

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

In Ingegneria Energetica

Ciclo XXI

Settore/i scientifico disciplinari di afferenza: ING-IND/18-19-20

TITOLO TESI

Costruzione di un Sistema Informativo Geografico Digitale
per analisi ambientali.

Presentata da: Ferrucci Barbara

Coordinatore Dottorato

Prof. Antonio Barletta

Relatore

Prof.Ing.Marco Sumini

Esame finale anno 2009

INTRODUZIONE

Questo lavoro ha come obiettivo l'utilizzo del Geographical Information System (GIS) [1, 2, 3] per analisi di sicurezza, monitoraggio e valutazioni di impatto ambientale.

Un Geographical Information System cattura, immagazzina, analizza, gestisce e presenta i dati che facciano riferimento ad una locazione. In senso generale, le applicazioni GIS sono strumenti che permettono agli utenti di formulare richieste interattive, analizzare informazioni spaziali, editare dati, mappe e presentare i risultati di tutte queste operazioni. La tecnologia del geographical information system può essere usata per investigazioni scientifiche, gestione delle risorse, gestione degli assetti, archeologia, analisi di impatto ambientale, pianificazione urbana, cartografia, criminologia, marketing, logistica e altri scopi. Per esempio, il GIS può permettere ai pianificatori delle emergenze di valutare facilmente le conseguenze di una situazione incidentale o di un disastro naturale, come pure può essere usato per valutare il livello di inquinamento di una regione.

Praticamente la totalità delle operazioni GIS possono essere svolte con software open source. Con l'uso generalizzato di formati dati non proprietari come gli Shape File, per dati vettoriali, e Geotiff, per dati raster, così come l'adozione di standar OGC per soluzioni server, lo sviluppo del software open source continua ad evolvere, specialmente per applicazioni rientrate ai servizi web. Soluzioni open source ben note includono GRASS GIS, Quantum GIS, MapServer, uDig, OpenJUMP, gvSIG e molti altri.

Tra le soluzioni open source, si è scelto di utilizzare il GRASS [4, 5, 6, 7, 10, 8, 9], disponibile nei termini della GNU public license (GPL), mostrando l'usabilità e le notevoli potenzialità di tale software, nonché la qualità dei pro-

dotti ottenibili, mai inferiori ai prodotti e agli strumenti messi a disposizione dai più radicati e diffusi programmi proprietari, spesso onerosi e inutilizzabili per produrre proprie e personalizzate soluzioni software.

Nel capitolo 4, vedremo l'applicazione all'analisi delle conseguenze di ipotetici incidenti, durante le operazioni di dismissione dell'impianto di processamento del combustibile nucleare, di Bosco Marengo (AL).

Nel capitolo 5, vedremo applicazioni nel campo del monitoraggio della qualità dell'aria tramite analisi di immagini satellitari. Si vedrà allora l'importanza della tecnologia GIS per contestualizzare al territorio i risultati delle rilevazioni, operazione indispensabile questa, per una corretta interpretazione del fenomeno, per mettere in relazione concentrazioni di particolato con insediamenti industriali o, più in generale, umani e per fare studi epidemiologici.

Capitolo 1

Introduzione ai GIS

1.1 Un po' di storia

Tra il 1958 e il 1961 il dipartimento di geografia dell'Università di Washington effettuava delle ricerche sui metodi statistici spaziali e crea i primi programmi per rappresentare la cartografia su un calcolatore. Poco tempo dopo, nel 1964, nasce l'Harvard Laboratory for Computer Graphics and Spatial Analysis, che si prefiggeva lo scopo di produrre un software apposito per la creazione e la manipolazione della cartografia, che fino agli anni '80 prenderà il nome di "cartografia numerica".

Tra i programmi prodotti dall'Harvard Laboratory è opportuno ricordare SYMAP [12], sviluppato nel 1964 contestualmente alla fondazione dell'Harvard Laboratory. Per svariati anni SYMAP rimase il software di riferimento per chiunque dovesse lavorare con la cartografia. Il programma utilizzava un modello di vettore e una collezione di oggetti (punti, linee ed aree) contraddistinti da coordinate spaziali, con associati un insieme di valori alfanumerici. Si potevano così generare mappe di varie dimensioni, caratterizzate da diversi simboli, legende, titoli e così via. Ovviamente al momento della nascita di SYMAP non vi era la possibilità di digitalizzare le immagini cartacee, perciò tutti i dati andavano inseriti manualmente all'interno del sistema. Quasi contemporaneamente, nel 1962, Roger F. Tomlinson sviluppava il Canada Geographic Information System (CGIS) [13], spesso definito "il primo

vero GIS", creato per gestire l'immensa quantità di dati prodotta dal Canada Land Inventory. Il CGIS fu largamente utilizzato per tutti gli anni '70, fino ai primi anni '80 e con il succedersi delle versioni fu introdotta la possibilità di scannerizzare le immagini cartacee, per poi vettorializzarle e descriverle con dei valori alfanumerici. Inoltre per la prima volta vi fu l'idea di separare i dati in file di attributi e file di georeferenziazione; nascono così i descriptor dataset (DDS) e gli image dataset (IDS). Un'ulteriore funzionalità innovativa offerta dal CGIS fu la possibilità di misurare distanze ed aree e quella di poter disegnare dei poligoni per interrogare gli attributi della cartografia sottostante.

Nel 1969 Jack Dangermond fonda l'Environmental Systems Research Institute (ESRI), i cui prodotti definiranno lo standard per la cartografia numerica nei vent'anni a seguire.

Nel 1980 ESRI sviluppa ARC/INFO che a tutt'oggi viene considerato il primo GIS moderno, in quanto per la prima volta si integra in una stessa soluzione software uno strumento CAD per manipolare la cartografia e un DBMS per gestire i dati relativi al territorio. E' doveroso precisare che nonostante esistessero già i database relazionali e il linguaggio SQL, tale soluzione non fu ancora adottata in ambiente GIS, ove si utilizzavano database proprietari (con un formato proprietario dei dati). Nel 1982 ARC/INFO veniva installato sui primi personal computer e per la prima volta alla piattaforma UNIX si aggiunse quella Windows. Uno dei fattori chiave del successo riscosso da tale software fu quello di essere corredato da un linguaggio di macro (Arc Macro Language) che permetteva agli utenti di utilizzare ARC/INFO come piattaforma per sviluppare applicazioni cartografiche specifiche per le loro esigenze.

In questo periodo i prodotti della ESRI dominano l'intero mercato dei GIS e costituiscono lo standard di riferimento. Parallelamente all'evoluzione dei software GIS, in questi anni assistiamo al proliferare degli strumenti CAD, necessari per disegnare in maniera vettoriale la cartografia. Nel 1980 l'Intergraph Corporation rilascia il primo sistema CAD interattivo, denominato IGDS (Intergraph Graphics Design System), che conquistò in poco tempo il mercato dei CAD.

Tra il 1990 e il 1995 si assiste alla nascita ed affermazione dei personal computer e delle reti locali di calcolatori, al miglioramento e semplificazione delle interfacce-utente, all' utilizzo diffuso della grafica a colori nonché ad un forte aumento delle prestazioni di elaborazione e memorizzazione dei dati. I database diventano aperti e questa svolta tecnologica segna l'inizio della fase dove i dati diventano i veri e propri soggetti dei sistemi informatici, che da ora in avanti possono essere denominati propriamente Sistemi Informativi Territoriali (SIT).

Nel 1994 viene fondato l'Open Geospatial Consortium (OGC), un consorzio internazionale no-profit che oggi conta più di 335 soggetti, tra società private, enti governativi ed università, che si prefigge lo scopo di portare avanti un processo aperto di consenso per incoraggiare lo sviluppo di servizi e contenuti di carattere geo-spaziale, per definire gli standard per l'elaborazione e l'interscambio dei dati GIS.

Per la prima volta si è intenzionati a rendere le diverse soluzioni software interoperabili tra di loro, affinché i dati rimangano i soggetti indiscussi dei Sistemi Informativi Territoriali. Per supportare la gestione di un elevato numero di servizi, i SIT non prevedono la creazione di un database centrale nel quale importare i dati provenienti dalle banche dati esterne, ma piuttosto operano l'integrazione di queste ultime, lasciando loro la completa autonomia di gestione ed aggiornamento. In questo modo ciascun ente preposto all'erogazione di un determinato servizio rimane l'unico proprietario e responsabile del mantenimento dei propri database e, grazie all'esistenza di standard aperti per i dati GIS, il Sistema Informativo Territoriale sarà in grado di accedere a tali dati per integrarli eventualmente con dati provenienti da differenti fonti e offrire così un servizio aggregato di complessità e utilità maggiore. In questo periodo quindi, si dispone di uno standard pubblico, rilasciato dall'OGC e denominato OpenGIS, che descrive la strutturazione dei dati GIS. Con tale presupposto inizia una fase di larga diffusione dei Sistemi Informativi Territoriali presso le Pubbliche Amministrazioni, che vedono nel SIT uno strumento indispensabile per gestire i numerosi servizi che devono offrire.

Il periodo tra il 1995 e il 2000 è caratterizzato da una forte espansione di

Internet, che entra nelle case della gente e diventa uno strumento indispensabile per la vita lavorativa, ciò porta ad un aumentano ulteriore delle prestazioni di elaborazione e memorizzazione dei calcolatori e conseguentemente la grafica 3D inizia a diventare un efficace strumento di rappresentazione e non soltanto di intrattenimento.

Questo periodo segna il momento di massima espansione dei SIT presso le Pubbliche Amministrazioni e gli Enti di pubblici servizi. L'espansione di Internet offre la possibilità di aprire alcune delle funzionalità del SIT direttamente verso i cittadini. Il Comune di Bologna è il primo in Italia (e tra i primi in Europa) ad offrire ai cittadini la possibilità di consultare la cartografia direttamente dal Web.

Nel 1996 nascono gli spatial database, ovvero delle banche dati ottimizzate per memorizzare e interrogare dati relativi ad oggetti collocati nello spazio, inclusi punti, linee e poligoni. Per gestire i tipi di dati spaziali sono necessarie delle funzionalità aggiuntive particolari che l'Open Geospatial Consortium ha definito e rese pubbliche. In questo periodo si cominciano ad utilizzare le foto aeree e da satellite come immagini raster, per poi disegnarvi sopra la cartografia. Sempre durante questo periodo si afferma una nuova tecnologia, che consente l'acquisizione diretta di immagini tridimensionali, si tratta del rilevamento aero-fotogrammetrico. Esso viene effettuato mediante riprese fotografiche aeree con centro di presa e focale nota, elaborando queste ultime mediante punti fotografici noti a terra, tramite un procedimento detto stereo-foto-restituzione.

Tra il 2000 e il 2005 continua l'aumento delle capacità di elaborazione e memorizzazione dei calcolatori, si ha : il miglioramento della connettività ed esplosione delle connessioni a banda larga, in particolare fibra ottica, la diffusione della tecnologia GPS, su automobili e dispositivi portatili, una larga diffusione di sensori automatici (es: autobus con la stima automatica dell'orario, sensori di traffico, ecc.) e di soggetti che mettono in rete dati georeferenziati.

I sistemi GIS e in generale i software che implementano i SIT raggiungono in questo periodo un alto grado di efficienza e ricoprono un ruolo centrale in pressoché tutte le funzioni cui la Pubblica Amministrazione deve assolvere.

Ad esempio i tecnici di diversi uffici comunali sono in grado di lavorare contemporaneamente sulla medesima cartografia, che resta sempre aggiornata in tempo reale e pubblicata automaticamente sul Web (ovviamente limitatamente ai layer che possono essere di pubblico dominio). Ormai le foto aeree e satellitari sono facilmente acquistabili e vengono integrate alla cartografia del SIT, la cui consultazione è sempre più realistica e dettagliata. Il SIT diventa anche uno strumento per promuovere il turismo e permette a potenziali visitatori di visualizzare la mappa di un determinato luogo, corredata di fotografie, luoghi di interesse culturali, sportivi, gastronomici e quant'altro. In ultimo è importante sottolineare la grande espansione dei software di navigazione GPS, che sotto certi aspetti si possono definire dei Sistemi Informativi Territoriali. Navtech e Teleatlas sono le prime aziende a realizzare una cartina stradale mondiale.

Gli anni successivi al 2005 sono caratterizzati da fattori come : diffusione capillare di Internet a banda larga, alto grado di mobilità degli utenti, reti wireless ampiamente diffuse, tecnologia GPS integrata nei dispositivi portatili (come orologi, palmari e fotocamere). Già da qualche anno si stavano moltiplicando i siti Web che offrivano la possibilità di cercare un indirizzo stradale e visualizzare la mappa relativa.

Nel 2004 Google Maps diventava il sito di riferimento in questo ambito, offrendo le mappe dell'intero pianeta corredate in maniera interattiva con tutti i punti d'interesse.

Quando nel 2005 venne rilasciata la prima versione di Google Earth, le modalità con le quali gli utenti erano abituati a navigare sul territorio cambiarono radicalmente. Per la prima volta la tecnologia di rendering 3D risultò semplice da utilizzare ed accessibile a tutti, senza la necessità di disporre di hardware particolare o di una connessione ad Internet particolarmente veloce. Le principali innovazioni che furono introdotte con Google Earth consistono nella possibilità di navigare della cartografia tridimensionale (ottenuta a partire da immagini da satellite) in modalità remota, con tempi d'attesa decisamente contenuti e con fluidità di movimento. Inoltre il controllo della navigazione avviene con un'interfaccia estremamente intuitiva ed essenziale e ciò ha consentito una rapida diffusione del software anche tra gli utenti meno

esperti. A queste funzionalità si aggiunge l'integrazione con diversi provider di contenuti (come Wikipedia).

Una svolta significativa nell'ambito dei SIT è avvenuta nel 2006 con il rilascio del Navigatore 3D della Regione Sardegna. Tale navigatore si presenta con un'interfaccia ed un modello di funzionamento molto simile a Google Earth, ma a differenza di quest'ultimo l'espedito della navigazione tridimensionale costituisce una modalità di accesso ai dati del SIT. Infatti la cartografia tridimensionale rappresenta il punto di contatto (ovvero l'interfaccia) tra il Sistema Informativo Territoriale della Regione Sardegna e l'utente, perciò, in accordo con il modello di funzionamento dei SIT, il territorio costituisce la chiave per accedere ai dati. Inoltre, a differenza di Google Earth, il Navigatore 3D della Regione Sardegna utilizza un vero e proprio modello tridimensionale del territorio sul quale sono state plasmate le fotografie satellitari, che a breve verranno sostituite con le fotografie aeree (con una precisione di 60 cm al suolo). Il 3D sul Web diventa il nuovo paradigma di accesso capillare alle informazioni territoriali.

Le tecnologie di mash-up che si stanno affermando in questo periodo sono destinate a proliferare in un numero sempre maggiore di ambiti. Esse permettono di collegare in maniera dinamica fonti di dati differenti fra loro (ad esempio se si visita il Monte Everest con Google Earth, è possibile cliccare sopra il nome del monte ed aprire la relativa scheda su Wikipedia, o per fare un altro esempio, mentre si naviga con Sardegna 3D viene visualizzata la condizione meteo di ciascun luogo in tempo reale). Inoltre, in accordo con il paradigma che è stato definito Web 2.0, i contenuti presenti sul Web non sono più statici, ma possono venire creati in maniera esplicita od automatica dagli utenti. Ad esempio già alcune fotocamere digitali dispongono di un dispositivo GPS per aggiungere agli attributi delle fotografie le coordinate geospaziali del luogo cui sono state scattate; in un futuro prossimo sarà possibile, mentre si naviga sulla cartografia tridimensionale, visualizzare le risorse di immagini, di video e di audio relative ai luoghi che si stanno visitando.

Rimanendo nel tema GPS (che costituisce una tecnologia chiave nell'evoluzione dei Sistemi Informativi Territoriali, in quanto consente di georeferenziare in modo automatico le risorse che vengono create), nei prossimi anni i

SIT di supporto alla mobilità, come i navigatori per le automobili, saranno aggiornati in tempo reale sulle condizioni del traffico, del meteo, degli eventi in corso e quant'altro.

Un importante sfida per il futuro consiste nel raggiungimento dell'interoperabilità tra i numerosi provider di dati geografici. Si tratta dunque di una sfida tecnologica ed è auspicabile che l'Open Geospatial Consortium detenga un ruolo centrale per portare a termine questo ambizioso obiettivo.

1.2 Definizione e caratteristiche del GIS

Il termine Sistema Informativo Geografico deriva dall'inglese Geographical Information System, da cui l'acronimo GIS con cui esso è correntemente indicato. Un GIS è un metodo informatico per catalogare ed analizzare aspetti, fenomeni ed immagini del territorio e per generare mappe. Attraverso le tecnologie GIS è possibile studiare e visualizzare, sottoforma di livelli tematici sovrapponibili, una moltitudine di informazioni geografiche (raster e vettoriali) che, opportunamente collegate al database, consentono di effettuare analisi statistiche, generare mappe tematiche derivate, analizzare elementi grafici o consentire valutazioni geo-ambientali altrimenti non ottenibili in tempi brevi.

In particolare un Sistema Informativo Geografico risulta uno strumento fondamentale nella gestione del territorio e di situazioni di emergenza ambientale sia di origine umana (nucleare, industriale, civile ecc.) che naturale (terremoti, inondazioni, incendi, frane, erosione, ecc), mettendo a disposizione potenti strumenti per l'analisi congiunta dei parametri coinvolti e per la previsione del livello di rischio. La successiva realizzazione di mappe del rischio consente facili considerazioni sull'urgenza di interventi e la programmazione di strategie di mitigazione nelle aree indagate.

Il GIS può essere visto come una forma di DBMS (Database Management System, Sistema di Gestione di basi di dati) e viene definito sistema in quanto costituito da un insieme di componenti hardware, software e umane, che interagiscono fra loro e insieme consentono di gestire dati geografici e georeferenziati, vale a dire dati relativi ad elementi od oggetti della superficie terrestre la cui posizione è definita da un insieme di coordinate.

E' quindi un sistema che offre ampie possibilità di interazione con chi lo utilizza e un insieme di strumenti che ne facilitano la personalizzazione e l'adesione a problematiche specifiche. Il modo in cui i dati sono archiviati nel GIS è profondamente diverso da quello in cui lo sono sulla carta. I dati nel GIS, siano essi punti, linee o aree, sono descritti da numeri ed archiviati sottoforma di coordinate x, y attraverso le quali il sistema è in grado di manipolarne la geometria e le relazioni tra essi esistenti per creare nuove informazioni.

Con il GIS non si è limitati alle informazioni che la carta mostra, ma è possibile associare ai dati geografici informazioni descrittive denominate *attributi*. Gli attributi possono comprendere un'ampia gamma di informazioni che sono organizzate in tabelle : ogni colonna o campo contiene i valori o singole caratteristiche dell'attributo ed ogni riga o record rappresenta i valori di tutti gli attributi relativi ad un singolo elemento geografico. La creazione, organizzazione e gestione delle tabelle è compito specializzato del sotto sistema del GIS di gestione dei dati. Il legame tra elementi geografici e attributi è la caratteristica base del funzionamento del GIS, questo legame è realizzato mediante un identificatore, ovvero un codice unico composto da numeri o da lettere, che viene assegnato contemporaneamente all'elemento geografico e al record che identifica l'insieme dei suoi attributi nella tabella.

La tabella degli attributi contiene in genere un campo denominato ID (o CAT) in cui sono archiviati tutti gli identificatori e l'esistenza di un tale legame permette di richiamare gli attributi a partire dai dati geografici e/o di accedere ai dati geografici partendo dagli attributi. Questo legame consente al GIS di visualizzare le carte in modo dinamico, sulla base di qualunque attributo prescelto, inoltre, lo stesso elemento geografico può essere rappresentato con simboli e grafiche differenti a seconda dell'attributo selezionato.

In sostanza il GIS può individuare elementi geografici selezionando qualunque tipo e numero di attributi e può renderli visibili in forma di carta grazie al legame tra essi instaurato.

1.3 Modello e tipologia dei dati geografici

In un GIS tutti gli oggetti presenti sulla superficie terrestre sono rappresentati mediante tre caratteristiche essenziali: la geometria, la topologia e gli attributi. La geometria riproduce la forma degli oggetti e viene ricondotta a tre elementi di base: punto, linea (o arco) e poligono (o area).

Un punto viene utilizzato per riprodurre elementi puntiformi, come ad esempio un punto quotato, un pozzo, o la posizione di una stazione meteorologica.

La linea definisce elementi a sviluppo lineare come una strada, una linea elettrica o un corso d'acqua.

Il poligono definisce aree chiuse, come un edificio, un lago o un affioramento geologico.

La topologia è l'insieme delle informazioni che riguardano le mutue relazioni spaziali tra i diversi elementi come la connessione, l'adiacenza o l'inclusione. Ad esempio viene specificato se un arco è connesso ad un altro, se è comune a due poligoni adiacenti, o se un poligono è completamente racchiuso all'interno di un altro.

Gli attributi rappresentano i dati descrittivi dei singoli oggetti reali. Per un elemento puntiforme rappresentante un pozzo, gli attributi possono ad esempio essere costituiti dalla profondità, l'anno di perforazione e il proprietario, per una stazione meteorologica la temperatura dell'aria o le precipitazioni, per una strada la larghezza, la categoria o il tipo di pavimentazione.

Il GIS prevede la gestione di queste informazioni in un database relazionale, ma l'aspetto che lo caratterizza è quello geometrico, esso memorizza la posizione del dato impiegando un sistema di proiezione reale che definisce la posizione geografica dell'oggetto e per farlo gestisce contemporaneamente i dati provenienti da diversi sistemi di proiezione e riferimento (es. UTM o Gauss Boaga). Quindi, a differenza della cartografia classica, la scala in un GIS più che parametro di visualizzazione è un parametro di qualità del dato.

La complessità della realtà ha bisogno di essere semplificata per poter essere interpretata e rappresentata nel GIS e questa semplificazione costituisce

il modello dei dati. Nei GIS possono essere utilizzate due diverse tecniche di rappresentazione dei dati: vettoriale e raster.

Il dato vettoriale : ha come elementi fondamentali di rappresentazione i punti, le linee e i poligoni, un punto è definito da una coppia di coordinate mentre una linea o un poligono dalle coordinate di un insieme di punti che quando connessi fra loro con segmenti retti, formano la rappresentazione grafica dell'oggetto (Fig. 1.1).

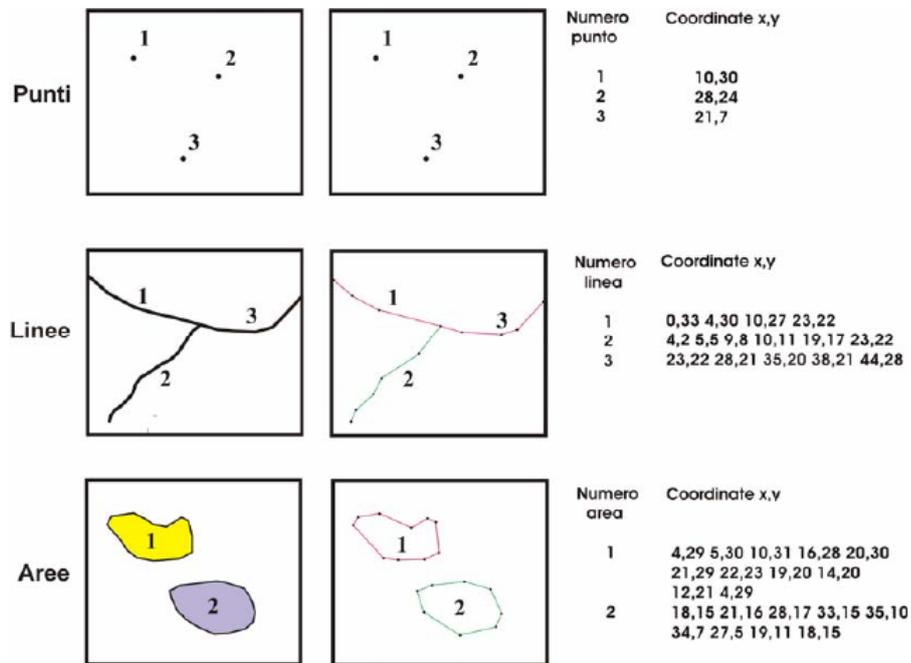


Figura 1.1: Rappresentazione vettoriale di punti, linee (archi) e aree (poligoni)

Un oggetto puntuale come un pozzo, un foro di sonda, una sorgente, può essere rappresentato da una coppia di coordinate (x, y), elementi lineari, come strade, fiumi, confini, sono descritti da una sequenza di punti, quindi coordinate (x, y), elementi areali, come bacini idrici, entità geografiche, parchi naturali, zone di vendita, sono definiti da poligoni, cioè da una sequenza di coordinate di punti che ritorna sul punto di origine. Tipici dati memorizzati in formato vettoriale sono quelli che provengono dalla digitalizzazione ma-

nuale di mappe, dai rilievi topografici con strumenti di campagna, dai CAD, dai GPS. I dati vettoriali GIS sono sempre accompagnati da informazioni topologiche.

Nella rappresentazione vettoriale Generalmente i due punti alle estremità di una linea vengono definiti nodi, i punti intermedi di una spezzata vengono definiti vertici.

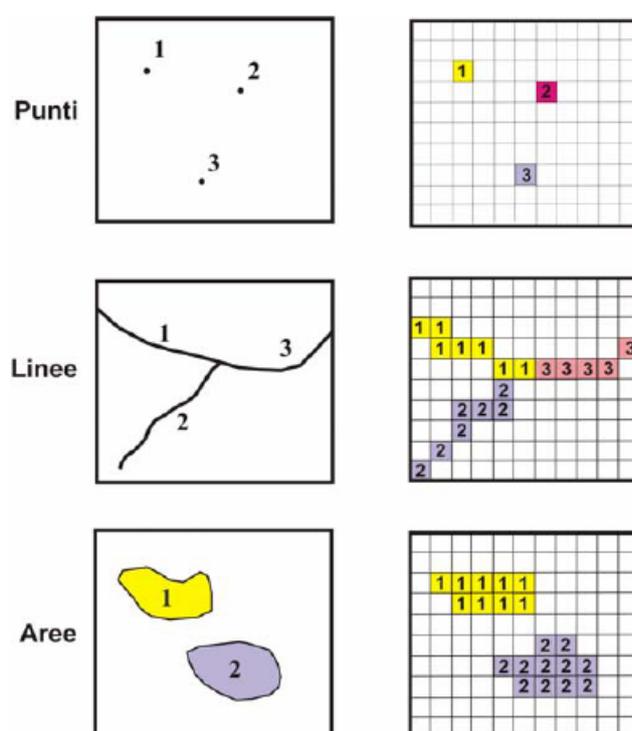


Figura 1.2: Rappresentazione raster di punti, linee (archi) e aree (poligoni)

Il dato raster : il termine raster indica un'immagine costituita da punti o pixel, la cui densità viene detta risoluzione ed è espressa in pixel/pollice, pixel/centimetro o nel mondo GIS in pixel/metro. Con i dati raster, il territorio viene riprodotto attraverso una matrice di pixel di forma quadrata o rettangolare, e a ciascuno è associato un attributo che definisce le caratteristiche e il valore dell'elemento rappresentato (Fig. 1.2). La dimensione del pixel è inversamente proporzionale alla precisione della carta mentre il numero è direttamente proporzionale. Tipici dati raster sono quelli generati

dagli scanner e dai programmi di interpretazione di immagini come quelli utilizzati per le immagini da satellite. In tale caso gli standard più comuni sono TIFF, RLC, LAN, BIP, GRASS e GRID, spesso utilizzati in diversi campi di applicazione. La qualità di un raster è determinata da tre fattori: la risoluzione, la compressione e la registrazione. La risoluzione dipende dalla fonte dei dati; ad esempio le immagini ottenute con il satellite per il telerilevamento LANDSAT, possono avere una risoluzione al suolo fino a 30 x 30 m mentre quelle ottenute dal satellite SPOT hanno una risoluzione fino a 10 x 10 m. Questi sono solo due esempi, dato che la risoluzione dei dati raster può variare moltissimo. Nel caso di immagini raster ottenute da scanner parleremo infatti di risoluzioni nell'ordine di centinaia di dpi (dots per inch) e quindi di risoluzione molto alta. Nel caso di griglie costruite appositamente per l'analisi territoriale, ad esempio contenenti dati relativi alla temperatura al suolo, la risoluzione dipenderà dalla metodologia scelta per la raccolta dei dati e dalla ampiezza della area di studio. Tipicamente in questi casi si parla di risoluzione dell'ordine delle decine di metri al suolo.

Per compressione si intende la capacità di comprimere i dati raster (generalmente essi richiedono da cento a mille volte più spazio di quelli vettoriali) per renderli più maneggevoli. Ad un primo livello, una tecnica di compressione consiste nel memorizzare il numero dei pixel uguali invece che memorizzarli in sequenza. Ad esempio, considerata una figura in bianco (B) e nero (N), invece che rappresentare la prima riga di un foglio con BBBBNNNBNNNNBB, la si memorizza come 3B3N1B4N2B. Il vantaggio ottenuto nella memorizzazione viene chiaramente compensato dai tempi di compressione e decompressione del dato che deve essere sempre decompresso prima dell'utilizzo.

Per registrazione si intendono le tecniche necessarie a georeferenziare e raddrizzare le immagini raster. Infatti le foto aeree e le immagini da satellite, oltre a dover essere posizionate correttamente (facendo collimare le coordinate dei punti noti a terra con quelle degli oggetti presenti nell'immagine), devono anche essere ortogonalizzate cioè ricalcolate tenendo conto dell'angolo da cui sono state prese.

Nel modello dei dati di un GIS, oggi i dati vettoriali e i dati raster coesisto-

no, si integrano a vicenda e sono generalmente usati i primi per dati discreti e i secondi per dati continui (ad esempio rete viaria vettoriale derivata dalla cartografia per i primi e umidità al suolo derivata da immagini da satellite per i secondi). Inoltre sono disponibili programmi in grado di convertire in modo più o meno automatico dati raster in vettoriali e viceversa. Entrambi i tipi di dati possono essere associati ad attributi: per i primi saranno legati alle primitive grafiche e agli oggetti, per i secondi ai singoli pixel.

In un GIS le diverse categorie di oggetti presenti sulla superficie terrestre sono distinti in elaborati diversi, o carte. Ogni elaborato contiene quindi una diversa caratteristica o tematismo, come l'idrografia, l'altimetria, ecc. (Fig. 1.3). E' tuttavia possibile l'inserimento di diversi tematismi in una stessa carta suddividendoli in layer distinti.

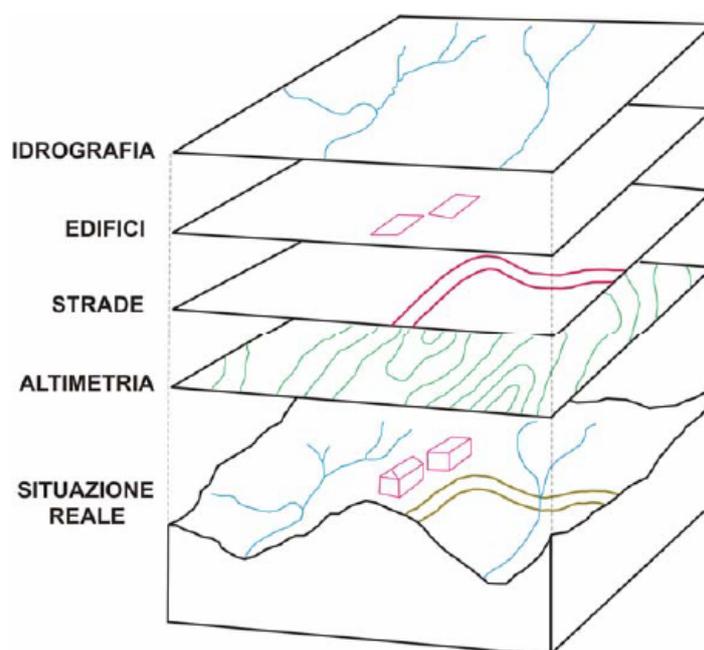


Figura 1.3: Rappresentazione del territorio mediante tematismi distinti, passibili di elaborazione tramite un GIS

I dati in forma raster occupano generalmente più memoria dei dati in forma vettoriale in quanto ad ogni cella viene assegnato un attributo, anche se

esistono tecniche di compattazione dei dati che limitano questo inconveniente. Il vantaggio è che lo spazio geografico risulta uniformemente definito in un modo semplice e prevedibile. In tal modo i sistemi raster hanno generalmente più potenza analitica dei sistemi vettoriali nell'analisi dello spazio continuo e sono pertanto adatti all'analisi di dati che presentano una continua variabilità nello spazio, come le temperature, le precipitazioni, l'altimetria ecc. Questi sistemi trovano quindi un'applicazione ottimale nella valutazione di problemi che includono numerose combinazioni matematiche di dati appartenenti a diversi tematismi. Sono quindi eccellenti nella valutazione di modelli ambientali. Infine, dal momento che le immagini da satellite impiegano una struttura raster, la maggior parte di questi sistemi può facilmente incorporare ed elaborare dati di questo tipo. I dati vettoriali sono particolarmente adatti all'analisi degli elementi lineari, come ad esempio all'analisi dei reticoli, come quello stradale o idrografico.

Un'altra importante caratteristica del modello dati di un GIS è la capacità di gestire oggetti tridimensionali: non stiamo parlando di un attributo che indichi l'altezza di un oggetto, ma di un sistema a tre coordinate reali. I modelli tridimensionali sono generati in due modi: quando si dispone di un insieme sparso di elementi quotati, si utilizza generalmente un algoritmo che crea un TIN (Triangulated Irregular Network), costruendo una rete di triangoli i cui vertici sono costituiti dai punti di cui si conoscono le tre coordinate; se si dispone invece di un insieme di punti quotati ordinati in griglie a passo regolare è possibile generare un DTM (Digital Terrain Model) o DEM (Digital Elevation Model). A partire da un TIN, un DTM o un DEM è possibile interpolare curve di livello, effettuare analisi di visibilità, generare profili longitudinali, effettuare analisi di pendenza e di esposizione, clivometrie, generare viste 3D, ecc. Ma soprattutto è possibile calcolare lunghezze reali e non ridotte all'orizzonte, dato che elementi come ad esempio le strade, hanno un andamento altimetrico, non sono "piatte".

In ogni caso l'elemento più importante del modello dati di un GIS rimangono sempre gli attributi. Infatti una applicazione per cartografia ha l'obiettivo principale di riprodurre su carta delle cartografie, mentre un GIS ha il suo obiettivo principale nell'analisi dei dati, per diventare uno strumento

di supporto alle decisioni.

L'utente di un GIS non ha solo bisogno di restituire una carta delle zone edificate, quanto di rappresentare un tematismo, ad esempio, retinarla in funzione dell'età media della popolazione residente. Per ottenere questo potrebbe interrogare una banca dati di tipo relazionale: per esempio, ad ogni edificio potrebbe essere associato un indirizzo, mentre in un altro data base (quello anagrafico) ad ogni indirizzo potrebbe essere associata la data di nascita delle persone che vi risiedono.

Tramite quindi l'informazione "Indirizzo" si potrebbe creare una relazione tra i due data base ed effettuare un'analisi relativa all'età media restituendola poi graficamente su carta.

Dalla stessa analisi si potrebbe poi derivare un nuovo tematismo relativo all'anzianità dei residenti e utilizzandolo in sovrapposizione al tematismo relativo alla carta dei bacini d'utenza dei centri anziani del Comune o dei servizi materno-infantili dell'Unità Sanitaria Locale, verificare la congruenza tra bacini d'utenza dei servizi (offerta) e potenziale utenza (domanda) per poi decidere eventualmente un nuovo piano d'azione o di sviluppo. Gli attributi che possono risiedere anche su più sistemi ed essere aggiornati da molti applicativi (nel caso precedente sono aggiornati dall'ufficio anagrafe e utilizzati da quello di pianificazione dei servizi), sono in genere memorizzati su dei data base relazionali ed interrogabili mediante linguaggi di tipo SQL (Structured Query Language). Una struttura relazionale del data base accuratamente progettata permette di effettuare diverse analisi sui dati senza essere costretti in percorsi obbligati. In altre parole, oggi possono essere estratte le date di nascita, domani si potrà condurre un'analisi sulla numerosità dei nuclei familiari e quindi una simulazione del carico della rete fognaria o della domanda di servizi scolastici.

La capacità di integrare, nel modello dati di un GIS, attributi provenienti da diversi data base anche raggiunti in rete locale o geografica mediante strutture relazionali e secondo formalismi SQL è oggi una necessità imprescindibile.

1.4 Algoritmi di elaborazione

La disponibilità di algoritmi di elaborazione dei dati è un parametro molto importante per valutare le potenzialità di uno strumento GIS. Gli algoritmi di overlay mapping, cioè di sovrapposizione ed integrazione di più strati informativi, sono presenti per entrambi i formati.

Nel formato raster sono più semplici dal punto di vista matematico e più intuitivi dal punto di vista utente e permettono infinite combinazioni di strati. Per il formato vettoriale sono più complesse in tutti i sensi, ma sicuramente più precise poiché elaborano dati sul continuo e non sul discreto.

Per gli algoritmi di buffering, cioè di generazione di aree di rispetto intorno ad elementi geografici, e per la famiglia di algoritmi per l'elaborazione di modelli digitali del terreno, valgono le stesse considerazioni sopra descritte.

Invece alcuni algoritmi sono disponibili solo sulla tipologia di formato che facilita l'approccio al problema.

Gli algoritmi per l'analisi di intorno o per la ricerca di percorsi di minimo costo sono tipici del formato raster, mentre gli algoritmi per l'allocazione di risorse, per la ricerca dei percorsi ottimali, di segmentazione dinamica, sono tipici del formato vettoriale.

In sintesi i due formati si distinguono proprio in base alla disponibilità di algoritmi di elaborazione che costituiscono il fulcro decisivo per la comparazione delle due tipologie di dati. Le necessità applicative richiedono tutte le funzionalità possibili di analisi senza distinzione e questo determina la naturale complementarità dei due formati e la necessità di generare delle sinergie.

Le funzionalità richieste affinché i due formati possano essere considerati in una logica integrata sono:

- facilità nel passaggio dei dati dal formato vettoriale al formato raster e viceversa;
- interfaccia utente studiata per rendere intuitivo l'uso integrato dei due formati;
- strutture omogenee di archivi di dati raster e vettoriali;

- gestione dei dati descrittivi associati alle due tipologie di formato attraverso l'uso di un comune data base relazionale;
- compatibilità del sistema di georeferenziazione dei due formati;
- possibilità di visualizzare ed interrogare simultaneamente e contemporaneamente dati raster e vettoriali;
- tecniche di elaborazione integrata.

L'insieme delle funzionalità sopra descritte rappresenta l'indirizzo attuale degli strumenti GIS più avanzati presenti sul mercato.

Negli ultimi anni da funzioni di conversione di dati da formato raster a vettoriale si è passati ad una integrazione orientata a visualizzare ed editare insieme le due tipologie di dati.

Le prossime frontiere sono verso interfacce utente sempre più indirizzate a rendere disponibili funzionalità e processi applicativi senza limiti di formato.

L'utente deve poter usufruire di tutte le potenzialità di elaborazione offerte dalle due tipologie, senza peraltro dover essere vincolato dai formati stessi. In conclusione la complementarietà dei due formati rende necessaria una sempre più completa integrazione all'interno di strumenti GIS orientati allo sviluppo di applicazioni territoriali.

1.5 Volume dei dati

La tecnologia hardware sta spingendo sempre più verso una riduzione sia delle dimensioni fisiche della memoria di massa che dei relativi costi. Nonostante questo, il volume dei dati geografici sia ancora un problema da non sottovalutare.

E' generalizzata l'idea che i dati in formato raster occupino molta memoria di massa. Infatti l'occupazione di una copertura raster è data dal numero di righe per il numero di colonne per il numero di bit necessari a contenere l'informazione descrittiva. Naturalmente per certe tipologie di dati raster (tematismi o cartografie) è possibile usare metodi di compressione che riducono di un certo fattore la quantità di memoria occupata.

Nonostante questo, quando parliamo di archivi di banche dati geografiche in formato raster, le quantità di memoria richieste sono sempre dell'ordine dei Gigabyte.

Il formato vettoriale è invece noto per essere più economico dal punto di vista del volume di dati. In questo caso però va posta molta attenzione alla risoluzione geografica richiesta, in quanto un dato ad alta risoluzione, acquisito per esempio vettorializzando da una mappa raster, può occupare una buona quantità di memoria.

In sintesi il formato raster può essere senz'altro considerato il formato che necessita di maggiore quantità di memoria di massa, tenendo presente comunque che in taluni casi anche il dato vettoriale può richiedere quantità al di sopra della media.

