

Gestió de la demanda elèctrica de les llars de Londres: Predicció i anàlisi dels consums.

Laura Guzman Boy
Màster de Ciència de dades
Àrea 3.2 Models Predictius

Sergio Trilles Oliver
Ferran Prados Carrasco

Gener de 2020



Aquesta obra està subjecta a una llicència de
[Reconeixement-NoComercial-SenseObraDerivada](#)
[3.0 Espanya de Creative Commons](#)

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Gestió de la demanda elèctrica de les llars de Londres: Predicció i anàlisi dels consums.</i>
Nom de l'autor:	<i>Laura Guzman Boy</i>
Nom del consultor/a:	<i>Sergio Trilles Oliver</i>
Nom del PRA:	<i>Ferran Prados Carrasco</i>
Data de lliurament (mm/aaaa):	<i>01/2020</i>
Titulació o programa:	<i>Màster en Ciència de Dades</i>
Àrea del Treball Final:	<i>Models predictius</i>
Idioma del treball:	<i>Català</i>
Paraules clau	<i>Gestió demanda, consums, predicció</i>
Resum del Treball (màxim 250 paraules): Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball	
<p>Donat la importància que té la demanda d'electricitat, en aquest projecte volem predir la quantitat d'energia que es consumirà en un futur a les llars de la ciutat de Londres amb precisió per a poder fer una gestió activa de la demanda.</p> <p>Per a fer-ho, primer farem un anàlisi dels patrons de la corba de demanda tant horària com diària i buscarem quins factors influeixen a la demanda d'electricitat.</p> <p>Llavors, farem diferents models de prediccions amb una xarxa neuronal LSTM utilitzant diferents agregats de dades, on realitzarem prediccions del consum de la hora següent, del dia següent, de la setmana següent, del mes següent i dels dos mesos següents.</p> <p>Durant el projecte hem vist com les variables que més afecten al consum diari són el propi consum passat, el dia de la setmana, mes i estació de l'any, el nombre de consumidors, el tipus de consumidor, la hora de posta de sol i factors meteorològics com la temperatura parent màxima i mínima, la humitat, pressió i cobertura dels núvols.</p> <p>També hem vist com el millor model diari obtingut utilitza les variables consum, temperatura màxima, temperatura mínima, humitat i dia de la setmana, obtenint un error mitjà de tant sols el 2.55% per a la predicció del consum del dia següent</p> <p>Com més acurada sigui la predicció, millor seran les polítiques de gestió de la demanda, més bons resultats obtindran els agents participants del mercat elèctric i es consumirà menys CO2 i energies no renovables.</p>	

Abstract (in English, 250 words or less):

Given the importance of electricity demand, in this project we want to predict how much energy will be consumed in London households in the future in order to be able to actively manage demand).

To do this, we will first analyse the patterns of the demand curve both hourly and daily, and look for what factors influence the electricity demand.

Then, we will make different predictions models with an LSTM neural network using different data aggregates, where we will make predictions of the consumption of the next hour, the next day, the following week, the next month and the following two months.

During the project we have seen how the variables that most affect daily consumption are the past consumption, the day of the week, month and season of the year, number of consumers, type of consumer, sunset time and meteorological factors such as maximum and minimum relative temperatures, humidity, pressure and cloud cover.

We have also seen how the best daily model obtained use the variables consumption, maximum and minimum temperature, humidity and day of the week, obtaining an average error of only 2.55% for the prediction of the day ahead's consumption.

As more accurate the forecast, the demand management policies will be better, the players in the electricity market will get better results, and less CO₂ and non-renewable energy will be consumed.

Índex

1.	Introducció	1
1.1	Context i justificació de la rellevància.....	1
1.1.1	Context: Introducció al mercat elèctric	1
1.1.2	Justificació de la rellevància del tema	2
1.1.3	Justificació de les dades que s'utilitzaran	3
1.2	Descripció de la proposta	4
1.3	Objectius	4
1.4	Enfocament i metodologia	5
1.5	Motivació	7
1.6	Planificació del Treball	7
1.7	Breu descripció dels altres capítols de la memòria	8
1.8	Breu sumari resultats obtinguts.....	9
2.	Estat de l'art.....	10
2.1	Literatura de mètodes estadístics	10
2.2	Literatura de mètodes d'intel·ligència artificial.....	10
2.3	Diferència entre predicció a curt termini i predicció a llarg termini	12
2.4	Literatura predicció d'energia a llars	13
2.5	Exemples utilització del nostre conjunt de dades	14
3.	Dades	16
3.1	Descripció dels datasets	17
3.1.1	Informació de les llars (Informations households)	17
3.1.2	Temps horari i diari (Weather hourly , weather daily)	17
3.1.3	Consum horari i diari (hourly dataset, daily dataset)	19
3.1.4	Dies festius (UK bank holidays)	20
3.1.5	Tarifes (tariffs)	20
3.2	Neteja i preparació de les dades.....	20
3.2.1	Informations_households.....	21
3.2.2	Weather_daily.....	25
3.2.3	Weather_hourly	27
3.2.4	UK Bank holidays	28
3.2.5	Tariffs	28
3.2.6	Consum diari (Daily_dataset_csv)	28
3.2.7	Hourly_dataset	29
3.3	Selecció de variables	29
4.	Anàlisi de dades	31
4.1	Anàlisi estadístic	31
4.1.1	Recompte per variables qualitatives	31
4.1.2	Quantitatius: Mitjana, mediana, quartils per variables quantitatives	32
4.1.3	Correlacions	35
4.2	Contrast d'hipòtesis	37
4.2.1	Consum segons tarifa és diferent?	37
4.2.2	Consum segons tipus de consumidor és diferent?.....	38

4.2.3	Consum segons tipus de tarifa i tipus de consumidor és diferent?	39
4.2.4	Consum segons dia és diferent?.....	41
4.2.5	Consum segons mes és diferent?.....	42
4.2.6	Consum segons estació és diferent?	43
4.2.7	Consum segons temperatura és diferent?	44
4.3	Anàlisi gràfic i patrons de consum visuals.....	44
4.3.1	Anàlisi dades diàries.....	44
4.3.2	Anàlisi dades horàries	62
4.4	Regles d'associació per l'anàlisi de dependències.....	70
4.5	Variables que afecten al consum. Resum resultats.....	71
5.	Predicció.....	74
5.1	Model a utilitzar	74
5.2	Predicció agregat horari.....	75
5.3	Predicció agregat diari.....	76
5.3.1	Predicció a un dia	77
5.3.1.1	Walkforward 365 dies.....	77
5.3.1.2	Walkforward augmentant més 365 dies	100
5.3.2	Predicció a més d'un dia (multistep)	126
5.4	Predicció agregat setmanal	155
5.5	Resum resultats.....	158
5.5.1	Agregat horari.....	158
5.5.2	Agregat diari	158
5.5.3	Agregat setmanal	159
6.	Conclusions	160
6.1	Descripció conclusions	160
6.1.1	Conclusions anàlisi.....	160
6.1.2	Conclusions prediccions	161
6.1.3	Conclusions generals	161
6.2	Assoliment objectius	162
6.2	Seguiment planificació i metodologia	163
6.3	Línes treball futur	164
7.	Glossari	165
8.	Bibliografia.....	166
9.	Annexos	171

Llista de figures

GRÀFIC 1 DIAGRAMA DE GANTT.....	7
GRÀFIC 2 RECOMPTE DE LLARS SEGONS ACORN, PINTAT SEGONS ACORN GROUPED	22
GRÀFIC 3 RECOMPTE DE USUARIS PER BLOCK, PINTAT SEGONS ACORN_GROUPED	23
GRÀFIC 4 RECOMPTE DE LLARS PER BLOC PINTAT SEGONS ACORN	24
GRÀFIC 5 DIAGRAMES DE CAIXA VARIABLES NUMÈRIQUES WEATHER_DAILY	26
GRÀFIC 6 DIAGRAMES DE CAIXA VARIABLES NUMÈRIQUES WEATHER_HOURLY	27
GRÀFIC 7 DIAGRAMES CAIXA DAILY_DATASET_CSV.....	29
GRÀFIC 8 CORRELACIÓ ENTRE VARIABLES WEATHER_DAILY	36
GRÀFIC 9 RECOMPTE DE MOSTRES PER DIA	42
GRÀFIC 10 RECOMPTE DE MOSTRES PER MES	43
GRÀFIC 11 A L'ESQUERRA, GRÀFIC DE RECOMPTE DE NOMBRE DE LLARS SEGONS TARIFA, A LA DRETA, GRÀFIC DE RECOMPTE DE NOMBRE DE LLARS SEGONS ACORN_GROUPED I A SOTA, GRÀFIC DE RECOMPTE DEL NOMBRE DE LLARS SEGONS ACORN.....	44
GRÀFIC 12 EVOLUCIÓ NOMBRE LLARS AL LLARG DEL TEMPS	45
GRÀFIC 13 A DALTA, GRÀFIC DE L'EVOLUCIÓ DEL CONSUM TOTAL , A SOTA, GRÀFIC DE L'EVOLUCIÓ DEL CONSUM MITJÀ PER LLAR.....	46
GRÀFIC 14 A SOBRE, GRÀFIC DE L'EVOLUCIÓ DEL CONSUM MITJÀ I LA TEMPERATURA MITJANA A ESCALA MENSUAL, A SOTA A ESCALA SETMANAL	47
GRÀFIC 15 EVOLUCIÓ DIFERENTS VARIABLES METEOROLÒGIQUES COM CLOUD COVER, DEW POINT, HUMIDITY, TEMPERATURE MIN, UV INDEX I VISIBILITY.....	47
GRÀFIC 16 CONSUM MITJÀ PER LLAR AL LLARG DE L'ANY	48
GRÀFIC 17 MITJANA CONSUM D'UN DIA, MITJANA CONSUM DE MITJA HORA DE UN DIA I MITJANA DEL CONSUM MÀXIM D'UN DIA AL LLARG DE L'ANY	49
GRÀFIC 18 EVOLUCIÓ CONSUM SEGONS ACORN AGRUPAT	50
GRÀFIC 19 EVOLUCIÓ CONSUM SEGONS ACORN	51
GRÀFIC 20 EVOLUCIÓ CONSUM SEGONS ACORN	52
GRÀFIC 21 EVOLUCIÓ CONSUM SEGONS ACORN ADEQUANT GRUPS.....	53
GRÀFIC 22 EVOLUCIÓ CONSUM SEGONS ACORN I TARIFA	54
GRÀFIC 23 EVOLUCIÓ CONSUM SEGONS TARIFA	55
GRÀFIC 24 A DALTA, GRÀFIC CONSUM MITJÀ PER LLAR PER DIA DE LA SETMANA, GRÀFIC CONSUM MITJÀ PER LLAR PER DIA DE LA SETMANA PINTAT SEGONS ACORN GROUPED I TARIFA, GRÀFIC CONSUM MITJÀ DIARI PER DIA DE LA SETMANA I ESTACIÓ PINTAT SEGONS ACORN GROUPED. A SOTA, GRÀFIC CONSUM MITJÀ PER LLAR PER DIA DE LA SETMANA SEGONS ESTACIÓ PINTAT SEGONS ICON I GRÀFIC CONSUM MITJÀ DIARI PER DIA DE LA SETMANA PINTAT SEGONS ACORN AGRUPAT I DIA FESTIU	56
GRÀFIC 25 CONSUM DIARI MITJÀ PER LLAR SEGONS MES DE L'ANY	56
GRÀFIC 26 CONSUM DIARI MITJÀ PER LLAR SEGONS ACORN	57
GRÀFIC 27 EVOLUCIÓ CONSUM MITJÀ DIARI PER LLAR SEGONS ACORN (LLARS QUE SÓN AFFLUENT).....	57
GRÀFIC 28 A DALTA A L'ESQUERRA, GRÀFIC CONSUM DIARI MITJÀ PER LLAR SEGONS HORA DE POSTA DE SOL PINTAT SEGONS ACORN AGRUPAT I TARIFA , A DALTA A LA DRETA, GRÀFIC CONSUM DIARI MITJÀ PER LLAR SEGONS HORA DE POSTA DE SOL PINTAT SEGONS ACORN AGRUPAT. A SOTA A L'ESQUERRA, GRÀFIC CONSUM DIARI MITJÀ PER LLAR SEGONS HORA DE POSTA DE SOL I PINTAT SEGONS TARIFA, A SOTA A LA DRETA, GRÀFIC CONSUM DIARI MITJÀ PER LLAR SEGONS HORA DE POSTA DE SOL PINTAT SEGONS ESTACIÓ.....	58
GRÀFIC 29 CONSUM DIARI MITJÀ PER LLAR SEGONS ESTACIÓ, ACORN AGRUPAT I TARIFA	59
GRÀFIC 30 CONSUM DIARI MITJÀ SEGONS TEMPERATURA MÀXIMA I MÍNIMA PINTAT EN FUNCió DE L'ESTACIó.....	59
GRÀFIC 31 GRÀFIC CONSUM DIARI MITJÀ SEGONS VARIS INDICADORS (DEWPoINT, APPARENTTEMPERATUREMAX, APPARENTTEMPERATUREMIN, PRESSURE, VISIBILITY, WINDBEARING, WINDSPEED, HUMIDITY, CLOUDCOVER I MOONPHASE) PINTAT EN FUNCió D'ESTACIó, Gràfic CONSUM DIARI MITJÀ SEGONS VARIS INDICADORS PINTAT EN FUNCió DE LA TARIFA I Gràfic CONSUM DIARI MITJÀ SEGONS VARIS INDICADORS PINTAT EN FUNCió DE ACORN AGRUPAT.....	61
GRÀFIC 32 EVOLUCIÓ CONSUM MITJÀ PER LLAR I VARIES VARIABLES DE TEMPERATURA.....	62
GRÀFIC 33 EVOLUCIÓ NOMBRE LLARS DATASET	62
GRÀFIC 34 CONSUM MITJÀ MIG-HORARI (CADA MITJA HORA) PER LLAR DURANT EL DIA.....	63
GRÀFIC 35 CONSUM MITJÀ MIG-HORARI PER LLAR DURANT EL DIA PINTAT SEGONS ESTACIó	64
GRÀFIC 36 CONSUM MITJÀ MIG-HORARI PER LLAR DURANT EL DIA PINTAT SEGONS DIA DE LA SETMANA.....	64

GRÀFIC 37 CONSUM MITJÀ MIG-HORARI PER LLAR DURANT EL DIA PINTAT SEGONS TARIFA.....	65
GRÀFIC 38 CONSUM MITJÀ MIG-HORARI PER LLAR DURANT EL DIA PINTAT SEGONS ACORN AGRUPAT.....	65
GRÀFIC 39 CONSUM MITJÀ MIG-HORARI PER LLAR DURANT EL DIA PINTAT SEGONS ACORN	66
GRÀFIC 40 CONSUM MITJÀ MIG-HORARI PER LLAR AL LLARG DEL DIA SEGONS DIA DE LA SETMANA, MES, ESTACIÓ I ACORN... 68	68
GRÀFIC 41 CONSUM MITJÀ MIG-HORARI PER LLAR DURANT EL DIA PINTAT AL LLARG DELS MESOS DE LA MOSTRA	68
GRÀFIC 42 CONSUM MITJÀ MIG-HORARI PER LLAR DURANT EL DIA PINTAT AL LLARG DE LES SETMANES DE LA MOSTRA.....	69
GRÀFIC 43 TARIFES TIME OF USE APLICADES AL LLARG DEL 2013 PER DIA I HORA	70
GRÀFIC 44 CONSUM HORARI REAL I PREDIT DE 14 DIES	75
GRÀFIC 45 ERROR PERCENTUAL DE LA PREDICCIÓ HORÀRIA	76
GRÀFIC 46 CONSUM MITJÀ PER LLAR DEL 15-03-2013 AL 27-02-2014.....	77
GRÀFIC 47: PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM UNIVARIATE UTILITZANT 7 DIES PASSATS ONE-STEP AMB WALK-FORWARD 365 DIES.....	77
GRÀFIC 48: ERRORS MODEL UNIVARIATE UTILITZANT 7 DIES PASSATS ONE-STEP AMB WALK-FORWARD 365 DIES	78
GRÀFIC 49 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MAX) ONE-STEP AMB WALK-FORWARD 365 DIES.....	78
GRÀFIC 50 ERRORS MODEL MULTIVARIATE (TEMPERATURA MAX) ONE-STEP AMB WALK-FORWARD 365 DIES	79
GRÀFIC 51 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MAX I HUMITAT) ONE-STEP AMB WALK-FORWARD 365 DIES.....	79
GRÀFIC 52 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MAX, HUMITAT, VISIBILITY I CLOUDCOVER) ONE-STEP AMB WALK-FORWARD 365 DIES.....	80
GRÀFIC 53 RMSE MODEL MULTIVARIATE (TEMPERATURA MAX, HUMITAT, VISIBILITY I CLOUDCOVER) ONE-STEP AMB WALK-FORWARD 365 DIES	81
GRÀFIC 54 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MAX I HUMITAT DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365 DIES.....	81
GRÀFIC 55 RMSE MODEL MULTIVARIATE (TEMPERATURA MAX I HUMITAT DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365 DIES	82
GRÀFIC 56 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MAX I HUMITAT DELS 28 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365.....	82
GRÀFIC 57 RMSE MODEL MULTIVARIATE (TEMPERATURA MAX I HUMITAT DELS 28 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365.....	83
GRÀFIC 58 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA I HUMITAT DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	83
GRÀFIC 59 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA I HUMITAT DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	84
GRÀFIC 60 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365 DIES.....	84
GRÀFIC 61 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I SUNSETTIMEHOUR DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	85
GRÀFIC 62 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I SUNSETTIMEHOUR DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	85
GRÀFIC 63 ERRORS.....	86
GRÀFIC 64 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I PRESSURE DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	86
GRÀFIC 65 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I PRESSURE DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	87
GRÀFIC 66 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I PRESSURE DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	87
GRÀFIC 67 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I DEWPOINT DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	88
GRÀFIC 68 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I DEWPOINT DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	88
GRÀFIC 69 ERROR MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I DEWPOINT DELS 14 DIES ANTERIOR) ONE-STEP AMB WALK-FORWARD 365	89

GRÀFIC 70 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	89
GRÀFIC 71 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	90
GRÀFIC 72 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	90
GRÀFIC 73 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	91
GRÀFIC 74 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	91
GRÀFIC 75 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	92
GRÀFIC 76 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	92
GRÀFIC 77 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	93
GRÀFIC 78 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	93
GRÀFIC 79 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	94
GRÀFIC 80 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ, SUNSETTIMEHOUR I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	94
GRÀFIC 81 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ, SUNSETTIMEHOUR I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	95
GRÀFIC 82 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	95
GRÀFIC 83 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	96
GRÀFIC 84 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	96
GRÀFIC 85 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ, SUNSETTIMEHOUR I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	97
GRÀFIC 86 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ, SUNSETTIMEHOUR I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	97
GRÀFIC 87 ERROR MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT , PRESSIÓ, SUNSETTIMEHOUR I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	98
GRÀFIC 88 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	98
GRÀFIC 89 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	99
GRÀFIC 90 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365	99
GRÀFIC 91 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I PRESSIÓ DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365 CANVIANT L'OPTIMITZADOR	100
GRÀFIC 92 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT I PRESSIÓ DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD 365 CANVIANT L'OPTIMITZADOR.....	100
GRÀFIC 93 PREDICCIÓS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN	101

GRÀFIC 94 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA I HUMITAT DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN	101
GRÀFIC 95 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	102
GRÀFIC 96 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	102
GRÀFIC 97 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	103
GRÀFIC 98 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	103
GRÀFIC 99 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	104
GRÀFIC 100 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I PRESSIÓ DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	104
GRÀFIC 101 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I PRESSIÓ DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	105
GRÀFIC 102 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	105
GRÀFIC 103 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	106
GRÀFIC 104 ERROR MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	106
GRÀFIC 105 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT PRESSURE I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	107
GRÀFIC 106 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT PRESSURE I CLOUDCOVER DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	107
GRÀFIC 107 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT PRESSURE I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	108
GRÀFIC 108 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT PRESSURE I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	108
GRÀFIC 109 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT PRESSURE, CLOUDCOVER I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	109
GRÀFIC 110 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT PRESSURE, CLOUDCOVER I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	109
GRÀFIC 111 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT, PRESSURE, WEEKDAY I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	110
GRÀFIC 112 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT, PRESSURE, WEEKDAY I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	110
GRÀFIC 113 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT, PRESSURE, WEEKDAY I SUNSETTIMEHOUR DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	111
GRÀFIC 114 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT, WEEKDAY I MONTH DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	111
GRÀFIC 115 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT, WEEKDAY I MONTH DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	112
GRÀFIC 116 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍMINA, HUMITAT, WEEKDAY I SEASON DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	112

GRÀFIC 117 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT, WEEKDAY I SEASON DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	113
GRÀFIC 118 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT, WEEKDAY, SEASON I MONTH DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	113
GRÀFIC 119 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT, WEEKDAY, SEASON I MONTH DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	114
GRÀFIC 120 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT, WEEKDAY I PRESSIÓ DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	114
GRÀFIC 121 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA, HUMITAT, WEEKDAY I PRESSIÓ DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	115
GRÀFIC 122 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	115
GRÀFIC 123 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	116
GRÀFIC 124 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	116
GRÀFIC 125 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ, SUNSETTIMEHOUR I CLOUDCOVER) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	117
GRÀFIC 126 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ, SUNSETTIMEHOUR I CLOUDCOVER) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	117
GRÀFIC 127 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ, SUNSETTIMEHOUR I CLOUDCOVER) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	118
GRÀFIC 128 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ, SUNSETTIMEHOUR I CLOUDCOVER) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	118
GRÀFIC 129 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ, SUNSETTIMEHOUR , CLOUDCOVER I WEEKDAY) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	119
GRÀFIC 130 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ, SUNSETTIMEHOUR , CLOUDCOVER I WEEKDAY) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	119
GRÀFIC 131 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ I SUNSETTIMEHOUR) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	120
GRÀFIC 132 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ I SUNSETTIMEHOUR) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	120
GRÀFIC 133 PERCENTATGE ERROR PREDICCIÓ SEGONS VALOR REAL MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS JUNTAMENT AMB VARIABLE EXTRETA DE PCA DE PRESSIÓ I SUNSETTIMEHOUR) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN.....	121
GRÀFIC 134 COMPARACIÓ ERRORS PREDICCIÓ AMB EL MOMENT DE L'ERROR MILLOR MODEL	121
GRÀFIC 135 PREDICCIONS DEL DIA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN UTILITZANT TANT SOLS LES DADES DE ACORN A.....	122
GRÀFIC 136 RMSE MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT TRAIN UTILITZANT TANT SOLS LES DADES DE ACORN A.	123
GRÀFIC 137 PREDICCIONS DELS 7 DIES SEGÜENTS I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT	128

GRÀFIC 138 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT. PREDICCIÓ A 7 DIES	129
GRÀFIC 139 PREDICCIÓNS DELS 30 DIES SEGÜENTS I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT.....	131
GRÀFIC 140 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT. PREDICCIÓ A 30 DIES	133
GRÀFIC 141 PREDICCIÓNS DELS 60 DIES SEGÜENTS I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT.....	137
GRÀFIC 142 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT. PREDICCIÓ A 60 DIES	139
GRÀFIC 143 PREDICCIÓ FINS A 7 DIES SEGÜENTS I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT.....	144
GRÀFIC 144 ERRORS MODEL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT. PREDICCIÓ A 7 DIES	145
GRÀFIC 145 PREDICCIÓNS DELS 30 DIES SEGÜENTS I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT.....	147
GRÀFIC 146 ERROR PERCENTUAL PREDICCIÓ A 30 DIES	148
GRÀFIC 147 PREDICCIÓ A 60 DIES	152
GRÀFIC 148 ERROR PREDICCIÓ A 60 DIES.....	153
GRÀFIC 149 PREDICCIÓNS DE LA SETMANA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA DE LES 7 SETMANES ANTERIORS) ONE-STEP AMB WALK-FORWARD ON TRAIN NO AUGMENTA	155
GRÀFIC 150 ERRORS DEL MODEL SETMANAL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA DE LES 7 SETMANES ANTERIORS) ONE-STEP AMB WALK-FORWARD ON TRAIN NO AUGMENTA.....	156
GRÀFIC 151 PREDICCIÓNS DE LA SETMANA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DE LES 7 SETMANES ANTERIORS) ONE-STEP AMB WALK-FORWARD ON TRAIN NO AUGMENTA.....	156
GRÀFIC 152 ERRORS DEL MODEL SETMANAL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DE LES 7 SETMANES ANTERIORS) ONE-STEP AMB WALK-FORWARD ON TRAIN NO AUGMENTA.....	156
GRÀFIC 153 PREDICCIÓNS DE LA SETMANA SEGÜENT I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DE LES 4 SETMANES ANTERIORS) ONE-STEP AMB WALK-FORWARD ON TRAIN NO AUGMENTA.....	157
GRÀFIC 154 ERRORS DEL MODEL SETMANAL MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA I HUMITAT DE LES 4 SETMANES ANTERIORS) ONE-STEP AMB WALK-FORWARD ON TRAIN NO AUGMENTA.....	157
IMATGE 1 VARIABLES DEL DATASET	30
IMATGE 2 ESTADISTICS FITXER DAILY TOTAL	31
IMATGE 3 ESTADISTICS INFORMATIONS_HOUSEHOLDS	32
IMATGE 4 ESTADÍSTICS WEATHER_DAILY	32
IMATGE 5 ESTADÍSTICS HOLIDAYS	32
IMATGE 6 ESTADÍSTICS DAILY	33
IMATGE 7 ESTADÍSTICS WEATHER DAILY	33
IMATGE 8 ESTADÍSTICS WEATHER DAILY	33
IMATGE 9 ESTADÍSTICS WEATHER DAILY	34
IMATGE 10 ESTADÍSTICS WEATHER DAILY	34
IMATGE 11 ESTADÍSTICS WEATHER DAILY	35
IMATGE 12 T-TEST CONSUM SEGONS TARIFA	38
IMATGE 13 DADES AGRUPADES PER ACORN_GROUPED	39
IMATGE 14 DADES AGRUPADES PER ACORN	39
IMATGE 15 CONSUM SEGONS TARIFA PER ACORN I CONSUM SEGONS ACORN PER TARIFA.....	40
IMATGE 16 TEST DE TUKEY CONSUM SEGONS DIA DE LA SETMANA.....	42
IMATGE 17 RMSE	126
IMATGE 18 ERROR PERCENTUAL.....	127

IMATGE 19 RMSE PREDICCIÓ A 30 DIES	130
IMATGE 20 ERRORS MODEL PREDICCIONS DELS 30 DIES SEGÜENTS I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT	133
IMATGE 21 RMSE 60 DIES	134
IMATGE 22 ERRORS MODEL PREDICCIONS DELS 60 DIES SEGÜENTS I CONSUM REAL D'UN ANY UTILITZANT XARXA NEURONAL LSTM MULTIVARIATE (TEMPERATURA MÀXIMA I MÍNIMA , HUMITAT I WEEKDAY DELS 14 DIES ANTERIORS) ONE-STEP AMB WALK-FORWARD AUGMENTANT	140
IMATGE 23 RMSE PREDICCIÓ 7 DIES	140
IMATGE 24 ERROR PREDICCIÓ 7 DIES.....	140
IMATGE 25 RMSE A 30 DIES IMATGE 26 ERROR A 30 DIES.....	141
IMATGE 27 RMSE A 60 DIES IMATGE 28 ERROR A 60 DIES.....	142
IMATGE 29 RMSE A 7 DIES.....	143
IMATGE 30 ERROR A 7 DIES.....	143
IMATGE 31 RMSE A 30 DIES IMATGE 32 ERROR A 30 DIES	146
IMATGE 33 RMSE 60 DIES IMATGE 34 ERROR 60 DIES	149
IMATGE 35 RMSE 7 DIES	150
IMATGE 36 ERROR 7 DIES	150
IMATGE 37 RMSE 30 DIES IMATGE 38 ERROR A 30 DIES.....	150
IMATGE 39 RMSE A 60 DIES IMATGE 40 ERROR A 60 DIES.....	151

1. Introducció

1.1 Context i justificació de la rellevància

1.1.1 Context: Introducció al mercat elèctric

Perquè el mercat elèctric sigui possible, hi ha diverses empreses i institucions que hi participen, on els principals agents són les generadores d'electricitat que a partir d'ara anomenarem productores, les comercialitzadores d'energia, l'operador del mercat i l'operador del sistema.

També hi ha altres agents que no participen directament en el mercat de l'energia, però fan que aquest tingui sentit, com les distribuïdors d'energia, que són les que transmeten l'energia a través de la xarxa des de la seva generació fins al punt de consum i també cal mencionar la recent figura de l'agregador de demanda energètica, que explicarem més endavant.

Les generadores són aquelles centrals que transformen energia primària (força hidràulica, calor, força eòlica...) en energia elèctrica, on l'energia primària utilitzada pot ser de diferents tipus com nuclear (centrals nuclears), combustibles fòssils com carbó, petroli i gas natural (centrals tèrmiques o de cicle combinat), geotèrmica (central geotèrmica), hidroelèctrica, eòlica i solar (central fotovoltaica o solar tèrmica). Algunes d'elles tenen un gran impacte mediambiental, sobretot les nuclears pels residus i les tèrmiques de carbó i petroli per l'emissió de CO₂ i altres gasos contaminants a l'atmosfera.

El consum d'energia elèctrica ha anat augmentant des de la seva aparició¹, afavorint al canvi climàtic, per tant, cal trobar solucions per a centrar-se en les energies renovables i l'eficiència energètica dels consumidors per reduir la petjada ecològica.

Un cop generada l'electricitat, les productores normalment venen l'electricitat al mercat majorista.

El mercat majorista està regulat per l'operador del sistema, en el cas de Gran Bretanya se'n encarrega National Grid Electricity System Operator (National Grid ESO, 2019), que són els que equilibren l'oferta amb la demanda d'electricitat a temps real.

Per evitar desequilibris entre demanda i oferta, tenen mecanismes per a la regulació a partir de la producció (oferta) i només una mesura per part de la demanda: el servei d'interrumpibilitat, que es basa en poder sol·licitar reduccions del consum d'energia en hores pic a grans consumidors, alleugerant així el consum total en els moments que no hi ha suficient generació per cobrir la demanda, on aquests grans consumidors reben una contraprestació pel servei.

¹ Al Regne Unit ha augmentat el consum total un 12% del 1970 al 2016. (Waters, Energy Consumption in the UK. July 2017, 2017)

Aquí és on entra la figura de l'agregador de demanda que hem mencionat a l'inici, que són els agents que gestionen la demanda d'energia dels petits consumidors. Aquesta figura va aparèixer a partir de l'augment de l'autoconsum, on per exemple, en una llar amb autoconsum, si tenen excedents d'energia, l'han de vendre al mercat i si per el contrari els en falta, l'han de comprar. L'agregador té el coneixement per predir quan es produirà més energia d'autoconsum i, per tant, poder aconsellar quan aquesta llar podrà consumir més. Un altre exemple seria si la llar necessita més consum del que produeix, llavors aquest consum d'energia que necessitin de més caldrà que el comprin al mercat (normalment a través de la comercialitzadora) i l'agregador pot aconsellar sobre quan serà més barata, per tant, podran saber quan consumir. Resumint, l'agregador és el que gestiona la demanda eficientment, fent que el client final deixi de ser un consumidor passiu.

Continuant amb els agents del mercat elèctric, a Gran Bretanya hi intervé l'operador de mercat Ofgem (Office of Gas and Electricity Markets), que regula les empreses que gestionen les xarxes d'electricitat i gas, prenen decisions sobre el control i execució de preus i protegint els interessos dels consumidors (ofgem, 2019).

Finalment, les comercialitzadores són qui compren l'energia al mercat majorista (i d'altres) i la subministren a través de les distribuïdors al client final. Les comercialitzadores, a part de comprar l'energia al mercat majorista, poden tenir generadores pròpies o contractes bilaterals amb agents de fora del mercat (altres productors).

1.1.2 Justificació de la rellevància del tema

Com hem vist, les comercialitzadores necessiten saber amb antelació quins seran els consums elèctrics dels seus clients per a poder establir les estratègies més eficaces -si la generació pròpia no es suficient- per generar o comprar amb anterioritat la quantitat d'energia necessària i no haver de recórrer a la seva compra a últim moment², podent eliminar els costos per fer aquestes compres a últim moment – ja que si tens menys energia de la que has de subministrar, l'operador del sistema et penalitzarà-, és a dir, si saben l'electricitat que hauran de subministrar, la poden comprar amb anterioritat a un preu més barat o més verda³. També podran evitar que els quedi molta energia sobrant, ja que l'electricitat a grans quantitats no es pot emmagatzemar.

És rellevant destacar que hi ha moments del dia que es consumeix més, on hi ha un pic de consum -anomenada hora punta-, i cal que les productores d'energia estiguin

² Comprar a últim moment significa comprar a mercat amb el preu que hi hagi en aquell moment i sense poder saber qui ha generat aquella energia, és a dir, és deixa de controlar el que es compra.

³ Amb comprar energia més verda ens referim a que si ja saben que hauran de comprar una quantitat determinada d'energia, que la comprin al productor que menys influeixi al canvi climàtic, per exemple, enlloc de comprar electricitat transformada a partir del carbó, comprar-la a una generadora d'energia amb biomassa o, encara millor, una energia renovable com l'eòlica, la solar o la hidràulica.

preparades per a poder subministrar tal quantitat d'energia al mateix moment, com també hi ha moments on la demanda és mínima, el que s'anomena vall.

En aquest context, veiem que la gestió de la demanda és molt important, ja que aquesta varia influïda per l'horari dels comerços, el canvi horari (primavera-tardor), la jornada laboral, la meteorologia, calendari laboral, estacionalitat i fins i tot pel preu, de manera que és quelcom incert i volàtil, tot i que per exemple hi ha previsions molt precises d'algunes d'aquestes variables, com les temperatures, vent i precipitacions.

Per tant, saber la demanda que hi haurà permetrà saber la corba de càrrega, de manera que l'anàlitzarem per poder saber quan hi ha pics de consum i demanda total per poder buscar solucions per rebaixar els períodes de pic i augmentar els de vall amb mesures tarifàries (que les hores de vall siguin més barates que les de pic -tarifa Time of Use-) o la implantació d'energia solar domèstica per exemple.

En conclusió, el primer pas per a poder fer una gestió de la demanda és dur a terme la previsió de la demanda del consum i l'anàlisi dels patrons de consum. Les previsions de demanda són imprescindibles perquè el mercat elèctric pugui funcionar ja que permeten estimar els costos i planificar inversions, alhora que permeten prendre decisions amb coneixement. En definitiva, amb les prediccions i l'anàlisi volem aconseguir veure com podem fer que la demanda s'adapti a la generació a partir dels hàbits de consum dels diferents consumidors i alhora que la generació s'adapti a la demanda a partir de les prediccions, descongestionant les xarxes de distribució i millorant la petjada ambiental.

En poques paraules, volem fer la part de predicción que necessiten els agregadors per poder gestionar la resposta de la demanda d'electricitat dels consumidors domèstics (Gkatzikis, Koutsopoulos i Salonidis 2013).

Com hem vist, la predicción de consums és rellevant des dels consumidors residencials i agregadors per a poder gestionar la demanda fins a les comercialitzadores, distribuïdores, generadores i operador del sistema per poder preveure el consum que es necessitarà i poder planificar amb antel·lació.

1.1.3 Justificació de les dades que s'utilitzaran

En l'estudi realitzat pel departament d'energia i canvi climàtic del govern del Regne Unit titulat "Energy Consumption in the UK (2015)" (Waters, Wilkes, & GoodRight, 2015), veiem com al 1970 l'ús de l'electricitat en el sector industrial significava un 40% de l'energia total, el transport un 18% i el sector domèstic un 24%.

Al 2013 l'ús de l'electricitat del sector industrial havia baixat al 16% de l'energia total, l'ús del transport havia pujat al 38% i el consum domèstic significava un 27% del total d'energia.

En l'estudi més actual que trobem (Waters, Energy Consumption in the UK. July 2017), veiem que al 2016 el transport ja significava un 40% del total del consum final, el consum domèstic un 29% i la indústria un 17%.

Per tant sabem que la majoria del consum demandat és del consum del transport i domèstic.

En aquest projecte ens centrarem en el consum domèstic d'electricitat i realitzarem una predicción de consums de les llars de Londres i un anàlisi dels patrons de consum dels consumidors.

Per a fer la predicción de consum i l'anàlisi de dades, utilitzarem les dades de consums elèctrics dels últims anys de les llars de Londres, dades meteorològiques, dades de la població i fins i tot dades de festivitats, conjunts de dades que extraurem de <https://www.kaggle.com/jeanmidev/smart-meters-in-london>.

1.2 Descripció de la proposta

La proposta del treball és predir els consums elèctrics de les llars de Londres a curt, mitjà i llarg termini, on cada predicción té la seva utilitat:

- La predicción horària permet saber la demanda horària durant tot el dia, permetent la utilització de mètodes per variar-la, com per exemple la variació de preus en funció de l'hora del consum.
- La predicción diària, setmanal i mensual permet saber l'energia que cal comprar, permetent a les comercialitzadores prendre les decisions de compra i generació d'energia a priori, estalviant-los costos i haver de recórrer a la compra d'energia amb més impacte ambiental.
- La predicción anual permet saber el consum que hi haurà en el futur i que les distribuïdores puguin tenir la xarxa preparada per poder transportar-la.

A part de fer les descrites prediccions, també analitzarem les dades per a poder trobar altres solucions per a la gestió de la demanda del consum.

La hipòtesis principal és que podem predir amb precisió els consums de les llars i gràcies a això podrem trobar més mètodes per la gestió de la demanda energètica.

1.3 Objectius

L'objectiu principal és dur a terme la predicción del consum de Londres, on es vol predir el consum horari i diari.

Si el temps ho permet, també volem fer la predicción de consums setmanal, mensual i anual.

Volem aconseguir una precisió d'almenys el 85% en les prediccions a curt termini (horari i diari). També voldríem aconseguir que la precisió de la predicción del consum horari feta el dia abans arribi al 95%, ja que entenem que la precisió serà diferent si volem saber el consum horari del dia després que el consum horari al cap de 5 dies.

En la predicció setmanal volem aconseguir almenys un 80% de precisió, en la mensual un 75% de precisió i en l'anual (llarg termini) un 70%.

Les precisions s'han decidit en funció de la dificultat de la predicció, ja que per exemple a llarg termini no es disposen de les mateixes dades per predir que a curt termini (prediccions de temperatura, prediccions de generació ...).

Amb l'objectiu principal volem respondre algunes preguntes com: Podem saber amb antelació quan es consumirà més energia durant el dia? I menys? El dia abans podem saber amb exactitud l'energia horària que es consumirà? Podem saber quanta quantitat d'energia necessitem per un dia en concret? Podem saber la quantitat d'energia que necessitarem un mes en concret? Es consumirà més energia d'aquí 3 anys?

Els objectius secundaris són:

- Saber com afecten les diferents variables (meteorologia, tipus de consumidor, vacances, preu...) al consum.
- Trobar patrons en la corba de càrrega diària a través de contrast d'hipòtesis, visualitzacions i models de clustering.

Per exemple volem saber si l'aplicació de la tarifa Time of Use⁴ fa canviar els hàbits de consum, reduint els pics i augmentant el consum en períodes vall.

Amb els objectius secundaris, a part d'ajudar a conèixer les dades que tenim per poder fer uns millors models, també volem respondre algunes preguntes com: Quan es consumeix més durant el dia? El consum baixa si s'augmenta el preu? Hi ha diferències de consum entre els diferents tipus de consumidors? Com influeix la temperatura al consum?

1.4 Enfocament i metodologia

L'estratègia a seguir per a dur a terme el projecte comença per una recerca dels estudis que hi ha actualment per a la predicció del consum, tant en termes d'empreses com de residencials, per a poder saber els models que s'utilitzen per a predir tant a curt com a llarg termini.

Llavors, un cop establerts els objectius i metodologia en funció dels estudis trobats, es procedirà a fer una neteja de les dades i agrupar-les en un sol dataset (conjunt de dades). En aquesta etapa utilitzarem R.

Continuarem amb un ànalisi estadístic i gràfic de les dades que tenim i les relacions que hi ha entre elles. S'utilitzarà R per a fer-ho.

⁴ Tarifa que té uns preus predefinits per un períodes determinats d'hores. Per exemple, en les hores que hi ha pic de demanda, el preu és més car, influint així a la demanda del consumidor.

Tot seguit, un cop coneudes les dades i les seves relacions, procedirem a realitzar varíes prediccions de consum utilitzant varis models supervisats.

Hem de diferenciar les prediccions a curt termini de les de llarg termini, ja que, per exemple, a curt termini podem utilitzar dades meteorològiques ja que en podem obtenir prediccions fiables, en canvi a llarg termini no les podem utilitzar. També a curt termini utilitzarem les dades de consum de cada mitja hora en canvi a llarg termini utilitzarem les dades de consum diari. A part, a llarg termini caldria també afegir-hi dades de prediccions de l'economia i la demografia.

Els models que utilitzarem per a les prediccions a curt i mitjà termini són models que ja s'han utilitzat en altres estudis, intentant millorar alguns paràmetres:

- Xarxes neuronals (H. Hippert 2001), (Amjady 2006) , (Bakirtzis, et al. 1996), (Charytoniuk i Chen 2000).
- Models utilitzant Suport Vector Machines (SVM) (Chen, Chang i Lin 2004), (H.Min. i Lee 2005) .

Centrant-nos sobretot amb models de xarxes neuronals, ja que són capaços d'aproximar-se a qualsevol funció no lineal i alhora determinar el model mitjançant el procés d'aprenentatge, tot i que en alguns estudis, el model amb SVM els hi ha funcionat millor que les xarxes neuronals (Achanta 2012).

Si ens queda temps sobrant, també voldríem provar els models:

- Models de regressió lineal (Song, et al. 2005).
- Models de prediccions de sèries temporals Box-Jenkins (model ARIMA) (Nogales, et al. 2002) (Espinoza, et al. 2005) (Contreras, et al. 2003).

Utilitzarem els mateixos models pel llarg termini, que també ja s'han utilitzat en altres estudis com:

- Xarxes neuronals (Ghods i Kalantar 2011).
- SVM (Achanta 2012).

Per a cada model provarem diferents paràmetres i anirem afinant les prediccions, quedant-nos amb els que ens donin millors resultats sense sobre-entrenar la mostra.

Aquests models seran testats amb la mateixa mostra de dades (utilitzarem una mostra per l'entrenament i una altra pel test, comprovant que la distribució dels dos grups sigui similar per evitar crear un model esbiaixat).

Per avaluar els models utilitzarem l'error absolut mitjà (MAE i MAPE) i l'error quadràtic mitjà, que penalitza els residus més grans. Per a veure l'error en la mateixa escala que les dades, també utilitzarem l'arrel de l'error quadràtic mitjà. En aquesta fase s'utilitzarà Python.

Un cop realitzat l'anàlisi i conegeudes les prediccions i precisió d'aquestes, es procedirà a realitzar les conclusions i a respondre les preguntes formulades en les diferents hipòtesis dels objectius.

Finalment, es realitzarà una reflexió crítica sobre la realització del treball, des de la metodologia utilitzada fins als models utilitzats.

1.5 Motivació

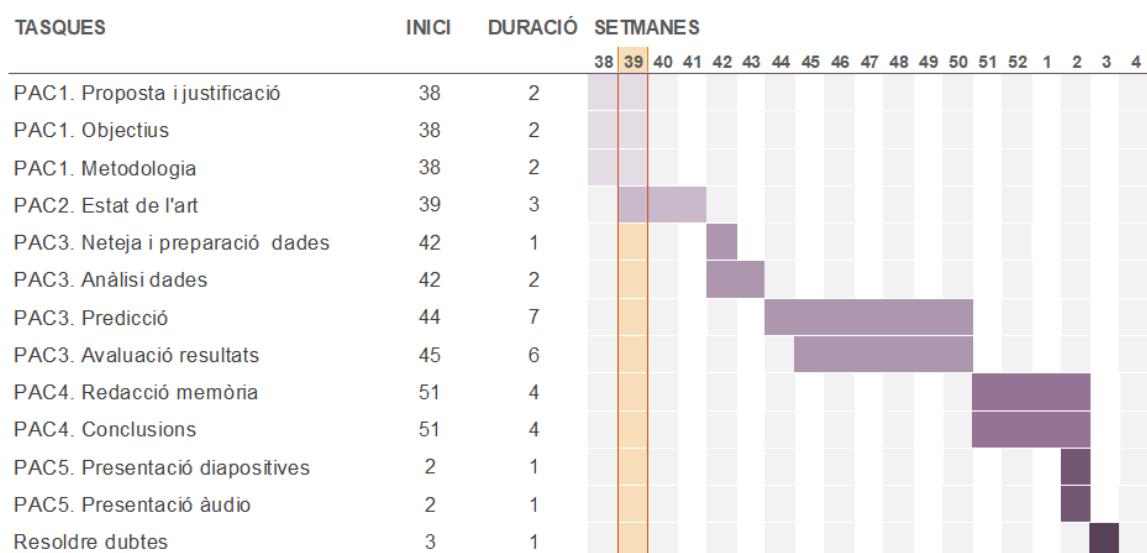
La motivació per a dur a terme aquest projecte es deu a dues principals raons: L'interès personal i l'experiència professional.

Després d'estar quatre anys treballant en una empresa d'enginyeria al departament d'eficiència energètica, on la principal funció era fer estudis elèctrics, per exemple d'optimització de potència i patrons de consum de grans empreses, ens varem interessar pel món de l'energia i en saber com funcionava.

Estar amb permanent contacte amb la previsió de preus a curt i llarg termini com també amb comercialitzadores i distribuïdores, va fer que ens preguntéssim com s'ho feien les comercialitzadores per predir la quantitat d'energia que havien de comprar , quan i a on la compraven, ja que moltes d'elles oferien tant sols energia verda, que és un punt a favor del medi ambient.

1.6 Planificació del Treball

Creem un diagrama de Gantt per poder planificar les tasques a realitzar, tal com veiem a la següent imatge:



Gràfic 1 Diagrama de Gantt

Les tasques es divideixen en 5 grans grups (cadascuna de les PAC).

En la primera part -PAC 1-, durant les dues primeres setmanes (38 i 39) realitzarem les següents tasques:

- Proposta i justificació: Explicarem la temàtica, posant-la en context i justificant la seva rellevància en la societat.
- Objectius: Donada la temàtica i la problemàtica a resoldre, determinarem què volem aconseguir i qui és l'abast.
- Metodologia: Explicarem l'estratègia a seguir, establint les tècniques i eines a utilitzar.

En la segona part -PAC 2-, de la setmana 39 a la 41, realitzarem la següent tasca:

- Estat de l'art: ens documentarem amb investigacions i estudis previs i explicarem els resultats que han aconseguit altres en la nostra línia de recerca.

En la tercera part -PAC 3-, li dedicarem de la setmana 42 a la 50, realitzant les següents tasques:

- Neteja i preparació de dades: Netejarem i prepararem les dades per al seu anàlisi, tractant valors buits i valors no vàlids.
- Anàlisi de dades: Analitzarem les dades i les relacions entre elles amb diferents tècniques.
- Predicció: Realitzarem la predicció a curt i a llarg termini del consum d'electricitat utilitzant diferents models.
- Avaluació de resultats: Avaluarem la precisió de cadascun d'ells i mirarem quins paràmetres es podrien canviar per obtenir millors resultats.

En la quarta part -PAC4-, a partir de la setmana 51 fins la setmana 2, realitzarem les següents tasques:

- Redacció memòria: aprofundirem en la redacció de la memòria, acabant de reescriure algunes parts i redactant els apartats que faltin.
- Conclusions: Escriptura de les conclusions extretes al llarg del projecte.

La cinquena part del projecte -PAC 5- la realitzarem la setmana 2 i conté :

- Presentació diapositives: Resum del treball en menys de 30 diapositives.
- Presentació àudio: Explicació de la presentació.

1.7 Breu descripció dels altres capítols de la memòria

En el capítol 2 – Estat de l'art – s'exposen els mètodes més utilitzats en la literatura per a la predicció del consum elèctric, començant per una visió global dels mètodes predictius utilitzats en l'electricitat i veient la diferència entre predicció a curt i predicció a llarg termini i finalment focalitzant en la predicció i comportament de les llars i exemples d'altres autors utilitzant el mateix conjunt de dades.

En el capítol 3 – Dades – es fa una descripció acurada dels grups de dades que tenim i les seves variables. Llavors es neteja i prepara cada dataset comprovant els valors de

les variables, els valors atípics i valors buits i tractant-los per al posterior analisi i prediccio, ja fent una pre-seleccio de les variables a utilitzar .

En el capítol 4 – Anàlisi de dades – es fa un anàlisi acurat de les dades que tenim i de les seves relacions. L'anàlisi consta primerament d'un anàlisi estadístic on es recompten les variables qualitatives i es veuen els principals estadístics de les variables quantitatives. També es miren les correlacions entre variables. Llavors, es fan diferents contrasts d'hipòtesis per respondre algunes preguntes com “el consum segons la tarifa és diferent?”, “el consum segons el dia és diferent?”. Després procedim amb l'anàlisi visual, on detectem patrons i relacions entre les variables, que ens serviran per seleccionar les variables per a la prediccio. Finalment també es comprova si hi ha alguna regla d'associació per a l'anàlisi de dependències entre variables.

En el capítol 5 – Predicció – s'utilitzen diferents models per predir el consum. En primer lloc es presenta el model a utilitzar per a la prediccio, LSTM i llavors es fan 3 tipus de prediccions: prediccio amb l'agregat horari, prediccio amb l'agregat diari i prediccio amb l'agregat setmanal. Per a cada agregat es proven diferents models on es canvien les variables per buscar el model que millor s'adapti a partir de l'RMSE.

En el capítol 6 – Conclusions – es fa una descripció de les conclusions del treball, una reflexió sobre l'assoliment d'objectius, un anàlisi de la planificació i metodologia i també s'exposen les línies de treball futur.

1.8 Breu sumari resultats obtinguts

En l'anàlisi de dades hem conclòs que de les variables que tenim, les que més expliquen el consum diàriament són el propi consum passat, el dia de la setmana, el mes i l'estació de l'any, el nombre de consumidors, el tipus de consumidor, la hora de posta de sol i factors relacionats amb la meteorologia com la temperatura apparent màxima i mínima, la humitat, la pressió, la cobertura dels núvols...

Sabent quines són les variables que més afecten al consum, hem desenvolupat models on introduíem diferents d'aquestes variables per a predir el consum futur. En la prediccio a una hora vista, hem obtingut un 2.83% d'error mitjà. En la prediccio a un dia vista, hem obtingut un 2.55% d'error mitjà. En la prediccio a 7 dies vista (una setmana), hem obtingut un error per la mitjana dels 7 dies de 4.21%. En la prediccio a 30 dies vista (un mes) hem obtingut un error mitjà pels 30 dies de 4.79%. En la prediccio a 60 dies vista (2 mesos) hem obtingut un error mitjà pels 60 dies de 6.05%

2. Estat de l'art

Molts mètodes han sigut utilitzats per a predir el consum d'electricitat, donada la importància que té l'eficient gestió de la demanda. Tot i haver-hi varietat de mètodes, en la majoria de literatura, per a mesurar la precisió del model utilitzat per a la predicción de demanda s'utilitza MAPE (Mean Absolute Percentage Error). En pocs casos hem trobat que utilitzin altres mètodes com MSE (mean square error) o RMSE (root mean square error), ja que penalitzen els residus més grans.

Els mètodes de predicción es poden classificar en dos grans grups: els mètodes estadístics i els mètodes d'intel·ligència artificial (Khan, et al. 2019) i les prediccions d'electricitat es poden classificar a grans trets per prediccions a curt termini i prediccions a llarg termini.

2.1 Literatura de mètodes estadístics

Alguns exemples dels mètodes estadístics utilitzats per a la predicción del consum són:

- Model de regressió lineal múltiple (MLR), utilitzat per varis autors com (Moghram i Rahman 1989), (Goia, May i Fusai 2010) i (Martínez 2015).
- Models econòmètrics com Markal-Time (Bianco, Manca i Nardini 2009).
- Models d'anàlisis de sèries temporals com el model ARIMA (Auto Regressive Integrated Moving Average).

Cal remarcar que dels models que trobem més literatura són els models ARIMA utilitzats per (R.E.Abdel-Aal i A.Z.Al-Garni 1997), (Conejo, et al. 2005), (Juberias, et al. 1999) i (Contreras, et al. 2003). També trobem extensions del model ARIMA, com ARIMAX⁵ utilitzat per (Goswami, Ganguly i Sil 2018) i (Arjmand, Samizadeh i Saryazdi 2018). També trobem models híbrids com ARMAX-GARCH⁶ utilitzat per (Tan, et al. 2010) .

2.2 Literatura de mètodes d'intel·ligència artificial

Dins el grup dels mètodes d'intel·ligència artificial els que destaquen en la literatura són els models Suport Vector Machine (SVM) , les Artificial Neural Networks (ANN) i les seves variants. Tant en els papers⁷ relacionats amb SVM i amb ANN veiem que tenen una precisió de predicción molt alta, i per tant són els models més utilitzats actualment d'entre tots els mètodes.

SVM és un model d'aprenentatge supervisat - és a dir, que requereix un conjunt de dades prèviament etiquetat – i permet resoldre problemes tant lineals com no lineals (gràcies a les funcions kernel) (Gironés Roig, et al. 2017). Trobem molta literatura en relació a la predicción de consums d'electricitat amb SVM, alguns exemples són:

⁵ ARIMAX: AutoRegressive Integrated Moving Average with eXogenous inputs.

⁶ ARMAX-GARCH: ARIMA - Generalized AutoRegressive Conditional Heteroskedastic.

⁷ papers: articles científics, en anglès.

- (Guo, Niu i Chen 2006) obtenen una precisió de més del 95% en la predicción mensual de tot un any utilitzant el kernel radial basis function (RBF).
- (Pandya 2018) compara 16 mètodes de predicción, on SVM és el que obté menys error. L'SVM utilitzat és SMOreg que utilitza el kernel polyKernel. En aquest cas no utilitzen dades meteorològiques, tant sols de consum.

Una de les variants utilitzades al SVM es el model LS-SVM (least squares suport vector Machine) utilitzat per varis autors (Huang, et al. 2019) , on en aquest paper (Kaytez, et al. 2015) obté més bons resultats que utilitzant el model amb ANN.

Tot i així, en el següent estudi (Theile, et al. 2018) on prediuen el consum a curt termini de cada mitja hora del dia següent, veiem que la xarxa neuronal recurrent (RNN), en aquest cas utilitzen el model NARX, té més precisió que SVM.

També trobem autors que comparen els resultats de SVM amb LS-SVM, on per exemple a (Wang i Yu 2008) s'obtenen més bons resultats utilitzant LS-SVM amb kernel RBF.

Les xarxes neuronals artificials són xarxes de neurones connectades a través d'enllaços. A cada neurona se li aplica una funció al valor de les seves entrades, produint un valor de sortida. La funció d'entrada combina les diferents entrades amb el seu pes i agrega tots els valors obtinguts de les connexions d'entrada per obtenir un valor (Gironés Roig, et al. 2017). Diferents exemples de la funció d'entrada són la funció suma ponderada, la funció màxim i la funció mínim. A part de la funció d'entrada, també s'aplica la funció d'activació per a poder dur a terme prediccions no lineals, i aquesta modifica el valor abans de passar-lo a la sortida. Alguns exemples de funció d'activació són la funció esgraó, la funció lineal, la funció sigmoide i la tangent hiperbòlica.

A part, les neurones es poden organitzar en diferents capes, podent modificar per a cada capa la funció d'entrada i d'activació.

Les connexions entre neurones poden ser de diferents tipologies:

- Feedforward Neural Networks (FNN): la informació sempre avança, la sortida d'una capa s'utilitza com a entrada de la següent.
- Recurrent Neural Networks (RNN): són possibles els bucles de retroalimentació

Per tant, veiem que les xarxes neuronals tenen molts tipus de models a utilitzar segons la combinació que es faci, i en la literatura en queda constància:

A (Asare-Bediako, Kling i Ribeiro 2013) utilitzen una xarxa neuronal per predir a curt termini utilitzant varies combinacions de variables i d'anys de consum, i queda plasmat que a més variables i a més anys d'historic de consum, més baix és l'error de la predicción. A (Osowski i Siwek 2018) comparen la precisió entre les xarxes neuronals MLP (Multi-Layer Perceptron), RBF (Radial Basis Function Network) i SVM amb un kernel Gaussià. La que té un error més baix és SVM i les dades utilitzades a part del consum són tipus de dia (laboral o cap de setmana) i l'estació.

Trobem molts estudis utilitzant les xarxes neuronals recurrents (RNN), on la més famosa i més utilitzada perquè funciona molt bé amb sèries de temps és la xarxa neuronal LSTM

(Long-Short Term Memory). Per exemple, en el següent estudi (Wang, et al. 2019) trobem que LSTM és el model que funciona millor, amb un 3% d'error, en comparació a una ANN (MLP) i SVM amb kernel RBF.

A (Hossen, et al. 2018) comparen l'error obtingut utilitzant varies RNN (RNN simple, LSTM i GRU -Gated Recurrent Unit-) i altres mètodes com ARIMA, GLM (Generalized linear model), random forest (RF) i SVM. Els resultats indiquen que dels altres mètodes el que té una millor precisió és SVM, tot i que no tant bona com la de les RNN. D'entre les RNN utilitzades, les que tenen una millor precisió són LSTM i GRU.

A (Almalaq 2019) comparen la precisió de la predicción diària i mensual de les xarxes neuronals recursives i no recursives. Les RNN utilitzades són LSTM i GRU i les ANN no recursives són RBFN (Radial Basis Function Network) i MLP (Multi Layer Perceptron). S'utilitza la funció d'activació sigmoide en tots els models menys en RBFN, que s'utilitza RBF. Tant en les prediccions del dia següent i de les prediccions del mes següent, les xarxes que obtenen millors resultats són les recursives.

No només és important el model a utilitzar sinó també els paràmetres a triar, ja que segons els paràmetres que s'utilitzin, per exemple els “learning rate” i “epochs”, s'obtindran uns resultats o uns altres, de manera que cal anar-los ajustant per obtenir un bon model que no estigui sobre-entrenat.

2.3 Diferència entre predicción a curt termini i predicción a llarg termini

Moltes variables afecten al consum elèctric : variables meteorològiques com la temperatura ambient, el cel cobert i la humitat, com altres variables com el tipus de dia (festiu o no), dia de la setmana , factors estacionals i fins i tot l'hora que es pon el sol. A curt termini es poden utilitzar totes aquestes variables (tant els valors passats com prediccions a curt termini) , però a llarg termini no podem utilitzar les variables meteorològiques, ja que ni tenim prediccions fiables per utilitzar ni les dades anteriors tenen una forta relació.

Trobem varis estudis que utilitzen models per predir la demanda d'electricitat a molt curt i a curt termini utilitzant les dades meteorològiques (Theile, et al. 2018) (Niu, Wang i Wu 2010) (Liu i Brown 2019) (Matsuo, et al. 2019) (Wang, Liu i Hong 2016). Per exemple, en el següent article (Gritsay, et al. 2019) trobem que a mesura que afegeixen variables al model, aquest millora la precisió.

Alguns articles científics han estudiat la influència de la meteorologia al consum i per exemple, els autors Zeng, et al. utilitzen SVM i ANN per integrar la correlació entre la meteorologia i el consum d'electricitat, aconseguint un 93% de precisió (Zeng, et al. 2016).

Tot i que en alguns models es millori la precisió amb les dades meteorològiques, a (Schachter i Mancarella 2014) veiem com la utilització de dades meteorològiques no millora la predicción amb la regressió lineal, ja que la relació entre temps i dades de

consum no és lineal. Sí que es millora la precisió al incloure les dades meteorològiques en la xarxa neuronal, ja que és capaç de modelar les relacions no lineals.

En el següent estudi (Twanabasu i Bremdal 2013), prediuen el consum a curt termini de la següent hora i comparen els 3 mètodes més utilitzats: ARIMA, SVM amb kernel radial i ANN amb la funció sigmoide a la capa oculta i funció lineal a la capa de sortida i la millor precisió l'obté l'ANN amb un 5.3% d'error seguit de l'ARIMA amb un 5.6% d'error i finalment l'SVM amb un 7,6% d'error.

Trobem molta més literatura sobre la predicción a curt termini (menys de dues setmanes) que a llarg, i això és degut a la gran importància i impacte que té la predicción a curt termini de la demanda d'electricitat, ja que, tal com hem vist a la introducció, una predicción de la demanda precisa permet l'eficient gestió de les comercialitzadores, distribuïdores, generadores d'electricitat i del sistema en general si parlem a gran escala i de la gestió de la demanda a petita escala (Wang, et al. 2019).

En quant al mitjà i llarg termini, els estudis que trobem ens indiquen la utilització de ARIMA, SVM i xarxes neuronals com LSTM.

Els autors (Agrawal, Muchahary i Tripathi 2018) obtenen un MAPE de 6.54 de mitjana per la predicción a 5 anys utilitzant LSTM.

Els autors (Sun 2010) obtenen un error del 3% en SVM i un error del 0.8% en un mètode híbrid que utilitza MG-CACO (mass recruitment and group recruitment based continuous ant colony optimization) per seleccionar els paràmetres de SVM.

Els autors (Mamun i Nagasaka 2004) obtenen un error del 3.46% utilitzant una xarxa neuronal, en concret RBFN (Radial Basis Function Network).

També és important comentar que la predicción a llarg termini és més difícil de realitzar amb precisió per múltiples factors com la influència de GDP (Gross Domestic Product), incidents naturals com tsunamis i volcans i incidents humans com crisis, inflació i inestabilitat política (Wang, et al. 2009).

2.4 Literatura predicción d'energia a llars

Els models predictius utilitzats enfocats al consum residencial són els mateixos exposats en els apartats anteriors, on predominen els estudis utilitzant SVM i ANN, ja que els consums residencials estan més influïts pels canvis de meteorologia i cal utilitzar models que ens permetin representar-ho.

A (Ihsane, et al. 2018) obtenen un error al voltant del 6% utilitzant una ANN, en concret MLP amb funció d'activació sigmoide i 3 capes (entrada, oculta i sortida), utilitzant un model pels dies laborals i un altre pels caps de setmana.

Un exemple d'estudi del comportament dels consumidors en funció del preu el trobem a (Yan, et al. 2015) on fan una predicción dels consums residencials de les futures 72h.

amb el model SVM per estudiar el comportament del consumidor als canvis de preu, on conclouen que ToU influeix el consumidor i que és una bona solució per la gestió de la demanda, podent fer disminuir els pics de consum.

Alguns dels articles exposats en apartats anteriors estan enfocats a la predicción residencial com els següents: (Almalaq 2019), (Asare-Bediako, Kling i Ribeiro 2013), (Cai, et al. 2019).

Trobem pocs estudis que utilitzin les dades dels consumidors per a fer prediccions, i a (Viegas, et al. 2015) fan una predicción dels segments del perfil de consum a partir de les dades del consum i les dades d'una encuesta (en aquest cas no utilitzen dades meteorològiques, però si estacionals). Utilitzen un model de regressió probit, un model SVM i un model GBT (Gradient Boosted Trees), on el que dona una millor precisió és SVM. També a (Hsiao 2015) fan una classificació del comportament per predir els diferents patrons de comportament de consum i les prediccions que obtenen utilitzant una BPN (back propagation neural network) són millors que altres models sense aquestes dades, com SVM, ARIMA i regressió lineal .

En relació al comportament dels consumidors trobem (Keitsch i Bruckner 2015) que estudia si la tarifa Time of Use influeix en el comportament dels consumidors i no pot verificar que sigui cert, al contrari del que s'ha trobat a altres estudis (Cosmo, Lyons i Nolan 2014).

A (Maigha i Crow 2014) proposen una metodologia basada en clustering per dissenyar correctament l'estructura dels Time of Use per obtenir un óptim rendiment.

També en relació als consumidors, a (Kwac, Flora i Rajagopal 2014) es fa una segmentació dels consumidors utilitzant els seus consums amb el mètode k-means per extreure patrons diaris de consum dels diferents grups de consumidors. Troben que el 87% de les mostres es poden resumir en 20 patrons, com per exemple pic al matí, pic a la nit, pic a la tarda. Gràcies a aquests patrons es poden saber quins seran els consumidors potencials per utilitzar programes de resposta a la demanda (DR : Demand Response). Per exemple els que tenen pic a la tarda, de 16h. a 22h., estant consumint a les hores de més consum, i per tant podrien ser objectius de DR. Un altre estudi de segmentació basada en el consum és (Sintov, Flora i Rajagopal 2013).

A (Wang, Chen, et al. 2015) es discuteixen varis mètodes de clustering per implementar l'aplicació de perfils de demanda per a la gestió de la resposta a la demanda.

2.5 Exemples utilització del nostre conjunt de dades

El conjunt de dades que utilitzarem, al ser unes dades obertes i penjades a Kaggle, altres autors les han utilitzat.

En primer lloc, l'usuari jeanmidev(Jean-Michel D.) de Kaggle, escriu un article en un bloc (D. 2017) on fa una descripció i exploració inicial del dataset. En l'article proposa algunes

idees per poder extreure coneixement com creuar les dades dels comptadors amb les de l'Acorn, predir el consum de les diferents llars i crear clústers dels perfils d'energia.

A Kaggle trobem dos apartats interessants: Kernels i Discussion, on a Kernels hi trobem codi que ha fet altres autors i a discussions varis tòpics i dubtes sobre les dades. A l'apartat Kernels, trobem clustering i anàlisis de les dades que han realitzat altres autors. A l'apartat Discussions, també hi ha un comentari de Jean-Michel D. on proposa altres idees com la segmentació del patró de consum i predicció a escala de Londres. L'autor Kaan Karabal també fa un anàlisi de les dades a (Karabal 2017).

3. Dades

En aquest apartat presentem els fitxers de dades que utilitzarem i prepararem les dades per al seu posterior anàlisi i predicció.

Tal com hem mencionat, les dades que utilitzarem estan extretes de Kaggle (<https://www.kaggle.com/jeanmidev/smart-meters-in-london>) publicades per Jean-Michel D. (usuari jeanmidev).

Aquestes dades, són una versió re-factoritzada de les dades de l'autoritat de Londres (Greater London Authority and UK Power Networks 2015) recollides per UK Power Networks.

Les dades consten de les lectures del consum d'energia d'una mostra de 5566 llars de Londres que van formar part del projecte "UK Power Networks led Low Carbon London" i les dades van ser enregistrades entre el Novembre del 2011 i el febrer del 2014 a través de Smart Meters (comptadors intel·ligents) .

En quant als consumidors, van ser seleccionats com una mostra equilibrada representativa de Londres, i hi trobem dos grups de consumidors.

El primer grup, 1123 consumidors, estan subjectes a preus d'electricitat dinàmics (Time of Use) durant el 2013. Els preus de la tarifa els donaven el dia abans a través del Smart Meter IHD (In Home Display) o de missatges de text. Les tarifes eren 3 (High - 67.20p/kWh-, Low -3.99p/kWh- i Normal -11.76p./kWh-) i en l'avís del dia anterior s'indicava el preu de les tarifes i les hores que s'aplicaria cadascuna.

L'altre grup de consumidors, els 4443 restants, eren clients que no estaven subjectes a la tarifa Time of Use i tenien una tarifa plana de 14.228 pence/kWh.

A la web de Kaggle hi trobem diferents datasets:

- Fitxers que contenen les dades horàries i diàries dels mencionats consumidors recollides per UK power networks (UK Power Networks 2019).
- Fitxer on trobem les dades del grup CACI Acorn (CACI Ltd. 2019) al que pertany cada consumidor. Acorn és una classificació de consumidors que segmenta la població de Regne Unit a partir de l'anàlisis de les dades demogràfiques, factors socials, població i comportament del consumidor. També ens trobem un fitxer amb els detalls de l'ACORN extret de la web de CACI Acorn.
- Fitxers dades horàries i diàries del temps (meteorologia) de l'àrea de Londres, que l'autor va recollir utilitzant l'api darksky (The Dark Sky Company, LLC 2012).
- Fitxer dies festius nacionals de Londres.

D'altra banda, a la pàgina original (Greater London Authority and UK Power Networks 2015) hi trobem també el fitxer amb les tarifes Time of Use i les hores d'aplicació de cadascuna.

Tot el codi generat en aquest apartat es troba en els fitxers annexos de R “Preparacio dades daily i total” i “Preparacio dades hourly”.

3.1 Descripció dels datasets

A continuació descriurem els datasets i les variables que contenen.

3.1.1 Informació de les llars (Informations households)

L’arxiu informations_households conté la informació relativa a la llar, és a dir, la tarifa que se li aplica, l’Acorn al que pertany i l’arxiu on trobem la informació dels consums d’aquesta llar. Les variables que tenim són:

- LCLid : Codi d’identificació de la llar.
- stdorToU: Tipus de tarifa (Std -tarifa plana- o ToU -tarifa Time of Use-). Factor.
- Acorn: Classificació Acorn a la que pertanyen els consumidors de la llar. Factor.
- Acorn_grouped: Classificació més global de l’Acorn (fusió de diferents grups). Factor.
- File: Nom de l’arxiu per trobar les dades de consum.

Aquest fitxer conté 5566 files (cada llar), on cada llar té un identificador diferent. En relació a la variable stdorToU, 4443 llars (79,82%) tenen tarifa "Std" i 1123 llars (20.17%) tenen tarifa "ToU" (Dynamic Time of Use).

La variable Acorn comprèn des de ACORN-, ACORN-A, ACORN-B ... fins ACORN-U (19 grups) i la variable Acorn_grouped té 5 atributs(ACORN-, ACORN-U, Adversity, Comfortable i Affluent), on Affluent és el que apareix amb més freqüència.

De la variable file, sabem que tenim 112 blocs, des del block_0 al block_111.

3.1.2 Temps horari i diari (Weather hourly , weather daily)

En l’arxiu weather_daily_darksky trobem les dades diàries en relació al temps com la temperatura màxima i mínima, velocitat del temps, humitat... Vegem les 32 variables i el seu tipus:

- temperatureMax : màxima temperatura del dia (C) . Número.
- temperatureMaxTime: dia i hora de la màxima temperatura del dia. Data (timestamp).
- windBearing: wind bearing. Número.
- icon: tipus de dia (clar, ennuvolat, emboirat, ventós...). Factor.
- dewPoint: Punt de rosada (temperatura a la qual, segons la pressió i l’humitat, l’aigua es condensa). Número.
- temperatureMinTime: dia i hora de la mínima temperatura del dia. Data (timestamp)
- cloudCover: percentatge de cel cobert per núvols. Número.
- windSpeed: velocitat del vent en km/h. Número.

- pressure: pressió atmosfèrica (mbar) . Número.
- apparentTemperatureMinTime: dia i hora de la minima temperatura aparent. Data (timestamp).
- apparentTemperatureHigh: alta temperatura aparent del dia. Número.
- precipType: tipus de precipitació (rain o snow). Factor.
- visibility: index de visibilitat. Número.
- humidity: index de humitat. Número.
- apparentTemperatureHighTime: dia i hora de l'alta temperatura aparent. Data (timestamp).
- apparentTemperatureLow: baixa temperatura aparent del dia. Número.
- apparentTemperatureMax: màxima temperatura aparent del dia. Número.
- uvIndex: index uv. Número.
- time: dia. Data.
- sunsetTime: dia i hora de la posta de sol. Data (timestamp).
- temperatureLow: Baixa temperatura del dia. Número.
- temperatureMin: Mínima temperatura del dia. Número.
- temperatureHigh Alta temperatura del dia. Número.
- sunriseTime: dia i hora de la sortida del sol. Data (timestamp).
- temperatureHighTime: dia i hora de l'alta temperatura. Data (timestamp)
- uvIndexTime: dia i hora de l'uv index. Data (timestamp)
- summary: Resum del dia (per exemple majoritàriament ennuvolat tot el dia, emboirat pel matí...). Cadena de text (String).
- temperatureLowTime: dia i hora de la baixa temperatura. Data (timestamp).
- apparentTemperatureMin: mínima temperatura aparent del dia. Número.
- apparentTemperatureMaxTime: dia i hora de la màxima temperatura aparent. Date (timestamp).
- apparentTemperatureLowTime: dia i hora de la baixa temperatura aparent. Date (timestamp).
- moonPhase: Fase de la lluna en percentatge. Número.

En l'arxiu weather_hourly_darksky trobem les dades diàries en relació al temps. En aquest cas tenim menys variables, 12:

- Visibility: Visibilitat. Número.
- windBearing: wind bearing. Número.
- temperature: temperatura. Número.
- Time: data i hora. Data (timestamp)
- dewPoint: Punt de rosada. Número.
- Pressure: Pressió atmosfèrica. Número.
- apparentTemperature: temperatura aparent. Número.
- windSpeed: velocitat del vent. Número.
- precipType: tipus de precipitació (pluja o neu). Factor.
- Icon: categorització dins les categories: clear-day, clear-night, cloudy, fog, wind, partly-cloudy-day i partly-cloudy-night. Factor.
- Humidity: Humitat. Número.

- Summary: Resum de l'hora en concret (majoritàriament ennuvolat, emboirat...).
Factor

3.1.3 Consum horari i diari (hourly dataset, daily dataset)

El consum horari el trobem en dos formats. Per una banda, tenim la carpeta hhblock_dataset que conté els blocs i per l'altra banda, la carpeta zip halfhourly_dataset que també conté blocs, però contenen diferents variables (estan organitzats de manera diferent).

En la primera, hhblock_dataset trobem una fila per cada dia i les variables són el consum de cada mitja hora, juntament amb el dia i l'identificador. En aquest cas, tenim el consum de totes les hores (mitjes hores) d'un dia d'una llar en una sola fila.

En la segona, halfhourly_dataset trobem una fila per cada dia i hora, per tant només conté les variables consum, dia (timestamp) i identificador : LCLid, tstamp i energy(kWh/hh). En aquest cas les hi ha 48 files per cada llar i dia (cada mitja hora).

Cadascun dels datasets té els seus pros i contres. Utilitzarem els dos, hhblock_dataset per calcular els valors buits de daily_dataset (la suma de totes les hores és el total del dia) i halfhourly_dataset per poder combinar amb les mesures del temps horàries.

El consum diari també el trobem en diferents formats. Per una banda tenim la carpeta daily_dataset que conté els blocs i l'arxiu daily_dataset.csv.

En la carpeta, en cada arxiu de block de daily_dataset trobem:
LCLid,day,energy_median,energy_mean,energy_max,energy_count,energy_std,energy_sum,energy_min.

En l'arxiu daily_dataset.csv trobem:

LCLid,day,energy_median,energy_mean,energy_max,energy_count,energy_std,energy_sum,energy_min.

Al examinar les dades carregades veiem que contenen la mateixa informació, per tant només utilitzarem un d'aquests datasets. En concret utilitzarem el que ja ens vé tot l'arxiu complet.

