

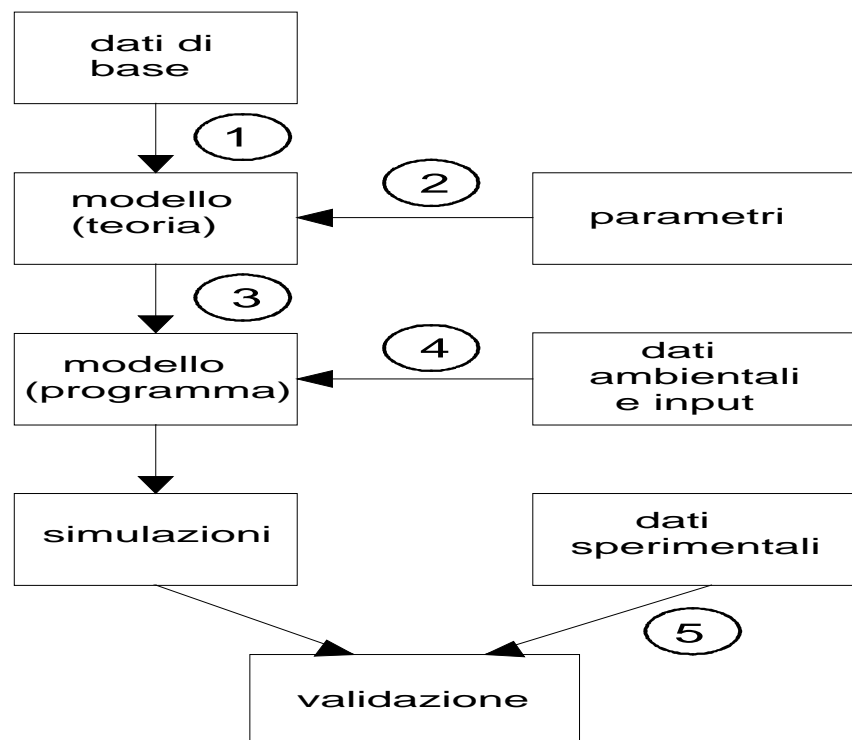
13 Valutazione dei modelli di simulazione

I modelli di simulazione non sostituiscono la conoscenza, ma sono piuttosto un mezzo per organizzarla. Quando il modello è utilizzato per analizzare un sistema attuando simulazioni, è importante capire come esso rappresenta i processi fisici, chimici e biologici coinvolti. La conoscenza del modello è infatti prerequisito fondamentale per l'utente che voglia valutarne le capacità previsionali in condizioni a lui familiari.

In questo capitolo esamineremo alcuni elementi utili nel valutare le capacità previsionali dei modelli di simulazione, valutando inoltre i possibili errori nell'uso degli stessi modelli.

Differenze tra dati simulati e misurati sperimentalmente

Come regola generale, quando si verifica una scarsa corrispondenza tra dati simulati e dati misurati, la struttura del modello dovrebbe essere l'ultima ad essere considerata responsabile per il risultato ottenuto. Le possibili ragioni per la mancata corrispondenza tra dati simulati e misurati possono essere diverse e spesso concomitanti nel determinare gli esiti delle simulazioni. Cerchiamo di schematizzare le operazioni che vanno dai dati che sono alla base del modello sino agli studi di validazione, nei quali si controlla la corrispondenza tra valori simulati dal modello e valori misurati in prove sperimentali:



Dall'esame di dati provenienti da uno o più prove sperimentali vengono formulate le teorie che costituiscono il modello e che definiscono i parametri necessari per la quantificazione dei processi. Le teorie alla base del modello sono codificate, attraverso un linguaggio di programmazione, nel programma da utilizzare su un elaboratore. Scelte una o più prove sperimentali i cui risultati si vogliono confrontare con le stime del modello, vengono forniti in input al modello stesso i dati che descrivono le condizioni in cui sono svolte le prove: dati climatici, dati relativi al suolo, valori di alcune variabili alla data di inizio della simulazione (ad

