

# Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components e FbProphet: Applicazione ai consumi energetici

Andrea Afify, Alessandro Maccario, Daniele Mingolla, Alessandro Risaro

*DISCo, Università degli Studi di Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy.*

---

## Abstract

In questo lavoro si sono utilizzati i dataset mensili relativi agli anni 2018, 2019 e 2020 dei consumi energetici degli edifici *U1* e *U6* dell'Università degli Studi di Milano-Bicocca, rilevati ogni 15 minuti, al fine di fornire delle analisi descrittive e predittive e valutare l'impatto che il lockdown dovuto alla pandemia di *SARS-CoV-2* ha avuto su di essi. In particolare, nella fase di preprocessing, viene utilizzato il *filtro di Kalman* per la sostituzione dei valori nulli. L'analisi predittiva viene eseguita utilizzando il modello *TBATS*, il quale tiene conto della multistagionalità presente nei nostri dati e il modello *FbProphet*. Per ogni edificio viene valutata l'influenza delle temperature sui consumi energetici attraverso le *cross-correlations* e, infine, vengono calcolate le *variazioni percentuali* dei consumi durante il lockdown da quelli previsti se non ci fosse stato.

*Keywords:* Consumi Energetici; *TBATS*; *Cross-correlation*; *SARS-CoV-2*; *FbProphet*.

---

## 1. Introduzione

Modelli che siano in grado di fornire risposte affidabili a situazioni di incertezza future rivestono un'importanza sempre più elevata nella società odierna.

L'Università Milano-Bicocca pone particolare attenzione al tema della sostenibilità, come dimostrato dalla sua posizione tra le prime 200 università nella classifica mondiale *Green Metric*, si rimanda al ranking *GreenMetric* (2020).

In questo lavoro si è cercato di fornire una previsione temporale degli andamenti futuri dei consumi di ener-

gia elettrica di due tra i principali edifici dell'Università Milano-Bicocca, l'*U1* e l'*U6*.

Per trovare un buon modello previsivo si è deciso di partire dalla decomposizione della serie temporale dei consumi energetici nelle sue componenti principali (trend, stagionalità e rumore). La decomposizione della serie, infatti, permette di stabilire quali modelli confrontare per fare previsioni permettendo quindi una prima scrematura tra i numerosi algoritmi presenti in letteratura.

Tali risultati aiuteranno la dirigenza nell'attuazione di azioni rivolte all'ottimizzazione energetica a cui conseguirà una maggiore sostenibilità ambientale. A tal fine si sono utilizzati e confrontati principalmente due modelli predittivi per time series: *TBATS* e *FbProphet*.

---

*Email addresses:* a.afify@campus.unimib.it (Andrea Afify), a.maccario1@campus.unimib.it (Alessandro Maccario), d.mingolla@campus.unimib.it (Daniele Mingolla), a.risaro@campus.unimib.it (Alessandro Risaro)

Inoltre, dopo aver testato la bontà dei due modelli, si è cercato di valutare l'effetto del primo lockdown dovuto alla pandemia di *SARS-CoV-2* confrontando i dati di quel periodo con le previsioni numeriche ottenute.

Infine si è ricercata l'eventuale presenza di correlazioni tra i consumi energetici giornalieri e la temperatura nel rispettivo giorno e a diversi lag temporali. Per raggiungere questo obiettivo si sono studiate le *cross-correlations* tra le due serie temporali ottenendo una visione quantitativa del fenomeno.

Il lavoro è stato suddiviso come segue. Nella Sez.2 vengono presentate le finalità di questa ricerca. Nella Sez.3 vengono introdotte le definizioni di base e le quantità statistiche di interesse discutendo i concetti teorici sui quali si fondano i modelli trattati. Nella Sez.4 vengono presentati i dataset a nostra disposizione e le modifiche ad essi apportate. Nella Sez.5 si discute la fase di preprocessing dei dati e le successive analisi statistiche effettuate. Nella Sez.6 vengono presentati i risultati ottenuti e le principali informazioni estrapolate dall'analisi dei dati. Nella Sez.7 infine vengono presentate le nostre conclusioni, i limiti riscontrati e i possibili sviluppi migliorativi che si possono apportare a questo lavoro.

## 2. Obiettivi

L'intento primario di questo lavoro è quello di fornire una rappresentazione descrittiva e delle analisi predittive basate sui dati relativi ai consumi energetici degli edifici *U1* ed *U6* dell'*Università di Milano-Bicocca*. In primo luogo sono state estratte tutte le informazioni utili che permettessero di comprendere lo stato attuale dell'andamento energetico di entrambi gli edifici ottenendone una loro comparazione, poi si sono ricercate

eventuali differenze significative. Per l'università, tale aspetto risulta essere importante tanto da un punto di vista ambientale quanto da un punto di vista economico. A questa prima fase descrittiva è stato affiancato uno studio della correlazione tra il consumo energetico e i dati meteorologici della zona *Milano-Bicocca* nello stesso periodo, studiando eventuali influenze della temperatura esterna sul dispendio energetico dei due edifici. Infine, sono state svolte analisi predittive di un anno in avanti rispetto ai dati relativi al periodo più recente (2020) e delle previsioni sul periodo segnato dal primo lockdown dovuto alla pandemia di *SARS-CoV-2* in modo da rapportare i consumi effettivi con quelli che si sarebbero ottenuti in mancanza della pandemia.

