

CAPITOLO IV- PARTICOLARITÀ SUI DATI

Descrivere su base generale tutte le particolarità dei dati è impensabile, ciò non toglie che alcune di esse vadano segnalate soprattutto in vista di applicazioni all'ambiente. Per questo, nel capitolo si parlerà dei problemi di scala spaziale che si incontrano quando si usano carte a scala diversa e quelli legati all'uso di strumenti di osservazione diversi. Si introdurranno i problemi di ergodicità che costituiscono la base per estendere all'ambiente l'uso dei metodi della meccanica statistica. Si illustreranno le opportunità connesse con la possibilità di osservare l'ambiente tramite strumenti remoti e la necessità di studiare la cinematica dei processi ambientali per arrivare a capirne la dinamica. Un paragrafo avvierà alla necessità di usare tutte le facilitazioni messe in campo dall'informatica nella rappresentazione e nella archiviazione dei dati. Il capitolo si concluderà con un accenno al problema di riconoscere la casualità dei dati e la causalità degli eventi, nonché la correlazione fra indicatori.

4.1. – Le scale delle carte geografiche

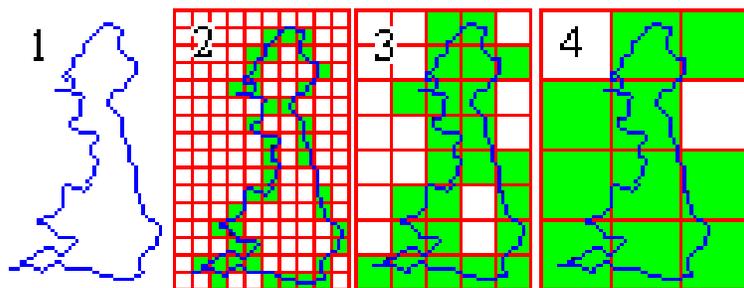


Fig. 4.1. – Misura delle coste dell'Inghilterra.

Si consideri un'immagine dell'Inghilterra (v. fig. 4.1) e le si sovrappongano tre reticoli, con maglie di lato da 1 (non esplicitato in figura), 5, 10 e 20 pixel⁽¹⁾ rispettivamente. In corrispondenza le aree di ciascuna maglia sono di 1, 25, 100 e 400 pixel.

Si contino i pixel di costa, (ovvero quelli che contengono punti di confine) per le quattro scale. Nella prima immagine di fig. 4.1 i pixel di costa risultano 225, nella seconda immagine sono 41 e nella terza 24 e nella quarta 10.

Tabella 4.1. Misura delle coste dell'Inghilterra.

Scala	Unità costiere (n. maglie)	$\Sigma_i(n. pixel) = n; \sigma_i^D$			
		σ (area)	D=0,5	D=0,6	D=0,7
1:1	225	1	225	225	225
1:5	41	25	205	283	390
1:10	24	100	240	380	603
1:20	10	400	200	364	603

La misura in pixel dell'area di costa alle tre scale risulta pertanto di 225, 1005, 2400 e 4000 pixel. Considerando le maglie costiere in termini di aree (corrispondenti ai singoli punti di contorno nel prima immagine di fig. 4.1 – elementi da 1 pixel di lato – e ai quadrati verdi nelle altre tre immagini della stessa figura) si riscontra

(v. tabella 4.1) che per D=0,7 l'area si stabilizza al valore 603. In effetti, disponendo di mappe accurate a più scale differenti, si potrebbe vedere che, variando il regolo, esistono valori dell'esponente D per i quali l'«area» diminuisce progressivamente tendendo a zero, ed altri per i quali l'«area» aumenta progressivamente tendendo ad infinito. Ma esiste anche un valore di D

⁽¹⁾ Pixel = *picture element*, cioè il componente pittorico elementare che costituisce un'immagine.

(e si dimostra essere unico – v. [Teorema di Hausdorff-Besicovitch](#)) per il quale l'«area» tende ad un valore finito. D è la «dimensione frattale» dell'oggetto e l'«area» la sua ampiezza. La dimensione frattale delle coste dell'Inghilterra è dell'ordine di 0,6-0,7 (più precisamente 0,65). Se si usa un regolo monodimensionale (ϵ_i) il ragionamento non cambia, a patto di tener conto che $\sigma_i = (\epsilon_i)^2$, cosicché la dimensione frattale in termini di lunghezze diventa doppia.

In definitiva la misura della caratteristica di un oggetto si può effettuare come descritto (o con altri metodi analoghi) definendo un valore della grandezza (finito e non nullo) e le dimensioni D relative al regolo usato per effettuare la misura. Le brevi considerazioni riportate stanno alla base degli spazi a dimensioni frazionarie e giustificano l'uso delle geometrie frattali nei problemi dell'ambiente.

4.2. – Scale e strumenti di misura

Un qualsiasi procedimento di misurazione agisce sulla grandezza da misurare come un filtro *passa banda* perché ogni misuratore ha un suo potere risolvete (risoluzione, r = estremo inferiore della banda di osservazione – ad esempio: dimensione del pixel) ed una sua ampiezza di osservazione (finestra, f = estremo superiore della banda di osservazione) cosicché la rappresentazione risulta composta da $n = f/r$ elementi significativi. In tab.4.2 sono riportati a titolo indicativo le bande di osservazione di alcuni strumenti di misura elementari.

È ragionevole ritenere che il rapporto n abbia un limite superiore dell'ordine di $10^6 \div 10^8$. La presenza di un tale limite suggerisce di approfondire la natura del rapporto n che potrebbe avere caratteristiche universali e fornire una relazione di indeterminazione intrinseca alla

Tabella 4.2. - Bande di osservazione (finestra f e risoluzione r) di alcuni strumenti elementari per misurare le grandezze fondamentali della meccanica.

