



GÖTEBORGS UNIVERSITET
HANDELSHÖGSKOLAN

Energikonsumtionsprediktion med expanding window approach

- En undersökning med syfte att skapa dimensioneringsunderlag för framtida elnätutbyggnader, i samarbete med Kungälv energi

Tim Emanuelsson

Kandidatuppsats i Statistik

Institutionen för nationalekonomi med statistik, Göteborgs universitet

Handledare: Mattias Sundén

Förord

Jag vill rikta ett stort tack till min handledare Mattias Sundén som under dessa distanstider alltid är snabb på att svara på mejl och skapa tid för möten. Utöver detta har Mattias även generöst delat med sig av sina erfarenheter av tidsserieanalyser. Studiens resultat hade inte blivit lika tillfredsställande utan Mattias vägledning. Vill även tacka mina kurskamrater som bidragit till intressanta diskussioner, bra opposition och trevliga pratstunder. Slutligen vill jag rikta ett stort tack till Kungälv energi som möjliggjorde uppsatsen. Det har varit en fröjd att samarbeta med er, uppskattar engagemanget och resurserna som ålagts för att kunna genomföra studien. Ett speciellt tack till min handledare på Kungälv energi, Björn Jansson som varit ett stöd under hela processen och bidragit med expertkunskap i ämnet.

Abstract

The impact of electric or hybrid vehicles and photovoltaic systems of a households energy consumption is a growing problem for energy grid owners. As the two increase in popularity the variance of energy consumption between two households in the same area is growing. This generates a major issue for grid owners as they are planning grid connections to future residential areas. To be able to cope with the growing changes in energy consumption the grid owners are dependent on predictions to ensure that the grid can handle the loads generated by the new residential area. In this thesis predictions are generated for the mean peak daily energy consumption, also the difference between households with and without electric vehicles.

The predictions are generated with an expanding window approach and aims to conclude if exogenous variables for the number of electric vehicles and/or mean photovoltaic energy generation improve the predictive power of an ARIMA-based model with only temperature as the exogenous variable. This thesis also aims to examine if there are any difference in mean peak daily energy consumption between households that has photovoltaic system, electric vehicle or neither and if there is a significant difference predicting it. This thesis can not prove that the introduction of variables for number of electric vehicles and/or mean photovoltaic energy generation improve the predictions compared to only using temperature as the exogenous variable. The difference between households with and without electric vehicles is shown to have a increasing trend, this was also predicted with expanding window approach and reached reasonable results.

Förkortningar

| | |
|--------------|--|
| ACF | Autocorrelation function |
| ANN | Artificial neural network |
| SVM | Support vector machine |
| ARIMA | Autoregressive integrated moving average |
| MAPE | Mean absolute percentage error |
| AIC | Akaike information criterion |
| OLS | Ordinary least squares |
| ML | Maximum likelihood |

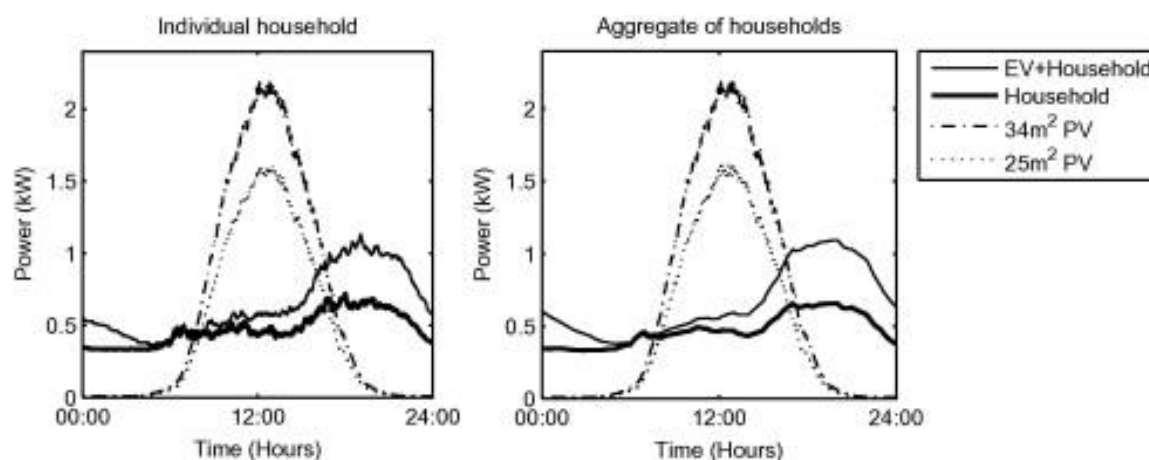
Innehållsförteckning

| | |
|--|----|
| 1. INTRODUKTION | 1 |
| 2. BAKGRUND | 3 |
| 3. DATA | 5 |
| 4. TIDIGARE RESULTAT..... | 8 |
| 5. METOD | 9 |
| 5.1 ICKE-STATIONÄRA OCH SÄSONGSMÄSSIGA TIDSSERIEMODELLER | 10 |
| 5.1.1 AR..... | 10 |
| 5.1.2 MA..... | 11 |
| 5.1.3 ARIMA..... | 11 |
| 5.1.4 SARIMA..... | 12 |
| 5.1.5 SARIMAX..... | 12 |
| 5.2 PARAMETERSKATTNINGAR..... | 13 |
| 5.2.1 Maximum likelihood-estimering..... | 14 |
| 5.3 EXPANDING WINDOW | 15 |
| 5.4 AVVIKELSEMÅTT | 15 |
| 6. RESULTAT | 16 |
| 6.1 FÖRBRUKNINGSPREDIKTIONER | 17 |
| 6.2 ENERGIKONSUMPTIONS DIFFERENSER | 20 |
| 7. DISKUSSION | 24 |
| 7.1 SUMMERING AV RESULTAT..... | 24 |
| 7.2 UTVÄRDERING AV RESULTAT | 24 |
| 7.3 UTVÄRDERING AV METOD SAMT JÄMFÖRELSE MED TIDIGARE STUDIER..... | 25 |
| 7.4 FÖRSLAG TILL FORTSATTA STUDIER | 26 |
| BILAGOR..... | 29 |
| B.1 ÖVRIGA FIGURER..... | 29 |
| B.2 ENKÄTUNDERSÖKNING..... | 31 |
| B.3 FÖRSTORADE PREDIKTIONSRISULTAT..... | 32 |

1. Introduktion

Hur kommer framtidens energibehov se ut? Ökad miljömedvetenhet och satsningar på nya tekniska lösningar har snabbt förändrat det generella konsumtionsbeteendet. Uppvärmningssystem, bilar med eldrift och solenergianläggningar skapar stora variationer i förbrukning. Förändringen är främst tydliga för villor eftersom möjligheten att installera denna typ av förbrukningsförändrande produkter sker på kort tid och ger stor effekt på förbrukningsmönstret och energikonsumtionen. Skillnaden i energiförbrukning mellan två intilliggande villor kan idag vara betydligt större än tidigare. Denna förändring leder till problem för elnätsägare som behöver dimensionera elledningar som skall hantera de svårförutsägbara lasterna.

Solenergianläggningar och el- eller hybridbilar har blivit tillgängliga för fler i takt med att produkternas pris sjunkit. Dessa produkter har en stor enskild inverkan på en villas energikonsumtion och dess belastning av elnätet.



Figur 1. Solenergianläggningar och elbilspåverkan på elnätslasten (Munkhammar, Grahn, & Widén, 2013).

I Figur 1 visas olika dagliga belastningar på elnätet beroende på om ett hushåll har elbil, solcellsanläggning eller inget av alternativen. Resultaten kommer från en svensk undersökning som granskat effekterna av elbil och solceller på en villas elnätsbelastning. Figur 1 understryker komplexiteten och framhäver problemet med att prediktera energikonsumtionen för en villa. Det är stora variationer mellan villor som har solenergianläggning och/eller elbil jämfört med dem som inte har. Exempelvis visar undersökning att elbilen ökar den totala

energiförbrukningen med 37% (Munkhammar, et.al. 2013). Laddning av el- eller hybridbil sker ofta under kvällen då hushåll generellt har sin konsumtionstopp. Det är även då efterfrågan på el är som störst och andelen förnyelsebara energikällor i elmixen minskar. Eftersom mycket energi behöver produceras under dessa timmar blir de energikällor som utefter efterfrågan kan öka sin produktion viktiga för att motsvara energibehovet. Detta för oss vidare till andra incitament för att förutspå energikonsumtionsbeteenden.

Elnätsägare är inte ensamma intressenter för energikonsumtionsprediktioner. Precisa prognoser är ett viktigt element för att nå länders energi och klimatmål. För att kunna nå målen är det viktigt att den förnyelsebara energin som produceras tas tillvara. Detta har genererat i mycket forskning kring tekniskt avancerade applikationer som ”smartgrids”. Syftet med dessa är att produkter där konstant elförsörjning inte är nödvändigt anpassar sig efter energitillgången. Låt säga att vindstyrkor tilltar efter den generella konsumtionstoppen under kvällen, med smarta applikationer tas detta i beaktning och laddningen av elbilen kan då avvakta tills elproduktionen ökar. Detta leder till att lasten från elbilsladdningen inte adderas till den generella konsumtionstoppen men även till en billigare laddning samt en möjlighet att ta vara på all den förnyelsebara energin som produceras.

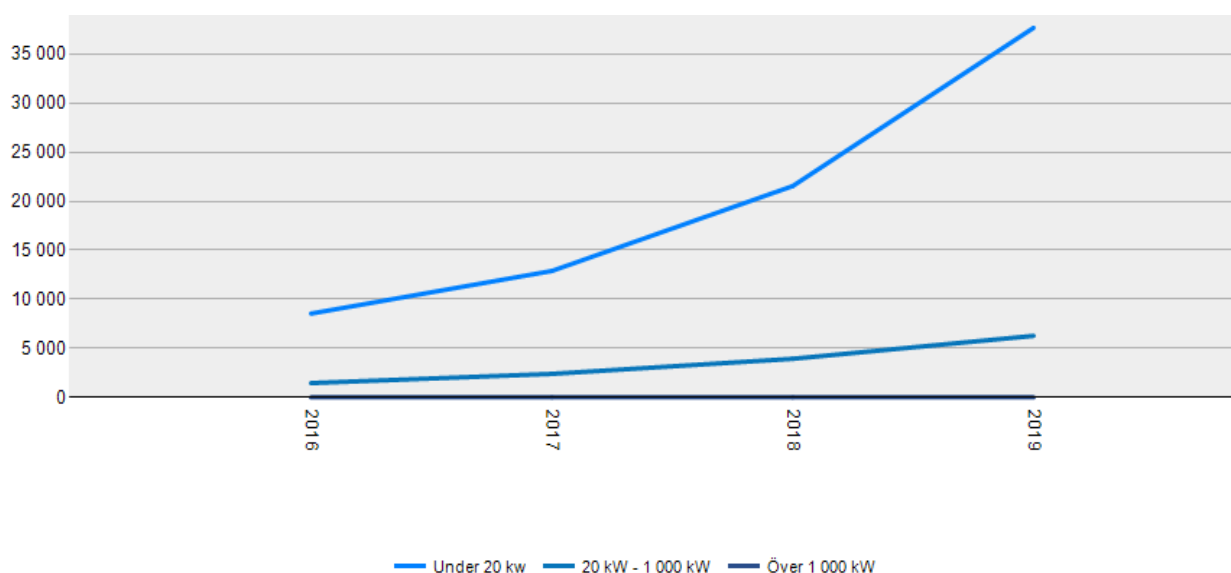
Syftet med denna studie är att skapa underlag för dimensionering av elnätsutbyggnader med prediktionsmodeller för villors energikonsumtion. Genom att fokusera på framtida utmaningar som elbilsladdning och solenergianläggningar är förhoppningen att kunna fånga och förklara den variation i elnätslast som de två faktorer adderar. Bra prediktioner är viktigt för att undgå dyra utbyggnader av elnätet i framtiden. Därav ligger studiens fokus i att prediktera för variabler som i hög grad bestämmer den framtida villaägarens konsumtion. Idag är det vanligt att se till hur stor last en kund kommer generera vid en viss temperatur. För att hitta en modell som kan prediktera framtida elnätslaster behövs även bilar med eldrift och solenergianläggningars påverkan tas med i beräkningen. Frågeställningarna för studien lyder därav:

- Predikterar en modell som har variabler för el- eller hybridbil och/eller solceller bättre än en modell med enbart temperatur?
- Om hushåll med el- eller hybridbil, solceller eller ingendera separeras, syns någon differens? Går denna att prediktera med god precision?

