

## Metodologia

La tecnica "dell'*optimal interpolation*" è alla base della metodologia statistica utilizzata per la creazione di un *dataset* di precipitazioni e temperature massime e minime giornaliere su una griglia regolare tridimensionale predefinita.

Questa tecnica soddisfa la necessità di effettuare un processo di interpolazione spaziale, su griglia regolare, dei dati a disposizione, derivanti da stazioni di rilevamento appartenenti a differenti reti di misurazione, sia con diversa copertura temporale sia con una distribuzione spaziale arbitraria e variabile sul territorio regionale, sfruttando il principio che solo le osservazioni più vicine possono influenzare il valore finale sul punto griglia.

Attraverso tale approccio, si rinuncia quindi ad avere un'informazione dettagliata sul singolo punto-stazione, ma si ottiene un campo di analisi su griglia, completo del maggior numero possibile di informazioni del passato, omogeneo ed integrabile con altri dati eventualmente disponibili in futuro, che può essere utilizzato per analisi statistiche sulla variazione relativa nel tempo dei parametri meteo-climatici. Il campo di analisi, ricavato con l'*optimal interpolation*  $x^a$  sui nodi di una griglia regolare, è ottenuto ponendo in relazione lineare l'incremento dell'analisi ( $x^a - x^b$ , dove  $x^b$  è il valore del campo di *background* in corrispondenza degli stessi nodi della griglia) ed il vettore innovazione ( $y^o - y^b$ ) (Kalnay<sup>(1)</sup> e Uboldi<sup>(2)</sup>):

$$x^a = x^b + K(y^o - y^b) \quad (1)$$

dove  $y^o$  rappresenta il vettore dei valori della variabile meteorologica misurati dalle stazioni e  $y^b$  è il valore che il campo di *background* assume sui punti corrispondenti alle stazioni stesse. Minimizzando la varianza dell'errore dell'analisi si ottiene la definizione della matrice guadagno K:

$$K = G(S + O)^{-1} \quad (2)$$

dove G è la matrice delle covarianze tra l'errore del campo di background sul punto griglia e l'errore del campo di background sul punto stazione, S è la matrice delle covarianze dell'errore del campo di background su coppie di punti stazione e O è la matrice delle covarianze dell'errore sull'osservazione.

## I dati osservati e la scelta del passo griglia

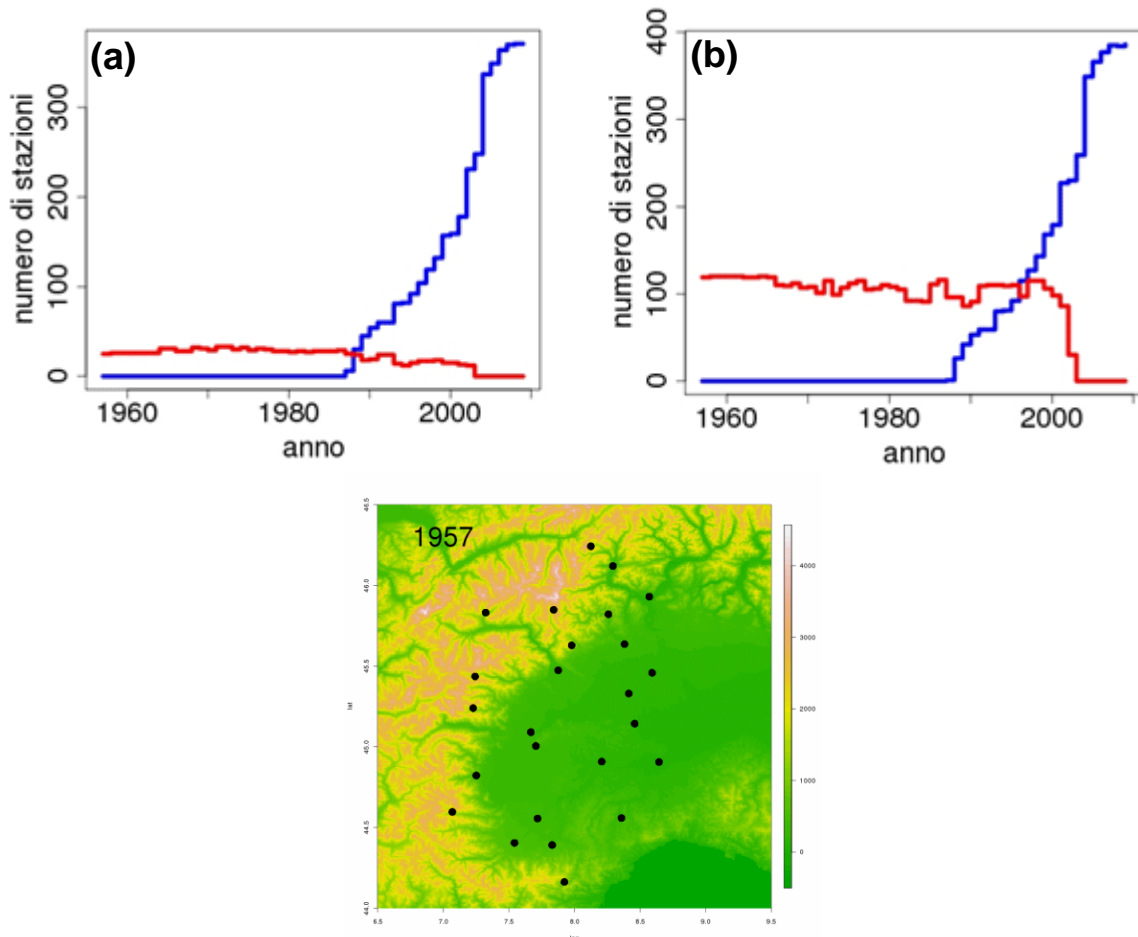
I dati osservati di temperatura e precipitazione, implementati nel *dataset*, derivano da due reti di misurazione che differiscono non solo per il numero di stazioni e per il periodo di funzionamento, ma presentano anche caratteristiche differenti nei sensori e nella loro gestione:

- Rete di stazioni meccaniche e manuali SIMN (disponibile dal 1913 al 2002)
- Rete di telemisura di Arpa Piemonte (disponibile a partire dal 1988).

Sono stati realizzati controlli di qualità per verificare la fruibilità di tali dati, sottoponendoli preventivamente a verifiche di tipo logico (che garantissero la coerenza del dato ed escludessero valori registrati fisicamente impossibili), di tipo climatologico (tramite i quali sono stati confrontati i singoli valori delle serie storiche con la climatologia delle stazioni stesse ed eventualmente con quelle vicine) e di tipo temporale (onde evitare persistenze anomale di un singolo valore).

Una prima analisi esplorativa è stata effettuata per valutare l'andamento della quantità di dati disponibili, consultando le anagrafiche dei pluviometri e dei termometri e ricavando, anno per anno, il numero di stazioni attive. Dalla figura 1 si evince infatti come il numero di sensori

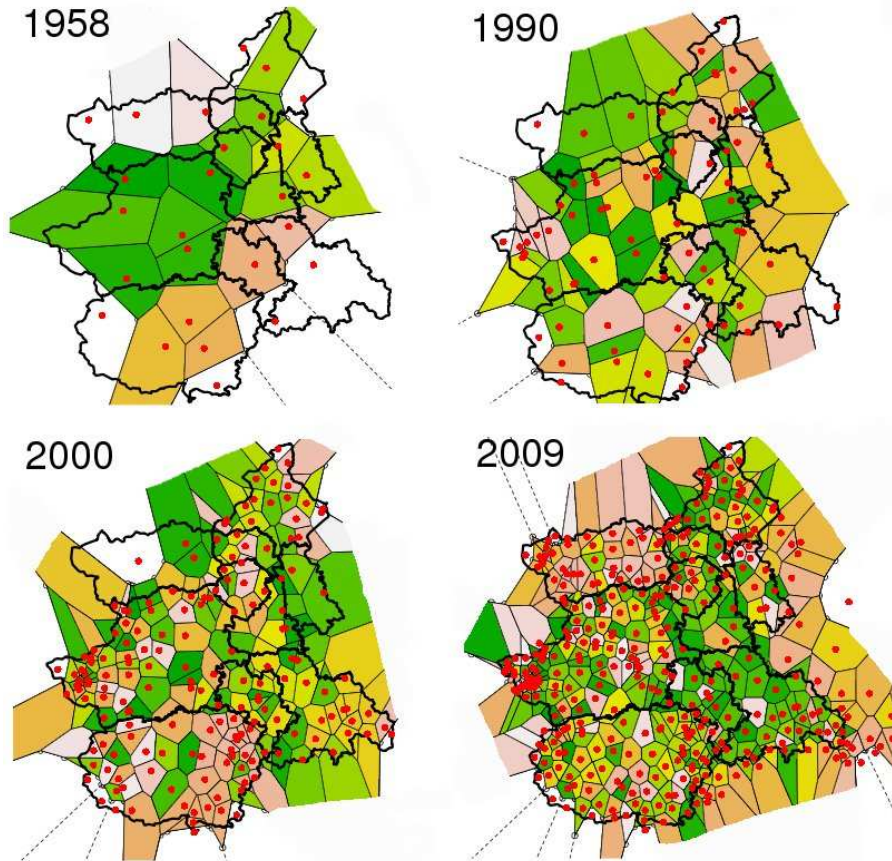
termometrici sia aumentato di oltre dieci volte nell'arco di 50 anni, partendo da circa 25 unità negli anni '50 fino ad arrivare a 371 sensori nel 2009, con un ripido incremento a partire dagli anni '90 in seguito all'introduzione della rete di telemisura. Per quanto riguarda le stazioni pluviometriche, si registra un aumento dei sensori da 119 unità negli anni '50 fino a 386 unità nel 2009.



