

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

FACOLTA' DI INGEGNERIA

**CORSO DI LAUREA SPECIALISTICA IN INGEGNERIA PER L'AMBIENTE E IL
TERRITORIO**

Indirizzo: Geoingegneria

*DIPARTIMENTO DI INGEGNERIA CHIMICA MINERARIA E DELLE TECNOLOGIE
AMBIENTALI*

TESI DI LAUREA

In

Modelli geostatistici per la geoingegneria LS

**CARATTERIZZAZIONE GEOSTATISTICA SPAZIO TEMPORALE DELLA
DISTRIBUZIONE DI BENZENE SUL TERRITORIO EUROPEO**

CANDIDATO
Monica Agostinone

RELATORE:
Chiar.mo Prof.
Roberto Bruno
CORRELATORE:
Alberto Pistocchi

Anno Accademico 2006/07
Sessione III

ABSTRACT

L'oggetto della presente tesi è costituito dalla caratterizzazione spazio-temporale della variabile ambientale concentrazione di benzene in atmosfera sul territorio europeo, attraverso diversi approcci e metodologie di stima.

L'analisi è stata condotta a partire dai dati dell'airbase comunitario della EEA (European Environment Agency) in collaborazione con il JRC di Ispra, Centro Comune di Ricerca della Commissione Europea. Lo studio geostatistico è stato svolto utilizzando principalmente in ambiente ILWIS (*Integrated Land and Water Information System*), un software open source adottato dallo stesso Jrc.

Lo studio si è concentrato sui dati relativi agli anni 2003, 2004, 2005, i più recenti tra quelli a disposizione nel database, e si è articolata in cinque fasi:

1. Analisi monovariata: studio della variabilità spaziale della concentrazione di benzene di ciascun anno sull'intera estensione europea e per diverse aree omogenee; cartografia mediante krigaggio ordinario, testando diversi modelli di variogramma.
2. Analisi spazio temporale: studio della correionalizzazione delle tre concentrazioni di benzene nei diversi anni; cartografia mediante cokrigaggio.
3. Integrazione di informazioni geografiche: studio di correlazione dei dati di benzene con informazioni fisiche e climatiche desunte da cartografie tematiche a scala europea (grandezze meteo, DEM, uso del suolo, ecc); analisi delle componenti principali delle variabili ausiliarie; cartografia mediante Krigaggio con Deriva Esterna.
4. Studio a scala locale: caratterizzazione spaziale della variabile per 5 aree omogenee coincidenti con i confini nazionali di Italia, Francia, Germania, Gran Bretagna e Spagna, al fine di mettere in luce la variabilità per differente scala spaziale.
5. Validazione confronto e analisi dei risultati ottenuti.

Le conclusioni raggiunte evidenziano e quantificano le caratteristiche spazio-temporali del benzene, nonché le correlazioni con differenti tematismi.

E' risultato inoltre che la cartografia risente della tecnica geostatistiche di stima utilizzata e della scala di lavoro.

INDICE

1. INTRODUZIONE	1
1.1.CONTESTO OBIETTIVI E STRUTTURA DELLO STUDIO	1
1.2.IMPORTANZA DELLE MAPPE TEMATICHE AMBIENTALI	3
1.2.1. Ruolo della comunità europea e quadro normativo per l'inquinamento atmosferico	4
1.2.2. L'agenzia europea dell'ambiente EEA (European Environment Agency)	6
1.3. L'INQUINANTE STUDIO: IL BENZENE A LIVELLO EUROPEO	8
1.4.STRUMENTI GEOSTATISTICI	10
1.4.1. Approccio geostatistico nella caratterizzazione spaziale dei fenomeni ambientali e ricostruzione spaziale dei fenomeni tramite stima.	10
1.4.1.1. Stime in caso di stazionarietà	12
<i>1.4.1.1.1. Krigaggio Ordinario</i>	13
<i>1.4.1.1.2. Stima con il contributo di variabili ausiliarie: cokrigaggio</i>	15
1.4.1.2. Stime nel caso di non stazionarietà	16
<i>1.4.1.2.1. Stima con il contributo di variabili ausiliarie:Kriging con deriva esterna</i>	16
1.4.2. Analisi in componenti principali	17
1.4.3. Software ILWIS	18
1.4.3.1. Geostatistica in ILWIS	19
2. CASO STUDIO	20
1.1. ANALISI DEI DATI DISPONIBILI: THE EUROPEAN AIR QUALITY DATABASE	20
2.1.1. Rete, stazioni, statistiche AirBase	21
1.2.ESTRAZIONE DEI DATI UTILI	22

2.2.1. Importazione dei dati in ILWIS	24
1.3. VISUALIZZAZIONE SPAZIALE DEI DATI	24
1.4. STATISTICHE DI BASE	28
3. STUDIO MONOVARIATO	33
1.1. ANALISI GEOST DELLA VARIABILITÀ SPAZIALE	33
3.1.1. Variogrammi sperimentali	33
3.1.2. Modellizzazione dei variogrammi	42
1.2. CARTOGRAFIA MEDIANTE KRIKAGGIO ORDINARIO	50
4. STUDIO SPAZIO TEMPORALE	58
1.1. INTRODUZIONE ALL'ANALISI SPAZIO TEMPORALE	58
1.2. L'APPROCCIO MULTIVARIATO ALL'ANALISI SPAZIO TEMPORALE	58
4.2.1. Correlazioni temporali	58
4.2.2. Modellizzazione dei variogrammi diretti e incrociati	61
1.3. CARTOGRAFIA MEDIANTE COKRIGAGGIO	65
5. INTEGRAZIONE DI INFORMAZIONI GEOGRAFICHE: LO STUDIO MULTIVARIATO	72
1.1. INFORMAZIONI AUSILIARIE DISPONIBILI	73
1.2. IDENTIFICAZIONE DELLE VARIABILI AUSILIARIE	74
5.2.1. Costruzione delle mappe climatiche	75
5.2.2. Formattazione dell'informazione topografica	78
5.2.3. Elaborazione della carta delle strade principali: densità di strade e distanza dalle strade principali	81
5.2.4. Cartografia dell'uso del suolo	84
5.2.5. Sottomappa e formattazione della mappa dell'intensità della luce notturna	84
1.3. ANALISI STATISTICA MULTIVARIATA	85

5.3.1. Nuvole di correlazione	85
5.3.2. Matrici di correlazione	92
5.3.3. Analisi delle componenti principali	94
1.4. ANALISI STRUTTURALE CON DERIVA ESTERNA	96
5.4.1. Variogrammi delle fluttuazioni	100
5.4.2. Cartografia mediante krigaggio con DE	102
6. STUDIO A SCALA LOCALE	107
1.1. FINALITÀ	107
1.2. DEFINIZIONE DELLE AREE OMOGENEE E STRUTTURA DELLO STUDIO	107
6.2.1. Italia	109
6.2.1.1. Analisi strutturale	109
6.2.1.2. Cartografia mediante Krigaggio Ordinario	111
6.2.2. Spagna	113
6.2.2.1. Analisi strutturale	113
6.2.2.2. Cartografia mediante Krigaggio Ordinario	113
6.2.3. Gran Bretagna	115
6.2.3.1. Analisi strutturale	115
6.2.3.2. Cartografia mediante Krigaggio Ordinario	116
6.2.4. Germania	117
6.2.4.1. Analisi strutturale	117
6.2.4.2. Cartografia mediante Krigaggio Ordinario	118
6.2.5. Francia	119
6.2.5.1. Analisi strutturale	119
6.2.5.2. Cartografia mediante Krigaggio Ordinario	120
7. ANALISI DEI RISULTATI	121
1.1. Analisi dei diversi modelli di variabilità spaziale	122
1.2. Raffronto fra le cartografie	128

8. CONCLUSIONI E DISCUSSIONI	137
-------------------------------------	-----

BIBLIOGRAFIA	142
---------------------	-----

ALLEGATO 1

ALLEGATO 2

Musica

Poesia

1. INTRODUZIONE

1.1. CONTESTO OBIETTIVI E STRUTTURA DELLO STUDIO

La presente tesi, avendo come oggetto la caratterizzazione spazio-temporale e il conseguente ottenimento di mappe tematiche per la variabile ambientale concentrazione di benzene in atmosfera, relativamente all'estensione europea, si colloca all'interno di un contesto sociale e politico sempre più attento alla salvaguardia e al risanamento dello stato dell'ambiente atmosferico, che necessita di elementi informativi e di supporto alle decisioni; quest'ultime in campo ambientale possono variare dalla selezione: definire un sito inquinato o meno, una zona a rischio o meno per la salute umana, a decisioni puramente normative: stabilire limiti di legge adeguati. In tale ottica si comprende l'importanza delle mappe tematiche ambientali come elemento d'informazione che può svilupparsi solo dopo un'attenta analisi del fenomeno, studio della variabilità e conseguente ricostruzione della grandezze in esame sul dominio in esame, spaziale e/o considerato; le analisi a loro volta si basano su dati provenienti da campagne di osservazione, monitoraggio e rilevamento dei fenomeni ambientali e delle sue variabili. A tale proposito sono nati corpi dedicati al reperimento ed alla gestione di tali banche dati e l'Unione Europea ha svolto in quest'ambito e a livello comunitario, un ruolo fondamentale sia nell'adozione di direttive ambientali che nella creazione di corpi ed istituzioni dedicati; tra questi, l'Agenzia Europea dell'Ambiente, si propone come un corpo dell'UE dedicato a fornire dati sicuri e indipendenti sull'ambiente; l'*Airbase*, database atmosferico della EEA è stato scelto per lo studio oggetto della presente tesi, che si è concentrato sull'inquinante Benzene in fase d'indagine dal JRC, che rientra all'interno del registro degli inquinanti europeo EPER.

Poste tali premesse, necessarie per inquadrare il lavoro e comprenderne le finalità, gli obiettivi possono essere così riassunti:

1. Il primo obiettivo è l'ottenimento di una mappa tematica della concentrazione di benzene in atmosfera per l'estensione europea, quindi la comprensione della correlazione spazio temporale e della dispersione dell'inquinante, a diverse scale spaziali e temporali, mettendone in luce caratteristiche.
2. Secondo obiettivo è stato quello di analizzare le correlazioni del benzene con una serie di materiale informativo ausiliario: fattori geografici, climatici, uso del suolo e distanze dalle strade principali, valutando così le relazioni dell'inquinante con diversi tematismi.
3. Il terzo obiettivo è legato alle metodologie di stima, dal momento che sono stati sviluppate diverse stime corrispondenti ad altrettanti approcci teorici e metodologici squisitamente geostatistici, oltre che a differenti scale spaziali, una finalità dello studio è stata quella di validare i differenti modelli assunti come rappresentativi della reale variabilità e quindi alla base delle cartografie ottenute, valutarne vantaggi e svantaggi, differenze, al fine di poter individuare la tecnica migliore in relazione ai fini dell'utilizzo delle informazioni cartografiche così elaborate.

Tali obiettivi sono stati perseguiti avvalendosi degli strumenti forniti dalla teoria geostatica e operativamente svolti con il software *ILWIS Integrated Land and Water Information System*, adottato dal JRC, che compendia funzionalità gis e geostatistiche. Lo studio è stato così strutturato:

1. Analisi monovariata: studio della variabilità spaziale della concentrazione di benzene di ciascun anno sull'intera estensione europea e per diverse aree omogenee; cartografia mediante krigaggio ordinario, testando diversi modelli di variogramma.
2. Analisi spazio temporale: studio della correionalizzazione delle tre concentrazioni di benzene nei diversi anni; cartografia mediante cokrigaggio.

3. Integrazione di informazioni geografiche: studio di correlazione dei dati di benzene con informazioni fisiche e climatiche desunte da cartografie tematiche a scala europea (grandezze meteo, DEM, uso del suolo, ecc); analisi delle componenti principali delle variabili ausiliarie; cartografia mediante Krigaggio con Deriva Esterna.
4. Studio a scala locale: caratterizzazione spaziale della variabile per 5 aree omogenee coincidenti con i confini nazionali di Italia, Francia, Germania, Gran Bretagna e Spagna, al fine di mettere in luce la variabilità per differente scala spaziale.
5. Validazione confronto e analisi dei risultati ottenuti.

Obiettivi e struttura conferiscono al lavoro un livello di completezza e approfondimento che lo inseriscono all'interno di uno stato dell'arte che in trattazioni singole ha preso in esame le diverse problematiche del caso studio, singolarmente in studi dedicati; a tal proposito si citano il report della EEA sullo stato dell'ambiente atmosferico finalizzato alla ricostruzione delle concentrazioni dell'inquinante sul suolo europeo, studi cofinanziati dall'UE relativamente all'inquinamento atmosferico e alla dispersione del benzene come MACBETH (*Monitoring of Atmospheric Concentration of Benzene in European Towns and Homes*) che ne ha approfondito le relazioni con tematismi esterni e in special modo con il vento, studi specificamente geostatistici incentrati sulle metodologie di stima, come quelli portati avanti da Geovariance.

1.2.IMPORTANZA DELLE MAPPE TEMATICHE AMBIENTALI

L'oggetto della presente tesi è costituito da una caratterizzazione spazio temporale, geostatisticamente basata, che ha portato all'ottenimento di una cartografia tematica ambientale per i livelli della concentrazione di benzene sul suolo europeo. Le mappe tematiche ambientali, che ricostruiscono l'andamento di una variabile ambientale su un determinato dominio spazio

temporale, oltre a rappresentare elementi oggettivi per l'interpretazione e comprensione dei fenomeni, trovano sempre più impiego nella messa a punto dei sistemi di supporto alle politiche decisionali e nella progettazione e gestione dei sistemi di monitoraggio; si inseriscono pertanto, all'interno di uno scenario sociale e politico che necessita di elementi conoscitivi adeguati per far fronte alla crescente necessità di proteggere le risorse fondamentali, la salute e di assicurare uno sviluppo economico responsabile che garantisca un futuro alle prossime generazioni. Solo se si dispone di un corredo informativo appropriato è possibile interpretare i fenomeni ambientali esaminati e agire con politiche specifiche nell'ottica della sicurezza e dello sviluppo sostenibile, oltre che del risanamento e le mappe tematiche vanno a costituire in tale contesto uno strumento fondamentale essendo il frutto dello studio delle variabili che rappresentano un determinato fenomeno d'attenzione ambientale. Nello specifico dell'inquinamento atmosferico l'attenzione allo studio delle variabili che lo rappresentano, deriva dai rischi per la salute osservati in cambiamenti nella mortalità e frequenza di malattie sia a breve sia a lungo termine, oltre che dai danni osservati agli ecosistemi e ai materiali, con particolare riguardo ai monumenti.

1.2.1. Ruolo della comunità europea e quadro normativo per l'inquinamento atmosferico.

L'indagine quali quantitativa delle concentrazioni di benzene sull'estensione europea oggetto della presente tesi, si inserisce all'interno di un quadro normativo sempre più attento alla salvaguardia e al risanamento dello stato dell'ambiente atmosferico, che prevede strumenti politici oltre a corpi e associazioni dedicati, che non possono prescindere da un'indagine accurata dei fenomeni ambientali, destinata all'adozione di misure e provvedimenti oltre che a scopi divulgativi e d'informazione.

L'adozione di una politica ambientale, che in campo internazionale si suole

far risalire alla Conferenza delle Nazioni Unite di Stoccolma del 1972, (“Human Environment” (ambiente umano)), a livello continentale europeo vede tra i suoi fautori la Comunità Europea; l’UE svolge, infatti, un ruolo fondamentale nell’adozione di una politica ambientale europea, che ingloba da un lato la definizione di normative volte a ridurre le concentrazioni di inquinanti nell’ambiente in tutti gli stati, dall’altro lo sviluppo e la gestione di un sistema d’informazione sui livelli di inquinamento europei e nazionali, che costituisce uno strumento rilevante per la gestione a livello locale rivolto oltre che alle autorità nazionali, alla cittadinanza pubblica.

Il quadro legislativo relativo alle emissioni ed alle concentrazioni ambientali degli inquinanti atmosferici ha subito negli anni un’evoluzione legata alla storia e alla gravità di quest’ultimi; se fino alla fine degli anni 80, l’attenzione era rivolta principalmente ai macroinquinanti, nel corso degli ultimi anni il campo dei composti soggetti a normative si è notevolmente ampliato. Il rilevamento della qualità dell’aria mediante sistemi automatici risale alla metà degli anni Settanta principalmente con l’obiettivo di controllare gli impianti industriali. È solo negli anni Ottanta che l’attenzione si sposta sulle «immissioni» attraverso l’introduzione di limiti sulla qualità dell’aria relativi alle particelle sospese.

Attualmente la direttiva quadro in materia di inquinamento atmosferico è rappresentata dalla direttiva 96/62/CE del 27 settembre 1996 sulla valutazione e gestione della qualità ambientale dell’aria, che prevede quattro direttive subordinate che stabiliscono gli obiettivi di qualità dell’aria e i valori limite per l’ossido di zolfo, diossido di azoto, particolato, ozono, benzene, monossido di carbonio, arsenico, cadmio, mercurio, nichel e idrocarburi aromatici policiclici. Obiettivo della direttiva è anche quello di armonizzare il monitoraggio, la misurazione e i metodi di valutazione per ottenere dati confrontabili e assicurare una buona informazione pubblica.

Nel 1999 il protocollo di Goteborg per la convenzione sull’inquinamento

atmosferico transfrontaliero a largo raggio (CRLTAP; UNECE, 1999) stabilisce le emissioni massime per il 2010 per il diossido di zolfo, l'ossido di azoto, i composti organici volatili e l'ammoniaca.

Due ulteriori protocolli specifici per il controllo delle emissioni sono stati convenuti nella convenzione di Arhus nel 1998: uno per i metalli pesanti e un altro per gli inquinanti organici persistenti. Questi ultimi protocolli richiedono una riduzione delle emissioni sotto i livelli del 1990.

La direttiva NEC (Ec, 2001c) stabilisce anche gli obiettivi di emissione per il 2010 per quattro inquinanti principali responsabili dell'acidificazione eutrofizzazione e livelli base dell'ozono ed è supportata da una legislazione specifica.

E' attualmente in vigore il sesto programma d'azione della comunità europea (2001-2010), che si occupa di stabilire 6 strategie tematiche, di cui una sull'inquinamento atmosferico che stabilisce gli obiettivi per raggiungere livelli di qualità dell'aria che non facciano aumentare significativamente gli impatti negativi e il rischio per la salute umana e dell'ambiente. All'interno di tale programma nel 2001 la commissione europea ha lanciato il programma "Clean Air for Europe"(CAFE). L'obiettivo è quello di rivedere le politiche attuali sulla qualità dell'aria e riscontrare i progressi verso gli obiettivi di qualità previsti dalla 6EAP.

1.2.2. L'agenzia europea dell'ambiente EEA (European Environment Agency)

I dati alla base della caratterizzazione spazio-temporale, sviluppata nella presente tesi, sono stati prelevati dal sito dell'agenzia europea dell'ambiente. La EEA *European Environment Agency*, è stata adottata con una direttiva della comunità europea nel 1990, è entrata in vigore a tutti gli effetti nel 1993 dopo la sua localizzazione a Copenhagen e all'inizio del 1994 è diventata operativa. Essa si propone come un corpo dell'unione europea dedicato a

fornire dati sicuri e indipendenti sull'ambiente delineandosi come un'ulteriore fonte di informazione per tutti coloro che sono coinvolti nello sviluppo, nell'adozione, nell'implementazione e nella valutazione della politica ambientale e in generale per la pubblica cittadinanza. I principali utenti della EEA sono rappresentati dagli attori principali della politica europea: la Commissione Europea, il Parlamento Europeo, il Consiglio europeo (specialmente attraverso i cambi di presidenza) e gli stati membri. L'attività della EEA è inoltre di supporto ad altre istituzioni europee secondarie come il comitato sociale ed economico, il comitato delle regioni e la banca d'investimento europea e al di fuori della rete delle istituzioni comunitarie alle compagnie commerciali, alle accademie, alle organizzazioni non governative ed alle altre parti civili della società che possono accedere e far uso delle informazioni messe a disposizione.

Le comunicazioni perseguono un doppio obiettivo, se da un lato devono costituire una risposta specifica alle effettive esigenze di informazione dei possibili utenti, dall'altro è necessario essere sicuri che tali informazioni siano comprese e interpretate nel modo più appropriato. Il sito della EEA costituisce uno dei portali d'informazione telematici più completo e dettagliato, inoltre i testi completi di ogni report ambientale, relazioni e documenti sono scaricabili dal sito così come i dati e le informazioni di supporto ad essi correlati, le pubblicazioni possono in alternativa essere ordinate dagli utenti. La EEA ha inoltre provveduto a creare e a collegare con il proprio sito il registro delle emissioni inquinanti europeo EPER (*European Pollutant Emission Register*): tale registro stabilito dalla decisione della commissione del 17 giugno 2000 si basa sull'articolo 15 della direttiva 96/61/EC su prevenzione e controllo dell'inquinamento, che obbliga gli stati membri a produrre una relazione triennale sulle emissioni industriali in aria e acqua di circa cinquanta inquinanti provenienti da specifiche attività industriali entrambi, opportunamente specificati negli annessi della

normativa.

1.3.L'INQUINANTE STUDIO: IL BENZENE A LIVELLO EUROPEO

L'inquinante oggetto della caratterizzazione spazio temporale sull'estensione europea è stato il benzene. Tale sostanza è inclusa nella lista degli inquinati dell'EPER: il *Registro Europeo delle Emissioni Inquinanti* da complessi industriali istituito dalla decisione della Commissione del 17 luglio 2000.

La sostanza è un liquido incolore, volatile e infiammabile, con un odore caratteristico. Evapora molto velocemente e, se esposto alle fiamme, costituisce un elevato pericolo di incendio. È solo debolmente solubile nell'acqua ma si miscela con la maggior parte dei solventi organici. Il benzene è uno dei prodotti chimici più semplici conosciuti come composti 'aromatici', caratterizzati da atomi di carbonio organizzati in anelli piuttosto che in catene.

Il benzene è utilizzato prevalentemente come reagente nella produzione di un'ampia gamma di sostanze chimiche, impiegate come: coloranti, detergenti, rivestimenti, materie plastiche, fibre, pesticidi, adesivi e rivestimenti, additivi lubrificanti del petrolio, agenti per la pulizia a secco, la colorazione e la produzione di alcuni tipi di gomma. Si trova anche nel petrolio grezzo, nel gasolio e nel fumo di sigaretta ed è anche classificato e regolato come un COV (Composto organico volatile). A livello del suolo, i COV reagiscono con gli altri inquinanti e contribuiscono alla formazione di ozono a concentrazioni potenzialmente tossiche nella bassa atmosfera.

I rilasci di benzene nell'ambiente derivano dalle emissioni dei veicoli, dall'evaporazione del petrolio durante le operazioni di rifornimento di carburante e dalle taniche di carburante, dalle emissioni durante il suo stesso processo produttivo, dal suo impiego come reagente/prodotto chimico intermedio e dalla combustione degli altri combustibili fossili. Le fonti antropiche rappresentano circa il 90% del benzene trovato in atmosfera e la

combustione incompleta è senza dubbio la più significativa, essendo il benzene, come visto, un additivo del petrolio.

Gli effetti negativi sono stati ampiamente studiati, è un genotossico. Alcuni studi su casi umani hanno mostrato effetti cromosomatici ad una media di esposizione sul posto di lavoro relativamente bassa, come 4-75 $\mu\text{g}/\text{m}^3$. In-vivo il benzene è indicato come un mutagenico, la cancerogenicità è stata riscontrata negli umani, con un incremento delle probabilità di leucemia dimostrate per esposizioni a lungo termine.

A livello normativo il benzene è affrontato nella direttiva: 2000/69/EC, seconda direttiva figlia della già citata direttiva quadro per l'inquinamento atmosferico e stabilisce una soglia annuale di concentrazione limite pari a 5 $\mu\text{g}/\text{m}^3$ per la concentrazione nell'aria, da raggiungere nel 2010. Un lavoro dell'unione europea è stato commissionato nel '98 per esaminare lo stato della concentrazione dell'inquinante e le fonti, propedeutico alla messa a punto della direttiva.

Il testo completo della direttiva si riporta nell'allegato 2 alla presente tesi.

A livello europeo, l'80, 85% delle emissioni di benzene è da attribuire al traffico veicolare; le concentrazioni sono più alte nelle aree densamente popolate con densità di traffico rilevante. Le combustioni domestiche coprono solo il 3, 7 % delle emissioni di benzene. Tuttavia possono rintracciarsi chiare correlazioni geografiche, quali le maggiori emissioni nei paesi del nord Europa, dove l'utilizzo della legna come combustibile domestico è molto comune; in Svezia ad esempio il contributo della combustione domestica rappresenta il 50%. L'evaporazione durante la distribuzione di carburante è anche correlata con la concentrazione dell'inquinante; l'unica fonte naturale significativa è costituita dalla combustione di biomasse e da incendi forestali che non vanno ad influenzare la qualità dell'aria nelle aree densamente popolate dell'unione europea.

La rimozione chimica dall'atmosfera può essere portata avanti solo dalla reazione con il radicale idrossido. La sua foto ossidazione contribuisce alla formazione dell'ozono, la reattività del benzene è tuttavia lenta. Un tempo di vita di qualche giorno in condizioni di concentrazioni rappresentative di radicali ossidrilici è sufficiente a trasportare il benzene per lunghe distanze, elemento che giustifica la presenza di elevate concentrazioni di benzene ravvisate in aree "remote" del suolo europeo.

L'abbattimento chimico è importante a livello regionale, ma assume una rilevanza maggiore a livello urbano. Il tempo di residenza di una massa d'aria varia dall'ordine di qualche minuto, a quello di qualche ora nelle aree urbane; sempre in condizioni di pressione stagnante che sono frequenti nel sud dell'Europa.

1.4. STRUMENTI GEOSTATISTICI

1.4.1. Approccio geostatistico nella caratterizzazione spaziale dei fenomeni ambientali e ricostruzione spaziale dei fenomeni tramite stima.

La tipologia dei fenomeni ambientali è molto ampia. Essi vanno da quelli più semplici, che si manifestano nello spazio e sono caratterizzati da una sola variabile, a quelli più complessi, regolati da un gran numero di variabili, che interagiscono tra loro e che sono definite non solo nello spazio, ma anche nel tempo. Le operazioni di tipo geostatistico sono fundamentalmente dirette alla caratterizzazione spaziale o temporale di tali fenomeni e alla loro modellizzazione e stima, ed i risultati delle operazioni, oltre a rappresentare elementi oggettivi per l'interpretazione e comprensione dei fenomeni, trovano impiego nella messa a punto dei sistemi di supporto alle decisioni e nella progettazione e gestione dei sistemi di monitoraggio. In molte applicazioni che riguardano l'igiene ambientale, spesso si pone il problema di stimare la probabilità che la concentrazione di un inquinante, o di un parametro da tenere sotto controllo, superi, in un punto del tempo e dello spazio, un determinato valore di soglia.

L'approccio della Geostatistica è di tipo probabilistico. In questo quadro, il particolare valore $z(x_0)$ assunto da una variabile di studio nel punto di posizione x_0 all'interno del campo di lavoro, è considerato essere una realizzazione di una Variabile Aleatoria (VA) $Z(x_0)$, definita in quel punto. Estendendo quest'assunzione a tutti i punti del campo, risulta definito un insieme di VA che costituisce una Funzione Aleatoria (FA). Poiché le VA che la definiscono hanno carattere spaziale, la FA risulterà caratterizzata da una legge spaziale, che deve essere in grado di interpretare la variabilità riscontrata nei punti di misura.

La funzione deterministica $z(x)$ che descrive l'andamento dei valori reali all'interno del campo, che nella terminologia geostatistica è chiamata Variabile Regionalizzata (VR), costituisce pertanto una realizzazione della FA $Z(x)$. Nella sua definizione più generale e più completa, la legge spaziale della FA $Z(x)$ definita nel campo S è espressa dall'insieme di tutte le funzioni di distribuzione multivariabili:

$$F(z_1, z_2, \dots, z_k) = \text{prob}\{Z_1 < z_1; Z_2 < z_2 \dots Z_k < z_k\}$$

che si possono definire per ogni valore intero di k e per ogni maniera di scegliere in S i k punti. Se, fissato k e la posizione dei k punti x_1, x_2, \dots, x_k , la funzione di distribuzione $F(z_1, z_2, z_k)$ è invariante per traslazione, cioè se si mantiene identica anche nei punti di posizione $x_1+b, x_2+b, \dots, x_k+b$, con b vettore qualsiasi, la FA è stazionaria. Una FA descritta da un variogramma stazionario costituisce operativamente il modello più semplice per interpretare in senso probabilistico una VR. La funzione variogramma viene stimata coi dati di campionatura attraverso il calcolo, nelle diverse direzioni dello spazio, dei variogrammi sperimentali, ed è così definito:

$$\gamma(x_0, h) = \frac{1}{2} \text{var}\{Z(x_0) - Z(x_0 + h)\}$$

Nel caso in cui vi è stazionarietà il variogramma, funzione solo di h , è dato da:

$$\gamma(h) = \frac{1}{2} E \left\{ [Z(x_0) - Z(x_0 + h)]^2 \right\}$$

Se i variogrammi sperimentali mostrano l'esistenza di un valore di soglia vuol dire che una FA stazionaria interpreta appropriatamente, quanto alla variabilità spaziale, il fenomeno di studio. Il comportamento nell'origine del variogramma è responsabile della dispersione alla piccola scala della variabile di studio. Esso può essere discontinuo, lineare o parabolico e indica, rispettivamente, un andamento di tipo discontinuo, continuo ma non derivabile o derivabile, il tutto in media quadratica, della VR. La discontinuità nell'origine è chiamata, e se ne intuisce la ragione, *effetto pepita*.

La variabilità dei parametri che descrivono i fenomeni ambientali è influenzata da fattori strutturali che spesso agiscono in forma anisotropa, cioè con modalità e caratteristiche dipendenti dalle direzioni dello spazio. Come conseguenza, si avrà che anche i parametri coinvolti nei processi avranno carattere anisotropo. Ecco il motivo per cui è stato introdotto il vettore h .

Nel caso di fenomeni multivariabili, l'approccio geostatistico non cambia concettualmente, mentre il modello di FA è più complesso, in quanto in ogni punto x_0 sono definite tante variabili aleatorie $Z_i(x_0)$ quante sono le variabili del problema.

Come per il caso monovariabile, si definisce il variogramma incrociato:

$$\gamma_{ij}(h) = E \left\{ [Z_i(x) - Z_i(x+h)] \cdot [Z_j(x) - Z_j(x+h)] \right\}$$

L'insieme dei variogrammi diretti (quando $i=j$) e incrociati (quando $i \neq j$) costituisce in questo caso la legge spaziale di un fenomeno multivariabile descritto da una FA stazionaria.

L'analisi di regionalizzazione e correionalizzazione, il kriging e il cokriging sono le metodologie della geostatistica lineare e stazionaria riconducibili all'approccio sopra descritto.

La costruzione di una carta che rappresenti l'andamento di una variabile, effettuata a partire da valori puntuali della stessa misurati in alcuni punti del campo, è un'operazione abbastanza usuale nella pratica della modellistica ambientale.

1.4.1.1. Stime in caso di stazionarietà

Il metodo geostatistico per la stima del valore di una variabile riferita a qualsiasi entità geometrica, puntuale o areale, è il *Kriging Ordinario* e si riferisce ad uno stimatore lineare, che per l'appunto è una combinazione lineare dei valori della variabile noti in prossimità dell'entità da stimare.

1.4.1.1.1. Krigaggio Ordinario

Se l'entità da stimare è il valore $Z(x_0)$ della variabile nel punto x_0 , e $Z(x_\alpha)$ sono i valori delle variabili nei punti di misura x_α ($\alpha=1, N$), lo stimatore lineare di $Z(x_0)$, indicato con $Z^*(x_0)$, sarà:

$$Z^*(x_0) = \sum_{\alpha=1}^N \lambda_\alpha Z(x_\alpha)$$

Nella precedente espressione λ_α sono i ponderatori della combinazione lineare. Si vuole innanzitutto che la stima sia corretta, e pertanto deve essere: $E[Z(x) - Z^*(x_0)] = 0$, il che implica:

$$\sum_{\alpha=1}^N \lambda_\alpha = 1.$$

I ponderatori possono essere scelti in maniera tale da rendere minima la suddetta varianza di stima con il rispetto della condizione precedente. L'applicazione del metodo di ottimizzazione condizionato porta al sistema seguente, espresso in termini di variogramma:

$$\sum_{\beta=1}^N \lambda_\beta \gamma(x_\alpha - x_\beta) + \mu = \gamma(x_\alpha - x_0) \quad \forall \alpha = 1, N$$

$$\sum_{\beta=1}^N \lambda_\beta = 1$$

Il precedente è un sistema lineare di $N+1$ equazioni in $N+1$ incognite. Quest'ultime sono i ponderatori λ_α e il parametro di ottimizzazione di Lagrange μ . I coefficienti $\gamma(x_\alpha - x_\beta)$ del sistema ed il termine noto $\gamma(x_\alpha - x_0)$ sono i valori della funzione variogramma per le distanze $(x_\alpha - x_\beta)$ e $(x_\alpha - x_0)$. La varianza di stima σ , diventa:

$$\sigma_k = \sum_{\alpha=1}^N \lambda_\alpha \gamma(x_\alpha - x_0) + \mu$$

Nel sistema precedente la matrice dei coefficienti $\gamma(x_\alpha - x_\beta)$ esprime la struttura dell'informazione, mentre i termini noti, caratterizzano l'entità da stimare. La varianza di stima è una misura della precisione delle stima. Essa varia con il punto da stimare e dipende dal numero e dalla disposizione geometrica dei punti campionati impiegati nella stima, oltre che dalla variabilità generale del fenomeno, espressa dal variogramma.

Se ciò che si vuole stimare, con la stessa base informativa, non è un'entità puntuale, ma il valore medio della variabile Z su un supporto areale

$$Z_A = \frac{1}{A} \int_A Z(x) dx$$

I ponderatori del kriging relativi alla stima:

$$[Z_A]^* = \sum_{\beta=1}^N \lambda_\beta(A) Z(x_\beta)$$

si ottengono dalla soluzione di un sistema strutturalmente analogo al precedente, ma che presenta, essendo cambiata l'entità da stimare, una differenza nei termini noti:

$$\sum_{\beta=1}^N \lambda_\beta(A) \gamma(x_\alpha - x_\beta) + \mu = \frac{1}{A} \int_A \gamma(x_\alpha - x) dx$$

$$\sum_{\beta=1}^N \lambda_\beta(A) = 1$$

La varianza di stima corrispondente è data da:

$$\sigma_k = \sum_{\alpha=1}^N \lambda_{\alpha} \gamma(x_{\alpha} - x_0) - \gamma(A, A) + \mu$$

**1.4.1.1.2. Stima con il contributo di variabili ausiliarie:
cokrigaggio**

Nel caso multivariabile è possibile effettuare la tipologia di stima vista sopra, facendo intervenire anche altre variabili, che si possono considerare ausiliarie, correlate con quella da stimare. Lo stimatore lineare corrispondente è il *cokriging*. Si consideri un insieme di nv variabili e N_i ($i=1, nv$) misure per ognuna di esse e si supponga di voler stimare il valore della variabile i_0 nel punto x_0 . Lo stimatore di cokriging è della forma:

$$Z_{i_0}^*(x_0) = \sum_{i=1}^{nv} \sum_{\alpha=1}^{N_i} \lambda_{\alpha}^i Z_i(x_{\alpha})$$

e il sistema di che consente di ottenere la matrice dei pondratori è il seguente:

$$\sum_{j=1}^{nv} \sum_{\beta=1}^{N_j} \lambda_{\beta}^j \gamma_{ij}(x_{\alpha} - x_{\beta}) + \mu_i = \gamma_{ii_0}(x_{\alpha} - x_0) \quad \text{per } i=1, nv \quad \text{e } \alpha=1, N_i$$

$$\sum_{\beta=1}^{N_{i_0}} \lambda_{\beta}^{i_0} = 1$$

$$\sum_{\beta=1}^{N_i} \lambda_{\beta}^i = 0 \quad \text{per } i=1, nv \quad \text{e } i \neq i_0$$

La corrispondente varianza di stima è data dalla seguente espressione.

$$\sigma_{ck}^2 = \sum_{i=1}^{nv} \sum_{\alpha=1}^{N_i} \lambda_{\alpha}^i \gamma_{ii_0}(x_{\alpha} - x_0) + \mu_{i_0}$$

Il ricorso al cokriging, per la ricostruzione di una variabile si rende necessario in due circostanze:

- per migliorare la precisione della stima di una variabile, quando si hanno a

disposizione misure di una variabile ausiliaria su campioni diversi da quelli su cui è misurata la variabile di interesse;

- per garantire alle stime relazioni lineari di coerenza, i valori stimati con il cokriging garantiscono questa proprietà, non garantita invece dalle stime effettuate con il kriging. Quando si è in presenza di variabili tra di loro non correlate, il cokriging di una qualunque di esse non presenta alcun miglioramento in precisione rispetto al kriging. I risultati della stima sono gli stessi e pertanto diventa inutile effettuare il cokriging.

1.4.1.2. Stime nel caso di non stazionarietà

Quando la variabile o le variabili di studio non possono essere descritte da FA stazionarie (il caso più classico è quando vi è la presenza di un *trend* o *deriva*) allora bisogna ricorrere ai metodi di stima della Geostatistica non stazionaria.

1.4.1.2.1. Stima con il contributo di variabili ausiliarie:Kriging con deriva esterna

Quando è disponibile, come informazione ausiliaria correlata alla variabile di studio, una variabile $s(x)$ nota su tutto il campo e quando non ricorrono le condizioni di stazionarietà (nel qual caso la metodologia di stima è, come si è visto, il *collocated cokriging*), il kriging con deriva esterna consente di operare la ricostruzione della variabile di interesse facendo intervenire la variabile ausiliaria. In questo caso la variabile ausiliaria contribuisce alla stima fornendo l'andamento generale su tutto il campo della variabile di studio, difficilmente ricostruibile, data la non stazionarietà, con le sole misure della variabile diretta. Deve valere questa condizione:

$$\sum_{\alpha=1}^N \lambda_{\alpha} s(x_{\alpha}) = s(x_0)$$

Questa condizione esprime che i ponderatori λ_{α} devono essere consistenti con una interpolazione esatta di $s(x)$.

La variabile ausiliaria nota su tutto il campo prende il nome di *deriva esterna*, da cui il metodo.

Si possono considerare anche più variabili ausiliarie note in tutto il campo, quindi più derive esterne in parallelo, purchè indipendenti. Siano esse $s_i(x)$, $i=1,ND$. Il sistema di kriging avrà la seguente forma:

$$\sum_{\beta=1}^N \lambda_{\beta} K(x_{\alpha} - x_{\beta}) + \sum_l \mu_l f^l(x_{\alpha}) + \sum_i \mu_i s_i(x_{\alpha}) = K(x_{\alpha} - x_0) \quad \forall \alpha = 1, n$$

$$\sum_{\alpha=1}^N \lambda_{\alpha} f^l(x_{\alpha}) = f^l(x_0) \quad \forall l = 1, L_k$$

$$\sum_{\alpha=1}^N \lambda_{\alpha} s_i(x_{\alpha}) = s_i(x_0) \quad \forall i = 1, ND$$

La varianza di stima è data da:

$$\sigma_k^2 = K(0) - \sum_{\alpha=1}^N \lambda_{\alpha}^u K(x_{\alpha} - x_0) - \sum_l \mu_l f^l(x_0) - \sum_i \mu_i s_i(x_0)$$

1.4.2. Analisi in componenti principali

L'analisi delle componenti principali è una tecnica per la semplificazione dei dati utilizzata in ambito della statistica multivariata. Insieme all'analisi delle corrispondenze e all'analisi delle corrispondenze multiple, appartiene all'analisi fattoriale.

Lo scopo primario di questa tecnica è la riduzione di un numero più o meno elevato di variabili (rappresentanti altrettante caratteristiche del fenomeno analizzato) in alcune variabili latenti. Ciò avviene tramite una trasformazione lineare delle variabili che proietta quelle originarie in un nuovo sistema cartesiano nel quale la nuova variabile con la maggiore varianza viene proiettata sul primo asse, la variabile nuova seconda per dimensione della varianza sul secondo asse e così via. La riduzione della complessità avviene limitandosi ad analizzare le principali (per varianza) tra le nuove variabili.

Diversamente da altre trasformazioni (lineari) di variabili praticate nell'ambito della statistica, in questa tecnica sono gli stessi dati che determinano i vettori di trasformazione.

Assumendo che a ciascuna delle variabili originarie venga sottratta la loro media e pertanto la nuova variabile (X_i) ha media nulla:

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} E \left\{ (\mathbf{w}^T \mathbf{x})^2 \right\}$$

Con i primi (k-1) componenti, il k-esimo componente può essere trovato sottraendo i primi (k-1) componenti principali a X :

$$\hat{\mathbf{x}}_{k-1} = \mathbf{x} - \sum_{i=1}^{k-1} \mathbf{w}_i \mathbf{w}_i^T \mathbf{x}$$

e sostituendo questo:

$$\mathbf{w}_k = \arg \max_{\|\mathbf{w}\|=1} E \left\{ (\mathbf{w}^T \hat{\mathbf{x}}_{k-1})^2 \right\}$$

Un metodo più semplice per calcolare la componente \mathbf{w}_i utilizza la matrice delle covarianze di \mathbf{x} . Trovando l'autovalore e l'autovettore della matrice di covarianza, si ottiene che l'autovettore con il maggiore autovalore corrisponde alla dimensione che ha la maggiore correlazione con l'insieme di dati. I dati originali sono infine proiettati nello spazio vettoriale ridotto.

1.4.3. Software ILWIS

ILWIS (Integrated Land and Water Information System) è un pacchetto Gis integrato sviluppato dall'ITC *International Institute of Geoinformation Science and Earth Observation*.

ILWIS è nato originariamente a scopi educativi e applicazioni a basso costo nei paesi in via di sviluppo; iniziato nel 1984, la prima versione risale al 1988 e se ne susseguono altre 1996, 2001, 2004 fino ad arrivare all'ultima del luglio 2007 open source. La versione più recente (3.4) offre un pacchetto di *image processing*, vettoriale, raster, geostatistico e statistico, operazioni di database e simili; l'utente può inoltre creare nuovi *script*, personalizzare le operazioni del menu e anche costruire applicazioni in *Visual Basic*, *Delphi* o

C++, che possono essere create esternamente da ILWIS e usare le sue funzioni interne.

Il principale vantaggio di ILWIS è rappresentato dall'essere un pacchetto compatto con diverse funzionalità gis raster o vettoriali. Il grande svantaggio sono virus, instabilità e necessità di importare i dati in formati ILWIS da altri e più popolari pacchetti gis.

1.4.3.1. Geostatistica in ILWIS

ILWIS contiene un pacchetto di funzioni statistiche e geostatistiche. Considerando le possibilità d'interpolazione, può essere usato per analizzare e produrre il variogramma, analizzare l'anisotropia dei dati (grazie al variogramma superficiale), sviluppare Kriging Ordinari e Cokriking (con una variabile ausiliaria), Kriking Universale con le coordinate come predittore e regressione lineare. ILWIS possiede inoltre un numero di algoritmi originali geostatistici che possono essere usati spesso manualmente.

Per esempio offre la possibilità di effettuare un Krigaggio direttamente a partire da mappe raster e il calcolo del variogramma sempre a partire da mappa raster, è anche funzionale allo svolgimento delle analisi statistiche basilari su un insieme di mappe raster (maplist): consente di fare l'analisi delle componenti principali a partire da mappe raster, l'analisi di correlazione tra mappe raster e mappe statistiche a più informazioni (minimo, massimo, media, deviazione standard).

Tuttavia ILWIS non annovera tra le sue potenzialità il regression kriking, ma permette di svolgere analisi simili. Per esempio una tabella può essere importata e convertita a mappa puntuale, che può essere sovrapposta a una mappa raster per analizzare se due variabili sono correlate. Questo può essere fatto estraendo dalla mappa raster de valori puntuali alle coordinate della point map. L'allegato 1 alla presente tesi, contiene un'ampia descrizione dell'architettura e delle funzionalità del programma.

2. CASO STUDIO

2.1. ANALISI DEI DATI DISPONIBILI: THE EUROPEAN AIR QUALITY DATABASE

I dati della concentrazione di benzene alla base del mio studio sono stati prelevati direttamente dal sito della EEA. Selezionando tra i temi dell'agenzia europea dell'ambiente quello dell'inquinamento atmosferico (air pollution), si accede ai collegamenti con i vari datasets disponibili:

- EEA aggregated and gap filled air emission data
- National Emission Ceilings (NEC) Directive Inventory
- National emissions reported to the Convention on Long-range Transboundary Air Pollution (LRTAP Convention)
- EMEP grids reprojected by EEA
- EPER - The European Pollutant Emission Register
- AirBase - The European air quality database
- Trends in emissions of ozone precursors (CLRTAP/EMEP)

Il dataset che è stato considerato è l'AirBase: database europeo della qualità dell'aria.

L'airbase è un sistema mantenuto dalla EEA attraverso i centri europei specializzati sull'aria e i cambiamenti climatici, tali centri sono costituiti da un consorzio di istituti dei diversi stati membri quindi da personale esperto nel settore ingaggiato dalla stessa EEA per assolvere funzioni specifiche all'interno della propria strategia globale. I centri assistono così l'agenzia europea dell'ambiente nella sua azione di supporto alla politica comunitaria, relazionano sugli sviluppi della politica ambientale in tema di qualità dell'aria, emissioni e cambi climatici, partecipando attivamente alle redazioni della EEA sia fornendo dati sullo stato dell'ambiente sia armonizzando la rete europea di monitoraggio.

2.1.1 Rete, stazioni, statistiche AirBase

I file contenenti le informazioni riguardanti i campionamenti atmosferici dell'AirBase sono disponibili sia per singolo paese che per l'intero territorio europeo e sono divisi in tre blocchi: rete (network), stazioni (station), statistiche (statistics).

Nell'AirBase network sono contenute le informazioni relative al nome e al codice delle stazioni di monitoraggio; nome, codice e tipo del network; data di inizio delle registrazioni; nome, indirizzo e fax dell'organizzazione responsabile della rete; nome e contatto e-mail del rappresentante dell'organizzazione.

Nell'AirBase stations oltre alla ripetizione del nome e codice delle stazioni ci sono altre informazioni relative alla posizione geografica delle stazioni stesse, longitudine, latitudine in deg e dms, tipologia della stazione e dell'area, nome della strada e della città di ubicazione.

Nell'AirBase statistics al codice della stazione è associato il nome del componente monitorato (inquinante), la sua formula e il suo codice, unità di misura, anno di campionamento, supporto temporale (es:orario o giornaliero), nome delle grandezze statistiche calcolate (es: media annuale, valore massimo, diversi percentili (50, 95, 98, percentile), quarto valore più alto ecc.), valore statistico di tali grandezze, percentuale dei valori validi.

2.2.ESTRAZIONE DEI DATI UTILI

La prima operazione del mio lavoro è stata quella di elaborare i file dati a disposizione, divisi nei tre blocchi sopraesposti, in modo da ottenere un unico file con caratteristiche appropriate per le successive analisi in ambiente ILWIS; tale operazione è stata svolta con l'ausilio di software per la gestione dei dati quali Microsoft Access e Microsoft Excel.

L'inquinante studio scelto in accordo con il Jrc è stato quindi il benzene, la grandezza statistica: la media annuale su base giornaliera, ossia a partire dai

valori medi giornalieri. La media annuale, rispetto alle altre grandezze disponibili nell’AirBase statistics, quali ad esempio percentili, valore massimo, quarto valore più alto ecc., è quella che meglio si presta alla caratterizzazione della concentrazione media dell’inquinante su scala europea; per quanto riguarda invece la decisione tra supporto temporale ora o giorno, premesso che non vi sono rilevanti differenze tra medie annuali calcolate a partire da medie orarie o da medie giornaliere, ho optato per quest’ultime.

Ho strutturato il file dati unico in modo tale che contenesse in una colonna i nomi di tutte le stazioni in cui è stato monitorato il benzene nell’arco dell’intero periodo di campionamento (dal 1993 al 2005), e in un numero di colonne pari agli anni nei quali è stata campionata tale variabile (13), i valori della media annuale su base giornaliera in corrispondenza delle stazioni monitorate in quell’anno e un -1 per quelle non monitorate.

La figura seguente mostra il formato del file dati ottenuto.

stazioni totali	benz 93	benz94	benz95	...
AT0002R	-1	-1	-1	...
AT0009A	-1	-1	-1	...
AT0012A	-1	-1	-1	...
AT0036A	-1	-1	-1	...
AT0039A	-1	-1	-1	...
AT0140A	-1	-1	-1	...
AT0217A	-1	-1	-1	...
BE0197A	-1	-1	-1	...
BE0228A	-1	-1	-1	...
BE0296A	-1	-1	-1	...
BE0323A	-1	11,071	8,404	...
BE0324A	-1	2,9	2,435	...
BE0325A	-1	2,587	2,211	...
BE0326A	-1	2,955	1,753	...
BE0327A	-1	3,426	3,122	...
BE0339A	-1	-1	3,341	...
BE0340A	-1	-1	2,225	...
BE0341A	-1	-1	1,461	...
BE0351A	-1	-1	-1	...
...
...
...

Alla colonna delle stazioni totali ho poi associato altre tre colonne contenenti le coordinate geografiche per la localizzazione spaziale, latitudine, longitudine e altitudine, informazioni estratte dall'AirBase stations. La tabella seguente riporta il formato definitivo assunto dal file.

stazioni totali	LAT	LON	ALT	benz 93	benz94	benz95	...
AT0002R	47,76944	16,76556	117	-1	-1	-1	-1 ...
AT0009A	48,209	16,374	173	-1	-1	-1	-1 ...
AT0012A	48,165	14,03556	316	-1	-1	-1	-1 ...
AT0036A	48,189	16,302	195	-1	-1	-1	-1 ...
AT0039A	48,27445	14,31472	265	-1	-1	-1	-1 ...
AT0140A	48,125	16,33167	194	-1	-1	-1	-1 ...
AT0217A	47,04194	15,43417	345	-1	-1	-1	-1 ...
BE0197A	51,05917	3,73056	5	-1	-1	-1	-1 ...
BE0228A	51,21056	4,43306	6	-1	-1	-1	-1 ...
BE0296A	50,94083	5,36944	41	-1	-1	-1	-1 ...
BE0323A	50,8297	4,3761	70	-1	11,071	8,404	...
BE0324A	51,3253	4,3583	4	-1	2,9	2,435	...
BE0325A	51,0625	5,095	32	-1	2,587	2,211	...
BE0326A	51,06889	5,11306	28	-1	2,955	1,753	...
BE0327A	51,19667	3,82417	6	-1	3,426	3,122	...
BE0339A	51,3178	4,2628	4	-1	-1	3,341	...
BE0340A	50,9656	5,4319	20	-1	-1	2,225	...
BE0341A	50,75806	5,07583	25	-1	-1	1,461	...
BE0351A	51,09083	4,35611	10	-1	-1	-1	...
...
...

