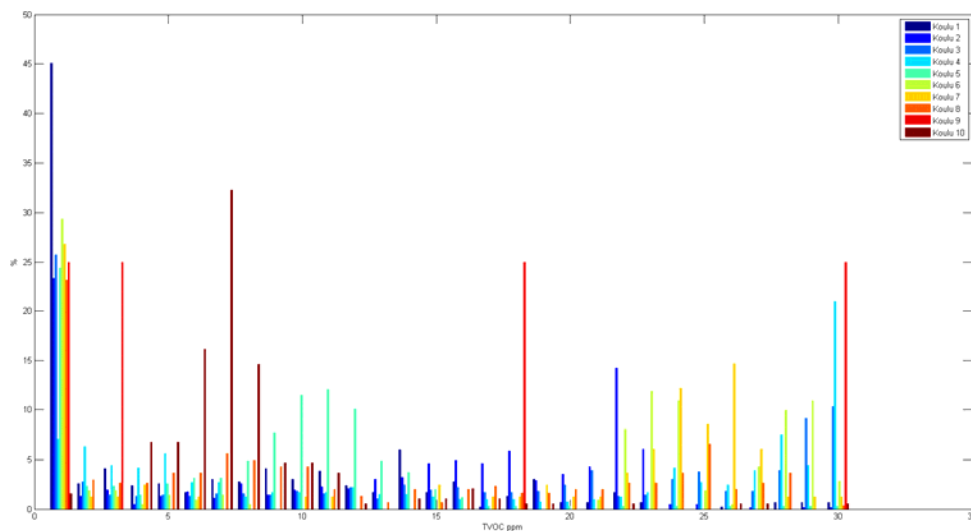


pitoisuudet ovat kouluissa pienet. Joissakin kouluissa ei ole juuri lainkaan haihtuvia orgaanisia yhdisteitä.

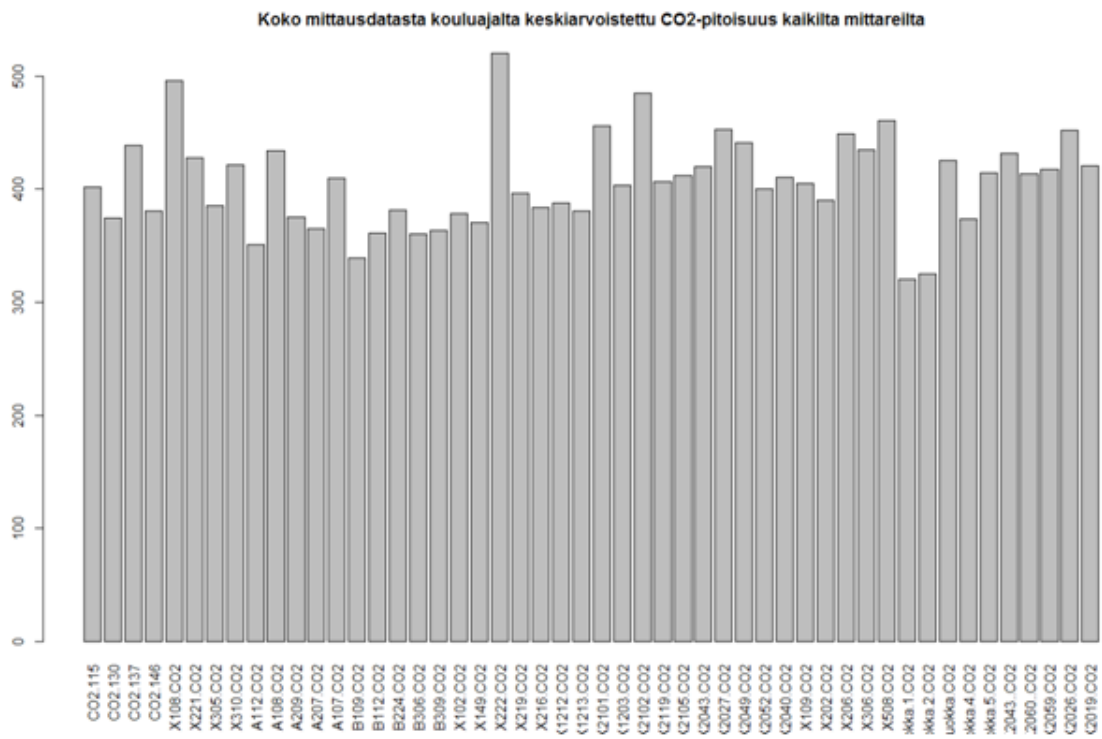


Kuva 17. Koulujen TVOC pitoisuuksien prosenttiosuudet

Kuvassa 18 on esitetty koulujen TVOC pitoisuuksien histogrammi. Kuvassa y akselilla on prosenttiosuus ja x akselilla TVOC pitoisuus. Histogrammi piirrettiin aineiston niistä riveistä, joista etsittiin samassa huoneessa tehdystä CO<sub>2</sub> pitoisuusmittauksista ne rivit joiden CO<sub>2</sub> pitoisuuden arvo oli yli 450 ppm. Eli yritettiin arvioida sitä, milloin luokissa on oppilaita. Histogrammista nähdään, että joissakin kouluissa TVOC pitoisuus on suurempi silloin kun oppilaita on koulussa. Tämä voi johtua siitä, että oppilaiden vaatteissa voi esimerkiksi pesuainejäämiä, jotka nostaa TVOC pitoisuutta tai luokissa käytetään paljon spriiiluokaisia tusseja.



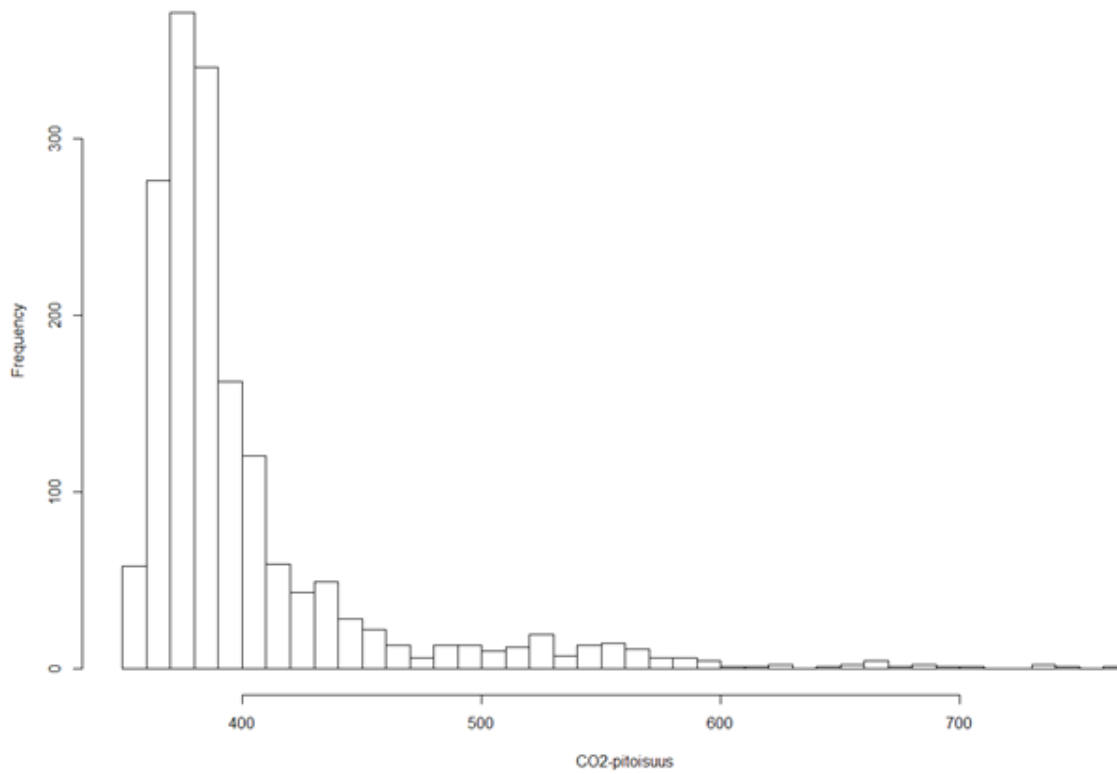
Kuva 18. Koulujen TVOC pitoisuuksien prosenttiosuudet



