



Predittore della domenica

Moroni Fabio
Belcore Daniel
Ferrara Francesco
Zanaboni Riccardo

Scopo progetto



Predittore della domenica

Il file Excel contiene due anni di dati orari, strutturati come segue:

- 1° colonna: giorno dell'anno [1, ..., 365];
- 2° colonna: ora del giorno [1, ..., 24];
- 3° colonna: giorno della settimana [1, ..., 7];
- 4° colonna: serie temporale.

Obiettivo: Identificare un modello per il valore della serie temporale della domenica in funzione del giorno dell'anno e dell'ora.

Si richiede di scrivere una funzione Matlab:

$$d_{\text{hat}} = \text{prediz}(h, d)$$

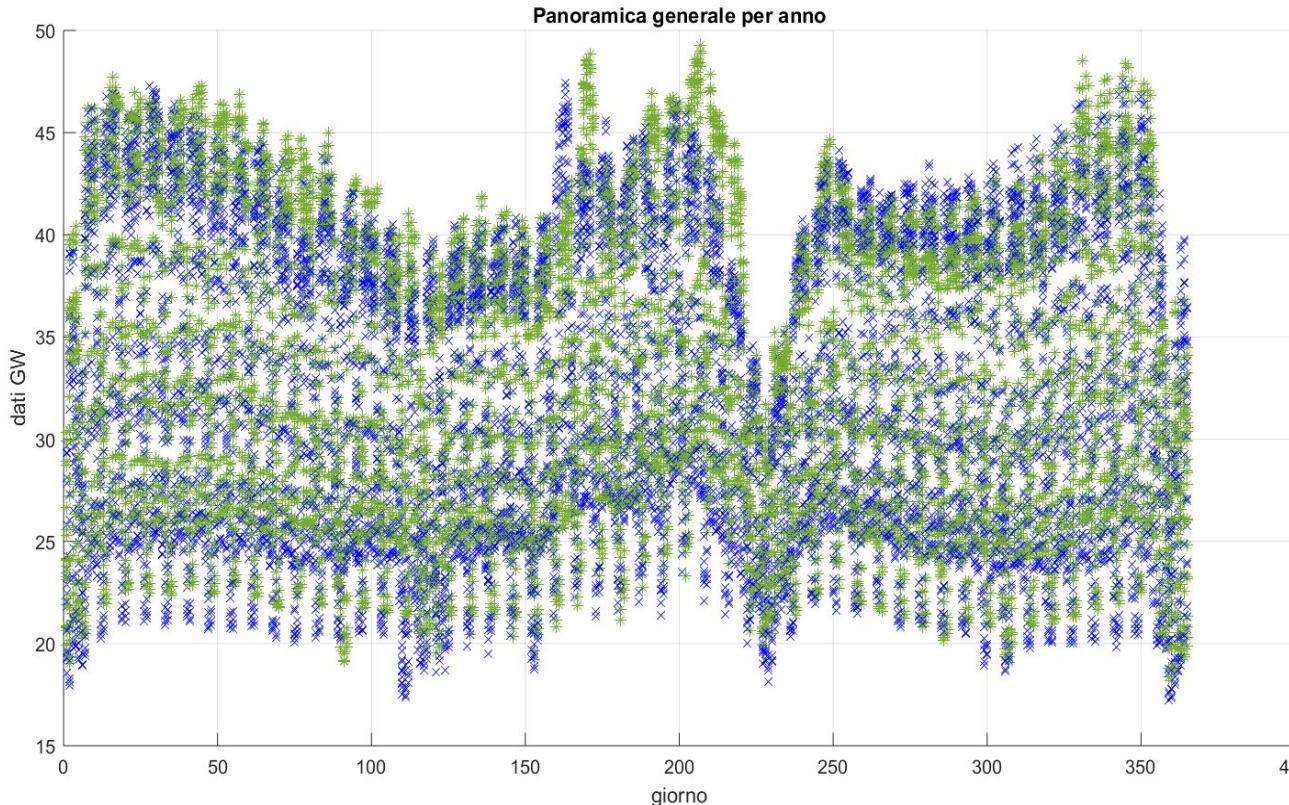
dove $h \in \{1, \dots, 24\}$ e $d \in \{1, \dots, 365\}$ sono scalari e d_{hat} è la previsione scalare (in un anno immediatamente successivo ai dati forniti) della serie nell'ora h del d -esimo giorno dell'anno. La funzione andrà consegnata al termine del progetto.



