

CONSIDERAZIONI SU TREND DATI, GRAFICI, MODELLI CLIMATICI E ALIASING

Quando vediamo grafici che mostrano dati catalogati su un lungo periodo (lunghezza relativa alla frequenza di campionamento dei dati stessi) siamo portati a credere che ciò che vediamo a grafico sia esattamente ciò che viene misurato, che tutti i dati campionati o registrati siano espressi nel grafico. Non è quasi mai così. La produzione di un grafico finale passa attraverso diverse fasi per cui, direttamente a grafico o precedentemente in tabella dati, vengono effettuate delle scremature dei dati registrati, scremature basate sulla teorica rappresentatività del dato. Questa rappresentatività può essere gestita in diversi modi, ma il metodo più usato è quello di verificare quando e quanto il dato si conforma alla media ed ai dati circostanti, precedenti e successivi. Esistono su questo metodo due ottiche diverse: quello della conformazione, e quello della eccezione.

L'ottica della conformazione tende a reputare non valido un dato che non si conforma a quelli circostanti e/o 'tipico', ad esempio se in una giornata ho 10 misurazioni di temperatura tra i 10°C ed i 12°C, ed improvvisamente in mezzo ho un picco di 15°C, l'ottica di trattamento basata sulla conformazione tenderà a sopprimere o smussare (smoothing) i 15°C per non deviare dalla media ottenuta con le 10 misurazioni coerenti. L'ottica della eccezione, al contrario, gestisce i dati dissonanti come eccezione particolare a cui dare più rilevanza; nel caso ipotizzato poc'anzi, quindi, il dato dei 15°C verrà non conformato ma esaltato nel trend dei dati. L'ottica della eccezione è quindi più conservativa e più rappresentativa della realtà.

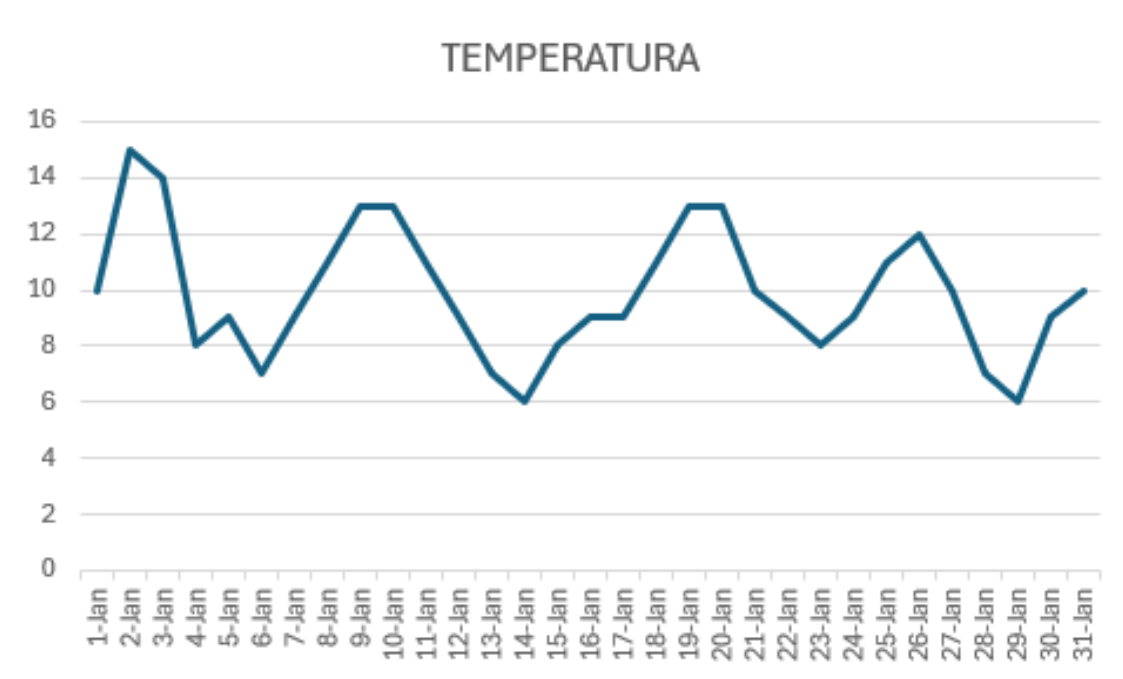
Parliamo ora di come questo si applichi ai grafici di lungo periodo.