GRANDEZZA	STRUMENTO	UNITA'	r	f	n
LUNGHEZZA	METRO	metro (m)	10^{-3}	10	10^4
	CALIBRO	metro (m)	10^{-4}	10^{-1}	10^3
TEMPO	OROLOGIO	secondo (s)	1	10^5	10^5
	CRONOMETRO	secondo (g)	10^{-1}	10^4	10^5
MASSA	BILANCIA	grammo (g)	1	10^4	10^4
	MICROBILANCIA	grammo (g)	10^{-3}	10^2	10^5

misurabilità di ciascun parametro⁽²⁾. Infatti, quando si vuol uscire dalla banda delle osservazioni di uno strumento (ampliare la finestra o disporre di una maggior risoluzione) si devono cambiare strumento e metodo di misura (un po' come per i passaggi di scala nel caso delle carte geografiche), compromettendo almeno in parte la continuità dei significati. Ovvero, i passaggi di scala implicano passaggi di significatività dei dati, problema perfino ovvio quando si vogliono confrontare misure a terra con misure da sensori remoti.

D'altronde anche i modelli hanno loro *bande di interpretazione* che sono proprie dei singoli modelli e che, in generale, differiscono da quelle delle osservazioni da interpretare. Da queste differenze deriva anche la diversa difficoltà e la diversa significatività dei processi di interpolazione rispetto a quelli di estrapolazione. Particolare non trascurabile è anche il fatto

⁽²⁾ Al riguardo si osserva che $1/r$ rappresenta il limite superiore dello spazio delle frequenze ottenuto trasformando secondo Fourier la variabile in esame e che le frequenze di Fourier possono avere il significato di variabile coniugata nel senso del principio quantistico di complementarità.

corrente che nelle previsioni si usa lo stesso modello prima per interpolare (quando si tara e si parametrizza il modello su un insieme di dati sperimentali) e poi per estrapolare (quando si procede alle previsioni) anche se le scale delle due operazioni non sono equivalenti.

In altre parole, ogni volta che si cambia lo strumento di osservazione, si deve assumere (implicitamente o esplicitamente) una legge f di convertibilità ($U=f(u)$) fra nuove (U) e vecchie (u) osservazioni, cioè si confida che si stia osservando la stessa cosa e lo si fa per ipotesi. Se l'ipotesi consiste nell'assumere omogenea la funzione di trasformazione (se si ammette, cioè, che f esista e che sia omogenea) si ha scienza (si possono confrontare solo grandezze commensurabili) altrimenti è opinione⁽³⁾.

L'operazione di misura, concepita come copertura con un regolo dell'oggetto da misurare, si generalizza concettualmente introducendo spazi a dimensione non intera ed i problemi di scala si superano considerando simmetrie di autosimilarità (sovrapposizioni per traslazione, rotazione e dilatazione). Nel caso deterministico diventano trattabili figure teoriche, “mostruose” per la geometria euclidea, quali la polvere di Cantor, i fiocchi di neve di Koch, il triangolo di Sierpinski, il bacino di Peano, ecc⁴.

Per i motivi descritti, ma anche per altri di tipo pratico (come l'inutilità di discutere variazioni relative di qualità ambientale con ordini di grandezza inferiori a 10^{-3} ÷ 10^{-4}) è importante che per l'ambiente il determinismo delle equazioni sia limitato e comunque tale da consentire previsioni adatte a chi deve decidere (dire: «il tempo domani è bello» è inutilmente presuntuoso. Corretto è invece stabilire che: «domani il tempo sarà bello con probabilità 0,8»). Non si può ignorare che la natura ci parla un linguaggio probabilistico e che i processi richiedono interpretazioni adeguate. Sotto quest'aspetto, alcuni *sistemi esperti* (v. capitolo 5) consentono di sviluppare un legame statistico con le osservazioni sperimentali, che può produrre numerosi vantaggi, come: (i) pesare la quantità di informazione effettivamente presente nei valori misurati; (ii) fornire risultati intrinsecamente dotati di margini d'incertezza; (iii) offrire opzioni equipollenti per rendere più agevole il compito di chi deve decidere. Anche per questo si tenta di superare le difficoltà degli strumenti deterministici con nuovi algoritmi. In particolare, la traduzione in formule di una legge costitutiva richiede l'identificazione delle scale spaziali e temporali coerenti con quelle dei dati e dei campionamenti.

D'altra parte, la sperimentazione ambientale è strettamente legata ai problemi delle scale di osservazione sia per quanto riguarda i protocolli di campionamento, sia per ciò che riguarda la significatività dei dati. E non basta: i modelli stessi sono condizionati dalla scelta delle scale temporali e spaziali che ne definiscono parametri ed algoritmi.

Non sempre, infatti, i risultati delle misure sono indifferenti alla durata ed all'ampiezza dei prelievi, né le medie temporali coincidono necessariamente con quelle d'insieme anche se, alla fine, l'ergodicità⁽⁵⁾ dei processi si assume per ipotesi. A questo si aggiunga che le scale spaziali e temporali dei fenomeni fisici, biologici, chimici, geologici, ecc., caratterizzanti un medesimo processo, possono interferire fra loro, sovrapponendosi fino anche a mascherare o

⁽³⁾È utile osservare che talvolta si tenta di rendere commensurabili grandezze diverse monetizzandole (pratica frequente in economia, ad esempio quando si valutano danni o si scambiano proprietà). In campo ambientale si ricorre alla monetizzazione degli oggetti nelle valutazioni di impatto ambientale per le quali deve essere chiaro che comunque le valutazioni sono sempre assolutamente soggettive (ad esempio, quanto vale la vita di una persona?).

⁽⁴⁾ Per approfondimenti si veda http://it.wikipedia.org/wiki/Lista_di_frattali_per_dimensione_di_Hausdorff

⁽⁵⁾ Per l'attuale contesto è ergodico un processo per il quale effettuare le medie su osservazioni ripetute nel tempo equivale ad effettuarle su repliche del processo (v. § 4.3).

alterare la percezione degli eventi.