Scaletta

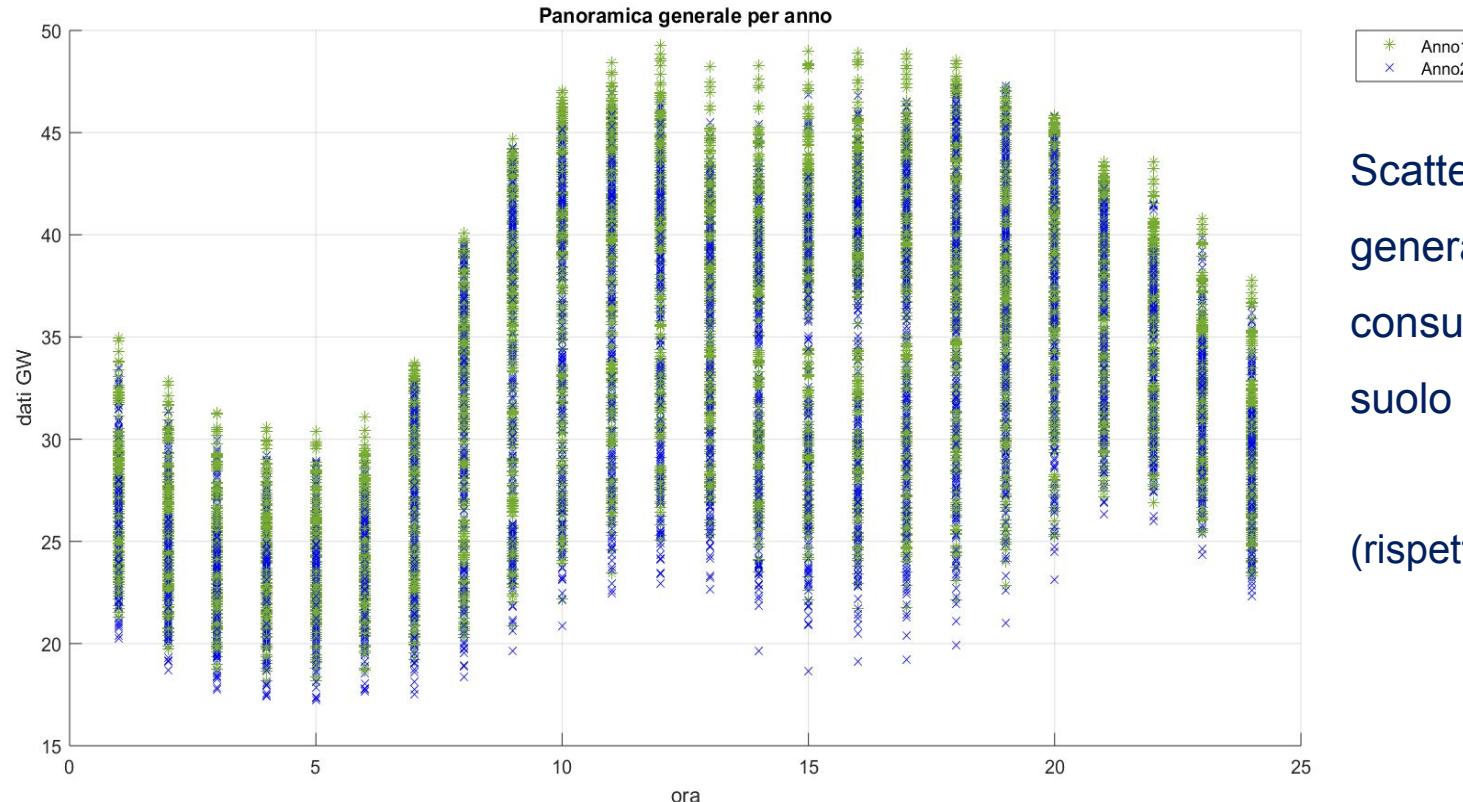
1. Prime osservazioni e considerazioni dei dati forniti
2. Primo approccio per identificazione del modello
3. Validazione, problemi riscontrati, nuovo approccio per identificazione
4. Conclusioni, mostrando la function MATLAB realizzata per il progetto

1. Dati a disposizione





1. Dati a disposizione



Scatter per panoramica generale dei due anni di consumo energetico sul suolo nazionale.

(rispetto ora del giorno)

1. Script MATLAB

```
1 - clear
2 - clc
3 - close all

4 -
5 - %% Lettura dati
6 - TabAnno1 = readtable('.\caricoITAhour.xlsx', 'Range', 'A2:D8762');
7 - TabAnno2 = readtable('.\caricoITAhour.xlsx', 'Range', 'A8763:D17522');

8 -
9 - TabAnno2.Properties.VariableNames = {'giorno_anno','ora_giorno', ...
10 -                                     'giorno_settimana','dati'};

11 -
12 - %% Selezione delle domeniche per anno e totali
13 -
14 - Anno1 = [TabAnno1.giorno_anno, TabAnno1.ora_giorno, TabAnno1.dat];
15 - Anno2 = [TabAnno2.giorno_anno, TabAnno2.ora_giorno, TabAnno2.dat];

16 - domeniche_anno1 = Anno1( (TabAnno1.giorno_settimana == 1), :, : );
17 - domeniche_anno2 = Anno2( (TabAnno2.giorno_settimana == 1), :, : );

18 - domeniche_totali = [domeniche_anno1; domeniche_anno2];

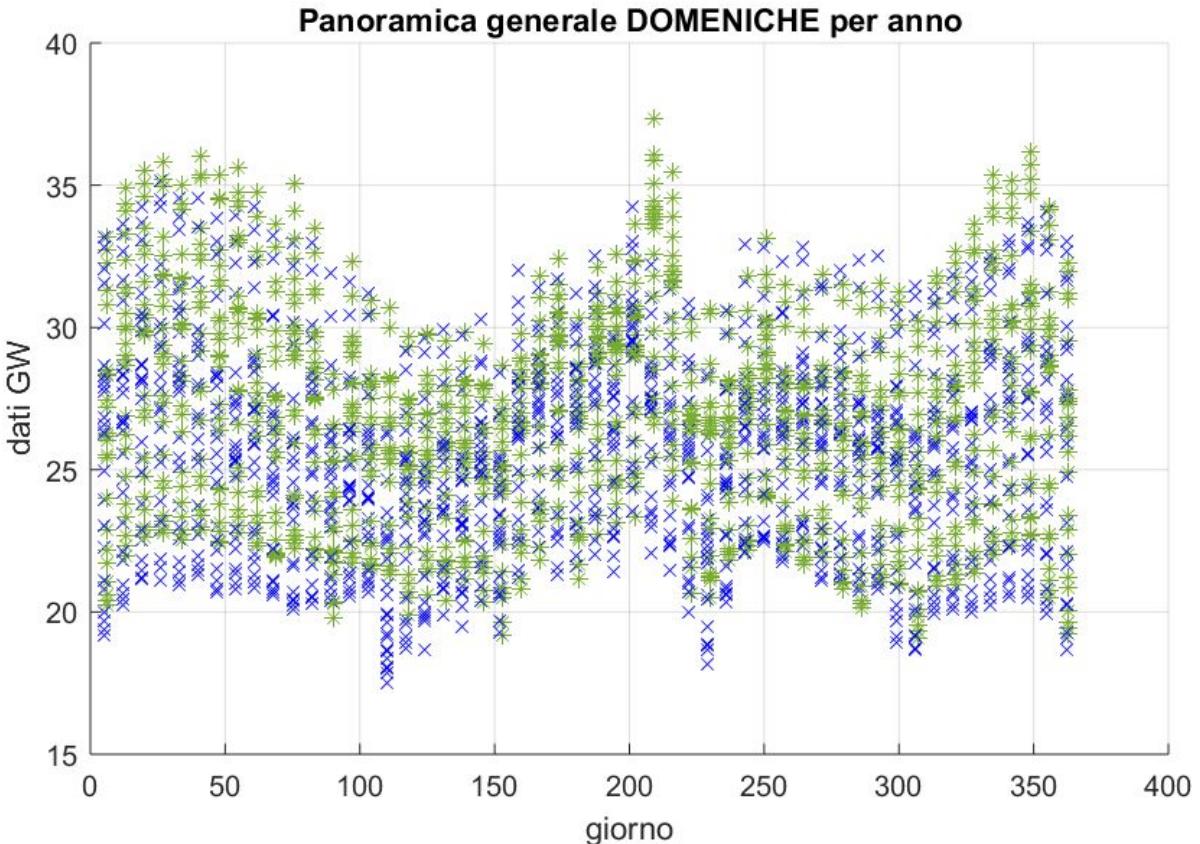
19 -
20 - domtot_giorno_anno = domeniche_totali(:,1) ./ 365 ;
21 - domtot_ora_giorno = domeniche_totali(:,2) ./ 24 ;
22 - domtot_dati = domeniche_totali(:,3);

23 -
24 - v1 = ones( length(domeniche_totali(:,3)) , 1 );

25 -
26 - v1 = ones( length(domeniche_totali(:,3)) , 1 );
```