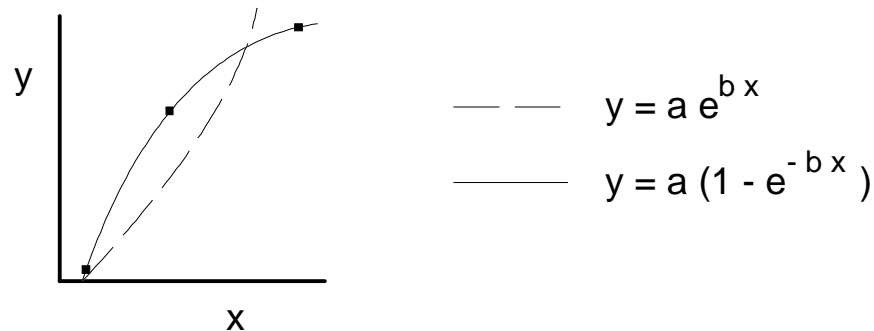
esempio: il contenuto di acqua nel suolo, il contenuto di nitrati ecc.), le operazioni colturali svolte (irrigazioni - date e quantità ; concimazioni - date, quantità e tipo di concime; ecc.), e infine i dati che descrivono il sito (latitudine, pendenza dei campi, la costante di Priestley-Taylor ecc.). Con questi dati in input vengono effettuate le simulazioni i cui risultati sono confrontati con i dati misurati nelle prove sperimentali (produzioni, contenuto di acqua nel profilo del suolo e nel tempo, superficie fogliare nel tempo ecc). E' frequente che i dati simulati non coincidano con i dati misurati. Nello schema della pagina precedente, i cinque punti indicano i tipi di errori che possono essere alla base delle discordanze osservate. Questi errori sono:

1. modello inadeguato
2. modello non calibrato
3. errori nel programma
4. errori negli input
5. errori nei risultati sperimentali

Benché questi errori possano egualmente incidere sugli esiti negativi di uno studio di validazione, è molto frequente che la causa della mancata concordanza tra dati stimati e misurati sia attribuita al punto 1. Vediamo ora quali possono essere le azioni possibili per individuare gli errori citati:

1. Valutazione del modello. Per poter esprimere effettivamente un parere sull'inadeguatezza del modello, dobbiamo preliminarmente verificare una serie di punti:

- Nel modello sono inclusi tutti i processi significativi? Ad esempio, un modello sulla ridistribuzione dell'acqua nel profilo che non tenga conto della crepacciabilità del suolo quando questa sia un processo evidente, probabilmente non può simulare correttamente i movimenti dell'acqua nel suolo in alcuni periodi dell'anno.
- I processi sono modellati correttamente? Come caso estremo ma utile per evidenziare il tipo di problema, consideriamo il grafico seguente:



E' evidente che la prima equazione non sarà mai adeguata nell'interpolare i dati, anche se potrà risultare significativa statisticamente.

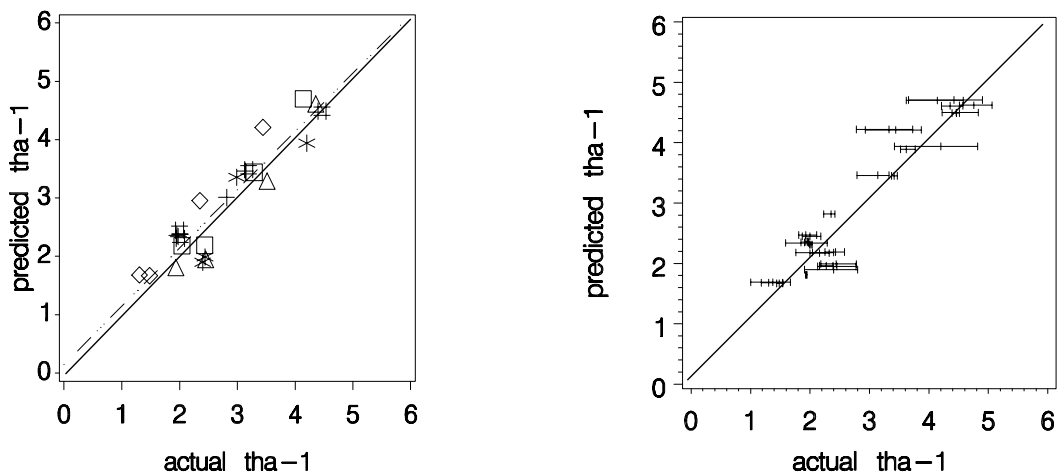
- L'intervallo di variazione dei dati che sono stati utilizzati per definire i modelli di processo era sufficientemente ampio? Nella validazione si potrebbe essere, per uno o più processi, al di fuori di tale intervallo; in tali condizioni il modello potrebbe fornire una previsione completamente errata.
2. Modello non calibrato. Sfortunatamente è necessario effettuare una calibrazione prima di poter fare un uso applicativo del modello, in quanto non esiste un modello universale che possa essere utilizzato in qualsiasi combinazione di suolo, coltura, clima e pratiche agricole. La necessità di calibrare il modello appare chiara quando si pensi a parametri che caratterizzano la specie coltivata, o come la durata del ciclo vegetativo per una coltura, che