La varietà delle scale possibili consente un loro uso gerarchico e condiziona l'algoritmo con il quale realizzare il modello. L'uso gerarchico delle scale è un modo molto efficace di procedere perché consente di graduare le difficoltà in relazione agli obiettivi, ma pone il problema del passaggio da un livello gerarchico a quello adiacente.

Ad esempio, nel caso del dilavamento di sostanze chimiche da un bacino imbrifero, sono stati costruiti modelli deterministici che stabiliscono i comportamenti annui. Questi vengono usati per scopi gestionali complessivi con l'obiettivo di indicare il contributo agricolo rispetto ad altri contributi o di confrontare gli effetti di quel bacino rispetto a quelli di altri bacini, ma anche per controllare all'interno del bacino aree a diversa destinazione. Le aree a maggior rischio vanno, invece, controllate a livello spaziale con dettaglio di campo o di sottobacino ed a livello temporale con dettaglio di evento meteorologico.

L'uso gerarchico dei modelli suggerisce di curare il trasferimento di informazioni da un livello di dettaglio all'altro della gerarchia seguendo schemi che in base alle scale definiscono il contenuto dei risultati e l'aggregazione dei parametri in modo da conferire continuità alle diverse famiglie di modelli (ad esempio: anno verso evento; bacino verso appezzamento). In altre parole, atteso che i parametri devono essere concentrati, quanto e come concentrare? Il "*quanto*" stabilisce il livello di concettualizzazione del modello; il "*come*" provvede alla possibilità di collegamenti fra diverse concettualizzazioni.

Un approccio ragionevole richiede connessioni dirette e bidirezionali fra esperimenti e modelli in modo che questi possano nascere e svilupparsi rappresentando una realtà misurabile e quelli, a loro volta, misurino una realtà simulabile, adatta a convalidare o smentire le ipotesi contenute nei modelli.

La concettualizzazione dei processi permette di trasformare un sistema in un altro che, al livello di dettaglio richiesto, rimane equivalente al primo. Questa operazione richiama quella nota in fisica col nome di *rinormalizzazione*, con la quale un sistema ad N gradi di libertà viene trasformato in un altro con $N' < N$ gradi di libertà.

In campo ambientale la concettualizzazione dei processi è oggi condotta per via intuitiva, ma nulla vieta di pensare che il legame fra sistema originale e sistema concettuale sia definibile attraverso una trasformazione funzionale di variabili macroscopiche. Questa trasformazione definisce passaggi di scala alla stessa maniera di quanto avviene per la rinormalizzazione e può essere ristretta a tipologie che garantiscono simmetrie come quelle dei frattali, già abbondantemente usati per descrivere processi dinamici o forme fantastiche con riferimenti reali (si pensi, ad esempio, alla verisimiglianza dei paesaggi sintetici di origine frattale).

Le scale, i frattali e la teoria della misura in ambiente naturale sono originariamente connessi ai problemi di geografia delle coste, ma possono dare spunti a problemi come quelli del drenaggio delle acque (ad esempio col bacino di Peano) o dello sviluppo degli agglomerati urbani.

Molte delle considerazioni effettuate avendo in mente un apparato fisico quale un miscelatore o un reattore chimico si possono estendere alle rilevazioni socio-economiche condotte con indagini campionarie. In queste intervengono anche effetti di pregiudizi psicologici, che derivano dal modo di porre le domande (dovuto a chi, come e quando l'intervistatore pone le domande e a chi, come e quando l'intervistato risponde), ed effetti di origine statistica (che derivano dalle stratificazioni del campione e dall'ipotesi di normalità delle risposte).

4.3. – Ergodicità

Gli insiemi di dati organizzati come valori consecutivi nel tempo vengono detti *serie temporali discrete* ed hanno la forma:

$$\{x\} \equiv x(t_1), x(t_2), \dots, x(t_n).$$

Spesso si indicano col nome di serie temporali gli insiemi ordinati per i quali sia:

$$t_1 < t_2 < \dots < t_n \quad \text{e} \quad \Delta t = t_2 - t_1 = t_3 - t_2 = \dots = t_n - t_{n-1}.$$

Il tempo t_i ($i = 1, 2, \dots, n$) indica l'istante al quale è riferito x che può rappresentare un valore istantaneo nell'intervallo $t+\Delta t$ o una qualche media di x in tale intervallo. Ad esempio, negli archivi di interesse meteorologico l'intervallo di media è talvolta un'ora, talvolta tre ore, talvolta un giorno, ecc.. Le medie sono ottenute con algoritmi diversi e non sempre "ortodossi".

I termini $x(t_i)$ di una serie temporale possono essere visti come *realizzazioni campionarie* di una popolazione infinita di grandezze x . Il *processo stocastico* che le genera dipende dalla risoluzione Δt , la cui ampiezza controlla il livello di correlazione fra valori successivi di x (aumentando Δt , la correlazione fra valori consecutivi di x tende a diminuire). Per analizzare i dati $\{x\}$ in termini di *pdf* (= *probability density function*) è necessario che esista una densità di probabilità invariante rispetto al tempo, cioè comune a tutti i termini $x(t_i)$. Condizione necessaria per una tale eventualità è che $\{x\}$ sia un processo *stazionario*, ovvero che abbia proprietà statistiche costanti. Inoltre, siccome le realizzazioni campionarie non sono ripetibili, per poter determinare i parametri della *pdf* dall'unico campione disponibile (la serie $\{x\}$), si deve assumere che il campione sia *ergodico*.

