

Aalto Yliopisto

MS-C2133 Operaatiotutkimuksen laboratoriotyöt I

Sähkönkulutuksen ennustaminen ja hankinnan optimointi

Daud Abucar, Bilqays Ayoub

Palautettu: 11. tammikuuta 2025

Sisällys

1	Johdanto	3
2	Aikasarjan analyysi	4
2.1	Mallin rakennus	9
2.2	Mallin diagnostiikka	12
3	Hankinnan optimointi	15
4	Parannusehdotukset	17
5	Yhteenveto	18
6	Liitteet	19

1 Johdanto

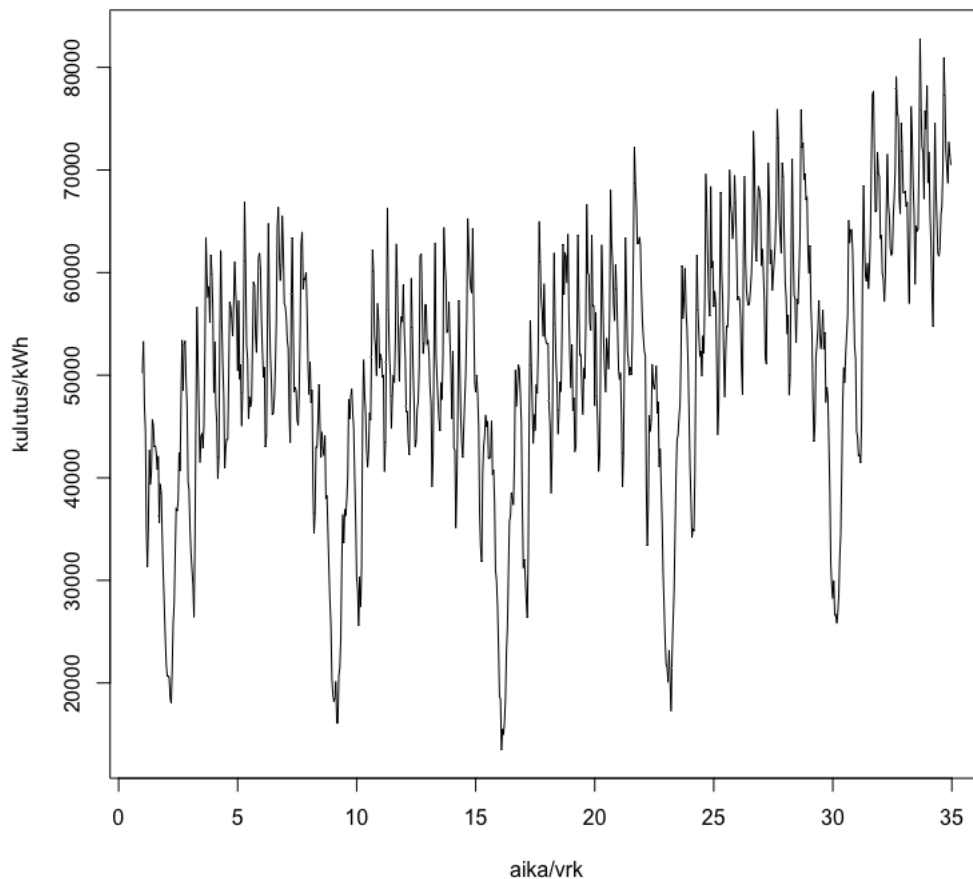
Loppuraportissa käsittelemme kurssin Operaatiotutkimuksen laboratoriotyöt I toisen työn eli Sähkönkulutuksen ennustaminen ja hankinnan optimointia. Työn kokonaisuus koostui mallin rakentamisesta, ennustamisesta ja optimoinnista. Tämän loppuraportin kirjoittamiseen on hyödynnetty työohjetta ja Prediction and Time Series Analysis luentoja. Raportin tukena on R:llä tehdyt hypoteesi testit, mikä validoi meidän tuloksiamme.

Laboratoriotyössä keskitytään SARIMAX- mallin rakentamiseen sähkönkulutukselle, jossa ulkoisena selittävänä tekijänä on ulkolämpötila. Mallin rakentamisessa käytettiin SARIMAX-mallia eli aikasarjaa, joka sisältää SAR, SMA ja ulkoisen tekijän X. Aikasarjan analyysissä hyödynnetään Box-Jenkins-menetelmää. Seasonal ARIMA- mallissa esiintyy kausivaihtelua, mikä näkyy aikasarjassa. Mallinnus prosessissa määritetään kausivaihtelun pituus stationaarisuuden takaamiseksi sekä määritellään kertalukujen parametrit autokorrelaatio- (ACF) ja osittaisautokorrelaatiofunktioiden (PACF) perusteella. Mallin rakenteen valmistuessa suoritetaan tarkastellaan residuaalisen satunnaisuutta suorittamalla Ljung-Boxin χ^2 testiä residuaalien arviointiin. Mallin valmistuessa ennustettiin seuraavan 24 tunnin sähkönkulutusta. Sähkön hankinnassa verrataan useita hankintavaihtoehtoja, jotka tarjoavat eri sähkön hankinja ja myyntihintoja. Tavoitteena on minimoida sähkön hankinta kustannuksia siten, että maksimoidaan voittoa. Tässä osuudessa korostuu riskinhallinta strategisen päätöksen teon seurauksena. Raportissa olevat kuvat ja parametriarvot ovat ennustamiseen ja optimoinnissa käytetty malli, mikä valittiin parhaaksi malliksi.

