

UNIVERSITÀ DEGLI STUDI DI
MILANO-BICOCCA

STREAMING DATA MANAGEMENT AND TIME
SERIES ANALYSIS

FINAL PROJECT

Modellizzazione di una serie temporale da aggregazione dei prezzi sul mercato energetico

Authors:

Louis Fabrice Tshimanga - 847529 -
l.tshimanga@campus.unimib.com

June 19th, 2020



Contents

1	Introduzione	1
1.1	Parametri di progetto	2
1.2	Caratteristiche della Serie	2
2	Modelli ARIMA, sARIMA, XARIMA	2
2.1	ARIMA, sARIMA	2
2.2	XARIMA	4
2.3	Test Forecast	4
3	Modelli a componenti non osservabili	6
3.1	Local Linear Trend	6
3.2	Random Walk "Trend"	6
3.3	Test Forecast	6
4	Modelli neurali	7
4.1	LSTM	7
4.2	Test Forecast	8
5	Predizione finale	9
5.1	Fitting rispetto al dataset	9
5.2	Forecasting	11
6	Conclusioni	14

1 Introduzione

I dati scaturiscono da un'aggregazione giornaliera a criterio ignoto dei prezzi del mercato energetico. Il file è quindi costituito da 2 colonne:

- Data [yyyy-mm-dd]
- Prezzo [€]

Il periodo di riferimento è dal 01-Gen-2010 al 31-Dic-2018, l'obiettivo è predire l'andamento della serie temporale (prezzo giornaliero) dal 1-Gen-2019 al 30-Nov-2019.

1.1 Parametri di progetto

Il dataset comprende 9 anni e richiede di prevedere 11 mesi di serie giornaliera. Per questo motivo si è ritenuto più opportuno utilizzare una suddivisione fra train set e test set di 8 anni e 1 anno rispettivamente (dunque $88.9\% - 11.1\%$), mirando ad acquisire le informazioni sulle caratteristiche annuali della serie, e utilizzando un test set il più simile possibile ai dati ignoti da prevedere (stando alle poche informazioni disponibili), in un compromesso tra stagionalità e uso completo dei dati.

Nella scelta di un particolare modello all'interno di una classe specifica, dove possibile si è preferito quello migliore per indice dell'Akaike Information Criterion (corretto, AICc), ossia un compromesso tra verosimiglianza e parsimonia. Per modelli simili, ai fini della predizione, e per modelli con AICc non disponibile, si sono confrontati sia il Root Mean Square Error (RMSE), sia il Mean Absolute Percentage Error (MAPE), il quale dà una misura facilmente interpretabile: "quanto sbaglia in percentuale, mediamente, il modello". L'ispezione rimane importante, poiché predittori costanti, conservativi e poco informativi dunque, potrebbero ottenere errori minori di modelli capaci invece di catturare movimenti stagionali o frastagliati. Non conoscendo una destinazione d'uso, la valutazione deve essere complessiva.

1.2 Caratteristiche della Serie

La serie in oggetto mostra su piccola scala un andamento settimanale piuttosto riconoscibile. Su ampia scala non è facilmente assimilabile né al White Noise, né ad altri processi stocastici caratteristici. Non c'è evidente relazione tra media e varianza, le quali cambiano di anno in anno, così come cambia la stima della funzione di Autocorrelazione (ACF). I test (aumentati) di Dickey-Fuller restituiscono p-value di valori inferiori all'1%, con statistiche-test più estreme per la serie destagionalizzata per il periodo di 7 giorni, indicando la viabilità per modelli stazionari.