## 3. Aspetti metodologici

I modelli utilizzati sono due in particolare: il modello *TBATS* e il modello *FbProphet*, sviluppato da *Facebook*. La scelta nell'utilizzo di tali strumenti è stata resa necessaria dall'incapacità dei più semplici modelli *ARMA* e *ARIMA* (il quale, per esempio, è in grado di valutare solamente una stagionalità) di considerare componenti di stagionalità multiple incluse nelle serie temporali (giornaliere, settimanali, mensili, annuali) a disposizione.

### 3.1. Modello TBATS

Il nome *TBATS* è un acronimo che rappresenta le principali caratteristiche del modello: **T**rigonometric seasonality, **B**ox-Cox transformation, **A**RMA errors, **T**rend and **S**easonal components.

Uno dei principali vantaggi di questo modello è quello di poter gestire stagionalità multiple nei dati e di riuscire a gestire effetti stagionali di lunghezze non intere, utile

al presentarsi per esempio di anni bisestili. In questa metodologia statistica ogni stagionalità è modellata su una rappresentazione trigonometrica basata sulla *serie di Fourier*.

Il modello ad un parametro può essere così rappresentato (si veda De Livera et al. (2011)):

$$y_t^{(w)} = \begin{cases} y_t^\omega - 1/\omega & \omega \neq 0 \\ \ln y_t & \omega = 0 \end{cases} \quad (1)$$

$$y_t^\omega = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t \quad (2)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \quad (3)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t \quad (4)$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t \quad (5)$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-1} + \sum_{i=1}^q \theta_i \epsilon_{t-1} + \epsilon_t \quad (6)$$

dove:

- l'equazione 1 rappresenta la trasformazione di *Box-Cox*;
- l'equazione 2 rappresenta l'andamento stagionale;
- l'equazione 3 rappresenta il trend locale;
- l'equazione 4 rappresenta il trend globale;
- l'equazione 5 rappresenta la  $i$ -esima componente stagionale al tempo  $t$ ;
- l'equazione 6 rappresenta un processo  $ARMA(p, q)$  dove  $\epsilon_t$  è un rumore bianco Gaussiano con media pari a zero e varianza costante  $\sigma^2$ .

### 3.2. Modello *FbProphet*

L'obiettivo primario del modello *FbProphet* [Taylor and Letham (2018)] è quello di semplificare il lavoro dell'analista il quale, conoscitore di dominio, non ha abilità professionali dal punto di vista dell'analisi delle serie storiche, o dei differenti metodi e modelli a disposizione. Infatti, tale approccio modellistico e implementativo ha il pregio di fornire parametri intuitivi alla comprensione dell'utente finale, rendendone il *tuning* semplice da applicare anche per l'utilizzatore non esperto nel modello applicativo sottostante.

La sua formulazione più generale è tale che:

$$y(t) = g(t) + s(t) + h(t) + e_t \quad (7)$$

dove:

- $g(t)$  è la funzione di **trend** (*tasso di crescita*) che modella i cambiamenti non periodici della serie temporale (il **trend** di una funzione a tratti);
- $s(t)$  rappresenta i cambiamenti periodici (stagionalità giornaliera, settimanali, mensili e annuali);
- $h(t)$  rappresenta i periodi di vacanza che avvengono irregolarmente durante uno o più giorni a seconda dell'anno;
- $e_t$  rappresenta infine ogni cambiamento che il modello non è in grado di catturare, ovvero l'**errore** (ipotizzato normalmente distribuito).

Tale modello viene definito simile ai modelli **GAM** (*Generalized Additive Model*), (si veda, e.g., Taylor and Letham (2018)), ovvero una classe di modelli di regressione nei quali vengono incorporate variabili predittive non lineari. Infatti, ad esempio, il trend potrebbe essere considerato:

- Non lineare: per problemi di previsione nei quali la capacità massima di crescita del parametro  $C$  non può essere definita, essa viene indicata come dipendente dal tempo  $t$ . Inoltre, il tasso di crescita non è costante. In tal caso, si avrà che:

$$g(t) = \frac{C(t)}{1 + e^{(-(k+a(t)^T \delta)(t-(m+a(t)^T \gamma)))}} \quad (8)$$

- Lineare, non costante: per problemi di previsione nei quali non si ha una saturazione nel parametro  $C$ , viene usato un tasso costante di crescita nella funzione. Il trend viene quindi definito come:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (9)$$

dove:

- $k$  è il tasso di crescita;
- $\delta$  è il tasso di aggiustamento;
- $m$  è il parametro di offset;
- $\gamma_j$  viene impostato pari a  $s_j \delta_j$  per rendere la funzione continua.

L'incertezza del trend futuro viene modellata considerando costante il suo tasso: si assume quindi che in futuro avremo la stessa frequenza e grandezza media nel tasso di cambiamento che è stato visto nei dati del passato.

In merito alla componente di stagionalità, essa viene specificata come serie di Fourier:

$$s(t) = \sum_{n=1}^N (a_n \cos(\frac{2\pi n t}{P}) + b_n \sin(\frac{2\pi n t}{P})) \quad (10)$$

Si stimano quindi i  $2N$  parametri definiti come:

$$\beta = [a_1, b_1, \dots, a_N, b_N]^T. \quad (11)$$

e costruendo una matrice composta da vettori di stagionalità per ciascun valore di  $t$  nei dati storici e futuri.