La variable energy_count és el nombre de vegades que el comptador ha recomptat energia, sent 48 el nombre de vegades que ho fa. En alguns casos no és 48 degut a que el comptador no funciona correctament, perquè la llar s'ha incorporat a la mostra a meitat del dia o ve per altres raons que desconeixem.

L'energy_sum és el consum diari d'una llar.

La variable energy_mean es refereix a la mitjana del consum mighorari d'un dia, és a dir, és el valor resultant de dividir la suma del que ha consumit una llar durant el dia entre

48 (hi ha 48 mitges hores durant el dia) o ve entre les vegades que s'ha comptat energia al dia si s'ha fet menys de 48 vegades (energy_count).

Les variables energy_max i energy_min reflecteixen el valor de consum de la mitja hora en que s'ha consumit més i menys respectivament.

3.1.4 Dies festius (UK bank holidays)

En aquest arxiu hi trobem els dies festius i el seu nom. Les variables que tenim són:

- Bank holidays: data del dia festiu. Data.
- Type: nom del dia festiu. Cadena de caràcters.

3.1.5 Tarifes (tariffs)

L'arxiu Tariffs conté les hores a les quals s'aplicava cada tarifa en el cas de Time of Use. El registre és cada mitja hora. Tantsols conté 2 variables:

- TariffDateTime: Data i hora.
- Tariff: Tipus de tarifa aplicable (Normal, High o Low).

3.2 Neteja i preparació de les dades

En aquest apartat realitzem una neteja i preparació de les dades per a cada dataset.

En primer lloc, revisem el tipus de variables que tenim i si corresponen amb el que els hi toca, i si no es correspon, llavors convertim el tipus de variable, per exemple revisem que les temperatures siguin numèriques, que les descripcions siguin caràcters i que les llistes de valors siguin factors.

També buscarem que no hi hagin errors a les dades i inconsistències com per exemple que hi hagi una temperatura de 200 graus. També procurarem que les variables estiguin normalitzades, per exemple en la variable Acorn, que no tinguem Acorn-A i Acorn-a, que seria el mateix valor, cal agrupar-los en un sol nom.

Per a fer-ho, ens ajudarem d'un estudi descriptiu de cadascuna de les variables.

Llavors, farem un estudi de valors atípics i buscarem si hi ha presència de valors buits per poder-los tractar.

Un cop realitzada la neteja en tots els grups de dades que tenim, els unirem tots en un sol dataset, eliminant les variables que no necessitem.

3.2.1 Informations_households

En primer lloc, revisem el tipus de variables que tenim. Al carregar-les a R, hem sol·licitat que no es doni per suposat que totes les cadenes són factors per fer-ho manualment. Veiem doncs que totes les variables són caràcter i té sentit discretitzar per a les variables stdorToU, Acorn i Acorn_grouped i file que contenen 2, 19 , 5 i 112 atributs respectivament. Discretizem les variables escollides perquè siguin factors.

Revisem el resum estadístic :

LCLid	stdorToU	Acorn	Acorn_grouped	file
Length:5566	Std:4443	ACORN-E:1567	ACORN- : 2	block_0 : 50
Class :character	ToU:1123	ACORN-Q: 831	ACORN-U : 49	block_1 : 50
Mode :character		ACORN-F: 684	Adversity :1816	block_10 : 50
		ACORN-H: 455	Affluent :2192	block_100: 50
		ACORN-L: 342	Comfortable:1507	block_101: 50
		ACORN-D: 292		block_102: 50
		(Other):1395		(Other) :5266

I el resum d'atributs de l'Acorn:

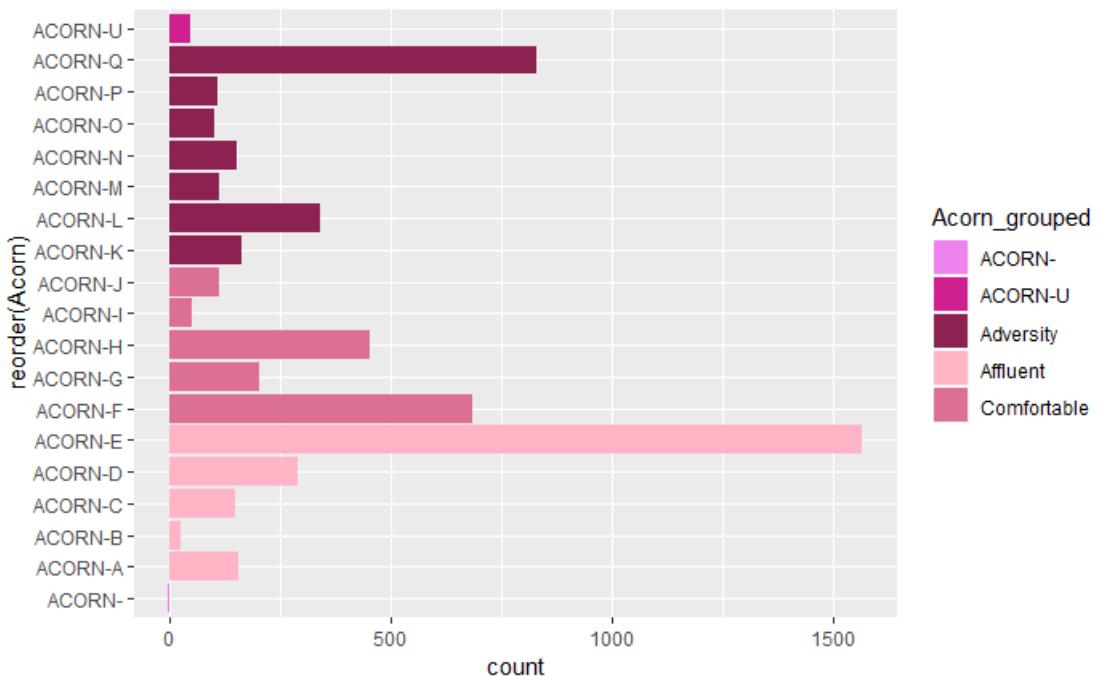
ACORN-	ACORN-A	ACORN-B	ACORN-C	ACORN-D	ACORN-E	ACORN-F	ACORN-G	ACORN-H
2	157	25	151	292	1567	684	205	455
ACORN-I	ACORN-J	ACORN-K	ACORN-L	ACORN-M	ACORN-N	ACORN-O	ACORN-P	ACORN-Q
51	112	165	342	113	152	103	110	831
ACORN-U								
49								

Ens adonem que ACORN-U i ACORN- de Acorn_grouped han de ser errors, ja que en la descripció de les dades original ens diuen que tenim 3 tipus d'Acorn_grouped, de manera que els hem de re-classificar dins d'una de les 3 classes. En l'estudi realitzat per l'autor de les dades a Kaggle, comenta que les llars de la mateixa ACORN estan en la mateixa fila. Amb aquestes dades, estudiem la relació que hi ha entre Acorn, Acorn_grouped i file per poder re-classificar els valors amb errors.

També ha de ser un error el valor ACORN- de la variable Acorn, ja que tant sols hi ha 2 llars d'aquest tipus.

Per comprovar quins valors haurien de tenir, fem taules de freqüències, tot i que ho veiem millor amb els gràfics:

Gràfic per saber a quin Acorn_grouped pertany cada Acorn:

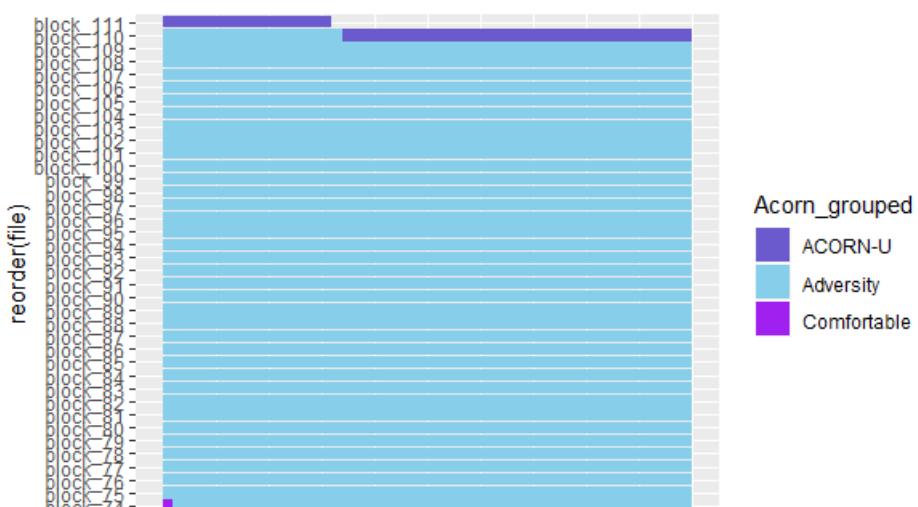


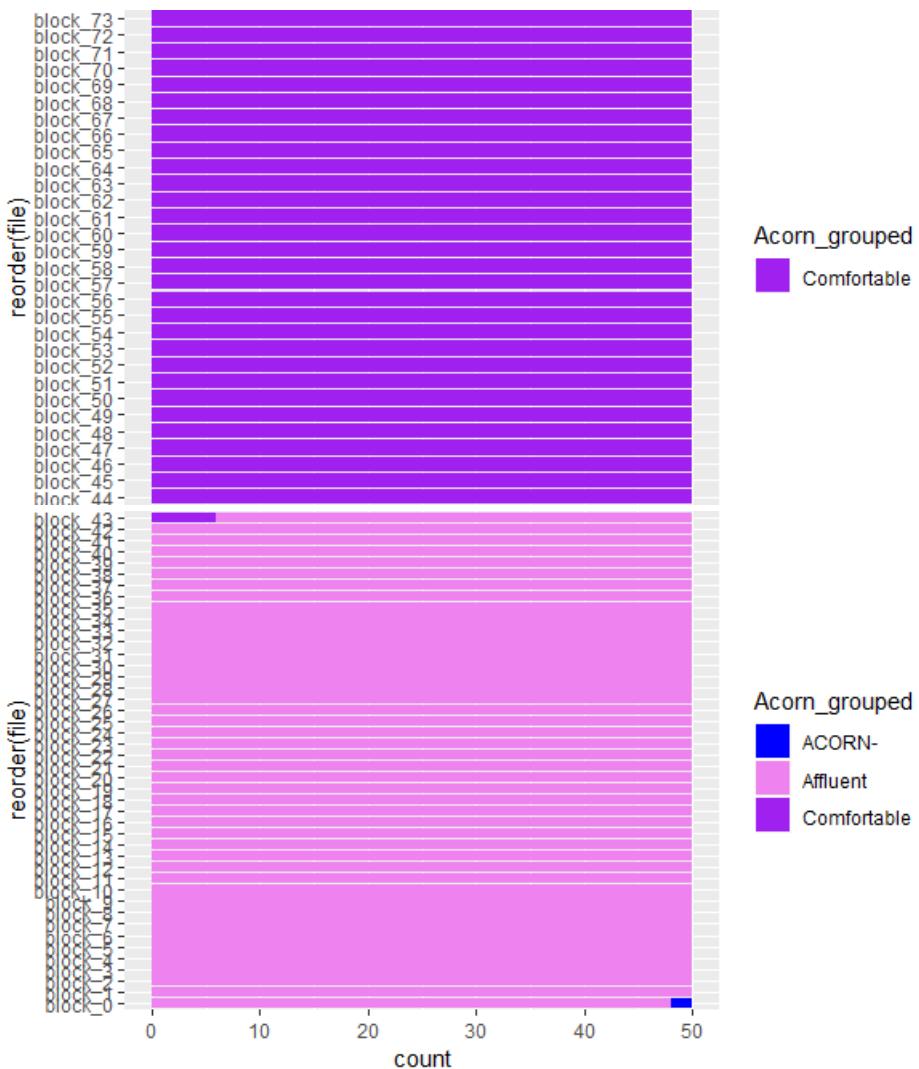
Gràfic 2 Recompte de llars segons Acorn, pintat segons Acorn grouped

En el gràfic 2 veiem que el valor ACORN- de Acorn grouped correspon a l'Acorn ACORN- i els ACORN-U de Acorn_grouped corresponen a l'Acorn ACORN-U.

També veiem que els primers Acorn consecutius (Acorn-A, Acorn-B, Acorn-C, Acorn-D i Acorn-E) pertanyen a Affluent, els següents (de l'Acorn-F a l'Acorn-J) pertanyen a Comfortable i els restants (de Acorn-K a Acorn-Q) pertanyen a Adversity. Això ens fa sospitar que "ACORN-" serà Affluent i "ACORN-U" serà Adversity.

Gràfic per saber a quin Acorn_grouped pertany cada block (file):

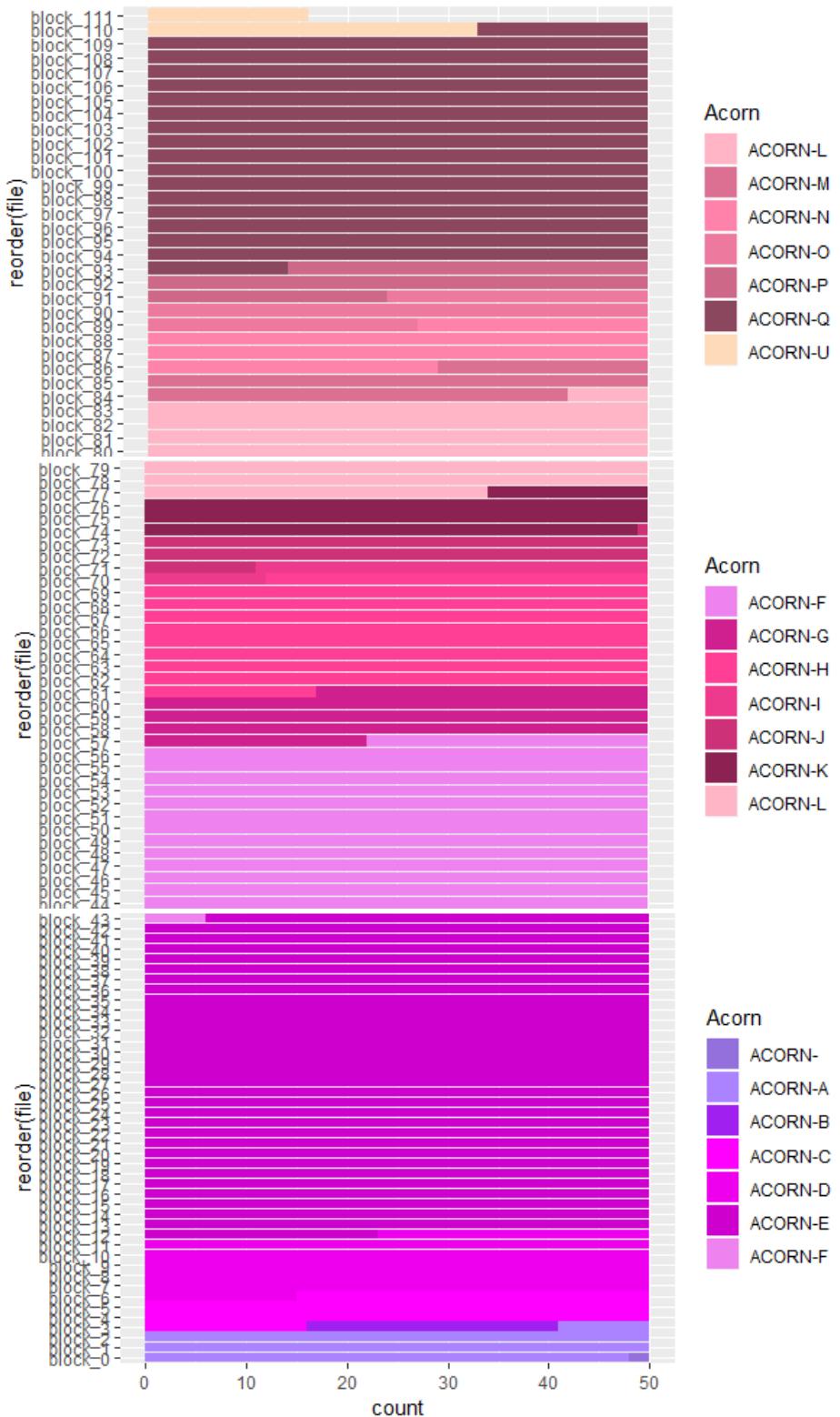




Gràfic 3 Recompte de usuaris per block, pintat segons Acorn_grouped

En el gràfic 3 veiem que els primers blocs (0 al 43) contenen Affluent , els següents(del 43 al 74) contenen comfortable i els últims (74 a 110) contenen Adversity. Això també ens fa pensar i confirmar que Acorn- és Affluent i Acorn-U és Adversity.

Gràfic per saber quin Acorn pertany a cada block (file):



Gràfic 4 Recompte de llars per bloc pintat segons Acorn

En el gràfic 4 veiem que a mesura que augmenta l'Acorn, augmenta el número de block, és a dir els de l'Acorn A estan al bloc_0 , block_1 i block_2, els de l'Acorn B estan al block_3 els de l'Acorn C al block_3, block_4, block_5 i block_6 i així successivament.

Per lògica doncs, intuïm que els de l'Acorn- que estan al block 0 han de ser ACORN-A. I com que els de l'ACORN-A són Affluent, per lògica entenem que aquests valors també ho seran.

El mateix amb els valors de l'ACORN-U, que estan tots als blocs 110 i 111, i els últims blocs són els que corresponen a Adversity.

Tractem aquests valors perquè tinguem 18 categories d'Acorn (Acorn-A fins Acorn-U) i 3 categories de Acorn_grouped (Adversity, Comfortable i Affluent).

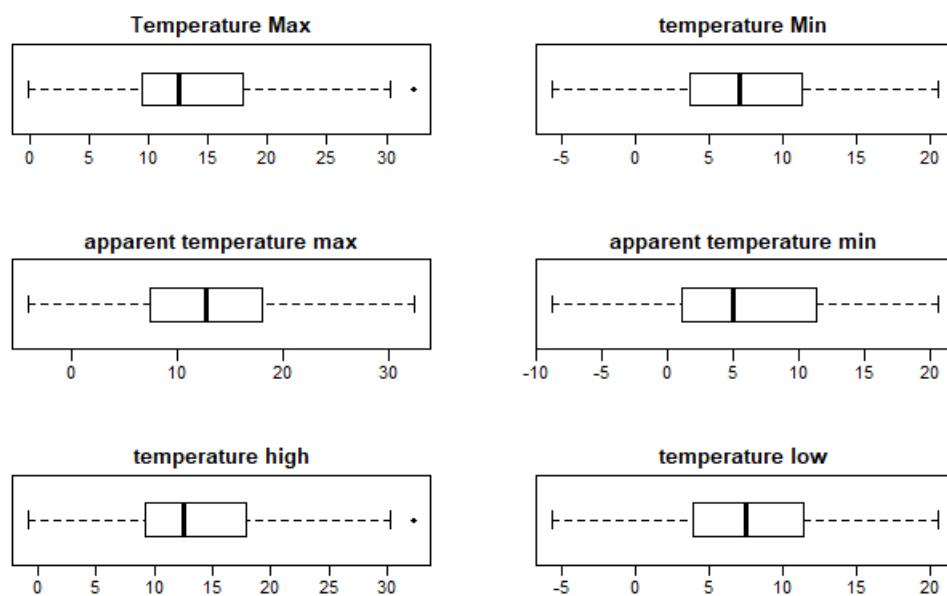
Procedim a buscar si hi ha valors buits i veiem que no conté valors buits.

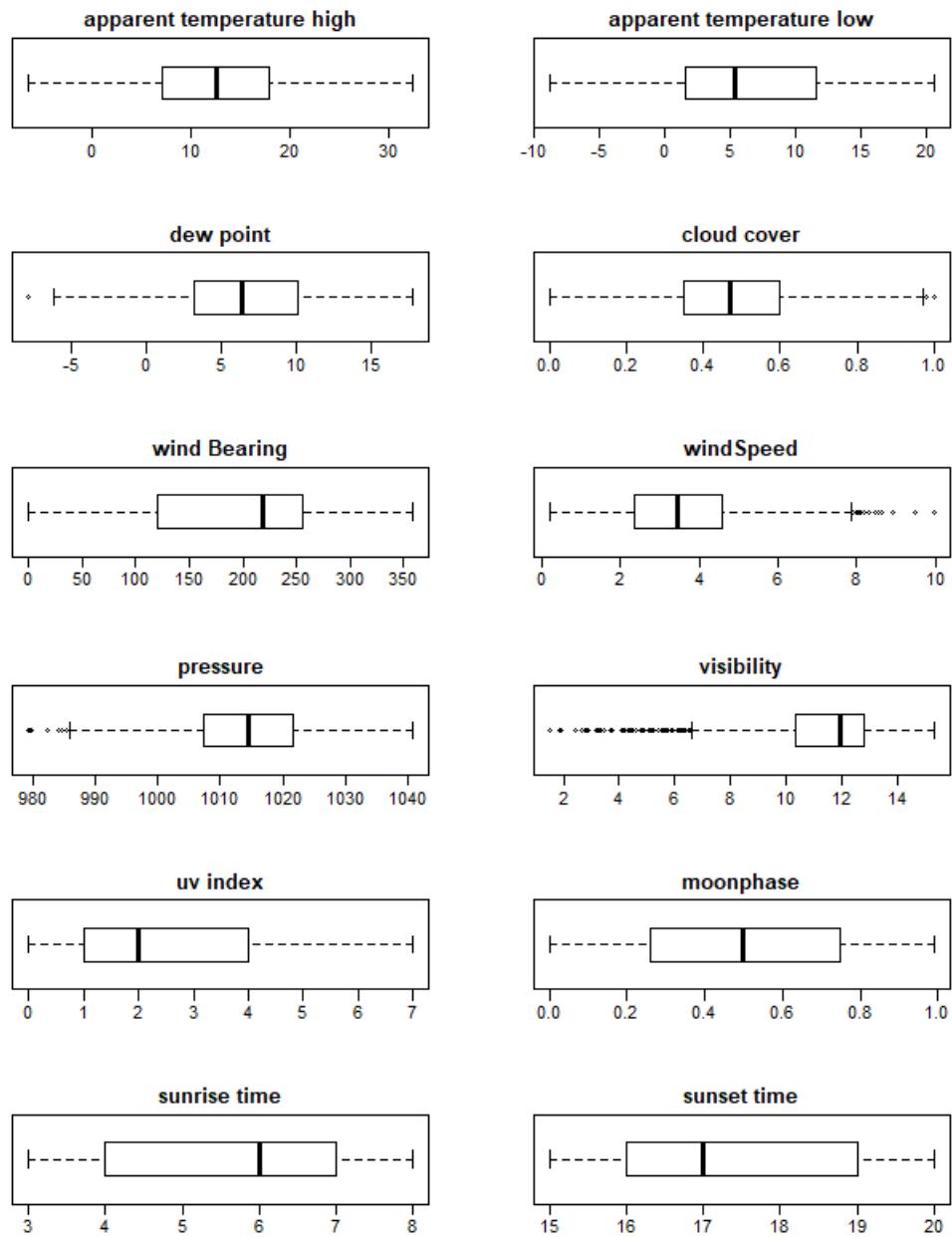
3.2.2 Weather_daily

En primer lloc, revisem que tots els tipus de dades siguin correctes i que no hi hagi dades errònies a partir del resum estadístic i dels gràfics de caixa .

Veiem que les dates les agafa com un factor, així que les convertirem a data (timestamp). A més a més, com que es repeteix la data a cada variable “Time”, decidim quedar-nos tant sols amb la hora, ja que tenim la variable data i creem les variables “Hour”.

Els valors de les variables quantitatives no estan fora de rang i comprovem que els valors de les variables qualitatives no estiguin repetits.





Gràfic 5 Diagrames de caixa variables numèriques weather_daily

Passem a comprovar si conté valors buits i veiem que conté valors buits a les variables cloudCover, uvIndex i uvIndexTime i que els valors es troben en el mateix dia (mateixa fila).

Utilitzem la regressió lineal per predir els valors de cloudCover i uvIndex. Per a seleccionar les variables utilitzades en la regressió, primer analitzem la correlació que hi ha entre elles. Obtenim que el valor de cloudCover que faltava era 0.64 i el valor de uvIndex 2.

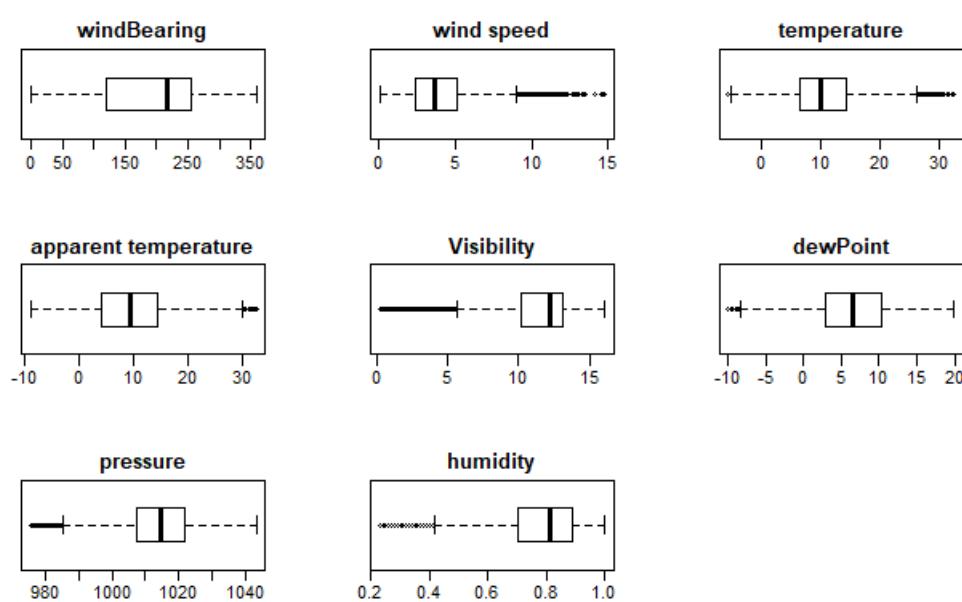
Per assignar el valor de uvIndexTime, primer mirem les hores que es produueixen els determinats uvIndex. Veiem que els uvIndex=2 s'han produït a les hores 9, 10, 11 i 12. La mitjana d'aquestes hores és 10.5, per tant assignem l'hora 11 a l'uvIndexTime faltant.

Finalment afegim algunes variables com l'any (year), el mes (month) i el dia de la setmana (weekday).

3.2.3 Weather hourly

En primer lloc, comprovem el tipus de dades i canviem a timestamp la variable time que s'agafava automàticament com a factor.

Llavors comprovem que els valors de les variables qualitatives no estiguin repetits ni mal escrits i després comprovem que les variables numèriques no estan fora del seu rang a partir de un boxplot.



Gràfic 6 Diagrames de caixa variables numèriques weather_hourly

Totes les variables són aparentment correctes.

Comprovem si alguna variable conté valors buits i veiem que la variable pressure conté valors buits.

Abans de tractar-la, creem les variables hora, data, any, mes i dia de la setmana.
Per a tractar els valors buits de pressió hem comprovat els valors buits en el seu context i hem vist que el valor de la pressió atmosfèrica es podia calcular aproximadament a partir del valor de pressió de l'hora anterior i l'hora següent al valor buit.

3.2.4 UK Bank holidays

En primer lloc unifiquem els noms dels valors que pot agafar la variable Type (tipus de festiu). Llavors factoritzem Type i convertim a data la variable Bank.holidays, ja que és la variable que conté la data. Comprovem que no hi hagi valors buits.

3.2.5 Tariffs

En primer lloc comprovem que el tipus de dades sigui correcte, i veiem que cal canviar la data a timestamp i la tarifa a factor. Ho convertim i comprovem si hi ha valors buits, concloent que no hi ha valors buits.

3.2.6 Consum diari (Daily_dataset_csv)

En aquest cas primer tractarem els valors buits i llavors comprovarem les variables.

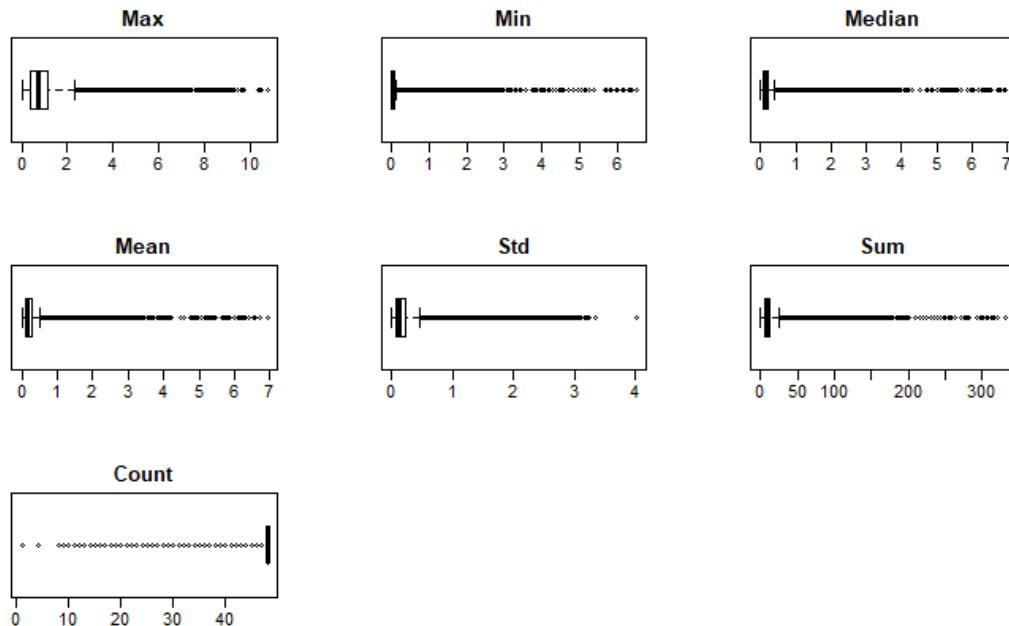
Veiem que tenim 30 valors buits a energy_median, energy_mean, energy_max, energy_sum i energy_mean. Aquests valors buits no són esporàdics sinó que els trobem tots en una mateixa fila (30 files), és a dir, hi ha 30 files que tenen els valors buits. Veiem que aquestes 30 files tenen energy_count a 0, que vol dir que no hi ha hagut cap mesura. També veiem que aquests valors s'han produït als extrems de la mesura, és a dir, si el primer registre comença un dia i acaba un altre dia, aquests registres són aquest primer i últim dia.

Comprovem al fitxer hhblock_dataset a partir de l'id i el dia, si trobem aquests registres, i no existeixen. De manera que els eliminem, ja que no ens aporten informació.

De valors buits també hi trobem 11331 corresponents a energy_std (desviació estàndard), que és una variable que en principi no utilitzarem per predir i que la podríem calcular si la necessitéssim, de manera que no la tractem ni la eliminem, ja que quan unim tots els datasets no la seleccionarem.

En quant al tipus de variables, convertim la variable day a data.

També creem diagrames de caixa per comprovar el rang dels valors numèrics i possibles outliers o valors atípics.



Gràfic 7 Diagrames caixa daily_dataset_csv

En el gràfic 7 veiem que es consideren molts outliers, tot i que no sembla haver-hi cap valor atípic. Sí que podem veure que per exemple a l'energia màxima consumida durant una hora (gràfic Max) hi ha algunes llars que superen els 10 KWh en una hora que estan més separats dels altres valors, i tenim un patró igual amb std, tot i que no els considerarem valors atípics ja que sempre poden haver-hi ocasions que es necessiti més energia en una hora per alguna cosa en concret (per exemple per un motor de piscina que s'engega sempre a la mateixa hora).

3.2.7 Hourly_dataset

En relació al primer grup de dades, hhblock_dataset, en principi només l'hem utilitzat per comprovar els valors buits de daily_dataset_csv, però el tractem i preparem per si el necessitéssim.

En primer lloc, tractem els valors buits que n'hi ha a les variables hh_19, hh_25, hh_26, hh_30 i hh_36. Per fer-ho farem la diferència entre la suma de les variables de hhblock_dataset de hh_0 a hh_47 (consum total diari sense el valor buit) i la variable energy_sum (consum total diari) del dataset daily_dataset. Un cop tenim tots els valors plens, convertim la variable day a data.

En relació al segon grup de dades, halfhourly_dataset, comprovem que no té valors buits. Llavors convertim la variable tstamp en tipus timestamp i afegim variables hora i dia.

3.3 Selecció de variables

Per a dur a terme l'estudi, combinem les dades dels fitxers informations_households, weather_daily, uk_bank_holidays i daily_dataset_csv.

Comencem combinant daily_dataset i uk_bank_holidays a partir del dia. Com que els dies que no són festius els marca com a valor buit, els convertim en un factor “No Holiday”. També afegim la variable “isHoliday” que indica si és un dia de vacances o no.

Llavors hi afegim el fitxer information_households a partir de l'id, abans havent seleccionat els atributs que volem d'information_households (LCLid, stdorToU, Acorn i Acorn_grouped).

Finalment hi afegim el conjunt weather_daily a partir de la data, on abans hem seleccionat les dades, obviant per exemple les variables timestamp.

També hi afegim la variable estació (season). Per tant, ens queda un dataset amb les següents variables:

"LCLid"	"date"
"energy_median"	"energy_mean"
"energy_max"	"energy_count"
"energy_std"	"energy_sum"
"energy_min"	"holidayType"
"isHoliday"	"stdorToU"
"Acorn"	"Acorn_grouped"
"temperatureMax"	"temperatureMaxHour"
"temperatureMin"	"temperatureMinHour"
"apparentTemperatureMax"	"apparentTemperatureMaxHour"
"apparentTemperatureMin"	"apparentTemperatureMinHour"
"temperatureHigh"	"temperatureHighHour"
"temperatureLow"	"temperatureLowHour"
"apparentTemperatureHigh"	"apparentTemperatureHighHour"
"apparentTemperatureLow"	"apparentTemperatureLowHour"
"windBearing"	"dewPoint"
"cloudCover"	"windspeed"
"pressure"	"humidity"
"visibility"	"moonPhase"
"icon"	"precipType"
"summary"	"uvIndex"
"uvIndexHour"	"sunsetTimeHour"
"sunriseTimeHour"	"year"
"month"	"weekday"
"season"	

imatge 1 Variables del dataset

4. Anàlisi de dades

Per a fer l'anàlisi utilitzarem gràfiques, correlacions per veure la relació entre variables, anàlisis estadístic per conèixer les variables (mitjana, mediana, quartils...), contrasts d'hipòtesis per respondre a algunes preguntes, regles d'associació per l'anàlisi de dependències i tècniques de visualització per buscar patrons de consum.

Tot el codi de realització i resultats executats que es mencionen en aquest apartat es troben al fitxer annex de R “Analisi dades daily”. La majoria de gràfics han estat creats amb Tableau i també es troben a l'annex.

4.1 Anàlisi estadístic

En l'arxiu daily tenim les dades diàries del consum i consta de 3517002 files i 48 columnes.

4.1.1 Recompte per variables qualitatives

En la imatge 2 veiem com en el fitxer daily hi tenim 3510403 files que contenen el consum de les diferents llars del 23-11-2011 al 28-02-2014.

```
LCLid      date
Length:3510403 Min. :2011-11-23
Class :character 1st Qu.:2012-10-21
Mode  :character Median :2013-03-30
                  Mean :2013-03-27
                  3rd Qu.:2013-09-10
                  Max. :2014-02-28
```

imatge 2 Estadistics fitxer daily total

Aquestes files consten de 5566 llars que tenen o bé tarifa Std (4443 llars) o bé tarifa ToU (1123 llars). Aquestes llars es classifiquen en diferents grups que anomenarem “Acorn”. Tenim de l’Acorn A a l’Acorn Q i l’Acorn U (18 grups). Veiem que els Acorn no engloben el mateix nombre de llars, sent els Acorn E, Q i F els més nombrosos i que conjuntament contenen més de la meitat de les llars.

Els Acorn es poden agrupar en una categoria més gran anomenada “Acorn grouped” on només hi ha 3 categories: Adversity, Comfortable i Affluent, sent els Affluent els que contenen quasi la meitat de les llars. Com ja hem dit anteriorment, sabem que els affluent tenen un millor nivell de vida que els adversity, i el grup comfortable seria entremig.

LCLid	stdorToU	Acorn	Acorn_grouped
Length:5566	Std:4443	ACORN-E:1567	ACORN- : 0
Class :character	ToU:1123	ACORN-Q: 831	ACORN-U : 0
Mode :character		ACORN-F: 684	Adversity :1865
		ACORN-H: 455	Affluent :2194
		ACORN-L: 342	Comfortable:1507
		ACORN-D: 292	
		(Other):1395	

imatge 3 Estadistics informations_households

En quant a les dades meteorològiques, veiem que de la mostra, la majoria de dies van estar parcialment ennuvolats. En quant a la precipitació, veiem que la majoria de dies ha estat pluja, sent només 20 dies els que ha nevat. En relació a summary, és una variable que desglossa la variable icon, de manera que les categories que més ens trobem és que el dia ha estat ennuvolat.

icon	precipType	summary
clear-day : 5	rain:809	Mostly cloudy throughout the day. :167
cloudy : 8	snow: 20	Partly cloudy throughout the day. :167
fog : 69		Partly cloudy until evening. :130
partly-cloudy-day :594		Mostly cloudy until evening. :111
partly-cloudy-night: 31		Foggy in the morning. : 36
wind :122		Partly cloudy starting in the afternoon, continuing until evening.: 16
		(Other) :202

imatge 4 Estadístics weather_daily

En relació als dies de festa, només suposen un 2% del total de dies. En quant a l'estació, veiem que tenim més dies de mostra de l'estació tardor.

holidayType	isHoliday	season
No Holiday :811	Mode :logical autumn:207	
New Years Day : 3	FALSE:811	spring:188
Boxing Day : 2	TRUE :18	summer:186
Christmas Day : 2		winter:248
Early May bank holiday: 2		
Easter Monday : 2		
(Other) : 7		

imatge 5 Estadistics holidays

4.1.2 Quantitatius: Mitjana, mediana, quartils per variables quantitatives

En relació a l'energia, veiem que la mitjana de l'energia mig-horaria diària (energy_mean) per llar és 0,212kWh i que el màxim és 6,928kWh (és a dir, es consumeix una mitjana de 0,212kWh cada mitja hora per llar i el màxim que es consumeix en una mitja hora és 6,928kWh)

Sabem que diàriament les llars gasten una mitjana de 10,12kWh al dia, arribant a un màxim de consum de 332,556kWh al dia.

Veiem que en alguns casos a l'energy_sum no tenim totes les hores d'un dia, ja sigui pel mal funcionament del comptador o ve perquè han començat a formar part de la mostra a meitat del dia, i ho sabem perquè energy_count no arriba a 48 en tots els casos.

energy_median	energy_mean	energy_max	energy_count	energy_std	energy_sum	energy_min
Min. :0.0000	Min. :0.00000	Min. : 0.0000	Min. : 1.0	Min. :0.000	Min. : 0.000	Min. :0.00000
1st Qu.:0.0670	1st Qu.:0.09808	1st Qu.: 0.3460	1st Qu.:48.0	1st Qu.:0.069	1st Qu.: 4.682	1st Qu.:0.02000
Median :0.1145	Median :0.16329	Median : 0.6880	Median :48.0	Median :0.133	Median : 7.815	Median :0.03900
Mean :0.1587	Mean :0.21173	Mean : 0.8345	Mean :47.8	Mean :0.173	Mean : 10.124	Mean :0.05963
3rd Qu.:0.1910	3rd Qu.:0.26246	3rd Qu.: 1.1280	3rd Qu.:48.0	3rd Qu.:0.229	3rd Qu.: 12.569	3rd Qu.:0.07100
Max. :6.9705	Max. :6.92825	Max. :10.7610	Max. :48.0	Max. :4.025	Max. :332.556	Max. :6.52400

Imatge 6 Estadístics daily

En relació a la meteorologia, veiem que ens els dies que s'ha pres la mostra s'ha arribat a una temperatura màxima de 32.4 graus i a una mínima de -5.64 graus, havent-hi dies que no s'han superat els 0 graus de màxima.

temperatureMax	temperatureMaxHour	temperatureMin	temperatureMinHour	apparentTemperatureMax
Min. :-0.060	Min. : 0.0	Min. :-5.640	Min. : 0.000	Min. :-4.110
1st Qu.: 9.502	1st Qu.:13.00	1st Qu.: 3.705	1st Qu.: 4.000	1st Qu.: 7.332
Median :12.625	Median :14.00	Median : 7.100	Median : 6.000	Median :12.625
Mean :13.660	Mean :13.79	Mean : 7.414	Mean : 9.141	Mean :12.929
3rd Qu.:17.920	3rd Qu.:15.00	3rd Qu.:11.277	3rd Qu.:18.000	3rd Qu.:17.920
Max. :32.400	Max. :23.00	Max. :20.540	Max. :23.000	Max. :32.420

Imatge 7 Estadístics weather daily

En quant a la temperatura apparent, veiem que tant la mínima temperatura com la màxima apparents són més baixes, i que les mitjanes de la temperatura mínima i màxima apparents són més baixes que les de temperatura.

apparentTemperatureMaxHour	apparentTemperatureMin	apparentTemperatureMinHour	temperatureHigh
Min. : 0.00	Min. :-8.880	Min. : 0.000	Min. :-0.810
1st Qu.:13.00	1st Qu.: 1.105	1st Qu.: 4.000	1st Qu.: 9.213
Median :14.00	Median : 4.885	Median : 5.500	Median :12.470
Mean :13.73	Mean : 5.738	Mean : 8.684	Mean :13.542
3rd Qu.:16.00	3rd Qu.:11.277	3rd Qu.:10.000	3rd Qu.:17.910
Max. :23.00	Max. :20.540	Max. :23.000	Max. :32.400

Imatge 8 Estadístics weather daily

Veiem que les temperatures High i Low estan entremig de les apparents màxima i mínima i de les temperatures màximes i mínimes.

temperatureHighHour	temperatureLow	temperatureLowHour	apparentTemperatureHigh	apparentTemperatureLowHour
Min. : 6.00	Min. :-5.64	Min. : 0.000	Min. :-6.460	Min. : 7.0
1st Qu.:13.00	1st Qu.: 3.99	1st Qu.: 4.000	1st Qu.: 7.032	1st Qu.:13.0
Median :14.00	Median : 7.54	Median : 5.000	Median :12.470	Median :14.0
Mean :14.18	Mean : 7.71	Mean : 6.531	Mean :12.724	Mean :14.2
3rd Qu.:15.00	3rd Qu.:11.47	3rd Qu.: 7.000	3rd Qu.:17.910	3rd Qu.:16.0
Max. :19.00	Max. :20.54	Max. :23.000	Max. :32.420	Max. :19.0

Imatge 9 Estadístics weather daily

Altres variables com cloud cover , humidity i moonphase són percentuals i van de 0 a 1, on la mitjana de tapament de núvol és 0.477 i la mitjana d'humitat és 0.78, tot i que aquesta variable en la mostra no arriba en cap moment als valors dels extrems 0 i 1. La variable sunsetTimeHour (hora de posta de sol) pren valors molt variables de les 15h. a les 20h., sent la mediana les 17h., i la variable sunriseTimeHour (hora de sortida del sol) també pren valors variables de les 3h. a les 8h., sent la mediana les 6h. La variable uvIndex només pot prendre del valor 0 a el 7, sent la mediana de 11.

Les altres variables que tenim (windBearing, dewPoint, windSpeed, pressure i visibility) prenen valors molt variats. Wind bearing va de 0 a 359 amb una mitjana de 185,7 , dewPoint (punt de rosada) va de -7.84 a 17.77 amb una mitjana de 6.53 , windSpeed va de 0,2 a 9,96 amb una mitjana de 3,58, pressure va de 979,2 fins 1040,9 amb una mitjana de 1014,1 i visibility va de 1.48 a 15.34 amb una mitjana de 11,17.

apparentTemperatureLow	apparentTemperatureLowHour	windBearing	dewPoint	cloudCover	windSpeed
Min. :-8.880	Min. : 0.000	Min. : 0.0	Min. :-7.84	Min. :0.0000	Min. :0.200
1st Qu.: 1.522	1st Qu.: 4.000	1st Qu.:120.5	1st Qu.: 3.18	1st Qu.:0.3500	1st Qu.:2.370
Median : 5.315	Median : 5.000	Median :219.0	Median : 6.38	Median :0.4700	Median :3.440
Mean : 6.085	Mean : 7.061	Mean :195.7	Mean : 6.53	Mean :0.4778	Mean :3.582
3rd Qu.:11.467	3rd Qu.: 7.000	3rd Qu.:255.0	3rd Qu.: 10.06	3rd Qu.:0.6000	3rd Qu.:4.577
Max. :20.540	Max. :23.000	Max. :359.0	Max. :17.77	Max. :1.0000	Max. :9.960

Imatge 10 Estadístics weather daily

pressure	humidity	visibility	moonPhase	uvIndex	uvIndexHour	sunsetTimeHour	sunriseTimeHour
Min. : 979.2	Min. : 0.4300	Min. : 1.48	Min. : 0.0000	Min. : 0.000	Min. : 0.00	Min. : 15.00	Min. : 3.000
1st Qu.: 1007.4	1st Qu.: 0.7200	1st Qu.: 10.33	1st Qu.: 0.2600	1st Qu.: 1.000	1st Qu.: 10.00	1st Qu.: 16.00	1st Qu.: 4.000
Median : 1014.6	Median : 0.7900	Median : 11.97	Median : 0.5000	Median : 2.000	Median : 11.00	Median : 17.00	Median : 6.000
Mean : 1014.1	Mean : 0.7819	Mean : 11.17	Mean : 0.5009	Mean : 2.542	Mean : 10.61	Mean : 17.44	Mean : 5.641
3rd Qu.: 1021.8	3rd Qu.: 0.8600	3rd Qu.: 12.83	3rd Qu.: 0.7500	3rd Qu.: 4.000	3rd Qu.: 12.00	3rd Qu.: 19.00	3rd Qu.: 7.000
Max. : 1040.9	Max. : 0.9800	Max. : 15.34	Max. : 0.9900	Max. : 7.000	Max. : 13.00	Max. : 20.00	Max. : 8.000

imatge 11 Estadístics weather daily

4.1.3 Correlacions

En primer lloc realitzem un heatmap amb els resultats de la correlació de les dades meteorològiques numèriques amb les d'energia diàries.

En el gràfic 8 veiem que amb el que estan més relacionades les variables d'energia és amb les variables de temperatura, on a més temperatura menys consum, ja que la relació és inversa (-0,17). També veiem que les següents variables que tenen més relació amb l'energia són dewPoint, uvindex, sunsetTimeHour i sunriseTimeHour on la relació és inversa menys amb sunriseTimeHour.

Veiem que la variable dewPoint té una relació directa amb la temperatura de fins el 0,94, de manera que creiem que no hauríem d'utilitzar aquesta variable per no crear esbiaixos. Les altres variables tenen una relació amb la temperatura de 0,7 uvindex, 0,68 sunsetTimeHour i -0,68 sunriseTimeHour. Aquestes variables també tenen una forta relació entre elles, així que utilitzarem tant sols una variable, sunsetTimeHour. Altres variables que tenen una mica de relació que també considerem interessants són humidity, cloudcover i visibility.

També mirem la correlació entre les variables qualitatives i les variables d'energia, on veiem que les variables d'energia semblen no tenir relació amb holidayType (0 o -0.01) ni amb isHoliday(0), ni amb season(0.02), ni amb icon (0.01). Tenen una mica de relació però molt poca amb preciptype (0.06) i amb summary (-0.05). Suposem que això és degut a que aquestes variables tenen poques categories de classificació i per tant no sembla haver-hi relació.

També mirem la correlació amb les variables de consumidor com són l'Acorn, l'Acorn agrupat i la tarifa, on veiem que stdorToU té una relació amb molt poca força tant amb energy_max, energy_mean i energy_sum (-0,03) com amb energy_min (-0,02). La variable Acorn_grouped té una relació igualment amb poca força, concretament 0,07 amb energy_sum i energy_mean i 0,05 amb energy_min i energy_max. La relació Acorn_grouped té una relació més forta, de -0.16 amb energy_sum i energy_mean, de -0.11 amb energy_max i de -0.13 amb energy_min.

	energy	energy	energy	temperat ure Max	temperat ure Min	Temperat ure Max	Temperat ure Min	temperat ure High	temperat ure Low	Temperat ure High	Temperat ure Low	dew Point	uv Index	sunset Time Hour	sunset visibility	month	pressure	sunrise Time Hour	sunrise humidity	cloud Cover	year	wind Speed	wind weekday	wind Bearing	moon Phase		
	mean	max	sum	ure Max	ure Min	Temperat ure Max	Temperat ure Min	ure High	ure Low	ure High	ure Low	-0,15	-0,15	-0,15	-0,15	-0,06	-0,04	0,15	0,08	0,06	0,08	0,03	-0,01	-0,02	-0,01		
energy_mean	1,00	0,76	0,99	-0,17	-0,16	-0,17	-0,17	-0,17	-0,16	-0,17	-0,16	-0,15	-0,15	-0,15	-0,15	-0,06	-0,04	0,15	0,08	0,06	0,08	0,03	-0,01	-0,02	-0,01		
energy_max	0,76	1,00	0,77	-0,15	-0,15	-0,15	-0,15	-0,14	-0,15	-0,14	-0,14	-0,13	-0,13	-0,13	-0,13	-0,05	-0,04	-0,03	0,13	0,07	0,05	0,02	0,03	-0,02	-0,01	-0,01	
energy_sum	0,99	0,77	1,00	-0,17	-0,16	-0,17	-0,17	-0,17	-0,16	-0,17	-0,16	-0,15	-0,14	-0,15	-0,15	-0,06	-0,04	-0,04	0,15	0,08	0,06	0,08	0,04	-0,02	-0,01	-0,01	
temperatureMax	-0,17	-0,15	-0,17	1,00	0,90	0,99	0,92	1,00	0,89	0,99	0,90	0,87	0,70	0,67	0,30	0,23	0,22	-0,68	-0,43	-0,37	-0,20	-0,18	-0,01	0,08	0,02		
temperatureMin	-0,16	-0,15	-0,16	0,90	1,00	0,90	0,98	0,90	0,89	0,90	0,90	0,94	0,60	0,61	0,20	0,25	0,11	-0,62	-0,23	-0,12	-0,21	-0,06	-0,01	0,08	0,00		
apparentTemperatureMax	-0,17	-0,15	-0,17	0,99	0,90	1,00	0,92	0,99	0,89	1,00	0,91	0,88	0,68	0,65	0,27	0,25	0,22	-0,67	-0,38	-0,35	-0,22	-0,20	-0,01	0,09	0,02		
apparentTemperatureMin	-0,17	-0,15	-0,17	0,92	0,98	0,92	1,00	0,92	0,90	0,92	0,92	0,94	0,62	0,62	0,17	0,27	0,16	-0,64	-0,22	-0,15	-0,24	-0,16	0,00	0,08	0,01		
temperatureHigh	-0,17	-0,15	-0,17	1,00	0,90	0,99	0,92	1,00	0,89	0,99	0,90	0,87	0,70	0,68	0,30	0,23	0,23	-0,69	-0,43	-0,37	-0,21	-0,19	-0,01	0,07	0,03		
temperatureLow	-0,16	-0,14	-0,16	0,89	0,89	0,89	0,90	0,89	1,00	0,89	0,98	0,90	0,59	0,58	0,18	0,25	0,16	-0,60	-0,21	-0,16	-0,20	-0,11	-0,01	0,06	0,01		
apparentTemperatureHigh	-0,17	-0,15	-0,17	0,99	0,90	1,00	0,92	0,99	0,89	1,00	0,91	0,88	0,69	0,66	0,27	0,25	0,23	-0,68	-0,39	-0,35	-0,22	-0,21	0,00	0,09	0,02		
apparentTemperatureLow	-0,16	-0,14	-0,16	0,90	0,90	0,91	0,92	0,90	0,98	0,91	1,00	0,90	0,61	0,60	0,18	0,27	0,19	-0,62	-0,21	-0,19	-0,23	-0,17	0,00	0,06	0,02		
dewPoint	-0,15	-0,14	-0,15	0,87	0,94	0,88	0,94	0,87	0,90	0,88	0,90	1,00	0,48	0,48	0,05	0,34	0,07	-0,51	0,03	-0,08	-0,24	-0,14	-0,01	0,11	0,01		
uvIndex	-0,15	-0,13	-0,14	0,70	0,60	0,68	0,62	0,70	0,59	0,69	0,61	0,48	1,00	0,92	0,30	-0,12	0,22	-0,92	-0,58	-0,27	-0,22	-0,14	0,00	-0,09	-0,08		
sunsetTimeHour	-0,15	-0,13	-0,15	0,67	0,61	0,65	0,62	0,68	0,58	0,66	0,60	0,48	0,92	1,00	0,26	-0,23	0,19	-0,94	-0,55	-0,14	-0,12	-0,12	0,00	0,04	-0,02		
visibility	-0,06	-0,05	-0,06	0,30	0,20	0,27	0,17	0,30	0,18	0,27	0,18	0,05	0,30	0,26	1,00	0,07	0,06	-0,25	-0,61	-0,38	-0,04	0,23	0,02	0,00	0,07		
month	-0,04	-0,04	-0,04	0,23	0,25	0,25	0,27	0,23	0,25	0,25	0,27	0,34	-0,12	-0,23	0,07	1,00	0,09	0,06	0,17	-0,13	-0,48	-0,12	-0,01	0,02	-0,03		
pressure	-0,04	-0,03	-0,04	0,22	0,11	0,22	0,16	0,23	0,16	0,23	0,19	0,07	0,22	0,19	0,06	0,09	1,00	-0,21	-0,29	-0,12	-0,14	-0,35	0,01	0,25	0,08		
sunriseTimeHour	0,15	0,13	0,15	-0,68	-0,62	-0,67	-0,64	-0,69	-0,60	-0,68	-0,62	-0,51	-0,92	-0,94	-0,25	0,06	-0,21	1,00	0,51	0,16	0,22	0,14	-0,01	-0,06	0,01		
humidity	0,08	0,07	0,08	-0,43	-0,23	-0,38	-0,22	-0,43	-0,21	-0,39	-0,21	0,03	-0,58	-0,55	-0,61	0,17	-0,29	0,51	1,00	0,47	-0,05	-0,05	-0,01	-0,10	-0,02		
cloudCover	0,06	0,05	0,06	-0,37	-0,12	-0,35	-0,15	-0,37	-0,16	-0,35	-0,19	-0,08	-0,27	-0,14	-0,38	-0,13	-0,12	0,16	0,47	1,00	0,06	0,17	-0,03	0,08	-0,02		
year	0,03	0,02	0,03	-0,20	-0,21	-0,22	-0,24	-0,21	-0,20	-0,22	-0,23	-0,24	-0,22	-0,12	-0,04	-0,48	-0,14	0,22	-0,05	0,06	1,00	0,16	0,00	-0,06	0,00		
windSpeed	0,03	0,03	0,04	-0,18	-0,06	-0,20	-0,16	-0,19	-0,11	-0,21	-0,17	-0,14	-0,14	-0,12	0,23	-0,12	-0,35	0,14	-0,05	0,17	0,16	1,00	0,03	0,20	0,03		
weekday	-0,01	-0,02	-0,02	-0,01	-0,01	-0,01	0,00	-0,01	0,00	0,00	-0,01	0,00	0,00	0,00	0,02	-0,01	0,01	-0,01	-0,01	-0,03	0,00	0,03	1,00	0,03	0,00		
windBearing	-0,02	-0,01	-0,01	0,08	0,08	0,09	0,08	0,07	0,06	0,09	0,06	0,11	-0,06	-0,10	0,25	0,20	0,00	0,08	0,02	-0,09	-0,05	0,04	0,03	1,00	0,04		
moonPhase	-0,01	-0,01	-0,01	0,02	0,00	0,02	0,01	0,03	0,01	0,02	0,02	0,01	0,01	0,01	-0,02	0,08	0,03	0,07	-0,02	-0,03	-0,08	0,00	-0,02	0,00	0,04	1,00	

Gràfic 8 Correlació entre variables weather daily

4.2 Contrast d'hipòtesis

Per a fer els contrasts d'hipòtesis, utilitzarem un contrast de dues mostres per la diferència de mitjanes quan la variable objectiu tingui només dos categories, en canvi utilitzarem l'anàlisi de variàncies quan tinguem més de dues categories o ve més d'una variable objectiu.

Per a triar el model per la diferència de mitjanes, comprovarem la normalitat de les dades dels dos grups i la igualtat de variàncies, i utilitzarem el test T de Student si les dades son normals i les variàncies iguals o diferents (Welch test), i si les dades no segueixen la distribució normal, utilitzarem el test U de Mann-Whitney, tot i que com que tenim una mostra molt gran, creiem que les dades no ens apareixeran com a normals però assumim el teorema del límit central, assumint que les dades tendeixen a seguir la distribució normal.

Per a triar el model d'anàlisi de variàncies quan tenim més de dues categories, comprovarem l'equilibri dels grups i la homoscedasticitat de les dades. Si no hi ha homoscedasticitat però els grups són equilibrats (hi ha un nombre semblant o igual de mostres per grup), utilitzarem ANOVA, en canvi si no hi ha homoscedasticitat i a més a més els grups no són equilibrats, llavors utilitzarem el test no paramètric de Kruskal Wallis. Per a saber quins grups son significativament diferents, utilitzarem anàlisis post-hoc per quantificar-ho, utilitzant Tukey per quan tinguem analisi ANOVA i Dunn amb ajustament de Bonferroni per Kruskal Wallis, ja que quan utilitzem Kruskal Wallis és perquè els grups no son equilibrats.

Per a triar el model d'anàlisi de variàncies quan tenim més de dues variables objectiu utilitzarem ANOVA si es compleixen les condicions d'aplicació i si no es compleixen utilitzarem el test no paramètric Friedman.

4.2.1 Consum segons tarifa és diferent?

Ens preguntem si els consumidors que tenen tarifa ToU consumeixen menys que els consumidors Std, i realitzem el contrast d'hipòtesis, on tenim:

Hipòtesi nul·la: $\text{mitjana_energy_sum_Std} = \text{mitjana_energy_sum_ToU}$

Hipòtesi alternativa: $\text{mitjana_energy_sum_Std} > \text{mitjana_energy_sum_ToU}$

Per a triar el test a utilitzar, assumim la normalitat de les dades pel teorema del límit central i comprovem la igualtat de variàncies. Utilitzarem un mètode per fer un contrast de dues mostres per la diferència de mitjanes amb distribució normal, variàncies desconegudes i diferents amb una hipòtesi alternativa unilateral utilitzant el test T de Student (Welch t-test) per a variàncies diferents.

```

Welch Two sample t-test

data: grup_std and grup_Tou
t = 70.889, df = 1234800, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.01606837      Inf
sample estimates:
mean of x mean of y
0.2150998 0.1986497

```

Imatge 12 T-test consum segons tarifa

Al obtenir un p-valor més petit que 0.05, rebutgem la hipòtesi nul·la que les mitjanes dels dos grups són iguals i acceptem la hipòtesi alternativa que la mitjana de la suma d'energia dels que tenen tarifa Std és més gran que la dels que tenen tarifa ToU.

Com que en aquest cas la normalitat no es compleix però la mostra si que hi tendeix, també utilitzem un altre test per confirmar els resultats (test U de Mann-Whitney).

També ens preguntem si els consumidors amb tarifa Std tenen consums màxims horaris més alts que els de tarifa ToU, on tenim:

H0: mitjana_energy_max_Std = mitjana_energy_max_ToU
H1: mitjana_energy_max_Std > mitjana_energy_max_ToU

I apliquem un contrast de dues mostres per la diferència de mitjanes amb una distribució normal amb variàncies desconegudes i diferents (T Student), que ens confirma que el consum màxim horari d'un dia dels consumidors amb tarifa Std és més alt que els dels consumidors amb tarifa ToU.

4.2.2 Consum segons tipus de consumidor és diferent?

Ens preguntem si és diferent el consum segons el tipus de consumidor, és a dir, si és diferent el consum d'Affluent que el de Comfortable, per exemple.

Com que les variables Acorn grouped tres nivells de categoria (Affluent, Comfortable i Adversity) , hem d'aplicar l'anàlisi de variància.

H0:mitjana_consum_Acorn_grouped_affluent=mitjana_consum_Acorn_grouped_comfortable=mitjana_consum_Acorn_grouped_adversity
H1: almenys dos grups són diferents

Veiem que no hi ha homoscedasticitat ja que en el test de Levene el p valor és menor de 0.05 de manera que es rebutja la hipòtesi nul·la que totes les variàncies són homogènies.

Com veiem a la imatge 13, els grups no són equilibrats i no podem aplicar ANOVA i realitzem el test no paramètric de Kruskal Wallis, on obtenim un p-valor inferior a 0.05 i rebutgem la hipòtesi nul·la, acceptant que almenys un dels grups és diferent. Per a saber quins són significativament diferents, apliquem el test de Dunn per a múltiples comparacions on veiem que tots els grups són significativament diferents entre ells.

Adversity	Affluent	Comfortable
1166999	1411215	938788

imatge 13 Dades agrupades per Acorn_grouped

Ens preguntem si és diferent el consum segons el tipus de consumidor, és a dir, si és diferent el consum d'Acorn A que el d'Acorn U, per exemple.

H0:

mitjana_energy_sum_AcornA=mitjana_energy_sum_Acorn_B=...=mitjana_energy_su
m_AcornU

H1: alguna de les mitjanes és diferent

Com que les variables Acorn tenen més de dos nivells de categoria, hem d'aplicar l'anàlisi de variància, i al no haver-hi homoscedasticitat ni equilibri entre els grups,

ACORN-A	ACORN-B	ACORN-C	ACORN-D	ACORN-E	ACORN-F	ACORN-G	ACORN-H
94046	14493	88840	189338	1024498	448035	123893	272360
ACORN-I	ACORN-J	ACORN-K	ACORN-L	ACORN-M	ACORN-N	ACORN-O	ACORN-P
29530	64970	106803	204338	66365	92156	61351	72306
ACORN-Q	ACORN-U						
534073	296073						

imatge 14 Dades agrupades per Acorn

realitzem el test de Kruskal Wallis amb la hipòtesi nul·la de que tots els Acorns són iguals i amb la hipòtesi alternativa que almenys un dels grups és diferent, obtenint un p-valor inferior a 0.05 i rebutjant la hipòtesi nul·la. Per a saber quins grups són significativament diferents, apliquem Dunn i resulta que tots són significativament diferents menys les parelles AcornH-AcornB, AcornF-AcornE, AcornJ-AcornI, AcornM-AcornI, AcornN-AcornI, AcornM-AcornJ, AcornN-AcornJ, AcornM AcornN, per tant veiem que hi ha un grup que és igual format pels Acorns I-M-N-J, un altre pels Acorns F-E i tots els altres individuals.

També ens preguntem si són diferents els pics de consum dels diferents consumidors, on :

Ho: energy_max_AcornA=energy_max_AcornB=...=energy_max_AcornU

H1: almenys un dels grups és diferent

Al no haver-hi homoscedasticitat ni grups equilibrats, apliquem Kruskal-Wallis i obtenim un p-valor inferior a 0.05, de manera que rebutgem la hipòtesi nul·la i acceptem la hipòtesi alternativa de que almenys un dels grups és diferent.

4.2.3 Consum segons tipus de tarifa i tipus de consumidor és diferent?

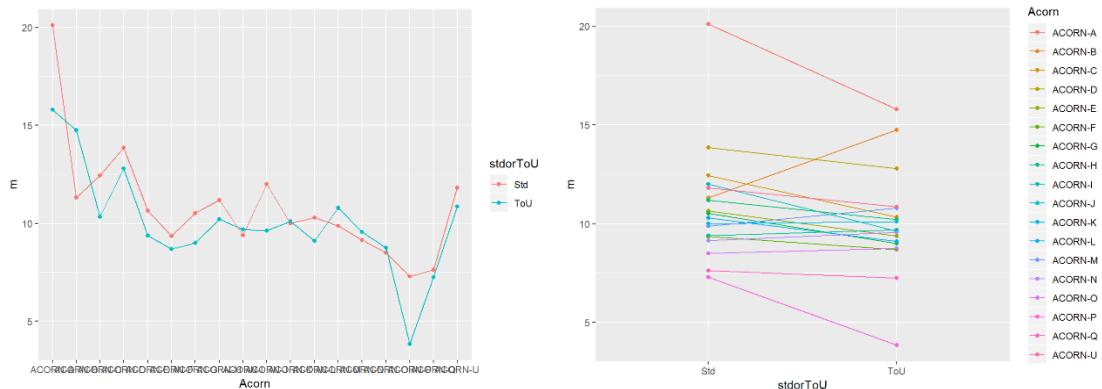
També ens preguntem si el tipus de tarifa afecta per igual al consum de tots els tipus de consumidors (Acorn), on:

H0:

mitjana_energy_sum_AcornA=mitjana_energy_sum_Acorn_B=...=mitjana_energy_su
m_AcornU

H1: alguna de les mitjanes és diferent

Abans de fer el contrast, analitzem i veiem que hi ha interaccions entre tarifa i tipus d'Acorn com veiem a la imatge 15.



Imatge 15 Consum segons tarifa per Acorn i consum segons Acorn per tarifa.

Al tenir més d'una variable objectiu (Acorn i tarifa), utilitzarem anàlisi multifactorial de variància. Com que els grups no són equilibrats i no hi ha homoscedasticitat, no podem utilitzar ANOVA ja que no es compleixen les condicions d'aplicació de l'ANOVA i utilitzem Friedman (test no paramètric), que rebutja la hipòtesi nul·la.

Aquest test també el podríem fer amb Kruskal-Wallis, creant una nova variable que conté totes les categories de les dues variables (per exemple, Acorn A Std, Acorn B ToU...) i realitzem un anàlisi de variància amb Kruskal Wallis, on obtenim que almenys un dels grups és diferent. També apliquem Dunn per veure quins són iguals o diferents.

Per exemple veiem que no són significativament diferents:

Els Acorns D, I, N i O no són significativament diferents en les diferents tarifes (ACORN-D ToU-ACORN-D Std, ACORN-I ToU-ACORN-I Std, ACORN-N ToU-ACORN-N Std, ACORN-O ToU-ACORN-O Std). També veiem que l'Acorn N i l'Acorn I no són significativament diferents entre ells (ACORN-N Std-ACORN-I Std, ACORN-N ToU-ACORN-I Std, ACORN-N Std-ACORN-I ToU, ACORN-N ToU-ACORN-I ToU).

Els altres grups que no són significativament són:

ACORN-B ToU-ACORN-A ToU
ACORN-G Std-ACORN-C ToU
ACORN-K Std-ACORN-C ToU
ACORN-U ToU-ACORN-C ToU
ACORN-G ToU-ACORN-E Std
ACORN-G ToU-ACORN-F Std
ACORN-J ToU-ACORN-F ToU
ACORN-I ToU-ACORN-G Std
ACORN-J Std-ACORN-G Std
ACORN-U ToU-ACORN-G Std
ACORN-M Std-ACORN-G ToU
ACORN-M ToU-ACORN-H Std
ACORN-L Std-ACORN-H ToU

ACORN-J Std-ACORN-I ToU
ACORN-K ToU-ACORN-I ToU
ACORN-U ToU-ACORN-I ToU
ACORN-K ToU-ACORN-J Std
ACORN-U ToU-ACORN-J Std
ACORN-O Std-ACORN-E ToU
ACORN-O ToU-ACORN-E ToU
ACORN-U ToU-ACORN-K Std
ACORN-U ToU-ACORN-K ToU
ACORN-N Std-ACORN-L ToU
ACORN-N ToU-ACORN-L ToU

4.2.4 Consum segons dia és diferent?

En primer lloc ens preguntrem si és més gran el consum el cap de setmana que entre setmana, de manera que fem:

$$H_0: \text{mitjana_energy_sum_finde} = \text{mitjana_energy_sum_laboral}$$
$$H_1: \text{mitjana_energy_sum_finde} > \text{mitjana_energy_sum_laboral}$$

I apliquem un contrast de dues mostres per la diferència de mitjanes amb una distribució normal amb variàncies desconegudes i diferents (T Student), que ens confirma que el consum en cap de setmana és més alt que en laboral.

Sabent que el consum en cap de setmana és més gran que en laboral, ens demanem si és més gran el consum en festiu que en laboral i si el consum de cap de setmana és més gran que al consum de holidays, i fem dos contrasts d'hipòtesis on:

$$H_0: \text{mitjana_energy_sum_capdesetmana} = \text{mitjana_energy_sum_isholiday}$$
$$H_1: \text{mitjana_energy_sum_capdesetmana} > \text{mitjana_energy_sum_isholiday}$$

Apliquem un contrast de dues mostres per la diferència de mitjanes amb una distribució normal amb variàncies desconegudes i diferents (T Student), que ens confirma que el consum en cap de setmana és més alt que en festiu, però amb un p-valor de 0.03 per una confiança del 95%.

$$H_0: \text{mitjana_energy_sum_isholiday} = \text{mitjana_energy_sum_laboral}$$
$$H_1: \text{mitjana_energy_sum_isholiday} > \text{mitjana_energy_sum_laboral}$$

Apliquem un contrast de dues mostres per la diferència de mitjanes amb una distribució normal amb variàncies desconegudes i diferents (T Student), que ens confirma que el consum en festiu és més alt que en laboral.

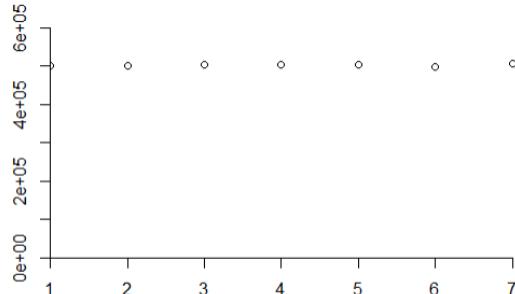
Per tant sabem que el consum en cap de setmana és més gran que en festiu i que el consum en festiu és més gran que el laboral.

Per detallar més, ens preguntrem si cada dia de la setmana es consumeix el mateix, és a dir, hi ha el mateix consum el dilluns i el dimarts per exemple? Realitzarem un anàlisi de variàncies per contrastar les hipòtesis:

H0: consum_dilluns=consum_dimarts=...=consum_diumenge

H1: algun dels grups és diferent

on no hi ha homoscedasticitat però si grups equilibrats, per tant apliquem ANOVA on obtenim un p-valor inferior a 0.05, de manera que rebutgem la hipòtesi nul·la i acceptem que almenys un dels grups és diferent dels altres.



Gràfic 9 Recompte de mostres per dia

Quantifiquem la diferència realitzant una comparació múltiple amb el test de Tukey:

```
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = energy_sum ~ weekday, data = dail)

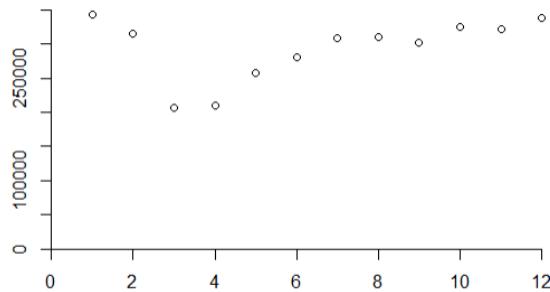
$`weekday`
    diff      lwr      upr   p adj
2-1 -0.13995706 -0.19373353 -0.08618059 0.0000000
3-1 -0.15186892 -0.20562381 -0.09811403 0.0000000
4-1 -0.17193486 -0.22566790 -0.11820181 0.0000000
5-1 -0.23987011 -0.29357699 -0.18616322 0.0000000
6-1  0.10631619  0.05247970  0.16015267 0.0000001
7-1  0.58564703  0.53199430  0.63929975 0.0000000
3-2 -0.01191186 -0.06564154  0.04181782 0.9949235
4-2 -0.03197780 -0.08568562  0.02173002 0.5784157
5-2 -0.09991305 -0.15359470 -0.04623140 0.0000009
6-2  0.24627325  0.19246193  0.30008456 0.0000000
7-2  0.72560409  0.67197663  0.77923155 0.0000000
4-3 -0.02006594 -0.07375215  0.03362027 0.9277452
5-3 -0.08800119 -0.14166122 -0.03434116 0.0000274
6-3  0.25818511  0.20439536  0.31197485 0.0000000
7-3  0.73751595  0.68391013  0.79112177 0.0000000
5-4 -0.06793525 -0.12157340 -0.01429710 0.0035599
6-4  0.27825104  0.22448313  0.33201896 0.0000000
7-4  0.75758189  0.70399797  0.81116580 0.0000000
6-5  0.34618629  0.29244452  0.39992806 0.0000000
7-5  0.82551713  0.77195945  0.87907482 0.0000000
7-6  0.47933084  0.42564320  0.53301848 0.0000000
```

imatge 16 test de Tukey consum segons dia de la setmana

On veiem que no hi ha diferència de consum entre els dies 2-3-4 (dimarts, dimecres i dijous tenen un consum igual) i si que hi ha diferencia per 1,5,6 i 7 (dilluns, divendres, dissabte i diumenge tenen un consum diferent entre ells i també diferent que el grup 2-3-4).

4.2.5 Consum segons mes és diferent?

Ens preguntem també si el consum de cada mes és igual o diferent, i al tenir grups no equilibrats i sense homoscedasticitat, apliquem Kruskal Wallis.



Gràfic 10 recompte de mostres per mes

En el test obtenim un pvalor inferior a 0.05, de manera que acceptem que els consums entre els mesos són diferents. Per quantificar aquesta diferència, utilitzem el test Dunn i apreciem que tots els mesos són significativament diferents entre ells menys 10 i 4 (octubre i abril)

4.2.6 Consum segons estació és diferent?

Ens preguntem si el consum és diferent entre estacions, amb les següents hipòtesis:

H0:

consum_estacio_hivern=consum_estacio_primavera=consum_estacio_estiu=consum_estacio_tardor

H1: almenys un grup és diferent

On al tenir grups no equilibrat i sense presència d'homoscedasticitat, apliquem el test de Kruskal Wallis obtenint una p inferior a 0.05, i per tant acceptant que almenys una de les estacions té un consum diferent. Utilitzem el test de Dunn per a saber quins són significativament diferents i veiem que totes les estacions són significativament diferents entre elles.

Ens preguntem també si es consumeix més a l'hivern que a les altres estacions, fent un contrast d'hipòtesis de dues mostres amb variàncies desconegudes i diferents amb Welch t-test, on amb la següent hipòtesis:

H0: mitjana_energy_sum_hivern=mitjana_energy_sum_altresestacions

H1: mitjana_energy_sum_hivern>mitjana_energy_sum_altresestacions

Obtenim un pvalor inferior a 0.05, de manera que afirmem que el consum a l'hivern és més gran que les altres estacions.

També fem contrastos de dues mostres per cada parella d'estacions, obtenint que el consum a l'hivern és més gran que a l'estiu, el consum a l'hivern és més gran que la tardor, el consum de la tardor és més gran que a la primavera i el consum de la primavera és més gran que a l'estiu. Per tant, quan es consumeix més és a l'hivern, seguit de la tardor, seguit de la primavera i finalment a l'estiu.