L'ergodicità richiede che il *valore atteso* (media di insieme) delle $x(t_i)$, indipendente da i , coincida con la media temporale (valore medio) delle osservazioni $x(t_i)$ al tendere ad infinito del loro numero. Si rimarca che, se in qualche modo è possibile stabilire con sufficiente approssimazione la stazionarietà dei dati, non esistono strumenti per analizzarne l'ergodicità, la quale va, quindi, sempre assunta. Ad esempio, per l'atmosfera molto spesso si assume la stazionarietà e l'ergodicità dei dati meteorologici, il determinismo dei processi di propagazione degli inquinanti, e la costanza del campo di emissione. In questo modo si assume anche la stazionarietà e l'ergodicità delle concentrazioni degli inquinanti.

I campi spaziali di grandezze scalari si caratterizzano in modo analogo, ma con la differenza che il riferimento invece di essere una sequenza di tempi è una matrice di coordinate a due indici. Un caso particolare dei campi spaziali sono i *transetti*, successioni di dati ordinati secondo i punti di una retta.

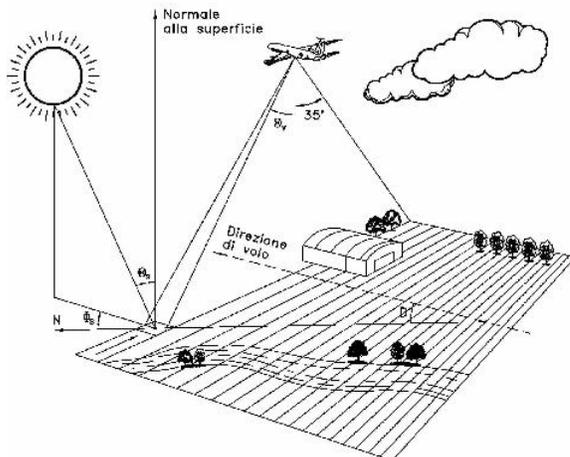


Fig.4.2 - Telerilevamento da aereo con scanner.

4.4. – Il telerilevamento

Il termine telerilevamento si riferisce a tecniche per l'acquisizione a distanza di dati relativi ad eventi puntuali, a profili, a scene

od a proprietà di oggetti. Queste tecniche riguardano misure di campi gravitazionali, elettromagnetici o barici ed utilizzano sistemi passivi (macchine fotografiche, fotometri, radiometri UV e IR, sismografi, magnetometri, gravimetri) e/o attivi (laser, sistemi a radio frequenza, radar, sonar, sismografi) installati su piattaforme terrestri, aeree, satellitari o su palloni (guidati, frenati o liberi) o su piattaforme marine, boe o navi. Per quanto riguarda l'ambiente è rilevante la possibilità di ottenere immagini del territorio e profili marini ed atmosferici. In fig.4.2 è schematizzato il funzionamento di un sensore a scansione montato su aereo.

Le immagini che si ottengono vengono analizzate con tecniche di interpretazione assistita da elaboratore permettono di studiare un'enorme quantità di fenomeni. Con il telerilevamento si possono ottenere informazioni difficilmente accessibili alle tecniche convenzionali che sono spesso incapaci di acquisire dati contemporanei e spazialmente distribuiti con altrettanta dovizia di dettaglio. Attualmente sono operativi sensori iperspettrali con risoluzioni spaziali inferiori al metro ed in grado di rilevare spettri con oltre 500 canali. Inoltre, sono possibili riprese 3-D e con satelliti geostazionari si possono effettuare analisi di una stessa scena in sequenze dotate di risoluzioni temporali molto elevate (cadenze ed esposizioni dell'ordine delle frazioni di secondo). Nei mezzi opachi alla radiazione elettromagnetica si usano onde elastiche (onde sonore, subsoniche ed ultrasoniche): in acqua con tecniche sonar e nei suoli e sottosuoli con tecniche sismografiche.

Con le immagini rilevate dai satelliti si sono raggiunti importanti risultati nella misura quantitativa dei parametri geofisici e chimico-biologici degli strati superficiali delle acque di mare, quali la salinità, la temperatura, la clorofilla, i solidi sospesi di differente origine, ecc..

Anche fenomeni di limitata estensione (ad es. le immissioni fluviali in mare) possono essere controllati da satellite, integrandone, eventualmente, i rilievi con sorvoli aerei.

Le osservazioni dallo spazio della superficie terrestre possono essere un utile complemento delle misure puntuali effettuate in loco, ma più spesso vale il senso opposto (le misure a terra supportano quelle satellitari).

Utilizzando dati da satellite e "dati verità" in loco sono stati effettuati riconoscimenti che vanno dalle varietà vegetali su suoli naturali o coltivati, fino a diverse forme di inquinamento idrico. Con l'analisi di serie temporali di immagini si può studiare sia l'evoluzione di ambienti inaccessibili, sia delicati processi di trasporto.

Il termine "elaborazione delle immagini" (*image processing*) si riferisce a tutte le operazioni eseguite su un'immagine per estrarre informazioni sugli oggetti in essa rappresentati. Le immagini possono essere digitali o analogiche, e le prime si ottengono dalle seconde con un procedimento di discretizzazione sia dell'immagine (pixel) che degli attributi.

La grande mole di dati che i sensori di telerilevamento consentono di acquisire ha reso indispensabile la messa a punto di procedure automatiche in grado di contenere i tempi di elaborazione in limiti ragionevoli e fornire informazioni tempestive sui fenomeni osservati. Le procedure automatiche consentono, inoltre, di standardizzare e rendere oggettivi i criteri di analisi così da sottoporre a precisi controlli quantitativi l'accuratezza dei risultati.

Un'immagine digitale (numerica) è una matrice bidimensionale di numeri che risulta definita dal campionamento su una griglia regolare ed ha valori generalmente espressi da numeri interi positivi. Ogni *pixel* risulta così caratterizzato da una terna di interi (m,n,I), dove m ed n sono il numero di riga e di colonna che individuano la posizione (all'interno dell'immagine digitale e sulla superficie terrestre) ed I è il valore dell'attributo associato (generalmente per taratura a terra) alla radianza (oppure alla riflettanza) misurata. La

dimensione dei pixel influisce sulla riproduzione dei dettagli della scena, ed è determinata dall'altezza del sistema di ripresa e dalle sue caratteristiche di funzionamento.