2.2.1. Importazione dei dati in ILWIS

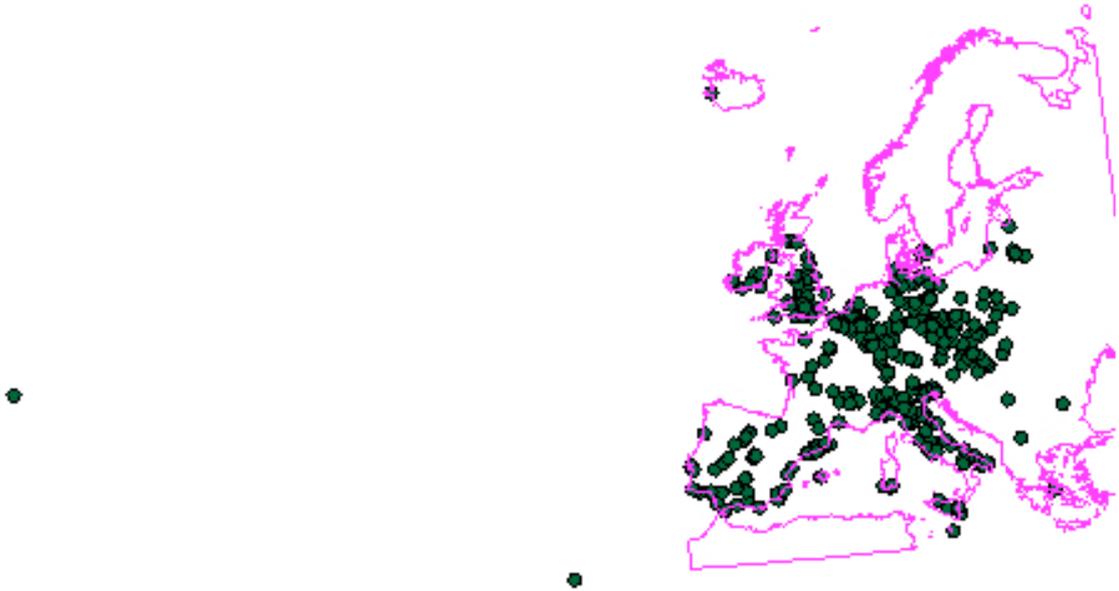
Tutte le operazioni descritte nel paragrafo precedente sono state propedeutiche all'importazione della tabella dati ottenuta in ILWIS, previa conversione in formato DBF.

La tabella in ILWIS è visualizzabile in una finestra dedicata e rimane fondamentalmente immutata. Il programma classifica ogni colonna in funzione di quello che contiene nella prima riga e non essendoci campi vuoti nel file input, opportunamente riempiti con dei -1, tutte le colonne vengono riconosciute come numeriche; tale formato è necessario per poter eseguire le elaborazioni statistiche a partire dai dati contenuti nella tabella e permette che per ogni colonna venga calcolato automaticamente lo *statistics pane*, ossia un insieme di grandezze statistiche basilari quali: minimo, massimo, media, deviazione standard, somma. La presenza dei -1 tuttavia falsifica i valori delle grandezze statistiche così ottenute e compromette l'utilizzo delle colonne stesse in altre operazioni; mediante un comando di "tab calc" ho quindi provveduto ad eliminare i dati se minori di zero, tale operazione ha consentito di ottenere un complesso di 13 colonne di formato numerico ma con valori indeterminati in corrispondenza delle stazioni di volta in volta non monitorate per i diversi anni.

2.3. VISUALIZZAZIONE SPAZIALE DEI DATI

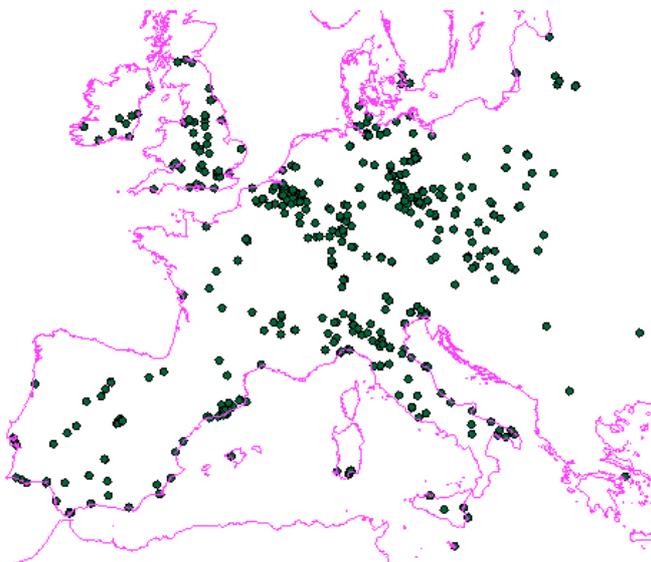
Il primo passo della mia analisi è stato quello di valutare la dislocazione spaziale sul suolo europeo delle stazioni complessive e di quelle monitorate nei diversi anni, al fine di visualizzare e analizzare la geografia del caso studio, la densità di campionamento, la distanza tra le stazioni, la presenza di eventuali stazioni non rappresentative. In ambiente ILWIS tale operazione è possibile mediante una conversione da *table* a *point map*; nello specifico della visualizzazione delle stazioni totali: impostando come dominio della tabella un record numerico e come sistema di coordinate il *lat-lon* in accordo con

quello del file input, si ottiene una mappa puntuale in cui ogni stazione è rappresentata da un punto.



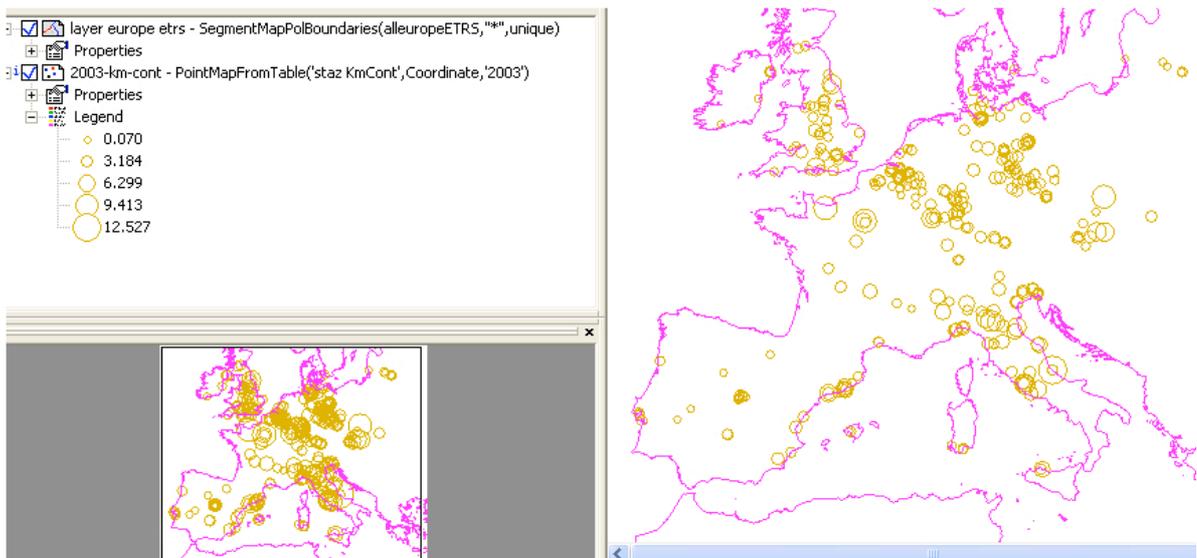
Point map delle stazioni totali: 534 stazioni monitorate nei 13 anni (dal 1993 al 2005)

Il profilo europeo è ben riconoscibile e si distinguono nettamente 3 stazioni non continentali quali: ES1131A, FR1317A, IS005A, (Canarie, Azzorre, Islanda), che ho scartato dall'analisi in quanto poco rappresentative e in modo da poter concentrare lo studio sull'Europa continentale.

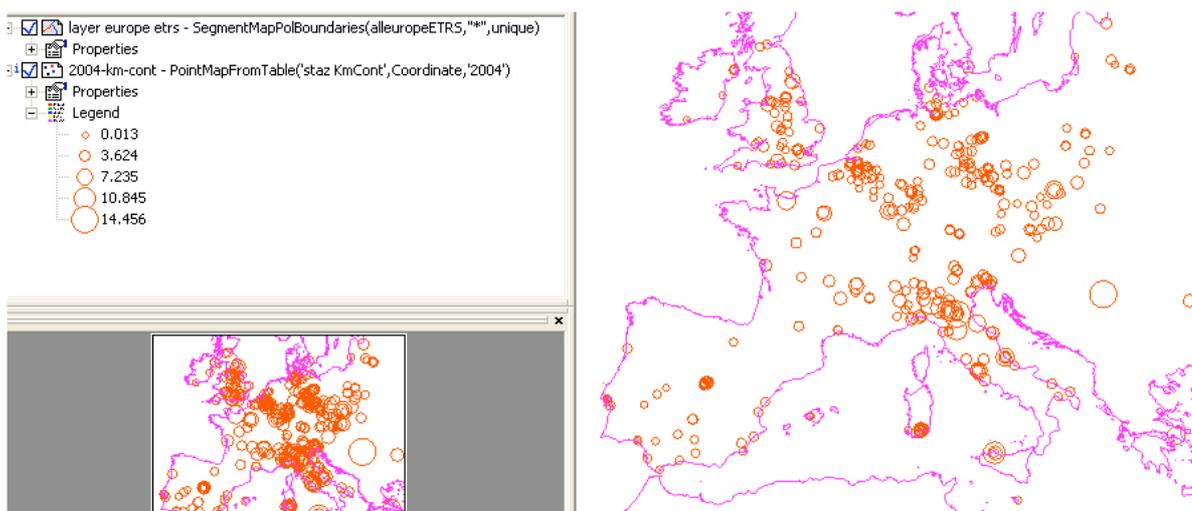


Point map delle stazioni totali continentali

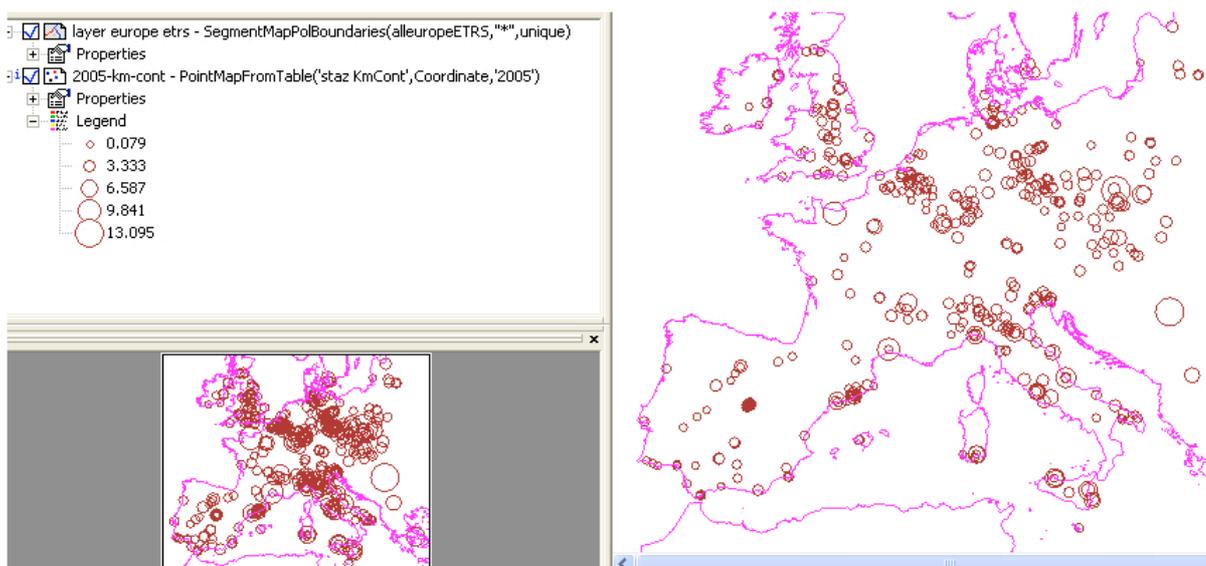
Sono stati presi in considerazione gli ultimi tre anni del database 2003, 2004, 2005, perché oltre ad essere i più recenti sono anche quelli più campionati rispettivamente con 294, 345, 423 dati su un totale di 534 stazioni monitorate nei 13 anni (dal 1993 al 2005). Il passo successivo è stato quindi quello di visualizzare spazialmente le stazioni monitorate per tali anni, che operativamente è consistito nel creare 3 *point map* a partire dalla tabella dati complessiva; contrariamente a quanto fatto nel passaggio da *table* a *point map* per la visualizzazione spaziale delle stazioni totali, ho impostato come dominio delle mappe una colonna della tabella (rispettivamente 2003, 2004, 2005) e non un record numerico. In tal modo, introducendo le colonne contenenti le coordinate delle stazioni nello specifico campo, si riescono ad ottenere tre mappe puntuali dove ogni punto assume oltre al valore geografico della stazione, il valore numerico della variabile campionata in quel punto; ovviamente vengono visualizzati solo i punti in corrispondenza dei quali si hanno valori numerici nella colonna di dominio, quindi solo le stazioni campionate per l'anno in esame.



Point map delle stazioni monitorate nel 2003



Point map delle stazioni monitorate nel 2004



Point map delle stazioni monitorate nel 2005

Il passaggio da *table* a *point map*, che per ogni punto forniscono oltre all'informazione spaziale anche quella quantitativa del valore della variabile, è fondamentale in ambiente ILWIS dal momento che partendo da tali mappe si possono sviluppare tutte le operazioni statistiche di cui il software dispone. Relativamente al sistema di coordinate adottato, ho scelto l' ETRS LAEA 1989 (sistema di coordinate chilometrico) e quindi convertito le mappe

inizialmente ottenute in coordinate geografiche (lat-lon), mediante l'operazione vettoriale di *transform point* disponibile in ILWIS.

2.4. STATISTICHE DI BASE

Alla valutazione geografica segue nell'analisi il calcolo di statistiche di base preliminari quali:

1. minimo, massimo, media, deviazione standard;
2. istogramma del numero di punti in funzione del valore assunto dalla variabile;
3. nuvole di correlazione;
4. matrice di varianza e covarianza;

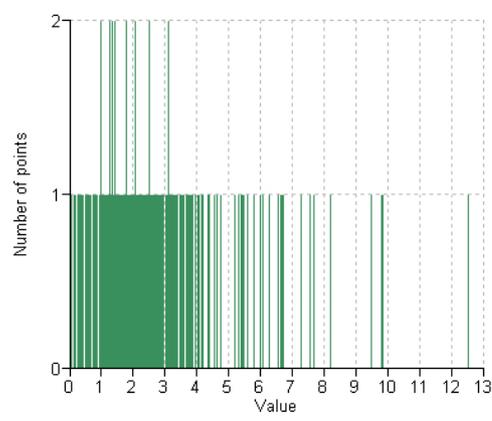
Il calcolo di tali grandezze, preliminare all'analisi strutturale mediante variogramma, rappresenta parte integrante e imprescindibile dello studio della variabilità spazio temporale in quanto consente una prima valutazione di massima sulla natura statistica della variabile e sulla correlazione tra i valori assunti nei diversi anni esaminati.

Le operazioni che ho svolto in ambiente ILWIS volte all'ottenimento delle statistiche di base sono di seguito riportate congiuntamente ai risultati ottenuti.

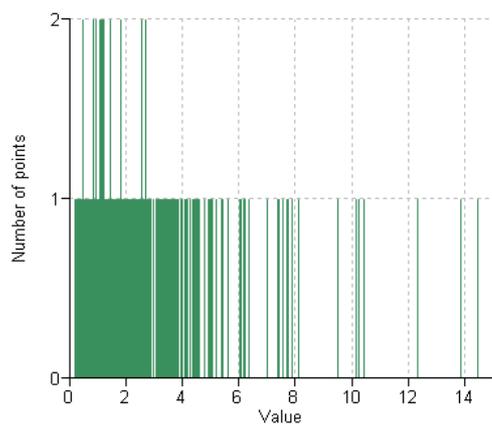
Minimo, massimo, media, e deviazione standard per i valori della variabile anno per anno, vengono calcolati direttamente nello *statistic pane* a partire dalla tabella dati per le colonne in formato *value*, come esposto contestualmente all'importazione e alla formattazione della tabella dati.

Statistiche	2003	2004	2005
Min	0.070	0.013	0.079
max	12.527	14.456	13.095
media	2.445	2.335	1.933
Dev.standard	1.777	2.047	1.662

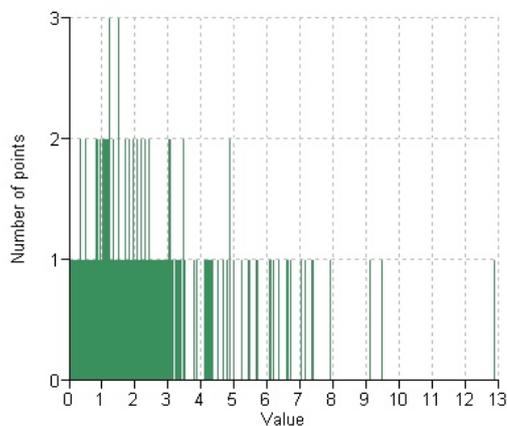
Come per lo *statistic pane* anche il calcolo e la visualizzazione degli istogrammi : valore-numero di punti, non necessita in ambiente ILWIS di operazioni o comandi; sono forniti direttamente in associazione alla creazione di *point map*.



Istogramma 2003



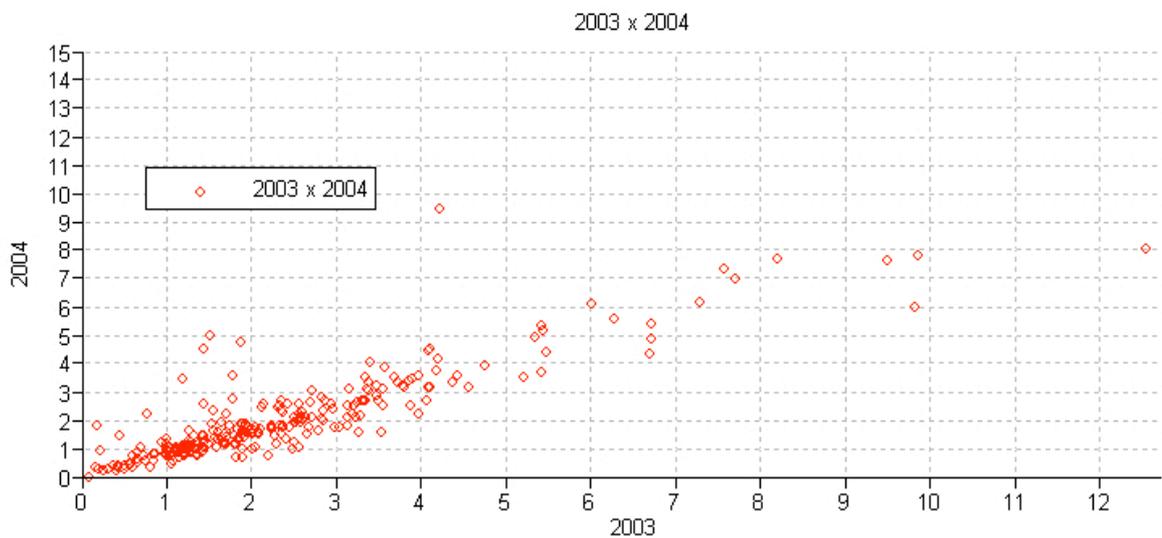
Istogramma 2004



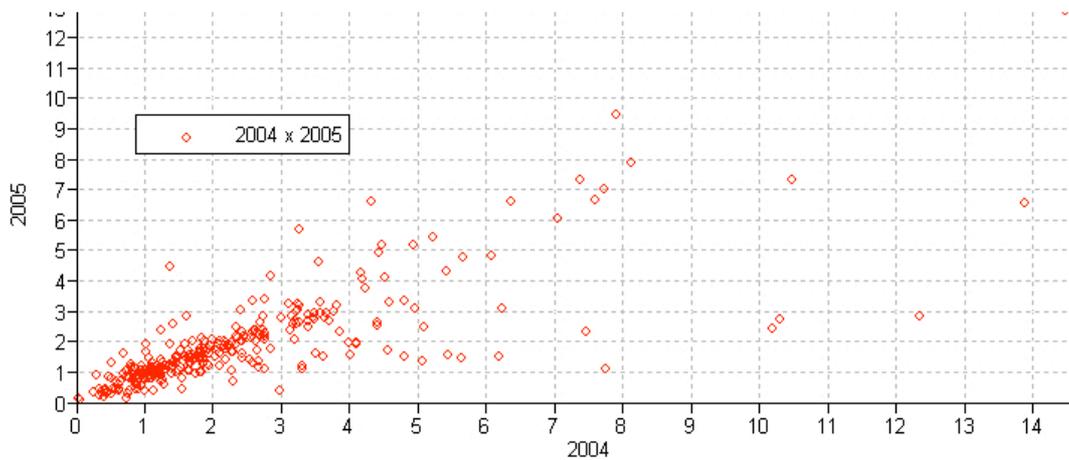
Istogramma 2005

Gli istogrammi riportati mostrano l'andamento del numero di punti in funzione del valore di concentrazione, forniscono così un'informazione relativa alla frequenza che la variabile assuma determinati valori e consentono un immediato riscontro grafico delle statistiche di massimo, minimo e media in accordo con quanto calcolato nello *statistic pane*.

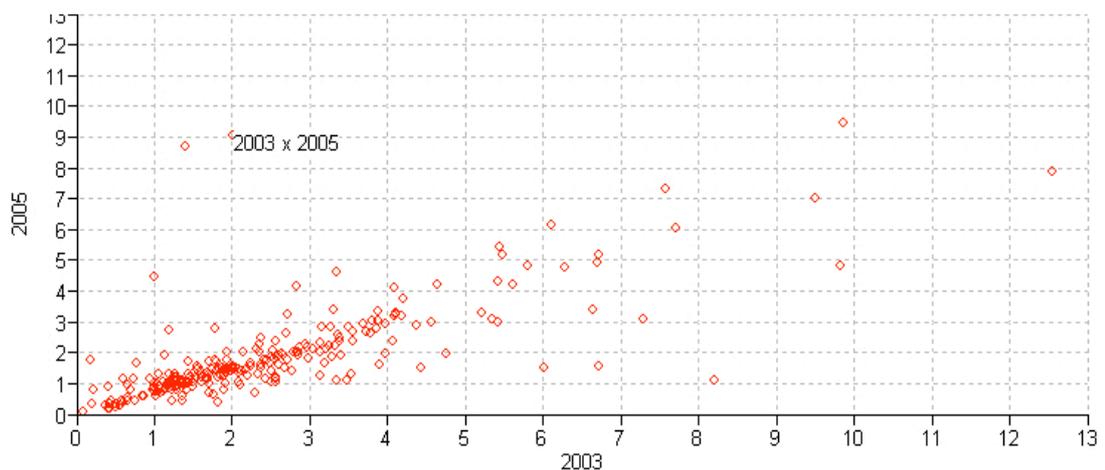
Relativamente ai valori assunti dalla variabile nei diversi anni, dal punto di vista di un'analisi preliminare è utile sviluppare le cosiddette nuvole di correlazione, ossia dei diagrammi cartesiani contenenti in ascissa e ordinata i valori della variabile riferiti a due anni diversi. In tali diagrammi vengono prese in considerazione e graficate solo le stazioni isotopiche dei due anni di volta in volta esaminati e dal momento che il 2003, 2004, 2005, sono gli anni più campionati, la significatività delle nuvole di correlazione ottenute non è compromessa; in tutte le combinazioni: 2003-2004, 2004-2005, 2003-2005, il numero di stazioni isotopiche risulta superiore a 100.



Nuvola di correlazione 2003/2004



Nuvola di correlazione 2004/2005



Nuvola di correlazione 2003/2005

L'andamento delle nuvole di correlazione evidenzia una forte approssimazione della retta a 45 gradi e nello studio 2003/2005 assume un'inclinazione leggermente inferiore, testimoniando così un abbassamento dei livelli di concentrazione dal 2003 al 2005, in accordo con quanto desumibile da una valutazione della media calcolata nello *statistic pane* che risulta in effetti inferiore per il 2005 rispetto al 2003.

Un'ulteriore grandezza utile alla valutazione della correlazione della variabilità temporale negli anni esaminati, è il calcolo della matrice di

varianza e covarianza; tale matrice contenente sulla diagonale principale i valori della varianza della variabile per diversi anni e nelle altre posizioni i valori della covarianza incrociata, fornisce una visualizzazione più efficace delle soglie di variabilità diretta e incrociata; normalizzando tali valori si ottiene il coefficiente di correlazione che essendo variabile tra 0 e 1 risulta d'interpretazione ancora più immediata.

In ILWIS la matrice di varianza e covarianza è direttamente ottenibile a partire da mappe raster, dal momento che i dati a disposizione non sono in questo formato ho provveduto al calcolo delle grandezze necessarie, varianze e covarianze, sfruttando le operazioni statistiche disponibili da colonna a partire dalla tabella dati complessivi. Il passaggio alla forma matriciale risulta poi immediato e si riportano di seguito i risultati ottenuti.

Matrice var-cov	2003	2004	2005
2003	3.182	2.415	2.065
2004	2.415	4.220	2.441
2005	2.065	2.441	2.773

Matrice di correlazione	2003	2004	2005
2003	1	0.877	0.788
2004	0.877	1	0.748
2005	0.788	0.748	1

3. STUDIO MONOVARIATO:

3.1. ANALISI GEOSTATISTICA DELLA VARIABILITÀ SPAZIALE

3.1.1. Variogrammi sperimentali

L'analisi strutturale della variabilità spaziale di una variabile è basata sul calcolo e sulla modellizzazione del variogramma come ampiamente discusso all'interno del capitolo strumenti geostatistici.

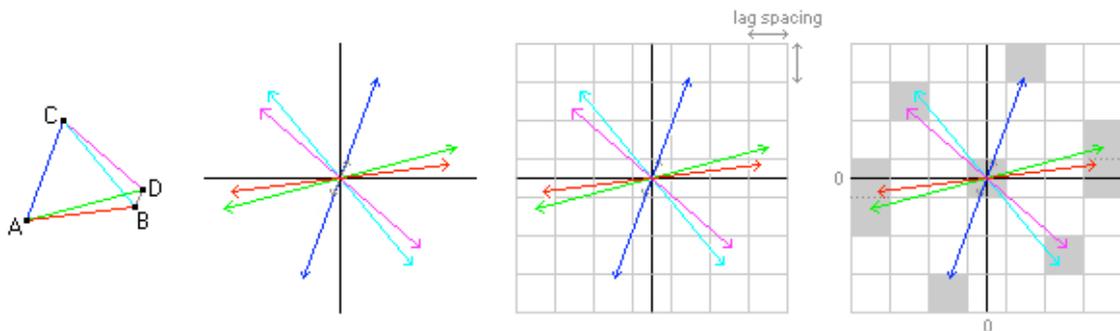
In ambiente ILWIS il valore del semivariogramma sperimentale è ottenibile a partire dalle *point map* mediante un'operazione di *spatial correlation*, che effettua il calcolo in funzione di due parametri imprescindibili quali passo e direzione, introdotti dall'utente a seconda delle caratteristiche del caso studio.

La prima operazione che ho svolto è stata quindi una valutazione della dislocazione delle stazioni sul territorio nei diversi anni, al fine di individuare un valore del passo plausibile con la geografia del caso in esame; a partire dalle *point map* del 2003, 2004, 2005 valutando approssimativamente la massima distanza tra le due stazioni tra loro più lontane, (direttamente sulla mappa mediante un comando di calcolo della distanza), e dividendo tale valore per il numero delle stazioni monitorate in quell'anno, ho ottenuto un valore orientativo del passo compreso tra i 10000 m ai 20000 m.

All'individuazione di tale intervallo di distanze plausibili per il passo, è seguito il calcolo delle mappe variografiche, che consistono in una rappresentazione a 360 gradi dei valori assunti dal variogramma in funzione della distanza e della direzione spaziale alla quale i punti vengono di volta in volta considerati per le diverse classi di distanza. Il calcolo delle mappe variografiche è propedeutico all'individuazione di un'eventuale anisotropia geometrica, dove per anisotropia geometrica si intende un andamento anisotropo della variabilità spaziale ravvisabile dal valore del *range* del variogramma variabile con la direzione, così da assumere un massimo in corrispondenza della direzione definita di anisotropia e un minimo in quella perpendicolare.

In ILWIS il calcolo delle mappe variografiche (*variogram surface*) è possibile a partire dalla *point map* mediante l'operazione statistica omonima: *variogram surface* che fornisce in output un grafico in formato mappa raster, con l'origine nel centro della mappa e in cui ogni cella ha la dimensione del passo specificato dall'utente che fornisce anche il numero di passi opportuno a coprire l'area in esame (a partire dalla cella centrale verso gli estremi degli assi X e Y, positivi e negativi). Nella mappa output ogni pixel contiene così una duplice informazione:

- la classe di distanza direzionale
- il valore del semivariogramma per tutte quelle coppie di punti il cui vettore di separazione termina proprio in quel pixel.



Classi di distanza direzionale nelle mappe variografiche

Nell'interpretazione della mappa ottenuta si parte dalla considerazione che i valori del semivariogramma vicino all'origine siano bassi: in punti a piccola distanza ci si aspetta, infatti, che la variabile assuma circa lo stesso valore. Quando non vi è anisotropia i valori del semivariogramma aumentano dal centro verso l'esterno ugualmente e in modo graduale in tutte le direzioni, come sarà ben visualizzabile dalla scala di colori della mappa raster, in cui si delineano circonferenze di colore concentriche; se siamo in presenza di un'anisotropia geometrica sulla mappa si individuano non più delle circonferenze ma delle ellissi di colore, si distingue quindi una determinata direzione in cui i valori del semivariogramma non aumentano di molto con la

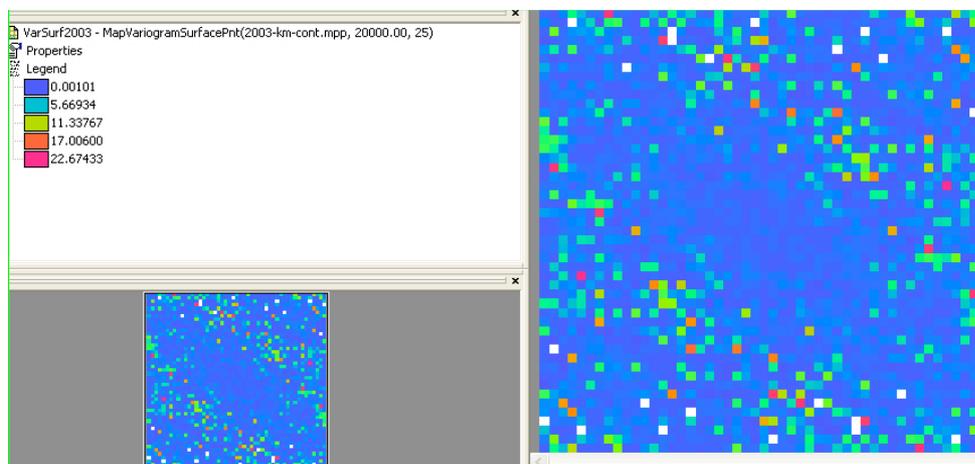
distanza, mentre crescono considerevolmente per piccole distanze in quella perpendicolare, come evidenzierà una rapida gradazione di colore. La direzione di anisotropia, quella per cui il valore del range è massimo, si può calcolare direttamente sulla mappa mediante il comando di misura della distanza che oltre a questa fornisce anche l'angolo tra la direzione della distanza tra i punti e l'asse Y, calcolato in senso orario.

Le mappe variografiche per la variabile concentrazione di benzene negli anni 2003, 2004, 2005, le ho ottenute due volte per ogni anno cambiando i parametri in ingresso, passo e numero di passi, al fine di trovare una soluzione che migliorasse la risoluzione della mappa minimizzando le celle vuote, dal momento che i pixel rappresentanti classi di distanza direzionale per le quali non si incontrano punti, restano indefiniti.

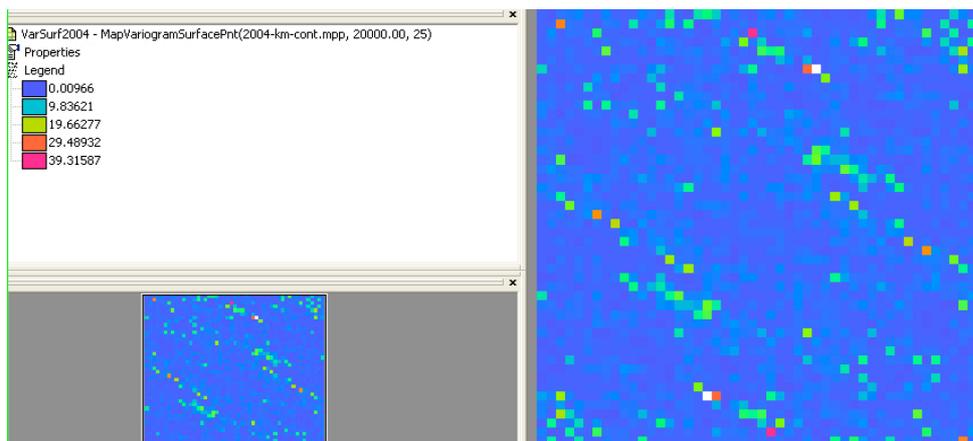
Il calcolo dei due *variogram surface* ha avuto in ingresso i seguenti parametri:

- il primo: passo di 15000 e numero di passi 35;
- il secondo: passo di 20000 m e numero di passi 25;

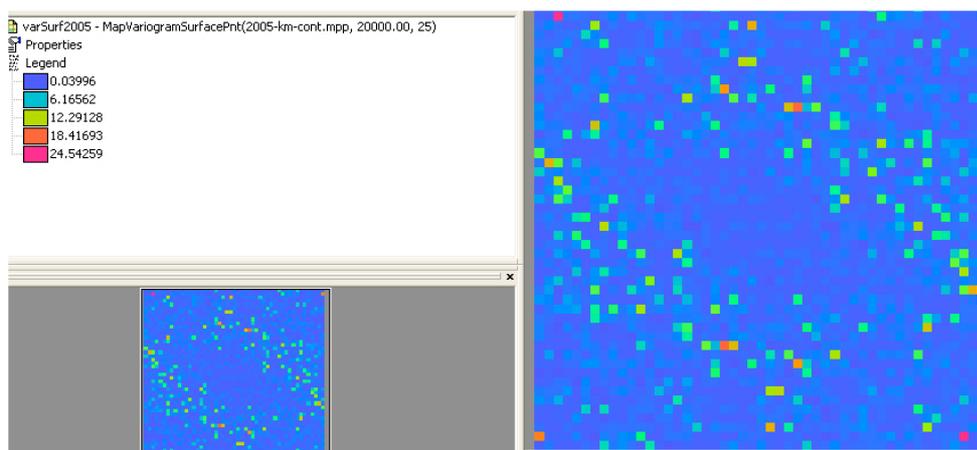
Le due mappe così ottenute non hanno evidenziato sostanziali differenze, la seconda, di seguito riportata, avendo una dimensione del pixel leggermente superiore risulta meglio definita e più facile da interpretare. Si ravvisa una netta anisotropia a 135° per i tre anni, in accordo con la direzione di dispersione atmosferica.



Variogram surface 2003: passo 20000m numero di passi 25



Variogram surface 2004: passo 20000m numero di passi 25



Variogram surface 2005: passo 20000m numero di passi 25

Delineati i parametri passo e direzione di anisotropia del caso studio, si può procedere nell'analisi strutturale alla caratterizzazione della variabilità spaziale mediante il calcolo dei variogrammi che in ambiente ILWIS è compendiato dalla già citata operazione di *spatial correlation* che consente due opzioni:

- Un calcolo omnidirezionale, in cui la distanza tra i punti, (quindi il variogramma), viene calcolata in tutte le direzioni;

- Un calcolo bidirezionale, in cui la distanza tra i punti viene calcolata in una specifica direzione fornita dall'utente (più o meno la tolleranza angolare) e nella direzione perpendicolare.

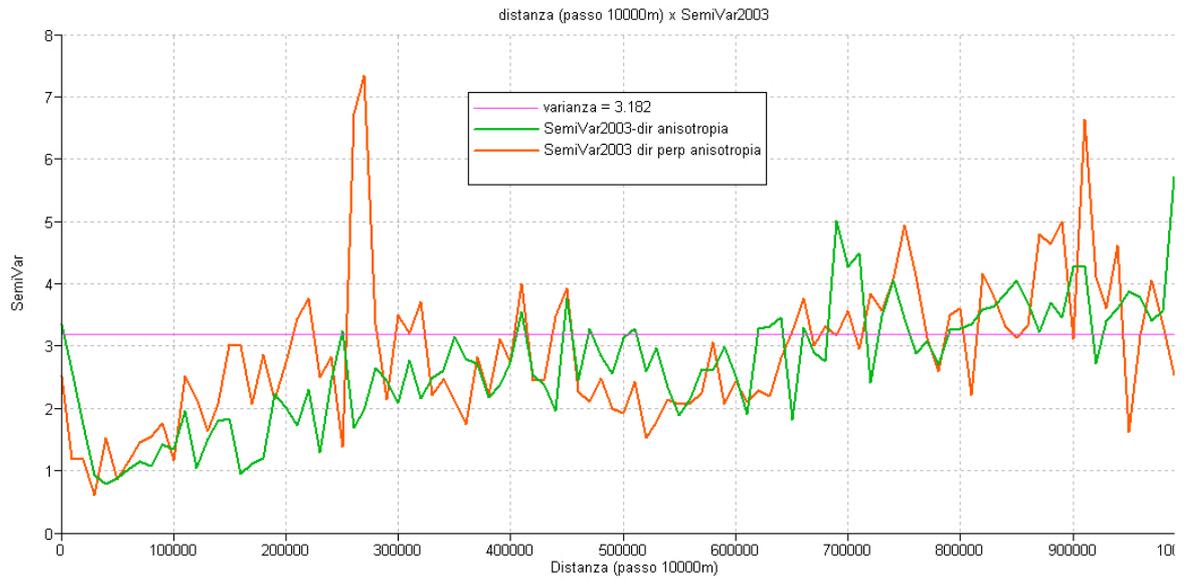
Alla luce dell'anisotropia riscontrata ho proceduto ad un calcolo bidirezionale per i tre anni, fornendo come direzione quella di anisotropia, 135° , una tolleranza angolare di 45° e un passo di 10000 m, ho sviluppato comunque anche il calcolo omnidirezionale sempre al passo di 10000 m, ottenendo il variogramma medio; tale grandezza, più robusta sfruttando la totalità dei dati a disposizione ma teoricamente penalizzante in quanto non conforme alla configurazione anisotropa del caso studio, costituisce un ulteriore elemento di valutazione dei margini d'incertezza e della sensitività del risultato alla delicata fase della modellizzazione.

Ho così ottenuto tre tabelle output dell'operazione di *spatial correlation* bidirezionale contenenti le informazioni ripartite nelle seguenti colonne:

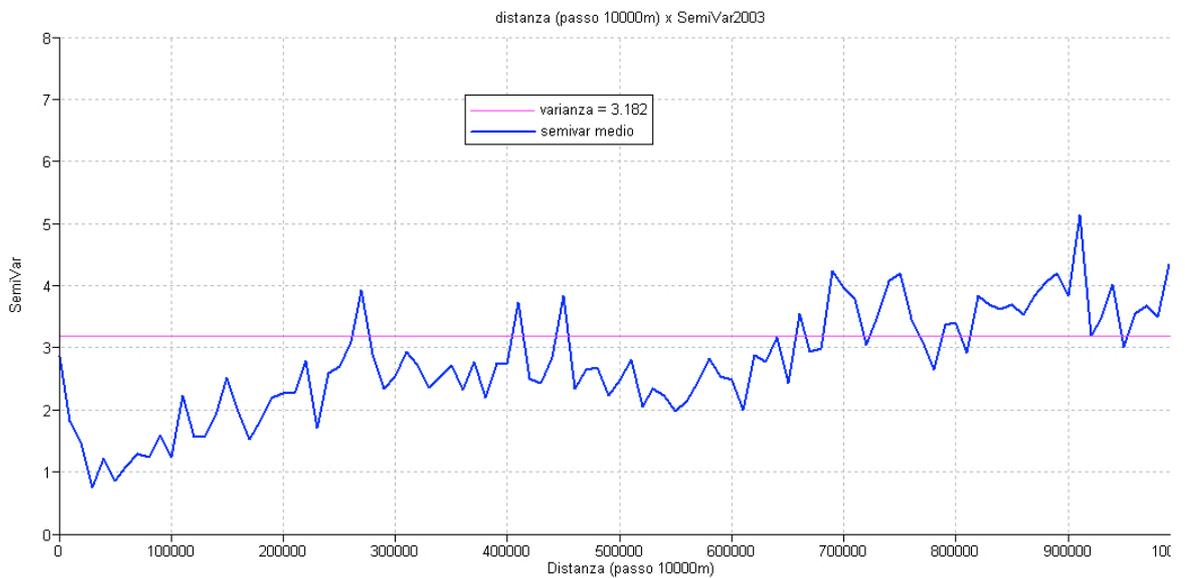
- colonna di distanza: con i valori medi degli intervalli di distanza, espressi in multipli del passo;
- colonna del numero di coppie: per ogni intervallo di distanza fornisce il numero di coppie di punti trovati a tale distanza gli uni dagli altri;
- colonna I: cataloga per ogni intervallo di distanza l'autocorrelazione spaziale delle coppie di punti contenute in quell'intervallo;
- colonna C: riporta per ogni intervallo di distanza, una stima della varianza spaziale delle coppie di punti contenute in quell'intervallo di distanza;
- colonna del passo medio 1 (per la direzione 1 o di anisotropia): riporta per ogni intervallo di distanza, la distanza media tra i punti delle coppie di punti contenuti in quell'intervallo;
- colonna del numero di coppie 1: cataloga per ogni intervallo di distanza il numero di coppie di punti trovati a tale distanza gli uni dagli altri e nella direzione stabilita dall'utente (direzione 1: di anisotropia);

- colonna del semivariogramma 1 (direzione 1: anisotropia): per ogni classe di distanza, contiene il valore del semivariogramma sperimentale delle coppie di punti trovate a tale distanza gli uni dagli altri e nella direzione 1;
- colonna del passo medio 2 (per la direzione 2 perpendicolare a quella di anisotropia): riporta per ogni intervallo di distanza, la distanza media tra i punti delle coppie di punti contenuti in quell'intervallo di distanza;
- colonna del numero di coppie 2 : cataloga per ogni intervallo di distanza il numero di coppie di punti trovati a tale distanza gli uni dagli altri e nella direzione stabilita dall'utente (direzione 2: perpendicolare a quella di anisotropia);
- colonna del semivariogramma 2 (direzione 2: perpendicolare a quella di anisotropia): per ogni classe di distanza, contiene il valore del semivariogramma sperimentale delle coppie di punti trovate a tale distanza gli uni dagli altri e nella direzione 2;

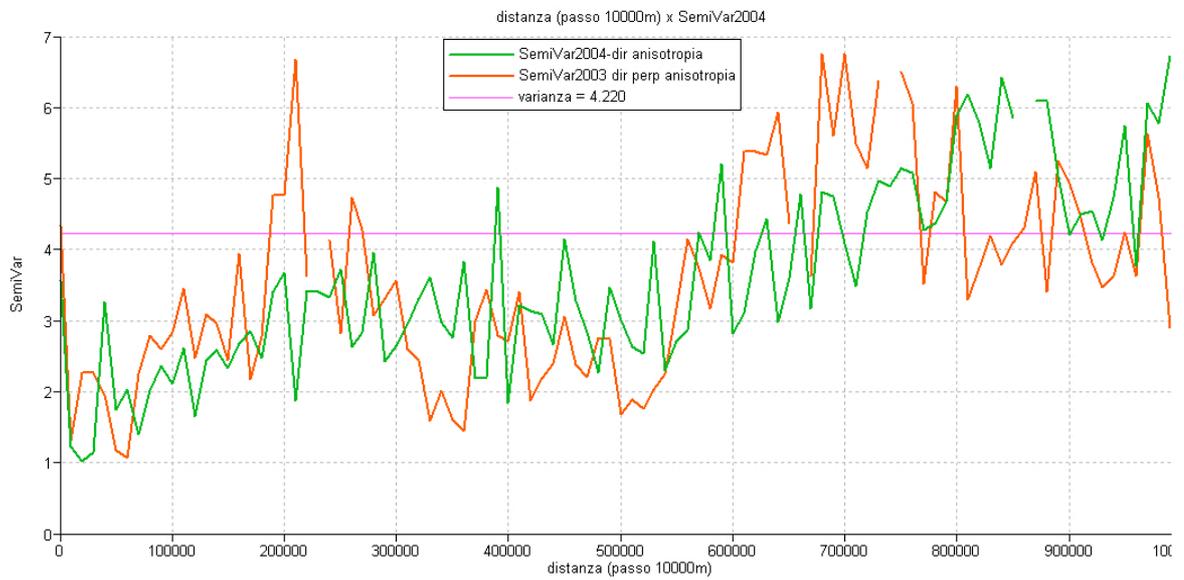
Nel caso omnidirezionale la tabella contiene solo da sei delle precedenti colonne: distanza, numero di coppie, colonna I, colonna c, colonna del passo medio, colonna del semivariogramma. Avendo a disposizione tale gamma di informazioni così ripartite, ho proceduto a visualizzare il semivariogramma creando dei grafici cartesiani a partire dalle operazioni sviluppabili da tabella, così strutturati: i valori della colonna di distanza sull'asse delle ascisse e quelli della colonna semivariogramma sull'asse delle ordinate; l'operazione è stata ripetuta per i tre valori di semivariogramma sperimentale calcolati (direzione di anisotropia, direzione perpendicolare, medio), oltre che per i tre anni in esame.



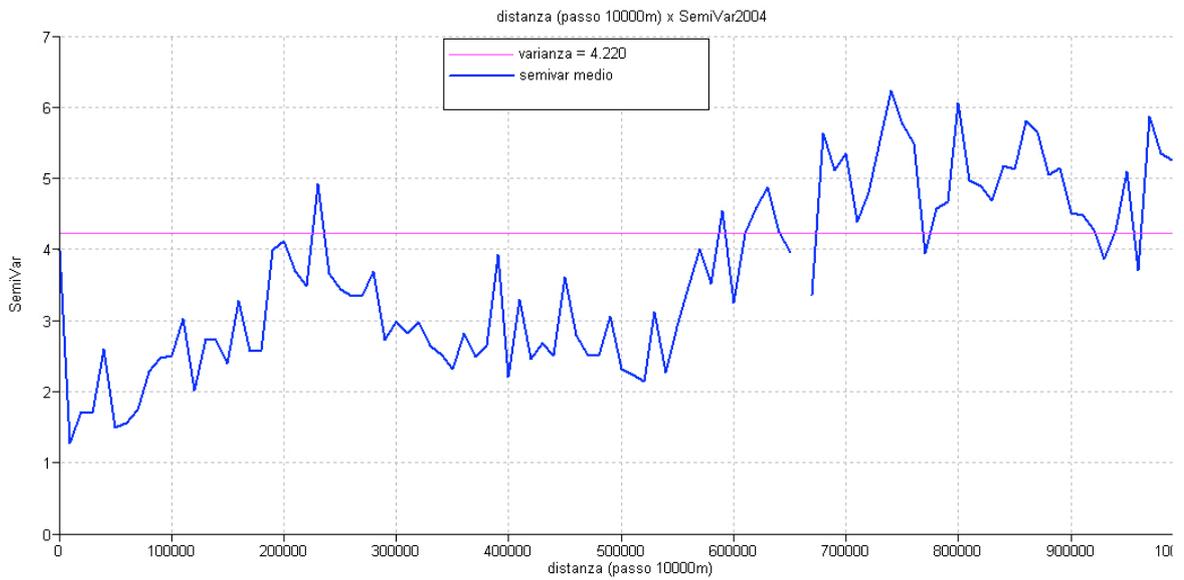
Semivariogramma anisotropo 2003



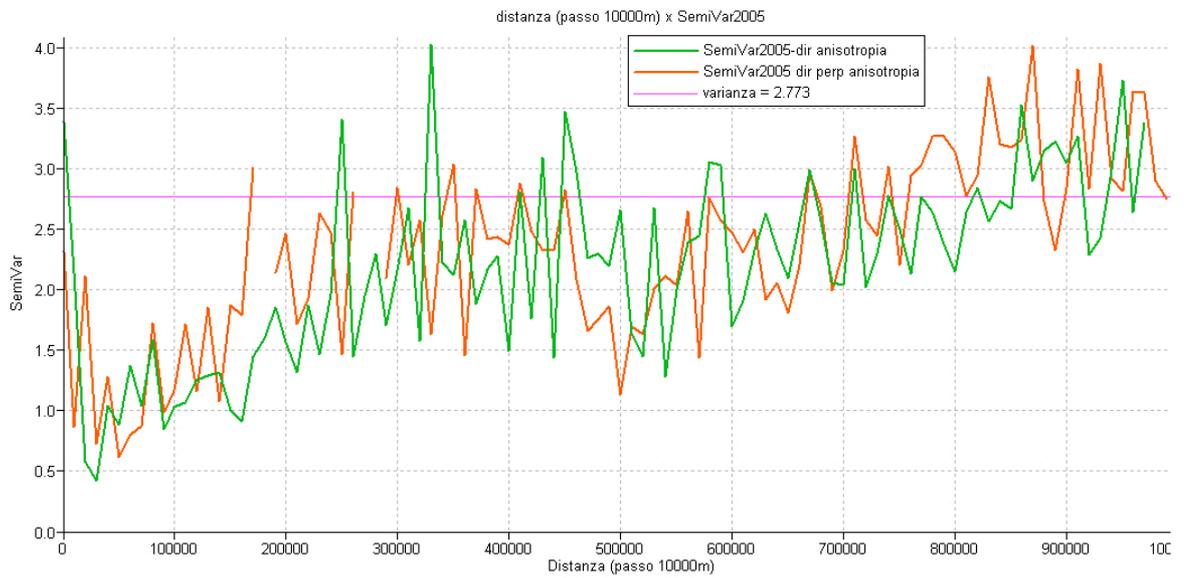
Semivariogramma medio 2003



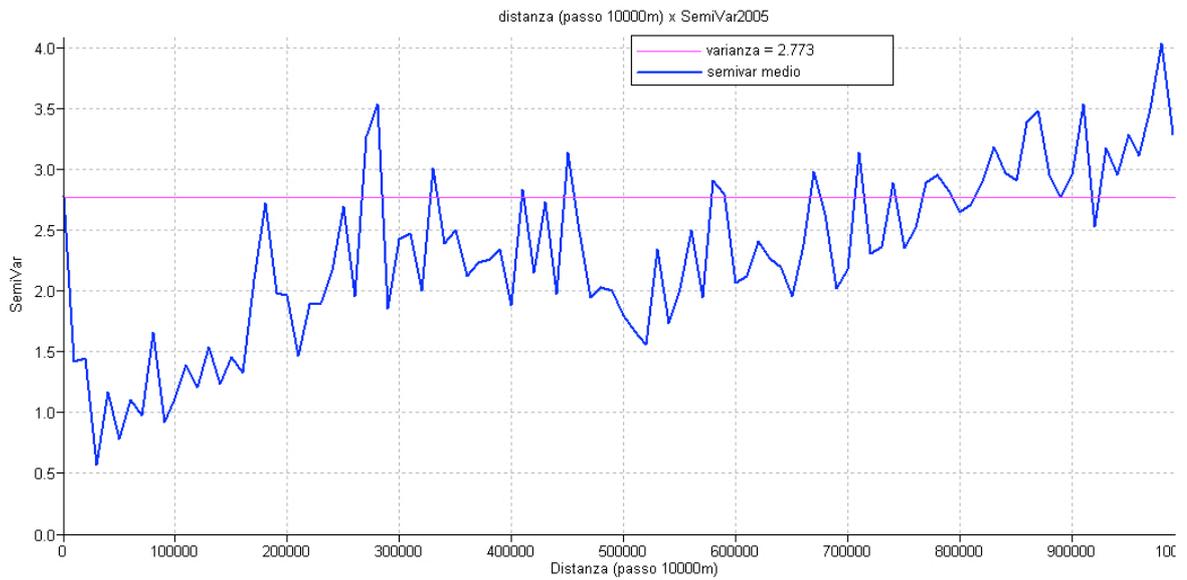
Semivariogramma anisotropo 2004



Semivariogramma medio 2004



Semivariogramma anisotropo 2005



Semivariogramma medio 2005

3.1.2. Modellizzazione dei variogrammi

Ottenuti i variogrammi sperimentali, il passo successivo nell'analisi di caratterizzazione della variabilità spaziale, propedeutico alla seguente interpolazione con metodi quali il krigaggio, è costituito dalla modellizzazione dei valori discreti del semivariogramma con una funzione continua.