**Figura 1:** (sopra) numero di stazioni attive della rete SIMN (rosso) e della rete Arpa Piemonte (blu) per le temperature (a) e le precipitazioni (b) nel periodo 1957-2009; (sotto) distribuzione delle stazioni di temperatura sul territorio piemontese all'inizio del periodo di analisi.

L'analisi del numero e della distribuzione spaziale delle stazioni è fondamentale anche per stimare un passo adeguato della griglia sulla quale rappresentare il campo di temperatura e di precipitazione ottenuto con l'*optimal interpolation*. A tal fine è stata effettuata un'analisi in ambiente GIS per ottenere una stima della media del raggio di influenza di ciascuna stazione. È stata quindi realizzata una tassellatura di Voronoi, attuando una decomposizione del territorio sulla base della distanza di ciascun punto dello spazio bidimensionale rispetto all'insieme delle localizzazioni delle stazioni di rilevamento. L'insieme dei punti più vicini a una determinata stazione rispetto a tutte le altre è definito come la "cella di Voronoi" di quella stazione e costituisce l'unità di base della tassellatura. Per ciascun anno compreso nel periodo 1957-2009 è stata effettuata una tassellatura sulla base delle stazioni attive e sono state calcolate le aree di ciascuna cella. La media di queste aree, insieme alla media del loro raggio (ottenuto approssimando le celle a delle superfici circolari), è indicativa della distanza media tra le stazioni. Con questo metodo è stato possibile ottenere una stima finale non influenzata da valori estremi a causa dell'aumento delle stazioni presenti, come avverrebbe invece considerando la media delle distanze di ciascuna stazione rispetto a tutte le altre. Non è inoltre necessario individuare a priori un numero soggettivo di k vicini, rispetto ai quali calcolare la distanza di ciascuna stazione, come

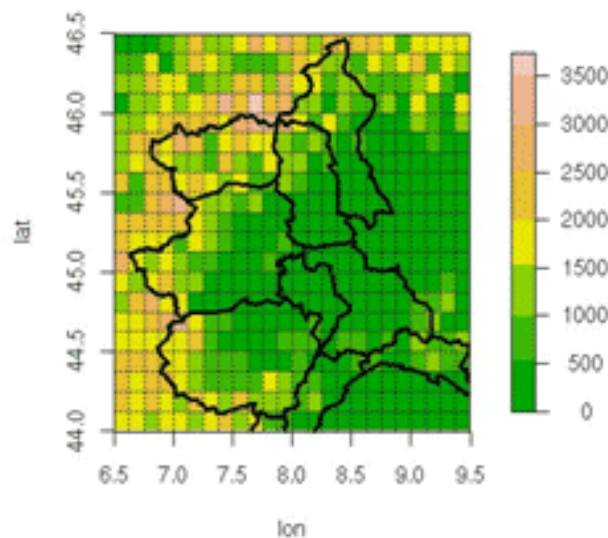
avverrebbe in analisi basate su algoritmi *nearest neighbour* delle semplici distanze tra le stazioni. Inoltre questo metodo ha il vantaggio di accompagnarsi ad una rappresentazione cartografica che può costituire uno strumento di analisi e di supporto per la localizzazione di criticità locali nel numero di stazioni presenti. I risultati ottenuti sono mostrati in figura 2.



**Figura 2:** tassellatura di Voronoi ricavata a partire dalle localizzazioni delle stazioni termometriche, indicate in rosso, negli anni 1958, 1990, 2000 e 2009. La valutazione delle aree e dei raggi medi delle celle per ciascun anno ha permesso di stimare la risoluzione ottimale dell'optimal interpolation per il periodo di studio 1957-2009.

Al termine di questa analisi preliminare, la risoluzione di 15 km è stata individuata come meglio rappresentativa dell'intero periodo preso in esame, 1957-2009.

L'altezza topografica di ciascuna cella della griglia è stata ottenuta con un *up-scaling* alla risoluzione di 15 km a partire da un DEM di risoluzione 100 m, applicando una tecnica di interpolazione bilineare. Successivamente, ove possibile, i valori della quota di ciascun punto griglia sono stati sostituiti dal valor medio delle quote delle stazioni reali presenti nel momento di massima densità della rete.



**Figura 3:** orografia dell'Italia nord-occidentale sulla quale sono stati assimilati i dati di temperatura minima e massima giornalieri e caratteristiche principali della griglia.