Un sistema computerizzato gestisce dati raccolti a frequenza variabile (teoricamente il più regolare possibile) salvandoli in tabelle, che non sono altro che files; come tali, essi occupano spazio, e quando richiamati in memoria per la rappresentazione grafica, occupano memoria. I trend grafici, che rappresentano lo storico dati, dovranno mostrare in una finestra di dimensioni prestabilite tutta la collezione di dati, che con l'aumento dei campionamenti assumerà un aspetto sempre più schiacciato.

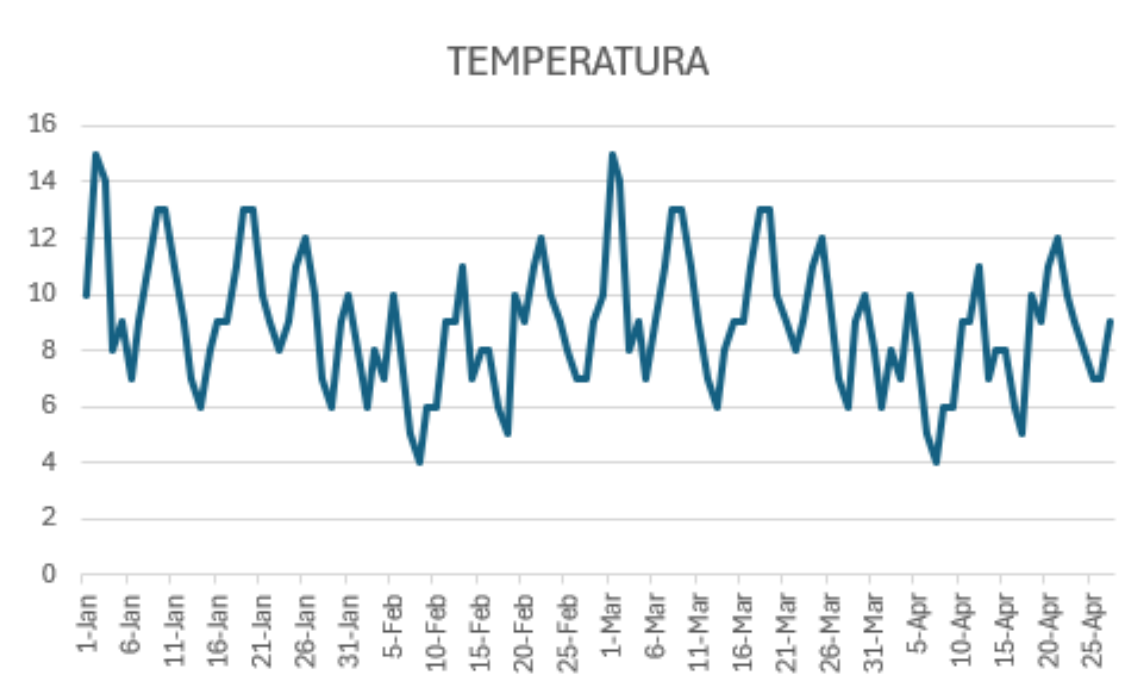
Si arriverà a situazioni in cui lo schiacciamento dei dati nel grafico è tale da rendere non identificabili correttamente i singoli punti di campionamento, producendo un grafico che assume sempre più l'aspetto di rumore.

Si considerino i seguenti grafici:

A)