2 Modelli ARIMA, sARIMA, XARIMA

2.1 ARIMA, sARIMA

Un'analisi dell'ACF e della Partial ACF per la serie complessiva evidenzia la stagionalità di periodo settimanale, con i picchi a ritardo-7 che cadono espo-

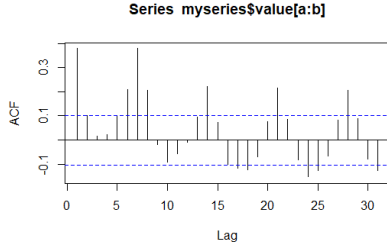


Figure 1: ACF stimata per l'anno 2010

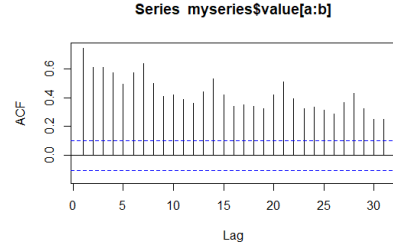


Figure 2: ACF stimata per l'anno 2018

nenzialmente nella PACF e linearmente nell'ACF. Inoltre, la PACF mostra relazioni significative almeno fino all'ottavo ritardo. Differenziando per destagionalizzare, si ottiene una serie a media nulla, con PACF indicativa di un processo AR(3) e ACF indicativa di un processo MA(4). Qui si opera una

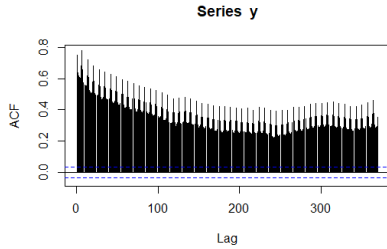


Figure 3: ACF con lag fino a 370, picchi settimanari a caduta quasi lineare

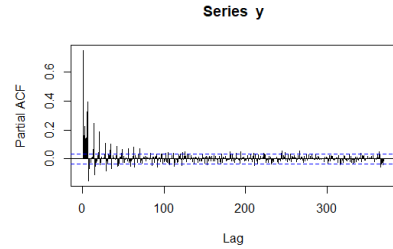


Figure 4: PACF con lag fino a 370, picchi settimanari a caduta esponenziale

grid search nello spazio dei processi $sARIMA(p, d, q)(1, 1, 1)_7$ con p tra 0 – 3 e q tra 0 – 4, provando d 0, 1. Tra questi si mantiene campione il modello con il miglior (minor) valore di AICc. Inoltre, seguendo l'intuizione dalla PACF stimata sulla serie non destagionalizzata, si provano alcuni modelli a memoria "lunga", con AR(6), MA(8) e circostanti, parte stagionale come sopra o annullando autoregressione e media mobile alternativamente. Da questi due processi emergono come migliori i modelli $sARIMA(3, 0, 4)(1, 1, 1)_7$ e $sARIMA(6, 0, 8)(1, 1, 1)_7$. Il modello breve raggiunge un MAPE di 9.63% sul training set, ma prolungando sul test set è evidente come la media nulla e l'assenza di multistagionalità diano previsioni troppo poco variabili. Di fatto

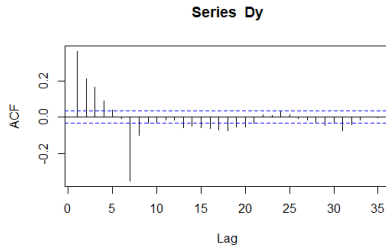


Figure 5: ACF sulla serie destagionalizzata dall'andamento settimanale, che suggerisce AR(3)

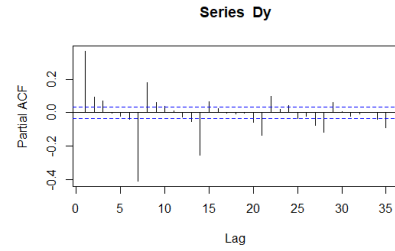


Figure 6: PACF sulla serie destagionalizzata dall'andamento settimanale, che suggerisce MA(4)

questi modelli catturano solo la variabilità stagionale. Il Ljung-Box test rifiuta per entrambi l'ipotesi di non autocorrelazione dei residui. Nonostante l'ACF non suggerisca componenti ARIMA ovvie, è evidente che i modelli non catturino tutto il segnale.

