

**METODOLOGIE OGGETTIVE PER
LA SPAZIALIZZAZIONE
DELLE MISURE AMBIENTALI**

**Roberto Sozzi
Andrea Bolignano**

Roma - 2018

Indice

1. PREMESSA

2. FORMULAZIONE GENERALE DEL PROBLEMA

2.1 Le misure

2.2 Geometria del Dominio di Analisi

2.3 Conoscenza a priori e il campo di background

3. METODI INTERPOLATIVI FUNZIONALI

4. METODI DI INTERPOLAZIONE OGGETTIVA

4.1 Metodi Semiempirici

4.2 Optimal Interpolation

1. PREMESSA

Immaginiamo di dover affrontare un problema ambientale complesso, per esempio la presenza di Diossine e Furani su una porzione del suolo regionale (per esempio la Valle del Sacco). Il **modo ideale di operare** sarebbe:

- **conoscere con certezza la natura** del problema (penseremmo subito alla deposizione secca ed umida di Diossine e Furani contenuti nell'aria)
- **conoscere con certezza le cause** che l'hanno provocata (*da dove, come, quando e per quanto si sono avute le emissioni*)
- **conoscere con certezza la meteorologia** che ha determinato la deposizione (variazione spazio-temporale del vento, temperatura, turbolenza, ecc.)

A questo punto, basta applicare le leggi fisiche che governano il fenomeno e si ha inevitabilmente il quadro quantitativo della situazione

Alla fine, **posso anche misurare, ma per avere conferma della correttezza delle informazioni di partenza.**

Per esempio, nel caso della **Qualità dell'Aria**, **Arpa Lazio** (nel **C.R.Q.A.**) realizza una **Ricostruzione Near-Realtime** e le **Previsioni a 5 giorni** proprio partendo dalla conoscenza:

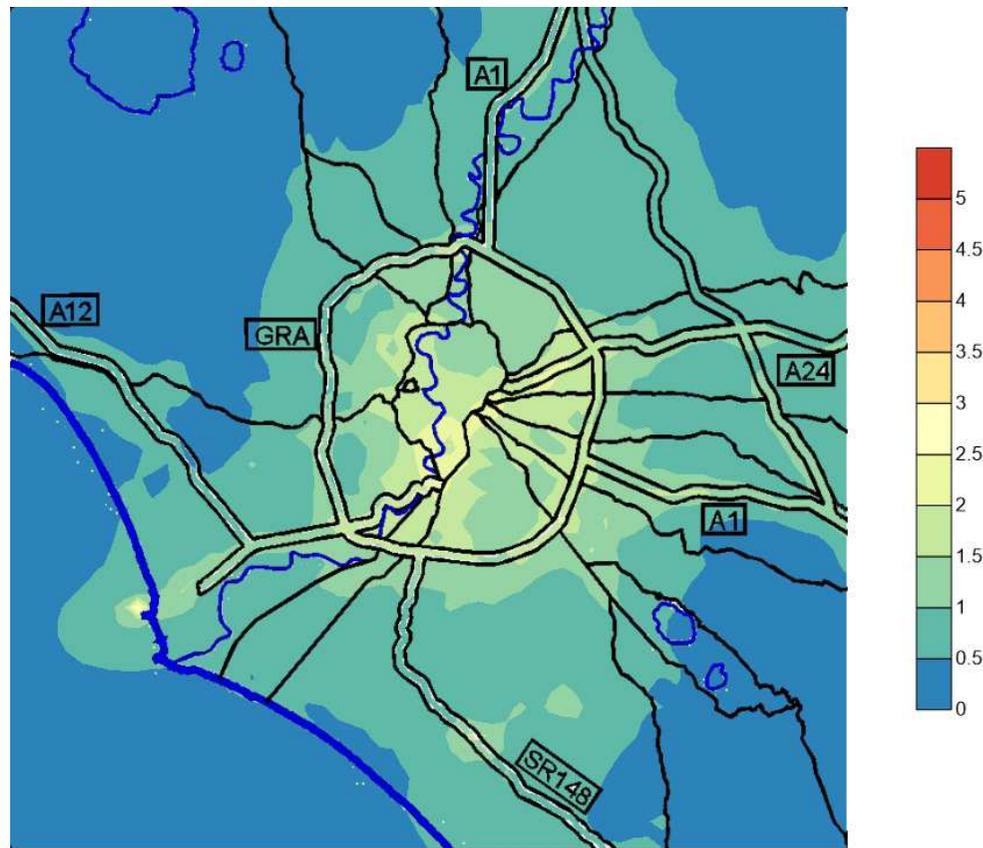
- delle cause meteorologiche e micrometeorologiche che determinano il trasporto, la dispersione, la deposizione e le trasformazioni chimiche degli inquinanti emessi;
- dalle emissioni regionali;
- della situazione a medio-grande scala (meteorologica e chimica).

Le misure ottenute dalle reti di Arpa (rete Q.A., rete micrometeorologica, ecc.) costituiscono effettivamente una verifica della correttezza dell'operato delle catene modellistiche.

In realtà, dato che nessun modello è perfetto e mai si è in grado di fornire ai modelli la conoscenza perfetta delle emissioni e della meteorologia, le misure vengono usate per correggere il modello e per fare ciò si usano tecniche simili a quelle che illustreremo nel corso (**Data Assimilation, Data Fusion**).

Concentrazione media annua del Benzene al suolo

La distribuzione spaziale della concentrazione è ottenuta facendo reagire chimicamente le sostanze emesse in atmosfera (direttamente o indirettamente legate ai processi di formazione del benzene) per ogni ora ed in ogni punto del dominio, e soggette ai meccanismi di trasporto e dispersione (meteorologia).



In una situazione ideale:

Cause → **Modello** → **Effetto**

A volte effettivamente si è in situazioni prossime a quelle ideali (Qualità dell'Aria di cui ormai si riesce ad ottenere una ricostruzione e una previsione affidabile).

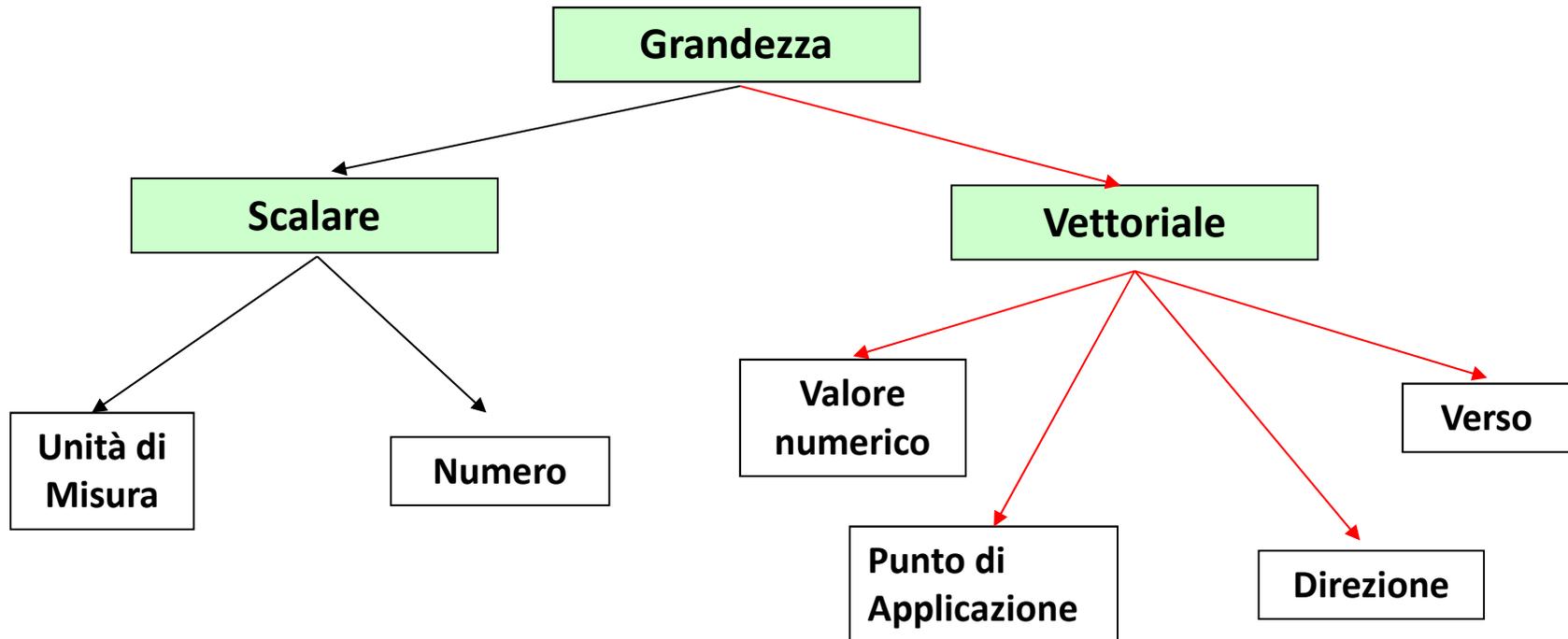
Purtroppo spesso ci si trova in situazioni altamente non ideali, nelle quali si ha la **necessità operativa di conoscere come alcune variabili di interesse ambientale variano almeno nello spazio:**

- Temperature e umidità dell'aria
- Velocità e direzione del vento
- Pressione
- ..
- Concentrazione in aria di specie gassose (CO, NO₂, O₃, NH₃, H₂S, ecc.)
- Concentrazione in aria di aerosol (PM₁, PM_{2.5}, PM₁₀)
- ..
- Deposizione al suolo di specie gassose e particellari
- ..
- Concentrazione di sostanze tossiche nei suoli.

Come si vede, gran parte delle grandezze di interesse ambientale sono

VARIABILI SCALARI

tranne il **vento** che è una **variabile vettoriale**.



In questo corso tratteremo solo della spazializzazione delle variabili scalari.

Riassumendo

la caratterizzazione dell'ambiente atmosferico o dei suoli

richiede la **conoscenza delle variabili che ne descrivono lo stato.**

Ma ...

spesso, l'**unica** cosa che conosciamo è il valore di queste variabili in **alcuni** punti del territorio (generalmente poco numerosi) e poche altre informazioni, spesso incomplete e a volte qualitative.

..... è sufficiente?

Il ricostruire i **Campi Ambientali**, cioè la **distribuzione spaziale** delle variabili di interesse a partire da un numero ridotto di misure è:

un problema intrinsecamente spaziale e 3D

Spazializzazione

Date N misure di una grandezza $Q \rightarrow$ Ottenere $Q(x, y, z)$



Quando è coinvolta l'atmosfera, l'oceano, il rumore, le Onde Elettromagnetiche, la spazializzazione si riferisce ad un certo istante di tempo t^* .

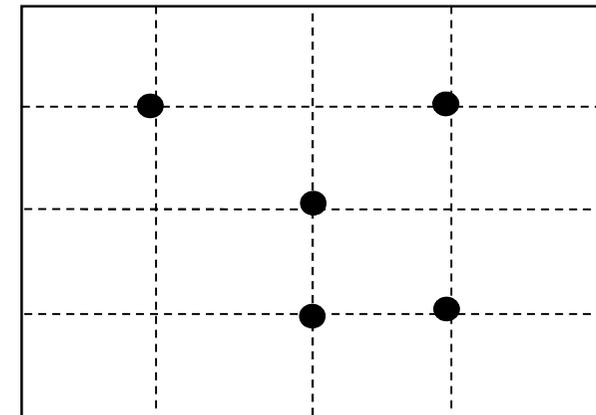
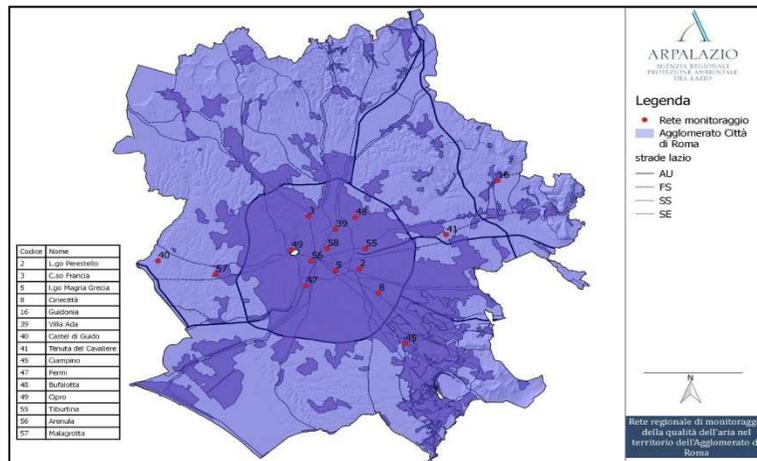
Per i suoli e gli altri comparti ambientali il tempo riveste minore importanza.

Nella *maggior parte* delle applicazioni pratiche l'interesse è limitato ai **campi orizzontali 2D** (come distribuzione della deposizione al suolo, etc....).

Limitandoci, per ora, alla **spazializzazione 2D di grandezze scalari**

Consideriamo misure che possono provenire da:

- Rete fissa di monitoraggio costituita da “*M*” stazioni di misura;
- Insieme di misure strumentali ottenute da una campagna sperimentale in “*M*” punti sul territorio;
- “*M*” prelievi simultanei di campioni di top-soil
-



Tutte le **metodologie di spazializzazione** sono nate intorno agli anni '50 ed i primi lavori scaturiscono da 2 necessità molto *lontane* tra loro:

- Interessi commerciali (fortissimi in quegli anni, soprattutto in Sud Africa) legati a importanti **giacimenti minerari** (oro, argento, uranio, diamanti)
 - difficoltà e onerosità delle prospezioni dirette
 - era necessario limitare il numero di prospezioni allo stretto necessario ed estrarre il maggior numero di informazioni possibile da esse;
- Avvio delle **previsioni meteorologiche** (dopo il tentativo fallito di Richardson del 1922) che richiedevano ad intervalli temporali ravvicinati la ricostruzione in tutto il dominio tridimensionale dei campi meteorologici (pressione, temperatura, umidità, vento) in ampie porzioni del globo in maniera automatica a partire dalle misure nei vari continenti e dai radiosondaggi.

Le linee di sviluppo sono state completamente indipendenti:

- **Linea mineraria**: Danie Krige ha introdotto le basi della famiglia dei metodi di Kriging, tutti basati su solide deduzioni Statistiche;
- **Linea meteorologica**: ha avuto uno sviluppo graduale:
 1. **Metodi interpolativi** di vario genere, esclusivamente di tipo geometrico, basati sull'uso di funzioni polinomiali, spline cubiche, sviluppi in Serie di Fourier, etc ... trascurando le proprietà statistiche dei campi e delle misure;
 2. **Metodi semiempirici basati su modelli lineari** che mettono in relazione *misure e campi* tenendo conto della statistica delle misure;
 3. **Metodi Statistici BLUE ottimali (Optimal Interpolation)**
 4. **Metodi variazionali (3D-var, 4D-var, filtro di Kalman)**

Salvo ad un breve accenno ad un metodo interpolativo del tipo 1), concentre-remo l'attenzione esclusivamente ai metodi del tipo 2) e 3).

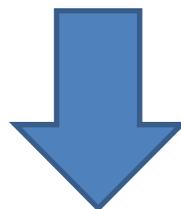
I metodi Variazionali (di tipo 4), pur essendo oggi i più promettenti ed usati, sono caratterizzati da una complessità matematica e statistica talmente elevata da impedirne la presentazione in un corso come questo.

Utilizzo in ARPA Lazio

- **Campagne di analisi top-soil per il rilevamento di Diossine, Furani, PCB ed altre sostanze tossiche;**
- **Campagne di qualità dell'aria ad hoc multi-punto;**
- **Spazializzazione delle misura della rete di monitoraggio (meteo, Q.A.)**
- **Predisposizione delle condizioni iniziali ed al contorno del modello meteorologico e di dispersione;**
- **Ricostruzione situazioni incidentali usando misure con o senza il supporto di un modello;**
- **Distribuzione spaziale di radionuclide.**

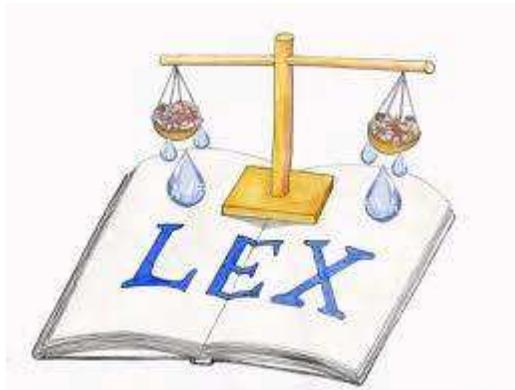
Motivazione

E' necessario conoscere esattamente **come operano questi metodi**, **come adeguarli al fenomeno in esame**, **come applicarli in maniera corretta** e non come semplici strumenti per ottenere belle figure



Per essere in grado di sostenere, con motivazioni convincenti, la spazializzazione realizzata anche in sedi *difficili* come quella giudiziaria.

La spazializzazione non è un problema grafico ma un problema modellistico anche se molto particolare



Per la Direttiva 2008/50/CE (recepita dal D. Lgs. 155/2010) le tecniche statistiche di spazializzazione sono identificate genericamente come:

Tecniche di Stima Obiettiva

Conviene chiarire subito un **Problema Semantico**

Interpolazione: con questo termine si intende un'operazione puramente matematica che, dati alcuni punti in cui è nota una funzione, determina il valore di questa funzione in un punto x interno al dominio spaziale entro cui si trovano i punti dati. Molti sono i metodi di questo tipo forniti dall'Analisi Numerica (polinomi interpolanti, spline bi-cubiche, polinomi trigonometrici di Fourier). Essi possono essere visti, nel contesto di questo corso, come strumenti utilizzabili o per realizzare spazializzazioni di prima approssimazione e per rappresentare elaborati grafici.

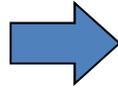
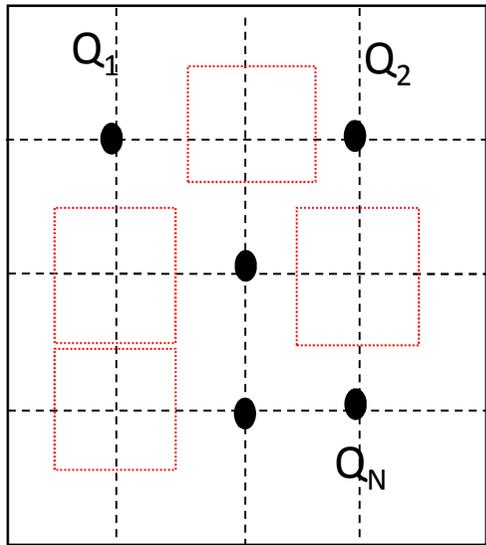
Spazializzazione: con questo termine intenderemo tutti i metodi presentati in questo corso che, partendo dalla conoscenza del valore di una grandezza fisica in un numero finito di punti del dominio spaziale di interesse e da un insieme, più o meno eterogeneo di informazioni a priori relative alla natura della variabile fisica d'interesse, alla statistica connessa al procedimento di misura e alla struttura del dominio di indagine stesso, stima il miglior valore possibile della variabile in punti differenti del dominio.

2. FORMULAZIONE GENERALE DEL PROBLEMA

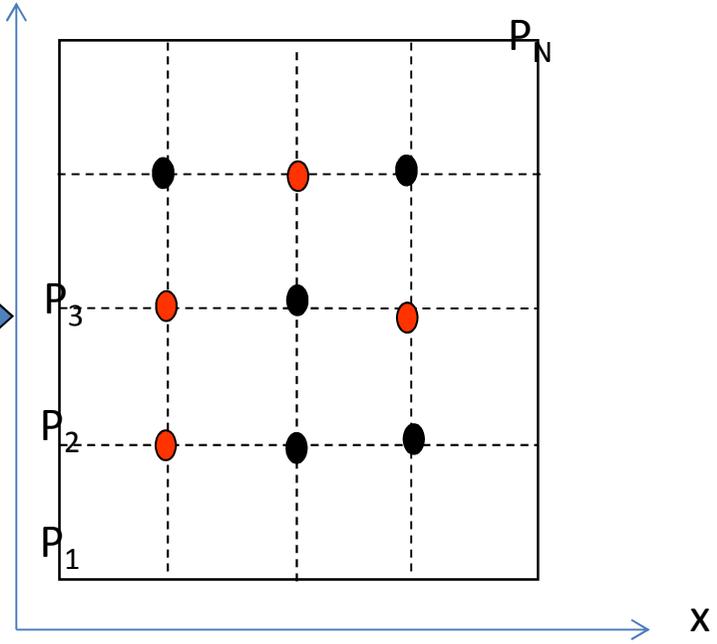
Si consideri una generica grandezza Q da spazializzare (per es. la temperatura dell'aria o la concentrazione in aria di un inquinante). Qualunque sia la metodologia di spazializzazione di un insieme discreto M di misure, **il punto di partenza e gli elementi del problema si riducono sempre ai seguenti:**

- “ M ” **Misure** Q_m ($m = 1, .. ,M$) distribuite irregolarmente nella spazio nelle posizioni $P_m(x_m, y_m)$;
- un **Campo di Background** $Q_b(x, y)$ noto virtualmente in ogni punto $P(x, y)$ di una porzione di territorio cui siamo interessati. Tale campo di background **rappresenta la prima approssimazione grossolana di come dovrebbe essere distribuita spazialmente la grandezza Q .**

Ogni metodologia di spazializzazione cerca di ottenere la stima più realistica possibile della variabile Q in un insieme di N punti dello spazio $P_i(i = 1, .. ,N)$ dove le misure non sono disponibili.



y



2.1 LE MISURE

Le **misure** costituiscono *la prima e più importante base conoscitiva e, nei casi più comuni, possono derivare:*

- da strumenti automatici come i sensori meteorologici, gli analizzatori di specie chimiche gassose e particellari (gli analizzatori automatici delle polveri a nastro o a filtro con misura di assorbimento Beta),
- da pesate (campionatori di polveri),
- da analisi chimiche di campioni ottenuti da prelievi in aria, dal suolo o dalle acque (mare, laghi, ecc.).

Attenzione: La Fisica (in particolare la Teoria della Misura) ricorda che quando si realizza una misura Q_m di una grandezza fisica Q (es. la concentrazione in aria di Diossine), di fatto è come se si **interrogasse** una **Variabile Stocastica Q avente** le seguenti caratteristiche:

- **non è mai possibile** accedere al suo Valore Vero,
- è solo possibile interrogarla col processo di misura, ottenendo una realizzazione a cui è associata una probabilità.

Quindi, il **PROCEDIMENTO DI MISURA** è un **PROCESSO STOCASTICO** che:

1. **Interroga** una variabile ambientale (che si comporta come fosse una variabile stocastica) in un punto dello spazio $P(x,y)$ e ad un istante (t). Il **Valore vero** posseduto da questa grandezza è totalmente inaccessibile;
2. dal **processo di misura** (sensore → acquirettore → convertitore A/D; campionamento → condizionamento campione → analisi chimica) ottengo alla fine **una o più realizzazioni della grandezza Q** che forniscono il valore della variabile con associata una certa probabilità, dipendente dalla legge statistica (funzione di Densità di Probabilità) cui ubbidisce la grandezza.

Quindi, ogni volta che si ottiene una misura di Q si pone in essere un procedimento di misura con cui cerchiamo di catturare questa grandezza in quel punto dello spazio e in quell'istante con un apparato (realizzazione) e da questa realizzazione cerchiamo di ottenere un numero (la misura in senso stretto).

In questo processo si infila del rumore vario che fa sì che l'ottenimento di diverse realizzazioni (misure) di Q sia assimilabile in pratica al campionamento discreto di una variabile stocastica.

Quindi se Q_m è una **misura** della grandezza Q ottenuta strumentalmente ed indichiamo proprio con Q il valore vero (inaccessibile) che tale misura ha in quel punto e in quell'istante, allora dalla Teoria degli Errori di Misura risulta che:

$$Q_m = Q + \varepsilon_m$$

dove ε_m è l'**errore di misura**.

La Teoria degli Errori dimostra che:

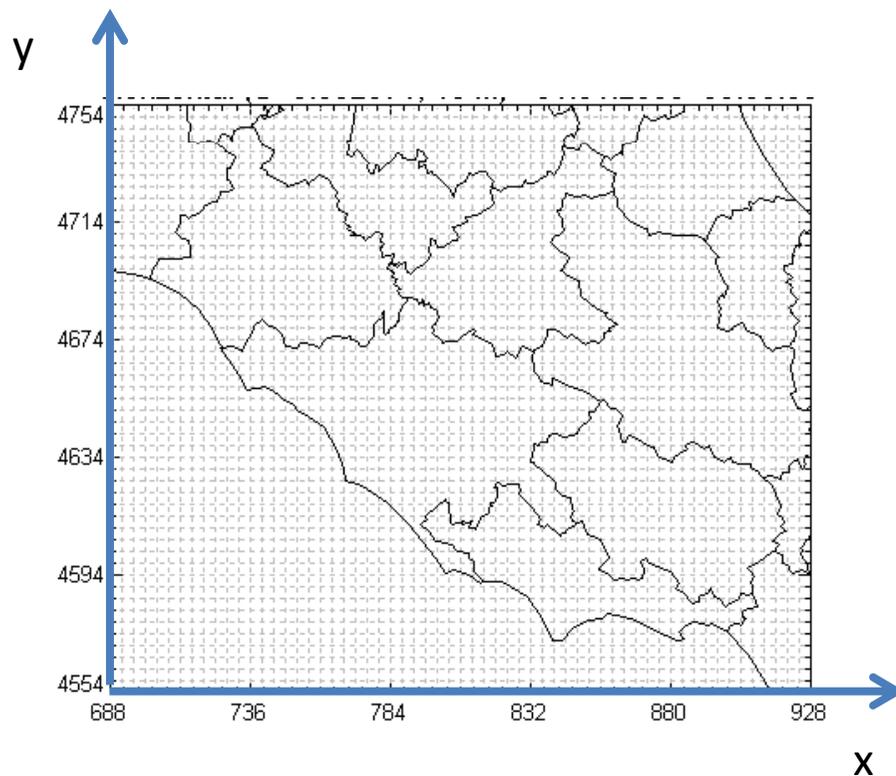
- ε_m è distribuito secondo una **Densità di Probabilità Gaussiana** (distribuzione completamente definita da due parametri: valor medio e varianza);
- Il **valore medio dell'errore** di misura è **nullo**;
- L'errore di misura possiede una data varianza σ^2_0 che dipende dalla struttura del sistema di misura (o da come viene condotta l'analisi chimica), dalle condizioni di acquisizione/campionamento e dalle caratteristiche del luogo di misura (interferenze locali). E' consuetudine individuare questa varianza come **incertezza intrinseca** (Norma UNI CEI ENV 13005 – 2000)

NOTA

1. L'incertezza intrinseca è propria di ogni singola misura e quindi, almeno in teoria, ogni misura avrebbe una propria incertezza e quindi avremmo tante incertezze quante sono le misure considerate.
2. Spesso però esse derivano da una rete di monitoraggio (per esempio Q.A. o meteo) dotata di postazioni che usano lo stesso strumento di misura per rilevare la grandezza Q. In questo caso si può asserire, almeno in prima approssimazione, che tutte le misure sono affette dalla stessa incertezza di misura strumentale σ^2_0 . In prima approssimazione perché le interferenze locali potrebbero essere diverse per le diverse misure.
3. Se le misure derivano invece da campionamenti e da successive analisi di laboratorio, l'incertezza intrinseca delle stesse potrebbe essere posta pari all'incertezza intrinseca dell'analisi effettuata, trascurando le inevitabili differenze di campionamento e le interferenze locali.

2.2 GEOMETRIA DEL DOMINIO DI ANALISI

Sia il **Campo di Background**, che costituisce tutta l'informazione a priori (non necessariamente corroborata da misure o da modelli fisici) che disponiamo della distribuzione spaziale $Q(x, y)$, sia l'**Analisi**, cioè il campo spaziale della variabile Q ottenuto dopo la spazializzazione delle misure disponibili, dovrebbero coprire con continuità l'intero territorio oggetto dello studio.



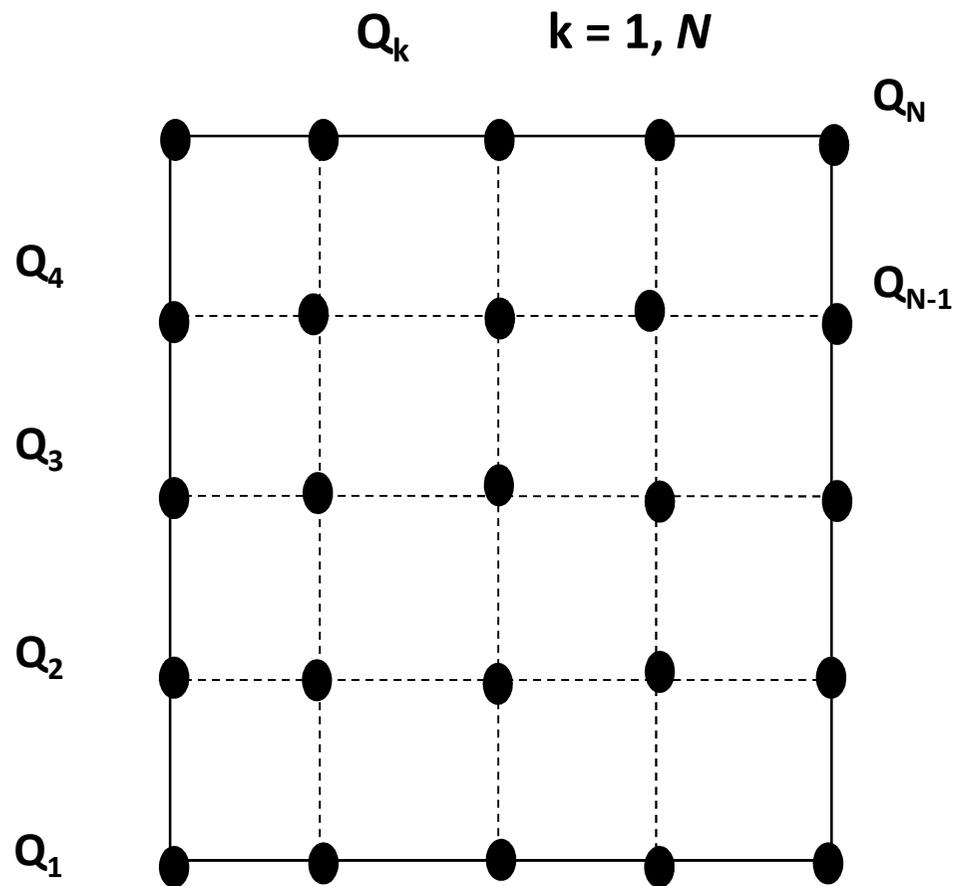
Virtualmente saremmo quindi interessati al valore più probabile della variabile Q in ogni punto $P(x, y)$ entro il dominio, ma, non potendo realizzare infinite stime, consideriamo solo un numero limitato di punti (N) opportunamente distribuiti nello spazio.

