



UNIVERSITÀ DEGLI STUDI DELLA TUSCIA - VITERBO

DIPARTIMENTO PER LA INNOVAZIONE NEI SISTEMI
BIOLOGICI, AGROALIMENTARI E FORESTALI
- DIBAF -

CORSO DI DOTTORATO DI RICERCA in
ECOLOGIA FORESTALE - XXIII ciclo

**Impatto dello schema di post
processamento dei dati eddy covariance
nella stima del bilancio del carbonio degli
ecosistemi terrestri**

Settore scientifico-disciplinare: BIO/07

Coordinatore: Prof. Paolo De Angelis

Firma:

Tutor: Ph.D. Dario Papale

Firma:

Dottorando: Dott. Carlo Trotta

Firma:

Indice

1	Introduzione	5
1.1	Incertezza	8
1.2	La teoria Eddy Covariance (EC)	9
1.3	Da alta frequenza a valori semiorari	12
1.4	Post Processamento	12
1.4.1	Spike	14
1.4.2	Storage	14
1.4.3	Soglia di u^* e criteri di applicazione	15
1.4.4	Gap filling	17
1.4.5	Partitioning	18
1.5	Obiettivi	20
2	Materiali e Metodi	21
2.1	Descrizione dei siti	21
2.2	Schemi di Post Processamento	21
2.3	Spike detection	23
2.4	Soglie di u^* e criteri di applicazione	25
2.4.1	metodo Reichstein	26
2.4.2	metodo Barr	26
2.4.3	Incertezza della soglia di u^*	30
2.5	Applicazione delle soglie di u^*	31
2.6	Gap Filling	31
2.6.1	Marginal Distribution Sampling	31
2.6.2	Reti Neurali Artificiali	32
2.7	Metodi di partitioning	34

2.7.1	Partitioning Reichstein	35
2.7.2	Partitioning Lasslop	37
2.7.3	Partitioning Van Gorsel	40
2.8	Approccio Statistico	42
2.8.1	Analisi Multifattoriale	43
2.8.2	ANalysis-Of-VARiance (ANOVA)	44
3	Risultati	51
3.1	Modello Additivo o ANOVA?	51
3.1.1	Verifica del modello	53
3.1.2	Verifica della Normalità	54
3.1.3	Verifica di Indipendenza	55
3.1.4	Modello additivo finale	56
3.2	Quantificazione dell'incertezza	57
3.3	Incertezza intrinseca	58
3.3.1	Incertezza intrinseca - uTH	58
3.3.2	Incertezza intrinseca - GF	63
3.4	Incertezza dei dati EC e Relazioni Funzionali	66
3.5	Quantificazione dell'incertezza Annuale	69
4	Conclusioni	85
	Bibliografia	103

Capitolo 1

Introduzione

Negli ultimi anni l'interesse per i cambiamenti climatici è aumentato notevolmente sia per un maggior coinvolgimento dell'opinione pubblica ai temi ambientali sia per la necessità, da parte della comunità scientifica, di una maggior comprensione dei meccanismi che regolano ed influenzano il clima. Al discorso sui cambiamenti climatici sempre più spesso è associato il fenomeno dell'aumento della concentrazione di anidride carbonica (CO_2) nell'atmosfera causato dall'attività antropica. Per poter stimare l'impatto che l'attività dell'uomo ha sull'aumento della CO_2 in atmosfera è necessario quantificare il ruolo delle altre componenti del ciclo del carbonio e calcolare il cosiddetto *Carbon Balance* (CB). Una delle componenti più importanti del CB è legata al bilancio tra fotosintesi e respirazione degli ecosistemi terrestri i quali sono in grado di assorbire una parte significativa di CO_2 immessa nell'atmosfera. L'accumulo o il rilascio di carbonio è strettamente associato alle diverse tipologie vegetazionali, ai loro mutamenti provocati dalle attività umane, come i cambi di uso del suolo, ed ai cambiamenti climatici stessi. Una tecnica utilizzata per misurare lo scambio netto di CO_2 tra la biosfera e l'atmosfera è quella della correlazione turbolenta (*Eddy Covariance*, EC). Per valutare l'importanza che il CB e la tecnica EC ricoprono nella comunità scientifica è possibile utilizzare il numero di articoli scientifici pubblicati. Prendendo, ad esempio, il numero di articoli scientifici che hanno, tra le parole chiave, il termine *carbon balance*, si è passati da un totale di 4 articoli tra il 1975 e 1993 a circa 42 nel solo 2011 (emeroteca virtuale periodici caspur).

Dall'analisi della cumulata del numero di articoli annui che riguardano il CB (figura 1.1) è evidente che, fino al 1995, la loro frequenza era molto ridotta mentre, da questa data in poi, la pendenza della curva mostra un sensibile aumento. L'anno 1995 coincide anche con le prime pubblicazioni che riportano, tra le parole chiave, il termine Eddy Covariance. Questa prima analogia sembra indicare che, l'applicazione della tecnica EC in ambito ecologico, abbia incrementato notevolmente la capacità di comprendere i fenomeni che influenzano il CB.

Un altro aspetto di sicuro interesse è quello dell'andamento delle due cumulate. Come è facile notare, dopo il 1995, le cumulate sembrano procedere quasi parallelamente, con la curva EC al di sotto di quella CB. Dal 2006 si ha un'inversione di posizione con la curva EC che sovrasta la cumulata del CB (figura 1.1).

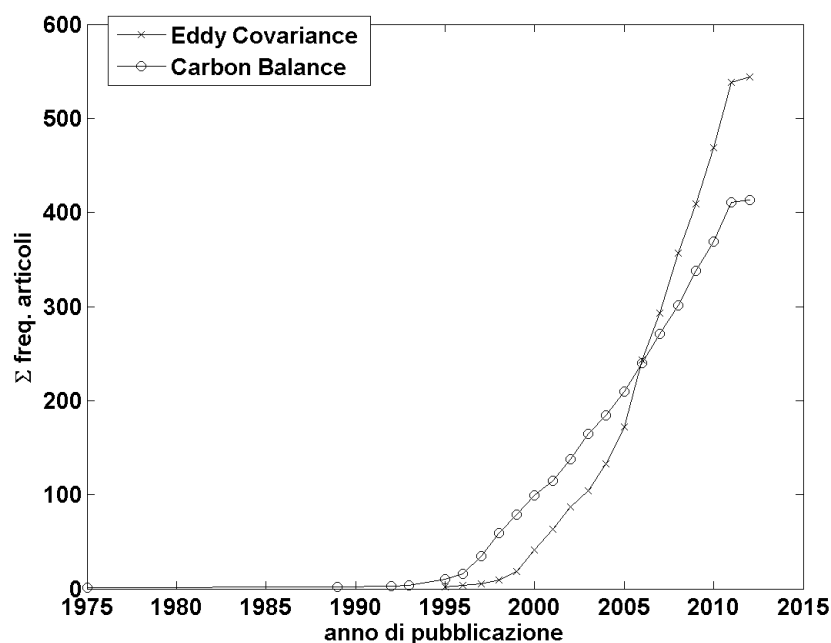


Figura 1.1: Cumulate delle frequenze assolute annue degli articoli che hanno tra le parole chiave Carbon Balance o Eddy Covariance.

Analizzando il numero di pubblicazioni annue che trattano del CB e di EC (figura 1.2) la divergenza è ancora più evidente infatti, ad eccezione del 2003,

negli anni successivi al 2000 il numero di articoli EC è nettamente superiore a quelli sul CB.

Una possibile spiegazione del notevole numero di articoli sulla tecnica EC

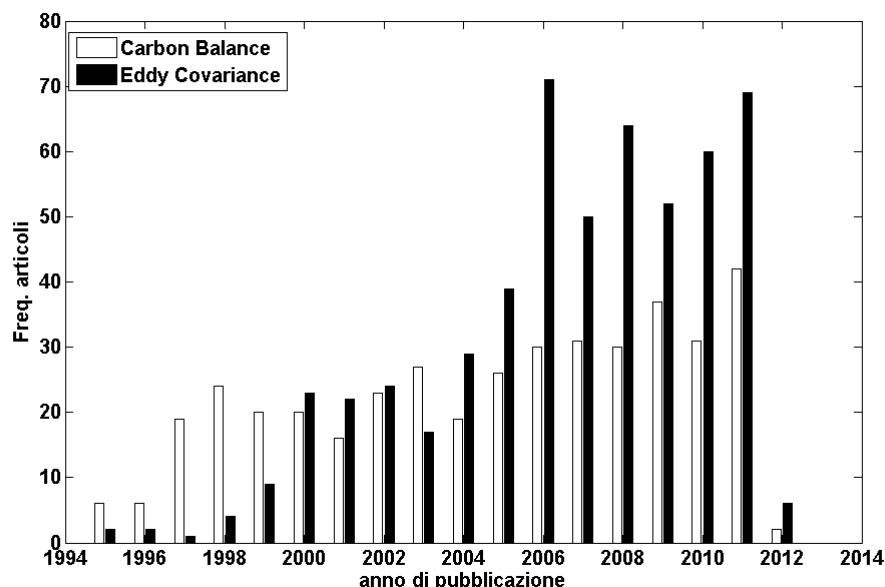


Figura 1.2: Frequenze assolute annue degli articoli che hanno tra le parole chiave Carbon Balance o Eddy Covariance.

può essere ricondotto, oltre alla valenza ecologica di queste misure, anche all’aspetto teorico di base. L’utilizzo della tecnica EC per scale temporali molto lunghe e in condizioni non ideali, secondo i classici criteri micrometeorologici, ha reso necessario un incremento degli studi teorici e delle applicazioni pratiche legate a questa tecnologia.

I dati misurati con la tecnica EC giocano un ruolo fondamentale in molte applicazioni. Ad esempio Luyssaert *et al.* (2008), utilizzando i dati EC, dimostrano che le foreste vetuste, contrariamente a quanto si pensava in precedenza, continuano ad assorbire CO₂. Molti autori utilizzando i dati di flusso per la calibrazione e validazione di modelli (Baldocchi, 1997; Law *et al.*, 2001b; Hanan *et al.*, 2002; Reichstein *et al.*, 2002b, 2003b; Hanson *et al.*, 2004). Negli ultimi anni i dati EC, congiuntamente ad altri dati ecologici, sono utilizzati per la cosiddetta “data-model fusion”. Ad esempio Richard-

son *et al.* (2010), utilizzando i dati EC, respirazione del suolo misurata con camerette ed altre misure ancillari ecologiche (LAI, quantità di lettiera e incremento di biomassa legnoso), stimano i parametri ed il valore iniziale del carbonio di un semplice modello (DALEC) riuscendo a ridurre l'incertezza dei risultati rispetto ai valori misurati e migliorando la capacità di previsione.

Sia nel caso dei modelli che nel confronto tra le misure di diversi siti, stimare l'incertezza è necessario infatti consente di conoscere i limiti di confidenza (con un certo livello di significatività statistica) di ciascuna osservazione (Aubinet *et al.*, 2012).

Oltre ai dati di flusso della CO₂, la tecnica EC è utilizzata anche per la stima dei flussi di vapore acqueo, calore latente e calore sensibile. Ad esempio Jung *et al.* (2010), utilizzando dati di evapotraspirazione derivati da misure EC, remote sensing e osservazioni meteorologiche, suggeriscono che la riduzione dell'evapotraspirazione globale sia principalmente dovuta ad una riduzione dell'umidità del suolo.

1.1 Incertezza

Una dei principali sbagli commessi, quando si parla di incertezza, è quello di confonderla con il termine errore. Seguendo quanto riportato da Aubinet *et al.* (2012), l'errore è un singolo valore che indica la differenza tra una misura e l'effettiva quantità che dovrebbe essere misurata mentre l'incertezza è un range di valori che definiscono i limiti entro i quali una quantità misurata dovrebbe cadere. Da queste definizioni è facile capire che l'errore può essere individuato e corretto mentre l'incertezza quantifica la precisione di una misura e, essendo un range e non un singolo valore, non può essere corretta. Le origini dell'incertezza possono essere molto diverse ad esempio errori da parte di un operatore, problemi di campionamento di una popolazione, errori strumentali, problemi di calibrazione ed infine condizioni di misura non sempre comparabili con quelle richieste dalla teoria (Aubinet *et al.*, 2012). Per capire l'importanza dell'incertezza è possibile considerare, ad esempio, gli esperimenti OptIC (Trudinger *et al.*, 2007) e REFLEX (Fox *et al.*, 2009) i quali hanno dimostrato che la distribuzione dei parametri di alcuni modelli

era influenzata direttamente dalla conoscenza dell'incertezza dei dati (Aubinet *et al.*, 2012).

Per definire l'importanza che ha l'incertezza è possibile anche citare quanto suggerito da Raupach *et al.* (2005): "l'incertezza dei dati è importante tanto quanto i valori dei dati stessi".

1.2 La teoria Eddy Covariance (EC)

Come precedentemente riportato tra le componenti del ciclo del carbonio un ruolo centrale è sicuramente quello degli ecosistemi terrestri infatti, uno dei principali obiettivi delle ricerche nelle bioscienze, è il bilancio netto del carbonio dell'ecosistema (Running *et al.*, 1999; Geider *et al.*, 2001). Tali studi richiedono la stima dei flussi di anidride carbonica a scale temporali orarie, giornaliere, stagionali ed annuali e a scale spaziali che vanno dalle foglie a singole piante a interi ecosistemi (Baldocchi, 2003).

In atmosfera i movimenti turbolenti ascendenti e discendenti d'aria trasportano i gas come l'anidride carbonica. La tecnica EC campiona questi movimenti turbolenti per determinare la differenza netta di materia nell'interfaccia canopy-atmosfera (Baldocchi, 2003).

La popolarità della tecnica EC è ascrivibile a 4 fattori: (i) consente di stimare lo scambio netto di carbonio degli ecosistemi (*Net Ecosystem carbon Exchange*, NEE) a livello di ecosistema in modo non invasivo; (ii) la NEE è misurata direttamente come scambio di CO₂ nell'interfaccia biosfera-atmosfera; (iii) l'area sorgente dei flussi misurati, detta *footprint*, è definita spazialmente e può variare da qualche centinaio di metri a diversi chilometri (Schmid, 1994); (iv) le misure EC hanno uno spettro temporale da orario ad annuale (Wofsy *et al.*, 1993; Baldocchi *et al.*, 2001) (Baldocchi, 2003).

Come ogni tecnica anche quella della correlazione turbolenta ha delle limitazioni infatti necessita di: (i) superfici pianeggianti; (ii) stabilità atmosferica; (iii) l'area della vegetazione all'interno della footprint non deve presentare discontinuità (Baldocchi, 2003).

La violazione di queste condizioni può causare un errore sistematico nell'interpretazione delle misure EC (Baldocchi *et al.*, 1988; Foken & Wichura, 1995; Massman & Lee, 2002) il quale si amplifica quando le misure vengono

integrate a scale temporali maggiori (ad esempi somme giornaliere e annuali) (Moncrieff *et al.*, 1996).

L'equazione di conservazione della massa è la base teorica per la tecnica EC (Baldocchi *et al.*, 1988; Paw *et al.*, 2000; Massman & Lee, 2002).

Prendendo in esame un volume di controllo la NEE è definita come il flusso di carbonio netto attraverso il piano superiore ed i piani laterali di questo volume ed al di sotto di una altezza di riferimento h_r . Assumendo l'incompressibilità e trascurando la divergenza orizzontale del flusso turbolento, il tasso di scambio netto suolo-foresta e atmosfera è dato da (Finnigan, 1999; Feigenwinter *et al.*, 2004):

$$\begin{aligned}
 NEE &= \overline{w'c'}(h_r) + \int_0^{h_r} \frac{\partial \overline{c(z)}}{\partial t} dz \\
 &+ \overset{I}{\int_0^{h_r}} \left(\overline{u(z)} \frac{\partial \overline{c(z)}}{\partial x} + \overline{v(z)} \frac{\partial \overline{c(z)}}{\partial y} \right) dz \quad \overset{II}{\quad} \\
 &+ \overset{III}{\int_0^{h_r}} \left(\overline{w(z)} \frac{\partial \overline{c(z)}}{\partial z} \right) dz \quad \overset{IV}{\quad} \quad (1.1)
 \end{aligned}$$

in cui x , y , z sono le coordinate cartesiane; u , v , w sono le componenti longitudinali, laterali e verticali del vento e c è la concentrazione molare di CO₂. La barra superiore indica la media temporale mentre gli apici gli scarti rispetto alla media. Solo il flusso turbolento verticale (termine I dell'equazione 1.1) sarà significativo in condizioni di forte vento e di sito pianeggiante e con copertura omogenea. Il termine II, il tasso di scambio dello storage della CO₂ negli strati d'aria al di sotto di h_r , è significativo durante condizioni di transizione tra instabilità e stabilità, in genere al crepuscolo, e di stabilità/instabilità, all'alba. Questo termine è insignificante per lunghi periodi di mediazione infatti, la CO₂ accumulata negli strati d'aria durante la notte, è dissipata con l'inizio della turbolenza del mattino seguente. I termini di avvezione (III e IV) rappresentano il trasporto orizzontale e verticale della CO₂ attraverso le facce del volume di controllo. Errori legati all'avvezione saranno molto alti in condizioni di vento debole e stratificazione stabile dell'atmosfera. Secondo molti autori (Aubinet *et al.*, 2005; Feigenwinter *et al.*, 2004; Staebler & Fitzjarrald, 2004) la somma del flusso turbolento e il cam-

biamiento dello storage sovrastima la NEE notturna a causa l'avvezione della CO₂ accumulata nello strato d'aria freddo al di sotto della strumentazione di misura.

In condizioni ideali di misura alcuni dei termini della 1.1 possono essere semplificati. Una volta semplificata, l'equazione di conservazione della massa è vista come il bilancio tra la divergenza del flusso verticale di CO₂ ($\partial F_z / \partial z$, con z quota degli strumenti di misura richiesti dalla tecnica EC) e le fonti biologiche di emissione/assorbimento S_B :

$$\frac{\partial F_z}{\partial z} = -S_B(z). \quad (1.2)$$

integrando l'equazione 1.2 rispetto all'altezza h , si ottiene un'uguaglianza tra il flusso verticale medio di densità misurato ad una certa altezza sopra la canopy, $F_z(h)$, e il flusso netto di densità di massa che viene emesso/assorbito dal suolo e vegetazione sottostante, $F_z(0)$:

$$F_z(h) = F_z(0) - \int_0^h S_B(z) dz. \quad (1.3)$$

La tecnica EC misura il termine $F_z(h)$ (Baldocchi, 2003).

Nella pratica la misura dei flussi turbolenti è accoppiata all'analisi statistica della densità di flusso verticale di massa istantanea ($F = \omega \rho_c$, $\mu\text{mol m}^{-2} \text{s}^{-1}$) utilizzando la media di Reynolds (Reynolds, 1895).

Il risultato di questa operazione è una relazione che esprime la densità di flusso di massa di CO₂ mediata in un intervallo di tempo (ad esempio mezz'ora) come la covarianza tra le fluttuazioni della velocità verticale del vento (ω) e la concentrazione della CO₂ (*mixing ratio*, $c = \frac{\rho_c}{\rho_a}$ dove ρ_c è la densità della CO₂ e ρ_a è la densità dell'aria):

$$F = \bar{\rho}_a \cdot \overline{\omega'c'}. \quad (1.4)$$

nell'equazione 1.4 la barra superiore indica la media temporale, gli apici le fluttuazioni rispetto alla media (ad esempio $c' = c - \bar{c}$). Se la covarianza è positiva il flusso è diretto verso l'atmosfera, se negativa il flusso è verso la superficie della terra (Baldocchi, 2003).

1.3 Da alta frequenza a valori semiorari

Per campionare tutti i vortici che contribuiscono al trasporto turbolento della CO₂ è necessario misurare frequentemente e per un sufficiente intervallo di tempo l'atmosfera. Di solito la frequenza di campionamento è dell'ordine di 10 Hz (10 misure al secondo) (Anderson *et al.*, 1984, 1986a; Goulden *et al.*, 1996a) per un intervallo di tempo che dovrebbe essere in funzione del periodo del giorno in cui si effettuano le misure (Lenschow, 1995) ma, in genere, è considerato costante e pari a 30 minuti (Anderson *et al.*, 1984, 1986a; Aubinet *et al.*, 2000; Massman & Lee, 2002) (Baldocchi, 2003).

Gli strumenti necessari per l'applicazione della tecnica EC sono l'anemometro ultrasonico e l'analizzatore chimico a risposta veloce. L'anemometro rileva, contemporaneamente e ad alta frequenza, le tre componenti della velocità del vento e la temperatura sonica dell'aria mentre l'analizzatore misura la concentrazione di una o più specie chimiche di interesse fisiologico (in genere le specie chimiche misurate sono la CO₂ e il vapore acqueo). I valori misurati da questi due strumenti vengono registrati su supporti fissi informatici e definiti *raw data*.

Per ottenere il dato finale di flusso F i raw data sono sottoposti ad una sequenza di elaborazioni. La figura 1.3 rappresenta un esempio di schema di processamento dei raw data.

1.4 Post Processamento

Una volta che i dati ad alta frequenza sono stati processati vengono aggregati come media semioraria. Anche per i dati semiorari sono necessarie una serie di correzioni e applicazione di filtri per escludere dalle elaborazioni valori non realistici o misure che non rispettano i presupposti teorici della tecnica EC.

Per escludere dati non buoni vengono applicati dei quality check (Papale *et al.*, 2006), la correzione per lo storage, la spike detection ed il filtraggio per la velocità di frizione (u^*) (Reichstein *et al.*, 2005). L'eliminazione dei dati misurati non buoni dalla serie temporale dei dati EC crea buchi (gap)

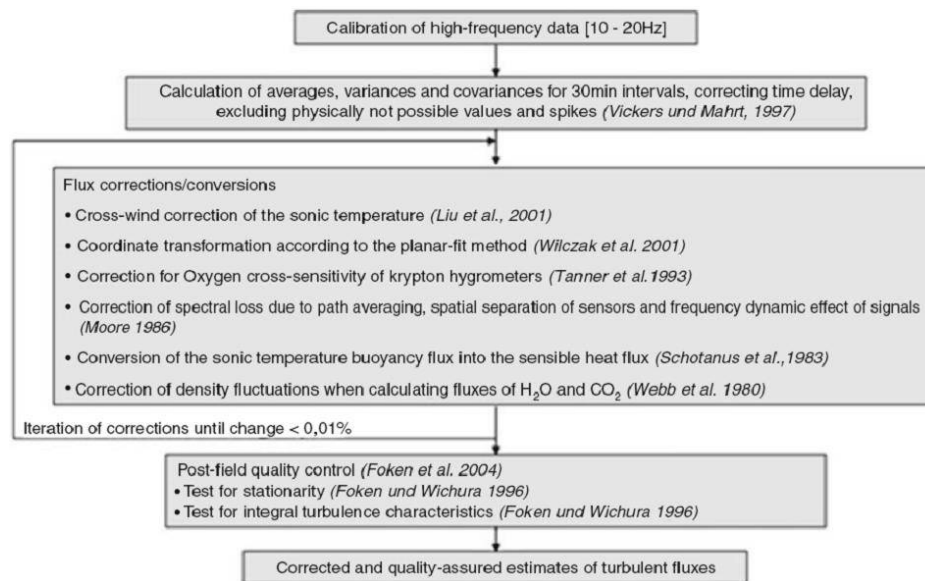


Figura 1.3: Esempio di schema per il processamento dei raw data (Mauder & Foken, 2006).

di lunghezza variabile (da mezz'ora a diversi giorni). In genere la copertura annuale dei dati EC varia dal 65 al 75% (Falge *et al.*, 2001a). Oltre al filtraggio alcuni periodi possono mancare per problemi strumentali.

Per alcune applicazioni dei dati EC, come ad esempio per la validazione di modelli ecologici o per il carbon balance, è necessario avere una serie temporale completa. Se i gap fossero distribuiti in maniera casuale potrebbero essere riempiti con la media ma, in genere, i dati mancanti comprendono soprattutto le notti ed i periodi invernali. Per avere la serie temporale completa è necessario adottare delle tecniche di riempimento (gap-filling).

Un'altra importante applicazione dei dati di NEE misurati con la tecnica EC è quella di poter estrapolare, con opportuni algoritmi di partizione ("flux partitioning algorithm"), la produttività primaria lorda (gross primary productivity, GPP) e la respirazione dell'ecosistema (Reco). Questa estrapolazione è utile per avere una maggior comprensione dei processi di assimilazione e rilascio di CO₂ che governano lo scambio netto dell'ecosistema (NEE) (Reichstein *et al.*, 2005; Lasslop *et al.*, 2010; van Gorsel *et al.*, 2007; Desai *et al.*, 2008).

Nel mondo le stazioni (torri) EC sono diverse centinaia, misurano i flussi

in diverse zone ecologiche, diverse tipologie di ecosistemi terrestri (colture, pascoli, foreste) e sono organizzate in reti continentali (ad esempio AMERIFLUX, CARBOEUROFLUX, ASIAFLUX) (Baldocchi, 2003).

In queste reti vengono applicati diversi metodi, tecniche e parametrizzazioni per filtrare i dati misurati (Papale *et al.*, 2006; Reichstein *et al.*, 2005), riempire i gap (Moffat *et al.*, 2007) e separare la NEE in GPP ed Reco (Reichstein *et al.*, 2005; Lasslop *et al.*, 2010; van Gorsel *et al.*, 2007; Desai *et al.*, 2008).

Poiché non è sempre possibile decidere a priori quale sia la corretta combinazione di opzione di elaborazione, le differenze nei risultati possono essere considerate incertezza e quindi tenute in considerazione nell'interpretazione dei dati misurati. Nelle sezioni seguenti saranno analizzati nel dettaglio le basi teoriche dei passaggi del post processamento.

1.4.1 Spike

Le misure EC sono spesso affette da spike dovuti sia a ragioni biofisiche (cambiamenti nella footprint o rapidi cambiamenti delle condizioni di turbolenza) che strumentali (ad esempio accumulo di acqua nell'anemometro sonico o nell'analizzatore di gas open-path). Gli spike influenzano una singola misura a alta frequenza e, in genere, vengono rimossi prima del calcolo del valore di flusso semiorario. Anche dopo il filtraggio ad alta frequenza la serie temporale semioraria può mostrare dei valori anomali quindi, anche per questa scala temporale, è necessario applicare una tecnica di eliminazione degli spike. La rimozione degli spike, di solito, non influenza direttamente la stima annuale della NEE ma può influenzare la qualità del processo di gap filling (Papale *et al.*, 2006).

1.4.2 Storage

Durante periodi di bassa turbolenza, soprattutto durante la notte, i flussi misurati dal sistema EC non rappresentano quelli reali infatti, in tali condizioni, i termini storage ed avvezione dell'equazione di conservazione della massa possono non essere considerati trascurabili (Lee, 1998; Aubinet *et al.*, 2003, 2005; Feigenwinter *et al.*, 2004; Marcolla *et al.*, 2005).