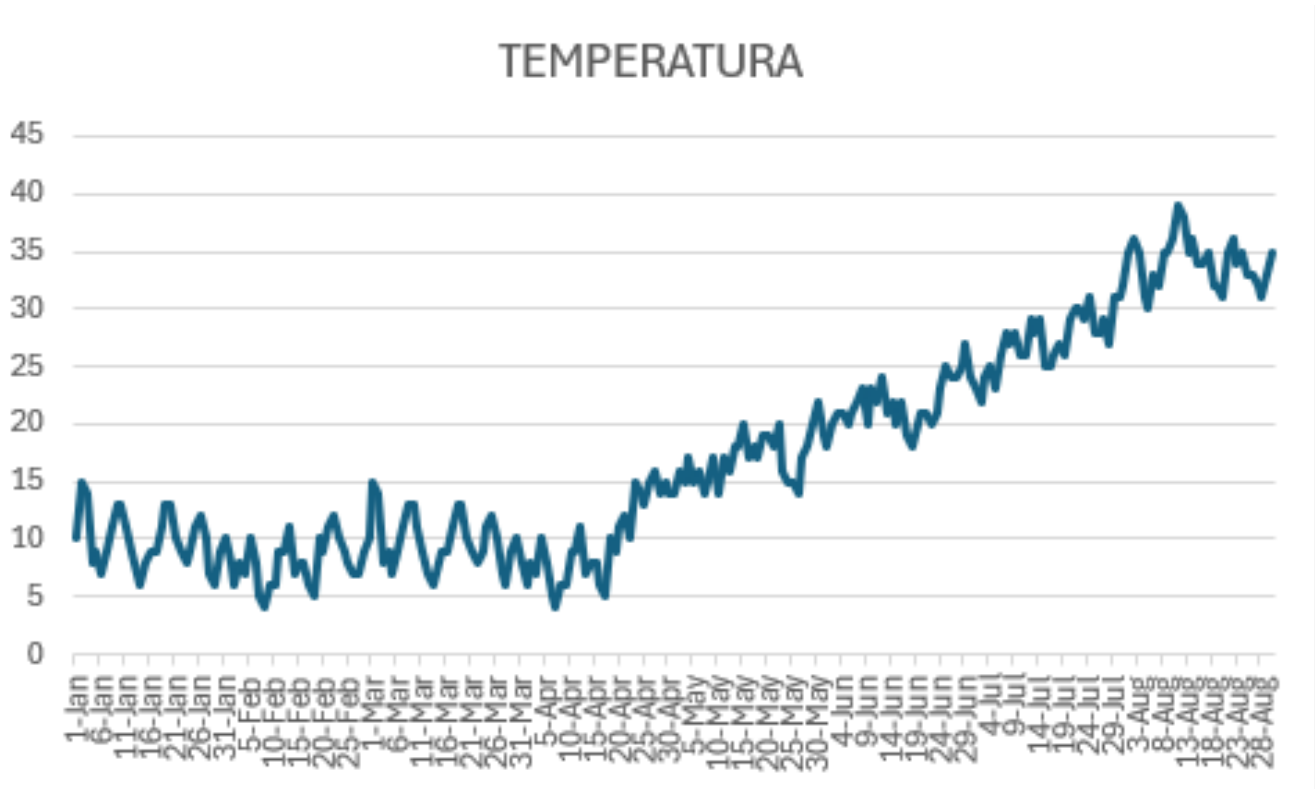
B)



Il grafico A) illustra un mese di misurazioni meteo, mentre il grafico B) ne mostra quattro mesi (in realtà due mesi, duplicati); è evidente lo schiacciamento di dati dovuti all'eccessivo numero di misurazioni da rappresentare nella stessa dimensione orizzontale.

Si consideri ora il grafico C):

C)



Il grafico C) rappresenta 8 mesi di misurazioni di dati, ed è evidente come, pur se il trend mostra chiaramente l'andamento al riscaldamento, i dati siano talmente compressi da creare una visuale generalizzata ed uniforme nascondendo la variabilità dei singoli giorni.

Ora immaginiamo cosa può succedere in un grafico che deve raccogliere e mostrare dati giornalieri di un periodo di 5 anni, di 10 anni, di 30 anni. Il grafico risultante sarà forzatamente caotico e simile a puro rumore, tanto da costringere a ridurre o selezionare la frequenza di rappresentazione, mostrando non più ogni singolo giorno, ma, ad esempio, un valore per settimana.

