

RICOSTRUZIONE DEI FLUSSI DI CO₂ DI UN ECOSISTEMA A MACCHIA MEDITERRANEA MEDIANTE RETI NEURALI

Asunis C.⁽¹⁾, Arca B.⁽²⁾, Duce P.⁽²⁾, Spano D.⁽¹⁾

⁽¹⁾Dipartimento di Economia e Sistemi Arborei, Università di Sassari, v. E. De Nicola, 9 – 07100, Sassari;

⁽²⁾CNR, Istituto per il Monitoraggio degli Agroecosistemi, v. Funtana di Lu Colbu, 4A - 07100, Sassari.

Riassunto: Nell'ultimo decennio sono state realizzate numerose reti di monitoraggio dei flussi di massa ed energia a scala regionale, che utilizzano come standard di misura la tecnica dell'Eddy Covariance. La continuità del monitoraggio dei flussi è una condizione difficilmente realizzabile a causa sia della complessità degli apparati di misura sia di alcune limitazioni di carattere modellistico: in generale dopo i controlli di congruità fisica e biologica, non più del 65% dei dati risulta utilizzabile per la costruzione dei bilanci annuali. Negli ultimi anni sono state sviluppate metodologie per la ricostruzione dei dati mancanti dei flussi di energia e materia negli ecosistemi terrestri; tuttavia, attualmente non esiste una metodologia di ricostruzione standardizzata. In questo lavoro vengono realizzati modelli basati sulle reti neurali per la ricostruzione dei flussi di CO₂ in un ecosistema a macchia mediterranea; le prestazioni fornite da tali modelli vengono confrontate con quelle di altre tecniche recentemente proposte dalla comunità internazionale. I dati sono stati raccolti nel corso del biennio 1998-99 in un sito sperimentale del progetto europeo MEDEFLEU situato nella Sardegna nord-occidentale. I modelli basati sulle reti neurali hanno fornito in genere prestazioni superiori rispetto agli altri metodi, con valori dell'errore quadratico medio compresi tra 0.9 e 2.9 $\mu\text{mol m}^{-2} \text{s}^{-1}$ di CO₂. Questi risultati confermano le possibilità applicative delle reti neurali nella ricostruzione dei flussi provenienti da reti di monitoraggio.

Parole chiave: reti neurali, eddy covariance, dati mancanti, modelli di stima.

Introduzione

La misura diretta degli scambi di massa ed energia fra biosfera e atmosfera fornisce importanti indicazioni quantitative sul contributo dei diversi ecosistemi terrestri al bilancio globale del carbonio.

Negli ultimi anni sono state realizzate diverse reti di monitoraggio dei flussi di massa ed energia a scala regionale (AmeriFlux, CarboEuroFlux, ecc.) nell'ambito del network internazionale FLUXNET. Queste reti di monitoraggio utilizzano la tecnica Eddy Covariance, basata sulla misura ad alta frequenza delle fluttuazioni degli scalari (CO₂, H₂O, ecc.) e della componente verticale della velocità del vento, come metodologia standard per la misura dei flussi di massa ed energia degli ecosistemi.

La continuità del monitoraggio dei flussi è una condizione difficilmente realizzabile a causa sia della complessità degli apparati strumentali sia di alcune limitazioni di carattere modellistico. L'analisi dei dati rilevati nei siti della rete EUROFLUX ha messo in evidenza che, nelle migliori condizioni di manutenzione del sistema, il monitoraggio copre non più dell'85% del periodo di misura (Aubinet et al. 2000); inoltre, il controllo di qualità dei dati porta a rigettare una parte delle osservazioni sperimentali, per cui mediamente, nel corso di un anno di misure, i valori dei flussi di carbonio coprono solo il 65% del periodo (Falge et al. 2001). Pertanto la realizzazione di un bilancio annuale del carbonio richiede la disponibilità di tecniche di ricostruzione dei dati. Affinché i dati provenienti da ecosistemi diversi siano confrontabili, è quindi

importante lo sviluppo di metodologie standardizzate per la ricostruzione dei dati mancanti. La comunità scientifica ha proposto varie tecniche di ricostruzione: alcune di tipo statistico basate sul calcolo, per periodi adiacenti a quello da ricostruire, dei valori medi delle grandezze (Falge et al. 2001), altre, definite semiempiriche, basate sull'individuazione del legame tra le grandezze di influenza sul fenomeno e le grandezze da ricostruire (Falge et al. 2001). A quest'ultima categoria appartengono le relazioni non lineari tra le grandezze e i fattori di controllo associati (Lloyd e Taylor, Arrhenius e Van't Hoff), che richiedono la parametrizzazione per l'ecosistema oggetto di studio (Lloyd e Taylor 1994, Falge et al. 2001).