2.2 XARIMA

A questo punto si decide di sviluppare le versioni XARIMA dei due modelli precedenti, inserendo dei regressori trigonometrici. Di fatto si cerca un segnale scomponibile in serie di Fourier, sommato ad un disturbo ARIMA. Il disturbo avrà le caratteristiche dei due campioni precedenti, la serie di seni e coseni sarà troncata ai termini di ordine K , dove K è scelto come prima via grid search dai valori 1 – 25, per esaurire un possibile rapporto con i mesi di un anno e avere abbastanza armoniche per forme d'onda non banali. I nuovi campioni saranno ancora i modelli col miglior AICc. A prescindere dalla forma ARIMA dei disturbi, il numero di regressori ottimo risulta 14. Con questa implementazione si ottiene per il modello migliore (disturbi a lunga memoria) un MAPE sul train set di 9.4%, 14.27% sul test set (il modello breve rimane sopra il 17%).

2.3 Test Forecast

Nonostante il miglioramento numerico ridotto sul train set, l'ispezione delle previsioni mostra chiaramente una miglior attinenza alle variazioni setti-

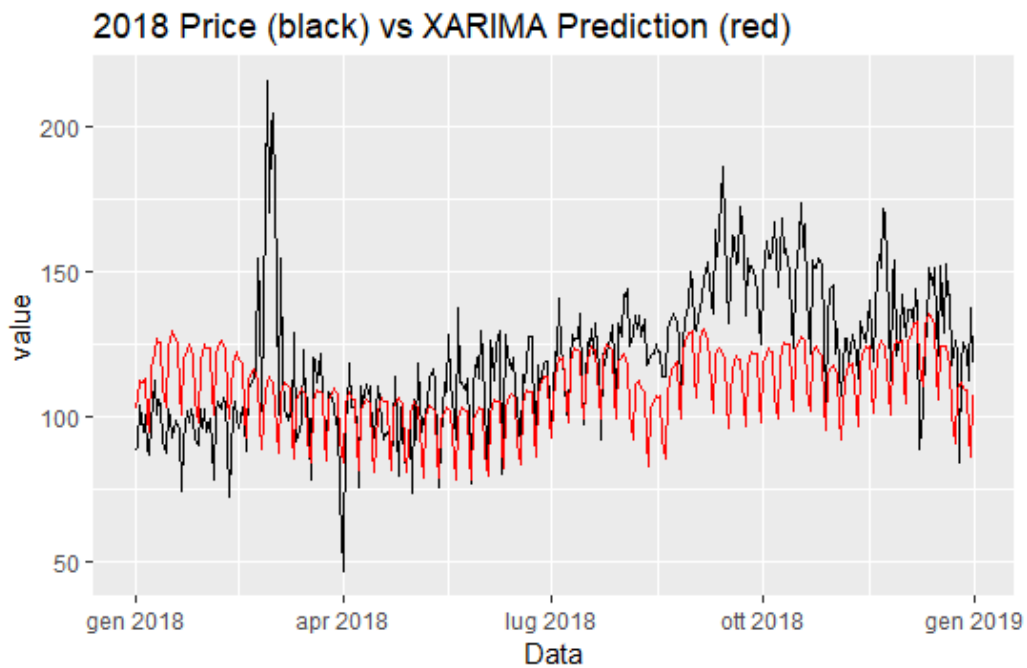


Figure 7: Confronto tra dati del test set e predizione ottenute dal modello $XARIMA(6, 0, 8)(1, 1, 1)_7$ a regressori sinusoidali

manali, mensili e stagionali. Questi modelli dunque colgono dei micro-trend importanti.