In generale, i dati raccolti dai sensori remoti richiedono preelaborazioni di “restauro” (*image restoration*), per eliminare le distorsioni introdotte durante l'acquisizione e la trasmissione. Le procedure di *correzione geometrica* e *radiometrica* che si usano servono: le prime ad eliminare le deformazioni introdotte dal sistema di ripresa e le seconde per eliminare l'errore introdotto dai sensori e l'influenza dello strato di atmosfera interposto tra i sensori e la scena investigata.

Ogni volta che un'immagine subisce una correzione o una trasformazione geometrica, i pixel vengono a trovarsi in una nuova posizione, più accurata, ma la cui radiometria non rappresenta più i valori reali misurati sulla scena. Una stima dei nuovi valori di radianza viene fatta mediante procedure di interpolazione matematica che sono dette *tecniche di ricampionamento*.

4.5. – Cinetica e dinamica

Immagine isolate non sono sufficienti a descrivere le dinamiche dell'ambiente per le quali è necessario realizzare sequenze di immagini. Anche in questo settore le tecnologie disponibili sono diverse, soprattutto in ordine alla visualizzazione (un po' meno per quanto riguarda l'analisi). Si pensi, ad esempio, di seguire l'evoluzione di una nuvola: per visualizzarla basta riprenderla con una cinepresa e proiettarne il film, ma se si prova a misurarne la «velocità di spostamento» si trova difficoltà già a partire dalla definizione del significato di tale spostamento (una nuvola infatti non trasla mai rimanendo identica a se stessa, ma cambia istante per istante di forma e consistenza). E questo nonostante gli attuali PC abbiano potenze (ampiezze di memoria e velocità di elaborazione) sufficienti per rendere accessibile a tutti, vedere, costruire ed analizzare episodi componendo e scomponendo sequenze di immagini. Questa possibilità di costruire e scomporre sequenze di immagini e di accelerare o rallentare a piacere la ripresa e la proiezione degli episodi è una possibilità che, aperta da poco, mette in campo prospettive didattiche eccezionali.

In ogni caso, l'interpretazione di sequenze di scene che costituiscono un episodio è necessaria per comprendere i processi intesi come evoluzione di stati. Si osservi, in particolare, che lo studio degli episodi consente di studiare la *cinematica* dei fenomeni, ma non ancora la loro *dinamica* (lo studio delle cause involve difficili se non impossibili analisi di causalità, che nell'attuale incertezza di conoscenza sugli spazi delle fasi scoraggia ogni ottimismo). In ogni caso, per introdurre alle difficoltà del problema, vale la pena di fare qualche considerazione esplicativa immaginando il caso, improbabile, di una proprietà scalare che evolve in un sistema continuo senza interferire con altre proprietà del sistema. Detta ψ la densità della proprietà, \underline{J} ($\underline{J} = \psi \underline{v}$) il flusso che propaga ψ nella scena (\underline{v} è la velocità di trasporto della proprietà) e g la velocità di produzione specifica (per unità di volume) della proprietà, il bilancio differenziale della proprietà (che vale per ogni punto del continuo) si scrive:

$$\frac{\partial \psi}{\partial t} + \nabla \cdot (\psi \underline{v}) + g = 0.$$

L'operatore scalare $\partial/\partial t$ e l'operatore vettoriale *nabla*, $\underline{\nabla} \equiv (\partial/\partial x, \partial/\partial y, \partial/\partial z)$, richiedono la conoscenza della distribuzione temporale e spaziale di ψ , \underline{v} e g . Su questa base e tenendo conto dei limiti di applicabilità dell'equazione (sistema continuo):

- se gli esperimenti sono in grado di misurare $\psi = \psi(x,y,z,t)$, $\underline{v} = \underline{v}(x,y,z,t)$ e $g = g(x,y,z,t)$, il bilancio fornisce la possibilità di effettuare verifiche di congruità sulle misure;
- se gli esperimenti permettono di misurare due delle tre distribuzioni spazio-temporali (si noti che di fatto le distribuzioni sono cinque perché \underline{v} è formato da tre componenti $-v_x, v_y, v_z$ che sono a loro volta distribuzioni spazio-temporali) il bilancio permette di determinare quella mancante;
- se gli esperimenti permettono di misurare solo una delle tre distribuzioni spazio-temporali, il problema rimane largamente indeterminato.

Il semplice caso illustrato mostra il percorso necessario per entrare nella cinematica degli eventi ambientali e rende ragione delle scarse prospettive disponibili a chi voglia studiare meccanismi evolutivi ignorando le tecniche di telerilevamento.

4.6. – Rappresentazione dei dati

I dati non sono mai conoscenza fine a se stessa e quelli ambientali non esauriscono i loro compiti con lo studio che li ha prodotti perché, oltre a supportare il lavoro originario, servono a formare una memoria di eventi che non torneranno più.

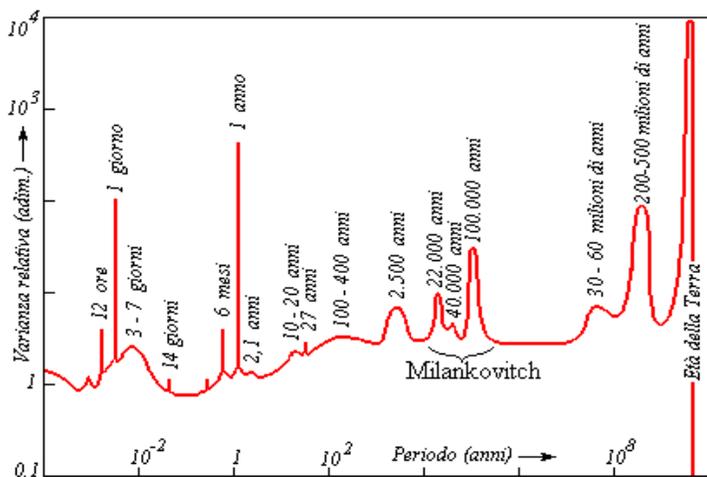


Fig. 4.3. Spettro di Fourier del clima terrestre.