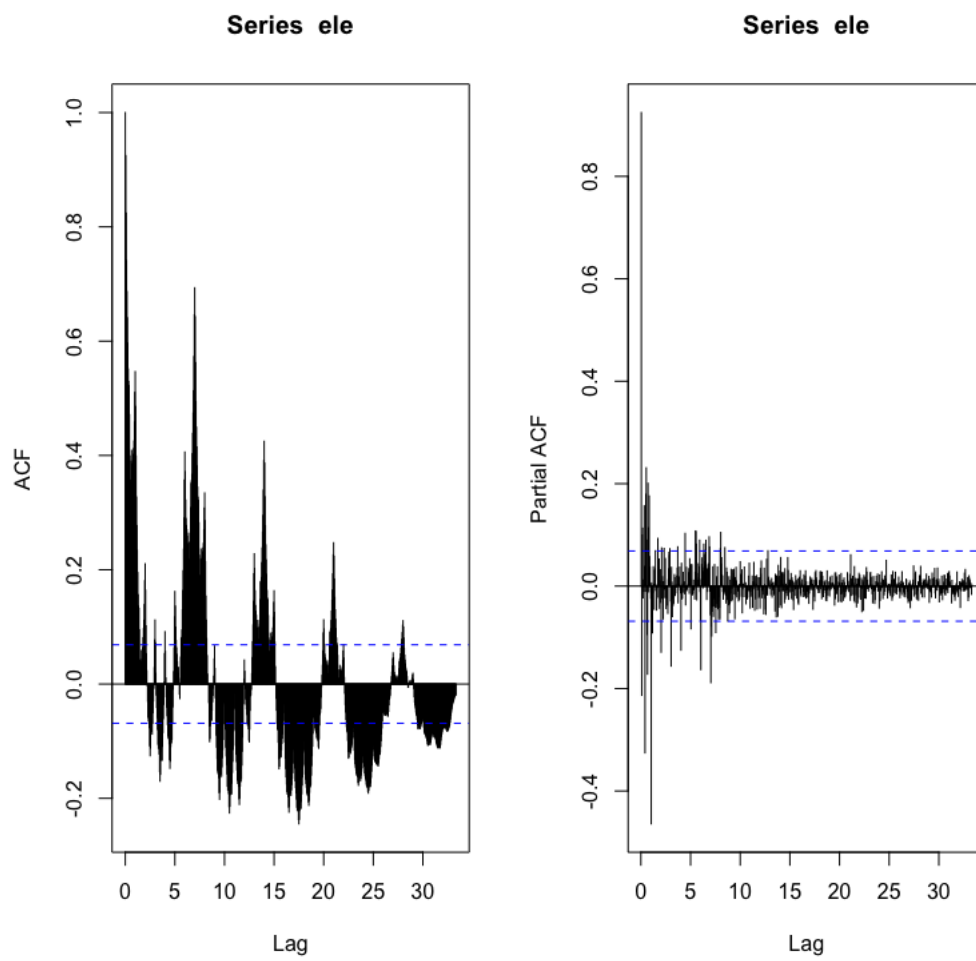
Raportin etenemisjärjestys vastaa projektin etenemisjärjestystä, jotta lukija ymmärtää tehtyjä valintoja johdonmukaisesti. Raportti on rakenteeltaan seuraava: Toisessa kappaleessa esitetään, kuinka SARIMAX- malli on valittu ja tehtyjen testien tulokset avataan johtopäätöksineen. Myös ennustukseen liittyvät parametrit on kerrottu tässä kappaleessa. Kolmannessa kappaleessa käydään läpi sähkön hankinnan optimointia ja tarkastellaan sen tuottamia tuloksia. Neljännessä kappaleessa käydään läpi parannusehdotuksia ja erikoistapaukset pidemmän ennustuksen varalta ja käsitellään yhteenvedo viidessä kappaleessa.

2 Aikasarjan analyysi

Data koostui kahdesta komponentista sähkönkulutuksesta ja lämpötilasta. Selittäjä muuttujana oli lämpötila ja selitettävä muuttuja oli sähkönkulutus 840 data pistettä, joista 24 käytettiin ennustamiseen, loput ja kaikki 816 selittäjämuuttujista käytettiin mallinrakennuksessa. Data havainnoillistaan plot työkalulla kuva 1, josta voidaan päätellä sähkönkustannuksen vaihtelevan viikoittain ja päivittäin. ACF ja PACF kuvaajista kuva 2 nähdään, että aikasarjaa täytyy differoida. Koska aikasarjalla on kaksi kausivaihtelua, kokeiltiin differointia arvoilla 1 ja kausivaihtelun differointia arvoilla 1, 24 ja 168.