Nel caso in cui lo storage non è trascurabile la CO₂ respirata dall'ecosistema non viene misurata dal sistema EC ma si accumula al di sotto del sistema stesso. Con la ripresa del regime turbolento il sistema EC rileva un repentino aumento del flusso causato dal trasporto della CO₂ accumulata come storage.

Lo storage può essere calcolato utilizzando le concentrazioni di CO₂ (c) misurate con un profilo ed applicando la formula:

$$S_c = \frac{P_a}{\Re T_a} \int_0^{h_m} \frac{dc(z)}{dt} dz \quad (1.5)$$

dove P_a è la pressione atmosferica; $\Re$ è la costante molare dei gas; T_a è la temperatura dell'aria; h_m sono le diverse altezze del profilo di misura (in genere m varia tra 4 e 7) (Aubinet *et al.*, 2001). Nella pratica le derivate sono approssimate dalle differenze tra due punti successivi di misura e l'integrale è approssimato con la somma pesata delle variabili misurate lungo il profilo (Aubinet *et al.*, 2001). Per i siti in cui il profilo non è disponibile l'unico modo per stimare lo storage è usando un approccio discreto in cui la concentrazione di CO₂ misurata al top della torre è considerata costante all'interno della canopy (Papale *et al.*, 2006).

Lo storage potrebbe non avere effetto nel calcolo dei flussi sul lungo periodo ma potrebbe indurre un certo ritardo sui flussi semiorari e, di conseguenza, sulle relazioni tra i flussi e le altre variabili climatiche (Aubinet *et al.*, 2012). Nel caso in cui la CO₂ respirata dall'ecosistema è rimossa dall'avvezione, quindi impossibile da rilevare dal sistema EC, è necessario un trattamento dei dati molto diverso (Aubinet *et al.*, 2012) e sarà spiegato nella sezione riguardante la soglia di u^* .

Nella realtà i due casi si presentano, spesso, simultaneamente quindi è necessaria sia un'operazione di correzione (storage) che una di filtraggio (secondo la soglia di u^*).

1.4.3 Soglia di u^* e criteri di applicazione

Da quanto già riportato nelle sezioni precedenti, la tecnica EC spesso sottoestima i flussi notturni di CO₂ in condizioni di bassa turbolenza. L'influenza

della sottostima è notevole quindi le cause e le motivazioni teoriche di questo fenomeno sono molto studiate (Goulden *et al.*, 1996b; Jarvis *et al.*, 1997; Black *et al.*, 2000; Valentini *et al.*, 2000; Lee, 1998; Finnigan, 1999; Baldocchi *et al.*, 2000; Massman & Lee, 2002).

Per l'applicazione pratica, molti autori filtrano le misure di flusso notturne in base alle condizioni di turbolenza atmosferica usando, come indicatore, la velocità di frizione (u^*). L'ipotesi alla base di questa procedura è che la NEE notturna è determinata biologicamente (dalla respirazione delle piante e del suolo) e quindi dovrebbe essere indipendente dai regimi di turbolenza (Wofsy *et al.*, 1993; Goulden *et al.*, 1996b; Massman & Lee, 2002). L'applicazione di questa tipologia di filtraggio implica la determinazione della cosiddetta soglia di u^* (u^* threshold) al di sotto della quale si considera che la bassa turbolenza influenzi le misure di flusso del sistema EC. In genere i flussi di CO₂ sono trascurabili quando la u^* tende a zero (Massman & Lee, 2002). In questo modo il filtro di u^* tende a aumentare la stima annuale della respirazione dell'ecosistema e, di conseguenza, a diminuire la stima dell'assorbimento della CO₂ da parte dell'ecosistema (Goulden *et al.*, 1996b; Aubinet *et al.*, 2000; Barr *et al.*, 2002; Saleska *et al.*, 2003).

Barford *et al.* (2001) mostrano che all'aumentare della soglia di u^* l'assorbimento del carbonio annuale è diminuisce.

Dal punto di vista teorico si assume che la soglia sia il punto in cui i flussi non aumentano all'aumentare del valore della velocità di frizione. In letteratura i primi valori di soglia di u^* erano definiti visivamente osservando lo scatter plot dei dati di flusso notturni rispetto ai valori di u^* . Spesso i flussi notturni, rispetto alla u^* , sono abbastanza rumorosi in particolare quando il range dei valori di u^* è limitato. Attualmente l'approccio visivo, sia per la soggettività dei risultati che per la difficoltà pratica, è stato sostituito da metodi automatici standardizzati (Gu *et al.*, 2005; Reichstein *et al.*, 2005). Tutti i metodi automatici proposti hanno l'obiettivo di definire la soglia u^* oltre la quale i flussi di NEE notturni sono indipendenti dal valore della velocità di frizione. La figura 1.4 rappresenta la condizione teorica ottimale per la stima della soglia di u^* . Nel grafico i singoli punti neri rappresentano i flussi notturni misurati in un ristretto intervallo di temperatura mentre i punti grigi indicano l'andamento teorico. Per valori di u^* inferiori a circa 0.2

m s^{-1} i flussi misurati sono in funzione della velocità di frizione mentre oltre questo valore le due variabili sono indipendenti. Secondo l'ipotesi espressa precedentemente i flussi notturni, per un intervallo di temperatura piccolo, dovrebbero essere indipendente dai regimi di turbolenza quindi, i valori di flusso dipendenti da u^* , devono essere eliminati. Per l'applicazione pratica la soglia è definita come il valore minimo di u^* tale che la differenza tra il corrispondente valore medio di NEE non è significativamente differente dalla media dei flussi delle classi successive (Gu *et al.*, 2005; Reichstein *et al.*, 2005). Il valore della soglia di u^* , essendo sito specifico, varia, in genere tra 0.1 e 0.5 m s^{-1} (Aubinet *et al.*, 2012).

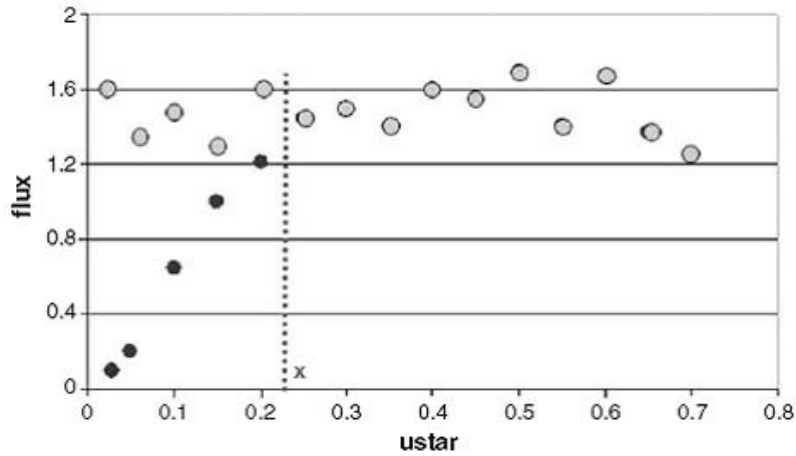


Figura 1.4: condizione teorica ed ottimale per la scelta della soglia di u^* . I punti neri sono i flussi notturni normalizzati (cioè flussi misurati in un ristretto range di temperatura); i punti grigi rappresentano, invece, l'andamento teorico dei flussi (indipendenti da u^*). Il valore della soglia di u^* è indicato dalla X (Aubinet *et al.*, 2012).

1.4.4 Gap filling

A causa di problemi legati alla turbolenza, malfunzionamenti o rotture di strumenti e l'applicazione dei filtri (despike ed u^*) la serie temporale di dati EC è discontinua. Questi fenomeni possono creare sia singoli valori mancanti ma, in genere, veri e propri periodi privi di misure. In genere i gap lunghi sono più difficili da riempire rispetto a quelli brevi (Falge *et al.*, 2001a; Ri-

Richardson *et al.*, 2006; Moffat *et al.*, 2007) infatti le proprietà dell'ecosistema possono variare nel tempo, creando così una serie storica non-stazionaria (Richardson & Hollinger, 2007).

La serie temporale discontinua può contenere informazioni sufficienti per creare dei modelli di riempimento di dati mancanti (*gap filling*) e quindi consentire, con serie temporali continue, di fare somme da scala giornaliera ad annuale (Moffat *et al.*, 2007).

Attualmente esistono numerose tecniche di gap filling (Barr *et al.*, 2004; Branswell *et al.*, 2005; Desai *et al.*, 2005; Falge *et al.*, 2001a; Gove & Hollinger, 2006; Hollinger *et al.*, 2004; Hui *et al.*, 2004; Knorr & Kattge, 2005; Noormets *et al.*, 2007; Ooba *et al.*, 2006; Papale & Valentini, 2003; Reichstein *et al.*, 2005; Richardson *et al.*, 2006; Schwalm *et al.*, 2007; Stauch & Jarvis, 2006) che si basano su un'ampia gamma di approcci come le interpolazioni, riempimenti su basi probabilistiche, look-up tables, regressioni non lineari, reti neurali artificiali, modelli.

Moffat *et al.* (2007) confrontano le performance di 15 diverse tecniche di gap filling applicate a 10 siti/anno Europei. Per calcolare le performance delle diverse tecniche di gap filling gli autori hanno creato una serie di 50 scenari di gap artificiali che hanno sovrapposto ai dati misurati. L'errore della tecnica di riempimento era quantificato confrontando statisticamente i dati misurati nel gap artificiale e quelli simulati. Moffat *et al.* (2007), dopo l'analisi dei risultati, raccomandano l'utilizzo delle tecniche proposte da Papale & Valentini (2003) e Reichstein *et al.* (2005) per il gap filling dei dati di flusso mancanti.

1.4.5 Partitioning

Dal punto di vista ecologico i flussi di NEE sono il risultato della seguente equazione:

$$NEE = Reco - GPP \quad (1.6)$$

dove GPP rappresenta la produzione primaria netta (gross primary production) ed Reco la respirazione totale dell'ecosistema (total ecosystem respiration). I due termini a destra dell'equazione racchiudono tutti i processi della componente terrestre del ciclo globale del carbonio. Attualmente non

è possibile avere una misura diretta di GPP ed Reco poichè questi processi rappresentano la combinazione di una moltitudine di risposte da parte degli organismi autotrofi ed eterotrofi (Desai *et al.*, 2008). Un metodo per stimare la GPP ed Reco potrebbe essere quello di riportare misure fatte con camere a livello di canopy ma, oltre ad richiede un intenso e pesante lavoro, comporta una notevole incertezza. Una valida alternativa a questo approccio è quella di utilizzare le misure dirette dei flussi di NEE applicando la tecnica EC.

I metodi per dividere (*“partition”*) la NEE nelle componenti GPP e Reco servono per stimare i percorsi del carbonio nell’ecosistema. Con l’adozione di procedure standardizzate di partitioning è possibile minimizzare le divergenze dovute al processamento ed enfatizzare quelle sito-specifiche (Foken & Wichura, 1995; Aubinet *et al.*, 2000; Baldocchi, 2003, 2008; Rebmann *et al.*, 2005; Reichstein *et al.*, 2005; Papale *et al.*, 2006; Göckede *et al.*, 2008). Alcuni metodi di partitioning proposti utilizzano i dati di flusso notturni. Questo approccio assume che la GPP è pari a zero durante le ore di buio (radiazione globale $< 20 \text{ W m}^{-2}$) quindi la NEE misurata è costituita interamente dalla Reco con la quale è possibile parametrizzare un modello ed estrapolare per i periodi di luce. Per descrivere la respirazione in funzione della temperatura si utilizza un modello esponenziale (Lloyd & Taylor, 1994). Utilizzando i dati notturni è possibile commettere errori per la frequente riduzione della turbolenza e la prevalenza dei flussi avvettivi non misurabili dai sistemi EC convenzionali (Goulden *et al.*, 1996a; Aubinet *et al.*, 2000; Feigenwinter *et al.*, 2004; Aubinet, 2008).

Per i motivi appena citati un altro approccio comunemente adottato è quello di estrapolare la respirazione dalle curve di luce, in questo caso viene considerato che la NEE varia sia in funzione della temperatura (soprattutto influenzando la Reco) sia in base al deficit di pressione di vapore (VPD), (influenza principalmente la GPP agendo sull’apertura stomatica), sia su altri fattori (Gilmanov *et al.*, 2007; Lasslop *et al.*, 2010).

1.5 Obiettivi

Da quanto riportato finora risulta evidente che, per ogni passaggio di post processamento dei dati EC, sono stati proposti e sviluppati una notevole quantità di approcci. Dal punto di vista teorico ognuno di questi ha delle solide basi quindi la scelta di un metodo o di un'altro è semplicemente a discrezione dell'utilizzatore. Come riportato nell'introduzione non essendo sempre possibile decidere a priori quale sia la corretta combinazione di opzione di elaborazione, le differenze nei risultati sono considerate incertezza ed interpretate come tali.

Gli scopi della presente tesi di dottorato sono: (i) applicare differenti schemi di post processamento dei dati EC; (ii) definire il peso che ogni fattore e loro combinazioni ha sul valore complessivo dell'incertezza; (iii) quantificare l'impatto che i diversi schemi di post processing hanno sul bilancio del carbonio annuale.

Dopo aver definito quali fattori del post processing causano la maggior divergenza verrà analizzata la loro incertezza intrinseca. Oltre allo studio strettamente legato all'analisi dei dati EC è interessante valutare anche l'impatto che l'incertezza dei dati EC può avere in ambito applicativo. Per indagare anche questo aspetto si è proceduto all'incertezza di alcune relazioni funzionali (ad es. curve di luce) indotta a sua volta dall'incertezza dei dati EC.

I dati utilizzati per la stima dell'incertezza sono stati misurati da siti EC presenti in Europa. La scelta dei siti è stata fatta per avere un'ampia rappresentazione di condizioni climatiche, ecologiche, vegetazionali ed edafiche. L'eterogeneità dei dati utilizzati per le elaborazioni rispecchia la necessità di valutare l'incertezza in un ampio range di ecosistemi terrestri i quali posso differire per numerosi aspetti.

Nella presenti tesi di dottorato l'indagine sull'incertezza è stata svolta secondo un approccio fattoriale con il quale si è quantificato non solo l'effetto che i singoli fattori hanno sul risultato finale ma soprattutto quello di stimare l'interazione tra i fattori. Dal punto di vista strettamente statistico l'eterogeneità dei dati utilizzati è uno dei requisiti fondamentali per condurre le analisi multifattoriali.

Capitolo 2

Materiali e Metodi

2.1 Descrizione dei siti

La tabella 2.1 riporta alcune delle informazioni principali dei siti utilizzati per la stima dell'incertezza dei dati EC. I siti cercano di comprendere un ampio range sia per quel che riguarda la variabilità ecologica ed di ecosistemi ma anche un ampio spettro di condizioni climatiche ed edafiche presenti nel continente europeo. Le analisi sono state svolte sia per diversi anni che in diverse tipologie di ecosistemi (Plant Functional Type, PFT): coltivi (Cropland, C); pascoli (Grassland, G); boschi misti (mixed forest, MF); foreste di latifoglie decidue (Deciduous Broadleaf Forest, DBF); foreste di aghifoglie sempreverdi (Evergreen Needleleaf Forest, ENF); savana (Dehesa) (savanna, S); foreste di latifoglie sempreverdi (Evergreen Broadleaf Forest, EBF).

Dal punto di vista spaziale i siti sono localizzati in Italia, Danimarca, Belgio, Francia, Finlandia, Spagna, Germania, Olanda.

2.2 Schemi di Post Processamento

Molte delle pubblicazioni sui dati EC spesso non riportano una stima dell'incertezza associata. La stima dell'incertezza è fondamentale per l'analisi multi-sito, per i lavori di sintesi e per una giusta interpretazione degli andamenti nel lungo periodo e della stima della variabilità interannuale (Elbers *et al.*, 2011).

Le stime dell'incertezza dei dati EC è importante anche per la loro appli-

nr	site	country	year	ID	Land Cover	PFT	reference
1	Lonzee	Belgio	2005, 2007	BE-Lon	Seminativo (rotazione)	C	
2	Vielsalm	Belgio	2000, 2006, 2007	BE-Vie	Fagus sylvatica, Pseudotsuga menziesii	MF	(Moureaux <i>et al.</i> , 2006)
3	Gebesee	Germania	2004, 2005	DE-Geb	-	C	(Anthoni <i>et al.</i> , 2004)
4	Grillenbourg	Germania	2008	DE-Gri	Prato	G	(Gilmanov <i>et al.</i> , 2007)
5	Hainich	Germania	2003	DE-Hai	Faggio	DBF	(Knohl <i>et al.</i> , 2003)
6	Klingenberg	Germania	2007	DE-Kli	Rotazione di colture	C	-
7	Rimi	Danimarca	2006, 2007	DK-Lva	Prateria	G	(Gilmanov <i>et al.</i> , 2007)
8	Sorø	Danimarca	2006, 2009	DK-Sor	Faggio	DBF	(Pilegaard <i>et al.</i> , 2003)
9	Las Majadas del Tietar (Caceres)	Spagna	2005	ES-LMa	Dehesa (Querceto rado)	S	-
10	Hyttälä	Finlandia	2007, 2008	FI-Hyy	Pinus sylvestris 41 yr	ENF	(Sumi <i>et al.</i> , 2003)
11	Fontainbleu	Francia	2007	FR-Fon	Querce	DBF	-
12	Castelporziano	Italia	2004	IT-Cpz	-	EBF	(Garbulsky <i>et al.</i> , 2008)
13	Lecceto	Italia	2006	IT-Lec	Vegetazione Mediterranea con prevalenza di latifoglie sempreverdi (Quercus ilex)	EBF	-
14	Roccarespampani 1	Italia	2005	ITRo1	Quercus cerris	DBF	(Rey <i>et al.</i> , 2002)
15	Roccarespampani 2	Italia	2006	ITRo2	Quercus cerris	DBF	(Tedeschi <i>et al.</i> , 2006)
16	Loobos	Olanda	2008, 2009	NL-Loo	Pinus sylvestris	ENF	(Dolman <i>et al.</i> , 2002)

Tabella 2.1: Caratteristiche generali dei siti

cazione nella fase di validazione dei modelli infatti finora la qualità dei dati EC è stata interpretata solo da un punto di vista qualitativo e non quantitativo.(Elbers *et al.*, 2011).

Un ulteriore fonte di incertezza è dovuta anche alla necessità di avere delle serie temporali complete. In questo caso sia i metodi di gap filling che quelli di partitioning contribuiscono ad amplificare l'incertezza complessiva. Pochi autori si sono occupati dell'incertezza globale dei dati EC. Ad esempio Rannik *et al.* (2006) hanno utilizzato i dati di NEE misurati in una pineta finlandese; Richardson & Hollinger (2007) analizzano l'effetto di lunghi gap e errori random delle misure; Oren *et al.* (2006) stimano l'incertezza derivata dal gap filling, da errori strumentali e dalla variabilità spaziale; Hagen *et al.* (2006) utilizzano due modelli per stimare l'incertezza statistica; Goulden *et al.* (1996b) studiano l'effetto combinato di errori sistematici, errori di campionamento e stime di respirazione durante calmi periodi notturni; Luyssaert *et al.* (2007) stimano l'incertezza basandosi su giudizi teorici; Elbers *et al.* (2011), assumendo che i flussi sono normalmente distribuiti e gli errori siano puramente statistici, additivi ed indipendenti, calcolano l'incertezza totale come somma dei quadrati delle singole incertezze assumendo, implicitamente, che la covarianza tra tutte le cause di incertezza sia trascurabile.

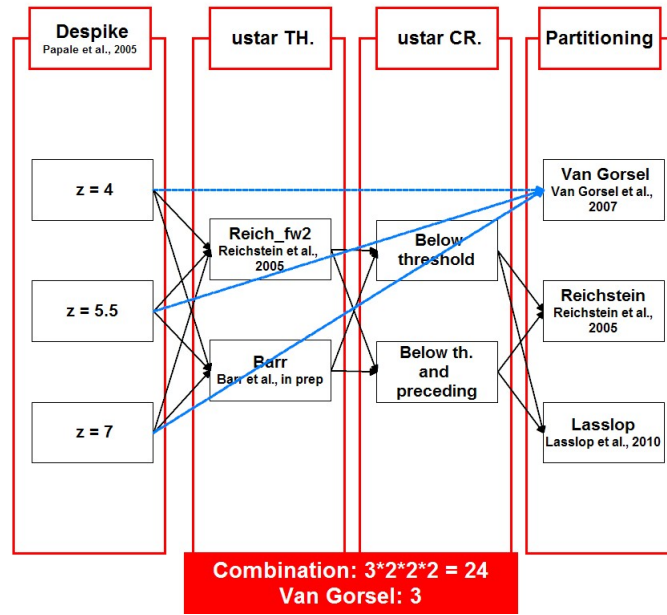
Nella presenti tesi di dottorato l'indagine sull'incertezza è stata svolta secondo un approccio fattoriale. Scopo fondamentale di questo approccio è quello di indagare e quantificare non solo l'effetto che i singoli fattori hanno sul risultato finale ma soprattutto quello di quantificare l'effetto dell'interazione tra i fattori.

Le figure 2.1(a) e 2.1(b) rappresentano, rispettivamente, gli schemi di processamento adottati per l'analisi fattoriale per i metodi di partitioning e di gap filling.

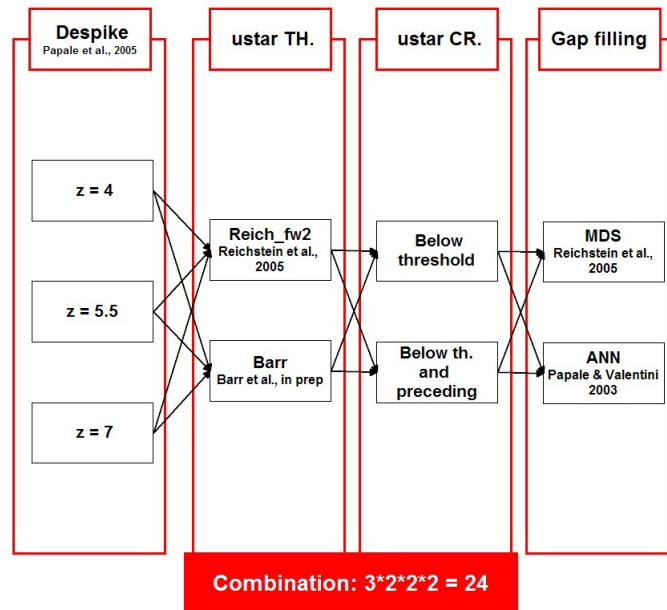
2.3 Spike detection

Come metodologia per l'eliminazione degli spike si è adottata quella proposta da Papale *et al.* (2006).

L'algoritmo utilizzato per definire gli spike è basato sulla posizione di ogni



(a)



(b)

Figura 2.1: schema post processing di Reco (a) ed NEE (b). I riquadri Despiking rappresentano il passaggio di despiking e le relative opzioni utilizzate. I riquadri ustar TH rappresentano i metodi per la definizione della soglia di u^* . I riquadri ustarCR rappresentano le due possibilità di applicazione della soglia di u^* . Il riquadro Partitioning rappresenta i metodi utilizzati per separare NEE in GPP ed Reco. Il riquadro Gap filling rappresenta i metodi utilizzati riempire i dati mancanti.

mezz'ora rispetto ai valori adiacenti ed è applicata a blocchi di 13 giorni separatamente per i dati notturni e diurni. Ai periodi notturni e diurni sono aggiunti un valore dai periodi adiacenti per poter calcolare la d_i ; i dati sono considerati notturni quando la radiazione globale era inferiore a 20 W m^{-2} , per i dati di alba e tramonto viene fatto un controllo incrociato tra ora locale e valori dell'angolo di inclinazione del sole (notte se è minore di zero). La spike detection è basata su una serie temporale di doppie differenze, usando la mediana della differenza assoluta rispetto alla mediana (MAD) come uno stimatore robusto degli spike (Sachs, 1996).

Per ciascun valore semiorari della NEE (NEE_i) viene calcolato un valore d :

$$d_i = (NEE_i - NEE_{i-1}) - (NEE_{i+1} - NEE_i) \quad (2.1)$$

e il valore è considerato spike se:

$$d_i < Md - \left(\frac{z \cdot MAD}{0.6745} \right) \quad (2.2)$$

o se

$$d_i > Md + \left(\frac{z \cdot MAD}{0.6745} \right) \quad (2.3)$$

dove Md è la mediana delle differenze, MAD è definito come

$$MAD = median(|d_i - Md|) \quad (2.4)$$

e z è un valore soglia.

Il valore di z non è unico ma è legato alla stima dell'effetto degli spike sui dati ed alla sensibilità della metodologia. In genere per le serie temporali di dati EC vengono usati 3 valori di z : 4, usato convenzionalmente perchè più conservativo, 5.5 e 7 (Papale *et al.*, 2006).

2.4 Soglie di u^* e criteri di applicazione

Tra i metodi proposti per il calcolo della soglia di u^* nel presente lavoro di dottorato sono stati scelti due, quello proposto da Barr *et al.*, (in prep.) e una versione leggermente modificata di quello proposto da Reichstein *et al.* (2005).

2.4.1 metodo Reichstein

Per la determinazione della soglia di u^* , l'insieme di dati notturni¹ è diviso in sette quantili di temperatura dell'aria e, per ciascuna di esse, i dati sono ripartiti in 20 quantili di velocità di frizione. Per ognuno dei 20 quantili viene calcolato il flusso medio di NEE notturna. La soglia è definita come il valore minimo di u^* tale che la differenza tra il corrispondente valore medio di NEE non è significativamente differente dalla media dei flussi delle classi 10 classi successive. Un'altra condizione affinché il valore della soglia sia accettato è la bassa correlazione tra classe di temperatura e u^* ($|r| < 0.4$). Il valore finale della soglia è definito come la mediana delle soglie delle classi di temperatura. Questo procedimento è applicato ad un sottogruppo di dati trimestrali (Gennaio-Marzo, Aprile-Giugno, Luglio-Settembre e Ottobre-Dicembre). Il motivo di queste progressive stratificazioni è quello di minimizzare le influenze che la stagionalità della struttura della vegetazione e la temperatura hanno sulla velocità di frizione.

La soglia di u^* è calcolata per ogni trimestre ma tutti i dati sono filtrati per la maggiore delle soglie trovate secondo un approccio conservativo. Nel caso in cui con questa procedura non è stata individuata in nessuno dei trimestri la soglia allora viene impostata al 90-esimo percentile dei dati (ad esempio almeno il 10% dei dati è conservato).

Per il processamento dei dati di questa tesi di dottorato è il metodo proposto da Reichstein *et al.* (2005) è stato leggermente modificato e denominato fw2. Per avere un valore di soglia più robusto il confronto con le 10 classi successive viene fatto anche per il valore medio successivo a quello attualmente considerato come soglia di u^* . Se anche questo valore non manifesta differenze significative con le classi successive allora la soglia di u^* è accettata.