2. Bakgrund

I detta avsnitt kommer bakgrundsinformation till behovet av att se effekterna av solceller och el- eller hybridbilars påverkan på energikonsumtionen. Vikten av dessa variabler kommer belysas för att kunna prediktera elnätsbelastningen för framtidens villaägare.

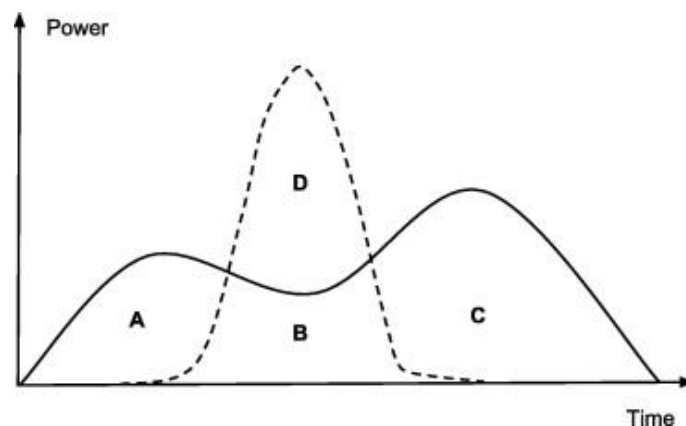
Antalet nätanslutna solcellsanläggningar har under de senaste tre åren ökat i Sverige. Störst ökning har skett för anläggningarna under 20 kW som visas i Figur 2. Denna anläggningsstorlek används oftast av privatpersoner och framförallt villaägare. I och med den minskade kostnaden per installerad kW för solcellsmoduler har tillgängligheten ökat och blivit ett allt vanligare tillägg för nybyggnationer eller i samband med omläggning av tak. Eftersom solceller i många fall kan ersätta annat takmaterial som tegelpannor kan den relativa kostnaden för installation av solceller reduceras (Energimyndigheten, 2019).



Figur 2. Antalet installerade solenergianläggningar per år (Energimyndigheten, 2020).

Den genererade solelen som inte förbrukas skickas ut till elnätet. Detta skapar energiflöden i elnätet som kan vara svåra att förutspå i planeringsstadiet av ett villaområde eftersom det är svårt att anta hur stor andel av villorna som kommer installera solenergianläggningar. En variabel som begränsar prognosen för solceller är dels hustakets area samt att den största intäkten som genereras av solceller är användning av den egenproducerade elen (Stridh, Larsson, 2017). Eftersom försäljningspriset av överskottselen är lägre än det pris som betalas för att köpa elen blir återbetalningstiden för solcellsanläggningen kortare om den genererar så litet överskott som möjligt. I Figur 3 illustreras den genererade solelen med den streckade linjen

och energikonsumtionen med den heldragna. Ur ett ekonomiskt perspektiv ska B vara så stor andel av summan av B och D som möjligt.



Figur 3. Typiska solenergiproducering- och energikonsumtionskurvor (Munkhammar, et.al. 2013).

För att minska de påfrestningarna som en överproducerande solcellsanläggning genererar kan en ökad egen konsumtion vara en lösning. Om hushållet har en el- eller hybridbil ökar konsumtionen av den egen-genererade solelen men eftersom solcellernas överproduktion och elbilsladdningen sällan sammanfaller påverkar detta enbart elnätslasten marginellt. En elbil ökar även hushållets totala elförbrukning och kan i sin tur leda till ytterligare påfrestningar på elnätet beroende på laddningsrutin. Som nämnts tidigare i studien och visas i Figur 1 ökar elbilsaddning den totala energiförbrukningen med 37% och höjer enbart andelen egenkonsumerade solel med 3% (Munkhammar et.al, 2013).

I Sverige och i andra kalla länder sammanfaller inte solenergiproduktionen och energibehov lika ofta som i varmare länder. I länder där energi används för att kyla ner hus kan större del av solenergiproduktionen användas eftersom då solen står som högst är den tid på dygnet som har högst temperatur (EIA, 2021). En faktor med inverkan på hur stor självkonsumtionen av den producerade energin är husets geografiska läge. Det är därav viktigt att konstatera att denna typ av analys är geografiskt begränsade.

Elnätet behöver kunna hantera ett villaområde där stor andel har solenergianläggningar som alla samtidigt kommer överproducera. Nätet behöver även hantera belastningarna under kvällen som nu belastas av allt fler el- eller hybridbilar som laddas hemma. Vikten av att dimensionera rätt är stor eftersom elnätsägarna måste kunna hantera topparna i efterfrågan samt produktionstopparna från solenergianläggningarna och slippa framtida dyra utbyggnationer av elnätet. Därav är det viktigt att prediktera skillnader som uppstår baserat på dessa två variabler.

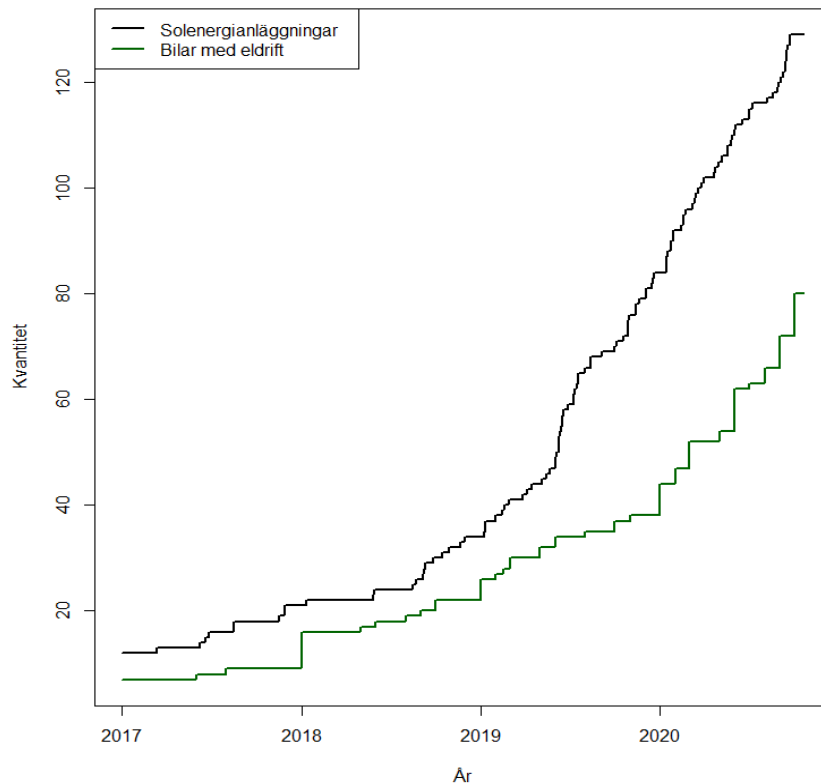
3. Data

Ursprunget av datan som används i analysen har distribuerats av Kungälvs energi. I datan presenteras timvärden för anläggningarnas förbrukning från 2017-01-01 till 2020-10-20. Mätvärdena är presenterade i kWh som avrundas till närmsta heltal för majoriteten av observationerna. Det förekommer även mått med två decimalers precision vilket är en följd av nya mätare som installerats. Anläggningarna har anonymiserats för att inte kunna spåras till de faktiska hushållen. Eftersom förbrukningsinformation kan användas för att förstå beteendemönster skyddas detta under GDPR. Förbrukningsdata kompletteras med lokal temperaturdata som även den framställs i timvärden och distribuerades av Kungälvs energi.

Tillsammans med det ovannämnda datamaterialet kommer även resultatet från en kundundersökning användas för att ytterligare förstå vad som påverkar anläggningarnas förbrukning. I enkäten besvaras frågor om anläggningens uppvärmningssystem, antalet vuxna i hushållet, om de har el- eller hybridbil samt om hushållet har installerade solceller. Enkäten skickades ut till 4 områden där Kungälvs energi distribuerar el. Områdena valdes eftersom dessa har stor andel villor och kunde med enkelhet inringas geografiskt. För att exkludera de anläggningar i områdena som inte var av intresse för undersökningen togs enbart anläggningar med som var registrerade hushåll eller lantbruk. Anläggningar där fakturaadressen skiljer sig från anläggningsadressen exkluderades eftersom det tyder på att detta inte är en primär bostad. Flera anläggningar på samma anslutningspunkt exkluderades eftersom dessa är flerbostadshus. Urvalet slutade med 1026 anläggningar där Kungälv Energi slumpvis valde 1000 anläggningar för utskick av enkäten varav 456 besvarade denna. Frågorna som inkluderades i enkätundersökningen visas i B.2 under bilagor.

I analysen används dagliga medelkonsumtionstoppar, dessa genereras genom att för varje anläggning spara det högsta konsumtionsvärdet för varje dygn. Konsumtionstopparna används sedan för att skapa dagliga medelvärden för att generera en tidsserie. Medelkonsumtionstopparna kommer i studien främst benämnas som toppar i energikonsumtion eller energikonsumtionstoppar. Förklaringsvariablerna som används är antalet el- eller hybridbilar som oftast laddas hemma samt medelvärdet av solenergiproduktionen per dag. Trenden för el- eller hybridbilar och solenergianläggningar i urvalet visar i Figur 4 att antalet el- eller hybridbilar har fördubblats under 2020. Temperaturen används även som en förklaringsvariabel. Eftersom temperaturdatan är angiven för varje timma används ett

medelvärde av de temperaturer som är kopplade till de enskilda hushållens dagliga konsumtionstoppar.

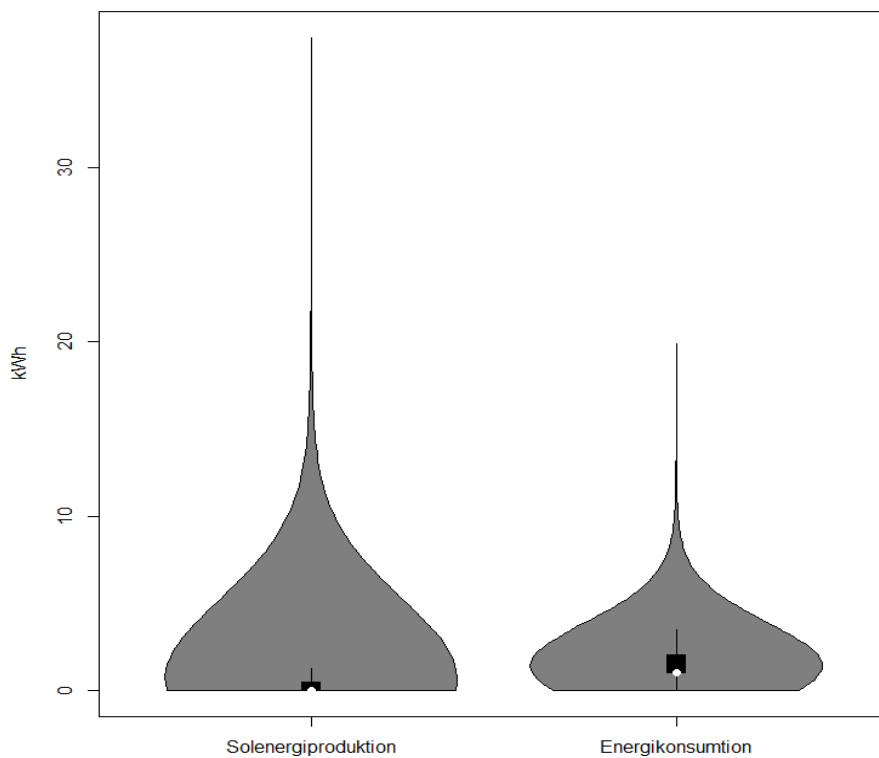


Figur 4. Trend för el- eller hybridbil och solcellsanläggningar från enkätundersökningen.