Infine, riguardo le festività, è l'utente ad inserirle tramite una lista di giorni di vacanza. Viene utilizzata una funzione indicatrice che rappresenta quando il tempo  $t$  si trova durante un giorno festivo  $i$ , al quale viene aggiunto un parametro  $k_i$  corrispondente al cambio nella previsione. Tale obiettivo viene raggiunto generando una matrice di regressori  $Z(t)$ , ottenendo:

$$h(t) = Z(t)k \quad (12)$$

L'effetto delle vacanze viene considerato usando una finestra intorno ad una particolare festività: i giorni vicini sono anch'essi definiti tali.

L'utilizzo di tale ulteriore modello oltreché del *TBA-TS* è stato reso necessario dall'incapacità di questo ultimo di modellare accuratamente la serie temporale dell'edificio *U6*. La scelta quindi, dopo alcuni test, è ricaduta su *FbProphet*.

Dopo aver cercato di fare delle previsioni sulla serie storica dei consumi si è voluto verificare quanto la temperatura giornaliera influisse su di essi. Per entrambi gli edifici si è quindi verificato il grado di correlazione tra le temperature giornaliere, calcolate come media sulle 24 ore, e la colonna *Consumo\_attiva\_prelevata*. A tal fine si sono calcolate numericamente le *cross-correlations* tra le due serie temporali, usando la funzione *acf* in R (si veda, e.g., Venables and Ripley (2002)).

Le *cross-correlations*, dati due segnali a valori reali  $x$  e  $y$  che differiscono solamente per uno spostamento sull'asse  $t$ , permettono di calcolare la correlazione incrociata per mostrare di quanto  $y$  deve essere anticipato

per renderlo identico ad  $x$ . Per valori discreti e serie temporali finite valgono le seguenti considerazioni:

Si assuma che la serie  $X_t$  sia infinita, ma la si osservi solo per valori discreti finiti:  $t = 1, \dots, n$ . La serie ha covarianza  $\gamma_t$  e correlazione  $\rho_t$  date rispettivamente da:

$$\gamma_t = \text{cov}(X_{t+\tau}, X_\tau) \quad (13)$$

$$\rho_t = \text{corr}(X_{t+\tau}, X_\tau) \quad (14)$$

non dipendenti da  $\tau$ .

La covarianza per  $t > 0$  è stimata dalle  $n - t$  coppie ordinate  $(X_{1+t}, X_1), \dots, (X_n, X_{n-t})$ . Se si prende la correlazione o la covarianza standard si usano differenti stime di media e varianza per ognuna delle sottoserie  $X_{1+t}, \dots, X_n$  and  $X_1, \dots, X_{n-t}$ , mentre sotto ipotesi di stazionarietà del secondo ordine queste dovrebbero avere stessa media e stessa varianza. Si introduce quindi il seguente stimatore:

$$c_t = \frac{1}{n} \sum_{s=\max(1, -t)}^{\min(n-t, n)} [X_{s+t} - \bar{X}][X_s - \bar{X}] \quad (15)$$

Inoltre si ha:

$$r_t = \frac{c_t}{c_0} \quad (16)$$

La funzione *acf* calcola e plotta di default sia  $c_t$  che  $r_t$  note come funzioni di autocovarianza e autocorrelazione. Queste definizioni sono facilmente estendibili a più serie temporali osservate nello stesso intervallo temporale. Ottenendo:

$$\gamma_{ij} = \text{cov}(X_i(t + \tau), X_j(\tau)) \quad (17)$$

$$c_{ij} = \frac{1}{n} \sum_{s=\max(1, -t)}^{\min(n-t, n)} [X_i(s + t) - \bar{X}_i][X_j(s) - \bar{X}_j] \quad (18)$$

che non sono simmetriche in  $t$  per  $i \neq j$ .

## 4. I dati

### 4.1. Descrizione del dataset

Per rispondere alle domande di ricerca poste durante la stesura di tale lavoro, sono stati considerati i seguenti dati:

- Per gli edifici *U1* e *U6*, i consumi energetici dell'*Università degli studi di Milano-Bicocca*;
- Le previsioni metereologiche riferite alla città di *Milano*.

#### 4.1.1. Dati Università Milano-Bicocca

Per l'intera analisi sono stati utilizzati i dati relativi agli anni 2018, 2019 e 2020 per entrambi gli edifici. Ogni file è suddiviso nei rispettivi mesi dell'anno corrispondente con un livello di dettaglio giornaliero e, nello specifico, con una misurazione delle informazioni ogni 15 minuti. Si è reso necessario un primo controllo e una pulizia o manipolazione dei dati (ad esempio, con la standardizzazione della colonna *ORA* nel formato *HH:MM:SS* (ad esempio, 00:15:00) per una corretta interpretazione e utilizzo in fase di aggregazione) per poi procedere all'unione dei diversi dataset in due file da cui partire con le successive analisi.

Le variabili di interesse per entrambi i dataset sull'energia sono:

- **POD**: identificativo del contatore di energia;
- **DATA**: corrispondente al giorno in cui viene effettuata la rilevazione;
- **ORA**: l'orario in cui viene effettuata la rilevazione (considerando come elemento di partenza la mezzanotte del primo giorno del mese. I dati sono poi ottenuti ogni 15 minuti);

- **FL\_ORA\_LEGALE:** in riferimento alle rilevazioni effettuate durante l'ora legale o durante l'ora solare;
- **CONSUMO\_ATTIVA\_PRELEVATA:** valore di kW nei 15 minuti considerati;
- **CONSUMO\_REATTIVA\_INDUTTIVA\_PRELEVATA:** consumo energetico (espresso in kW\h) prelevato ma non consumato direttamente dall'utilizzatore;
- **POTENZA\_MASSIMA:** limite massimo di potenza del contatore (espresso in kW) prelevabile dalla rete telefonica;
- **TIPO\_DATO:** identifica la tipologia dei dati.