Risoluzione	Lat x Lon	n° punti Dominio	n° punti Piemonte	Estremi Dominio
0.125°	24 x 20	480	297	Longitudine 6.5-9.5 W Latitudine 44.0-46.5 N

**Tabella 1:** caratteristiche principali della griglia di interpolazione dei dati scelta.

## Il campo di *background*

Come precedentemente citato, nel processo di interpolazione l'analisi è calcolata sulla base di un campo di appoggio denominato campo di *background*. Nel caso delle temperature è stato possibile utilizzare un campo modellistico, ottenuto mediante un *downscaling* del set di reanalisi ERA-40 elaborato dall'ECMWF – *European Centre for Medium range Weather Forecast* (disponibile a risoluzione 1.125°) per il periodo 1957-2001 e del set di analisi dell'ECMWF (disponibile a risoluzione 0.500° fino al 2006 e successivamente a risoluzione 0.250°) per il periodo 2002-2009. Preventivamente si è verificato che i segnali climatologici principali, forniti da ERA-40 per il territorio regionale sulle temperature, fossero congruenti con quelli derivanti da un sottoinsieme omogeneo, in termini di tipologia di sensore e, laddove possibile, di densità spaziale, di stazioni presenti sul territorio nel periodo tra il 1950 e il 2000 (Ciccarelli).

Nel caso della precipitazione non è stato possibile usufruire di un campo esterno ed il *background* è stato quindi calcolato a partire dalle osservazioni stesse, mediante una procedura di *detrending* (Uboldi<sup>(2)</sup>).

## Il campo IDI e la definizione dei parametri di scala

Uno dei principali obiettivi del lavoro svolto è stato quello di mantenere la matrice guadagno K dell'algoritmo dell'*optimal interpolation* il più costante possibile, al fine di compensare la grande variabilità nella densità delle stazioni di rilevamento sul territorio nell'intero periodo in esame. Mantenere la matrice guadagno costante permette, infatti, di conservare il più possibile l'omogeneità temporale del segnale e garantisce che il *dataset* finale sia affetto solo marginalmente dalla non uniformità della distribuzione dei dati.

E' stata quindi svolta un'analisi approfondita del campo IDI, *Integral Data Influence*, cioè l'analisi dei risultati che si ottengono ponendo i valori del campo di *background* pari a zero ed il valore delle osservazioni pari a uno. L'equazione (1) si riduce a:

$$x^a = K \quad (3)$$

$$K = G(S + O)^{-1} \quad (4)$$

Assumendo che le correlazioni dell'errore di *background* nelle matrici G e S possano essere separate nel prodotto tra la correlazione orizzontale e quella verticale (Kalnay) e utilizzando funzioni gaussiane per definire le correlazioni in funzione della sola distanza tra le stazioni ed i punti griglia, la matrice guadagno dipende fundamentalmente dai parametri di scala  $\sigma_{oriz}$  e  $\sigma_{vert}$ , oltre che da  $\varepsilon^2$ , che è legato alla matrice di covarianza degli errori delle osservazioni e che stima la confidenza nelle osservazioni rispetto al campo di background:

$$G(ii, jj) = \exp(-d_{ijoriz}^2 / 2\sigma_{oriz}^2) * \exp(-d_{ijvert}^2 / 2\sigma_{vert}^2) \quad (5)$$

dove  $d_{ij}$  rappresenta la distanza tra il punto griglia i-esimo e la stazione j-esima

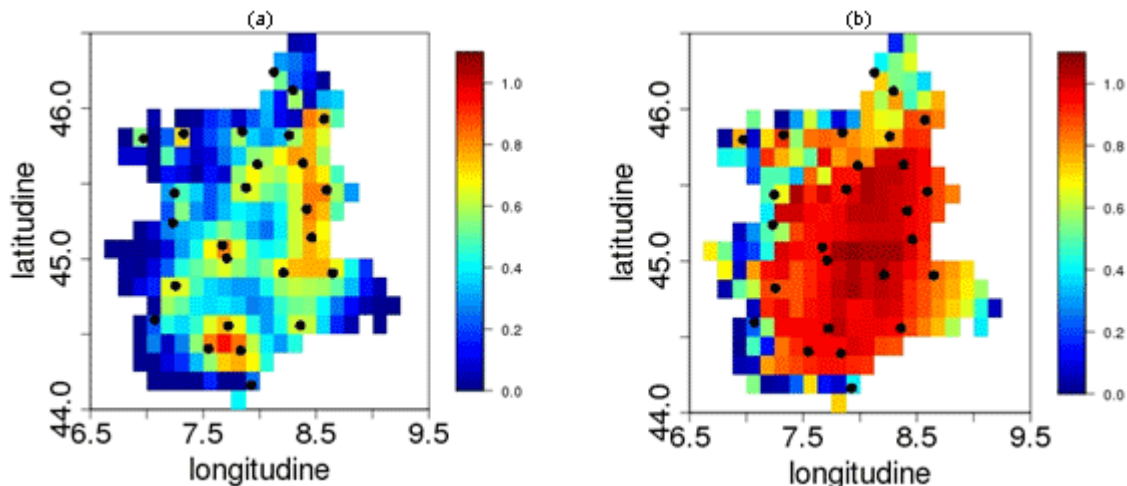
$$S(ii, jj) = \exp(-d_{ijoriz}^2 / 2\sigma_{oriz}^2) * \exp(-d_{ijvert}^2 / 2\sigma_{vert}^2) \quad (6)$$

dove  $d_{ij}$  rappresenta la distanza tra le coppie di stazioni i-esima e j-esima

$$O = \varepsilon^2 I \quad (7)$$

dove I è la matrice identità.