- Normalmente nelle applicazioni si considerano **porzioni rettangolari di territorio** con lati orientati in direzione W-E e N-S, di dimensioni rispettivamente L_x e L_y .
- Nello spigolo SW (in basso a sinistra) si colloca un **sistema di riferimento Cartesiano ortogonale (x,y)** con asse x in direzione E e asse y in direzione N;
- Il dominio si ripartisce in **celle elementari** di lato Δx e Δy , sovrapponendo ad esso una **griglia regolare** che ha n_x nodi lungo x e n_y nodi lungo y . Questo dominio discreto, che ci accompagnerà in tutte le metodologie di spazializzazione, prende il nome di **Dominio di Analisi**;
- Ci **limitiamo** quindi a ricostruire il **campo $Q(x,y)$ in un numero discreto di punti** coincidenti con i nodi $P_i(x_i, y_i)$ definiti come:

$$\begin{aligned}x_i &= (i-1) \cdot \Delta x & i &= 1, \dots, n_x \\y_i &= (j-1) \cdot \Delta y & j &= 1, \dots, n_y\end{aligned}$$

- Il campo discreto $Q(x_i, y_i)$ sarà una tabella bidimensionale contenente in totale $N = n_x n_y$ valori corrispondenti ad ogni nodo di griglia.

- ✓ Per ragioni matematiche, negli algoritmi che seguono questi valori verranno riorganizzati in un vettore di N valori, $[Q]$ con elementi:



2.3 LA CONOSCENZA A PRIORI ED IL CAMPO DI BACKGROUND

Quale può essere la conoscenza, anche a livello di prima approssimazione, che possediamo *a priori* del campo Q ?

La risposta dipende dal tipo di problema ambientale che viene trattato. Consideriamo questi due casi tipici.

- a) La variabile Q è una *variabile meteorologica o di Q.A.* (o una variabile oceanografica) e disponiamo di una sua ricostruzione, **diretta o indiretta, realizzata, magari in maniera approssimata, con un modello matematico di previsione/ricostruzione, che spesso usa una griglia di calcolo a risoluzione diversa (diversi Δx e Δy rispetto alla griglia di Analisi):**
- interpolo con uno dei metodi dell'Analisi Numerica bi-dimensionale dai valori prodotti dal modello sulla propria griglia alla griglia di Analisi voluta,
 - quanto ottenuto è il campo di background che effettivamente tiene conto di tutte le conoscenze *a priori* possedute ed utilizzate dal modello per la ricostruzione.

b) La variabile Q è una variabile meteorologica o di Q.A (o una variabile oceanografica), ma non dispongo di informazioni modellistiche a priori; dispongo solo delle misure prodotte da un insieme di sensori posti irregolarmente nel territorio e spesso organizzati in una rete.

→ In questo caso il Campo di Background lo ottengo interpolando numericamente le misure disponibili con uno dei metodi dell'Analisi Numerica.

NOTA

Sia nel caso *a)* che nel caso *b)* **ipotizziamo, implicitamente o esplicitamente**, che **le informazioni disponibili tengano già conto delle cause che presumibilmente hanno determinato la variazione nello spazio della grandezza**, cioè:

→ nel caso in cui si disponga di modelli: le emissioni (luogo di emissione e tassi di emissione) ed i campi meteorologici specifici;

→ i criteri con cui sono stati localizzati i punti di monitoraggio che dovrebbero aver già tenuto conto delle sorgenti e della meteorologia media.

SPESSE, PERÒ, CAPITANO SITUAZIONI TOTALMENTE DIVERSE

ESEMPIO 1

È stata individuata la presenza di sostanze tossiche sul terreno, ma non è noto alcun incidente che ne potrebbe essere la causa.

1. Cerco sul territorio potenziali sorgenti,
2. Determino il movimento medio delle masse d'aria e le caratteristiche meteorologiche tipiche,
3. Individuo un certo numero di siti che potrebbero essere influenzati da questa possibile emissione
4. Individuo un certo numero di siti che non dovrebbero essere influenzati
5. Eseguo campionamenti in tutti questi siti. A questo punto tutte le misure raccolte dovrebbero contenere tutte le informazioni a priori umanamente ottenibili

→ A questo punto il Campo di Background può essere l'interpolazione sui nodi del dominio di Analisi di tutte le misure ottenute dalla campagna.

ESEMPIO 2

È avvenuto un incidente (esplosione, incendio, emissione accidentale di sostanza tossica) quindi ho qualche informazione a corredo delle ricerche della concentrazione (o deposizione) della grandezza Q (luogo e momento dell'incidente, natura dell'incidente)

a) Se posso, eseguo/faccio eseguire, una *simulazione modellistica preliminare* anche se le informazioni emmissive sono solo ipotetiche, mentre quelle meteo sono dettagliate.

→ Il campo di background potrebbe essere proprio la ricostruzione modellistica nei nodi del dominio di analisi.

b) Avendo a disposizione questa informazione a priori, posso individuare due insiemi di punti di campionamento:

→ quelli influenzati con elevata probabilità

→ quelli non influenzati con elevata probabilità

consentendo di realizzare un piano di campionamento equilibrato.

Ovviamente in pratica sono possibili numerose situazioni differenti che si dovranno affrontare mettendo in campo ogni conoscenza geografica, impiantistica, fisica e chimica possibile.

La **differenza** tra i due casi sta nella **quantità e qualità di informazioni** disponibili sull'evento e quindi sulla **realisticità quantitativa e qualitativa di una eventuale ricostruzione modellistica**.

Spesso, a fronte di informazioni a priori scarsissime (magari limitate solo a quando e dove si è verificato l'incidente e quale attività industriale ha coinvolto), si rinuncia alla modellizzazione preliminare.

Errore!

Una modellizzazione in atmosfera può **sempre** contare su una conoscenza meteorologica e micrometeorologica buona sul territorio regionale. Quindi, una **modellizzazione con emissioni non molto realistiche** o anche totalmente irrealistiche (per esempio unitarie) **conduce ad una ricostruzione che comunque per ragioni fisiche ha qualitativamente caratteristiche molto simili a quelle reali**

.... e ciò consente di individuare, come detto, i punti di misura essendo ragionevolmente sicuri di non trascurare zone importanti.

Si ricordi che se le misure vengono condotte in siti non appropriati, si rischia di perdere, senza possibilità di recupero, informazioni preziose, cosa che porta poi a spazializzazioni irrealistiche indipendentemente dal modello utilizzato (più o meno complesso).

Prendiamo in considerazione un **esempio verosimile** in cui possediamo l'informazione preziosa secondo cui:

1. Era noto il luogo dell'incidente
2. L'incidente (o comunque l'evento) è avvenuto quando il movimento medio delle masse d'aria avveniva prevalentemente nella direzione NW-SE.

Una simulazione modellistica realizzata in maniera estremamente semplificata (condizioni di turbolenza adiabatica, emissioni unitarie al suolo) conduce all'individuazione di un'Area di Influenza Probabile dell'evento individuabile in **quella porzione del territorio in cui la concentrazione al suolo risulta essere non inferiore ad una data frazione del suo valore massimo (es. 10%)**

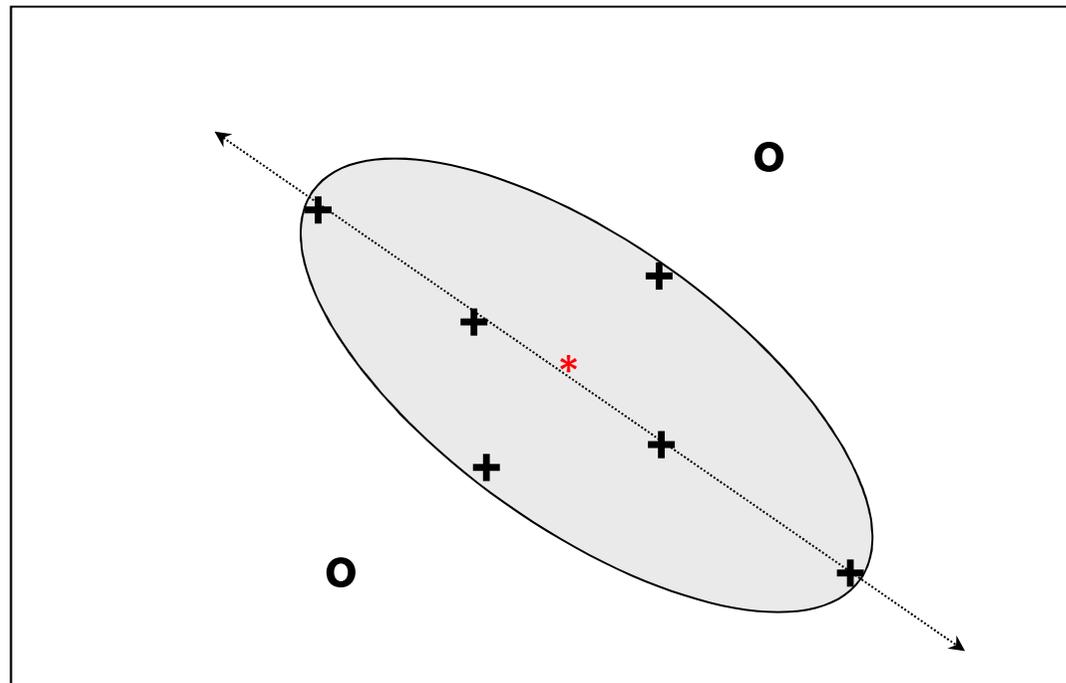
Nelle figure seguenti, questa area di influenza è rappresentata semplicemente come un'ellisse.

Caso in cui si è tenuto conto dell'informazione a priori ottenuta dal modello e, sulla base di questa conoscenza, sono stati individuati i possibili punti di misura della campagna sperimentale, che verranno spazializzati a fine di ottenere la rappresentazione spaziale del fenomeno.

*** Posizione possibile sorgente della Nota**

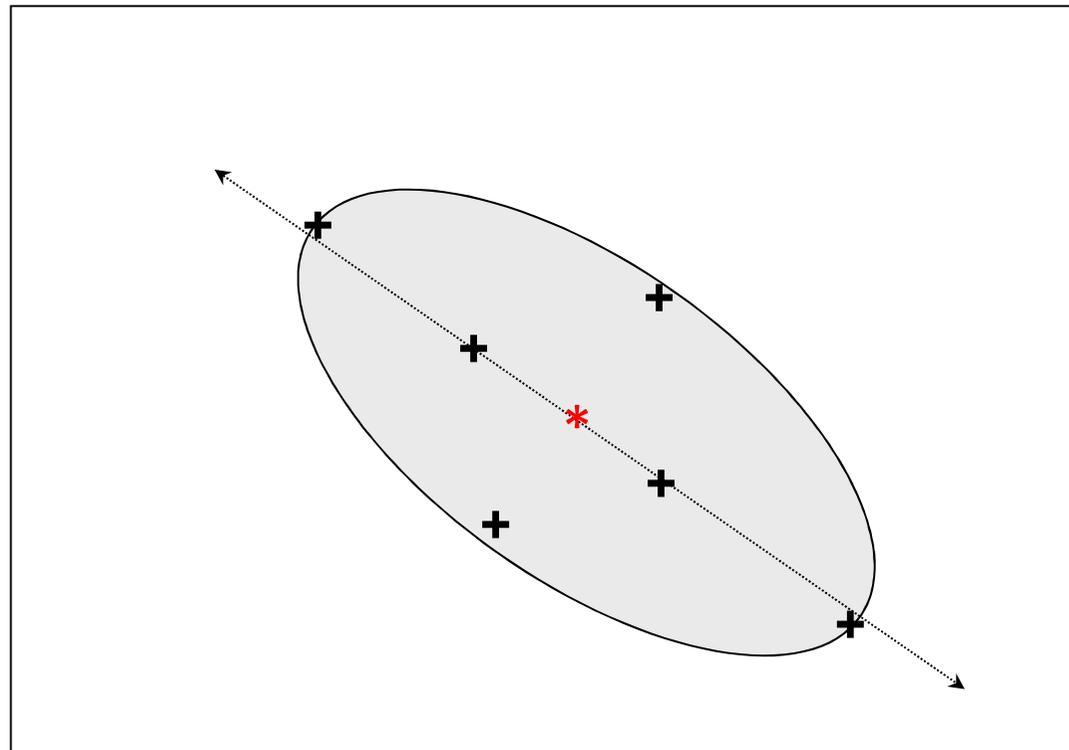
+ Possibili punti di misura probabilmente influenzati

O Possibili punti di misura probabilmente non influenzati



In questo caso, invece, i punti di misura sono stati individuati in maniera scorretta, visto che mancano misure che possano indicare il valore di fondo della variabile Q.

*** Posizione possibile sorgente della Nota**
+ Possibili punti di misura



ESEMPIO 3

È segnalata la presenza della variabile Q nel territorio, per esempio emersa a fronte di controlli casuali o programmati, ma non c'è altra informazione disponibile.

→ Non posso fare altro che distribuire le misure in maniera spazialmente omogenea ed ottenerle strumentalmente

In questo caso il campo di background più ragionevole è la media delle misure ottenute.

In questo caso non ci sono elementi neppure di natura qualitativa per poter stabilire la localizzazione dei diversi punti di misura.



Presumibilmente conviene:

- Realizzare poche misure distribuite uniformemente sul territorio e spazializzarle, individuando la distribuzione di prima approssimazione della variabile;
- Infittire le misure là dove sembrerebbero esserci zone non influenzate dal fenomeno, là dove sembrano esserci i maggiori gradienti e là dove sembrerebbero risiedere i massimi;
- Iterare la procedura

Come si nota, per il **successo della spazializzazione** è necessaria una **progettazione logica e sinergica**:

1. **Del progetto di monitoraggio** (dove, quando, per quanto e con che strumenti fare le misure). Se le misure vengono realizzate in punti idonei (punti che evidenziano il fondo, punti che evidenziano i gradienti, punti che evidenziano i massimi, sulla base delle informazioni modellistiche a priori (Campo di Background), allora la spazializzazione successiva darà i risultati desiderati ed i più realistici possibili. *In caso contrario, non solo le misure non contribuiranno alla conoscenza complessiva del fenomeno, ma addirittura potrebbero distorcere le informazioni plausibili contenute nel Campo di Background.*
2. **Della metodologia per la predisposizione del campo di background**, cioè della conoscenza preliminare (a priori) del fenomeno, ancorché grossolana e/o qualitativa). In sostanza, l'assenza di un Campo di Background non contribuisce alla predisposizione di un'adeguata localizzazione dei punti di misura e l'unica strategia possibile è quella di realizzare campagne di misura iterate nel tempo.

È **fondamentale**, nella realizzazione delle misure e nella predisposizione del campo di background, **tenere conto della Scala Temporale T** del fenomeno indagato. **In sostanza, per scala temporale si intende l'intervallo di tempo in cui l'evento varia significativamente:**

- **Variabile meteo e concentrazioni in aria:** $T \in \text{ora} \div \text{giorno}$
- **Deposizione al suolo:** vista la lentezza della deposizione secca, T può essere dell'ordine della settimana o mese, che si riduce però al giorno se intervengono fenomeni piovosi che innescano la deposizione umida
- **Contaminazione dei suoli:** T può essere ritenuto tendente all'infinito

Bisogna **tener conto del tempo di scala del fenomeno perché:**

- **tutte le misure**
- **tutte le informazioni che concorrono alla costruzione del campo di back**

devono essere contemporanee, cioè tra esse non deve trascorre un intervallo di tempo superiore al Tempo di Scala T.

In caso contrario si considererebbero misure afferenti a fenomeni virtualmente differenti.

ERRORI DI BACKGROUND

Associato al **Campo di Background** c'è sempre un **Errore di Background** (più o meno grande, a volte anche grandissimo), anch'esso **Variabile Stocastica** caratterizzata da una propria ed appropriata statistica.

Consideriamo un generico *Nodo di griglia k*:

- a) Valore vero della variabile $\rightarrow Q_k$
- b) Valore del Campo di Background $\rightarrow Q_{b,k}$
- c) Errore del Campo di Background $\rightarrow \varepsilon_{b,k} = Q_k - Q_{b,k}$

L'errore è distribuito secondo una **densità di probabilità non necessariamente Gaussiana**. Comunque per semplicità si assume che:

- Il valore medio di $\varepsilon_{b,k}$ è = 0
- L'intero comportamento statistico dell'errore di background è rappresentato dalla varianza degli errori $\sigma_{b,k}^2$

A priori la varianza degli errori di background potrebbe variare da nodo a nodo.

Possibilità pratiche per il calcolo del Campo di Background

Per quanto detto fin qui, consideriamo due casi:

1. **il campo di background deriva da un modello impiegato con meteorologia ed emissioni realistiche.** In questo caso è possibile che sia noto l'errore associato alle ricostruzioni modellistiche e quindi può risultare possibile una sua stima diretta nei diversi nodi della griglia di analisi.
2. **Sempre in questo caso, ma senza informazioni sull'incertezza modellistica, una strategia potrebbe essere la seguente:**
 - $Q_{m,k}$ valore interpolato con metodi di Analisi Numerica nel nodo di griglia k a partire dalle misure circostanti,
 - Q_k valore prodotto dal modello

L'errore intrinseco del campo di background potrebbe essere ritenuto spazialmente omogeneo e pari a:

$$\sigma_b^2 = \frac{1}{N} \sum_{j=1}^N (Q_{m,k} - Q_k)^2$$

3. Se il campo di background non deriva da considerazioni modellistiche, ma lo si deduce dalle sole misure realizzate e disponibili, allora:

→ Valore Medio (i = postazione di misura)

$$\bar{Q}_m = \frac{1}{M} \sum_{i=1}^M Q_{m,i}$$

→ **Errore di Background** potrebbe essere pari alla varianza delle misure disponibili,

$$\sigma_b^2 = \frac{1}{M} \sum_{i=1}^M (Q_{m,i} - \bar{Q}_m)^2$$

anche se una scelta migliore sembrerebbe essere il **quadrato dell'errore standard della media**:

$$\sigma_b^2 = (\varepsilon_b)^2 = \frac{1}{M} \sum_{i=1}^M (Q_{m,i} - \bar{Q}_m)^2$$

Come si vede, la **determinazione del campo di background** (che a volte, ed erroneamente si evita di fare, ponendolo identicamente nullo):

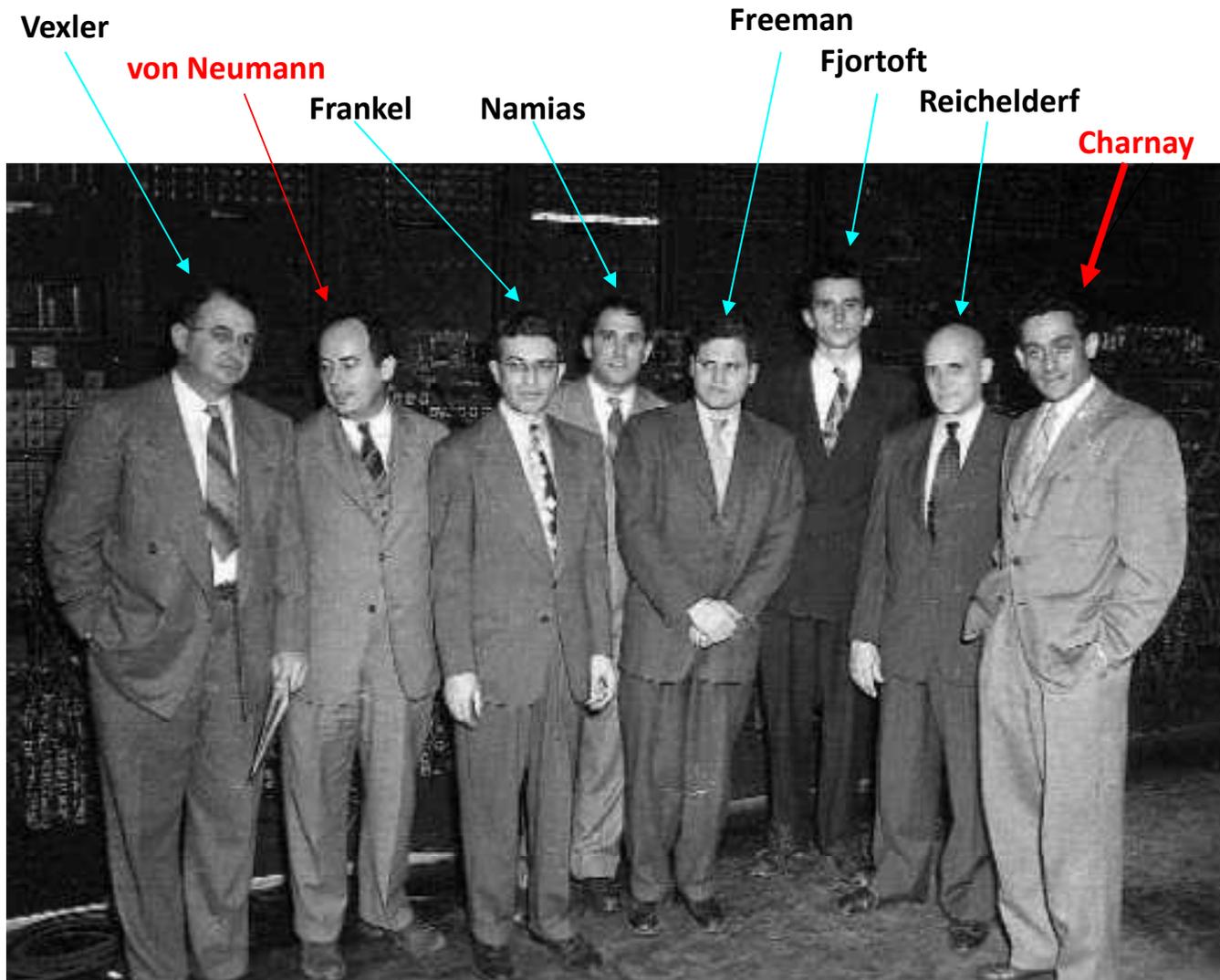
1. **Costringe a riflettere attentamente sulla natura fisica e chimica del problema,**
2. **Costringe a raccogliere, organizzare ed interiorizzare tutte le informazioni disponibili a priori**
3. **Permette di avere validi elementi per valutare a posteriori il grado di attendibilità della spazializzazione.**

3. METODI INTERPOLATIVI FUNZIONALI

E' questa una **famiglia eterogenea** di metodi che interpolano analiticamente i valori Q_k attribuiti ad una generica grandezza fisica Q in K punti $P_k(x_k, y_k)$ mediante misure dirette della stessa, in modo da ottenere per via analitica la stima della grandezza Q in un punto di misura $P(x, y)$ in cui la grandezza fisica non è stata misurata. Questi metodi sono materia descritta nei normali corsi di Analisi Numerica. Di essi fanno parte, per esempio, le interpolazioni polinomiali di vario tipo, le spline bi-cubiche e le serie di Fourier.

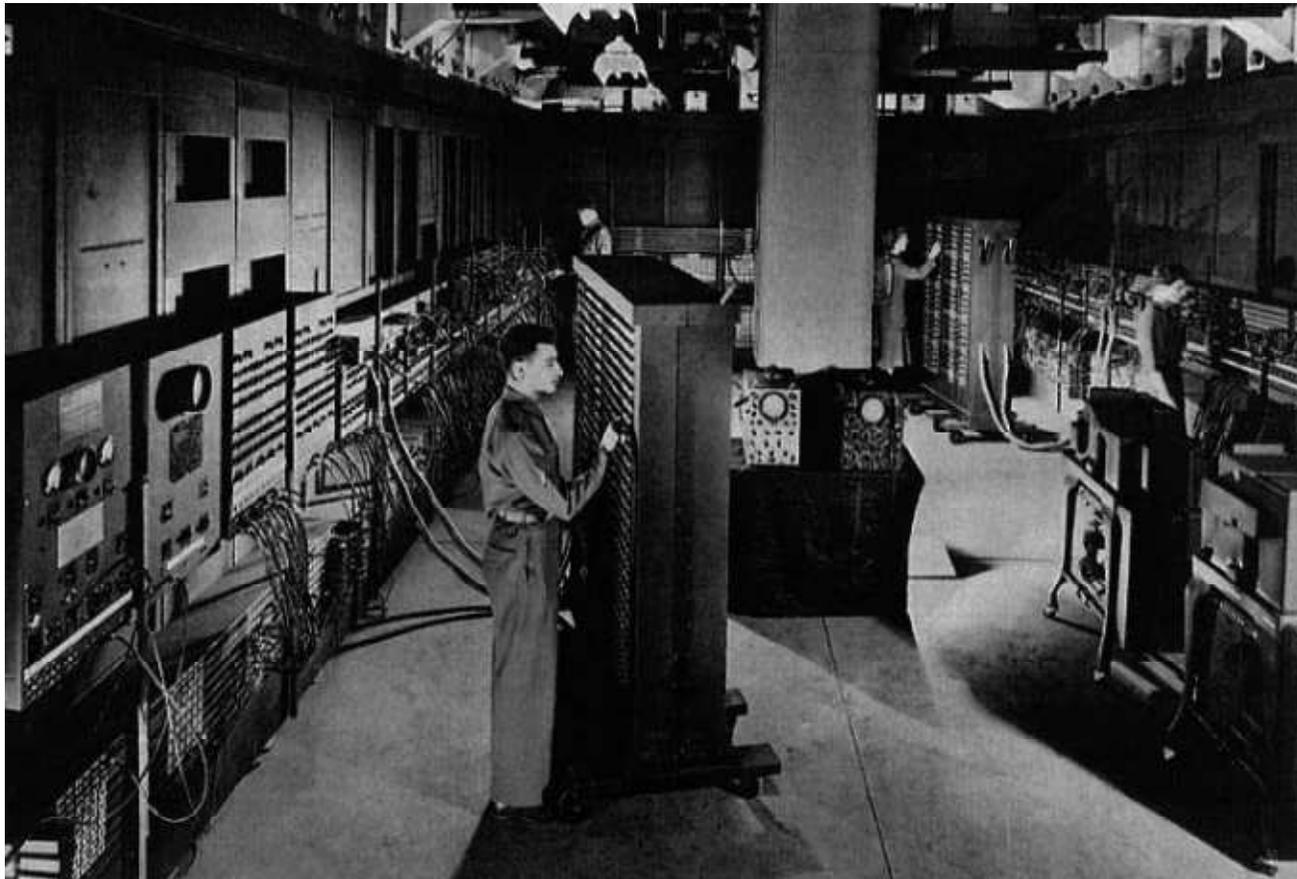
A parte la lunga storia matematica legata al problema dell'interpolazione monodimensionale e bidimensionale, queste tecniche sono state utilizzate ampiamente dal 1950 in poi per la **preparazione automatica** delle condizioni iniziali dei modelli meteorologici prognostici (**Analisi Meteorologiche**).

Obiettivo finale: ricostruire automaticamente (cioè senza apprezzabile intervento umano) i campi al suolo ed in quota delle principali variabili meteorologiche (temperatura, umidità, componenti cartesiane del moto, pressione, geopotenziale, ecc.) a partire dalle misure rilevate sull'intero globo.



4 Aprile 1950: l'equipe di ricercatori che ha realizzato la prima previsione meteorologica numerica usando come strumento di calcolo

... il Supercomputer ENIAC (*Electronic Numerical Integrator and Calculator*) dei laboratori Militari di Aberdeen (Maryland), creatura di von Neumann durante il progetto Manhattan.



