

1 Analisi serie economica: Previsione della produzione di energia dal vento negli USA

L'analisi della serie economica propone uno studio della produzione di energia elettrica, tramite il vento, negli USA. Il settore dell'energia eolica nasce negli USA negli anni Settanta: il governo, tramite la Nasa, inizia a condurre una ricerca sulle grandi turbine eoliche ai fini commerciali, più precisamente in Ohio, uno degli Stati dove la disponibilità di vento è maggiore. In quegli anni vengono messe in funzione tredici turbine sperimentali, ponendo le basi per lo sviluppo di tecnologie ancora oggi utilizzate¹. Questo settore ha visto un grande sviluppo a partire dagli anni 2009-2010, a seguito di una nuova politica energetica inaugurata dall'amministrazione Obama. Nel 2020, inoltre, sono stati installati, in territorio americano, più impianti di energia eolica rispetto a qualsiasi altra forma di energia. Tale sviluppo è dovuto principalmente alla diminuzione dei costi di produzione degli impianti. Attualmente gli USA sono il secondo paese, dopo la Cina, per la produzione di energia da questa fonte rinnovabile, con una produzione nel 2021 pari a un quinto di quella mondiale. L'obiettivo dell'attuale amministrazione americana è quello di arrivare entro il 2035 a una produzione di energia totalmente proveniente da fonti rinnovabili².

Obiettivo dell'analisi è quello di prevedere la produzione di energia dal vento per un periodo futuro, più precisamente un anno, rispetto all'ultima rilevazione dei dati a disposizione.

I dati per lo svolgimento dell'analisi provengono dalla EIA³, sezione Electricity Data Browser.

I dati originali sono stati rilevati mensilmente, a partire dal dicembre del 2000 fino al novembre del 2022, l'unità di misura riportata è migliaia di MWh. In tutto si osservano 264 rilevazioni. Per lo svolgimento dell'analisi si fa riferimento alla modellizzazione della serie storica mediante *ARIMA*.

1.1 Analisi Preliminare e Modellizzazione

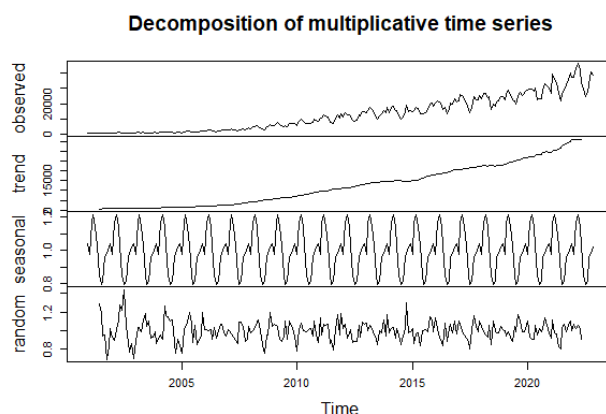


Figure 1: Decomposizione moltiplicativa

La serie storica è inizialmente decomposta in forma moltiplicativa, come mostrato in Figura 1. Si evidenzia la presenza di un trend crescente ed una componente stagionale. Si noti che, a partire dal biennio 2009-2010, la produzione di energia ha registrato un incremento, dovuto probabilmente al cambio di politica energetica, che ha portato nuovi investimenti nel settore. Infine, si possono notare oscillazioni della componente random più accentuate nei primi anni delle rilevazioni, dal 2000 al 2005.

Al fine di determinare la stazionarietà del processo, si procede all'analisi dell'ACF e la PACF⁴. Si evince una non stazionarietà della serie originale, come suggerito dall'andamento lineare decrescente dell'ACF, notando anche la presenza di una componente stagionale. Si certifica la presenza di radici unitarie, causa di non stazionarietà, attraverso i test DF e KPSS. Questi sono in accordo sui dati non differenziati, tuttavia sono in contrasto dopo aver posto $d = 1$. Ciò avviene in virtù del fatto che i due test sono poco potenti e conservativi, facendo sì che entrambi tendano ad accettare la propria ipotesi nulla, che però viene formulata in maniera opposta. Difatti l'ipotesi nulla del DF è la presenza di radice unitaria, contrariamente al KPSS. Si riportano in Tabella 1 valori con i risultati dei due Test⁵:

¹Fonte: Timeline: The history of wind power, The Guardian: <https://www.theguardian.com/environment/2008/oct/17/wind-power-renewable-energy>

²Fonte: Dipartimento Americano per l'Energia: <https://www.energy.gov/eere/wind/wind-market-reports-2021-edition>

³US Energy Information Administration, link alla sezione fonte dei dati: <https://www.eia.gov/electricity/data/browser/>