Dovendo prevedere solo le domeniche, filtriamo i dati a nostra disposizione tenendo conto delle sole domeniche, indicate grazie al numero 1 nella colonna dedicata 'giorno della settimana'.

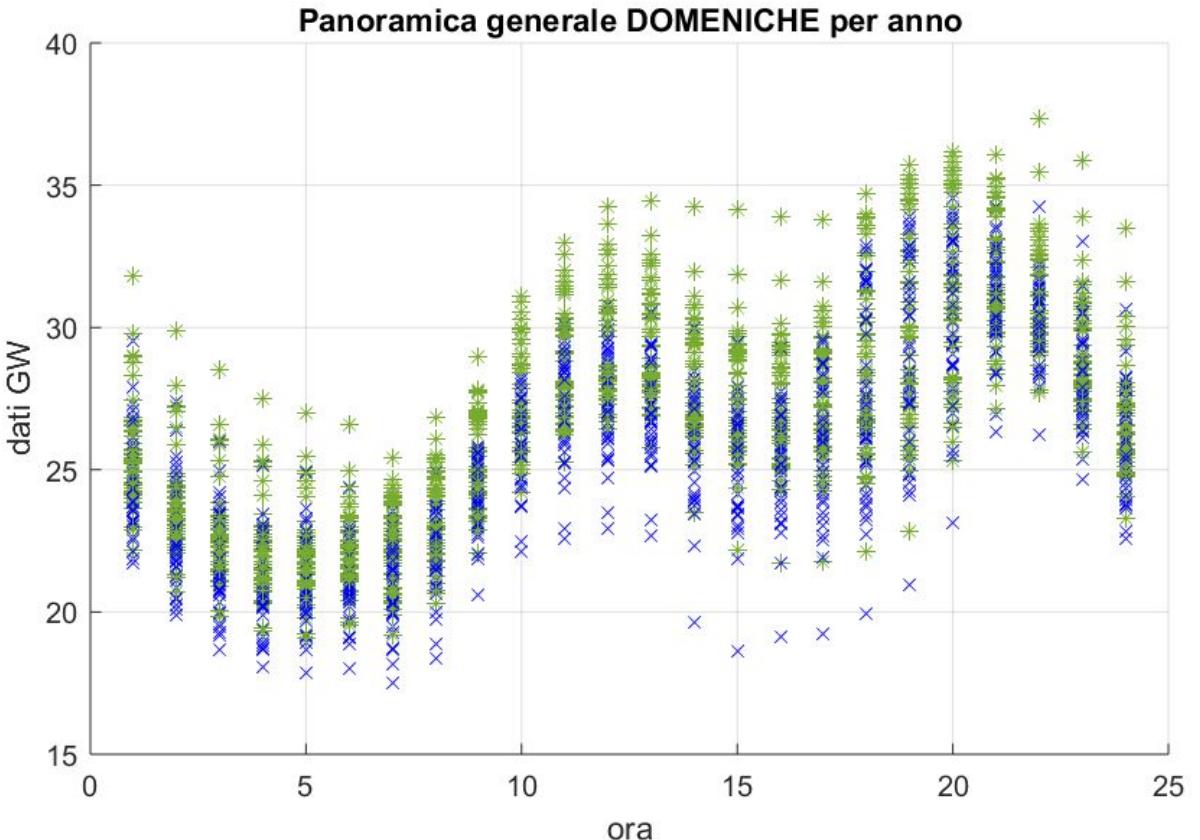
1. Dati a disposizione



Tendenzialmente nel secondo anno i valori sono più bassi, fattore non facile da notare negli scatter precedenti, considerando tutti i giorni della settimana.

(Rispetto giorno dell'anno)

1. Dati a disposizione



Anno1
Anno2

Tendenzialmente nel secondo anno i valori sono più bassi, fattore non facile da notare negli scatter precedenti, considerando tutti i giorni della settimana.