Alcuni dati interessanti:

30 tonnellate di peso,

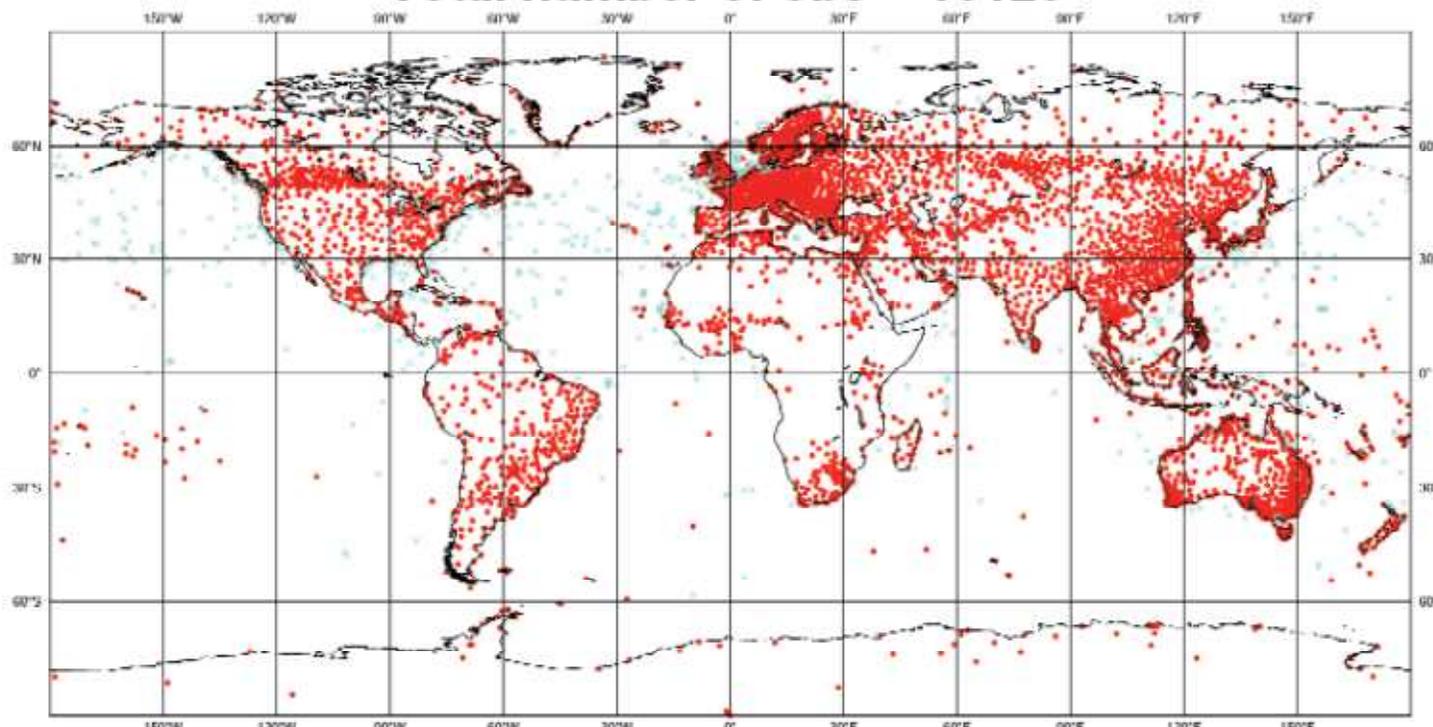
180 m² di superficie occupata,

140 kW di potenza elettrica richiesta

Alimentato a schede perforate

ECMWF Data Coverage (All obs) - SYNOP/SHIP
16/DEC/2003; 00 UTC
Total number of obs = 15123

Obs Type	
•	13672 SYNOP
•	1451 SHIP



Disponibilità di misure: meteorologiche per le previsioni meteorologiche (2006)

15.000 stazioni meteo sinottiche terrestri

37.500 osservazioni aeree

600 radiosondaggi

3.300 boe strumentate

300.000 osservazioni termiche satellitari

Dopo questo inizio da pionieri, fino al 1997 le sorti della previsione meteorologica numerica vanno ad una *ragazzina* di Buenos Aires, **Eugenia Kalnay**, che ha conseguito poco più che ventenne il PhD presso il MIT con Charney (primo PhD assoluto del MIT ad una donna e per giunta straniera).

E' lei che per oltre 10 anni dirige con energia e competenza Environmental Modeling Center del NCEP (National Centers for Environmental Prediction), spingendo la ricerca nella realizzazione delle Analisi e nella Assimilazione delle misure nei modelli alle vette vertiginose raggiunte oggi, e questo poderoso lavoro di ricerca ed organizzativo l'ha iniziato da ventenne.



A quei tempi, per poter ottenere la **distribuzione spaziale** delle **principali variabili meteorologiche** ad ore fisse (**ore sinottiche 00, 06, 12, 18, 21 GMT**), cioè i campi iniziali o le **Analisi**, venivano utilizzati questi metodi interpolativi di tipo funzionale.

Questi metodi sono puramente matematici/geometrici.

Punto di partenza:

1. Scelta di una famiglia di funzioni interpolanti

- Rette
- Polinomi di vario grado
- Funzioni trigonometriche (Fourier)
- ecc.

2. le misure Q_k della variabile Q in K punti P_k distribuiti irregolarmente nello spazio

Punto di arrivo → stima puramente analitica del valore di Q in un punto generico P (interno alla porzione di spazio contenente i punti sedi di misura) basata esclusivamente:

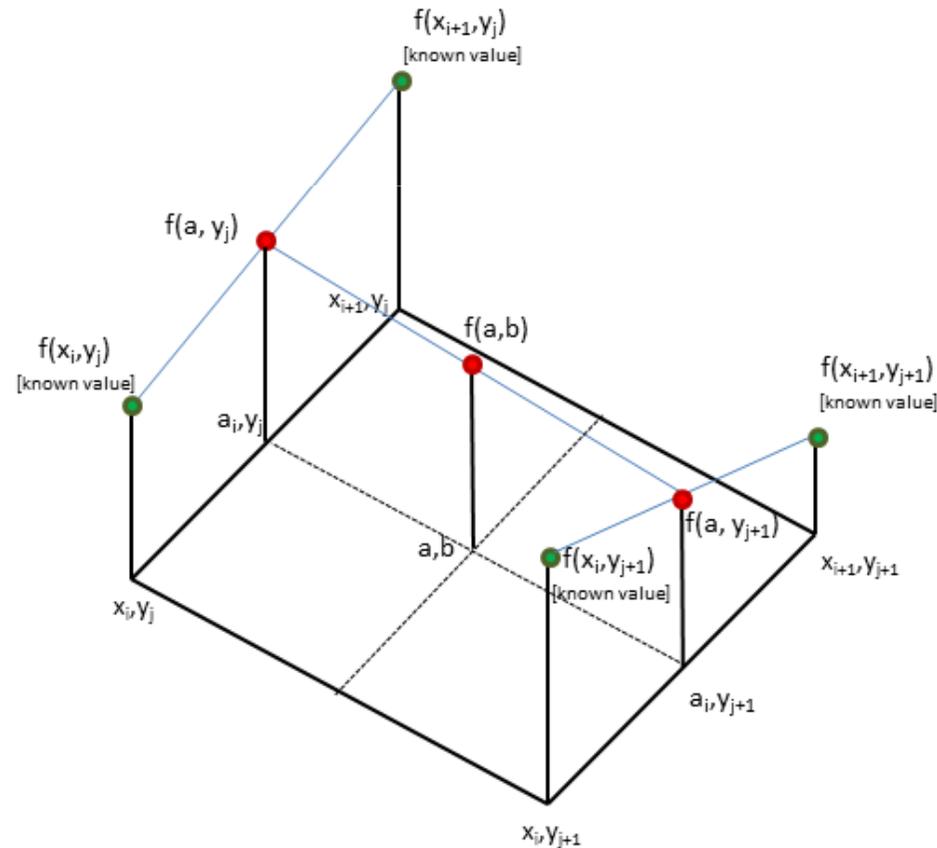
- sulla posizione dei punti di misura
- Sulla distanza reciproca tra punti di misura e punto di interpolazione.

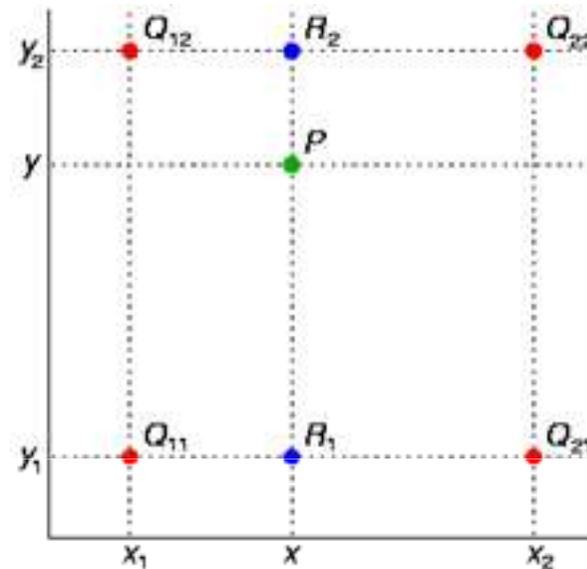
Dopo una decina di anni dalla loro introduzione operativa, questi metodi puramente geometrici sono stati abbandonati dai meteorologi nell'Analisi dei campi sinottici a favore di metodi basati non solo sulle informazioni geometriche del problema ma anche sulle informazioni statistiche relative alle misure ed ai fenomeni fisici stessi (metodi che vedremo tra pochissimo)

Di fatto sopravvive il metodo di interpolazione Bi-Lineare che, come vedremo tra poco, costituisce uno strumento matematico di notevole utilità per effettuare le spazializzazioni di primo tentativo richiesta da metodi più complessi e statisticamente più rigorosi, quelli che ora vengono effettivamente impiegati nella previsione meteorologiche e nelle previsioni della qualità dell'aria.

IL METODO DI INTERPOLAZIONE BI-LINEARE

Immaginiamo di conoscere una generica grandezza fisica f in 4 punti distribuiti attorno ad un punto $P(a,b)$. Questa situazione è quella in cui la grandezza f la si conosce su una griglia regolare a bassa risoluzione (esempio dimensione cella 10 km) e si vuole da essa ottenere il valore della grandezza in un punto (a,b) non coincidente con un nodo di griglia.





Il metodo procede:

1. **Prima** interpola linearmente nella direzione x , ottenendo la grandezza Q nei due punti R_1 e R_2
2. **Poi** interpola linearmente tra R_1 e R_2 ottenendo il valore di Q desiderato.

→ **Interpolazione Lineare Monodimensionale**
(lungo l'asse x)

- conosco i valori y_1 e y_2 che la grandezza y assume nei punti x_1 e x_2
- voglio il valore y che la grandezza assume nel punto x contenuto nell'intervallo $x_1 \div x_2$.

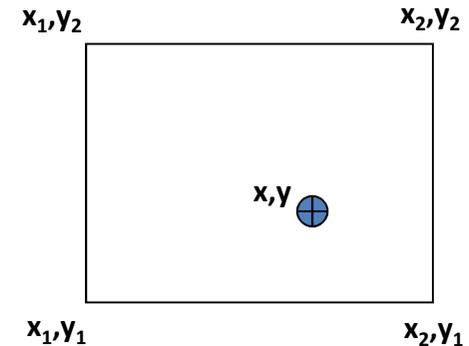
$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} \cdot (x - x_1)$$

Finalmente, **Interpolazione bi-Lineare**

1. Variabile da interpolare f

2. Valori noti:

- Q_{11} nel punto (x_1, y_1)
- Q_{12} nel punto (x_1, y_2)
- Q_{21} nel punto (x_2, y_1)
- Q_{22} nel punto (x_2, y_2)



3. Q è il valore interpolato nel punto (x, y) ottenuto con

la relazione bi-lineare

espressa in forma matriciale come:

$$Q(x, y) = \frac{1}{(x_2 - x_1) \cdot (y_2 - y_1)} \cdot \begin{bmatrix} (x_2 - x) & (x - x_1) \end{bmatrix} \cdot \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \cdot \begin{bmatrix} (y_2 - y) \\ (y - y_1) \end{bmatrix}$$

Breve richiamo alle regole per il calcolo matriciale

Prodotto matrice-vettore

$$\mathbf{y} := \mathbf{A} * \mathbf{x}, \quad \mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & \ddots & & \vdots \\ \vdots & & & \\ a_{n,1} & \cdots & & a_{n,n} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n \\ \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,n}x_n \end{bmatrix}$$

(n-1 somme + n prodotti) x n = 2n²-n operazioni

Esempio

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} \boxed{2} & \boxed{3} & \boxed{2} \\ 2 & 3 & 1 \\ 2 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \boxed{3} & 1 \\ \boxed{3} & 2 \\ \boxed{1} & 0 \end{pmatrix} = \begin{pmatrix} 17 & 8 \\ 16 & 8 \\ 9 & 4 \end{pmatrix}$$

4. METODI DI INTERPOLAZIONE OGGETTIVA

(Spazializzazione Oggettiva delle Misure)

Come anticipato, le due famiglie di Metodi di Interpolazione Oggettiva che consideriamo:

1. Partono dalla disponibilità di misure di una generica variabile ambientale scalare Q (es. meteorologica o di Q.A.) realizzata in M punti disposti irregolarmente sul territorio (ignoriamo le differenze di quota);
2. Partono dalla conoscenza di prima approssimazione (di background) della variabile Q negli N nodi di una Griglia di Analisi regolare, cioè partono dalla conoscenza di campo discreto di *first-guess*;
3. Producono la migliore stima possibile Q_a di Q che indichiamo col termine di **Analisi**, stima che differirà più o meno da un irraggiungibile Valore Vero. L'Analisi è quanto di meglio possiamo sperare di ottenere per il campo della variabile Q ai nodi della griglia di Analisi.

Quindi, in estrema sintesi:

- Se si dispone di misure Q_m negli M punti di misura,
- Se si conosce la statistica degli errori di misura ,
- Se si conosce il campo di background Q_b negli N punti della griglia di Analisi,
- Se si conosce la statistica degli errori di background

tutti i metodi delle due famiglie possono essere espressi formalmente come:

$$Q_a(x,y)$$

=

Funzione (Q_m ($m=1,..,M$), Statistica di Q_m , Q_{bi} ($i = 1, .., N$), statistica di Q_{bi})

Il **Modello Analitico** che lega l'Analisi in un punto i -esimo della griglia di Analisi alle M misure disponibili nei siti k ($k = 1, \dots, M$) potrebbe essere qualsiasi, ma seguendo una lunga tradizione vigente nel mondo dei Modelli Statistici, viene sempre adottato un **modello lineare**.

Per rendere più chiare le idee, consideriamo tre casi a complessità crescente.

- 1) Consideriamo un solo nodo i -esimo della griglia di Analisi. Se $Q_{m,k}$ sono le M misure disponibili, la relazione lineare che lega l'Analisi alle misure sarà di questo tipo:

$$Q_{a,i} = \sum_{k=1}^M h_{i,k} \cdot Q_{m,k}$$

dove $h_{i,k}$ è il peso che la misura k -esima ha nell'Analisi relativa al nodo di griglia i -esimo. Le due famiglie di metodi di spazializzazione differiscono proprio nel modo con cui vengono determinati questi pesi.

2) Consideriamo ora il caso in cui si voglia l'Analisi in 2 nodi di griglia (indicati come nodo 1 e 2) avendo a disposizione 3 misure, allora il modello lineare tra Analisi e misure può essere scritto nella forma seguente:

$$\begin{bmatrix} Q_{a,1} \\ Q_{a,2} \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{2,1} & h_{3,1} \\ h_{1,2} & h_{2,2} & h_{3,2} \end{bmatrix} \cdot \begin{bmatrix} Q_{m,1} \\ Q_{m,2} \\ Q_{m,3} \end{bmatrix}$$

Se si definisce la matrice dei pesi [H] come:

$$[H] = \begin{bmatrix} h_{1,1} & h_{2,1} \\ h_{2,1} & h_{2,2} \\ h_{3,1} & h_{3,2} \end{bmatrix}$$

mentre i vettori delle Analisi e delle Misure sono dati rispettivamente da:

$$[Q_a] = \begin{bmatrix} Q_{a,1} \\ Q_{a,2} \end{bmatrix}$$

$$[Q_m] = \begin{bmatrix} Q_{m,1} \\ Q_{m,2} \\ Q_{m,3} \end{bmatrix}$$

allora il modello lineare risulta formulato come:

$$[Q_a] = [H]^T [Q_m]$$

- 3) In generale**, se si desidera l'Analisi in N punti della griglia di Analisi usando le M misure disponibili, il modello lineare che lega analisi e misure è ancora data dalla relazione matriciale:

$$[Q_a] = [H]^T [Q_m]$$

dove, però, il vettore dell'Analisi $[Q_a]$ ha N elementi, il vettore delle Misure $[Q_m]$ ha M elementi e la matrice dei pesi $[H]$ ha tante colonne quante sono i punti di Analisi e tante righe quante sono le misure.

Entrambe le famiglie di Metodi di Interpolazione Oggettiva, anche se non detto esplicitamente e in una maniera o nell'altra, si basano sulla **congettura** seguente:

Prima Legge della Geografia (Legge di Tobler)

Tutto è correlato, ma oggetti vicini sono più correlati di oggetti lontani

(Everything is related to everyting else, but near things are more related then dinstant things).

Questa congettura emerge come naturale conseguenza dalla visione del mondo naturale dove le discontinuità risultano soprattutto dovute all'azione dell'uomo.

Le due famiglie di metodi si differenziano nei criteri che usano per giungere all'Analisi:

1) METODI SEMI-EMPIRICI: sono metodi che

- 1. Usano le M misure disponibili tenendo conto in modo semplificato delle problematiche relative agli errori di misura;**
- 2. Usano la Congettura di Tobler;**
- 3. Usano in modo semi-empirico similitudini deducibili da altri campi della Fisica Classica (campo elettromagnetico, campo gravitazionale, ecc.) che implicitamente vengono presi come modello funzionale di riferimento;**
- 4. Non usano alcun criterio di Ottimalità Statistica.**

2) **METODI BLUE (Best Linear Unbiased Estimator)**: sono metodi che partono da un'ipotesi di lavoro che vede:

- Il campo di Analisi Q_a agli N punti della Griglia di Analisi pari al campo di background, che viene elevato al rango di *vero first-guess* dell'Analisi
- **cui si somma un termine correttivo** che dipende:
 - dalle M misure disponibili,
 - dagli M valori assunti da campo di background nei siti di misura,
 - da opportuni pesi che, a loro volta, derivano:
 - dalla Statistica delle misure (riassunte per es. in σ^2_o)
 - dalla Statistica del background (riassunte per es. in σ^2_b)
 - dal fatto di rendere **Statisticamente Minima** la differenza quadratica media tra l'Analisi negli N punti della griglia ed il valore vero della variabile negli N punti della griglia.

Questi metodi si ripartiscono **in due famiglie distinte**, ma matematicamente e statisticamente equivalenti:

→ Metodi di **Optimal Interpolation** (O.I.) (Gandin, 1963)

→ Metodi di **Kriging** (Krige, 1950)

Anche se formalmente riconducibili l'uno all'altro, questi metodi ormai vengono applicati, soprattutto per consuetudine, in ambiti fisici e ambientali differenti:

- **Suolo, sottosuolo, prospezioni geologiche e studi minerari** vengono normalmente condotti usando i diversi metodi di Kriging disponibili,
- **Meteorologia, climatologia, qualità dell'aria** usano normalmente metodi di Optimal Interpolation (O.I.),

anche se non mancano eccezioni nell'impiego pratico delle due famiglie di metodi di spazializzazione.

NOTA

Il metodo O.I. fino a pochi anni orsono ha costituito la metodologia pressoché unica per realizzare:

- 1. Le Analisi Meteorologiche Automatiche ,**
- 2. La predisposizione delle condizioni iniziali ed al contorno per le previsioni meteorologiche numeriche (a partire dalla prima realizzata nel 1950 dal gruppo di Charney),**
- 3. La realizzazione delle Analisi a posteriori dei campi meteorologici assimilati con le misure,**
- 4. Le Analisi a posteriori e le ricostruzioni assimilate con le misure dei campi che definiscono la qualità dell'aria**
- 5. La predisposizione delle condizioni iniziali ed al contorno necessarie per la previsione della qualità dell'aria a piccola, media e grande scala)**

e ancor oggi il Centro Regionale della Qualità dell'Aria di ARPA Lazio usa questa metodologia operativa di Analisi nelle proprie Catene Modellistiche.

Nei grandi Centri Meteorologici Mondiali (ECMWF, NCEP, ecc.) ora sono in uso tecniche molto più avanzate (quasi tutte sviluppate e/o applicate operativamente da Eugenia Kalnay), in qualche modo riconducibili all'O.I., che spazio-temporizzano l'Assimilazione delle misure utilizzando i metodi matematici propri dell'**Analisi Variazionale**.

Vanno ricordati i metodi:

→ 3D-Var

→ 4D-Var

→ Filtro di Kalman

Queste tecniche, per la loro estrema complessità matematica e statistica, non possono essere presentate in questa sede neppure in forma semplificata.

4.1 METODI SEMIEMPIRICI

Successive Correction Methods (SCM)

Questi metodi, pur con varianti significative, presentano tutti la medesima **struttura algoritmica** che, in sostanza, è un **insieme di iterazioni con cui si migliora e si perfeziona progressivamente il campo di Analisi individuato all'iterazione precedente.**

→ Nodi della griglia di Analisi N . Nodo generico i ($i = 1, \dots, N$)

→ Punti di misura M . Punto di misura generico m ($m = 1, \dots, M$)

1. $x_{a,i}(k)$ = analisi nel nodo i -esimo all'iterazione k -esima
2. $x_{a,m}(k)$ = analisi nel punto di misura m -esimo all'iterazione k -esima
3. $x_{b,i}$ = valore del campo di background per il nodo di griglia i -esimo
4. $x_{b,m}$ = valore del campo di background per il punto di misura m -esimo
5. y_m^0 = misura disponibile nel punto m -esimo
6. $W_{i,m}$ = peso che la misura rilevata nel punto m -esimo ha sull'analisi per il punto di griglia i -esimo
7. ε^2 = stima del rapporto tra la varianza degli errori di misura σ^2_0 e la varianza del campo di background σ^{2b}

ALGORITMO

Passo 0: il campo di Analisi sia nei nodi i della griglia di Analisi che nei siti m di misura è pari al campo di background in tali punti:

$$\begin{aligned}x_{a,i}(0) &= x_{b,i} \\ x_{a,m}(0) &= x_{b,m}\end{aligned}$$

Passo 1: l'Analisi nei nodi i della griglia di Analisi è data in questa prima iterazione dalla relazione:

$$x_{a,i}(1) = x_{a,i}(0) + \frac{\sum_{m=1}^M [W_{i,m}^{(0)} \cdot (y_m^0 - x_{a,m}(0))]}{\sum_{m=1}^M W_{i,m}^{(0)} + \epsilon^2}$$

e una relazione del tutto analoga vale per l'Analisi relativa ai punti di misura.

L'espressione analitica con cui si calcolano i pesi $W_{i,m}$ (che come si vede, possono variare con l'iterazione) è l'elemento che differenzia i diversi metodi *SCM* e di essi discuteremo tra poco.

Passo 2÷K: ad ogni iterazione successiva ($k+1$) viene aggiornata l'Analisi prodotta alla iterazione (k) nel modo seguente, per quanto riguarda i nodi i di Analisi:

$$x_{a,i}(k+1) = x_{a,i}(k) + \frac{\sum_{m=1}^M [W_{i,m}^{(k)} \cdot (y_m^0 - x_{a,m}(k))]}{\sum_{m=1}^M W_{i,m}^{(k)} + \varepsilon^2}$$

e la stessa equazione, in cui l'indice del nodo di griglia i viene sostituito dall'indice m del punto di misura, vale anche per determinare l'Analisi all'iterazione ($k+1$) relativa ai punti di misura.

La convergenza del metodo iterativo la si raggiunge in genere dopo poche iterazioni (tipicamente 4)

La struttura algoritmica e funzionale di tutti i metodi SCM è la medesima ed è quella qui presentata.

I vari metodi si differenziano però per come vengono calcolati i vari pesi $W_{i,m}$ e come essi vengono fatti variare da iterazione ad iterazione.

METODO DI CRESSMAN (1959)

Questo metodo parte dal presupposto che ad ogni sito di misura m sia associato un Raggio di Influenza $R_m(k)$ che può variare con l'iterazione k .

Ogni punto di osservazione fornisce una misura della grandezza Q che:

- Naturalmente costituisce il **valore locale di Q** nell'istante di osservazione
- ma, in *qualche modo*, fornisce anche un'indicazione del valore di Q nell'**intorno del punto di osservazione stesso**.

Quindi, ogni sito di osservazione sperimentale si ipotizza che rappresenti un'intera porzione dello spazio (Area di Influenza) che, se immaginata di forma circolare, può essere descritta proprio da un **Raggio di Influenza**.



In pratica, la misura in un sito si estende all'intera Area di Influenza ad esso legata e tale Area presenterà valori di Q *parenti* di quelli rilevati nel sito di misura e legati ad essi da un opportuno Modello di Propagazione.

Ma questa parentela, e quindi l'estensione dell'**Area di Influenza**, non potrà essere un concetto assoluto, ma **dipenderà da cosa effettivamente voglio osservare**.

Prima di approfondire il concetto di Raggio di Influenza, conviene introdurre almeno qualitativamente, il concetto di Scala Spaziale della variabile Q e a questo scopo conviene considerare come esempio la concentrazione di PM_{10} , variabile decisamente familiare a tutti noi e che osserviamo quotidianamente con la rete di monitoraggio regionale della Q.A.

Se osserviamo attentamente il *near-realttime* del CRQA o un suo *forecast*, notiamo che la distribuzione spaziale del PM_{10} su tutta la regione può essere vista come la somma di alcuni contributi tipici:

- 1) Un valore poco variabile sull'intera regione, che rappresenta il fondo regionale, cioè un valore medio che non evidenzia le peculiarità locali, ma rappresenta a grana grossa (o a bassa soluzione, che è lo stesso) la presenza del PM_{10} dovuto al complesso delle emissioni regionali trasportate e disperse dalle masse d'aria e dalla turbolenza dell'atmosfera media del territorio, dovute anche ad una provenienza extra regionale (dalle altre regioni confinanti, da lunga distanza come le polveri sahariane). Questo contributo, più o meno, lo sentono tutte le centraline di monitoraggio della rete regionale e lo possiamo indicare come contributo di fondo.

2) A questo valore lentamente variabile si viene a sommare un contributo legato alle peculiarità geografiche regionali che possono esercitare effetti di confinamento o di disaccoppiamento. Nello specifico, l'Agglomerato di Roma, la Zona Litoranea, la Zona Appenninica e la Zona Valle del Sacco (in cui viene zonizzata la regione ai fini della Q.A.) hanno caratteristiche geografiche e meteorologiche tali da differenziarsi tra loro e da potersi disaccoppiare. Spesso avviene che ad un innalzamento di concentrazione nella Valle del Sacco non corrisponde un analogo innalzamento di concentrazione nella zona litoranea. Il valore tipico (per esempio medio) di concentrazione in queste macro-aree potremmo indicarlo col termine contributo a grande-scala.

3) Se ora consideriamo con più attenzione (a maggior risoluzione) una delle porzioni di territorio omogenee a grande scala, per esempio l'Agglomerato di Roma, vediamo che il PM_{10} varia a seconda di si trovi in un' area occupata dalla campagna attorno alla città, in porzioni di aree occupate da comuni importanti presenti nell'agglomerato e nella porzione di area occupata dalla città di Roma. Se nell'agglomerato separiamo questa porzioni di territorio e mediamo le concentrazioni rilevate in ognuna di esse otteniamo il contributo di concentrazione a meso-scala.

- 4) A questo punto consideriamo una di queste porzioni di territorio, per esempio la città di Roma: le centraline che in esse sono attive evidenziano come ci siano **grosse differenze da luogo a luogo**. La concentrazione rilevata a Villa Ada è molto diversa da quella rilevata a Cinecittà che è diversa da quella rilevata a Magna Grecia e così via. Questa diversità è dovuta a varie cause locali: la presenza e l'estensione di un parco, la dimensione e la geometria dei canyon urbani, la geometria degli edifici in una piazza, ecc. Se ripartiamo l'intero agglomerato in aree omogenee tipologicamente (parco, canyon di diverse orientazioni, pizze, ecc.) e mediamo la concentrazione in ciascuna di queste aree evidenziamo a questo punto il contributo locale del PM_{10} .
- 5) Ma se consideriamo nel dettaglio una qualsiasi di queste aree elementari, ci accorgiamo che al suo interno la variabilità media di PM_{10} esiste ancora a causa delle **peculiarità tipiche del sito** (per esempio, se consideriamo il canyon dove è posta la centralina di via Tiburtina, troveremmo una bella differenza tra il PM_{10} sul marciapiede in cui sta la centralina e la concentrazione al marciapiede opposto). Quindi abbiamo evidenziato i contributi a microscala.

.. e il processo potrebbe continuare aumentando costantemente la risoluzione con cui osserviamo il fenomeno.

Ciò che abbiamo detto, *mutatis mutandis*, vale per qualsiasi variabile di interesse ambientale (per esempio le variabili meteorologiche e della qualità dell'aria).

Quindi possiamo asserire che la **misura ottenuta in una postazione può essere vista come la somma di diversi contributi additivi e indipendenti** (almeno in prima approssimazione):

1. un **contributo di fondo** (che costituisce la visione a bassissima risoluzione del fenomeno);
2. un **contributo a grande scala** (in cui emergono almeno le grosse e più evidenti peculiarità territoriali);
3. un **contributo a meso-scala** (in cui la risoluzione è ulteriormente aumentata e si comincia a tener conto della orografia e delle caratteristiche geografiche)
4. un **contributo a scala locale** (ora il fenomeno la si vede ad elevata risoluzione e cominciano ad emergere le peculiarità locali legate al land-use, per esempio)
5. un **contributo a microscala** in cui sono completamente evidenziate le pur minime caratteristiche locali.

Nella ricostruzione della distribuzione spaziale della generica variabile Q conviene **lasciarsi guidare dalla fisica del problema e dalle caratteristiche geografiche**, sfruttando completamente le possibilità offerte dall'algoritmo:

- si **definiscono alcuni Raggi di Influenza** (per esempio 4) che rappresentano grossolanamente le scale spaziali tipiche del fenomeno (il fondo, la grande scala, la meso-scala e la scala locale)
- si definisce **come l'informazione legata ad una misura si propaga** nello spazio attorno al punto in cui è stata rilevata se si osserva il fenomeno prendendo a riferimento una delle scale spaziali considerate. Ciò si può fare in tanti modi diversi, ma comunque **ritenendo valida la Congettura di Tobler**.