För att undersöka skillnader mellan hushåll som har el- eller hybridbil, solenergianläggningar och ingendera delas datasett upp baserat på den tid som angetts som inköpsdatum i enkätundersökningen. Om ett hushåll exempelvis skaffar en elbil 2019-10-01 används konsumtionstopparna som föregår datumet i det dataset där hushåll med ingendera placeras. Observationerna efter denna tidpunkt adderas till datamängden för el- eller hybridbil. Om samma hushåll senare skulle skaffa solenergianläggning adderas den data som observeras efter den tidpunkten till den datamängd som har både el- eller hybridbil och solenergianläggning. Detta gjordes för att undvika att den ena variabeln påverkade den andra. Variablerna som används i analysen är antalet el- eller hybridbil samt den absoluta temperaturdifferensen mellan de temperaturer som är kopplade till konsumtionstopparna för hushållen med ingendera jämfört med temperaturerna kopplade till topparna för hushållen med el- eller hybridbil.

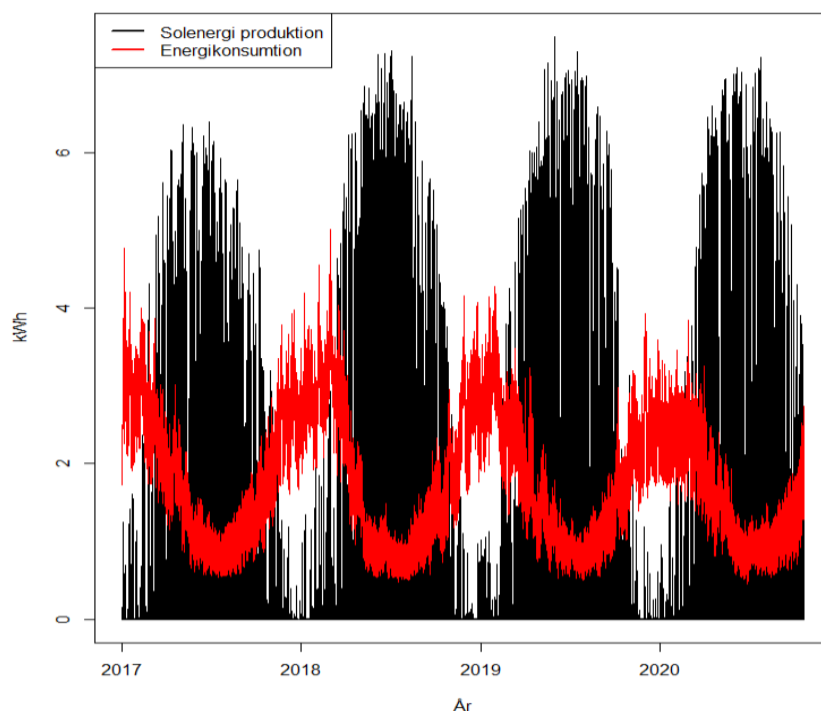
Resterande del av avsnittet ger deskriptiv statistik av datamaterialet. På grund av säsongvariationerna används grafer istället för de vanligt förekommande numeriska värdena

för att bättre visa datans egenskaper. I Figur 5 visas en violinplot över konsumtion- samt produktionsvärdena för solcellsanläggningarna. Violinploten visar andelen av datan som befinner sig vid ett givet värde. Den skeva fördelningen för solenergiproduktionen beror på att de inte producerar något då det är mörkt och lågt om solcellerna inte utsätts för direkt solljus. De höga värdena kommer sannolikt från ett fåtal enskilda solenergianläggningar eftersom anläggningen måste vara stor om den skall producera över 30 kWh. Liknande förklaring kan appliceras på energikonsumtionen. Ett fåtal hushåll har den typen av installationer som kan generera en förbrukning över 15 kWh.



Figur 5. Violinplot som beskriver fördelningen av konsumtionsvärden och produktionsvärden från solcellsanläggningar.

I Figur 6 visas variationen över säsonger, grafen är baserad på dagliga medelvärden för konsumtion som visas av den röda linjen och dagliga medelvärden för energiproduktionen från solceller. Som förklarades i avsnitt 2 producerar solenergianläggningarna ofta mest då konsumtionen är låg vilket görs tydligt i grafen nedan som visar två trender som har sina toppar under motsatta årstider.



Figur 6. Dagliga medelvärden för konsumtion och produktion.

4. Tidigare resultat

I detta avsnitt kommer resultat från tidigare studier på energikonsumtionsprediktioner analyseras. Studierna vars resultat kommer att undersökas har använt ARIMA-baserade modeller för att prediktera energikonsumtion likt denna studie. De tidigare resultaten belyses för att visa andra tillvägagångssätt för hantering och prediktering av energikonsumtionsdata men även för att kunna jämföra med de resultat som nås i denna studie. Resultaten redovisas ofta i branschen med Mean Absolut Percentage Error förkortat som MAPE. Den matematiska förklaringen till MAPE presenteras i avsnitt 5.4.

Tserkezos (1992) skapade en ARIMA-baserad modell för att förutspå hushållens energiefterfrågan i Grekland genom att använda månads- och kvartalsdata insamlade över 15 år. Energidata beskriver i MWh och motsvarade den totala konsumtionen för hushållen. Med hjälp av variabler som elpris och hushållsinkomst kan modellen prediktera energikonsumtionen för de sista 24 månaderna av datainsamlingsperioden. Två modeller testades, en som genererade prediktioner baserat på kvartalsdata och en baserat på månadsdata. Resultaten visar att den modell som applicerats på månadsdata predikterar bättre än den modell som applicerats på kvartalsdata, prediktionskraften i MAPE resulterade i 3.78% respektive 7.69%.

Rallapalli & Ghosh (2012) använde en SARIMA-modell för att prediktera månadsvisa konsumtionstoppar i Indien. Konsumtionsdatan är angiven månadsvis 6 år tillbaka i tiden, de fem första åren används för träning av modellen och det sista året för validering. Studiens syfte är att visa hur en SARIMA-modell presterar bättre än de verktyg som Indiens Central Electricity Authority (CEA) använder idag. Resultatet stödjer det författarna antyder, SARIMA-modellen genererar MAPE mellan 1.59% och 2.05% för de 5 regionerna och är tydligt bättre än de som används i Indien idag.

Hutama, Akbar & Candra (2018) jämförde skillnaderna i prediktiv förmåga mellan en SARIMAX-modell jämfört med en ANN-baserad modell med flera hidden layers (MLP). I studien används dagligen insamlad data från föregående år för att skapa prediktioner för fyramånadersperioden juni – september året därpå för dagliga energikonsumtionstoppar. Resultatet visar att SARIMAX-modellen presterar bättre med en MAPE på 2.4% medan MLP-modellen nådde 2.7%. De belyser även skillnaden i tid för att skapa modellerna. SARIMAX går snabbare att skapa samt presterar bättre i denna studie.

5. Metod

I energibranschen används förbrukningsprofiler, dessa beskriver hur stor last en kund genererar vid en given temperatur. I modellen som används i denna analys kommer temperaturen tillsammans med variabler för el- eller hybridbil och solenergiprodukten användas i ett försök att förbättra prediktioner för hushålls energiförbrukning. Modellerna som används i analysen är ARIMAX och SARIMAX, nedan kommer SARIMAX beskrivas eftersom det enbart är en förlängning av ARIMAX. Modellerna är en kombination av flera, därav kommer de modeller som SARIMAX bygger på förklaras nedan för att förstå hur den fungerar. Expanding window approach är ett sätt att hantera träning- och testdatan där SARIMAX- eller ARIMAX-modellerna sedan appliceras. Expanding window förklaras närmare under avsnitt 5.3.

En ARIMA-baserad modell valdes på grund av att de skapar tillförlitliga konfidensintervall för prediktionerna (Deb et.al, 2017). Det valdes även eftersom ARIMA har visat god prediktiv förmåga utan förklaringsvariabler (Abdel-Aal, Al-Garni. 1997). Eftersom enbart temperaturen och antalet el- eller hybridbilar samt solenergiproduktionen är de tillgängliga variablerna som förändras över tid behövs en modell som kan prediktera väl utan eller med få exogena variabler.

ARIMA-baserade modellers popularitet vid tidsserieprognoser kommer bland annat från dess flexibilitet. Den innehåller flera olika tidsseriemodeller vilket gör att den kan hantera tidsrier väl som kan förklaras med någon av de inkluderade modellerna. Om tidsrier kan förklaras genom enbart Autoregressive, enbart Moving average eller kombinationen av dem båda kan detta hanteras av modellen (Zhang. 2003). I det första delavsnittet förklaras SARIMAX och dess delar, i det andra delavsnittet går vi på djupet i parameterskattningen för att därefter förklara expanding window och slutligen belysa avvikelsemåtten som används i studien.

5.1 Icke-stationära och säsongsmässiga tidsseriemodeller

För att kunna prediktera en tidsrie som är icke-stationär och innehåller säsongsmässiga komponenter används SARIMAX. En SARIMAX-modell består precis som en ARIMA-modell av AR och MA modeller, integrering samt differentiering. Det som skiljer dem åt är att SARIMAX även har Säsongsmässiga AR och MA modeller, en säsongskomponent samt exogena variabler för att kunna prediktera tidsrier med säsongsmässig variation och använda förklaringsvariabler. Då SARIMAX-modellen bygger på flera enskilda modeller för hantering av tidsrier kommer de underliggande modellerna presenteras fram till det att en SARIMAX-modell definierats.

5.1.1 AR

Autoregressiva (AR) modeller bygger på iden att det nuvarande värdet av serien x_t , kan förklaras av de p föregående värdena $x_{t-1}, x_{t-2}, \dots, x_{t-p}$, där p är antalet steg tillbaka i tiden som används för att prediktera det nuvarande värdet. En AR-modell av p laggade termer kan definieras som:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + z_t \quad (1)$$

Där x_t är stationärt, $\phi_1, \phi_2, \dots, \phi_p$ är konstanter ($\phi_p \neq 0$). Vi antar även att z_t är diskret vitt brus med medelvärde 0 och varians σ_z^2 . Att x_t är stationärt innebär att den inte har en trend. Det kan även förklaras som att medelvärdet är konstant över tidsrien. För att förenkla ekvationen används backshift operator B, denna definierar x_{t-p} som $B^p x_t$ och ekvation 1 kan skrivas som

$$\phi_p(B)x_t = z_t \quad (2)$$

Där $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ (Shumway, Stoffer. 2011, s 85).

5.1.2 MA

Som ett alternativ till den autoregressiva representationen i ekvation 2 där x_t på ekvationens vänstra sida antas kombineras linjärt, antar Moving average (MA) modellen av termen q , att det diskreta vita bruset z_t på den högra sidan av ekvation 2 kombineras linjärt för att generera den observerade tidsseriedatan. En MA-modell av termen q kan definieras som

$$x_t = z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_q z_{t-q} \quad (3)$$

Detta kan precis som ekvation 1 förenklas med backshift operator B och ekvation 3 kan skrivas som

$$x_t = \theta_q(B)z_t \quad (4)$$

Där $\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$. Dessa två modeller kombinerat skapar en ARMA(p, q) -modell som kan användas på stationär data (Shumway, Stoffer. 2011, s 90).

5.1.3 ARIMA

För att hantera icke-stationär data används istället en ARIMA-modell. I står för integrated och modellen differentierar tidsserien för att kunna uppnå antagandet om stationaritet. Resultatet integreras tillbaka för att inte skapa prediktioner för den differentierade datan utan på original-datan. Om d är ett positivt heltal är x_t en ARIMA(p, d, q) -process om $y_t = (1 - B)^d x_t$ är en kausal ARMA(p, q) process enligt

$$\phi_p^*(B)x_t = \phi_p(B)(1 - B)^d x_t = \theta_q(B)z_t, \quad \{z_t\} \sim WN(0, \sigma^2) \quad (5)$$

Där $\phi_p(z)$ och $\theta_q(z)$ är polynom av graden p och q och processen är kausal om $\phi_p(z) \neq 0$ för $|z| \leq 1$ (Brockwell, Davis. 2002, s 85). Polynomet $\phi_p^*(z)$ har 0 av graden d vid $z = 1$. Vi har även att x_t är stationär endast då $d = 0$, då den reduceras till en ARMA(p, q) -process. Notera även att om $d \geq 1$, kan vi lägga till en godtycklig polynomtrend av graden $(d - 1)$ till x_t utan att bryta likheterna i ekvation 5 (Brockwell, Davis. 2002, s 180).