Kuva 19. Kouluaikojen keskiarvoistetut CO<sub>2</sub> pitoisuudet

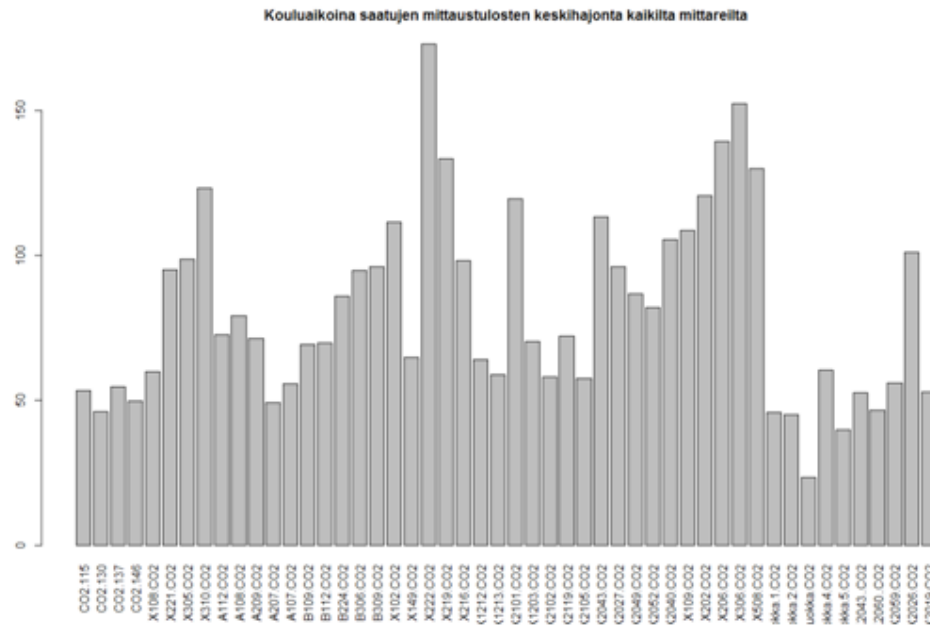
Tästä kuvasta päätellen CO<sub>2</sub>-pitoisuudet ovat ihan hyvällä mallilla keskimääräisesti.

Kaikista mittauspisteistä keskiarvoistetun aikasarjan histogrammi

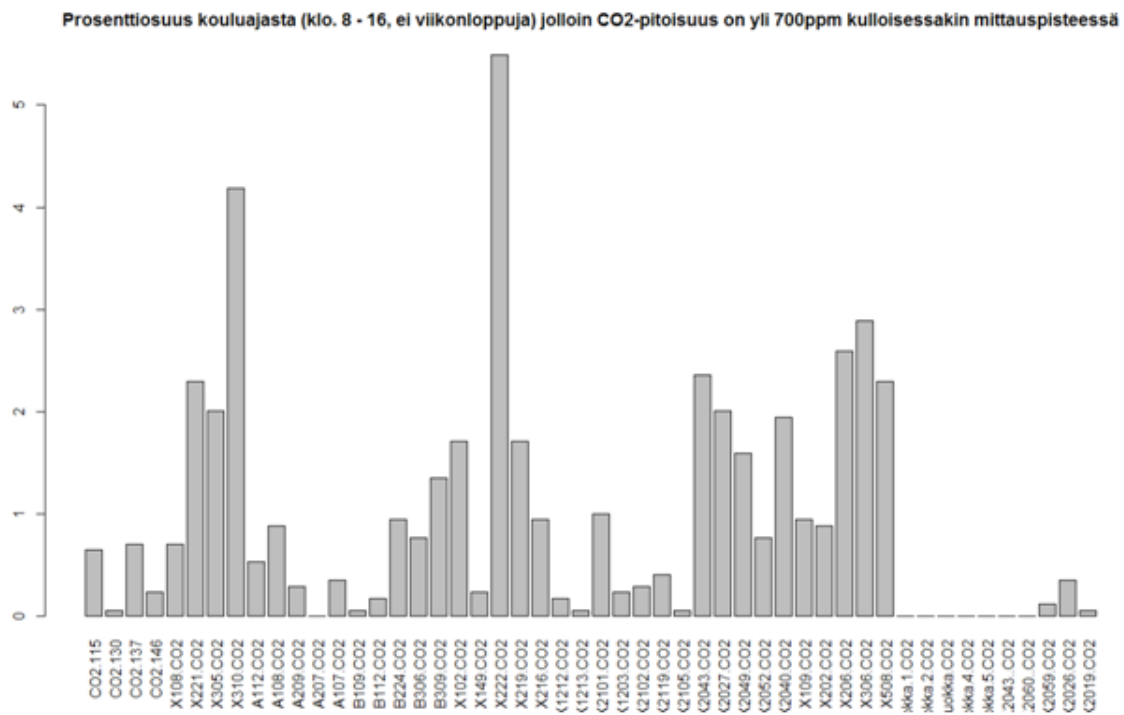


Kuva 20. Kouluaikeiden keskiarvoistetut CO2 pitoisuudet histogrammina

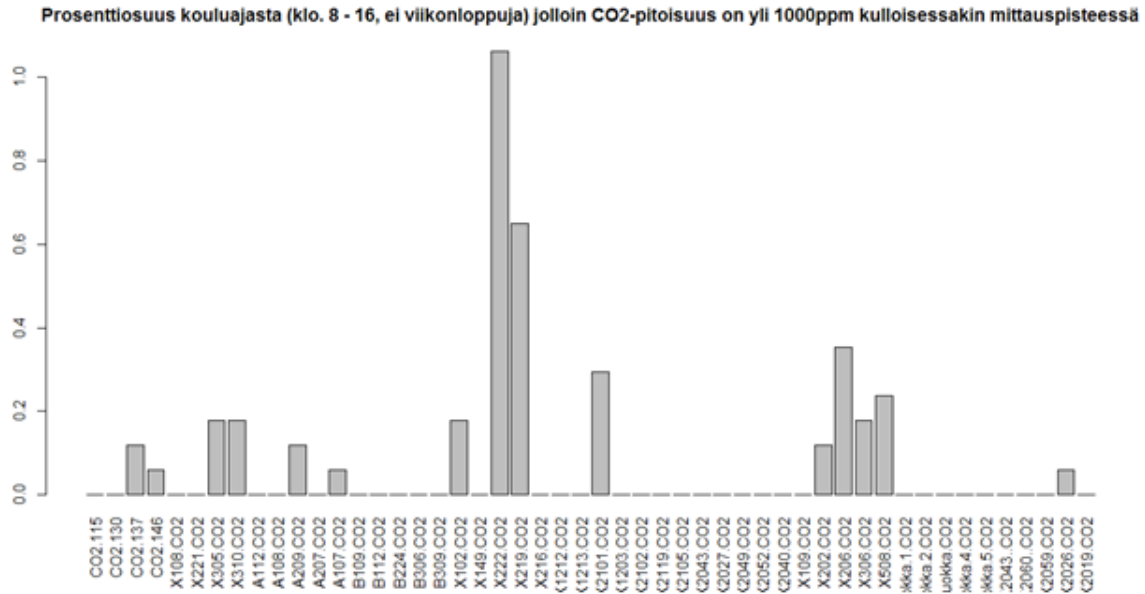
Tässä histogrammi kaikkien mittareiden keskiarvoistetusta aikasarjasta. Tämä tukee edellisen kuvan perusteella tekemääni päätelmää siitä, että CO2-pitoisuus on keskimäärin mittaushetkeissä hyvä.