⁴Rispettivamente, Autocorrelation function e Partial autocorrelation function

⁵I valori inseriti come riferimento sono ricavati al livello di significatività 5%

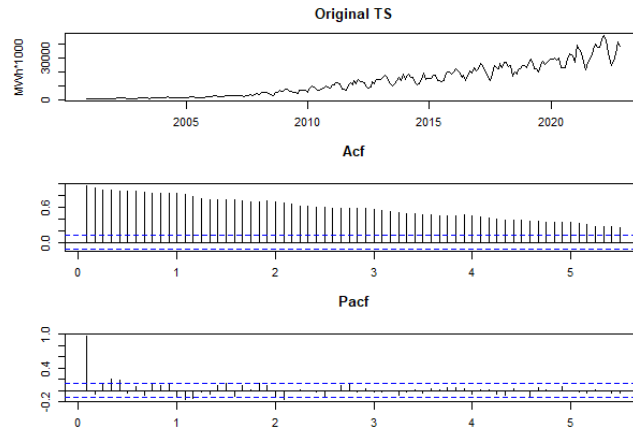


Figure 2: Serie originale, ACF, PACF

	DF stat.	DF rif.	KPSS stat.	KPSS rif.
$d = 0$	2.8443	-1.95	1.680	0.468
$d = 1$	-0.2932	-1.95	0.1778	0.463

Table 1: Risultati dei Test DF e KPSS

Si decide perciò di differenziare i dati di partenza per rimuovere la parte non stazionaria stagionale e non. Si imposta dunque l'ordine $d = 1$ e $D = 1$.

Si osserva dalla ACF in Figura 3 che il decadimento lineare è stato rimosso insieme alla componente non stazionaria stagionale.

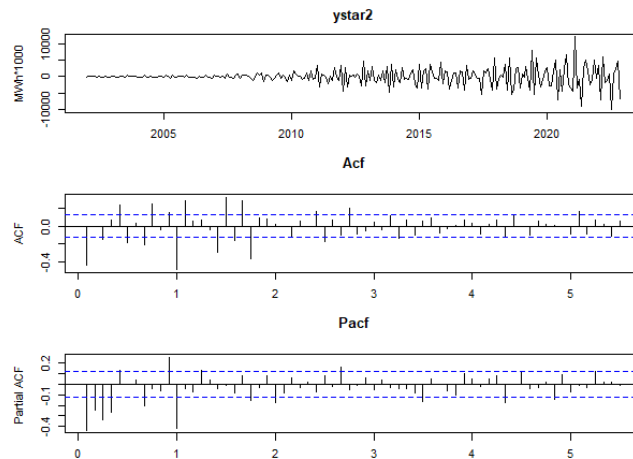


Figure 3: Serie con $d = 1$ e $D = 1$, ACF e PACF

Ora che l'ipotesi di stazionarietà è verificata, si può procedere alla modellazione tramite ARIMA.

Inizialmente si fa ricorso a un modello $ARIMA(0, 1, 0) \times (0, 1, 0)_{12}$, da cui si nota l'eteroschedasticità dei residui, come da Figura 4: per gli anni 2010, la serie dei residui presenta un andamento molto più irregolare rispetto a quello degli anni 2000, come riportato anche dalla serie dei residui al quadrato e in valore assoluto. Si opta per una trasformazione logaritmica, per ricondursi ad una situazione di omoschedasticità.

Dopo aver trasformato la serie applicando il logaritmo, si ottengono residui con un profilo che suggerisce omoschedasticità, o comunque la presenza di un andamento approssimativamente omogeneo lungo tutto il periodo considerato.

Il modello che si propone è un $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ sulle y_t trasformate in scala logaritmica. Si opta per questo modello in quanto ACF e PACF di Figura 5 mostrano una bassa correlazione seriale, o comunque, entro le bande di accettazione.

Si riportano alcuni indici statistici di interesse⁶:

⁶Rispettivamente: Akaike Information Criterion, Akaike Information Criterion corretto, Bayesian Information Criterion