5.1.4 SARIMA

Om tidsserien även har en säsongsvariation kan denna hanteras genom att addera en säsongskomponent som differentierar serien av period s . Om vi passar en $ARMA(p, q)$ -modell $\phi_p(B)y_t = \theta_q(B)z_t$ på den differentierade tidsserien $y_t = (1 - B^s)x_t$, då kommer modellen av den ursprungliga tidsserien vara $\phi_p(B)(1 - B^s)x_t = \theta_q(B)z_t$. Detta är ett specialfall av en generell SARIMA-modell som kan förklaras enligt följande. Om d och D är positiva heltal är x_t en $SARIMA(p, d, q)(P, D, Q)_s$ -process av period s om den differentierade tidsserien $y_t = (1 - B)^d(1 - B^s)^D x_t$ förklaras en kausal ARMA-process av

$$\phi_p(B)\Phi_P(B^s)y_t = \theta_q(B)\Theta_Q(B^s)z_t, \quad \{z_t\} \sim WN(0, \sigma^2) \quad (6)$$

Där $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$, och $\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}$. $\Phi_1, \Phi_2 \dots \Phi_P$ är de säsongsmässiga AR koefficienter (SAR) och $\Theta_1, \Theta_2 \dots \Theta_Q$ är koefficienter av den säsongsmässiga MA (SMA) (Brockwell, Davis. 2002, s 203). Delarna $\phi_p(B)$ och $\theta_q(B)$ förklaras efter ekvation 2 respektive 4. Den differentierade serien y_t är kausal endast om $\phi(z) \neq 0$ och $\Phi(z) \neq 0$ för $|z| \leq 1$ (Brockwell, Davis. 2002, s 85).

5.1.5 SARIMAX

SARIMAX möjliggör användning av linjär regression då antagandet om oberoende feltermar inte uppfylls. I prediktioner av tidsserier finns en autokorrelation mellan observationerna och därav även i feltermerna. En SARIMAX-modell kan ses som en linjär regression med SARIMA-fel. SARIMA-modellen appliceras alltså på den linjära regressionens feltermar för att hantera den autokorrelation som uppstår. Detta kan förklaras som

$$x_t = \mathbf{u}'_t \beta + w_t \quad (7)$$

Där x_t är observationerna vid tiden $t = (1, 2, \dots, n)$, \mathbf{u}'_t innehåller värdena av den exogena variabeln vid tiden t och β innehåller regressionskoefficienten. Slutligen w_t är värdena från en kausal SARIMA $(p, d, q)(P, D, Q)_s$ -process av period s som uppfyller

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D w_t = \theta_q(B)\Theta_Q(B^s)z_t, \quad \{z_t\} \sim WN(0, \sigma^2) \quad (8)$$

(Brockwell, Davis. 2002, s 211). Parametrarna i ekvation 8 beskrivs i tidigare avsnitt.

5.2 Parameterskattningar

Vid sökandet efter den optimala SARIMA-modellen användes funktionen ”auto.arima”. Funktionen kan hantera såväl ARIMA som SARIMAX-modeller och är ett automatiserat sätt att hitta de optimala parametrarna för (p, d, q) (P, D, Q) och skatta dess koefficienter. Detta sker genom att modellen med lägst AIC är den som används (Hyndman, Khandakar, 2008). Akaike information criterion (AIC) används enligt följande för att få fram parametrarna p, q, P och Q och dess koefficienter givet att d och D är kända:

$$AIC = -2\log(L) + 2(p + q + P + Q + k) \quad (9)$$

Där k är antalet exogena variabler samt ett eventuellt intercept. Beteckningen L är maximum likelihood-funktionen av modellen passad till den differentierade datan $(1 - B)^d(1 - B^s)^D x_t$. Därav är likelihood för hela modellen x_t inte definierad vilket leder till att modeller med olika differentiering inte är jämförbara (Hyndman, Khandakar, 2008). Detta eftersom parameterskattning av D och d genom minimering av AIC tenderar att leda till överdifferentierade parametrar. Överdifferentiering leder i sin tur till sämre prediktionsförmåga och vidare prediktionsintervall (Smith, Yadav. 1994).

För att skatta parametrarna D och d används därför unit-root tests. Unit-root test är test för stationaritet. Om modellen är en ARIMA används KPSS unit-root test där nollhypotesen innebär att tidsserien är stationär såtillvida att tidsseriens trend kan tas bort och kvar blir en stationär tidsserie, trenden behöver inte vara linjär. KPSS-test är därmed mindre sannolik att överdifferentiera jämfört med exempelvis Dickey-Fullers test som har en nollhypotes där en unit-root existerar och tidsserien alltså inte är stationär. KPSS testet görs för olika värden på d med start på $d = 0$ fram tills testet blir insignifikant (Hyndman, Khandakar, 2008). Testet kan beskrivas som:

$$a_t = \xi t + r_t + \varepsilon_t \quad (10)$$

Där r_t är en random walk:

$$r_t = r_{t-1} + m_t \quad (11)$$

Där m_t är oberoende och likafördelad $(0, \sigma_m^2)$. Nollhypotesen innebär att $\sigma_m^2 = 0$, detta eftersom ε_t antas vara stationär och därav är a_t stationär runt trenden. Det finns även specialfall av ekvation 9 där vi sätter $\xi = 0$, i detta fall är nollhypotesen att a_t är stationär vid en nivå (r_0)

snarare än runt en trend (Kwiatkowski, Phillips, Schmidt, Shin. 1992). Om tidsserien har säsongsmässig variation och en SARIMA-modell används görs först ett Canova-Hansen test som likt KPSS är ett test för stationaritet för att bestämma den säsongsmässiga differentieringsparametern D . Därefter används KPSS för att ta fram den bästa parametern d (Hyndman, Khandakar, 2008). Om differentieringen är korrekt utförd görs det tydligt i ACF-ploten. ACF-ploten visar autokorrelationen vid olika lagg av tidsserien. Om det finns en unit-root avtar ACF linjärt och sakta. Om tidsserien däremot är stationär avtar ACF fort (Brockwell, Davis. 2002, s 182).

5.2.1 Maximum likelihood-estimering

Maximum likelihood-funktionen som i ekvation 9 benämns som L kan utöver att estimeras regressionskoefficienten β samt parameterskattningarna av ϕ, Φ, θ och Θ även användas för att välja parametrarna p, q, P och Q givet att d och D är kända. Detta tillvägagångssätt används i denna studie då parametrarna p, q, P och Q är okända. Innan ML appliceras behöver ekvation 7 differentieras enligt

$$e_t = (1 - B)^d (1 - B^s)^D (x_t - \mathbf{u}'_t \beta) \quad (12)$$

Där e_t kan skrivas om till matrisform där $e = (e_1, e_2 \dots e_n)$. Nu kan parametrarna, parameterskattningarna och regressionskoefficienten β estimeras av följande ekvation

$$L(\beta, \phi, \Phi, \theta, \Theta, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\det \Gamma_n)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} e' \Gamma_n^{-1} e \right\} \quad (13)$$

Där $\Gamma_n(\beta, \phi, \Phi, \theta, \Theta, \sigma^2)$ är kovariansmatrisen av e . Eftersom e_t är en SARMA(p, q)(P, Q) - process med parametrarna $(\phi, \Phi, \theta, \Theta, \sigma^2)$ generas ML-estimaterna $\hat{\beta}, \hat{\phi}, \hat{\Phi}, \hat{\theta}$ och $\hat{\Theta}$ genom att minimera

$$\ell(\beta, \phi, \Phi, \theta, \Theta) = \ln(n^{-1} S(\beta, \phi, \Phi, \theta, \Theta)) + n^{-1} \sum_{t=1}^n \ln r_{t-1} \quad (14)$$

Där

$$S(\beta, \phi, \Phi, \theta, \Theta) = \sum_{t=1}^n (e_t - \hat{e}_t)^2 / r_{t-1} \quad (15)$$

\hat{e}_t är den bästa enstegs prediktorn för e_t och $r_{t-1}\sigma^2$ är det genomsnittliga kvadratfelet för \hat{e}_t . ML-estimatet för σ^2 ges av

$$\hat{\sigma}^2 = n^{-1}S(\beta, \phi, \Phi, \theta, \Theta) \quad (16)$$

(Brockwell, Davis. 2002, s 213).

5.3 Expanding window

Expanding window metoden används vid tidsserieanalyser och innebär att träningsdatan innehåller alla tillgängliga observationer fram till den horisont som skall predikteras (Lazzeri, F. 2020, s 57). Applicerat på denna studie används först alla observationer från 2017-01-01 fram till 2018-12-31 för att träna den första modellen som predikterar värdet för 2019-01-01. Det faktiska värdet från 2019-01-01 adderas sedan till träningsdatan som nu skall prediktera värdet för 2019-01-02. Metoden fortsätter att prediktera för en given horisont med nya faktiska värden vid varje ny prediktion.

Eftersom träningsdatan konstant uppdateras med faktiska värden driver inte prediktionerna ifrån den verkliga tidsserien vilket förbättrar prediktionsförmågan. I undersökningen används ARIMA-baserade modeller där parametrarna (p, d, q) eller för SARIMA $(p, d, q)(P, D, Q)_s$ är permanenta genom alla prediktioner. Parametrarna hålls permanenta eftersom tidsserien inte visar större förändring över den observerade tidsserien därav finns inget som antyder att parametrarna behöver omvärderas. Även eftersom denna studie genererar 658 enstegsprediktioner, om då parametrarna skulle omprövats vid varje steg hade beräkningsbördan blivit många gånger större och ofta resulterat i samma val av parametrar. Däremot skattas ML om för parametervärdena i varje enskild prediktion, som förklarats under avsnitt 5.2.1.

5.4 Avvikelsemått

I resultatet framställs modellernas prediktionsfel med ett antal avvikelsemått som i detta avsnitt matematiskt förklaras. Till att börja med det mått som används mest frekvent inom energibranschen, MAPE kan förklaras som:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y(t) - \hat{y}(t)|}{y(t)} \right) \quad (17)$$

Där $y(t)$ är det faktiska värdet vid tiden t och $\hat{y}(t)$ är det predikterade värdet vid tiden t och n är antalet observationer (Hutama, Akbar, Candra, 2018). Övriga avvikelsemått som används i studien är

$$MAE = \frac{1}{n} \sum_{i=1}^n |y(t) - \hat{y}(t)| \quad (18)$$

som mäter den absoluta avvikelsen där $y(t)$ är det faktiska värdet vid tiden t och $\hat{y}(t)$ är det predikterade värdet vid tiden t och n är antalet observationer (Cortinhas, Black. 2012, s 63). Slutligen används även följande

$$RMSE = \frac{1}{n} \sum_{i=1}^n \sqrt{(y(t) - \hat{y}(t))^2} \quad (19)$$

där $y(t)$ är det faktiska värdet vid tiden t och $\hat{y}(t)$ är det predikterade värdet vid tiden t och n är antalet observationer (Cortinhas, Black. 2012, s 92). Både RMSE och MAE är beroende av storleken på observationerna vilket bör vara i beaktning vid jämförelser mellan resultat med olika numerisk storlek.

6. Resultat

I kommande avsnitt kommer studiens resultat framställas. De metoder som använts för att nå resultaten beskrevs ingående i avsnitt 5. I det första delavsnittet kommer den första frågeställningen besvaras där fokus är på att undersöka om variabler för el- eller hybridbilar och solceller förbättrar prediktionsförmågan för dagliga toppar i energikonsumtion. I det andra delavsnittet kommer den andra frågeställningen stå i fokus. Där separeras hushållen beroende på om de har el- eller hybridbil, solceller, båda eller ingendera för att sedan undersöka om det finns någon differens mellan dem. Mer information kring incitamentet för uppdelningen och hur den genomförts tekniskt ges i avsnitt 3. För att vara användbart som underlag vid framtida dimensionering av elnät skapas en modell som med 95% prediktionsintervall predikterar skillnader mellan hushåll med tidigare uppräknade installationer.