Il problema della memoria storica è recente e peculiare dell'ambiente. Alcuni archivi di eventi passati sono stati realizzati direttamente dalla natura con le stratificazioni geologiche, e biologiche che parlano di forme di vita, di clima e spesso di catastrofi. Durante il secolo scorso sono state effettuate numerose scoperte di indicatori

indiretti (indicati genericamente con il nome di *proxy*) correlati con il clima della Terra e si è ricostruita una “storia termica” della Terra che sviluppata in serie di Fourier ha fornito lo spettro riportato in fig. 4.3 il quale mette in evidenza alcune frequenze particolari. Si tratta di eventi che si ripetono con il periodo indicato nella figura. Così, ad esempio, si notano i cicli diurni dovuti al succedersi del giorno e della notte, quelli annuali e semestrali legati alle stagioni, quelli di 10÷20 anni, che vengono attribuiti all'attività solare, e le periodicità di Milankovitch (v. http://en.wikipedia.org/wiki/Milankovitch_cycles), che sono associate a movimenti della Terra. Altri picchi sono ancora privi di assegnazione certa, ma tutti mostrano come, scelta una rappresentazione dei dati disponibili, se la scelta è stata appropriata si possono mettere in mostra informazioni che altre rappresentazioni non sono in grado di dare. Queste affermazioni stanno a significare come sia importante non solo disporre di dati, non solo avere documentazione degli eventi passati, ma anche scegliere la giusta ‘vetrina’ e, soprattutto, tener presente che dati di diversa origine e diversa accuratezza possono essere utilizzati in diversi contesti.

Per conservare i dati raccolti (e, simmetricamente, per ricostruirne altri da loro *proxy*) si devono tener presenti alcune precauzioni e fra queste in particolare:

- la qualità di un dato non è proprietà intrinseca del dato stesso, ma piuttosto dell'uso che se ne fa. In altre parole, non ha senso parlare in termini assoluti di «dato buono» o di «dato cattivo» perché la qualifica di «buono» o «cattivo» dipende dall'uso che se ne vuol fare, ovvero chi decide se un dato è «buono» o meno è chi lo utilizza (l'utente) e non chi lo ha prodotto (il rilevatore). Quest'ultimo, però, deve mettere il primo nelle condizioni di capirne i significati ed i limiti. Ad esempio, se si determinano i colibatteri nella laguna di Venezia non ha senso rendere pubblica la lista dei valori riscontrati, senza indicare dove, come e quando sono stati rilevati (metadati). Altrimenti un utente di quei dati (sia esso un politico, un ricercatore o un cittadino qualsiasi) non potrà giovarsene minimamente. Peggio ancora se si dice che, ad esempio, il Golfo di Napoli è inquinato senza precisare in quale punto, in che momento, in quale modo e quale sostanza sia stata rilevata. Eppure affermazioni di questo genere se ne trovano a iosa tutti i giorni, tutte prive di valore informativo anche qualora siano originariamente supportate da dati precisi e puntuali, ma occultati all'utente finale;
- ogni dato ha valore scientifico solo se è in qualche modo riproducibile, ovvero se è descritto in dettaglio il percorso seguito per ottenerlo. In altre parole, **solo un metadato esauriente rende scientifico il dato** e non la reputazione dell'operatore che l'ha realizzato. Negli esempi precedenti il «come» deve essere approfondito a livelli di pedanteria soprattutto con riferimento alla formazione del campione analizzato perché, mentre le analisi dispongono quasi sempre di metodi standardizzati, i campionatori, le operazioni di campionamento e quelle di conservazione dei campioni sono in parte lasciate alla discrezione dei singoli operatori;
- la formazione di un dato qualsiasi richiede risorse umane ed economiche spesso rilevanti (soprattutto se riguarda l'ambiente) ed è comunque uno spreco non cercare di conservarlo e di renderlo pubblico in modo da permettere anche usi diversi (generalmente imprevisi ed imprevedibili) da quelli che ne hanno indotto la produzione. Lo “spreco” in campo ambientale è anche più grave se si considera che gli eventi che hanno generato il dato non sono riproducibili;
- l'opportunità di conservare e rendere pubblica la disponibilità dei dati ha posto, fino a poco tempo fa, rilevanti problemi di memoria e di comunicazione, che ora, con i bassi costi, le grandi capacità delle memorie elettroniche e la diffusione di *internet*, si possono considerare ampiamente risolti;
- la pubblicizzazione pone a sua volta problemi di organizzazione dei dati, della facilitazione degli accessi e della automatizzazione della ricerca (motori di ricerca).

Come nota a margine, ma non per questo marginale, è bene mettere in guardia dall'uso incondizionato dell'elettronica perché l'oltranzismo anche in questo settore può produrre danni inimmaginabili. Un caso reale per tutti. È entrato nella didattica l'uso di rappresentazioni eseguite tramite calcolatore e presentate con sistemi elettronici molto eleganti (ad esempio *Power Point*) che fanno fare un salto di qualità alle conferenze scientifiche. Anche la didattica delle scienze si può giovare di tali strumenti perché si possono mostrare agli allievi disegni di una precisione e di una chiarezza impossibili da ottenere alla lavagna, ma si deve stare molto attenti alle velocità che questi apparati consentono. L'utilizzo di simili supporti per la didattica deve prevedere una struttura oculata e rispettare i tempi dell'apprendimento.