Aubinet et al. (2000) propongono l'utilizzo delle serie storiche dei dati di grandezze di influenza sul fenomeno (temperatura e umidità dell'aria, radiazione globale, ecc.) per la messa a punto di modelli di regressione non-lineare da utilizzare per la ricostruzione dei dati mancanti: a tal fine suggeriscono l'utilizzo di tecniche di modellizzazione basate sulle reti neurali. Le reti neurali possono essere definite come strutture di elaborazione numerica di tipo adattivo, dotate cioè della capacità di apprendere le relazioni tra un set di variabili di input (x_1, \dots, x_n) e un set di variabili di output (z_1, \dots, z_n), a partire dai dati sperimentali che vengono forniti al modello come esempio. Questo strumento modellistico si è affermato rapidamente a partire dalla seconda metà degli anni ottanta, quando alcuni studi (Hecht-Nielsen 1989, Hornik et al. 1989) hanno evidenziato la superiorità delle reti neurali, rispetto agli

altri metodi di regressione, nello sfruttare l'informazione presente nei dati sperimentali. Reti neurali relativamente semplici, purché dotate di almeno uno strato intermedio di neuroni, sono in grado di apprendere qualsiasi funzione continua a un qualsiasi grado di approssimazione e vengono pertanto definite "approssimatori universali" (Hornik et al. 1989, Huntington e Cox 1997). Questa proprietà è notevolmente importante per l'individuazione delle relazioni non lineari tra variabili, che caratterizzano molti fenomeni di interesse ecologico e ambientale (Smith 1996). Rispetto ai metodi statistici basati ad esempio su regressioni polinomiali o logistiche, l'utilizzo delle reti neurali consente di evitare sia la scelta a priori di un modello non lineare da adattare ai dati sperimentali (Lek 1996, Paruelo e Tomasel 1997) sia la definizione di assunzioni teoriche (distribuzione delle varianze e degli errori) relativamente al fenomeno in studio. Nel caso dei processi che avvengono a livello di ecosistema, la relazione analitica fra le variabili non è sempre ben definibile, anche a causa dell'elevata variabilità delle caratteristiche dell'ecosistema (specie, suolo, clima, ecc.). Tuttavia, per scopi applicativi, come ad esempio quello della ricostruzione dei dati mancanti, non è indispensabile avere informazioni analitiche sul fenomeno, mentre è utile essere in grado di modellizzarlo nelle condizioni reali in cui questo si svolge, anche se ciò può limitare l'applicabilità dei modelli realizzati in aree differenti da quelle in cui sono stati messi a punto. Altra interessante proprietà delle reti neurali è la robustezza, o resistenza al rumore, che indica la capacità delle reti di operare, anche se con minore accuratezza, in presenza di dati incompleti o affetti da errori (Patterson 1996). Questa proprietà, che può essere spiegata con la struttura distribuita che caratterizza le reti neurali, è molto importante proprio per le difficoltà di misura che si riscontrano nel monitoraggio dei sistemi naturali (Vickers e Mahrt 1996, Moncrieff et al. 1997a, Moncrieff et al. 1997b). Le caratteristiche funzionali appena descritte giustificano i tentativi di realizzare modelli neurali per la ricostruzione dei valori mancanti e di determinare l'effettiva superiorità di questi modelli rispetto ad altri metodi di ricostruzione (Lek et al. 1996, Tang et al. 1996, Xia et al. 1999).

Obiettivo di questo lavoro è sviluppare modelli neurali per la ricostruzione dei flussi di CO₂ in un ecosistema a macchia mediterranea, valutarne le prestazioni e confrontarle con quelle fornite da altri metodi di ricostruzione.

Materiali e metodi

I dati sperimentali sono stati raccolti nel biennio 1998-1999 in un sito, realizzato nell'ambito del progetto europeo MEDEFU (Carbon and water fluxes of Mediterranean forests and impacts of land use/cover changes), ubicato nella Sardegna nord-occidentale (Alghero, 40° 36' N, 8° 9' E, 50 m s.l.m.), all'interno

della riserva naturale Foreste Porto Conte dell'Azienda Foreste Demaniali della Regione Sardegna.

La zona di studio è costituita da un pianoro con limitata pendenza caratterizzato da suoli poco profondi ed elevata presenza di scheletro, talvolta con evidenti rocce affioranti.

Il soprassuolo è caratterizzato da una vegetazione che rientra nell'orizzonte delle boscaglie e delle macchie litoranee delle zone costiere della Sardegna (Arrigoni 1968). Il popolamento vegetale è composto da macchia termofila mediterranea e da macchia bassa degradata con una netta predominanza di *Juniperus phoenicea* L., *Pistacia lentiscus* L., *Chamaerops humilis* L. e *Phillyrea angustifolia* L. L'altezza massima della copertura vegetale è superiore a 2.5 m per il ginepro; le altezze medie sono comprese tra 0.7 m per la palma nana e 1.4 m per il ginepro licio (Paci et al. 1999).

L'area, percorsa da incendio negli anni sessanta, è attualmente caratterizzata, oltre che dalla vegetazione a macchia bassa, da zone di rimboscimento con essenze riconducibili alle specie *Pinus halepensis* Mill. e *Pinus pinea* L.