4.2.7 Consum segons temperatura és diferent?

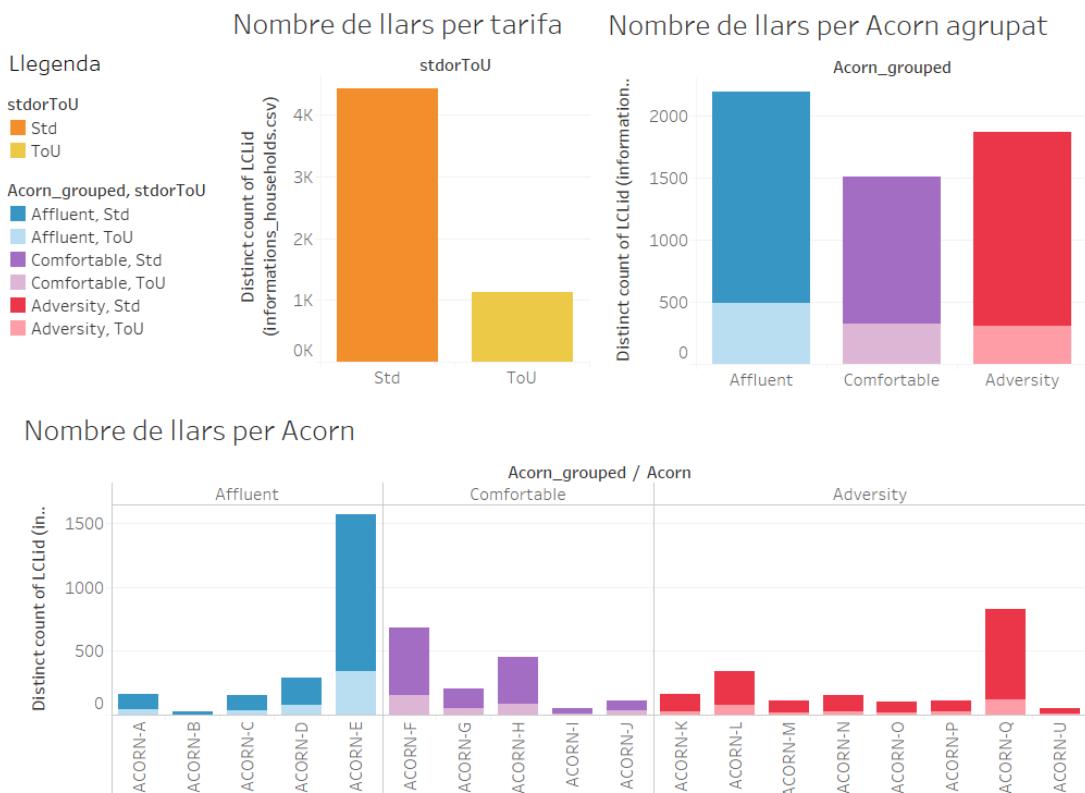
Ens preguntem si el consum varia amb el canvi de temperatura i fem un anàlisi de variància amb Kruskal Wallis ja que no es compleix la homoscedasticitat ni l'equilibri entre grups. Per a fer-ho abans hem transformat la variable temperatura a factor amb les categories "Molt Alta", "Alta", "Normal", "Baixa" i "Molt Baixa", i hem obtingut una p inferior a 0.05, de manera que acceptem que almenys un dels grups té un consum diferent. Quantifiquem amb el test de Dunn i veiem que totes les categories de temperatura son significativament diferents menys les categories alta-baixa, que són iguals.

4.3 Anàlisi gràfic i patrons de consum visuals

En aquest apartat representem les variables que tenim per veure'n informació bàsica de les dades i veurem el consum des de diferents perspectives. Els gràfics d'aquest apartat s'han realitzat amb Tableau.

4.3.1 Anàlisi dades diàries

En primer lloc, veiem representades les llars en funció de la tarifa que tenen i en funció de l'Acorn agrupat i de l'Acorn.



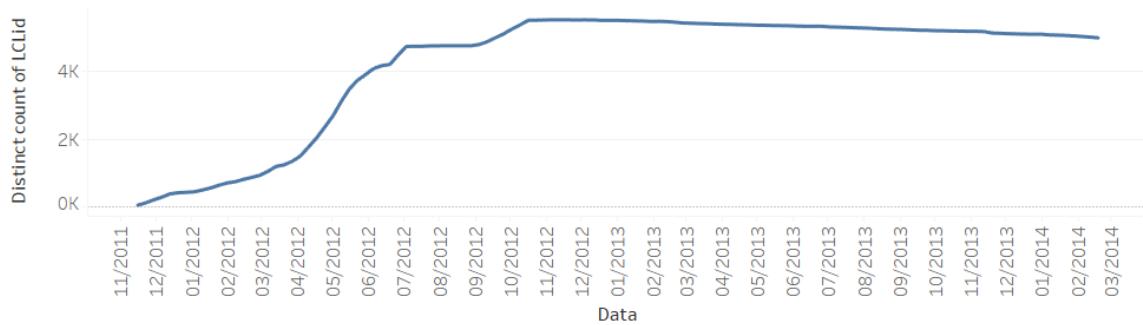
Gràfic 11 A l'esquerra, gràfic de recompte de nombre de llars segons tarifa, a la dreta, gràfic de recompte de nombre de llars segons Acorn_grouped i a sota, gràfic de recompte del nombre de llars segons Acorn.

Com ja sabíem, en la mostra hi ha més llars amb la tarifa Std que ToU. En relació a l'Acorn agrupat, les llars Affluent són les que més hi ha, seguides de Adversity i finalment Comfortable. També la proporció de Std i ToU és diferent en cada Acorn agrupat.

En relació a l'Acorn, veiem que almenys una cinquena part de la mosta és del grup Acorn-E. També veiem que la proporció de Std i ToU és diferent en cada Acorn. També mirant l'Acorn, veiem que la proporció de llars dins de cada Acorn agrupat és molt desigual, per exemple, dins de l'Acorn agrupat Affluent, veiem que la majoria de les llars són de l'Acorn E, de manera que quan mirem les dades de manera agrupada segons l'Acorn agrupat, com que Acorn E té més pes, el grup tendirà a reflectir les dades de l'Acorn-E.

També un valor a tenir en compte és el nombre de llars participants en la mostra en cada moment, ja que si ho mirem temporalment, el nombre de llars ha anat augmentant fins a l'octubre del 2012 arribar a ser el màxim i a partir d'aquí disminueix suavament.

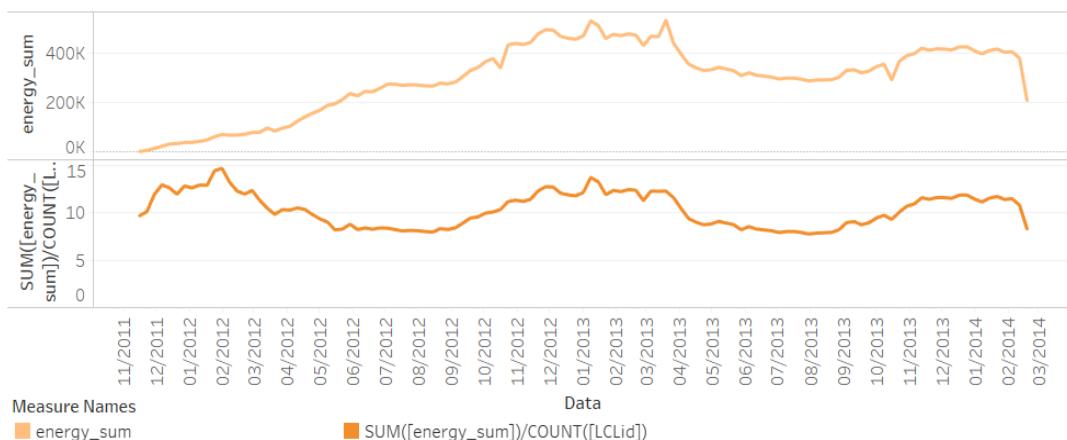
Evolució nombre de llars



Gràfic 12 Evolució nombre llars al llarg del temps

Aquest fet cal tenir-lo en compte, ja que quan estudiem el consum, és important tenir en compte el nombre de llars que estan consumint, ja que ens podria fer pensar que varia el consum per la influència d'alguna altre variable, quan realment és pel nombre de llars que estan consumint. Tal com veiem en la següent imatge, les dades relacionades amb el consum les hem dividit pel nombre de llars en cada moment, de manera que observem el consum mitjà per llar i no la suma de totes les llars.

Evolució consum. Diferència entre no tenir en compte el nombre de llars i si tenir-la en compte

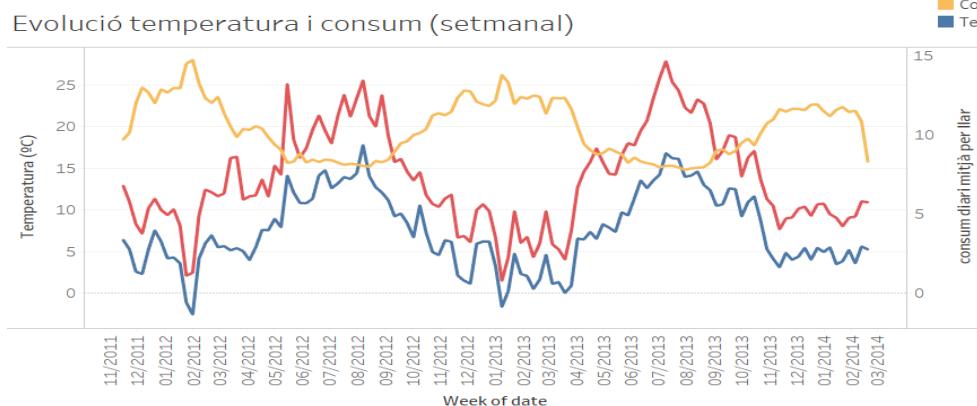
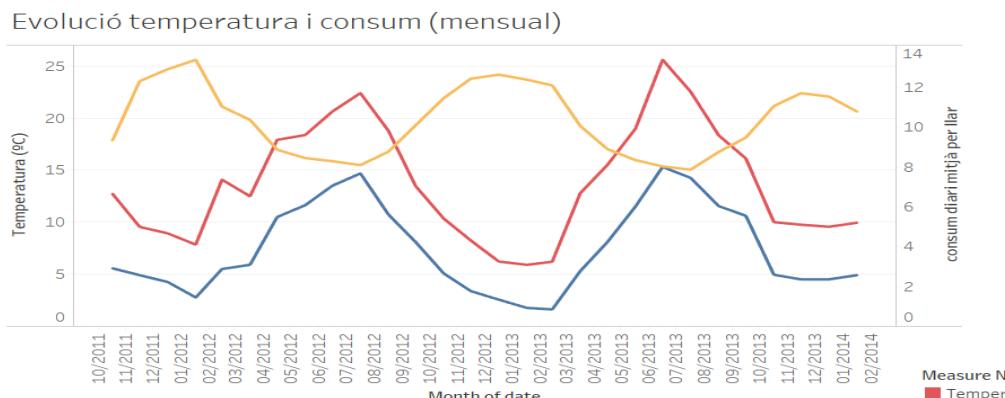


Gràfic 13 A dalt, gràfic de l'evolució del consum total , a sota, gràfic de l'evolució del consum mitjà per llar.

Gràcies a aquest fet, podem veure l'estacionalitat de les dades, on veiem que els mesos d'hivern (desembre, gener i febrer) és quan més es consumeix i que els mesos d'estiu (juny, juliol i agost) són els mesos que menys es consumeix.

Una variable que influeix perquè això passi és la temperatura. En el següent gràfic veiem l'evolució del consum juntament amb l'evolució de les temperatures màximes i mínimes, veient que el consum és inversament proporcional a la temperatura, és a dir, a menys temperatura, més consum.

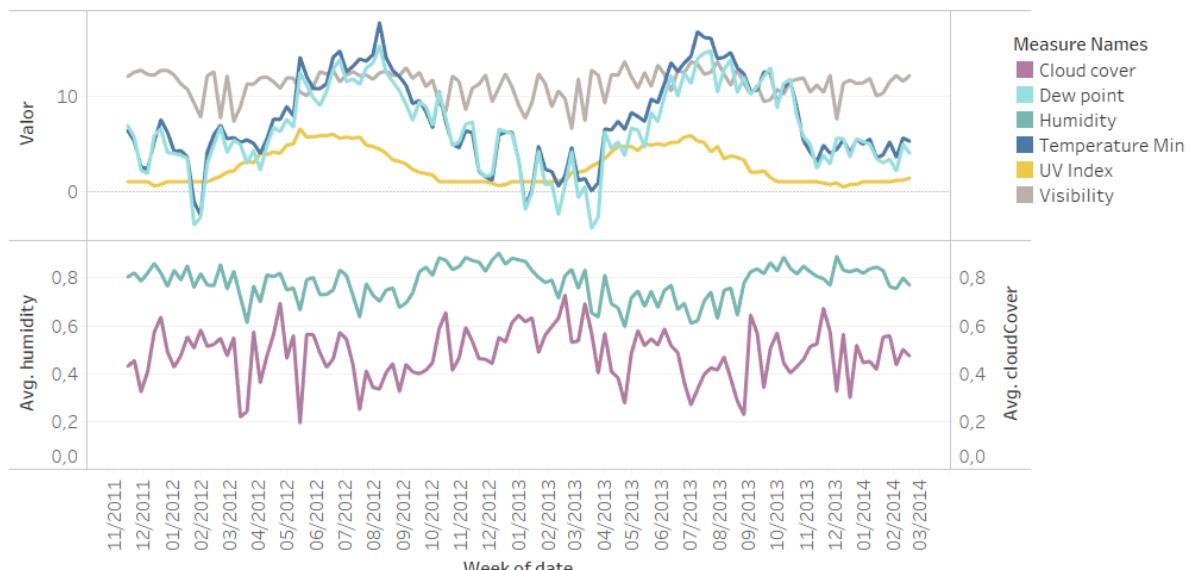
És interessant també veure que depenen de l'agregat que utilitzem per la data, veiem més o menys detall. En el primer gràfic hem utilitzat detall mensual i en el segon detall setmanal, de manera que veiem més clarament la variació. Si féssim el detall en diari, llavors el gràfic és massa confús.



Gràfic 14 A sobre, gràfic de l'evolució del consum mitjà i la temperatura mitjana a escala mensual, a sota a escala setmanal

Entenem que a part de la temperatura, hi ha altres fets meteorològics que poden afectar al consum, de manera que també els representem per veure gràficament la relació. En el gràfic següent veiem per exemple que l'atribut dewpoint (punt de rosada) quasi ressegueix la línia de la temperatura mínima i també uv index segueix la mateixa tendència. La visibilitat, humitat i la cobertura de núvols, tot i a priori semblar tenir relació amb la manera de consumir, gràficament no s'hi veu cap relació amb el consum.

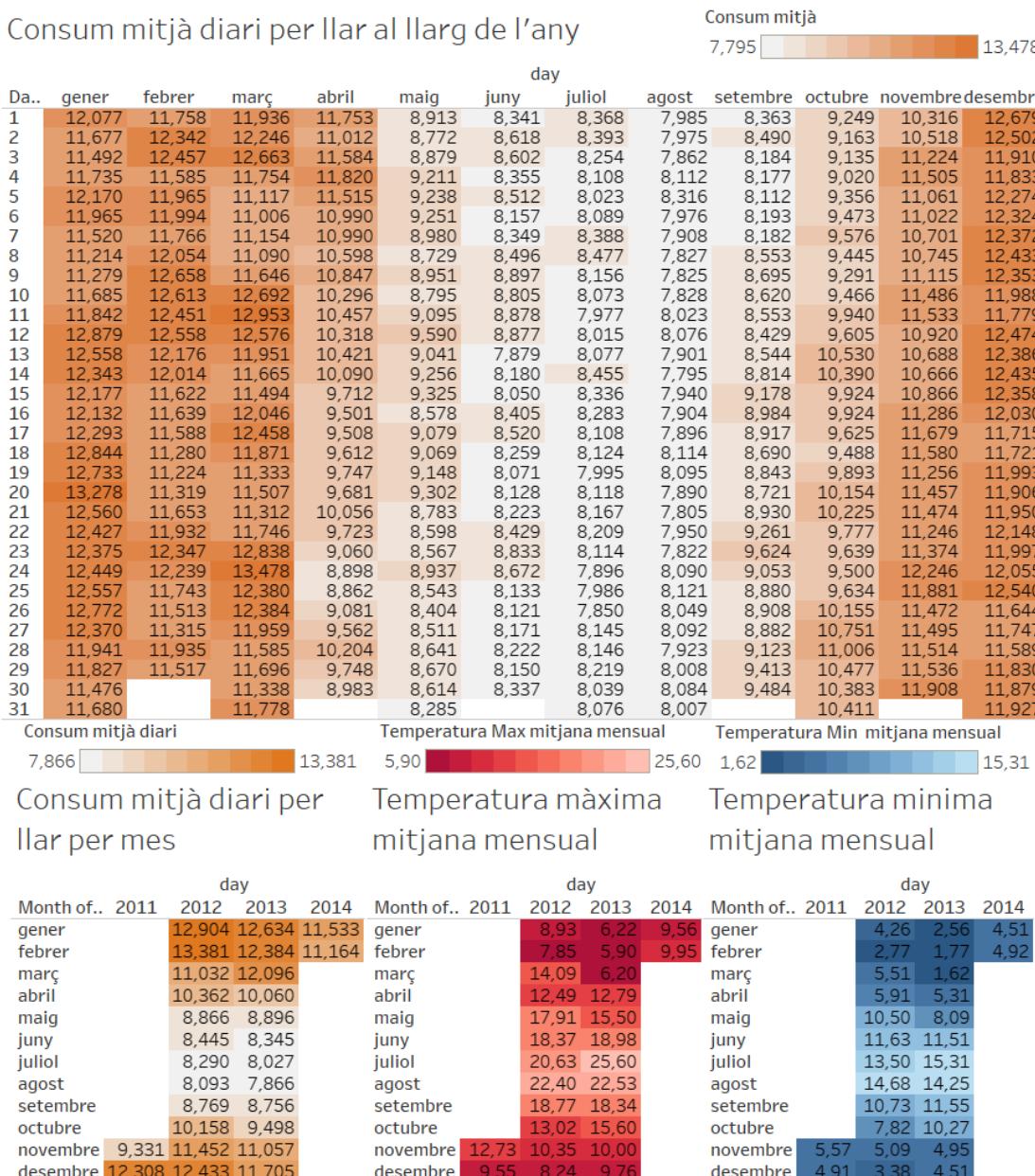
Evolució altres indicadors (setmanal)



Gràfic 15 Evolució diferents variables meteorològiques com cloud cover, dew point, humidity, temperature min, uv index i visibility

Per ser una mica més específics i veure el consum amb més detall, preparam un heatmap per a poder representar el consum anual de cada dia de l'any. En el heatmap podem confirmar que els mesos que es consumeixen més són desembre, gener, febrer i març, seguits pels mesos de novembre, abril i octubre, on el dia amb més consum de l'any és el 24 de març seguit pel 20 de gener. Els mesos amb menys consum són juliol i agost, on el dia amb menys consum de l'any és el 14 d'agost.

Si ens fixem en els heatmaps de temperatura, veiem que les temperatures més altes coincideixen amb els mesos de menys consum i que les temperatures més baixes coincideixen amb els mesos de més consum.

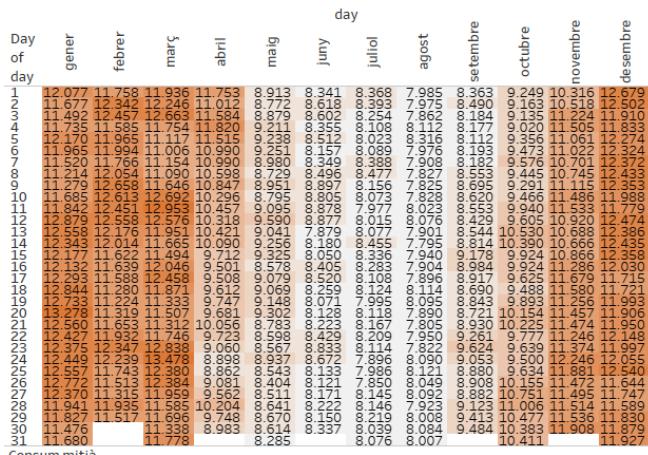


Gràfic 16 Consum mitjà per llar al llarg de l'any.

Veiem doncs que la temperatura és una variable que hem de tenir en compte.

També s'ha representat en un heatmap el consum des de diferents perspectives: consum mitjà diari, consum mitjà mig-horari(de mitja hora d'un dia) i consum màxim mig-horari (consum de la mitja hora que ha gastat més) , on veiem que els tres segueixen el mateix patró sigui en termes diaris o en termes d'estació.

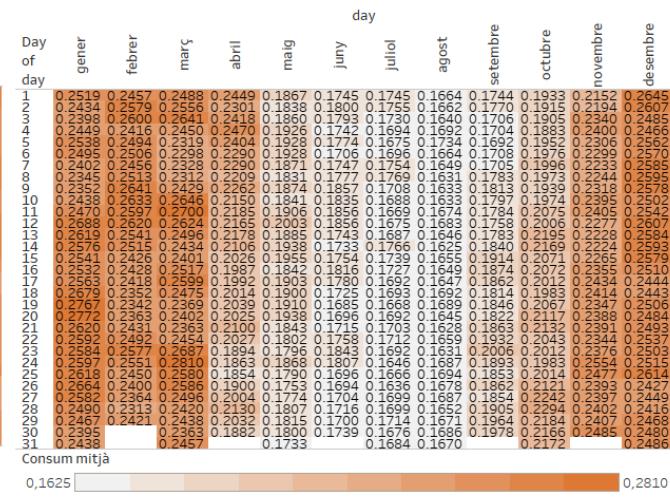
Consum mitjà diari per llar al llarg de l'any



Consum mitjà

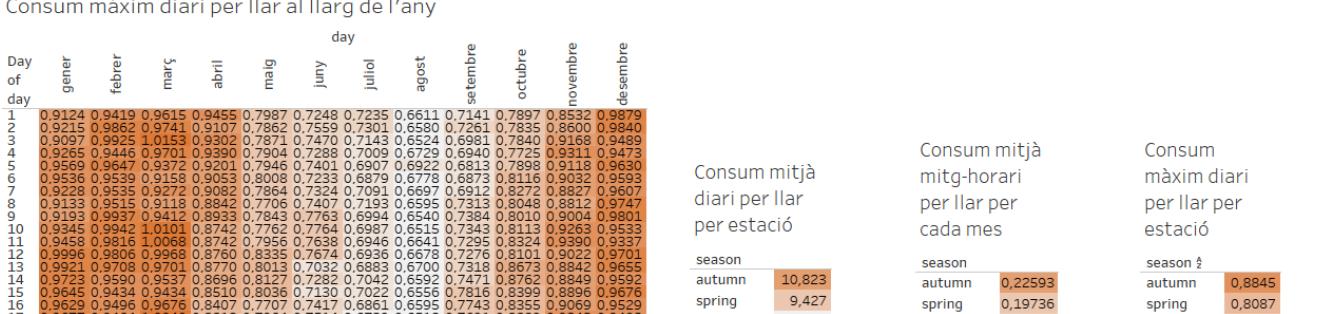
7,795

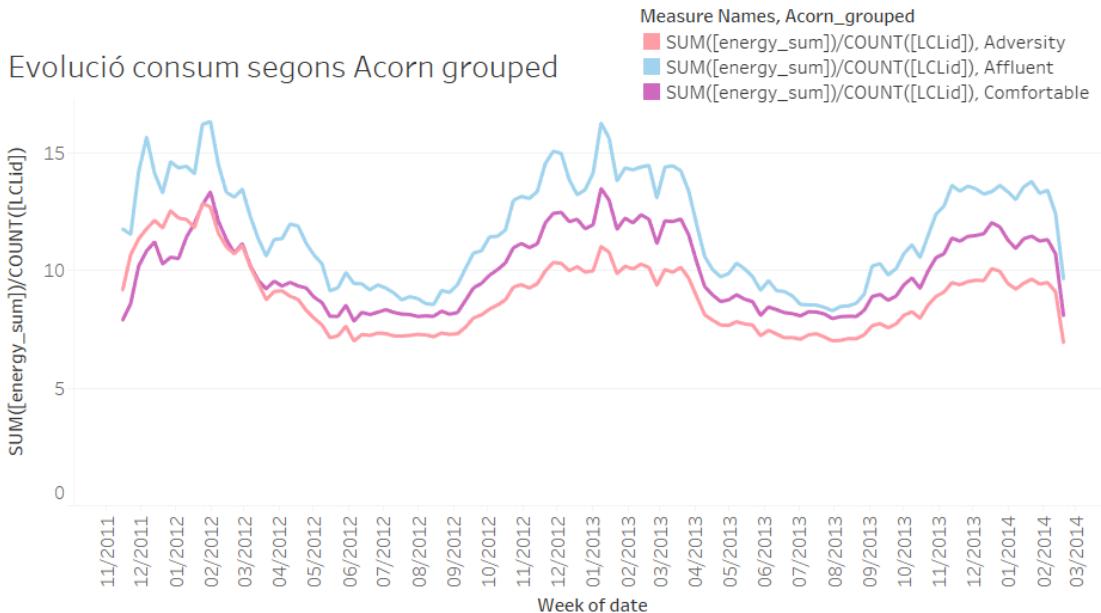
Consum mitjà mitjà horari per llar al llarg de l'any



Consum mitjà

0,1625





Gràfic 18 Evolució consum segons Acorn Agrupat

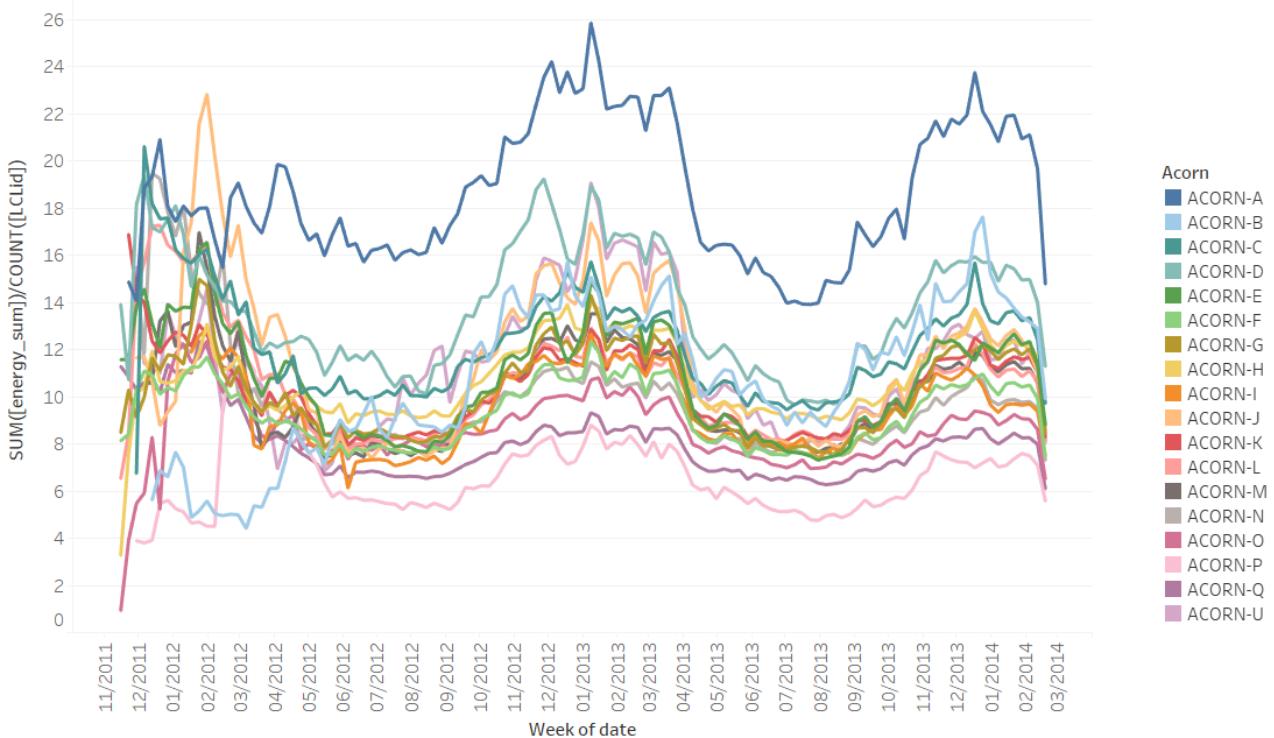
Mirem l'evolució del consum segons l'Acorn, i veiem de seguida que hi ha Acorn que tenen un consum molt diferent dels altres Acorns, però tot i així segueixen el mateix patró. Per exemple, veiem que els de l'Acorn A consumeixen molt, arribant a consumir fins a 26 kW diaris al gener del 2013. Per l'altra banda, veiem que Acorn P i Acorn Q són els tipus de consumidors que consumeixen menys en èpoques de gran demanda, consumint tant sols entre 8 i 9 kW al gener del 2013.

El següent gràfic no sembla molt entenedor, però si s'utilitza Tableau per explorar-lo, és molt més fàcil ja que permet destacar les línies que interessen i en mostra els valors.

En el mateix gràfic veiem que tot i utilitzar el consum diari per llar i no el total per totes les llars, veiem com del novembre del 2011 al març del 2012 no tots els Acorns segueixen el mateix patró, suposem pel fet de que hi hagi masses poques mostres i que sigui poc representatiu.

Això ens fa pensar que no ens podem fiar de les dades del 2011 i principis del 2012 ja que les mostres no són representatives i no ens servirien per a fer prediccions, ja que podrien sortir esbiaixades per aquest fet, de manera que no les utilitzarem per entrenar el model.

Evolució consum segons Acorn



Gràfic 19 Evolució consum segons Acorn

Com hem comentat, el gràfic anterior és complicar de visualitzar si no s'utilitza Tableau, de manera que realitzem el mateix gràfic filtrant els Acorn per Acorn agrupat, de manera que obtenir els tres següents gràfics.

En el primer gràfic on es mostren els del grup Affluent, que conté de l'Acorn A a l'Acorn E, veiem clarament com l'Acorn A destaca per sobre els altres pel seu consum més elevat i que l'Acorn E, que és el que té més mostres dins del grup d'Affluent, és el que té menys consum, consumint menys fins i tot que alguns Acorn de Comfortable.

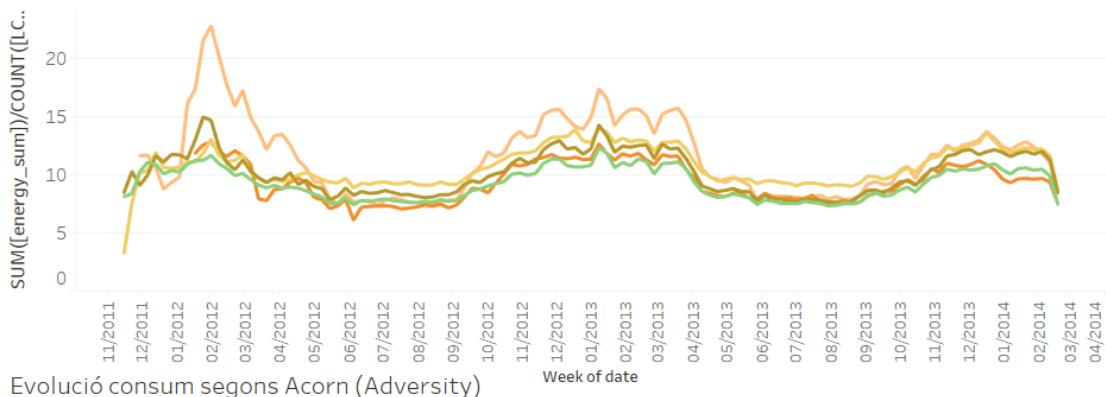
En el segon gràfic on es mostren els del grup Comfortable, que conté de l'Acorn F a l'Acorn J , veiem que l'Acorn J destaca pel seu consum més elevat en èpoques de més consum però molt baix en èpoques de baix consum. L'Acorn F i H, que són els que tenen més pes, en aquest cas sí que són representatius de la mostra, sobretot si eliminéssim d'aquest grup a l'Acorn J, que té un patró diferent. Si posem l'Acorn J al grup Affluent, sembla que hi encaixa més.

En l'últim gràfic on es mostra el grup Adversity, que conté de l'Acorn K a l'Acorn U, veiem que l'Acorn U destaca per sobre els altres pel seu consum més elevat. I sembla que també correspondria al grup Affluent. Veiem que l'Acorn que té més representació en aquest grup és l'Acorn Q, tot i no ser el que millor defineix el grup.

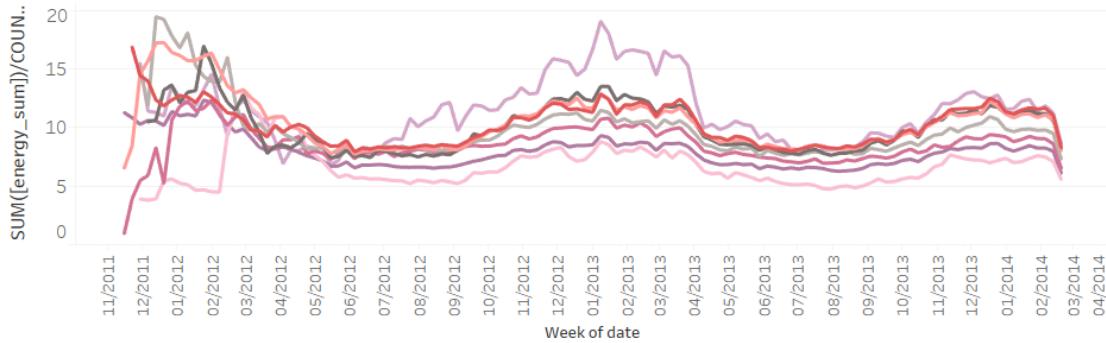
Evolució consum segons Acorn (Affluent)



Evolució consum segons Acorn (Comfortable)



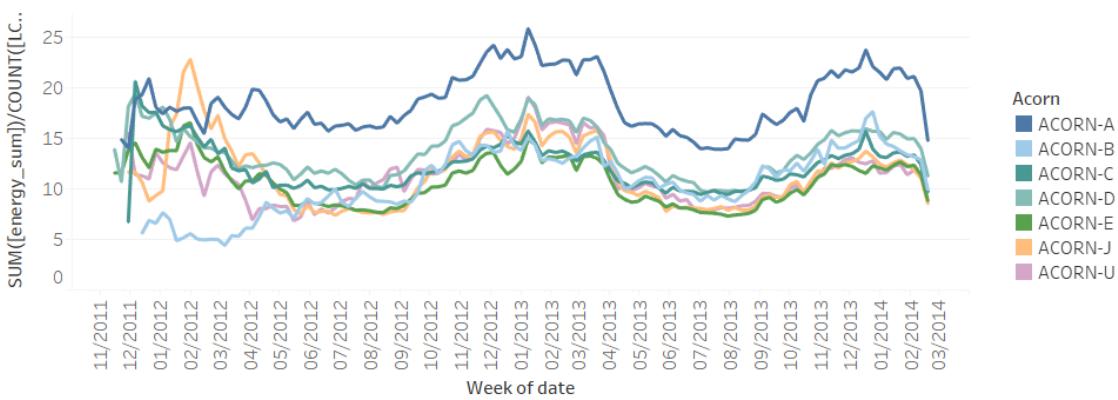
Evolució consum segons Acorn (Adversity)



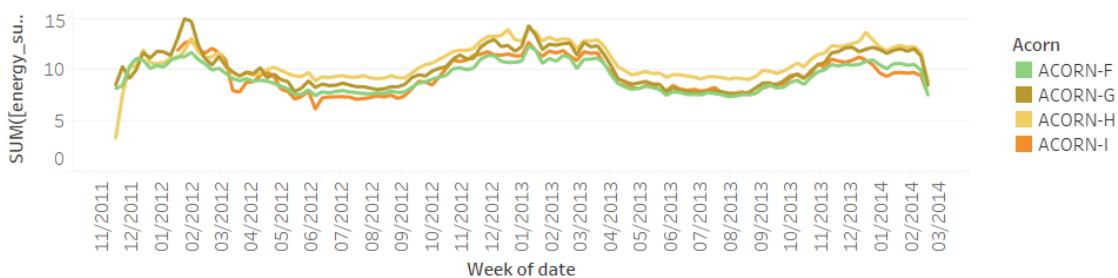
Gràfic 20 Evolució consum segons Acorn

Seguidament mostrem el mateix gràfic però amb els Acorn J i U canviats al grup Affluent, de manera que ara sí que tots els grups segueixen el mateix patró. De cara a les prediccions, fins i tot es podria fer un grup per l'Acorn A.

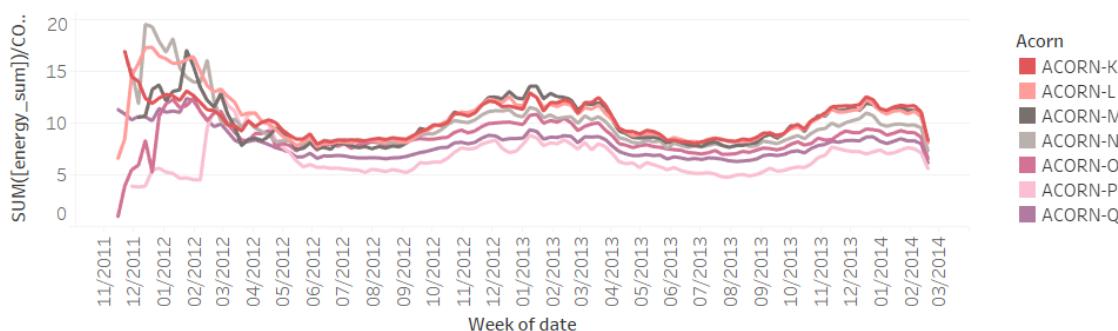
Evolució consum segons Acorn (Affluent)



Evolució consum segons Acorn (Comfortable)



Evolució consum segons Acorn (Affluent)



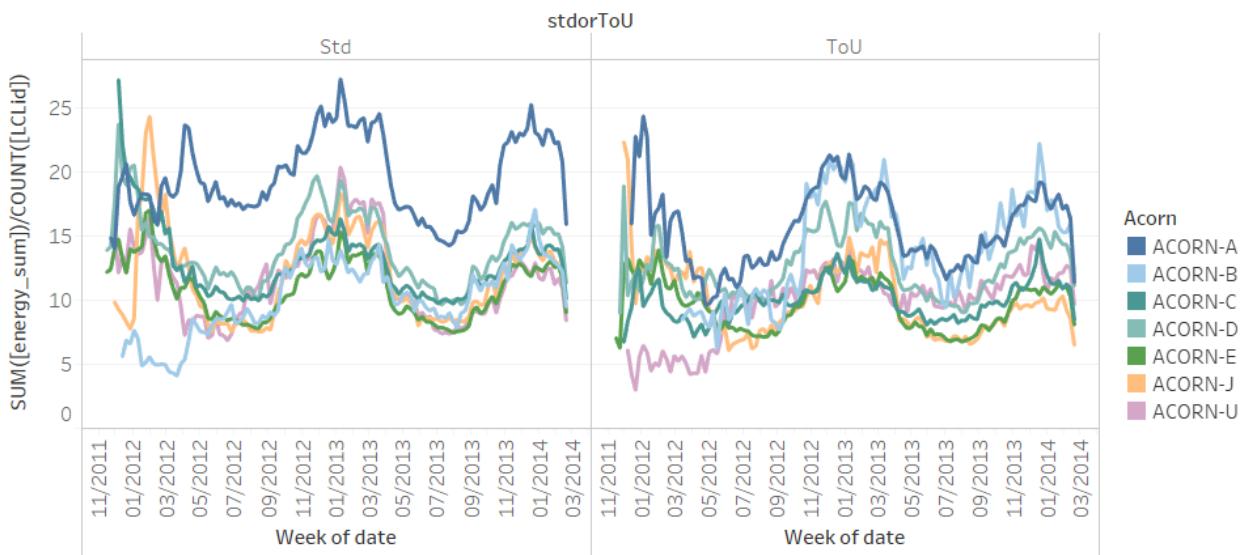
Gràfic 21 Evolució consum segons Acorn adequant grups

Amb aquests grups creats, els separam per tipus de tarifa, i veiem que dins d'algun Acorn el consum és diferent segons la tarifa. Per exemple veiem que en el grup Affluent, l'Acorn A que destacava en el grup, ara només destaca en el grup amb tarifa Std i no en el grup amb tarifa ToU. També veiem que varien molt els consums en moment de gran demanda dels Acorns U i B, on l'Acorn U amb tarifa Std consumeix més que el de tarifa ToU, en canvi els de l'Acorn B, consumeixen més els del grup ToU.

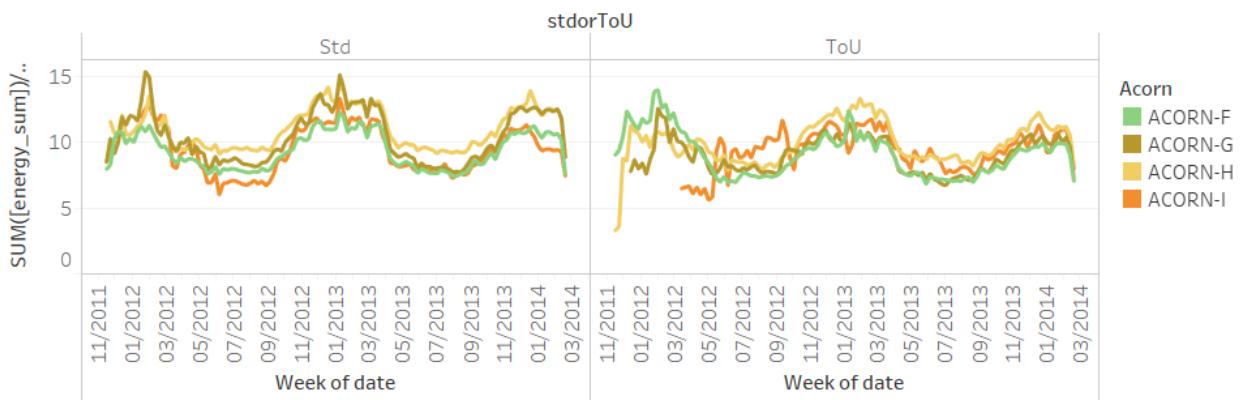
També veiem que del grup Comfortable, l'Acorn que varia més el consum és el G.

En quant al grup d'Adversity, veiem que el patró de consum de l'Acorn P canvia dràsticament segons la tarifa, i també és curiós veure que hi ha alguns Acorns que amb la tarifa ToU consumeixen més, com l'Acorn K.

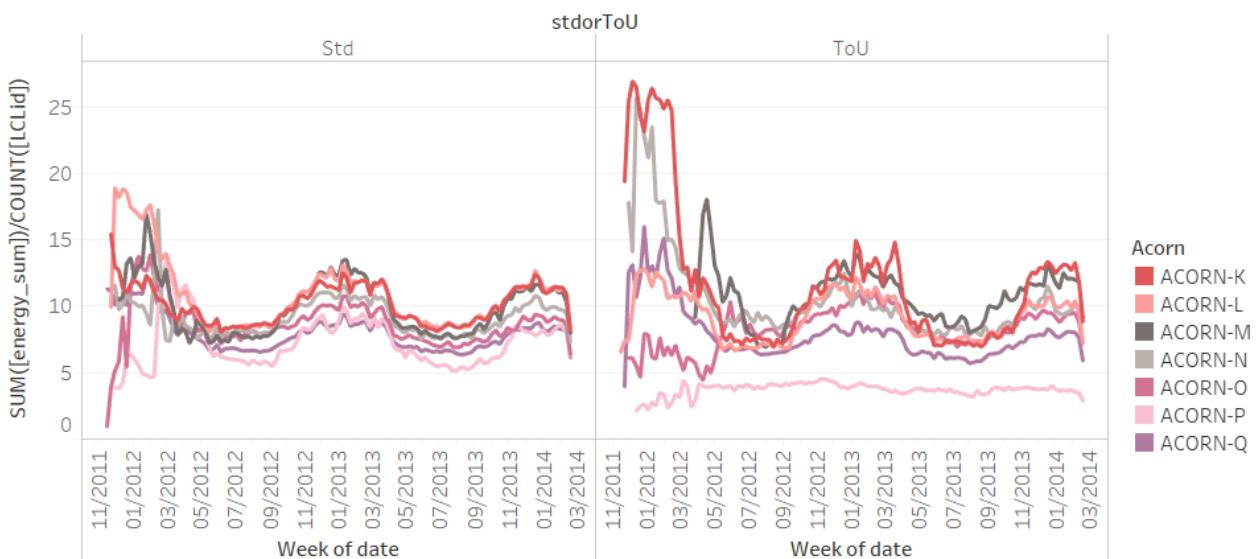
Evolució consum segons Acorn (Affluent)



Evolució consum segons Acorn (Comfortable)

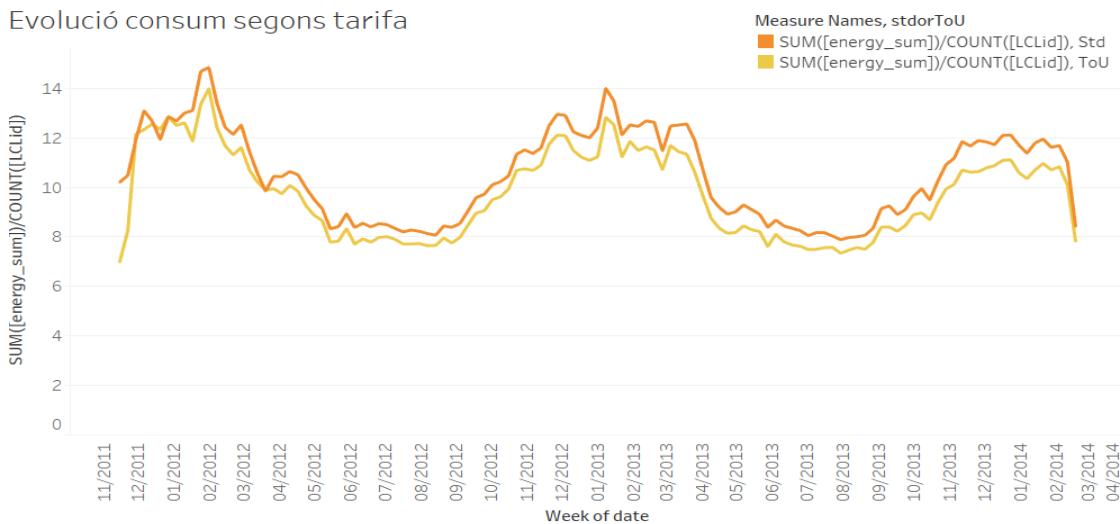


Evolució consum segons Acorn (Affluent)



Gràfic 22 Evolució consum segons Acorn i tarifa

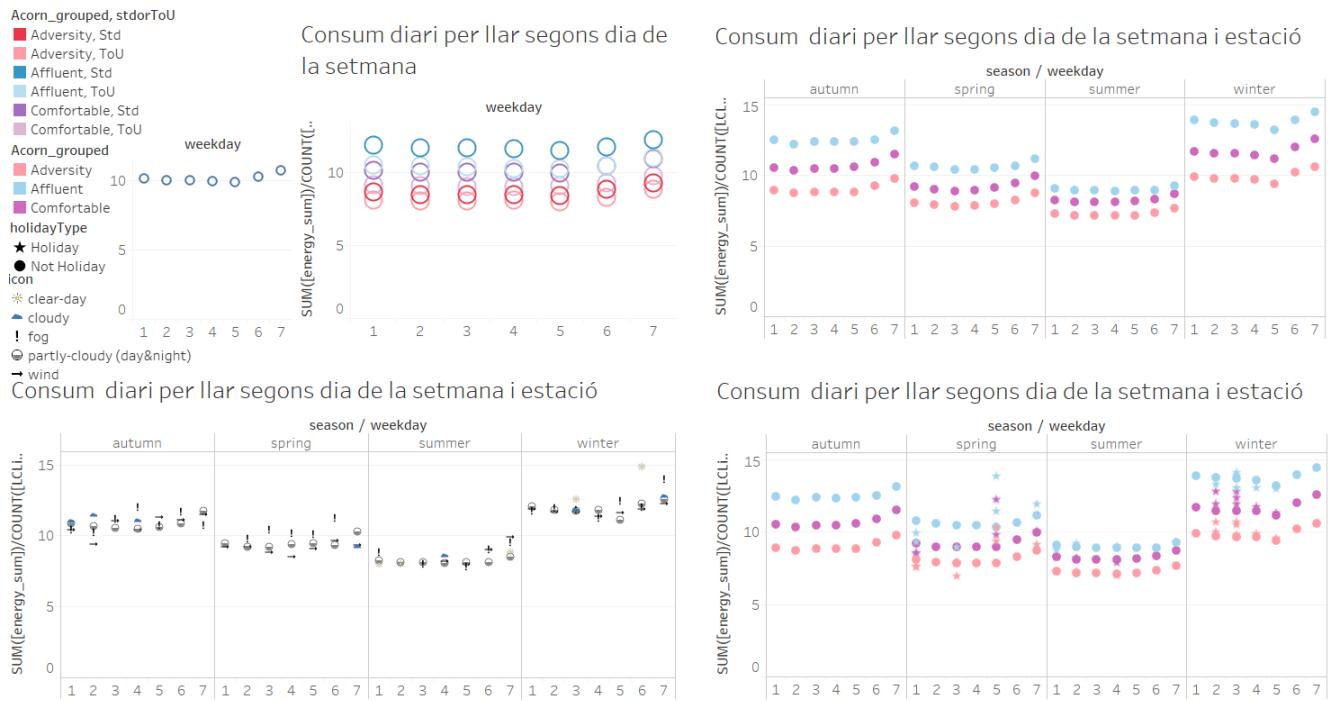
Tot i així, de mitjana els que tenen tarifa Std consumeixen més que els que tenen tarifa ToU.



Gràfic 23 Evolució consum segons tarifa

Altres variables que influeixen en el consum són el dia de la setmana, el mes, l'estació i fins i tot l'hora en la que es pon el sol.

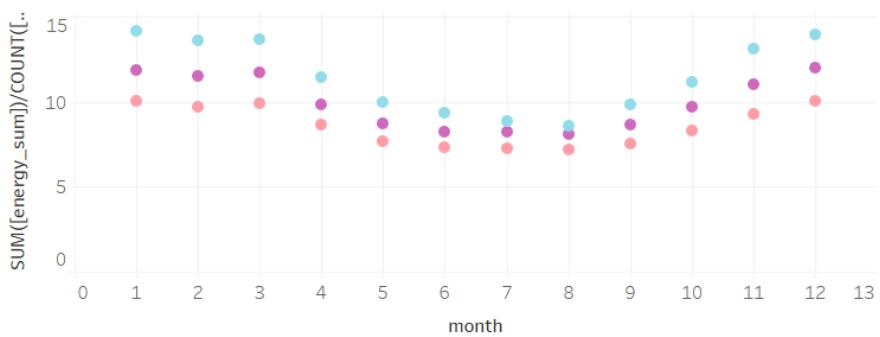
Si mirem la següent imatge, veiem que hi ha diferents gràfics que representen el consum en funció del dia de la setmana. En general veiem que el dia que les llars consumeixen més és el diumenge i el dia que consumeixen menys el divendres. Si utilitzem l'estació (gràfic de la dreta a dalt). Veiem clarament com l'estació que es consumeix menys és l'estiu seguit de la primavera, la tardor i l'hivern. També en aquest gràfic veiem com la diferència entre Acorn grouped a l'estiu quasi és indiferent en canvi a l'hivern hi ha molta diferència. També hem provat de classificar els consums per icon i per si és un dia festiu, no obtenint cap relació.



Gràfic 24 A dalt, Gràfic consum mitjà per llar per dia de la setmana, gràfic consum mitjà per llar per dia de la setmana pintat segons Acorn grouped i tarifa, gràfic consum mitjà diari per dia de la setmana i estació pintat segons Acorn grouped. A sota, gràfic consum mitjà per llar per dia de la setmana segons estació pintat segons icon i gràfic consum mitjà diari per dia de la setmana pintat segons Acorn agrupat i dia festiu

Si ens fixem en el següent gràfic on tenim el consum diari mitjà mensual, tornem a veure que els mesos de juliol i agost el consum és mínim i que els mesos de desembre i gener són els que presenten més consum. També veiem el mateix que passava per dia de setmana on quan hi ha menys consum, els tres grups d'Acorn agrupat tenen consums molt semblants i a mesura que augmenta el consum, els grups es separen i tenen un consum diferent.

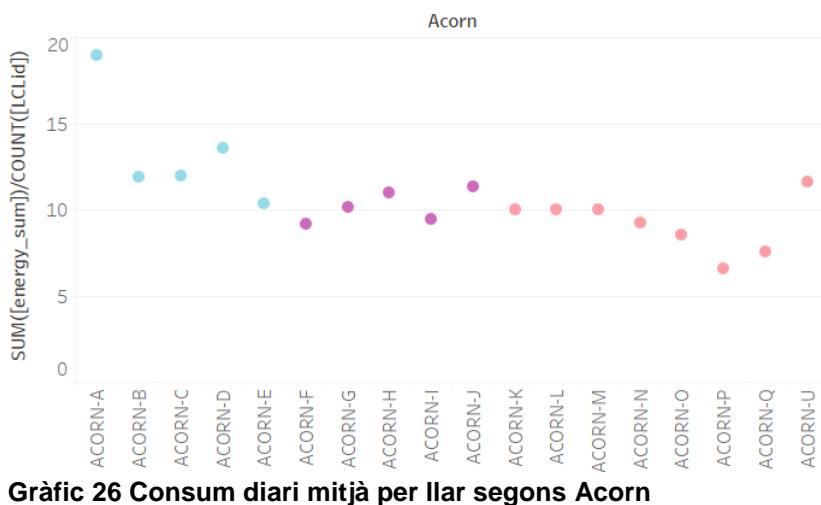
Consum diari per llar segons mes de l'any



Gràfic 25 Consum diari mitjà per llar segons mes de l'any

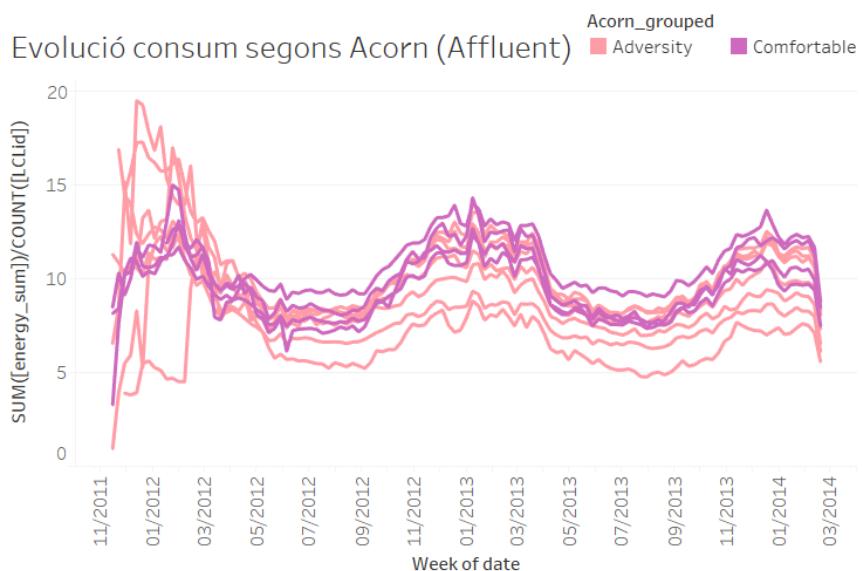
En el següent gràfic veiem el consum diari mitjà de cada Acorn, sent l'Acorn A el que consumeix més i l'Acorn P el que menys. Veiem que el consum dins de cada Acorn agrupat és desigual, i que de l'Acorn K a l'Acorn O podrien ser perfectament del grup Comfortable.

Consum diari per llar segons Acorn



Gràfic 26 Consum diari mitjà per llar segons Acorn

Si ho mirem temporalment, veiem que certament els valors d'aquests Acorn són semblants com es veu al següent gràfic, on es veu perfectament que els grups adversity i affluent es solapen menys els dos Acorns amb menys consum, Q i P.



Gràfic 27 Evolució consum mitjà diari per llar segons Acorn (llars que són Affluent)

En els següents gràfics veiem representat el consum segons l'hora de posta de sol, on es veu clarament que com més d'hora es pon el sol, més es consumeix.

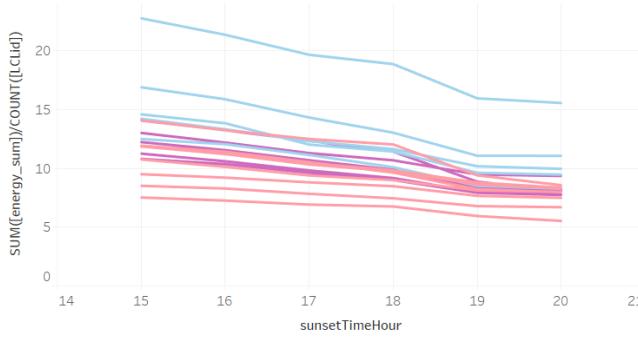
En el gràfic de dalt a l'esquerra, veiem el consum dels diferents Acorn segons l'hora de posta de sol, on veiem que segueixen un patró molt semblant. Quan ho mirem segons

l'Acorn agrupat al grafic de dalt a la dreta, veiem que la diferència entre el consum dels Acorn agrupats es fa més petita a mesura que augmenta l'hora a la que es pon el sol, alhora que es consumeix menys.

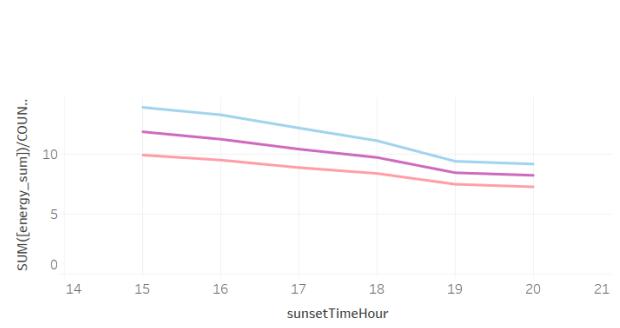
En el gràfic de baix a l'esquerra, veiem també la diferència del consum segons l'hora de posta de sol des de la perspectiva de la tarifa, on la ToU sempre està en global una mica per sota de la tarifa Std.

En el gràfic de baix a la dreta, utilitzem la perspectiva d'estació per a veure la diferència de consum entre les diferents hores de posta de sol. Veiem que a l'hivern el consum no varia encara que canviï l'hora a la que es pon el sol, i a l'estiu la variació és mínima. En canvi el consum de la tardor i primavera es veu molt afectat pel canvi d'hora de posta de sol.

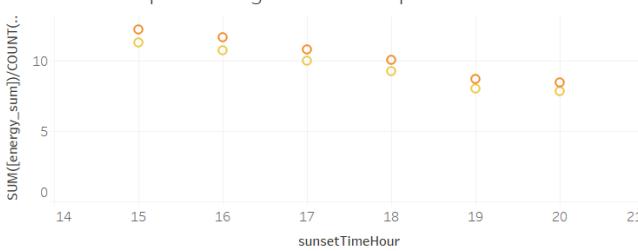
Consum diari per llar segons l'hora de posta de sol



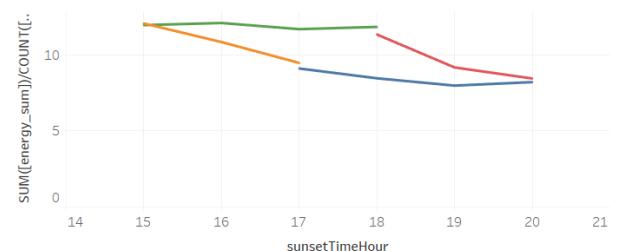
Consum diari per llar segons l'hora de posta de sol



Consum diari per llar segons l'hora de posta de sol



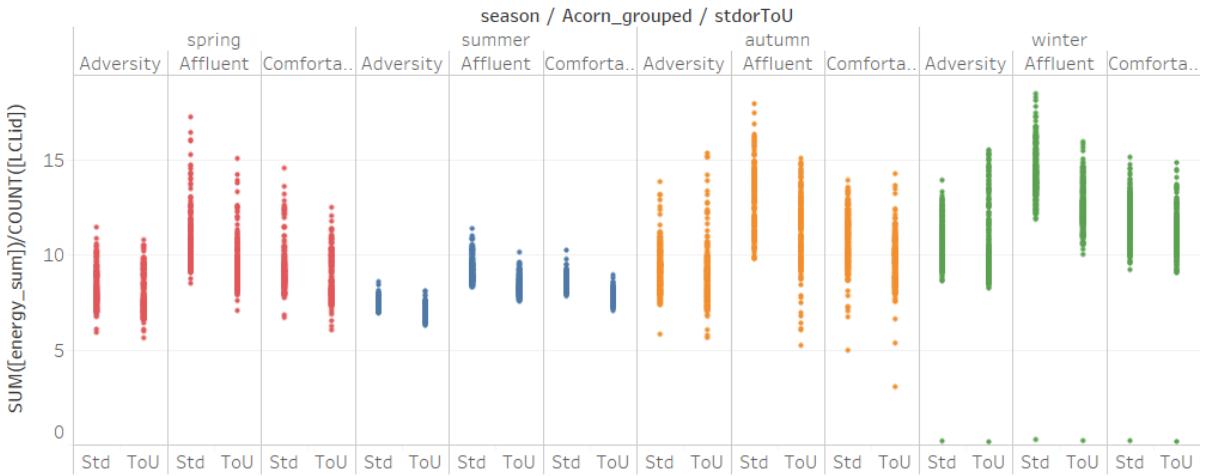
Consum diari per llar segons l'hora de posta de sol



Gràfic 28 A dalt a l'esquerra, gràfic consum diari mitjà per llar segons hora de posta de sol pintat segons Acorn agrupat i tarifa , a dalt a la dreta, gràfic consum diari mitjà per llar segons hora de posta de sol pintat segons Acorn agrupat. A sota a l'esquerra, gràfic consum diari mitjà per llar segons hora de posta de sol i pintat segons tarifa, a sota a la dreta, gràfic consum diari mitjà per llar segons hora de posta de sol pintat segons estació.

Si ens fixem en el següent gràfic, on es representa el consum segons l'estació, l'Acorn agrupat i la tarifa, veiem com el consum dels que tenen tarifa ToU quasi sempre és inferior als que tenen tarifa Std, menys en el cas dels adversity i comfortable a la tardor i els d'adversity i comfortable a l'hivern. També veiem que el consum dels affluent és més alt en totes les estacions, seguit de comfortable i adversity, tot i que durant la tardor i l'hivern els comfortable i adversity consumeixen de manera molt semblant. Destacar que es veu clarament que a l'estiu el consum és mínim.

Consum diari segons estació, Acorn agrupat i tarifa



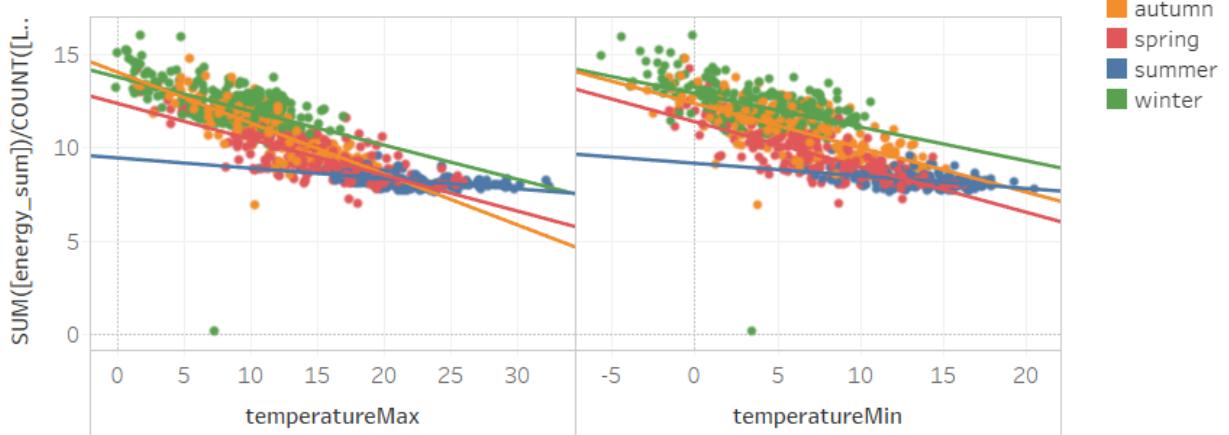
Gràfic 29 Consum diari mitjà per llar segons estació, Acorn agrupat i tarifa

Fins ara hem fet un resum dels fets que hem trobat amb les variables descriptives, i ara ens centrarem en les variables quantitatives (les meteorològiques).

Com ja hem vist, la temperatura és un factor que afecta totalment al consum, de manera que en el següent gràfic representem el consum respecte la temperatura segons l'estació. En primer lloc veiem clarament que les temperatures són més altes a l'hivern, seguit per la tardor, la primavera i finalment l'estiu.

També veiem que el canvi de temperatura a l'estiu quasi no afecta al consum (la línia de tendència és quasi horitzontal). En el cas de temperatura màxima, l'estació que veu el seu consum més afectat per la temperatura és amb diferència la tardor (la línia taronja és la que té més pendent). En el cas de la temperatura mínima, l'estació que es veu més afectada és la primavera, però amb molt poca diferència amb la tardor.

Consum diari segons temperatura



Gràfic 30 Consum diari mitjà segons temperatura màxima i mínima pintat en funció de l'estació

A part de la temperatura, volem saber quines altres variables qualitatives afecten al consum, de manera que les representem en funció del consum, pintant-les segons estació, tarifa i Acorn agrupat.

En els tres gràfics en conjunt veiem clarament com les variables dewpoint, temperatura màxima apparent i temperatura mínima apparent són les que influeixen més en el consum des de les tres perspectives (estació, tarifa i Acorn agrupat). Entenem que dewpoint, que és el valor de punt de rosada, al estar totalment relacionat amb la temperatura mínima, segueix la mateixa tendència, per tant tampoc serà una variable que tinguem en compte. També globalment veiem que la variable de fase de la lluna tampoc té relació amb el consum.

En relació a la variable pressió, veiem que quasi no influeix en el consum per tarifa i el consum per Acorn agrupat, però si influeix en les estacions hivern i primavera, on veiem que a l'hivern a menys pressió menys consum i a més pressió més consum i a la primavera, a menys pressió més consum i a més pressió menys consum.

En quant a la visibilitat, veiem que a menys visibilitat més consum i a més visibilitat, menys consum. Observem que aquesta variable influeix més als mesos d'hivern, primavera i estiu, quasi no tenint influència els mesos de tardor.

Veiem que les variables windBearing i cloudCover quasi no tenen influència, veient tant sols una mica de pendent a cloudCover on a menys tapat, menys consum i a més tapat més consum, influint de manera igual a totes les estacions.

En quant a la velocitat del vent, veiem com influeix sobretot en la estació primavera, on a menys velocitat del vent, menys consum, i a més velocitat del vent, més consum.

En quant a la humitat, en els gràfics de tarifa i Acorn agrupat veiem com a menys humitat menys consum i que a més humitat més consum, i la humitat afecta de manera diferent a l'estació hivern, on veiem que el consum quasi no varia pel canvi en la humitat.

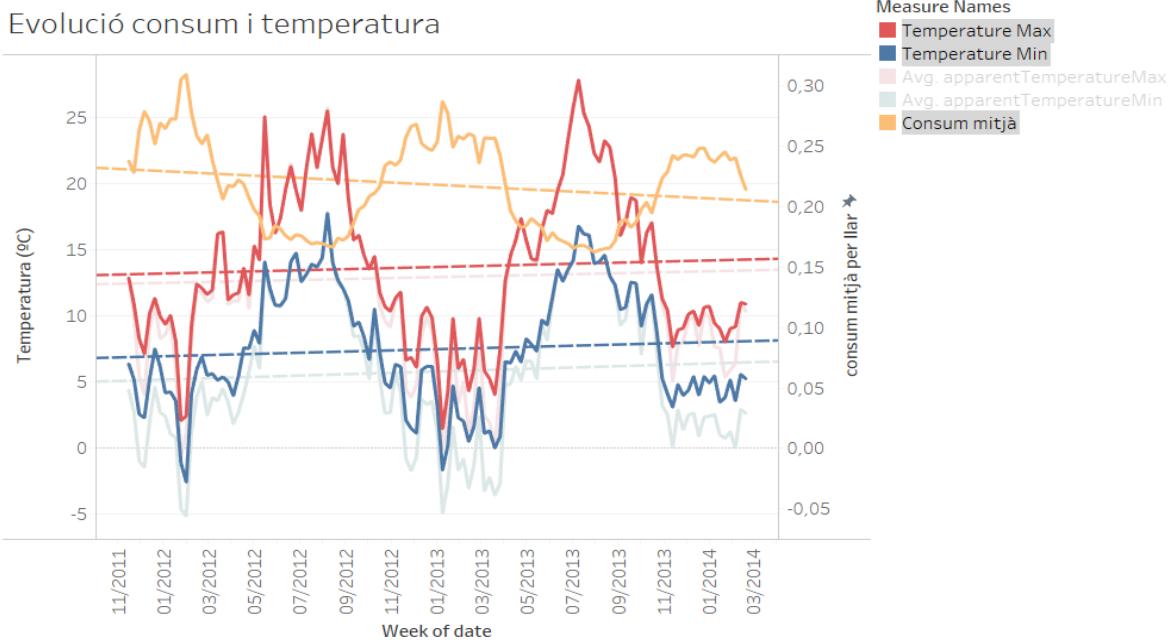


Gràfic 31 Gràfic consum diari mitjà segons varis indicadors (dewPoint, apparentTemperatureMax, apparentTemperatureMin, pressure, visibility, windbearing, windspeed, humidity, cloudcover i moonphase) pintat en funció d'estació, Gràfic consum diari mitjà segons varis indicadors pintat en funció de la tarifa i Gràfic consum diari mitjà segons varis indicadors pintat en funció de Acorn agrupat.

En veure que la temperatura apparent té més influència que la temperatura, ja que les línies de tendència tenen més pendent, representem altre cop el gràfic de la evolució del consum i la temperatura. En color més clar hi ha representades les temperatures apparents, que la diferència amb la temperatura normal rau en que en els moments de menys temperatura, la temperatura apparent és més baixa.

Una altre observació interessant és que veiem que la tendència anual és que cada cop les temperatures siguin més altes i, per tant, que no es consumeixi tant en períodes de fred. Amb tants pocs anys com a mostra no podem afirmar que el consum anual seguirà

disminuint, ja que no sabem si ha estat casualitat dels anys que tenim en la mostra o que realment la temperatura serà cada vegada més elevada.

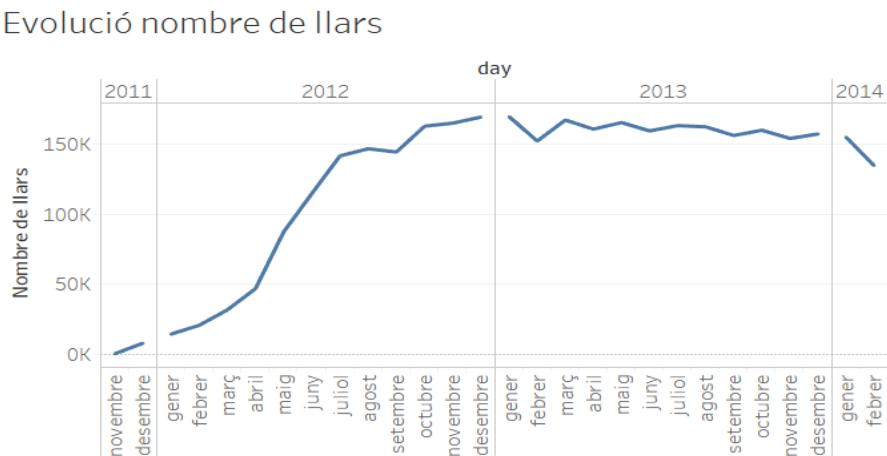


Gràfic 32 Evolució consum mitjà per llar i variables de temperatura.

En alguns dels gràfics hem vist un punt aïllat, que és del consum del 28 de febrer del 2014, últim dia que s'agafaven dades de la mostra, on només s'ha recomptat una mitja hora i per això el consum total del dia és tant baix. També en els gràfics de línies temporals veiem com l'últim dia el consum baixa en tots els grups.

4.3.2 Anàlisi dades horàries

Abans de començar l'anàlisi, fem èmfasi en que el nombre de llars utilitzat durant l'estudi varia al llarg del temps com podem veure al següent gràfic, així que hem representat el consum mitjà horari per llar enlloc del consum horari total de totes les llars.



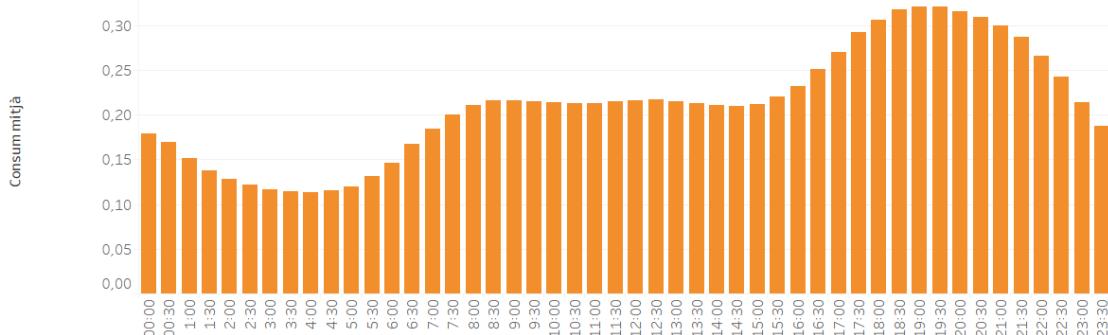
Gràfic 33 Evolució nombre llars dataset

Tradicionalment, el consum de l'energia horària es divideix en punta pla i vall, on el pla són de 8.00 h a 18.00 h i de 22.00 h a 00.00 h, la punta de 18.00 h a 22.00 h i el vall de 0.00 h a 8.00 h. Aquests són els horaris d'hivern (a l'estiu canvi degut al canvi horari), i són els que utilitzarem com a notació. Per començar amb l'anàlisi visual, ens fixem en el patró horari que segueix el consum durant el dia.

Veiem que el patró horari segueix perfectament aquest tipus de patró que hem descrit, i en vall és quan es consumeix menys, en pla és el consum mitjà i en punta quan es consumeix més

Com veiem al següent gràfic, el consum més gran durant el dia es produeix entre les 18.30 h i les 20.00 h, on es consumeix uns 0.3kWh cada mitja hora per cada llar, i a partir d'aquesta hora baixa fins arribar al consum més baix del dia, entre les 3.00 h i les 4.30 h de la matinada, que es consumeixen uns 0.12kWh, i llavors va pujant fins a les 8.30 h que s'estabilitza a un consum d'uns 0.22kWh fins les 15.30 h, que torna a pujar fins arribar al pic.