3 Modelli a componenti non osservabili

3.1 Local Linear Trend

Vista la sensibilità acquisita precedentemente, si tenta un modello semplice le cui componenti non osservabili siano un trend localmente lineare (LLT) o random walk (RW), una stagionalità settimanale, nonché una variabilità annuale per com'è stata catturata dai regressori nel miglior modello XARIMA. Il modello che usa un trend integrato, considerato ottimo punto di partenza per stime "agnostiche", ha RMSE e MAPE di 17.27 e 12.60% sul train set, sensibilmente maggiori rispetto al campione XARIMA, ma teoricamente potrebbe performare altrettanto bene sul test set. I risultati mostrano però che sul test non c'è miglioramento netto rispetto al meno accurato dei modelli XARIMA (quello a memoria breve).

3.2 Random Walk "Trend"

Ponendo la stessa struttura ma riducendo la componente trend ad un processo random walk, si ottiene un MAPE sul test set quasi identico (17.69% con RW, 17.23% con LLT), e l'unica varianza ignota aggiunta dal LLT non cambia percettibilmente l'onere computazionale dell'ottimizzazione.

3.3 Test Forecast

Pertanto il modello con i regressori mutuati dalla stima XARIMA, un LLT e una stagionalità settimanale sarà considerato campione di categoria, senza aggiunte di stagionalità annuali (che dovrebbero essere catturate dai regressori) o cicli (di cui andrebbe stimata una periodicità che ad ispezione sembra appena colta negli 8 anni di train set, si veda il problema dell'aliasing).

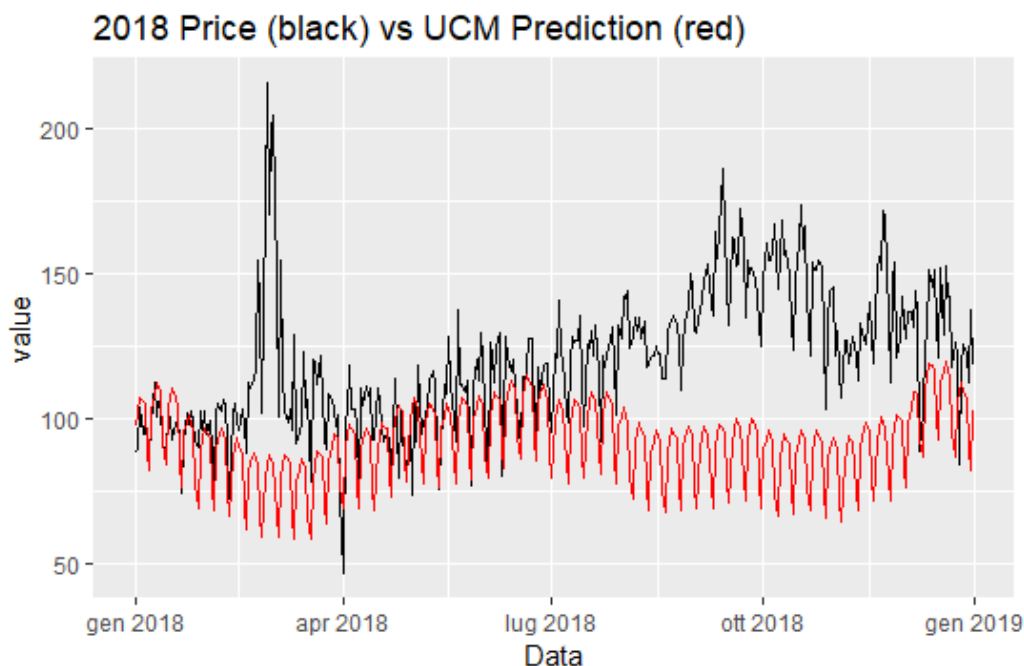


Figure 8: Confronto tra dati del test set e predizioni ottenute dal modello UCM(regressori+LLT+seasonal-7)

4 Modelli neurali

Esistono diversi metodi di Machine Learning applicabili alle serie temporali, tra cui spiccano le reti neurali, nelle architetture comuni come in quelle adattate ai problemi sequenziali, ossia la famiglia delle Recurrent Neural Networks (RNNs). Il vantaggio di questi metodi predittivi è che a fronte di un numero di parametri elevato, l'imputazione riguarda un numero ristretto di iperparametri, utili affinché il modello apprenda soluzioni localmente ottime del problema. Ciononostante le inizializzazioni non sono sempre equivalenti tra loro, come non lo sono gli ottimi globali, ed il costo computazionale dei modelli più performanti va soppesato rispetto alle specifiche del problema.