Il clima della stazione è di tipo mediterraneo, con precipitazioni annue medie di 643 mm concentrate principalmente nel periodo autunno-invernale e periodo siccitoso esteso da maggio ad agosto. La temperatura media annua è di 16.8 °C, con escursione tra gennaio e agosto di 14 °C. Il regime dei venti presenta una prevalenza dei quadranti III e IV con una frequenza del 28% per la direzione W, del 17% per quella S, e del 20% per la provenienza da NW.

I dati sperimentali utilizzati per lo studio sono stati raccolti tramite un sistema Eddy Covariance (EC) costituito da un anemometro sonico (Gill-Solent Ultrasonic Anemometer, Gill Instruments, UK) e da un analizzatore ad infrarossi di tipo chiuso (Li-COR 6262, Li-COR, USA); il sistema ha consentito la misura della densità dei flussi di anidride carbonica, acqua, calore sensibile e calore latente.

Tramite una stazione meteorologica, sono stati inoltre rilevati i seguenti parametri: temperatura dell'aria, umidità dell'aria, radiazione globale, direzione e intensità del vento, intensità della pioggia. I valori di questi parametri sono stati acquisiti e memorizzati con frequenza di 30 minuti.

I dati registrati dal sistema EC sono stati acquisiti tramite il software EDISOL v. 0.39 (IERM, Institute of Ecology and Resource Management Edinburgh University, UK) e successivamente elaborati tramite il software EdiRe v. 1.4.3 (IERM, Institute of Ecology and Resource Management Edinburgh University, UK). L'elaborazione è stata effettuata seguendo la metodologia utilizzata nell'ambito della rete di monitoraggio EUROFLUX (Aubinet et al. 2000). Mediante un database relazionale è stato realizzato un archivio contenente i dati sperimentali provenienti dai differenti sensori allo scopo di consentire l'estrazione dei dati per il controllo di qualità e per le successive elaborazioni.

Al fine di effettuare il controllo sulla qualità dei dati istantanei è stato condotto il test sull'instazionarietà del valore medio (Foken e Wichura 1996). Sui dati mediati su un periodo di 30 minuti è stato eseguito un controllo preliminare che ha consentito di escludere i valori sperimentali non compresi all'interno dei range di congruità fisica di ciascuna grandezza. Un ulteriore controllo di qualità sui dati mediati è stato condotto analizzando i valori della velocità di frizione (Fan et al. 1995, Goulden et al. 1996b, Law et al. 1999) e individuando i dati sperimentali rilevati in condizioni di insufficiente rimescolamento turbolento ($u^* < 0.2$). Inoltre, è stato applicato il test della chiusura del bilancio di energia (Moncrieff et al. 1997a, Moncrieff et al. 1997b) imponendo una soglia arbitraria (100 W m^{-2}) alla differenza tra flusso turbolento (H+LE) ed energia disponibile (Rn-G). Infine, sui dati rilevati in condizioni di instabilità atmosferica è stato condotto il test sulle caratteristiche della turbolenza (Foken e Wichura 1996). L'analisi delle caratteristiche funzionali dell'ecosistema ha consentito l'individuazione di quattro periodi stagionali caratterizzati da una differente attività: la prima fase (I), corrispondente alla stagione invernale, da dicembre a febbraio; la seconda (II), corrispondente a quella primaverile, da marzo a giugno; la stagione estiva (III) compresa tra i mesi di luglio e agosto e quella autunnale (IV), compresa tra i mesi di settembre e novembre. In questo lavoro non vengono presentati i risultati relativi alla stagione invernale in quanto il numero di dati raccolti nel biennio sperimentale è limitato dal fatto che la sperimentazione ha avuto inizio nella primavera del 1998.

Per ciascun periodo stagionale del 1998 è stato estratto, con criterio casuale, il 75% dei dati, che sono stati utilizzati per calibrare i metodi (50%) e per i test di funzionamento (25%). Il restante 25% dell'archivio ha permesso la simulazione dei periodi con dati mancanti, durante la fase di validazione dei metodi di ricostruzione; la validazione dei metodi è stata effettuata anche su tutti i dati relativi al 1999.

I modelli neurali realizzati sono basati su reti neurali (RN) del tipo feedforward a tre strati; in Tabella 1 si riportano le principali caratteristiche strutturali delle reti neurali realizzate. Lo strato di input è costituito da cinque neuroni che ricevono in ingresso i valori delle seguenti grandezze: temperatura dell'aria, radiazione globale, deficit di tensione di vapore, ora del giorno e giorno dell'anno. Il numero di neuroni dello strato intermedio (4-10), detto anche hidden, è stato fissato inizialmente mediante regole empiriche e successivamente modificato mediante la tecnica cascade correlation (Patterson 1996). Lo strato di output è costituito da un unico neurone che fornisce i valori della densità di flusso di CO_2 . I neuroni di ciascuno strato sono connessi con tutti i neuroni dello strato adiacente. La forza di queste connessioni (peso) è stata modificata durante la fase di istruzione, attraverso un procedimento iterativo che ha portato all'individuazione del minimo livello raggiunto dall'errore quadratico medio della