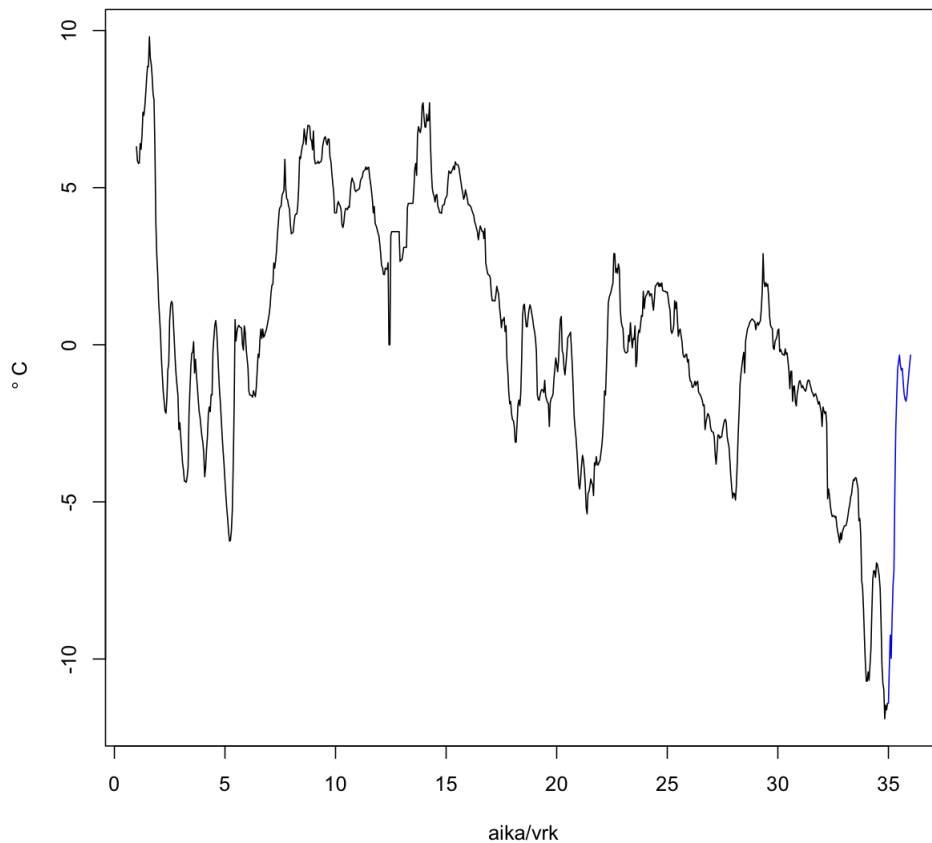


Kuva 1: Sähkönkulutus

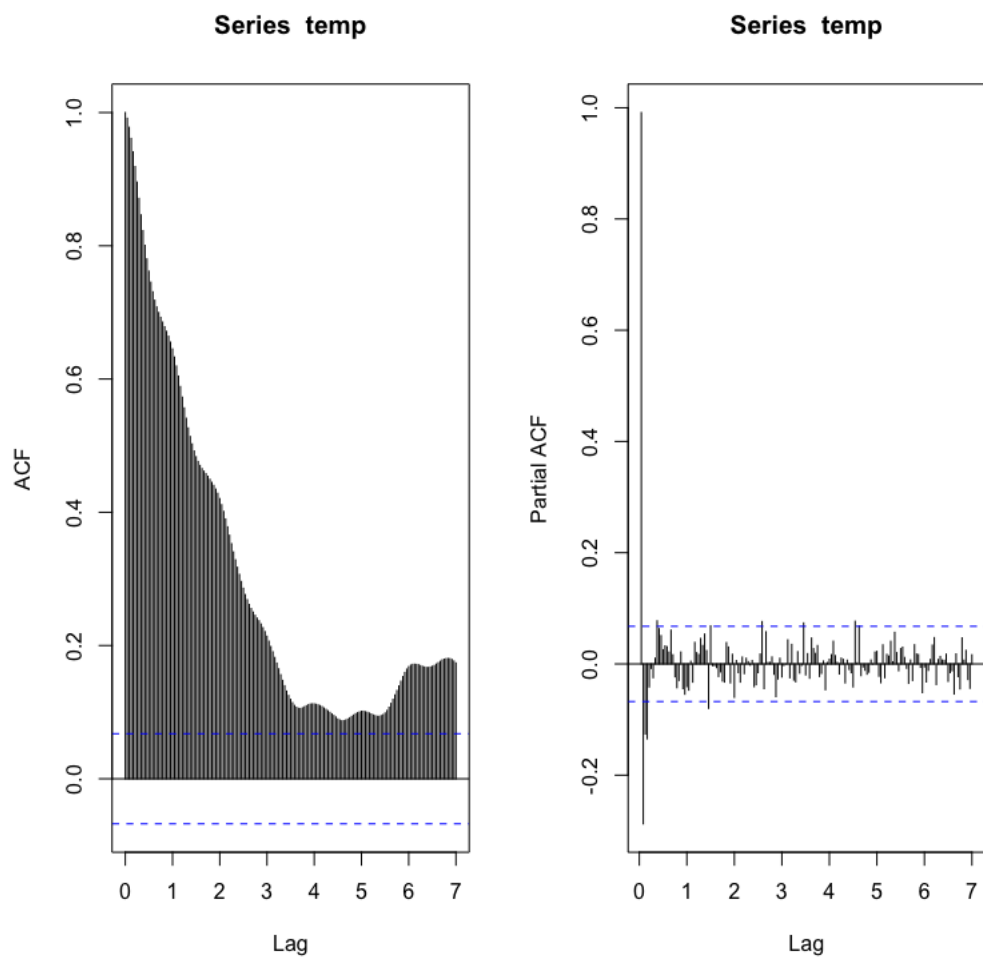


Kuva 2: Kulutus aikasarjan ACF ja PACF kuvat

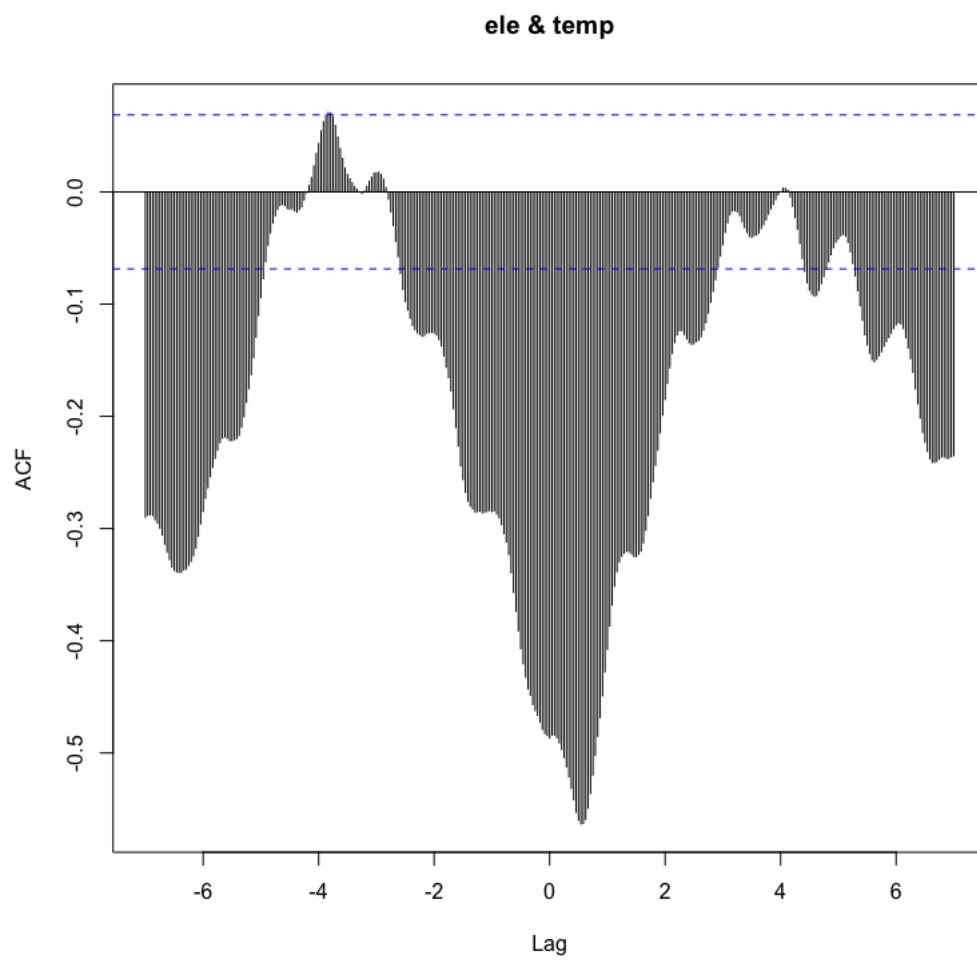
Lämpötila vaihteli vain päivittäin ja laskee koko aikavälillä lineaarisesti kuva 3. Piirrettiin ACF ja PACF kuvat ja huomataan, että mitä lähempänä arvo on nykyistä arvoa, sitä enemmän se vaikuttaa nykyiseen arvoon kuva 4. SARIMAX- mallissa lämpötila on ulkoinen tekijä, joten lähdettiin tutki-
maan ristikorrelaatiota. Ristikorrelaatio kuvajasta nähdään selkeästi, että viiveen arvolla 1 lämpötila ja kulutus korreloi negatiivisesti eli riippuvuus on kääntäen verrannollista. Myös viiveen arvolla 7 ja -7 on hieman korrelaatiota. Mallin estimoinnissa, lämpötilan viive on yksi. Lämpötilan viive voi mahdollisesti johtua ihmisen reaktio nopeudesta. Kun lämpötila muuttuu, ihminen tai laite ei välttämättä heti reagoi siihen vaan reaktio tulee hetken viiveellä perässä.



