

## Alcune considerazioni sull'analisi dei dati ambientali

Iacopo Borsi

TEA Sistemi SpA - Pisa  
iacopo.borsi@tea-group.com

L'analisi dei processi ambientali, compresi quelli che riguardano la risorsa idrica e il suo sfruttamento, richiede necessariamente un'ampia attività di monitoraggio dei principali parametri in gioco e, conseguentemente, l'archiviazione e l'elaborazione dei dati raccolti. La crescente disponibilità di tecnologie avanzate per l'acquisizione dei dati, la loro trasmissione ed archiviazione, spesso anche ad un costo relativamente limitato, fa sì che oggi in molte situazioni la quantità dei dati raccolti possa essere molto buona. Se è vero infatti che esistono numerosi casi di *data scarcity*, è altrettanto vero che spesso essi sono relativi a regioni in cui il monitoraggio ambientale risente di numerose difficoltà, per evidenti problemi socio-economici o geografici (regioni dell'Africa centrale, Asia, ad esempio). Potremmo dunque azzardare la tesi secondo la quale in certi casi più che la mancanza di dati il problema reale è il mal utilizzo o sotto utilizzo dei dati esistenti. Questo porta inevitabilmente ad un grado di conoscenza nettamente inferiore a quello che potenzialmente si potrebbe avere applicando procedure avanzate di analisi. Poiché la modellistica (e non solo quella idrologica e idrogeologica) non è altro che un processo più sofisticato di elaborazione dati, anche essa rientra in questo panorama.

Affrontare questo tema nella sua completezza non è certo possibile nel breve spazio di una rubrica, né per mano di una sola persona: tuttavia in questa sede vogliamo sottolineare alcuni aspetti importanti, sempre con l'obiettivo di lanciare la possibilità di proficue collaborazioni professionali, oltre che scientifiche e culturali.

Il primo punto da ricordare, forse il più importante, è quello riguardante la qualità: ogni dato, in quanto tale, è sempre affetto da un errore, derivante dallo strumento e dalla procedura con cui è stato acquisito. Sebbene possa sembrare banale ricordare questo aspetto, non lo diventa quando capita di leggere risultati di elaborazione dati (anche modellistiche) che mancano completamente della fase di analisi di confidenza del dato, che tipicamente poi rappresenta l'input per il modello (sia esso deterministico o statistico). Un esempio di scuola è l'osservazione che troviamo in Hill e Tiedeman (2007) dove si sottolinea come non sempre un modello non lineare sia necessariamente più capace di simulare il reale andamento di un fenomeno. Riprendendo l'osservazione della citata monografia, mostriamo in Fig. 1 il grafico di alcuni dati (a puro titolo di esempio didattico e privi di reale significato fisico) e della loro barra degli errori. Non considerando l'errore da cui sono affette le misure, si ipotizza un modello non lineare (linea verde) che a prima vista interpola molto bene la serie di dati in nostro possesso. Tuttavia i dati sono campioni di un processo che nella realtà è lineare (linea blu) e dunque la stima non lineare che abbiamo ottenuto senza considerare la barra degli errori restituisce un'informazione errata e fuorviante.

Ovviamente con ciò non vogliamo affermare che i modelli debbano necessariamente mantenersi il più semplici possibili, visto che spesso esistono processi in cui la non linearità deve essere mantenuta, pena, ancora una volta, la mal comprensione del fenomeno. Negli ambienti di settore, questo tipo di valutazione è spesso chiamata "*management of model nonlinearity*".

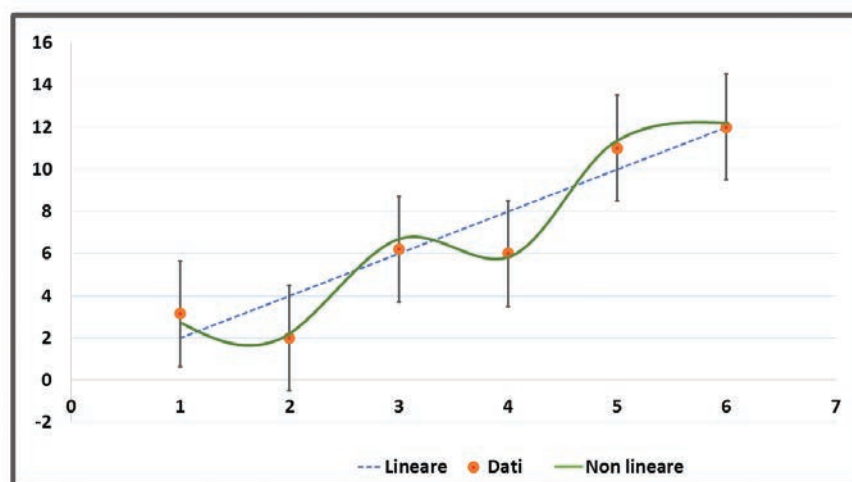


Fig. 1: Grafico di una (errata) interpolazione non lineare di dati provenienti da un processo lineare. Figura liberamente tratta da Hill e Tiedeman (2014), (Fig. 11.1, p. 269) su gentile concessione delle autrici.