dipende dalla varietà o ibrido utilizzata. Deve essere comunque tenuto conto che nella calibrazione possono esser fatti variare i parametri entro l'intervallo di variazione noto per ciascuno di essi. Far assumere ai parametri valori al di fuori dell'intervallo noto degrada il modello al livello dei modelli regressivi di tipo statistico (crf. cap. 9). I modelli, anche se costruiti rappresentando i processi reali, hanno comunque una base di empirismo (crf. con cap.2) che può rendere necessario un aggiustamento di parametri per rappresentare effettivamente il sistema in studio. Inoltre, talvolta la variabile indipendente usata come regressore in un submodello è variabile surrogata di altre non indipendenti tra di loro, il che rende il submodello stesso tanto più instabile quanto maggiore è la correlazione tra le variabili che sono surrogate nel regressore (il problema è sostanzialmente lo stesso che si ha nella regressione multipla quando si è in presenza di multicollinearità tra regressori).

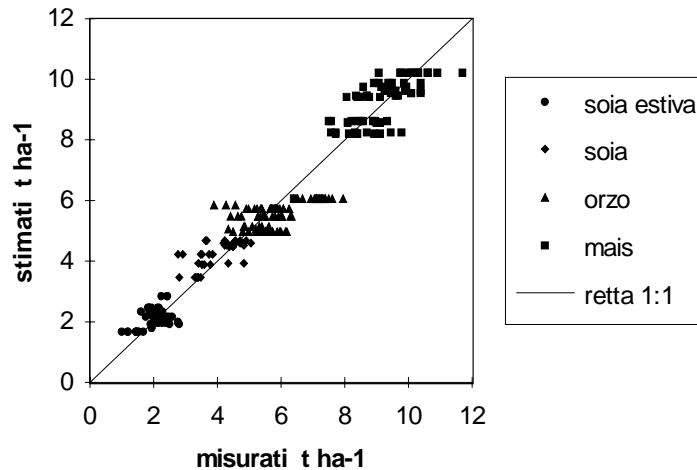
3. Errori nel programma. Con questi intendiamo gli errori di programmazione, errori sempre presenti nelle prime versioni dei programmi e che in genere vengono individuate ed eliminate con l'uso del modello. Le operazioni da eseguire per cercare di risolvere il problema sono:
 - Effettuare calcoli "a mano".
 - Effettuare bilanci di massa e energia. Se, ad esempio, il modello "crea" acqua, ne consegue che c'è un errore nel programma.
 - Verificare se i risultati delle simulazioni rientrano nella realtà biologica e fisica. Ad esempio, accumulo di biomassa con temperature sotto allo zero o evapotraspirazioni di 30 mm al giorno non sono possibili.
 - Provare il modello con un input molto variabile. Un modello che produce simulazioni verosimili con un data set in input può produrre risultati totalmente inaccettabili con un altro, il che porta a sospettare un errore nella codifica.
4. Errori negli input. L'effetto dell'errore dei valori di una variabile di input è tanto maggiore quanto più il modello è sensibile a quella variabile. Due esempi possono essere rappresentati dai dati di radiazione, o dal dato di densità apparente del suolo. E' sicuramente da evitare l'uso di files di dati meteorologici senza aver prima prodotto dei grafici delle variabili vs. il tempo, controllando andamenti ed eventuali anomalie rispetto alla norma. Il fatto che i dati siano stati raccolti da una stazione automatizzata non garantisce dall'assenza di errori anche macroscopici, così come il fatto che i dati vi siano stati forniti da esperti nel settore non garantisce sulla loro qualità. Nei dati utilizzati come input possono comunque esserci errori di microscala, per esempio quando si usino dati di una stazione meteorologica anche non molto distante dal sito nel quale sono stati condotti gli esperimenti, ma che però non è rappresentativa del clima nella località in studio.
5. Errori nei risultati sperimentali. I dati sperimentali misurati sono affetti da errore. E' in qualche modo sorprendente come in genere di questo sia tenga sempre conto nelle analisi tradizionali dei dati per le quali sono utilizzati test molto severi per separare, ad esempio, due medie, e, al tempo stesso, come la presenza dell'errore sperimentale sia del tutto ignorata quando si pretende l'assoluta concordanza tra dati simulati e dati sperimentali. Se da un lato considerare un intervallo fiduciario per i dati sperimentali porterebbe troppo spesso alla conclusione che il modello comunque è adeguato, dall'altro il considerare le medie dei dati misurati come se fossero le medie della popolazione in luogo di considerarle come *stime* delle medie conduce praticamente sempre alla valutazione che il modello ha capacità previsionali nulle. Si può affermare che non esiste via per valutare oggettivamente un modello quale che sia l'impegno posto nel farlo; peraltro vedremo nel prosieguo del capitolo alcuni criteri numerici per confrontare i modelli. In questo paragrafo si è solo voluto puntualizzare che non si può operare come se i dati sperimentali fossero esenti da errore.