In ambiente ILWIS la modellizzazione è possibile a partire direttamente dalla finestra di visualizzazione del grafico aggiungendo un modello di variogramma (*Add Graph Semi-variogram Model*) a scelta tra quelli disponibili:

- sferico;
- esponenziale;
- gaussiano;
- wave;
- quadratico razionale;
- circolare;
- modello di potenza;

Alla selezione del modello segue l'individuazione dei parametri di modello, da inserire opportunamente negli appositi campi, quali:

- *nugget* o effetto pepita: si ha quando si riscontra un salto verticale dal valore zero per passo nullo a un valore qualsiasi del semivariogramma localizzato a piccolissima distanza dall'origine, tale salto una volta individuato dev'essere opportunamente quantificato e specificato;
- *sill*: a differenza del nugget specificato se presente, l'introduzione del sill è imprescindibile per la modellizzazione del semivariogramma rappresentando la soglia massima di variabilità per il modello selezionato;

- *Range*: anch'esso parametro imprescindibile rappresentando la distanza alla quale viene raggiunta la soglia e quindi alla quale i valori del semivariogramma rimangono costanti;
- *Slope e potenza*: esclusivamente per il modello di potenza bisogna definire il coefficiente di direzione slope (parametro k del modello), il grado dell'esponente di potenza (parametro m del modello);

E' possibile sviluppare al massimo la combinazione di due modelli, di cui uno costituito dall'effetto pepita e l'impossibilità di sviluppare strutture annidate rappresenta senza dubbio un limite del software; tali strutture, che permetterebbero di definire un miglior modello e di conseguenza una migliore stima, arricchirebbero l'analisi strutturale di una valutazione dei fenomeni fisici associati ad ogni struttura caratterizzante variabilità spaziale. Fatta questa precisazione, in accordo con il modello e i parametri scelti, viene rappresentata immediatamente sul grafico una linea continua che mostra i valori del semivariogramma così modellizzato; l'intento è ovviamente quello di trovare il modello e i parametri (in termini grafici la linea continua) che meglio approssimano l'andamento del variogramma sperimentale; questo può richiedere diversi tentativi e il criterio di scelta basato fondamentalmente sul confronto grafico, può avvalersi del calcolo di grandezze ausiliarie, in ambiente ILWIS fondamentalmente R^2 .

R^2 si ottiene sottraendo all'unità una frazione il cui numeratore contiene la somma del quadrato delle differenze tra i valori del semivariogramma sperimentale e i valori del semivariogramma calcolati con il modello adottato e al denominatore la somma del quadrato delle differenze tra i valori del semivariogramma sperimentale e la media dei valori del semivariogramma sperimentale calcolata rispetto al numero degli intervalli. Il massimo valore che R^2 può assumere è quindi 1 e corrisponde al caso di un'esatta coincidenza tra i valori del semivariogramma calcolati usando un determinato modello, e i valori del semivariogramma sperimentale. Il calcolo di R^2 è possibile grazie

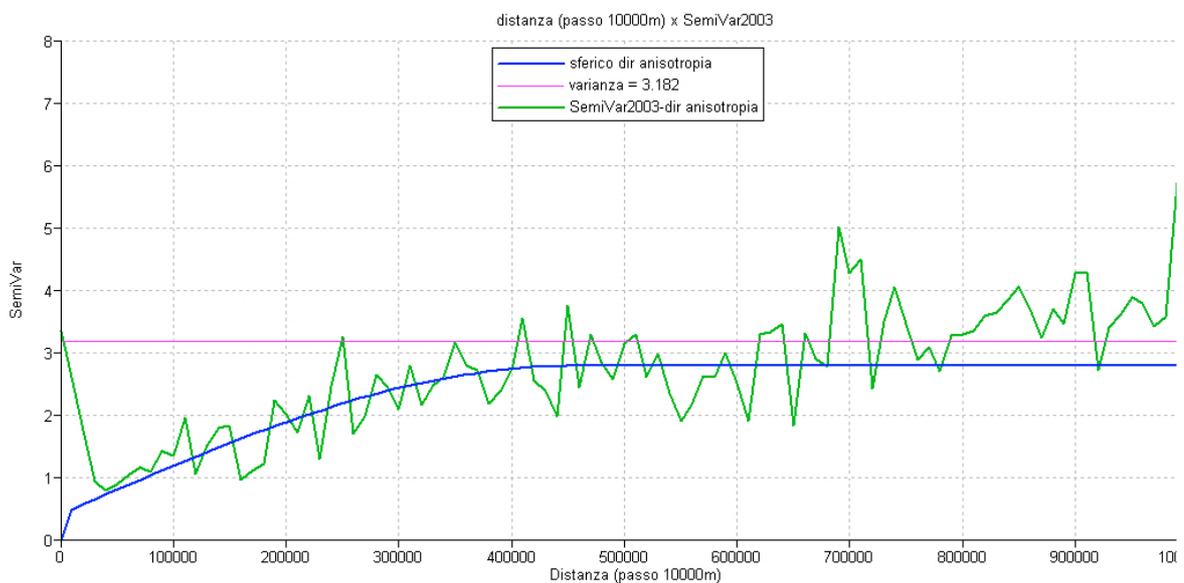
all'operazione *Column Semi-variogram* che registra all'interno di una colonna output della tabella *Spatial Correlation* i valori del semivariogramma calcolato in funzione dei valori di distanza contenuti nella colonna di distanza e del modello e parametri di semivariogramma specificati dall'utente. In tal modo si possono registrare in diverse colonne i valori del semivariogramma relativi a differenti modelli e determinare R^2 con semplici operazioni dalla barra di comando.

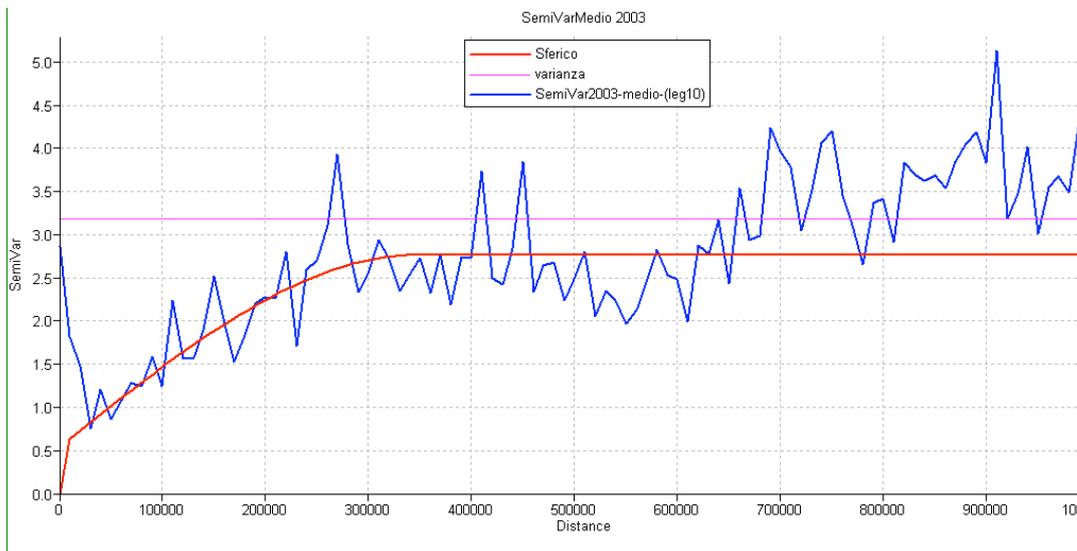
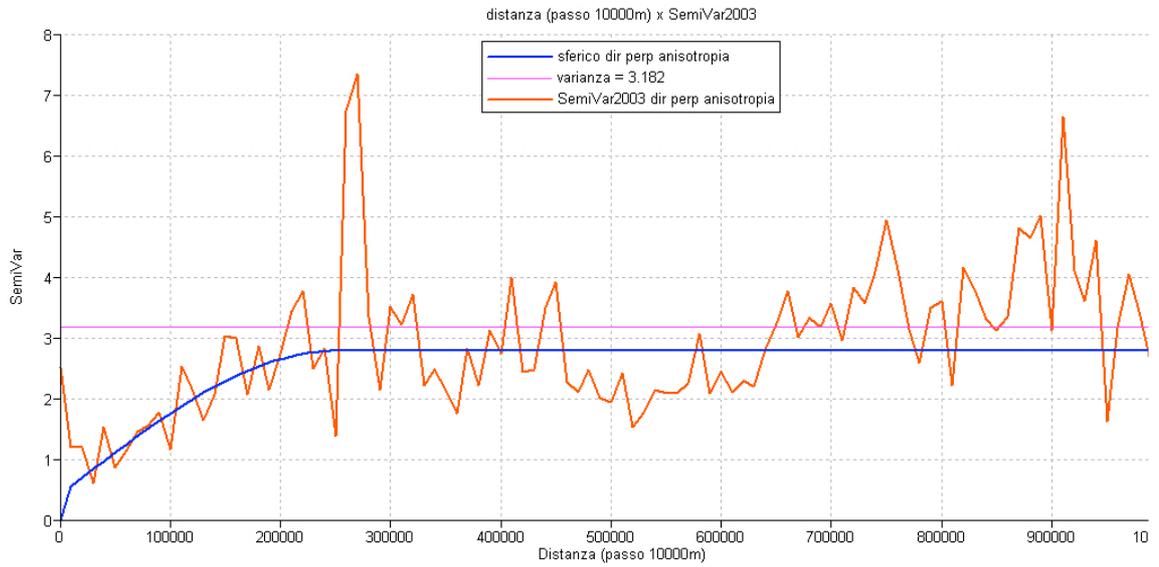
Relativamente al caso in esame, la fase di modellizzazione ha riguardato i semivariogrammi sperimentali anisotropi e medi della media annua della concentrazione di benzene nei tre anni considerati. Il modello che ho ritenuto meglio si prestasse alla modellizzazione della variabile è lo sferico con effetto pepita, i parametri adottati sono riassunti schematicamente nella seguente tabella, dove per ogni anno esaminato si riportano: modello, *nugget*, *sill* e *range* in funzione del semivariogramma considerato: medio o anisotropo.

ANNO	Semivariogramma	Modello	Nugget	Sill	Range (m)
2003	Dir. anisotropia	Sferico	0.4	2.780	450000
	Dir.perpendicolare	Sferico	0.4	2.780	250000
	Medio	Sferico	0.54	2.770	350000
2004	Dir. anisotropia	Sferico	0.9	3.6	450000
	Dir.perpendicolare	Sferico	0.9	3.6	250000
	Medio	Sferico	1.13	3.75	350000
2005	Dir. anisotropia	Sferico	0.37	2.370	450000
	Dir.perpendicolare	Sferico	0.37	2.370	300000
	Medio	Sferico	0.57	2.430	350000

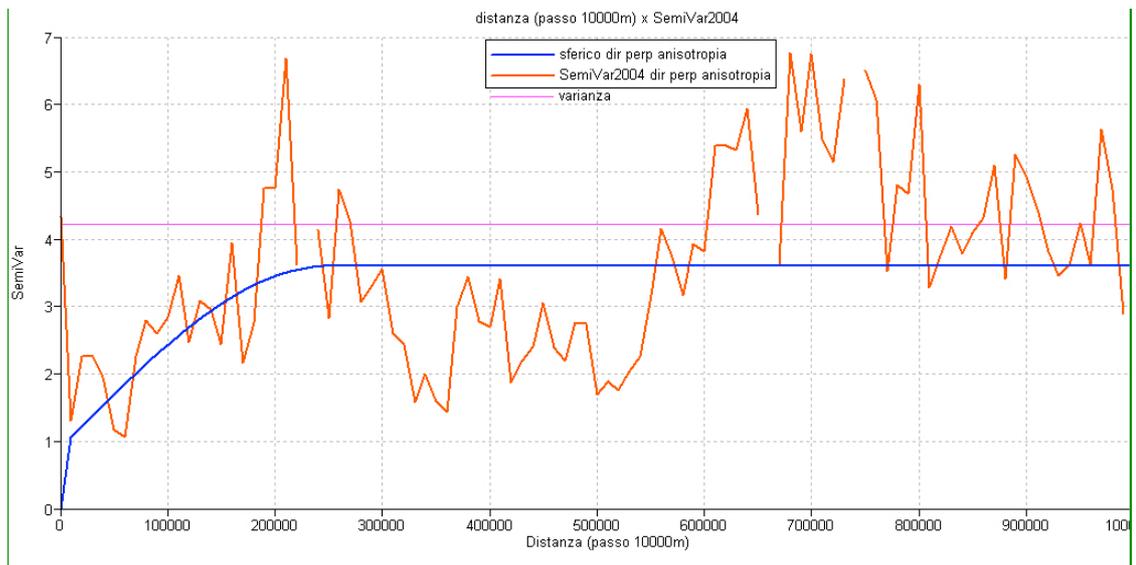
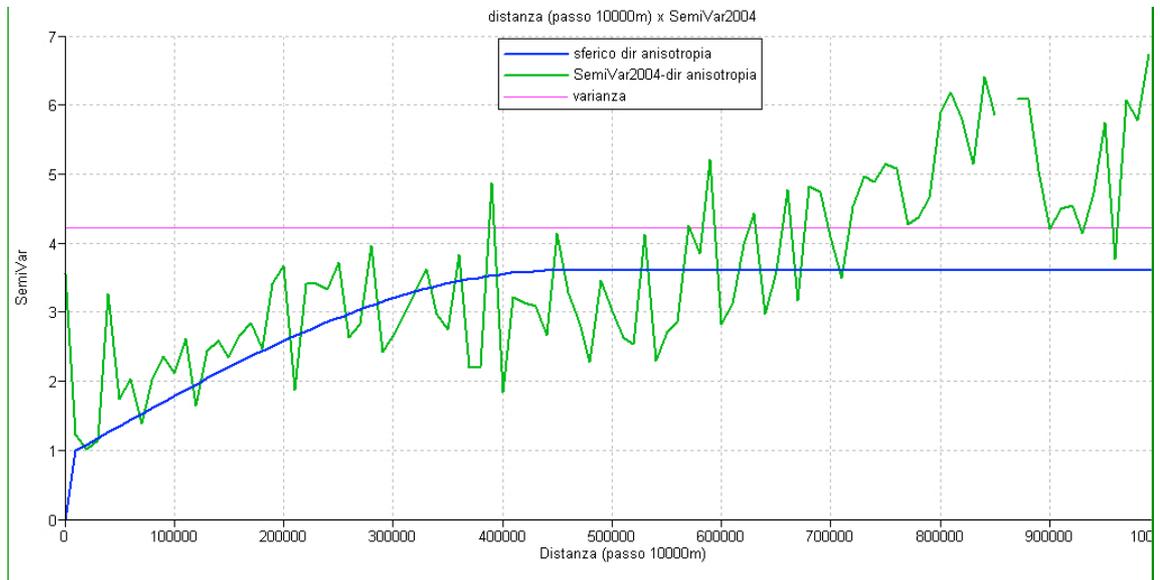
Il criterio che mi ha portato all'individuazione di tali parametri è stato principalmente quello del confronto grafico tra semivariogramma sperimentale e semivariogramma modellizzato; i grafici riportati di seguito mostrano l'effettiva adeguatezza delle soluzioni adottate. Per avere ulteriori elementi di valutazione ho provveduto a calcolare i valori di R^2 per diversi tentativi di modellizzazione preliminari a quella finale, e il valore assunto da tale grandezza non è andato oltre lo 0.3; ho inoltre esportato il file dati originario e calcolato e modellizzato i variogrammi all'interno del software *multigeo* che consente lo sviluppo di strutture annidate, senza tuttavia riscontrare sostanziali differenze dal modello sferico più pepita individuato e parametrizzato in ILWIS.

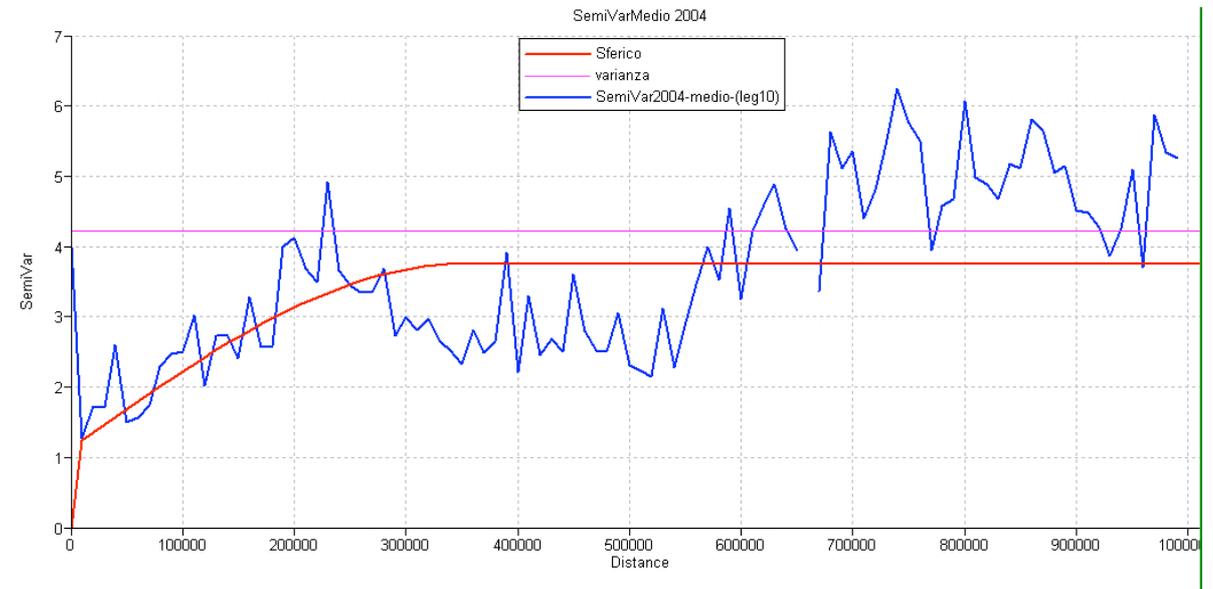
Modellizzazione 2003



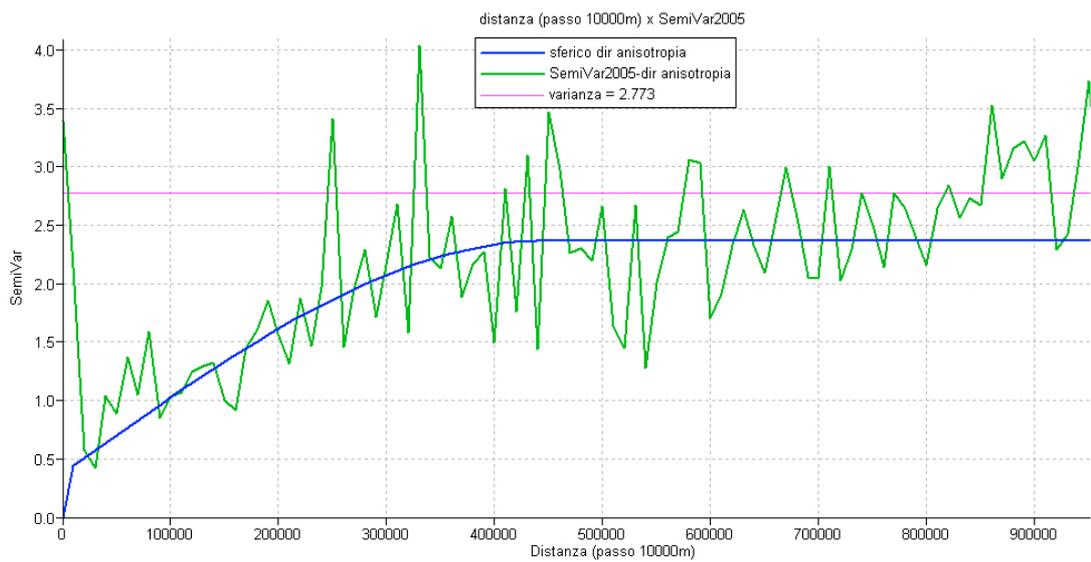


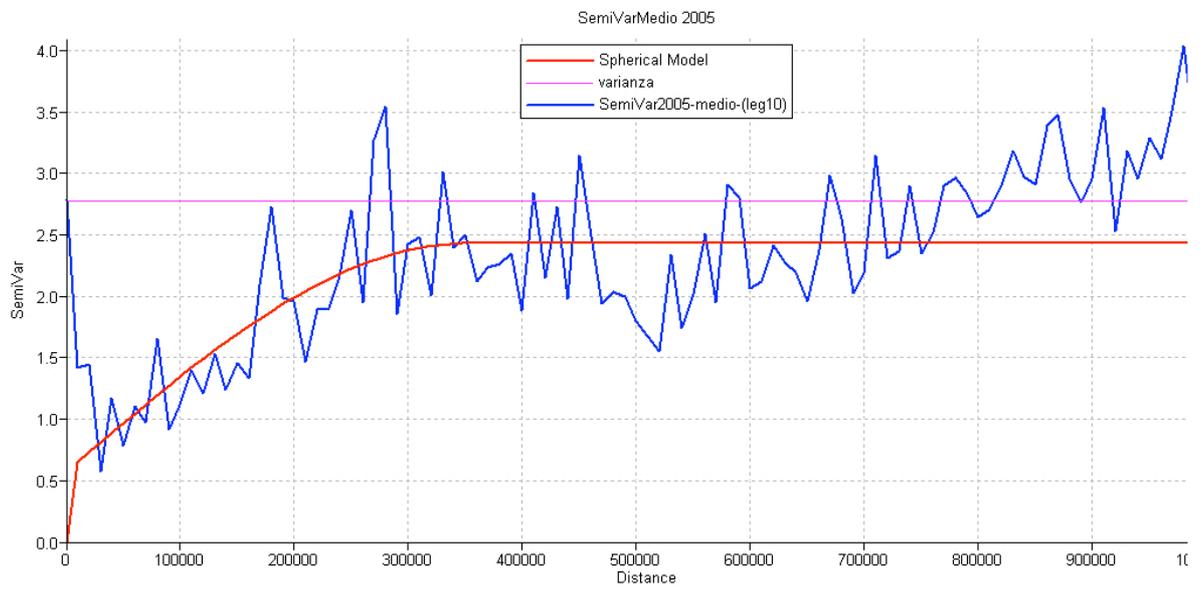
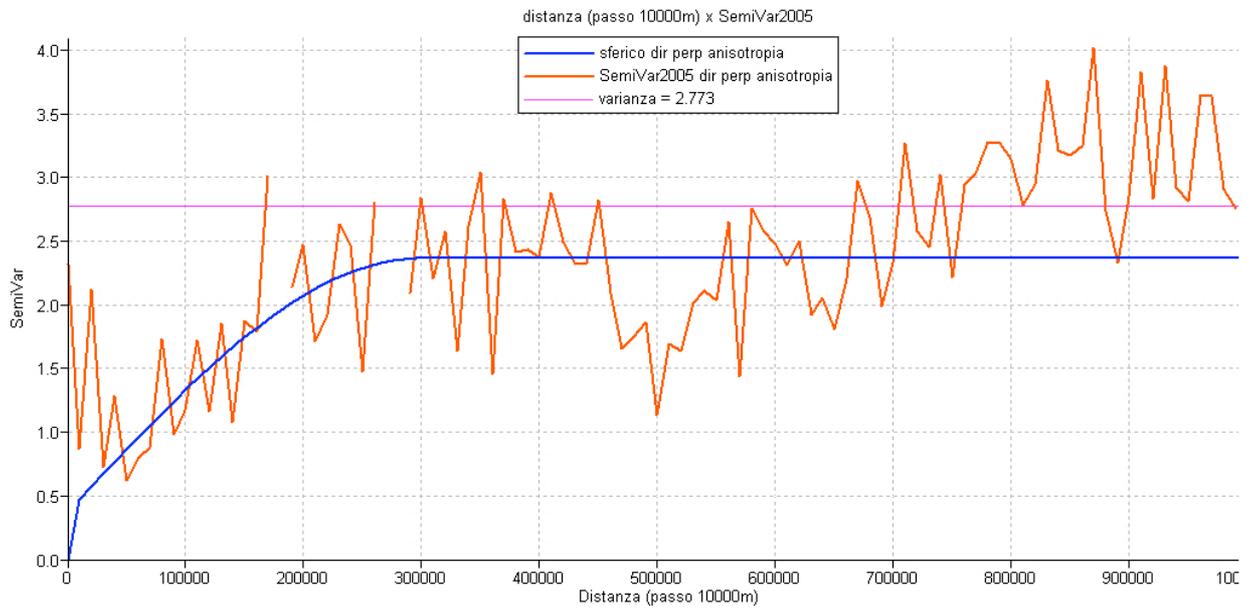
Modellizzazione 2004





Modellizzazione 2005

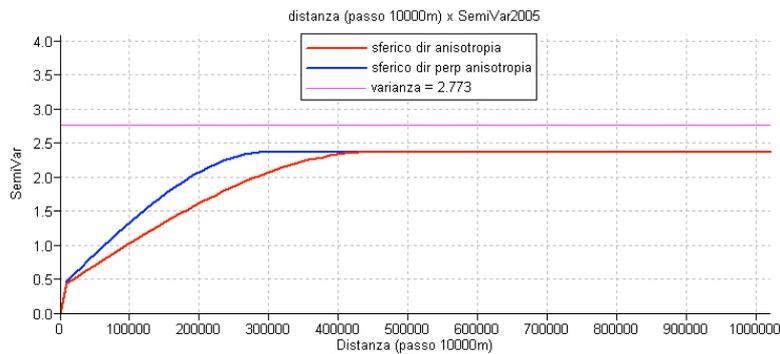
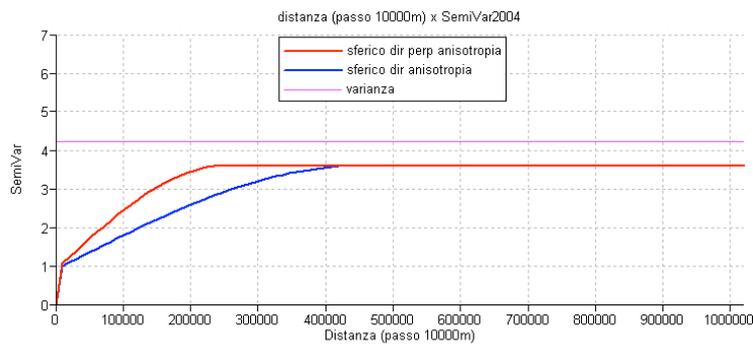
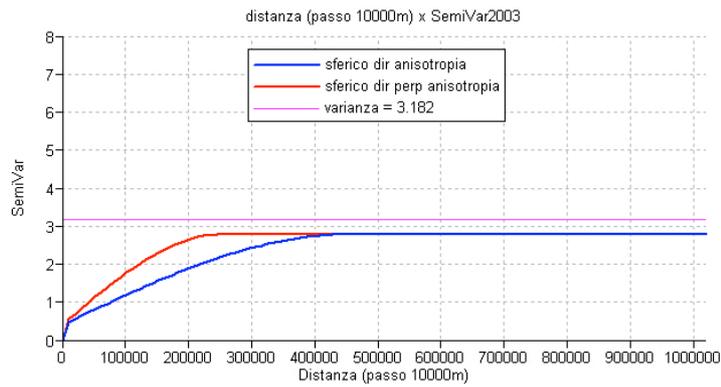




L'andamento dei variogrammi sperimentali evidenzia un forte effetto pepita, con valori del semivariogramma elevati in corrispondenza dell'origine. Tale andamento è da attribuire alla variabilità a piccola scala della variabile, per punti molto vicini ad esempio nei centri cittadini. Ho trascurato quindi questi valori alla scala dell'analisi europea, salvo poi riconsiderarli in una valutazione per aree omogenee (cinque aree coincidenti con le estensioni nazionali di Italia, Francia, Germania, Gran Bretagna, Spagna) quindi uno studio a scala locale, descritto in seguito in un sottocapitolo dedicato della presente tesi.

3.2. CARTOGRAFIA MEDIANTE KRIGAGGIO ORDINARIO

Il passo finale della prima fase di analisi, lo studio monovariato, è stato stimare la concentrazione di benzene mediante Krigaggio ordinario, implicando quindi l'ipotesi di un comportamento della grandezza in esame rappresentato da una funzione aleatoria stazionaria; come esposto nel capitolo strumenti geostatistici, le stime con il Krigaggio ordinario sono medie pesate dei dati in ingresso, il fattore di peso è determinato in funzione del modello di semivariogramma prescelto come rappresentativo della variabilità spaziale ed il calcolo è sviluppato in modo da minimizzare l'errore di stima in ogni punto. I valori stimati sono così combinazione lineare dei dati in ingresso e hanno il minimo errore di stima che può essere inoltre visualizzato in mappe d'errore, elemento quest'ultimo che consente di valutare l'attendibilità dei valori stimati contraddistinguendo il krigaggio da altri metodi di stima; ovviamente dal momento che il krigaggio utilizza il modello funzionale per calcolare il valore del semivariogramma in funzione del passo, la precisione di stima è strettamente connessa all'accuratezza dall'analisi strutturale e all'attenzione posta nell'individuazione del modello più appropriato a rappresentare la variabilità spaziale calcolata a partire dai dati sperimentali. La scelta del modello che per i tre anni del caso studio si riassume nei tre grafici seguenti

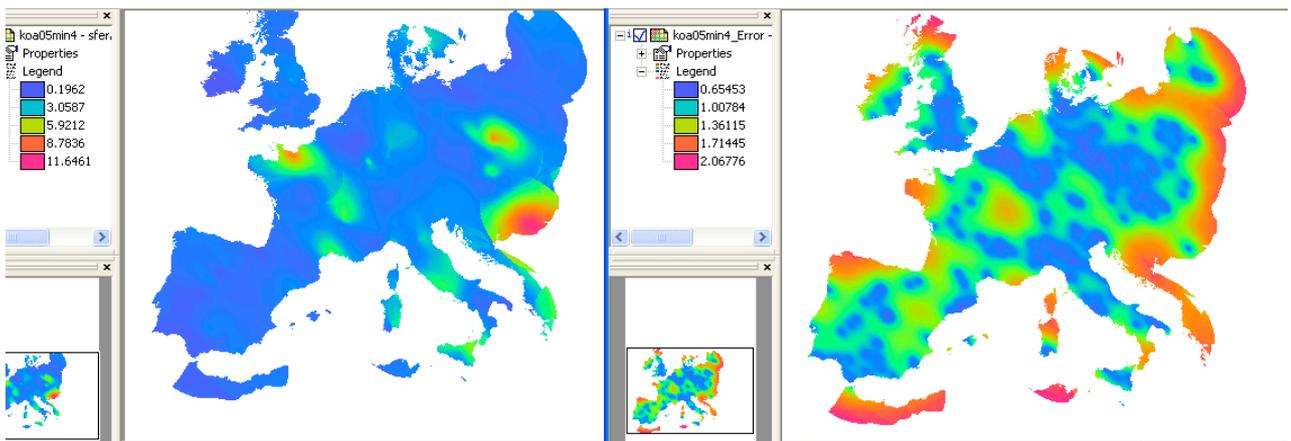


Basandomi su tali modelli, e rispettivi parametri, ho proceduto alla stima della concentrazione di benzene conclusiva dell'analisi strutturale stazionaria monovariata, mediante krigaggio ordinario anisotropo. In ambiente ILWIS il krigaggio è attuabile a partire dalle *point map*, mediante l'operazione di *kriging*: interpolazione puntuale che a partire da una *point map* fornisce una mappa raster contenente in ogni pixel la stima del valore assunto dalla variabile e opzionalmente anche una mappa raster dell'errore di stima di ogni pixel; a differenza del krigaggio semplice, che usa tutti i dati in input nella

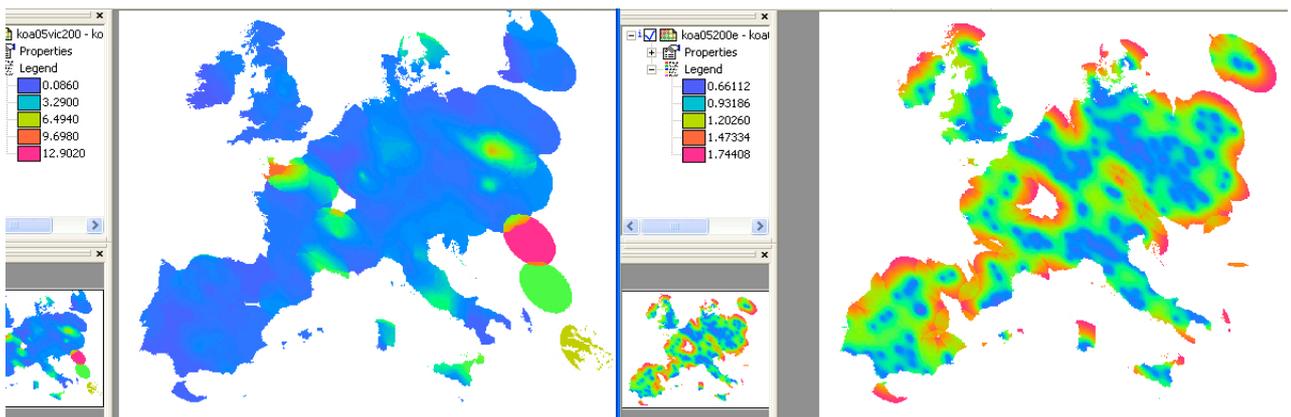
stima di ogni pixel, nel krigaggio ordinario si prendono in considerazione solo i dati all'interno di un opportuno vicinaggio di stima del pixel da stimare e contenuti entro un numero minimo e massimo di punti, specificati dall'utente insieme all'estensione del vicinaggio. Il krigaggio anisotropo, operazione *Anisotropic kriging* in ILWIS, è una variante dell'ordinario in cui si tiene conto dell'influenza della direzione di anisotropia e risulta essere il più adeguato alla stima della variabile studio alla luce dell'anisotropia riscontrata e precedentemente commentata. Ho quindi svolto l'operazione di *Anisotropic kriging* per i 3 anni in esame, specificando negli appositi campi i parametri necessari alla stima: la direzione di anisotropia, il rapporto tra i range nelle due direzioni (di anisotropia e perpendicolare), i parametri del modello del variogramma nella direzione di anisotropia, il vicinaggio di stima, il numero minimo e massimo di punti da considerare all'interno del vicinaggio riassunti nella seguente tabella.

	Direzione anisotropia	Rapporto range	Vicinaggio	Modello	Parametri (nugget, sill, range(m))
2003	135	1.8	400000 (min=1,max=100)	Sferico	0.400, 2.780, 450000
2004	135	1.8	400000 (min=1,max=100)	Sferico	0.900, 3.600, 450000
2005	135	1.5	400000 (min=1,max=100)	Sferico	0.370, 2.370, 450000

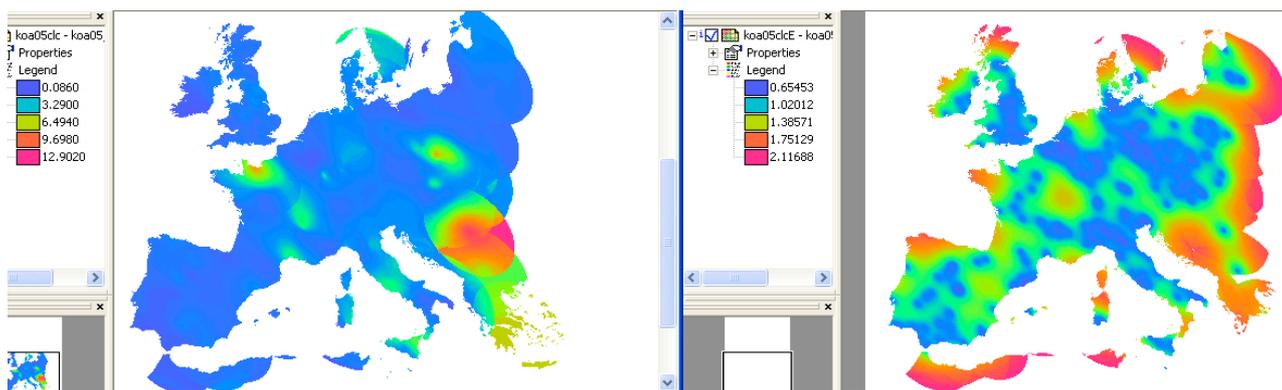
Ho scelto un vicinaggio di 400000 abbastanza ampio e un numero minimo di punti pari a 1 in modo da ottenere una mappatura di stima completa del suolo europeo; per il 2005, anno maggiormente campionato, ho effettuato diversi tentativi di stima variando i parametri del numero minimo di punti (da 1 a 4) e del vicinaggio (da 400 km a 200 km), i risultati ottenuti sono di seguito riportati.



Mappa del benzene 2005 e rispettiva varianza di stima: KO anisotropo con numero minimo di punti 4.



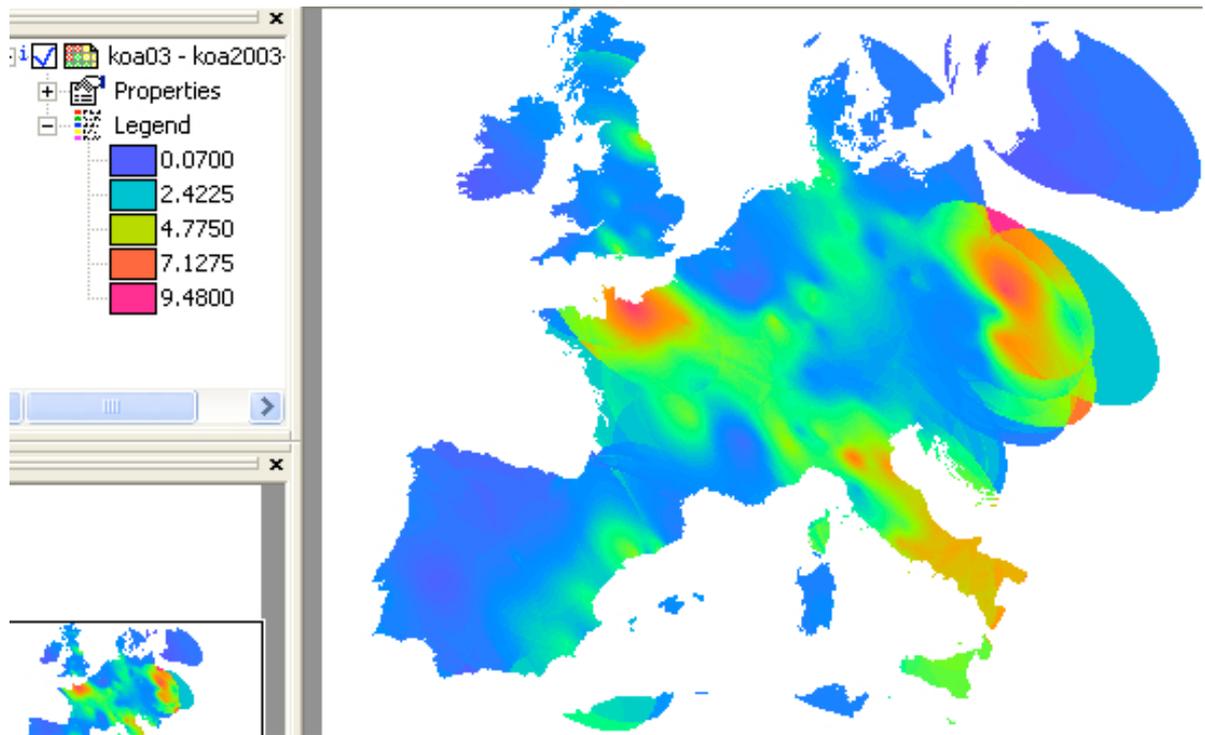
Mappa del benzene 2005 e rispettiva varianza di stima: KO anisotropo con vicinaggio 200 km.



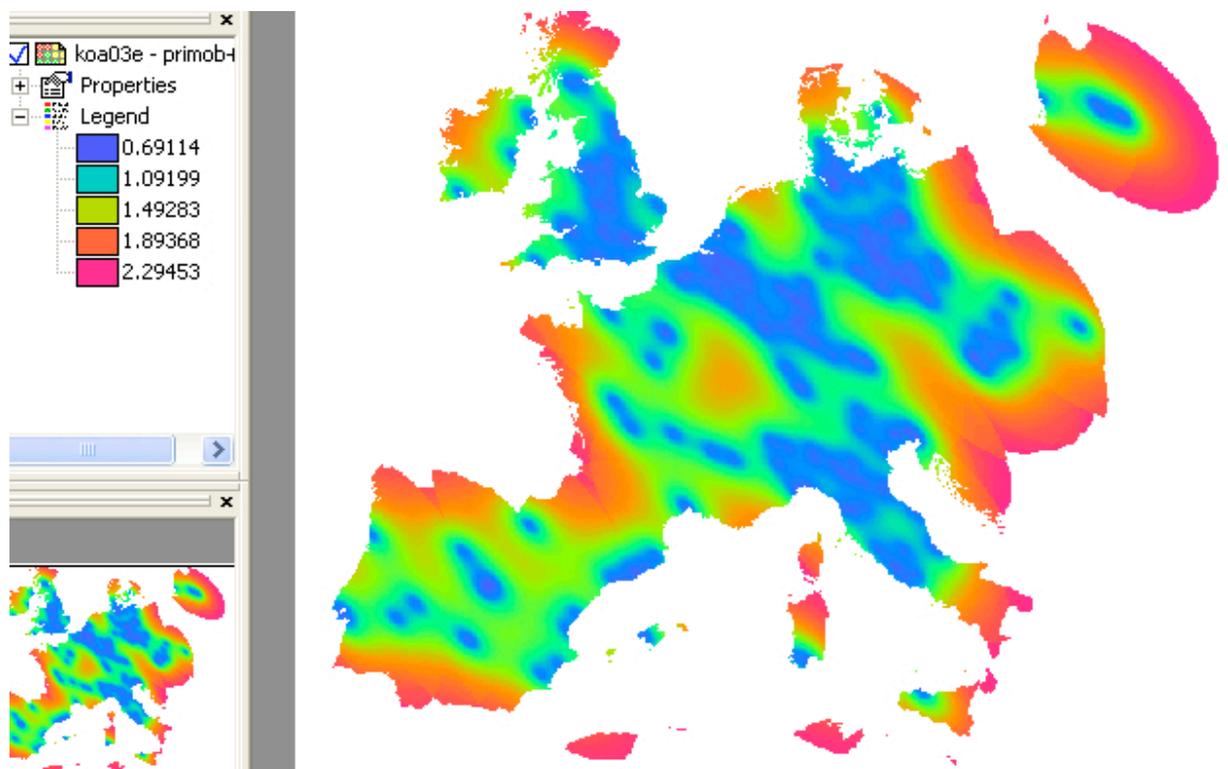
Mappa del benzene 2005 e rispettiva varianza di stima; KO anisotropo con i parametri numero minimo di punti 1 e vicinaggio 400km.

Appare evidente come la scelta dei parametri di vicinaggio e numero minimo di punti influenzi la stima e soprattutto escluda, nel caso di un numero minimo di punti maggiore e di un vicinaggio minore, la stima di parte del territorio continentale in cui invece è d'interesse determinare la concentrazione del benzene. Si nota poi come la densità di campionatura influisca sulla stima e soprattutto sulle mappe di varianze di stima, a tal proposito prima di mostrare i risultati del Krigaggio ordinario anisotropo per i tre anni si riporta una mappa dell'ubicazione delle stazioni.

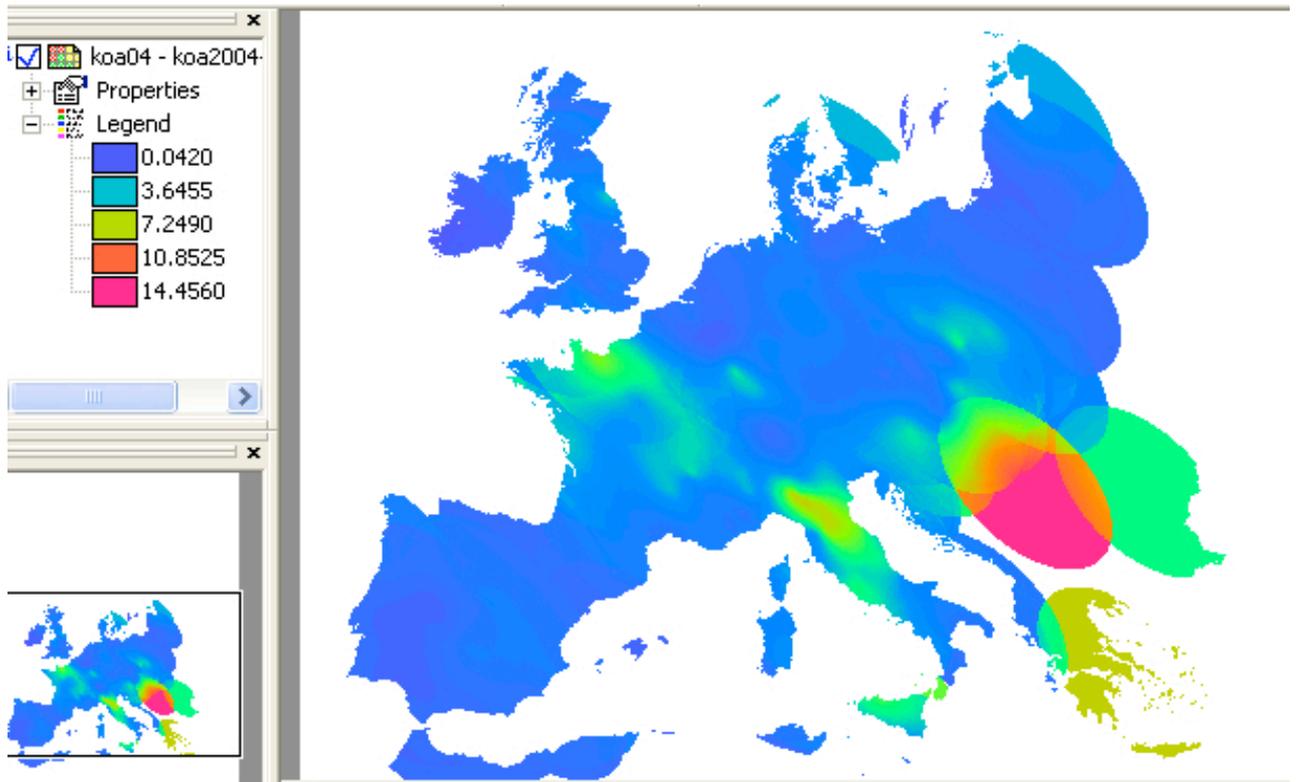
Seguono le 6 mappe di stima e varianza di stima mediante krigaggio ordinario anisotropo per il 2003, 2004, 2005 . I valori di massimo minimo per i diversi anni variano rispettivamente da 0.007 e 9.48 per il 2003, a 0.04 e 14.45 per il 2004, a 0.08 e 12.09 per il 2005. Emerge chiaramente come la maggior campionatura per l'anno 2005 porti ad una stima che copre aree più estese, e inoltre la varianza di stima varia tra un minimo e un massimo inferiori rispetto a quelli del 2003 e del 2004. Ovviamente i valori di tale grandezza raggiungono il valori più elevati in corrispondenza dei confini continentali e manifestano un risentono dell'ubicazione delle stazioni.