Si gestiscono le diverse iterazioni utilizzando il modello lineare in maniera da far emergere progressivamente le varie scale del fenomeno. In particolare:

1. **prima si ricostruisce la distribuzione di Q a bassa risoluzione**
2. **poi la si raffina nelle iterazioni successive introducendo gradualmente le informazioni alle scale più fini rilevate dalle postazioni di misura.**

Dal punto di vista operativo il metodo di Cressman, che concretizza questa logica di ricostruzione del campo a risoluzione crescente, procede nel modo seguente:

1. Le varie iterazioni del metodo, cui corrispondono altrettante Analisi della variabile Q a risoluzione crescente, verranno condotte con il modello lineare che vede l'Analisi come la somma pesata delle misure. Noto ad ogni iterazione il raggio di influenza $R_m^{(k)}$ (che progressivamente si va riducendo per aumentare la risoluzione dell'Analisi), il **peso** $W_{i,m}$ (variabile con l'iterazione k) che la misura m ha sull'Analisi relativa al nodo di griglia i **dipende dalla distanza Euclidea** $d_{i,m}$ **tra i due nodi** secondo la relazione seguente:

$$W_{i,m}^{(k)} = \begin{cases} \frac{\left(R_m^{(k)}\right)^2 - d_{i,m}^2}{\left(R_m^{(k)}\right)^2 + d_{i,m}^2} & \text{se } d_{i,m}^2 \leq \left(R_m^{(k)}\right)^2 \\ 0.0 & \text{se } d_{i,m}^2 > \left(R_m^{(k)}\right)^2 \end{cases}$$

2. Inizialmente l'Analisi deve contenere una visione d'insieme per non perdere il fenomeni a grande scala. Quindi bisogna che l'influenza di ogni singola misura sia la massima possibile (il raggio di influenza è molto elevato e dell'ordine delle dimensioni dell'intero dominio spaziale di interesse). In questo modo il campo di Analisi tiene conto di tutto quanto viene rilevato dall'insieme dei punti di misura e quindi del *livello di fondo*. A questa Scala Spaziale corrisponderà un raggio di influenza $R_m^{(1)}$ che può essere posto pari alla diagonale del dominio di Analisi. Così facendo, si vede immediatamente come nella prima Analisi a bassa risoluzione tutte le postazioni di misura influenzano ogni nodo di griglia visto che tutte le misure contengono l'informazione del fondo.

3. Nelle iterazioni successive, progressivamente riducendo il raggio di influenza, si inseriscono nell'Analisi i fenomeni a scala via via più piccola fino ad inserire nelle immediate vicinanze del punto di misura gli effetti locali.

Per esempio: Servizio Meteorologico Svedese (1980)

$$R(1) = 1500 \text{ km}$$

$$R(2) = 1200 \text{ km}$$

$$R(3) = 750 \text{ km}$$

$$R(4) = 400 \text{ km}$$

NOTA

- a) Ogni variabile ambientale è un caso a sé e va vista considerando la Fisica (o la Chimica) che l'ha generata e la scala spaziale a cui si vuole analizzarla (scala regionale o di un territorio di dimensioni inferiori; es. la distribuzione di temperatura sull'Agglomerato di Roma o la distribuzione di Diossine e Furani sui suoli della Valle del Sacco attorno a Frosinone). Quindi, dato il problema, sarà cura di chi fa l'Analisi definire quale è la Scala Spaziale (quindi il Raggio di Influenza) per il fondo, quale per la grande scala, quale per la meso-scala e quale per la scala locale.
- b) L'Algoritmo prevede Raggi di Influenza variabili non solo con le scale spaziali e quindi con le iterazioni, ma anche con le diverse postazioni di misura. Tuttavia, se gli strumenti ed i metodi di campionamento sono gli stessi nelle diverse postazioni di misura, per una data iterazione il raggio di influenza sarà indipendente dalla specifica postazione m -esima considerata. Può però capitare che, ad una data scala, peculiarità geografiche o meteorologiche possano consigliare scelte diverse e quindi in questo caso, cui accenneremo brevemente nel seguito, sarà opportuno introdurre leggi semi-empiriche di differenziazione dei Raggi di Influenza in base alle informazioni disponibili.

ALGORITMO

Passo 0: il campo di Analisi sia nei nodi i della griglia di Analisi che nei siti m di misura è pari al campo di background in tali punti:

$$\begin{aligned}x_{a,i}(0) &= x_{b,i} \\ x_{a,m}(0) &= x_{b,m}\end{aligned}$$

Passo 1÷K: l'Analisi nei nodi i della griglia di Analisi è data in questa prima iterazione dalla relazione:

$$x_{a,i}^{(k+1)} = x_{a,i}^{(k)} + \frac{\sum_{m=1}^M W_{i,m}^k \cdot [y_m^0 - x_{a,m}^{(k)}]}{\sum_{m=1}^M W_{i,m}^k + \varepsilon^2}$$

e una relazione del tutto analoga vale per l'Analisi relativa ai punti di misura.

Tutte le conoscenze statistiche relative agli errori di misura (condensate nella varianza σ^2_0) e agli errori di background (condensate nella varianza σ^2_b) sono raggruppate nell'unico parametro ε^2 che è l'unico parametro di *tuning* consentito dal metodo.

- Se $\varepsilon^2 = 1$ si ipotizza una **piena sfiducia nei confronti delle misure** e si addossa tutto il peso della spazializzazione al campo di background tenendo conto comunque al meglio delle misure disponibili
- Se, però, $\varepsilon^2 = 0$, allora **si ritiene inesistente l'errore di misura** (di fatto una pazzia e un falso) o **si ritiene di non avere informazioni sul campo di background** (cosa difficile per quanto abbiamo detto in precedenza).

Nelle applicazioni più antiche di questo metodo spesso si è posto $\varepsilon^2 = 0$, riponendo eccessiva fiducia nelle misure ed ottenendo spesso campi altamente irregolari.

Il numero di iterazioni è importante anche se il metodo converge velocemente.

In applicazioni poco critiche e spesso per semplicità si può eseguire solo la prima iterazione.

METODO DI BARNES (1964, 1978)

Questo metodo *SCM* effettua il calcolo dei pesi $W_{i,m}$ in maniera spazialmente meno rigida del metodo di Cressman. In pratica, l'influenza di un punto di misura non si esaurisce mai con la distanza tra punto di griglia e punto di misura, ma diminuisce con essa progressivamente e molto rapidamente, senza però esaurirsi mai.

Questa diminuzione è pilotata da una *Lunghezza di Scala* caratteristica L_m^k che, come sempre, può variare con il punto di misura, per adeguarsi alle diverse entità e statistiche degli errori di misura, e con l'iterazione k .

- Nella prima iterazione la lunghezza di scala L_m^k è tale da distribuire su una scala molto vasta l'influenza di una singola misura. Ciò comporterà che nel punto di analisi verranno *sentite* praticamente tutte le misure, con intensità (peso) variabile con la rispettiva distanza, ma tutte queste influenze verranno miscelate filtrando così gli effetti a piccola scala e locali.
- Nelle iterazioni successive, la riduzione della Lunghezza di Scala L_m^k fa sì che, per la Congettura di Tobler, le peculiarità viste da una singola postazione di misura vengono trasferite progressivamente a punti sempre più prossimi al sito stesso.

In pratica il metodo di Barnes realizza tutto ciò utilizzando questa relazione funzionale per i pesi $W_{i,m}$ che, come si vede, prevede un **peso di tipo Gaussiano centrato sul punto di misura**:

$$W_{i,m}(k) = \exp \left[-\frac{d_{i,m}^2}{2(L_m^k)^2} \right]$$

dove L_m^k è la Lunghezza di Scala all'iterazione k -esima.

Operativamente le scelte che si possono fare sono analoghe a quelle fatte per il raggio di influenza su cui si basa il metodo di Cressman.

Per esempio, per una ricostruzione meteorologica a scala limitata (regione) si potrebbero adottare le Lunghezze di Scala seguenti:

$$L(1) = 400 \text{ km}$$

$$L(2) = 200 \text{ km}$$

$$L(3) = 150 \text{ km}$$

$$L(4) = 50 \text{ km}$$

NOTA 1

Tutti i metodi *SCM* (quelli illustrati ed i numerosissimi proposti in alternativa nella letteratura Scientifica):

- **si basano tutti sulla Congettura di Tobler,**
- **sono sostanzialmente ed intrinsecamente geometrici,**
- **tengono conto parzialmente delle proprietà statistiche delle misure e del campo di background solamente attraverso il rapporto ε^2 delle rispettive varianze**

NOTA 2

Come si vede, l'Analisi nei punti di misura non coincide col valore della misura locale e ciò è ragionevole visto che la misura è inevitabilmente affetta da un errore o incertezza intrinseca.

Solo se $\varepsilon^2 = 0$, allora c'è coincidenza tra analisi e misura, anche se, come abbiamo visto, questa è sostanzialmente una falsità.

NOTA 3

Tutti i metodi *SCM* presentano una riduzione con il procedere delle iterazioni del Raggio di Influenza (Metodo di Cressman) o della Lunghezza di Scala (Metodo di Barnes). Non ci sono regole specifiche per realizzare ciò. L'unica cosa dimostrabile è che la velocità di convergenza del metodo cresce con il tasso di riduzione del raggio di Influenza (o della Lunghezza di Scala).

Se, per esempio, consideriamo la Lunghezza di Scala L_{k+1} per il modello di Barnes all'iterazione $k+1$ (ragionamenti analoghi possono essere fatti per il Raggio di Influenza del Metodo di Cressman) la convergenza del metodo è garantita sempre, ma è decisamente rapida se:

$$\gamma = \frac{L_{k+1}^2}{L_k^2}$$

$$0 \leq \gamma \leq 1$$

$$L_{k+1} = \sqrt{\gamma} \cdot L_k$$

Il numero M e la distribuzione spaziale delle postazioni di misura condiziona ovviamente la possibile scelta delle Scale alle diverse iterazioni.

In particolare non è conveniente effettuare un'Analisi ad una scala inferiore alla **distanza minima tra le diverse stazioni R_{\min}** . Se le M stazioni di misura sono distribuite uniformemente in un'area A , allora una buona scelta di R_{\min} potrebbe essere:

$$R_{\min} = \sqrt{\frac{A}{M}}$$

La scelta di γ dipende dal problema e dall'operatore. Si potrebbe per esempio:

- stabilire L_1 come diagonale del dominio di Analisi
- stabilire $L_k = R_{\min}$

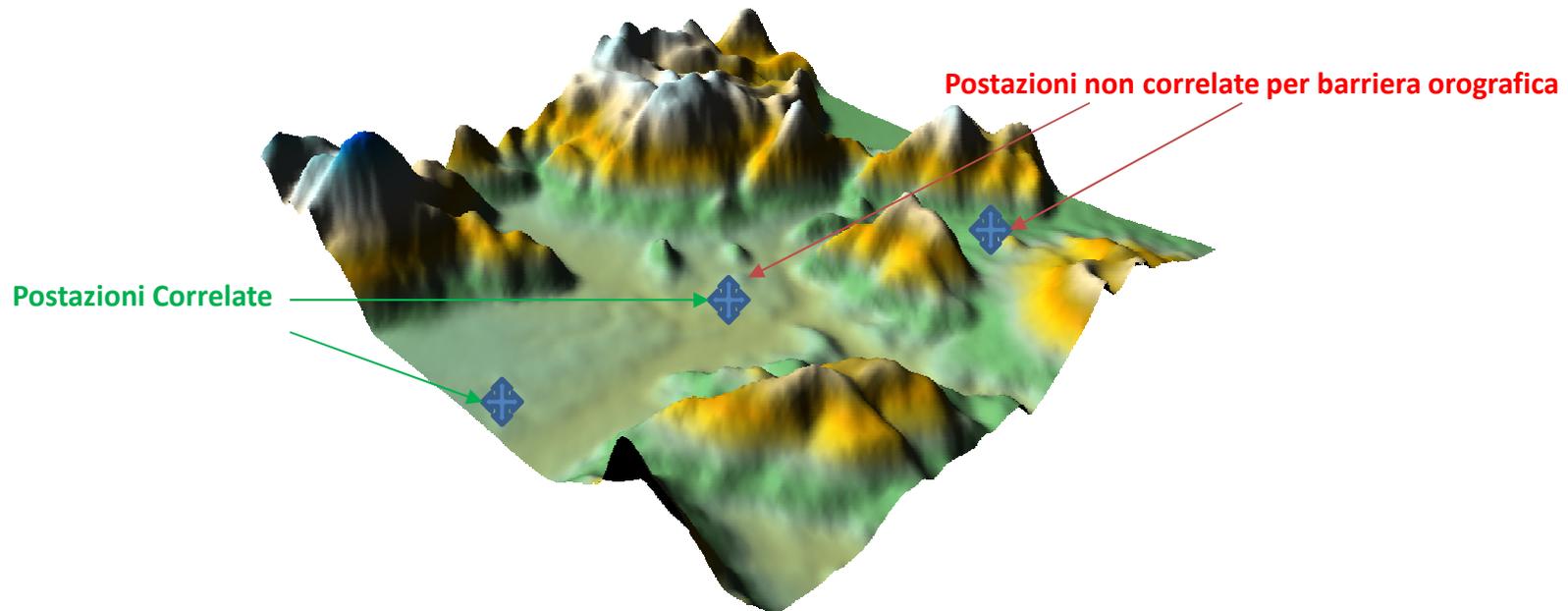
Se si vuole che ciò sia vero bisogna che valga la relazione che lega tra loro Numero di Iterazione K e parametro γ :

$$\gamma = \left(\frac{L_K}{L_1} \right)^{\frac{2}{K-1}}$$

Come si vede, se si passa in 4 iterazioni da L_1 a L_K , γ deve valere 0.25.

NOTA 3

Questi metodi sembrerebbe non possano tener conto esplicitamente di eventuali peculiarità caratteristiche del dominio di calcolo (es. valli, catene montuose, barriere varie).

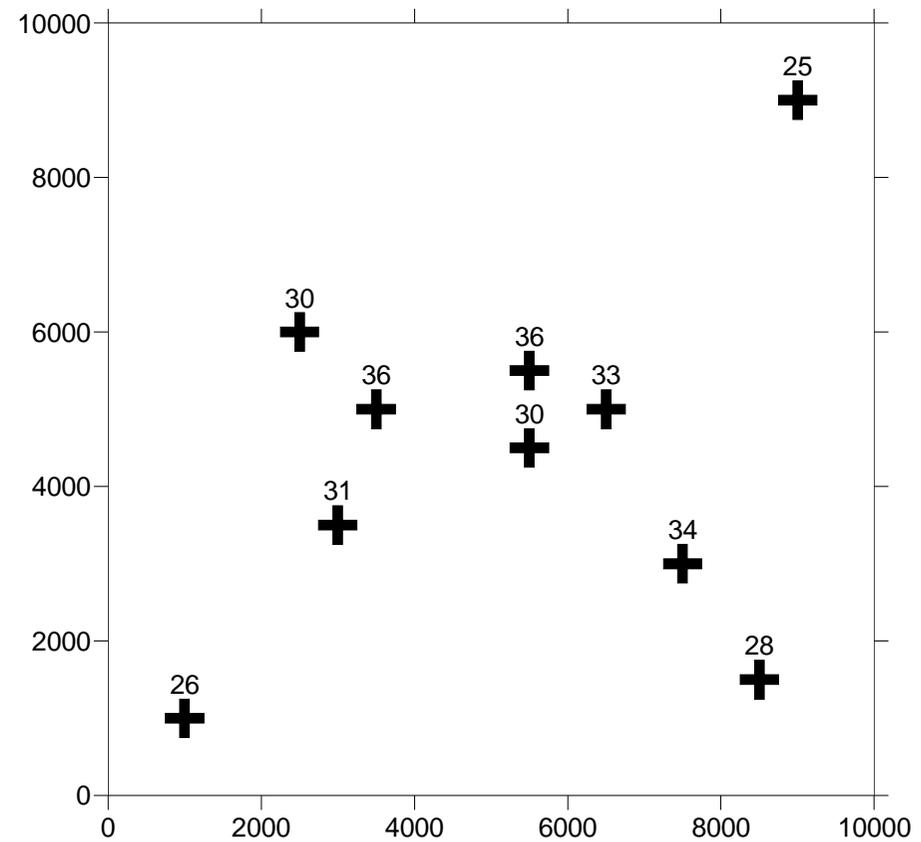


In realtà è **sempre possibile** agire sulle relazioni che esprimono i vari pesi $W_{i,m}$ in modo da tener conto di barriere che possono impedire la correlazione tra due punti, anche se la loro distanza reciproca risulta inferiore al raggio di influenza.

ESEMPIO 1

In questo esempio l'Analista ha a disposizione la misura della temperatura dell'aria a 2 m dal suolo rilevata in 10 stazioni meteorologiche distribuite su un dominio spaziale di 10 x 10 km. e nessuna informazione aggiuntiva.

X	Y	T(C)
1000	1000	26
8500	1500	28
3000	3500	31
7500	3000	34
3500	5000	36
5500	4500	30
6500	5000	33
5500	5500	36
2500	6000	30
9000	9000	25



A) Informazioni Statistiche sulle misure

In tutte le stazioni di misura la temperatura è misurata con **sensori** (termometri) dello stesso tipo.

L'**incertezza intrinseca dichiarata** del termometro è pari a 0.5 °C

La **varianza associata alle misura** è identica per ogni postazione di misura ed è pari a $\sigma^2_0 = 0.25 \text{ °C}^2$.

B) Informazioni a priori (Campo di Background)

Non si dispone di alcuna informazione sul *land-use* del territorio né sull'*orografia*. Non si può fare altro che costruire un **campo di background costante e pari al valore medio delle temperature rilevate**.

Valore di Background = 30.9 °C (costante ovunque)

Varianza dei valori che hanno creato la media = 3.7 °C²

Errore standard nel Valore di Background = 1.16 °C

La **Varianza di Background** è identica in tutti i punti del dominio di Analisi ed è pari a $\sigma^2_b = 1.34 \text{ °C}^2$

C) Statistica Complessiva

$\varepsilon^2 = 0.19$

D) Informazioni di Scala

Distanza massima tra Stazioni di Misura = 11.313 m

Distanza minima tra Stazioni di Misura = 1.000 m

Applicazione SCM (Barnes)

Numero Iterazioni = 5

Scale spaziali:

iterazione 1 → 5157 m

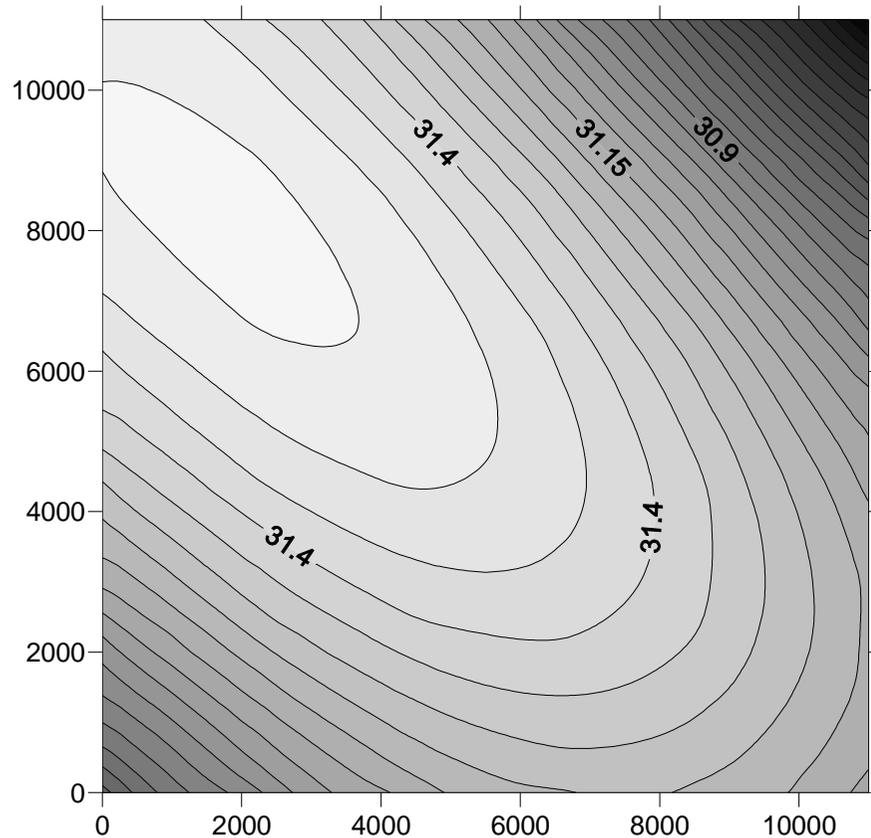
iterazione 2 → 2578 m

iterazione 3 → 1289 m

iterazione 4 → 645m

iterazione 5 → 322 m





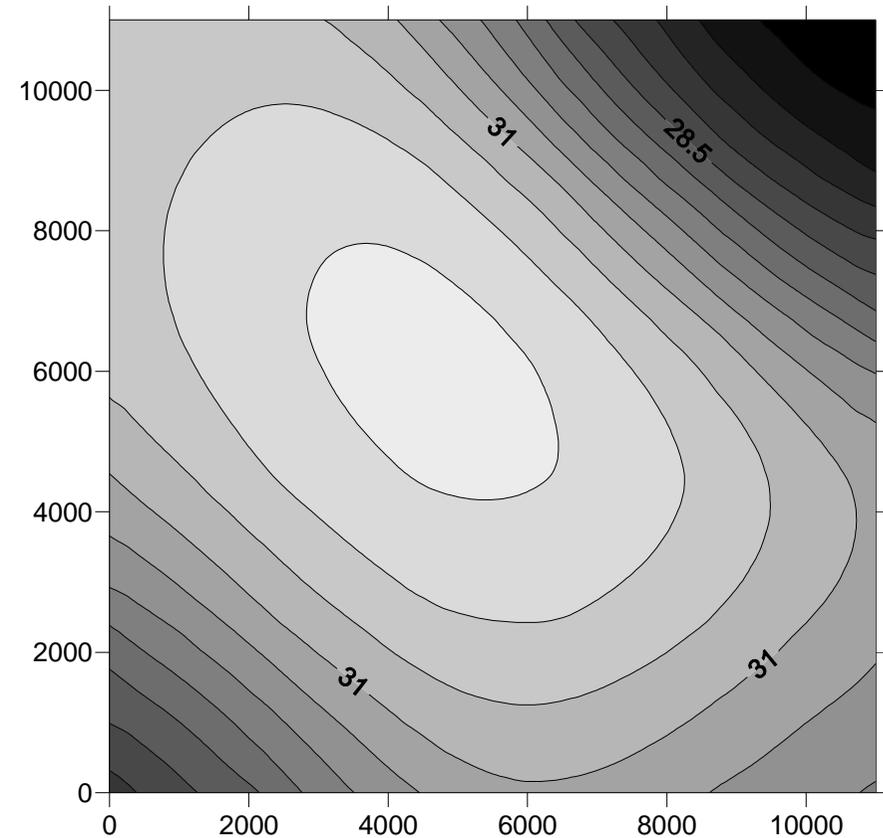
⇒ Iterazione 1

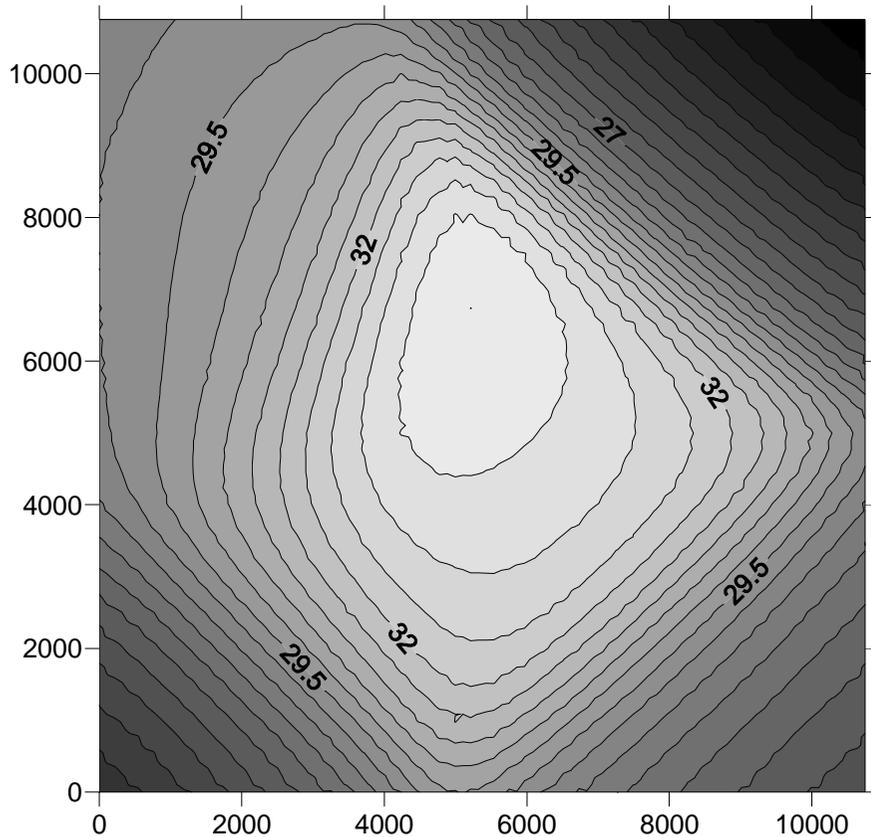
Le misure inseriscono nel Background (inizialmente costante ovunque) le prime informazioni relative ad una **Scala Spaziale dell'ordine dei 5000 m**.

Quindi, le informazioni a grande scala evidenziano una **prima variabilità spaziale della temperatura**

⇒ Iterazione 2

L'inserimento delle **informazioni a media scala (dell'ordine dei 2500 m)** cominciano a mostrare i **primi dettagli spaziali del campo di temperatura della zona**





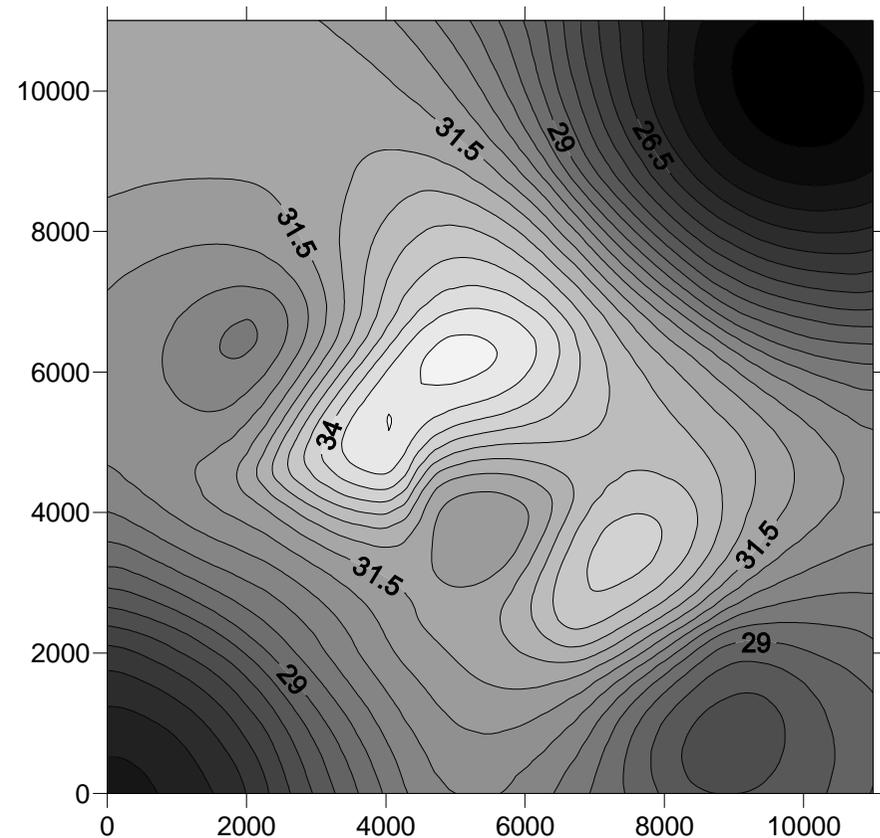
⇒ Iterazione 3

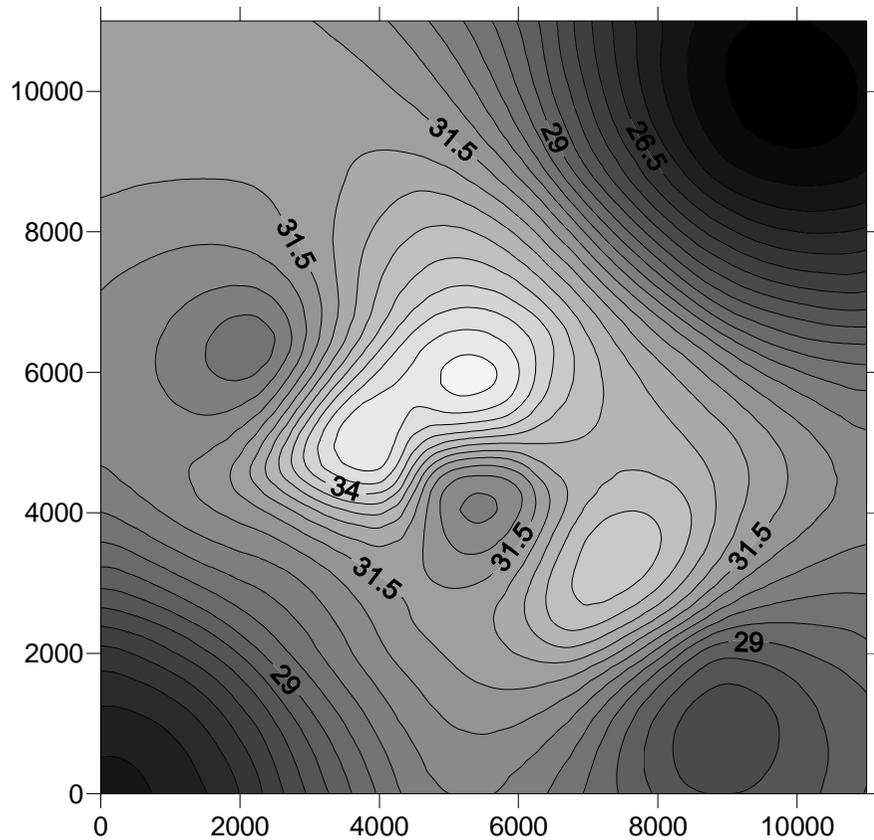
Ora vengono trasferiti al campo di analisi i dettagli corrispondenti ad una **Scala Spaziale dell'ordine di 1300 m**.

Emerge a questo punto una porzione di territorio con temperature elevate, circondata da zone più fredde.

⇒ Iterazione 4

La zona a temperatura elevata, che pareva omogenea all'iterazione precedente, una volta inserite le informazioni alla **Scala Spaziale dell'ordine dei 650 m** mostra una **struttura spaziale decisamente più complessa**.