Il grafico valori stimati vs. valori misurati

Una rappresentazione grafica molto efficace della concordanza tra dati simulati e dati misurati è data dai grafici di tipo valori stimati vs. dati misurati. In questi grafici è possibile valutare la dispersione dei dati intorno alla retta identità, essendo i dati al di sopra di questa retta sovrastime del modello ed essendo i dati al di sotto della retta sottostime. Possono quindi essere individuati trend particolari in rapporto alla variabile in studio o può essere individuato un intervallo di valori nel quale le capacità previsionali del modello possono essere scarse o insufficienti. Nei grafici che seguono sono rappresentati dati stimati con il modello CropSyst relativi alla produzione di soia a S. Prospero (MO), contro i dati misurati. Nel grafico di destra sono rappresentati gli stessi dati, ma in luogo delle sole medie misurate sono tracciati gli intervalli min-max. Sebbene questo tipo di rappresentazione non sia comune, la si ritiene più corretta in quanto fornisce anche un'indicazione sulla variabilità dei dati sperimentali.



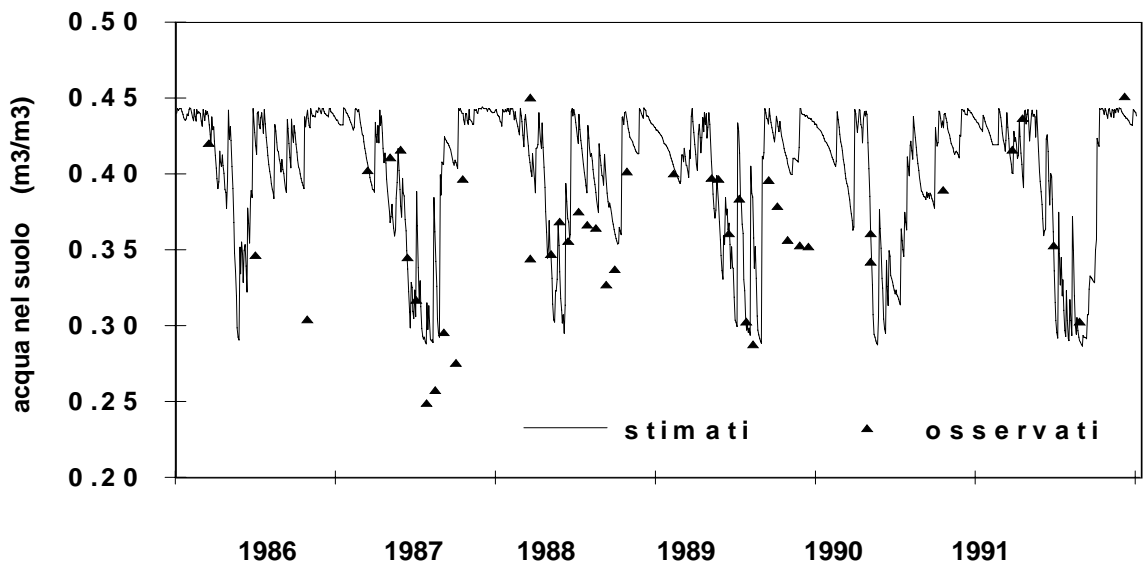
In questo tipo di grafico il confronto tra dati simulati e misurati è talvolta presentato in modo non corretto e da ciò ne consegue una valutazione errata sulle capacità previsionali del modello del modello. Ciò accade quando vengono presentati gruppi di dati diversi sia come origine (ad esempio, produzioni di colture diverse, contenuto d'acqua di suoli diversi), sia come intervallo di variazione (ad esempio si considerino i livelli produttivi tipici di soia, frumento e mais, o i contenuti d'acqua di un suolo sabbioso, di un franco e di un argilloso). I dati individuati da un valore simulato e uno misurato tendono a collocarsi sulla retta identità (o per lo meno così sembra anche per un fattore di scala), tanto che spesso vengono calcolati i parametri della retta stimati vs. misurati. La pendenza è spesso vicina ad uno (se la calibrazione del modello è stata effettuata correttamente) e r^2 assume valori elevati (benché ciò possa non avere rilevanza in assoluto, come vedremo nel paragrafo relativo agli indici numerici per la valutazione dei modelli); nel commento che ne segue talvolta viene detto che il modello globalmente è capace di stimare la variabile di risposta esaminata. Questo tipo di conclusione, sulla base di dati come quelli descritti e rappresentati esemplificativamente nel grafico che segue, è completamente erronea.