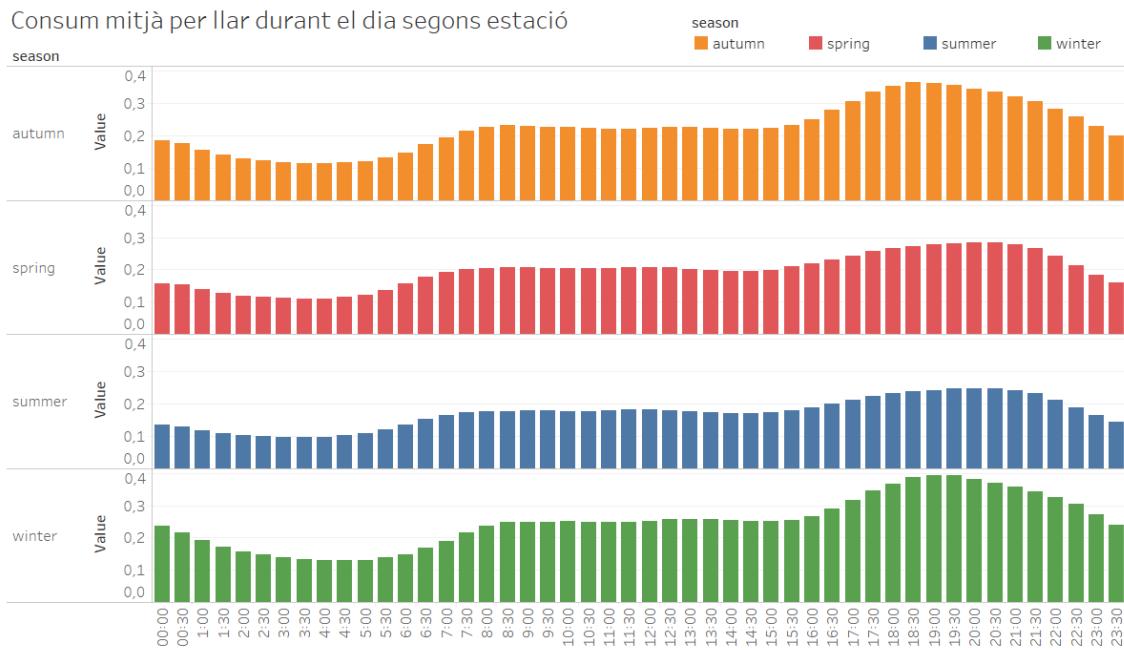
Consum mitjà per llar durant el dia



Gràfic 34 Consum mitjà mig-horari (cada mitja hora) per llar durant el dia

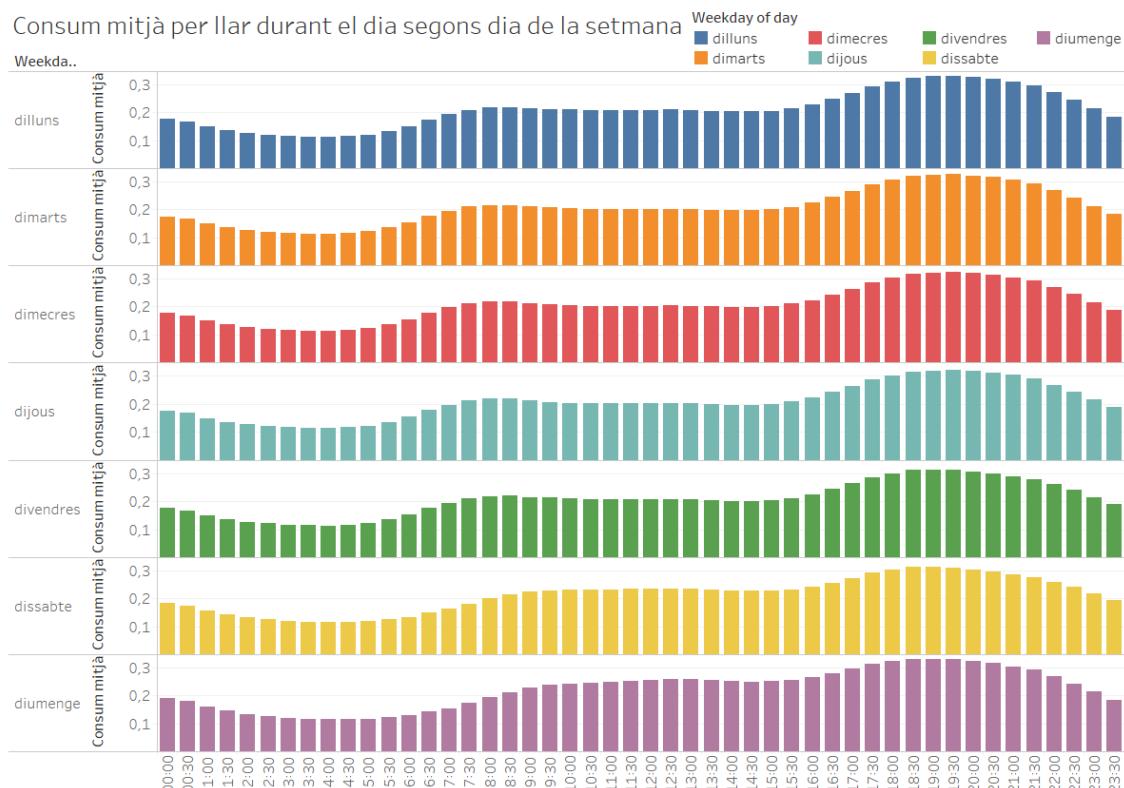
També verifiquem si aquest patró horari de consum és el mateix per a totes les estacions, i en el següent gràfic observem que hi ha dos patrons diferenciats però semblants, els de tardor i hivern i els de primavera i estiu.

Tot i veure que en hivern i tardor hi ha més consum que en primavera i estiu, la diferència principal del patró horari rau en que els períodes vall i punta són més pronunciats.



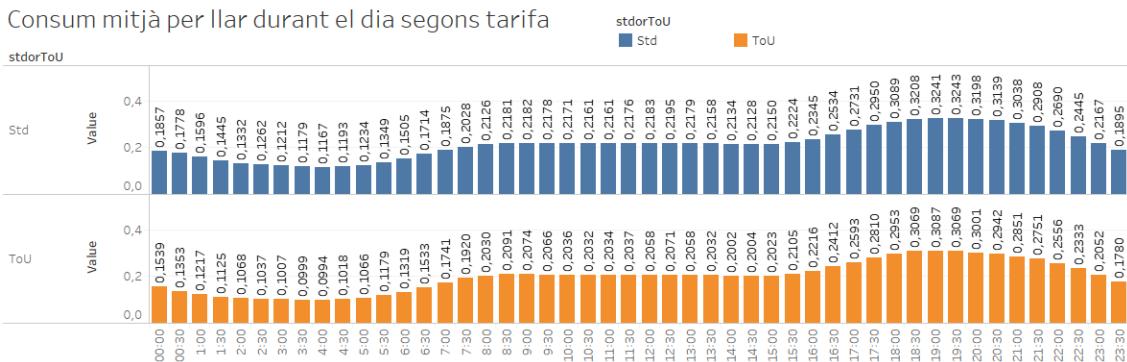
Gràfic 35 Consum mitjà mig-horari per llar durant el dia pintat segons estació

Analitzem també el patró del consum horari segons el dia de la setmana, on en el següent gràfic observem que els patrons són quasi iguals, amb la única diferència que dissabte i diumenge en el període pla hi ha més consum.



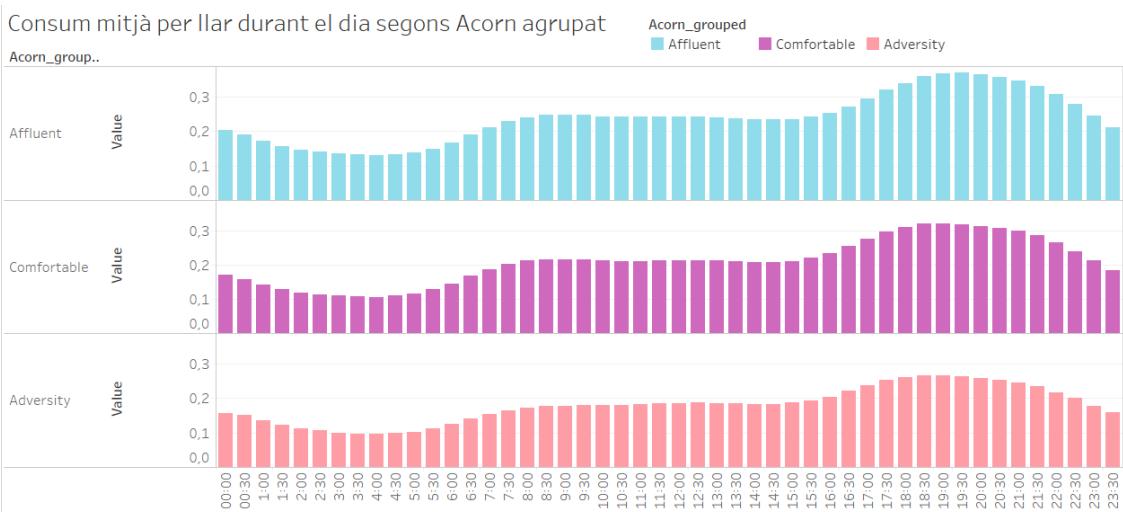
Gràfic 36 Consum mitjà mig-horari per llar durant el dia pintat segons dia de la setmana

També ens interessa saber si hi ha un patró diferent segons la tarifa de la llar, i en el següent gràfic confirmem que el patró de consum és el mateix, on la única diferència entre tarifes és que les llars que tenen ToU consumeixen menys que els que tenen Std.



Gràfic 37 Consum mitjà mig-horari per llar durant el dia pintat segons tarifa

Un altre atribut és el tipus de consumidor, on si mirem el següent gràfic on veiem el patró de consum horari segons l'Acorn agrupat, es veu clarament com els Affluent consumeixen més que els Comfortable, i que els Comfortable consumeixen més que els Adversity. També veiem que el període punta de Adversity és més pla i que el de Affluent és més pronunciat.



Gràfic 38 Consum mitjà mig-horari per llar durant el dia pintat segons Acorn agrupat

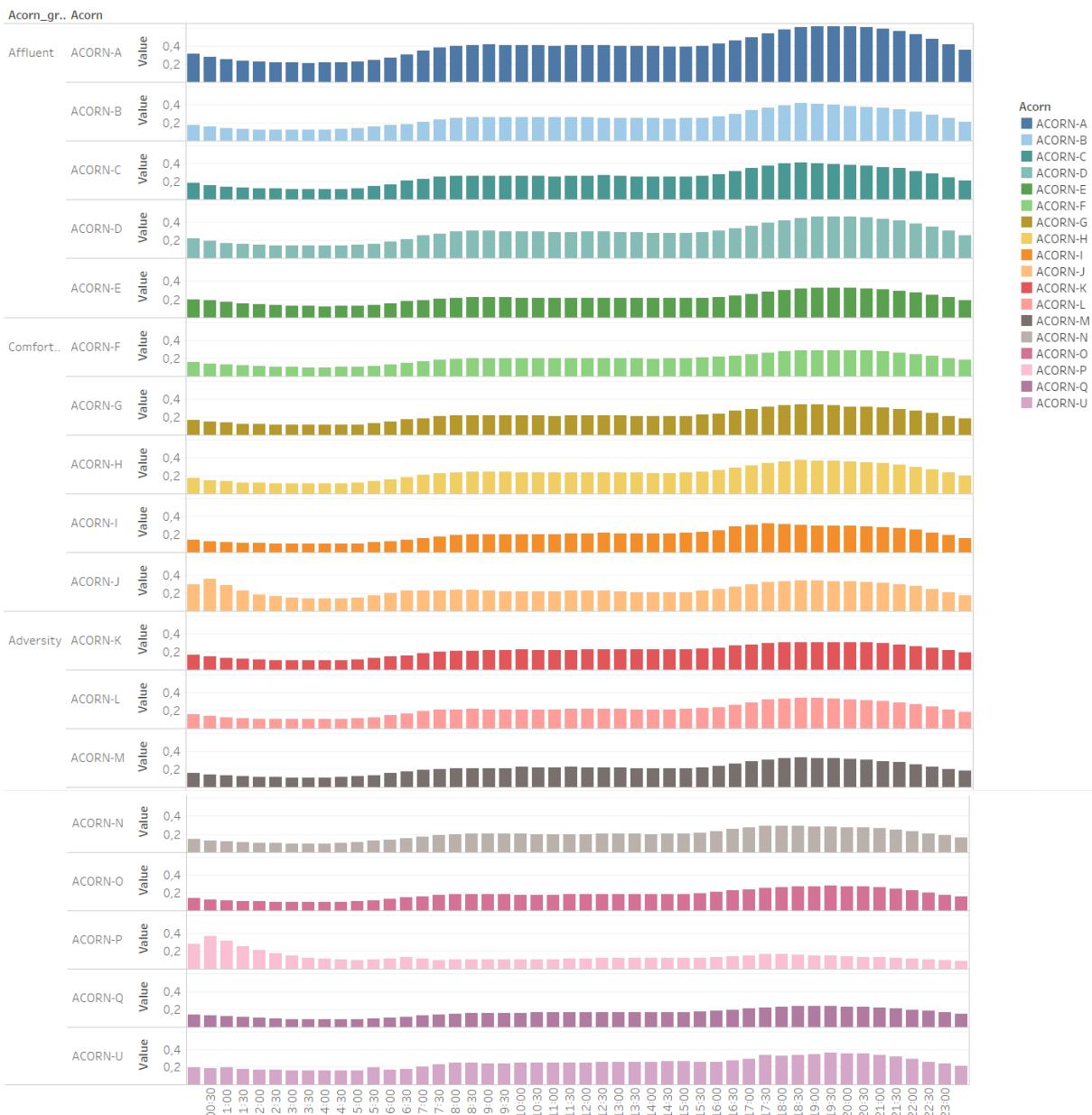
Desglossem els consums en Acorn per veure les diferències entre consumidors i en el següent gràfic veiem que hi ha varis patrons horaris diferenciats.

En primer lloc, hi ha dos Acorn que destaquen per tenir molt consum entre les 0 h i les 2 h de la matinada, els Acorn J i P. Tot i compartir aquest consum inusual en el període vall, veiem que J té un patró normal en pla i punta, en canvi P té un període punta igual que el pla.

El següent Acorn que té un patró inusual és l'Acorn U, que per les seves característiques arribem a la conclusió que és l'Acorn "Unknown" on s'hi engloben els consumidors que no estan en cap de les altres categories.

Un altre Acorn que crida l'atenció és l'Acorn A per tenir un consum molt més elevat que la resta.

Consum mitjà per llar durant el dia segons Acorn



Gràfic 39 Consum mitjà mig-horari per llar durant el dia pintat segons Acorn

Fins ara hem vist els patrons horaris segons varis atributs, i en el següents heatmaps veiem aquests atributs analitzats però el color ens ajuda a identificar si hi ha alguna diferència no vista.

El primer gràfic ens confirma la diferència de consum entre els dies laborals i els cap de setmana.

En el segon gràfic veiem que desembre i gener tenen un patró de consum molt igual, sent els mesos que més es consumeix, sobretot amb un consum més elevat en pla i punta. També veiem que juny juliol i agost segueixen el mateix patró i són els mesos amb menys consum. En aquest gràfic veiem perfectament com el consum comença a

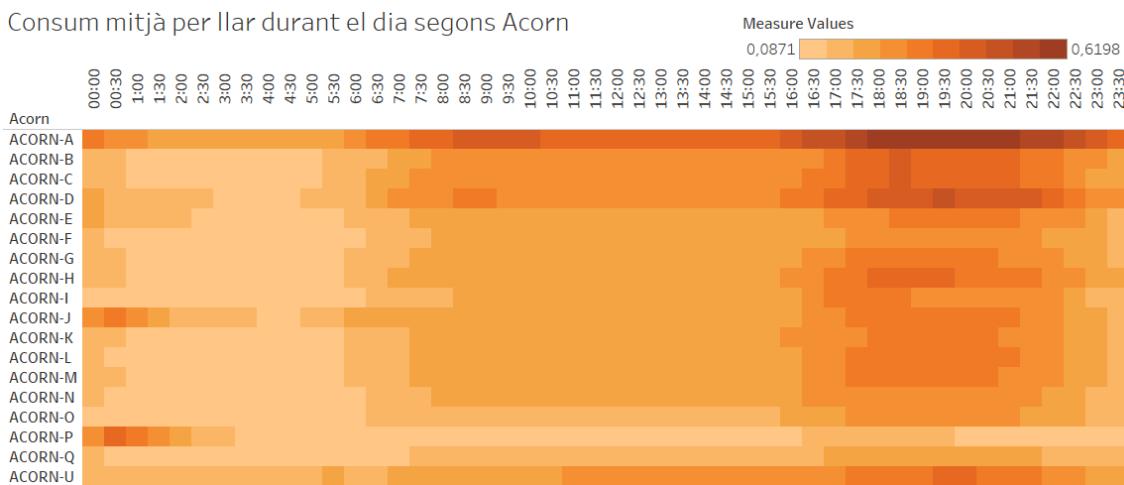
disminuir a partir de gener arribant al mínim a l'agost i llavors tornant a augmentar fins gener.

En el tercer gràfic veiem el consum segons l'estació, que ens confirma que el patró de primavera i estiu és molt igual, tot i que amb una mica més de consum en primavera, però també veiem que el consum de tardor i hivern és força diferent.

En l'últim gràfic del consum segons l'Acorn ens confirma que J i P tenen un patró totalment diferent a la resta i que U també té un consum diferenciat. També veiem que A té un consum més elevat i que comparteix un patró molt semblant amb D. Els Acorns B i C també tenen un consum una mica més elevat que la resta.

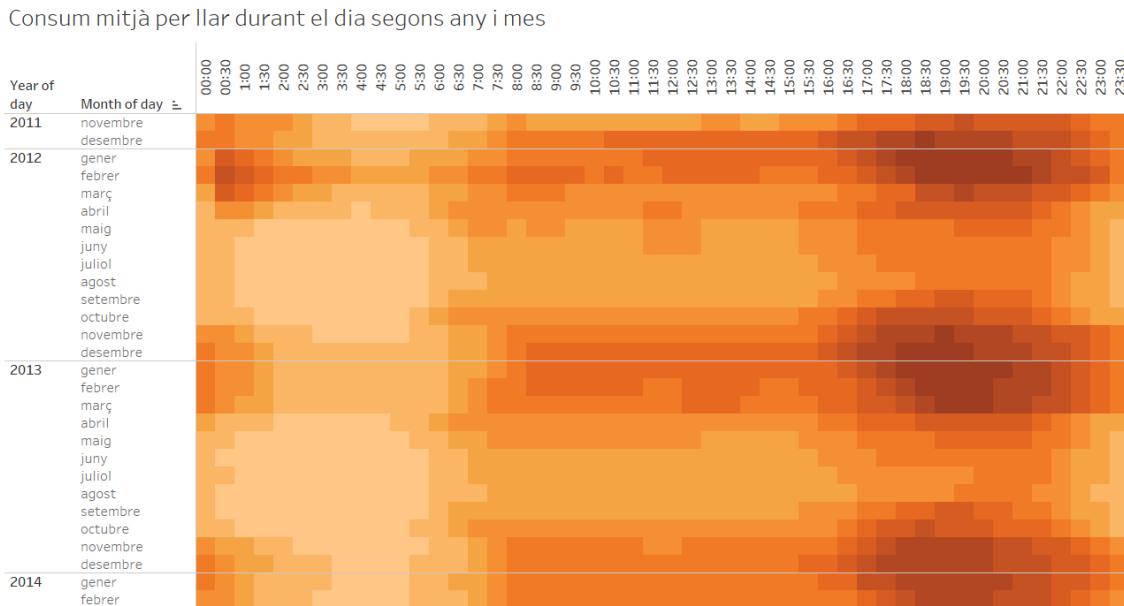
També veiem que de E a N tenen un patró igual (excluint J) i que O i Q tenen un patró equiparable.





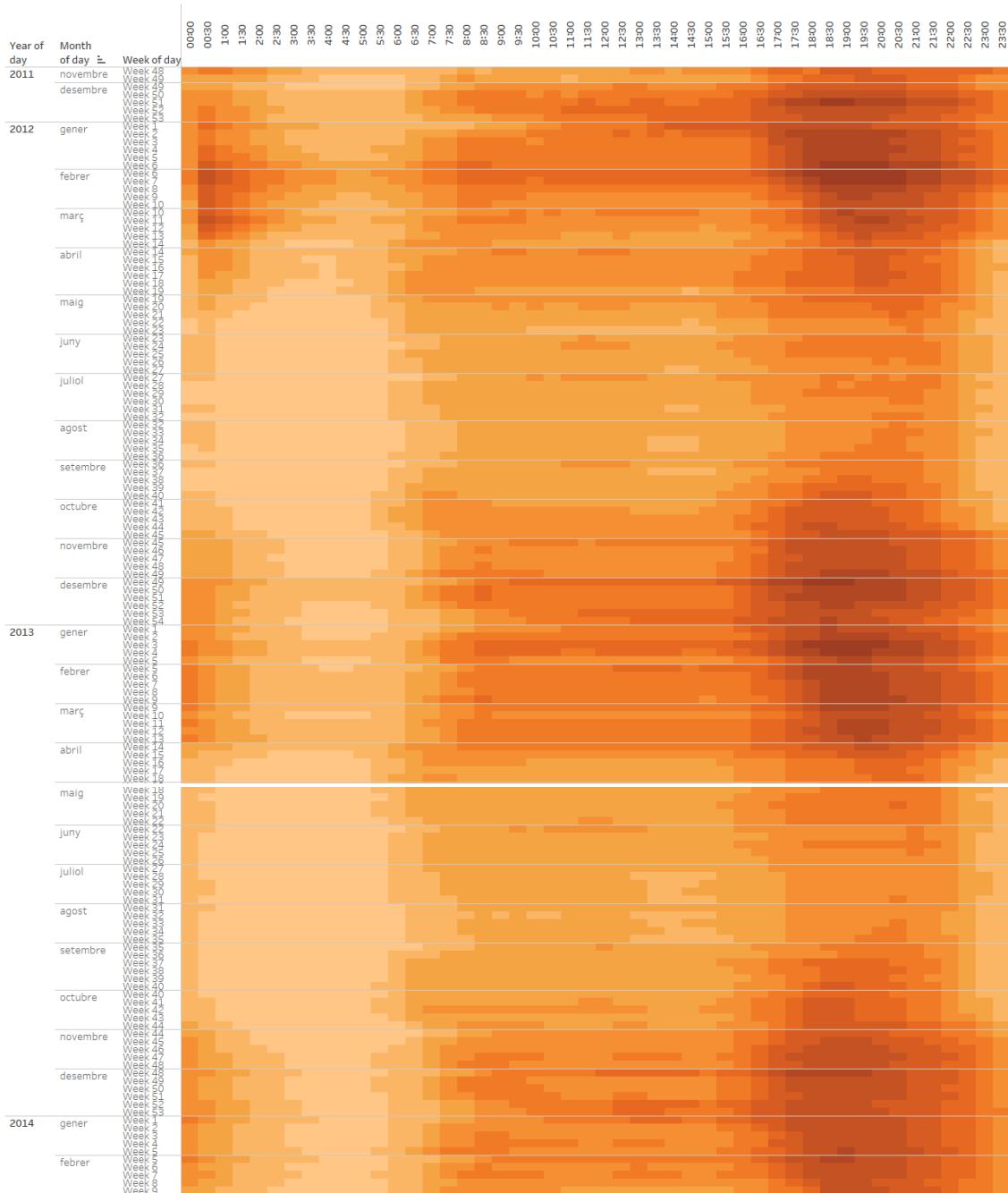
Gràfic 40 Consum mitjà mig-horari per llar al llarg del dia segons dia de la setmana, mes, estació i Acorn.

Per finalitzar amb l'anàlisi, mirem l'evolució temporal dels patrons horaris, veient que cada any es repeteix el mateix patró.



Gràfic 41 Consum mitjà mig-horari per llar durant el dia pintat al llarg dels mesos de la mostra

Consum mitjà per llar durant el dia segons any , mes i setmana

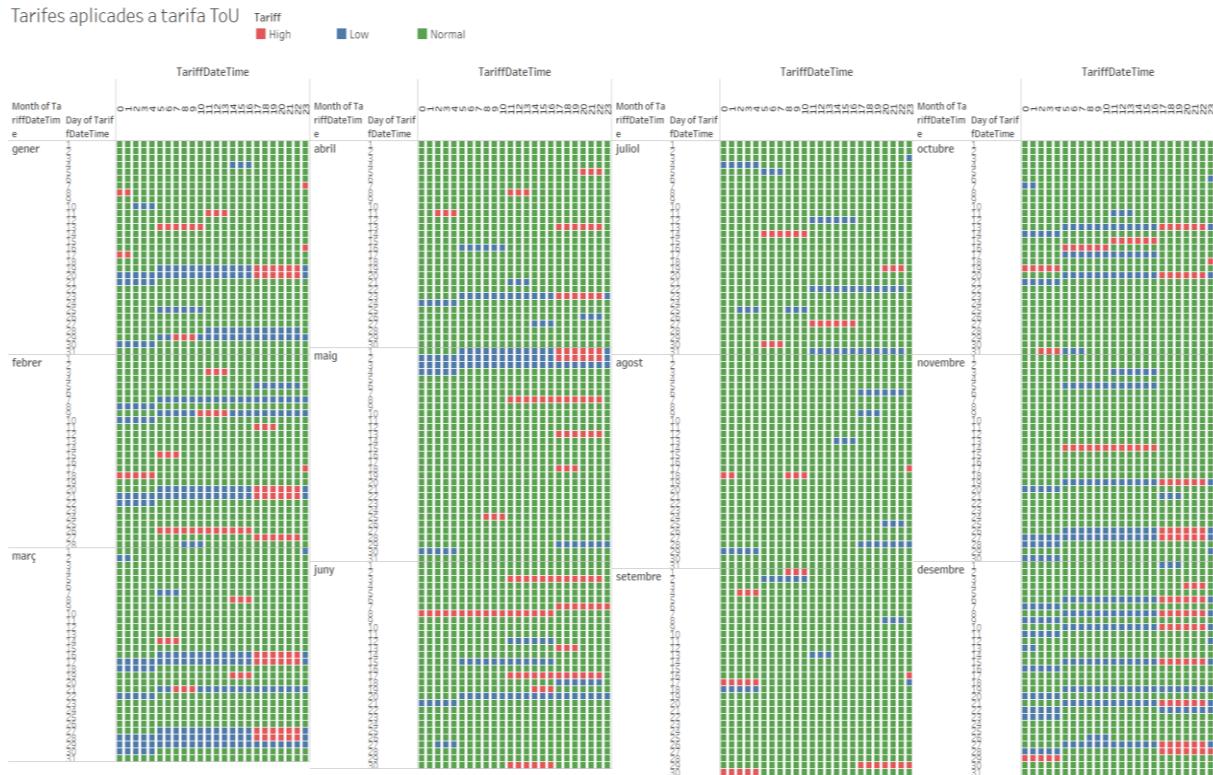


Gràfic 42 Consum mitjà mig-horari per llar durant el dia pintat al llarg de les setmanes de la mostra

Un cop vistes les diferències de patrons segons varis atributs, ens sorprèn que no hi hagi diferència entre el consum de les tarifes Std i ToU, de manera que comprovem les tarifes que s'han aplicat a ToU.

En el gràfic veiem que la tarifa cara (“High”, en vermell) no sempre s’ha aplicat en els períodes de més consum (punta) i també veiem que la tarifa barata (“Low”, en blau) no sempre ha estat aplicada en els períodes de menys consum fins i tot sent aplicada en

períodes de punta, de manera que no ha ajudat a estimular el consum en períodes vall ni a reduir-lo en punta.



Gràfic 43 Tarifes Time of Use aplicades al llarg del 2013 per dia i hora

4.4 Regles d'associació per l'anàlisi de dependències

Utilitzant tant sols variables qualitatives com season, stdorToU, Acorn, icon, month, weekday, tempMaxFactor i sunsetTimeHour buscarem quina relació tenen amb l'energia. Després de fer alguns models, al ser moltes variables les que expliquen el model i al haver-hi tants tipus de clients, estació ... per explicar el consum, els resultats no són concluents.

Per exemple utilitzant season, Acorn, month, weekday, tempMaxFactor i sunsetTimeHour juntament amb el consum factoritzat, ens trobem que la única regla que obtenim amb una confiança de més del 80% és que quan la temperatura és alta, l'estació és estiu, amb un suport del 14%.

Si hi afegim la variable stdorToU, llavors obtenim 9 regles amb una confiança de més del 80%, on hi ha regles que no tenen sentit com que si la posta de sol és a les 18, llavors la tarifa és Std i d'altres que són més interessants com si el consum d'energia és normal, llavors té tarifa Std, però tant sols amb un suport del 24%.

També hi hem afegit la variable icon i la variable isholiday, obtenint 172 regles amb una confiança de més del 80%. Al examinar-les, les més interessants són:

[31]{tempMaxFactor=Normal,energy_sum_factor=Baix} =>	{isHoliday=FALSE}	0.1339433 0.9808036 1.0038169
[32]{stdorToU=Std,tempMaxFactor=Normal,energy_sum_factor=Baix} =>	{isHoliday=FALSE}	0.1058592 0.9807541 1.0037662
[45]{energy_sum_factor=Baix} => {isHoliday=FALSE}		0.3146842 0.9787370 1.0017018
[46]{stdorToU=Std,energy_sum_factor=Baix} => {isHoliday=FALSE}		0.2485483 0.9787345 1.0016993

[47]{stdorToU=Std,energy_sum_factor=Normal} =>	{isHoliday=FALSE}	0.2426766	0.9783781	1.0013345
[48]{energy_sum_factor=Normal} =>	{isHoliday=FALSE}	0.3033083	0.9783374	1.0012928
[66]{stdorToU=Std,icon=partly-cloudy-day,energy_sum_factor=Normal} =>	{isHoliday=FALSE}	0.1758421	0.9754311	0.9983184
[67]{icon=partly-cloudy-day,energy_sum_factor=Normal} =>	{isHoliday=FALSE}	0.2194346	0.9754295	0.9983167
[70]{stdorToU=Std,energy_sum_factor=Alt} =>	{isHoliday=FALSE}	0.1423659	0.9746840	0.9975537
[72]{energy_sum_factor=Alt} =>	{isHoliday=FALSE}	0.1725305	0.9746190	0.9974872
[75]{energy_sum_factor=Molt baix} =>	{isHoliday=FALSE}	0.1865379	0.9745045	0.9973700
[76]{stdorToU=Std,energy_sum_factor=Molt baix} =>	{isHoliday=FALSE}	0.1469672	0.9744586	0.9973230
[80]{icon=partly-cloudy-day,energy_sum_factor=Alt} =>	{isHoliday=FALSE}	0.1184958	0.9727763	0.9956012
[82]{stdorToU=Std,icon=partly-cloudy-day,energy_sum_factor=Molt baix} =>	{isHoliday=FALSE}	0.1104688	0.9725764	0.9953967
[83]{tempMaxFactor=Baixa,energy_sum_factor=Baixa} =>	{isHoliday=FALSE}	0.1102163	0.9717522	0.9945532
[153]{icon=partly-cloudy-day,energy_sum_factor=Alt} =>	{stdorToU=Std}	0.1005791	0.8256916	1.0335554
[154]{isHoliday=FALSE,energy_sum_factor=Alt} =>	{stdorToU=Std}	0.1423659	0.8251636	1.0328945
[155]{energy_sum_factor=Alt} =>	{stdorToU=Std}	0.1460636	0.8251086	1.0328256
[156]{isHoliday=FALSE,icon=partly-cloudy-day,energy_sum_factor=Normal} =>	{stdorToU=Std}	0.1758421	0.8013416	1.0030755
[157]{icon=partly-cloudy-day,energy_sum_factor=Normal} =>	{stdorToU=Std}	0.1802712	0.8013403	1.0030738
[170]{isHoliday=FALSE,energy_sum_factor=Normal} =>	{stdorToU=Std}	0.2426766	0.8000986	1.0015195
[171]{energy_sum_factor=Normal} =>	{stdorToU=Std}	0.2480397	0.8000653	1.0014778

On veiem que una de les regles que té més suport (regla 45, 31% de suport) és que si l'energia és baixa és que no és festiu. També amb un suport del 30% la regla 48 ens diu que si el consum d'energia és normal, no és festiu.

4.5 Variables que afecten al consum. Resum resultats.

Per a fer la prediccó de sèries temporals, si utilitzéssim ARIMA, abans hauríem de comprovar que hi hagi estacionaritat (stationarity), és a dir, que hi hagi una mitjana i variància constant de manera que no hi hagi tendències ni estacionalitat (seasonality).

En el nostre cas veiem que hi ha una tendència d'augment del consum, varis efectes estacionals com la periodicitat diària (cada dia es repeteix el mateix patró), setmanal i anual .

Com que no utilitzarem ARIMA i utilitzarem LSTM, no ens cal tocar la estacionaritat ni fer diferències ni logaritmes a les dades.

En resum, hem vist:

En l'anàlisi estadístic hem vist que teníem el consum diari de 5556 llars diferents del 23-11-2011 al 28-02-2014. Aquestes llars tenen característiques com el tipus de llar (Acorn), el tipus de llar agrupat (Acorn_grouped) i la tarifa. Per cada llar i dia tenim les característiques del dia, tant meteorològiques com de festivitat i d'estació.

Sabem que les llars tenen un consum mitjà de 10,12 kWh al dia, podent arribar a consumir fins 332kWh al dia. La temperatura màxima mitjana del període ha estat 13.66 graus i la minima mitjana de 7.41 graus.

En quant a la correlació, hem vist que hi ha molta relació entre temperatura i consum. Algunes variables correlacionades amb el consum són: dewPoint, uvindex, sunsetTimeHour i sunriseTimeHour on la relació és inversa menys amb sunriseTimeHour. Totes aquestes variables menys sunsetTimeHour estan molt correlacionades amb la temperatura. Altres variables relacionades amb el consum però no tant intens són: humidity, cloudcover i visibility.

Tot i sortir que no hi ha correlació entre consum i estació, tant en el contrast d'hipòtesis com en l'anàlisi gràfic veiem com la estació afecta de diferents maneres al consum.

En contrasts d'hipòtesis, hem vist :

- El consum segons la tarifa és diferent (tarifa Std consumeix més que ToU)
- El consum segons tipus de consumidor és diferent (Acorns I-M-N-J són diferents de Acorns F-E i alhora diferents de tots els altres Acorn individuals)
- El consum segons el dia és diferent (es consumeix més en cap de setmana que en festiu, i es consumeix més en festiu que en laboral i també sabem que dimarts, dimecres i dijous es consumeix el mateix i en l'anàlisi gràfic veiem com el divendres és el dia que es consumeix menys, seguit de dimarts-dimecres-dijous, llavors dilluns seguit de dissabte i el dia que es consumeix més el diumenge)
- El consum entre mesos és diferent (octubre i abril es consumeix el mateix i tots els altres mesos consumeixen diferent)
- El consum entre estació és diferent (totes les estacions tenen un consum significativament diferent entre elles, si les ordenem de més a menys consum tindríem: hivern, tardor, primavera, estiu)
- El consum segons la temperatura és diferent

En l'anàlisi gràfic diari hem vist com tenim molts més consumidors amb tarifa Std que amb tarifa ToU, i que la majoria de llars forma part de l'Acorn E (Affluent), Q (Adversity) o F (Comfortable). També hem vist que el nombre de llars varia molt al llarg del temps, de manera que s'utilitzarà la mitjana del consum per llar i no el consum total de totes les llars. Hem vist com l'evolució del consum segueix un patró anual, on a l'hivern té el pic més alt i a l'estiu el pic més baix de consum. També hem desglossat el consum per Acorn i hem vist com es podien classificar en tres grups (Affluent, Comfortable i Affluent). També hem vist com el mes i l'estació afecten de manera diferent a cada tipus de consumidor (gràfics 24 i 25). També hem analitzat com afecta al consum diferents variables respecte l'hora del sol, veient que per exemple la tardor i la primavera afecten molt al consum segons l'hora de posta de sol.

També hem analitzat com afecten diferents variables meteorològiques al consum segons la estació, tarifa o tipus de consumidor (Acorn agrupat) on hem vist per exemple en el gràfic 31, que el consum es veu molt afectat per la temperatura apparent (tant la màxima com la mínima), sobretot a la primavera i la tardor (ja que tenen un pendent més fort). També en el mateix gràfic hem vist com la visibilitat afecta de diferent manera al consum de diferents estacions, on totes les estacions a més visibilitat tenen menys consum, menys la tardor, on a més visibilitat, més consum. En el mateix gràfic també veiem que la humitat afecta de manera diferent al consum de l'estació de hivern que les altres estacions i que la pressió afecta de manera diferent a l'hivern i a la tardor.

En l'anàlisi gràfic horari hem vist que hi ha un patró horari clarament definit, on de 8.00 h a 15.00 h es consumeix el mateix (als voltants de 0.21kWh de mitjana cada mitja hora), llavors va pujant el consum fins arribar a les 19.00 h al pic de consum de més de 0.31kWh., i llavors comença a baixar gradualment fins arribar al pic més baix a les 4.00 h (als voltants de 0.11kWh) (Veure gràfic 34) . També hem vist que aquest patró horari

canvia segons l'estació, on tardor i hivern tenen uns períodes punta i vall més pronunciats que primavera i estiu, a part de la diferència de consum entre estacions.

També hem vist que els dies laborals tenen un patró horari i que els festius tenen un patró una mica diferent, amb més consum en període pla. Si mirem el consum horari segons Acorn agrupat, veiem que els affluent tenen una vall i una punta molt alta en comparació als altres grups, i Adversity és el que té un consum més “planer”, ja que no hi ha tanta diferència entre punta, pla i vall.

També hem analitzat visualment els horaris d'aplicació de la tarifa ToU, on hem vist les hores on s'ha aplicat la tarifa High, la Low i la normal i que aquestes no tenien molt sentit amb els patrons horaris vistos i per tant no poden tenir influència en la reducció de consum en punta i augment de consum en vall com pensàvem.

Per concloure aquest apartat, en l'anàlisi hem vist que variables que afectaven al consum diari son:

- Propi consum
- Mes i estació de l'any
- Laboralitat (dia de la setmana)
- Nombre de consumidors
- Tipus de consumidor
- Meteorologia: Temperatura apparent màxima i mínima, humitat, pressió, cloudCover...
- Hora de posta de sol

5. Predicció

Utilitzarem varis models per a fer varies prediccions combinant diferents variables en cada model.

No podem fer predicció anual degut a les poques dades. Considerem predicció a curt les horàries i diàries i la predicció a mig termini és setmanal.

El codi dels models el trobem als annexes a la carpeta de Python, anomenats tal i com mostrem les taules resum (cada arxiu és un model). Aquests arxius estan estructurats de la següent manera:

Primer es carrega el dataset a utilitzar i es seleccionen les variables a utilitzar. Llavors es fa shift a les dades (s'agafen els valors dels dies anteriors) i s'esborren els valors nuls. Després normalitzem tots els valors de les variables de 0 a 1 i seleccionem la X(variables per predir) i la y(valor a predir). Llavors passem les dades a format “mostres, dies anteriors per la predicció, variables” i definim el model de xarxa neuronal amb LSTM. Després s'executa el model amb test i train walk forward i es calculen els errors. Llavors fem un gràfic amb les prediccions, però aquestes estan amb valors normalitzats, així que les tornem als valors originals i fem un gràfic per comparar el valor predit amb el valor real. Finalment calculem l'error percentual i fem un gràfic.

Els datasets utilitzats es troben a l'annex “dades finals” i són extrets dels fitxers de codi de preparació de dades de R i Estacionalitat de R.

5.1 Model a utilitzar

Pels models utilitzarem una xarxa neuronal LSTM (long short-term memory).

LSTM és una xarxa neuronal artificial recurrent que pot processar seqüències de dades, com són les sèries temporals.

Podem desenvolupar 2 tipus de models:

- Univariate time series, on només 1 variable és observada en cada moment, en el nostre cas el consum per cada dia.
- Multivariate time series, on 2 o més variables són observades al mateix temps, com seria el consum i la temperatura diaris.

I podem decidir el nombre de temps a predir:

- One-step forecast, on predim el següent temps ($t+1$), per exemple el consum del dia següent.
- Multi-step forecast, on predim 2 o més temps futurs.

No ens centrarem en els paràmetres del model sinó en les variables que hi introduïm, de manera que farem varies prediccions introduint diferents variables de diferents temps passats i utilitzant diferents mètodes de test.

En quant a les dades d'entrenament i de test, utilitzarem el mètode walk forward, on per a cada predicción, afegim el valor real de la última predicción feta. És a dir, per exemple en el cas de dades diàries, utilitzarem els consums del 15-03-2012 al 15-03-2013 per entrenament, per poder predir el consum del 16-03-2013. Llavors per a la predicción del dia següent(17-03-2013), utilitzem de 16-03-2012 a 16-03-2013 (sempre s'utilitza un any per a train).

També utilitzarem un walkforward que a mesura que es vagi avançant en el dia predict, augmenti el train. Per exemple, per a predir el consum de 16-03-2013, utilitzarà de 15-03-2013 a 15-03-2013 i per predir 17-03-2013 utilitzarà de 15-03-2013 a 16-03-2013, de manera que la data d'inici de train sempre és 15-03-2012.

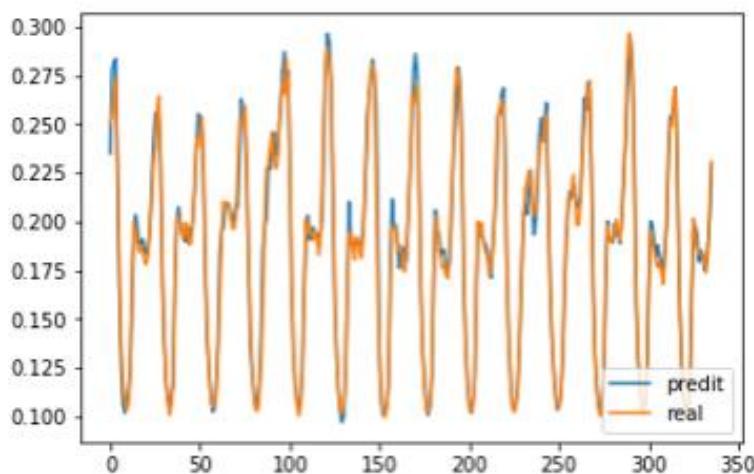
Els models els evaluarem a partir de RMSE.

5.2 Predicció agregat horari

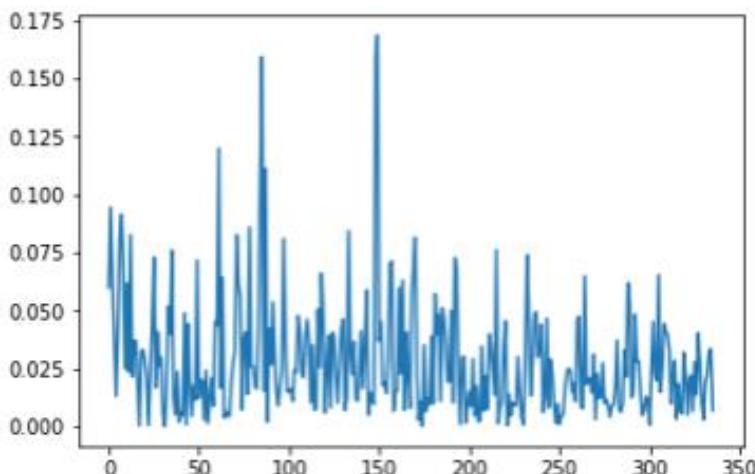
Amb el fitxer horari hem tingut problemes per tractar, analitzar i modelar. Provant amb R no el podíem tractar, i utilitzant Python podem tractar-lo, analitzar-lo i modelar-lo. Al fer un model amb totes les dades disponibles, la execució és molt farregosa i tarda més de 72 hores a executar (l'hem parat ja que sinó no podíem avançar).

Finalment fem un model de mostra per veure si amb menys dades també aconseguim un bon model.

Utilitzem només les dades de consum passades, en concret 48 hores passades, i predirem la següent hora. Utilitzarem walkforward on entrenarem amb 14 dies (336 hores) i farem test en els 14 dies següents. Obtenim un RMSE de 0.012557 i un error del 2.83%.



Gràfic 44 Consum horari real i predict de 14 dies



Gràfic 45 Error percentual de la predicción horària

Degut a les limitacions que suposa treballar amb aquestes dades sense la maquinaria adequada, decidim centrar-nos en les prediccions diàries.

Taula resum models horaris:

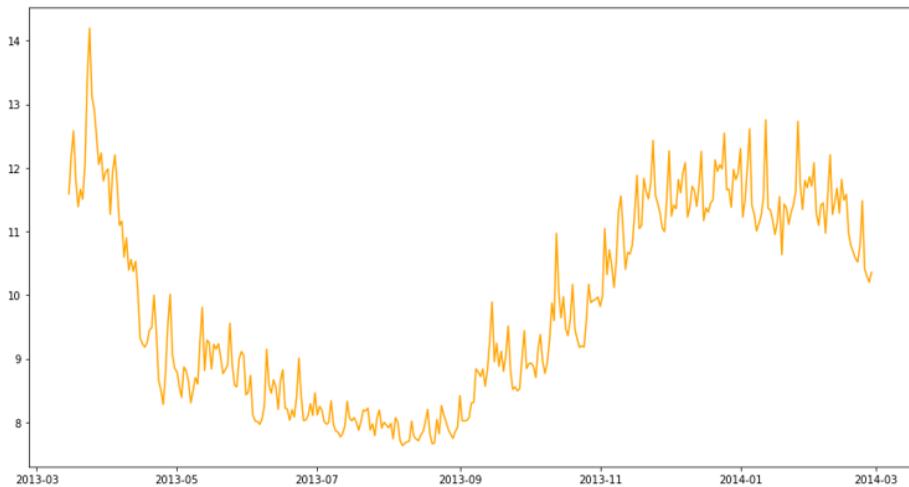
	Dades passades	Train	RMSE
MH	48 hores	Totes les dades	No executa
MH1	48 hores	14 dies	0.01255

5.3 Predicció agregat diari

Com ja hem comentat, començarem utilitzant les dades del 15-03-2012 al 15-03-2013 per a entrenament i a mesura que faci la predicció del dia següent, llavors s'avançarà un dia.

També ens interessa veure el rendiment que té el model al llarg de tot un any i per tant utilitzarem tot un any per a test (sempre amb la tècnica de walk forward).

Seguidament veiem el consum que volem predir:

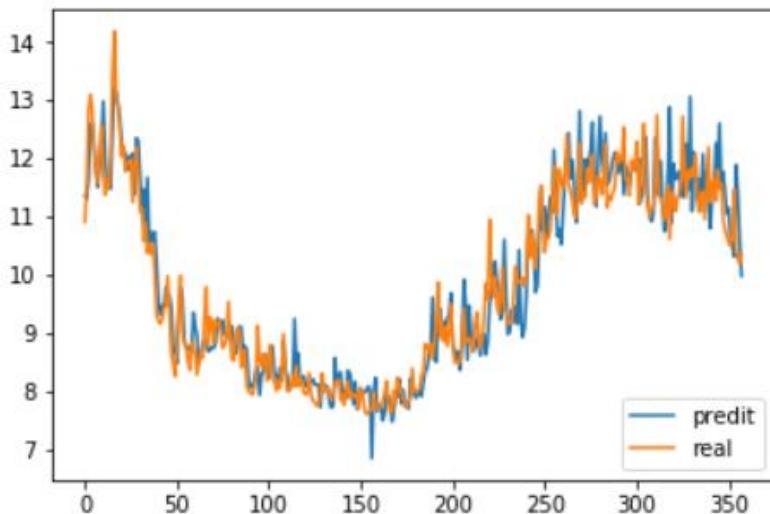


Gràfic 46 Consum mitjà per llar del 15-03-2013 al 27-02-2014

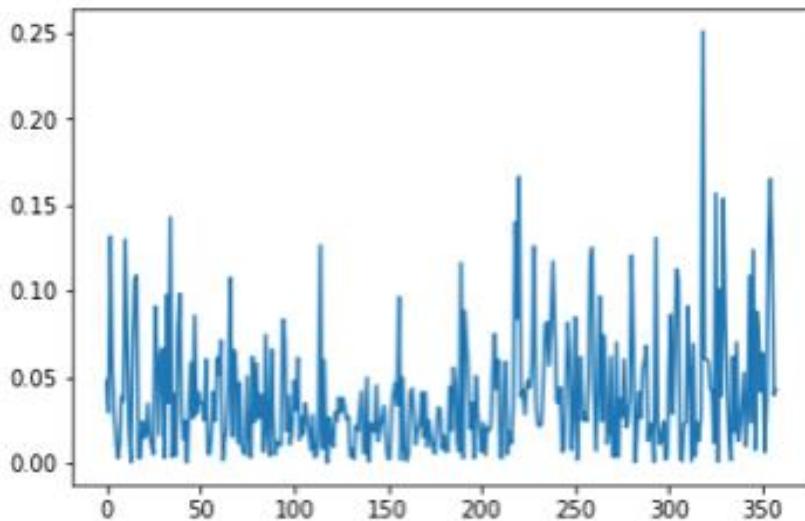
5.3.1 Predicció a un dia

5.3.1.1 Walkforward 365 dies

Comencem realitzant una predicció diària a un dia utilitzant tant sols les dades de consum passades, el que seria el model univariate one-step. Per a fer la predicció del dia següent, utilitzem el consum dels 7 dies anteriors amb els valors normalitzats en una xarxa neuronal LSTM amb 50 neurones a la capa d'entrada amb activació relu i llavors la capa de sortida amb 1 neurona (ja que estem predint 1 valor). Obtenim un error mitjà (RMSE) de 0.0393.



Gràfic 47: Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM univariate utilitzant 7 dies passats one-step amb walk-forward 365 dies.

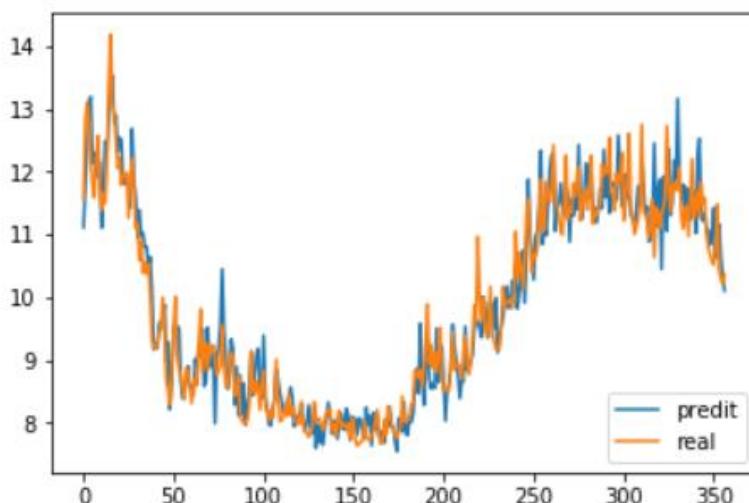


Gràfic 48: Errors model univariate utilitzant 7 dies passats one-step amb walk-forward 365 dies

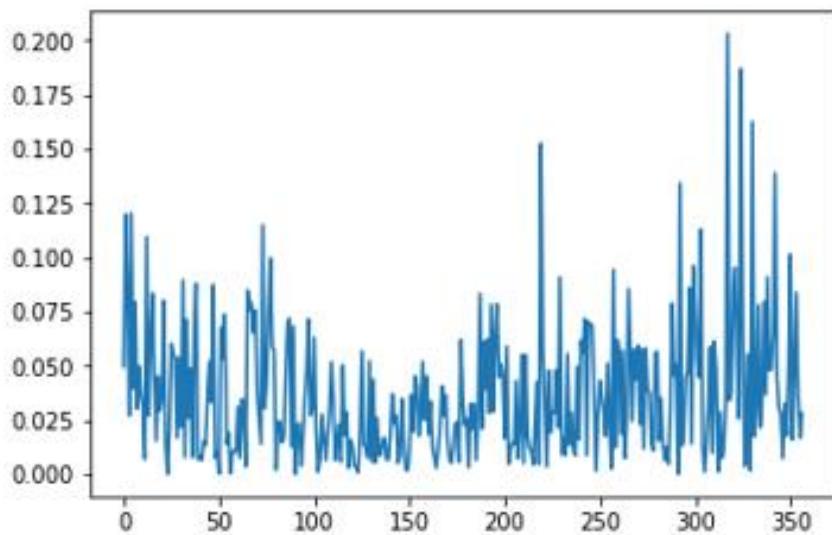
Les dades de consum utilitzades són la mitjana de consum diari per consumidor (suma consum diari tots consumidors respecte nombre consumidors dia) i no la suma del consum de tots els consumidors, ja que durant el temps el nombre de consumidors varia, i si sabem la predicción del consum mitjà, si el multipliquem pel nombre de consumidors, tindrem una predicción del consum total.

Veiem que els primers 50 dies fa una molt bona predicción (el que serien gener i febrer), i a partir d'aquí fins el dia 200 s'acosta força, tot i que hi ha algun pic que no concorda, i a partir de 220, quan ja s'acaba l'estiu, la predicción fa errors més alts.

Creem un nou model introduint la temperatura màxima de dies passats (7 dies anteriors, com el consum) al model per mirar si respon millor i no fa tants pics d'errors, obtenint una petita millora amb tant sols un 0.0368 d'error (RMSE) de mitjana, tot i que al representar-ho, tot i haver millorat la mitjana de l'error general, veiem que continuen havent-hi pics de consum que no haurien de ser-hi:



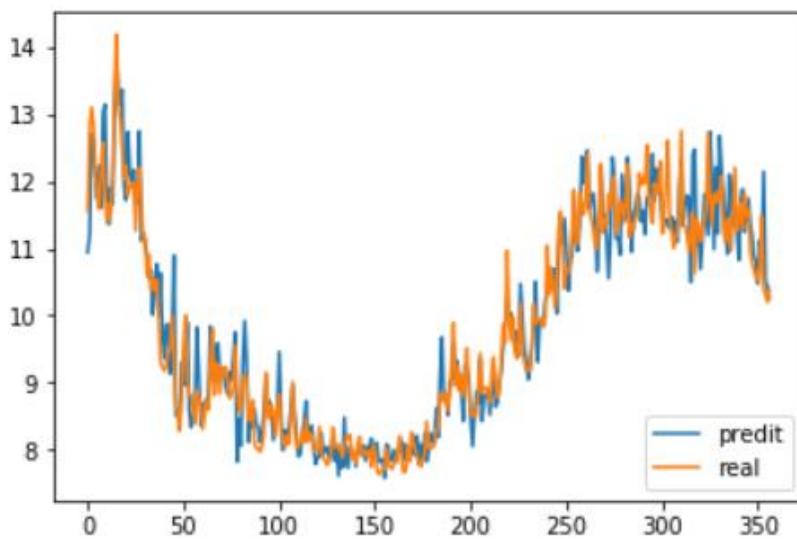
Gràfic 49 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura max) one-step amb walk-forward 365 dies.



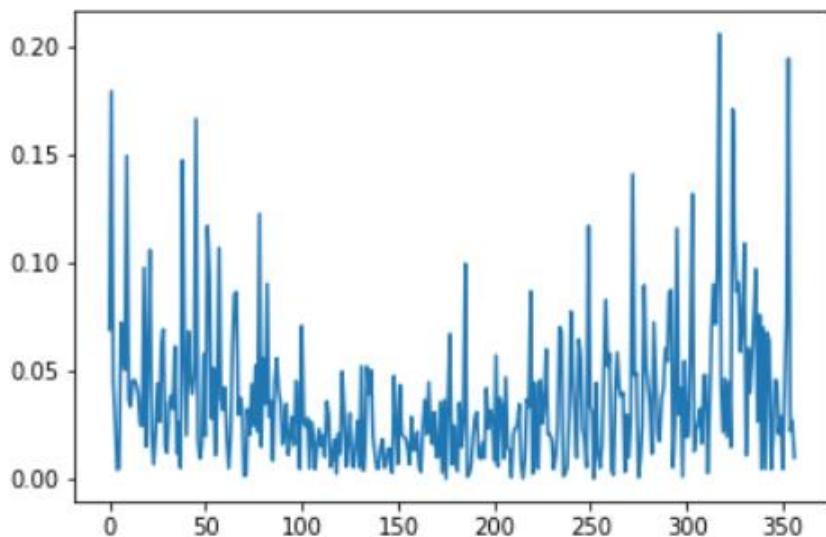
Gràfic 50 Errors model multivariate (temperatura max) one-step amb walk-forward 365 dies

En veure que amb la temperatura màxima ha millorat la predicción global, introduïm una altre variable al model, la humitat.

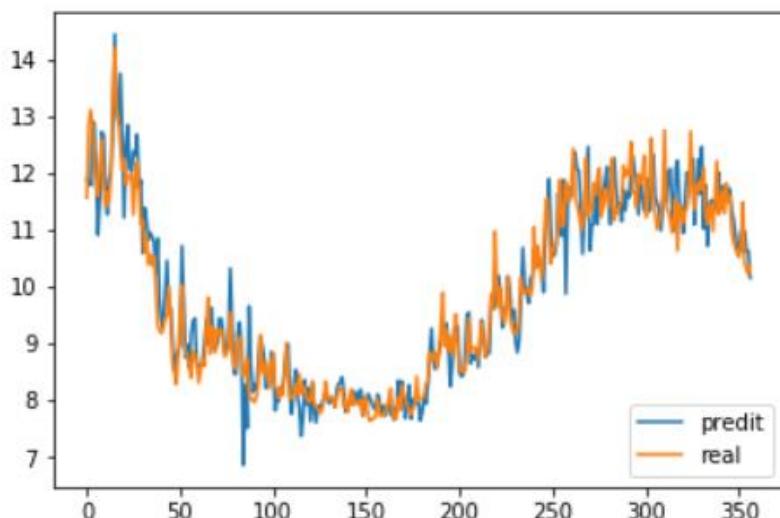
Veiem que es millora una mica la precisió global de la predicción ja que obtenim un error del 0.0363, però només pel fet que els errors més grans no són tant alts.



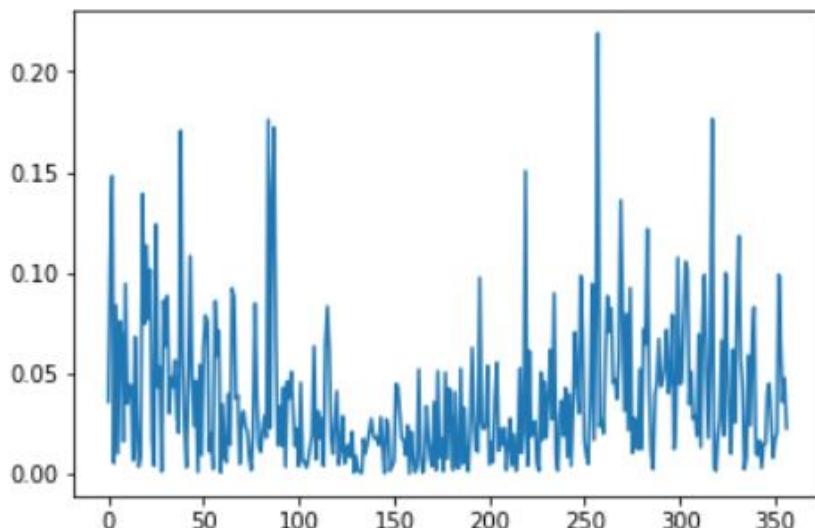
Gràfic 51 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura max i humitat) one-step amb walk-forward 365 dies.



Finalment, creem un nou model introduint també les variables visibility i cloudCover, tot i que en aquest cas no millora la predicción i obtenim un error (RMSE) mitjà de 0.0378.



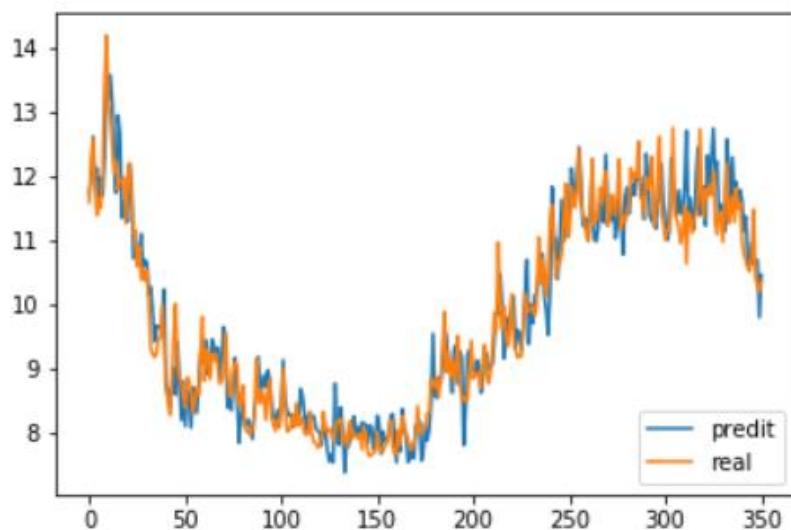
Gràfic 52 Predicccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura max, humitat, visibility i cloudCover) one-step amb walk-forward 365 dies.



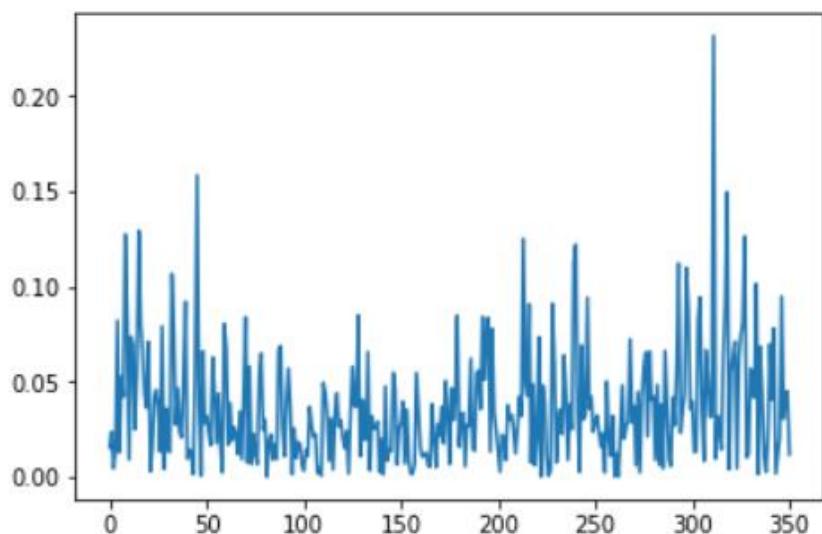
Gràfic 53 RMSE model multivariate (temperatura max, humitat, visibility i cloudCover) one-step amb walk-forward 365 dies.

Veient que la millor predicció l'hem obtingut a partir de la xarxa neuronal multivariate amb temperatura i humitat, provem el mateix model utilitzant 14 dies anteriors en lloc dels 7 dies anteriors per l'entrenament.

Obtenim una petita millora en la predicció, on obtenim un error mitjà de 0.0360.

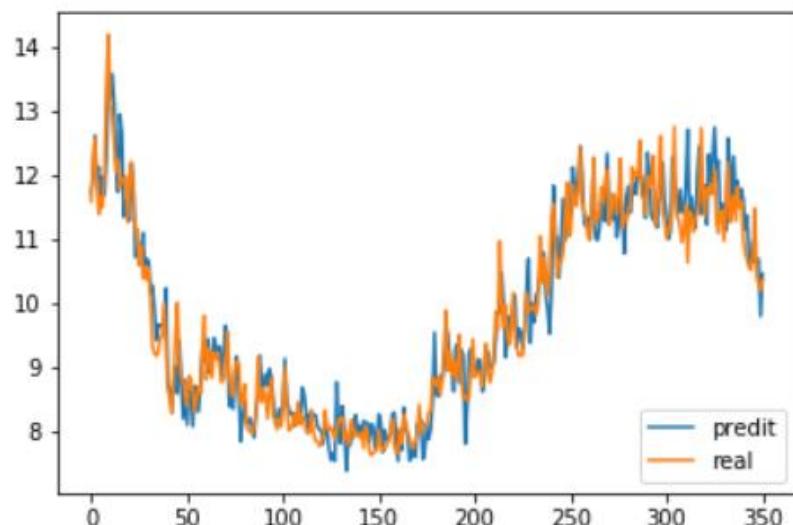


Gràfic 54 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura max i humitat dels 14 dies anteriors) one-step amb walk-forward 365 dies.

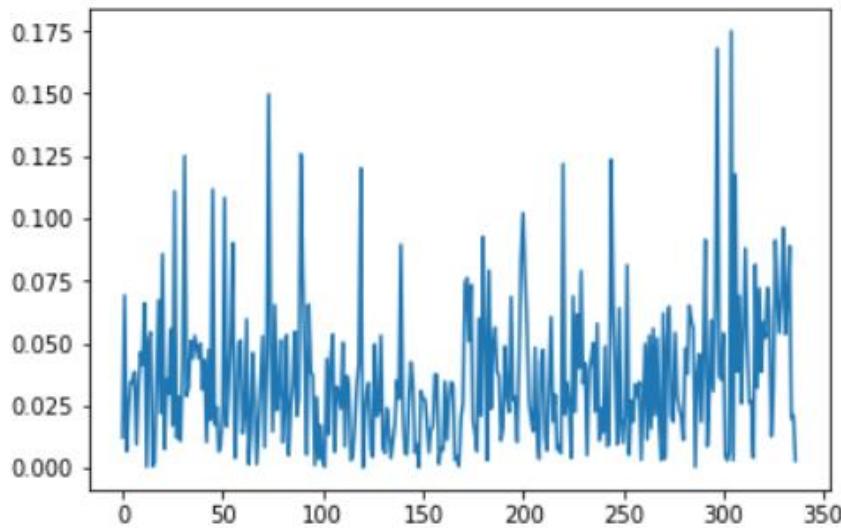


Gràfic 55 RMSE model multivariate (temperatura max i humitat dels 14 dies anteriors) one-step amb walk-forward 365 dies

Al veure que al augmentar els dies de 7 a 14 dies millora la predició, provem utilitzant els valors de 28 dies anteriors, obtenint un error més alt, del 0.0363.

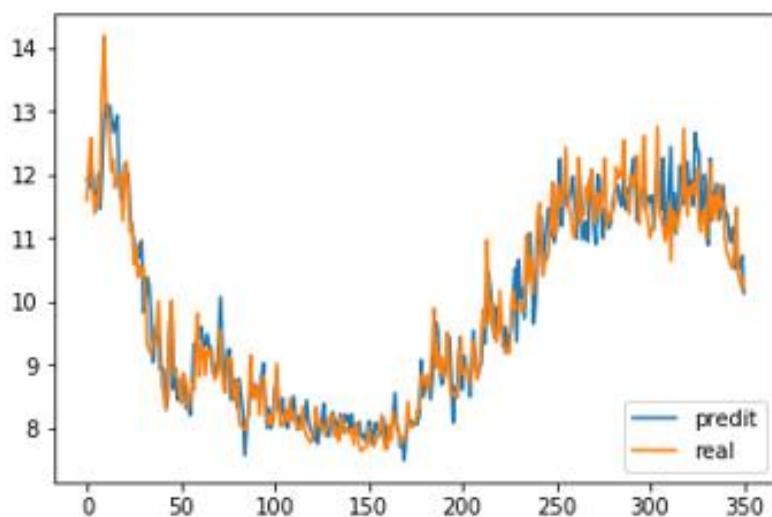


Gràfic 56 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura max i humitat dels 28 dies anteriors) one-step amb walk-forward 365.

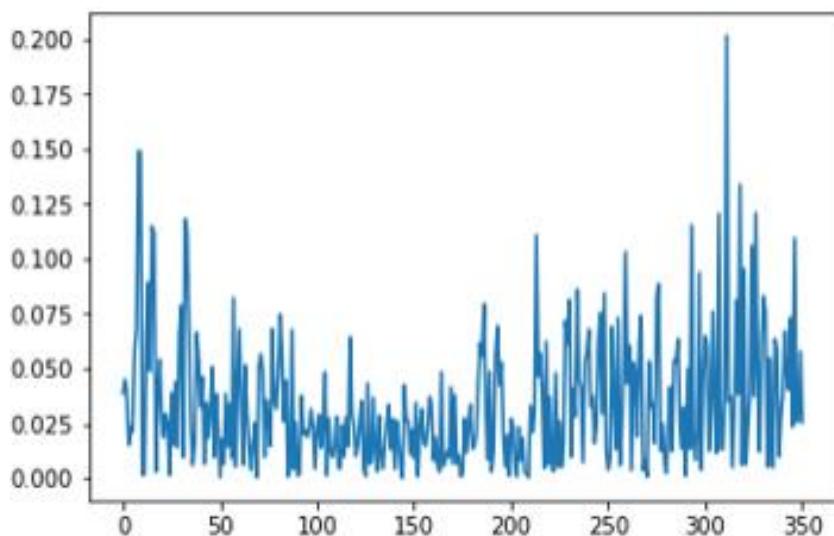


Gràfic 57 RMSE model multivariate (temperatura max i humitat dels 28 dies anteriors) one-step amb walk-forward 365.

A la millor predicció fins ara, el model multivariate consum, temperatura max i humitat amb 14 dies passats, provem d'introduir-hi la temperatura minima i obtenim un error RMSE d'un 0.03393

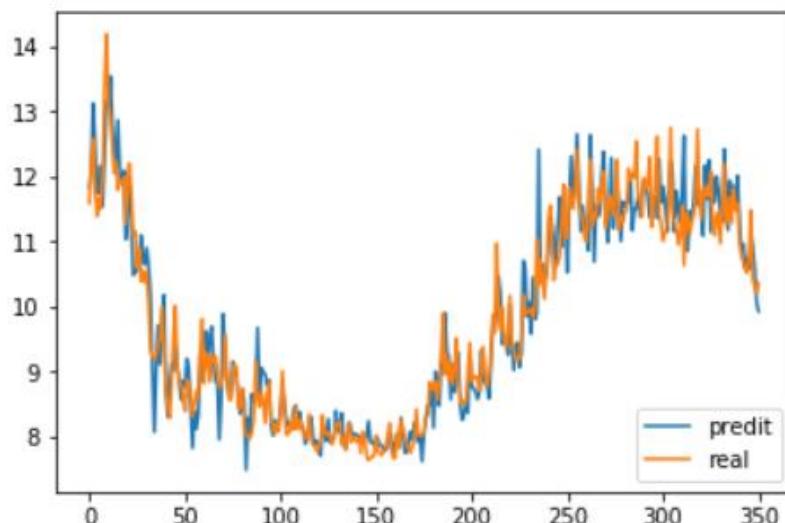


Gràfic 58 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors) one-step amb walk-forward 365



Gràfic 59 RMSE model multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors) one-step amb walk-forward 365

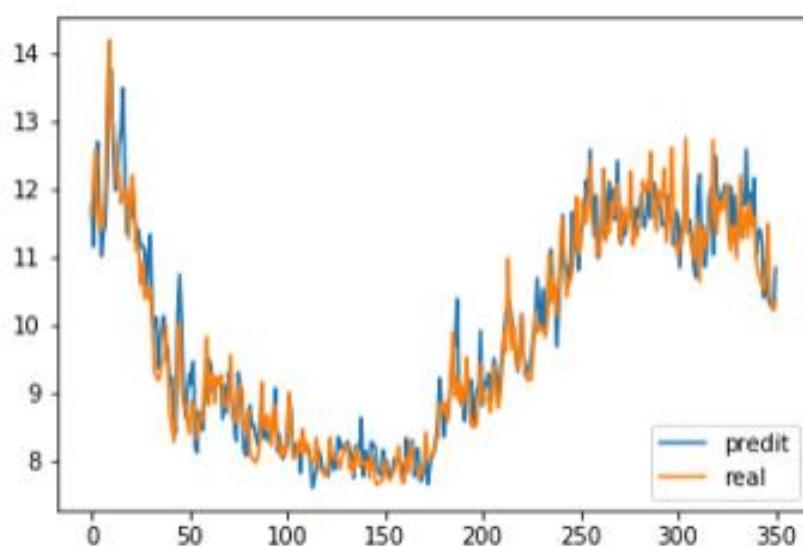
Provem de fer aquest mateix model però sense les dades de la humitat i el resultat és pitjor, obtenint un 0.0359 d'error (RMSE).



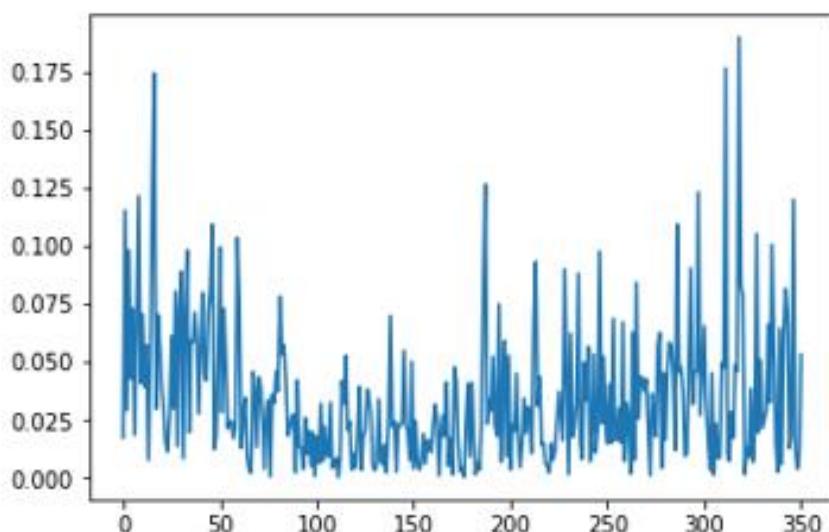
Gràfic 60 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima dels 14 dies anteriors) one-step amb walk-forward 365 dies

Partint del millor model fins ara, (multivariate amb consum, temperatures màxima i minima i humitat 14 dies passats), hi introduïm 5 variables per separat:

- Sunset time hour: model amb consum, temperatures màxima i minima , humitat i hora de posta de sol dels 14 dies passats, on obtenim un error RMSE de 0.03459, és a dir, aquest model és pitjor que el model de partida sense la hora de posta de sol.