Kuva 3: Lämpötila



Kuva 4: Lämpötilan ACF ja PACF kuvat



Kuva 5: Kulutuksen ja lämpötilan ristikorrrelaatio funktio

2.1 Mallin rakennus

On saatu selville, millaisen datan kanssa ollaan tekemisissä. Seuraava vaihe on differoiminen. Kokeiltiin differointi eri arvoilla. Differoinnissa on kolme parametria.

$d = 1$ Ei-kausivaihtelu osan differointi,
 $D = 1$ Kausivaihtelu osan differointi,
 $S_{lampotila} = 24$ Kausivaihtelun pituus (lämpötila) ja
 $S_{kulutus} = 24 * 7 = 168$ Kausivaihtelun pituus (kulutus)

Kokeiltiin myös arvoilla $S_{kulutus} = 24$ ja $d = 0$. Differoinnin jälkeen piirettiin ACF ja PACF kuvat.

Keskitytään kulutuksen ACF ja PACF kuviin (kuva 6). ACF:ssä nähdään selkeä piikki lag = 7 kohdalla, voidaan todeta, että SMA osa on nolla ja MA osa on 1. PACF huomataan piikit lag arvoilla 7 ja 14 ja näyttää, että piikit vaimenevat eksponentiaalisesti. Tämä viittaa vahvasti siihen, että SAR ja AR osa eivät ole nollia. AR osa on 1 ja SAR osa saa arvot 1 tai 2.