differenza tra i valori stimati dalla rete e quelli misurati; questo procedimento è stato applicato utilizzando i dati di test. Le reti neurali sono state istruite mediante l'algoritmo backpropagation (Rumelhart et al. 1986). Per facilitare la ricerca del minimo assoluto (Patterson 1996) e quindi ottenere il minimo errore in fase di istruzione, al coefficiente di apprendimento (η), uno dei parametri dell'algoritmo backpropagation, è stato assegnato un valore iniziale oscillante tra 0.5 e 0.9 (Tabella 1) che, durante l'istruzione, si è dimezzato automaticamente ogni 15000 iterazioni. Il momento (α), altro importante parametro progettuale che influenza la velocità di apprendimento, è stato mantenuto costante su un valore di 0.4-0.8. La trasmissione del segnale fra neuroni è stata regolata tramite funzioni di trasferimento di tipo sigmoidale per le unità dello strato intermedio e di tipo lineare per le unità dello strato di output. I modelli neurali sono stati realizzati attraverso il software NeuralWorks Professional II/Plus (NeuralWare, Pittsburg, USA). Le prestazioni delle reti neurali sono state valutate sul set di dati di validazione del 1998 e del 1999.

Le prestazioni fornite dai modelli neurali sono state confrontate con quelle fornite da altri due metodi di ricostruzione recentemente proposti dalla comunità scientifica (Falge et al. 2001). Il primo è basato sulla realizzazione di tavole a doppio ingresso (look-up table, LUT) su base stagionale. Il secondo metodo è basato sulla ricostruzione degli andamenti medi giornalieri (mean daily variation, MDV) su una scala temporale che è stata fissata in dieci giorni.

Il metodo LUT è stato applicato raggruppando i valori della radiazione solare (R_g) in 22 classi, con intervalli di 50 W m^{-2} e i valori della temperatura (T) dell'aria in 17 classi con intervalli di $2 \text{ }^\circ\text{C}$: per ciascuna classe è stato calcolato il valore medio del flusso di CO_2 . Il metodo è stato validato ricostruendo i valori mancanti con i valori medi della tavola sinottica per le medesime classi di R_g e di T.

Il metodo MDV è stato applicato calcolando i valori medi dei flussi e la loro deviazione standard per ogni mezz'ora del giorno e per ciascuna decade della stagione considerata. Il metodo è stato validato attribuendo alle mezze ore prive di dati il valore medio della mezz'ora calcolato su una finestra temporale della stessa decade del dato da ricostruire.

Le prestazioni dei metodi sono state valutate confrontando i valori stimati e quelli misurati e calcolando i valori medi, l'errore quadratico medio (RMSE), gli errori assoluti (MAE) e i coefficienti di determinazione (R^2) della retta passante per l'origine. Per i modelli neurali è stata condotta l'analisi della sensibilità mediante un metodo empirico, ovvero analizzando la variazione dell'output dei modelli conseguente a una variazione percentuale a carico di ciascuna grandezza di input; l'analisi della sensibilità è stata condotta sui dati di validazione.

Tabella 1 – Parametri delle reti neurali realizzate.

Periodo	II	III	IV
N° neuroni input	5	5	5
N° neuroni hidden	8	7	10
N° neuroni output	1	1	1
N° iterazioni	291000	203000	285000
η	0.50	0.90	0.50
α	0.80	0.40	0.80
N° record istruzione	1052	920	1334
N° record test	372	296	440
RMSE istruzione	0.08	0.05	0.04
RMSE test	0.07	0.04	0.03

η , coefficiente di apprendimento; α , momento; RMSE, errore quadratico medio

Risultati

Il controllo di qualità ha avuto come obiettivo l'ottenimento di un set di dati sperimentali affidabile da utilizzare per la messa a punto e per la validazione dei metodi di ricostruzione; i test sono stati effettuati imponendo limiti restrittivi che mediamente hanno portato all'eliminazione di circa il 54% dei dati sperimentali (Tabella 2). In particolare, il 34.8% dei dati è stato eliminato per le condizioni di insufficiente rimescolamento turbolento ($u^* < 0.2$), il 17.4% in base alla congruità fisica dei valori dei flussi di CO_2 e l'1.5% in base al test sulla chiusura del bilancio energetico; le soglie applicate ai dati meteorologici non hanno determinato l'eliminazione di alcun valore. Gran parte dei dati eliminati sono riferibili ai periodi notturni, caratterizzati da condizioni di insufficiente rimescolamento turbolento. L'eliminazione di questi dati dal dataset ha evitato che la sottostima dei flussi misurati dal sistema determinasse un peggioramento delle prestazioni dei metodi di ricostruzione. I dati sperimentali non hanno consentito di applicare le tecniche di correzione tramite il metodo della CO_2 accumulata (storage) in quanto i dati della concentrazione di CO_2 lungo il profilo sono risultati disponibili solo per brevi intervalli di tempo. Inoltre, si è preferito non sostituire i dati dei flussi di CO_2 registrati a bassi valori di u^* con quelli verificatisi in presenza di una maggiore turbolenza nello stesso periodo e nelle medesime condizioni meteorologiche (Grelle e Lindroth 1996, Goulden et al. 1996, Law et al. 1999) in quanto non tutti gli autori sono concordi sulla validità di questa tecnica (Lee 1998, Falge et al. 2001). Le prestazioni dei differenti metodi di ricostruzione, applicati al campione di dati sperimentali relativo al 1998 sono riportate in Tabella 3. I dati dei parametri statistici considerati mostrano che i metodi di ricostruzione basati sulle reti neurali hanno fornito i migliori risultati in tutti i periodi stagionali considerati: il RMSE, che oscilla tra 0.9 e 1.9 $\mu mol m^{-2} s^{-1}$, è minore rispetto a quello fornito dagli altri metodi (1.0-2.4 $\mu mol m^{-2} s^{-1}$); così anche il MAE, che oscilla tra 0.6 e 1.3 $\mu mol m^{-2} s^{-1}$. I coefficienti di determinazione spiegano mediamente una quota della varianza oscillante tra il 67% e l'81%. I buoni risultati