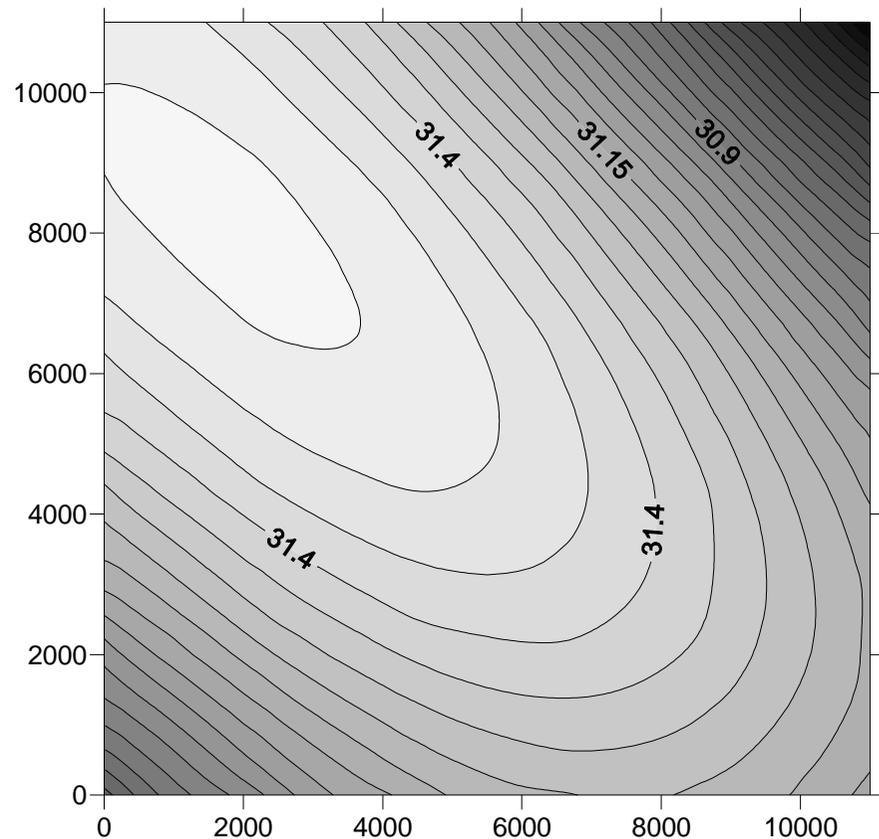


Punto di Arrivo del SCM di Barnes

Punto di Partenza del SCM di Barnes ⇒

⇒ **Iterazione 5 (Finale)**

L'inserimento delle informazioni alla **Scala Spaziale di circa 300 m** perfeziona, anche se di poco, il campo ottenuto all'iterazione precedente.



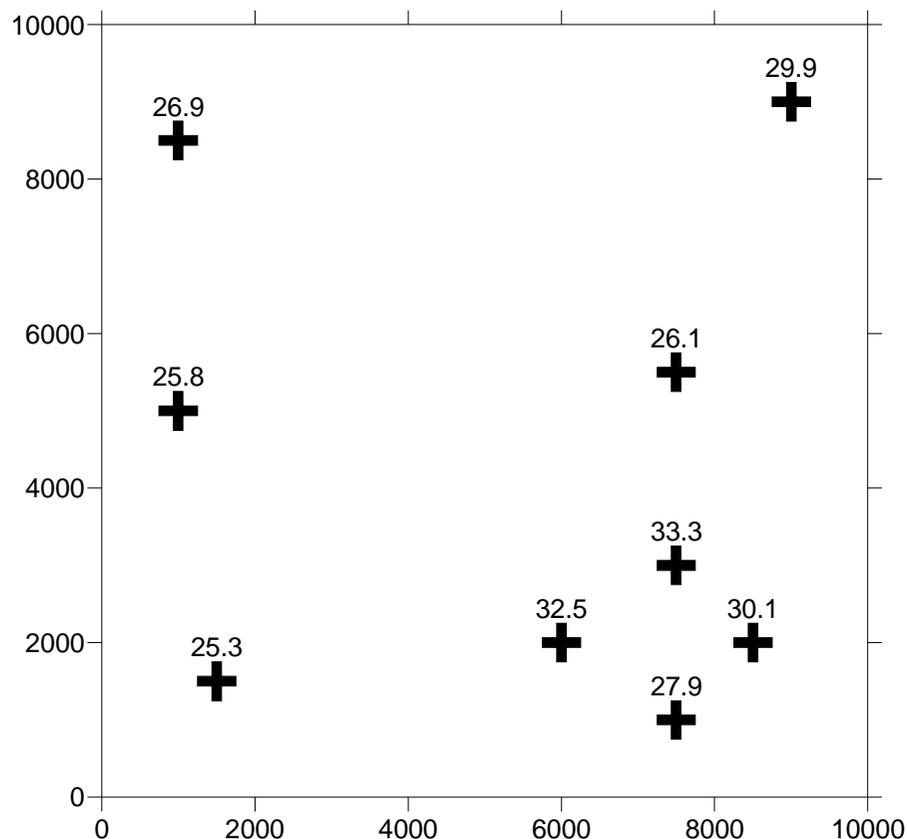
ESEMPIO 2

In questo caso si hanno a disposizione per l'Analisi 9 stazioni di misura di temperatura a 2 m distribuite su un dominio spaziale di 10 x 10 km.

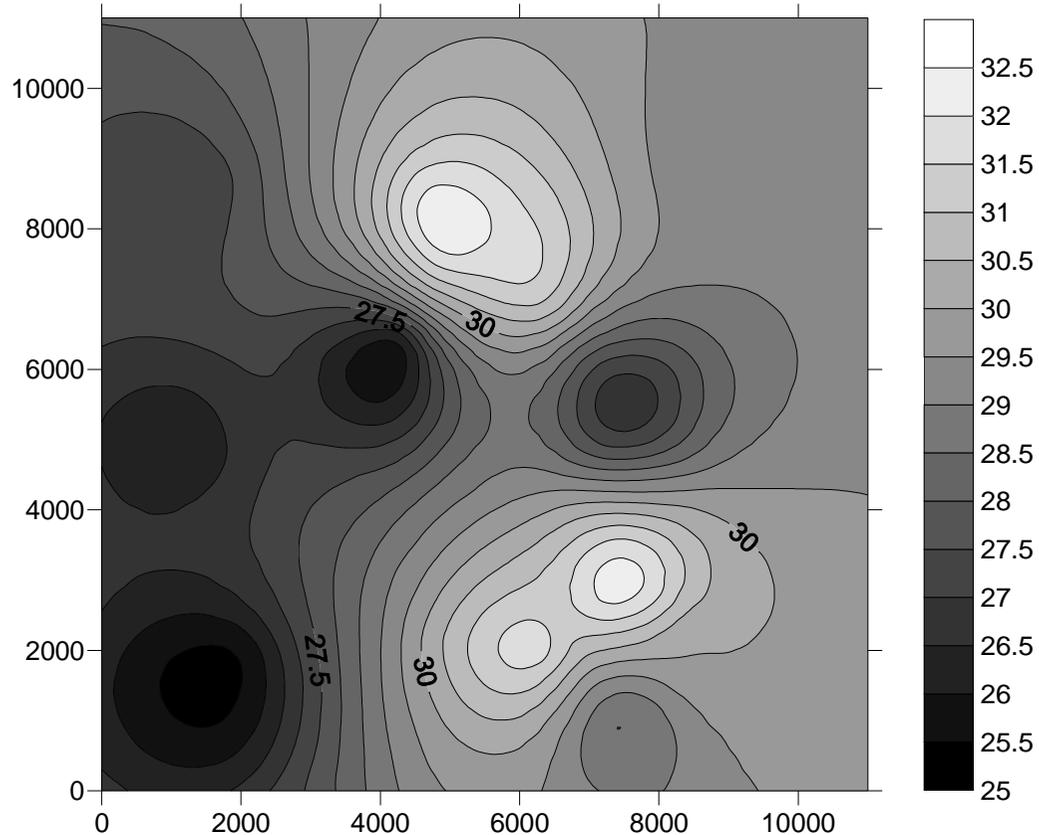
Le misure disponibili anche in questo caso tutte le stazioni impiegano lo stesso termometro caratterizzato da un'incertezza intrinseca di 0.2 °C.

Quindi la **Varianza delle Misure risulta pari a $\sigma_0^2 = 0.05 \text{ } ^\circ\text{C}^2$**

x	y	T(C)
1500	1500	25,3
7500	1000	27,9
6000	2000	32,5
8500	2000	30,1
7500	3000	33,3
1000	5000	25,8
7500	5500	26,1
1000	8500	26,9
9000	9000	29,9



Cosa non comune nella realtà, in questo caso la **distribuzione reale di temperatura** risulta quella riportata sotto, **ma non è nota all'Analista**. Con questo a disposizione sarà possibile verificare la bontà della ricostruzione del campo realizzata dal SCM di Barnes.

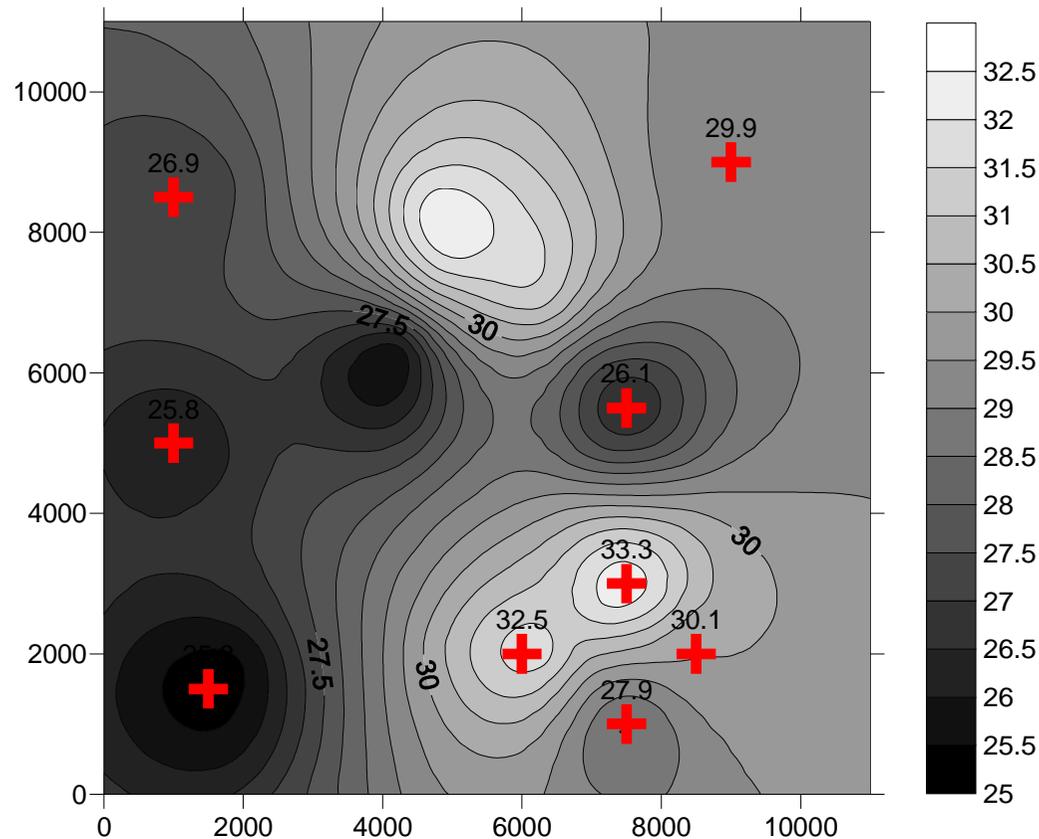


Per varie ragioni l'Analista non ha alcuna informazione a priori su cui basare la propria analisi se non i dati stessi.

Senza che l'Analista lo sappia o lo sospetti, **nella realtà la disposizione dei punti di misura nello spazio e il campo di temperatura sono quelli riportati qui sotto.**

Si vede immediatamente come nella zona Nord del dominio in cui il campo di temperatura reale presenta un **massimo relativo** non è stato collocato alcun punto di misura.

Difficilmente il SCM potrà ricostruire informazioni perse in origine!



..... ma tutte queste cose l'Analista non le può sapere!

A) Campo di Background

L'Analista, non disponendo di informazioni a priori, anche in questo caso non può che utilizzare un campo di Background costante nello spazio e pari al valor medio delle temperature misurate. Quindi il Campo di Background risulta in ogni punto pari a 29 °C.

La Varianza di Background (pari al quadrato dell'errore standard associato alla media) è pari a $\sigma_b^2 = 0.89 \text{ °C}^2$

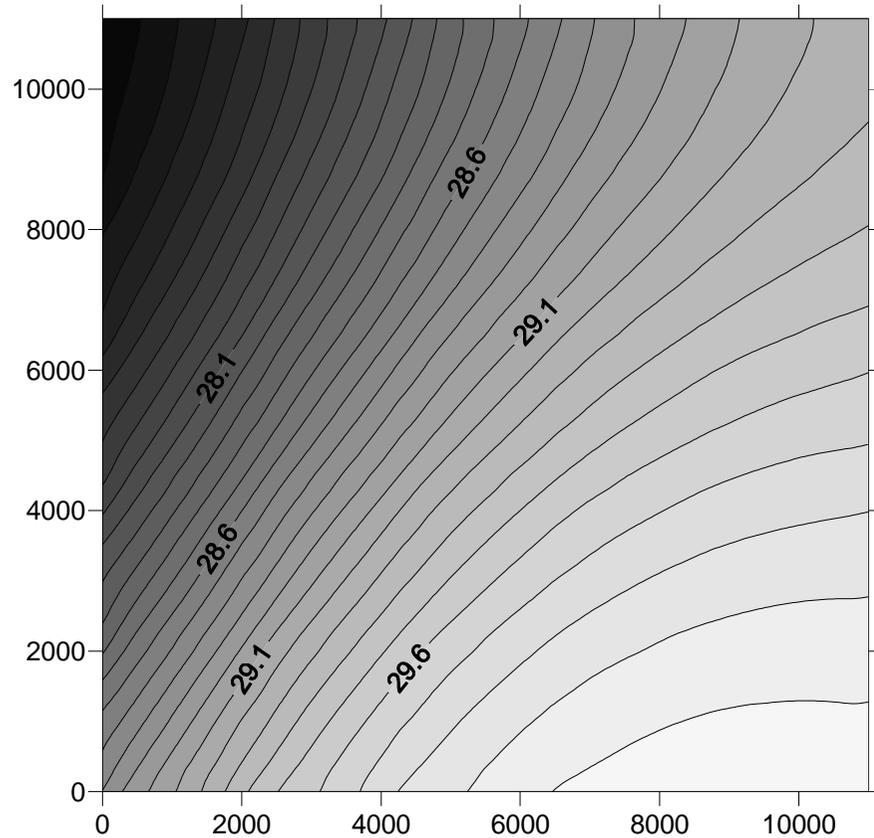
B) Misure

La Varianza associata alle misure è pari a $\sigma_0^2 = 0.05 \text{ °C}^2$

C) Scale Spaziali

Sulla base della distribuzione spaziale dei punti di misura, le Scale Spaziali considerate nelle 5 iterazioni del SCM di Barnes sono:

Iterazione 1:	4255 m
Iterazione 2:	2127 m
Iterazione 3:	1063 m
Iterazione 4:	531 m
Iterazione 5:	266 m

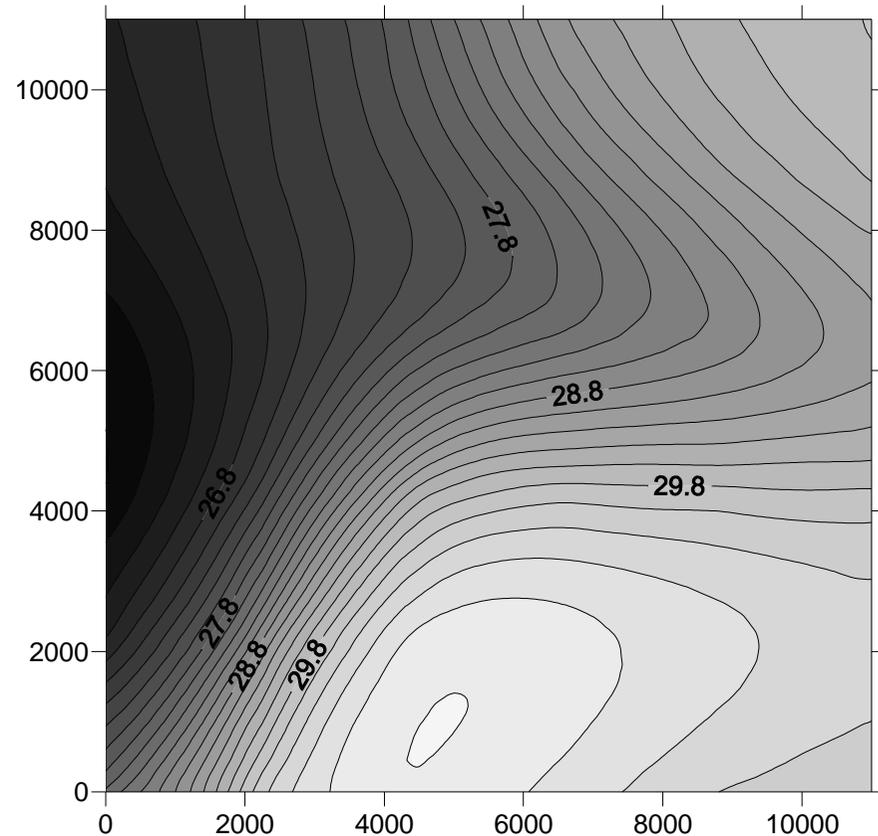


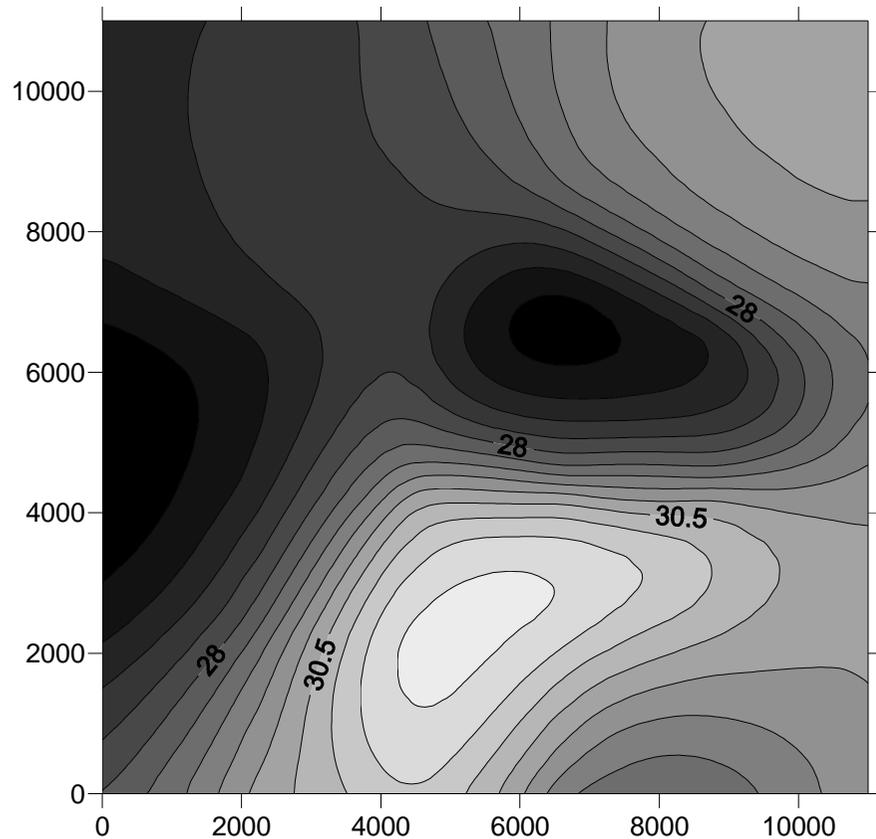
⇒ Iterazione 2

L'introduzione delle informazioni alla **Scala Spaziale di 2100 m** produce una distorsione del campo termico e lo trasforma da campo decrescente monotonicamente da NW a SE ad un campo decisamente più articolato con valori minimi a NE e SE.

⇒ Iterazione 1

Le informazioni alla **Scala Spaziale dell'ordine di 4200 m** conducono ad un campo di temperatura decrescente da NW a SE, cosa non immediatamente evidente dalle misure e che costituisce una caratteristica meteorologica a grande scala.



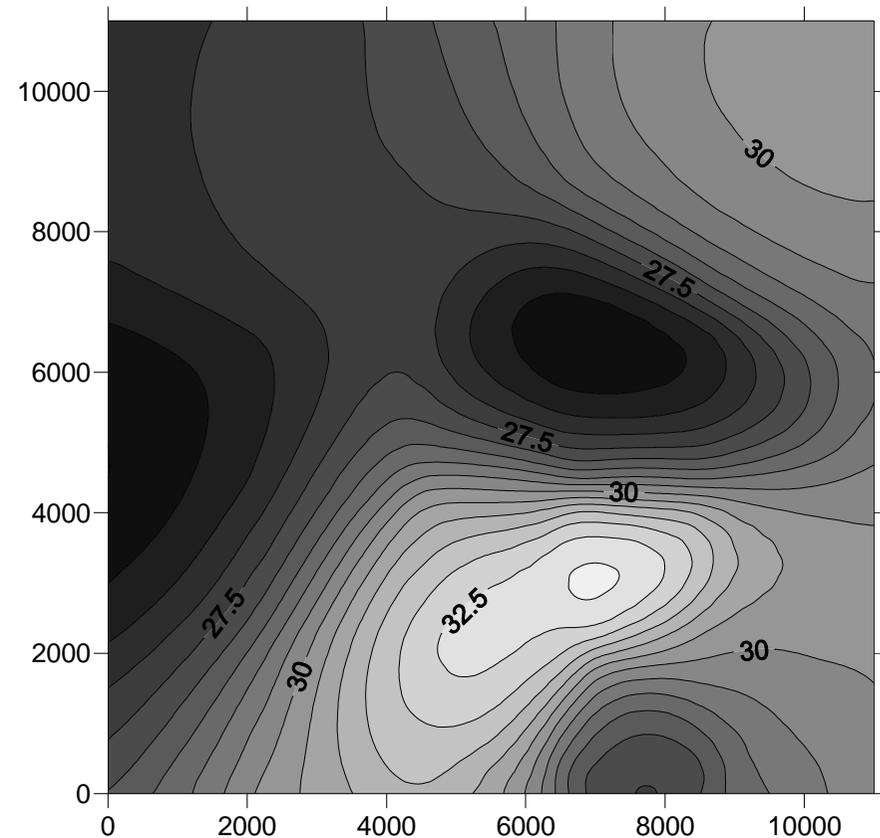


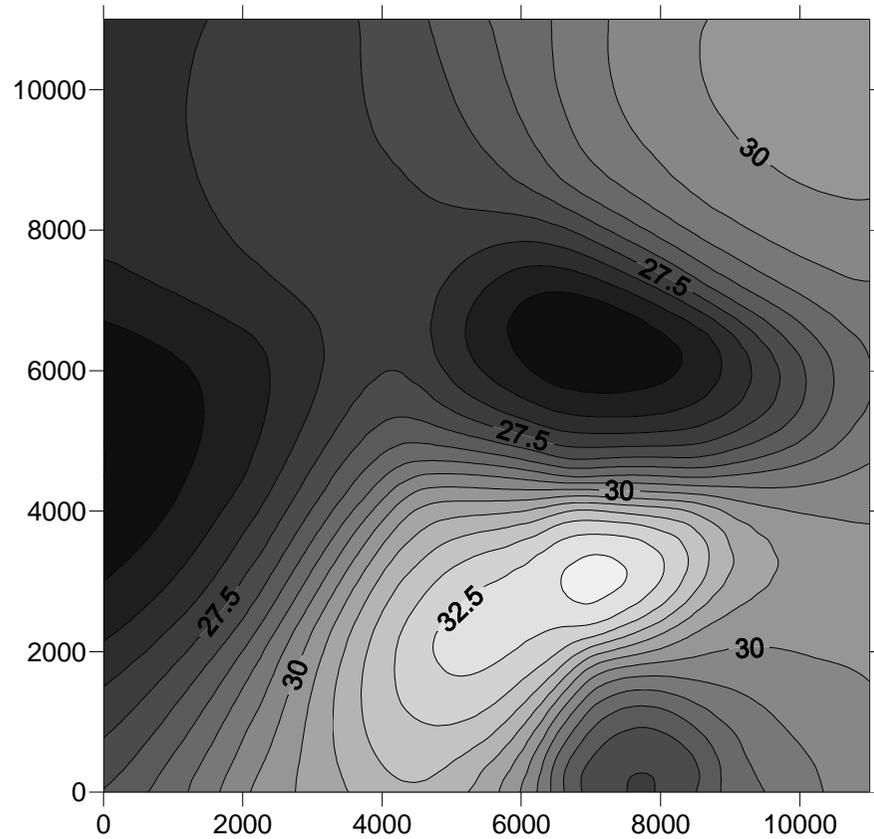
⇒ Iterazione 3

Le informazioni alla **Scala Spaziale di 1050 m** contenute nelle misure dettagliano meglio il promontorio di temperatura a E del dominio e la zona di minimo locale a Sud

⇒ Iterazione 4

L'unico effetto visibile legato all'introduzione nell'Analisi delle informazioni possedute dalle misure **alla Scala Spaziale di 530 m** è la descrizione più dettagliata del minimo locale nella parte S del dominio.



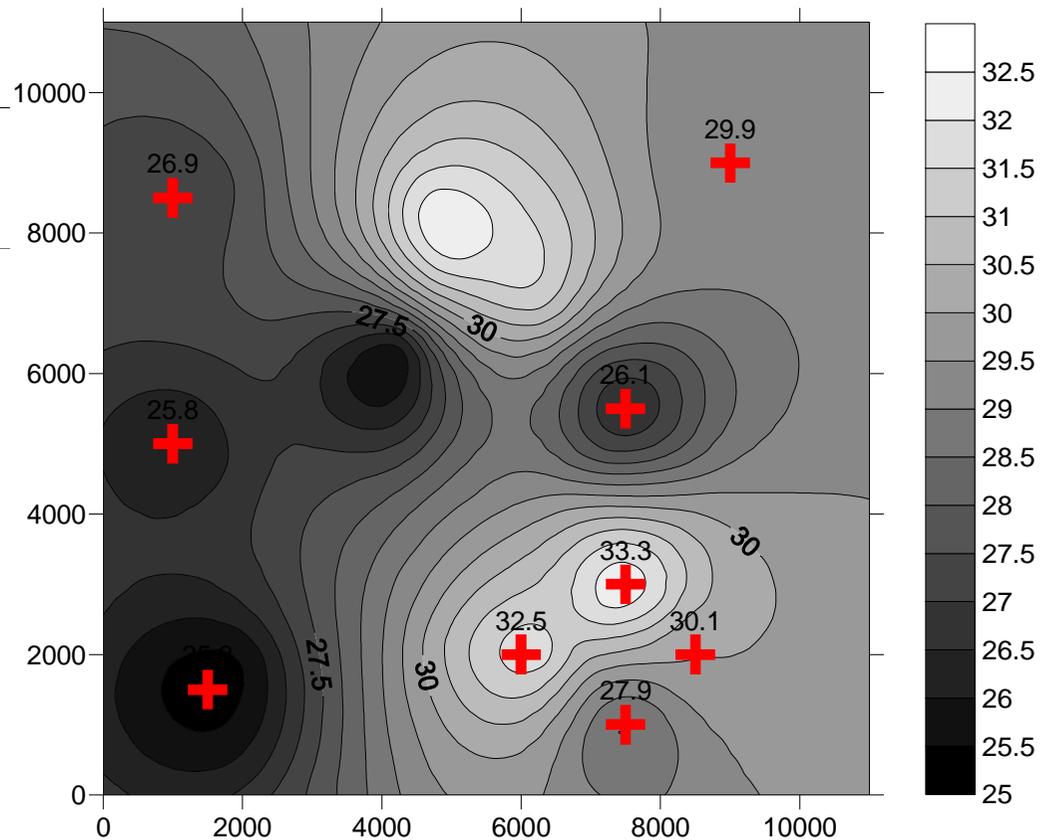


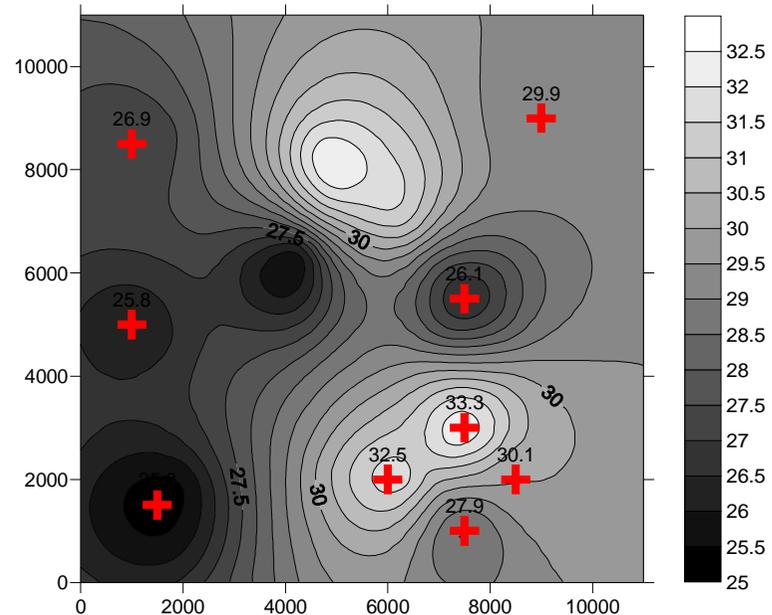
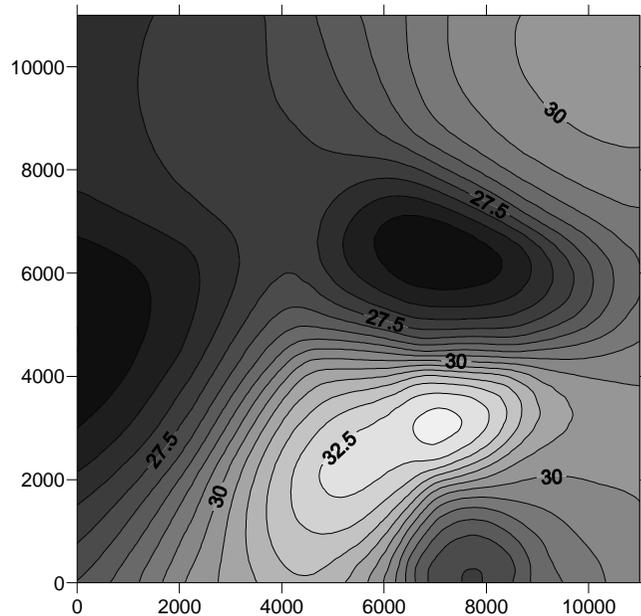
Questa è la realtà che l'Analista non poteva conoscere!