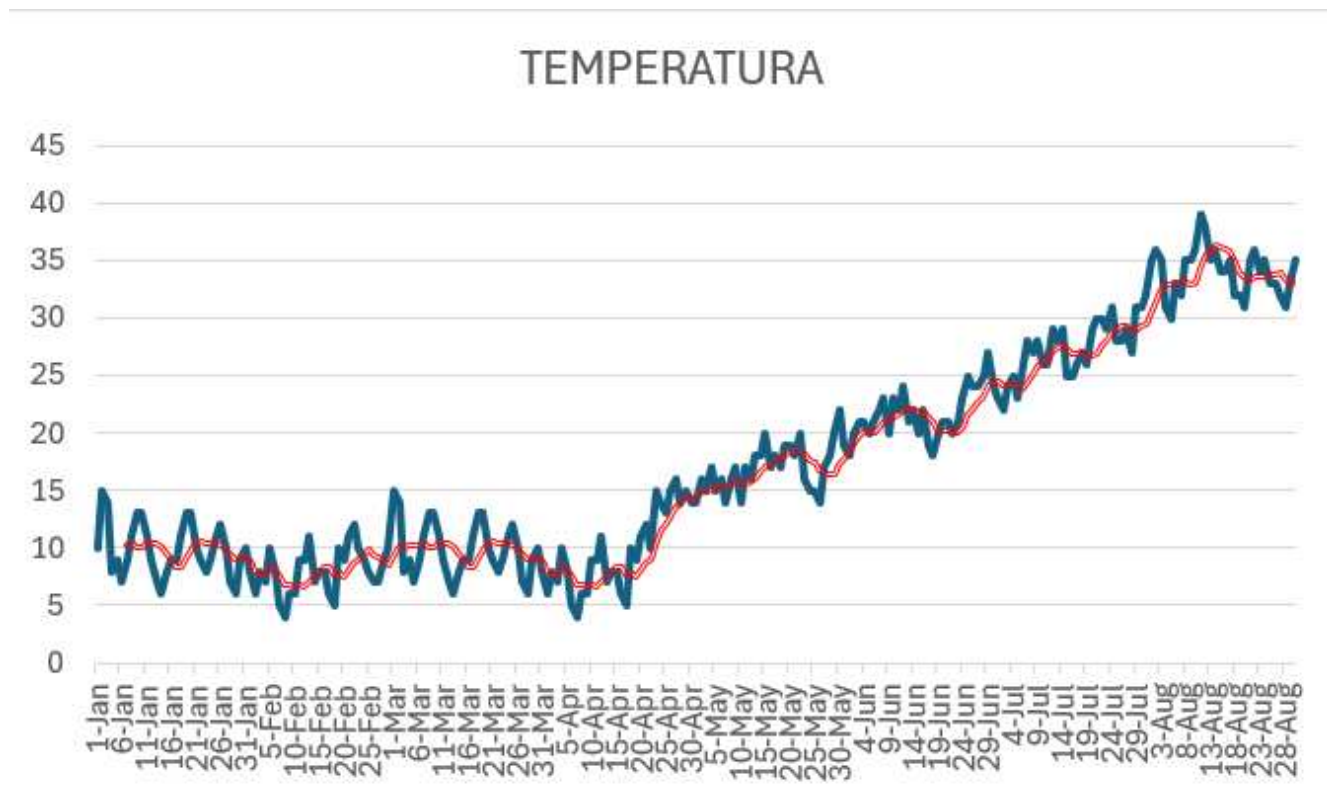
Come verrà scelto quel valore? Facendo una media di tutti i valori di ogni settimana.

E come verrà creata questa media? Ecco qui che entrano in gioco i due concetti di conformità ed eccezione.

Quasi tutti i modelli climatici esistenti sono costituiti seguendo un'ottica alla conformità, poco conservativa, e che tende ad appiattire i dati quando questi differiscono dalla serie circostante.

Si consideri il grafico C-bis):

C-bis)



Questo grafico mostra la 'Moving Average' (media mobile) a sette giorni, rappresentata da una doppia linea rossa. E' evidente come in questa rappresentazione media la variabilità ed i picchi che differiscono dai dati circostanti siano sacrificati a vantaggio della uniformità di rappresentazione.

Questo è il tipico esempio di grafico prodotto con l'ottica della conformazione, che è lo standard in tutti i sistemi di rappresentazione grafica di dati, siano essi semplici come Microsoft Excel o complessi come i modelli climatici. I dati così catalogati, trovano dunque una rappresentazione grafica che non li rappresenta affatto, e che piega il loro andamento al concetto di uniformità.

Ma c'è un fattore ancora più importante da considerare: lo smussamento dei dati e la loro non rappresentatività possono essere causati anche da problematiche tecniche più complesse e legate ai metodi di campionamento ed agli algoritmi di gestione. Si parla in questi casi di *Effetto Schiacciamento*.

L'effetto schiacciamento dei dati nei trend di misurazione strumentale è un fenomeno che può verificarsi quando il periodo di campionamento si allunga. Questo può essere legato a vari fattori, tra cui *l'aliasing* e *l'errore di quantizzazione*.