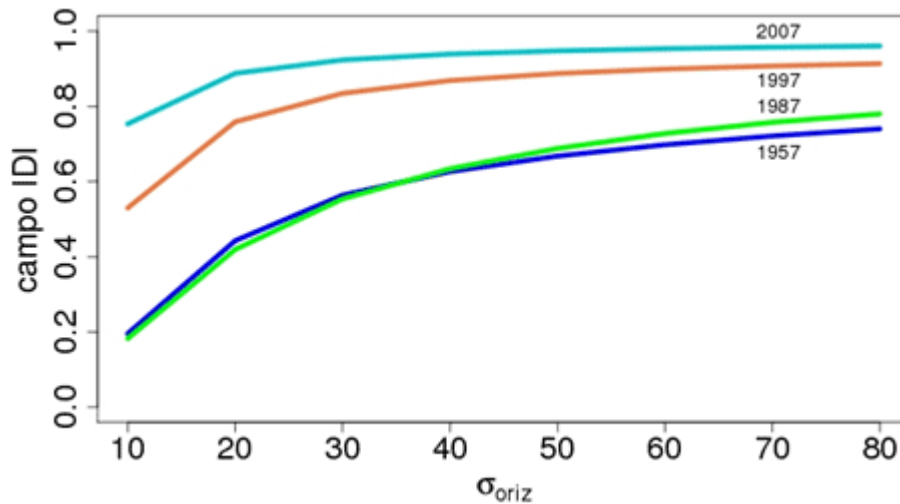
Ne consegue che, se le funzioni di correlazione si approssimano a zero al crescere della distanza dal punto stazione, il campo IDI sarà approssimativamente pari a 1 nelle porzioni di territorio prossime alle stazioni mentre tenderà a zero nelle zone prive di sensori, e l'area di "influenza" di ciascuna stazione sarà proporzionale al valore dei parametri di scala  $\sigma_{oriz}$  e  $\sigma_{vert}$ .



**Figura 4:** temperatura - variazione del campo IDI osservata per uno stesso anno (1958) al variare del parametro di scala  $\sigma_{oriz}$  con  $\sigma_{vert}=500$  m e  $\varepsilon^2=0.5$ . I punti neri rappresentano le 26 stazioni di rilevamento attive nell'anno in esame. (a)  $\sigma_{oriz} = 15$  km, (b)  $\sigma_{oriz} = 60$  km.

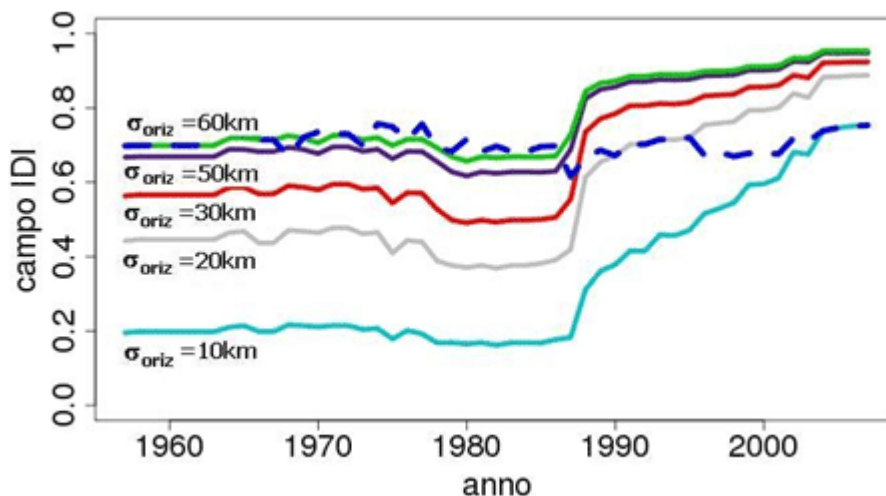
La scelta dei parametri di scala rappresenta dunque un importante elemento per mascherare o amplificare il contributo totale delle stazioni presenti sul territorio e filtrare la disomogeneità del segnale dovuta al numero variabile di sensori.