offerti dalle reti neurali, oltre che in termini di errore di stima, sono

Tabella 2 – Dati eliminati (%) in seguito all'applicazione dei test per il controllo di qualità.

Test	1998	1999
Congruità fisica di C, H, LE	21.3	13.6
Soglia u^* ($> 0.2 m s^{-1}$)	35.4	34.1
Bilancio energetico	2.2	0.8
Totale	58.9	48.5

confermati soprattutto dai bassi valori del coefficiente di regressione (0.1-3.6%), rispetto a quelli ottenuti con gli altri metodi di ricostruzione (3.5-13.0%). La Figura 1 mostra la relazione tra i valori ricostruiti mediante i differenti metodi e quelli misurati nel periodo autunnale del 1998; è evidente la maggiore dispersione che caratterizza sia il metodo LUT sia, in particolare, il metodo MDV, che ha fornito in questo periodo le prestazioni meno soddisfacenti (Tabella 3).

Tabella 3 – Prestazioni dei metodi di ricostruzione dei valori delle densità di flusso di CO_2 relativi ai periodi primaverile (II), estivo (III) ed autunnale (IV) del 1998.

Periodo	Metodo	R^2	RMSE	MAE	B	n
II	LUT	0.55	2.4	1.8	4.9	363
	MDV	0.67	2.2	1.5	12.1	412
	RN	0.75	1.9	1.3	3.6	440
III	LUT	0.42	1.0	0.7	8.3	340
	MDV	0.58	1.1	0.7	10.9	398
	RN	0.67	0.9	0.6	0.5	416
IV	LUT	0.67	1.8	1.4	3.5	359
	MDV	0.66	2.1	1.3	13.4	568
	RN	0.81	1.5	1.0	-0.1	608

R^2 , coeff. di determinazione della retta passante per l'origine; RMSE, errore quadratico medio ($\mu mol m^{-2} s^{-1}$); MAE, media degli errori assoluti ($\mu mol m^{-2} s^{-1}$); B, coeff. di regressione della retta passante per l'origine (%); n, numero di osservazioni

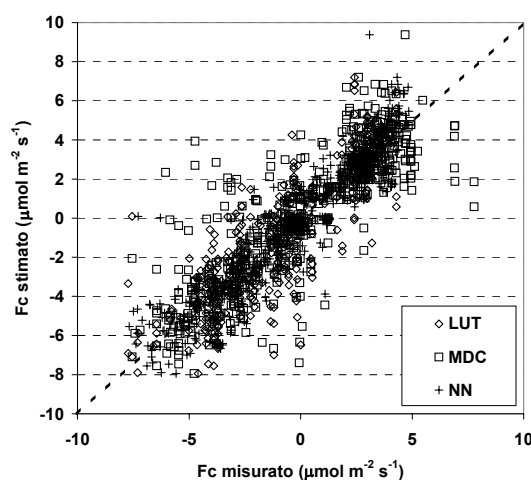


Figura 1 – Confronto tra i valori delle densità di flusso di CO₂ misurate e ricostruite mediante i metodi RN, LUT ed MDV, relativi all'autunno 1998.

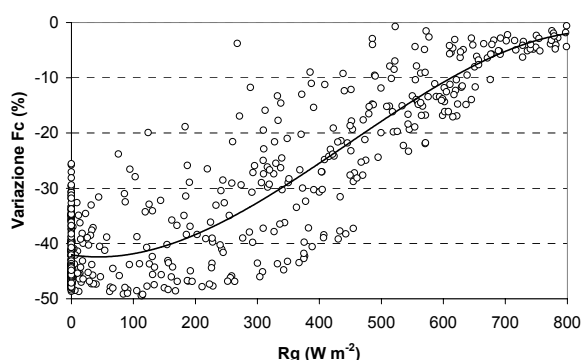


Figura 2 – Variazione percentuale della densità di flusso di CO₂ (F_c) indotta, per ciascun livello radiativo, da una variazione del 5% della radiazione globale (R_g): autunno 1998.