6.1 Förbrukningsprediktioner

Första steget är att hitta parametrarna (p, d, q) eller för SARIMAX $(p, d, q)(P, D, Q)_s$ samt vilka exogena variabler som genererar lägre AIC. Till detta används de faktiska värdena för de exogena variablerna för att inte prediktionsresultatet skall ha någon inverkan. I Tabell 1 visas vilka variabler som är inkluderade i den modell med lägst AIC.

| Variabel 1 | Variabel 2 | Variabel 3 | AIC |
|------------|-----------------------------|-----------------------------|---------------|
| Temperatur | | | -647,9 |
| Temperatur | Solenergiproduktion | | -657,4 |
| Temperatur | Antal el- eller hybridbilar | | -645,9 |
| Temperatur | Solenergiproduktion | Antal el- eller hybridbilar | -655,4 |

Tabell 1. ARIMAX (6,0,0) för toppar i energikonsumtion med olika kombinationer av exogena variabler.

Som visas i Tabell 1 ger kombinationen temperatur och solenergiproduktion det lägsta AIC-värdet. Förvånande, baserat på teorin inom ämnet genererar inte variabeln för antalet el- eller hybridbilar någon förbättrad AIC. All information som behövs för att skapa den slutgiltiga modellen är nu tillgänglig. Modellen som kommer användas är en ARIMAX (6,0,0) med prediktionerna för temperatur och solenergiproduktion som exogena variabler. ARIMAX (6,0,0) innebär att modellen använder AR(6) alltså värden från de 6 föregående dagarna för att prediktera det kommande. Baserat på teorin skulle den mest väntade modellen vara en AR(7) eftersom den även inkluderar samma veckodag föregående vecka men en AR(6) är inte helt förvånande.

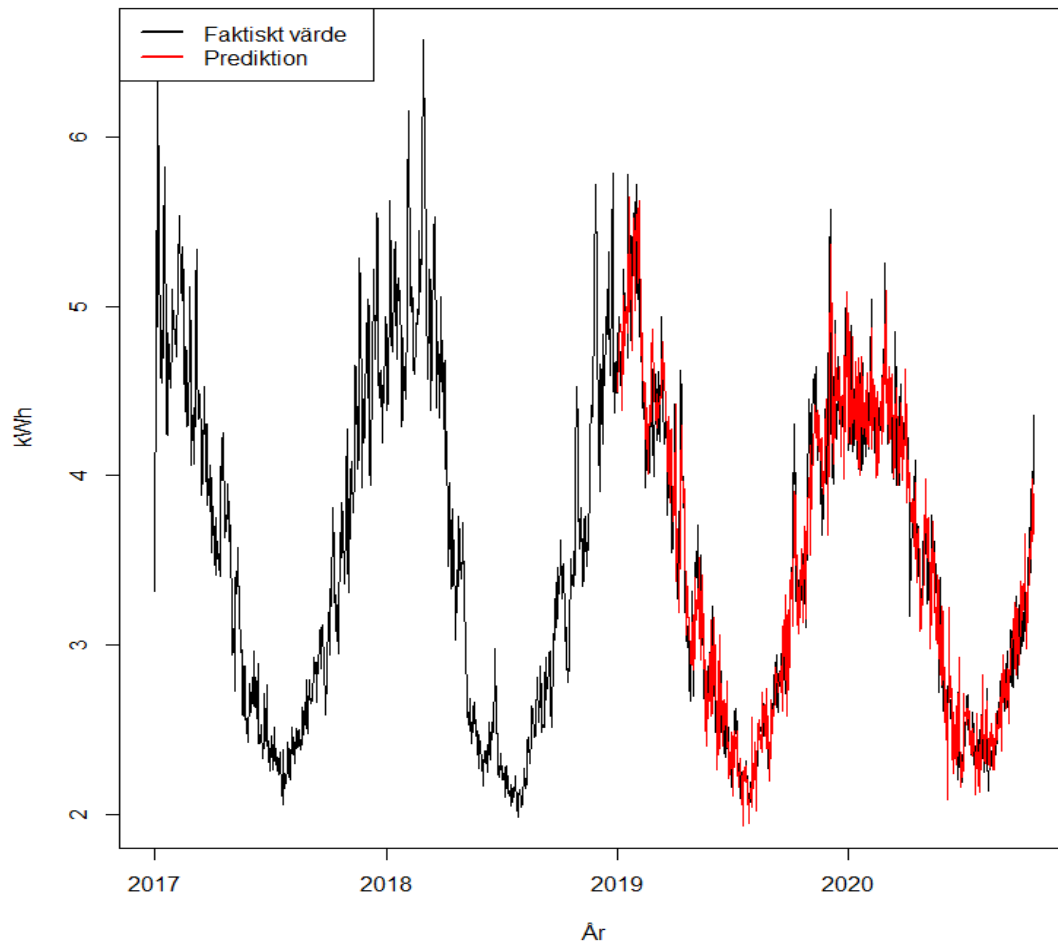
För att möjliggöra de expanding window predikteringen behöver de exogena variablerna predikteras på liknande vis som den faktiska modellen. Genom att först hitta den SARIMA eller ARIMA modell som når lägst AIC används "auto.arima". När denna sedan är fastslagen används expanding window approach som beskrivs i avsnitt 5.3 för att generera prediktioner med konstant återanknytning till verkliga värden. De två exogena variablerna som genererade lägst AIC visas i Tabell 2 nedan.

| Variabel | Modell | AIC | RMSE | MAE |
|---------------------|-------------------------------------|---------|--------|--------|
| Temperatur | SARIMA(3,0,0)(0,1,0) ₃₆₅ | 5185.57 | 23.076 | 18.169 |
| Solenergiproduktion | ARIMA(0,1,2) | 1786.63 | 0.3463 | 0.2576 |

Tabell 2. Prediktionsresultaten för de exogena variablerna.

För att prediktera temperaturen användes en SARIMA(3,0,0)(0,1,0)₃₆₅ vilket innebär att en AR(3) modell med en säsongsmässig differentiering var den som gav lägst AIC. AR(3) modellen använder temperaturvärdena från de 3 föregående dagarna. 365 motsvarar att tidsserien består av dagliga värden. I solenergiproduktions predikteringen genererade en ARIMA(0,1,2) lägst AIC-värde. Denna modell använder en MA(2) där det vita bruset från de två föregående dagarna används för predikteringen samt en differentiering för att uppnå stationaritet.

Prediktionsresultaten som visas i Tabell 2 kan vid en första anblick se ut att vara bristfälliga. Plotar för hur prediktionerna följer de faktiska värdena finns grafiskt i Figur B1 och Figur B2 under bilagor. Där syns hur prediktionerna till stor del följer de mönster som finns i datan. Tabell 2 är även ett exempel på att RMSE och MAE är känsliga för storleken på variabeln som nämndes i avsnitt 5.4. All delar som krävs för att generera prediktionerna är nu färdigställda. Efter ett första test av ARIMAX modellen visar det sig däremot att koefficienten för solenergiproduktionen i ARIMAX modellen är positiv. Detta betyder att vid ökad solenergiproduktion ökar energikonsumtionstopparna. Sambandet bör vara omvänt, om solenergiproduktionen ökar minskar topparna i energikonsumtion. På grund av detta kan inte variabeln användas eftersom den visar ett samband som inte har teoretiskt eller logiskt stöd. Därav kommer enbart temperatur användas som exogen variabel i modellen.



Figur 7. Prediktionsresultat för toppar i energiförbrukning (förstorad bild över prediktion finns i Figur B5 i bilagor)

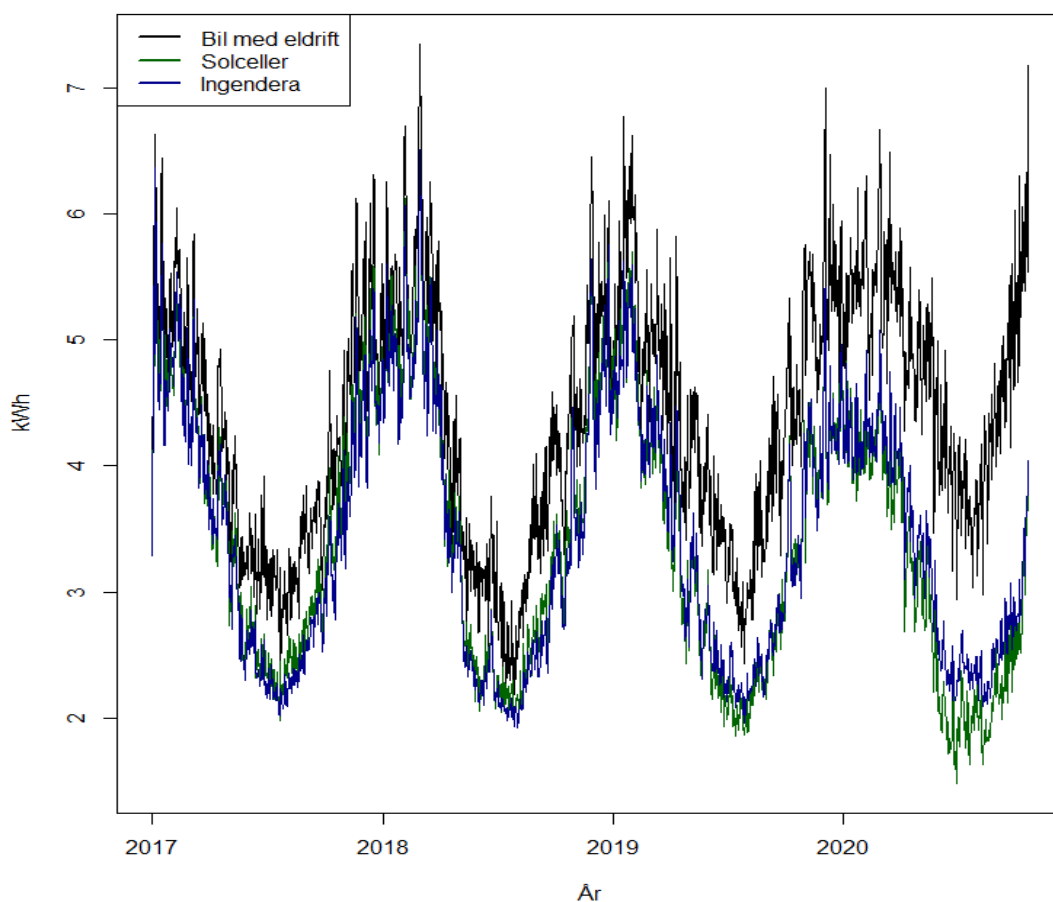
I Figur 7 framställs resultatet av prediktionerna för toppar i energikonsumtion. Trots att endast en exogen variabel kunde användas följer prediktionen de faktiska värdena väl. ACF-plotten för modellen visar på snabbt avtagande autokorrelation för laggen vilket tyder på att modellen uppnår antagandet om stationaritet. ACF-plotten för modellen visas i Figur B3 i bilagor. I tabell 3 nedan visas avvikelsemått för prediktionerna. Avvikelsemått bekräftar det Figur 7 visar, modellen lyckas fånga större del av variationerna och når MAPE på 4.15%.

| Modell | RMSE | MAE | MAPE |
|---------------|--------|--------|-------|
| ARIMAX(6,0,0) | 0.1842 | 0.1433 | 4.20% |

Tabell 3. Avvikelsemått för prediktioner av toppar i energikonsumtion med temperatur som exogen variabel.