#### 4.1.2. Dati Meteo

I dati meteorologici sono stati acquisiti attraverso l'*API Meteostat*, la quale permette di accedere a diversi dati storici relativi alle condizioni meteorologiche di una determinata zona specificata attraverso le coordinate di longitudine e latitudine. I dati vengono forniti dall'*API* ad una granularità oraria e sono stati scaricati in formato *csv*. Le variabili d'interesse sono:

- **time:** identificativo dell'ora di rilevazione;
- **temp:** temperatura dell'aria espressa in °C;
- **rhum:** L'umidità relativa espressa in percentuale (%);
- **wspd:** La velocità del vento espressa in km\h;
- **coco:** Codice identificativo delle condizioni meteorologiche.

## 4.2. Data Quality

### 4.2.1. Dati Consumi energetici

- **Accuratezza:** l'attendibilità delle informazioni energetiche si suppone essere elevata in quanto i dati utilizzati derivano da rilevazioni effettuate tramite strumentazione caratterizzata da errori di misura non significativi;
- **Completezza:** i principali dati mancanti si sono rilevati in egual modo nella variabile *ORA* e in *CONSUMO\_ATTIVA\_PRELEVATA*. Per la risoluzione di tale inconveniente, si è proceduto alla scrittura di alcuni snippet di codice in grado di integrarli e imputarli. Inoltre, si presentavano totalmente mancanti i dati relativi al mese di giugno 2020 per l'edificio U6;
- **Consistenza:** I dataset originali relativi ai due edifici presentavano una struttura pressoché identica in tutti i singoli file disponibili, garantendo la consistenza delle informazioni. Il problema immediatamente individuato è stato relativo al mese di giugno 2020, come già specificato e visibile dalla Figura 1. In particolare in questo periodo, i dati risultavano identici tra l'edificio U1 e U6. A seguito di successive analisi, si è compreso come i dati fossero riferiti al primo edificio, rendendo ridondanti quelli del secondo.

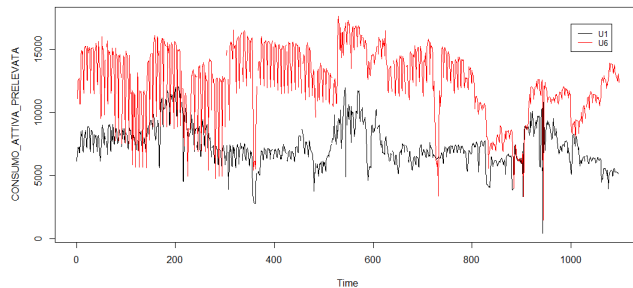


Figura 1: Confronto dei dati originali degli edifici U1 (linea nera) e U6 (linea rossa)

#### 4.2.2. Dati meteo

- **Accuratezza:** I dati acquisiti da Meteostat provengono da diverse interfacce pubbliche, tra cui molti di questi sono forniti dalle autorità governative come *The National Oceanic and Atmospheric Administration* (NOAA) e *Germany's national meteorological service* (DWD);
- **Completezza:** Si è riscontrato un problema di completezza dei dati, in quanto vi erano 42 righe che presentavano *NaN* per tutti gli attributi d'interesse;
- **Consistenza:** Si è riscontrato anche un problema di consistenza dei dati, in quanto il giorno 3 Gennaio 2018 alle ore 12 PM è stata rilevata una variazione di temperatura di 30 gradi rispetto alla rilevazione precedente.

## 5. Analisi e trattamento dei dati

### 5.1. Data Cleaning

#### 5.1.1. Dati consumi energetici

I problemi riscontrati riguardo i dati energetici hanno fatto riferimento a:

- **Valori mancanti:** nel caso dei *missing value* relativi alla colonna **ORA**, sono stati rilevati quarti d'ora mancanti e imputati tramite codice in *Python*. Per le colonne rimanenti, è stato scelto il metodo di imputazione di *Kalman* dopo averlo considerato migliore in comparazione ad altri metodi valutati. Per fare ciò, sono state confrontate differenti metodologie di inserimento dati (ad esempio: interpolazione, media, *last observation carried forward* (locf)). Tale confronto è stato eseguito creando dei valori nulli casuali rispetto ai dataset originali (come se fossero valori da imputare realmente); infine, tramite diverse misure d'errore (ad esempio, il MAPE - *Mean absolute percentage error*), il metodo di *Kalman* è risultato tra i migliori in questi termini, come visibile dalla Tabella 1.

| Metodo                  | MAPE U1 | MAPE U6 |
|-------------------------|---------|---------|
| Interpolazione          | 1.8     | 1.2     |
| Kalman                  | 1.9     | 1.2     |
| Seadec                  | 2.1     | 2.5     |
| Media                   | 9.3     | 25.5    |
| Locf                    | 2.1     | 1.6     |
| Seasplit                | 3.9     | 10.9    |
| Weighted Moving Average | 1.9     | 1.5     |

Tabella 1: Valori MAPE del test per l'imputazione dei valori mancanti

- **Valori nulli:** anche i valori pari a 0 sono stati considerati come valori mancanti. Per questo motivo, sono stati convertiti in valori *NA* per poi imputarli col metodo sopracitato. Ciò ha permesso di correggere gli errori derivanti da eventuali mancate rilevazioni dello strumento fisico.