L'analisi di sensibilità, condotta sui modelli neurali allo scopo di evidenziare la variazione dei valori della densità di flusso di CO₂ in risposta alla variazione dei valori di ciascuna grandezza di input, ha evidenziato che l'output fornito dai modelli è sensibile soprattutto alle variazioni della radiazione solare: incrementi del 5% di questa grandezza si propagano all'interno della rete neurale determinando decrementi medi dei valori forniti dai modelli neurali, ovvero maggiori assimilazioni, del 40%, per bassi livelli radiativi, mentre per elevati livelli radiativi (> 650 W m⁻²) l'output del modello è poco sensibile alle variazioni di R_g (Figura 3). L'analisi di sensibilità ha anche evidenziato lo scarso effetto delle grandezze giorno dell'anno ed ora del giorno.

Al fine di valutare le prestazioni dei metodi di ricostruzione in presenza di condizioni sperimentali differenti da quelle utilizzate per la loro realizzazione, i dati sperimentali relativi al 1999 sono stati utilizzati per la validazione dei metodi RN e LUT già messi a punto con i dati del 1998. Il metodo MDV non è stato applicato poiché esso richiede la disponibilità di dati contigui a quelli da ricostruire. I modelli neurali confermano le buone prestazioni anche sui dati del 1999 (Tabella 4), anche se si osserva una sovrastima nel periodo estivo (B = 5.8%) e, più limitatamente, nel periodo primaverile (B = 2.3%); nel periodo autunnale si osserva invece una sottostima (B = - 3.6%). Il metodo LUT fornisce una elevata sovrastima dei valori misurati nel periodo invernale (B = 11.1%), ma fornisce risultati migliori negli altri due periodi considerati. Tuttavia, in termini sia di RMSE sia di MAE, questo metodo fornisce risultati meno soddisfacenti delle reti neurali.

Discussione

L'analisi dei risultati sperimentali evidenzia che tra tutti i metodi di ricostruzione dei dati di densità di flusso di

CO₂ misurati nel biennio 1998-99 su un ecosistema a macchia mediterranea, i modelli basati sulle reti neurali hanno fornito le migliori prestazioni, mostrando i valori più bassi degli errori assoluti, degli errori standard delle stime e dei coefficienti di regressione; inoltre, i modelli

Tabella 4 - Prestazioni dei metodi di ricostruzione dei valori delle densità di flusso di CO₂ relativi ai periodi primaverile (II), estivo (III) ed autunnale (IV) del 1999.

Periodo	Metodo	R ²	RMSE	MAE	B	n
II	LUT	0.05	3.7	2.9	1.0	2069
	RN	0.59	2.9	2.1	2.3	3159
III	LUT	0.35	1.1	0.8	-2.7	1250
	RN	0.67	0.9	0.6	5.8	1609
IV	LUT	0.35	2.3	1.8	11.1	1551
	RN	0.82	1.4	1.0	-3.6	2771

R², coefficiente di determinazione della retta passante per l'origine; RMSE, errore quadratico medio (μmol m⁻² s⁻¹); MAE, media degli errori assoluti (μmol m⁻² s⁻¹); B, coefficiente di regressione della retta passante per l'origine (%); n, numero di osservazioni

neurali non evidenziano sistematiche sovrastime o sottostime dei dati osservati. Le prestazioni fornite dai modelli neurali concordano con quelle ottenute da Aubinet et al. (2000) su coperture vegetali ad alto fusto: anche in questo caso i modelli neurali hanno fornito nella ricostruzione dei dati prestazioni superiori rispetto a un modello di regressione non lineare.

In accordo con altri lavori sulla ricostruzione dei dati (Aubinet et al. 2000, Falge et al. 2001) si ritiene che uno degli aspetti più importanti sulle possibilità applicative delle reti neurali sia rappresentato dall'ampiezza della serie storica utilizzata nella fase di apprendimento. Probabilmente, le migliori prestazioni fornite dai modelli neurali sono dovute alla capacità di utilizzare in maniera più efficiente le informazioni contenute nei dati. Ciò conferma una delle principali proprietà delle reti, in particolare di quelle feedforward multistrato, ovvero la capacità di approssimare funzioni altamente non-lineari tra variabili (Patterson 1996, Gardner e Dorling 1997). I buoni risultati forniti dai modelli neurali applicati ai dati del 1999 sono una conferma della proprietà di generalizzazione, ovvero della capacità di operare anche in presenza di condizioni sperimentali diverse da quelle utilizzate durante la fase di istruzione (Patterson 1996). Tuttavia, analogamente ad altri metodi, anche le reti neurali non possono funzionare per estrapolazione: l'utilizzo di un set di istruzione sufficientemente ampio da considerare tutte le condizioni meteorologiche e colturali che caratterizzano l'ecosistema è determinante per ottenere buoni risultati (Aubinet et al. 2000) e per evitare che la rete, apprendendo per così dire a memoria su un limitato set di dati sperimentali, mostri limitate capacità di generalizzazione (overtraining delle reti) (Smith 1996, Schultz 2001).