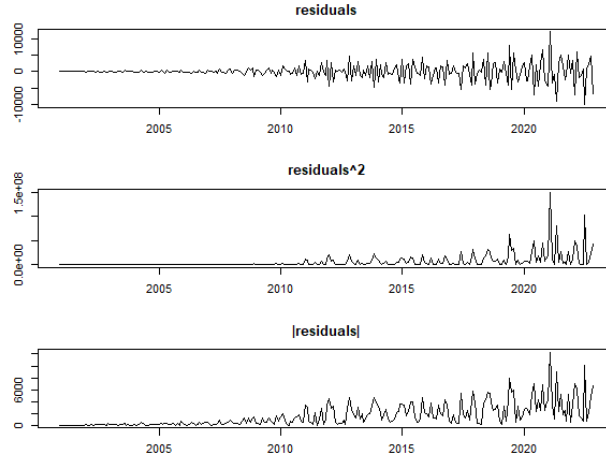


Figure 4: Residui di $ARIMA(0, 1, 0) \times (0, 1, 0)_{12}$

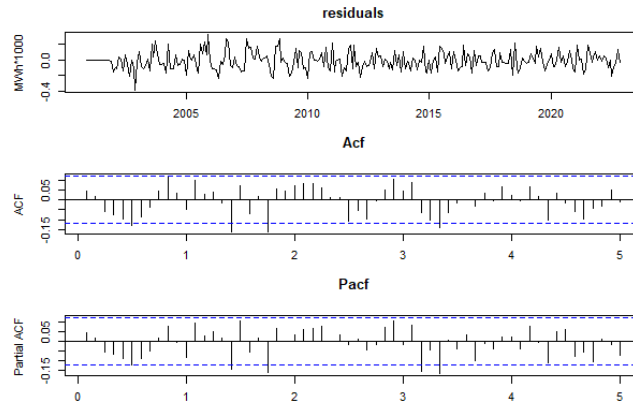


Figure 5: Residui del modello proposto, ACF, PACF

- $AIC = -343.43$
- $AIC_c = -343.33$
- $BIC = -332.85$

I coefficienti associati alle componenti *moving average* e *seasonal moving average* risultano entrambi altamente significativi. Nel modello proposto non si inseriscono variabili di calendario: si assume implicitamente che la produzione di energia eolica sia indipendente dal giorno della settimana. Inoltre, la ricerca di anomalie non ha prodotto alcun risultato.

Dopo aver stimato il modello, si procede con la fase di diagnostica.

Da Figura 6 si osserva che i residui in valore assoluto presentano un andamento approssimativamente omogeneo, mentre i residui al quadrato, per loro costruzione, tendono ad amplificare i valori più elevati. A supporto di ciò si propongono i risultati del Test Arch, a due differenti *lag*. In entrambi i casi, sia per $h = 12$, sia per $h = 24$, i *p-value* associati ai test non risultano significativi.

Come evidenziato nella Figura 7 si rifiuta l'ipotesi di normalità dei residui, affermazione supportata anche dal *p-value* altamente significativo fornito dal test di Shapiro-Wilk. In particolare, dal qq-plot si può notare un cattivo adattamento alla retta teorica, soprattutto sulle code della distribuzione.

1.2 Analisi Previsionale

Conclusa la scelta del modello opportuno e la sua diagnostica, si passa alla fase previsionale.

In ottica della previsione *ex-ante*, si stimano le quantità di energia prodotta tramite eolico nei $H = 12$ mesi successivi all'ultima osservazione rilevata.

In Figura 8 il grafico riporta, oltre alle previsioni, anche le corrispettive bande di previsione. Si nota, inoltre, che la

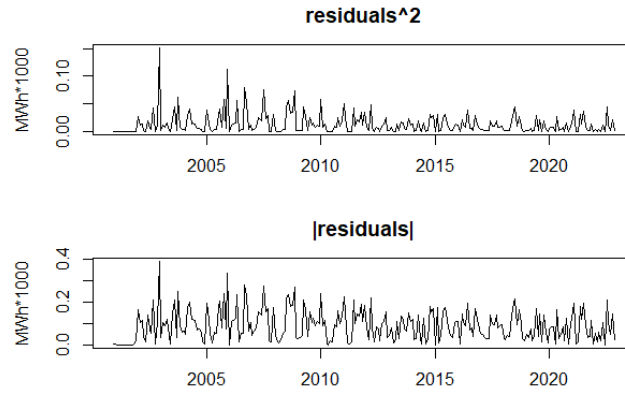


Figure 6: Residui in valore assoluto e al quadrato

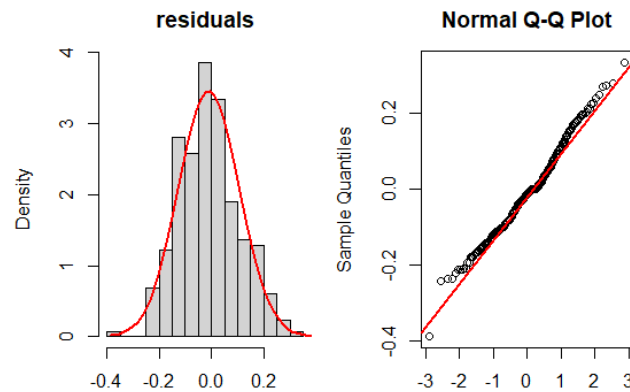


Figure 7: Istogramma dei residui e qq-plot

distanza tra l'estremo superiore e l'estremo inferiore aumenta, all'aumentare dell'orizzonte previsionale. A causa della trasformazione logaritmica dei dati, l'intervallo risultante non è simmetrico rispetto alle previsioni.

