

IMPLEMENTAZIONE DEL MODULO STATISTICO ALL'INTERNO DI GALEMYS FIELDSTAT



La funzionalità FieldStat è stata elaborata con il contributo teorico dell'Università di Bergamo e permette una prima rapida analisi statistica dei dati letti dai dispositivi di acquisizione.

Gli obiettivi principali di tale funzionalità sono 2:

- 1) Rimuovere dalle letture l'eventuale effetto della temperatura sullo strumento.
- 2) Rimuovere il rumore dallo strumento.

Il primo obiettivo è affidato al cosiddetto Robust Covariate Adjustment, il secondo al Robust Smoothing.

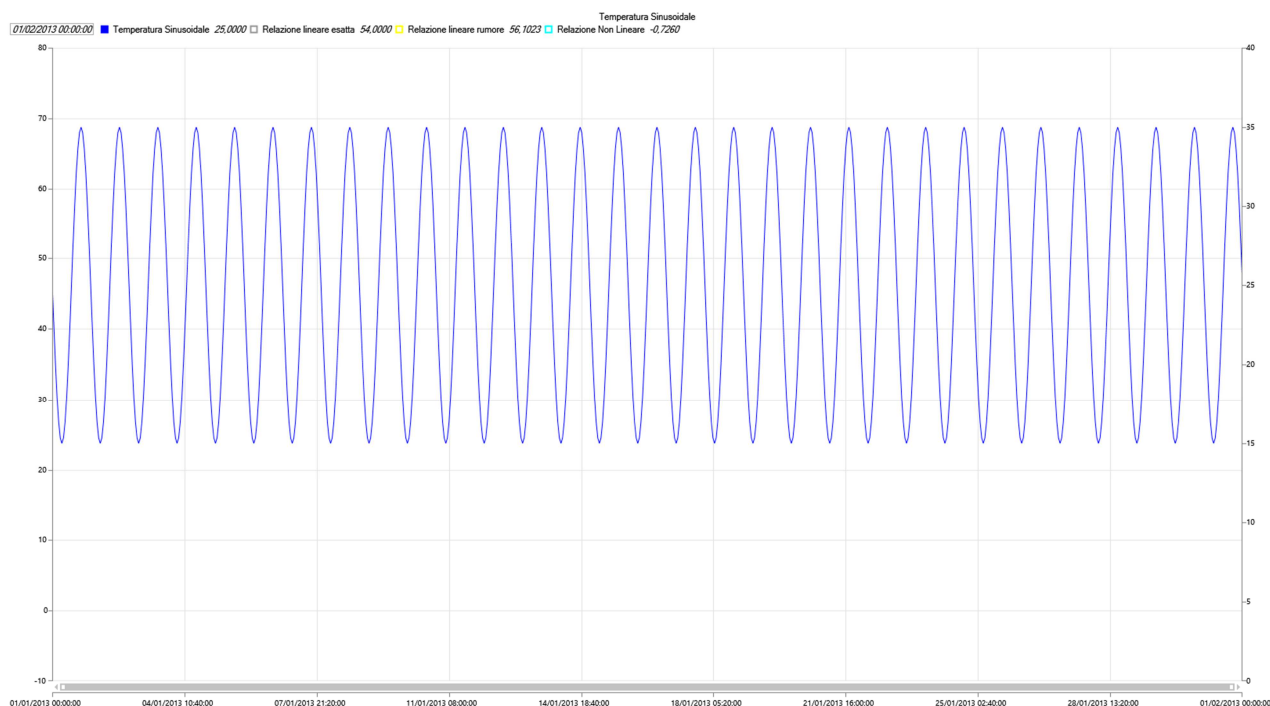
Robust Covariate Adjustment

La correzione delle Coovariate avviene tramite determinazione dell'eventuale correlazione lineare tra le letture e la temperatura. Il termine ROBUST indica un'analisi statistica cosiddetta robusta in luogo della regressione lineare standard, si veda oltre il capitolo "Analisi Robusta" per approfondimenti.