Kuva 21. Kouluaikojen keskihajonta CO2 pitoisuuksista

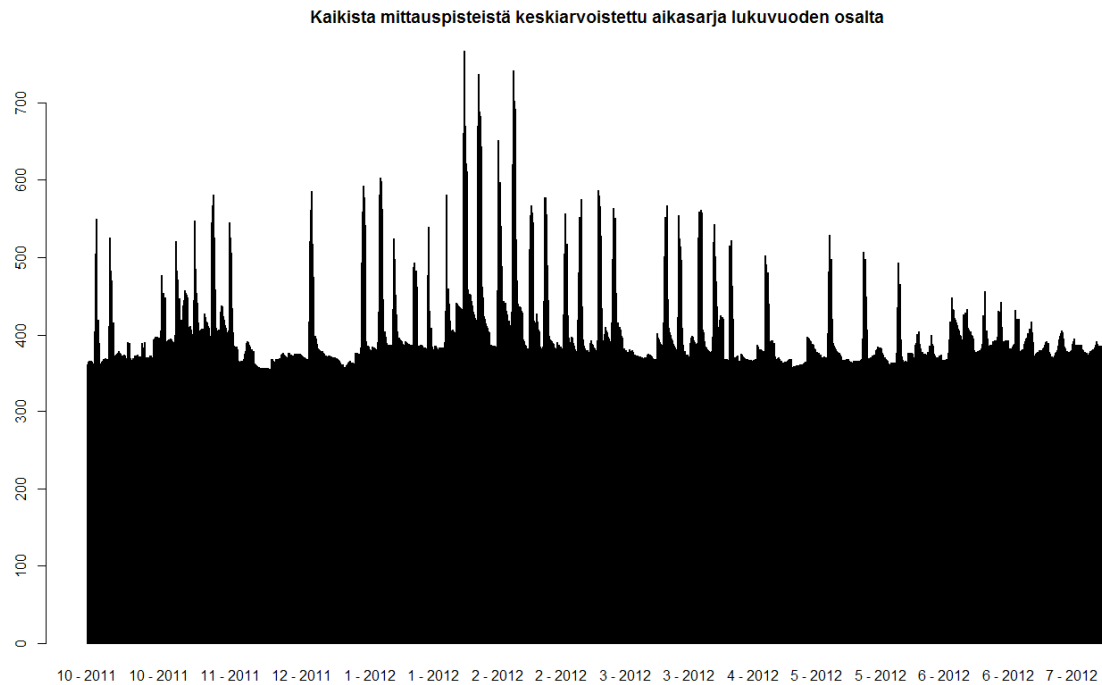


Kuva 22. Kouluaikojen yli 700 ppm CO2 pitoisuuksien prosenttiosuudet



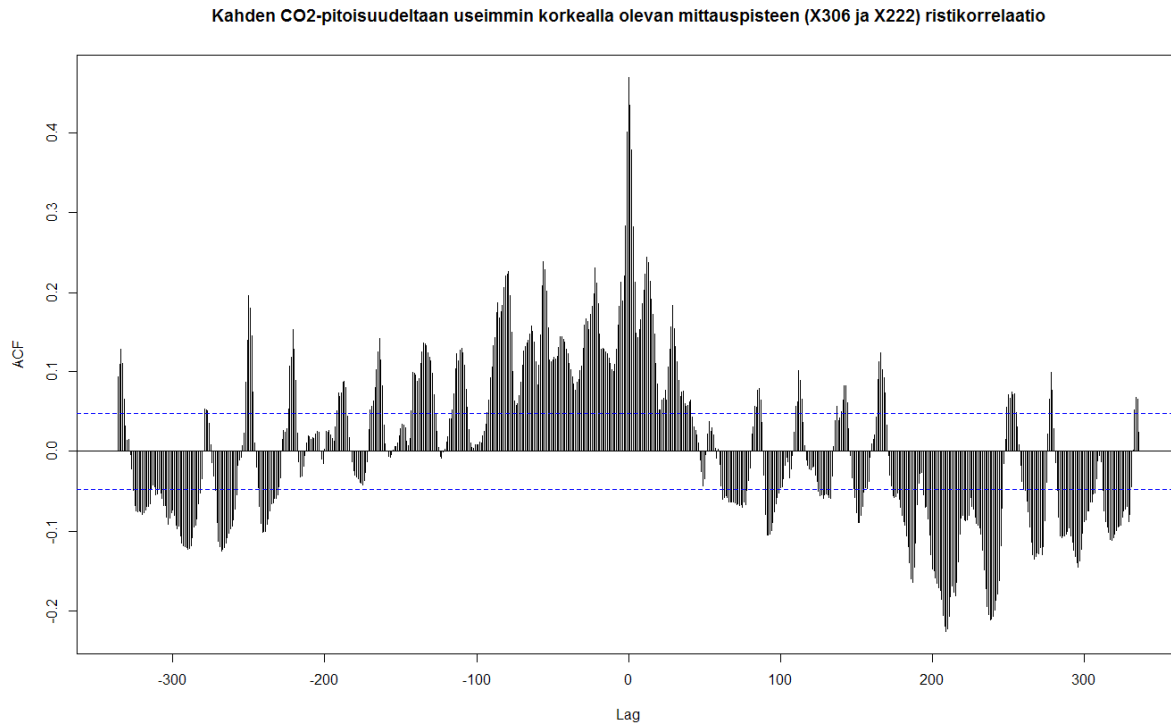
Kuva 23. Kouluaikeojen yli 1000 ppm CO<sub>2</sub> pitoisuuksien prosenttiosuudet

Näistä kolmesta yllä olevista kuvista (kuvat 21-23) on havaittavissa että samoissa mittauspisteissä joissa keskihajonta on korkea, on myös CO<sub>2</sub>-pitoisuus useammin lähellä jopa haitallisia arvoja (tosin vielä normien rajoissa). Tämä viittaa mahdollisesti suureen kuormitukseen (paljon ihmisiä) ja/tai riittämättömään ilmanvaihtoon. Ilmanvaihto ei kykene pitämään CO<sub>2</sub>-pitoisuutta alhaalla, jolloin se vaihtelee voimakkaammin.



**Kuva 24. Aikasarjakuvaaja kaikkien CO<sub>2</sub> mittauksien keskiarvoista lukuvuoden ajalta**

Kuvassa 24 on esitetty kaikista mittareista keskiarvoistettu aikasarja lukuvuoden osalta. Tässäkin kuvaajassa on pelkästään kouluajat mukana. Mielenkiintoista tässä kuvaajassa on se, että korkeimmat huiput näyttävät sattuvan helmikuun kohdalle.