4.1 LSTM

Nel caso corrente si è ricorso ad una Long Short Term Memory, capace di modulare la quantità di informazione passata nel suo stato, con un modello:

- Input \rightarrow LSTM(200 unità) \rightarrow LSTM(200) \rightarrow Dense (40) \rightarrow Dense(10) \rightarrow Output

dove Input è una sequenza scomposta in tutte le possibili finestre di 112 ritardi della serie (16 settimane, 4 mesi lunari, più di una stagione astronomica), Output è la sequenza delle previsioni in ciascuna finestra. Questa forma ideale per il training è inadeguata per il testing, se confrontato alle metodologie precedenti, ed il forecasting puro: bisogna infatti integrare ogni nuova previsione nella sequenza di Input per generare una serie futura multi-step (esistono modelli il cui output prevede già una finestra determinata). Il modello è stato ottimizzato per 400 epoche su batch di 64 finestre, minimizzando il Mean Absolute Error (il MAPE può portare a denominatori nulli) con l'algoritmo Adam.

4.2 Test Forecast

Con questi iperparametri la rete mostra chiaramente di aver appreso l'andamento settimanale della serie, specie i picchi negativi. Seguono in figura le predizioni con il test set ad Input aggiornato con i dati reali, e con Input aggiornato con i dati predetti, poiché questa sarà la modalità per la predizione finale. Infine

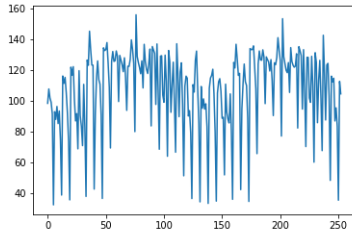


Figure 9: Predizione con input da test set

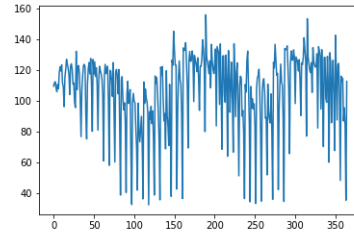


Figure 10: Predizione con input dalle predizioni precedenti

il confronto diretto con il test set per la predizione ricorsiva, con RMSE di 33.21% e MAPE di 19.4% (Loss:1.15e-04).