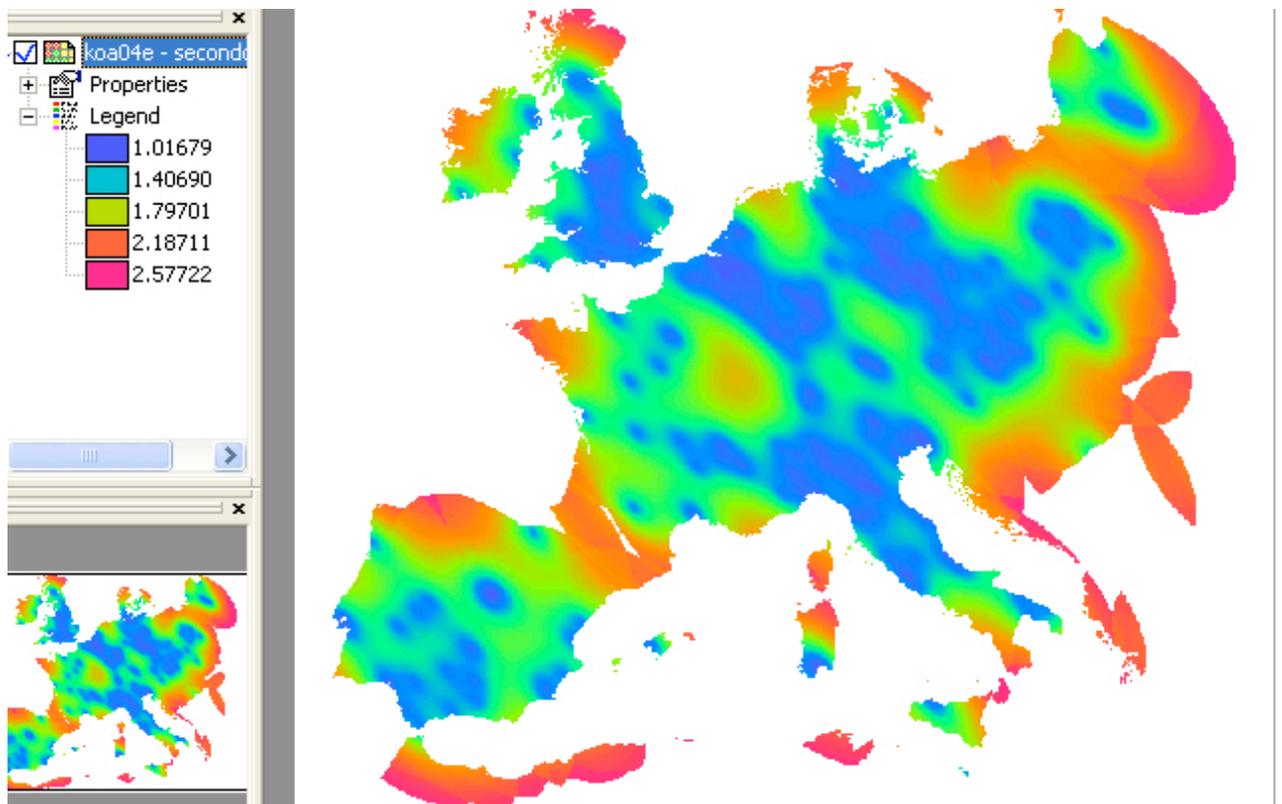
Mappa benzene 2003:KO anisotropo



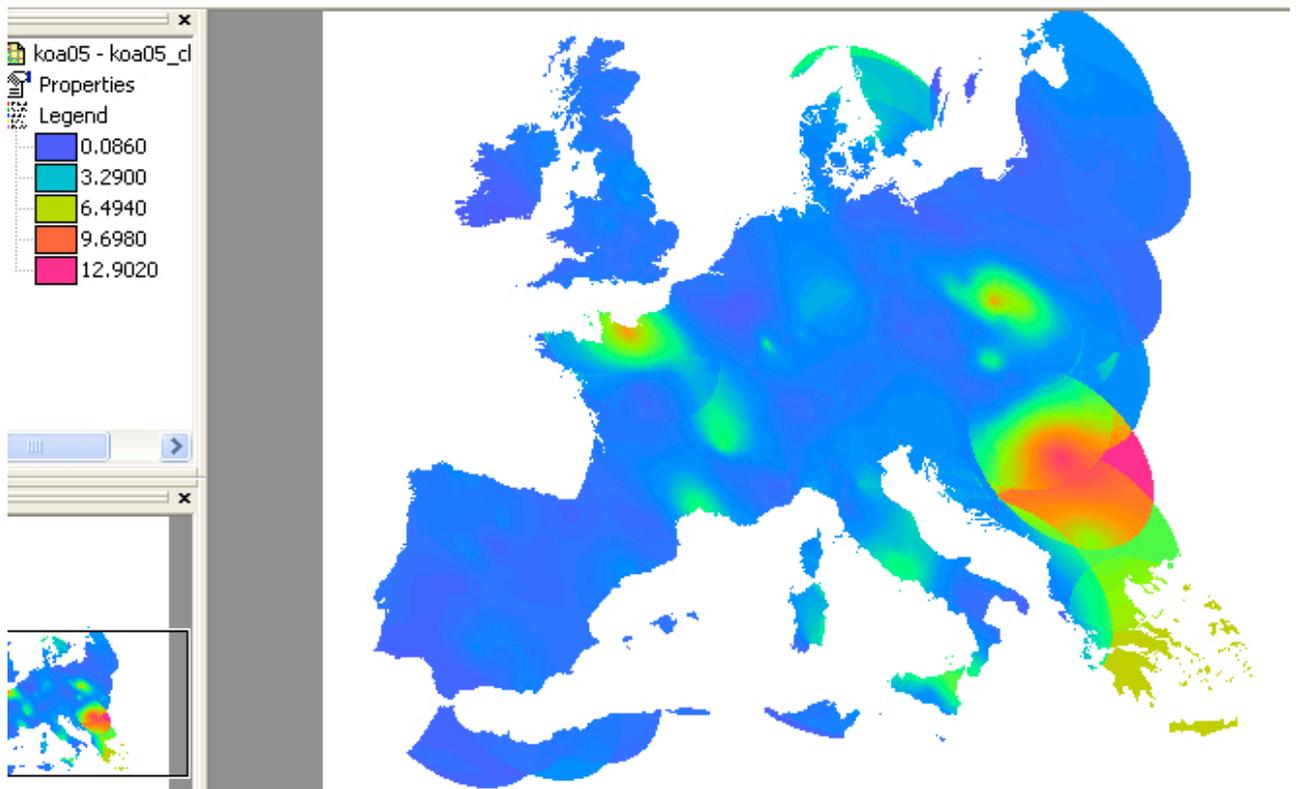
Mappa della varianza di stima del benzene 2003:KO anisotropo



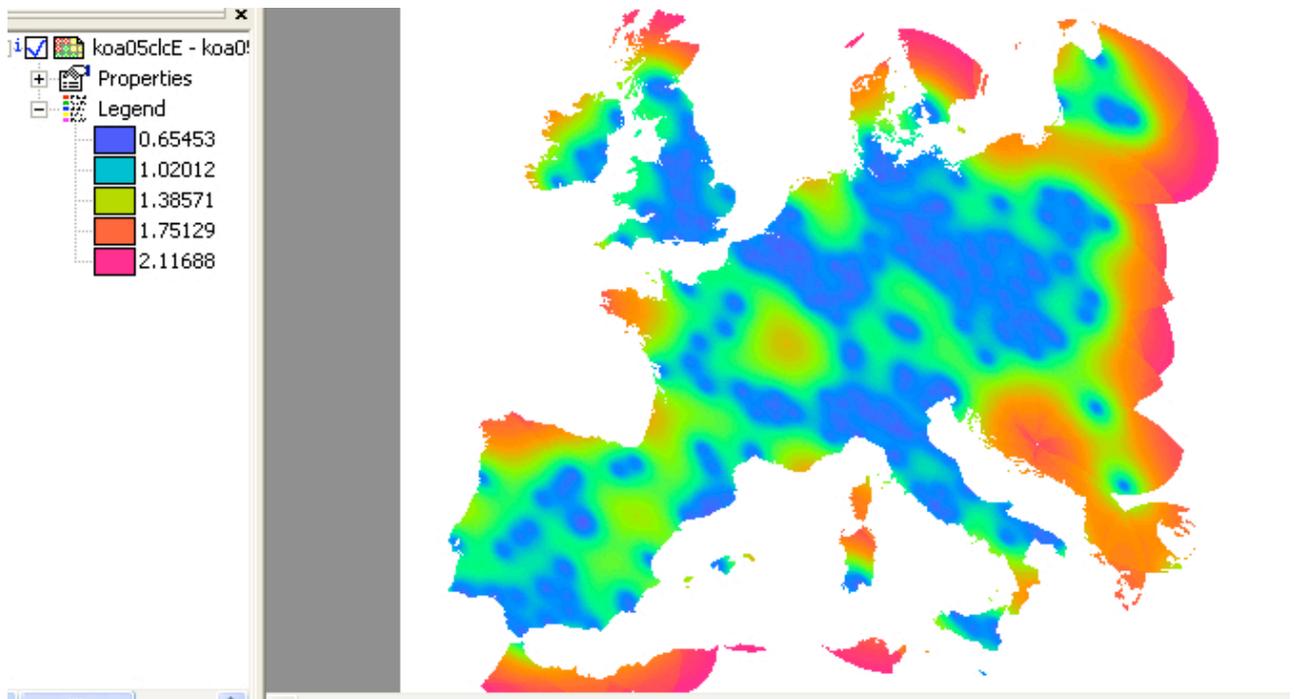
Mappa benzene 2004:KO anisotropo



Mappa della varianza di stima del benzene 2004:KO anisotropo



Mappa benzene 2005:KO anisotropo



Mappa della varianza di stima del benzene 2005:KO anisotropo

4.STUDIO SPAZIO TEMPORALE

4.1.INTRODUZIONE ALL'ANALISI SPAZIO TEMPORALE

Le metodologie della geostatistica multivariata, come esposto all'interno del capitolo strumenti geostatistici, studiano i fenomeni multivariati ossia quei fenomeni spaziali e spazio-temporali che sono descritti da più di una variabile; si ha inoltre che singole variabili possano configurarsi, ai fini di un'analisi geostatistica, come un sistema multivariabile ed è il caso del presente studio, dove la stessa variabile benzene riferita ad anni di campionamento precedenti o successivi a quello di volta in volta preso in esame, ha costituito la variabile ausiliaria in uno studio multivariato stazionario; tale parte dell'analisi ha consentito di mettere in luce la variabilità e la correlazione temporali della variabile e si è articolata in tre fasi che costituiscono il corrispettivo multivariato di quelle esaminate per lo studio monovariato: una prima fase di analisi strutturale multivariata e calcolo dei variogrammi diretti e incrociati, una seconda fase di modellizzazione di tali variogrammi, una terza fase di stima mediante cokrigaggio.

Si sottolinea che una prima valutazione di massima sulla correlazione tra i valori assunti nei diversi anni esaminati, preliminare allo studio multivariato, è stata già effettuata all'interno delle statistiche di base, precedenti le fasi mono e multi variate dello studio, con il calcolo delle nuvole di correlazione e la matrice di varianza covarianza.

4.2. L'APPROCCIO MULTIVARIATO ALL'ANALISI SPAZIO TEMPORALE

4.2.1. Correlazioni temporali

L'analisi strutturale multivariata è basata sul calcolo e sulla modellizzazione dei variogrammi diretti e incrociati. In ambiente ILWIS il valore del variogramma incrociato sperimentale è ottenibile a partire da *point map* dipendenti da tabelle dati con almeno due variabili in altrettante colonne,

mediante l'operazione di *Cross Variogram*, che effettua il calcolo in funzione del passo e delle variabili introdotte; tale operazione costituisce la forma multivariata di quella di *Spatial Correlation* analizzata precedentemente per il calcolo dei variogrammi diretti ma non consente, come quest'ultima, di tener conto della direzione della distanza tra i punti e quindi dell'anisotropia. Si precisa pertanto che nelle fasi di calcolo e modellizzazione dei variogrammi propedeutiche alla successiva stima della variabile, sono stati considerati i variogrammi medi delle quattro direzioni per le coppie di variabili ,(anni), di volta in volta esaminate, non tenendo quindi conto del comportamento anisotropo, emerso e commentato contestualmente al calcolo delle mappe variografiche; l'analisi dei risultati di tale fase dello studio va quindi effettuata alla luce di tale omissione dovuta al limite del software, che non consente di inglobare l'informazione di anisotropia nello studio multivariato. L'operazione di *Cross Variogram* fornisce come risultato una tabella con le informazioni ripartite in colonne del tutto analoghe a quelle risultanti dall'operazione di *Spatial Correlation*:

- colonna di distanza;
- passo medio;
- numero di coppie per la prima variabile (A);
- numero di coppie per la seconda variabile (B);
- semivariogramma sperimentale medio della prima variabile (A);
- semivariogramma sperimentale medio della seconda variabile (B);

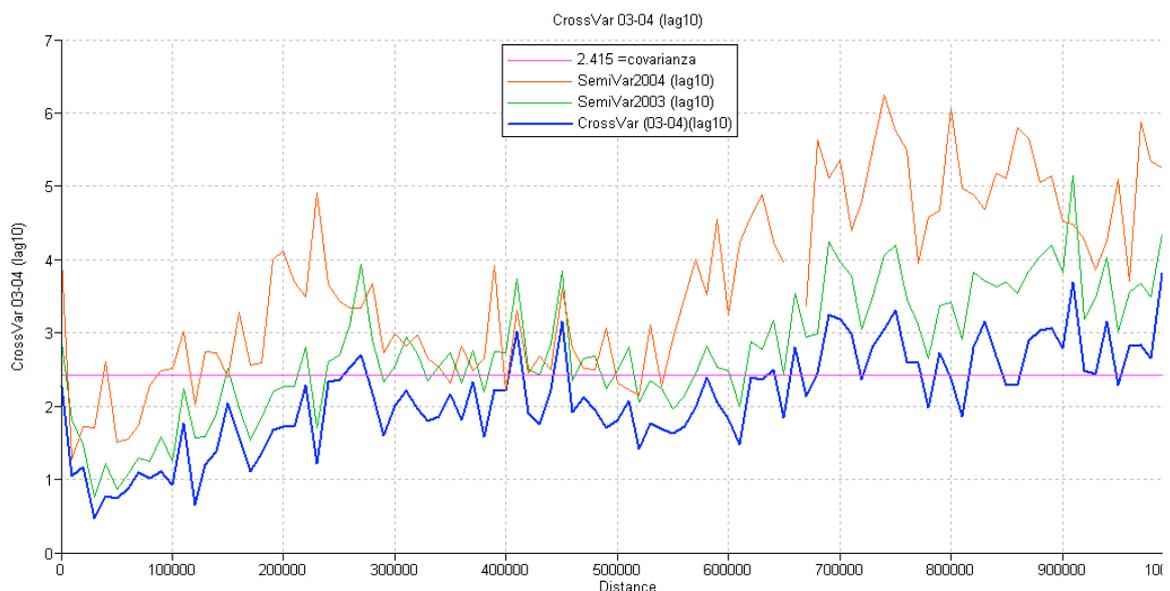
il cui contenuto è stato precedentemente descritto nello studio monovariato, e in aggiunta fornisce due colonne con informazioni specifiche dell'analisi multivariata:

- colonna del numero di coppie AB: riporta per ogni intervallo di distanza il numero di punti con valori validi della prima e della seconda variabile, trovati rispettivamente a tale distanza gli uno dagli altri;

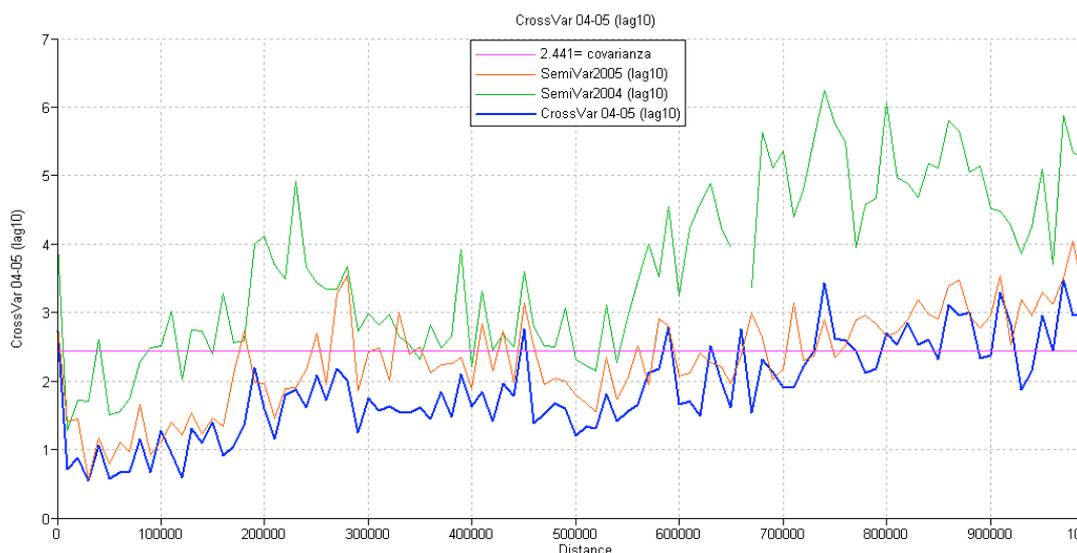
- colonna del variogramma incrociato AB: contiene per ogni intervallo di distanza i valori sperimentali del variogramma incrociato per le coppie di punti contenute nell'intervallo di distanza considerato relativamente alle variabili A e B.

Il primo passo dell'analisi strutturale multivariata è stato quindi quello di calcolare i variogrammi incrociati per le due coppie di anni consecutivi: 2003-2004, 2004-2005, a passo 10000 m, mediante l'operazione di *Cross Correlation* a partire dalle *point map*, ottenendo così tre tabelle con il formato e i valori precedentemente descritti; avendo a disposizione tale gamma di informazioni ho proceduto alla visualizzazione dei variogrammi diretti e incrociati in grafici cartesiani sviluppabili a partire da tabella, strutturati in modo da avere i valori della colonna di distanza sull'asse delle ascisse e quelli del variogramma sia diretto che incrociato sull'asse delle ordinate.

Si ricorda che il variogramma incrociato per definizione è calcolato sui dati isotopici ed è quindi rappresentativo per numeri adeguati di stazioni isotopiche, almeno uguali o superiori a 100, condizione che risulta soddisfatta nel caso studio per entrambe le coppie considerate: 2003-2004, 2004-2005.



Variogrammi incrociati e diretti 2003-2004



Variogrammi incrociati e diretti 2004-2005

4.2.2. Modellizzazione dei variogrammi sperimentali diretti e incrociati:

Ottenuti i variogrammi sperimentali diretti e incrociati (dall'operazione di *Cross Correlation*), il passo successivo nell'analisi strutturale multivariata, propedeutico alla seguente interpolazione con metodi quali il cokrigaggio, è costituito dalla modellizzazione dei valori discreti dei variogrammi diretti e incrociati con una funzione continua.

In ambiente ILWIS la modellizzazione del variogramma incrociato è possibile sempre a partire dalla finestra di visualizzazione del grafico con l'operazione *Add Graph Semi-variogram Model*, con le stesse modalità e possibilità descritte e commentate precedentemente per i variogrammi diretti. La modellizzazione del variogramma incrociato di due variabili è del tutto indipendente dalla modellizzazione dei variogrammi diretti delle stesse variabili e non essendo consentita la definizione di strutture annidate, non è di conseguenza possibile sviluppare un modello di corregeionalizzazione lineare così come definito nel capitolo strumenti geostatistici. Affinché i modelli

scelti garantiscano una corretta stima mediante cokrigaggio è opportuno quindi, se non sviluppare esternamente il modello di corregeionalizzazione, quantomeno verificare che l'equazione di Cauchy-Schwartz, espressa dalla seguente formula, sia soddisfatta:

$$|\gamma_{AB}| \leq (\gamma_A * \gamma_B)^{1/2}$$

dove:

- $|\gamma_{AB}|$ è il valore assoluto valore del variogramma incrociato;
- γ_A rappresenta il valore del semivariogramma della prima variabile che concorre al calcolo del variogramma incrociato;
- γ_B rappresenta il valore del semivariogramma della seconda variabile che concorre al calcolo del variogramma incrociato;

Dal momento che i valori del variogramma incrociato tra due variabili possono aumentare o diminuire con il passo in funzione della correlazione tra le variabili, mentre i valori del variogramma diretto delle due variabili sono positivi per definizione, quando si mette a punto un modello per il variogramma sperimentale incrociato, bisogna verificare tale disuguaglianza per garantirne l'attendibilità.

All'interno della mia analisi il criterio di modellizzazione del variogramma incrociato sperimentale è stato inizialmente grafico, salvo poi verificare, con operazioni matematiche attuabili direttamente dalla tabella output dell'operazione *Cross Variogram*, che la disuguaglianza di Cauchy-Schwartz fosse soddisfatta. Ho inoltre provveduto a sviluppare il modello di corregeionalizzazione esportando i dati e importandoli all'interno del software *multigeo* che consente la modellizzazione con strutture annidate e tale calcolo; l'unico inconveniente è stato che, operando tale software una media dei punti vicini, ho ottenuto per le stesse variabili variogrammi identici per andamento ma con soglie di variabilità inferiori a quelle calcolate in ILWIS; questo ha impedito di immettere in ILWIS tout-cour i parametri del modello di corregeionalizzazione ottenuti in *multigeo*, ma quantificando le differenze

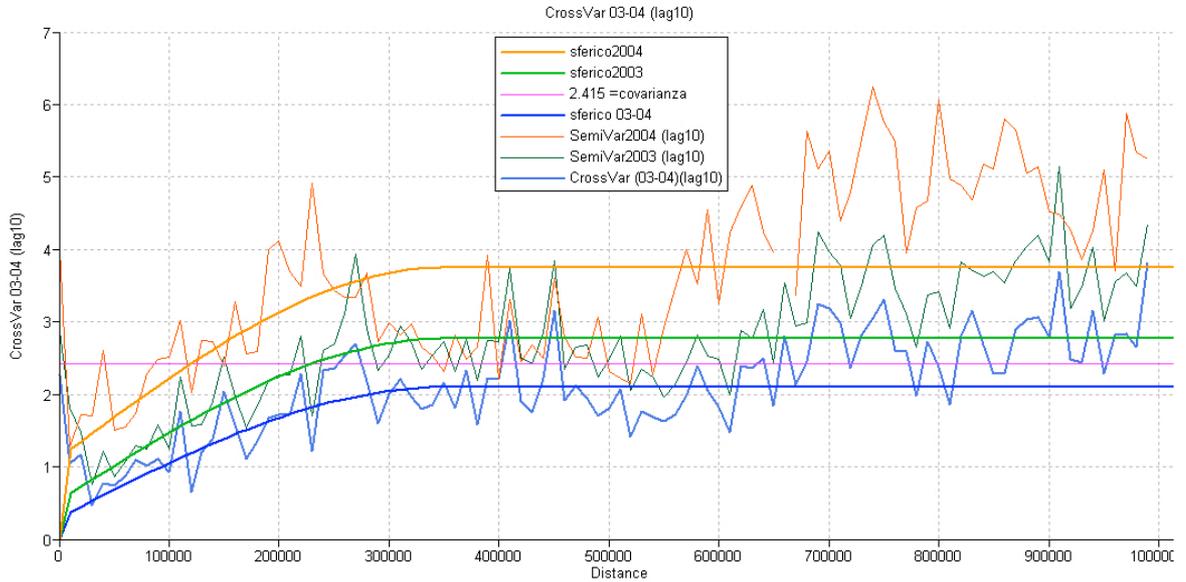
tra i valori delle varianze e delle covarianze, ha permesso comunque di rivedere e correggere le parametrizzazioni sviluppate in ILWIS sulla sola base dell'approssimazione grafica del variogramma incrociato sperimentale.

La fase di modellizzazione ha riguardato i variogrammi sperimentali diretti e incrociati della media annua della concentrazione di benzene per le coppie di anni: 2003-2004, 2004-2005. Il modello che ho ritenuto meglio si prestasse alla modellizzazione della variabile, sulla base delle considerazioni precedenti, (quindi verifica della disegualianza di Cauchy-Shwarz e modello di corregeionalizzazione ottenuto da *multigeo* opportunamente interpretato), è lo sferico con effetto pepita, i parametri adottati sono riassunti schematicamente nella seguente tabella.

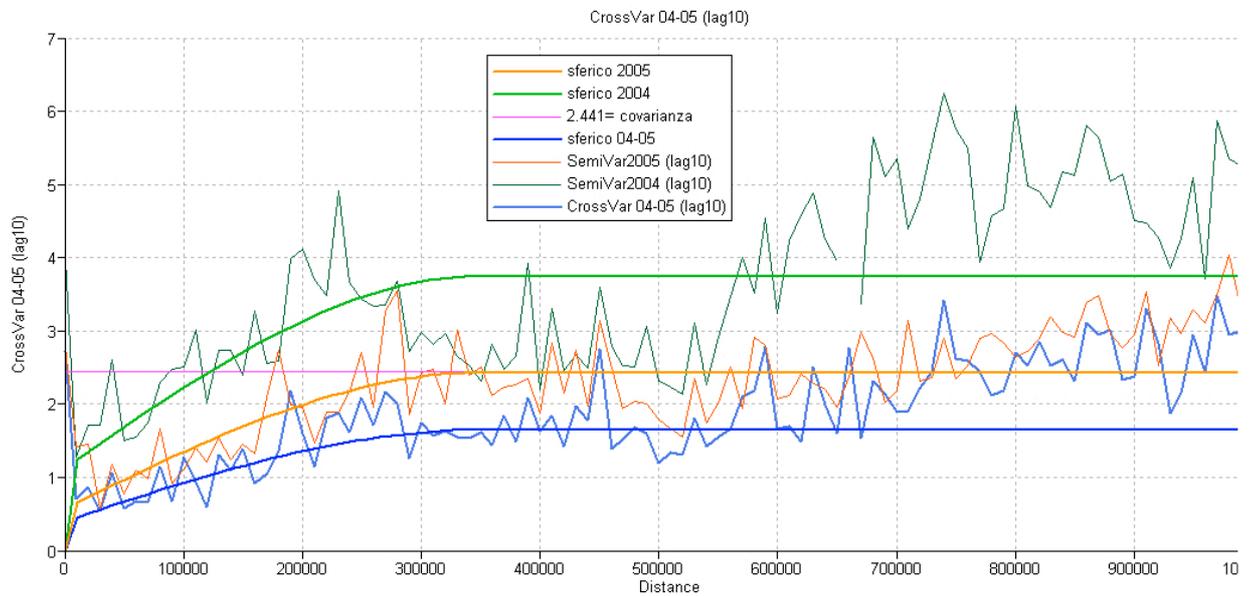
Coppia anni	Semivariogramma	Modello	Nugget	Sill	Range (m)
2003-2004	2003	Sferico	0.54	2.770	350000
	2004	Sferico	1.13	3.75	350000
	2003-2004	Sferico	0.30	2.10	350000
2004-2005	2004	Sferico	1.13	3.75	350000
	2005	Sferico	0.57	2.43	350000
	2004-2005	Sferico	0.40	1.65	350000

Le immagini seguenti mostrano l'adeguatezza di modelli e parametri adottati.

Modellizzazione 2003-2004

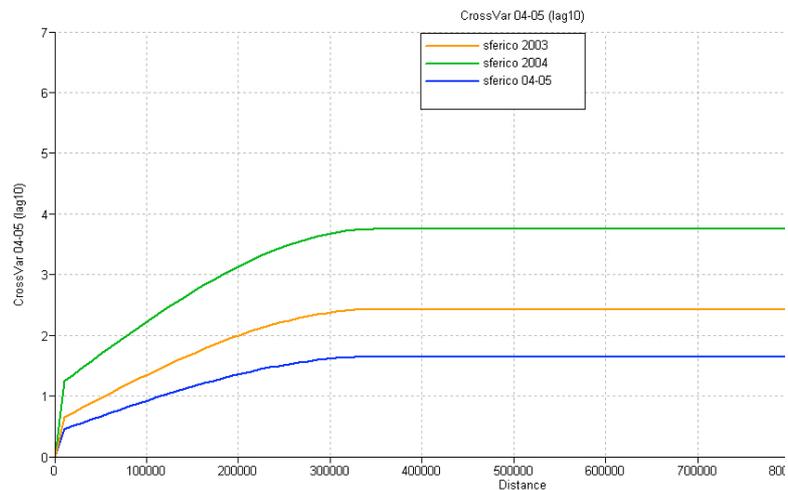
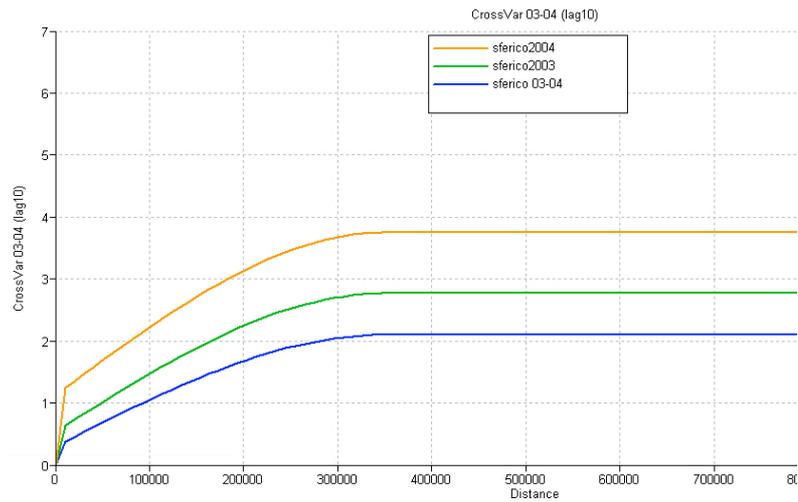


Modellizzazione 2004-2005



4.3. CARTOGRAFIA MEDIANTE COKRIGAGGIO

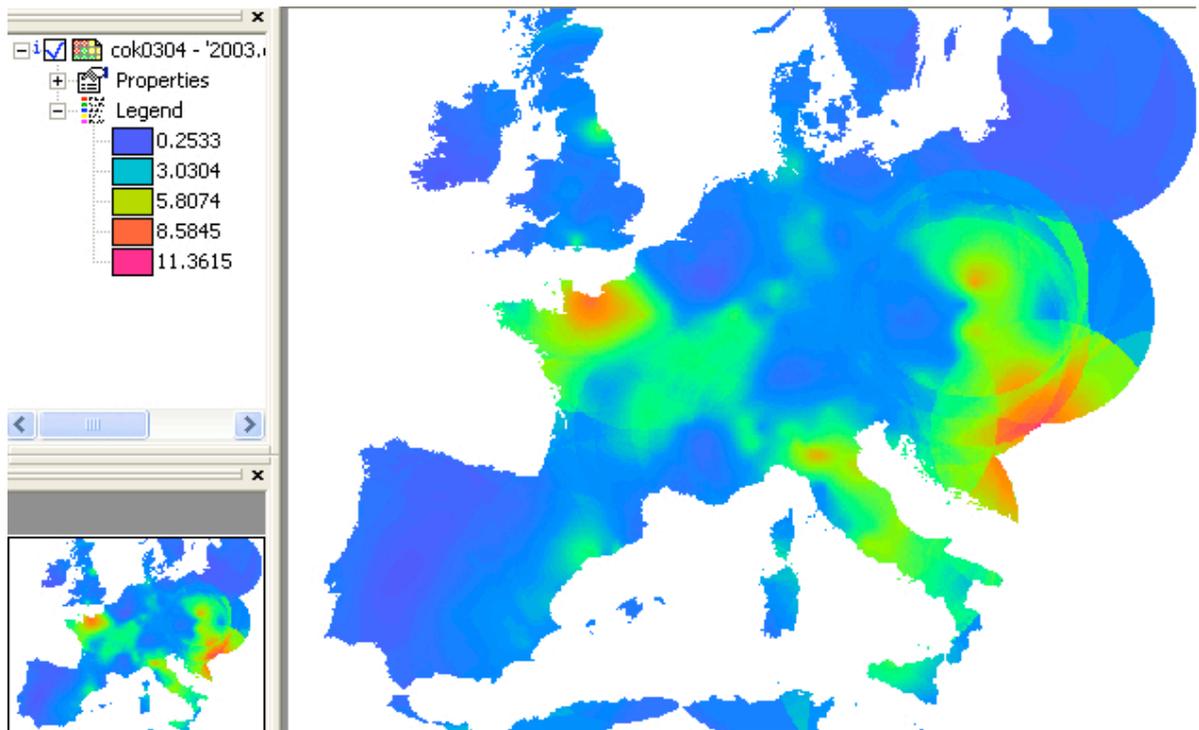
Il passo finale dello studio della correlazione temporale in multivariato, con l'ipotesi di comportamento della variabile rappresentato da una funzione aleatoria stazionaria, è stato stimare la variabile in esame mediante cokrigaggio, sfruttando così oltre ai dati della variabile da stimare anche quelli di una covariabile altamente correlata, positivamente o negativamente; come esposto nel capitolo strumenti geostatistici, il cokrigaggio può quindi essere visto come la variante multivariata dell'operazione di Krigaggio Ordinario in caso stazionario: le stime sono medie pesate dei dati in ingresso della variabile da stimare e di quella ausiliaria, i fattori di peso per i valori delle due variabili sono determinati in funzione dei modelli dei variogrammi diretti e incrociati, prescelti come rappresentativi della variabilità spaziale e della mutua correlazione tra le variabili; il calcolo è sviluppato in modo da minimizzare l'errore di stima in ogni punto, funzione solo dei modelli di variogramma diretto e incrociato adottati e visualizzabile in mappe d'errore. Come per il krigaggio ordinario anche nel cokrigaggio quindi la precisione di stima è strettamente connessa all'accuratezza dall'analisi strutturale e all'attenzione posta nell'individuazione dei modelli più appropriati a rappresentare la variabilità spaziale delle singole variabili e la loro correlazione. I grafici riportati di seguito riassumono i modelli adottati per le due coppie di variabili considerate nello studio multivariato, stazionario nell'ordine: 2003-2004, 2004-2005; basandomi su tali modelli e rispettivi parametri, ho proceduto alla stima della concentrazione di benzene conclusiva dell'analisi strutturale multivariata.



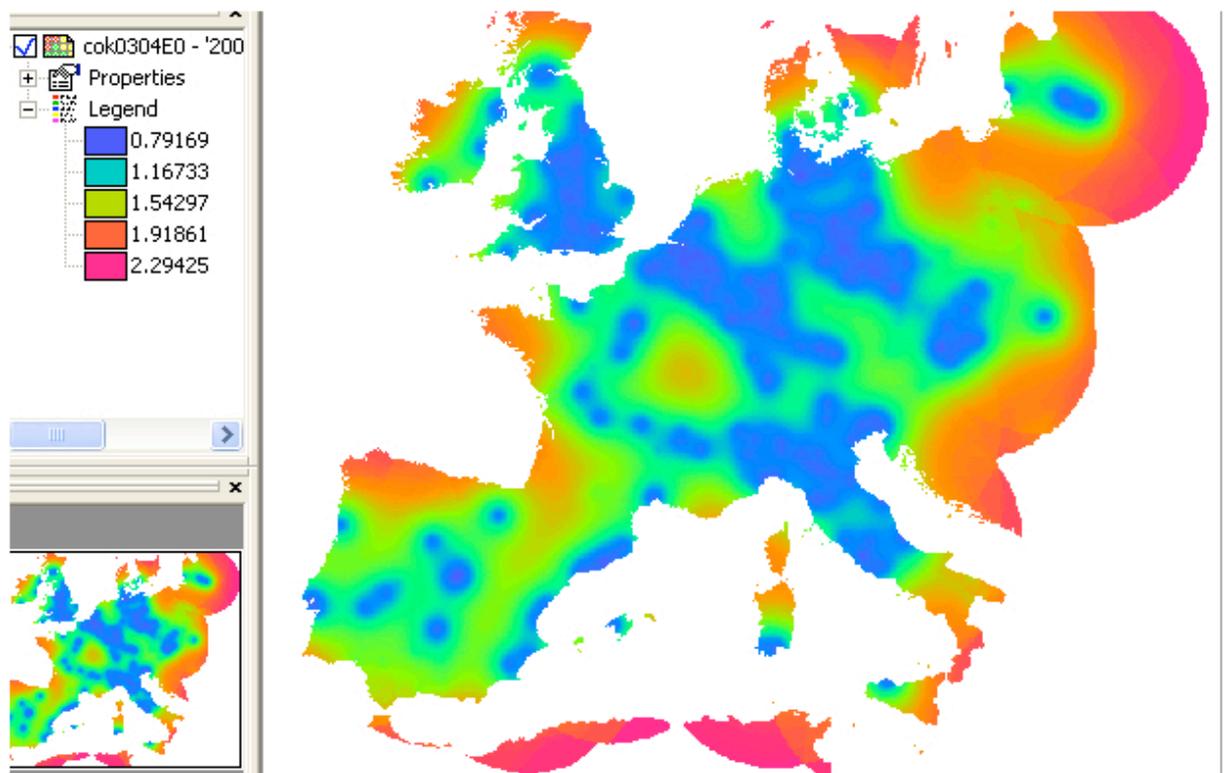
In ambiente ILWIS il cokrigaggio è attuabile a partire dalle *point map*, mediante l'operazione di *cokriging*: interpolazione puntuale che a partire da una *point map* fornisce una mappa raster contenente in ogni pixel la stima del valore assunto dalla variabile e opzionalmente anche una mappa raster dell'errore di stima di ogni pixel; come nel krigaggio ordinario si prendono in considerazione solo i dati all'interno di un opportuno vicinaggio di stima del pixel da stimare e contenuti entro un numero minimo e massimo di punti, specificati dall'utente insieme all'estensione del vicinaggio. Ho quindi svolto l'operazione di *Cokriging* per i 3 anni in esame, stimando il 2003 con covariabile il 2004, il 2004 due volte: una con covariabile il 2003 e una con covariabile il 2005, il 2005 con covariabile il 2004, e specificando negli

appositi campi i parametri necessari alla stima: modelli di semivariogramma, vicinaggio, numero minimo e massimo di punti da considerare all'interno di tale vicinaggio di stima, riassunti nella seguente tabella.

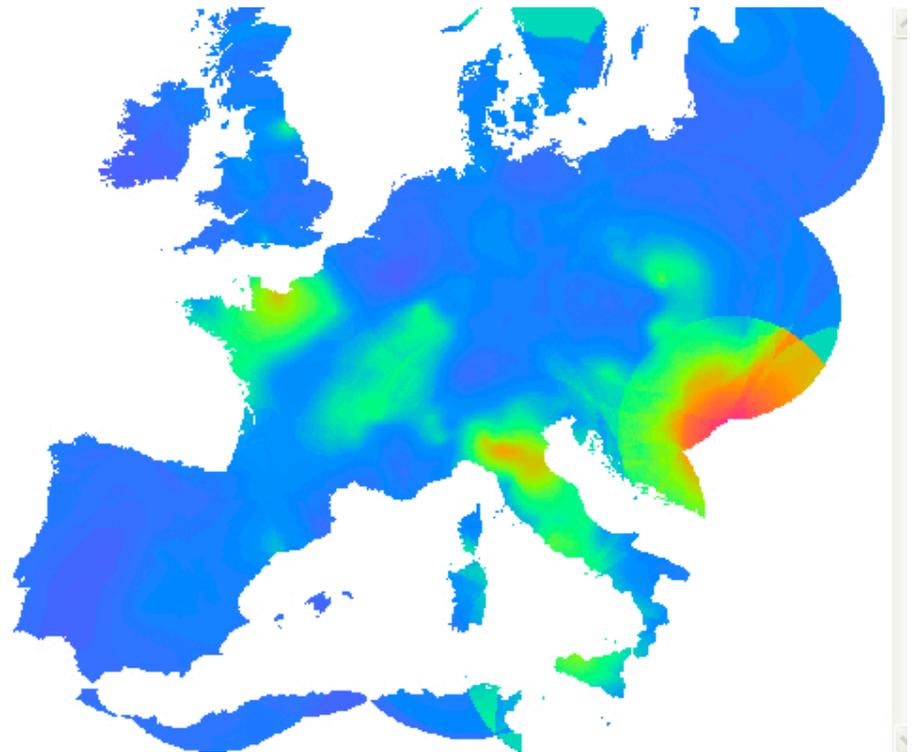
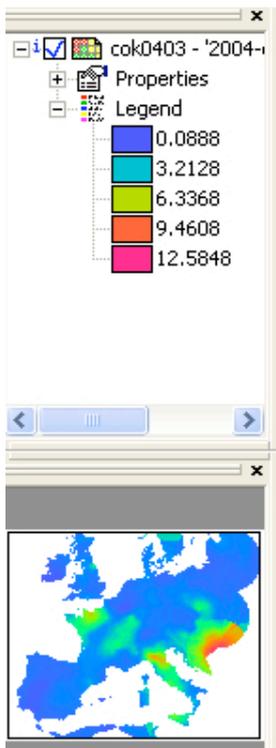
Variabile stimata	Variabile ausiliaria	Vicinaggio	Modello e parametri (nugget, sill, range)					
2003	2004	400000	2003	Sferico	0.54	2.77	350000	
		(min=1,	2004	Sferico	1.13	3.75	350000	
		max=100)	'03/'04	Sferico	0.30	2.10	350000	
2004	2003	400000	2004	Sferico	1.13	3.75	350000	
		(min=1,	2003	Sferico	0.54	2.77	350000	
		max=100)	'04/'03	Sferico	0.30	2.10	350000	
	2005	2004	400000	2004	Sferico	1.13	3.75	350000
			(min=1,	2005	Sferico	0.57	2.43	350000
			max=100)	'04/'05	Sferico	0.40	1.65	350000
2005	2004	400000	2005	Sferico	0.57	2.43	350000	
		(min=1,	2004	Sferico	1.13	3.75	350000	
		max=100)	'05/'04	Sferico	0.40	1.65	350000	



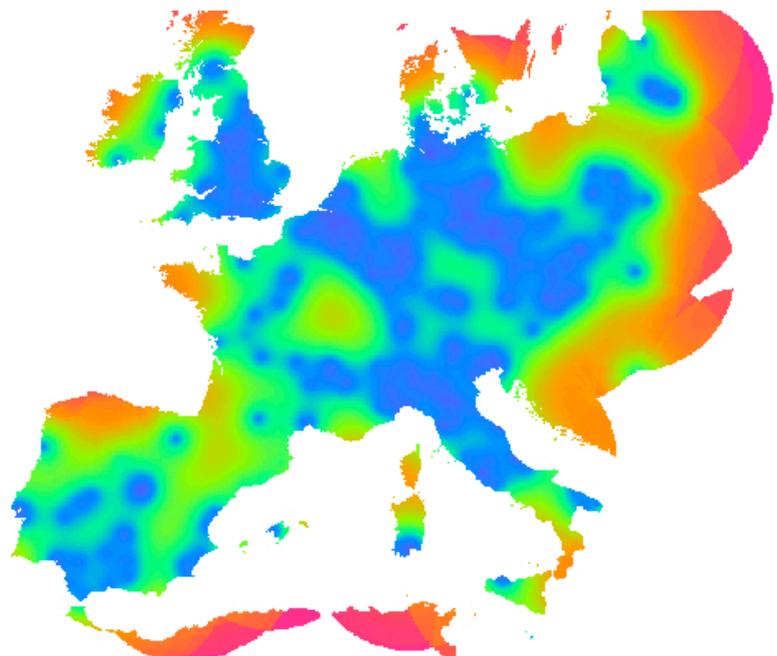
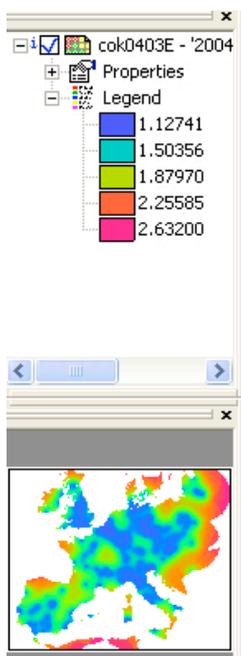
Cokrigaggio 2003-2004



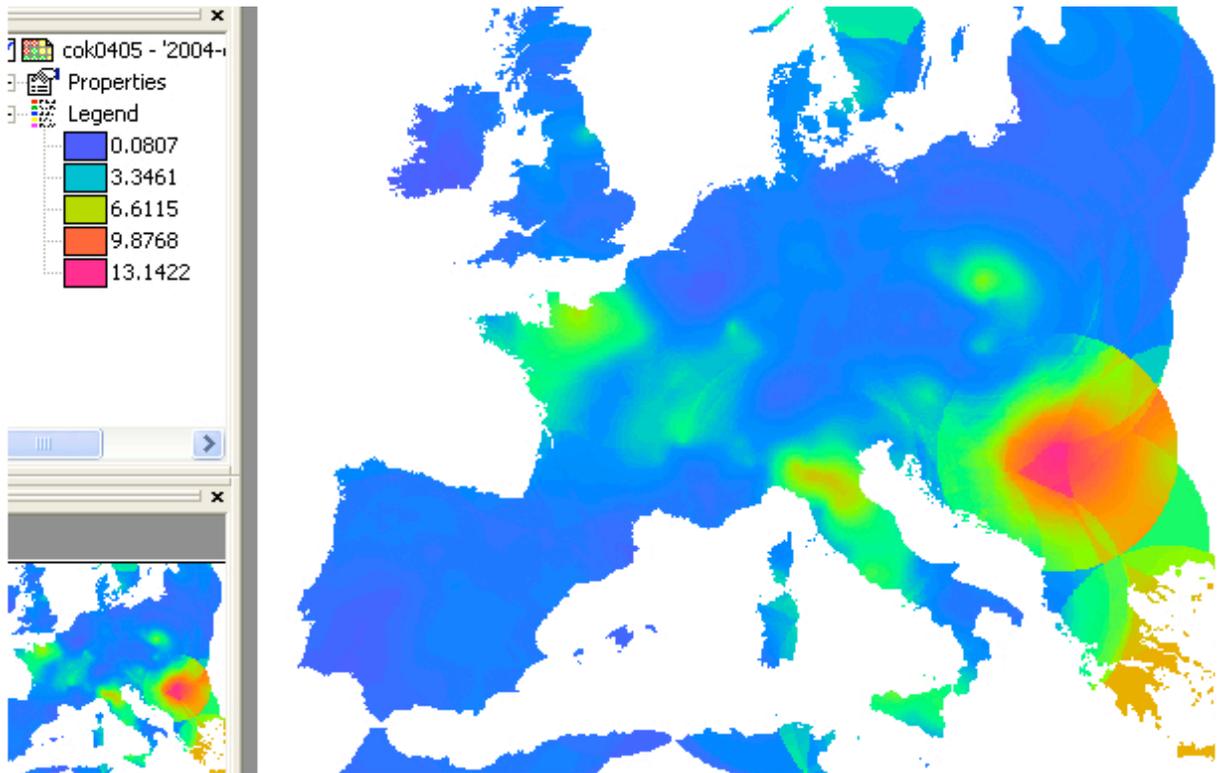
Varianza di stima Cokrigaggio 2003-2004



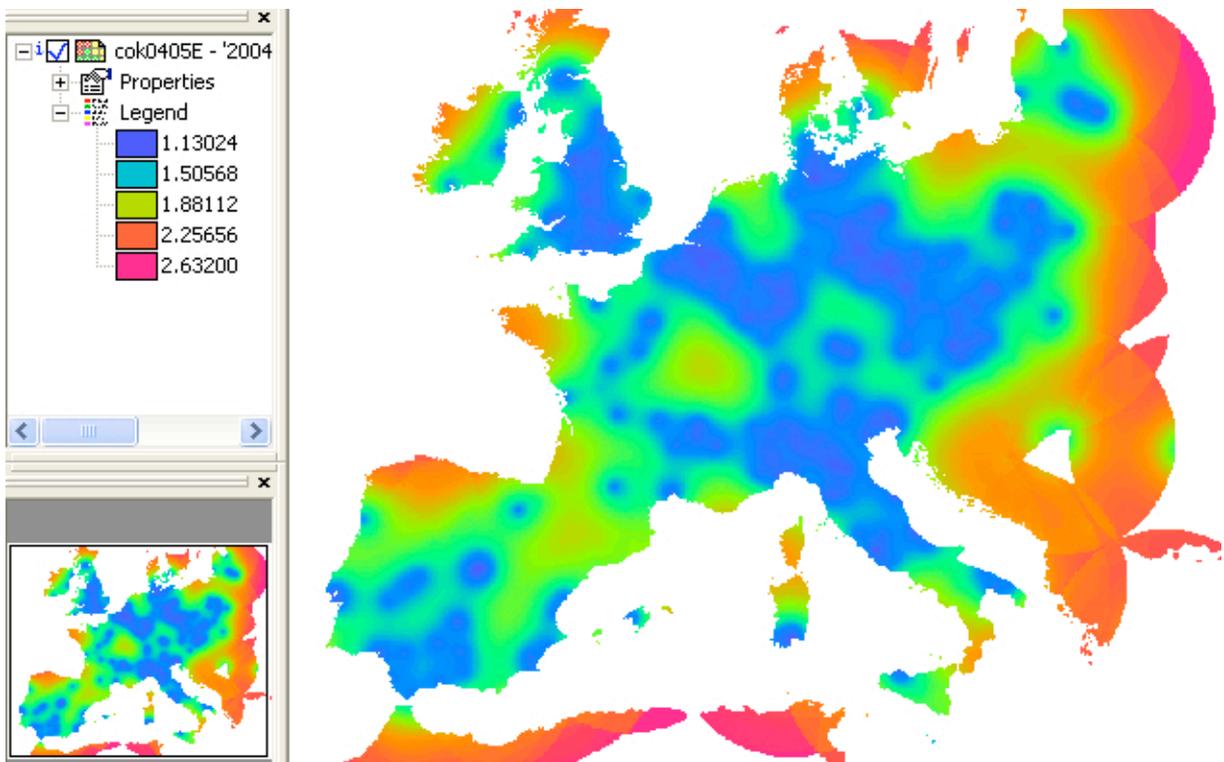
Cokrigaggio 2004-2003



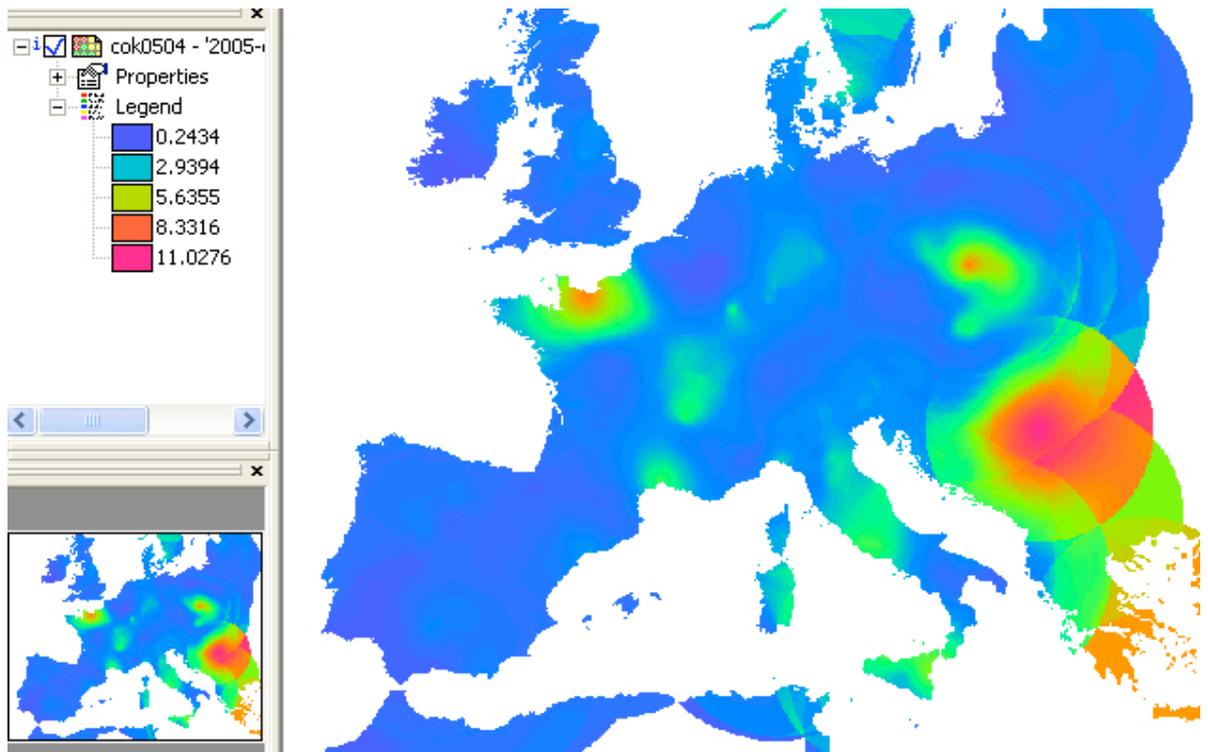
Varianza di stima Cokrigaggio 2004-2003



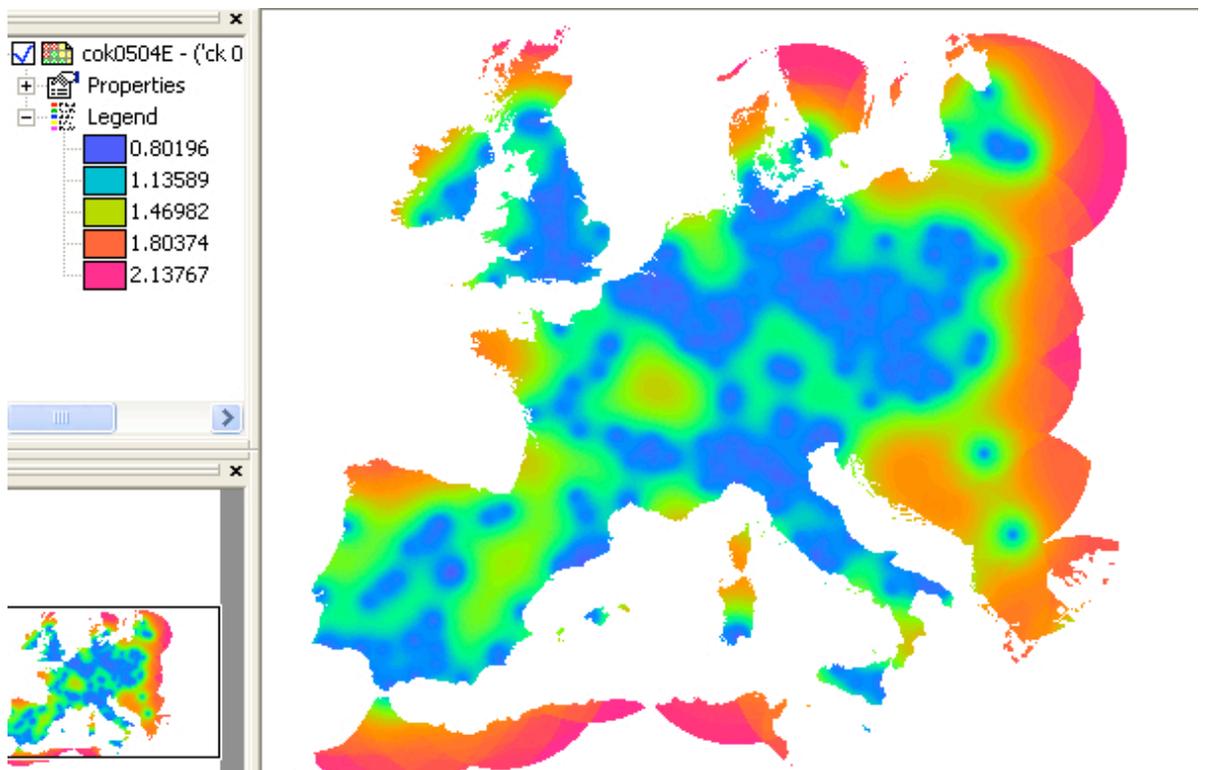
Cokrigaggio 2004-2005



Varianza di stima Cokrigaggio 2004-2005



Cokrigaggio 2005-2004



Varianza di stima Cokrigaggio 2005-2004

5. INTEGRAZIONE DI INFORMAZIONI GEOGRAFICHE: LO STUDIO MULTIVARIATO

La terza fase in cui si è articolato il caso studio ha visto l'integrazione di grandezze fisico, climatiche, geografiche desumibili da set di dati e cartografie tematiche a scala europea, potenzialmente correlabili con la dispersione dell'inquinante come ancora in fase di studio. L'analisi della correlazione tra il benzene e tali variabili ausiliarie opportunamente ricavate, omogeneizzate e processate con metodi quali l'analisi delle componenti principali, al fine di ridurle, ha consentito di:

- Studiare il comportamento della variabile in esame in funzione di vari fattori (clima, uso del suolo, distanza dalle strade principali..ecc) e quindi sviluppare ipotesi sulle principali fonti e cause di dispersione oltre che sulla variabilità di concentrazione;
- Ottenere una cartografia del benzene a scala europea, mediante krigaggio con deriva esterna, avvalendosi delle grandezze maggiormente correlate quindi con una stima multivariata e ipotizzando un comportamento non stazionario della variabile;
- Fornire elementi di confronto con le mappe precedentemente ottenute dagli studi monovariato e multivariato temporale stazionari, per la successiva fase di confronto e validazione mediante test kriging.