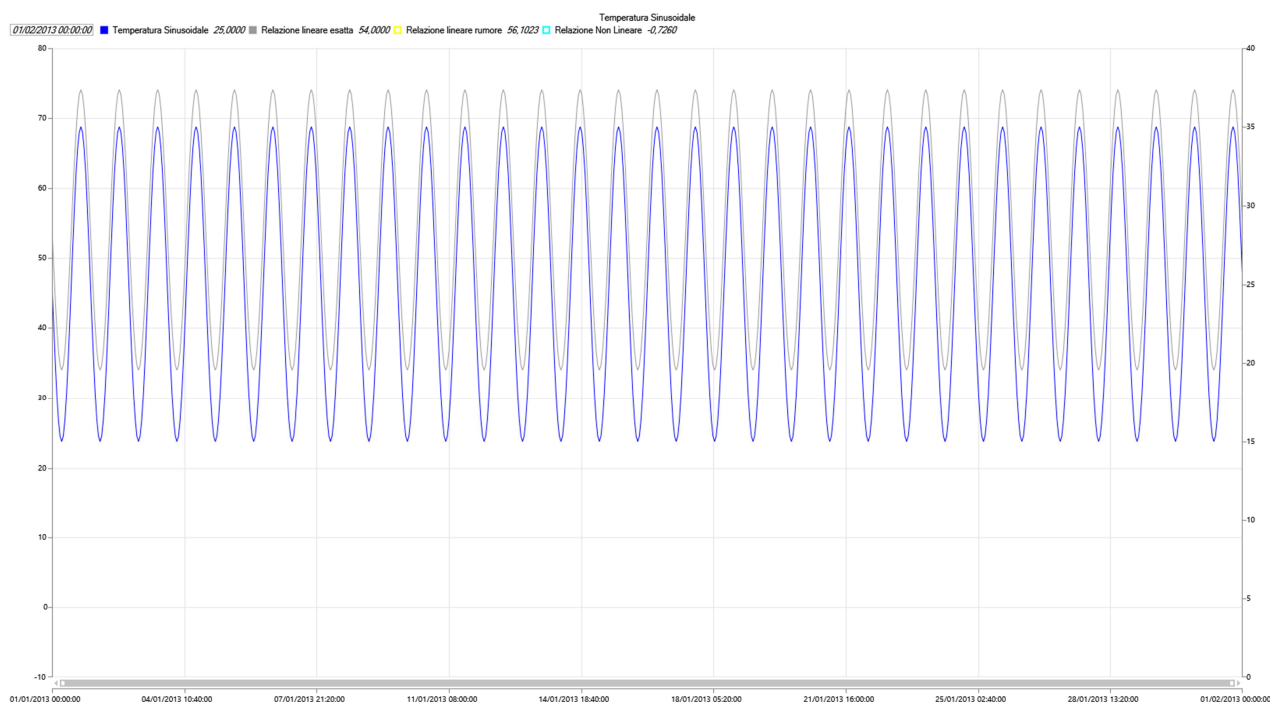
Ogni strumento risente delle condizioni ambientali e, di fatto si comporta in misura maggiore o minore anche come termometro. Tanto più lo strumento risente della temperatura tanto più un grafico delle letture comparato con un grafico delle temperature mostrerà una correlazione. Al crescere della temperatura si osserverà una crescita (o una decrescita) del dato letto e al decrescere della temperatura si osserverà una decrescita (o una crescita) del dato letto.

Rimuovere questo effetto al fine di visualizzare il dato "vero" è l'obiettivo di questo processo.

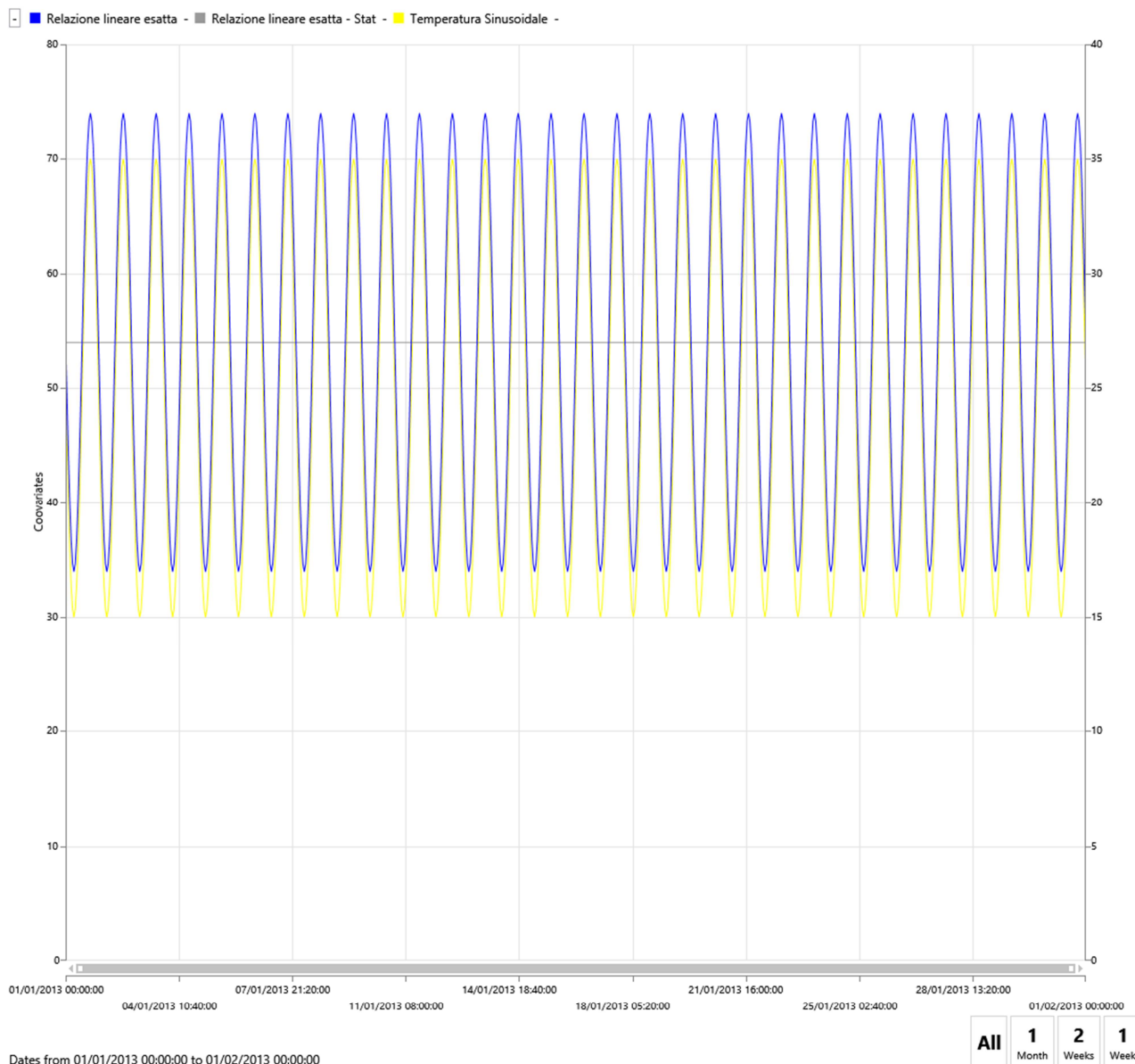
Si consideri a scopo d'esempio una temperatura perfettamente sinusoidale che raggiunga la massima di 25° alle 18:00 e la minima di 10° alle 6:00:



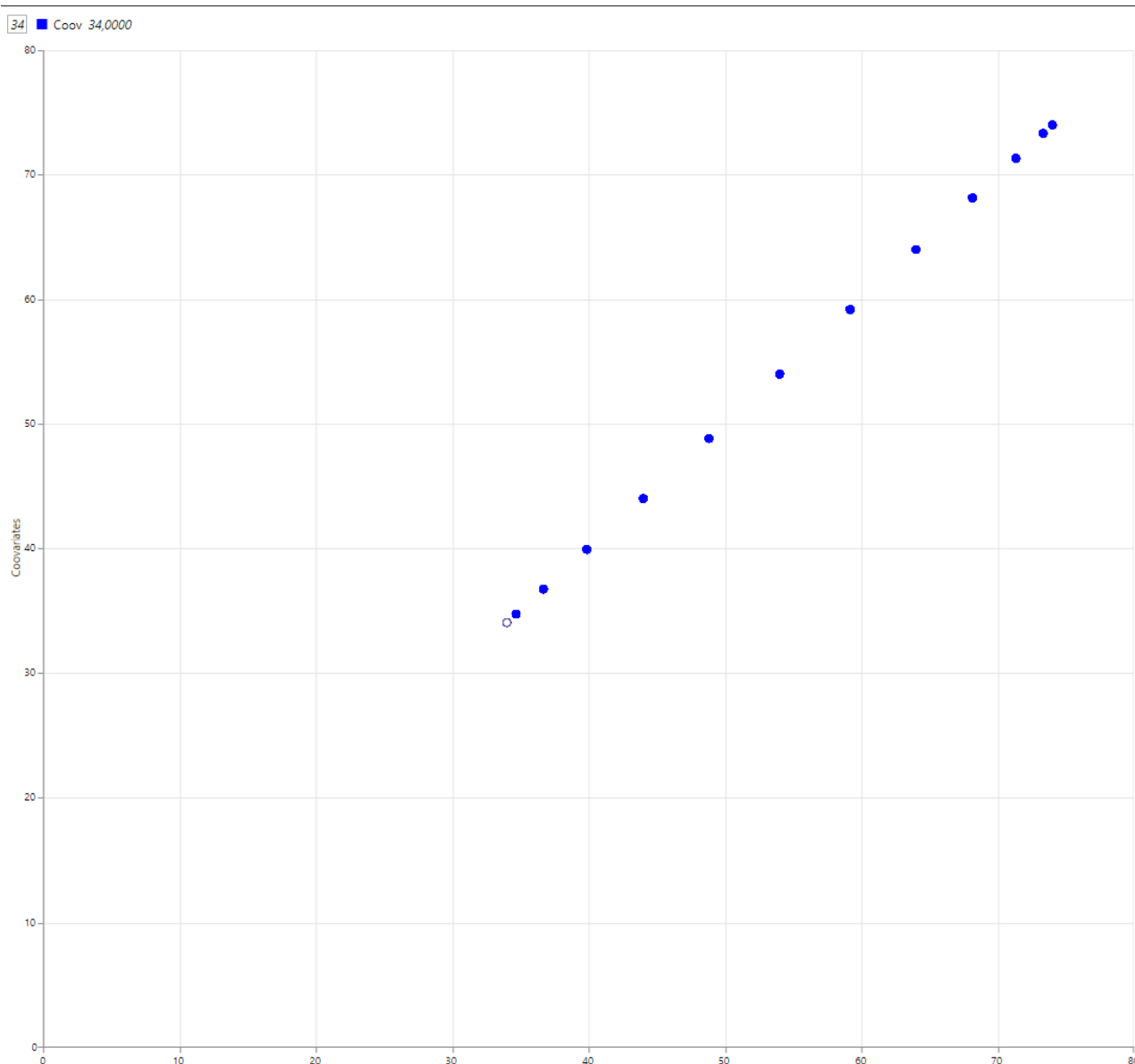
Si consideri ora uno strumento che non rilevi alcuna variazione del proprio dato “reale” ma che subisca in modo perfettamente lineare l’effetto della temperatura secondo la relazione $y = 2 \cdot T + 4$, il grafico risultante avrà un aspetto analogo al seguente:



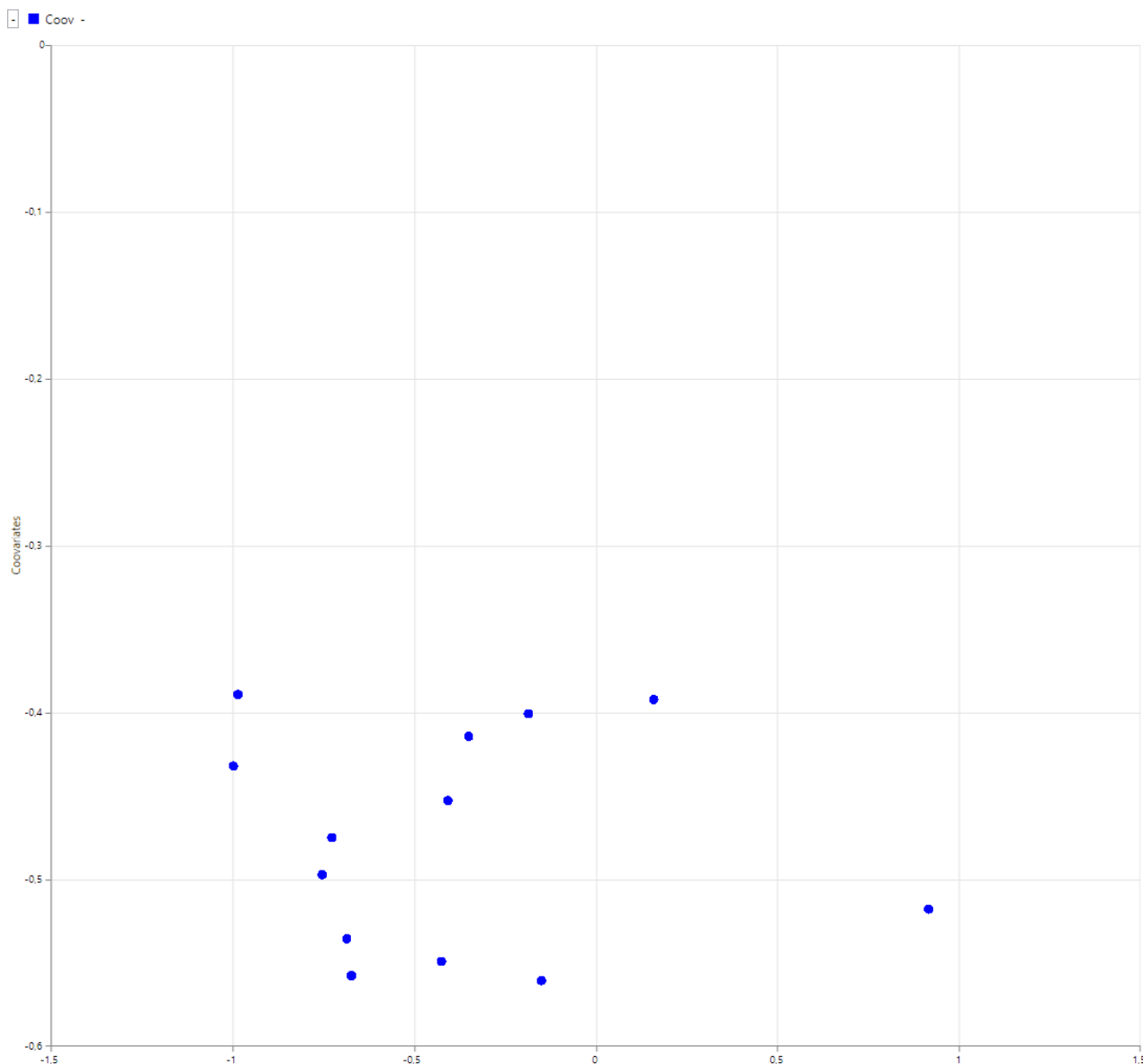
In un caso così idealizzato, la funzione Coovariate Adjustment di FieldStat permette di rimuovere completamente l’effetto della temperatura ottenendo, come atteso, una linea retta:



FieldStat offre inoltre una visualizzazione denominata Scatter Plot che permette di ottenere in forma visuale la correlazione tra la lettura e le temperature. Tanto più i punti si collocano lungo una direzione lineare e definita, tanto maggiore è la correlazione tra i dati, tanto più il grafico mostra punti diffusi, tanto minore risulta la correlazione. Nel caso forzatamente perfetto in esempio di correlazione esatta, il grafico di Scatter ha il seguente aspetto:



Un esempio di Scatter Plot di uno strumento che NON mostra correlazione tra lettura e temperature e quindi scarsa sensibilità alle variabili d'ambiente è il seguente:

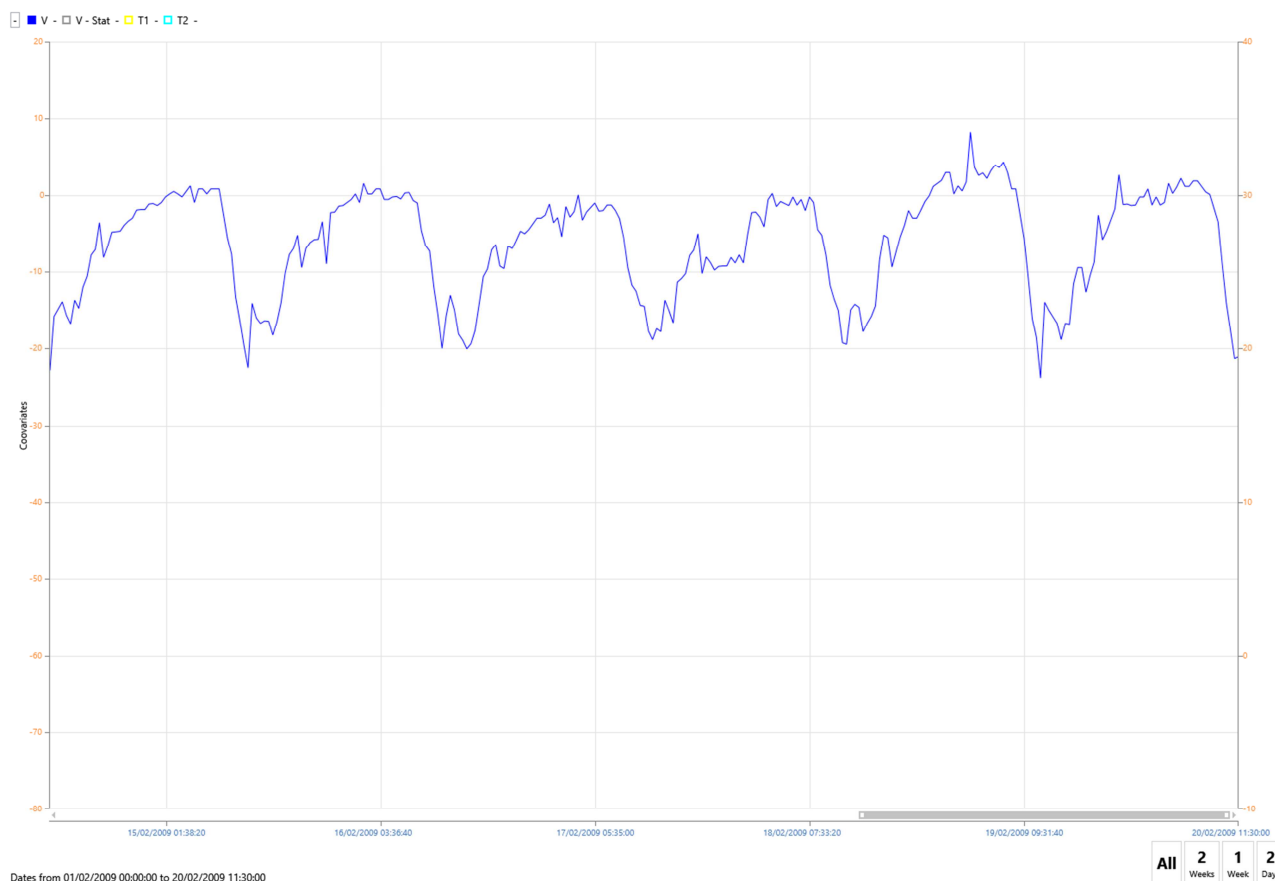


In un caso come questo risulta sostanzialmente superfluo eseguire il Covariate Adjustment.

Robust Smoothing

Lo Smoothing è una tecnica di “addolcimento” delle asperità di un grafico. L’obiettivo principale dello Smoothing è quello di attenuare il rumore naturale di uno strumento al fine di rendere palese l’andamento reale del dato. FieldStat esegue un calcolo cosiddetto Robusto dello Smoothing; si veda oltre il capitolo “Analisi Robusta” per approfondimenti.

Consideriamo il seguente grafico di un dato reale, non ha importanza in questa sede di che tipo di strumento si tratti:

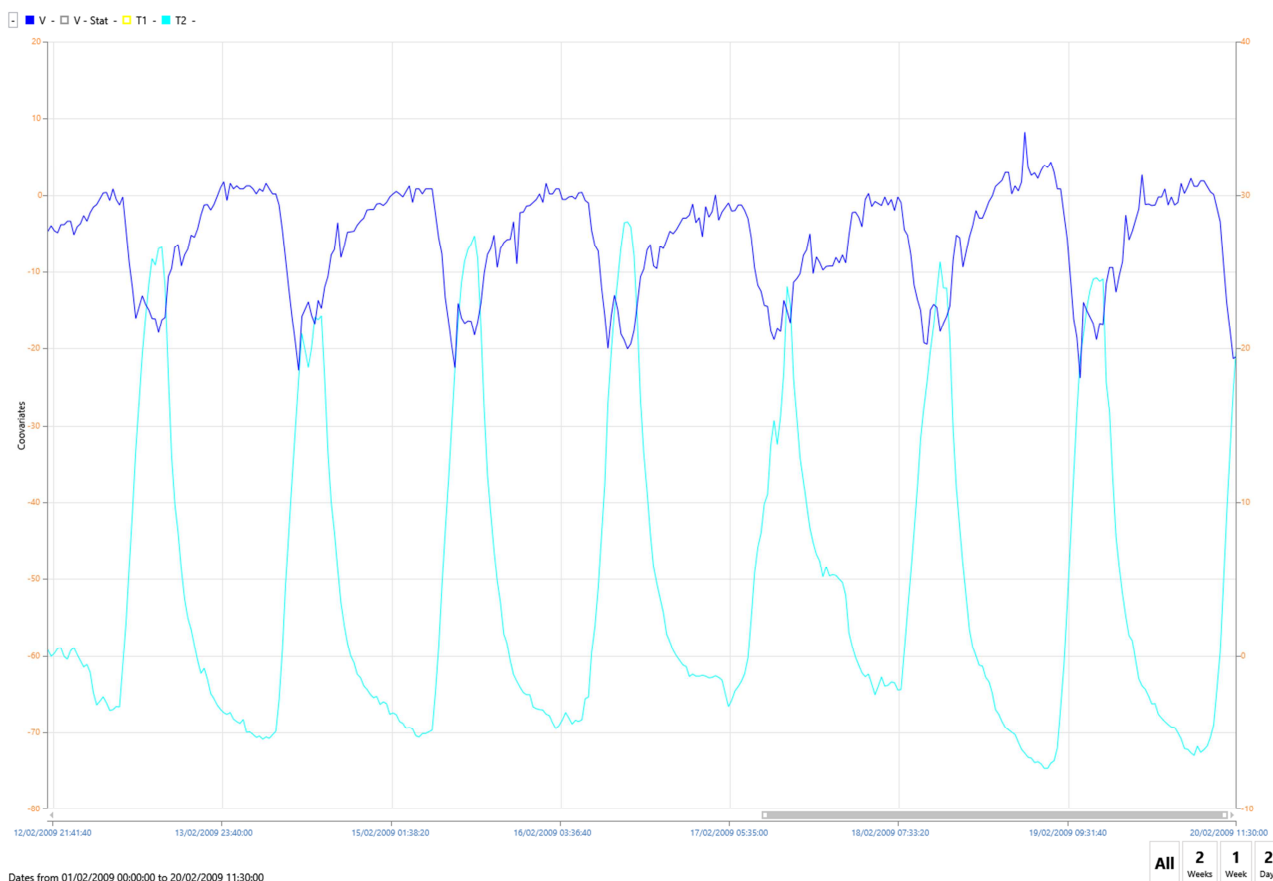


Lo Smoothing agisce rimuovendo le asperità e riducendo il grafico al suo andamento meno puntuale. Matematicamente l'effetto è ottenuto con una regressione lineare locale ai minimi quadrati che ricostruisce la nuova posizione di ogni punto (x,y) in funzione di tutto un intorno $[x-a, x+a]$ del punto stesso. L'effetto ottenuto è analogo a:

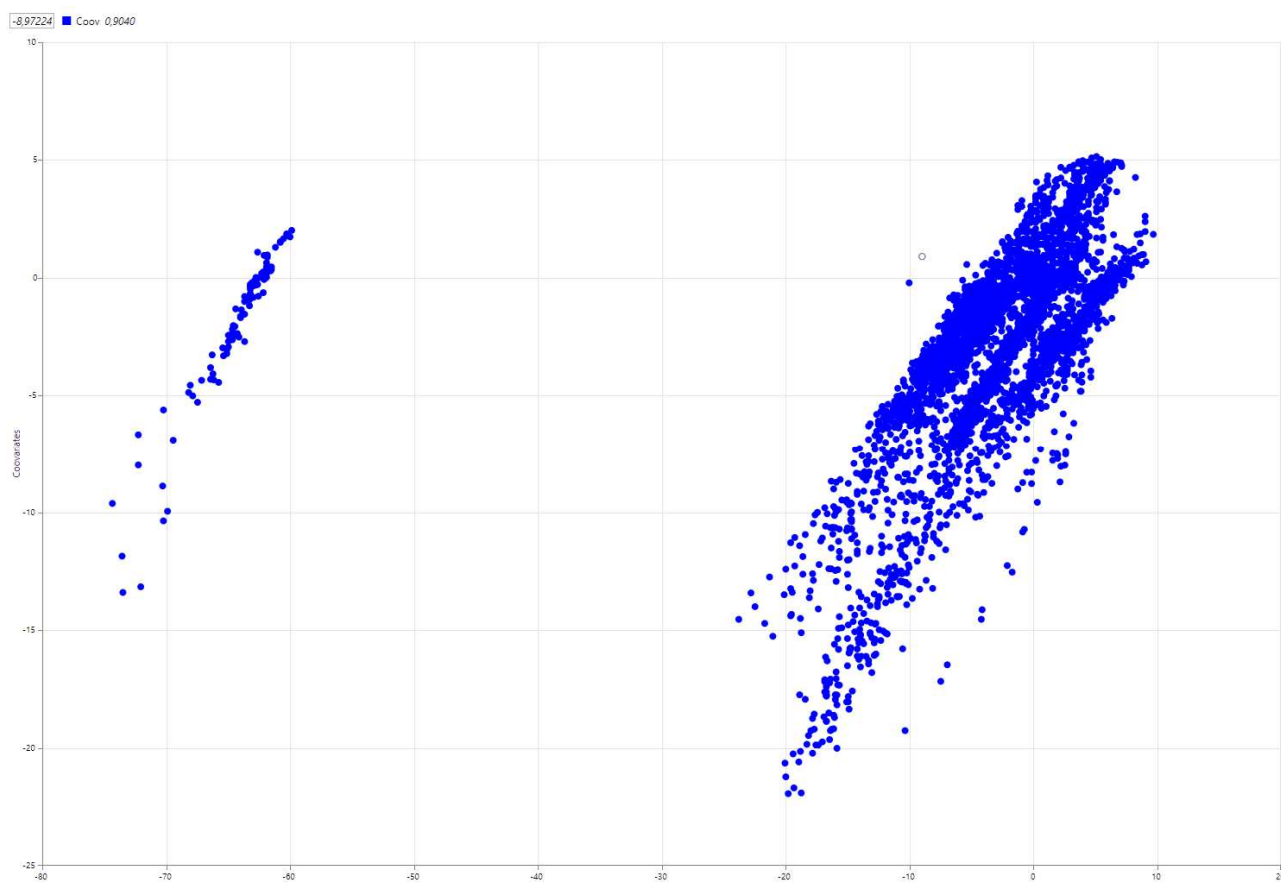


Il caso reale di cui al punto precedente mostra un curioso andamento di crescita e decrescita a intervallo giornaliero, sintomo di una probabile dipendenza del dato dalla temperatura. L'aggiunta della linea di temperatura e il grafico di Scatter confermano tale tesi:

  **CQOP**  **SOA**
COSTRUTTORI QUALIFICATI OPERE PUBBLICHE

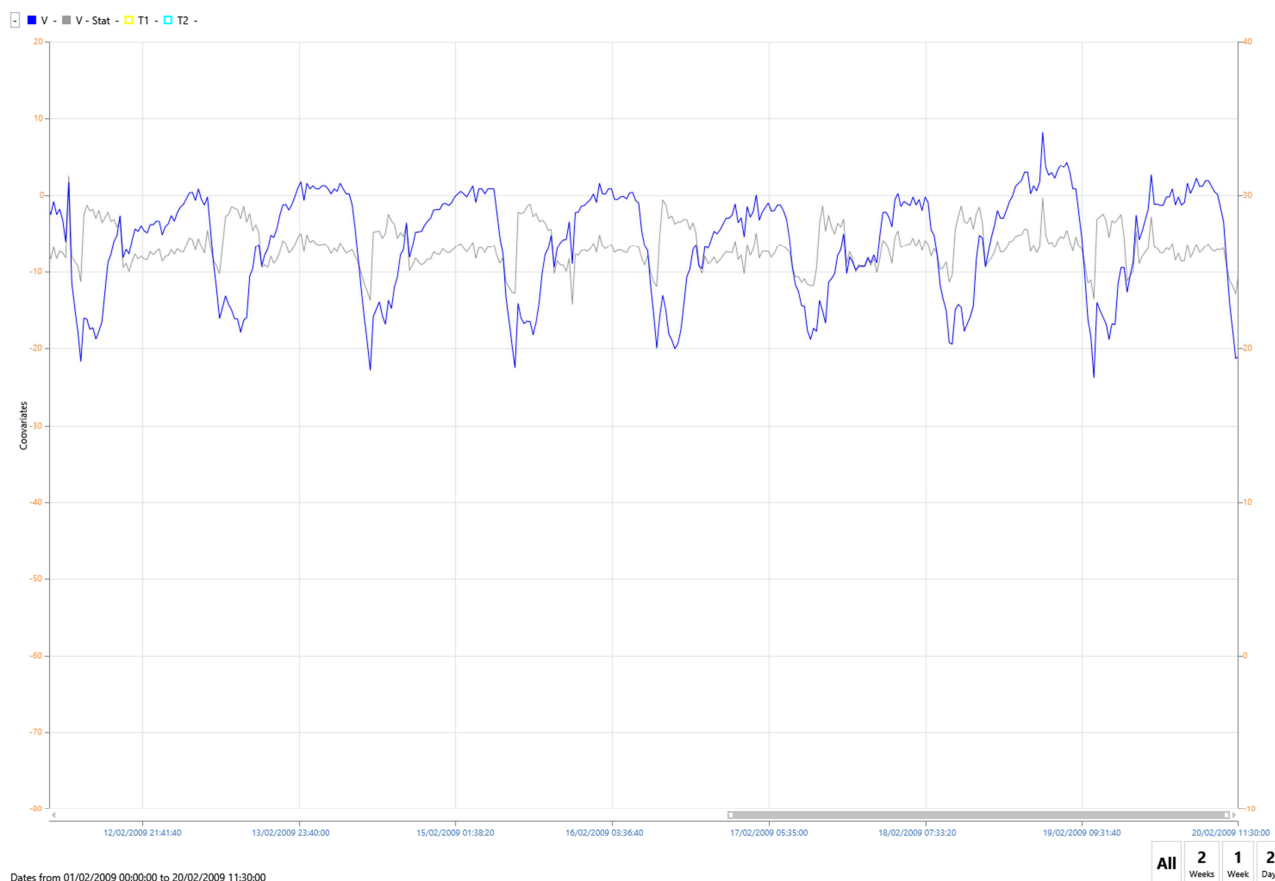


FIELD S.r.l. - Via Provinciale, 44 - 24040 Lallio (BG) - Italy - tel.: +39/035203471; fax: +39/035203448
 Cap Soc. 50.000 EURO i.v. - P. IVA./VAT No. IT 02718860162 Reg. Imprese BG n. 02718860162
 e-mail: info@field srl.it - <http://www.field srl.it>