2.4.2 metodo Barr

Come per il metodo della sezione precedente (2.4.1), anche per il metodo proposto da Barr *et al.*, (in prep.) i dati notturni sono stratificati in stagioni, classi di temperatura e classi di u^* .

¹i dati di flusso sono considerati notturni se al momento della misura la radiazione globale era minore di 20 W m^{-2} .

Fatta questa stratificazione la soglia di u^* viene valutata per ciascuno degli strati stagione $\times$ strati temperatura. La soglia annua della u^* è definita come il valore massimo stagionale delle medie delle soglie stimate per i diversi strati di temperatura di ogni stagione. Per definire il valore di soglia di u^* Barr et al. (in prep.) propongono di applicare una metodologia simile a quella proposta da Solow (1987).

Secondo Solow (1987) è possibile individuare un punto sconosciuto x_c , in una serie di dati (x, y) utilizzando una regressione lineare in due fasi:

$$y_i = \begin{cases} a_0 + a_1 x_i + \varepsilon, & 1 \leq i \leq c \\ a_0 + a_1 x_c + a_2 (x_i - x_c) + \varepsilon, & c < i \leq n \end{cases} \quad (2.5)$$

in cui c e x_c sono i punti di cambio di pendenza (*change point*, CP). L'equazione 2.5 impone che i punti CP siano continui.

Per applicare questa metodologia al calcolo della soglia di u^* , Barr et al. modificano la funzione imponendo un valore all'intercetta:

$$y_i = \begin{cases} b_0 + b_1 x_i + \varepsilon, & 1 \leq i \leq c \\ b_0 + b_1 x_c + \varepsilon, & c < i \leq n \end{cases} \quad (2.6)$$

riducendo il numero di parametri da 3 a 2. La funzione 2.6 è utilizzata per il calcolo della soglia di u^* mentre quella 2.5 è utilizzata come strumento di diagnosi per la bontà della soglia trovata. La funzione di diagnosi 2.5 e quella operativa 2.6 possono essere utilizzate in due modi: (i) il caso più comune in cui si ha un deficit di NEE a bassi valori di u^* , ($a_1 > a_2$ 2.5 o $b_1 > 0$ 2.6); (ii) il caso meno comune in cui si ha un eccesso di NEE a bassi valori di u^* , ($a_1 < a_2$ 2.5 o $b_1 < 0$ 2.6).

Il valore più probabile di x_c è quello che massimizza la F_c (Solow, 1987; Wang, 2003):

$$F_{\max} = \max_{1 \leq c \leq n} (F_c) \quad (2.7)$$

dove F_c è il valore calcolato per c che varia tra 2 ed $n-1$. $F_{\max}$ stima la bontà della parametrizzazione del modello (2.5 o 2.6) considerando non continui i punti CP.

Nel caso in cui i dati non contengano CP, il modello 2.5 diventa:

$$y = \alpha_0 + \alpha_1 x + \varepsilon \quad (2.8)$$

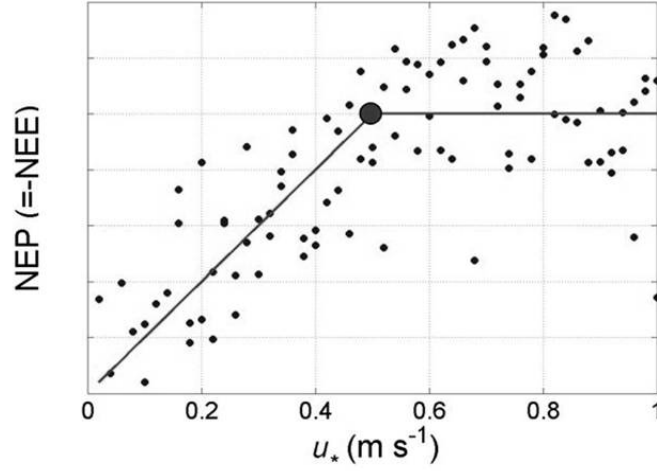


Figura 2.2: Esempio di applicazione del metodo Barr per il calcolo della soglia di u^* . Il punto grande indica il *change point*.

mentre quello 2.6:

$$y = \beta_0 + \varepsilon. \quad (2.9)$$

Utilizzando queste due ipotesi è possibile calcolare i valori di F_c del modello 2.5:

$$F_c = \frac{SSE_{2.9} - SSE_{2.5}}{SSE_{2.5}/(n - n_a)} \quad (2.10)$$

e del modello 2.6:

$$F_c = \frac{SSE_{2.9} - SSE_{2.5}}{SSE_{2.9}/(n - n_b)} \quad (2.11)$$

dove n è il numero dei CP e i termini SSE sono le somme degli errori al quadrato:

$$SSE_{2.5} = \sum_{i=1}^c (y_i - a_0 - a_1 x_i)^2 + \sum_{i=c+1}^n (y_i - a_0 - a_1 x_c - a_2 (x_i - x_c))^2 \quad (2.12)$$

$$SSE_{2.8} = \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_i)^2 \quad (2.13)$$

$$SSE_{2.6} = \sum_{i=1}^c (y_i - b_0 - b_1 x_i)^2 + \sum_{i=c+1}^n (y_i - b_0 - b_1 x_c)^2 \quad (2.14)$$

$$SSE_{2.9} = \sum_{i=1}^n (y_i - \beta_0)^2 \quad (2.15)$$

Per testare le ipotesi nulle, i dati non contengono CP significativi, sono necessari dei valori di F_{max} critici e modello specifici (Lund & Reeves, 2002). Per il confronto tra 2.5 e 2.9 può essere utilizzato il valore critico di F_{max} proposto da Wang (2003). Per il confronto tra 2.8 e 2.9, il valore critico di F_{max} è riportato nella tabella 2.2 calcolato utilizzando 100 ripetizioni Monte-Carlo come consigliato da Lund & Reeves (2002) e da Wang (2003).

n	percentili di F_{max}		
	$F_{max,0.90}$	$F_{max,0.95}$	$F_{max,0.99}$
10	6.299	9.147	18.266
15	5.699	7.877	13.810
20	5.517	7.443	12.648
30	5.322	7.031	11.446
50	5.303	6.876	10.664
70	5.348	6.888	10.503
100	5.447	6.918	10.453
150	5.524	6.981	10.386
200	5.614	7.062	10.560
300	5.739	7.201	10.687
500	5.873	7.342	10.675
700	6.059	7.563	11.007
1000	6.274	7.783	11.232

Tabella 2.2: percentili della distribuzione di F_{max} per l'equazione 2.5

Ulteriori modifiche riguardano anche la stratificazione stagionale la quale può essere fatta in due modi: (i) trimestralmente (con 4 stagioni e 3-8 classi di temperatura in base al numero di dati) con punti di inizio e fine delle stagioni flessibili in modo tale da avere lo stesso numero di dati in ogni stagione; (ii) utilizzando una finestra temporale mobile (con 4 classi di temperatura e 7-17 classi di stagioni in funzione del numero di dati misurati) con una sovrapposizione del 50% tra finestre adiacenti per aumentare la risoluzione temporale e le dimensioni del campione. In entrambe le modalità il mese di Dicembre è spostato all'inizio dell'anno.

Per i valori soglia di u^* stimati con l'equazione 2.8 sono applicati 4 criteri di

controllo della qualità. I valori soglia non sono validi se il CP non era statisticamente significativo (F-test, equazione 2.10 e tabella 2.2, con un livello di significatività del 5%). Per ciascun sito/anno, i CP significativi di tutte le classi e i valori bootstrap erano divisi in modalità D (deficit: $b_1 > 0$) e E (eccesso: $b_1 = 0$) in base ai parametri b_1 dell'equazione 2.5, e le soglie stimate di u^* per la modalità meno frequente sono eliminate. L'analisi del sito/anno non viene applicata se gli strati combinati per tutti i passaggi di bootstrap (vedi sezione 2.4.3) contengono meno di 4000 valori di CP accettabili o se la percentuale di CP è meno del 20%. Gli anni anomali di ciascun sito sono identificati in base ai parametri a_1 e a_2 del modello diagnostico del CP 2.5 ed eliminati. Quando un sito/anno supera i criteri di validità, i valori di soglia della u^* accettati sono mediati tra i diversi stagione $\times$ strati di temperatura per ogni passaggio di bootstrap (vedi sezione 2.4.3), restituendo 1000 stime del valore medio annuale.

2.4.3 Incertezza della soglia di u^*

Per definire l'incertezza della soglia di u^* le procedure dei metodi fw2 e Barr sono ripetute secondo una tecnica di bootstrapping applicata ai dati di un anno intero (Efron & Tibshirani, 1993). Ad ogni ripetizione del bootstrapping dalle misure effettuate in un anno è estratto, con eventuale ripetizione, un campione di 17520 valori. Anche se, dal punto di vista teorico, le estrazioni con ripetizioni potrebbero creare un campionamento sbilanciato verso una particolare stagione (o a particolari condizioni meteorologiche), dato l'elevato numero di dati estratti (17520) questa eventualità è molto improbabile. Ad esempio la probabilità che meno di 4000 punti siano estratti dall'estate è pari a $1,14 \times 10^{-11}$. Il vantaggio del bootstrapping è soprattutto quello che i parametri possono essere stimati senza fare ipotesi sulla normalità della distribuzione e utilizzando anche campioni di piccole dimensioni. L'effetto dei dati mancanti è anch'esso incluso dato che i punti dei dati mancanti sono campionati con la stessa probabilità (Papale *et al.*, 2006). Per il metodo proposto da Barr *et al.*, l'incertezza è definita in base alla distribuzione dei passaggi di bootstrap ed è espressa in un intervallo di confidenza del 95% sulla base dei percentili 2.5 e 97.5 o come coefficiente di

variazione (CV), rapporto tra la deviazione standard e la media.

2.5 Applicazione delle soglie di u^*

Oltre all'eliminazione dei valori di flusso con u^* al di sotto della soglia individuata è possibile adottare un altro tipo di approccio. Con l'approccio ancora più conservativo, oltre all'eliminazione dei dati misurati durante periodi con u^* al di sotto della soglia, è possibile eliminare anche la prima misura con buona turbolenza ma successiva ad un periodo con u^* al di sotto della soglia.

2.6 Gap Filling

I metodi di gap filling da applicare nella stima dell'incertezza sono stati scelti seguendo quanto suggerito da Moffat *et al.* (2007).

2.6.1 Marginal Distribution Sampling

Per il gap filling dei dati EC e di quelli meteorologici con il metodo MDS si applica un metodo simile a quello utilizzato da Falge *et al.* (2001a). Falge *et al.* (2001a) propongono di riempire il valore misurato mancante con la media dei valori misurati nello stesso momento in giorni adiacenti (metodo Mean Diurnal Variation, MDV). Falge *et al.* (2001a), in alternativa al metodo MDV, propongono l'applicazione del metodo delle lookup table secondo il quale sono create delle tabelle con valori di flusso misurati in determinate condizioni meteorologiche. I valori di flusso mancanti possono essere “looked-up” dalla tabella in funzione delle condizioni meteorologiche associate ai dati mancanti.

Il metodo di gap filling proposto da Reichstein *et al.* (2005) combina la metodologia MDV con quella delle lookup table considerando sia la covarianza dei flussi rispetto alle variabili meteorologiche sia l'autocorrelazione temporale dei flussi. In questo metodo di gap filling sono diversificate 3 casi: (1) solo i dati di interesse sono mancanti ma tutte i dati meteo sono disponibili; (2) sia la temperatura dell'aria (T_{air}) che il deficit di pressione di vapore (VPD) sono mancanti ma è presente la radiazione globale (R_g); (3) anche la R_g è mancante.

Nel caso (1) i valori mancanti sono riempiti dalla media dei valori che si trovano in condizioni meteorologiche simili e all'interno di una finestra di 7 giorni. Le condizioni meteorologiche sono considerate simili se che la R_g , T_{air} e VPD differiscono, al massimo, rispettivamente, di 50 W m^{-2} , $2.5 \text{ }^\circ\text{C}$ e 5.0 hPa . Se queste condizioni meteorologiche non si verificano allora la finestra temporale viene allargata a 14 giorni.

Per il caso (2) è seguito un approccio simile al precedente ma la condizione meteorologica simile è definita solo per la R_g la quale deve differire, al massimo, di 50 W m^{-2} e la dimensione della finestra temporale non viene ingrandita.

Nel caso (3) i valori mancanti sono riempiti dalla media dei valori misurati nella stessa ora del giorno ± 1 ora. In questo caso la grandezza della finestra temporale inizia con 0.5 giorni e interpolando linearmente i dati disponibili delle ore adiacenti. Se anche dopo questo passaggio i valori mancanti persistono, il procedimento è ripetuto aumentando la dimensione della finestra temporale fino a che tutti i valori mancanti sono riempiti. La figura 2.3 rappresenta i passaggi e le grandezze delle finestre considerate nell'algoritmo di gap filling. Oltre al riempimento dei dati mancanti vengono registrati anche la grandezza della finestra temporale, il numero e la deviazione standard dei valori mediati, in questo modo è possibile definire, per ogni sostituzione, il grado di accuratezza.

Per comodità i dati sostituiti sono classificati in tre categorie (A, B, C) in base al metodo (1, 2, 3) ed alla grandezza della finestra temporale (figura 2.3). La classificazione è basata sull'idea che la stima del valore mancante migliora con la conoscenza delle condizioni meteorologiche e con l'utilizzo dell'autocorrelazione temporale della variabile che è migliore per finestre temporali di piccole dimensioni Reichstein *et al.* (2005).

2.6.2 Reti Neurali Artificiali

Le reti neurali artificiali (Artificial Neural Network, ANN) sono state ampiamente utilizzate nei casi in cui sia richiesta la soluzione di problemi complessi e che sono difficilmente risolvibili con i metodi tradizionali Mohandes *et al.* (1998). Rispetto all'approccio tradizionale, le ANN hanno il vantaggio di

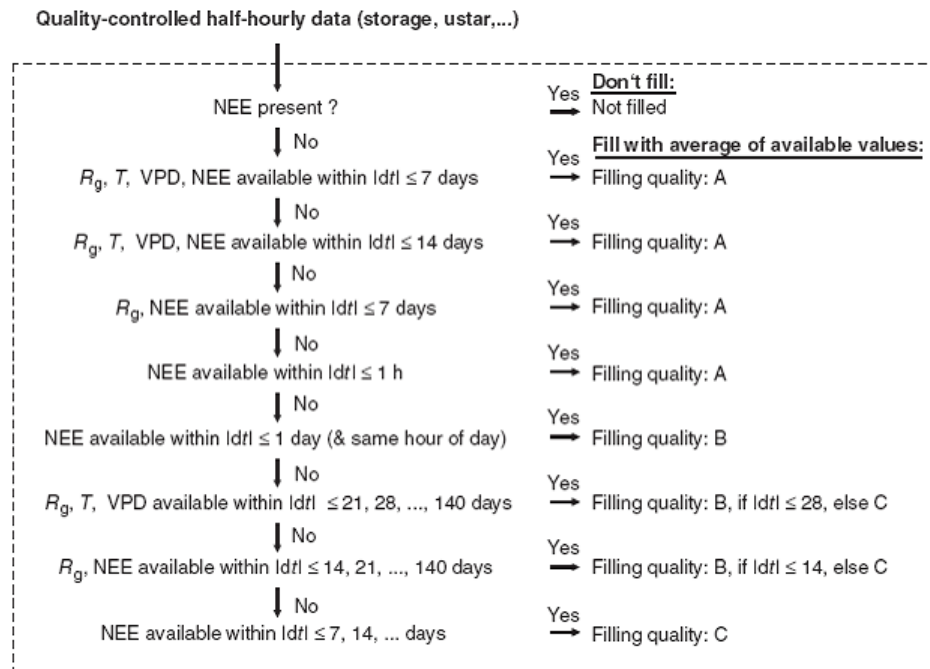


Figura 2.3: Diagramma di flusso per il gap filling con il metodo MDS. NEE rappresenta lo scambio netto di CO₂ dell'Ecosistema; R_g, radiazione globale; T, temperatura dell'aria; VPD, deficit di pressione di vapore; |dt|, differenza nel tempo. Per gli indici di qualità: A, alta; B, media; C, bassa. (Reichstein *et al.*, 2005).

estrarre le informazioni desiderate utilizzando i dati di input senza avere una specifica equazione. Di contro sono necessari un certo numero di coppie di dati input-output. Le ANN possono essere ri-addestrate e adattarsi facilmente a nuovi dati.

Una ANN è un paradigma di elaborazione delle informazioni che si ispira ai sistemi nervosi biologici, come il cervello, elaboratori di informazioni. L'elemento chiave di questo paradigma è la nuova struttura del sistema di elaborazione delle informazioni. La struttura è composta da un gran numero di elementi di elaborazione altamente interconnessi (neuroni) che lavorano all'unisono per risolvere problemi specifici. La rete è costituita di solito da uno strato di input, uno strato nascosto e uno strato di output (Esen *et al.*, 2009).

La ANN impara dagli esempi e corregge i pesi delle connessioni sinaptiche

presenti tra i neuroni. La ANN funziona come un modello “black box” in cui non sono necessarie informazioni dettagliate sulle funzioni. Il neurone artificiale ha diversi input e un output. Il neurone può funzionare in due modalità: (i) in modalità addestramento (training) in cui il neurone può essere addestrato per un certo schema di input; (ii) in modalità applicazione, se lo schema di input è stato acquisito ed a ogni input viene associato l’output che sarà l’attuale risultato (Zhang, 2005). Molti degli algoritmi usati per l’addestramento delle ANN sono basati sul concetto di gradiente discendente il più popolare dei quali è l’algoritmo di retro propagazione (back propagation algorithm). Dato che è molto difficile sapere quale algoritmo di addestramento sarà il più veloce per un dato problema, il migliore è di solito scelto per tentativi ed errori (Esen *et al.*, 2009). Anche nell’applicazione delle ANN per riempire i dati mancanti della serie temporale della NEE è necessario un insieme di dati per addestrare la rete e fissare i pesi delle interconnessioni tra i neuroni. In genere nel dataset di addestramento per input si considerano le variabili meteo mentre come output si considera la NEE. Una volta che la rete è stata addestrata con il dataset completo di input ed output, i pesi delle interconnessioni vengono bloccati. A questo punto la rete viene utilizzata per calcolare i flussi di NEE mancanti fornendo solo i dati meteo di input.

2.7 Metodi di partitioning

Per l’applicazione e la validazione di modelli spesso è utile separare (*partition*) la NEE in GPP e Reco attraverso opportuni “algoritmi di partizione”. Come detto in precedenza questo aiuta a capire meglio quali, tra i processi di assorbimento e respirazione, i modelli rappresentano meno fedelmente (Falge *et al.*, 2001a; Reichstein *et al.*, 2002b). Se, ad esempio, un modello sovrastima sia la GPP che la Reco della stessa quantità, l’errore del modello non si manifesterebbe se venissero confrontate le NEE misurate con quelle modellizzate. Inoltre, la partizione di NEE è necessaria per capire meglio a cosa è dovuta la variabilità interannuale e quella tra siti della NEE (Valentini *et al.*, 2000).

Anche se i metodi di *partitioning* possono essere utilizzati per riempire i

dati mancanti (*gap-filling*) della serie temporale EC in realtà sono concettualmente diversi infatti il gap filling è, essenzialmente, un procedimento di interpolazione dei dati mancanti (Falge *et al.*, 2001a,b; Hui *et al.*, 2003) mentre molti metodi di partitioning sono un'estrapolazione dai dati diurni e notturni (Reichstein *et al.*, 2005).

2.7.1 Partitioning Reichstein

L'approccio proposto da Reichstein *et al.* (2005) considera la relazione temperatura dell'aria Reco secondo il modello esponenziale (Lloyd & Taylor, 1994):

$$Reco = Rref \exp \left(E0 \left(\frac{1}{(Tref - T0)} - \frac{1}{(T - T0)} \right) \right) \quad (2.16)$$

dove T0 è considerato costante e pari a 46.02 °C; E0, l'energia di attivazione che essenzialmente determina la sensibilità della temperatura; Tref è fissata 10 °C. I dati misurati nell'anno vengono divisi in sottoperiodi e la parametrizzazione del modello viene fatta per ciascuno di essi. La lunghezza dei sottoperiodi è pari a 15 giorni con uno shift tra una finestra e la successiva di 5 giorni (quindi una sovrapposizione di 10 giorni con la precedente). Prima della parametrizzazione viene fatto un controllo sul numero di dati disponibili all'interno della finestra infatti se sono presenti almeno 6 dati e se la temperatura varia di almeno 5 °C allora la parametrizzazione del modello è considerata accettabile.

La lunghezza della finestra è di 15 giorni perché ritenuta sufficientemente breve per evitare gli effetti di disturbo stagionali e abbastanza lunga per fornire un numero di dati sufficiente e di un range di temperatura adeguato per la parametrizzazione (Reichstein *et al.*, 2005). La scelta della lunghezza della finestra pari a 15 giorni è supportata sia dalla massimizzazione dello spettro di frequenza della temperatura dell'aria sia da precedenti risultati che mostrano un gap spettrale di 15 giorni nei dati di flusso (Baldocchi *et al.*, 2001b). Come variabile di input il modello utilizzata la temperatura dell'aria invece della temperatura del suolo perché, in genere, riesce a spiegare meglio molta della varianza della Reco (Reichstein *et al.*, 2005). Per ogni sottoperiodo i parametri del modello e le relative statistiche sono registrate e valutate

dopo la parametrizzazione. I risultati dei sottoperiodi sono considerati validi solo se l'errore relativo standard delle stime del parametro $E0$ è inferiore del 50% e quando il valore stimato è compreso tra 0 e 450 K. Nel caso in cui per $E0$ non è possibile definire un andamento stagionale la stima di $E0$ rappresentativo dell'anno è fatta seguendo due strade: (i) vengono mediati i valori di $E0$ dei 3 periodi in cui è stata registrato il minor errore standard e (ii) viene calcolata la media ponderata di tutti gli $E0$ considerando l'inverso dell'errore standard come peso (Reichstein *et al.*, 2005).

Dopo aver stimato il parametro $E0$ è necessario definire il valore del parametro che stimi il livello di Respirazione indipendente dalla temperatura ($Rref$). Dato che anche questo parametro varia durante l'anno viene stimato considerando periodi di 4 giorni consecutivi. In questi 4 giorni il modello 2.16 viene parametrizzato considerando fissi tutti i parametri tranne $Rref$. Questa stima è effettuata una volta con $E0$ stimato dal sottoperiodo e una volta con $E0$ stimato per tutto l'anno. Le stime $Rref$ sono assegnate a un punto nel tempo infatti se, ad esempio, i dati di Reco sono disponibili solo nella quarta notte e nelle ore 18:00-06:00, allora il valore di $Rref$ viene assegnato alle 18:00 della quarta notte. Una volta che i parametri di $Rref$ sono stati stimati per ciascun periodo, i valori mancanti vengono definiti interpolando linearmente le stime calcolate. Questa procedura costituisce una serie temporale del parametro $Rref$ del modello 2.16 con un valore di $E0$ nel sottoperiodo costante (Reichstein *et al.*, 2005).

Con questa procedura è possibile stimare la Reco per ogni valore della serie temporale secondo l'equazione

$$Reco(t) = Rref(t) \exp \left(E0 \left(\frac{1}{(Tref - T0)} - \frac{1}{(T(t) - T0)} \right) \right). \quad (2.17)$$

Anche se all'equazione 2.17 è simile alla 2.16 in realtà i parametri e le variabili indicate con il simbolo t sono in funzione del tempo (Reichstein *et al.*, 2005).

La stima dei parametri è fatta utilizzando il metodo dei minimi quadrati ordinari il quale massimizza la probabilità dei valori del parametro assumendo che i residui sono distribuiti normalmente. Reichstein *et al.* (2005) affermano che la distribuzione dei residui (differenza tra valori di flusso notturni misurati ed Reco calcolata dal modello) è quasi sempre simmetrica, ma più

stretta rispetto alla Gaussuiana. Questa leggera differenza tra distribuzione dei residui e curva Gaussiana implica solo un piccolo errore nell'applicazione del metodo dei minimi quadrati ordinari (Reichstein *et al.*, 2005).

2.7.2 Partitioning Lasslop

Un altro approccio comunemente adottato per il partitioning è quello di estrapolare la respirazione dalle curve di luce, in questo caso si considera che la NEE varia in funzione sia della temperatura (soprattutto influenzando la Reco) sia del deficit di pressione di vapore (VPD) (influenza principalmente la GPP agendo sull'apertura stomatica), sia per altri fattori.

Secondo Desai *et al.* (2008) la grande variabilità osservata tra i metodi che stimano il ciclo diurno della Reco potrebbe essere causata dagli effetti confondenti introdotti da alcuni fattori. Ad esempio durante periodi aridi (alti valori di VPD) il ciclo diurno della NEE assume una forma asimmetrica a causa dell'elevata respirazione del pomeriggio ma anche dalla limitazione stomatica della GPP. In queste condizioni VPD massimo viene misurato dopo il massimo della radiazione giornaliera massima. Conseguenza di questa interazione l'accumulo di carbonio ad uno stesso livello di irraggiamento può essere sostanzialmente minore nel pomeriggio rispetto alle ore mattutine.

Lasslop *et al.* (2010), per fare il partitioning, utilizzano la classica curva di luce iperbole equilatera (Falge *et al.*, 2001a):

$$NEE = \frac{\alpha\beta R_g}{\alpha R_g + \beta} + \gamma \quad (2.18)$$

dove α ($\mu\text{mol m}^{-2} \text{s}^{-1}$) è la pendenza iniziale della curva di luce e rappresenta l'efficienza quantica della canopy; β ($\mu\text{mol m}^{-2} \text{s}^{-1}$) rappresenta il valore massimo di assorbimento di CO₂ a saturazione di luce, γ ($\mu\text{mol m}^{-2} \text{s}^{-1}$) è la Reco e R_g è la radiazione globale (W m^{-2}).

Per considerare l'influenza della temperatura sulla respirazione, il termine γ è rimpiazzato con un modello di respirazione (Gilmanov *et al.*, 2003), in questo caso con il modello di Lloyd & Taylor (1994)

$$NEE = \frac{\alpha\beta R_g}{\alpha R_g + \beta} + R_{ref} \exp \left(E0 \left(\frac{1}{(T_{ref} - T0)} - \frac{1}{(T - T0)} \right) \right). \quad (2.19)$$

Una ulteriore modifica alla curva di luce è quella che considera il VPD come fattore limitante della GPP. In questo caso il parametro β dell'equazione 2.19 è rimpiazzato con una funzione esponenziale decrescente (Körner, 1995) per elevati VPD:

$$\beta = \begin{cases} \beta_0 \exp(-k(\text{VPD} - \text{VPD}_0)), & \text{VPD} > \text{VPD}_0 \\ \beta = \beta_0, & \text{VPD} < \text{VPD}_0 \end{cases} \quad (2.20)$$

Per l'applicazione delle relazioni 2.20 dovrebbe essere utilizzato il VPD fogliare. Dato che, in genere, i siti EC misurano solamente il VPD atmosferico è necessario adottare opportuni accorgimenti. Il VPD fogliare è maggiore o minore di quello atmosferico se la temperatura fogliare è superiore o inferiore a quella dell'aria. Per ignorare le differenze di potenziale specifiche della vegetazione è stato considerato il valore soglia VPD_0 pari a 10 hPa (in accordo con i risultati degli studi a livello fogliare di Körner (1995)).