Basilare è stata l'estrazione dei valori puntuali delle variabili ausiliarie, una volta note su tutto il territorio europeo e in formato di mappe raster, nei punti di campionamento della variabile benzene, in modo da poter quantificare le correlazioni e determinare la variabile da utilizzare come deriva esterna nella stima conclusiva, quindi come rappresentativa della variabilità sistematica, o trend della variabile studio.

Di seguito verranno riportati tutti i passaggi logici, operativi e i risultati conseguentemente ottenuti che hanno caratterizzato tale studio multivariato,

partendo dalla natura del materiale informativo ausiliario, analizzando le operazioni propedeutiche al suo utilizzo, quali conversioni dei formati, elaborazioni su gis, omogeneizzazione dei georiferimenti, fino ad arrivare allo studio della correlazione con il benzene e alla stima finale.

5.1. INFORMAZIONI AUSILIARIE DISPONIBILI

Il materiale informativo per l'analisi multivariata mi è stato fornito dal JRC di Ispra ed è così costituito:

1) Informazione climatica: informazione puntuale, in formato di tabella contenete nelle proprie colonne le coordinate geografiche, in sistema lat-lon, e i valori assunti dalle seguenti grandezze:

- precipitazione mensile (prc) (mm/mese);
- giorni piovosi medi mensili (rd0) (-);
- velocita' del vento media mensile a 10 m (wind) (m/s);
- temperatura media mensile (tmp) (deg C);

Per un totale di quattro tabelle con informazioni climatiche riferite ai valori medi nei 12 mesi dell'anno

2) Informazione topografica:

- DEM: Digital Elevation Model, modello digitale delle quote, nel formato di mappa raster in coordinate chilometriche INSPIRE e pixel di dimensione 1 km, che copre l'estensione europea;
- Openness: un indicatore di conformazione del terreno (landform) ottenuto dividendo il valore del DEM per il valore della media dei pixel circostanti (9 km di raggio), che evidenzia bene i fondovalle e i crinali; il formato è identico al DEM: mappa raster in coordinate chilometriche INSPIRE e dimensione del pixel di 1 km, che copre l'estensione europea;

3) Carta delle strade principali (*road network dalla digital chart of the world ESRI*), nel formato di una mappa di segmenti e in coordinate geografiche

lat-lon, che copre l'estensione europea;

- 4) Cartografia dell'uso del suolo, nel formato di 11 mappe raster, ognuna rappresentante una determinata classe di uso del suolo, con dimensione del pixel di 1km; le mappe sono frutto di un'aggregazione a 1km dal momento che la cartografia originale era relativa a una dimensione del pixel di 100 m, pertanto in tali mappe ogni pixel mostrerà la percentuale di territorio corrispondente alla classe da 1 a 11; coprono l'estensione europea e sono in coordinate chilometriche ETRS LAEA.

Le 11 classi sono le seguenti:

- Continuous urban fabric
 - Discontinuous urban fabric
 - Industrial or commercial units
 - Road and rail networks and associated land
 - Port areas
 - Airports
 - Mineral extraction sites
 - Dump sites
 - Construction sites
 - Green urban areas
 - Sport and leisure facilities
- 5) Intensità della luce notturna: *lights at night intensity* riferita all'anno 2003 formato di mappa raster in coordinate geografiche lat-lon e dimensione del pixel di 30 primi, copre l'estensione mondiale.

5.2. IDENTIFICAZIONE DELLE VARIABILI AUSILIARIE

A partire da tale set di informazioni ausiliarie, ho svolto una serie di operazioni propedeutiche allo studio della correlazione e alla successiva fase di stima con deriva esterna, operazioni quali conversione del formato dati in formato raster, omogeneizzazione dei supporti temporali a quello annuale del

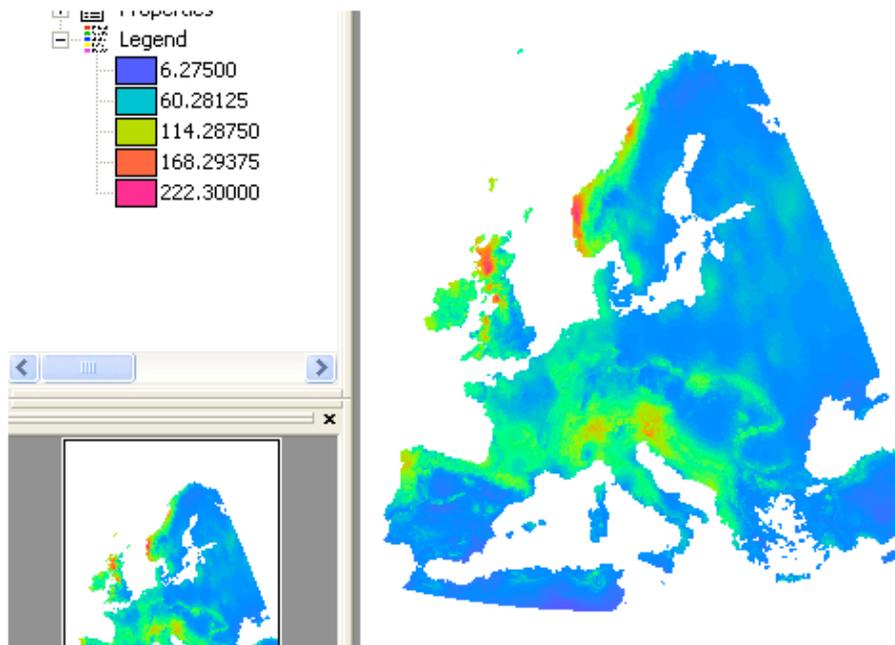
benzene, elaborazioni GIS per estrarre informazioni di interesse nel formato di mappa raster e conversione delle mappe ottenute ad un unico georeference (quindi a stesse dimensione del pixel, sistema di coordinate e di proiezione). Si precisa che in ambiente ILWIS il formato di mappa raster è indispensabile nello studio della correlazione, in quanto oltre a permettere l'estrazione di valori puntuali in corrispondenza delle stazioni di campionamento del benzene, consente di raggruppare le mappe in *map-list*, da cui è possibile accedere a statistiche quali: analisi delle componenti principali, matrici di varianza-covarianza, matrici di correlazione; ovviamente le mappe all'interno della *map list* devono essere riferite ad uno stesso sistema di georeferenziazione.

Si riporta di seguito una breve descrizione delle operazioni di formattazioni svolte, divisa in funzione della tipologia delle grandezze ausiliarie precedentemente esposta e mantenendo lo stesso ordine.

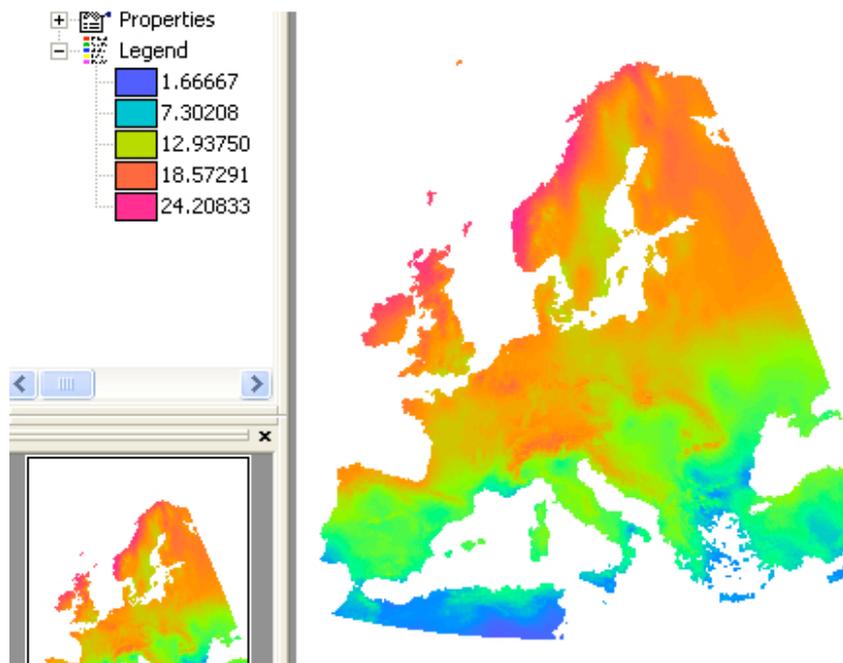
5.2.1. Costruzione delle mappe climatiche

Essendo i dati climatici in formato di 4 tabelle contenenti ognuna in 12 colonne i valori delle medie mensili delle diverse grandezze (temperatura media mensile, velocità del vento media mensile a 10 m, giorni piovosi medi mensili, precipitazione mensile), la prima operazione che ho svolto è stata quella di calcolare una media mensile su supporto annuale, valore quindi confrontabile con la variabile concentrazione media annua di benzene. Da tali valori in formato tabulare ho poi creato delle *point map*, convertito il sistema di riferimento da geografico (lat-lon) a ETRS LAEA, adottato anche in questa fase della trattazione in accordo con gli studi mono e multivariato, mediante l'operazione vettoriale di *transform point*; ho convertito le mappe puntuali così ottenute al formato raster (operazione *rasterize*), impostando un georeference con dimensione del pixel di 1 km, in accordo con le caratteristiche del formato raster delle mappe DEM e uso del suolo, e sistema

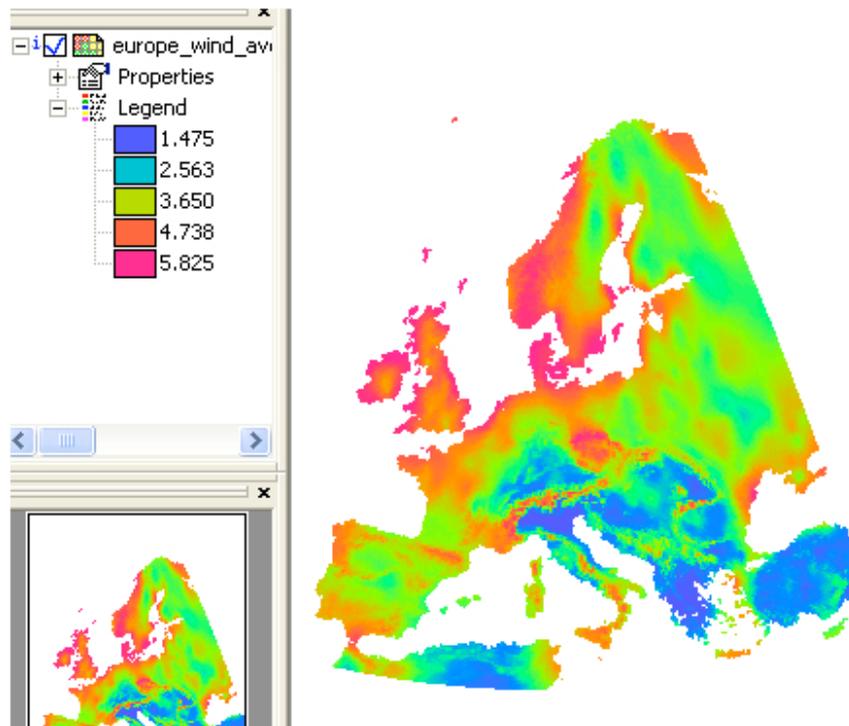
di riferimento ETRS LAEA. Per quanto riguarda la scelta del numero di pixel atto a rappresentare il punto della *point map* originaria nella mappa raster, il criterio è stato quello di ottenere una mappa continua delle variabili climatiche sul suolo europeo, ed ha portato, dopo diversi tentativi, all'individuazione di un valore di 20 pixel. I risultati ottenuti sono di seguito riportati.



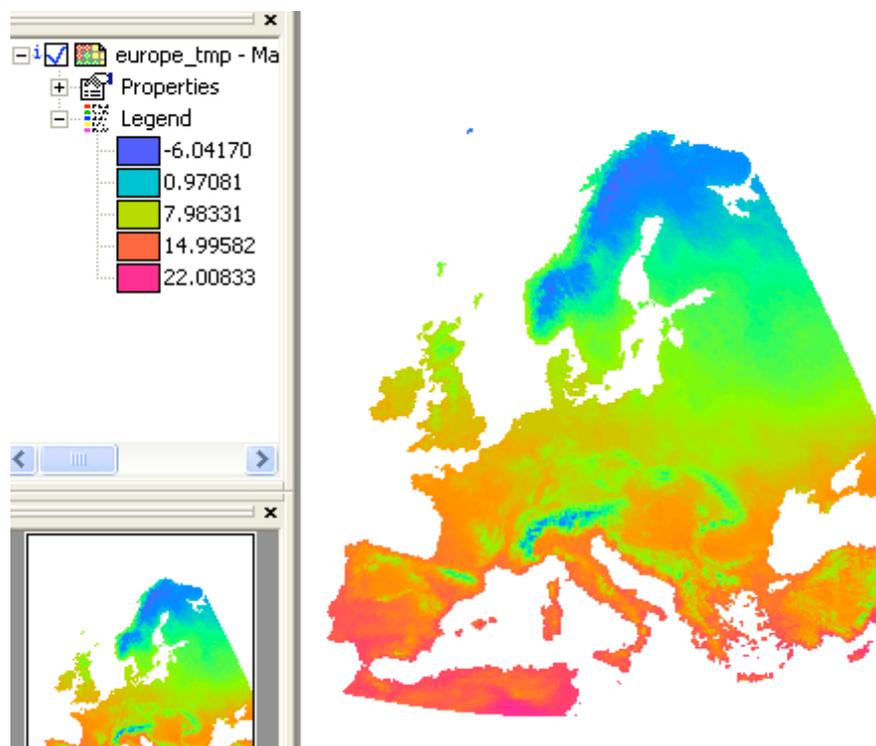
Mappa della media su base annuale delle precipitazioni mensili.



Mappa della media su base annuale dei giorni piovosi medi mensili.



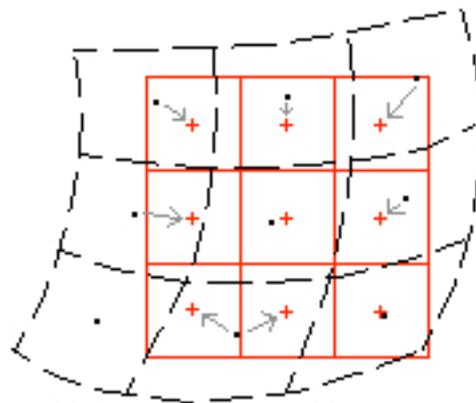
Mappa della media su base annuale della velocità del vento media mensile a 10 m.



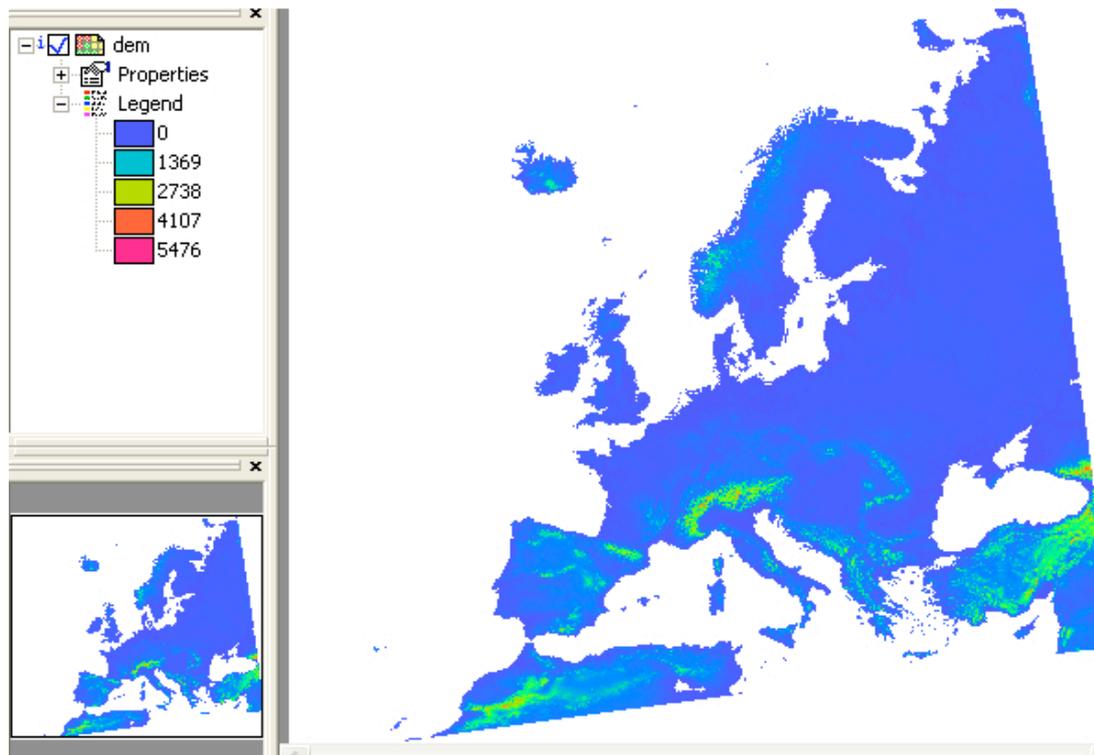
Mappa della media su base annuale della velocità della temperatura media mensile.

5.2.2. Formattazione dell'informazione topografica.

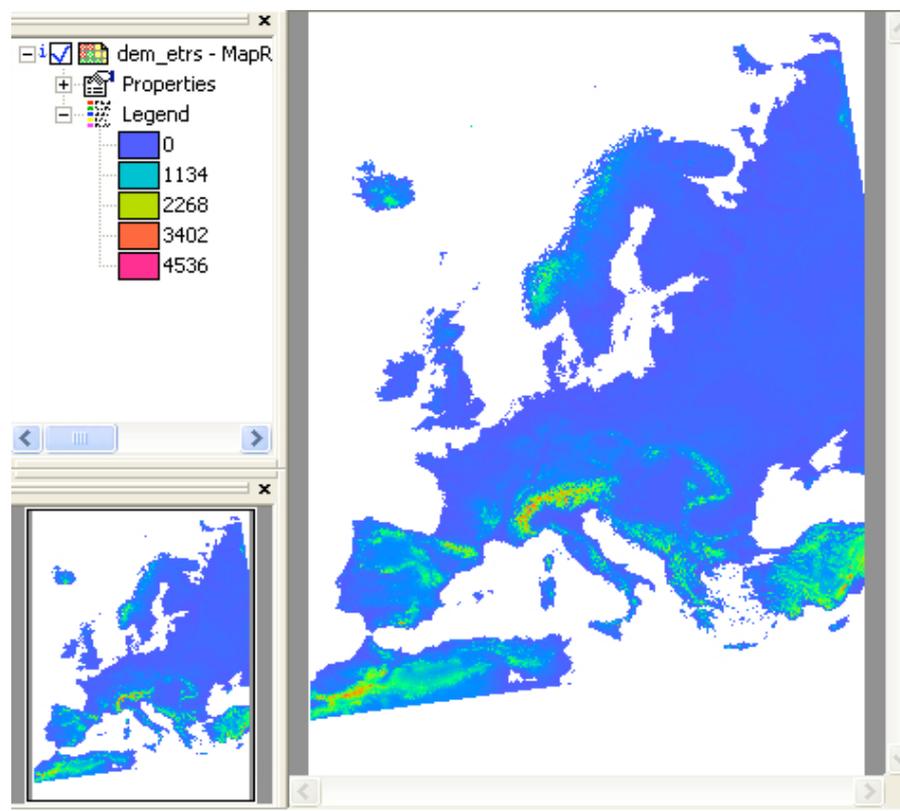
L'informazione topografica a mia disposizione rappresentata dalle mappe DEM e Openness, è già disponibile in formato raster con dimensione del pixel 1 km e sistema di coordinate INSPIRE, l'operazione che ho svolto è stata quindi un cambio di georeference in modo da ottenere una mappa nel sistema di coordinate prescelto per il caso studio: ETRS LAEA. Tale operazione a partire da mappa raster è stata possibile mediante l'operazione di *Resample Nearest Neighbour*, ossia un cambio di georefernce che assegna ad ogni pixel della mappa output il valore del pixel originario maggiormente compreso entro i limiti di un nuovo pixel, come esemplifica l'immagine seguente.



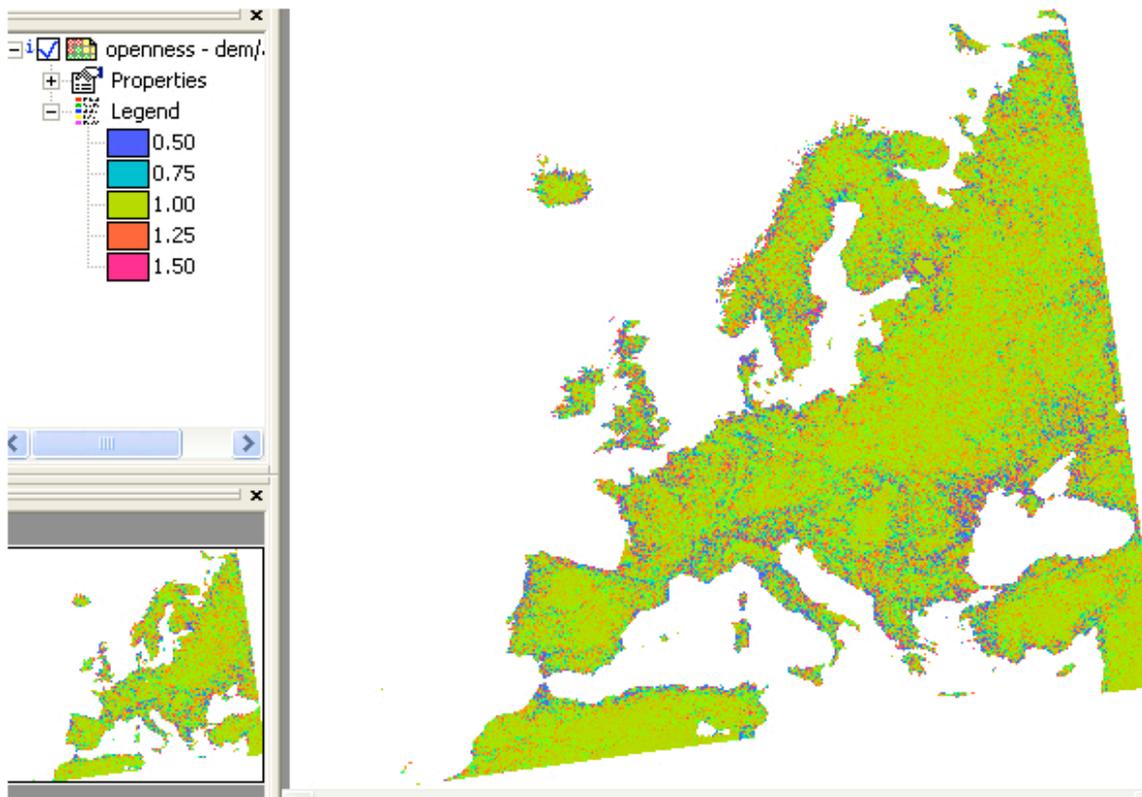
Nello specifico *Resample Nearest Neighbour* delle mappe DEM e Openness l'operazione non modifica di molto le informazioni originarie contenute in tali mappe, dal momento che il passaggio è effettuato tra due georefernce aventi la stessa dimensione del pixel 1 km, e entrambi un sistema di coordinate kilometrico (piano). Si riportano di seguito le mappe ottenute.



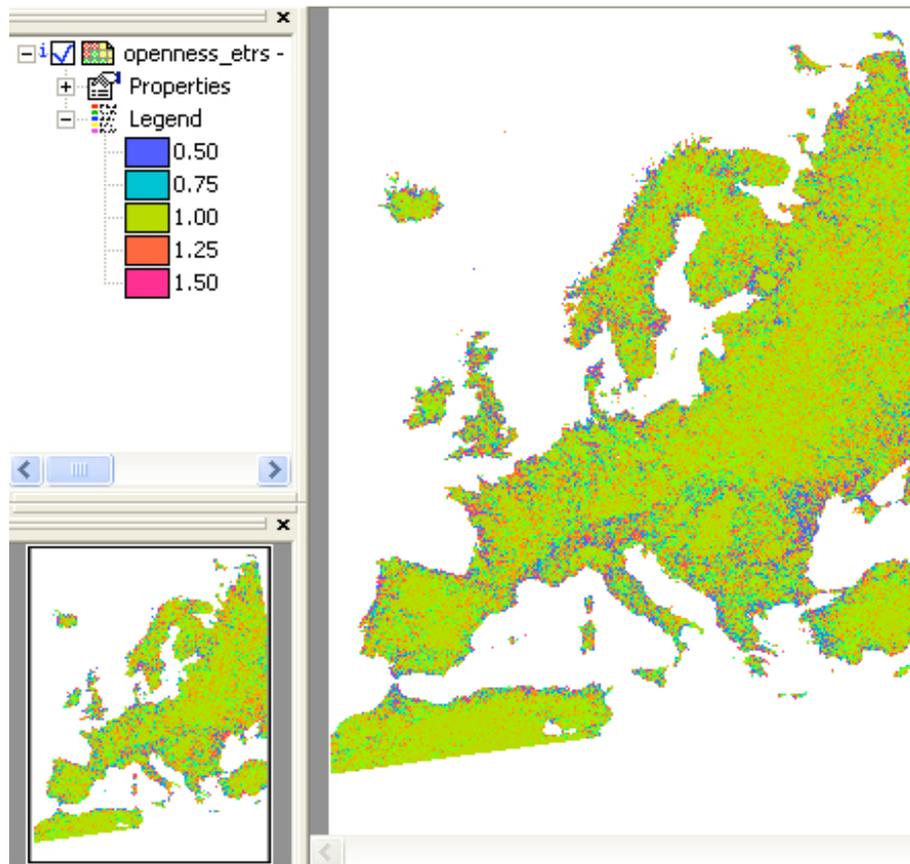
DEM : mappa originale



Dem dopo l'operazione di resample



Openness: mappa originale



Dem dopo l'operazione di resample

5.2.3. Elaborazione della carta delle strade principali: densità di strade e distanza dalle strade principali

Il materiale informativo a disposizione consiste nella *segment map* delle strade principali: “*Road network dalla digital chart of the world ESRI*”; da tale carta in formato di *segment map* ho svolto delle elaborazioni GIS per estrarre due informazioni, e quindi due variabili ausiliarie di interesse, nel formato di mappa raster:

- Densità di strade;
- Distanza dalle strade principali;

La densità di strade è facilmente ottenibile in ambiente ILWIS a partire dal formato di *Segment Map* sfruttando l’operazione di *Segment Density* che rasterizza la mappa di segmenti attribuendo ad ogni pixel la somma della lunghezza dei segmenti contenuti in quel pixel; bisogna introdurre un georeference per la mappa raster così calcolata che ho scelto coincidente con quello precedentemente adottato ed essendo la dimensione del pixel di 1km ne segue che la variabile “Densità di strade” così creata fornisce informazioni relative alla lunghezza totale delle strade contenute in un 1 km².

Per quanto concerne il calcolo della distanza dalle strade principali in ambiente ILWIS è possibile una *Distance Operation* che a partire da una mappa fonte in formato raster, con pixel definiti e pixel indefiniti, ne fornisce un'altra, sempre raster, che contiene in ogni pixel la distanza più piccola in metri verso i pixel definiti della mappa fonte; ossia per tutti i pixel che nella mappa fonte hanno un valore indefinito viene calcolato un valore di distanza.

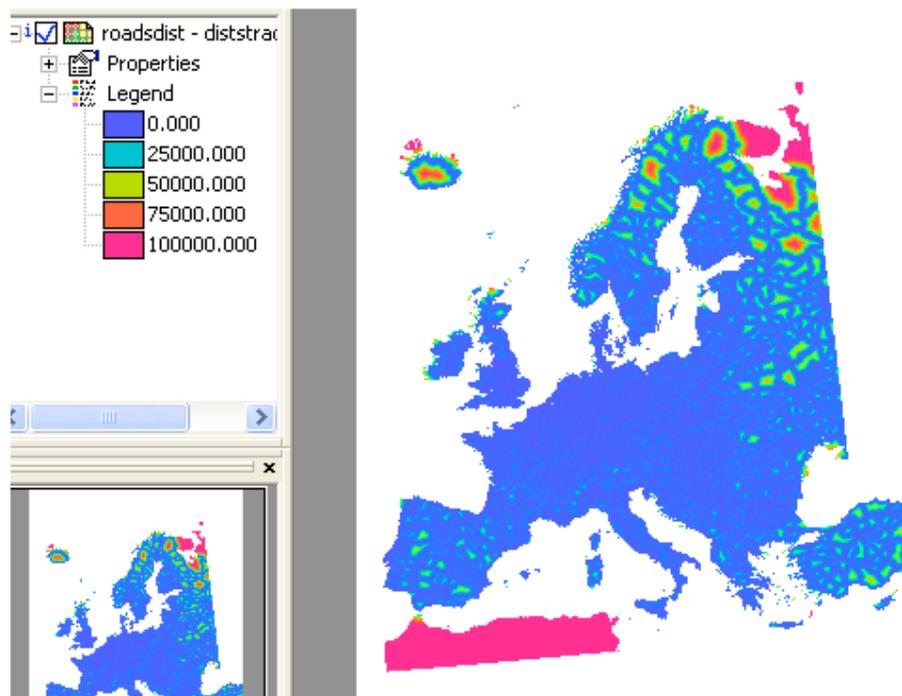
Dal momento che l’informazione relativa alle strade principali è nel formato di *segment map* e coordinate geografiche lat-lon, il calcolo della variabile “distanza dalle stradi principali” ha richiesto una rasterizzazione della mappa di segmenti, previa conversione del sistema di coordinate da geografiche a ETRS LAEA. Nel calcolo della mappa raster delle distanze ho inserito

ovviamente il georeference comune a tutte le mappe fin qui esaminate, con dimensione del pixel di 1km e sistema di coordinate ETRS LAEA.

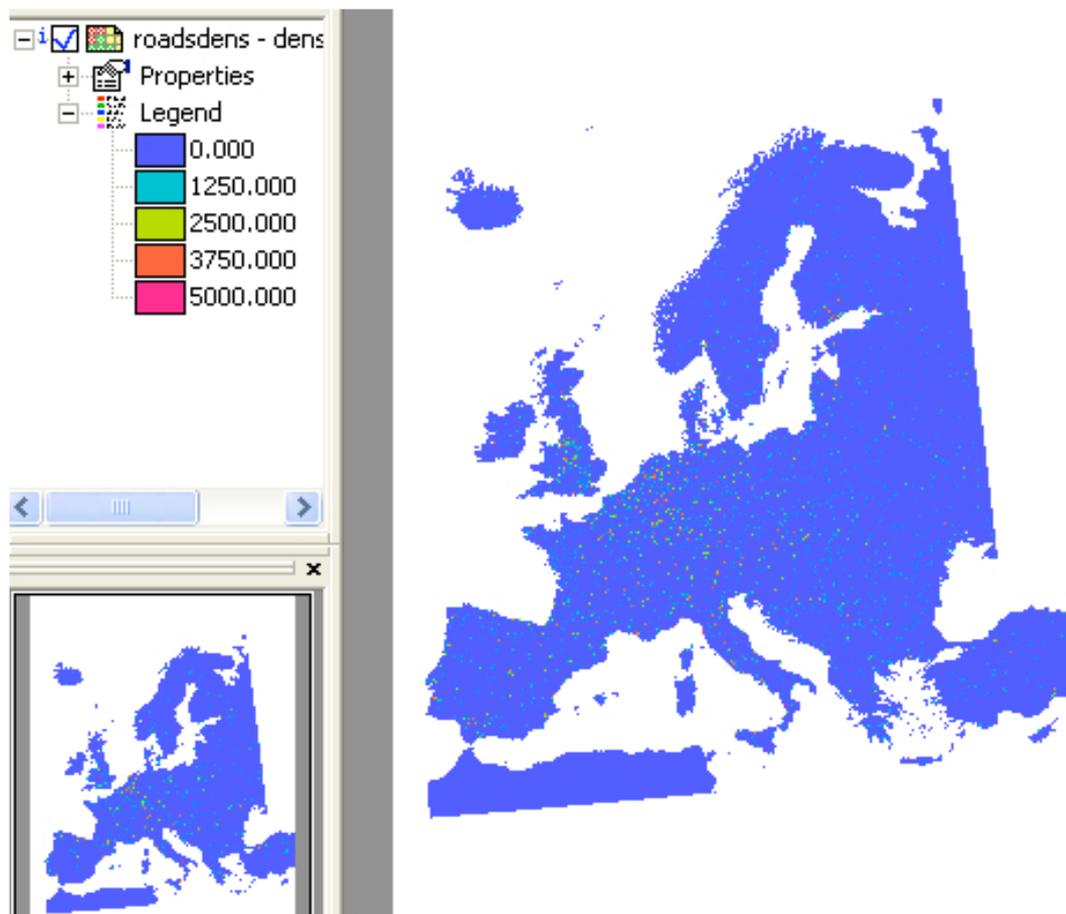
Le immagini seguenti mostrano la *Segment Map* delle strade principali in input e le due mappe raster, elaborate e calcolate a partire da questa, corrispondenti alle variabili ausiliarie così create: densità delle strade e distanza dalle strade principali.



Segment map delle strade principali



Distanza dalle strade principali



Densità di strade

Le variabili ausiliarie desunte dall'elaborazione dell'informazione delle strade principali mostrate nelle mappe sopra riportate hanno le seguenti soglie di variabilità:

- Distanza dalle strade principali: variabile da un minimo di zero a un massimo di 675276.82129 m.
- Densità di strade: variabile da un minimo di zero a un massimo di 14335.773300 m.

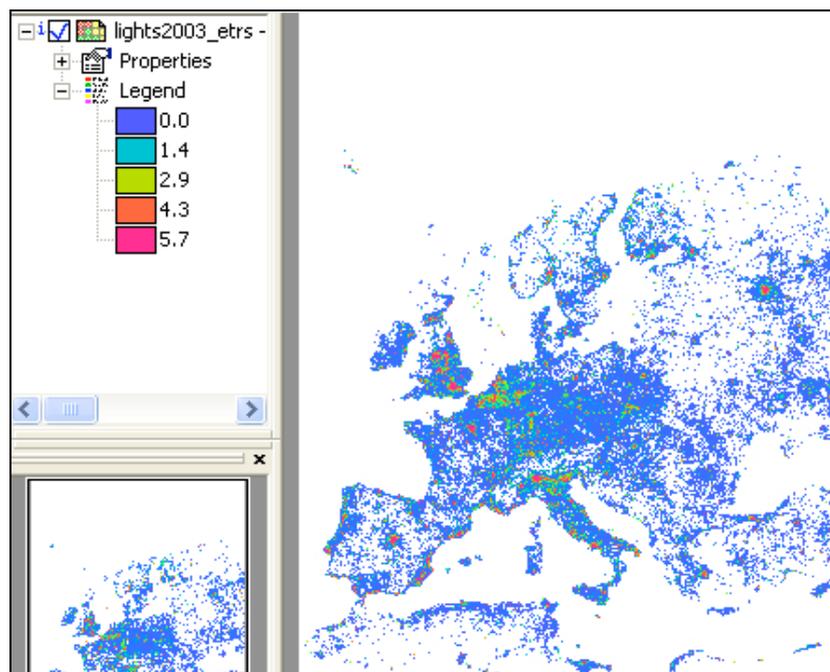
Tuttavia essendo la variabilità della prima variabile per la maggiorparte dei pixel compresa entro i 100000 m, e della seconda entro 5000 m, ho scelto di considerare tali intervalli al fine di ottimizzare la visualizzazione delle mappe e migliorare la comprensione delle informazioni che contengono.

5.2.4. Cartografia dell'uso del suolo.

Le 11 mappe dell'uso del suolo, non hanno richiesto operazioni preliminari al loro utilizzo nella valutazione delle correlazioni, dal momento che contengono già le informazioni nel formato e georeference richiesto.

5.2.5. Sottomappa e formattazione della mappa dell'intensità della luce notturna.

La variabile *Lights at night intensity* in formato di mappa raster copre l'estensione mondiale, la prima operazione è stata quindi quella di creare una sottomappa che contenesse solo l'estensione europea, in accordo con il caso studio; questo è stato possibile mediante l'operazione di *Sub Map*, che crea una sottomappa in funzione dei limiti geografici specificati in termini di coordinate. Successivamente ho effettuato un'operazione di *Resample Nearest Neighbour*, la stessa descritta per la formattazione dell'informazione topografica, che ha consentito il passaggio dal sistema di coordinate geografico con dimensione del pixel di 30 primi della mappa input, a quello comune alle altre mappe e prescelto per le successive elaborazioni.



Lights at night intensity

5.3. ANALISI STATISTICA MULTIVARIATA

Una volta ricavate le variabili ausiliarie e ottenute le mappe nel formato desiderato, ho proceduto ad analizzare la correlazione di tali grandezze con la variabile studio, concentrazione media annua di benzene, avvalendomi fondamentalmente di tre strumenti:

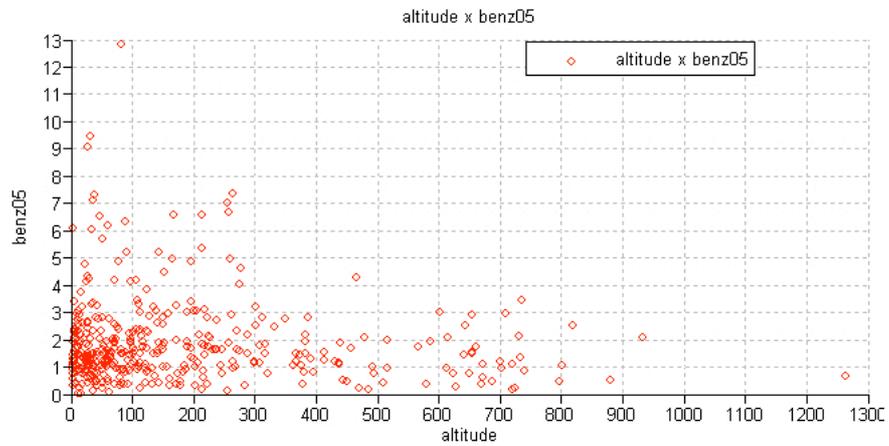
- Nuvole di correlazione, quindi diagrammi cartesiani contenenti in ordinata il benzene e in ascissa la variabile ausiliaria di volta in volta esaminata;
- Matrice di correlazione complessiva con tutte le variabili ausiliare, che permette, variando l'indice di correlazione tra -1 e 1 , di individuare le correlazioni esistenti;
- Analisi delle componenti principali, per la cui teoria si rimanda al capitolo strumenti geostatistici, che ha permesso di studiare le correlazioni tra i due gruppi omogenei di variabili, quali le 4 variabili climatiche e le 11 classi di uso del suolo, al fine di valutare se le componenti principali rappresentative della variabilità di tali gruppi omogenei di variabili, avessero con il benzene una correlazione maggiore che le variabili considerate singolarmente.

Si sottolinea inoltre che tali operazioni sono state svolte per i tre anni presi in esame nel caso studio, 2003, 2004, 2005, e hanno fornito pertanto elementi utili ad una valutazione temporale delle correlazioni.

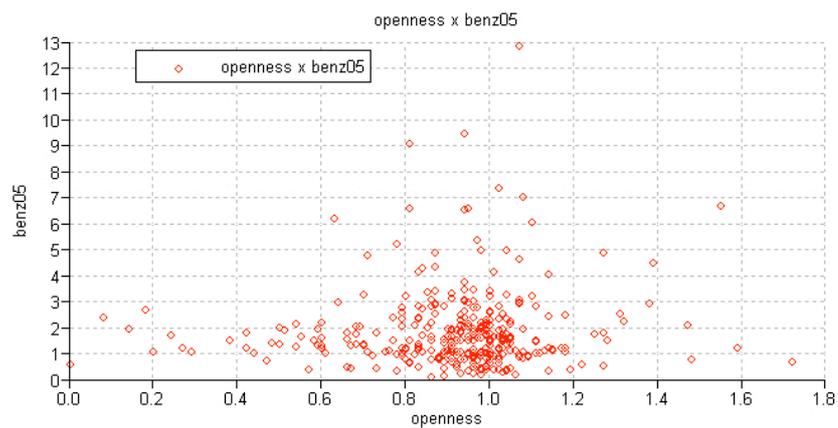
5.3.1. Nuvole di correlazione

In ambiente ILWIS le nuvole di correlazione sono state ottenute a partire dalla tabella associata alle *point map* del benzene per i diversi anni, mediante l'operazione di *map value* che ha consentito di estrarre dalle mappe raster delle variabili ausiliarie i valori assunti in corrispondenza dei punti di campionamento del benzene, e di salvare tali valori in una colonna dedicata. In questo modo a partire dalle operazioni di tabella è stato immediato ottenere

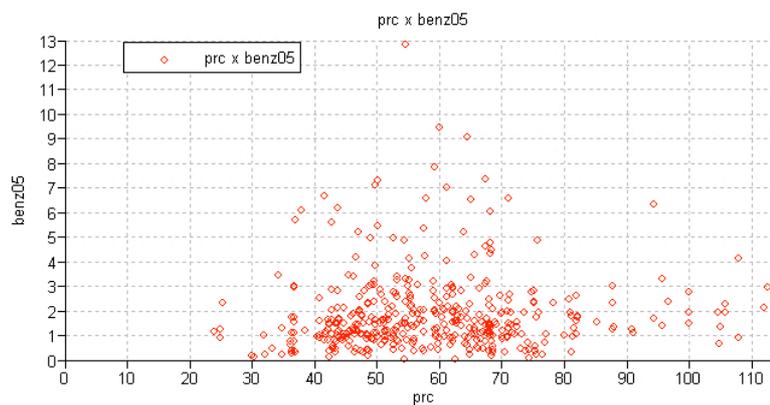
i grafici cartesiani delle nuvole di correlazione per i 3 anni, che evidenziano, come prevedibile, lo stesso andamento; si riportano pertanto le nuvole di correlazione solo per un anno, il 2005.