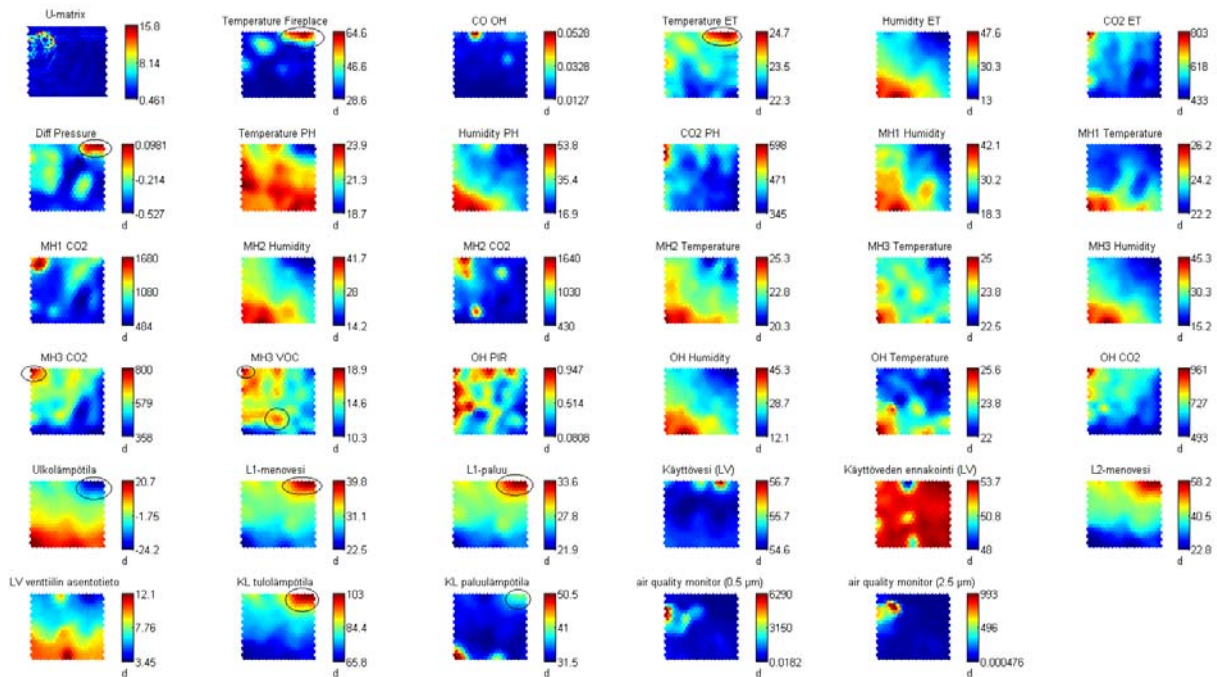
Kuva 25. Kahden useimmin korkeimpien CO<sub>2</sub> pitoisuuksien ristikorrelaatio

Kuvassa 25 on esitetty kahden useimmin kohonneita arvoja mitanneen mittarin ristikorrelaatio. Tästä ei näe oikein mitään selvää yhteyttä, paitsi että korrelaatio näyttäisi olevan vahvin ilman viivettä, eli CO<sub>2</sub>-pitoisuudet näiden luokkien välillä seuraavat jossain määrin toisiaan.

Markus

### 4.3 Itseorganisoituva kartta

Omakotitalon mittausaineistosta muodostettiin itseorganisoituva kartta. Kartan opetuksessa muuttujat skaalattiin varianssiskaalauksella. Itseorganisoutuvalla kartalla voi visualisoida muuttujat kuvaksi, josta nähdään helpommin muuttujien väliset riippuvuudet. Kuvassa 26 esitetty itseorganisoituvan kartan komponenttitasot muuttujittain. Muuttujien arvot ovat denormalisoitu.



Kuva 26. Itseorganisoituvan kartan komponenttitasot muuttujittain omakotitalon mittausaineistosta

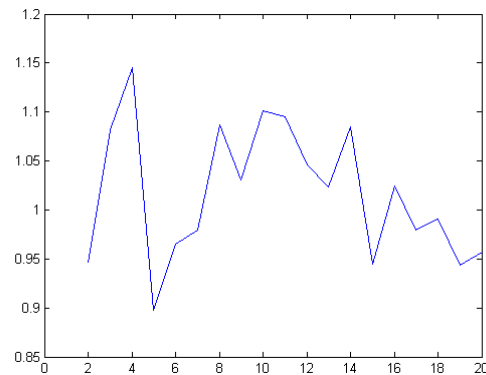
Itseorganisoituvan kartan komponenttitasosta nähdään samoja ilmiötä kuin muilla menetelmillä. Joitakin ilmiöitä on merkitty kuvaan ympyrällä. Esimerkiksi TVOC histogrammien tulokset (kuvat 13 ja 14) eli TVOC pitoisuudet ovat koholla kun henkilöitä ei ole paikalla ja myös kun henkilöitä ei ole paikalla. Ulkolämpötilan ollessa alhaalla takan lämpötila on korkeampi kuin muulloin. Kartasta nähdään myös takan käytöstä johtuva paine-eron kasvu, vaikka takan käytössä paine-eron pitäisi laskea. Tämä voi johtua siitä, että takanlämpötilaa mitataan takan pinnasta, niin lämmön nousu tulee viiveellä. Paine-eron kasvu johtuu ilmeisesti siitä, kun pellit laitetaan kiinni niin takka alkaa kunnolla lämmittämään sisäilmaa, joka lämmitessään laajentaa ilmaa ja siten talon ilmanpaine menee ylipaineiseksi.



Markus

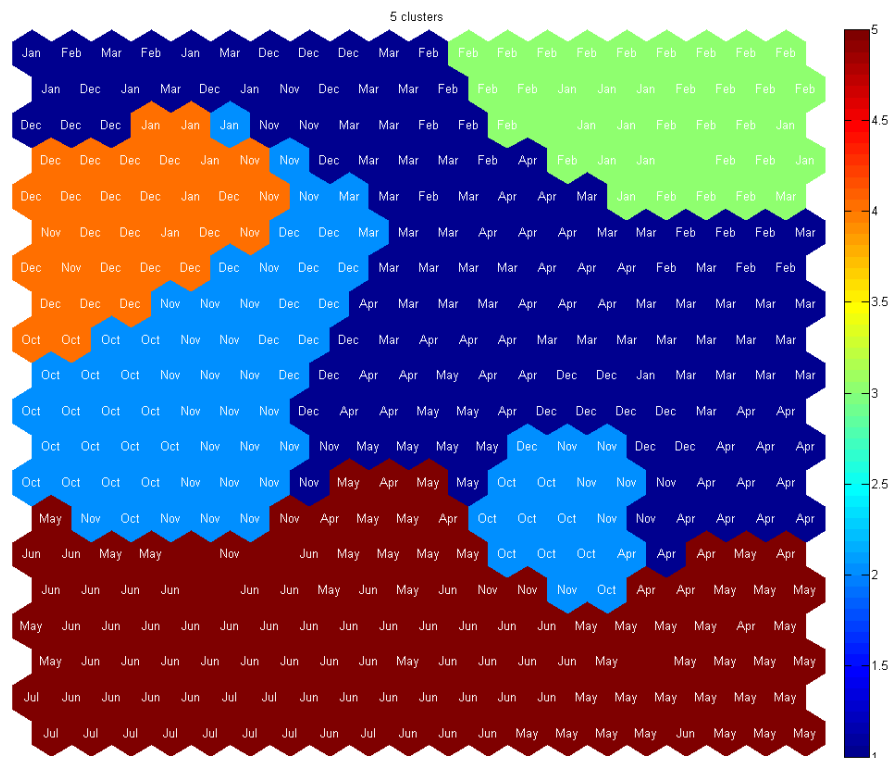
#### 4.4 Itseorganisoituvan kartan klusterointi k-means menetelmällä

Itseorganisoitua kartta klusteroitiin k-means menetelmällä siten, että klusteroinnin optimaalista klusterimäärää tarkasteltiin Davies-Bouldin indeksin avulla. Kuvassa 27 on esitetty Davies-Bouldin indeksi, jonka perusteella kartta klusteroitiin viiteen klusteriin.