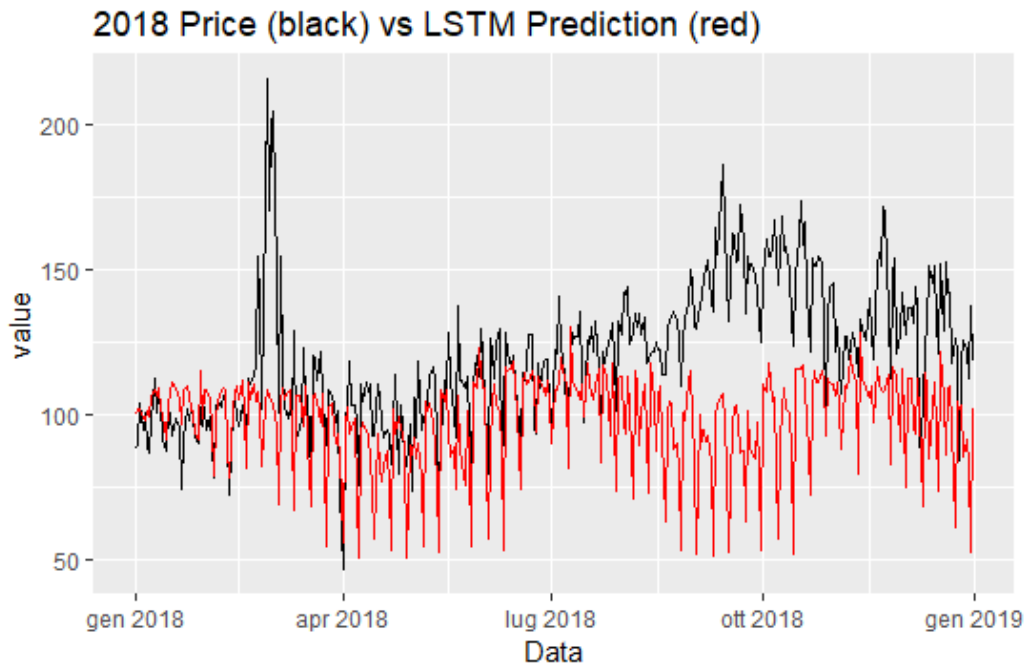


Figure 11: Confronto tra test set e predizione da modello LSTM in ricorsione, o "cieco"

5 Predizione finale

5.1 Fitting rispetto al dataset

Tutti i 3 modelli campioni sono stati ri-stimati sulla totalità dei dati per generare le predizioni degli 11 mesi successivi a quelli del dataset. In questo caso solo i dati di adattamento al training set sono disponibili, e per la LSTM viene riportata soltanto la funzione di perdita con il suo andamento.

- train set XARIMAfinal error: RMSE 14.14; MAPE 9.17
- train set UCMfinal error: RMSE 16.66 MAPE 12.09
- train loss: $9.34e-06$;

Nella storia dell'ottimizzazione, i picchi potrebbero derivare da una superficie degli errori irregolare per la dimensione relativamente ridotta dei batch, o dalla natura dei dati e del modello, ma l'andamento è incoraggiante.

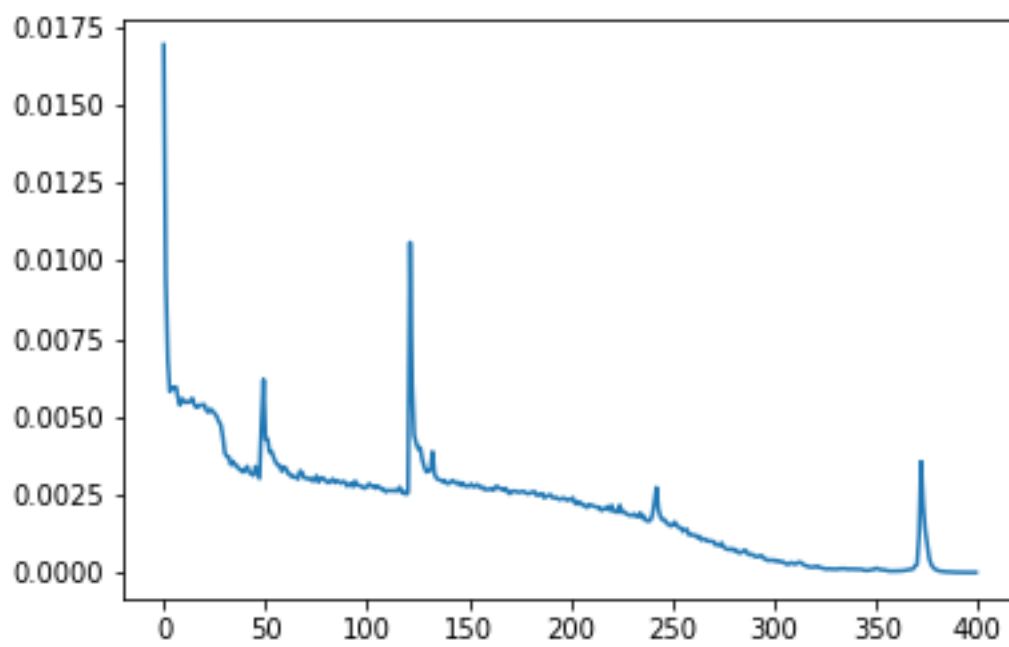


Figure 12: Ottimizzazione della loss function per il modello LSTM su tutto il dataset.