1.6 I parametri e gli standard dei dati GIS

Il crescente numero di Sistemi Informativi Territoriali ha generato una crescita complementare di dati cartografici.

Questi ultimi, contrariamente alla cartografia tradizionale, realizzata per un'utenza decisamente vasta e diversificata, sono stati acquisiti, nella maggior parte dei casi, ad uso e consumo del solo utente richiedente.

Tale proliferazione di banche dati, se pur acquisite con metodologie differenti, ha generato duplicazioni e ripetizioni.

D'altro canto l'informatizzazione della cartografia è una modalità talmente nuova che per essa non sono ancora stati raggiunti e standardizzati criteri omogenei di produzione e di controllo di qualità, a differenza della cartografia tradizionale che ha da tempo degli standard riconosciuti a livello internazionale.

In questi ultimi anni, comunque, studi approfonditi hanno definito gli aspetti generali da considerare per la definizione degli standard per la cartografia numerica e si sta lavorando alacremente, soprattutto sulla base delle esperienze già effettuate, alla individuazione e alla determinazione di tutti i parametri necessari.

La letteratura, in materia di standard di cartografia numerica, individua tre aspetti essenziali: contenuti metrici e semantici dei dati, struttura e formato dei dati, profili di qualità.

Un parametro fondamentale per definire il contenuto metrico di un dato di cartografia numerica è la scala nominale.

Questo è forse l'unico parametro che permette oggi un minimo di conoscenza sull'aspetto metrico del dato.

Si conviene infatti che un prodotto di cartografia numerica, realizzato con certi requisiti di precisione metrica, abbia come scala nominale la scala di una cartografia tradizionale realizzata seguendo gli stessi requisiti di precisione metrica.

Ma, a parte questo caso, si riscontrano notevoli difficoltà a trovare dei parametri standardizzati. Il problema è generato essenzialmente dalle due logiche che distinguono i dati nati per sistemi di cartografia numerica derivati dal CAD, da quelli realizzati per sistemi GIS: i primi, dedicati essenzialmente alla restituzione, seguono maggiormente le logiche della cartografia tradizionale rappresentando il dato grafico in modo geometrico e il dato descrittivo in modo gerarchico-numerativo; i secondi, orientati all'analisi, rappresentano la parte grafica del dato in modo topologico e strettamente connessa alla sua descrizione attraverso tecniche relazionali.

Oltre che una maggiore chiarezza sulla differenza tra queste due logiche, è necessario introdurre alcuni parametri di qualità.

Per caratterizzare la qualità globale (dell'intero prodotto) si definiscono: l'esaustività come misura dell'eccesso o della mancanza di informazioni, l'attualità, come percentuale di cambiamenti intervenuti tra la data della rilevazione e quella attuale, la genealogia come insieme di informazioni sulle fonti e sui trattamenti.

Per la qualità locale (relativa cioè a ciascun singolo oggetto cartografico) si definiscono la precisione metrica, definita come differenza della posizione di un punto sulla cartografia rispetto alla sua posizione reale nel sistema di riferimento cartografico utilizzato, la risoluzione, data dalle dimensioni del più piccolo particolare rappresentato, la precisione semantica, definita come corrispondenza tra la realtà e l'attributo qualitativo associato all'oggetto, la

coerenza o congruenza logica, definita come assenza di incongruenze rilevabili indipendentemente da una verifica nella realtà (un edificio su una strada, due curve di livello che si intersecano, ecc.) e in ultimo la congruenza geometrica, cioè l'assenza di errori di forma e posizione che non possono essere rilevati senza una verifica diretta sul campo.

Questi parametri sono ancora oggi scarsamente utilizzati per descrivere i dati.

Generalmente ci si limita ad una descrizione complessiva del contenuto geometrico e degli attributi associati, alla scala nominale, alla fonte e alla precisione metrica.

L'accuratezza delle rappresentazioni digitali di dati spaziali dipende dai requisiti imposti dall'utente ed è legata alla sorgente dei dati (cartografia, immagini, ecc.) e alle metodologie e strumenti usati per acquisirli.

I dati vettoriali vengono generalmente acquisiti attraverso rilievi aerei o direttamente da cartografia. In tal caso l'accuratezza è fortemente relazionata alle regole di interpretazione del dato sorgente e alla precisione in fase di digitalizzazione dell'informazione.

Regole poco precise comportano eccessiva discrezionalità da parte dell'operatore in fase di interpretazione; anche la scelta di metodologie e di strumenti di acquisizione appropriati al tipo di informazione è di fondamentale importanza per assicurare una buona accuratezza del dato.

Sicuramente l'accuratezza è assicurata comunque da metodologie di controllo di qualità del dato molto onerose, ma necessarie per questa tipologia di formato. D'altro canto l'accuratezza ottenibile va a favore delle tecniche di elaborazione dei dati che sfruttano la continuità del formato vettoriale. L'accuratezza del dato raster dipende invece, nella maggior parte dei casi, dagli strumenti di acquisizione.

Per esempio, per le immagini telerilevate, l'accuratezza è determinata dalle piattaforme satellitari ed aeree. Le oscillazioni dei mezzi che trasportano i sensori provocano deformazioni delle immagini che possono essere corrette con più o meno precisione in fasi successive di elaborazione del dato. Anche l'informazione radiometrica influenza l'accuratezza del dato.

Per quanto riguarda l'accuratezza dei dati di tipo cartografico, acquisiti

con strumenti di scansione (scanner), essa è legata essenzialmente alla precisione strumentale senza peraltro sottovalutare l'influenza determinata dalle tecniche software di discretizzazione.

In sintesi possiamo considerare l'accuratezza dei dati raster più sistematica e meno discrezionale, pur non sottovalutando la difficoltà nel controllo dell'errore presente per entrambi i formati.

La risoluzione del dato spaziale descrive la relazione tra la distanza come misura nel reale e la distanza come misura della sua rappresentazione digitale o su carta, è una caratteristica legata al fattore di scala cartografica anche se non può essere direttamente relazionata alla dimensione fisica di ogni elemento grafico.

In pratica, mentre la dimensione del lato della cella elementare nel formato raster definisce la risoluzione del dato (cella di 10 x 10 metri = risoluzione di 10 metri), nel formato vettoriale la risoluzione può essere associata alla distanza minima fra due vertici di una spezzata pur essendo comunque fortemente influenzata dalla scala di rappresentazione cartografica del documento sorgente.

Questo metodo può essere opinabile in quanto un lungo tratto rettilineo può essere acquisito esclusivamente con i due vertici estremi senza per questo ridurre la precisione.

Invece, con tecniche di scansione e vettorializzazione automatica, tale distanza, come parametro di risoluzione, può avere maggior senso. Infatti, in tal caso, il dato vettoriale è ricavato da un formato raster, pertanto la distanza fra i nodi di una spezzata è legata alla risoluzione del dato raster.

Per quanto riguarda i dati rilevati direttamente sul terreno con tecniche topografiche, la precisione è legata agli strumenti utilizzati ed alle tecniche di rilievo.

In sintesi, poichè il parametro di risoluzione è così diversificato nei due formati, deve essere necessariamente comparato insieme agli altri fattori indicati in questo capitolo.

1.7 Proiezioni e Sistemi di Riferimento

Caratteristica fondamentale di un GIS è la sua capacità di georeferenziare i dati; ovvero di attribuire ad ogni elemento le sue coordinate spaziali reali. In altre parole, le coordinate di un oggetto non sono memorizzate relativamente ad un sistema di riferimento arbitrario (ad esempio 12 centimetri dal bordo inferiore e 5 da quello sinistro di una mappa) né relativamente al sistema di coordinate della periferica usata, come la tavoletta digitalizzatrice o il video, ma sono memorizzate secondo le coordinate del sistema di riferimento in cui realmente è situato l'oggetto (come $121^{\circ} 27'$ lat. E e $41^{\circ} 53'$ long. N utilizzando il sistema geografico) e nelle reali dimensioni, non in scala. La scala di rappresentazione diventa a questo punto solamente un parametro per definire il grado di accuratezza e la risoluzione delle informazioni grafiche, e quindi utilizzabili ad esempio per definire la densità di rappresentazione: ad una scala minore, elementi come gli edifici non vengono rappresentati e compaiono gli isolati o le aree urbanizzate.

L'aver introdotto il concetto di sistema di riferimento porta immediatamente ad accennare l'uso delle proiezioni cartografiche. Mediante le proiezioni o rappresentazioni siamo in grado di rappresentare la superficie approssimativamente sferica della Terra su di un piano pur mantenendo alcune proprietà geometriche quali l'isogonia, l'equivalenza o l'equidistanza; tra le più utilizzate, oltre alla rappresentazione geografica (che in effetti non è una proiezione ma solo un sistema di riferimento) vi sono l'U.T.M., la Gauss-Boaga, la Lambert (queste utilizzate anche in Italia) oltre alla conica, polare, stereografica e diverse altre.

Le carte geografiche sono piatte ma le superfici che esse rappresentano sul globo sono curve. Il concetto di proiezione è, quindi, richiesto per rappresentare uno spazio tridimensionale su una carta a due dimensioni. Una carta geografica accurata dovrebbe indicare in legenda il nome ed i parametri del sistema di proiezione adottato.

Durante il processo di proiezione dei dati reali su un foglio di carta sono introdotti inevitabilmente degli errori. Anche i più accurati sistemi di proiezione comportano distorsioni di almeno una delle caratteristiche geografiche:

forma, area, direzione, distanza.

Una distinzione tra i diversi sistemi di proiezione viene operata proprio sulla base delle caratteristiche del mondo reale rispettate: le proiezioni equivalenti preservano le aree, le proiezioni conformi gli angoli, quelle equidistanti le distanze tra punti determinati. Ne consegue che non esiste un sistema di proiezione preferibile in assoluto e che l'adozione di un sistema piuttosto che un altro dipende dall'uso cui è destinata la cartografia e dalla zona da rappresentare. Il primo passo nel processo di proiezione è quello di individuare uno o più punti di contatto tra la terra e la superficie di proiezione, orientando quest'ultima in maniera secante o tangente rispetto al globo. La localizzazione del punto o della linea di contatto tra le due superfici è importante poiché si tratta del punto o della linea a distorsione zero. In generale, le distorsioni aumentano con la distanza dalla zona di contatto.

Per ogni proiezione viene definito anche un sistema di riferimento, utilizzato per il calcolo delle coordinate. Ad esempio nel sistema UTM si utilizzano spicchi predeterminati ampi sei gradi in latitudine detti fusi con un sistema di coordinate ortogonali all'interno di ogni fuso (l'Italia è a cavallo dei fusi 32, 33 e 34); nella Gauss-Boaga, il riferimento è il meridiano passante per Monte Mario (a Roma) e vengono utilizzate coordinate chilometriche misurate convenzionalmente partendo da 1500 a sinistra e da 2520 a destra del meridiano di riferimento.

La maggior parte delle carte geografiche riporta l'indicazione del sistema di coordinate adottato. Un sistema di coordinate è, in pratica, un sistema di misurazione degli elementi su una superficie, esistono fondamentalmente due standard: un sistema di riferimento basato su una superficie piana ed un altro basato su una sfera.

Il sistema di coordinate rettangolare, o cartesiano (spesso usato per individuare localizzazioni nella geometria piana) utilizza come riferimento due assi ortogonali (x e y), la cui intersezione costituisce l'origine del sistema e la localizzazione $x=0$, $y=0$. Tutte le posizioni sul piano sono individuate da due valori (positivi o negativi), che specificano rispettivamente la posizione orizzontale e verticale con riferimento all'origine del sistema. Questi due valori sono indicati come coordinata x e coordinata y .

Le coordinate sferiche sono misurate in latitudine e longitudine: considerando la terra come una sfera, latitudine e longitudine sono angoli misurati dal centro della terra ad un punto sulla superficie. Latitudine e longitudine sono tradizionalmente misurati in gradi. In latitudine: 01 indica l'Equatore, 90· il Polo Nord e -90· il Polo Sud. In longitudine: 01 indica il Primo Meridiano, che parte dal Polo Nord, passa per Greenwich (in Inghilterra) e termina al Polo Sud. La longitudine è misurata positivamente fino a 1801, spostandosi da Greenwich verso est, e negativamente in caso contrario.

Come si può immaginare leggendo gli esempi sopra riportati, la necessità di rispondere a diverse esigenze cartografiche ha reso necessario lo sviluppo di programmi di conversione da un sistema di coordinate ad un altro, disponibili in un numero sempre maggiore.

1.8 Panoramica delle funzioni GIS

1.8.1 Overlay

Mediante l'analisi spaziale, partendo dalle informazioni esistenti in un database geografico, possono essere creati nuovi livelli informativi, associando i dati in maniera da identificare relazioni prima non chiaramente visibili. Tipico esempio sono le analisi effettuate con la semplice sovrapposizione di più livelli informativi (overlay) come l'uso del suolo ed i dati catastali, allo scopo di identificare, per ogni proprietà (particella catastale), la tipologia di uso del suolo. Probabilmente le funzioni di overlay sono state le prime ad essere implementate in un GIS, e rimangono ancora oggi le funzioni di base in questi sistemi. Concettualmente si tratta di funzionalità molto semplici ma solamente una struttura dei dati completamente topologica permette di realizzarle in modo efficace. Infatti il risultato della sovrapposizione di diversi livelli informativi non deve essere solamente visuale, cioè si possono vedere gli elementi sovrapposti ma devono soprattutto essere a livello degli attributi, che devono essere riportati da un livello informativo all'altro, in corrispondenza degli elementi corrispondenti.

Le sovrapposizioni (overlay) possono essere suddivise in tre categorie prin-

cipali: punti su poligoni, linee su poligoni e poligoni su poligoni: comune a tutte le operazioni di overlay è che almeno uno dei due tematismi considerati sia di tipo poligonale. Se ad esempio poniamo, come livello poligonale, una carta dell'uso del suolo, mediante le funzioni di overlay possiamo attribuire la tipologia del suolo in cui cade ad esempio un traliccio elettrico (punto su poligono), un tratto di strada (linea su poligono), un fabbricato (poligono su poligono). Il nuovo livello informativo conterrà, oltre che tutti gli attributi rispettivamente del traliccio, della strada o del fabbricato, anche tutte le informazioni che sono associate alla carta poligonale dell'uso del suolo. Mediante questi operatori sono possibili sofisticate analisi di tipo ambientale e scientifico: si pensi, in campo geologico, alla possibilità di sovrapporre carte diverse riportanti informazioni sulla geologia, la copertura vegetale, l'acclività, la fratturazione della roccia, per determinare le zone potenzialmente predisposte al dissesto; tali analisi, tipiche delle scienze ambientali, risultano grandemente facilitate dalla presenza degli operatori di overlay.

1.8.2 Buffering

Spesso si devono determinare delle aree di rispetto intorno a specifici elementi geografici, le linee elettriche ad alta tensione, per esempio, prevedono delle fasce di rispetto di oltre 250 metri per parte; le leggi ambientaliste prevedono l'inedificabilità entro una specifica distanza dalle rive di corsi d'acqua, laghi e mare; un ponte ripetitore copre un'area con un determinato raggio.

La risposta a questi specifici problemi viene agevolmente data da un GIS che disponga delle funzioni di buffering, in grado cioè di creare un'area di rispetto intorno agli elementi geografici che sono presenti nel database. La possibilità di modulare questa operazione a seconda delle necessità dell'operatore dà modo di risolvere, con pochi passaggi, problemi altrimenti difficilmente risolvibili; ad esempio la capacità di effettuare buffering asimmetrici rispetto, ad esempio ai due lati di un elemento lineare oppure di effettuare un buffering parametrizzato a seconda delle caratteristiche dell'elemento: per creare fasce di rispetto intorno alle linee elettriche, si può usare come parametro discriminante la tensione di esercizio, creando fasce maggiori per

tensioni maggiori della linea. Una volta creata la fascia di rispetto che sia intorno ad un punto, linea o poligono, il risultato è sempre un livello informativo di tipo poligonale, che può essere utilizzato per successive analisi; ad esempio, effettuando l'overlay della carta della vegetazione con le fasce di rispetto create intorno ad una strada in costruzione, può essere valutata la superficie e la tipologia della vegetazione coinvolta nel progetto.

1.8.3 Analisi di rete

Lo spostamento delle persone, il trasporto e la distribuzione di beni e servizi, la distribuzione dell'energia, le comunicazioni: tutte queste attività prevedono lo spostamento di materia o di informazioni mediante dei sistemi di reti, che sempre più costituiscono una delicata e vitale infrastruttura del mondo di oggi.

La forma, la capacità e l'efficienza delle reti ha quindi un impatto non trascurabile sull'attuale standard di vita e comunque influenza la nostra visione del mondo.

Le funzioni principali che possono essere effettuate mediante un GIS sulle reti, di qualsiasi tipo esse siano (trasporto, distribuzione, comunicazione, ecc.) sono fondamentalmente la ricerca del minimo percorso su una rete o comunque del percorso meno costoso; l'allocazione di porzioni della rete ad un fornitore o consumatore di risorse; la verifica delle connettività tra due punti della rete. Le funzioni di analisi di rete offrono la soluzione ad un problema di base: determinare il percorso minimo o comunque il più efficace per attraversare una rete passando per determinate località di questa. Il costo può essere determinato utilizzando un qualsiasi attributo presente sugli elementi costituenti la rete (tipicamente, una rete a grafo è rappresentata mediante archi connessi tramite nodi; questa può anche essere la rappresentazione di una rete stradale, simboleggiata con gli archi che indicano l'asse stradale ed i nodi che indicano gli incroci) che possa essere quantificato in maniera numerica: tipico esempio la distanza o il tempo di percorrenza lungo un arco del grafo può essere utilizzato per identificare il minimo percorso in termini di distanza o di tempo, ad esempio, lungo una rete viaria. Allocazione di

risorse: per effettuare queste analisi bisogna disporre di un centro che offra risorse e di una domanda di queste sulla rete, o viceversa, di una offerta sulla rete e di una capacità ricettiva su di un centro.

A titolo di esemplificazione, mediante l'uso delle funzioni di allocazione, secondo questa ultima modalità può essere affrontato e risolto un problema tipico della gestione pubblica di una città, ovvero la raccolta dei rifiuti solidi urbani.

Questa analisi viene effettuata rappresentando una eventuale discarica come centro di raccolta, e riportando la distribuzione dei cassonetti sul grafo che rappresenta la rete viaria cittadina: in questo modo si può determinare il numero massimo di cassonetti che, svuotati mediante i mezzi di raccolta, determinano la saturazione della discarica, ed anche identificare quali parti della rete stradale contribuiscono ad alimentare la discarica fino alla sua saturazione.

Operazioni analoghe possono essere effettuate per allocare studenti su scuole oppure determinare piani di evacuazione in caso di calamità, allocando la popolazione sulle strutture di ricovero, ecc.

Connettività: le funzioni per la verifica della connettività servono ad identificare se e quali porzioni di una rete sono connesse. Ad esempio, una società per la gestione delle reti elettriche può avere bisogno di sapere quale impianto presente sulla rete elettrica serve un determinato utente, e nel caso di interruzione o guasto quale percorso alternativo può essere adottato per mantenere la funzionalità del sistema. Analogamente un idrologo può utilizzare tali funzionalità per identificare tutti i rami ed i corsi d'acqua a monte di un determinato punto di un fiume.

1.8.4 La segmentazione dinamica

Gli elementi lineari, in una visione del mondo di tipo geografico, raffigurano oggetti come strade, fiumi o limiti amministrativi.

Utilizzando l'estensione al modello dati di tipo georelazionale chiamata segmentazione dinamica, è possibile rappresentare tali elementi lineari ed

associare delle informazioni, gli attributi, a qualsiasi porzione degli archi che rappresentano tali elementi.

La segmentazione dinamica è la capacità di associare diversi insiemi di attributi a qualsiasi segmento di un elemento geografico lineare senza dover cambiare la struttura fisica di questo: posso quindi attribuire informazioni diverse a porzioni diverse di un arco senza dover effettivamente spezzare fisicamente l'arco.

Infatti, la rappresentazione all'interno di un GIS di elementi geografici come un reticolo stradale o fluviale viene effettuata memorizzando una serie di coordinate X, Y in un dato sistema di riferimento. Ma un altro metodo per identificare la posizione di un elemento geografico potrebbe essere anche la sua posizione lungo una strada od un fiume (ad es. "al 12esimo chilometro dell'autostrada A14"). Questo metodo semplifica enormemente l'acquisizione dei dati, in quanto consente di memorizzare e gestire solamente una coordinata di posizione invece che due e di usare un sistema di riferimento più vicino alla realtà dell'utilizzatore finale.

Le funzionalità di segmentazione dinamica (nota: si tratta di funzionalità e non di modalità di memorizzazione di dati, in quanto la struttura dei dati rimane la medesima di quella utilizzata per memorizzare dati di tipo lineare ma viene estesa mediante opportune procedure software che implementano queste nuove potenzialità) permettono quindi di rappresentare e gestire in maniera estremamente efficace delle informazioni associabili ad elementi geografici lineari.

Ad esempio, i dati relativi ad un reticolo stradale, quali lo stato della pavimentazione, la frequenza degli incidenti, i limiti di velocità possono essere tutti associati al medesimo grafo che rappresenta la rete stradale.

Alla base della segmentazione dinamica vi è la definizione della route, che rappresenta un elemento lineare a cui possono essere associati degli attributi; la route può essere costituita da un insieme di archi, ma senza la costrizione che il suo inizio o fine debbano coincidere con il nodo iniziale o finale di un arco, cioè può comprendere anche parti di un arco.

In questo modo possono essere associati attributi a più archi oppure a parte di un arco, nel caso che la route sia definita solo su una porzione di

esso. La disponibilità di strumenti per creare, memorizzare, modificare, interrogare e rappresentare le routes completano le funzionalità di segmentazione dinamica.

1.8.5 Operatori raster e tridimensionali

La necessità di operare da un unico ambiente su diverse tipologie di dati hanno spinto l'evoluzione dei sistemi GIS verso dei sistemi integrati in cui fossero disponibili operatori in grado di effettuare analisi su dati bidimensionali e tridimensionali, oltre che raster.

Le funzionalità di visualizzazione dei dati raster in diversissimi formati hanno aperto la strada all'integrazione di dati telerilevati, anche da satellite, con i dati di tipo vettoriale; ma la successiva implementazione di una serie completa di operatori in grado di elaborare dati di tipo raster ha praticamente portato all'interno dei GIS funzionalità che erano prerogativa esclusiva dei sistemi di image processing.

Tali funzionalità, in grado di operare su dati di tipo matriciale, hanno aperto la strada ad una serie di applicazioni specifiche difficilmente realizzabili mediante l'approccio vettoriale: la modellazione idrogeologica, la ricerca di superfici di minimo costo (analoghe alla ricerca del minimo percorso, ma in tre dimensioni), l'interpolazione di dati puntuali per la generazione di superfici tridimensionali o la derivazione di curve di livello, sono tutte operazioni che sono state implementate in maniera straordinariamente efficace su strutture dati di tipo matriciale.

Mediante una sola specifica funzione ad esempio, è possibile derivare la struttura di un reticolo fluviale partendo da un modello digitale del terreno, e quindi codificare i vari rami del reticolo così identificati o calcolare l'esposizione di una superficie rispetto ad una sorgente luminosa, potendo modificare in tempo reale tutti i parametri relativi all'orientamento della sorgente e del modello tridimensionale sotto osservazione.

Alcune di queste funzionalità che operano su dati raster, trovano una controparte nella funzionalità di modellazione tridimensionale in cui troviamo operatori utili per effettuare calcoli di lunghezze vere (ad esempio, la

lunghezza vera di una strada considerando anche l'andamento altimetrico e non quella proiettata su di una superficie piana come nella rappresentazione cartografica bidimensionale), per determinare la reciproca visibilità di due punti oppure per effettuare profili morfologici o sezioni.

Le funzioni che determinano la visibilità di due punti, ad esempio, possono essere utilizzate per le operazioni di valutazione di impatto ambientale, per determinare se e da quali punti è visibile un manufatto in costruzione come una strada od un insediamento abitativo, identificando le possibili soluzioni ottimali.

La crescente disponibilità di informazioni sul territorio, tra cui quelle relative alla morfologia, rendono maggiormente fruibili le funzioni che operano sui dati tridimensionali aprendo la strada a delle analisi assai sofisticate che erano prima appannaggio di sistemi specializzati di sola modellazione tridimensionale.

Tutte le funzionalità sopra descritte, per quanto potenti possano essere, troverebbero una forte limitazione nel loro utilizzo se non ci fosse una facile ed adeguata modalità di interazione tra l'uomo e la macchina in grado di svolgerle.

Analogamente, il risultato di una analisi o di una modellazione spesso si ottiene effettuando le stesse operazioni in maniera iterativa su diversi set di dati oppure modificando dei parametri e reiterando l'analisi. La disponibilità di un linguaggio di programmazione e di adeguati strumenti di interazione con l'utente sono la risposta a queste esigenze. Mediante un linguaggio di programmazione di alto livello, dotato di strutture di controllo del flusso delle operazioni e di comandi per la realizzazione di maschere e menù, è possibile realizzare sofisticate interfacce che rendono agevole anche ad un operatore non specializzato l'accesso e la manipolazione dei dati.

La diffusione di standard internazionali per quanto riguarda le GUI (Graphical User Interface), come X-Windows, ha di fatto unificato le modalità di sviluppo delle interfacce rendendo agevole la loro realizzazione e la loro portabilità tra macchine anche di diversi costruttori.

Una ulteriore sofisticazione, di grande utilità in sede di definizione ed implementazione di una interfaccia è la disponibilità di editor grafici inte-

rattivi, mediante i quali possono essere costruiti dall'utente dei menù dotati di tutti i widgets tipici di queste GUI: cursori, bottoni di scelta, scroll-list ecc. sono alla portata anche del programmatore casuale che viene sgravato dalla necessità di conoscere comandi complessi e può quindi concentrarsi sulle problematiche dell'applicazione a cui sta lavorando.

1.9 Applicazioni presenti e future

Secondo il più recente studio di Teknibank (marzo 1993), le applicazioni GIS più diffuse in assoluto sono quelle di gestione di reti tecnologiche (29%), di pianificazione e gestione urbanistica (23%) e del territorio (22%).

In generale, le Regioni sono in prevalenza dotate di applicazioni per la gestione del territorio, mentre si intravede uno sforzo per le applicazioni di monitoraggio ambientale e per attività GIS in relazione alla pianificazione e gestione della rete dei trasporti.

Le Provincie sono in prevalenza dotate di applicazioni di pianificazione e omogenizzazione degli strumenti urbanistici e nel medio periodo dovrebbero orientarsi verso soluzioni per la pianificazione e gestione del territorio, incluso il controllo delle realizzazioni ambientali e le reti tecnologiche connesse ai trasporti.

I Comuni evidenziano il ruolo prioritario delle applicazioni di pianificazione e gestione urbanistica (piani regolatori comunali) e al secondo posto quelle connesse alle reti tecnologiche.

Le aziende di servizi pubblici, nazionali e locali, hanno per lo più realizzato applicazioni per interventi di progettazione, gestione e controllo di reti tecnologiche per l'erogazione di gas, elettricità, acqua e di monitoraggio ambientale.

In particolare i diversi settori che si avvalgono della tecnologia GIS possono essere raggruppati come segue:

- Ambiente: salvaguardia ambientale studi di impatto ambientale, localizzazione di aree, destinazioni di uso, gestione degli spazi verdi, defini-

zione di zone a rischio, raccolta, controllo, analisi dei dati riguardanti l'inquinamento.

- Difesa del suolo.
- Erosione potenziale dei suoli nell'area vesuviana .
- Progetto Campi Flegrei.
- Calcolo dell'apporto solido fluviale del bacino idrografico del Tusciano (SA) attraverso l'utilizzo del modello di Gavrilovic: teoria e applicazione tramite un sistema GIS.
- Gestione dei rifiuti .
- Progetto Casagiove.
- Produzione di servizi e studi sul territorio.
- Progetto LabUSIP
- Beni Culturali: l'informazione geografica è una componente essenziale delle problematiche legate alla gestione dei beni architettonici e paesaggistici, del patrimonio storico ed artistico, dei beni archeologici; le attività di tutela e di valorizzazione, di studio e di ricerca connesse a tali settori risultano tanto più efficaci se impostate secondo un processo dinamico, che abbracci la dimensione globale del contesto. Il GIS permette di indirizzare in modo efficace ed innovativo le strategie di gestione dei Beni Culturali nell'ambito della pianificazione territoriale.
- Emergenze sanitarie animali.
- Sorveglianza epidemiologica Blue Tongue in Campania: uso dei GIS per la creazione di mappe di rischio

Altri settori di utilizzo del GIS sono:

- Agricoltura: agro meteorologia, agronomia;

- Analisi e gestione di rischi: definizione dei rischi di incendi, terremoti, frane, valanghe, alluvioni, fenomeni di erosione e dissesto;
- Architettura ;
- Catasto: gestione piani urbanistici e territoriali ;
- Cartografia informatizzata;
- Demografia: evoluzione demografica, evoluzione occupazionale, espansione urbana e residenziale;
- Elettricità: localizzazione delle centrali ;
- Geologia ;
- Geomarketing;
- Gestione delle reti: funzioni necessarie alla gestione delle reti idriche, elettriche, del gas: cartografia, programmazione della manutenzione tecnica, simulazione di nuovi circuiti, localizzazione rapida di guasti e organizzazione degli interventi di ripristino dei servizi ;
- Prevenzione del crimine;
- Sanità;

Un ulteriore settore in cui l'utilizzo del GIS si è imposto, è la Pubblica Amministrazione:

L'informazione geografica è un riferimento essenziale per numerosi settori di primario interesse della Pubblica Amministrazione.

Per le Regioni: per la programmazione e pianificazione territoriale.

Per le Province: per i piani territoriali di coordinamento, finalizzati all'indicazione degli indirizzi generali di assetto del territorio: destinazione del suolo, localizzazioni infrastrutturali, linee di intervento per il consolidamento del suolo, definizione degli ambiti da destinare a parco o a riserva naturale.

Per i Comuni: per i piani regolatori generali, i piani di recupero, i piani degli insediamenti produttivi, i piani particolareggiati, i piani del traffico, i

piani per l'edilizia economica e popolare; per la gestione del territorio: i piani per le opere pubbliche, le concessioni, i servizi a rete. Senza un supporto appropriato, la complessità dei dati territoriali ne rende difficile la gestione a più livelli operativi.

Il GIS offre il supporto per gestire la complessità dei dati territoriali; permette di avere una visione complessiva e precisa delle diverse componenti del territorio e di anticiparne le evoluzioni e le trasformazioni; garantisce un'informazione omogenea, coerente e costantemente aggiornata, migliorando i costi di gestione.

Il GIS fornisce agli enti amministrativi le risposte più adatte per: eseguire analisi geografiche, progettare interventi nel territorio, prefigurare scenari congruenti con le modifiche che la realtà subisce, simulare effetti di eventi naturali (alluvioni, incendi), simulare interventi umani (come azioni di pianificazione), monitorare il territorio, sostenere processi decisionali, migliorare la gestione delle proprie risorse.

1.10 L'integrazione tecnologica

Un'ultima considerazione in merito alle applicazioni GIS riguarda l'integrazione con altre tecnologie. La finalizzazione di un GIS a specifiche esigenze applicative rende utile, possibile e a volte necessario ricorrere anche ad altre tecnologie quali centraline di raccolta dati (ad esempio per la misura della qualità dell'aria), GPS (per la radiolocalizzazione di mezzi mobili), Videotel (per l'aggiornamento remoto degli attributi), ATM (per la distribuzione al pubblico delle informazioni per mezzo di sportelli tipo bancomat), reti locali e/o geografiche (per la condivisione della cartografia di base con altri enti e/o dipartimenti), ipertesti (per la lettura intelligente delle disposizioni amministrative relative agli oggetti geografici), multimedialità (per l'integrazione di basi dati diverse, documenti, fotografie, filmati, ecc.), RDBMS (per l'accesso geografico ad informazioni contenute in altri sistemi), fino ad ipotizzare, per specifiche applicazioni, sconfinamenti nella robotica, nella realtà virtuale, nelle autostrade elettroniche, nei sistemi esperti o nella televisione interattiva. Tutto ciò, pur apparendo futuribile, può essere invece realtà di oggi se

ritagliata attorno ad una applicazione GIS che, non ci stancheremo mai di ripetere, ha una domanda precisa a cui rispondere e quindi può mettere in campo risorse dedicate.

Va di moda, in altri settori, pensare ad applicazioni verticali, e quindi, per contro, ad applicazioni orizzontali. Nel GIS si potrebbero ad esempio definire verticali le applicazioni mirate ad uno specifico settore (monitoraggio, valutazione di impatto, simulazione del traffico) ed orizzontali quelle legate ad una tecnologia (mosaicatura di mappe, fotointerpretazione, condivisione di data base). In realtà un'applicazione è sempre specifica e ciò che ci deve guidare è sempre la domanda di servizi (simulazioni, previsioni, dimensionamenti) cui risponde.

Sono applicazioni ad esempio quelle che calcolano i percorsi ottimali per la raccolta di rifiuti o la distribuzione di merci, quelle che monitorizzano l'inquinamento, quelle che calcolano domanda e offerta di servizi, quelle che simulano l'effetto di piani di riconversione, ecc.

Non vale quindi tanto la pena tentare di individuare delle tipologie, quanto dare una panoramica generale, non certo esaustiva, della ricchezza di utilizzi che presenta oggi questa tecnologia.

In futuro possiamo ipotizzare che l'utente, con semplici e potenti prodotti integrati object-oriented, possa estrarre dati da un server, rimodellarli, arricchirli con dati prelevati da altri archivi, organizzarli in un proprio schema, lanciare in cascata una serie di funzioni e memorizzare il tutto (connessioni, modelli, dati e procedure) in un "oggetto", magari legato ad un'icona e riprocessato quotidianamente.

Queste potrebbero essere le applicazioni del futuro: un futuro che è già cominciato, dato che la tecnologia GiS presente oggi sul mercato già consente di fare cose simili su un semplice personal computer.

1.11 Diffusione dei dati GIS

Gli ultimi studi sui costi/benefici dei Sistemi Informativi Territoriali effettuati nei paesi d'oltre oceano hanno evidenziato una netta riduzione del costo attribuito ai dati rispetto al costo complessivo del sistema. Si sta gradual-

mente decrementando la spesa sui dati dal 70% al 20%, favorendo in tal modo la diffusione della tecnologia GIS.

La riduzione della spesa sui dati è dovuta essenzialmente alla grande disponibilità di basi territoriali in forma digitale: dai dati di pubblico dominio degli enti governativi alle banche dati, generate da società private, immesse sul mercato a prezzi decisamente interessanti.

Anche in Italia molti enti pubblici e società private hanno acquisito basi territoriali di particolare interesse per molti potenziali utenti.

In questo senso, una recente iniziativa italiana è stata quella della pubblicazione del catalogo ArcData Italia, nel quale sono presenti dati di diversa tipologia, da coperture geografiche a tabelle di attributi, da modelli digitali del terreno a immagini da satellite.

Tali banche dati sono state sviluppate da ESRI ITALIA, da vari enti pubblici e società private e sono potenzialmente utili in diversi campi di applicazione quali: gestione delle risorse naturali, analisi socio-demografiche, trasporti, pianificazione territoriale, ricerche di mercato, ecc.

Tra i primi dati presenti a catalogo figurano grafi stradali (con la toponomastica e i numeri civici), foto da satellite, indicatori di marketing, cartografia a varie scale di alcune regioni (limiti amministrativi, idrografia, strade, ferrovie, ecc.) e dati statistici georeferenziati, mentre altre banche dati sono in preparazione.

Capitolo 2

Software per la Geostatistica

2.1 Introduzione alla Geostatistica

La geostatistica è quella branca della statistica che si occupa dell'analisi di dati spaziali [22], strettamente collegata con i Sistemi Informativi Territoriali. Il suo campo classico di applicazione sono le Scienze della Terra, in particolar modo nella Geologia, Geologia Ambientale, Ecologia, Meteorologia, Agronomia.

La geostatistica si occupa di valutare l'autocorrelazione spaziale dei dati, cercando di verificare se osservazioni effettuate su punti vicini presentino effettivamente una minore variabilità rispetto ad osservazioni effettuate tra punti distanti. L'obiettivo è quindi valutare l'effetto della posizione del punto di misura sulla variabilità del dato osservato. Tale variabilità viene di solito elaborata con lo strumento del semivariogramma, che valuta il grado di variabilità di punti a distanze crescenti.

Lo studio della variabilità spaziale è necessario per la successiva fase di predizione spaziale con cui si possono fornire delle stime sul valore assunto e sugli errori commessi da una variabile in una posizione in cui la misurazione non è stata effettuata. Il metodo di interpolazione statistico più noto è il kriging.

Esistono vari tipi di kriging:

- Kriging ordinario (per processi stazionari)

- Kriging universale (per processi non stazionari)

2.1.1 Autocorrelazione

L'autocorrelazione definisce il grado di dipendenza spaziale tra i valori assunti da una variabile campionata.

Una caratteristica intuitiva dell'ambiente è che le sue proprietà sono in relazione fra di loro in una qualche scala, grande o piccola che sia. Questa situazione è definita autocorrelazione, in Figura 2.1, A raffigura il rapporto tra correlazione e distanza, B l'autocorrelazione. Questo significa che valori campionati in luoghi vicini tra di loro, tendono ad avere comportamenti simili, mentre valori di una stessa variabile misurati in campioni raccolti in luoghi lontani tra di loro tendono ad avere comportamenti differenti, o almeno tendono a differire dai valori medi che si riscontrano nei due luoghi stessi. In tal senso, la correlazione fra i valori della variabile tende a diminuire con l'aumentare della distanza.

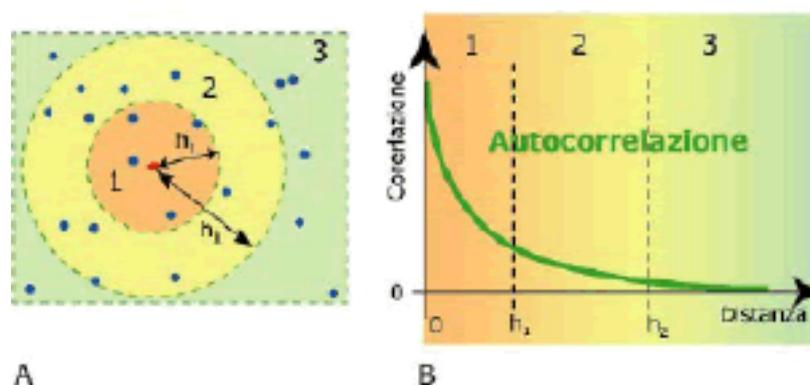


Figura 2.1: Correlazione e Autocorrelazione

Adottando il punto di vista stocastico, ogni punto dello spazio non ha un solo valore per una proprietà, ma un intero insieme di valori. In questo modo il valore osservato diventa un valore estratto a caso da un infinito numero di valori possibili, assunti da qualche distribuzione di probabilità per una qualche legge. Questo significa che in ogni punto dello spazio c'è una variazione. Così, ad ogni punto x_0 , una proprietà, $Z(x_0)$, è trattata

come una variabile casuale, generalmente continua e non discreta, con una media simbolo, una varianza (varianza), momenti di ordine più alto ed una funzione di probabilità di densità cumulativa. Questa variabile ha quindi una distribuzione di probabilità da cui viene estratto il valore reale. Il set di valori reali (misurati) di Z può considerarsi come una variabile regionalizzata in quanto, è una variabile il cui valore è fortemente condizionato dalla posizione spaziale. Detta $Z(x)$ la variabile regionalizzata nella posizione x , potremmo scrivere:

$$Z(x) = a + R(x) \quad (2.1)$$

in cui a è la componente aleatoria e $R(x)$ la componente regionalizzata. Quando a è dominante rispetto a $R(x)$ si studia la variabile con i metodi classici della statistica. In caso contrario si ricade nel campo della geostatistica.

I valori delle variabili regionalizzate tendono ad essere in relazione tra loro e, come detto, due valori vicini ad un terzo tendono ad essere simili, mentre quelli più lontani lo sono meno.

2.1.2 Ipotesi di stazionarietà

Un processo stazionario è un processo stocastico in cui la funzione di densità di probabilità di una qualche variabile casuale Z non cambia né nel tempo né nello spazio. Quindi anche i parametri media e varianza non cambiano nel tempo e nello spazio, di conseguenza la distribuzione del processo casuale mantiene gli stessi attributi ovunque.