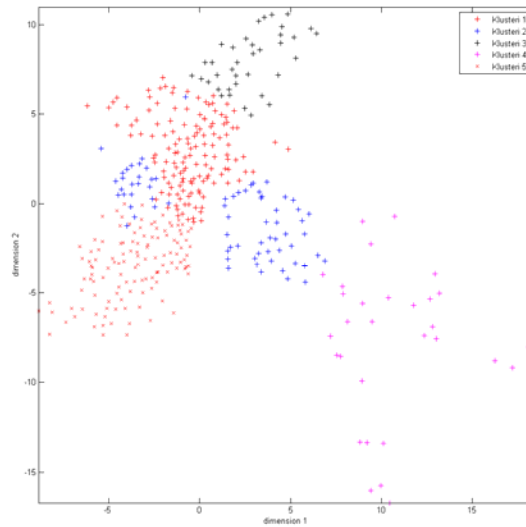
Kuva 27. Davies-Bouldin indeksi itseorganisoituvan kartan klusteroinnista

Kuvassa 28 on esitetty klusteroidun kartan komponenttitaso, jossa eri värit kertovat klusterin numeron ja alueen. Klusteroidun kartan pinnalle visualisoitiin kuukaudet, jotka mahdollisesti selittävät mitä kuukautta kartan yksi neuroni esittää. Esimerkiksi klusteri 3, näyttäisi selittävän talvikauden ja klusteri 5 kesäkauden.



Kuva 28. Klusteroitu itseorganisoitua kartta

Klusteroinnin hyvyttä ja ominaisuuksia tarkasteltiin Sammonin kuvauksella. Kuvassa 29 on esitetty Sammonin kuvaus klusteroidusta kartasta. Kuvasta nähdään miten klusteri 4 on suhteessa muihin klustereihin kauempana ja klusteri 2 on jakaantunut kahteen osaan. Tämän saman näkee myös itseorganisoituvan kartan pinnalle visualisoiduista klustereita, mutta Sammonin kuvauksella näkee klustereiden etäisyydet paremmin.

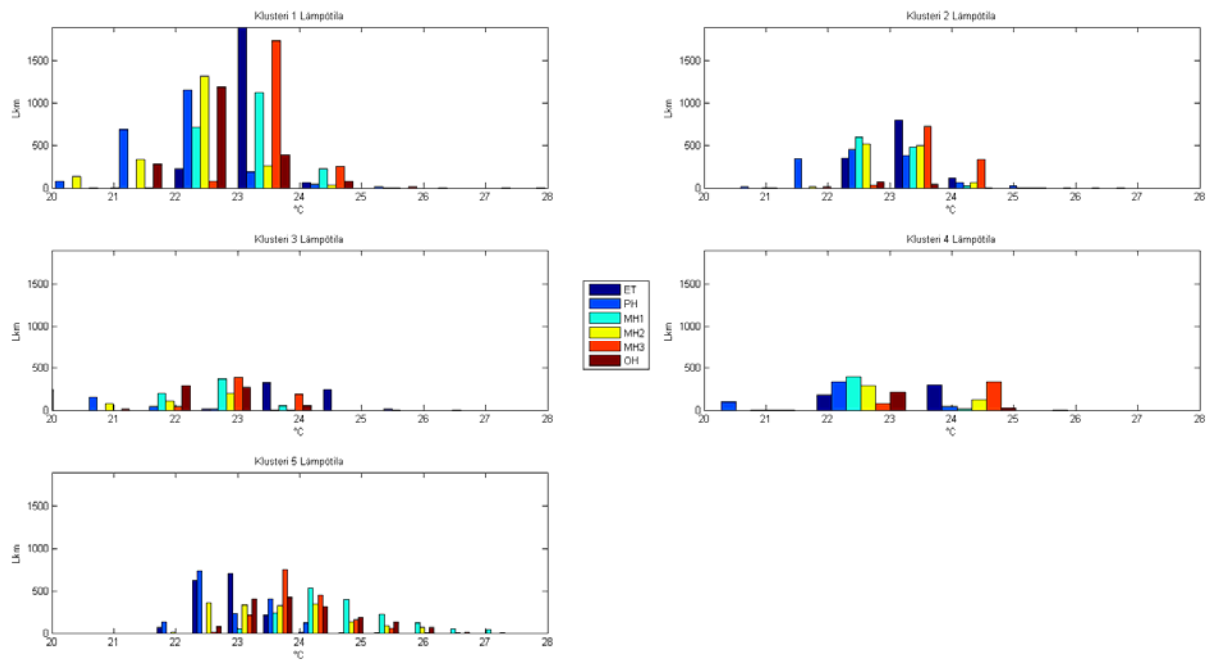


Kuva 29. Sammonin kuvaus klusteroinnista

Markus, Kimmo

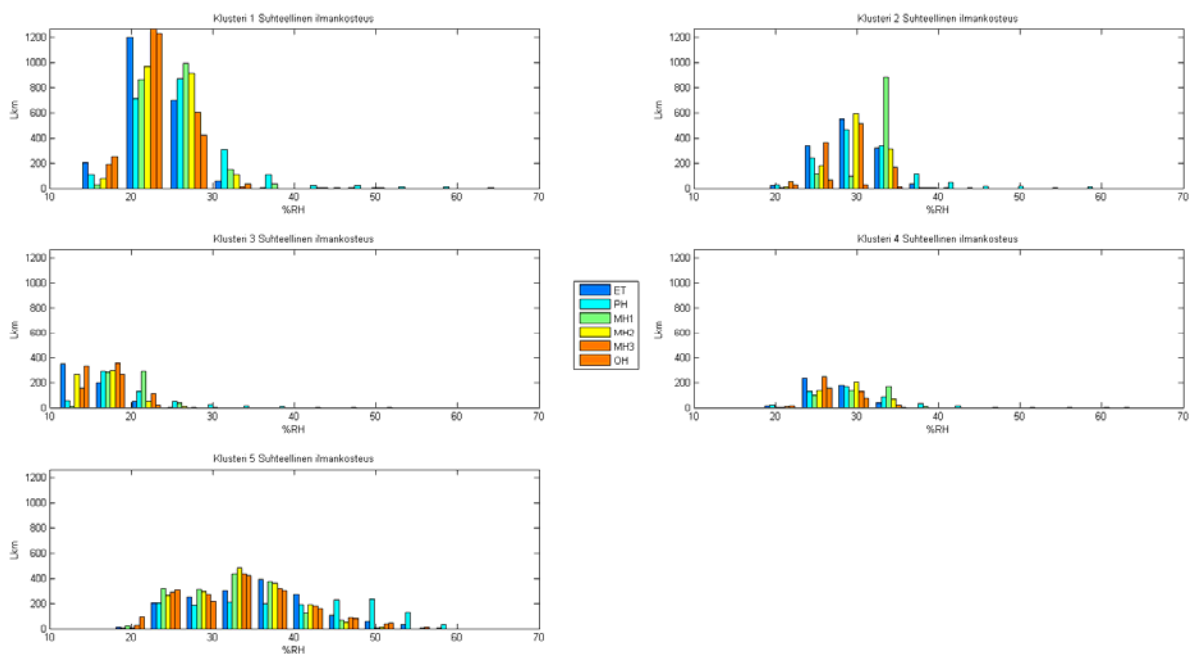
#### 4.5 Klusterien histogrammit

Klusteroidusta itseorganisoituvasta kartasta muodostettiin histogrammit. Kuvassa 30 on esitetty eri klustereiden lämpötilojen histogrammit. Kuvassa y akseli on lukumäärä eli montako tapausta kuuluu kyseiseen klusteriin ja x akseli on lämpötila.