Per quanto riguarda i dati mancanti del mese di giugno 2020, si è scelto di utilizzare il modello *TBATS* al-

lenato sui dati precedenti a tale periodo. I valori predetti sul mese di *giugno* da parte di questo modello sono stati utilizzati per sostituire i dati nulli di *U6* relativi allo stesso arco temporale.

### 5.1.2. Dati meteo

Si è proceduto alla sostituzione dei valori *NaN* per le colonne d'interesse utilizzando il metodo di imputazione di *Kalman* implementato attraverso la libreria *imputeTS* di *R*. Per quanto riguarda la riga che presentava un valore di temperatura non consistente, questa è stata sostituita con la media tra il valore di temperatura della rilevazione precedente e della successiva.

### 5.2. Data Integration

I dataset riguardanti i consumi energetici di *Bicocca* e i dati meteorologici sono stati integrati attraverso un'operazione di **join** sulla colonna **Data** e la colonna **time**. In quanto le due colonne presentavano rilevazioni ad una differente granularità, infatti i dati relativi ai consumi erano registrati ogni 15 minuti, mentre i dati relativi alle condizioni meteo venivano fornite ogni ora; si è proceduto all'aggregazione dei primi. La colonna *Consumo\_reattiva\_induttiva\_prelevata* e *Consumo\_attiva\_prelevata* sono state aggregate tramite la somma, mentre la colonna *Potenza\_massima* è stata aggregata utilizzando la media.

Al fine di rispondere agli obiettivi di questo lavoro si è deciso di fare un'ulteriore aggregazione dei dati, arrivando ad un livello di definizione giornaliero. I dati relativi ai consumi energetici sono stati aggregati come sopra riportato, mentre i dati relativi al meteo sono stati aggregati attraverso la media.

### 5.3. Analisi Predittiva

L'analisi predittiva effettuata sui consumi elettrici dei due edifici ha riguardato l'anno 2021 ed è stata svolta attraverso l'utilizzo del modello *TBATS*. Nelle due figure successive, sono riportati i risultati ottenuti rispettivamente per l'edificio *U1* e l'edificio *U6*:

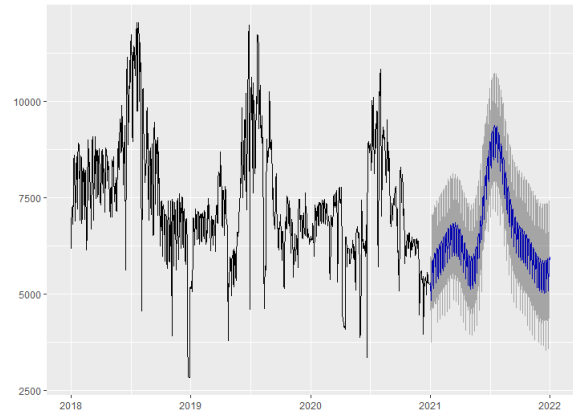


Figura 2: In Figura è mostrata la previsione sui consumi per l'edificio *U1* relativa all'anno 2021 ottenuta tramite il modello *TBATS*.

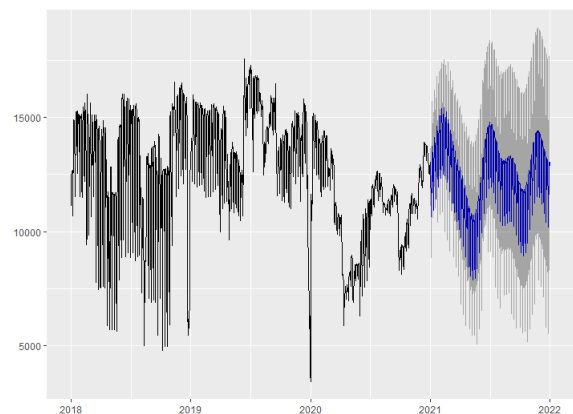


Figura 3: In Figura è mostrata la previsione sui consumi per l'edificio *U6* relativa all'anno 2021 ottenuta tramite il modello *TBATS*.

Come è possibile notare dalle figure precedenti, il modello *TBATS* è in grado di modellare la multistagionalità dei nostri dati e di fornirci con un grado di affidabilità molto elevato, i possibili andamenti futuri dei con-



sumi. Nel grafico di *U1*, il picco dei consumi nell'anno 2021 avviene attorno a metà *luglio* rimanendo coerente a quanto accaduto nello stesso periodo negli anni precedenti. Per quanto riguarda il grafico dei consumi di *U6*, notiamo come il modello preveda un incremento nei primi mesi del 2021 che tenderà a diminuire fino a raggiungere un picco di minimo attorno a fine aprile per poi risalire nel mese di *giugno*. Tali informazioni risultano essere importanti nell'analisi dell'efficienza energetica dei due edifici in quanto denotano dei livelli di consumi particolarmente diversi nonostante il periodo temporale sia lo stesso. Sarebbe necessario avere a disposizione ulteriori dati per poterne comprendere le cause.