Supponiamo di misurare una variabile regionalizzata $Z(x)$ in una particolare direzione dello spazio ottenendo alcuni valori per ogni punto in questa direzione. Si potrebbe ottenere un andamento di valori simile a quello mostrato.

Come si vede, è presente un cambiamento sistematico del valore medio della variabile, rilevante alla scala dell'osservazione, e allora non si può più assumere una stazionarietà della media, o stazionarietà del modello. Questa

condizione di non stazionarietà dei valori della variabile e data dalla somma di due componenti:

- drift (aumento sistematico del valore medio)
- residuo (variabilità attorno alla componente sistematica)

Il drift è essenzialmente il valore medio della variabile in funzione della localizzazione in cui la variabile è stata misurata e può essere scritto come $m(x)$. Il valore della variabile Z per ogni punto x può essere espresso allora come la somma della media del punto x e la deviazione dalla media del valore misurato, $Y(x)$:

$$Z(x) = Y(x) + m(x) \quad (2.2)$$

Le assunzioni fatte sono le seguenti:

- il residuo $Y(x)$ è una variabile casuale stazionaria, in cui la media è costante;
- il drift sia una funzione deterministica della localizzazione;
- il drift e il residuo sono non correlati

2.1.3 Semivariogramma

Il semivariogramma è un algoritmo geostatistico che viene impiegato per valutare l'autocorrelazione spaziale di dati osservati in punti georeferiti.

La funzione semivariogramma interpola la semivarianza dei valori osservati in gruppi di coppie di punti a determinate distanze (lag) secondo una certa direzione.

La semivarianza è pari a:

$$\gamma(h) = \frac{1}{(2m(h))_{i=1}} \sum_{i=1}^{(h)} (Z(x_i + h) - Z(x_i))^2 \quad (2.3)$$

in cui Z è il valore di una misura in un particolare punto, h è un intervallo di distanza tra punti di misurazione (lag) e $m(h)$ rappresenta il numero di coppie di osservazioni effettuate alla distanza h .

Gli assi del semivariogramma sperimentale sono distanze tra coppie di dati (asse x) e semivarianza (asse y) (Figura 2.1). Il grafico ottenuto dall'applicazione della formula 2.3 è formato da una serie di punti distinti, distanti tra loro un certo lag definito. Il semivariogramma sperimentale (punti discreti) deve essere interpolato con diverse funzioni matematiche, in modo da determinare il tipo di autocorrelazione spaziale della variabile misurata, che saranno successivamente usate nell'interpolazione con il kriging. Nel processo di kriging è infatti richiesta una funzione continua per l'assegnamento dei pesi in tutti i punti, che è legata al valore della semivarianza. Tale modello di semivariogramma si desume a partire dall'osservazione del semivariogramma sperimentale.

Si può immaginare una funzione che meglio approssimi i punti del variogramma sperimentale; il problema è però capire quanto questa curva seguirà i dati sperimentali. Per modellare bene il variogramma bisogna quindi porre attenzione sull'andamento generale dei punti e non sulle singole fluttuazioni, quindi il tipo di funzione da scegliere dovrà essere la più semplice possibile, in relazione sempre alla complessità dell'andamento del variogramma.

I parametri stimati sono:

- Nugget: descrive il livello di variabilità casuale;
- Sill: valore massimo della semivarianza quando si ha stazionarietà (esso approssima per eccesso la varianza campionaria);
- Range: rappresenta la distanza massima entro la quale si manifesta correlazione tra semivarianza e lag;

La funzione semivariogramma è una funzione di correlazione spaziale, con determinate caratteristiche grafiche e quantitative, dette proprietà strutturali, quali: simmetria, continuità, zona d'influenza, comportamento vicino all'origine, anisotropia e drift.

La modellazione del variogramma è un passo molto delicato in quanto consiste nello scegliere la curva che meglio approssima la serie discreta di punti ottenuti dall'analisi del variogramma sperimentale. La procedura di adattamento dei modelli (fitting model) al variogramma è difficile poiché l'accuratezza delle semivarianze osservate non è costante e il variogramma può contenere molte fluttuazioni punto a punto

Perciò è raccomandabile una procedura di fitting del modello che inglobi sia l'aspetto visuale nell'andamento della funzione che quello statistico, come segue:

1. va tracciato il variogramma sperimentale;
2. si sceglie uno o più modelli tra quelli disponibili, i quali abbiano approssimativamente una forma corretta e con sufficiente dettaglio, in modo da rispettare le tendenze principali dei valori sperimentali;
3. si adatta il modello attraverso procedure statistiche di minimizzazione degli errori;
4. si ispeziona il risultato grafico, per valutare qualitativamente la bontà del risultato.

Esistono vari tipi di modelli per l'approssimazione dei variogrammi sperimentali: esponenziale, sferico, gaussiano e lineare.

2.1.3.1 Calcolo del variogramma

La stima della funzione variogramma viene effettuata sulla base dei dati provenienti dal campionamento del fenomeno oggetto di studio. Se si hanno dati campionati secondo una maglia regolare il calcolo è molto semplice poiché, data la stazionarietà dell'incremento $Z(x+h) - Z(x)$, risulta immediato calcolare la funzione variogramma per una certa direzione e per un determinato lag h . Il calcolo del variogramma si basa sul considerare le differenze dei valori della variabile regionalizzata in due localizzazioni differenti, separate da una distanza h . La procedura da seguire è questa:

1. si parte facendo la differenza tra i valori $Z(x_1)$ e $Z(x_2)$, poi tra $Z(x_2)$ e $Z(x_3)$, fino alla coppia $Z(x_{i-1})$ e $Z(x_i)$; le differenze saranno uguali a $m(x)$, ovvero al numero di coppie di campioni per questo lag ;
2. il risultato di ogni differenza si eleva al quadrato;
3. si sommano tutti i quadrati;
4. si divide questa somma per $2m(h)$;
5. si ripete la stessa procedura da 1 a 4 per il secondo lag (il doppio del primo);
6. si ripete la stessa procedura da 1 a 4 per il terzo lag (tre volte il primo);
7. si ripete la procedura fino all'ultima distanza di lag che si vuole.

Generalmente si sceglie un numero di lag che varia da 10 a 20. Infatti, un basso numero di lag (minore di 10) potrebbe far perdere di significato al variogramma, abbassando il potere risolutivo del variogramma. Al contrario, un alto numero di lag porterebbe a considerare troppi lag oltre il range massimo di correlazione spaziale tra i dati.

Purtroppo però non sempre si ha una maglia regolare di campionature, bisogna allora aggiustare la tecnica per il calcolo del variogramma. In questo caso infatti potrebbe accadere che in una certa direzione r non cada nessun punto campionato, perché disposti in modo irregolare. Per ovviare a ciò allora si considera una direzione r , individuata tramite un angolo a partire da una direzione di riferimento, con una tolleranza angolare $\Delta\phi$. Una tolleranza Δh deve anche essere data sulla distanza. Così facendo tutte le coppie aventi distanza compresa tra $(h - \Delta h)$ e $(h + \Delta h)$ e allineate secondo la direzione compresa tra simbolo $(\phi - \Delta\phi)$ e $(\phi + \Delta\phi)$ contribuiscono al calcolo del variogramma.

I valori delle tolleranze da adottare dipendono ovviamente dalla quantità di campioni di cui si dispone: più sono numerosi i campioni e più piccole possono essere le tolleranze, consentendo una maggior precisione nel calcolo dei variogrammi.

2.1.4 Mappa del variogramma

La mappa del variogramma è un grafico che rappresenta l'andamento della varianza nello spazio. Viene costruito, utilizzando una griglia con celle quadrate (o con settori circolari), sulla base dei valori delle ordinate del variogramma sperimentale. L'ampiezza delle celle è determinata dall'ampiezza dei lag, più è alto il numero di lag tanto più le celle sono piccole.

La mappa del variogramma fa sì che sia subito comprensibile come vari nello spazio l'anisotropia per poter decidere su quali direzioni principali concentrare l'attenzione con il variogramma. Ogni cella della mappa è rappresentata infatti con una simbologia di colori variabile a seconda del valore del variogramma secondo una scala cromatica propria (Fig. 2.2).

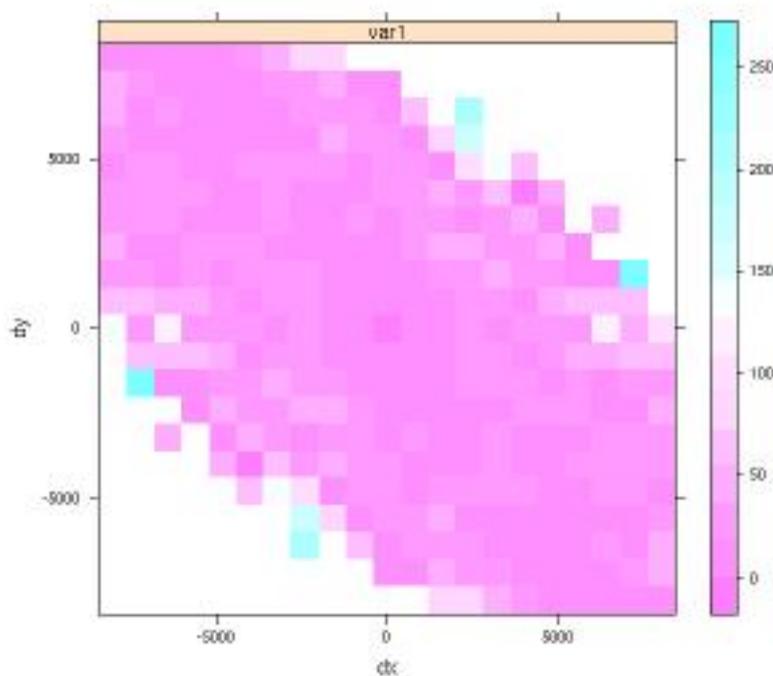


Figura 2.2: Esempio di mappa del variogramma

Questo grafico, come anticipato, permette di individuare la direzione di maggiore continuità del dato semplicemente osservandolo dal centro ed individuando la direzione per cui si hanno valori più bassi di varianza (quadrantini

rosa scuro). Questo differente andamento della variabilità spaziale è noto come anisotropia. L'anisotropia è la proprietà per la quale un determinato oggetto ha caratteristiche che dipendono dalla direzione lungo la quale esse sono considerate.

2.1.5 Cross-Validation

La cross-validation è una procedura per cui iterativamente ogni campione viene escluso dal dataset ed interpolato attraverso il valore degli altri, utilizzando il modello di variogramma che si vuole testare. Il confronto tra il valore stimato ed il valore reale è detto residuo della cross-validation.

Lo studio del dataset dei residui ci indica il comportamento del modello. I principali parametri da studiare sono:

1. La media dei residui, che indica l'accuratezza della stima, il quale deve essere prossimo a zero.
2. L'errore minimo quadrato (RMSE): $\sqrt{\frac{1}{n} \sum_{i=1}^n (Z_i - Z')^2}$ che deve essere il più piccolo possibile.
3. La deviazione standard del kriging (MSDR) $\frac{1}{n} \sum_{i=1}^n \frac{(Z_i - Z')^2}{\sigma^2}$, la quale permette di confrontare la grandezza degli errori predetti con gli errori realmente commessi, si confrontano cioè la varianza vera e quella calcolata nella cross-validation.

dove Z_i è il valore stimato i -esimo, Z' il valore reale, σ^2 varianza la varianza dell'errore ed n il numero di campioni.

2.1.6 Kriging e Cokriging

Per poter ricostruire la superficie cercata è necessario interpolare i dati disponibili per stimare i valori dove non si hanno campioni. In questo senso il kriging ci viene incontro, in quanto dà una soluzione al problema della stima basato su un modello continuo di variazione spaziale stocastica. Esso

fa il miglior uso della conoscenza della variabile, prendendo in considerazione il modo in cui una proprietà varia nello spazio attraverso il modello del variogramma che è stato scelto e validato.

Esistono diversi tipi di kriging, tra cui kriging ordinario (ordinary kriging) e kriging universale (universal kriging), per diverse tipologie di variabili. Ciò che li differenzia è il tipo di variabile usata: il kriging ordinario può lavorare solo con variabili stazionarie del secondo ordine (presentano cioè media costante e varianza dipendente solo dal lag muovendosi da punto a punto), il kriging universale può invece lavorare anche con variabili non stazionarie (che presentano cioè un drift). Come anticipato una delle assunzioni fatte nel kriging ordinario è la stazionarietà del dato da stimare. Questo significa che muovendosi da una zona ad un'altra del campo la media dei valori è pressoché costante. Quando invece esiste un significativo trend spaziale del dato (caratteristica intrinseca del dato stesso che fa sì che la media dei valori non sia costante ma vari da punto a punto) questa assunzione viene meno. La condizione di stazionarietà del dato può essere comunque ristabilita attraverso l'introduzione di una funzione deterministica che descriva il drift, cioè l'andamento della media, in modo da poter isolare il residuo, cioè la parte aleatoria del dato. Il kriging universale quindi modella e sottrae il drift presente nel dato tramite una funzione deterministica, ed analizza la sola componente aleatoria (residuo).

Dato che il drift ed il residuo sono non correlati è possibile modellare il drift tramite una funzione deterministica della localizzazione, si può scrivere quindi il valore della variabile $Z(x)$ come:

$$Z(x) = \sum_{k=0}^K \alpha_k f_k(x) + m(x) \quad (2.4)$$

dove $\sum_{k=0}^K \alpha_k f_k(x)$ è la somma di un set di funzioni $f_k(x)$ polinomiali di ordine 1 o 2 ed $f_0(x) = 1$ e il termine $m(x)$ è il componente stocastico (variogramma).

L'universal kriging, utilizzato in condizioni di non stazionarietà, fornisce la stima puntuale (nel punto x_0) attraverso la formula:

$$Z_{UK}^*(x_0) = \sum_{k=0}^K \sum_{i=1}^n \alpha_k \lambda_i f_k(x_i) \quad (2.5)$$

la quale fornirà una stima corretta solo se si rispetta la condizione:

$$\left[\sum_{i=1}^n \lambda_i f_k(x_i) = f_k(x_0) \right] \forall k = 1, \dots, K \quad (2.6)$$

La soluzione del sistema qui di seguito permette di quantificare i pesi simbolo per la stima del generico punto:

$$\left[\sum_{i=1}^n \lambda_i \gamma(x_i, x_j) + \psi_0 + \sum_{k=0}^K \psi_k f_k(x_i) = \gamma(x_i, x_0) \right] \forall i = 1, \dots, N \quad (2.7)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad (2.8)$$

$$\left[\sum_{i=1}^n \lambda_i f_k(x_i) = f_k(x_0) \right] \forall k = 1, \dots, K \quad (2.9)$$

dove il simbolo (x_i, x_j) è la semivarianza dei residui tra i punti x_i e x_j e il simbolo (x_i, x_0) è la semivarianza tra l' i -esimo punto campionato ed il punto bersaglio della stima x_0 .

Nel kriging, sia ordinario che universale, come appena visto, tramite un sistema di equazioni lineari, si assegnano i pesi per i campioni circostanti al punto analizzato che minimizzano la varianza totale degli errori. Solitamente i quattro cinque campioni più vicini contribuiscono per l'80% del peso totale, il restante 20% viene assegnato all'incirca ad altri dieci punti vicini. Si hanno vari fattori che influenzano i pesi tra cui:

- i campioni vicini portano più peso di quelli lontani e l'entità del peso dipende dalla loro posizione e dal variogramma;
- i campioni raggruppati in un certo intorno portano meno peso di quelli isolati.

Nell'ambito della statistica multivariata il metodo di interpolazione usato è il cokriging. Questo metodo è molto simile al ordinary kriging, l'unica differenza risiede nel fatto che, nello stimare la variabile principale (target) non si basa solo sui valori della variabile esaminata ma prende in considerazione anche altre variabili ausiliarie.

Si consideri il caso in cui si abbiano due variabili spazialmente correlate $Z_1(x)$ e $Z_2(x)$, campionate rispettivamente in un set di n_1 e n_2 locazioni, volendo effettuare la stima del valore della variabile $Z_1(x_0)$ tramite i valori circostanti di $Z_1(x)$ e $Z_2(x)$ utilizzeremo:

$$Z^*(x_0) = \sum_{i=1}^{n_1} \lambda_i Z_1(x_i) + \sum_{i=1}^{n_2} \omega_i Z_2(x_i) \quad (2.10)$$

dove λ_i e ω_i della variabile rappresentano i pesi delle due variabili considerate.

Il valore dei pesi da assegnare si trova risolvendo il sistema qui di seguito:

$$\sum_{j=1}^{n_1} \lambda_j \gamma_1(x_i, x_j) + \sum_{j=1}^{n_2} \omega_j \gamma_{12}(x_i, x_j) + \mu_1 = \gamma_1(x_i, x_0) \quad (2.11)$$

$$\sum_{j=1}^{n_1} \lambda_j \gamma_{12}(x_i, x_j) + \sum_{j=1}^{n_2} \omega_j \gamma_2(x_i, x_j) + \mu_2 = \gamma_2(x_i, x_0) \quad (2.12)$$

$$\sum_{j=1}^{n_1} \lambda_j = 1 \quad (2.13)$$

$$\sum_{j=1}^{n_2} \omega_j = 0 \quad (2.14)$$

Il cokriging è quindi utile quando la variabile principale è sottocampionata rispetto alle altre, come nel nostro caso in cui si ha che la variabile $pozzi_{01}$ (principale), che rappresenta i valori di falda misurati al 2001, ha un terzo dei campioni della variabile $pozzi_{91}$ (di appoggio).

2.2 'R' : storia e caratteristiche

Esiste in circolazione una gamma assai vasta di software specializzati nell'analisi statistica dei dati, basta ricordare SAS, SPSS, STATA, STATGRAPHICS PLUS, SHAZAM, S PLUS, MINITAB, GAUSS, etc., solo per citarne alcuni che sono in commercio. Sono prodotti che costituiscono senz'altro un fondamentale ed insostituibile ausilio per il lavoro dello statistico, tuttavia molti di questi programmi sono anche alquanto costosi ed è consentito l'uso su licenza da parte del produttore.

Da alcuni anni a questa parte, soprattutto in ambito universitario (ma non solo) si sta sempre più diffondendo un nuovo package che merita di sicuro una debita trattazione e l'interessamento da parte degli statistici e di coloro che fanno analisi dei dati e che costituisce anche una valida alternativa ai software sopraccitati. Ci stiamo riferendo al software R [20, 21], un ambiente statistico open source.

Esso è il frutto del lavoro collettivo svolto da un gruppo, sempre più folto, di ricercatori in campo statistico ed informatico a livello mondiale. R più che un software statistico può essere definito come un ambiente, costituito da una varietà di strumenti, orientato alla gestione, all'analisi dei dati e alla produzione di grafici, basato sul linguaggio S creato da AT&T Bell Laboratories, ambiente dal quale è nato un altro software commerciale più noto, S-Plus, prodotto prima da MathSoft e ora da Insightful.

R, a differenza di quest'ultimo è un GNU-Software, ossia è disponibile gratuitamente nei termini della GPL (General Public License): chiunque può accedere al codice sorgente e modificarlo, migliorarlo, integrarlo e mettere a disposizione della comunità scientifica il proprio lavoro nella logica e nella filosofia del software opensource. La versione iniziale di R venne sviluppata nel 1996 per il sistema operativo MacOS da Ross Ihaka e Robert Gentleman del dipartimento di Statistica dell'Università di Auckland in Nuova Zelanda.

Successivamente altri ricercatori iniziarono ad aggiungersi e a fornire il loro contributo alla scrittura del codice sorgente e al miglioramento dell'applicazione, sviluppata in seguito anche per l'ambiente UNIX. Nel 1997 è nato l'R Development Core Team, gruppo formato da statistici di tutto il mondo

che si occupano dello sviluppo e della distribuzione del programma e, ancora più recentemente, nel 2003 è stata costituita dagli stessi membri dell'R Development Core Team la R Foundation for Statistical Computing.

Si tratta di una organizzazione non profit che lavora per il pubblico interesse allo scopo di promuovere lo sviluppo (attuale e futuro) e la diffusione del software, di fornire un punto di riferimento per persone, istituzioni ed imprese che desiderano sostenere il progetto e interagire con la community che ruota intorno ad R, di gestire e tutelare il copyright di R e della relativa documentazione.

2.2.1 Funzionalità dell'ambiente R

L'ambiente R è basato sul concetto di "package" tradotto di solito in italiano con pacchetto. Un package è un insieme di strumenti che svolgono determinate funzioni, ma può anche contenere solo dati oppure solo documentazione. Attualmente è disponibile una vasta gamma di packages utilizzabili per la risoluzione di specifici problemi o per analisi statistiche molto particolareggiate.

Il cuore di R è rappresentato dal modulo base (che offre gli strumenti fondamentali per l'analisi statistica) e attorno a questo modulo "ruotano" una serie di altre librerie addizionali, alcune delle quali sono già comprese nel programma R al momento in cui lo si installa, mentre altre librerie ancora, in relazione alle esigenze e necessità, possono essere aggiunte e installate dall'utente.

R è un ambiente computazionale molto dinamico ed in continua evoluzione e frequentemente nuovi ulteriori packages sono realizzati e messi a disposizione dei ricercatori. Esaminiamo sinteticamente i principali packages di R puntualizzando alcune delle funzionalità che questi consentono di operare. Quella che segue è la situazione valida sino alla versione 1.8.1.

1. *Package "base"*: questo modulo è il motore dell'ambiente R, viene caricato automaticamente e fornisce all'utente gli strumenti per le più importanti e diffuse analisi statistiche come: ANOVA, regressione lineare, statistica descrittiva ed inferenziale, analisi esplorativa dei dati,

grafici elementari, modelli lineari generalizzati, generazione di campioni delle più comuni variabili casuali, operazioni su matrici e vettori.

2. *Package "ctest"*: consente di effettuare tutti i principali test statistici per la verifica delle ipotesi (test t, test F, test di normalità, test non parametrici, test per l'omoscedasticità delle varianze, test Chi-quadro, etc.)
3. *Packages "ts" e "tseries"*: sono due packages dedicati all'analisi delle serie temporali
4. *Package "spatial"*: analisi dei dati di serie spaziali
5. *Packages "grid", "lattice", "rgl" e "scatterplot3d"*: permettono di realizzare e manipolare grafici avanzati e tridimensionali
6. *Packages "mva", "amap", "multidim" e "multiv"*: permettono di eseguire le principali analisi su dati multidimensionali (analisi delle componenti principali, analisi fattoriale, correlazione canonica, scaling multidimensionale, clustering gerarchico)
7. *Package "cluster"*: è il modulo specializzato nell'analisi dei gruppi (cluster analysis)
8. *Package "nls"*: effettua l'analisi della regressione con modelli non lineari applicando il metodo dei minimi quadrati (Nonlinear Least Squares)
9. *Package "matrix"*: strumenti per l'uso avanzato di matrici e vettori con metodi numerici per l'algebra lineare
10. *Packages "survival" e "survrec"*: pacchetti specializzati nell'analisi della sopravvivenza
11. *Package "nlme"*: modelli lineari e non lineari con effetti misti (Linear and Nonlinear mixed effects models)
12. *Package "foreign"*: contiene tutta una serie di funzioni utili per importare file da software statistici quali Minitab, S, SAS, SPSS, Stata.

Con il rilascio della versione 1.9.0. avvenuta il 12 aprile 2004 ci sono state importanti modifiche nella struttura organizzativa dei packages all'interno di R. In particolare il package che nella precedente release era denominato "base" è stato diviso in quattro pacchetti denominati: "base", "graphics", "stats", "utils". Tutti e quattro sono caricati in una installazione standard, ma tale divisione alleggerisce il lavoro svolto dal sistema, consentendo, anche grazie ad altre modifiche apportate, migliori prestazioni. Inoltre si è riusciti ad aumentare la velocità di start up di una sessione di R: il tempo impiegato con l'ultima versione è circa i due terzi di del tempo della versione 1.8.1.

Il package "base" continua a contenere i tools strettamente di base del software, mentre i packages "ctest", "eda", "modreg", "mva", "nls", "stepfun" and "ts" sono confluiti nel pacchetto "stats". Anche tutti gli archivi (datasets) contenenti serie storiche sono stati spostati in quest'ultimo package. "graphics" contiene tutti gli strumenti di grafica di base, invece "utils" contiene delle utility del sistema tra cui i comandi per help, l'installazione e la rimozione di nuovi packages, il lancio di sessioni dimostrative delle funzionalità del programma (demo), etc.. Con la release 2.0.0 disponibile dal 4 ottobre 2004 si sono avute ulteriori modifiche e aggiornamenti. Tra le principali novità

- è stato implementato un nuovo sistema di caricamento del programma e dei packages detto 'lazy loading': al posto di mantenere in memoria gli oggetti di R, questi sono memorizzati in un database su disco rigido e caricati solo al primo uso. Ciò consente un più veloce start up con una riduzione del 40% rispetto alla versione 1.9.x e un uso più efficiente della memoria dell'elaboratore;
- tutti i datasets che prima si trovavano nei packages "base" e "stats" sono stati inseriti in un nuovo package denominato "datasets";
- il package "graphics" è stato diviso in due pacchetti "grDevices" (dispositivi grafici condivisi tra i grafici di base e quelli del package "grid") e "graphics" (grafici di base).

2.2.2 R e la statistica ambientale

In questo paragrafo si vogliono presentare brevemente alcuni packages di R di particolare interesse per la statistica ambientale [21]. Per quanto riguarda le applicazioni della geostatistica si possono considerare:

- geoR*: fornisce una serie di funzioni per l'analisi di dati geostatistici;
- geoRglm*: modelli spaziali lineari generalizzati;
- gstat*: modelli geostatistici univariati e multivariati, previsioni e simulazioni;
- GRASS*: interfaccia tra GRASS ed il software R;
- spatial*: funzioni per kriging e point pattern analysis;
- spatstat*: spatial Point Pattern analysis, stima del modello e simulazione;
- splanacs*: Spatial and Space-Time Point Pattern Analysis;
- sgeostat*: un object-oriented framework per i modelli geostatistici;
- vardiag*: un package per la diagnostica dei variogrammi.

Relativamente alla problematica dei valori estremi è molto valido il package *evd* nel quale sono disponibili una serie di funzioni relative alle seguenti distribuzioni :

- Distribuzione del minimo e del massimo
- Distribuzione di Frechet
- Generalized Extreme Value Distribution
- Generalized Pareto Distribution
- Distribuzione di Gumbel
- Distribuzione di statistiche ordinali

con stima di massima verosimiglianza dei parametri;

Per l'utilizzo del metodo CART (Classification and Regression Tree) e degli alberi di regressione generalizzati nello studio dei fenomeni ambientali possono essere utili i packages:

tree: Classification and regression trees;

rpart: Recursive partitioning and regression trees

Capitolo 3

Software GIS

3.1 Requisiti e funzionalità

Ciò che si richiede ad un programma applicativo e di cui si dovrebbe tener conto al momento della sua scelta è, o dovrebbe essere, funzione degli obiettivi che si intendono raggiungere attraverso il suo utilizzo. Tali obiettivi possono variare anche in modo considerevole, ma si può tentare di schematizzarli secondo tre tipologie:

minimo: costruzione di una o più carte sulle quali vengono rappresentati i temi di interesse; tali carte rappresentano una fotografia al tempo t del sistema geografico studiato. Il contenuto informativo deve essere particolarmente ricco in quanto queste carte, che non si propongono come strumenti diretti di previsione dei diversi fenomeni cui fanno riferimento, devono comunque essere utilizzabili dagli addetti ai lavori per poter sviluppare progetti di intervento;

intermedio: si pone ad un gradino superiore rispetto al precedente in quanto se al livello minimo l'analisi si limita alla fotografia, in questo caso lo scopo è quello di affrontare lo studio delle variazioni nel tempo (dinamica) dei fenomeni di interesse. Per fare ciò è necessario avvalersi di dati raccolti ad intervalli successivi (serie storiche) e di tecniche di analisi appropriate;

alto: a questo livello si collocano le applicazioni che hanno come obiettivo quello di approdare ad una conoscenza sul possibile sviluppo dei fenomeni

analizzati attraverso l'utilizzo della modellizzazione.

Naturalmente lo sviluppo complessivo di un GIS può prevedere l'accostamento di tutti e tre i livelli menzionati. Alcuni dei temi studiati con l'ausilio del GIS possono essere affrontati al livello minimo (per la mancanza dei dati necessari ad un superiore approfondimento o perché il tema in questione non richiede un'analisi più dettagliata); per altri invece potrebbero essere previsti livelli di analisi più approfondite nel senso indicato in precedenza. Pertanto la scelta del software GIS dovrebbe rispondere alle reali esigenze, ovvero le sue caratteristiche dovrebbero essere più che sufficienti per rispondere alle necessità degli utenti ma non eccessivamente sovradimensionate (la qual cosa potrebbe portare a notevoli aggravii in termini di costi di acquisizione e di formazione del personale tecnico) rispetto agli obiettivi previsti. In effetti esistono pacchetti GIS un po' per tutte le esigenze (e le "tasche"), ma già i software come ad esempio ArcView o Grass sono in grado di soddisfare la maggior parte delle possibili richieste degli utenti medi. L'elenco delle principali funzioni che un software GIS mette a disposizione possono essere sintetizzate come segue.

Innanzitutto è possibile individuare un insieme di funzionalità che dovrebbero far parte di un qualsiasi pacchetto applicativo GIS e che rappresentano sostanzialmente il livello minimo di funzionalità richieste.

Inserimento dati grafici In genere solo una parte (più o meno consistente) dei dati utilizzati in un GIS sono già disponibili in formato digitale (perché prodotti dall'utilizzatore stesso o acquisibili sul mercato presso fornitori che li distribuiscono nei formati più diffusi). Il più delle volte è necessario procedere alla costruzione di nuovi database o alla conversione, in un opportuno formato digitale, dei dati disponibili. Si pensi ad esempio al processo, denominato digitalizzazione, che consente di convertire in formato digitale le mappe cartacee. Un software GIS deve prevedere le procedure che consentono di automatizzare questo processo utilizzando il sistema della scansione o il ricorso alla digitalizzazione manuale, oltre naturalmente alle funzioni di editing grafico più comuni. · **Funzionalità di Database:** la gestione (inserimento, aggiornamento, modifica) dei dati geometrici e degli eventuali attributi as-

sociati, richiede molto spesso che il software preveda funzionalità proprie di un DBMS (DataBase Management System).

L'utilizzo di tali sistemi è consigliabile quando la quantità di dati presenti nel GIS è particolarmente rilevante e la loro integrità è un requisito primario. In genere, anche i software che non dispongono di un proprio modulo interno per la gestione dei database, prevedono almeno la possibilità di interfacciamento con database esterni dove possono essere registrati tanto i dati alfanumerici che quelli geometrici.

Output grafico I risultati delle elaborazioni devono essere interpretati e comunicati. Proprio per questa ragione è necessario che il software sia in grado di produrre output grafici di alta qualità, dalla visualizzazione sullo schermo di mappe (anche tridimensionali) e grafici, alla rappresentazione cartografica, alla produzione di report alfanumerici (tabelle, documenti standard, ecc.) e di documentazione multimediale.

Selezione ed analisi dei dati Un software GIS dovrebbe essere in grado di aiutare l'utente nella risposta a domande riguardanti: selezione/localizzazione in base alle relazioni spaziali ed agli attributi degli oggetti geografici (quali oggetti corrispondono a queste precise caratteristiche e dove si trovano?); analisi temporale (dove si sono verificati cambiamenti e di quale tipo?); analisi spaziale (che relazioni esistono tra gli oggetti geografici?); idoneità - prodotto finale della pianificazione - obiettivi (qual è il posto migliore per ...?) modellazione di un sistema - relazioni causa/effetto (che effetti produce un determinato evento?) procedimenti iterativi di analisi di idoneità e di modellazione per la valutazione comparata delle possibili soluzioni/scenari (cosa accade se?) Alcuni esempi relativi a casi concreti: Chi è il proprietario dell'edificio all'angolo tra via Roma e via Verdi? Quali sono le abitazioni che si trovano ad una distanza inferiore di 100 metri dalla tangenziale? Quali sono le zone ad uso industriale presenti nel territorio comunale? Ecc. Con il software GIS si possono eseguire sia semplici ricerche (grafiche e alfanumeriche) sia sofisticate analisi spaziali - per la creazione di modelli, l'individuazione

di tendenze, la predisposizione di scenari - nell'elaborazione delle quali la tecnologia GIS dimostra tutte le sue potenzialità.

Ambiente di sviluppo E' molto probabile che, dopo il periodo iniziale dedicato all'utilizzo di base ed eventualmente all'approfondimento delle funzionalità del software, gli utilizzatori si propongano di produrre elaborazioni ed ottenere risultati non esplicitamente previsti dagli strumenti standard presenti nel pacchetto software. In questi casi le uniche possibilità per superare questo tipo di difficoltà, sono quelle di rivolgersi ad un prodotto che preveda le funzioni richieste oppure di integrare il software già posseduto con procedure scritte "ad hoc". In questo secondo caso, la presenza di un ambiente di sviluppo interno al software GIS, può diventare elemento fondamentale per ovviare alle "carenze" del software di base.

In sintesi le caratteristiche peculiari che dovrebbero essere sempre presenti in un software GIS sono:

- Interfaccia utente di semplice uso; ·
- Ambiente integrato per la gestione di mappe, tabelle, grafici, layout di stampa;
- Funzionalità avanzate per la visualizzazione delle mappe, l'analisi dei dati, la produzione di cartografia ;
- Procedure guidate per effettuare analisi geografiche complesse;
- Utilità per la proiezione cartografica dei dati;
- Accesso e scambio dati attraverso l'uso dei formati più diffusi;
- Editing dei dati geografici e tabellari;
- Integrazione di dati vettoriali (es. CAD) e raster (es. immagini TIFF);
- Personalizzazione dell'ambiente di lavoro;
- Ambiente di sviluppo integrato per estendere le funzionalità standard;
- Possibilità di espandere ulteriormente le potenzialità del software .

3.2 ArcView GIS

ArcGis è strutturato dal punto di vista logico in applicazioni separate che si occupano di visualizzazione e gestione mappe, gestione dei dati, geoprocessing, ha un'architettura modulare e scalare, che lo rende flessibile nella configurazione del sistema., la modularità consiste nella possibilità di acquisire il sistema nel tempo: ArcReader, ArcView, ArcEditor, ArcInfo, le estensioni, ArcGIS Server, ArcIMS ed ArcPAD. ArcView è in grado di visualizzare, interrogare, analizzare i dati ed inoltre mette a disposizione numerose funzioni per l'editing, il supporto per le annotazioni, etc.

ArcEditor è principalmente orientato alla creazione e all'editing del geodatabase. ArcInfo include tutte le funzionalità di ArcEditor ed in aggiunta ha avanzati strumenti di geoprocessing. La tecnologia ArcSDE all'interno di ArcGIS Server fornisce servizi di database, mentre ArcIMS fornisce servizi di Internet. ArcGIS è scalabile, perchè può essere utilizzato su un singolo desktop o attraverso la rete. E' costruito con tecnologia ad oggetti e tutti i prodotti software della famiglia ArcGIS, condividono le stesse applicazioni, la stessa interfaccia utente e concetti operativi.

In questa ottica, conoscendo uno dei prodotti sarà semplice capire come funzionano gli altri, e personalizzazioni fatte su uno di essi funzioneranno anche con gli altri.

Funzionalità ArcGIS comprende le funzionalità necessarie per un GIS professionale:

- automazione dei dati ed editing supportata da strumenti avanzati per la costruzione delle feature;
- gestione di dati attraverso un'applicazione del tipo di Gestione Risorse, per la visualizzazione di tutti i dati geografici e per la gestione del metadata;
- geoprocessing e manipolazione dei dati possibili attraverso wizards e strumenti per la conversione ed analisi;

- i prodotti ArcGIS Desktop (ArcView, ArcEditor, e ArcInfo) sono abilitati ad Internet e possono facilmente integrare dati da qualsiasi server ArcIMS per analisi con dati in locale;
- aggiungere dati dal Geography Network attraverso una voce nel menu File.

ArcGIS include un numero di elementi unici che sono diventati uno standard per gli utenti:

- proiezioni immediate sia con dati vettoriali che con dati raster;
- produzione di mappe di alta qualità, grazie anche alla possibilità di regolare la trasparenza di qualsiasi layer, sia vettoriale che raster;
- funzionalità di report e grafici di alta qualità, rende più semplice comunicare le idee rappresentate nella mappa; per molti utenti, la migliore qualità cartografica è più significativa dato che fornisce opzioni che prima non erano disponibili;
- ArcGIS lavora con i formati dati esistenti e geodatabase; il geodatabase è come un negozio di informazioni geografiche, all'interno di un DBMS con un modello dati che supporta il versioning e che supporta anche oggetti con attributi e comportamenti; questo permette una gestione integrata dei dati ed un ambiente di editing ed aggiornamenti multiutente.

ArcGIS Desktop comprende tre applicazioni integrate chiave: ArcMap, ArcCatalog, e ArcToolbox.

ArcMap è l'ambiente di editing, di interrogazione, di analisi e di restituzione cartografica ;

ArcCatalog ha la duplice funzione di gestire i dati GIS (e relativi metadati) e di essere lo schema editor, permettendo la creazione di entità quali feature class (linee, punti, poligoni), tabelle, relazioni, domini, ecc...

ArcToolbox consente l'esecuzione di operazioni, anche complesse, di geoprocessing (come overlay, buffer, append) e l'export/import verso altri formati dati.

I Modelli Dati ArcGIS forniscono dei Template per settori industriali ed applicazioni specifiche che includono delle strutture di base non proprietarie che permettono di astrarre e modellare il comportamento di "oggetti" del mondo reale all'interno del Geodatabase.

3.3 Il GIS nel panorama delle applicazioni Open-Source.

Con Open Source (termine inglese che significa sorgente aperto) si indica un software rilasciato con un tipo di licenza che consente a eventuali sviluppatori di accedere liberamente al codice sorgente, in modo che con la collaborazione di più utenti il prodotto finale possa raggiungere una complessità maggiore di quanto si potrebbe ottenere dal lavoro di un singolo programmatore. In particolare i vantaggi dell'utilizzo del software libero si possono riassumere così [11]:

- costi ridotti, non solo sulla spesa iniziale (costi delle licenze), ma anche sulle spese dei servizi di supporto, della formazione, dei costi di migrazione, d'installazione e di gestione;
- indipendenza dai fornitori;
- maggiore sicurezza;
- maggiori possibilità di personalizzazione e di espansibilità.

Un limite riguarda l'incompatibilità di diversi open source con alcune piattaforme, ma questo problema va riducendosi grazie al diffondersi e alla continua ottimizzazione del software libero che negli ultimi anni ha visto un sempre maggiore interesse da parte delle Pubbliche Amministrazioni Europee e di

grosse compagnie inducendo programmatori e traduttori a concentrare gli sforzi per migliorarne le caratteristiche ed elevarne la qualità.

Per quanto concerne i software il GIS la gamma delle applicazioni open source suddivisa per categorie: Desktop, Analisi GIS e Telerilevamento, per ogni categoria citiamo alcuni tra i più rappresentativi.