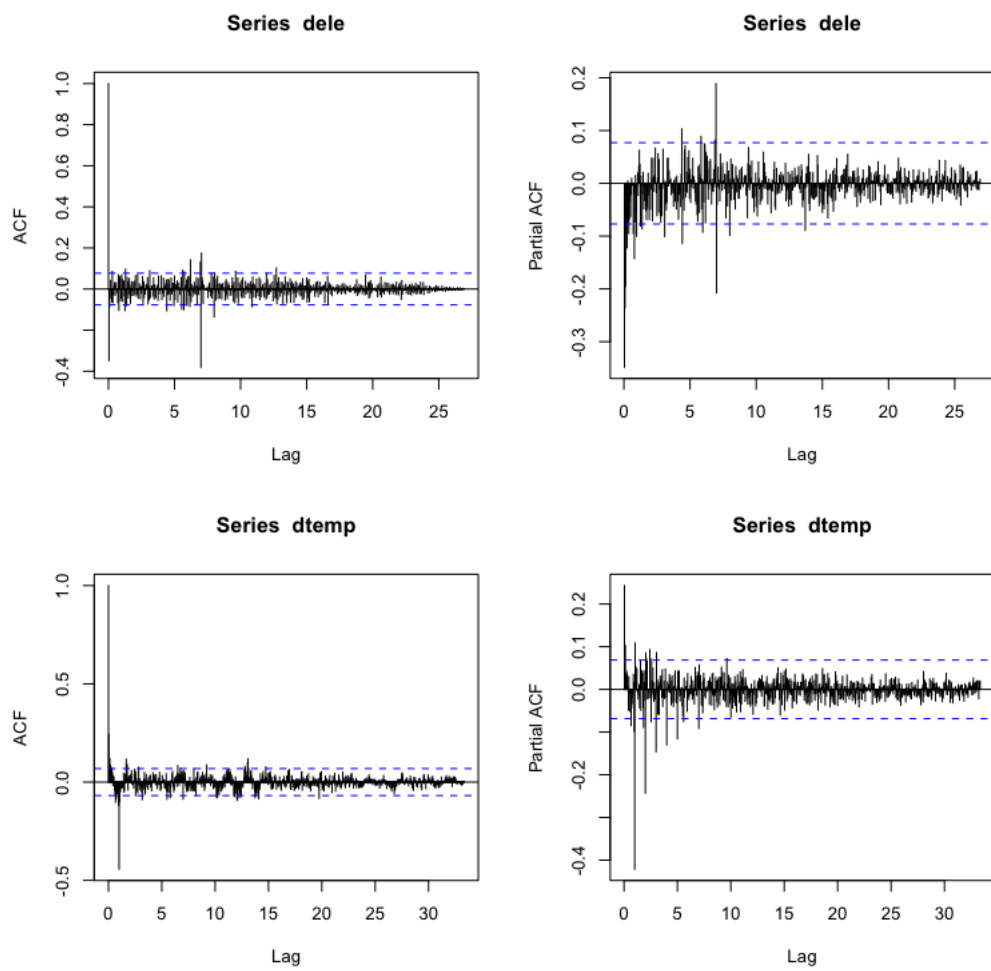
Differoidun lämpötilan ACF ja PACF kuvasta pääteltiin samoja päätelmiä, kun kulutuksen kohdalla, sillä kuvat ovat melkein samanlaiset. Ainoa ero on, että SAR kausivaihtelu lämpötila mallissa on päivittäistä eikä viikoittaista, sillä eksponentiaalinen vaimentuminen tapahtuu lag arvoilla 1, 2, 3, 4, 5, 6 ja 7. Aikasarjan stationaarisuutta testattiin Augmented Dickey–Fuller testillä, jos p-value on pienempi kuin 0.05 niin voidaan hylätä nolla hypoteesi, mikä vihjaa stationaarisuuteen. Kaikki seuraavat mallit läpäisivät tämän testin ja kuvaajista voidaan päätellä, että data on differoinnin jälkeen stationaarista.

Augmented Dickey–Fuller Test

```
data: dtemp
Dickey–Fuller = -7.6085, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Augmented Dickey–Fuller Test

```
data: dele
Dickey–Fuller = -12.298, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary
```



Kuva 6: ACF ja PACF differoitu

Yksi malleista tehtiin Auto.arima komennolla. Auto.arima itse estimoi miten differoidaan aikasarja ja valitsee parametriin niin, että se minimoi residuaalien AIC ja BIC arvot. AIC ja BIC ovat eräänlaisia funktioita, mitkä mittaavat mallin hyvyyden niin, että mitä yksinkertaisempi malli on kysyessä, sen parempi malli se on.

Taulukko 1: Mallit

Malli 1 (auto.arima)	Malli 2	Malli 3
$p = 1$	$p = 1$	$p = 1$
$P = 2$	$P = 2$	$P = 1$
$q = 1$	$q = 1$	$q = 1$
$Q = 0$	$Q = 0$	$Q = 0$
$d = 0$	$d = 1$	$d = 1$
$D = 1$	$D = 1$	$D = 1$
$S_{lampotila} = 24$	$S_{lampotila} = 24$	$S_{lampotila} = 24$
$S_{kulutus} = 24$	$S_{kulutus} = 168$	$S_{kulutus} = 168$

2.2 Mallin diagnostiikka

Näin meillä on kolme vahvaa mallia. Näistä pitää valita yksi. Mallin valinnassa käytettiin Box-Ljung testiä. Testi testaa onko residuaalien välillä korrelaatiota. Residuaalien täytyy olla riippumattomia toisistaan ja jotta ne ovat valkoista kohinaa täytyy residuaalien olla, myös normaalijakaantuneita.

Box-Ljung χ^2 testin testisuure on,

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}, \quad (1)$$

missä n on otoksen määrä, $\hat{\rho}$ on estimoinnin pearson korrelaatio ja k viive arvo .

Box testin hypoteesit ovat seuraavat.

H_0 = Residuaalit ovat riippumattomia ja

H_1 = Residuaalit korreloivat, eli meidän mallin selittävä osa ei selitä kaikkea selitettävän muuttujan käytöstä.

Testisuureen viive arvo oli kaikilla 14.

Box-Ljung test