I parametri E_0 , R_{ref} , α , β e k sono stimati in due differenti modi: (i) E_0 è stimato usando i dati notturni ($R_g < 4 \text{ W m}^{-2}$) poi, fissata E_0 , vengono stimati R_{ref} , α , β e k utilizzando i dati diurni; (ii) tutti i parametri (E_0 , R_{ref} , α e β e k) sono stimati utilizzando i dati diurni. In questa modalità il limite superiore di E_0 , riportato nella tabella 2.3, non viene utilizzato altrimenti il parametro E_0 viene spesso rifiutato nei periodi con elevato VPD.

Lasslop *et al.* (2010), nel processo di partitioning, considerano trascurabili i processi non biologici nel processo di separazione (Hofmeister, 1997; Kowalski *et al.*, 2008). I modelli sono parametrizzati con una breve finestra temporale per considerare la variabilità stagionale dello stato vegetativo dell'ecosistema. I parametri sono stimati ogni due giorni in una finestra mobile di 4 giorni per i dati diurni ed una finestra di 12 giorni per i dati notturni². La relazione tra misure di NEE e temperatura è considerata indipendente da errori sistematici durante i periodi notturni di bassa turbolenza. Il parametro che determina l'entità della respirazione, R_{ref} , è stimato dai dati diurni ed è quindi indipendente da questi errori.

Affinché le stime annuali di GPP ed Reco siano affidabili sono necessari un alto numero di stime o brevi periodi di parametri mancanti durante l'anno. I settaggi dei parametri durante il procedimento di stima sono raccolti nella

²per notte si intende i dati che hanno una radiazione globale inferiore a 20 W m^{-2} .

tabella 2.3.

	Valore iniziale	valori accettabili	se il parametro non è accettabile
E0	100	50-400	il valore è settato come la finestra precedente, se $\nexists$ o se $E0 < 50 \Rightarrow E0 = 50$, se $E0 > 400 \Rightarrow E0 = 400$
Rref	Media della NEE notturna	> 0	parametro non usato
α	0.01	$\geq 0, < 0.22$	il valore è settato come la finestra precedente, se $\nexists$ o se $\alpha < 0 \Rightarrow \alpha = 0$
β	0.03 quantile - 0.97 quantile di NEE	$\geq 0, < 250$ se $> 100 \Rightarrow \sigma(\beta) < \beta$	se negativo $\beta = 0$, altrimenti β non usato
k	0	≥ 0	settato a zero

Tabella 2.3: Settaggio dei parametri durante il processo di stima

L'algoritmo di parametrizzazione è sensibile alla scelta dei valori iniziali di β , il valore di saturazione della fotosintesi. Per affrontare questo problema i parametri vengono stimati 3 volte, dimezzando e raddoppiando i valori iniziali riportati in tabella 2.3 rispettivamente nella seconda e terza parametrizzazione. Tra tutti i parametri sono considerati quelli con il minor RMSE. I parametri sono accettati solo se compresi nei range tabella 2.3. Se i parametri erano oltre questi range, alcuni venivano fissati, come riportato nell'ultima colonna della tabella 2.3, e gli altri riparametrizzati. Se anche dopo questo passaggio i parametri erano fuori dai range allora i parametri non venivano usati. In questo caso i flussi sono stati calcolati per i due set di parametri vicini e poi interpolati linearmente utilizzando il reciproco della distanza dei parametri come peso. Il parametro k è stimato per ogni finestra di 4 giorni per quantificare il massimo assorbimento del carbonio in funzione del VPD.

2.7.3 Partitioning Van Gorsel

Come già esposto in precedenza (paragrafo 1.4.3) per eliminare i dati di flusso ottenuti durante periodi di bassa turbolenza è stato proposto l'utilizzo di una soglia per la velocità di frizione (Goulden *et al.*, 1996b). La definizione di una adeguata soglia di u^* è fondamentale infatti, come riportato ad esempio da Miller *et al.* (2004), variazioni relativamente piccole della soglia di u^* comportano un cambio di segno della NEE di una foresta tropicale.

van Gorsel *et al.* (2007) propongono un metodo alternativo a quello della soglia di u^* utilizzando le diverse scale temporali della riduzione dei flussi turbolenti, del tasso di scambio dello storage e dell'insorgenza dell'avvezione per lo sviluppo di una stratificazione stabile dell'aria entro la canopy.

La figura 2.4 mostra l'andamento medio diurno della somma dei flussi turbolenti verticali (F_c) e del termine tasso di scambio dello storage (F_s). In genere, all'inizio della sera, le serie temporali EC mostrano un aumento di concentrazione di CO₂ causato dal un cambiamento nel tasso di scambio dello storage (il termine storage dell'equazione di massa non è più trascurabile). Raggiunto un valore massimo (R_{max}), l'accumulo si riduce (il termine storage è piccolo o tende a zero) e le concentrazioni rimangono relativamente costanti.

van Gorsel *et al.* (2007) propongono di utilizzare il valore R_{max} per svilup-

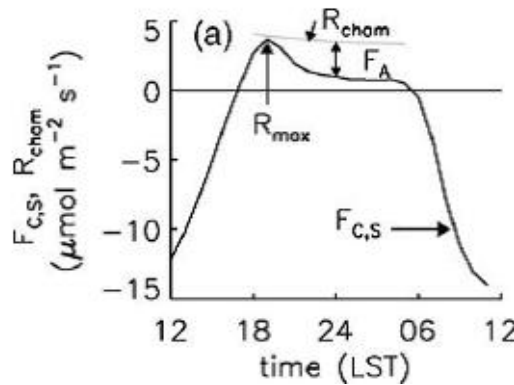


Figura 2.4: Andamento medio diurno della somma dei flussi turbolenti verticali e del termine tasso di scambio dello storage ($F_{C,S}$) e della Reco derivata da misure con camerette (R_{chamb}). F_A rappresenta l'avvezione orizzontale e verticale. (van Gorsel *et al.*, 2009).

pare un modello di respirazione (van't Hoff, 1898):

$$NEE_{night} = \varepsilon_1 \exp(\varepsilon_2 T_s) \quad (2.21)$$

dove ε_1 e ε_2 sono i coefficienti della regressione. T_s rappresenta la temperatura dello strato superiore del suolo (al di sotto di -0.2 m di profondità).

Il modello è poi applicato a tutti i dati con una finestra temporale mobile di 30 giorni. L'utilizzo di una finestra temporale di 30 giorni è giustificato dall'effetto delle variazioni stagionali della respirazione dovuto, ad esempio, a cambiamenti del contenuto idrico e di sostanza organica del suolo.

Il momento in cui si verifica R_{max} è dato, considerando l'intervallo di 30 giorni, dall'andamento medio giornaliero di F_c e F_s . Successivamente è scelto un periodo di 3 ore che inizia un'ora prima del momento di R_{max} . Per valutare la qualità del punto di inizio all'interno delle 3 ore devono essere considerate le seguenti condizioni:

- assicurare che le condizioni atmosferiche sopra la canopy non siano fortemente stabili e favorevoli alla turbolenza intermittente.
- per eliminare gli spike della R_{max} , sono utilizzate solo le stime di respirazione entro un range pari a $r = [0.5\bar{R}_{LRC}, 2\bar{R}_{LRC}]$, in cui $\bar{R}_{LRC}$ è la media della respirazione ottenuta dall'applicazione di una curva di luce.
- per eliminare le notti con elevata turbolenza intermittente, sono utilizzati solo i dati in cui il rapporto tra la media di NEE e la relativa deviazione standard ($\frac{\bar{E}_{NEE}}{\sigma_{E_{NEE}}}$) supera l'unità. In cui $\bar{R}_{LRC}$ è utilizzato come miglior stimatore di $\bar{E}_{NEE}$ mentre $\sigma_{E_{NEE}}$ è la deviazione standard dei termini F_c e F_s nella stessa finestra temporale.
- infine i dati misurati durante le precipitazioni sono esclusi poichè sono, in genere, di bassa qualità e anche perchè la respirazione è probabilmente mal rappresentata infatti al respirazione aumenta nel caso in cui il suolo è bagnato dopo periodi di aridità oppure si riduce a causa della riduzione della diffusione in terreni bagnati (Lee *et al.*, 2004; Savage & Davidson, 2003).

2.8 Approccio Statistico

La statistica è la scienza che risolve i problemi in presenza di variabilità. Questa definizione implica che se per la soluzione di un problema è necessaria la registrazione di dati, la statistica deve essere una componente integrata nel processo di risoluzione. Nel procedimento di risoluzione del problema è insito un certo grado di incertezza causato dalla naturale variazione dei risultati virtualmente presente in tutti gli esperimenti (Manson *et al.*, 2003). Gli esperimenti sono eseguiti per registrare dei dati (campioni) dai quali estrarre l'inferenza delle popolazioni o dei processi³. Di solito le inferenze sono conclusioni estratte attraverso parametri di modelli o distribuzioni di probabilità, le conclusioni sono basate sul calcolo di statistiche dalle singole risposte (Manson *et al.*, 2003).

Tutti i dati sono soggetti ad una varietà di fonti che inducono variazioni nelle misure. Queste variazioni possono manifestarsi per differenze fisse tra strumenti, differenze casuali (random) dovute a cambiamenti delle condizioni ambientali, errori di misura durante la registrazione da parte dello strumento, ma anche per effetti dovuti ad altre cause conosciute o sconosciute (Manson *et al.*, 2003).

Lo studio della variabilità include la costruzione di un disegno sperimentale e lo sviluppo di modelli statistici⁴ che descrivano la variabilità (Manson *et al.*, 2003). A volte gli esperimenti sono condotti per confermare, caratterizzare o confermare/rifiutare un modello e, soprattutto, le ipotesi alla base del modello stesso (Manson *et al.*, 2003).

I modelli possono avere molte forme ma, in genere, tutti cercano di individuare una o più variabili di risposta⁵, attraverso rapporti con uno o più fattori⁶. In alcuni casi le variabili di risposta sono definite in accordo con

³Il processo è una serie di operazioni ripetibili il cui risultato è una caratteristica osservabile o misura.

⁴Un modello è definito *statistico* se è derivato da dati che sono soggetti a diversi tipi di specifiche, osservazione, esperimento e/o errori di misura. In statistica per errore dei modelli non è indice che il modello è sbagliato ma che sono presenti fonti sconosciute di variazione incontrollate, spesso errori di misure (Manson *et al.*, 2003).

⁵variabili di risposta o risultato di un esperimento

⁶il fattore è una variabile di controllo dell'esperimento la quale può influenzare i valori osservati delle variabili di risposta

un modello di probabilità il quale è una funzione di alcune (in genere sconosciute) costanti. In altri casi il modello contiene uno o più fattori oltre alle (sconosciute) costanti (Manson *et al.*, 2003).

2.8.1 Analisi Multifattoriale

Quando un esperimento considera due o più fattori, questi possono influenzare le variabili risposta sia singolarmente che congiuntamente. Spesso, come nel caso degli esperimenti con un fattore per volta, non è possibile valutare adeguatamente gli effetti⁷ congiunti dei fattori (Manson *et al.*, 2003). Per questo motivo gli esperimenti fattoriali includono, nel disegno sperimentale, tutte le possibili combinazioni dei livelli-fattoriali⁸ (Manson *et al.*, 2003).

I dati multifattoriali derivano da due situazioni:

1. Esperimenti Fattoriali⁹: in molti esperimenti è opportuno esaminare l'effetto¹⁰ di due o più fattori contemporaneamente;
2. Disegno sperimentale: in genere è preferibile suddividere le unità sperimentali in gruppi ("blocchi") prima di attribuire i diversi livelli di fattore. I gruppi sono definiti in modo tale da ridurre la stima della varianza utilizzata per inferenze;

un insieme di dati può avere contemporaneamente una struttura fattoriale e includere fattori a "blocchi". Nell'esperimento fattoriale sono considerati diversi fattori contemporaneamente in ogni unità sperimentale, la quale è considerata come una unità di osservazione (Freund & Wilson, 2003).

L'analisi statistica dei dati di un esperimento fattoriale è rivolta ad esaminare come la variabile di risposta è influenzata dal diverso livello dei fattori.

⁷Per effetto si intende un cambiamento nella risposta media corrispondente a un cambiamento nella combinazione del livello di un fattore o ad un cambiamento delle condizioni sperimentali.

⁸livello-fattoriale o livello è lo specifico valore di un fattore. A volte il livello è detto anche trattamento

⁹Un esperimento in cui sono considerate unità sperimentali con tutte le combinazioni dei fattori è chiamato esperimento fattoriale.

¹⁰Per effetto si intende un cambiamento nella risposta media corrispondente a un cambiamento nella combinazione del livello di un fattore o ad un cambiamento delle condizioni sperimentali.

Quando esiste un'interazione questa può modificare e, a volte, annullare le influenze dei singoli fattori per questo è importante condurre esperimenti per definire le interazioni. In genere esperimenti con un solo fattore alla volta (o che considerano solo certe combinazioni di fattori) non sono in grado di rilevare le interazioni, lo scopo degli esperimenti fattoriali è proprio quello di considerare contemporaneamente tutte le combinazioni dei fattori (Freund & Wilson, 2003).

Se i livelli di un fattore influenzano la medie delle risposte, le influenze dei livelli del fattore sono definite come *effetti fissi* (costanti). Se i livelli di un fattore influenzano la variabilità delle risposte, gli effetti del fattore sono definiti come *effetti casuali* (random) (Manson *et al.*, 2003).

2.8.2 ANalysis-Of-Variance (ANOVA)

L'analisi della varianza (ANalysis-Of-Variance, ANOVA) separa la variabilità osservata nei risultati di un esperimento in due componenti: (i) variabilità dovuta a cause conosciute¹¹; (ii) variabilità casuale (random) per cause sconosciute¹². Nella componente random delle variabilità, oltre ad includere le variazioni occasionali e gli errori di misura, sono inclusi anche tutti gli effetti dei fattori che non sono stati nè controllati nè misurati quindi, anche questi, vengono etichettati come variazione "random" (Berthouex & Brown, 2002). L'analisi della varianza affronta il problema dell'identificazione dei fattori che contribuiscono significativamente alla variabilità complessiva delle misure. L'idea di base è dividere la variabilità totale dei dati e attribuire il peso ai fattori studiati nell'esperimento e alle loro interazioni (Berthouex & Brown, 2002).

La tecnica ANOVA è utilizzata per testare l'ipotesi che le medie di due o più gruppi siano uguali, assumendo che le popolazioni campionate siano distribuite normalmente (Yang & Trewn, 2004).

La validità di ogni inferenza dipende dal rispetto di determinate assunzioni riguardanti la natura dei dati (Freund & Wilson, 2003). Per l'applicazione dell'ANOVA le assunzioni sono espresse, in genere, in termini degli elementi

¹¹dovuta a cambiamenti dei fattori sperimentali o misure che covariano.

¹²dovuta a effetti incontrollati come fenomeni occasionali e errori di misura

di un modello lineare e, in particolare, per il termine dell'errore casuale ε . Le assunzioni sono:

1. il modello adottato e i suoi parametri rappresentano adeguatamente il comportamento dei dati;
2. i termini ε sono variabili casuali normalmente distribuite con media pari a zero e varianza σ^2 ;
3. i termini ε sono indipendenti in senso probabilistico cioè un valore di ε non influenza i valori gli altri ε .

Per spiegare l'ANOVA viene riportato il caso di un esperimento fattoriale a 2 fattori.

Esperimenti fattoriali a 2 fattori

Il più semplice esperimento fattoriale è quello che implica due fattori, ad esempio i fattori A e B che hanno rispettivamente a livelli e b livelli. In questo caso le n repliche rappresentano tutte le combinazioni dei livelli a e b ($a \times b$). L'osservazione nella cella ij per la k replica è rappresentata come y_{ijk} . Per l'applicazione dell'analisi ANOVA le abn osservazioni dovrebbero essere casuali.

Adottando l'analisi ANOVA per gli esperimenti con 2 fattori le osservazioni possono essere descritte con un modello statistico lineare:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases} \quad (2.22)$$

dove μ è la media generale, τ_i è l'effetto dell' i -esimo livello del fattore A, β_j è l'effetto dell' j -esimo livello del fattore B, $(\tau\beta)_{ij}$ è l'effetto dell'interazione tra A e B, infine ε_{ijk} è la componente dell'errore random avente una distribuzione normale, media pari a zero e varianza pari a σ^2 (Montgomery & Runger, 2002).

In questo caso l'interesse è rivolto a verificare l'ipotesi che sia il fattore A che il fattore B che la loro interazione non abbiano effetto. Si ipotizza che: (i) i fattori A e B sono considerati fissi; (ii) che l'inferenza è confinata a

livelli di questi due fattori; (iii) nel modello statistico di solito gli effetti τ_i , β_j e $(\tau\beta)_{ij}$ sono definiti come deviazioni rispetto alla media in modo tale che $\sum_{i=1}^a \tau_i = 0$, $\sum_{j=1}^b \beta_j = 0$, $\sum_{i=1}^a (\tau\beta)_{ij} = 0$ e $\sum_{j=1}^b (\tau\beta)_{ij} = 0$ (Montgomery & Runger, 2002).

Per testare l'ipotesi sugli effetti di A, B e loro interazioni è possibile utilizzare l'analisi della varianza. Per le formule seguenti verranno adottare le seguenti simbologie:

$y_{i..}$: somma delle osservazioni per l' i -esimo livello del fattore A;

$y_{.j.}$: somma delle osservazioni per l' j -esimo livello del fattore B;

$y_{ij.}$: somma delle osservazioni per l' ij -esimo livelli dei fattori A, B;

$y_{...}$: somma di tutte le osservazioni;

$\bar{y}_{i..}$: media delle osservazioni per l' i -esimo livello del fattore A;

$\bar{y}_{.j.}$: media delle osservazioni per l' j -esimo livello del fattore B;

$\bar{y}_{ij.}$: delle osservazioni per l' ij -esimo livello dei fattori A, B;

$\bar{y}_{...}$: media di tutte le osservazioni.

Adottando questa simbologia le formule adottate per il calcolo dei termini del modello sono:

$$\begin{aligned}
 y_{i..} &= \sum_{j=1}^b \sum_{k=1}^n y_{ijk} & \bar{y}_{i..} &= \frac{y_{i..}}{bn} & i &= 1, 2, \dots, a \\
 y_{.j.} &= \sum_{i=1}^a \sum_{k=1}^n y_{ijk} & \bar{y}_{.j.} &= \frac{y_{.j.}}{an} & j &= 1, 2, \dots, b \\
 y_{ij.} &= \sum_{k=1}^n y_{ijk} & \bar{y}_{ij.} &= \frac{y_{ij.}}{n} & \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases} \\
 y_{...} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk} & \bar{y}_{...} &= \frac{y_{...}}{abn}
 \end{aligned}$$

Le ipotesi da verificare sono:

1. $H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0$ (il fattore A non ha effetto)

H_1 : almeno un $\tau_i \neq 0$

2. $H_0 : \beta_1 = \beta_2 = \dots = \beta_b = 0$ (il fattore B non ha effetto)

H_1 : almeno un $\beta_i \neq 0$

3. $H_0 : (\tau\beta)_{11} = (\tau\beta)_{12} = \dots = (\tau\beta)_{ab} = 0$ (nessuna interazione)

H_1 : almeno un $(\tau\beta)_{ij} \neq 0$

per testare le ipotesi è necessario decomporre la variabilità totale dei dati nelle sue componenti e poi confrontare i vari elementi della decomposizione. La variabilità totale è misurata dalla somma degli scarti al quadrato (sum of squares, SS) delle osservazioni:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 \quad (2.23)$$

e la decomposizione degli scarti al quadrato è definita come:

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 &= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 \\ &+ an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &+ n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \\ &+ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ijk} - \bar{y}_{ij.})^2 \end{aligned} \quad (2.24)$$

o simbolicamente:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \quad (2.25)$$

l'equazione 2.25 esprime la SS_T totale come la somma degli SS per il fattore A (SS_A), la somma degli SS per il fattore B (SS_B), la somma degli SS per l'interazione dei fattori A e B (SS_{AB}) e un errore delle SS (SS_E) (Montgomery & Runger, 2002). Ad ogni SS sono associati dei gradi di libertà (*degrees*

of freedom, df) anch'essi additivi secondo la formula:

$$abn - 1 = (a - 1) + (b - 1) + ((a - 1)(b - 1)) + ab(n - 1)$$

$abn - 1$ sono i df di SS_T ; $(a - 1)$ e $(b - 1)$ sono i df di SS_A e SS_B ; $(a - 1)(b - 1)$ sono i df di SS_{AB} ; $ab(n - 1)$ sono i df di SS_E .

Dato che le SS sono influenzate dal numero di livelli e dal numero di repliche di ogni livello, in genere per testare le ipotesi vengono calcolate anche le medie degli scarti al quadrato (*mean square*, MS). Se ogni addendo dell'equazione 2.25 è diviso per il corrispondente grado di libertà, è possibile ottenere la media degli scarti al quadrato (MS) per A, B, interazione ed errore:

$$MS_A = \frac{SS_A}{a - 1} \quad MS_B = \frac{SS_B}{b - 1} \quad MS_{AB} = \frac{SS_{AB}}{(a - 1)(b - 1)} \quad MS_E = \frac{SS_E}{ab(n - 1)}$$

A questo punto è possibile calcolare anche i valori attesi per ogni MS quindi:

$$\begin{aligned} E(MS_A) &= E\left(\frac{SS_A}{a - 1}\right) = \sigma^2 + \frac{bn \sum_{i=1}^a \tau_i^2}{a - 1} \\ E(MS_B) &= E\left(\frac{SS_B}{b - 1}\right) = \sigma^2 + \frac{an \sum_{j=1}^b \beta_j^2}{b - 1} \\ E(MS_{AB}) &= E\left(\frac{SS_{AB}}{(a - 1)(b - 1)}\right) = \sigma^2 + \frac{n \sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{(a - 1)(b - 1)} \\ E(MS_E) &= E\left(\frac{SS_E}{ab(n - 1)}\right) = \sigma^2 \end{aligned} \quad (2.26)$$

Dall'esame dei MS attesi è chiaro che se le ipotesi nulle $H_0 : \tau_i = 0$, $H_0 : \beta_j = 0$ e $H_0 : (\tau\beta)_{ij} = 0$ sono tutte vere, tutte e quattro le MS sono stime imparziali di σ^2 (Montgomery & Runger, 2002).

Per verificare che gli effetti del fattore A sono tutti pari a zero ($H_0 : \tau_i = 0$) è possibile utilizzare il rapporto:

$$F_0 = \frac{MS_A}{MS_E} \quad (2.27)$$

il quale è una distribuzione F con $(a - 1)$ e $ab(n - 1)$ gradi di libertà se $H_0 : \tau_i = 0$ è vera. L'ipotesi $H_0 : \tau_i = 0$ è rifiutata, con un certo livello α di significatività, se $f_0 > f_{\alpha, a-1, ab(n-1)}$.

Analogamente testare l'ipotesi $H_0 : \beta_j = 0$ si utilizza il rapporto:

$$F_0 = \frac{MS_B}{MS_E} \quad (2.28)$$

il quale è una distribuzione F con $(b-1)$ e $ab(n-1)$ gradi di libertà se $H_0 : \beta_j = 0$ è vera. L'ipotesi $H_0 : \beta_j = 0$ è rifiutata, con un certo livello α di significatività, se $f_0 > f_{\alpha, b-1, ab(n-1)}$.

Infine, per verificare l'ipotesi $H_0 : (\tau\beta)_{ij} = 0$ si utilizza il rapporto:

$$F_0 = \frac{MS_{AB}}{MS_E} \quad (2.29)$$

il quale è una distribuzione F con $(a-1)(b-1)$ e $ab(n-1)$ gradi di libertà se $H_0 : (\alpha\beta)_{ij} = 0$ è vera. L'ipotesi $H_0 : (\tau\beta)_{ij} = 0$ è rifiutata, con un certo livello α di significatività, se $f_0 > f_{\alpha, (a-1)(b-1), ab(n-1)}$.

In genere conviene prima verificare se esiste interazione e poi valutare gli effetti dei singoli fattori. Quando l'interazione è significativa, l'effetto dei singoli fattori coinvolti nell'interazione non può avere un valore interpretativo molto pratico. La conoscenza delle interazioni, di solito, è più importante della conoscenza sugli effetti dei fattori.

Riassumendo quanto detto finora le formule per il calcolo degli SS sono:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn} \quad SS_A = \sum_{i=1}^a \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn} \quad SS_B = \sum_{j=1}^b \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

$$SS_{AB} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn} - SS_A - SS_B \quad SS_E = SS_T - SS_{AB} - SS_A - SS_B$$

In genere i risultati sono raccolti in una tabella ANOVA, nel caso di esperimento fattoriale a 2 fattori la tabella ANOVA è come quella 2.4:

Fonte di variazione	Somma degli scarti al quadrato	Gradi di libertà	Medie dei quadrati	F_0
Fattore A	SS_A	a-1	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Fattore B	SS_B	b-1	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
A*B	SS_{AB}	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
Errore	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Totale	SS_T	$abn-1$		

Tabella 2.4: tabella ANOVA per un esperimento fattoriale a 2 fattori.

Capitolo 3

Risultati

3.1 Modello Additivo o ANOVA?

Come riportato nelle sezioni precedenti (2.8.2 la validità di ogni inferenza dipende dal rispetto di determinate assunzioni riguardanti la natura dei dati.

Quando si parla di analisi fattoriale spesso si applica l'analisi della varianza senza valutare attentamente che le assunzioni dell'analisi stessa siano verificate. Nelle pagine seguenti verranno analizzate in dettaglio le assunzioni per l'applicazione dell'ANOVA ai dati EC.

I risultati sono calcolati su scala giornaliera perché questa scala temporale è utilizzata per la parametrizzazione e validazione di molti modelli (Hanson *et al.*, 2004; Moffat *et al.*, 2007). Per ridurre l'effetto dei cicli a scala temporale superiore al giorno (ad esempio l'alternarsi delle stagioni, trend interannuali, periodi di siccità, etc.) l'incertezza è calcolata per ogni singolo giorno dell'anno. La stima dell'incertezza su scala giornaliera, oltre alle ragioni ecologiche, sembra giustificabile soprattutto nei siti gestiti (*management*) in cui la rimozione del soprassuolo e gli interventi di tipo agronomico (arature, distribuzione di fertilizzanti, etc.) influiscono sull'entità e andamento dei flussi. Oltre al management e alle pratiche agronomiche l'analisi dell'incertezza su singoli giorni è utile anche per limitare l'effetto di eventi eccezionali e non sempre prevedibili (ad esempio attacchi di parassiti, incendi, schianti da vento).