Desktop *Quantum GIS* : (spesso abbreviato come QGIS) è un'applicazione desktop GIS Open Source molto simile nell'interfaccia utente e nelle funzioni alle release di pacchetti GIS commerciali equivalenti, è supportato continuamente da un attivo gruppo di sviluppatori volontari che emettono con regolarità aggiornamenti e correzioni. Al momento Quantum Gis ha una interfaccia tradotta in 14 lingue, il software viene rilasciato come multi-piattaforma e dal sito possono essere scaricate le compilazioni per Mac OS X , Linux,UNIX , e Microsoft Windows. Le performances sembrano essere migliori su macchine non Windows. I vantaggi per gli utenti Windows e Macintosh rispetto al più noto GRASS GIS è di non dover installare l'emulatore X Window X11, girando Quantum GIS nativamente sulle piattaforme non-linux e non-unix, e di avere una interfaccia utente più semplice e intuitiva, nonostante Quantum non disponga del set esteso di funzioni tipico di GRASS. Paragonato ad altri Desktop GIS della stessa classe, Quantum GIS è un software di minore dimensione e mediamente a parità di operazioni non necessita della stessa quantità di RAM. Essendo distribuito come pacchetto Open Source, il codice sorgente di Quantum GIS è liberamente messo a disposizione dagli sviluppatori e può essere scaricato e modificato. Questo permette la sua riprogrammazione per rispondere a specifiche esigenze. Per aumentare le funzionalità e la compatibilità possono altresì essere compilati dei Plug-In, piccole estensioni caricate al momento del lancio del programma. Quantum GIS permette di far confluire dati provenienti da diverse fonti in un unico progetto di analisi territoriale. I dati, divisi in Layers, possono essere analizzati e da essi viene creata l'immagine mappa con il graficismo che può essere personalizzato dall'utente e eventualmente rispondere alle analisi tipiche del GIS: gradazione di colori, sfumatura di colore, valore unico. La mappa può essere arricchita da icone e da etichette dipendenti dagli attri-

buti degli elementi cartografici. Un sistema di scripting può essere invocato per gestire operazioni ripetitive sui dati. Se installato il supporto GRASS, i Tools di GRASS sono richiamabili da Quantum GIS per essere applicati sui dati del progetto. I dati possono essere riproiettati dinamicamente.

gvSIG : frutto dell'iniziativa intrapresa in Spagna dal Dipartimento delle Infrastrutture e dei Trasporti della Comunità Autonoma di Valencia (Generalitat Valenciana) a partire dal 2003. Il progetto è finanziato mediante il Fondo Europeo di Sviluppo Regionale (FERS/FEDER). Presenta un'interfaccia user friendly ed un rapido accesso ai formati raster e vettoriali più comuni. E' in grado di integrare dati locali e remoti all'interno della stessa vista utilizzando servizi WMS, WFS, WCS, ArcIMS e sorgenti di dati JDBC, PostGIS, Oracle Spatial, ecc. e, inoltre, include una serie di funzioni fra cui l'editing avanzato, il geoprocessing, la gestione dei CRS, il supporto a diversi linguaggi di scripting, il georeferencing, il network analyst, ecc. E' anche un client SDI (Spatial Data Infrastructure), poichè prevede tra le varie funzionalità anche strumenti di ricerca di geodati quali Gazetteer e Catalog. Attualmente gvSIG si è arricchito dell'estensione Sextante, che raccoglie più di 150 strumenti di analisi dati geografici (raster e vettoriali).

Vis5D: è un sistema per la visualizzazione interattiva di grandi grid data sets 5D come quelli prodotti dai modelli numerici atmosferici. Si possono gestire isosuperfici, contour line slices, colored slices, volume renderings, etc. di dati in una griglia 3D, quindi ruotare ed animare immagini in tempo reale.

Analisi GIS *R*: è un ambiente statistico distribuito sotto licenza GNU GPL e sviluppato da un team di ricercatori in ambito statistico e informatico di fama mondiale. Esistono versioni di R per diverse piattaforme e quindi anche per i più diffusi sistemi operativi nell'ambito del personal computer, cioè GNU/Linux, Microsoft Windows e Apple MacOS. Inoltre, R, costituendo un vero e proprio ambiente di programmazione, permette una elevatissima flessibilità nell'implementazione di funzioni di calcolo e di rappresentazione grafica statistica. R è facilmente interfacciabile in un ambiente GIS e consente di effettuare analisi geostatistiche complesse su dati GRASS, tramite il

modulo `spgrass6`. Inoltre, tramite moduli appositi, può essere interfacciato anche con RDMBS come Postgresql e Mysql, permettendo così di realizzare analisi direttamente su geodatabase (es. Postgresql con l'estensione Postgis).

Telerilevamento *Ossim*: Software altamente performante nel campo del remote sensing, image processing, GIS e fotogrammetria. Attivamente sviluppato a partire dal 1996. OSSIM è stato fondato da agenzie governative dell' Intelligence e della Difesa statunitense, attualmente utilizzato in settori di ricerca e in siti operativi. Sviluppato su una serie di librerie software ad alte prestazioni, è scritto in C++ impiegando le più recenti tecnologie di programmazione ad oggetti.

Anche se non è stato citato nelle varie categorie, esiste un software che appartiene a tutte: GRASS. Data la potenza di questo software, paragonabile a a quella di altri di tipo commerciale, è risultato un mezzo idoneo per gli scopi prefissi e che verranno esposte nei prossimi capitoli.

3.4 GRASS

3.4.1 Storia e diffusione

GRASS (Geographic Resources Analysis Support System) è un Sistema Informativo Geografico (GIS) utilizzato per la gestione, elaborazione, modellamento spaziale e visualizzazione di dati geografici bi- e tri-dimensionali. E' un software libero (Open Source Free Software) che può essere acquisito, modificato e ridistribuito nei termini del GNU GPL(General Public License).

GRASS ha subito un continuo sviluppo dal 1982 (in realtà è stato distribuito al pubblico solo nel 1989 ma la vera grande diffusione l'ha avuta con l'avvento di Internet negli anni 90) ed ha coinvolto un gran numero di Uffici federali degli USA, Università e Compagnie private: NASA, NOAA, USDA, the National Park Service, the U.S. Census Bureau, USGS. Lo sviluppo del progetto GRASS e gli aggiornamenti che via via si realizzavano sono stati portati avanti dall'ideatore originario lo U.S. Army Corps of Engineers e in particolar modo dal CERL (Construction Engineering Research Laboratory)

Champaign, Illinois fino alla versione 4.1 nel 1992, alla quale aggiunse cinque aggiornamenti e varie patches fino al 1995. Dal 1997 il progetto è stato portato avanti dal GRASS Research Group della Baylor University, Waco (Texas), U.S.A. che ha rilasciato la versione la 4.2. Sotto il coordinamento dell'USACERL e di altri siti di sviluppo, il GRASS Research Group ha continuato ad aggiornare GRASS finché ha passato il testimone ad un gruppo di ricercatori dell'Università di Hannover coordinato da Markus Neteler che ha curato lo sviluppo dalla versione 5 in poi. Attualmente lo sviluppo continua ad essere coordinato da Neteler presso l'ITC di Trento ed è uscita la versione 6.3 A causa della repentina crescita e popolarità di GRASS, il "Development Team" è diventato un team internazionale che conta sviluppatori in tutto il mondo. Lo sviluppo sta producendo nuove versioni di GRASS, correzione dei bugs, nuovi manuali e documentazione.

GRASS si è diffuso nell'ultimo decennio soprattutto in ambito accademico, ma la potenza e flessibilità del sistema e la migliorata facilità di installazione e d'uso nelle ultime versioni, e ancora l'esistenza di versioni di Linux di facile installazione ed utilizzo l'hanno reso appetibile anche per gli utilizzatori professionali e le Pubbliche Amministrazioni.

Agli inizi GRASS era disponibile solo in ambiente Unix o Unix like significativamente diffuso in ambito accademico ed era di indubbio interesse per la possibilità di intervenire direttamente sul codice sorgente. I primi utilizzatori professionali e delle P.A. non conoscevano Unix (a parte gli utenti ARC/INFO) e necessitavano invece di strumenti in grado di svolgere un numero limitato di operazioni in ambienti più conosciuti come Dos o Windows. La disponibilità del codice sorgente recava loro pochi vantaggi (non erano in grado di manipolarlo od anche soltanto di interpretarlo); la documentazione disponibile era in lingua inglese e non c'erano tutorial; la limitata diffusione dei collegamenti internet spesso non consentiva di consultare la già allora nutrita mailing list degli utilizzatori di GRASS; la limitata velocità di collegamento rendeva impensabile scaricare i file di GRASS con i collegamenti allora esistenti.

Attualmente fattori come: la disponibilità gratuita e completa del codice sorgente, la nutrita presenza di documentazione tecnica, la diffusione capil-

lare ed il miglioramento delle prestazioni della rete, rendono più semplice l'adozione di GRASS.

3.4.2 Caratteristiche

Il sistema è organizzato su tre livelli: core, moduli e interfaccia grafica. È scritto principalmente in C, con alcuni moduli in FORTRAN.

L'interfaccia grafica normalmente utilizzato usa il linguaggio interpretato Tcl/Tk, ma esiste una (vecchia) versione non-free Motif. I moduli sono organizzati per gruppi:

- d.** comandi per la gestione del display grafico;
- g.** comandi per la gestione dei file;
- i.** comandi per il processamento di immagini;
- r.** comandi per l'elaborazione di dati in formato raster;
- v.** comandi per l'elaborazione di dati in formato vettoriale;
- r3.** comandi per l'elaborazione di dati raster tridimensionali;
- p./ps.** comandi per la creazione e gestione di file postscript per la stampa di mappe.

Comprende oltre 350 moduli per l'elaborazione di dati vettoriali (2D/3D), raster e voxel. Ha diverse interfacce per l'integrazione con altri programmi di geostatistica, basi di dati, applicazioni geografiche su internet e altri pacchetti GIS. E' il più grande progetto GIS nell'ambito degli Open Source e può essere utilizzato sia come GIS Desktop che come elemento principale di una più completa infrastruttura GIS.

GRASS è un software Libero (Free & Open Source Software) che può essere acquisito, modificato e ridistribuito nei termini del GNU General Public License come pubblicato dalla Free Software Foundation. Originariamente scritto, sviluppato, gestito e diffuso dall'U.S. Army Construction Engineering Research Laboratories (USA-CERL), sezione dell'US Army Corp of

Engineers, come strumento per la gestione del territorio e la pianificazione ambientale per scopi militari, GRASS si è evoluto divenendo rapidamente un potente strumento in un ampio campo di applicazioni in molte differenti aree della ricerca scientifica. E' attualmente utilizzato in tutto il mondo in ambienti accademici e commerciali, in molti settori governativi, che includono NASA, NOAA, USDA, il National Park Service, l'U.S. Census Bureau, USGS, e in molte compagnie di consulenza ambientale. GRASS è sviluppato da una rete mondiale di operatori che continuano a rilasciare con regolarità nuove versioni di GRASS. Il quartier generale di GRASS è presso l'ITC-irst di Trento (ITC - Trento) ove viene mantenuto lo sviluppo coordinato di GRASS. Le piattaforme supportate includono Linux, Sun Solaris, Silicon Graphics Irix, HP-UX, DEC-Alpha, MacOS X e MS Windows (NT, 2000 Professional, XP). Ecco alcune funzionalità:

- Ortho-Rettifica
- Visualizzazione 3-D
- Creazione e modifica DEM (Digital Elevation Model)
- Rilievi ombreggiati con precisione 16-bit
- Generazione ed etichettatura delle isometriche (Contour Lines)
- Generazione automatica, Digital Line Graph (DLG)
- Analisi statistica
- Conversione di coordinate Ri-proiezione di dati Raster e vettoriali in piu' di 120 proiezioni supportate
- Classificazione e Riclassificazione
- Analisi dati Geostatistici e Spaziali :Varianza, Neighbor, Surface, Prediction .
- Poligoni tematici generati autonomamente da immagini legate a Remote Sensing Esporta e importa dati GIS Data, completi degli attributi

40 formati file GIS compresi: GeoTiff, Tiff, DXF, SDTS, DEM, Shape, ELAS, PNG, ARC, MIF, Tiger, Garmin, ASCII e tanti altri

3.4.3 GRASS e sistemi operativi

GRASS è sviluppato in ambiente UNIX e portato su altri sistemi, tra cui:

- UNIX-like (Solaris, IRIX, Linux, BSD) con interfaccia grafica e terminale in XWindows;
- MSWindows (NT/2000/XP con Cygwin);
- MacOS X;
- Java (JGRASS, sperimentale).

Le differenze riguardano soprattutto l'efficienza e stabilità dei sistemi operativi a prescindere da GRASS. Le figure 3.1, 3.2, 3.3 mostrano come si presenta l'interfaccia grafica di GRASS sotto i diversi sistemi operativi.

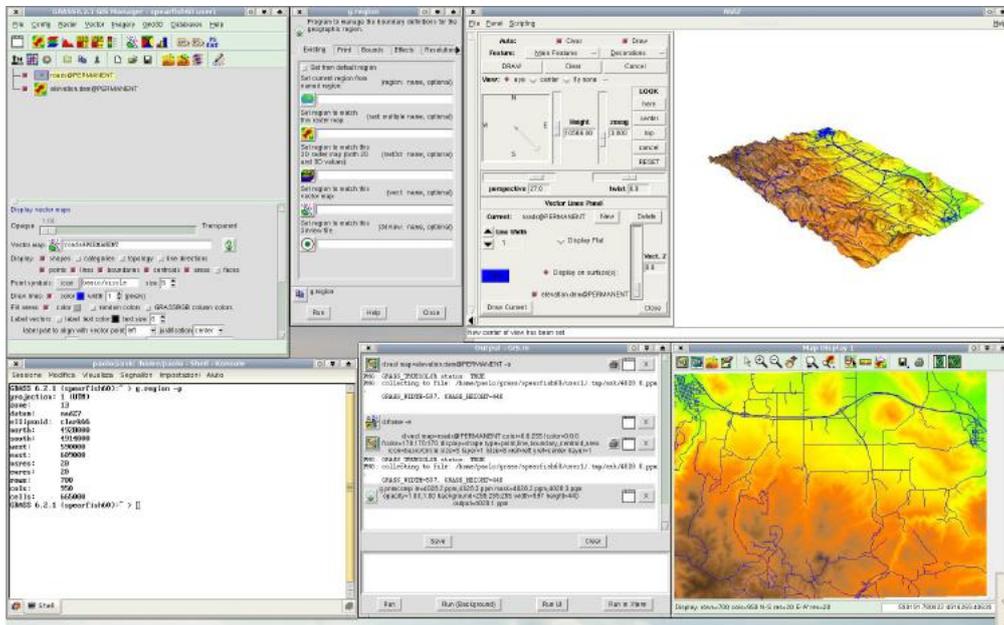


Figura 3.1: GRASS sotto Linux

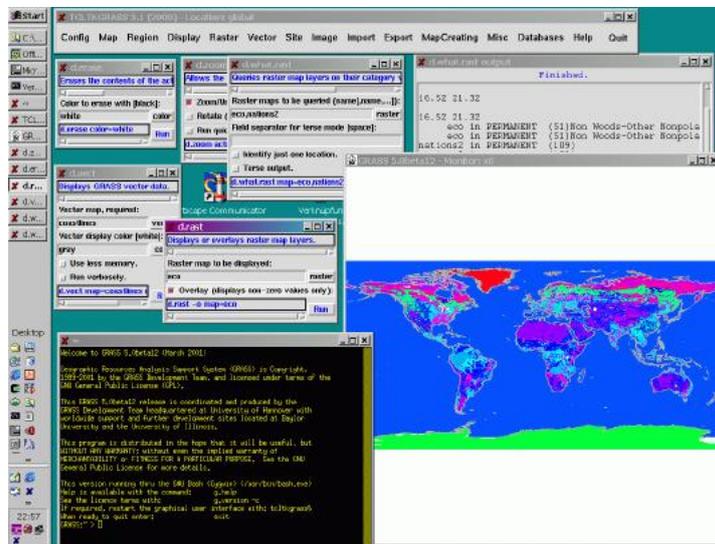


Figura 3.2: GRASS sotto Windows/Cygnus

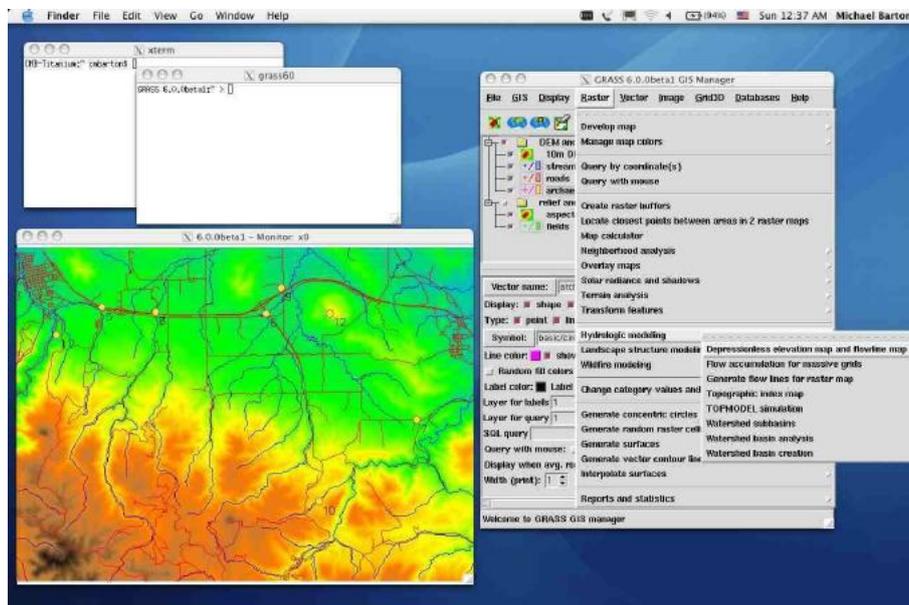


Figura 3.3: GRASS sotto Mac OS X

3.4.4 Organizzazione dei dati in GRASS

In GRASS i dati sono contenuti in una directory Unix definita DATABASE (ad esempio \$HOME/grass.data). Il DATABASE è suddiviso in sottodirettori definiti LOCATIONS. Ciascuna LOCATION rappresenta un DATABASE indipendente la cui denominazione viene definita dall'utente.

Un criterio per definire la suddivisione di un DATABASE in LOCATIONS potrebbe essere ad esempio quello geografico. Pertanto uno studio condotto sull'intero territorio italiano, potrebbe avere le LOCATIONS rappresentate dalle diverse Regioni italiane. Ciascuna LOCATION è a sua volta suddivisa in sottodirettori definiti MAPSETS (Fig. 3.4):

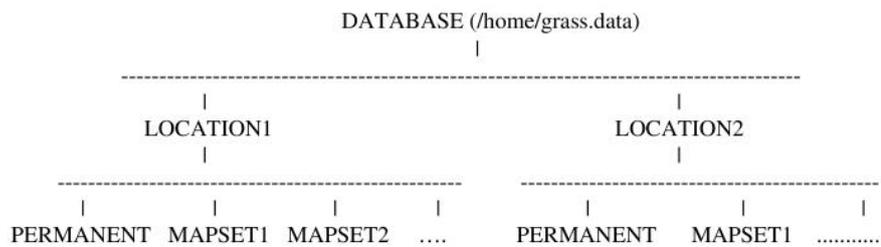


Figura 3.4: Schema di un database Grass

Ogni volta che viene creata dall'utente una nuova LOCATION, viene creato automaticamente un Mapset denominato PERMANENT, che può essere modificato solo dal creatore della LOCATION e contiene diverse informazioni, quali ad es. le coordinate della regione di lavoro e la risoluzione delle carte raster. Gli altri MAPSET della LOCATION sono creati dall'utente, in numero illimitato e con qualunque nome. La suddivisione in MAPSET può essere utilizzata per separare i diversi tematismi all'interno di una stessa LOCATION. Nell'esempio precedente per ogni Regione potrebbe risultare utile destinare alle caratteristiche altimetriche, vegetazionali o idrografiche specifici MAPSET. Oppure può essere utilizzato per distinguere l'accesso a diversi utenti, a ciascuno dei quali può essere destinato uno specifico MAPSET.

La regione di lavoro (definita region in GRASS) è l'area geografica di forma rettangolare, definita dalle coordinate dei bordi del rettangolo, all'interno

della quale hanno effetto le operazioni eseguite da GRASS. Una regione deve essere definita in concomitanza con la creazione di una Location e viene assunta come regione di default. Tale regione non può essere modificata ed è sempre disponibile. E' tuttavia possibile definire altre regioni durante una sessione di GRASS, che possono essere salvate e richiamate in qualunque momento.

3.4.5 Rappresentazione dei dati

Per l'importazione e l'esportazione dei dati raster GRASS utilizza principalmente un modulo valido per numerosi formati che vengono gestiti tramite la libreria OGR/GDAL ed altri moduli specializzati per i file ASCII Grid, binary, ESRI, ma fondamentale premessa: la risoluzione e l'estensione geografica di un raster ottenuto in GRASS è determinata da come queste sono state impostate precedentemente tramite il comando *g.region* e sono quindi indipendenti da quelle dei vettori o dei raster di input che vengono ricampionati in automatico per adattarne la risoluzione in base alle impostazioni definite.

In GRASS ad ogni oggetto di una carta raster, o meglio alle celle che compongono l'oggetto, viene attribuito un solo valore di categoria ed eventualmente un'etichetta descrittiva. La non assegnazione di una categoria comporta l'assegnazione automatica di un valore particolare definito *null*. Un'immagine vettoriale è costituita da un insieme di caratteristiche geometriche e di attributi.

La geometria di un oggetto, riproducibile mediante le forme geometriche elementari punto, linea e poligono, è riconducibile ad una delle cinque tipologie seguenti:

point: per rappresentare gli elementi puntiformi (punti quotati, pozzi, ecc.); ogni punto è rappresentato da una coppia di coordinate;

line: per rappresentare elementi lineari (strade, fiumi, linee elettriche, ecc.); ogni linea è costituita da una spezzata; i due punti alle estremità sono definiti nodi, mentre i punti intermedi sono definiti vertici; ogni linea viene memorizzata mediante una sequenza di coppie di coordinate in corrispondenza dei nodi e dei vertici della spezzata;

boundary: per rappresentare, anche parzialmente; i bordi di aree chiuse (laghi, piante di edifici, appezzamenti di terreno, ecc.); è una spezzata memorizzata mediante una sequenza di coppie di coordinate in corrispondenza dei nodi e dei vertici;

centroid: è un elemento puntiforme all'interno di un'area chiusa in corrispondenza del quale vengono inseriti gli attributi dell'area; Ogni centroide è rappresentato da una coppia di coordinate;

area: è l'insieme di un bordo di area (boundary) completo, vale a dire chiuso, e del relativo centroide (centroid).

La geometria è salvata in uno specifico formato vettoriale di GRASS (formato nativo). E' comunque possibile importare e trasformare molti altri formati (ArcInfo-Coverages, CSV, DGN, SHAPE files, GIS ManagerL, MapInfo, MySQL, ODBC, OGD, PostgreSQL/PostGIS, S57, SDTS, TIGER, UK .NTF,VRT). In alternativa, la geometria può anche essere salvata in un database PostGIS. Inoltre alcuni formati GIS, come ad es. i files SHAPE, possono essere utilizzati direttamente senza la conversione di formato.

Nel file della geometria ogni oggetto è caratterizzato da un numero di categoria, che può essere comune a più oggetti. Tutti gli attributi delle immagini vettoriali sono salvati in table all'interno di un database e collegati alle caratteristiche geometriche mediante una DBMI (DataBase Management Interface). Attualmente sono gestibili in GRASS i seguenti DBMS (Database Management Systems): DBF (DataBase File), PostgreSQL, MySQL, ODBC, mediante i driver denominati dbf, pg e ogr. Il formato DBF il cui driver, dbf, è sempre presente, rappresenta il driver di default.

Una table è costituita da un insieme di colonne, ciascuna delle quali rappresenta un attributo, e da tante righe quanti sono gli oggetti presenti nella carta (o i gruppi di oggetti che nel file della geometria hanno lo stesso valore di categoria). La connessione fra ciascun oggetto vettoriale (punto, linea, poligono) nel file della geometria ed i suoi attributi nella table è effettuato tramite il numero di categoria (Fig. 3.5).

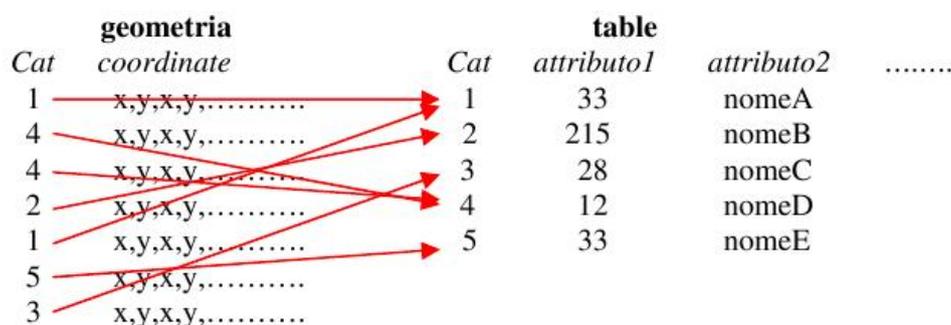


Figura 3.5: Connessioni tra oggetti Vettoriali e categorie

In molte circostanze, quando si crea una carta vettoriale viene automaticamente creata la table degli attributi a cui viene assegnato lo stesso nome della carta. Dal momento che il linguaggio di interrogazione delle table, SQL, non consente l'utilizzo del punto all'interno del nome di una table, anche il nome della carta vettoriale è soggetto alla stessa restrizione. I formati vettoriale e raster sono salvati in direttori diversi, e sono gestiti mediante comandi diversi, per cui è possibile assegnare lo stesso nome ad una carta vettoriale e ad una raster senza problemi di conflittualità.

3.4.6 Strumenti

Analisi Spaziale Operazioni di notevole interesse nel campo raster sono quelle relative all'elaborazione dei modelli digitali del terreno (DTM) ad esempio per la costruzione di carte della pendenza e dell'esposizione. Come esempio proponiamo l'uso del comando *r.slope.aspect* da shell o da interfaccia grafica .

Come primo passo occorre definire la risoluzione di lavoro e quindi del raster da creare (ad esempio *g.region res=50*). Nella relativa finestra occorrerà inserire il nome del raster relativo alla morfologia del territorio, il nome del raster delle pendenze, il modo in cui esprimere la pendenza (format for reporting the slope: degrees o percent), la precisione numerica dei raster da creare (type of output...) ed il nome del raster delle esposizioni (output aspect file name). È possibile utilizzare tale modulo anche per calcolare i

raster relativi al profilo della curvatura, della tangente, la derivata parziale di primo e secondo grado per le variazioni lungo l'asse delle x ed delle y e la derivata di secondo grado rispetto ad entrambe le coordinate.

La Geostatistica Peculiarità della geostatistica è la presenza di un software dedicato, affidabile e robusto, con cui svolgere tali analisi. Il mondo Open Source ormai mette a disposizione tutto questo con l'ambiente statistico R. Questo software è attualmente totalmente integrato in GRASS.

L'unione GRASS + R risultano uno strumento formidabile per qualsiasi tipo di analisi, paragonabile se non migliore agli applicativi proprietari, permettendo all'utente qualsiasi tipo di elaborazione a costi praticamente nulli.

È possibile avviare R dalla shell di GRASS, importare e esportare dati dai formati GRASS e lanciare comandi per entrambi i programmi. Questa modalità di lavoro permette di ottenere il massimo delle potenzialità su entrambi i fronti, non ultimo programmare le elaborazioni in un codice unico. Il plugin GRASS infatti mette a disposizione la maggior parte degli strumenti di GRASS in modo user-friendly, consentendo di sviluppare applicazioni anche complesse e di renderle disponibili a utenti finali non necessariamente interessati a un utilizzo diretto di GRASS o di R.

Geoprocessing Il termine *geoprocessing* è in realtà un termine tipico del linguaggio Arcview che indica appunto le operazioni di analisi spaziale. In realtà anche l'operazione di dissoluzione è un'operazione spaziale. La caratteristica forse più interessante dei GIS non è quella legata alla semplice archiviazione, interrogazione e rappresentazione cartografica delle informazioni, ma quella di generare, attraverso operazioni di analisi spaziale, delle nuove informazioni sul sistema insediativo e sul territorio, derivate da informazioni di base presenti nel sistema, ma che aggiungono nuove conoscenze, non disponibili prima dell'elaborazione. Gli strumenti di analisi spaziale possono essere di tipo:

- analisi semplici: utilizzando funzioni di base presenti sul software, fornendo delle informazioni importanti e nuove, ma di tipo statico;

- analisi complesse: fino a configurare un modello di simulazione o, nei casi più evoluti (quando si introduce la possibilità di interagire con la procedura, e di introdurre ipotesi del tipo: cosa succederebbe se ...) un vero e propri Sistema di Supporto alle Decisioni (SSD, o DSS).

Le operazioni di Geoprocessing sono tecniche di analisi basate sul formato vettoriale e utilizzate per la derivazione di nuovi dati a partire dai dati in entrata. Tra le diverse tecniche di geoprocessing, *dissolve*, *clip* e *merge* sono molto utilizzate dagli analisti GIS.

Dissolve : questa operazione permette di aggregare degli oggetti basandoci sui valori (ripetuti) presenti in un campo. Ad esempio, consideriamo il seguente tema (Fig. 3.6, Fig. 3.7, Tab. 3.2).

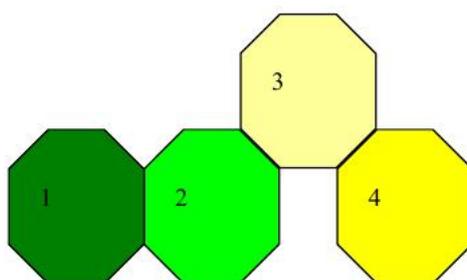


Figura 3.6: Classi specifiche

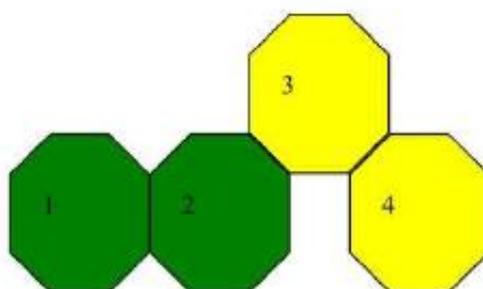


Figura 3.7: Classi generali

Ogni oggetto appartiene sia a una classe specifica che a una generale. La legenda è basata sulle classi specifiche.

Creiamo ora una legenda .

ID	Classe specifica	Classe generale
1	Bosco di latifoglie	Bosco
2	Bosco di conifere	Bosco
3	Prateria	Formazione aperta
4	Seminativo	Formazione aperta

Tabella 3.2: Classi di appartenenza

Tra poligoni appartenenti a una stessa classe c'è un confine che li separa. Questo avviene quando applichiamo una legenda general a oggetti gerarchicamente più specifici, ovvero quando viene applicata una legenda di ordine gerarchico superiore a quello degli oggetti digitalizzati. Possiamo dissolvere tali linee di confine (outline) attraverso l'operazione dissolve, basata sul campo Classe generale. I poligoni presenti sono ora solo due (fig. 3.8, tab. 3.3).

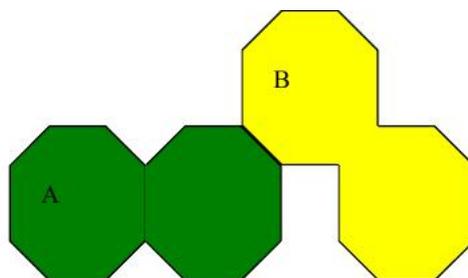


Figura 3.8: Operazione dissolve

ID	Classe generale
A	Bosco
B	Formazione aperta

Tabella 3.3: Legenda generale

Clip: questa operazione permette di “tagliare” uno strato (tema) di oggetti geografici rispetto a un altro strato (tema) (fig. 3.9).

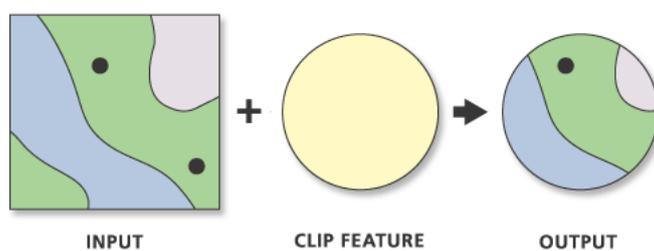


Figura 3.9: Clip

Merge: l'operazione di Merge permette di creare un unico strato (tema) da due o più strati in entrata (fig. 3.10).

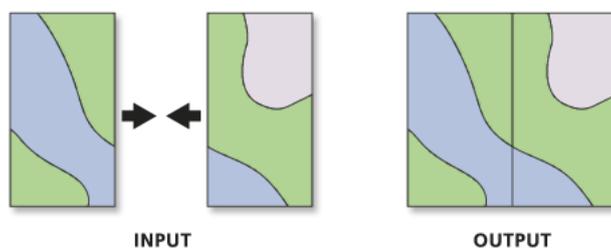


Figura 3.10: Merge

L'operazione di Merge rappresenta l'inverso della "selezione di oggetti geografici"; in quel caso da un set di oggetti veniva creato un subset condizionato (tramite selezione) di oggetti. In questo caso vengono fusi due strati (adiacenti o meno) in un unico strato.

Capitolo 4

Studi sul sito SOGIN di Bosco Marengo

4.1 Premessa

L'impianto ex FN di Bosco Marengo, in provincia di Alessandria, di proprietà SOGIN, ha operato dal 1973 al 1995 per il riprocessamento di combustibili per le centrali nucleari italiane (ricariche della centrale di Garigliano, prima carica e ricariche per Caorso, ricariche per Trino) e per reattori esteri. Alla fine del 1995 l'ENEA, azionista pressoché esclusivo della FN, allora proprietaria dell'impianto, decise di non proseguire ulteriormente le attività di fabbricazione di combustibile nucleare e di procedere alla disattivazione dell'impianto. Nel 1996 è stato presentato un piano di disattivazione, reiterato da SOGIN nel luglio 2003 ed ora prossimo alla conclusione. Le operazioni di decommissioning riguardano lo smantellamento, la decontaminazione meccanica e il trattamento di apparecchiature che servivano per la produzione di combustibile nucleare. Nell'impianto, quando sono state fermate le attività di fabbricazione, vi erano stoccate circa 112 tonnellate di combustibile nucleare. Il materiale è stato tutto allontanato e trasferito all'estero. L'ultimo trasporto è avvenuto nel novembre 2006. Il programma attuale prevede il completamento dello smantellamento degli impianti e il condizionamento dei rifiuti entro il 2009. L'impianto FN di Bosco Marengo sarà il primo impianto

gestito da Sogin a concludere le attività di decommissioning.

Questo dipartimento si è occupato di stimare dosi da radiazione e rischio per operatori durante le fasi di decommissioning. Si sono anche analizzate possibili situazioni incidentali durante le operazioni di dismissione, le quali potessero portare al rilascio di radionuclidi all'esterno dell'impianto, valutando i rischi potenziali per la popolazione. Gli strumenti GIS si sono rivelati a questo fine uno strumento imprescindibile.

4.2 Dati utilizzati

La CTR del Piemonte Si tratta della versione destrutturata del database della Carta Tecnica Regionale (CTR) in versione di "Puro Disegno" destinata alla produzione di output di qualità e alla consultazione al continuo della carta. E' costituita da 800 Sezioni della CTR del Piemonte in scala 1/10.000 in formato raster e in formato vettoriale, la realizzazione delle quali è stata eseguita attraverso la digitalizzazione delle carte tecniche del territorio regionale nel sistema di riferimento WGS84/UTM. In sostanza costituisce una rielaborazione dal formato DXF organizzata in file di disegno per AutoCAD (DWG) trutturata in pochissimi layer tematici, già correttamente predisposta in un formato standard di vestizione cartografica (simboli, spessori, campiture, ..).

RA_alluvione2000 Si tratta di fotogrammi della ripresa aerea a colori con copertura stereoscopica, in scala 1/15.000 realizzata dalla Regione Piemonte a seguito dell'alluvione dell'ottobre 2000 (periodo di acquisizione: ottobre 2000-ottobre 2001; quota d'acquisizione 2880 metri). I fotogrammi, di risoluzione 600 dpi, sono distribuiti in formato ECW (Enhanced Compression Wavelet, della Earth Resource Mapping), una particolare struttura dati, impiegata nella memorizzazione di immagini georiferite di grande estensione che consente elevati rapporti di compressione associati a notevoli velocità di elaborazione. Nell'insieme, forniscono una copertura totale del territorio del Piemonte. Organizzati in un database relazionale, possono essere visualizzati interattivamente attivando col mouse i singoli tasselli.

Dati del Consorzio per il Sistema Informativo del Piemonte (CSI)

Sono layer relativi alle caratteristiche territoriali di tutto il Piemonte, contenenti informazioni grafiche, georeferenziate nel sistema di riferimento WGS84/UTM e in scala 1:10000, che rappresentano strade, aree edificate, fiumi ecc., importanti nel processo di caratterizzazione del territorio.

4.3 Importazione e georeferenziazione delle immagini Satellitari

GRASS presenta numerosi moduli che lo rendono in grado di georeferenziare raster ed immagini di vario tipo:

- mappe scannerizzate non referenziate: tramite quattro punti definiti (*i.group*, *i.target*, *i.points*, *i.rectify*);
- immagini satellitari da sensori ottici e radar non referenziate: tramite un certo numero di punti definiti di controllo sul terreno (*i.group*, *i.target*, *i.points*, *i.rectify*);
- ortofoto basandosi su modelli digitali del terreno (DEM): *i.ortho.photo*;
- georeferenziazione di foto digitali ordinarie: procedura modificata di *i.ortho.photo*.

I punti di controllo sul terreno (GCPs) sono scelti manualmente (un metodo automatizzato di ricerca dei punti di controllo è attualmente in sviluppo) e le immagini sono rettificate in modo sequenziale nel database di geocodifica. Sia mappe raster che vettoriali possono essere usate come mappe di riferimento.

Le immagini satellitari pervenute in formato ECW sono state convertite nel formato TIFF e successivamente importate e georeferenziate, prendendo come punti di riferimento quelli appartenenti alla Carta Tecnica Regionale.

Per la georeferenziazione delle immagini satellitari (raster) [23, 24], si prende come riferimento la CTR, l'obiettivo è quello di georeferenziare la prima sulla seconda individuando dei 'control points' che individuino gli stessi punti sia sull'immagine raster sia su quella vettoriale. Si procede prima

all'individuazione della zona di riferimento sulla CTR e con zoom successivi si isola la zona (Fig. 4.1)

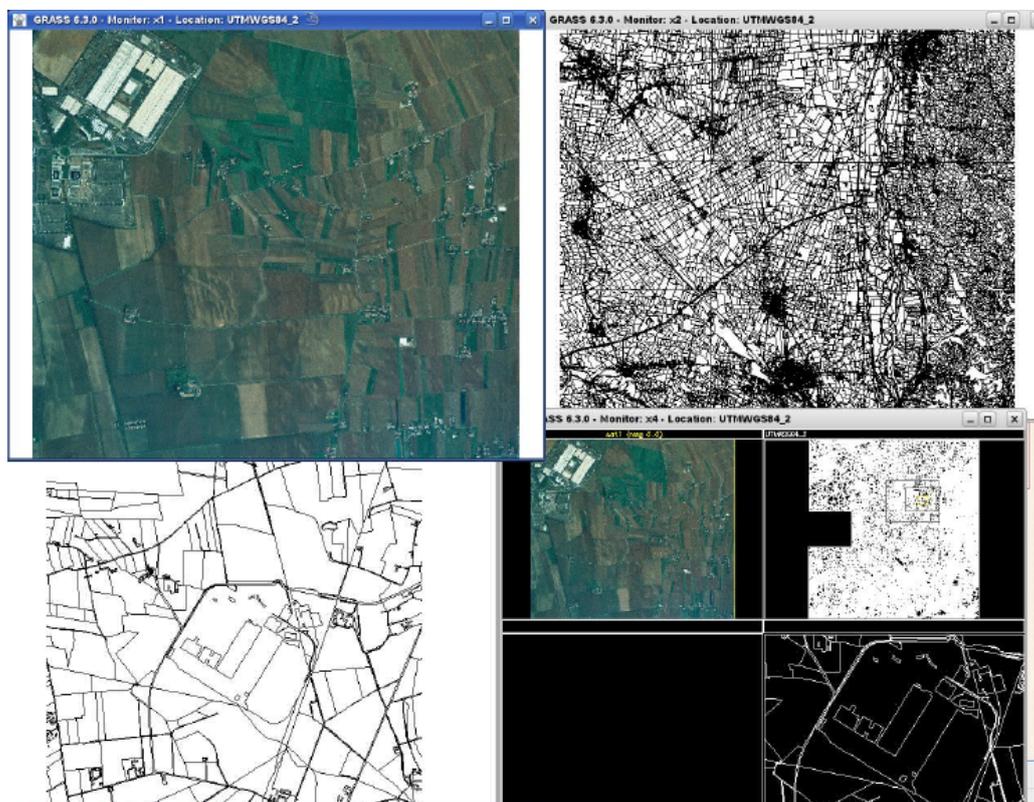


Figura 4.1: Percorso di georeferenziazione

Si procede poi ad individuare il maggior numero possibile di contro points, che in figura 4.2 sono evidenziati in verde.