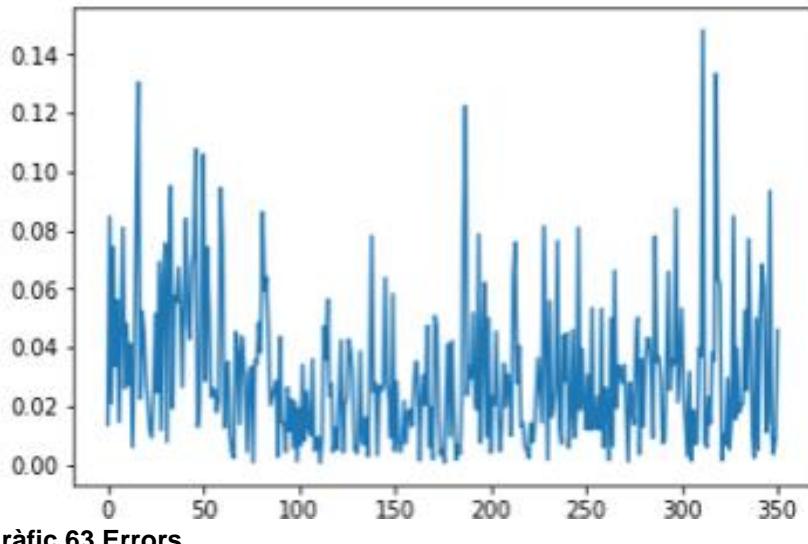


Gràfic 61 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward 365



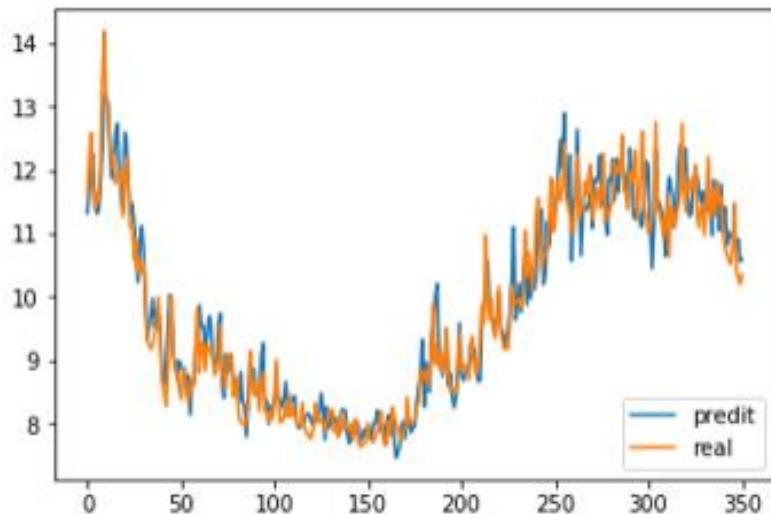
Gràfic 62 RMSE model multivariate (temperatura màxima i mínima, humitat i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward 365

Com en altres models, a on s'adapta millor és a l'estiu. Si ens fixem en el gràfics d'errors percentuals, calculat a partir de la diferència entre el valor real i el predit i això dividit entre el valor real. Veiem que hi ha 4 errors que destaquen entre el 12 i 14% d'error. L'error mitjà és de 3.09%

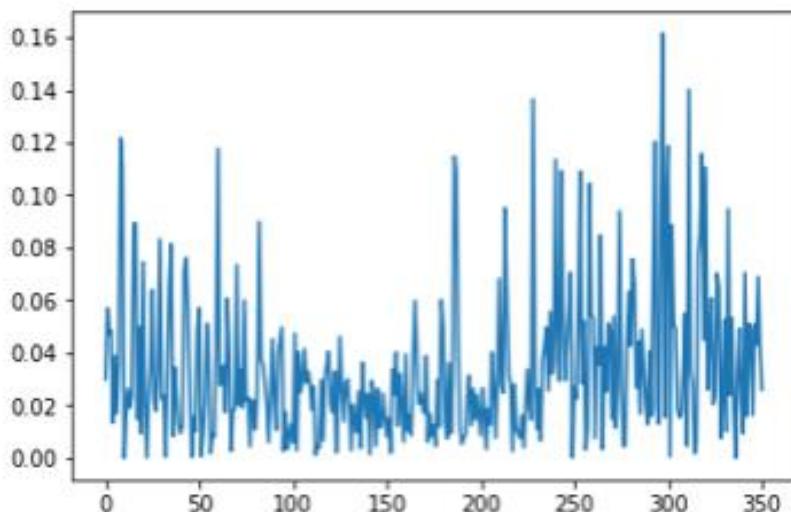


Gràfic 63 Errors

- Pressure: model amb consum, temperatures màxima i minima , humitat i pressió dels 14 dies passats, on obtenim un error RMSE de 0.03389, és a dir, aquest model és una mica millor que el model de partida sense la pressió.

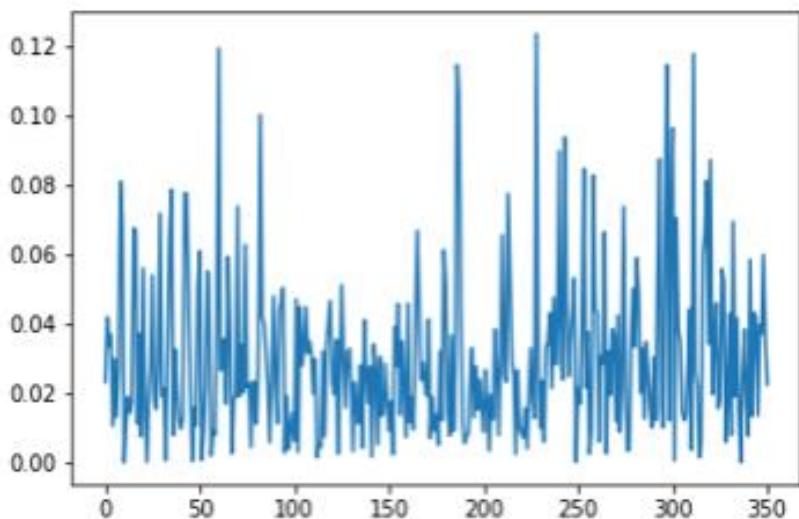


Gràfic 64 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i pressure dels 14 dies anteriors) one-step amb walk-forward 365



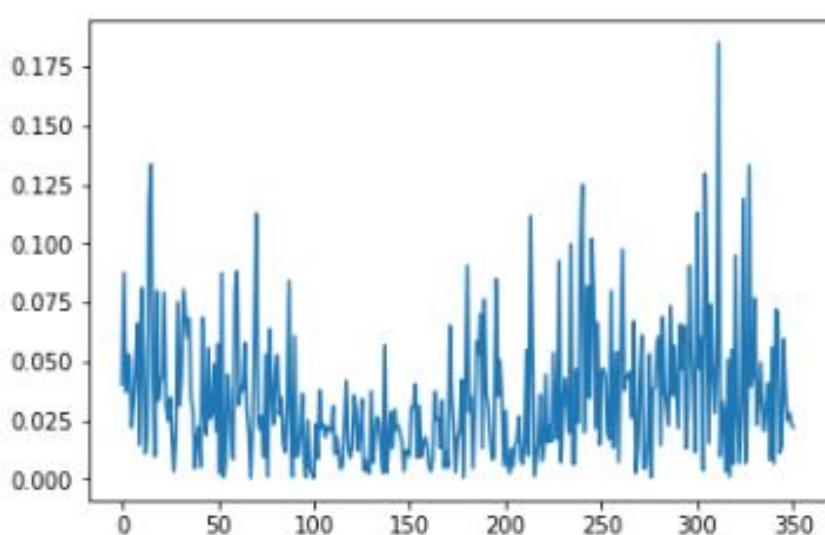
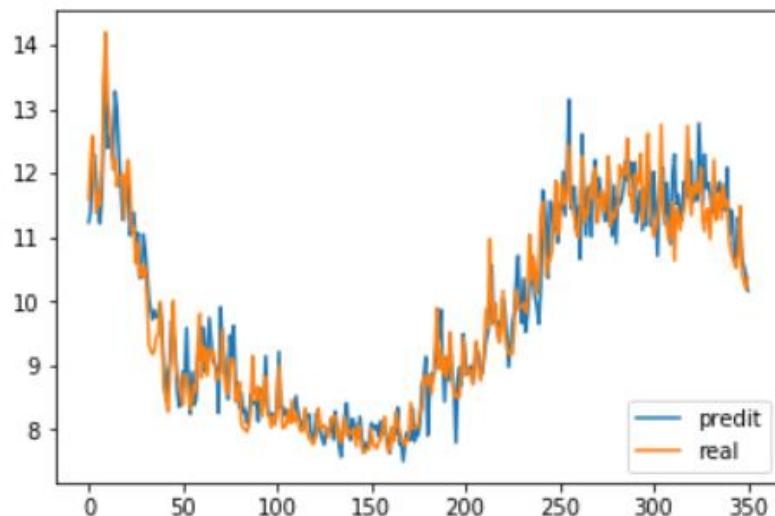
Gràfic 65 RMSE model multivariate (temperatura màxima i mínima, humitat i pressure dels 14 dies anteriors) one-step amb walk-forward 365

L'error percentual mitjà és de 3.02% i els errors màxims estan al voltant del 12%.

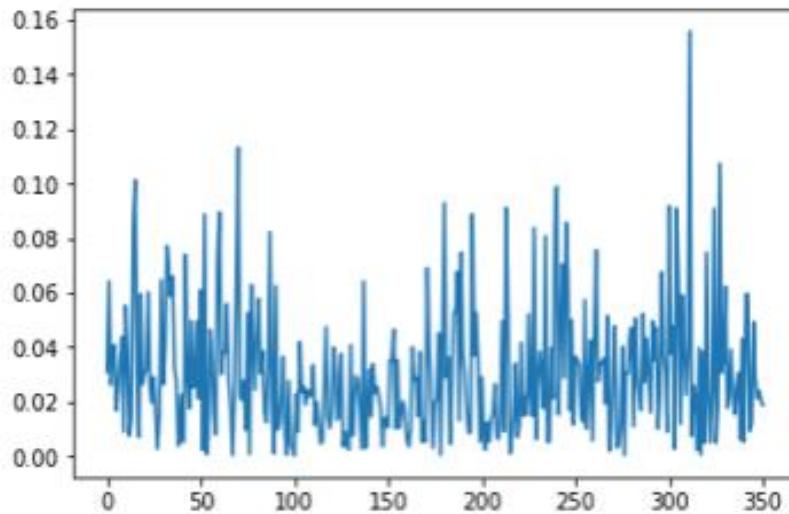


Gràfic 66 Errors model multivariate (temperatura màxima i mínima, humitat i pressure dels 14 dies anteriors) one-step amb walk-forward 365

- DewPoint: model amb consum, temperatures màxima i minima , humitat i punt de rosada dels 14 dies passats, on obtenim un error RMSE de 0.03438, és a dir, aquest model és pitjor que el model de partida sense el punt de rosada.

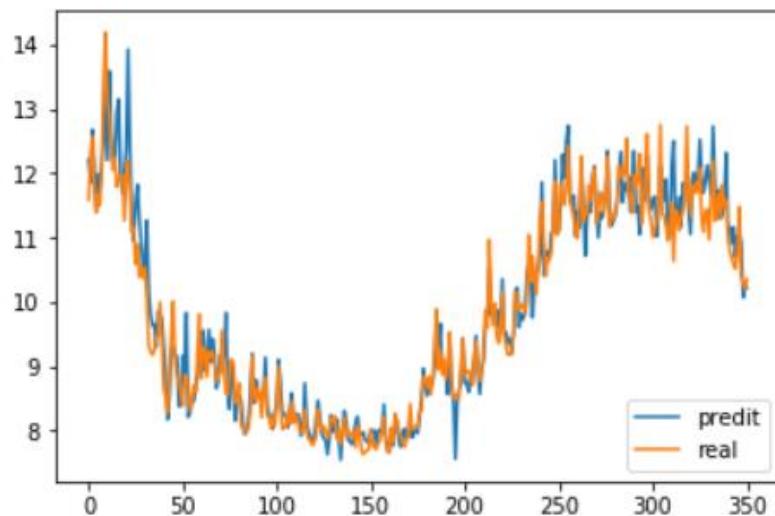


L'error percentual mitjà és de 3.06% i l'error més alt (només un) està al voltant del 16%, els altres no superen el 12%.

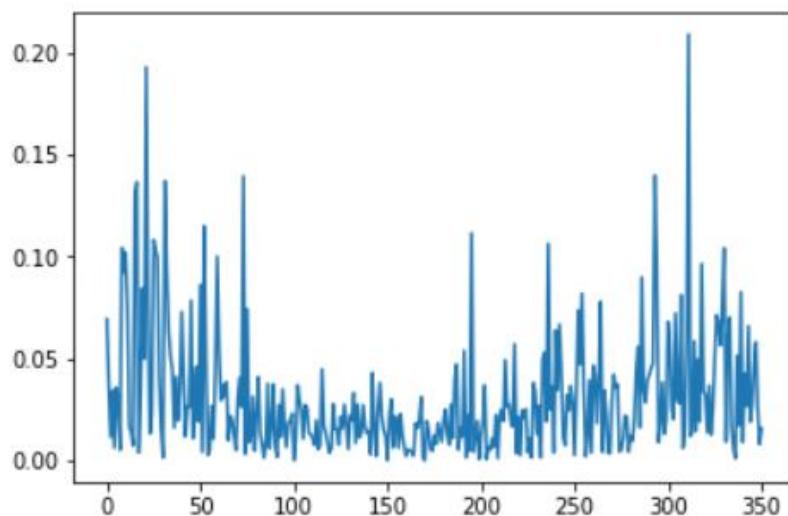


Gràfic 69 Error model multivariate (temperatura màxima i mínima, humitat i dewPoint dels 14 dies anteriors) one-step amb walk-forward 365

- Weekday: model amb consum, temperatures màxima i minima , humitat i dia de la setmana dels 14 dies passats, on obtenim un error RMSE de 0.03079, és a dir, aquest model és millor que el model de partida sense el dia de la setmana.

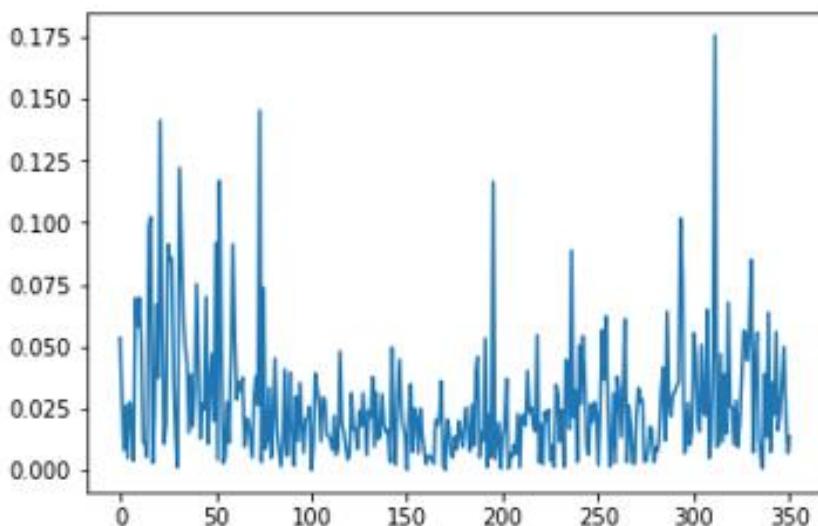


Gràfic 70 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i weekday dels 14 dies anteriors) one-step amb walk-forward 365



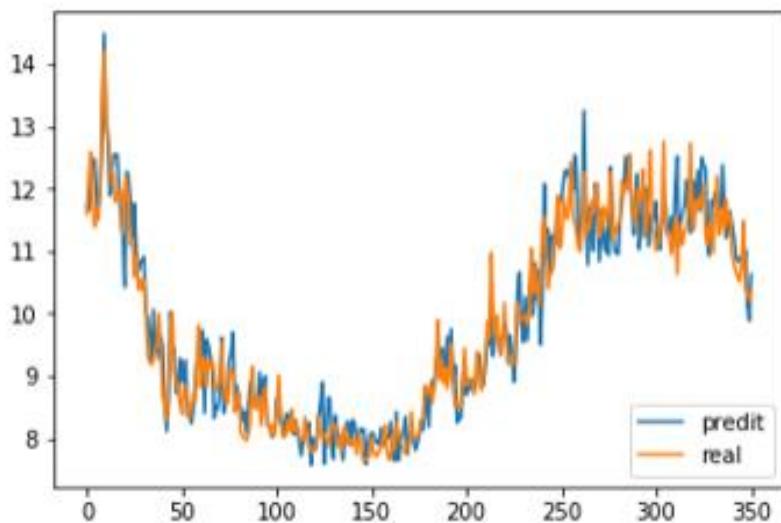
Gràfic 71 RMSE model multivariate (temperatura màxima i mínima, humitat i weekday dels 14 dies anteriors) one-step amb walk-forward 365

L'error percentual mitjà és del 2.72%, on la diferència amb els models anteriors és que només 14 dies tenen un error més elevat del 7,5%, tot i que aquells errors són més elevats que en els altres models.

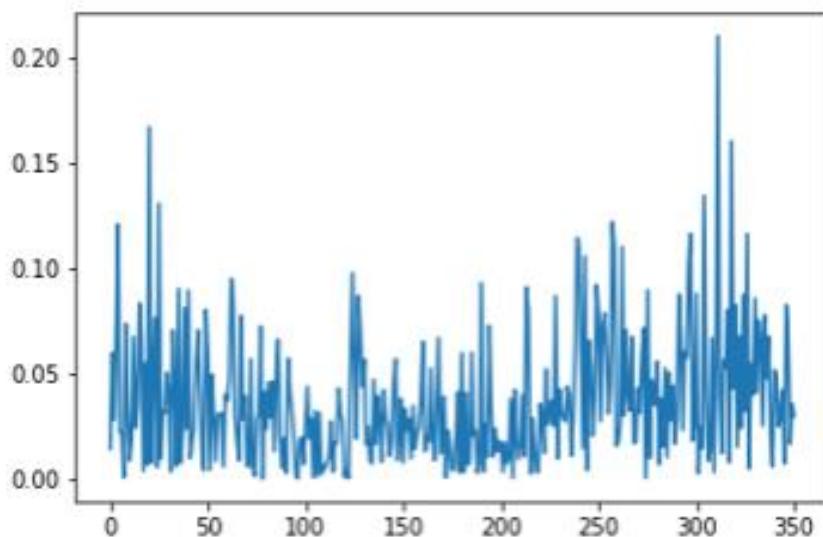


Gràfic 72 Errors model multivariate (temperatura màxima i mínima, humitat i weekday dels 14 dies anteriors) one-step amb walk-forward 365

- CloudCover: model amb consum, temperatures màxima i minima , humitat i cobertura dels núvols dels 14 dies passats, on obtenim un error RMSE de 0.03741, és a dir, aquest model és pitjor que el model de partida sense la cobertura de núvols.

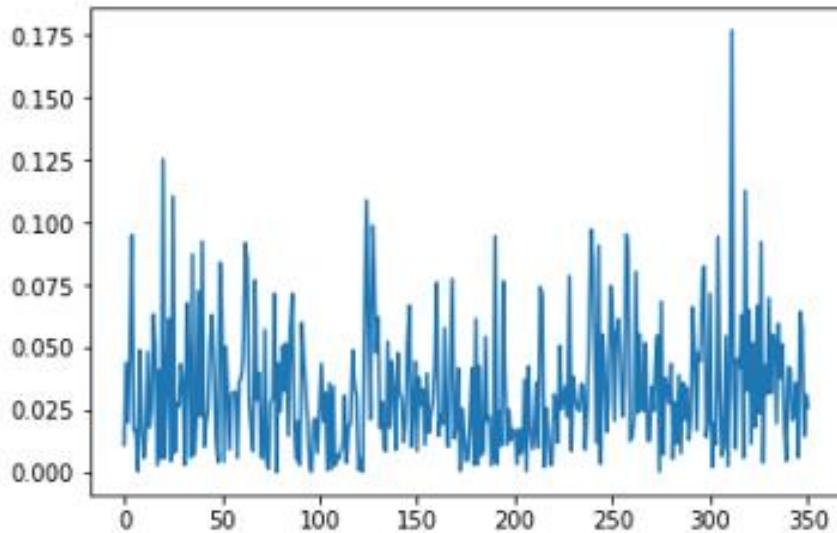


Gràfic 73 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i cloudCover dels 14 dies anteriors) one-step amb walk-forward 365



Gràfic 74 RMSE model multivariate (temperatura màxima i mínima, humitat i cloudCover dels 14 dies anteriors) one-step amb walk-forward 365

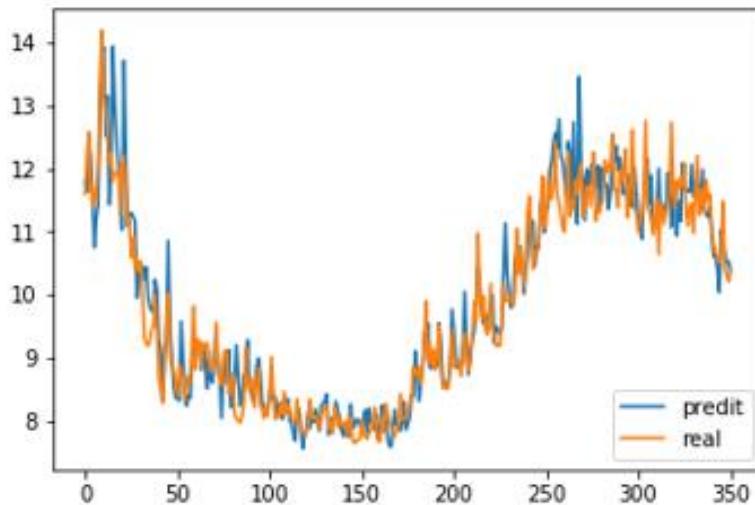
L'error mitjà és del 3.35%.



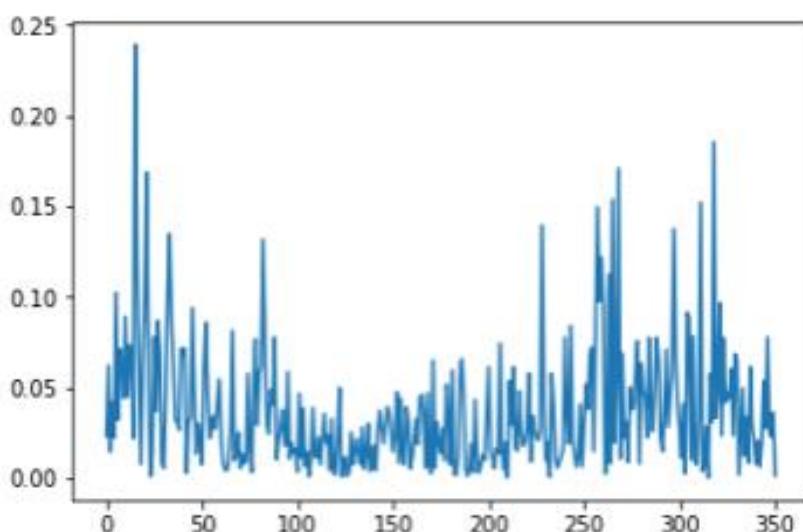
Gràfic 75 Errors model multivariate (temperatura màxima i mínima, humitat i cloudCover dels 14 dies anteriors) one-step amb walk-forward 365

Només amb el model que hi hem introduït la pressió i en el que hem introduït weekday ha millorat l'error global. Tot i així provem altres combinacions per veure si amb les diferents variables combinades millora.

- Pressió i sunsetTimeHour: model amb consum, temperatures màxima i minima , humitat, pressió i hora de posta de sol dels 14 dies passats, on obtenim un error RMSE de 0.03696, és a dir, aquest model és pitjor que el model de partida sense la hora de posta de sol i la pressió.

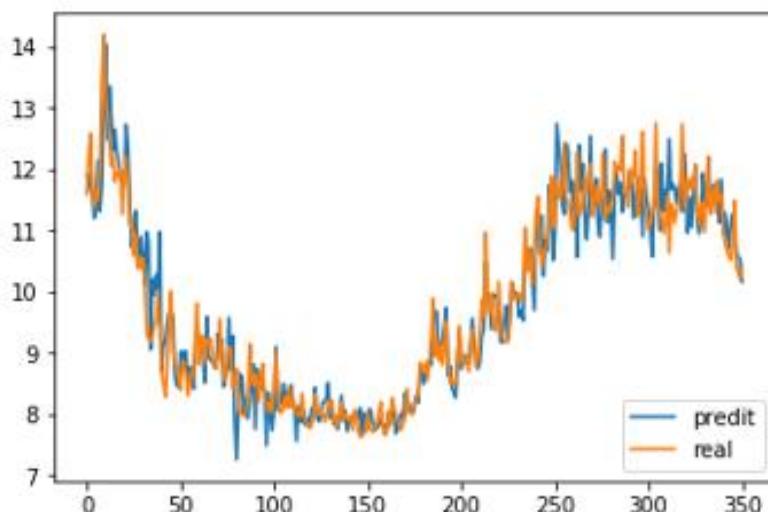


Gràfic 76 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat , pressió i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward 365

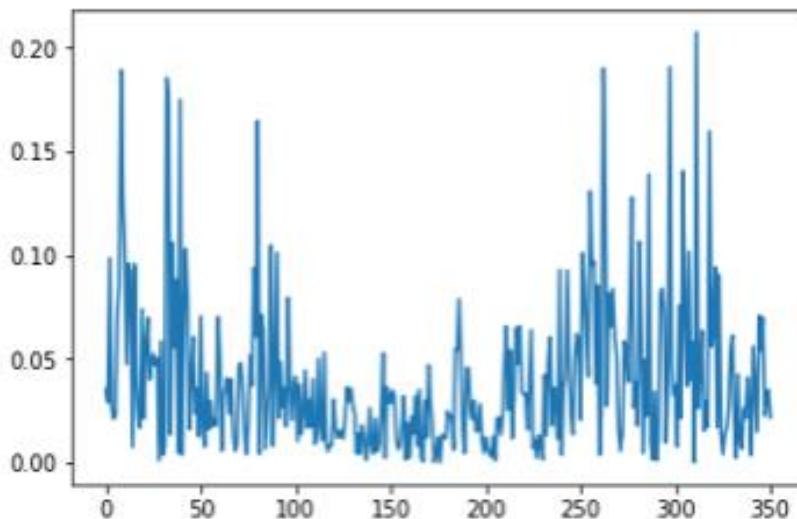


Gràfic 77 RMSE model multivariate (temperatura màxima i mínima, humitat , pressió i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward 365

- Pressió i cloudCover: model amb consum, temperatures màxima i minima , humitat, pressió i cobertura dels núvols dels 14 dies passats, on obtenim un error RMSE de 0.03946, és a dir, aquest model és pitjor que el model de partida sense la cobertura dels núvols i la pressió.

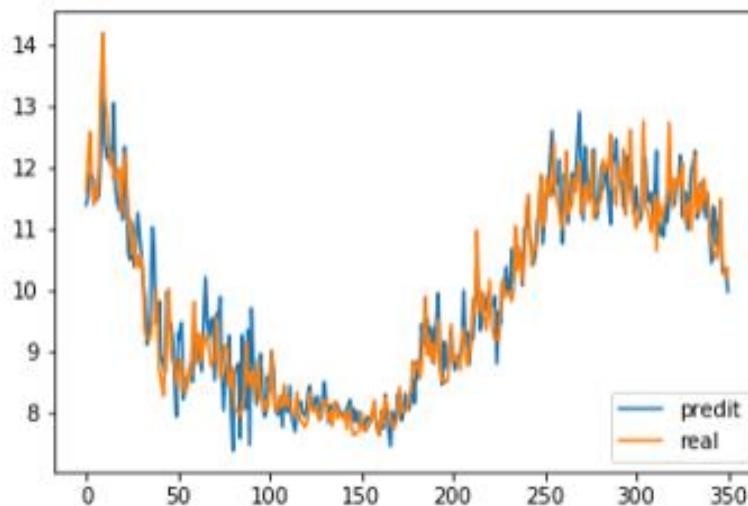


Gràfic 78 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat , pressió i cloudCover dels 14 dies anteriors) one-step amb walk-forward 365

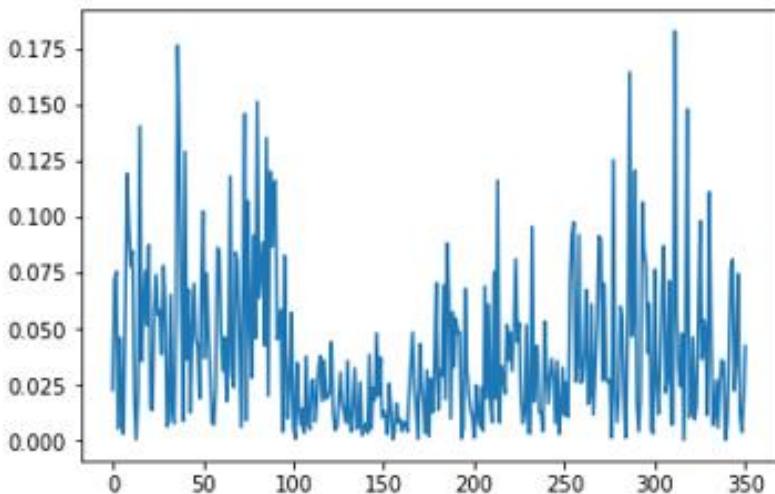


Gràfic 79 RMSE model multivariate (temperatura màxima i mínima, humitat , pressió i cloudCover dels 14 dies anteriors) one-step amb walk-forward 365

- Pressió, sunsetTimeHour i cloudCover: model amb consum, temperatures màxima i minima , humitat, pressió , hora de posta de sol i cobertura dels núvols dels 14 dies passats, on obtenim un error RMSE de 0.03957, és a dir, aquest model és pitjor que el model de partida sense la hora de posta de sol , pressió i cobertura dels núvols..

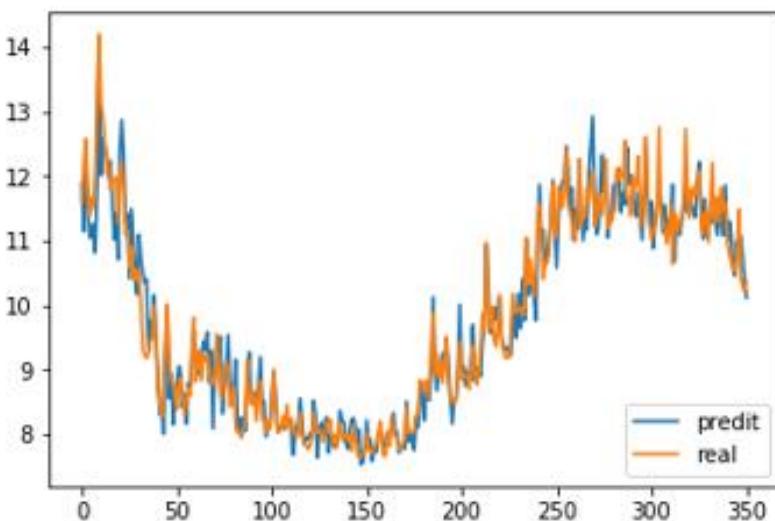


Gràfic 80 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat , pressió, sunsetTimeHour i cloudCover dels 14 dies anteriors) one-step amb walk-forward 365

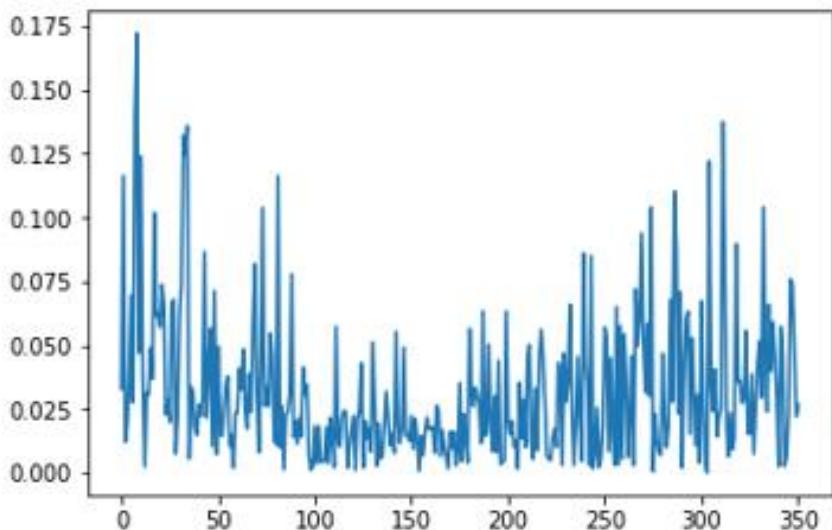


Gràfic 81 RMSE model multivariate (temperatura màxima i mínima, humitat , pressió, sunsetTimeHour i cloudCover dels 14 dies anteriors) one-step amb walk-forward 365

- Pressió i weekday: model amb consum, temperatures màxima i minima , humitat, pressió i dia de la setmana dels 14 dies passats, on obtenim un error RMSE de 0.03171, és a dir, aquest model és millor que el model de partida sense el dia de la setmana i la pressió, tot i que és pitjor que el model que utilitza weekday sense pressió.

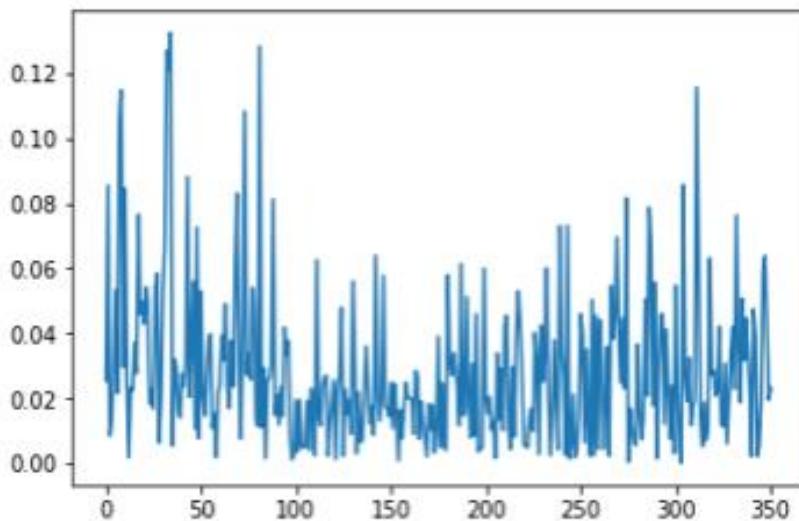


Gràfic 82 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat , pressió i weekday dels 14 dies anteriors) one-step amb walk-forward 365



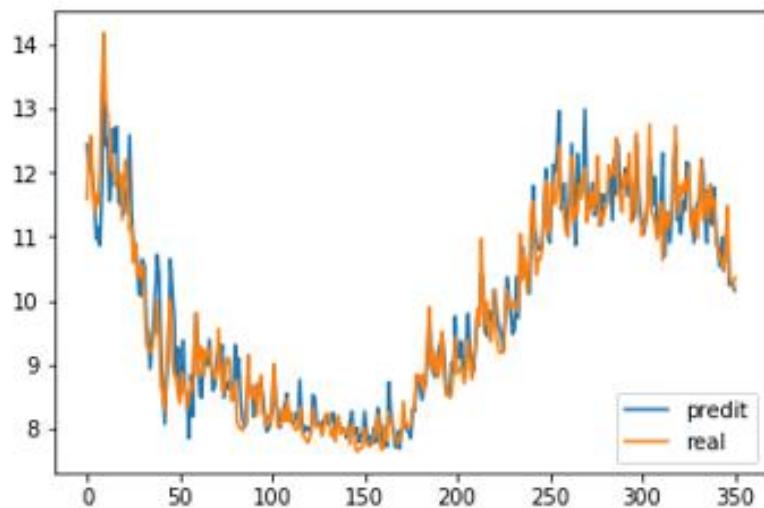
Gràfic 83 RMSE model multivariate (temperatura màxima i mínima, humitat , pressió i weekday dels 14 dies anteriors) one-step amb walk-forward 365

La mitjana de l'error percentual és de 2.82%, on hi ha 6 errors prop del 12% i els altres errors no superen o igualen el 8%.

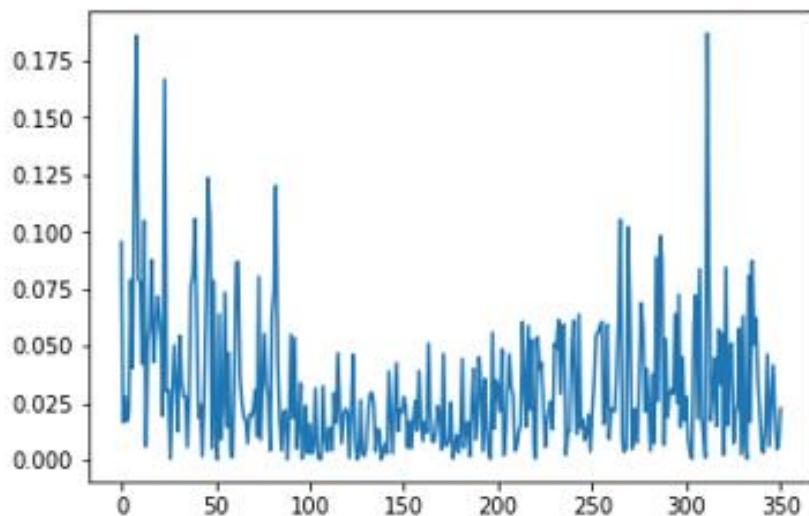


Gràfic 84 Errors model multivariate (temperatura màxima i mínima, humitat , pressió i weekday dels 14 dies anteriors) one-step amb walk-forward 365

- Pressió, weekday, sunsetTimeHour: model amb consum, temperatures màxima i minima , humitat, pressió, dia de la setmana i hora de posta de sol dels 14 dies passats, on obtenim un error RMSE de 0.03134, és a dir, aquest model és millor que el model de partida sense el dia de la setmana , la pressió i weekday, tot i que és pitjor que el model que utilitza weekday sense pressió ni hora de posta de sol.

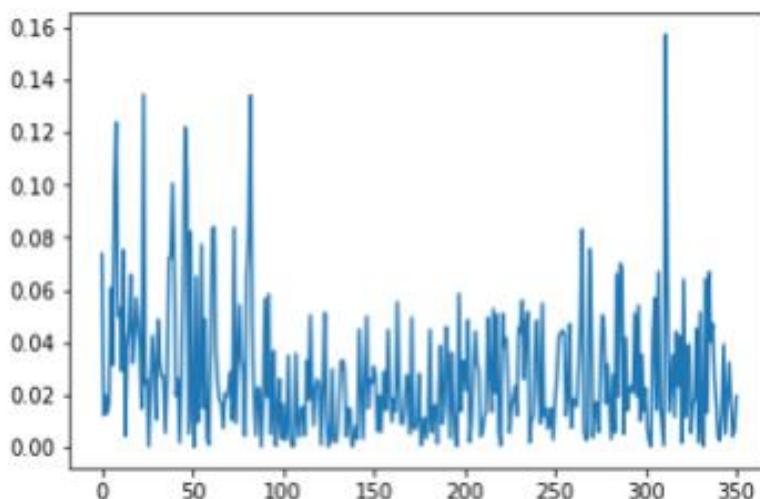


Gràfic 85 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat , pressió, sunsetTimeHour i weekday dels 14 dies anteriors) one-step amb walk-forward 365



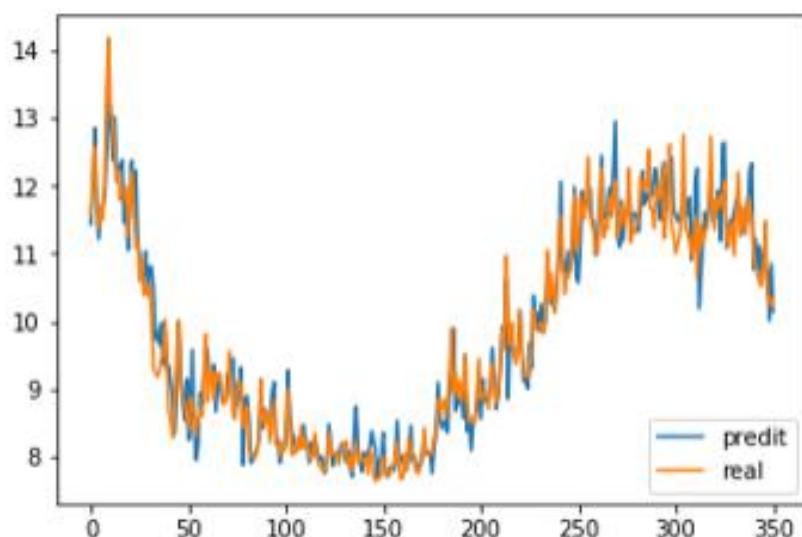
Gràfic 86 RMSE model multivariate (temperatura màxima i mínima, humitat , pressió, sunsetTimeHour i weekday dels 14 dies anteriors) one-step amb walk-forward 365

Té un 2,78% d'error mitjà percentual

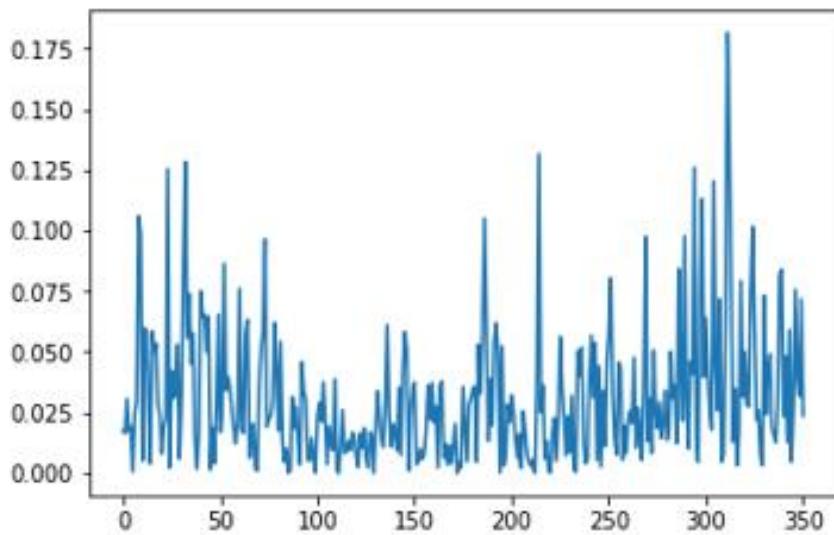


Gràfic 87 Error model multivariate (temperatura màxima i mínima, humitat , pressió, sunsetTimeHour i weekday dels 14 dies anteriors) one-step amb walk-forward 365

Recordem que el millor model fins ara és el que utilitza el consum, temperatures màximes i mínimes, humitat i dia de la setmana amb 14 dies anteriors. Si eliminem la variable humitat d'aquest model, millorem el model obtenint un 0.03083 de RMSE.

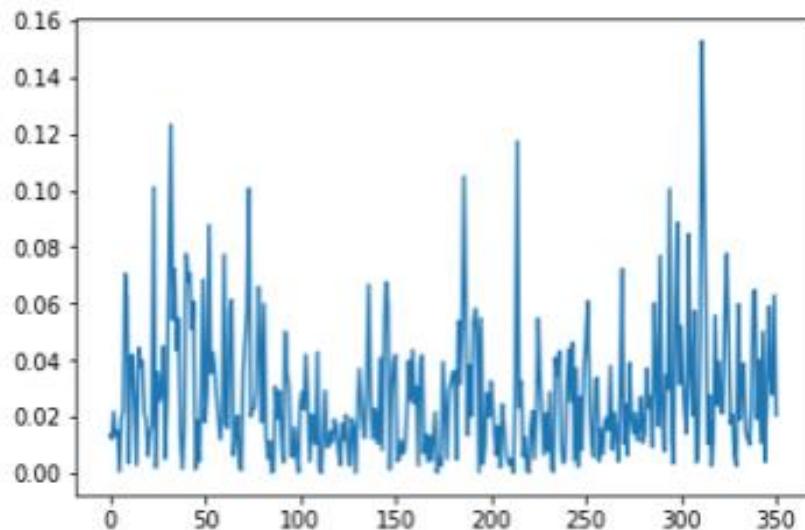


Gràfic 88 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward 365



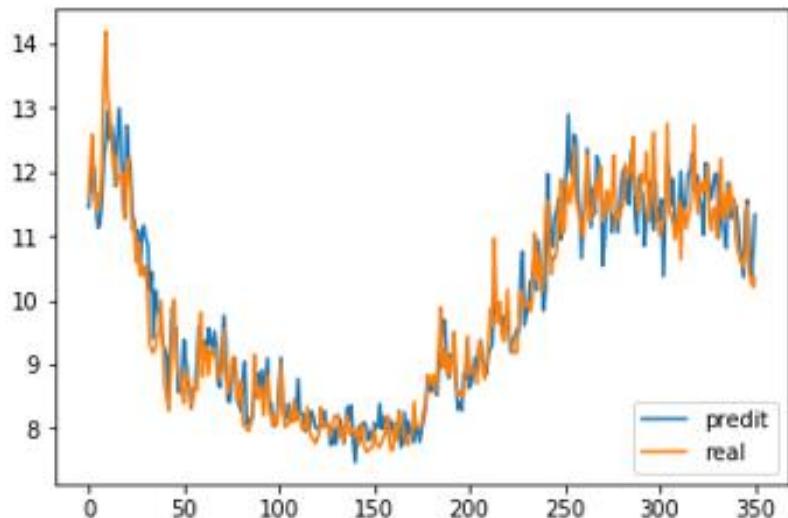
Gràfic 89 RMSE model multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward 365

Té un 2,76% d'error percentual mitjà,

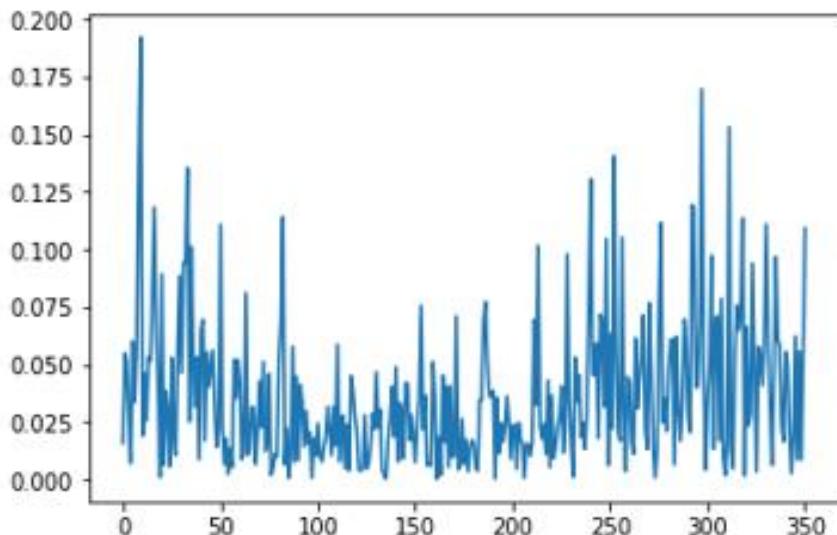


Gràfic 90 Errors model multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward 365

Fins ara hem utilitzat l'optimitzador Adam en el model, agafem un dels models que hem fet i canviem l'optimitzador per veure si millorem el model. Per fer-ho agafem el model que utilitza el consum, les temperatures màxima i mínima, la humitat i la pressió, que tenia un 0.03389 de RMSE, i al canviar l'optimitzador obtenim un 0.03665 RMSE, és a dir, l'optimitzador Adam funciona millor amb les nostres dades.



Gràfic 91 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i pressió dels 14 dies anteriors) one-step amb walk-forward 365 canviant l'optimitzador



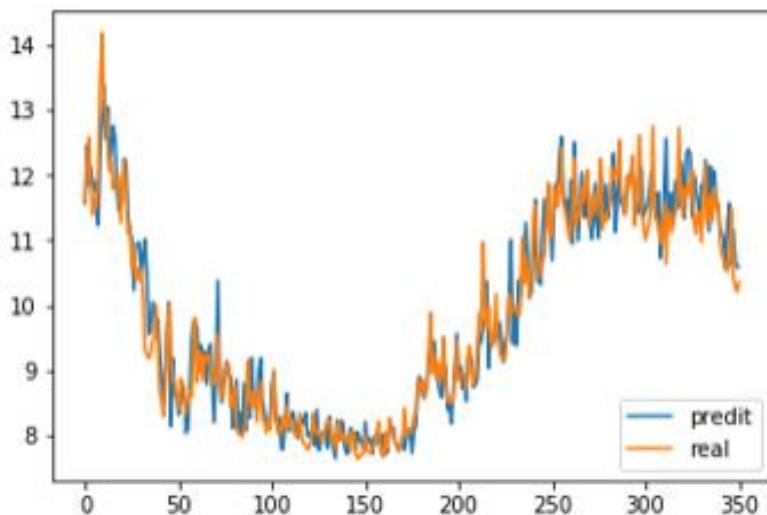
Gràfic 92 RMSE model multivariate (temperatura màxima i mínima, humitat i pressió dels 14 dies anteriors) one-step amb walk-forward 365 canviant l'optimitzador

5.3.1.2 Walkforward augmentant més 365 dies

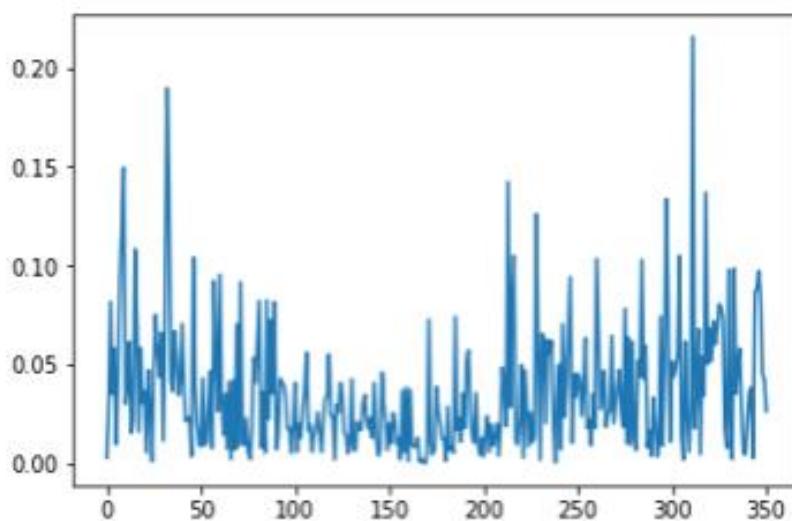
Fins ara els models utilitzats agafaven sempre els 365 dies previs per a l'entrenament, per exemple per predir el 16-03-2013 utilitzem un train del 15-03-2012 al 15-03-2013, i per predir el dia següent, el 17-03-2013, utilitzem del 16-03-2012 al 16-03-2013.

Volem provar si a mesura que anem avançant en les prediccions, que el train no avanci, de manera que sempre sigui del 15-03-2012 al dia abans del dia que volem fer la predicción. És a dir, fem un walk forward i enllac d'utilitzar sempre 365 per train, train sempre va creixent.

Comencem provant aquest model utilitzant les variables consum, temperatures màxima i mínima i humitat amb 14 dies anteriors, obtenint un 0.03598 de RMSE, un valor més alt que el del model amb un train de 365 dies.



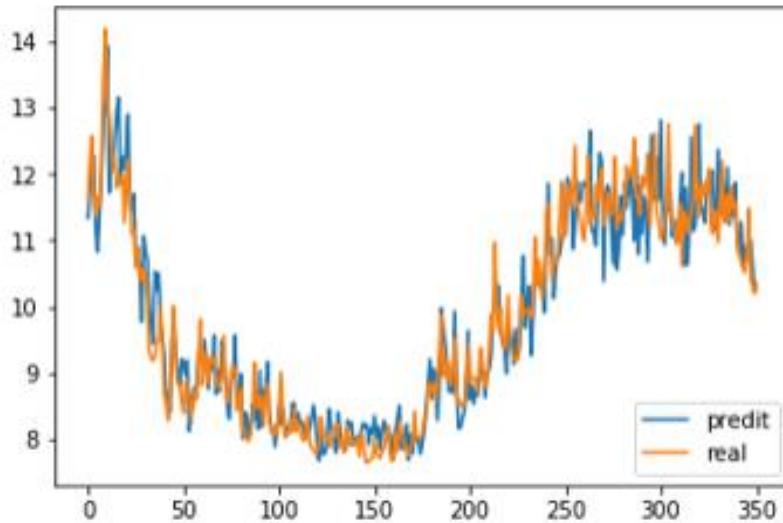
Gràfic 93 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors) one-step amb walk-forward augmentant train



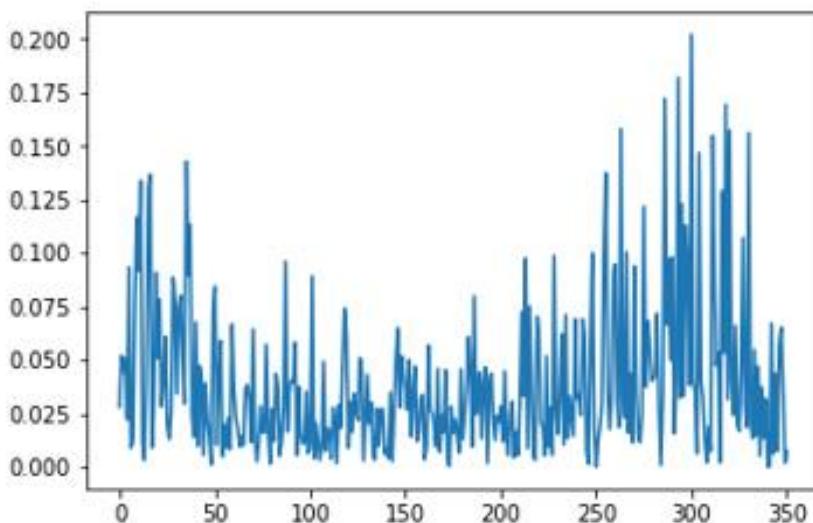
Gràfic 94 RMSE model multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors) one-step amb walk-forward augmentant train

A aquest model “base” amb consum, temperatura màxim i mínima i humitat dels 14 dies anteriors hi afegim varies variables individualment:

- cloudCover:
Obtenim un RMSE de 0.03915, més alt que amb el model amb les mateixes variables que utilitza walkforward de 365 dies de train.



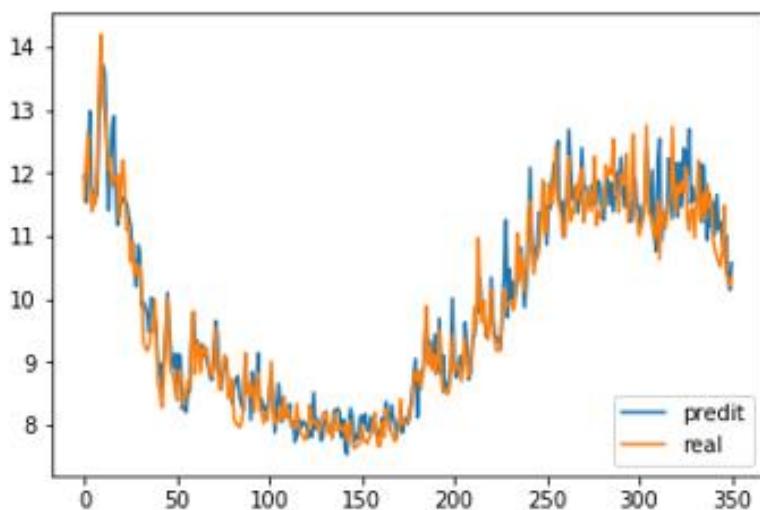
Gràfic 95 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i cloudCover dels 14 dies anteriors) one-step amb walk-forward augmentant train.



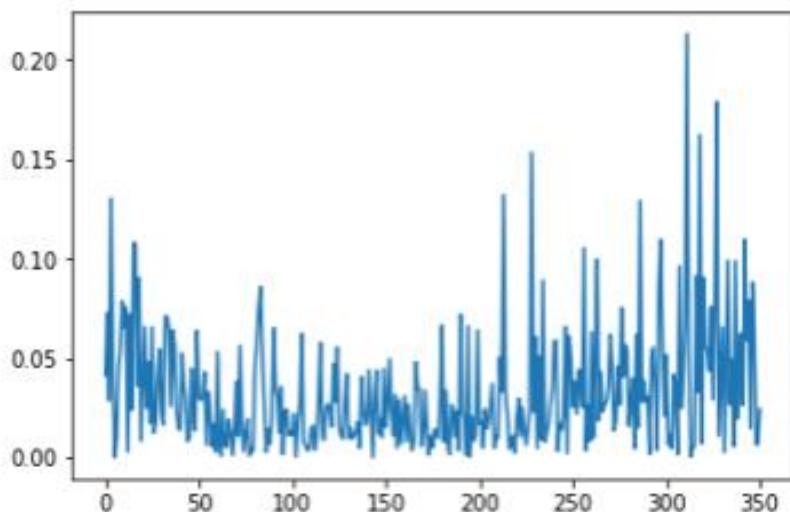
Gràfic 96 RMSE model multivariate (temperatura màxima i mínima, humitat i cloudCover dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- sunsetTimeHour

Obtenim un RMSE de 0.03299, més baix que amb el model amb les mateixes variables que utilitza walkforward de 365 dies de train

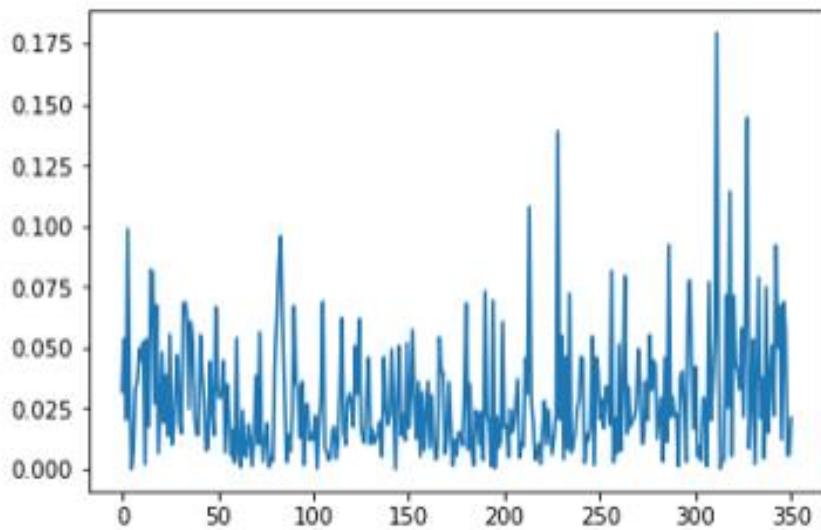


Gràfic 97 Predicções del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.



Gràfic 98 RMSE model multivariate (temperatura màxima i mínima, humitat i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

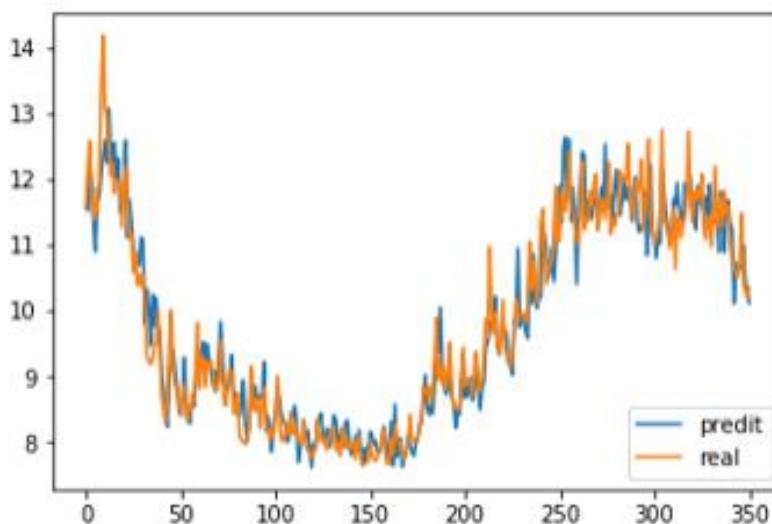
Error mitjà de 2.93%



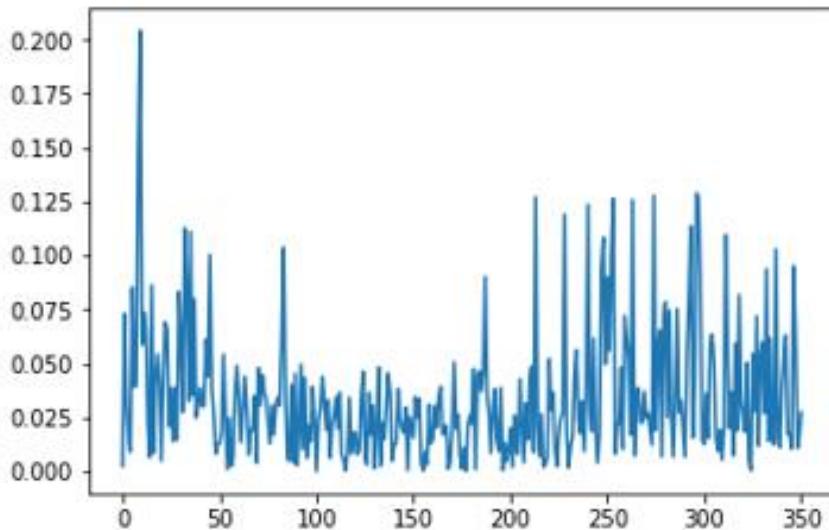
Gràfic 99 Errors model multivariate (temperatura màxima i mínima, humitat i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Pressure

Obtenim un RMSE de 0.03402, més alt que amb el model amb les mateixes variables que utilitza walkforward de 365 dies de train



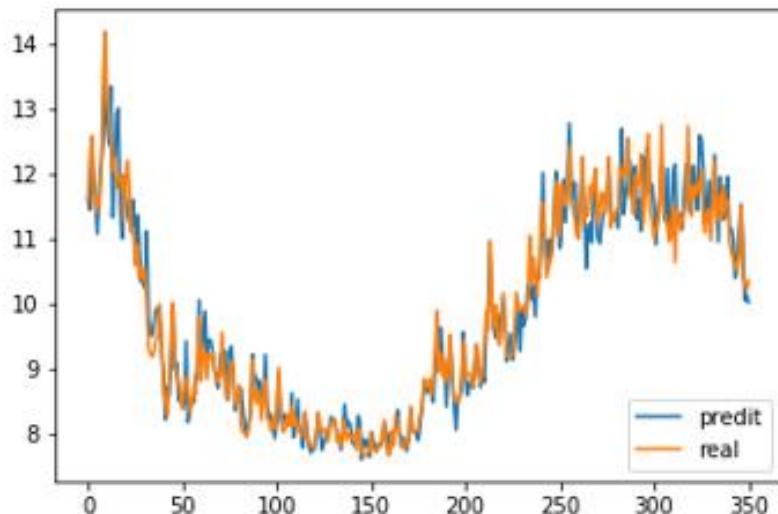
Gràfic 100 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i pressió dels 14 dies anteriors) one-step amb walk-forward augmentant train.



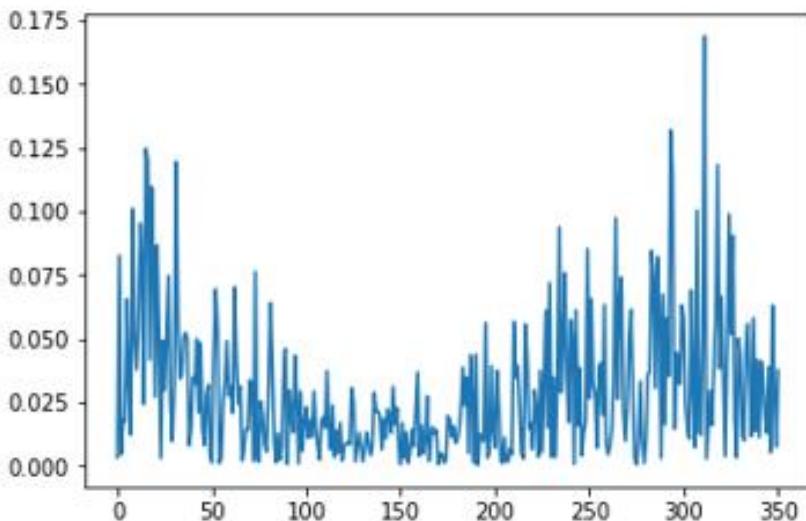
Gràfic 101 RMSE model multivariate (temperatura màxima i mínima, humitat i pressió dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Weekday

Obtenim un RMSE de 0.02920, més baix que amb el model amb les mateixes variables que utilitza walkforward de 365 dies de train

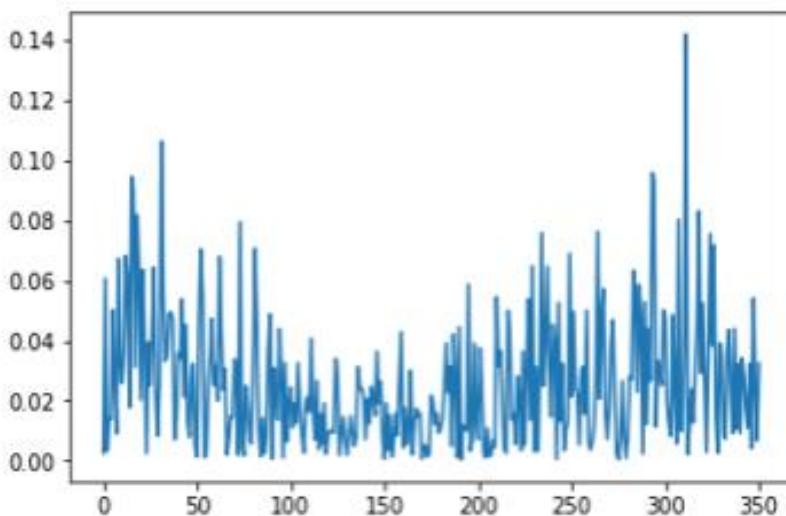


Gràfic 102 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train.



Gràfic 103 RMSE model multivariate (temperatura màxima i mínima, humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train.

2.55% error

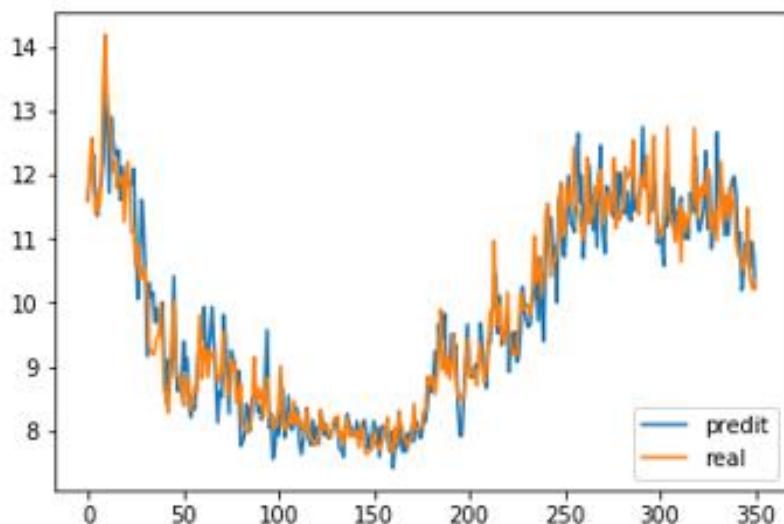


Gràfic 104 Error model multivariate (temperatura màxima i mínima, humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train.

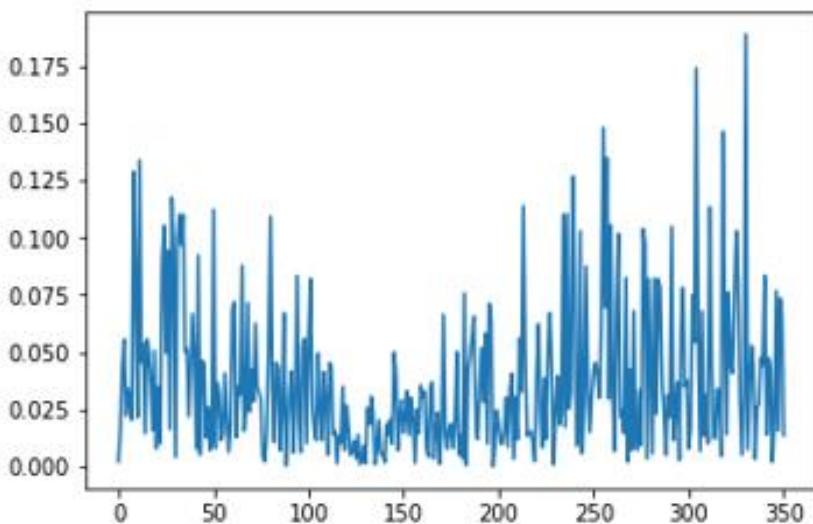
Veiem que els models amb posta de sol i el de dia de la setmana milloren utilitzant el train que va augmentant, sent el model amb el dia de la setmana el que millor funciona.

Provem altres combinacions a partir del model amb consum, temperatures i humitat:

- Pressure i cloudCover
Obtenim un RMSE de 0.03724

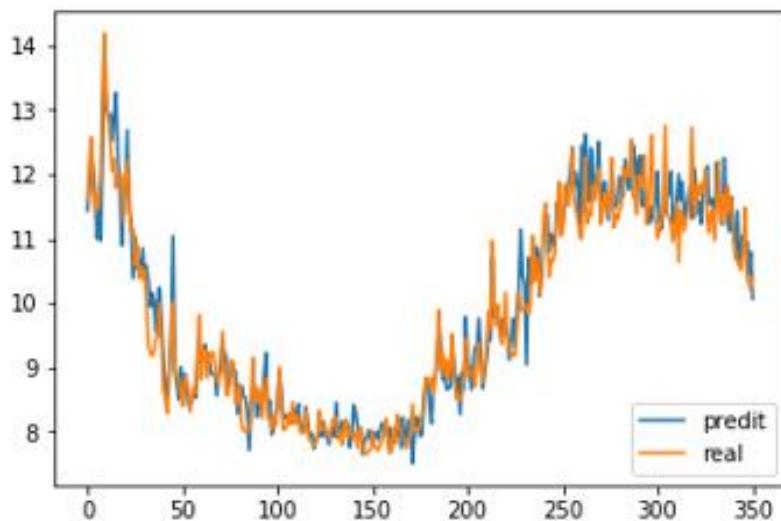


Gràfic 105 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat pressure i cloudCover dels 14 dies anteriors) one-step amb walk-forward augmentant train.

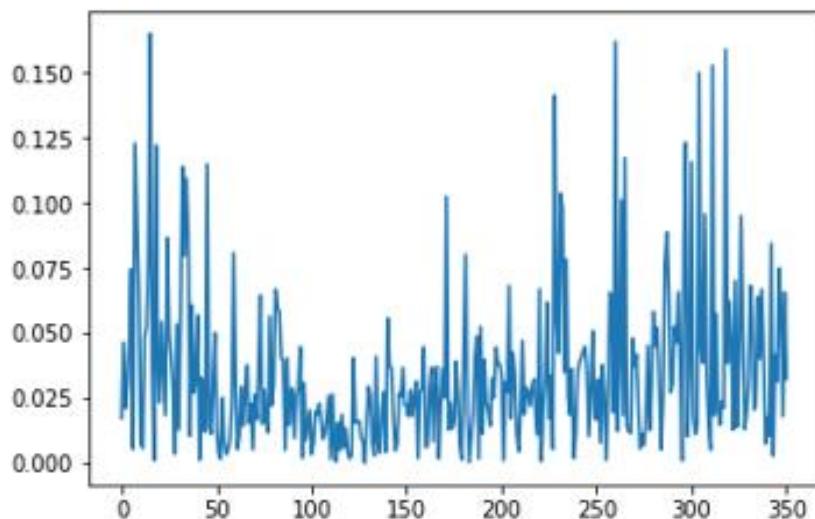


Gràfic 106 RMSE model multivariate (temperatura màxima i mínima, humitat pressure i cloudCover dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Pressure i sunsetTimeHour
Obtenim un RMSE de 0.03398

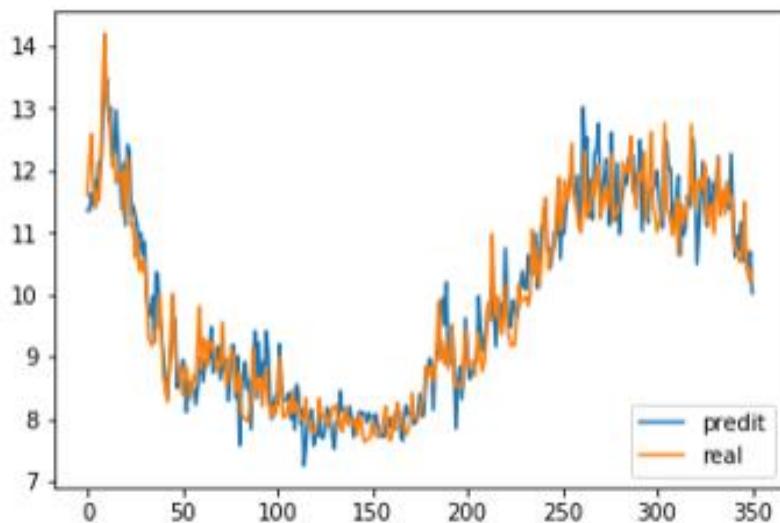


Gràfic 107 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat pressure i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

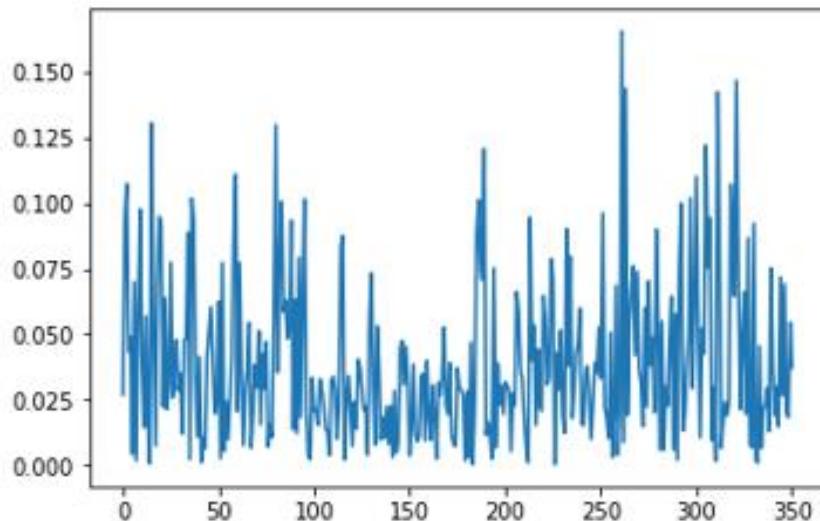


Gràfic 108 RMSE model multivariate (temperatura màxima i mínima, humitat pressure i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Pressure, sunsetTimeHOur i cloudCover
Obtenim un RMSE de 0.03896

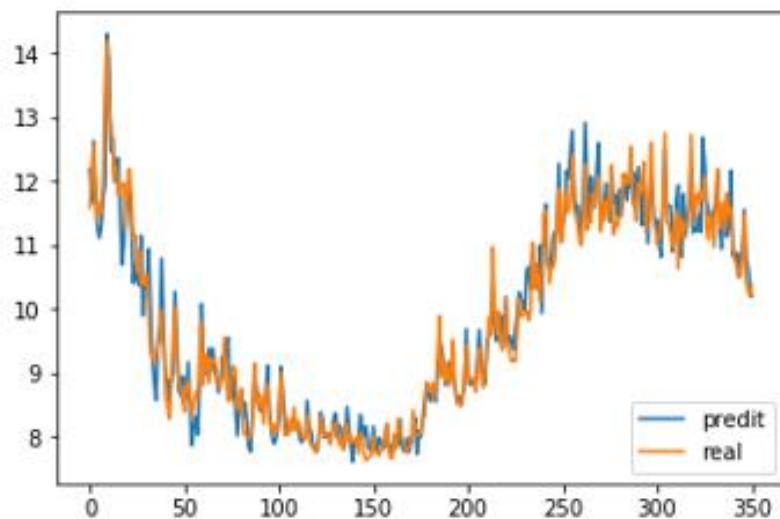


Gràfic 109 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat pressure, cloudCover i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

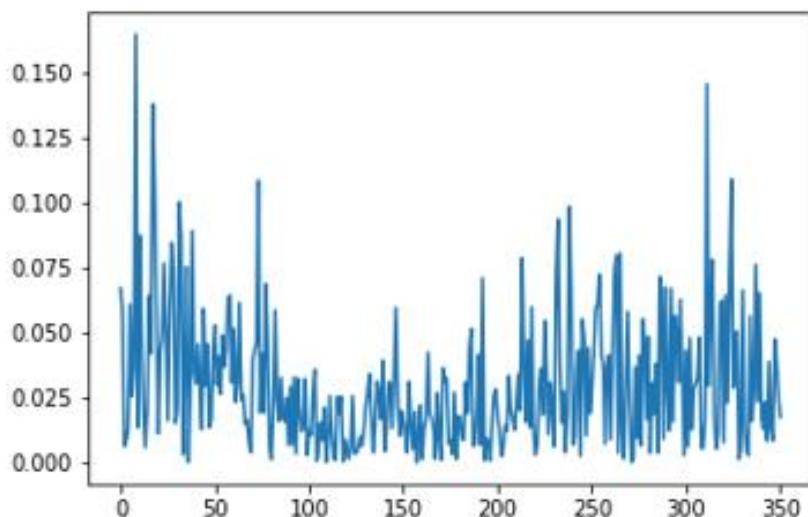


Gràfic 110 RMSE model multivariate (temperatura màxima i mínima, humitat pressure, cloudCover i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Pressure, sunsetTimeHour i weekday
Obtenim un RMSE de 0.2998

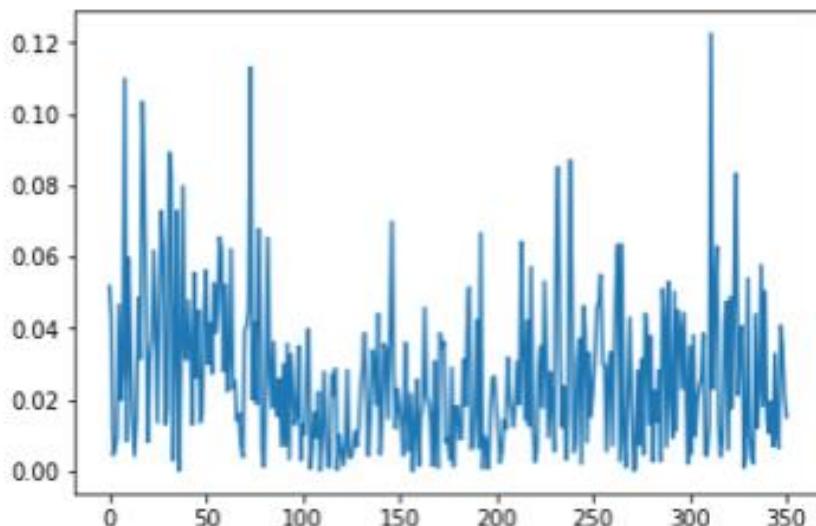


Gràfic 111 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat, pressure, weekday i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.