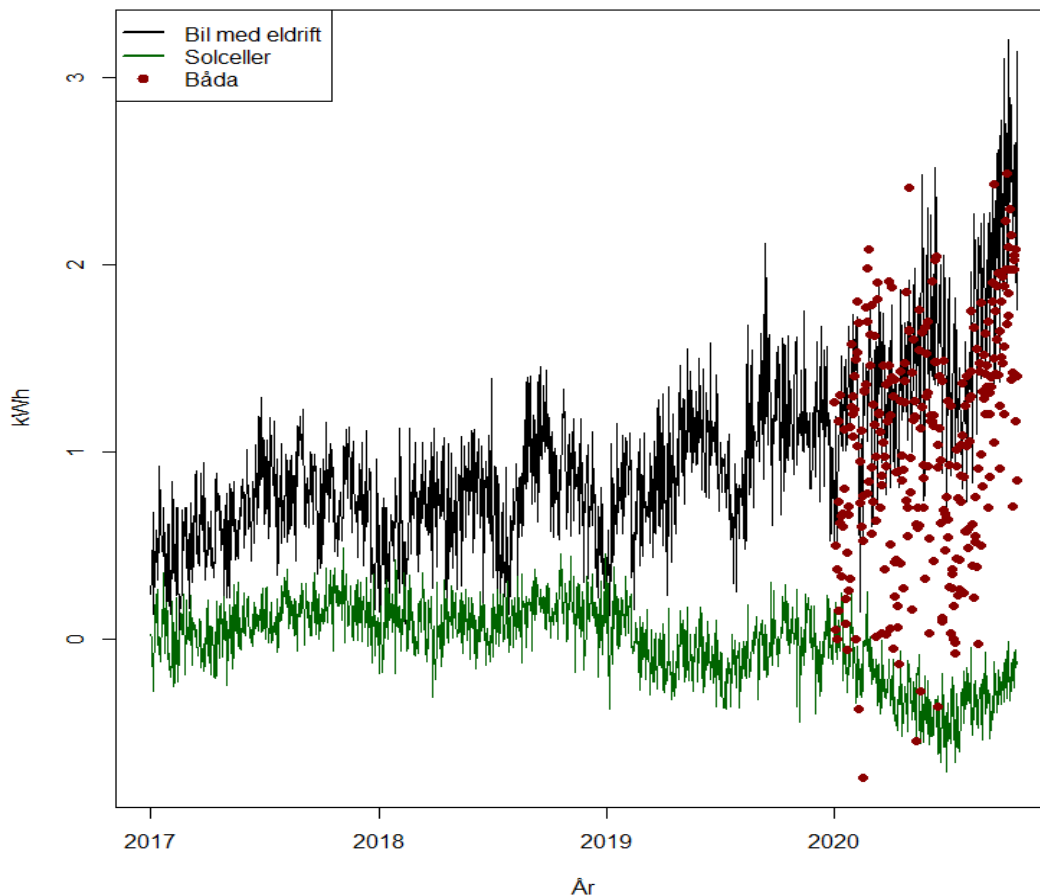
6.2 Energikonsumtions differenser

För att skapa en förståelse över hur stor effekt inköp av bil med eldrift eller solenergianläggning har på ett hushåll delas datan upp i subgrupper baserat på vilken av de två tekniska installationerna de har. Under Figur 4 i avsnitt 3 beskrivs hur datan som visas i Figur 8 framställts.



Figur 8. Årlig variation baserat på hushållens tekniska installationer.

I Figur 8 syns en antydning till en tilltagande skillnad mellan de hushåll som har el- eller hybridbil och de som inte har. Det visas även att sommaren 2020 var första året som hushåll med solenergianläggningar hade marginellt lägre toppar i energikonsumtion än de som har ingendera. För att förtydliga skillnaderna visar Figur 9 differenserna för energikonsumtionstoppar mellan hushåll med el- eller hybridbil, solenergianläggningar eller båda. Hushåll med ingendera används i Figur 9 som referenspunkt.



Figur 9. Skillnad i energikonsumtionstoppar beroende på tekniska installationer med hushåll utan nämnda installationer som referenspunkt.

I Figur 9 kan utvecklingen av skillnaden i energikonsumtionstoppar följas. Det ploten visar är hur hushållens toppar i energikonsumtion skiljer sig från de hushåll som har varken el- eller hybridbil eller solceller. Den svarta linjen motsvarar ökningen i energikonsumtionstoppar om hushållet har en el- eller hybridbil jämfört med hushåll som har ingendera. Denna skillnad har varit tilltagande sedan 2019 och skillnaden mellan hushåll med och utan el- eller hybridbil ser en allt tydligare trend. Figur 9 visar även att om ett hushåll har solceller installerat genererar detta inte till någon större förändring vad gällande toppar i energikonsumtion. De röda prickarna motsvarar de hushåll som har både el- eller hybridbil och solceller jämfört med hushåll med ingendera. I detta fall kan det konstateras att de högsta värdena är på samma nivå som för hushållen med enbart el- eller hybridbil men variationen är relativt sätt mycket stor. Vilket kan vara ett problem då prediktioner skall genereras för kortsiktig energiefterfrågan i smartgrid-syfte.

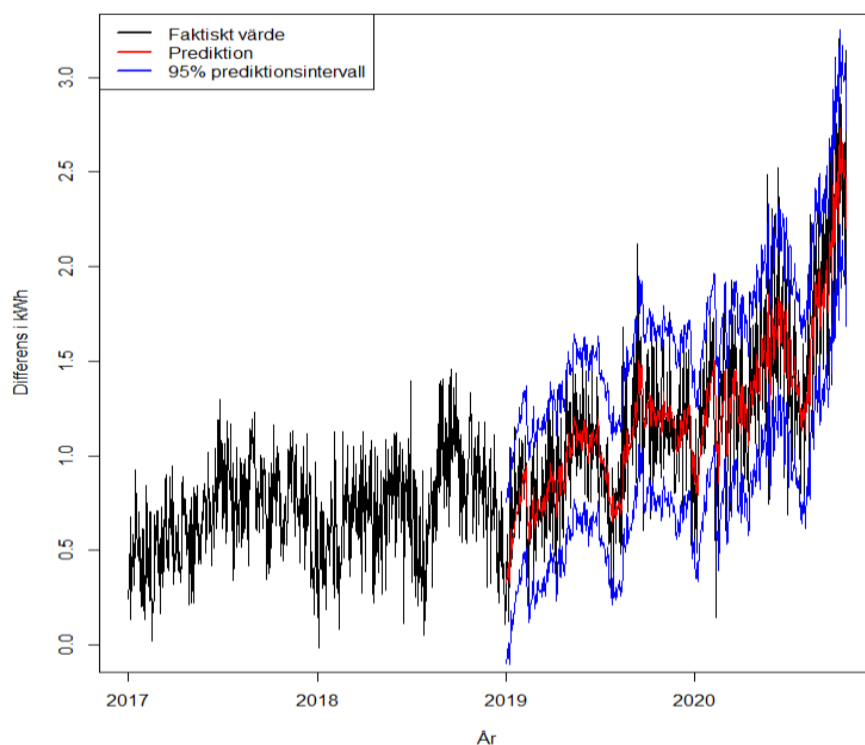
Resultatet av denna studie skall användas som dimensioneringsunderlag och kan därmed enbart fokusera på den subgrupp som genererar högst toppar i energikonsumtion. Därav kommer skillnaden mellan hushåll som har ingendera och el- eller hybridbil predikteras. De exogena variablerna som här blir aktuella är temperaturskillnaden och antalet el- eller hybridbilar. Eftersom temperaturvariablerna för de två subgrupperna är den medeltemperatur som är kopplad till topparna i energikonsumtion blir skillnaden extra intressant. En större skillnad antyder att andra variabler än väderförhållanden påverkar energikonsumtionstopparna. Även att de rent tidsmässigt frångår de klassiska konsumtionstopparna som hushåll utan el- eller hybridbil bör följa.

För att se vilken kombination av exogena variabler som skall användas testas vilken som genererar lägst AIC. Precis som i avsnitt 6.1 används de faktiska värdena för de exogena variablerna vid val av parametrarna (p, d, q) eller $(p, d, q)(P, D, Q)$. I Tabell 4 visas att de exogena variablerna i detta fall inte bättre resultat. En ARIMA (4,1,1) når ett lägre AIC än modellerna med olika kombinationer av exogena variabler. Detta betyder att modellen har en AR(4) del där de 4 föregående värdena används vid predikteringen. Modellen har även behövt differentieras för att uppnå stationaritet. Även en MA(1) del som använder det vita bruset från föregående dag. Den teoretiska uppbackningen av modellvalet är inte lika självklart som i föregående prediktion eftersom det här är beroende av individers laddningsrutiner.

| Variabel 1 | Variabel 2 | AIC |
|-----------------------------|-----------------------------|---------------|
| | | 246.96 |
| Temperatur | | 247.19 |
| Antal el- eller hybridbilar | | 248.12 |
| Temperatur | Antal el- eller hybridbilar | 248.28 |

Tabell 4. ARIMAX (4,1,1) för skillnader i energikonsumtionstoppar med olika kombinationer av exogena variabler.

I Figur 10 visas resultatet av prediktionerna med 95% prediktionsintervall som genererats av ARIMA (4,1,1) med expanding window approach. Prediktionerna följer de faktiska värdena bra sett till den variation som finns i datan. Att modellen kan hantera stora variationer är viktigt eftersom hushåll med el- eller hybridbil ofta generera detta på grund av inkonsekventa laddningsrutiner.



Figur 10. Prediktionsresultat för skillnader i energiförbrukningstoppar (förstorad bild över prediktion finns i Figur B6 i bilagor)

Prediktionsintervallet kan även valideras i Figur 10. 95% av de faktiska värdena skall vara inom intervallet. 32 faktiska värden är utanför intervallet, eftersom modellen genererat 658 prediktioner är 95.14% av de faktiska värdena inom intervallet. MAPE-resultatet som visas i Tabell 5 är sämre än för den tidigare prediktionen. Ett väntat resultat givet att denna tidsserie har högre varians än tidigare. Även ACF-ploten för denna modell visar på snabbt avtagande autokorrelation för laggen vilket tyder på att modellen uppnår antagandet om stationaritet. ACF-ploten för modellen visas i Figur B4 i bilagor.

| Modell | RMSE | MAD | MAPE |
|--------------|--------|--------|--------|
| ARIMA(4,1,1) | 0.2326 | 0.1819 | 18.58% |

Tabell 5. Avvikelsemått för prediktioner av skillnader i energiförbrukningstoppar

Det kan även konstateras att under 2020 har differensen mellan hushåll med el- eller hybridbil och ingendera förändrats i snabbare takt än tidigare. Den nedre blåa linjen i Figur 10 som visar att värdet med 97.5% sannolikhet bör befinna sig över denna har under andra delen av 2020 mer än fördubblats och befinner sig runt 2 kWh skillnad. Vilket säger att med 97.5% sannolikhet är skillnaden mellan hushåll med el- eller hybridbil och hushåll med ingendera över 2kWh i oktober 2020.

7. Diskussion

Detta slutliga avsnitt av studien ämnar att diskutera resultatet med utgångspunkt i frågeställningarna. För att sedan studera framtidsutsikter för denna typ av energikonsumtionsanalyser och förändringar hos energikonsumtionen. Diskussionen kommer även ta upp metodval, jämförelse med tidigare studier samt förslag på framtida.

7.1 Summering av resultat

Predikterade en modell med variabler för antalet el- eller hybridbilar och solenergiproduktion bättre än den ARIMA-baserade modellen som enbart använde temperatur? Resultatet visade att detta inte var fallet. Att addera variabeln för solenergiproduktion genererade lägre AIC men variabeln visade sig ha motsatt inverkan än vad som teoretiskt kan uppbackas. Därav kunde inte variabeln användas då den uppenbarligen fångade något annat underliggande samband.

Skillnaderna mellan hushållens toppar i energikonsumtion baserat på om de installerat solceller, el- eller hybridbil eller ingendera visade ett intressant resultat. Skillnaderna som uppstod då ett hushåll hade solceller jämfört med ingendera genererade inte någon större minskning av topparna i energikonsumtion. Differensen då ett hushåll hade en el- eller hybridbil däremot visade på en tilltagande trend. Denna trend kunde predikteras med tillfredsställande resultat då över 95% av de faktiska värdena är innanför de genererade prediktionsintervall.

7.2 Utvärdering av resultat

Sett till den första modellen som predikterar hushållens toppar i energikonsumtion kan det konstateras att variablerna som var tillgängliga inte kunde fånga den variation som sedan visade sig vid uppdelningen av hushållen. Den faktiska skillnaden som finns mellan hushåll som har el- eller hybridbil och ingendera är markant och bör om rätt variabel hittas kunna förklara en del av den variation som finns i de dagliga topparna i energikonsumtion. Variabeln för solenergiproduktion minskade AIC-resultatet men konstaterades vara oanvändbar då den visade motsatt inverkan på modellen än vad som förväntades. Som visats under de tidigare avsnitten av denna studie, solenergiproduktion och energikonsumtionstoppas sammanfaller sällan vilket tyder på att solenergiproduktionen saknar större inverkan på topparna i energikonsumtion. I Figur 8 respektive 9 kan en tendens till minskade energikonsumtionstoppas under sommaren anses men ingen markant skillnad synliggörs.