⇒ Iterazione 5

La procedura iterativa SCM di Barnes si conclude con il campo di temperatura riportato qui di fianco, molto simile a quello ottenuto all'iterazione precedente, ma profondamente differente dal campo iniziale.

Quanto è vicino alla realtà?





La ricostruzione fatta dal SCM (a sinistra) e la realtà (a destra) portano alle seguenti riflessioni:

1. La spazializzazione SCM, nonostante la mancanza totale di informazioni a priori, comunque ha condotto ad una **spazializzazione realistica** che cattura, almeno in prima approssimazione, l'essenziale della distribuzione;
2. **Ovviamente nella parte N del dominio l'assenza sia di misure che di informazioni a priori ha impedito al SCM di operare correttamente, come fatto nella parte S.**

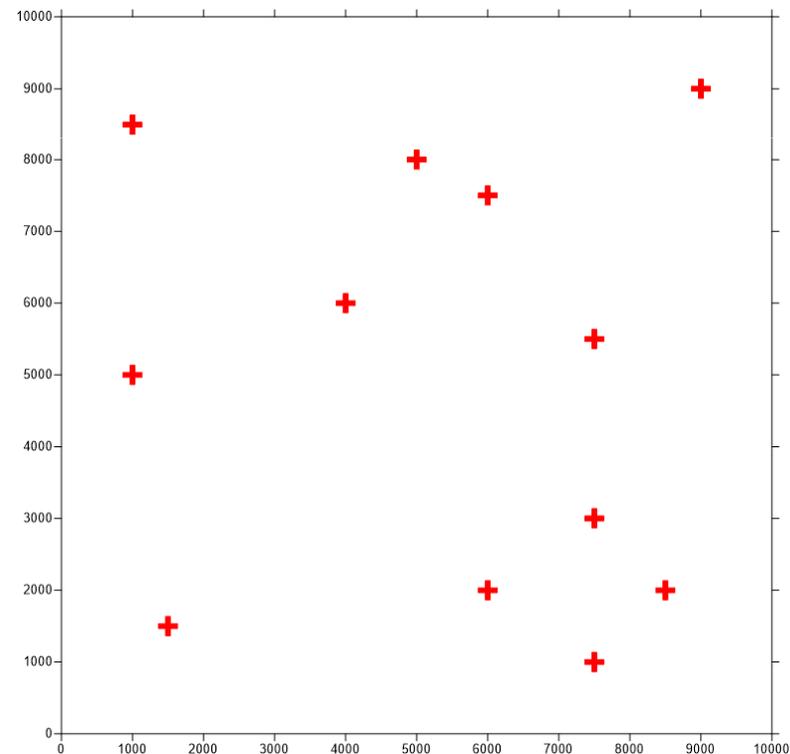
Le informazioni non si creano se non ci sono, ma si possono distruggere se male utilizzate.

ESEMPIO 3

Si hanno a disposizione per l'Analisi le stesse 9 stazioni di misura di temperatura a 2 m distribuite su un dominio spaziale di 10 x 10 km dell'Esempio 2 cui si aggiungono altre 3 stazioni di misura del tutto identiche poste nella parte N del dominio. In tutte le stazioni si impiega lo stesso termometro caratterizzato da un'incertezza intrinseca di 0.2 °C.

Quindi la **Varianza delle Misure** risulta pari a $\sigma_0^2 = 0.05 \text{ } ^\circ\text{C}^2$

X	Y	T(°C)
1500	1500	25,3
7500	1000	27,9
6000	2000	32,5
8500	2000	30,1
7500	3000	33,3
1000	5000	25,8
7500	5500	26,1
1000	8500	26,9
9000	9000	29,9
4000	6000	25,3
6000	7500	32,2
5000	8000	32,5



..... ma tutte queste cose l'Analista non le può sapere!

A) Campo di Background

L'Analista, non disponendo di informazioni a priori, anche in questo caso non può che utilizzare un campo di Background costante nello spazio e pari al valor medio delle temperature misurate. Quindi il Campo di Background risulta in ogni punto pari a 29 °C.

La Varianza di Background (pari al quadrato dell'errore standard associato alla media) è pari a $\sigma_b^2 = 0.89 \text{ °C}^2$

B) Misure

La Varianza associata alle misure è pari a $\sigma_0^2 = 0.05 \text{ °C}^2$

C) Scale Spaziali

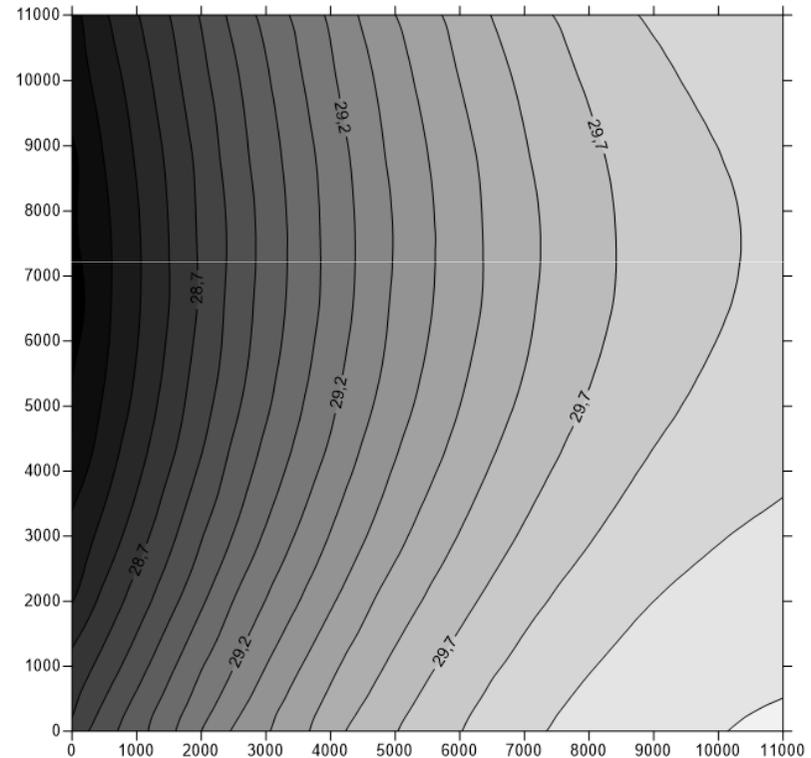
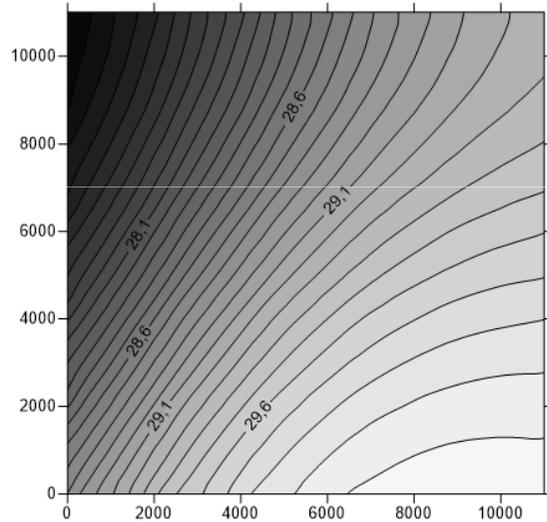
Sulla base della distribuzione spaziale dei punti di misura, le Scale Spaziali considerate nelle 5 iterazioni del SCM di Barnes sono:

Iterazione 1:	4255 m
Iterazione 2:	2127 m
Iterazione 3:	1063 m
Iterazione 4:	531 m
Iterazione 5:	266 m

⇒ **Iterazione 1**

La nuova ricostruzione (a destra) rispetto a quella dell'Esempio 2 (a sinistra)

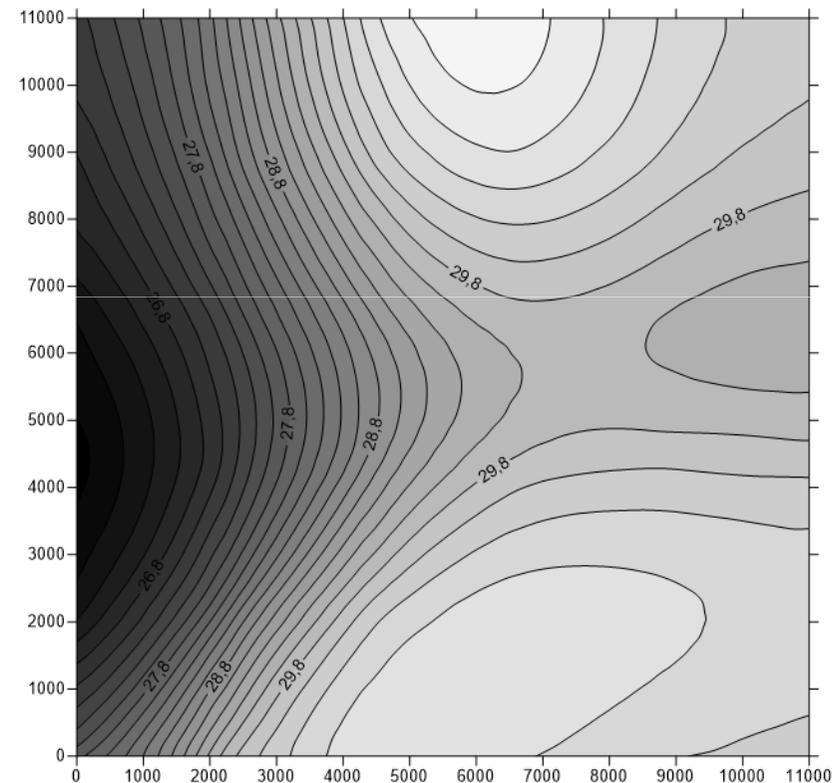
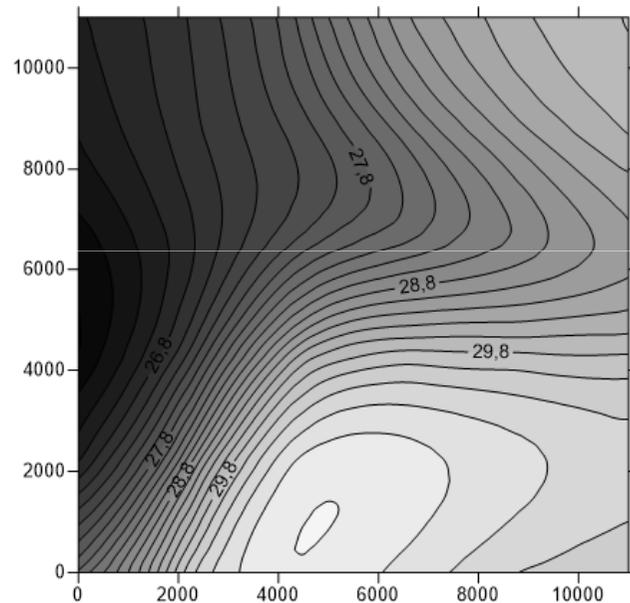
Pur essendoci ancora una tendenza a grande scala della temperatura ad aumentare da W a E, la presenza delle nuove postazioni di misura mostra fin da subito una struttura termica autonoma nel settore N.



⇒ Iterazione 2

La nuova ricostruzione (a destra) rispetto a quella dell'Esempio 2 (a sinistra)

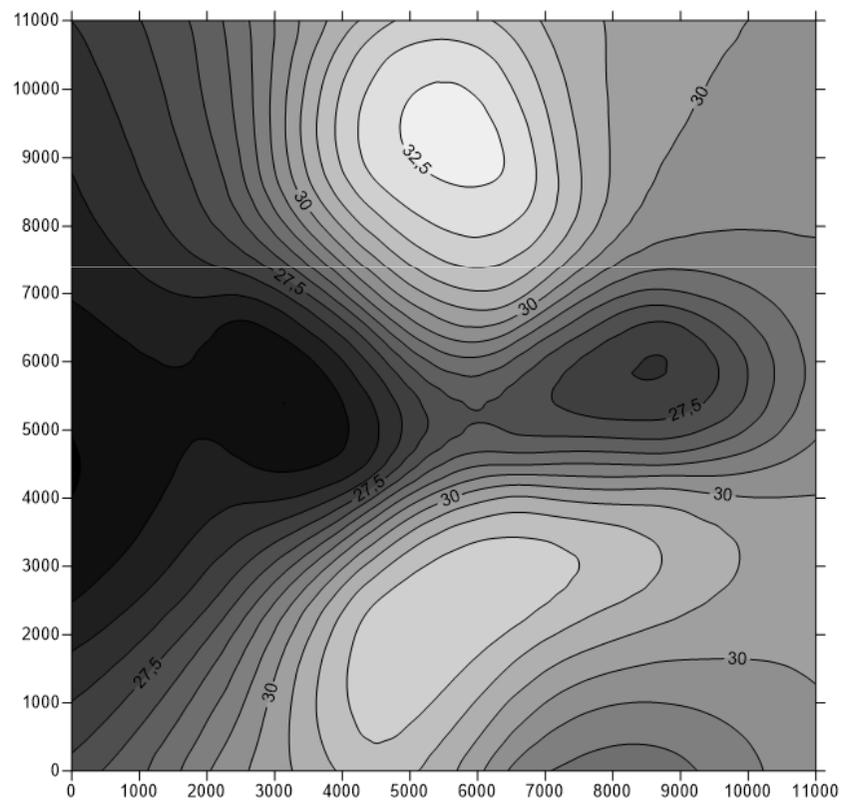
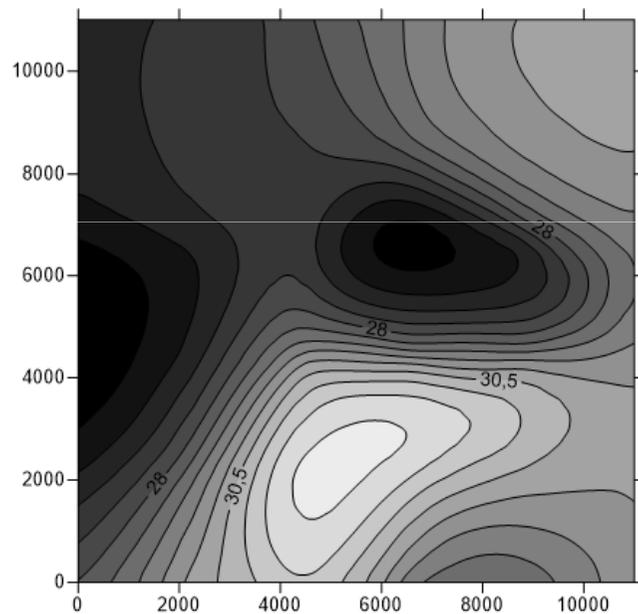
La nuova iterazione introduce informazioni di scala capaci di evidenziare la struttura termica nella parte N del dominio che presenta fin da subito una certa complessità.



⇒ Iterazione 3

La nuova ricostruzione (a destra) rispetto a quella dell'Esempio 2 (a sinistra)

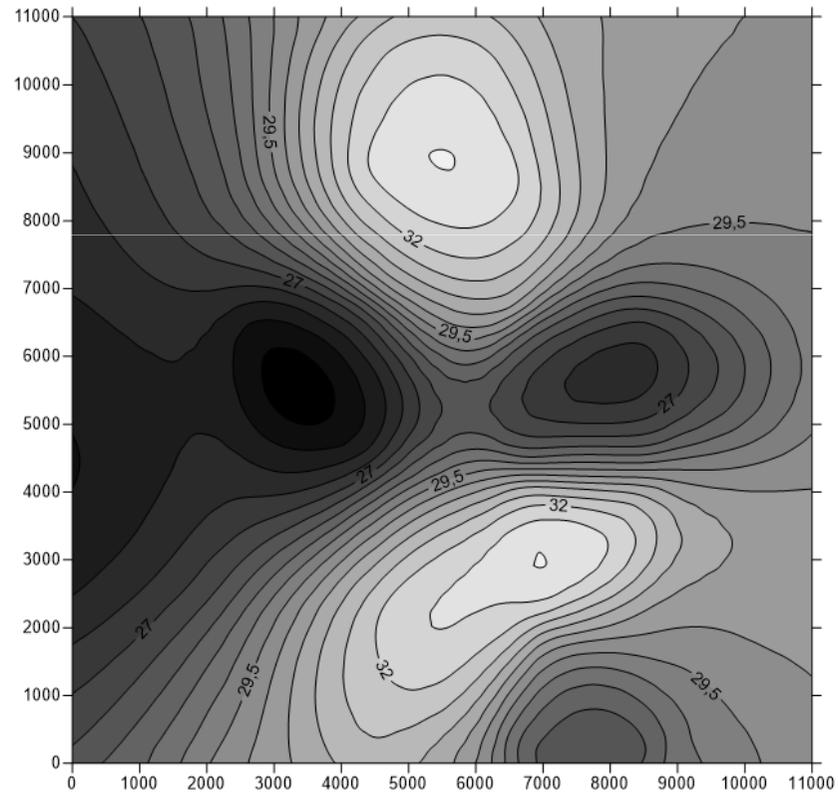
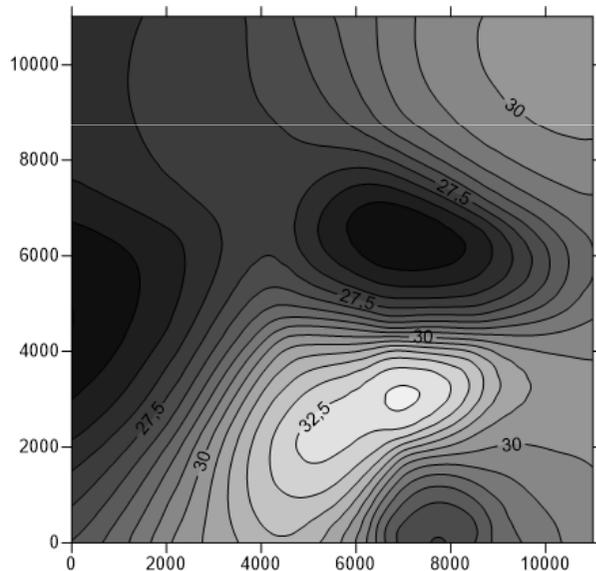
Già a questa iterazione i dettagli del massimo relativo di temperatura a N del dominio emergono con notevole dettaglio



⇒ Iterazione 4

La nuova ricostruzione (a destra) rispetto a quella dell'Esempio 2 (a sinistra)

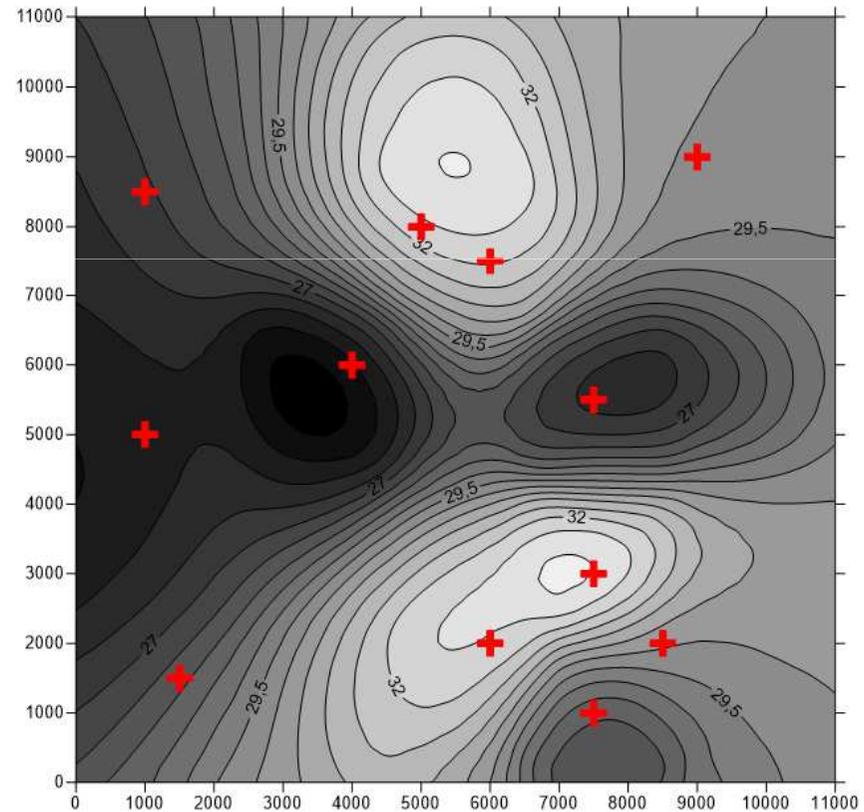
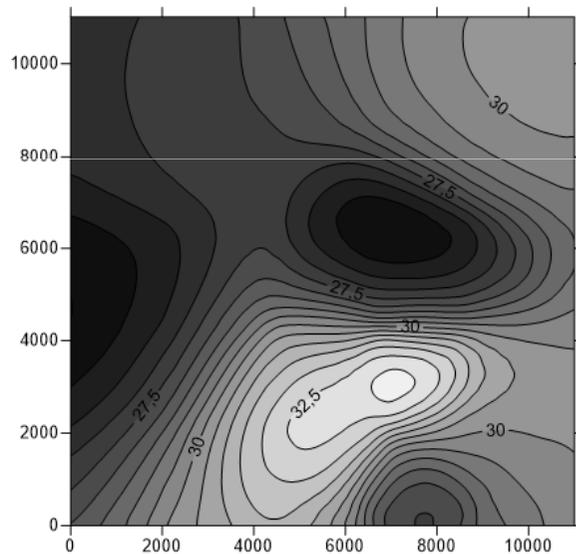
A questo punto la ricostruzione del massimo locale di temperatura nella parte N del dominio risulta pressoché completa e le differenze in questa parte del dominio con quanto ricostruito nell'Esempio 2 sono sostanziali.

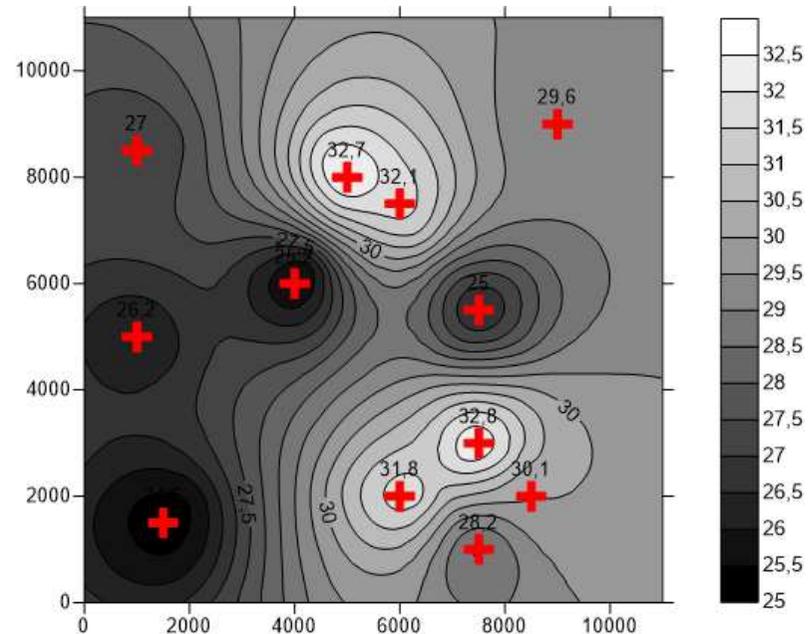
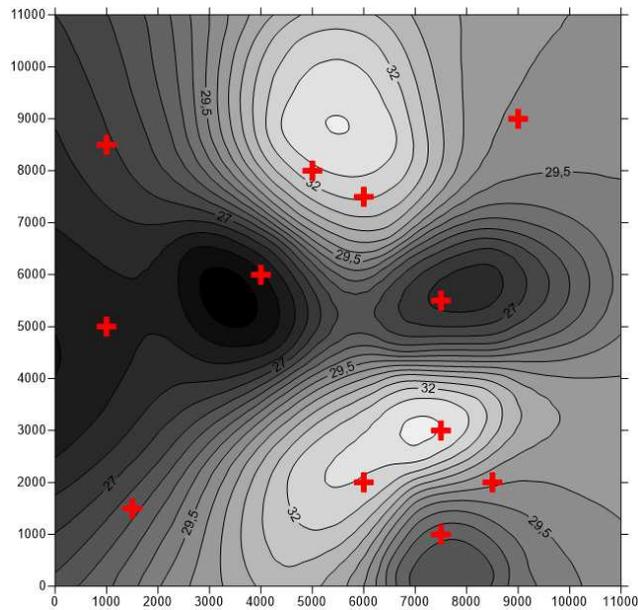


⇒ Iterazione 5

La nuova ricostruzione (a destra) rispetto a quella dell'Esempio 2 (a sinistra)

La distribuzione cui giunge l'algoritmo risulta molto differente (soprattutto nella parte N del dominio) rispetto alla distribuzione finale cui si giunge nell'Esempio 2. E ciò dipende esclusivamente dalla presenza di informazioni in questa parte del dominio.





La ricostruzione fatta dal *SCM* (a sinistra) e la realtà (a destra) portano alle seguenti riflessioni:

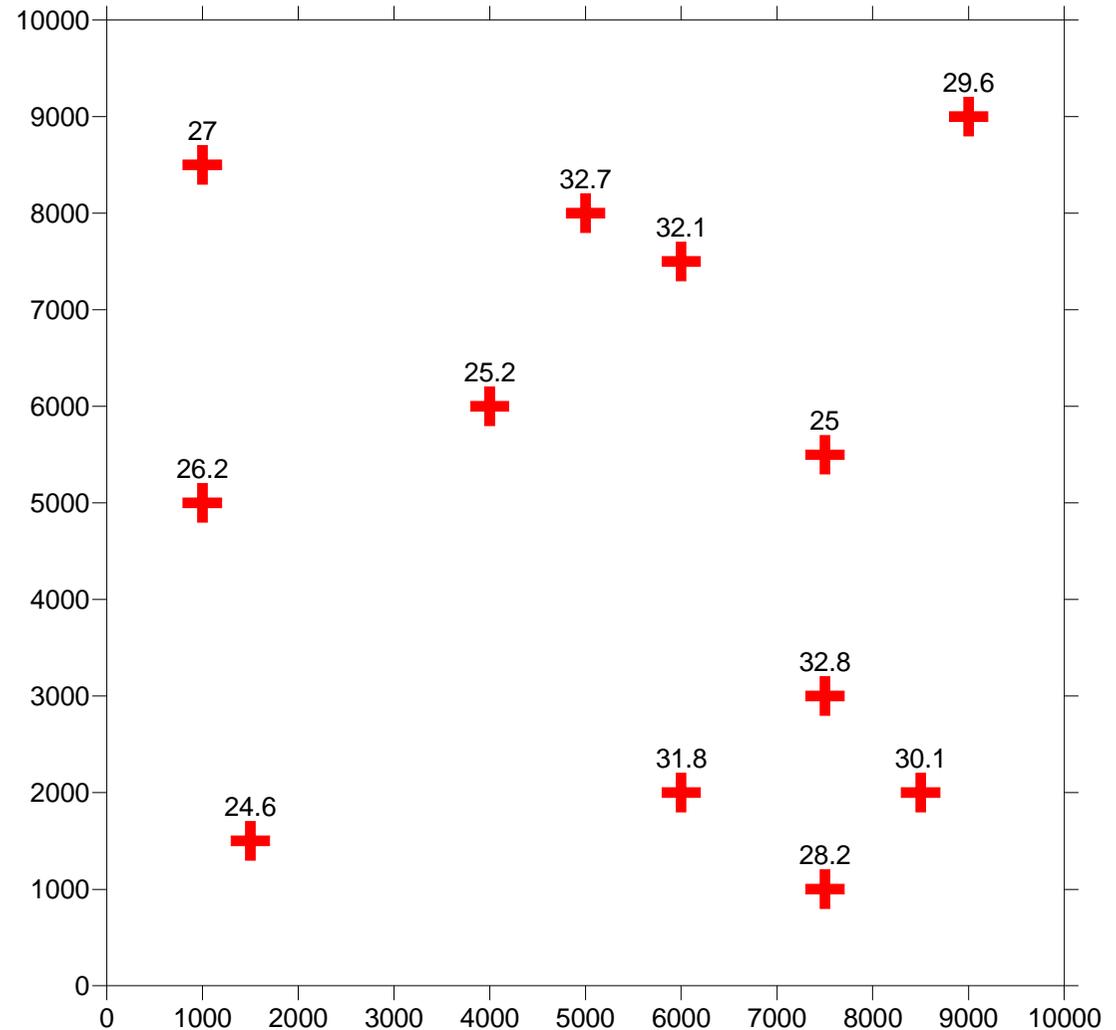
1. La spazializzazione *SCM*, nonostante la mancanza totale di informazioni a priori, comunque ha condotto ad una **spazializzazione realistica** che cattura la maggior parte della distribuzione reale;
2. La presenza delle postazioni di misura nella parte N del dominio consente anche in questa porzione del territorio una ricostruzione reale del fenomeno.