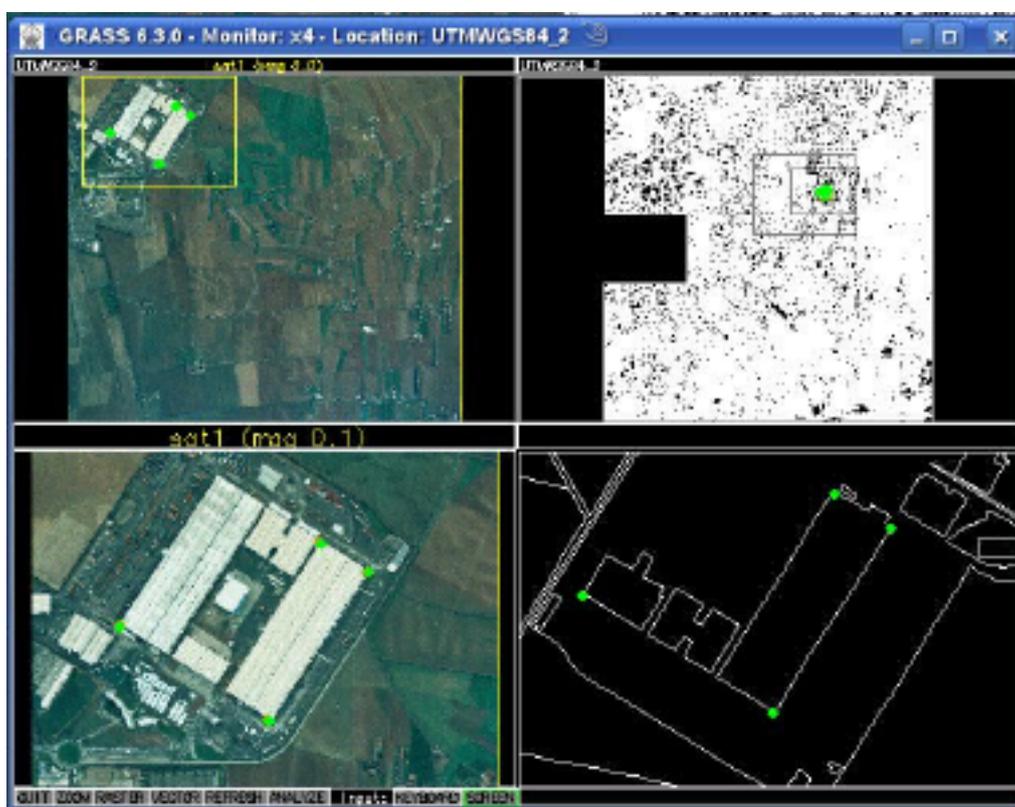


Figura 4.2: Selezione dei 'control points'

Conclusa questa operazione si applica la funzione *i.rectify* e si ottiene il risultato voluto (Fig. 4.3).



Figura 4.3: Immagine georeferenziata

Ripetendo la stessa procedura per tutte le immagini che occorrono, si ottiene la copertura di tutta l'area d'interesse. pertanto per quanto riguarda la georeferenziazione, GRASS offre un metodo più laborioso rispetto ad ArcWiew, ma ugualmente efficace.

4.4 Overlay dei layer significativi e creazione del database

I layer generalmente in formato vettoriale, rappresentano le carte tematiche alle quali è possibile relazionare delle tabelle contenenti gli attributi. Nel caso in esame abbiamo a disposizione i seguenti vettoriali:

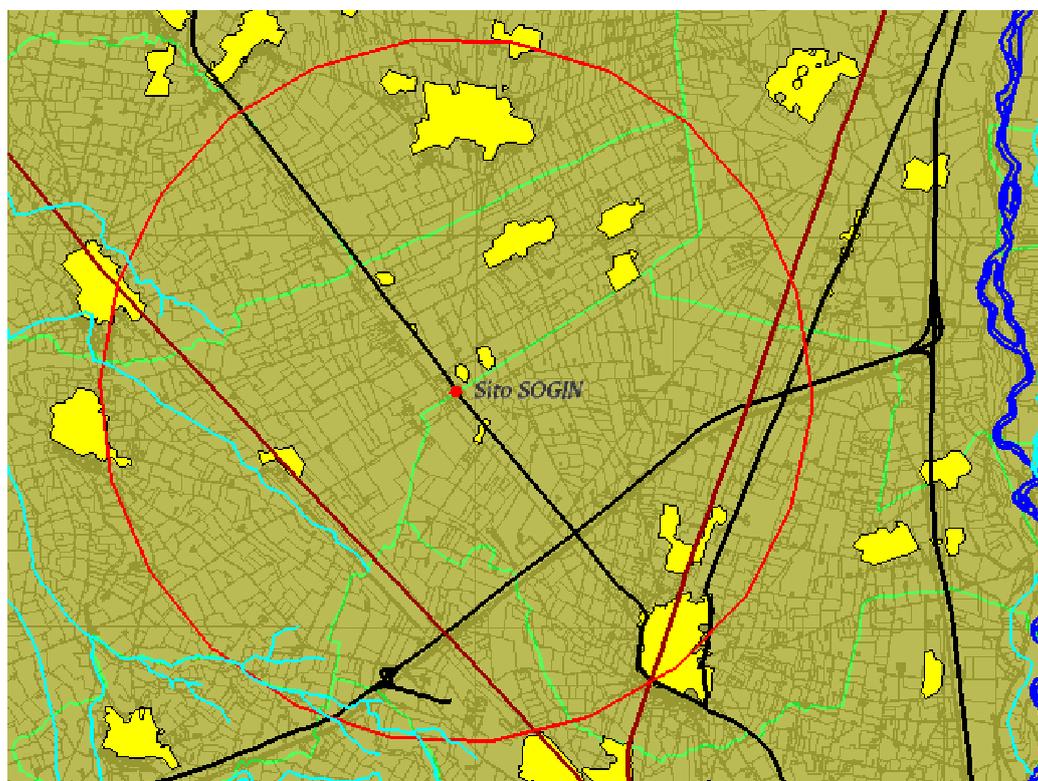


Figura 4.4: Overlay dei layer vettoriali

1. Strade principali e autostrade.
2. Comuni
3. Fiumi principali e secondari
4. Zone edificate
5. Ferrovie

Effettuando un overlay die layer appena menzionati includendo la CTR, si ottiene la carta in figura 4.4:

Come si può osservare, è stato creato un layer che individua la posizione del sito SOGIN ed è stata creata attorno a questo un area del raggio di

cinque chilometri (evidenziata in rosso), all'interno della quale sarà necessario individuare il maggior numero di informazioni possibili relativamente ad ogni layer, in modo da poter popolare il database corrispondente.

Il database utilizzato è PostgreSQL che può essere interfacciato con GRASS anche direttamente. Una volta creato il database lo si connette alla Location in cui stiamo lavorando e tutti i file creati successivamente avranno una tabella corrispondente nel database. Ogni tabella può essere manipolata a piacere, aggiungendo o rimuovendo sia dati sia colonne o righe. Pertanto per il layer Comuni la tabella Comuni conterrà informazioni del tipo: estensione, numero di abitanti, tipo di coltivazioni e allevamenti. La tabella relativa alle zone edificate invece sarà più specifica riguardo il numero di abitanti in base al sesso e all'età, la distanza dal sito, ecc.

Una volta compilato il database sarà possibile interrogare la mappa interattivamente, in modo da ottenere informazioni puntuali o generali in base al layer interrogato.

Per completare questa fase si è reso necessario creare un layer vettoriale che riproducesse in ogni sua parte l'impianto costituito da diversi edifici. In tal modo assegnando ad ogni edificio una categoria diversa si è creata una table che riporta tutti i dati d'interesse per ogni area (Fig. 4.5).



Figura 4.5: Layer dell'impianto

4.5 Implementazione dei dati relativi alle dosi di radiazione e creazione del layer Rosa dei Venti

Il codice utilizzato per la determinazione dei valori di dose è il GENII-LIN sviluppato presso il Laboratorio di Montecuccolino dell'Università di Bologna, in collaborazione con il Pacific Northwest National Laboratory (Richland WA).

Il sistema GENII-LIN offre: la possibilità di calcolare la dose di radiazione sia per i rilasci acuti sia per quelli cronici, con le opzioni per la dose totale, la dose impegnata e la dose equivalente; la possibilità di fare stime in base alle vie di esposizione (Exposure Pathways), compresa l'esposizione diretta, attraverso l'acqua (spiagge, canottaggio e pesca), il terreno e l'aria (nube semi-infinita e modello limitato della nube), l'inalazione e l'ingestione.

I tipi di rilascio considerati sono: rilascio acuto in aria, dal livello del suolo, da punti elevati o dall'acqua; rilascio cronico in aria, dal livello del suolo, da punti elevati o dall'acqua; contaminazione iniziale del terreno o della superficie. Il codice calcola i valori di dose secondo le sedici direzioni della rosa dei venti a diverse distanze, per questo motivo è stato necessario creare il layer Rosa dei Venti. La rosa dei venti è la rappresentazione schematica dei punti cardinali: Nord, Sud, Est e Ovest e delle sedici direzioni da questi determinate: N, NNE, NE, NEE, ecc...

Per creare questo layer è stato scritto un programma in Python che, permette di volta in volta di inserire le coordinate della sorgente e crea il vettoriale della rosa dei venti centrato esattamente nel punto voluto (Fig. 4.6).

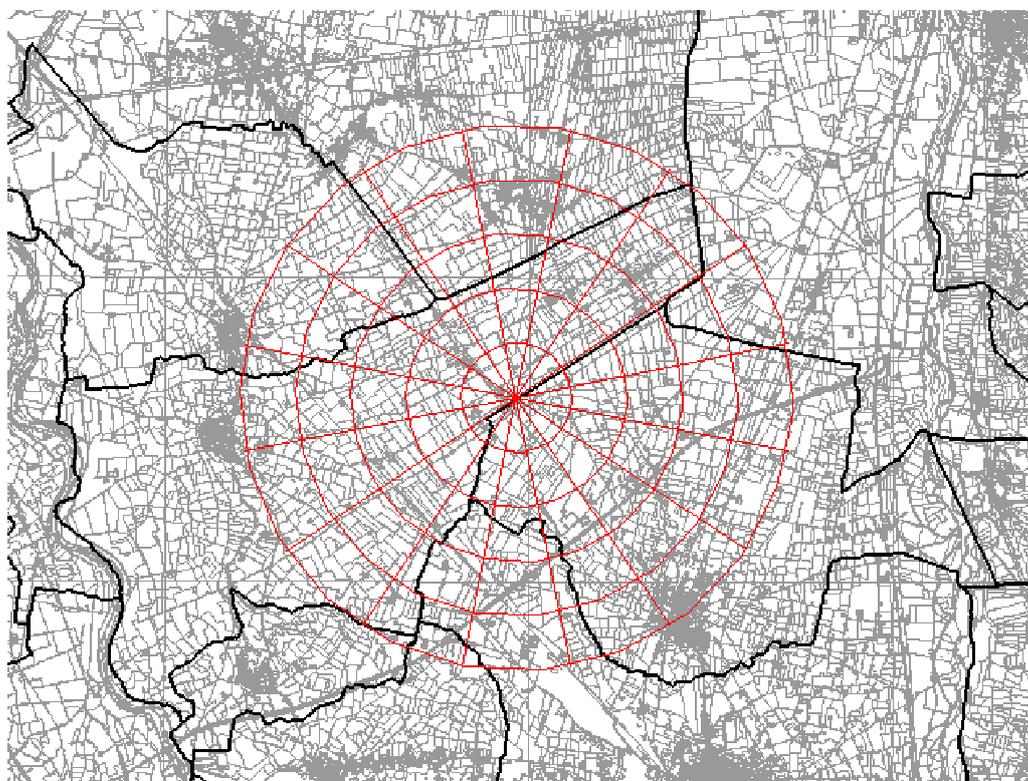


Figura 4.6: Layer Rosa dei Venti

I valori delle dosi, relativi ad ogni direzione della rosa dei venti, e posti in in un'apposita cartella, vengono letti e trasformati in attributi puntuali. Quindi il file vettoriale di punti creato, viene trasformato in un file raster i cui punti vengono successivamente interpolati attraverso il modulo IDW (Inverse Distance Weighted), che è un interpolatore deterministico.

In figura 4.7 è mostrato il risultato finale.

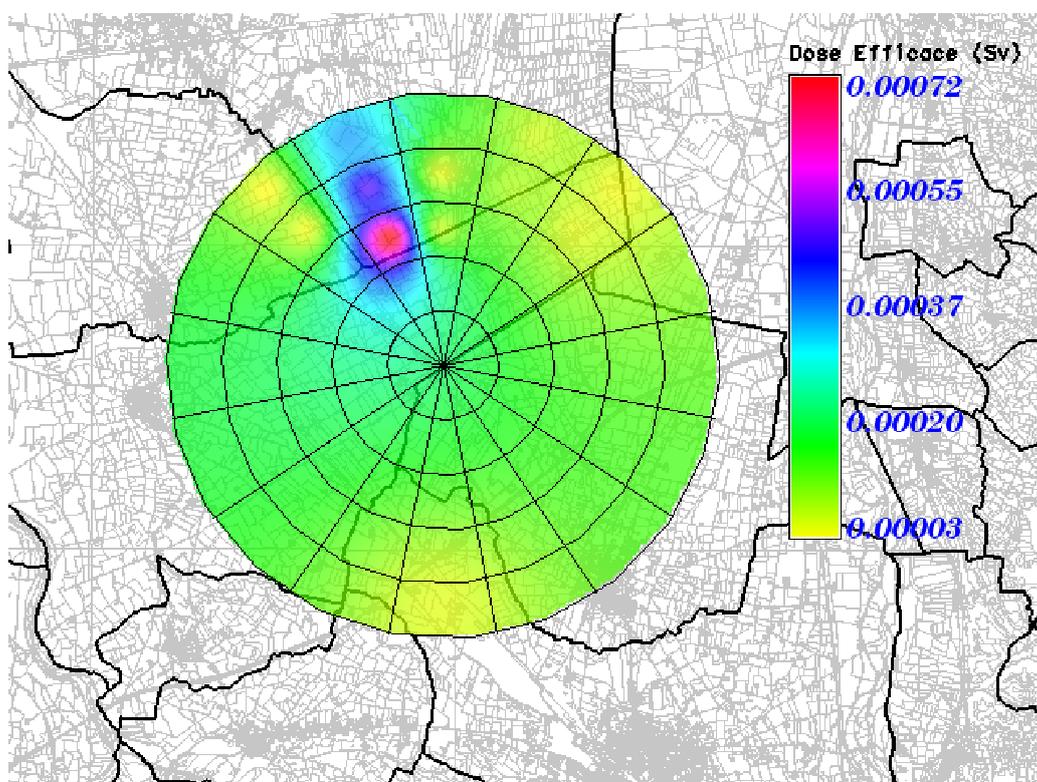


Figura 4.7: Concentrazione di dosi in seguito ad un rilascio acuto in aria

4.6 Conclusioni

Volutamente, non siamo entrati nel dettaglio delle situazioni incidentali analizzate, né abbiamo voluto quantificare i risultati delle nostre analisi. Non è lo scopo di questa trattazione. Abbiamo invece voluto mostrare come gli strumenti GIS, in particolare il GRASS, siano efficacissimi nel gestire le informazioni.

Vale la pena aggiungere che lo stesso studio è stato condotto con il software commerciale ArcGIS. GRASS si è dimostrato parimenti efficace e ha permesso di raggiungere i risultati con un notevole abbattimento dei costi.

Capitolo 5

Valutazione della Qualità dell'Aria

5.1 Il ruolo dell'Aerosol nelle problematiche di inquinamento atmosferico

L'aerosol o particolato atmosferico è composto da particelle e corpuscoli in sospensione all'interno dell'atmosfera la cui natura chimica è variabile e dipende dall'origine dell'aerosol, può essere costituito da aggregati molecolari o da minutissimi frammenti di materiale solido di diversa natura fissati insieme per adesione. Può essere immesso in atmosfera dall'azione dei venti sui deserti con conseguente trasporto di sabbie sottilissime, oppure dall'azione dei venti sugli oceani con conseguente trasporto di spume marine. Grandi quantità di aerosol vengono immesse dai vulcani oppure assieme alle emissioni industriali e da combustibile dovute all'attività umana. Hanno la capacità di assorbire e diffondere la radiazione solare e agiscono da centro di nucleazione per il vapore acqueo, contribuendo così alla formazione delle nubi modificandone le proprietà ottiche.

Gli aerosol giocano un ruolo importantissimo nel processo di scattering della radiazione solare, essi infatti rappresentano i centri di diffusione di questa che, quando li incontra, viene "deviata" in direzione diversa da quella di incidenza e questo in base alla forma, concentrazione e composizione del-

le particelle, e all'andamento della loro distribuzione spaziale e temporale. Dunque contribuiscono in maniera determinante a una redistribuzione del campo energetico-radiativo di natura solare mentre soltanto in piccola parte contribuiscono all'assorbimento di tale radiazione.

Lo studio della composizione chimica dell'aerosol in funzione della sua distribuzione dimensionale può fornire informazioni sui processi di formazione, trasporto, diffusione e abbattimento del particolato atmosferico nell'ambiente. La dimensione delle particelle è fortemente correlata agli effetti tossicologici delle sostanze contenute, controllandone i tempi di persistenza nell'atmosfera e la loro capacità di penetrare all'interno dell'apparato respiratorio. Inoltre, è di rilevante interesse poter correlare il carico e la composizione del particolato atmosferico alle condizioni meteorologiche, con particolare riguardo all'individuazione del trasporto delle masse d'aria dalle aree sorgenti e alla valutazione dei processi di abbattimento (effetto del vento e delle precipitazioni).

Gli effetti che gli aerosol hanno sulla salute sono strettamente legati alla dimensione e composizione delle particelle che li compongono.

Il Particolato Totale Sospeso (PTS) è costituito da particelle con dimensioni differenti: il diametro delle particelle può variare da alcune frazioni di micron ad alcune decine di micron, le particelle ritenute dannose a livello sanitario sono quelle più fini perché riescono a penetrare nelle parti più profonde delle vie respiratorie.

Il Particolato Fine è uno degli inquinanti seguito con maggiore attenzione per le implicazioni sanitarie ad esso collegate.

La sigla PM10 (particulate mater) identifica materiale presente nell'atmosfera in forma di particelle microscopiche respirabili, il cui diametro aerodinamico medio è uguale o inferiore a 10 μm , quindi in grado di penetrare nel tratto superiore dell'apparato respiratorio (dal naso alla laringe). Esse variano molto sia per forma che per composizione chimica in quanto dipendono dalle fonti di emissioni dominanti, cioè dal tipo di insediamenti della zona ma, poiché possono rimanere sospese nell'aria per parecchi giorni ed essere quindi trasportate anche a grandi distanze, il carico di PM10 in una città dipende non solo da fonti locali ma anche dalla quota trasportata. In

inverno i valori di PM10 sono superiori a quelli estivi così come aumentano con la nebbia e con l' assenza di vento . La loro rimozione avviene , nel caso di particelle grossolane , per deposizione secca , cioè esse cadono per gravità . Per le altre si tratta di deposizione umida ad opera delle nubi e della pioggia. Le principali fonti di PM10 sono:

- Sorgenti naturali: l'erosione del suolo, gli incendi boschivi, le eruzioni vulcaniche, la dispersione di pollini, il sale marino.
- Sorgenti legate all'attività dell'uomo: processi di combustione (tra cui quelli che avvengono nei motori a scoppio, negli impianti di riscaldamento, in molte attività industriali, negli inceneritori e nelle centrali termoelettriche), usura di pneumatici, freni ed asfalto.

Una parte rilevante del PM10 presente in atmosfera deriva dalla trasformazione in particelle liquide di alcuni gas (composti dell'azoto e dello zolfo) emessi da attività umane.

La nocività delle polveri sottili dipende dalle loro dimensioni e dalla loro capacità di raggiungere le diverse parti dell'apparato respiratorio:

- oltre i 7 μm : cavità orale e nasale
- fino a 7 μm : laringe
- fino a 4,7 μm : trachea e bronchi primari
- fino a 3,3 μm : bronchi secondari
- fino a 2,1 μm : bronchi terminali
- fino a 1,1 μm : alveoli polmonari

I valori limite sono definiti in Italia dal decreto-legge nr. 60 del 2 aprile 2002; tale decreto fissa due limiti accettabili di PM10 in atmosfera:

- Il primo è un valore limite di 50 $\mu\text{g}/\text{m}^3$ come valore medio misurato nell'arco di 24 ore da non superare più di 35 volte/anno.
- Il secondo come valore limite di 40 $\mu\text{g}/\text{m}^3$ come media annuale

Il PM_{2.5} è la frazione più fine del PM₁₀, costituita dalle particelle con diametro uguale o inferiore a 2,5 µm, è il particolato più pericoloso per la salute infatti è in grado di penetrare profondamente nei polmoni. Questo particolato può rimanere sospeso nell'atmosfera per giorni o settimane a differenza delle particelle maggiori che rimangono in atmosfera da poche ore a pochi giorni.

5.2 Il Telerilevamento da satellite come supporto per le Valutazioni della Qualità dell'Aria : il MODIS

La conoscenza del contenuto di particolato fine ed ultra fine presente in atmosfera ricopre un ruolo fondamentale per una valutazione adeguata della qualità dell'aria, finalizzata ad azioni mirate alla mitigazione degli effetti sulla salute umana e sull'impatto ambientale, da parte degli enti preposti .

Attualmente in Italia l'attività di controllo e valutazione della qualità dell'aria è gestita dall'ARPA (Agenzia Regionale per la Protezione Ambientale) e si affida principalmente all'utilizzo di una serie di stazioni di rilevamento distribuite sul territorio nazionale, che sono connesse via rete ai centri provinciali di elaborazione e validazione dati, e ai diversi centri di archiviazione delle informazioni a livello regionale. In ognuna delle stazioni sono installati degli strumenti che registrano in continuo i principali inquinanti presenti nell'aria; un computer industriale e un modem provvedono all'acquisizione, alla memorizzazione e all'invio dei dati al Centro di Gestione che, dopo aver effettuato il controllo di qualità, li diffonde al pubblico, alla Regione e agli Enti locali.

Questo sistema di valutazione effettua misurazioni puntuali di concentrazione e non possiede gli strumenti necessari per fornire la distribuzione di inquinanti in un'intera area.

Negli ultimi anni è stata sviluppata una nuova metodologia di rilevamento che contrariamente alle misurazioni effettuate da postazioni fisse a terra, è in grado di fornire vere e proprie mappe di concentrazione di particolato relative

ad intere zone geografiche. Il metodo consiste nell'analisi dello spessore ottico dell'aerosol o AOT (Aerosol Optical Thickness) utilizzando strumentazioni satellitari.

Creati appositamente per il programma EOS (Earth Observing System), che ha come principale obiettivo migliorare la conoscenza sui cambiamenti globali terrestri attraverso l'osservazione continua del pianeta, i satelliti TERRA e AQUA sono il risultato di una joint venture tra l'agenzia spaziale americana NASA e quella giapponese NASDA. Il primo sorvola la Terra da nord a sud, passando l'Equatore di mattina, mentre Aqua passa da sud a nord nel pomeriggio. Volano su un'orbita eliosincrona quasi-polare, ad un'altezza di 705 km dalla Terra. Tale orbita comporta due passaggi giornalieri per ciascun satellite, approssimativamente alla stessa ora solare ogni giorno. AQUA passa sull'Italia all'incirca alle 01:30 a.m. e p.m., e i suoi passaggi sono complementari a quelli del satellite TERRA, che con strumenti simili fornisce misure alle 10:30 a.m. e p.m. Questi orari rappresentano un buon campionamento per il ciclo giornaliero della temperatura.

Su entrambi i satelliti è montato il MODIS (Moderate Resolution Imaging Spectroradiometer) [26, 27, 28, 29], uno spettroradiometro a 36 bande che misura la radiazione nel visibile e nell'infrarosso, derivandone prodotti relativi alla vegetazione, alla tipologia della superficie, alla fluorescenza della clorofilla sugli oceani, alle proprietà delle nubi e degli aerosol, agli incendi, alla neve su terra e ai ghiacci su mare. Il primo MODIS fu lanciato a bordo del satellite Terra nel Dicembre 1999, il secondo su Aqua nel maggio 2002. Da allora i loro dati globali sono disponibili e utilizzati negli studi per la comprensione delle dinamiche globali e dei processi relativi alla terra, agli oceani e agli strati più bassi dell'atmosfera. L'utilizzo del MODIS riveste inoltre un ruolo fondamentale nello sviluppo di modelli interattivi capaci di produrre previsioni riguardo i cambiamenti globali. E' stata dedicata una cura particolare alla creazione e alla distribuzione dei prodotti, cercando di facilitarne al massimo la fruizione degli utenti finali nei settori della ricerca e delle applicazioni. A tale scopo, i ricercatori della NASA, riuniti nel MODIS Team, hanno sviluppato prodotti specifici, direttamente dai dati grezzi ricevuti dai due sensori, suddividendoli secondo le diverse applicazioni. Quindi, esistono

prodotti MODIS Atmosphere, Land, Ocean. Alcuni di tali prodotti possono essere subito disponibili e scaricabili tramite browser come immagini, altri necessitano di ulteriori elaborazioni da parte dell'utente.

Negli ultimi anni sono state sviluppate diverse tecniche e strumentazioni in grado di ottenere informazioni sulle caratteristiche ottiche e fisiche delle particelle.

Vi sono diversi studi che prendono in considerazione il fenomeno degli aerosol e ne studiano caratteristiche e modelli: una combinazione di rilevamenti simultanei su terra e oceano possono ad esempio determinare il ruolo dell'aerosol in ambiente atmosferico e conseguentemente a ricavare quanto l'attività umana influisce nella formazione dell'aerosol stesso. Mediante i dati del satellite Terra è possibile studiare la variabilità interannuale della qualità dell'aria in ambiente troposferico, grazie a strumenti specifici quali MOPITT e AOD (Aerosol Optical Depth) installati su MODIS. Considerando diversi periodi annuali e diverse regioni è possibile definire differenti emissioni in diverse aree con determinate caratteristiche che danno risposte significative in termini di comparazione: aree urbane o in cui vi è presenza massiccia di insediamenti industriali producono un significativo quantitativo di polveri ed altri inquinanti rilevabili dagli strumenti sul satellite, così come le aree soggette ad incendi boschivi. Dalla raccolta dei dati in tali porzioni di territorio è successivamente possibile creare modelli che permettano di interpretare il dato e produrre risposte significative sul fenomeno.

5.2.1 Architettura del sistema software

PM MAPPER® è un sistema di monitoraggio dell'inquinamento atmosferico da piattaforma satellitare, in grado di generare mappe di concentrazione di particolato fine Pm10 e Pm2.5 (in $\mu\text{g}/\text{m}^3$) e mappe di Qualità dell'Aria basate sulla classificazione di pericolosità dell'Ente di Prevenzione Ambientale americano US EPA 2006. Attualmente il sistema genera mappe con una risoluzione a terra di 3km x 3km, con la possibilità di selezionare il formato di uscita tra un'ampia gamma di formati per garantire la massima compatibilità con i sistemi di visualizzazione e con i sistemi GIS comunemente utilizzati.

Center Wavelength	Ne Δ L	Ne $\Delta\rho$	Maximum Reflectance	SNR	Pixel Size at Nadir (m)
470	0.145	$2.35 \cdot 10^{-4}$	0.96	243	500
550	0.127	$2.11 \cdot 10^{-4}$	0.86	228	500
659	0.169	$3.39 \cdot 10^{-4}$	1.38	128	250
865	0.123	$3.99 \cdot 10^{-4}$	0.92	201	250
1240	0.045	$3.12 \cdot 10^{-4}$	0.47	120	500
1640	0.027	$3.63 \cdot 10^{-4}$	0.94	275	500
2130	0.009	$3.06 \cdot 10^{-4}$	0.75	110	500

Tabella 5.1: Caratteristiche dei canali MODIS usati per il rilevamento degli aerosol sull'oceano; Ne $\Delta\rho$ corrisponde al sole allo zenit ($\vartheta_s = 0^\circ$).

PM MAPPER è un sistema software costituito da due step di elaborazione in grado di gestire l'input di dati multispettrali acquisiti dai sensori MODIS.

In tabella 5.1 sono riportati i valori delle sette bande spettrali che sono potenzialmente utili per il telerilevamento dell'aerosol.

Il primo step è dedicato all'elaborazione del dato satellitare multispettrale per estrarre l'informazione di Aerosol Optical Thickness (AOT), ovvero il contenuto di particolato atmosferico presente nella colonna di atmosfera osservata dal sensore satellitare. Il segnale ricevuto da un sensore satellitare ad una fissata lunghezza d'onda (Figura 5.1) è composto da diversi contributi di radiazione (diretta, diffusa, scatterata) che possono interagire con il mezzo che attraversano (atmosfera, nubi, superficie).

Dalle caratteristiche di interazione della radiazione con l'atmosfera, e separando i contributi di radianza/rifletanza provenienti da superficie e atmosfera, è possibile quantificare il contenuto di aerosol presente nella colonna di atmosfera osservata dal satellite. L'approccio utilizzato per estrarre il contenuto di AOT è di tipo Look-up table, ovvero i valori di radianza/rifletanza osservati dal sensore satellitare vengono confrontati con una serie di valori calcolati in funzione delle diverse condizioni atmosferiche (temperatura, umidità, ecc.), superficiali (suolo, acqua) e contenuto di aerosol.

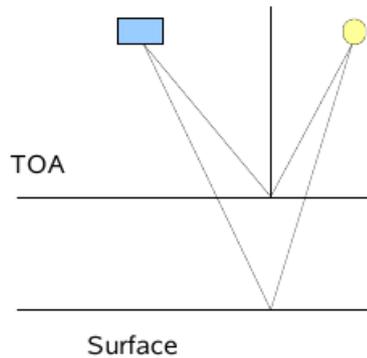


Figura 5.1: Contributi alla Radianza/Riflettanza ricevuta da un sensore satellitare

In Figura 5.2 è rappresentato uno diagramma semplificato del sistema PM MAPPER[®] per il calcolo dell'AOT: inizialmente vengono eseguite una serie di procedure di pre-processamento (identificazione dei pixel di nube, di acqua e di suolo; calcolo degli angoli di osservazione; caricamento di dati ancillari; ecc.) che estraggono tutte le informazioni necessarie e preparano il dato per l'applicazione dell'algoritmo di inversione.

La variabilità delle proprietà della superficie di tipo Suolo rispetto a quella di tipo Acqua determina due algoritmi differenziati per tipologia di suolo osservato (PROCESS_ocean, PROCESS_Land). PM MAPPER[®] lavora su cluster 3km x 3km di pixel, e per consentire la corretta applicazione dell'algoritmo devono essere rispettati i seguenti requisiti: - eliminazione dei pixel di nube; - i valori di radianza/riflettanza osservati devono essere al di sotto del valore di soglia di 0.25. Mentre il rispetto del primo requisito risulta evidente in quanto il sensore non ha la capacità di penetrare attraverso le nubi, il rispetto del secondo è richiesto per minimizzare l'errore nella procedura di calcolo dell'AOT.

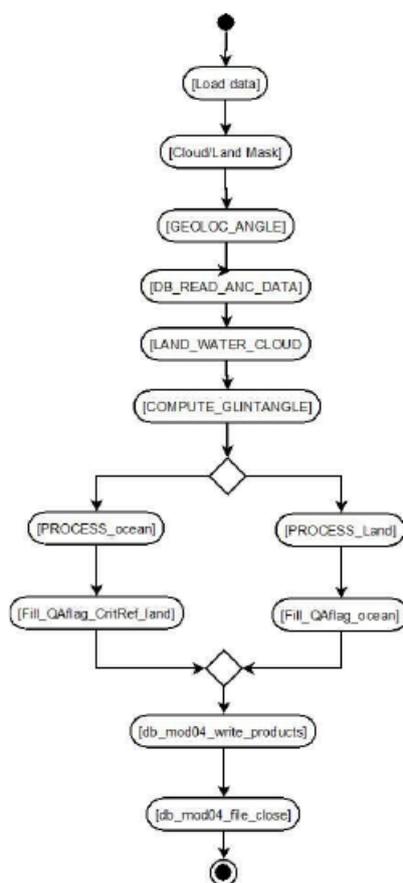


Figura 5.2: Diagramma funzionale del modulo di elaborazione per il calcolo dell'AOT

Il secondo modulo di elaborazione consente di stimare la concentrazione di particolato fine PM_{2.5} e PM₁₀ alla superficie (nei primi 250m dell'atmosfera) da valori di AOT. Inoltre, utilizzando i valori di riferimento per l'appartenenza alle diverse classi di pericolosità, è possibile ottenere come prodotto di secondo livello la mappa di Qualità dell'Aria: PM MAPPER® utilizza gli standard della classificazione US EPA 2006 con la definizione di cinque classi di pericolosità: Buona, Moderata, Non salutare per individui sensibili, Non salutare, Non salutare intensa.

Molti studi sono stati effettuati per valutare la possibilità di utilizzare mappe di PM misurate da sensori satellitari per integrare o sostituire le misure da strumentazione al suolo. In recenti studi è stato dimostrato che esiste una buona correlazione tra valori di PM_{2.5} misurati al suolo e valori

di AOT, con un coefficiente di correlazione pari a 0.7 per misure istantanee, e che migliora fino a valori di 0.91 – 0.95 su medie mensili calcolate su un arco temporale di un anno. Le performance di PM MAPPER sono state valutate su scala regionale in differenti periodi tra Ottobre 2005 e Giugno 2008, mostrando una deviazione dell'ordine del 30% dalle misure di particolato al suolo.

Le recenti attività di validazione hanno messo a confronto PM MAPPER (mappe di AOT con risoluzione 3km x 3km) con il prodotto originale MODIS (mappe di AOT con risoluzione 10 km x 10 km). Su un periodo di 8 mesi (Novembre 2007 – Giugno 2008) sono state raccolti dati MODIS provenienti dai satelliti TERRA ed AQUA. Per garantire la correttezza del calcolo dei coefficienti di correlazione, le mappe di concentrazione di sono state ricampionate alla risoluzione comune di 1 km x 1 km.

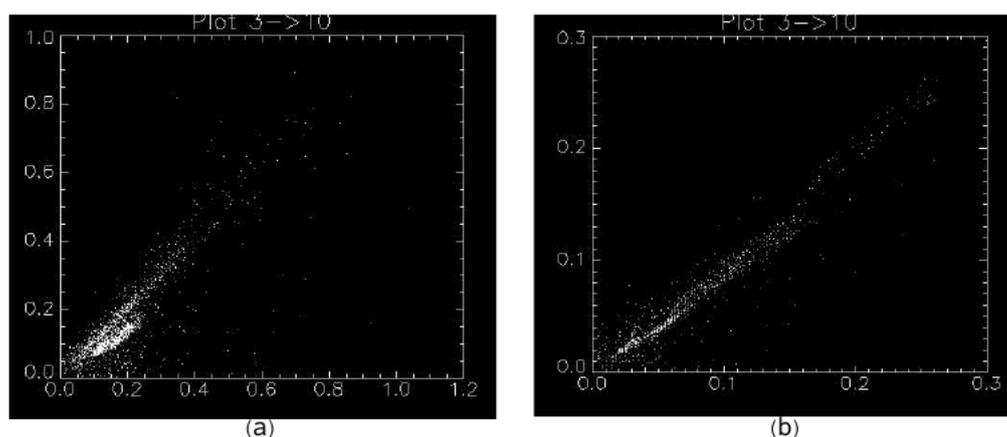


Figura 5.3: Scatterplot dei valori di AOT estratti da PM MAPPER rispetto al prodotto MODIS sul Nord Italia nel periodo Novembre 2007 – Giugno 2008: a sinistra è mostrato il caso peggiore riscontrato (coefficiente di correlazione pari a 0.80), mentre a destra è mostrato il caso migliore (coefficiente di correlazione pari a 0.95).

5.2.2 PM MAPPER : esempio di monitoraggio e uso delle mappe

In Figura 5.4 (b) è mostrata la mappa di concentrazione di PM_{2.5} sull'Italia, relative alla giornata del 14 Febbraio 2008. All'informazione quantitativa

($\mu\text{g}/\text{m}^3$) delle mappe di concentrazione di polveri sottili PM2.5 e PM10 viene associata una scala di colori per consentire un'interpretazione qualitativa della mappa. E' stata utilizzata una scala di colori RAINBOW che rende immediata l'identificazione di zone interessate da condizioni di inquinamento elevato al passaggio del colore dal blu verso il rosso. I sistemi di visualizzazione applicano in modo lineare la scala di colore selezionata ai valori presenti nella mappa: nel caso di scala RAINBOW, ad esempio, al valore minimo presente nella mappa viene associato il colore nero, mentre al valore massimo viene associato il colore rosso. L'applicazione automatica della scala di colore non consente quindi di confrontare direttamente mappe di concentrazione di polveri sottili relative a differenti periodi di acquisizione se i valori minimo e massimo delle diverse mappe sono differenti. Per rendere la visualizzazione delle mappe confrontabile PM MAPPER applica la scala di colore RAINBOW in modo assoluto a valori di concentrazione compresi tra $0 \mu\text{g}/\text{m}^3$ (colore nero) e $\geq 250.4 \mu\text{g}/\text{m}^3$ (colore rosso), per le mappe di PM2.5, e valori di concentrazione compresi tra $0 \mu\text{g}/\text{m}^3$ (colore nero) e $\geq 424 \mu\text{g}/\text{m}^3$ (colore rosso), per le mappe di PM10.

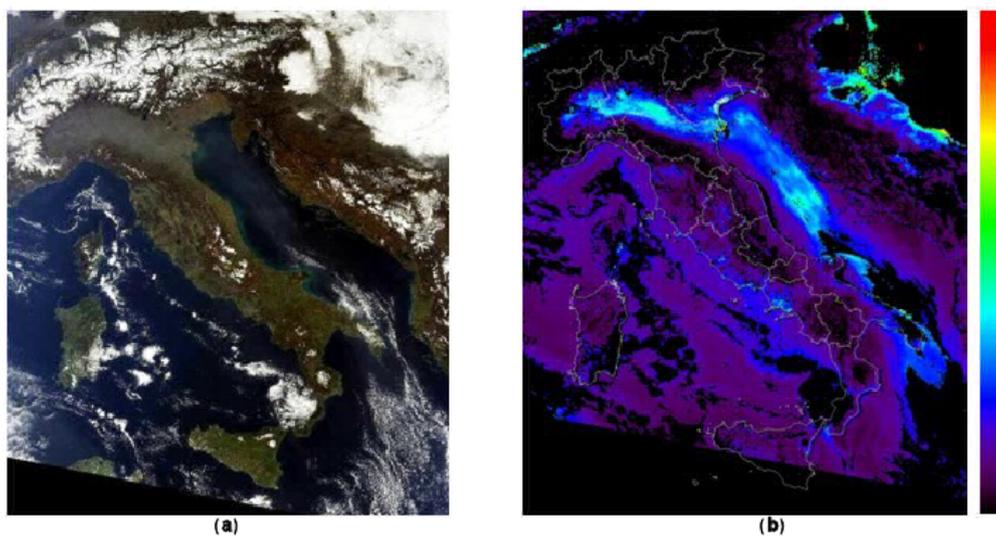


Figura 5.4: Immagine MODIS acquisita sull' Italia il 14/02/2008 (a) , relativa mappa di concentrazione di polveri sottili PM2.5 (b).

In Figura 5.6 è rappresentata l'informazione di Qualità dell'Aria associata

alla mappa di concentrazione di polveri sottili del 14 Febbraio 2008 di Figura 5.4 (b), utilizzando la classificazione US EPA 2006 mostrata in Figura 5.5.

Standard EPA2006 di qualità dell'aria (PM 10 e PM 2.5)

Index	Category	PM _{2.5} (µg/m ³)	PM ₁₀ (µg/m ³)
0-50	Good	0-15.4	0-54
51-100	Moderate	15.5-40.4	55-154
101-150	Unhealthy for sensitive groups	40.5-65.4	155-254
151-200	Unhealthy	65.6-150.4	255-354
201-300	Very unhealthy	150.5-250.4	355-424

Figura 5.5: Classi di pericolosità

Rispetto alle mappe di concentrazione di polveri sottili in cui l'informazione è di tipo continuo e quantitativo, la mappa di Qualità dell'Aria da un'informazione qualitativa che consente immediatamente di identificare le zone soggette a stress di inquinamento.