Nell'analisi dei gruppi effettuata singolarmente risulta infatti spesso che la media dei dati misurati è migliore stima dei risultati ottenibili attraverso simulazioni del modello.

Il grafico valori stimati e misurati vs. il tempo

Un ulteriore metodo per confrontare graficamente gli esiti delle simulazioni con i dati misurati è dato dai grafici in cui sono riportati i valori simulati e misurati della variabile sotto esame in ordinata e il tempo, generalmente espresso in giorni, in ascissa. Dal momento che l'intervallo d'integrazione nei modelli è in genere piuttosto breve, l'output del modello contiene una serie continua di dati che viene rappresentata con una linea; viceversa, essendo i dati misurati molti di meno, essi vengono rappresentati come punti. Un grafico di questo tipo è quello riportato di seguito:



Questi grafici permettono di valutare la presenza di eventuali "effetti deriva" nelle simulazioni e, se la scala è sufficientemente ampia, eventuali effetti stagionali. Nella valutazione di modelli dinamici è pertanto opportuno rappresentare il confronto tra dati simulati e misurati anche rispetto al tempo. Anche in questi grafici sarebbe opportuno indicare la dispersione dei dati osservati intorno alla media.

Indici numerici per la valutazione di modelli

Gli indici numerici per la valutazione di modelli appartengono essenzialmente a due categorie: quelli che misurano la correlazione tra i valori stimati (P_i) e valori osservati (O_i) e quelli che misurano la differenza tra P_i e O_i . Prima del calcolo di un qualsivoglia indice sembra opportuno riprendere quanto affermato in apertura del capitolo sulla necessità che l'utilizzatore del modello capisca capacità e limiti del modello. Se, ad esempio, un modello non simula il ciclo dell'azoto, non potrà essere effettuata la validazione dello stesso con dati sperimentali relativi a situazioni in cui ci sia stato uno stress azotato.