#### 5.4. Selezione del modello

Per valutare se la scelta del modello *TBATS* fosse da considerarsi adeguata, si è deciso di valutarne la capacità previsiva rispetto a dati già in possesso, come ad esempio quelli dell'anno 2019. Per fare ciò, si sono utilizzati come dati di training i valori relativi al 2018, prevedendo poi quelli dell'anno successivo, come visibile in Figura 4 (la preferenza è ricaduta su tale anno in particolare, e non sulla previsione nell'anno 2020 a causa dell'impatto della pandemia).

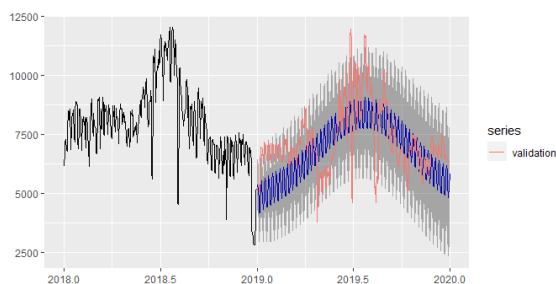


Figura 4: In Figura si mostra la serie temporale dei consumi energetici dell'edificio *U1* (linea nera e linea rossa). La linea rossa evidenzia i dati reali per l'anno 2019, la previsione del *TBATS* è rappresentata dalla linea blu.

Poiché il *MAPE* risultante è stato pari a 16.4, si è ritenuta accettabile l'abilità previsionale del modello *TBATS*.

Infatti, utilizzando anche il modello *FbProphet* di *Facebook* si è deciso di valutarne la correttezza rispetto allo stesso periodo temporale, come si può vedere dalla Figura 5.

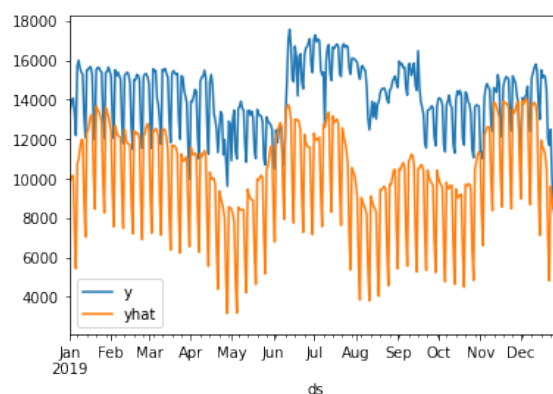


Figura 5: In Figura si mostra il confronto tra i dati reali per l'anno 2019 (linea arancione) e la previsione del modello *FbProphet* (linea blu).

In questo caso il *MAPE* risultante, ovvero la variazione percentuale tra i valori previsti e i valori reali, è stato pari a 29%, un risultato nettamente peggiore rispetto a quello del modello *TBATS*.

Come visibile dalla Figura 6, similmente al *MAPE* ottenuto precedentemente, si è ottenuto un *MAPE* pari a 26.3%, minore rispetto al valore ottenuto dal modello *FbProphet*.

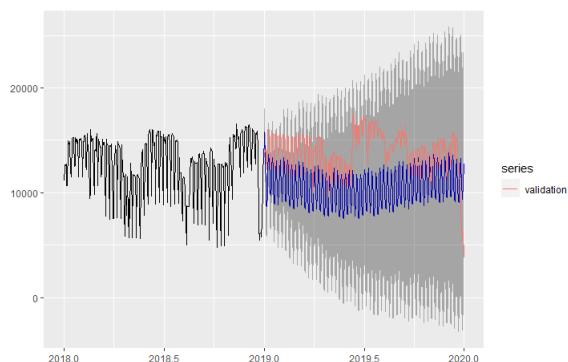


Figura 6: In Figura si mostra la serie temporale dei consumi energetici dell'edificio *U6* (linea nera e linea rossa). La linea rossa evidenzia i dati reali per l'anno 2019, la previsione del *TBATS* è rappresentata dalla linea blu.

### 5.5. Analisi impatto SARS-CoV-2

Per analizzare l'impatto che il *SARS-CoV-2* ha avuto sui consumi energetici degli edifici *U1* e *U6* si è usato il modello *TBATS*, sfruttando i dati *pre-Covid* per allenare il modello e facendo poi un'analisi predittiva sul periodo dal 23 *Febbraio* (giorno di chiusura dell' università causa emergenza sanitaria) al 10 settembre 2020, così da considerare il periodo del primo *lockdown*. Una volta ottenute le predizioni si sono confrontati i dati con i veri valori registrati nel periodo pandemico (nelle immagine sono rappresentati in rosso con il nome di *validation*).

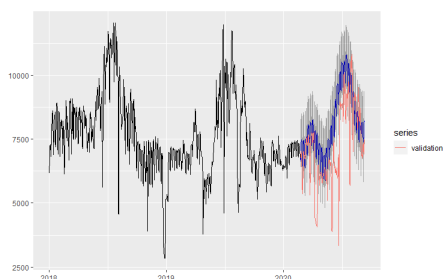


Figura 7: In Figura si mostra la serie temporale dei consumi per l'edificio *U1* (linea nera e linea rossa), in particolare si evidenziano i dati dei consumi relativi al periodo *Covid* (linea rossa) e si evidenzia la previsione del modello *TBATS* (linea blu)

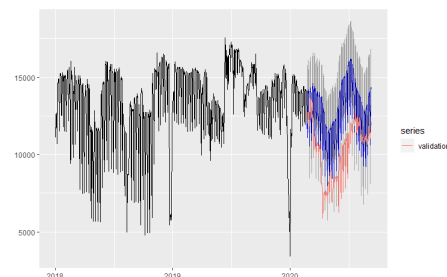


Figura 8: In Figura si mostra la serie temporale dei consumi per l'edificio *U6* (linea nera e linea rossa), in particolare si evidenziano i dati dei consumi relativi al periodo *Covid* (linea rossa) e si evidenzia la previsione del modello *TBATS* (linea blu)

Sia per l'edificio *U1* che per l'edificio *U6* si è osservato che la pandemia *Covid-19* ha avuto un impatto riduttivo sui consumi energetici, portando a minori consumi rispetto a quelli previsti dal modello *TBATS*.