4.7. – Banche Dati

Il sapere, sia individuale che collettivo, si conquista per gradi operando su due fronti: quello dell'acquisizione di conoscenza e quello della sua memorizzazione. Nella seconda metà del secolo scorso la velocità della ricerca scientifica crebbe enormemente, producendo accumuli sempre più difficili da gestire, soprattutto per quanto riguarda la reperibilità (non poche idee rimasero dormienti in archivi prima di essere riscoperte). Finalmente, verso la fine del secolo, si è costituito un binomio formidabile, che solo di recente si è cominciato ad utilizzare, formato dalla scrittura su carta e da quella su supporti elettromagnetici. Ora l'accoppiata libri e memorie magnetiche consente di annullare il ritardo fra acquisizione e registrazione (riducendo o eliminando dalla scienza le leggende) e riducono o annullano i tempi ed i disagi del reperimento, mentre i libri costringono alle sintesi.

La natura ha lasciato le sue tracce di memoria nell'ambiente, come quelle presenti nelle strutture geologiche, chimiche, biologiche, e anche paesaggistiche. Alcune di queste memorie sono leggibili dall'uomo (come ad esempio la tettonica, la geomorfologia, la geochimica, la biologia, la genetica, la dendrologia, e così via), altre ancora probabilmente lo diventeranno. Ma tutto questo è in codice, come pure lo sono le immagini del telerilevamento, quindi va decodificato prima di poter essere depositato in archivi di memoria digitale.

Il patrimonio di conoscenza ottenuto decifrando i segnali della natura passati e presenti è sempre più ponderoso e non deve andare perduto né per abbandono, né perché dimenticato in archivi cartacei assolutamente inadeguati alla sua conservazione. Ora, che esistono le reti di calcolatori, non è ammissibile addurre scuse e tutte le conoscenze raccolte dovrebbero essere rese collettive e liberamente accessibili. Purtroppo, oltranzismi di interessi particolaristici, ignoranza, e giochi di potere insistono a violare le leggi nazionali ed europee esplicitamente dedicate ad affermare la pubblicità dei dati ambientali.

D'alto canto va anche rilevato che internet è ricchissima di dati, soprattutto statunitensi, e che l'abbondanza produce talvolta ridondanza, il che può anche disorientare, inoltre è necessario vigilare costantemente sui siti perché è effettivo il pericolo di controinformazione.

4.8. – Causalità e casualità

Ogni misura ambientale è costituita da un valore, X , associato ad un punto del territorio, $\underline{r} \equiv (x,y,z) - o$, meglio, ad un suo intorno, e ad un istante, t (o , meglio, ad un suo intorno). Ha quindi la struttura:

$$X = X(\underline{r};t) = X(x,y,z;t).$$

Una sola o poche misure forniscono scarse informazioni sui sistemi ambientali e, comunque, troppo poche per entrare nei loro funzionamenti.

Per studiare i meccanismi evolutivi di un ambiente è necessario considerare più indicatori e raccogliere misure distribuite nello spazio e nel tempo in modo da disporre di dati significativi, che si caratterizzano nel modo seguente:

$$X_i = X(\underline{r}_j;t_k) = X(x_j,y_j,z_j;t_k)$$

dove: $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; $k = 1, 2, \dots, K$, essendo I il numero degli indicatori, J il numero dei punti di prelievo e K gli istanti di prelievo.

Ma compiere ripetute misure su campioni prelevati in numerosi punti e svariate volte non dà alcuna garanzia, perché si rischia di fare misure inutili e comunque di non fare misure indispensabili. In altre parole, per studiare l'ambiente si devono effettuare campagne di misura organizzate in modo funzionale agli obiettivi che si vogliono raggiungere. Talvolta può diventare indispensabile addirittura realizzare osservatori fissi e farli funzionare con continuità nel tempo o raccogliere immagini come avviene per monitorare la meteorologia o per controllare la produttività dei terreni agricoli.

In chiave metodologica è necessario stabilire come scegliere:

- A – i siti da monitorare: quali e quanti punti e come effettuare i prelievi;
- B - le cadenze temporali per i campionamenti: quanti e quali istanti, con che frequenza e con quale durata campionare;
- C – gli indicatori: quanti, quali e come effettuare le misure.

I numeri I , J e K non sono definibili su base oggettiva, pur valendo il criterio generale secondo il quale maggiore è il loro valore, maggiore è l'informazione raccolta con la campagna di misure. Di fatto, il contenuto informativo dipende anche dall'architettura della campagna di campionamento (ad esempio, è inutile conoscere i camini dei palazzi veneziani per controllare la qualità dell'aria a Porto Marghera, mentre è indispensabile sapere numero e caratteristiche dei camini industriali di Porto Marghera per controllare la qualità dell'aria nel centro storico veneziano) che va progettata con cura cercando un ragionevole compromesso fra quantità e varietà di informazioni richieste.

In ordine ai vincoli quantitativi sul numero delle misure si usano spesso discriminanti statistici che fanno riferimento a criteri di casualità e ricorrono sia per individuare i siti ed i tempi di campionamento sia per qualificare l'informazione raccolta in campagna.

La casualità dei dati (ovvero degli eventi) è un concetto che va riferito alla teoria dei giochi d'azzardo, per la quale un sistema che non bara conserva nel tempo le chance di vittoria. Ovvero, una sequenza di numeri è casuale se non esistono regole per prevedere il numero successivo.

Formulata in questi termini la definizione è ambigua. Infatti, se si considerano le due sequenze formate da 0 e 1:

a) 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0

b) 1 0 0 1 1 0 1 1 0 1 1 1 0 0 1 0 0 0 1

si è portati ad affermare che la sequenza a) non è casuale poichè è evidente una regola per costruirla (successione di 1 e 0), mentre è casuale la sequenza b) che è stata generata lanciando una moneta ed assegnando il valore 1 a «testa» e 0 a «croce». In altre parole, la sequenza b) è casuale perché i suoi elementi hanno tutti la stessa probabilità di realizzarsi.