Si prosegue l'analisi con una previsione *ex-post* a dodici mesi. In Figura 9 si mostrano le stime a confronto con i valori osservati. Si aggiunge, inoltre, la previsione del modello *naive* per un raffronto con il modello scelto. Si nota, inoltre, che la spezzata del modello *naive* è in prevalenza al di sotto delle altre due curve.

Infine, per fornire un giudizio sul modello stimato in fase di previsione *ex-post*, si introducono delle misure di errore, riportate in Tabella 2. Le misure di errore prodotte dal modello ARIMA sono, come atteso, più piccole rispetto a quelle del *naive*.

Misura di errore	ARIMA	Naive
ME	-6.984e+02	4.675e+03
MAE	2.310e+03	5.539e+03
MPE	-3.166e-02	1.188e-01
RMSE	2.940e+03	6.506e+03
MAPE	7.041e-02	1.498e-01
RMSPE	9.667e-02	1.736e-01
LLE	4.020e-03	2.115e-02

Table 2: Misure di errore previsione *ex-post*

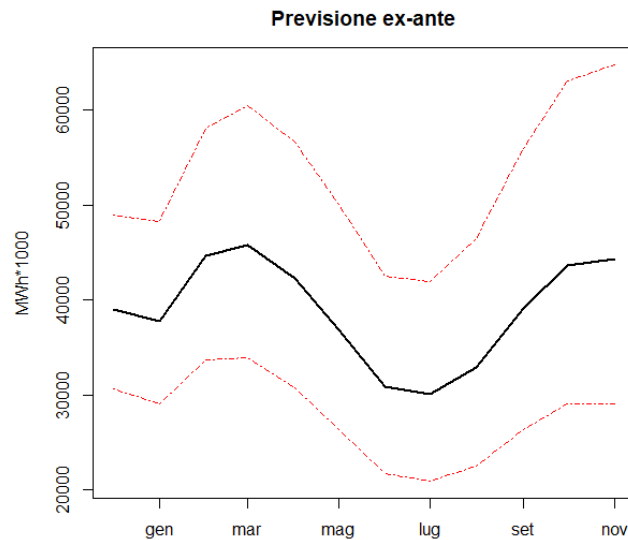


Figure 8: Previsione a dodici mesi

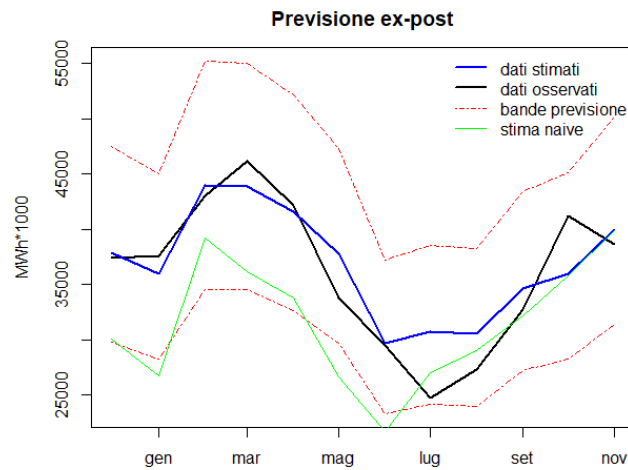


Figure 9: Previsioni per i 12 mesi del 2022

1.3 Conclusioni

Il lavoro svolto è stato presentato in due fasi. Nella prima fase di analisi preliminare, ci si è ricondotti alle ipotesi di lavoro, in seguito si è stimato il modello ARIMA, lavorando su una trasformazione logaritmica dei dati di partenza. Nella seconda parte si è proceduto all'analisi previsionale sia *ex-ante* che *ex-post*.

Le previsioni *ex-ante* mostrano un andamento sinusoidale delle stime, il cui trend appare crescente. Il primo è dato dalla componente stagionale, di cui si parla a partire dalla fase preliminare; il secondo, invece, è in continuità col trend osservato per tutta la serie. Tenendo conto, inoltre, degli obiettivi dell'amministrazione americana, ciò è coerente con l'attuale politica energetica. La variabilità che si accompagna alle previsioni si può attribuire, in parte, a fattori esogeni, quali il tipo di governo e alle leggi che quest'ultimo promulga. Se si catturasse questa variabilità, tramite inserimento di un regressore esterno, sarebbe possibile ottenere stime più accurate.