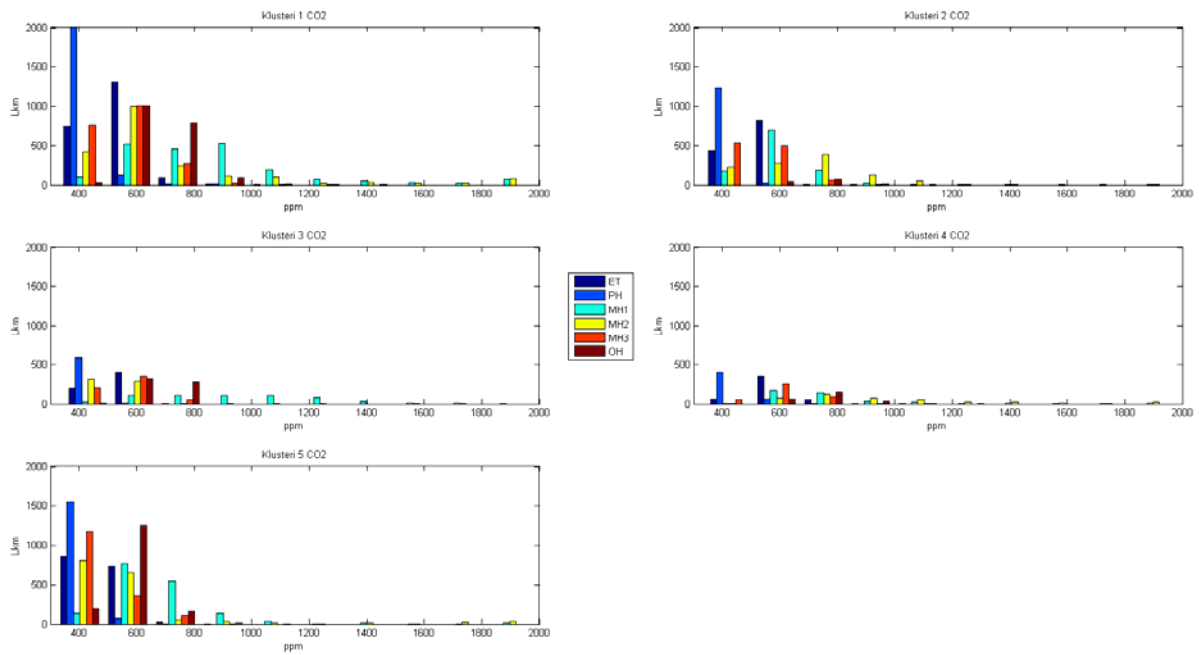
Kuva 30. Klustereiden histogrammit lämpötiloista

Kuvassa 31 on esitetty eri klustereiden ilmankosteuksien histogrammit. Kuvassa y akseli on lukumäärä eli montako tapausta kuuluu kyseiseen klusteriin ja x akseli on ilmankosteus.



Kuva 31. Klustereiden histogrammit ilmankosteuksista

Kuvassa 32 on esitetty eri klustereiden CO<sub>2</sub> pitoisuuksien histogrammit. Kuvassa y akseli on lukumäärä eli montako tapausta kuuluu kyseiseen klusteriin ja x akseli on pitoisuus.



Kuva 32. Klustereiden histogrammit CO<sub>2</sub> pitoisuuksista

Markus, Mikko, Kimmo

## 5 JOHTOPÄÄTÖKSET

Koulu ja omakotitaloaineistosta pystyi tekemään monipuolisia analyysyjä. Sisäilmanlaadun analysointia näyttäisi tuloksien perusteella tekemään hyvin aikasarjakuvaajilla, autokorrelaatioilla ja histogrammeilla. Tarkempaan analysointiin voi käyttää itseorganisoituvaa karttaa, jonka tuloksia ei tässä harjoitustyössä kuitenkaan tarkemmin analysoitu. Sisäilmanmittausaineistossa oli erityyppisiä mittauksia. Näitä mittauksia analysoidessa monimuuttujamenetelmillä, kuten itseorganisoituvilla kartoilla, voi mittauksien välisiä riippuvuuksia analysoida helpommin.

Tuloksien perusteella kaikissa kymmenessä koulussa oli sisäilman laatu CO<sub>2</sub> pitoisuuden ja TVOC pitoisuuden osalta hyvä. Koulujen välistä eroa ei ollut paljoa, koska suurimmaksi osaksi sisäilman laatu oli hyvä. Aineistosta olisi pitänyt vielä analysoida lämpötila ja ilmankosteus. Kouluissa näyttäisi CO<sub>2</sub> pitoisuus nousevan varsin nopeasti oppituntien aikana, mutta välituntien aikana luokat kerkeävät tuulettua riittävästi joten ilmanvaihdon tehostamista ei välttämättä tarvitse lisätä.

Omakotitalossa makuuhuoneiden CO<sub>2</sub> pitoisuus on suurimmaksi osaksi hyvä, mutta näyttäisi kasvavan yö aikaan liian suureksi eli yli sisäilmastoluokan S3 rajan yli. Omakotitalon ilmankosteus on myös hyvä, mutta talviaikaan liian pieni, tarvittaessa olisi syytä lisätä ilmankosteutta ilmankostuttajalla. Omakotitalon tuloksista nähtiin, että lämpötilat ovat liian suuret, tämä vähentää asumisviihtyvyyttä varsinkin yöllä ja lisäksi CO<sub>2</sub> pitoisuudet olivat liian suuret yöaikaan. Itseorganisoituvan kartan avulla olisi saanut selvitettyä muiden mittauksien ja mahdollisesti vuodenaikojen vaikutukset sisäilmanlaatuun.

Lähteet

Asumisterveysohje

[http://www.finlex.fi/pdf/normit/14951-asumisterveysohje\\_pdf.pdf](http://www.finlex.fi/pdf/normit/14951-asumisterveysohje_pdf.pdf)

sisäilmayhdistys

[http://www.sisailmayhdistys.fi/portal/terveelliset\\_tilat/sisailmasto/](http://www.sisailmayhdistys.fi/portal/terveelliset_tilat/sisailmasto/)