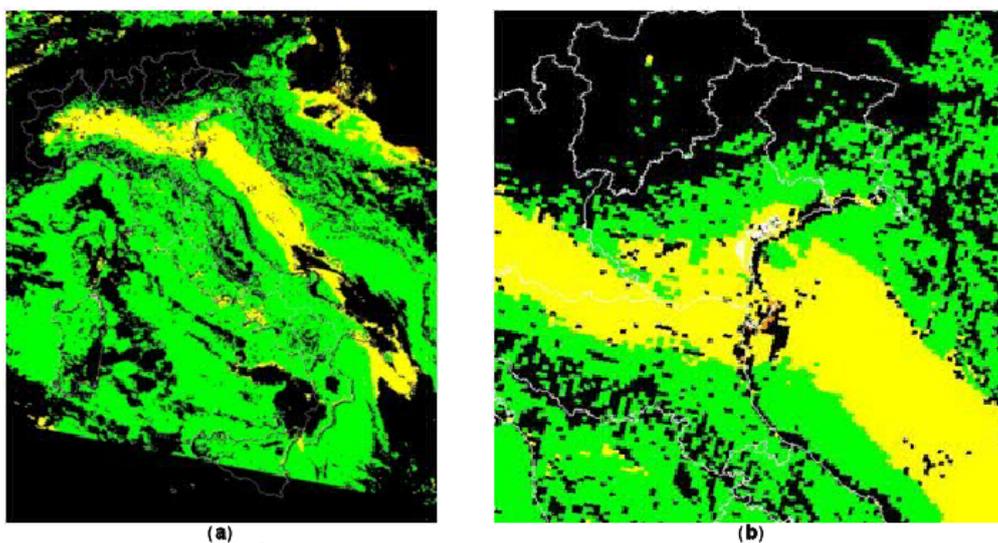


Figura 5.6: Mappa di Qualità dell' Aria sull' Italia relativa all' acquisizione del 14 Febbraio 2008 (a), e particolare sul nord-est (b).

Utilizzando i colori associati alle classi US EPA2006 è semplice riconoscere nell'esempio di Figura 5.6(a) una situazione di Qualità dell'Aria Buona

diffusa su tutta l'Italia, che nella Pianura Padana diventa Moderata. Dallo zoom in Figura 5.6(b) è poi possibile identificare un numero ridotto di pixel in cui la condizione diventa Non salubre per individui sensibili.

5.3 Analisi e Gestione di mappe di Concentrazione di particolato con GRASS

5.3.1 Dati utilizzati

Per l'analisi delle concentrazioni di particolato sulla zona di Alessandria, abbiamo utilizzato una serie di immagini raster in formato Geotiff relative a tutta l'Italia Settentrionale forniteci da MEEO (Meteorological and Environmental Earth Observation MEEO) azienda che si occupa di sviluppare e commercializzare servizi e prodotti nell'ambito dell'Osservazione della Terra.

Diversamente dai dati ottenuti attraverso le centraline a terra, la qualità di quelli satellitari, è strettamente legata alla copertura nuvolosa della zona d'interesse, i valori di concentrazione in corrispondenza delle nubi risulta nullo per evidenti motivi legati all'impossibilità dei sensori di analizzare l'aerosol in quelle zone. Per questo motivo tra le immagini a disposizione, sono state utilizzate quelle con la minore copertura nuvolosa, che a differenza delle altre offrono una visione meno frammentaria dell'andamento delle concentrazioni di particolato. la differenza tra i due tipi di immagine può essere osservata in figura 5.7 e 5.8.

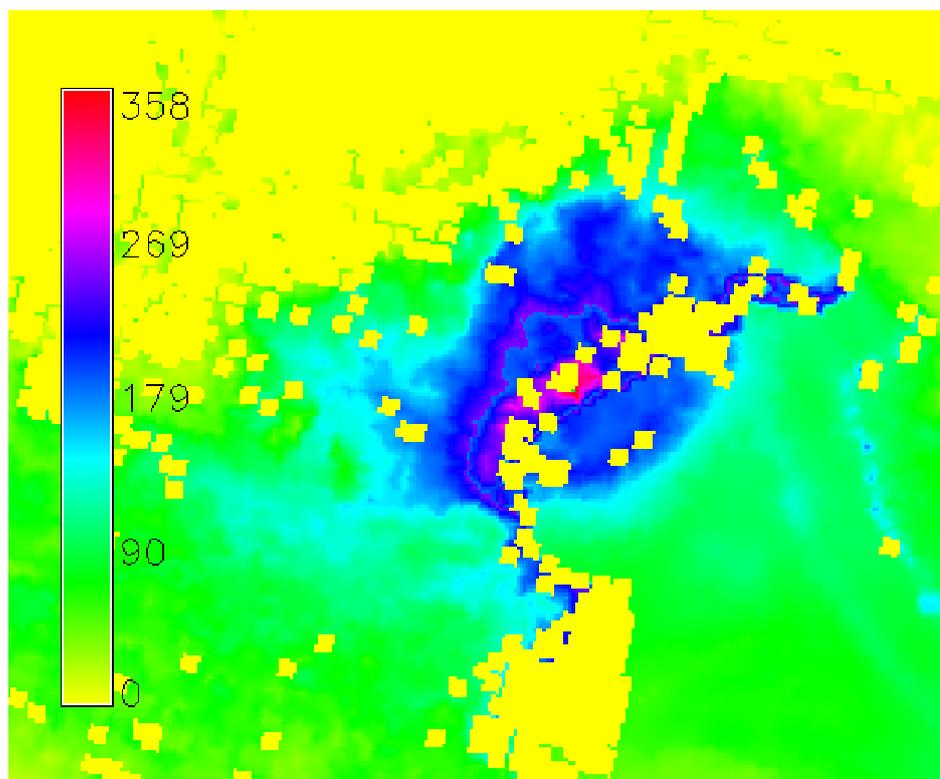


Figura 5.7: Esempio di immagine con scarsa copertura nuvolosa

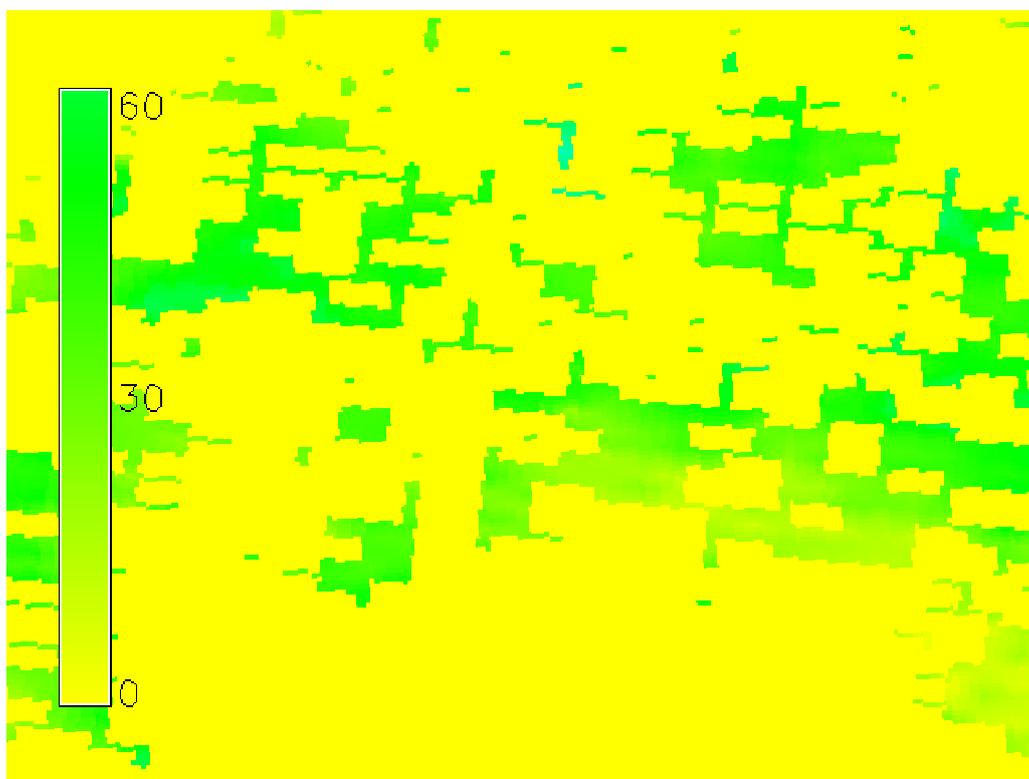


Figura 5.8: Esempio di immagine con eccessiva copertura nuvolosa

Le zone evidenziate in giallo sono quelle con intensa copertura nuvolosa che implicano la mancata rilevazione del dato che quindi viene posto uguale a zero. La presenza di nuvole nelle rilevazioni satellitari è un fattore costante, per questo motivo la maggior parte delle immagini utilizzate presentano zone con valori nulli. Questi 'buchi' potrebbero essere riempiti eseguendo delle interpolazioni sui valori di pm calcolati nelle aree più vicine, in questa sede però, si è preferito non ricorrere a tale soluzione, in modo da utilizzare solo dati effettivamente rilevati.

L'altra serie di immagini utilizzate è quella relativa alla Qualità dell'Aria (QA), ad ogni coppia di raster pm10 e pm2.5, di un dato giorno, corrisponde un raster di QA. A differenza degli altri, a questi raster non sono associati va-

lori puntuali corrispondenti a quantità definite, ma ad ogni pixel corrisponde il valore della categoria del colore rappresentato. Infatti la loro utilità sta nel rappresentare la QA in maniera strettamente visiva, seguendo la scala di colori delle classi di pericolosità definite dall'EPA, come mostrato in figura 5.6.

Come si può notare, le immagini così come ci sono pervenute, non sono di nessuna utilità se non vengono riferite ad una zona ben precisa, pertanto il primo passo è stato quello di contestualizzare e riproiettare i dati raster.

5.3.2 Contestualizzazione dei dati

L'utilità delle immagini raster relative alle concentrazioni di PM10 e PM2.5 è strettamente legata ad una loro specifica contestualizzazione, quindi un inquadramento dell'area d'interesse nell'ambito di aree comunali o regionali, urbane o verdi, abitative o industriali, tutti elementi indispensabili per eventuali analisi.

Il sistema di riferimento geografico utilizzato in questa sede è il sistema di coordinate UTM (Universale Trasverso di Mercatore) che si basa sull'ellissoide di riferimento (datum) europeo del 1950 e a cui fa riferimento la cartografia ufficiale dell'Istituto Geografico Militare (IGM). Tale sistema consiste in una proiezione su un piano della superficie terrestre (cioè della superficie esterna dell'ellissoide di riferimento) e nella successiva divisione in "zone". Ogni zona è delimitata da un meridiano occidentale e da un parallelo meridionale e, a partire da questi, si estende per un certo numero di gradi. L'Italia è individuata dalle zone 32 e 33 e il Piemonte è situato tutto nella zona 32.

Dopo l'importazione delle immagini, molte risultavano 'spostate' rispetto alla posizione reale, poichè georeferenziate rispetto alle zone 31 e 33 del sistema UTM. La riproiezione è stata fatta con GRASS, sono state create due nuove location rispettivamente nelle zone 31 e 33 del sistema di coordinate UTM, importate in ognuno le immagini riferite alle rispettive zone e successivamente, attraverso il modulo `r.proj` ogni immagine è stata riproiettata nella zona 32.

In figura 5.9 è riportata una delle immagini di cui sopra, alla quale per dare un'idea della collocazione spaziale, abbiamo sovrapposto i layer dei limiti nazionali che da solo conferisce subito all'immagine un valore significativo, trasformandola da semplice tavola di colori ad un mezzo di analisi scientifica attraverso cui è possibile fare una prima, anche se grossolana, valutazione della qualità dell'aria in una qualsivoglia zona rappresentata.

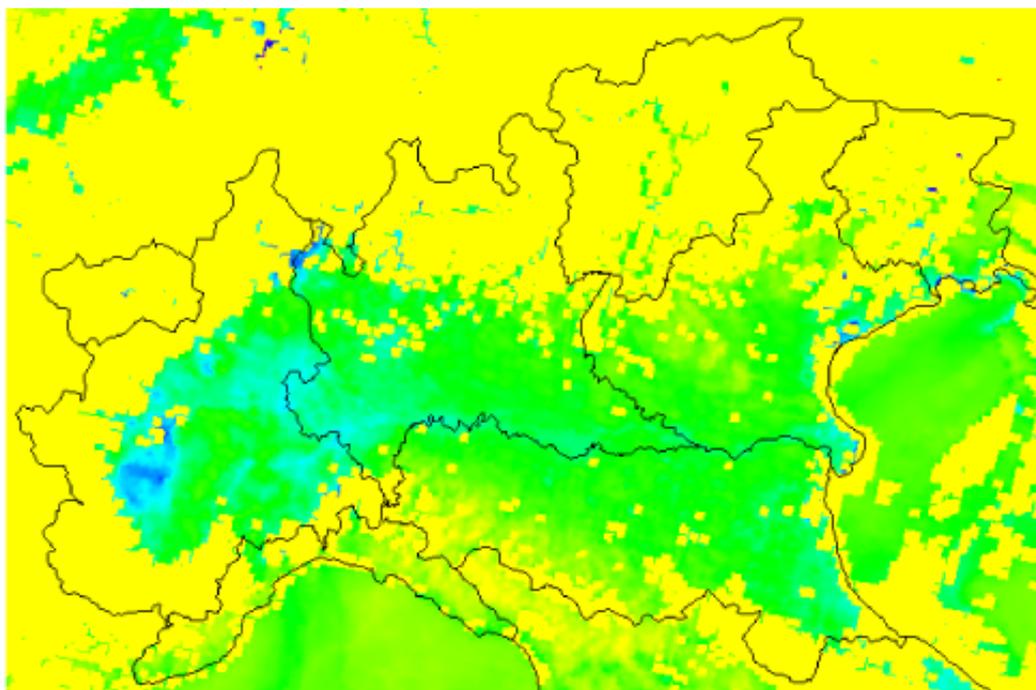


Figura 5.9: Overlay dei confini nazionali

In primo luogo quindi, abbiamo individuato la zona d'interesse, che corrisponde ad un'area delimitata della provincia di Alessandria, evidenziata in rosso (Fig. 5.10) e che include diverse zone appartenenti a comuni limitrofi.

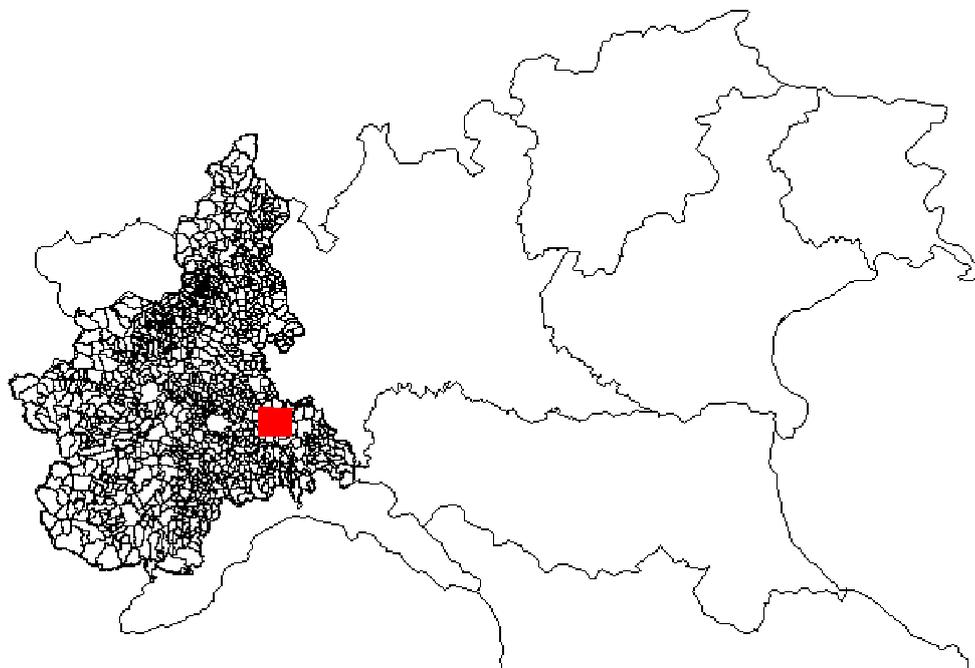


Figura 5.10: Delimitazione dell'area di interesse

Avendo a disposizione i dati vettoriali CSI della Regione Piemonte, sono stati effettuati una serie di overlay successivi attraverso i quali, si è ottenuta una caratterizzazione particolareggiata dell'area oggetto di studio, che permette di effettuare analisi prendendo in considerazione le caratteristiche della zona. In particolare sono stati utilizzati i layer vettoriali di strade e autostrade, ferrovie, fiumi principali, zone abitate, limiti comunali e carta tecnica regionale di Alessandria. In figura 5.11 è mostrata la carta risultante.

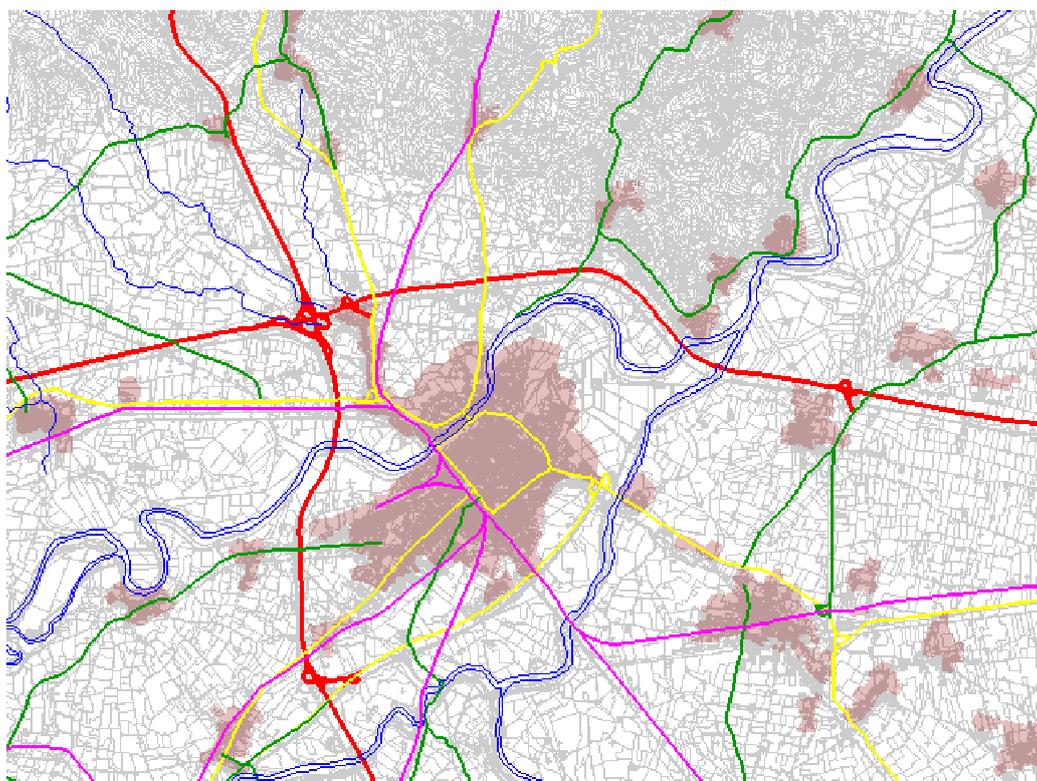


Figura 5.11: Overlay vettoriale

Alla carta in figura 5.11 è possibile sovrapporre in semitrasparenza i raster delle concentrazioni di pm, questa operazione può essere fatta in modo da osservare in successione e alla velocità desiderata tutte le immagini a disposizione, in tal modo è possibile fare delle valutazioni immediate, utili per un'approccio iniziale al problema, per esempio si può determinare:

- se le concentrazioni di pm siano più rilevanti in corrispondenza delle vie di comunicazione e dei centri abitati piuttosto che in aree più isolate.
- se esistono delle aree critiche in corrispondenza delle quali le concentrazioni di pm raggiungono valori elevati nella maggior parte dei giorni considerati.

Pertanto è stato creato un database di mappe di concentrazione di pm10 e pm2.5 e qualità dell'aria, nel quale ogni mappa è stata arricchita con tutti i layers di interesse, come mostato in Figura 5.12.

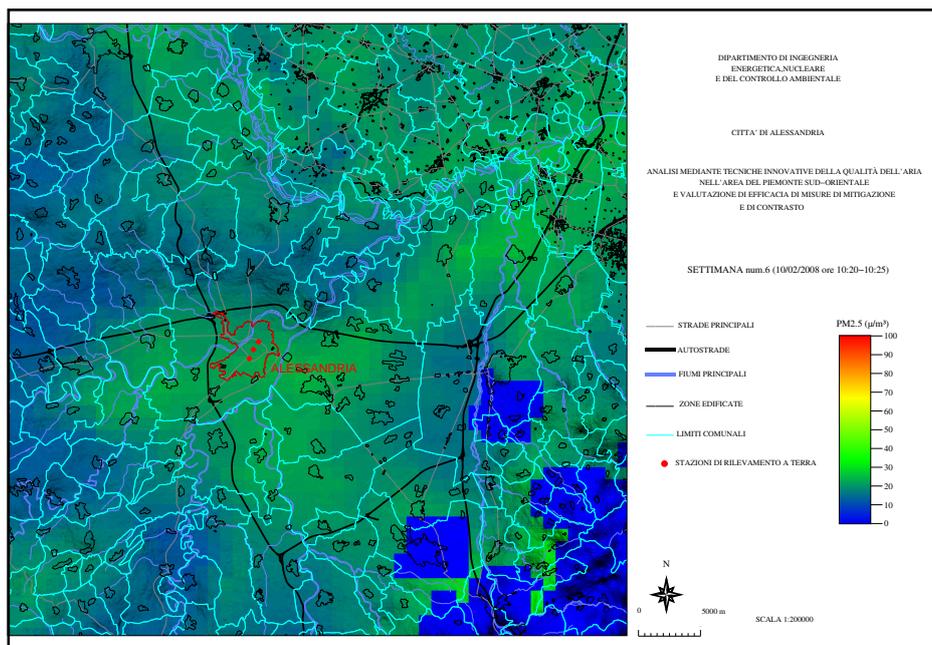


Figura 5.12: Contestualizzazione delle Mappe.

Un simile approccio consente di avere subito una visione generale e di individuare la metodologia più appropriata per un'analisi accurata, infatti se risultasse che durante tutto il periodo esaminato le zone più critiche sono esattamente quelle in corrispondenza delle autostrade, si potrebbe subito pensare che queste rappresentino la fonte principale di pm, quindi ogni successivo passo sarebbe fatto partendo da questo dato.

In base ad una valutazione iniziale del tutto generale, basata su informazioni prettamente visive dei dati, è possibile stabilire quali e in che modo archiviare le informazioni numeriche associate alle diverse mappe e soprattutto che tipo di nuovi dati ci servono per poter costruire un database ben strutturato e funzionale, che metta a disposizione elementi utili per diversi tipi di analisi.

5.3.3 Creazione del Database

La creazione del database consiste nell'associare ad ogni layer una tabella che riassume nel modo più dettagliato possibile le opportune caratteristiche in termini alfanumerici. Innanzi tutto è stato necessario estrapolare dalle immagini raster GeoTiff i valori numerici delle concentrazioni di pm10 e pm2.5. Con GRASS l'interrogazione di una mappa sia raster sia vettoriale può essere fatta interattivamente cliccando sul punto d'interesse, ma a differenza del formato raster, ogni file vettoriale presenta una tabella degli attributi alla quale si può accedere da PostgreSQL, che ci consente di copiarla, modificarla, esportarla ecc. Per le immagini raster la procedura di estrapolazione dei dati è diversa, per ogni raster è stata creata una carta vettoriale che riveste esattamente l'area d'interesse e la suddivide in celle aventi ognuna un'area di un chilometro quadrato. Attraverso il modulo `v.rast.stats`, ad ogni tabella degli attributi dei file vettoriali creati, è stata aggiunta una nuova colonna contenente i valori delle concentrazioni di pm. In tal modo, per ogni mappa, otteniamo una matrice 50x50 di valori nella quale ad ogni cella, relativa ad un'area ben precisa sulla carta raster, corrisponde l'esatto valore di concentrazione di pm. Come mostrato in figura 5.13 interrogando la griglia vettoriale appaiono in una finestra gli attributi dell'area interessata, tra cui la concentrazione del pm e le coordinate x,y del punto nel sistema di riferimento utilizzato.

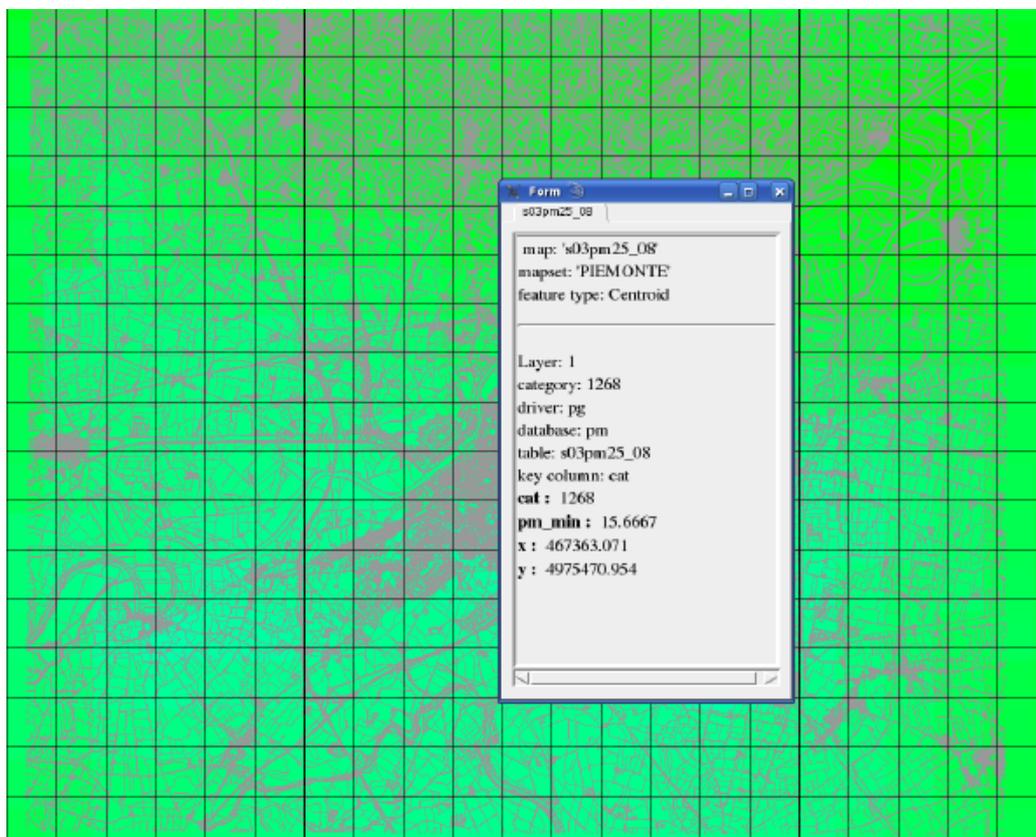


Figura 5.13: Matrice vettoriale

Successivamente, attraverso PostgreSQL tutte le matrici sono state esportate e raggruppate in ordine temporale in un file di calcolo.

Per un'analisi più specifica dei valori di particolato, è stata individuata all'interno della matrice iniziale, una zona più piccola delimitata da una matrice 8x7, mostrata in figura 5.14, che ricopre l'area urbana di Alessandria.

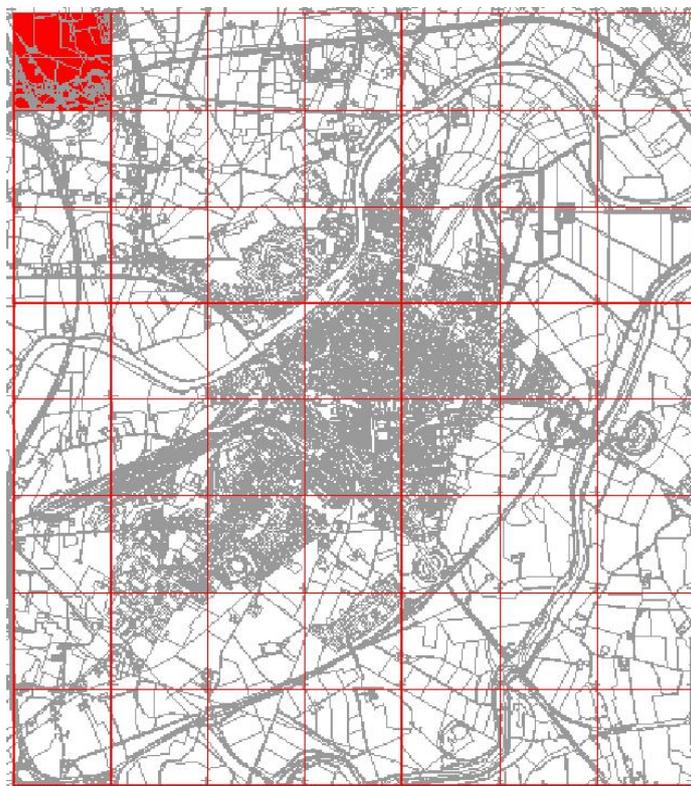


Figura 5.14: Matrice di riferimento per la creazione del database

Ad ogni chilometro quadrato individuato da quest'ultima griglia, è stato associato un diagramma che descrive l'andamento nel tempo delle concentrazioni di particolato. Quindi per esempio per la cella evidenziata in rosso in figura 5.14, per i due tipi di inquinanti, abbiamo un diagramma che mostra l'andamento delle concentrazioni negli anni 2006, 2007, 2008, in figura 5.15 è mostrato un esempio relativo al 2007. La stessa cosa è stata ottenuta per tutte le celle della griglia.

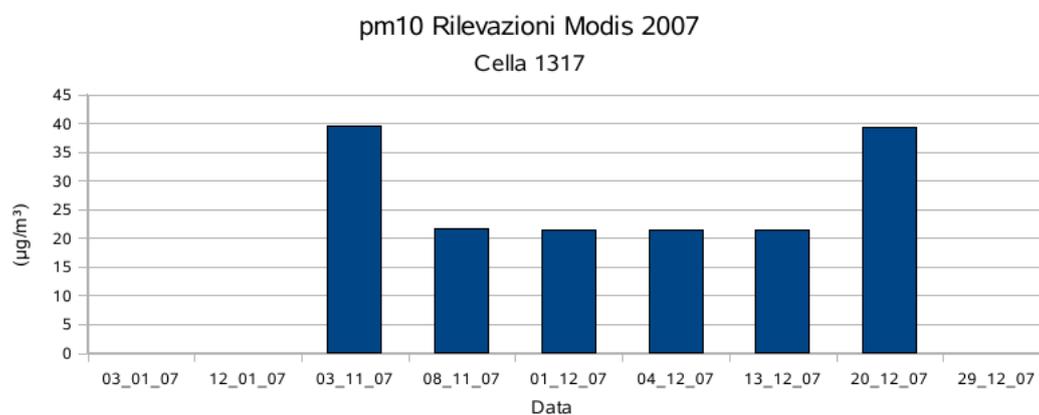


Figura 5.15: Esempio di diagramma relativo ad una cella della griglia di riferimento

5.3.4 Serie temporali

GRASS offre un supporto per elaborazioni di serie temporali (*r.series*), che permette di ottenere statistiche a partire da un set di mappe di input co-registrate, come le immagini satellitari multi temporali. Il modulo permette di calcolare la statistica univariata e la regressione lineare, inoltre possono essere automatizzate analisi comprese in script di testo in modo da generare indicatori climatici.

Avendo a disposizione un numero apprezzabile di immagini dell'Italia settentrionale relative agli anni 2006 e 2008, è possibile fare un'analisi temporale dell'andamento medio delle concentrazioni di particolato. Prendendo ogni anno singolarmente abbiamo creato:

- Carta dei valori medi relativi a tutto l'anno .
- Carta dei valori medi estivi.
- Carta dei valori medi invernali.

Per quanto riguarda i mesi centrali de 2007 non abbiamo a disposizione nessuna informazione.

5.3.4.1 Valori medi Pm10 e Pm2.5 2006-2007

Come spiegato nel capitolo precedente le mappe di concentrazione ottenute da telerilevamento presentano una qualità che risulta strettamente legata alle condizioni di copertura nuvolosa e illuminazione della superficie terrestre. Infatti nei casi in cui la copertura nuvolosa risulti eccessiva o l'irraggiamento solare sia tale da creare una condizione di luminosità sfavorevole, i sensori non sono in grado di produrre dati utili e la mappa risulta inutilizzabile, pertanto per gli anni presi come riferimento, non abbiamo a disposizione tutti i 365 giorni.

Di seguito nelle figure 5.16, 5.17, 5.18, 5.19, 5.20, 5.21, sono mostrate le mappe di concentrazione media di pm10 e pm2.5 relative a tutto l'anno, alla stagione estiva e a quella invernale del 2006:

- per la media dei valori annuali sono stati presi in considerazione il 10,14,24,22 Gennaio, 06,15 Marzo, 03,19,22,23 Aprile, 11,14 Maggio, 12,14,15,19,24,26 Giugno, 22 Agosto, 10 Novembre, 25 Dicembre.
- per la media della stagione estiva sono stati considerati i giorni compresi tra Giugno e Agosto 2007
- per la media della stagione invernale sono stati considerati i giorni compresi tra Dicembre 2007 a Marzo 2008.

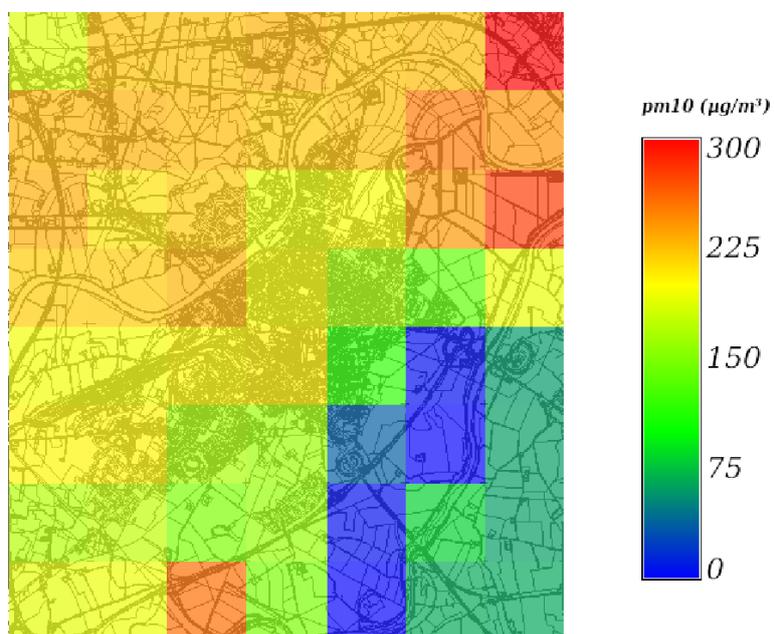


Figura 5.16: Matrice dei valori medi di pm_{10} anno 2006

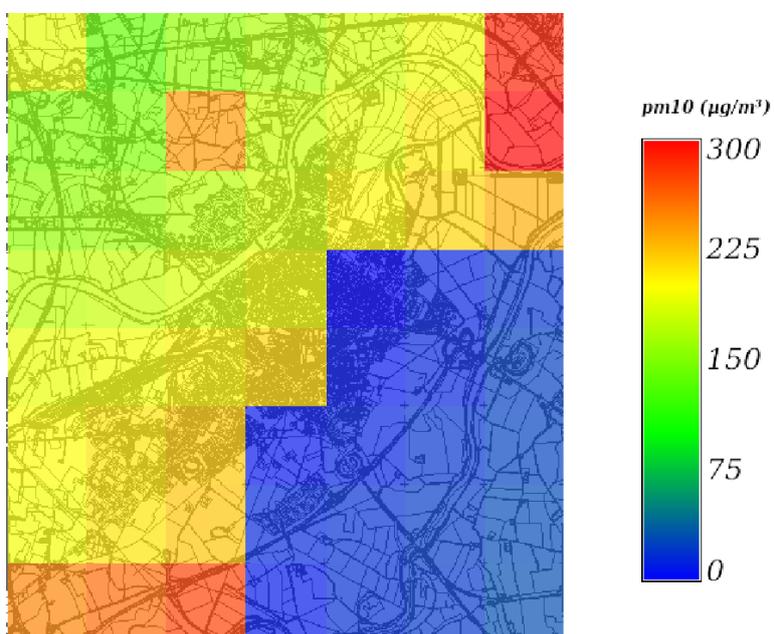


Figura 5.17: Valori medi estivi di pm_{10} anno 2006

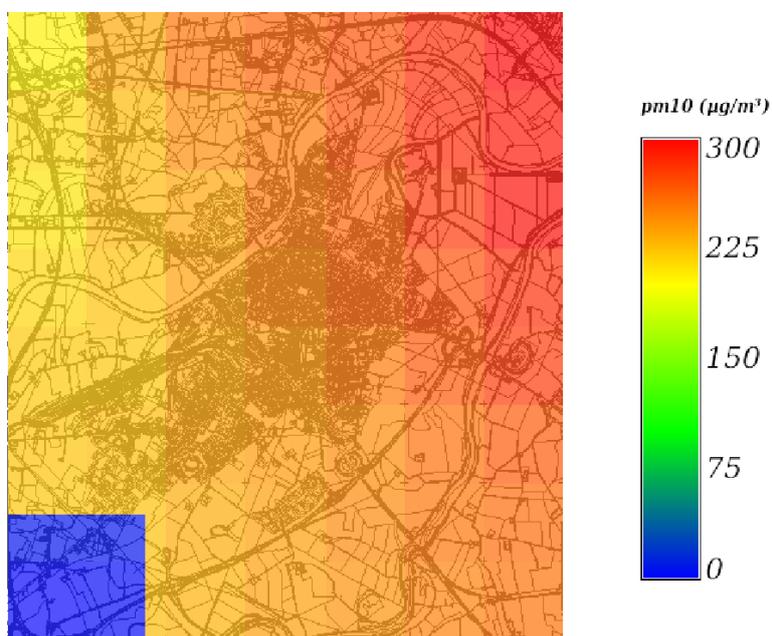


Figura 5.18: Valori medi pm_{10} inverno 2006/2007

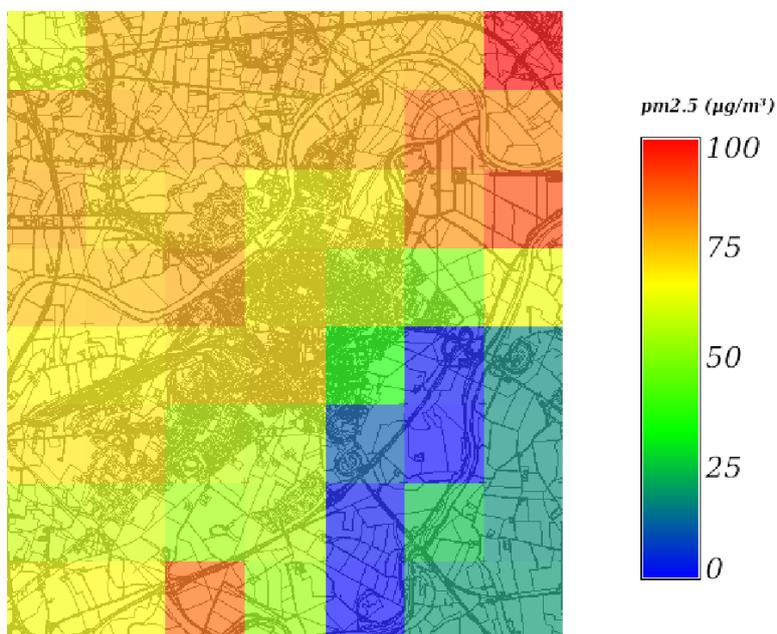


Figura 5.19: Matrice dei valori medi di $pm_{2.5}$ anno 2006

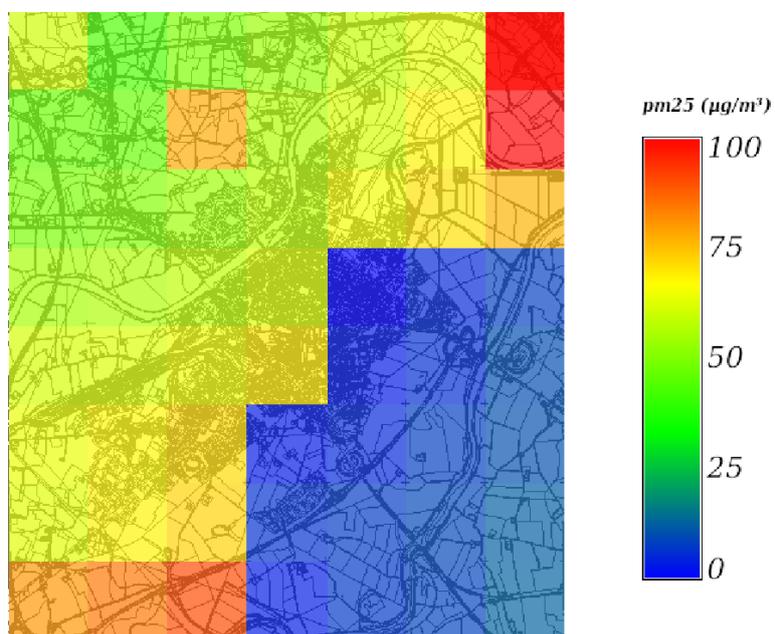


Figura 5.20: Matrice dei valori medi estivi di pm2.5 anno 2006

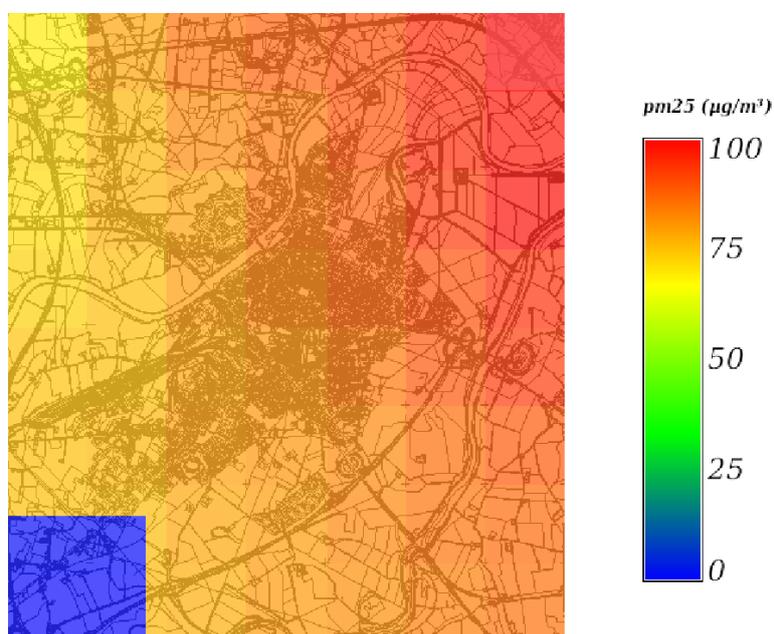


Figura 5.21: Matrice dei valori medi di pm2.5 inverno 2006/2007

5.3.4.2 Valori medi Pm10 e Pm2.5 anno 2007

Per il calcolo dei valori medi dell'anno 2007 (figure 5.22, 5.23) sono stati considerati i giorni: 3, 12 Gennaio, 3, 8 Novembre e 1, 4, 13, 20, 29 Dicembre.