(Rispetto ora del giorno)



Sono presenti outliers?

```
%% Verifica outliers giornaliero per OGNI domenica
% NESSUN OUTLIER RILEVATO IN NESSUNA DOMENICA

j = 0;
for i = 1 : 52
    domenica = domeniche_totali( j+1 : j+24 , 1:3);
    domenica_outliers = isoutlier(domenica(:,3), 'quartiles');
    outliers_tot(i) = sum(domenica_outliers);
    j = j + 24;
end

Somma_totale_outliers = sum(outliers_tot)      % = 0
```

Considerando tutti i dati delle domeniche di tutto l'anno: poiché è risultato non esserci alcun outlier, abbiamo approfondito la ricerca provando ad analizzare se fossero presenti outlier domenica per domenica.

Tuttavia, anche in questo caso non sono stati individuati outliers.

2. Primi modelli

Non essendoci particolari richieste nella consegna, come primo tentativo abbiamo optato per un modello di tipo matrioska polinomiale.

```
phi_mod3 = [ v1, domtot_giorno_anno, domtot_ora_giorno, ...
    (domtot_giorno_anno.^2), (domtot_ora_giorno.^2), ...
    (domtot_giorno_anno .* domtot_ora_giorno), ...
    (domtot_giorno_anno.^3), (domtot_ora_giorno.^3), ...
    ( (domtot_giorno_anno.^2) .* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^2) .* domtot_giorno_anno ) ] ;
```

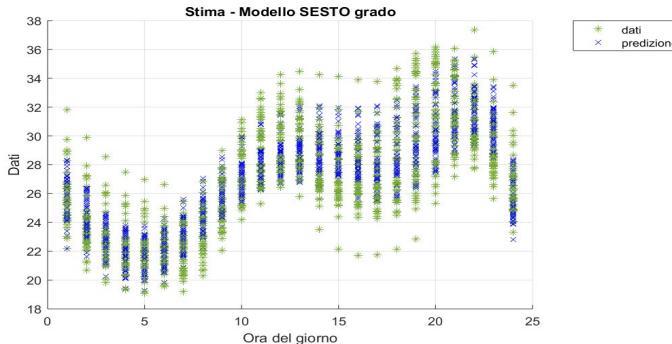
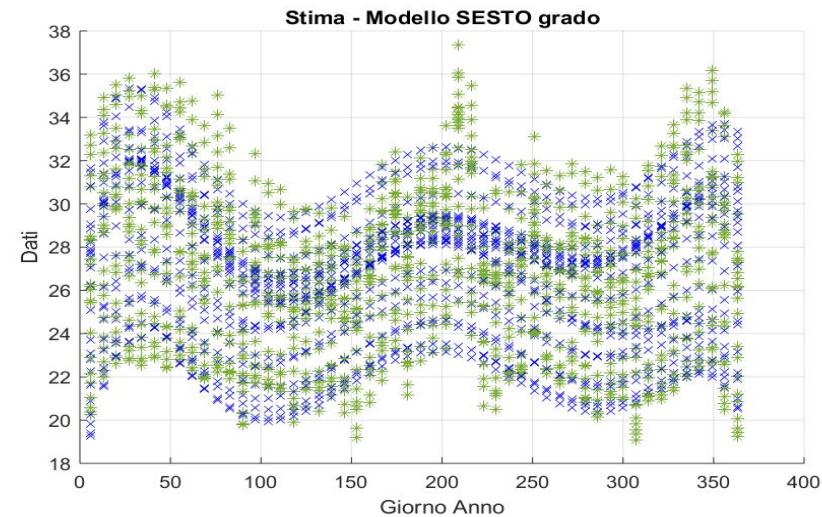
```
LS_mod3 = lscov (phi_mod3 , domtot_dati);
mod3 = phi_mod3 * LS_mod3 ;
```

```
phi_mod4 = [ v1, domtot_giorno_anno, domtot_ora_giorno, ...
    (domtot_giorno_anno.^2), (domtot_ora_giorno.^2), ...
    (domtot_giorno_anno .* domtot_ora_giorno), ...
    (domtot_giorno_anno.^3), (domtot_ora_giorno.^3), ...
    ( (domtot_giorno_anno.^2) .* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^2) .* domtot_giorno_anno ), ...
    (domtot_giorno_anno.^4), (domtot_ora_giorno.^4), ...
    ( (domtot_giorno_anno.^2).* (domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^3) .* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^3) .* domtot_giorno_anno ) ] ;
```

```
phi_mod5 = [ v1, domtot_giorno_anno, domtot_ora_giorno, ...
    (domtot_giorno_anno.^2), (domtot_ora_giorno.^2), ...
    (domtot_giorno_anno .* domtot_ora_giorno), ...
    (domtot_giorno_anno.^3), (domtot_ora_giorno.^3), ...
    ( (domtot_giorno_anno.^2) .* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^2) .* domtot_giorno_anno ), ...
    (domtot_giorno_anno.^4), (domtot_ora_giorno.^4), ...
    ( (domtot_giorno_anno.^2).*(domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^3) .* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^3) .* domtot_giorno_anno ), ...
    ( (domtot_giorno_anno.^4) .* (domtot_ora_giorno.^4) ), ...
    ( (domtot_giorno_anno.^3) .* (domtot_giorno_anno.^2) ), ...
    ( (domtot_ora_giorno.^4) .* domtot_giorno_anno) ] ;
```

e andando avanti ...