```
data:  malli1 residuals
```

```
X-squared = 75.481, df = 10, p-value = 3.835e-12
```

```
data:  malli2 residuals
```

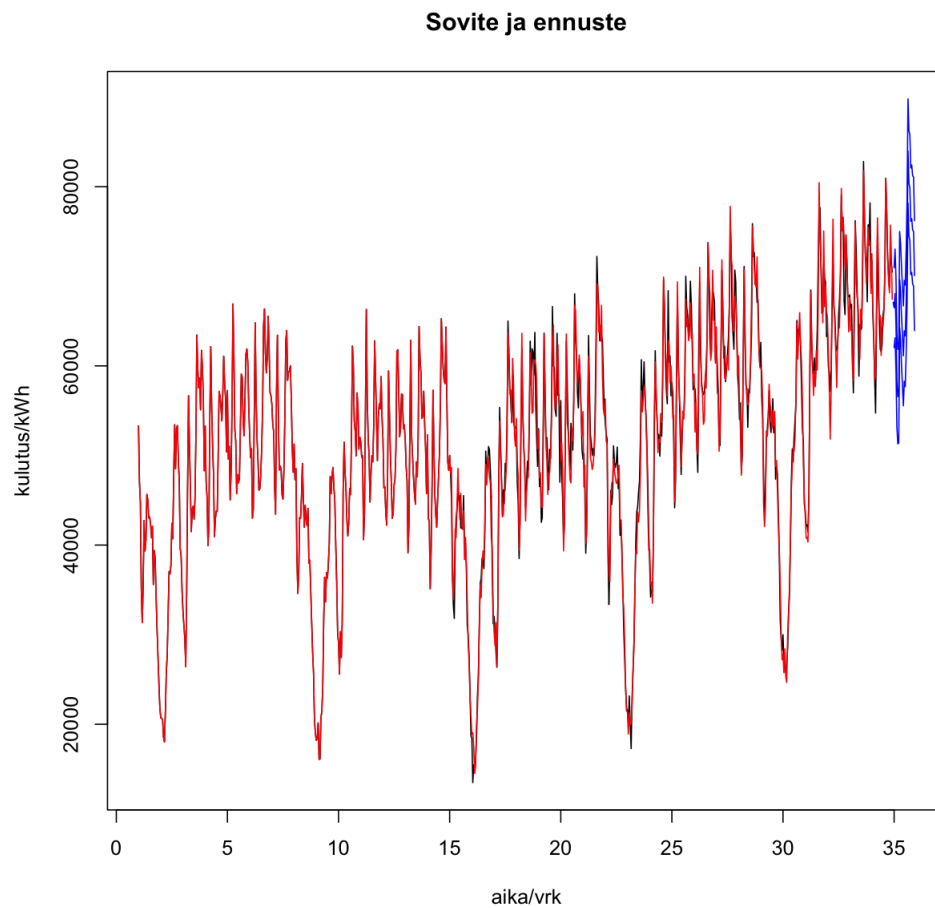
```
X-squared = 20.321, df = 10, p-value = 0.02636
```