Den växande differensen kan komma att bli ett stort problem för elnätsägare de kommande åren eftersom det är svårt för dem att prediktera el- eller hybridbilsmarknaden eller de tillhörande laddarna. Den ökade belastningen uppstår när ett hushåll laddar bilen, det är alltså inte bilen i sig som förklarar lasten utan även laddaren som används. Eftersom majoriteten av el- eller hybridbilar behöver laddas mer än en timma beror toppkonsumtionen på laddarens effekt. För att mer precist kunna prediktera toppar i energikonsumtion bör information kring laddarna insamlas. Det finns ett hushåll vars konsumtionsmönster antyder att de har två laddare eller bytt till en laddare som ger dubbel effekt jämfört med den som tidigare använts. Om ett hushåll har två el- eller hybridbilar genereras ingen skillnad jämfört med de som har en. Möjligtvis blir topparna i energikonsumtion mer stabila eftersom billaddningen kommer ske mer frekvent. Det är först när ett hushåll skaffar ytterligare en laddare som det uppstår skillnader i energikonsumtionstoppar. Det kan också vara så att det är huvudsäkringen som sätter begränsning för effekt i uttaget. Det kostar pengar att säkra upp vilket många vill undvika.

Variabeln för antalet el- eller hybridbilar som användes under analysen var bristfällig. Variabeln indikerade enbart vilken dag en ny el- eller hybridbil införskaffades. Det finns flera problem med detta. Dels beror variabeln av kvalitén på svaret som angetts i enkätundersökningen och hur precisa de varit. Det finns även en viss förvirring då individer enbart angett inköpsdatum för den nuvarande bilen men har i verkligheten haft el- eller hybridbil under en längre period innan den nuvarande. Eftersom många bilar står på leasingföretag kan dem inte kopplas till området och bidrar därmed tyvärr ingen utväg. Inköpsdatumerna i sig är inte det intressanta, det som faktiskt påverkar energikonsumtionen är när bilen laddas. Eftersom vi inte har kännedom om individers laddningsvanor säger variabeln bara att efter denna punkt kan variationerna i energikonsumtionstoppar vara högre eftersom en el- eller hybridbil inte laddas varje dag.

7.3 Utvärdering av metod samt jämförelse med tidigare studier

I jämförelse med de studier som nämnts i avsnitt 4 når denna studie liknande MAPE-resultat. I Majoriteten av studierna som studerades användes däremot aggregerad energikonsumtion och visade mindre varierande tidsserier vilket ger bättre förutsättningar för att generera bra prediktioner. Klimatet och de tekniska installationerna som är vanliga på de platser som studierna i avsnitt 4 gjorts skiljer sig även dem i stor utsträckning från de förutsättningar som vi har i Sverige. Trots detta genererar denna studie liknande resultat med en tidsserie där

variationen är högre. Att MAPE-resultatet blev liknande i denna studie beror på att en window approach användes. Med enbart användning av ARIMA-baserad modeller hade prediktionerna drivit bort från de faktiska värdena. Med expanding window får modellen de faktiska värdet av de som i föregående steg predikterades och kan på så vis alltid generera kommande prediktion med all tillgänglig information. Att modellen konstant fick möjlighet att återknyta till de faktiska värdena möjliggjorde prediktionerna för differenspredikteringen som har alltför stor variation för att predikteras över en längre horisont med god precision.

7.4 Förslag till fortsatta studier

Fortsatta studier inom ämnet behöver dedikerad mer tid till att generera fler exogena variabler som har signifikant förklaringskraft. En av studies tillkortakommanden var bristen på information som var tidsserieanpassad. En stor del av den information som gavs var demografisk och stationär över tid vilket gjorde dem meningslösa att använda i denna typ av modell. Med fler variabler som varierar över tid kan även maskininlärningsmetoder som ANN och SVM användas som komplettering till de ARIMA-baserade modellerna. En av de stora utmaningarna är att förstå och prediktera när en el- eller hybridbil laddas, hur många som kommer ladda samtidigt och laddarens effekt. Dessa variabler är svårpredikterade men viktiga för att i framtiden kunna prediktera energikonsumtion. För att förhindra dyra utbyggnader av elnätet bör potentiella begränsningar för hemmaladdning undersökas. Eventuellt bör möjligheter för lagring av ett områdes solenergiöverskott undersökas. På så vis kan den energin användas under de timmar som har högst energikonsumtionstoppar, avlasta elnätet och förhindra utbyggnader. Under vintertid då topparna i energikonsumtion är höga och solenergin nästintill obefintlig kan energilagringen laddas under den tid på dygnet effektbehovet är lågt och på så vis jämna ut energiflödet under dagen.

Referenser

- Abdel-Aal, R. E., & Al-Garni, A. Z, 1997, Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. *Energy*, 22(11), 1059-1069.
- Brockwell, P. Davis, R, 2002, *Introduction to Time Series and Its Forecasting*. Springer.
- Cortinhas, C. Black, K, 2012, *Statistics for business and economics*. Wiley
- Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W, 2017, A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74, 902-924.
- EIA (U.S. Energy Information Administration), Hourly electricity consumption varies throughout the day and across seasons, 2020, *available at* <https://www.eia.gov/todayinenergy/detail.php?id=42915> (last visited January 24, 2020)
- Energimyndigheten, Solelportalen, 2019, *available at* <http://www.energimyndigheten.se/fornybart/solelportalen/har-mitt-hus-ratt-forutsattningar/sa-undersoker-du-taket/> (last visited January 19, 2021)
- Energimyndigheten, Statistikdatabas ,2020, *available at* https://pxexternal.energimyndigheten.se/pxweb/sv/N%3%a4tanslutna%20solcellsanl%3%a4ggningar/N%3%a4tanslutna%20solcellsanl%3%a4ggningar/EN0123_1.px/table/tableViewLayout2/(last visited October 24, 2020)
- Hutama, A. H., Akbar, S., & Candra, M. Z. C, 2018, Medium Term Power Load Forecasting for Java and Bali Power System Using Artificial Neural Network and SARIMAX. *2018 5th International Conference on Data and Software Engineering (ICoDSE)* (pp. 1-6). IEEE.
- Hyndman, RJ and Khandakar, Y ,2008, Automatic time series forecasting: The forecast package for R, *Journal of Statistical Software*, 26(3).
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of econometrics*, 54(1-3), 159-178.

Lazzeri, F, 2020, *Machine Learning for Time Series Forecasting with Python*. Wiley

Munkhammar, J., Grahn, P., & Widén, J, 2013, Quantifying self-consumption of on-site photovoltaic power generation in households with electric vehicle home charging. *Solar energy*, 97, 208-216.

Rallapalli, S. R., & Ghosh, S. ,2012, Forecasting monthly peak demand of electricity in India—A critique. *Energy policy*, 45, 516-520.

Shumway, R., Stoffer, D, 2011, *Time Series Analysis and Its Applications*. Springer.

Smith, J., & Yadav, S, 1994, Forecasting costs incurred from unit differencing fractionally integrated processes. *International Journal of Forecasting*, 10(4), 507-514.

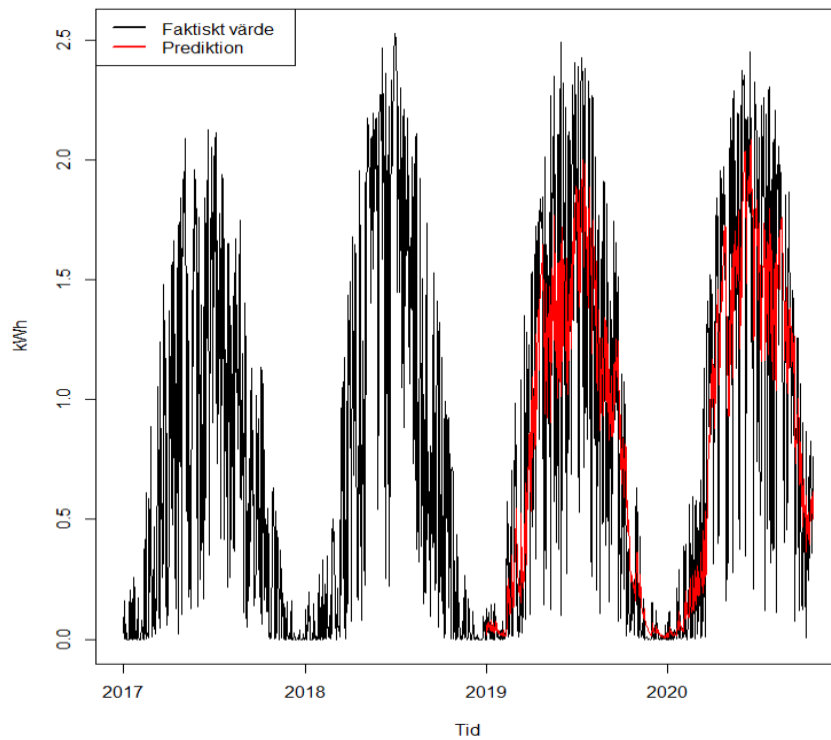
Stridh, B., & Larsson, D, 2017, Investeringskalkyl för solceller. *E2B2 Forskning och innovation för energieffektivt byggande och boende*.

Tserkezos, E. D, 1992, Forecasting residential electricity consumption in Greece using monthly and quarterly data. *Energy Economics*, 14(3), 226-232.

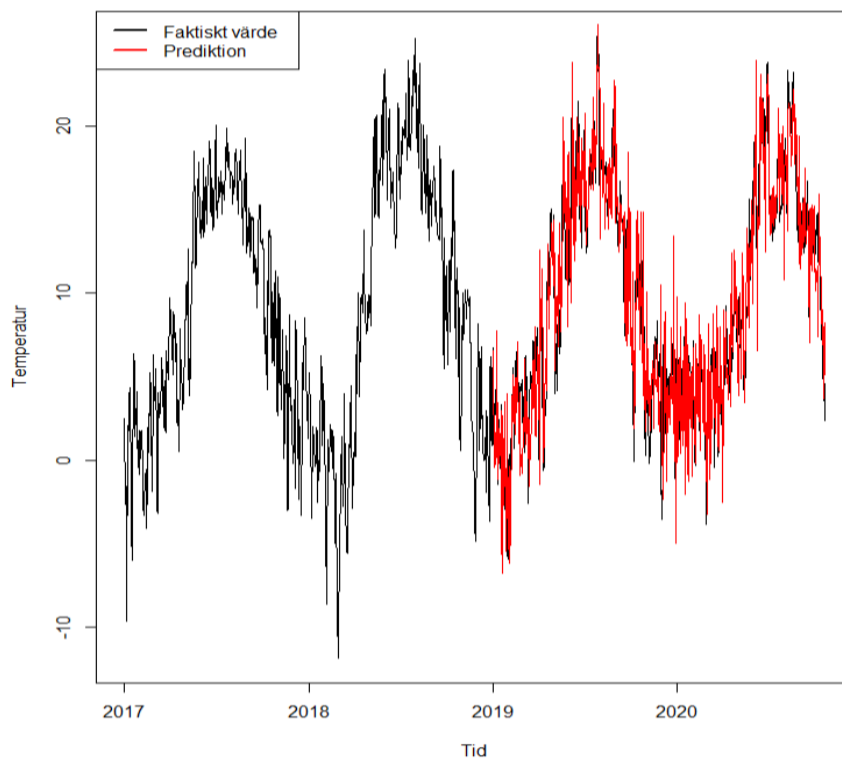
Zhang, G. P, 2003, Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.