### 5.6. Analisi Cross-Correlations

Di seguito sono rappresentati graficamente i risultati ottenuti dalle analisi delle *cross-correlation* svolte rispettivamente sui consumi e le temperature per singolo edificio e tra i consumi dei due edifici. Nella Figura 9 è importante notare come la correlazione oltre ad assumere valori statisticamente significativi, si attesti attorno a valori superiori a 0.4 indicando un livello di correlazione positivo moderato.

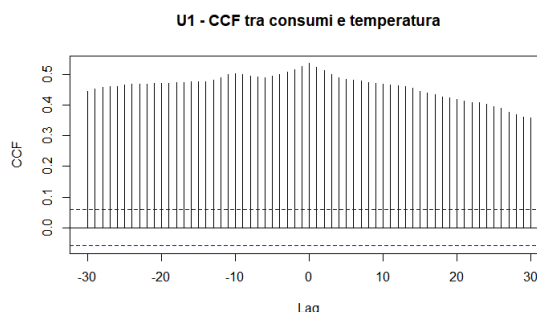


Figura 9: In Figura si mostrano le cross-correlations a diversi lag temporali tra consumi energetici e temperatura per l'edificio *U1*.

In Fig.10 si nota un fenomeno contrario a quello presente nella figura precedente. Infatti dallo studio della *cross-correlation* tra i consumi e la temperatura dell'edificio U6, si sono ottenuti dei valori statisticamente non significativi attorno allo 0, indicando un livello di correlazione trascurabile.

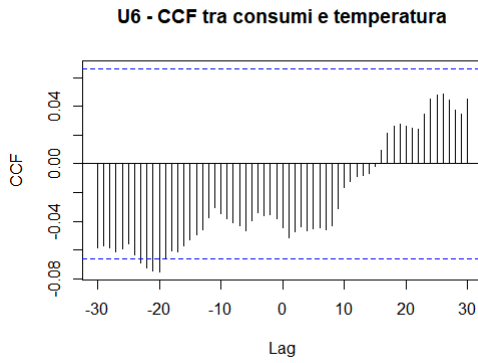


Figura 10: In Figura si mostrano le cross-correlations a diversi lag temporali tra consumi energetici e temperatura per l'edificio U6.

In Fig.11 si nota la presenza di una correlazione attorno al lag 0 che decresce all'aumentare e al diminuire del valore del lag, rimanendo sempre significativa. I consumi di U1 e U6 risultano essere correlati ma non vuol dire necessariamente che tra di essi sussista un legame diretto di *causa-effetto*, potendo essere tale *correlazione spuria*, ovvero dipendente da una terza variabile in comune, in assenza di meccanismo *logico-causale* plausibile che li metta in relazione tra loro.

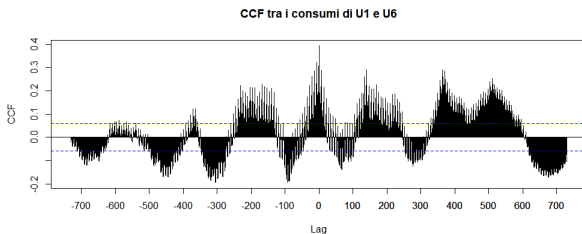


Figura 11: In Figura si mostrano le cross-correlations a diversi lag temporali tra consumi energetici degli edifici U1 e U6.

## 6. Risultati

In Sez.5 si sono presentate le metodologie adoperate per raggiungere diversi obiettivi: si sono confrontati due modelli statistici per fare previsioni sui consumi energetici per il breve-medio termine. Inoltre dopo aver indicato come metodologia più adeguata ai fini previsionali il modello *TBATS*, si è confrontata la previsione per il periodo di lockdown con i corrispondenti dati reali dimostrando come i consumi siano effettivamente diminuiti rispetto a quanto ci si sarebbe aspettati in assenza della pandemia. Infine, si sono calcolate le *cross-correlations* tra temperature e consumi e tra consumi per edifici diversi.

Per raggiungere il primo obiettivo si sono confrontati i modelli *FbProphet* e *TBATS*. La metrica usata per il confronto è stato il *Mean Absolute Percentage Error (MAPE)*, che misura la accuratezza delle previsioni secondo la seguente formula:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (19)$$

dove  $A_t$  rappresenta la serie dei valori reali mentre  $F_t$  rappresenta la serie fittizia delle previsioni. Questa metrica è stata testata sui dati prima della lockdown per non avere distorsioni dovute all'influenza della pandemia sui consumi come mostrato successivamente. In particolare si è usato come training set l'insieme i cui dati corrispondono all'anno 2018 e come test set l'insieme dei dati corrispondenti al 2019, ottenendo i seguenti risultati:

| Metodo           | MAPE U1 | MAPE U6 |
|------------------|---------|---------|
| <b>FbProphet</b> | 37.3    | 29.0    |
| <b>TBATS</b>     | 16.4    | 26.3    |

Tabella 2: Valori MAPE per i modelli *FbProphet* e *TBATS*.