Per ogni anno sono stati quindi modulati opportunamente i parametri di scala in funzione del numero di stazioni, per mantenere il campo IDI prossimo ad un valore di riferimento individuato pari a 0.8. Si tenga presente che incrementare notevolmente ed arbitrariamente i parametri di scala non porta conseguentemente ad un valor medio del campo IDI sul dominio pari a 1, poiché l'IDI tende ad un valore asintotico che è strettamente legato al numero delle stazioni. Per questo motivo il valore di riferimento 0.8 è stato definito come il minimo tra i valori asintotici calcolati per ogni anno (vedi figura 5).



**Figura 5:** temperatura - andamento del campo IDI in funzione del fattore di scala  $\sigma_{oriz}$ , fissati  $\sigma_{vert}=500$  m e  $\varepsilon^2=0.5$ . Il valore massimo che il campo IDI può raggiungere è funzione anche del numero di stazioni attive in quell'anno: all'aumentare del fattore di scala  $\sigma_{oriz}$ , il campo IDI medio tende ad un valore asintotico.

Sono stati inizialmente fissati i parametri  $\sigma_{vert}$  e  $\varepsilon^2$  pari a valori suggeriti dalla letteratura (Uboldi<sup>(2)</sup>) ed è stato variato  $\sigma_{oriz}$  in un intervallo compreso tra 10 e 60 km, in modo tale che il valore del corrispondente campo IDI, mediato sull'intero territorio regionale, fosse prossimo al valore di riferimento 0.8. Successivamente è stato ripetuto lo stesso metodo utilizzando il set di  $\sigma_{oriz}$  precedentemente trovato e variando  $\varepsilon^2$  in un intervallo tra 0.2 e 0.8. Il parametro  $\sigma_{vert}$  è stato mantenuto fisso a 500 m, in quanto si è verificato operativamente che la sua variazione non influisce significativamente sul campo risultante (figura 6).

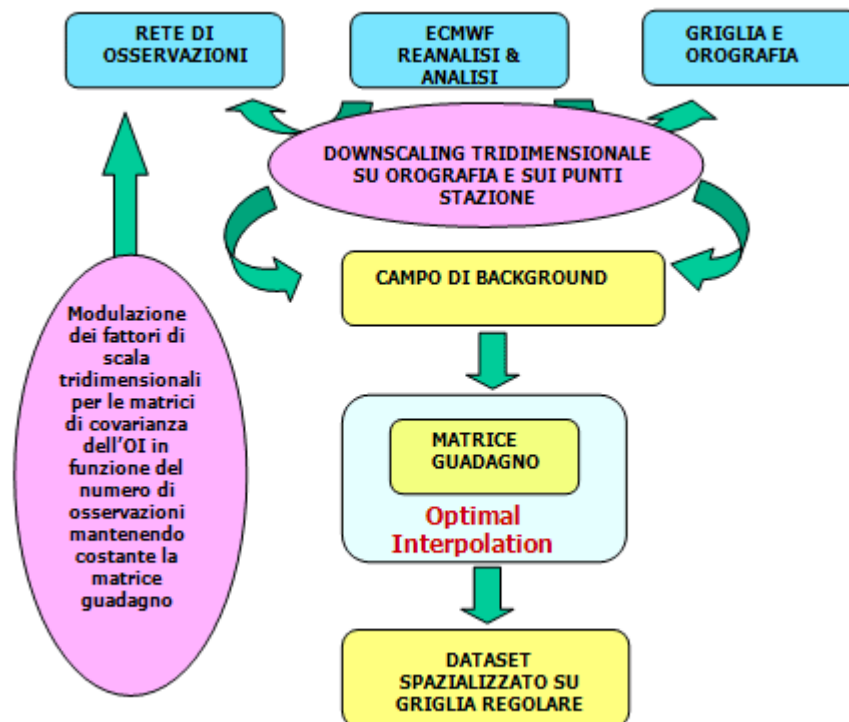


**Figura 6:** temperatura - andamento del campo IDI nel periodo 1957-2009 al variare dei fattori di scala  $\sigma_{oriz}$ , fissati  $\sigma_{vert}=500$  m e  $\varepsilon^2=0.5$ . La linea tratteggiata rappresenta l'andamento del campo IDI utilizzando il set di valori di  $\sigma_{oriz}$  tali da rendere la matrice guadagno costante nell'arco temporale di studio.

Per quanto riguarda la precipitazione, si è ripetuto il procedimento per la determinazione dei parametri di scala del campo IDI in maniera analoga alla temperatura, variando il termine  $\sigma_{\text{oriz}}$  da 10 a 30 km, a causa del maggior numero di sensori pluviometrici disponibili nel periodo. Inoltre l'analisi del campo IDI è stata effettuata impostando un campo quota costante, con valore pari a 2m, per annullare nell' algoritmo dell' *optimal interpolation* la relazione tra il segnale e la quota.