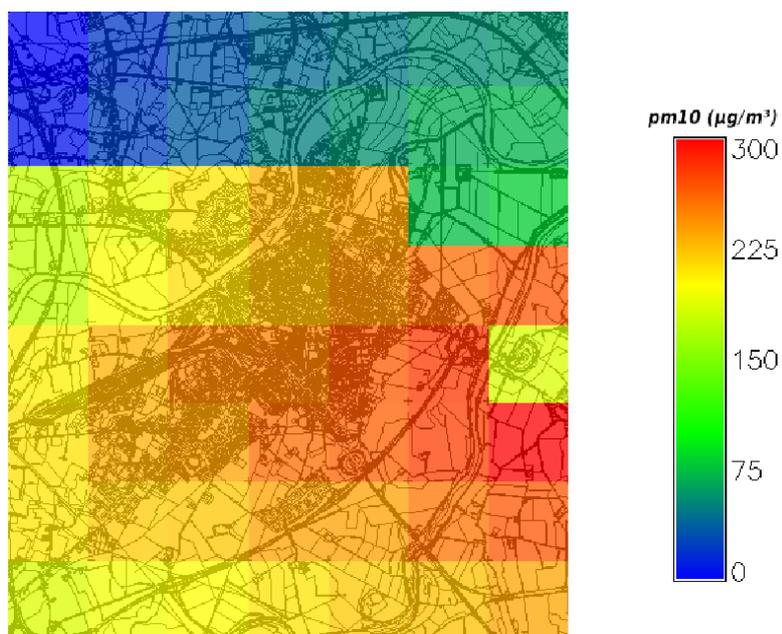


Figura 5.22: Matrice dei valori medi di pm10 anno 2007

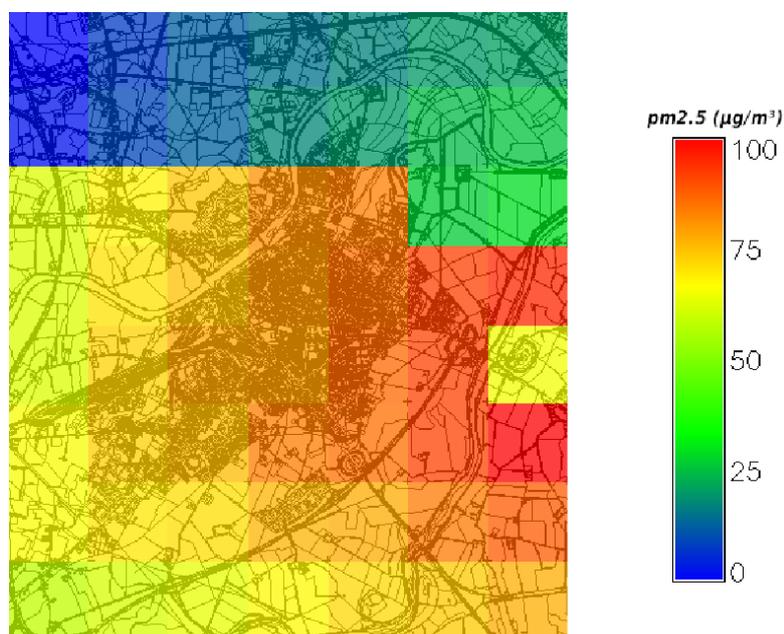


Figura 5.23: Matrice dei valori medi di pm2.5 anno 2007

5.3.4.3 Valori medi Pm10 e Pm2.5 anno 2008

Per quanto riguarda il 2008 sono state utilizzate.

- per i valori medi annuali (figure 5.24,5.27) sono stati considerati i giorni 18, 23 Gennaio, 10, 14, 19 Febbraio, 02, 03, 14, 19, 29 Marzo, 01, 19, 26 Aprile, 03, 07, 12 Maggio, 22, 29 Giugno, 10, 24 Agosto.
- per la stagione estiva (figure 5.25, 5.28) sono stati considerati i giorni compresi tra Giugno e Agosto .
- per la stagione invernale (figure 5.26, 5.29) sono stati considerati i giorni compresi tra Dicembre 2007 e Marzo 2008

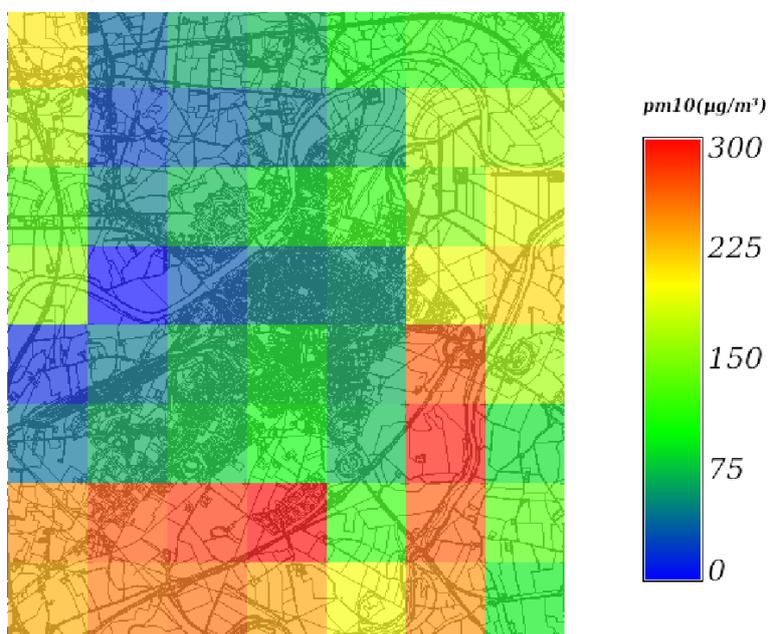


Figura 5.24: Matrice dei valori medi di pm10 anno 2008

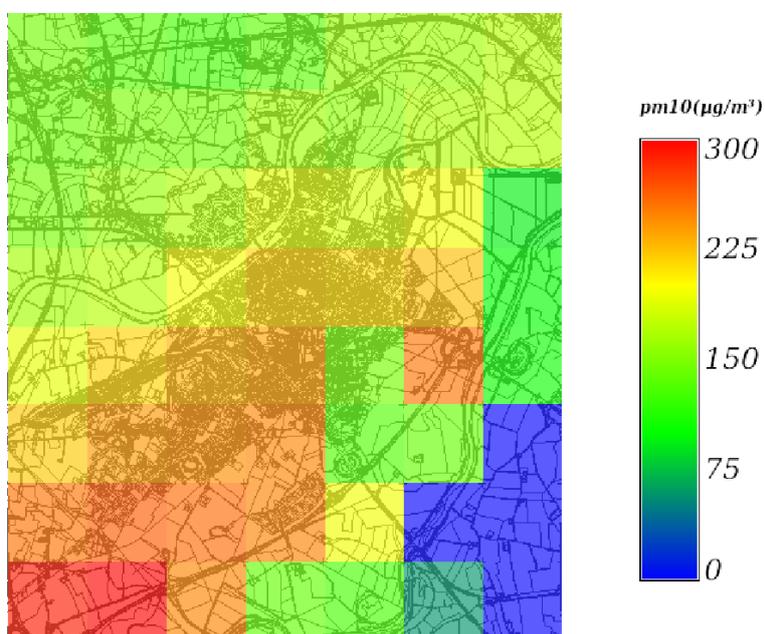
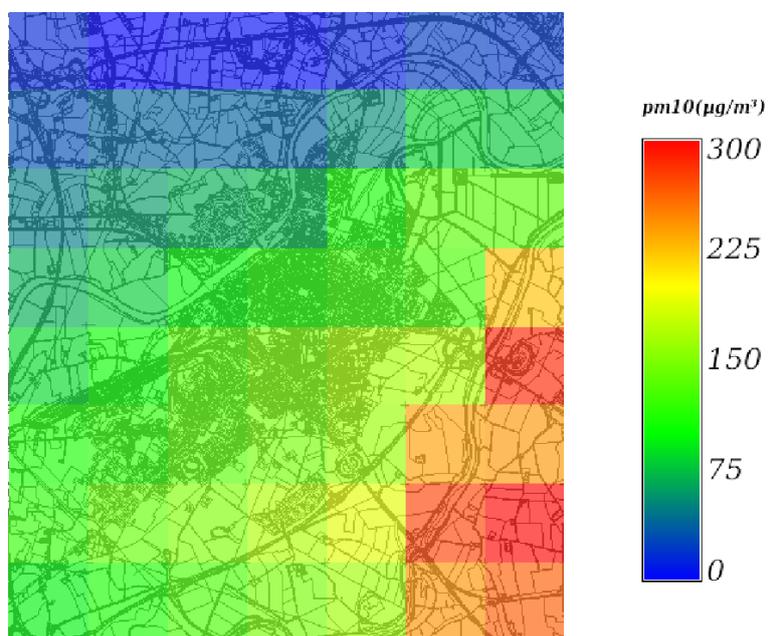
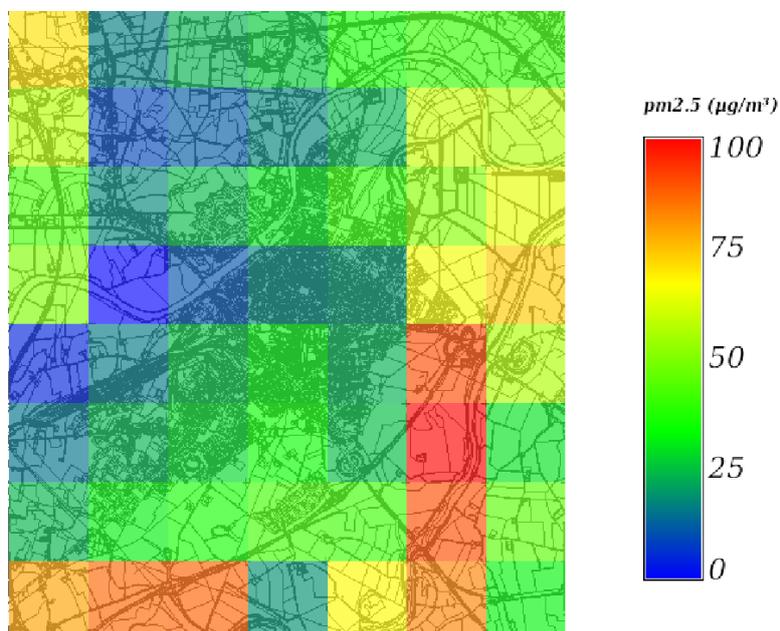


Figura 5.25: Matrice dei valori medi estivi di pm10 anno 2008

Figura 5.26: Matrice dei valori medi di pm_{10} inverno 2007/2008Figura 5.27: Matrice dei valori medi di $pm_{2.5}$ anno 2008

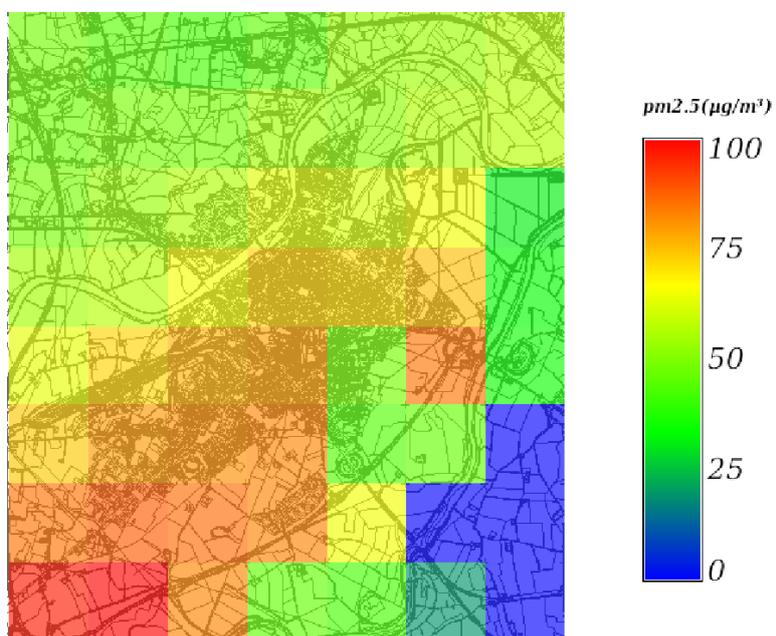


Figura 5.28: Matrice dei valori medi estivi di pm2.5 anno 2008

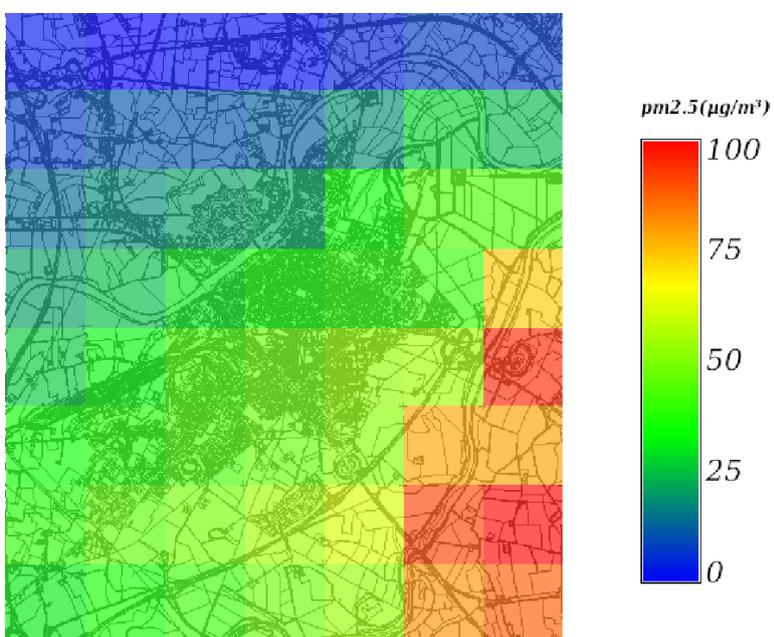


Figura 5.29: Matrice dei valori medi di pm2.5 nell'inverno 2007/2008

Naturalmente la visualizzazione dei valori di concentrazione in GRASS è interattiva e consente di fare dei confronti e delle valutazioni abbastanza rapidamente, in questa sede tali valori saranno mostrati attraverso delle tabelle opportunamente realizzate. I valori medi annuali e stagionali sono mostrati in tabella 5.2, 5.3:

Periodo	Valore Max pm10($\mu\text{g}/\text{m}^3$)	Valore Min pm10($\mu\text{g}/\text{m}^3$)
Anno 2006	85.0	44.0
Estate 2006	123.0	68.0
Inverno 2006/2007	46.0	16.0
Anno 2007	33.2	15.1
Estate 2007	-	-
Inverno 2007/2008	54.8	36.4
Anno 2008	36.1	24.9
Estate 2008	22.2	0.0

Tabella 5.2: Sintesi dei valori medi di pm10 del periodo 2006-2008

Periodo	Valore Max pm2.5($\mu\text{g}/\text{m}^3$)	Valore Min pm2.5($\mu\text{g}/\text{m}^3$)
Anno 2006	24.8	12.8
Estate 2006	39.4	24.0
Inverno 2006/2007	13.2	4.4
Anno 2007	9.2	4.6
Estate 2007	-	-
Inverno 2007/2008	14.9	9.9
Anno 2008	9.7	6.7
Estate 2008	6.3	0.0

Tabella 5.3: Sintesi dei valori medi di pm2.5 del periodo 2006-2008

Per gli stessi periodi abbiamo a disposizione i valori di pm10 misurati a terra dalle centraline (ARPA di Alessandria) relativi alle rilevazioni effettuate da tre stazioni di monitoraggio presenti nella zona urbana di Alessandria, posizionate nella zona centrale della matrice presa come riferimento ed evidenziate in verde in figura 5.30 e denominate rispettivamente 1, 2, 3.

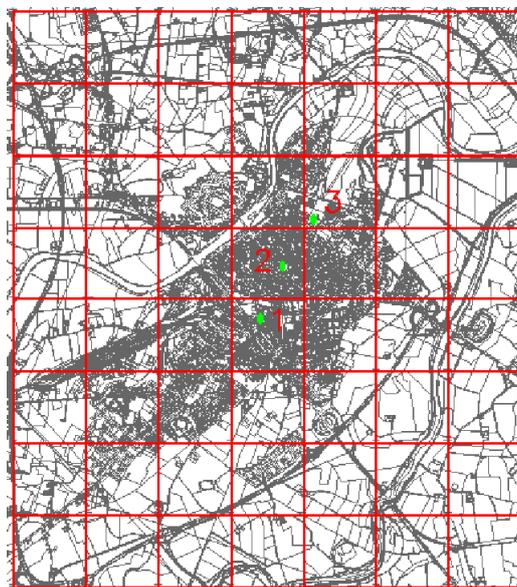


Figura 5.30: Posizione delle Centraline di monitoraggio

In tabella 5.4 sono riportati i valori medi del periodo 2006-2008 rilevati a terra, come si può notare i valori relativi all'estate del 2008 non sono disponibili. In tabella 5.5 gli stessi dati rilevati da satellite.

Periodo	centr.1 pm10 ($\mu\text{g}/\text{m}^3$)	centr.2 pm10 ($\mu\text{g}/\text{m}^3$)	centr.3 pm10($\mu\text{g}/\text{m}^3$)
Anno 2006	54.6	57.7	50
Estate 2006	31.47	39.69	23.94
Inverno 2006/2007	76.90	52	66.04
Anno 2007	60.9	37.2	43
Estate 2007	42.31	22.45	18.7
Inverno 2007/2008	74.92	57.30	61.64
Anno 2008	64.02	49.05	40
Estate 2008	-	-	-

Tabella 5.4: Valori medi di pm10 rilevati dalle centraline : 2006-2008

Periodo	Posiz.1 pm10 ($\mu\text{g}/\text{m}^3$)	Posiz.2 pm10 ($\mu\text{g}/\text{m}^3$)	Posiz.3 pm10 ($\mu\text{g}/\text{m}^3$)
Anno 2006	72.35	71.25	71.26
Estate 2006	108	104	105
Inverno 2006/2007	40.54	41.25	42.53
Anno 2007	30.16	28.78	29.33
Estate 2007	-	-	-
Inverno 2007/2008	45.38	43.14	42.94
Anno 2008	28.32	26.79	29.21
Estate 2008	17.49	16	13.66

Tabella 5.5: Valori medi di pm10 2006-2008 rilevati da satellite nei punti in cui sono posizionate le centraline

In figura 5.31 è riportato un grafico che rappresenta gli andamenti relativi alle rilevazioni a terra e satellitari, in questo caso sono stati messi a confronto i valori medi massimi delle rilevazioni satellitari, con quelli massimi registrati dalle centraline a terra.

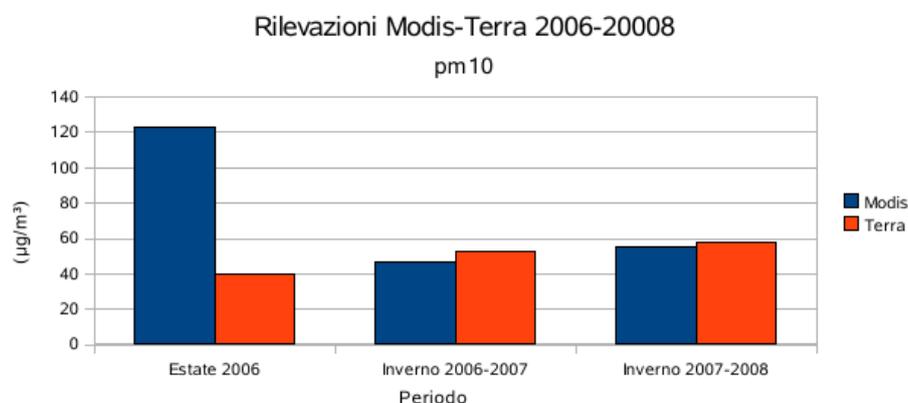


Figura 5.31: Andamenti pm10 terra-modis

Come si può notare i valori rilevati da satellite relativi all'estate del 2006 presentano un picco considerevole che si discosta marcatamente dalle rilevazioni a terra. Tuttavia i due periodi successivi mostrano andamenti quasi coincidenti.

Analizzando singolarmente le mappe di concentrazione di pm10 dell'estate 2006, si è notato che il picco riscontrato è dovuto ai valori molto elevati

relativi ai giorni al 19, 24 e 26 Giugno (figura 5.32), durante i quali si sono registrati picchi di concentrazione pari a $220 \mu\text{g}/\text{m}^3$.

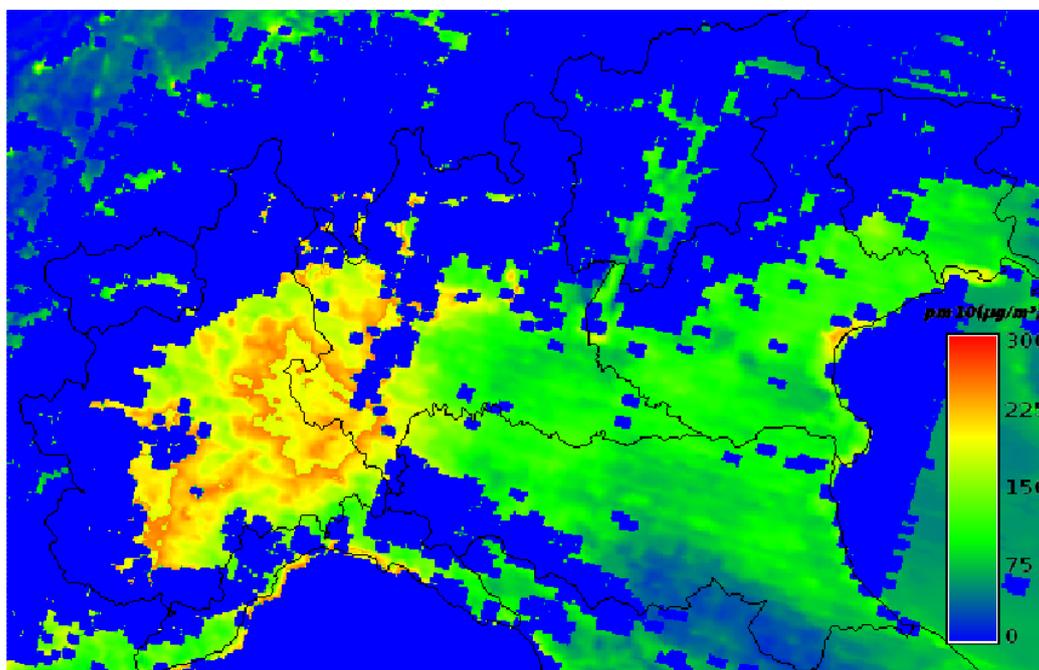


Figura 5.32: Mappa di concentrazione di pm10, Nord-Italia 24/06/06

Il brusco innalzamento di valori di concentrazione di pm10 registrato in questo periodo, come si può notare in figura 5.32 ha coinvolto tutta l'Italia Nord-Occidentale e potrebbe essere attribuito ad una situazione climatica particolare. L'estate 2006 infatti, è stata caratterizzata da una netta discontinuità termica passando dal freddo esordio di Giugno, subito seguito da un lungo periodo caldissimo che ha interessato anche tutto il mese di Luglio e che, nella zona del Piemonte, ha fatto registrare valori sopra la media di $+2^\circ\text{C}$, per poi concludersi bruscamente con una nuova fase fresca e piovosa ad Agosto.

Responsabile di queste anomalie è ritenuto l'Anticiclone Africano che può essere indice di condizioni di tempo stabile, soleggiato e molto caldo con temperature ben oltre le medie del periodo anche di parecchi gradi.

La discrepanza tra i valori misurati a terra e quelli da satellite in questo caso specifico, non è da considerarsi indicativa di una inaffidabilità del

sistema di rilevazione, ma è molto probabilmente legata al fatto che mentre i valori di terra corrispondono ad una media oraria giornaliera delle rilevazioni totali, i dati da satellite vengono prodotti a partire da un'unica osservazione giornaliera effettuata in una determinata ora, che nel caso specifico potrebbe corrispondere ad un momento in cui la stabilità atmosferica e le particolari condizioni di pressione e temperatura abbiano determinato una condizione di criticità .

Se si escludono i picchi in Giugno, il valore medio, relativo all'estate 2006, si aggira attorno a $45 \mu\text{g}/\text{m}^3$, avvicinandosi a quello rilevato a terra, in figura 5.33 è mostrato il diagramma che rappresenta gli andamenti dei valori di pm10 rilevati con le due metodologie, nel quale sono stati omessi i picchi estivi del 2006.

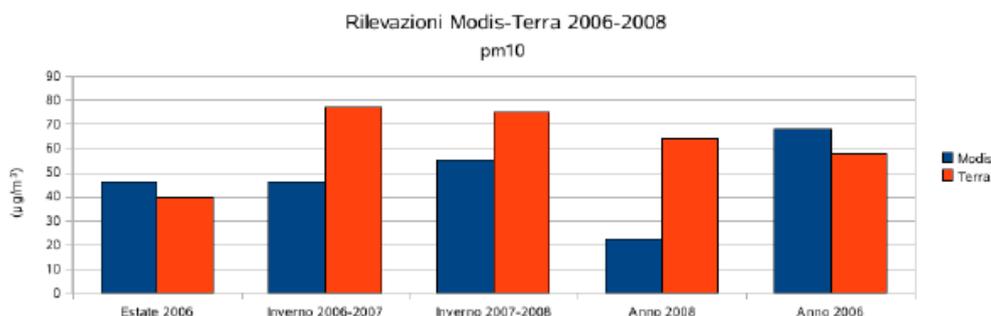


Figura 5.33: Andamenti pm10 Modis-Terra 2006-2008

Pur non essendoci una completa corrispondenza dei valori, si può notare che per i periodi relativi all'Inverno 2006-2007, Inverno 2007-2008 e l'anno 2008, sussiste uno stesso andamento nei due casi. I valori ottenuti da satellite relativi all'anno 2008 risultano inferiori più del 40% rispetto a quelli rilevati a terra, la causa di questa discrepanza potrebbe essere dovuta al fatto che mentre la media dei valori a terra è stata calcolata considerando 90 giorni che vanno da Gennaio a Marzo, mentre nel primo caso i giorni considerati sono stati 20 che vanno da Gennaio ad Agosto.

Un ulteriore confronto può essere fatto considerando i valori di ogni centralina e confrontandoli con i valori registrati dal satellite in quegli stessi punti, nell'arco di ogni anno, i diagrammi sono mostrati in figura 5.34:

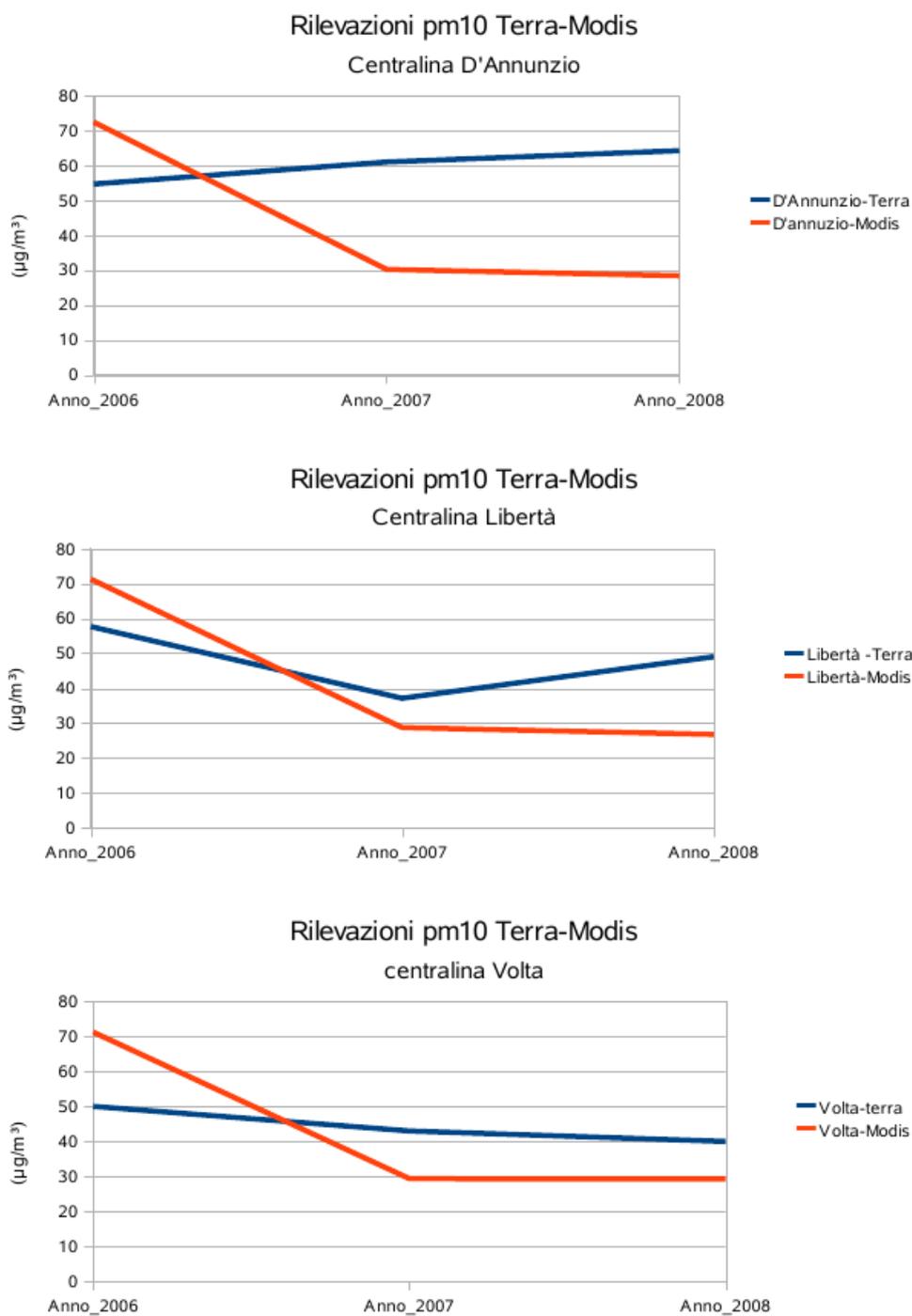


Figura 5.34: Valori Modis-Terra rilevati in corrispondenza delle centraline

Ad eccezione degli andamenti relativi alle rilevazioni in corrispondenza della centralina D'Annunzio, per i quali si nota una discrepanza totale tra le due misure, gli altri due casi sembrano seguire i medesimi andamenti fino al 2007 per poi presentare una tendenza opposta nel 2008.

Se consideriamo la media dei tre valori relativi all'ubicazione delle centraline, derivanti dai due tipi di rilevazione, l'andamento è mostrato in figura 5.35:

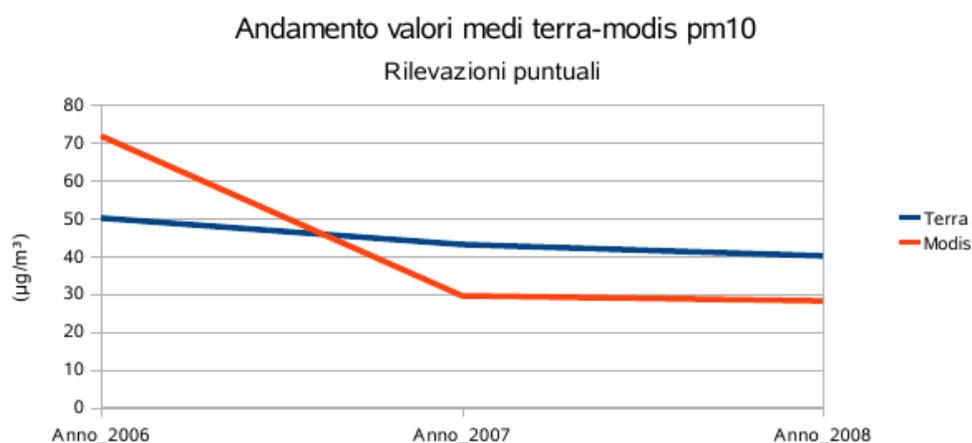


Figura 5.35: Valori Medi Modis-Terra rilevati in corrispondenza delle centraline

Come si può notare i valori medi presentano lo stesso trend. Bisogna tener presente che in questo caso sono stati considerati tre singoli valori derivanti dalla media stagionale dei valori corrispondenti alle rilevazioni fatte nella postazione fissa delle tre centraline. In entrambi i casi si nota un andamento decrescente dei valori dal 2006 al 2008.

L'andamento varia se invece di considerare la media dei tre valori, si tiene conto del valore massimo rilevato sia da terra che da satellite in ogni posizione (fig. 5.36).

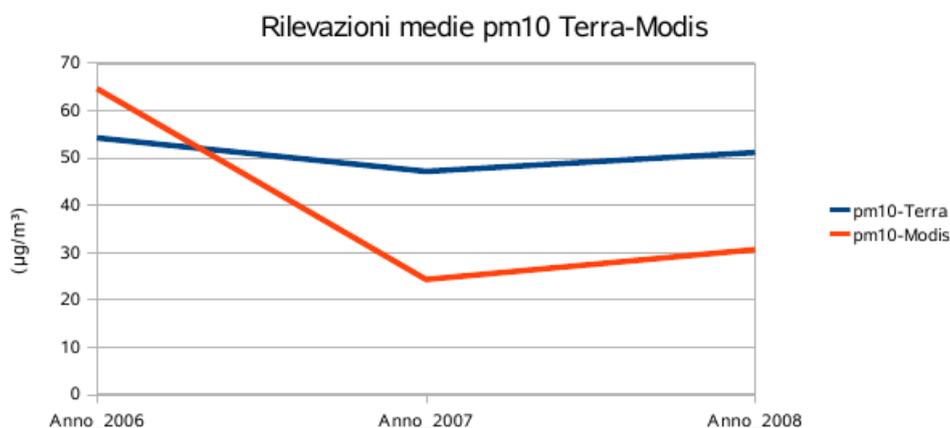


Figura 5.36: Andamenti medi annuali pm10 Terra-Modis

Le rilevazioni modis, come nel caso precedente, presentano dei valori più alti nel 2006, mentre negli anni 2007 e 2008 prevalgono i valori rilevati a terra, tuttavia si può notare che nel secondo caso entrambi gli andamenti seguono una curva discendente che presenta un massimo nel 2006 e un minimo nel 2007. Nel primo caso invece (Fig. 5.35) il minimo si registra nel 2008. Questo risultato mette in evidenza che nell'arco dei tre anni considerati la qualità dell'aria ha comunque subito dei miglioramenti passando dal 2006 al 2008.

Questa prima analisi è stata fatta allo scopo di avere delle indicazioni sulla validità o meno del telerilevamento nell'ambito delle valutazioni della qualità dell'aria, in prima approssimazione, quindi, visti i risultati, possiamo considerare il telerilevamento come un mezzo efficace per questo tipo di valutazioni.

5.3.5 Analisi delle differenze di concentrazione giornaliere di pm10: Terra-Modis

Nel paragrafo precedente abbiamo effettuato dei confronti, tra i dati da telerilevamento e i dati a terra, prendendo in considerazione le medie dei valori calcolate su lunghi periodi. In questa fase invece, la stessa analisi verrà fatta giorno per giorno, considerando solo i valori rilevati a terra dei giorni per i

quali disponiamo dei dati da satellite. In figura 5.30 è mostrata la posizione delle centraline rispetto alla zona in esame. Come valori Modis di riferimento considereremo quelli relativi alle tre aree, ciascuna di 1 km², nelle quali sono ubicate le centraline. Quindi non considereremo più valori medi ma valori puntuali. Le centraline che precedentemente abbiamo indicato con i numeri 1, 2, 3 sono denominate rispettivamente D'Annunzio, Libertà e Volta, nomi che compariranno nei diagrammi 5.37, 5.38, 5.39.

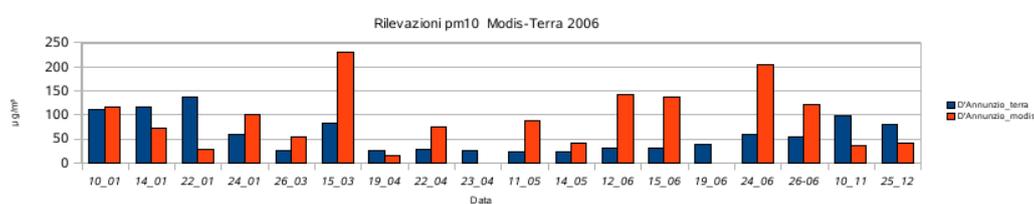


Figura 5.37: Rilevazioni giornaliere D'Annunzio-Modis-Terra 2006

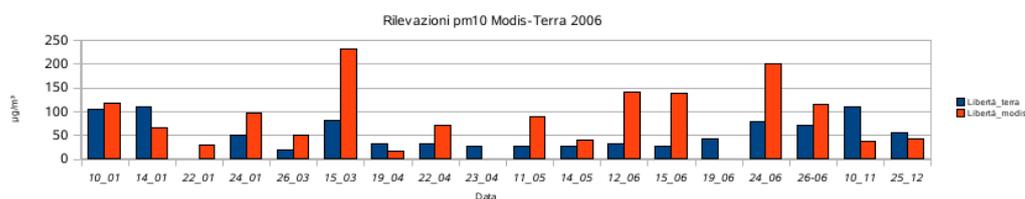


Figura 5.38: Rilevazioni giornaliere Libertà-Modis-Terra 2006

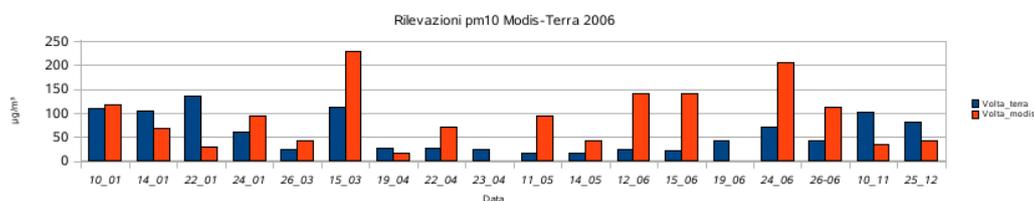


Figura 5.39: Rilevazioni giornaliere Volta-Modis-Terra 2006

La normativa vigente in materia di qualità dell'aria, costituita dal DM 60/02, riguardo all'inquinante pm10 stabilisce valori standard di riferimento

in forma di due diversi indicatori: la media annuale e il numero di giorni nell'anno solare in cui viene rilevata una concentrazione media superiore al valore 50. Per il primo indicatore, la media annuale, è stato fissato il valore $40 \mu\text{g}/\text{m}^3$ mentre per il secondo indicatore, numero di giorni con concentrazione $>50 \mu\text{g}/\text{m}^3$, è stato fissato il valore 35 giorni (pari a circa il 10 % dei giorni dell'anno).