Oltre all'immediata modifica dell'entità dei flussi, molti dei fenomeni riportati hanno effetto anche sul lungo periodo. Molti studi, soprattutto quelli a scala interannuale, hanno attribuito la variabilità dei flussi a variazioni di alcuni *drivers* climatici (variazioni della temperatura dell'aria, del suolo ect.) (Barford *et al.*, 2001; Hollinger *et al.*, 2004). Inoltre la variabilità dei flussi è dovuta anche a cambiamenti di natura biologica (ad esempio fenomeni di acclimatazione al regime di luce prevalente o cambiamenti del LAI come riportato da Schimel *et al.* (2001); Wang *et al.* (2004); Flanagan *et al.* (2002)) e alle caratteristiche qualitative e quantitative del substrato disponibile per la respirazione (Janssens *et al.*, 2001).

Dal punto di vista strettamente statistico l'analisi fattoriale scompone la varianza complessiva in diversi addendi attribuendo a ciascuno un peso. In figura 3.1 è riportato il profilo della PDF (*Probability Density Function*) delle deviazioni standard dei flussi per diverse classi di flusso medio per tutti i siti/anno considerati. La figura mostra che l'andamento della deviazione standard è riconducibile all'entità dei flussi misurati. Questo risultato sembra confermare quanto riportato da alcuni autori (Hagen *et al.*, 2006; Hollinger & Richardson, 2005). Secondo Hollinger & Richardson (2005)) i dati EC ed i modelli parametrizzati su tali dati presentano, oltre ad una varianza non costante, le seguenti caratteristiche: (i) residui non indipendenti; (ii) errori non-Gaussiani; (iii) un potenziale *bias* nel campionamento dovuto alla distribuzione non causale dei dati mancanti. Le prime tre proprietà menzionate sarebbero riconducibili alla natura stocastica della turbolenza, ad errori strumentali ed alla non uniforme distribuzione delle condizioni ambientali (ad es. un maggior numero di valori con radiazione solare pari a zero rispetto a quelli con radiazione massima) (Hagen *et al.*, 2006).

Riassumendo, data la difficoltà di isolare gli effetti che alcuni fenomeni hanno sulla serie temporale e la non costanza della varianza durante l'anno l'analisi dell'incertezza con il modello additivo è fatta per ogni singolo giorno dell'anno. Per minimizzare l'effetto dell'eteroschedasticità della varianza nel confronto degli addendi del modello additivo, i singoli addendi sono stati normalizzati rispetto alla SS totale giornaliera. Oltre alla minimizzazione dell'effetto dell'eteroschedasticità della varianza la normalizzazione rende il confronto delle incertezze più facile esprimendo tutti i valori in percentuale.

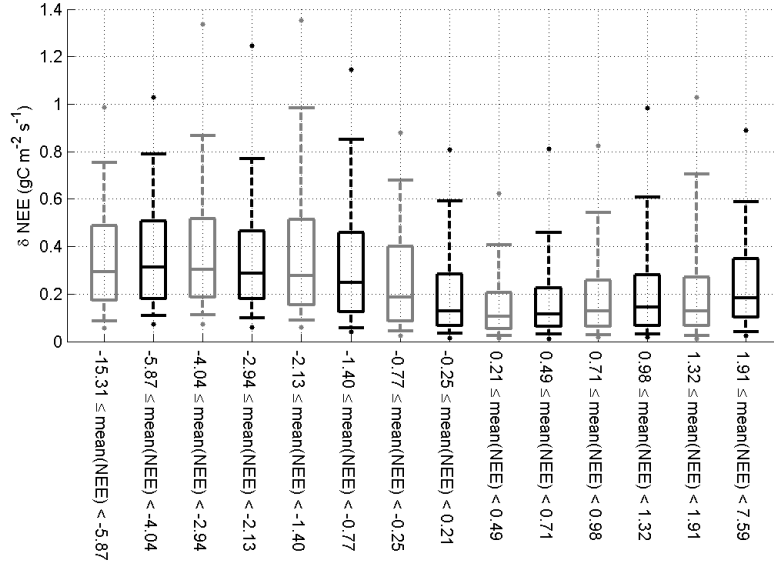


Figura 3.1: Distribuzioni dei valori della deviazione standard (σ^2) rispetto a classi di valori medi di NEE (mean NEE) per le aggregazioni giornaliere di tutti i siti anno.

3.1.1 Verifica del modello

La prima ipotesi che viene fatta per l'analisi ANOVA è quella che il modello adottato e i suoi parametri rappresentino adeguatamente il comportamento dei dati.

La verifica di questa prima ipotesi è fondamentale infatti, se il disegno sperimentale è più complesso di quello ipotizzato per l'analisi fattoriale, allora il modello lineare dovrebbe avere un numero di parametri ed una complessità maggiore. Nel caso in cui i parametri del modello non siano adeguati all'analisi dei dati le conclusioni saranno non valide per la presenza di risultati spuri. Inoltre, se alcuni parametri non sono stati inclusi, allora la somma dei quadrati ad essi associati verrà considerata nel termine ε anche se l'errore non è assolutamente casuale. L'utilizzo di un modello errato può anche pro-

vocare stime distorte dei parametri inclusi nel modello (Freund & Wilson, 2003).

Alla luce di quanto appena riportato risulta chiaro che sono necessari opportuni accorgimenti sia per l'applicazione dell'analisi fattoriale sia nelle fasi di discussione e conclusioni.

3.1.2 Verifica della Normalità

La seconda assunzione dell'analisi della variabilità ipotizza che i termini ε siano variabili casuali normalmente distribuite con media pari a zero e varianza σ^2 .

Questa assunzione è importante per il confronto tra la distribuzione F e la distribuzione del rapporto MS_B / MS_W . Anche se la violazione di questa assunzione non influenza la rappresentatività della distribuzione F rispetto ai rapporti MS_B / MS_W in casi di estrema non normalità, la presenza di outlier può provocare errori infatti le medie possono anche non essere un insieme appropriato di parametri per la descrizione e l'inferenza (Freund & Wilson, 2003).

L'assunzione sulla varianza è necessaria per considerare la varianza dei gruppi come un estimatore della varianza. Se questa ipotesi è vera il rapporto MS_B / MS_W può essere utilizzato come test di verifica. Se le differenze tra le σ^2 sono notevoli allora l'utilizzo delle medie può risultare dubbio (Freund & Wilson, 2003).

Per verificare la normalità della distribuzione dei dati è stato condotto il test del χ^2 . Il test definisce la bontà di adattamento all'ipotesi nulla che i dati sono un campione casuale estratto da una distribuzione normale con la stessa media e varianza stimata dai dati analizzati. Per l'ipotesi alternativa, invece, i dati non sono normalmente distribuiti.

Il risultato del test è 1 se l'ipotesi nulla può essere rifiutata con un 5% di livello di significatività. Il test restituisce come risultato 0 se l'ipotesi nulla non può essere rifiutata. Per il test i valori sono divisi in gruppi e per ciascuno vengono calcolate le frequenze assolute e attese. Fatto questo viene calcolato il χ^2 :

$$\chi^2 = \sum_{i=1}^n \frac{O_i - E_i}{E_i} \quad (3.1)$$

dove O_i rappresenta le frequenze assolute osservate e E_i quelle attese. Se il campione è numeroso la distribuzione può essere approssimata a quella del χ^2 , i gruppi con meno di 5 valori sono accorpati a quelli più vicini fino a quando i gruppi estremi non hanno almeno 5 valori. Per questo test sono stati creati 10 gruppi ed il test confronta le frequenze dei dati osservati con quelle della distribuzione del χ^2 con 7 gradi di libertà.

La figura 3.2 rappresenta le percentuali di risultati del test χ^2 su scala giornaliera. Per ciascun giorno di ogni sito/anno è stato svolto il test per la verifica di Normalità. Nella maggior parte dei casi (77%) i dati giornalieri non mostrano una distribuzione Normale. Tenendo presente questo aspetto i test statistici che assumono la distribuzione Normale dei dati possono essere applicati ma con le dovute cautele. Nel caso dell'analisi fattoriale questo implica tutti i problemi e dubbi esposti in precedenza.

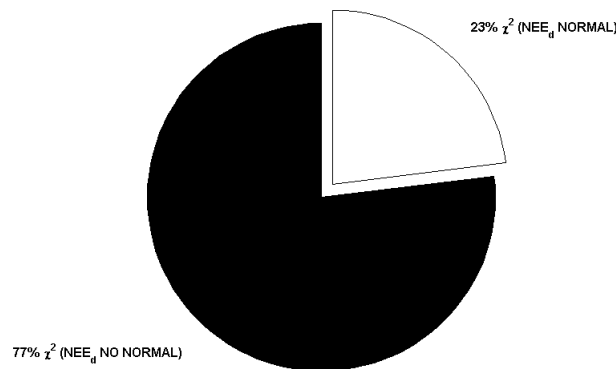


Figura 3.2: Percentuale di dati NEE giornalieri (NEE_d) di tutti i siti anno che hanno una distribuzione Normale secondo il test del χ^2 .

3.1.3 Verifica di Indipendenza

L'ultima assunzione è che i termini ε siano indipendenti in senso probabilistico¹. La verifica di questa ipotesi è importante perché MS_B e MS_W sono

¹un valore di ε non influenza i valori gli altri ε

considerati due estimatori indipendenti per una stessa varianza. La più frequente violazione di questa assunzione si verifica quando le osservazioni sono registrate in un certo tempo o spazio e misure adiacenti tendono ad essere collegate tra loro (Freund & Wilson, 2003).

I dati EC sembrerebbero violare l'ipotesi di indipendenza infatti, avendo in comune il valore misurato e una serie di passaggi di processamento, non sono indipendenti.

Per le analisi svolte con l'adozione della scala temporale giornaliera ed l'analisi di ciascun valore di quest'ultima si ipotizza che l'influenza tra i dati adiacenti sia trascurabile.

3.1.4 Modello additivo finale

Alla luce delle motivazioni delle sezioni precedenti si sconsiglia l'utilizzo della ANOVA nell'analisi fattoriale e nella stima del livello di significatività delle differenze delle medie. Per stimare l'incertezza legata al post processing sarà applicato un modello additivo che consente di valutare quantitativamente l'influenza di tutti i fattori e loro interazioni. Il vantaggio del modello additivo è legato soprattutto al minor numero di assunzioni necessarie per la sua applicazione e per la verifica dei risultati.

Il modello additivo applicato sarà:

$$\begin{aligned}
 y_{ijkz} = & \mu + \tau_i + \beta_j + \gamma_k + \rho_z + \\
 & + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + \\
 & + (\tau\rho)_{iz} + (\beta\gamma)_{jk} + (\beta\rho)_{jz} + (\gamma\rho)_{kz} + \\
 & + (\tau\beta\gamma)_{ijk} + (\tau\beta\rho)_{ijz} + (\tau\gamma\rho)_{ikz} + (\beta\gamma\rho)_{jkz} + \\
 & + (\tau\beta\gamma\rho)_{ijkz} + \varepsilon_{ijkz} \left\{ \begin{array}{l} i = 1, 2, 3 \\ j = 1, 2 \\ k = 1, 2 \\ z = 1, 2 \end{array} \right. \quad (3.2)
 \end{aligned}$$

in questo modello esistono 4 fattori principali (legati ai 4 fattori); 6 interazioni di primo livello (2 fattori); 5 interazione interazioni di secondo livello (3 fattori); 1 interazione di terzo livello (4 fattori) e il termine di errore.

Nella formula:

τ rappresenta il despiking con i che va da 1 a 3 (3 soglie: 4, 5.5 e 7) (sezione 2.3, pagina 23);

β rappresenta la metodologia per il calcolo della soglia di u^* con j che va da 1 a 2 (2 metodi per il calcolo della soglia: Reichstein (sezione 2.4.1, pagina 26) e Barr (sezione 2.4.2, pagina 26);

γ rappresenta l'applicazione della soglia di u^* con k che varia tra 1 e 2 (2 possibili applicazioni della soglia di u^*)

ρ rappresenta i metodi di gap filling o di partitioning con z che varia tra 1 e 2 (due metodi di gap filling o due di partitioning).

Dei 3 metodi di partitioning applicati la stima dell'incertezza con il modello additivo ne considera solo due: Reichstein (sezione 2.7.1), Lasslop (sezione 2.7.2). Il metodo proposto da van Gorsel *et al.* (2007) (sezione 2.7.3) non utilizza tutti i fattori e, di conseguenza, non rientra nell'analisi fattoriale in cui tutti i livelli devono incrociarsi tra loro.

3.2 Quantificazione dell'incertezza

Le figura 3.3(a) e 3.3(b) mostrano le PDF dei rapporti tra tutti i termini del modello additivo e la SS totale rispettivamente per la variabile NEE_d ed $Reco_d$. Dall'analisi dei boxplot è evidente che per la variabile NEE_d l'incertezza globale è essenzialmente dovuta a 2 singoli fattori: (i) la metodologia di gap filling adottata (GF); (ii) metodologia per il calcolo della soglia di u^* (uTH). Il contributo del despiking e del criterio di applicazione della soglia di u^* sembrano avere un'importanza è molto inferiore. Analizzando le interazioni è evidente che in tutti i casi in cui più fattori interagiscono il loro contributo è molto basso (al di sotto del 10%), unica lieve eccezione è per l'interazione uTH * GF in cui il 90-esimo percentile è superiore al 10%.

Per la variabile $Reco_d$ il fattore che maggiormente influenza l'incertezza globale è la metodologia di partitioning (PT) la cui mediana è quasi al 50%. In questo caso l'incertezza ha valori estremamente alti sia rispetto ai percentili della $Reco_d$ sia rispetto a tutti i boxplot della figura 3.3(a). Altro fattore che contribuisce, in maniera evidente anche se molto ridotta rispetto al PT,

all'incertezza è, come in precedenza, la metodologia di soglia di u^* . Anche per la Reco_d le interazioni tra fattori hanno un contributo al di sotto del 10% con unica eccezione per l'interazione $u\text{TH}^*\text{PT}$ in cui, nel 90-95% dei casi, si raggiungono valori superiori al 25%.

3.3 Incertezza intrinseca

Dai risultati ottenuti è evidente che in tutti gli schemi di processamento applicati l'incertezza globale è dovuta principalmente dai fattori soglia di $u^*(u\text{TH})$, metodi di gap filling (GF) e metodi di partitioning (PT). Questi fattori, in realtà, non sono totalmente indipendenti tra loro ma, come si dimostrerà nelle sezioni successive, hanno una certa dipendenza. Dato il notevole impatto che questi fattori hanno sull'incertezza globale dei dati EC è necessario valutare la loro incertezza intrinseca.

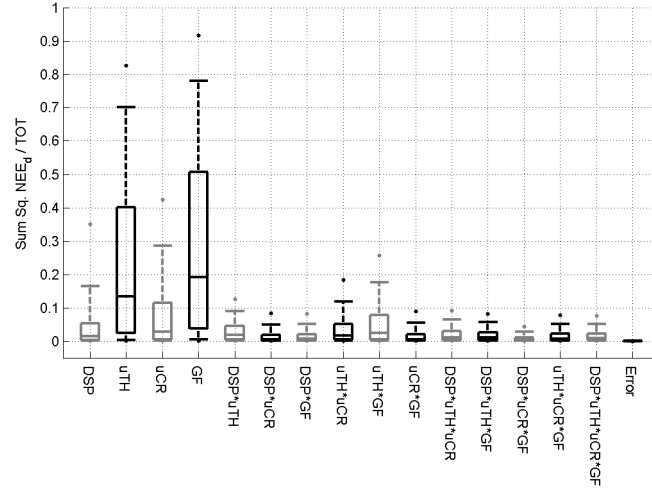
3.3.1 Incertezza intrinseca - $u\text{TH}$

Per quel che riguarda il fattore $u\text{TH}$ oltre alla metodologia adottata per definirne il valore è ulteriormente interessante stimare l'effetto che la sua incertezza intrinseca può avere sull'incertezza globale.

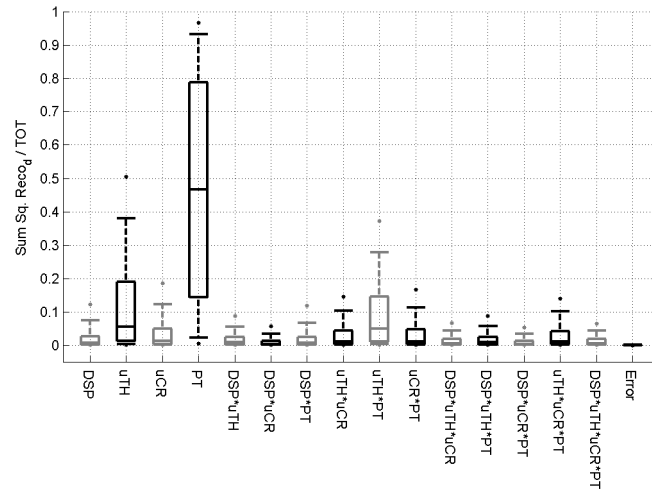
Per valutare l'incertezza intrinseca del fattore $u\text{TH}$ si è adottato un unico schema di post processamento tra tutti quelli possibili mantenendo fissi i livelli di ogni fattore:

- soglia di despiking: $z = 5.5$;
- metodologia per il calcolo della soglia di u^* : fw2;
- eliminazione dei soli valori al di sotto della soglia di u^* ;
- metodo di gap filling: MDS.

Per definire l'incertezza della soglia di u^* la procedura di calcolo è ripetuta 100 volte secondo una tecnica di bootstrapping applicata ai dati di un anno intero. Alla fine del bootstrapping, avendo 100 soglie di u^* , è possibile calcolare alcuni percentili (1, 5, 10, 25, 50, 75, 90, 95, 99). In genere di tutti i percentili calcolati, per soglia di u^* si considera il valore in corrispondenza della mediana (50-esimo percentile). Per valutare l'incertezza intrinseca



(a)



(b)

Figura 3.3: PDF dei rapporti tra gli addendi del modello additivo e la SS totale per tutti i siti/anno insieme. L'asterisco indica l'interazione. (a) Reco_d ; (b) NEE_d . DSP indica il fattore despikes; uTH il fattore metodologia per la soglia di ustar; uCR la modalità di applicazione della soglia di ustar; GF metodologia di gap filling; PT metodologia di partitioning.

della soglia di u^* , mantenendo fisso lo schema riportato in precedenza, sono stati utilizzati, come valore soglia, i diversi percentili.

La tabella 3.1 riporta i percentili delle soglie di u^* per due foreste di aghifoglie sempreverdi (ENF), Hyytiälä e Loobos. Dai valori riportati in tabella è possibile notare che, come già detto in precedenza, i valori dei percentili non sono, per uno stesso sito/anno, molto diversi tra loro.

percentili	FIHyy2007	FIHyy2008	NLLoo2008	NLLoo2009
1th	0.3416	0.332063	0.111967	0.0815667
5th	0.390109	0.39006	0.116011	0.1104
10th	0.456178	0.405865	0.12292	0.118667
25th	0.491881	0.441843	0.136825	0.13541
50th	0.556392	0.480718	0.147638	0.149133
75th	0.582176	0.530622	0.179575	0.1631
90th	0.611213	0.566378	0.221857	0.1797
95th	0.6425	0.61464	0.23165	0.1897
99th	0.710485	0.690144	0.288375	0.1981

Tabella 3.1: percentili delle soglie di u^* definite utilizzando il metodo fw2. La soglia di despiking è impostata a 5.5.

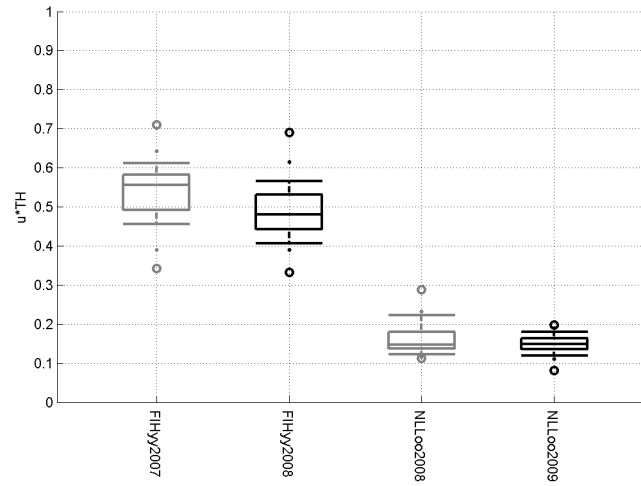
Confrontando le soglie di u^* riportate dai boxplot in figura 3.4(a) per il sito di Hyytiälä e Loobos è evidente che, anche se entrambe i siti sono classificati nella stessa PFT, i valori delle soglie sono estremamente diversi. Per il sito finlandese (Hyytiälä) tutti i percentili sono superiori a 0.3 mentre per Loobos le soglie sono tutte inferiori a tale valore. Questo risultato potrebbe essere riconducibile alle caratteristiche stazionali o alle caratteristiche del sistema di misura adottato. La tipologia di soprassuolo dei due siti è molto simile infatti nel caso di Hyytiälä, si tratta di una pineta di *Pinus sylvestris* di circa 41 anni mentre la copertura vegetale di Loobos è anch'essa una pineta di *Pinus sylvestris* ma sono presenti anche Pino nero, Betulla e Douglasia.

L'impatto che le soglie di u^* hanno sull'incertezza è rappresentato dalla figura 3.4(b). L'incertezza legata ai percentili delle soglie di u^* è stata quantificata come differenza rispetto al valore annuale di NEE ottenuto con il 50-esimo percentile. Anche se le soglie di u^* non sono estremamente diverse

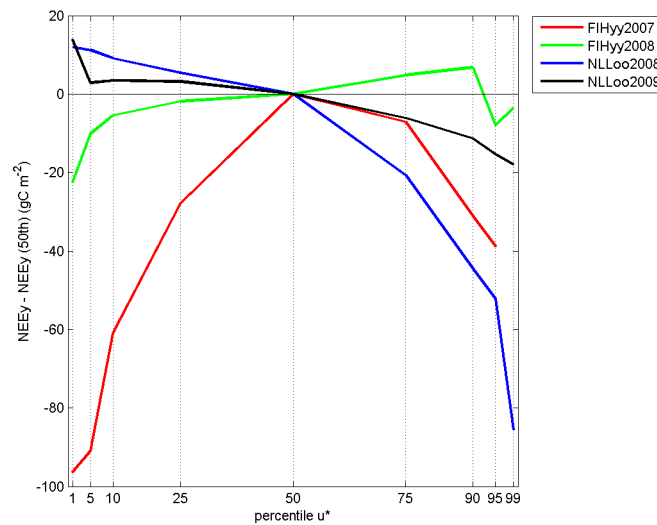
tra loro l'impatto che hanno sull'incertezza globale è veramente notevole. Ad esempio nel caso di Hyytiälä anno 2007 scegliere come soglia di u^* il primo o il quinto percentile comporta una differenza, rispetto al valore annuo di riferimento, di circa -100 gC m^{-2} . Prendendo i percentili diametralmente opposti (95-esimo e 99-esimo) le differenze sono notevoli se non addirittura critiche. Infatti scegliendo come soglia il 95-esimo percentile la differenza rispetto al valore annuale di riferimento è di circa -40 gC m^{-2} mentre con la soglia del 99-esimo percentile non è possibile applicare il gap filling con il metodo MDS per la mancanza di troppi dati misurati. Analizzando i risultati dello stesso sito ma per l'altro anno, 2008, è interessante notare che la scelta del percentile ha un impatto compreso tra -20 e $+10 \text{ gC m}^{-2}$. La differenza tra le incertezze può essere ricondotta all'effetto della variabilità interannuale. In effetti per il sito di Hyytiälä l'anno 2007 ha delle soglie di u^* più alte rispetto a quelle dell'anno 2009 (figura 3.4(a)). Queste differenze influiscono sul grado di incertezza che, infatti, è superiore per l'anno 2007 rispetto al 2008.

Per quanto riguarda la pineta di Loobos i risultati sono leggermente diversi infatti, per i percentili inferiori alla mediana, le differenze sono sempre positive (comprese tra $+18$ e $+8 \text{ gC m}^{-2}$) mentre per i percentili superiori le differenze sono sempre negative (tra -20 e -80 gC m^{-2}). Come già mostrato in figura 3.4(a), le soglie di u^* di Loobos sono molto inferiori rispetto a quelle del sito di Hyytiälä, questo potrebbe indurre a pensare che l'incertezza sia ridotta. In realtà questa supposizione può essere fuorviante infatti, come è possibile vedere dai risultati mostrati in figura 3.4(b) l'anno 2008 e l'anno 2009 hanno, nel caso del 99-esimo percentile, una incertezza molto diversa pari, rispettivamente, a -20 e -80 gC m^{-2} . Questa notevole differenza è dovuta alla notevole differenza dei valori del percentile infatti il 99-esimo percentile è pari a 0.3 m s^{-1} per l'anno 2008 mentre per il 2009 è 0.2 m s^{-1} . In sintesi una differenza interannuale della soglia di u^* pari a 0.1 m s^{-1} implica un'incertezza della NEE annuale pari a circa 60 gC m^{-2} .

Alla luce di quanto finora esposto è chiaro che, anche applicando lo stesso schema di processamento, la scelta della soglia di u^* ha un notevole impatto sull'incertezza globale.



(a)



(b)

Figura 3.4: (a) PDF delle soglie di u^* delle foreste di aghifoglie sempreverdi (ENF). I cerchi indicano il primo ed il 99-esimo percentile. (b) Differenze tra i valori annuali di NEE calcolati con diversi percentili di soglie di u^* . Le differenze sono calcolate rispetto al valore annuale di NEE ottenuto applicando il 50-esimo percentile delle soglie di u^* .

3.3.2 Incertezza intrinseca - GF

Come già riportato in precedenza i metodi di gap filling sono delle procedure di interpolazione che cercano di riempire i dati mancanti nelle serie temporali di dati EC. Come per tutte le procedure d'interpolazione, anche nel caso dei dati EC, maggiore è il numero di dati mancanti minore sarà l'affidabilità del risultato.