(Hill, 2014). Tuttavia, siamo voluti partire da questo esempio banale per sottolineare l'importanza e l'efficacia di un'accurata analisi sull'incertezza del dato. Le procedure di *uncertainty analysis* sono fondamentali per ottenere una stima dell'errore da cui i risultati finali di un modello sono affetti. Nel campo delle acque sotterranee, ad esempio, la maggior parte del livello di incertezza deriva dalla eterogeneità che caratterizza la maggior parte del sottosuolo: i parametri derivano da misure che sono necessariamente puntuali e che vengono attribuite a zone più estese, con il grande rischio di rendere omogeneo ed isotropo ciò che nella realtà non è (Bear e Cheng, 2010).

Questo tipo di problemi sono ben noti, ed infatti sono l'oggetto principale di specifiche sotto-discipline della statistica e delle scienze della terra (come la geostatistica e la *uncertainty analysis*, appunto). Anche questo aspetto, però, comporta delle criticità. Infatti, l'esistenza di metodi avanzati utilizzati senza un opportuno grado di conoscenza può comportare l'ottenimento di risultati tutt'altro che veritieri. Questo rischio è anche indotto dall'esistenza di molti software (ad es. STATA, SAGA, AQUARIUS, etc.) che facilitano il loro utilizzo, ma che pur essendo di alta qualità tecnico-scientifica, se applicati senza ben conoscere la teoria che vi sta dietro, inducono a conclusioni del tutto prive di fondamento e, di conseguenza, a valutazioni errate. Un esempio classico (per fortuna, dobbiamo dirlo, non riscontrato nella lettura dei lavori inviati a questa rivista) è quello dell'utilizzo di metodi e test statistici "parametrici" (basati cioè su ipotesi di normalità della distribuzione dei dati) applicati a campioni statistici con distribuzioni non normali. I test di normalità, e l'eventuale necessario processo di normalizzazione, dovrebbero essere invece il primo passo di ogni analisi statistica, specialmente in campo ambientale, dove i dati che soddisfano le ipotesi di distribuzione normale, di omogeneità, di non-persistenza e stazionarietà (nel caso delle serie temporali) sono più un'eccezione che la regola (Machiwal e Jha, 2012). Altro esempio classico (da evitare) è quello dell'utilizzo di metodi di *Kriging* per l'in-

terpolazione spaziale dei dati, applicati utilizzando campioni che non sono spazialmente correlati tra loro, o meglio quando le distanze fra i punti di monitoraggio interpolati siano più grandi della distanza alla quale le proprietà misurate sono ancora realmente correlate.

In conclusione, abbiamo voluto utilizzare questo breve spazio per ricordare, ancora una volta, come la multidisciplinarietà e l'incontro di professionisti con competenze e specializzazioni diverse possa fornire un valore aggiunto agli studi e alle caratterizzazioni ambientali che vengono effettuate sul nostro territorio, siano essi a titolo di ricerca che di consulenza tecnica. Le avanzate tecnologie software per il rilevamento, la trasmissione e il salvataggio dei dati (La Vigna, 2014), assieme ad un alto livello della teoria tecnico-scientifica sull'analisi ed elaborazione dati, forniscono un panorama veramente proficuo che dovrebbe più spesso incontrarsi con i cosiddetti *data users*, cioè con tutti coloro che hanno necessità di reperire informazioni ed arricchire il quadro conoscitivo del processo ambientale studiato. Nel caso delle acque sotterranee ciò si traduce, senza dubbio, anche in un utilizzo più consapevole e sostenibile della risorsa idrica.

## BIBLIOGRAFIA

- Bear J., Cheng A.H.D. (2010), *Modeling Groundwater Flow and Contaminant Transport, Theory and Applications of Transport in Porous Media 23*, New York: Springer.
- Hill M.C e Tiedeman C.R. (2007). *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*. New York: Wiley.
- Hill M.C. (2014). Comunicazione privata (giugno 2014).
- La Vigna F. (2014). Il design delle reti di monitoraggio idrogeologiche, *Acque Sotterranee – Italian Journal of Groundwater*, Vol. 3, n. 1, 51-52.
- Machiwal, D. e Jha, M.K. (2012), *Hydrologic Time Series Analysis: Theory and Practice*. New Delhi and New York: Springer.