Come si nota, la chiave del successo di una ricostruzione sta nel disporre di un insieme di informazioni strumentali adeguate.

ESEMPIO 4

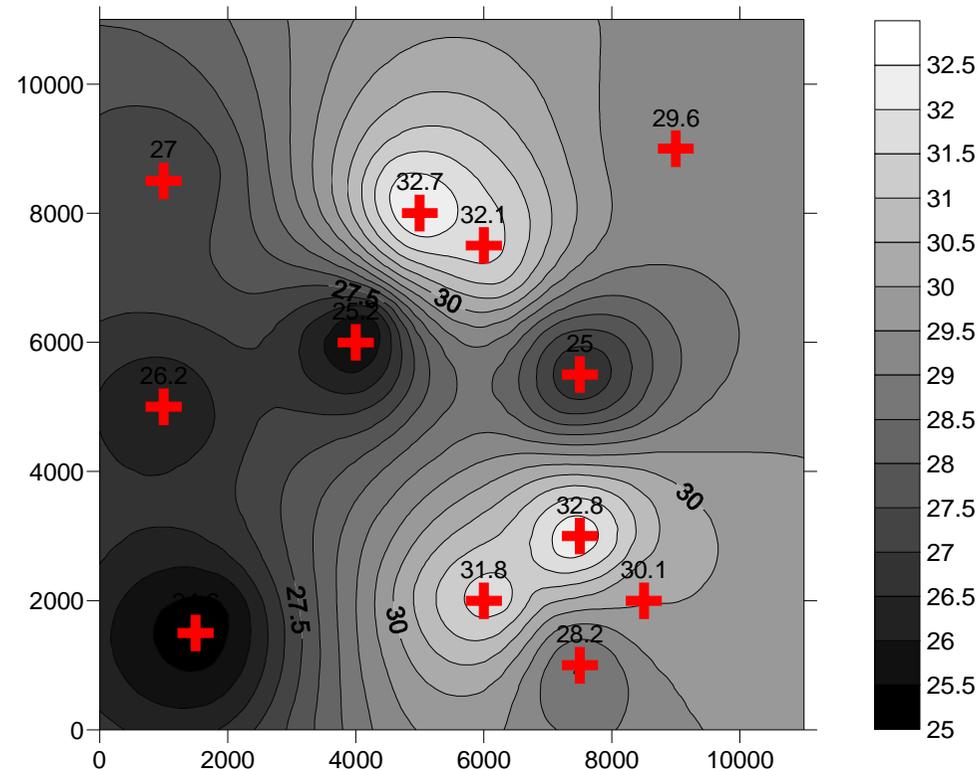
Si immagini che l'Analista disponga ora di 12 misure della temperatura a 2 metri dal suolo distribuite abbastanza uniformemente nel dominio di analisi di dimensioni 10 x 10 km.

X	Y	T(C)
1500	1500	24,6
7500	1000	28,2
6000	2000	31,8
8500	2000	30,1
7500	3000	32,8
1000	5000	26,2
7500	5500	25
4000	6000	25,2
6000	7500	32,1
5000	8000	32,7
1000	8500	27
9000	9000	29,6



Questa volta, però, l'analista dispone di una **immagine satellitare contemporanea alle misure**. Una volta riportata la temperatura radiativa vista dal satellite alla temperatura a 2 m, l'informazione di background disponibile a priori risulta essere quella qui sotto illustrata.

Quindi, invece di un valore di Background costante usato negli esempi precedenti, ora ad ogni nodo della griglia di Analisi abbiamo un valore di temperatura che sappiamo essere prossimo, ma concettualmente differente dalla temperatura a 2 m caratteristica del nodo.



L'applicazione del SCM di Barnes a questo caso comporta che:

1. La **Varianza di Background** la si debba ottenere dalle informazioni disponibili. In questo caso si è scelto di porla pari alla varianza tra valori di temperatura satellitari rilevati nei nodi in cui sono collocati i punti di misura e le misure a tali punti. Il risultato ottenuto è che $\sigma_b^2 = 0.15 \text{ } ^\circ\text{C}^2$
2. L'incertezza intrinseca dichiarata dei termometri impiegati nelle stazioni risulta pari a $0.25 \text{ } ^\circ\text{C}$, quindi la **Varianza delle Misure** è $\sigma_0^2 = 0.05 \text{ } ^\circ\text{C}^2$
3. Perciò **$\varepsilon^2 = 0.33$**
4. La disposizione delle stazioni di misura sul dominio di analisi è tale per cui:

Distanza massima stazioni = 10606 m

Distanza minima = 1118 m

e quindi le **Scale Spaziali** considerate dal SCM sono:

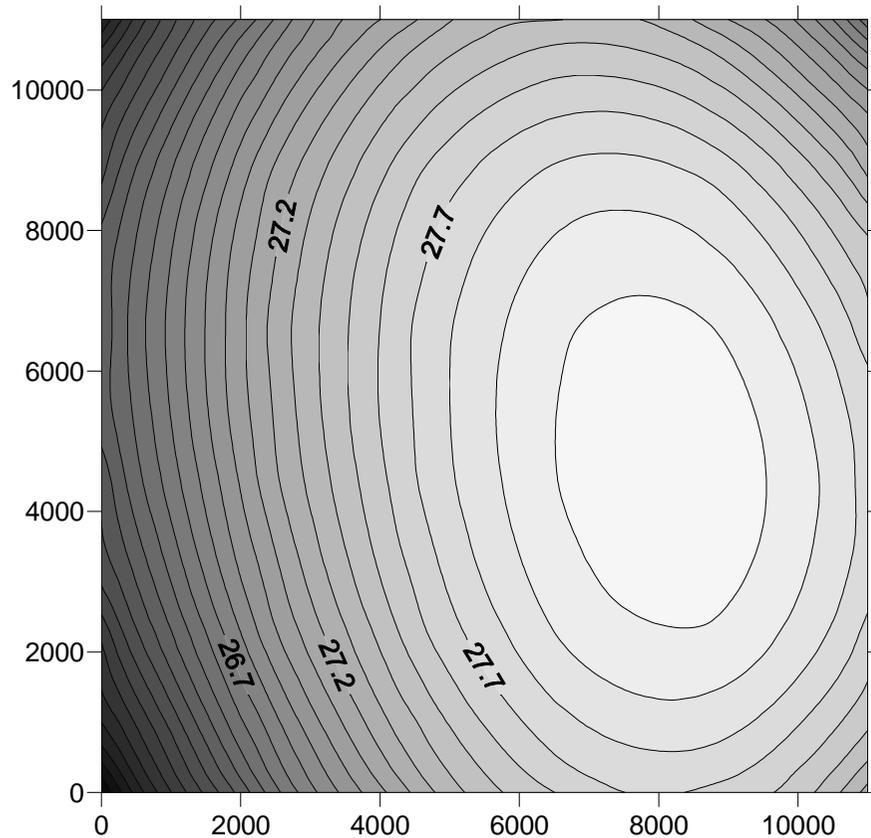
Iterazione 1: 4744 m

Iterazione 2: 2372 m

Iterazione 3: 1186 m

Iterazione 4: 593 m

Iterazione 5: 296 m

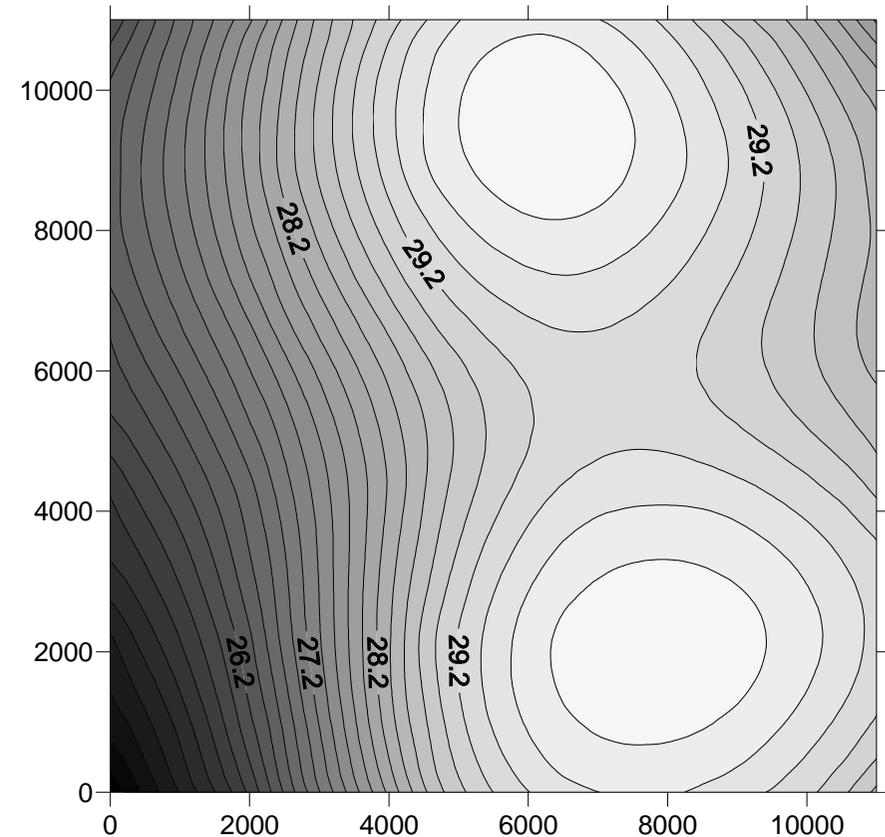


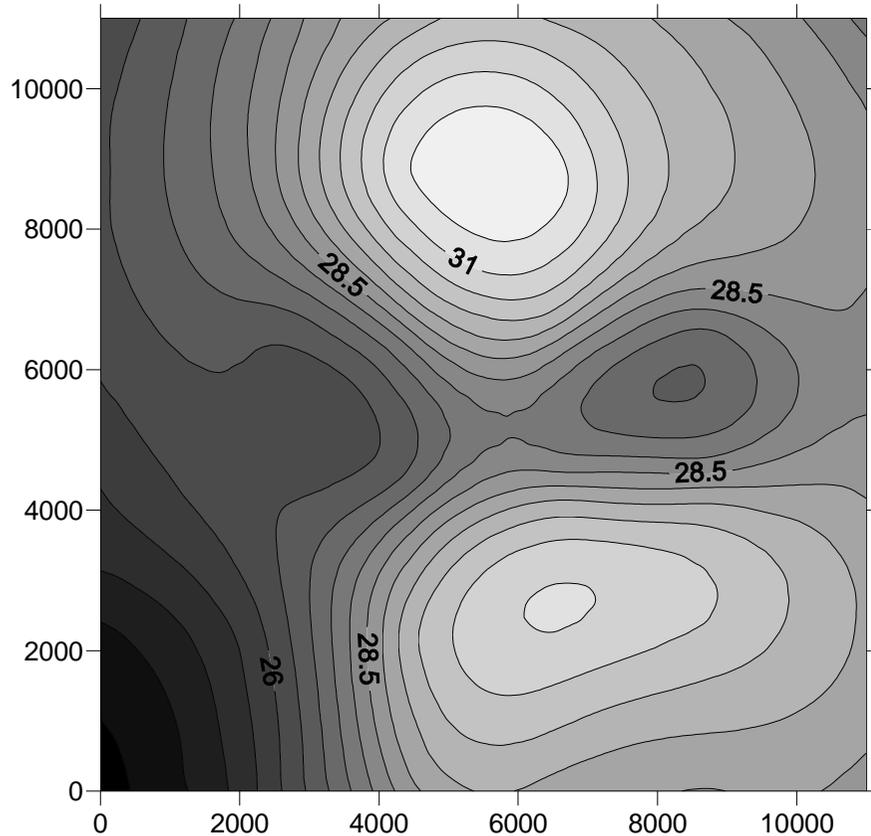
⇒ Iterazione 1

Le informazioni alla **Scala Spaziale di 5000 m** contenute nelle misure con il campo di background, consentono di ottenere subito una distribuzione di temperatura non uniforme, ma che evidenzia come nella parte E del dominio la temperatura presenti valori maggiori

⇒ Iterazione 2

Le informazioni alla Scala Spaziale di circa **2500 m** iniziano a far vedere i primi dettagli del campo di temperatura nella zona E del dominio



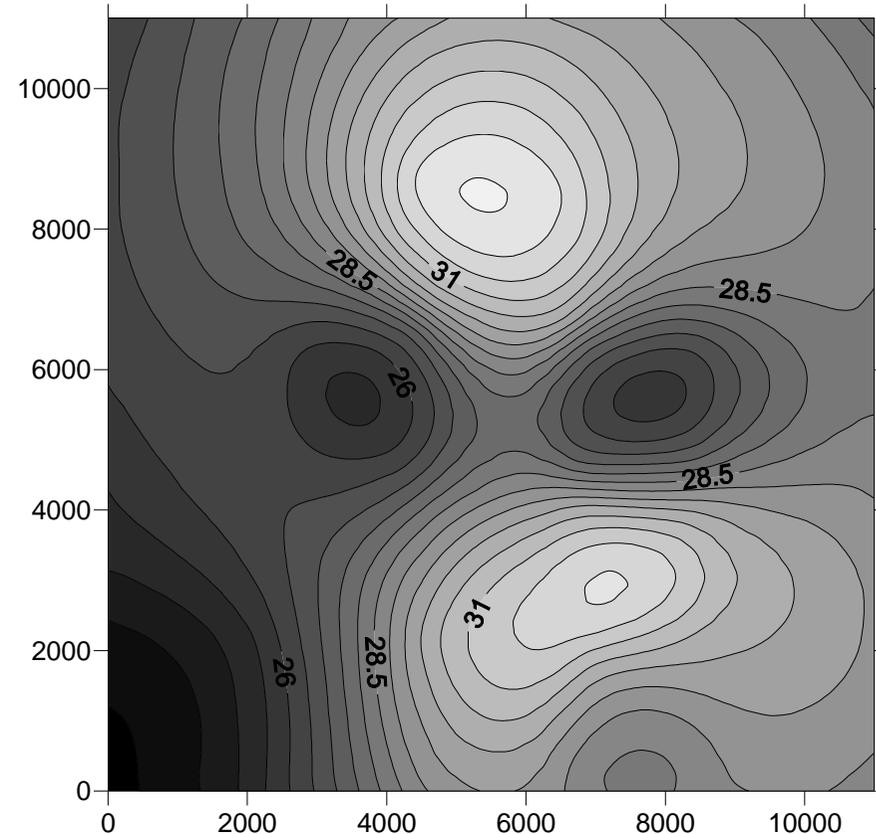


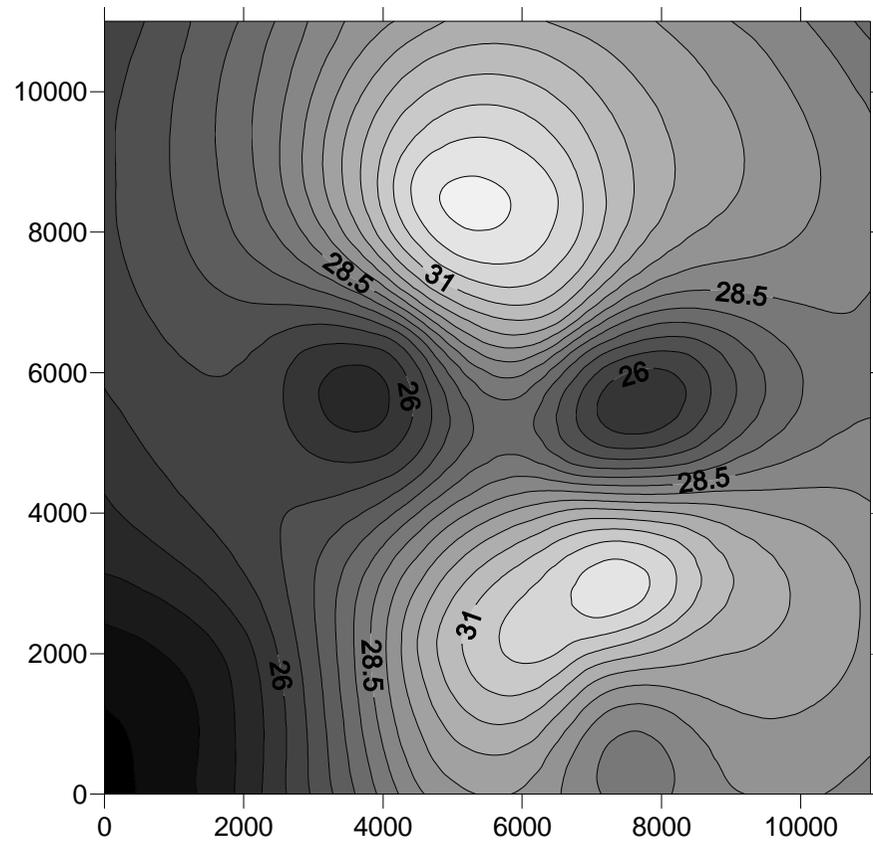
⇒ Iterazione 3

Le informazioni alla **Scala Spaziale di circa 1200 m** contenute nelle misure fanno emergere dettagli interessanti del campo di temperatura in cui si distinguono chiaramente due zone di massimo locale e due di minimo locale.

⇒ Iterazione 4

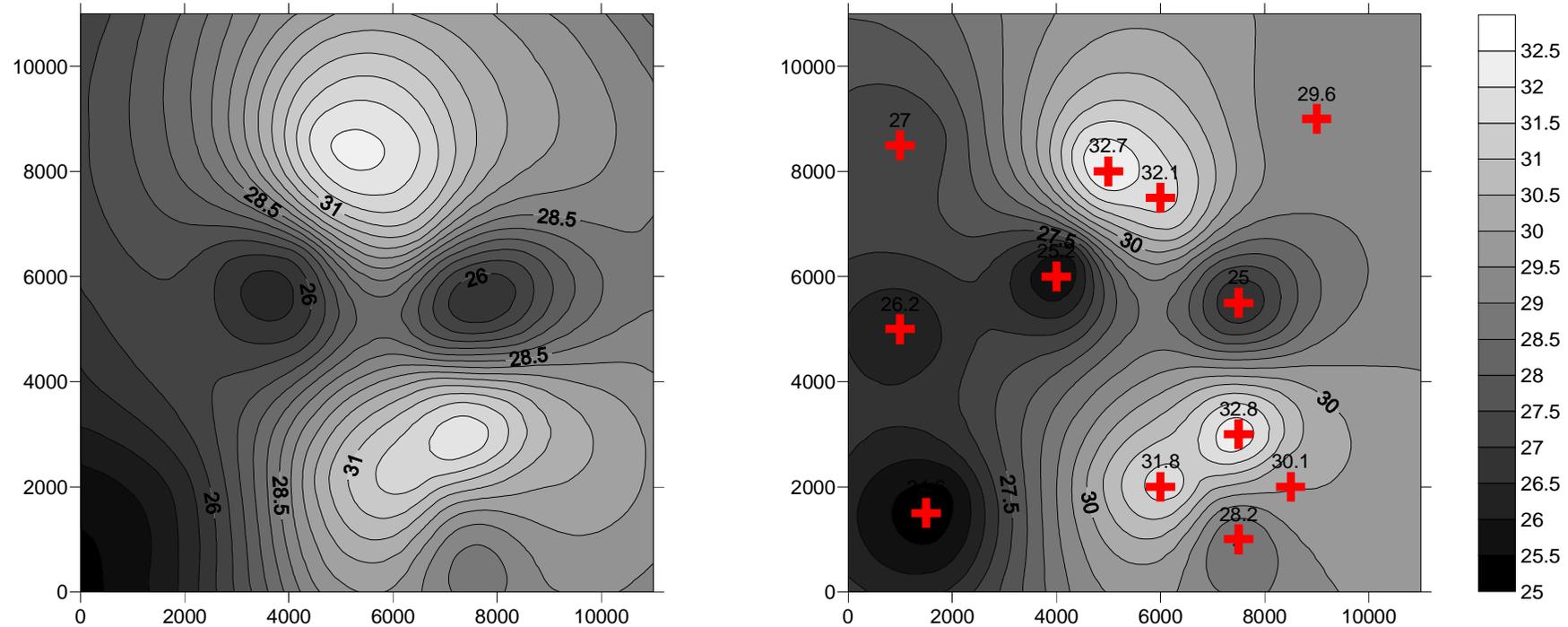
Le informazioni alla **Scala Spaziale di circa 600 m** sorprendentemente, pur mantenendo due zone di massimo locale evidenziano una zona di minimo locale in più rispetto all'iterazione precedente, oltre che mostrare maggiori dettagli nelle varie parti del campo.





⇒ **Iterazione 5**

L'ultima iterazione, in cui vengono inseriti i dettagli relativi alla **Scala Spaziale di circa 300 m** conferma a grandi linee il campo individuato all'iterazione precedente.



Sopra a sinistra è stato riportato il risultato finale del SCM di Barnes con 12 punti di misura e il campo di Background mostrato sopra a destra e che è un'elaborazione di un'immagine satellitare all'infrarosso. **È innegabile che le due distribuzioni mostrano la stessa realtà fisica.**

L'immagine satellitare non è detto che sia la realtà, visto il processo di inversione e di elaborazione necessario a trasformare la radianza in temperatura a 2 m.

Sicuramente il risultato del SCM è il meglio che si possa fare con i SCM avendo a disposizione misure ben distribuite e un campo di Background credibile.

4.2 OPTIMAL INTERPOLATION

Questa tecnica di spazializzazione ha come base un modello statistico lineare che vede l'Analisi in un punto di una generica **grandezza Q** come:

La Combinazione Lineare, con pesi opportuni, delle misure disponibili.

Fin qui nulla di differente dai precedenti metodi di spazializzazione. In realtà, tutta la potenza del metodo sta proprio nei pesi che vengono determinati su solide basi statistiche.

Il modello lineare, in corrispondenza di un generico nodo i -esimo della griglia di Analisi, determina l'Analisi $Q_{a,i}$ che inevitabilmente differirà dal Valore Vero Q .

Quindi \rightarrow Errore di Analisi $\varepsilon_{a,i}$

$$\varepsilon_{a,i} = Q_{a,i} - Q$$

I pesi vengono determinati in modo tale che, complessivamente, lo scarto quadratico medio degli Errori di Analisi sia il minimo possibile. (Modello Statistico BLUE).

Il metodo è molto potente, ma lo **sviluppo analitico dei calcoli** e l'ottenimento delle varie relazioni che lo caratterizzano è **piuttosto laborioso**, richiedendo una conoscenza approfondita di Analisi Vettoriale, Algebra Lineare e Analisi Differenziale in forma matriciale.

Qui di seguito introdurremo l'argomento usando come semplificazione introduttiva due esempi proposti da E. Kalnay nel suo libro, in modo da comprendere, al di là dei formalismi matematici, l'essenza del problema.

Inoltre si eviterà di fornire le dimostrazioni delle relazioni che costituiscono l'essenza del metodo Optimal Interpolation (O.I.)

Primo passo

Consideriamo una **tipica proprietà termodinamica dell'aria** \Rightarrow **Temperatura T**

T varia nello spazio e nel tempo \Rightarrow **Campo Spaziale variabile nel tempo $T(x,y,z;t)$**

In un punto $P(x,y)$ nei pressi del suolo (a 2 m) si misuri ad un istante t la temperatura media oraria con due strumenti diversi $\Rightarrow T_1, T_2$.

Dalla Teoria dell'Errore non abbiamo dubbi che:

- i due valori di temperatura misurati non coincidono con la **temperatura vera** T_v che è:

$$T_v = T(x,y,2;t)$$

- la temperatura vera T_v è inaccessibile strumentalmente, anche se è l'obiettivo vero delle misure che si sono realizzate.

Essendo inaccessibile, il valore vero T_v lo dobbiamo stimare dalle informazioni sperimentalmente disponibili, cioè da T_1 e T_2 e tale stima la indichiamo col nome di Analisi e la indichiamo come T_a .

E' un dato di fatto che entrambe le temperature misurate T_1 e T_2 siano parenti di T_v ma differiranno dal valore vero per un errore, quindi:

$$\begin{aligned}T_1 &= T_v + \varepsilon_1 \\T_2 &= T_v + \varepsilon_2\end{aligned}$$

Se immaginiamo di ripetere N volte la misura di T_1 e di T_2 , otterremmo ogni volta una coppia di errori ε_1 e ε_2 **sempre diversi**.

Ciò che la Teoria della Misura ci garantisce, se le misure non sono affette da un errore sistematico, è che

..... sia ε_1 che ε_2 sono due **Variabili Stocastiche** distribuite, ciascuna, secondo una Funzione di Densità di Probabilità Gaussiana.

In particolare

- il valore medio di ε_1 e il valore medio di ε_2 (cioè i rispettivi *Valori Attesi*) sono entrambi nulli, quindi **le due Gaussiane che descrivono statisticamente i due errori di misura sono entrambe a media nulla;**

$$E(\varepsilon_1) = \overline{\varepsilon_1} = 0$$

$$E(\varepsilon_2) = \overline{\varepsilon_2} = 0$$

- la **varianza** σ_1^2 e σ_2^2 dei due errori **non è nulla**, quindi:

$$E([\varepsilon_1 - E(\varepsilon_1)]^2) = \overline{(\varepsilon_1 - \overline{\varepsilon_1})^2} = \overline{\varepsilon_1^2} = \sigma_1^2$$

$$E([\varepsilon_2 - E(\varepsilon_2)]^2) = \overline{(\varepsilon_2 - \overline{\varepsilon_2})^2} = \overline{\varepsilon_2^2} = \sigma_2^2$$

- le due misure T_1 e T_2 sono tra loro **indipendenti** e non interferenti tra loro, quindi i due errori non possono avere alcuna parentela. Ciò comporta che gli *errori* siano tra loro *indipendenti* e quindi che:

$$E(\varepsilon_1, \varepsilon_2) = \overline{([\varepsilon_1 - \overline{\varepsilon_1}] \cdot [\varepsilon_2 - \overline{\varepsilon_2}])} = Cov(\varepsilon_1, \varepsilon_2) = 0$$

Quindi gli ingredienti conoscitivi di cui disponiamo sono dunque:

- ⇒ le due **misure** T_1 e T_2 la cui conoscenza deriva direttamente dalle misure realizzate;
- ⇒ gli **errori di misura** ε_1 e ε_2 che sono variabili stocastiche di tipo Gaussiano completamente caratterizzate da un valore medio nullo degli errori e dalle due standard deviation σ_1 e σ_2 che possiamo stimare, per esempio, in laboratorio o da reiterate misure in campo con strumenti campione, comunque seguendo la norma UNI CEI ENV 13005;
- ⇒ i diversi **errori di misura** sono **tra loro indipendenti** e quindi la **Covarianza tra gli errori di misura relativi alle due misure risulta nulla a priori**.

Sulla base solo di queste informazioni vogliamo ottenere la migliore stima possibile T_a (cioè la migliore Analisi possibile) della temperatura vera T_v .

Per fare ciò, formuliamo il seguente **Problema BLUE** (*Best Linear Unbiased Estimation*):

⇒ La **stima** del valore vero, cioè l'**analisi** T_a , la si ottiene con un modello lineare del tipo

$$T_a = a \cdot T_1 + b \cdot T_2$$

dove a e b sono coefficienti numerici che **pesano** le due misure disponibili (cioè l'informazione strumentale disponibile) e che dovranno tener conto delle informazioni statistiche note (gli errori sono tra loro statisticamente indipendenti e la distribuzione Gaussiana che li caratterizza ha media nulla e varianza σ_1^2 e σ_2^2 rispettivamente);

⇒ Il **valore atteso** (cioè il valore medio) dell'**analisi** T_a (che è una variabile stocastica, essendo la combinazione lineare di variabili stocastiche) **deve coincidere col valore vero** T_v (*stimatore unbiased*)

⇒ La **varianza** σ_a dell'**analisi** deve essere la minima possibile.

Primo requisito

Il **valore medio dell'analisi T_a** deve coincidere col **valore vero T_v** .

Sviluppiamo i conti:

$$E(T_a) = \overline{T_a} = T_v$$

$$\begin{aligned}\overline{T_a} &= \overline{a \cdot T_1 + b \cdot T_2} = \overline{a \cdot (T_v + \varepsilon_1) + b \cdot (T_v + \varepsilon_2)} = \\ &= a\overline{T_v} + a\overline{\varepsilon_1} + b\overline{T_v} + b\overline{\varepsilon_2} = \\ &= a \cdot T_v + b \cdot T_v = T_v\end{aligned}$$

Quindi, **perché ciò sia vero**, è necessario che **tra i due pesi a e b valga la relazione seguente**:

$$a + b = 1$$

Secondo requisito

Il **varianza dell'analisi σ_a^2** deve essere la minima possibile.

Ricordando che la somma dei due pesi è unitaria, si ha che:

$$\begin{aligned}\sigma_a^2(a,b) &= \overline{(T_a - T_v)^2} = \overline{[(a \cdot T_1 + b \cdot T_2) - T_v]^2} = \\ &= \overline{[a \cdot T_v + a \cdot \varepsilon_1 + (1-b) \cdot T_v + (1-b) \cdot \varepsilon_2]^2} = \\ &= \overline{a^2 \cdot \varepsilon_1^2 + (1-b)^2 \cdot \varepsilon_2^2} = \\ &= a^2 \sigma_1^2 + (1-b)^2 \sigma_2^2 = \sigma_a^2(a)\end{aligned}$$

Perché la varianza dell'analisi sia minima è necessario (ed anche sufficiente visto che stiamo trattando un Problema ai Minimi Quadrati che ha un funzionale da minimizzare sicuramente convesso) **che la sua derivata rispetto al peso a sia nulla**, cioè che:

$$\frac{d\sigma_a^2}{da} = 2a\sigma_1^2 + 2a\sigma_2^2 - 2\sigma_2^2 = 0$$

Ricordando che la somma dei pesi a e b è unitaria (è ciò garantisce la natura unbiased dell'algoritmo) ed azzerando la derivata si ottiene l'espressione analitica per i pesi ottimali a minima varianza di analisi:

$$a = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

$$b = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Come si nota immediatamente, le relazioni che esprimono analiticamente i pesi ottimali dipendono esclusivamente dalla statistica degli errori di misura dei due sensori, sintetizzati nelle due rispettive varianze degli errori.