5.2 Forecasting

Serie storiche previste per il periodo 01-Jan-2019 / 30-Nov-2019:

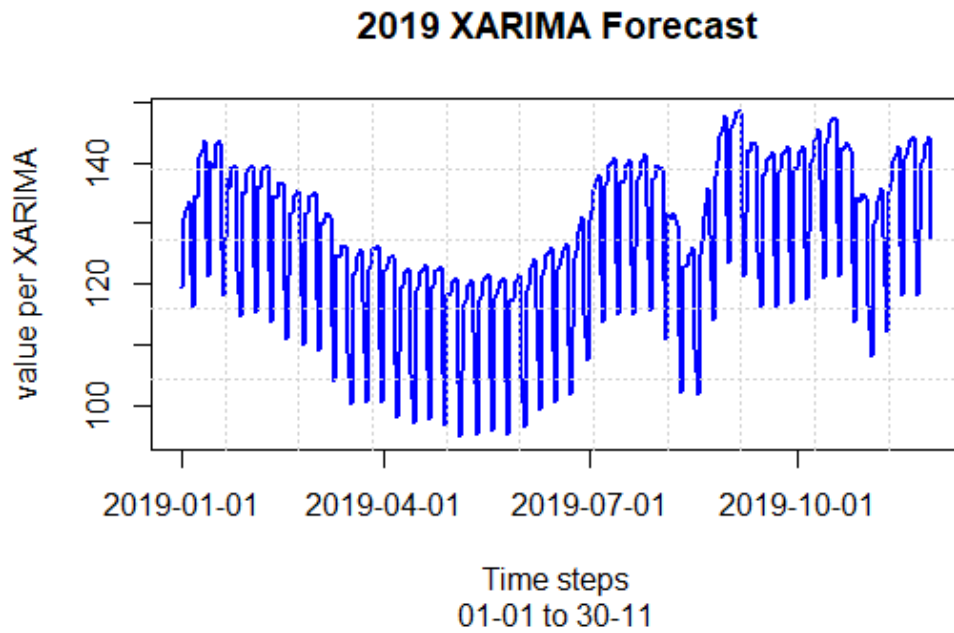


Figure 13: Predizione finale del modello $XARIMA(6,0,8)(1,1,1)_7$ stimato sulla totalità dei dati