La distribuzione delle precipitazioni in Piemonte evidenzia indubbiamente una relazione con l'altezza topografica del territorio. I flussi umidi provenienti dal Mediterraneo, ai quali è esposto il territorio regionale, interagiscono con la catena montuosa e sono costretti ad un moto ascensionale, determinando una precipitazione di tipo orografico più intensa in corrispondenza della zona di interfaccia con la zona pianeggiante. Tuttavia, dal momento che le diverse aree montane e pedemontane non sono egualmente esposte ai flussi umidi e presentano, a parità di quota, diversa distribuzione spaziale delle precipitazioni anche a causa della differente esposizione dei versanti alpini e della conformazione delle vallate interne, si è preferito eliminare direttamente nell'interpolazione ogni relazione tra precipitazione e quota.

Si è quindi ottenuto, sia per la precipitazione sia per la temperatura, un set di parametri tali da mantenere la matrice guadagno pressoché costante nell'arco temporale studiato e, utilizzando tali parametri nel processo di interpolazione, si è ottenuta la miglior stima possibile del campo di analisi in quelle porzioni di territorio in cui non sono presenti sensori di rilevazione e, al contempo, non è stato sommato arbitrariamente un falso segnale laddove la densità di stazioni varia molto nel tempo.

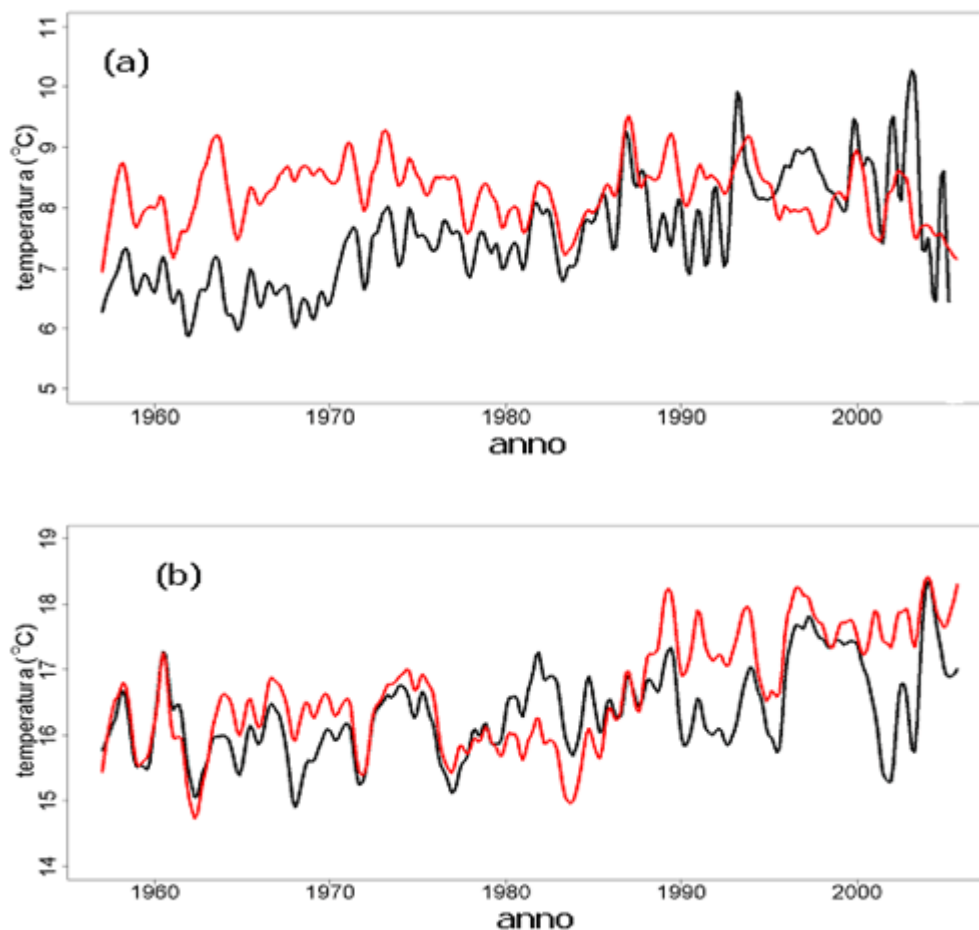


**Figura 7:** schematizzazione compatta della metodologia sviluppata ed applicata, finalizzata alla creazione di un dataset giornaliero di temperature massime e minime e di precipitazioni sul Piemonte nel periodo compreso tra il 1957 sino ad oggi. Per le precipitazioni va escluso l'utilizzo dei dati provenienti dall'ECMWF.

Una volta ottenuti i campi di temperatura e precipitazione spazializzati sulla griglia a 15 km di risoluzione, sono stati effettuati alcuni controlli per valutare la bontà del processo di interpolazione.