Per poter dimostrare questa affermazione è stata utilizzata la serie temporale di dati di un sito anno (NLLoo 2009) ed è stato applicato un unico schema di processamento. I filtri applicati sono:

- soglia di despiking: $z = 5.5$;
- metodologia per il calcolo della soglia di u^* : fw2;
- eliminazione dei soli valori al di sotto della soglia di u^* ;

Oltre all'eliminazione di valori in base allo schema di filtraggio, prima del passaggio GF, sono stati inseriti dei gap artificiali. I gap artificiali hanno una lunghezza crescente (dal 5% al 50% dei dati misurati) e sono posizionati in maniera casuale (20 posizioni per ciascuna lunghezza) all'interno della serie temporale. La lunghezza crescente dei gap artificiali serve per vedere l'effetto che la riduzione di dati disponibili per il gap filling ha sulle performance di riempimento. Il posizionamento casuale è motivato dalla stretta relazione esistente tra performance del GF e posizione temporale dei dati mancanti. Se i dati mancanti sono concentrati durante la stagione non vegetativa la loro interpolazione sarà più semplice a causa della ridotta variabilità, al contrario, se i dati mancanti ricadono nella stagione vegetativa, l'interpolazione sarà più difficile per la maggior variabilità dei flussi.

Per stimare le performance del GF i dati simulati e quelli misurati, tolti artificialmente, sono messi a confronto. La quantificazione delle performance è data dal coefficiente di correlazione (r^2) e dall'errore quadratico medio assoluto (RMSEabs). Le figure 3.6(a) e 3.6(b) rappresentano l'andamento delle performance in funzione della lunghezza dei gap artificiali. Quando il gap artificiale è breve (5%) la correlazione tra dati misurati e simulati è, per più del 25%, superiore a 0.77. Man mano che i gap artificiali si allungano la correlazione scende ad esempio, per i gap artificiali che eliminano il 50% dei

dati misurati, tutti i valori sono al di sotto del 0.78. Per gap brevi (5%) il 75% degli esperimenti ha valori di RMSEabs inferiori a $3.65 \mu\text{mol m}^{-2} \text{s}^{-1}$ mentre per i gap con lunghezza pari al 50% l'errore è, in tutti i casi, superiore a $3.65 \mu\text{mol m}^{-2} \text{s}^{-1}$. Un altro aspetto da mettere in evidenza è quello delle distanze tra i percentili, quando i gap sono brevi la PDF dell'errore e della correlazione sono molto ampie ma si riducono progressivamente con l'aumento del numero di dati mancanti. Questo fenomeno potrebbe essere riconducibile all'effetto che la posizione del gap artificiale ha sulle performance, se il gap breve elimina pochi dati nel periodo vegetativo la performance sarà peggiore rispetto ad un gap con la medesima lunghezza ma che elimina i dati del periodo non vegetativo. Nel caso dei gap molto lunghi (50%) la probabilità che i dati mancanti possano ricadere nella stagione vegetativa è molto più alta.

Un ulteriore aspetto da tener presente è anche la strategia di gap filling alla base del metodo applicato. Alla fine del GF la serie temporale sarà sempre completa ma, nel caso di molti dati misurati mancanti, con una percentuale molto alta di dati gap filled di scarsa qualità. Nel caso del metodo MDS i dati di scarsa qualità sono quelli calcolati con la variazione media diurna (MDV). Avendo pochi dati da mediare le medie diurne non rispecchieranno né la variabilità né l'entità dei flussi misurati. Inoltre dato che la media è calcolata con un ridotto numero di dati le medie saranno, molto probabilmente, identiche tra loro. I filtri applicati che riducono maggiormente la numerosità dei dati sono quelli per il calcolo della soglia di u^* . La figura 3.7 rappresenta le PDF del numero di dati rimanenti, rispetto a tutti quelli misurati, dopo l'applicazione di alcuni dei filtri del post processamento. Per questa figura la soglia di u^* è calcolata solo con la metodologia fw2. L'effetto che le soglie di despiking hanno sulla numerosità è molto ridotto infatti in tutti i casi i valori sono superiori allo 0.9 o addirittura, come nel caso di $z = 7$, il despiking sembra non togliere nessun valore. Applicando la soglia di u^* nelle due modalità, eliminazione dei valori inferiori alla soglia oppure eliminazione di quelli sotto la soglia e del primo valore seguente ad un periodo di bassa turbolenza, il numero di dati rimanenti dopo il filtraggio è molto ridotto. Ad esempio, nel caso di despiking $z = 7$ ed eliminazione dei valori al di sotto della soglia di u^* , la percentuale di dati presenti passa da 1 a circa 0.65 per il 75% dei

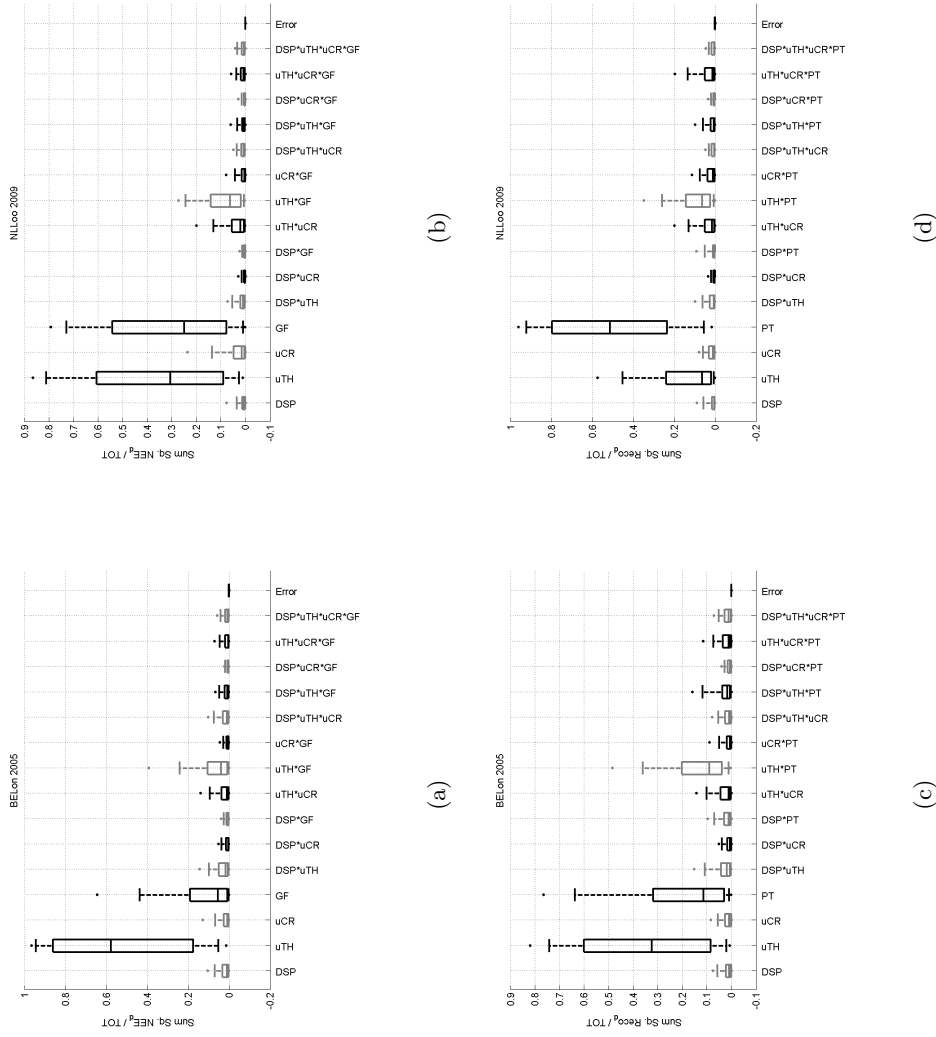


Figura 3.5: PDF dei rapporti tra gli addendi del modello additivo e la devianza totale. (a) e (c) per flussi di NEE_d ed $Reco_d$ per il sito/anno di Lonzee 2005; (b) e (d) per il sito/anno di Lonzee 2009. DSP indica il fattore despike; uTH il fattore metodologia per la soglia di ustar; uCR la modalità di applicazione della soglia di ustar; GF metodologia di gap filling; PT metodologia di partitioning.



(b)

Alla luce di questi risultati sembra verificata l'ipotesi che il peso del fattore GF, nell'incertezza globale del processamento dei dati EC, sia una conseguenza del valore della soglia di u^* .

3.4 Incertezza dei dati EC e Relazioni Funzionali

Dal punto di vista applicativo molto spesso i dati EC sono utilizzati sia per la parametrizzazione che per la validazione di modelli. Questo aspetto ha una valenza ecologica fondamentale perché analizzando i risultati di un modello

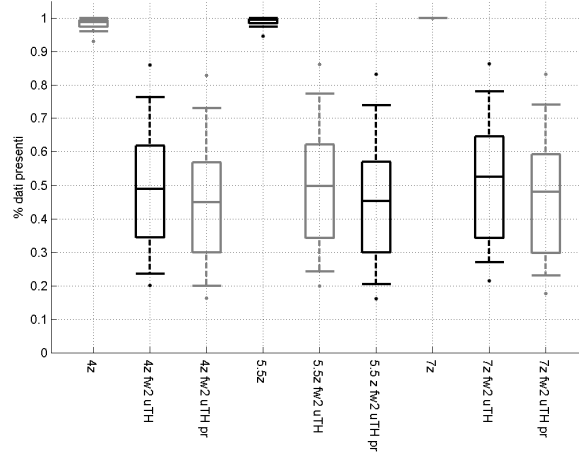


Figura 3.7: PDF del numero di dati rimanenti, rispetto a quelli iniziali, dopo l'applicazione dei filtri dello schema di post processamento. 4 SD, 5.5 SD ed 7 SD rappresentano le 3 soglie di despiking. fw2 il metodo di Reichstein per definire la soglia di u^* ; uTH rappresenta l'eliminazione dei valori inferiori alla soglia di u^* ; uTHpr all'eliminazione dei valori inferiori alla soglia di u^* ed anche di quelli successivi al periodo di bassa turbolenza.

o i parametri dello stesso è possibile capire il funzionamento biologico di molti processi.

Un esempio di applicazione delle misure EC per studi ecologici è quello delle curve di luce. Le curve di luce esprimono, in forma matematica, il funzionamento biologico della fotosintesi e, se parametrizzate con i dati EC, ad una scala spaziale di ecosistema. La formula della curva di luce (Michaelis & Menten, 1913; Falge *et al.*, 2001a) è:

$$NEE = \frac{\alpha \beta R_g}{\alpha R_g + \beta} + \gamma \quad (3.3)$$

dove α ($\mu\text{mol m}^{-2} \text{s}^{-1}$) è la pendenza iniziale della curva di luce; β ($\mu\text{mol m}^{-2} \text{s}^{-1}$) rappresenta il valore massimo di assorbimento di CO_2 a saturazione di luce, γ ($\mu\text{mol m}^{-2} \text{s}^{-1}$) è l'intercetta della curva con l'asse Y; R_g è la radiazione globale (W m^{-2}). In sintesi la curva di luce è un modello funzionale che mette in relazione gli scambi di CO_2 e la radiazione globale.

Ogni parametro dell'equazione 3.3 ha un preciso significato ecologico infatti

α rappresenta l'efficienza quantica della chioma; β indica l'andamento della fotosintesi a saturazione di luce; γ rappresenta la respirazione. Approfondendo ulteriormente la valenza ecologica dei parametri della curva di luce è possibile estrarre ulteriori conclusioni.

Dalla forma del modello e dalla definizione dei parametri è possibile capire che il processo biologico di fotosintesi non è sempre crescente all'aumentare della radiazione globale infatti, oltre una certa soglia di flusso radiativo, l'assorbimento rimane costante (β). Un altro aspetto interessante è quello legato al parametro γ infatti, quando il flusso radiativo è molto basso, la pianta non assorbe CO₂ ma, se l'intercetta è positiva, emettere CO₂ in atmosfera. Il parametro α , esprimendo l'efficienza quantica, consente di quantificare l'efficienza dell'ecosistema nell'utilizzo l'energia luminosa per convertire la CO₂ in biomassa.

Considerato che la capacità di fotosintesi varia durante l'anno l'applicazione della curva di luce è fatta per finestre temporali ridotte. Anche in questo caso il confronto dei parametri in diversi periodi dell'anno consente di quantificare le modificazioni biologiche dell'ecosistema.

Dato che il modello può essere parametrizzato con i dati EC, è interessante indagare quale potrebbe essere l'impatto che l'incertezza legata al post processamento può avere su questa relazione funzionale. Tenendo presente la variabilità annuale dei flussi misurati la parametrizzazione è stata fatta considerando finestre temporali di lunghezza pari a 15 giorni. I dati NEE utilizzati per la parametrizzazione sono derivati da tutti gli schemi di processamento possibili. Applicando le stesse finestre temporali e gli stessi valori di radiazione globale l'incertezza dei parametri è dovuta principalmente ai valori di NEE. Le figure 3.8(a), 3.8(b) e 3.8(c) rappresentano le incertezze dei parametri della curva di luce rispettivamente per α , β e γ per diverse PFT. Per rendere confrontabili sia i valori dei parametri all'interno di uno stesso anno sia tra siti/anno di una stessa PFT, i valori originali dei parametri

sono stati normalizzati secondo le formule:

$$\alpha_{Norm\ ij} = \frac{\alpha_{ij}}{\bar{\alpha}_i} \quad (3.4)$$

$$\beta_{Norm\ ij} = \frac{\beta_{ij}}{\bar{\beta}_i} \quad (3.5)$$

$$\gamma_{Norm\ ij} = \frac{\gamma_{ij}}{\bar{\gamma}_i} \quad (3.6)$$

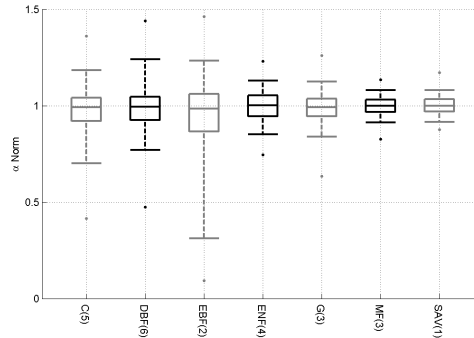
dove l'indice i rappresenta la finestra di 15 giorni; j indica lo schema di processamento per i valori di NEE; la barra orizzontale indica la media di tutti gli j -esimi parametri all'interno della i -esima finestra. Le PDF dei parametri (figure 3.8(a), 3.8(b) e 3.8(c)) mostrano che le incertezze dei flussi di NEE derivanti dai diversi schemi di processamento dei dati EC hanno un notevole impatto sui parametri.

Prendendo come esempio il parametro α per la PFT dei coltivi (cropland, C), questo può differire dal valore medio da 1.5 a 0.5 volte. L'interpretazione di questi valori in chiave ecologica potrebbe far pensare che, per una qualsiasi causa esclusivamente biologica, gli ecosistemi C abbiano aumentato (o diminuito) la capacità di fotosintesi di circa il 50% rispetto al valore medio. L'incertezza dei dati EC sembra influenzare soprattutto il parametro β infatti il valore di saturazione può essere superiore di 1.4 o ridursi a 0.1 volte rispetto al valore medio. Anche la componente γ risente, maniera molto ridotta, dell'incertezza dei dati EC infatti per questo parametro i valori oscillano sempre intorno ad un $\pm 20\%$, con l'unica eccezione per i siti anno classificati come foreste sempreverdi di latifoglie (EBF) in cui l'incertezza di γ varia da 1.8 a 0.4 volte il valore medio.

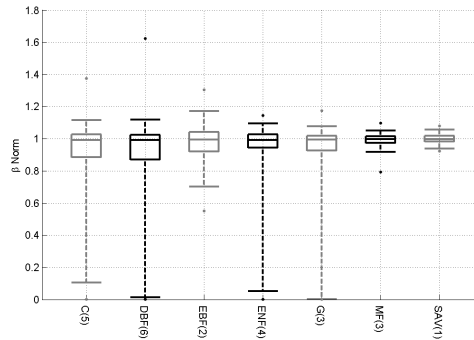
Dai risultati ottenuti è evidente che l'incertezza del post processamento contribuisce all'incertezza dei parametri delle curve di luce. Questa influenza potrebbe rendere difficile sia l'interpretazione dei parametri del modello sia, soprattutto, l'applicazione del modello per la stima del bilancio del carbonio.

3.5 Quantificazione dell'incertezza Annuale

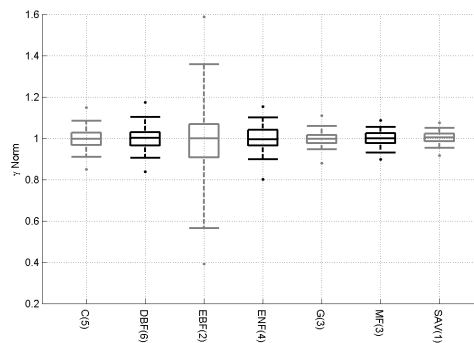
L'oggetto di quest'ultima sezione è quello di riportare sinteticamente i risultati ottenuti focalizzando l'attenzione soprattutto nel definire quantitativa-



(a)



(b)



(c)

Figura 3.8: PDF dei rapporti tra i parametri della curva di luce ed il valore medio. (a) parametro α ; (b) parametro β ; (c) parametro γ . Le sigle sull'asse X rappresentano: coltivi (Cropland, C); pascoli (Grassland, G); boschi misti (mixed forest, MF); foreste di latifoglie decidue (Deciduous Broadleaf Forest, DBF); foreste di aghifoglie sempreverdi (Evergreen Needleleaf Forest, ENF); savana (Dehesa) (savanna, S); foreste di latifoglie sempreverdi (Evergreen Broadleaf Forest, EBF).

mente l'incertezza annuale associata ai diversi schemi di post processamento dei dati EC. Stimare l'incertezza è un aspetto fondamentale sia per le numerose applicazioni che questi dati hanno negli studi in campo ecologico sia per le stime del bilancio del carbonio degli ecosistemi terrestri.

La quantificazione dell'incertezza dei valori annuali è calcolata seguendo questa metodologia. Per ciascun sito/anno il valore annuale di NEE (NEEy) del i -esimo schema di processamento è pari a:

$$NEEy_i = \sum_{j=1}^d NEE_{d\ ij} \quad (3.7)$$

similmente per la Reco:

$$Recoy_i = \sum_{j=1}^d Reco_{d\ ij} \quad (3.8)$$

dove la sommatoria rappresenta l'aggregazione di tutti i j -esimi valori giornalieri presenti nell'anno dell' i -esimo schema di processamento. Per confrontare i valori annuali sia tra i diversi schemi di processamento, sia per diversi anni ma anche per confronti inter-PFT e intra-PFT è necessario standardizzare i valori calcolati secondo un criterio comune. Il fattore di standardizzazione adottato è la media di tutti i valori annuali calcolati per ciascun sito/anno quindi:

$$NEEy_{media} = \frac{\sum_{i=1}^m NEEy_i}{m} \quad (3.9)$$

oppure:

$$Recoy_{media} = \frac{\sum_{j=1}^n Reco_{y_j}}{n} \quad (3.10)$$

dove m indica tutti gli i -esimi schemi di post processamento possibili per la NEE mentre n rappresenta tutti i j -esimi schemi possibili per la Reco. L'incertezza rispetto alla media è calcolata sia come rapporto tra i singoli valori annuali ed il fattore di standardizzazione sia come differenza tra questi due termini. In entrambe i casi i valori di ciascun sito/anno risultano standardizzati quindi è possibile aggregarli in base alla PFT e avere una stima dell'incertezza non per singolo anno ma generalizzata per PFT. Con il calcolo del rapporto è possibile avere un valore percentuale dell'incertezza mentre con le differenze l'incertezza è espressa quantitativamente.

Combinando tutti i risultati degli i -esimi schemi di processamento per la NEE e dei j -esimi schemi per la Reco è possibile calcolare la GPP applicando la formula:

$$GPP_{ij} = Reco_j - NEE_i \begin{cases} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{cases} \quad (3.11)$$

dove m ed n rappresentano tutti gli schemi possibili per NEE ed Reco. Le tabelle 3.2, 3.3 e 3.4 mostrano l'incertezza dei dati annuali di NEE, Reco e GPP calcolata sia quantitativa che in percentuale. I valori delle tabelle sono rappresentati graficamente nelle figure 3.9 3.10, 3.11 3.12 e 3.13 3.14. Dai percentili riportati in figura 3.9 è possibile notare che tutti i 25-esimi e 75-esimi quantili sono compresi tra 0.9 e 1.1 quindi la distanza dei quartili è, in generale, pari al 20%. La PFT che manifesta la maggior variabilità è quella delle G (con il decimo e novantesimo percentile pari a 0.7 e 1.2 volte il valor medio), di contro la variabilità più ridotta è quella delle MF. Passando all'analisi delle differenze tra le NEE annuali ed il valore medio (figura 3.10) si nota che le PDF delle variabilità non sono analoghe a quelle viste in figura 3.10 infatti, nel caso delle differenze, la variabilità maggiore viene riscontrata per le ENF con una differenza interquartile pari a circa 100 gC m⁻² (-50 gC m⁻² per il 25-esimo percentile e +50 gC m⁻² per il 75-esimo). Nel range delle ENF (± 50 gC m⁻²) sono compresi tutti i quartili delle altre PFT. Per i rapporti $Reco_y / Reco_{media}$ la maggior variabilità si manifesta per le EBF con una distanza tra i quartili di circa 0.25 (1.1 - 0.85) mentre per le altre PFT è compresa tra 1.1 e 0.9 volte il valore medio. Le conclusioni tratte dall'analisi dell'incertezza percentuale sembrano trovare conferma anche nell'analisi dell'incertezza quantitativa (figura 3.12) infatti anche in questo caso le EBF hanno il maggior grado di variabilità dell'incertezza. Nel caso dell'incertezza quantitativa i quartili di molti siti/anno sono compresi tra +100 e -100 gC m⁻² con l'unica eccezione delle foreste sempreverdi di latifoglie in cui il 25% dei valori è prossimo a -200 gC m⁻². L'incertezza della NEE_y e della Reco_y si combinano con il calcolo la GPP. L'incertezza percentuale della GPP (figura 3.13) sembra anch'essa compresa tra 0.9 e 1.1 volte il valore medio nella maggior parte dei casi infatti, in quasi tutte le PFT considerate, il decimo e novantesimo percentile sono all'interno di

PFT	NEE _y - NEE _{y_{media}} (gC m ⁻²)							
	5 th (%)	10 th (%)	25 th (%)	50 th (%)	75 th (%)	90 th (%)	95 th (%)	
C (5)	-25.25 (0.72)	-20.41 (0.84)	-11.75 (0.95)	-2.54 (1.01)	10.83 (1.04)	26.14 (1.09)	32.34 (1.12)	
DBF (6)	-52.85 (0.82)	-40.70 (0.89)	-20.56 (0.95)	0.30 (1.00)	21.25 (1.05)	37.35 (1.13)	53.20 (1.18)	
EBF (2)	-149.35 (0.72)	-112.65 (0.79)	-48.66 (0.93)	-2.35 (1.00)	40.69 (1.09)	117.49 (1.20)	158.76 (1.27)	
ENF (4)	-65.42 (0.59)	-45.24 (0.85)	-26.11 (0.92)	-10.35 (1.04)	23.28 (1.07)	71.57 (1.11)	84.63 (1.22)	
G (3)	-37.90 (0.51)	-30.53 (0.70)	-15.29 (0.92)	-5.02 (1.02)	17.71 (1.08)	44.62 (1.21)	116.03 (1.26)	
MF (3)	-19.44 (0.95)	-13.63 (0.96)	-7.02 (0.98)	0.61 (1.00)	8.63 (1.02)	15.16 (1.03)	20.37 (1.04)	
SAV (1)	-16.99 (0.90)	-15.21 (0.91)	-2.42 (0.92)	1.08 (0.99)	6.77 (1.03)	7.56 (1.17)	8.41 (1.19)	

Tabella 3.2: Incertezza globale per la NEE annuale espressa come rapporto tra i valori annuali e le medie di ogni sito/anno. I numeri tra parentesi vicino alle PFT indicano quanti siti/anno sono stati utilizzati per i calcoli. I numeri tra parentesi indicano il rapporto $NEE_y / NEE_{y_{media}}$.

PFT	Recoy / Recoymedia (gC m ⁻²)						
	5 th (%)	10 th (%)	25 th (%)	50 th (%)	75 th (%)	90 th (%)	95 th (%)
C (5)	-110.58 (0.90)	-97.13 (0.91)	-57.48 (0.95)	-0.86 (1.00)	45.01 (1.04)	102.41 (1.09)	111.80 (1.11)
DBF (6)	-140.69 (0.86)	-103.74 (0.90)	-58.07 (0.95)	-9.15 (0.99)	55.99 (1.04)	116.62 (1.11)	155.96 (1.15)
EBF (2)	-455.99 (0.59)	-299.18 (0.73)	-172.82 (0.83)	44.39 (1.04)	105.66 (1.10)	331.29 (1.30)	388.11(1.35)
ENF (4)	-293.81 (0.76)	-223.82 (0.83)	-100.31 (0.92)	11.54 (1.01)	93.73 (1.09)	186.08 (1.16)	217.83 (1.22)
G (3)	-164.27 (0.89)	-142.10 (0.90)	-64.78 (0.96)	11.93 (1.01)	65.65 (1.04)	137.96 (1.09)	179.20 (1.12)
MF (3)	-122.51 (0.89)	-92.49 (0.92)	-71.95 (0.94)	-44.76 (0.96)	62.73 (1.05)	140.12 (1.12)	182.00 (1.16)
SAV (1)	-118.23 (0.84)	-40.19 (0.95)	-16.17 (0.98)	10.44 (1.01)	35.05 (1.05)	37.37 (1.05)	38.33 (1.05)

Tabella 3.3: Incertezza globale per la Reco annuale espressa come rapporto tra i valori annuali e le medie di ogni sito/ anno. I numeri tra parentesi vicino alle PFT indicano quanti siti/anno sono stati utilizzati per i calcoli. I numeri tra parentesi indicano il rapporto Recoymedia / Recoymedia.

PFT	GPP _y / GPP _{y_{media}} (gC m ⁻²)						
	5 th (%)	10 th (%)	25 th (%)	50 th (%)	75 th (%)	90 th (%)	95 th (%)
C (5)	-116.08 (0.90)	-98.62 (0.93)	-55.10 (0.96)	-2.46 (1.00)	50.75 (1.03)	100.77 (1.08)	117.70 (1.10)
DBF (6)	-149.94 (0.92)	-113.99 (0.93)	-65.48 (0.96)	-8.87 (1.00)	62.30 (1.04)	121.24 (1.07)	160.94 (1.09)
EBF (2)	-471.46 (0.72)	-338.58 (0.80)	-146.75 (0.91)	30.44 (1.02)	122.91 (1.08)	337.00 (1.20)	424.75 (1.25)
ENF (4)	-284.78 (0.83)	-237.98 (0.87)	-108.17 (0.93)	11.81 (1.01)	97.39 (1.07)	197.39 (1.12)	247.94 (1.18)
G (3)	-185.40 (0.89)	-158.27 (0.91)	-90.18 (0.95)	-5.77 (1.00)	76.81 (1.04)	151.58 (1.09)	194.89 (1.12)
MF (3)	-122.45 (0.93)	-96.51 (0.94)	-74.23 (0.96)	-40.55 (0.98)	62.08 (1.04)	135.48 (1.08)	184.48 (1.11)
SAV (1)	-117.07 (0.86)	-43.19 (0.95)	-20.33 (0.98)	10.43 (1.01)	31.22 (1.04)	39.76 (1.05)	48.35 (1.06)

Tabella 3.4: Incertezza globale per la GPP annuale espressa come rapporto tra i valori annuali e le medie di ogni sito/anno. I numeri tra parentesi vicino alle PFT indicano quanti siti/anno sono stati utilizzati per i calcoli. I numeri tra parentesi indicano il rapporto GPPy / GPPy_{media}.