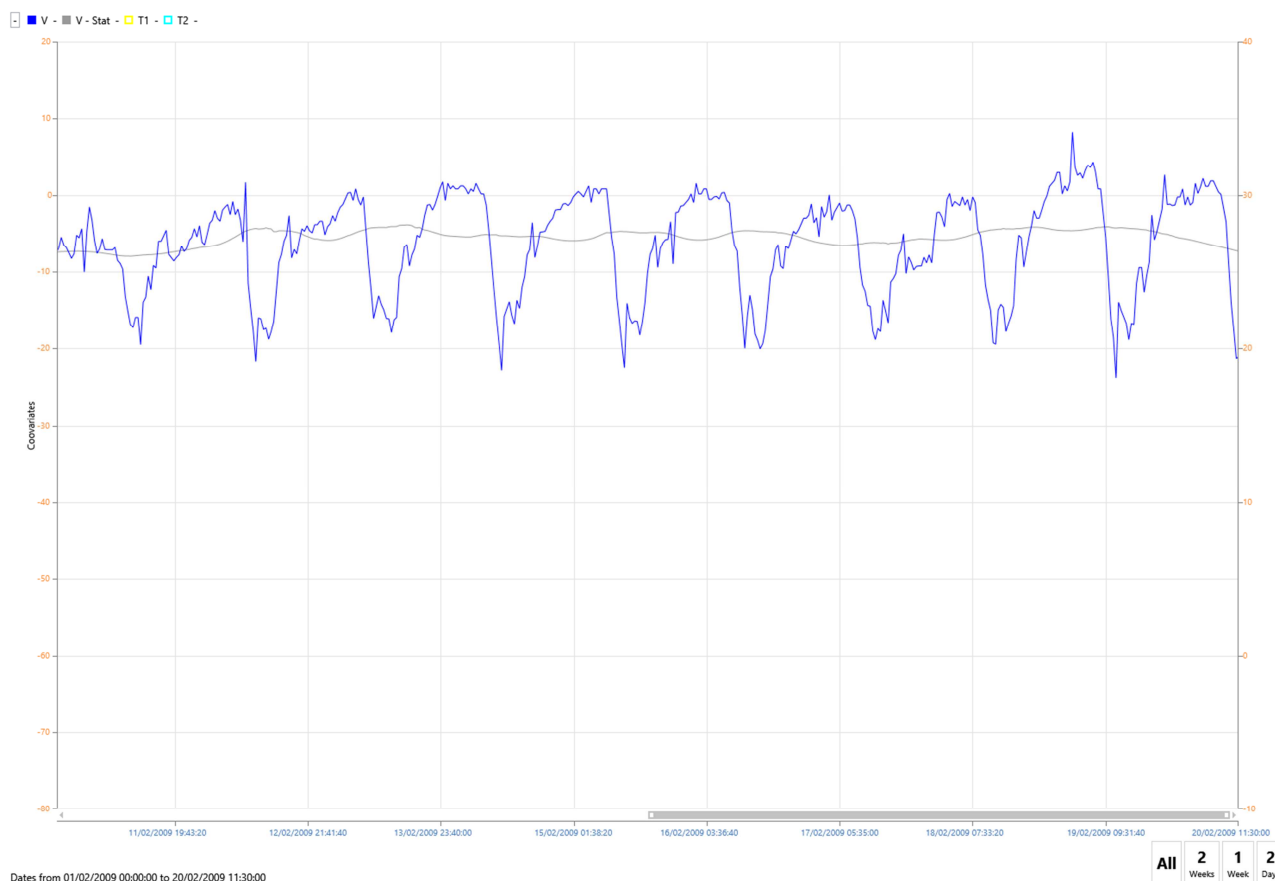


Contestualmente si osservi come il grafico di Scatter mostri una chiara relazione tra temperatura e valori con una piccola serie di punti che “sfuggono” alla linearità. Scopo dell’analisi robusta è proprio trattare adeguatamente questi valori anomali. Si veda oltre il capitolo “Analisi Robusta” per approfondimenti.

Data l’evidente correlazione con la temperatura si procede applicando l’aggiustamento delle Coovariate, ottenendo un chiaro smorzamento dell’effetto:



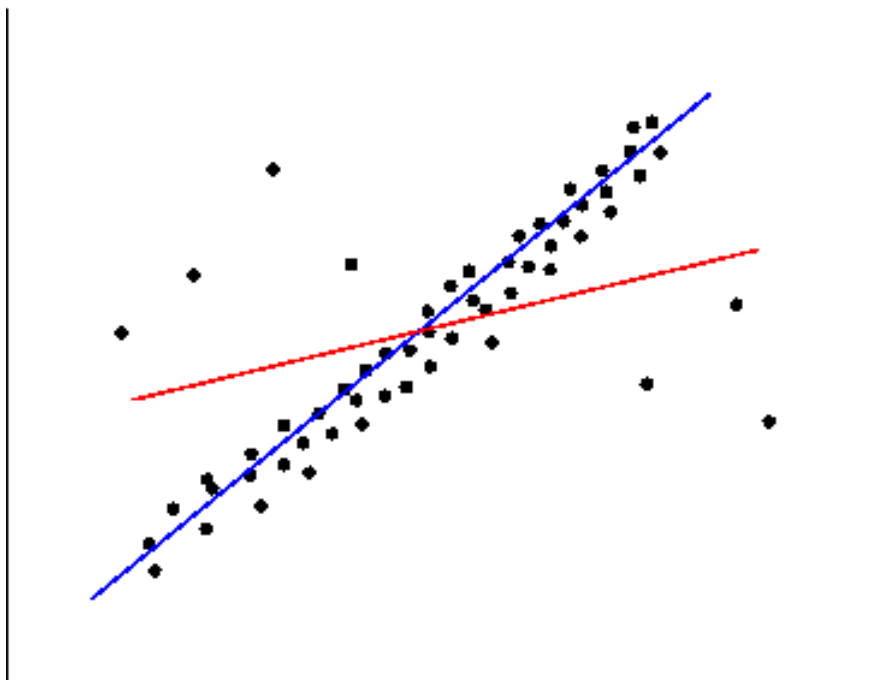
Applicando quindi lo Smoothing è possibile rimuovere il rumore che agisce sulla curva e renderne più leggibile l'andamento:



Si osservi come la linea in grigio risulti più semplificata della serie originale in blu ma in compenso, tale “semplicità” permetta una valutazione più efficace del dato “reale” o, quantomeno, sia utile nella valutazione di tale “reale stato dei fatti”.

Analisi Robusta

L’analisi statistica standard risulta naturalmente essere teoricamente esatta. Tuttavia questo tipo di analisi soffre talvolta della propria stessa esattezza. Si consideri il seguente caso: Data una serie di punti determinare la retta che ne approssima meglio l’andamento.



L'analisi standard, tramite regressione ai minimi quadrati, permette di determinare la linea in rosso come soluzione del problema. Ad un umano la soluzione migliore appare invece essere la linea in blu poiché il cervello è immediatamente in grado di dare “meno peso” ai punti sparsi esternamente considerandoli in qualche misura “eccezioni” e si concentra istintivamente sul fatto che il grosso dei punti risulta allineato lungo la linea azzurra.

L'analisi robusta, detto in due parole, cerca di riprodurre questo atteggiamento umano istintivo sfruttando l'analisi iterata del problema e riuscendo a pesare il contributo di ogni singolo punto in relazione all'insieme dei punti.

L'analisi robusta risulta ovviamente enormemente più onerosa in termini di capacità di calcolo del pc.

In internet è possibile approfondire questo argomento particolarmente tecnico e complesso partendo da qui:

<https://www.google.it/search?q=robust+analysis>