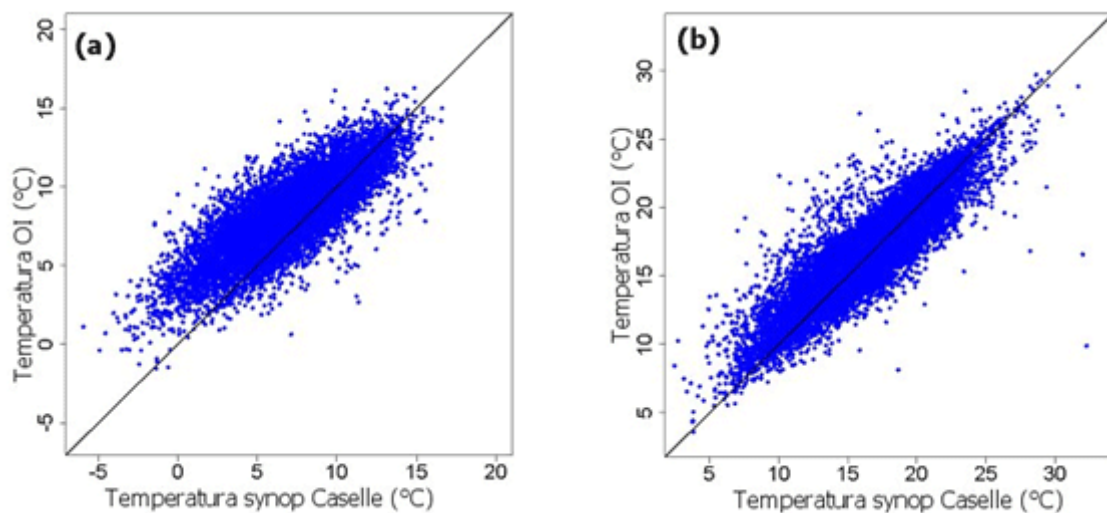
Sono stati quindi confrontate statisticamente alcune serie storiche di stazioni synop della rete GTS, non incluse nel processo di assimilazione, con il corrispondente punto griglia dell'analisi ottenuta con l'*optimal interpolation*. Nel confronto è stato tenuto conto che, trattandosi di una tecnica di spazializzazione, il punto griglia non riproduce fedelmente i dati della stazione corrispondente in quanto non può tenere conto delle caratteristiche della stazione stessa quali quota, esposizione, ecc. L'obiettivo è stato quindi quello di verificare che nelle due serie storiche a confronto si mantenessero le stesse caratteristiche nelle distribuzioni dei dati.

Come esempio, si mostra nelle figure 8 e 9 l'andamento e la dispersione della temperatura massima e minima della stazione SYNOP di Caselle (quota pari a 287 m) e della temperatura del corrispondente punto griglia (quota pari a 273 m), dopo che le serie storiche giornaliere sono state de-stagionalizzate attraverso il filtro di loess.



**Figura 8:** andamento della temperatura minima (a) e massima (b) giornaliera nel periodo 1957-2009 registrata dalla stazione synop di Caselle (nero) e dall'analisi dell'*optimal interpolation* del corrispondente punto griglia (rosso).





**Figura 9:** dispersione della serie storica della stazione synop di Caselle vs la serie storica del corrispondente punto griglia dell'optimal interpolation per le temperature minime (a) e massime (b) nel periodo 1957-2009. I segnali sono stati precedentemente destagionalizzati.

Da entrambe le figure si evidenzia come l'andamento della temperatura massima assegnata al punto griglia in corrispondenza della stazione di Caselle segua meglio lo sviluppo del SYNOP corrispondente, rispetto alla temperatura minima. In particolare, il segnale interpolato evidenzia un andamento mediamente superiore all'osservato nel periodo che arriva fino a fine anni '70, oltre che una dispersione visivamente maggiore nello *scatter-plot* (figura 9).

Va sottolineato tuttavia che il dato interpolato sulla cella della griglia rappresenta una media elaborata di tutti i punti presenti nella cella stessa, e quindi è ragionevole aspettarsi discrepanze contenute quando la corrispondente serie storica viene confrontata con quella di un punto-stazione singolo, seppur facente parte della cella che lo rappresenta.

## References

- (1) Kalnay, E. (2003): *Atmospheric modeling, data assimilation and predictability*. Cambridge Univ. Press, 341 pp.
- (2) Uboldi, F., Lussana, C. and Salvati, M. (2008), *Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks*. *Meteorological Applications*, 15: 331-345, doi: 10.1002/met.76.
- (3) N.Ciccarelli, J. von Hardenberg, A. Provenzale, C. Ronchi, A. Vargiu, R. Pelosini (2008), *Climate variability in north-western Italy during the second half of the 20th century*. *Global and Planetary Change*, 63: 185-195.