Ricapitolando

Il **valore** ottimale **dell'analisi** T_a (cioè quello che più avvicina l'analisi al valore vero) lo si ottiene pesando linearmente le due misure T_1 e T_2 disponibili in modo tale che:

$$a = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

$$b = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

E' immediato verificare che vale la relazione seguente:

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Se definiamo **precisione di analisi** $1/\sigma_a^2$ ed analogamente **precisione delle misure** rispettivamente i rapporti $1/\sigma_1^2$ e $1/\sigma_2^2$, allora si può asserire che:

la precisione dell'analisi risulta pari alla somma delle precisioni delle misure che l'hanno determinata.

Un breve Esempio

- La **temperatura T_1** la rileviamo con un termometro (per esempio una PT100) la cui incertezza intrinseca dichiarata risulta pari a $0.5 \text{ }^\circ\text{C} \rightarrow \sigma_1^2 = 0.25 \text{ }^\circ\text{C}^2$
- La **temperatura T_2** la rileviamo con un anemometro-termometro ultrasonico la cui incertezza intrinseca dichiarata (nella rilevazione della temperatura) risulta pari a $0.05 \text{ }^\circ\text{C} \rightarrow \sigma_2^2 = 0.25 \cdot 10^{-2} \text{ }^\circ\text{C}^2$

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{0.25 \cdot 10^{-2}}{0.25 + 0.25 \cdot 10^{-2}} = 0.0990$$

$$a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{0.25}{0.25 + 0.25 \cdot 10^{-2}} = 0.9901$$

$$T_a = a_1 \cdot T_1 + a_2 \cdot T_2 = 0.0990 \cdot 27.5 + 0.9901 \cdot 26.9 = 26.91$$

Precisione

È immediato vedere che:

- Precisione di $T_1 = 4 \text{ }^\circ\text{C}^{-2}$
- Precisione di $T_2 = 400 \text{ }^\circ\text{C}^{-2}$
- Precisione di $T_a = 404 \text{ }^\circ\text{C}^{-2}$

Per cercare di introdurre gradualmente il *misterioso* simbolismo Vettoriale, possiamo riformulare brevemente tutto quanto detto introducendo Matrici e Vettori.

Si può raggruppare l'insieme delle misure T_1 e T_2 , l'insieme dei pesi a e b e l'Analisi T_a rispettivamente nel **Vettore delle Misure** $[y_0]$, nella **Matrice dei pesi** $[\alpha]$ e nel **Vettore dell'Analisi** $[x_a]$ dati rispettivamente da:

$$\boxed{[x_a] = [T_a]} \quad \boxed{[\alpha] = \begin{bmatrix} a \\ b \end{bmatrix}} \quad \boxed{[y_0] = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}}$$

Quindi, la relazione vettoriale che esprime l'Analisi diventa:

$$\boxed{[x_a] = [\alpha]^T \cdot [y_0]}$$

La Matrice dei pesi in realtà in questo semplice esempio si riduce ad un vettore. In generale la matrice $[\alpha]$ ha tante colonne quante sono i punti in cui si realizza l'Analisi (nel nostro caso 1) e tante righe quante sono le informazioni strumentali disponibili (nel nostro caso 2). Inoltre, si noti che in questo semplice esempio il vettore dell'Analisi possiede un solo elemento.

Secondo Passo

Consideriamo una tipica proprietà termodinamica dell'aria \Rightarrow **Temperatura T**

T varia nello spazio e nel tempo \Rightarrow **Campo Spaziale variabile nel tempo $T(x,y,z;t)$**

Ci poniamo in un punto $P(x,y)$ nei pressi del suolo (a 2 m) e misuriamo ad un istante t la temperatura media oraria T_0 con uno strumento.

Abbiamo una **coscienza a priori** del campo termico $T(x,y,z;t)$, per esempio fornita da un modello di previsione meteorologica (come WRF), oppure ipotizziamo che la temperatura nel punto sia in qualche modo parente del valore medio delle temperature misurate in un insieme di punti circostanti. Comunque sia, questa misura di temperatura derivante da una **conoscenza a priori** la chiamiamo **valore di background** a cui associamo un errore ε_b .

T_0 misura caratterizzata da un errore ε_0

T_b valore di background caratterizzato da un errore ε_b .

A partire dall'unica misura di temperatura disponibili e da una sua conoscenza modellistica, vogliamo ottenere la miglior stima (quella che meno differirà dalla realtà) della temperatura in quel punto.

Quello che abbiamo appena formulato è di fatto lo stesso problema del passo precedente.

Basta fare le sostituzioni seguenti:

$$\begin{array}{l} T_1 \Rightarrow T_b \\ T_2 \Rightarrow T_0 \end{array} \qquad \begin{array}{l} \varepsilon_1 \Rightarrow \varepsilon_b \\ \varepsilon_2 \Rightarrow \varepsilon_0 \end{array}$$

Quindi, il **valore dell'analisi** T_a , cioè quel **valore che statisticamente meno si discosta dal valore vero**, lo si ottiene usando le medesime relazioni viste prima in cui vengono fatte le sostituzioni di cui sopra.

In pratica si ha che:

$$T_a = a \cdot T_b + b \cdot T_0$$

$$T_a = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_b^2} \cdot T_b + \frac{\sigma_b^2}{\sigma_0^2 + \sigma_b^2} \cdot T_0$$

Riorganizzando opportunamente i termini della relazione precedente si ha che:

$$T_a = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_b^2} \cdot T_b + \frac{\sigma_b^2}{\sigma_0^2 + \sigma_b^2} \cdot T_0$$

cioè:

$$T_a = T_b + W \cdot (T_0 - T_b)$$

↑
Innovation

$$W = \frac{\sigma_b^2}{\sigma_0^2 + \sigma_b^2}$$

Gain

L'Errore di analisi è dato da:

$$\sigma_a^2 = \frac{1}{1/\sigma_b^2 + 1/\sigma_0^2}$$

$$\sigma_a^2 = (1 - W) \cdot \sigma_b^2$$

In sintesi

- ⇒ L'Analisi la si ottiene **aggiungendo** al **background (first guess)** l'**innovazione** (cioè la differenza tra la misura ed il background) pesata in maniera statisticamente ottimale;
- ⇒ Il peso ottimale (**Gain**) è la varianza di background divisa per la somma della varianza di background e quella di misura. Maggiore è la varianza di background, maggiore è la correzione del valore di background;
- ⇒ La **precisione dell'analisi** è la **somma delle precisioni di background e di misura**;
- ⇒ La **varianza di analisi** è **quella di background ridotta del fattore (1-W)**. Ciò sta a significare che minore è l'incertezza della misura più si *trascina* il valore di background verso il valore vero.

In Generale

Modifichiamo la situazione considerata nel Secondo Passo rendendola più realistica.

1) Postazioni di Misura

Mentre nel secondo Passo è stata considerata una sola postazione di misura attiva

..... **qui ora consideriamo M postazioni di misura.**

Ogni postazione j -esima ($j = 1, .. ,M$) \Rightarrow $T_{0,j}$ $\epsilon_{0,j}$ $\sigma_{0,j}^2$

E' consuetudine definire un vettore y_0 (**vettore delle misure**) come:

$$[y_0] = \begin{bmatrix} T_{0,1} \\ T_{0,2} \\ \dots \\ T_{0,M} \end{bmatrix}$$

Gli M punti di misura **possono essere distribuiti in modo qualsiasi nello spazio** e la loro posizione normalmente non coincide con altrettanti nodi della griglia di Analisi.

A rigore, **la caratterizzazione statistica** delle M misure non si esaurisce con l'errore di misura associato ad ognuna di esse, **ma deve comprendere anche la Correlazione/Covarianza (esistente o meno) tra i diversi errori di misura.**

In sostanza bisogna rispondere alla domanda:

Un errore associato alla misura j -esima è correlato all'errore associato alla misura k -esima?

Anche se non lo si può escludere a priori, normalmente si è ipotizzato per semplicità che non ci sia correlazione tra i vari errori di misura.

Quindi è possibile definire una matrice di Varianza-Covarianza $[R]$ che risulta quindi diagonale:

$$\mathbf{R} = \begin{bmatrix} \sigma_{0,1}^2 & 0 & \dots & 0 \\ & \sigma_{0,2}^2 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & \sigma_{0,M}^2 \end{bmatrix}$$

Molto spesso le **misure considerate vengono prodotte da una rete fissa** (rete meteorologica, micrometeorologica o rete di Qualità dell'Aria). In questo caso, molto frequente, **tutti gli strumenti utilizzati sono gli stessi e se le modalità di acquisizione ed elaborazione delle misure sono le medesime**, allora gli errori di misura statisticamente saranno gli stessi e quindi ogni misura della rete meteorologica sarà caratterizzata dalla medesima varianza σ_0^2 .

In questo caso, la **Matrice di Varianza-Covarianza [R]**, i cui elementi diagonali sono le varianze degli errori di misura per ogni singola postazione, mentre gli elementi non diagonali sono le Covarianze tra errori di misura appartenenti a diverse postazioni, per l'**indipendenza statistica delle misure**, prende la **forma diagonale** seguente:

$$[\mathbf{R}] = \sigma_0^2 \cdot \begin{bmatrix} 1 & 0 & \dots & 0 \\ & 1 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma_0^2 \cdot [\mathbf{1}]$$

dove con [1] si è indicata la matrice (M,M) unitaria.

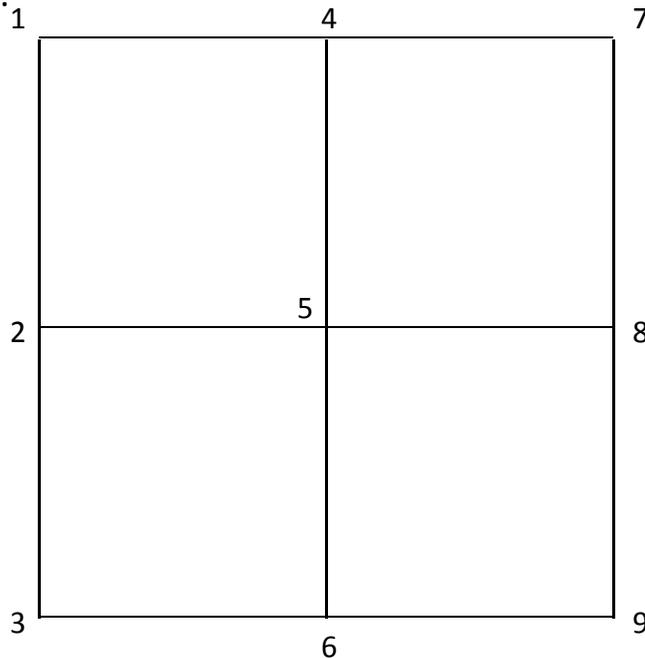
2) Punti di Analisi

Nel Secondo Passo è stato considerato un solo punto in cui era richiesta l'analisi e per giunta questo punto coincideva col punto di misura.

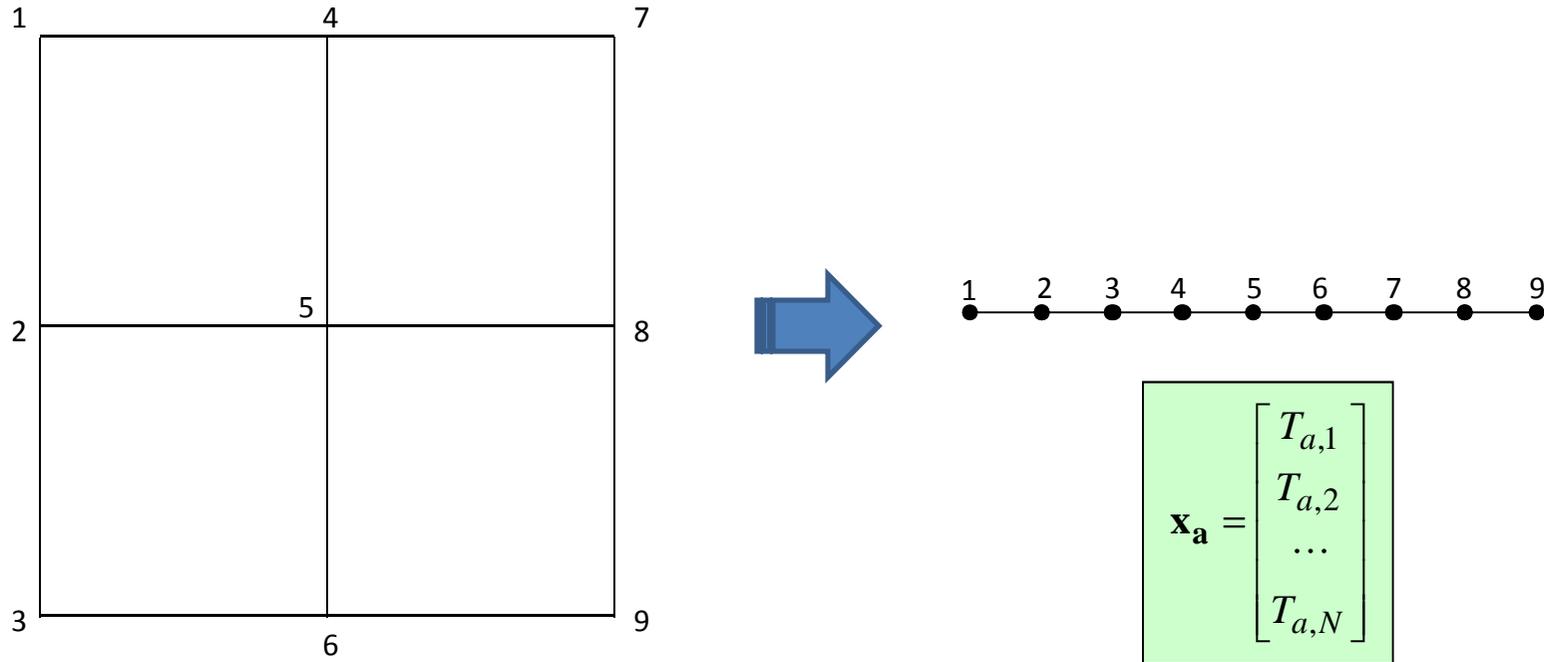
Qui ora ipotizziamo di **volere l'Analisi** in **N punti distinti del territorio**.

Anche se questi punti potrebbero essere distribuiti in maniera qualsiasi nel dominio, è normale che essi siano i nodi di una griglia di calcolo.

Semplice esempio di griglia.



Ad ogni punto i -esimo ($i = 1, \dots, N$) della griglia sarà associato un valore $T_{a,i}$ e, per semplicità matematica, tutti questi valori vengono di consueto raggruppati nel **Vettore di Analisi $[x_a]$**



Il Vettore di Analisi rappresenta il risultato della spazializzazione delle misure (rappresentate dal vettore delle misure $[y_0]$).

3) Background (First Guess)

Nel Secondo Passo si era considerato un valore di background per la posizione in cui veniva fatta l'Analisi (tra l'altro, coincidente con punto di misura).

Ora si assume che **per ogni punto del territorio** sia possibile **ottenere dalle informazioni a priori disponibili** una prima stima della variabile di analisi (qui, la temperatura a 2 m dal suolo). Quindi si ritiene disponibile **una metodologia (modello) con cui** stimare in prima approssimazione la variabile di Analisi.

I modi operativi per fare ciò possono essere vari:

- **può essere disponibile** (se la variabile è normalmente considerata da un modello meteorologico o di simulazione della dispersione/deposizione di una sostanza gassosa o aeriforme in aria) **una simulazione numerica che si ritiene non in grado di dar conto dei fenomeni locali che si intendono studiare**. Da qui la necessità dell'Analisi.
- **il first guess può essere derivato direttamente da opportuna elaborazione delle misure disponibili** (esempio: il background è costante spazialmente e pari al valore medio delle misure)

E' consuetudine definire due vettori che contengono i valori di background:

1. Il **Vettore** $[x_b]$ di dimensione N che contiene i valori di background per gli N punti in cui si considera l'Analisi

$$\mathbf{x}_b = \begin{bmatrix} T_{b,1} \\ T_{b,2} \\ \dots \\ T_{b,N} \end{bmatrix}$$

2. Il **Vettore** $[y_b]$ di dimensione M che contiene i valori di background per gli M punti in cui sono attive le misure

$$\mathbf{y}_b = \begin{bmatrix} T_{b0,1} \\ T_{b0,2} \\ \dots \\ T_{b0,M} \end{bmatrix}$$

Dato che il Background è noto nei nodi della griglia di Analisi, per ottenere il valore di background nei punti di misura è necessario predisporre una metodologia che consenta la loro stima sulla base del background noto nei nodi.

3. La **matrice [S]**, di dimensione **(M,M)** che è la Matrice di Varianza-Covarianza degli errori di misura.

In essa l'elemento generico $S_{j,k}$ è la Covarianza tra l'errore di background al sito di misura j e quello al sito k .

Se $j = k$, l'elemento (che sta sulla diagonale), è la varianza dell'errore.

The diagram illustrates the structure of the matrix $[S]$. A box labeled "Colonna k" has an arrow pointing to the k -th column of the matrix. A box labeled "Riga j" has an arrow pointing to the j -th row of the matrix. The matrix is shown as:

$$[S] = \begin{bmatrix} \overline{\varepsilon_{b01} \cdot \varepsilon_{b01}} & \overline{\varepsilon_{b01} \cdot \varepsilon_{b02}} & \cdots & \overline{\varepsilon_{b01} \cdot \varepsilon_{b0M}} \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \overline{\varepsilon_{b0j} \cdot \varepsilon_{b0k}} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \overline{\varepsilon_{b0M} \cdot \varepsilon_{b0M}} \end{bmatrix}$$

4. La **matrice [G]** che è la più complessa da illustrare.

Se consideriamo il generico **nodo di Analisi j**, l'errore di background relativo è ε_{bj} .

Se consideriamo il generico **punto di misura k**, il relativo errore di background è ε_{b0k} .

Il generico elemento della matrice [G], dimensione (N,M) è la covarianza tra questi due errori di background.

$$[G] = \begin{bmatrix} \varepsilon_{b1} \cdot \varepsilon_{b01} & \varepsilon_{b1} \cdot \varepsilon_{b02} & \cdots & \varepsilon_{b1} \cdot \varepsilon_{b0M} \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \varepsilon_{bj} \cdot \varepsilon_{b0k} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \varepsilon_{bN} \cdot \varepsilon_{b0M} \end{bmatrix}$$

Riga j

Colonna k

La stima operativa degli elementi delle matrici [S] e [G] è piuttosto complessa e non si parlerà di ciò nei dettagli. Basti dire che sono state individuate delle relazioni funzionali che esprimono analiticamente tutte le covarianza (e varianze) richieste, che dipendono da parametri deducibili dalle serie storiche disponibili o dalla natura del problema trattato.

4) L'Analisi

La deduzione dell'equazione che quantifica l'Analisi relativa agli N punti del dominio di calcolo sulla base di un **campo di background** e delle M misure realizzate in altrettante postazioni di misura è concettualmente analoga a quella fatta per il Passo 2, **ma dal punto di vista pratico i dettagli matematici (di Algebra Lineare) di questa deduzione sono complessi e noiosi e non necessari in questa sede.**

Per semplificare, partiamo dal **quanto ottenuto al Passo 2** per un solo punto di misura e per un solo punto di analisi (coincidenti tra loro). In particolare se:

$$x_a = T_a$$

$$x_b = T_b$$

$$y_0 = T_0$$

$$y_b = T_b \text{ (riferito al punto di misura)}$$

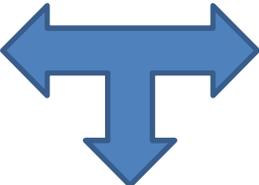
Allora l'Analisi, come detto, è data da:

$$x_a = x_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_b^2} \cdot (y_0 - y_b)$$

Se operiamo le sostituzioni seguenti:

- x_a , analisi in un punto → Vettore delle analisi $[x_a]$ negli N punti del dominio
- x_b , background in un punto → Vettore del background $[x_b]$ negli N punti del dominio
- Innovazione ($y_0 - y_b$) → Vettore delle M innovazioni $[y_0] - [y_b]$
- Varianza σ^2_0 → Matrice di Varianza-Covarianza delle misure [R]
- Varianza σ^2_b a denominatore → Matrice di Varianza-Covarianza di background nelle postazioni di misura [S]
- Varianza σ^2_b a numeratore → Matrice di Varianza-Covarianza degli errori di background tra postazioni di misura e punti di analisi [G]

Si ottiene la **Relazione Fondamentale dell'Optimal Interpolation** in **forma vettoriale** che agevolmente si può confrontare col risultato ottenuto al **Passo 2** e riportata per confronto.

$$x_a = x_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_b^2} \cdot (y_0 - y_b)$$


Sostituzioni

$$[\mathbf{x}_a] = [\mathbf{x}_b] + [\mathbf{G}] \cdot ([\mathbf{S}] + [\mathbf{R}])^{-1} \cdot ([\mathbf{y}_0] - [\mathbf{y}_b])$$

Se è noto il campo background, le misure, la statistica relativa e le matrici [S] e [G], il vettore di analisi lo si ottiene immediatamente dalla relazione precedente.

Questa relazione, però, presenta un problema matematico legato alla stima dell'inversa della matrice ([S]+[R]), risolvibile con metodi numerici, ma comunque con notevole sforzo di calcolo.

L'Analisi ottenuta con la O.I. è la **Migliore Analisi Possibile** (in termini statistici), tenuto conto:

1. delle informazioni a priori (campo di Background) disponibili sul fenomeno indagato sia nei nodi della griglia di Analisi, sia (ricostruite per interpolazione o modellisticamente) nei punti di misura;
2. della informazioni relative agli errori connessi al Campo di Background
3. delle misure disponibili
4. degli errori connessi alle misure e al loro ottenimento

La O.I., oltre a fornire una relazione matematica che consente di stimare i pesi con cui le misure entrano nel modello lineare di Analisi, **consente pure di individuare una matrice $[P_a]$** la cui diagonale contiene la Varianza degli errori di Analisi per gli N nodi della griglia, mentre ogni elemento $P_{i,j}$ rappresenta la Covarianza tra l'errore di Analisi nei punti i e j . L'espressione analitica della matrice degli errori di analisi risulta pari a:

$$[P_a] = [R] \cdot ([R] + [S])^{-1} [S]$$

Quella presentata è la Teoria di base dell'Optimal Interpolation che nel recente passato ha costituito la base per la preparazione delle condizioni iniziali per il run dei Modelli di Previsione Meteorologica.

Nonostante abbia fornito ottimi risultati, attualmente è stata sostituita da metodologie matematiche molto più complesse e potenti che sfruttano maggiormente le proprietà statistiche e delle misure e dei modelli (3D-Var, 4D-Var, Filtro di Kalman).

Parallelamente si è consolidata un'altra celebre metodologia di spazializzazione che, pur partendo apparentemente da basi matematiche differenti, si è dimostrato che di fatto si riconduce all'Optimal Interpolation. Questa metodologia o meglio questa famiglia di metodologie è quella del Kriging, molto usata soprattutto nelle applicazioni ingegneristiche e minerarie.

Nelle applicazioni pratiche di spazializzazione, meno critiche e complesse di quelle legate alla predisposizione delle condizioni iniziali per i modelli meteorologici e di dispersione degli inquinanti, si è cercata una semplificazione operativa della metodologia base dell'OI.

IL METODO DI BRATSETH

La **complessità del metodo O.I.** e l'**onerosità computazionale** ne scoraggiano la sua applicazione al di fuori dei Centri di Previsione Meteorologica.

Nelle applicazioni più comuni, sarebbe auspicabile poter disporre di un metodo di spazializzazione delle misure discrete (per esempio ottenute in una campagna sperimentale) che avesse le seguenti caratteristiche:

- fosse computazionalmente efficiente e concettualmente semplice,
- utilizzasse al meglio le **informazioni a priori** e le **informazioni statistiche** disponibili,
- **non fosse totalmente semiempirico,**
- fosse **legato concettualmente all'OI,**
- però fosse algoritmicamente simile ai metodi delle Successive Correzioni(SCM).

Un metodo di Spazializzazione che possiede tutte queste caratteristiche esiste ed è il **Metodo di Bratseth**.

Questo metodo **parte** dalla medesima base matematica usata dall'OI per ottenere l'analisi, ma, **impiegando complessi ragionamenti di Algebra Lineare** (che non tenteremo di presentare, nemmeno in una forma semplificata), alla fine perviene ad una formulazione matematicamente del tutto equivalente all'O.I, ma che:

- che si presenta come una procedura di Correzione successive di un campo iniziale
- è formalmente identica a quella usata da ogni tipico algoritmo di analisi con Successive Correzioni.

Gli **ingredienti** di cui ha bisogno il metodo sono sempre gli stessi:

- un **campo di background** che consente all'algoritmo di effettuare la prima iterazione, ottenendo così il primo valore di Analisi,
- una serie di **misure** distribuite irregolarmente nello spazio.

Forma semplificata del metodo di Bratseth usata da ARPA LAZIO nella Data Assimilation continua delle misure nel modello previsionale della Qualità dell'Aria.

Campo di background

- ⇒ Quando si dispone di conoscenze a priori del campo (es. si ha una previsione numerica dello stesso), il background è proprio questo campo.
- ⇒ Quando si hanno solo le misure in un numero discreto di punti, li si interpola in modo semplice, per esempio usando un'interpolazione bi-lineare.

Informazioni statistiche del fenomeno

Le informazioni statistiche richieste sono una semplificazione di quelle usate da O.I. In particolare:

- ⇒ La statistica degli errori di misura riassunta, nell'ipotesi realistica di una rete dotata dei medesimi sensori, dalla varianza σ_0^2 delle misure;
- ⇒ La statistica del campo di background, riassunta nella varianza intrinseca al campo di background σ_b^2

⇒ Disponiamo di M misure nel dominio di calcolo → $y_{0,m}$ ($m = 1, \dots, M$)

⇒ Vogliamo l'Analisi $x_{a,i}$ in N punti comunque disposti nello spazio ($i = 1, \dots, N$)

ALGORITMO

⇒ Alla prima iterazione, l'Analisi nei punti di misura e nei punti di analisi viene posta pari al corrispondente valore di background

⇒ All'iterazione k-esima:

1. L'Analisi nel generico punto i -esimo sarà data dalla relazione seguente:

$$x_{a,i}^k = x_{a,i}^{k-1} + \sum_{m=1}^M \alpha_{i,m} \cdot [y_{0,m} - x_{a,m}^{k-1}]$$

2. Una volta ottenuta l'Analisi per tutti i punti i ($i = 1, \dots, N$) della griglia regolare di Analisi, si calcola anche l'Analisi stessa in tutti i punti j ($j = 1, \dots, M$) in cui si dispone delle misure e lo si fa con la stessa relazione di prima:

$$x_{a,j}^k = x_{a,j}^{k-1} + \sum_{m=1}^M \alpha_{i,j} \cdot [y_{0,j} - x_{a,j}^{k-1}]$$

Queste relazioni che conducono all'Analisi si presentano del tutto identiche a quelle dei SCM.

In realtà, dopo complessi passaggi matematici qui non riportati, queste relazioni si ottengono dalla Teoria Base dell'O.I., una volta introdotte semplificazioni nelle proprietà statistiche del Campo di Background.

In pratica, quello che si è ottenuto, è che la matematica dei SCM non è semi-empirica se si scelgono opportunamente i pesi $\alpha_{i,j}$ delle innovazioni e proprio qui sta la connessione con O.I. e la potenza del metodo.

STIMA DEI PESI

Gli **elementi essenziali** per determinare i pesi $\alpha_{i,j}$ sono:

⇒ La distanza Euclidea tra i punti considerati e

⇒ la correlazione caratteristica del campo stesso.

- 1) L'ipotesi pratica utilizzata è che la **correlazione** tra il valore della variabile in Analisi in un due punti **diminuisce esponenzialmente** con la distanza orizzontale Δr e verticale Δz secondo una legge del tipo

$$\rho_{i,j} = \exp\left[-\frac{|\Delta r|^2}{R_h^2}\right] \cdot \exp\left[-\frac{|\Delta z|^2}{R_z^2}\right]$$

dove R_h e R_z sono Lunghezze di Scala da definire in base alla fisica del fenomeno e che si riducono progressivamente con l'effettuazione delle iterazioni.

$$(R_{h,z} = kk \cdot R_{h,z}^0 \quad \text{per esempio } kk = 2.0, 1.0, 0.8, 0.4)$$

- 2) Gli errori di misura (riassunti dalla varianza σ_o^2) e quelli del campo di background (riassunti dalla varianza σ_b^2) possono essere complessivamente considerati col rapporto $\varepsilon_n^2 = \sigma_o^2 / \sigma_b^2$.

$$x_{a,j}^k = x_{a,j}^{k-1} + \sum_{m=1}^M \alpha_{i,j} \cdot [y_{0,j} - x_{a,j}^{k-1}]$$

In termini Analitici i Pesi sono dati dalla relazioni seguenti

$$\alpha_{i,m} = \frac{\rho_{i,m}}{\mu_m}$$

$$\mu_i = \varepsilon_n^2 + \sum_{m=1}^M \rho_{i,m}$$

$$\alpha_{i,j} = \begin{cases} \frac{\rho_{i,j}}{\mu_m} & \text{se } i \neq j \\ \frac{\rho_{i,j} + \varepsilon_n^2}{\mu_m} & \text{se } i = j \end{cases}$$

$$x_{a,i}^k = x_{a,i}^{k-1} + \sum_{m=1}^M \alpha_{i,m} \cdot [y_{0,m} - x_{a,m}^{k-1}]$$

Per concludere

.. Il Metodo di Bratseth è basato sulle stesse basi matematiche e statistiche dell'OI, ma

.... si presenta a tutti gli effetti come un Metodo delle Successive Correzioni, quindi di semplice implementazione ed uso pratico

.... oltre che essere più facilmente adattabile ai problemi specifici (anche alle misure raccolte in una campagna di monitoraggio).

Grazie per la Pazienza e l'Attenzione!