## LIITE 1 Koulujen mittauksien tilastolliset ominaisuudet

	yksikkö	minimi	maksimi	keskiarvo	mediaani	keskihajonta
<b>Humidity 115</b>	%RH	11,7	164,1	62,0	58,1	26,5
<b>Temperature 115</b>	°C	20,4	25,6	23,0	22,8	0,9
<b>CO2 115</b>	ppm	357,8	1040,7	401,7	387,5	54,5
<b>Humidity 130</b>	%RH	5,4	63,9	25,4	24,0	10,5
<b>Temperature 130</b>	°C	18,2	25,7	22,1	22,1	1,3
<b>CO2 130</b>	ppm	312,4	849,0	375,2	358,1	49,4
<b>Humidity 137</b>	%RH	6,6	67,5	27,3	26,6	9,7
<b>Temperature 137</b>	°C	21,1	25,6	22,7	22,6	0,6
<b>CO2 137</b>	ppm	392,4	1035,7	441,3	424,7	57,0
<b>Humidity 144</b>	%RH	5,3	66,2	25,6	24,7	10,3
<b>Temperature 144</b>	°C	18,6	24,7	21,2	21,2	0,9
<b>CO 2144</b>	ppm	299,8	1073,4	366,7	343,1	64,6
<b>Humidity 146</b>	%RH	4,1	63,9	24,7	22,9	10,6
<b>Temperature 146</b>	°C	18,5	28,2	21,1	21,1	0,8
<b>CO2 146</b>	ppm	326,2	1109,9	378,9	368,3	48,7
<b>VOC 137</b>	ppm	0,0	19,4	4,9	4,4	4,6
<b>Koulu 1 end</b>		1,0	1,0	1,0	1,0	0,0
<b>115 Humidity</b>	%RH	5,1	47,5	21,3	21,0	7,8
<b>115 Temperature</b>	°C	19,1	25,1	21,9	22,0	1,3
<b>115 CO2</b>	ppm	337,3	898,9	402,7	378,8	71,6
<b>108 Humidity</b>	%RH	6,4	50,3	23,9	23,3	8,3
<b>108 Temperature</b>	°C	19,7	22,9	21,1	21,1	0,4
<b>108 CO2</b>	ppm	384,3	994,7	499,7	491,0	64,2
<b>221 Humidity</b>	%RH	6,5	43,6	22,3	22,5	6,7
<b>221 Temperature</b>	°C	20,6	25,8	22,8	22,7	0,9
<b>221 CO2</b>	ppm	359,2	1930,5	428,1	393,8	98,4
<b>305 Humidity</b>	%RH	5,6	44,7	21,7	21,6	7,0
<b>305 Temperature</b>	°C	19,5	26,0	22,3	22,0	1,1
<b>305 CO2</b>	ppm	307,2	1199,0	385,3	357,4	97,5
<b>310 Humidity</b>	%RH	6,5	44,3	22,0	22,1	6,7
<b>310 Temperature</b>	°C	19,4	26,5	22,0	21,6	1,4
<b>310 CO2</b>	ppm	333,6	1242,0	415,9	374,3	119,1
<b>108 VOC</b>	ppm	0,0	29,6	4,4	3,2	5,0
<b>Koulu 2 end</b>		2,0	2,0	2,0	2,0	0,0
<b>A112 Temperature</b>	°C	17,2	24,6	20,9	20,7	1,0
<b>A112 Humidity</b>	%RH	3,9	46,9	22,4	21,9	8,3
<b>A112 CO2</b>	ppm	301,9	994,0	348,6	330,9	59,6
<b>A108 Humidity</b>	%RH	3,6	46,4	21,7	21,2	8,0
<b>A108 Temperature</b>	°C	18,8	24,8	21,9	21,7	0,9
<b>A108 CO2</b>	ppm	368,4	1097,6	425,6	400,7	71,6
<b>A209 Humidity</b>	%RH	3,5	49,8	21,3	20,9	9,0

<b>A209 Temperature</b>	°C	16,1	24,4	21,0	20,9	1,3
<b>A209 CO2</b>	ppm	318,6	1289,5	375,9	357,9	61,0
<b>A207 Humidity</b>	%RH	4,2	47,4	22,2	22,1	8,0
<b>A207 Temperature</b>	°C	17,6	26,3	20,9	20,6	1,3
<b>A207 CO2</b>	ppm	312,9	744,2	364,5	350,3	49,6
<b>A107 Humidity</b>	%RH	3,9	52,8	22,5	22,0	8,5
<b>A107 Temperature</b>	°C	18,2	25,5	22,0	21,8	0,9
<b>A107 CO2</b>	ppm	364,0	1018,0	405,1	392,0	51,5
<b>A112 VOC</b>	ppm	0,0	18,1	3,5	3,0	3,7
<b>A107 VOC</b>	ppm	0,0	23,3	6,4	6,5	3,8
<b>Koulu 3 end</b>		3,0	3,0	3,0	3,0	0,0
<b>B109 Temperature</b>	°C	19,0	24,6	21,4	21,2	0,9
<b>B109 Humidity</b>	%RH	5,6	48,3	23,8	23,5	7,0
<b>B109 CO2</b>	ppm	275,4	1074,2	346,8	318,0	71,3
<b>B112 Humidity</b>	%RH	0,1	43,4	17,7	20,6	9,9
<b>B112 Temperature</b>	°C	19,8	25,4	22,0	21,7	1,1
<b>B112 CO2</b>	ppm	298,9	907,0	363,7	339,0	65,1
<b>B224 Humidity</b>	%RH	5,8	51,5	23,4	23,3	6,5
<b>B224 Temperature</b>	°C	19,9	27,1	22,5	22,3	1,1
<b>B224 CO2</b>	ppm	296,9	988,2	385,8	357,6	78,5
<b>B306 Humidity</b>	%RH	5,7	47,8	25,0	25,0	6,5
<b>B306 Temperature</b>	°C	19,3	28,4	22,1	21,6	1,4
<b>B306 CO2</b>	ppm	256,0	1245,8	365,1	332,9	96,8
<b>B309 Humidity</b>	%RH	4,4	41,8	21,3	21,5	6,0
<b>B309 Temperature</b>	°C	19,8	26,6	22,0	21,5	1,3
<b>B309 CO2</b>	ppm	282,9	1430,8	377,2	340,2	103,6
<b>B309 VOC</b>	ppm	0,0	26,9	10,7	10,7	8,1
<b>Koulu 4 end</b>		4,0	4,0	4,0	4,0	0,0
<b>149 VOC</b>	ppm	0,0	10,7	0,1	0,0	0,4
<b>102 Humidity</b>	%RH	5,6	44,7	21,6	22,4	7,3
<b>102 Temperature</b>	°C	14,4	25,3	21,6	21,4	1,2
<b>102 CO2</b>	ppm	298,7	1160,4	366,3	340,9	82,7
<b>149 Humidity</b>	%RH	2,6	43,1	17,6	17,9	7,3
<b>149 Temperature</b>	°C	20,0	25,0	22,7	22,6	0,6
<b>149 CO2</b>	ppm	302,1	1237,7	366,7	347,3	65,0
<b>222 Humidity</b>	%RH	2,9	46,3	19,1	19,5	7,7
<b>222 Temperature</b>	°C	18,8	23,9	22,1	22,0	0,5
<b>222 CO2</b>	ppm	403,8	1611,4	511,2	452,5	158,2
<b>219 Humidity</b>	%RH	3,6	48,0	20,1	20,6	7,3
<b>219 Temperature</b>	°C	18,2	24,5	21,4	21,2	0,8
<b>219 CO2</b>	ppm	304,3	1909,4	386,8	348,8	121,4
<b>216 Humidity</b>	%RH	4,0	51,9	21,4	22,2	7,4
<b>216 Temperature</b>	°C	17,0	21,8	19,6	19,6	0,7
<b>216 CO2</b>	ppm	308,0	1780,5	379,2	345,0	92,3
<b>216 VOC</b>	ppm	0,0	28,3	3,8	3,3	3,8