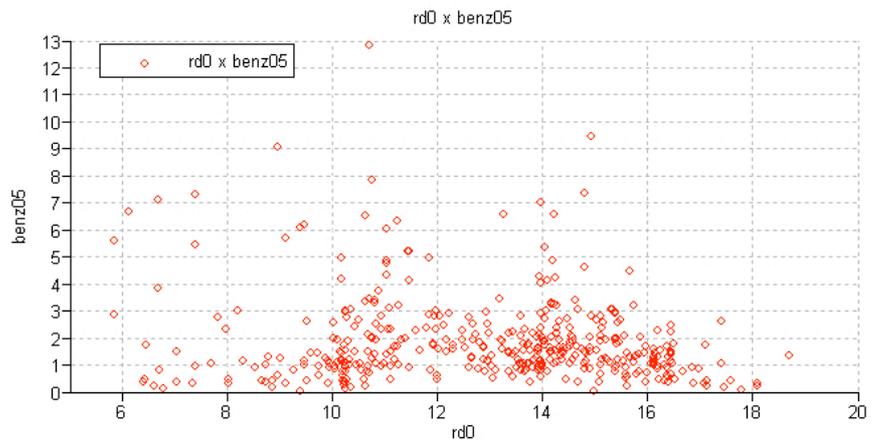
Correlazione altitudine-benzene '05



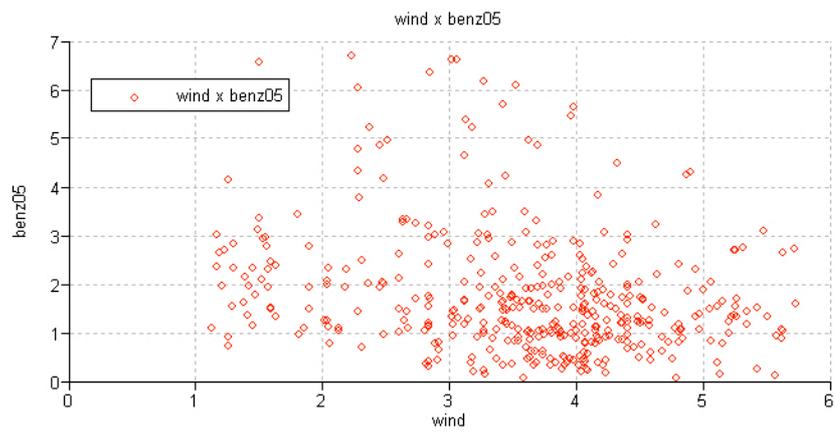
Correlazione openness-benzene '05



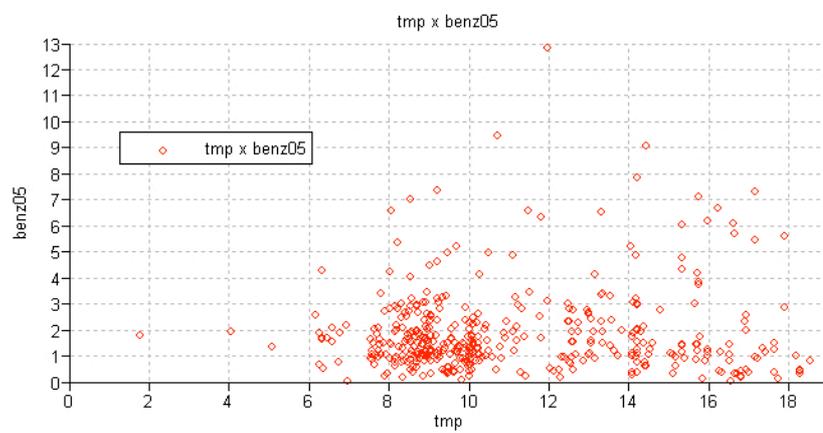
Correlazione prc-benzene '05



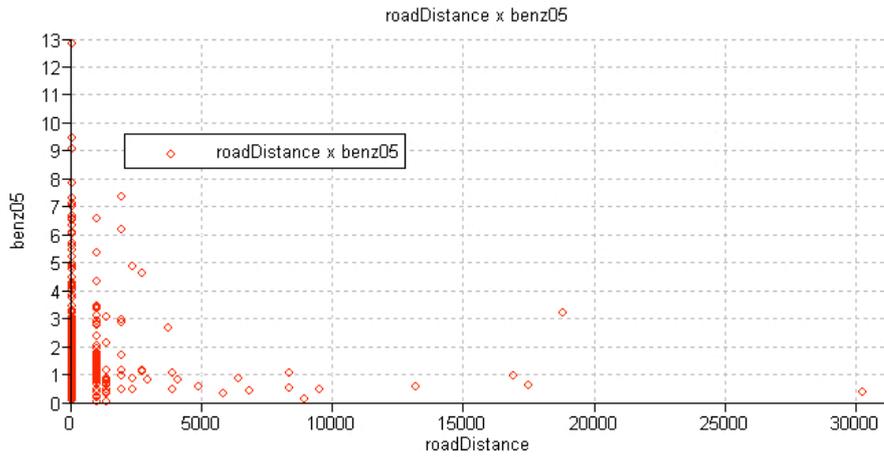
Correlazione rd0-benzene '05



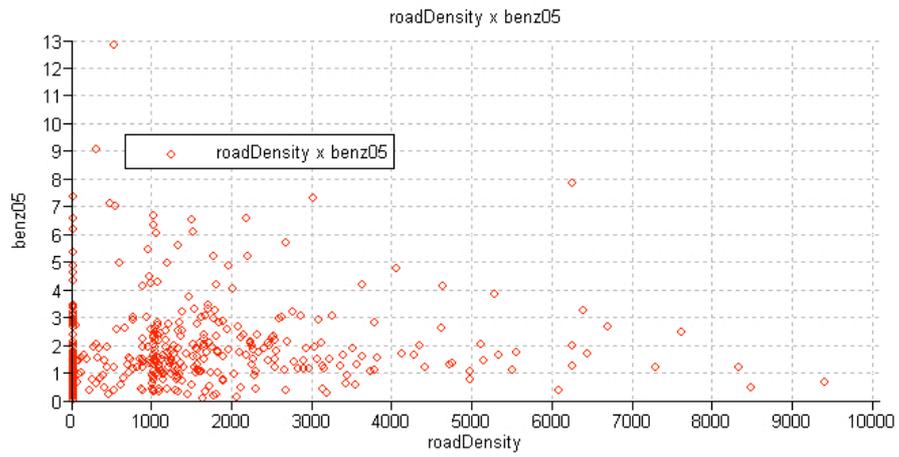
Correlazione wind-benzene '05



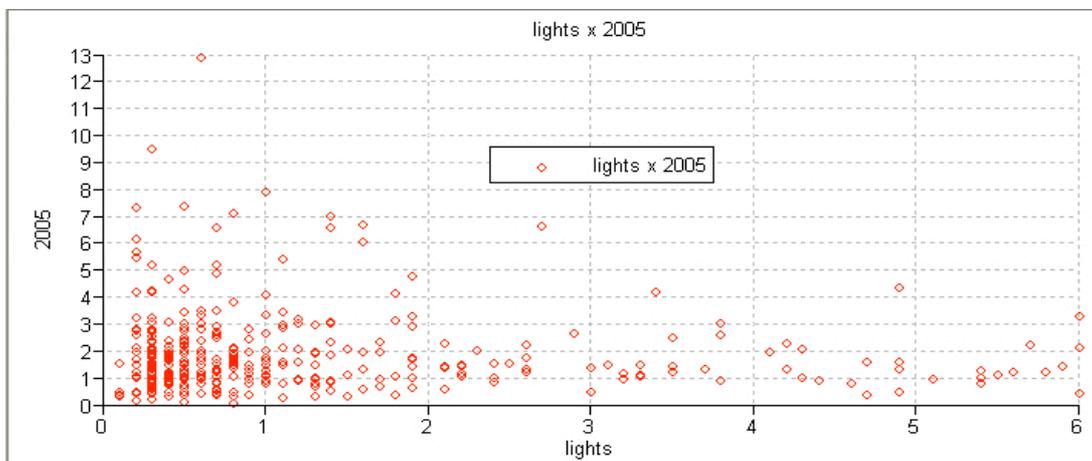
Correlazione tmp -benzene '05



Correlazione roadDistance-benzene '05

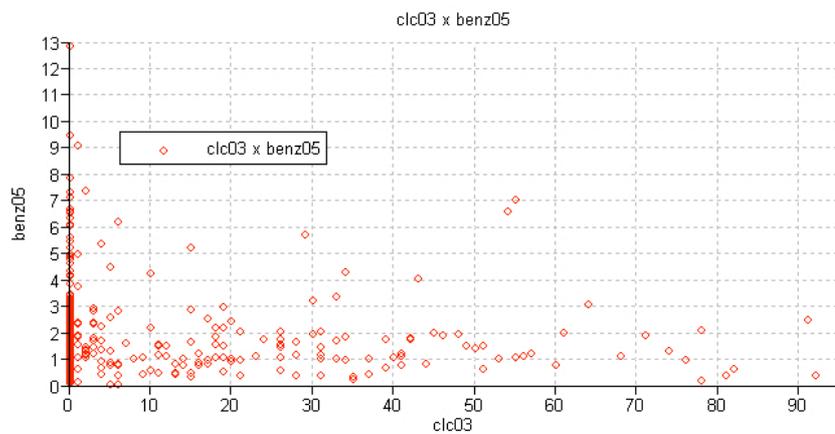
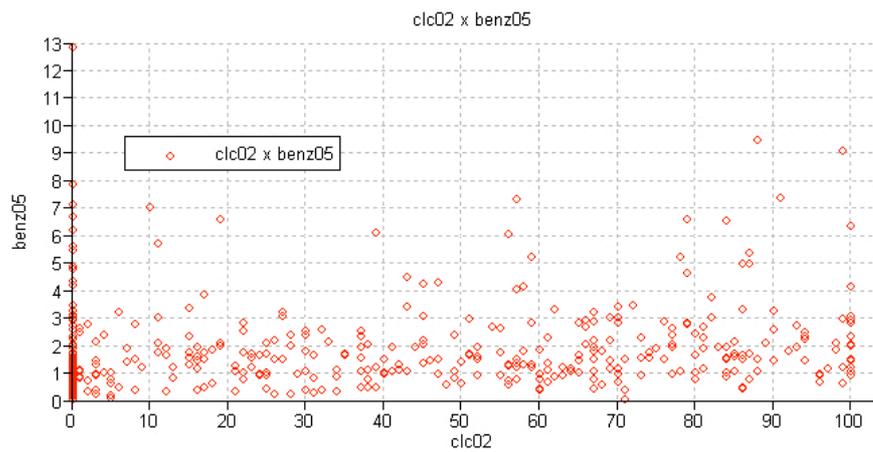
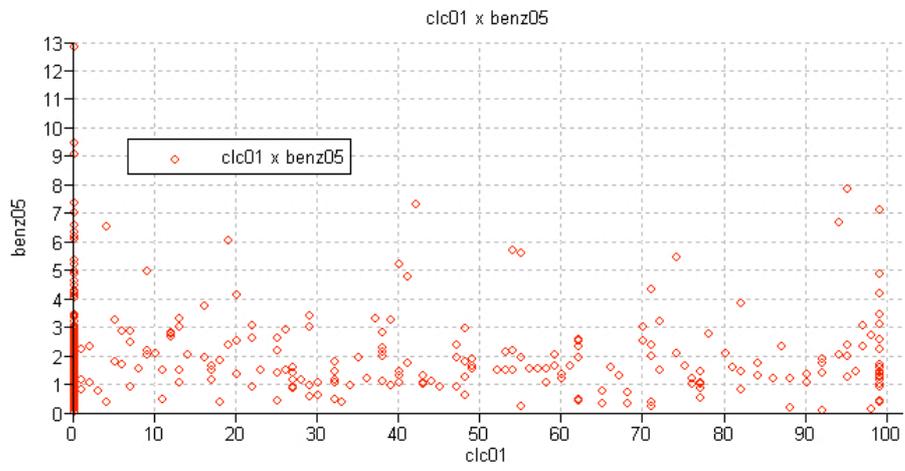


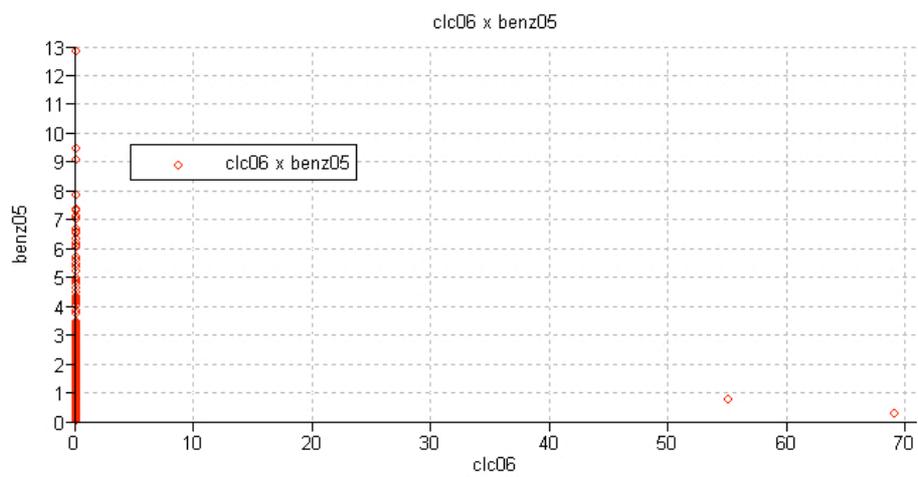
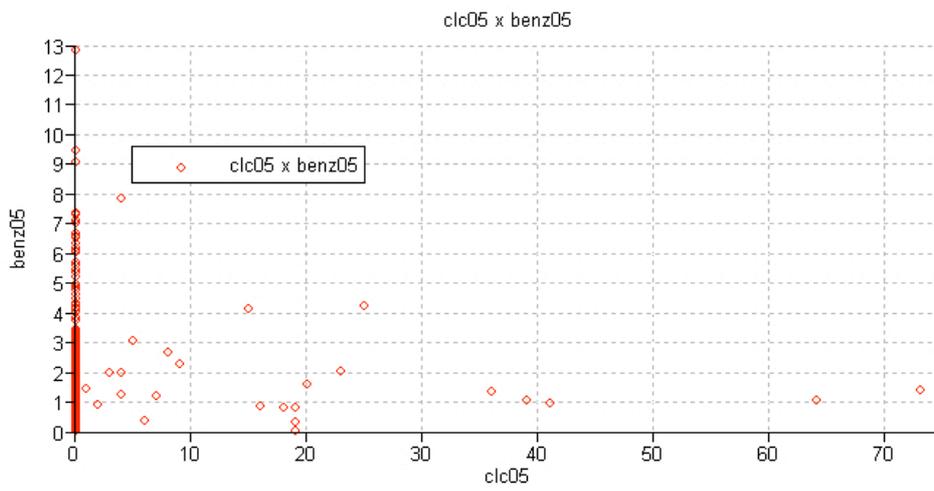
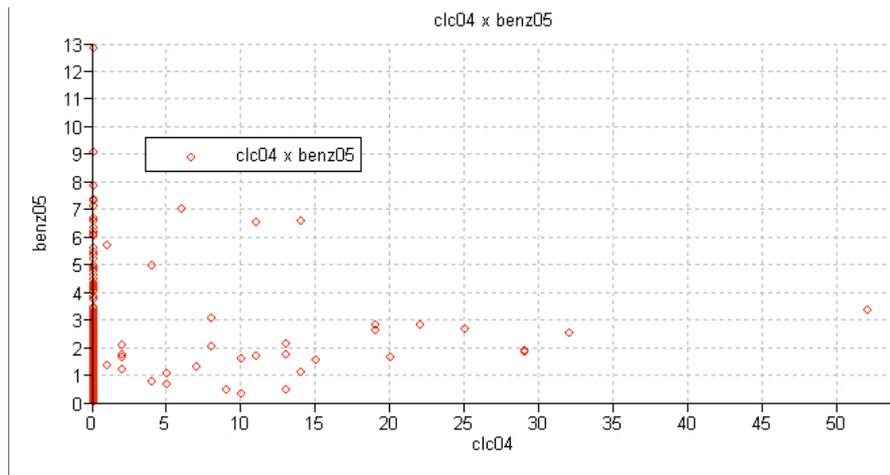
Correlazione roadDensity-benzene '05



Correlazione Lights-benzene '05

Correlazione mappe di uso del suolo 'clc'-benzene '05.





Commentando brevemente i risultati ottenuti: relativamente alla nuvola di correlazione tra il valore dell'altitudine e la concentrazione di benzene, si ravvisa che si ha una forte densità di valori compresi tra 0- 3 $\mu\text{g}/\text{m}^3$, per altitudini comprese entro i 200m , parallelamente però queste sono anche le altitudini alle quali si riscontrano i valori più elevati, dal momento che per valori superiori ai 300m, si nota che il livello di concentrazione dell'inquinante non supera i a $\mu\text{g}/\text{m}^3$.

Relativamente alla correlazione con l'openness, parametro che esprime della conformazione del territorio, si nota come la maggior parte dei valori sia contenuta per valori di tale parametro compresi tra 0,8 e 1 . Si ricorda che l'openness è ottenuto dividendo i valori della mappa DEM per i valori un'area spaziale circostante di 9 km, e evidenzia bene i fondovalle e i crinali, valori tra 0.8 e 1 di tale rapporto indicano quindi un'altitudine della zona circostante quella esaminata leggermente superiore o uguale, quindi zone pianeggianti.

La correlazione con la variabile *prc*, non mette in luce particolari trend. Sia i valori più alti che quelli più bassi di concentrazione sono localizzati infatti all'interno dello stesso range di valori del prc, tra 35 e 70 mm/mese, si può notare comunque come per precipitazioni maggiori i valori di concentrazione di benzene rimangono inferiori a $3\mu\text{g}/\text{m}^3$.

La correlazione con il numero di giorni piovosi medi mensili, risulta approssimabile invece da una retta con coefficiente angolare negativo; si nota infatti come al diminuire del valore del parametro *rd0*, i valori di concentrazione aumentino e lo stesso si può affermare per il vento, per il quale si riscontra una correlazione negativa. Per le temperature si nota che i valori maggiori del benzene si hanno per valori maggiori rispetto al range dentro il quale ricadono la maggior parte dei valori (8 e 10 °C).

La distanza dalle strade mette in luce un netto andamento iperbolico con valori di concentrazione d'inquinante che aumentano al diminuire della distanza delle strade e diminuiscono invece all'aumentare della distanza.

La densità di strade così come le luci notturne non evidenziano particolari correlazioni.

Le nuvole di correlazioni per le classi di uso del suolo mostrano una corrispondenza fino alla classe di uso 4, dopo diventano poco significative; la nuvole per le classi di uso del suolo dalla 6 in poi, con andamento del tutto simile a quello per la classe 6, sono quindi state omesse per tale ragione. Zone urbane continue e discontinue e zone commerciali o industriali (clc01, clc02, clc03), mostrano comunque un andamento piuttosto omogeneo dal quale è difficile ricavare andamenti delineati.

5.3.2. Matrici di correlazione

Le nuvole di correlazione forniscono così una panoramica complessiva delle correlazioni e soprattutto permettono una valutazione qualitativa delle possibili relazioni funzionali tra benzene e variabili ausiliarie (es: lineari, logaritmiche, iperboliche..ecc); risulta tuttavia fondamentale un'analisi quantitativa della correlazione, in modo da poter determinare in maniera univoca le variabili maggiormente correlate. Il passo successivo è stato quindi quello di calcolare le matrici di correlazione che in ambiente ILWIS sono ottenibili a partire da *map list*, quindi per variabili in formato di mappa raster come le variabili ausiliarie, ma non per le variabili ausiliarie e la variabile studio, il benzene, la cui informazione è disponibile in mappe puntuali. Ho allora provveduto al calcolo di tali coefficienti, partendo dalla tabella associata alle *point map* e sfruttando le operazioni statistiche eseguibili da colonna e quindi le colonne di valori in input della concentrazione di benzene congiuntamente ai valori estratti per le variabili ausiliarie, ottenendo i seguenti risultati riassunti in forma tabulare (per le 11 variabili di uso del suolo i coefficienti di correlazione si riportano in una tabella dedicata).

	2003	2004	2005
DEM	-0.109	-0.080463	-0.091107
openness	0.055945	0.084706	0.066540
prc	0.037451	0.127	0.050002
rd0	-0.262	-0.259	-0.142
wind	-0.317	-0.386	-0.212
tmp	0.193	0.157	0.060456
roadDistance	-0.120	-0.135	-0.119
roadDensity	0.143	0.080250	0.067316
Lights	-0.026524	-0.071128	-0.031555

Matrice di correlazione benzene variabili ausiliarie

	2003	2004	2005
clc 01	0.169	0.030300	0.032351
clc 02	0.081578	0.134	0.129
clc 03	-0.119	-0.024945	-0.091210
clc 04	0.036550	0.158	0.086964
clc 05	0.005283	-0.050536	-0.043018
clc 06	-0.038938	-0.040836	-0.057540
clc 07	-0.030803	-0.026585	0.007065
clc 08	-0.039564	-0.030803	-0.022807
clc 09	-0.052920	-0.054381	-0.055621
clc 10	0.036360	0.051412	-0.036676
clc 11	0.027433	-0.020559	-0.012217

Matrice di correlazione benzene variabili ausiliarie 'clc'

Si nota che i valori dei coefficienti di correlazione non superano il – 0.386 tra la variabile ausiliaria ‘wind’ e la concentrazione di benzene nel 2004, questo mi ha portato ad individuare come soglia minima di correlazione, oltre la

quale considerare le correlazioni significative, un valore di $|0.1|$; (ovviamente in valore assoluto potendo essere le correlazioni sia positive che negative). Nelle tabelle sopra riportate sono evidenziate in rosso le righe contenenti valori del coefficiente di correlazione superiore alla soglia stabilita per tutti e tre gli anni considerati, (velocità del vento media mensile a 10 m, giorni piovosi medi mensili, distanza dalle strade principali, uso del suolo 2: discontinuous urban fabric).

5.3.3. Analisi delle componenti principali

La scelta delle variabili da utilizzare nella successiva fase di stima mediante krigaggio con deriva esterna, è stata effettuata solo dopo un'ulteriore studio della correlazione mediante un'analisi delle componenti principali, volta a verificare per due gruppi omogenei di variabili, quali climatiche e uso del suolo, se le componenti principali siano più correlate con la concentrazione di benzene rispetto alle variabili ausiliarie considerate singolarmente.

In ambiente ILWIS l'analisi delle componenti principali è possibile a partire da *map list* mediante l'operazione omonima *PCA* e fornisce come risultato:

- una *map list* delle componenti principali, espresse quindi in formato di mappe raster;
- la matrice di varianza-covarianza tra variabili e componenti principali;
- le percentuali di variabilità da attribuire ad ogni componente.

Ho così provveduto a svolgere due analisi delle componenti principali, la prima a partire dalla *map list* delle 4 variabili climatiche (precipitazione mensile (prc), giorni piovosi medi mensili (rd0), velocità del vento media mensile a 10 m (wind), temperatura media mensile;) la seconda con in input la *map list* delle 11 variabili di uso del suolo, indicate con la sigla 'clc' dal nome del satellite *Corine Land Cover* dal quale sono state ottenute; si riportano di seguito i risultati ottenuti in termini di matrice di correlazione e di percentuali di variabilità, omettendo di mostrare le mappe delle varie

componenti, che non aggiungendo elementi ai fini dell'interpretazione, risultano poco significative.

Variance per band:				
351.95	30.50	2.59	0.58	
Variance percentages per band:				
91.27	7.91	0.67	0.15	
	europa_prc_media_etr	europa_rd0_media_etr	europa_tmp	europa_wind_avg_etr
PC 1	0.994	0.101	-0.046	0.014
PC 2	0.091	-0.501	0.860	-0.045
PC 3	-0.065	0.833	0.504	0.219
PC 4	0.004	-0.211	-0.073	0.975

PCA variabili climatiche

Variance per band:											
24.94	3.13	1.21	1.12	1.06	0.67	0.40	0.19	0.17	0.16	0.15	
Variance percentages per band:											
75.10	9.43	3.63	3.37	3.19	2.02	1.21	0.58	0.52	0.49	0.46	
	clc02	clc03	clc04	clc05	clc06	clc07	clc08	clc09	clc10	clc11	clc01
PC 1	0.998	0.055	0.005	0.002	0.002	0.002	0.001	0.002	0.018	0.015	0.014
PC 2	0.056	-0.997	-0.018	-0.009	-0.009	-0.011	-0.006	-0.005	-0.018	-0.021	-0.02
PC 3	-0.015	-0.024	0.005	0.002	0.011	0.025	0.001	0.003	0.018	0.998	0.048
PC 4	0.013	0.025	-0.010	-0.012	0.000	-0.002	0.000	-0.006	-0.025	0.049	-0.99
PC 5	0.001	0.011	-0.002	0.000	-0.005	-1.000	-0.009	-0.001	0.002	0.025	0.003
PC 6	-0.001	-0.009	0.001	0.000	1.000	-0.005	-0.000	0.003	0.003	-0.011	-0.00
PC 7	0.016	0.019	-0.024	-0.010	0.004	-0.001	-0.002	-0.006	-0.999	0.018	0.027
PC 8	0.000	0.007	-0.041	-0.027	-0.000	0.009	-0.998	-0.021	0.004	0.001	0.001
PC 9	0.002	0.013	-0.271	-0.953	0.001	0.000	0.040	-0.129	0.018	0.003	0.015
PC 10	-0.003	-0.014	0.956	-0.282	-0.002	-0.002	-0.033	0.062	-0.021	-0.004	-0.00
PC 11	-0.001	-0.002	-0.097	-0.107	-0.003	-0.001	-0.014	0.989	-0.003	-0.002	-0.00

PCA variabili climatiche

Dal calcolo della correlazione tra tali componenti principali e la variabile benzene considerata per i tre anni in esame è emerso che le correlazioni non sono maggiori rispetto a quelle con le variabili maggiormente correlate considerate singolarmente; quali : wind e rd0 e clc02 evidenziate contestualmente al calcolo dei coefficienti di correlazione.

5.4. ANALISI STRUTTURALE CON DERIVA ESTERNA

Il passo conclusivo dello studio multivariato, come per le precedenti fasi in cui si è articolata la presente analisi, è rappresentato dalla fase di analisi strutturale e successiva stima, questa volta nell'ipotesi di situazione non stazionaria; generalmente ciò si verifica quando la variabile studio mostra, in accordo con la fisica del fenomeno, una variabilità sistematica, una tendenza che potrebbe essere spiegata da una variabile ausiliaria se opportunamente correlata. Il kriging con deriva esterna è il metodo che consente la stima in tale condizione: opera infatti la ricostruzione della variabile di interesse facendo intervenire la variabile ausiliaria nota su tutto il campo, che fornisce l'andamento generale della variabile di studio, difficilmente ricostruibile, data la non stazionarietà, con le sole misure della variabile diretta. Semplificando la stima con krigaggio con deriva esterna può essere considerata come quella vista per il caso stazionario, con l'accortezza di considerare la variabile come somma di due componenti: una deriva, con legge nota su tutto il campo in funzione dei valori della variabile ausiliaria, e una fluttuazione dei valori della variabile rispetto alla deriva, rappresentabile da una variabile aleatoria stazionaria.

In ambiente ILWIS non è consentita l'analisi in condizioni non stazionarie, tuttavia, alla luce della suddetta semplificazione è stato comunque possibile ottenere tale stima mediante Krigaggio con deriva esterna articolata nei seguenti 3 passi elementari:

- 1) Determinazione del trend o della deriva, ottenuta mediante una regressione lineare sui dati della concentrazione di benzene Z , noti dal database, in funzione delle variabili ausiliarie ($X, X_1, X_2 \dots$ ecc.) note su tutto il campo:

$$Z^* = a + b X + c X_1 + d X_2$$

2) Calcolo e successiva stima con Krigaggio Ordinario delle fluttuazioni definite come segue:

$$Y = Z - Z^*$$

3) Somma del risultato del Krigaggio Ordinario delle fluttuazioni alla deriva nota su tutto il campo, che restituisce la stima della variabile studio cercata.

La regressione lineare, che ha consentito di individuare la deriva, è stata svolta in ambiente ILWIS a partire dalla tabella associata alle *point map* del benzene per i diversi anni, sfruttando le operazioni da colonna *least squares fit* che a partire dai valori discreti contenuti in due colonne, (dati e variabile ausiliaria), calcola e fornisce in una colonna dedicata i valori interpolati, in accordo con una legge che si presume spieghi la relazione tra le variabili delle colonne in ingresso, studio e ausiliaria. Sono attuabili diversi tipi di interpolazione mediante regressione, dalla polinomiale a uno o più termini, alla logaritmica, al modello di potenza o all'esponenziale. I coefficienti del modello funzionale rappresentante l'interpolazione sono stabiliti in modo tale da garantire la migliore approssimazione. Una volta noti i valori stimati della variabile studio, risultato dell'operazione di *least squares fit*, è immediato il calcolo delle fluttuazioni mediante sottrazione valore vero, valore stimato, a partire dalla barra di comando. A partire dai valori discreti delle fluttuazioni in formato colonna, si hanno tutti gli strumenti in ambiente ILWIS per poterne effettuare uno studio monovariato e stazionario identico metodologicamente a quello visto nello studio monovariato, che si conclude con il krigaggio ordinario. Si ottiene così il primo termine che contribuisce alla somma della stima con deriva esterna, in formato di una mappa raster. Il secondo termine, rappresentato dal trend o deriva, avendo a disposizione le mappe delle variabili ausiliarie e i coefficienti e la relazione funzionale forniti

dall'operazione di *least squares fit*, si calcola direttamente a partire dalla barra di comando e sarà anch'esso nel formato di mappa raster.

Nello specifico dei dati a disposizioni ho deciso di utilizzare in una stima con deriva esterna che ha coinvolto i tre anni in esame e il vento come variabile ausiliaria dal momento che ha evidenziato le maggiori correlazioni con il benzene, correlazioni negative. Ho poi effettuato un ulteriore krigaggio con deriva esterna solo per il 2005, anno maggiormente campionato, sfruttando le 4 variabili più correlate quindi : velocità del vento media mensile a 10 m, giorni piovosi medi mensili, distanza dalle strade principali, uso del suolo 2, in una regressione lineare multipla.

Si riportano tabelle e grafici riassuntivi delle derivate esterne utilizzate nei diversi tentativi e per diversi anni.

	Variabile ausiliaria	Coefficienti di correlazione	Relazione funzionale
2003	wind	- 0,317	$Y = 4.422 + (-0.541 * X)$
2004		- 0,386	$Y = 4.937 + (-0.728 * X)$
2005		- 0,212	$Y = 3.154 + (-0.338 * X)$

Tabella relazioni variabile ausiliaria 'wind'-benzene per anno

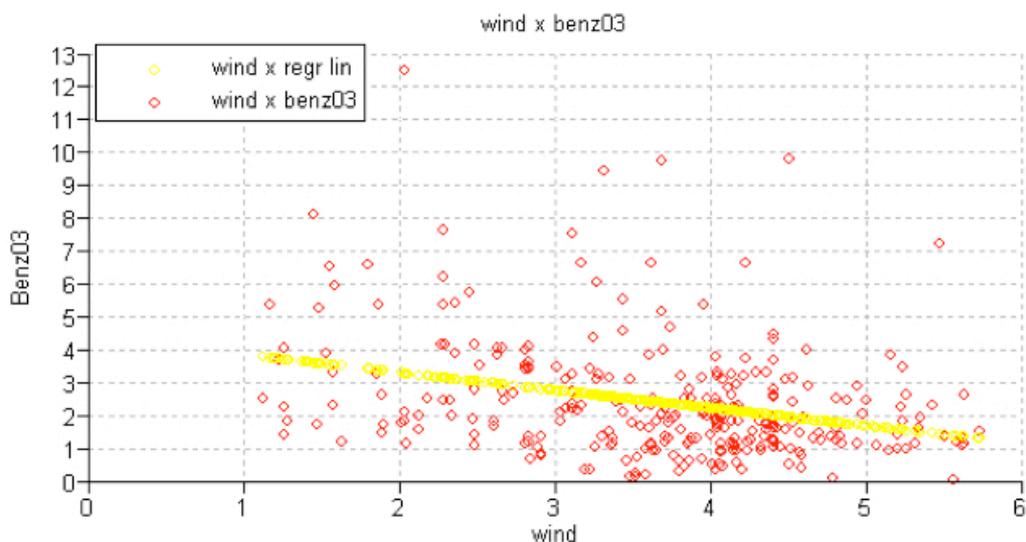


Grafico della regressione lineare con la variabile ausiliaria 'wind' per il 2003

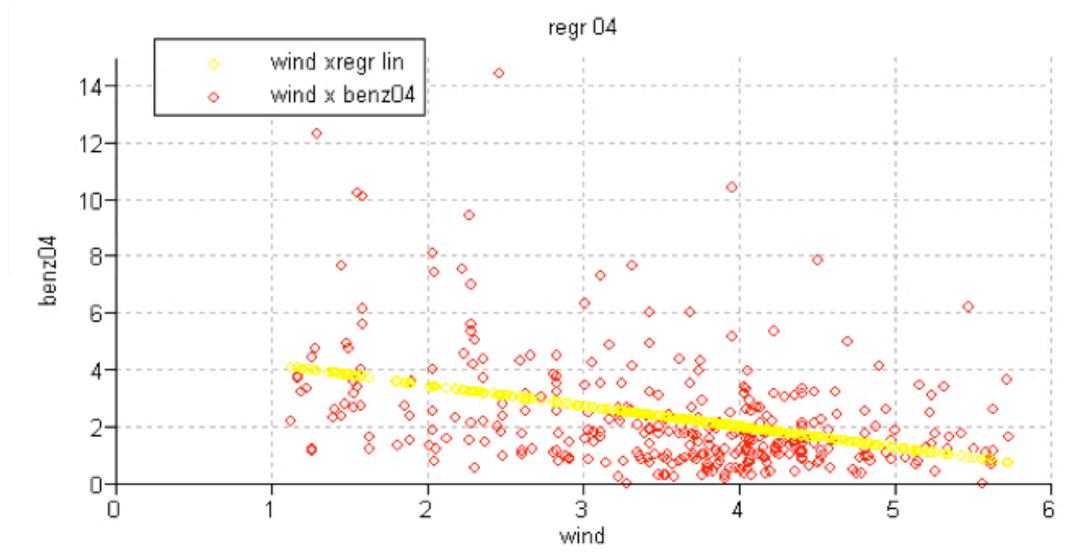


Grafico della regressione lineare con la variabile ausiliaria 'wind' per il 2004

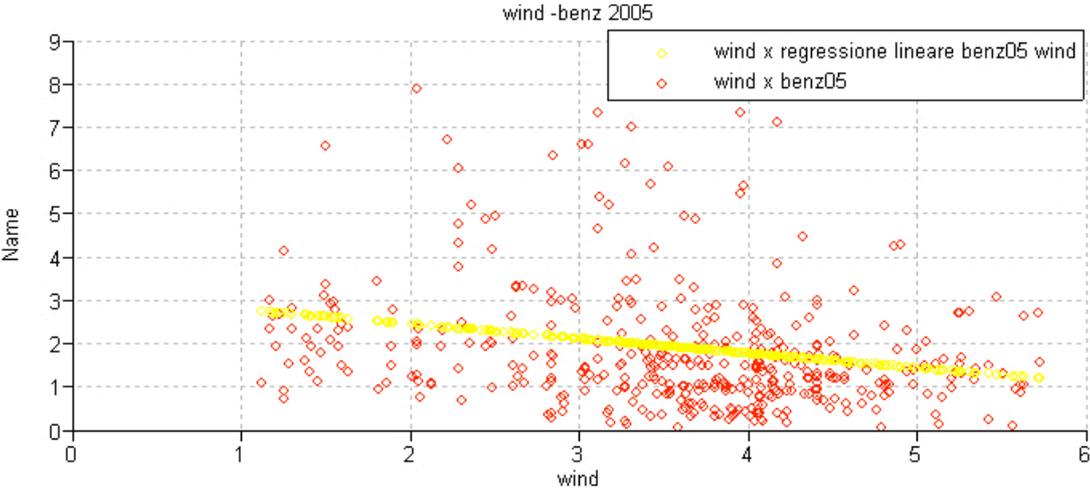


Grafico della regressione lineare con la variabile ausiliaria 'wind' per il 2005

	Variabile ausiliaria	Coefficienti di correlazione	Relazione funzionale
2005	wind	-0.142	3.611206577-
	rd0	-0.212	-0.26028*wind -0.07555*rd0
	roadDistance	-0.119	-0.0000720961*roadDistance
	clc02	0.129	+clc02*0.007139

Tabella relazioni variabili ausiliarie-benzene per il secondo tentativo di KDE

5.4.1. Variogramma delle fluttuazioni

Come esposto, la stima con deriva esterna di base sull'analisi strutturale monovariata stazionaria delle fluttuazioni, svolta negli stessi passaggi logici e operativi visti per lo studio monovariato. I variogrammi sperimentali calcolati e i rispettivi modelli vengono di seguito riportati nella forma di grafici e tabelle riassuntive delle scelte e dei parametri adottati. Si sottolinea che è stata riscontrata la stessa anisotropia geometrica a 135° vista per la variabile studio, pertanto ho proceduto al calcolo dei variogrammi anisotropi, sempre a passo 10000m, che ha mostrato un andamento del tutto simile a quello visto per la variabile studio relativamente allo studio monovariato, seppure con una soglia di variabilità leggermente inferiore, tale scarto tra le soglie di variabilità è infatti assorbito dalla deriva esterna. Si ricorda a tal proposito che i valori dei sill per i variogrammi anisotropi nel caso monovariato, sono di 2.78, 3.6, 2.370 µg/m³, rispettivamente per il 2003, 2004, 2005.

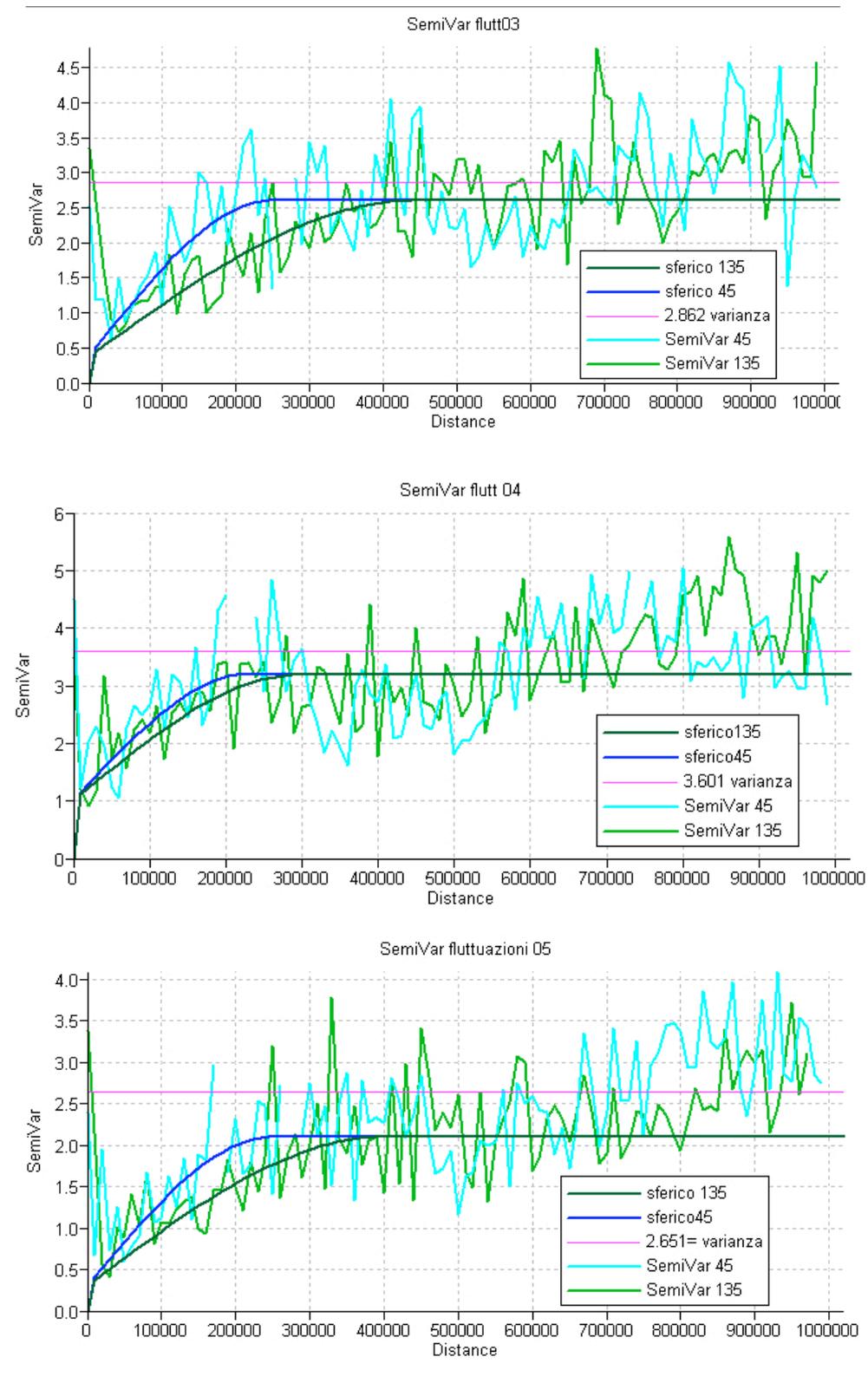
; la seguente tabella riassume i parametri di modellizzazione per i variogrammi delle fluttuazioni riportati subito dopo parallelamente ai modelli.

ANNO	Semivariogramma Fluttuazioni	Modello	Nugget	Sill µg/m ³	Range (m)
2003	Dir. anisotropia	Sferico	0.37	2.60	440000
	Dir.perpendicolare	Sferico	0.37	2.60	250000
2004	Dir. anisotropia	Sferico	1.00	3.20	300000
	Dir.perpendicolare	Sferico	1.00	3.20	230000
2005	Dir. anisotropia	Sferico	0.30	2.10	400000
	Dir.perpendicolare	Sferico	0.30	2.10	250000

Tabella riassuntiva: modello e parametri variogrammi fluttuazioni

Si ricorda che i valori del *sill* nel caso monovariato sono di 2.78, 3.6, 2.370, rispettivamente per il 2003, 2004, 2005.

Semivariogrammi anisotropi e modelli per le fluttuazioni nei tre anni



Si sottolinea che quest'ultimo grafico di semivariogramma e rispettivo modello, per il primo tentativo di deriva esterna del 2005 coincide con quello del secondo tentativo di deriva esterna, che pertanto essendo identico non viene riportato.

5.4.2. Cartografia mediante krigaggio con DE

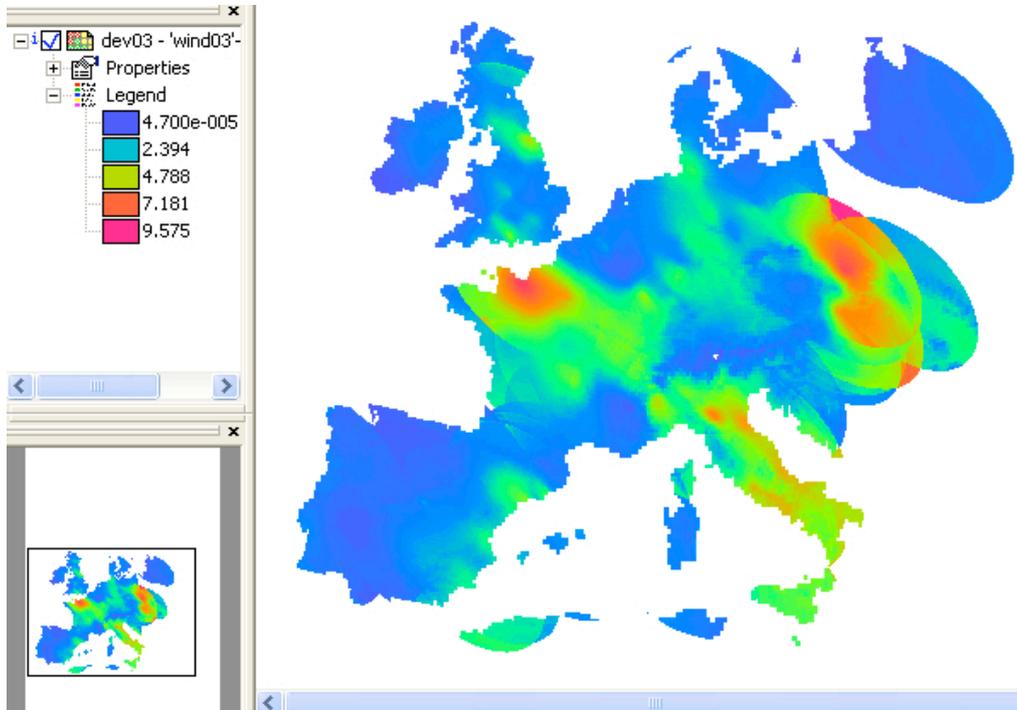
Modellizzato opportunamente il variogramma delle fluttuazioni, ho poi provveduto alla stima delle fluttuazioni con il Krigaggio Ordinario Anisotropo, alla luce della sopra menzionata anisotropia riscontrata a 135°; I parametri di stima delle fluttuazioni sono riassunti nella seguente tabella.

	Direzione anisotropia	Rapporto range	Vicinaggio
2003	135	1.76	400000 (min=1,max=100)
2004	135	1.3	400000 (min=1,max=100)
2005	135	1.6	400000 (min=1,max=100)

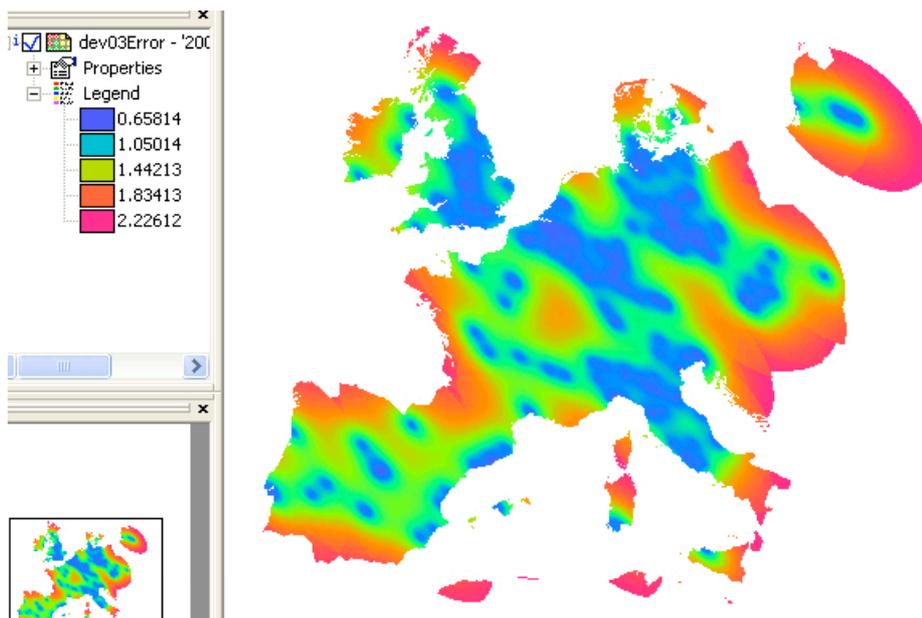
Tabella riassuntiva: parametri di stima delle fluttuazioni con KO anisotropo

Le mappe ottenute sono poi state sommate alla deriva, che, avendo a disposizione le mappe delle variabili ausiliarie e i coefficienti della relazione funzionale tra queste e il benzene, si calcola direttamente a partire dalla barra di comando e sarà anch'esso nel formato di mappa raster.

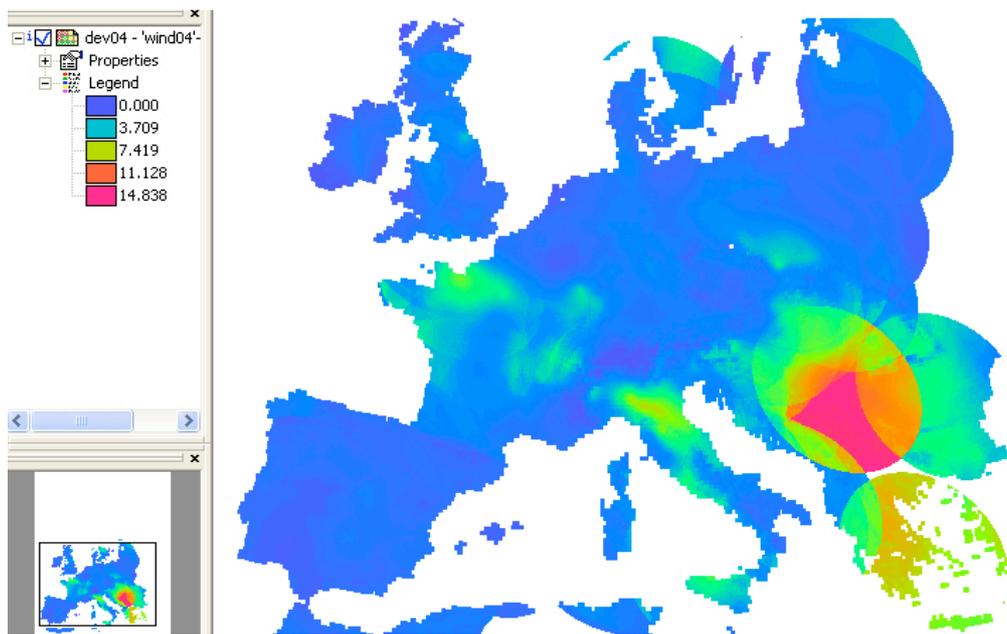
I risultati ottenuti dalla somma, quindi della stima con Krigaggio Ordinario con deriva esterna sono di seguito riportati. Si sottolinea che le mappe delle varianze di stima sono coincidenti con quelle ottenute dal Krigaggio Ordinario anisotropo nella stima delle fluttuazioni.



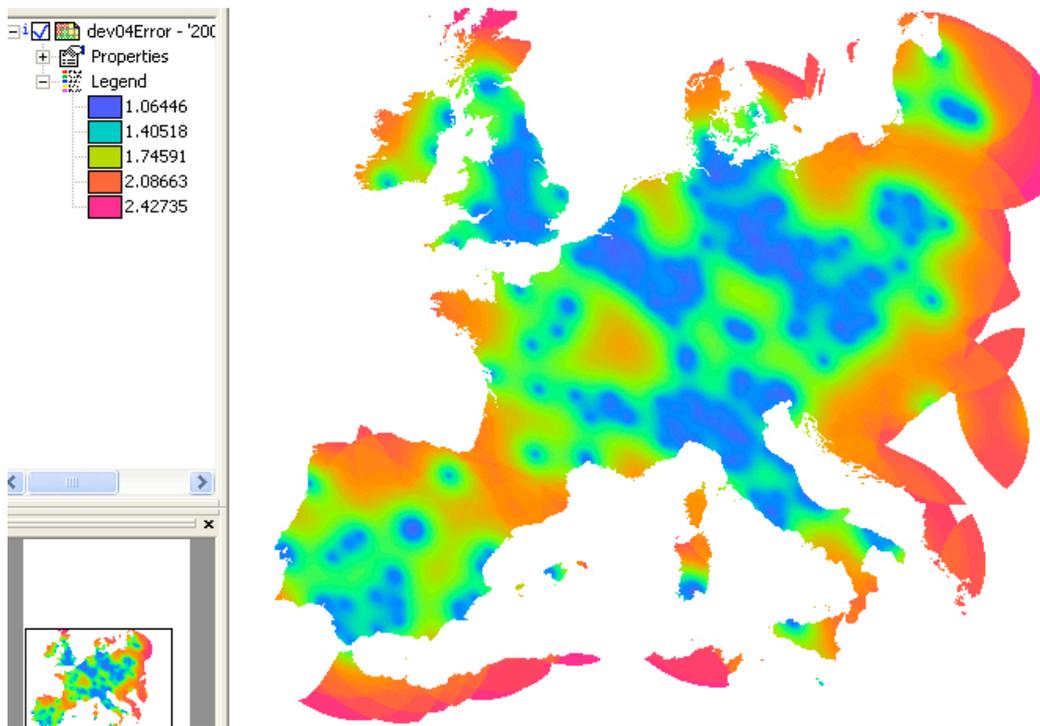
Krigaggio con DE 2003



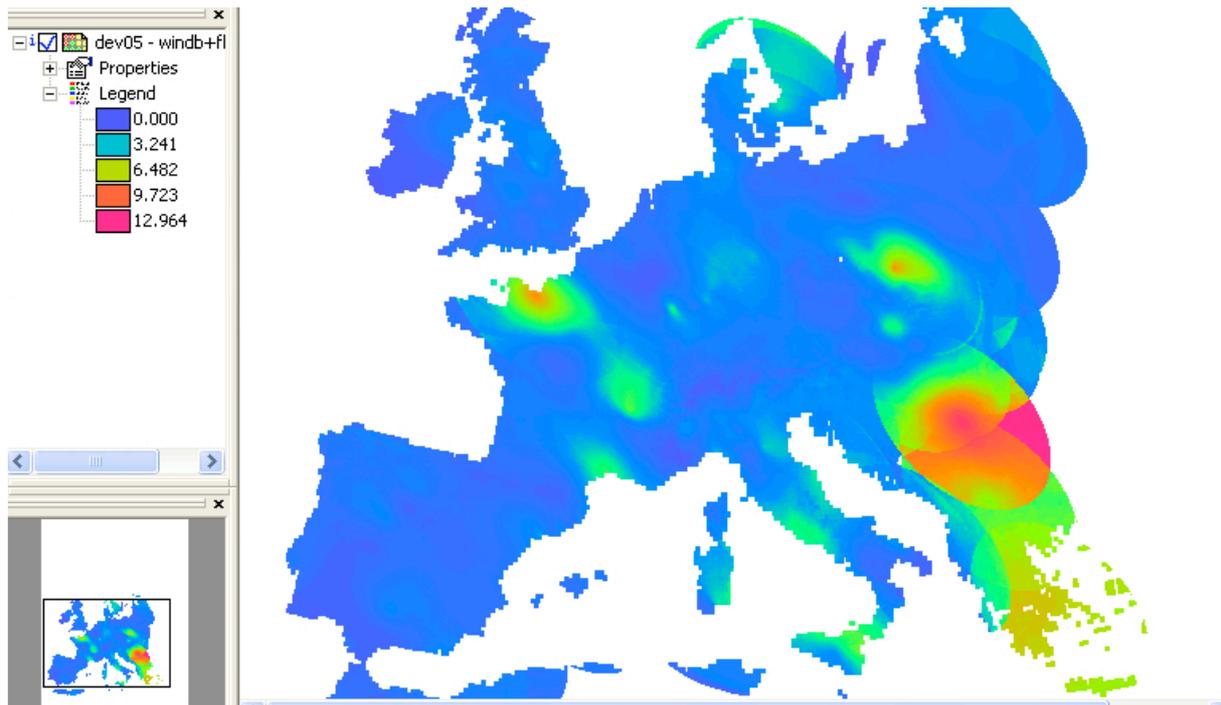
Varianza di stima Krigaggio con DE 2003



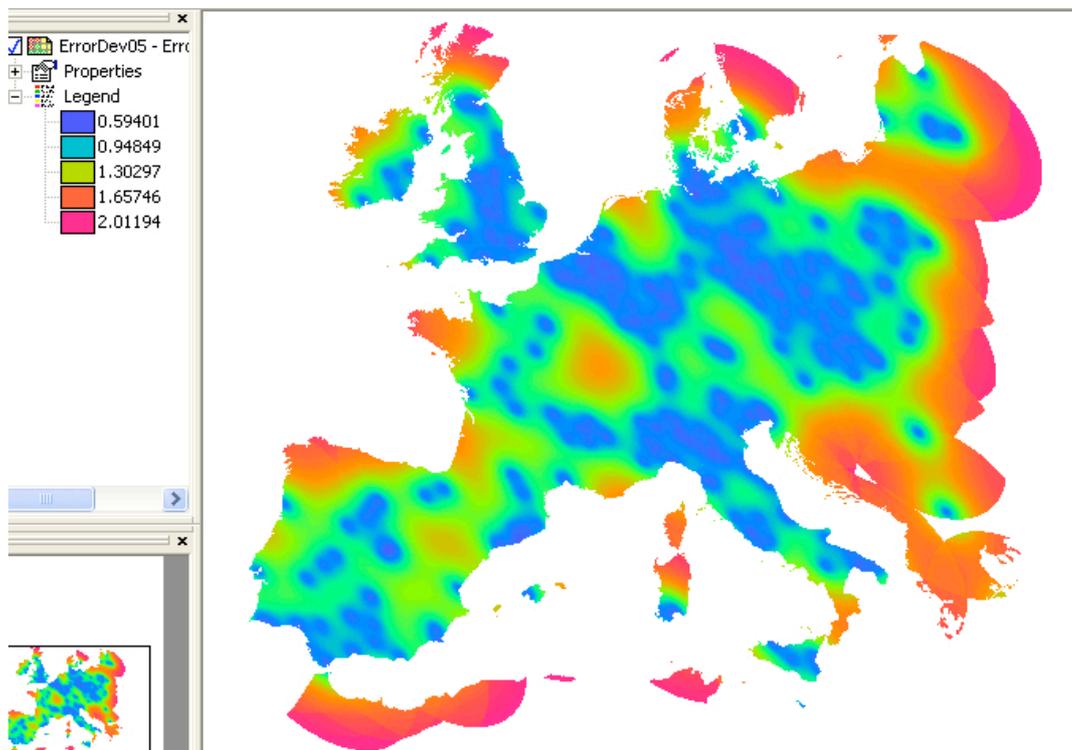
Krigaggio con DE 2004



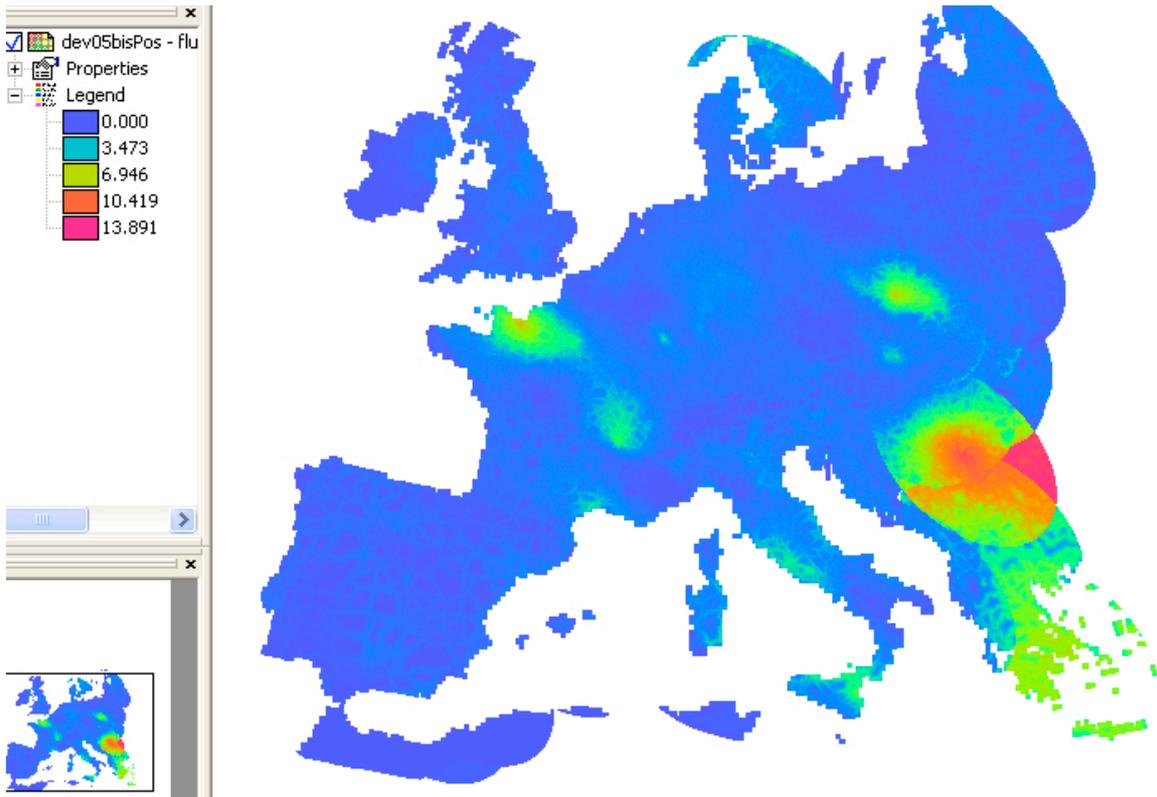
Varianza di stima Krigaggio con DE 2004



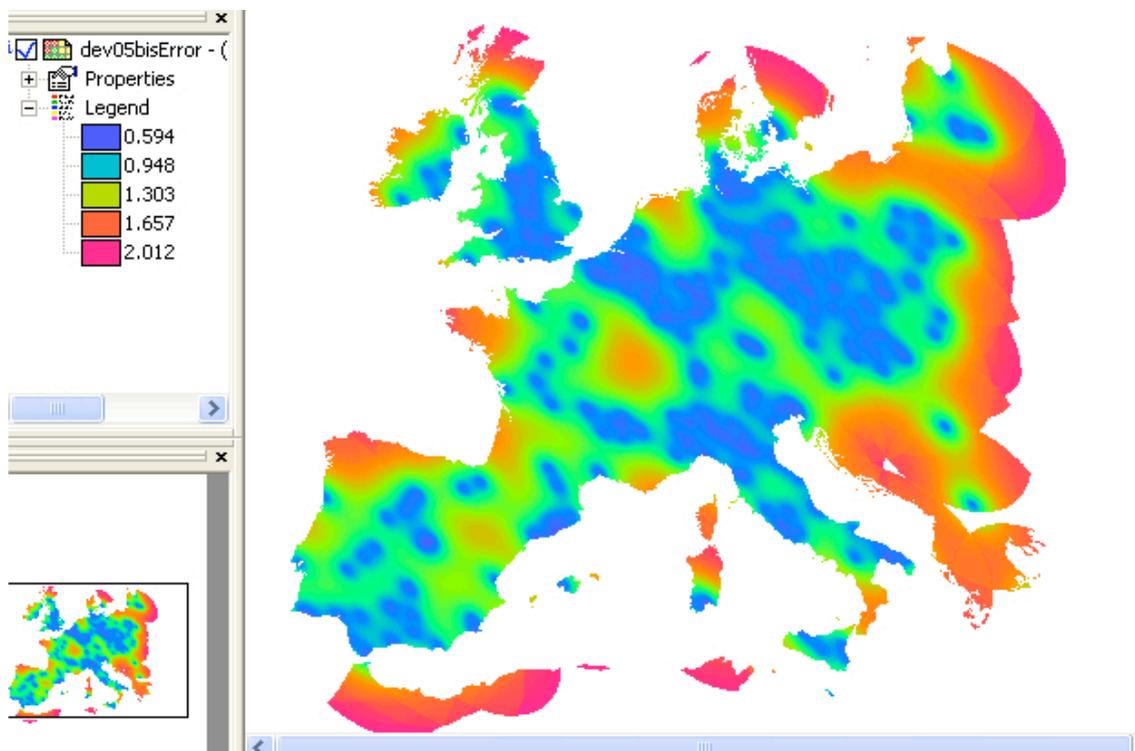
Krigaggio con DE 2005



Varianza di stima Krigaggio con DE 2005



Secondo tentativo di Krigaggio con DE 2005



Varianza di stima secondo tentativo di Krigaggio con DE 2005

6. STUDIO A SCALA LOCALE

6.1. FINALITÀ

L'ultima fase in cui si è articolato il caso studio è stata un'analisi della variabilità spaziale a piccola scala, relativamente a 5 aree omogenee coincidenti con le nazioni più campionate, Italia, Francia, Spagna, Germania, Gran Bretagna.