Come mostrato in Tab. 2 il modello *TBATS* si adatta meglio ai dati sui consumi.

Per valutare l'effetto pandemico sui consumi energetici si sono calcolate le variazioni percentuali dei dati previsti con il modello *TBATS* dai dati reali ottenendo per i due edifici i seguenti risultati:

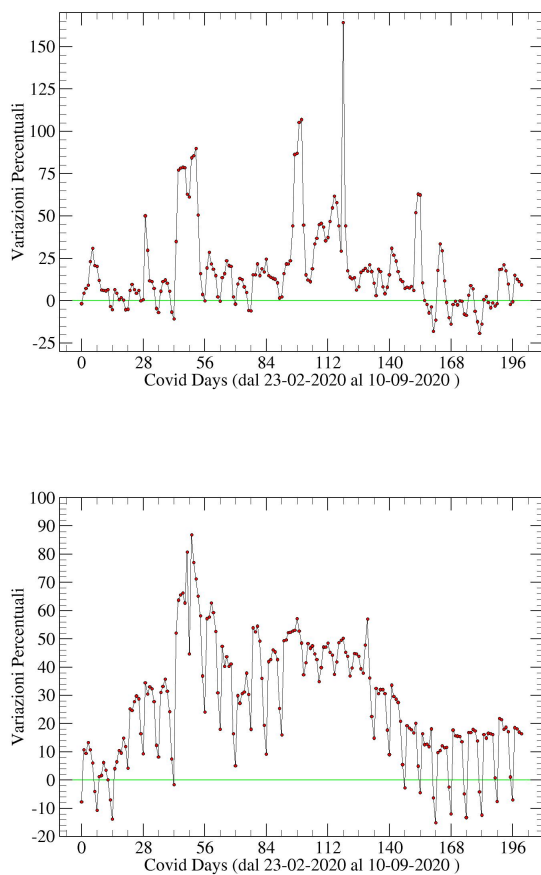


Figura 12: In Figura si mostrano le variazioni percentuali giornaliere tra le previsioni dei consumi energetici (modello *TBATS*) e i dati reali del periodo dell'edificio *U6* (panel superiore) e dell'edificio *U1* (panel inferiore). Si noti come nel periodo del lockdown esse siano quasi tutte positive e molto elevate con una media per gli edifici *U1* e *U6* al 25 luglio 2020 pari rispettivamente al 22,6% e al 32,7%. Invece a partire da tale data tornano a stabilizzarsi e a oscillare intorno allo zero.

Come si può notare dalla Fig.12 le variazioni percentuali indicano un elevato scostamento dei valori previsti, ottenuti con un modello efficace come il *TBATS*, dai valori reali durante il *lockdown*. Come facilmente intuibile il periodo di chiusura ha determinato un inevitabile risparmio energetico.

## 7. Conclusione e possibili sviluppi

In questo lavoro da una parte si è mostrato quantitativamente come il modello *TBATS* risulti essere un buon modello predittivo per i consumi energetici dopo averlo confrontato con il modello *FbProphet*. Infatti i due modelli, pur essendo tra le principali metodologie statistiche presenti in letteratura che permettono di trattare la *multistagionalità*, presentano due valori di *MAPE* molto differenti che dimostrano la migliore performance del *TBATS*.

Inoltre, dopo aver selezionato il *TBATS* come metodologia statistica adatta ai fini previsivi per i nostri dati, si è usato questo modello per valutare l'effetto della pandemia di *SARS-CoV-2*. Le variazioni percentuali infatti mostrano come i consumi siano drasticamente diminuiti durante il periodo pandemico come ci si aspettava.

Si è infine riusciti a misurare quantitativamente, attraverso le correlazioni incrociate, due fenomeni: il primo riguarda la correlazione su diversi lag temporali tra consumi energetici e temperatura, il secondo invece riguarda la variazione ciclica delle correlazioni incrociate su differenti lag temporali tra i consumi energetici dei due edifici.

I principali risultati di questa analisi mostrano come la temperatura sia correlata su diversi lag temporali con i consumi dell'edificio *U1*, e sia invece incorrelata con i consumi dell'edificio *U6*. Questo comportamento diffe-

rente delle correlazioni incrociate su diversi lag temporali tra consumi e temperature andrebbe ulteriormente approfondito in un successivo lavoro. La ricerca delle cause infatti può essere dovuta a molteplici fattori come ad esempio una differente efficienza energetica dei due edifici o alla dimensione degli stessi. Un'altra valutazione che va oltre gli scopi del presente lavoro e che sarà approfondita altrove riguarda la ciclicità delle correlazioni incrociate a differenti lag temporali tra i consumi dei due edifici. La valutazione di questo fenomeno, infatti, deve tener conto di una possibile correlazione spuria dovuta all'influenza di altre variabili che portano correlazioni e anticorrelazioni cicliche per i due edifici.

### **Riferimenti bibliografici**

- De Livera, A.M., Hyndman, R.J., Snyder, R.D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association* 106, 1513–1527.
- GreenMetric, U., 2020. Overall rankings 2020. URL: <https://greenmetric.ui.ac.id/rankings/overall-rankings-2020>.
- Taylor, S.J., Letham, B., 2018. Forecasting at scale. *The American Statistician* 72, 37–45.
- Venables, W., Ripley, B., 2002. *Modern applied statistics with s* fourth edition. world.