Gràfic 112 RMSE model multivariate (temperatura màxima i mínima, humitat, pressure, weekday i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

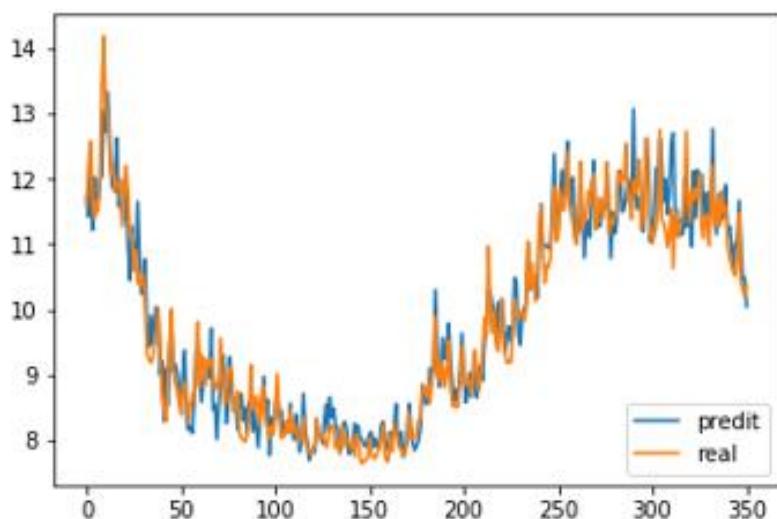
Tenim un error percentual mitjà del 2.67%



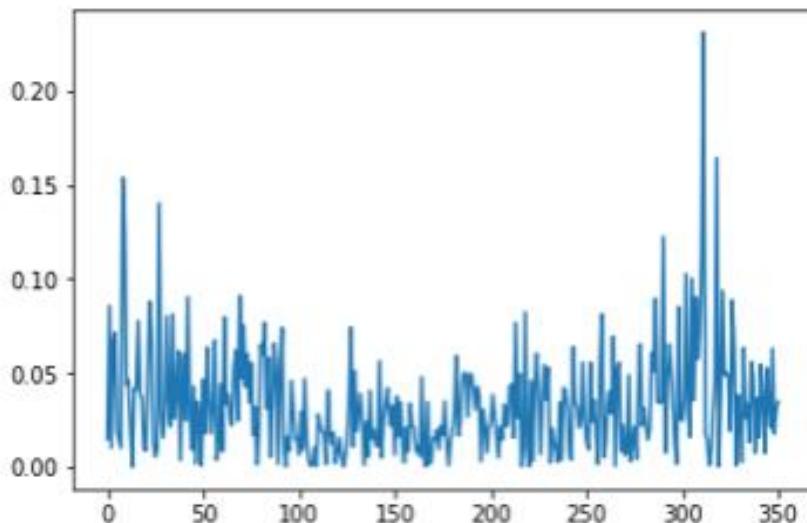
Gràfic 113 Errors model multivariate (temperatura màxima i mínima, humitat, pressure, weekday i sunsetTimeHour dels 14 dies anteriors) one-step amb walk-forward augmentant train.

El millor model fins ara és el que incorpora weekday. Provem aquest model (amb consum, temperatures mínima i màxima, humitat i weekday) afegint-hi algunes variables:

- Month
- Obtenim un 0.03317 de RMSE



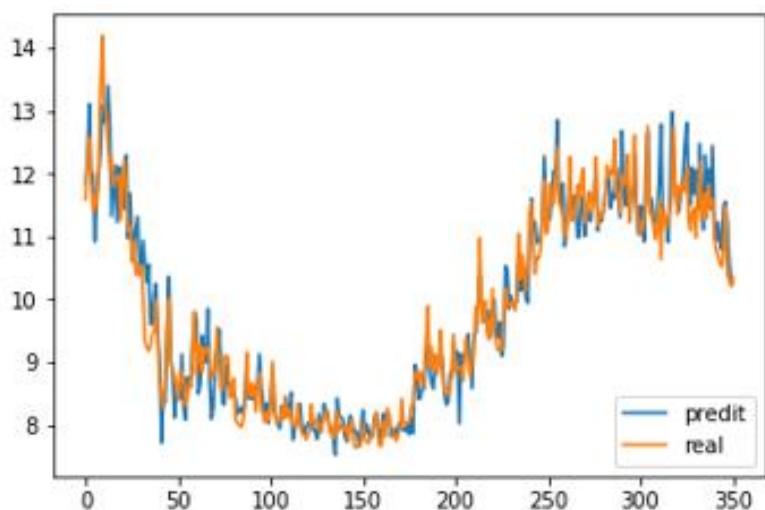
Gràfic 114 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat, weekday i month dels 14 dies anteriors) one-step amb walk-forward augmentant train.



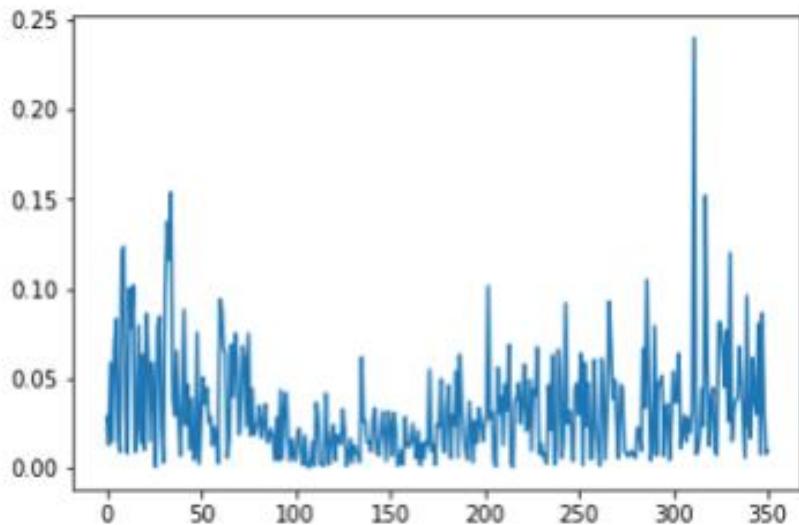
Gràfic 115 RMSE model multivariate (temperatura màxima i mínima, humitat, weekday i month dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Season

Obtenim un 0.03213 de RMSE

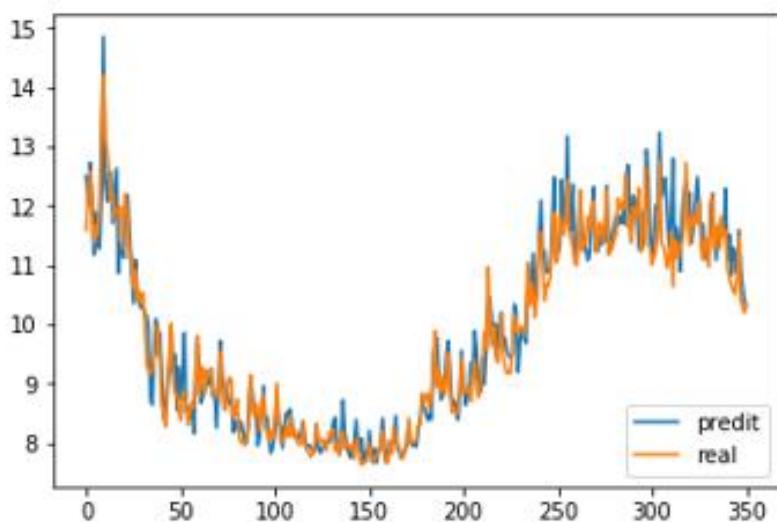


Gràfic 116 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat, weekday i season dels 14 dies anteriors) one-step amb walk-forward augmentant train.

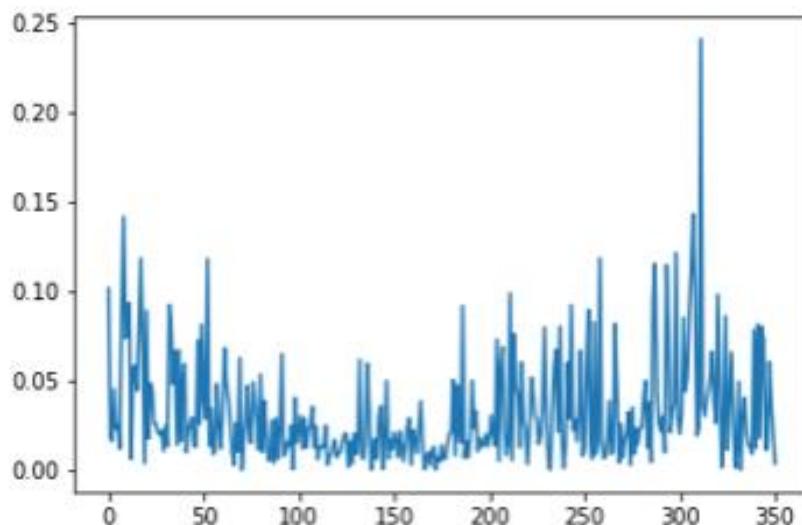


Gràfic 117 RMSE model multivariate (temperatura màxima i mínima, humitat, weekday i season dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Season i month
Obtenim un 0.03250 de RMSE

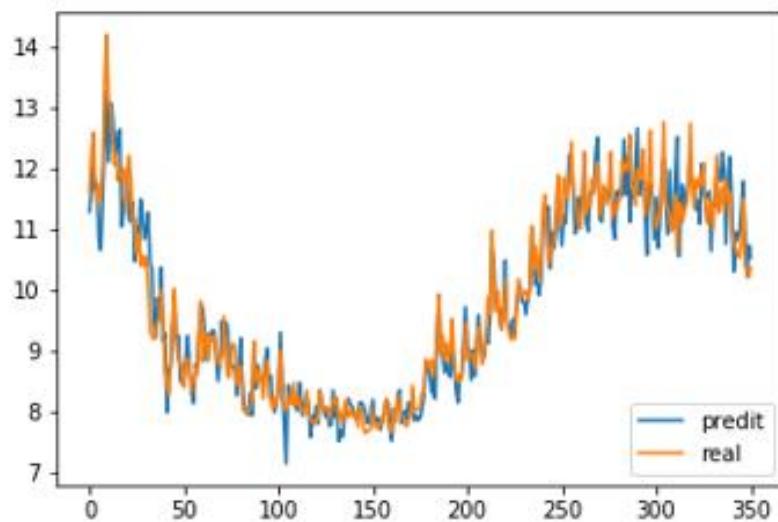


Gràfic 118 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat, weekday, season i month dels 14 dies anteriors) one-step amb walk-forward augmentant train.

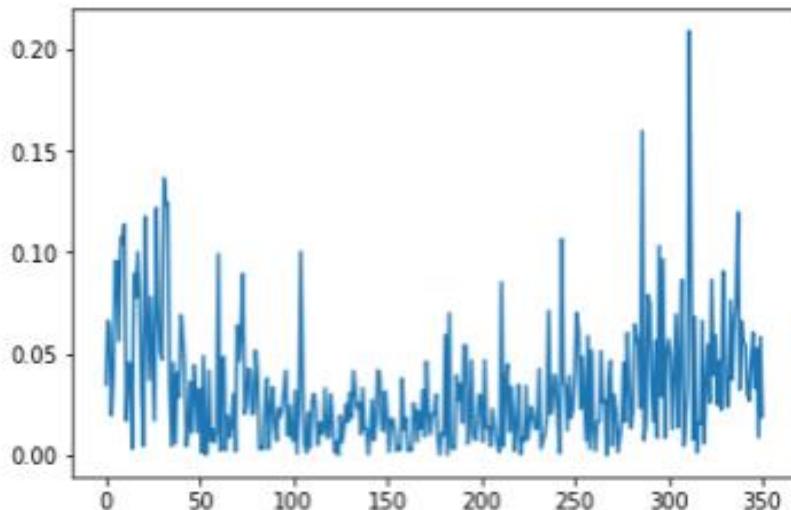


Gràfic 119 RMSE model multivariate (temperatura màxima i mínima, humitat, weekday, season i month dels 14 dies anteriors) one-step amb walk-forward augmentant train.

- Pressure
Obtenim un 0.03283 de RMSE



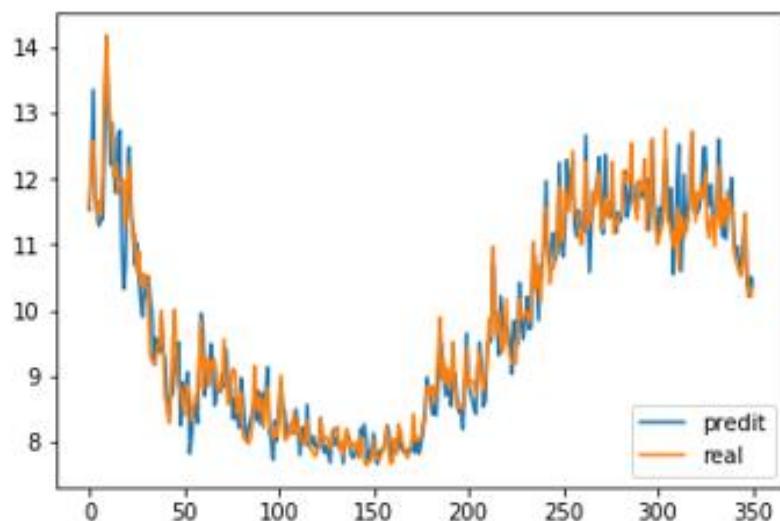
Gràfic 120 Prediccions del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima, humitat, weekday i pressió dels 14 dies anteriors) one-step amb walk-forward augmentant train.



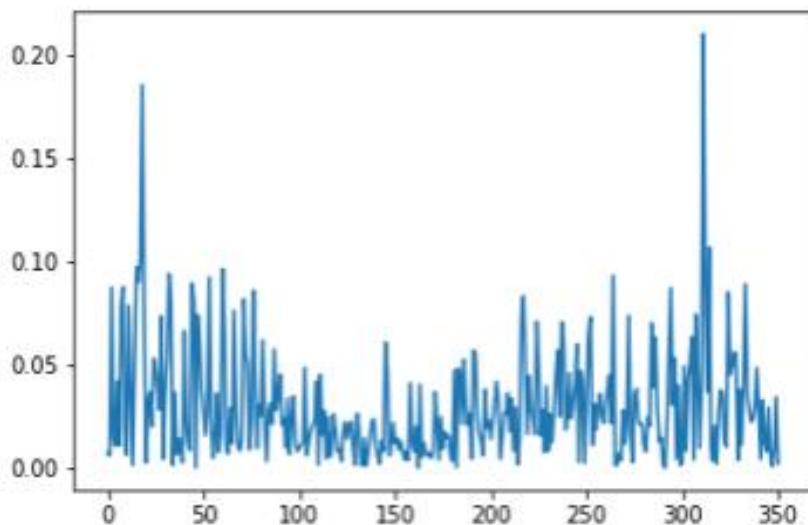
Gràfic 121 RMSE model multivariate (temperatura màxima i mínima, humitat, weekday i pressió dels 14 dies anteriors) one-step amb walk-forward augmentant train.

Cap d'aquests models millora el model amb weekday.

Provem també un model on tinguem el consum, les temperatures màxima i mínima i el dia de la setmana, sense la humitat, i obtenim un bon resultat de 0.02972 de RMSE.

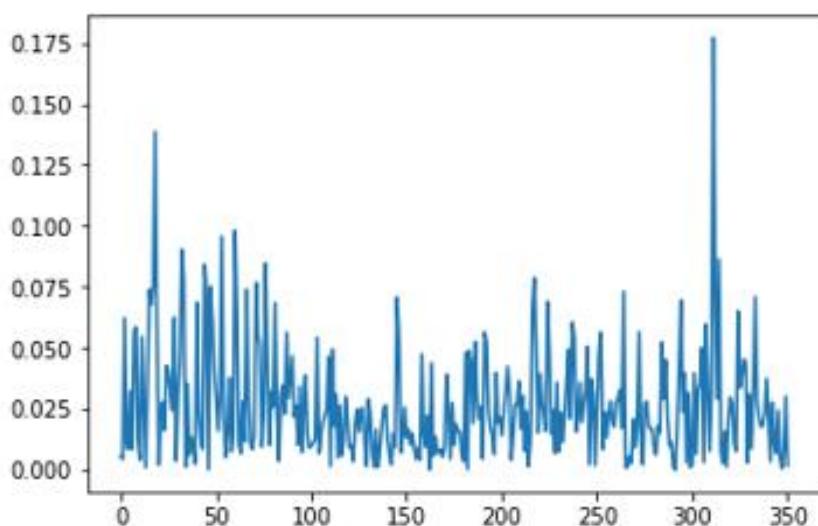


Gràfic 122 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train.



Gràfic 123 RMSE model multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train.

Error mitjà de 2.66%



Gràfic 124 Errors model multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train.

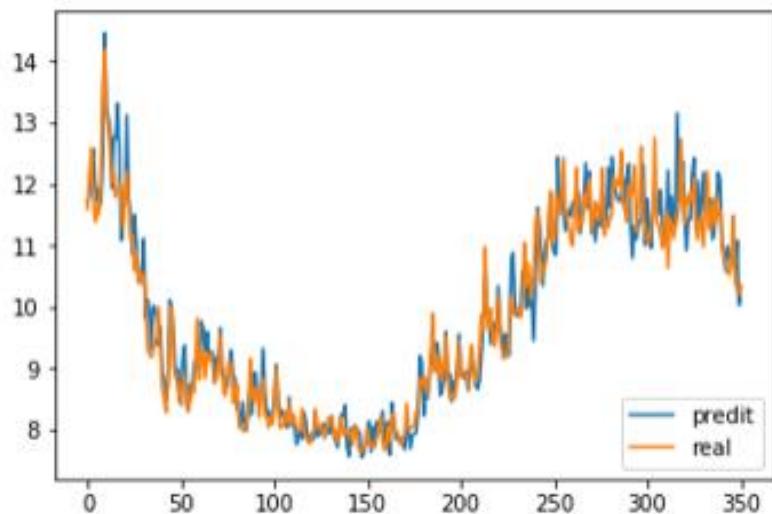
Finalment, aplicarem PCA a algun model per veure si millora els que ja tenim.

A partir del model base amb consum, temperatura màxima i mínima i humitat, hi afegim:

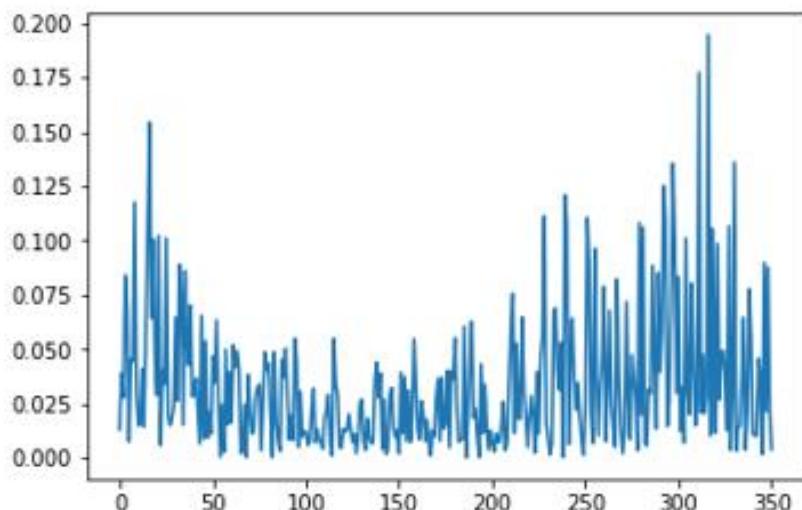
- PCA(pressió, sunsetTimeHour i cloudCover):

Creem una nova variable a partir de la reducció amb PCA de les variables pressió, sunsetTimeHour i cloudCover. Amb la variable reduïda juntament amb les variables consum, temperatures màxima i mínima i humitat amb les dades de 14 dies anteriors creem el model.

Obtenim un RMSE de 0.03331.



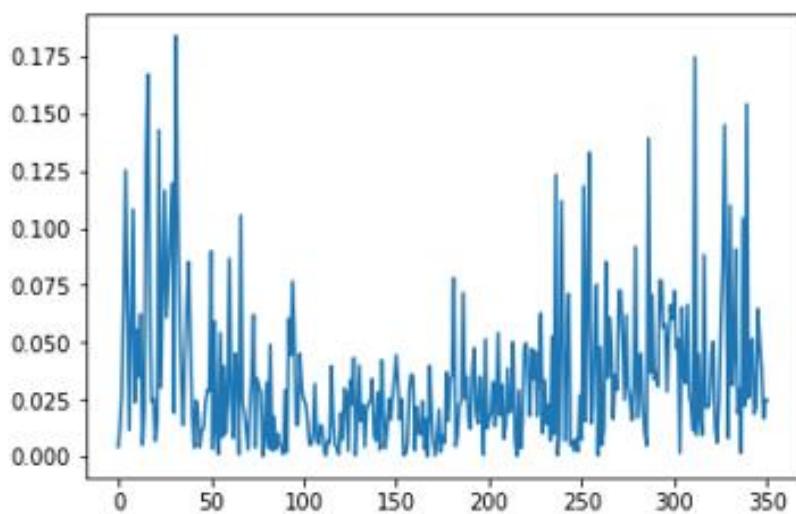
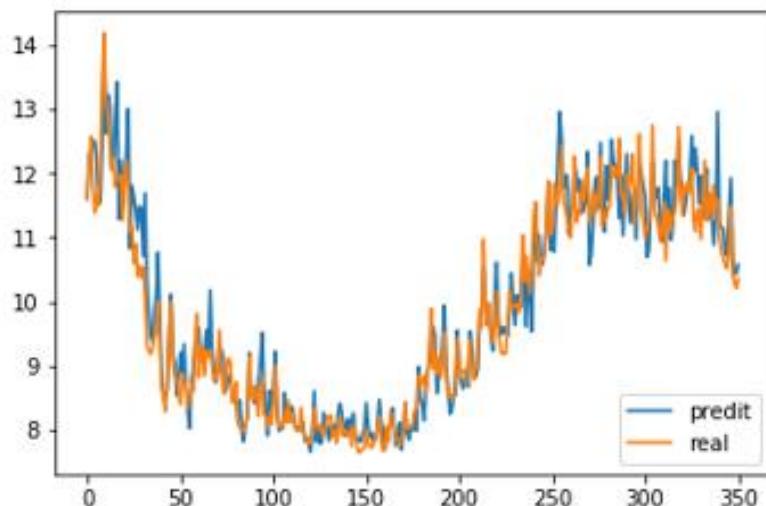
Gràfic 125 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors juntament amb variable extreta de PCA de pressió, sunsetTimeHour i cloudCover) one-step amb walk-forward augmentant train.



Gràfic 126 RMSE model multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors juntament amb variable extreta de PCA de pressió, sunsetTimeHour i cloudCover) one-step amb walk-forward augmentant train.

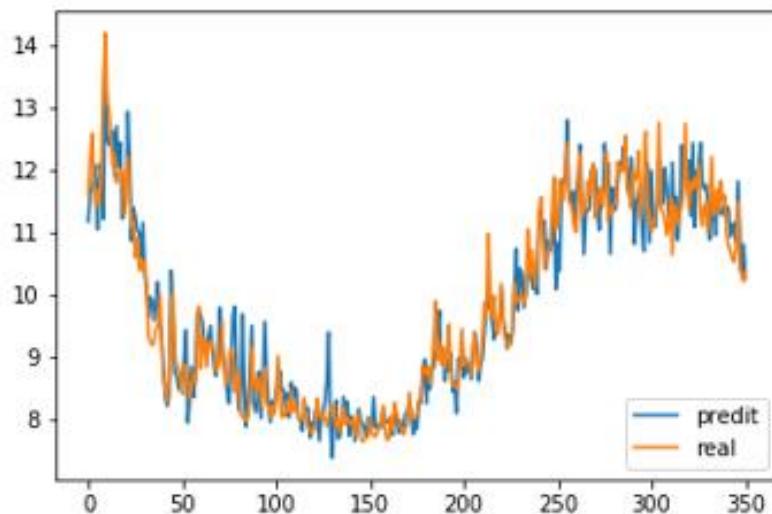
- Weekday, PCA(pressió, sunsetTimeHour, cloudCover): Creem una nova variable a partir de la reducció amb PCA de les variables pressió, sunsetTimeHour i cloudCover. Amb la variable reduïda juntament amb les variables consum, temperatures màxima i mínima , humitat i weekday amb les dades de 14 dies anteriors creem el model.

Obtenim un RMSE de 0.03457.

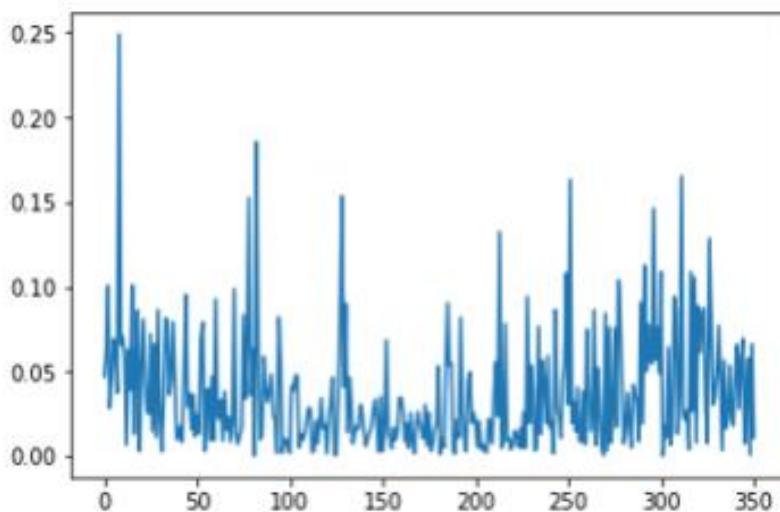


- PCA (weekday, pressió, sunsetTimeHour, cloudCover): Creem una nova variable a partir de la reducció amb PCA de les variables pressió, sunsetTimeHour, cloudCover i weekday. Amb la variable reduïda juntament amb les variables consum, temperatures màxima i mínima i humitat amb les dades de 14 dies anteriors creem el model.

Obtenim un RMSE de 0.03682.



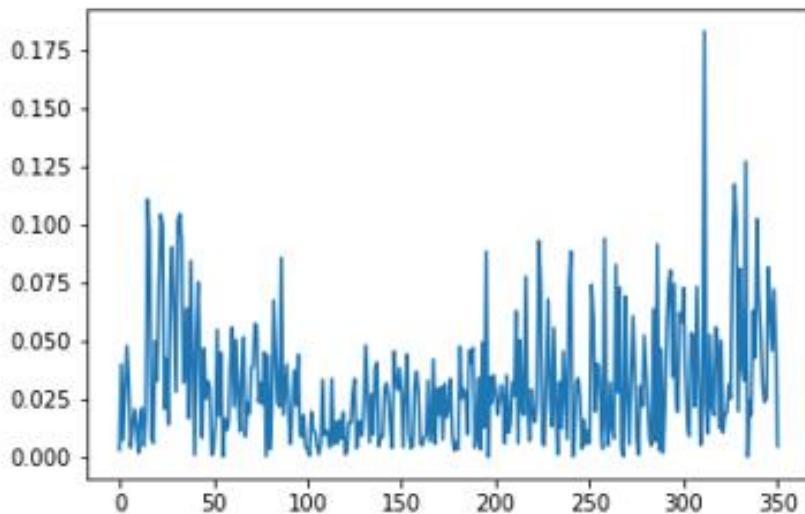
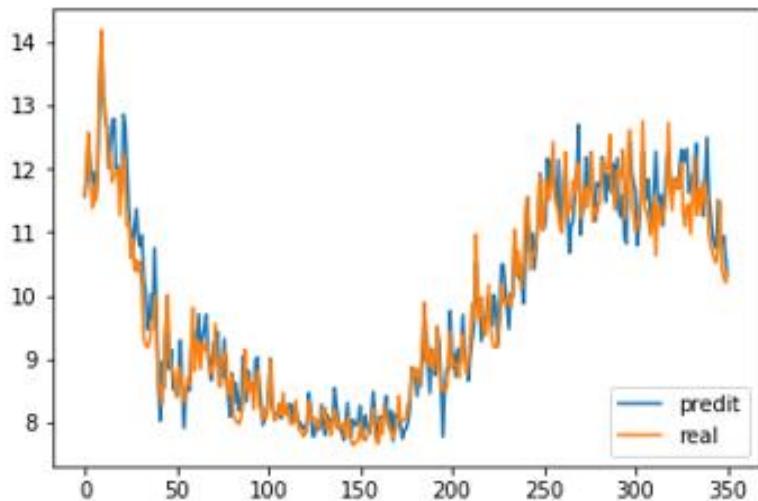
Gràfic 129 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors juntament amb variable extreta de PCA de pressió, sunsetTimeHour , cloudCover i weekday) one-step amb walk-forward augmentant train.



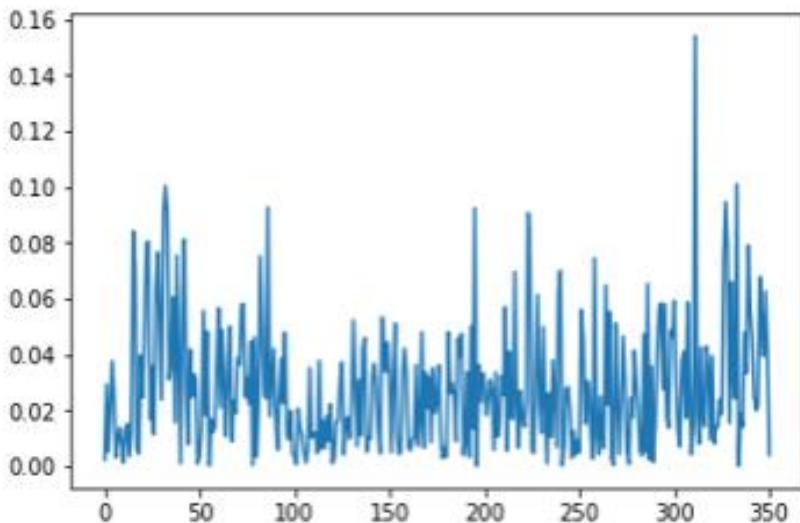
Gràfic 130 RMSE model multivariate (temperatura màxima i mínima i humitat dels 14 dies anteriors juntament amb variable extreta de PCA de pressió, sunsetTimeHour , cloudCover i weekday) one-step amb walk-forward augmentant train.

- Weekday, PCA (pressió, sunsetTimeHour): Creem una nova variable a partir de la reducció amb PCA de les variables pressió i sunsetTimeHour. Amb la variable reduïda juntament amb les variables consum, temperatures màxima i mínima, humitat i weekday amb les dades de 14 dies anteriors creem el model.

Obtenim un RMSE de 0.03130.

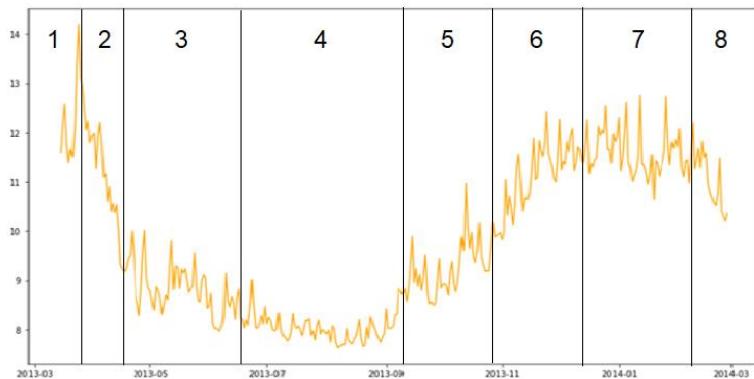
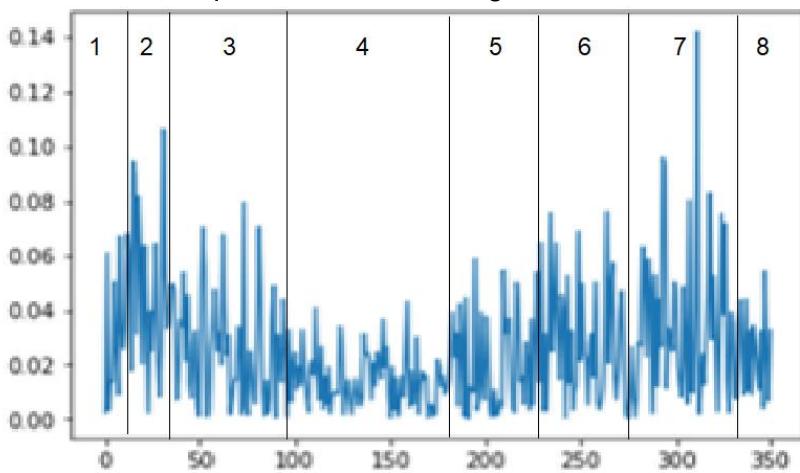


Error mitjà percentual de 2.82%



Gràfic 133 Percentatge error predicción segons valor real model multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors juntament amb variable extreta de PCA de pressió i sunsetTimeHour) one-step amb walk-forward augmentant train.

En resum, la millor predicció l'hem obtingut amb el model que utilitzava els 14 dies anteriors de les variables consum, temperatura màxima i mínima, humitat i weekday amb el walkforward on train va augmentant a mesura que avancen les prediccions. Hem dividit els errors d'aquest model en 8 franges:



Gràfic 134 Comparació errors predicción amb el moment de l'error millor model

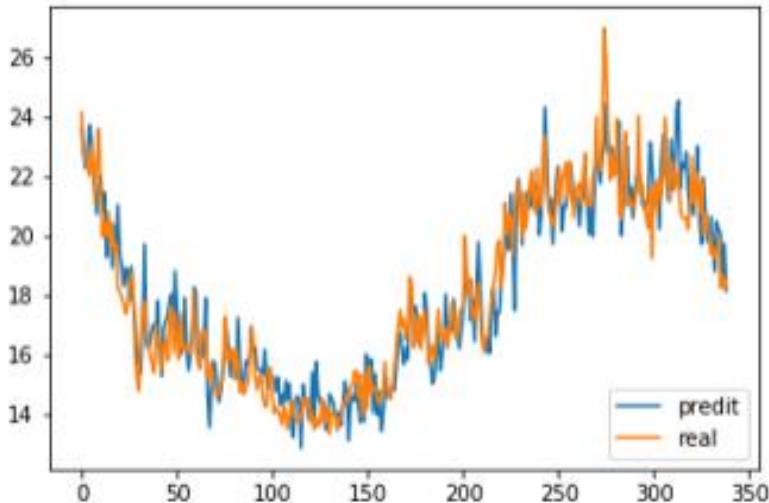
Veiem que hi ha vàries tendències/patrons d'error marcades, i que el principi o fi de cada tendència coincideix amb el canvi de tendència/patró del consum. En les franges 2 i 3 hi ha primavera, a la franja 4 tenim l'estiu, franges 5 i 6 tardor, i franges 7,8 i 1 hivern. A l'hivern és quan més variabilitat hi ha de consum i això es reflecteix en les prediccions, ja que a l'hivern és quan hi ha errors més elevats.

Veiem que els errors de l'estiu (franja 4) no superen el 4% d'error, a la tardor hi ha un màxim de 6-8% d'error i a la primavera pot arribar fins un 7-10% d'error. Com dèiem, a l'hivern hi ha uns errors màxims més elevats, al voltant del 8-10%, tenint un pic d'error d'un dia d'un 14%. Tot i haver-hi errors alts, veiem com hi ha molts errors que estan al voltant del 0-1%.

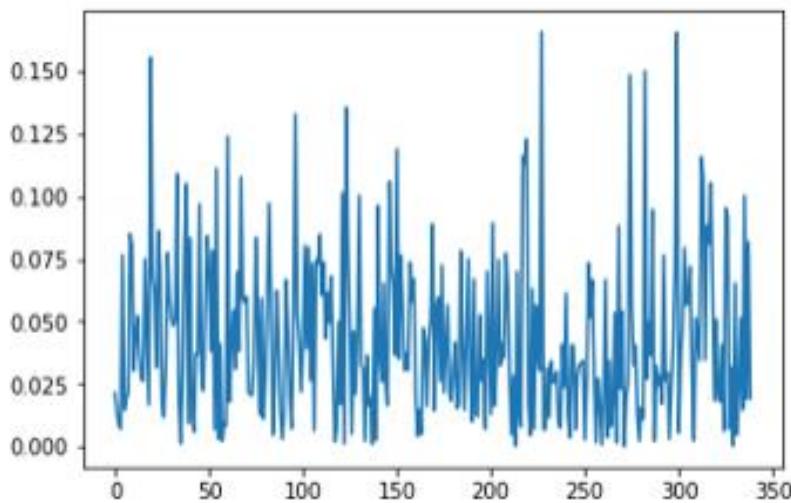
Aquest patró d'error el trobem repetit en tots els models.

En el millor model, que utilitzem la temperatura màxima i mínima, la humitat i el dia de la setmana amb 14 dies anteriors i el walkforward on el train va augmentant, provem de introduir-hi no totes les dades del consum total mitjà, sinó només les dades de un tipus de consumidor per mirar si s'adapta millor el model. Provarem amb les dades diàries de l'Acorn A.

Pensàvem que el model s'adaptaria bé a un grup de consumidors, però obtenim un RMSE de 0.04886 i un error de 4.31%, més alt que el model que conté la mitjana de totes les llars.



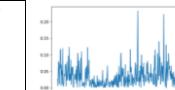
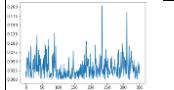
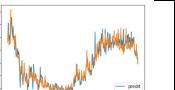
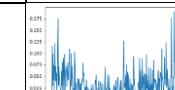
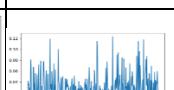
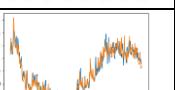
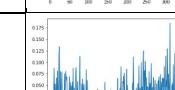
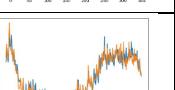
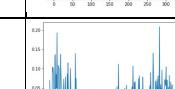
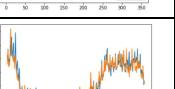
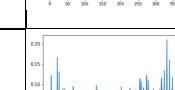
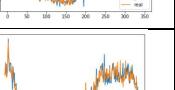
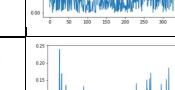
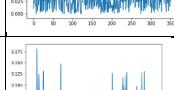
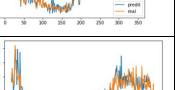
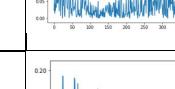
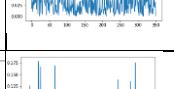
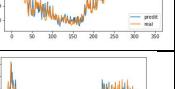
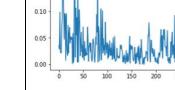
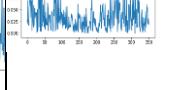
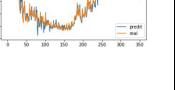
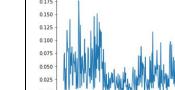
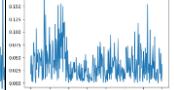
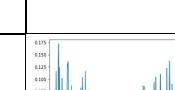
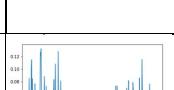
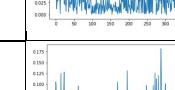
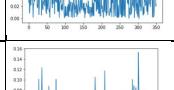
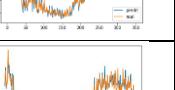
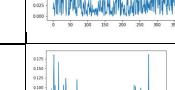
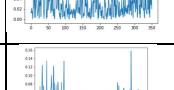
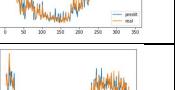
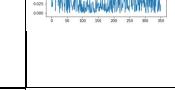
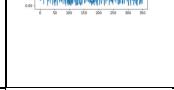
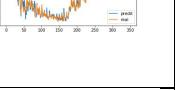
Gràfic 135 Predicció del dia següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train utilitzant tant sols les dades de Acorn A.

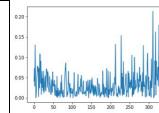
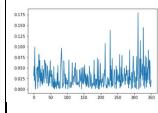
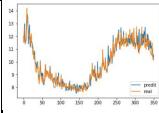
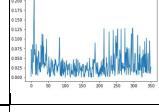
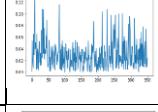
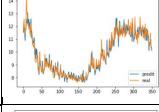
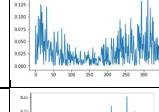
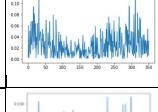
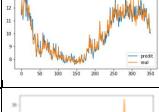
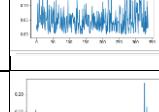
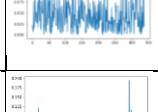
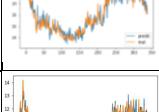
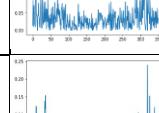
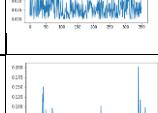
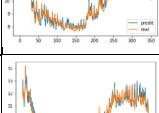
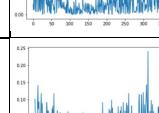
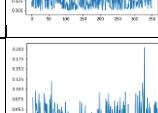
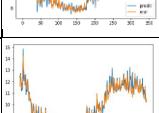
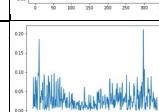
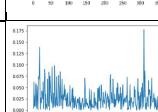
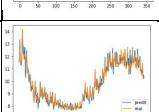
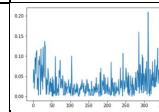
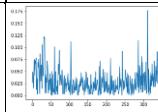
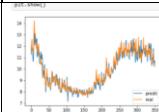
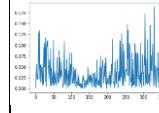
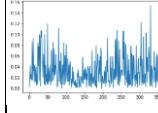
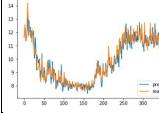
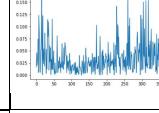
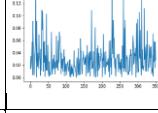
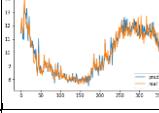
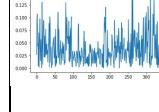
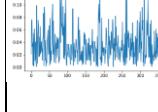
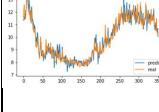
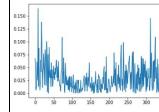
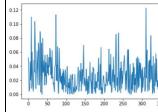
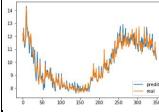
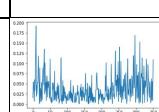
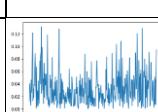
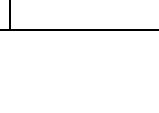
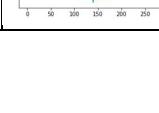


Gràfic 136 RMSE model multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant train utilitzant tant sols les dades de Acorn A.

Taula resum resultats models amb les variables utilitzades, on els 3 millors models estan marcats en verd:

	Variables	Dies passats	Walkforward	RMSE	Error mitjà	RMSE	Error	Valor real vs valor predit
M1		7	Walkforward 365 dies	0.0393				
M2	Temp. max	7	Walkforward 365 dies	0.0368	3.274 %			
M3	Temp. max, humitat	7	Walkforward 365 dies	0.0363				
M4	Temp. max, humitat, visibility, cloudCover	7	Walkforward 365 dies	0.0378				
M5	Temp max, humitat	14	Walkforward 365 dies	0.036				
M6	Temp max, humitat	28	Walkforward 365 dies	0.0363				
M7	Temp max, temp min, humitat	14	Walkforward 365 dies	0.03393	3.009 3%			

M8	Temp max, temp min	14	Walkforward 365 dies	0.0359	3.187 %			
M9	Temp max, temp min, humitat, sunsetTimeHour	14	Walkforward 365 dies	0.03459	3.091 %			
M10	Temp max, temp min, humitat, pressure	14	Walkforward 365 dies	0.03389	3.024 %			
M11	Temp max, temp min, humitat, dewPoint	14	Walkforward 365 dies	0.03438	3.062 %			
M12	Temp max, temp min, humitat, weekday	14	Walkforward 365 dies	0.03079	2.716 1%			
M13	Temp max, temp min, humitat, cloudCover	14	Walkforward 365 dies	0.03741	3.345 4%			
M14	Temp max, temp min, humitat, pressure, sunsetTimeHour	14	Walkforward 365 dies	0.03696	3.274 9%			
M15	Temp max, temp min, humitat, pressure, cloudCover	14	Walkforward 365 dies	0.03946	3.486 6%			
M16	Temp max, temp min, humitat, pressure, sunsetTimeHour, cloudCover	14	Walkforward 365 dies	0.03957	3.586 6%			
M17	Temp max, temp min, humitat, pressure ,weekday	14	Walkforward 365 dies	0.03171	2.818 %			
M18	Temp max, temp min, weekday	14	Walkforward 365 dies	0.03083	2.759 %			
M19	Temp max, temp min, humitat, pressure, sunsetTimeHour,weekd ay	14	Walkforward 365 dies	0.03134	2.785 %			
M20	Temp max, temp min, humitat	14	Walkforward + 365 dies (augmentant)	0.03598	3.203 8%			
M21	Temp max, temp min, humitat, cloudCover	14	Walkforward + 365 dies (augmentant)	0.03915	3.464 9%			

M22	Temp max, temp min, humitat, sunsetTimeHour	14	Walkforward + 365 dies (augmentant)	0.03299	2.920 9%			
M23	Temp max, temp min, humitat, pressure	14	Walkforward + 365 dies (augmentant)	0.03402	3.006 6%			
M24	Temp max, temp min, humitat, weekday	14	Walkforward + 365 dies (augmentant)	0.02920	2.549 1%			
M24A	Temp max, temp min, humitat, weekday Només Acorn A	14	Walkforward + 365 dies (augmentant)	0.04886	4.31%			
M25	Temp max, temp min, humitat, weekday MONTH	14	Walkforward + 365 dies (augmentant)	0.03317	2.988 %			
M26	Temp max, temp min, humitat, weekday, season	14	Walkforward + 365 dies (augmentant)	0.03213	2.868 %			
M27	Temp max, temp min, humitat, weekday, season, month	14	Walkforward + 365 dies (augmentant)	0.03250	2.879 %			
M28	Temp max, temp min, weekday	14	Walkforward + 365 dies (augmentant)	0.02972	2.664 %			
M29	Temp max, temp min, humitat, pressure, weekday	14	Walkforward + 365 dies (augmentant)	0.03283	2.893 %			
M30	Temp max, temp min, humitat, pressure, cloudCover	14	Walkforward + 365 dies (augmentant)	0.03724	3.310 2%			
M31	Temp max, temp min, humitat, pressure, sunsetTimeHour	14	Walkforward + 365 dies (augmentant)	0.03398	3.024 9%			
M32	Temp max, temp min, humitat, pressure, sunsetTimeHour, cloudCover	14	Walkforward + 365 dies (augmentant)	0.03896	3.529 5%			
M33	Temp max, temp min, humitat, pressure, sunsetTimeHour, weekday	14	Walkforward + 365 dies (augmentant)	0.02998	2.673 %			
M34	Temp max, temp min, humitat, pressió (Canviant optimitzador)	14	Walkforward 365 dies	0.03665	3.246 %			

M35	Temp max, temp min, humitat, PCA(pressió, sunsetTimeHour, cloudCover)	14	Walkforward + 365 dies (augmentant)	0.03331	2.928 %	
M36	Temp max, temp min, humitat, weekday, PCA(pressió, sunsetTimeHour, cloudCover)	14	Walkforward + 365 dies (augmentant)	0.03457	3.048 %	
M37	Temp max, temp min, humitat, PCA(weekday, pressió, sunsetTimeHour, cloudCover)	14	Walkforward + 365 dies (augmentant)	0.03682	3.286 %	
M38	Temp max, temp min, humitat, weekday, PCA(pressure, sunsetTimeHour)	14 dies	Walkforward + 365 dies (augmentant)	0.03130	2.820 %	

5.3.2 Predicció a més d'un dia (multistep)

Veient que la millor predicció l'hem obtingut a partir de la xarxa neuronal multivariate amb consum, temperatura màxima i mínima, humitat i weekday (amb les dades de 14 dies anteriors) i utilitzant el walkforward on train cada vegada és més gran, provem de predir més d'un dia amb aquest model. S'anomena multistep. La variació que aplicarem al model és que en la capa de sortida, enllloc de 1 neurona hi tindrem el nombre de neurones igual al nombre de dies per endavant que volem predir.

- Predicció 7 dies (a una setmana)

Obtenim un RMSE de 0.056288 pels 7 dies, on per cada dia és:

```
Error predicció 1 dia següent: 0.04440560194722621
Error predicció 2 dia següent: 0.05305027129342371
Error predicció 3 dia següent: 0.05878823185519177
Error predicció 4 dia següent: 0.06708710186404082
Error predicció 5 dia següent: 0.07098120810029249
Error predicció 6 dia següent: 0.07406943970167362
Error predicció 7 dia següent: 0.07479595370174597
```

imatge 17 RMSE

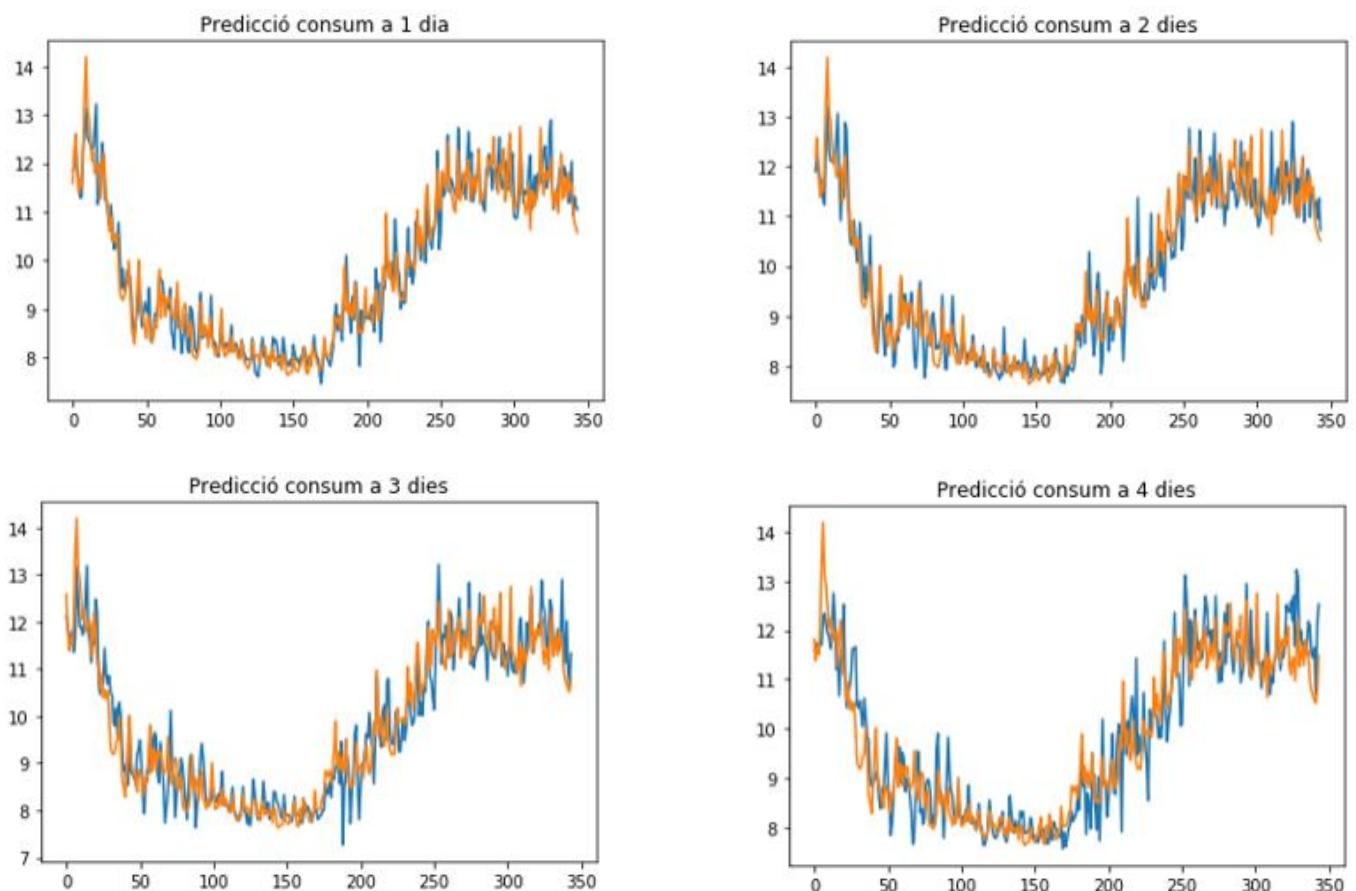
Veiem que, com era d'esperar, l'RMSE augmenta a mesura que la predicció va més enllà. Si ens fixem en la següent imatge amb l'error per dia, veiem com l'error de predicció pel dia següent té una mitjana d'un 3.06% i l'error per la predicció del setè dia següent és del 5.05%.

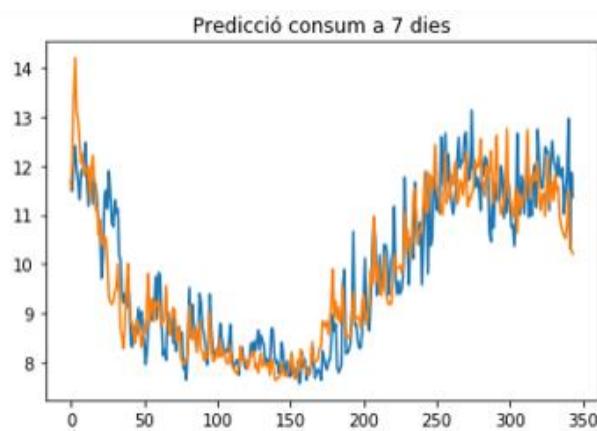
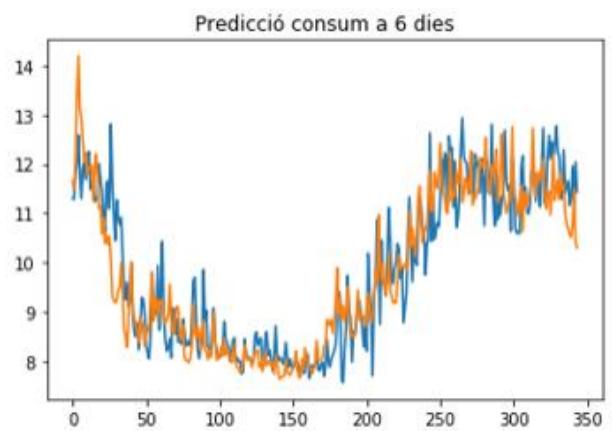
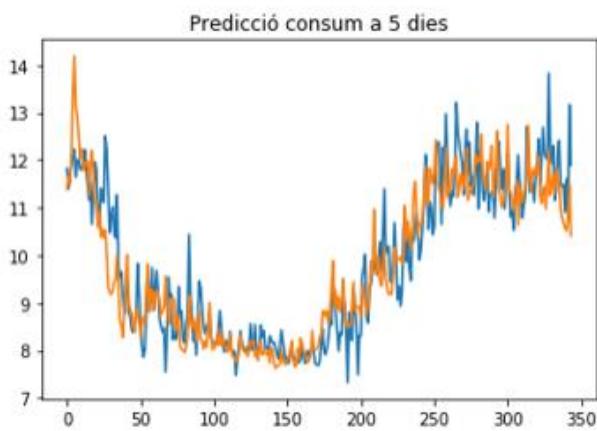
L'error mitjà a 1 dia és de 3.0568365749951054 %
 L'error mitjà a 2 dies és de 3.533745618463487 %
 L'error mitjà a 3 dies és de 4.002815889728186 %
 L'error mitjà a 4 dies és de 4.439177931557118 %
 L'error mitjà a 5 dies és de 4.804251124624361 %
 L'error mitjà a 6 dies és de 5.017627755303729 %
 L'error mitjà a 7 dies és de 5.053565397722889 %

Imatge 18 Error percentual

Tot i haver utilitzat el mateix model que el que dona millors resultats per a la predicció a un dia, veiem que l'error a un dia és més elevat en aquest model.

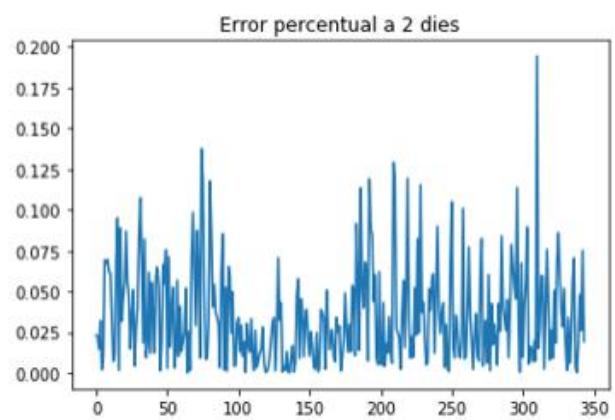
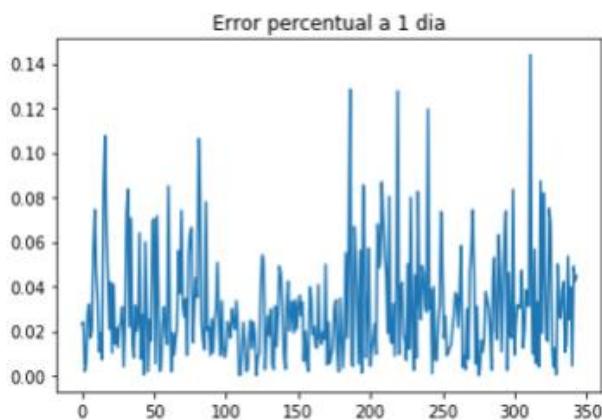
Vegem-ho gràficament:

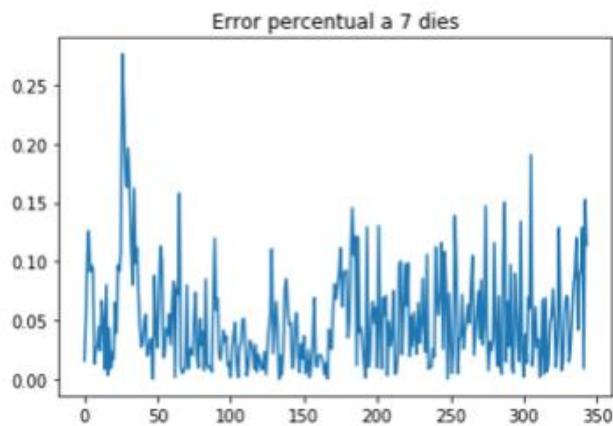
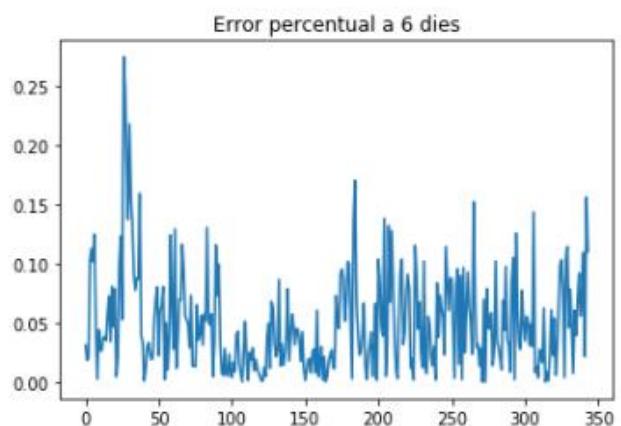
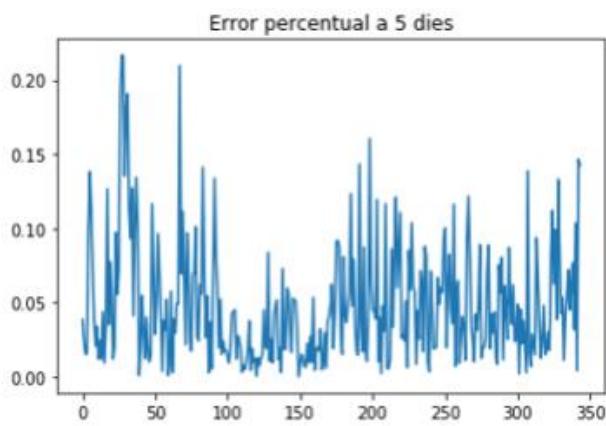
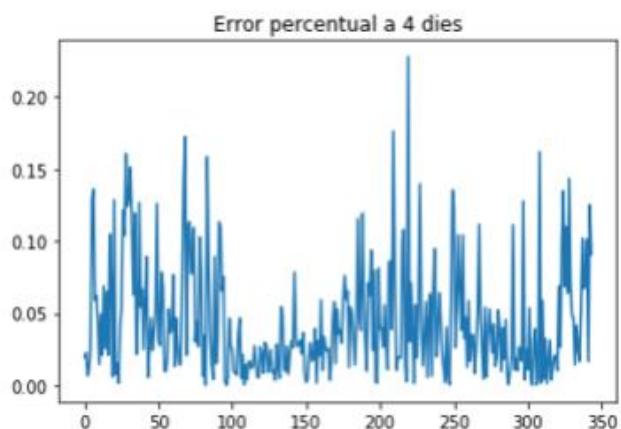
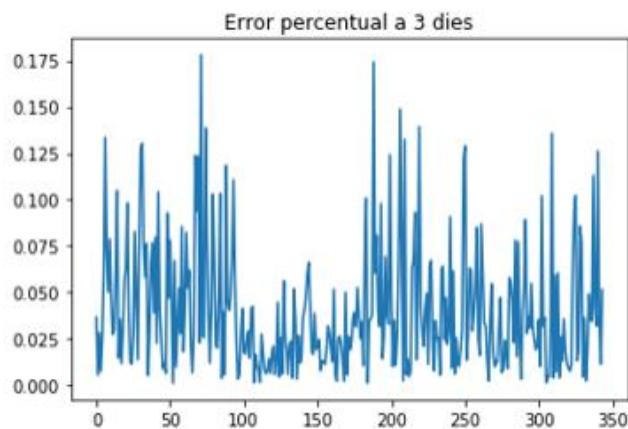




Gràfic 137 Predicccions dels 7 dies següents i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant

Error:





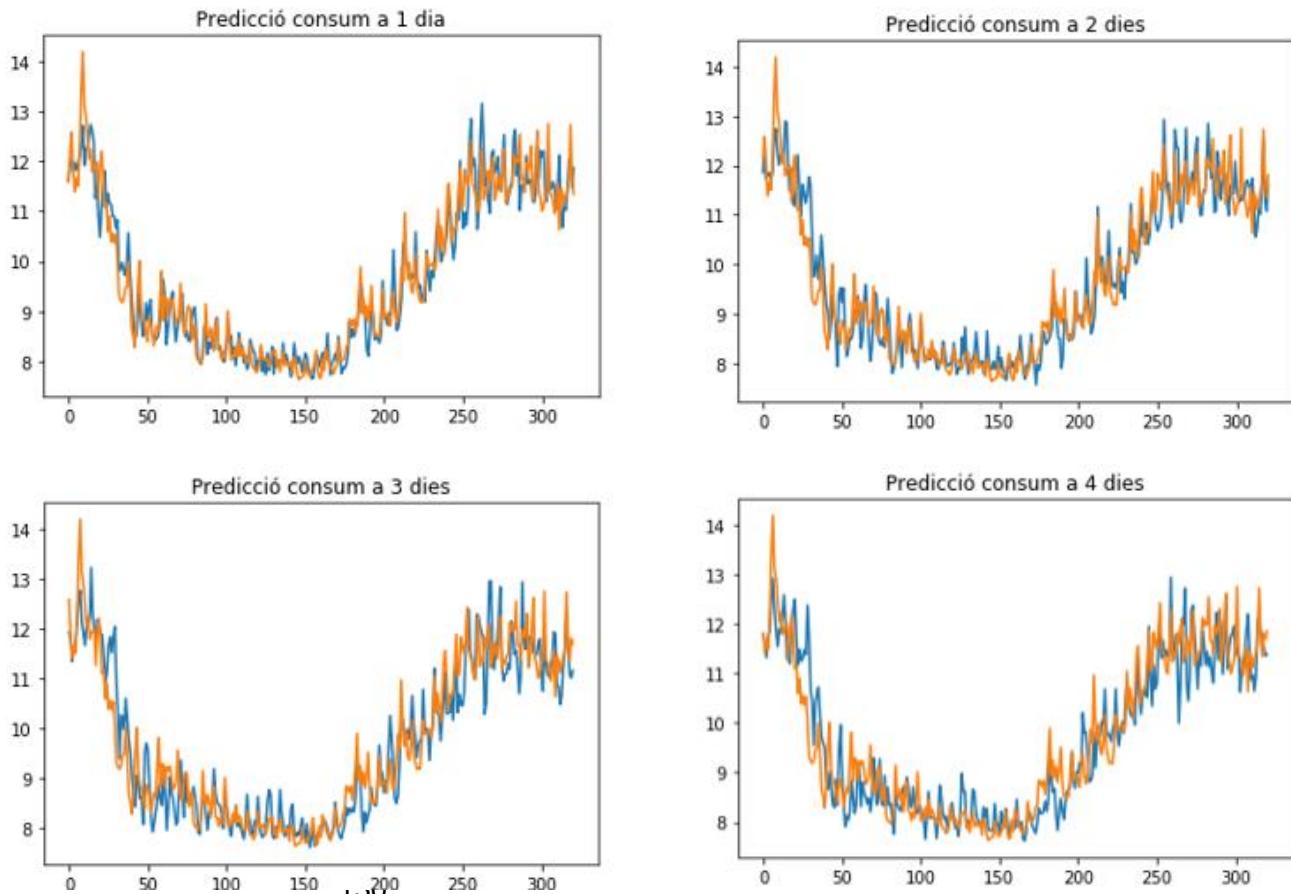
Gràfic 138 Errors model multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant. Predicció a 7 dies.