Gli studi monovariato, temporale e multivariato svolti finora, hanno esaminato la correlazione spaziale della concentrazione di benzene a livello europeo, mettendone in luce la variabilità in funzione del tempo e di fattori esterni che possono pregiudicarne la presenza e la dispersione, senza apportare però elementi conoscitivi sull'eventuale presenza di diverse scale di variabilità spaziale, vagliata all'interno della presente analisi che va così a completare lo studio fornendo importanti elementi di riflessione; un approccio complessivo europeo, necessario per la valutazione del trasporto di lungo raggio che molti inquinanti possono avere ma anche per le misure economiche e politiche prese a livello comunitario, va comunque affiancato da un'analisi locale; in tal modo si possono valutare le differenze e soprattutto se sia il caso di ottenere il trend globale come somma di studi locali con opportune modifiche per le zone di confine.

6.2. DEFINIZIONE DELLE AREE OMOGENEE E STRUTTURA DELLO STUDIO

Alla luce di quanto emerso contestualmente allo studio dei variogrammi diretti della variabile nello studio monovariato, il forte disturbo all'origine dei variogrammi, trascurato nel caso di un modello globale, è indice di un disturbo della variabilità a piccola scala che è stata così presa in esame mediante un approccio per aree omogenee. Il primo passo è stato quello di delineare 5 aree omogenee, e quindi 5 sottoinsiemi del database principale. Le aree sono state scelte coincidenti con i confini nazionali delle 5 nazioni più

campionate. Italia, Francia, Spagna, Germania, Gran Bretagna, e per tali aree è stato analizzato l'ultimo anno a disposizione nel database: il 2005.

Individuate le aree, ho provveduto all'analisi della variabilità spaziale monovariata, negli stessi passaggi logici e operativi visti relativamente allo studio monovariato, quindi calcolo delle mappe variografiche, individuazione del passo e della direzione di calcolo dei variogrammi, calcolo e modellizzazione di quest'ultimi, propedeutici alla successiva fase di Krigaggio Ordinario. Non avendo preso in esame per tale analisi più anni e non avendo riscontrato correlazioni consistenti con le variabili ausiliarie, si sono così escluse le stime con cokrigaggio e krigaggio con deriva esterna svolte sul suolo europeo.

Tra le operazioni preliminari allo studio a scala locale, oltre alle formattazioni dei file dati e conversioni in *point map* nazionali, indispensabili in ambiente ILWIS alle fasi di caratterizzazione spaziale, si evidenzia la creazione di file dati subordinati a quelli originali, ottenuti aggregando tra loro, per coordinate geografiche e valore di concentrazione di benzene, le stazioni distanti meno di 5000 m. Tale aggregazione, e successivo studio a partire dai dati aggregati, trova giustificazione nel disturbo che emerge nella variabilità a piccola scala, disturbo che si suppone sia dovuto alla disomogeneità delle stazioni e quindi alla presenza di stazioni ravvicinate ai centri urbani; l'aggregazione ha portato a semplificare l'interpretazione della variabilità, regolarizzandola (in maniera variabile a seconda dell'area), e di conseguenza ha reso più agevole individuare dei modelli per i variogrammi, da utilizzare nella stima del totale dei dati a disposizione (comprese le stazioni distanti meno di 5000 m).

Fatte queste premesse sull'assetto generale dell'analisi a scala locale e sulle aree studio, si riportano di seguito divisi per aree, l'analisi strutturale e la cartografia. Si precisa che verranno omesse le informazioni relative alle mappe variografiche, dal momento che non si riscontrano anisotropie ad eccezione dell'Italia, dove è dovuta alla conformazione geografica.

Di seguito si riporta una tabella che riassume i risultati delle analisi preliminari allo studio dei variogrammi, per le diverse aree e per ogni area, per la totalità dei dati a disposizione e per i dati aggregati.

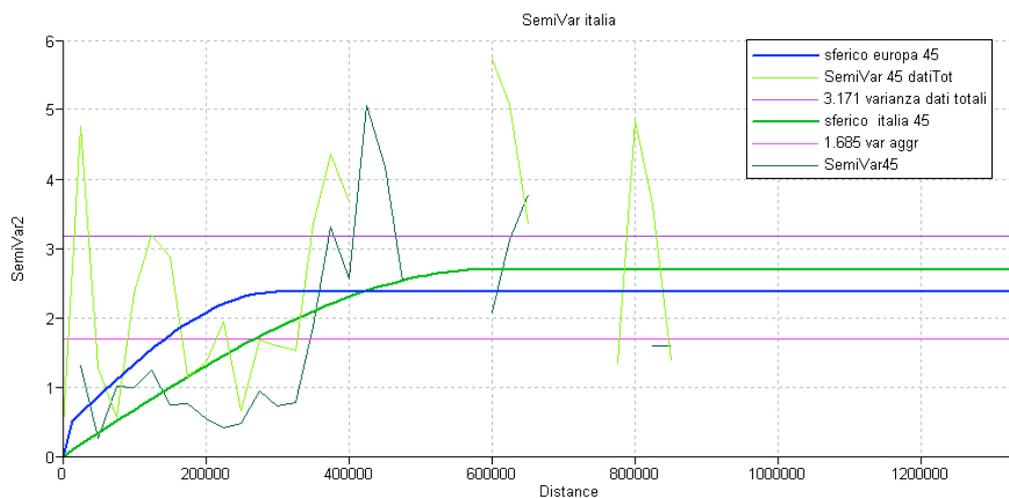
<i>Statistiche di base</i>	min	max	media	dev stand	num campioni
italia	0,079	7,917	2,543	1,781	89
italia aggr	0,079	5,512	2,307	1,298	63
spagna	0,175	6,207	1,359	1,121	64
spagna aggr	0,175	3,714	1,196	0,833	52
francia	0,418	9,507	2,501	2,462	32
francia aggr	0,418	9,507	2,337	2,179	28
granbretagna	0,416	4,514	1,554	0,82	41
granbretagna aggr	0,416	4,514	1,547	0,805	37
germania	0,494	4,202	1,842	0,778	74
germania aggr	0,494	4,202	1,872	0,758	59

Tabella riassuntiva delle statistiche di base e dell'aggregazione dei dati vicini

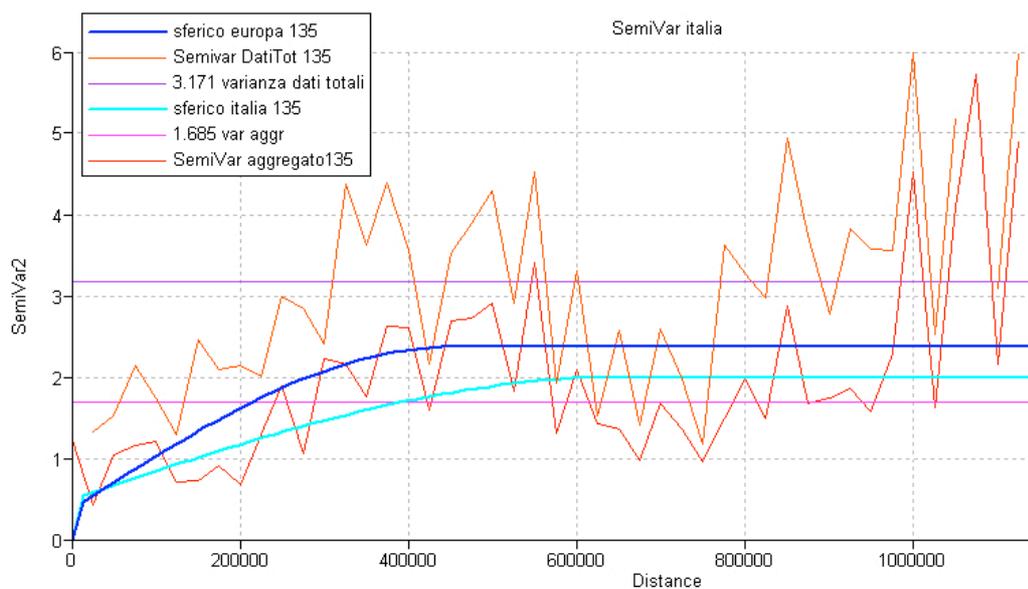
6.2.1. Italia

6.2.1.1. Analisi strutturale

I variogrammi e le mappe variografiche sono stati calcolati a passo 25000 m, e hanno messo una luce, come accennato contestualmente alla struttura del lavoro, una netta anisotropia a 135°, che risulta dall'estensione geografica e della quale si è tenuto conto nell'operazione si *Spatial Correlation*. I variogrammi sono stati calcolati dopo una verifica di eventuali anisotropie a un passo di 25000m; la tabella seguente riassume quelli che sono stati i parametri di modellizzazione dei variogrammi sperimentali a piccola scala, si precisa che il modello è stato individuato a partire dai file dati aggregati così come definiti precedentemente e per confronto si riportano anche i parametri di modellizzazione del semivariogramma sperimentale europeo.



Semivariogramma Italia 45°



Semivariogramma Italia 135°

Si nota chiaramente come il semivariogramma a 45° non vada oltre i 400000, 500000 m, dovuto chiaramente a limiti di natura geografica; relativamente all'effetto dell'aggregazione dei dati vicini emerge che il variogramma dei dati aggregati ha soglia di variabilità inferiore a quello calcolato a partire dai dati totali. Analizzando i modelli, quelli adottati per l'Italia ricalcano quello europeo, essendo sempre uno sferico più effetto pepita e, mentre a 45° la

variabilità nazionale risulta essere superiore a quella europea, a 135° la soglia italiana risulta invece minore di quella individuata scala comunitaria.

La seguente tabella riassume i parametri e il modello dei semivariogrammi anisotropi calcolati a partire dai dati aggregati relativi all'area dell'Italia, si ribadisce che sono tali parametri ad essere alla base della successiva stima con Krigaggio Ordinario a partire dai dati totali.

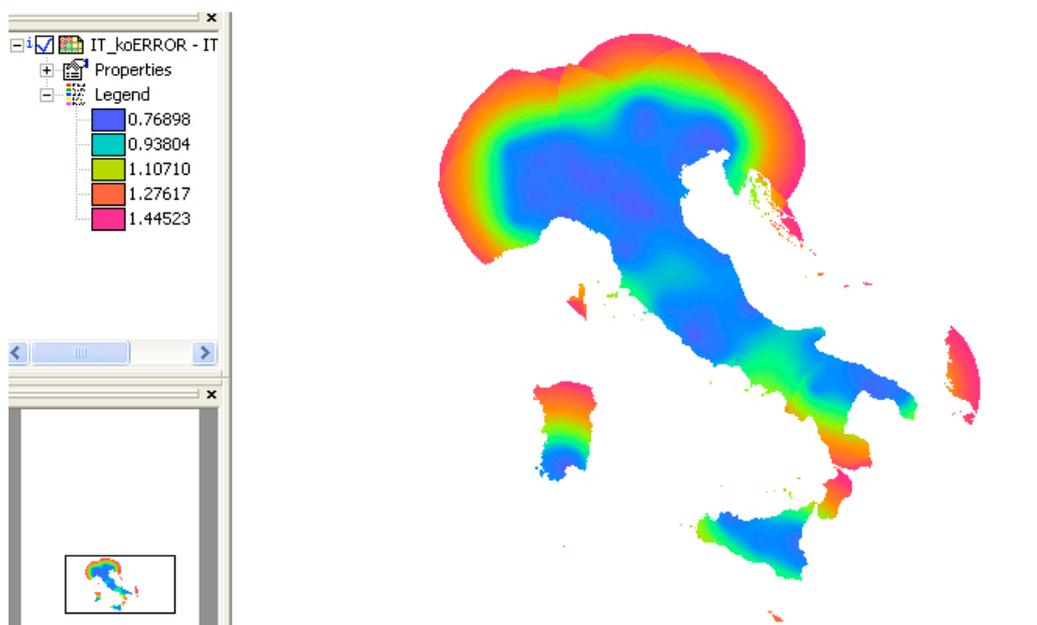
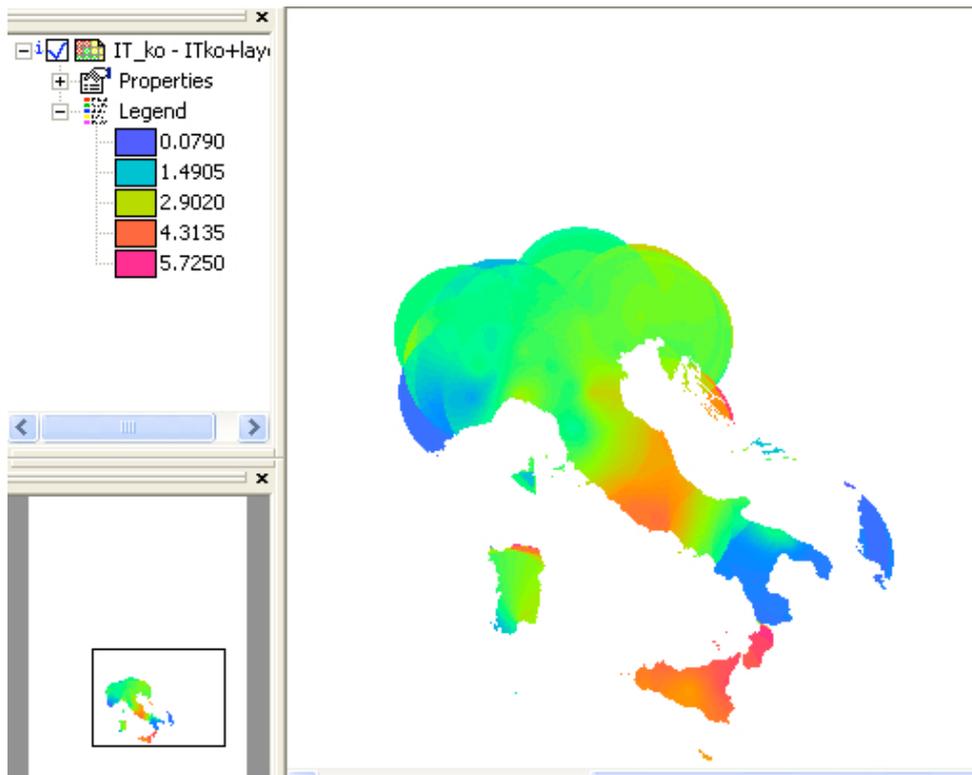
<i>Variogrammi: modello e parametri</i>	modello	nugget	sill	range (m)
Italia 45°	sferico	0	2,7	600000
Italia 135°	sferico	0,5	2	650000

A

6.2.1.2. Cartografia mediante krigaggio Ordinario

A partire dai modelli e parametri individuati e sopra esposti, per i variogrammi a partire dai dati aggregati, il passaggio successivo è stato quello della stima della stima mediante Krigaggio Ordinario anisotropo. I parametri adottati nella stima e le mappe ottenute sono di seguito riportati.

<i>Parametri del krigaggio ordinario</i>		
vicinaggio (m)		200000
num minimo punti		1
num massimo punti		100
<i>Parametri del Krigaggio ordinario anisotropo</i>		
direzione di anisotropia		135°
rapporto range		1,083
vicinaggio (m)		200000
num minimo punti		1
num massimo punti		100

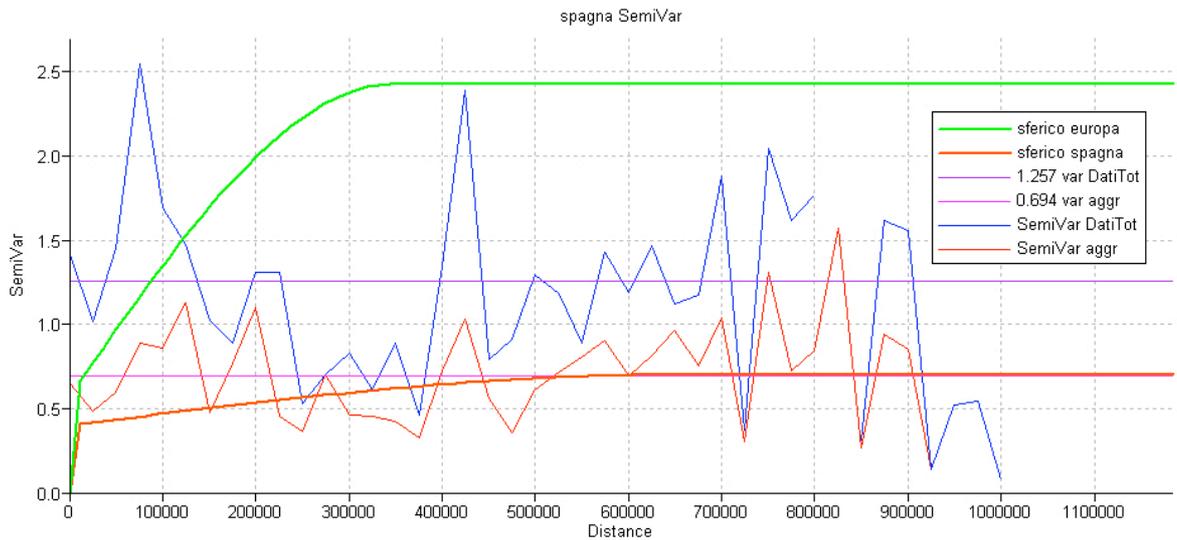


Mappe del KO anisotropo e relativa varianza di stima: Italia

6.2.2. Spagna

6.2.2.1. Analisi strutturale

Il semivariogramma relativo alla Spagna, che è stato preso in esame è quello medio delle 4 direzioni, sempre a passo 25000 m, non essendosi riscontrata alcuna anisotropia per tale area.



Semivariogramma medio Spagna

In analogia con quanto visto per l'Italia, l'aggregazione dei dati distanti meno di 5000 m, contribuisce ad abbassare considerevolmente le soglie. Il modello anche in questo caso ricalca quello europeo, essendo uno sferico più pepita, ma la variabilità per l'area in esame risulta nettamente minore.

La seguente tabella riassume i parametri e il modello del semivariogramma medio calcolato a partire dai dati aggregati relativi all'area della Spagna

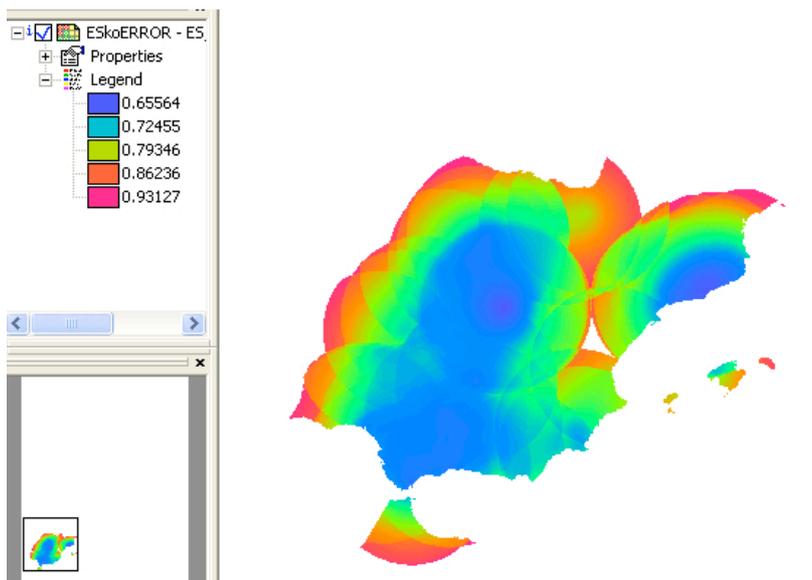
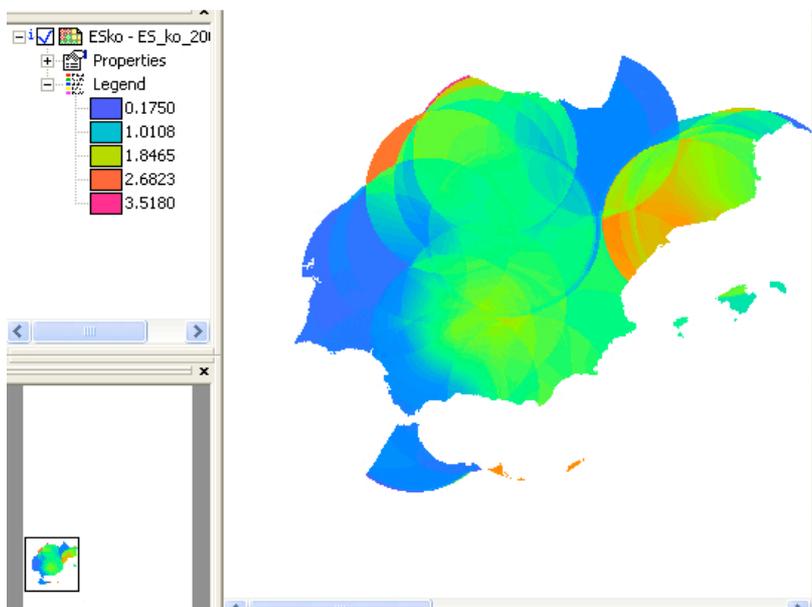
<i>Variogrammi: modello e parametri</i>	modello	nugget	sill	range (m)
Spagna medio.	sferico	0,4	0,7	650000

6.2.2.2. Cartografia mediante Krigaggio Ordinario

A partire dai modelli e parametri individuati e sopra esposti, per i variogrammi a partire dai dati aggregati, il passaggio successivo è stato

quello della stima della stima mediante Krigaggio Ordinario. I parametri adottati nella stima e le mappe ottenute sono di seguito riportati.

<i>Parametri del krigaggio ordinario</i>		
vicinaggio (m)		200000
num minimo punti		1
num massimo punti		100

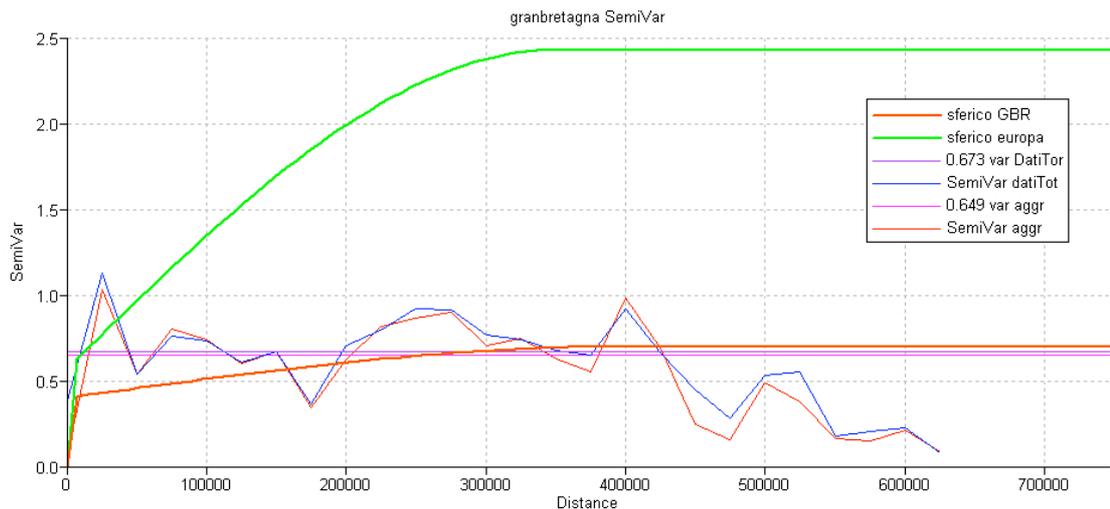


Mappe del KO anisotropo e relativa varianza di stima: Spagna

6.2.3. Gran Bretagna

6.2.3.1. Analisi strutturale

Il semivariogramma relativo alla Gran Bretagna, che è stato preso in esame è quello medio delle 4 direzioni, sempre a passo 25000 m, non essendosi riscontrata alcuna anisotropia per tale area.



Semivariogramma medio Gran Bretagna

Anche per la Gran Bretagna il variogramma ha un andamento individuato dalle stesse strutture di quello europeo, la soglia del variogramma europeo è però nettamente superiore alla variabilità locale. In questo caso si nota come l'aggregazione non contribuisca a regolarizzare l'andamento del variogramma, nè a eliminare il disturbo all'origine.

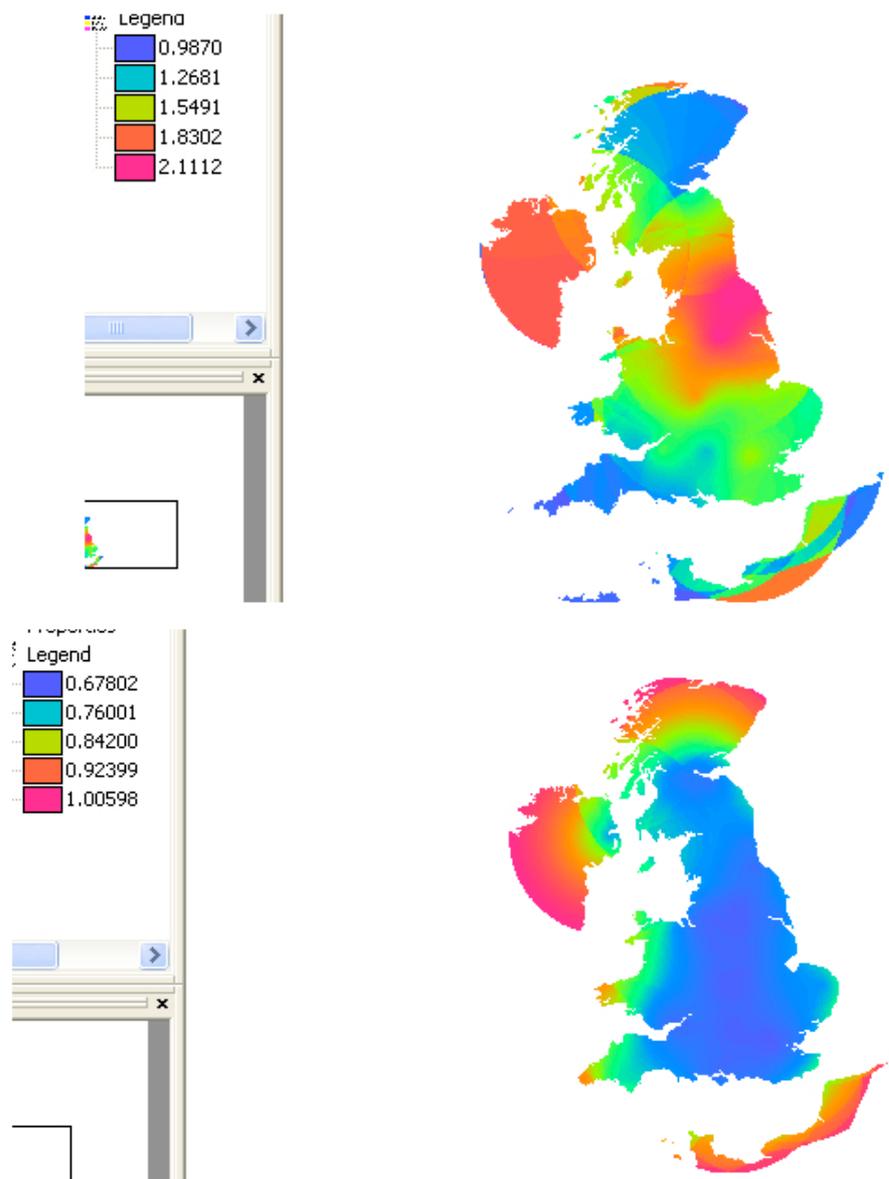
La seguente tabella riassume i parametri e il modello del semivariogramma medio calcolato a partire dai dati aggregati relativi all'area della Gran Bretagna.

<i>Variogrammi: modello e parametri</i>	modello	nugget	sill	range (m)
Granbretagna medio	sferico	0,4	0,7	400000

6.2.3.2. Cartografia mediante Krigaggio Ordinario

A partire dai modelli e parametri individuati e sopra esposti, per i variogrammi a partire dai dati aggregati, il passaggio successivo è stato quello della stima della stima mediante Krigaggio Ordinario. I parametri adottati nella stima e le mappe ottenute sono di seguito riportati.

<i>Parametri del krigaggio ordinario</i>		
vicinaggio (m)		200000
num minimo punti		1
num massimo punti		100

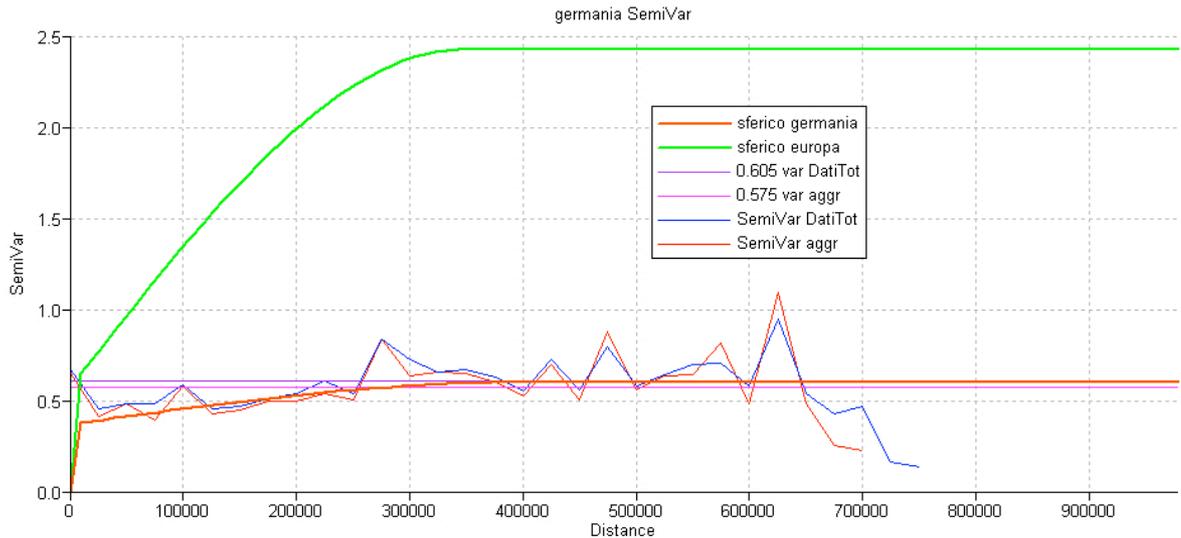


Mappe del KO anisotropo e relativa varianza di stima: Gran Bretagna

6.2.4. Germania

6.2.4.1. Analisi strutturale

Il semivariogramma relativo alla Germania che è stato preso in esame è quello medio delle 4 direzioni, sempre a passo 25000 m, non essendosi riscontrata alcuna anisotropia per tale area.



Semivariogramma medio Germania

Per la Germania si può ripetere quanto appena considerato per la Gran Bretagna: il modello di variogramma è in linea con quello europeo ma soglia di variabilità nettamente minore; anche in questo caso l'aggregazione non contribuisce in modo significativo nell'abbassare la variabilità del variogramma.

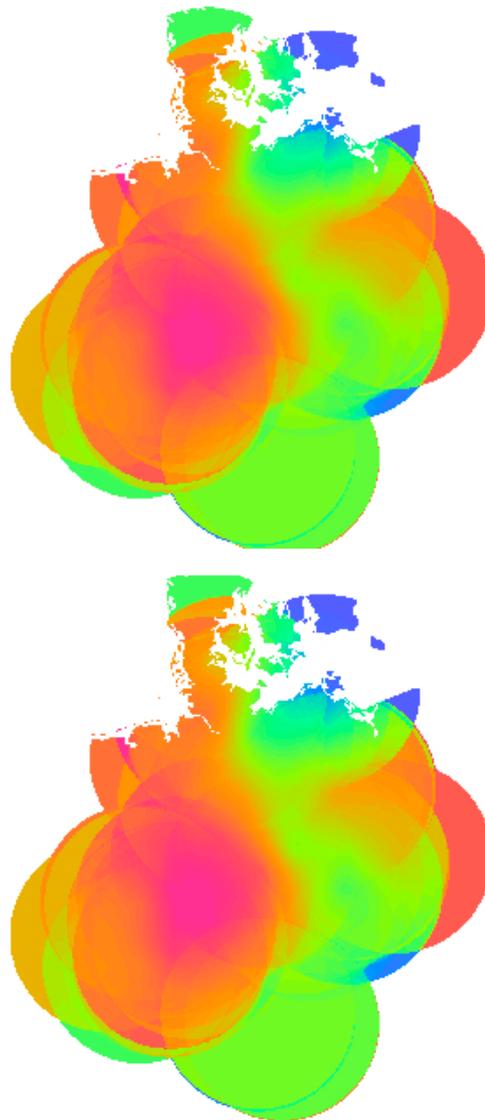
La seguente tabella riassume i parametri e il modello del semivariogramma medio calcolato a partire dai dati aggregati relativi all'area della Germania.

<i>Variogrammi: modello e parametri</i>	modello	nugget	sill	range (m)
Germania medio	sferico	0,37	0,6	400000

6.2.4.2. Cartografia mediante Krigaggio Ordinario

A partire dai modelli e parametri individuati e sopra esposti, per i variogrammi a partire dai dati aggregati, il passaggio successivo è stato quello della stima della stima mediante Krigaggio Ordinario. I parametri adottati nella stima e le mappe ottenute sono di seguito riportati.

<i>Parametri del krigaggio ordinario</i>		
vicinaggio (m)		200000
num minimo punti		1
num massimo punti		100

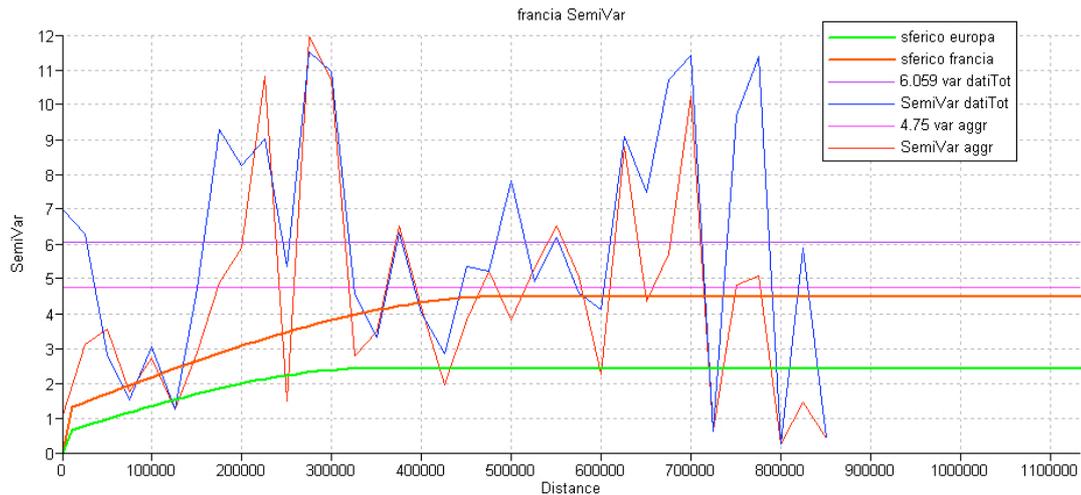


Mappe del KO anisotropo e relativa varianza di stima: Germania

6.2.5. Francia

6.2.5.1. Analisi strutturale

Il semivariogramma relativo alla Francia che è stato preso in esame è quello medio delle 4 direzioni, sempre a passo 25000 m, non essendosi riscontrata alcuna anisotropia per tale area.



Semivariogramma medio Francia

L'area omogenea della Francia è la prima tra quelle esaminate in cui i valori di variabilità locale superano quelli totali, pur avendo sempre lo stesso andamento, individuato dalla somma di un modello sferico e di un pepita. L'aggregazione dei dati che distano meno di 5000 m, anche in questo caso non contribuisce ad una drastica diminuzione delle soglie del variogramma sperimentale, come per Italia e Spagna, ma si nota che i picchi di variabilità a piccolissima scala vengano ridimensionati.

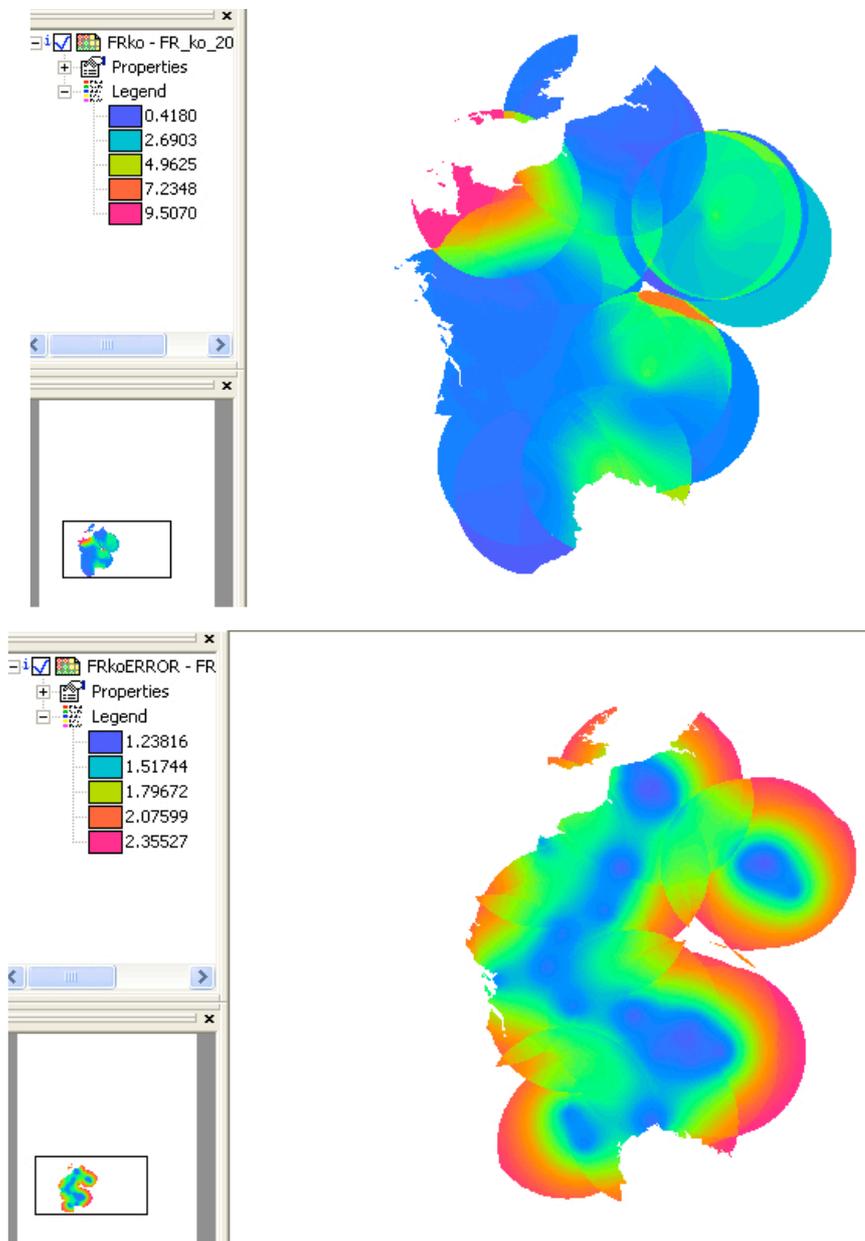
La seguente tabella riassume i parametri e il modello del semivariogramma medio calcolato a partire dai dati aggregati relativi all'area della Germania.

<i>Variogrammi: modello e parametri</i>	modello	nugget	sill	range (m)
Francia medio	sferico	1,2	4,5	500000

6.2.5.2. Cartografia mediante Krigaggio Ordinario

A partire dai modelli e parametri individuati e sopra esposti, per i variogrammi a partire dai dati aggregati, il passaggio successivo è stato quello della stima della stima mediante Krigaggio Ordinario. I parametri adottati nella stima e le mappe ottenute sono di seguito riportati.

<i>Parametri del krigaggio ordinario</i>		
vicinaggio (m)		200000
num minimo punti		1
num massimo punti		100



Mappe del KO anisotropo e relativa varianza di stima: Francia

7. ANALISI DEI RISULTATI

Il passo conclusivo dello studio della presente tesi è stato quello di analizzare i risultati ottenuti, mostrati e commentati in precedenza; l'analisi dei risultati si è articolata in due fasi:

- una validazione dei diversi modelli di variabilità spaziale alla base delle diverse stime, mediante strumenti funzionali a tale operazione, quali il test kriging e calcolo di grandezze specifiche, opportunamente introdotti spiegati nel seguente paragrafo;
- raffronto fra le diverse cartografie, ottenendo mappe delle differenze tra le diverse stime e valutando le statistiche elementari di tali differenze.

Prima di entrare nella trattazione delle due fasi dell'analisi dei risultati si ricapitolano brevemente i vari metodi di stima utilizzati corrispondenti ad altrettanti approcci teorici in cui si è articolato studio, che hanno condotto all'ottenimento delle cartografie:

- Krigaggio Ordinario anisotropo: analisi strutturale, stima e conseguente cartografia, basate su un approccio monovariato e stazionario.
- Cokrigaggio Ordinario: analisi strutturale, stima e conseguente cartografia, basate su un approccio multivariato stazionario, dove la configurazione multivariata è ottenuta dal considerare la stessa variabile in anni precedenti o successivi a quello in esame, come variabile ausiliaria; approccio finalizzato ad inquadrare la correlazione temporale.
- Krigaggio con deriva esterna: analisi strutturale, stima e conseguente cartografia basate su un approccio multivariato, non stazionario; l'approccio multivariato è giustificato dallo studio delle correlazioni con materiale informativo geografico, dal quale si sono estratte le variabili maggiormente correlate. Sono state sviluppate due analisi che si differenziano in base al numero di variabili ausiliarie considerate che in un caso coincidono solo con il vento e nell'altro con le 4 più correlate, velocità media del vento a 10 m, giorni piovosi medi mensili, uso del suolo urbano discontinuo, distanza dalle strade principali.

- Krigaggio Ordinario per aree omogenee: analisi strutturale, stima e conseguente cartografia relativa a 5 aree omogenee considerate, alla luce di un approccio a piccola scala.

7.1. Analisi dei diversi modelli di variabilità spaziale

L'analisi dei diversi modelli di variabilità spaziale alla base delle stime effettuate, risulta indispensabile al fine di poter valutare quale tra i modelli e quindi tra gli approcci di stima adottati, porta ad un valore stimato della concentrazione di benzene più simile a quello reale.

L'analisi si è basata fondamentalmente sulla tecnica del test Kriging, che consiste nello stimare la variabile con il modello adottato, in punti in cui essa è nota, in modo da poter quantificare in tali punti gli scarti tra valori veri e valori stimati; la stima parte così da un set dati, sottoinsieme di quello iniziale, dal quale si escludono le informazioni che si andranno a stimare e che consentiranno il test: la valutazione della differenza tra veri e stimati.

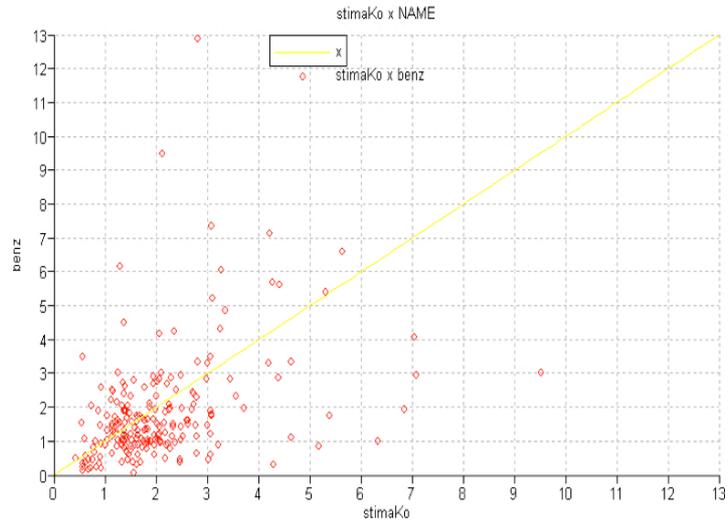
Dal momento che in ambiente ILWIS non è contemplata tale operazione, è stata sviluppata con una serie di passaggi elementari. Operativamente, ho quindi provveduto a dividere il set dati del 2005 in due set dati omogenei dal punto di vista delle concentrazioni di benzene e uniformemente distribuiti sul territorio. A partire da un set, ho effettuato le stime con i diversi modelli individuati e presentati in precedenza contestualmente alle varie fasi dello studio, quindi rispettivamente per il Krigaggio Ordinario anisotropo, Cokrigaggio, Krigaggio con deriva esterna 1 (1° tentativo, variabile ausiliaria vento), Krigaggio con deriva esterna 2 (2° tentativo, regressione polinomiale con più variabili ausiliarie).

Dalle mappe stimate dal primo set ho estratto, con la già citata operazione di *mapvalue*¹, i valori in corrispondenza delle stazioni del secondo set di dati che

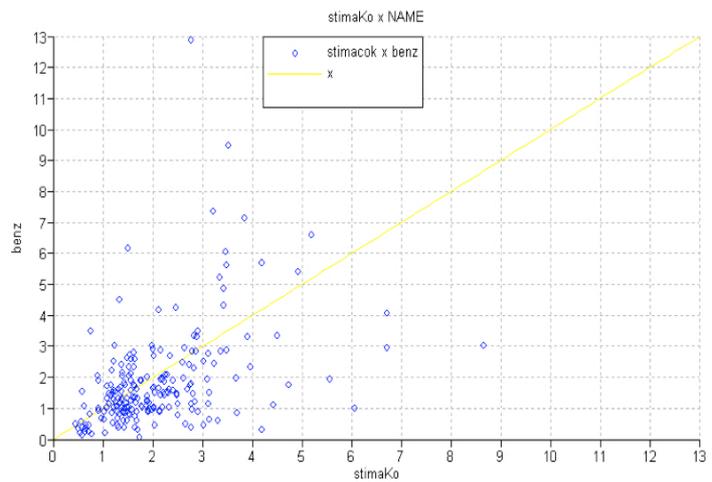
¹ L'operazione di *MapValue* è stata presentata nello studio multivariato (cap 5) contestualmente all'ottenimento delle nuvole di correlazione.

non ha contribuito alla stima e dove la variabile è nota. In tal modo avendo a disposizione per il secondo set di dati, i valori veri e gli stimati, ho provveduto a creare dei grafici cartesiani con in ascissa i valori stimati e in ordinata i veri.