Nei diagrammi delle rilevazioni 2006 si osservano i dati riportati nelle tabelle 5.6, 5.7, 5.8:

Rilevazioni	gg $>50 \mu\text{g}/\text{m}^3$	gg $\leq 40 \mu\text{g}/\text{m}^3$	Medie($\mu\text{g}/\text{m}^3$)
D'annunzio-Terra	50%	44.4%	57.56
D'Annunzio-Modis	62.5%	11.1%	82.74

Tabella 5.6: Rilevazioni giornaliere 2006: D'Annunzio

Rilevazioni	gg $>50 \mu\text{g}/\text{m}^3$	gg $\leq 40 \mu\text{g}/\text{m}^3$	Medie($\mu\text{g}/\text{m}^3$)
Libertà-Terra	41.2%	29.4%	50.72
Libertà-Modis	62.5%	12.5%	81.5

Tabella 5.7: Rilevazioni giornaliere 2006: Libertà

Rilevazioni	gg $>50 \mu\text{g}/\text{m}^3$	gg $\leq 40 \mu\text{g}/\text{m}^3$	Medie($\mu\text{g}/\text{m}^3$)
Volta-Terra	44.4%	44.4%	57.22
Volta-Modis	62.5	12.5%	81.43

Tabella 5.8: Rilevazioni giornaliere 2006:Volta

Da questi dati si nota che le medie annuali in entrambe le rilevazioni risultano superiori al limite previsto. Il valor medio calcolato per le rilevazioni Modis risulta molto più elevato di quello rilevato a terra a causa di due picchi rilevati il 15 Marzo e il 24 Giugno, che superano i $200 \mu\text{g}/\text{m}^3$.

Se però consideriamo gli intervalli di concentrazione mostrati in tabella 5.3, osserviamo che per le tre postazioni considerate si osserva che le rilevazioni sia terra sia modis seguono uno stesso trend, che può essere osservato nella figura seguente:

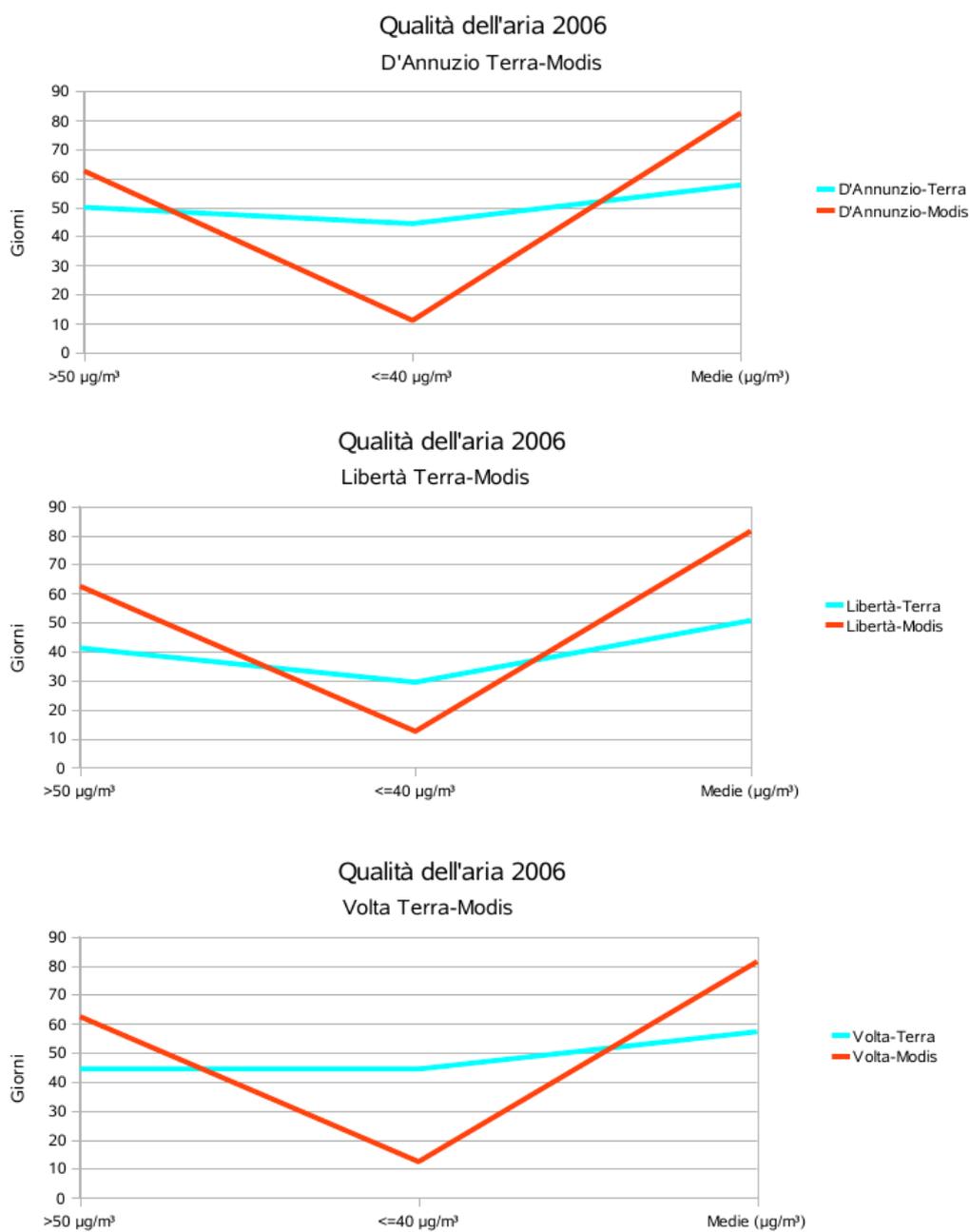


Figura 5.40: Qualità dell'aria 2006 Terra-Modis

Pertanto ai fini della valutazione della qualità dell'aria, si potrebbe dire

che entrambe le metodologie portano ad una stessa conclusione, inoltre il valor medio, sia per i dati a terra, sia per i dati Modis, rientra nel limite di una qualità dell'aria "Moderata", essendo inferiore ai $154 \mu\text{g}/\text{m}^3$ (Tab. 5.9). Da ciò si può dedurre che, per quanto riguarda i valori medi, se avessimo avuto a disposizione solo i dati ottenuti da satellite, avremmo avuto un'indicazione sulla qualità dell'aria in sintonia con quella che mostrano le rilevazioni a terra.

Qualità dell'aria	Livelli($\mu\text{g}/\text{m}^3$)
Buona	0-54
Moderata	55-154
Pericolosa per categorie a rischio	155-254
Pericolosa	255-354
Molto pericolosa	355-424

Tabella 5.9: Standard EPA di qualità dell'aria (pm10)

Per quanto riguarda il numero di giorni nell'anno solare in cui viene rilevata una concentrazione media superiore al valore di $50 \mu\text{g}/\text{m}^3$, si può osservare che in entrambi i casi il limite del 10 % dei giorni dell'anno è abbondantemente superato, infatti se consideriamo che su 18 giorni considerati abbiamo riscontrato un superamento del limite in circa la metà (nei due casi), rapportando questo valore a 365 giorni si ottiene che di questi circa 160 dovrebbero presentare valori fuori soglia, quindi il limite è superato di almeno 3 volte.

Per quanto riguarda la soglia dei $40 \mu\text{g}/\text{m}^3$, si osserva che le rilevazioni a terra presentano una percentuale di giorni con qualità dell'aria buona sempre molto superiore a quella ottenuta dai dati satellitari, si nota anche che circa il 60% delle misurazioni Modis è superiore a quelle a terra, questo dato non è verificabile, nel senso che come spiegato in precedenza i picchi registrati possono riferirsi ad un ora del giorno caratterizzata da particolari condizioni al contorno. Per questo bisogna ricordare che i valori satellitari sono reali e riferiti ad un preciso momento di un giorno, mentre quelli a terra derivano da una media giornaliera calcolata sulle 24 ore.

Per concludere possiamo dire che, dai risultati ottenuti si evince che per una valutazione di qualità dell'aria basata su un'analisi statistica dei dati

riferita a lunghi periodi, i dati ottenuti da telerilevamento e da terra risultano convergenti.

5.3.6 Analisi delle differenze di concentrazione mensile di pm10: Terra-Modis

Di seguito riportiamo i diagrammi relativi alle medie mensili relative al 2006 calcolate per i valori delle rilevazioni terra e Modis (5.41, 5.42, 5.43).

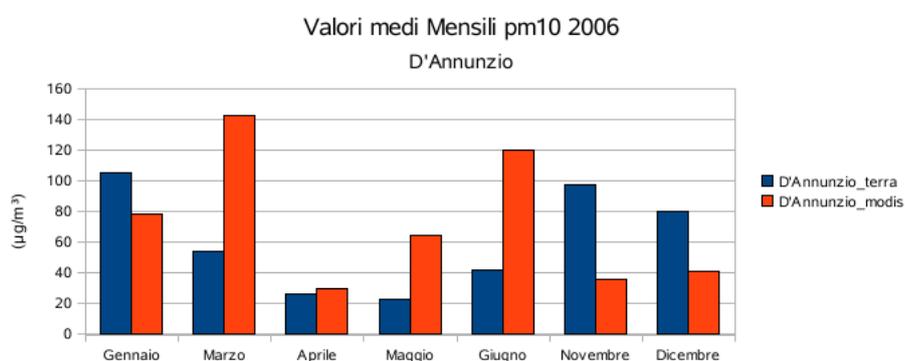


Figura 5.41: Rilevazioni mensili D'Annunzio-Modis-Terra 2006

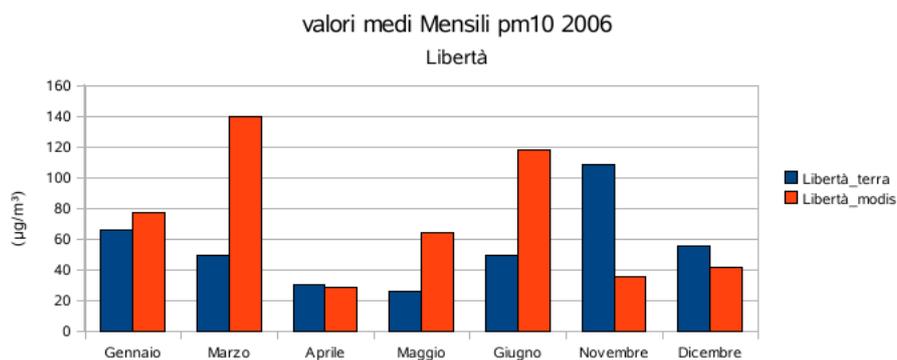


Figura 5.42: Rilevazioni mensili Libertà-Modis-Terra 2006

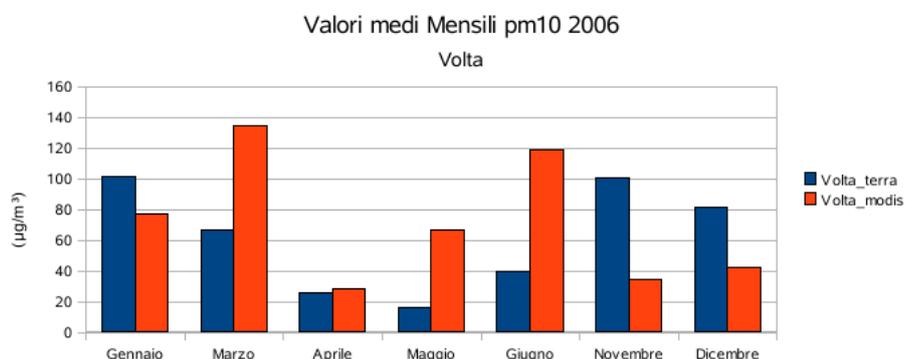


Figura 5.43: Rilevazioni mensili Volta-Modis-Terra 2006

Rilevazioni	Gennaio	Marzo	Aprile	Maggio	Giugno	Novembre	Dicembre
D'Ann.-Terra	Moderata	Moderata	Buona	Buona	Buona	Moderata	Moderata
D'Ann.-Modis	Moderata	Moderata	Buona	Moderata	Moderata	Buona	Buona

Tabella 5.10: Qualità dell'aria mensile D'Annunzio 2006

Rilevazioni	Gennaio	Marzo	Aprile	Maggio	Giugno	Novembre	Dicembre
Libertà-Terra	Moderata	Buona	Buona	Buona	Buona	Moderata	Moderata
Libertà-Modis	Moderata	Moderata	Buona	Moderata	Moderata	Buona	Buona

Tabella 5.11: Qualità dell'aria mensile Libertà 2006

Rilevazioni	Gennaio	Marzo	Aprile	Maggio	Giugno	Novembre	Dicembre
Volta-Terra	Moderata	Moderata	Buona	Buona	Buona	Moderata	Moderata
Volta-Modis	Moderata	Moderata	Buona	Moderata	Moderata	Buona	Buona

Tabella 5.12: Qualità dell'aria mensile Volta 2006

Da quanto mostrato sopra (tab. 5.10, 5.11, 5.12) si rileva che in tre casi su sette le rilevazioni Terra-Modis coincidono in un range di valori, mentre in quattro casi sono presenti delle discordanze rilevanti. Ciò che si può dedurre è che il sistema Modis rileva dei valori più alti nei mesi estivi, mentre le rilevazioni a terra risultano più alte in quelli invernali.

Tuttavia osserviamo che nonostante la discordanza esistente, questa non va oltre una certa soglia passando solo da buona a moderata e non da buona a pericolosa o viceversa, rientrando quindi in un range accettabile di errore.

Per avere una conferma di queste osservazioni possiamo esaminare gli stessi diagrammi relativi agli anni 2007 e 2008 (Fig.5.44, 5.45). Purtroppo per questi anni le date a disposizione sono poche e relative solo ai mesi invernali. Comunque l'andamento per i mesi invernali rimane confermato, si osserva che a parte il mese di Febbraio del 2008, i valori delle centraline superano, anche di molto, quelli rilevati da satellite, ma che comunque il range di valori di qualità dell'aria misurati rimane sempre tra i primi due livelli (buona-moderata).

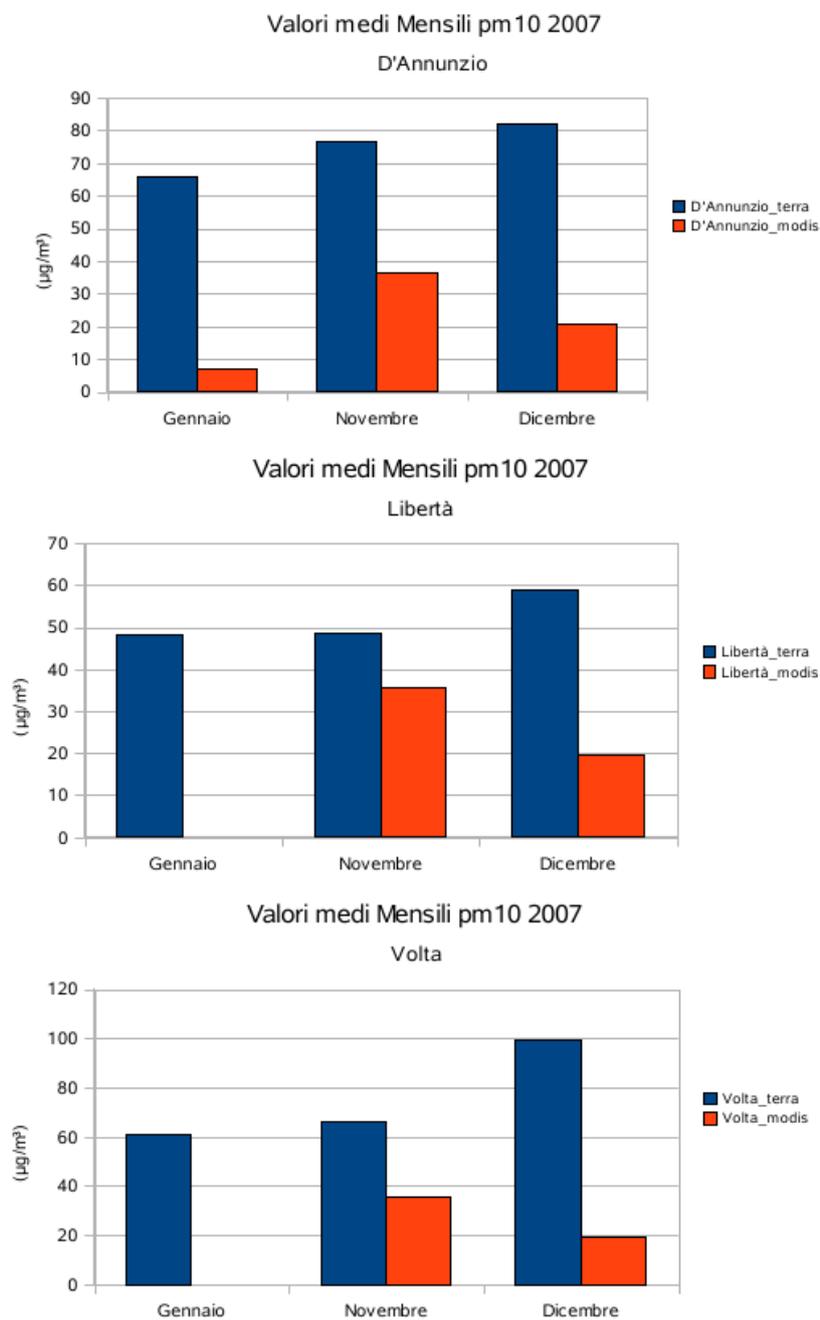


Figura 5.44: Rilevazioni mensili Modis-Terra 2007

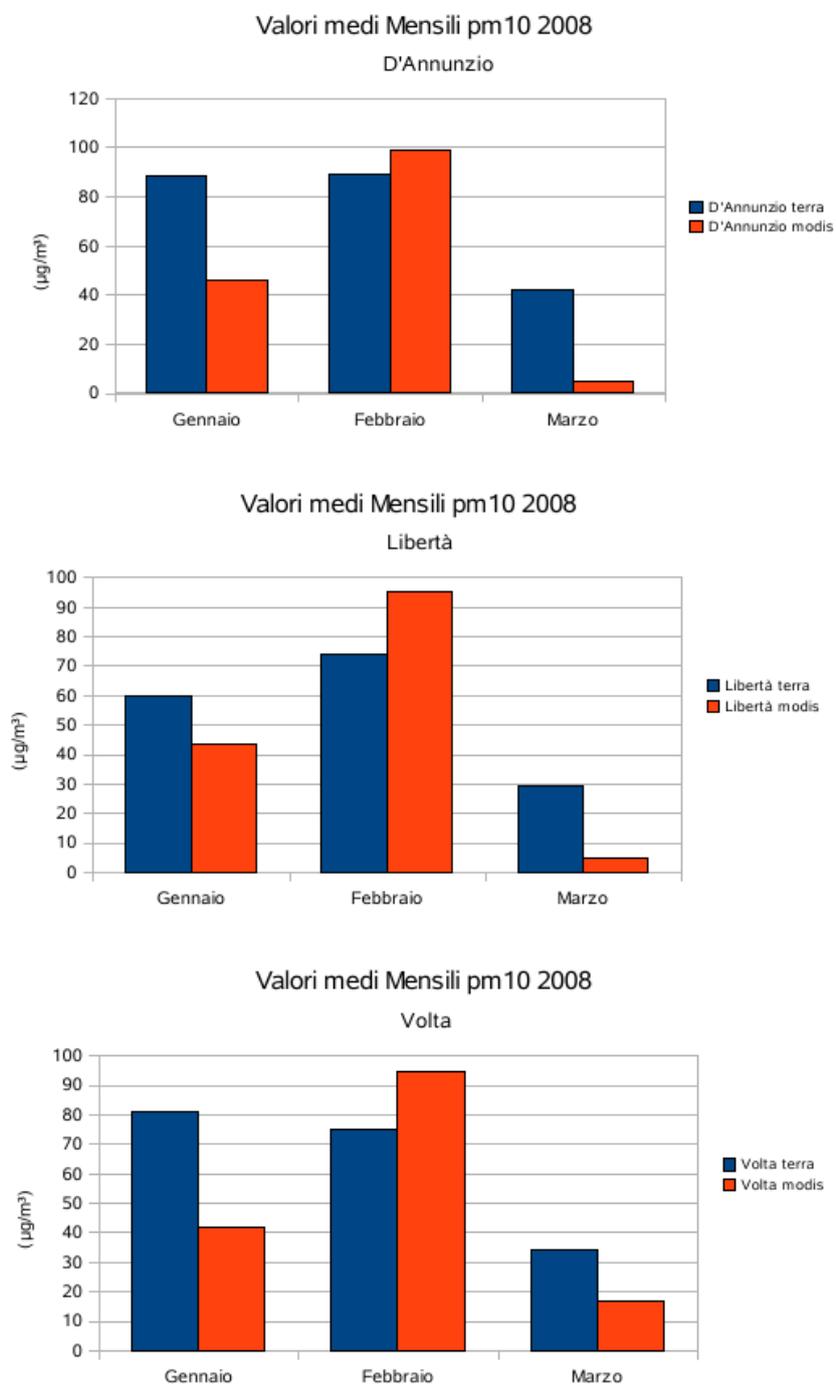


Figura 5.45: Rilevazioni mensili Modis-Terra 2008

5.4 Conclusioni

Dai confronti effettuati tra il metodo di rilevazione dati di qualità dell'aria convenzionale e quello più recente che si affida a sensori montati su satellite, si può giungere ad un primo risultato importante che da un lato non esclude l'uso di una metodologia rispetto all'altra, e dall'altro evidenzia come, almeno in un primo tempo, l'applicazione congiunta dei due metodi possa portare a risultati più completi nell'ambito di una valutazione più particolareggiata dei dati a disposizione.

Fin dall'inizio non ci si poteva aspettare che i dati ottenuti con i due metodi coincidessero perfettamente, non solo per il diverso sistema di rilevazione ma soprattutto per la natura dell'aerosol, che non si presenta come un oggetto fisso, che varia le sue caratteristiche lentamente e con continuità, come potrebbe essere la vegetazione, ma è più paragonabile ad un 'gas' in continua evoluzione, le cui caratteristiche dipendono da una vasta gamma di fenomeni sia naturali che antropici. Nonostante queste premesse sono emersi due risultati significativi:

1. i valori di concentrazione di pm10 ottenuti con i due sistemi, utilizzati per ottenere nei due casi le medie annuali, seguono lo stesso trend.
2. calcolando con gli stessi valori le medie mensili si ottengono dei livelli di qualità dell'aria che nel 40% dei casi coincidono e nel restante 60% non presentano discordanze preoccupanti.

Pertanto nel primo caso possiamo affermare che il risultato è ottimo perchè i due sistemi di rilevazione, indipendentemente l'uno dall'altro mostrano lo stesso andamento (crescente o decrescente) della qualità dell'aria passando da un anno all'altro.

Il secondo invece ci rassicura sul fatto che nell'ambito meno generale delle medie mensili, i due risultati non portano a valutazioni divergenti, ma rientrano entrambi in un range di errore trascurabile l'uno rispetto all'altro.

In questa sede ci siamo serviti di una modesta quantità di dati (Terra-Modis) per effettuare i diversi confronti, ottenendo risultati che portano ad

una valutazione positiva del metodo di rilevamento satellitare, ed è presumibile pensare che avendo a disposizione una quantità di rilevazioni molto superiore si possa giungere a conclusioni uguali o addirittura migliori, che possano validare definitivamente un metodo di rilevazione efficace ed economico finalizzato a valutazioni globali della qualità dell'aria.

C'è da dire che le centraline forniscono un dato locale, molto puntuale, mentre le osservazioni satellitari hanno il vantaggio di osservare simultaneamente aree molto vaste, pertanto sono solo queste ultime, che permettono una corretta interpretazione degli andamenti spazio-temporali delle concentrazioni di particolato. Il fatto, ad esempio, che mai si osservi un massimo di concentrazione in corrispondenza di centri abitati, insediamenti industriali o arterie principali di comunicazione, fa pensare di rimettere in discussione il presunto legame tra attività umane e concentrazioni di particolato.

Conclusioni

Tra i vari strumenti che la rivoluzione informatica ha prodotto negli ultimi anni i Sistemi Informativi Geografici (Geographic Information System, GIS), rappresentano una innovazione epocale nella gestione e nella produzione cartografica. Questi sistemi si basano sulla fusione di due capisaldi dell'innovazione informatica: i sistemi di disegno computerizzato (CAD) e i data base relazionali (DBMS), i quali peraltro sono tra le prime creazioni dell'informatica. Il primo sistema ha permesso il disegno computerizzato delle entità geografiche il secondo l'immagazzinamento dei dati e delle informazioni legate a queste entità. La fusione di questi due sistemi nei GIS ha permesso il superamento del compromesso insito in ogni rappresentazione cartografica: infatti ogni rappresentazione di entità geografiche è sempre in qualche misura simbolica e in scala, ovvero si basa su paradigmi di rappresentazione secondo i quali un determinato simbolo (es. un piccolo rettangolo) nella carta rappresenta un oggetto reale con determinate proprietà geometriche (es. una casa). Sebbene i cartografi nelle loro storia ultrasecolare abbiano sviluppato delle raffinatissime e ormai consolidate rappresentazioni simboliche, la rappresentazione simbolica di una carta geografica o tematica tradizionale rappresenta sempre un limite per una conoscenza completa di tutte le informazioni legate alle entità geografiche; ad esempio, sebbene sia abbastanza semplice rappresentare con un simbolo i contorni di un edificio, non è facile e conveniente rappresentare in forma simbolica il numero di piani dell'edificio, la lista degli inquilini che vi abitano, la presenza di garage e cantine nei seminterrati.

Il superamento di questo limite è rappresentato dalla diffusione dei Sistemi Informativi geografici (GIS), che permettono di analizzare una entità geografica, sia per la sua completa natura geometrica (e simbolica), sia per

il suo totale contenuto informativo; in pratica questi sistemi realizzano attraverso l'ingegneria del software un legame tra ogni entità geografica di una carta e un record di un database.

Le potenzialità di queste tecniche nel campo dell'analisi di sicurezza e di impatto ambientale di una qualunque installazione industriale sono enormi, al punto che ormai il GIS è uno strumento imprescindibile, potendosi con esso creare scenari ambientali dinamici e interattivi che possono essere aggiornati in tempo reale a rappresentare la situazione del territorio, che nel tempo può variare significativamente.

Una delle innovazioni più evidenti nel campo dell'Information Technology nel corso dell'ultimo decennio è il progressivo affermarsi del software Open Source. Anche il mondo della geoinformatica e dei Sistemi Informativi Geografici (GIS) ha usufruito largamente delle possibilità offerte dallo sviluppo di software Open Source, sia grazie alla qualità offerta dal peer-review del codice da parte dei moltissimi sviluppatori, sia grazie alla possibilità di migliorare e adattare il codice sorgente alle necessità personali. GRASS rappresenta, per dimensione del codice, uno dei progetti Free Software più articolati a livello mondiale. In questa sede abbiamo cercato di mettere in evidenza la versatilità e potenzialità del GRASS nella gestione di problematiche ambientali diverse, l'applicazione dei vari moduli ha permesso di ottenere un database ricco di informazioni grafiche e alfanumeriche che costituiscono la base per un'analisi dettagliata degli argomenti affrontati.

Bibliografia

- [1] *Chang K.*, “Introduction to Geographic Information System”, 4th Edition. McGraw Hill, (2007)
- [2] *M. N. DeMers*, “Fundamentals of Geographic Information Systems”, Wiley Ed. (2003)
- [3] *M. Ioannilli, U. M.A. Schiavoni*, “Fondamenti di Sistemi Informativi Geografici”, T_EXmat Ed. (2002)
- [4] *M. Neteler and H. Mitasova*, “Open Source GIS: A GRASS GIS Approach.”, Third Edition. The International Series in Engineering and Computer Science: Volume 773, (2008).
- [5] *S. Fontanari, M. Neteler*, “GRASS: un GIS open source”, ITC-irst, Trento, (2003).
- [6] *A. Clerici*, “GRASS 6.2 Manuale pratico introduttivo”, (2008)
- [7] *Ciolli, Tattoni, Vitti, Zottele, Zatelli*, “Tutorial GRASS 6.2”, risorsa online (2008)
- [8] *M. Negretti, M. Brovelli*, “Introduzione a GRASS 6”, risorsa online (2006)
- [9] *Ciolli, P. Zatelli*, “Tutorial GRASS 6.0 M”, risorsa online (2006)
- [10] *L. Lami, P. Cavallini, R. Blazek*, “L’analisi geografica Open Source: GRASS GIS parte 1^a - Vettori”, Mondogis 60-65 (51), (2005)

- [11] *Mitchell S., Csillag F., Tague C.*, “Advantages of opensource GIS to improve spatial environmental modelling”, Proceedings of the Open Source GIS-GRASS user conference 2002, Trento, Italy, 11-13, (2002).
- [12] *C. Soderlund, W. Nelson, A. Shoemaker and A. Paterson*, “SyMAP: A system for discovering and viewing syntenic regions of FPC maps”, *Genome Research* 16:1159-1168, (2006)
- [13] *Peuquet D.J.*, “Raster Data Handling in Geographic Information Systems”, Buffalo, New York: Geographic Information Systems Laboratory, State University of New York, (1977)
- [14] *S. Kaitala, A. Shavykin, V.A. Volkov*, “Environmental GIS database for the White Sea. Environmental GIS database for the White Sea”
- [15] *Hofierka J, Suri M .*, “The solar radiation model for Open source GIS. Implementation and applications”, Proceedings of the Open Source GIS-GRASS user conference 2002, Trento, Italy, 11-13, (2002)
- [16] *Eccel E., Rea R.*, “Utilizzo di GRASS ed R per la spazializzazione di dati meteorologici in una valle alpina”, *Geomatics Workbooks*, Unità operativa Agrometeorologia e Clima Istituto Agrario di S. Michele all’Adige, (2004)
- [17] *A. Clerici*, “Manuale per l’apprendimento delle caratteristiche e delle modalità d’uso di v.digit”, risorsa online, (2007)
- [18] *A. Clerici*, “Manuale pratico per l’uso di nviz (GRASS6.0)”, risorsa online, (2006)
- [19] *Chiraz P., De Rosa P., Marchesini I., Mirabella F.*, Dispense del corso tenuto a Terni dal 26 al 30 giugno 2006
- [20] *A. F. Zuur, E. N. Ieno, and E. Meesters*, “A Beginner’s Guide to R Use”, R. Springer, (2009).

- [21] *C. Cencetti, P. De Rosa, A. Fredduzzi, I. Marchesini*, “GRASS, R, GSTAT: La geostatistica per l’interpolazione della superficie piezometrica”, VIII Meeting utenti GRASS, (2007)
- [22] *Bernabei M.S.*, “Dispense: Geostatistica Distribuzione di punti”, Università di Camerino, Dipartimento di matematica e Informatica
- [23] *L. Penasa*, “La georeferenziazione di carte topografiche con GRASS GIS (GRASS 6.0)”, (2005)
- [24] *M. Neteler, L. Lami, P. Cavallini, O. Dassau*, “L’analisi geografica Open Source: GRASS GIS parte 3 Image Processing”, *MondoGIS* 69-73 (53), (2006)
- [25] *D. Penchini, C. Cencetti*, “Ricostruzione tramite GIS della superficie piezometrica della falda alluvionale nell’Alta valle del Tevere”, Università Degli Studi di Perugia, (2006)
- [26] *A. Altobelli, R. Napolitano, E. Bressan, K. Mignozzi, T. Hubina e E. Feoli*, “Prodotti MODIS per lo studio della vegetazione: Teoria, applicazione pratica e problemi di scala”, Dipartimento di Biologia – Università degli Studi di Trieste, (2008)
- [27] *Santorelli E., Dietrich S.*, “Il telerilevamento da satellite come supporto dei Sistemi Informativi Territoriali: il MODIS”, *Italian Journal of Agrometeorology* 72 - 73 (1), (2005)
- [28] *J.A. Richards*, “Remote Sensing Digital Image Analysis: An Introduction”, Terza Edizione, New York: Springer-Verlag, (1999).
- [29] *T.M Lillesand, R.W. Kiefer*, “Remote Sensing and Image Interpretation”, New York: John Wiley, (1994).
- [30] *G. Mauro, F. Carli*, Nuove tecnologie per il monitoraggio della vegetazione, indici di vegetazione a confronto per un sensore satellitare a bassa risoluzione spaziale., Dipartimento di Scienze Geografiche e Storiche – Università degli Studi di Triestel, (2006)

Indice

1	Introduzione ai GIS	4
1.1	Un po' di storia	4
1.2	Definizione e caratteristiche del GIS	10
1.3	Modello e tipologia dei dati geografici	12
1.4	Algoritmi di elaborazione	19
1.5	Volume dei dati	20
1.6	I parametri e gli standard dei dati GIS	21
1.7	Proiezioni e Sistemi di Riferimento	25
1.8	Panoramica delle funzioni GIS	27
1.8.1	Overlay	27
1.8.2	Buffering	28
1.8.3	Analisi di rete	29
1.8.4	La segmentazione dinamica	30
1.8.5	Operatori raster e tridimensionali	32
1.9	Applicazioni presenti e future	34
1.10	L'integrazione tecnologica	37
1.11	Diffusione dei dati GIS	38
2	Software per la Geostatistica	40
2.1	Introduzione alla Geostatistica	40
2.1.1	Autocorrelazione	41
2.1.2	Ipotesi di stazionarietà	42
2.1.3	Semivariogramma	43
2.1.3.1	Calcolo del variogramma	45

<i>INDICE</i>	151
2.1.4	Mappa del variogramma 47
2.1.5	Cross-Validation 48
2.1.6	Kriging e Cokriging 48
2.2	'R' : storia e caratteristiche 52
2.2.1	Funzionalità dell'ambiente R 53
2.2.2	R e la statistica ambientale 56
3	Software GIS 58
3.1	Requisiti e funzionalità 58
3.2	ArcView GIS 62
3.3	Il GIS nel panorama delle applicazioni Open-Source. 64
3.4	GRASS 67
3.4.1	Storia e diffusione 67
3.4.2	Caratteristiche 69
3.4.3	GRASS e sistemi operativi 71
3.4.4	Organizzazione dei dati in GRASS 73
3.4.5	Rappresentazione dei dati 74
3.4.6	Strumenti 76
4	Studi sul sito SOGIN di Bosco Marengo 81
4.1	Premessa 81
4.2	Dati utilizzati 82
4.3	Importazione e georeferenziazione delle immagini Satellitari . . 83
4.4	Overlay dei layer significativi e creazione del database 86
4.5	Implementazione dei dati relativi alle dosi di radiazione e crea- zione del layer Rosa dei Venti 89
4.6	Conclusioni 91
5	Valutazione della Qualità dell'Aria 92
5.1	Il ruolo dell'Aerosol nelle problematiche di inquinamento at- mosferico 92
5.2	Il Telerilevamento da satellite come supporto per le Valutazio- ni della Qualità dell'Aria : il MODIS 95
5.2.1	Architettura del sistema software 97

5.2.2	PM MAPPER : esempio di monitoraggio e uso delle mappe	101
5.3	Analisi e Gestione di mappe di Concentrazione di particolato con GRASS	104
5.3.1	Dati utilizzati	104
5.3.2	Contestualizzazione dei dati	107
5.3.3	Creazione del Database	112
5.3.4	Serie temporali	115
5.3.4.1	Valori medi Pm10 e Pm2.5 2006-2007	116
5.3.4.2	Valori medi Pm10 e Pm2.5 anno 2007	120
5.3.4.3	Valori medi Pm10 e Pm2.5 anno 2008	121
5.3.5	Analisi delle differenze di concentrazione giornaliere di pm10: Terra-Modis	133
5.3.6	Analisi delle differenze di concentrazione mensile di pm10: Terra-Modis	138
5.4	Conclusioni	143

Elenco delle figure

1.1	Rappresentazione vettoriale di punti, linee (archi) e aree (poligoni) . . .	13
1.2	Rappresentazione raster di punti, linee (archi) e aree (poligoni)	14
1.3	Rappresentazione del territorio mediante tematismi distinti, passibili di elaborazione tramite un GIS	16
2.1	Correlazione e Autocorrelazione	41
2.2	Esempio di mappa del variogramma	47
3.1	GRASS sotto Linux	71
3.2	GRASS sotto Windows/Cygnus	72
3.3	GRASS sotto Mac OS X	72
3.4	Schema di un database Grass	73
3.5	Connessioni tra oggetti Vettoriali e categorie	76
3.6	Classi specifiche	78
3.7	Classi generali	78
3.8	Operazione dissolve	79
3.9	Clip	80
3.10	Merge	80
4.1	Percorso di georeferenziazione	84
4.2	Selezione dei 'control points'	85
4.3	Immagine georeferenziata	86
4.4	Overlay dei layer vettoriali	87
4.5	Layer dell'impianto	88
4.6	Layer Rosa dei Venti	90

4.7	Concentrazione di dosi in seguito ad un rilascio acuto in aria	91
5.1	Contributi alla Radianza/Riflettanza ricevuta da un sensore satellitare .	99
5.2	Diagramma funzionale del modulo di elaborazione per il calcolo dell'AOT	100
5.3	Scatterplot dei valori di AOT estratti da PM MAPPER rispetto al prodotto MODIS sul Nord Italia nel periodo Novembre 2007 – Giugno 2008: a sinistra è mostrato il caso peggiore riscontrato (coefficiente di correlazione pari a 0.80), mentre a destra è mostrato il caso migliore (coefficiente di correlazione pari a 0.95).	101
5.4	Immagine MODIS acquisita sull' Italia il 14/02/2008 (a) , relativa mappa di concentrazione di polveri sottili PM2.5 (b).	102
5.5	Classi di pericolosità	103
5.6	Mappa di Qualità dell' Aria sull' Italia relativa all' acquisizione del 14 Febbraio 2008 (a), e particolare sul nord-est (b).	103
5.7	Esempio di immagine con scarsa copertura nuvolosa	105
5.8	Esempio di immagine con eccessiva copertura nuvolosa	106
5.9	Overlay dei confini nazionali	108
5.10	Delimitazione dell'area di interesse	109
5.11	Overlay vettoriale	110
5.12	Contestualizzazione delle Mappe.	111
5.13	Matrice vettoriale	113
5.14	Matrice di riferimento per la creazione del database	114
5.15	Esempio di diagramma relativo ad una cella della griglia di riferimento .	115
5.16	Matrice dei valori medi di pm10 anno 2006	117
5.17	Valori medi estivi di pm10 anno 2006	117
5.18	Valori medi pm10 inverno 2006/2007	118
5.19	Matrice dei valori medi di pm2.5 anno 2006	118
5.20	Matrice dei valori medi estivi di pm2.5 anno 2006	119
5.21	Matrice dei valori medi di pm2.5 inverno 2006/2007	119
5.22	Matrice dei valori medi di pm10 anno 2007	120
5.23	Matrice dei valori medi di pm2.5 anno 2007	121
5.24	Matrice dei valori medi di pm10 anno 2008	122
5.25	Matrice dei valori medi estivi di pm10 anno 2008	122

5.26	Matrice dei valori medi di pm10 inverno 2007/2008	123
5.27	Matrice dei valori medi di pm2.5 anno 2008	123
5.28	Matrice dei valori medi estivi di pm2.5 anno 2008	124
5.29	Matrice dei valori medi di pm2.5 nell'inverno 2007/2008	124
5.30	Posizione delle Centraline di monitoraggio	126
5.31	Andamenti pm10 terra-modis	127
5.32	Mappa di concentrazione di pm10, Nord-Italia 24/06/06	128
5.33	Andamenti pm10 Modis-Terra 2006-2008	129
5.34	Valori Modis-Terra rilevati in corrispondenza delle centraline	131
5.35	Valori Medi Modis-Terra rilevati in corrispondenza delle centraline	132
5.36	Andamenti medi annuali pm10 Terra-Modis	133
5.37	Rilevazioni giornaliere D'Annunzio-Modis-Terra 2006	134
5.38	Rilevazioni giornaliere Libertà-Modis-Terra 2006	134
5.39	Rilevazioni giornaliere Volta-Modis-Terra 2006	134
5.40	Qualità dell'aria 2006 Terra-Modis	136
5.41	Rilevazioni mensili D'Annunzio-Modis-Terra 2006	138
5.42	Rilevazioni mensili Libertà-Modis-Terra 2006	138
5.43	Rilevazioni mensili Volta-Modis-Terra 2006	139
5.44	Rilevazioni mensili Modis-Terra 2007	141
5.45	Rilevazioni mensili Modis-Terra 2008	142