```
data:  malli3 residuals
```

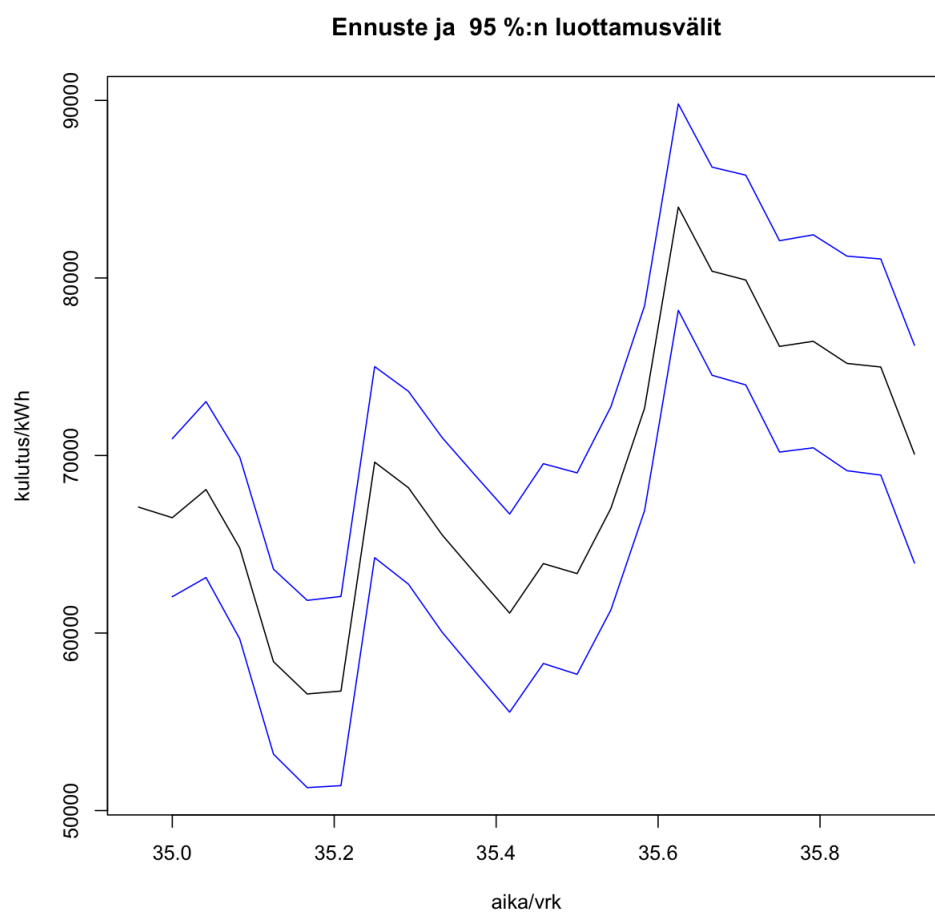
```
X-squared = 18.573, df = 11, p-value = 0.06919
```

Jos Box testin p-value on suurempi kuin 0.05 niin ei ole tarpeeksi todisteita nollahypoteesin hylkäämiseen. Näistä voidaan hylätä malli 1 ja malli

2. Näin meille jäi Malli 3 mitä lähdettiin ennustamaan. Mallia estimoitiin SARIMAX-mallilla. Mallia ennustettiin 24 tuntia eteenpäin ja tulokset näyttävät järkeviltä. Malli 3:n Ex-post saa oikean kulutuksen kanssa lähestulkoon samat arvot (kuva 7)



Kuva 7: Sovite ja ennuste



Kuva 8: Ennuste ja 95 %:n luottamusvälit

3 Hankinnan optimointi

Sähköä sai kolmelta eri taholta. Nord Pool Elspot, Nord Pool Meklari ja SVK. Nord pool Elspotille pystyi myös myymään sähköä. Lähdettiin vertaamaan hintoja keskenään. Suunnitelma oli, jos hinta on pienempi kuin Elspotin myynti hinta niin ostettiin maksimi määrä sähköä tältä taholta ja myytiin ylimääräinen pois Elspotille. Pohdittiin myös Meklariilta ostamista. Meklariilla on halpa hinta mutta blokkituotteen ostaminen aiheutti sen, että voitot eivät korvanneet 2000€ sopimuspreemiota, joten Meklariilta ei ostettu ollenkaan sähköä.

SVK:lla oli kaksi eri hinta luokka. Jos ostaa alle 40000 kWh niin saa halvemmalla ja yli menevästä maksetaan huippuhintaa. Samalla tavalle edettiin eli verrattiin hintoja keskenään.

	est vs alle 40	mekla vs est	est vs alle 70
0:00	5,85	6	6,03
1:00	5,85	6	6,03
2:00	5,85	5,9	5,9
3:00	5,85	5,9	5,9
4:00	5,85	5,85	5,85
5:00	5,85	6	6
6:00	5,85	6	6,03
7:00	9,84	9,84	9,84
8:00	10,08	10	10,23
9:00	10,08	10	10,23
10:00	10,08	10	10,23
11:00	10,08	10	10,23
12:00	9,99	9,99	9,99
13:00	8,98	8,98	8,98
14:00	8,63	8,63	8,63
15:00	9,18	9,18	9,18
16:00	9,59	9,59	9,59
17:00	9,69	9,69	9,69
18:00	9,54	9,54	9,54
19:00	9,49	9,49	9,49
20:00	9,59	9,59	9,59
21:00	9,13	9,13	9,13
22:00	5,85	6	6,03
23:00	5,85	6	6,03

Kuva 9: Hintojen vertailu

Kuva 9 nähdään milloin kannattavampaa myydä Estpolille. Keltaiset solut indikoivat, että kannattaa ostaa maksimi määrä ja myydä ylimääräinen pois. Tällä tavalla saatiin laskettua paljon ja keneltä ostetaan. Meklarin eliminomisessa verrattiin kahta hankintastrategiaa, joko ostettiin meklarilta x määrä tai ei ostettu meklarilta ollenkaan ja ostettiin SVK:lta ja Estpotilta. Meklarilta ostaminen maksimissaan säästi noin 1700€. Koska sopimuspremio oli suurempi kuin tämä, niin päästiin halvemmalla ostamalla vain SVK:lta ja Estpotilta.

4 Parannusehdotukset

Palataan sähkönkulutus ja lämpötila kuvaajaan. Huomaamme, että loppu päässä kulutus nousee. Tämä tapahtuu äkillisesti kuva 1. Myös lämpötila lähtee jyrkkään laskuun samaan aikaan. Koska lämpötilan kausivaihtelu on yksi vuosi, niin alle vuoden pituisella datalla on mahdotonta mallintaa sitä täydellisesti. Tästä huolimatta on olemassa keinoja ilman, että tarvitsee kerätä lisää dataa, kuten esimerkiksi vaihtoregressiomalli. Kyseisessä mallissa voidaan käyttää eri aikasarjoja tietyillä ajanhetkillä. Toteutus tapahtuisi niin, että kun kulutus ja lämpötila on tasaista, käytetään yhtä mallia ja heti muutoksen tapahtuessa vaihdetaan toiseen malliin.

Optimoinnissa voitaisiin muodostaa lineaarinen ohjelmointi ja käyttää esimerkiksi Simplex algoritmia ratkaisemaan optimi osto- ja myyntiarvo. Ennustaminen oli tässä tapauksessa vielä mahdollista, koska 95-luottamusvälit olivat suhteellisen pieniä. Jos haluaa ennustaa pitkälle, niin täytyy parantaa mallia käyttämällä vaihtoregressiomallia tai tasaisen siirtymän autoregressiivista mallia (STAR). Vaihtoregressio malli on erikoistapaus STAR mallista, missä vaihto ei tapahdu tasaisesti vaan heti. Myös kuvassa 1 näkyy että viikonloppuisen kulutus on vähimmillään, joka saattaa johtua siitä, että ihmiset ovat vapailla ja yritykset ovat kiinni. Jos haluaa ennustaa keskiviikon sijaan lauantai tai sunnuntai, täytyy katsoa nykyistä arvoa ja viime viikonlopun arvot ja verrata niitä keskenään.

5 Yhteenveto

Raportissa käytiin läpi, miten ja mistä malli rakennettiin. Myös avattiin miten hypoteesi testit toimivat ja miten voidaan valita eri malleista se paras.

Sähkönkulutuksen ennustamisen ja hankinnan optimoinnin raportoinnissa rakennettiin SARIMAX- malli. Mallin rakentamiseen käytettiin annetun sähkönkulutuksen sekä lämpötila datan aikasarjan analysointia identifioimaan mallin rakennetta ja parametreja. Mallin rakennuksessa oli tärkeää, että tutkitaan diagnostiikkaa ja katsoa monta vaihtoehtoista mallin parametreja parhaimman mallin arvoimiseksi, mikä vaati eri näkökulmia ja testauksia. Mallin rakennuksessa auttoi R ohjelman kirjastot, jotka sisältävät laaja valikoima eri funktioita. Työohjeessa ja kurssilla jaetut esimerkki pohjat R-komennoille olivat suuri apu mallinnus prosessia, sillä mallinnus ja kuvien piirrokset tapahtui R-ohjelman kautta. Koska data ei ollu stationaarinen, piti se differoida. Suoritettuumme Box-Jenkins-menetelmän avulla aikasarja analyysia saatiin muutama vaihtoehto kausivaihtelu pituuden määrittelymiseen. Koska data ei ollu stationaarinen, piti se differoida. ACF ja PACF avulla laaditut kuvat mahdollistivat kertaluvun parametrien valinnat. Differoinnin jälkeen testattiin mallin hyvyttä eri parametrien arvoilla, joista saatiin kolme hyvää mallia. Näistä kolmesta valittiin Box-Ljung χ^2 testin perusteella paras, mikä paljastui olevan kolmas malli. Ennustaminen tapahtui predict-funktion avulla. Näin saatiin ennuste ja optimointiin käytettiin exceliä.

6 Liitteet