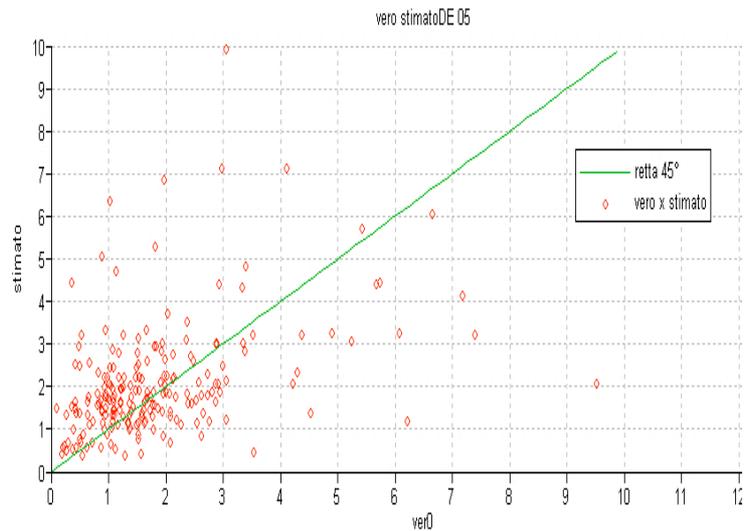
I risultati ottenuti vengono di seguito riportati



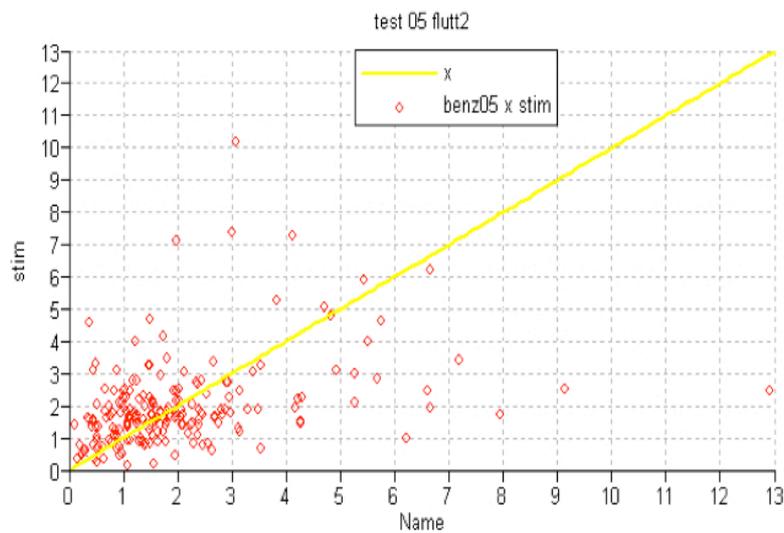
Test Kriging KO anisotropo '05



Test CoKriging '04-'05



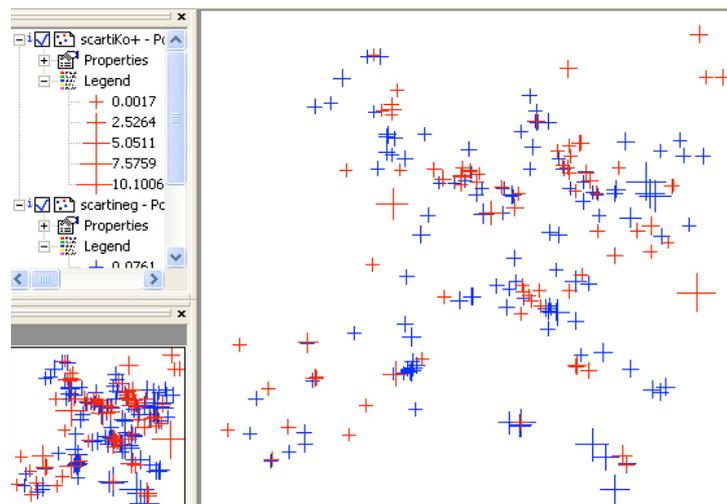
Test Kriging con deriva esterna 1°tentativo '05.



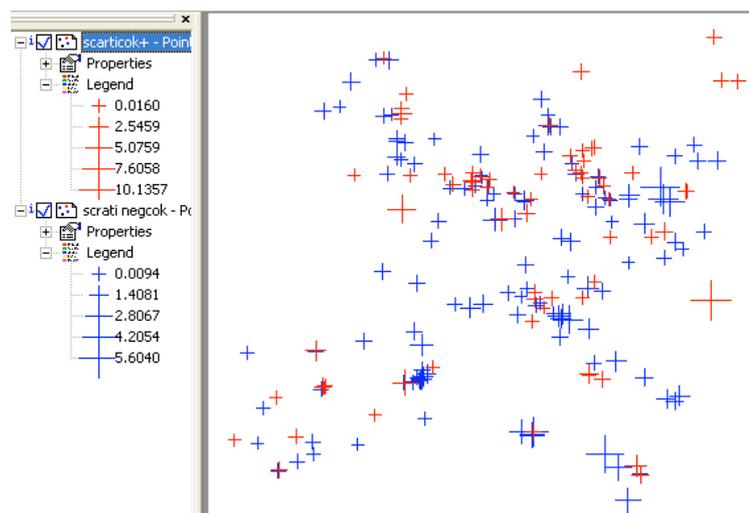
Test Kriging con deriva esterna 1°tentativo

Le nuvole di correlazione hanno delineato un comportamento simile per i diversi approcci metodologici e rispettivi modelli, dopo un'iniziale approssimazione della retta a 45°, si discostano da tale andamento, che significherebbe la perfetta coincidenza tra valore vero e stimato. Si nota come le differenze più consistenti si hanno in corrispondenza dei valori più alti assunti dalla variabile.

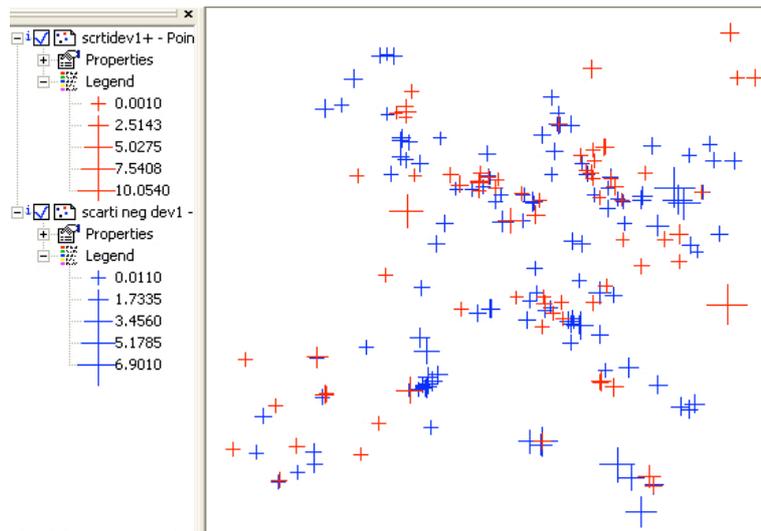
Le nuvole di correlazione così ottenute, sono utili per avere una panoramica globale della relazione tra veri e stimati, quindi ad una valutazione qualitativa che deve essere completata dalla collocazione spaziale degli errori; a partire dai valori graficati, veri e stimati, ho provveduto quindi a calcolare gli scarti e a visualizzarli spazialmente in *point map*, dividendo gli scarti negativi (in blu), dagli scarti positivi (in rosso). Le mappe ottenute sono di seguito riportate.



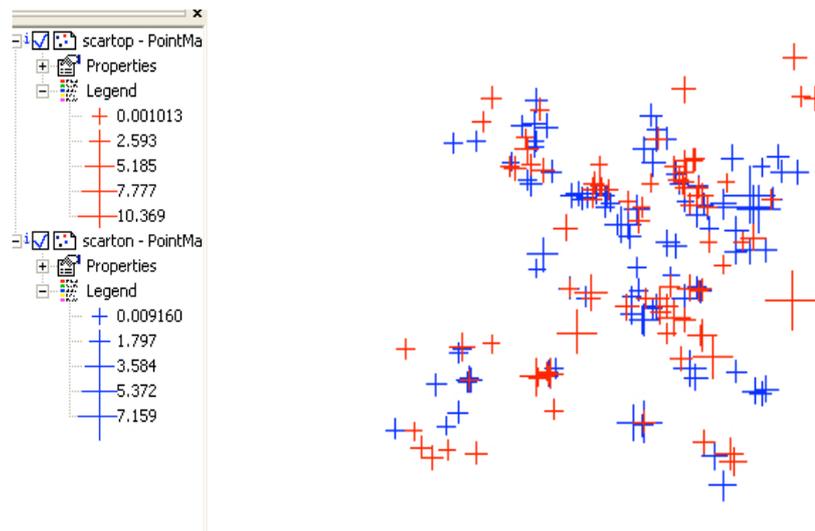
Mappa degli errori per il Krigaggio Ordinario



Mappa degli errori per il Cokrigaggio



Mappa degli errori per il Krigaggio con DE tentativo 1



Mappa degli errori per il Krigaggio con DE tentativo 2

Si nota che gli scarti positivi sono maggiori di quelli negativi, variano tra 0 e 10 mentre quest'ultimi tra 0 e 6/7, e entrambi sono distribuiti uniformemente. L'analisi qualitativa è stata accompagnata dal calcolo di una grandezza che permette un raffronto quali-quantitativo diretto: gli scarti al quadrato diviso la varianza di stima. Tale valore teoricamente dovrebbe coincidere con l'unità, se minore di uno significa che si stimano errori maggiori di quelli che si commettono, se maggiore di uno che si stimano errori minori di quelli che si commettono. Ho quindi provveduto ad estrarre dalle mappe della varianza di

stima ottenute e mostrate per i diversi metodi, i valori in corrispondenza delle coordinate del set di dati per il quale si hanno a disposizione i valori veri e gli stimati e quindi gli scarti al quadrato. I risultati ottenuti sono contenuti nella seguente tabella.

Metodologia di stima	Scarti al quadrato/varianza di stima
Cokrigaggio	2,1270
Kriging Ordinario Anisotropo	2,7140
Kriging con DE 1	2,9613
Kriging con DE 2	3,2251

Dalla valutazione di tale grandezza la stima con Cokrigaggio del 2005, sfruttando come variabile ausiliaria il 2004, risulta la migliore. Si nota come la stima con deriva esterna, pur essendo svolta con grandezze correlate non apporti un miglioramento, questo si può correlare con la difficoltà di individuare un trend per la variabile a scala europea, ossia una deriva esterna che assorba una quota di variabilità consistente. Nel primo tentativo di Krigaggio con deriva esterna infatti, il *sill* della deriva, che sommato a quello della fluttuazione dev'essere minore o uguale a quello della variabile iniziale, assorbe uno 0.210 di una variabilità totale di 2,773, nel secondo tentativo ancora meno, uno 0.121.

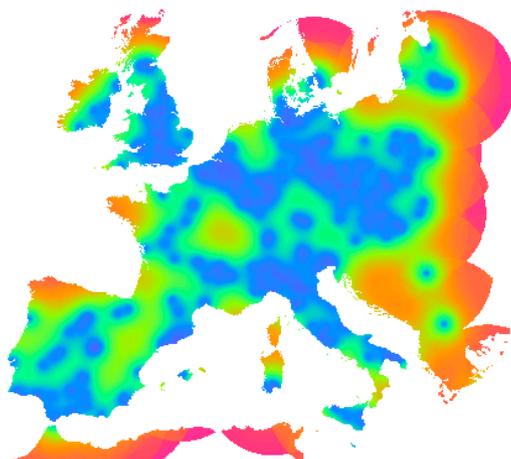
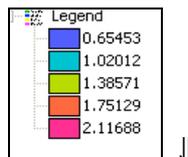
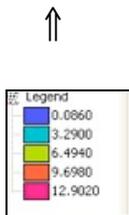
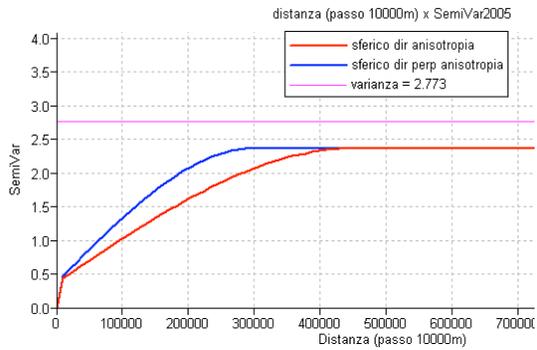
Si deve comunque tenere conto che gli errori reali, anche nel caso del Cokrigaggio, sono sistematicamente maggiori della varianza di stima teorica, il che trova spiegazione dal momento che un modello ad una scala come quella europea e a partire da dati eterogenei come possono essere quelli di traffico, urbani e rurali, porta a regolarizzare l'andamento della variabilità reale delle concentrazioni.

7.2.Raffronto fra le cartografie

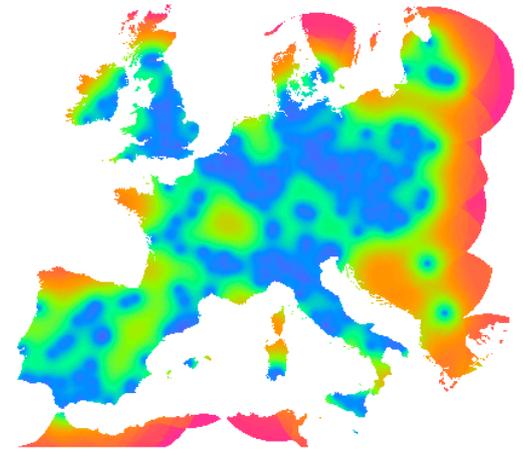
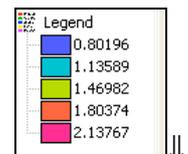
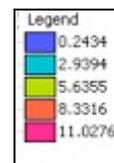
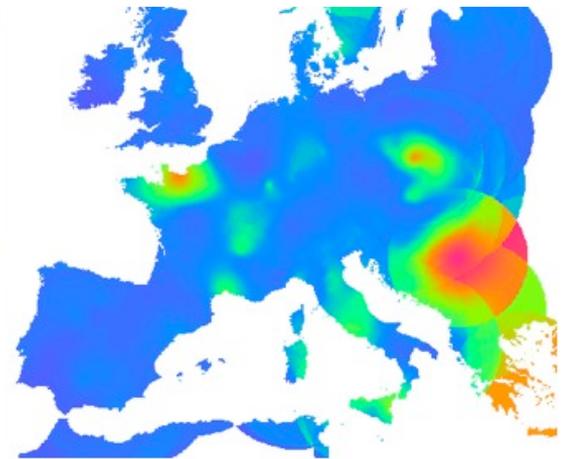
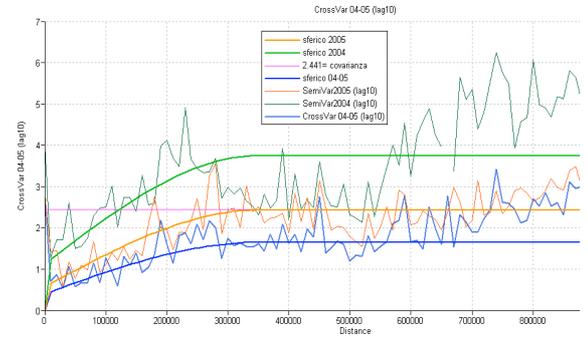
Fase finale dell'analisi dei risultati è stato un raffronto tra le diverse cartografie, in modo tale da quantificare le differenze in termini di valori di concentrazione, massimi e minimi di concentrazione, valori della varianza di stima. Sempre a partire dalla cartografia ottenuta per il 2005, l'operazione basilare di tale fase, oltre al raffronto in parallelo delle diverse mappe, è stata quella di ottenere delle mappe delle differenze fra le cartografie; Prendendo come riferimento la cartografia mediante Krigaggio Ordinario, che corrisponde all'approccio basilare e più diffuso quando non si hanno variabili ausiliarie e nell'ipotesi di stazionarietà, ho effettuato rispetto a questa le differenze con le cartografie ottenute dagli altri metodi di stima.

Si riportano di seguito variogrammi, mappe di stima e varianza di stima relativi al Krigaggio Ordinario anisotropo preso come riferimento e al modello di volta in volta confrontato con questo. Segue a tale raffronto grafico dei risultati ottenti e già mostrati in precedenza nel corso della trattazione, indipendenti l'uno dall'altro, la mappa delle differenze e le statistiche ad essa correlate.

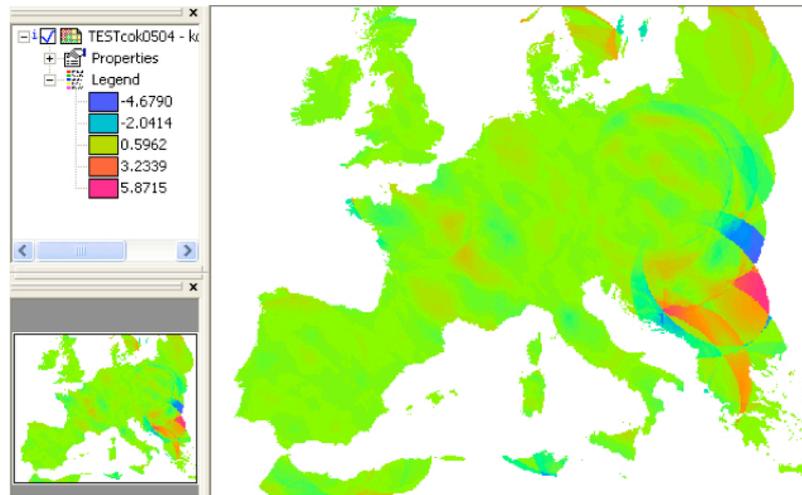
Variogramma, mappa di stima e varianza di stima per il Krigaggio Ordinario anisotropo '05'.



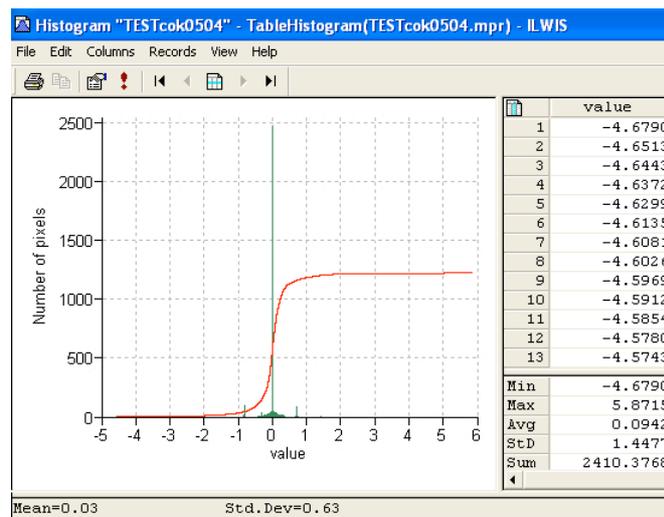
Variogramma, mappa di stima e varianza di stima per il cokrigaggio '04-05'.



Mappa della differenza KO anisotropo- Cokrigaggio 2004-2005

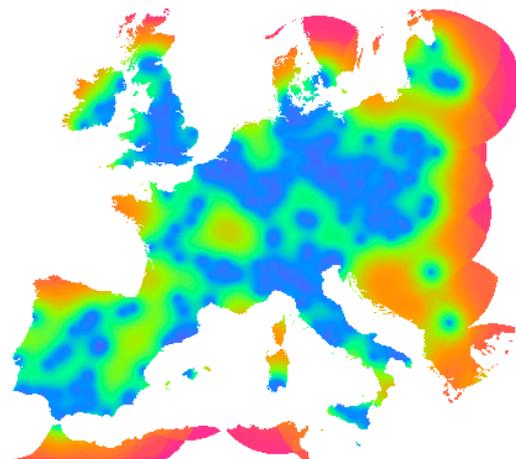
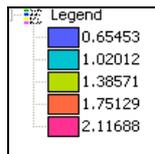
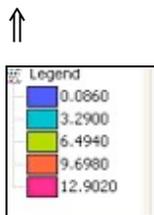
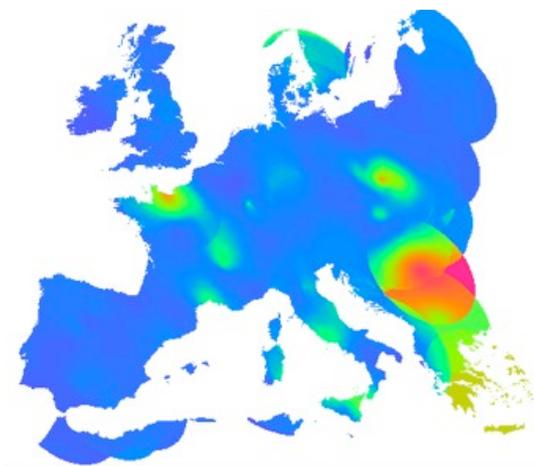
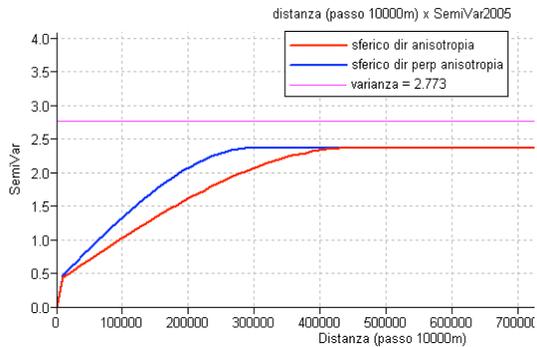


Statistiche della differenza KO anisotropo- Cokrigaggio 2004-2005

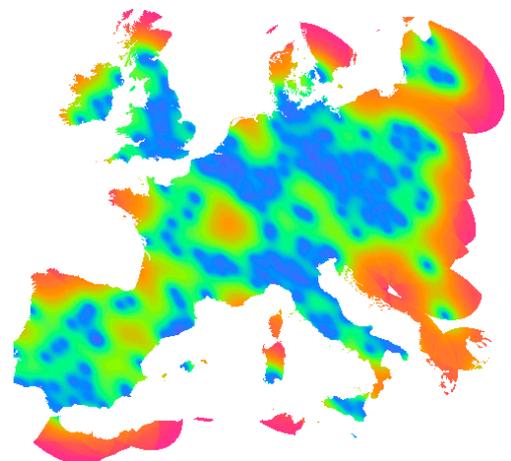
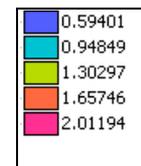
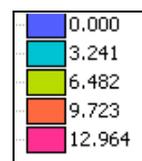
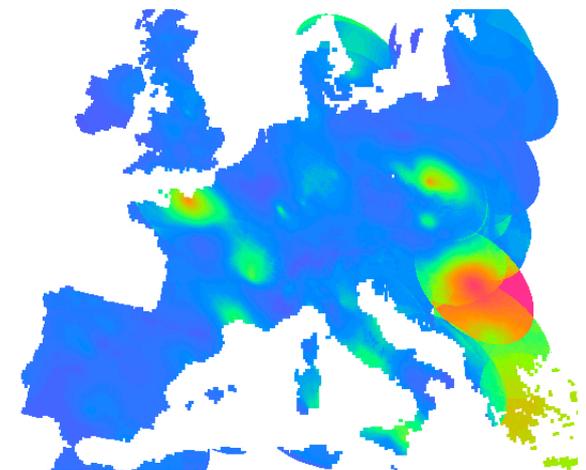
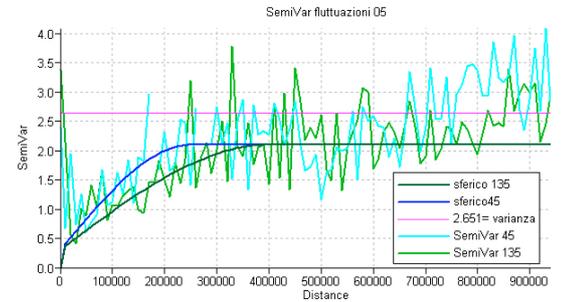


Si nota come il Krigaggio Ordinario porti a stimare concentrazioni leggermente maggiori per il 2005 rispetto a quelle ottenute con il cokrigaggio, un massimo di 12,9 per il KO contro 11 del cokrigaggio. I valori della varianza di stima oscillano tra gli stessi valori. Le differenze variano da -4,6 a 5,8 e hanno media pari a 0,03. E' evidente graficamente come nel Krigaggio Ordinario non si riesca a tener conto dell'anisotropia e come il fatto di avere a disposizione i dati della variabile ausiliaria porti ad una mappatura migliore, più variabile.

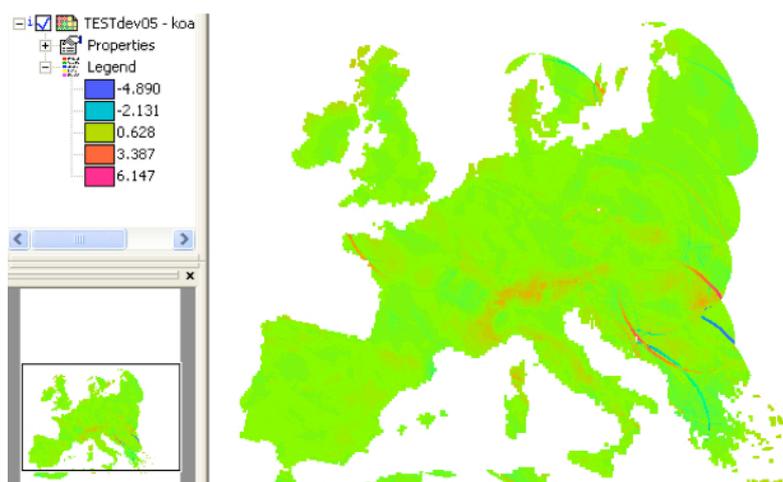
Variogramma, mappa di stima e varianza di stima per il Krigaggio Ordinario anisotropo 05'.



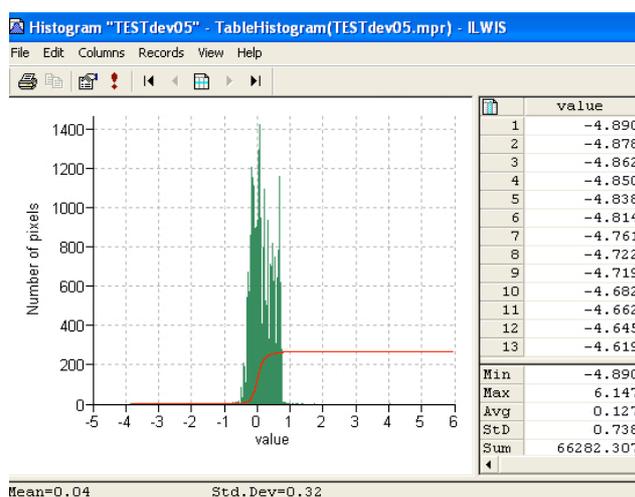
Variogramma, mappa di stima e varianza di stima per il Krigaggio con deriva Esterna 1.



Mappa della differenza KO anisotro- KDE (1) 2005

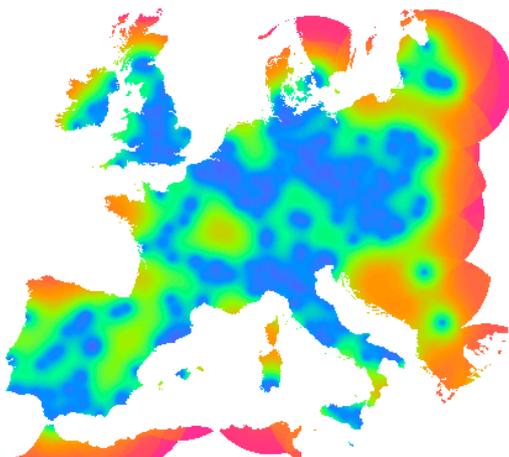
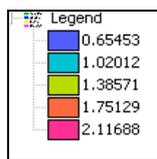
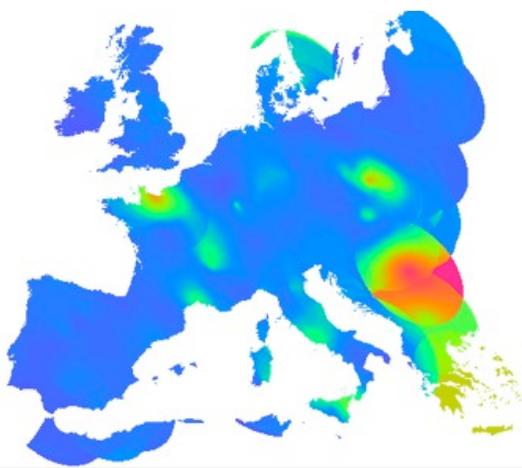
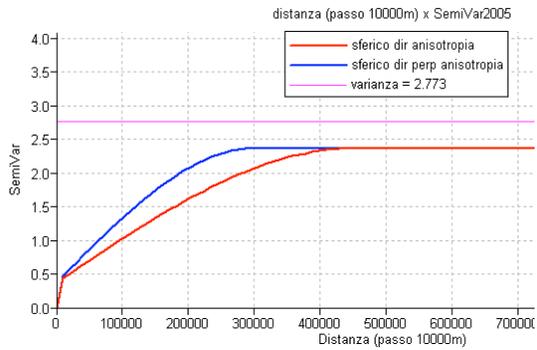


Statistiche della differenza KO anisotro- KDE (1) 2005

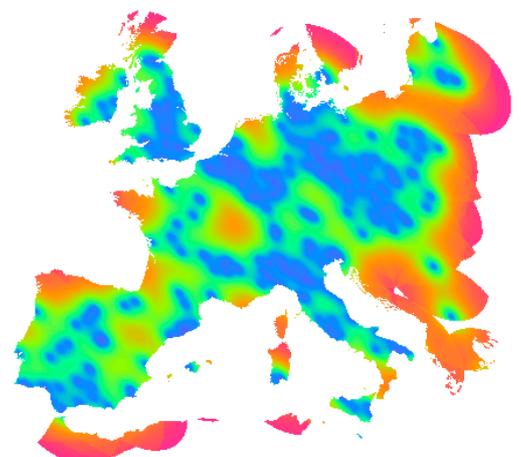
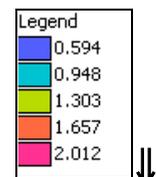
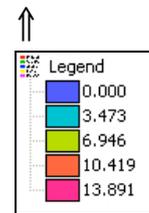
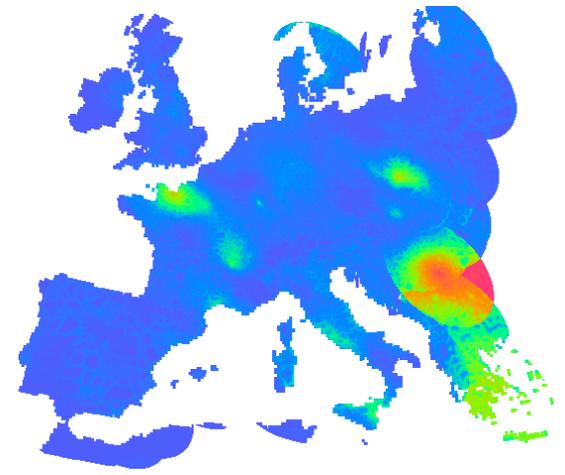
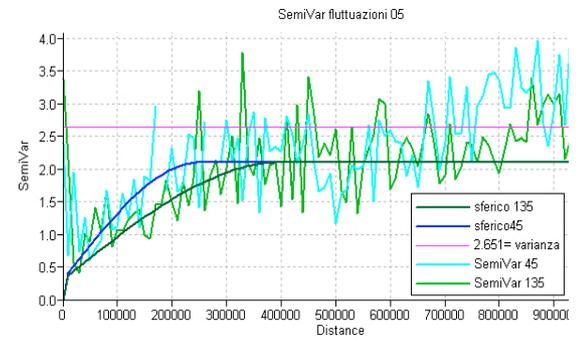


Si nota che il KDE porta a stimare concentrazioni in linea con quelle del KO, che i massimi sono circa coincidenti, così come la varianza di stima varia nello stesso range valori. Le differenze variano da -4,8 a 6,1 e hanno media pari a 0,003, sono maggiori quindi rispetto al caso precedente. E' evidente l'influenza dell'anisotropia anche nel KDE essendo il variogramma delle fluttuazioni anisotropo. Nella mappa del KDE le concentrazioni pari a 0 sono estremamente evidenti e ricalcano i massimi della carta tematica del vento

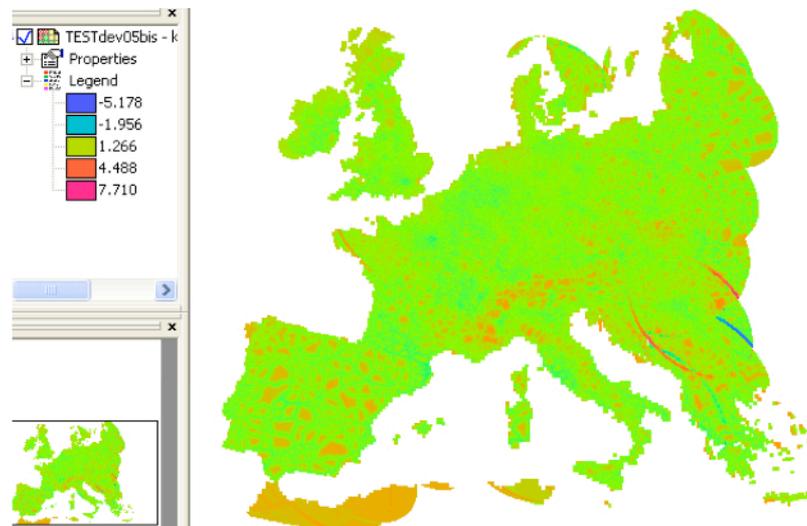
Variogramma, mappa di stima e varianza di stima per il Krigaggio Ordinario anisotropo 05'.



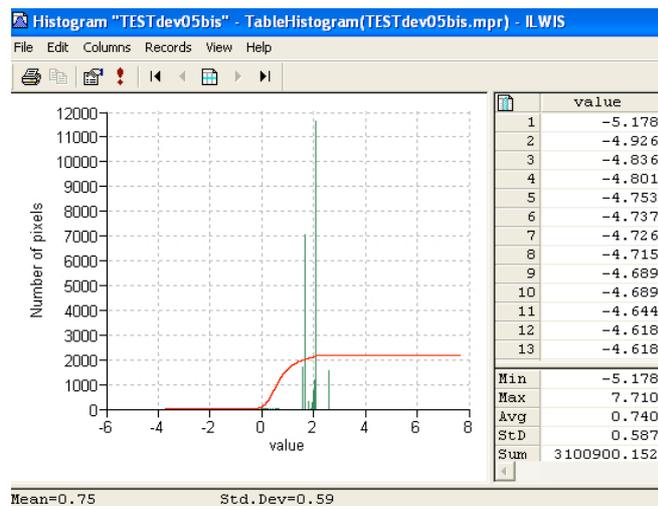
Variogramma, mappa di stima e varianza di stima per il Krigaggio con deriva Esterna 2.



Mappa della differenza KO anisotropo KDE (1) 2005

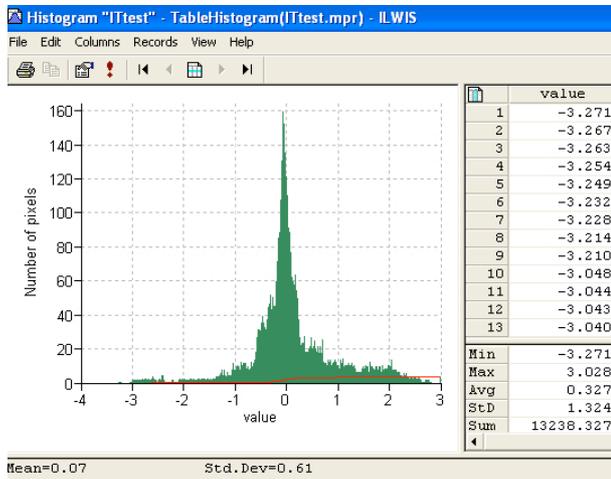


Statistiche della differenza KO anisotropo- KDE (2) 2005

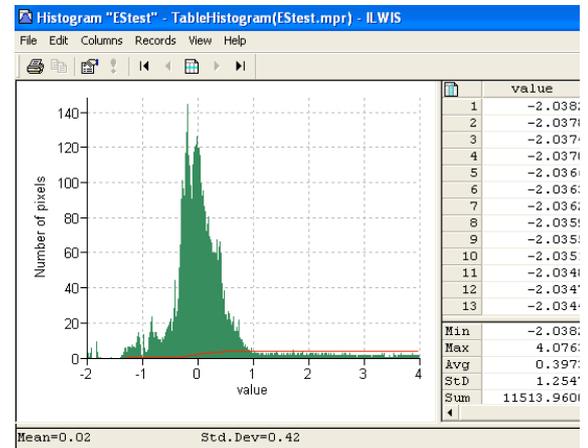


Si nota che il secondo tentativo di KDE porta a stimare concentrazioni maggiori di quelle del KO, mentre la varianza di stima varia nello stesso range valori. Le differenze variano da -5,2 a 7,7 e hanno media pari a 0,75 , valore più elevato tra quelli esaminati. E' evidente l'influenza dell'anisotropia anche nel KDE essendo il variogramma delle fluttuazioni anisotropo. Il reticolo stradale appare nettamente in rilievo nella mappa con valori compresi tra 3 e 4.

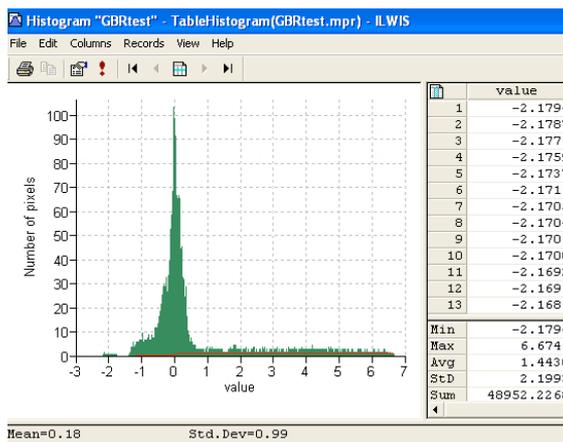
Relativamente all'analisi a scala locale trattata nel capitolo 6, si riportano le statistiche delle differenze tra KO a scala europea e KO locale.



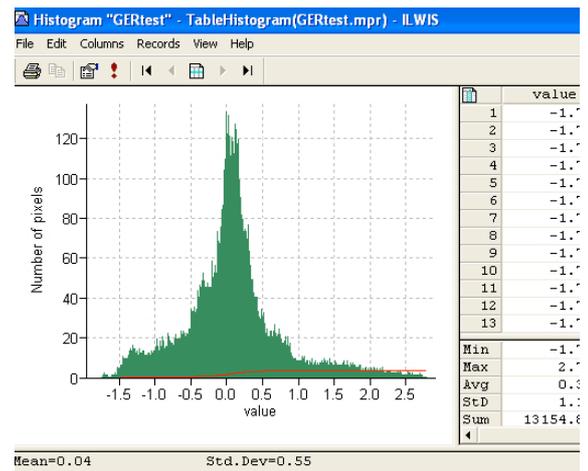
Mappa delle differenze KOeuropa-KO Italia



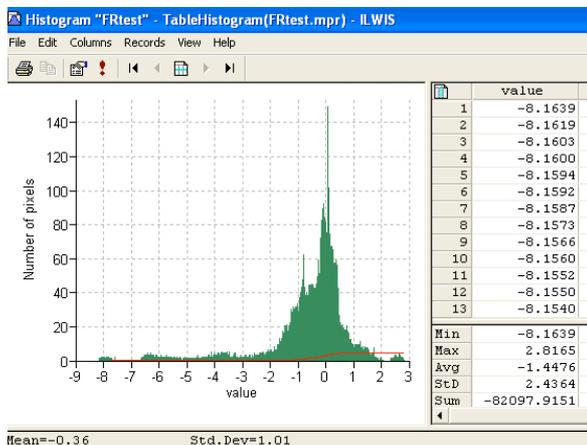
Mappa delle differenze KOeuropa-KO Spagna



Mappa delle differenze KOeuropa-KO Granbretagna



Mappa delle differenze KOeuropa-KO Germania



Mappa delle differenze KOeuropa-KO Francia

Lo scarto tra le due mappe è in questo caso indicativo dal momento che il meccanismo di stima è lo stesso, a cambiare è solo la scala. Le medie delle differenze, come prevedibile dall'analisi dei variogrammi, per la quale si rimanda al capitolo 6, sono più consistenti; per l'Italia, la cui soglia di variabilità era in linea con quella europea, la media degli scostamenti è coerentemente (si ricorda che il metodo è lo stesso: KO anisotropo) bassa: 0,07 e lo stesso ordine di grandezza si mantiene per la mappa delle differenze di Germania e Spagna rispettivamente: 0,04 , 0,02. Per la Gran Bretagna, che aveva una soglia di variabilità molto inferiore a quella europea, la media delle differenze tra le 2 mappe alla due diverse scale è di 0,18, per la Francia coerentemente alla variabilità interna alla nazione maggiore di quella a livello europeo, il valore è negativo e pari a $-0,36$.

Tali risultati pongono quindi importanti temi di riflessione sull'interpretazione delle mappe tematiche in funzione del metodo di stima e della scala di lavoro, la scelta di tali parametri deve essere effettuata in base alla finalità della carta, e tali argomentazioni saranno oggetto del seguente capitolo conclusivo.

8. Conclusioni e discussioni

Ricapitolando quelli che sono stati gli obiettivi alla base del presente studio:

1. capire correlazioni spazio temporali e la dispersione del benzene a diverse scale spaziali temporali: costruzione di una mappa tematica di concentrazione sul suolo europeo;
2. studio con grandezze ausiliarie per individuare le correlazione con diversi tematismi e integrare le variabili significative nella ricostruzione spaziale della distribuzione;
3. Validazione e confronto dei modelli e dei metodi alla base della cartografia prodotta, al fine di individuare la tecnica più appropriata in relazione alla finalità dell'utilizzo dell'informazione cartografica prodotta;

Si possono trarre, relativamente ad ognuno di essi, le seguenti conclusioni:

1. la prima informazione utile al fine di interpretare la correlazione spaziale della variabile benzene, è emersa dall'analisi strutturale a scala europea, che ha mostrato nelle mappe variografiche, una netta anisotropia della variabilità, con direzione principale di anisotropia a 135° ; tale direzione, così individuata, trova riscontro in quella che è la direzione di dispersione degli inquinanti a livello europeo. E' da sottolineare infatti, che a scala locale, relativamente alle 5 aree omogenee prese in esame, il comportamento della variabile è isotropo, ad eccezione dell'Italia dove l'anisotropia è dovuta alla conformazione geografica del territorio.

Sempre all'interno dell'analisi strutturale, contestualmente al calcolo dei variogrammi anisotropi, si è rilevato un effetto pepita consistente all'origine, molto probabilmente da attribuire alle stazioni ravvicinate dei centri urbani, come approfondito all'interno dell'analisi a scala locale, mediante l'operazione di aggregazione dei "dati vicini", ossia dei dati che distano meno di 5000 m. Tale operazione ha mostrato un netto abbassamento della soglia di variabilità per le aree omogenee di Italia e Spagna, e la riduzione di un consistente effetto pepita all'origine per l'area della Francia, evidenza che conferma quanto ipotizzato relativamente al

disturbo creato da stazioni ravvicinate.

Continuando con l'approccio locale al problema, è emerso che a tale scala la variabilità ha sempre lo stessa struttura individuata a scala europea, ma soglie differenti, che possono scostarsi dal trend globale sia per eccesso, come nel caso della Francia, che per difetto come nel caso delle altre aree omogenee.

Relativamente ai valori numerici delle stime, i risultati ottenuti sono in pieno accordo con quanto contenuto nel report dell'Agenzia Europea dell'Ambiente che riporta lo stato dell'inquinamento atmosferico dal 1994 al 2000. Trova conferma in tale report la distribuzione dei massimi della concentrazione di benzene coincidente per tutte le stime con la zona dei balani compresa tra Grecia, Romania e Repubblica Ceca, riportando come uniche stazioni con concentrazioni eccedenti ai valori di normativa: una in Grecia (traffic), una in Romania (urbana) e una in Repubblica Ceca. Deduzioni interessanti possono trarsi anche sul trasporto dell'inquinante: emerge la presenza di benzene in zone remote del suolo europeo anche in concentrazioni rilevanti, e questo è da attribuire al trasporto a scala europea.

Alla scala temporale, dal 2003 a 2005, si è riscontrata una diminuzione delle concentrazioni, e questo già dall'analisi delle nuvole di correlazione; come prevedibile poi, le concentrazioni di benzene in anni di campionamento diversi sono risultate molto correlate, con indici del coefficiente di correlazione dell'ordine di 0,8 , e tale forte correlazione, insieme alla non isotopia delle stazioni nei diversi anni, ha fatto sì che la stima mediante cokrigaggio effettuata nello studio di correlazione temporale, abbia fornito, come si argomenterà nel seguito delle conclusioni, i risultati migliori.

2. Lo studio della correlazione con le variabili ausiliarie ha evidenziato

diverse informazioni qualitative: quasi la totalità dei valori della concentrazione di benzene è contenuto entro i 200 m di quota sul livello del mare; le concentrazioni maggiori si individuano in zone pianeggianti; per precipitazioni molto elevate, maggiore ai 70 mm/mese, i valori di concentrazione di benzene sono bassi; la velocità media del vento a 10 m e i giorni medi piovosi mensili sono in relazione di proporzionalità inversa con la concentrazione di benzene; per valori di temperatura superiori ai 10°C le concentrazioni di benzene sono maggiori, il che potrebbe essere correlato con le alte pressioni che si hanno a temperature elevate; la distanza dalle strade principali mostra una correlazione iperbolica con la concentrazione di benzene, tale risultato è in linea con quella che è la maggior fonte dell'inquinante: il traffico; l'uso del suolo urbano discontinuo risulta maggiormente correlato con l'inquinante dell'uso del suolo urbano continuo; da un punto di vista quantitativo le variabili maggiormente correlate risultano: velocità del vento media mensile a 10 m, giorni piovosi medi mensili, distanza dalle strade principali, uso del suolo urbano discontinuo, utilizzate all'interno del Krigaggio con Deriva Esterna.

3. La validazione e il confronto dei diversi approcci teorici metodologici e delle corrispondenti cartografie ottenute, ha mostrato che la stima più attendibile risulta quella ottenuta dal cokrigaggio che sfrutta come variabile ausiliaria la concentrazione di benzene in anni precedenti o successivi a quello in esame; tale risultato trova spiegazione nella forte correlazione tra i valori della variabile per i diversi anni, (che però non sono tra loro in correlazione intrinseca, condizione che renderebbe il cokrigaggio coincidente con il krigaggio ordinario), e nel fatto che le stazioni non sono isotopiche e che quindi la variabile ausiliaria apporta informazioni spaziali in punti in cui la variabile da stimare non è nota, questo implica anche una migliore cartografia, più variabile e graduata.

Il risultato che il Krigaggio con deriva esterna non apporti miglioramenti alla stima può essere interpretato come dovuto alla difficoltà di individuare un trend efficace ad una scala come quella europea e all'eterogeneità dei dati in ingresso, relativi ad aree urbane, rurali, stazioni *traffic*.

Le cartografie per aree omogenee, sebbene ottenute con Krigaggio ordinario anisotropo, si discostano come prevedibile dall'analisi dei variogrammi, dal risultato del krigaggio anisotropo a scala europea.

Tali risultati aprono una discussione sulla scala di lavoro, emerge che il modello europeo oltre ad essere necessario allo scopo di valutazione di inquinamento transfrontaliero e di politica comunitaria, permette di ravvisare l'anisotropia a 135° e le relazioni con le variabili ausiliarie geografiche, che a piccola scala non emergono. Lavorando alla scala locale, pur mantenendo questa la stessa struttura di quella europea per le diverse aree, non mantiene gli stessi parametri e quindi essendo diverse le soglie di variabilità, anche le stime sono diverse, e stimare il locale con il modello globale comporterebbe inevitabili incongruenze; dal momento che l'analisi a piccola scala è indispensabile per capire i trend di variabilità locali e quindi più adatta a valutazioni di rischio connesse all'inquinante, mentre lo studio a scala globale permette di mettere in luce tematismi, quali correlazioni con fattori esterni, anisotropia, che invece a piccola scala non emergono, i due studi andrebbero affiancati o quantomeno scelti in funzione dell'aspetto che si desidera studiare e con la consapevolezza di quanto esposto.

Un'operazione utile invece ai fini di migliorare l'informazione in ingresso, potrebbe essere quella di differenziare gli studi in base alla tipologia di stazioni, per vedere se questo riesce ad eliminare il disturbo che ne consegue sulle stime, rendendo il file dati omogeneo, quindi migliorando le correlazioni con grandezze esterne che potrebbero fungere più efficacemente da derive nella stima.

Scegliendo come mappa di riferimento il cokrigaggio e confrontando i valori ottenuti con i limiti di legge di concentrazione, sottolineando che all'interno della normativa non si fa riferimento al supporto spaziale per l'analisi, nell'intervallo di confidenza probabilistico associato alle decisioni, i trend a meno di massimi localizzati, stanno sotto il valore limite.

Si può concludere quindi affermando che lo studio e la caratterizzazione spazio temporale di variabili ambientali, deve essere visto come uno strumento versatile che deve adeguarsi ogni volta in termini di scala e correlazioni con fattori esterni, a quelli che sono i fenomeni analizzati; si deve tener conto infatti di tutte le ipotesi che stanno dietro i rispettivi approcci e associare l'incertezza di ogni stima anche alle decisioni che vengono prese basandosi su questi.

BIBLIOGRAFIA

Wackernagel H. 1995. *Multivariate Geostatistics*. Springer

Bruno, R. and G. Raspa. 1993. Integration between geostatistical methodologies and GIS environment geodata: the external drift. *Conf. Proc. of fourth European Conference and Exhibition on Geographical Information Systems*,

Tomislav Hengl 2007: *A Practical Guide to Geostatistical Mapping of Environmental Variables* JRC Scientific&Technical Reports

EEA Report No 2/2007 :*Air pollution in Europe 1990–2004*

Dr. James Franklin 2006 *Long-Range Transport of Chemicals in the Environment* Science Dossier.

A.Pistocchi, S.Galmarini *Evaluation of a simple spatially explicit model of atmospheric transport of pollutants in Europe* EC, DG JRC, Institute for Environment and Sustainability

Edzer J.Pebesma, Kor de Jong and David Briggs 2006 *Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example* International Journal of Geographical Information Science

International Institute for Geo-Information Science and Earth Observation
ILWIS 3.0 User's Guide

Siti internet

<http://www.eea.europa.eu/>

<http://www.eper.cec.eu.int/eper/>

<http://nuke.georesources.org/>

<http://www.itc.nl/>

<http://www.geovariances.com/>