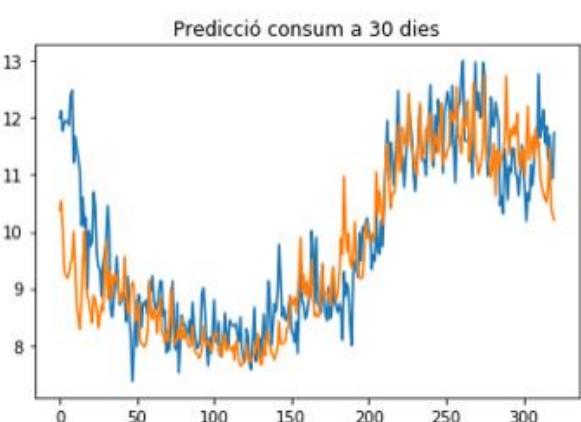
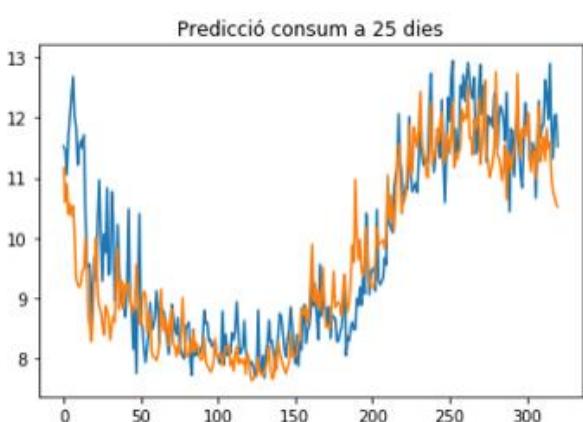
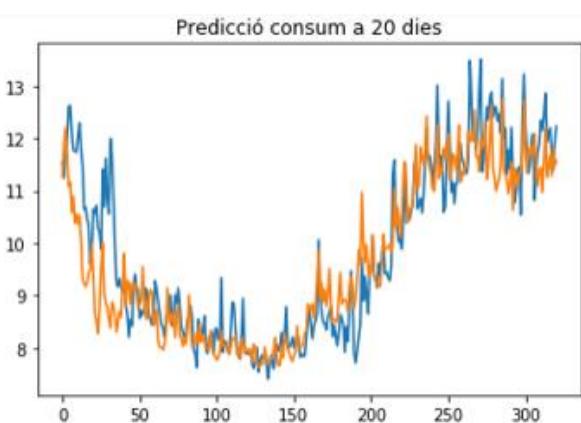
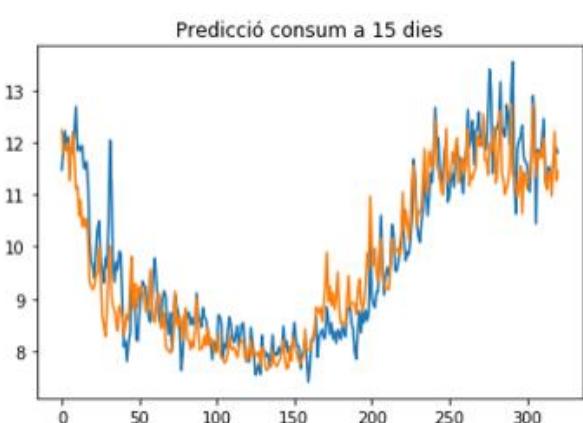
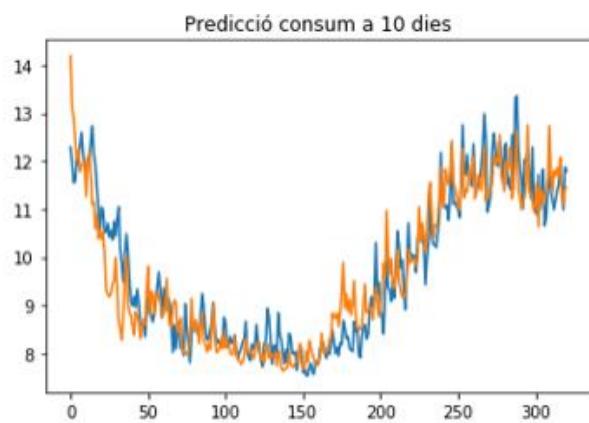
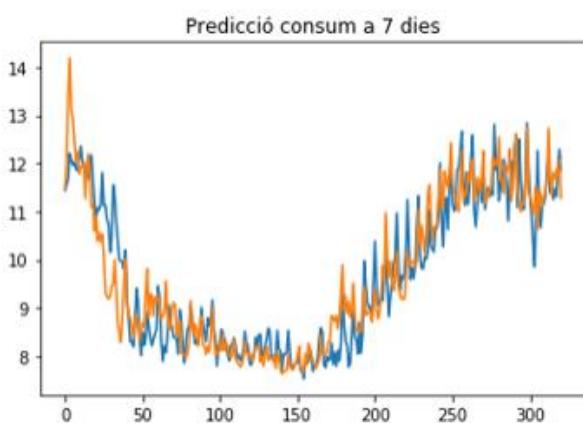
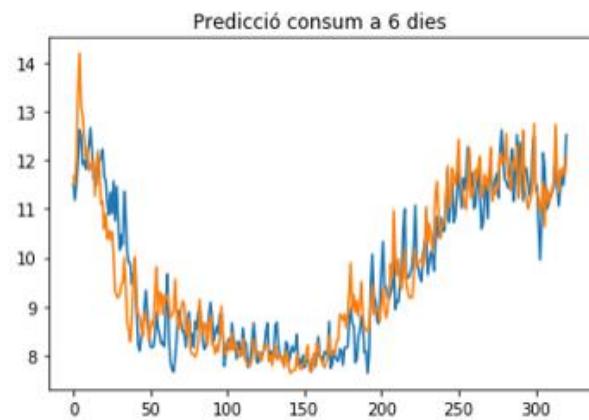
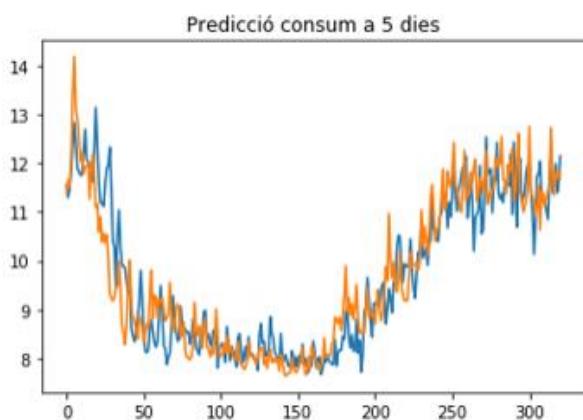
- Predicció 30 dies (a un mes)

Obtenim un RMSE de 0.06336 pels 30 dies, on per cada dia és :

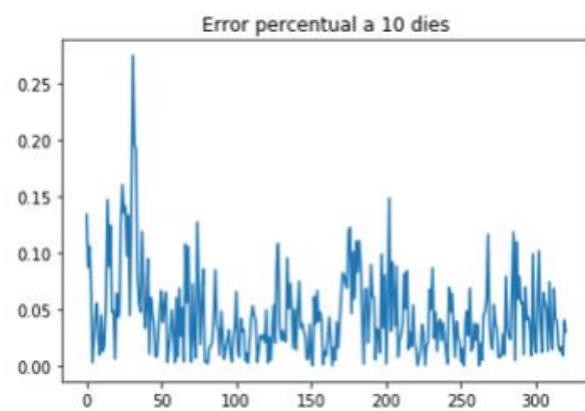
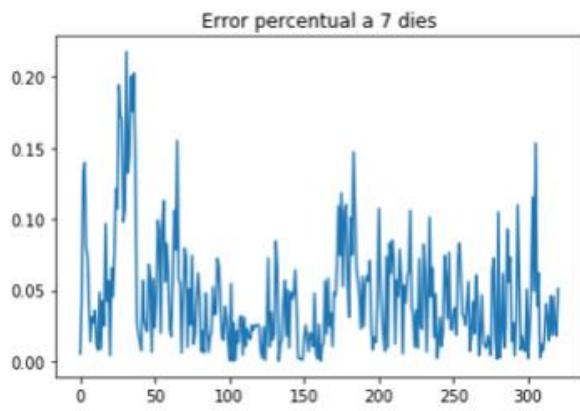
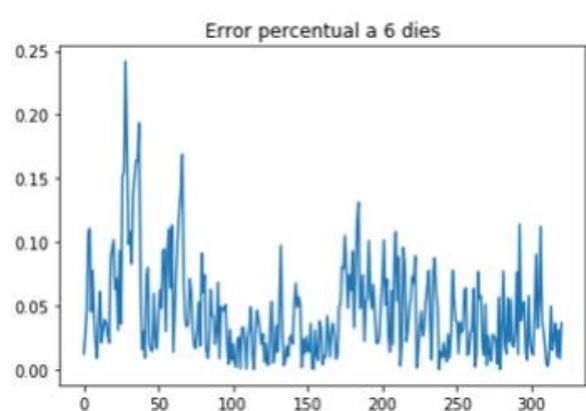
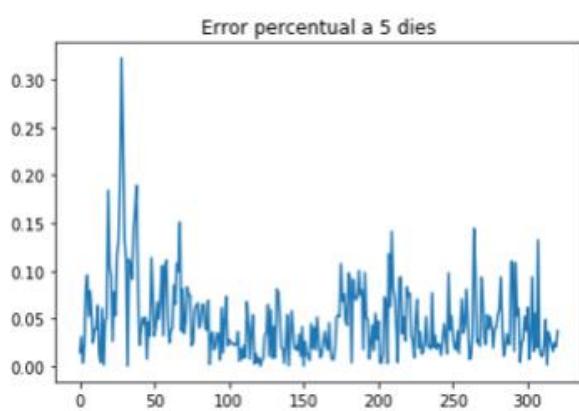
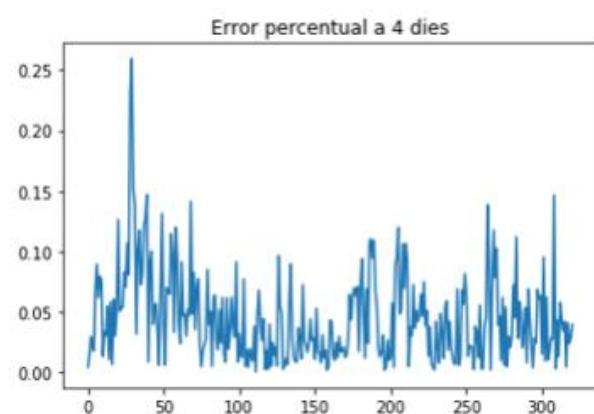
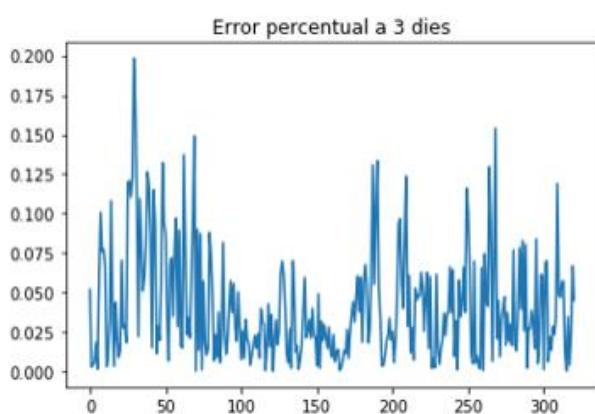
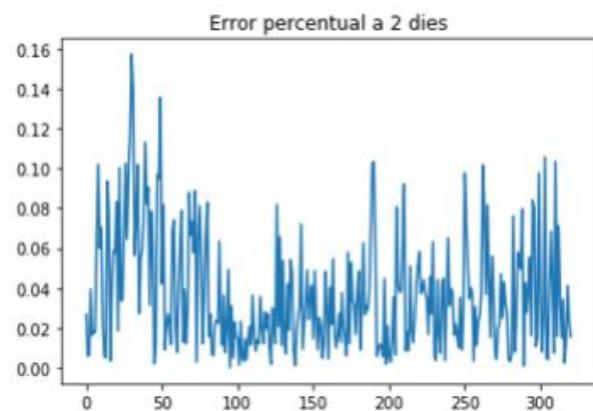
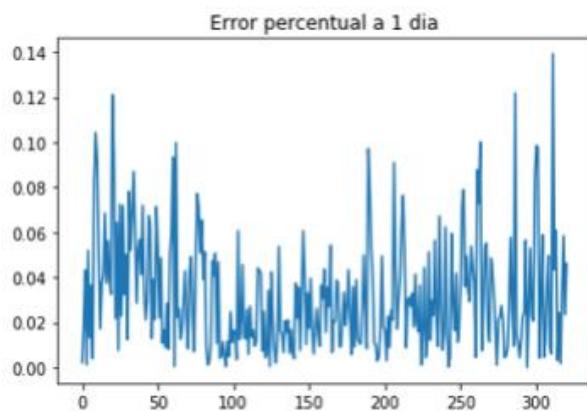
```
Error predicción 1 dia següent: 0.047713475502803276
Error predicción 2 dia següent: 0.05334677125316044
Error predicción 3 dia següent: 0.06105267280071678
Error predicción 4 dia següent: 0.06456381451782518
Error predicción 5 dia següent: 0.06856752725109426
Error predicción 6 dia següent: 0.06379745139010491
Error predicción 7 dia següent: 0.06577296181561096
Error predicción 8 dia següent: 0.06328989459823108
Error predicción 9 dia següent: 0.06122628399775239
Error predicción 10 dia següent: 0.06412780628932768
Error predicción 11 dia següent: 0.0625173249896466
Error predicción 12 dia següent: 0.06445061541670374
Error predicción 13 dia següent: 0.06923830537711873
Error predicción 14 dia següent: 0.06769823228481564
Error predicción 15 dia següent: 0.06650784420586008
Error predicción 16 dia següent: 0.06954945689190016
Error predicción 17 dia següent: 0.07204898652168114
Error predicción 18 dia següent: 0.07331459177650745
Error predicción 19 dia següent: 0.07687798756444528
Error predicción 20 dia següent: 0.08455152716041575
Error predicción 21 dia següent: 0.08543587906060066
Error predicción 22 dia següent: 0.08306432296521903
Error predicción 23 dia següent: 0.07742337668520707
Error predicción 24 dia següent: 0.07768617764032275
Error predicción 25 dia següent: 0.08094524735123418
Error predicción 26 dia següent: 0.08726743943351789
Error predicción 27 dia següent: 0.0868921352453962
Error predicción 28 dia següent: 0.08628596014469317
Error predicción 29 dia següent: 0.08662548745096152
Error predicción 30 dia següent: 0.09317324678772622
```

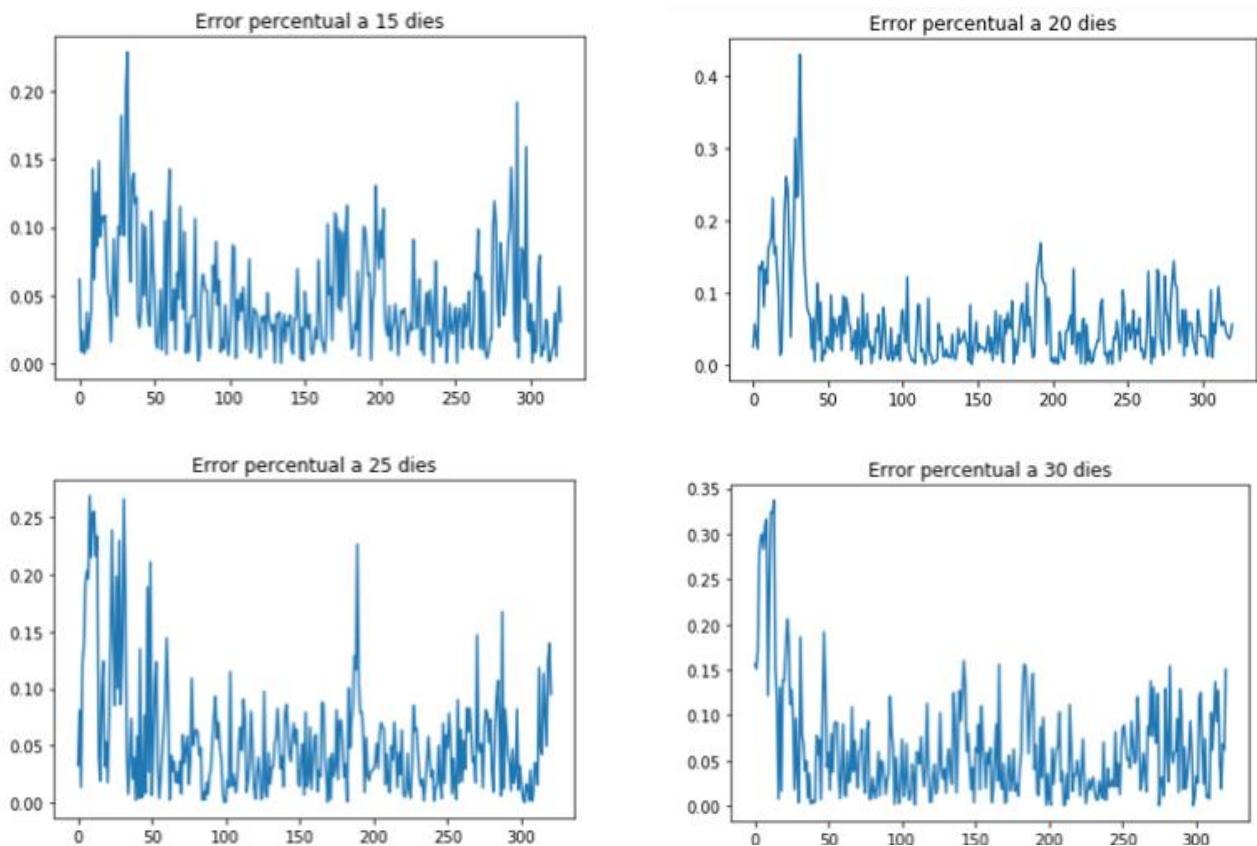
Imatge 19 RMSE predicción a 30 dies.





Gràfic 139 Prediccions dels 30 dies següents i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant





Gràfic 140 Errors model multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant. Predicció a 30 dies

L'error mitjà a 1 dia és de 3.201301904539095 %
 L'error mitjà a 2 dies és de 3.695479031626149 %
 L'error mitjà a 3 dies és de 4.13242978174084 %
 L'error mitjà a 4 dies és de 4.4861972085003625 %
 L'error mitjà a 5 dies és de 4.741936371808464 %
 L'error mitjà a 6 dies és de 4.515537597028588 %
 L'error mitjà a 7 dies és de 4.502326551129249 %
 L'error mitjà a 10 dies és de 4.492387138130774 %
 L'error mitjà a 15 dies és de 4.736990140327543 %
 L'error mitjà a 20 dies és de 5.587884468908786 %
 L'error mitjà a 25 dies és de 5.479946538037582 %
 L'error mitjà a 30 dies és de 6.337070606244632 %

Imatge 20 Errors model Prediccions dels 30 dies següents i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant

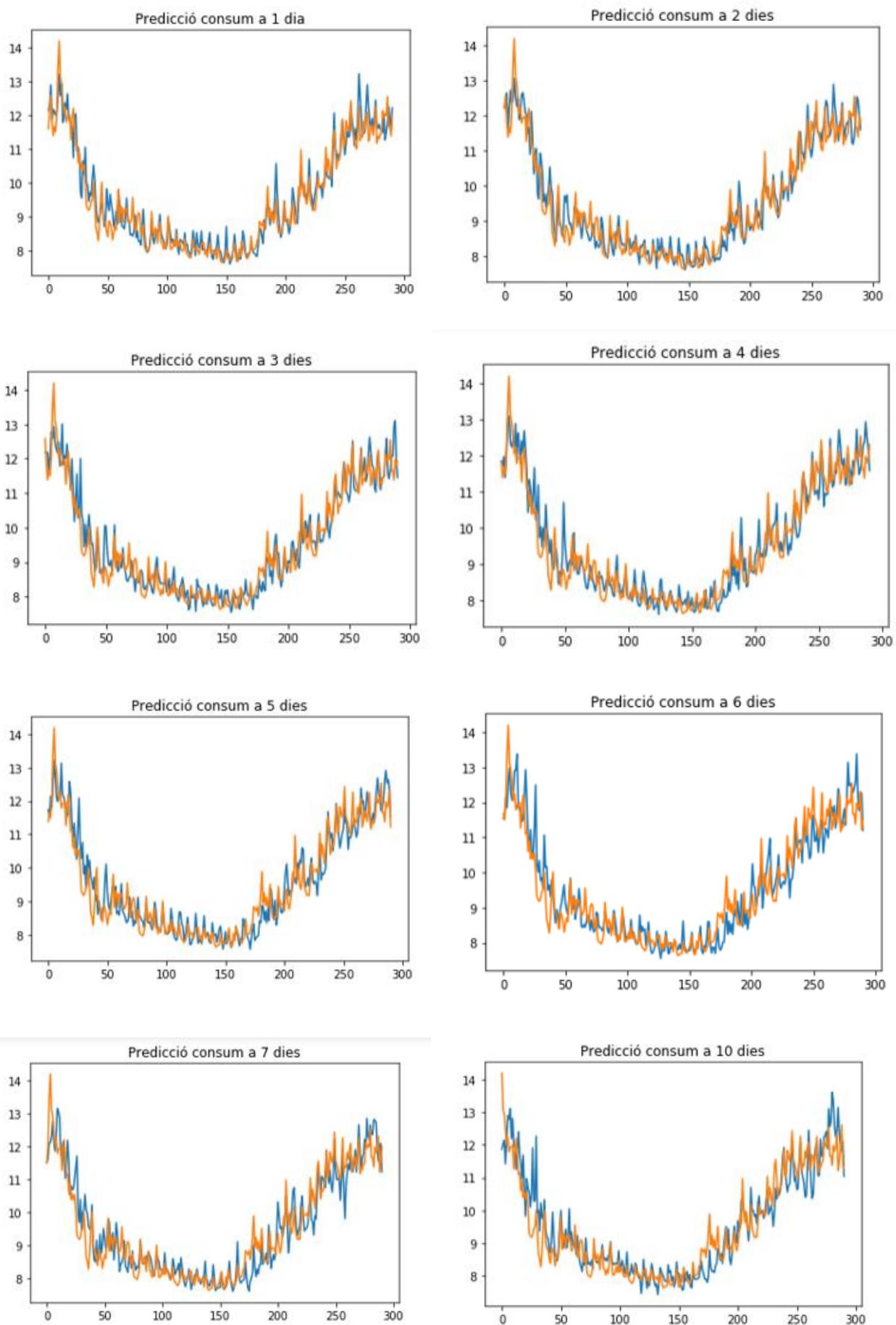
Veiem com l'error mitjà va augmentant a mesura que es fa una predició més enllà, sent del 3.20% a un dia, del 4.49% a 10 dies, del 5.58% a 20 dies i de 6.33% a 30 dies (un mes). L'error a un dia és més alt que en el model on només predim a 7 dies vista.

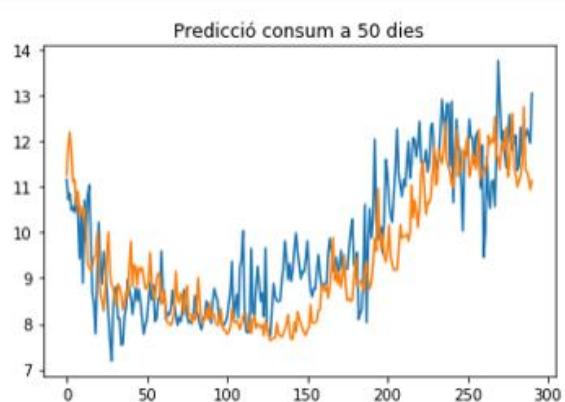
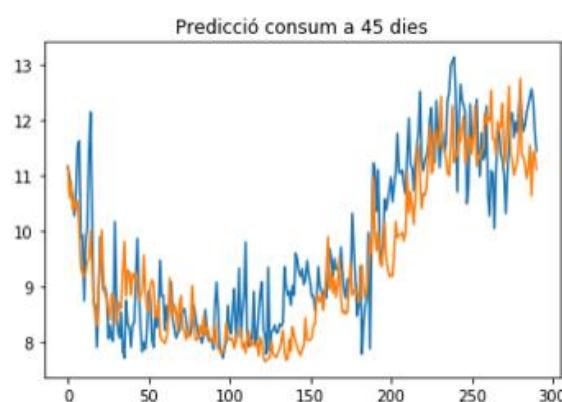
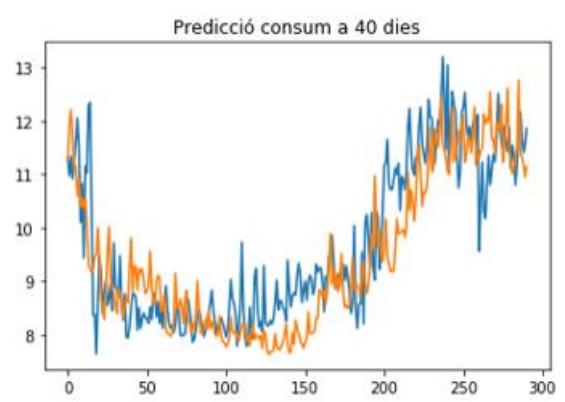
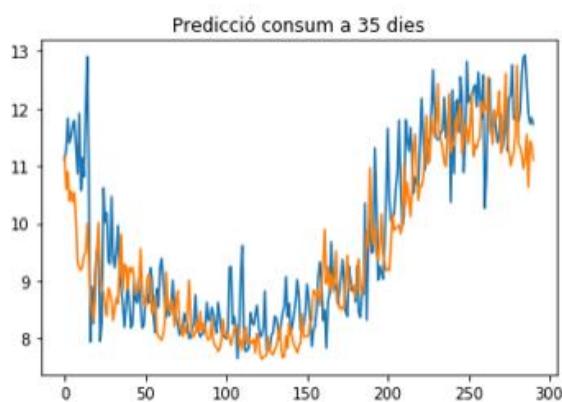
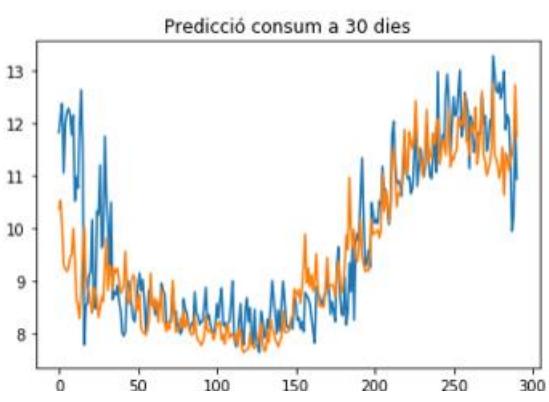
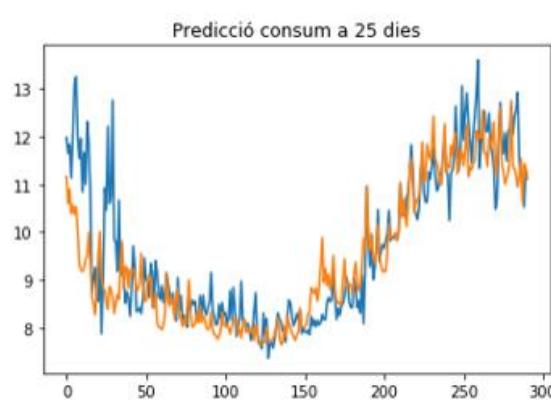
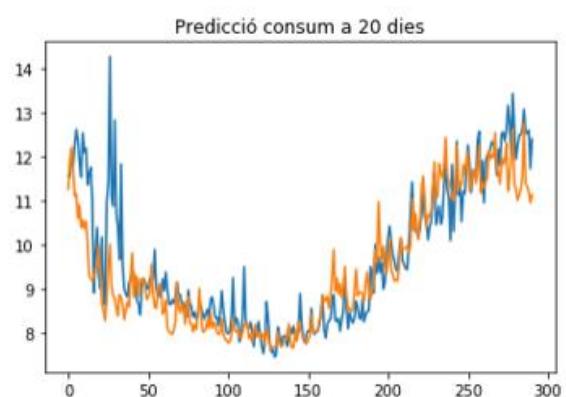
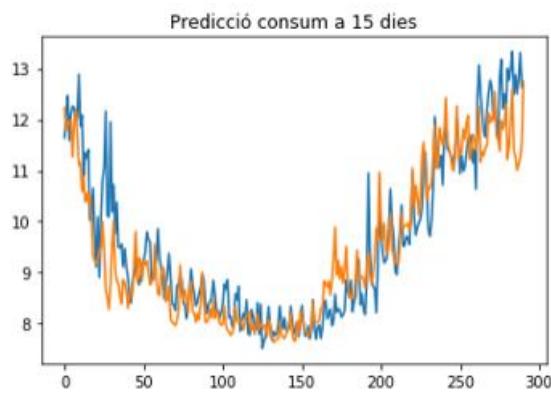
- Predicció 60 dies (a dos mesos)

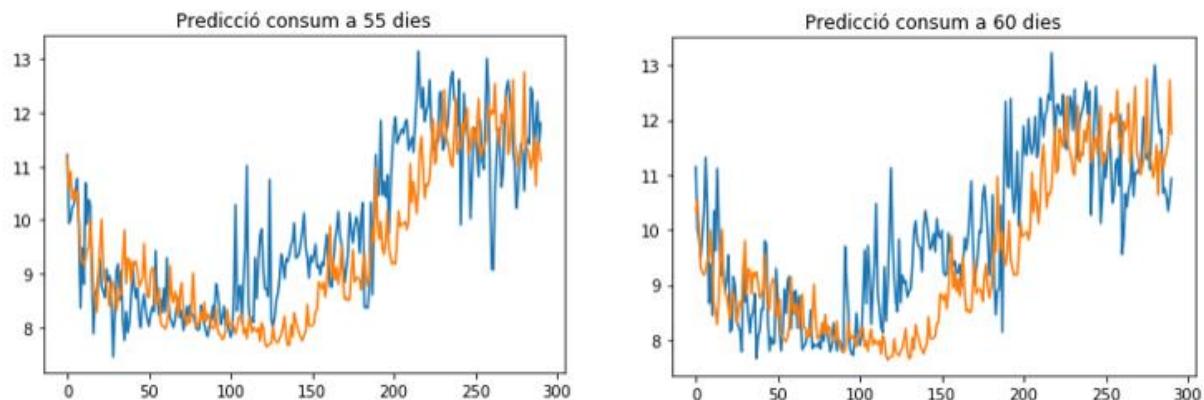
RMSE mitjà dels 60 dies 0.070878

```
Error predicció 1 dia següent: 0.04554174800183954
Error predicció 2 dia següent: 0.04874270774416747
Error predicció 3 dia següent: 0.058902349316397595
Error predicció 4 dia següent: 0.05212259374675393
Error predicció 5 dia següent: 0.055651889339802196
Error predicció 6 dia següent: 0.06072442898485177
Error predicció 7 dia següent: 0.062116660019193366
Error predicció 8 dia següent: 0.06062435411701553
Error predicció 9 dia següent: 0.06292869324902442
Error predicció 10 dia següent: 0.06811261199768653
Error predicció 11 dia següent: 0.06765964476845497
Error predicció 12 dia següent: 0.06669606282954284
Error predicció 13 dia següent: 0.07077233511928248
Error predicció 14 dia següent: 0.07288828274078953
Error predicció 15 dia següent: 0.07476190749844664
Error predicció 16 dia següent: 0.07560852105057292
Error predicció 17 dia següent: 0.0844046207888959
Error predicció 18 dia següent: 0.0882887867151327
Error predicció 19 dia següent: 0.09036555811032823
Error predicció 20 dia següent: 0.08804968510772046
Error predicció 21 dia següent: 0.08594535137863074
Error predicció 22 dia següent: 0.08737583623242912
Error predicció 23 dia següent: 0.09062524860568202
Error predicció 24 dia següent: 0.09053227984773812
Error predicció 25 dia següent: 0.09034830304778812
Error predicció 26 dia següent: 0.09311310782806986
Error predicció 27 dia següent: 0.094427409689534
Error predicció 28 dia següent: 0.09598833102444151
Error predicció 29 dia següent: 0.09684883446295388
Error predicció 30 dia següent: 0.09919502548586263
Error predicció 31 dia següent: 0.09664831948104185
Error predicció 32 dia següent: 0.09587492411655841
Error predicció 33 dia següent: 0.09441682392097041
Error predicció 34 dia següent: 0.09572420979993616
Error predicció 35 dia següent: 0.09104415262953595
Error predicció 36 dia següent: 0.09107620987059303
Error predicció 37 dia següent: 0.0934511536564005
Error predicció 38 dia següent: 0.0912182476619563
Error predicció 39 dia següent: 0.0882273215942756
Error predicció 40 dia següent: 0.08886177466410716
Error predicció 41 dia següent: 0.08680421134728288
Error predicció 42 dia següent: 0.08894176976988748
Error predicció 43 dia següent: 0.08789205554532034
Error predicció 44 dia següent: 0.08836877397953237
Error predicció 45 dia següent: 0.08329205216397706
Error predicció 46 dia següent: 0.08263968138064384
Error predicció 47 dia següent: 0.08247183259234075
Error predicció 48 dia següent: 0.0833310506962542
Error predicció 49 dia següent: 0.08384997353056574
Error predicció 50 dia següent: 0.08676269438515125
Error predicció 51 dia següent: 0.08689362927741899
Error predicció 52 dia següent: 0.08652880154816349
Error predicció 53 dia següent: 0.08645283518354474
Error predicció 54 dia següent: 0.0858665760936956
Error predicció 55 dia següent: 0.08936058420523006
Error predicció 56 dia següent: 0.08916086288526218
Error predicció 57 dia següent: 0.08908628872692025
Error predicció 58 dia següent: 0.09086623519440076
Error predicció 59 dia següent: 0.09075403845155135
Error predicció 60 dia següent: 0.09068754042445028
```

imatge 21 RMSE 60 dies

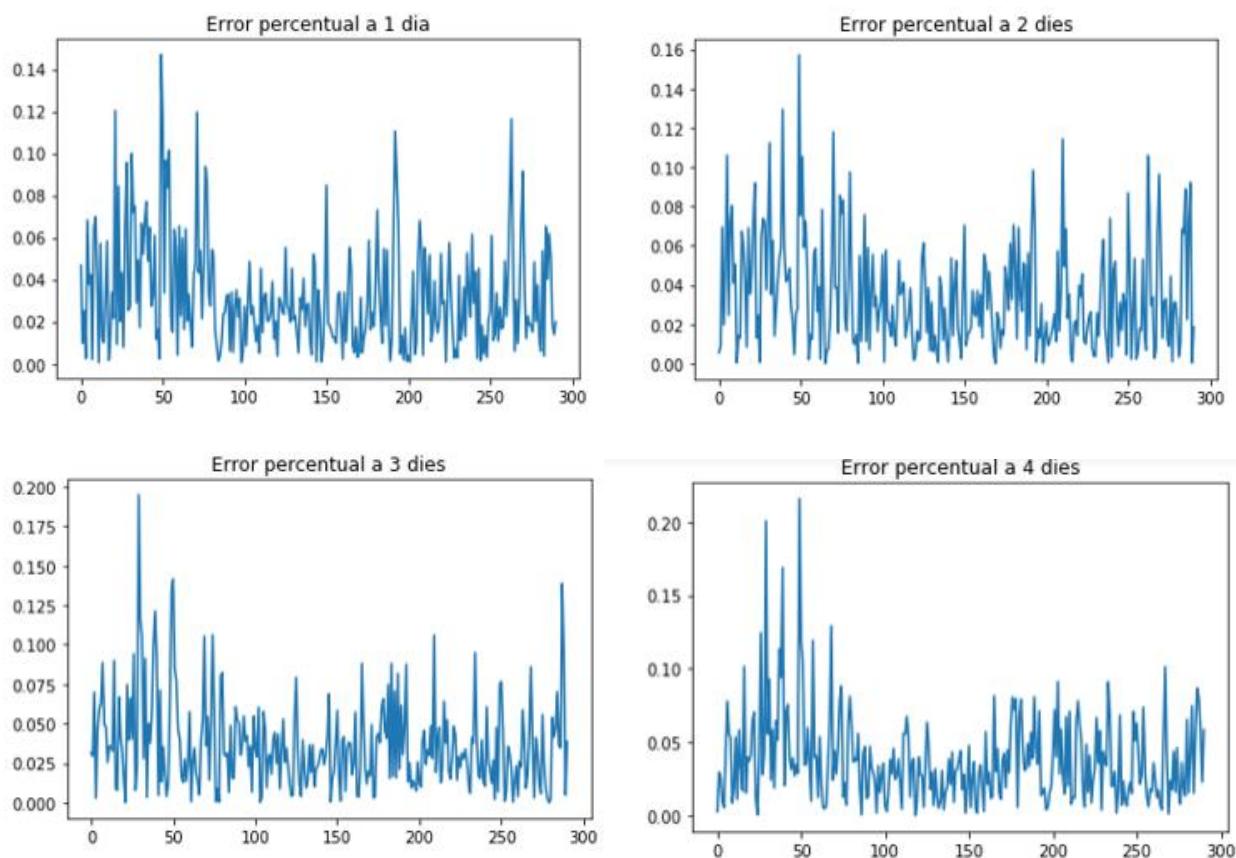


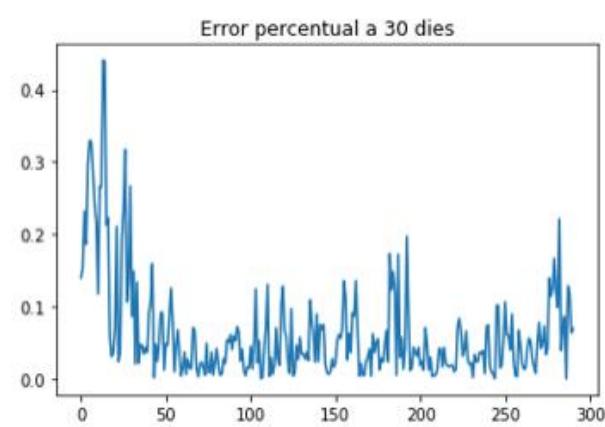
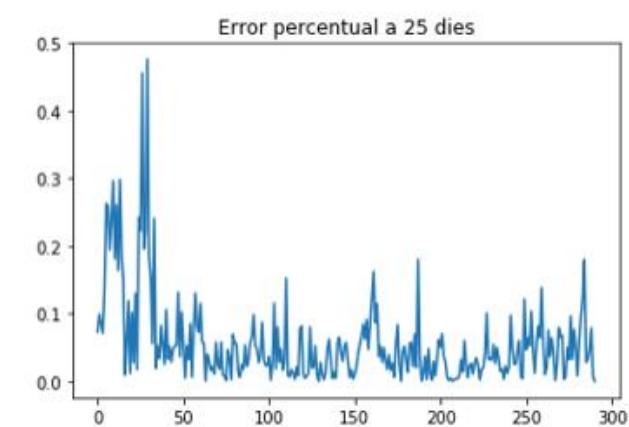
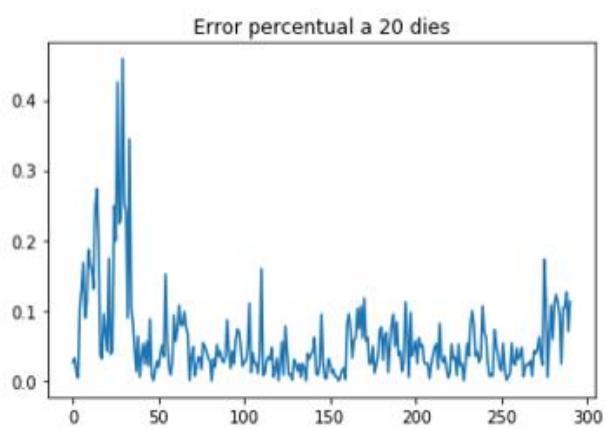
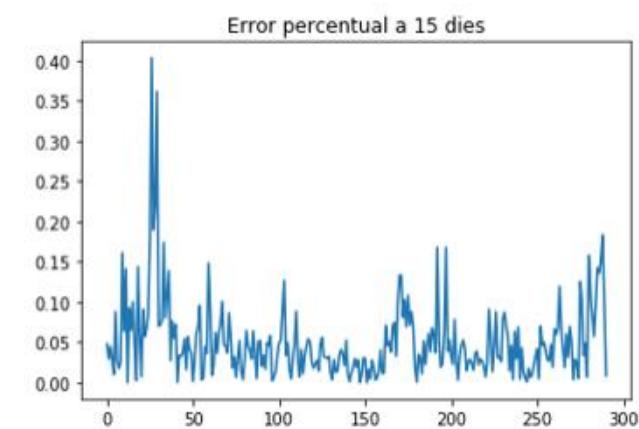
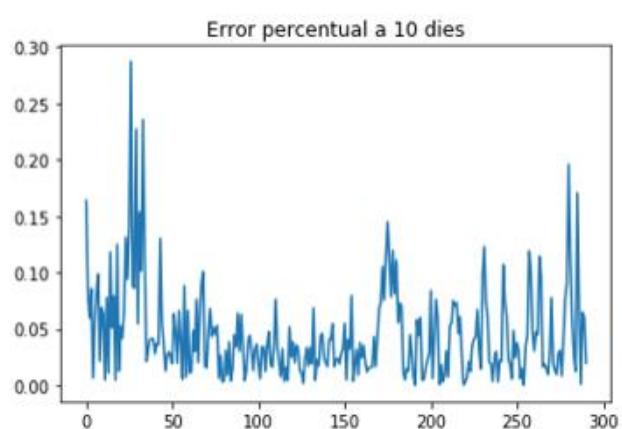
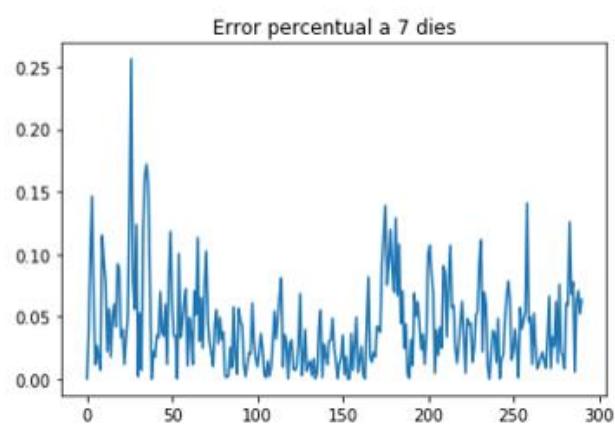
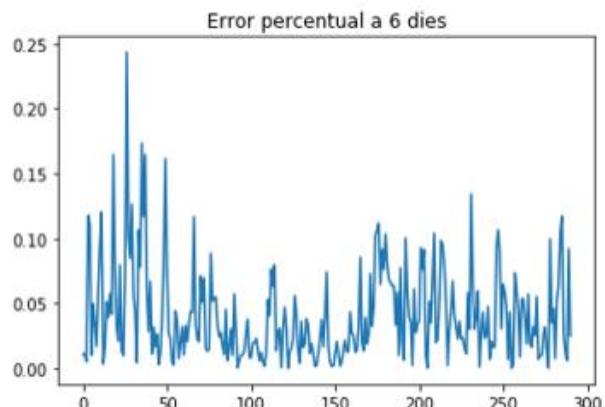
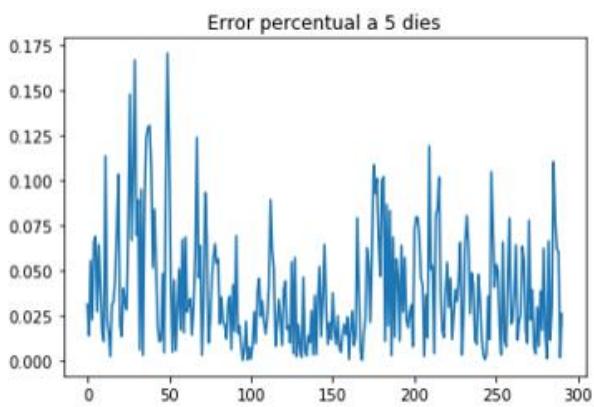


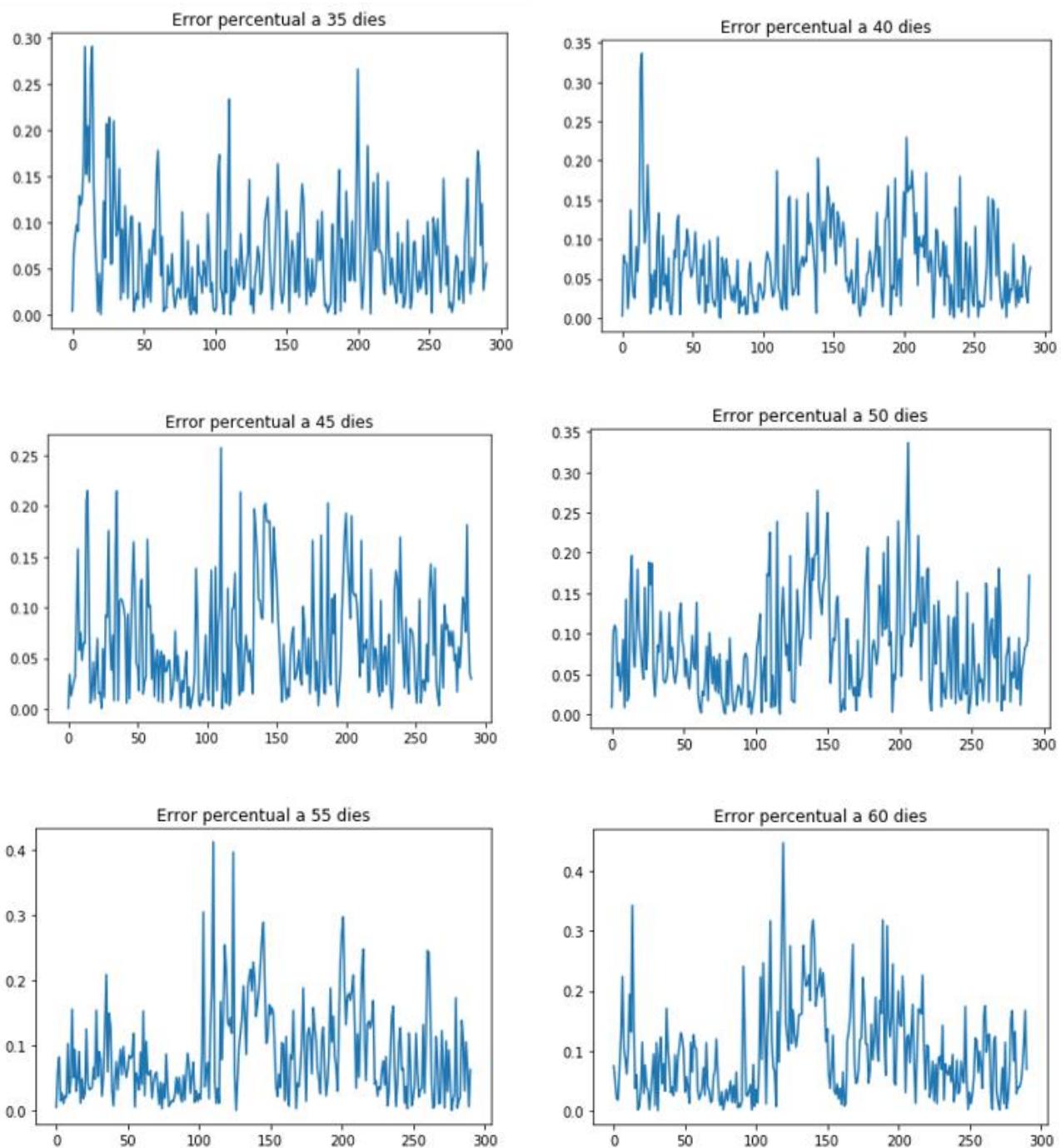


Gràfic 141 Prediccions dels 60 dies següents i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant

Veiem com els primers dies fa una predicció acurada però a partir dels 40 dies la predicció està més lluny que a prop de la realitat. Ens fixem que al contrari que en les prediccions a un dia, a on es falla més és a l'estiu i no a l'hivern.







Gràfic 142 Errors model multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant. Predicció a 60 dies

En el gràfic dels errors es veu clarament el fet de que a mesura que es prediu més enllà en el temps, s'adapta pitjor a les dades d'estiu i les d'hivern les prediu aproximadament amb uns errors semblants.

En la següent imatge veiem com l'error a un dia és del 3.22%, a 10 dies del 4.51%, a 20 dies del 6.60%, a 30 dies del 6.35%, a 40 dies del 6.74, a 50 dies a 8.29% i a 60 dies del 9.71%.

L'error mitjà a 1 dia és de 3.2269765896675553 %
L'error mitjà a 2 dies és de 3.4439064063941855 %
L'error mitjà a 3 dies és de 3.6711913643319716 %
L'error mitjà a 4 dies és de 3.7829875192103812 %
L'error mitjà a 5 dies és de 3.951424789673647 %
L'error mitjà a 6 dies és de 4.120618795467848 %
L'error mitjà a 7 dies és de 4.297009147206339 %
L'error mitjà a 10 dies és de 4.513654726238839 %
L'error mitjà a 15 dies és de 5.049189639509716 %
L'error mitjà a 20 dies és de 5.604116131921684 %
L'error mitjà a 25 dies és de 5.708291719831363 %
L'error mitjà a 30 dies és de 6.352893822113277 %
L'error mitjà a 35 dies és de 6.564774591495934 %
L'error mitjà a 40 dies és de 6.74872864798812 %
L'error mitjà a 45 dies és de 6.720113138640084 %
L'error mitjà a 50 dies és de 8.294681290687343 %
L'error mitjà a 55 dies és de 8.577243606043028 %
L'error mitjà a 60 dies és de 9.71099008784458 %

Imatge 22 Errors model prediccions dels 60 dies següents i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima , humitat i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant

En els models que provarem a partir d'ara, només visualitzarem els millors models. Es poden comprovar els gràfics dels models que no són tant bons a l'annex corresponent del model.

Provarem un model que utilitzi les variables temperatura màxima i mínima i humitat (sense weekday) agafant els 14 dies anteriors amb walkforward on va augmentant train a mesura que avancen les prediccions.

- Predicció a 7 dies: Obtenim RMSE mitjà de 0.063718 i un error percentual mitjà del 4.81%.

RMSE:

Error predicció 1 dia següent: 0.06008873530580592
Error predicció 2 dia següent: 0.07327357748109317
Error predicció 3 dia següent: 0.07127651816786797
Error predicció 4 dia següent: 0.0749016743107488
Error predicció 5 dia següent: 0.07316006699906583
Error predicció 6 dia següent: 0.07164354700716427
Error predicció 7 dia següent: 0.07197292803048952

Imatge 23 RMSE predicció 7 dies

Error:

L'error mitjà a 1 dia és de 3.745392632777331 %
L'error mitjà a 2 dies és de 4.930147418662605 %
L'error mitjà a 3 dies és de 4.778695164991031 %
L'error mitjà a 4 dies és de 5.178522813644043 %
L'error mitjà a 5 dies és de 5.086857159731961 %
L'error mitjà a 6 dies és de 4.968317886291285 %
L'error mitjà a 7 dies és de 4.951264218245851 %

Imatge 24 Error predicció 7 dies

L'error mitjà és més alt que en el model anterior, però l'error a 7 dies és més baix en aquest model. De totes maneres, si el que volem és predir a una setmana, llavors necessitarem el menys error possible global.

- Predicció a 30 dies

Obtenim RMSE de 0.06972 i error de 5.27%

RMSE

```
Error predicció 1 dia següent: 0.056103689874360804
Error predicció 2 dia següent: 0.05710674233873536
Error predicció 3 dia següent: 0.059119409048586966
Error predicció 4 dia següent: 0.06111307970695543
Error predicció 5 dia següent: 0.06445491974106266
Error predicció 6 dia següent: 0.06322421069774836
Error predicció 7 dia següent: 0.06311594537060823
Error predicció 8 dia següent: 0.06611861566327414
Error predicció 9 dia següent: 0.06194482460230059
Error predicció 10 dia següent: 0.06263701870049584
Error predicció 11 dia següent: 0.0654830215460122
Error predicció 12 dia següent: 0.07028894697885268
Error predicció 13 dia següent: 0.06967064010916385
Error predicció 14 dia següent: 0.07124594400677113
Error predicció 15 dia següent: 0.07913951570606297
Error predicció 16 dia següent: 0.07636503652215111
Error predicció 17 dia següent: 0.07767325091136612
Error predicció 18 dia següent: 0.07762102673454513
Error predicció 19 dia següent: 0.07956504021842596
Error predicció 20 dia següent: 0.07491749989000919
Error predicció 21 dia següent: 0.07966580085786579
Error predicció 22 dia següent: 0.08402193844439133
Error predicció 23 dia següent: 0.08017889570793427
Error predicció 24 dia següent: 0.08606957491049597
Error predicció 25 dia següent: 0.08842463976330374
Error predicció 26 dia següent: 0.08862392055633181
Error predicció 27 dia següent: 0.0935944456476905
Error predicció 28 dia següent: 0.09507475153077723
Error predicció 29 dia següent: 0.10009366755628664
Error predicció 30 dia següent: 0.10312766484610622
```

Imatge 25 RMSE a 30 dies

Error

```
L'error mitjà a 1 dia és de 3.8289216740980048 %
L'error mitjà a 2 dies és de 3.8675230945750956 %
L'error mitjà a 3 dies és de 4.0981341791126935 %
L'error mitjà a 4 dies és de 4.247771944399659 %
L'error mitjà a 5 dies és de 4.395242776564957 %
L'error mitjà a 6 dies és de 4.381943825766524 %
L'error mitjà a 7 dies és de 4.414350741978539 %
L'error mitjà a 8 dies és de 4.461292656607943 %
L'error mitjà a 9 dies és de 4.180183765908975 %
L'error mitjà a 10 dies és de 4.336295171154531 %
L'error mitjà a 11 dies és de 4.630348359283441 %
L'error mitjà a 12 dies és de 4.857024089754767 %
L'error mitjà a 13 dies és de 4.838667795760833 %
L'error mitjà a 14 dies és de 5.012184786894556 %
L'error mitjà a 15 dies és de 5.5914046779207585 %
L'error mitjà a 16 dies és de 5.30236314089655 %
L'error mitjà a 17 dies és de 5.351456771020501 %
L'error mitjà a 18 dies és de 5.536209695159893 %
L'error mitjà a 19 dies és de 5.555519425621439 %
L'error mitjà a 20 dies és de 5.256176896629322 %
L'error mitjà a 21 dies és de 5.579002078814822 %
L'error mitjà a 22 dies és de 5.889179033367152 %
L'error mitjà a 23 dies és de 5.6989814686607065 %
L'error mitjà a 24 dies és de 6.127466729084459 %
L'error mitjà a 25 dies és de 6.284565075279462 %
L'error mitjà a 26 dies és de 6.2721430335661275 %
L'error mitjà a 27 dies és de 6.736528111806712 %
L'error mitjà a 28 dies és de 6.8158688809028005 %
L'error mitjà a 29 dies és de 7.140539469070347 %
L'error mitjà a 30 dies és de 7.3565616621563095 %
```

Imatge 26 Error a 30 dies

- Predicció a 60 dies

Obtenim un RMSE 0.076162 de i un error de 6.36%

RMSE

Error predicció 1 dia següent: 0.0559280003110196
Error predicció 2 dia següent: 0.05844069392289882
Error predicció 3 dia següent: 0.05419684176732098
Error predicció 4 dia següent: 0.05772695685651792
Error predicció 5 dia següent: 0.06335974735826629
Error predicció 6 dia següent: 0.06246182440452529
Error predicció 7 dia següent: 0.06401918938157053
Error predicció 8 dia següent: 0.0722543814283145
Error predicció 9 dia següent: 0.07540300439358928
Error predicció 10 dia següent: 0.06731205902646666
Error predicció 11 dia següent: 0.06164318992013771
Error predicció 12 dia següent: 0.06600742537311174
Error predicció 13 dia següent: 0.06829876823544793
Error predicció 14 dia següent: 0.0736145023016248
Error predicció 15 dia següent: 0.08313432640270002
Error predicció 16 dia següent: 0.08459921711032094
Error predicció 17 dia següent: 0.07859141086760126
Error predicció 18 dia següent: 0.07517651707377669
Error predicció 19 dia següent: 0.08033295532338801
Error predicció 20 dia següent: 0.07890175494205329
Error predicció 21 dia següent: 0.08091913844315257
Error predicció 22 dia següent: 0.08798862902590861
Error predicció 23 dia següent: 0.08599579003336116
Error predicció 24 dia següent: 0.0809991348631659
Error predicció 25 dia següent: 0.08036145409650076
Error predicció 26 dia següent: 0.07872938391293757
Error predicció 27 dia següent: 0.0800652197241126
Error predicció 28 dia següent: 0.08618008968708492
Error predicció 29 dia següent: 0.09151253519816722
Error predicció 30 dia següent: 0.09156480957911002
Error predicció 31 dia següent: 0.08789967541095382
Error predicció 32 dia següent: 0.0864669257833003
Error predicció 33 dia següent: 0.087601402390837874
Error predicció 34 dia següent: 0.088917722478405
Error predicció 35 dia següent: 0.09375154833834188
Error predicció 36 dia següent: 0.09554728748230988
Error predicció 37 dia següent: 0.09585624845514189
Error predicció 38 dia següent: 0.09268083090857526
Error predicció 39 dia següent: 0.091344990898073324
Error predicció 40 dia següent: 0.09003518759959035
Error predicció 41 dia següent: 0.09416988538501864
Error predicció 42 dia següent: 0.09707323825844427
Error predicció 43 dia següent: 0.10146112474798466
Error predicció 44 dia següent: 0.098377098960842272
Error predicció 45 dia següent: 0.09347814890186669
Error predicció 46 dia següent: 0.09230496100086848
Error predicció 47 dia següent: 0.09403348366442764
Error predicció 48 dia següent: 0.09424122274851919
Error predicció 49 dia següent: 0.09550196264197496
Error predicció 50 dia següent: 0.10129431968475468
Error predicció 51 dia següent: 0.1047473591941168
Error predicció 52 dia següent: 0.10257622091203498
Error predicció 53 dia següent: 0.1002371233760128
Error predicció 54 dia següent: 0.100846872204707141
Error predicció 55 dia següent: 0.100868428501991558
Error predicció 56 dia següent: 0.10577550451773723
Error predicció 57 dia següent: 0.11530133313587787
Error predicció 58 dia següent: 0.114194483949961
Error predicció 59 dia següent: 0.107650821779140088
Error predicció 60 dia següent: 0.1057217852905353

Imatge 27 RMSE a 60 dies

Veiem com l'error a 1 dia és del 3.81%, a 7 dies del 4.64%, a 30 dies del 6.32% i a 60 dies del 9.12%.

Error

L'error mitjà a 1 dia és de 3.8180360746965833 %
L'error mitjà a 2 dies és de 4.021558355059615 %
L'error mitjà a 3 dies és de 3.823315277142801 %
L'error mitjà a 4 dies és de 4.002828499436973 %
L'error mitjà a 5 dies és de 4.384652215941731 %
L'error mitjà a 6 dies és de 4.524344998221985 %
L'error mitjà a 7 dies és de 4.640300206411297 %
L'error mitjà a 8 dies és de 5.20358785853008 %
L'error mitjà a 9 dies és de 5.194588325872169 %
L'error mitjà a 10 dies és de 4.820609584068314 %
L'error mitjà a 11 dies és de 4.434926852870058 %
L'error mitjà a 12 dies és de 4.81248688649175 %
L'error mitjà a 13 dies és de 5.0041684926517425 %
L'error mitjà a 14 dies és de 5.320460299726722 %
L'error mitjà a 15 dies és de 5.541704367089151 %
L'error mitjà a 16 dies és de 5.804199072515519 %
L'error mitjà a 17 dies és de 5.648214629218328 %
L'error mitjà a 18 dies és de 5.44408625982978 %
L'error mitjà a 19 dies és de 5.697707570970561 %
L'error mitjà a 20 dies és de 5.466465287158722 %
L'error mitjà a 21 dies és de 5.825708159292685 %
L'error mitjà a 22 dies és de 6.184565273094766 %
L'error mitjà a 23 dies és de 5.917870051131651 %
L'error mitjà a 24 dies és de 5.6626559533278416 %
L'error mitjà a 25 dies és de 5.461038489512068 %
L'error mitjà a 26 dies és de 5.465777019943531 %
L'error mitjà a 27 dies és de 5.4915202880128545 %
L'error mitjà a 28 dies és de 5.8028136708702924 %
L'error mitjà a 29 dies és de 6.1145358357521395 %
L'error mitjà a 30 dies és de 6.329581147212698 %
L'error mitjà a 31 dia és de 5.949284428645927 %
L'error mitjà a 32 dies és de 6.162981455136602 %
L'error mitjà a 33 dies és de 6.182123679808553 %
L'error mitjà a 34 dies és de 6.321651533505875 %
L'error mitjà a 35 dies és de 6.455692188534395 %
L'error mitjà a 36 dies és de 6.115268308199881 %
L'error mitjà a 37 dies és de 6.442279874809209 %
L'error mitjà a 38 dies és de 6.467795482500434 %
L'error mitjà a 39 dies és de 6.539954748138895 %
L'error mitjà a 40 dies és de 6.675776298998337 %
L'error mitjà a 41 dies és de 7.049002328990378 %
L'error mitjà a 42 dies és de 7.688172886156128 %
L'error mitjà a 43 dies és de 7.498407326477663 %
L'error mitjà a 44 dies és de 7.802944871157428 %
L'error mitjà a 45 dies és de 6.819314775099253 %
L'error mitjà a 46 dies és de 6.9928338695999 %
L'error mitjà a 47 dies és de 7.396857193268208 %
L'error mitjà a 48 dies és de 7.58243110088742376 %
L'error mitjà a 49 dies és de 7.981467315519682 %
L'error mitjà a 50 dies és de 8.31966789627357 %
L'error mitjà a 51 dies és de 8.551658342138573 %
L'error mitjà a 52 dies és de 8.252460998123505 %
L'error mitjà a 53 dies és de 8.066875045588276 %
L'error mitjà a 54 dies és de 8.2257721797374 %
L'error mitjà a 55 dies és de 8.728432815253961 %
L'error mitjà a 56 dies és de 9.306081940956123 %
L'error mitjà a 57 dies és de 9.638811961213346 %
L'error mitjà a 58 dies és de 9.29849976026591 %
L'error mitjà a 59 dies és de 9.893762423450075 %
L'error mitjà a 60 dies és de 9.124222568687895 %

Imatge 28 Error a 60 dies

Un altre model que volem provar és utilitzant les variables temperatura màxima i mínima i weekday -sense humitat- agafant els 14 dies anteriors amb walkforward on va augmentant train a mesura que avancen les prediccions.

- Predicció a 7 dies

Obtenim RMSE mitjà de 0.055378 i un error percentual mitjà del 4.21%.

RMSE

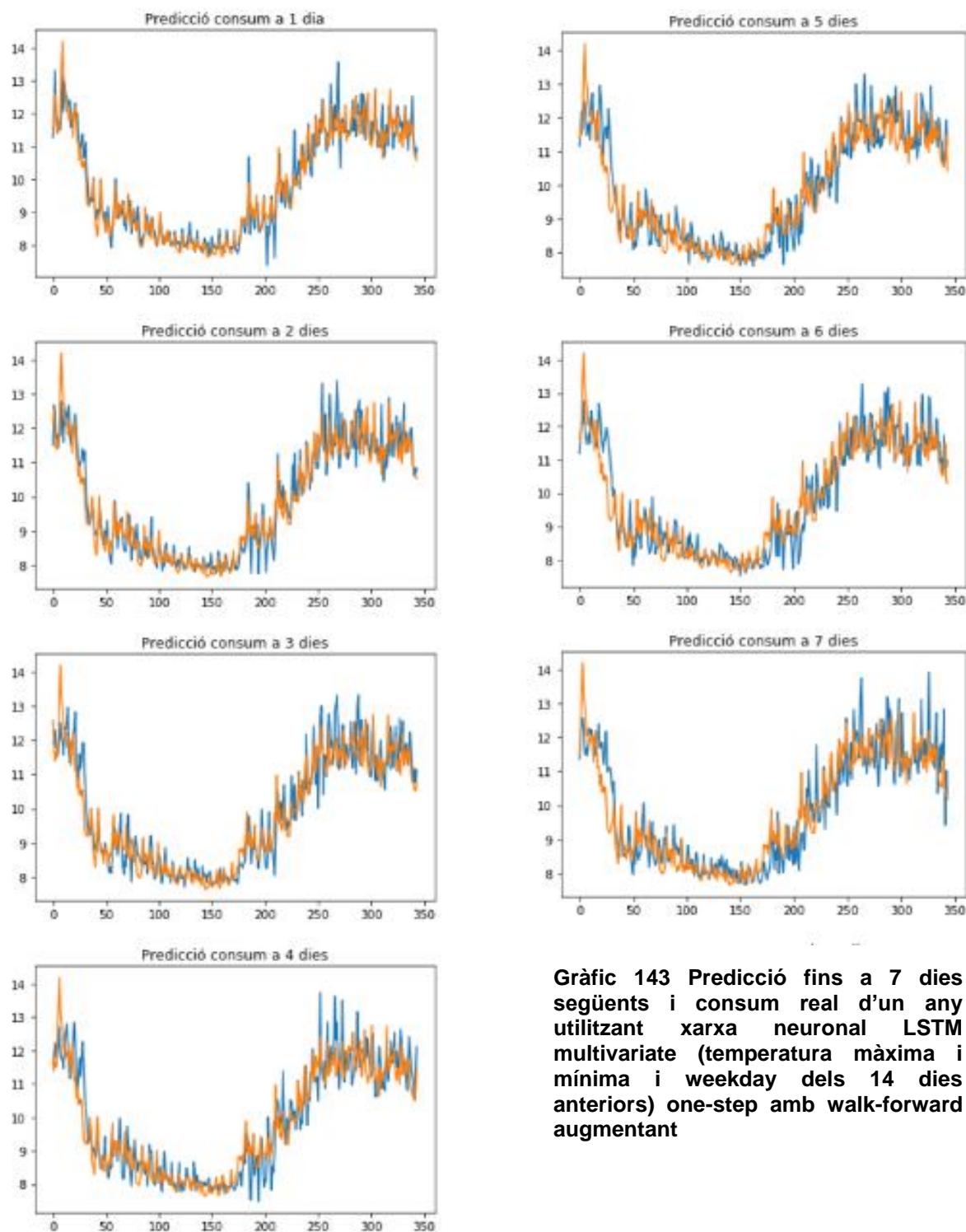
```
Error predicción 1 dia següent: 0.04766261526847275  
Error predicción 2 dia següent: 0.05282975109263351  
Error predicción 3 dia següent: 0.060496747220503855  
Error predicción 4 dia següent: 0.065340869904399  
Error predicción 5 dia següent: 0.06805877510044427  
Error predicción 6 dia següent: 0.07140896449848305  
Error predicción 7 dia següent: 0.07340117080446443
```

Imatge 29 RMSE a 7 dies

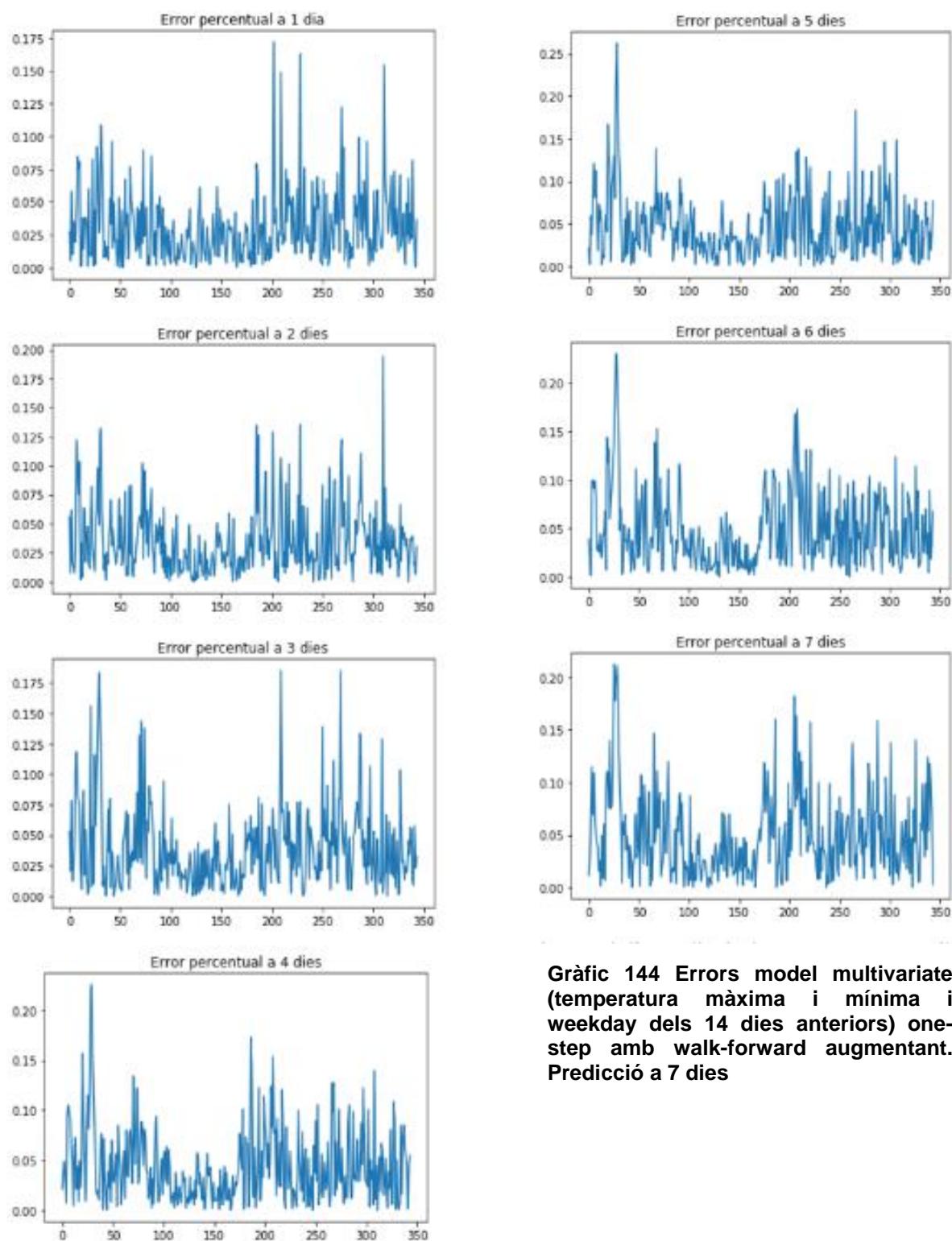
Error

```
L'error mitjà a 1 dia és de 3.1109857795752553 %  
L'error mitjà a 2 dies és de 3.5156806482526712 %  
L'error mitjà a 3 dies és de 3.9584861329651986 %  
L'error mitjà a 4 dies és de 4.37295518682416 %  
L'error mitjà a 5 dies és de 4.590828068433436 %  
L'error mitjà a 6 dies és de 4.9181066806700455 %  
L'error mitjà a 7 dies és de 5.021845306397759 %
```

Imatge 30 Error a 7 dies



Gràfic 143 Predicció fins a 7 dies següents i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant



Gràfic 144 Errors model multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant. Predicció a 7 dies

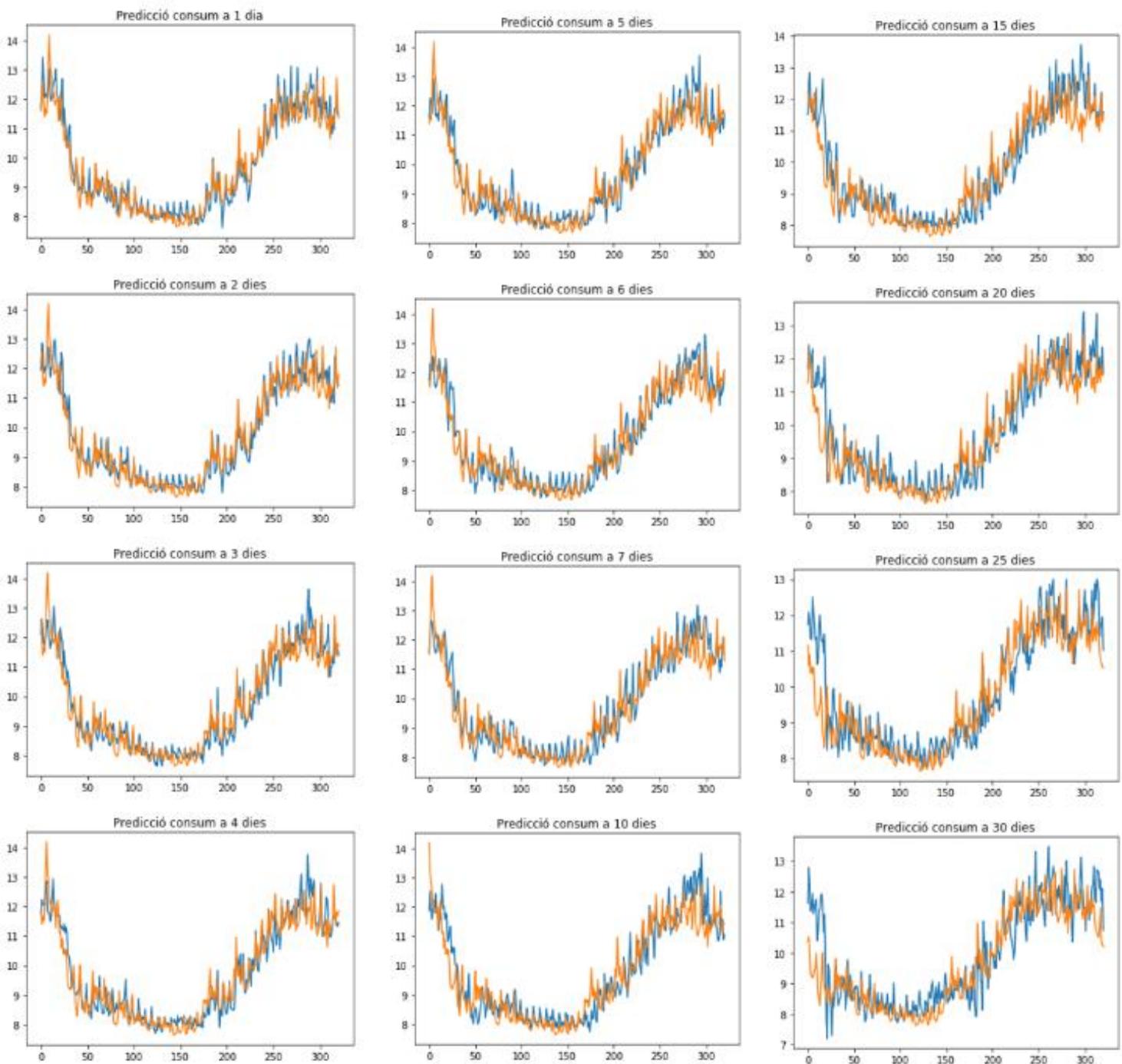
- Predicció a 30 dies

Obtenim un de RMSE 0.06336 i un d'error de 4.79%.