```
# Asetetaan työhakemisto SET Working Directory
# Muokkaa tarvittaessa oman työhakemistosi polku
# RStudioissa tämä voi tehdä myös valikosta Session->Set working directory

# Luetaan sähkönkulutus- ja lämpötiladata, hypothesis headerin yli
eletemp = read.table(file = "sahko.csv",
                     sep = ";",
                     dec = ",",
                     skip = 1,
                     col.names = c('kWh', 'Celcius'))

# Sähkönkulutus aikasarjaksi
ele = ts(eletemp$kWh[1:816], start = 1, frequency = 24)

# Lämpötila kahdeksi aikasarjaksi: 816 ensimmäistä havaintoa kytetty
# ja 24 viimeistä havaintoa ennustamiseen.
temp = ts(eletemp$Celcius, start = 1, frequency = 24)
temp816 = ts(eletemp$Celcius[1:816], start = 1, frequency = 24)
# start parametrina vektori: 817. tunti = 35. päivä ensimmäinen tunti
temp24 = ts(eletemp$Celcius[817:840], start = c(35,1), frequency = 24)

# Plotataan aikasarjat
ts.plot(ele,
        xlab = "aika/vrk",
        ylab = "kulutus/kWh")
ts.plot(temp816, temp24,
        xlab = "aika/vrk",
        ylab = expression(~degree~C),
        col = c("black", "blue"))

# Meritellään 2x2 plottausruudukko.
par(mfrow=c(2,2))

# Plotataan autokorrelaatio-, osittaisautokorrelaatio- ja ristikorrelaatio
acf(ele, lag.max=168)
acf(ele, lag.max=168, type = "partial")
```

```

acf(temp, lag.max=168)
acf(temp, lag.max=168, type = "partial")
# Piirretään ristikorrelaatiofunktio omaan kuvaan
par(mfrow=c(1,1))
ccf(ele, temp, lag.max=168)

# Stationarisoidaan aikasarjat. Määrittele parametrit d,S,D
d = 1 # Differoinnin kertaluku d
S = 24 # Kausidifferoinnin jakso S
D = 1 # Kausidifferensoinnin kertaluku D
S2 = 168 # viikottainen vaihtelu

dele = ele
dtemp = temp816
if (d > 0) {
  dele = diff(dele, lag = 1, differences = d)
  dtemp = diff(dtemp, lag = 1, differences = d)
}
if (D > 0) {
  dele = diff(dele, lag = S2, differences = D)
  dtemp = diff(dtemp, lag = S, differences = D)
}

# Differoitujen aikasarjojen autokorrelaatio-, osittaisautokorrelaatio-
acf(dele, lag.max=168)
acf(dele, lag.max=168, type = "partial")
acf(dtemp, lag.max=168)
acf(dtemp, lag.max=168, type = "partial")
ccf(dele, dtemp, lag.max=168)

adf.test(dtemp)
adf.test(dele)

'now we can assume that our time series is stationary'

# Estimoidaan malli ja lasketaan ennusteet ilman ulkoista muuttujaa.
p = 1
q = 1

```

```

P = 1
Q = 0
malli = arima(ele ,
               order = c(p,d,q) ,
               seasonal = list(order = c(P, D, Q), period = S2) ,
               method = "CSS")
enne = predict(malli , n.ahead = 24)

# Estimoidaan malli l mp tilan kanssa. M rit l mp tilan mahd
L = 1

tempestimointi = eletemp$Celcius[1:(816-L)]
tempennuste = eletemp$Celcius[(816-L+1):(816-L+24)]
eleestimointi = ts(eletemp$kWh[(1+L):816] , start =1, frequency = 24)
malli2 = arima(eleestimointi ,
               order = c(p,d,q) ,
               seasonal = list(order = c(P, D, Q) , period = S2) ,
               xreg = tempestimointi ,
               method = "CSS")

summary(malli2)

enne2 = predict(malli2 ,
                n.ahead = 24,
                newxreg = tempennuste)

# Esimerkki Portmanteau-testist . Onko residuaaliaikasarjaluossa nol
Box.test(malli2$residuals ,
         lag = 14,
         type = "Ljung-Box" ,
         fitdf = p + q + P + Q)

# Palautetaan plottaus normaaliin 1x1 ruutuun
par(mfrow=c(1,1))

# Plotataan kuva s hk nkulutusaikasarjasta , mallin (2) sovitteesta ,
# ennusteesta ja ennusteen 95 %:n eli 1,96*sigma-luottamusv leist .
ts.plot(eleestimointi ,
        eleestimointi - malli2$residuals ,

```

```

enne2$pred,
enne2$pred + 1.96*enne2$se,
enne2$pred - 1.96*enne2$se,
ylab = "kulutus/kWh",
xlab = "aika/vrk",
col = c("black", "red", "blue", "blue", "blue"),
main = "Sovite ja ennuste")

# Plotataan kuva pelk st ennusteesta.
ts.plot(enne2$pred,
        enne2$pred + 1.96*enne2$se,
        enne2$pred - 1.96*enne2$se,
        ylab = "kulutus/kWh",
        xlab = "aika/vrk",
        col = c("black", "blue", "blue"),
        main = "Ennuste ja 95 %:n luottamusv lit")

# Kirjoitetaan ennuste ja luottamusv lit .csv-tiedostoon, jonka voi a
output = cbind(enne2$pred,
               enne2$pred + 1.96*enne2$se,
               enne2$pred - 1.96*enne2$se)
write.csv2(output, file = 'ennuste.csv')

```