<b>Koulu 5 end</b>		5,0	5,0	5,0	5,0	0,0
<b>1212 Temperature</b>	°C	19,6	24,9	21,0	20,9	1,0
<b>1212 Humidity</b>	%RH	5,7	38,4	21,8	22,6	6,6
<b>1212 CO2</b>	ppm	311,4	848,6	389,1	368,1	67,9
<b>1213 Humidity</b>	%RH	4,0	42,8	20,7	21,3	7,1
<b>1213 Temperature</b>	°C	20,0	25,2	21,5	21,3	0,9
<b>1213 CO2</b>	ppm	318,8	839,8	379,8	360,9	60,8
<b>1203 Humidity</b>	%RH	5,7	40,2	23,4	24,0	7,1
<b>1203 Temperature</b>	°C	19,4	24,4	20,6	20,4	0,9
<b>2101 Humidity</b>	%RH	5,5	44,8	22,6	22,9	7,2
<b>2101 Temperature</b>	°C	20,0	23,8	21,3	21,2	0,8
<b>2101 CO2</b>	ppm	359,6	1534,5	439,7	412,3	83,1
<b>1203 CO2</b>	ppm	336,8	1010,4	405,2	380,9	71,1
<b>2102 Humidity</b>	%RH	3,7	54,6	21,2	21,2	7,8
<b>2102 Temperature</b>	°C	19,9	24,0	21,6	21,6	0,8
<b>2102 CO2</b>	ppm	421,7	1018,0	486,8	470,2	66,3
<b>2119 Humidity</b>	%RH	4,8	41,5	21,6	22,2	6,7
<b>2119 Temperature</b>	°C	18,7	26,9	21,1	21,0	1,2
<b>2119 CO2</b>	ppm	344,8	961,0	402,8	382,0	63,2
<b>2105 Humidity</b>	%RH	4,3	42,9	20,6	20,9	6,8
<b>2105 Temperature</b>	°C	21,2	25,4	22,6	22,5	0,7
<b>2105 CO2</b>	ppm	357,3	921,9	410,5	394,3	56,8
<b>1213 VOC</b>	ppm	0,0	16,3	2,7	2,4	3,1
<b>2105 VOC</b>	ppm	0,0	6,8	0,0	0,0	0,3
<b>Koulu 6 end</b>		6,0	6,0	6,0	6,0	0,0
<b>2043 Humidity</b>	%RH	3,9	43,6	19,0	18,9	5,9
<b>2043 Temperature</b>	°C	20,5	26,3	23,2	23,3	0,9
<b>2043 CO2</b>	ppm	286,1	1067,5	418,0	394,5	110,0
<b>2027 Humidity</b>	%RH	4,8	48,4	19,5	18,8	7,3
<b>2027 Temperature</b>	°C	21,0	24,4	22,4	22,3	0,7
<b>2027 CO2</b>	ppm	335,1	1060,4	452,1	420,4	97,5
<b>2049 Humidity</b>	%RH	5,2	45,6	20,1	19,5	6,5
<b>2049 Temperature</b>	°C	20,1	25,9	22,6	22,4	0,9
<b>2049 CO2</b>	ppm	344,1	866,6	432,9	414,5	82,2
<b>2052 Humidity</b>	%RH	5,4	45,1	20,0	19,5	6,3
<b>2052 Temperature</b>	°C	20,8	24,4	22,2	22,1	0,7
<b>2052 CO2</b>	ppm	311,8	859,5	402,3	378,1	87,9
<b>2040 Humidity</b>	%RH	4,4	46,1	19,6	19,1	6,3
<b>2040 Temperature</b>	°C	20,8	24,9	22,5	22,3	0,8
<b>2040 CO2</b>	ppm	319,1	958,7	408,1	384,0	99,7
<b>2043 VOC</b>	ppm	0,0	24,8	8,1	6,9	7,1
<b>2049 VOC</b>	ppm	0,0	20,1	4,8	2,7	5,7
<b>Koulu 7 end</b>		7,0	7,0	7,0	7,0	0,0
<b>109 Humidity</b>	%RH	4,8	45,8	21,6	21,9	6,6
<b>109 Temperature</b>	°C	20,0	24,4	22,2	22,2	0,7



109 CO2	ppm	323,0	1047,7	410,8	364,1	104,0
202 Humidity	%RH	3,2	46,8	20,1	19,8	7,3
202 CO2	ppm	315,4	1065,8	397,2	343,6	129,2
202 Temperature	°C	20,4	25,6	23,0	23,0	0,6
206 Humidity	%RH	4,4	44,1	20,4	20,1	6,5
206 Temperature	°C	19,5	26,6	23,8	23,8	1,0
206 CO2	ppm	356,8	1525,2	454,7	403,1	125,8
306 Humidity	%RH	5,6	48,3	22,1	21,9	6,5
306 Temperature	°C	20,3	25,0	21,7	21,4	0,8
306 CO2	ppm	328,7	1502,1	444,1	368,5	156,0
508 Humidity	%RH	5,8	46,8	23,4	23,7	6,6
508 Temperature	°C	20,6	24,8	22,4	22,5	0,9
508 CO2	ppm	351,5	1380,4	464,4	419,2	114,5
202 VOC	ppm	0,0	30,1	4,2	2,7	5,4
Koulu 8 end		8,0	8,0	8,0	8,0	0,0
luokka 1 Humidity	%RH	6,1	53,2	29,0	28,4	7,6
luokka 1 Temperature	°C	19,0	24,9	21,7	21,5	1,1
luokka 1 CO2	ppm	271,3	790,5	336,1	317,4	64,7
luokka 2 Humidity	%RH	5,4	50,4	27,0	24,9	8,1
luokka 2 Temperature	°C	20,0	24,8	22,0	21,8	0,9
luokka 2 CO2	ppm	280,2	826,6	337,5	320,2	57,6
pienluokka Humidity	%RH	5,5	67,3	32,1	30,6	10,4
pienluokka Temperature	°C	17,1	23,0	21,3	21,6	1,1
pienluokka CO2	ppm	392,2	778,9	434,5	429,5	34,9
luokka 4 Humidity	%RH	6,2	52,1	28,0	27,0	7,7
luokka 4 Temperature	°C	19,2	24,7	21,7	21,5	0,9
luokka 4 CO2	ppm	330,1	879,0	390,0	361,9	80,3
luokka 5 Humidity	%RH	8,5	46,8	29,2	29,3	6,1
luokka 5 Temperature	°C	19,3	28,4	22,0	21,7	1,4
luokka 5 CO2	ppm	378,4	828,0	421,2	402,0	50,1
luokka 2 VOC	ppm	0,0	24,0	6,2	5,6	4,2
pienluokka VOC	ppm	0,0	21,7	2,7	0,2	4,3
Koulu 9 end		9,0	9,0	9,0	9,0	0,0
2043 Humidity	%RH	7,4	36,4	20,0	19,6	5,0
2043 Temperature	°C	21,3	24,9	22,7	22,3	0,8
2043 CO2	ppm	344,2	1118,7	447,9	429,1	74,6
2060 Humidity	%RH	6,8	31,0	17,9	17,7	4,5
2060 Temperature	°C	21,4	25,3	22,6	22,4	0,7
2060 CO2	ppm	298,9	1078,2	429,1	410,9	84,7
2059 Humidity	%RH	7,6	34,4	20,0	19,5	5,0
2059 Temperature	°C	21,3	25,0	22,5	22,2	0,8
2059 CO2	ppm	342,8	1186,3	428,3	410,0	79,1
2026 Humidity	%RH	7,9	31,5	18,6	18,3	4,5
2026 Temperature	°C	21,9	24,9	22,7	22,6	0,4
2026 CO2	ppm	387,9	1037,8	440,4	432,1	41,4

<b>2019 Humidity</b>	%RH	8,6	32,6	19,9	19,4	4,2
<b>2019 Temperature</b>	°C	21,0	25,0	22,6	22,5	0,7
<b>2019 CO2</b>	ppm	332,2	1499,4	436,4	421,6	76,5
<b>2060 VOC</b>	ppm	0,0	21,9	5,2	2,7	6,8
<b>2019 VOC</b>	ppm	0,0	21,2	2,0	0,0	5,0
<b>Koulu 10 end</b>		10,0	10,0	10,0	10,0	0,0