L'analisi della sensibilità condotta sui modelli neurali ha fornito informazioni sull'effetto, in termini di output,

della variazione di ciascuna grandezza di input. Queste informazioni, unitamente ad altri test statistici, possono in parte attenuare una delle principali critiche avanzate nei confronti di questi metodi, ovvero quella dello scarso potere esplicativo dei parametri delle reti, in particolare dei pesi delle connessioni (White 1992).

Relativamente agli altri metodi di ricostruzione, l'analisi dei risultati del lavoro evidenzia che il metodo MDV, applicato su una finestra temporale di 10 giorni ai dati del 1998, fornisce i risultati meno soddisfacenti. Come già evidenziato da altri autori (Falge et al. 2001), la validità di questo metodo è influenzata notevolmente dall'ampiezza della finestra temporale di calcolo: periodi molto lunghi determinano elevate deviazioni standard dalla media, dovute al fatto che possono variare le condizioni meteorologiche e quindi la risposta dell'ecosistema, mentre periodi troppo brevi possono rendere inapplicabile il metodo per la mancanza di sufficienti dati sperimentali. Inoltre, essendo il metodo MDV di tipo statistico, non considera gli andamenti delle grandezze ambientali che hanno influenza sull'andamento dei flussi (radiazione globale, temperatura, deficit di pressione di vapore, ecc.). Queste grandezze sono invece utilizzate per l'applicazione del metodo LUT, che ha fornito risultati più vicini a quelli delle reti neurali. Falge et al. (2001) considerano questo metodo, unitamente al metodo basato sulla parametrizzazione delle funzioni di Michaelis-Menten, Lloyd e Taylor, Arrhenius e Van't Hoff, sufficientemente attendibile e proponibile come metodo standardizzato per la ricostruzione dei dati mancanti; probabilmente tale metodo fornisce buoni risultati in presenza di serie storiche molto ampie, tali da ottenere tavole in cui siano rappresentati tutti i livelli delle grandezze di influenza sui flussi, evitando in tal modo la ricostruzione per interpolazione dei livelli mancanti. Le prestazioni meno soddisfacenti fornite dal metodo LUT, rispetto a quelle riportate da Falge et al. (2001), sono probabilmente dovute alla minore ampiezza del set di dati sperimentali utilizzato per la costruzione delle tavole; occorre anche ricordare che il metodo LUT non considera altri fattori che possono fornire informazioni aggiuntive sul fenomeno (deficit di tensione di vapore, ora del giorno, ecc.) (Falge et al. 2001). Ulteriori miglioramenti delle prestazioni dei modelli neurali potrebbero essere ottenuti utilizzando come grandezze di input anche la temperatura del suolo, l'umidità del suolo e il gradiente della concentrazione di CO₂ lungo il profilo verticale della vegetazione. La disponibilità di questi dati consentirebbe di migliorare la qualità dei dati in ingresso nei modelli e realizzare reti neurali più complete, in grado di ricostruire con maggiore accuratezza anche i flussi notturni. Infatti, in accordo con quanto osservato da altri autori (Aubinet et al. 2000, Falge et al. 2001) i risultati evidenziano che l'accuratezza dei metodi di ricostruzione dipende in primo luogo dal controllo di qualità effettuato sui dati da utilizzare per la realizzazione e la calibrazione dei metodi; in particolare, è stata evidenziata l'importanza

del controllo di qualità sui dati notturni e, eventualmente, della loro correzione per mezzo di misure aggiuntive. La mancata applicazione di queste metodologie limita la validità dei metodi di ricostruzione ai soli dati diurni e, di fatto, ne impedisce l'utilizzo per il calcolo dei bilanci annuali del carbonio. In accordo con quanto proposto da Wijk e Bouten (1999) i risultati del lavoro suggeriscono la possibilità di utilizzare le reti neurali per la modellizzazione dei flussi di CO₂ a partire dalle serie storiche di grandezze meteorologiche, quando non è disponibile un modello analitico o non si hanno informazioni sui parametri del modello.

In conclusione, i risultati sperimentali confermano l'applicabilità dei modelli neurali nella ricostruzione dei dati mancanti in ecosistemi a macchia mediterranea e rappresentano uno strumento per l'ampliamento delle serie di dati dei flussi e delle conoscenze utili alla comprensione della funzionalità di questi ecosistemi. E' evidente che ulteriori studi dovranno ampliare questi aspetti e, soprattutto, individuare le migliori tecniche da proporre come standard per la ricostruzione dei dati mancanti.