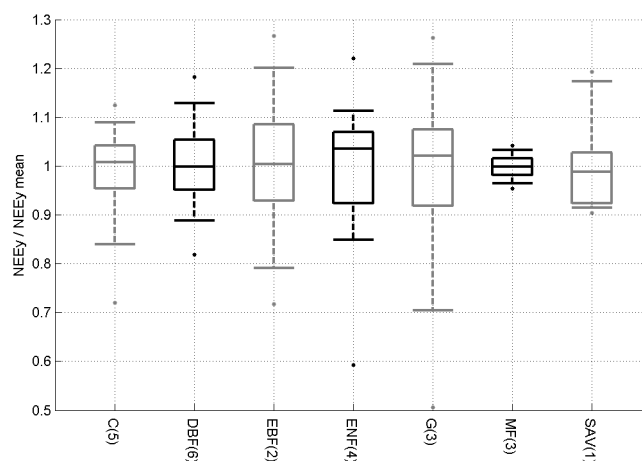


Figura 3.9: PDF dei rapporti tra i valori annuali di NEE e la loro media per ciascuna delle PFT.

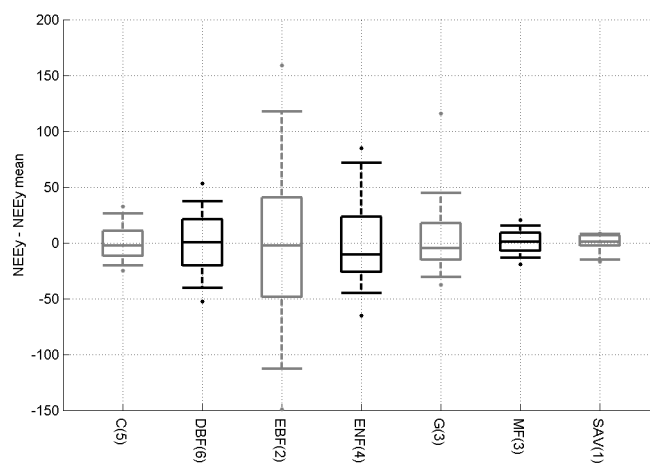


Figura 3.10: PDF delle differenze tra i valori annuali di NEE e la loro media per ciascuna delle PFT.

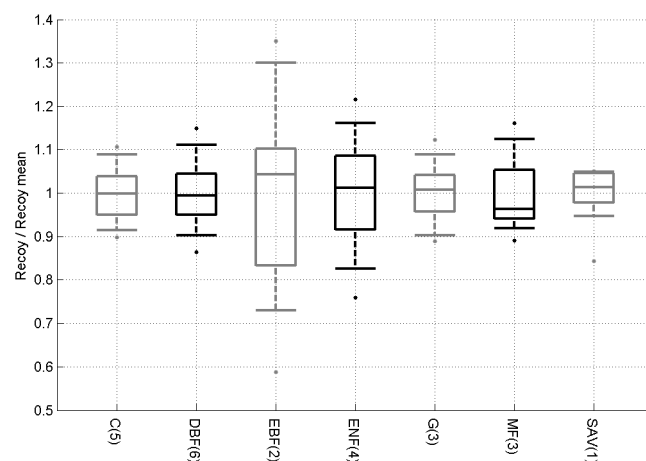


Figura 3.11: PDF dei rapporti tra i valori annuali di Reco e la loro media per ciascuna delle PFT.

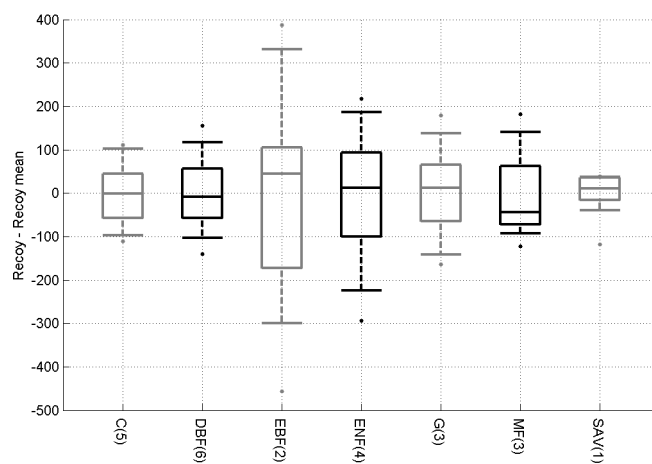


Figura 3.12: PDF delle differenze tra i valori annuali di Reco e la loro media per ciascuna delle PFT.

questo range. Le uniche eccezioni sono mostrate dalle EBF e ENF dove il decimo e 90-esimo percentile sono superiori al range 0.9 - 1.1. Per quanto riguarda l'incertezza quantitativa della GPP (figura 3.14) la distanza interquartile è compresa, come già visto per la RecoY, tra -100 e + 100 gC m⁻² per quasi tutti i siti con l'eccezione delle EBF in cui i valori sono leggermente superiori a questo range. Anche per l'incertezza quantitativa di GPP la maggior variabilità è mostrata per le EBF con valori che vanno da oltre 300 gC m⁻² (90-esimo percentile) a meno di -300 gC m⁻² (10-esimo percentile).

Per i boxplot sulle incertezze quantitative e percentuali di NEEy, RecoY e GPPy è interessante notare che, in alcuni casi, la mediana e la media sono molto diverse infatti, come nel boxplot delle MF mostrato in figura 3.14, la mediana delle differenze è circa 80 gC m⁻² inferiore alla media. Questo risultato potrebbe essere riconducibile alle distribuzioni asimmetriche delle incertezze dei singoli siti/anno compresi nella PFT. Per verificare questa ipotesi si analizzano, ad esempio, le PDF dell'incertezza quantitativa di GPP nelle MF (figura 3.15). Questa PFT comprende un solo sito (Vielsalm, in Belgio) il cui soprassuolo è costituito principalmente da Douglasia e Faggio associati con Abete bianco, Abete rosso, Pino silvestre, Farnia. Le differenze di GPP interannuali potrebbero essere causate da differenze nelle soglie di u^* infatti le medie delle soglie sono 0.3466, 0.3951 e 0.3693 rispettivamente per gli anni 2000, 2006 e 2007. L'impatto delle diverse soglie di u^* potrebbe essere anche al causa delle distanze tra i percentili infatti l'anno con la media delle soglie più alta (2006) mostra il maggior interquartile, l'anno con la media delle soglie più bassa (2000) ha il minor interquartile mentre quello con la media delle soglie intermedia ha un interquartile intermedio agli altri due. La marcata asimmetria nelle distribuzioni delle differenze di GPPy, soprattutto per gli anni 2000 e 2006, si riflette nell'asimmetria mostrata in figura 3.13. In questo sito, non essendo stati registrati né particolari cambiamenti di natura ecologica né interventi di origine antropica, le differenze interannuali potrebbero essere giustificate quanto riportato in precedenza. Analizzando anche le PDF per i siti classificati come C è possibile estrarre ulteriori conclusioni. In figura 3.14 la distribuzione dell'incertezza quantitativa sembra perfettamente simmetrica quindi, per quanto ipotizzato per il sito di Vielsalm, si potrebbe ipotizzare che le PDF per i siti/anno classificati

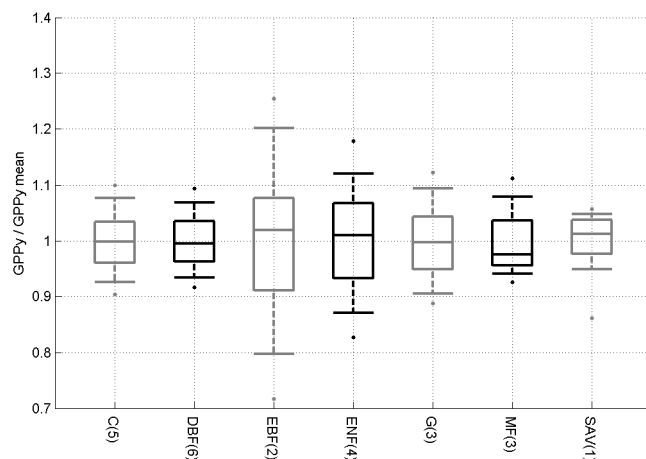


Figura 3.13: PDF dei rapporti tra i valori annuali di GPP e la loro media per ciascuna delle PFT.

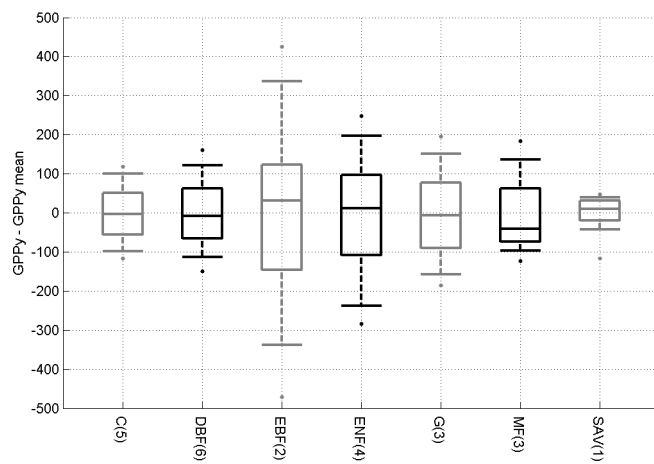


Figura 3.14: PDF delle differenze tra i valori annuali di GPP e la loro media per ciascuna delle PFT.

	2004	2005
temperatura media annua ($^{\circ}\text{C}$)	9.5	9.58
Precipitazione annua (mm/anno)	506.6	521.9
Spessore strato nevoso (m)	0.04	0.05

Tabella 3.5: Caratteristiche climatiche del sito di Gebesee.

come C siano tendenzialmente simmetriche. Analizzando queste distribuzioni in dettaglio (figura 3.16) si nota che la simmetria della PDF della figura 3.14) è dovuta ad un effetto di annullamento tra i percentili che hanno valori simili ma segno opposto. In sintesi anche se le PDF dei siti/anno mostrano delle distribuzioni asimmetriche ma valori assoluti simili, l'aggregazione per PFT ed il calcolo dei percentili risulta essere simmetrica. Per valutare l'incertezza interannuale per la PFT C si analizzeranno nel dettaglio gli anni 2004 e 2005 del sito di Gebesee. Dal punto di vista climatico questi due anni sono abbastanza simili (tabella 3.5) mentre per il soprassuolo il discorso è molto diverso. Dal 22 agosto 2003 al 3-5 agosto 2004 il soprassuolo era costituito da colza mentre dal 16 settembre 2004 al 16 luglio 2005 il suolo era coperto da orzo invernale. Nel periodo tra le due colture il terreno ha subito solo delle lavorazioni preparatorie (aratura e scarificazione). Le coltivazioni sono, in entrambe gli anni considerati, sottoposte a interventi di diserbo e prodotti fitosanitari. Il cambiamento della copertura vegetale, in questo caso, potrebbe essere la causa delle differenze nelle medie delle soglie di u^* , 0.1735 m s^{-1} per l'anno 2004 e 0.1288 m s^{-1} per il 2005. La specie vegetale che costituisce il soprassuolo e gli interventi antropici potrebbero spiegare la notevole differenza interannuale della GPP ed anche la forma della PDF (figura 3.16).

Per avere un valore generale dell'incertezza del post processamento, i valori annuali delle differenze e dei rapporti di tutte le PFT sono stati uniti e le PDF riportata nelle figure 3.17(a) e 3.17(b).

Le PDF della figura 3.17(a) mostrano che i rapporti dei diversi flussi medi annuali con quelli di ogni schema di post processamento possono variare da 0.7 a 1.2. Dei tre flussi quello che manifesta la maggior variabilità è NEE mentre quello con la minor divergenza tra i percentili è GPP. Questo risul-

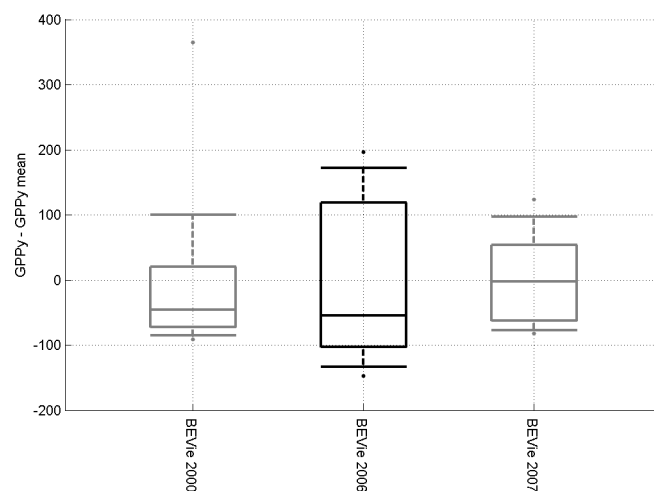


Figura 3.15: PDF delle differenze tra i valori annuali di GPP e la loro media per le foreste miste.

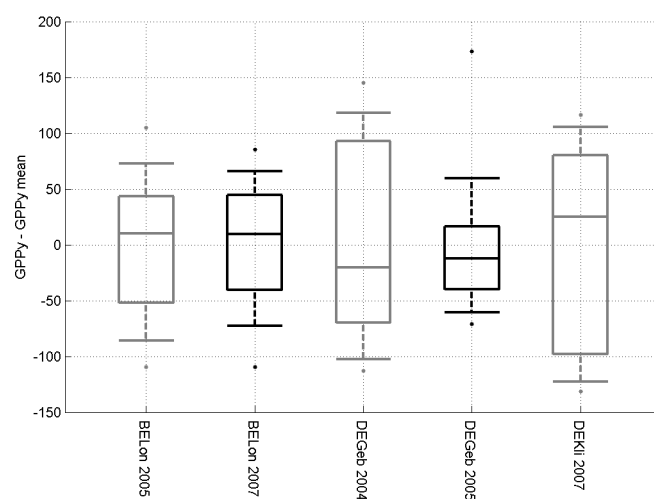
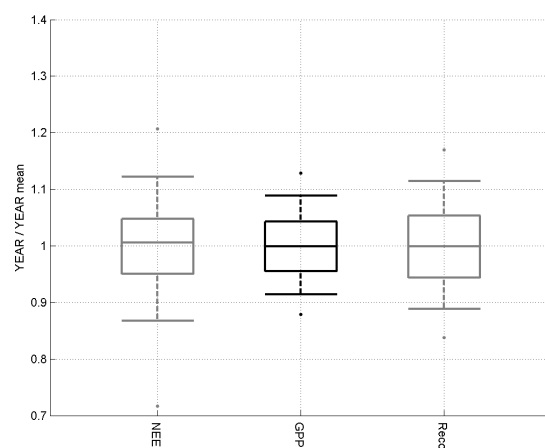


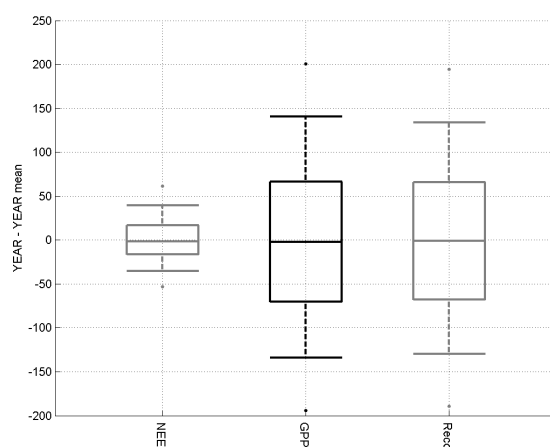
Figura 3.16: PDF delle differenze tra i valori annuali di GPP e la loro media per le campi coltivati.

tato potrebbe suggerire che le incertezze dei flussi di NEE ed Reco hanno una direzione simile ma, calcolando la GPP come differenza, l'incertezza complessiva si riduce. Analizzando la forma delle distribuzioni sembra che il bilancio tra NEE e Reco sia principalmente guidato dalla respirazione infatti la PDF della GPP ha una forma più simile a quella della respirazione che della NEE.

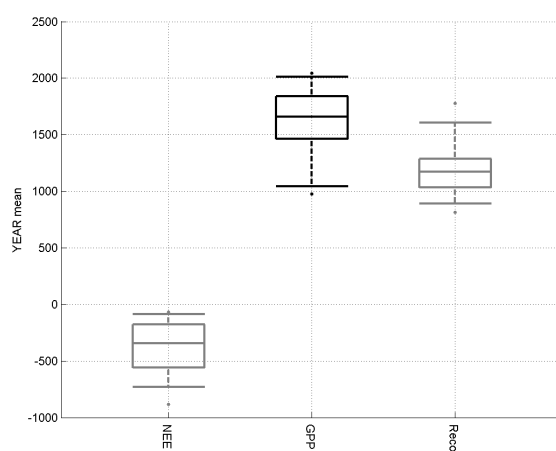
La figura 3.17(b) mostra l'incertezza dei flussi annuali come differenze valore di ogni schema di post processamento e il valore medio di tutti gli schemi. A differenza della figura precedente la PDF della NEE mostra la minor distanza tra percentili mentre i valori massimi di differenze sono quelle della GPP. Dal punto di vista quantitativo la NEE mostra valori che variano tra $\pm 50 \text{ gC m}^{-2}$ mentre per Reco e GPP le differenze sono intorno a $\pm 150 \text{ gC m}^{-2}$. Questo risultato, in apparente contrasto con quello della figura 3.17(a), in realtà può essere giustificato considerando il termine di paragone adottato, cioè il valore medio annuale. La figura 3.17(c) mostra le PDF dei valori medi adottati per la normalizzazione dei valori annuali. Tutti i valori medi di NEE sono compresi tra -60 e -900 gC m^{-2} mentre per GPP ed Reco le medie sono molto più alte. Per la Reco valori medi sono compresi tra +800 e +1800 gC m^{-2} mentre per GPP le medie hanno un range tra +1000 e +2000 gC m^{-2} . Questa notevole differenza di valori medi ha un notevole impatto quando vengono calcolati i rapporti. Ad esempio le differenze di $\pm 50 \text{ gC m}^{-2}$ per i flussi di NEE, confrontate con valori medi che oscillano tra -60 e -900 gC m^{-2} , hanno sicuramente un peso percentuale superiore rispetto alle differenze di GPP ($\pm 150 \text{ gC m}^{-2}$) rapportate a flussi medi che variano tra +1000 e +2000 gC m^{-2} .



(a)



(b)



(c)

Figura 3.17: PDF dei valori medi di tutti gli schemi di post processamento. (a) rapporto tra valori annuali e la loro media; (b) differenza tra valori annuali e la loro media; (c) valori medi annuali.

Capitolo 4

Conclusioni

In questo lavoro è stata utilizzata una serie di schemi di post processamento per calcolare l'incertezza dei dati EC. Le differenze tra gli schemi di post processamento possono essere di natura strettamente statistica, come ad esempio nella scelta di una diversa soglia di despiking, o di tipo teorico, ad esempio il metodo di gapfilling o partitioning.

Dato che ogni opzione dei diversi passaggi di post processing ha solide basi teoriche, lo schema di elaborazione è una scelta legata solamente all'utilizzatore. Non potendo decidere a priori quale delle combinazioni di opzioni di elaborazione sia più corretta, le differenze dei risultati del post processamento sono da considerare ed interpretare come incertezza.

Dato che l'incertezza definisce il range di valori possibili che una quantità può assumere, oltre alla conoscenza dell'origine dell'incertezza è fondamentale dare una stima quantitativa della sua entità. Dai risultati ottenuti, oltre a definire le principali cause di incertezza del post processing dei dati EC è stato possibile attribuire, a ogni fattore e loro combinazioni, un peso relativo rispetto all'incertezza globale. Considerando le numerose applicazioni che i dati EC hanno in ambito ecologico e, soprattutto, nella stima del bilancio del carbonio, definire le cause dell'incertezza e quantificarne l'entità è un aspetto fondamentale. Conoscere l'incertezza di un valore consente di definire il livello di confidenza con cui deve essere utilizzato. Ad esempio conoscendo l'incertezza dei valori utilizzati per parametrizzare un modello, il livello di confidenza dei parametri sarà in funzione dell'incertezza dei dati

iniziali. L'influenza dell'incertezza, nel caso dei modelli, non si esaurisce con la fase di parametrizzazione ma deve essere considerata anche nella fase applicativa e di validazione. Per stimare l'influenza dell'incertezza delle misure EC, dovuta ai diversi schemi di post processamento, sulla parametrizzazione è stato utilizzato un semplice modello (curva di luce). I risultati mostrano che la scelta dello schema di post processamento causa una notevole incertezza dei parametri del modello.

A differenza dei molti studi sull'incertezza presenti in letteratura (Elbers *et al.*, 2011; Richardson & Hollinger, 2007; Rannik *et al.*, 2006; Baldocchi, 2003), in questo studio è stato adottato un approccio di tipo fattoriale. L'applicazione dell'analisi fattoriale ha il notevole vantaggio di poter stimare l'effetto che i singoli fattori e le loro interazioni hanno sull'incertezza globale. Un altro vantaggio dell'analisi fattoriale è quello di poter attribuire, a ogni fattore ed interazione, non solo un effetto qualitativo ma, soprattutto, un valore quantitativo. Quantificando l'incertezza è stato possibile definire quali passaggi del post processamento sono più influenti sull'incertezza globale. Per la stima dell'incertezza si è preferito adottare un modello additivo. I risultati del modello additivo mostrano che, per le serie temporali analizzate, il fattore che maggiormente contribuisce all'incertezza complessiva è la definizione della soglia di u^* . Questo primo risultato sembra confermare quanto affermato da Papale *et al.* (2006) e da Elbers *et al.* (2011). Oltre alla soglia di u^* anche il partitioning influenza notevolmente la GPP.

Un altro fattore che sembra avere un notevole impatto sull'incertezza globale è quello del gap filling. In realtà, come dimostrano i risultati delle incertezze intrinseche, l'incertezza del gap filling è funzione della soglia di u^* . La soglia di u^* è il fattore che elimina il maggior numero di dati considerati non buoni e, di contro, è anche il fattore che riduce maggiormente il numero di dati che possono essere utilizzati come riferimento per i metodi di gap filling.

In sintesi i risultati ottenuti indicano che per la stima dell'incertezza del post processamento l'attenzione deve essere rivolta soprattutto ad alcune fasi dell'elaborazione: definizione della soglia di u^* ; metodi di gap filling; metodi di partitioning.

La scelta di uno schema di processamento e, soprattutto, dei passaggi più critici, è fondamentale anche per gli studi sul confronto tra diversi ecosistemi

e sulla variabilità interannuale. Ad esempio il confronto della Reco tra siti che utilizzano metodi diversi di partitioning deve essere fatto con molta attenzione infatti le differenze e analogie potrebbe essere causate non solo da fattori ecologici ma dai metodi adottati per la separazione dei flussi. Come consigliato da Papale *et al.* (2006) per evitare problemi di questo genere è necessario standardizzare le fasi di elaborazione.

Per avere dei valori indicativi dell'incertezza totale legata ai diversi schemi di post processamento, sono stati confrontati, per ogni PFT, i valori annuali dei diversi schemi di post processamento con il loro valore medio, rispettivamente, per NEE, Reco e GPP. I risultati mostrano che i valori delle medie annuali di NEE sono compresi tra -60 e -900 gC m^{-2} mentre per GPP ed Reco variano, rispettivamente, tra $+800$ e $+1800 \text{ gC m}^{-2}$ e $+1000$ e $+2000 \text{ gC m}^{-2}$. Le differenze tra i valori annuali e le medie per i flussi di NEE sono pari a $\pm 50 \text{ gC m}^{-2}$ mentre per Reco e GPP sono intorno a $\pm 150 \text{ gC m}^{-2}$. Questi risultati sembrano confermare quanto riportato da diversi autori (Elbers *et al.*, 2011; Richardson & Hollinger, 2007; Rannik *et al.*, 2006; Papale *et al.*, 2006; Baldocchi, 2003).

I risultati dimostrano che la scelta di un diverso schema di post processamento dei dati EC ha un notevole impatto sull'incertezza globale e sul bilancio del carbonio.

I risultati presentati quantificano l'incertezza dei dati EC considerando i passaggi del post processamento. In realtà l'incertezza dei dati EC è costituita anche da altre componenti: (i) errori casuali nella misura dei flussi causati dai sistemi di misura e dal trasporto turbolento; (ii) variazioni della footprint, soprattutto in ecosistemi con disomogeneità spaziale; (iii) errori per limitazioni strumentali (ad esempio frequenza di acquisizione, separazione dei sensori, etc); (iv) scelta degli schemi di processamento dei dati ad alta frequenza.

Per avere una stima dell'incertezza complessiva delle misure EC sarebbe necessario effettuare uno studio integrato di tutti questi fattori. Conoscendo l'incertezza complessiva potrebbe essere definita una configurazione strumentale ed uno schema di elaborazione standard per tutti i dati EC.

Bibliografia

- Anderson, DE, Verma, SB, & Rosenberg, NJ. 1984. Eddy correlation measurements of co₂, latent heat and sensible heat fluxes over crop surface. *Boundary layer meteorology*, 167–183.
- Anderson, DE, Verma, SB, Clement, RE, *et al.* 1986a. Turbulence spectra of co₂, water vapour, temperature and velocity over a deciduous forest. *Agricultural and forest meteorology*, 81–99.
- Anthoni, P.M., Freibauer, A., Kolle, O., & Schulze, E.D. 2004. Winter wheat carbon exchange in thuringia, germany. *Agricultural and forest meteorology*, 55–67.
- Aubinet, M. 2008. Eddy covariance co₂ flux measurements in nocturnal conditions: an analysis of the problem. *Ecological applications*, 1368–1378.
- Aubinet, M., Grelle, A., Ibrom, A., Rannik, U., Moncrieff, J., Foken, T., Kowalski, P., Martin, P., Berbigier, P., Bernhofer, Ch., Clement, R., Elbers, J., Granier, A., Grunwald, T., Morgenster, K., Pilegaard, K., Rebmann, C., W., Snijders, Valentini, R., & Vesala, T. 2000. Estimates of the annual net carbon and water exchange of european forests: The euroflux methodology. *Adv. ecol. res.*, 113–174.
- Aubinet, M., Chermanne, B., Vandenhaute, M., Longdoz, B., Yernaux, M., & Laitat, E. 2001. Long term carbon dioxide exchange above a mixed forest in the belgian ardennes. *Agricultural and forest meteorology*, 293–315.
- Aubinet, M., Heinesch, B., & Yernaux, M. 2003. Horizontal and vertical co₂ advection in a sloping forest. *Boundary layer meteorology*, 397–417.