L'aliasing è un fenomeno che si verifica quando la frequenza di campionamento è inferiore al doppio della massima frequenza del segnale (*frequenza di Nyquist*). In questo caso, le frequenze più alte possono apparire come frequenze più basse, creando distorsioni nel segnale campionato.

L'errore di quantizzazione si verifica quando un segnale continuo viene convertito in un segnale discreto. Durante questo processo, i valori continui vengono approssimati a valori discreti, introducendo un errore.

Inoltre, l'effetto di schiacciamento dei dati può essere influenzato dalla sincronizzazione del campionamento con i massimi e i minimi del segnale. Se i campioni non sono esattamente sincronizzati con le variazioni del segnale, la ricostruzione del segnale può non essere fedele.

A noi, ai fini di questo lunghissimo discorso, interessa in modo particolare il problema dell'aliasing applicato ai modelli climatici. Può essere evitato? E se sì, come? Con che confidenza?

Nei modelli climatici, l'aliasing viene teoricamente mitigato attraverso diverse tecniche. Una di queste è l'uso di una frequenza di campionamento adeguata. Secondo il teorema del campionamento di *Nyquist-Shannon*, la frequenza di campionamento dovrebbe essere almeno il doppio della frequenza massima presente nel segnale per evitare l'aliasing. Un'altra strategia è l'uso di filtri *passa-basso* per rimuovere le frequenze superiori alla frequenza di Nyquist prima del campionamento. Questo può aiutare a prevenire l'aliasing, ma può anche ridurre la risoluzione dei dati.

Inoltre, i modelli climatici utilizzano tecniche di interpolazione e smoothing per ridurre gli effetti dell'aliasing. Queste tecniche possono aiutare a riempire i dati mancanti tra i campioni e a ridurre il rumore nei dati.

Ciò che deve essere chiaro, è che queste tecniche di anti-aliasing possono solo mitigare, non eliminare completamente, l'aliasing, ed al costo di falsare i dati ed il loro risultato grafico in altre maniere, ottenendo quindi in maniera diversa proprio il risultato che vogliamo evitare: l'appiattimento dei dati.

In modo particolare l'utilizzo di filtri passa basso può rivelarsi addirittura avere un effetto opposto a quello desiderato. Perché? Un filtro passa-basso applicato a una serie di misurazioni di temperatura con molti picchi può avere vari effetti. Innanzitutto, un filtro passa-basso permette il passaggio di frequenze al di sotto di una data soglia, detta frequenza di taglio, bloccando le alte frequenze. Pertanto, i picchi di temperatura che si verificano a frequenze superiori alla frequenza di taglio del filtro potrebbero essere attenuati o eliminati. Questo può avere il doppio e contrastante effetto di preservare la misurazione di temperatura (cioè evitare di eliminarla dalla media calcolata) ma di smussarne i picchi, rendendo la serie di misurazioni più liscia o uniforme. Questo è ritenuto utile (secondo noi erroneamente) quando si desidera eliminare le fluttuazioni di temperatura a breve termine e concentrarsi sulle tendenze a lungo termine.

E' importante notare che l'uso di un passa-basso può introdurre una certa distorsione del segnale; ad esempio, se la frequenza di taglio del filtro è troppo bassa, potrebbe eliminare alcune delle variazioni di temperatura che sono effettivamente importanti.

E' dunque evidente come, in modelli climatici che devono essere attendibili e rappresentativi dei dati raccolti, la scelta di applicare un filtro passa-basso dovrebbe essere relegata solo a catalogazioni di dati riguardanti località geografiche e/o periodi di temperatura caratterizzati da naturale ridotta variabilità. E' altresì evidente che un grafico del clima dovrebbe, per essere attendibile e rappresentativo, soddisfare determinati requisiti minimi:

- Essere prodotto con un'ottica rivolta alla eccezione, e non alla conformità dei dati
- Essere prodotto mediando i dati basandosi su medie prodotte tramite la salvaguardia dei picchi di valori
- Essere prodotto implementando tecniche di anti-aliasing non deformanti

22 febbraio 2024, Roma

Alessandro Demontis

Perito Chimico (1992) e Tecnico per la Gestione delle Acque e delle Risorse Ambientali (CEE - 1997)