Bibliografia

- Arrigoni PW (1968). Fitoclimatologia della Sardegna. *Webbia* 23: 1-100.
- Aubinet M, Grelle A, Ibrom A, Rannik U, Moncrieff J, Foken T, Kowalski AS, Martin PH, Berbigier P, Bernhofer C, Clement R, Elbers J, Granier A, Grunwald T, Morgenstern K, Pilegaard K, Rebmann C, Snijders W, Valentini R, Vesala T (2000). Estimates of the annual net carbon and water exchange of European forests: the EUROFLUX methodology. *Advanced in Ecological Research* 30: 113-175.
- Falge E, Baldocchi D, Olson R, Anthoni P, Aubinet M, Bernhofer C, Burba G, Ceulemans R, Clement R, Dolman H, Granier A, Gross P, Grünwald T, Hollinger D, Jensen N, Katul G, Keronen P, Kowalski A, CTa Lai, Law BE, Meyers T, Moncrieff J, Moors E, Munger JW, Pilegaard K, Rannik Ü, Rebmann C, Suyker A, Tenhunen J, Tu K, Verma S, Vesala T, Wilson K, Wofsy S (2001). Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology* 107: 43-69.
- Fan SM, Goulden ML, Munger JW, Daube BC, Bakwin PS, Wofsy SC, Amthor JS, Fitzjarrald DR, Moore KE, Moore TR (1995). Environmental controls on the photosynthesis and respiration of a boreal lichen woodland: a growing season of whole-ecosystem exchange measurements by eddy correlation. *Oecologia* 102: 443-452.
- Foken T, Wichura B (1996). Tools for quality assessment of surface based flux measurement. *Agriculture and Forest Meteorology* 78: 83-105.
- Gardner MW e Dorling SR (1997). Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. *Atmospheric Environment* 32, 14/15: 2627-2636.

- Goulden ML, Munger JW, Fan SM, Daube BC, Wofsy SC (1996a). Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability. *Science* 271: 1576-1578.
- Goulden ML, Munger J, Fan W, Song-Miao, Daub BC, Wofsy SC (1996b). Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy. *Global Change Biology* 2: 169-182.
- Grelle A, Lindroth A (1996). Eddy-correlation for long term monitoring of fluxes of heats, water vapour and CO₂. *Global Change Biology* 2: 297-307.
- Hecht-Nielsen R. (1989). *Neurocomputing*, Addison-Wesley, Reading.
- Hornik K, Stinchcombe M, White H (1989). Multilayer feedforward network are universal approximators, *Neural Networks* 2: 359-366.
- Huntingford C, Cox PM (1997). Use of statistical and neural network techniques to detect how stomatal conductance responds to changes in the local environment. *Ecological Modelling* 97: 217-246.
- Law BE, Ryan MG, Anthoni PM (1999). Seasonal and annual respiration of a ponderosa pine ecosystem. *Global Change Biology* 5: 169-182.
- Lee X (1998). On micrometeorological observations of surface-air exchange over tall vegetation. *Agricultural and Forest Meteorology* 91: 39-50.
- Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S (1996). Application of neural networks to modelling non-linear relationships in ecology. *Ecological Modelling* 90: 39-52.
- Lloyd J, Taylor JA (1994). On the temperature dependence of soil respiration. *Functional Ecology* 8: 315-323.
- Moncrieff J, Valentini R, Greco S, Seufert G, Ciccioli P (1997a). Trace gas exchange over terrestrial ecosystems: methods and perspectives in micrometeorology. *Journal of Experimental Botany* 48: 1133-1142.
- Moncrieff JB, Massheder JM, De Bruin H, Elbers J, Friborg T, Heusinkveld B, Kabat P, Scott S, Soegaard H, Verhoef A (1997b). A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. *Journal of Hydrology* 188-189: 589-611.
- Paci M, Pellizzaro G, Spano D, Asunis C, Cesaraccio C, Sirca C (1999). Analisi della distribuzione spaziale della vegetazione: un esempio di applicazione alla macchia mediterranea nella penisola di Capo Caccia (Sassari). *Monti e Boschi* 5: 17-25.
- Paruelo M, Tomasel F (1997). Prediction of functional characteristics of ecosystems: a comparison of artificial neural network and regression models. *Ecological Modelling* 98: 173-186.
- Patterson DW (1996). *Artificial neural networks: theory and applications*. Simon & Schuster, Singapore.
- Rumelart DE, Hinton GE, Williams RJ (1986). Learning Internal Representations by Error Propagation. In: *Parallel Distributed Processing: exploration in the microstructure of cognition*. Vol. 1, Foundations, MIT Press, Cambridge.
- Schultz A, Wieland R, Lutze G (2000). Neural networks in agroecological modelling - stylish application or helpful tool. *Computers and Electronics in Agriculture* 29: 73-97.
- Smith M (1996). *Neural networks for statistical modeling*. ITC Press, Boston.
- Tang WY, Kassim AHM, Abubakar SH (1996). Comparative studies of various missing data treatment methods – Malaysian experience. *Atmospheric Research* 42: 247-262.
- Vickers D, Mahrt L (1997). Quality control and fluxes sampling problems for tower and aircraft data. *Journal of Oceanic Technology* 14: 512-526.
- White H (1992). *Artificial neural networks: approximation and learning*. Blackwell, Cambridge.
- Xia Y, Fabian P, Stohl A, Winterhalter M (1999). Forest climatology: reconstruction of mean climatological data for Bavaria, Germany. *Agricultural And Forest Meteorology* 96: 117-129.