- Aubinet, M., Berbigier, P., Bernhofer, Ch., A., Cescatti, Feigenwinter, Ch., *et al.* 2005. Comparing co₂ storage and advection conditions at night at different carboeuroflux sites. *Bound. layer meteorol.*, 63–94.
- Aubinet, M., Vesala, T., & Papale, D. 2012. *Eddy covariance a practical guide to measurement and data analysis*. Springer.
- Baldocchi, D. 2008. Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian journal of botany*, 1–26.
- Baldocchi, D, Falge, E, & Wilson, K. 2001b. A spectral analysis of biosphere-atmosphere trace gas flux densities and meteorological variables across hour to multi-year time scales. *Agricultural and forest meteorology*, 1–27.
- Baldocchi, D.D. 1997. Measuring and modeling carbon dioxide and water vapor exchange over a temperate broad-leaved forest during the 1995 summer drought. *Plant cell environ.*, 1108–1122.
- Baldocchi, DD., Hicks, BB., & Meyers, TP. 1988. Measuring biosphere-atmosphere exchange of biologically related gas with micrometeorological methods. *Ecology*, 1331–1340.
- Baldocchi, DD., Falge, E., Gu, L., *et al.* 2001. Fluxnet: a new tool to study the temporal and spatial variability of ecosystem scale carbon dioxide, water vapour and energy flux densities. *Bulletin of american meterological society*, 2415–2434.
- Baldocchi, Dennis, Finnigan, John, Wilson, Kell, U, K.T. Paw, & Falge, Eva. 2000. On measuring net ecosystem carbon exchange over tall vegetation on complex terrain. *Boundary-layer meteorology*, 257–291.
- Baldocchi, Dennis D. 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange of ecosystem: past, present and future. *Global change biology*, 479–492.
- Barford, C.C., Wofsy, S.C., GouldenL., M., Munger, J.W., Pyle, E.H., Urbanski, S.P., Hutyra, L., Saleska, S.R., Fitzjarrald, D., & Moore, K. 2001.

- Factors controlling long- and short-term sequestration of atmospheric co₂ in a mid-latitude forest. *Science*, 1688–1691.
- Barr, A.G., Griffis, T.J., Black, T.A., Lee, X., Staebler, R.M., Fuentes, J.D., Chen, Z., & Morgenstern, K. 2002. Comparing the carbon budgets of boreal and temperate deciduous forest stands. *Can. j. forest res.*, 813–822.
- Barr, A.G., Black, T.A., Hogg, E.H., Kljun, N., Morgenstern, K., & Nesic, Z. 2004. Interannual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. *Agricultural and forest meteorology*, 237–255.
- Berthouex, Paul Mac, & Brown, Linfield C. 2002. *Statistics for environmental engineers*. Second edition edn. Lewis Publishers.
- Black, T.A., Chen, W.J., Barr, A.G., Arain, M.A., Chen, Z., Nesic, Z., Hogg, E.H., Neumann, H.H., & Yang, P.C. 2000. Increased carbon sequestration by a boreal deciduous forest in years with a warm spring. *Geophys. res. lett.*, 1271–1274.
- Braswell, B.H., Sacks, B., Linder, E., & Schimel, D.S. 2005. Estimating ecosystem process parameters by assimilation of eddy flux observations of net. *Global change biology*, 335–355.
- Desai, A.R., Bolstad, P., Cook, B.D., Davis, K.J., & Carey, E.V. 2005. Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper midwest, usa. *Agricultural and forest meteorology*, 33–55.
- Desai, A.R., Richardson, A.D., Moffat, A.M., Kattge, J., Hollinger, D.Y., Barr, A., Falge, E., Noormets, A., Papale, D., Reichstein, M., & Stauch, V.J. 2008. Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. *Agricultural and forest meteorology*, 821–838.
- Dolman, A.J., Moors, E.J., & Elbers, J.A. 2002. The carbon uptake of a mid latitude pine forest growing on sandy soil. *Agric. and forest. met.*, 157–170.
- Efron, B., & Tibshirani, R.J. 1993. *An introduction to the bootstrap*.

- Elbers, Jan A., Jacobs, Cor M.J., Kruijt, Bart, Jans, Wilma W.P., & Moors, Eddy J. 2011. Assessing the uncertainty of estimated annual totals of net ecosystem productivity: A practical approach applied to a mid latitude temperate pine forest. *Agricultural and forest meteorology*, 1823–1830.
- Esen, H., Ozgen, F., Esen, M., & Sengur, A. 2009. Artificial neural network and wavelet neural network approaches for modelling of a solar air heater. *Expert systems with applications*, 11240–11248.
- Falge, E., Baldocchi, D.D., Olson, R., *et al.* 2001a. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and forest meteorology*, 43–69.
- Falge, E., Baldocchi, D.D., Olson, R., *et al.* 2001b. Seasonality of ecosystem respiration and gross primary production as derived from fluxnet measurements. *Agricultural and forest meteorology*, 43–69.
- Feigenwinter, Ch., Bernhofer, Ch., & Vogt, R. 2004. The influence of advection on the short term CO₂ budget in and above a forest canopy. *Bound. layer meteorol.*, 201–224.
- Finnigan, J. 1999. A comment on the paper by Lee (1998): on micrometeorological observations of surface-air exchange over tall vegetation. *Agricultural and forest meteorology*, 55–64.
- Flanagan, L.B., Wever, N.E., & Carlson, P.J. 2002. Seasonal and interannual variation in the carbon dioxide exchange and carbon balance in a northern temperate grassland. *Global change biology*, 599–615.
- Foken, T.H., & Wichura, B. 1995. Tools for quality assessment of surface-based flux measurements. *Agricultural and forest meteorology*, 83–105.
- Fox, A., *et al.* 2009. The reflex project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and forest meteorology*, 1597–1615.
- Freund, Rudolf J., & Wilson, William J. 2003. *Statistical methods*. Second edition edn. Academic Press.

- Garbulsky, M.F., Penuelas, J., & adn I. Filella, D. Papale. 2008. Remote estimation of carbon dioxide uptake by a mediterranean forest. *Global change biology*, 2860–2867.
- Geider, R.J., Delucia, E.H., *et al.* 2001. Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. *Global change biology*, 849–882.
- Gilmanov, T.G., Johnson, D.A., & Saliendra, N.Z. 2003. Growing season co₂ fluxes in a sagebrush-steppe ecosystem in idaho: bowen ratio/energy balance measurements and modeling. *Basic and applied ecology*, 167–183.
- Gilmanov, T.G., Soussana, J.E., Aires, L., *et al.* 2007. Partitioning european grassland net ecosystem co₂ exchange into gross primary productivity and ecosystem respiration using light response function analysis. *Agriculture ecosystems and environment*, 93–120.
- Göckede, M., Foken, T., Aubinet, M., Aurela, M., Banza, J., Bernhofer, C., Bonnefond, J.-M., Brunet, Y., Carrara, A., Clement, R., Dellwik, E., Elbers, J., Eugster, W., Fuhrer, J., Granier, A., Grünwald, T., Heinesch, B., Janssens, I., Knohl, A., Koeble, R., Laurila, T., Longdoz, B., Manca, G., Marek, M., Markkanen, T., Mateus, J., Matteucci, G., Mauder, M., Migliavacca, M., Minerbi, S., Moncrieff, J., Montagnani, L., Moors, E., Ourcival, J.M., Papale, D., Pereira, J., Pilegaard, K., Pita, G., Rambal, S., Valentini, R., Vesala, T., Verbeck, H., & Yakir, D. 2008. Quality control of carboeurope flux data. 1. coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems. *Biogeosciences*, 433–450.
- Goulden, M.L., Munger, J.W., Fan, S.M., Daube, B.C., & Wofsy, S.C. 1996a. Exchange of carbon dioxide by a deciduous forest: Response to interannual climate variability. *Science*, 1576–1578.
- Goulden, M.L., Munger, J.W., Fan, S.M., Daube, B.C., & Wofsy, S.C. 1996b. Measurement of carbon storage by long-term eddy correlation: Methods and a critical assessment of accuracy. *Global change biology*, 169–182.

- Gove, J.H., & Hollinger, D.Y. 2006. Application of a dual unscented kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. *J. geophys. res.*
- Gu, Lianhong, Falge, Eva M., Bodena, Tom, Baldocchi, Dennis D., Black, T.A., Saleska, Scott R., Suni, Tanja, Verma, Shashi B., Vesala, Timo, Wofsy, Steve C., & Xu, Liukang. 2005. Objective threshold determination for nighttime eddy flux filtering. *Agricultural and forest meteorology*, 179–197.
- Hagen, S.C., Braswell, B.H., Linder, E., Frolking, S., Richardson, A.D., & Hollinger, D.Y. 2006. Statistical uncertainty of eddy flux-based estimates of gross ecosystem carbon exchange at howland forest, maine. *J. geophys. res.*, doi:10.1029/2005JD006154.
- Hanan, NP, Burba, G., Verma, S., *et al.* 2002. Inversion of net ecosystem co2 flux measurements for estimation of canopy par absorption. *Global change biology*, 563–574.
- Hanson, PJ, Amthor, JS, Wullschleger, SD, *et al.* 2004. Carbon and water cycle simulations for an upland oak forest using 13 stand-level models: intermodel comparisons and evaluations against independent measurements. *Ecological monographs*, 443–489.
- Hofmeister, H. 1997. *Lebensraum wald. pflanzengesellschaften und ihre Ökologie.*
- Hollinger, D.Y., & Richardson, A.D. 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree physiology*, 873–885.
- Hollinger, D.Y., Abel, J., Dail, B., Davidson, S.M., Goltz, H., & Hughes, H. 2004. Spatial and temporal variability in forest-atmosphere co2 exchange. *Global change biology*, 1689–1706.
- Hui, D, Wan, S, Su, B, *et al.* 2003. Gap-filling missing data in eddy covariance measurements by multiple imputation (mi) for annual estimations. *Agricultural and forest meteorology*, 93–111.

- Hui, D., Wan, S., Su, B., Katul, G., Monson, R., & Luo, Y. 2004. Gapfilling missing data in eddy covariance measurements using multiple imputation (mi) for annual estimations. *Agricultural and forest meteorology*, 93–111.
- Janssens, I.A., Lankreijer, H., Matteucci, G., *et al.* 2001. Productivity overshadows temperature in determining soil and ecosystem respiration across european forests. *Global change biology*, 269–278.
- Jarvis, P.G., Massheder, J.M., Hale, S.E., Moncrieff, J.B., Rayment, M., & Scott, S.L. 1997. Seasonal variation of carbon dioxide, water vapor, and energy exchanges of a boreal black spruce forest. *J. geophys. res. atmospheres*, **D24**(102).
- Jung, Martin, Reichstein, Markus, Ciais, Philippe, Seneviratne, Sonia I., Sheffield, Justin, Goulden, Michael L., Bonan, Gordon, Cescatti, Alessandro, Chen, Jiquan, de Jeu, Richard, JohannesDolman, A., Eugster, Werner, Gerten, Dieter, Gianelle, Damiano, Gobron, Nadine, Heinke, Jens, Kimball, John, Law, Beverly E., Montagnani, Leonardo, Mu, Qiaozhen, Mueller, Brigitte, Oleson, Keith, Papale, Dario, Richardson, Andrew D., Rouspard, Olivier, Running, Steve, Tomelleri, Enrico, Viovy, Nicolas, Weber, Ulrich, Williams, Christopher, Wood, Eric, Zaehle, Sönke, & Zhang, Ke. 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 951–954.
- Knohl, Alexander, Schulze, Ernst-Detlef, Kolle, Olaf, & Buchmann, Nina. 2003. Large carbon uptake by an unmanaged 250-year-old deciduous forest in central germany. *Agricultural and forest meteorology*, 151–167.
- Knorr, W., & Kattge, J. 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by monte carlo sampling. *Global change biology*, 1333–1351.
- Körner, C. 1995. *Ecophysiology of photosynthesis*. eds schulze e-d, caldwell mm edn. Chap. Leaf diffusive conductances in the major vegetation types of the globe, pages 463–490.

- Kowalski, AS, Serrano-Ortiz, P, Janssens, IA, *et al.* 2008. Can flux tower research neglect geochemical co₂ exchange? *Agricultural and forest meteorology*, 1045–1054.
- Lasslop, Gitta, Reichstein, Markus, Papale, Dario, Richardson, Andrew D., Arneth, Almut, Barr, Alan, Stoy, Paul, & Wohlfahrt, Georg. 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global change biology*, 187–208.
- Law, BE, Thornton, P, Irvine, J, *et al.* 2001b. Carbon storage and fluxes in ponderosa pine forests at different developmental stages. *Global change biology*, 755–777.
- Lee, X. 1998. On micrometeorological observations of surface-air exchange over tall vegetation. *Agric. for. meteorol.*, 39–50.
- Lee, X., Wu, H.-J., Sigler, J., Oishi, C., & Siccama, T. 2004. Rapid and transient response of soil respiration to rain. *Global change biology*, 1017–1026.
- Lenschow, DH. 1995. *biogenic trace gases: measuring emissions from soil and water*. Matson pa and harriss rc edn. Chap. Micrometeorological techniques for measuring biosphere-atmosphere trace gas exchange, pages 126–163.
- Lloyd, J, & Taylor, JA. 1994. On the temperature dependence of soil respiration. *Functional ecology*, 315–323.
- Lund, R., & Reeves, J. 2002. Detection of undocumented change-points: A revision of the two-phase regression model. *J. climate*, 2547–2554.
- Luyssaert, S., Inglima, I., Jung, M., Richardson, A.D., Reichstein, M., Papale, D., Piao, S.L., Schulze, E.D., Wingate, L., Matteucci, G., Aragao, L., Aubinet, M., Beer, C., Bernhofer, C., Black, K.G., Bonal, D., Bonnefond, J.M., Chambers, J., Ciais, P., Cook, B., Davis, K.J., Dolman, A.J., Gielen, B., Goulden, M., Grace, J., Granier, A., Grelle, A., Griffis, T., Grünwald, T., Guidolotti, G., Hanson, P.J., Harding, R., Hollinger, D.Y.,

- Hutyra, L.R., Kolari, P., Kruijt, B., Kutsch, W., Lagergren, F., Laurila, T., Law, B.E., Maire, G. Le, Lindroth, A., Loustau, D., Malhi, Y., Mateus, J., Migliavacca, M., Misson, L., Montagnani, L., Moncrieff, J., Moors, E., Munger, J.W., Nikinmaa, E., Ollinger, S.V., Pita, G., Rebmann, C., Rouspard, O., Saigusa, N., Sanz, M.J., Seufert, G., Sierra, C., Smith, M.L., Tang, J., Valentini, R., Vesala, T., & Janssens, I.A. 2007. Co2 balance of boreal, temperate, and tropical forests derived from a global database. *Global change biology*, 2509–2537.
- Luyssaert, Sebastiaan, Schulze, E.-Detlef, Börner, Annett, Knohl, Alexander, Hessenmöller, Dominik, Law, Beverly E., Ciais, Philippe, & Grace, John. 2008. Old-growth forests as global carbon sinks. *Nature*, 213–215.
- Manson, Robert, F., L. Gunst Richard, & L., Hess James. 2003. *Statistical design and analysis of experiments*. Second edition edn. Wiley Series in Probability and Statistics. John Wiley and Sons.
- Marcolla, B., Cescatti, A., Montagnani, L., Manca, G., Kerschbaumer, G., & Minerbi, S. 2005. Role of advective fluxes in the carbon balance of an alpine coniferous forest. *Agricultural and forest meteorology*, 193–206.
- Massman, WJ., & Lee, X. 2002. Eddy covariance flux corrections and uncertainty in long term studies of carbon and energy exchanges. *Agricultural and forest meteorology*, 121–144.
- Mauder, M., & Foken, T. 2006. Impact of post-field data processing on eddy covariance flux estimates and energy balance closure. *Meteorol z*, 597–609.
- Michaelis, L., & Menten, M.L. 1913. Die kinetik der invertinwirkung. *Biochemische zeitschrift*, 333.
- Miller, S.D., Goulden, L., M., Menton, M.C., da Rocha, H.R., de Freitas, H.C., et al. 2004. Biometric and micrometeorological measurements of tropical forest carbon balance. *Ecological applications*, 114–126.
- Moffat, AM., Papale, D., Reichstein, M., Hollinger, DY., Richardson, AD., Barr, AG., Beckstein, C., Braswell, BH., Churkina, G., Desai, AR., Falge, E., Gove, JH., Heimann, M., Hui, D., Jarvis, AJ., Kattge, J., Noor-

- ments, A., & Stauch, V.J. 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and forest meteorology*, 209–232.
- Mohandes, M., Rehman, S., & Halawani, T.O. 1998. Estimation of global solarradiation using artificial neural-networks. *Renewable energy*, 179–184.
- Moncrieff, J.B., Mahli, Y., & Leuning, R. 1996. The propagation of errors in long term measurements of land atmosphere fluxes of carbon and water. *Global change biology*, 231–240.
- Montgomery, Douglas C., & Runger, George C. 2002. *Applied statistics and probability for engineers*. Third edition edn. United States of America: John Wiley and Sons.
- Moureaux, C., Debacq, A., Bodson, B., Heinesch, B., & Aubinet, M. 2006. Annual net ecosystem carbon exchange by a sugar beet crop. *Agricultural and forest meteorology*, 25–39.
- Noormets, A., Chen, J., & Crow, T.R. 2007. Age-dependent changes in ecosystem carbon fluxes in managed forests in northern wisconsin, usa. *Ecosystems*, 187–203.
- Ooba, M., Hirano, T., Mogami, J.-I., R., R. Hirata, & Fujinumba, Y. 2006. Comparisons of gap-filling methods for carbon flux dataset: a combination of a genetic algorithm and an artificial neural network. *Ecol. modell.*, 473–486.
- Oren, R., Hsieh, C.I., Stoy, P., Albertson, J., McCarthy, H.R., Harrell, P., & Katul, G.G. 2006. Estimating the uncertainty in annual net ecosystem carbon exchange: spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements. *Global change biology*, 883–896.
- Papale, D., & Valentini, R. 2003. A new assessment of european forest carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global change biology*, 525–535.
- Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Longdoz, B., Kutsch, W., Rambal, S., Valentini, R., Vesala, T., & Yakir,

- D. 2006. Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, 571–583.
- Paw, UK, Baldocchi, DD, Meyers, TP, *et al.* 2000. Correction of eddy covariance measurements incorporating both advective effects and density fluxes. *Boundary layer meteorology*, 487–511.
- Pilegaard, K., Mikkelsen, T.N., Beier, C., N.O.Jensen, Ambus, P., & Ro-Poulsen, H. 2003. Field measurements of atmosphere-biosphere interactions in a danish beech forest. *Boreal environment research*, 315–333.
- Rannik, U., Kolari, P., Vesala, T., & Hari, P. 2006. Uncertainties in measurement and modelling of net ecosystem exchange of a forest. *Agricultural and forest meteorology*, 244–257.
- Raupach, MR., *et al.* 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global change biology*, 378–397.
- Rebmann, C, Gockede, M, Foken, T, *et al.* 2005. Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling. *Theoretical and applied climatology*, 121–141.
- Reichstein, M, Tenhunen, JD, Ourcival, J-M, *et al.* 2002b. Severe drought effects on ecosystem co₂ and h₂o fluxes at three mediterranean sites: revision of current hypothesis? *Global change biology*, 999–1017.
- Reichstein, M, Tenhunen, JD, Ourcival, J-M, *et al.* 2003b. Inverse modelling of seasonal drought effects on canopy co₂/h₂o exchange in three mediterranean ecosystems. *Journal of geophysical research*, doi: 10.1029/2003JD003430.
- Reichstein, M., Falge, E., Baldocchi, DD., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila,

- T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., & Valentini, R. 2005. On separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global change biology*, 1424–1439.
- Rey, A., Pegoraro, E., Tedeschi, V., Parri, I.D., Jarvis, P.G., & Valentini, R. 2002. Annual variation in soil respiration and its components in a coppice oak forest in central Italy. *Global change biology*, 851–866.
- Reynolds, O. 1895. On the dynamical theory of incompressible viscous fluids and the determination of criterion. *Philosophical transactions of royal society of London*, 935–982.
- Richardson, A.D., & Hollinger, D.Y. 2007. A method to estimate the additional uncertainty in gap-filled net resulting from long gaps in the CO₂ flux record. *Agricultural and forest meteorology*, 199–208.
- Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Burman, P., Davidson, E.A., Evans, R.S., Flanagan, L.B., Munger, J.W., Savage, K., Urbanski, S.P., & Wofsy, S.C. 2006. Comparing simple respiration models for eddy flux and dynamic chamber data. *Agricultural and forest meteorology*, 219–234.
- Richardson, A.D., Williams, M., Hollinger, D.Y., Moore, D.J., Dail, D.B., Davidson, E.A., Scott, N.A., Evans, R.S., Hughes, H., Lee, J.T., Rodrigues, C., & Savage, K. 2010. Estimating parameters of a forest ecosystem model with measurements of stocks and fluxes as joint constraints. *Oecologia*, 25–40.
- Running, S.W., Baldocchi, D.D., *et al.* 1999. A global terrestrial monitoring network, scaling tower fluxes with ecosystem modelling and EOS satellite data. *Remote sensing environment*, 108–127.
- Sachs, L. 1996. *Angewandte statistik: Anwendung statistischer methoden*.
- Saleska, S.R., Miller, S.D., Matross, D.M., Goulden, L.M., Wofsy, S.C., da Rocha, H.R., de Camargo, P.B., P., Crill, P., Daube, B.C., de Freitas, H.C.,

- Hutyra, L., Keller, M., Kirchhoff, V., Menton, M., Munger, J.W., Pyle, E.H., Rice, A.H., & Silva, H. 2003. Carbon in amazon forests: unexpected seasonal fluxes and disturbance induced losses. *Science*, 1554–1557.
- Savage, K.E., & Davidson, E.A. 2003. A comparison of manual and automated systems for soil co₂ flux measurements: trade-offs between spatial and temporal resolution. *J. exp. botany*, 891–899.
- Schimel, D.S., House, J.I., Hibbard, K.A., *et al.* 2001. Recent patterns and mechanisms of carbon exchange by terrestrial ecosystem. *Nature*, 169–172.
- Schmid, HP. 1994. Source areas for scalar fluxes. *Boundary layer meteorology*, 293–318.
- Schwalm, C.R., Black, T.A., Morgenstern, K., & Humphreys, E.R. 2007. A method for deriving net primary productivity and component respiratory fluxes from tower-based eddy covariance data: a case study using a 17-year data record from a douglas-fir chronosequence. *Global change biology*, 370–385.
- Solow, A.R. 1987. Testing for climate change: An application of the two-phase regression model. *J. climate appl. meteor.*, 1401–1405.
- Staebler, R.M., & Fitzjarrald, D.R. 2004. Observing subcanopy co₂ advection. *Agric.for.meteorol.*, 139–156.
- Stauch, V.J., & Jarvis, A.J. 2006. A semi-parametric model for eddy covariance co₂ flux time series data. *Global change biology*, 1707–1716.
- Suni, T., Rinne, J., Reissell, A., *et al.* 2003. Long-term measurements of surface fluxes above a scots pine forest in hyytiala, southern finland, 1996–2001. *Boreal environment research*, 287–301.
- Tedeschi, V., Rey, A., Manca, G., Valentini, R., Jarvis, P.G., & Borghetti, M. 2006. Soil respiration in a mediterranean oak forest at different developmental stages after coppicing. *Global change biology*, 110–121.
- Trudinger, CM, *et al.* 2007. Optic project: an intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *J. geophys res biogeosci.*

- Valentini, R., Matteucci, G., Dolman, A.J., Schulze, E.D., Rebmann, C., Moors, E.J., Granier, A., Gross, P., Jensen, N.O., Pilegaard, K., Lindroth, A., Grelle, A., Bernhofer, C., Grunwald, T., Aubinet, M., Ceulemans, R., Kowalski, A.S., Vesala, T., Rannik, U., Berbigier, P., Loustau, D., Guomundsson, J., Thorgeirsson, H., Ibrom, A., Morgenstern, K., Clement, R., Moncrieff, J., Montagnani, L., Minerbi, S., & Jarvis, P.G. 2000. Respiration as the main determinant of carbon balance in european forests. *Nature*, 861–865.
- van Gorsel, Eva, Leuning, Ray, Helen, A. Cleugh, Keith, Heater, Miko, U.F. Kirschbaum, & Suni, Tanja. 2007. Nocturnal carbon efflux:reconciliation of eddy covariance and chamber measurements using an alternative to u^* -threshold filtering technique. *Tellus B*, 397–403.
- van Gorsel, Eva, Delpierre, Nicolas, Leuning, Ray, Black, Andy, Munger, J. William, Wofsy, Steven, Aubinet, Marc, Feigenwinter, Christian, Beringer, Jason, Bonal, Damien, Chen, Baozhang, Chen, Jiquan, Clement, Robert, Davis, Kenneth J., Desai, Ankur R., Dragoni, Danilo, Etzold, Sophia, Grünwald, Thomas, Gu, Lianhong, Heinesch, Bernhard, Hutya, Lucy R., Jans, Wilma W.P., Kutsch, Werner, Law, B.E., Leclerc, Monique Y., Mammarella, Ivan, Montagnani, Leonardo, Noormets, Asko, Rebmann, Corinna, & Wharton, Sonia. 2009. Estimating nocturnal ecosystem respiration from the vertical turbulent flux and change in storage of CO_2 . *Agricultural and forest meteorology*, 1919–1930.
- van't Hoff, J.H. 1898. *Lectures on theoretical and physical chemistry. part 1. chemical dynamics*.
- Wang, Q., Tenhunen, J., Falge, E., Bernhofer, C., Granier, A., & Vesala, T. 2004. Simulation and scaling of temporal variation in gross primary production for coniferous and deciduous temperate forests. *Global change biology*, 37–51.
- Wang, X.L. 2003. Comments on “detection of undocumented change-points: A revision of the two-phase regression model”. *J. climate*, 3383–3385.

-
- Wofsy, SC., Goulden, ML., *et al.* 1993. Net exchange of CO₂ in mid-latitude forest. *Science*, 1314–1317.
- Yang, Kai, & Trewn, Jayant. 2004. *Multivariate statistical methods in quality management*. McGraw-Hill Engineering Referecence. McGraw-Hill.
- Zhang, C.L. 2005. Generalized correlation of refrigerant mass flow rate through adiabatic capillary tubes using artificial neural network. *International journal of refrigeration*, 506–514.