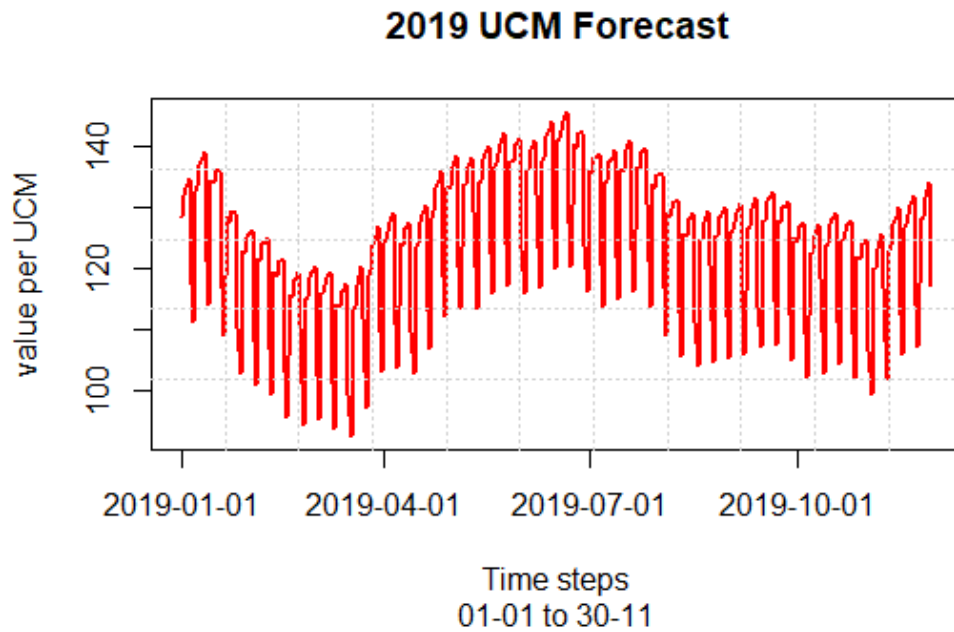


Figure 14: Predizione finale del modello UCM(regressori+LLT+seasonal-7) stimato sulla totalità dei dati

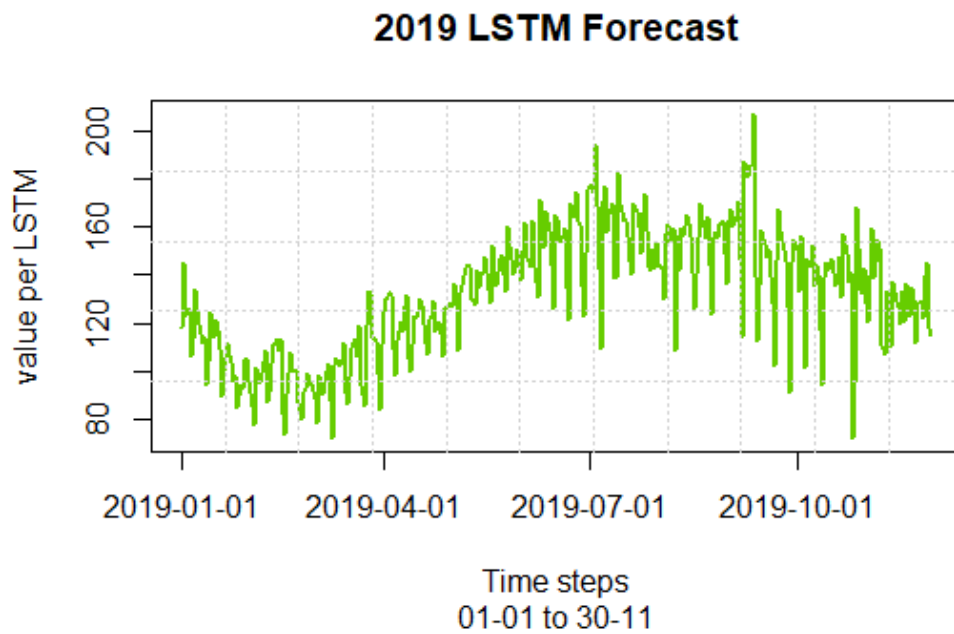


Figure 15: Predizione finale del modello LSTM stimato sulla totalità dei dati

6 Conclusioni

Le previsioni del 2019 dai 3 modelli finali risultano abbastanza simili quantomeno per la depressione dei prezzi nella fase primaverile, anche se c'è uno shift nella risalita dei prezzi. I modelli XARIMA e UCM condividono i regressori che spiegherebbero parte di questa deviazione dei prezzi, ma le serie previste non sembrano particolarmente più simili tra loro che alla serie della LSTM. Quest'ultima sembra ancora più in linea con il dataset, rispetto alla serie test: probabilmente 1 anno in più su 10 è una quantità di informazione rilevante per il modello in questione. L'architettura usata è promettente e quelle simili sono interessanti, soltanto la forte differenza di costo computazionale rispetto ai modelli UCM e (X)ARIMA (in mancanza di conoscenze pregresse sulla serie e sulla destinazione d'uso delle previsioni o l'impatto dei tipi di errore) fa sì che si sia evitata una sperimentazione ulteriore a riguardo.