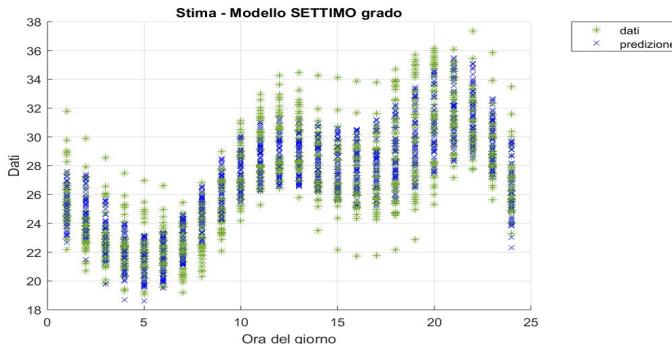
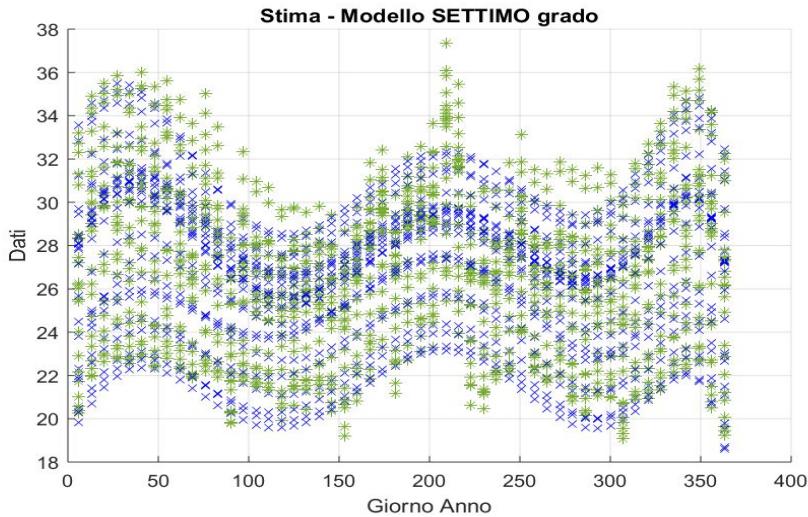
2. Primi modelli



* dati
x predizione

```
phi_mod6 = [ v1, domtot_giorno_anno, domtot_ora_giorno, ...
    (domtot_giorno_anno.^2), (domtot_ora_giorno.^2), ...
    (domtot_giorno_anno .* domtot_ora_giorno), ...
    (domtot_giorno_anno.^3), (domtot_ora_giorno.^3), ...
    ( (domtot_giorno_anno.^2) .* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^2) .* domtot_giorno_anno ), ...
    (domtot_giorno_anno.^4), (domtot_ora_giorno.^4), ...
    ( (domtot_giorno_anno.^2).* (domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^3).* (domtot_ora_giorno ) ), ...
    ( (domtot_ora_giorno.^3) .* domtot_giorno_anno ), ...
    (domtot_giorno_anno.^5), (domtot_ora_giorno.^5), ...
    ( (domtot_giorno_anno.^3).* (domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^4).* (domtot_ora_giorno ) ), ...
    ( (domtot_ora_giorno.^3) .* (domtot_giorno_anno.^2) ), ...
    ( (domtot_giorno_anno.^4) .* domtot_giorno_anno ), ...
    ( (domtot_ora_giorno.^4) .* (domtot_giorno_anno.^2) ), ...
    ( (domtot_giorno_anno.^5) .* domtot_giorno_anno ) ] ;
```