Bilagor

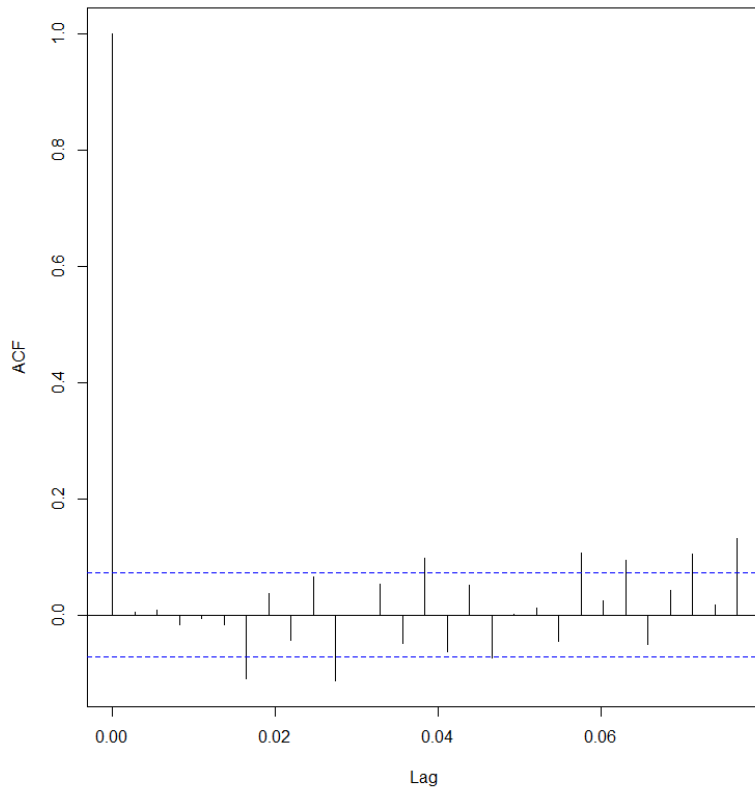
B.1 Övriga figurer



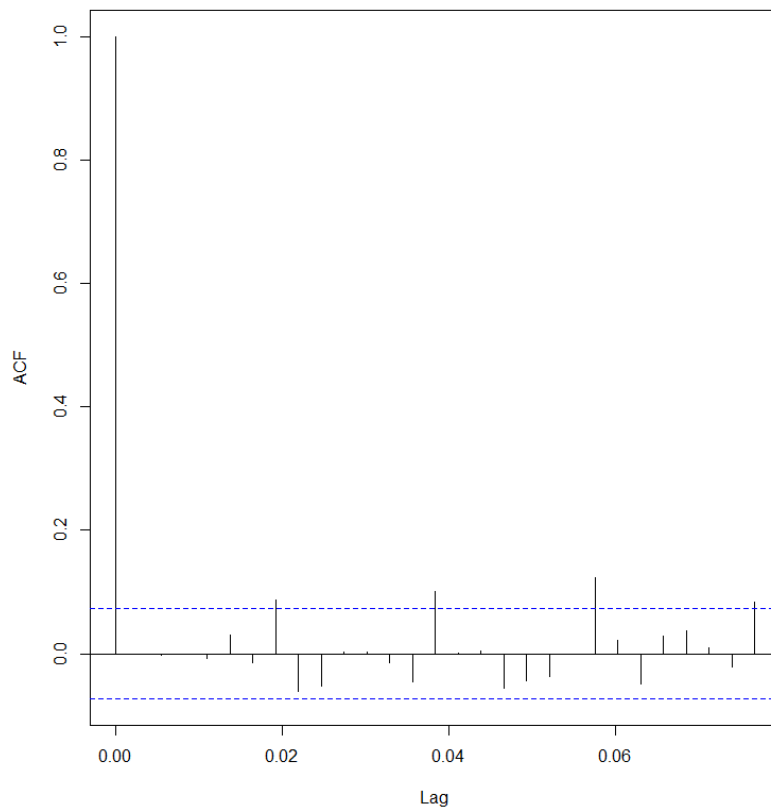
Figur B1. Prediktionsresultat för variabeln solenergiproduktion



Figur B2. Prediktionsresultat för variabeln Temperatur



Figur B3. ACF plot energikonsumtionstoppar.



Figur B4. ACF plot skillnader i energikonsumtionstoppar

B.2 Enkätundersökning

- Direktverkande el
- Fjärrvärme
- Bergvärmepump
- Luft/luft- eller luft/vatten-värmepump
- Frånluftsvärmepump
- Pellets/ved
- *Annat (var god specificera nedan)

*Specificera uppvärmningssystem

Hur många vuxna personer är ni i hushållet?

Använder ni bil med eldrift som laddas hemma?

- Nej
- Ja, och laddar mestadels hemma
- Ja, men laddar mestadels inte hemma

Om ja, vilken typ av elbil är det på bilen/bilarna?

Vilket inköpsår och månad (ungefär)?

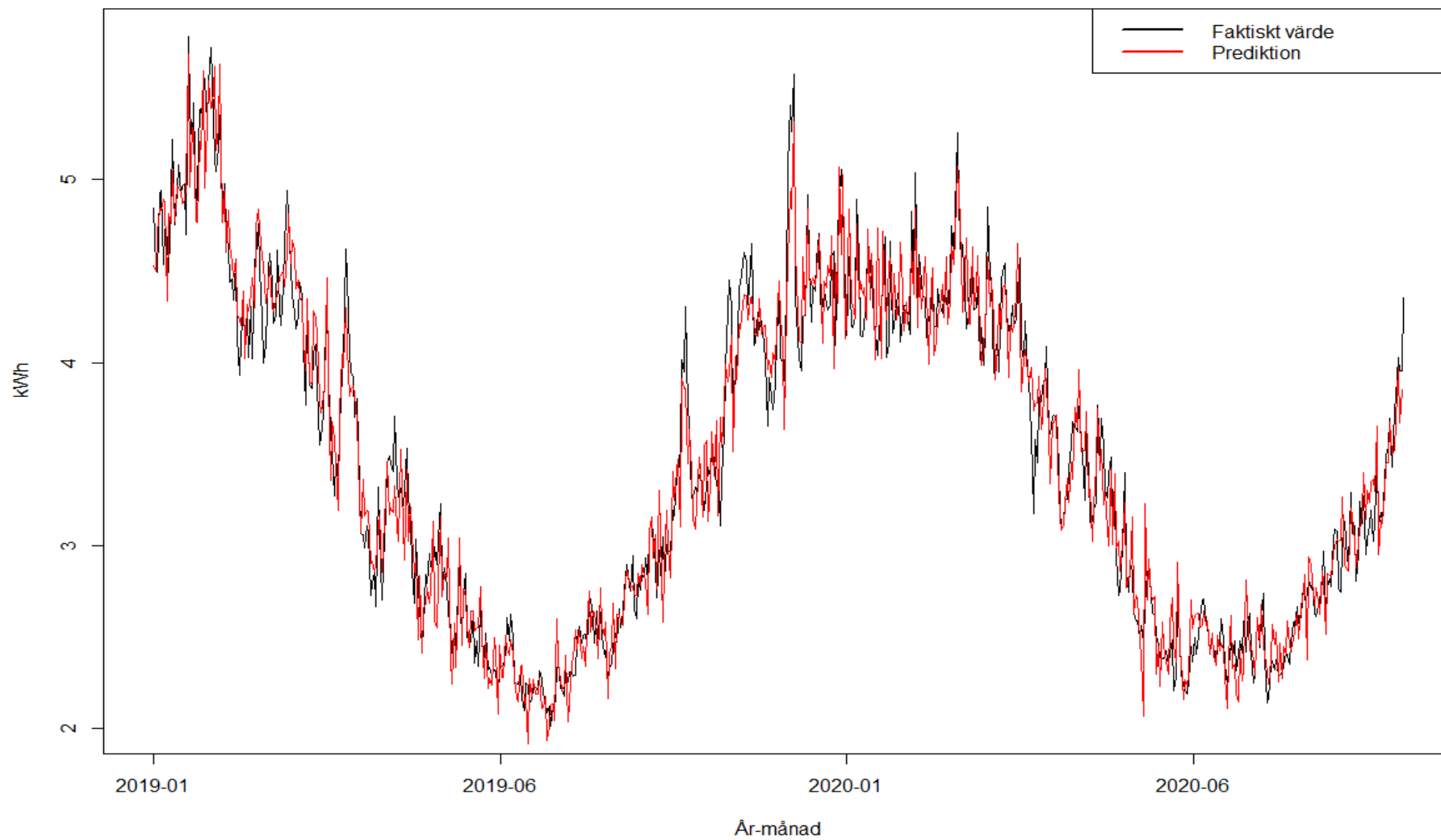
Kommentar kring frågorna om elbilar

Har ni solceller installerade?

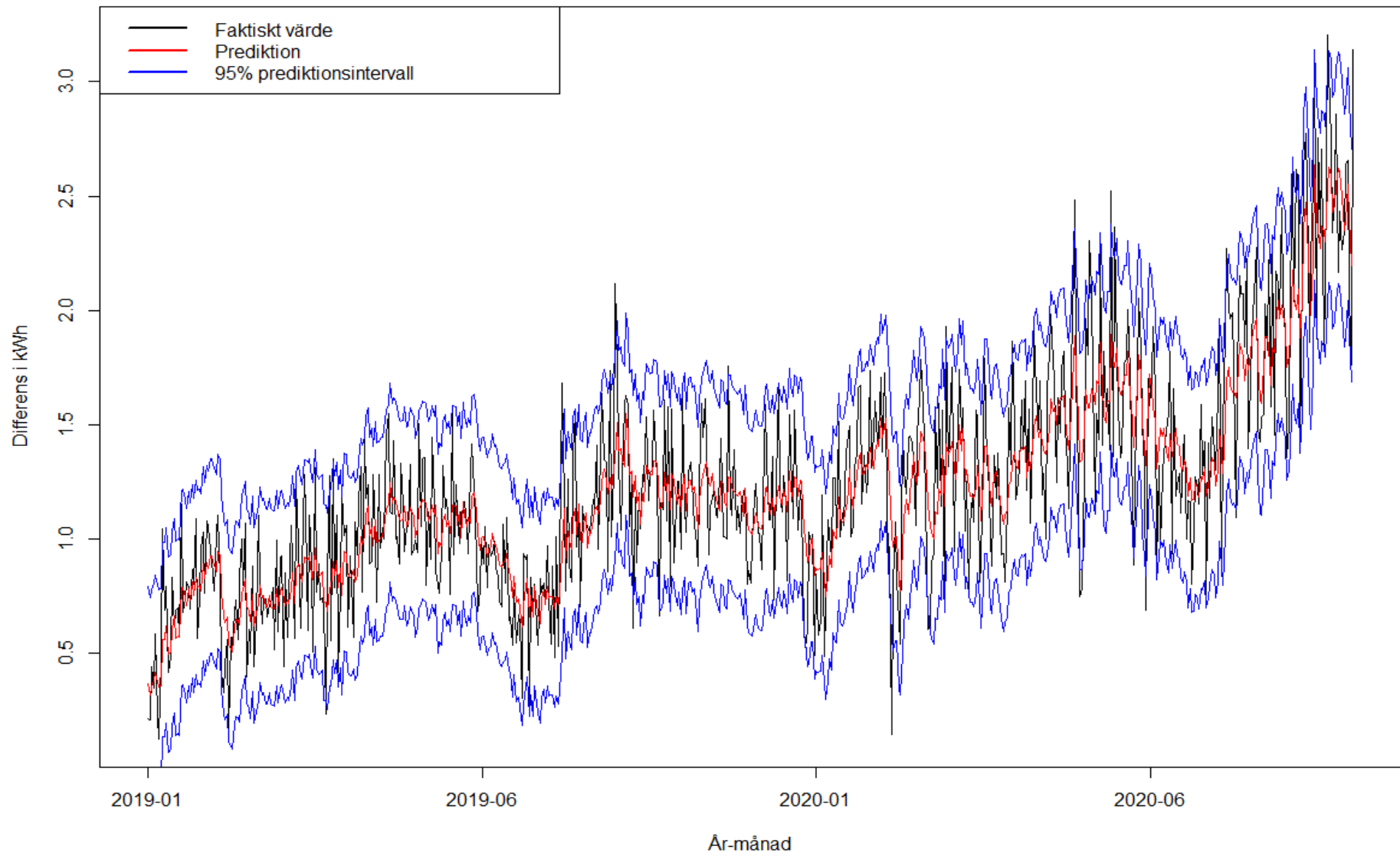
- Ja
- Nej

Om ja, vilket år och månad installerades de?

B.3 Förstorade prediktionsresultat



Figur B5. Förstorat prediktionsresultat från Figur 7



Figur B6. Förstorat prediktionsresultat från Figur 10