Si faccia attenzione che, riguardando le sequenze a) e b) come numeri di 20 cifre, entrambe fanno parte dell'insieme di 2^{20} elementi⁽⁶⁾ tutti equiprobabili e che, quindi, in tale ottica la sequenza a) rientra fra quelle casuali.

È bene anche tener presente che i «numeri casuali» vanno distinti da quelli di «rumore» che indicano valori distribuiti con legge di probabilità definita non uniforme. Così un «rumore

⁽⁶⁾ Infatti ciascun numero con $n-1$ cifre genera due numeri da n cifre. Ad esempio, 1 genera 10 e 11, mentre 0 genera 00 e 01.

bianco» è un insieme di valori distribuiti in probabilità secondo una distribuzione normale (gaussiana).

Infine, sono da stabilire criteri relativi alla causalità dei dati per evitare di confondere le cause con gli effetti e le coincidenze con le conseguenze. In questo caso ci si basa su due principi: il primo afferma che a cause uguali corrispondono effetti uguali e viceversa, mentre il secondo stabilisce che il passato ed il presente possono causare il futuro e non viceversa. Questo settore che è dominio della cibernetica è di grande interesse per i problemi delle Scienze Ambientali per le quali, però, le applicazioni sono ancora marginali.

4.9. – Correlazioni e regressioni

La quota di informazione presente nei dati ambientali varia in relazione a circostanze che spesso si possono cogliere solo a posteriori, quando si analizzano i risultati di una campagna di misure. In particolare, se si dispone di due serie di misure fra le quali esiste (o si sospetta che esista) un qualche tipo di corrispondenza legata ad un indice $i = 1, 2, \dots, N$, $x = \{x_{ij}\}$

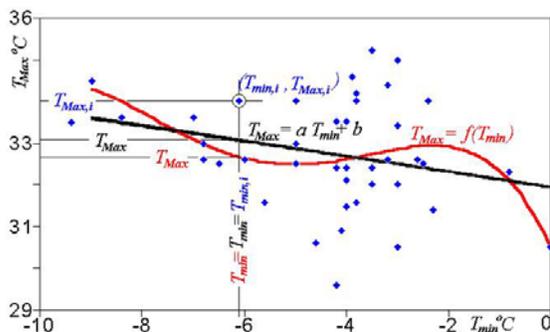


Fig.4.4 – I punti danno le temperature massime e minime degli anni 1960÷1999, le curve mostrano le relative regressioni lineare e cubica.

e $y = \{y_{ij}\}$, (ad esempio le temperature dell'aria in due punti di un territorio) si dice che le due serie di dati sono correlate intendendo che $y_i = f(x_i)$, così che le misure $\{y_{ij}\}$ non aggiungono informazione (o ne aggiungono poca) a quella contenuta nelle misure $\{x_{ij}\}$.

In effetti un legame di correlazione è meno rigido di quello espresso da una relazione funzionale, potendo andare dalla completa indipendenza delle due variabili fino alla loro completa mutua dipendenza. In particolare, stabilire il grado di correlazione fra serie di dati è importante per capire:

- quanto una serie di dati possa funzionare come *proxy* di un'altra;
- quanto, in una stessa serie di dati, due valori "prossimi vicini" siano collegati. Questa circostanza (autocorrelazione) è, a sua volta, importante per: (a) valutare la prevedibilità di un dato; (b) stabilire la possibilità di trattare un insieme ordinato di numeri (tipicamente serie temporali) con i metodi della statistica che richiedono l'indipendenza dei singoli dati dell'insieme.

Il termine «regressione» indica un'equazione matematica con la quale si rappresenta un legame di correlazione fra due insiemi ordinati di variabili casuali. In fig. 4.4 sono riportate le temperature massime annue misurate a Venezia presso l'Osservatorio Cavanis nel periodo 1960÷1999 come funzione delle corrispondenti temperature minime (<http://www.istitutoveneto.it>). Le curve (linee nera e rossa) sono le regressioni che si ottengono minimizzando i quadrati degli scarti fra valori misurati e valori calcolati con una retta e con una cubica.

L'analisi comparata di due serie temporali (analisi bivariata) si effettua a stadi nella successione seguente:

1. – si sceglie un indice di correlazione fra i dati;
2. – si calcolano i valori dell'indice di correlazione (statistiche);
3. – si applica un test di significatività alle statistiche;
4. – si calcolano i parametri di un'equazione di regressione;
5. – si valutano i limiti di confidenza della regressione.

I passi dell'analisi sono sostanzialmente soggettivi. Di fatto, non esiste un indice generale, ma si usa in prevalenza il coefficiente di correlazione $\rho(x, y)$ definito come:

$$\rho(x, y) \equiv \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}}$$

dove: $\text{cov}(x, y) \equiv \frac{1}{N} \sum_{i=1}^N [(x_i - \mu_x)(y_i - \mu_y)]$, $\text{var}(\xi) \equiv \frac{1}{N} \sum_{i=1}^N (\xi_i - \mu_\xi)^2$ e $\mu_\xi \equiv \frac{1}{N} \sum_{i=1}^N \xi_i$.

Per correlazioni lineari è: $\rho = -1$ per anticorrelazione; $\rho = +1$ per correlazione; $\rho = 0$ per assenza di correlazione. Naturalmente sono previsti tutti i casi intermedi e, nel caso lineare, il valore di ρ classifica il grado di correlazione. Nel caso di correlazioni non lineari il valore dell'indice ρ non è robusto come per quelle lineari, ma in tal caso è possibile considerare trasformazioni che linearizzino la correlazione.

Infine, si osserva che le considerazioni svolte possono essere estese a correlazioni multivariate senza grandi difficoltà, anche se si devono complicare un poco i formalismi.