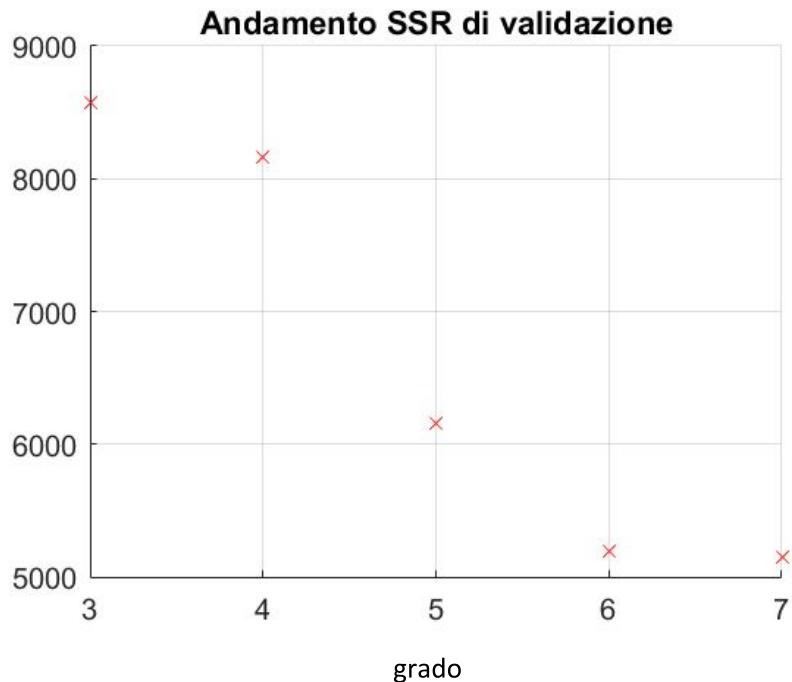
2. Primi modelli



* dati
x predizione

```
phi_mod7 = [ v1, domtot_giorno_anno, domtot_ora_giorno, ...
    (domtot_giorno_anno.^2), (domtot_ora_giorno.^2), ...
    (domtot_giorno_anno .* domtot_ora_giorno), ...
    (domtot_giorno_anno.^3), (domtot_ora_giorno.^3), ...
    ( (domtot_giorno_anno.^2) .* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^2) .* domtot_giorno_anno ), ...
    (domtot_giorno_anno.^4), (domtot_ora_giorno.^4), ...
    ( (domtot_giorno_anno.^2).* (domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^3).* domtot_ora_giorno ), ...
    ( (domtot_ora_giorno.^3).* domtot_giorno_anno ), ...
    (domtot_giorno_anno.^5), (domtot_ora_giorno.^5), ...
    ( (domtot_giorno_anno.^3).* (domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^4).* (domtot_ora_giorno) ), ...
    ( (domtot_ora_giorno.^3).* (domtot_giorno_anno.^2) ), ...
    ( (domtot_ora_giorno.^4).* domtot_giorno_anno ), ...
    (domtot_giorno_anno.^6), (domtot_ora_giorno.^6), ...
    ( (domtot_giorno_anno.^4).* (domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^5).* (domtot_ora_giorno) ), ...
    ( (domtot_giorno_anno.^3).* (domtot_ora_giorno.^3) ), ...
    ( (domtot_ora_giorno.^4).* (domtot_giorno_anno.^2) ), ...
    ( (domtot_ora_giorno.^5).* domtot_giorno_anno), ...
    (domtot_giorno_anno.^7), (domtot_ora_giorno.^7), ...
    ( (domtot_giorno_anno.^6).* (domtot_ora_giorno.^1) ), ...
    ( (domtot_giorno_anno.^5).* (domtot_ora_giorno.^2) ), ...
    ( (domtot_giorno_anno.^4).* (domtot_ora_giorno.^3) ), ...
    ( (domtot_ora_giorno.^4).* (domtot_giorno_anno.^3) ), ...
    ( (domtot_ora_giorno.^5).* (domtot_giorno_anno.^2) ), ...
    ( (domtot_ora_giorno.^6).* (domtot_giorno_anno.^1) ) ] ;
```

2. Cross-Validazione (polinomiale)



I dati presentano variazioni molto rapide che dei semplici polinomi faticano a seguire.

Dovremo aumentare sempre di più il grado massimo dei monomi impiegati, come si può notare osservando l'andamento degli SSR di validazione.



3. Modello serie di Fourier

```
%% Modello Serie di Fourier
for N = 3 : 25
    phi_mod = [ c, v1 , cos( 2 * pi/24 * domtot_ora_giorno(:)) , ...
                cos(2 * pi/365 * domtot_giorno_anno(:)), ...
                sin(2 * pi/24 * domtot_ora_giorno(:)) , ...
                sin(2 * pi/365 * domtot_giorno_anno(:))];
    for k = 2 : N
        phi_mod = [ phi_mod , cos( k * 2 * pi/24 * domtot_ora_giorno(:)) , .
                    cos( k * 2 * pi/365 * domtot_giorno_anno(:)), ...
                    sin( k * 2 * pi/24 * domtot_ora_giorno(:)) , ...
                    sin( k * 2 * pi/365 * domtot_giorno_anno(:)) ];
    end
    %Stima LS, calcolo della predizione e SSR di validazione
    LS_mod = lscov (phi_mod , domtot_dati);
    mod = phi_mod * LS_mod ;
```

Inizialmente abbiamo usato lo stesso numero di armoniche per entrambe le variabili, scrivendo tutto in un ciclo unico.

v1 = vettore di 1;

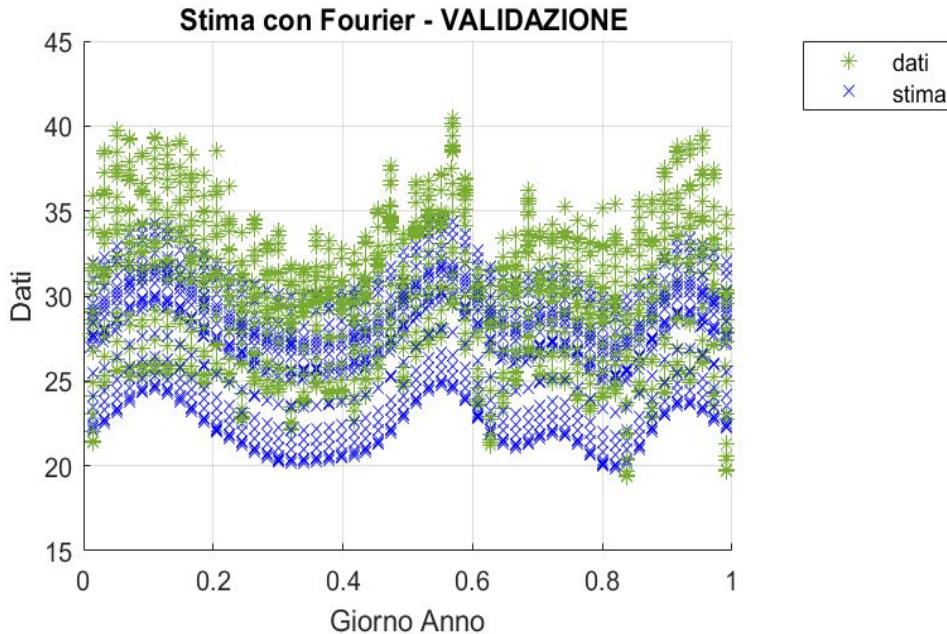
c = vettore con sola variabile giorno anno per provare a prevedere anche il trend annuale.