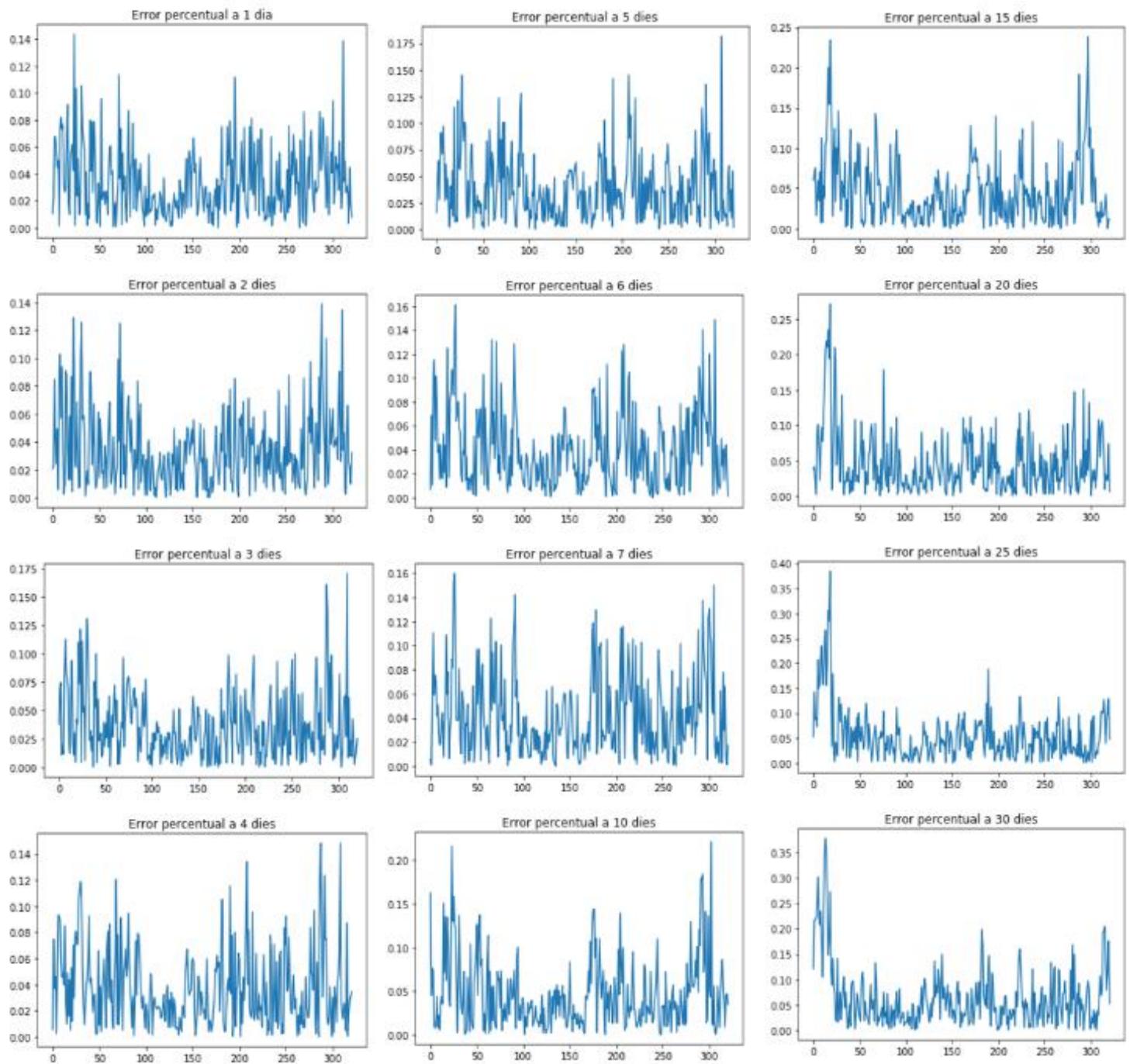
RMSE	Error
Error predicció 1 dia següent: 0.0491192815924	L'error mitjà a 1 dia és de 3.4155629885365033 %
Error predicció 2 dia següent: 0.0507005606687	L'error mitjà a 2 dies és de 3.435038616018267 %
Error predicció 3 dia següent: 0.0540494580935	L'error mitjà a 3 dies és de 3.6272238547226237 %
Error predicció 4 dia següent: 0.0543232858033	L'error mitjà a 4 dies és de 3.7341992245670785 %
Error predicció 5 dia següent: 0.0569058150000	L'error mitjà a 5 dies és de 3.922543222430784 %
Error predicció 6 dia següent: 0.0592166339281	L'error mitjà a 6 dies és de 4.098318260534777 %
Error predicció 7 dia següent: 0.0613506089097	L'error mitjà a 7 dies és de 4.27063737383024 %
Error predicció 8 dia següent: 0.0642029788235	L'error mitjà a 8 dies és de 4.319841789147668 %
Error predicció 9 dia següent: 0.0667145677306	L'error mitjà a 9 dies és de 4.585037871068335 %
Error predicció 10 dia següent: 0.070145816850	L'error mitjà a 10 dies és de 4.630526471102679 %
Error predicció 11 dia següent: 0.068722243278	L'error mitjà a 11 dies és de 4.74178437159315 %
Error predicció 12 dia següent: 0.067849157131	L'error mitjà a 12 dies és de 4.690657502771326 %
Error predicció 13 dia següent: 0.072864725782	L'error mitjà a 13 dies és de 4.86098757983995 %
Error predicció 14 dia següent: 0.069308136246	L'error mitjà a 14 dies és de 4.692380482134837 %
Error predicció 15 dia següent: 0.070394246699	L'error mitjà a 15 dies és de 4.707857514332385 %
Error predicció 16 dia següent: 0.072517561883	L'error mitjà a 16 dies és de 4.9138177207568 %
Error predicció 17 dia següent: 0.075323373813	L'error mitjà a 17 dies és de 5.0166324347546425 %
Error predicció 18 dia següent: 0.074407938339	L'error mitjà a 18 dies és de 5.090880963978517 %
Error predicció 19 dia següent: 0.071919683531	L'error mitjà a 19 dies és de 4.87361041373152 %
Error predicció 20 dia següent: 0.069578301098	L'error mitjà a 20 dies és de 4.672330915157581 %
Error predicció 21 dia següent: 0.070290007618	L'error mitjà a 21 dies és de 4.9317061580084281 %
Error predicció 22 dia següent: 0.072923432220	L'error mitjà a 22 dies és de 5.0438676326233765 %
Error predicció 23 dia següent: 0.073903124814	L'error mitjà a 23 dies és de 5.114543004885975 %
Error predicció 24 dia següent: 0.076886199700	L'error mitjà a 24 dies és de 5.275172035039056 %
Error predicció 25 dia següent: 0.080109478502	L'error mitjà a 25 dies és de 5.415003327021586 %
Error predicció 26 dia següent: 0.081417709236	L'error mitjà a 26 dies és de 5.439261632304306 %
Error predicció 27 dia següent: 0.084349650862	L'error mitjà a 27 dies és de 5.638453581876803 %
Error predicció 28 dia següent: 0.088973743180	L'error mitjà a 28 dies és de 5.989429692277341 %
Error predicció 29 dia següent: 0.093267907591	L'error mitjà a 29 dies és de 6.204044733464207 %
Error predicció 30 dia següent: 0.094442777183	L'error mitjà a 30 dies és de 6.384730483524267 %

imatge 31 RMSE a 30 dies

imatge 32 Error a 30 dies



Gràfic 145 Predicccions dels 30 dies següents i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i weekday dels 14 dies anteriors) one-step amb walk-forward augmentant



Gràfic 146 Error percentual predicción a 30 dies

- Predicció a 60 dies

Obtenim un RMSE 0.070017 de i un error de .6.0769

RMSE

Error predicció 1 dia següent: 0.047426721609315846
 Error predicció 2 dia següent: 0.051477039658305745
 Error predicció 3 dia següent: 0.0510570088416192
 Error predicció 4 dia següent: 0.05713255024308317
 Error predicció 5 dia següent: 0.05925391329625415
 Error predicció 6 dia següent: 0.0610183544867185
 Error predicció 7 dia següent: 0.06434462482935731
 Error predicció 8 dia següent: 0.06841048649879698
 Error predicció 9 dia següent: 0.06777364279281463
 Error predicció 10 dia següent: 0.07077347084777413
 Error predicció 11 dia següent: 0.06964998287211543
 Error predicció 12 dia següent: 0.07161635987188451
 Error predicció 13 dia següent: 0.07079040253621885
 Error predicció 14 dia següent: 0.07047948767114583
 Error predicció 15 dia següent: 0.07010867353826797
 Error predicció 16 dia següent: 0.07077346232015823
 Error predicció 17 dia següent: 0.07269983488038752
 Error predicció 18 dia següent: 0.06914378861234487
 Error predicció 19 dia següent: 0.06886230233801179
 Error predicció 20 dia següent: 0.06739058148875043
 Error predicció 21 dia següent: 0.06853165011164186
 Error predicció 22 dia següent: 0.06668668506648707
 Error predicció 23 dia següent: 0.06928965674687486
 Error predicció 24 dia següent: 0.07330814378345413
 Error predicció 25 dia següent: 0.07364812210983827
 Error predicció 26 dia següent: 0.06918494235901879
 Error predicció 27 dia següent: 0.070540892327933248
 Error predicció 28 dia següent: 0.070694987577785479
 Error predicció 29 dia següent: 0.07364106811652503
 Error predicció 30 dia següent: 0.07214982291949486
 Error predicció 31 dia següent: 0.07240328236708428
 Error predicció 32 dia següent: 0.07793526783622794
 Error predicció 33 dia següent: 0.08369031763760137
 Error predicció 34 dia següent: 0.08838768293250438
 Error predicció 35 dia següent: 0.08122098350963226
 Error predicció 36 dia següent: 0.08090638894582052
 Error predicció 37 dia següent: 0.07965747139839638
 Error predicció 38 dia següent: 0.08204252037340757
 Error predicció 39 dia següent: 0.08884437345844868
 Error predicció 40 dia següent: 0.09347904616511639
 Error predicció 41 dia següent: 0.08771855412286649
 Error predicció 42 dia següent: 0.09895103058643714
 Error predicció 43 dia següent: 0.09187188768053889
 Error predicció 44 dia següent: 0.093260849386248014
 Error predicció 45 dia següent: 0.09573497642081126
 Error predicció 46 dia següent: 0.09962862232584871
 Error predicció 47 dia següent: 0.10131628774304995
 Error predicció 48 dia següent: 0.09873658077597949
 Error predicció 49 dia següent: 0.09767663688958021
 Error predicció 50 dia següent: 0.1018326227798295
 Error predicció 51 dia següent: 0.10348051234804032
 Error predicció 52 dia següent: 0.10621973183569575
 Error predicció 53 dia següent: 0.10334306505863514
 Error predicció 54 dia següent: 0.10547811101075967
 Error predicció 55 dia següent: 0.10171243428568191
 Error predicció 56 dia següent: 0.10119088287716847
 Error predicció 57 dia següent: 0.10287660557976439
 Error predicció 58 dia següent: 0.10147058011773248
 Error predicció 59 dia següent: 0.09912403855186336
 Error predicció 60 dia següent: 0.09842710510855089

Error

L'error mitjà a 1 dia és de 3.1604447888601532 %
 L'error mitjà a 2 dies és de 3.464385676917455 %
 L'error mitjà a 3 dies és de 3.5805887239782073 %
 L'error mitjà a 4 dies és de 4.064903567644374 %
 L'error mitjà a 5 dies és de 4.155215071815508 %
 L'error mitjà a 6 dies és de 4.185795441829246 %
 L'error mitjà a 7 dies és de 4.476025727342001 %
 L'error mitjà a 8 dies és de 4.628067407888385 %
 L'error mitjà a 9 dies és de 4.580228498751644 %
 L'error mitjà a 10 dies és de 4.7891015220041435 %
 L'error mitjà a 11 dies és de 4.83315783747827 %
 L'error mitjà a 12 dies és de 4.868641647277406 %
 L'error mitjà a 13 dies és de 4.893875143989665 %
 L'error mitjà a 14 dies és de 4.93686532777373 %
 L'error mitjà a 15 dies és de 4.964319213650899 %
 L'error mitjà a 16 dies és de 5.144468956153503 %
 L'error mitjà a 17 dies és de 5.22918772740669 %
 L'error mitjà a 18 dies és de 4.9252484762506175 %
 L'error mitjà a 19 dies és de 4.812543083917279 %
 L'error mitjà a 20 dies és de 4.822614402196003 %
 L'error mitjà a 21 dies és de 4.892429688742379 %
 L'error mitjà a 22 dies és de 4.793916799883175 %
 L'error mitjà a 23 dies és de 4.935175125462188 %
 L'error mitjà a 24 dies és de 5.092982800887137 %
 L'error mitjà a 25 dies és de 4.995700251154802 %
 L'error mitjà a 26 dies és de 4.676161498891386 %
 L'error mitjà a 27 dies és de 4.986950011169034 %
 L'error mitjà a 28 dies és de 4.931788784158473 %
 L'error mitjà a 29 dies és de 5.114608137456565 %
 L'error mitjà a 30 dies és de 4.981571685250897 %
 L'error mitjà a 31 dia és de 6.191048443642301 %
 L'error mitjà a 32 dies és de 6.509236181185099 %
 L'error mitjà a 33 dies és de 6.276834363996911 %
 L'error mitjà a 34 dies és de 5.998815655336182 %
 L'error mitjà a 35 dies és de 5.997451716863105 %
 L'error mitjà a 36 dies és de 6.162358054116744 %
 L'error mitjà a 37 dies és de 6.238372282452784 %
 L'error mitjà a 38 dies és de 6.341203881953648 %
 L'error mitjà a 39 dies és de 6.565244448492666 %
 L'error mitjà a 40 dies és de 6.78474138812833 %
 L'error mitjà a 41 dies és de 6.70827122881783 %
 L'error mitjà a 42 dies és de 6.7564592397669 %
 L'error mitjà a 43 dies és de 6.829335048592573 %
 L'error mitjà a 44 dies és de 7.060667315287692 %
 L'error mitjà a 45 dies és de 7.153373844256226 %
 L'error mitjà a 46 dies és de 7.002829398601474 %
 L'error mitjà a 47 dies és de 7.099920008781681 %
 L'error mitjà a 48 dies és de 7.334090318439893 %
 L'error mitjà a 49 dies és de 7.638068619098444 %
 L'error mitjà a 50 dies és de 8.724221501559965 %
 L'error mitjà a 51 dies és de 8.524959362485482 %
 L'error mitjà a 52 dies és de 8.769660248428962 %
 L'error mitjà a 53 dies és de 8.738257845390047 %
 L'error mitjà a 54 dies és de 8.958789929367572 %
 L'error mitjà a 55 dies és de 9.14775845507389 %
 L'error mitjà a 56 dies és de 9.182399300755133 %
 L'error mitjà a 57 dies és de 9.016827408816363 %
 L'error mitjà a 58 dies és de 8.825837217822096 %
 L'error mitjà a 59 dies és de 8.992607297254484 %
 L'error mitjà a 60 dies és de 9.1697755795739 %

Imatge 33 RMSE 60 dies

Imatge 34 Error 60 dies

També provarem un model que utilitzi les variables temperatura màxima i mínima, la humitat, weekday i l'estació, agafant els 14 dies anteriors amb walkforward on va augmentant train a mesura que avancen les prediccions.

- Predicció a 7 dies

Obtenim RMSE mitjà de 0.061569 i error de 4.73%

RMSE

```
Error predicció 1 dia següent: 0.05146226737728889
Error predicció 2 dia següent: 0.06427961550676087
Error predicció 3 dia següent: 0.07047695992418186
Error predicció 4 dia següent: 0.0729435021088956
Error predicció 5 dia següent: 0.07387084787639028
Error predicció 6 dia següent: 0.07592566299040113
Error predicció 7 dia següent: 0.07798694442206162
```

imatge 35 RMSE 7 dies

Error

```
L'error mitjà a 1 dia és de 3.4457691113463613 %
L'error mitjà a 2 dies és de 4.379628225399063 %
L'error mitjà a 3 dies és de 4.635675119622094 %
L'error mitjà a 4 dies és de 4.9615905778450475 %
L'error mitjà a 5 dies és de 5.139195232117659 %
L'error mitjà a 6 dies és de 5.299377130256948 %
L'error mitjà a 7 dies és de 5.260351754615196 %
```

imatge 36 Error 7 dies

- Predicció a 30 dies

Obtenim un RMSE de 0.066835 i un error de 4.99%

RMSE

```
Error predicció 1 dia següent: 0.049337510671949955
Error predicció 2 dia següent: 0.05897883062919062
Error predicció 3 dia següent: 0.06308163453592287
Error predicció 4 dia següent: 0.06076977549774085
Error predicció 5 dia següent: 0.06185609714249502
Error predicció 6 dia següent: 0.06583318079577589
Error predicció 7 dia següent: 0.06630949526493879
Error predicció 8 dia següent: 0.06487907805059497
Error predicció 9 dia següent: 0.06176333974508744
Error predicció 10 dia següent: 0.06167442294976592
Error predicció 11 dia següent: 0.06489166695618023
Error predicció 12 dia següent: 0.0633890872948724
Error predicció 13 dia següent: 0.06410856965239496
Error predicció 14 dia següent: 0.06740741075519648
Error predicció 15 dia següent: 0.06579254830515392
Error predicció 16 dia següent: 0.06781236356485676
Error predicció 17 dia següent: 0.06973841197166078
Error predicció 18 dia següent: 0.07329898283603126
Error predicció 19 dia següent: 0.08014305436219984
Error predicció 20 dia següent: 0.0790957178746485
Error predicció 21 dia següent: 0.07736549340943488
Error predicció 22 dia següent: 0.08074311309114227
Error predicció 23 dia següent: 0.08154709379866976
Error predicció 24 dia següent: 0.08542792623423501
Error predicció 25 dia següent: 0.09000969423883756
Error predicció 26 dia següent: 0.09035396831295313
Error predicció 27 dia següent: 0.09385145297983044
Error predicció 28 dia següent: 0.08878927255861549
Error predicció 29 dia següent: 0.09002641503730122
Error predicció 30 dia següent: 0.09706075845981398
```

imatge 37 RMSE 30 dies

Error

```
L'error mitjà a 1 dia és de 3.325398446635793 %
L'error mitjà a 2 dies és de 4.101790795521142 %
L'error mitjà a 3 dies és de 4.210171430343687 %
L'error mitjà a 4 dies és de 4.109373479128708 %
L'error mitjà a 5 dies és de 4.291637096228313 %
L'error mitjà a 6 dies és de 4.740265902585702 %
L'error mitjà a 7 dies és de 4.82505085872098 %
L'error mitjà a 8 dies és de 4.561298866865552 %
L'error mitjà a 9 dies és de 4.21540157432929 %
L'error mitjà a 10 dies és de 4.359303538439615 %
L'error mitjà a 11 dies és de 4.537997439170452 %
L'error mitjà a 12 dies és de 4.408950309076382 %
L'error mitjà a 13 dies és de 4.49343722537355 %
L'error mitjà a 14 dies és de 4.610933931604066 %
L'error mitjà a 15 dies és de 4.66599421524011 %
L'error mitjà a 16 dies és de 4.696702815930394 %
L'error mitjà a 17 dies és de 4.666288818104336 %
L'error mitjà a 18 dies és de 4.917265857972283 %
L'error mitjà a 19 dies és de 5.4098107015445285 %
L'error mitjà a 20 dies és de 5.418904616601345 %
L'error mitjà a 21 dies és de 5.222471154735223 %
L'error mitjà a 22 dies és de 5.303528685983539 %
L'error mitjà a 23 dies és de 5.2861112355653646 %
L'error mitjà a 24 dies és de 5.677370059297436 %
L'error mitjà a 25 dies és de 6.076114073370857 %
L'error mitjà a 26 dies és de 6.121206076843288 %
L'error mitjà a 27 dies és de 6.367682312757138 %
L'error mitjà a 28 dies és de 5.99030297453761 %
L'error mitjà a 29 dies és de 6.221411181424961 %
L'error mitjà a 30 dies és de 6.828205237881082 %
```

imatge 38 Error a 30 dies

- Predicció a 60 dies

Obtenim un RMSE 0.06614 de i un error de 6.05%

RMSE

Error predicció 1 dia següent: 0.05206173804033639
 Error predicció 2 dia següent: 0.05393983976715141
 Error predicció 3 dia següent: 0.05672352601765881
 Error predicció 4 dia següent: 0.05926723995368844
 Error predicció 5 dia següent: 0.05965725365634359
 Error predicció 6 dia següent: 0.06091777601229184
 Error predicció 7 dia següent: 0.063576380124772524
 Error predicció 8 dia següent: 0.06297438064497894
 Error predicció 9 dia següent: 0.06363021117563711
 Error predicció 10 dia següent: 0.0612219509679283
 Error predicció 11 dia següent: 0.06268844235988946
 Error predicció 12 dia següent: 0.05735358857978129
 Error predicció 13 dia següent: 0.05868812920183576
 Error predicció 14 dia següent: 0.06028130166356804
 Error predicció 15 dia següent: 0.06020254805338228
 Error predicció 16 dia següent: 0.060465494114093354
 Error predicció 17 dia següent: 0.061528922362915824
 Error predicció 18 dia següent: 0.06843888442027361
 Error predicció 19 dia següent: 0.070976480502262544
 Error predicció 20 dia següent: 0.06978219811122893
 Error predicció 21 dia següent: 0.0715808623979674
 Error predicció 22 dia següent: 0.07467178109868448
 Error predicció 23 dia següent: 0.07480885867196478
 Error predicció 24 dia següent: 0.0762677019498332
 Error predicció 25 dia següent: 0.0785779334727503
 Error predicció 26 dia següent: 0.07625661565995989
 Error predicció 27 dia següent: 0.0807545624769193
 Error predicció 28 dia següent: 0.07997294593170816
 Error predicció 29 dia següent: 0.08018465226909498
 Error predicció 30 dia següent: 0.08325554875928607
 Error predicció 31 dia següent: 0.0842455127294678
 Error predicció 32 dia següent: 0.0826168075875881
 Error predicció 33 dia següent: 0.08089490917735336
 Error predicció 34 dia següent: 0.08411909152589746
 Error predicció 35 dia següent: 0.082521881545721
 Error predicció 36 dia següent: 0.08294882986511436
 Error predicció 37 dia següent: 0.08025409023819083
 Error predicció 38 dia següent: 0.078787185580472438
 Error predicció 39 dia següent: 0.07743888741460075
 Error predicció 40 dia següent: 0.0773325905163542
 Error predicció 41 dia següent: 0.0752589280786795
 Error predicció 42 dia següent: 0.07769228973275048
 Error predicció 43 dia següent: 0.07683015009051225
 Error predicció 44 dia següent: 0.07453287047408728
 Error predicció 45 dia següent: 0.07539137258524659
 Error predicció 46 dia següent: 0.0758957728264142
 Error predicció 47 dia següent: 0.07673834449738788
 Error predicció 48 dia següent: 0.07946258852271368
 Error predicció 49 dia següent: 0.08099985855494976
 Error predicció 50 dia següent: 0.0783686164581228
 Error predicció 51 dia següent: 0.07802336393522602
 Error predicció 52 dia següent: 0.07784445547483125
 Error predicció 53 dia següent: 0.0754548879277676
 Error predicció 54 dia següent: 0.07794880227902654
 Error predicció 55 dia següent: 0.07803730748858009
 Error predicció 56 dia següent: 0.07890730601766524
 Error predicció 57 dia següent: 0.08193374842161447
 Error predicció 58 dia següent: 0.082687468560888525
 Error predicció 59 dia següent: 0.08430835603306758
 Error predicció 60 dia següent: 0.08636886273729785

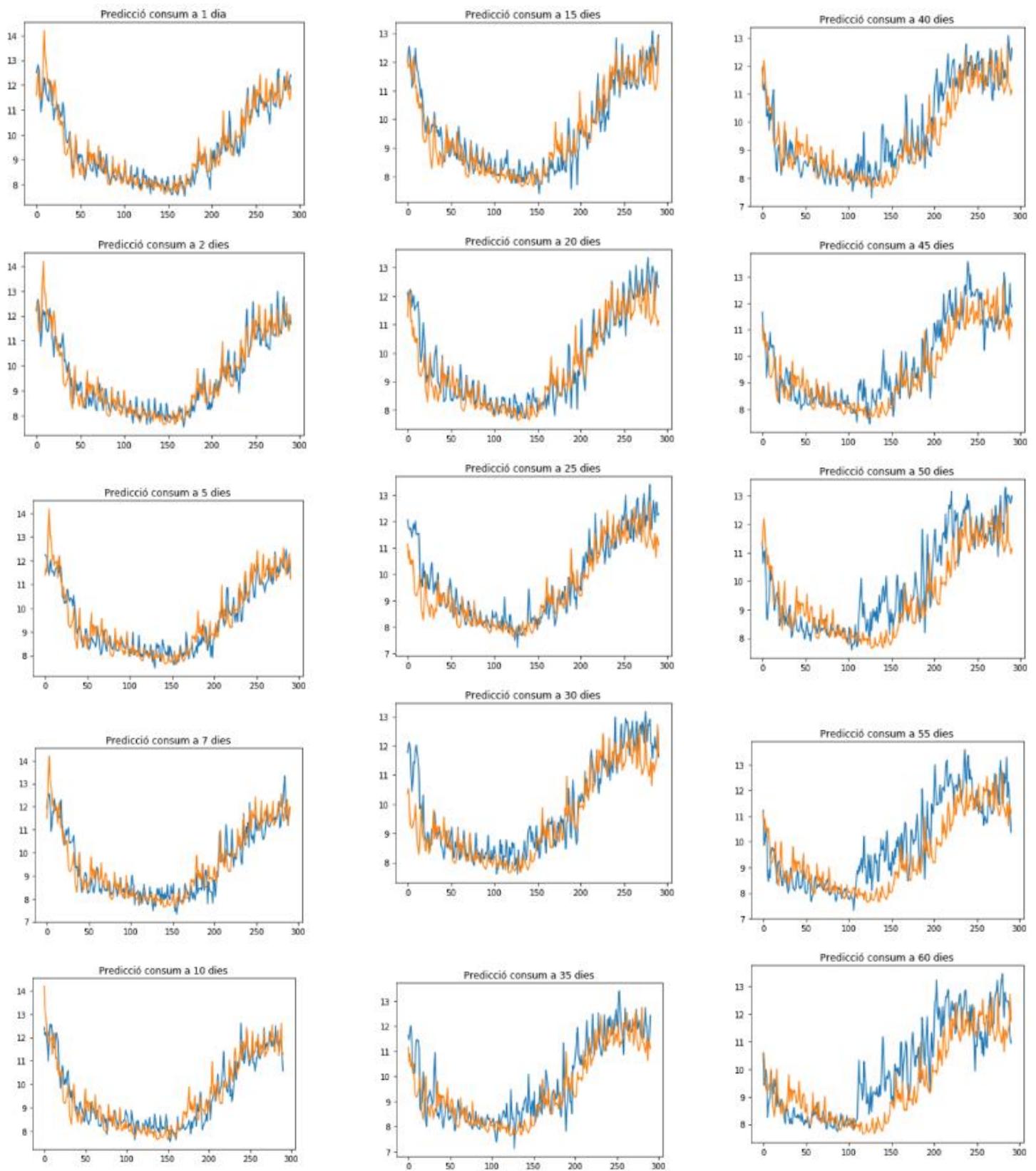
Imatge 39 RMSE a 60 dies

Error

L'error mitjà a 1 dia és de 3.578364877637653 %
 L'error mitjà a 2 dies és de 3.8088180734959605 %
 L'error mitjà a 3 dies és de 3.9880868017708515 %
 L'error mitjà a 4 dies és de 4.066230835399487 %
 L'error mitjà a 5 dies és de 4.236671342071782 %
 L'error mitjà a 6 dies és de 4.406047613275763 %
 L'error mitjà a 7 dies és de 4.613064606611421 %
 L'error mitjà a 8 dies és de 4.688426854880731 %
 L'error mitjà a 9 dies és de 4.761125871966178 %
 L'error mitjà a 10 dies és de 4.559090317026345 %
 L'error mitjà a 11 dies és de 4.601762655912542 %
 L'error mitjà a 12 dies és de 4.343120070942558 %
 L'error mitjà a 13 dies és de 4.463662867831812 %
 L'error mitjà a 14 dies és de 4.495800439919593 %
 L'error mitjà a 15 dies és de 4.413768284611584 %
 L'error mitjà a 16 dies és de 4.424651101470647 %
 L'error mitjà a 17 dies és de 4.511696213918755 %
 L'error mitjà a 18 dies és de 4.770433071407211 %
 L'error mitjà a 19 dies és de 4.9239683445473466 %
 L'error mitjà a 20 dies és de 4.894214428878567 %
 L'error mitjà a 21 dies és de 5.068182364752452 %
 L'error mitjà a 22 dies és de 5.13825391784946 %
 L'error mitjà a 23 dies és de 4.998351099817669 %
 L'error mitjà a 24 dies és de 5.110101921462872 %
 L'error mitjà a 25 dies és de 5.388140111952774 %
 L'error mitjà a 26 dies és de 5.268926556105512 %
 L'error mitjà a 27 dies és de 5.79595074161711 %
 L'error mitjà a 28 dies és de 5.639658043761658 %
 L'error mitjà a 29 dies és de 5.521716493047065 %
 L'error mitjà a 30 dies és de 5.549817487473326 %
 L'error mitjà a 31 dia és de 6.389031330813144 %
 L'error mitjà a 32 dies és de 6.42635582993137 %
 L'error mitjà a 33 dies és de 6.472439875170774 %
 L'error mitjà a 34 dies és de 6.902582644245282 %
 L'error mitjà a 35 dies és de 6.783951657588317 %
 L'error mitjà a 36 dies és de 6.663898676449636 %
 L'error mitjà a 37 dies és de 6.583113693294758 %
 L'error mitjà a 38 dies és de 6.29405044648471 %
 L'error mitjà a 39 dies és de 6.422765271118778 %
 L'error mitjà a 40 dies és de 6.180573544935269 %
 L'error mitjà a 41 dies és de 6.4431364967241525 %
 L'error mitjà a 42 dies és de 6.476602441478506 %
 L'error mitjà a 43 dies és de 6.25553923148803555 %
 L'error mitjà a 44 dies és de 6.427071540790533 %
 L'error mitjà a 45 dies és de 6.6264245382811355 %
 L'error mitjà a 46 dies és de 6.4475088819860199 %
 L'error mitjà a 47 dies és de 6.238230883414947 %
 L'error mitjà a 48 dies és de 6.593273630190143 %
 L'error mitjà a 49 dies és de 6.990289416831466 %
 L'error mitjà a 50 dies és de 8.095185482328223 %
 L'error mitjà a 51 dies és de 8.440912768717862 %
 L'error mitjà a 52 dies és de 8.637332794453588 %
 L'error mitjà a 53 dies és de 8.584930722293516 %
 L'error mitjà a 54 dies és de 9.059700726321696 %
 L'error mitjà a 55 dies és de 9.216744294102472 %
 L'error mitjà a 56 dies és de 8.984997159137546 %
 L'error mitjà a 57 dies és de 8.94695561557581 %
 L'error mitjà a 58 dies és de 9.07172487121218 %
 L'error mitjà a 59 dies és de 9.202666554691943 %
 L'error mitjà a 60 dies és de 9.2918591155622 %

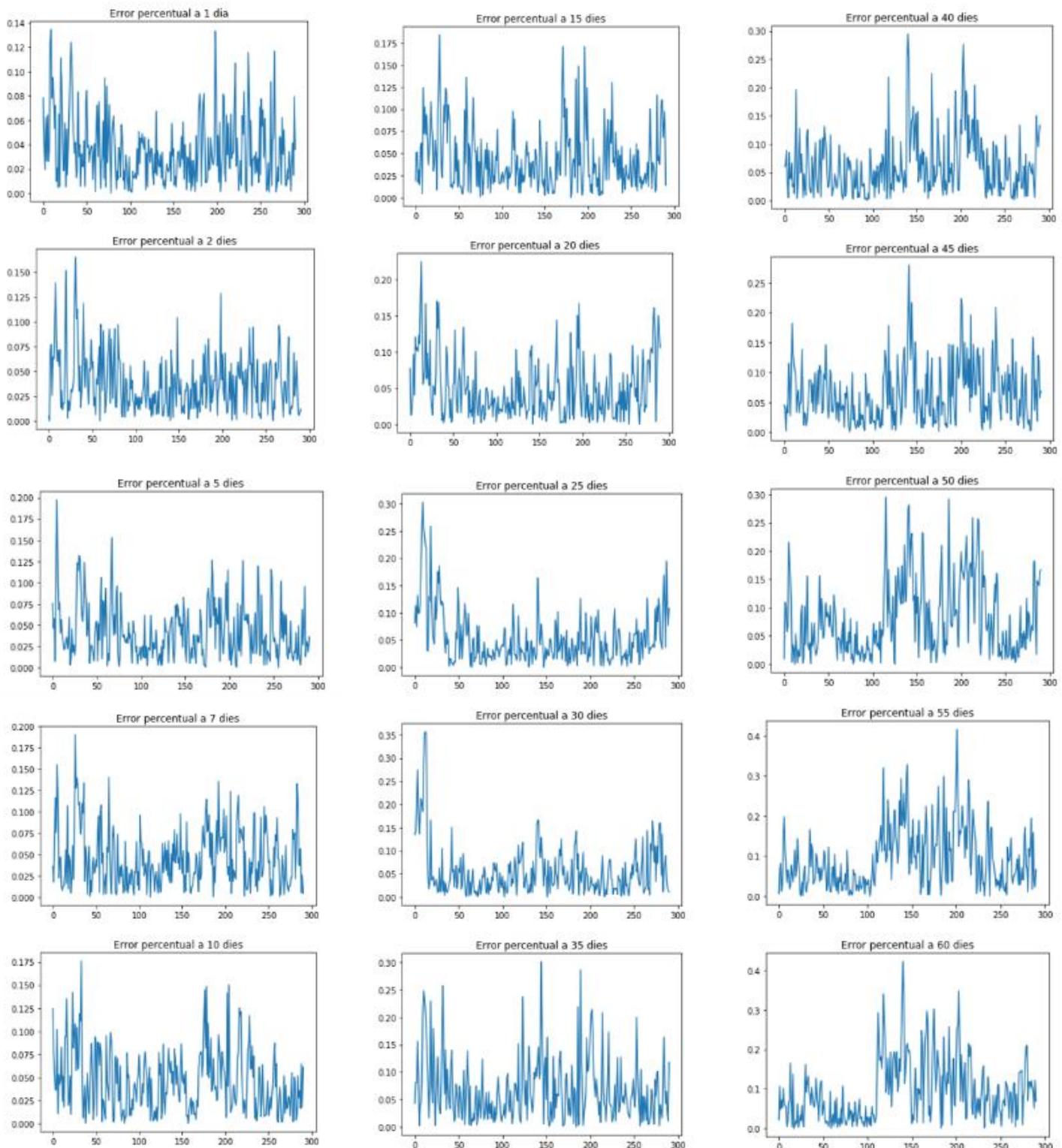
Imatge 40 Error a 60 dies

Visualitzem-ho:



Gràfic 147 Predicció a 60 dies

Aquest és el millor model per a 60 dies, i veiem que és millor amb la variable season, en canvi a menys dies és millor sense la variable season.



Gràfic 148 Error predicció a 60 dies

A més dies fem la prediccó, pitjor és l'error a l'estiu.

També hem fet altres models utilitzant el walkforward amb un train de 365 dies:

Per a predir 7 dies endavant:

- Temp max, humidity, 7 dies anteriors, obtenint un 0.06034 RMSE
- Temp max, humidity, 8 dies anteriors, obtenint un 0.06175 RMSE
- Temp max, humidity, 14 dies anteriors, obtenint un 0.05973 RMSE
 - Temp max, humidity, visibility, cloudCover, 8 dies anteriors 0.06789 RMSE

Per a predir 30 dies endavant:

- Temp max, humidity, 14 dies anteriors, obtenint un 0.06893 RMSE

Per a predir 60 dies endavant

- Temp max, humidity, 14 dies anteriors, obtenint un 0.07343 RMSE

No entrem en gran detall en aquests models ja que no milloren els models anteriors.

Taula resum multistep:

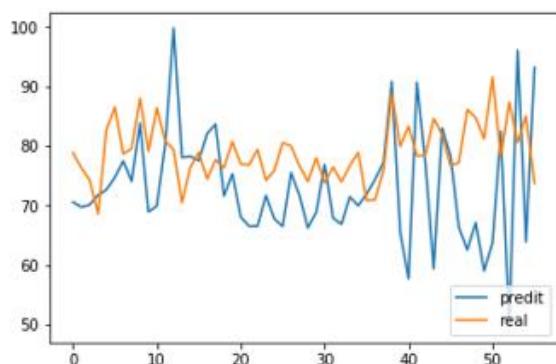
	Variables	Dies passats	Dies predits	Walkforward	RMSE	Error mitjà
MM1	Temp max, temp min, humidity, weekday	14	7	Walkforward +365 dies (augmentant)	0.05628	4.27%
MM1a	Temp max, temp min, humidity	14	7	Walkforward +365 dies (augmentant)	0.06372	4.81%
MM1b	Temp max, temp min, weekday	14	7	Walkforward +365 dies (augmentant)	0.05537	4.21%
MM1c	Temp max, temp min, humidity, weekday, season	14	7	Walkforward +365 dies (augmentant)	0.061569	4.73%
MM2	Temp max, temp min, humidity, weekday	14	30	Walkforward 365 dies (augmentant)	0.06461	4.96%
MM2a	Temp max, temp min, humidity	14	30	Walkforward +365 dies (augmentant)	0.06972	5.26%
MM2b	Temp max, temp min, weekday	14	30	Walkforward +365 dies (augmentant)	0.06336	4.79%
MM2c	Temp max, temp min, humidity, weekday, season	14	30	Walkforward +365 dies (augmentant)	0.066835	4.99%
MM3	Temp max, temp min, humidity, weekday	14	60	Walkforward 365 dies (augmentant)	0.07087	6.28%

MM3a	Temp max, temp min, humidity	14	60	Walkforward +365 dies (augmentant)	0.076162	6.36%
MM3b	Temp max, temp min, weekday	14	60	Walkforward +365 dies (augmentant)	0.070017	6.07%
MM3c	Temp max, temp min, humidity, weekday, season	14	60	Walkforward +365 dies (augmentant)	0.066145	6.05%
MM4	Temp max, humidity	7	7	Walkforward 365 dies	0.06034	
MM5	Temp max, humidity	8	7	Walkforward 365 dies	0.061748	
MM6	Temp max, humidity	14	7	Walkforward 365 dies	0.05973	
MM7	Temp max, humidity	14	30	Walkforward 365 dies	0.068926	
MM8	Temp max, humidity	14	60	Walkforward 365 dies	0.073428	
MM9	Temp max, humidity, visibility, cloudCover	8	7	Walkforward 365 dies	0.067887	

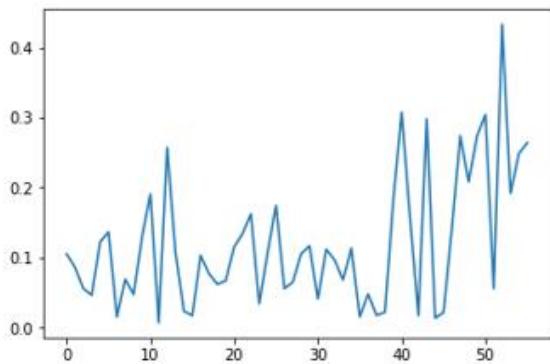
5.4 Predicció agregat setmanal

Ja hem fet una predicció diària a una setmana vista, però en aquest cas el que farem és agrupar les dades diàries que tenim (mitjana consum diari per llar) obtenint la suma de consum setmanal mitjà per llar.

Comencem amb un model amb una xarxa neuronal LSTM univariate one-step, és a dir, només utilitzem les dades passades , en aquest cas 7 setmanes, del consum, temperatura màxima i mínima, obtenint un error mitjà (RMSE) de 0.1856 i un error percentual del 12%, molt més alt que en el model diari on predim setmanalment.

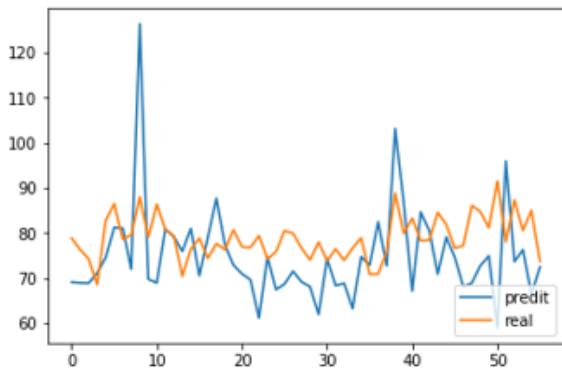


Gràfic 149 Predicció de la setmana següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima de les 7 setmanes anteriors) one-step amb walk-forward on train no augmenta

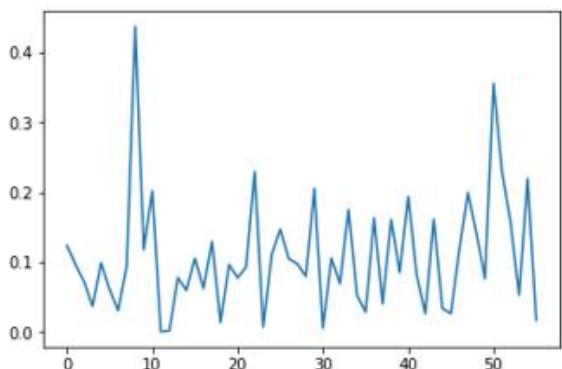


Gràfic 150 Errors del model setmanal multivariate (temperatura màxima i mínima de les 7 setmanes anteriors) one-step amb walk-forward on train no augmenta

Provem un altre model afegint-hi la humitat, i la precisió millora però l'error continua sent alt, ja que obtenim un 0.1663 de RMSE i un 10.82% d'error.



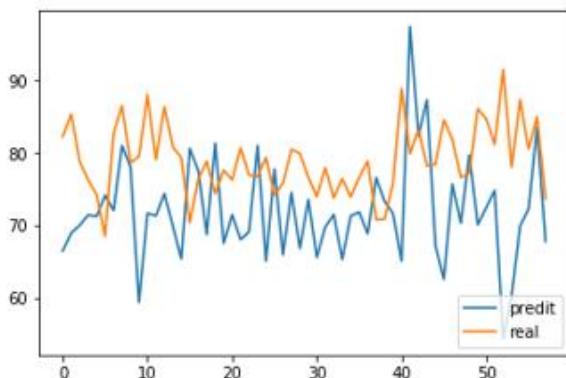
Gràfic 151 Prediccions de la setmana següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i humitat de les 7 setmanes anteriors) one-step amb walk-forward on train no augmenta



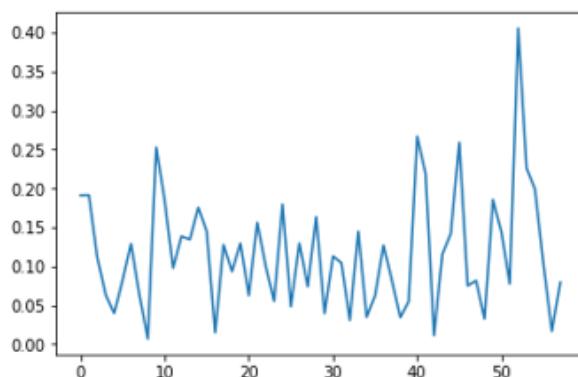
Gràfic 152 Errors del model setmanal multivariate (temperatura màxima i mínima i humitat de les 7 setmanes anteriors) one-step amb walk-forward on train no augmenta

Provem aquest últim model però enlloc d'utilitzar les 7 setmanes anteriors, utilitzem 4 setmanes anteriors. Obtenim un 0.15868 de RMSE i un 11.74% d'error, que tot i millorar els models anteriors, no millora (ni s'acosta) a l'error obtingut en el model diari on fèiem

la predicción setmanal (a 7 dies), de manera que descartem aquest tipus de model agregat setmanalment.



Gràfic 153 Prediccions de la setmana següent i consum real d'un any utilitzant xarxa neuronal LSTM multivariate (temperatura màxima i mínima i humitat de les 4 setmanes anteriors) one-step amb walk-forward on train no augmenta



Gràfic 154 Errors del model setmanal multivariate (temperatura màxima i mínima i humitat de les 4 setmanes anteriors) one-step amb walk-forward on train no augmenta

Taula resum models setmanals:

MW	Temp max, temp min,	7 setmanes	Walkforward 365	0.18526	12%
MW1	Temp max, temp min, humitat	7 setmanes	Walkforward 365	0.16663	10.82%
MW2	Temp max, temp min, humitat	4 setmanes	Walkforward 365	0.15868	11.74%

A cada apartat s'ha fet una taula resum dels models vistos, i la primera columna és el nom de l'arxiu per trobar el model a l'annex.

5.5 Resum resultats

Hem utilitzat 3 tipus d'agregats de dades per a les prediccions: horari, diari i setmanal. Sempre hem utilitzat el mateix model, una xarxa neuronal LSTM amb una capa d'entrada amb activació relu i una capa de sortida.

5.5.1 Agregat horari

Amb l'agregat horari no hi hem pogut treballar molt ja que tenia execucions molt llargues, de manera que hem fet un únic model utilitzant 48 hores passades de consum com a variables del model, entrenant amb walkforward 14 dies passats (336 hores) i fent test en els 14 dies següents, obtenint un RMSE de 0.012557 i un error percentual mitjà (error respecte valor real) d'un 2.83%.

5.5.2 Agregat diari

Amb l'agregat diari és amb el que hem treballat més i hem fet prediccions a un dia, a set dies, a 30 dies i a 60 dies vista.

5.5.2.1 Predicció a un dia

Els 3 millors models utilitzen walkforward amb train de 365 dies que va augmentant a mesura que es fan les prediccions, tot i que no tots els models d'aquest tipus milloren el walkforward que no augmenta (sempre té train de 365 dies).

El millor model és el que utilitza les variables consum, temperatura màxima, temperatura mínima, humitat i weekday fent servir 14 dies passats, que obté un RMSE mitjà de 0.2920 i un error percentual mitjà de 2.55%.

El segon millor model és el que utilitza les variables consum, temperatura màxima, temperatura mínima i weekday fent servir 14 dies passats, que obté un RMSE mitjà de 0.2972 i un error percentual mitjà de 2.66%.

El tercer millor model és el que utilitza les variables consum, temperatura màxima, temperatura mínima, humitat, pressió, sunsetTimeHour i weekday fent servir 14 dies passats, que obté un RMSE mitjà de 0.2998 i un error percentual mitjà de 2.67%.

En quant a l'error, en tots els models trobem que s'adapten millor al període d'estiu. En el millor model, la predicció de l'estiu no supera el 4% d'error, en canvi a l'hivern té un error que arriba fins al 14% i d'altres al voltant del 9%, però tot i així moltes de les prediccions tenen un error d'entre el 0-1%.

5.5.2.2 Predicció a més de 1 dia

Predicció a 7 dies vista

El millor model és el que utilitza les variables consum, temperatura màxima, temperatura mínima i weekday fent servir 14 dies passats amb el walkforward que augmenta, on obtenim un RMSE mitjà dels 7 dies de 0.05537 i un error percentual mitjà de 4.21%.

Si el que ens interessa és predir el consum de tota la setmana (els 7 dies) aquest és el model amb menys error global. L'error mitjà de predicción del primer dia és de 3.11 i del setè dia de 5.02%. Si mirem el gràfic d'errors del setè dia, veiem com el patró d'error és diferent i que tots els errors han augmentat proporcionalment menys la tardor, que l'error ha augmentat més que proporcionalment.

Si el que volguéssim només és predir el setè dia (i tant ens fan els altres), llavors hi ha el model que utilitza consum, temperatura màxima, temperatura minima i humitat amb walkforward que va augmentant, que tot i tenir un RMSE més alt, de 0.06371, i un error mitjà de 4.81%, l'error per al primer dia és més alt, del 3.75% però l'error per al setè dia és més baix, del 4.95%.

Predicció a 30 dies vista

El millor model és el que utilitza les variables consum, temperatura màxima, temperatura mínima i weekday fent servir 14 dies passats amb el walkforward que augmenta, on obtenim un RMSE mitjà dels 30 dies de 0.06336 i un error percentual mitjà de 4.79%. L'error a un dia és del 3.20% i va augmentant l'error fins la predicción a 30 dies, que és del 6.33%. En aquest cas, el millor model global ja és el model que té un error més baix per al trentè dia.

Predicció a 60 dies vista

El millor model és el que utilitza les variables consum, temperatura màxima, temperatura mínima, weekday, humitat i estació fent servir 14 dies passats amb el walkforward que augmenta, on obtenim un RMSE mitjà dels 60 dies de 0.066145 i un error percentual mitjà de 6.05%. L'error a un dia és del 3.57%, l'error a 30 dies del 5.54% i l'error a 60 dies del 9.29%. Si ens mirem l'error de la predicción a 60 dies gràficament, veiem com les estacions que menys adapta són l'estiu i la tardor (al contrari que la predicción a un dia), i aquests errors poden arribar a ser d'entre el 20 i 40%, no superant el 20% en les altres estacions. Aquest model és el millor model global, és a dir, si volem predir tot el consum a dos mesos vista utilitzariem aquest, en canvi si el que volem és la millor predicción per al seixantè dia, llavors hi ha models millors com per exemple el que utilitza les variables consum, temperatura màxima, temperatura mínima i humitat, que té un error mitjà a 60 dies de 9.12%

5.5.3 Agregat setmanal

Amb l'agregat setmanal hem fet una predicción a 4 setmanes (1 mes) utilitzant les variables consum, temperatura màxima temperatura minima i humitat, obtenint un RMSE de 0.1586 i un error del 11.74%, molt més elevat que la predicción a 30 dies (un mes) en el model agregat diari.

6. Conclusions

6.1 Descripció conclusions

Desglossem aquest apartat en les conclusions que hem extret de l'anàlisi, les conclusions que hem extret de les prediccions i les conclusions generals on expliquem les lliçons apreses en el projecte.

6.1.1 Conclusions anàlisi

Hem vist com el consum diari de les llars de Londres té un patró anual on a l'hivern és on es consumeix més i a l'estiu on es consumeix menys, com també hem vist que el consum entre setmana és més baix que en cap de setmana. A part del dia de la setmana, mes i estació, hem vist com el tipus de consumidor també afecta en el consum diari com també el tipus de tarifa. Altres variables que hem vist que afecten al consum són la hora de posta de sol, les temperatures apparents màximes i mínimes, humitat, pressió i cobertura dels núvols.

En quant al consum horari, hem vist com té un patró diari, on de 8.00 h a 15.00 h es consumeix el mateix (als voltants de 0.21kWh de mitjana cada mitja hora per llar), llavors va pujant el consum fins arribar a les 19.00 h al pic de consum de més de 0.31kWh., i llavors comença a baixar gradualment fins arribar al pic més baix a les 4.00 h (als voltants de 0.11kWh). Aquest patró es veu afectat per la estació i el tipus de consumidor.

Pensàvem que la tarifa afectaria al patró del consum horari i que les llars amb tarifa ToU tindrien un vall més elevat i una punta més baixa que els consumidors amb tarifa Std, però realment ens hem trobat amb que la diferència entre tarifa Std i ToU és que les llars amb tarifa Std consumeixen més que les llars amb tarifa ToU.

Veient que els consumidors que tenen tarifa Std consumeixen més de mitjana que els ToU, sembla que els consumidors amb tarifa ToU estiguin més conscienciatos pel consum, de manera que creiem que si s'apliqués d'una altra manera la tarifa ToU (per exemple amb uns horaris diferents als vistos i que quan hi hagués previsió de més consum augmentés el preu i quan hi hagués previsió de menys consum disminuís el preu, sobretot augmentant el preu en els períodes punta i disminuir el preu en els períodes vall), es podria disminuir el consum en el període punta i augmentar-lo en el vall, intentant així equilibrar la demanda. També es podria utilitzar aquesta política de preus per estimular el consum (baixar el preu) en moments que hi ha molta producció renovable i fer decréixer el consum (augmentar preu) en moments que no hi hagi producció renovable.

Amb la gestió del preu conjuntament a les prediccions de consum i prediccions de generació d'energies renovables es podria fer una bona gestió de la demanda i ajudar a que el consum sigui més verd i eficient.

6.1.2 Conclusions prediccions

Gràcies a l'anàlisi previ de les dades, hem pogut saber quines eren les variables més idònies per a realitzar la prediccó de consum. Depenen si la prediccó és més a curt o a llarg termini, s'adapta millor amb unes o altres variables:

- En la prediccó a un dia vista, el millor model obtingut utilitza consum, temperatura màxima, temperatura mínima, humitat i weekday
- En la prediccó a 7 dies i a 30 dies vista, el millor model és el que utilitza consum, temperatura màxima, temperatura mínima i weekday (sense humitat)
- En la prediccó a 60 dies vista, el millor model és el que utilitza consum, temperatura màxima, temperatura mínima, weekday, humitat i estació

Les variables que es troben a tots els models, i que són les més importants, són el consum, la temperatura màxima, la temperatura mínima i el dia de la setmana (weekday). Veiem que la prediccó a curt (a un dia) s'adapta millor si s'hi introduceix la humitat, i que la prediccó a llarg (a 60 dies) s'adapta millor si s'hi introduceixen les variables humitat i estació.

Amb aquests models hem obtingut uns errors mitjans molt baixos, de 2.55% per la prediccó a un dia vista, 4.21% de mitjana per la prediccó a 7 dies vista, 4.79% de mitjana per la prediccó a 30 dies vista i 6.05% de mitjana per la prediccó a 60 dies vista.

6.1.3 Conclusions generals

Com més coneixem el consum i les variables que l'influeixen, millors prediccions farem, no només és important el model a utilitzar. Com més acurades siguin les prediccions, millor gestió de la demanda es podrà fer.

La prediccó horària, que ens permet saber el consum que hi haurà en una hora en concret, té moltes aplicacions per a la gestió de la demanda a nivell particular.

En primer lloc, es pot fer una política de preus en funció de la demanda prevista per equilibrar el consum de les diferents hores que, com hem vist, varia molt durant el dia. Això és important sobretot quan parlem de les hores pic, ja que el sistema durant aquestes hores ha de generar energia addicional, amb la conseqüència que és energia ràpida però molt contaminant, de manera que si s'aconsegueix rebaixar el consum en aquestes hores, es reduiria significativament la generació addicional. També es important estimular el consum en hores vall, de manera que es pugui aprofitar tota la energia generada per energies renovables. De manera que és un guany econòmic pel consumidor, ja que consumeix en les hores més barates amb el corresponent estalvi i un guany pel sistema i pel medi ambient. En definitiva, la gestió de la demanda permet flexibilitat al mercat.

En segon lloc, l'agregador energètic les pot utilitzar per poder estructurar amb antelació la demanda dels seus clients com també per gestionar la oferta energètica (excedents d'autoconsum dels seus clients). Per exemple, si l'agregador ja sap prèviament el consum dels clients i el que generaran, sabrà amb antelació el consum excedent, podent decidir emmagatzemar-lo en bateries i vendre l'energia en el moment que en donin més preu per exemple o també per exemple decidir no vendre l'excedent ja que se sap que més tard ja es consumirà.

Les prediccions diàries, setmanals i mensuals permeten saber el consum que hi haurà un dia en concret, una setmana en concret o un mes en concret.

Per una banda, això és útil a les comercialitzadores per poder saber l'energia que els cal comprar, permetent-los-hi prendre les decisions de compra i generació d'energia a priori, estalviant-los-hi costos i no haver de recórrer a la compra d'energia a últim moment amb el conseqüent impacte econòmic i ambiental, ja que les centrals de reserva que s'utilitzen quan falta energia són de carbó ja que tenen la capacitat de generar energia al moment, i aquestes tenen moltes emissions de CO₂. De manera que si es sap amb exactitud el consum amb antelació, es disminueix el volum d'energia a generar per aquestes centrals i també s'estalvia ja que s'haurà comprat l'energia en un moment en que hi hagués millor preu.

D'altra banda, és útil per a l'operador de mercat i per a les distribuïdors, ja que sabran amb antelació el volum total que es necessita en un moment determinat i poder-ho planificar. Això també és útil per a la predicción del preu futur, ja que el preu és la cassació de la oferta i la demanda d'energia.

6.2 Assoliment objectius

L'objectiu principal era realitzar una predicción horària, diària, setmanal, mensual i anual fiable del consum de les llars de Londres. Hem realitzat totes les prediccions menys la predicción anual, degut a les poques dades (2 anys).

Volíem aconseguir almenys una precisió del 85% en prediccions a curt termini (horari i diari) i hem aconseguit un error mitjà del 2.83% per a la predicción del consum de la següent hora i un error mitjà del 2.55% per a la predicción del consum del dia següent.

També volíem aconseguir que la precisió de la predicción del consum horari feta el dia abans arribés al 95% (és a dir, volíem aconseguir un error de menys del 5% per a la predicción de les següents 24 hores), però al centrar-nos en la predicción diària, no hem fet aquest tipus de predicción (ja que les dades horàries tardaven molt a executar -en més de 72 hores no havia acabat d'executar- i era difícil treballar-hi i es van descartar).

En la predicción setmanal volíem aconseguir un 80% de precisió i hem aconseguit un error mitjà del 4,21%. En la predicción mensual volíem aconseguir un 75% de precisió i hem aconseguit un 4.79% d'error mitjà.

Les preguntes que ens havíem fet ja es poden respondre:

- Podem saber amb antelació quan es consumirà més energia durant el dia? I menys?
Sí, ho podem saber per una banda gràcies als patrons horaris que hem fet i per altra ho podríem saber si féssim una predicción horària a 24 hores vista.
- El dia abans podem saber amb exactitud l'energia horària que es consumirà?
Sí que podríem saber-ho amb una predicción horària a 24 hores vista.
- Podem saber quanta quantitat d'energia necessitem per un dia en concret?
Sí. Podem saber quant es consumirà al dia següent amb un error del 2.55%, saber quant es consumirà al setè dia següent amb un error del 4.95%, saber quant es consumirà al trentè dia amb un error del 6.33% i saber quan es consumirà al cap de 60 dies amb un error del 9.29%... A mesura que predim un dia més enllà en el temps, l'error va pujant, però tot i així és un error molt acceptable.
- Podem saber la quantitat d'energia que necessitarem un mes en concret?
Sí. Podem saber quant es consumirà al mes següent amb un error de 4.79% o quan es consumirà al cap de dos mesos amb un error del 6.05%.

En quant als objectius secundaris, al ser un pas “entremig” per assolir l'objectiu principal, els hem assolit. Hem vist com afecten al consum les diferents variables, els patrons de corba horaris, diaris, setmanals i anuals, hem vist com l'aplicació de la tarifa Time of Use no tenia repercussió en canviar els hàbits de consum, hem vist que el tipus de consumidor influeix molt en el patró de consum, també la temperatura influeix molt al consum...

Donem per assolit l'objectiu principal de fer prediccions fiables del consum. Els objectius mencionats no assolits han sigut degut a les característiques de la màquina utilitzada i degut a les característiques de les dades.

6.2 Seguiment planificació i metodologia

La planificació s'ha seguit adequadament. Potser s'han dedicat masses hores a la neteja i preparació de dades i al seu anàlisi i llavors les hores pensades per a dedicar a les prediccions no han sigut suficients.

La metodologia ha estat prou adequada. Sí que hagués estat millor revisar les dades que teníem primer per llavors poder-nos marcar uns objectius en funció de les dades i no només en funció dels estudis trobats.

En relació als models plantejats per utilitzar en les prediccions, havíem previst utilitzar xarxes neuronals i SVM però finalment es va decidir centrar-nos en les xarxes neuronals i variar les variables en funció de l'anàlisi previ.

En relació al programari utilitzat, vàrem tenir problemes amb el fitxer amb els consums horaris amb R, ja que al executar qualsevol canvi en el fitxer no teníem prou memòria RAM suficient. Vàrem provar dividint el fitxer, obtenint els mateixos problemes i també creant un dataset virtual, però llavors teníem problemes en que no ens detectava els nombres com a nombres i no ens permetia canviar-ho. Després de dedicar-hi moltes hores sense sortida, vàrem decidir provar de tractar el fitxer en Python, i aquest no va donar problemes. Llavors a l' hora de modelar la xarxa neuronal i executar-la en el fitxer horari, vàrem veure que la execució era molt llarga, ja que després de 72 hores no havia executat, de manera que vàrem decidir no treballar més amb les dades horàries i centrar-nos en les dades diàries.

6.3 Línies treball futur

Degut a que el consum d'electricitat està relacionat amb molts temes i té molts agents participants, seria massa extens realitzar totes les prediccions necessàries per a tots en un sol projecte.

En aquest projecte volíem però no hem pogut explorar:

- Predicció horària a 48 hores vista
- Introducció del tipus de consumidor al model (Acorn o Acorn agrupat)

També seria interessant variar el model, per exemple:

- Provar altres models de LSTM amb diferents paràmetres
- Provar altres models de xarxes neuronals
- Provar altres models , per exemple SVM

Una altre línia seria utilitzar les prediccions com a input d'altres prediccions, per exemple en la figura de l'agregador d'energia, que necessita saber el consum futur per poder gestionar la demanda o per les comercialitzadores, que a partir de el consum futur i de prediccions de generació es podria predir el preu futur de l'energia.

7. Glossari

Acorn	Tipus de consumidor
ANOVA	Anàlisis variància
ARIMA	Auto Regressive Integrated Moving Average
ARIMAX	AutoRegressive Integrated Moving Average with eXogenous inputs
ARMAX-GARCH ARIMA	Generalized AutoRegressive Conditional Heteroskedastic
Dunn	Anàlisis post-hoc ANOVA no paramètric
Friedman	ANOVA no paramètric
Kruskal-Wallis	ANOVA no paramètric
LS-SVM	Least squares suport vector machine
LSTM	Long short-term memory
MLR	Multiple linear regression
PCA	Principal component analysis
RMSE	Root mean square error
SVM	Support Vector MACHines
Std	Tarifa Standard
Test	Conjunt de dades de test
Train	Conjunt de dades d'entrenament
ToU	Tarifa Time of Use

8. Bibliografia

- Achanta, Renuka. «Long Term Electric Load Forecasting using Neural Networks and Support Vector Machines.» *IJCST (International Journal of Computer Science And Technology)* 3, núm. 1 (Jan-March 2012).
- Agrawal, R. K., F. Muchahary, i M. M. Tripathi. «Long term load forecasting with hourly predictions based on long-short-term-memory networks.» *2018 IEEE Texas Power and Energy Conference (TPEC)*. College Station, TX, 2018. 1-6.
- Almalaq, Abdulaziz. «Comparison of Recursive and Non-Recursive ANNs in Energy Consumption Forecasting in Buildings.» *2019 IEEE Green Technologies Conference(GreenTech)*. Lafayette, LA, USA, 2019. 1-5.
- Amjadi, N. «Day-ahead price forecasting of electricity markets by a new fuzzy neural network.» *IEEE Trans. Power Syst.* 21, núm. 2 (May 2006): 887-896.
- Arjmand, Azadeh, Reza Samizadeh, i Mohammad Dehghani Saryazdi. «Improved Forecasting of Short Term Electricity Demand by using Integrated Data Preparation and Input Selection Methods.» *Journal of Energy Management and Technology (JEMT)* 3, núm. 1 (2018): 48-57.
- Asare-Bediako, B., W. L. Kling, i P. F. Ribeiro. «Day-ahead residential load forecasting with artificial neural networks using smart meter data.» *2013 IEEE Grenoble Conference*. Grenoble, 2013. 1-6.
- Bakirtzis, A.G., V. Petridis, S.J. Kartzis, M.C. Alexiadis, i A.H. Maassis. «A neural network short term load forecasting model for the Greek power system.» *IEEE Transactions on Power Systems* 11, núm. 2 (May 1996): 858-863.
- Bianco, Vincenzo, Oronzio Manca, i Sergio Nardini. «Electricity consumption forecasting in Italy using linear regression models.» *Elsevier Energy* 34, núm. 9 (Sep 2009): 1413-1421.
- CACI Ltd. Acorn. 10 / 2019. <https://acorn.caci.co.uk> (últim accés: 05 / 10 / 2019).
- Cai, Huiling, Shoupeng Shen, Qingcheng Lin, Xuefeng Li, i Hui Xiao. «Predicting the Energy Consumption of Residential Buildings for Regional Electricity Supply-Side and Demand-Side Management.» *IEEE Access* 7 (2019): 30386-30397.
- Charytoniuk, W., i M.-S. Chen. «Very short-term load forecasting using artificial neural networks.» *IEEE Transactions on Power Systems* 15, núm. 1 (02 2000): 263-268.
- Chen, Bo-Juen, Ming-Wei Chang, i Chih-Jen Lin. «Load Forecasting Using Support Vector Machines:..» *IEEE TRANSACTIONS ON POWER SYSTEMS* 19, núm. 4 (Nov 2004): 1821-1830.
- Conejo, A.J., M.A. Plazas, R. Espinola, i A.B. Molina. «Day-ahead electricity price forecasting using the wavelet transform and ARIMA models.» *IEEE Transactions on Power Systems* 20, núm. 2 (May 2005): 1035-1042.
- Contreras, J., R. Espinola, F. J. Nogales, i A. J. Conejo. «ARIMA models to predict next-day electricity prices.» *IEEE Trans. Power Syst* 18, núm. 3 (Aug 2003): 1014-1020.
- Cosmo, Valeria Di, Sean Lyons, i Anne Nolan. «Estimating the Impact of Time-of-Use Pricing on Irish Electricity Demand.» *Energy journal. The Quarterly Journal of the IAEE's Energy Economics Education Foundation*. 35, núm. 2 (2014).
- D., Jean Michel. *Energy in the UK - Analysis of smart meter data in London*. 2017. <http://the-odd-dataguy.com/Londonsmartmeter> (últim accés: 09 / 10 / 2019).
- Espinosa, Marcelo, Caroline Joye, Ronnie Belmans, i B. De Moor. «Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series.» *IEEE TRANSACTIONS ON POWER SYSTEMS* 20, núm. 3 (Aug 2005): 1622-1630.

- Ghods, L., i M. Kalantar. «Different Methods of Long-Term Electric Load Demand.» *Iranian Journal of Electrical and Electronic Engineering* 7, núm. 4 (Dec 2011): 249-259.
- Gironés Roig, Jordi, Jordi Casas Roma, Julià Minguillón Alfonso, i Ramon Caihuelas Quiles. *Minería de datos. Modelos y algoritmos*. Barcelona: Editorial UOC, 2017.
- Gkatzikis, Lazaros, Iordanis Koutsopoulos, i Theodoros Saloniidis. «The Role of Aggregators in Smart Grid Demand Response Markets.» *IEEE journal on selected areas in communications* 31, núm. 7 (July 2013): 1-11.
- Goia, Aldo, Caterina May, i Gianluca Fusai. «Functional clustering and linear regression for peak load forecasting.» *International Journal of Forecasting* 26, núm. 4 (Oct-Dec 2010): 700-711.
- Goswami, Kuheli, Ayandep Ganguly, i Arindam Kumar Sil. «Day Ahead Forecasting and Peak Load Management using Multivariate Auto Regression Technique.» *Proceedings of 2018 IEEE Applied Signal Processing Conference (ASPCON)*. 2018. 279-282.
- Greater London Authority and UK Power Networks. *London Datastore - SmartMeter Energy Consumption Data in London Households*. 2015. <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households> (últim accés: 05 / 10 / 2019).
- Gritsay, Aleksandr S., Vladimir V. Makarov, Rustam N. Khamitov, Andrey A. Tatevosyan, i Sergey N. Gritsay. «The Method of Short-term Forecast Electricity Load with Combined a Sinusoidal Function and an Artificial Neural Network.» *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*. Saint Petersburg and Moscow, Russia, 2019. 523-526.
- Guo, Ying-Chun, Dong-Xiao Niu, i Yan-Xu Chen. «Support Vector Machine model in electricity load forecasting.» *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*. Dalian, 2006. 2892-2896.
- H. Hippert, C. Pedreira, R. Souza. «Neural networks for short-term load forecasting: A review and evaluation.» *IEEE Trans. Power Syst.* 16, núm. 1 (Feb 2001): 44-45.
- H.Min., Jae, i Young-chan Lee. «Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters.» *Expert Systems with Applications*, 2005: 603-614.
- Hossen, Tareq, Arun Sukumaran Nair, Radhakrishnan Angamuthu Chinnathambi, i Prakash Ranganathan. «Residential Load Forecasting Using Deep Neural Networks (DNN).» *2018 North American Power Symposium (NAPS)*. Fargo, ND, 2018. 1-5.
- Hsiao, Yu-Hsiang. «Household Electricity Demand Forecast Based on Context Information and User Daily Schedule Analysis From Meter Data.» *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS* 11, núm. 1 (Feb 2015): 33-43.
- Huang, Jianqiang, Ye Liang, Haodong Bian, i Xiaoying Wang. «Using Cluster Analysis and Least Square Support Vector Machine to Predicting Power Demand for the Next-Day.» *IEEE Access* 7 (2019): 82681-82692.
- Ihsane, Imane, Laurence Miegeville, Nadia Aït-Ahmed, i Patrick Guerin. «New Evaluation Metrics for Electrical Demand Forecasting: Application to the Residential Sector.» *2018 AEIT International Annual Conference*. Bari, 2018. 1-6.
- Jason, Wilson, Tao Hong, i Jingrui Xie. «Long Term Probabilistic Load Forecasting and Normalization With Hourly Information.» *IEEE Transactions on Smart Grid*. Volume: 5 , Issue: 1 (Jan 2014): 456 - 462.
- Juberias, G., R. Yunta, J. Garcia Moreno, i C. Mendivil. «A new ARIMA model for hourly load forecasting.» *IEEE Transmission and Distribution Conference*. New Orleans, 1999. 314-319.
- Karabal, Kaan. *BDA503_Course_Project_SPACELondon Smart Meter Analysis Project*. 2017. <https://storage.googleapis.com/kaggle-forum-message->

- attachments/266676/8231/BDA503_Project_SPACE.html (últim accés: 09 / 10 / 2019).
- Kaytez, Fazil, M. Cengiz Taplamacioglu, Ertugrul Cam, i Firat Hardalac. «Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines.» *International Journal of Electrical Power & Energy Systems* 67 (May 2015): 431-438.
- Keitsch, Krischan A., i Thomas Bruckner. «Influence of Demand Response Tariffs on the Electrical Load of Households.» *2015 IEEE Electrical Power and Energy Conference (EPEC)*. London, 2015. 215-219.
- Khan, Sajjad, Nadeem Javaid, Annas Chand, Raza Abid Abbasi, Abdul Basit Majeed Khan, i Hafiz Muhammad Faisal. «Forecasting day, week and month ahead electricity load consumption of a building using empirical mode decomposition and extreme learning machine.» *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2019.
- Kwac, Junsuk, June A Flora, i Ram Rajagopal. «Household Energy Consumption Segmentation Using Hourly Data.» *IEEE Transactions on Smart Grid*, Jan 2014: 420-430.
- Liu, Jinxiang, i Laura E. Brown. «Prediction of Hour of Coincident Daily Peak Load.» *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. Washington, DC, USA, 2019. 1-5.
- Maigha, i M. L. Crow. «Clustering-based methodology for optimal residential time of use design structure.» *2014 North American Power Symposium (NAPS)*. Pullman, WA, 2014. 1-6.
- Mamun, M.A., i K. Nagasaka. «Artificial neural networks applied to long-term electricity demand forecasting.» *Fourth International Conference on Hybrid Intelligent Systems (HIS'04)*. Kitakyushu, Japan, 2004. 204-209.
- Martínez, Alfredo Pastor. *Modelo de predicción de la demanda eléctrica mediante regresión dinámica*. Treball final de grau, Universidad Politécnica de Madrid, 2015.
- Matsuo, Yuji, Kimiya Otani, Tomofumi Shibata, Yasuo Yorita, Yasuaki Kawakami, i Yu Nagatomi. «Short-term Load Forecasting Using Artificial Neural Network.» *IEEJ*, 2019: 1-22.
- Moghram, Ibrahim, i Salfur Rahman. «Analysis and evaluation of five short-term load forecasting techniques.» *IEEE Transactions on Power Systems* 4, núm. 4 (Oct 1989): 1484-1491.
- National Grid ESO. *National Grid ESO*. 09 / 2019. <https://www.nationalgrideso.com/>.
- Niu, D., Y. Wang, i D. D. Wu. «Power load forecasting using support vector machine and ant colony optimization.» *Expert Systems with Applications* 37 (2010): 2531-2539.
- Nogales, F. J., J. Contreras, A. J. Conejo, i R. Espinola. «Forecasting next-day electricity prices by time series models.» *IEEE Trans.Power Syst.* 17, núm. 2 (May 2002): 342-348.
- Ofgem. *Ofgem*. 09 / 2019. <https://www.ofgem.gov.uk/about-us/who-we-are>.
- Osowski, Stanislaw, i Krzysztof Siwek. «Data Mining of Electricity Consumption in Small Power Region.» *2018 19th International Conference "Computational Problems of Electrical Engineering" (CPEE)*. 2018. 1-4.
- Pandya, Rajnikant. «Predicting Domestic Electricity Consumption Using Data Mining for Rural Area of Gujarat.» *IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018*. 2018. 167-173.
- Puspita, Verilly, i Ermatita. *IOP Conf. Series: Journal of Physics: Conference Series* 1196 (2019) 1196, núm. 1 (2019): 1-7.
- R.E.Abdel-Aal, i A.Z.Al-Garni. «Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis.» *Energy* 22, núm. 11 (Nov 1997): 1059-1069.

- Schachter, Jonathan, i Pierluigi Mancarella. «A Short-term Load Forecasting Model for Demand.» *11th International Conference on the European Energy Market (EEM14)*. Krakow, 2014. 1-5.
- Sintov, Nicole D, June A Flora, i Ram Rajagopal. «Utility customer segmentation based on smart meter data: Empirical study.» *IEEE SmartGridComm 2013 Symposium - Support for Storage, Renewable Resources and Micro-grids*. 2013. 720-725.
- Song, Kyung-Bin, Young-Sik Baek, Dug Hun Hong, i Gilsoo Jang. «Short-Term Load Forecasting for the Holidays Using.» *IEEE TRANSACTIONS ON POWER SYSTEMS* 20, núm. 1 (Feb 2005).
- Sun, Wei. «Mid-long term power load forecasting based on MG-CACO and SVM method.» *2010 2nd International Conference on Future Computer and Communication*. Wuha, 2010. V1-118-V1-121.
- Tan, IZhongfu, Jinliang Zhang, Jianhui Wang, i Jun Xu. «Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models.» *Applied Energy* 87, núm. 11 (Nov 2010): 3606-3610.
- The Dark Sky Company, LLC. *Dark Sky API*. 2012. <https://darksky.net/dev> (últim accés: 05 / 10 / 2019).
- Theile, Philipp, et al. «Day-ahead electricity consumption prediction of a population of households: analyzing different machine learning techniques based on real data from RTE in France.» *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 2018. 1-6.
- Twanabasu, Satya Ram, i Bernt A. Bremdal. «Load forecasting in a Smart Grid oriented building.» *22nd International Conference on Electricity Distribution*. Stockholm: CIRED, 2013.
- UK Power Networks. *UK Power Networks*. 2019. <https://www.ukpowernetworks.co.uk/> (últim accés: 05 / 10 / 2019).
- Viegas, Joaquim L., Susana M. Vieira, Sousa João M.C, R. Melício, i V.M.F. Mendez. «Electricity demand profile prediction based on household characteristics.» *2015 12th International Conference on the European Energy Market (EEM)*. Lisbon, 2015. 1-5.
- Wang, Pu, Bidong Liu, i Tao Hong. «Electric load forecasting with recency effect: A big data approach.» *International Journal of Forecasting* 32, núm. 3 (Sep 2016): 585-597.
- Wang, Xin, Fang Fang, Xiaoning Zhang, Yajuan Liu, Le Wei, i Yang Shi. «LSTM-based Short-term Load Forecasting for Building Electricity Consumption.» *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*. Vancouver, BC, Canada, 2019. 1418-1423.
- Wang, Yi, i Songqing Yu. «Annual Electricity Consumption Forecasting with Least Squares Support Vector Machines.» *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*. 2008.
- Wang, Yi, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, i Yun Zhao. «Load Profiling and Its Application to Demand Response: A Review.» *Tsinghua science and technology* 20, núm. 2 (2015): 117-119.
- Wang, Yong, Dawu Gu, Jianping Xu, i Jing Li. «Back Propagation Neural Network for Short-term Electricity Load Forecasting with Weather Features.» *2009 International Conference on Computational Intelligence and Natural Computing*. Wuhan, 2009. 58-61.
- Waters, Liz. «Energy Consumption in the UK. July 2017.» Department for Business, Energy & Industrial Strategy, Govern UK, 2017.
- Waters, Liz, Emily Wilkes, i Victoria GoodRight. *Energy Consumption in the UK (2015). Energia i canvi climàtic*, Govern Regne Unit, Department of Energy and Climate Change, 2015.

Yan, Xing, Dustin Wright, Sunil Kumar, Gordon Lee, i Yusuf Ozturk. «Enabling consumer behavior modification through real time energy pricing.» *First IEEE Workshop on Pervasive Energy Services 2015 (PerCom Workshops)*. 2015. 311-316.

Zeng, Qi, et al. «An optimum regression approach for analyzing weather influence on the energy consumption.» *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. Beijing, 2016. 1-6.

9. Annexos

Els annexes del codi creat amb R (tractament i anàlisi de dades) i del codi creat amb Python (models predictius) es presenten en arxiu adjunt, igual que els arxius de dades utilitzats.

Dins la carpeta de “Dades Final” hi trobem les dades ja tractades i utilitzades pels models de predicción. Les dades originals no s’adjunten degut al seu pes, però es poden trobar a <https://www.kaggle.com/jeanmidev/smart-meters-in-london>.

Dins la carpeta “Python. Models” hi trobem dues carpetes (on a les dues hi ha els mateixos arxius però en diferents formats):

- Carpeta Jupyter notebook: fitxers de cadascun dels models de predicción creats en format ipynb
- Carpeta PDF: fitxers de cadascun dels models de predicción creats en format PDF.

Dins la carpeta “R. Analisi i tractament de les dades” hi trobem dues carpetes (on a les dues hi ha els mateixos arxius però en diferents formats):

- Carpeta HTML: fitxers anàlisi i tractament de les dades en format HTML
- Carpeta RMD: fitxers anàlisi i tractament de les dades en format RMD

Aquests arxius també estan penjats a un repositori GitHub que es pot accedir en el següent link: <https://github.com/lguzmanb/TFM>

Dins la carpeta “Tableau. Visualitzacions anàlisis de dades” hi trobem els fitxers Tableau Extract i Tableau Workbook de les visualitzacions creades en l’anàlisi de dades.

Les visualitzacions creades amb Tableau també es poden trobar en els següents links:

- Visualitzacions analisi horari:
https://public.tableau.com/profile/laura5448#!vizhome/consum_horari/Dashboard1?publish=yes
- Visualitzacions anàlisi diari:
https://public.tableau.com/profile/laura5448#!vizhome/weather_consum_households/DashboardAcornevolucio2?publish=yes
- Visualització tarifa
https://public.tableau.com/profile/laura5448#!vizhome/tarifa_horari/Dashboard1?publish=yes

Quan s’obre el link apareix una visualització, i si es va a baix de tot de la pàgina web, es pot accedir a les altres visualitzacions a través dels links situats a la dreta (Metadata).