Misure di correlazione

Il calcolo di un indice quantitativo di associazione, covariazione o correlazione tra un valore stimato ed uno osservato può assumere una quantità di forme; tuttavia, l'indice usato quasi esclusivamente è il coefficiente di correlazione di Pearson (r) o il suo quadrato (r^2). Willmott (1984) obietta sull'uso di questo coefficiente in quanto esso non è coerentemente legato all'accuratezza delle previsioni, dove per accuratezza delle previsioni si intende la misura con cui le previsioni si avvicinano al dato osservato. Sempre Willmott (1981) ha dimostrato che insiemi di dati stimati vs. osservati assai diversi possono tutti raggiungere valori di r prossimi ad 1; inoltre, diversi lavori di comparazione tra modelli (Powell, 1980; Davies, 1981; Mc Laren Limited *et al.*, 1980) illustrano come r e r^2 non siano validi per confrontare le capacità previsionali di modelli. Accade inoltre che differenze modeste tra P_i e O_i diano luogo a valori molto bassi o negativi di r . In conclusione, l'uso di r o r^2 non sembra appropriato nel valutare le capacità previsionali di modelli di simulazione.

Misure di differenze

Fox (1981) indica diversi indici per la valutazione delle capacità previsionali dei modelli e tra questi individua come più importanti la deviazione standard ($RMSE$) e l'errore assoluto medio (MAE). Questi due indici sono calcolati come:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad [13.1]$$

$$MAE = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \quad [13.2]$$

L'importanza di questi due indici è legata al fatto che essi riassumono in essi la differenza tra P_i e O_i . Il MAE è meno sensibile a valori estremi di quanto non sia $RMSE$. Talvolta $RMSE$ è riportato come misura relativa, dividendo $RMSE$ per \bar{O} ; questo è il modo con cui lo propongono Loague e Green (1991), indicandolo ancora come $RMSE$ anche se è più propriamente un coefficiente di variabilità:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} * \frac{100}{\bar{O}} \quad [13.3]$$

Wilmott (1982) sostiene però che il rapportare l'*RMSE* alla media dei dati osservati è in generale assai discutibile, in quanto l'indice è instabile in conseguenza del fatto che \bar{O} può assumere valori prossimi allo 0. Per contro, rapportare l'errore alla media consente di confrontare le capacità previsionali di un modello rispetto a variabili i cui ordini di grandezza sono diversi. Loague e Green (1991) propongono, fra gli altri, due indici che appaiono estremamente utili: l'efficienza della modellazione (*EF*) e il coefficiente di massa residua (*CRM*). Questi due indici sono calcolati secondo le seguenti formule:

$$EF = \frac{\sum_{i=1}^n (O_i - \bar{O})^2 - \sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad [13.4]$$

$$CRM = \frac{\sum_{i=1}^n O_i - \sum_{i=1}^n P_i}{\sum_{i=1}^n O_i} \quad [13.5]$$

Il valore ottimale per *EF* è 1; valori maggiori di 0 indicano che le stime del modello sono migliori della media delle osservazioni, mentre valori negativi ci informano che la media delle osservazioni è predittore migliore del modello. Il valore ottimale di *CRM* è invece 0; valori positivi ci informano che il modello tende a sottostimare i dati misurati, mentre valori negativi indicano il contrario.

Altri indici che appare opportuno calcolare sono i parametri *a* e *b* della regressione $P_i = a + b O_i$, verificando l'ipotesi nulla $H_0 : a = 0, b = 1$. Deve però essere richiamato quanto osservato in chiusura del paragrafo relativo ai grafici valori stimati vs. valori osservati. In quel caso il valore stimato di *a* e *b* non fornisce alcuna indicazione nel valutare la capacità previsionale del modello.

In conclusione *n*, *a*, *b*, *RMSE*, *EF* e *CRM* dovrebbero comunque essere riportati quando si vogliono esprimere gli esiti della valutazione della capacità previsionale dei modelli.