Problemi riscontrati

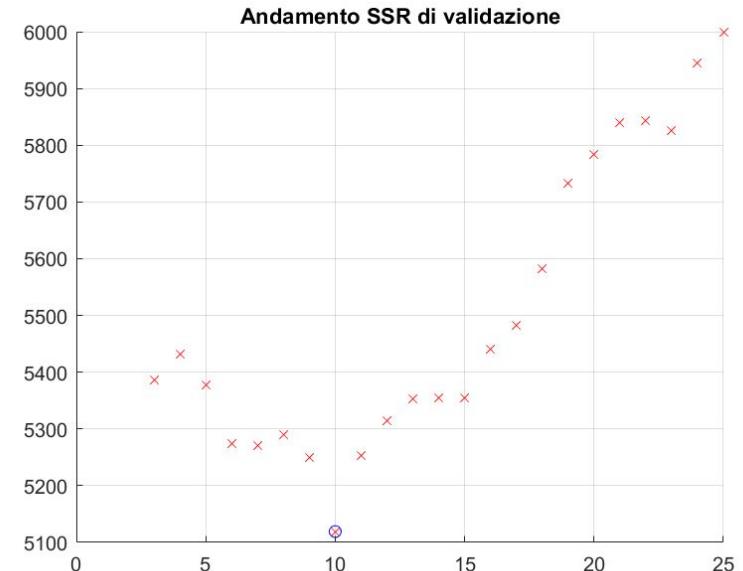
PROBLEMA TREND (1) :

Come notiamo dallo scatter, il modello non stima correttamente il trend annuale, ottenendo così un grafico "traslato".



PROBLEMA ARMONICHE (2) :

Vengono utilizzate un numero eccessivo di armoniche per ognuna delle variabili.





Soluzioni adottate

SOLUZIONE TREND (1) :

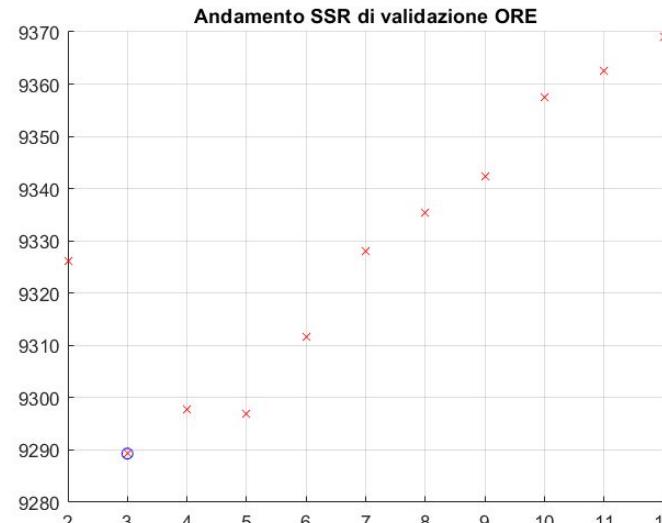
Viene stimato il trend annuale e sottratta la stima così ottenuta ai dati di partenza, su cui si calcola il modello di Fourier

```
%% Stima del TREND e sistemazione dati
```

```
phi_trend = [v1, domtot_giorno_anno];  
LS_trend = lscov(phi_trend, domtot_dati);  
trend = phi_trend * LS_trend;  
domtot_dati = domtot_dati - trend; ←
```

SOLUZIONE ARMONICHE (2) :

Si è separata la validazione dei modelli per ogni variabile e valutare con la cross-validazione il numero migliore di armoniche. Ad esempio, sono necessarie meno armoniche per le ore del giorno.





Risultati delle soluzioni

```
% Modello Serie di Fourier con accortezze finali

phi_mod = [];
NH = 3;

for NG = 3 : 25
    phi_hh = [ cos( 2 * pi/24 * domtot_ora_giorno(:)) , ...
                sin(2 * pi/24 * domtot_ora_giorno(:)) ];

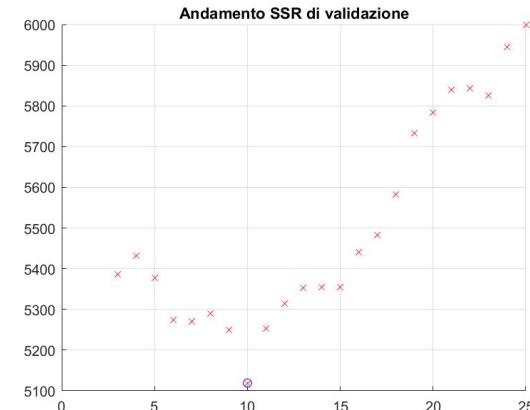
    phi_gg = [ cos(2 * pi/365 * domtot_giorno_anno(:)) , ...
                sin(2 * pi/365 * domtot_giorno_anno(:)) ];

    for k = 2 : NH
        phi_hh = [ phi_hh , cos( k * 2 * pi/24 * domtot_ora_giorno(:)) , ...
                    sin( k * 2 * pi/24 * domtot_ora_giorno(:)) ];
    end

    for k = 2 : NG
        phi_gg = [ phi_gg , cos( k * 2 * pi/365 * domtot_giorno_anno(:)), ...
                    sin( k * 2 * pi/365 * domtot_giorno_anno(:)) ];
    end

domtot_dati = domeniche_annol(:,3) - trend;
LS_mod_hh = lscov (phi_hh , domtot_dati);
mod_hh = phi_hh * LS_mod_hh;
domtot_dati = domtot_dati - mod_hh;
```

Nonostante la sottrazione del modello ottenuto con le sole ore del giorno (3 armoniche), il numero di armoniche da usare per i giorni dell'anno non è variato, rimanendo così a 10.

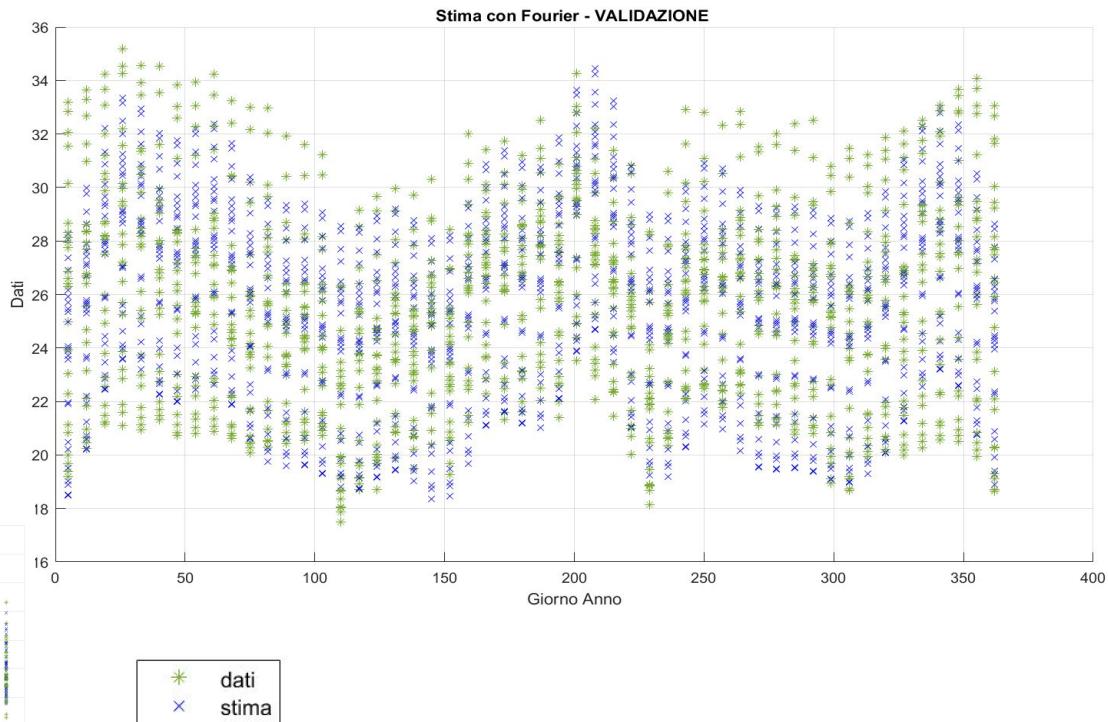
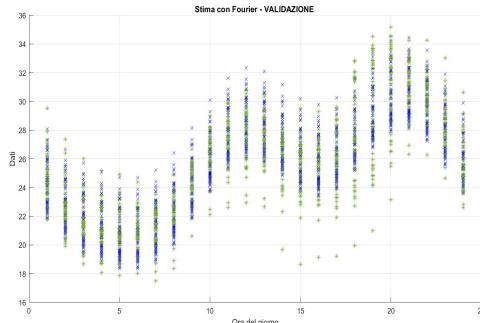




- Sottraendo volta per volta dei modelli ‘parziali’, abbiamo ridotto il numero di armoniche necessarie per ogni variabile, semplificando e riducendo le dimensioni della matrice PHI.

Da $10 \times 4 = 40$ colonne a $(3 \times 2 + 10 \times 2) = 26$ colonne

- Anche in validazione è notevolmente migliorata la predizione.



4. Function richiesta

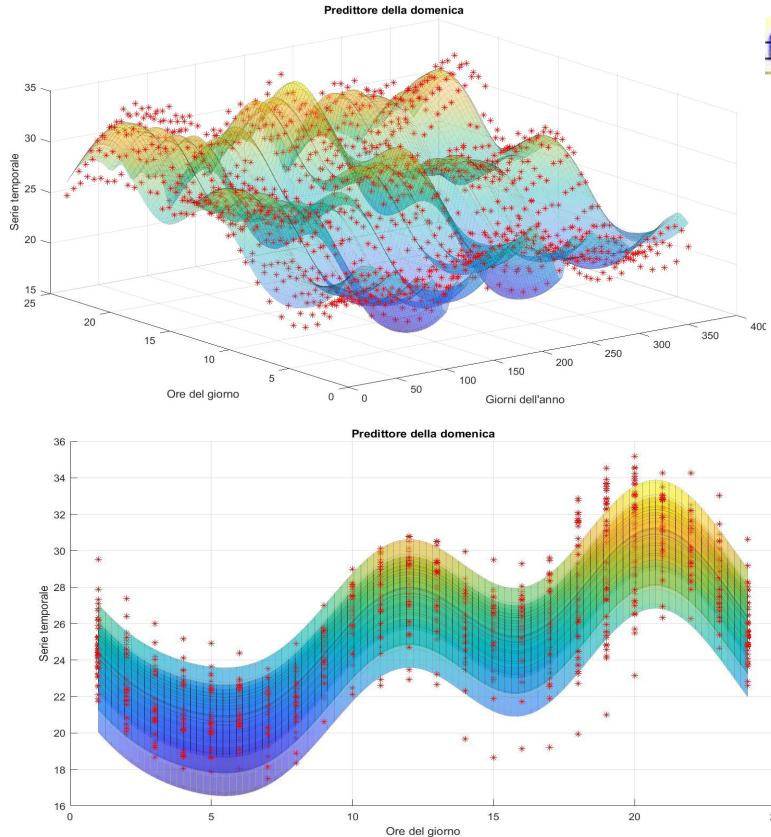
Si richiede di scrivere una funzione Matlab:

$$d_{\text{hat}} = \text{prediz}(h, d)$$

dove $h \in \{1, \dots, 24\}$ e $d \in \{1, \dots, 365\}$ sono scalari e d_{hat} è la previsione scalare (in un anno immediatamente successivo ai dati forniti) della serie nell'ora h del d -esimo giorno dell'anno.

```
function predizione = prediz(ora_giorno, giorno_anno)
% Predizione : per un corretto calcolo, vengono sommati i valori
%                 dei due modelli per le singole variabili e del
%                 trend annuale inizialmente calcolato.
v2 = domeniche_anno2(:,3);
v1 = domeniche_anno1(:,3);
t = mean(v2) - mean(v1);

% CALCOLO FINALE
predizione = mod_hh_pred + mod_gg_pred + trend_pred + t ;
```



4. Grafico finale con mesh

```
function myMesh(giorno_anno, ora_giorno, dati)
```

